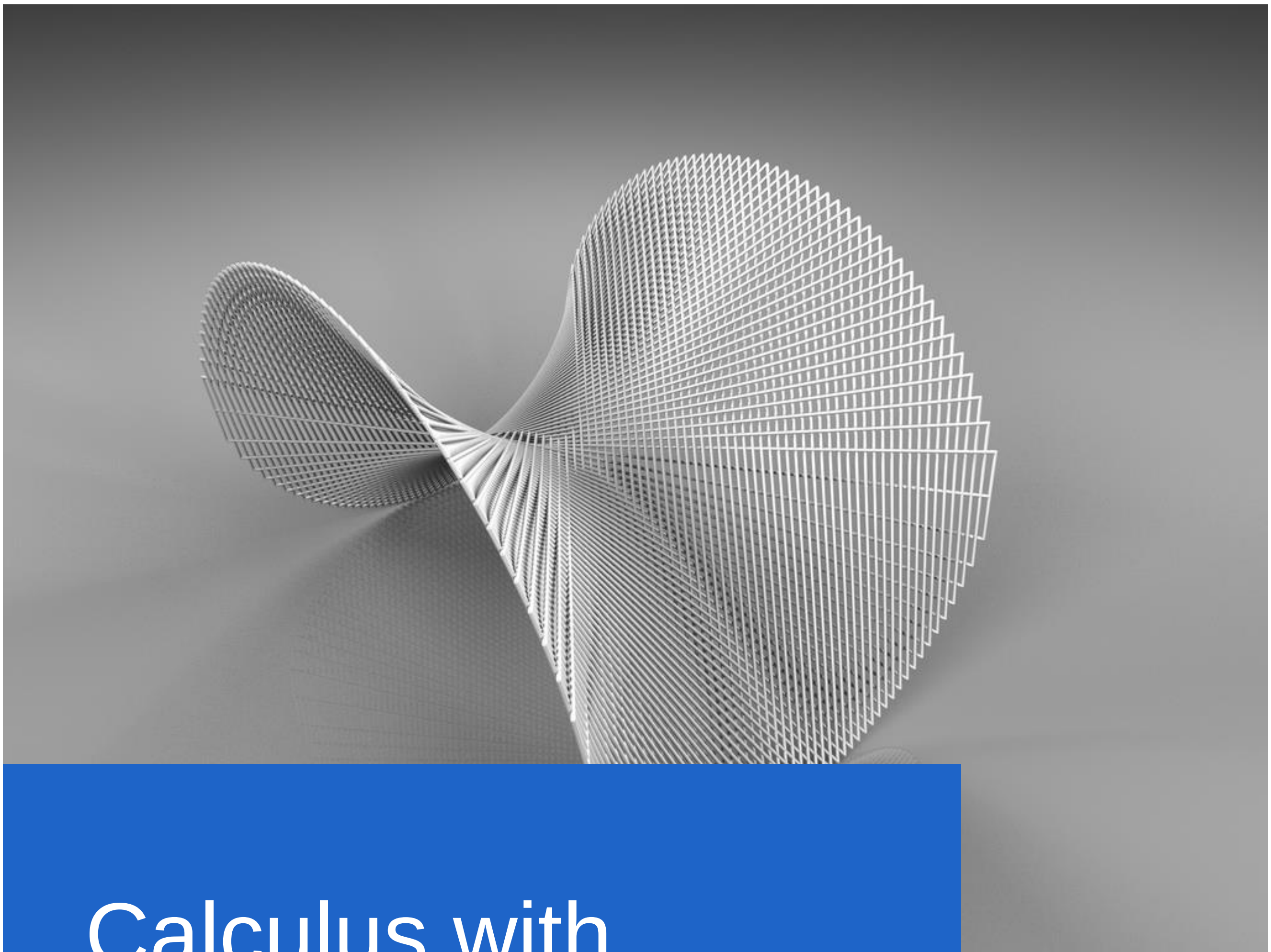
FACULTY OF
BIOSCIENCE ENGINEERING



Calculus with Differential Equations

Prof. dr. ir. Jan Baetens
M.Sc. Elien Van de Walle
M.Sc. An Schelfaut

Academic year 2022–2023 – version 1

GHENT
UNIVERSITY

Calculus with Differential Equations

Prof. dr. ir. Jan M. Baetens
M.Sc. Elien Van de Walle
M.Sc. An Schelfaut

The cover photo represents a hyperbolic paraboloid whose standard equation is given by

$$z = \frac{x^2}{a^2} - \frac{y^2}{b^2}\,.$$

Preface

The purpose of this course is to present mathematics as the science of deductive reasoning and not as the art of manipulation. Unfortunately, many students feel mathematics is incomprehensible and is riddled with complex and abstract jargon. Our goal is to impose a lasting understanding of and appreciation for calculus on the student. Our course is intended to give the student an understanding of what calculus is truly about. It does not take more intelligence than that of a parrot to be able to go through a list of theorems and equations; but only when one understands their origins can one correctly and confidently apply them in the real world.

The over-emphasis on the calculator and foremostly the computer is definitely a point of confusion for the student. The computer is only a time-saving machine whose usefulness depends on the knowledge of the user. We do admit the computer is a remarkable machine, and we will make use of it whenever appropriate, yet it is this fascination that gives students a false sense of what they are doing. The confidence gained from all the correct answers leads to an inseparable dependence where the student is absolutely helpless without it.

Throughout the textbook we constantly refer to science and engineering. The purpose of this is to show how the scientific method applies to all disciplines and to understand that mathematics is an expression of one's observations and hypothesis. For that reason, several examples and exercises were chosen because of their relevance in reality, such that the reader can get a good feel of why and how this course is so important for future engineers. Note that because of its engineering viewpoint, we always indicate the dimensions of the used base quantities, being mass [M], time [T], temperature [Θ] and length [L]. Throughout this course the icon in the margin indicates that there's a supporting You Tube video available. The QR-code below takes you directly to the appropriate You Tube video online. At the end of every chapter one can find an extensive list of exercises linked to the topics discussed in the corresponding chapter.

Even though much time and efforts have been spent in compiling this text, it cannot be free of errors, and the authors would be grateful if these would be reported to them so that the quality of this text can be improved even further.

It goes without saying that many people have contributed to this course in addition to its authors, namely, Demir Ali Köse, Janos Coquyt, Tinne De Boeck, Diego De Gusem, Lander De Visscher, Wannes Dewulf, Jeroen Galle, Jelle Hustinx, Hanna Jaspaert, Linde Lambrecht, Robin Simoens, Ward Van Belle, Caitlin Vanden Bussche, Victor Vanthilt and Hilder Vernieuwe.

Finally, we are grateful that Ben Orlin, author of the book *Math with Bad Drawings* and the blog `https://mathwithbaddrawings.com/`, granted us permission to include his cartoons at the end of some of

the chapters.

Ghent, September 10, 2020

The authors

Contents

Preface iii

1 Introduction 1

I Precalculus 3

2 Sets and numbers 5

2.1 Sets . . . 5

2.1.1 Logic operators . . . 5

2.1.2 Definition and representation of sets . . . 6

2.1.3 Set operations . . . 7

2.1.4 Set properties . . . 9

2.2 The set of real numbers . . . 10

2.2.1 Definition . . . 10

2.2.2 Real number arithmetic . . . 11

2.2.2.1 Addition and multiplication . . . 11

2.2.2.2 Exponentiation . . . 13

2.2.2.3 Arithmetic in $\overline{\mathbb{R}}$ . . . 16

2.2.3 Intervals in $\mathbb{R}$ . . . 17

2.3 Complex numbers . . . 19

2.3.1 Definition . . . 19

2.3.2 Complex number arithmetic . . . 19

2.4 Exercises . . . 22

3 Functions **27**
3.1 The Cartesian coordinate plane 27
3.2 Functions 28
3.2.1 Relations 28
3.2.2 Functions in $\mathbb{R}$ 32
3.2.2.1 Definition 32
3.2.2.2 Function graphs 34
3.2.2.3 Domain, codomain and range 34
3.2.3 Function arithmetic 36
3.2.4 Function properties 40
3.2.4.1 Injections, surjections and bijections 40
3.2.4.2 Symmetry 42
3.2.4.3 Periodicity 43
3.2.4.4 Function behaviour 44
3.2.5 Transformations 46
3.2.6 Piecewise-defined functions 52
3.2.7 Function families 52
3.3 Absolute value functions 54
3.3.1 Definition and properties 54
3.3.2 Absolute value functions 55
3.4 Inverse functions 57
3.4.1 Definition and properties 57
3.4.2 Domain restriction 58
3.5 Exercises 63

4 Algebraic functions **69**
4.1 Polynomial functions 69
4.1.1 Constant and linear functions 69
4.1.2 Quadratic functions 70
4.1.3 General polynomial functions 74
4.1.3.1 Definition 74
4.1.3.2 Factorization of polynomials 74
4.2 Rational functions 78
4.2.1 Definition 78
4.2.2 Graphs of rational functions 80
4.3 Irrational functions 82
4.4 Conic sections 85

4.4.1 Overview 85
4.4.2 Circles 85
4.4.3 Ellipses 87
4.4.4 Parabolas 92
4.4.5 Hyperbolas 96
4.5 Exercises 101

5 Transcendental functions 107
5.1 Definition 107
5.2 Exponential and logarithmic functions 107
5.2.1 Definitions 107
5.2.1.1 Exponential functions 107
5.2.1.2 Logarithmic functions 111
5.2.2 Properties 113
5.2.3 Exponential and logarithmic equations and inequalities 115
5.2.4 Applications 119
5.2.4.1 Growth models 119
5.2.4.2 Newton’s law of cooling 121
5.2.4.3 Radioactive decay 121
5.3 Trigonometric functions 122
5.3.1 Foundations of trigonometry 122
5.3.2 Circular motion 124
5.3.3 The six trigonometric functions 125
5.3.3.1 Cosine and sine on the unit circle 125
5.3.3.2 Beyond the unit circle 127
5.3.3.3 The other trigonometric functions 129
5.3.4 Trigonometric identities 130
5.3.4.1 Pythagorean identities 130
5.3.4.2 Even/odd, cofunction and difference and sum identities 132
5.3.4.3 Double angle identities and power reduction formulas 135
5.3.4.4 Product to sum formulas and vice versa 136
5.3.5 Graphs of cosine and sine functions 136
5.3.5.1 Domain and range 136
5.3.5.2 Graphs 136
5.3.6 Graphs of the other trigonometric functions 140
5.3.6.1 Domain and range 140
5.3.6.2 Graphs of tangent and cotangent functions 141

5.3.6.3 Applications . . . 144
5.4 Inverse trigonometric functions . . . 146
5.4.1 Inverse cosine and sine functions . . . 146
5.4.2 Inverse tangent and cotangent functions . . . 149
5.4.3 Solving equations involving trigonometric functions . . . 151
5.5 Conic sections continued . . . 157
5.5.1 Translation of axes . . . 157
5.5.2 Rotation of axes . . . 160
5.6 Exercises . . . 164

6 Vector math **169**
6.1 Definition and representation . . . 169
6.2 Vector arithmetic . . . 170
6.2.1 Addition . . . 170
6.2.2 Scalar multiplication . . . 174
6.2.3 Magnitude and direction . . . 175
6.3 Unit vectors . . . 177
6.4 The dot product . . . 178
6.5 Vectors in three-dimensional space and beyond . . . 181
6.5.1 Cartesian coordinates and distance in space . . . 181
6.5.2 Vector representation and operations in space . . . 182
6.6 The cross product . . . 184
6.7 Exercises . . . 188

7 Three-dimensional analytical geometry **191**
7.1 Lines . . . 191
7.1.1 Definition . . . 191
7.1.2 Relative position of lines . . . 194
7.2 Planes . . . 196
7.2.1 Definition . . . 196
7.3 Three-dimensional objects . . . 201
7.3.1 Spheres and cylinders . . . 201
7.3.2 Quadratic surfaces . . . 203
7.4 Exercises . . . 210

II Single variable calculus 215

8 Limits and continuity 217

8.1 An intuitive introduction . . . 218
8.1.1 Limits and their approximation . . . 218
8.1.2 Existence of limits . . . 219
8.1.3 Limits of difference quotients . . . 221
8.2 Epsilon-delta definition of a limit . . . 223
8.3 Finding limits analytically . . . 224
8.3.1 Properties of limits . . . 224
8.3.2 The squeeze theorem . . . 227
8.3.3 Limits of functions equal at all but one point . . . 230
8.4 One-sided limits . . . 232
8.5 Continuity . . . 236
8.5.1 Definition . . . 236
8.5.2 Intermediate value theorem . . . 239
8.6 Limits involving infinity . . . 241
8.6.1 Limits of infinity and vertical asymptotes . . . 242
8.6.2 Indeterminate forms . . . 245
8.6.3 Limits at infinity and horizontal asymptotes . . . 246
8.6.4 Slant asymptotes . . . 250
8.7 Exercises . . . 253

9 Derivatives and their applications 259

9.1 Definition . . . 259
9.1.1 Intuitive introduction . . . 259
9.1.2 Formalism . . . 261
9.1.3 Differentiability on closed intervals . . . 267
9.1.4 Interpretations of the derivative . . . 269
9.1.4.1 Instantaneous rate of change . . . 269
9.1.4.2 The slope of the tangent line . . . 270
9.1.5 Higher-order derivatives . . . 271
9.1.6 Smoothness . . . 272
9.2 Basic differentiation rules . . . 273
9.2.1 Derivatives of algebraic and transcendental functions . . . 273
9.2.2 Properties of the derivative . . . 275
9.3 The chain rule . . . 278

9.4 Implicit differentiation . . . 282
9.4.1 First derivative . . . 282
9.4.2 Higher-order derivatives . . . 285
9.4.3 Logarithmic differentiation . . . 286
9.5 Derivatives of inverse functions . . . 287
9.6 L'Hôpital's rule . . . 290
9.6.1 Indeterminate forms 0/0 and ∞/∞ . . . 290
9.6.2 Indeterminate forms $0 \cdot \infty$ and $\infty - \infty$ . . . 292
9.6.3 Indeterminate forms 0^0, 1^∞ and ∞^0 . . . 293
9.7 Applications of the derivative . . . 295
9.7.1 Newton's method . . . 295
9.7.2 Differentials . . . 297
9.8 Taylor series . . . 301
9.8.1 Taylor polynomials . . . 301
9.8.2 Taylor's theorem . . . 302
9.8.3 Taylor series . . . 305
9.9 Exercises . . . 311
9.9.1 Analytical exercises . . . 311
9.9.2 Root finding algorithms . . . 315
9.9.2.1 Bisection method . . . 315
9.9.2.2 Newton's method . . . 317

10 The graphical behaviour of functions 319

10.1 Extreme values . . . 319
10.2 The mean value theorem . . . 325
10.2.1 Introduction . . . 325
10.2.2 The theorems . . . 326
10.3 Increasing and decreasing functions . . . 329
10.4 Concavity and the second derivative . . . 336
10.4.1 Concavity . . . 336
10.4.2 The second derivative test . . . 341
10.5 Curve sketching . . . 343
10.6 Optimization . . . 347
10.7 Exercises . . . 353

11 Polar coordinates and parametric equations **359**
11.1 Polar coordinates . . . 359
11.1.1 Definition . . . 359
11.1.2 Linking polar and rectangular coordinates . . . 360
11.1.3 Graphs of polar equations . . . 364
11.2 Parametric equations . . . 371
11.2.1 Definition . . . 371
11.2.2 Graphing parametric equations . . . 371
11.2.3 Parametrising curves . . . 375
11.2.4 Conic sections continued even further . . . 377
11.3 Derivatives and parametric and polar equations . . . 378
11.3.1 Parametric equations . . . 378
11.3.2 Polar equations . . . 381
11.3.3 Smoothness . . . 382
11.3.4 Concavity . . . 384
11.4 Exercises . . . 386

12 Integration **393**
12.1 Antiderivatives and (in)definite integration . . . 393
12.1.1 Antiderivatives and indefinite integration . . . 393
12.1.2 The definite integral . . . 397
12.2 Riemann sums . . . 401
12.2.1 Approximating areas . . . 401
12.2.2 Riemann sums . . . 402
12.2.3 Limits of Riemann sums . . . 405
12.3 The fundamental theorem of calculus . . . 406
12.3.1 Mean value theorem for definite integrals . . . 406
12.3.2 Main theorems . . . 407
12.3.3 Motion and the fundamental theorem of calculus . . . 410
12.3.4 The fundamental theorem of calculus and the chain rule . . . 411
12.3.5 Average value . . . 412
12.4 Techniques of antidifferentiation . . . 413
12.4.1 Substitution . . . 414
12.4.1.1 Rationale . . . 414
12.4.1.2 Integrals involving trigonometric functions . . . 416
12.4.1.3 Integrals leading to inverse trigonometric functions . . . 418
12.4.1.4 Substitution and definite integration . . . 419

12.4.2 Integration by parts . . . 421
12.4.3 Trigonometric integrals . . . 423
12.4.3.1 Integrals of the form $\int \sin^m(x)\cos^n(x)\,dx$ . . . 424
12.4.3.2 Integrals of products of sines and cosines of differing period . . . 427
12.4.3.3 Integrals of the form $\int \tan^m(x)\sec^n(x)\,dx$. . . . 427
12.4.4 Trigonometric substitution . . . 429
12.4.5 Partial fraction decomposition . . . 433
12.5 Improper Integration . . . 439
12.5.1 Improper integrals with infinite bounds . . . 439
12.5.2 Improper integrals with infinite range . . . 442
12.6 Exercises . . . 445
12.6.1 Analytical exercises . . . 445
12.6.2 Numerical integration . . . 449
12.6.2.1 The midpoint method . . . 449
12.6.2.2 Approximation error . . . 451
12.6.2.3 The trapezium method . . . 452
12.6.2.4 Simpson's rule . . . 454

13 Applications of integration 457
13.1 Area between curves . . . 457
13.1.1 Rectangular coordinates . . . 457
13.1.2 Polar coordinates . . . 461
13.2 Volume by cross-sectional area . . . 465
13.2.1 Volumes by slicing . . . 465
13.2.2 Solids of revolution . . . 468
13.3 The shell method . . . 474
13.4 Arc length . . . 480
13.4.1 Rectangular coordinates . . . 480
13.4.2 Parametric and polar equations . . . 483
13.5 Surface area . . . 486
13.5.1 Rectangular coordinates . . . 486
13.5.2 Parametric and polar equations . . . 489
13.6 Exercises . . . 492

14 Vector–valued functions **497**
14.1 Vector–valued functions 497
14.1.1 Definition 497
14.1.2 Algebra of vector–valued functions 498
14.2 Calculus and vector–valued functions 501
14.2.1 Limits of vector-valued functions 501
14.2.2 Continuity 502
14.2.3 Derivatives 503
14.2.3.1 Definition and properties 503
14.2.3.2 Tangent vector and lines 506
14.2.3.3 Smoothness 507
14.2.4 Integration 508
14.2.5 The calculus of motion 510
14.3 Unit tangent and normal vectors 514
14.4 Arc length and curvature 517
14.4.1 Arc length 517
14.4.2 Curvature 520
14.5 Exercices 525

III Multivariable calculus **529**

15 Functions of several variables **531**
15.1 Introduction to multivariable functions 531
15.1.1 Functions of two variables 531
15.1.2 Functions of three variables 535
15.2 Limits and continuity of multivariable functions 536
15.2.1 Introductory concepts and definitions 536
15.2.2 Limits 538
15.2.3 Continuity 543
15.2.4 Functions of three variables 544
15.3 Partial derivatives 545
15.3.1 First partial derivatives 545
15.3.2 Second partial derivatives 548
15.3.3 Higher-order partial derivatives 551
15.3.4 Functions of three variables 552
15.4 Total differential and differentiability 552
15.4.1 Total differential 552

15.4.2 Differentiability . . . 553
15.4.2.1 Definition . . . 553
15.4.2.2 Approximating with differentials . . . 555
15.4.3 Functions of three variables . . . 556
15.5 The multivariable chain rule . . . 557
15.5.1 Rationale . . . 557
15.5.2 Implicit functions . . . 560
15.6 Directional derivatives . . . 562
15.6.1 Definition . . . 562
15.6.2 The gradient . . . 564
15.6.3 Functions of three variables . . . 570
15.7 Tangent lines, normal lines, and tangent planes . . . 571
15.7.1 Tangent and normal lines . . . 571
15.7.2 Tangent plane . . . 576
15.8 Taylor series expansions . . . 579
15.9 Extreme values . . . 581
15.9.1 Necessary conditions for extrema . . . 582
15.9.2 Sufficient conditions for extrema . . . 588
15.9.2.1 Functions of two variables . . . 588
15.9.2.2 Functions of three variables . . . 590
15.10 Constrained optimisation . . . 591
15.10.1 Extrema in a restricted domain . . . 592
15.10.2 Linear programming . . . 595
15.11 Exercices . . . 600

16 Double and line integrals **609**
16.1 Iterated integrals and area . . . 609
16.1.1 Iterated integrals . . . 609
16.1.2 Area of a plane region . . . 611
16.2 Double integration and volume . . . 616
16.2.1 Definition . . . 616
16.2.2 Properties . . . 622
16.2.3 Double integration with polar coordinates . . . 627
16.3 Centre of mass . . . 632
16.3.1 Mass and weight . . . 632
16.3.2 Centre of mass . . . 636
16.4 Surface area . . . 640

16.5 Line integrals over a scalar field . . . 644
16.5.1 Definition . . . 644
16.5.2 Properties . . . 649
16.5.3 Centre of mass . . . 649
16.6 Vector fields . . . 651
16.6.1 Definition . . . 651
16.6.2 The del operator . . . 652
16.6.3 Divergence and curl . . . 654
16.7 Line Integrals over vector fields . . . 658
16.7.1 Definition . . . 658
16.7.2 Properties . . . 662
16.7.3 The fundamental theorem of line integrals . . . 664
16.8 Exercises . . . 668

IV Differential equations 675

17 Introduction 677
17.1 Mathematical modelling and simulation . . . 677
17.1.1 Mathematical models . . . 677
17.1.2 Model development . . . 678
17.2 Differential and difference equations . . . 681
17.2.1 Differential equations . . . 681
17.2.1.1 Nomenclature . . . 681
17.2.1.2 Solutions of differential equations . . . 682
17.2.2 Difference equations . . . 687
17.3 Exercises . . . 690

18 Qualitative analysis of first-order differential equations 693
18.1 Direction and vector fields . . . 693
18.2 Equilibria and stability . . . 696
18.2.1 Equilibrium points . . . 696
18.2.2 Stability . . . 697
18.3 Bifurcation . . . 702
18.4 Exercises . . . 707
18.4.1 Direction and vector fields . . . 707
18.4.2 Equilibria and stability . . . 710
18.4.3 Bifurcation . . . 711

19 Methods of solution for first-order differential equations 715
19.1 Linear first-order differential equations . . . 715
19.1.1 Existence and uniqueness of solutions . . . 715
19.1.2 Homogeneous linear first-order differential equations . . . 716
19.1.3 Non-homogeneous linear first-order differential equations . . . 718
19.2 Solutions of non-linear first-order differential equations . . . 723
19.3 Separable differential equations . . . 724
19.4 Exact differential equations . . . 728
19.5 Change of variables . . . 731
19.6 Euler's method . . . 735
19.6.1 Motivation and rationale . . . 735
19.6.2 The method . . . 736
19.6.3 Error analysis . . . 738
19.7 Runge-Kutta methods . . . 740
19.7.1 Improving the Euler method . . . 740
19.7.2 The methods of Runge and Kutta . . . 741
19.8 Exercises . . . 743

V Appendices 747

A Proofing Techniques 749
A.1 Definitions . . . 749
A.2 Theorems . . . 749

B Calculus in Mathematica and Wolfram|Alpha 751
B.1 Mathematica Notebooks . . . 751
B.1.1 Input cell . . . 752
B.1.2 Text and layout . . . 752
B.2 Mathematica for dummies . . . 752
B.2.1 Operations, evaluations and lists . . . 752
B.2.1.1 Mathematical and realtional operations . . . 752
B.2.1.2 Evaluation of expressions . . . 753
B.2.1.3 Assignements . . . 753
B.2.1.4 Lijsten . . . 754
B.2.2 Functions . . . 755
B.2.3 Solving (un)equality's . . . 757
B.2.4 Visualisatie . . . 757

B.3 Calculus specific intstructions 761
B.3.1 Limits 761
B.3.2 Derivatives 762
B.3.3 Integrals 764
B.3.4 Series 765

C Python Tutorial 767
C.1 What are Python and Jupyter Notebooks? 767
C.2 Python for dummies 768
C.2.1 Objects 768
C.2.2 Variables 769
C.2.3 Mathematical operations 770
C.2.4 Logical operations 770
C.2.5 List operations 770
C.2.6 Python functions 771
C.2.7 Control statements 773
C.2.7.1 While-loop 773
C.2.7.2 For-loop 774
C.2.7.3 If, else-control statement 775
C.3 Packages 775
C.3.1 Numpy 776
C.3.2 Sympy 776
C.3.3 Other Packages 776

D Answers to the exercises 777

E Review exercises 875
E.1 Exam January 2019 875
E.2 Exam August 2019 878
E.3 Exam January 2020 882
E.4 Exam August 2020 885
E.5 Examination January/March 2021 886
E.6 Exam August 2021 889
E.7 Oplossingen 890

F List with definitions 901

G List with theorems 907

1
Introduction

Imagine the following situation.

Sam and Alex are travelling in the car, but its speedometer is broken. Still, Alex wants to know how fast they are going, so he asks Sam. The latter says that they covered 1.2 kilometres in the last minute, so he argues that they are driving 72 km/h. Alex is not, however, satisfied with this answer because he does not want to know the average speed [LT^{-1}] for the last minute, or even the last second, rather he wants to know the speed right now. As they are approaching a road sign, Sam says that they will measure it up there. He observes that they were AT the sign for zero seconds, and the distance was zero meters, so their speed is:

$$\frac{0\ \mathrm{m}}{0\ \mathrm{s}} = \frac{0}{0}\,,$$

and he wisely says that he does not know. He argues that he needs to know some distance over some time, so keeping the time should zero cannot be done.

Actually, Alex wants to know their instantaneous speed, and this might seem pretty amazing, but it is not easy to work out the speed of a car at any point in time. Even the speedometer of a car just shows us an average of how fast we were going for the last (very short) amount of time.

Now, consider we drop a ball from the roof top terrace of the main building at Campus Ledeganck. For the sake of simplicity, we use the following simplified formula to find the distance d [L], measured in metres, fallen:

$$d = 5\,t^2\,,$$

where t [T] is time, measures in seconds. Clearly, after one second, the distance fallen is five metres, but how fast is that? We know

$$\text{speed} = \frac{\text{distance}}{\text{time}}\,,$$

so at one second we get a speed of 5 m/s, but as in the previous situation this constitutes an average speed. If we like to know the instantaneous speed, we run in exactly the same problem as before, as we get for the speed at $t = 1$s:

$$\text{speed} = \frac{0\text{ m}}{0\text{ s}}\,.$$

Let us try to circumvent this problem by inventing a time Δt so short it will not matter. Let us work out the difference in distance between t and $t + \Delta t$. At 1 second the ball has fallen

$$5\,t^2 = 5 \cdot (1)^2 = 5\text{ m}.$$

At $t + \Delta t$ seconds the ball has fallen

$$\begin{aligned} 5\,t^2 &= 5\,(1 + \Delta t)^2\text{ m}, \\ &= 5\,(1 + 2\Delta t + (\Delta t)^2)\text{ m}, \\ &= 5 + 10\Delta t + 5(\Delta t)^2\text{ m}. \end{aligned}$$

Consequently, the difference in distance between t and $t + \Delta t$ is

$$10\Delta t + 5(\Delta t)^2\text{m}\,,$$

while we get the corresponding speed by dividing this change in distance by the time elapse Δt:

$$\begin{aligned} \text{speed} &= \frac{10\Delta t + 5(\Delta t)^2\text{m}}{\Delta t\text{ s}}, \\ &= 10 + 5\Delta t\text{m/s}. \end{aligned}$$

Now if we want Δt to be so small it will not matter, we shrink it to zero and get 10 m/s.

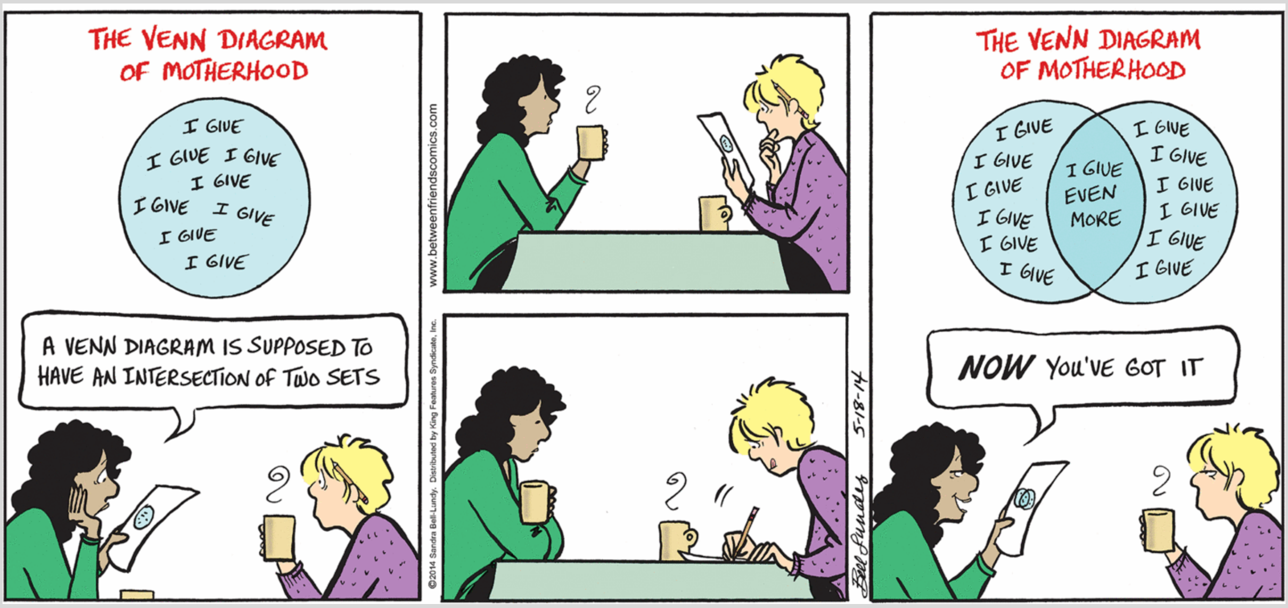
Without really paying attention to it, we just used calculus to cut time and distance into such small pieces that a pure answer came out. The fundamental idea of calculus is to study change by studying instantaneous change, by which we mean changes over tiny intervals of time. It turns out that such changes tend to a be lot simpler to analyse than changes over finite intervals of time.

The goal of this course is to get a comprehensive understanding of what calculus exactly is, and even more importantly, what we can do with it.

In Part I we present the preliminaries that one should master before even trying to move on to the study of calculus itself. The latter is the subject of Parts II and III. More precisely, Part II introduces differential and integral calculus of functions of one variable, while multivariable functions are covered in Part III.

PART I

PRECALCULUS



The only way to learn mathematics is to do mathematics.

— Paul Halmos —

2

Sets and numbers

To make sure that everyone has an understanding of the basic concepts that are at the basis of subsequent chapters, we begin this part with a brief summary of set theory and some of the associated vocabulary and notations we use throughout the text.

2.1 Sets

2.1.1 Logic operators

Throughout this course, and especially when defining new mathematical objects, we will often make use of notation that contains one or more logic operators. An overview of them is given in Table 2.1. Note the difference between $X \Rightarrow Y$ and $X \Leftrightarrow Y$. $X \Rightarrow Y$ states that if X is true, Y is also true, but if Y is true, X is not necessarily true. $X \Leftrightarrow Y$, on the other hand, states that X and Y are both true or both not true and thus equivalent. For instance, we may write

$$\text{Garfield is a cat} \ \wedge \ \text{all cats are mammals} \ \Rightarrow \ \text{Garfield is a mammal}\,,$$

though

$$\text{Garfield is a mammal} \ \not\Rightarrow \ \text{Garfield is a cat}\,!$$

Table 2.1: Overview of important logic operators. X and Y denote logic statements, which are either true or false.

Notation	Reads as
$X \Rightarrow Y$	X implies Y; if X then Y
$X \Leftrightarrow Y$	X if and only if Y
$X \wedge Y$	X and Y
$X \vee Y$	X or Y
$\forall x$	for all (elements) x
$\exists x$	there exists an (element) x
$\exists!\, x$	there exists just one (element) x
$.\vert.$	. so that .
$:.$	it holds that .

2.1.2 Definition and representation of sets

We start with a definition of a set.

Definitie 2.1 (Set)
A **set** (*verzameling*) is a well-defined collection of objects which are called the elements of the set.

In this definition, 'well-defined' means that it is possible to determine if something belongs to the collection or not, without prejudice. For example, the collection of letters that make up the word 'smolko' is well defined and is a set, but the collection of the worst math teachers in the world is not well defined, and so is not a set. In general, there are three ways to describe sets.

1. The **verbal method**: use a sentence to define a set.
2. The **roster method**: begin with a left brace '{', list each element of the set only once and then end with a right brace '}'.
3. The **set-builder method**: a combination of the verbal and roster methods.

For example, let S be the set described verbally as the set of letters that make up the word 'smolko'. A roster description of S would be $S = \{s, m, o, l, k\}$. Note that sets do not allow for repeated elements while they are also orderless, so $\{k, l, m, o, s\}$ is also a roster description of S. A set-builder description of S is:

$$S = \{x \,|\, x \text{ is a letter in the word 'smolko'}\}.$$

In this notation we call x a **dummy variable** and 'x is a letter in the word 'smolko'' the **predicate**. The way to read this is: 'The set of elements x such that x is a letter in the word 'smolko.' Clearly m is in S and q is not in S, i.e. $m \in S$ and $q \notin S$. Moreover, the empty set is written as $A = \{\}$ or $A = \emptyset$, and a set containing a single element is reffered to as a **singleton** (*singleton*).

Graphically, sets are typically represented by means of so-called **Venn diagrams** (*Venn-diagram*), enclosed areas in the plane. For instance, Figure 2.1 shows the Venn diagram of the set $S = \{s, m, o, l, k\}$.

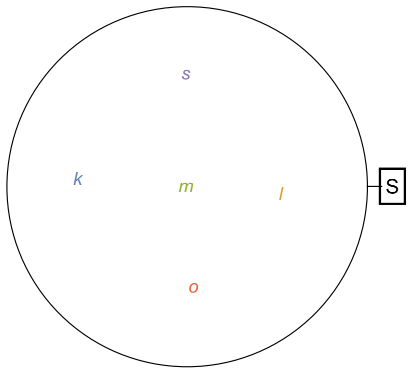


Figure 2.1: Venn diagram of the set $S = \{s, m, o, l, k\}$.

2.1.3 Set operations

Having defined sets, we can now devise some operations that can be performed with them. Let us consider two sets A and B. First, we define that A and B are equal if and only if they have the same elements, that is

$$(A = B) \quad \Leftrightarrow \quad \big(\forall x \,|\, (x \in A) \Leftrightarrow (x \in B)\big)\,.$$

The equality of two sets can be expressed alternatively upon introducing the concept of **subsets** (*deelverzameling*). We say that A is a subset of the set B, if all elements of A are in B; that is if and only if $\forall x : x \in A \Rightarrow x \in B$. We write this as $A \subseteq B$. It immediately follows that two sets A and B are equal if and only if $A \subseteq B$ and $B \subseteq A$. A **proper subset** (*strikte deelverzameling*) A of a set B, denoted $A \subset B$, is a subset that is strictly contained in B and so necessarily excludes at least one element of B.

Equality

Robert Recorde, a Welsh mathematician, introduced the equal sign in 1557. He motivated his choice, by stating: *And to avoid the tedious repetition of these words: is equal to: I will set as I do often in work use, a pair of parallels, or Gemowe lines of one length, thus: $=$, because no 2 things, can be more equal.*

For instance, the regular polygons make up a proper subset of the set of polygons.

Finally, we can define a **universal set** (*universum*), often denoted by U, which contains all objects, including itself. Actually, in the set-builder description of the exemplary set $S = \{s, m, o, l, k\}$, it is implicitly understood that the predicate should be interpreted with respect to the letters available in the European alphabet. This alphabet may hence be understood as the universal set in which the predicate must be interpreted. In the case of ambiguity this can made more explicit in the set-builder description as

$$S = \{x \in \text{European alphabet} \,|\, x \text{ is a letter in the word ‘smolko’}\}.$$

Let us now envisage the following operations on the sets A and B, subsets of the universal set U, which are illustrated in Figure 2.2.

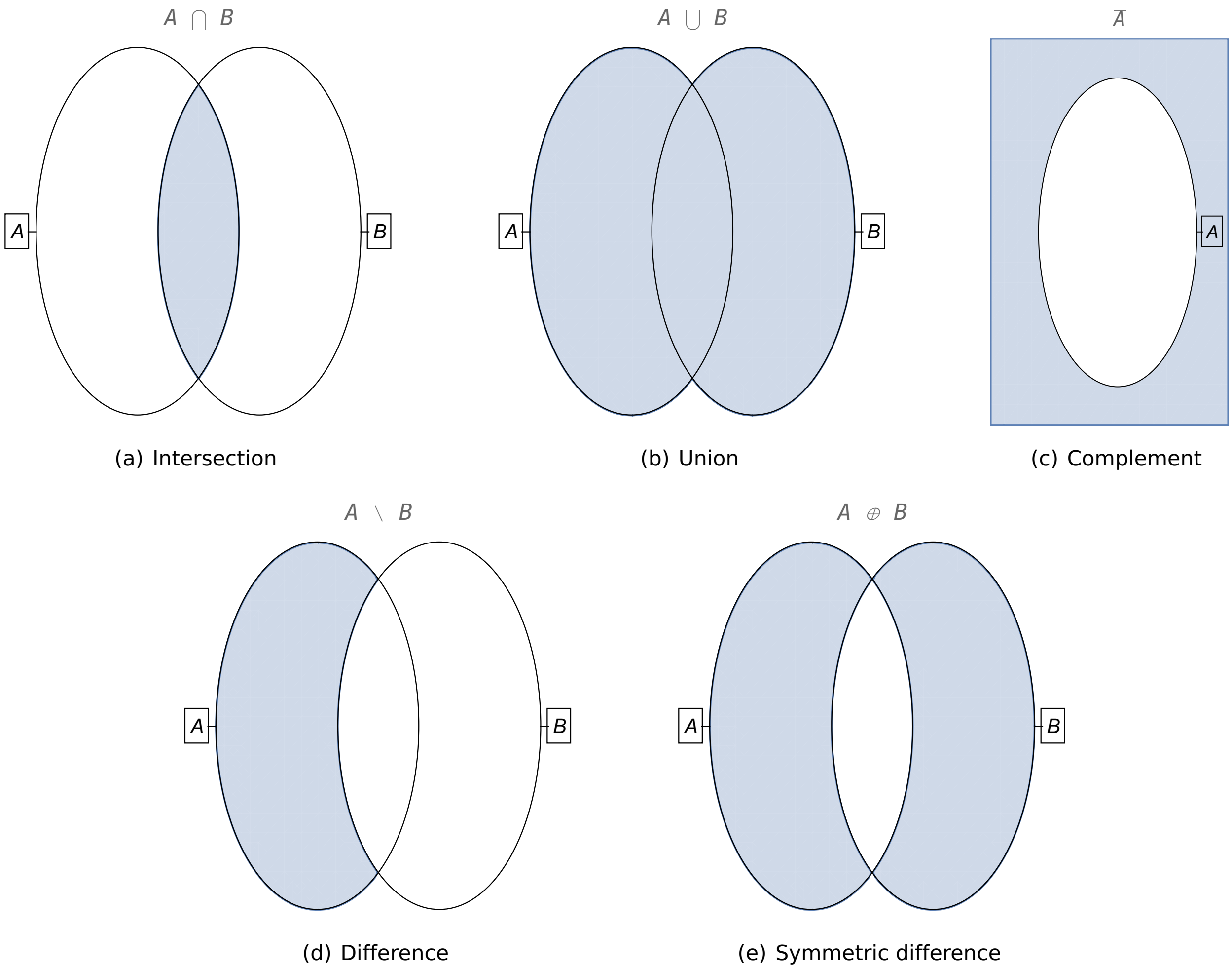


Figure 2.2: Set operations involving two sets A and B.

- **Intersection** (*doorsnede*):
$$A \cap B = \{x \mid (x \in A) \wedge (x \in B)\}.$$
This yields the common elements of A and B. Two sets are called **disjoint** (*disjunct*) if $A \cap B = \varnothing$. This operation generalises directly to n sets.

- **Union** (*unie*):
$$A \cup B = \{x \mid (x \in A) \vee (x \in B)\}.$$
This yields the set of elements that belong to either of the two sets. This operation generalises directly to n sets.

- **Complement** (*complement*):
$$\overline{A} = \{x \in U \mid x \notin A\}.$$
This yields the set of elements in the universal set U that do not belong to a set A.

- **Difference** (*verschil*):
$$A \setminus B = \{x \mid (x \in A) \wedge (x \notin B)\} = A \cap \overline{B}.$$
This yields the set of elements that belong to set A but not to set B.

- **Symmetric difference** (*symmetrisch verschil*):
$$A \oplus B = \{x \mid (x \in A \cup B) \wedge (x \notin A \cap B)\}.$$

This yields the set of elements that belong to either one or the other set but not both.

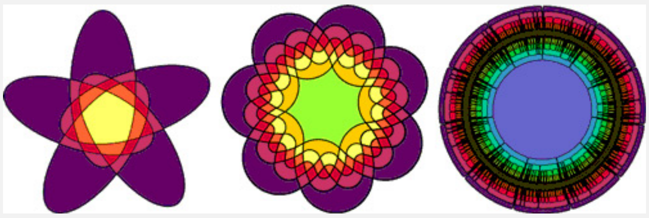
- **Cartesian product** (*Cartesisch product*):

$$A \times B = \{(a,b) \mid (a \in A) \wedge (b \in B)\}.$$

This yields the set of ordered pairs of (a,b) of all elements a and b, that belong to set A and B, respectively. When taking the Cartesian product of A with itself, i.e. $A \times A$, this is also denoted as A^2.

Venn-diagrams

Venn diagrams were named after John Venn (1843–1923), who studied and standardised these diagrams. A major issue that he tried to tackle, was finding symmetrical diagrams of partially multiple overlapping sets. Venn only got as far as 4 sets and it took until 1975 for mathematicians to extend this to 5 and more. The figure below illustrates this for 5, 7 and 11 sets, respectively.



2.1.4 Set properties

Below we list the most important properties of sets, most of which can be understood intuitively or using a Venn diagram representation.

- **Commutativity** (*commutativiteit*) of intersection and union:

$$A \cap B = B \cap A \quad \text{and} \quad A \cup B = B \cup A.$$

- **Associativity** (*associativiteit*) of intersection and union:

$$A \cap (B \cap C) = (A \cap B) \cap C \quad \text{and} \quad A \cup (B \cup C) = (A \cup B) \cup C.$$

- **Distributivity** (*distributiviteit*) with respect to intersection and union:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad \text{and} \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

- **Identity laws**:

$$A \cup \varnothing = A \quad \text{and} \quad A \cap U = A.$$

- **Complement laws**:

$$A \cup \overline{A} = U \quad \text{and} \quad A \cap \overline{A} = \varnothing.$$

For instance, if we are looking for the words that are common to Dutch, English and German, it does not matter that we first determine the words that are common to Dutch and English and then look which of those also are used in German, or start by first determining the words that are common to German and English and finally verify which of those are also used in Dutch.

For completeness, we also mention the **De Morgan's laws** (*regels van De Morgan*):

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \quad \text{and} \quad \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

2.2 The set of real numbers

2.2.1 Definition

Throughout your mathematical upbringing, you have encountered several famous sets of numbers:

- The set of **natural numbers** (*natuurlijke getallen*): $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$.
- The set of **integers** (*gehele getallen*): $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.
- The set of **rational numbers** (*rationale getallen*):

$$\mathbb{Q} = \left\{ \frac{a}{b} \,\middle|\, a, b \in \mathbb{Z} \wedge b \neq 0 \right\}.$$

Essentially, rational numbers are the ratios of integers, provided the denominator is not zero. For instance,

$$\frac{3}{4} = 0.75, \qquad \text{and} \qquad \frac{1}{3} = 0.333333\ldots$$

are just two exemplary rational numbers. Looking at those, it is clear that another way to describe the rational numbers is:

$$\mathbb{Q} = \{x \,|\, x \text{ possesses a repeating or terminating decimal representation}\}.$$

Indeed, it can be proofed that any decimal number with a repeating or terminating decimal representation can be written as a ratio of integers, so as a rational number.

There are of course numbers with a decimal that neither repeats nor terminates, e.g.

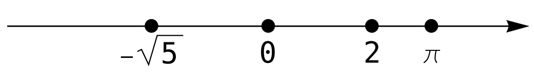
$$\pi = 3.141592654\ldots, \qquad \text{and} \qquad 0.123456789101112123\ldots$$

Such numbers are called **irrational numbers** (*irrationale getallen*) and they form the set of the irrational numbers, denoted $\mathbb{I}$. Now, we can define a new set, namely the set of so-called **real numbers** (*reële getallen*) as follows:

$$\mathbb{R} = \mathbb{I} \cup \mathbb{Q}.$$

Figure 2.3 shows how the sets of natural, rational and real numbers are nested. It clearly holds that the set of natural numbers is a proper subset of the one of the integers, which on its turn is again a proper subset of the set of rational numbers, and so on. Besides, this Venn diagram emphasizes that the sets of rational and irrational numbers are disjoint.

The set $\mathbb{R}$ may be visualized as a line because its elements can be ordered using an order relation. More precisely, the real numbers can be identified with the points on an infinitely long line once its origin, unit of length and orientation have been chosen:



With every real number x corresponds one point on this line, and vice versa, every point on this line represents one real number. This line is called the **real number line** (*reële getallenas*).

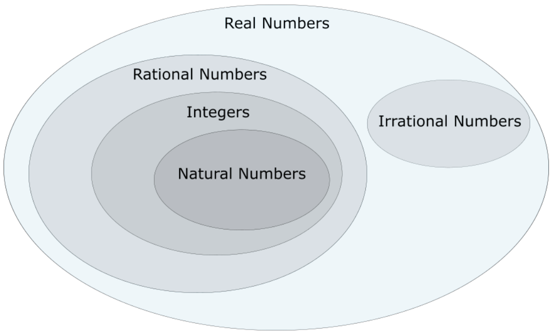


Figure 2.3: Venn diagram of the sets of natural, integer, rational, irrational and real numbers.

In addition to the set of real numbers, we often make reference to one of the following subsets for the sake of brevity:

$$\begin{aligned}
\mathbb{R}_0 &= \mathbb{R}\backslash\{0\},\\
\mathbb{R}^+ &= \{x \mid x \in \mathbb{R} \wedge x \geq 0\},\\
\mathbb{R}^- &= \{x \mid x \in \mathbb{R} \wedge x \leq 0\},\\
\mathbb{R}_0^+ &= \{x \mid x \in \mathbb{R} \wedge x > 0\},\\
\mathbb{R}_0^- &= \{x \mid x \in \mathbb{R} \wedge x < 0\}.
\end{aligned}$$

Moreover, it is possible to extend $\mathbb{R}$ with two more elements, namely positive infinity ($+\infty$) and negative infinity ($-\infty$), which are defined as:

$$\forall x \in \mathbb{R}: \quad -\infty < x < +\infty\,,$$

and which do not belong to $\mathbb{R}$. Doing so, one arrives at the set of **extended real numbers** (*uitgebreide reële getallen*):

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}\,.$$

For completeness, it should be mentioned that there is a further extension of the set of real numbers possible to the set of **complex numbers** (*complexe getallen*), defined by

$$\mathbb{C} = \left\{a + i\,b \mid (a, b \in \mathbb{R}) \wedge \left(i^2 = -1\right)\right\}.$$

This extension allows us, for instance, to compute the square root of a negative number. The complex numbers are discussed in Section 2.3. Throughout this course we will, however, mostly restrict our discussion to the set of real numbers.

2.2.2 Real number arithmetic

2.2.2.1 Addition and multiplication

In the set of real numbers, we can define to main operations namely, addition ($+$) and multiplication ($\cdot$). If a, b and c are three real numbers, we have the following five axioms:

1. **Algebraic closure** (*algebraïsch gesloten*):

$$a + b \in \mathbb{R} \quad \text{and} \quad a \cdot b \in \mathbb{R}\,.$$

2. **Commutativity** (*commutativiteit*):

$$a+b=b+a \quad \text{and} \quad a\cdot b=b\cdot a\,.$$

3. **Associativity** (*associativiteit*):

$$a+(b+c)=(a+b)+c \quad \text{and} \quad a\cdot(b\cdot c)=(a\cdot b)\cdot c\,.$$

4. **Identity property**:

$$a+0=0+a=a \quad \text{and} \quad a\cdot 1=1\cdot a=a\,.$$

5. **Inverse property**:

$$a+(-a)=(-a)+a=0 \quad \text{and} \quad a\cdot a^{-1}=a^{-1}\cdot a=1$$

Essentially, the identity property indicates that 0 is the **neutral element** (*neutraal element*) of the addition operation and 1 is neutral element of the multiplication operation, while the inverse property shows that there is always an **opposite element** (*tegengesteld element*) in the case of addition and an **inverse element** (*invers element*) in the case of multiplication. Finally, according to the sixth axiom of real numbers, multiplication **distributes** over addition:

$$a\cdot(b+c)=a\cdot b+a\cdot c \quad \text{and} \quad (a+b)\cdot c=a\cdot c+b\cdot c\,.$$

The operations of subtraction and division are not listed above because they fail to possess many of the aforementioned properties. More precisely, subtraction and division are not commutative, nor associative, as for instance $4-1\neq 1-4$ and likewise $(4-1)-2\neq 4-(1-2)$.

Throughout this course we will sometimes be confronted with problems involving the summation of many numbers, e.g.

$$a_0+a_1+a_2+\cdots+a_n\,,$$

where n is some natural number. Clearly, writing such sums can become quite cumbersome, so a shorthand notation thereof has been established using the **capital sigma notation**; that is

$$a_0+a_1+a_2+\cdots+a_n=\sum_{i=0}^{n}a_i\,,$$

where i is the **summation index** (*index*) and a_i is a **generic term** (*algemene term*) in the summation. This expression should be read as the sum of the numbers a_i for i going from 0 to n. Here, the index i starts at 0, but this is not a requisite.

We can formulate the following useful properties of summations:

- Sum or difference of summations:

$$\sum_{i=0}^{n}a_i\pm\sum_{i=0}^{n}b_i=\sum_{i=0}^{n}(a_i\pm b_i)\ .$$

- Splitting a summation:

$$\sum_{i=0}^{n}a_i=\sum_{i=0}^{m}a_i+\sum_{i=m+1}^{n}a_i\,,\text{ where } m<n\,.$$

- Scalar multiplication:

$$c \cdot \sum_{i=0}^{n} a_i = \sum_{i=0}^{n} c \cdot a_i \,, \text{where}\, c \in \mathbb{R}\,.$$

- Constant summation:

$$\sum_{i=0}^{n} c = c \cdot (n+1)\,, \text{where}\, c \in \mathbb{R}\,.$$

Moreover, we recall (without proof) two useful identities, that will be used in later chapters and involve the sum of the first n (squared) natural numbers:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \qquad \text{and} \qquad \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}. \tag{2.1}$$

Similarly to the capital sigma notation for summation, we can use the **capital pi notation** for the product of n numbers:

$$\prod_{i=0}^{n} a_i = a_0 \cdot a_1 \cdot a_2 \cdots a_n\,.$$

Using the capital pi notation, it becomes easy to define the so-called **factorial** (*faculteit*) of a natural number n, denoted $n!$:

$$n! = \prod_{k=1}^{n} k = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n.$$

From here on, we will mostly drop the dot-notation to denote multiplication and use a blank space instead. Only in case of ambiguity (e.g. when multiplying two numbers), we will hold on the dot-notation for clarity.

2.2.2.2 Exponentiation

In addition to the four elementary operations in $\mathbb{R}$, namely addition, subtraction, multiplication and division, we can also define **exponentiation** (*machtsverheffing*). It involves two numbers, the **base** (*grondtal*) $b \in \mathbb{R}_0$ and the **exponent** (*exponent*) n and is written as b^n. When n is a strictly positive integer, exponentiation corresponds to repeated multiplication of the base: that is, b^n is the product of multiplying n bases:

$$b^n = \underbrace{b\,b\,b \cdots b}_{n \text{ factors}}\,.$$

This expression should be read as b raised to the power of n. For negative powers, we have

$$b^{-n} = \frac{1}{b^n}\,.$$

By convention, it holds that any non-zero number raised to the 0 power is 1, i.e. $b^0 = 1$ if $b \neq 0$. The expression b^2 is often called the square of b or b squared, while b^3 is frequently called the cube of b or b cubed.

The exponent does not necessarily have to be an integer, it can as well be a rational number, such as $1/n$, where $n \in \mathbb{N}_0$. More precisely, we can have

$$x = b^{\frac{1}{n}}\,,$$

which should be interpreted as the number x for which the n-th power equals b. This implies that $b^{1/n}$ is a solution to the equation

$$x^n = b\,.$$

Example 2.1

Moore's law concerns the observation that the number of transistors in a dense integrated circuit doubles about every two years. The observation is named after Gordon Moore, whose 1965 paper described a doubling every year in the number of components per integrated circuit, and projected this rate of growth would continue for at least another decade. Mathematically, we can state this law as

$$P_n = P_0\, 2^n\,,$$

where P_0 [–] is the number of transistors in some reference year, n [–] the number of two-year periods, and P_n [–] the number transistors n two-year periods passed the reference year. In 1988, the number of transistors in the Intel 386 SX microprocessor was 275 000. What were the transistors counts of the Pentium II Intel microprocessor in 1997?

Solution

Between 1988 and 1997 there are 9 years, so 4.5 periods of two years. Hence, if Intel doubles the number of transistors every two years, the new processor would have

$$P_{9/2} = 275\,000 \cdot 2^{\frac{9}{2}} = 275\,000 \cdot 22.63 = 6\,223\,250$$

transistors in 1997. In fact, in 1997, the Pentium II had 7.5 million transistors (Figure 2.4). In other words, since 1988 up until 1997, Intel had been doubling the number of transistors in its microprocessors in less than every two years.

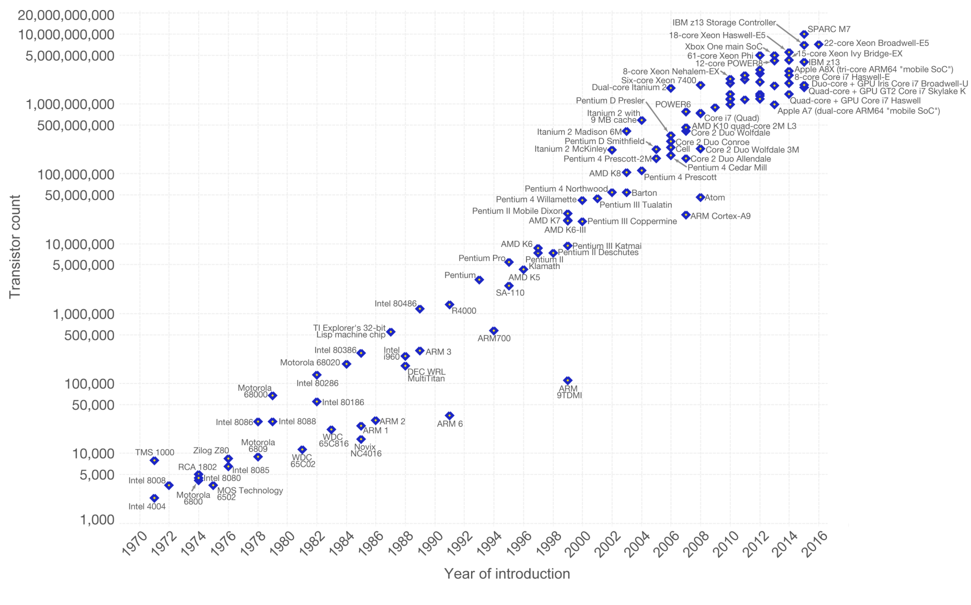


Figure 2.4: Number of transistors in computer hardware manufactured since 1970.

Alternatively, $b^{1/n}$ is often written using the radical symbol as $\sqrt[n]{b}$. It is called the principal n-th root of b. If n is even and b is positive, then $x^n = b$ has two real solutions because even powers of real numbers are always positive. These solutions are the positive and negative n-th roots, i.e.

$$\sqrt[n]{a} \qquad \text{and} \qquad -\sqrt[n]{a}\,.$$

If b is negative, the equation has no solution in the set real numbers for even n. On the other hand, if n is odd, then $x^n = b$ has exactly one real solution $b^{1/n}$ that is positive if b is positive and negative if b is negative. Finally, taking a positive real number b to a rational exponent m/n, where m is an integer and n is a positive integer, and considering principal roots only, yields

$$b^{\frac{m}{n}} = \left(b^m\right)^{\frac{1}{n}} = \sqrt[n]{b^m} = \left(\sqrt[n]{b}\right)^m .$$

The following basic identities hold for the operation of exponentiation for every $a, b \in \mathbb{R}$ and $p, q \in \mathbb{Q}$:

- $a^p\, a^q = a^{p+q}$,
- $(a^p)^q = a^{p\,q}$,
- $(a\,b)^p = a^p\, b^p$,

and recalling that $1/a^p = a^{-p}$, we also have:

- $\dfrac{a^p}{a^q} = a^{p-q}$, if $a \neq 0$,
- $\left(\dfrac{a}{b}\right)^p = \dfrac{a^p}{b^p}$, if $b \neq 0$.

However, be aware that exponentiation is not commutative, nor associative, unlike addition and multiplication. Indeed, it is clear that $2^3 = 8 \neq 3^2 = 9$ and likewise

$$\left(2^3\right)^4 = 8^4 = 4096 \neq 2^{(3^4)} = 2^{81} = 2417851639229258349412352 .$$

Example 2.2

Yearly, there are about three consecutive generations of box moths in Belgium. Suppose the number of box moths (*Cydalima perspectalis*) in generation i can be described as follows:

$$B_i = r\, B_{i-1} ,$$

where r [T^{-1}] represents the growth rate of the box moth population. Assuming that we know the initial number of box moths (generation 0), we can compute the number of box moths in generation 1 as

$$B_1 = r\, B_0 .$$

Then, we can compute the number of box moths in generation 2:

$$B_2 = r\, B_1 = r\, r\, B_0 = r^2\, B_0 ,$$

and so on.

The total number of box moths that has seen daylight up to and including generation n, can then be written as

$$\begin{aligned} T_n &= \sum_{i=0}^{n} B_i , \\ &= B_0 + B_1 + B_2 + \cdots + B_n , \\ &= B_0 + r\, B_0 + r^2\, B_0 + \cdots + r^n\, B_0 . \end{aligned}$$

For instance, if $B_0 = 5$, $r = 1.1$ and $n = 5$, we find $T_5 \approx 39$ individuals.

For what concerns the square and cube of the sum and difference of two real numbers, we can infer the following well-known identities:

$$\begin{aligned}
(a+b)^2 &= a^2+2ab+b^2 \\
(a-b)^2 &= a^2-2ab+b^2 \\
(a+b)(a-b) &= a^2-b^2 \\
(a+b)^3 &= a^3+3a^2b+3ab^2+b^3 \\
(a-b)^3 &= a^3-3a^2b+3ab^2-b^3 \\
(a+b)(a^2-ab+b^2) &= a^3+b^3 \\
(a-b)(a^2+ab+b^2) &= a^3-b^3
\end{aligned}$$

2.2.2.3 Arithmetic in $\overline{\mathbb{R}}$

Obviously, the rules of arithmetic that apply to $\mathbb{R}$ apply to $\overline{\mathbb{R}}$ as well, but we need to introduce a few more rules involving a real number $a \in \mathbb{R}$ and/or $+\infty$ and/or $-\infty$. With regard to addition, we have

$$\begin{aligned}
a+(+\infty) &= (+\infty)+a = +\infty\,, \\
a+(-\infty) &= (-\infty)+a = -\infty\,, \\
(+\infty)+(+\infty) &= +\infty\,, \\
(-\infty)+(-\infty) &= -\infty\,,
\end{aligned}$$

while for subtraction we have

$$\begin{aligned}
a-(+\infty) &= -\infty\,, \\
a-(-\infty) &= +\infty\,, \\
(+\infty)-a &= +\infty\,, \\
(-\infty)-a &= -\infty\,, \\
(+\infty)-(-\infty) &= +\infty\,, \\
(-\infty)-(+\infty) &= -\infty\,.
\end{aligned}$$

For multiplication, we accordingly have (considering $a \in \mathbb{R}_0$)

$$a\cdot(+\infty) = (+\infty)\cdot a = \begin{cases} +\infty\,, & \text{if} \quad a>0\,, \\ -\infty\,, & \text{if} \quad a<0\,, \end{cases}$$

and

$$a\cdot(-\infty) = (-\infty)\cdot a = \begin{cases} -\infty\,, & \text{if} \quad a>0\,, \\ +\infty\,, & \text{if} \quad a<0\,. \end{cases}$$

Likewise, for products involving $+\infty$ and/or $-\infty$ only:

$$\begin{aligned}
(+\infty)\cdot(+\infty) &= +\infty\,, \\
(-\infty)\cdot(-\infty) &= +\infty\,, \\
(+\infty)\cdot(-\infty) &= (-\infty).(+\infty) = -\infty\,.
\end{aligned}$$

And for the division, we have

$$\frac{a}{+\infty} = \frac{a}{-\infty} = 0\,,$$

irrespective of the sign of $a \in \mathbb{R}_0$, and

$$\frac{+\infty}{a} = \begin{cases} +\infty\,, & \text{if} \quad a > 0\,, \\ -\infty\,, & \text{if} \quad a < 0\,, \end{cases}$$

$$\frac{-\infty}{a} = \begin{cases} -\infty\,, & \text{if} \quad a > 0\,, \\ +\infty\,, & \text{if} \quad a < 0\,. \end{cases}$$

And finally, for what concerns principal n-th roots, for $n \in \mathbb{N}_0$:

$$\begin{aligned} \sqrt[n]{+\infty} &= +\infty\,, \\ \sqrt[2n+1]{-\infty} &= -\infty\,. \end{aligned}$$

2.2.3 Intervals in $\mathbb{R}$

Segments of the real number line are called **intervals** (*interval*) of numbers. Table 2.2 gives a summary of the interval notation for real numbers. If the endpoint is included in the interval, we use closing square brackets, '[' or ']', when defining the interval and use a filled dot to indicate membership in the interval on the real number line. Otherwise, we use opening square brackets, ']' or '[', and a circle to indicate that the endpoint is not part of the set. If the interval does not have finite endpoints, we use the symbols $-\infty$ to indicate that the interval extends infinitely to the left and $+\infty$ to indicate that the interval extends infinitely to the right. Since infinity is a concept, and not a number, we always use opening square brackets when using these symbols in interval notation.

It should not be forgotten that any interval in $\mathbb{R}$ corresponds with a certain set of real numbers, so that we may apply the set operations introduced in Section 2.1.3 directly to intervals. For example, if $A = [-5, 3\,[$ and $B = \,]\,1, +\infty\,[$, then we easily find $A \cap B = \,]\,1, 3\,[$ and $A \cup B = [-5, +\infty\,[$. Likewise, we find $A \setminus B = [-5, 1]$.

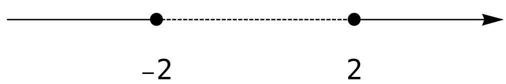
Example 2.3

Express the following sets of numbers using interval notation.

1. $\{x \,|\, x \leq -2 \ \vee \ x \geq 2\}$
2. $\{x \,|\, x \neq \pm 3\}$
3. $\{x \,|\, -1 < x \leq 3 \ \vee \ x = 5\}$

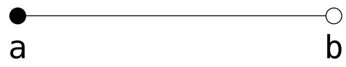
Solution

1. The best way to proceed here is to graph the set of numbers on the number line. The inequality $x \leq -2$ corresponds to the interval $]-\infty, -2]$ and the inequality $x \geq 2$ corresponds to the interval $[2, +\infty\,[$. Since we are looking to describe the real numbers x in one of these or the other, we have $\{x \,|\, x \leq -2 \ \vee \ x \geq 2\} = \,]-\infty, -2] \cup [2, +\infty\,[$.



2. For the set $\{x \,|\, x \neq \pm 3\}$, we exclude both $x = 3$ and $x = -3$ from our set. This breaks the number line into three intervals, $]-\infty, -3\,[$, $]-3, 3\,[$ and $]\,3, +\infty\,[$, so $\{x \,|\, x \neq \pm 3\} = \,]-\infty, -3\,[\, \cup \,]-3, 3\,[\, \cup \,]\,3, +\infty\,[$.

Table 2.2: Interval notation for two real numbers a and b for which it holds that $a < b$.

Set of real numbers	Interval notation	Region on the real number line
$\{x \mid a < x < b\}$	$]\,a, b\,[$	
$\{x \mid a \leq x < b\}$	$[a, b\,[$	
$\{x \mid a < x \leq b\}$	$]\,a, b]$	
$\{x \mid a \leq x \leq b\}$	$[a, b]$	
$\{x \mid x < b\}$	$]-\infty, b\,[$	
$\{x \mid x \leq b\}$	$]-\infty, b]$	
$\{x \mid x > a\}$	$]\,a, +\infty\,[$	
$\{x \mid x \geq a\}$	$[a, +\infty\,[$	
$\mathbb{R}$	$]-\infty, +\infty\,[$	

3. Graphing the set $\{x \mid -1 < x \leq 3 \ \vee \ x = 5\}$, we get one interval, $]-1, 3]$ along with a single number, or point, $\{5\}$. Consequently, we have $\{x \mid -1 < x \leq 3 \ \vee \ x = 5\} = \,]-1, 3] \cup \{5\}$.

2.3 Complex numbers

We leave a detailed discussion of complex numbers to the course 'Algebra', and restrict here to a basic introduction to complex which suffices for the scope of this course.

2.3.1 Definition

Consider the polynomial $p(x) = x^2 + 1$. The zeros of p are the solutions to $x^2 + 1 = 0$, or $x^2 = -1$. This equation has no real solutions, but we can formally extract the square roots of both sides to get $x = \pm\sqrt{-1}$. The quantity $\sqrt{-1}$ is usually relabeled i, the so-called **imaginary unit** (*imaginaire eenheid*). The number i, while not a real number, plays along well with real numbers, and acts very much like any other radical expression. For instance, $3(2i) = 6i$, $7i - 3i = 4i$, $(2-7i)+(3+4i) = 5-3i$, and so forth. The key property that distinguishes i from the real numbers is the fact that

$$i^2 = -1\,.$$

Hence, if c is a real number with $c \geq 0$, then we can write

$$\sqrt{-c} = i\sqrt{c}\,.$$

Having defined the imaginary unit, we are now in the position to define the complex numbers.

Definitie 2.2 (Complex number)
A **complex number** (*complex getal*) is a number of the form

$$a + bi\,,$$

where a and b are real numbers and i is the **imaginary unit** (*imaginaire eenheid*).

Do not forget that a or b could be zero, which means numbers like $3i$ and 6 are also complex numbers. In other words, do not forget that the complex numbers include the real numbers, so 0 and $\pi - \sqrt{21}$ are both considered complex numbers (See Figure 2.5).

2.3.2 Complex number arithmetic

The arithmetic of complex numbers is as you would expect, as long as you remember that $i^2 = -1$.

Example 2.4

Perform the indicated operations. Write your answer in the form $a + bi$.

1. $(1-2i)-(3+4i)$
2. $(1-2i)(3+4i)$
3. $\dfrac{1-2i}{3-4i}$
4. $\big(x-(1+2i)\big)\big(x-(1-2i)\big)$

Solution

1. We combine like terms to get $(1-2i)-(3+4i) = 1-2i-3-4i = -2-6i$.

2. Using the distributive property, we get

$$(1-2i)(3+4i) = 3+4i-6i-8i^2\,.$$

Since $i^2 = -1$, we get $3+4i-6i-8i^2 = 11-2i$.

3. First we deal with the denominator $3-4i$ as we would any other denominator containing a radical, and multiply both numerator and denominator by $3+4i$. Doing so produces

$$\frac{1-2i}{3-4i}\,\frac{3+4i}{3+4i} = \frac{(1-2i)(3+4i)}{(3-4i)(3+4i)} = \frac{11-2i}{25} = \frac{11}{25} - \frac{2}{25}\,i.$$

4. We can rely on the fact that $(a-b)(a+b) = a^2-b^2$ and see that

$$\begin{aligned}\big(x-(1+2i)\big)\big(x-(1-2i)\big) &= \big((x-1)-2i\big)\big((x-1)+2i\big)\\ &= \big((x-1)^2-(2i)^2\big)\\ &= x^2-2x+5.\end{aligned}$$

A couple of remarks about the last example are in order. First, the **conjugate** (*(complex) toegevoegde*) of a complex number $a+bi$ is the number $a-bi$. The notation commonly used for conjugation is a bar; that is

$$\overline{a+bi} = a-bi\,.$$

For example, $\overline{3+2i} = 3-2i$ and $\overline{6} = 6$. The properties of the conjugate are summarized below, for z and w complex numbers.

- $\overline{\overline{z}} = z$
- $\overline{z}+\overline{w} = \overline{z+w}$
- $\overline{z}\,\overline{w} = \overline{zw}$
- $\big(\overline{z}\big)^n = \overline{z^n}$, for any natural number n
- z is a real number if and only if $\overline{z} = z$.

To form the **opposite** (*tegengestelde*) of a complex number, take the opposite of each part:

$$-(a+bi) = -a+(-b)i = -a-bi.$$

For example, the opposite of $6-2i$ is $-6+2i$.

Although we will only rarely have to resort to complex numbers throughout this course, their importance in engineering cannot be underestimated since complex numbers have essential applications in a variety of scientific and related areas such as signal processing, control theory, electromagnetism, fluid dynamics, quantum mechanics, cartography, and vibration analysis. For that reason, you will often encounter them in more advanced mathematics courses, such as differential equations. Besides, using complex numbers one can construct arty and intriguing graphics, like the so-called Julia set depicted in Figure 2.5.

Quaternions

The quaternions are a number system that extends the complex numbers, dating back to the 1840s only. They are generally represented as:

$$a + b\ \mathbf{i} + c\ \mathbf{j} + d\ \mathbf{k},$$

where a, b, c, and d are real numbers, and $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are the fundamental quaternion units for which it holds that $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1$. Amongst other things, the spin of an electron can be described using quaternions.

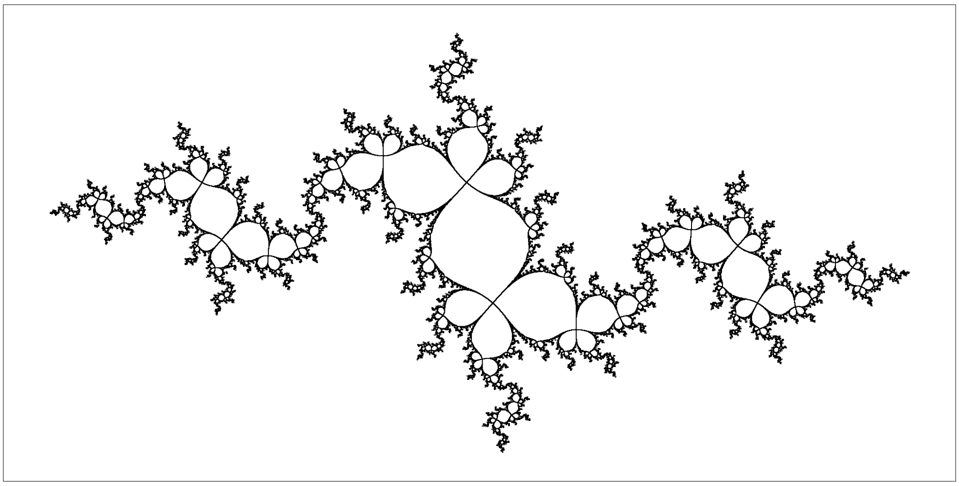


Figure 2.5: Exemplary Julia set.

2.4 Exercises

Sets and Logic operators

Assignment 2.1 — Determine the negation of the expressions below.

(a) $\exists\, x \in \mathbb{Z},\ \forall\, y \in \mathbb{Z} : x < y$

(b) $\forall\, y \in \mathbb{Z},\ \exists\, x \in \mathbb{Z} : x < y$

(c) $\forall\, x \in \mathbb{Z},\ \exists\, y \in \mathbb{Z} : x + y = 0$

(d) $\exists\, y \in \mathbb{Z},\ \forall\, x \in \mathbb{Z} : x + y = 0$

Assignment 2.2 — Define the following sets with words and write them out completely.

(a) $A = \{a \mid a \in \mathbb{N} \,\wedge\, 2 < a < 6\}$

(b) $B = \{x \mid x \in \mathbb{Q}^+ \,\wedge\, 2x^2 + x - 6 = 0\}$

(c) $C = \{x \mid x \in \mathbb{Z}^+ \,\wedge\, x^2 - 5 = 0\}$

Assignment 2.3 — Write in a concise manner that A is a set of;

(a) all even numbers bigger than 100.

(b) all pairs of integers whose first and second elements are even and odd, respectively.

(c) all integers, different from zero, that are multiples of 3.

(d) all positive rational numbers whose square root is greater than 3.

(e) all numbers that when divided by 6 result in a remainder of 2.

Assignment 2.4 — Fill in the correct symbols. Choose from $\subset, \not\subset, =, \neq, \in, \notin, \ni, \not\ni$. Multiple answers might be possible.

(a) $\{1, 3, 5, 7, 9, 11, \ldots\} \ldots \{x \in \mathbb{N} \mid x \text{ is an even number}\}$

(b) $\{x \mid x \text{ is a rose}\} \ldots \{x \mid x \text{ is a flower}\}$

(c) $\{1, 3, 5, 7, 9\} \ldots 2$

(d) $\{1\} \ldots \{1, 3, 5, 7, 9\}$

(e) $\{1, 3\} \ldots \{\{1\}, \{3\}, \{5\}, \{7\}, \{9\}\}$

(f) $\{1, 3\} \ldots \{1, 3, 5, 7, 9\}$

(g) $\{1, 3, 5, 7, 9\} \ldots \emptyset$

(h) $\{1\} \ldots \{\{1\}, \{3\}, \{5\}, \{7\}, \{9\}\}$

(i) $\{1, 3, \{5, 7, 9\}\} \ldots 5$

Assignment 2.5 — Assume $A = \{1, \{1\}, \{2\}\}$. Which from the following statements is true?

(a) $1 \in A$

(b) $\{1\} \in A$

(c) $\{1\} \subseteq A$

(d) $\{\{1\}\} \subseteq A$

(e) $2 \in A$

(f) $\{\{2\}\} \subseteq A$

(g) $\{\{2\}\} \subset A$

(h) $\{2\} \subseteq A$

Assignment 2.6 — Given $U = \{1, 2, 3, \ldots, 9, 10\}$, assume $A = \{1, 2, 3, 4, 5\}$, $B = \{1, 2, 3, 4\}$, $C = \{3, 5, 7\}$ and $D = \{2, 4, 6, 8\}$. Describe each of the following sets:

(a) $(A \cup B) \cap C$

(b) $A \cup (B \cap C)$

(c) $\overline{C} \cup \overline{D}$

(d) $\overline{C \cap D}$

(e) $(A \cup B) \setminus C$

(f) $A \cup (B \setminus C)$

(g) $(B \setminus C) \setminus D$

(h) $B \setminus (C \setminus D)$

(i) $(A \cup B) \setminus (C \cap D)$

Assignment 2.7 — Simplify the following expressions:

(a) $A \cap (B \setminus A)$

(b) $(A \setminus B) \cup (A \cap B)$

(c) $\overline{A} \cup \overline{B} \cup \left(A \cap B \cap \overline{C}\right)$

(d) $(A \cap B) \cup \left(A \cap B \cap \overline{C} \cap D\right) \cup \left(\overline{A} \cap B\right)$

(e) $(A \cap B) \cup \left(B \cap \left((C \cap D) \cup \left(C \cap \overline{D}\right)\right)\right)$

Assignment 2.8 — Use set notation to define the shaded areas in Figure 2.6.

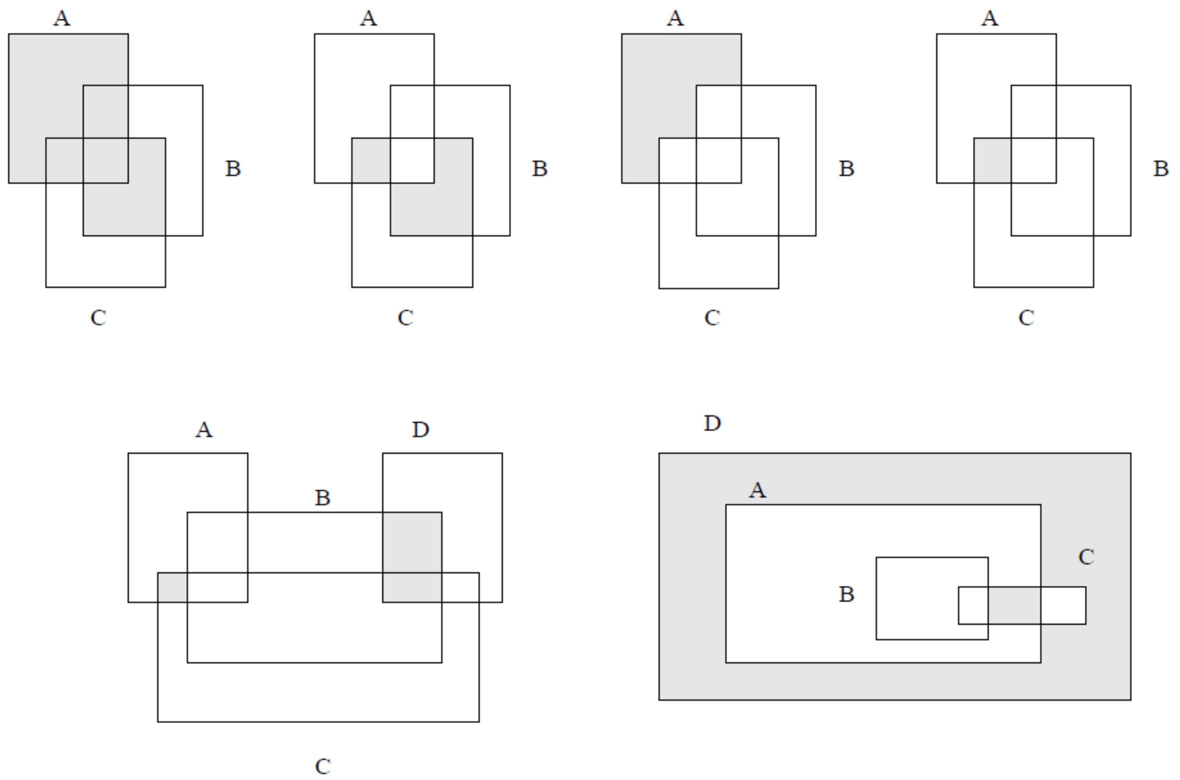


Figure 2.6: Shaded areas used in exercise 2.8.

The set of real numbers

Assignment 2.9 — Which of the numbers below are rational or irrational?

(a) 5.369

(b) $\frac{12}{7}$

(c) $\sqrt{13}$

(d) $1.232345456767...$

(e) $3.0236363636...$

(f) $\sqrt{121}$

Assignment 2.10 — Rewrite the following expressions using a sum or multiplication sign.

(a) $x+x^2+x^3+x^4+...+x^{99}$

(b) $\sqrt{3}+\sqrt{5}+\sqrt{7}+\sqrt{9}+...+\sqrt{51}$

(c) $\frac{1}{a+1}\cdot\frac{4}{a+2}\cdot\frac{9}{a+3}\cdot\frac{16}{a+4}\cdots\frac{169}{a+13}$

Assignment 2.11 — Calculate the following sums.

(a) $\displaystyle\sum_{j=0}^{3} 2^j$

(b) $\displaystyle\sum_{j=0}^{3} j^2$

(c) $\displaystyle\sum_{j=0}^{4} \frac{24}{j!}$

(d) $\displaystyle\sum_{j=1}^{4} \frac{(-1)^j}{j}$

(e) $\displaystyle\sum_{j=1}^{5} (2j-1)^2$

(f) $\displaystyle\sum_{j=1}^{100} -7j^2$

(g) $\displaystyle\sum_{j=1}^{90} \left(-2j^2+3j-5\right)$

(h) $\displaystyle\sum_{\substack{0<k<10\\ k \text{ is oneven}}} k^2$

Assignment 2.12 — Simplify each sum or product to an expression without sigma or pi notation.

(a) $\displaystyle\sum_{j=1}^{n} (3j-2)$

(b) $\displaystyle\sum_{j=1}^{n} \left((3j-5)\,\frac{1}{n^2}\right)$

(c) $\displaystyle\sum_{j=1}^{n} (3j-4)^2$

(d) $\displaystyle\sum_{j=1}^{n} \left((j-2)^2\,\frac{1}{n^3}\right)$

(e) $\displaystyle\prod_{j=1}^{n} j^3$

(f) $\displaystyle\prod_{k=2}^{n} \left(1-\frac{1}{k^2}\right)$

Assignment 2.13 — Calculate or simplify the following algebraic forms.

(a) $\left(64\, a^{6m} b^{12n} c^{18p}\right)^{\frac{1}{6}}$

(b) $\dfrac{1-x}{1-\sqrt{x}}$

(c) $\left(1-\sqrt{2}-\sqrt{3}\right)^2$

(d) $b\sqrt{\dfrac{4a}{b^4}}-\sqrt{\dfrac{9a}{b^2}}+\dfrac{1}{b}\sqrt{\dfrac{a}{4}}+2b\sqrt{\dfrac{25a}{b^4}}$

(e) $\left(\sqrt{\dfrac{x+1}{x-1}}\right)\left(\sqrt[3]{\dfrac{x-1}{x+1}}\right)$

(f) $\sqrt[3]{a^3+\dfrac{3}{2}a^2b+\dfrac{3}{4}ab^2+\dfrac{1}{8}b^3}$

(g) $\dfrac{a^2+b^2}{(b-a)^2}\;\dfrac{(a-b)^3}{a+b}\;\dfrac{(-a-b)^2}{a^2-b^2}$

(h) $\left(x^{\frac{1}{3}}-x^{-\frac{1}{3}}\right)^3+3\left(x^{\frac{1}{3}}-x^{-\frac{1}{3}}\right)$

(i) $\dfrac{\left(x^2\right)^3 x^{-4}\sqrt[3]{x^5}}{\sqrt[3]{x^2}\,\sqrt[3]{\sqrt[4]{\left(x^2\right)^3}}}$

(j) $\left(\dfrac{16^{-2}a^{\frac{1}{2}}b^{-3}}{81^{-1}a^{-\frac{1}{2}}b^{3}}\right)\sqrt{ab^{\frac{9}{4}}\left(ab^{\frac{3}{2}}\right)^{\frac{1}{2}}}$

Complex numbers

Assignment 2.14 — Determine and simplify

- $z+w$
- zw
- z^2
- z^{-1}
- $\dfrac{z}{w}$
- $\dfrac{w}{z}$
- $\bar{z}$
- $z\bar{z}$
- $(\bar{z})^2$

rewrite each pair of complex numbers in standard form: $a+bi$.

(a) $z=2+3i, \quad w=4i$

(b) $z=1+i, \quad w=-i$

(c) $z=3-5i, \quad w=2+7i$

(d) $z=\sqrt{2}-\sqrt{2}i, \quad w=\sqrt{2}+\sqrt{2}i$

(e) $z=1-\sqrt{3}i, \quad w=-1-\sqrt{3}i$

(f) $z=-\dfrac{\sqrt{2}}{2}+\dfrac{\sqrt{2}}{2}i, \quad w=-\dfrac{\sqrt{2}}{2}-\dfrac{\sqrt{2}}{2}i$

Assignment 2.15 — Write the following numbers in standard form: $a+bi$.

(a) $(4+8i)+(15-12i)$

(b) $(2+4i)-(6-7i)$

(c) $(2+3i)+(-5+i)$

(d) $(2+i)^2$

(e) $\overline{(5+6i)}(5+6i)$

(f) $\dfrac{1}{5+2i}$

(g) $\dfrac{1+i}{2+3i}$

(h) $\dfrac{1+2i}{3-4i}+\dfrac{2-i}{5i}$

Mathematics knows no races or geographic boundaries; for mathematics, the cultural world is one country.

— David Hilbert —

3 Functions

3.1 The Cartesian coordinate plane

In order to visualize the pure excitement that is calculus, we need to unite algebra and geometry. Simply put, we must find a way to draw algebraic things. Let us start with possibly the greatest mathematical achievement of all time: the **Cartesian coordinate plane** (*Cartesisch coördinatenstelsel*). So named in honour of René Descartes.

Imagine two real number lines crossing at a right angle at 0. The horizontal number line is usually called the ***x*-axis** (*x-as*), while the vertical number line is usually called the ***y*-axis** (*y-as*). For example, consider the point P in Figure 3.1. To use the numbers on the axes to label this point, we project the point P to the x- (respectively y-) axis. We then describe the point P using the **ordered pair** (*geordend koppel*) $(2,-4)$. The first number in the ordered pair is called the **abscissa** (*abscis*) or ***x*-coordinate** and the second is called the **ordinate** (*ordinaat*) or ***y*-coordinate**. When we speak of the Cartesian coordinate plane, we mean the set of all possible ordered pairs (x,y) as x and y take values from the real numbers. The ordered pair $(2,-4)$ comprise the **Cartesian coordinates** (*Cartesische coördinaten*) of the point P. In practice, the distinction between a point and its coordinates is blurred. We can think of $(2,-4)$ as instructions on how to reach P from the **origin** (*oorsprong*) $(0,0)$.

The axes divide the plane into four regions called **quadrants** (*kwadrant*). They are labelled with Roman numerals and proceed counter-clockwise around the plane . If a point other than the origin happens to lie on the axes, we typically refer to that point as lying on the positive or negative x-axis (if $y=0$) or on the positive or negative y-axis (if $x=0$). Such points do not belong to any of the four quadrants.

Using Cartesian coordinates, we can introduce the three main types of **symmetry** (*symmetrie*), namely symmetry about the x-axis, symmetry about the y-axis, and finally, symmetry about the origin.

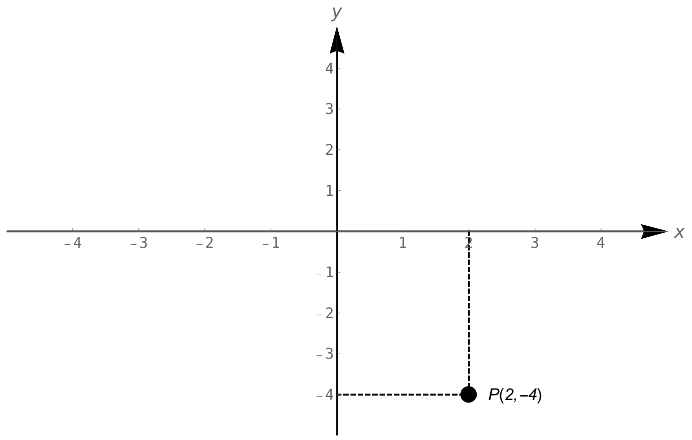


Figure 3.1: The point P located in the Cartesian coordinate plane.

3.2 Functions

3.2.1 Relations

Definitie 3.1 (Relation)
A **relation** (*relatie*) is a set of points in the plane. Hence, a relation R in $\mathbb{R}$ is a subset of the Cartesian product $\mathbb{R}^2$.

Since relations are sets, we can describe them using the techniques presented in Chapter 2. That is, we can describe a relation verbally, using the roster method, or using set-builder notation. Most frequently, the latter kind of description is preferred and a relation is defined using a specific predicate that depends on x and y, $P(x,y)$, i.e.

$$R = \left\{(x,y) \in \mathbb{R}^2 \mid P(x,y)\right\}.$$

The predicate $P(x,y)$ is the rule that allows us to select the ordered pairs (x,y) that make up the relation. Here, we call x the **argument** (*argument*) of the relation R and y its corresponding **image** (*beeld*). Since the elements in a relation are points in the plane, we often try to describe the relation graphically as well. Doing so produces the **graph** (*grafiek*) of the relation R.

Example 3.1

Graph the following relations.

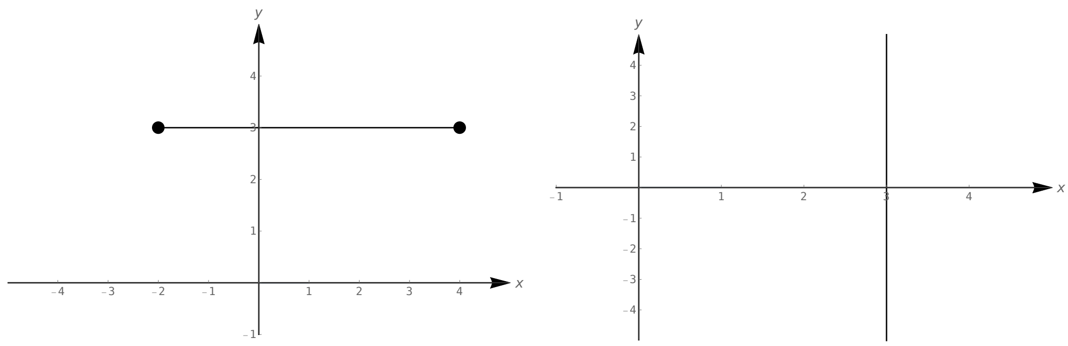
1. $B = \{(x,3) \mid -2 \leq x \leq 4\}$
2. $C = \{(3,y) \mid y \text{ is a real number}\}$

Solution

1. In words, $\{(x,3) \mid -2 \leq x \leq 4\}$ reads 'the set of points $(x,3)$ such that $-2 \leq x \leq 4$. Plotting several representative points should convince you that B describes the horizontal line segment from the point $(-2,3)$ up to and including the point $(4,3)$ (Figure 3.2(a)).

2. The relation C is described as the set of points $(3,y)$ such that y is a real number. All of these points have an x-coordinate of 3, but the y-coordinate is free to be whatever it wants to be,

without restriction. Hence, all the points of C lie on the vertical line $x = 3$ (Figure 3.2(b)).



(a) Set $B = \{(x,3) \,|\, -2 \leq x \leq 4\}$ (b) Set $C = \{(3,y) \,|\, y \text{ is a real number}\}$

Figure 3.2: Graphs of different relations.

The relation C in the previous example lead us to our final way to describe relations: **algebraically** (*algebraïsch*). We can more succinctly describe the points in C as those points which satisfy the equation $x = 3$. Let us now study the graphs of equations in a more general setting. For that purpose, we rely on the so-called fundamental graphing principle.

Definitie 3.2 (Fundamental graphing principle)
The graph of an equation is the set of points which satisfy the equation.

It is at this point that we gain some insight into the word 'relation'. If the equation to be graphed contains both x and y, then the equation itself is what is relating the two variables. For instance, in the next example, we consider the graph of the equation $x^2 + y^3 = 1$ by graphing the relation $R = \{(x,y) \,|\, x^2 + y^3 = 1\}$.

Example 3.2

Graph the equation $x^2 + y^3 = 1$.

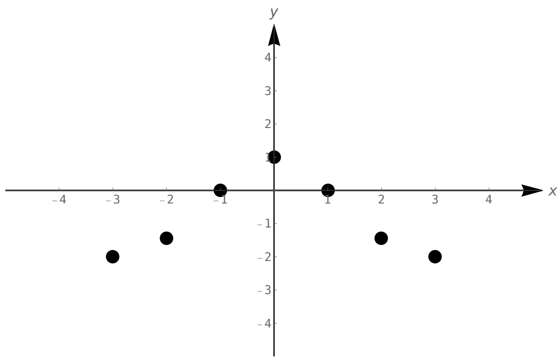
Solution

To efficiently generate points on the graph of this equation, we first solve this equation for y:

$$\begin{aligned} & x^2 + y^3 &= &\ 1 \\ \Leftrightarrow\quad & y^3 &= &\ 1 - x^2 \\ \Leftrightarrow\quad & \sqrt[3]{y^3} &= &\ \sqrt[3]{1 - x^2} \\ \Leftrightarrow\quad & y &= &\ \sqrt[3]{1 - x^2}\,. \end{aligned}$$

We now substitute a value in for x, determine the corresponding value y, and plot the resulting point (x,y) in the Cartesian coordinate plane. We first generate a table of points which are on the graph of the equation. These points are then plotted in the plane as shown below.

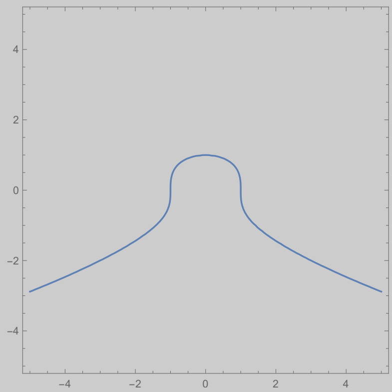
x	y
-3	-2
-2	$-\sqrt[3]{3}$
-1	0
0	1
1	0
2	$-\sqrt[3]{3}$
3	-2



Alternatively, we can construct a graph of this equation in Mathematica using the built-in function **`ContourPlot`**:

In[1]:=
```
ContourPlot[x^2+y^3==1,{x,-5,5},{y,-5,5}]
```

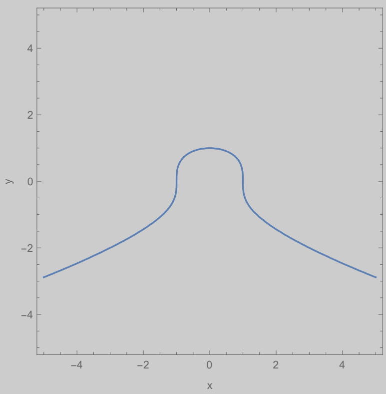
Out[1]=



Of course, we should add frame labels to this plot, which can be done as follows.

In[2]:=
```
ContourPlot[x^2+y^3==1,{x,-5,5},{y,-5,5}, FrameLabel →{"x","y"}]
```

Out[2]=



The places where the graph of an equation crosses or touches the axes are called the **intercepts** (*intercept*).

Another fact which you may have noticed about the graph in Example 3.2 is that it seems to be symmetric about the y-axis. To actually prove this analytically, we assume (x, y) is a generic point on the graph of the equation. That is, we assume $x^2 + y^3 = 1$ is true. To show that the graph as a whole is symmetric about the y-axis, we need to show that $(-x, y)$ satisfies the equation $x^2 + y^3 = 1$, too.

Substituting $(-x, y)$ into the equation gives

$$\begin{array}{rrcl} & (-x)^2 + (y)^3 & \stackrel{?}{=} & 1 \\ \Leftrightarrow & x^2 + y^3 & \stackrel{\checkmark}{=} & 1. \end{array}$$

Since we are assuming that the original equation $x^2 + y^3 = 1$ is true, we have shown that $(-x, y)$ satisfies the equation and hence is on the graph. In this way, we can check whether the graph of a given equation possesses any of the symmetries introduced in Section 3.1.

- About the y-axis: substitute $(-x, y)$ into the equation and verify whether or not the result is equivalent to the original equation.
- About the x-axis: substitute $(x, -y)$ into the equation and verify whether or not the result is equivalent to the original equation.
- About the origin: substitute $(-x, -y)$ into the equation and verify whether or not the result is equivalent to the original equation.

Intercepts and symmetry are two tools which can help us sketch the graph of an equation analytically, as demonstrated in the next example.

Example 3.3

Find the x- and y-intercepts (if any) of the graph of $(x-2)^2 + y^2 = 1$. Test for symmetry. Plot additional points as needed to complete the graph.

Solution

To look for x-intercepts, we set $y = 0$ and solve

$$\begin{array}{rrcl} & (x-2)^2 + 0^2 & = & 1 \\ \Leftrightarrow & \sqrt{(x-2)^2} & = & \sqrt{1} \\ \Leftrightarrow & x-2 & = & \pm 1 \\ \Leftrightarrow & x = 3 & \wedge & x = 1. \end{array}$$

We get two answers for x which correspond to two x-intercepts: $(1, 0)$ and $(3, 0)$. Turning our attention to y-intercepts, we set $x = 0$ and solve

$$\begin{array}{rrcl} & (0-2)^2 + y^2 & = & 1 \\ \Leftrightarrow & 4 + y^2 & = & 1 \\ \Leftrightarrow & y^2 & = & -3. \end{array}$$

Consequently, the graph has no (real-valued!) y-intercepts.

Moving along to symmetry, we can immediately dismiss the possibility that the graph is symmetric about the y-axis or the origin. If the graph possessed either of these symmetries, then the fact that $(1, 0)$ is on the graph would mean $(-1, 0)$ would have to be on the graph. The only symmetry left to test is symmetry about the x-axis. To that end, we substitute $(x, -y)$ into the equation and simplify

$$\begin{array}{rrcl} & (x-2)^2 + (-y)^2 & = & 1 \\ \Leftrightarrow & (x-2)^2 + y^2 & = & 1. \end{array}$$

Since we have obtained our original equation, the graph is symmetric about the x-axis. Proceeding as we did in the Example 3.2, we obtain the graph shown in Figure 3.3.

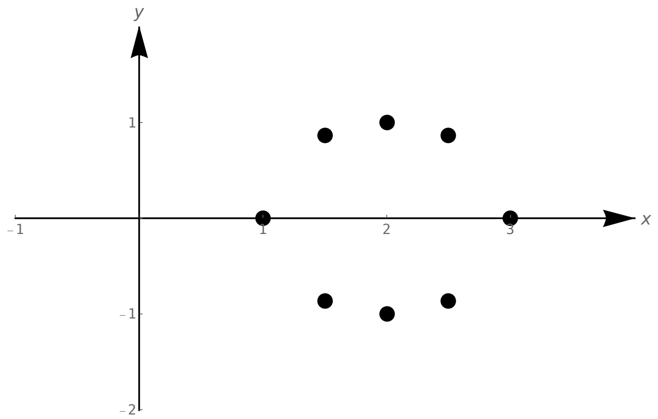


Figure 3.3: Graph of the equation $(x-2)^2+y^2=1$.

3.2.2 Functions in $\mathbb{R}$

3.2.2.1 Definition

Definitie 3.3 (Function)
A relation in which each x-coordinate in $\mathbb{R}$ is matched with at most only one y-coordinate is said to describe y as a **function** (*functie*) of x. A function f that maps x to y is denoted as

$$f : x \mapsto f(x)\,,$$

with $y = f(x)$.

The notation $y = f(x)$ (read: y equals f of x) means that the pair (x,y) belongs to the set of pairs defining the function f. x is called the **argument** or **input** (*input*) of the function f and $f(x)$ is called the **value** taken by the function when evaluated at a point x, or its **output** (*output*) or **image**. In the framework of applications, x is often called the **independent variable** (*onafhankelijke veranderlijke*), while y is called the **dependent variable** (*afhankelijke veranderlijke*). Loosely speaking, a function may be envisaged as a black box that returns for each input a corresponding output (Figure 3.4).

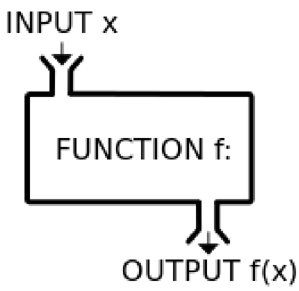


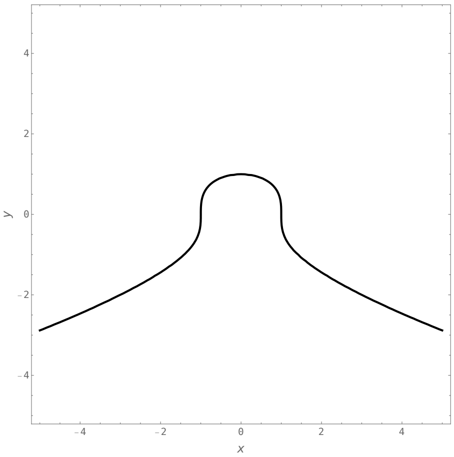
Figure 3.4: Describing a function as a black box.

To (graphically) determine whether or not a relation is function, we can use the following theorem, which is an immediate consequence of Definition 3.3.

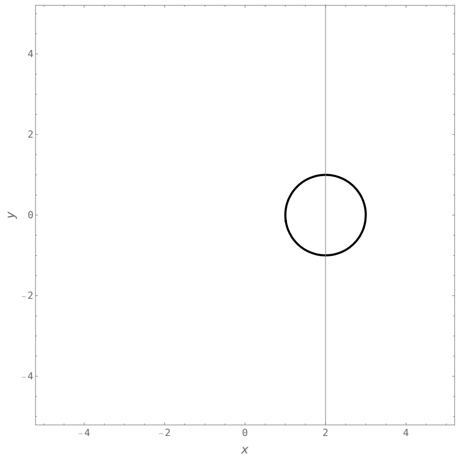
> **Theorem 3.1 (Vertical line test)**
> *A set of points in the plane represents y as a function of x if and only if no two points lie on the same vertical line.*

Example 3.4

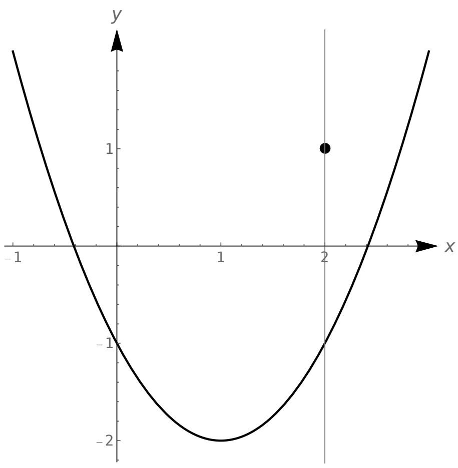
Determine graphically which of the following relations actually represents a function.



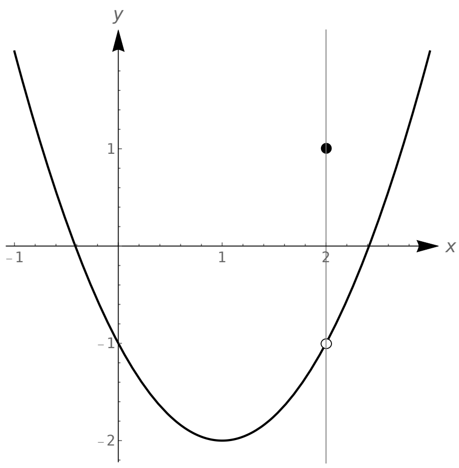
(a) Graph of relation A



(b) Graph of relation B



(c) Graph of relation C



(d) Graph of relation D

Solution

(a) In the graph of A, every vertical line crosses the graph at most once, so A does represent y as a function of x.

(b) Looking at the graph of B, we can easily imagine a vertical line crossing the graph more than once. Hence, B does not represent y as a function of x.

(c) In C, there is a point on the curve with x-coordinate 2 just below $(2, 1)$, which means that both $(2, 1)$ and this point on the curve lie on the vertical line $x = 2$. Hence, the graph of C fails the vertical line test, so y is not a function of x here.

(d) In D notice that the point with x-coordinate 2 on the curve has been omitted, leaving an open circle there. Hence, the vertical line $x = 2$ crosses the graph of D only at the point $(2,1)$. Indeed, any vertical line will cross the graph at most once, so we have that the graph of D passes the vertical line test. Thus it describes y as a function of x.

Finally, it is important to note that a function for which the dependent variable can be written explicitly in terms of the independent variable, i.e. as $y = f(x)$, is called an **explicit function** (*expliciete functie*). On the other hand, if the function is defined by

$$F(x,y) = 0\,,$$

which does not contain y explicitly at one side of the equation, it is called an **implicit function** (*impliciete functie*). For instance,

$$x^2 + y^2 - 4 = 0$$

constitutes an implicit function, which defines two explicit functions, namely

$$y = \sqrt{4 - x^2}, \qquad \text{and} \qquad y = -\sqrt{4 - x^2}\,.$$

3.2.2.2 Function graphs

It is often useful to draw the graph of a function for getting a global view of its properties. Formally, we define the graph of a function as follows.

Definitie 3.4 (Fundamental graphing principle for functions)
The graph G of a function f is the set of points which satisfy the equation $y = f(x)$. More formally,

$$G = \left\{(x, f(x)) \mid x \in \mathbb{R} \wedge y = f(x)\right\}\,.$$

The x-coordinates of the x-intercepts of the graph of $y = f(x)$ can be found by solving $f(x) = 0$. For this reason, they are called the **zeros** (*nulpunt*) of f.

3.2.2.3 Domain, codomain and range

When defining a function f as

$$f : x \mapsto f(x)\,,$$

we not only need to specify how f maps an argument x to its image y, but also the sets to which x and $y = f(x)$ belong. In other words, we need to specify what are the possible in- and outputs of our black box (Figure 3.4). We call the set of possible inputs the **domain** (*domein*) of the function, denoted as $\operatorname{dom} f$. The set of possible outputs is called the **codomain** (*codomein*). Suppose X is the domain of a function f and Y is its codomain, then we can define this function using arrow notation to make explicit the domain and codomain:

$$f : X \to Y\,,$$

which should be read as f maps domain X to codomain Y. The set to which f actually maps the domain, is called the **range** (*bereik*) or **image** (*beeld*) of f, denoted by $\operatorname{im} f$. The range is thus defined as

$$\operatorname{im} f = \left\{y = f(x) \mid x \in X \wedge y \in Y\right\}\,.$$

These concepts are illustrated in Figure 3.5. The difference between range and codomain may seem subtle, but can be very important, as will be shown in Example 3.5.

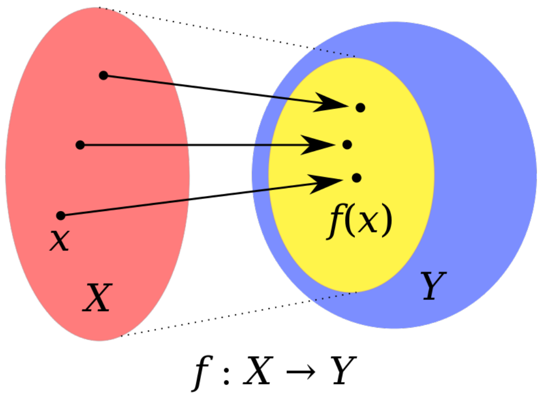


Figure 3.5: The domain X, codomain Y and range $\{f(x) \mid x \in X\}$ of a function f.

Throughout this course we will focus on so-called **real functions** (*reële functie*), which are real-valued functions of a real variable. Such functions can be written as

$$f : \mathbb{R} \to \mathbb{R} \quad : \quad x \mapsto f(x)\,.$$

In most cases, we can relatively easily determine the domain, codomain and range of a function f by investigating the formula defining it. Doing so, we sometimes run into functions whose domain consists of two ore more consecutive open intervals, such as for

$$f(x) = \frac{1}{x}\,,$$

for which $\operatorname{dom} f = \mathbb{R} \setminus \{0\} =]-\infty, 0[\cup]0, +\infty[$. The points at which a function in such a case is not defined are called the function's **singularities** (*singulariteiten*). So, $f(x) = \frac{1}{x}$ has a singularity at $x = 0$.

Example 3.5

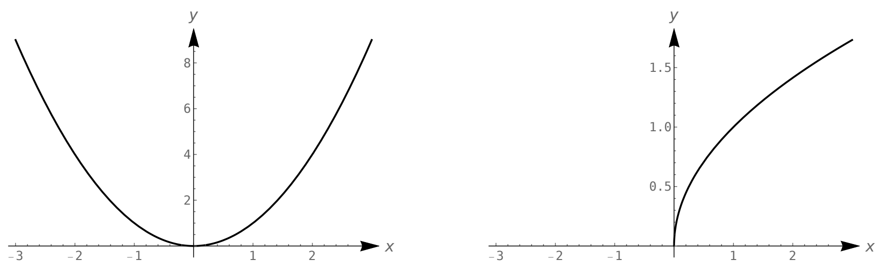
Determine the domain, codomain and range of the following functions

1. $f : \mathbb{R} \to \mathbb{R} \quad : \quad x \mapsto x^2\,,$
2. $g : \mathbb{R} \to \mathbb{R}^+ \quad : \quad x \mapsto x^2\,,$
3. $h : \mathbb{R}^+ \to \mathbb{R}^+ \quad : \quad x \mapsto x^{\frac{1}{2}}\,.$

Solution

1. From the function definition of f, we infer that both the domain and codomain are $\mathbb{R}$. Yet, since f does not map to any negative number, the range of f is $\mathbb{R}^+$. Its graph is shown in Figure 3.6(a).
2. From the function definition of g, we infer that the domain is $\mathbb{R}$, but its codomain is $\mathbb{R}^+$. Again, since g does not map to any negative number, the range of g is $\mathbb{R}^+$ and its graph is the same as the of f shown in Figure 3.6(b).

3. From the function definition of h, we infer that both the domain and codomain are $\mathbb{R}^+$. Since the mapping is done by the square root, this has important implications. Firstly, the square root of a negative number does not exist in $\mathbb{R}$, which sets a restriction on the domain. Secondly, the square root of $x \in \mathbb{R}^+$ maps x to $\sqrt{x}$ and $-\sqrt{x}$. Hence, if the codomain would be defined as $\mathbb{R}$, h would not be a function (see Definition 3.3)! From this we can also infer that the range of h is $\mathbb{R}^+$. Its graph is shown in Figure 3.6(b).



(a) Graph of f and g (b) Graph of h

Figure 3.6: Graphs of the functions f, g and h in Example 3.5.

3.2.3 Function arithmetic

It seems natural that functions should have their own arithmetic which is consistent with the arithmetic of real numbers. The following definitions allow us to add, subtract, multiply and divide functions using the arithmetic we already know for real numbers. Suppose f and g are functions and $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$, then we can define the following operations on $\operatorname{dom} f \cap \operatorname{dom} g$:

- The **sum** (*som*) of f and g:
$$(f+g)(x) = f(x) + g(x).$$
- The **difference** (*verschil*) of f and g:
$$(f-g)(x) = f(x) - g(x).$$
- The **product** (*product*) of f and g:
$$(fg)(x) = f(x)g(x).$$
- The **quotient** (*quotiënt*) of f and g:
$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)},$$
provided $g(x) \neq 0$.

Note that while the formula $(f+g)(x) = f(x) + g(x)$ looks suspiciously like some kind of distributive property, it is nothing of the sort; the addition on the left hand side of the equation is function addition, and we are using this equation to define the output of the new function $f+g$ as the sum of the real number outputs from f and g.

Example 3.6
Let $f(x) = 6x^2 - 2x$ and $g(x) = 3 - \frac{1}{x}$.

1. Find the domain of $g-f$. Then find and simplify a formula for $(g-f)(x)$.
2. Find the domain of g/f. Then find and simplify a formula for $\left(\frac{g}{f}\right)(x)$.

Solution

1. To find the domain of $g-f$ we need to find the domain of g and of f separately, then find the intersection of these two sets. Owing to the denominator in the expression $g(x) = 3-\frac{1}{x}$, we get that the domain of g is $\mathbb{R}_0$. Since $f(x) = 6x^2 - 2x$ is valid for all real numbers, we have no further restrictions. Thus the domain of $g-f$ matches the domain of g, namely, $\mathbb{R}_0$.

   Moving along, we need to simplify a formula for $(g-f)(x)$. In this case, we get common denominators and attempt to reduce the resulting fraction. Doing so, we get

$$\begin{aligned}(g-f)(x) &= g(x)-f(x)\\ &= \left(3-\frac{1}{x}\right)-\left(6x^2-2x\right)\\ &= \frac{-6x^3+2x^2+3x-1}{x}.\end{aligned}$$

2. First, we find the domain of g and f separately, and find the intersection of these two sets. In addition, since $\left(\frac{g}{f}\right)(x) = \frac{g(x)}{f(x)}$, we are introducing a new denominator, namely $f(x)$, so we need to guard against this being 0 as well. The domain of g is $\mathbb{R}_0$ and the domain of f is $\mathbb{R}$. Setting $f(x) = 0$ gives $6x^2 - 2x = 0$ or $x = 0$ or $x = \frac{1}{3}$. So, the domain of g/f is $\mathbb{R} \setminus \{0, \frac{1}{3}\}$.

   Next, we find and simplify a formula for $\left(\frac{g}{f}\right)(x)$.

$$\begin{aligned}\left(\frac{g}{f}\right)(x) &= \frac{g(x)}{f(x)}\\ &= \frac{3-\frac{1}{x}}{6x^2-2x} = \frac{3x-1}{\left(6x^2-2x\right)x}\\ &= \frac{3x-1}{2x^2(3x-1)} \qquad \left(\text{Assuming } x \neq \frac{1}{3} \text{ and } x \neq 0\right)\\ &= \frac{1}{2x^2}\end{aligned}$$

In addition to operations on functions, we can also compose functions. Function composition is defined below.

Definitie 3.5 (Composite function)
Suppose f and g are two functions. The **composite** (*samenstelling*) of g with f, denoted $g \circ f$, is defined by

$$(g \circ f)(x) = g(f(x)),$$

provided $x \in \operatorname{dom} f$ and $f(x) \in \operatorname{dom} g$.

The quantity $g \circ f$ is also read g composed with f or, more simply g of f. At its most basic level, Definition 3.5 tells us to obtain the formula for $(g \circ f)(x)$, we replace every occurrence of x in the formula for $g(x)$ with the formula we have for $f(x)$. If we take a step back and look at this from a procedural, inputs and outputs perspective, Defintion 3.5 tells us the output from $g \circ f$ is found by taking the output from f, $f(x)$, and then making that the input to g. The result, $g(f(x))$, is the output from $g \circ f$. This is illustrated in Figure 3.7 for a setting where $f : x \mapsto x^2$ and $g : x \mapsto x+1$. Clearly, the notion of function composition can easily be generalised to an arbitrary number of functions. For instance, suppose f, g and h are three functions, then we may consider

$$\big((h \circ g) \circ f\big)(x) = h\Big(g\big(f(x)\big)\Big).$$

In the expression $g(f(x))$, the function f is often called the inside function while g is often called the outside function. There are two ways to go about evaluating composite functions - inside out and outside in - depending on which function we replace with its formula first. Both ways are demonstrated in the following example.

Example 3.7

Let $f(x) = x^2 - 4x$, $\quad g(x) = 2 - \sqrt{x+3}$ and $h(x) = \dfrac{2x}{x+1}$.

Find and simplify the indicated composite functions. State the domain of each.

1. $(g \circ f)(x)$
2. $\big(h \circ (g \circ f)\big)(x)$

Solution

1. By definition, $(g \circ f)(x) = g(f(x))$. We now illustrate two ways to approach this problem.

   - Inside out: We insert the expression $f(x)$ into g first to get

   $$(g \circ f)(x) = g(f(x)) = g\left(x^2 - 4x\right) = 2 - \sqrt{\left(x^2 - 4x\right) + 3} = 2 - \sqrt{x^2 - 4x + 3}.$$

   - Outside in: We use the formula for g first to get

   $$(g \circ f)(x) = g(f(x)) = 2 - \sqrt{f(x) + 3} = 2 - \sqrt{\left(x^2 - 4x\right) + 3} = 2 - \sqrt{x^2 - 4x + 3}.$$

   We get the same answer as before.

   To find the domain of $g \circ f$, we need to find the elements in the domain of f whose outputs $f(x)$ are in the domain of g. We accomplish this by determining the domain before we simplify the formula for the composite function. To that end, we examine $(g \circ f)(x) = 2 - \sqrt{\left(x^2 - 4x\right) + 3}$. To ensure that the argument of the square root is positive, it should hold that $x^2 - 4x + 3 \geq 0$. We find the zeros of $x^2 - 4x + 3$ to be $x = 1$ and $x = 3$. Consequently, the domain of $g \circ f$, is $]-\infty, 1] \cup [3, +\infty[$. Figure 3.8 shows the graph of $f(x)$, $g(x)$ and $(g \circ f)$.

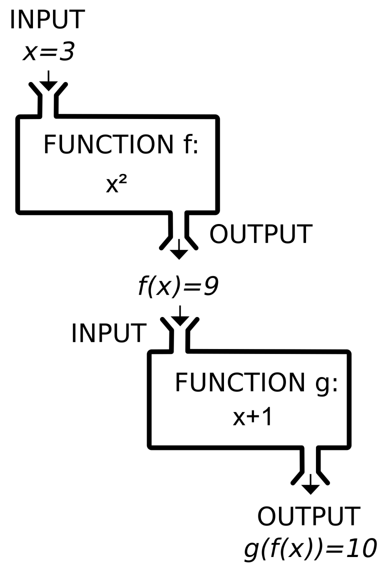


Figure 3.7: Function composition $(g \circ f)(x)$, where $f : x \mapsto x^2$ and $g : x \mapsto x+1$.

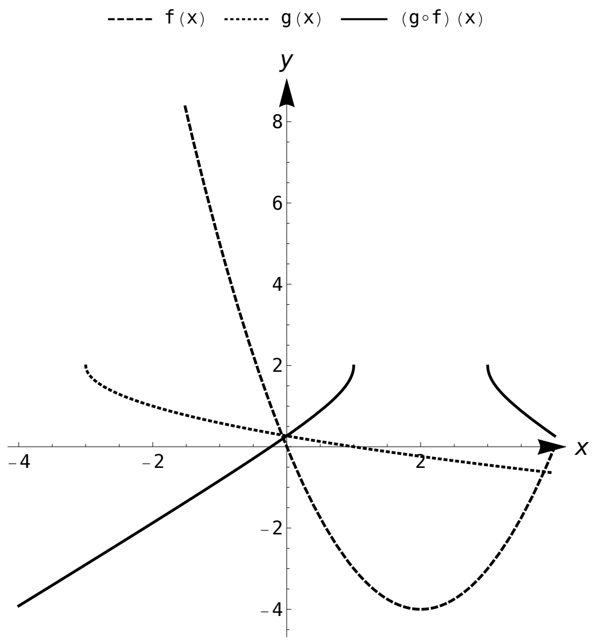


Figure 3.8: Graph of $f(x)$, $g(x)$ and $(g \circ f)$ in Example 3.7.

2. The expression $\big(h \circ (g \circ f)\big)(x)$ indicates that we first find the composite, $g \circ f$ and compose the function h with the result. We know already that $(g \circ f)(x) = 2 - \sqrt{x^2 - 4x + 3}$. We now proceed as usual.

   - Inside out: We insert the expression $(g \circ f)(x)$ into h first to get

$$\begin{aligned} \big(h \circ (g \circ f)\big)(x) &= h\big((g \circ f)(x)\big) = h\left(2 - \sqrt{x^2 - 4x + 3}\right) \\ &= \frac{2\left(2 - \sqrt{x^2 - 4x + 3}\right)}{\left(2 - \sqrt{x^2 - 4x + 3}\right) + 1} = \frac{4 - 2\sqrt{x^2 - 4x + 3}}{3 - \sqrt{x^2 - 4x + 3}}. \end{aligned}$$

- Outside in: We use the formula for $h(x)$ first to get

$$\begin{aligned}\big(h\circ(g\circ f)\big)(x) &= h\big((g\circ f)(x)\big)=\frac{2\big((g\circ f)(x)\big)}{\big((g\circ f)(x)\big)+1}\\ &= \frac{2\left(2-\sqrt{x^2-4x+3}\right)}{\left(2-\sqrt{x^2-4x+3}\right)+1}=\frac{4-2\sqrt{x^2-4x+3}}{3-\sqrt{x^2-4x+3}}.\end{aligned}$$

To find the domain of $(h\circ(g\circ f))$, we look at the step before we began to simplify,

$$(h\circ(g\circ f))(x)=\frac{2\left(2-\sqrt{x^2-4x+3}\right)}{\left(2-\sqrt{x^2-4x+3}\right)+1}.$$

For the square root, we need $x^2-4x+3\geq 0$, which requires $]-\infty,1]\cup[3,+\infty[$. Next, we set the denominator to zero and solve: $\left(2-\sqrt{x^2-4x+3}\right)+1=0$. We get $\sqrt{x^2-4x+3}=3$, and, after squaring both sides, we have $x^2-4x+3=9$. To solve $x^2-4x-6=0$, we use the quadratic formula and get $x=2\pm\sqrt{10}$. Hence we must exclude these numbers from the domain of $h\circ(g\circ f)$. Consequently our final domain for $h\circ(f\circ g)$ is

$$\left]-\infty,2-\sqrt{10}\right[\cup\left]2-\sqrt{10},1\right]\cup\left[3,2+\sqrt{10}\right[\cup\left]2+\sqrt{10},\infty\right[.$$

.

From this example, we learn that function composition is not commutative, so $(g\circ f)(x)\neq(f\circ g)(x)$, though it is associative, i.e. it holds that

$$\big((h\circ g)\circ f\big)(x)=\big(h\circ(g\circ f)\big)(x).$$

Also note the importance of finding the domain of the composite function before simplifying.

3.2.4 Function properties

When graphing functions, we will typically investigate whether or not there is some kind of symmetry to its graph, whether or not it is periodic, and so on.

3.2.4.1 Injections, surjections and bijections

Injections, surjections and bijections are classes of functions distinguished by the manner in which arguments and images are mapped to each other.

Definitie 3.6 (Injective, surjective and bijective)
A function $f: X\to Y$ is called

- **injective** (*injectief*) (one-to-one) if each element of the codomain is mapped to by at most one element of the domain. Mathematically, we may write:

$$\forall x_1,x_2\in\operatorname{dom}f: f(x_1)=f(x_2)\Rightarrow x_1=x_2.$$

  This implies that two different elements belonging to the function's domain cannot be mapped to the same element in its codomain, so every straight line parallel to the x-axis intersects the graph of f in at most one point. An injective function is an **injection** (*injectie*);

- **surjective** (*surjectief*) (onto) if each element of the codomain is mapped to by at least one element of the domain. That is, the range and the codomain of the function are equal. This implies that every straight line parallel to the x-axis intersects the graph of f in at least one point. A surjective function is a **surjection** (*surjectie*); and
- **bijective** (*bijectief*) (one-to-one correspondence) if each element of the codomain is mapped to by exactly one element of the domain. So, the function is both injective and surjective. This implies that every straight line parallel to the x-axis intersects the graph of f in at exactly one point. A bijective function is a **bijection** (*bijectie*).

The four possible combinations of injective and surjective features are illustrated in Table 3.1. An injective function does not need to be surjective because not all elements of the codomain may be associated with arguments, and likewise a surjective function does not need to be injective as some images may be associated with more than one argument.

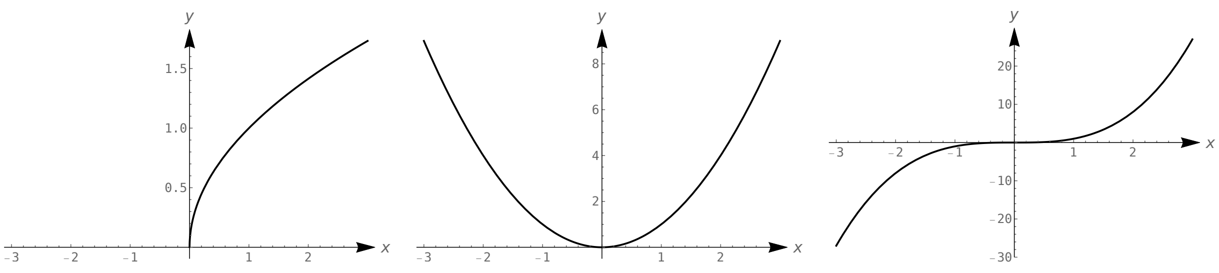
Example 3.8

Determine whether the following real functions are injections, surjections and/or bijections.

1. $f : \mathbb{R}^+ \to \mathbb{R} : x \mapsto \sqrt{x}$
2. $g : \mathbb{R} \to \mathbb{R} : x \mapsto x^2$
3. $h : \mathbb{R} \to \mathbb{R} : x \mapsto x^3$

Solution Figure 3.9 depicts the graphs of the considered functions, which were generated in Mathematica.

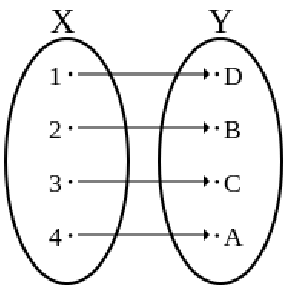
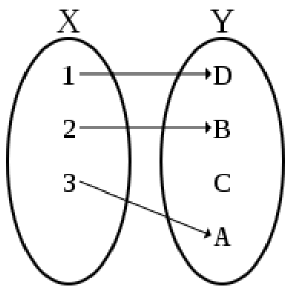
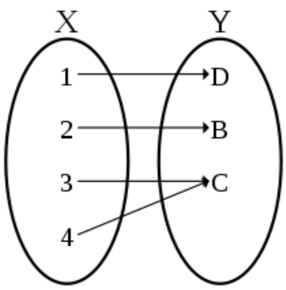
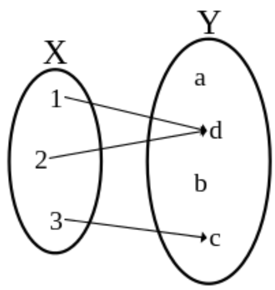
1. For any two elements x_1 and x_2 of this function's domain, it holds that if $\sqrt{x_1} = \sqrt{x_2}$, then $x_1 = x_2$. This means that every straight line parallel to the x-axis intersects with the function's graph in at most one point (Figure 3.9(a)), so the function f is an injection. Since the range of this function is restricted to $\mathbb{R}^+$, whereas its codomain is $\mathbb{R}$, this function is non-surjective, and hence it cannot be a bijection.
2. Since we have for any two elements x_1 and x_2 of this function's domain that if $x_1^2 = x_2^2$ then $x_1 = \pm x_2$, this function is non-injective. Besides, it is neither a surjection because its range is $\mathbb{R}^+$, whereas its codomain is $\mathbb{R}$. Consequently, it is not a bijection (Figure 3.9(b)).
3. For any two elements x_1 and x_2 of this function's domain, it holds that if $x_1^3 = x_2^3$, then $x_1 = x_2$. So the function h is an injection. Moreover, it is also a surjection because its range is $\mathbb{R}$; that is every straight line parallel to the x-axis intersects with the function's graph in at least one point. Since h is both an injection and surjection, it is a bijection (Figure 3.9(c)).



(a) Graph of f (b) Graph of g (c) Graph of h

Figure 3.9: Graphs of the functions f, g and h in Example 3.8.

Table 3.1: The four possible combinations of injective and surjective features.

	Surjective	Non-surjective
Injective		
Non-injective		

3.2.4.2 Symmetry

Of the three symmetries discussed in Section 3.2.1, only two are of significance to functions: symmetry about the y-axis and symmetry about the origin.

Definitie 3.7 (Even/odd functions)
A function f is

- an **even** (*even*) function if and only if $f(-x) = f(x)$ for all $x \in \text{dom} f$. The graph of f is symmetric about the y-axis.
- an **odd** (*oneven*) function if and only if $-f(-x) = f(x)$, or, equivalently, $f(-x) = -f(x)$ for all $x \in \text{dom} f$. The graph of f is symmetric about the origin.

Example 3.9

Determine analytically if the following functions are even, odd, or neither even nor odd. Verify your result with Mathematica.

1. $f(x) = \dfrac{5}{2-x^2}$
2. $g(x) = \dfrac{5x}{2-x^2}$

Solution

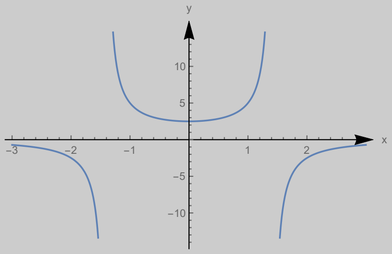
1. The first step is to replace x with $-x$ and simplify.

$$f(-x) = \frac{5}{2-(-x)^2} = \frac{5}{2-x^2} = f(x)$$

Hence, f is even. This conclusion can be verified in Mathematica using the function **`Plot`**.

```
In[3]:= Plot[5/(2-x^2),{x,-3,3},AxesLabel→{"x","y"},AxesStyle→Arrowheads[{0,0.05`}]]
```

Out[3]=

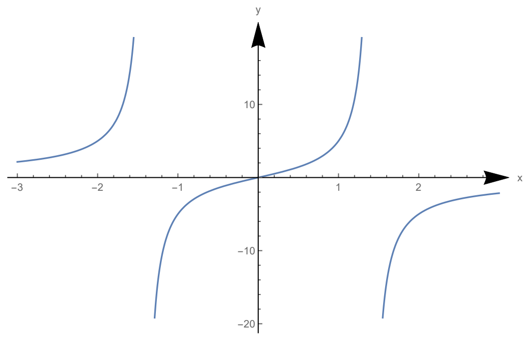


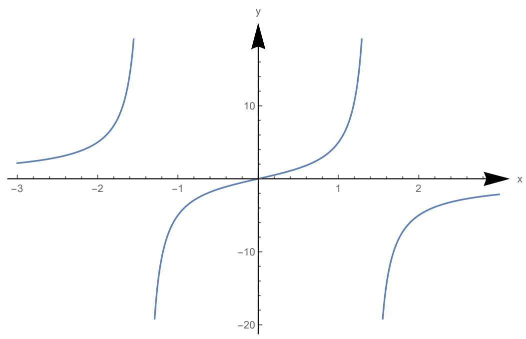
Here, the option **`AxesLabel`** was used to label the axes, while using the option `Arrowheads` we add arrowheads to these axes.

2. Again, we replace x with $-x$ and simplify.

$$g(-x) = \frac{5(-x)}{2-(-x)^2} = \frac{-5x}{2-x^2} = -g(x)$$

Clearly, g is odd. This is confirmed by the graph of this function generated in Mathematica:





3.2.4.3 Periodicity

Definitie 3.8 (Periodic function)
A function f is said to be **periodic** (*periodiek*) with **period** (*periode*) P ($P \in \mathbb{R}_0^+$), if

$$f(x+P) = f(x)\,,$$

for all $x \in \operatorname{dom} f$. If there exists a least positive constant P with this property, it is called the **fundamental period**.

A function with period P will repeat on intervals of length P, and these intervals are referred to as **periods**.

3.2.4.4 Function behaviour

As you shall see in Chapters 4 and 5, each family of functions has its own unique attributes and we will study them all in great detail. The purpose of this section is to lay the foundation for that further study by investigating aspects of function behaviour which apply to all functions. To start, we will examine the concepts of **increasing** (*stijgend*), **decreasing** (*dalend*) and **constant** (*constant*). Before defining the concepts algebraically, it is instructive to first look at them graphically. For that purpose, consider the graph of a function f in Figure 3.10.

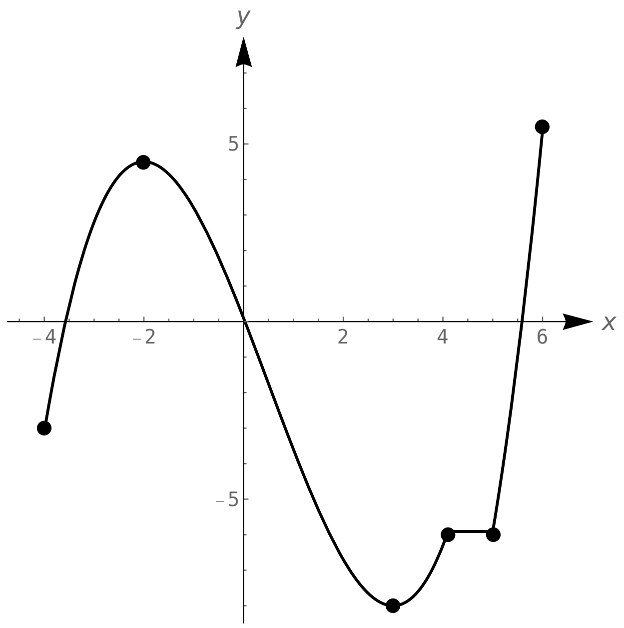


Figure 3.10: The graph of $y = f(x)$.

For the x values between -4 and -2 (inclusive), the y-coordinates on the graph are increasing, as we move from left to right. Hence the function f is increasing on the interval $[-4, -2]$. Analogously, we say that f is decreasing on the interval $[-2, 3]$, increasing once more on the interval $[3, 4]$, constant on $[4, 5]$, and finally increasing once again on $[5, 6]$.

Let us now introduce more formal algebraic definitions of what it means for a function to be increasing, decreasing or constant.

Definitie 3.9 (Function behaviour)
Suppose f is a function defined on an interval $I \subset \operatorname{dom} f$. We say f is:

- **increasing** on I if and only if $f(x_1) \leq f(x_2)$ for all x_1, x_2 in I with $x_1 < x_2$.
- **decreasing** on I if and only if $f(x_1) \geq f(x_2)$ for all x_1, x_2 in I with $x_1 < x_2$.
- **constant** on I if and only if $f(x_1) = f(x_2)$ for all x_1, x_2 in I.

If the order $\leq$ in the definition of an increasing function is replaced by $<$, we say that f is **strictly increasing** (*strikt stijgend*) on the interval I, and likewise for a **strictly decreasing** (*strikt dalend*) function. Clearly, if f is either strictly increasing or decreasing on an interval I, it must hold that f is an injective function .

We say that functions are **monotonically** (*monotoon*) increasing or decreasing on the interval I if they are entirely non-decreasing or entirely non-increasing, respectively. For instance, a function that increases monotonically does not exclusively have to increase, it simply must not decrease (Figure 3.11(a)). On the other hand, a function is **strictly monotone** (*strikt monotoon*) on the interval I if it is either strictly increasing or decreasing on that interval.

Now let us turn our attention to a few of the points on the graph in Figure 3.10. Clearly, the point $(-2, 4.5)$ does not have the largest y-value of all of the points on the graph of f but $(-2, 4.5)$ is on the top of the hill between $x = -4$ and $x = 3$. We say that the function f has a **local maximum** (*lokaal maximum*) at the point $(-2, 4.5)$, because the y-coordinate 4.5 is the largest y-value (hence, function value) on the curve near $x = -2$. Similarly, we say that the function f has a **local minimum** (*lokaal minimum*) at the point $(3, -8)$, since the y-coordinate -8 is the smallest function value near $x = 3$.

If we look at the entire graph, we see that the largest y-value (the largest function value) is 5.5 at $x = 6$. In this case, we say the **maximum** (*maximum*) of f is 5.5, sometimes also called the absolute or global maximum. Similarly, the **minimum** (*minimum*) of f is -8. This is also sometimes referred to as the absolute or global maximum.

We formalize these concepts in the following definitions.

Definitie 3.10 (Local and global extrema)
Suppose f is a function with $f(a) = b$.

- We say f has a **local maximum** at the point (a, b) if and only if there is an open interval I containing a for which $f(a) \geq f(x)$ for all x in I. The value $f(a) = b$ is called a local maximum value of f in this case.
- We say f has a **local minimum** at the point (a, b) if and only if there is an open interval I containing a for which $f(a) \leq f(x)$ for all x in I. The value $f(a) = b$ is called a local minimum value of f in this case.
- The value b is called the **(absolute or global) maximum** of f if $b \geq f(x)$ for all x in the domain of f. We may write
$$b = \max f.$$
- The value b is called the **(absolute or global) minimum** of f if $b \leq f(x)$ for all x in the domain of f. We may write
$$b = \min f.$$

It is important to note that not every function will have all of these features. Indeed, it is possible to have a function with no local or absolute extrema at all!

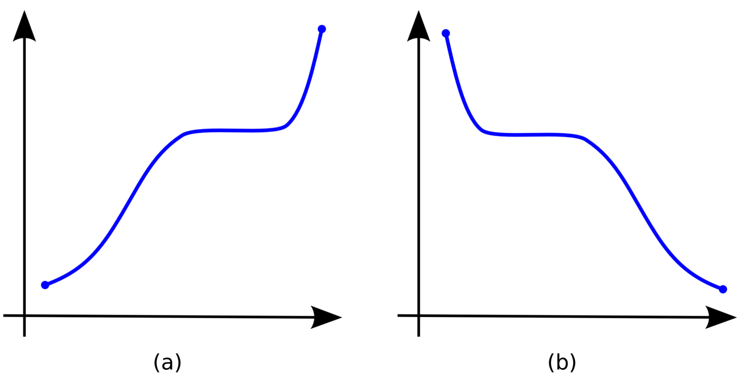


Figure 3.11: Graph of a monotonically increasing (a) and decreasing (b) function.

3.2.5 Transformations

We may change or transform the graphs of functions by making certain modifications to their formulas. The transformations we will study fall into three broad categories: shifts, reflections and scalings. Suppose the graph in Figure 3.12 is the complete graph of a function f.

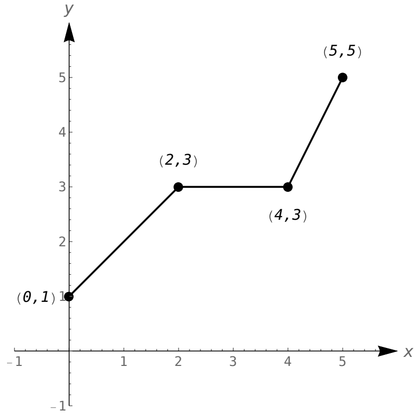


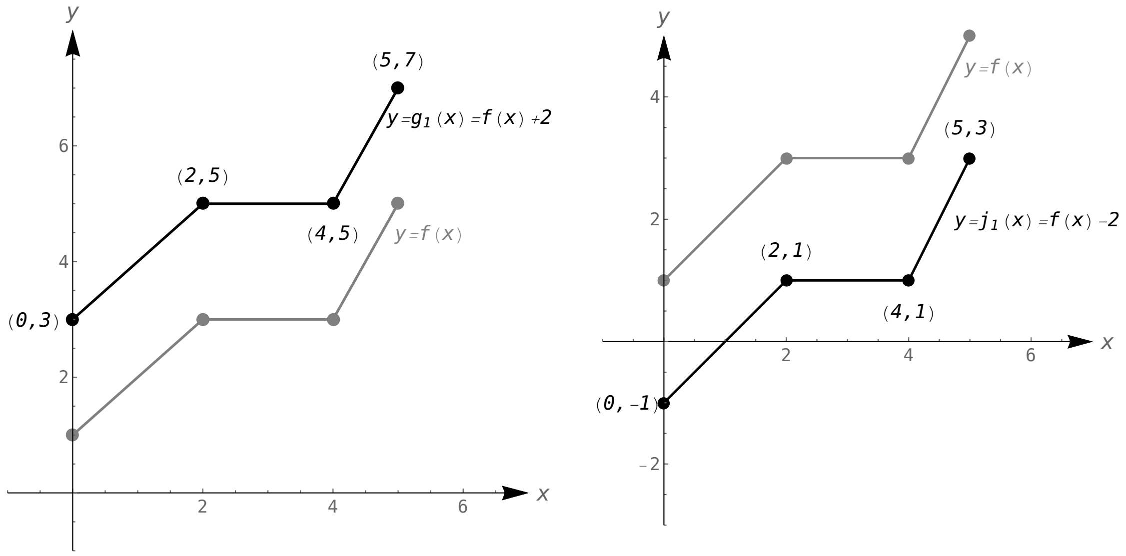
Figure 3.12: The graph of $y = f(x)$.

Suppose we wanted to graph the function defined by the formula $g_1(x) = f(x) + 2$. In order to graph g_1, we need to graph the points $(x, g_1(x))$. For example, using the points indicated on the graph of f, we can make the following table.

x	$f(x)$	$g_1(x) = f(x) + 2$
0	1	3
2	3	5
4	3	5
5	5	7

Hence, to obtain the graph of g_1, we just add 2 to the y-coordinate of each point on the graph of f (Figure 3.13(a)). Geometrically, we are 'shifting the graph up 2 units'. It is important to note that the domain of f and the domain of g are the same, but that the range of f is $[1, 5]$ while the range of g_1 is $[3, 7]$. In general, **shifting a function vertically** (*verticale verschuiving*) like this will leave the domain unchanged, but could very well affect the range.

You can easily imagine what would happen if we wanted to graph the function $j_1(x) = f(x) - 2$. Geometrically, we would then shift the graph down 2 units (Figure 3.13(b)).



(a) $y = g_1(x) = f(x) + 2$ (b) $y = j_1(x) = f(x) - 2$

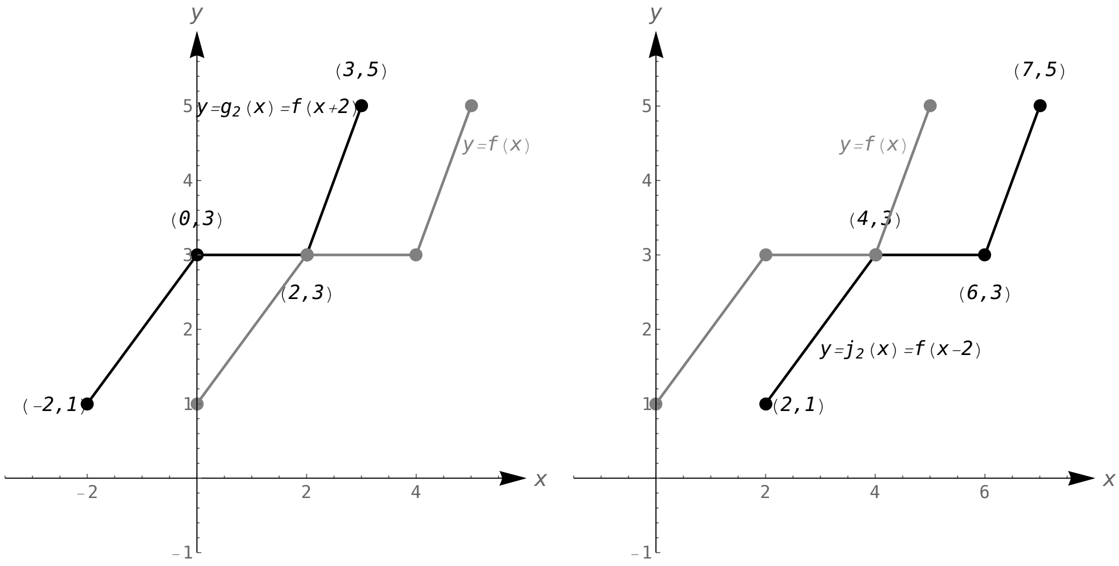
Figure 3.13: Vertical shifts of the graph of $y = f(x)$: two units up (a) and down (b).

Now, what happens if we add to or subtract from the input of the function? For instance, suppose we wanted to graph $g_2(x) = f(x+2)$. We know, for instance, $f(0) = 1$. To determine the corresponding point on the graph of g_2, we need to figure out what value of x we must substitute into $g_2(x) = f(x+2)$ so that the quantity $x+2$, works out to be 0. Solving $x+2 = 0$ gives $x = -2$, so $(-2,1)$ is on the graph of g_2. Continuing in this fashion, we get the following table.

x	$g_2(x)$	$=$	$f(x+2)$
-2	$g_2(-2)$	$=$	$f(0) = 1$
0	$g_2(0)$	$=$	$f(2) = 3$
2	$g_2(2)$	$=$	$f(4) = 3$
3	$g_2(3)$	$=$	$f(5) = 5$

In summary, the points $(0,1)$, $(2,3)$, $(4,3)$ and $(5,5)$ on the graph of $y = f(x)$ give rise to the points $(-2,1)$, $(0,3)$, $(2,3)$ and $(3,5)$ on the graph of $y = g_2(x)$, respectively. In general, if (a,b) is on the graph of $y = f(x)$, then $(a-2,b)$ is on the graph of $y = g_2(x)$. The point $(a-2,b)$ is exactly 2 units to the left of the point (a,b) so the graph of $y = g_2(x) = f(x+2)$ is obtained by shifting the graph $y = f(x)$ to the left 2 units (Figure 3.14(a)).

Note that while the ranges of f and g_2 are the same, the domain of g_2 is $[-2,3]$ whereas the domain of f is $[0,5]$. In general, when we **shift the graph horizontally** (*horizontale verschuiving*), the range will remain the same, but the domain could change. Similarly, if we set out to graph $j_2(x) = f(x-2)$, we would effect a shift to the right 2 units (Figure 3.14(b)).



(a) $y = g_2(x) = f(x+2)$

(b) $y = j_2(x) = f(x-2)$

Figure 3.14: Horizontal shifts of the graph of $y = f(x)$: two units to the left (a) and right (b).

We now turn our attention to **reflections** (*spiegeling*). We know from Section 3.1 that the graph of $y = -f(x)$ is the graph of f reflected across the x-axis (Figure 3.15(a)). Similarly, the graph of $y = f(-x)$ is the graph of f reflected across the y-axis (Figure 3.15(b)).

Finally, we turn our attention to our last class of transformations known as **scalings** (*schaling*). A thorough discussion of scalings can get complicated because they are not as straightforward as the previous transformations. The transformations covered so far are known as **rigid transformations** (*directe isometrie*) because they do not change the shape of the graph, only its position and orientation in the plane. If, however, we wanted to make a new graph twice as tall as a given graph, we would be changing the shape of the graph. This type of transformation is called **non-rigid** (*indirecte isometrie*). Not only will it be important for us to differentiate between modifying inputs versus outputs, we must also pay close attention to the magnitude of the changes we make.

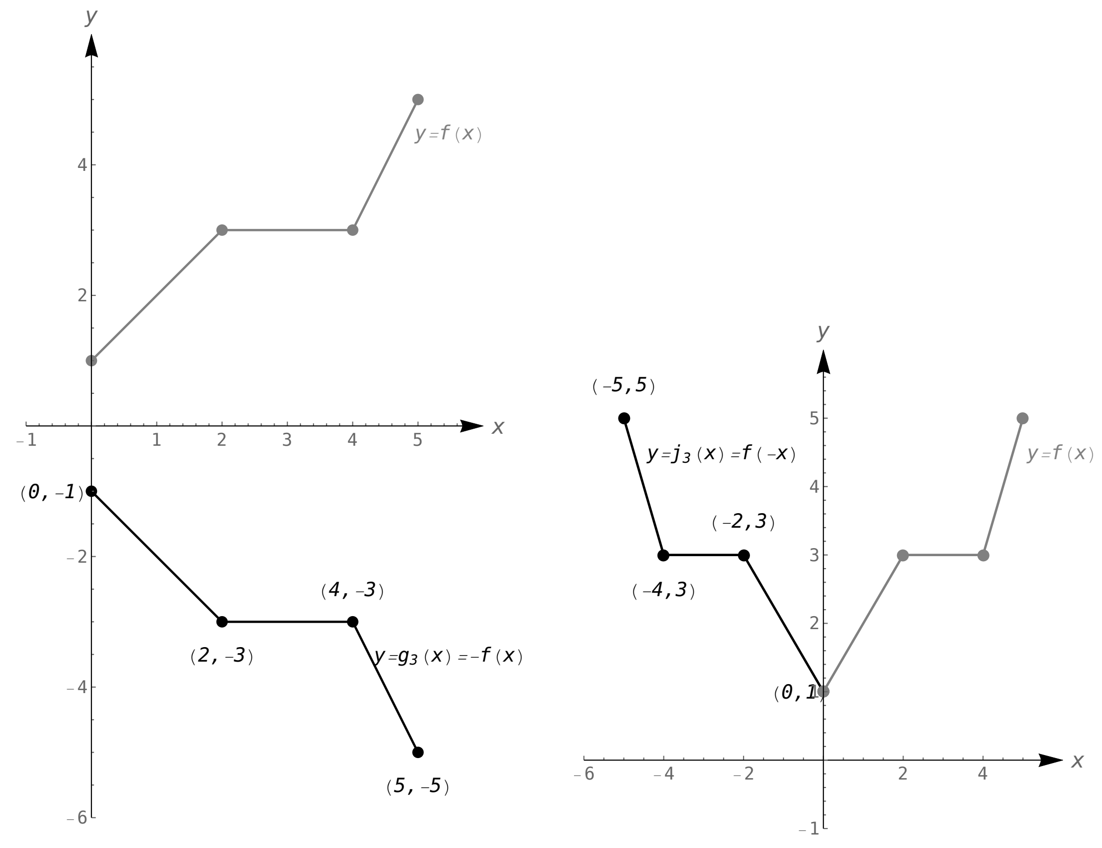
Suppose we wish to graph the function $g_4(x) = 2f(x)$. From its graph, we can build a table of values for g_4 as before.

x	$f(x)$	$g_4(x) = 2f(x)$
0	1	2
2	3	6
4	3	6
5	5	10

If (a, b) is on the graph of f, then $(a, 2b)$ is on the graph of g_4. In other words, to obtain the graph of g_4, we multiply all of the y-coordinates of the points on the graph of f by 2. This is known as a vertical scaling by a factor of 2 (Figure 3.16(a)). Likewise, if we wish to graph $y = \frac{1}{2}f(x)$, we multiply all of the y-coordinates of the points on the graph of f by $\frac{1}{2}$. This creates a vertical scaling by a factor of $\frac{1}{2}$ (Figure 3.16(b)).

In general, suppose f is a function and $a > 0$, then to obtain the graph of $y = af(x)$, we have to vertically scale the graph of f by a factor of a.

- If $a > 1$, we say the graph of f has undergone a **vertical stretching (expansion, dilation)** by a factor of a.



(a) $y = g_3(x) = -f(x)$ (b) $y = j_3(x) = f(-x)$

Figure 3.15: Reflections of the graph of $y = f(x)$: across x-axis (a) and y-axis (b).

- If $0 < a < 1$, we say the graph of f has undergone a **vertical shrinking (compression, contraction)** by a factor of $\frac{1}{a}$.

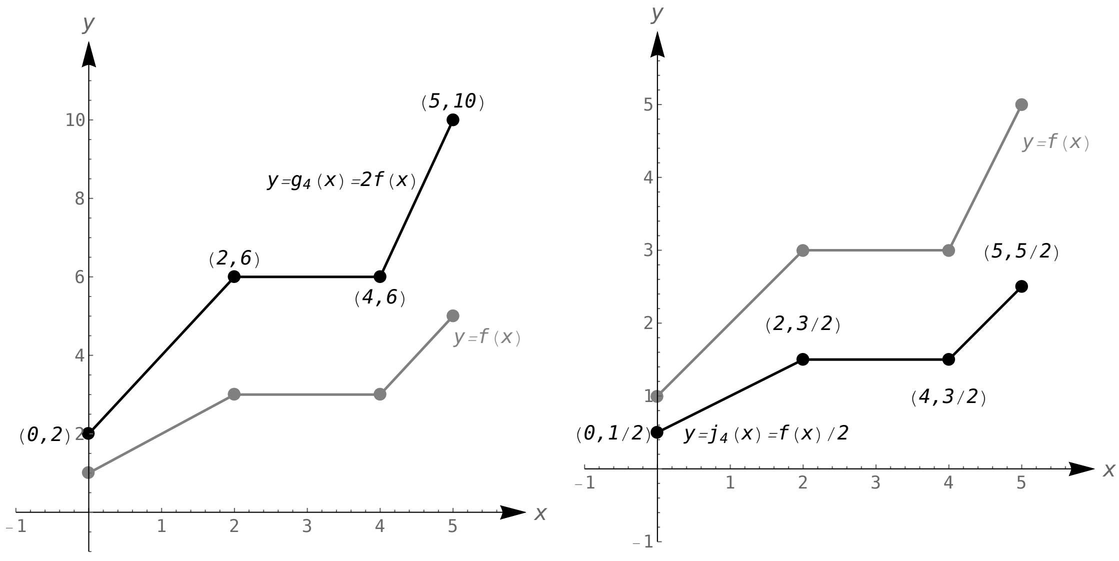
In terms of inputs and outputs, multiplying the outputs from a function by positive number a causes the graph to be vertically scaled by a factor of a. It is natural to ask what would happen if we multiply the inputs of a function by a positive number. This leads us to our last transformation.

Suppose we want to graph $g_5(x) = f(2x)$. If we want to determine the point on g_5 which corresponds to the point $(2,3)$ on the graph of f, we set $2x = 2$ so that $x = 1$. Substituting $x = 1$ into $g_5(x)$, we obtain $g_5(1) = f(2 \cdot 1) = f(2) = 3$, so that $(1,3)$ is on the graph of g_5. Continuing in this fashion, we obtain the following table.

x	$g_5(x)$	$=$	$f(2x)$
0	$g_5(0)$	$=$	$f(0) = 1$
1	$g_5(1)$	$=$	$f(2) = 3$
2	$g_5(2)$	$=$	$f(4) = 3$
$\frac{5}{2}$	$g_5(5/2)$	$=$	$f(5) = 5$

In general, if (a,b) is on the graph of f, then $\left(\frac{a}{2}, b\right)$ is on the graph of g. This results in a horizontal scaling by a factor of $\frac{1}{2}$ (Figure 3.17(a)). If, on the other hand, graphing $y = f\left(\frac{1}{2}x\right)$, results in a horizontal scaling by a factor of 2 (Figure 3.17(b)).

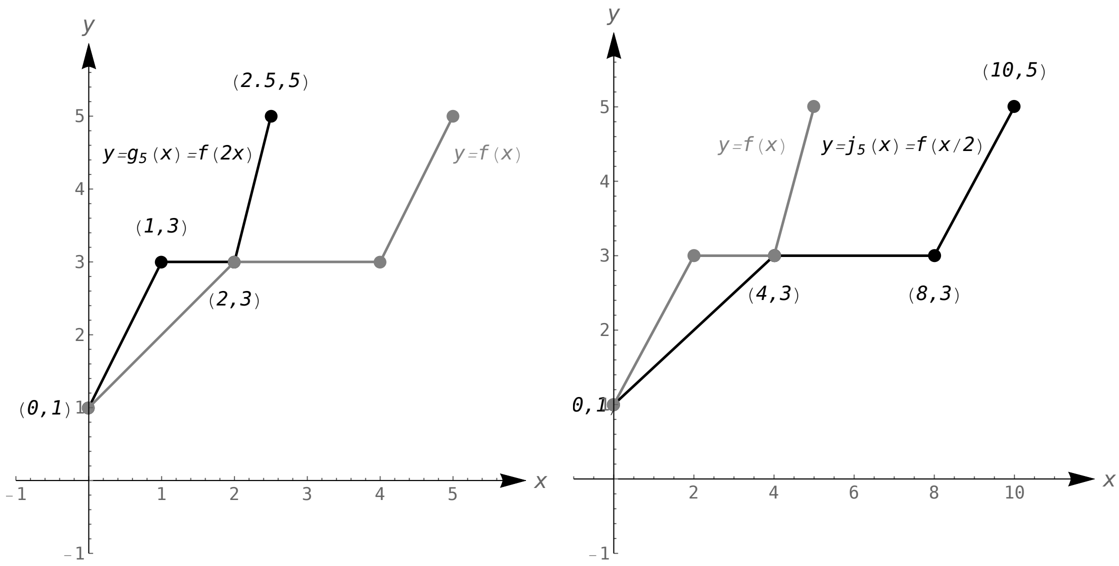
In general, suppose f is a function and $b > 0$, then to obtain the graph of $y = f(bx)$, we have to horizontally scale the graph of f by a factor of $\frac{1}{b}$.



(a) $y = g_4(x) = 2f(x)$ (b) $y = j_4(x) = \frac{1}{2}f(x)$

Figure 3.16: Vertical scalings of the graph of $y = f(x)$: stretching by a factor 2 (a); shrinking by a factor 1/2 (b).

- If $0 < b < 1$, we say the graph of f has undergone a **horizontal stretching (expansion, dilation)** by a factor of $\frac{1}{b}$.
- If $b > 1$, we say the graph of f has undergone a **horizontal shrinking (compression, contraction)** by a factor of b.



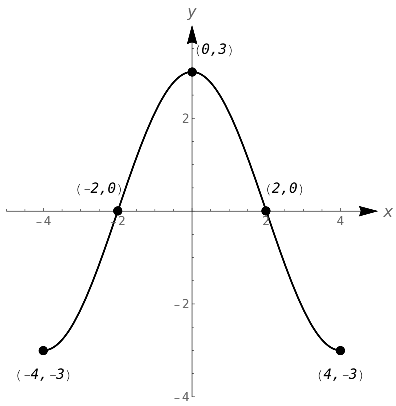
(a) $y = g_5(x) = f(2x)$ (b) $y = j_5(x) = f(\frac{x}{2})$

Figure 3.17: Horizontal scalings of the graph of $y = f(x)$: shrinking by a factor 2 (a); streching by a factor 1/2 (b).

Example 3.10

Below is the complete graph of $y = f(x)$. Use it to graph

$$g(x) = \frac{4-3f(1-2x)}{2}.$$



Solution

We track the five key points $(-4,-3)$, $(-2,0)$, $(0,3)$, $(2,0)$ and $(4,-3)$ indicated on the graph of f to their new locations. We first rewrite $g(x)$ as

$$g(x) = -\frac{3}{2}f(-2x+1)+2\,.$$

Let us first focus on $f(-2x+1)$. To get from $f(x)$ to $f(-2x+1)$, we need a horizontal shift with one unit, i.e. $f(x+1)$, followed by a horizontal shrinking by a factor of 2, i.e. $f(2x+1)$, followed on its turn by a reflection across the y-axis. So, we set $-2x+1$ equal to the x-coordinates of the key points and solve. For example, solving $-2x+1=-4$, we get $x=\frac{5}{2}$. We summarize the results in the table below.

a	-4	-2	0	2	4
$x = \frac{a-1}{2}$	$\frac{5}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{3}{2}$

Next, we take each of the x values and substitute them into $g(x) = -\frac{3}{2}f(-2x+1)+2$ to get the corresponding y-values. Substituting $x=\frac{5}{2}$, and using the fact that $f(-4)=-3$, we get

$$g\left(\frac{5}{2}\right) = -\frac{3}{2}f\left(-2\left(\frac{5}{2}\right)+1\right)+2 = -\frac{3}{2}f(-4)+2 = -\frac{3}{2}(-3)+2 = \frac{9}{2}+2 = \frac{13}{2}.$$

We see that the output from f is first multiplied by $-\frac{3}{2}$. Thinking of this as a two step process, multiplying by $\frac{3}{2}$ and then by -1, we have a vertical stretching by a factor of $\frac{3}{2}$ followed by a reflection across the x-axis. Adding 2 results in a vertical shift up 2 units. Continuing in this manner, we get the following table.

x	$\frac{5}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{3}{2}$
$g(x)$	$\frac{13}{2}$	2	$-\frac{5}{2}$	2	$\frac{13}{2}$

To graph g, we plot each of the points in the table above and connect them in the same order and fashion as the points to which they correspond (Figure 3.18(b)). The reader is strongly encouraged to graph the series of functions which shows the gradual transformation of the graph of f into the graph of g.

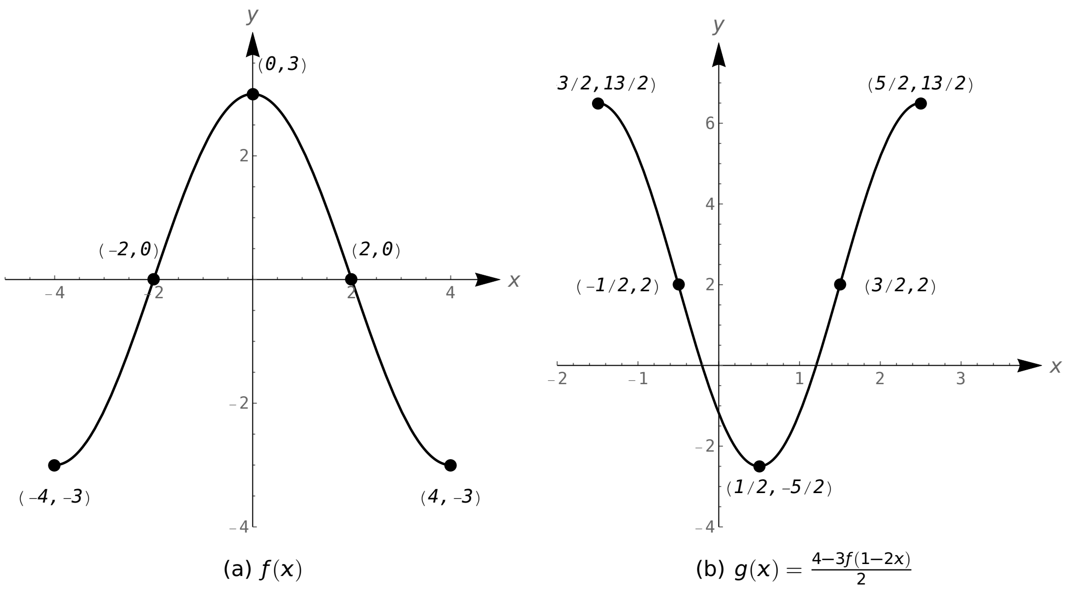


Figure 3.18: The graph of the original function $f(x)$ (a), alongside the transformed function $g(x)$ (b).

3.2.6 Piecewise-defined functions

In many applications, one will encounter functions that are defined on a sequence of intervals. Such functions are referred to as **piecewise-defined functions**, or **piecewise functions** (*stuksgewijze functie*) for short. For instance,

$$f(x) = \begin{cases} (x+1)^2, & \text{if} \quad x < -1, \\ -x, & \text{if} \quad -1 \leq x < 1, \\ \sqrt{x-1}, & \text{if} \quad x \geq 1, \end{cases} \tag{3.1}$$

is a piecewise function. Its graph is given in Figure 3.19

3.2.7 Function families

Throughout the remainder of this course we will focus our attention on the so-called **elementary functions** (*elementaire functie*), which are functions that are compositions of a finite number of arithmetic operations, exponentials, logarithms, constants, and solutions of algebraic equations. Two important

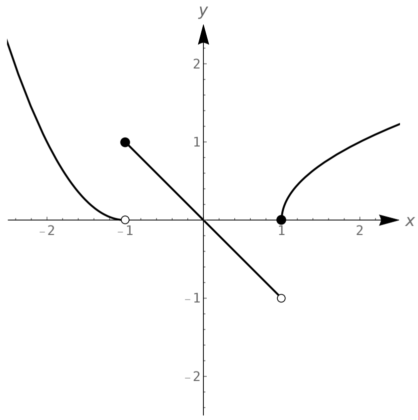


Figure 3.19: The graph of Equation (3.1).

families can be distinguished among the elementary functions, namely the **algebraic** (*algebraïsche*) and **transcendental functions** (*transcendente functie*).

An algebraic function is a function that can be defined as the root of a polynomial equation. Quite often algebraic functions are algebraic expressions using a finite number of terms, involving only the algebraic operations addition, subtraction, multiplication, division, and raising to a fractional power. Examples of such functions are:

- power functions, e.g. $f(x) = 2x^3$,
- polynomial functions, e.g. $f(x) = 1 + x + x^3$,
- rational functions, e.g.
$$f(x) = \frac{1+x}{1+x+x^3},$$
- irrational functions, e.g. $f(x) = \sqrt{1+x+x^3}$,
- and any compositions thereof, e.g.
$$f(x) = \frac{1+x}{\sqrt{1+x+x^3}}.$$

A transcendental function is a function that does not satisfy a polynomial equation, in contrast to an algebraic function. In other words, a transcendental function transcends algebra in that it cannot be expressed in terms of a finite sequence of the algebraic operations of addition, multiplication, and root extraction. Examples of transcendental functions include

- exponential functions, e.g. $f(x) = 2^x$,
- logarithmic functions, e.g. $f(x) = \ln(x)$,
- trigonometric functions, e.g. $f(x) = \sin(x)$,
- hyperbolic functions, e.g. $f(x) = \sinh(x)$,
- and most compositions thereof, e.g.
$$f(x) = \frac{2^x}{\sin(x)}.$$

Algebraic and transcendental functions are studied in detail in Chapter 4 and 5, respectively.

3.3 Absolute value functions

3.3.1 Definition and properties

Throughout this course we adopt the following definition of the absolute value.

Definitie 3.11 (Absolute value)
The **absolute value** (*absolute waarde*) of a real number x, denoted $|x|$, is given by

$$|x| = \begin{cases} -x, & \text{if} \quad x < 0, \\ x, & \text{if} \quad x \geq 0. \end{cases}$$

In this definition, we define $|x|$ using a piecewise-defined function. Other ways to define the absolute value are that $|x|$ is the distance from the real number x to 0 on the number line, or by the equation $|x| = \sqrt{x^2}$. We first remind ourselves of the properties of the absolute value.

Let a, b and x be real numbers and let n be an integer. Then the following arithmetic properties hold.

- **Product rule:**
$$|ab| = |a||b|,$$
- **Power rule:**
$$\left|a^n\right| = |a|^n,$$
  whenever a^n is defined,
- **Quotient rule:**
$$\left|\frac{a}{b}\right| = \frac{|a|}{|b|},$$
  provided $b \neq 0$.

Besides, we have that $|x| = 0$ if and only if $x = 0$, and

- for $c > 0$, $|x| = c$ if and only if $x = c$ or $-x = c$,
- for $c < 0$, $|x| = c$ has no solution.

Example 3.11

Solve each of the following equations.

1. $|3x-1| = 6$
2. $|x| = x^2 - 6$
3. $|x-2| + 1 = x$

Solution

1. The equation $|3x-1| = 6$ is of the form $|x| = c$ for $c > 0$, so $|3x-1| = 6$ is equivalent to $3x-1 = 6$ or $3x-1 = -6$. Solving the former, we arrive at $x = \frac{7}{3}$, and solving the latter, we get $x = -\frac{5}{3}$.
2. For $x < 0$, we have that $|x| = -x$, so for $x < 0$, the equation $|x| = x^2 - 6$ is equivalent to $-x = x^2 - 6$. Rearranging this gives us $x^2 + x - 6 = 0$, or $(x+3)(x-2) = 0$. We get $x = -3$ or $x = 2$. Since only $x = -3$ satisfies $x < 0$, this is the answer we keep.

For $x \geq 0$, we have that $|x| = x$, so the equation $|x| = x^2 - 6$ becomes $x = x^2 - 6$. From this, we get $x^2 - x - 6 = 0$ or $(x-3)(x+2) = 0$. Our solutions are $x = 3$ or $x = -2$, and since only $x = 3$ satisfies $x \geq 0$, this is the one we keep. All together, our two solutions to $|x| = x^2 - 6$ are $x = -3$ and $x = 3$.

3. To solve $|x-2| + 1 = x$, we first isolate the absolute value and get

$$|x-2| = x - 1\,. \tag{3.2}$$

Since we see x both inside and outside of the absolute value, we break the equation into cases:

$$|x-2| = \begin{cases} -(x-2), & \text{if} \quad (x-2) < 0, \\ (x-2), & \text{if} \quad (x-2) \geq 0. \end{cases}$$

Simplifying yields

$$|x-2| = \begin{cases} -x+2, & \text{if} \quad x < 2, \\ x-2, & \text{if} \quad x \geq 2. \end{cases}$$

So, for $x < 2$, we have that $|x-2| = -x+2$ and Equation (3.2) becomes $-x+2 = x-1$, which gives $x = \frac{3}{2}$. Since this solution satisfies $x < 2$, we keep it. Next, for $x \geq 2$, it holds that $|x-2| = x-2$, so Equation (3.2) becomes $x-2 = x-1$. Here, the equation reduces to $-2 = -1$, which signifies we have no solutions here. Hence, our only solution is $x = \frac{3}{2}$.

3.3.2 Absolute value functions

Next, we turn our attention to graphing **absolute value functions** (*absolute waarde functie*). Our strategy in the next example is to make liberal use of Definition 3.11.

Example 3.12

Graph each of the following functions.

1. $f(x) = |x|$
2. $h(x) = \dfrac{|x|}{x}$

Find the zeros of each function and the x- and y-intercepts of each graph, if any exist. From the graph, determine the domain and range of each function, list the intervals on which the function is increasing, decreasing or constant, and find the relative and absolute extrema, if they exist.

Solution

1. To find the zeros of f, we set $f(x) = 0$. We get $|x| = 0$, which gives us $x = 0$. So we get $(0,0)$ as our x-intercept. To find the y-intercept, we set $x = 0$, and find $y = f(0) = 0$, so that $(0,0)$ is our y-intercept as well. Using Definition 3.11, we get

$$f(x) = |x| = \begin{cases} -x, & \text{if} \quad x < 0, \\ x, & \text{if} \quad x \geq 0. \end{cases}$$

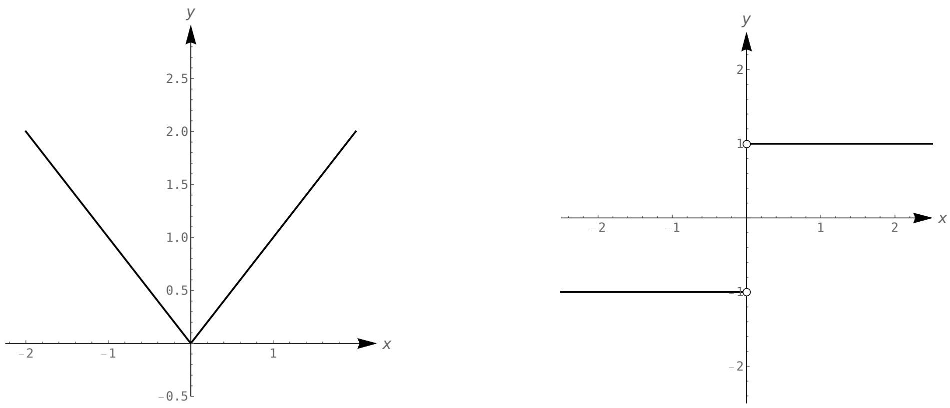
Hence, for $x < 0$, we are graphing the line $y = -x$; for $x \geq 0$, we have the line $y = x$. In this way, we get the graph shown in Figure 3.20(a).

By projecting the graph to the x-axis, we see that the domain is $\mathbb{R}$. Projecting to the y-axis gives us the range $[0, +\infty[$. The function is increasing on $[0, +\infty[$ and decreasing on $]-\infty, 0]$. The relative minimum value of f is the same as the absolute minimum, namely 0 which occurs at $(0,0)$. There is no relative maximum value of f. There is also no absolute maximum value of f, since the y-values on the graph extend infinitely upwards.

2. We first note that, due to the fraction in the formula of $h(x)$ it should hold that $x \neq 0$. Thus the domain is $\mathbb{R}_0$. To find the zeros of h, we set $h(x) = \frac{|x|}{x} = 0$. This last equation implies $|x| = 0$, which implies $x = 0$. However, $x = 0$ is not in the domain of h, which means we have, in fact, no x-intercepts. We have no y-intercepts either, since $h(0)$ is undefined. Re-writing the absolute value in the function gives

$$h(x) = \begin{cases} \dfrac{-x}{x}, & \text{if } \quad x < 0, \\ \dfrac{x}{x}, & \text{if } \quad x > 0, \end{cases} = \begin{cases} -1, & \text{if } \quad x < 0, \\ 1, & \text{if } \quad x > 0. \end{cases}$$

To graph this function, we graph two horizontal lines: $y = -1$ for $x < 0$ and $y = 1$ for $x > 0$. We have open circles at $(0, -1)$ and $(0, 1)$ because the domain of h excludes 0. The range consists of just two y-values: $\{-1, 1\}$. The function h is constant on $]-\infty, 0[$ and $]0, +\infty[$. The local minimum value of h is the absolute minimum value of h, namely -1; the local maximum and absolute maximum values for h also coincide: they both are 1. Every point on the graph of h is simultaneously a relative maximum and a relative minimum.



(a) (b)

Figure 3.20: Graph of $f(x) = |x|$ (a) and $h(x) = \frac{|x|}{x}$ (b).

For what concerns inequalities involving the absolute value, we have the following properties, which follow easily from Definition 3.11.

- If $c > 0$, then $|x| < c$ is equivalent to $-c < x < c$.
- If $c > 0$, then $|x| \leq c$ is equivalent to $-c \leq x \leq c$.
- If $c \leq 0$, then $|x| < c$ has no solution, while if $c < 0$, then $|x| \leq c$ has no solution.
- If $c \geq 0$, then $|x| > c$ is equivalent to $x < -c$ or $x > c$.
- If $c \geq 0$, then $|x| \geq c$ is equivalent to $x \leq -c$ or $x \geq c$.

- If $c < 0$, then $|x| > c$ and $|x| \geq c$ are true for all real numbers.

We can understand each of these statements graphically, so do not learn them by heart. For instance, if $c > 0$, the graph of $y = c$ is a horizontal line which lies above the x-axis through $(0, c)$. Essentially, to solve $|x| < c$, we are looking for the x values where the graph of $y = |x|$ is below the graph of $y = c$. Both graphs are shown in Figure 3.21. We know that the graphs intersect when $|x| = c$, which happens when $x = c$ or $x = -c$. We see that the graph of $y = |x|$ is below $y = c$ for x between $-c$ and c, and hence we get $|x| < c$ is equivalent to $-c < x < c$. The other properties can be shown similarly.

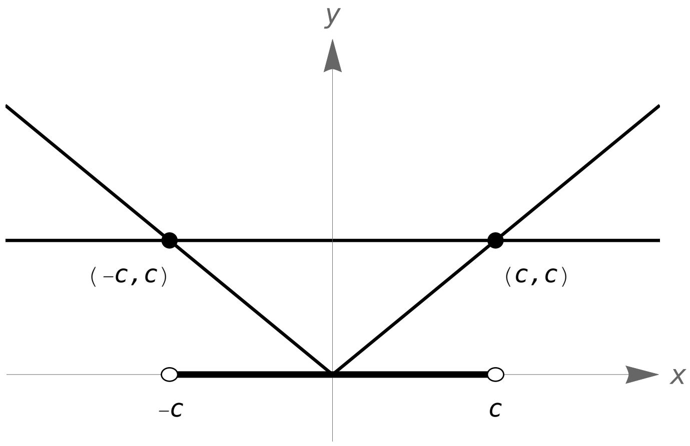


Figure 3.21: The graphs of $y = c$ ($c > 0$) and $y = |x|$.

3.4 Inverse functions

3.4.1 Definition and properties

We can define an **inverse function** (*inverse functie*) as follows.

Definitie 3.12 (Inverse function)
Consider a function $f : X \to Y$, defined by

$$f = \{(x, y) \mid x \in X \wedge y = f(x)\} .$$

Then, the relation

$$f^{-1} = \{(y, x) \mid x \in X \wedge y = f(x)\}$$

is the inverse relation f^{-1} of the function f (Figure 3.22). If and only if this inverse relation is a function on the range Y, this inverse relation f^{-1} is the **inverse of** f (*inverse van* f) and the function f is called **invertible** (*inverteerbaar*).

At this point it is important to recall that a relation constitutes a function if and only if each x-coordinate is matched with at most one y-coordinate (Definition 3.3).

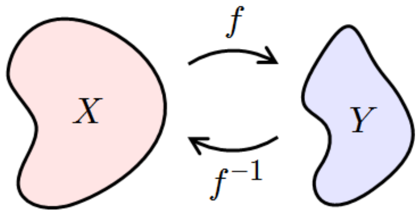


Figure 3.22: If f maps X to Y, then f^{-1} maps Y back to X.

Suppose now f and f^{-1} are inverse functions, then we obviously have the following properties.

- The range of f is the domain of f^{-1} and the domain of f is the range of f^{-1}, i.e.

$$\operatorname{im} f = \operatorname{dom} f^{-1} \quad \text{and} \quad \operatorname{dom} f = \operatorname{im} f^{-1}.$$

- $(f^{-1} \circ f)(x) = x$ for all x in $\operatorname{dom} f$ and $(f \circ f^{-1})(x) = x$ for all x in $\operatorname{dom} f^{-1}$.
- $\left(f^{-1}\right)^{-1}(x) = f(x)$ for all x in $\operatorname{dom} f$.
- $f(a) = b$ if and only if $f^{-1}(b) = a$.
- (a, b) is on the graph of f if and only if (b, a) is on the graph of f^{-1}.

The last property tells us that the graphs of inverse functions are reflections about the line $y = x$. Using the properties of inverse functions, it can be shown that there exists exactly one inverse function f^{-1} for an invertible function f.

Let us now turn our attention to the function $f(x) = x^2$. Is f invertible? A likely candidate for the inverse is the function $g(x) = \sqrt{x}$. Checking the composition yields $(g \circ f)(x) = g(f(x)) = \sqrt{x^2} = |x|$, which is not equal to x for all x in the domain $]-\infty, \infty[$. For example, when $x = -2$, $f(-2) = (-2)^2 = 4$, but $g(4) = \sqrt{4} = 2$, which means g failed to return the input -2 from its output 4. Since a function matches a number with exactly one other number it is impossible to construct a function which takes 4 back to both $x = 2$ and $x = -2$ (Definition 3.3). Still, from a graphical standpoint, we know that if $y = f^{-1}(x)$ exists, its graph can be obtained by reflecting $y = x^2$ about the line $y = x$ (Figure 3.23).

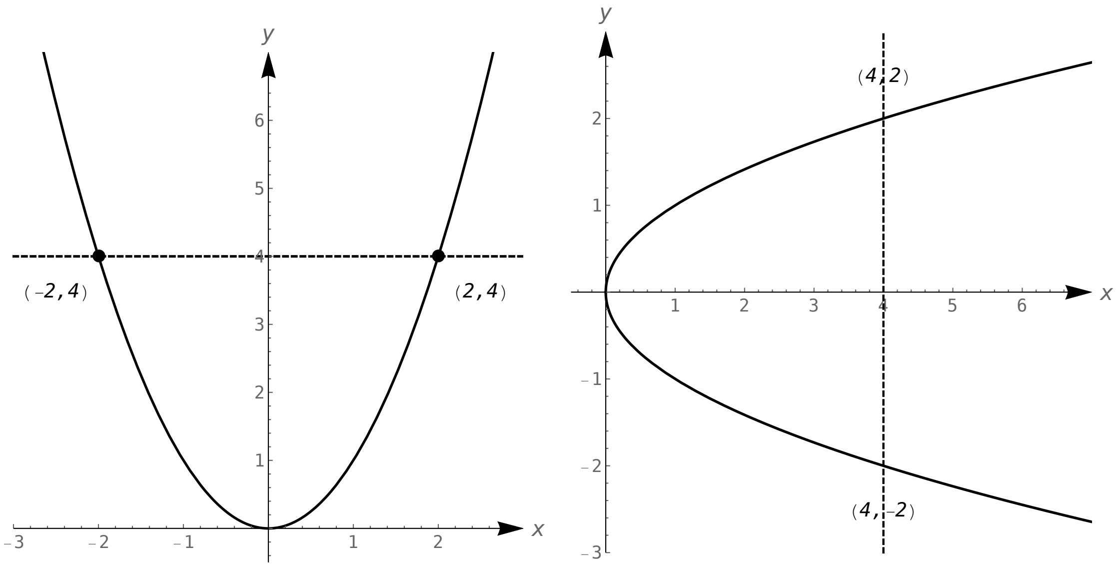
We see that the graph of the supposed inverse function fails the vertical line test (Theorem 3.1), and as such, does not represent y as a function of x. The vertical line $x = 4$ on the graph on the right corresponds to the horizontal line $y = 4$ on the graph of $y = f(x)$. The fact that the horizontal line $y = 4$ intersects the graph of f twice means that two different inputs, namely $x = -2$ and $x = 2$, are matched with the same output, 4, which is the cause of all of the trouble. Consequently, a function f is invertible if and only if f is an injective function (one-to-one), and this makes that the corresponding inverse function f^{-1} is an injection as well.

To find the inverse of an invertible function, we may follow the following steps.

1. Write $y = f(x)$.
2. Interchange x and y.
3. Solve $x = f(y)$ for y to obtain $y = f^{-1}(x)$.

3.4.2 Domain restriction

Let us return to the function $f(x) = x^2$. We know that f is not one-to-one, and thus, is not invertible. However, if we restrict the domain of f, we can produce a new function g which is one-to-one. If we



(a) (b)

Figure 3.23: Graph of $y = f(x) = x^2$ (a) and the reflection of $y = x^2$ about the line $y = x$ (b).

define $g(x) = x^2$ for $x \geq 0$, we can investigate the graph of this function (Figure 3.24). It is clear that g is an injective function, so we can try find its inverse. We proceed as follows

$$\begin{array}{llll} 1. & y & = & g(x) \\ & & = & x^2, \;\; x \geq 0 \\ 2. & x & = & y^2, \;\; y \geq 0 \qquad\qquad \text{(Switch } x \text{ and } y\text{.)} \\ 3. & y & = & \pm\sqrt{x}, \, y \geq 0 \\ \Rightarrow & y & = & \sqrt{x} \end{array}$$

We get $g^{-1}(x) = \sqrt{x}$. At first it looks like we will run into the same trouble as before, but when we check the composition, the domain restriction on g saves the day. We get

$$\left(g^{-1} \circ g\right)(x) = g^{-1}(g(x)) = g^{-1}\left(x^2\right) = \sqrt{x^2} = |x| = x,$$

since $x \geq 0$, and likewise

$$\left(g \circ g^{-1}\right)(x) = g\left(g^{-1}(x)\right) = g\left(\sqrt{x}\right) = \left(\sqrt{x}\right)^2 = x.$$

Graphing g and g^{-1} on the same set of axes shows that they are reflections about the line $y = x$.

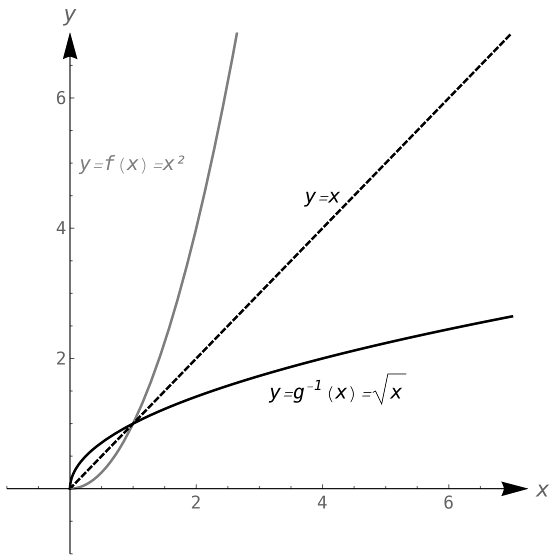


Figure 3.24: The graph of $g(x) = x^2$ for $x \geq 0$ and its inverse $g^{-1}(x) = \sqrt{x}$.

Example 3.13

Determine the inverse of the following function, if it exists, and check your answer graphically:

$$f(x) = x^2 - 2x + 4,$$

where $x \leq 1$.

Solution

We can reformulate the function definition as

$$f(x) = (x-1)^2 + 3\,,$$

from which we infer that it represents the parabola displayed in Figure 3.23, but shifted to the right by one unit and up three units. Moreover, since this function's domain is restricted to $x \leq 1$, we are selecting only the left half of the parabola. Consequently, f is an injective function and thus it is invertible.

We now derive the formula for $f^{-1}(x)$.

$$\begin{array}{lrcll}
1. & y & = & f(x) & \\
 & & = & (x-1)^2+3,\ x \leq 1 & \\
2. & x-3 & = & (y-1)^2,\ y \leq 1 & \text{(Switch } x \text{ and } y.\text{)} \\
\Leftrightarrow & \pm\sqrt{x-3} & = & y-1 & \\
\Rightarrow & y & = & 1-\sqrt{x-3} & \text{(Since } y \leq 1.\text{)}
\end{array}$$

We have $f^{-1}(x) = 1-\sqrt{x-3}$. In order to check our answer graphically, we graph $y = f^{-1}(x)$ and $y = f(x)$ in Figure 3.25. From this figure it is clear that we obtained the correct inverse function since its graph is the reflection of the graph of f about the first bisector.

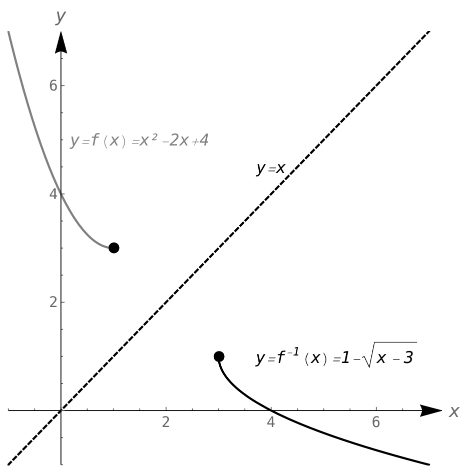
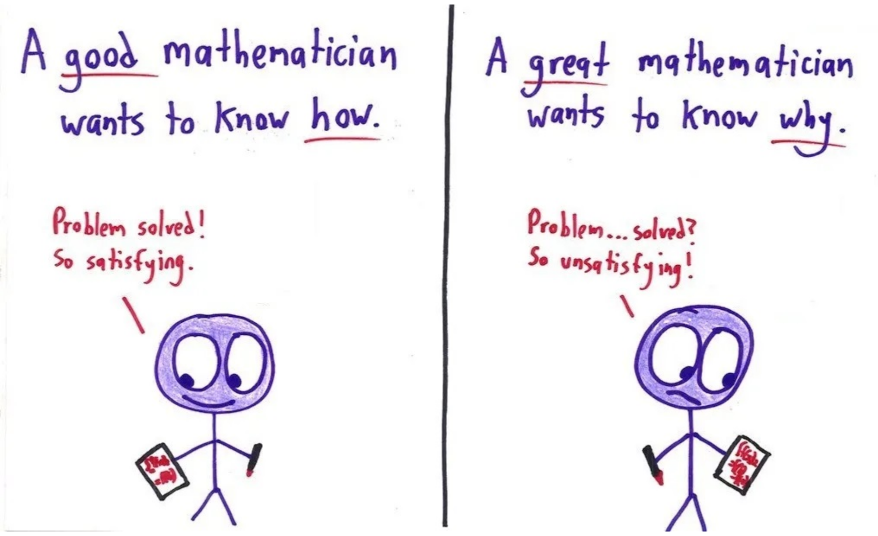


Figure 3.25: The graph of $f(x) = x^2-2x+4$ for $x \leq 1$ and its inverse $f^{-1}(x) = 1-\sqrt{x-3}$.



From *Math with Bad Drawings*, used by permission of Ben Orlin.

3.5 Exercises

Functions in $\mathbb{R}$

Assignment 3.1 — Which of the following graphs corresponds with a function?

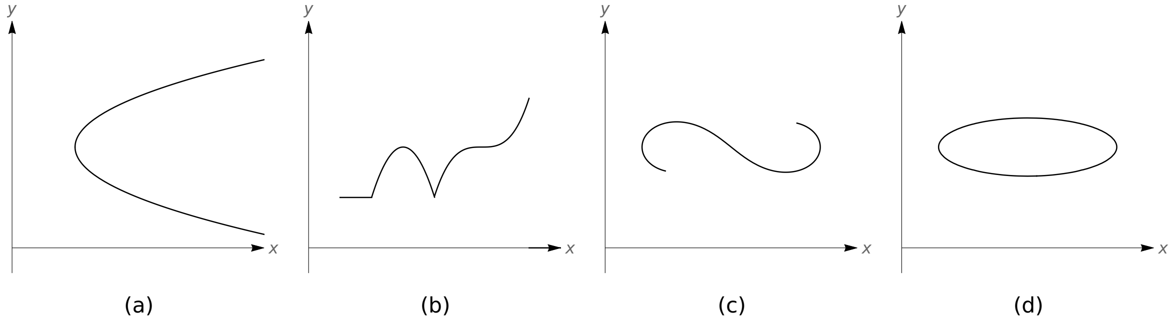


Figure 3.26: The graphs from exercise 3.1.

Function arithmetic

Assignment 3.2 — Consider the functions below and determine an expression for and the domain of a) $(f+g)(x)$, b) $(f-g)(x)$, c) $(fg)(x)$ and d) $\left(\frac{f}{g}\right)(x)$.

(a) $f(x)=x^3-1 \quad$ and $\quad g(x)=\dfrac{x+1}{x-1}$

(b) $f(x)=\dfrac{x}{2} \quad$ en $\quad g(x)=\dfrac{2}{x}$

(c) $f(x)=x \quad$ and $\quad g(x)=\sqrt{x+1}$

Assignment 3.3 — Determine the following for the functions $f(x)=x+5$ and $g(x)=x^2-3$:

(a) $(f\circ g)(0)$

(b) $g(f(0))$

(c) $(f\circ f)(-5)$

(d) $g(g(2))$

Assignment 3.4 — Determine $g\circ f$ and $f\circ g$. Check the domain of each composite function as well.

(a) $f(x)=5x \quad$ and $\quad g(x)=\dfrac{1}{x-2}$

(b) $f(x)=x^2 \quad$ and $\quad g(x)=\sqrt{x}$

(c) $f(x)=x^3 \quad$ and $\quad g(x)=\sqrt[3]{1-x}$

(d) $f(x)=x^2-4 \quad$ and $\quad g(x)=|x-1|$

(e) $f(x)=|x| \quad$ and $\quad g(x)=\sqrt{4-x}$

Assignment 3.5 — Consider the function

$$g(x) = \frac{1-x}{1+x}.$$

Calculate $(g \circ g)(x)$ and determine the domain.

Assignment 3.6 — Consider the function $f(x)$ with domain $[0,2]$ and image $[0,1]$ (Figure 3.27).

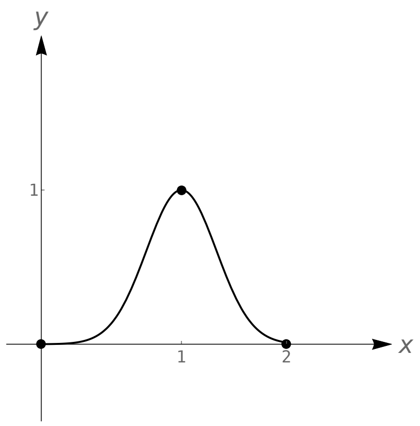


Figure 3.27: The graph from the function $y = f(x)$ from Exercise 3.6.

Sketch the graph of the transformations of $f(x)$ below and determine their domain and image.

(a) $y = f(x) + 2$

(b) $y = f(x) - 1$

(c) $y = f(x+2)$

(d) $y = f(x-1)$

(e) $y = -f(x)$

(f) $y = f(-x)$

(g) $y = f(4-x)$

(h) $y = 2f(x)$

(i) $y = -\dfrac{f(x)}{2}$

(j) $y = f(2x)$

(k) $y = f\left(\dfrac{x}{3}\right)$

(l) $y = 1 - f(1-x)$

(m) $y = 1 + f\left(-\dfrac{x}{2}\right)$

Piecewise-defined functions

Assignment 3.7 — Sketch the graphs of the piecewise functions below.

(a) $f(x) = \begin{cases} x^2, & \text{als} \quad x \leq -2, \\ 3-x, & \text{als} \quad -2 < x < 2, \\ 4, & \text{als} \quad x \geq 2. \end{cases}$

(b) $f(x) = \begin{cases} x+5, & \text{als} \quad x \leq -3, \\ \sqrt{9-x^2}, & \text{als} \quad -3 < x \leq 3, \\ -x+5, & \text{als} \quad x > 3. \end{cases}$

(c) $g(x) = \begin{cases} x^2, & \text{als} \quad x \leq -1, \\ \sqrt{1-x^2}, & \text{als} \quad -1 < x \leq 1, \\ x, & \text{als} \quad x > 1. \end{cases}$

(d) $f(x) = \begin{cases} \dfrac{1}{x}, & \text{als} \quad -6 < x < -1, \\ x, & \text{als} \quad -1 < x < 1, \\ \sqrt{x}, & \text{als} \quad 1 < x < 9. \end{cases}$

Assignment 3.8 — Determine the function $f(x)$ depicted in Figure 3.31.

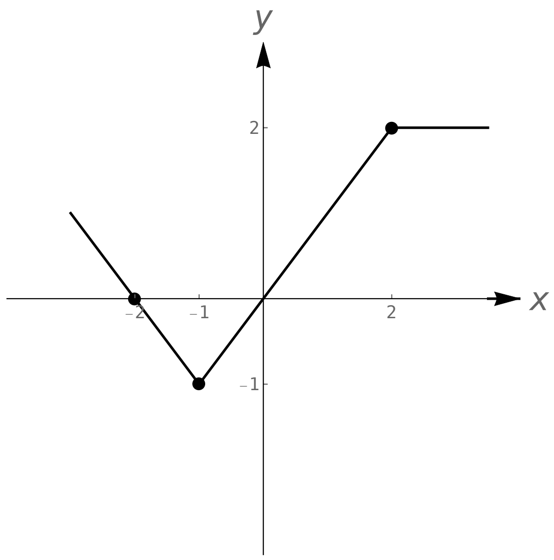


Figure 3.31: The graph of the piecewise function $f(x)$ from Exercise 3.8

Absolute value functions

Assignment 3.9 — If $|x-4| < 2$, check which of the statements below are true or false. Demonstrate the correct statements and provide a counterexample for the incorrect statements.

(a) $2 < x < 6$

(b) $0 < x < 4$

(c) $-1 < -\dfrac{x}{2} < 0$

(d) $1 < \dfrac{6}{x} < 3$

(e) $0 < x-2 < 4$

(f) $0 < \dfrac{1}{x} < \dfrac{1}{6}$

(g) $-6 < -x < 2$

(h) $-6 < -x < -2$

Assignment 3.10 — Solve the equations below.

(a) $\frac{2}{3}|5-2x|-\frac{1}{2}=5$

(b) $|2-5x|=5|x+1|$

(c) $|x+3|=|2x+1|$

(d) $4|2x-1|-2=10$

(e) $7|x-2|+7=-2|x-2|+2$

(f) $|2x+3|+9\leq 7$

(g) $|3x-7|<2$

(h) $\left|\frac{x}{2}-1\right|\leq 1$

(i) $|x-|x||=10$

(j) $-2\left(|-9x|+|1-9x|\right)=-100$

(k) $|x|-|x-2|=2$

(l) $|x+1|>|x-3|$

(m) $|x-3|<2|x|$

(n) $|x|-|2-x|<2$

(o) $\frac{|3-5x|}{x-2}>6$

(p) $||x+2|-5|>3$

(q) $\left|\frac{1-2x}{1-x}\right|>2$

(r) $\left|\frac{x}{2+x}\right|<1$

(s) $||x|-|7-x||=21$

(t) $|1-x|+|2x-1|-|x+1||x-1|=x$

(u) $-3|x-1|+|3x-1|\leq x-1$

(v) $|5-x|+|3x-1|\geq x+2$

(w) $|3-|2-x||\leq 2x$

Assignment 3.11 — Sketch the graphs of the functions below.

(a) $f(x)=|x-2|$

(b) $f(x)=1+|x-2|$

(c) $f(x)=\frac{x+|x|}{2}$

(d) $f(x)=\frac{x-|x|}{2}$

(e) $f(x)=|x+2|-|x-3|+1$

(f) $f(x)=3|x+4|-4$

(g) $f(x)=||x|+3|$

(h) $f(x)=||x|-3|$

(i) $|x|+|y|=1$

Inverse functions

Assignment 3.12 — Determine the inverse f^{-1} of f if f is defined as below. Is f^{-1} a function? Verify this both algebraically and graphically.

(a) $f(x)=2x+8$

(b) $f(x)=1-\frac{4+3x}{5}$

(c) $f(x)=\frac{x+6}{x+5}$

(d) $f(x)=x^3+1$

(e) $f(x)=x^2+x$

(f) $f(x)=\sqrt{1-x^2}$

(g) $f(x)=3\sqrt{x-1}-4$

(h) $f(x)=x^2-6x+5$

Review exercises

Assignment 3.13 — For the following functions, determine the domain, codomain and image in each case. Then examine whether they are injective, surjective, and/or bijective. Also check whether the functions are periodic, even/odd and increasing/decreasing. If possible, determine any maxima and/or minima.

(a) $f(x) = \sqrt{4 - x^2}$

(b) $f(x) = \sqrt{4 + x^2}$

(c) $f(x) = \dfrac{1}{1 - x^2}$

(d) $f(x) = \dfrac{1}{1 + x^2}$

(e) $f(x) = x^5 + 1$

(f) $f(x) = x^4 + 1$

(g) $f(x) = \sqrt{1 - x}$

(h) $f(x) = |x|$

(i) $f(x) = \dfrac{x}{x + 1}$

(j) $f(x) = -\dfrac{4}{x^3}$

(k) $f(x) = 3 - 2\sqrt{x + 2}$

(l) $f(x) = \sqrt{x^2 + 4x + 4}$

Obvious is the most dangerous word in mathematics

— Eric Temple Bell —

4

Algebraic functions

4.1 Polynomial functions

4.1.1 Constant and linear functions

Many of the functions we already encountered in the preceding chapter were either constant or linear. Here, we first of all give a more formal definition of a linear function.

Definitie 4.1 (Linear function)
A **linear function** (*lineaire functie*) is a function of the form

$$f(x) = a\,x + b,$$

where a and b are real numbers with $a \neq 0$. The domain of a linear function is $\mathbb{R}$.

For the case $a = 0$, we get $f(x) = b$, which is referred to as a **constant function** (*constante functie*).

Recall that to graph a function f, we graph the equation $y = f(x)$. Hence, the graph of a linear function is a line with slope a and y-intercept $(0, b)$; the graph of a constant function is a horizontal line (a line with slope $a = 0$) and an y-intercept of $(0, b)$. For that reason, Definition 4.1 is therefore often referred to as the slope-intercept definition of a line in the plane. In general, given two points in the plane (x_1, y_1) and (x_2, y_2) the equation of straight line is given by

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}\,(x - x_1)\ , \tag{4.1}$$

where

$$\frac{y_2 - y_1}{x_2 - x_1}$$

is the slope of the line.

4.1.2 Quadratic functions

Definitie 4.2 (Quadratic function)
A **quadratic function** (*kwadratische functie*) is a function of the form

$$f(x) = ax^2 + bx + c,$$

where a, b and c are real numbers with $a \neq 0$. The domain of a quadratic function is $\mathbb{R}$.

The most basic quadratic function is $f(x) = x^2$, which is shown in Figure 4.1. Its shape is called a **parabola** (*parabool*). The point $(0,0)$ is called the **vertex** (*top*) of the parabola.

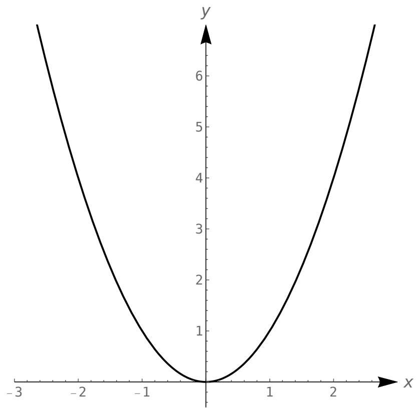


Figure 4.1: The graph of $y = x^2$.

Definition 4.2 uses the general form of a quadratic function, though a quadratic function f may as well be defined using its so-called standard form, being

$$f(x) = a(x-h)^2 + k\,,$$

where a, h and k are real numbers with $a \neq 0$. The vertex of the graph of $y = f(x)$ is in this notation given by (h,k). Any quadratic function can be rewritten in standard form by completing the square.

The graph of $y = a(x-h)^2 + k$ is a parabola opening upwards if $a > 0$, and opening downwards if $a < 0$. Moreover, the symmetry enjoyed by the graph of $y = x^2$ about the y-axis is translated to a symmetry about the vertical line $x = h = -\frac{b}{2a}$ which is the vertical line through the vertex. This line is called the **axis of symmetry** (*symmetrieas*) of the parabola.

Our next example is a classic application of quadratic functions.

Example 4.1

Donnie inherits a large parcel of land near Tielt from one of his relatives. The time is finally right for him to pursue his dream of farming some cannabis (*Cannabis sativa*). He wishes to build a rectangular pasture, and estimates that he has enough money for 200 linear metres of fencing material. If he makes the pasture adjacent to a stream (so no fencing is required on that side), what are the dimensions of the pasture which maximize the area? What is the maximum area? If an average cannabis plant needs 2.5 square metres of grazing area, how many cannabis can Donnie keep in his pasture? It is always helpful to sketch the problem situation (Figure 4.2).

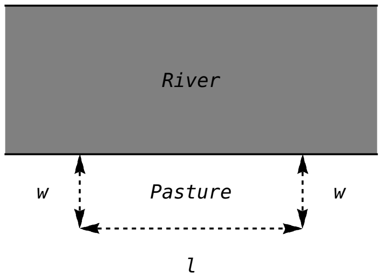


Figure 4.2: Donnie's pasture.

Solution

We are tasked to find the dimensions of the pasture which would give a maximum area. We let w [L] denote the width of the pasture and we let l [L] denote the length of the pasture. Since the units given to us in the statement of the problem are metres, we assume w and l are measured in metre. The area of the pasture, which we will call A, is related to w and l by the equation $A = wl$. Since w and l are both measured in metre, A [L^2] has units of square metre. We are given the total amount of fencing available is 200 metres, which means $w + l + w = 200$, or, $l + 2w = 200$. We now have two equations,

$$\begin{cases} A &= wl\,, \\ l+2w &= 200\,. \end{cases}$$

In order to maximize A, we need to use the information given to write A as a function of just one variable, either w or l:

$$A = wl = w(200-2w) = 200w - 2w^2.$$

We now have A as a function of w:

$$A(w) = -2w^2 + 200w.$$

Before we go any further, we need to find the applied domain of A so that we know what values of w make sense in this problem situation. Since w represents the width of the pasture, $w > 0$. Likewise, l represents the length of the pasture, so $l = 200 - 2w > 0$. Solving this latter inequality, we find $w < 100$. Hence, the function we wish to maximize is $A(w) = -2w^2 + 200w$ for $0 < w < 100$. Since A is a quadratic function of w, we know that the graph of $y = A(w)$ is a parabola. Since the coefficient of w^2 is -2, we know that this parabola opens downwards. This means that there is a maximum value to be found, and we know it occurs at the vertex. Using the vertex formula, we find

$$\begin{cases} w &= -\dfrac{200}{2(-2)} = 50\,, \\ A(50) &= -2(50)^2 + 200(50) = 5000\,. \end{cases}$$

Since $w = 50$ lies in the applied domain, $0 < w < 100$, we have that the area of the pasture is maximized when the width is 50 metres. To find the length, we use $l = 200 - 2w$ and find $l = 200 - 2(50) = 100$, so the length of the pasture is 100 metres. The maximum area is $A(50) = 5000$, or 5000 m^2. If an average cannabis plant requires 2.5 square metres of pasture, Donnie can raise $\frac{5000}{2.5} = 2000$ such plants.

In practice, quadratic functions often pop up in inequalities, which we can solve graphically.

One of the classic applications of inequalities is the notion of **tolerances**. Recall that for real numbers

x and c, the quantity $|x-c|$ may be interpreted as the distance from x to c. Solving inequalities of the form $|x-c| \leq d$ for $d \geq 0$ can then be interpreted as finding all numbers x which lie within d units of c. We can think of the number d as a tolerance and our solutions x as being within an accepted tolerance of c. We use this principle in the next example.

Example 4.2

The area A ([L^2]) of a square piece of particle board which measures x [L] centimetres on each side is $A(x) = x^2$. Suppose a manufacturer needs to produce a 24 centimetre by 24 centimetre square piece of particle board as part of a home office desk kit. How close does the side of the piece of particle board need to be cut to 24 centimetres to guarantee that the area of the piece is within a tolerance of 0.25 square centimetres of the target area of 576 square centimetres?

Solution

Mathematically, we express the desire for the area $A(x)$ to be within 0.25 square centimetres of 576 as $|A-576| \leq 0.25$. Since $A(x) = x^2$, we get $|x^2-576| \leq 0.25$, which is equivalent to $-0.25 \leq x^2-576 \leq 0.25$. Recalling the increasing property of the square root; that is if $0 \leq a \leq b$, then $\sqrt{a} \leq \sqrt{b}$, we proceed

$$\begin{array}{rrcll}
 & -0.25 \leq & x^2-576 & \leq 0.25 & \\
\Leftrightarrow & 575.75 \leq & x^2 & \leq 576.25 & \text{(Add 576 across the inequalities.)} \\
\Leftrightarrow & \sqrt{575.75} \leq & \sqrt{x^2} & \leq \sqrt{576.25} & \text{(Take square roots.)} \\
\Leftrightarrow & \sqrt{575.75} \leq & |x| & \leq \sqrt{576.25} & (\sqrt{x^2} = |x|)
\end{array}$$

Consequently, we find the solution to $\sqrt{575.75} \leq |x|$ to be $]-\infty, -\sqrt{575.75}\,] \cup [\,\sqrt{575.75}, +\infty[$ and the solution to $|x| \leq \sqrt{576.25}$ to be $[-\sqrt{576.25}, \sqrt{576.25}\,]$. To solve $\sqrt{575.75} \leq |x| \leq \sqrt{576.25}$, we intersect these two sets to get $[-\sqrt{576.25}, -\sqrt{575.75}\,] \cup [\,\sqrt{575.75}, \sqrt{576.25}\,]$. Since x represents a length, we discard the negative answers and get $[\sqrt{575.75}, \sqrt{576.25}\,]$. This means that the side of the piece of particle board must be cut between $\sqrt{575.75} \approx 23.995$ and $\sqrt{576.25} \approx 24.005$ centimetres, a tolerance of (approximately) 0.005 centimetres of the target length of 24 centimetres.

Our last example in the section demonstrates how inequalities can be used to describe regions in the plane.

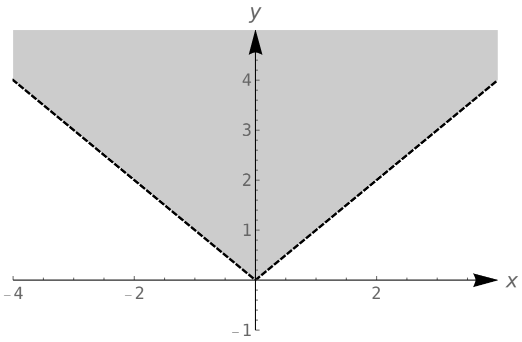
Example 4.3

Sketch the following relations.

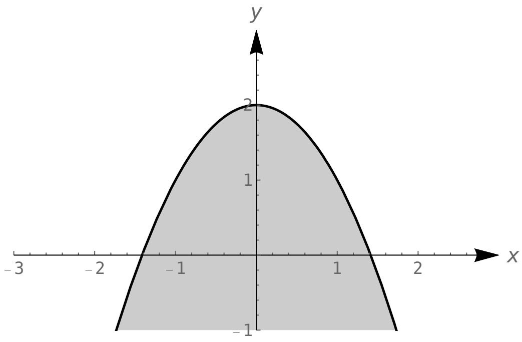
1. $R = \{(x,y) : y > |x|\}$
2. $S = \{(x,y) : y \leq 2-x^2\}$

Solution

1. The relation R consists of all points (x,y) whose y-coordinate is greater than $|x|$. If we graph $y = |x|$, then we want all of the points in the plane above the points on the graph. Dotting the graph of $y = |x|$ to indicate that the points on the graph itself are not in the relation, we get the shaded region in Figure 4.3(a).
2. For a point to be in S, its y-coordinate must be less than or equal to the y-coordinate on the parabola $y = 2-x^2$. This is the set of all points below or on the parabola $y = 2-x^2$ (Figure 4.3(b)).



(a)



(b)

Figure 4.3: Graph of the relation R (a) and S (b).

Many quadratic equations $ax^2+bx+c=0$ cannot be solved by factoring them. This is generally true when the roots are not rational numbers. An other method of solving quadratic equations involves the use of the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}, \tag{4.2}$$

where $a \neq 0$. In relation to quadratic equations, complex numbers come in when the value under the radical portion of the quadratic formula is negative. When this occurs, the equation has no roots in $\mathbb{R}$. The roots belong to $\mathbb{C}$, will be called **complex roots** (*complexe wortels*) (or imaginary roots), and are expressible as $a \pm bi$. When using the quadratic formula to solve a quadratic equation with real coefficients, there are three possibilities depending on the discriminant $D = b^2-4ac$:

1. Two different real roots if $D > 0$.
2. One real root if $D = 0$.
3. Two complex roots, complex conjugates, if $D < 0$.

Example 4.4

Solve the following quadratic equations.

1. $x^2+2x+2=0$
2. $x^2-4x+13=0$

Solution

1. Substituting in Equation (4.2), we get

$$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a} = \frac{-2 \pm \sqrt{-4}}{2}.$$

Since the discriminant $D = b^2-4ac$ is negative, this equation has no solution in $\mathbb{R}$. But if you were to express the solution using imaginary numbers, the solutions would be

$$x = \frac{-2 \pm 2i}{2} = -1 \pm i.$$

2. Notice that after substitution in Equation (4.2), we are left with a negative value under the

square root radical.

$$x = \frac{4 \pm \sqrt{-36}}{2}$$

There are two complex roots

$$x = \frac{4 \pm 6i}{2} = 2 \pm 3i.$$

4.1.3 General polynomial functions

4.1.3.1 Definition

Three of the families of functions studied thus far – constant, linear and quadratic – belong to a much larger group of functions called polynomials. We begin our formal study of general polynomials with a definition and some examples.

Definitie 4.3 (Polynomial function)
A **polynomial function** (*veeltermfunctie*) is a function of the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0,$$

where a_0, a_1, ..., a_n are real numbers, $a_n \neq 0$ and $n \in \mathbb{N}$. The domain of a polynomial function is $\mathbb{R}$.

Given $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$ with $a_n \neq 0$, we say

- The natural number n is called the **degree** (*graad*) of the polynomial f.
- $a_n x^n$ is called the **leading term** (*hoogstegraadsterm*) of the polynomial f.
- The real number a_n is called the **leading coefficient** of the polynomial f.
- The real number a_0 is called the **constant term** (*constante term*) of the polynomial f.

Moreover, if $f(x) = a_0$, and $a_0 \neq 0$, we say f has degree 0, while if $f(x) = 0$, we say f has no degree.

In Part II, we will introduce the tools that are needed to graph polynomial functions and understand their behaviour. Anyhow, we will often have to determine the zeros of a polynomial equation. For that reason, we recall the most important facts about the factorization of polynomials.

4.1.3.2 Factorization of polynomials

Suppose we wish to find the zeros of $f(x) = x^3 + 4x^2 - 5x - 14$. Setting $f(x) = 0$ results in the polynomial equation $x^3 + 4x^2 - 5x - 14 = 0$. It is easy to see that $f(2) = 0$, but possible other zeros seem less obvious. Now, if $x = 2$ is a zero, there should be a factor of $(x-2)$ lurking around in the factorization of $f(x)$. In other words, we should expect that $x^3 + 4x^2 - 5x - 14 = (x-2)\,q(x)$, where $q(x)$ is some second degree polynomial. We can find this polynomial through **polynomial division** (*Euclidische deling*). Dividing $x^3 + 4x^2 - 5x - 14$ by $x - 2$ gives

$$
\begin{array}{r|l}
x^3 + 4x^2 - 5x - 14 & x-2 \\ \cline{2-2}
-\left(x^3 - 2x^2\right) & x^2+6x+7 \\ \cline{1-1}
6x^2 - 5x & \\
-\left(6x^2 - 12x\right) & \\ \cline{1-1}
7x - 14 & \\
-\left(7x - 14\right) & \\ \cline{1-1}
0 &
\end{array}
$$

This means $x^3+4x^2-5x-14 = (x-2)\left(x^2+6x+7\right)$, so to find the zeros of f, we now solve $(x-2)\left(x^2+6x+7\right)=0$. We get $x-2=0$ (which gives us our known zero, $x=2$) as well as $x^2+6x+7=0$. The latter leads to additional zeros, namely $x=-3\pm\sqrt{2}$.

First of all, we should remember what we may expect when dividing polynomials in general.

Definitie 4.4 (Polynomial division)
Suppose $d(x)$ and $p(x)$ are nonzero polynomials where the degree of p is greater than or equal to the degree of d. There exist two unique polynomials, $q(x)$ and $r(x)$, such that

$$p(x) = d(x)\,q(x) + r(x),$$

where either $r(x)=0$ or the degree of r is strictly less than the degree of d.

As you may recall, all of the polynomials in Definition 4.4 have special names. The polynomial p is called the **dividend** (*deeltal*); d is the **divisor** (*deler*); q is the **quotient** (*quotiënt*); r is the **remainder** (*rest*). If $r(x)=0$ then d is called a **factor** of p.

The additional finding that $x-2$ is a factor of $x^3+4x^2-5x-14$ as $x=2$ is a zero of $x^3+4x^2-5x-14=0$ can be generalized in the following theorem.

Theorem 4.1 (The factor theorem)
Suppose $p(x)$ is a nonzero polynomial. The real number c is a zero of $p(x)$ if and only if $x-c$ is a factor of $p(x)$.

For completeness we also mention the related remainder theorem.

Theorem 4.2 (The remainder theorem)
Suppose p is a polynomial of degree at least 1 and c is a real number. When $p(x)$ is divided by $x-c$ the remainder is $p(c)$.

Sometimes $x-c$ is not only a factor of a given polynomial $p(x)$, but as well of the quotient resulting from dividing $p(x)$ by $x-c$. In that case, we say that the **multiplicity** (*multipliciteit*) of the factor $x-c$ is 2, and hence the multiplicity of $x=c$ as a zero of $p(x)$ is 2 as well. Clearly, the multiplicity of a factor $x-c$ can be at most equal to the degree of the polynomial $p(x)$.

Clearly, a polynomial division can become quite tedious, but when dividing polynomials by quantities of the form $x-c$, we can rely on **Horner's approach** (*methode van Horner*). Essentially, this boils down to constructing a synthetic division tableau for the polynomial division problem. Let us rework our division problem using this tableau to see how it greatly streamlines the division process. To divide

$x^3+4x^2-5x-14$ by $x-2$, we write 2 in the place of the divisor and the coefficients of $x^3+4x^2-5x-14$ in for the dividend. Then bring down the first coefficient of the dividend.

$$\begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & & & & \\ \hline & & & & \end{array} \qquad \begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & & & \\ \hline & 1 & & & \end{array}$$

Next, take the 2 from the divisor and multiply by the 1 that was 'brought down' to get 2. Write this underneath the 4, then add to get 6.

$$\begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & & \\ \hline & 1 & & & \end{array} \qquad \begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & & \\ \hline & 1 & 6 & & \end{array}$$

Now take the 2 from the divisor times the 6 to get 12, and add it to the -5 to get 7.

$$\begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & 12 & \\ \hline & 1 & 6 & & \end{array} \qquad \begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & 12 & \\ \hline & 1 & 6 & 7 & \end{array}$$

Finally, take the 2 in the divisor times the 7 to get 14, and add it to the -14 to get 0.

$$\begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & 12 & 14 \\ \hline & 1 & 6 & 7 & \end{array} \qquad \begin{array}{r|rrrr} & 1 & 4 & -5 & -14 \\ 2 & \downarrow & 2 & 12 & 14 \\ \hline & 1 & 6 & 7 & \boxed{0} \end{array}$$

The first three numbers in the last row of our tableau are the coefficients of the quotient polynomial. Remember, we started with a third degree polynomial and divided by a first degree polynomial, so the quotient is a second degree polynomial. Hence the quotient is x^2+6x+7. The number in the box is the remainder. Synthetic division is our tool of choice for dividing polynomials by divisors of the form $x-c$. It is important to note that it works only for these kinds of divisors. Besides, when doing a synthetic division, do not forget to insert 0's in the division tableau to account for any missing powers of x.

In Mathematica, a polynomial can be factored over integers using the built-in function **Factor** as follows.

```
In[4]:= Factor[x^3 +4*x^2 -5*x -14]

Out[4]= (-2 +x) (7 +6x +x^2)
```

Of course, if we allow for irrational numbers in our factorisation, the last factor in $(-2+x)(7+6x+x^2)$ can be factored as $(x+3-\sqrt{2})(x+3+\sqrt{2})$. For that purpose, one can look for the roots of $7+6x+x^2=0$ in Mathematica using **Solve**.

The following theorem gives us an upper bound on the number of real zeros.

Theorem 4.3 (Zeros and multiplicity)
Suppose f is a polynomial of degree $n \geq 1$. Then f has at most n real zeros, counting multiplicities.

In many cases, however, the factorization of a polynomial f will lead to quadratic terms with complex zeros. In that case it is important to note that complex roots of a polynomial f occur as complex conjugate pairs, as indicated by the following theorem.

Theorem 4.4 (Conjugate pairs theorem)
If f is a polynomial function with real number coefficients and z is a zero of f, then so is $\overline{z}$.

Even though a polynomial f has complex roots, we can still write down a real factorization involving linear factors corresponding to the real zeros of f and irreducible quadratic factors that give the complex zeros of f, as formalized in the following theorem.

Theorem 4.5 (Real factorization theorem)
Suppose f is a polynomial function with real coefficients. Then $f(x)$ can be factored into a product of linear factors corresponding to the real zeros of f and irreducible quadratic factors which give the complex zeros of f.

The value of Theorem 4.5 is illustrated in the next example.

Once we determined the zeros of an equation, we will often have to determine on which intervals the corresponding function is positive and on which intervals it is negative. For that purpose, we will use a so-called sign chart. For instance, consider the quadratic function $f(x) = x^2 - 3x - 4$. The zeros of $x^2 - 3x - 4 = 0$ are $x = -1$ and $x = 4$, and by choosing a few test values in the resulting intervals $]-\infty, -1]$, $[-1, 4]$ and $[4, +\infty[$ we can determine the sign of the function value at those points. This leads to the following sign diagram:

$x^2 - 3x - 4$	$+$	$\bullet$	$-$	$\bullet$	$+$
		-1		4	

where the dots indicate that the zeros belong to the function's domain, otherwise circles are used.

Splines

A spline is a function defined piecewise by polynomials and is used for addressing interpolating problems. The term spline comes from the flexible spline devices used by shipbuilders and draftsmen to draw smooth shapes.

Een spline of degree n on the interval $[a, b]$ is a function S on that interval consisting of a concatenation of k polynomials S_i defined on the subinterval $[x_{i-1}, x_i]$ in $[a, b]$. Hence,

$$S(x) = \begin{cases} S_0(x) & x \in [x_0, x_1] \\ S_1(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ S_{k-1}(x) & x \in [x_{k-1}, x_k], \end{cases}$$

where $a = x_0 < x_1 < \ldots < x_{k-1} < x_k = b$ and $S_i(x)$ is a polynomial of a degree not higher than n.

4.2 Rational functions

4.2.1 Definition

If we add, subtract or multiply polynomial functions according to the function arithmetic rules, we will produce another polynomial function. If, on the other hand, we divide two polynomial functions, the result may not be a polynomial. Here we study we study rational functions - functions which are ratios of polynomials.

Definitie 4.5 (Rational function)
A **rational function** (*rationale functie*) is a function which is the ratio of polynomial functions. Said differently, h is a rational function if it is of the form

$$h(x) = \frac{p(x)}{q(x)},$$

where p and q are polynomial functions.

Note that according to this definition, all polynomial functions are also rational functions. By taking $q(x) = 1$. Rational functions are used in numerical analysis for interpolation and approximation of functions. Approximations in terms of rational functions are well suited for computer algebra systems and other numerical software because they can be evaluated straightforwardly. They are also used to approximate or model more complex equations in science and engineering including fields and forces in physics, spectroscopy in analytical chemistry, enzyme kinetics in biochemistry, medicine concentrations in vivo, wave functions for atoms and molecules, and so on.

Obviously, we have domain issues anytime the denominator of a fraction is zero. In the example below, we review this concept as well as some of the arithmetic of rational expressions.

Example 4.5

Find the domain of the following rational functions. Write them in the form $\frac{p(x)}{q(x)}$ for polynomial functions p and q and simplify.

1. $f(x) = \dfrac{2x-1}{x+1}$
2. $g(x) = 2 - \dfrac{3}{x+1}$
3. $h(x) = \dfrac{2x^2-1}{x^2-1} - \dfrac{3x-2}{x^2-1}$
4. $r(x) = \dfrac{2x^2-1}{x^2-1} \div \dfrac{3x-2}{x^2-1}$

Solution

1. To find the domain of f, we find the zeros of the denominator and exclude them from the domain. Setting $x+1=0$ results in $x=-1$. Hence, our domain is $\mathbb{R}\backslash\{-1\}$. The expression $f(x)$ is already in the form requested and when we check for common factors among the numerator and denominator we find none, so we are done.

2. Proceeding as before, we determine the domain of g by solving $x+1=0$. As before, we find the domain of g is $\mathbb{R}\backslash\{-1\}$. To write $g(x)$ in the form requested, we need to get a common denominator.

$$\begin{aligned} g(x) &= 2-\frac{3}{x+1} &&= \frac{2(x+1)}{x+1}-\frac{3}{x+1} \\ &= \frac{(2x+2)-3}{x+1} &&= \frac{2x-1}{x+1} \end{aligned}$$

This formula is now completely simplified.

3. The denominators in the formula for $h(x)$ are both x^2-1 whose zeros are $x=\pm1$. As a result, the domain of h is $\mathbb{R}\backslash\{-1,1\}$. We now proceed to simplify $h(x)$:

$$\begin{aligned} h(x) &= \frac{2x^2-1}{x^2-1}-\frac{3x-2}{x^2-1} \\ &= \frac{2x^2-1-3x+2}{x^2-1} &&= \frac{2x^2-3x+1}{x^2-1} \\ &= \frac{(2x-1)(x-1)}{(x+1)(x-1)} &&= \frac{(2x-1)\cancel{(x-1)}}{(x+1)\cancel{(x-1)}} \\ &= \frac{2x-1}{x+1} \end{aligned}$$

Note that it is important to find the domain of h before simplifying the expression defining it. Otherwise, we would get that the domain of h is $\mathbb{R}\setminus\{-1\}$ instead of the correct $\mathbb{R}\setminus\{-1,1\}$.

4. To find the domain of r, it may help to temporarily rewrite $r(x)$ as

$$r(x)=\frac{\dfrac{2x^2-1}{x^2-1}}{\dfrac{3x-2}{x^2-1}}.$$

We need to set all of the denominators equal to zero which means we need to solve not only $x^2-1=0$, but also $\frac{3x-2}{x^2-1}=0$. We find $x=\pm1$ for the former and $x=2/3$ for the latter. Our domain is $\mathbb{R}\backslash\{-1,2/3,1\}$. We simplify $r(x)$:

$$\begin{aligned} r(x) &= \frac{2x^2-1}{x^2-1}\div\frac{3x-2}{x^2-1} &&= \frac{2x^2-1}{x^2-1}\cdot\frac{x^2-1}{3x-2} \\ &= \frac{\left(2x^2-1\right)\cancel{\left(x^2-1\right)}}{\cancel{\left(x^2-1\right)}(3x-2)} &&= \frac{2x^2-1}{3x-2} \end{aligned}$$

A few remarks about Example 4.5 are in order. Note that the expressions for $f(x)$, $g(x)$ and $h(x)$ work out to be the same. However, only two of these functions are actually equal. Recall that functions are ultimately sets of ordered pairs (Definition 3.3), so for two functions to be equal, they need, among other things, to have the same domain. Since $f(x)=g(x)$ and f and g have the same domain, they are equal functions. Even though the formula $h(x)$ is the same as $f(x)$, the domain of h is different than the domain of f, and thus they are different functions.

4.2.2 Graphs of rational functions

In Part II we will introduce the tools that are needed to graph rational functions and understand their behaviour. Still, here we already want to underline a few distinctive features of the graph of any rational function. For that purpose, consider the graph of the following rational function

$$f(x) = \frac{2x-1}{x+1}, \tag{4.3}$$

which is shown in Figure 4.4.

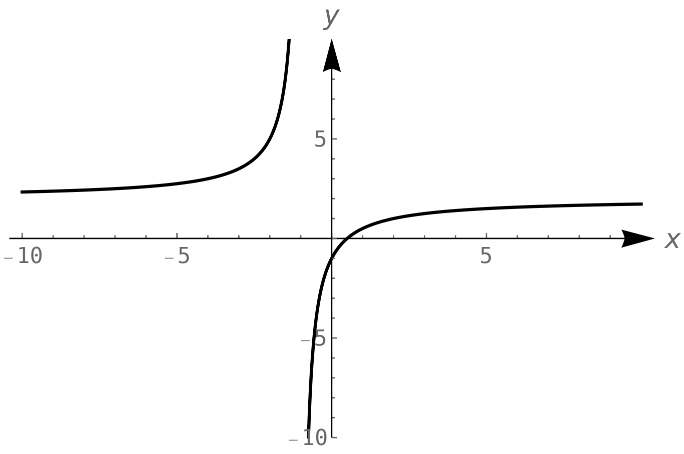


Figure 4.4: The graph of $f(x) = \frac{2x-1}{x+1}$.

First, note that the graph appears to break at $x = -1$. We know from Example 4.5 that $x = -1$ is not in the domain of f which means $f(-1)$ is undefined. We see that we can get near $x = -1$ from two directions.

As the x-values approach -1 from the left, the function values become larger and larger positive numbers. We express this symbolically by writing $x \underset{<}{\to} -1$, $f(x) \to +\infty$, or alternatively $x \to -1^-$, $f(x) \to +\infty$. Similarly, using analogous notation, we conclude that as $x \underset{>}{\to} -1$, $f(x) \to -\infty$, or equivalently $x \to -1^+$, $f(x) \to -\infty$. Here the > means approaching from above and < means approaching from below. For this type of unbounded behavior, we say the graph of $y = f(x)$ has a **vertical asymptote** (*verticale asymptoot*) of $x = -1$.

The other feature worthy of note about the graph of $y = f(x)$ is that it seems to level off on the left and right hand sides of the plot window. We see that as $x \to -\infty$, $f(x)$ approaches 2 coming from values larger than 2 (from above) and as $x \to +\infty$, $f(x)$ approaches 2 though coming from values smaller than 2 (from below). In this case, we say the graph of $y = f(x)$ has a **horizontal asymptote** (*horizontale asymptoot*) of $y = 2$. We formalize the concepts of vertical and horizontal asymptotes in Chapter 8. We then also introduce the notion of a slant asymptote.

Example 4.6

A mathematical model for the population P [–], in thousands, of a certain species of bacteria, t [+] days after it is introduced to an environment is given by

$$P(t) = \frac{100}{(5-t)^2},$$

for $0 \leq t < 5$.

1. Find and interpret $P(0)$.
2. When will the population reach 100 000?
3. Determine the behavior of P as $t \underset{<}{\to} 5$. Interpret this result graphically and within the context of the problem.

Solution

1. Substituting $t = 0$ gives $P(0) = \frac{100}{(5-0)^2} = 4$, which means 4000 bacteria are initially introduced into the environment.
2. We remember that $P(t)$ is measured in thousands, so, 100 000 bacteria corresponds to $P(t) = 100$. Substituting for $P(t)$ gives
$$\frac{100}{(5-t)^2} = 100\,,$$
whose solution is $t = 4$ or $t = 6$. Of these two solutions, only $t = 4$ lies in our domain, so this is the solution we keep. Hence, it takes 4 days for the population of bacteria to reach 100 000.
3. To determine the behaviour of P as $t \underset{<}{\to} 5$, we can make a table

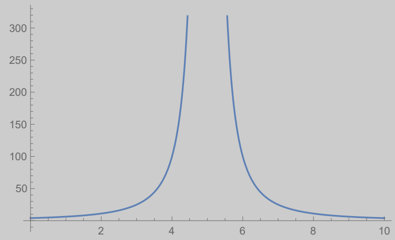
t	$P(t)$
4.9	10 000
4.99	1000 000
4.999	100 000 000
4.9999	10 000 000 000

In other words, as $t \underset{<}{\to} 5$, $P(t) \to +\infty$. The line $t = 5$ is a vertical asymptote of the graph of $y = P(t)$. This means that the population of bacteria is increasing without bound as we near 5 days, which cannot actually happen. For this reason, $t = 5$ is called the doomsday for this population. There is no way any environment can support infinitely many bacteria, so shortly before $t = 5$ the environment would collapse.

In Mathematica, we can verify the correctness of our reasoning above by plotting the studied function using the built-in function **`Plot`** as follows.

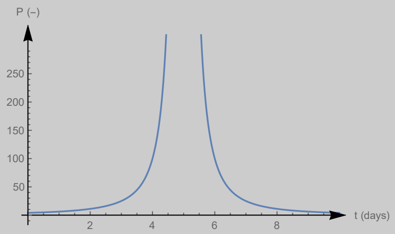
In[5]:=
```
Plot[(100)/(5-t)^2,{t,0,10}]
```

Out[5]=



This plot becomes, however, more informative upon adding the axis labels and directions as follows.

```
In[6]:= Plot[(100)/(5-t)^2,{t,0,10},     AxesLabel →{"t (days)","P (-)"},
        AxesStyle→Arrowheads[{0,0.05}]]
```

Out[6]=


4.3 Irrational functions

This section serves as a watershed for functions which are combinations of polynomial, and more generally, rational functions, with the operations of radicals, such as

$$f(x) = \sqrt{1-x^2},$$

$$g(x) = \sqrt[4]{\frac{16x}{x^2-9}},$$

and

$$h(x) = \sqrt[3]{x^3+3x^2-6x-8}.$$

It is business of Part II to introduce the tools to better understand the behaviour of such functions in all the detail. Here we restrict to the basics to help shore up the requisite skills needed for a good understanding of the subsequent parts of this course. In literature, functions containing radicals are sometimes referred to as **irrational functions** (*irrationale functie*).

It is worth remarking that, in the light of Section 3.4, we could define $f(x) = \sqrt[n]{x}$ functionally as the inverse of $g(x) = x^n$ with the stipulation that when n is even, the domain of g is restricted to $\mathbb{R}^+$. From what we know about $g(x) = x^n$ from Section 4.1, we can produce the graphs of $f(x) = \sqrt[n]{x}$ by reflecting the graphs of $g(x) = x^n$ across the line $y = x$. Figures 4.5(a)-4.5(c) show the graphs of $y = \sqrt{x}$, $y = \sqrt[4]{x}$ and $y = \sqrt[6]{x}$ obtained in this way from the graphs of $y = x^2$, $y = x^4$ and $y = x^6$, respectively. We can see the vertical steepening near $x = 0$ and the horizontal flattening as $x \to +\infty$. Likewise, Figures 4.5(d)-4.5(f) show the graphs of $y = \sqrt[3]{x}$, $y = \sqrt[5]{x}$ and $y = \sqrt[7]{x}$ obtained by reflecting the graphs of $y = x^3$, $y = x^5$ and $y = x^7$ about the axis $y = x$, respectively. These odd-indexed radical functions also follow a predictable trend - steepening near $x = 0$ and flattening as $x \to \pm\infty$.

Example 4.7

For the following functions, state their domains and determine their zero. Check your answer graphically using Mathematica.

1. $f(x) = 3x\sqrt[3]{2-x}$

Solution

1. As far as the domain is concerned, $f(x)$ has no denominators and no even roots, which means its domain is $\mathbb{R}$. Its zeros can be found as follows.

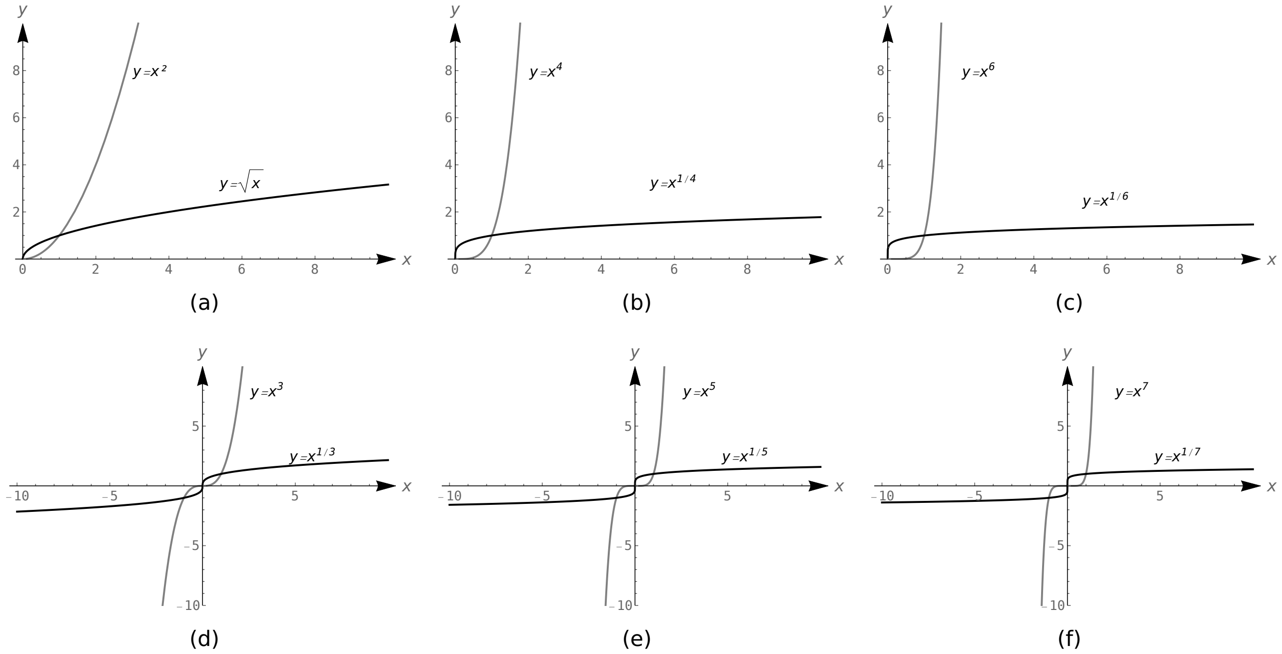


Figure 4.5: Graphs of $y=\sqrt{x}$ (a), $y=\sqrt[4]{x}$ (b), $y=\sqrt[6]{x}$ (c), $y=\sqrt[3]{x}$ (d), $y=\sqrt[5]{x}$ (e) and $y=\sqrt[7]{x}$ (f).

$$
\begin{array}{rrcl}
 & f(x) & = & 0 \\
\Leftrightarrow & 3x\sqrt[3]{2-x} & = & 0 \\
\Leftrightarrow & 3x=0 & \text{or} & \sqrt[3]{2-x}=0 \\
\Leftrightarrow & x=0 & \text{or} & 2-x=0 \\
\Leftrightarrow & x=0 & \text{or} & x=2
\end{array}
$$

The zeros 0 and 2 divide the real number line into three intervals, where the function may have a different behaviour. The graph of this function is shown in Figure 4.6.

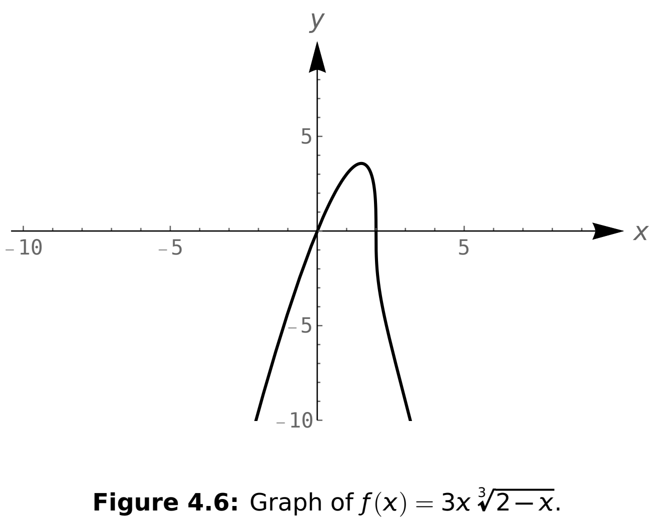


Figure 4.6: Graph of $f(x)=3x\sqrt[3]{2-x}$.

We conclude this section with an application in which an irrational function pops up naturally.

Example 4.8

Carl wishes to get high speed internet service installed in his remote bird observation post located

30 metres from the E34. The nearest junction box is located 50 metres downroad (Figure 4.7). Suppose it costs €15 per metre to run cable along the road and €20 per metre to run cable off of the road.

Express the total cost C of connecting the Junction Box to the Outpost as a function of x, the number of metres the cable is run along the E34 before heading off road directly towards the Outpost. Determine a reasonable applied domain for the problem.

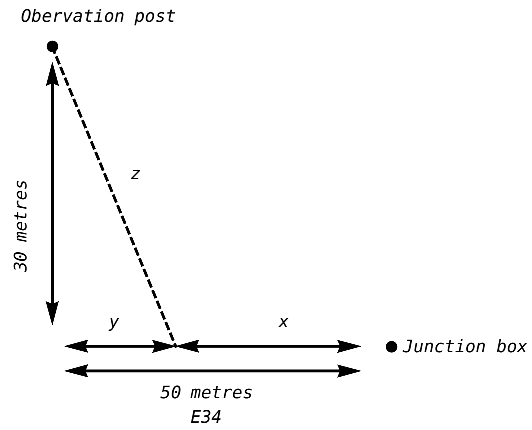


Figure 4.7: Map showing the observation post and junction box in Example 4.8.

Solution

The cost is broken into two parts: the cost to run cable along the E34 at €15 per metre, and the cost to run it off road at €20 per metre. Since x represents the metres of cable run along the E34, the cost for that portion is $15x$. From Figure 4.7, we see that the number of metres the cable is run off road is z, so the cost of that portion is $20z$.

Hence, the total cost is $C = 15x + 20z$. Our next goal is to determine z as a function of x. The diagram suggests we can use the Pythagorean theorem to get $y^2 + 30^2 = z^2$. But we also see $x + y = 50$ so that $y = 50 - x$. Hence, $z^2 = (50 - x)^2 + 900$. Solving for z, we obtain $z = \pm\sqrt{(50-x)^2 + 900}$. Since z represents a distance, we choose $z = \sqrt{(50-x)^2 + 900}$ so that our cost as a function of x only is given by

$$C(x) = 15x + 20\sqrt{(50-x)^2 + 900}.$$

From the context of the problem, we have $0 \leq x \leq 50$.

4.4 Conic sections

In this section, we will investigate the two-dimensional figures that are formed when a right circular cone is intersected by a plane.

History of conic sections

The Greek mathematician Menaechmus (c. 380-c. 320 BCE) is generally credited with discovering the shapes formed by the intersection of a plane and a right circular cone. Depending on how he tilted the plane when it intersected the cone, he formed different shapes at the intersection – beautiful shapes with near-perfect symmetry. It was also said that Aristotle may have had an intuitive understanding of these shapes, as he observed the orbit of the planet to be circular. He presumed that the planets moved in circular orbits around Earth, and for nearly 2000 years this was the commonly held belief.

It was not until the Renaissance movement that Johannes Kepler noticed that the orbits of the planet were not circular in nature. His law of planetary motion in the 1600s changed our view of the solar system forever. He claimed that the sun was at one end of the orbits, and the planets revolved around the sun in an oval-shaped path.

4.4.1 Overview

The name **conic section** (*kegelsnede*) is used to refer to any of the shapes that can be formed by intersecting a double-napped right circular cone with a plane. There are indeed several ways to intersect such a cone by a plane. Let us first consider a plane that does not contain the cone's vertex. Then, we can slice its top nappe, for instance, with a horizontal plane. This produces a circle (Figure 4.8(a)). Tilting this plane only slightly produces an ellipse (Figure 4.8(b)), while tilting the plane even further leads to a parabola (Figure 4.8(c)). If we continue increasing the tilting angle like this, the plane will at some point cut through both nappes, giving rise to a hyperbola (Figure 4.8(d)).

If the slicing plane contains the vertex of the cone, we get the so-called degenerate conics, namely a point (Figure 4.9(a)), a line (Figure 4.9(b)) or two intersecting lines (Figure 4.9(c)).

In the remainder of this section we will review the non-degenerate cases in detail. Then, in Chapter 5 we will show how any quadratic equation of the form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

corresponds to a conic section.

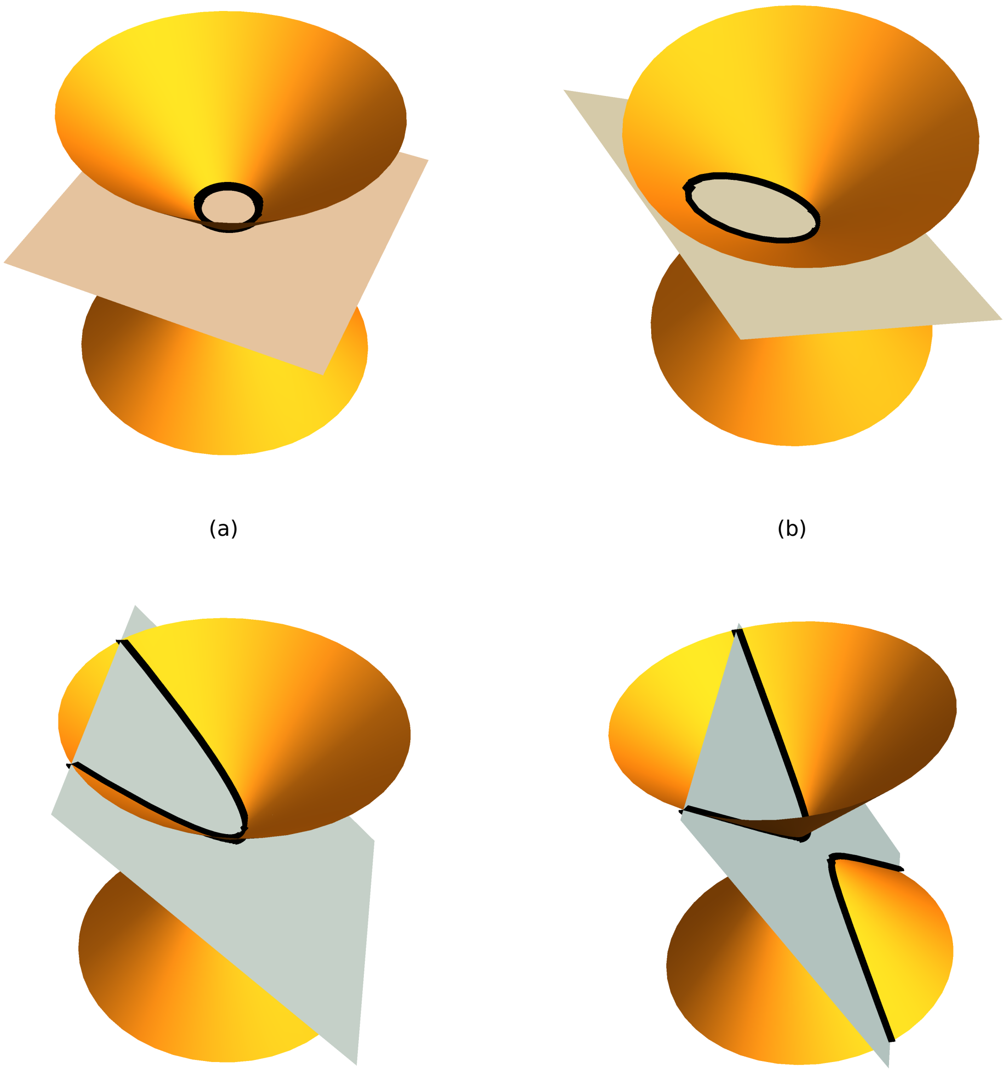
4.4.2 Circles

In geometry, a circle is defined as follows.

Definitie 4.6 (Circle)
A **circle** (*cirkel*) with **centre** (*middelpunt*) (x_0, y_0) and **radius** (*straal*) $r > 0$ is the set of all points (x, y) in the plane whose distance to (x_0, y_0) is r.

We express this definition algebraically using the distance formula as

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2}\,.$$



(a) (b)

(c) (d)

Figure 4.8: The intersection of a double-napped right circular cone with a plane with varying tilting angle: a circle (a), ellipse (b), parabola (c) and hyperbola (d).

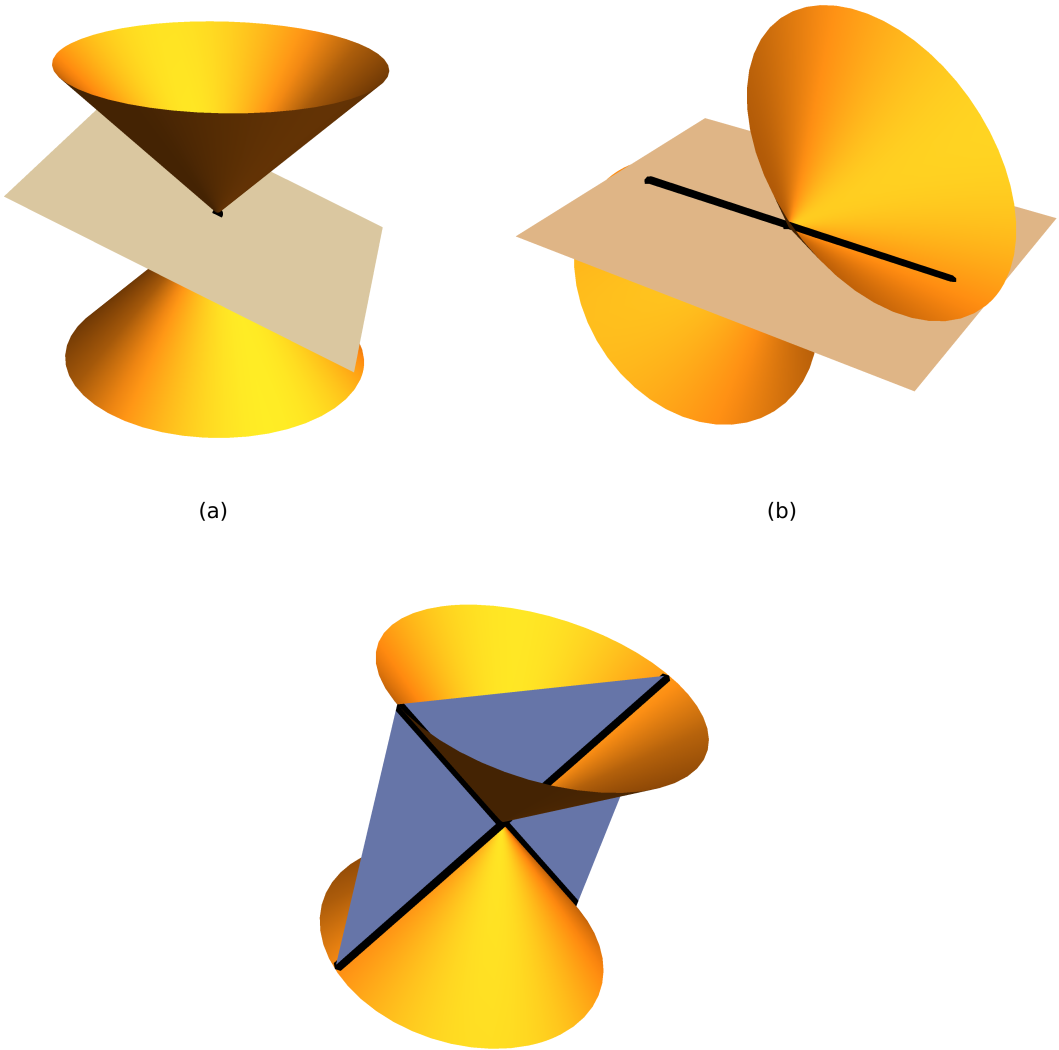
By squaring both sides of this equation, we get an equivalent equation (since $r > 0$) which gives us the standard equation of a circle with centre (x_0, y_0) and radius r:

$$(x - x_0)^2 + (y - y_0)^2 = r^2. \tag{4.4}$$

We close our introduction to circles with the most important circle in all of mathematics: the unit circle.

Definitie 4.7 (Unit circle)
The **unit circle** (*eenheidscirkel*) is the circle centred at $(0, 0)$ with a radius of 1. The standard equation of the unit circle is

$$x^2 + y^2 = 1.$$



(a) (b)

(c)

Figure 4.9: The intersection of a double-napped right circular cone with a plane containing the vertex of the cone and with varying tilting angle: a point (a), line (b) and two intersecting lines (c).

4.4.3 Ellipses

In the definition of a circle, (Definition 4.6), we fixed a point called the centre and considered all of the points which were at a fixed distance r from that one point. For the ellipse, we fix two distinct points and a distance d.

Definitie 4.8 (Ellipse)
Given two distinct points F_1 and F_2 in the plane and a fixed distance d, an **ellipse** (*ellips*) is the set of all points (x, y) in the plane such that the sum of each of the distances from F_1 and F_2 to (x, y) is d. The points F_1 and F_2 are called the **foci** (*brandpunten*) of the ellipse.

We may imagine taking a length of string and anchoring it to two points on a piece of paper. The curve traced out by taking a pencil and moving it so the string is always taut is an ellipse (Figure 4.10).

The centre of the ellipse is the midpoint of the line connecting the two foci. The **major axis** (*grote as*) of the ellipse is the line segment connecting two opposite ends of the ellipse which also contains the

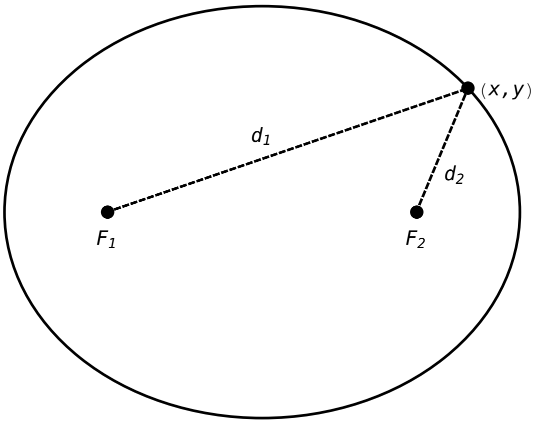


Figure 4.10: $d_1 + d_2 = d$ for all (x, y) on the ellipse.

centre and foci. The **minor axis** (*kleine as*) of the ellipse is the line segment connecting two opposite ends of the ellipse which contains the centre but is perpendicular to the major axis. The **vertices** (*top*) of an ellipse are the points of the ellipse which lie on the major axis. Notice that the centre is also the midpoint of the major axis (Figure 4.10) and that the major axis is the longer of the two axes through the centre, and likewise, the minor axis is the shorter of the two.

In order to derive the standard equation of an ellipse, we assume that the ellipse has its centre at $(0,0)$, its major axis along the x-axis, and has foci $(c,0)$ and $(-c,0)$ and vertices $(-a,0)$ and $(a,0)$ (Figure 4.11). We will label the y-intercepts of the ellipse as $(0,b)$ and $(0,-b)$. Moreover, we assume a, b, and c are all positive numbers.

Note that since $(a,0)$ is on the ellipse, it must satisfy the conditions of Definition 4.8. That is, the distance from $(-c,0)$ to $(a,0)$ plus the distance from $(c,0)$ to $(a,0)$ must equal the fixed distance d. Since all of these points lie on the x-axis, we get

$$\begin{aligned} (a+c)+(a-c) &= d \\ \Leftrightarrow \quad 2a &= d. \end{aligned}$$

We now use that fact $(0,b)$ is on the ellipse, along with the fact that $d = 2a$ to get

$$\begin{array}{rrcl} & \text{distance from } (-c,0) \text{ to } (0,b) + \text{distance from } (c,0) \text{ to } (0,b) &=& 2a \\ \Leftrightarrow & \sqrt{(0-(-c))^2+(b-0)^2}+\sqrt{(0-c)^2+(b-0)^2} &=& 2a \\ \Leftrightarrow & 2\sqrt{b^2+c^2} &=& 2a \\ \Leftrightarrow & \sqrt{b^2+c^2} &=& a\,. \end{array}$$

From this, we get $a^2 = b^2 + c^2$, or $b^2 = a^2 - c^2$, which will prove useful later. Now consider an arbitrary point (x,y) on the ellipse. Applying Definition 4.8, we get

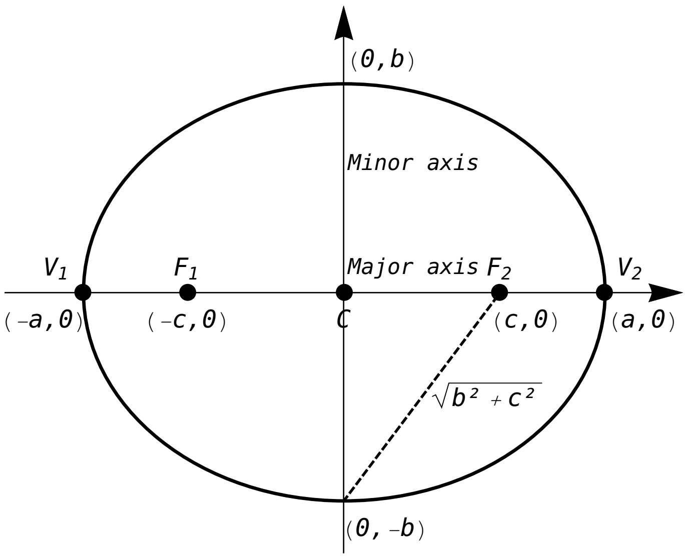


Figure 4.11: An ellipse with centre C $(0,0)$, foci F_1 $(-c,0)$, F_2 $(c,0)$ and vertices V_1 $(-a,0)$ and V_2 $(a,0)$.

$$
\begin{array}{rrcll}
 & \sqrt{(x+c)^2+y^2}+\sqrt{(x-c)^2+y^2} & = & 2a & \\
\Leftrightarrow & \sqrt{(x+c)^2+y^2} & = & 2a-\sqrt{(x-c)^2+y^2} & \\
\Rightarrow & (x+c)^2+y^2 & = & 4a^2-4a\sqrt{(x-c)^2+y^2}+(x-c)^2+y^2 & \text{(Square both sides.)} \\
\Leftrightarrow & 4a\sqrt{(x-c)^2+y^2} & = & 4a^2+(x-c)^2-(x+c)^2 & \\
\Leftrightarrow & 4a\sqrt{(x-c)^2+y^2} & = & 4a^2-4cx & \\
\Leftrightarrow & a\sqrt{(x-c)^2+y^2} & = & a^2-cx & \\
\Rightarrow & a^2\left((x-c)^2+y^2\right) & = & a^4-2a^2cx+c^2x^2 & \text{(Square both sides.)} \\
\Leftrightarrow & a^2x^2-2a^2cx+a^2c^2+a^2y^2 & = & a^4-2a^2cx+c^2x^2 & \\
\Leftrightarrow & a^2x^2-c^2x^2+a^2y^2 & = & a^4-a^2c^2 & \\
\Leftrightarrow & \left(a^2-c^2\right)x^2+a^2y^2 & = & a^2\left(a^2-c^2\right). &
\end{array}
$$

Recall now that $b^2=a^2-c^2$ so that

$$
\begin{array}{rrcl}
 & \left(a^2-c^2\right)x^2+a^2y^2 & = & a^2\left(a^2-c^2\right) \\
\Leftrightarrow & b^2x^2+a^2y^2 & = & a^2b^2 \\
\Leftrightarrow & \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2} & = & 1.
\end{array}
$$

This equation is for an ellipse centred at the origin. To get the formula for the ellipse centred at (x_0,y_0), we could use the transformations from Section 3.2.5 to obtain:

$$\frac{(x-x_0)^2}{a^2}+\frac{(y-y_0)^2}{b^2}=1. \tag{4.5}$$

Note that if $a>b$, then we have an ellipse whose major axis is horizontal, and hence, the foci lie to the left and right of the centre. If $b>a$, the roles of the major and minor axes are reversed, and the foci lie above and below the centre. Finally, it is worth mentioning that if $a=b$, we arrive at the standard equation of a circle. This indicates that it is fine to think of an ellipse as the deformation of a circle in which the circle is stretched farther in one direction than the other.

As with circles, an equation may be given which is an ellipse, but is not in the standard form of

Equation (4.5). In those cases, we will need to massage the given equation into the standard form by taking the following steps:

1. Group the same variables together on one side of the equation and position the constant on the other side.
2. Complete the square in both variables as needed.
3. Divide both sides by the constant term so that the constant on the other side of the equation becomes 1.

Example 4.9

Graph $x^2+4y^2-2x+24y+33=0$. Find the centre, the major and minor axes, the vertices, the endpoints of the minor axis, and the foci.

Solution

Since we have a sum of squares and the squared terms have unequal coefficients, it is a good bet we have an ellipse on our hands. We need to complete both squares, and then divide, if necessary, to get the right-hand side equal to 1.

$$\begin{array}{rrcl} & x^2+4y^2-2x+24y+33 &=& 0 \\ \Leftrightarrow & x^2-2x+4y^2+24y &=& -33 \\ \Leftrightarrow & x^2-2x+4\left(y^2+6y\right) &=& -33 \\ \Leftrightarrow & \left(x^2-2x+1\right)+4\left(y^2+6y+9\right) &=& -33+1+4(9) \\ \Leftrightarrow & (x-1)^2+4(y+3)^2 &=& 4 \\ \Leftrightarrow & \dfrac{(x-1)^2}{4}+\dfrac{(y+3)^2}{1} &=& 1 \end{array}$$

Now that this equation is in the standard form, we see that $x-x_0$ is $x-1$ so $x_0=1$, and $y-y_0$ is $y+3$ so $y_0=-3$. Hence, our ellipse is centred at $(1,-3)$. We see that $a^2=4$ so $a=2$, and $b^2=1$ so $b=1$. Consequently, the major axis will lie along the horizontal line $y=-3$, which means the minor axis lies along the vertical line $x=1$. The vertices are the points on the ellipse which lie along the major axis so in this case, they are the points $(-1,-3)$ and $(3,-3)$, and the endpoints of the minor axis are $(1,-2)$ and $(1,-4)$. To find the foci, we find $c=\sqrt{4-1}=\sqrt{3}$, which means the foci lie $\sqrt{3}$ units from the centre. Since the major axis is horizontal, the foci lie $\sqrt{3}$ units to the left and right of the centre, at $(1-\sqrt{3},-3)$ and $(1+\sqrt{3},-3)$ (Figure 4.12).

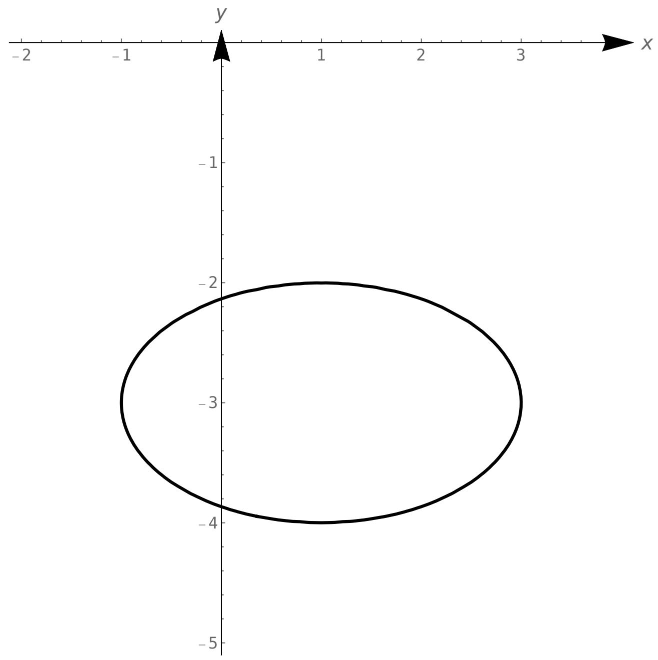


Figure 4.12: Graph of $x^2+4y^2-2x+24y+33=0$.

Johannes Kepler discovered that the orbits along which the planets travel around the Sun are ellipses with the Sun approximately at one focus (Figure 4.13). A key feature of such planetary orbits are their eccentricity, which is a measure of the roundness of an ellipse and can be quantified as below.

Definitie 4.9 (Eccentricity)
The **eccentricity** (*excentriciteit*) of an ellipse, denoted e, is the following ratio:

$$e = \frac{\text{distance from the centre to a focus}}{\text{distance from the centre to a vertex}}.$$

From this definition, we infer that for a circle $e=0$, while we have for an ellipse that $e<1$.

Finally, it is important to underline that ellipses have a reflective property. If we imagine the dashed lines in Figure 4.10 representing sound waves, then the waves emanating from one focus reflect off the top of the ellipse and head towards the other focus. Such geometry is exploited in the construction of so-called whispering galleries (Figure 4.14). If a person whispers at one focus, a person standing at the other focus will hear the first person as if they were standing right next to them. We explore this in our last example.

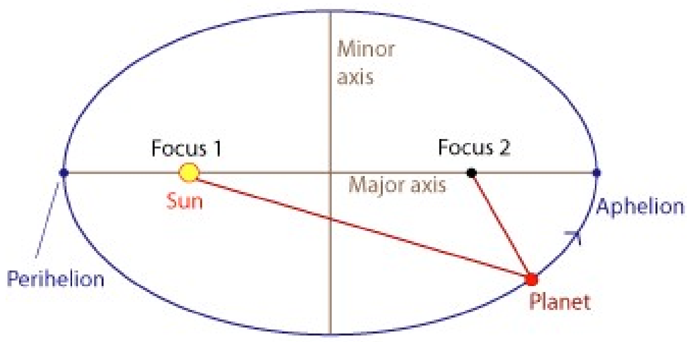


Figure 4.13: An elliptical orbit of a planet.

Example 4.10

Lisa and Jason want to exchange secrets from across a crowded whispering gallery. If the room is 40 metres high at the centre and 100 metres wide at the floor, how far from the outer wall should each of them stand so that they will be positioned at the foci of the ellipse?

Solution

It is most convenient to imagine this ellipse centred at $(0,0)$. Since the ellipse is 100 units wide and 40 units tall, we get $a = 50$ and $b = 40$. Hence, our ellipse has the equation

$$\frac{x^2}{50^2} + \frac{y^2}{40^2} = 1.$$

We are looking for the foci, and we get $c = \sqrt{50^2 - 40^2} = \sqrt{900} = 30$, so that the foci are 30 units from the centre. That means they are $50 - 30 = 20$ units from the vertices. Hence, Jason and Lisa should stand 20 metres from opposite ends of the gallery.

Figure 4.14: Whispering gallery at Grand Central station, New York, United states.

4.4.4 Parabolas

We have already learned in Section 4.1 that the graph of a quadratic function $f(x) = ax^2 + bx + c$ ($a \neq 0$) is called a parabola. We may also define parabolas in terms of distance.

Definitie 4.10 (Parabola)
Let F be a point in the plane and d be a line not containing F. A **parabola** (*parabool*) is the set of all points equidistant from F and d. The point F is called the **focus** (*brandpunt*) of the parabola and the line d is called the **directrix** (*richtlijn*) of the parabola.

Essentially, in Figure 4.15, each dashed line from the point F to a point on the curve has the same length as the dashed line from the point on the curve to the line d. The point V is the **vertex** (*top*). The vertex is the point on the parabola closest to the focus.

We want to use only the distance definition of parabola to derive the equation of a parabola and we should get an expression much like those studied in Section 4.1. Let p denote the directed distance from the vertex to the focus, which by definition is the same as the distance from the vertex to the directrix. For simplicity, assume that the vertex is $(0,0)$ and that the parabola opens upwards. Hence, the focus is $(0,p)$ and the directrix is the line $y = -p$. All this is presented schematically in Figure 4.16.

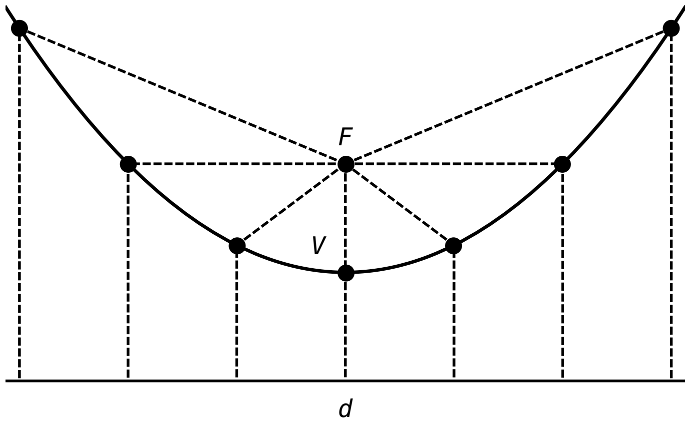


Figure 4.15: Geometric construction a a parabola.

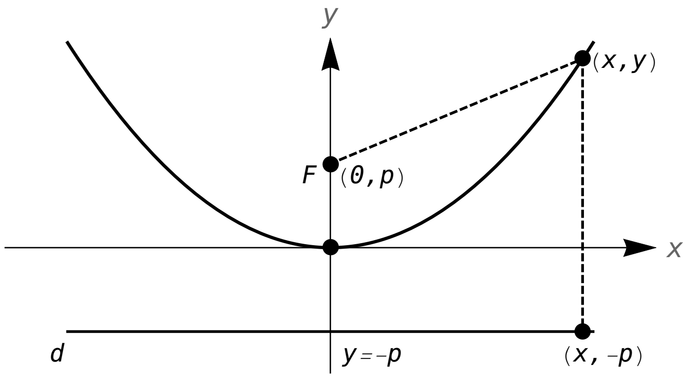


Figure 4.16: Parabola with vertex in $(0,0)$, directrix $d: y = -p$ and focus $F(0,p)$.

From the definition of parabola, we know that the distance from $(0,p)$ to (x,y) is the same as the distance from $(x,-p)$ to (x,y). Using the distance formula, we get

$$\begin{array}{rrcll}
 & \sqrt{(x-0)^2+(y-p)^2} & = & \sqrt{(x-x)^2+(y-(-p))^2} & \\
\Leftrightarrow & \sqrt{x^2+(y-p)^2} & = & \sqrt{(y+p)^2} & \\
\Leftrightarrow & x^2+(y-p)^2 & = & (y+p)^2 & \text{(Square both sides.)} \\
\Leftrightarrow & x^2+y^2-2py+p^2 & = & y^2+2py+p^2 & \text{(Expand quantities.)} \\
\Leftrightarrow & x^2 & = & 4py. & \text{(Gather like terms.)}
\end{array}$$

Solving for y yields $y = \frac{x^2}{4p}$, which is a quadratic function of the form found in Definition 4.2 with $a = \frac{1}{4p}$ and vertex $(0,0)$.

When $p < 0$, the parabola opens downwards. In our formulation, we say that p is a directed distance from the vertex to the focus: if $p > 0$, the focus is above the vertex; if $p < 0$, the focus is below the vertex. If we choose to place the vertex at an arbitrary point (x_0, y_0), we arrive at the standard equation of a vertical parabola using the transformations from Section 3.2.5:

$$(x-x_0)^2 = 4p(y-y_0)\,. \tag{4.6}$$

Notice that in this standard equation, only one of the variables, x, is squared. This is a quick way to distinguish an equation of a parabola from that of a circle or ellipse because in the equation of a circle or ellipse, both variables are squared.

If we interchange the roles of x and y, we can produce horizontal parabolas: parabolas which open to the left or to the right. The directrices of these would be vertical lines and the focus would either lie to the left or to the right of the vertex, as seen in Figure 4.17. The standard equation of a horizontal parabola with vertex (x_0, y_0) is

$$(y-y_0)^2 = 4p(x-x_0)\,. \tag{4.7}$$

If $p > 0$, the parabola opens to the right; if $p < 0$, it opens to the left.

As with circles and ellipses, not all parabolas will come to us in the forms in Equations (4.6) or (4.7). If we encounter an equation with two variables in which exactly one variable is squared, we can, however, put the equation into a standard form using the following steps.

1. Group the variable which is squared on one side of the equation and position the non-squared variable and the constant on the other side.
2. Complete the square if necessary and divide by the coefficient of the perfect square.
3. Factor out the coefficient of the non-squared variable.

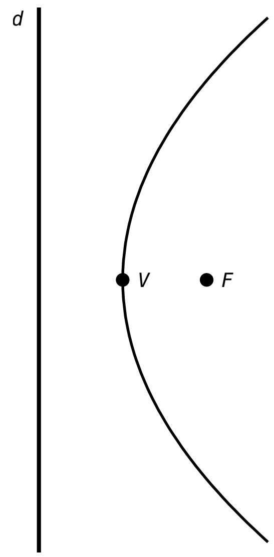


Figure 4.17: Horizontal parabola opening to the right.

Example 4.11

Consider the equation $y^2 + 4y + 8x = 4$. Put this equation into standard form and graph the parabola. Find the vertex, focus, and directrix.

Solution

We need a perfect square on the left-hand side of the equation and factor out the coefficient of the non-squared variable (x) on the other.

$$\begin{array}{rrcll}
 & y^2+4y+8x & = & 4 & \\
\Leftrightarrow & y^2+4y & = & -8x+4 & \\
\Leftrightarrow & y^2+4y+4 & = & -8x+4+4 & \text{(Complete the square in } y \text{ only.)} \\
\Leftrightarrow & (y+2)^2 & = & -8x+8 & \text{(Factor.)} \\
\Leftrightarrow & (y+2)^2 & = & -8(x-1) &
\end{array}$$

Now that the equation is in the form given in Equation (4.7), we see that $x - x_0$ is $x - 1$ so $x_0 = 1$, and $y - y_0$ is $y+2$ so $y_0 = -2$. Hence, the vertex is $(1,-2)$. We also see that $4p = -8$ so that $p = -2$. Since $p < 0$, the focus will be the left of the vertex and the parabola will open to the left. The distance from the vertex to the focus is $|p| = 2$, which means the focus is 2 units to the left of 1, so if we start at $(1,-2)$ and move left 2 units, we arrive at the focus $(-1,-2)$. The directrix, then, is 2 units to the right of the vertex, so if we move right 2 units from $(1,-2)$, we would be on the vertical line $x = 3$. Moreover, the parabola is 8 units wide at the focus.

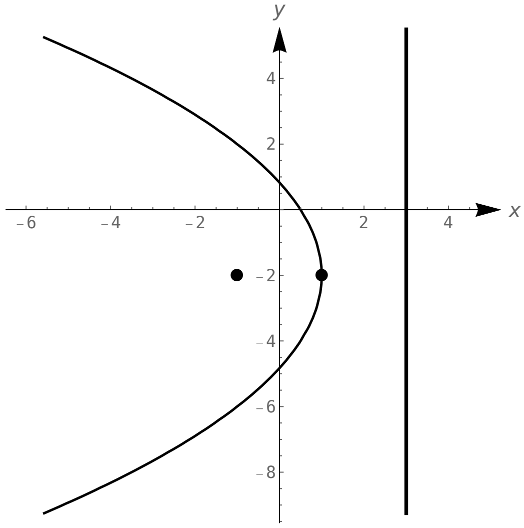
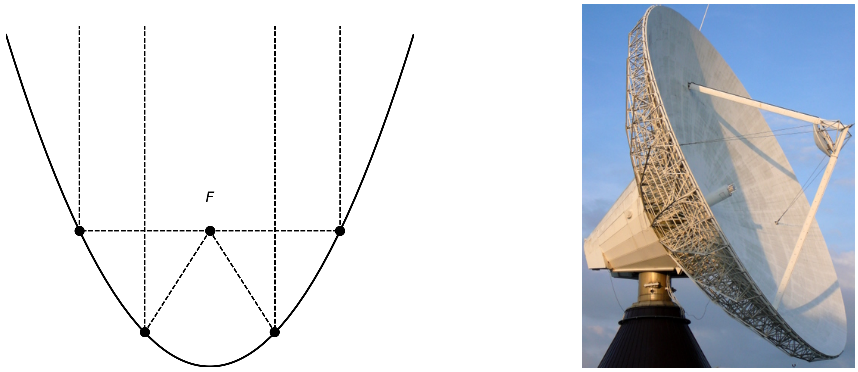


Figure 4.18: Graph of $y^2 + 4y + 8x = 4$.

Parabolas are used to model physical phenomena such as the trajectories of projectiles. Other applications of the parabola concern its reflective property. If we imagine the dashed lines in Figure 4.19(a) as emanating from the focus, we see that the waves are reflected off the parabola in a coherent fashion as in the case in a flash light. Here, the bulb is placed at the focus and the light rays are reflected off a parabolic mirror to give directional light. This reasoning also works the other way around if we imagine the dashed lines below as electromagnetic waves heading towards a parabolic dish. It turns out that the waves reflect off the parabola and concentrate at the focus which then becomes the optimal place for the receiver (Figure 4.19(b)).



(a) (b)

Figure 4.19: Reflective property of parabola (a) and a parabolic antenna making use thereof (b).

Example 4.12

A satellite dish is to be constructed in the shape of a paraboloid of revolution. If the receiver placed at the focus is located 2 metres above the vertex of the dish, and the dish is to be 12 metres wide, how deep will the dish be?

Solution

One way to approach this problem is to determine the equation of the parabola suggested to us by this data. For simplicity, we will assume the vertex is $(0,0)$ and the parabola opens upwards. Our standard form for such a parabola is $x^2 = 4py$. Since the focus is 2 units above the vertex, we know $p = 2$, so we have $x^2 = 8y$. Since the parabola is 12 metres wide, we know the edge is 6 metres from a vertical line through the vertex. To find the depth, we are looking for the y value when $x = 6$. Substituting $x = 6$ into the equation of the parabola yields $6^2 = 8y$ or $y = 36/8 = 4.5$. Hence, the dish will be 4.5 metres deep.

Examples of parabolas occurring in nature are also manifold. For instance, rainbows and many natural bridges have a parabolic shape. Moreover, the trajectory traversed by jumping fish or aquatic mammals also approaches a parabola (Figure 4.20).

(a) (b) (c)

Figure 4.20: Parabolas in nature: rainbow (a), natural bridge (b) and trajectories of jumping dolphins.

4.4.5 Hyperbolas

In the definition of an ellipse, Definition 4.8, we fixed two points called foci and looked at points whose distances to the foci always added to a constant distance d. But what, if any, curve we would generate if we replaced added with subtracted. The answer is a hyperbola.

Definitie 4.11 (Hyperbola)
Given two distinct points F_1 and F_2 in the plane and a fixed distance d, a **hyperbola** (*hyperbool*) is the set of all points (x, y) in the plane such that the absolute value of the difference of each of the distances from F_1 and F_2 to (x, y) is d. The points F_1 and F_2 are called the **foci** (*brandpunten*) of the hyperbola.

Note that the hyperbola has two parts, called **branches** (*tak*). The **centre** (*middelpunt*) of the hyperbola is the midpoint of the line connecting the two foci. The **transverse axis** (*hoofdas*) of the hyperbola is the line segment connecting two opposite ends of the hyperbola which also contains the centre and foci. The **vertices** (*top*) of a hyperbola are the points of the hyperbola which lie on the transverse axis. In addition, there are lines called **asymptotes** (*asymptoot*) which the branches of the hyperbola approach for large x- and y-values (Figure 4.21). The **conjugate axis** (*nevenas*) of a hyperbola is the line through the centre which is perpendicular to the transverse axis. It contains two **imaginary vertices** (*imaginaire toppen*).

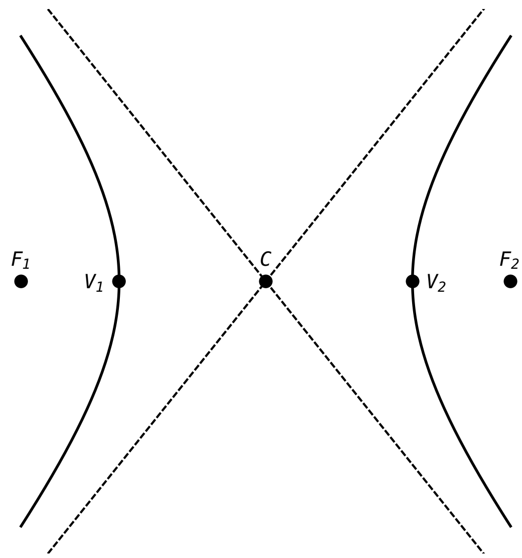


Figure 4.21: A hyperbola with centre C; foci F_1, F_2; and vertices V_1, V_2 and asymptotes (dashed)

Suppose now we wish to derive the equation of a hyperbola. For simplicity, we shall assume that the centre is $(0,0)$, the vertices are $(a,0)$ and $(-a,0)$ and the foci are $(c,0)$ and $(-c,0)$. We label the endpoints of the conjugate axis $(0,b)$ and $(0,-b)$. Although b does not enter into our derivation, we will have to justify this choice as you shall see later. As before, we assume a, b, and c are all positive numbers. Schematically we get the picture shown in Figure 4.22.

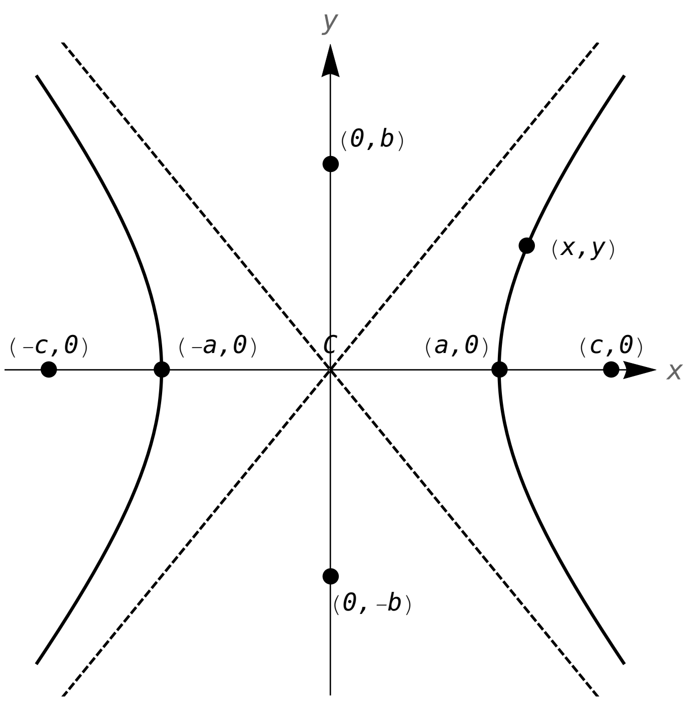


Figure 4.22: A hyperbola with centre in $(0,0)$; foci $F_1(-c,0)$, $F_2(c,0)$; and vertices $V_1(-a,0)$, $V_2(a,0)$, imaginary vertices $(0,-b)$ and $(0,b)$, and asymptotes (dashed).

Since $(a,0)$ is on the hyperbola, it must satisfy the conditions of Definition 4.11. That is, the distance from $(-c,0)$ to $(a,0)$ minus the distance from $(c,0)$ to $(a,0)$ must equal the fixed distance d. Since all these points lie on the x-axis, we get

$$\begin{array}{rrcl} & \text{distance from } (-c,0) \text{ to } (a,0) - \text{distance from } (c,0) \text{ to } (a,0) & = & d \\ \Leftrightarrow & (a+c)-(c-a) & = & d \\ \Leftrightarrow & 2a & = & d. \end{array}$$

Hence, d is the distance between the vertices V_1 and V_2.

Now consider a point (x,y) on the hyperbola. Applying Definition 4.11, we get

$$\begin{array}{rrcl} & \left|\text{distance from } (-c,0) \text{ to } (x,y) - \text{distance from } (c,0) \text{ to } (x,y)\right| & = & 2a \\ \Leftrightarrow & \left|\sqrt{(x-(-c))^2+(y-0)^2} - \sqrt{(x-c)^2+(y-0)^2}\right| & = & 2a \\ \Leftrightarrow & \left|\sqrt{(x+c)^2+y^2} - \sqrt{(x-c)^2+y^2}\right| & = & 2a\,. \end{array}$$

Following the same procedure as when deriving the standard formula of an ellipse (Equation (4.5)), we arrive at:

$$\left(a^2-c^2\right)x^2+a^2y^2 \quad = \quad a^2\left(a^2-c^2\right).$$

What remains is to determine the relationship between a, b and c. To that end, we note that since a and c are both positive numbers with $a < c$, we get $a^2 < c^2$ so that a^2-c^2 is a negative number. Hence, c^2-a^2 is a positive number. Let us rewrite the equation by solving for y^2/x^2 to get

$$\begin{array}{rrcl} & \left(a^2-c^2\right)x^2+a^2y^2 & = & a^2\left(a^2-c^2\right) \\ \Leftrightarrow & -\left(c^2-a^2\right)x^2+a^2y^2 & = & -a^2\left(c^2-a^2\right) \\ \Leftrightarrow & a^2y^2 & = & \left(c^2-a^2\right)x^2-a^2\left(c^2-a^2\right) \\ \Leftrightarrow & \dfrac{y^2}{x^2} & = & \dfrac{\left(c^2-a^2\right)}{a^2}-\dfrac{\left(c^2-a^2\right)}{x^2}\,. \end{array}$$

As x and y attain very large values, the quantity $\frac{(c^2-a^2)}{x^2} \to 0$ so that $\frac{y^2}{x^2} \to \frac{(c^2-a^2)}{a^2}$. By setting $b^2 = c^2-a^2$ we get $\frac{y^2}{x^2} \to \frac{b^2}{a^2}$. This shows that $y \to \pm\frac{b}{a}x$ as $|x|$ grows large. Thus $y = \pm\frac{b}{a}x$ are the asymptotes to the graph. In our equation of the hyperbola we can substitute $a^2-c^2 = -b^2$ which yields

$$\begin{array}{rrcl} & \left(a^2-c^2\right)x^2+a^2y^2 & = & a^2\left(a^2-c^2\right) \\ \Leftrightarrow & -b^2x^2+a^2y^2 & = & -a^2b^2 \\ \Leftrightarrow & \dfrac{x^2}{a^2}-\dfrac{y^2}{b^2} & = & 1\,. \end{array}$$

The equation above is for a hyperbola whose centre is the origin and which opens to the left and right.

If the hyperbola were centred at a point (x_0,y_0), we would get the following:

$$\frac{(x-x_0)^2}{a^2}-\frac{(y-y_0)^2}{b^2} = 1\,. \tag{4.8}$$

If the roles of x and y were interchanged, then the hyperbola's branches would open upwards and downwards and we would get a vertical hyperbola. Its standard equation is given by

$$\frac{(y-y_0)^2}{b^2} - \frac{(x-x_0)^2}{a^2} = 1\,. \tag{4.9}$$

By convention, a always refers to the coefficient in the denominator of the term containing x^2, while b refers to the coefficient appearing in the denominator of the term containing y^2.

The distance from the centre to the foci, c, as seen in the derivation, can be found by the formula $c = \sqrt{a^2+b^2}$. Also note that we can quickly distinguish the equation of a hyperbola from that of a circle or ellipse because the hyperbola formula involves a difference of squares where the circle and ellipse formulas both involve the sum of squares.

Determining the location of a known event has many practical uses such as: locating the epicentre of an earthquake, an airplane crash site, the position of the person speaking in a large room, etc.. To determine the location of an earthquake's epicentre, seismologists use trilateration. A seismograph allows one to determine how far away the epicentre was; using three separate readings, the location of the epicentre can be approximated.

Example 4.13

Consider three microphones at positions A, B and C which all record a noise (a person's voice, an explosion, etc.) created at unknown location D. The microphone does not know when the sound was created, only when the sound was detected. How can the location be determined in such a situation?

Solution

If each location has a clock set to the same time, hyperbolas can be used to determine the location. Suppose the microphone at position A records the sound at exactly 12:00, location B records the time exactly 1 second later, and location C records the noise exactly 2 seconds after that. We are interested in the difference of times. Since the speed of sound is approximately 340 m/s, we can conclude quickly that the sound was created 340 meters closer to position A than position B. If A and B are a known distance apart (as shown in Figure 4.23(a)), then we can determine a hyperbola on which D must lie.

The difference of distances between A and B is 340 metres; this is also the distance between vertices of the hyperbola. So we know $2a = 340$. Positions A and B lie on the foci, so $2c = 1000$. From this we can find $b \approx 470$ and can sketch the hyperbola, with equation

$$\frac{x^2}{170^2} - \frac{y^2}{470^2} = 1\,,$$

whose graph is shown in Figure 4.23(b). We only care about the side closest to A because the sound was first heard at that location.

We can also find the hyperbola defined by positions B and C. In this case, $2b = 680$ as the sound travelled an extra 2 seconds to get to C. We still have $2c = 1000$, centring this hyperbola at $(-500, 500)$. We find $a \approx 367$. This hyperbola has equation

$$\frac{(y-500)^2}{340^2} - \frac{(x+500)^2}{367^2} = 1\,,$$

and is sketched in Figure 4.23(c). The intersection point of the two graphs is the location of the sound, at approximately $(188, -222.5)$.

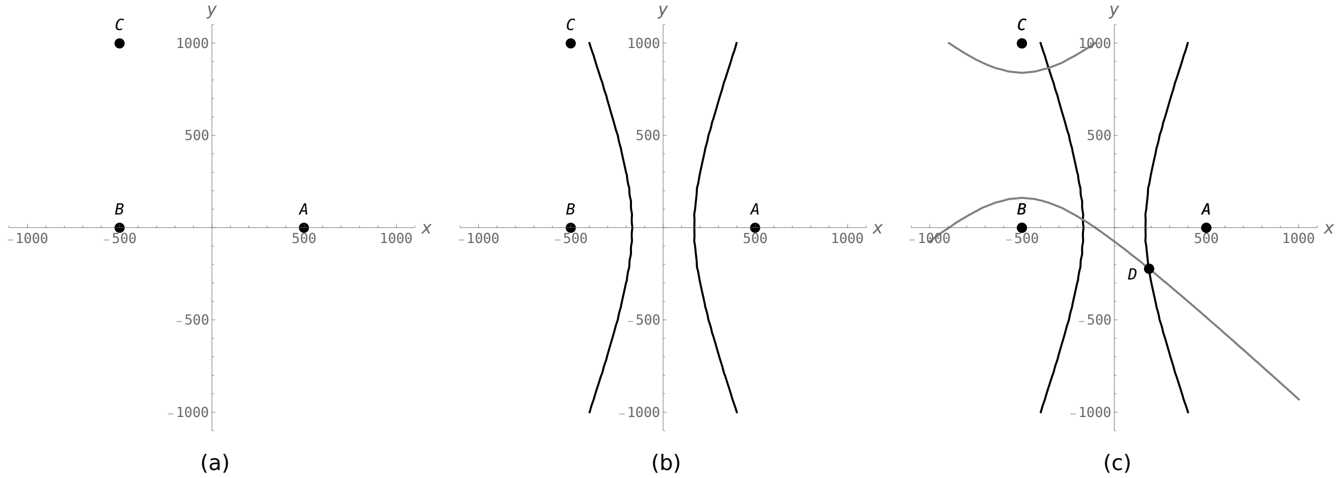


Figure 4.23: An example of trilateration using hyperbolas.

From *Math with Bad Drawings*, used by permission of Ben Orlin.

4.5 Exercises

Polynomial functions

Assignment 4.1 — Draw and describe the following areas.

(a) $y \leq x - 1$

(b) $|x| - 4 < y < 2 - x$

(c) $x^2 \leq y < x + 2$

(d) $x^2 + 2x + y^2 < 8$

(e) $x^2 + y^2 < 2x, \quad x^2 + y^2 < 2y$

(f) $x^2 + y^2 - 4x + 2y > 4, \quad x + y > 1$

Assignment 4.2 — Factorize the polynomials below into real factors.

(a) $2x^6 - 128$

(b) $8x^3 + 12x^2 + 6x + 1$

(c) $2x^3 + 3x^2 + 2x + 3$

(d) $x^4 - 7x^3 + 18x^2 - 20x + 8$

(e) $x^4 + x^3 - 2x^2 - 4x - 8$

(f) $2x^3 - 3x^2 - 3x + 2$

(g) $8a^3 - 60a^2b + 150ab^2 - 125b^3$

(h) $a^2 - 2a + 1 - b^2 - 4bc - 4c^2$

(i) $x^3 - 4x^2y + 7xy^2 - 4y^3$

(j) $2(x^2 + 6x + 1)^2 + 5(x^2 + 6x + 1)(x^2 + 1) + 2(x^2 + 1)^2$

Assignment 4.3 — Write each polynomial as a product of real factors.

(a) $16x^4 - 8x^2 + 1$

(b) $x^4 - 1$

(c) $x^5 - x^4 - 16x + 16$

(d) $x^5 + x^3 + 8x^2 + 8$

(e) $x^4 + 6x^3 + 9x^2$

(f) $x^6 - 3x^4 + 3x^2 - 1$

(g) $x^9 - 4x^7 - x^6 + 4x^4$

Assignment 4.4 — Solve the equations below in $\mathbb{C}$ and determine both the real and complex decomposition.

(a) $x^3 - 3x^2 + 20 = 0$

(b) $2x^3 - 4x^2 - 10x + 12 = 0$

(c) $x^6 - 16x^3 + 64 = 0$

(d) $8x^4 - 20x^3 + 18x^2 - 7x + 1 = 0$

(e) $x^3 - 16x^2 + 48x + 72 = 0$

(f) $4x^3 - 14x^2 + 8x + 8 = 0$

(g) $x^5 + 6x^4 + x^3 - 26x^2 - 32 = 0$

(h) $-2x^6 - 10x^5 - 16x^4 - 8x^3 = 0$

Assignment 4.5 — Solve the equation below in $\mathbb{R}$. Write the set of solutions as an interval.

(a) $-2x^3+19x^2-49x+20>0$

(b) $x^4-9x^2 \leq 4x-12$

(c) $3x^2+2x<x^4$

(d) $\dfrac{x^3+2x^2}{2}<x+2$

(e) $2x^4>5x^2+3$

Rational functions

Assignment 4.6 — Perform the divisions below.

(a) $\dfrac{1-5x^4+4x^5}{1-x}$

(b) $\dfrac{x^3-1}{x^2-2}$

(c) $\dfrac{x^2}{x^2+5x+3}$

(d) $\dfrac{x^3}{x^2+2x+3}$

(e) $\dfrac{2x^3-3x^2+4x-5}{x^2-6x+7}$

Assignment 4.7 — Solve the following rational equations in $\mathbb{R}$.

(a) $\dfrac{x^2-2x+1}{x^3+x^2-2x}=1$

(b) $\dfrac{x}{x^2-1}>0$

(c) $\dfrac{4x}{x^2+4}\geq 0$

(d) $\dfrac{3x^2-5x-2}{x^2-9}<0$

(e) $\dfrac{x^4-4x^3+x^2-2x-15}{x^3-4x^2}\geq x$

(f) $\dfrac{5x^3-12x^2+9x+10}{x^2-1}\geq 3x-1$

Assignment 4.8 — Determine the domain and intersections with the x-axis and y-axis of the rational functions below. Also determine any vertical and horizontal asymptotes.

(a) $f(x)=\dfrac{3x+2}{x^2+2x+2}$

(b) $f(x)=\dfrac{x^2-9}{x^3-x}$

(c) $f(x)=\dfrac{4}{x^3+x^2}$

(d) $f(x)=\dfrac{x^3+3x^2+6}{x^2+x-1}$

(e) $f(x)=\dfrac{x^2-x-12}{x^2+x-6}$

(f) $f(x)=\dfrac{x^2-x-6}{x+1}$

Irrational functions

Assignment 4.9 — Solve the following irrational equations in $\mathbb{R}$.

(a) $1+\dfrac{x+1}{\sqrt{x^2+2x}}=0$

(b) $\sqrt{2x+1}-\sqrt{x-1}=2$

(c) $\sqrt{x^2-3x+2}=|x|-2$

(d) $\sqrt[3]{x}+\sqrt[3]{x^2}+x>0$

(e) $\sqrt{3x+1}-\sqrt{x-4}=\sqrt{x+1}$

(f) $\sqrt{-x^2-x}\leq 2x+1$

Assignment 4.10 — In the theory of relativity, the mass m [M] of an object is not a constant quantity, but a variable quantity depending on the velocity v [LT^{-1}] of the object according to

$$m(v) = m_0 \frac{c}{\sqrt{c^2 - v^2}},$$

with $c = 299792.458$ km/s (the speed of light) and m_0 [M] the mass of the object at rest. What speed must an object have in order for its mass to be twice its resting mass?

Assignment 4.11 — Determine the domain and intersections with the x-axis and y-axis of the irrational functions below. Also determine any vertical asymptotes.

(a) $f(x) = |5 - \sqrt{8+2x}|$

(b) $f(x) = \sqrt[4]{\dfrac{16x}{x^2-9}}$

(c) $f(x) = \dfrac{5x}{\sqrt[3]{x^3+8}}$

(d) $f(x) = x^{\frac{2}{3}}(x-7)^{\frac{1}{3}}$

(e) $f(x) = \dfrac{1}{x\sqrt{x^2-4}}$

(f) $f(x) = \dfrac{\sqrt{x-3}}{\sqrt{2x+2} - \sqrt{x-1} - 2}$

(g) $f(x) = \dfrac{\sqrt{2x^2-3x-2}}{\sqrt[3]{x^3+3x^2+7} - x - 1}$

Assignment 4.12 — Determine which function belongs to which graph in Figure 4.24.

(a) $f(x) = \sqrt{\dfrac{1}{9-x^2}}$

(b) $f(x) = \sqrt{-2x+6}$

(c) $f(x) = \sqrt{x^3 - 3x^2 + 2x}$

(d) $f(x) = \sqrt{3x^2 - x + 2}$

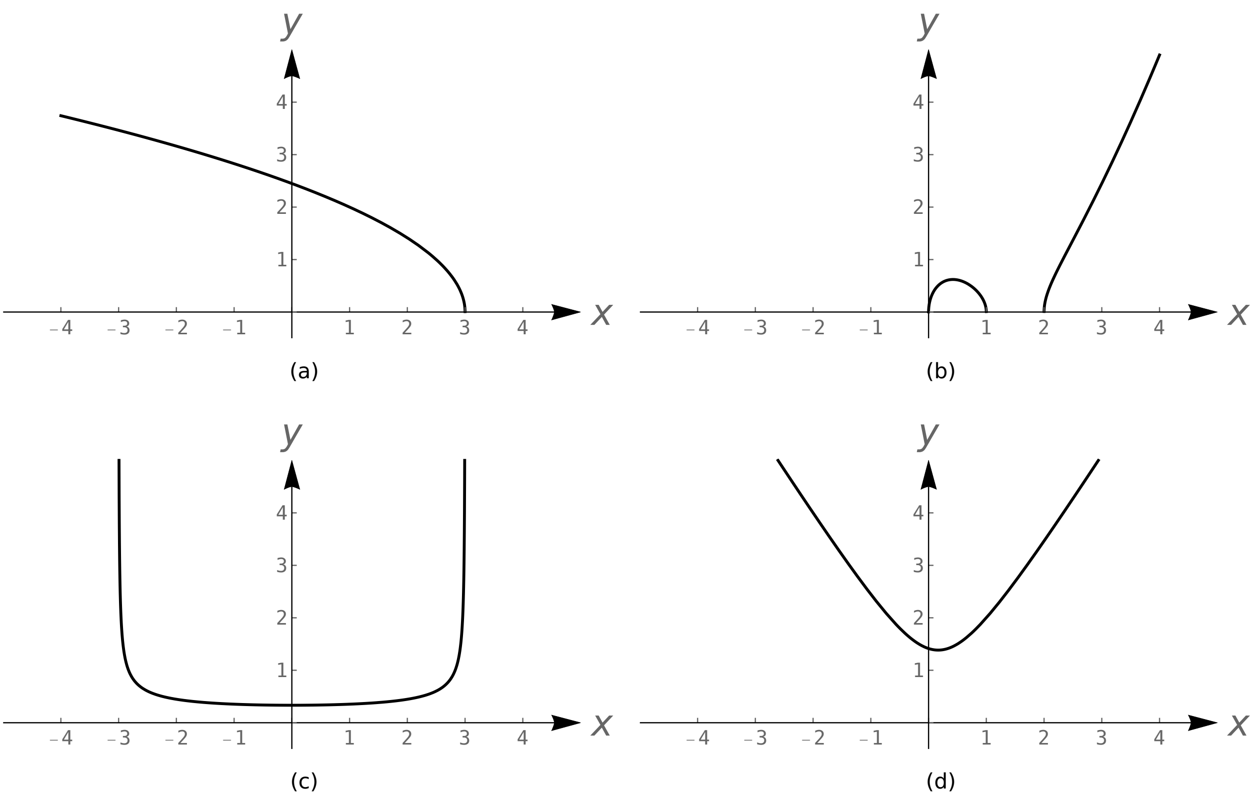


Figure 4.24: Graphs of the irrational functions in Exercise 15.

Conic sections

Assignment 4.13 — Determine the shape of the parabolas below. Find the top, the focal point, the axis of symmetry, the directrix, and the points of intersection with the x-axis and the y-axis. Draw the graph and determine the image.

(a) $y = x^2 - 4x + 3$

(b) $y = x^2 - 2x$

(c) $y^2 + 2y + 2x = 0$

(d) $x^2 + x + y = 0$

(e) $y = -3x^2 + 5x + 4$

Assignment 4.14 — Figure 4.25 shows the graph of $y = x^2$ and four shifts. For each graph, write the equation of the function that is shown.

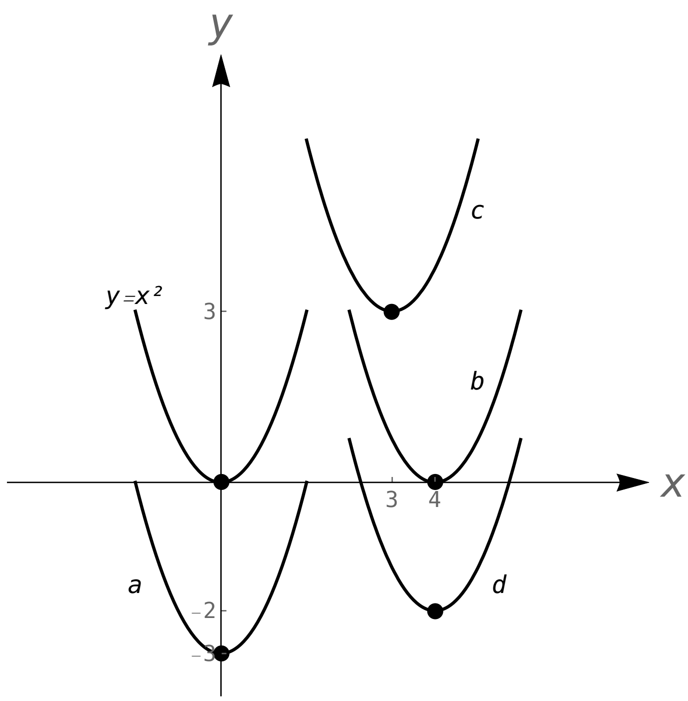


Figure 4.25: The graph of $y = x^2$ and four shifts.

Assignment 4.15 — Identify and sketch the graph of the conic sections with the following standard equations.

(a) $2x^2 + 3y^2 - 2 = 0$

(b) $2y^2 + 3x = 0$

(c) $x^2 - 2y^2 + 3 = 0$

(d) $3x^2 - 4y = 0$

(e) $-2x^2 - 4y^2 + 3 = 0$

(f) $2x^2 + 3y = 0$

(g) $y^2 - 3x = 0$

(h) $x^2 - y^2 - 3 = 0$

(i) $-2x^2 + 3y^2 + 3 = 0$

Assignment 4.16 — Determine the equation of the given conic sections.

(a) Ellipse with focal points in $(0,\pm 2)$ and length of half major axis equal to 3.

(b) Ellipse with focal points in $(0,1)$ en $(4,1)$ and with a distance from the center to the top that is twice as great as the distance from the center to the focal point.

(c) Ellipse with center in the origin, through $(3,1)$ and with vertical minor axis with length 4.

(d) Parabola with focal point in $(2,3)$ and top in $(2,4)$.

(e) Parabola through the origin, with focal point in $(0,-1)$ and upper tangent $y=0$.

(f) Parabola with focal point in $(-2,0)$ and directrix $x=2$.

(g) Hyperbola with center in origin and through $(1,5)$ and $(2,7)$.

(h) Hyperbola with focal points in $(0,\pm 2)$ and tops in $(0,\pm 1)$.

(i) Hyperbola with focal points in $(\pm 5,1)$ and asymptotes $x=\pm(y-1)$.

Assignment 4.17 — Identify and sketch the graph of the following conic sections.

(a) $x^2+y^2+2x-3y=0$

(b) $3x^2+3y^2+2x+7y=3$

(c) $3x^2-2y^2+3x+4y=0$

(d) $2x^2+3y^2-4x-12y+10=0$

(e) $2y^2+3x-4y+2=0$

(f) $2x^2-x-y+7=0$

(g) $3x^2+7y^2-14y+5=0$

(h) $2x^2-y^2-4x+3=0$

(i) $4x^2-y^2-4y=0$

(j) $9x^2+4y^2-18x+8y=23$

Assignment 4.18 — A hyperbolic mirror is used in some telescopes. Such a mirror has the property that an incoming light beam directed to one focal point will be reflected back to the other focal point. Use Figure 4.26 to construct the equation that models the hyperbolic mirror.

Assignment 4.19 — Long-range navigation (LORAN)is a radio navigation system developed during World War II. This system allows an aircraft to be controlled by maintaining a constant difference between the distance of the aircraft from two fixed points: a master station and a slave station. Determine an equation for the hyperbola in Figure 4.27 that describes this fixed difference.

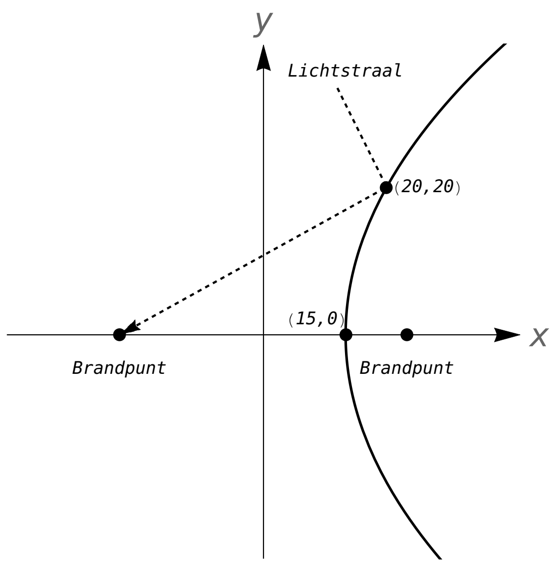


Figure 4.26: Reflection of an incoming light beam on a hyperbolic mirror.

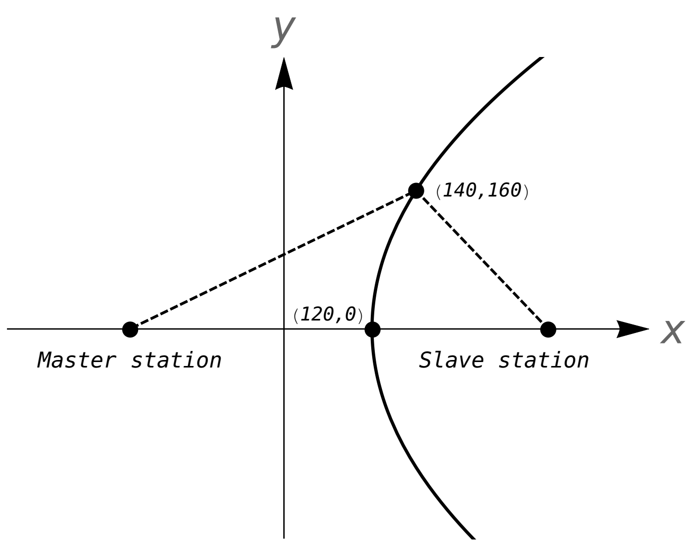


Figure 4.27: LORAN navigation

Review Exercises

Assignment 4.20 — Solve the inequalities below algebraically and graphically.

(a) $1 \leq (2x-3)^2 \leq 4$

(b) $3x^2+2x-8>0$

(c) $\frac{3}{4}x^2 > 4(x-2)$

(d) $3(x-1)^2 > 4(x-1)$

(e) $5x+4 \leq 3x^2$

(f) $2 \leq |x^2-9| < 9$

(g) $x^2 \leq |4x-3|$

(h) $x|x+5| \geq -6$

(i) $x|x-3| < 2$

(j) $\dfrac{x^2}{x|x|+1} < \dfrac{1}{2}x$

Mathematics is a game played according to certain simple rules with meaningless marks on paper.

— David Hilbert —

5

Transcendental functions

5.1 Definition

A transcendental function is a function that does not satisfy a polynomial equation, in contrast to an algebraic function. In other words, a **transcendental function** (*transcendente functie*) transcends algebra in that it cannot be expressed in terms of a finite sequence of the algebraic operations of addition, multiplication, and root extraction. A function that is not transcendental is algebraic.

The most familiar transcendental functions are the logarithmic, exponential (with any non-trivial base), trigonometric, and hyperbolic functions, and the inverses of all of these. Less familiar are the special functions, such as the gamma, elliptic, and zeta functions. Besides, the generalized hypergeometric and Bessel functions are transcendental in general, but algebraic for some special parameter values.

5.2 Exponential and logarithmic functions

5.2.1 Definitions

5.2.1.1 Exponential functions

Up to this point, we have dealt with functions that involve terms of the form x^p where the base of the term, x, varies but the exponent of each term, p, remains constant. Here, we study functions of the form $f(x) = b^x$ where the base b is a constant and the exponent x is the variable. We start our exploration of these functions with $f(x) = 2^x$, whose graph is shown in Figure 5.1(a).

A few remarks about the graph of $f(x) = 2^x$ are in order. As $x \to -\infty$, the function $f(x) = 2^x$ takes on values that are increasingly closer to 0. In other words, as $x \to -\infty$, $f(x) \to 0^+$ and the x-axis is a

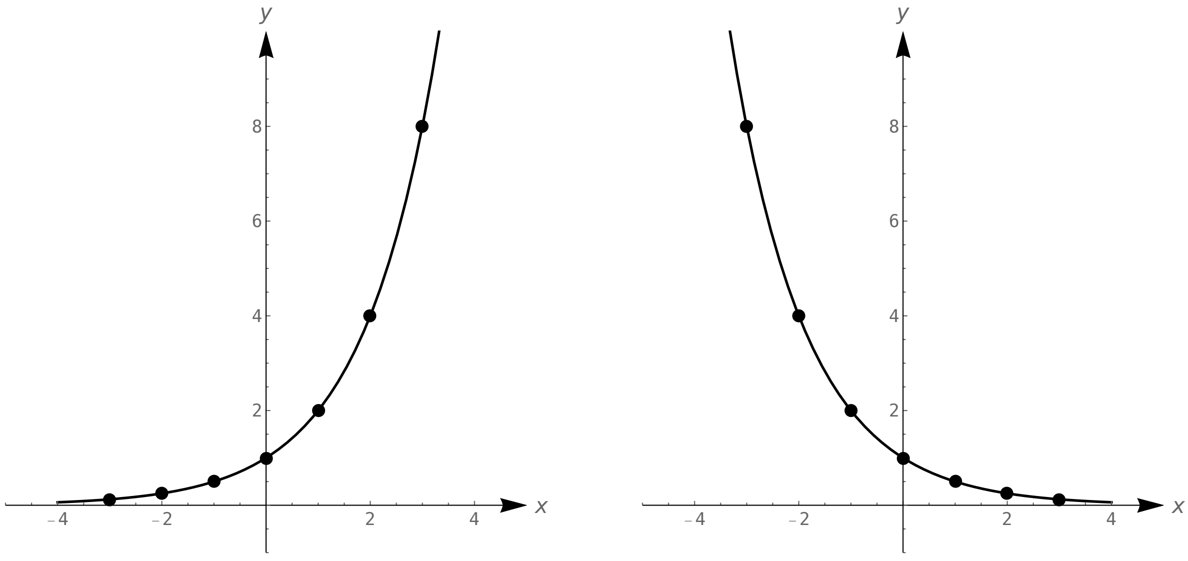


Figure 5.1: The graph of $y = f(x) = 2^x$ (a) and $y = g(x) = \left(\frac{1}{2}\right)^x = 2^{-x}$ (b).

horizontal asymptote. On the flip side, as $x \to +\infty$, we find $f(x) \to +\infty$. As a result, our graph suggests the range of f is $\mathbb{R}_0^+$. Besides, it is clear that f is injective and hence invertible, while $\operatorname{dom} f = \mathbb{R}$.

Here, we wish to study the family of functions $f(x) = b^x$, but which bases b make sense to study? We find that we run into difficulty if $b < 0$. For example, if $b = -2$, then the function $f(x) = (-2)^x$ has trouble, because, for instance, at $x = \frac{1}{2}, f(x) = \sqrt{-2}$ is not a real number. So we must restrict our attention to bases $b \geq 0$. What about $b = 0$? The function $f(x) = 0^x$ is undefined for $x \leq 0$ because we cannot divide by 0 and 0^0 is an indeterminate form. For $x > 0$, $0^x = 0$ so the function $f(x) = 0^x$ is the same as the function $f(x) = 0$ for $x > 0$. We know everything we can possibly know about this function, so we exclude it from our investigations. The only other base we exclude is $b = 1$, since the function $f(x) = 1^x = 1$ is, once again, a function we have already studied (see Chapter 4). Bearing this in mind, we are now ready to give a more formal definition of exponential functions.

Definitie 5.1 (Exponential function)
A function of the form

$$f(x) = b^x$$

where b is a strictly positive fixed real number ($b > 0$) and $b \neq 1$ is called a **base** b **exponential function** (*exponentiële functie met grondtal b*). Moreover, such a function is called exponentially increasing if $b > 1$ and exponentially decreasing if $0 < b < 1$.

Now, we could wonder what the graph of an exponential function with $0 < b < 1$ looks like. For instance, consider $g(x) = \left(\frac{1}{2}\right)^x$. Naively, we could certainly build a table of values and connect the points, but more wisely we could take a step back and note that $g(x) = \left(\frac{1}{2}\right)^x = \left(2^{-1}\right)^x = 2^{-x} = f(-x)$, where $f(x) = 2^x$. Thinking back to Section 3.2.5, the graph of $f(-x)$ is obtained from the graph of $f(x)$ by reflecting it across the y-axis (Figure 5.1(b)). We see that the domain and range of g match that of f, namely $\mathbb{R}$ and $\mathbb{R}_0^+$, respectively. Like f, g is also injective. Whereas f is always increasing, g is always decreasing. As a result, as $x \to -\infty$, $g(x) \to +\infty$, and on the flip side, as $x \to +\infty$, $g(x) \to 0^+$.

In literature, one very often comes across the wording **exponential growth** (*exponentiële groei*), but what exactly does it mean? Let us contrast exponential growth with linear growth in the following table.

x	$f(x) = 2^x$	$h(x) = 2x$
0	1	0
1	2	2
2	4	4
3	8	6
4	16	8
5	32	10

From this table we can infer that for these two functions, exponential growth dwarfs linear growth. More specifically, the former implies that original value from the range increases by the same percentage over equal increments found in the domain, whereas the latter refers to the original value from the range that increases by the same amount over equal increments found in the domain. For exponential growth, over equal increments, the constant multiplicative rate of change resulted in doubling the output whenever the input increased by one. For linear growth, the constant additive rate of change over equal increments resulted in adding 2 to the output whenever the input was increased by one.

Of all of the bases for exponential functions, two occur the most often in scientific circles. The first, base 10, is often called the **common base** (*tiendelige basis*). The second base is an irrational number, $e \approx 2.718$, called the **natural base**. The following examples give us an idea how these functions are used in the wild.

Example 5.1

The value of a tractor can be modelled by $V(x) = 25\left(\frac{4}{5}\right)^x$, where $x \geq 0$ is age of the vehicle in years and $V(x)$ is the value in thousands of euros.

1. Find and interpret $V(0)$.
2. Sketch the graph of $y = V(x)$ using transformations.
3. Find and interpret the horizontal asymptote of the graph of $y = V(x)$.

Solution

1. To find $V(0)$, we replace x with 0 to obtain $V(0) = 25\left(\frac{4}{5}\right)^0 = 25$. Since x represents the age of the tractor in years, $x = 0$ corresponds to the tractor being brand new. Since $V(x)$ is measured in thousands of euros, $V(0) = 25$ corresponds to a value of €25,000. Putting it all together, we interpret $V(0) = 25$ to mean the purchase price of the tractor was €25 000.
2. To graph $y = 25\left(\frac{4}{5}\right)^x$, we start with the basic exponential function $f(x) = \left(\frac{4}{5}\right)^x$. Since the base $b = 4/5$ is between 0 and 1, the graph of $y = f(x)$ is decreasing. We plot the y-intercept $(0, 1)$ and two other points, $(-1, 5/4)$ and $(1, 4/5)$, and notice the horizontal asymptote $y = 0$ (Figure 5.2(a)). To obtain $V(x) = 25\left(\frac{4}{5}\right)^x$, we multiply the output from f by 25, which results in a vertical stretch by a factor of 25. We multiply all of the y-values in the graph by 25 and obtain the points $(-1, 125/4)$, $(0, 25)$ and $(1, 20)$. The horizontal asymptote remains 0. Finally, we restrict the domain to $\mathbb{R}^+$ to fit with the applied domain given to us (Figure 5.2(b)).

3. We see from the graph of V that its horizontal asymptote is $y = 0$. This means as the tractor gets older, its value diminishes to 0.

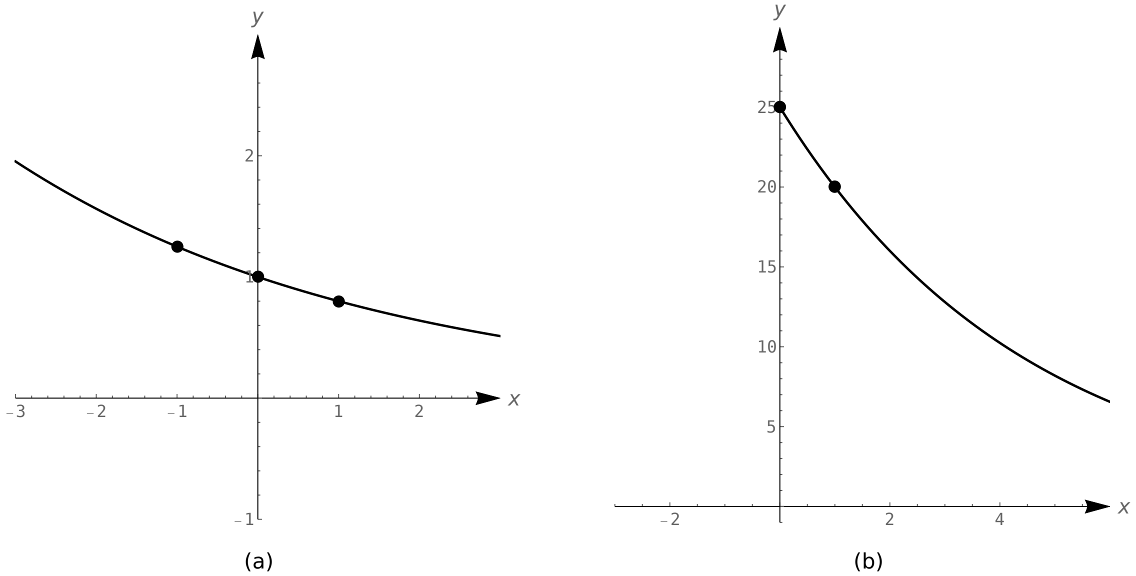


Figure 5.2: The graph of $y = f(x) = \left(\frac{4}{5}\right)^x$ (a) and $y = V(x) = 25\left(\frac{4}{5}\right)^x$ (b) in Example 5.1.

The function in the previous example is often called a **decay curve**. In contrast, increasing exponential functions are used to model growth curves and we shall see several different examples of those later. We present another common decay curve in the following example. Although it may look more complicated than the previous example, it is actually just a basic exponential function which has been modified by a few transformations from Section 3.2.5.

Example 5.2

According to Newton's Law of cooling the temperature of coffee T $[\Theta]$ in degrees Celsius t [T] minutes after it is served can be modelled by

$$T(t) = 21 + 50e^{-0.1t}.$$

1. Find and interpret $T(0)$.
2. Sketch the graph of $y = T(t)$ using transformations.
3. Find and interpret the horizontal asymptote of the graph.

Solution

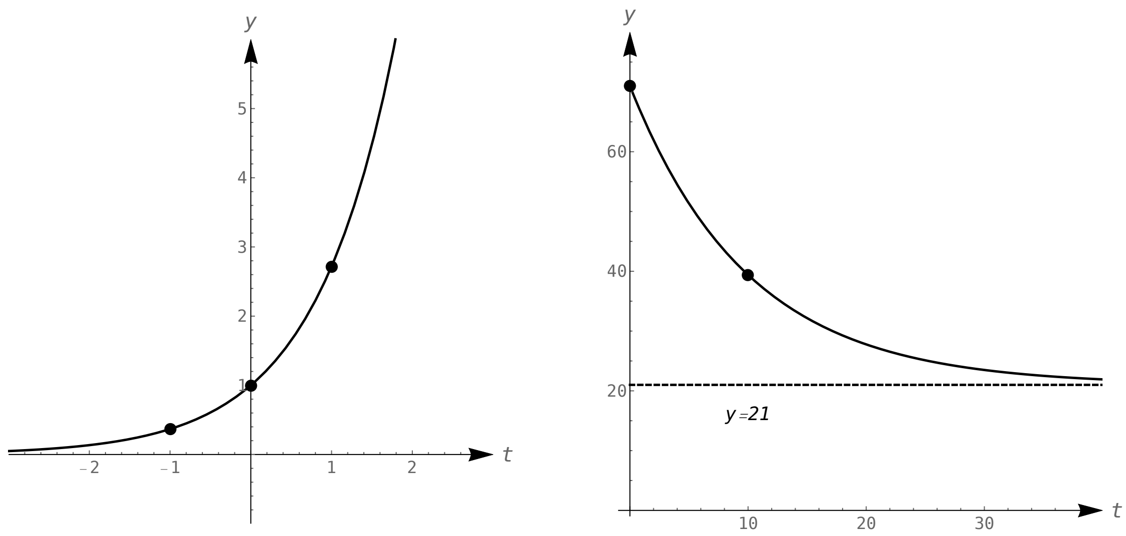
1. To find $T(0)$, we replace every occurrence of the independent variable t with 0 to obtain $T(0) = 21 + 50e^{-0.1(0)} = 71$. This means that the coffee was served at 71°C.
2. To graph $y = T(t)$ using transformations, we start with the basic function, $f(t) = e^t$. Since $e \approx 2.718 > 1$, the graph of f is an increasing exponential with y-intercept $(0, 1)$ and horizontal asymptote $y = 0$. The points $\left(-1, e^{-1}\right) \approx (-1, 0.37)$ and $(1, e) \approx (1, 2.72)$ are also

on the graph (Figure 5.3(a)). To use this information on $f(t)=e^t$, we rewrite $T(t)$ as

$$T(t)=21+50f(-0.1t)\,.$$

Multiplication of the input to f, t, by -0.1 results in a horizontal expansion by a factor of 10 as well as a reflection about the y-axis. We divide each of the x-values of our points by -0.1 to obtain $\left(10,e^{-1}\right)$, $(0,1)$, and $(-10,e)$. Since none of these changes affected the y-values, the horizontal asymptote remains $y=0$. Next, we see that the output from f is being multiplied by 50. This results in a vertical stretch by a factor of 50. We multiply the y-coordinates by 50 to obtain $\left(10,50e^{-1}\right)$, $(0,50)$, and $(-10,50e)$. Obviously, the horizontal asymptote remains $y=0$. Finally, we add 21 to all of the y-coordinates, which shifts the graph upwards to obtain $\left(10,50e^{-1}+21\right)\approx(10,39.39)$, $(0,71)$, and $(-10,50e+21)\approx(-10,156.91)$. Adding 21 to the horizontal asymptote shifts it upwards as well to $y=21$. We connect these three points and, last but not least, we restrict the domain to match the applied domain $\mathbb{R}^+$ (Figure 5.3(b)).

3. From the graph, we see that the horizontal asymptote is $y=21$. As $t\to+\infty$, the term $50e^{-0.1t}$ becomes very small. Hence, the graph of T is approaching the horizontal line $y=21$ from above. This means that as time goes by, the temperature of the coffee is cooling to 21°C, presumably room temperature.



(a) (b)

Figure 5.3: The graph of $y=f(t)=e^t$ (a) and $y=T(t)=21+50e^{-0.1t}$ (b).

5.2.1.2 Logarithmic functions

As we have already remarked, the function $f(x)=b^x$ is injective, and hence invertible. We now turn our attention to these inverses, the logarithmic functions.

Definitie 5.2 (Logarithmic function)
The inverse of the exponential function $f(x)=b^x$ is called the **base** b **logarithm function** (*logaritmische functie met grondtal b*), and is denoted

$$f^{-1}(x)=\log_b(x)\,.$$

We read $\log_b(x)$ as log base b of x.

The **common logarithm** (*tiendelige logaritme, Briggse logaritme)* of a real number x is $\log_{10}(x)$ and is usually written $\log(x)$. The **natural logarithm** (*natuurlijke logaritme, Neperiaanse logaritme* of a real number x is $\log_e(x)$ and is usually written $\ln(x)$.

Since logarithmic functions are defined as the inverses of exponential functions, we can use the findings of Section 3.4 to tell us something about logarithmic functions. For example, we know that the domain of a logarithmic function is the range of an exponential function, namely $\mathbb{R}_0^+$, and that the range of a logarithmic function is the domain of an exponential function, namely $\mathbb{R}$. Since we know the basic shapes of $y = f(x) = b^x$ for the different cases of b, we can obtain the graph of $y = f^{-1}(x) = \log_b(x)$ by reflecting the graph of f across the line $y = x$ as shown below. The y-intercept $(0,1)$ on the graph of f corresponds to an x-intercept of $(1,0)$ on the graph of f^{-1}. The horizontal asymptotes $y = 0$ on the graphs of the exponential functions become vertical asymptotes $x = 0$ on the graphs of the logarithmic functions. All this is illustrated in Figure 5.4 for the functions $f_1(x) = e^x$, $f_2(x) = 2^x$, $f_3(x) = \left(\frac{1}{e}\right)^x$ and $f_4(x) = \left(\frac{1}{2}\right)^x$ and their corresponding inverses $f_1^{-1}(x) = \ln(x)$, $f_2^{-1}(x) = \log_2(x)$, $f_3^{-1}(x) = \log_{\frac{1}{e}}(x)$ and $f_4^{-1}(x) = \log_{\frac{1}{2}}(x)$, respectively.

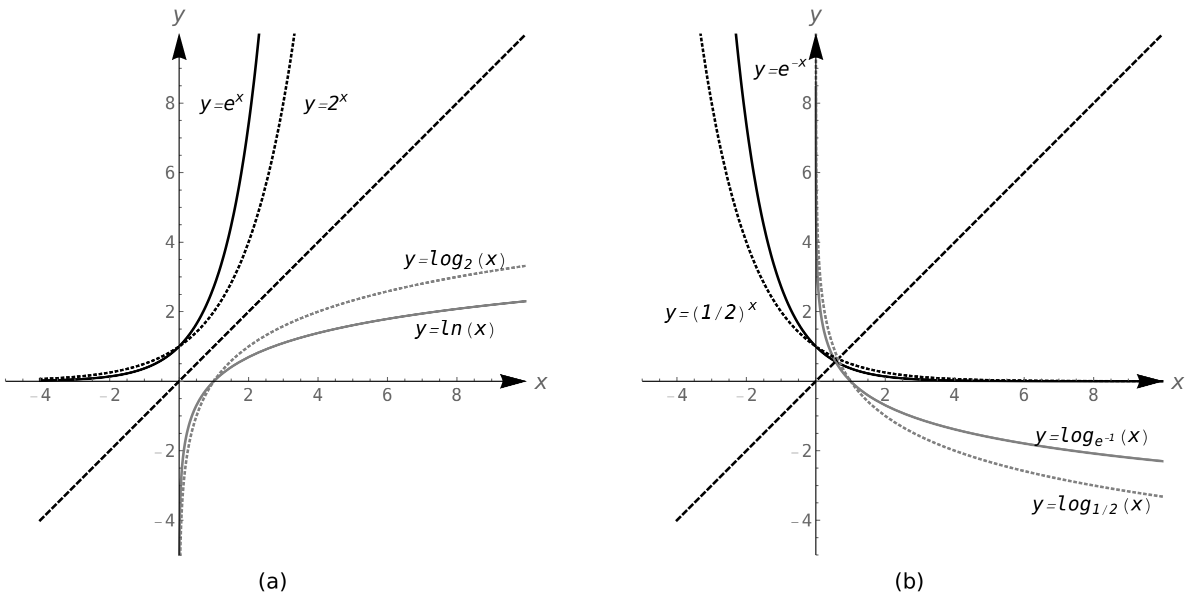


Figure 5.4: The graph of $f_1(x) = e^x$ (a), $f_2(x) = 2^x$ (a), $f_3(x) = \left(\frac{1}{e}\right)^x$ (b) and $f_4(x) = \left(\frac{1}{2}\right)^x$ (b) and their corresponding inverses $f_1^{-1}(x) = \ln(x)$ (a), $f_2^{-1}(x) = \log_2(x)$ (a), $f_3^{-1}(x) = \log_{\frac{1}{e}}(x)$ (b) and $f_4^{-1}(x) = \log_{\frac{1}{2}}(x)$ (b), respectively.

Logarithms and the human psyche

Logarithms occur in several laws describing human perception. For instance, Hick's law proposes a logarithmic relation between the time individuals take to choose an alternative and the number of choices they have, while Fitts's law predicts that the time required to rapidly move to a target area is a logarithmic function of the distance to and the size of the target.

Interestingly, psychological studies found that individuals with little mathematics education tend to estimate quantities logarithmically, that is, they position a number on an unmarked line according to its logarithm, so that 10 is positioned as close to 100 as 100 is to 1000. Increasing education shifts this to a linear estimate that involves positioning 1000 10 times as far away.

Up until this point, restrictions on the domains of functions came from avoiding division by zero and keeping negative numbers from beneath even radicals. With the introduction of logarithmic functions, we now have another restriction. Since the domain of $f(x) = \log_b(x)$ is $\mathbb{R}_0^+$, the argument of the logarithmic function must be strictly positive.

Example 5.3

Find the domain of the following functions. Check your answers graphically using Mathematica.

1. $f(x) = 2\log(3-x)-1$
2. $g(x) = \ln\left(\dfrac{x}{x-1}\right)$

Solution

1. We have to make sure that the argument of the involved logarithm is strictly positive, so we set $3-x>0$ to obtain $]-\infty,3[$. The graph shown in Figure 5.5(a) confirms this.
2. To find the domain of g, we need to solve the inequality $\frac{x}{x-1}>0$. First, we define $r(x)=\frac{x}{x-1}$, and find that r is undefined at $x=1$ and $r(x)=0$ when $x=0$. Choosing some test values, we easily find that $\frac{x}{x-1}>0$ on $]-\infty,0[\,\cup\,]1,+\infty[$ to get the domain of g. The graph of $y=g(x)$ confirms this (Figure 5.5(b)).

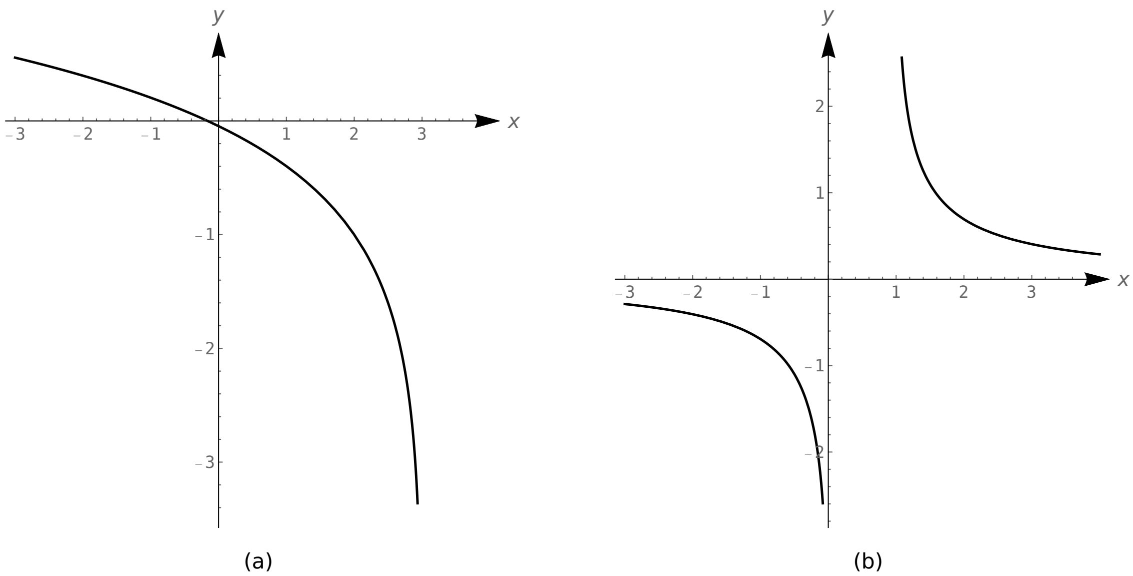


Figure 5.5: The graph of $f(x) = 2\log(3-x)-1$ (a) and $g(x) = \ln\left(\frac{x}{x-1}\right)$ (b).

5.2.2 Properties

As we shall see shortly, exponential functions inherit analogs of all of the properties of exponents you encountered in Chapter 2. First, we look at the consequence of exponential and logarithmic functions to be injective.

Let $f(x) = b^x$ and $g(x) = \log_b(x)$ where $b>0$, $b\neq 1$. Then f and g are injective functions and

- $b^u = b^w$ if and only if $u = w$ for all real numbers u and w.
- $\log_b(u) = \log_b(w)$ if and only if $u = w$ for all real numbers $u>0$, $w>0$.

We now state the algebraic properties of exponential functions which will serve as a basis for the properties of logarithms. While these properties may look identical to the ones you learned in Chapter 2, they apply to real number exponents, not just rational exponents.

Theorem 5.1 (Algebraic properties of exponential functions)
Let $b > 0$, $b \neq 1$ and let u and w be real numbers, then

- ***Product rule***: $b^{u+w} = b^u \, b^w$
- ***Quotient rule***: $b^{u-w} = \dfrac{b^u}{b^w}$
- ***Power rule***: $(b^u)^w = b^{u\,w}$

To each of these properties of exponential functions corresponds an analogous property of logarithmic functions. We list these below in our next theorem.

Theorem 5.2 (Algebraic properties of logarithmic functions)
Let $b > 0$, $b \neq 1$ and let $u > 0$ and $w > 0$ be real numbers.

- ***Product rule***: $\log_b(u\,w) = \log_b(u) + \log_b(w)$
- ***Quotient rule***: $\log_b\left(\dfrac{u}{w}\right) = \log_b(u) - \log_b(w)$
- ***Power rule***: $\log_b(u^w) = w\log_b(u)$

From a purely functional approach, we can see the properties in Theorem 5.2 as an example of how inverse functions interchange the roles of inputs in outputs. For instance, the product rule for exponential functions given in Theorem 5.1, $f(u+w) = f(u)f(w)$, says that adding inputs results in multiplying outputs. Hence, whatever f^{-1} is, it must take the products of outputs from f and return them to the sum of their respective inputs. Since the outputs from f are the inputs to f^{-1} and vice-versa, we have that f^{-1} must take products of its inputs to the sum of their respective outputs. This is precisely what the product rule for logarithmic functions states in Theorem 5.2.

Example 5.4

Use the properties of logarithms to write the following as a single logarithm.

1. $\log_3(x-1) - \log_3(x+1)$
2. $\log(x) + 2\log(y) - \log(z)$

Solution

1. The difference of logarithms requires the quotient rule:

$$\log_3(x-1) - \log_3(x+1) = \log_3\left(\frac{x-1}{x+1}\right).$$

2. We first apply the power rule, and then the product/quotient rule to get the following.

$$\begin{aligned}
\log(x) + 2\log(y) - \log(z) &= \log(x) + \log\left(y^2\right) - \log(z) && \text{(Power rule.)} \\
&= \log\left(xy^2\right) - \log(z) && \text{(Product rule.)} \\
&= \log\left(\frac{xy^2}{z}\right) && \text{(Quotient rule.)}
\end{aligned}$$

We observe that using log properties to reassemble logarithms can increase the domain of the expression. For example, we leave it to the reader to verify the domain of $f(x) = \log_3(x-1) - \log_3(x+1)$ is $]\,1, +\infty\,[$ but the domain of $g(x) = \log_3\left(\frac{x-1}{x+1}\right)$ is $]-\infty, -1\,[\,\cup\,]\,1, +\infty\,[$. We will need to keep this in mind when we solve equations involving logarithms

In many cases it is convenient to change the base of the governing exponential or logarithmic functions. For that purpose, we may rely on the following theorem.

Theorem 5.3 (Change of base formulas)
Let $a, b > 0$, and $a, b \neq 1$. Then, we have

- $a^x = b^{x \log_b(a)}$, *for all real numbers x;*
- $\log_a(x) = \dfrac{\log_b(x)}{\log_b(a)}$, *for all real numbers $x > 0$.*

5.2.3 Exponential and logarithmic equations and inequalities

In this section we will briefly recall techniques for solving equations involving exponential or logarithmic functions. We first summarize below the two common ways to solve exponential equations.

1. Isolate the exponential function.
2. (a) If convenient, express both sides with a common base and equate the exponents.
   (b) Otherwise, take the natural log of both sides of the equation and use the power rule.

Likewise, the steps for solving an equation involving logarithmic functions are:

1. Isolate the logarithmic function.
2. (a) If convenient, express both sides as logs with the same base and equate the arguments of the log functions.
   (b) Otherwise, rewrite the log equation as an exponential equation.

Dual meaning of $\log(x)$

Throughout this text we adopted the notation $\ln(x)$ and $\log(x)$ to refer to the natural algorithm and common logarithm of x, respectively. In literature and on the Internet, however, one often finds that $\log(x)$ is used to refer to the natural logarithm of x, so beware when consulting other sources of information!

Example 5.5

Solve the following exponential and logarithmic equations.

1. $2^{3x} = 16^{1-x}$
2. $9 \cdot 3^x = 7^{2x}$
3. $75 = \dfrac{100}{1+3e^{-2t}}$
4. $\log_{117}(1-3x) = \log_{117}\left(x^2-3\right)$
5. $\log_7(1-2x) = 1-\log_7(3-x)$
6. $1+2\log_4(x+1) = 2\log_2(x)$

Solution

1. Since 16 is a power of 2, we can rewrite the equation as $2^{3x} = \left(2^4\right)^{1-x}$. Using properties of exponents, we get $2^{3x} = 2^{4(1-x)}$. Given the one-to-one property of exponential functions, we get $3x = 4(1-x)$, which gives $x = 4/7$.

2. We first note that we can rewrite the equation as $3^2 \cdot 3^x = 7^{2x}$ to obtain $3^{x+2} = 7^{2x}$. Since it is not convenient to express both sides as a power of 3 (or 7 for that matter) we use the natural log: $\ln\left(3^{x+2}\right) = \ln\left(7^{2x}\right)$. The power rule gives $(x+2)\ln(3) = 2x\ln(7)$. This equation is linear and can be solved for x:

$$\begin{array}{rrcl} & (x+2)\ln(3) & = & 2x\ln(7) \\ \Leftrightarrow & x\ln(3)+2\ln(3) & = & 2x\ln(7) \\ \Leftrightarrow & 2\ln(3) & = & 2x\ln(7)-x\ln(3) \\ \Leftrightarrow & 2\ln(3) & = & x(2\ln(7)-\ln(3)) \\ \Leftrightarrow & x & = & \dfrac{2\ln(3)}{2\ln(7)-\ln(3)} \end{array}$$

3. First, we isolate the exponential:

$$\begin{aligned} 75 = \frac{100}{1+3e^{-2t}} \quad &\Leftrightarrow \quad 75\left(1+3e^{-2t}\right) = 100 \\ &\Leftrightarrow \quad e^{-2t} = \tfrac{1}{9}\,. \end{aligned}$$

Taking the natural log of both sides gives

$$\ln\left(e^{-2t}\right) = \ln\left(\frac{1}{9}\right) \quad \Leftrightarrow \quad t = \ln(3)\,.$$

4. Since we have the same base on both sides of this equation , we equate what is inside the logs to get $1-3x = x^2-3$. Solving $x^2+3x-4 = 0$ gives $x = -4$ and $x = 1$. To check whether none of these is an extraneous solution, we substitute, for instance $x = 1$, into our original equation to obtain $\log_{117}(-2) = \log_{117}(-2)$. While these expressions look identical, neither is a real number, which means $x = 1$ is not in the domain of the original equation, and is not a solution. Similarly, we can verify that $x = -4$ is indeed a solution in the domain of the original equation.

5. We first collect the logarithms on the same side and then use the product rule to get

$$\log_7[(1-2x)(3-x)] = 1 \quad \Leftrightarrow \quad 7^1 = (1-2x)(3-x)\,,$$

which leads to $2x^2-7x-4=0$ whose solution is $x=-1/2$ or $x=4$. However, checking $x=4$ in the original equation produces $\log_7(-7)=1-\log_7(-1)$, which is a clear domain violation.

6. We gather the logs to one side to get $1=2\log_2(x)-2\log_4(x+1)$. Before we can combine the logarithms, however, we need a common base. Since 4 is a power of 2, we change the base
$$\log_4(x+1)=\frac{\log_2(x+1)}{\log_2(4)}=\frac{1}{2}\log_2(x+1).$$
Hence, our original equation becomes
$$\begin{array}{rrcll} & 1 & = & 2\log_2(x)-2\left(\frac{1}{2}\log_2(x+1)\right) & \\ \Leftrightarrow & 1 & = & 2\log_2(x)-\log_2(x+1) & \\ \Leftrightarrow & 1 & = & \log_2\left(x^2\right)-\log_2(x+1) & \text{(Power rule.)} \\ \Leftrightarrow & 1 & = & \log_2\left(\dfrac{x^2}{x+1}\right) & \text{(Quotient rule.)} \end{array}$$
Rewriting this in exponential form, we get $\frac{x^2}{x+1}=2$ or $x^2-2x-2=0$. Using the quadratic formula, we get $x=1\pm\sqrt{3}$. Yet, for what concerns the solution $x=1-\sqrt{3}$, it holds that is negative so if substituted into the original equation, the term $2\log_2\left(1-\sqrt{3}\right)$ is undefined, and hence we should discard it as solution.

This example demonstrates the importance of checking for extraneous solutions when solving equations involving logarithms. These are easy to spot - any supposed solution which causes a negative number inside a logarithm needs to be discarded.

Just as we encountered for algebraic functions, we can also run into inequalities involving exponential or logarithmic functions.

Example 5.6

Solve the following inequalities.

1. $\dfrac{e^x}{e^x-4}\le 3$
2. $2^{x^2-3x}-16\ge 0$
3. $\dfrac{1}{\ln(x)+1}\le 1$
4. $x\log(x+1)\ge x$

Solution

1. The first step we need to take to solve $\frac{e^x}{e^x-4}\le 3$ is to get 0 on one side of the inequality. To that end, we subtract 3 from both sides and get .
$$\begin{array}{rcl} & \dfrac{e^x}{e^x-4}-3 & \le & 0 \\ \Leftrightarrow & \dfrac{e^x}{e^x-4}-\dfrac{3\left(e^x-4\right)}{e^x-4} & \le & 0 \\ \Leftrightarrow & \dfrac{12-2e^x}{e^x-4} & \le & 0 \end{array}$$

We set $r(x) = \frac{12-2e^x}{e^x-4}$ and we note that r is undefined when its denominator $e^x - 4 = 0$, or when $e^x = 4$. Solving this gives $x = \ln(4)$, so the domain of r is $\mathbb{R} \setminus \{\ln(4)\}$. To find the zeros of r, we solve $r(x) = 0$ and obtain $12 - 2e^x = 0$. Solving for e^x, we find $e^x = 6$, or $x = \ln(6)$. When we build our sign diagram, finding test values may be a little tricky since we need to check values around $\ln(4)$ and $\ln(6)$. Recall that the function $\ln(x)$ is increasing which means $\ln(3) < \ln(4) < \ln(5) < \ln(6) < \ln(7)$. While the prospect of determining the sign of $r\big(\ln(3)\big)$ may be very unsettling, remember that $e^{\ln(3)} = 3$, so

$$r\big(\ln(3)\big) = \frac{12-2e^{\ln(3)}}{e^{\ln(3)}-4} = \frac{12-2(3)}{3-4} = -6$$

We determine the signs of $r\big(\ln(5)\big)$ and $r\big(\ln(7)\big)$ similarly and obtain the following sign diagram.

$\dfrac{12-2e^x}{e^x-4}$	$-$	$\ln(4)$ (○)	$+$	$\ln(6)$ (●)	$-$

From this sign diagram, we find our answer to be $]-\infty, \ln(4)[\, \cup [\ln(6), +\infty[$.

2. Since we already have 0 on one side of the inequality, we set $r(x) = 2^{x^2-3x} - 16$. The domain of r is all real numbers, so in order to construct our sign diagram, we need to find the zeros of r. Setting $r(x) = 0$ gives

$$2^{x^2-3x} = 16 \quad \Leftrightarrow \quad 2^{x^2-3x} = 2^4\,.$$

So, $x^2 - 3x = 4$ and the solutions of this equation are $x = 4$ and $x = -1$. From the sign diagram,

x^2-3x-4	$+$	-1 (●)	$-$	4 (●)	$+$

we see $r(x) \geq 0$ on $]-\infty, -1] \cup [4, +\infty[$ because 2^x is increasing everywhere and $x^2 - 3x + 4$ is positive there.

3. We start solving this inequality by getting 0 on one side. Getting a common denominator yields

$$\frac{1}{\ln(x)+1} - \frac{\ln(x)+1}{\ln(x)+1} \leq 0 \quad \Leftrightarrow \quad \frac{\ln(x)}{\ln(x)+1} \geq 0.$$

We define $r(x) = \frac{\ln(x)}{\ln(x)+1}$ and set about finding the domain and the zeros of r. Due to the appearance of the term $\ln(x)$, we require $x > 0$. In order to keep the denominator away from zero, we solve $\ln(x) + 1 = 0$ so $\ln(x) = -1$, so $x = e^{-1}$. Hence, the domain of r is $]\,0, e^{-1}\,[\, \cup \,]\, e^{-1}, +\infty\,[$. To find the zeros of r, we set $r(x) = 0$ so that $\ln(x) = 0$, and we find $x = e^0 = 1$. In order to determine test values for r, we need to find numbers between 0, e^{-1}, and 1 which have a base of e. Since $e \approx 2.718 > 1$, $0 < e^{-2} < e^{-1} < e^{-1/2} < 1 < e$. To determine the sign of $r\left(e^{-2}\right)$, we use the fact that $\ln\left(e^{-2}\right) = -2$, and find $r\left(e^{-2}\right) = \frac{-2}{-2+1} = 2$, which is positive. The rest of the test values are determined similarly.

$\dfrac{\ln(x)}{\ln(x)+1}$	$+$	e^{-1} (○)	$-$	1 (●)	$+$

From our sign diagram, we find the solution to be $]0, e^{-1}[\, \cup [1, +\infty[$.

4. We begin by subtracting x from both sides to get $x\log(x+1) - x \geq 0$. We define $r(x) = x\log(x+1) - x$ and due to the presence of the logarithm, we require $x + 1 > 0$, or $x > -1$. To find the zeros of r, we solve $r(x) = 0$ for x:

$$x\log(x+1) - x = 0 \quad \Leftrightarrow \quad x\big(\log(x+1) - 1\big) = 0\,.$$

This gives $x = 0$ or $\log(x+1) - 1 = 0$, where the latter means that $x = 9$. We select test values x so that $x + 1$ is a power of 10, and we obtain $-1 < -0.9 < 0 < \sqrt{10} - 1 < 9 < 99$:

$x\log(x+1)-x$	$+$	0	$-$	9	$+$

Our sign diagram gives the solution to be $]-1, 0] \cup [9, +\infty[$.

5.2.4 Applications

Exponential and logarithmic functions are used to model a wide variety of behaviours in the real world.

5.2.4.1 Growth models

The law of uninhibited - Malthusian - growth states as its premise that the instantaneous rate at which a population increases at any time is directly proportional to the population at that time. In other words, the more organisms there are at a given moment, the faster they reproduce. Formulating the law as stated results in a so-called differential equation, which is the logic of Mathematics III. Anyhow, solving this differential equation leads to the following model equation for the number of organisms N [-] at time t [T]:

$$N(t) = N_0 e^{kt}, \tag{5.1}$$

where $N(0) = N_0$ [–] is the initial number of organisms and $k > 0$ is the constant of proportionality, and represents the Malthusian growth rate [T^{-1}].

Example 5.7

In order to perform atherosclerosis research, epithelial cells are harvested from discarded umbilical tissue and grown in the laboratory. A technician observes that a culture of 12 000 cells grows to 5 000 000 cells in one week. Assuming that the cells follow the law of uninhibited growth, find a formula for the number of cells, N, in thousands, after t days.

Solution

Since N is to give the number of cells in thousands, we have $N_0 = 12$, so $N(t) = 12e^{kt}$. In order to complete the formula, we need to determine the growth rate k. We know that after one week, the number of cells has grown to five million. Since t measures days and the units of N are in thousands, this translates mathematically to $N(7) = 5000$. We get the equation $12e^{7k} = 5000$ which gives $k = \frac{1}{7}\ln\left(\frac{1250}{3}\right)$. Hence, we get

$$N(t) = 12e^{\frac{t}{7}\ln\left(\frac{1250}{3}\right)}.$$

Of course, in practice, we would approximate k to some desired accuracy, say $k \approx 0.8618$, which we can interpret as an 86.18% daily growth rate for the cells.

Obviously, the law of uninhibited growth will in most practical settings not hold because the availability of resources in the environment that are needed for organisms to grow and persist is limited. This effect can, however, be incorporated in a logistic - Verhulst - growth model, which incorporates that the rate of growth of a population varies jointly with the population itself as well as the room the population has to grow. More specifically, if a population behaves according to the assumptions of **logistic growth** (*logistische groei*), the number of organisms N [-] at time t [T] is given by

$$N(t) = \frac{L}{1 + Ce^{-kLt}}, \tag{5.2}$$

where $N(0) = N_0$ is the initial population, L [–] is the limiting population, C [–] is a measure of how much room there is to grow given by

$$C = \frac{L}{N_0} - 1.$$

and $k > 0$ is the constant of proportionality [T^{-1}].

The logistic function is used not only to model the growth of organisms, but is also to model the spread of disease and rumours.

Example 5.8

The number of people N, in hundreds, at a local community college who have heard the rumour 'Carl is afraid of Virginia Woolf' can be modelled using the logistic equation

$$N(t) = \frac{84}{1 + 2799e^{-t}},$$

where $t \geq 0$ is the number of days after April 1, 2009.

1. Find and interpret $N(0)$.
2. Find and interpret the end behaviour of $N(t)$.
3. How long until 4200 people have heard the rumour?

Solution

1. We find $N(0) = \frac{84}{1+2799e^0} = \frac{84}{2800} = \frac{3}{100}$. Since $N(t)$ measures the number of people who have heard the rumour in hundreds, $N(0)$ corresponds to 3 people. Since $t = 0$ corresponds to April 1, 2009, we may conclude that on that day, 3 people have heard the rumour.
2. We could simply note that $N(t)$ is written in the form of Equation (5.2), and identify $L = 84$. However, to see why the answer is 84, we proceed analytically. Since the domain of N is restricted to $t \geq 0$, the only end behaviour of significance is $t \to +\infty$. As we have seen before, as $t \to +\infty$, we have $2799e^{-t} \to 0^+$ and so $N(t) \approx 84$. Hence, as $t \to +\infty$, $N(t) \to 84$. This means that as time goes by, the number of people who will have heard the rumour approaches 8400.
3. To find how long it takes until 4200 people have heard the rumour, we set $N(t) = 42$. Solving $\frac{84}{1+2799e^{-t}} = 42$ gives $t = \ln(2799) \approx 7.937$. Consequently, it takes around 8 days until 4200 people have heard the rumour.

5.2.4.2 Newton's law of cooling

We first encountered Newton's Law of cooling in Example 5.2. In that example we had a cup of coffee cooling from 71°C to room temperature 21°C according to the formula $T(t) = 21 + 50e^{-0.1t}$, where t was measured in minutes. In this situation, we know the physical limit of the temperature of the coffee is room temperature and the differential equation which gives rise to our formula for $T(t)$ takes this into account. Newton's Law of Cooling states that the rate of cooling of the coffee at a given time t is directly proportional to how much of a temperature gap exists between the coffee at time t and room temperature, not the temperature of the coffee itself. Of course, if we take an item from the refrigerator and let it sit out in the kitchen, the object's temperature will rise to room temperature, and since the physics behind warming and cooling is the same, we combine both cases in the equation below.

The temperature T [Θ] of an object at time t [T] is given by the formula

$$T(t) = T_a + (T_0 - T_a)\, e^{-kt}, \tag{5.3}$$

where $T(0) = T_0$ [T^{-1}] is the initial temperature of the object, T_a [Θ] is the ambient temperature and $k > 0$ is the constant of proportionality.

Example 5.9

Recall from Example 5.2 that the temperature of coffee T (in degrees Celcius) t minutes after it is served can be modelled by $T(t) = 21 + 50e^{-0.1t}$. How long will the coffee be warmer than 50°C?

Solution

We need to find when $T(t) > 50$:

$$21 \;+\; 50e^{-0.1t} > 50 \quad \Leftrightarrow \quad 50e^{-0.1t} - 29 > 0\,.$$

Hence we set $r(t) = 50e^{-0.1t} - 29$. The domain of r is artificially restricted due to the context of the problem to $\mathbb{R}^+$. Solving r(t)=0 results in $e^{-0.1t} = \frac{29}{50}$ so that $t = -10\ln\left(\frac{29}{50}\right)$, or equivalently $t = 10\ln\left(\frac{50}{29}\right)$. Now, we construct the sign diagram:

$$\begin{array}{lc|c} 50e^{-0.1t} - 29 & + & - \\ \hline & \multicolumn{2}{c}{10\ln\left(\frac{50}{29}\right)} \end{array}$$

This means it takes approximately 5.5 minutes for the coffee to cool to 50°C. Until then, the coffee is warmer than that.

5.2.4.3 Radioactive decay

Another real-world phenomenon that can be described by means of an exponential function is radioactive decay. There are elements which are unstable and emit energy spontaneously. In doing so, the amount of the element itself diminishes. The assumption behind the underlying model is that the rate of decay of an element at a particular time is directly proportional to the amount of the element present at that time. In other words, the more of the element there is, the faster the element decays. This is precisely the same kind of hypothesis which drives the law of uninhibited growth, and as such, the equation governing radioactive decay is similar to Equation (5.1) with the exception that the rate constant k is negative.

More precisely, the amount of a radioactive element A [M] at time t [T] is given by the formula

$$A(t) = A_0 e^{kt}, \tag{5.4}$$

where $A(0) = A_0$ is the initial amount of the element and $k < 0$ is the constant of proportionality, also referred to as the decay rate $[T^{-1}]$. In this context, one often uses the so-called **half-life** (*halfwaarde-tijd*), which is the time required for the amount of a radioactive element to reduce to half of its initial amount.

Example 5.10

Iodine-131 is a commonly used radioactive isotope used to help detect how well the thyroid is functioning. Suppose the decay of Iodine-131 follows the model given in Equation (5.4), and that the half-life of Iodine-131 is approximately 8 days. If 5 grams of Iodine-131 is present initially, find a function which gives the amount of Iodine-131, A, in grams, t days later.

Solution

Since we start with 5 grams initially, Equation (5.4) gives $A(t) = 5e^{kt}$. Since the half-life is 8 days, it takes 8 days for half of the Iodine-131 to decay, leaving half of it behind. Hence, $A(8) = 2.5$ which means $5e^{8k} = 2.5$. Solving, we get $k = \frac{1}{8}\ln(1/2) = -\ln(2)/8 \approx -0.08664$, which we can interpret as a loss of material at a rate of 8.664% daily. Hence, $A(t) = 5e^{-\frac{t\ln(2)}{8}} \approx 5e^{-0.08664t}$.

5.3 Trigonometric functions

5.3.1 Foundations of trigonometry

When two half-lines share a common initial point they form an **angle** (*hoek*) and the common initial point is called the **vertex** (*hoekpunt*) of the angle (Figure 5.6).

One commonly used system to measure angles is **degree measure** (*graden*). Quantities measured in degrees are denoted by the familiar '°' symbol. One complete revolution is 360°, and parts of a revolution are measured proportionately. Thus half of a revolution (a **straight angle**) (*gestrekte hoek*) measures $\frac{1}{2}(360°) = 180°$, a quarter of a revolution (a **right angle**) (*rechte hoek*) measures $\frac{1}{4}(360°) = 90°$ and so on. Recall that if an angle measures strictly between 0° and 90° it is called an **acute angle** (*scherpe hoek*) (Figure 5.6(a)) and if it measures strictly between 90° and 180° it is called an **obtuse angle** (*stompe hoek*) (Figure 5.6(b)). In practice, the distinction between the angle itself and its measure is blurred so that the sentence α is an angle measuring 42° is often abbreviated as $\alpha = 42°$.

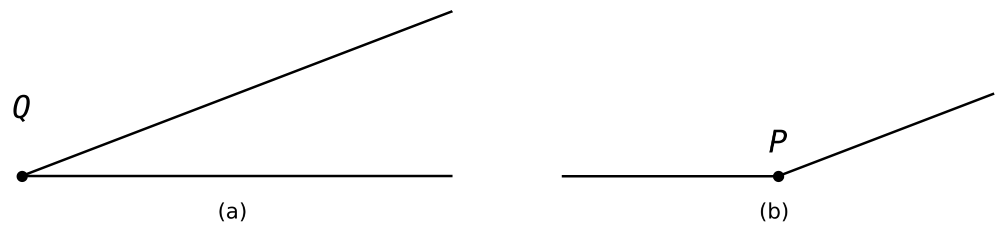


Figure 5.6: An acute angle with vertex Q (a) and obtuse angle with vertrex P (b).

Using our definition of the degree measure, we have that 1° represents the measure of an angle which constitutes $\frac{1}{360}$ of a revolution. Now, there are two ways to further subdivide degrees. The first, and most familiar, is **decimal degrees** (*decimale graden*). The second way to divide degrees is the **Degree - Minute - Second** (**DMS**) system. In this system, one degree is divided equally into sixty minutes, and in turn, each minute is divided equally into sixty seconds. Recall that two acute angles are called **complementary angles** (*complementaire hoeken*) if their measures add to 90°. Two angles, either a pair of right angles or one acute angle and one obtuse angle, are called **supplementary**

angles (*supplementaire hoeken*) if their measures add to 180°. In Figure 5.7, the angles α and β are supplementary angles while the pair γ and δ are complementary angles.

When the direction of rotation matters, we will typically use an **oriented angle** (gerichte hoek). We imagine the angle being swept out starting from an **initial side** and ending at a **terminal side**. When the rotation is counter-clockwise, we say that the angle is **positive** (*positief*); when the rotation is clockwise, we say that the angle is **negative** (*negatief*).

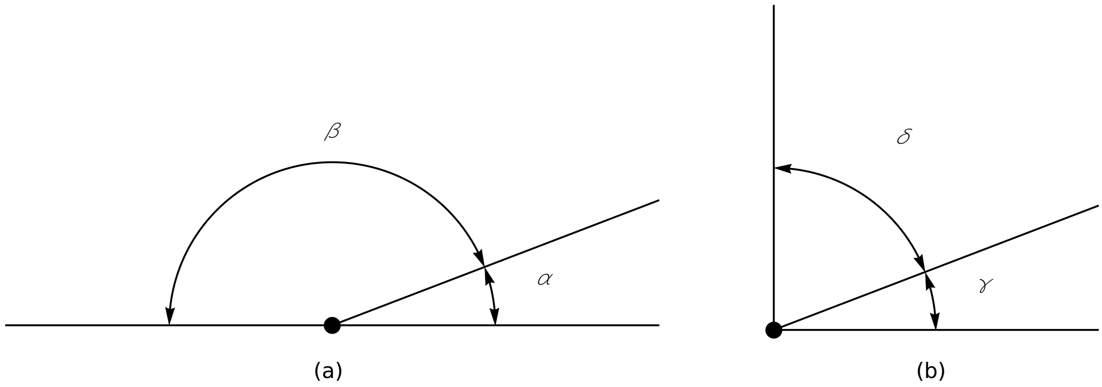


Figure 5.7: Supplementary (a) and complementary angles (b).

An angle is said to be in **standard position** if its vertex is the origin and its initial side coincides with the positive x-axis. Furthermore, two angles in standard position are called **coterminal** if they share the same terminal side. Such angles always differ by a multiple of 360° and since there are infinitely many integers, any given angle has infinitely many coterminal angles.

Remember that the real number π is defined to be the ratio of a circle's circumference C to its diameter d. It is a mathematical constant. In terms of the radius, we equivalently have $2\pi = \frac{C}{r}$. This tells us that for any circle, the ratio of its circumference to its radius is also always constant; in this case the constant is 2π. Suppose now we take a portion of the circle, so instead of comparing the entire circumference C to the radius, we compare some **arc** (*boog*) measuring s units in length to the radius (Figure 5.8). Let θ be the angle whose vertex is the centre of the circle and whose determining rays pass through the endpoints of the arc. Using proportionality arguments, it stands to reason that the ratio $\frac{s}{r}$ should also be a constant among all circles, and it is this ratio which defines the **radian measure** (*radialen*) of an angle.

We note that an angle with radian measure 1 means the corresponding arc length s equals the radius of the circle r, hence $s = r$. When the radian measure is 2, we have $s = 2r$; when the radian measure

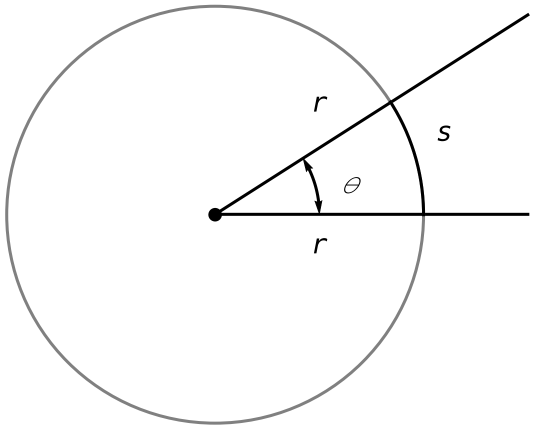


Figure 5.8: The radian measure of θ is $\frac{s}{r}$.

is 3, $s = 3r$, and so forth. Thus the radian measure of an angle θ tells us how many 'radius lengths' we need to sweep out along the circle to subtend the angle θ. Since one revolution sweeps out the entire circumference $2\pi r$, one revolution has radian measure $\frac{2\pi r}{r} = 2\pi$. Note that, by definition, the radian measure of an angle is a length divided by another length so that these measurements are actually dimensionless and are considered dimensionless numbers.

Since one revolution counter-clockwise measures 360° and the same angle measures 2π radians, we can use the proportion $\frac{2\pi \text{ radians}}{360°}$, or equivalently, $\frac{\pi \text{ radians}}{180°}$, as the conversion factor between degrees and radians. For example, to convert 60° to radians we find $\frac{\pi}{3}$ radians.

The merit of the radian measure lies in how easily angles in this measure can be identified with real numbers. Consider the unit circle, the angle θ in standard position, and the corresponding arc measuring s units in length. By definition, and the fact that the unit circle has radius 1, the radian measure of θ is $\frac{s}{r} = \frac{s}{1} = s$ so that we have $\theta = s$. In order to identify real numbers with oriented angles, we make good use of this fact by essentially wrapping the real number line around the unit circle and associating to each real number t an oriented arc on the unit circle with initial point $(1, 0)$.

5.3.2 Circular motion

Suppose an object is moving as along a circular path of radius r from the point P to the point Q in an amount of time t [T] (Figure 5.9).

Here s represents a **displacement** (*verplaatsing*) so that $s > 0$ means the object is traveling in a counter-clockwise direction and $s < 0$ indicates movement in a clockwise direction. The average velocity of the object, denoted $\overline{v}$ $[T^{-1}]$, is defined as the average rate of change of the position of the object with respect to time. As a result, we have $\overline{v} = \frac{s}{t}$. The quantity $\overline{v}$ $[LT^{-1}]$ conveys two ideas: the direction in which the object is moving and how fast the position of the object is changing. The contribution of direction in the quantity $\overline{v}$ is either to make it positive (in the case of counter-clockwise motion) or negative (in the case of clockwise motion), so that the quantity $|\overline{v}|$ quantifies how fast the object is moving - it is the speed of the object. Measuring θ in radians we have $\theta = \frac{s}{r}$ thus $s = r\theta$ and

$$\overline{v} = \frac{s}{t} = \frac{r\theta}{t} = r\frac{\theta}{t}.$$

The quantity $\frac{\theta}{t}$ is called the **average angular velocity** (*gemiddelde hoeksnelheid*) of the object, denoted by $\overline{\omega}$. The quantity $\overline{\omega}$ is the average rate of change of the angle θ with respect to time. If $\overline{\omega}$ is constant throughout the duration of the motion, then it can be shown that the average velocities are the same as their instantaneous counterparts, v and ω, respectively. In this case, v is simply called the velocity of the object and is the instantaneous rate of change of the position of the object with respect to time and likewise for ω.

If the path of the object were uncurled from a circle to form a line segment, then the velocity of the object on that line segment would be the same as the velocity on the circle. For this reason, the quantity v is often called the linear velocity of the object in order to distinguish it from the angular velocity, ω. Consequently, for an object moving on a circular path of radius r with constant angular velocity ω, the (linear) velocity of the object is given by $v = r\omega$.

Example 5.11

Assuming that the surface of the Earth is a sphere, any point on the Earth can be thought of as an object travelling on a circle which completes one revolution in (approximately) 24 hours. The path traced out by the point during this 24 hour period is the Latitude of that point. Campus Coupure of Ghent University is at 51.05325° north latitude, and the radius of the earth at this Latitude is

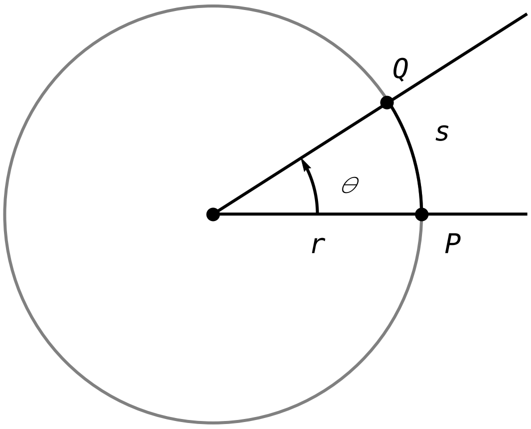


Figure 5.9: An object moving along a circular path of radius r from P to Q.

approximately 6365 km. Find the linear velocity of the campus as the world turns.

Solution

To use the formula $v = r\omega$, we first need to compute the angular velocity ω. The earth makes one revolution in 24 hours, and one revolution is 2π radians, so $\omega = \frac{2\pi\,\text{radians}}{24\,\text{hours}} = \frac{\pi}{12\,\text{hours}}$. We are also assuming that we are viewing the rotation of the earth as counter-clockwise so $\omega > 0$. Hence, the linear velocity is

$$v = 6365\,\text{km} \cdot \frac{\pi}{12\,\text{h}} \approx 1666\,\frac{\text{km}}{\text{h}}.$$

5.3.3 The six trigonometric functions

5.3.3.1 Cosine and sine on the unit circle

The sine and cosine of an acute angle are defined in the context of a right triangle. Consider for that purpose the generic right triangle with corresponding acute angle θ (Figure 5.10). The side with length a is called the **side of the triangle adjacent to** θ (*aanliggende rechthoekzijde*); the side with length b is called the **side of the triangle opposite** θ (*overstaande rechthoekzijde*); and the remaining side of length c (the side opposite the right angle) is called the **hypotenuse** (*hypothenusa*). We now imagine drawing this triangle in Quadrant I so that the angle θ is in standard position with the adjacent side to θ lying along the positive x-axis (Figure 5.11).

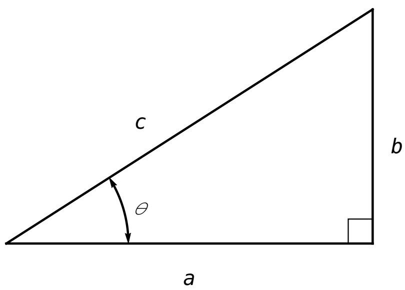


Figure 5.10: A right triangle with acute angle θ.

Then the **cosine** (*cosinus*) and **sine** (*sinus*) of θ are defined as

$$\cos(\theta) = \frac{a}{c} \tag{5.5}$$

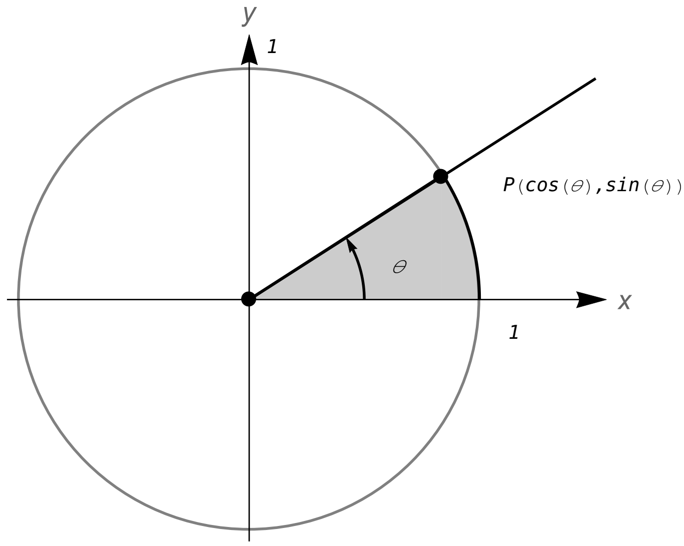


Figure 5.11: Position of a point P on the unit circle expressed in terms of $\cos(\theta)$ and $\sin(\theta)$.

and

$$\sin(\theta) = \frac{b}{c}, \tag{5.6}$$

so we have determined the cosine and sine of θ in terms of the lengths of the sides of the right triangle.

We now imagine drawing this right triangle in Quadrant I so that the angle θ is in standard position with the adjacent side to θ lying along the positive x-axis and the point P lies on the unit circle (Figure 5.11). Given Equations (5.5) and (5.6), we see that the x-coordinate of the point P is the cosine of θ, while the y-coordinate of P is the sine of θ (Figure 5.11). Since, for each angle θ, there is only one associated value of $\cos(\theta)$ and only one associated value of $\sin(\theta)$, both the sine and cosine may be interpreted as functions. It can be verified easily that the gray area in Figure 5.11 equals $\theta/2$; that is halve the circular angle θ. Besides, given how we defined angles, it is easy to see that $\cos(\theta) = \cos(\theta + 2\,\pi)$ for all $\theta \in [0, 2\pi]$, from which it follows that the cosine function is periodic with period 2π. Similarly, we find that also the sine function is periodic with period 2π.

Having introduced the sine and cosine, we should recall one of the most important identities in trigonometry.

Theorem 5.4 (The Pythagorean identity)
For any angle θ, it holds that

$$\cos^2(\theta) + \sin^2(\theta) = 1. \tag{5.7}$$

We summarize the cosine and sine values for certain common angles in Table 5.1.

To determine the cosines and sines of angles not given in Table 5.1 we may exploit the symmetry inherent in the unit circle. Suppose, for instance, we wish to know the cosine and sine of $\theta = 5\pi/6$. We plot θ in standard position below and, as usual, let $P(x, y)$ denote the point on the terminal side of θ which lies on the unit circle. Note that the terminal side of θ lies $\pi/6$ radians short of one half revolution (Figure 5.12(b)). Knowing from Table 5.1 that $\cos\left(\frac{\pi}{6}\right) = \frac{\sqrt{3}}{2}$ and $\sin\left(\frac{\pi}{6}\right) = \frac{1}{2}$. This means that the point on the terminal side of the angle $\frac{\pi}{6}$, when plotted in standard position, is $\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$. From Figure 5.12(b), it is clear that the point $P(x, y)$ we seek can be obtained by reflecting that point about the y-axis. Hence, $\cos\left(\frac{5\pi}{6}\right) = -\frac{\sqrt{3}}{2}$ and $\sin\left(\frac{5\pi}{6}\right) = \frac{1}{2}$.

Table 5.1: Cosine and sine value of common angles

θ (degrees)	θ (radians)	$\cos(\theta)$	$\sin(\theta)$
$0°$	0	1	0
$30°$	$\frac{\pi}{6}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$
$45°$	$\frac{\pi}{4}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$
$60°$	$\frac{\pi}{3}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$
$90°$	$\frac{\pi}{2}$	0	1

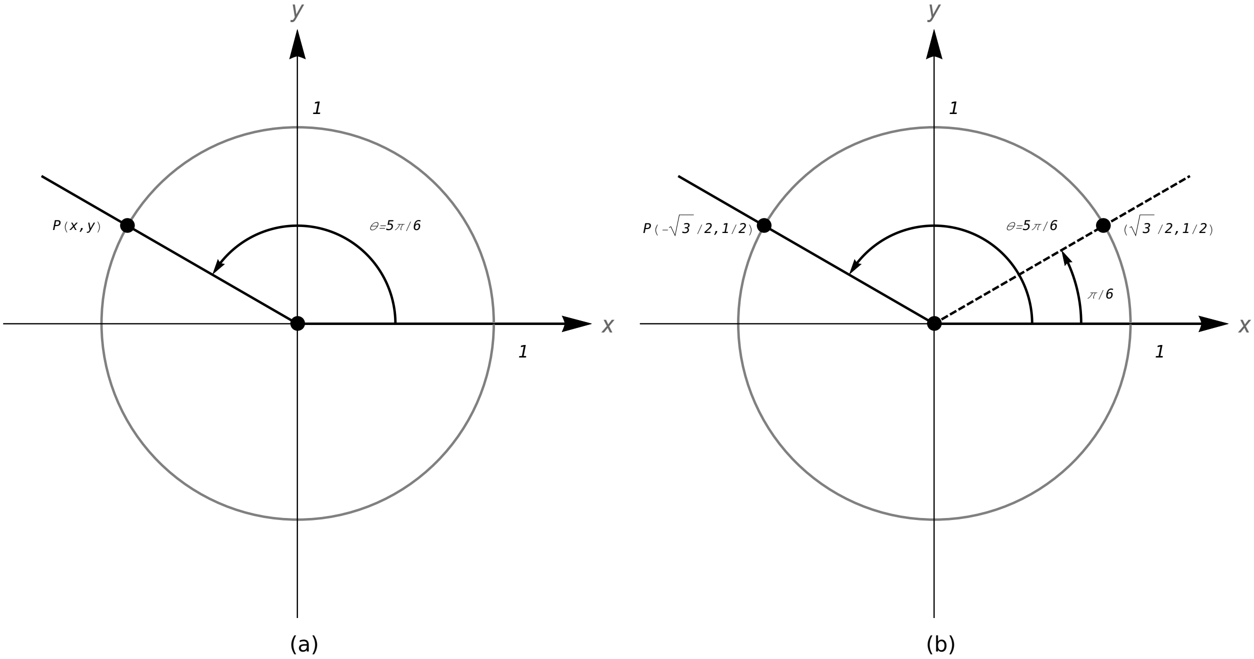


Figure 5.12: Determining the $\cos\left(\frac{5\pi}{6}\right)$ and $\sin\left(\frac{5\pi}{6}\right)$.

5.3.3.2 Beyond the unit circle

In defining the cosine and sine functions, we assigned to each angle a position on the unit circle. Here, we broaden our scope to include circles of radius r centred at the origin. Consider for the moment the acute angle θ drawn in Figure 5.13 in standard position. Let $Q(x,y)$ be the point on the terminal side of θ which lies on the circle $x^2+y^2=r^2$, and let $P(\widetilde{x},\widetilde{y})$ be the point on the terminal side of θ which lies on the unit circle. Now consider dropping perpendiculars from P and Q to create two right triangles, ΔOPA and ΔOQB. These triangles are **similar** (*gelijkvormig*) , thus it follows that $\frac{x}{\widetilde{x}}=\frac{r}{1}=r$, so $x=r\widetilde{x}$ and, similarly, we find $y=r\widetilde{y}$. Since, by definition, $\widetilde{x}=\cos(\theta)$ and $\widetilde{y}=\sin(\theta)$, we get the coordinates of Q to be $x=r\cos(\theta)$ and $y=r\sin(\theta)$. By reflecting these points through the x-axis, y-axis and origin, we obtain the result for all other angles θ.

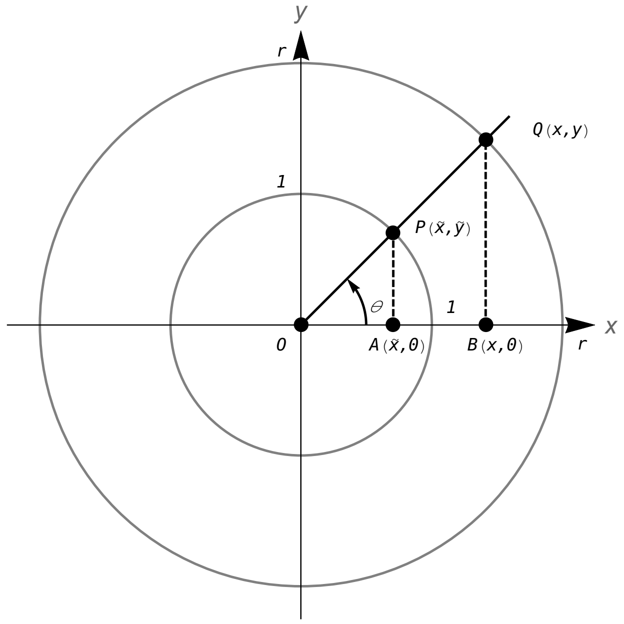


Figure 5.13: Relation between the coordinates of a point P on the unit circle and a point Q on the circle $x^2 + y^2 = r^2$, where both P and Q lie on the terminal side of θ.

The result is summarized in the following theorem.

Theorem 5.5 (Sine and cosine in the unit circle)
If $Q(x, y)$ is the point on the terminal side of an angle θ, plotted in standard position, which lies on the circle $x^2 + y^2 = r^2$ then $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Moreover,

$$\cos(\theta) = \frac{x}{r} = \frac{x}{\sqrt{x^2 + y^2}} \quad \textit{and} \quad \sin(\theta) = \frac{y}{r} = \frac{y}{\sqrt{x^2 + y^2}}.$$

Theorem 5.5 also gives us what we need to describe the position of an object travelling in a circular path of radius r with constant angular velocity ω. Suppose that at time t, the object has swept out an angle measuring θ radians. If we assume that the object is at the point $(r, 0)$ when $t = 0$, the angle θ is in standard position. By definition, $\omega = \frac{\theta}{t}$ which we rewrite as $\theta = \omega t$. According to Theorem 5.5, the location of the object $Q(x, y)$ on the circle is found using the equations $x = r\cos(\theta) = r\cos(\omega t)$ and $y = r\sin(\theta) = r\sin(\omega t)$. Hence, at time t, the object is at the point $(r\cos(\omega t), r\sin(\omega t))$, where $\omega > 0$ indicates a counter-clockwise direction and $\omega < 0$ indicates a clockwise direction.

Example 5.12

Suppose we are in the situation of Example 5.11. Find the equations of motion of Campus Coupure as the earth rotates.

Solution

From Example 5.11, we take $r = 6365$ km and and $\omega = \frac{\pi}{12\,\text{hours}}$. Hence, the equations of motion are $x = r\cos(\omega t) = 6365\cos\left(\frac{\pi}{12}t\right)$ and $y = r\sin(\omega t) = 6365\sin\left(\frac{\pi}{12}t\right)$, where x and y are measured in km and t is measured in hours.

5.3.3.3 The other trigonometric functions

Starting from the sine and cosine functions we introduced earlier, we may define four more trigonometric functions, being the **secant** (*secans*) of x:

$$\sec(x) = \frac{1}{\cos(x)}, \qquad \text{provided} \cos(x) \neq 0,$$

the **cosecant** (*cosecans*) of x:

$$\csc(x) = \frac{1}{\sin(x)}, \qquad \text{provided} \sin(x) \neq 0,$$

the **tangent** (*tangens*) of x:

$$\tan(x) = \frac{\sin(x)}{\cos(x)}, \qquad \text{provided} \cos(x) \neq 0,$$

and finally the **cotangent** (*cotangens*) of x:

$$\cot(x) = \frac{\cos(x)}{\sin(x)}, \qquad \text{provided} \sin(x) \neq 0.$$

Note that of the six trigonometric functions, only cosine and sine are defined for all angles. Table 5.2 lists the tangent and cotangent for certain common angles.

Table 5.2: Tangent and cotangent of common angles

x (degrees)	x (radians)	$\tan(x)$	$\cot(x)$
0°	0	0	undefined
30°	$\frac{\pi}{6}$	$\frac{\sqrt{3}}{3}$	$\sqrt{3}$
45°	$\frac{\pi}{4}$	1	1
60°	$\frac{\pi}{3}$	$\sqrt{3}$	$\frac{\sqrt{3}}{3}$
90°	$\frac{\pi}{2}$	undefined	0

By combining Equations (5.5) and (5.6) with the definition of the four other trigonometric functions, we can also express these in terms of the sides of a right triangle

$$\sec(\theta) = \frac{c}{a}, \qquad \csc(\theta) = \frac{c}{b}, \qquad \tan(\theta) = \frac{b}{a}, \qquad \cot(\theta) = \frac{a}{b},$$

where a, b and c are defined as in Figure 5.10.

Example 5.13

In order to determine the height of the maple tree (*Acer pseudoplatanus*) in the inner yard of Campus Coupure, two sightings from the ground, one 20 metres directly behind the other, are made. If the angles of inclination were 45° and 30°, respectively, how tall is the tree to the nearest foot?

Solution

Sketching the problem situation in Figure 5.14, we find ourselves with two unknowns: the height h of the tree and the distance x from the base of the tree to the first observation point.

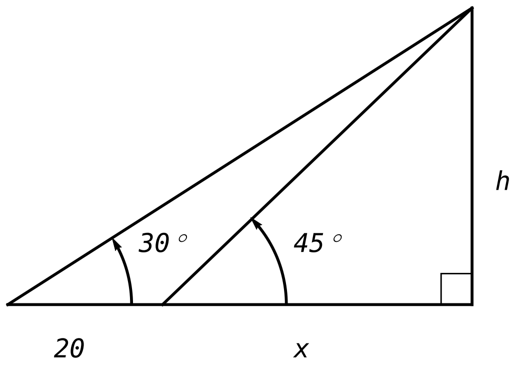


Figure 5.14: Determining the height of a maple tree.

Using the definition of the tangent function in terms of the sides of a right triangle, we get a pair of equations: $\tan(45°) = \frac{h}{x}$ and $\tan(30°) = \frac{h}{x+20}$. Since $\tan(45°) = 1$, the first equation gives $\frac{h}{x} = 1$, or $x = h$. Substituting this into the second equation gives $\frac{h}{h+20} = \tan(30°) = \frac{\sqrt{3}}{3}$. Clearing fractions, we get $3h = (h+20)\sqrt{3}$. The result is a linear equation for h, so we proceed to expand the right hand side and gather all the terms involving h to one side.

$$\begin{aligned} 3h &= (h+20)\sqrt{3} \\ \Leftrightarrow \quad (3-\sqrt{3})h &= 20\sqrt{3} \\ \Leftrightarrow \quad h &= \frac{20\sqrt{3}}{3-\sqrt{3}} \approx 27.32 \end{aligned}$$

Hence, the tree is approximately 27 metres tall.

5.3.4 Trigonometric identities

Here, we recall several collections of identities which have uses in this course and beyond.

5.3.4.1 Pythagorean identities

Given the four newly defined trigonometric functions, it makes sense to combine their definitions with the Pythagorean identity (Theorem 5.4) to derive new Pythagorean-like identities for the remaining four trigonometric functions. Assuming $\cos(x) \neq 0$, we may start with $\cos^2(x) + \sin^2(x) = 1$ and divide both sides by $\cos^2(x)$ to obtain $1 + \frac{\sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)}$. This reduces to

$$1 + \tan^2(x) = \sec^2(x)\,. \tag{5.8}$$

If $\sin(x) \neq 0$, we can divide both sides of the identity $\cos^2(x) + \sin^2(x) = 1$ by $\sin^2(x)$ to obtain

$$\cot^2(x) + 1 = \csc^2(x)\,. \tag{5.9}$$

These identities play an important role in not just trigonometry, but in calculus as well. We will use them later find the values of the trigonometric functions of an angle and solve equations and inequalities. In calculus, they are needed to simplify otherwise complicated expressions.

Example 5.14

Verify the following identities. Assume that all quantities are defined.

1. $\Big(\sec(\theta)-\tan(\theta)\Big)\Big(\sec(\theta)+\tan(\theta)\Big)=1$
2. $\dfrac{\sec(\theta)}{1-\tan(\theta)}=\dfrac{1}{\cos(\theta)-\sin(\theta)}$
3. $6\sec(\theta)\tan(\theta)=\dfrac{3}{1-\sin(\theta)}-\dfrac{3}{1+\sin(\theta)}$
4. $\dfrac{\sin(\theta)}{1-\cos(\theta)}=\dfrac{1+\cos(\theta)}{\sin(\theta)}$

Solution

In verifying identities, we typically start with the more complicated side of the equation and use known identities to transform it into the other side of the equation.

1. Expanding the left hand side of the equation gives:

$$\big(\sec(\theta)-\tan(\theta)\big)\big(\sec(\theta)+\tan(\theta)\big)=\sec^2(\theta)-\tan^2(\theta)\,,$$

   which equals 1 according to Equation (5.8).

2. While both sides of our last identity contain fractions, the left side affords us more opportunities to use our identities. Substituting $\sec(\theta)=\frac{1}{\cos(\theta)}$ and $\tan(\theta)=\frac{\sin(\theta)}{\cos(\theta)}$, we get:

$$\begin{aligned}\frac{\sec(\theta)}{1-\tan(\theta)} &= \frac{\dfrac{1}{\cos(\theta)}}{1-\dfrac{\sin(\theta)}{\cos(\theta)}}=\frac{\left(\dfrac{1}{\cos(\theta)}\right)\cos(\theta)}{\left(1-\dfrac{\sin(\theta)}{\cos(\theta)}\right)\cos(\theta)}\\ &= \frac{1}{\cos(\theta)-\sin(\theta)}\,,\end{aligned}$$

   which is exactly what we had set out to show.

3. The right hand side of the equation seems to hold more promise. We get common denominators and add:

$$\begin{aligned}\frac{3}{1-\sin(\theta)}-\frac{3}{1+\sin(\theta)} &= \frac{3(1+\sin(\theta))-3(1-\sin(\theta))}{(1+\sin(\theta))(1-\sin(\theta))}\\ &= \frac{(3+3\sin(\theta))-(3-3\sin(\theta))}{1-\sin^2(\theta)}\\ &= \frac{6\sin(\theta)}{1-\sin^2(\theta)}\,.\end{aligned}$$

   Since we wish to transform this expression into $6\sec(\theta)\tan(\theta)$, we use a reciprocal and quotient identity and find

$$6\sec(\theta)\tan(\theta)=6\left(\frac{1}{\cos(\theta)}\right)\left(\frac{\sin(\theta)}{\cos(\theta)}\right).$$

In other words, we need to get cosines in our denominator. From Equation (5.7) we have $1-\sin^2(\theta) = \cos^2(\theta)$ so we get:

$$\begin{aligned}
\frac{3}{1-\sin(\theta)} - \frac{3}{1+\sin(\theta)} &= \frac{6\sin(\theta)}{1-\sin^2(\theta)} = \frac{6\sin(\theta)}{\cos^2(\theta)} \\
&= 6\left(\frac{1}{\cos(\theta)}\right)\left(\frac{\sin(\theta)}{\cos(\theta)}\right) = 6\sec(\theta)\tan(\theta)\,.
\end{aligned}$$

4. It is debatable which side of the identity is more complicated. One thing which stands out is that the denominator on the left hand side is $1-\cos(\theta)$, while the numerator of the right hand side is $1+\cos(\theta)$. This suggests the strategy of starting with the left hand side and multiplying the numerator and denominator by the quantity $1+\cos(\theta)$:

$$\begin{aligned}
\frac{\sin(\theta)}{1-\cos(\theta)} &= \frac{\sin(\theta)}{(1-\cos(\theta))}\cdot\frac{(1+\cos(\theta))}{(1+\cos(\theta))} \\
&= \frac{\sin(\theta)(1+\cos(\theta))}{1-\cos^2(\theta)} = \frac{\sin(\theta)(1+\cos(\theta))}{\sin^2(\theta)} \\
&= \frac{\cancel{\sin(\theta)}(1+\cos(\theta))}{\cancel{\sin(\theta)}\sin(\theta)} = \frac{1+\cos(\theta)}{\sin(\theta)}\,.
\end{aligned}$$

5.3.4.2 Even/odd, cofunction and difference and sum identities

Our first set of identities is even/odd identities, which are summarized in the following theorem.

Theorem 5.6 (Even/Odd identities)
For all applicable angles θ, it holds that

- $\cos(-\theta) = \cos(\theta)$
- $\sin(-\theta) = -\sin(\theta)$
- $\tan(-\theta) = -\tan(\theta)$
- $\sec(-\theta) = \sec(\theta)$
- $\csc(-\theta) = -\csc(\theta)$
- $\cot(-\theta) = -\cot(\theta)$

Proof To proof these identities it suffices to show $\cos(-\theta) = \cos(\theta)$ and $\sin(-\theta) = -\sin(\theta)$. For that purpose, consider angles θ and $-\theta$ plotted in standard position, where $0 \le \theta \le 2\pi$ (Figure 5.15). Let P and Q denote the points on the terminal sides of θ and $-\theta$, respectively, which lie on the unit circle. Hence, the coordinates of P are $(\cos(\theta), \sin(\theta))$ and the coordinates of Q are $(\cos(-\theta), \sin(-\theta))$. It follows that the points P and Q are symmetric about the x-axis, such that $\cos(-\theta) = \cos(\theta)$ and $\sin(-\theta) = -\sin(\theta)$. □

The even/odd identities can be used to derive the sum and difference identities for cosine.

Theorem 5.7 (Sum and difference identities for cosine)
For all angles α and β, it holds that

$$\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)\,, \tag{5.10}$$

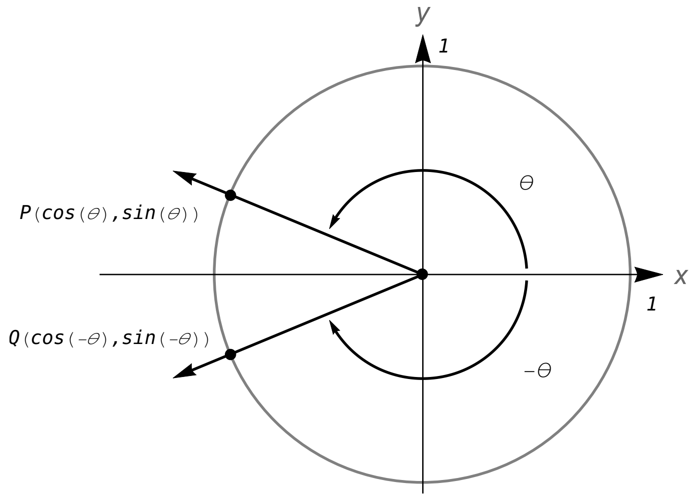


Figure 5.15: Even/Odd identity for cosine and sine.

$$\cos(\alpha-\beta) \quad = \quad \cos(\alpha)\cos(\beta)+\sin(\alpha)\sin(\beta)\,. \tag{5.11}$$

We can use the sum and difference identities for cosine to derive the so-called cofunction identities. The results are summarized in the following theorem.

Theorem 5.8 (Cofunction identities)
For all applicable angles θ, it holds that

- $\cos\left(\frac{\pi}{2}-\theta\right)=\sin(\theta)$
- $\sec\left(\frac{\pi}{2}-\theta\right)=\csc(\theta)$
- $\tan\left(\frac{\pi}{2}-\theta\right)=\cot(\theta)$
- $\sin\left(\frac{\pi}{2}-\theta\right)=\cos(\theta)$
- $\csc\left(\frac{\pi}{2}-\theta\right)=\sec(\theta)$
- $\cot\left(\frac{\pi}{2}-\theta\right)=\tan(\theta)$

Proof Consider for instance $\cos\left(\frac{\pi}{2}-\theta\right)$. Straightforward application of the difference identity for cosine yields:

$$\begin{aligned}
\cos\left(\frac{\pi}{2}-\theta\right) &= \cos\left(\frac{\pi}{2}\right)\cos(\theta)+\sin\left(\frac{\pi}{2}\right)\sin(\theta)\\
&= (0)\big(\cos(\theta)\big)+(1)\big(\sin(\theta)\big)\\
&= \sin(\theta)\,.
\end{aligned}$$

Moreover, from this result we immediately get that

$$\sin\left(\frac{\pi}{2}-\theta\right)=\cos\left(\frac{\pi}{2}-\left[\frac{\pi}{2}-\theta\right]\right)=\cos(\theta),$$

which says, in words, that the cosine of an angle is the sine of its complement. Now that these identities have been established for cosine and sine, the remaining trigonometric functions follow suit. □

With the cofunction identities in place, we are now in the position to derive the sum and difference formulas for sine. To achieve this, we convert to cosines using a cofunction identity, then expand using the difference formula for cosine

$$\begin{aligned}
\sin(\alpha+\beta) &= \cos\left(\frac{\pi}{2}-(\alpha+\beta)\right) \\
&= \cos\left(\left[\frac{\pi}{2}-\alpha\right]-\beta\right) \\
&= \cos\left(\frac{\pi}{2}-\alpha\right)\cos(\beta)+\sin\left(\frac{\pi}{2}-\alpha\right)\sin(\beta) \\
&= \sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta)\,.
\end{aligned}$$

We can derive the difference formula for sine by rewriting $\sin(\alpha-\beta)$ as $\sin(\alpha+(-\beta))$ and using the sum formula and the even / odd identities.

Theorem 5.9 (Sum and difference identities for sine)
For all angles α and β, it holds that

$$\sin(\alpha+\beta) = \sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta)\,, \tag{5.12}$$

$$\sin(\alpha-\beta) = \sin(\alpha)\cos(\beta)-\cos(\alpha)\sin(\beta)\,. \tag{5.13}$$

Also for the tangent function, one may derive sum and difference identities.

Example 5.15

Derive a formula for $\tan(\alpha+\beta)$ in terms of $\tan(\alpha)$ and $\tan(\beta)$.

Solution

We can start expanding $\tan(\alpha+\beta)$ using a quotient identity and our sum formulas

$$\begin{aligned}
\tan(\alpha+\beta) &= \frac{\sin(\alpha+\beta)}{\cos(\alpha+\beta)} \\
&= \frac{\sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta)}{\cos(\alpha)\cos(\beta)-\sin(\alpha)\sin(\beta)}\,.
\end{aligned}$$

Since $\tan(\alpha)=\frac{\sin(\alpha)}{\cos(\alpha)}$ and $\tan(\beta)=\frac{\sin(\beta)}{\cos(\beta)}$, it looks as though if we divide both numerator and denominator by $\cos(\alpha)\cos(\beta)$ we will have what we want

$$\begin{aligned}
\tan(\alpha+\beta) &= \frac{\sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta)}{\cos(\alpha)\cos(\beta)-\sin(\alpha)\sin(\beta)}\cdot\frac{\frac{1}{\cos(\alpha)\cos(\beta)}}{\frac{1}{\cos(\alpha)\cos(\beta)}} \\
&= \frac{\frac{\sin(\alpha)\cancel{\cos(\beta)}}{\cos(\alpha)\cancel{\cos(\beta)}}+\frac{\cancel{\cos(\alpha)}\sin(\beta)}{\cancel{\cos(\alpha)}\cos(\beta)}}{\frac{\cancel{\cos(\alpha)}\cancel{\cos(\beta)}}{\cancel{\cos(\alpha)}\cancel{\cos(\beta)}}-\frac{\sin(\alpha)\sin(\beta)}{\cos(\alpha)\cos(\beta)}} \\
&= \frac{\tan(\alpha)+\tan(\beta)}{1-\tan(\alpha)\tan(\beta)}\,.
\end{aligned}$$

Naturally, this formula is limited to those cases where all of the tangents are defined.

Theorem 5.10 (Sum and difference identities for tangent)
For all applicable angles α and β, it holds that

$$\tan(\alpha+\beta) = \frac{\tan(\alpha)+\tan(\beta)}{1-\tan(\alpha)\tan(\beta)}, \tag{5.14}$$

$$\tan(\alpha-\beta) = \frac{\tan(\alpha)-\tan(\beta)}{1+\tan(\alpha)\tan(\beta)}. \tag{5.15}$$

5.3.4.3 Double angle identities and power reduction formulas

Theorem 5.11 (Double angle identities)
For all applicable angles θ, it holds that

$$\cos(2\theta) = \cos^2(\theta)-\sin^2(\theta), \tag{5.16}$$

$$\sin(2\theta) = 2\sin(\theta)\cos(\theta), \tag{5.17}$$

$$\tan(2\theta) = \frac{2\tan(\theta)}{1-\tan^2(\theta)}. \tag{5.18}$$

Proof Specialize the sum formulas for cosine (Theorem 5.7), sine (Theorem 5.9) and tangent (Theorem 5.10) to the case where $\alpha = \beta$. □

Note that the Pythagorean identity (Theorem 5.4) allows to express $\cos(2\theta)$ alternatively as

$$\cos(2\theta) = 2\cos^2(\theta)-1, \tag{5.19}$$

and

$$\cos(2\theta) = 1-2\sin^2(\theta). \tag{5.20}$$

In calculus, we will often be confronted with situations where it is useful to reduce the power of cosine and sine. Solving Equation (5.19) for $\cos^2(\theta)$ and the Equation (5.20) for $\sin^2(\theta)$ results in the aptly-named power reduction formulas below. These are also known as **Carnot's formulas** (*formules van Carnot*).

Theorem 5.12 (Power reduction formulas)
For all angles θ, it holds that

$$\cos^2(\theta) = \frac{1+\cos(2\theta)}{2}, \tag{5.21}$$

$$\sin^2(\theta) = \frac{1-\cos(2\theta)}{2}. \tag{5.22}$$

Not only can these formulas be used to reduce the power of cosine, they can as well be used to derive expressions for the cosine, sine and tangent of half angle. To start, we apply the power reduction formula to $\cos^2\left(\frac{\theta}{2}\right)$

$$\cos^2\left(\frac{\theta}{2}\right) = \frac{1+\cos\left(2\left(\frac{\theta}{2}\right)\right)}{2} = \frac{1+\cos(\theta)}{2}.$$

5.3.4.4 Product to sum formulas and vice versa

The product to sum formulas, are easily verified by expanding each of the right hand sides in accordance with Theorems 5.7 and 5.9. They are of particular use in calculus, and we list them here for reference. These are also known as **Simpson's formulas** (*formules van Simpson*).

Theorem 5.13 (Product to sum formulas)
For all angles α and β, it holds that

$$\cos(\alpha)\cos(\beta) = \frac{1}{2}\big(\cos(\alpha-\beta)+\cos(\alpha+\beta)\big)\,, \tag{5.23}$$

$$\sin(\alpha)\sin(\beta) = \frac{1}{2}\big(\cos(\alpha-\beta)-\cos(\alpha+\beta)\big)\,, \tag{5.24}$$

$$\sin(\alpha)\cos(\beta) = \frac{1}{2}\big(\sin(\alpha-\beta)+\sin(\alpha+\beta)\big)\,. \tag{5.25}$$

Related to the product to sum formulas are the sum to product formulas. These are easily verified using the product to sum formulas, and as such, their proofs are left as exercises.

Theorem 5.14 (Sum to product formulas)
For all angles α and β, it holds that

$$\cos(\alpha)+\cos(\beta) = 2\cos\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right)\,, \tag{5.26}$$

$$\cos(\alpha)-\cos(\beta) = -2\sin\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right)\,, \tag{5.27}$$

$$\sin(\alpha)\pm\sin(\beta) = 2\sin\left(\frac{\alpha\pm\beta}{2}\right)\cos\left(\frac{\alpha\mp\beta}{2}\right)\,. \tag{5.28}$$

5.3.5 Graphs of cosine and sine functions

5.3.5.1 Domain and range

In the same way we studied polynomial, rational, exponential, and logarithmic functions, we will study the trigonometric functions $f(x)=\cos(x)$ and $g(x)=\sin(x)$. The first order of business is to find the domains and ranges of these functions. Whether we think of identifying the real number x with the angle $\theta=x$ radians, or think of wrapping an oriented arc around the unit circle to find coordinates on the unit circle, it should be clear that both the cosine and sine functions are defined for all real numbers x. So $\operatorname{dom}\cos(x)=\operatorname{dom}\sin(x)=\mathbb{R}$, and likewise $\operatorname{im}\cos(x)=\operatorname{im}\sin(x)=[-1,1]$.

5.3.5.2 Graphs

The even/odd identities in Theorem 5.6 tell us $\cos(-x)=\cos(x)$ for all real numbers x and $\sin(-x)=-\sin(x)$ for all real numbers x. This means $f(x)=\cos(x)$ is an even function, while $g(x)=\sin(x)$ is an odd function (see Chapter 3). Another important property of these functions is that $\cos(x+2\pi k)=$

$\cos(x)$ and $\sin(x+2\pi k) = \sin(x)$, for all real numbers x and any integer k; that is the sine and cosine functions are periodic with period 2π. One last property of the functions $f(x) = \cos(x)$ and $g(x) = \sin(x)$ is worth pointing out: the graphs of both of these functions have no jumps, gaps, holes in the graph, asymptotes, corners or cusps.

To graph $y = \cos(x)$, we make a table using some of the common values of x in the interval $[0, 2\pi]$ (Table 5.3). This generates a portion of the cosine graph, which we call the **fundamental cycle** (*fundamentele cyclus*) of $y = \cos(x)$ (Figure 5.16(a)).

Table 5.3: Cosine and sine of some common angles.

x	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{4}$	$\frac{3\pi}{2}$	$\frac{7\pi}{4}$	2π
$\cos(x)$	1	$\frac{\sqrt{2}}{2}$	0	$-\frac{\sqrt{2}}{2}$	-1	$-\frac{\sqrt{2}}{2}$	0	$\frac{\sqrt{2}}{2}$	1
$\sin(x)$	0	$\frac{\sqrt{2}}{2}$	1	$\frac{\sqrt{2}}{2}$	0	$-\frac{\sqrt{2}}{2}$	-1	$-\frac{\sqrt{2}}{2}$	0

To get the graph of the cosine function for intervals stretching beyond the fundamental cycle, we way imagine copying and pasting this graph end to end infinitely in both directions (left and right) on the x-axis (Figure 5.16(c)).

We can plot the fundamental cycle of the graph of $y = \sin(x)$ similarly, with similar results (Figures 5.16(b) and 5.16(d)). It is of course no accident that the graphs of $y = \cos(x)$ and $y = \sin(x)$ are so similar. Using a cofunction identity (Theorem 5.8) along with the even property of cosine, we have

$$\sin(x) = \cos\left(\frac{\pi}{2} - x\right) = \cos\left(-\left(x - \frac{\pi}{2}\right)\right) = \cos\left(x - \frac{\pi}{2}\right)$$

Recalling Section 3.2.5, we see from this formula that the graph of $y = \sin(x)$ is the result of shifting the graph of $y = \cos(x)$ to the right $\frac{\pi}{2}$ units.

Now that we know the basic shapes of the graphs of $y = \cos(x)$ and $y = \sin(x)$, we can rely on the tools from Section 3.2.5 to graph more complicated curves. To do so, we need to keep track of the movement of some key points on the original graphs, such as $x = 0$, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$ and 2π.

Example 5.16

Graph one cycle of the following functions. State the period of each.

1. $f(x) = 3\cos\left(\frac{\pi x - \pi}{2}\right) + 1$
2. $g(x) = \frac{1}{2}\sin(\pi - 2x) + \frac{3}{2}$

Solution

1. We set the argument of the cosine, $\frac{\pi x - \pi}{2}$, equal to each of the values: 0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$, 2π and solve for x. This leads to x-values 1, 2, 3, 4 and 5. Next, we substitute each of these x-values into $f(x)$ to determine the corresponding y-values and connect the dots in a pleasing wavelike fashion (Figure 5.17(a)). One cycle is graphed on $[1, 5]$ so the period is the length

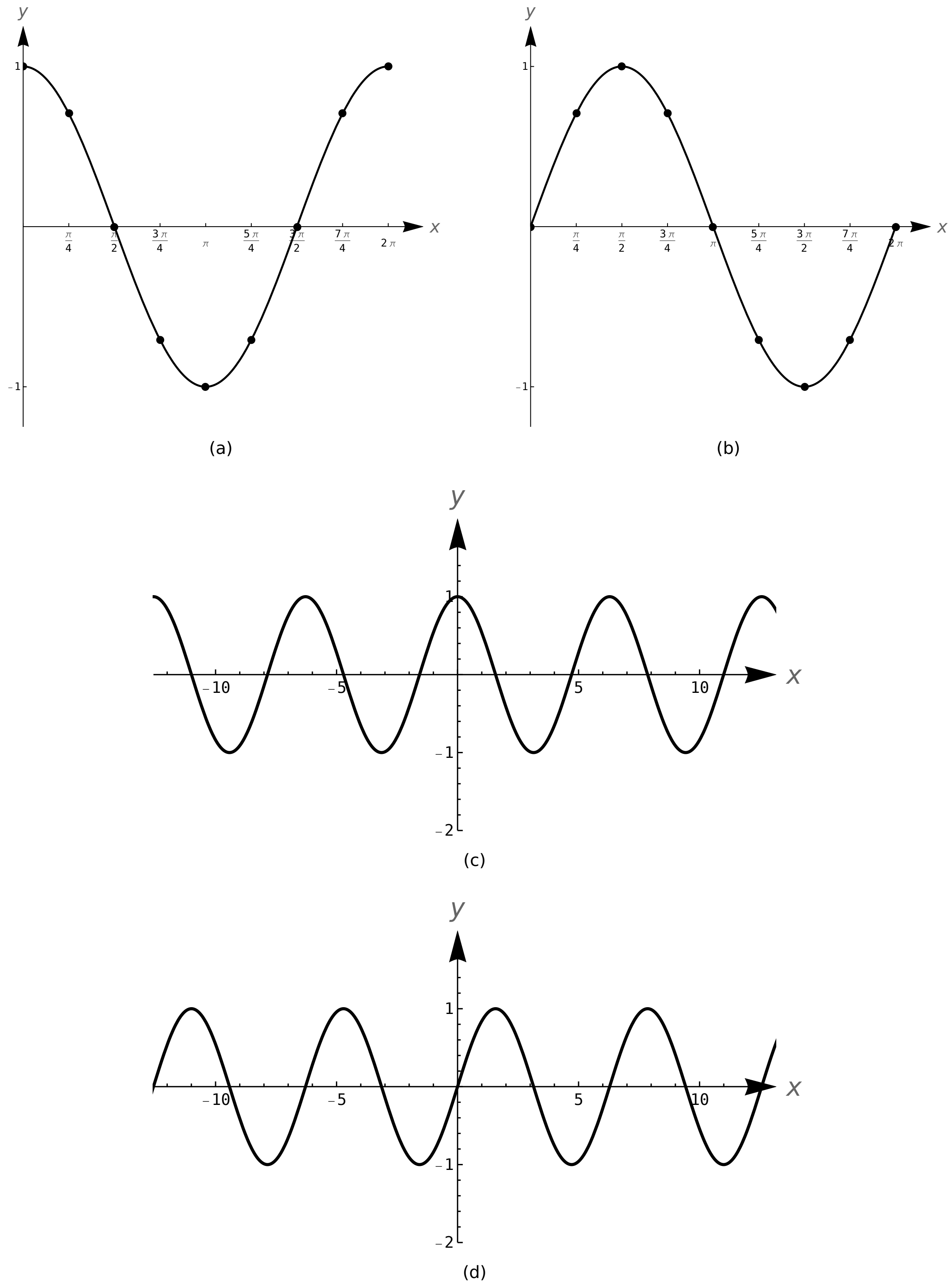


Figure 5.16: Graphs of the fundamental cycle (a,b) and four such cycles (c,d) of $y = \cos(x)$ (a,c) and $y = \sin(x)$ (b,d).

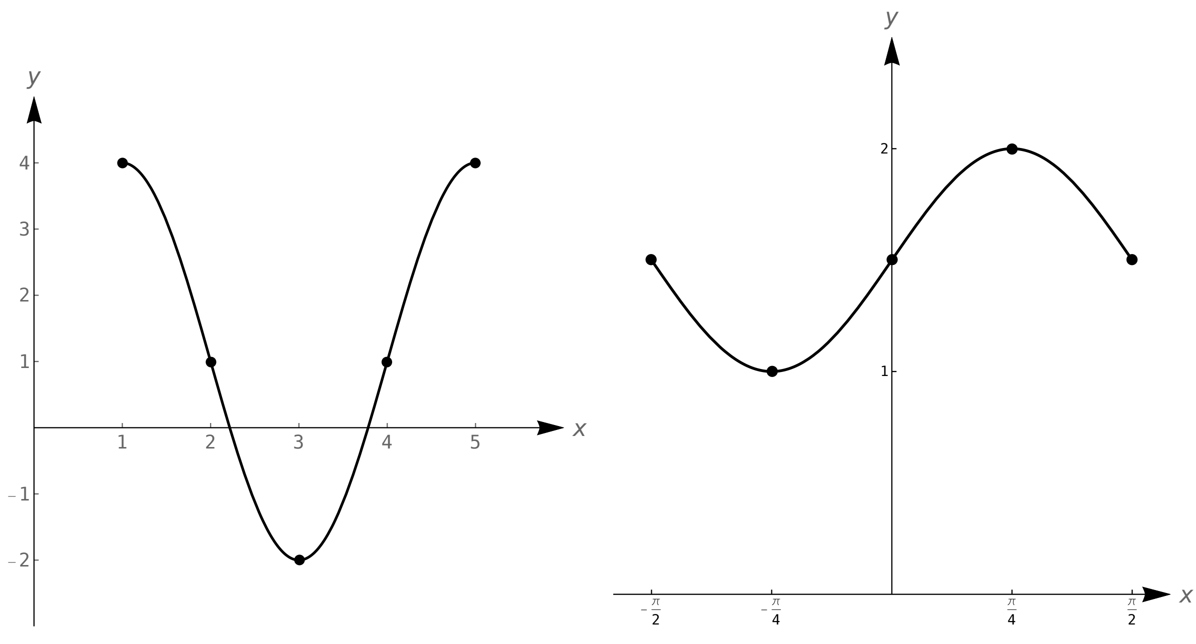
of that interval which is 4.

x	1	2	3	4	5
$\cos\left(\frac{\pi x - \pi}{2}\right)$	1	0	-1	0	1
$f(x)$	4	1	-2	1	4

2. Again, we set the argument of the sine, $\pi - 2x$, equal to each of our quarter marks, being 0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$, 2π and solve for x, which leads to x-values $\frac{\pi}{2}$, $\frac{\pi}{4}$, 0, $-\frac{\pi}{4}$ and $-\frac{\pi}{2}$. We now find the corresponding y-values on the graph by substituting each of these x-values into $g(x) = \frac{1}{2}\sin(\pi - 2x) + \frac{3}{2}$.

x	$\frac{\pi}{2}$	$\frac{\pi}{4}$	0	$-\frac{\pi}{4}$	$-\frac{\pi}{2}$
$\sin(\pi - 2x)$	0	1	0	-1	0
$f(x)$	$\frac{3}{2}$	2	$\frac{3}{2}$	1	$\frac{3}{2}$

Once again, we connect the dots in a wavelike fashion (Figure 5.17(b)). One cycle was graphed on the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ so the period is $\frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi$.



(a) (b)

Figure 5.17: The of $f(x) = 3\cos\left(\frac{\pi x - \pi}{2}\right) + 1$ (a) and $g(x) = \frac{1}{2}\sin(\pi - 2x) + \frac{3}{2}$ (b).

The functions in Example 5.16 are examples of **sinusoids** (*sinusoïde*) . Roughly speaking, a sinusoid is the result of taking the basic graph of $f(x) = \cos(x)$ or $g(x) = \sin(x)$ and performing any of the transformations mentioned in Section 3.2.5. Sinusoids can be characterized by four properties: period, amplitude, phase shift and vertical shift. More specifically, for the functions

$$y = A\cos(\omega x + \phi) + B \quad \text{and} \quad y = A\sin(\omega x + \phi) + B,$$

or equivalently,

$$y = A\cos\left[\omega\left(x + \frac{\phi}{\omega}\right)\right] + B \quad \text{and} \quad y = A\sin\left[\omega\left(x + \frac{\phi}{\omega}\right)\right] + B,$$

we have:

- period $\dfrac{2\pi}{\omega}$
- amplitude $|A|$
- phase shift $-\dfrac{\phi}{\omega}$
- vertical shift B

for $\omega > 0$. Here, ϕ is called the **phase** (*fase*) of the sinusoid, while ω is nothing but the angular velocity. It is the number of cycles the sinusoid completes over a 2π interval.

5.3.6 Graphs of the other trigonometric functions

Before constructing the graphs of the other trigonometric functions, we first determine their domains and ranges.

5.3.6.1 Domain and range

Starting from the domain and range of the cosine and sine functions, we can now determine the domains and ranges of the other trigonometric functions.

For what concerns the function $f(x) = \sec(x) = \frac{1}{\cos(x)}$. We know that it is undefined whenever $\cos(x) = 0$. Since we know $\cos(x) = 0$ whenever $x = \frac{\pi}{2} + \pi k$ for integers k, the domain of this function, in set builder notation is

$$\text{dom } \sec(x) = \left\{x : x \neq \frac{\pi}{2} + \pi k, \forall k \in \mathbb{Z}\right\}.$$

Using interval notation to describe this set, we get

$$\text{dom } \sec(x) = \ldots \cup \left]-\frac{5\pi}{2}, -\frac{3\pi}{2}\right[\cup \left]-\frac{3\pi}{2}, -\frac{\pi}{2}\right[\cup \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[\cup \left]\frac{\pi}{2}, \frac{3\pi}{2}\right[\cup \left]\frac{3\pi}{2}, \frac{5\pi}{2}\right[\cup \ldots$$

This is, however, a very cumbersome notation. In order to write this in a more compact way, we note that from the set-builder description of the domain, the kth point excluded from the domain, which we will call x_k, and which is given by $x_k = \frac{(2k+1)\pi}{2}$. Hence, the domain of $f(x) = \sec(x)$ consists of the intervals determined by successive points x_k:

$$]x_k, x_{k+1}[= \left]\frac{(2k+1)\pi}{2}, \frac{(2k+3)\pi}{2}\right[,$$

where $k = 0, \pm1, \pm2, \ldots$. The union of infinitely many such intervals can be written as

$$\operatorname{dom}\sec(x) = \bigcup_{k=-\infty}^{+\infty} \left]\frac{(2k+1)\pi}{2}, \frac{(2k+3)\pi}{2}\right[.$$

To determine the range of $f(x) = \sec(x)$, we appeal to the definition $\sec(x) = \frac{1}{\cos(x)}$ and recall that the range of $f(x) = \cos(x)$ is $[-1, 1]$. Since $f(x) = \sec(x)$ is undefined when $\cos(x) = 0$, we split our discussion into two cases: when $0 < \cos(x) \le 1$ and when $-1 \le \cos(x) < 0$. If the former case we can divide the inequality $\cos(x) \le 1$ by $\cos(x)$ to obtain $\sec(x) = \frac{1}{\cos(x)} \ge 1$. Moreover, we have that as $\cos(x) \to 0^+$, $\sec(x) \to +\infty$. If, on the other hand, if $-1 \le \cos(x) < 0$, then dividing by $\cos(x)$ causes a reversal of the inequality so that $\sec(x) = \frac{1}{\cos(x)} \le -1$. In this case, as $\cos(x) \to 0^-$, we get $\sec(x) \to -\infty$. Since $f(x) = \cos(x)$ admits all of the values in $[-1, 1]$, the function $f(x) = \sec(x)$ admits all of the values in $]-\infty, -1] \cup [1, +\infty[$. Using set-builder notation, the range of $f(x) = \sec(x)$ can be written as

$$\operatorname{im}\sec(x) = \{u : u \le -1 \lor u \ge 1\} = \{u : |u| \ge 1\}.$$

Similar arguments can be used to determine the domains and ranges of the remaining three circular functions: $\csc(x)$, $\tan(x)$ and $\cot(x)$. The reader is encouraged to do so. The results are summarized in Table 5.4.

Table 5.4: Domains and ranges of the trigonometric functions.

Function	Domain	Range
$f(x) = \sin(x)$	$\mathbb{R}$	$[-1, 1]$
$f(x) = \cos(x)$	$\mathbb{R}$	$[-1, 1]$
$f(x) = \sec(x)$	$\bigcup_{k=-\infty}^{\infty} \left]\frac{(2k+1)\pi}{2}, \frac{(2k+3)\pi}{2}\right[$	$\{u : \|u\| \ge 1\} =]-\infty, -1] \cup [1, +\infty[$
$f(x) = \csc(x)$	$\bigcup_{k=-\infty}^{\infty}]k\pi, (k+1)\pi[$	$\{u : \|u\| \ge 1\} =]-\infty, -1] \cup [1, +\infty[$
$f(x) = \tan(x)$	$\bigcup_{k=-\infty}^{\infty} \left]\frac{(2k+1)\pi}{2}, \frac{(2k+3)\pi}{2}\right[$	$\mathbb{R}$
$f(x) = \cot(x)$	$\bigcup_{k=-\infty}^{\infty}]k\pi, (k+1)\pi[$	$\mathbb{R}$

5.3.6.2 Graphs of tangent and cotangent functions

Finally, we turn our attention to the graphs of the tangent and cotangent functions. When constructing a table of values for the tangent function, we see that $y = \tan(x)$ is undefined at $x = \frac{\pi}{2}$ and $x = \frac{3\pi}{2}$. As $x \underset{<}{\to} \frac{\pi}{2}$, $\sin(x) \to 1^-$ and $\cos(x) \to 0^+$, so that $\tan(x) = \frac{\sin(x)}{\cos(x)} \to +\infty$ producing a vertical asymptote at $x = \frac{\pi}{2}$. Using a similar analysis, we get that as $x \underset{>}{\to} \frac{\pi}{2}$, $\tan(x) \to -\infty$; as $x \underset{<}{\to} \frac{3\pi}{2}$, $\tan(x) \to +\infty$; and as

$x \underset{>}{\rightarrow} \frac{3\pi}{2}$, $\tan(x) \rightarrow -\infty$. Plotting this information yields the graph in Figures 5.18(a) and 5.18(c).

From this graph, it appears as if the tangent function is periodic with period π. To prove that this is the case, we appeal to the sum formula for tangents. We have:

$$\tan(x+\pi) = \frac{\tan(x)+\tan(\pi)}{1-\tan(x)\tan(\pi)} = \frac{\tan(x)+0}{1-(\tan(x))(0)} = \tan(x),$$

which tells us the period of $\tan(x)$ is at most π. To show that it is exactly π, suppose p is a positive real number so that $\tan(x+p) = \tan(x)$ for all real numbers x. For $x = 0$, we have $\tan(p) = \tan(0+p) = \tan(0) = 0$, which means p is a multiple of π. The smallest positive multiple of π is π itself, so we have established the result. We take as our fundamental cycle for $y = \tan(x)$ the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

It should be no surprise that $y = \cot(x)$ behaves similarly to $y = \tan(x)$. It clearly appears from Figures 5.18(b) and 5.18(d) as if the period of $\cot(x)$ is π, and we leave it to the reader to prove this. We take as one fundamental cycle the interval $[0, \pi]$.

Example 5.17

Graph one cycle of the following functions. State the period of each.

1. $f(x) = 1 - 2\sec(2x)$
2. $g(x) = 2\cot\left(\frac{\pi}{2}x + \pi\right) + 1$

Solution

1. To graph $f(x) = 1-2\sec(2x)$, we first set the argument of secant, $2x$, equal to 0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$ and 2π and solve for x, to get 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$, π. Next, we substitute these x-values into $f(x)$. If $f(x)$ exists, we have a point on the graph in Figure 5.19(a); otherwise, we have found a vertical asymptote. In addition to these points and asymptotes, we have graphed the associated cosine curve (dashed curve), being $y = 1-2\cos(2x)$. Since one cycle is graphed over the interval $[0, \pi]$, the period is $\pi - 0 = \pi$.

x	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π
$\sec(2x)$	1	undefined	-1	undefined	1
$f(x)$	-1	undefined	3	undefined	-1

2. We take 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$ and π as quarter marks for constructing the fundamental cycle of the cotangent curve. To graph this function, we begin by setting $\frac{\pi}{2}x + \pi$ equal to each such mark and solve for x, to arrive at -2, $\frac{-3}{2}$, -1, $\frac{-1}{2}$ and 0. We now use these x-values to generate our graph (Figure 5.19(b)). We find the period to be $0-(-2) = 2$.

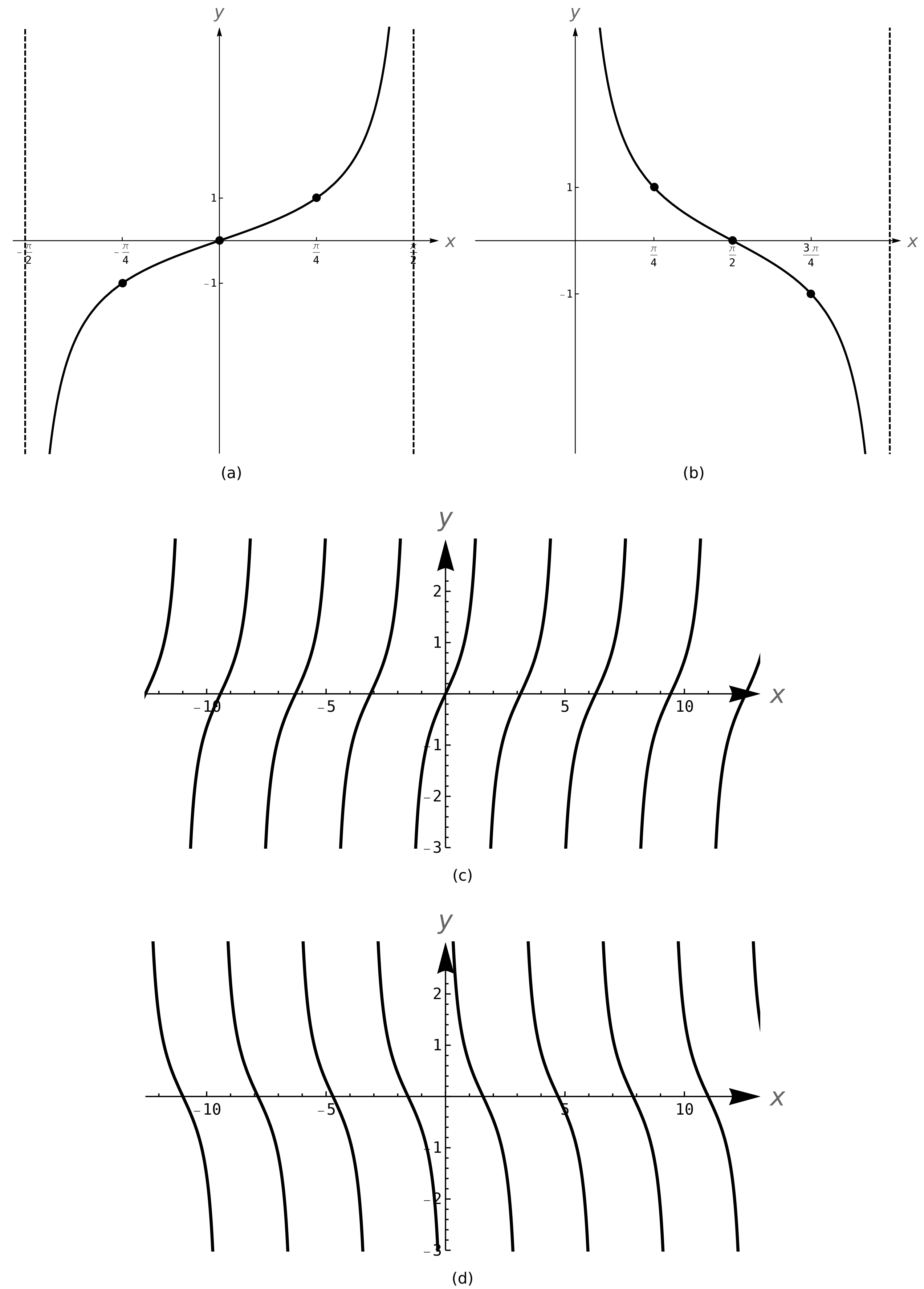
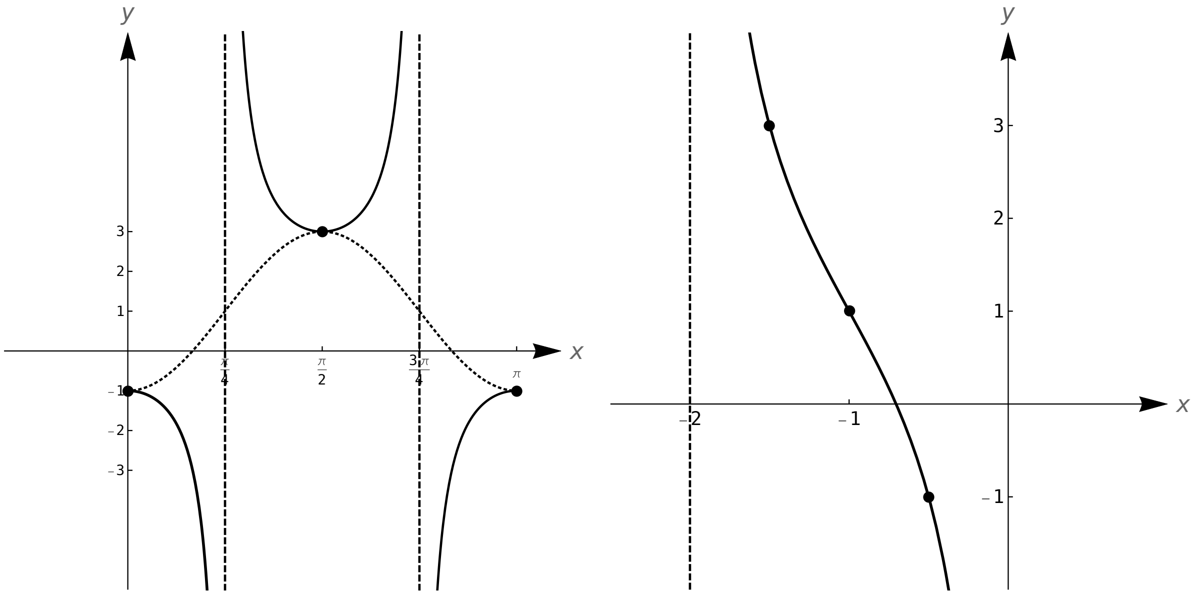


Figure 5.18: The fundamental cycle (a,b) and eight cycles (c,d) of $y = \tan(x)$ (a,c) and $y = \cot(x)$ (b,d).

x	-2	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0
$\cot\left(\frac{\pi}{2}x+\pi\right)$	undefined	1	0	-1	undefined
$g(x)$	undefined	3	1	-1	undefined



(a) (b)

Figure 5.19: One cycle of $y = 1-2\sec(2x)$ (a) and $y = 2\cot(\frac{\pi}{2}x+\pi)+1$ (b) over $[0, 2\pi]$.

5.3.6.3 Applications

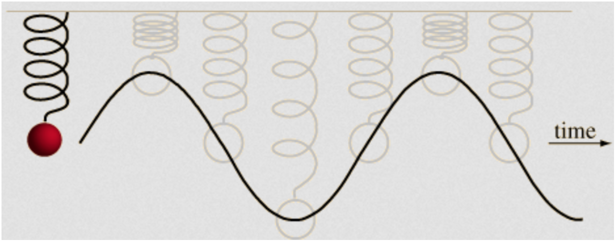
Trigonometric functions are often used to solve problems that involve periodic behaviour, such as

- problems that involve circular movement on a repetitive nature;
- problems involving repetitive motion, such as spring motion (Figure 5.20(a)), oscillating waves and tides (Figure 5.20(b));
- problems involving environmental fluctuations (Figure 5.20(c)).

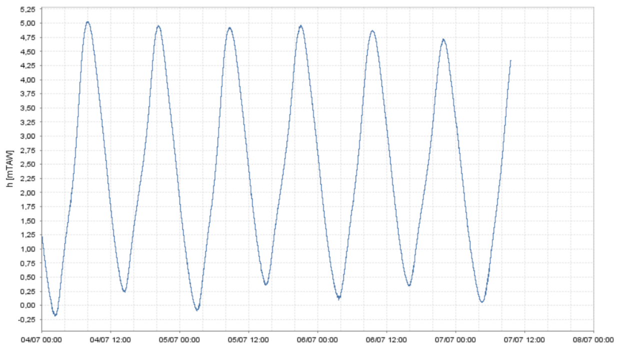
Example 5.18

According to the U.S. Naval Observatory website, the number of hours H of daylight that Fairbanks, Alaska received on the 21st day of the nth month of 2009 is given below. Here $t = 1$ represents January 21, 2009, $t = 2$ represents February 21, 2009, and so on.

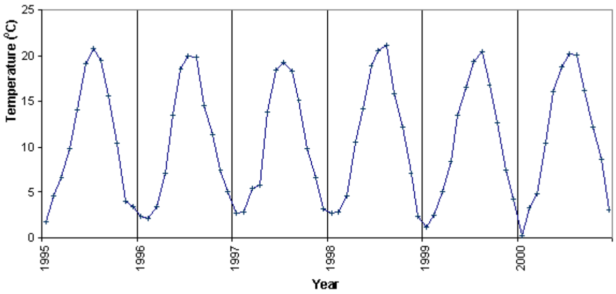
Month	1	2	3	4	5	6	7	8	9	10	11	12
Hours of Daylight	5.8	9.3	12.4	15.9	19.4	21.8	19.4	15.6	12.4	9.1	5.6	3.3



(a) Motion of an object attached to a spring.



(b) Water height above sea level of the river Scheldt at the monitoring station Prosperpolder.



(c) Variation in annual cycle of river temperature at the Struma River in Bulgaria at the site of Boboshevo.

Figure 5.20: Applications of trigonometric functions illustrated.

Find a sinusoid that models these data and use Mathematica to graph your answer along with the data.

— Solution — To get a feel for the data, we first plot it using the Mathematica built-in function **`ListPlot`**. The result is plotted in Figure 5.21.

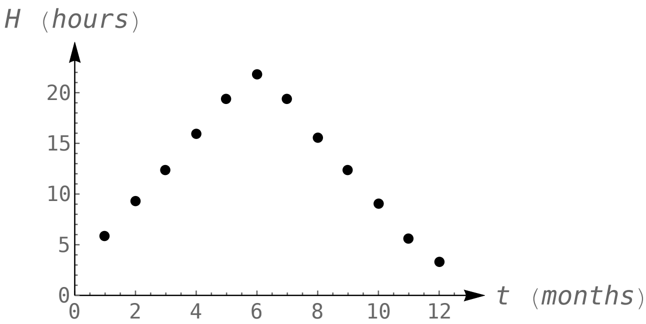


Figure 5.21: Hours of daylight on the 21st day of the month of 2009 in Fairbanks, Alaska

We do our best to find the constants A, ω, ϕ and B so that $H(t) = A\sin(\omega t + \phi) + B$ closely matches the data. We first go after the vertical shift B whose value determines the baseline. In a typical sinusoid, the value of B is the average of the maximum and minimum values. So here we take $B = \frac{3.3+21.8}{2} = 12.55$. Next is the amplitude A which is the displacement from the baseline to the maximum (and minimum) values. We find $A = 21.8 - 12.55 = 12.55 - 3.3 = 9.25$. At this point, we have $H(t) = 9.25\sin(\omega t + \phi) + 12.55$. Next, we go after the angular frequency ω. Since the data collected is over the span of a year (12 months), we take the period $T = 12$ months. This means $\omega = \frac{2\pi}{T} = \frac{2\pi}{12} = \frac{\pi}{6}$. The last quantity to find is the phase ϕ. It is easy to find the phase shift $-\frac{\phi}{\omega}$. Since we picked $A > 0$, the phase shift corresponds to the first value of t with $H(t) = 12.55$ (the baseline value). Here, we choose $t = 3$, since its corresponding H value of 12.4 is closer to 12.55 than the next value, 15.9, which corresponds to $t = 4$. Hence, $-\frac{\phi}{\omega} = 3$, so $\phi = -3\omega = -3\left(\frac{\pi}{6}\right) = -\frac{\pi}{2}$. We have

$$H(t) = 9.25\sin\left(\frac{\pi}{6}t - \frac{\pi}{2}\right) + 12.55,$$

whose graph is shown together with the data in Figure 5.22.

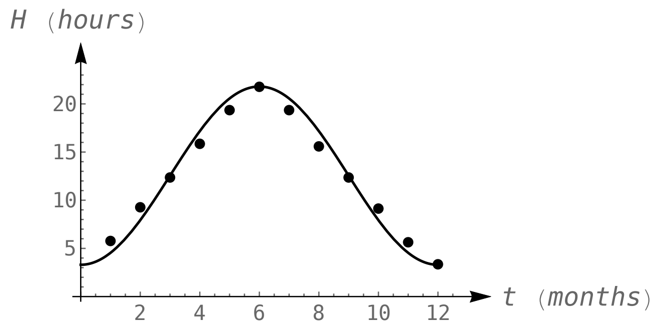


Figure 5.22: Hours of daylight on the 21st day of the month of 2009 in Fairbanks, Alaska

Forgotten trigonometric functions

In addition to the trigonometric functions we encountered in this chapter, there are a few functions that were common historically, but are nowadays sometimes forgotten, such as the chord, given by

$$\text{crd}(\theta) = 2\sin\left(\frac{\theta}{2}\right),$$

the versine

$$\text{versin}(\theta) = 1 - cos(\theta),$$

and the haversine

$$\text{haversin}(\theta) = \frac{1}{2}\text{versin}(\theta).$$

Still these functions are used in several fields. For instance, one period of a versine or haversine is commonly used in signal processing and control theory as the shape of a pulse.

5.4 Inverse trigonometric functions

In this section we concern ourselves with finding inverses of the trigonometric functions, which are also referred to as **arcus**, **antitrigonometric** or **cyclometric functions** (*cyclometrische functies*). Our immediate problem is that, owing to their periodic nature, none of the six trigoniometric functions is injective. To remedy this, we restrict the domains of the circular functions in the same way we restricted the domain of the quadratic function in Example 3.13 to obtain an injective function.

5.4.1 Inverse cosine and sine functions

We first consider $f(x) = \cos(x)$. Choosing the interval $[0,\pi]$ allows us to keep the range as $[-1,1]$ as well as the properties of being smooth and continuous. Recall from Section 3.4 that the inverse of a function f is typically denoted f^{-1}. For this reason, some textbooks use the notation $f^{-1}(x) = \cos^{-1}(x)$

for the inverse of $f(x) = \cos(x)$. The obvious pitfall here is our convention of writing $(\cos(x))^2$ as $\cos^2(x)$, $(\cos(x))^3$ as $\cos^3(x)$ and so on. It is far too easy to confuse $\cos^{-1}(x)$ with $\frac{1}{\cos(x)} = \sec(x)$, so we will not use this notation. Instead, we use the notation $f^{-1}(x) = \arccos(x)$, read **arccosine** (*boogcosinus*) of x. As a consequence of the imposed domain restriction, we have that

$$\cos(\arccos(x)) = x\,,$$

provided $-1 \leq x \leq 1$, and

$$\arccos(\cos(x)) = x\,,$$

provided $0 \leq x \leq \pi$, which is important to keep in mind when solving equations involving (inverse) trigonometric functions.

To understand the arc in arccosine, recall that an inverse function, by definition, reverses the process of the original function. The function $f(x) = \cos(x)$ takes a real number input x, associates it with the angle $\theta = x$ radians, and returns the value $\cos(\theta)$. Digging deeper, we have that $\cos(\theta) = \cos(x)$ is the x-coordinate of the terminal point on the unit circle of an oriented arc of length $|x|$ whose initial point is $(1, 0)$. Hence, we may view the inputs to $f(x) = \cos(x)$ as oriented arcs and the outputs as x-coordinates on the unit circle. The function f^{-1}, then, would take x-coordinates on the unit circle and return oriented arcs, hence the arc in arccosine. In Figure 5.23 are the graphs of $f(x) = \cos(x)$ and $f^{-1}(x) = \arccos(x)$, where we obtain the latter from the former by reflecting it across the line $y = x$, in accordance with what we learned in Section 3.4.

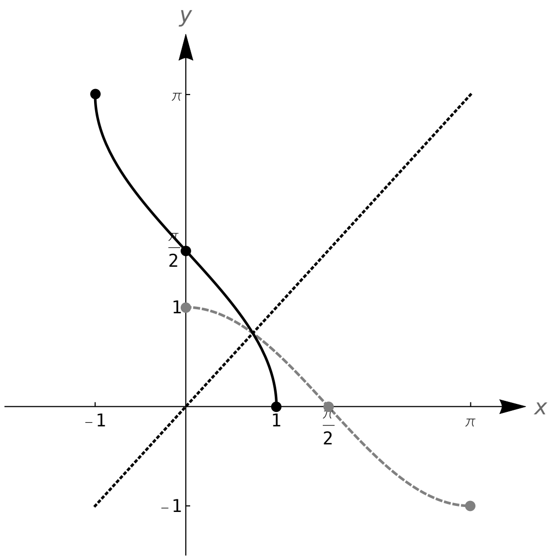


Figure 5.23: The graph of $y = \cos(x)$ for $x \in [0, \pi]$ (dashed) and $y = \arccos(x)$ for $x \in [-1, 1]$ (solid).

We restrict $g(x) = \sin(x)$ in a similar manner, although the interval of choice is $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ (Figure 5.24). It should be no surprise that we call $g^{-1}(x) = \arcsin(x)$, which is read **arcsine** (*boogsinus*) of x. From Figure 5.24, we may infer that the arcsine is an odd function. Besides, due to the domain restriction, we have that

$$\sin(\arcsin(x)) = x\,,$$

provided $-1 \leq x \leq 1$ and

$$\arcsin(\sin(x)) = x\,,$$

provided $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$.

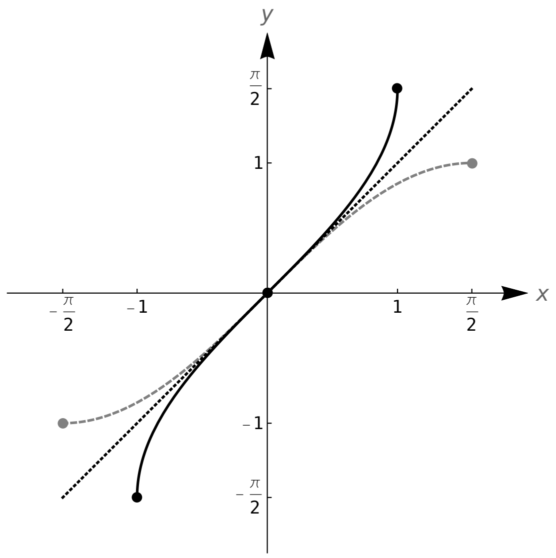


Figure 5.24: The graph of $y = \sin(x)$ for $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ (dashed) and $y = \arcsin(x)$ for $x \in [-1, 1]$ (solid).

Example 5.19

Find the exact values of the following.

1. $\arccos\left(\frac{1}{2}\right)$
2. $\arccos\left(-\frac{\sqrt{2}}{2}\right)$
3. $\arcsin\left(-\frac{1}{2}\right)$
4. $\arccos\left(\cos\left(\frac{\pi}{6}\right)\right)$
5. $\arccos\left(\cos\left(\frac{11\pi}{6}\right)\right)$
6. $\sin\left(\arccos\left(-\frac{3}{5}\right)\right)$

Solution

1. To find $\arccos\left(\frac{1}{2}\right)$, we need to find the real number x that lies between 0 and π for which it holds that $\cos(x) = \frac{1}{2}$. We know $x = \frac{\pi}{3}$ meets these criteria, so $\arccos\left(\frac{1}{2}\right) = \frac{\pi}{3}$.

2. The number $x = \arccos\left(-\frac{\sqrt{2}}{2}\right)$ lies in the interval $[0, \pi]$ with $\cos(x) = -\frac{\sqrt{2}}{2}$. Our answer is $\arccos\left(-\frac{\sqrt{2}}{2}\right) = \frac{3\pi}{4}$.

3. To find $\arcsin\left(-\frac{1}{2}\right)$, we seek the number x in the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ with $\sin(x) = -\frac{1}{2}$. The answer is $x = -\frac{\pi}{6}$ so that $\arcsin\left(-\frac{1}{2}\right) = -\frac{\pi}{6}$.

4. Since $0 \le \frac{\pi}{6} \le \pi$, we simply have that $\arccos\left(\cos\left(\frac{\pi}{6}\right)\right) = \frac{\pi}{6}$. However, in order to make sure we understand why this is the case, we can use the definition of arccosine. Working from the inside out, $\arccos\left(\cos\left(\frac{\pi}{6}\right)\right) = \arccos\left(\frac{\sqrt{3}}{2}\right)$. Now, $\arccos\left(\frac{\sqrt{3}}{2}\right)$ is the real number x with $0 \le x \le \pi$ and $\cos(x) = \frac{\sqrt{3}}{2}$. We find $x = \frac{\pi}{6}$, so that $\arccos\left(\cos\left(\frac{\pi}{6}\right)\right) = \frac{\pi}{6}$.

5. Since $\frac{11\pi}{6}$ does not fall between 0 and π, we are forced to work from the inside out starting

with $\arccos\left(\cos\left(\frac{11\pi}{6}\right)\right) = \arccos\left(\frac{\sqrt{3}}{2}\right)$. From the previous problem, we know $\arccos\left(\frac{\sqrt{3}}{2}\right) = \frac{\pi}{6}$. Hence, $\arccos\left(\cos\left(\frac{11\pi}{6}\right)\right) = \frac{\pi}{6}$.

6. We let $x = \arccos\left(-\frac{3}{5}\right)$ so that $\cos(x) = -\frac{3}{5}$ for some x where $0 \leq x \leq \pi$. Since $\cos(x) < 0$, we can narrow this down a bit and conclude that $\frac{\pi}{2} < x < \pi$, so that x corresponds to an angle in Quadrant II. In terms of x, then, we need to find $\sin\left(\arccos\left(-\frac{3}{5}\right)\right) = \sin(x)$. Using the Pythagorean identity, we get $\left(-\frac{3}{5}\right)^2 + \sin^2(x) = 1$ or $\sin(x) = \pm\frac{4}{5}$. Since x corresponds to a Quadrants II angle, we choose $\sin(x) = \frac{4}{5}$. Hence, $\sin\left(\arccos\left(-\frac{3}{5}\right)\right) = \frac{4}{5}$.

Most of the common errors encountered in dealing with the inverse trigonometric functions come from the need to restrict the domains of the original functions so that they are injective. One instance of this phenomenon is the fact that $\arccos\left(\cos\left(\frac{11\pi}{6}\right)\right) = \frac{\pi}{6}$ as opposed to $\frac{11\pi}{6}$. This is the exact same phenomenon discussed in Section 3.4 when we saw $\sqrt{(-2)^2} = 2$ as opposed to -2.

5.4.2 Inverse tangent and cotangent functions

The next pair of functions we discuss are the inverses of tangent and cotangent, which are named **arctangent** (*boogtangens*) and **arccotangent** (*boogcotangens*), respectively. First, we restrict $f(x) = \tan(x)$ to its fundamental cycle on $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ to obtain $f^{-1}(x) = \arctan(x)$. Among other things, note that the vertical asymptotes $x = -\frac{\pi}{2}$ and $x = \frac{\pi}{2}$ of the graph of $f(x) = \tan(x)$ become the horizontal asymptotes $y = -\frac{\pi}{2}$ and $y = \frac{\pi}{2}$ of the graph of $f^{-1}(x) = \arctan(x)$ (Figure 5.25). Observe that the arctangent is an odd function and due to the domain restrition that

$$\tan\left(\arctan(x)\right) = x\,,$$

for all real numbers x, whereas

$$\arctan(\tan(x)) = x\,,$$

$-\frac{\pi}{2} < x < \frac{\pi}{2}$ only.

Next, we restrict $g(x) = \cot(x)$ to its fundamental cycle on $]0, \pi[$ to obtain $g^{-1}(x) = \operatorname{arccot}(x)$. Once again, the vertical asymptotes $x = 0$ and $x = \pi$ of the graph of $g(x) = \cot(x)$ become the horizontal asymptotes $y = 0$ and $y = \pi$ of the graph of $g^{-1}(x) = \operatorname{arccot}(x)$ (Figure 5.26). Moreover, as a consequence of the imposed domain restriction, we have

$$\cot\left(\operatorname{arccot}(x)\right) = x\,,$$

for all real numbers x and

$$\operatorname{arccot}(\cot(x)) = x\,,$$

provided $0 < x < \pi$ only. Finally, we also have that

$$\arctan(x) = \operatorname{arccot}\left(\frac{1}{x}\right),$$

for $x > 0$ and likewise

$$\operatorname{arccot}(x) = \arctan\left(\frac{1}{x}\right),$$

for $x > 0$.

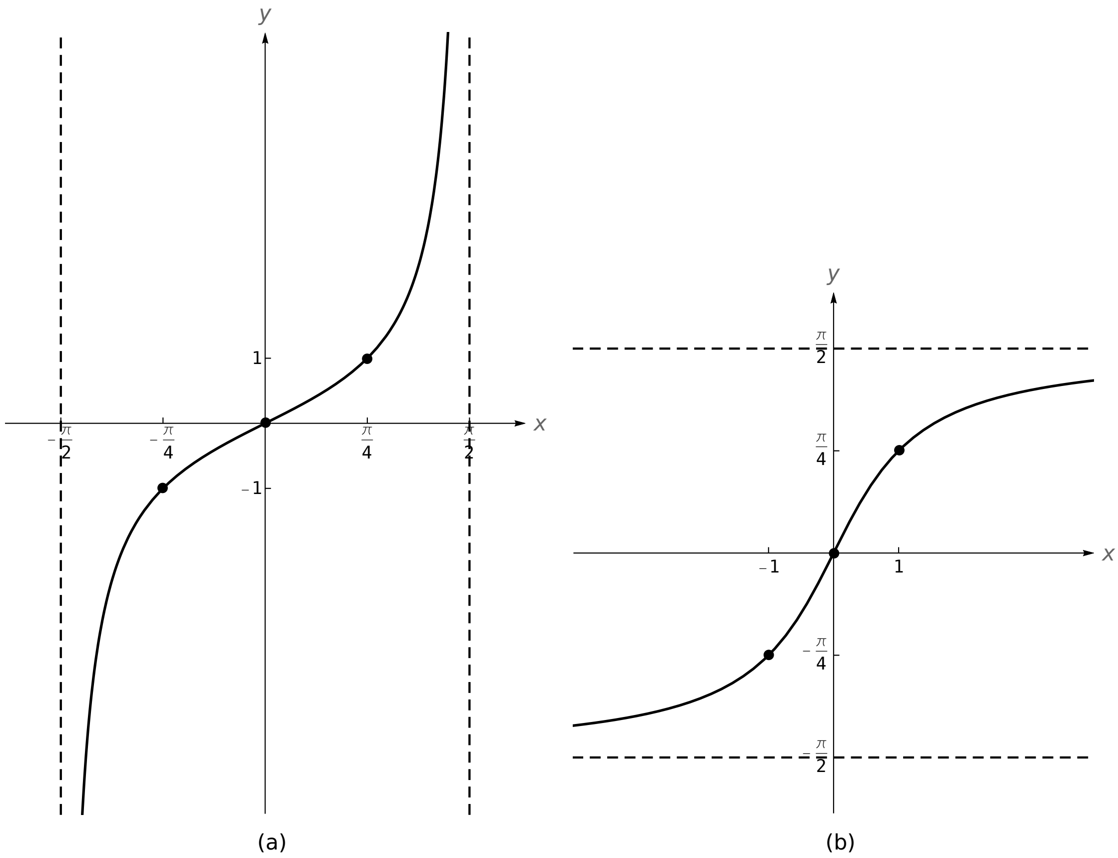


Figure 5.25: One cycle of $y = \tan(x)$ (a) and $y = \arctan(x)$ (b).

Example 5.20

Rewrite the following as algebraic expressions of x and state the domain on which the equivalence is valid.

1. $\tan(2\arctan(x))$
2. $\cos(\operatorname{arccot}(2x))$

Solution

1. If we let $t = \arctan(x)$, then $-\frac{\pi}{2} < t < \frac{\pi}{2}$ and $\tan(t) = x$. We look for a way to express $\tan(2\arctan(x)) = \tan(2t)$ in terms of x. Before we get started using identities, we note that $\tan(2t)$ is undefined when $2t = \frac{\pi}{2} + \pi k$ for integers k. Dividing both sides of this equation by 2 tells us we need to exclude values of t where $t = \frac{\pi}{4} + \frac{\pi}{2}k$, where k is an integer. The only members of this family which lie in $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ are $t = \pm\frac{\pi}{4}$, which means the values of t under consideration are $\left]-\frac{\pi}{2}, -\frac{\pi}{4}\right[\cup \left]-\frac{\pi}{4}, \frac{\pi}{4}\right[\cup \left]\frac{\pi}{4}, \frac{\pi}{2}\right[$. Returning to $\tan(2t)$, we note the double angle identity $\tan(2t) = \frac{2\tan(t)}{1-\tan^2(t)}$, is valid for all the values of t under consideration, hence we get
$$\tan(2\arctan(x)) = \tan(2t) = \frac{2\tan(t)}{1-\tan^2(t)} = \frac{2x}{1-x^2}\,.$$
To find where this equivalence is valid we check back with our substitution $t = \arctan(x)$. Since the domain of $\arctan(x)$ is all real numbers, the only exclusions come from the values of t we discarded earlier, $t = \pm\frac{\pi}{4}$. Since $x = \tan(t)$, this means we exclude $x = \tan\left(\pm\frac{\pi}{4}\right) = \pm 1$. Hence, the equivalence $\tan(2\arctan(x)) = \frac{2x}{1-x^2}$ holds for all x in $\mathbb{R} \setminus \{-1, 1\}$.

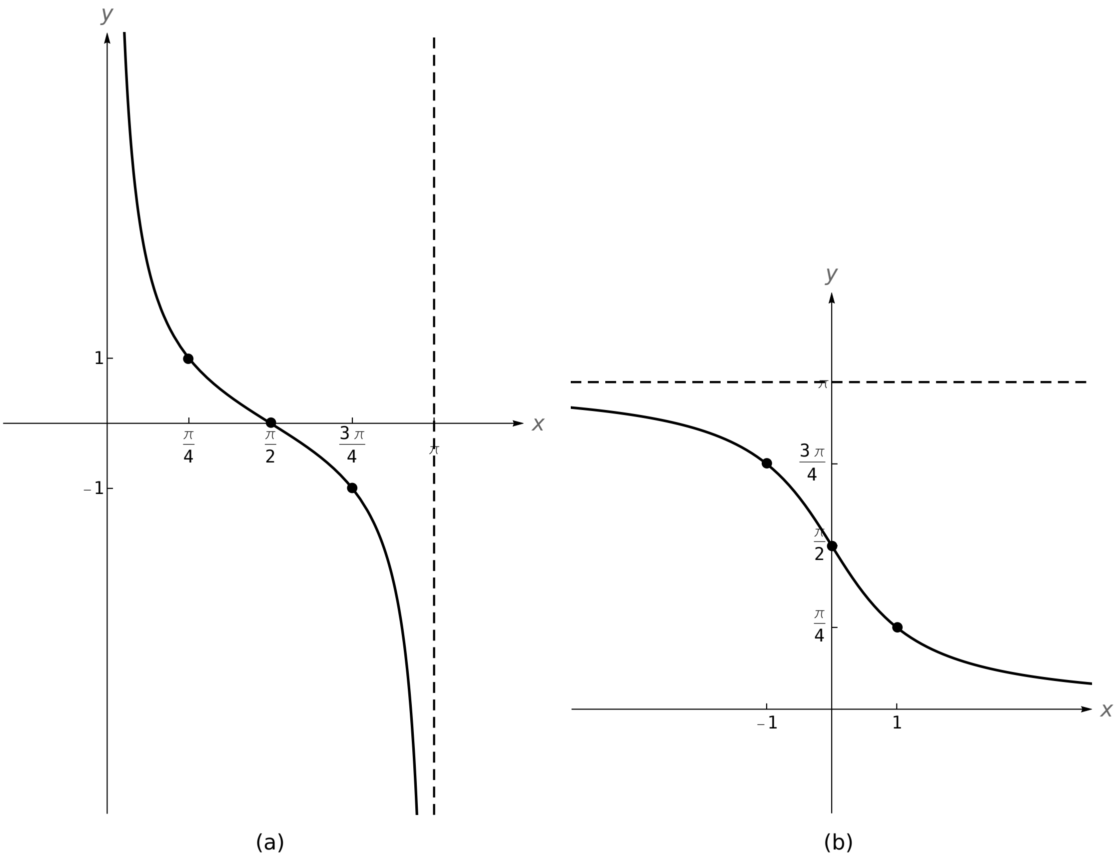


Figure 5.26: One cycle of $y = \cot(x)$ (a) and $y = \operatorname{arccot}(x)$ (b).

2. To get started, we let $t = \operatorname{arccot}(2x)$ so that $\cot(t) = 2x$ where $0 < t < \pi$. In terms of t, $\cos(\operatorname{arccot}(2x)) = \cos(t)$, and our goal is to express the latter in terms of x. Since $\cos(t)$ is always defined, there are no additional restrictions on t, so we can begin using identities to relate $\cot(t)$ to $\cos(t)$. The identity $\cot(t) = \frac{\cos(t)}{\sin(t)}$ is valid for t in $]0, \pi[$, so our strategy is to obtain $\sin(t)$ in terms of x, then write $\cos(t) = \cot(t)\sin(t)$. The identity $1 + \cot^2(t) = \csc^2(t)$ holds for all t in $]0, \pi[$ and relates $\cot(t)$ and $\csc(t) = \frac{1}{\sin(t)}$. Substituting $\cot(t) = 2x$, we get $1 + (2x)^2 = \csc^2(t)$, or $\csc(t) = \pm\sqrt{4x^2+1}$. Since t is between 0 and π, $\csc(t) > 0$, so $\csc(t) = \sqrt{4x^2+1}$ which gives $\sin(t) = \left(4x^2+1\right)^{-1/2}$. Hence,

$$\cos(\operatorname{arccot}(2x)) = \cos(t) = \cot(t)\sin(t) = \frac{2x}{\sqrt{4x^2+1}}.$$

Since $\operatorname{arccot}(2x)$ is defined for all real numbers x and we encountered no additional restrictions on t, we have $\cos\left(\operatorname{arccot}(2x)\right) = \frac{2x}{\sqrt{4x^2+1}}$ for all real numbers x.

For comprehensiveness, Table 5.5 lists the domain restrictions and corresponding range of the trigonometric and inverse trigonometric functions that will be used throughout the remainder of this course.

5.4.3 Solving equations involving trigonometric functions

We have already seen how to solve equations like $\tan(x) = -1$ for real numbers x by appealing to the unit circle and relying on the fact that the answers corresponded to a set of common angles listed in Tables 5.1 and 5.2. If, on the other hand, we had been asked to solve $\tan(x) = -2$, we would have been hard-pressed to do so. With the introduction of the inverse trigonometric functions, however,

Table 5.5: Domains and ranges of the trigonometric and inverse trigonometric functions.

Function	Domain	Range
$\sin(x)$	$[-\pi/2, \pi/2]$	$[-1, 1]$
$\cos(x)$	$[0, \pi]$	$[-1, 1]$
$\tan(x)$	$]-\pi/2, \pi/2[$	$]-\infty, +\infty[$
$\cot(x)$	$]0, \pi[$	$]-\infty, +\infty[$
$\arcsin(x)$	$[-1, 1]$	$[-\pi/2, \pi/2]$
$\arccos(x)$	$[-1, 1]$	$[0, \pi]$
$\arctan(x)$	$]-\infty, +\infty[$	$]-\pi/2, \pi/2[$
$\text{arccot}(x)$	$]-\infty, +\infty[$	$]0, \pi[$

we are now in a position to solve these equations. We will illustrate this functionality in the following example.

Example 5.21

Solve the following equations.

1. $\sin(x) = \dfrac{1}{3}$
2. $\tan(x) = -2$

Solution

1. If $\sin(x) = \frac{1}{3}$, then the terminal side of the angle corresponding to x, when plotted in standard position, intersects the unit circle at $y = \frac{1}{3}$. Geometrically, we see that this happens at two places: in Quadrant I and Quadrant II (Figure 5.27(a)). Since $\frac{1}{3}$ is not the sine of any of the common angles in Table 5.1, we use the arcsine functions to express our answers:

$$\sin(x) = \frac{1}{3} \quad \Leftrightarrow \quad \begin{cases} x = \arcsin\left(\frac{1}{3}\right) + 2k\pi, \text{ of} \\ x = \pi - \arcsin\left(\frac{1}{3}\right) + 2k\pi \end{cases}$$

   for integers k.

2. We may visualize the solutions to $\tan(x) = -2$ as angles. Since tangent is negative only in Quadrants II and IV, we focus our efforts there. Since -2 is not the tangent of any of the common angles listed in Table 5.2, we need to use the arctangent function to express our answers. The real number $t = \arctan(-2)$ satisfies $\tan(t) = -2$ and $-\frac{\pi}{2} < t < 0$, so we see that all solutions are of the form $x = t + \pi k = \arctan(-2) + \pi k$ for integers k.

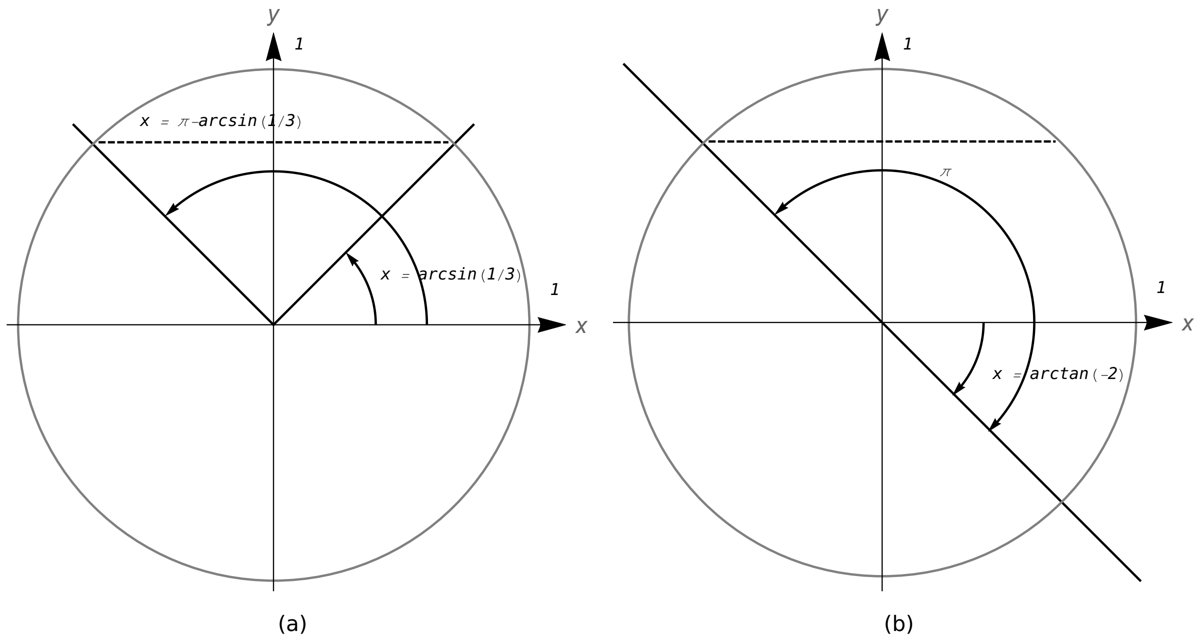


Figure 5.27: Solving $\sin(x)=\frac{1}{3}$ (a) and $\tan(x)=-2$ (b).

In Example 5.21 we solved some basic equations involving the trigonometric functions. With our comprehension of the unit circle and hence our understanding between the output of a trigonometric function and its input, we may infer the following strategies for solving basic trigonometric equations.

- To solve $\cos(x)=c$ or $\sin(x)=c$ for $-1\leq c\leq 1$, first solve for x in the interval $[0,2\pi[$, thereby not forgetting that opposite and supplementary angles are also solutions of $\cos(x)=c$ and $\sin(x)=c$, respectively. Finally, add integer multiples of the period 2π. If $c<-1$ or of $c>1$, there are no real solutions.
- To solve $\sec(x)=c$ or $\csc(x)=c$ for $c\leq -1$ or $c\geq 1$, convert to cosine or sine, respectively, and solve as above. If $-1<c<1$, there are no real solutions.
- To solve $\tan(x)=c$ for any real number c, first solve for x in the interval $\left]-\frac{\pi}{2},\frac{\pi}{2}\right[$ and add integer multiples of the period π.
- To solve $\cot(x)=c$ for $c\neq 0$, convert to tangent and solve as above. If $c=0$, the solution to $\cot(x)=0$ is $x=\frac{\pi}{2}+\pi k$ for integers k.

The question remains, however, how do we solve something like $\sin(3x)=\frac{1}{2}$? Since this equation has the form $\sin(u)=\frac{1}{2}$, we know the solutions take the form $u=\frac{\pi}{6}+2\pi k$ or $u=\frac{5\pi}{6}+2\pi k$ for integers k. Since the argument of sine here is $3x$, we have $3x=\frac{\pi}{6}+2\pi k$ or $3x=\frac{5\pi}{6}+2\pi k$ for integers k. To solve for x, we divide both sides of these equations by 3, and obtain $x=\frac{\pi}{18}+\frac{2\pi}{3}k$ or $x=\frac{5\pi}{18}+\frac{2\pi}{3}k$ for integers k. For what concerns equations involving two different trigonometric functions or equations containing the same trigonometric functions but with different arguments, we will need to use the identities we introduced in Section 5.3 identities and some algebra to manipulate the equation up to a point that we reach a solvable one.

Example 5.22

Solve the following equations and inequalities and list the solutions which lie in the interval $[0, 2\pi[$.

1. $\cos(2x) = -\dfrac{\sqrt{3}}{2}$
2. $\csc\left(\dfrac{1}{3}x - \pi\right) = \sqrt{2}$
3. $\sec^2(x) = 4$
4. $3\sin^3(x) = \sin^2(x)$
5. $\cos(2x) = 3\cos(x) - 2$
6. $\sin(2x) > \cos(x)$

Solution

1. From $\cos(2x) = -\frac{\sqrt{3}}{2}$, we immediately infer that the solutions are given by

$$\begin{array}{rrcl} & 2x = \frac{5\pi}{6} + 2\pi k & \vee & 2x = \frac{7\pi}{6} + 2\pi k \\ \Leftrightarrow & x = \frac{5\pi}{12} + \pi k & \vee & x = \frac{7\pi}{12} + \pi k \end{array}$$

for integers k. The solutions that lie in $[0, 2\pi[$ are $x = \frac{5\pi}{12}, \frac{7\pi}{12}, \frac{17\pi}{12}$ and $\frac{19\pi}{12}$.

2. Since this equation has the form $\csc(u) = \sqrt{2}$, we rewrite this as $\sin(u) = \frac{\sqrt{2}}{2}$ and we directly find $u = \frac{\pi}{4} + 2\pi k$ or $u = \frac{3\pi}{4} + 2\pi k$ for integers k. Since the argument of cosecant here is $\left(\frac{1}{3}x - \pi\right)$, we have

$$\begin{array}{rrcl} & \frac{1}{3}x - \pi = \dfrac{\pi}{4} + 2\pi k & \vee & \dfrac{1}{3}x - \pi = \dfrac{3\pi}{4} + 2\pi k \\ \Leftrightarrow & x = \frac{3\pi}{4} + 6\pi k + 3\pi = \dfrac{15\pi}{4} + 6\pi k & \vee & x = \frac{9\pi}{4} + 6\pi k + 3\pi = \dfrac{21\pi}{4} + 6\pi k \end{array}$$

None of the solutions lie in $[0, 2\pi[$.

3. We start by extracting square roots to get $\sec(x) = \pm 2$. Converting to cosines, we have $\cos(x) = \pm\frac{1}{2}$. For $\cos(x) = \frac{1}{2}$, we get $x = \frac{\pi}{3} + 2\pi k$ or $x = -\frac{\pi}{3} + 2\pi k$ for integers k. For $\cos(x) = -\frac{1}{2}$, we get $x = \frac{2\pi}{3} + 2\pi k$ or $x = -\frac{2\pi}{3} + 2\pi k$ for integers k. These solutions can be combined as $x = \frac{\pi}{3} + \pi k$ and $x = \frac{2\pi}{3} + \pi k$ for integers k. The solutions that lie in $[0, 2\pi[$ are $x = \frac{\pi}{3}, \frac{2\pi}{3}, \frac{4\pi}{3}$ and $\frac{5\pi}{3}$.

4. We may not divide both sides of $3\sin^3(x) = \sin^2(x)$ by $\sin^2(x)$. Instead we gather all of the terms to one side of the equation and factor.

$$\begin{array}{rrcl} & 3\sin^3(x) & = & \sin^2(x) \\ \Leftrightarrow & 3\sin^3(x) - \sin^2(x) & = & 0 \\ \Leftrightarrow & \sin^2(x)(3\sin(x) - 1) & = & 0 \end{array}$$

So, we get $\sin^2(x) = 0$ or $3\sin(x) - 1 = 0$, from which we deduce that $\sin(x) = 0$ or $\sin(x) = \frac{1}{3}$. The solution to the first equation is $x = \pi k$, with $x = 0$ and $x = \pi$ being the two solutions which lie in $[0, 2\pi[$. To solve $\sin(x) = \frac{1}{3}$, we use the arcsine function to get $x = \arcsin\left(\frac{1}{3}\right) + 2\pi k$ or $x = \pi - \arcsin\left(\frac{1}{3}\right) + 2\pi k$ for integers k. We find the two solutions here which lie in $[0, 2\pi[$, manely $x = \arcsin\left(\frac{1}{3}\right)$ and $x = \pi - \arcsin\left(\frac{1}{3}\right)$.

5. We have cosine on both sides, but the arguments differ. Using the identity $\cos(2x) = 2\cos^2(x) - 1$, we obtain a quadratic in disguise, which can be solved:

$$
\begin{array}{rrcll}
 & \cos(2x) & = & 3\cos(x)-2 & \\
\Leftrightarrow & 2\cos^2(x)-1 & = & 3\cos(x)-2 & \\
\Leftrightarrow & 2\cos^2(x)-3\cos(x)+1 & = & 0 & \\
\Leftrightarrow & 2u^2-3u+1 & = & 0 & \text{(Let } u=\cos(x).\text{)} \\
\Leftrightarrow & (2u-1)(u-1) & = & 0\,. &
\end{array}
$$

This gives $u = \frac{1}{2}$ or $u = 1$. Since $u = \cos(x)$, we get $\cos(x) = \frac{1}{2}$ or $\cos(x) = 1$. Solving $\cos(x) = \frac{1}{2}$, we get $x = \frac{\pi}{3} + 2\pi k$ or $x = -\frac{\pi}{3} + 2\pi k$ for integers k. From $\cos(x) = 1$, we get $x = 2\pi k$ for integers k. The answers which lie in $[0, 2\pi[$ are $x = 0$, $\frac{\pi}{3}$, and $\frac{5\pi}{3}$.

6. We first rewrite $\sin(2x) > \cos(x)$ as $\sin(2x) - \cos(x) > 0$ and let $f(x) = \sin(2x) - \cos(x)$. Our original inequality is thus equivalent to $f(x) > 0$. The domain of f is all real numbers, so we can advance to finding the zeros of f. Setting $f(x) = 0$ yields

$$\sin(2x)-\cos(x)=0 \quad \Leftrightarrow \quad 2\sin(x)\cos(x)-\cos(x)=0 \quad \Leftrightarrow \quad \cos(x)(2\sin(x)-1)=0\,.$$

From $\cos(x) = 0$, we get $x = \frac{\pi}{2} + \pi k$ for integers k of which only $x = \frac{\pi}{2}$ and $x = \frac{3\pi}{2}$ lie in $[0, 2\pi[$. For $\sin(x) = \frac{1}{2}$ we get $x = \frac{\pi}{6} + 2\pi k$ or $x = \frac{5\pi}{6} + 2\pi k$ for integers k. Of those, only $x = \frac{\pi}{6}$ and $x = \frac{5\pi}{6}$ lie in $[0, 2\pi[$. Next, we choose our test values. For $x = 0$ we find $f(0) = -1$; when $x = \frac{\pi}{4}$ we get $f\left(\frac{\pi}{4}\right) = 1 - \frac{\sqrt{2}}{2} = \frac{2-\sqrt{2}}{2}$; for $x = \frac{3\pi}{4}$ we get $f\left(\frac{3\pi}{4}\right) = -1 + \frac{\sqrt{2}}{2} = \frac{\sqrt{2}-2}{2}$; when $x = \pi$ we have $f(\pi) = 1$, and lastly, for $x = \frac{7\pi}{4}$ we get $f\left(\frac{7\pi}{4}\right) = -1 - \frac{\sqrt{2}}{2} = \frac{-2-\sqrt{2}}{2}$. The leads to the following sign diagram:

$$
\begin{array}{l|ccccccccc}
\sin(2x)-\cos(x) & - & | & + & | & - & | & + & | & - \\
\hline
 & & \frac{\pi}{6} & & \frac{\pi}{2} & & \frac{5\pi}{6} & & \frac{3\pi}{2} &
\end{array}
$$

We see $f(x) > 0$ on $\left]\frac{\pi}{6}, \frac{\pi}{2}\right[\cup \left]\frac{5\pi}{6}, \frac{3\pi}{2}\right[$, so this is our answer.

Our next example puts solving equations and inequalities to good use, namely for finding domains of functions.

Example 5.23

Express the domain of the following functions using extended interval notation.

1. $f(x) = \csc\left(2x + \dfrac{\pi}{3}\right)$
2. $g(x) = \sqrt{1 - \cot(x)}$

Solution

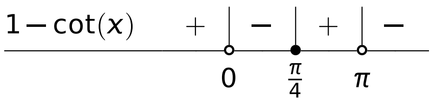
1. We rewrite f in terms of sine as $f(x) = \frac{1}{\sin\left(2x+\frac{\pi}{3}\right)}$. Since the sine function is defined everywhere, our only concern comes from zeros in the denominator. Solving $\sin\left(2x + \frac{\pi}{3}\right) = 0$, we get for integers k:

$$
\begin{array}{rrcl}
 & 2x+\dfrac{\pi}{3} & = & k\pi \\
\Leftrightarrow & x & = & -\dfrac{\pi}{6}+\dfrac{\pi}{2}k\,.
\end{array}
$$

In set-builder notation, our domain is $\left\{x : x \neq -\frac{\pi}{6}+\frac{\pi}{2}k\,, \forall k \in \mathbb{Z}\right\}$. If we now let x_k denote the kth number excluded from the domain, we have $x_k = -\frac{\pi}{6}+\frac{\pi}{2}k = \frac{(3k-1)\pi}{6}$ for integers k. The intervals which comprise the domain are of the form $]x_k, x_{k+1}[= \left]\frac{(3k-1)\pi}{6}, \frac{(3k+2)\pi}{6}\right[$ as k runs through the integers. Using extended interval notation, we have that the domain is

$$\bigcup_{k=-\infty}^{+\infty} \left]\frac{(3k-1)\pi}{6}, \frac{(3k+2)\pi}{6}\right[.$$

2. We first note that, due to the presence of the $\cot(x)$ term, $x \neq \pi k$ for integers k. Next, we recall that for the square root to be defined, we need $1-\cot(x) \geq 0$. Our strategy is to solve this inequality over $]0,\pi[$, i.e. interval which generates a fundamental cycle of cotangent, and then add integer multiples of the period, in this case, π. We let $r(x) = 1-\cot(x)$ and set about making a sign diagram for r over the interval $]0,\pi[$ to find where $r(x) \geq 0$. We note that r is undefined for $x = \pi k$ for integers k, in particular, at the endpoints of our interval $x = 0$ and $x = \pi$. Next, we look for the zeros of r. Solving $r(x) = 0$, we get $\cot(x) = 1$ or $x = \frac{\pi}{4}+\pi k$ for integers k and only one of these, $x = \frac{\pi}{4}$, lies in $]0,\pi[$. Choosing the test values $x = \frac{\pi}{6}$ and $x = \frac{\pi}{2}$, we get $r\left(\frac{\pi}{6}\right) = 1-\sqrt{3}$, and $r\left(\frac{\pi}{2}\right) = 1$, and the following sign diagram.



We find $g(x) \geq 0$ on $\left[\frac{\pi}{4},\pi\right[$. Adding multiples of the period we get our solution to consist of the intervals

$$\left[\frac{\pi}{4}+\pi k, \pi+\pi k\right[= \left[\frac{(4k+1)\pi}{4}, (k+1)\pi\right[,$$

or more briefly using extended interval notation:

$$\bigcup_{k=-\infty}^{\infty} \left[\frac{(4k+1)\pi}{4}, (k+1)\pi\right[.$$

We close this section with an example that demonstrates how to solve equations and inequalities involving inverse trigonometric functions.

Example 5.24

Solve the following equations and inequalities analytically.

1. $\arcsin(2x) = \dfrac{\pi}{3}$
2. $4\arctan^2(x) - 3\pi\arctan(x) - \pi^2 = 0$
3. $\pi^2 - 4\arccos^2(x) < 0$

Solution

1. We first note that $\frac{\pi}{3}$ is in the range of the arcsine function, so a solution exists! Next, we

exploit the inverse property of sine and arcsine

$$\begin{aligned} \arcsin(2x) &= \frac{\pi}{3} \\ \Rightarrow \quad \sin(\arcsin(2x)) &= \sin\left(\frac{\pi}{3}\right) \\ \Leftrightarrow \quad 2x &= \frac{\sqrt{3}}{2} \qquad \text{(Since } \sin(\arcsin(u)) = u.\text{)} \\ \Leftrightarrow \quad x &= \frac{\sqrt{3}}{4}. \end{aligned}$$

2. With the presence of both $\arctan^2(x)$ ($= (\arctan(x))^2$) and $\arctan(x)$, we substitute $u = \arctan(x)$. The equation becomes

$$\begin{aligned} 4u^2 - 3\pi u - \pi^2 = 0 \quad &\Leftrightarrow \quad (4u+\pi)(u-\pi) = 0 \\ &\Leftrightarrow \quad u = \arctan(x) = -\frac{\pi}{4} \quad \vee \quad u = \arctan(x) = \pi. \end{aligned}$$

Since $-\frac{\pi}{4}$ is in the range of arctangent, but π is not, we only get solutions from the first equation. To solve for x, we may write

$$\begin{aligned} \arctan(x) &= -\frac{\pi}{4} \\ \Rightarrow \quad \tan(\arctan(x)) &= \tan\left(-\frac{\pi}{4}\right) \\ \Leftrightarrow \quad x &= -1. \qquad \text{(Since } \tan(\arctan(u)) = u.\text{)} \end{aligned}$$

3. Since the inverse trigonometric functions are continuous on their domains, we can solve inequalities featuring these functions using sign diagrams. Since all of the nonzero terms of $\pi^2 - 4\arccos^2(x) < 0$ are on one side of the inequality, we let $f(x) = \pi^2 - 4\arccos^2(x)$ and note that the domain of f is limited by the $\arccos(x)$ to $[-1,1]$. Next, we find the zeros of f by setting $f(x) = \pi^2 - 4\arccos^2(x) = 0$. We get $\arccos(x) = \pm\frac{\pi}{2}$, and since the range of arccosine is $[0,\pi]$, we focus our attention on $\arccos(x) = \frac{\pi}{2}$. Consequently, we get $x = \cos\left(\frac{\pi}{2}\right) = 0$ as our only zero. Hence, we have two test intervals, $[-1,0[$ and $]0,1]$. Choosing test values $x = \pm 1$, we get $f(-1) = -3\pi^2 < 0$ and $f(1) = \pi^2 > 0$, which leads us to conclude that $f(x) < 0$ on $[-1,0[$ and $f(x) > 0$ on $]0,1]$. So, our answer is $[-1,0[$.

5.5 Conic sections continued

5.5.1 Translation of axes

In Section 4.4, we introduced the standard equations of conic sections. The careful reader might have noticed that, for instance, the standard equation of a circle R with radius r and centre in (x_0, y_0) (Equation (4.4)) dictates a shift of the circle

$$x^2 + y^2 = r^2$$

with centre in (0,0) x_0 units to the right along the x-axis and y_0 units up along the y-axis. In other words, the studied circle R is just a translated version of a circle with the same radius r but centred at the origin. Besides, upon expanding Equation (4.4), i.e.

$$x^2 - 2x_0x + x_0^2 + y^2 - 2yy_0 + y_0^2 = r^2$$

we see that linear terms in x and y appear in the left-hand side of the equation.

These two facts hint that we might in general be able to get rid of the linear terms in the quadratic equation

$$ax^2 + cy^2 + dx + ey + f = 0 \tag{5.29}$$

by introducing a change of variables; that is by letting

$$\begin{cases} \tilde{x} &= x - x_0 \\ \tilde{y} &= y - y_0\,, \end{cases}$$

where x_0 and y_0 should be determined in such a way that the squares in Equation (5.29) can be completed, and hence the linear terms vanish. Basically, this change of variables is equivalent to a shift of the original xy-Cartesian coordinate system to an $\tilde{x}\tilde{y}$-Cartesian coordinate system in which the $\tilde{x}$-axis is parallel to the x-axis and x_0 units away, and the $\tilde{y}$- axis is parallel to the y-axis and y_0 units away. This means that the origin $\tilde{o}$ of the new coordinate system has coordinates (x_0, y_0) in the original system (Figure 5.28).

In order to find a suitable change of variables for Equation (5.29), let us first rewrite this equation as

$$a\left(x^2 + \frac{d}{a}x\right) + c\left(y^2 + \frac{e}{c}y\right) + f = 0\,,$$

so that the squares can be completed

$$a\left(\left(x + \frac{d}{2a}\right)^2 - \frac{d^2}{4a^2}\right) + c\left(\left(y + \frac{e}{2c}\right)^2 - \frac{e^2}{4c^2}\right) + f = 0\,.$$

Working out the outer parentheses gives

$$a\left(x + \frac{d}{2a}\right)^2 + c\left(y + \frac{e}{2c}\right)^2 - \frac{d^2}{4a} - \frac{e^2}{4c} + f = 0\,,$$

which, by dividing it by ac and introducing the following change of variables

$$\begin{cases} \tilde{x} &= x + \dfrac{d}{2a} \\ \tilde{y} &= y + \dfrac{e}{2c}\,, \end{cases}$$

can be recast in the standard form of a conic section:

$$\frac{\tilde{x}^2}{c} + \frac{\tilde{y}^2}{a} = \frac{d^2}{4a^2c} + \frac{e^2}{4c^2a} - \frac{f}{ac}\,. \tag{5.30}$$

Example 5.25

Rewrite the following quadratic equation in standard form by eliminating the linear terms in x and

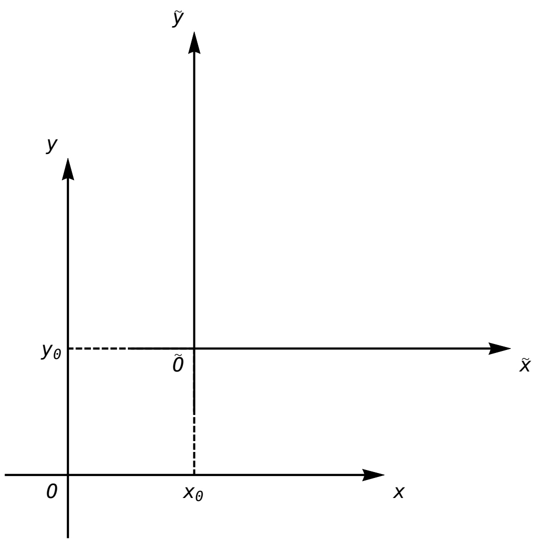


Figure 5.28: Shift of the xy- Cartesian coordinate system to a system $\tilde{x}\tilde{y}$ with origin $\tilde{o}$ with coordinates (x_0, y_0) in the xy-system.

y.

$$3x^2 - 7y^2 + 12\sqrt{2}x - 28\sqrt{2}y - 30 = 0$$

Then, draw the corresponding conic section.

Solution

The linear terms can be eliminated by completing the squares in x and y as follows.

$$\begin{aligned}
& 3x^2 - 7y^2 + 12\sqrt{2}x - 28\sqrt{2}y - 30 = 0 \\
\Leftrightarrow\quad & 3(x^2 + 4\sqrt{2}x) - 7(y^2 + 4\sqrt{2}y) - 30 = 0 \\
\Leftrightarrow\quad & 3(x^2 + 4\sqrt{2}x + 8) - 7(y^2 + 4\sqrt{2}y + 8) = 30 + 24 - 56 \\
\Leftrightarrow\quad & 3(x + 2\sqrt{2})^2 - 7(y + 2\sqrt{2})^2 = -2
\end{aligned}$$

Now, let

$$\begin{cases} \tilde{x} = x + 2\sqrt{2} \\ \tilde{y} = y + 2\sqrt{2} \end{cases}$$

then we get

$$3\tilde{x}^2 - 7\tilde{y}^2 = -2\,,$$

or equivalently

$$-\frac{3\tilde{x}^2}{2} + \frac{7\tilde{y}^2}{2} = 1\,.$$

This is the standard equation of a vertical hyperbola centred in $(-2\sqrt{2}, -2\sqrt{2})$.

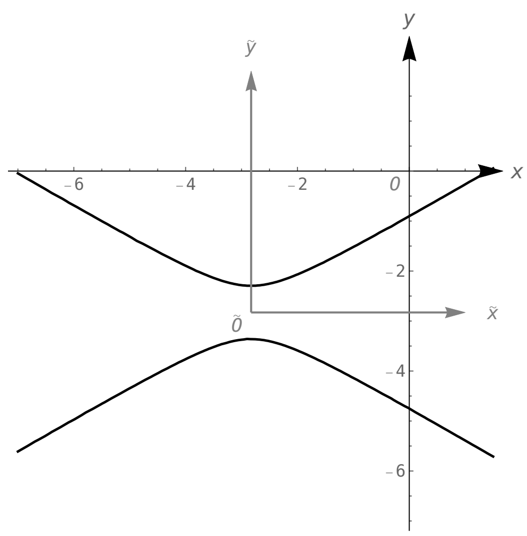


Figure 5.29: Graph of $3x^2 - 7y^2 + 12\sqrt{2}x - 28\sqrt{2}y - 30 = 0$, together with the original and translated coordinate axes.

5.5.2 Rotation of axes

We have just seen that a translation of axes can be used to eliminate linear terms in x and/or y in quadratic, but what about the term in xy that might appear in the general quadratic equation

$$ax^2 + bxy + cy^2 + dx + ey + f = 0\,? \tag{5.31}$$

We shall see that the term bxy can be eliminated by rotating the coordinate axes. For that purpose, we need the trigonometric fuctions that were introduced in Section 5.3.

In Figure 5.30(a) the x- and y-axes have been rotated about the origin through an acute angle θ to produce the $\tilde{x}$- and $\tilde{y}$-axes. Thus, a given point P has coordinates (x, y) in the first coordinate system and $(\tilde{x}, \tilde{y})$ in the new coordinate system. To see how $\tilde{x}$ and $\tilde{y}$ are related to x and y we observe from Figure 5.30(b) that

$$\begin{cases} \tilde{x} &= r\cos(\phi) \\ x &= r\cos(\theta+\phi)\,, \end{cases}$$

and likewise

$$\begin{cases} \tilde{y} &= r\sin(\phi) \\ y &= r\sin(\theta+\phi)\,, \end{cases}$$

The addition formula for the cosine then gives

$$\begin{aligned} x &= r\cos(\theta+\phi) = r\left(\cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi)\right) \\ &= \left(r\cos(\phi)\right)\cos(\theta) - \left(r\sin(\phi)\right)\sin(\theta) \\ &= \tilde{x}\cos(\theta) - \tilde{y}\sin(\theta)\,. \end{aligned}$$

A similar computation gives y in terms of $\tilde{x}$ and $\tilde{y}$, so that we arrive at the following formulas relating x and y and $\tilde{x}$ and $\tilde{y}$:

$$\begin{cases} x &= \tilde{x}\cos(\theta) - \tilde{y}\sin(\theta) \\ y &= \tilde{x}\sin(\theta) + \tilde{y}\cos(\theta)\,. \end{cases} \tag{5.32}$$

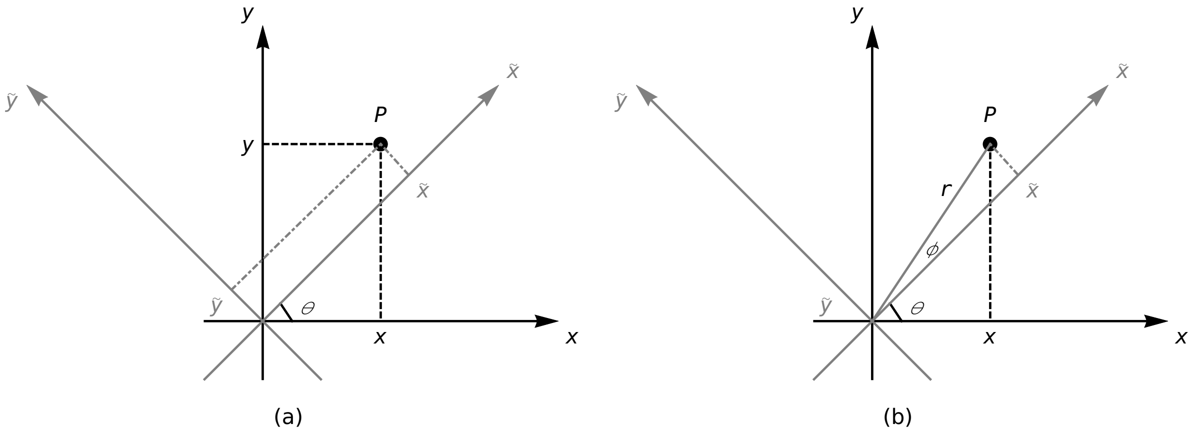


Figure 5.30: Rotation of axes.

By solving these equations for $\tilde{x}$ and $\tilde{y}$, we obtain formulas to determine the coordinates $\tilde{x}$ and $\tilde{y}$ of a point P that has coordinates x and y in the original Cartesian coordinate system:

$$\begin{cases} \tilde{x} & = & x\cos(\theta) + y\sin(\theta) \\ \tilde{y} & = & -x\sin(\theta) + y\cos(\theta)\,. \end{cases} \tag{5.33}$$

Now let us try to determine an angle θ such that the term bxy in Equation (5.31) disappears when the axes are rotated through the angle θ. If we substitute the expressions from Equation (5.32) in Equation (5.31), we get

$$\begin{aligned} a\big(\tilde{x}\cos(\theta) - \tilde{y}\sin(\theta)\big)^2 + b\big(\tilde{x}\cos(\theta) - \tilde{y}\sin(\theta)\big)\big(\tilde{x}\sin(\theta) + \tilde{y}\cos(\theta)\big) + c\big(\tilde{x}\sin(\theta) + \tilde{y}\cos(\theta)\big)^2 \\ + d\big(\tilde{x}\cos(\theta) - \tilde{y}\sin(\theta)\big) + e\big(\tilde{x}\sin(\theta) + \tilde{y}\cos(\theta)\big) + f = 0 \end{aligned} \tag{5.34}$$

Expanding and collecting terms, we obtain an equation of the form

$$\tilde{a}\tilde{x}^2 + \tilde{b}\tilde{x}\tilde{y} + \tilde{c}\tilde{y}^2 + \tilde{d}\tilde{x} + \tilde{e}\tilde{y} + f = 0\,,$$

where the coefficient $\tilde{b}$ of $\tilde{x}\tilde{y}$ is

$$\begin{aligned} \tilde{b} &= 2\,(c-a)\sin(\theta)\cos(\theta) + b\Big(\cos^2(\theta) - \sin^2(\theta)\Big) \\ &= (c-a)\sin(2\theta) + b\cos(2\theta)\,. \end{aligned}$$

To eliminate the term $\tilde{x}\tilde{y}$ we choose θ in such a way that $\tilde{b} = 0$, i.e.

$$(a-c)\sin(2\theta) = b\cos(2\theta)\,,$$

or

$$\cot(2\theta) = \frac{a-c}{b}\,. \tag{5.35}$$

Note that, after choosing θ in this way, we end up with an ellipse if $\tilde{a}$ and $\tilde{b}$ have the same sign, whereas we obtain a hyperbola if these coefficients have a different sign. Besides, when one of them is zero, we have a parabola.

Example 5.26

Identify and sketch the graph of

$$73x^2+72xy+52y^2+30x-40y-75=0\,. \tag{5.36}$$

Solution

This equation is in the form of Equation (5.31) with $a=73$, $b=72$ and $c=52$. Consequently, by choosing

$$\cot(2\theta)=\frac{a-c}{b}=\frac{73-52}{72}=\frac{7}{24}$$

we can eliminate the term $\widetilde{x}\widetilde{y}$ in the rotated counterpart of Equation (5.36). Using the Pythagorean theorem (Figure 5.31), we observe that

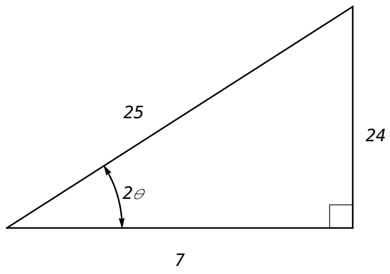


Figure 5.31: Using the Pythagorean theorem to compute 2θ.

$$\cos(2\theta)=\frac{7}{25}\,.$$

This observation allow us to compute $\cos(\theta)$ and $\sin(\theta)$ using the half-angle formulas:

$$\begin{aligned}\cos(\theta) &= \sqrt{\frac{1+\cos(2\theta)}{2}}=\sqrt{\frac{1+\frac{7}{25}}{2}}=\frac{4}{5}\\ \sin(\theta) &= \sqrt{\frac{1-\cos(2\theta)}{2}}=\sqrt{\frac{1-\frac{7}{25}}{2}}=\frac{3}{5}\,.\end{aligned}$$

Consequently, the rotation equations become

$$\begin{cases} x &= \frac{4}{5}\widetilde{x}-\frac{3}{5}\widetilde{y}\\ y &= \frac{3}{5}\widetilde{x}+\frac{4}{5}\widetilde{y}\,,\end{cases}$$

and substituting these in Equation (5.36), yields

$$73\left(\frac{4}{5}\widetilde{x}-\frac{3}{5}\widetilde{y}\right)^2+72\left(\frac{4}{5}\widetilde{x}-\frac{3}{5}\widetilde{y}\right)\left(\frac{3}{5}\widetilde{x}+\frac{4}{5}\widetilde{y}\right)+52\left(\frac{3}{5}\widetilde{x}+\frac{4}{5}\widetilde{y}\right)^2$$
$$+30\left(\frac{4}{5}\widetilde{x}-\frac{3}{5}\widetilde{y}\right)-40\left(\frac{3}{5}\widetilde{x}+\frac{4}{5}\widetilde{y}\right)-75=0\,.$$

This simplifies to

$$4\tilde{x}^2 + \tilde{y}^2 - 2\tilde{y} = 3\,.$$

Completing the square in $\tilde{y}$ leads to:

$$\tilde{x}^2 + \frac{\left(\tilde{y}-1\right)^2}{4} = 1\,,$$

and we recognize this as being an ellipse with centre in $(0,1)$ with respect to the rotated coordinate axes $\tilde{x}$ and $\tilde{y}$ (Figure 5.32).

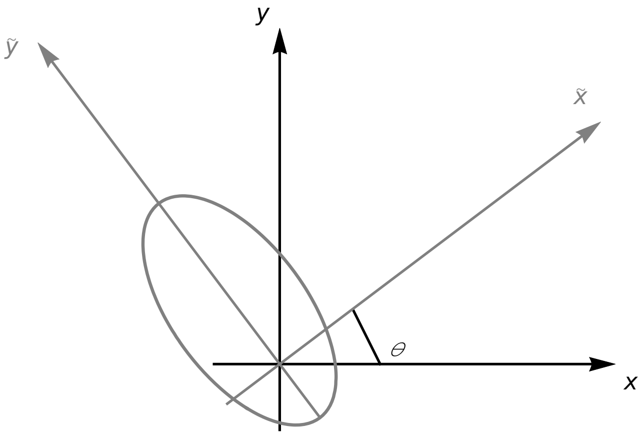


Figure 5.32: Graph of $73x^2 + 72xy + 52y^2 + 30x - 40y - 75 = 0$, together with the original and rotated coordinate axes.

Actually, we even do not have to translate and/or rotate the coordinate axes when we are merely interested in which conic section a given quadratic equations represents by relying on the following theorem.

Theorem 5.15 (Classification of conic sections)
Suppose the equation $ax^2 + bxy + cy^2 + dx + ey + f = 0$ describes a non-degenerate conic section.

- *If $b^2 - 4ac > 0$ then the graph of the equation is a hyperbola.*
- *If $b^2 - 4ac = 0$ then the graph of the equation is a parabola.*
- *If $b^2 - 4ac < 0$ then the graph of the equation is an ellipse or circle.*

As you may expect, the quantity $b^2 - 4ac$ mentioned in Theorem 5.15 is called the **discriminant** (*discriminant*)

5.6 Exercises

Exponential and logarithmic functions

Assignment 5.1 — Simplify the expressions below.

(a) $\log_4\left(\frac{1}{8}\right)$

(b) $\log_{a^2}\left(a^3\right)$

(c) $e^{(3\ln(9))/2}$

(d) $\log_{\frac{1}{3}}\left(3^{2x}\right)$

(e) $\log_x\left(x\left(\log_y\left(y^2\right)\right)\right)$

(f) $\log_{15}(75)+\log_{15}(3)$

(g) $2\log_3(12)-4\log_3(6)$

(h) $2\ln(x)+5\ln(x-2)$

Assignment 5.2 — Prove the equalities below.

(a) $\log_a(b)\cdot\log_b(c)\cdot\log_c(a)=1$

(b) $\log_b(a)\cdot\log_c(b)=\log_c(a)$

(c) $\dfrac{1}{\log_a(x)}+\dfrac{1}{\log_b(x)}=\dfrac{1}{\log_{ab}(x)}$

Assignment 5.3 — Solve the equations below

(a) $3^{2x+1}=\sqrt[3]{3}$

(b) $4^{\frac{1}{x}}\cdot 16^{\frac{1}{x+2}}=64^{\frac{1}{x+1}}$

(c) $\log_4(x+4)-2\log_4(x+1)=\dfrac{1}{2}$

(d) $2\log_3(x)+\log_9(x)=10$

(e) $\dfrac{1}{2^x}=\dfrac{5}{8^{x+3}}$

(f) $(\ln x)^2+\ln(x^2)=\ln(x)$

(g) $4^{x-2}\cdot 8^{3x-1}=\sqrt{2}$

(h) $5^{2x-1}-3\cdot 5^{x+2}+6250=0$

(i) $\ln(x)-5\ln(2)=\ln(3x+84)+\ln(12)$

(j) $8^x+4^x=5\cdot 2^{x-4}$

(k) $5^{2x}+5^3=5^{x+2}+5^{x+1}$

(l) $4\cdot 8^{x-1}+1=4^x+2^{x-1}$

(m) $\dfrac{\log_4(x)}{\log_4(3)}\cdot\log_3\left(\dfrac{x}{9}\right)=3\cdot 27^{\log_3(2)}$

(n) $\log_3(x+4)+\log_3(x-2)=2\log_3(x)$

(o) $\log_2(3)\cdot\log_3(5)\cdot\log_5(7)\cdot\log_7(x)=4$

Assignment 5.4 — Solve the inequalities below.

(a) $\ln(x^2-2)\le\ln(x)$

(b) $9^{-x}<\dfrac{2+3^{x+1}}{3^x}$

(c) $\log\left(\dfrac{x(10-x)}{16}\right)<0$

(d) $\log_{\frac{1}{3}}(4x)<\log_{\frac{1}{3}}(x-1)-2$

(e) $x\ln(x)-x>0$

(f) $2^{3x-x^2}>\left(\dfrac{1}{8}\right)^{1-x}$

Assignment 5.5 — Determine the domain and intersections with the x- and y-axis of the (logarithmic) functions below.

(a) $f(x) = \log\left(\dfrac{x+2}{x^2-1}\right)$

(b) $f(x) = \ln(4x-20) + \ln\left(x^2+9x+18\right)$

(c) $f(x) = \ln\left(\sqrt{x-4}-3\right)$

(d) $f(x) = \dfrac{\sqrt{-1-x}}{\log_{\frac{1}{2}}(x)}$

(e) $f(x) = \log_2\left(\dfrac{1-x}{1+x}\right)$

(f) $f(x) = \log_6\left(\log_{\frac{1}{5}}(x)\right)$

Assignment 5.6 — The pressure p of a gas with molar mass M at height h is

$$p = p_0\, e^{-Mgh/RT},$$

where p_0 is the pressure at sea level and g and R are constants. Write h as a function of the other variables.

Assignment 5.7 — A bacterial culture grows at a rate proportional to the number of cells present. Suppose the culture initially contains 500 cells, while 800 cells are present still after 24 hours. How many more cells will there be another 12 hours later?

Assignment 5.8 — A radioactive material has a half-life of 1200 years. What percentage of the initial radioactivity is left after 10 years? How many years does it take to decrease the level of radioactivity by 10%?

Assignment 5.9 — A scientist inoculates a bacterial culture with 20 bacteria. The number of bacteria increases by a constant factor every hour such that after 1 day the number of bacteria equals 220.

(a) Determine how many bacteria there will be after 1 week.

(b) How long does it take to reach 5 million bacteria?

Assignment 5.10 — A thermometer is removed from an oven at 72°C and placed in a room at 20°C . After 1 minute, the thermometer gives 48°C. What temperature will the thermometer indicate after 5 minutes?

Assignment 5.11 — An object is placed in a freezer of −5°C. In 40 minutes, the object cools from 45°C to 20°C. How much longer is needed for the object to cool to 0°C?

Trigonometric functions

Assignment 5.12 — Determine the trigonometric numbers below.

(a) $\cos\left(\frac{3\pi}{4}\right)$

(b) $\tan\left(-\frac{3\pi}{4}\right)$

(c) $\sin\left(\frac{2\pi}{3}\right)$

(d) $\sin\left(\frac{7\pi}{12}\right)$

(e) $\cos\left(\frac{5\pi}{12}\right)$

(f) $\sin\left(\frac{11\pi}{12}\right)$

(g) $\cot\left(\frac{13\pi}{12}\right)$

(h) $\sec\left(-\frac{\pi}{12}\right)$

(i) $\csc\left(-\frac{\pi}{3}\right)$

(j) $\cot\left(\frac{7\pi}{6}\right)$

Assignment 5.13 — Below, a list of trigonometric numbers corresponding to an angle θ is given. Determine the other trigonometric numbers.

(a) $\sin(\theta) = \frac{3}{5}$ met $\theta \in \left[\frac{\pi}{2}, \pi\right]$

(b) $\tan(\theta) = 2$ met $\theta \in \left[0, \frac{\pi}{2}\right[$

(c) $\sec(\theta) = 3$ met $\theta \in \left]-\frac{\pi}{2}, 0\right]$

(d) $\cos(\theta) = -\frac{5}{13}$ met $\theta \in \left[\frac{\pi}{2}, \pi\right]$

(e) $\csc(\theta) = -2$ met $\theta \in \left]\pi, \frac{3\pi}{2}\right]$

(f) $\tan(\theta) = \frac{1}{2}$ met $\theta \in \left[\pi, \frac{3\pi}{2}\right[$

Assignment 5.14 — Prove the identities below.

(a) $\cos^4(x) - \sin^4(x) = \cos(2x)$

(b) $\dfrac{1-\cos(x)}{\sin(x)} = \dfrac{\sin(x)}{1+\cos(x)} = \tan\left(\dfrac{x}{2}\right)$

(c) $\dfrac{1-\cos(x)}{1+\cos(x)} = \tan^2\left(\dfrac{x}{2}\right)$

(d) $\dfrac{\cos(x)-\sin(x)}{\cos(x)+\sin(x)} = \sec(2x) - \tan(2x)$

(e) $\dfrac{\tan^2(x)-\sin^2(x)}{1-\sin^2(x)} = \tan^4(x)$

(f) $\sin(x+y)\sin(x-y) = \cos^2(y) - \cos^2(x)$

(g) $\cos(x+y)\cos(x) + \sin(x+y)\sin(x) = \cos(y)$

(h) $\dfrac{\sin(x+y)}{\sin(x-y)} = \dfrac{\tan(x)+\tan(y)}{\tan(x)-\tan(y)}$

(i) $\sin^2(2x) - \sin^2(x) = \sin(3x)\sin(x)$

Assignment 5.15 — Consider the general sinusoids below. Determine for each them the amplitude, period, phase shift, vertical shift, domain, image, and zeros.

(a) $f(x) = 3\sin\left(\frac{x}{2\pi}\right)$

(b) $f(x) = \frac{2}{3}\sin\left(\frac{2}{3}\left(x - \frac{\pi}{4}\right)\right) - 11$

(c) $f(x) = \sin\left(10\pi\left(x + \frac{1}{2}\right)\right) + 3$

(d) $f(x) = 2\sin(3x-2)+1$

Assignment 5.16 — Solve the trigonometric equations below.

(a) $\cos(3x) = \sin(7x)$

(b) $\sin(x) - \sin^3(x) = -\cos^2(x)$

(c) $\sin^2(x) + 3\sin(x)\cos(x) = 1$

(d) $\dfrac{1+\sin(x)}{1-\sin(x)} = 2$

(e) $\sin(x)(\tan(x)+\cot(x)) = 2$

(f) $\tan(2x) = 5\tan(x)$

(g) $\cos(3x) = \cos(2x) - \cos(x)$

(h) $2\tan^2(x) + 6 = 5\sec^2(x)$

(i) $2\cos(x)\cos(3x) = -1$

(j) $\sin^3(x) + \cos^3(x)$
$= \sin(x)\cos(x)(\sin(x)+\cos(x))$

Assignment 5.17 — Solve the trigonometric equations below.

(a) $\tan(2x) < \frac{1}{3}$

(b) $2\sin\left(2x - \frac{\pi}{3}\right) + 1 < 0$

(c) $4\cos\left(3\left(x + \frac{\pi}{5}\right)\right) + 2 > 4$

(d) $\tan\left(2x + \frac{\pi}{6}\right) < \sqrt{3}$

(e) $2\cos^2(2x) + (\sqrt{3}+2)\cos(2x) + \sqrt{3} < 0$

(f) $\dfrac{\sin(x)}{2\sin(x)-1} > \dfrac{1-\sin(x)}{4\sin^2(x)-1}$

Inverse trigonometric functions

Assignment 5.18 — Determine the domain of the functions below.

(a) $\arctan(x^3+1)$

(b) $\arcsin\left(\frac{1}{x}\right)$

(c) $\arccos(x^2-3)$

(d) $\arccos\left(\frac{6}{\pi}\arcsin(x)\right)$

(e) $\arcsin\left(\frac{4}{\pi}\arccos(2x)\right)$

Assignment 5.19 — Calculate the values below.

(a) $\arcsin\left(-\frac{\sqrt{2}}{2}\right)$

(b) $\arctan\left(\frac{\sqrt{3}}{3}\right)$

(c) $\operatorname{arccot}\left(\sqrt{3}\right)$

(d) $\sin\left(\arccos\left(\frac{\sqrt{3}}{2}\right)\right)$

(e) $\cot\left(\arctan(-1)\right)$

(f) $\sin\left(2\arctan\left(\frac{1}{\sqrt{5}}\right)\right)$

(g) $\cos\left(2\operatorname{arccot}\left(\sqrt{7}\right)\right)$

(h) $\tan\left(4\arcsin\left(\frac{1}{3}\right)\right)$

(i) $\cot\left(3\arccos\left(-\frac{1}{\sqrt{2}}\right)\right)$

(j) $\cos\left(\arcsin\left(\frac{\sqrt{2}}{2}\right)\right)$

(k) $\tan\left(\arccos\left(-\frac{\sqrt{2}}{2}\right)\right)$

(l) $\cos\left(2\arcsin\left(\frac{2}{5}\right)\right)$

(m) $\arccos\left(\sin\left(\arctan(-1)\right)\right)$

Conic sections continued

Assignment 5.20 — Investigate the nature of the following conic sections and reduce the equation to a standard form.

(a) $9y^2+12y+4=0$

(b) $3x^2-y^2-12x-2y+6=0$

Q: What do you get when you cross a mosquito with a mountain climber?
A: Nothing. You can't cross a vector and a scalar.

6 Vector math

For many applications, real numbers suffice, but are other times, these do not suffice. Perhaps it is important to know, for instance, how close the nearest cuckoo's nest is as well as the direction in which it lies. To answer questions like these which involve both a quantitative answer, or magnitude, along with a direction, we use the mathematical objects called **vectors** (*vector*).

6.1 Definition and representation

A vector is represented geometrically as a directed line segment where the magnitude of the vector is taken to be the length of the line segment and the direction is made clear with the use of an arrow at one endpoint of the segment. When referring to vectors in this text, we shall adopt the bold-faced arrow notation, so the symbol $\vec{\mathbf{v}}$ is read as the vector v.

In Figure 6.1(a) is a typical vector $\vec{\mathbf{v}}$ with endpoints $P(1, 2)$ and $Q(4, 6)$. The point P is called the **initial point** or **head** (*aangrijpingspunt*) of $\vec{\mathbf{v}}$ and the point Q is called the **terminal point** or **tail** of $\vec{\mathbf{v}}$. Since we can reconstruct $\vec{\mathbf{v}}$ completely from P and Q, we write $\vec{\mathbf{v}} = \overrightarrow{PQ}$, where the order of points P (initial point) and Q (terminal point) is important.

While it is true that P and Q completely determine $\vec{\mathbf{v}}$, it is important to note that since vectors are defined in terms of their two characteristics, **magnitude** (*grootte*) and **direction** (*richting*), any directed line segment with the same length and direction as $\vec{\mathbf{v}}$ is considered to be the same vector as $\vec{\mathbf{v}}$, regardless of its initial point. In the case of our vector $\vec{\mathbf{v}}$ above, any vector which moves three units to the right and four up from its initial point to arrive at its terminal point is considered the same vector as $\vec{\mathbf{v}}$. The notation we use to capture this idea is the **component form** of the vector, $\vec{\mathbf{v}} = (3, 4)$, where the first number, 3, is called the ***x*-component** (*x-component*) of $\vec{\mathbf{v}}$ and the second number, 4, is called the ***y*-component** (*y-component*) of $\vec{\mathbf{v}}$. If we wanted to reconstruct $\vec{\mathbf{v}} = (3, 4)$ with initial point $\widetilde{P}(-2, 3)$, then we would find the terminal point of $\vec{\mathbf{v}}$ by adding 3 to the x-coordinate and adding 4 to the y-coordinate to obtain the terminal point $\widetilde{Q}(1, 7)$ (Figure 6.2).

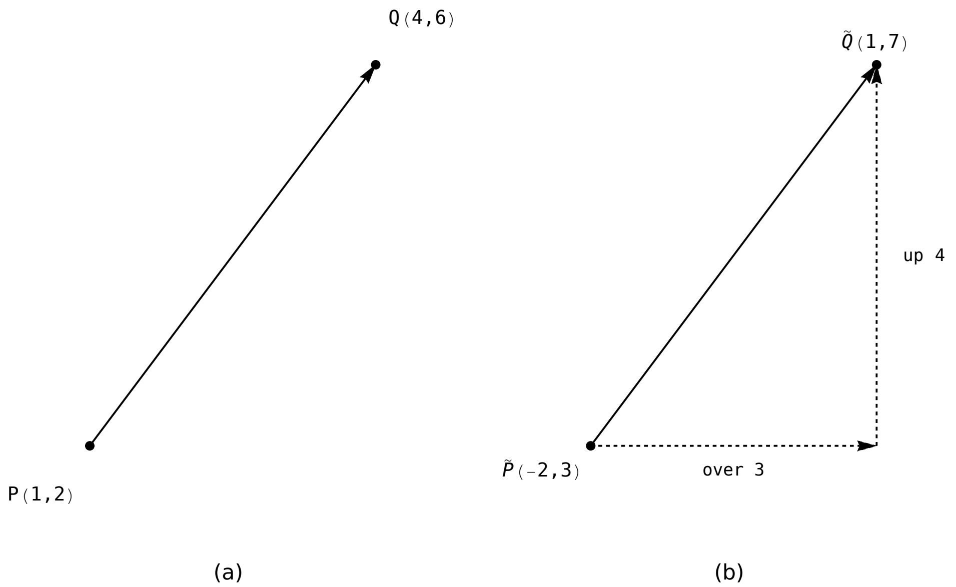


Figure 6.1: $\vec{\boldsymbol{v}} = (3,4)$ with initial point $P(1,2)$ (a) and $\widetilde{P}(-2,3)$ (b).

This idea is formalized in the following definition.

Definitie 6.1 (Component form of a vector)
Suppose $\vec{\boldsymbol{v}}$ is represented by a directed line segment with initial point $P(x_0, y_0)$ and terminal point $Q(x_1, y_1)$. The component form of $\vec{\boldsymbol{v}}$ is given by

$$\vec{\boldsymbol{v}} = \overrightarrow{PQ} = (x_1 - x_0, y_1 - y_0).$$

Using the language of components, we have that two vectors are equal if and only if their corresponding components are equal. That is, $(v_1, v_2) = (\widetilde{v}_1, \widetilde{v}_2)$ if and only if $v_1 = \widetilde{v}_1$ and $v_2 = \widetilde{v}_2$.

6.2 Vector arithmetic

6.2.1 Addition

We are now set to define operations on vectors. Suppose we are given two vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$. The sum $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ is obtained as illustrated in Figure 6.2. First, plot $\vec{\boldsymbol{v}}$. Next, plot $\vec{\boldsymbol{w}}$ so that its initial point is the terminal point of $\vec{\boldsymbol{v}}$. To plot the vector $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ we begin at the initial point of $\vec{\boldsymbol{v}}$ and end at the terminal point of $\vec{\boldsymbol{w}}$. It is helpful to think of the vector $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ as the net result of moving along $\vec{\boldsymbol{v}}$ then moving along $\vec{\boldsymbol{w}}$.

Example 6.1

A plane leaves an airport with an airspeed of 175 kilometres per hour at a bearing of N40°E. A 35 kilometres per hour wind is blowing at a bearing of S60°E. Find the speed and bearing of the plane.

Solution

For both the plane and the wind, we are given their speeds and their directions. Coupling speed

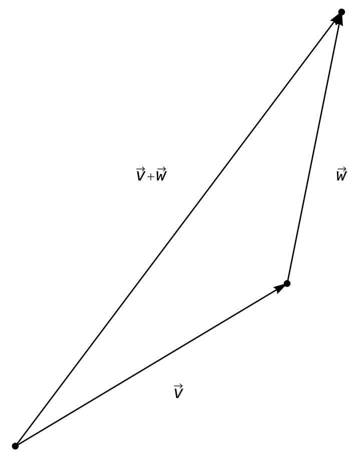


Figure 6.2: The vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ and their sum $\vec{\boldsymbol{v}}+\vec{\boldsymbol{w}}$.

(as a magnitude) with direction is the concept of velocity. We let $\vec{\boldsymbol{v}}$ denote the plane's velocity and $\vec{\boldsymbol{w}}$ denote the wind's velocity in the diagram below. The true speed and bearing is found by analysing the vector $\vec{\boldsymbol{v}}+\vec{\boldsymbol{w}}$ (Figure 6.3(a)). From the vector diagram, we get a triangle, the lengths of whose sides are the magnitude of $\vec{\boldsymbol{v}}$, which is 175, the magnitude of $\vec{\boldsymbol{w}}$, which is 35, and the magnitude of $\vec{\boldsymbol{v}}+\vec{\boldsymbol{w}}$, which we call c (Figure 6.3(b)). From the given bearing information, we go through the usual geometry to determine that the angle between the sides of length 35 and 175 measures 100°.

From the law of cosines, we determine $c=\sqrt{31850-12250\cos(100°)}\approx 184$, which means the true speed of the plane is (approximately) 184 kilometres per hour. To determine the true bearing of the plane, we need to determine the angle α. Using the law of cosines once more, we find $\cos(\alpha)=\frac{c^2+29400}{350c}$ so that $\alpha\approx 11°$. Given the geometry of the situation, we add α to the given 40° and find the true bearing of the plane to be (approximately) N51°E.

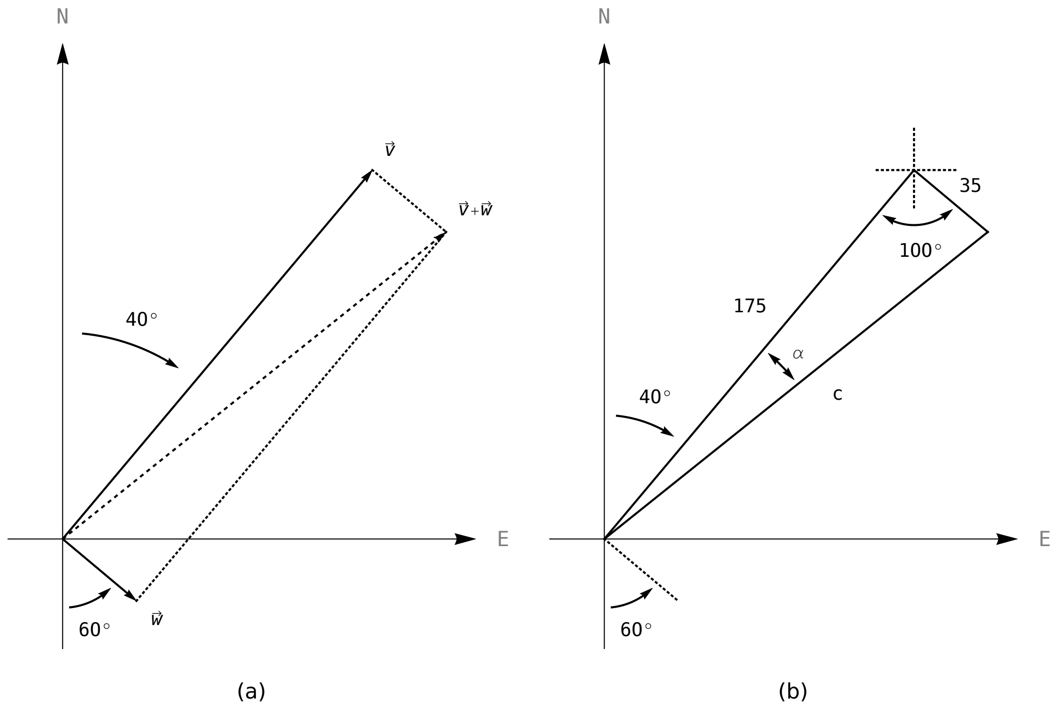


Figure 6.3: Finding the speed and bearing of a plane travelling at 175 kilometres per hour at a bearing of N40°E under a 35 kilometre per hour wind is blowing at a bearing of S60°E.

Having now a geometric understanding of the addition of vectors, we are ready to give its algebraic counterpart.

Definitie 6.2 (Vector addition)
Suppose $\vec{\boldsymbol{v}} = (v_1, v_2)$ and $\vec{\boldsymbol{w}} = (w_1, w_2)$. The vector $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ is defined by

$$\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}} = (v_1 + w_1, v_2 + w_2).$$

In order for vector addition to enjoy the same kinds of properties as real number addition, it is necessary to extend our definition of vectors to include a **zero vector** (*nulvector*) , $\vec{\boldsymbol{0}} = (0,0)$. Geometrically, $\vec{\boldsymbol{0}}$ represents a point, which we can think of as a directed line segment with the same initial and terminal points. The direction of $\vec{\boldsymbol{0}}$ is in fact undefined. Having introduced the vector counterpart of the real number 0, we can go ahead and list the properties of vector addition, which are completely in line with those for real numbers. Essentially, for all vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ we have

- **Commutative property:**
$$\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}} = \vec{\boldsymbol{w}} + \vec{\boldsymbol{v}},$$
- **Associative property:**
$$(\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}}) + \vec{\boldsymbol{w}} = \vec{\boldsymbol{u}} + (\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}),$$
- **Identity property:**
$$\vec{\boldsymbol{v}} + \vec{\boldsymbol{0}} = \vec{\boldsymbol{0}} + \vec{\boldsymbol{v}} = \vec{\boldsymbol{v}},$$
- **Inverse property:** for every vector $\vec{\boldsymbol{v}}$, there is a vector $-\vec{\boldsymbol{v}}$ so that
$$\vec{\boldsymbol{v}} + (-\vec{\boldsymbol{v}}) = (-\vec{\boldsymbol{v}}) + \vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}.$$

For what concerns the additive inverse $-\vec{\boldsymbol{v}} = (-v_1, -v_2)$ of a vector $\vec{\boldsymbol{v}} = (v_1, v_2)$, it is clear that both vectors have the same length, but opposite directions. As a result, when adding the vectors geometrically, the sum $\vec{\boldsymbol{v}} + (-\vec{\boldsymbol{v}})$ results in starting at the initial point of $\vec{\boldsymbol{v}}$ and ending back at the initial point of $\vec{\boldsymbol{v}}$, or in other words, the net result of moving $\vec{\boldsymbol{v}}$ then $-\vec{\boldsymbol{v}}$ is not moving at all. Using the additive inverse of a vector, we can define the difference of two vectors, $\vec{\boldsymbol{v}} - \vec{\boldsymbol{w}} = \vec{\boldsymbol{v}} + (-\vec{\boldsymbol{w}})$. If $\vec{\boldsymbol{v}} = (v_1, v_2)$ and $\vec{\boldsymbol{w}} = (w_1, w_2)$ then

$$\begin{aligned}
\vec{\boldsymbol{v}} - \vec{\boldsymbol{w}} &= \vec{\boldsymbol{v}} + (-\vec{\boldsymbol{w}}) \\
&= (v_1, v_2) + (-w_1, -w_2) \\
&= \big(v_1 + (-w_1), v_2 + (-w_2)\big) \\
&= (v_1 - w_1, v_2 - w_2).
\end{aligned}$$

In other words, like vector addition, vector subtraction works component-wise. From the diagram in Figure 6.5, we see that $\vec{\boldsymbol{v}} - \vec{\boldsymbol{w}}$ may be interpreted as the vector whose initial point is the terminal point of $\vec{\boldsymbol{w}}$ and whose terminal point is the terminal point of $\vec{\boldsymbol{v}}$ as depicted below. It is also worth mentioning that in the parallelogram determined by the vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$, the vector $\vec{\boldsymbol{v}} - \vec{\boldsymbol{w}}$ is one of the diagonals – the other being $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$.

Example 6.2

Let $\vec{\boldsymbol{v}} = (3,4)$ and suppose $\vec{\boldsymbol{w}} = \overrightarrow{PQ}$ where $P(-3,7)$ and $Q(-2,5)$. Find $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ and interpret this sum geometrically.

Solution

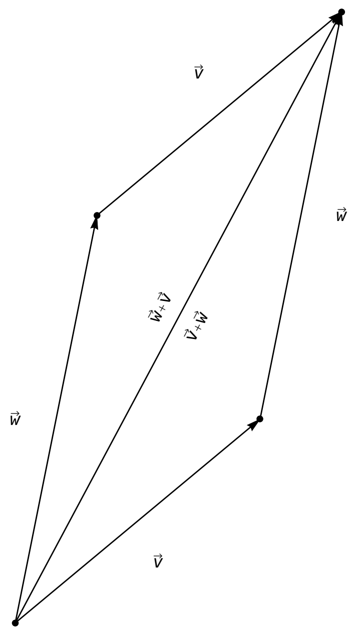


Figure 6.4: Proving the commutative property of vector addition.

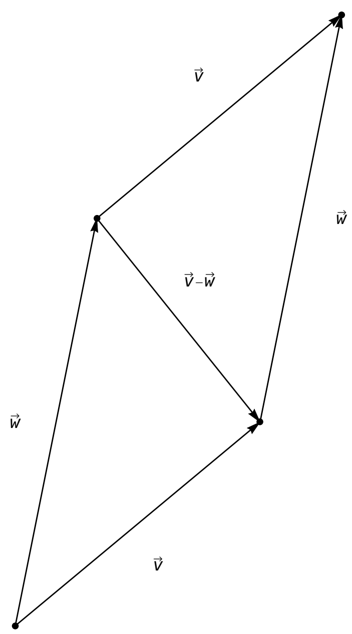


Figure 6.5: The vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ and their difference $\vec{\boldsymbol{v}} - \vec{\boldsymbol{w}}$.

First, we need to write $\vec{\boldsymbol{w}}$ in component form, which yields $\vec{\boldsymbol{w}} = \big(-2-(-3), 5-7\big) = (1,-2)$. Thus $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}} = (3,4) + (1,-2) = \big(3+1, 4+(-2)\big) = (4,2)$.

To visualize this sum in Figure 6.6, we draw $\vec{\boldsymbol{v}}$ with its initial point at $(0,0)$ for convenience so that its terminal point is $(3,4)$. Next, we graph $\vec{\boldsymbol{w}}$ with its initial point at $(3,4)$. Moving one to the right and two down, we find the terminal point of $\vec{\boldsymbol{w}}$ to be $(4,2)$. We see that the vector $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ has initial point $(0,0)$ and terminal point $(4,2)$ so its component form is $(4,2)$, as required.

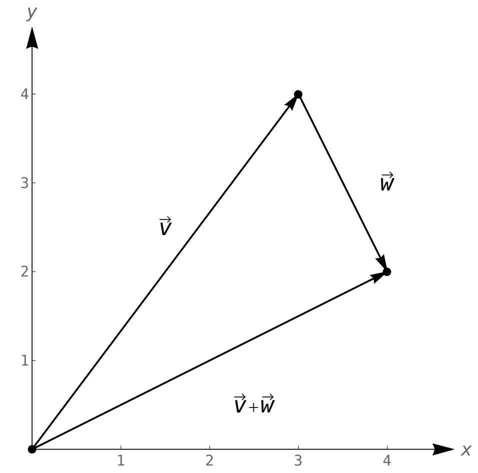


Figure 6.6: The vectors $\vec{\mathbf{v}} = (3, 4)$ and $\vec{\mathbf{w}} = (1, -2)$ and their sum.

6.2.2 Scalar multiplication

Next, we discuss **scalar multiplication** (*scalaire vermenigvuldiging*) – that is, taking a real number times a vector.

Definitie 6.3 (Scalar multiplication)
If k is a real number and $\vec{\mathbf{v}} = (v_1, v_2)$, we define $k\vec{\mathbf{v}}$ by

$$k\vec{\mathbf{v}} = k\,(v_1, v_2) = (kv_1, kv_2)\,.$$

Scalar multiplication by k in vectors can be understood geometrically as scaling the vector (if $k > 0$) or scaling the vector and reversing its direction (if $k < 0$) as demonstrated in Figure 6.7.

The properties of scalar multiplication are summarized below for vectors $\vec{\mathbf{v}}$ and $\vec{\mathbf{w}}$ and scalars k and r.

- **Associative property:**
$$(kr)\vec{\mathbf{v}} = k(r\vec{\mathbf{v}})\,,$$
- **Identity property:**
$$1\vec{\mathbf{v}} = \vec{\mathbf{v}}\,,$$
- **Additive inverse property:**
$$-\vec{\mathbf{v}} = (-1)\vec{\mathbf{v}}\,,$$
- **Distributive property of scalar multiplication over scalar addition:**
$$(k+r)\vec{\mathbf{v}} = k\vec{\mathbf{v}} + r\vec{\mathbf{v}}\,,$$
- **Distributive property of scalar multiplication over vector addition:**
$$k(\vec{\mathbf{v}} + \vec{\mathbf{w}}) = k\vec{\mathbf{v}} + k\vec{\mathbf{w}}\,,$$
- **Zero product property:**
$$k\vec{\mathbf{v}} = \vec{\mathbf{0}} \quad \Leftrightarrow \quad k = 0 \quad \vee \quad \vec{\mathbf{v}} = \vec{\mathbf{0}}\,.$$

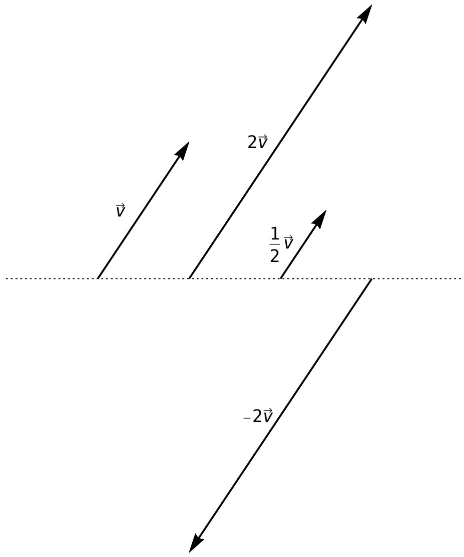


Figure 6.7: The scalar multiplication of a vector $\vec{\boldsymbol{v}}$.

Our next example demonstrates how Definition 6.3 allows us to do the same kind of algebraic manipulations with vectors as we do with variables – multiplication and division of vectors notwithstanding. If the pedantry seems familiar, it should. We spell out the solution of the following example in excruciating detail to encourage the reader to think carefully about why each step is justified.

Example 6.3

Solve $5\vec{\boldsymbol{v}}-2\left(\vec{\boldsymbol{v}}+(1,-2)\right)=\vec{\boldsymbol{0}}$ for $\vec{\boldsymbol{v}}$.

Solution

$$\begin{aligned}
5\vec{\boldsymbol{v}}-2\left(\vec{\boldsymbol{v}}+\ (1,-2)\right) &= \vec{\boldsymbol{0}} \\
5\vec{\boldsymbol{v}}-2\vec{\boldsymbol{v}}-2\,(1,-2) &= \vec{\boldsymbol{0}} \\
(5-2)\vec{\boldsymbol{v}}+\left((-2)1,(-2)(-2)\right) &= \vec{\boldsymbol{0}} \\
3\vec{\boldsymbol{v}}+(-2,4) &= \vec{\boldsymbol{0}} \\
3\vec{\boldsymbol{v}} &= \vec{\boldsymbol{0}}-(-2,4) \\
\vec{\boldsymbol{v}} &= \frac{1}{3}(2,-4) \\
\vec{\boldsymbol{v}} &= \left(\frac{2}{3},\frac{-4}{3}\right)
\end{aligned}$$

6.2.3 Magnitude and direction

A vector whose initial point is $(0,0)$ is said to be in **standard position** (*standaardvoorstelling*). If $\vec{\boldsymbol{v}}=(v_1,v_2)$ is plotted in standard position, then its terminal point is necessarily (v_1,v_2) (Figure 6.8).

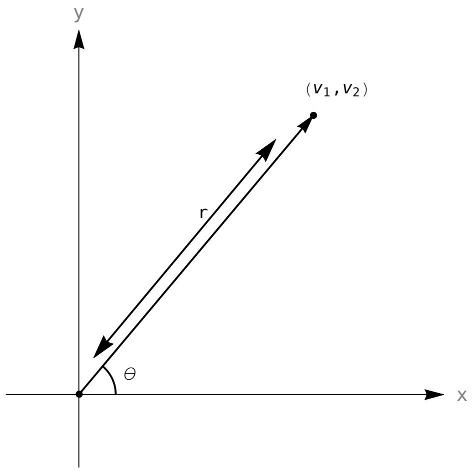


Figure 6.8: The $\vec{\mathbf{v}} = (v_1, v_2)$ in standard position.

Plotting a vector in standard position enables us to more easily quantify the concepts of magnitude and direction of the vector. We can convert the point (v_1, v_2) in rectangular coordinates to a pair (r, θ) in polar coordinates where $r \geq 0$. The magnitude of $\vec{\mathbf{v}}$, which we said earlier was the length of the directed line segment, is $r = \sqrt{v_1^2 + v_2^2}$ and is denoted by $\|\vec{\mathbf{v}}\|$. From Theorem 5.5, we know $v_1 = r\cos(\theta) = \|\vec{\mathbf{v}}\|\cos(\theta)$ and $v_2 = r\sin(\theta) = \|\vec{\mathbf{v}}\|\sin(\theta)$. From the definition of scalar multiplication and vector equality, we get

$$\begin{aligned} \vec{\mathbf{v}} &= (v_1, v_2) \\ &= \big(\|\vec{\mathbf{v}}\|\cos(\theta), \|\vec{\mathbf{v}}\|\sin(\theta)\big) \\ &= \|\vec{\mathbf{v}}\|\big(\cos(\theta), \sin(\theta)\big). \end{aligned}$$

This motivates the following definition.

Definitie 6.4 (Magnitude and direction of a vector)
Suppose $\vec{\mathbf{v}}$ is a vector with component form $\vec{\mathbf{v}} = (v_1, v_2)$. Let (r, θ) be a polar representation of the point with rectangular coordinates (v_1, v_2) with $r \geq 0$.

- The **magnitude** (*grootte*) of $\vec{\mathbf{v}}$, denoted $\|\vec{\mathbf{v}}\|$, is given by

$$\|\vec{\mathbf{v}}\| = r = \sqrt{v_1^2 + v_2^2}\,.$$

- If $\vec{\mathbf{v}} \neq \vec{\mathbf{0}}$, the **(vector) direction** (*richting*) of $\vec{\mathbf{v}}$, denoted $\hat{\mathbf{v}}$ is given by

$$\hat{\mathbf{v}} = \big(\cos(\theta), \sin(\theta)\big)\,.$$

Both magnitude and direction of a vector come along with a few important properties.

- It holds that $\|\vec{\mathbf{v}}\| \geq 0$ and $\|\vec{\mathbf{v}}\| = 0$ if and only if $\vec{\mathbf{v}} = \vec{\mathbf{0}}$.
- For all scalars k, it holds that

$$\|k\,\vec{\mathbf{v}}\| = |k|\,\|\vec{\mathbf{v}}\|\,.$$

- If $\vec{\mathbf{v}} \neq \vec{\mathbf{0}}$ then $\vec{\mathbf{v}} = \|\vec{\mathbf{v}}\|\,\hat{\mathbf{v}}$, so that

$$\hat{\mathbf{v}} = \left(\frac{1}{\|\vec{\mathbf{v}}\|}\right)\vec{\mathbf{v}}\,. \tag{6.1}$$

Example 6.4

1. Find the polar representation of the vector $\vec{\boldsymbol{v}} = \left(3, -3\sqrt{3}\right)$, assuming that $0 \le \theta < 2\pi$.

2. For the vectors $\vec{\boldsymbol{v}} = (3, 4)$ and $\vec{\boldsymbol{w}} = (1, -2)$, find the following.

   (a) $\hat{\boldsymbol{v}}$ (b) $\|\vec{\boldsymbol{v}}\| - 2\|\vec{\boldsymbol{w}}\|$ (c) $\|\vec{\boldsymbol{v}} - 2\vec{\boldsymbol{w}}\|$ (d) $\|\hat{\boldsymbol{w}}\|$

Solution

1. For $\vec{\boldsymbol{v}} = \left(3, -3\sqrt{3}\right)$, we get $\|\vec{\boldsymbol{v}}\| = \sqrt{(3)^2 + (-3\sqrt{3})^2} = 6$. We can find the θ we are after by converting the point with rectangular coordinates $(3, -3\sqrt{3})$ to polar form (r, θ) where $r = \|\vec{\boldsymbol{v}}\| > 0$. This leads to $\tan(\theta) = -3\sqrt{3}/3 = -\sqrt{3}$. Since $(3, -3\sqrt{3})$ is a point in Quadrant IV, θ is a Quadrant IV angle. Hence, we pick $\theta = \frac{5\pi}{3}$.

2. (a) Since we are given the component form of $\vec{\boldsymbol{v}}$, we will use Equation (6.1). For $\vec{\boldsymbol{v}} = (3, 4)$, we have $\|\vec{\boldsymbol{v}}\| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$. Hence, $\hat{\boldsymbol{v}} = \frac{1}{5}(3, 4) = \left(\frac{3}{5}, \frac{4}{5}\right)$.

   (b) We already know that $\|\vec{\boldsymbol{v}}\| = 5$, so to find $\|\vec{\boldsymbol{v}}\| - 2\|\vec{\boldsymbol{w}}\|$, we need only find $\|\vec{\boldsymbol{w}}\|$. Since $\vec{\boldsymbol{w}} = (1, -2)$, we get $\|\vec{\boldsymbol{w}}\| = \sqrt{1^2 + (-2)^2} = \sqrt{5}$. Hence, $\|\vec{\boldsymbol{v}}\| - 2\|\vec{\boldsymbol{w}}\| = 5 - 2\sqrt{5}$.

   (c) Our first step is to find the component form of the vector $\vec{\boldsymbol{v}} - 2\vec{\boldsymbol{w}}$. As such, we get $\vec{\boldsymbol{v}} - 2\vec{\boldsymbol{w}} = (3, 4) - 2(1, -2) = (1, 8)$. Hence, $\|\vec{\boldsymbol{v}} - 2\vec{\boldsymbol{w}}\| = \|(1, 8)\| = \sqrt{1^2 + 8^2} = \sqrt{65}$.

   (d) To find $\|\hat{\boldsymbol{w}}\|$, we first need $\hat{\boldsymbol{w}}$. Using Equation (6.1) along with $\|\vec{\boldsymbol{w}}\| = \sqrt{5}$, we get

$$\hat{\boldsymbol{w}} = \frac{1}{\sqrt{5}}(1, -2) = \left(\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right) = \left(\frac{\sqrt{5}}{5}, -\frac{2\sqrt{5}}{5}\right).$$

   Hence,

$$\|\hat{\boldsymbol{w}}\| = \sqrt{\left(\frac{\sqrt{5}}{5}\right)^2 + \left(-\frac{2\sqrt{5}}{5}\right)^2} = \sqrt{\frac{5}{25} + \frac{20}{25}} = \sqrt{1} = 1.$$

The process exemplified in Example 6.4 by which we take information about the magnitude and direction of a vector and find the component form of a vector is called **resolving** a vector into its components.

6.3 Unit vectors

Vectors with length 1 are called unit vectors and are very important in algebra.

Definitie 6.5 (Unit vector)
Let $\vec{\boldsymbol{v}}$ be a vector. If $\|\vec{\boldsymbol{v}}\| = 1$, we say that $\vec{\boldsymbol{v}}$ is a **unit vector** (*eenheidsvector*).

If $\vec{\boldsymbol{v}}$ is a unit vector, then necessarily,

$$\vec{\boldsymbol{v}} = \|\vec{\boldsymbol{v}}\|\hat{\boldsymbol{v}} = 1 \cdot \hat{\boldsymbol{v}} = \hat{\boldsymbol{v}}.$$

Conversely, it can be shown that

$$\hat{\boldsymbol{v}} = \left(\frac{1}{\|\vec{\boldsymbol{v}}\|}\right)\vec{\boldsymbol{v}}$$

is a unit vector for any nonzero vector $\vec{\boldsymbol{v}}$. The process of multiplying a nonzero vector by the factor $\|\vec{\boldsymbol{v}}\|^{-1}$ to produce a unit vector is called **normalizing** (*normeren*) the vector and the resulting vector $\hat{\boldsymbol{v}}$ is called the 'unit vector in the direction of $\vec{\boldsymbol{v}}$'. The terminal points of unit vectors, when plotted in standard position, lie on the unit circle. As a result, we may visualize normalizing a nonzero vector $\vec{\boldsymbol{v}}$ as shrinking its terminal point, when plotted in standard position, back to the unit circle. In practice, if $\vec{\boldsymbol{v}}$ is a unit vector we write it as $\hat{\boldsymbol{v}}$ (Figure 6.9).

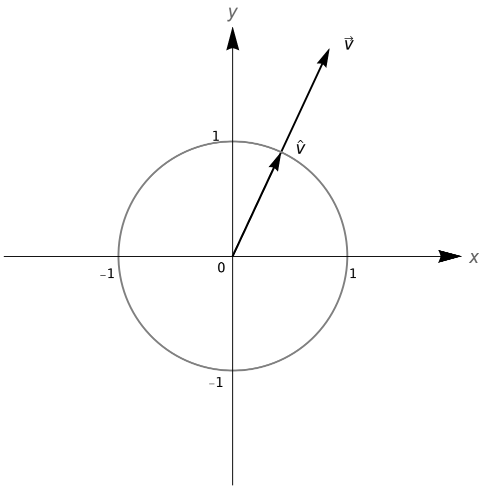


Figure 6.9: Vector normalisation $\hat{\boldsymbol{v}} = \left(\frac{1}{\|\vec{\boldsymbol{v}}\|}\right)\vec{\boldsymbol{v}}$.

Of all of the unit vectors, the so-called **principal unit vectors** (*basis eenheidsvector*) deserve special attention. In two dimensions, they are given by

- The vector $\hat{\boldsymbol{\imath}} = (1, 0)$,
- The vector $\hat{\boldsymbol{\jmath}} = (0, 1)$.

We may think of the vector $\hat{\boldsymbol{\imath}}$ as representing the positive x-direction, while $\hat{\boldsymbol{\jmath}}$ represents the positive y-direction. Consequently, the coordinate axes x and y are the axes in the direction of $\hat{\boldsymbol{\imath}}$ and $\hat{\boldsymbol{\jmath}}$, respectively. Together, $\hat{\boldsymbol{\imath}}$ and $\hat{\boldsymbol{\jmath}}$ make up the so-called **standard basis** (*standaardbasis*) for the Euclidean plane. Having introduced principal unit vectors, we are now ready to get up to the following decomposition theorem.

Theorem 6.1 (Principal vector decomposition theorem)
Let $\vec{\boldsymbol{v}}$ be a vector with component form $\vec{\boldsymbol{v}} = (v_1, v_2)$. Then $\vec{\boldsymbol{v}} = v_1\hat{\boldsymbol{\imath}} + v_2\hat{\boldsymbol{\jmath}}$.

6.4 The dot product

In Section 6.2, we learned how add and subtract vectors and how to multiply vectors by scalars. Here, we will see how to multiply vectors. We begin with the following definition.

Definitie 6.6 (Dot product)
Suppose $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are vectors whose component forms are $\vec{\boldsymbol{v}} = (v_1, v_2)$ and $\vec{\boldsymbol{w}} = (w_1, w_2)$. The **dot product** (*scalair product*) of $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ is given by

$$\vec{v}\cdot\vec{w} = (v_1, v_2)\cdot(w_1, w_2) = v_1 w_1 + v_2 w_2.$$

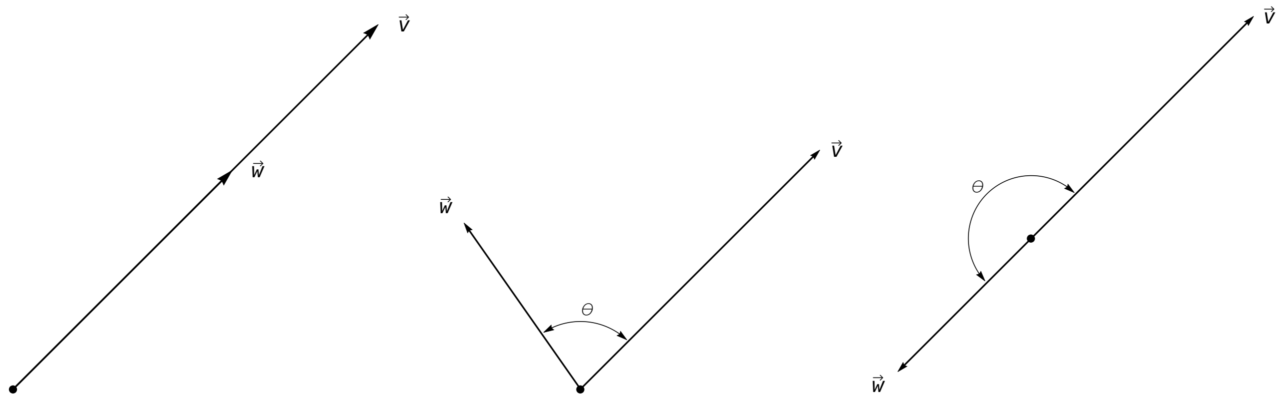
.

For example, let $\vec{v} = (3,4)$ and $\vec{w} = (1,-2)$. Then $\vec{v}\cdot\vec{w} = (3,4)\cdot(1,-2) = (3)(1)+(4)(-2) = -5$. Note that the dot product takes two vectors and produces a scalar. The dot product enjoys the following properties for all vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ and scalar k.

- **Commutative property:**
$$\vec{v}\cdot\vec{w} = \vec{w}\cdot\vec{v}\,,$$
- **Distributive property:**
$$\vec{u}\cdot(\vec{v}+\vec{w}) = \vec{u}\cdot\vec{v}+\vec{u}\cdot\vec{w}\,,$$
- **Scalar property:**
$$(k\vec{v})\cdot\vec{w} = k(\vec{v}\cdot\vec{w}) = \vec{v}\cdot(k\vec{w})\,,$$
- **Relation to magnitude:**
$$\vec{v}\cdot\vec{v} = \|\vec{v}\|^2\,.$$

Like most of the theorems involving vectors, the proof of these properties amounts to using the definition of the dot product and properties of real number arithmetic.

We now explore the geometric properties of the dot product. Suppose for that purpose $\vec{v}$ and $\vec{w}$ are two nonzero vectors. If we draw $\vec{v}$ and $\vec{w}$ with the same initial point, we define the angle between $\vec{v}$ and $\vec{w}$ to be the angle θ determined by the rays containing the vectors $\vec{v}$ and $\vec{w}$, as illustrated in Figure 6.10. We require $0 \le \theta \le \pi$.



(a) $\theta = 0$ (b) $0 < \theta < \pi$ (c) $\theta = \pi$

Figure 6.10: Two vectors $\vec{v}$ and $\vec{w}$ with the same initial point and the angle θ between them.

The following theorem gives us some insight into the geometric role the dot product plays.

Theorem 6.2 (Geometric interpretation of dot product)
If $\vec{v}$ and $\vec{w}$ are nonzero vectors then

$$\vec{v}\cdot\vec{w} = \|\vec{v}\|\|\vec{w}\|\cos(\theta), \tag{6.2}$$

where θ is the angle between $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$.

This theorem states that taking a dot product of two vectors boils down to projecting one vector onto the direction of the second vector and subsequently scaling it with the magnitude of the latter.

An immediate consequence of Theorem 6.2 is the following.

Theorem 6.3 (Angle between vectors)
Let $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ be nonzero vectors and let θ the angle between $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$. Then

$$\theta = \arccos\left(\frac{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}}{\|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\|}\right) = \arccos(\hat{\boldsymbol{v}} \cdot \hat{\boldsymbol{w}}). \tag{6.3}$$

Proof We arrive at Equation (6.3) by solving Equation (6.2) for θ. Since $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are nonzero, so are $\|\vec{\boldsymbol{v}}\|$ and $\|\vec{\boldsymbol{w}}\|$. Hence, we may divide both sides of $\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = \|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\| \cos(\theta)$ by $\|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\|$. Since $0 \leq \theta \leq \pi$ by definition, the values of θ exactly match the range of the arccosine function, so we get Equation (6.3). We can rewrite the argument of the arccosine function as

$$\frac{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}}{\|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\|} = \left(\frac{1}{\|\vec{\boldsymbol{v}}\|}\vec{\boldsymbol{v}}\right) \cdot \left(\frac{1}{\|\vec{\boldsymbol{w}}\|}\vec{\boldsymbol{w}}\right) = \hat{\boldsymbol{v}} \cdot \hat{\boldsymbol{w}},$$

giving us the alternative formula $\theta = \arccos(\hat{\boldsymbol{v}} \cdot \hat{\boldsymbol{w}})$. □

Example 6.5

Find the angle between the following pairs of vectors.

1. $\vec{\boldsymbol{v}} = \left(3, -3\sqrt{3}\right)$ and $\vec{\boldsymbol{w}} = \left(-\sqrt{3}, 1\right)$
2. $\vec{\boldsymbol{v}} = (2, 2)$ and $\vec{\boldsymbol{w}} = (5, -5)$

Solution

We use Equation (6.3) in each case below.

1. We have $\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = \left(3, -3\sqrt{3}\right) \cdot \left(-\sqrt{3}, 1\right) = -3\sqrt{3} - 3\sqrt{3} = -6\sqrt{3}$. As $\|\vec{\boldsymbol{v}}\| = \sqrt{3^2 + (-3\sqrt{3})^2} = 6$ and $\|\vec{\boldsymbol{w}}\| = \sqrt{(-\sqrt{3})^2 + 1^2} = \sqrt{4} = 2$, we find that

$$\theta = \arccos\left(\frac{-6\sqrt{3}}{12}\right) = \arccos\left(-\frac{\sqrt{3}}{2}\right) = \frac{5\pi}{6}.$$

2. For $\vec{\boldsymbol{v}} = (2, 2)$ and $\vec{\boldsymbol{w}} = (5, -5)$, we find $\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = (2, 2) \cdot (5, -5) = 10 - 10 = 0$. Hence, it does not matter what $\|\vec{\boldsymbol{v}}\|$ and $\|\vec{\boldsymbol{w}}\|$ are, and

$$\theta = \arccos\left(\frac{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}}{\|\vec{\boldsymbol{v}}\|\|\vec{\boldsymbol{w}}\|}\right) = \arccos(0) = \frac{\pi}{2}.$$

Note that the vectors $\vec{\boldsymbol{v}} = (2, 2)$, and $\vec{\boldsymbol{w}} = (5, -5)$ in Example 6.5 are called **orthogonal** (*orthogonaal*) and we write $\vec{\boldsymbol{v}} \perp \vec{\boldsymbol{w}}$, because the angle between them is $\frac{\pi}{2}$ radians $= 90°$. Geometrically this means that when orthogonal vectors are sketched with the same initial point, the lines containing the vectors are perpendicular. We state the relationship between orthogonal vectors and their dot product in the following theorem.

Theorem 6.4 (Orthogonality of vectors)
Let $\vec{\mathbf{v}}$ and $\vec{\mathbf{w}}$ be nonzero vectors. Then $\vec{\mathbf{v}} \perp \vec{\mathbf{w}}$ if and only if $\vec{\mathbf{v}} \cdot \vec{\mathbf{w}} = 0$.

6.5 Vectors in three-dimensional space and beyond

Vectors are of course not limited to the plane as most processes involving forces happen in three-dimensional space. For that reason, we will move our mathematics out of the plane and into space. That is, we begin to think mathematically not only in two dimensions, but in three, and even more. With this foundation, we can explore vectors in space.

6.5.1 Cartesian coordinates and distance in space

Up to this point we have considered mathematics in a two–dimensional world. We have plotted graphs on the xy-plane using rectangular and polar coordinates and found the area of regions in the plane. Here, we introduce Cartesian coordinates in space.

Each point P in space can be represented with an ordered triple, $P = (x, y, z)$, where x, y and z represent the relative position of P along the x-, y- and z-axes, respectively. Each axis is perpendicular to the other two. For plotting shapes in space a first convention is that the axes must conform to the **right hand rule** (*rechterhandregel*). This rule states that when the index finger of the right hand is extended in the direction of the positive x-axis, and the middle finger (bent inward so it is perpendicular to the palm) points along the positive y-axis, then the extended thumb will point in the direction of the positive z-axis (Figure 6.11).

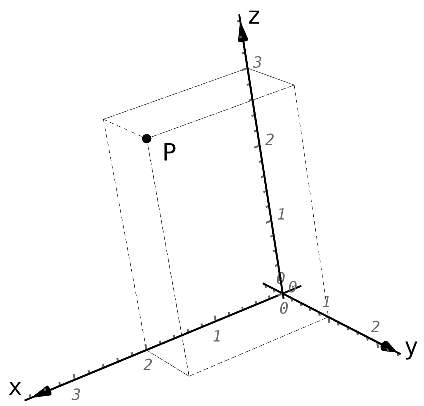


Figure 6.11: Plotting the point $P = (2, 1, 3)$ in space.

As long as the coordinate axes are positioned so that they follow this rule, it does not matter how the axes are drawn on paper. Throughout this text we will adhere to the convention with xy-plane as being a horizontal plane, where the positive z-axis is pointing up. This point of view is preferred by most mathematicians. Just as the x- and y-axes divide the plane into four quadrants, the x-, y-, and z-coordinate planes divide space into eight **octants** (*octant*). The octant in which x, y, and z are positive is called the first octant. The Euclidean distance between two points $P(x, y, z)$ and $Q(\tilde{x}, \tilde{y}, \tilde{z})$ in such a three-dimensional space is given by

$$d(P, Q) = \sqrt{(x - \tilde{x})^2 + (y - \tilde{y})^2 + (z - \tilde{z})^2}\,. \tag{6.4}$$

Higher-dimensional mathematics in art

The concept of four-dimensional space and the difficulties involved in trying to visualize it helped inspire many modern artists in the first half of the twentieth century, such as Picasso and Weber. The former got introduced to the topic through Poincaré's *Elementary Treatise on the Geometry of Four Dimensions*. He incorporated some aspects of four-dimensional mathematics in his painting *Portrait of Daniel-Henry Kahnweiler* (Figure 6.12).

Figure 6.12: Picasso's portrait of Daniel-Henry Kahnweiler.

6.5.2 Vector representation and operations in space

Essentially, all of the definitions and vector operations we introduced in Sections 6.1 and 6.2 may be generalized intuitively to three or more dimensions.

For instance, in three dimensions the component form of a vector is defined as follows.

Definitie 6.7 (Component form of a vector in three dimensions)
Suppose $\vec{\boldsymbol{v}}$ is represented by a directed line segment with initial point $P(x_0, y_0, z_0)$ and terminal point $Q(x_1, y_1, z_1)$. The component form of $\vec{\boldsymbol{v}}$ is given by

$$\vec{\boldsymbol{v}} = \overrightarrow{PQ} = (x_1 - x_0, y_1 - y_0, z_1 - z_0).$$

Likewise, vector addition, scalar multiplication and the dot product, as well as their corresponding properties, generalize naturally to three dimensions.

Definitie 6.8 (Vector addition in three dimensions)
Suppose $\vec{\boldsymbol{v}} = (v_1, v_2, v_3)$ and $\vec{\boldsymbol{w}} = (w_1, w_2, w_3)$. The vector $\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}$ is defined by

$$\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}} = (v_1 + w_1, v_2 + w_2, v_3 + w_3).$$

Definitie 6.9 (Scalar multiplication in three dimensions)
If k is a real number and $\vec{\boldsymbol{v}} = (v_1, v_2, v_3)$, we define $k\vec{\boldsymbol{v}}$ by

$$k\vec{\boldsymbol{v}} = k(v_1, v_2, v_3) = (kv_1, kv_2, kv_3).$$

Definitie 6.10 (Dot product in three dimensions)
Suppose $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are vectors whose component forms are $\vec{\boldsymbol{v}} = (v_1, v_2, v_3)$ and $\vec{\boldsymbol{w}} = (w_1, w_2, w_3)$. The dot product of $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ is given by

$$\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = (v_1, v_2, v_3) \cdot (w_1, w_2, w_3) = v_1 w_1 + v_2 w_2 + v_3 w_3.$$

Besides, in three dimensions, we can define three principal unit vectors, which make up a standard basis.

- The vector $\hat{\boldsymbol{\imath}} = (1, 0, 0)$
- The vector $\hat{\boldsymbol{\jmath}} = (0, 1, 0)$
- The vector $\hat{\boldsymbol{k}} = (0, 0, 1)$

Example 6.6

Given $\vec{\boldsymbol{v}} = (1, 2, 2)$ and $\vec{\boldsymbol{w}} = (-1, 0, 3)$, find

1. The unit vector in the direction of $\vec{\boldsymbol{v}}$.
2. $\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}$.

Solution

1. We find $\|\vec{\boldsymbol{v}}\| = 3$, so the unit vector $\hat{\boldsymbol{z}}$ in the direction of $\vec{\boldsymbol{v}}$ is

$$\hat{\boldsymbol{z}} = \frac{1}{3}\vec{\boldsymbol{v}} = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right).$$

2. Using Definition 6.10 we immediately find

$$\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} = 1(-1) + 2(0) + 2(3) = 5.$$

Example 6.7

Let $\vec{\boldsymbol{u}} = (1, 1, 1)$, $\vec{\boldsymbol{v}} = (-1, 3, -2)$ and $\vec{\boldsymbol{w}} = (-5, 1, 4)$. Find the angle between each pair of vectors.

Solution

For this purpose, we use Equation (6.3).

1. Between $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$:

$$\begin{aligned}\theta &= \arccos\left(\frac{\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}}}{\|\vec{\boldsymbol{u}}\|\,\|\vec{\boldsymbol{v}}\|}\right) \\ &= \arccos\left(\frac{0}{\sqrt{3}\sqrt{14}}\right)\end{aligned}$$

$$= \frac{\pi}{2}.$$

2. Between $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{w}}$:

$$\begin{aligned}\theta &= \arccos\left(\frac{\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{w}}}{\|\vec{\boldsymbol{u}}\|\,\|\vec{\boldsymbol{w}}\|}\right)\\ &= \arccos\left(\frac{0}{\sqrt{3}\sqrt{42}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

3. Between $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$:

$$\begin{aligned}\theta &= \arccos\left(\frac{\vec{\boldsymbol{v}}\cdot\vec{\boldsymbol{w}}}{\|\vec{\boldsymbol{v}}\|\,\|\vec{\boldsymbol{w}}\|}\right)\\ &= \arccos\left(\frac{0}{\sqrt{14}\sqrt{42}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

6.6 The cross product

Orthogonality is immensely important. A quick scan of your current environment will undoubtedly reveal numerous surfaces and edges that are perpendicular to each other. The dot product provides a quick test for orthogonality: vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ are perpendicular if and only if, $\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{v}} = 0$.

Given two non–parallel, nonzero vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ in space, it is very useful to find a vector $\vec{\boldsymbol{w}}$ that is perpendicular to both $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$. There is a operation, called the cross product, that creates such a vector.

Definitie 6.11 (Cross product)
Let $\vec{\boldsymbol{u}} = (u_1, u_2, u_3)$ and $\vec{\boldsymbol{v}} = (v_1, v_2, v_3)$ be vectors in $\mathbb{R}^3$. The **cross product** (*kruisproduct of vectorieel product*) of $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$, denoted $\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}}$, is the vector

$$\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}} = \big(u_2v_3 - u_3v_2, -(u_1v_3 - u_3v_1), u_1v_2 - u_2v_1\big).$$

Example 6.8

Let $\vec{\boldsymbol{u}} = (2, -1, 4)$ and $\vec{\boldsymbol{v}} = (3, 2, 5)$. Find $\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}}$, and verify that it is orthogonal to both $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$.

Solution

Using Definition 6.11, we have

$$\vec{u} \times \vec{v} = \Big((-1)5 - (4)2, -\big((2)5 - (4)3\big), (2)2 - (-1)3\Big) = (-13, 2, 7).$$

We now test whether or not $\vec{u} \times \vec{v}$ is orthogonal to $\vec{u}$ and $\vec{v}$ using the dot product:

$$\big(\vec{u} \times \vec{v}\big) \cdot \vec{u} = (-13, 2, 7) \cdot (2, -1, 4) = 0,$$

$$\big(\vec{u} \times \vec{v}\big) \cdot \vec{v} = (-13, 2, 7) \cdot (3, 2, 5) = 0.$$

Since both dot products are zero, $\vec{u} \times \vec{v}$ is indeed orthogonal to both $\vec{u}$ and $\vec{v}$.

Given the vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$ and the scalar c, the following properties hold for the cross product, each of which can be verified by referring to the definition.

- **Anticommutative property:**
$$\vec{u} \times \vec{v} = -(\vec{v} \times \vec{u}),$$
- **Distributive properties:**
$$\begin{aligned} (\vec{u} + \vec{v}) \times \vec{w} &= \vec{u} \times \vec{w} + \vec{v} \times \vec{w}, \\ \vec{u} \times (\vec{v} + \vec{w}) &= \vec{u} \times \vec{v} + \vec{u} \times \vec{w}, \end{aligned}$$
- **Compatibility with scalar multiplication:**
$$c(\vec{u} \times \vec{v}) = (c\vec{u}) \times \vec{v} = \vec{u} \times (c\vec{v}),$$
- **Orthogonality properties:**
$$\begin{aligned} (\vec{u} \times \vec{v}) \cdot \vec{u} &= 0, \\ (\vec{u} \times \vec{v}) \cdot \vec{v} &= 0, \end{aligned}$$
- **Zero cross product:**
$$\vec{u} \times \vec{0} = \vec{0},$$
- **Self cross product:**
$$\vec{u} \times \vec{u} = \vec{0},$$
- **Triple Scalar Product**:
$$\vec{u} \cdot (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \cdot \vec{w}.$$

Another wonderful property of the cross product, as defined, is that it follows the right hand rule. Given $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ with the same initial point, point the index finger of your right hand in the direction of $\vec{u}$ and let your middle finger point in the direction of $\vec{v}$. Your thumb will naturally extend in the direction of $\vec{u} \times \vec{v}$. If you switch, and point the index finger in the direction of $\vec{v}$ and the middle finger in the direction of $\vec{u}$, your thumb will now point in the opposite direction, allowing you to visualize the anticommutative property of the cross product.

The property $\vec{u} \times \vec{u} = \vec{0}$ reveals something more interesting than the cross product of a vector with itself is $\vec{0}$. Let $\vec{u}$ and $\vec{v}$ be parallel vectors; that is, let there be a scalar c such that $\vec{v} = c\vec{u}$. Consider their cross product:

$$\vec{u} \times \vec{v} = \vec{u} \times (c\vec{u})$$

$$
\begin{aligned}
&= c(\vec{\mathbf{u}} \times \vec{\mathbf{u}}) \\
&= \vec{\mathbf{0}}.
\end{aligned}
$$

This hints at something deeper. Theorem 6.2 related the angle between two vectors and their dot product; there is a similar relationship relating the cross product of two vectors and the angle between them, given by the following theorem.

Theorem 6.5 (The cross product and angles)
Let $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ be nonzero vectors in $\mathbb{R}^3$. Then

$$\|\vec{\mathbf{u}} \times \vec{\mathbf{v}}\| = \|\vec{\mathbf{u}}\|\,\|\vec{\mathbf{v}}\|\sin(\theta),$$

where θ, $0 \le \theta \le \pi$, is the angle between $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$.

Note that this theorem makes a statement about the magnitude of the cross product. When the angle between $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ is 0 or π (i.e., the vectors are parallel), the magnitude of the cross product is 0. The only vector with a magnitude of 0 is $\vec{\mathbf{0}}$, hence the cross product of parallel vectors is $\vec{\mathbf{0}}$.

We demonstrate the truth of this theorem in the following example.

Example 6.9

Let $\vec{\mathbf{u}} = (1,3,6)$ and $\vec{\mathbf{v}} = (-1,2,1)$. Find the angle between $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$, and the magnitude of $\vec{\mathbf{u}} \times \vec{\mathbf{v}}$.

Solution

We use Theorem 6.2 to find the angle between $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$:

$$
\begin{aligned}
\theta &= \arccos\left(\frac{\vec{\mathbf{u}} \cdot \vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|\,\|\vec{\mathbf{v}}\|}\right) \\
&= \arccos\left(\frac{11}{\sqrt{46}\sqrt{6}}\right) \\
&\approx 0.8471 = 48.54^\circ.
\end{aligned}
$$

Since $\vec{\mathbf{u}} \times \vec{\mathbf{v}} = (-9,-7,5)$ we have $\|\vec{\mathbf{u}} \times \vec{\mathbf{v}}\| = \sqrt{155}$. The question now is whether or not $\|\vec{\mathbf{u}} \times \vec{\mathbf{v}}\| = \|\vec{\mathbf{u}}\|\,\|\vec{\mathbf{v}}\|\sin(\theta)$? Using numerical approximations, we find:

$$
\begin{aligned}
\|\vec{\mathbf{u}} \times \vec{\mathbf{v}}\| &= \sqrt{155} & \qquad \|\vec{\mathbf{u}}\|\,\|\vec{\mathbf{v}}\|\sin(\theta) &= \sqrt{46}\sqrt{6}\sin(0.8471) \\
&\approx 12.45, & &\approx 12.45.
\end{aligned}
$$

Numerically, they seem equal. Using a right triangle, one can show that

$$\sin\left(\arccos\left(\frac{11}{\sqrt{46}\sqrt{6}}\right)\right) = \frac{\sqrt{155}}{\sqrt{46}\sqrt{6}},$$

which allows us to verify Theorem 6.5 exactly.

Computational geometry

In computational geometry of the plane, the cross product is used to determine the sign of the acute angle defined by three points $P_1 = (x_1, y_1), P_2 = (x_2, y_2)$, and $P_3 = (x_3, y_3)$. It corresponds to the direction (upward or downward) of the cross product of the two vectors defined by the two pairs of points (P_1, P_2) and (P_1, P_3). The sign of the acute angle is

Computational geometry

the sign of the expression

$$(x_2 - x_1)(y_3 - y_1) - (y_2 - y_1)(x_3 - x_1),$$

which is the signed length of the cross product of the two vectors. If the result is 0, the points are collinear; if it is positive, the three points constitute a positive angle of rotation around P_1 from P_2 to P_3, otherwise a negative angle.

6.7 Exercises

Vector arithmetic

Assignment 6.1 — Draw the vectors

$$\vec{\boldsymbol{a}} = (3, -2), \quad \vec{\boldsymbol{b}} = \left(-\frac{4}{3}, 0\right), \quad \vec{\boldsymbol{c}} = (0, 2) \quad \text{en} \quad \vec{\boldsymbol{d}} = (1, -1),$$

and determine their size and direction.

Assignment 6.2 — Solve the equations below to $\vec{\boldsymbol{x}}$. What are the coordinates for $\vec{\boldsymbol{x}}$ if $A = (3, 2)$ and $B = (1, 4)$?

(a) $\vec{\boldsymbol{a}} + \vec{\boldsymbol{x}} + \vec{\boldsymbol{b}} = \vec{\boldsymbol{0}}$

(b) $\vec{\boldsymbol{a}} - \vec{\boldsymbol{b}} = 2\vec{\boldsymbol{b}} + \vec{\boldsymbol{x}} - \vec{\boldsymbol{a}}$

(c) $3\,(\vec{\boldsymbol{x}} - \vec{\boldsymbol{a}}) = \vec{\boldsymbol{x}} - \vec{\boldsymbol{b}}$

(d) $2\,(\vec{\boldsymbol{x}} - \vec{\boldsymbol{a}}) = 3\big(\vec{\boldsymbol{x}} - \vec{\boldsymbol{b}}\big)$

Unit vectors

Assignment 6.3 — Consider the following points: $A = (-1, 2)$, $B = (2, 0)$, $C = (1, -3)$ and $D = (0, 4)$. Express the vectors below using the basis vectors $\hat{\boldsymbol{\imath}}$ and $\hat{\boldsymbol{\jmath}}$.

(a) $\overrightarrow{AB}$

(b) $\overrightarrow{BA}$

(c) $\overrightarrow{AC}$

(d) $\overrightarrow{AB} - \overrightarrow{BC}$

(e) $\overrightarrow{AC} - 2\overrightarrow{AB} + 3\overrightarrow{CD}$

(f) $\dfrac{\overrightarrow{AB} + \overrightarrow{AC} + \overrightarrow{AD}}{3}$

The dot product

Assignment 6.4 — Consider the vectors: $\vec{\boldsymbol{a}} = (3, 2)$ and $\vec{\boldsymbol{b}} = (x, 2 - x)$. Determine x such that $\vec{\boldsymbol{a}} \perp \vec{\boldsymbol{b}}$.

Assignment 6.5 — Determine the dot product and the angle between the vectors $\vec{\boldsymbol{a}} = (2, 1)$ and $\vec{\boldsymbol{b}} = (1, 2)$.

Assignment 6.6 — Consider the vectors $\vec{\boldsymbol{a}} = (a_1, a_2)$, $\vec{\boldsymbol{b}} = (b_1, b_2)$ and $\vec{\boldsymbol{c}} = (c_1, c_2)$. Calculate the coordinates of $\big(\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}}\big)\,\vec{\boldsymbol{c}}$ and $\vec{\boldsymbol{a}}\,\big(\vec{\boldsymbol{b}} \cdot \vec{\boldsymbol{c}}\big)$.

Assignment 6.7 — Prove that the points $A = (2, -1)$, $B = (1, 3)$ and $C = (-3, 2)$ are the vertices of a square and determine the fourth vertex.

Vectors in three-dimensional space and beyond

Assignment 6.8 — Determine the parameter h such that the given vectors are orthogonal.

(a) $\vec{\boldsymbol{v}} = (-1, 2, 3)$ en $\vec{\boldsymbol{w}} = (1, 2, h)$

(b) $\vec{\boldsymbol{a}} = (\sqrt{3}, h, 8)$ en $\vec{\boldsymbol{b}} = (h, -4, 2)$

The cross product

Assignment 6.9 — Determine $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$ with

(a) $\vec{\boldsymbol{u}} = \hat{\boldsymbol{\imath}} - 2\hat{\boldsymbol{\jmath}} + 3\hat{\boldsymbol{k}}$ and $\vec{\boldsymbol{v}} = 3\hat{\boldsymbol{\imath}} + \hat{\boldsymbol{\jmath}} - 4\hat{\boldsymbol{k}}$,

(b) $\vec{\boldsymbol{u}} = \hat{\boldsymbol{\jmath}} + 2\hat{\boldsymbol{k}}$ and $\vec{\boldsymbol{v}} = -\hat{\boldsymbol{\imath}} - \hat{\boldsymbol{\jmath}} + \hat{\boldsymbol{k}}$.

Assignment 6.10 — Determine $\vec{\boldsymbol{u}} \times (\vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}})$ and $(\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}) \times \vec{\boldsymbol{w}}$ with $\vec{\boldsymbol{u}} = \hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}} + 3\hat{\boldsymbol{k}}$, $\vec{\boldsymbol{v}} = 2\hat{\boldsymbol{\imath}} - 3\hat{\boldsymbol{\jmath}}$ and $\vec{\boldsymbol{w}} = \hat{\boldsymbol{\jmath}} - \hat{\boldsymbol{k}}$. Give an explanation as to why both results are not equal to each other.

Review exercises

Assignment 6.11 — Consider the vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$:

(a) $\vec{\boldsymbol{u}} = \hat{\boldsymbol{\imath}} - \hat{\boldsymbol{\jmath}}$ and $\vec{\boldsymbol{v}} = \hat{\boldsymbol{\jmath}} + 2\hat{\boldsymbol{k}}$,

(b) $\vec{\boldsymbol{u}} = 3\hat{\boldsymbol{\imath}} + 4\hat{\boldsymbol{\jmath}} - 5\hat{\boldsymbol{k}}$ and $\vec{\boldsymbol{v}} = 3\hat{\boldsymbol{\imath}} - 4\hat{\boldsymbol{\jmath}} - 5\hat{\boldsymbol{k}}$.

Determine the following for the given vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$.

(a) $\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}}$, $\vec{\boldsymbol{u}} - \vec{\boldsymbol{v}}$, $2\vec{\boldsymbol{u}} - 3\vec{\boldsymbol{v}}$

(b) $\|\vec{\boldsymbol{u}}\|$ and $\|\vec{\boldsymbol{v}}\|$

(c) $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$

(d) $\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}}$

(e) the angle between $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$

Q: Why did the chicken cross the road?
A: The answer is trivial and is left as an exercise for the reader.

7

Three-dimensional analytical geometry

In Chapter 6 we already introduced Cartesian coordinates in space when discussing vectors in space. Of course, we can make use of that framework to extend also analytical geometry to three dimensions. We start our brief discussion with lines and planes, after which we turn to more involved objects such as quadratic surfaces.

7.1 Lines

7.1.1 Definition

To find the equation of **a line** (*rechte*) in the xy-plane, we need a point on the line and the direction of the line. This also holds true for lines in space. Let P be a point in space, let $\vec{\boldsymbol{p}}$ be the vector with initial point at the origin and terminal point at P (i.e., $\vec{\boldsymbol{p}}$ points to P), and let $\vec{\boldsymbol{d}}$ be a vector. Consider the points on the line through P in the direction of $\vec{\boldsymbol{d}}$. Clearly one point on the line is P; we can say that the vector $\vec{\boldsymbol{p}}$ lies at this point on the line. To find another point on the line, we can start at $\vec{\boldsymbol{p}}$ and move in a direction parallel to $\vec{\boldsymbol{d}}$. For instance, starting at $\vec{\boldsymbol{p}}$ and travelling one length of $\vec{\boldsymbol{d}}$ places one at another point on the line. Consider Figure 7.1 where certain points along the line are indicated. The figure illustrates how every point on the line can be obtained by starting with $\vec{\boldsymbol{p}}$ and moving a certain distance in the direction of $\vec{\boldsymbol{d}}$. That is, we can define the line as a function of t using a **vector equation** (*vectorvergelijking van een rechte*):

$$\vec{\boldsymbol{\ell}}(t) = \vec{\boldsymbol{p}} + t\,\vec{\boldsymbol{d}}. \tag{7.1}$$

In many ways, this is not a new concept. Compare Equation (7.1) to the familiar $y = mx + b$ equation of a line:

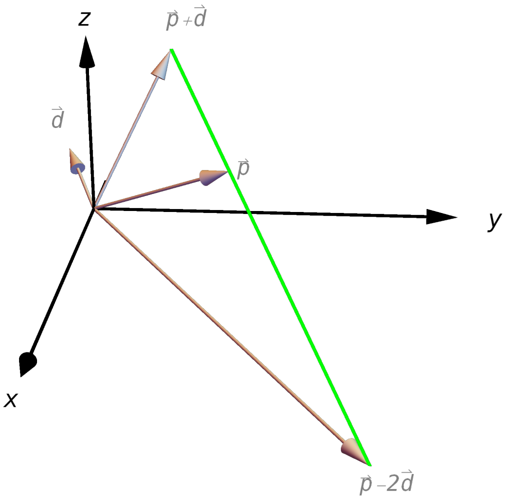


Figure 7.1: Defining a line in space.

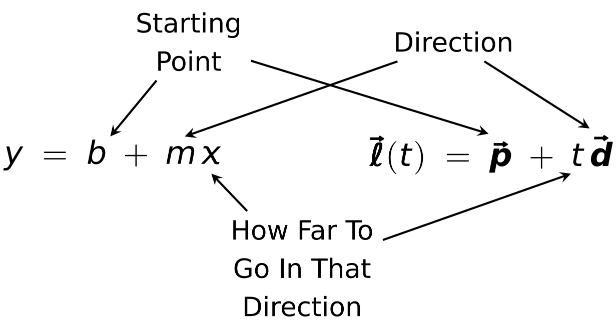


Figure 7.2: Understanding the vector equation of a line.

The equations exhibit the same structure: they give a starting point, define a direction, and state how far in that direction to travel.

Equation (7.1) is an example of a **vector–valued function** (*vectorfunctie*); the input of the function is a real number and the output is a vector. We will cover vector–valued functions extensively in Chapter 14.

There are other ways to represent a line. Let $\vec{\boldsymbol{p}} = (x_0, y_0, z_0)$ and let $\vec{\boldsymbol{d}} = (a, b, c)$. Then the equation of the line through $\vec{\boldsymbol{p}}$ in the direction of $\vec{\boldsymbol{d}}$ is:

$$\begin{aligned}\vec{\boldsymbol{\ell}}(t) &= \vec{\boldsymbol{p}} + t\,\vec{\boldsymbol{d}}\\ &= (x_0, y_0, z_0) + t(a, b, c)\\ &= (x_0 + at, y_0 + bt, z_0 + ct).\end{aligned}$$

The last line states that the x-values of the line are given by $x = x_0 + at$, the y-values are given by $y = y_0 + bt$, and the z-values are given by $z = z_0 + ct$. These three equations, taken together, are the **parametric equations of the line** (*parametervoorstelling van een rechte*) through $\vec{\boldsymbol{p}}$ in the direction of $\vec{\boldsymbol{d}}$.

Finally, each of the equations for x, y and z above contain the variable t:

$$\begin{cases} x &= x_0 + at,\\ y &= y_0 + bt,\\ z &= z_0 + ct.\end{cases}$$

We can solve for t in each equation to obtain:

$$\begin{cases} t = \dfrac{x-x_0}{a}, \\ t = \dfrac{y-y_0}{b}, \\ t = \dfrac{z-z_0}{c}. \end{cases}$$

assuming $a,b,c \neq 0$. Since t is equal to each expression on the right, we can set these equal to each other, forming the **Cartesian equations of the line** (*cartesische vergelijkingen van een rechte*) through $\vec{\boldsymbol{p}}$ in the direction of $\vec{\boldsymbol{d}}$:

$$\frac{x-x_0}{a} = \frac{y-y_0}{b} = \frac{z-z_0}{c}.$$

Each representation has its own advantages, depending on the context. We summarize these three forms in the following definition, then give examples of their use.

Definitie 7.1 (Equations of lines in space)
Consider the **line** in space that passes through $\vec{\boldsymbol{p}} = (x_0, y_0, z_0)$ in the direction of $\vec{\boldsymbol{d}} = (a,b,c)$.

1. The **vector equation** of the line is

$$\vec{\boldsymbol{\ell}}(t) = \vec{\boldsymbol{p}} + t\vec{\boldsymbol{d}}.$$

2. The **parametric equations** of the line are

$$\begin{cases} x = x_0 + at, \\ y = y_0 + bt, \\ z = z_0 + ct. \end{cases}$$

3. The **symmetric equations** of the line are

$$\frac{x-x_0}{a} = \frac{y-y_0}{b} = \frac{z-z_0}{c}.$$

Example 7.1

Give all three equations, as given in Definition 7.1, of the line through $P = (2,3,1)$ in the direction of $\vec{\boldsymbol{d}} = (-1,1,2)$.

Solution

We identify the point $P = (2,3,1)$ with the vector $\vec{\boldsymbol{p}} = (2,3,1)$. Following the definition, we have

- the vector equation of the line is $\vec{\boldsymbol{\ell}}(t) = (2,3,1) + t(-1,1,2)$;
- the parametric equations of the line are

$$\begin{cases} x = 2-t, \\ y = 3+t, \\ z = 1+2t; \end{cases}$$

and

- the symmetric equations of the line are

$$\frac{x-2}{-1} = \frac{y-3}{1} = \frac{z-1}{2}.$$

The resulting line is graphed in Figure 7.3.

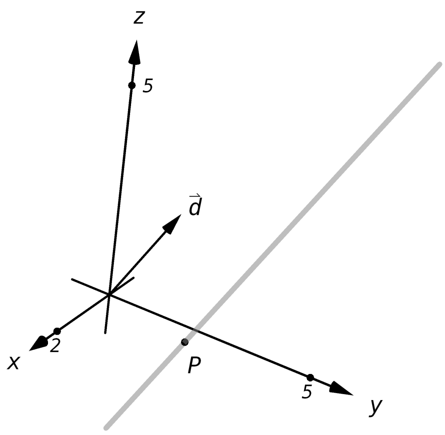


Figure 7.3: Graphing the line from Example 7.1.

The first two equations of the line are useful when a t value is given: one can immediately find the corresponding point on the line. These forms are good when calculating with a computer; most software programs easily handle equations in these formats.

7.1.2 Relative position of lines

In the plane, two distinct lines can either be parallel or they will intersect at exactly one point. In space, given the equations of two lines, it can sometimes be difficult to tell whether the lines are distinct or not (i.e., the same line can be represented in different ways). Given lines $\vec{\ell}_1(t) = \vec{p}_1 + t\vec{d}_1$ and $\vec{\ell}_2(t) = \vec{p}_2 + t\vec{d}_2$, we have four possibilities: $\vec{\ell}_1$ and $\vec{\ell}_2$ are

the **same line** (*samenvallend*)	they share all points;
intersecting (*snijdend*) lines	share only 1 point;
parallel (*evenwijdig*) lines	$\vec{d}_1 \parallel \vec{d}_2$, no points in common;
skew (*kruisend*) lines	$\vec{d}_1 \nparallel \vec{d}_2$, no points in common.

Example 7.2

Consider lines ℓ_1 and ℓ_2, given in parametric equation form

$$\ell_1 : \begin{cases} x = 1+3t \\ y = 2-t \\ z = t \end{cases} \quad \text{and} \quad \ell_2 : \begin{cases} x = -2+4s \\ y = 3+s \\ z = 5+2s\,. \end{cases}$$

Determine whether ℓ_1 and ℓ_2 are the same line, intersect, are parallel, or skew.

Solution

We start by looking at the directions of each line. Line ℓ_1 has the direction given by $\vec{d}_1 = (3, -1, 1)$ and line ℓ_2 has the direction given by $\vec{d}_2 = (4, 1, 2)$. It should be clear that $\vec{d}_1$ and $\vec{d}_2$ are not parallel, hence ℓ_1 and ℓ_2 are not the same line, nor are they parallel. Figure 7.4 verifies this fact. It shows the points and directions indicated by the equations of each line are identified.

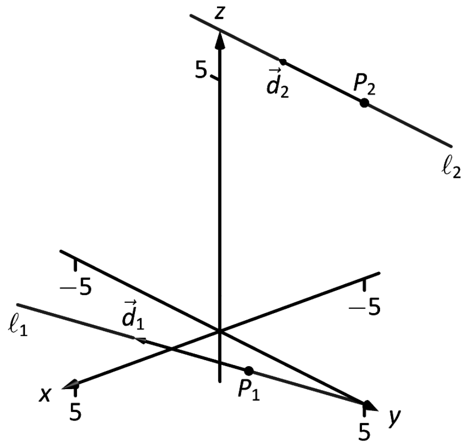


Figure 7.4: Sketching the lines from Example 7.2.

We next check to see if they intersect (if they do not, they are skew lines). To find if they intersect, we look for t and s values such that the respective x, y and z values are the same. That is, we want s and t such that:

$$\begin{cases} 1+3t = -2+4s \\ 2-t = 3+s \\ t = 5+2s\,. \end{cases}$$

This is a relatively simple system of linear equations. Since the last equation is already solved for t, we substitute that value of t into the equation above it:

$$2-(5+2s) = 3+s \quad \Rightarrow \quad s = -2, \quad t = 1\,.$$

A key to remember is that we have three equations; we need to check if $s = -2$, $t = 1$ satisfies the first equation as well:

$$1+3(1) \neq -2+4(-2).$$

It does not. Therefore, we conclude that the lines ℓ_1 and ℓ_2 are skew.

One of the key points to understand from this section is this: to describe a line, we need a point and a direction. Lines are one of two fundamental objects of study in space. The other fundamental object is the plane, which we study in detail in the next section. Many complex three dimensional objects are studied by approximating their surfaces with lines and planes.

7.2 Planes

7.2.1 Definition

Any flat surface, such as a wall, table top or stiff piece of cardboard can be thought of as representing part of a **plane** (*vlak*). Consider a piece of cardboard with a point P marked on it. One can take a nail and stick it into the cardboard at P such that the nail is perpendicular to the cardboard. This nail provides a handle for the cardboard. Moving the cardboard around moves P to different locations in space. Tilting the nail but keeping P fixed tilts the cardboard. Both moving and tilting the cardboard defines a different plane in space. In fact, we can define a plane by: 1) the location of P in space, and 2) the direction of the nail.

The previous section showed that one can define a line given a point on the line and the direction of the line. One can make a similar statement about planes: we can define a plane in space given a point on the plane and the direction the plane faces. Once again, the direction information will be supplied by a vector, called a **normal vector** (*normaalvector*), that is orthogonal to the plane.

What exactly does orthogonal to the plane mean? Choose any two points P and Q in the plane, and consider the vector $\overrightarrow{PQ}$. We say a vector $\vec{\boldsymbol{n}}$ is orthogonal to the plane if $\vec{\boldsymbol{n}}$ is perpendicular to $\overrightarrow{PQ}$ for all choices of P and Q; that is, if $\vec{\boldsymbol{n}}\cdot\overrightarrow{PQ}=0$ for all P and Q. This gives us way of writing an equation describing the plane. Let $P=(x_0,y_0,z_0)$ be a point in the plane and let $\vec{\boldsymbol{n}}=(a,b,c)$ be a normal vector to the plane. A point $Q=(x,y,z)$ lies in the plane defined by P and $\vec{\boldsymbol{n}}$ if and only if, $\overrightarrow{PQ}$ is orthogonal to $\vec{\boldsymbol{n}}$. Knowing $\overrightarrow{PQ}=(x-x_0,y-y_0,z-z_0)$, consider:

$$\begin{aligned} & \overrightarrow{PQ}\cdot\vec{\boldsymbol{n}}=0 \\ \Leftrightarrow\quad & (x-x_0,y-y_0,z-z_0)\cdot(a,b,c)=0 \\ \Leftrightarrow\quad & a(x-x_0)+b(y-y_0)+c(z-z_0)=0. \end{aligned} \tag{7.2}$$

Equation (7.2) defines an implicit function describing the plane. More algebra produces:

$$ax+by+cz=ax_0+by_0+cz_0.$$

The right hand side is just a number, so we replace it with d:

$$ax+by+cz=d. \tag{7.3}$$

As long as $c\neq 0$, we can solve for z:

$$z=\frac{1}{c}(d-ax-by). \tag{7.4}$$

Equation (7.4) is especially useful as many computer programs can graph functions in this form. Equations (7.2) and (7.3) have specific names, given next.

Definitie 7.2 (Equations of a plane)
The **plane** passing through the point $P=(x_0,y_0,z_0)$ with normal vector $\vec{\boldsymbol{n}}=(a,b,c)$ can be described by an **equation with standard form**

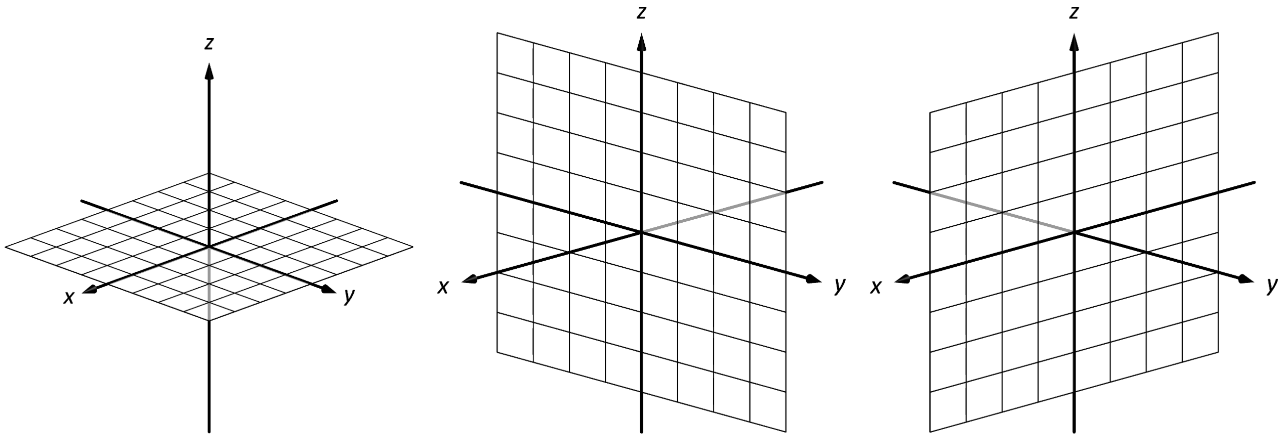
$$a(x-x_0)+b(y-y_0)+c(z-z_0)=0;$$

the **equation's general form** is

$$ax+by+cz=d.$$

Clearly, the coordinate axes naturally define three planes (shown in Figure 7.5), the **coordinate planes** (*coördinaatvlak*): the xy-plane, the yz-plane and the xz-plane. The xy-plane is characterized as the set of all points in space where the z-value is 0. This, in fact, gives us an equation that describes this plane: $z = 0$. Likewise, the xz-plane is all points where the y-value is 0, characterized by $y = 0$. Furthermore, the equation $x = 2$ describes all points in space where the x-value is 2.

A key to remember throughout this section is this: to find the equation of a plane, we need a point and a normal vector. We will give several examples of finding the equation of a plane, and in each one different types of information are given. In each case, we need to use the given information to find a point on the plane and a normal vector.



(a) The xy-plane. (b) The yz-plane. (c) The xz-plane.

Figure 7.5: The coordinate planes.

Example 7.3

Write the equation of the plane that passes through the points $P = (1,1,0)$, $Q = (1,2,-1)$ and $R = (0,1,2)$ in standard form.

Solution

We need a vector $\vec{n}$ that is orthogonal to the plane. Since P, Q and R are in the plane, so are the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$; $\overrightarrow{PQ} \times \overrightarrow{PR}$ is orthogonal to $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ and hence the plane itself.

It is straightforward to compute $\vec{n} = \overrightarrow{PQ} \times \overrightarrow{PR} = (2,1,1)$. We can use any point we wish in the plane (any of P, Q or R will do) and we arbitrarily choose P. Following Definition 7.2, the equation of the plane in standard form is

$$2(x-1)+(y-1)+z = 0.$$

The plane is sketched in Figure 7.6.

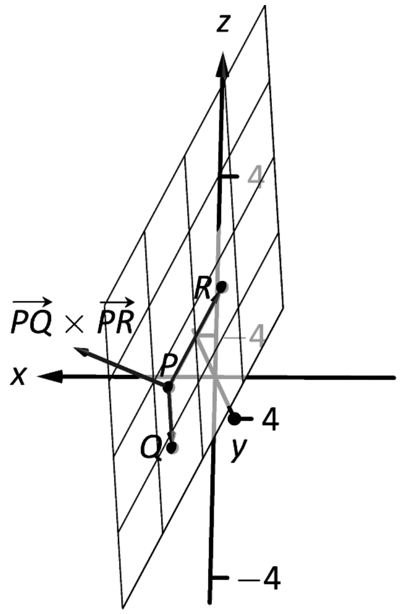


Figure 7.6: Sketching the plane in Example 7.3.

We have just demonstrated the fact that any three non-collinear points define a plane. This is why a three-legged stool does not rock; its three feet always lie in a plane. A four-legged stool will rock unless all four feet lie in the same plane.

Example 7.4

Verify that lines ℓ_1 and ℓ_2, whose parametric equations are given below, intersect, then give the equation of the plane that contains these two lines in general form.

$$\ell_1 : \begin{cases} x &= -5+2s \\ y &= 1+s \\ z &= -4+2s \end{cases} \qquad \ell_2 : \begin{cases} x &= 2+3t \\ y &= 1-2t \\ z &= 1+t \end{cases}$$

Solution

The lines clearly are not parallel. If they do not intersect, they are skew, meaning there is not a plane that contains them both. If they do intersect, there is such a plane.

To find their point of intersection, we set the x, y and z equations equal to each other and solve for s and t:

$$\begin{cases} -5+2s &= 2+3t \\ 1+s &= 1-2t \\ -4+2s &= 1+t \end{cases} \quad \Rightarrow \quad s=2, \quad t=-1.$$

When $s=2$ and $t=-1$, the lines intersect at the point $P=(-1,3,0)$.

Let $\vec{\boldsymbol{d}}_1=(2,1,2)$ and $\vec{\boldsymbol{d}}_2=(3,-2,1)$ be the directions of lines ℓ_1 and ℓ_2, respectively. A normal vector to the plane containing these the two lines will also be orthogonal to $\vec{\boldsymbol{d}}_1$ and $\vec{\boldsymbol{d}}_2$. Thus we find a normal vector $\vec{\boldsymbol{n}}$ by computing $\vec{\boldsymbol{n}}=\vec{\boldsymbol{d}}_1 \times \vec{\boldsymbol{d}}_2=(5,4-7)$.

We can pick any point in the plane with which to write our equation; each line gives us infinite choices of points. We choose P, the point of intersection. We follow Definition 7.2 to write the plane's equation in general form:

$$\begin{aligned} 5(x+1)+4(y-3)-7z &= 0 \\ \Leftrightarrow \quad 5x+5+4y-12-7z &= 0 \\ \Leftrightarrow \quad 5x+4y-7z &= 7. \end{aligned}$$

The plane is $5x+4y-7z=7$; it is sketched in Figure 7.7.

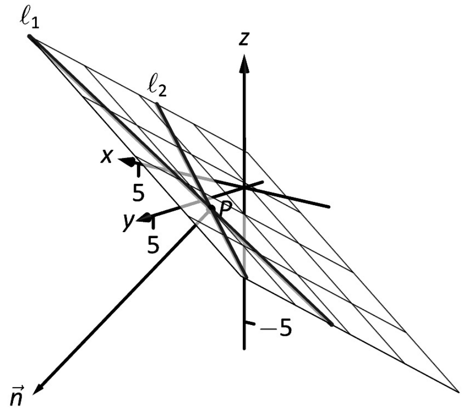


Figure 7.7: Sketching the plane in Example 7.4.

Having now defined lines and planes in space, it makes of course sense to look for the intersection between planes or between a plane and a line.

Example 7.5

Give the parametric equations of the line that is the intersection of the following planes:

$$\begin{aligned} p_1: &\quad x-(y-2)+(z-1)=0, \\ p_2: &\quad -2(x-2)+(y+1)+(z-3)=0. \end{aligned}$$

Solution

To find an equation of a line, we need a point on the line and the direction of the line. We can find a point on the line by solving each equation of the planes for z:

$$\begin{aligned} p_1: &\quad z=-x+y-1 \\ p_2: &\quad z=2x-y-2. \end{aligned}$$

We can now set these two equations equal to each other to find values of x and y where the planes have the same z value:

$$\begin{aligned} -x+y-1 &= 2x-y-2 \\ \Leftrightarrow \quad 2y &= 3x-1 \\ \Leftrightarrow \quad y &= \frac{1}{2}(3x-1). \end{aligned}$$

We can choose any value for x; we choose $x=1$. This determines that $y=1$. We can now use the equations of either plane to find z: when $x=1$ and $y=1$, $z=-1$ on both planes. We have found a point P on the line: $P=(1,1,-1)$.

We now need the direction of the line. Since the line lies in each plane, its direction is orthogonal to a normal vector for each plane. Considering the equations for p_1 and p_2, we can quickly determine their normal vectors. For p_1, $\vec{\boldsymbol{n}}_1 = (1,-1,1)$ and for p_2, $\vec{\boldsymbol{n}}_2 = (-2,1,1)$. A direction orthogonal to both of these directions is their cross product: $\vec{\boldsymbol{d}} = \vec{\boldsymbol{n}}_1 \times \vec{\boldsymbol{n}}_2 = (-2,-3,-1)$.

The parametric equations of the line through $P = (1,1,-1)$ in the direction of $d = (-2,-3,-1)$ is:

$$\ell : \begin{cases} x &= 1-2t \\ y &= 1-3t \\ z &= -1-t. \end{cases}$$

The planes and line are graphed in Figure 7.8.

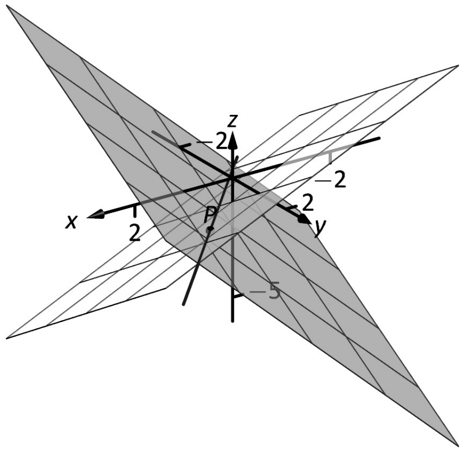


Figure 7.8: Graphing the planes and their line of intersection in Example 7.5.

Finite element method

One of the most popular numerical method for solving problems of engineering that relies on an approximation of surfaces by means of small planes is the finite element method. It proceeds by dividing the domain of the problem into a collection of subdomains (i.e. small planes), with each subdomain represented by a set of element equations to the original problem. For instance, the stresses on the surface of a human knee joint can be assessed by means of this method.

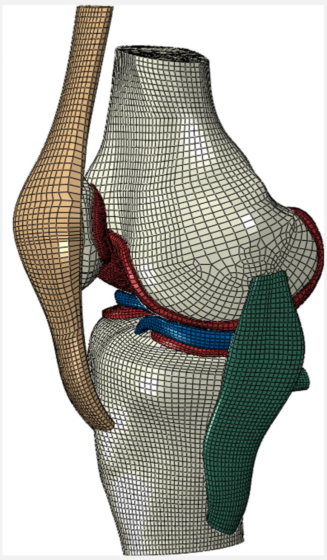


Figure 7.9: Illustrating finding the distance from a point to a plane.

In the final section of this chapter we will investigate more complex three-dimensional objects.

7.3 Three-dimensional objects

7.3.1 Spheres and cylinders

Just as a circle is the set of all points in the plane equidistant from its centre, a sphere is the set of all points in space that are equidistant from a given point. Equation (6.4) allows us to write an equation of the sphere.

We start with a point $C = (a,b,c)$ which is to be the centre of a sphere with radius r. If a point $P = (x,y,z)$ lies on the sphere, then P is r units from C; that is,

$$||\overrightarrow{PC}|| = \sqrt{(x-a)^2 + (y-b)^2 + (z-c)^2} = r.$$

Squaring both sides, we get the standard equation of a sphere in space with center at $C = (a,b,c)$ with radius r, as given in the following definition.

Definitie 7.3 (Standard equation of a sphere)
The standard equation of the **sphere** (*bol*) with radius r, centred at $C = (a,b,c)$, is

$$(x-a)^2 + (y-b)^2 + (z-c)^2 = r^2.$$

The equation of a sphere is an example of an implicit function defining a surface in space. In the case of a sphere, the variables x, y and z are all used. We now consider a situation where surfaces are defined where one or two of these variables are absent, in addition to the coordinate planes that we encountered before.

The equation $x = 1$ obviously lacks the y and z variables, meaning it defines points where the y and z coordinates can take on any value. Now consider the equation $x^2 + y^2 = 1$ in space. In the plane, this equation describes a circle of radius 1, centred at the origin. In space, the z coordinate is not specified, meaning it can take on any value. In Figure 7.10(a), we show part of the graph of the equation $x^2 + y^2 = 1$ by sketching 3 circles: the bottom one has a constant z-value of -1.5, the middle one has a z-value of 0 and the top circle has a z-value of 1. By plotting all possible z-values, we get the surface shown in Figure 7.10(b). This surface is a cylinder.

Definitie 7.4 (Cylinder)
Let C be a curve in a plane and let L be a line not parallel to C. A **cylinder** (*cilinder*) is the set of all lines parallel to L that pass through C. The curve C is the **directrix** (*richtkromme*) of the cylinder, and the lines are the **rulings** (*beschrijvenden*).

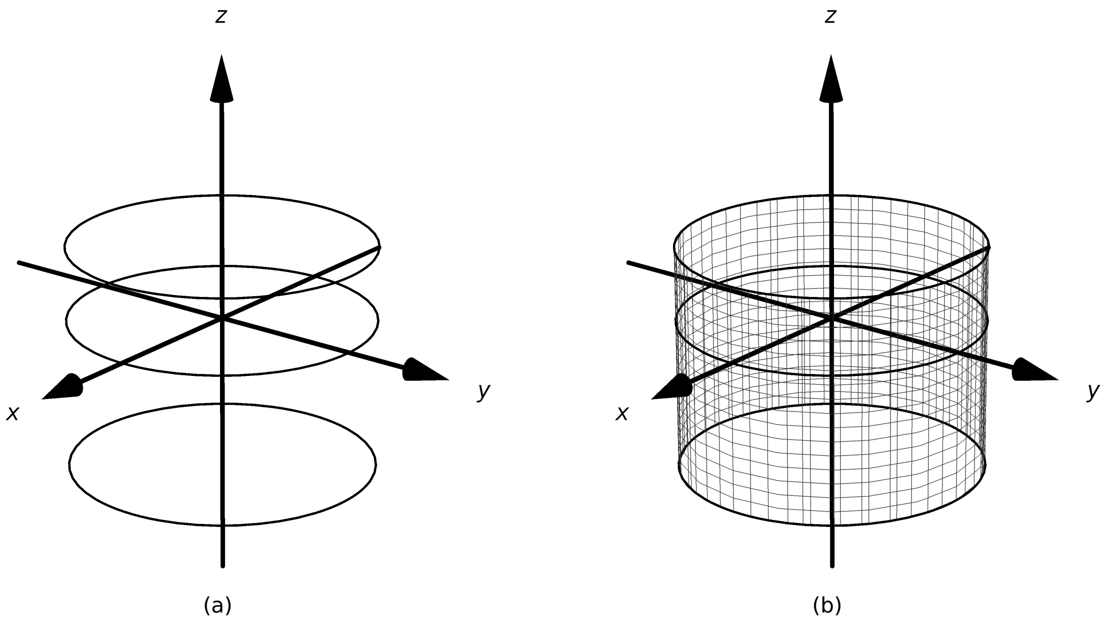


Figure 7.10: Sketching $x^2 + y^2 = 1$.

In this text, we consider curves C that lie in planes parallel to one of the coordinate planes, and lines L that are perpendicular to these planes, forming **right cylinders** (*rechte cilinder*). Thus the directrix can be defined using equations involving 2 variables, and the rulings will be parallel to the axis of the third variable.

In the example preceding the definition, the curve $x^2 + y^2 = 1$ in the xy-plane is the directrix and the rulings are lines parallel to the z-axis. Actually, any circle shown in Figure 7.10(a) can be considered a directrix; we simply choose the one where $z = 0$.

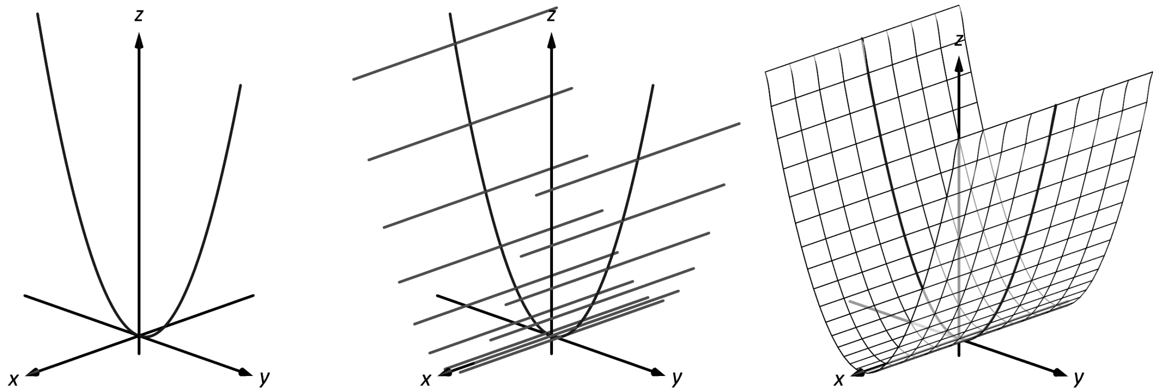
Example 7.6

Graph the following cylinders.

1. $z = y^2$
2. $x = \sin(z)$

Solution

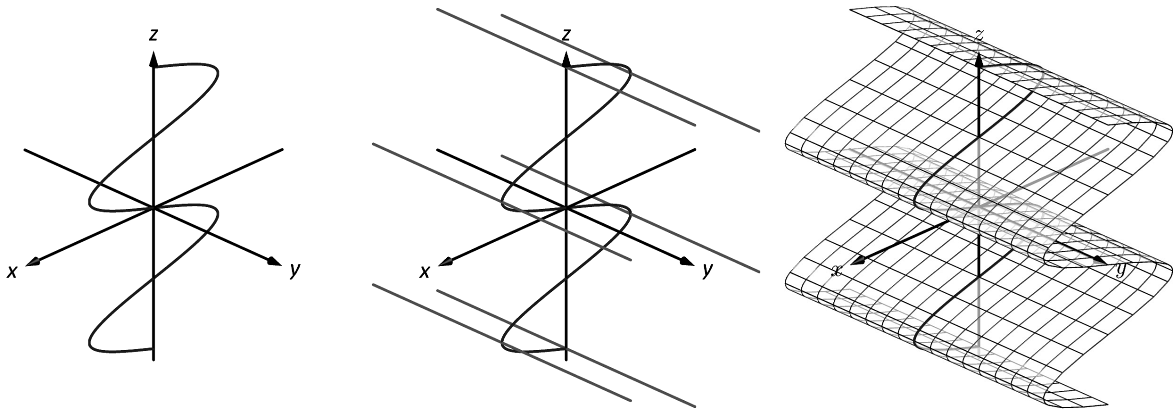
1. We can view the equation $z = y^2$ as a parabola in the yz-plane, as illustrated in Figure 7.11(a). As x does not appear in the equation, the rulings are lines through this parabola parallel to the x-axis, shown in Figure 7.11(b). These rulings give an idea as to what the surface looks like, drawn in Figure 7.11(c).



(a) (b) (c)

Figure 7.11: Sketching the parabolic cylinder defined by $z = y^2$.

2. We can view the equation $x = \sin(z)$ as a sine curve that exists in the xz-plane, as shown in Figure 7.12(a). The rules are parallel to the y-axis as the variable y does not appear in the equation $x = \sin(z)$; some of these are shown in Figure 7.12(b). The surface is shown in Figure 7.12(c).



(a) (b) (c)

Figure 7.12: Sketching the cylinder defined by $x = \sin(z)$.

7.3.2 Quadratic surfaces

Spheres, planes and cylinders are important surfaces to understand. We now consider one last type of surface, a quadric surface. Essentially, they are the three-dimensional extension of the conic sections we discussed in Section 4.4. Their definition is given below.

Definitie 7.5 (Quadric surface)
A **quadric surface** (*kwadratisch oppervlak*) is the graph of the general second–degree equation in three variables:

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0.$$

When the coefficients d, e or f are not zero, the basic shapes of the quadric surfaces are rotated in space. We will focus on quadric surfaces where these coefficients are 0; we will not consider rotations.

There are six basic quadric surfaces: the elliptic paraboloid, elliptic cone, ellipsoid, hyperboloid of one sheet, hyperboloid of two sheets, and the hyperbolic paraboloid.

We study each shape by considering **traces**, that is, intersections of each surface with a plane parallel to a coordinate plane. For instance, consider the elliptic paraboloid $z = x^2/4 + y^2$, shown in Figure 7.13. If we intersect this shape with the plane $z = d$ (i.e., replace z with d), we have the equation:

$$d = \frac{x^2}{4} + y^2.$$

Divide both sides by d:

$$1 = \frac{x^2}{4d} + \frac{y^2}{d}.$$

This describes an ellipse – so cross sections parallel to the x-y coordinate plane are ellipses. This ellipse is drawn in the figure. Now consider cross sections parallel to the xz-plane. For instance, letting $y = 0$ gives the equation $z = x^2/4$, clearly a parabola. Intersecting with the plane $x = 0$ gives a cross section defined by $z = y^2$, another parabola. These parabolas are also sketched in the figure.

Thus we see where the elliptic paraboloid gets its name: some cross sections are ellipses, and others are parabolas.

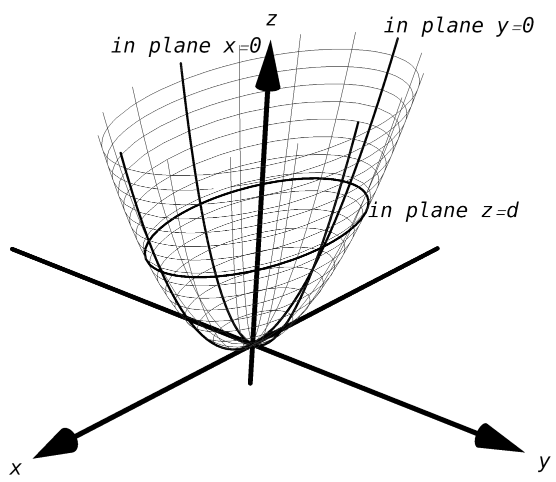
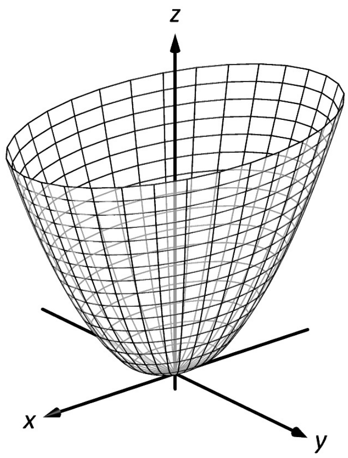


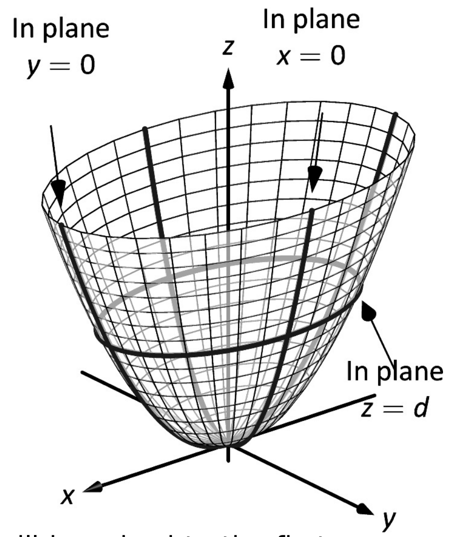
Figure 7.13: The elliptic paraboloid $z = x^2/4 + y^2$.

Such an analysis can be made with each of the quadric surfaces. We give a sample equation of each, provide a sketch with representative traces, and describe these traces.

Elliptic paraboloid (*elliptische paraboloïde*): $z = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$



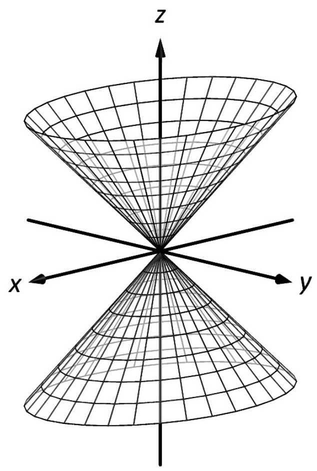
Plane	Trace
$x = d$	Parabola
$y = d$	Parabola
$z = d$	Ellipse



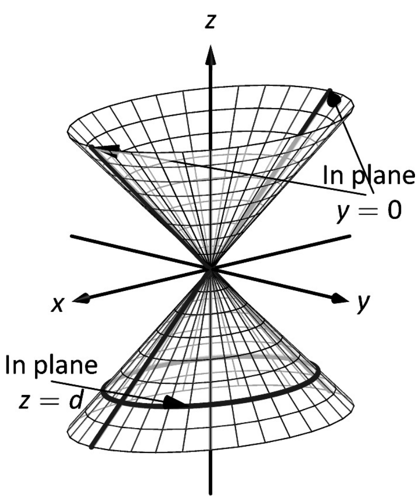
One variable in the equation of the elliptic paraboloid will be raised to the first power; above, this is the z variable. The paraboloid will open in the direction of this variable's axis. Thus $x = y^2/a^2 + z^2/b^2$ is an elliptic paraboloid that opens along the x-axis.

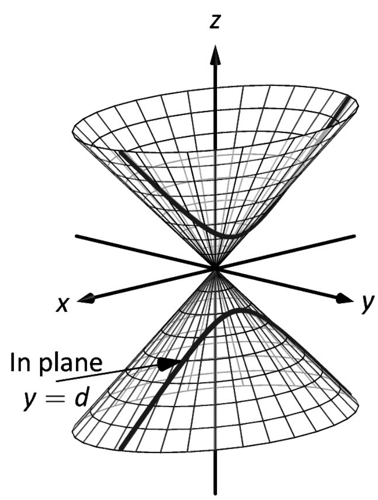
Multiplying the right hand side by (-1) defines an elliptic paraboloid that opens in the opposite direction.

Elliptic cone (*elliptische kegel*): $z^2 = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$



Plane	Trace
$x = 0$	Crossed Lines
$y = 0$	Crossed Lines
$x = d$	Hyperbola
$y = d$	Hyperbola
$z = d$	Ellipse



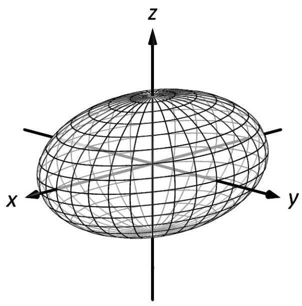


One can rewrite the governing equation as

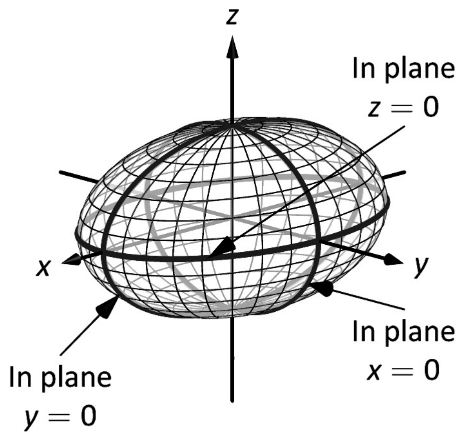
$$z^2 - \frac{x^2}{a^2} - \frac{y^2}{b^2} = 0.$$

The one variable with a positive coefficient corresponds to the axis that the cones open along.

Ellipsoid (*ellipsoïde*): $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$



Plane	**Trace**
$x = d$	Ellipse
$y = d$	Ellipse
$z = d$	Ellipse



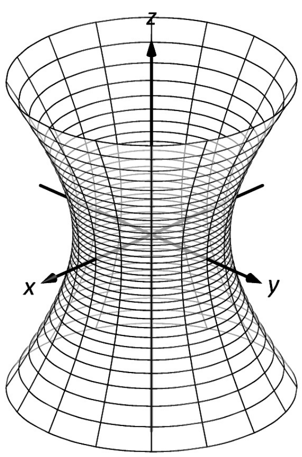
If $a = b = c \neq 0$, the ellipsoid is a sphere with radius a.

Earth ellipsoid

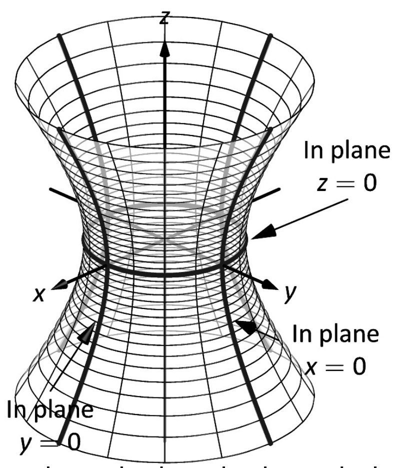
An Earth ellipsoid is a mathematical figure approximating the Earth's form, used as a reference frame for computations in the geosciences. Various different ellipsoids have been used as approximations. It is an ellipsoid of revolution whose minor axis, which connects the geographical North Pole and South Pole, is approximately aligned with the Earth's axis of rotation.

Many methods exist for determination of the axes of an Earth ellipsoid, but several ellipsoids are of special importance, such as the Bessel ellipsoid of 1841 and (for GPS positioning) the WGS84 ellipsoid. In the latter, the semi major axis measures 6 378 137.0 m, while its semi minor axis measures approximately 6 356 752.314 245 m.

Hyperboloid of one sheet (*eenbladige hyperboloïde*): $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$

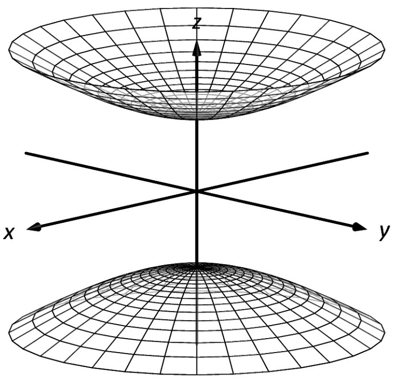


Plane	**Trace**
$x = d$	Hyperbola
$y = d$	Hyperbola
$z = d$	Ellipse

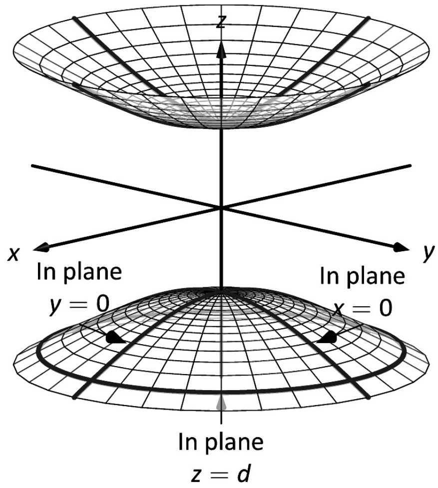


The one variable with a negative coefficient corresponds to the axis that the hyperboloid opens along.

Hyperboloid of two sheets (*tweebladige hyperboloïde*): $\dfrac{z^2}{c^2} - \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$

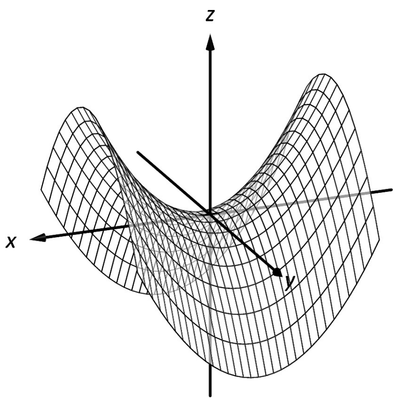


Plane	Trace
$x = d$	Hyperbola
$y = d$	Hyperbola
$z = d$	Ellipse

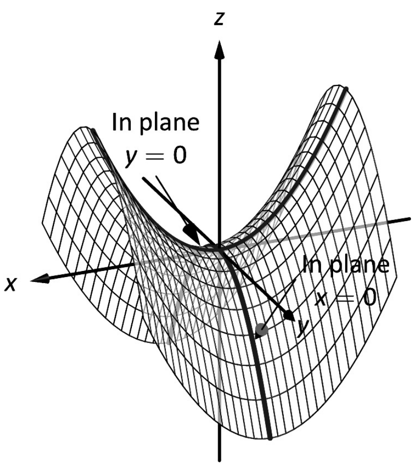


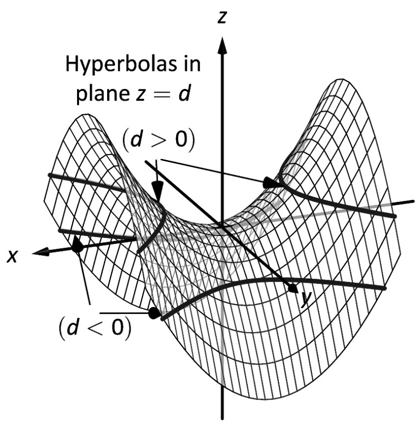
The one variable with a positive coefficient corresponds to the axis that the hyperboloid opens along. In the case illustrated, when $|d| < |c|$, there is no trace.

Hyperbolic Paraboloid (*hyperbolische paraboloïde*): $z = \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2}$



Plane	Trace
$x = d$	Parabola
$y = d$	Parabola
$z = d$	Hyperbola





The parabolic traces will open along the axis of the one variable that is raised to the first power.

Example 7.7

Sketch the quadratic surface defined by the given equation.

1. $y = \dfrac{x^2}{4} + \dfrac{z^2}{16}$
2. $x^2 + \dfrac{y^2}{9} + \dfrac{z^2}{4} = 1$
3. $z = y^2 - x^2$.

Solution

1. We first identify the quadratic by pattern–matching with the equations given previously. Only two surfaces have equations where one variable is raised to the first power, the elliptic paraboloid and the hyperbolic paraboloid. In the latter case, the other variables have different signs, so we conclude that this describes an elliptic paraboloid. As the variable with the first power is y, we note the paraboloid opens along the y-axis.

   To make a decent sketch by hand, we need only draw a few traces. In this case, the traces $x = 0$ and $z = 0$ form parabolas that outline the shape.

   $x = 0$: The trace is the parabola $y = z^2/16$.

   $z = 0$: The trace is the parabola $y = x^2/4$.

   Graphing each trace in the respective plane creates a sketch as shown in Figure 7.14(a). This is enough to give an idea of what the paraboloid looks like. The surface is filled in Figure 7.14(b).

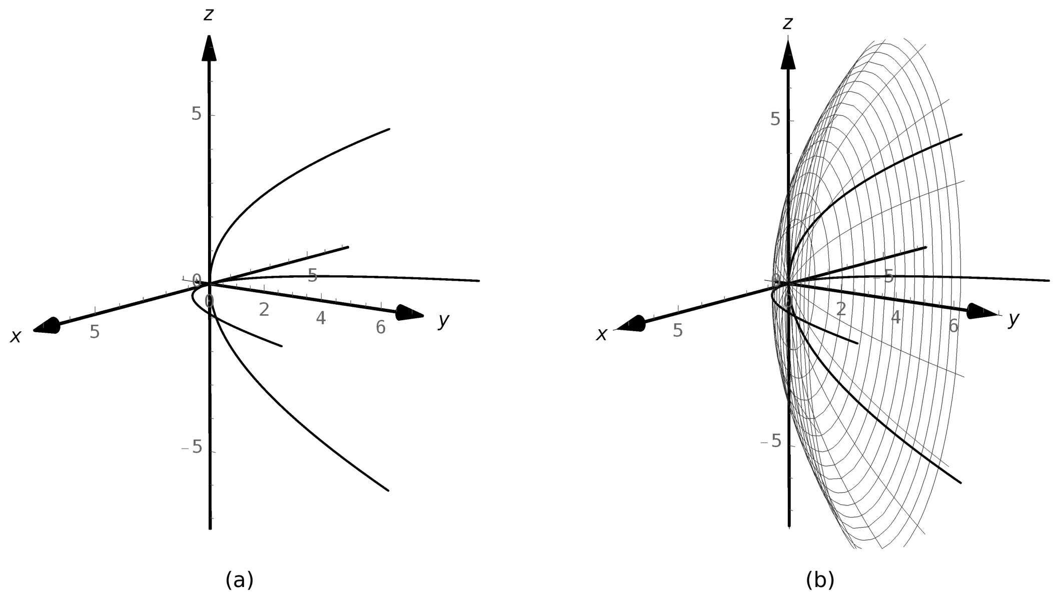


Figure 7.14: Sketching the elliptic paraboloid from Example 7.7.1.

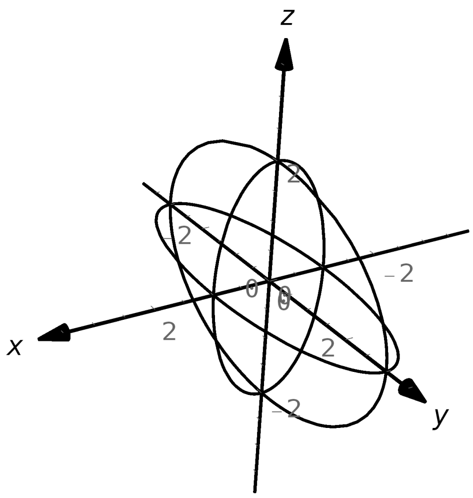
2. This is an ellipsoid. We can get a good idea of its shape by drawing the traces in the coordinate planes.

   $x = 0$: The trace is the ellipse $y^2/9 + z^2/4 = 1$. The major axis is along the y–axis with length 6 the minor axis is along the z-axis with length 4.

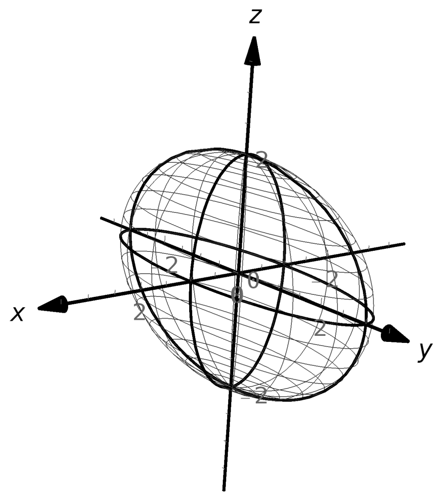
$y = 0$: The trace is the ellipse $x^2 + z^2/4 = 1$. The major axis is along the z-axis, and the minor axis has length 2 along the x-axis.

$z = 0$: The trace is the ellipse $x^2 + y^2/9 = 1$, with major axis along the y-axis.

Graphing each trace in the respective plane creates a sketch as shown in Figure 7.15(a). Filling in the surface gives Figure 7.15(b).



(a)



(b)

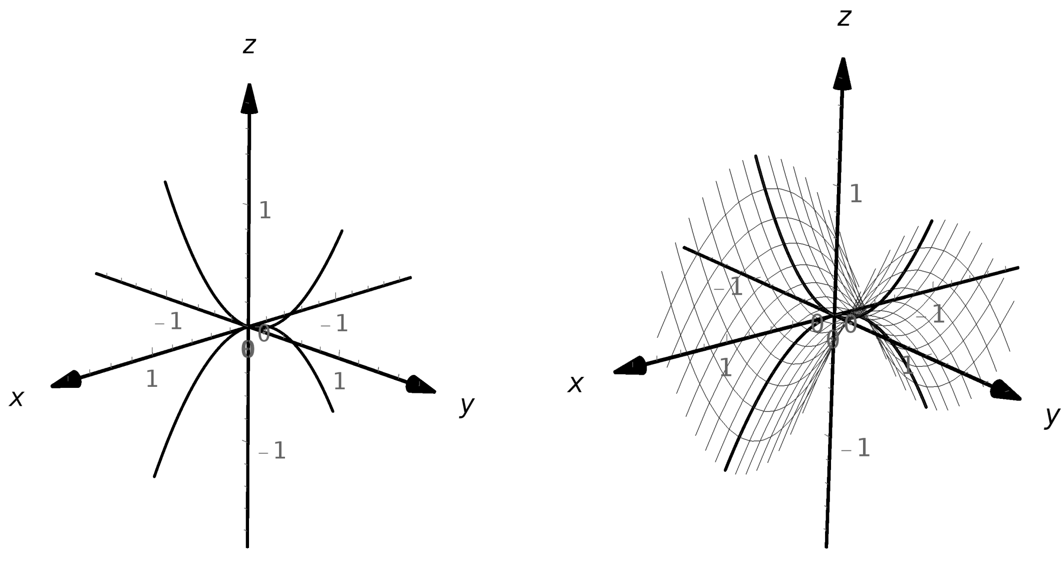
Figure 7.15: Sketching the ellipsoid from Example 7.7.2.

3. This defines a hyperbolic paraboloid. Consider the traces in the yz- and xz-planes:

$x = 0$: The trace is $z = y^2$, a parabola opening up in the yz-plane.

$y = 0$: The trace is $z = -x^2$, a parabola opening down in the xz-plane.

Sketching these two parabolas gives a sketch like that in Figure 7.16(a), and filling in the surface gives a sketch like in Figure 7.16(b).



(a) (b)

Figure 7.16: Sketching the hyperbolic paraboloid from Example 7.7.3.

7.4 Exercises

Lines

Assignment 7.1 — Prove that the points $A=(1,2,3)$, $B=(2,3,4)$ and $C=(3,4,5)$ are on the same line.

Assignment 7.2 — Determine the parametric representation of the line that is the intersection of the following planes:

$$\begin{aligned} p_1: &\quad x-2y+3z=0,\\ p_2: &\quad 2x+3y-4z=4. \end{aligned}$$

Assignment 7.3 — Determine the cartesian equations, parameter equations, and vector equation of the line ℓ_1 in each of the following cases.

(a) through $O=(0,0,0)$ and in the direction of vector $\vec{\boldsymbol{d}}=(1,2,3)$

(b) through $A=(3,4,1)$ and $B=(1,4,0)$

(c) through $A=(1,2,0)$ and parallel to

$$\ell_2: \frac{2x+2}{3}=\frac{y-1}{2}=\frac{2z+3}{1}$$

(d) through $A=(1,2,3)$ and parallel to $\vec{\boldsymbol{d}}=(2,-3,-4)$

(e) through $A = (-1,0,1)$ and perpendicular to $p: 2x - y + 7z = 12$

(f) through $O = (0,0,0)$ and parallel to the intersection of $p_1: x + 2y - z = 2$ and $p_2: 2x - y + 4z = 5$

(g) through $A = (2,-1,-1)$ and parallel with $p_1: x + y = 0$ and $p_2: x - y + 2z = 0$.

Assignment 7.4 — Determine for all cases below if the lines ℓ_1 and ℓ_2 are parallel, (perpendicular) intersecting or skew. If they are intersecting, determine the coordinates of their intersection.

(a) $\ell_1: \dfrac{x-2}{3} = \dfrac{y-2}{4} = \dfrac{z}{2}$ and $\ell_2: \dfrac{x}{6} = \dfrac{3y+2}{24} = \dfrac{3z+4}{12}$

(b) $\ell_1: \begin{cases} 2x - 3y + z = 0 \\ x + y + z = 0 \end{cases}$ and $\ell_2: \begin{cases} 2x - y = 1 \\ x + 2y - z = 0 \end{cases}$

(c) $\ell_1: \begin{cases} 2x + 3y - z = 4 \\ x - y + 3z = 1 \end{cases}$ and $\ell_2: \dfrac{x-2}{8} = \dfrac{y-1}{-7} = \dfrac{-z+5}{5}$

(d) $\ell_1: \begin{cases} x + y + z = 1 \\ x + 2z = 0 \end{cases}$ and $\ell_2: \begin{cases} x + y = 4 \\ z = 0 \end{cases}$

Planes

Assignment 7.5 — Determine the cartesian equation of the plane p in each of the following cases.

(a) through $P = (1,4,2)$ and with normal vector $\vec{\boldsymbol{n}} = (3,1,-4)$

(b) through $P_1 = (1,-2,1)$, $P_2 = (2,0,3)$ and $P_3 = (0,1,-1)$

(c) through $P = (1,2,-3)$ and perpendicular to

$$\ell: \begin{cases} x = t \\ y = -2 - 2t \\ z = 1 + 3t \end{cases},$$

with parameter $t \in \mathbb{R}$.

(d) through $P = (0,0,1)$ and perpendicular to

$$\ell: \frac{2x+2}{1} = \frac{y-1}{3} = \frac{z+1}{-2}$$

(e) through the lines

$$\ell_1: \frac{x+1}{2} = \frac{y-2}{3} = \frac{z-1}{1} \quad \text{and} \quad \ell_2: \frac{x+1}{1} = \frac{y-2}{-1} = \frac{z-1}{2}$$

(f) through $P_1 = (1,1,1)$ and $P_2 = (2,0,3)$ and perpendicular to $p: x + 2y - 3z = 0$

(g) through the cross-section of $p_1 : 2x+3y-z=0$ and $p_2 : x-4y+2z=-5$ and through $P=(-2,0,-1)$

Assignment 7.6 — We consider the planes $p_1 : 3x+2y-z+3=0$, $p_2 : -x+2y+z-3=0$ and $p_3 : 6x+4y-2z-3=0$.

(a) Show that p_1 en p_2 are perpendicular.

(b) Show that p_1 en p_3 are parallel.

(c) What can you conclude about the location of p_2 and p_3 relative to each other? Solve this question without doing any math.

Assignment 7.7 — Determine the angle between $p_1 : 3x-2y+z-4=0$ and $p_2 : x+4y-3z-2=0$.

Three-dimensional objects

Assignment 7.8 — Name and draw the surfaces below.

(a) $3x^2-2y^2+z^2+3=0$

(b) $-4x^2+2z^2-3=0$

(c) $3x^2-y^2=z^2$

(d) $(x-1)^2+(y-2)^2=(z-4)^2$

(e) $2x^2+3z^2=1$

(f) $y^2+2z^2=x$

(g) $x^2+4y^2+9z^2=36$

(h) $\frac{25}{9}x^2-25y^2+z^2=25$

(i) $y=4z^2-x^2$

(j) $x^2+y^2+4z^2+2x=0$

(k) $9x^2+y^2-4z^2+2y=0$

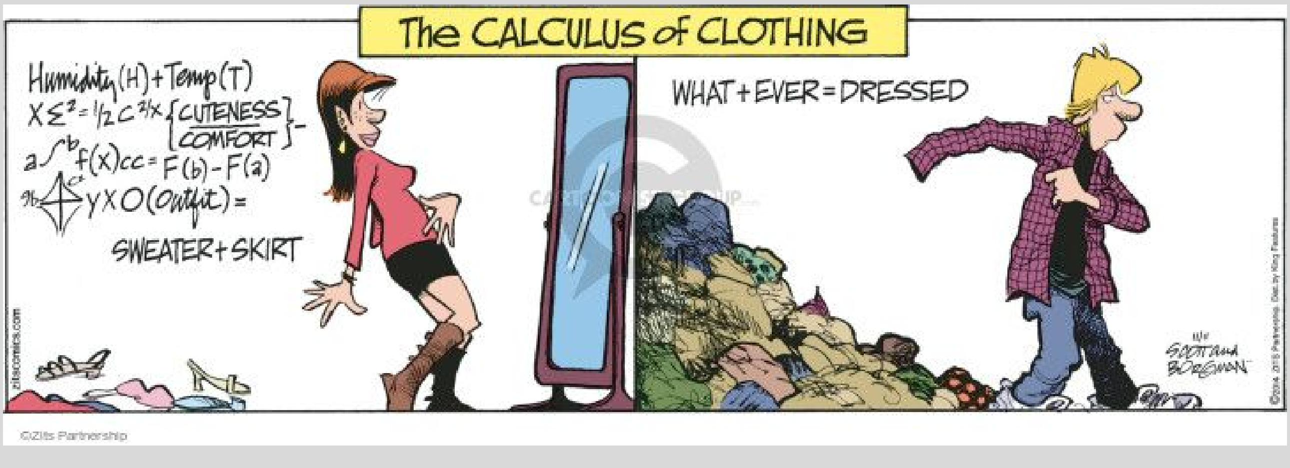
(l) $x-2z^2=0$

Review exercises

Assignment 7.9 — Determine the equation of the line through $P=(2,-1,5)$ and perpendicular to $p: 3x+2y-2z-7=0$.

PART II

SINGLE VARIABLE CALCULUS



Mathematics as an expression of the human mind reflects the active will, the contemplative reason, and the desire for aesthetic perfection. Its basic elements are logic and intuition, analysis and construction, generality and individuality.

— Richard Courant —

8

Limits and continuity

Calculus means a method of calculation or reasoning. When one computes the sales tax on a purchase, one employs a simple calculus. Proving a theorem in geometry employs another calculus. Despite the wonderful advances in mathematics that had taken place into the first half of the 17^{th} century, mathematicians and scientists were keenly aware of what they could not do. In particular, two important concepts eluded mastery by the great thinkers of that time: area and rates of change.

Area seems innocuous enough; areas of circles, rectangles, parallelograms, etc., are standard topics of study for students today just as they were then. However, the areas of arbitrary shapes could not be computed, even if the boundary of the shape could be described exactly. Rates of change were also important. When an object moves at a constant rate of change, then distance = rate × time. But what if the rate is not constant – can distance still be computed?

It turns out that these two concepts were related. Two mathematicians, Sir Isaac Newton and Gottfried Leibniz, are credited with independently formulating a system of computing that solved the above problems and showed how they were connected. Their system of reasoning was “a” calculus. However, as the power and importance of their discovery took hold, it became known to many as “the” calculus. Today, we generally shorten this to discuss calculus. The foundation of the calculus is the **limit** (*limiet*). It is a tool to describe a particular behaviour of a function. This chapter begins our study of the limit by approximating its value graphically and numerically. After a formal definition of the limit, properties are established that make finding limits tractable. Once the limit is understood, then the problems of area and rates of change can be approached.

Leibniz versus Newton

As stated today, both Newton and Leibniz are credited for developing calculus independently. Yet, the question “Who was the first?” was a major intellectual controversy back in the beginning of the Eighteenth century. It is known as the Leibniz–Newton calculus controversy and divided the mathematical community into two bickering groups for years after.

8.1 An intuitive introduction

8.1.1 Limits and their approximation

Consider the function

$$y = \frac{\sin(x)}{x}. \tag{8.1}$$

We could ask ourselves the question: When x is near the value 1, what value (if any) is y near?

While our question is not precisely formed (what constitutes "near the value 1"?), the answer does not seem difficult to find. One might think first to look at a graph of this function to approximate the appropriate y values. Consider Figure 8.1(a), where this function is graphed. For values of x near 1, it seems that y takes on values near 0.85. In fact, when $x = 1$, then $y = \sin(1)/1 \approx 0.84$, so it makes sense that when x is near 1, y will be near 0.84.

Consider this again at a different value for x. When x is near 0, what value (if any) is y near? By considering Figure 8.1(b), one can see that it seems that y takes on values near 1. But what happens when $x = 0$? We have

$$y \to \frac{\sin(0)}{0} \to \frac{0}{0}.$$

The expression $0/0$ has no value; it is **indeterminate** (*onbepaald*). Such an expression gives no information about what is going on with the function nearby. We cannot find out how y behaves near $x = 0$ for this function simply by letting $x = 0$.

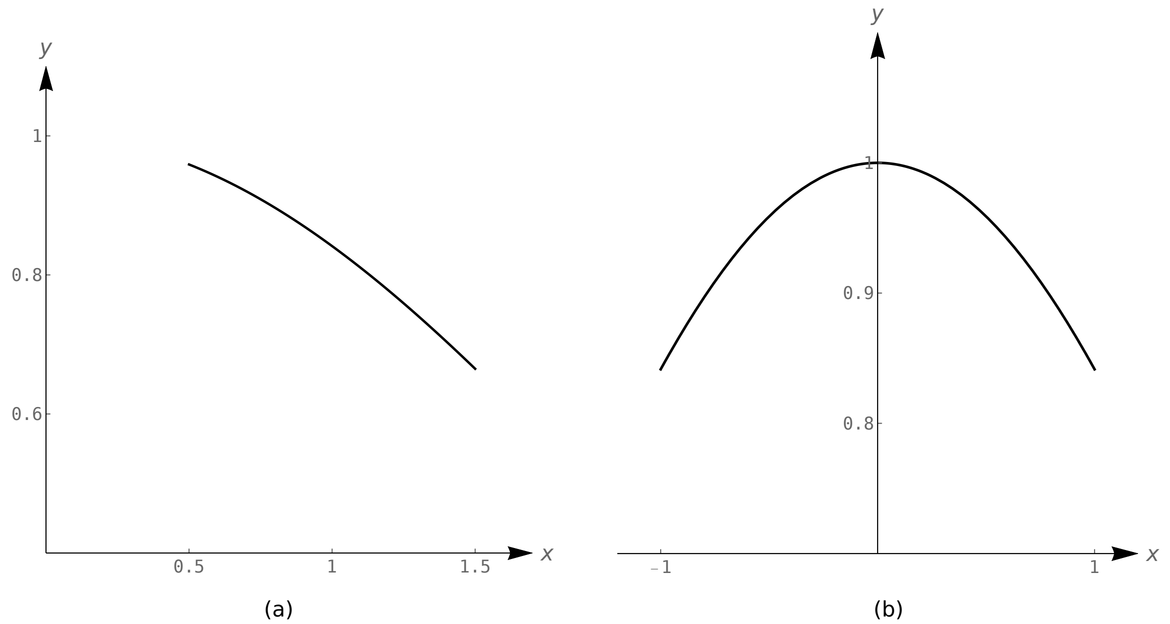


Figure 8.1: The graph of $y = \frac{\sin(x)}{x}$ near $x = 1$ (a) and $x = 0$ (b).

Finding a limit entails understanding how a function behaves near a particular value of x. Before continuing, it will be useful to establish some notation. Let $y = f(x)$; that is, let y be a function of x for some function f. The expression "the limit of y as x approaches 1" describes a number, often referred to as L, that y nears as x nears 1. We write all this as

$$\lim_{x\to 1} y = \lim_{x\to 1} f(x) = L.$$

This is not a formal definition, but an intuitive one to settle the mind. It allows us to approximate limits both graphically and numerically.

For what concerns the function defined by Equation (8.1), we approximated graphically that

$$\lim_{x\to 1}\frac{\sin(x)}{x}\approx 0.84 \quad \text{and} \quad \lim_{x\to 0}\frac{\sin(x)}{x}\approx 1.$$

8.1.2 Existence of limits

A function may not have a limit for all values of x. That is, we cannot say $\lim_{x\to c} f(x) = L$ for some numbers L for all values of c, for there may not be a number that $f(x)$ is approaching. There are three common ways in which a limit may fail to exist.

1. The function $f(x)$ may approach different values on either side of c.
2. The function may grow without upper or lower bound as x approaches c.
3. The function may oscillate as x approaches c without approaching a specific value.

Each of these cases is illustrated in the following examples

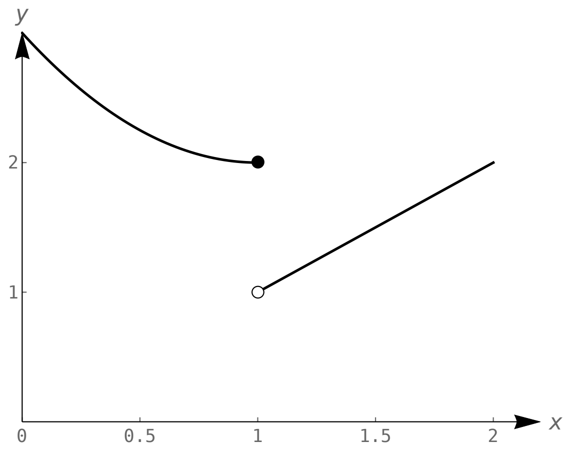
Example 8.1

Let us consider the function

$$f(x) = \begin{cases} x^2-2x+3, & \text{if} \quad x\le 1, \\ x, & \text{if} \quad x>1, \end{cases} \tag{8.2}$$

and try to determine $\lim_{x\to 1} f(x)$.

A graph of $f(x)$ around $x = 1$ and a corresponding table are given in Figure 8.2(a) and 8.2(b), respectively. It is clear that as x approaches 1, $f(x)$ does not seem to approach a single number. Instead, it seems as though $f(x)$ approaches two different numbers. When considering values of x less than 1, so approaching 1 from the left, it seems that $f(x)$ is approaching 2; when considering values of x greater than 1 ,so approaching 1 from the right, it seems that $f(x)$ is approaching 1. Consequently, the limit does not exist since $f(x)$ is not approaching one value as x approaches 1.



(a)

x	$f(x)$
0.9	2.01
0.99	2.0001
0.999	2.000001
1.001	1.001
1.01	1.01
1.1	1.1

(b)

Figure 8.2: Graphically (a) and numerically (b) approximating $\lim_{x\to 1} f(x)$ for f given by Equation (8.2).

Example 8.2

Let us now have a closer look at

$$\lim_{x\to 1} \frac{1}{(x-1)^2}.$$

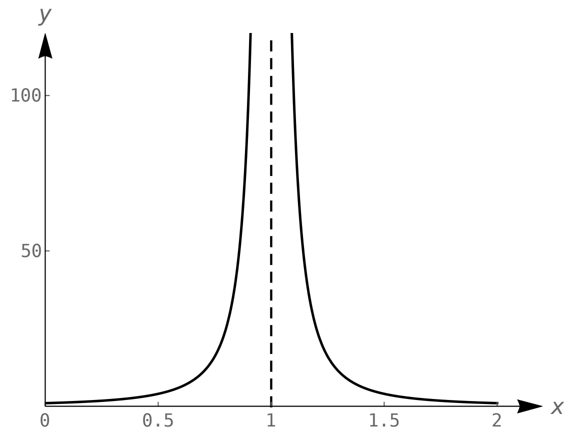
A graph and table of $f(x) = 1/(x-1)^2$ are given in Figure 8.3(a) and 8.3(b), respectively. Both show that as x approaches 1, $f(x)$ grows larger and larger. Indeed, if x is near 1, then $(x-1)^2$ is very small, and:

$$\frac{1}{\text{very small number}} = \text{very large number}.$$

Since $f(x)$ is not approaching a single number, we conclude that

$$\lim_{x\to 1} \frac{1}{(x-1)^2}$$

does not exist.



(a)

x	$f(x)$
0.9	100.
0.99	10000.
0.999	$1. \times 10^6$
1.001	$1. \times 10^6$
1.01	10000.
1.1	100.

(b)

Figure 8.3: Graphically (a) and numerically (b) approximating $\lim_{x\to 1} \frac{1}{(x-1)^2}$.

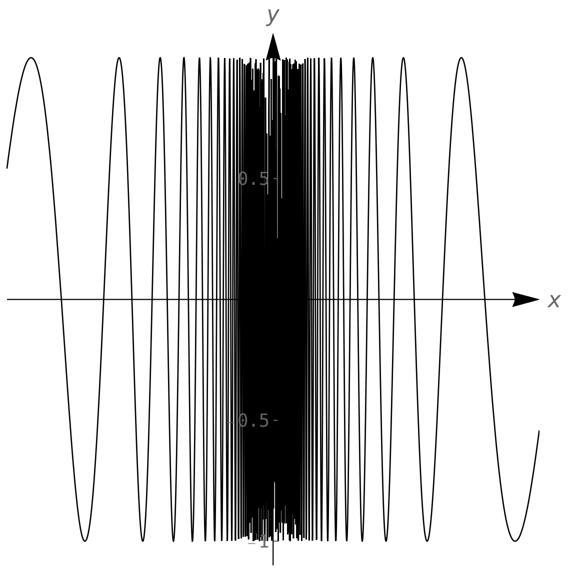
Example 8.3

Let us finally explore why

$$\lim_{x\to 0} \sin\left(\frac{1}{x}\right)$$

does not exist.

For that purpose, Figure 8.4(a) shows $f(x)$ on the interval $[-0.1, 0.1]$; notice how $f(x)$ clearly seems to oscillate near $x = 0$. This is confirmed in Table 8.4(b), where we see $\sin(1/x)$ evaluated for values of x near 0. As x approaches 0, $f(x)$ does not appear to approach any value. It can be shown that in reality, as x approaches 0, $\sin(1/x)$ takes on all values between -1 and 1 infinitely many times. Because of this oscillation, so the considered limit does not exist.



(a)

x	$\sin(1/x)$
0.1	−0.544021
0.01	−0.506366
0.001	0.82688
0.0001	−0.305614
$1. \times 10^{-5}$	0.0357488
$1. \times 10^{-6}$	−0.349994
$1. \times 10^{-7}$	0.420548

(b)

Figure 8.4: Graphically (a) and numerically (b) approximating $\lim_{x \to 0} \sin\left(\frac{1}{x}\right)$.

8.1.3 Limits of difference quotients

Let $f(x)$ represent the position function, in metres, of some particle that is moving in a straight line, where x is measured in seconds. Let us say that when $x = 1$, the particle is at position 10 m, and when $x = 5$, the particle is at 20 m. Another way of expressing this is to say

$$f(1) = 10 \quad \text{and} \quad f(5) = 20.$$

Since the particle travelled 10 metres in 4 seconds, we can say the particle's average velocity was 2.5 m/s. We write this using a quotient of differences, or, a **difference quotient** (*differentiequotiënt*):

$$\frac{f(5) - f(1)}{5 - 1} = \frac{10}{4} = 2.5\text{m/s}.$$

In fact, we are finding in this way the slope of the **secant line** (*snijlijn*) through those two points.

Now consider finding the average speed on another time interval. We again start at $x = 1$, but consider the position of the particle h seconds later. That is, consider the positions of the particle when $x = 1$ and when $x = 1 + h$. The difference quotient is now

$$\frac{f(1+h) - f(1)}{(1+h) - 1} = \frac{f(1+h) - f(1)}{h}. \tag{8.3}$$

Let now

$$f(x) = -1.5x^2 + 11.5x;$$

for which it holds that $f(1) = 10$ and $f(5) = 20$. We can compute this difference quotient for all values of h except $h = 0$, for then we get 0/0, an indeterminate form. For all values $h \neq 0$, the difference quotient computes the average velocity of the particle over an interval of time of length h starting at $x = 1$. For small values of h, i.e., values of h close to 0, we get average velocities over very short time periods and compute secant lines over small intervals (Figure 8.5). This leads us to wonder what the limit of the difference quotient is as h approaches 0. That is,

$$\lim_{h\to 0} \frac{f(1+h)-f(1)}{h} = \ ? \ .$$

As we do not yet have a true definition of a limit nor an exact method for computing it, we settle for approximating the value in Table 8.1. This table gives us reason to assume the value of the limit is about 8.5.

Table 8.1: The difference quotient given by Equation (8.3) evaluated at values of h near 0.

h	$\frac{f(1+h)-f(1)}{h}$
−0.5	9.25
−0.1	8.65
−0.01	8.515
0.01	8.485
0.1	8.35
0.5	7.75

Proper understanding of limits is key to understanding calculus. With limits, we can accomplish seemingly impossible mathematical things, like adding up an infinite number of numbers (and not get infinity) and finding the slope of a line between two points, where the two points are actually the same point. These are not just mathematical curiosities; they allow us to link position, velocity and acceleration together, connect cross-sectional areas to volume, find the work done by a variable force, and much more.

In the next section we give the formal definition of the limit and begin our study of finding limits analytically.

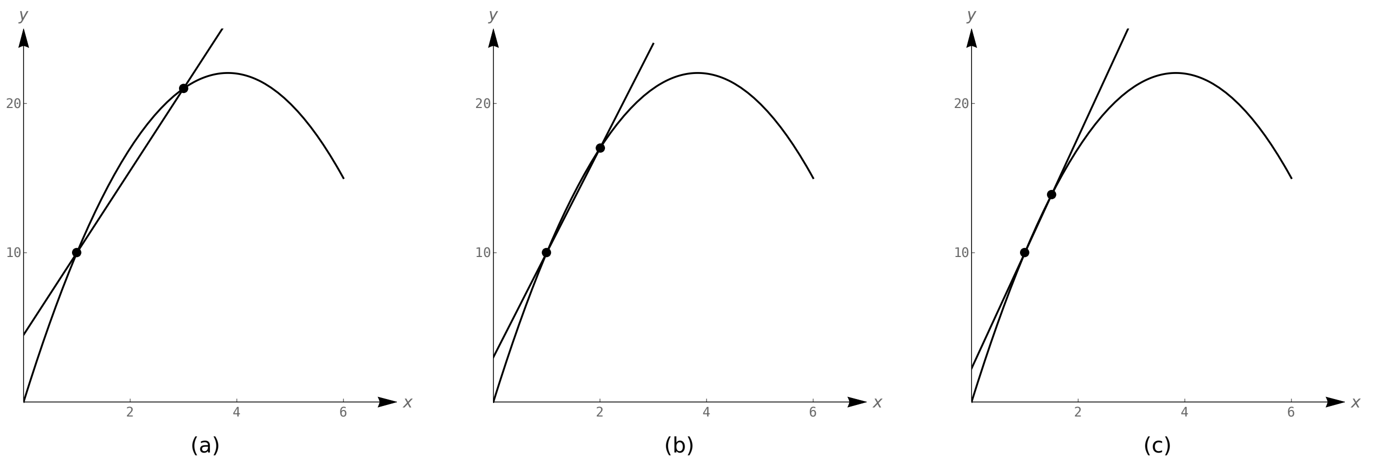


Figure 8.5: Secant lines of $f(x)$ at $x=1$ and $x=1+h$, for shrinking values of h.

8.2 Epsilon-delta definition of a limit

This section introduces the formal definition of a limit, typically called the **epsilon-delta definition** (*epsilon-delta definitie*), referring to the letters ε and δ of the Greek alphabet.

Given a function $y = f(x)$ and an x-value, c, an informal way of describing a limit based on our findings in Section 8.1 could be:

> The limit of the function f, as x approaches c, is a value L if y is near L whenever x is near c.

Or formulated in a more quantitative way:

> If x is within a certain tolerance level of c, then the corresponding value $y = f(x)$ is within a certain tolerance level of L.

Since the traditional notation for the x-tolerance is the lowercase Greek letter delta (δ), and the y-tolerance is denoted by lowercase epsilon, (ε) we may reformulate our definition as:

> If x is within δ units of c, then the corresponding value of y is within ε units of L.

Using the absolute value to express the tolerance 'x is within δ units of c' mathematically, i.e. as

$$|x - c| < \delta\,,$$

we arrive at following formal statement to define a limit

$$|x - c| < \delta \longrightarrow |y - L| < \varepsilon\,.$$

Note that δ and ε, being tolerances, can be any positive (but typically small) values. The full formal definition of a limit we arrive at in this way is given below.

Definitie 8.1 (The limit of a function f)
Let I be an open interval containing c, and let f be a function defined on I, except possibly at c. The limit of $f(x)$, as x approaches c, is L, denoted by

$$\lim_{x \to c} f(x) = L,$$

and means that given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all x in I, where $x \neq c$, if $|x - c| < \delta$, then $|f(x) - L| < \varepsilon$.

Note the order in which ε and δ are given. In the definition, the y-tolerance ε is given first and then the limit will exist if we can find an x-tolerance δ that works. The (ε, δ) definition is illustrated in Figure 8.6.

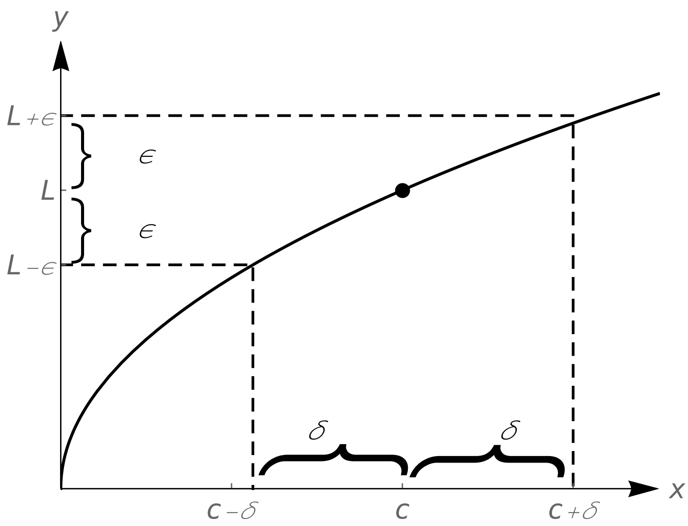


Figure 8.6: Illustrating the (ε, δ) definition.

8.3 Finding limits analytically

Recognizing that (ε, δ)-proofs are cumbersome, this section gives a series of theorems which allow us to find limits much more quickly and intuitively.

8.3.1 Properties of limits

The following properties of limits indicate that already established limits do behave nicely. For that purpose, let b, c, L_1 and L_2 be real numbers, let n be a positive integer, and let f and g be functions defined on an open interval I containing c with the following limits:

$$\lim_{x\to c} f(x) = L_1 \qquad \text{and} \qquad \lim_{x\to c} g(x) = L_2.$$

Then, the following limits hold.

1. **Constants**: $\lim_{x\to c} b = b$
2. **Identity**: $\lim_{x\to c} x = c$
3. **Sums/Differences**: $\lim_{x\to c} (f(x) \pm g(x)) = L_1 \pm L_2$
4. **Scalar multiples**: $\lim_{x\to c} b f(x) = bL_1$
5. **Products**: $\lim_{x\to c} f(x)g(x) = L_1 L_2$
6. **Quotients**: $\lim_{x\to c} \dfrac{f(x)}{g(x)} = \dfrac{L_1}{L_2}$, $(L_2 \neq 0)$
7. **Powers**: $\lim_{x\to c} f(x)^n = L_1^n$
8. **Roots**: $\lim_{x\to c} \sqrt[n]{f(x)} = \sqrt[n]{L_1}$, where $f(x) \geq 0$ on I if n is even.

For what concerns function composition, we get

$$\lim_{x\to c} g(f(x)) = N\,,$$

provided

$$\lim_{x\to c} f(x) = M, \qquad \lim_{x\to M} g(x) = N \qquad \text{and} \qquad g(M) = N.$$

We use these properties in the following example.

Example 8.4

Let

$$\lim_{x\to 2} f(x) = 2, \quad \lim_{x\to 2} g(x) = 3 \quad \text{and} \quad p(x) = 3x^2 - 5x + 7.$$

Find the following limits:

1. $\lim_{x\to 2} \big(f(x) + g(x)\big)$
2. $\lim_{x\to 2} \big(5f(x) + g(x)^2\big)$
3. $\lim_{x\to 2} p(x)$.

Solution

1. Using the sum/difference rule, we know that

$$\lim_{x\to 2} \big(f(x) + g(x)\big) = 2 + 3 = 5.$$

2. Using the scalar multiple and sum/difference rules, we find that

$$\lim_{x\to 2} \big(5f(x) + g(x)^2\big) = 5\cdot 2 + 3^2 = 19.$$

3. Here we combine the power, scalar multiple, sum/difference and constant rules:

$$\begin{aligned}\lim_{x\to 2} p(x) &= \lim_{x\to 2} (3x^2 - 5x + 7)\\ &= \lim_{x\to 2} 3x^2 - \lim_{x\to 2} 5x + \lim_{x\to 2} 7\\ &= 3\cdot 2^2 - 5\cdot 2 + 7\\ &= 9.\end{aligned}$$

We can also verify this result with Mathematica, using the built-in command **`Limit`** as follows.

```
In[7]:=  Limit[3*x^2-5*x+7,x→2]

Out[7]=  9
```

Part 3 of Example 8.4 demonstrates how the limit of a quadratic polynomial can be determined using the properties of limits. Not only that, recognize that

$$\lim_{x\to 2} p(x) = 9 = p(2);$$

i.e., the limit at 2 was found just by plugging 2 into the function. This holds true for all polynomials, and also for rational functions , as stated in the following theorem.

Theorem 8.1 (Limits of polynomials and rational functions)
Let $p(x)$ and $q(x)$ be polynomials and c a real number. Then:

1. $\lim\limits_{x\to c} p(x) = p(c)$
2. $\lim\limits_{x\to c} \dfrac{p(x)}{q(x)} = \dfrac{p(c)}{q(c)}$, *where $q(c) \neq 0$.*

Likewise, for what concerns irrational functions we have the following theorem.

Theorem 8.2 (Limits of irrational functions)
Let f be an irrational function and c a real number. Then:

$$\lim_{x\to c} f(x) = f(c),$$

provided $c \in \operatorname{dom} f$.

It was likely frustrating in Section 8.2 to do a lot of work to prove that

$$\lim_{x\to 2} x^2 = 4,$$

as this seemed fairly obvious. Theorem 8.1 shows, however, that polynomial and rational functions behave in an obvious fashion.

Polynomial and rational functions are not the only functions to behave in such a predictable way. The same holds true for the power, exponential, logarithmic and trigonometric functions we studied in Chapters 4 and 5.

Example 8.5

Evaluate the following limits.

1. $\lim\limits_{x\to \pi} \cos(x)$
2. $\lim\limits_{x\to 3} \left(\sec^2(x) - \tan^2(x)\right)$
3. $\lim\limits_{x\to 1} e^{\ln(x)}$
4. $\lim\limits_{x\to 0} \dfrac{\sin(x)}{x}$

Solution

1. We easily see
$$\lim_{x\to\pi} \cos(x) = \cos(\pi) = -1.$$
2. We immediately have:
$$\lim_{x\to 3} \left(\sec^2(x) - \tan^2(x)\right) = \sec^2(3) - \tan^2(3).$$
Using the Pythagorean identity given by Equation (5.8), this last expression is 1; therefore
$$\lim_{x\to 3} \left(\sec^2(x) - \tan^2(x)\right) = 1.$$

3. We use the exponential/logarithmic identity that $e^{\ln(x)} = x$ and evaluate

$$\lim_{x\to 1} e^{\ln(x)} = \lim_{x\to 1} x = 1.$$

4. We encountered this limit in Section 8.1. Applying our theorems, we attempt to find the limit as

$$\lim_{x\to 0} \frac{\sin(x)}{x} \to \frac{\sin(0)}{0} \to \frac{0}{0}.$$

This, of course, violates the condition of quotient property, as the limit of the denominator is not allowed to be 0. Therefore, we are still unable to evaluate this limit with tools we currently have at hand.

8.3.2 The squeeze theorem

By knowing certain limits of functions, we can find limits involving sums, products, powers, etc., of these functions. We further the development of such comparative tools with the squeeze theorem, a clever and intuitive way to find the value of some limits.

Before stating this theorem formally, suppose we have functions f, g and h where g always takes on values between f and h; that is, for all x in an interval,

$$f(x) \leq g(x) \leq h(x).$$

If f and h have the same limit at c, and g is always squeezed between them, then g must have the same limit as well. That is essentially what the **squeeze theorem** (*insluitstelling*) states.

Theorem 8.3 (Squeeze theorem)
Let f, g and h be functions on an open interval I containing c such that for all x in I,

$$f(x) \leq g(x) \leq h(x).$$

If

$$\lim_{x\to c} f(x) = L = \lim_{x\to c} h(x),$$

then

$$\lim_{x\to c} g(x) = L.$$

It can take some work to figure out appropriate functions by which to squeeze a given function. However, that is generally the only place where work is necessary; the theorem makes the evaluating the limit part very simple. We use this theorem in the following example.

Example 8.6

Show that

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1.$$

Solution

We begin by considering the unit circle (Section 5.3). Remember that each point on the unit circle has coordinates $(\cos(\theta), \sin(\theta))$ for some angle θ (Figure 8.7). Using similar triangles, we can extend the line from the origin through the point to the point $(1, \tan(\theta))$, as shown. Here we are assuming that $0 \le \theta \le \pi/2$. Later we will show that we can also consider $\theta \le 0$.

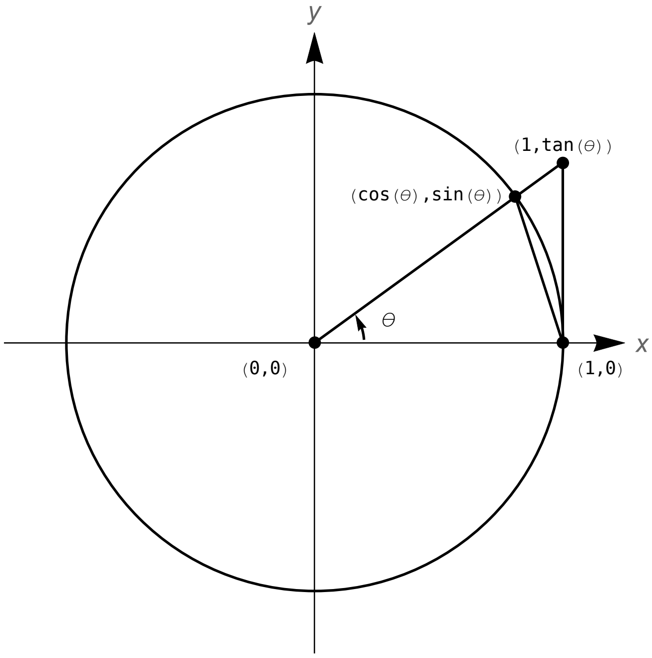
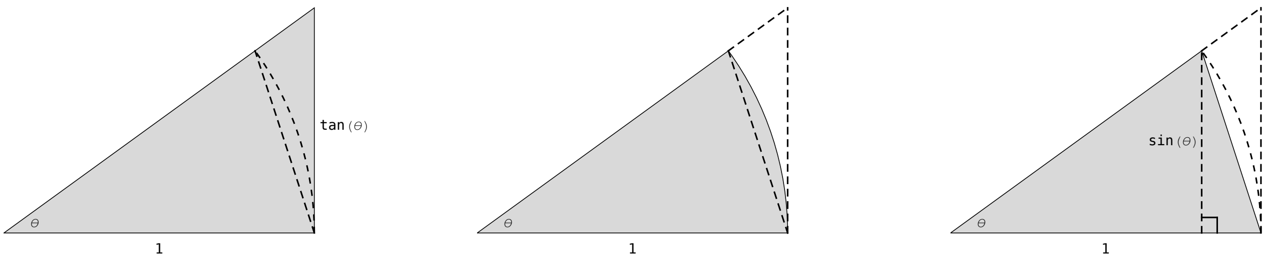


Figure 8.7: The unit circle and related triangles.

Figure 8.7 shows three regions have been constructed in the first quadrant, two triangles and a sector of a circle, which are also drawn below. The area of the large triangle is $\tan(\theta)/2$; the area of the sector is $\theta/2$; the area of the triangle contained inside the sector is $\sin(\theta)/2$ (Figure 5.11). It is then clear from the diagram that



$$\frac{\tan(\theta)}{2} \qquad \ge \qquad \frac{\theta}{2} \qquad \ge \qquad \frac{\sin(\theta)}{2}$$

Multiplying all terms in this inequality by $2\sin^{-1}(\theta)$, yields

$$\frac{1}{\cos(\theta)} \ge \frac{\theta}{\sin(\theta)} \ge 1.$$

Taking reciprocals reverses the inequalities, giving

$$\cos(\theta) \le \frac{\sin(\theta)}{\theta} \le 1.$$

Not that these inequalities hold for all values of θ near 0, even negative values, since $\cos(-\theta) = \cos(\theta)$ and $\sin(-\theta) = -\sin(\theta)$.

Now take limits for $\theta \to 0$.

$$\begin{aligned}
\lim_{\theta\to 0}\cos(\theta) &\leq \lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} \leq \lim_{\theta\to 0} 1\\
\cos(0) &\leq \lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} \leq 1\\
1 &\leq \lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} \leq 1
\end{aligned}$$

Clearly, Theorem 8.3 guarantees that

$$\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} = 1.$$

Actually, this limit tells us more than just that as x approaches 0, $\sin(x)/x$ approaches 1. Both x and $\sin(x)$ are approaching 0, but the ratio of x and $\sin(x)$ approaches 1, meaning that they are approaching 0 in essentially the same way. So for small x, the functions $y = x$ and $y = \sin(x)$ are essentially indistinguishable.

We include this special limit, along with three others, which can be determined in a similar way using Theorem 8.3, in the following theorem.

Theorem 8.4 (Special limits)

1. $\displaystyle\lim_{x\to 0}\frac{\sin(x)}{x} = 1$
2. $\displaystyle\lim_{x\to 0}\frac{\cos(x)-1}{x} = 0$
3. $\displaystyle\lim_{x\to 0}(1+x)^{\frac{1}{x}} = e$
4. $\displaystyle\lim_{x\to 0}\frac{e^x-1}{x} = 1$

A short word on how to interpret the latter three limits in Theorem 8.4. We know that as x goes to 0, $\cos(x)$ goes to 1. So, in the second limit, both the numerator and denominator are approaching 0. However, since the limit is 0, we can interpret this as saying that $\cos(x)$ is approaching 1 faster than x is approaching 0.

In the third limit in Theorem 8.4, inside the parentheses we have an expression that is approaching 1 (though never equalling 1), and we know that 1 raised to any power is still 1. At the same time, the power is growing toward infinity. What happens to a number near 1 raised to a very large power? In this particular case, the result approaches **Euler's number** (*Eulergetal*), e, approximately 2.718. Upon an appropriate change of variables, we can also write this as

$$e = \lim_{x\to\pm\infty}\left(1+\frac{1}{x}\right)^x.$$

In the fourth limit in Theorem 8.4, we see that as $x \to 0$, e^x approaches 1 just as fast as $x \to 0$, resulting in a limit of 1.

Euler's number and interests

Although the symbol e was introduced by Leonhard Euler around 1727, it was Jacob Bernoulli who already discovered this constant in 1683 through the third limit in Theorem 8.4. He came across this special limit while studying a question about compound interest:

A bank account starts with \$1.00 and pays 100 percent interest per year. If the interest is credited once, at the end of the year, the value of the account at year-end will be \$2.00. What happens if the interest is computed and credited more frequently during the year?

If there are n compounding intervals, the interest for each interval will be $100\%/n$ and the value at the end of the year will be $\$1.00\,(1+1/n)^n$.

8.3.3 Limits of functions equal at all but one point

Consider the following limit:

$$\lim_{x\to 1}\frac{x^2-1}{x-1}.$$

We begin by attempting to apply Theorem 8.1 and substituting 1 for x in the quotient. This, however, gives:

$$\lim_{x\to 1}\frac{x^2-1}{x-1}=\frac{1^2-1}{1-1}=\frac{0}{0},$$

an indeterminate form. We cannot apply the theorem.

By graphing the function $y=\frac{x^2-1}{x-1}$ (Figure 8.8), we see that the function seems to be linear, implying that the limit should be easy to evaluate. Recognize that the numerator of our quotient can be factored:

$$\frac{x^2-1}{x-1}=\frac{(x-1)(x+1)}{x-1}.$$

The function is not defined when $x=1$, but for all other x,

$$\frac{x^2-1}{x-1}=\frac{(x-1)(x+1)}{x-1}=\frac{\cancel{(x-1)}(x+1)}{\cancel{x-1}}=x+1.$$

Clearly $\lim\limits_{x\to 1}(x+1)=2$. Recall that when considering limits, we are not concerned with the value of the function at 1, only the value the function approaches as x approaches 1. Since $(x^2-1)/(x-1)$ and $x+1$ are the same at all points except $x=1$, they both approach the same value as x approaches 1. Therefore we may conclude that

$$\lim_{x\to 1}\frac{x^2-1}{x-1}=2.$$

This finding is formalized in the following theorem

Theorem 8.5 (Equality of functions and limits)
Let $g(x)=f(x)$ for all x in an open interval, except possibly at c, and let $\lim\limits_{x\to c}g(x)=L$ for some real number L. Then

$$\lim_{x\to c}f(x)=L.$$

So, when dealing with a rational function of the form $g(x)/f(x)$ and directly evaluating the limit

$$\lim_{x\to c}\frac{g(x)}{f(x)}$$

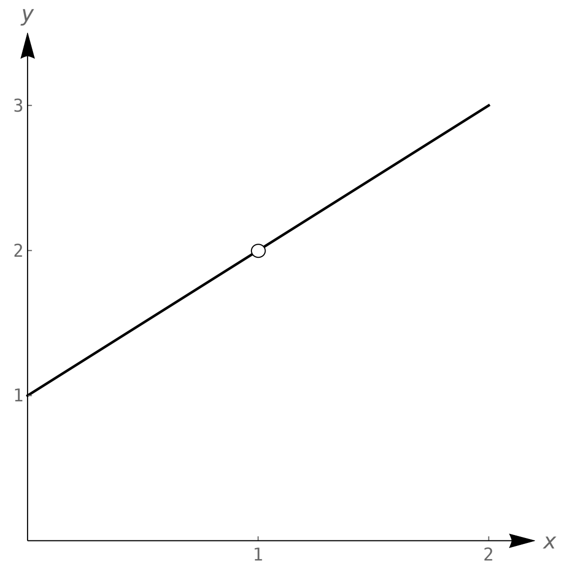


Figure 8.8: The graph of $y = \frac{x^2-1}{x-1}$.

returns 0/0, the fundamental theorem of algebra tells us that $(x-c)$ is a factor of both $g(x)$ and $f(x)$. One can then use algebra to factor this term out, cancel, and then apply Theorem 8.5.

We end this section by revisiting a limit first seen in Section 8.1, a limit of a difference quotient. Let $f(x) = -1.5x^2 + 11.5x$; we approximated the limit

$$\lim_{h\to 0} \frac{f(1+h)-f(1)}{h} \approx 8.5.$$

We now formally evaluate this limit in the following example.

Example 8.7

Let $f(x) = -1.5x^2 + 11.5x$, then find

$$\lim_{h\to 0} \frac{f(1+h)-f(1)}{h}.$$

Solution

Since f is a polynomial, our first attempt should be to employ Theorem 8.1 and substitute 0 for h. However, we see that this gives us $0/0$.

Knowing that we have a rational function hints that some algebra will help. Consider the following:

$$\begin{aligned}
\lim_{h\to 0} \frac{f(1+h)-f(1)}{h} &= \lim_{h\to 0} \frac{-1.5(1+h)^2 + 11.5(1+h) - \left(-1.5(1)^2 + 11.5(1)\right)}{h} \\
&= \lim_{h\to 0} \frac{-1.5(1+2h+h^2) + 11.5 + 11.5h - 10}{h} \\
&= \lim_{h\to 0} \frac{-1.5h^2 + 8.5h}{h} \\
&= \lim_{h\to 0} \frac{h(-1.5h + 8.5)}{h} \\
&= 8.5.
\end{aligned}$$

This matches our previous approximation (see Table 8.1).

This section contains several valuable tools for evaluating limits. One of the main results is that many functions behave in a very nice, predictable way. In Section 8.5 we give a name to this nice behaviour; we label such functions as continuous. Defining that term will require us to look again at what a limit is and what causes limits to not exist.

For the sake of comprehensiveness, We list below the steps that should be taken when evaluating the limit $\lim\limits_{x\to a} f(x)$:

1. Compute $f(a)$.
2. You arrive at one of the following cases:
   - $f(a) \in \mathbb{R}$: the limit is computed.
   - $f(a) = \left(\frac{0}{0}\right)$: try to get $x-a$ as a common factor in the nominator and denominator, possibly after multiplying with its conjugate binomial, then simplify and return to Step 1.
   - $f(a) = \left(\frac{c}{0}\right) = \pm\infty \, (c \neq 0)$: $x = a$ is a vertical asymptote of the function f.

8.4 One-sided limits

Remember from Section 8.1 that one of the ways in which limits of functions fail to exist is when the function approaches different values from the left and right. To explore in depth the concepts underlying this we introduce in this section the **one-sided limit** (*eenzijdige limiet*). We begin with formal definitions that are very similar to Definition 8.1, but the notation is slightly different and $x \neq c$ is replaced with either $x < c$ or $x > c$.

Definitie 8.2 (One-sided limits)
Left-hand Limit (*linkerlimiet*)

Let f be a function defined on $]a, c[$ for some $a < c$ and let L be a real number.

The limit of $f(x)$, as x approaches c from the left, is L, or, the left-hand limit of f at c is L, denoted by

$$\lim_{\substack{x\to c \\ <}} f(x) = L,$$

means given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $a < x < c$, if $|x-c| < \delta$, then $|f(x)-L| < \varepsilon$.

Right-hand Limit (*rechterlimiet*)

Let f be a function defined on $]c, b[$ for some $b > c$ and let L be a real number.

The limit of $f(x)$, as x approaches c from the right, is L, or, the right-hand limit of f at c is L, denoted by

$$\lim_{\substack{x\to c \\ >}} f(x) = L,$$

means given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $c < x < b$, if $|x-c| < \delta$, then $|f(x)-L| < \varepsilon$.

Practically speaking, when evaluating a left-hand limit, we consider only values of x to the left of c, i.e., where $x < c$. The notation $x \underset{<}{\rightarrow} c$ is used to imply that we look at values of x to the left of c. A similar statement holds for evaluating right-hand limits; there we consider only values of x to the right of c, i.e., $x > c$. We can use the theorems from previous sections to help us evaluate these limits; we just restrict our view to one side of c.

We practice evaluating left- and right-hand limits through a series of examples.

Example 8.8

Let the function f_1 be defined by

$$f_1(x) = \begin{cases} x, & \text{if } \ 0 \le x \le 1, \\ 3-x, & \text{if } \ 1 < x < 2. \end{cases}$$

Its graph is shown in Figure 8.9. Find each of the following:

1. $\lim\limits_{x \underset{<}{\rightarrow} 1} f_1(x)$
2. $\lim\limits_{x \underset{>}{\rightarrow} 1} f_1(x)$
3. $\lim\limits_{x \rightarrow 1} f_1(x)$
4. $f_1(1)$
5. $\lim\limits_{x \underset{>}{\rightarrow} 0} f_1(x)$
6. $f_1(0)$
7. $\lim\limits_{x \underset{<}{\rightarrow} 2} f_1(x)$
8. $f_1(2)$

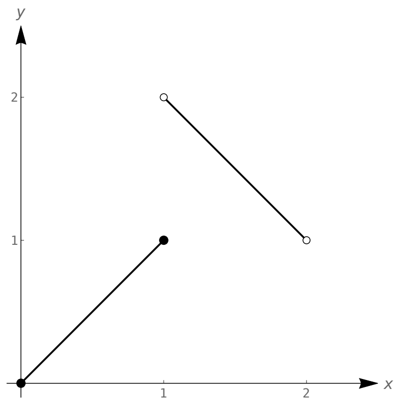


Figure 8.9: The graph of f_1 in Example 8.8.

Solution

For these problems, the visual aid of the graph is likely more effective in evaluating the limits than using f_1 itself.

1. As x goes to 1 from the left, we see that $f_1(x)$ is approaching the value of 1. Therefore $\lim\limits_{x \underset{<}{\rightarrow} 1} f_1(x) = 1$.
2. As x goes to 1 from the right, we see that $f_1(x)$ is approaching the value of 2. Recall that it does not matter that there is an open circle there; we are evaluating a limit, not the value of the function. Therefore $\lim\limits_{x \underset{>}{\rightarrow} 1} f_1(x) = 2$.
3. The limit of f_1 as x approaches 1 does not exist. The function does not approach one particular value, but two different values from the left and the right.
4. Using the definition and by looking at the graph we see that $f_1(1) = 1$.

5. As x goes to 0 from the right, we see that $f_1(x)$ is also approaching 0. Therefore $\lim\limits_{\substack{x\to 0\\ >}} f_1(x) = 0$. Note we cannot consider a left-hand limit at 0 as f_1 is not defined for values of $x < 0$.
6. Using the definition and the graph, $f_1(0) = 0$.
7. As x goes to 2 from the left, we see that $f_1(x)$ is approaching the value of 1. Therefore $\lim\limits_{\substack{x\to 2\\ <}} f_1(x) = 1$.
8. The graph and the definition of the function show that $f_1(2)$ is not defined.

Alternatively, we could again make use of Mathematica to determine the one-sided limits. The option **`Direction`** specifies which one-sided limit should be computed ("FromBelow" and "FromAbove" for left-hand and right-hand limits, respectively). For example, we can compute the limit in 7 as follows:

```
In[8]:= Limit[Piecewise[{{x, 0≤x≤1}, {3-x, 1<x<2}}], x → 2, Direction → "FromBelow"]

Out[8]= 1
```

Note how the left and right-hand limits in the previous examples were different at $x = 1$. This, of course, causes the limit to not exist. The following theorem states what is fairly intuitive: the limit exists precisely when the left and right-hand limits are equal.

Theorem 8.6 (Limits and one sided limits)
Let f be a function defined on an open interval I containing c. Then

$$\lim_{x\to c} f(x) = L$$

if and only if,

$$\lim_{\substack{x\to c\\ <}} f(x) = L \quad \textit{and} \quad \lim_{\substack{x\to c\\ >}} f(x) = L.$$

Throughout these examples pay attention to the fact that the value of the function may/may not be equal to the value(s) of its left/right-hand limits, even when these limits agree.

Example 8.9

Let

$$f_2(x) = \begin{cases} 2-x, & \text{if} \quad 0 < x < 1, \\ (x-2)^2, & \text{if} \quad 1 < x < 2. \end{cases}$$

A graph of this function is shown in Figure 8.10. Evaluate the following.

1. $\lim\limits_{\substack{x\to 1\\ <}} f_2(x)$
2. $\lim\limits_{\substack{x\to 1\\ >}} f_2(x)$
3. $\lim\limits_{x\to 1} f_2(x)$
4. $f_2(1)$
5. $\lim\limits_{\substack{x\to 0\\ >}} f_2(x)$
6. $f_2(0)$
7. $\lim\limits_{\substack{x\to 2\\ <}} f_2(x)$
8. $f_2(2)$

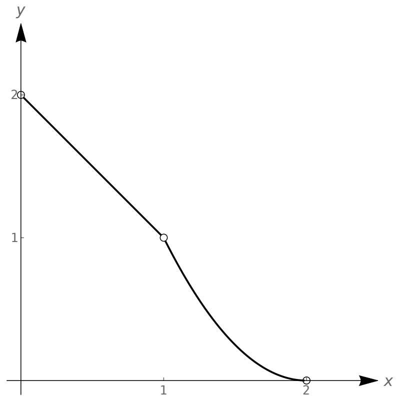


Figure 8.10: A graph of f_2 in Example 8.9.

Solution

We will evaluate each using both the definition of f_2 and its graph.

1. As x approaches 1 from the left, we see that $f_2(x)$ approaches 1. Therefore $\lim\limits_{\substack{x\to 1\\ <}} f_2(x) = 1$.

2. As x approaches 1 from the right, we see that $f_2(x)$ approaches 1. Therefore $\lim\limits_{\substack{x\to 1\\ >}} f_2(x) = 1$.

3. The limit of f_2 as x approaches 1 exists and is 1, as f_2 approaches 1 from both the right and left. Therefore $\lim\limits_{x\to 1} f_2(x) = 1$.

4. $f_2(1)$ is not defined. Note that 1 is not in the domain of f_2 as defined by the problem, which is indicated on the graph by an open circle when $x = 1$.

5. As x goes to 0 from the right, $f_2(x)$ approaches 2. So $\lim\limits_{\substack{x\to 0\\ >}} f_2(x) = 2$.

6. $f_2(0)$ is not defined as 0 is not in the domain of f_2.

7. As x goes to 2 from the left, $f_2(x)$ approaches 0. So $\lim\limits_{\substack{x\to 2\\ <}} f_2(x) = 0$.

8. $f_2(2)$ is not defined as 2 is not in the domain of f_2.

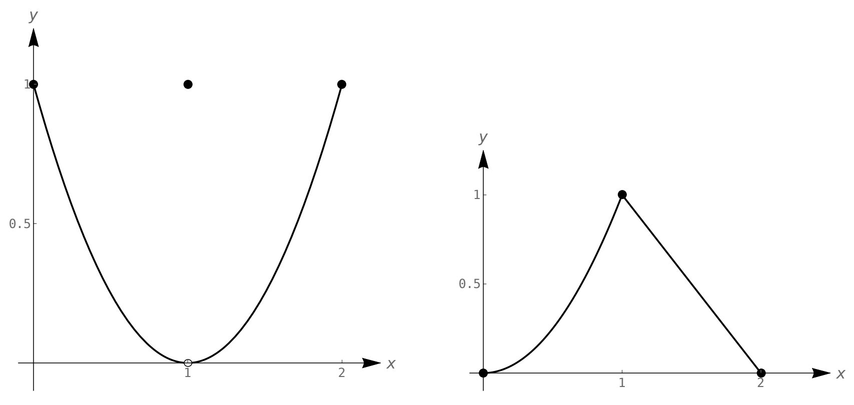
Example 8.10

Consider the following piecewise-defined functions:

1.
$$f_3(x) = \begin{cases} (x-1)^2, & \text{if} \quad 0 \le x \le 2, x \ne 1, \\ 1, & \text{if} \quad x = 1, \end{cases}$$

2.
$$f_4(x) = \begin{cases} x^2, & \text{if} \quad 0 \le x \le 1, \\ 2-x, & \text{if} \quad 1 < x \le 2. \end{cases}$$

Their graphs are shown in Figure 8.11(a) and 8.11(b), respectively. Evaluate for both functions $\lim\limits_{\substack{x\to 1\\ <}} f_i(x)$, $\lim\limits_{\substack{x\to 1\\ >}} f_i(x)$, $\lim\limits_{x\to 1} f_i(x)$ and $f_i(1)$.



(a) (b)

Figure 8.11: A graph of f_3 (a) and f_4 (b) in Example 8.10.

Solution

1. It is clear by looking at the graph that both the left and right-hand limits of f_3, as x approaches 1, are 0. Thus it is also clear that the limit is 0; i.e., $\lim_{x\to 1} f_3(x) = 0$. It is also clearly stated that $f_3(1) = 1$.

2. It is clear from the definition of the function and its graph that all of the following are equal:

$$\lim_{\substack{x\to 1\\ <}} f_4(x) = \lim_{\substack{x\to 1\\ >}} f_4(x) = \lim_{x\to 1} f_4(x) = f_4(1) = 1.$$

In Examples 8.8 – 8.10 we were asked to find both $\lim_{x\to 1} f_i(x)$ and $f_i(1)$. Consider the following table:

	$\lim_{x\to 1} f_i(x)$	$f_i(1)$
Example 8.8	does not exist	1
Example 8.9	1	not defined
Example 8.10.1	0	1
Example 8.10.2	1	1

Only in Example 8.10.2 do both the function and the limit exist and agree. This seems nice; in fact, it seems normal. This is in fact an important situation which we explore in the next section and refers to a function being continuous. In short, a continuous function is one in which when a function approaches a value as $x \to c$ (i.e., when $\lim_{x\to c} f(x) = L$), it actually attains that value at c. Such functions behave nicely as they are very predictable.

8.5 Continuity

8.5.1 Definition

As we have studied limits, we have gained the intuition that limits measure where a function is heading. That is, if $\lim_{x\to 1} f(x) = 3$, then as x is close to 1, $f(x)$ is close to 3. We have seen, though, that this is not necessarily a good indicator of what $f(1)$ actually is. This can be problematic; functions can tend to one value but attain another. This section focuses on functions that do not exhibit such behaviour.

Definitie 8.3 (Continuous function)
Let f be a function defined on an open interval I containing c.

1. f is **continuous** (*continu*) at c if $\lim_{x\to c} f(x) = f(c)$.
2. f is continuous on I if f is continuous at c for all values of c in I. If f is continuous on $\mathbb{R}$, we say f is continuous everywhere.

Note that this definition of continuity (currently) only applies to open intervals.

To establish whether or not a function f is continuous at c one should verify:

1. $\lim_{x\to c} f(x)$ exists,
2. $f(c)$ is defined, and
3. $\lim_{x\to c} f(x) = f(c)$.

Example 8.11

Let f (a) and g (b) be defined as shown in Figures 8.12(a) and 8.12(b), respectively. Give the interval(s) on which these functions are continuous.

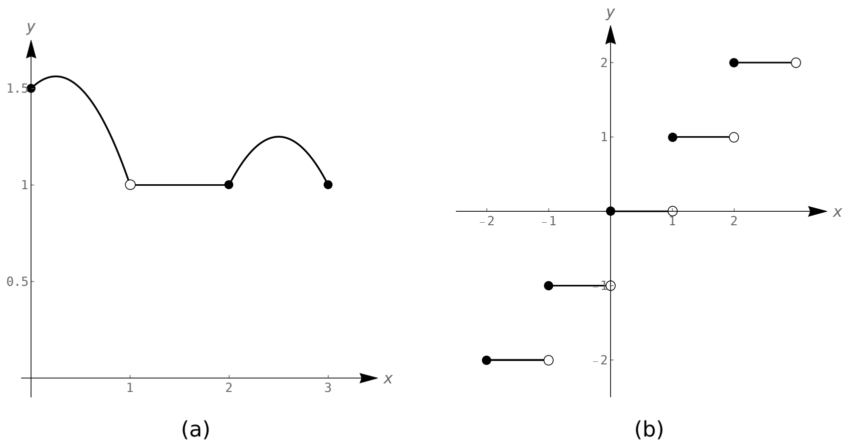


Figure 8.12: The graph of f (a) and g (b) in Example 8.11.

Solution

a. We proceed by examining the three criteria for continuity.
   (a) The limits $\lim_{x\to c} f(x)$ exists for all c between 0 and 3.
   (b) $f(c)$ is defined for all c between 0 and 3, except for $c = 1$. We know immediately that f cannot be continuous at $x = 1$.
   (c) The limit $\lim_{x\to c} f(x) = f(c)$ for all c between 0 and 3, except, of course, for $c = 1$.

   We conclude that f is continuous at every point of $]0,3[$ except at $x = 1$. Therefore f is continuous on $]0,1[$ and $]1,3[$.

b. We examine the three criteria for continuity.

(a) The limits $\lim_{x\to c} g(x)$ do not exist at the jumps from one step to the next, which occur at all integer values of c. Therefore the limits exist for all c except when c is an integer.

(b) The function is defined for all values of c.

(c) The limit $\lim_{x\to c} g(x) = g(c)$ for all values of c where the limit exist, since each step consists of just a line.

We conclude that g is continuous everywhere except at integer values of c. So the intervals on which g is continuous are

$$\ldots,]-2,-1[,]-1,0[,]0,1[,]1,2[, \ldots .$$

Our definition of continuity on an interval specifies the interval is an open interval. We can extend the definition of continuity to closed intervals by considering the appropriate one-sided limits at the endpoints.

Definitie 8.4 (Continuity on closed intervals)
Let f be defined on the closed interval $[a,b]$ for some real numbers $a < b$. f is continuous on $[a,b]$ if:

1. f is continuous on $]a,b[$,
2. $\lim_{\substack{x\to a \\ >}} f(x) = f(a)$ and
3. $\lim_{\substack{x\to b \\ <}} f(x) = f(b)$.

We can of course make the appropriate adjustments to talk about continuity on half–open intervals such as $[a,b[$ or $]a,b]$ if necessary. Also note that we call the function f **right-continuous** (*rechtscontinu*) at a if

$$\lim_{\substack{x\to a \\ >}} f(x) = f(a)\,,$$

and **left-continuous** (*linkscontinu*) at b if

$$\lim_{\substack{x\to b \\ <}} f(x) = f(b)\,,$$

where we of course assumed that the respective one-sided limits exist.

Using this new definition, we can adjust our answer in Example 8.11 by stating that f is continuous on $[0,1[$ and $]1,3]$. Likewise, the function g in Example 8.11 is continuous on the following half–open intervals

$$\ldots, [-2,-1[, [-1,0[, [0,1[, [1,2[, \ldots .$$

Most of the functions you have likely seen in the past are continuous on their domains. This is demonstrated in the following example where we examine the intervals of continuity of a variety of common functions.

Example 8.12

For each of the following functions, give the domain of the function and the interval(s) on which it is continuous.

1. $f(x) = \dfrac{1}{x}$
2. $f(x) = \sqrt{x}$
3. $f(x) = \sqrt{1-x^2}$
4. $f(x) = |x|$

Solution

We examine each in turn.

1. The domain of $f(x) = 1/x$ is $\mathbb{R}_0$. As it is a rational function, we apply Theorem 8.1 together with Definition 8.3 to recognize that f is continuous on all of its domain.

2. The domain of $f(x) = \sqrt{x}$ is $\mathbb{R}^+$. It follows that $f(x) = \sqrt{x}$ is continuous on its domain of $\mathbb{R}^+$.

3. The domain of $f(x) = \sqrt{1-x^2}$ is $[-1,1]$. Using properties of limits shows that f is continuous on all of its domain, $[-1,1]$.

4. The domain of $f(x) = |x|$ is $\mathbb{R}$. We can define the absolute value function as

$$f(x) = \begin{cases} -x, & \text{if} \quad x < 0, \\ x, & \text{if} \quad x \geq 0. \end{cases}$$

   Each piece of this piecewise defined function is continuous on all of its domain, giving that f is continuous on $]-\infty, 0[$ and $[0, +\infty[$. We cannot assume this implies that f is continuous on $\mathbb{R}$; we need to check that $\lim\limits_{x\to 0} f(x) = f(0)$, as $x = 0$ is the point where f transitions from one piece of its definition to the other. It is easy to verify that this is indeed true, hence we conclude that $f(x) = |x|$ is continuous everywhere.

Continuous functions can be combined to form other continuous functions, which is an immediate consequence of the properties of limits. So, if we let f and g be continuous functions on an interval I, c be a real number and n be a positive integer, then the following functions are continuous on I.

1. **Sums/Differences**: $f \pm g$
2. **Constant Multiples**: $c \cdot f$
3. **Products**: $f \cdot g$
4. **Quotients**: f/g (As long as $g \neq 0$ on I.)
5. **Powers**: f^n
6. **Roots**: $\sqrt[n]{f}$ (If n is even then require $f(x) \geq 0$ on I.)

For what concerns function compositions, we consider a function f which is continuous on I, whose range on I is J, and a function g which is continuous on J. Then $g \circ f$, i.e., $g(f(x))$, is continuous on I.

A function f that is not continuous at c is called **discontinuous** (*discontinu*) at c, i.e. there is a **discontinuity** (*discontinuïteit*) at c.

8.5.2 Intermediate value theorem

A common way of thinking of a continuous function is that its graph can be sketched without lifting your pencil. That is, its graph forms a continuous curve, without holes, breaks or jumps. This pseudo–definition glosses, however, over some of the finer points of continuity. Very strange functions are continuous that one would be hard pressed to actually sketch by hand, an example of this being for instance the Weierstrass function (Figure 8.13).

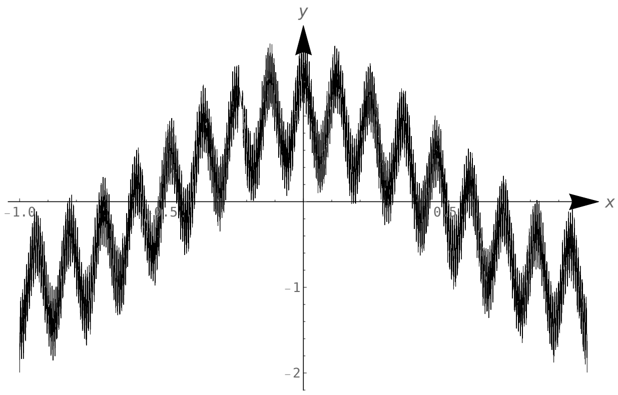


Figure 8.13: The Weierstrass function.

This intuitive notion of continuity does nonetheless help us understand another important concept as follows. Suppose f is defined on $[1,2]$ and $f(1) = -10$ and $f(2) = 5$. If f is continuous on $[1,2]$ (i.e., its graph can be sketched as a continuous curve from $(1,-10)$ to $(2,5)$) then we know intuitively that somewhere on $[1,2]$ f must be equal to -9, and -8, and -7, -6, $\ldots$, 0, $1/2$, etc. In short, f takes on all intermediate values between -10 and 5. It may take on more values; f may actually equal 6 at some time, for instance, but we are guaranteed all values between -10 and 5 will be covered.

This notion seems intuitive and its importance will turn out to be profound. Therefore the concept is stated in the form of a theorem, the so-called **intermediate value theorem** (*tussenwaardestelling*).

Theorem 8.7 (Intermediate value theorem)
Let f be a continuous function on $[a,b]$ and, without loss of generality, let $f(a) < f(b)$. Then for every value u, where $f(a) < u < f(b)$, there is at least one value c in $]\,a,b\,[$ such that $f(c) = u$.

One important application of the intermediate value theorem is root finding. Given a function f, we are often interested in finding values of x where $f(x) = 0$. These roots may be very difficult to find exactly. Good approximations can be found through successive applications of this theorem. Suppose through direct computation we find that $f(a) < 0$ and $f(b) > 0$, where $a < b$. The intermediate value theorem states that there is at least one c in $]\,a,b\,[$ such that $f(c) = 0$. The theorem does not give us any clue as to where to find such a value in the interval $]\,a,b\,[$, just that at least one such value exists.

There is a technique that produces a good approximation of c. Let d be the midpoint of the interval $[a,b]$ and consider $f(d)$. There are three possibilities:

1. $f(d) = 0$: We got lucky and stumbled on the actual value. We stop as we found a root.
2. $f(d) < 0$: Then we know there is a root of f on the interval $[d,b]$ – we have halved the size of our interval, hence are closer to a good approximation of the root.
3. $f(d) > 0$: Then we know there is a root of f on the interval $[a,d]$ – again,we have halved the size of our interval, hence are closer to a good approximation of the root.

Successively applying this technique is called the **bisection method** (*halveringsmethode*) of root finding. We continue until the interval is sufficiently small. We demonstrate this in the following example.

Example 8.13

Approximate the root of $f(x) = x - \cos(x)$, accurate to three places after the decimal.

Solution

Consider the graph of $f(x) = x - \cos(x)$, shown in Figure 8.14(a). It is clear that the graph crosses the x-axis somewhere near $x = 0.8$. To start the bisection method, pick an interval that contains 0.8. We choose $[0.7, 0.9]$. Note that all we care about are signs of $f(x)$, not their actual value, so this is all we display.

Iteration 1: $f(0.7) < 0$, $f(0.9) > 0$, and $f(0.8) > 0$. So replace 0.9 with 0.8 and repeat.

Iteration 2: $f(0.7) < 0$, $f(0.8) > 0$, and at the midpoint, 0.75, we have $f(0.75) > 0$. So replace 0.8 with 0.75 and repeat. Note that we do not need to continue to check the endpoints, just the midpoint. Thus we put the rest of the iterations in Table 8.14(b).

Notice that in the 12^{th} iteration we have the endpoints of the interval each starting with 0.739. Thus we have narrowed the zero down to an accuracy of the first three places after the decimal. Using a computer, we have

$$f(0.7390) = -0.00014, \quad f(0.7391) = 0.000024.$$

Either endpoint of the interval gives a good approximation of where f is 0. The intermediate value theorem states that the actual zero is still within this interval. While we do not know its exact value, we know it starts with 0.739.

This type of exercise is rarely done by hand. Rather, it is simple to program a computer to run such an algorithm and stop when the endpoints differ by a preset small amount.

It is a simple matter to extend the bisection method. For instance, we can find x, where $f(x) = 1$. It actually works very well to define a new function g where $g(x) = f(x) - 1$. Then use the bisection method to solve $g(x) = 0$. Similarly, given two functions f and g, we can use this method to solve $f(x) = g(x)$. Once again, create a new function h where $h(x) = f(x) - g(x)$ and solve $h(x) = 0$.

8.6 Limits involving infinity

In Definition 8.1 we stated that in the equation $\lim_{x\to c} f(x) = L$, both c and L were real numbers. In this section we relax that definition a bit by considering situations when it makes sense to let c and/or L be infinity. Essentially, we allow c and/or L to be in the set of extended real number $\overline{\mathbb{R}}$ (Section 2.2).

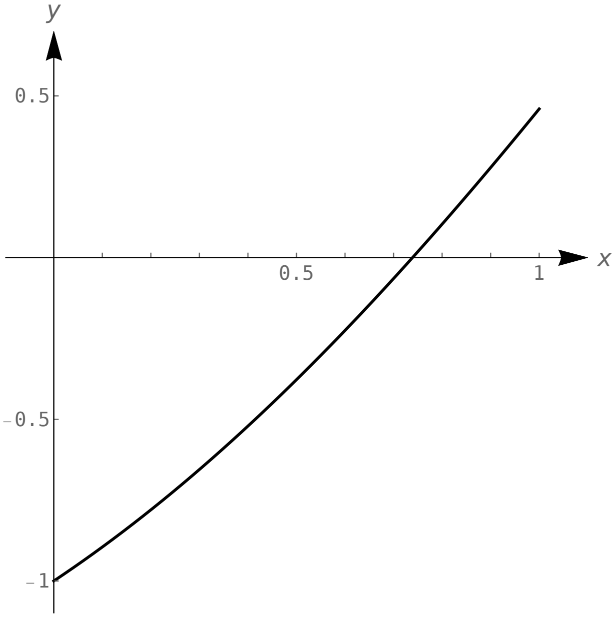
As a motivating example, consider $f(x) = \frac{1}{x^2}$, as shown in Figure 8.15. Note how, as x approaches 0, $f(x)$ grows very, very large – in fact, it grows without bound. It seems appropriate, and descriptive, to state that

$$\lim_{x\to 0} x^{-2} = +\infty.$$

Also note that as x gets very large, $f(x)$ gets very, very small. We could represent this concept with notation such as

$$\lim_{x\to+\infty} \frac{1}{x^2} = 0.$$

We explore both types of use of infinity in turn.



[h]

(a) Graph of $f(x) = x - \cos x$ near a root.

Iteration #	Interval	Midpoint Sign
1	$[0.7, 0.9]$	$f(0.8) > 0$
2	$[0.7, 0.8]$	$f(0.75) > 0$
3	$[0.7, 0.75]$	$f(0.725) < 0$
4	$[0.725, 0.75]$	$f(0.7375) < 0$
5	$[0.7375, 0.75]$	$f(0.7438) > 0$
6	$[0.7375, 0.7438]$	$f(0.7407) > 0$
7	$[0.7375, 0.7407]$	$f(0.7391) > 0$
8	$[0.7375, 0.7391]$	$f(0.7383) < 0$
9	$[0.7383, 0.7391]$	$f(0.7387) < 0$
10	$[0.7387, 0.7391]$	$f(0.7389) < 0$
11	$[0.7389, 0.7391]$	$f(0.7390) < 0$
12	$[0.7390, 0.7391]$	

(b)

Figure 8.14: Finding a root of $f(x) = x - \cos x$.

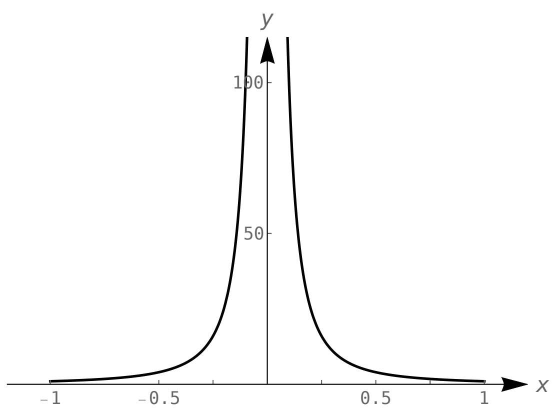


Figure 8.15: Graphing $f(x) = \frac{1}{x^2}$ for values of x near 0.

8.6.1 Limits of infinity and vertical asymptotes

Definitie 8.5 (Limit of infinity)
Let I be an open interval containing c, and let f be a function defined on I, except possibly at c.

- The limit of $f(x)$, as x approaches c, is positive infinity, denoted by

$$\lim_{x\to c} f(x) = +\infty,$$

  means that given any $N > 0$, there exists $\delta > 0$ such that for all x in I, where $x \neq c$, if $|x - c| < \delta$, then $f(x) > N$.

- The limit of $f(x)$, as x approaches c, is negative infinity, denoted by

$$\lim_{x\to c} f(x) = -\infty,$$

means that given any $N < 0$, there exists $\delta > 0$ such that for all x in I, where $x \neq c$, if $|x-c| < \delta$, then $f(x) < N$.

The first definition is similar to the (ε,δ)-definition (Definition 8.1). In that definition, given any (small) value ε, if we let x get close enough to c (within δ units of c) then $f(x)$ is guaranteed to be within ε of L. Here, given any (large) value N, if we let x get close enough to c (within δ units of c), then $f(x)$ will be at least as large as N. In other words, if we get close enough to c, then we can make $f(x)$ as large as we want.

We define one-sided limits that approach infinity in a similar way.

Definitie 8.6 (One-sided limits of infinity)

- Let f be a function defined on $]a,c[$ for some $a < c$.

  The limit of $f(x)$, as x approaches c from the left, is infinity, or, the left-hand limit of f at c is positive infinity, denoted by

$$\lim_{\substack{x\to c\\ <}} f(x) = +\infty,$$

  means given any $N > 0$, there exists $\delta > 0$ such that for all $a < x < c$, if $|x-c| < \delta$, then $f(x) > N$.

- Let f be a function defined on $]c,b[$ for some $b > c$.

  The limit of $f(x)$, as x approaches c from the right, is positive infinity, or, the right-hand limit of f at c is infinity , denoted by

$$\lim_{\substack{x\to c\\ >}} f(x) = +\infty,$$

  means given any $N > 0$, there exists $\delta > 0$ such that for all $c < x < b$, if $|x-c| < \delta$, then $f(x) > N$.

- The left- (or, right-) hand limit of f at c is negative infinity is defined as in Definition 8.5.

Example 8.14

Find

1. $\displaystyle \lim_{x\to 1} \frac{1}{(x-1)^2}$,
2. $\displaystyle \lim_{x\to 0} \frac{1}{x}$,
3. $\displaystyle \lim_{\substack{x\to -2\\ >}} \frac{x-1}{\sqrt{x+2}}$,
4. $\displaystyle \lim_{x\to 1} \frac{\sqrt{x^2+3}-2x}{x-1}$.

Solution

1. In Example 8.2, by inspecting values of x close to 1 we concluded that this limit does not exist. That is, it cannot equal any real number. But the limit could be infinite. And in fact, we see that the function does appear to be growing larger and larger, as $f(0.99) = 10^4$, $f(0.999) = 10^6$, $f(0.9999) = 10^8$. A similar thing happens on the other side of 1. In general,

let a large value N be given. Let $\delta = 1/\sqrt{N}$. If x is within δ of 1, i.e., if $|x-1| < 1/\sqrt{N}$, then:

$$\begin{aligned} |x-1| &< \frac{1}{\sqrt{N}} \\ \Leftrightarrow \quad (x-1)^2 &< \frac{1}{N} \\ \Leftrightarrow \quad \frac{1}{(x-1)^2} &> N, \end{aligned}$$

which is what we wanted to show. So we may say

$$\lim_{x\to 1} \frac{1}{(x-1)^2} = +\infty.$$

2. It is easy to see that the function grows without bound near 0, but it does so in different ways on different sides of 0. Since its behaviour is not consistent, we cannot say that

$$\lim_{x\to 0} \frac{1}{x} = \infty.$$

However, we can make a statement about one–sided limits. We can state that

$$\lim_{\substack{x\to 0 \\ >}} \frac{1}{x} = +\infty, \qquad \text{and,} \qquad \lim_{\substack{x\to 0 \\ <}} \frac{1}{x} = -\infty.$$

3. Here, we are confronted with the composition of an rational and irrational function. The point $x = -2$ does not belong to the corresponding function's domain, as it is a zero of the denominator only, so we expect to find positive or minus infinity. Indeed, we easily find

$$\lim_{\substack{x\to -2 \\ >}} \frac{x-1}{\sqrt{x+2}} = \frac{-3}{0} = -\infty.$$

4. In this case $x = 1$ is a zero of both the numerator and denominator, but we can simplify the expression as follows.

$$\begin{aligned} \lim_{x\to 1} \frac{\sqrt{x^2+3}-2x}{x-1} &= \lim_{x\to 1} \frac{x^2+3-4x^2}{(x-1)\left(\sqrt{x^2+3}+2x\right)} \\ &= \lim_{x\to 1} \frac{-3(x^2-1)}{(x-1)\left(\sqrt{x^2+3}+2x\right)} \\ &= \lim_{x\to 1} \frac{-3(x+1)(x-1)}{(x-1)\left(\sqrt{x^2+3}+2x\right)} = -\frac{3}{2} \end{aligned}$$

In Mathematica, these limits are computed as any other (one-sided) limit. For example, to compute

$$\lim_{\substack{x\to 0 \\ >}} \frac{1}{x},$$

we write the following.

```
In[9]:= Limit[1/x, x → 0, Direction → "FromAbove"]

Out[9]= ∞
```

If a function f has a limit (or, left- or right-hand limit) of infinity at $x = c$, then the graph of f looks similar to a vertical line near $x = c$. This observation leads to a definition.

Definitie 8.7 (Vertical asymptote)
Let I be an interval that either contains c or has c as an endpoint, and let f be a function defined on I, except possibly at c.

If the limit of $f(x)$ as x approaches c from either the left or right (or both) is $+\infty$ or $-\infty$, then the line $x = c$ is a **vertical asymptote** (*verticale asymptoot*) of f.

Example 8.15

Find the vertical asymptotes of

$$f(x) = \frac{3x}{x^2 - 4}.$$

Solution

Vertical asymptotes occur where the function grows without bound; this can occur at values of c where the denominator is 0. When x is near c, the denominator is small, which in turn can make the function take on large values. In the case of the given function, the denominator is 0 at $x = \pm 2$. Substituting in values of x close to 2 and -2 seems to indicate that the function tends toward ∞ or $-\infty$ at those points. We can graphically confirm this by looking at Figure 8.16. Thus the vertical asymptotes are at $x = \pm 2$.

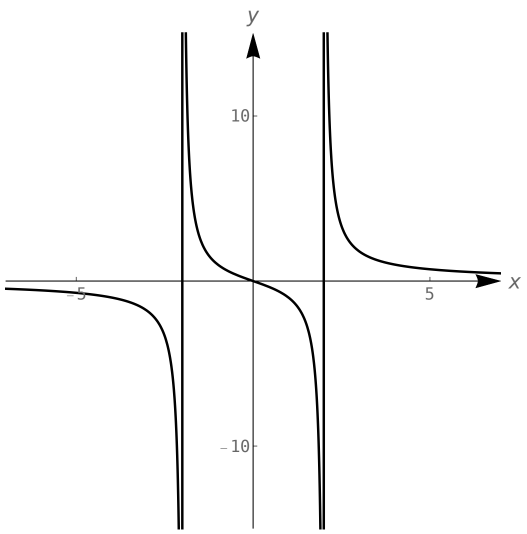


Figure 8.16: Graphing $f(x) = \frac{3x}{x^2 - 4}$.

If the denominator of a rational function is 0 at a certain point but the numerator is not, then there will usually be a vertical asymptote at that point. On the other hand, if the numerator and denominator are both zero at that point, then there may or may not be a vertical asymptote at that point. This case where the numerator and denominator are both zero returns us to an important topic.

8.6.2 Indeterminate forms

We have seen how the limits

$$\lim_{x \to 0} \frac{\sin(x)}{x} \quad \text{and} \quad \lim_{x \to 1} \frac{x^2 - 1}{x - 1}$$

each return the indeterminate form $0/0$ when we blindly plug in $x = 0$ and $x = 1$, respectively. However, $0/0$ is not a valid arithmetical expression. With a little cleverness, one can come up with $0/0$ expressions which have a limit of ∞, 0, or any other real number. That is why this expression is called **indeterminate** (*onbepaald*).

A key concept to understand is that such limits do not really return $0/0$. Rather, keep in mind that we are taking limits. What is really happening is that the numerator is shrinking to 0 while the denominator is also shrinking to 0. The respective rates at which they do this are very important and determine the actual value of the limit.

An indeterminate form indicates that one needs to do more work in order to compute the limit. That work may be algebraic (such as factoring and cancelling) or it may require a tool such as the squeeze theorem. In Chapter 9 we will learn a technique called l'Hôpital's Rule that provides another way to handle indeterminate forms.

Some other common indeterminate forms are $+\infty-\infty$, $\infty\cdot 0$, ∞/∞, 0^0, ∞^0 and 1^∞. Again, keep in mind that the expression $\infty-\infty$ does not really mean subtract infinity from infinity. Rather, it means one quantity is subtracted from the other, but both are growing without bound. What is the result? It is possible to get every value between $-\infty$ and $+\infty$.

8.6.3 Limits at infinity and horizontal asymptotes

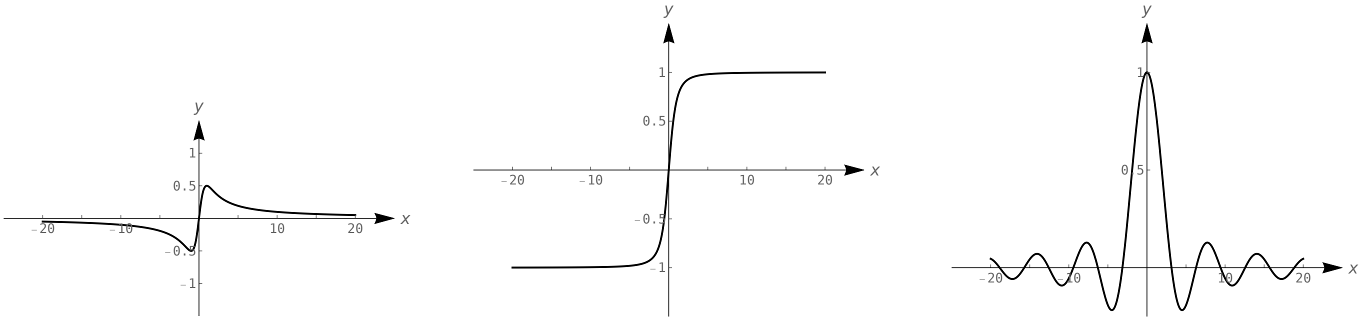
In Figure 8.15 we briefly considered what happens to $f(x) = x^{-2}$ as x grew very large. Graphically, it concerns the behaviour of the function to the far right of the graph. We make this notion more explicit in the following definition.

Definitie 8.8 (Limits at infinity and horizontal asymptotes)
Let L be a real number.

1. Let f be a function defined on $]a, +\infty[$ for some number a. The limit of f at infinity is L, or $\lim\limits_{x\to+\infty} f(x) = L$, means for every $\varepsilon > 0$ there exists $M > a$ such that if $x > M$, then $|f(x)-L| < \varepsilon$.
2. Let f be a function defined on $]-\infty, b[$ for some number b. The limit of f at negative infinity is L, or $\lim\limits_{x\to-\infty} f(x) = L$, means for every $\varepsilon > 0$ there exists $M < b$ such that if $x < M$, then $|f(x)-L| < \varepsilon$.
3. If $\lim\limits_{x\to+\infty} f(x) = L$ or $\lim\limits_{x\to-\infty} f(x) = L$, we say the line $y = L$ is a **horizontal asymptote** (*horizontale asymptoot*) of f.

Horizontal asymptotes can take on a variety of forms. Figure 8.17(a) shows that $f(x) = x/(x^2+1)$ has a horizontal asymptote of $y = 0$, where 0 is approached from both above and below. On the other hand, Figure 8.17(b) shows that $f(x) = x/\sqrt{x^2+1}$ has two horizontal asymptotes; one at $y = 1$ and the other at $y = -1$. Figure 8.17(c) shows that $f(x) = (\sin(x))/x$ has even more interesting behaviour than at just $x = 0$; as x approaches $\pm\infty$, $f(x)$ approaches 0, but oscillates as it does this.

We can analytically evaluate limits at infinity for rational functions once we understand $\lim\limits_{x\to+\infty} 1/x$. As x gets larger and larger, $1/x$ gets smaller and smaller, approaching 0. We can, in fact, make $1/x$ as small as we want by choosing a large enough value of x. Given ε, we can make $1/x < \varepsilon$ by choosing



(a) (b) (c)

Figure 8.17: A graph of $f(x) = x/(x^2+1)$ (a), $f(x) = x/\sqrt{x^2+1}$ (b) and $f(x) = (\sin(x))/x$.

$x > 1/\varepsilon$. Thus we have $\lim_{x\to+\infty} 1/x = 0$. It is now not much of a jump to conclude the following:

$$\lim_{x\to+\infty} \frac{1}{x^n} = 0 \quad \text{and} \quad \lim_{x\to-\infty} \frac{1}{x^n} = 0.$$

Now suppose we need to compute the following limit:

$$\lim_{x\to+\infty} \frac{x^3+2x+1}{4x^3-2x^2+9}.$$

A good way of approaching this is to divide through the numerator and denominator by x^3, which is the largest power of x to appear in the function. Doing this, we get

$$\lim_{x\to+\infty} \frac{x^3+2x+1}{4x^3-2x^2+9} = \lim_{x\to+\infty} \frac{x^3\left(1+2/x^2+1/x^3\right)}{x^3\left(4-2/x+9/x^3\right)}.$$

Then using the rules for limits (which also hold for limits at infinity), as well as the fact about limits of x^{-n}, we see that the limit becomes

$$\frac{1+0+0}{4-0+0} = \frac{1}{4}.$$

This procedure works for any rational function. In fact, it gives us the following theorem.

Theorem 8.8 (Limits of rational functions at infinity)
Let $f(x)$ be a rational function of the following form:

$$f(x) = \frac{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0}{b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0},$$

where any of the coefficients may be 0 except for a_n and b_m.

1. *If $n = m$, then $\lim_{x\to+\infty} f(x) = \lim_{x\to-\infty} f(x) = \frac{a_n}{b_m}$.*
2. *If $n < m$, then $\lim_{x\to+\infty} f(x) = \lim_{x\to-\infty} f(x) = 0$.*
3. *If $n > m$, then $\lim_{x\to+\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$ are both infinite.*

Intuitively, as x gets very large, all the terms in the numerator are small in comparison to $a_n x^n$, and likewise all the terms in the denominator are small compared to $b_n x^m$. If $n = m$, looking only at these two important terms, we have $(a_n x^n)/(b_n x^m)$. This reduces to a_n/b_m. If $n < m$, the function behaves like $a_n/(b_m x^{m-n})$, which tends toward 0. If $n > m$, the function behaves like $a_n x^{n-m}/b_m$, which will tend to either $+\infty$ or $-\infty$ depending on the values of n, m, a_n, b_m and whether you are looking for $\lim\limits_{x\to+\infty} f(x)$ or $\lim\limits_{x\to-\infty} f(x)$.

With care, we can quickly evaluate limits at infinity for a large number of functions by considering the largest powers of x. This is, for instance, the case for irrational functions. As an example, consider again

$$\lim_{x\to\pm\infty} \frac{x}{\sqrt{x^2+1}},$$

graphed in Figure 8.17(b). When x is very large, $x^2+1 \approx x^2$. Thus

$$\sqrt{x^2+1} \approx \sqrt{x^2} = |x|, \quad \text{and} \quad \frac{x}{\sqrt{x^2+1}} \approx \frac{x}{|x|}.$$

This expression is 1 when x is positive and -1 when x is negative. Hence we get asymptotes of $y=1$ and $y=-1$, respectively. In general, when evaluation limits at infinity involving irrational functions we should bear in mind that

$$\begin{aligned} \sqrt{x^2} &= x, && \forall x \in \mathbb{R}^+, \\ \sqrt[3]{x^3} &= x, && \forall x \in \mathbb{R}, \\ \sqrt{x^2} &= -x, && \forall x \in \mathbb{R}^-, \\ \sqrt[3]{(-x)^3} &= -x, && \forall x \in \mathbb{R}. \end{aligned}$$

Example 8.16

Evaluate each of the following limits.

1. $\lim\limits_{x\to-\infty} \dfrac{x^2+2x-1}{x^3+1}$
2. $\lim\limits_{x\to+\infty} \dfrac{x^2-1}{3-x}$
3. $\lim\limits_{x\to-\infty} \dfrac{\sqrt{x^2+5}+7x}{2x-3}$
4. $\lim\limits_{x\to-\infty} \left(\sqrt{x^2+8x}+x\right)$

Solution

1. The highest power of x is in the denominator. Therefore, the limit is 0.
2. We see that

$$\begin{aligned} \lim_{x\to+\infty} \frac{x^2-1}{3-x} &= \lim_{x\to+\infty} \frac{x^2\left(1-1/x^2\right)}{x\,(3/x-1)} \\ &= \lim_{x\to+\infty} \frac{x\left(1-1/x^2\right)}{3/x-1} = -\infty\,. \end{aligned}$$

3. We first should realize that the highest power of x in both the numerator and denominator is the same. Hence, we proceed as follows:

$$\lim_{x\to-\infty} \frac{\sqrt{x^2+5}+7x}{2x-3} = \lim_{x\to-\infty} \frac{|x|\sqrt{1+\frac{5}{x^2}}+7x}{x\left(2-\frac{3}{x}\right)}$$
$$= \lim_{x\to-\infty} \frac{x\left(-\sqrt{1+\frac{5}{x^2}}+7\right)}{x\left(2-\frac{3}{x}\right)}$$
$$= \frac{-1+7}{2} = 3.$$

4. At first sight, we would say that this limit leads to the indeterminate form $\infty-\infty$, but this can be overcome by multiplying both numerator and denominator by the conjugate expression.

$$\lim_{x\to-\infty}\left(\sqrt{x^2+8x}+x\right) = \lim_{x\to-\infty} \frac{x^2+8x-x^2}{\sqrt{x^2+8x}-x}$$
$$= \lim_{x\to-\infty} \frac{8x}{-x\left(\sqrt{1+\frac{8}{x}}+1\right)} = \frac{8}{-2} = -4$$

In Mathematica, these limits are again computed as any other limit. To specify that the limit is at (minus) infinity, we simply write `(-)Infinity`. For example,

$$\lim_{x\to-\infty} \frac{x^2+2x-1}{x^3+1},$$

is computed as follows.

```
In[10]:= Limit[(x^2 + 2 x - 1) (x^3 + 1), x → -Infinity]

Out[10]= 0
```

For the sake of comprehensiveness, we list below the steps that should be taken when evaluating the limit $\lim_{x\to\pm\infty} f(x)$:

1. Compute $f(\pm\infty)$.
2. You arrive at one of the following cases:
   - $f(\pm\infty) = \pm\infty$: the function values approaches ∞ as $x \to \pm\infty$
   - $f(\pm\infty) = b \in \mathbb{R}$: $y = b$ is a horizontal asymptote of the function f.
   - $f(\pm\infty) = \left(\frac{\infty}{\infty}\right)$: factor out the highest-degree term in both the nominator and denominator, then simplify and return to Step 1.
   - $f(\pm\infty) = (\infty-\infty)$: multiply with the conjugate binomial, then factor out the highest-degree term and return to Step 1.

8.6.4 Slant asymptotes

In addition to vertical and horizontal asymptotes, we can also define **slant or oblique asymptotes** (*schuine asymptoot*). These are diagonal lines such that the difference between the curve and the line approaches 0 as x tends to $+\infty$ or $-\infty$.

Definitie 8.9 (Slant asymptotes)
The line $y = ax + b$ is a **slant asymptote** for the function f if and only if

$$\lim_{x\to+\infty}\left[f(x)-(ax+b)\right]=0,$$

or/and

$$\lim_{x\to-\infty}\left[f(x)-(ax+b)\right]=0\,.$$

From this definition, it follows that

$$\lim_{x\to\pm\infty}\left[\frac{f(x)}{x}-\left(a+\frac{b}{x}\right)\right]=0,$$

so

$$a=\lim_{x\to\pm\infty}\frac{f(x)}{x}\qquad\text{and}\qquad b=\lim_{x\to\pm\infty}(f(x)-ax)\,.$$

Example 8.17

Determine the asymptotes, if any, of the following functions.

1. $f(x)=\dfrac{x^3+2}{x^2-9}$
2. $f(x)=\sqrt{x^2-4x+3}$

Solution

1. (a) Vertical asymptotes

   Since $f(x)$ tends to infinity as x approaches 3 or -3, the vertical asymptotes are $x=-3$ and $x=3$. More precisely, we find that

   $$\lim_{\substack{x\to-3\\<}}f(x)=-\infty\qquad\lim_{\substack{x\to3\\<}}f(x)=-\infty$$
   $$\lim_{\substack{x\to-3\\>}}f(x)=+\infty\qquad\lim_{\substack{x\to3\\>}}f(x)=+\infty$$

   This allows us to conclude that $f(x)$ tends towards $-\infty$ as x is approaching -3 from the left, whereas $f(x)$ tends towards ∞ when approaching -3 from the right, and likewise for what concerns the vertical asymptote at $x=3$.

   (b) Horizontal asymptotes

   There are none because

   $$\lim_{x\to\pm\infty}\frac{x^3+2}{x^2-9}=\pm\infty.$$

(c) Slant asymptotes

We verify that

$$a = \lim_{x\to\pm\infty} \frac{f(x)}{x} = \lim_{x\to\pm\infty} \frac{x^3+2}{x^3-9x} = 1,$$

$$b = \lim_{x\to\pm\infty} \big(f(x) - ax\big) = \lim_{x\to\pm\infty} \left(\frac{x^3+2}{x^2-9} - x\right) = \lim_{x\to\pm\infty} \left(\frac{2+9x}{x^2-9}\right) = 0.$$

Consequently, the function has slant asymptote $y = x$ for $x \to \pm\infty$. The position of this asymptote with respect to the graph of the function f can be found be determining the sign of

$$g(x) = f(x) - (ax+b) = \frac{x^3+2}{x^2-9} - x = \frac{9x+2}{x^2-9}.$$

The sign diagram of the function g is:

x		-3		$-\frac{2}{9}$		3	
$g(x)$	$-$	|	$+$	0	$-$	|	$+$

Consequently, we may conclude that the graph of f lies above the slant asymptote for $x \to +\infty$ because then $g(x) > 0$, whereas the graph of f lies below the slant asymptote for $x \to -\infty$ because then $g(x) < 0$. This is confirmed by the graph of the function f shown in Figure 8.18(a).

2. $f(x) = \sqrt{x^2-4x+3}$

(a) Vertical asymptotes

The function $f(x)$ only tends to infinity if x does so, so there are no vertical asymptotes.

(b) Horizontal asymptotes

There are none because

$$\lim_{x\to\pm\infty} \sqrt{x^2-4x+3} = +\infty.$$

(c) Slant asymptotes

We verify that

- $x \to +\infty$

$$a = \lim_{x\to+\infty} \frac{f(x)}{x} = \lim_{x\to+\infty} \frac{\sqrt{x^2-4x+3}}{x} = 1,$$

$$\begin{aligned} b &= \lim_{x\to+\infty} \big(f(x) - ax\big) = \lim_{x\to+\infty} \left(\sqrt{x^2-4x+3} - x\right) \\ &= \lim_{x\to+\infty} \left(\frac{-4x+3}{\sqrt{x^2-4x+3}+x}\right) = -2. \end{aligned}$$

- $x \to -\infty$

$$a = \lim_{x\to-\infty} \frac{f(x)}{x} = \lim_{x\to-\infty} \frac{\sqrt{x^2-4x+3}}{x} = -1,$$

$$
\begin{aligned}
b &= \lim_{x\to-\infty}\big(f(x)-ax\big) = \lim_{x\to-\infty}\left(\sqrt{x^2-4x+3}+x\right) \\
&= \lim_{x\to-\infty}\left(\frac{-4x+3}{\sqrt{x^2-4x+3}-x}\right) = 2.
\end{aligned}
$$

Hence, $y = x-2$ is a slant asymptote of f for $x \to +\infty$, while $y = -x+2$ is a slant asymptote of f for $x \to -\infty$. Again, the position of the slant asymptote with respect to the graph of f can be determined by verifying the sign of

$$
g(x) = \sqrt{x^2-4x+3}-(x-2) \qquad \text{and} \qquad h(x) = \sqrt{x^2-4x+3}-(-x+2).
$$

It follows that as $x \to +\infty$, then we have that $g(x) < 0$, so the graph of f lies below the slant asymptote $y = x-2$. Similarly, as $x \to -\infty$, then we have that $h(x) < 0$, so the graph of f lies below the slant asymptote $y = -x+2$. This is confirmed by the graph of the function f shown in Figure 8.18(b).

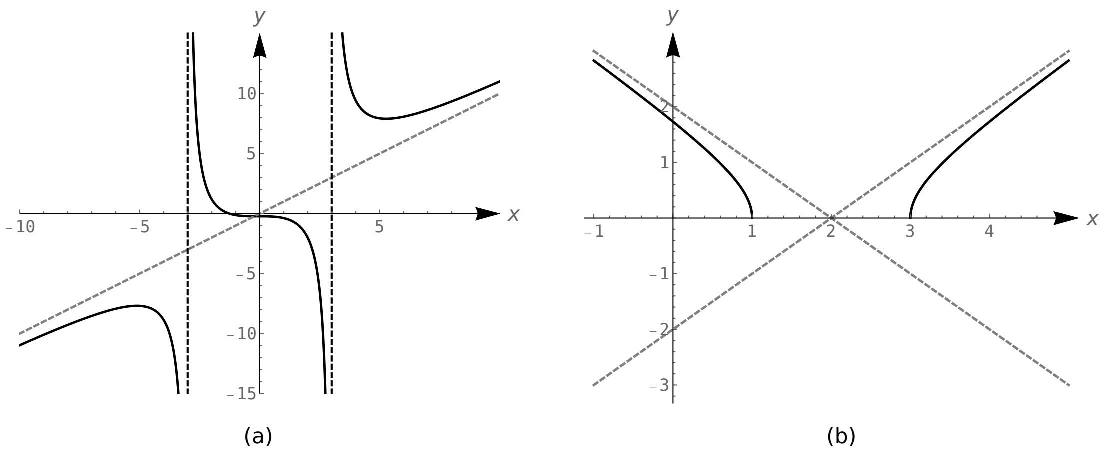


Figure 8.18: A graph of $f(x) = \frac{x^3+2}{x^2-9}$ (a) and $f(x) = \sqrt{x^2-4x+3}$ (b).

8.7 Exercises

Assignment 8.1 — Calculate the following limits.

(a) $\lim\limits_{x \to -1} \left(4x^2 - 5x + 3\right)$

(b) $\lim\limits_{\substack{x \to 0 \\ >}} x^2 \ln(x)$

(c) $\lim\limits_{x \to 1} \dfrac{x^2 - 1}{x^3 - 1}$

(d) $\lim\limits_{x \to 2} \dfrac{\sqrt{x+2} - 2}{x - 2}$

(e) $\lim\limits_{\substack{x \to 1 \\ >}} \dfrac{\sqrt{(x-1)^2}}{x - 1}$

(f) $\lim\limits_{\substack{x \to -4 \\ >}} \dfrac{2x + 8}{|x + 4|}$

(g) $\lim\limits_{x \to 1} \dfrac{(x-1)^3}{x^2 - 4x + 3}$

(h) $\lim\limits_{x \to -1} \dfrac{x^3 + 2x^2 + x}{x^8 - 2x^4 + 1}$

(i) $\lim\limits_{x \to 1} \dfrac{x - \sqrt{x}}{2 - \sqrt{x+3}}$

(j) $\lim\limits_{x \to 5} \dfrac{\sqrt{x+4} + x - 8}{x^2 - 8x + 15}$

(k) $\lim\limits_{x \to 1} \dfrac{\sqrt{8x+1} + \sqrt{2x-1} - 4}{x - 1}$

(l) $\lim\limits_{x \to 0} \dfrac{|3x - 1| - |3x + 1|}{x}$

(m) $\lim\limits_{x \to 3} \dfrac{x^2 - 6x + 9}{x^2 - 9}$

(n) $\lim\limits_{h \to 0} \dfrac{\sqrt{4+h} - 2}{h}$

(o) $\lim\limits_{x \to 1} \dfrac{x^2 - 1}{\sqrt{x+3} - 2}$

(p) $\lim\limits_{x \to 2} \dfrac{x^4 - 16}{x^3 - 8}$

Assignment 8.2 — Determine $\lim\limits_{x \to 0} \left(x^2 \sin\left(\dfrac{1}{x}\right)\right)$ by using the squeeze theorem.

One-sided limits

Assignment 8.3 — Consider the function $y = f(x)$ as given in Figure 8.19 and calculate the limits.

(a) $\lim\limits_{\substack{x \to 0 \\ >}} f(x)$

(b) $\lim\limits_{x \to 1} f(x)$

(c) $\lim\limits_{\substack{x \to 2 \\ >}} f(x)$

(d) $\lim\limits_{\substack{x \to 2 \\ <}} f(x)$

(e) $\lim\limits_{\substack{x \to 3 \\ <}} f(x)$

(f) $\lim\limits_{\substack{x \to 3 \\ >}} f(x)$

(g) $\lim\limits_{\substack{x \to 4 \\ >}} f(x)$

(h) $\lim\limits_{\substack{x \to 4 \\ <}} f(x)$

(i) $\lim\limits_{\substack{x \to 5 \\ <}} f(x)$

(j) $\lim\limits_{\substack{x \to 5 \\ >}} f(x)$

(k) $\lim\limits_{x \to +\infty} f(x)$

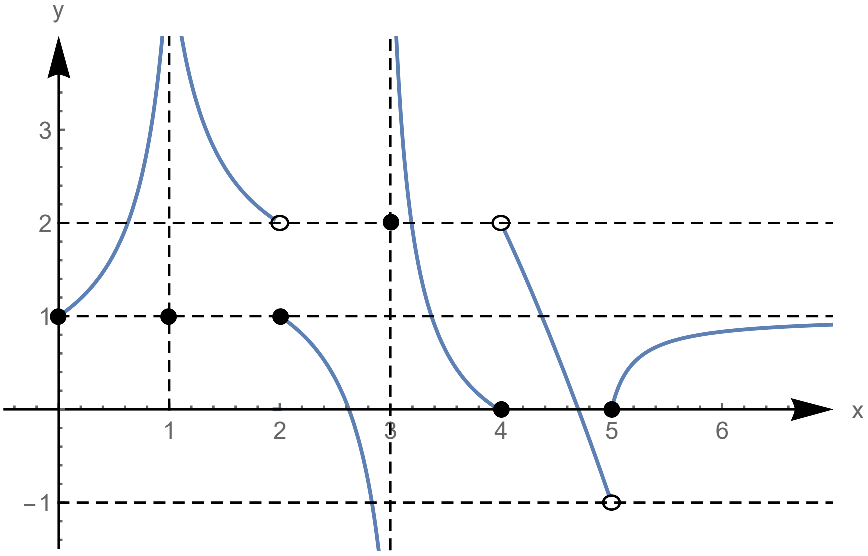


Figure 8.19: The function $y = f(x)$ from Exercise 8.3 .

Assignment 8.4 — Consider the function

$$f(x) = \begin{cases} x-1, & \text{if } x \le -1, \\ x^2+1, & \text{if } -1 < x \le 0, \\ (x+\pi)^2, & \text{if } x > 0. \end{cases}$$

Compute the following limits

(a) $\lim\limits_{\substack{x \to -1 \\ <}} f(x)$

(b) $\lim\limits_{\substack{x \to -1 \\ >}} f(x)$

(c) $\lim\limits_{\substack{x \to 0 \\ >}} f(x)$

(d) $\lim\limits_{\substack{x \to 0 \\ <}} f(x)$

Continuity

Assignment 8.5 — Investigate whether the given functions in the given interval I contains at least one zero. If so, determine the zero(s) using the bisection method.

(a) $f(x) = x^5 - 5x + 1, \quad I = [0, 1]$

(b) $f(x) = x^3 + 8x^2 - 3, \quad I = [-10, -1]$

(c) $f(x) = \dfrac{|x|}{x}, \quad I = [-1, 1]$

(d) $f(x) = 1 - x + \sin(x), \quad I = [0, \pi]$

(e) $f(x) = x^3 - 15x + 1, \quad I = [-4, 4]$

Limits involving infinity

Assignment 8.6 — Determine the following limits:

(a) $\displaystyle\lim_{x\to-\infty}\left(x^5-x^4+x^3\right)$

(b) $\displaystyle\lim_{x\to\pm\infty}\frac{x^2-1}{x+1}$

(c) $\displaystyle\lim_{x\to\pm\infty}\frac{\sqrt{x^2+5}+2x}{3x}$

(d) $\displaystyle\lim_{x\to\pm\infty}\frac{6x^3+4x^2-x-1}{x^4+5}$

(e) $\displaystyle\lim_{x\to\pm\infty}\frac{\sqrt{x^2-1}+2x}{x+7}$

(f) $\displaystyle\lim_{x\to+\infty}\frac{2x-3}{\sqrt{x^2+4}-\sqrt{2x}}$

(g) $\displaystyle\lim_{x\to\pm\infty}\frac{\sqrt{x^4+x^3+8}}{1-x^2}$

(h) $\displaystyle\lim_{x\to\pm\infty}\left(\sqrt{4x^2+3}-2x\right)$

(i) $\displaystyle\lim_{x\to\pm\infty}\left(2x-1-\sqrt{4x^2+x}\right)$

(j) $\displaystyle\lim_{x\to+\infty}\frac{3x^3-5x^2+7}{8+2x-5x^3}$

(k) $\displaystyle\lim_{x\to\pm\infty}\frac{2x-1}{\sqrt{3x^2+x+1}}$

(l) $\displaystyle\lim_{x\to-\infty}\frac{2x-5}{|3x+2|}$

(m) $\displaystyle\lim_{x\to\pm\infty}\left(x+\sqrt{x^2-4x+1}\right)$

Assignment 8.7 — Determine the asymptotes of the graph of the functions below. Also, determine how these graphs approach the asymptotes.

(a) $f(x) = \dfrac{1-x}{2x-1}$

(b) $f(x) = \dfrac{x^2-x-2}{x-3}$

(c) $f(x) = \dfrac{3x^2-5}{x^2+7}$

(d) $f(x) = \dfrac{2x^3+x}{x^2+1}$

(e) $f(x) = x - \dfrac{7}{x}$

(f) $f(x) = \dfrac{x^4-1}{3x^3-12x}$

(g) $f(x) = \sqrt{x^2-4} - 3$

(h) $f(x) = \sqrt{x^2+2x-3}$

(i) $f(x) = x - \sqrt{x^2-4}$

(j) $f(x) = \dfrac{x^2+x-6}{\sqrt{x^2-x-6}}$

Review Exercises

Assignment 8.8 — Calculate the following limits.

(a) $\lim\limits_{x\to+\infty} \dfrac{\sin(x)}{x}$

(b) $\lim\limits_{x\to-\infty} (x+\cos(x))$

(c) $\lim\limits_{x\to+\infty} \dfrac{1+\cos(x)}{x^2}$

(d) $\lim\limits_{x\to 0} \dfrac{\sin(3x)}{x}$

(e) $\lim\limits_{x\to 0} \dfrac{\tan(x)}{x}$

(f) $\lim\limits_{x\to 0} \dfrac{\sin(2x)}{\tan(5x)}$

(g) $\lim\limits_{x\to 0} \dfrac{\sin^2(x)}{x.\tan(2x)}$

(h) $\lim\limits_{x\to 0} \dfrac{\sin^2(7x)}{\tan^2(2x)}$

(i) $\lim\limits_{x\to 0} \dfrac{1-\cos(4x)}{x^2}$

(j) $\lim\limits_{x\to 0} \dfrac{1-\sqrt{1-\sin(x)}}{x}$

Assignment 8.9 — Calculate the following limits.

(a) $\lim\limits_{x\to+\infty} \left(\dfrac{2x+1}{2x}\right)^x$

(b) $\lim\limits_{x\to 0} (1+3x)^{1/x}$

(c) $\lim\limits_{x\to 0} \dfrac{e^{-x}-1}{x}$

(d) $\lim\limits_{x\to 0} \dfrac{e^{ax}-1}{x}$

Assignment 8.10 — Investigate the continuity of the following functions within their domain.

(a) $f(x) = \sin(4x)$

(b) $f(x) = \dfrac{x+2}{x^2-9}$

(c) $f(x) = \sqrt{x^2+5}$

(d) $f(x) = \sqrt{x^2-1}$

(e) $f(x) = \begin{cases} x^2, & \text{als} \quad x \in \mathbb{R}^+, \\ -x, & \text{als} \quad x \in \mathbb{R}^-, \end{cases}$

(f) $f(x) = \dfrac{|x|}{x}$

Assignment 8.11 — Consider fiure 8.19 again. In which point is $f(x)$ discontinuous? In which point is $f(x)$ left/right continuous

Assignment 8.12 — Determine the value of a such that the functions below are continuous for all x.

(a) $f(x) = \begin{cases} x-a, & \text{als} \quad x < 3, \\ 1-ax, & \text{als} \quad x \geq 3, \end{cases}$

(b) $f(x) = \begin{cases} x^3, & \text{als } x \leq 2, \\ ax^2, & \text{als } x > 2. \end{cases}$

Assignment 8.13 — Determine whether the graphs of the rational functions below have one or more vertical asymptotes and/or perforations.

(a) $f(x) = \dfrac{x^2+5x+6}{x+3}$

(b) $f(x) = \dfrac{x^2+3x-4}{x^2+x-6}$

(c) $f(x) = \dfrac{x^2-9}{x^2-2x-3}$

Q: What is the first derivative of a cow?
A: Prime Rib!

— Queen Elizabeth II —

9

Derivatives and their applications

The previous chapter introduced the most fundamental of calculus topics: the limit. This chapter introduces the second most fundamental of calculus topics: the derivative. Limits describe where a function is going; derivatives describe how fast the function is going.

9.1 Definition

9.1.1 Intuitive introduction

A common amusement park ride lifts riders to a height then allows them to freefall a certain distance before safely stopping them. Suppose such a ride drops riders from a height of 150 metres. From physics we know that the height (in metres) of the riders, t seconds after freefall (and ignoring air resistance, etc.) can be accurately modelled by $f(t) = -16t^2 + 150$. It allows us to verify that, without intervention, the riders will hit the ground at $t = 2.5\sqrt{1.5} \approx 3.06$ seconds, but how fast will the riders be travelling after two seconds?

We have been given a position function, but what we want to compute is a velocity at a specific point in time, i.e., we want an instantaneous velocity. We do not currently know how to calculate this. However, we do know how to calculate an average velocity using the difference quotient introduced in Section 8.1. More specifically, we have

$$\frac{\text{change in distance}}{\text{change in time}} = \frac{\text{rise}}{\text{run}} = \text{average velocity.}$$

We can approximate the instantaneous velocity at $t = 2$ by considering the average velocity over some time period containing $t = 2$. If we make the time interval small, we will get a good approximation. For instance, consider the interval from $t=2$ to $t=3$. On that interval, the average velocity is

$$\frac{f(3)-f(2)}{3-2}=\frac{f(3)-f(2)}{1}=-80 \text{ m/s},$$

where the minus sign indicates that the riders are moving down. By narrowing the considered interval, we get a better approximation of the instantaneous velocity. On $[2,2.5]$ we have

$$\frac{f(2.5)-f(2)}{2.5-2}=\frac{f(2.5)-f(2)}{0.5}=-72 \text{ m/s}.$$

We can do this for smaller and smaller intervals of time. For instance, over a time span of 1/10$^{\text{th}}$ of a second, i.e., on $[2,2.1]$, we have

$$\frac{f(2.1)-f(2)}{2.1-2}=\frac{f(2.1)-f(2)}{0.1}=-65.6 \text{ m/s}.$$

Likewise, over a time span of 1/100$^{\text{th}}$ of a second, on $[2,2.01]$, the average velocity is

$$\frac{f(2.01)-f(2)}{2.01-2}=\frac{f(2.01)-f(2)}{0.01}=-64.16 \text{ m/s}.$$

Essentially, we are computing the average velocity on the interval $[2,2+h]$ for small values of h. That is, we are computing

$$\frac{f(2+h)-f(2)}{h},$$

where h is small. Still, we really want to use $h=0$, but this, of course, returns the indeterminate form $0/0$. Computing this limit directly gives

$$\begin{aligned}
\lim_{h\to 0}\frac{f(2+h)-f(2)}{h} &= \lim_{h\to 0}\frac{-16(2+h)^2+150-(-16(2)^2+150)}{h}\\
&= \lim_{h\to 0}\frac{-64h-16h^2}{h}\\
&= \lim_{h\to 0}(-64-16h)\\
&= -64.
\end{aligned}$$

Graphically, we can view the average velocities we computed numerically as the slopes of secant lines on the graph of f going through the points $(2,f(2))$ and $(2+h,f(2+h))$. In Figure 9.1(a), the secant line corresponding to $h=1$ is shown. Notice how well it approximates f between those two points – it is a common practice to approximate functions with straight lines. As $h\to 0$, these secant lines approach the **tangent line** (*raaklijn*), a line that goes through the point $(2,f(2))$ with the special slope of -64 (Figure 9.1(b)). It is clear that this tangent line approximates the function f even better than the secant line.

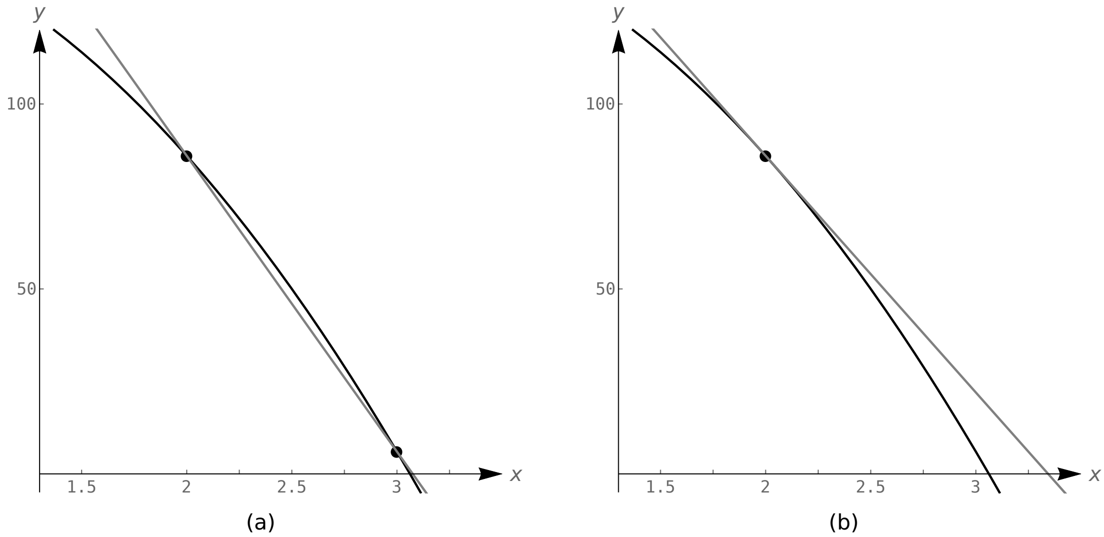


Figure 9.1: The secant line to $f(x)$ with $h=1$ (a) and the tangent line to f at $x=2$.

9.1.2 Formalism

Having introduced the derivative in an intuitive way, let us now turn to its formal definition.

Definitie 9.1 (Derivative at a point)
Let f be a continuous function on an open interval I and let c be in I. The **derivative** (*afgeleide*) of f at c, denoted $f'(c)$, is

$$\lim_{h\to 0}\frac{f(c+h)-f(c)}{h},$$

provided the limit exists. If the limit exists, we say that f is differentiable at c; if the limit does not exist, then f is not differentiable at c. If f is **differentiable** (*afleidbaar*) at every point in I, then f is differentiable on I. Furthermore, we call f continuously differentiable over I if f' is continuous over I.

Using this definition, we can also formally define a tangent line to the graph of a function f.

Definitie 9.2 (Tangent line)
Let f be continuous on an open interval I and differentiable at c, for some c in I. The line with equation $y=\ell(x)$

$$y=f'(c)(x-c)+f(c),$$

is the **tangent line** (*raaklijn*) to the graph of f at c; that is, it is the line through $(c,f(c))$ whose slope is the derivative of f at c.

When $f'(c)=0$, the tangent line is the horizontal line through $\big(c,f(c)\big)$; that is, $y=f(c)$. Moreover, the larger $f'(c)$ the more the tangent lines becomes oriented vertically.

Clearly, from the derivative we can also construct the normal line. It is perpendicular to the tangent line, hence its slope is the opposite–reciprocal of the tangent line's slope.

Definitie 9.3 (Normal line)
Let f be continuous on an open interval I and differentiable at c, for some c in I. The **normal line**

(*normaal*) to the graph of f at c is the line with equation

$$n(x) = \frac{-1}{f'(c)}(x-c)+f(c),$$

where $f'(c) \neq 0$.

When $f'(c) = 0$, the normal line is the vertical line through $\big(c, f(c)\big)$; that is, $x = c$.

Some examples will help us understand these definitions.

Example 9.1

Let $f(x) = 3x^2 + 5x - 7$. Find:

1. $f'(1)$.
2. The equation of the tangent line to the graph of f at $x = 1$.
3. The equation of the normal line to the graph of f at $x = 1$.

Solution

1. We compute this directly using Definition 9.1.

$$\begin{aligned} f'(1) &= \lim_{h\to 0} \frac{f(1+h)-f(1)}{h} \\ &= \lim_{h\to 0} \frac{3(1+h)^2+5(1+h)-7-(3(1)^2+5(1)-7)}{h} \\ &= \lim_{h\to 0} \frac{3h^2+11h}{h} \\ &= \lim_{h\to 0}(3h+11) = 11 \end{aligned}$$

2. The tangent line at $x = 1$ has slope $f'(1)$ and goes through the point $(1, f(1)) = (1, 1)$. Thus the tangent line has equation, in point-slope form, $y = 11(x-1)+1$. In slope-intercept form, we have $y = 11x - 10$.
3. Since $f'(1) = 11$. Hence at $x = 1$, the normal line will have slope $-1/11$. An equation for the normal line is

$$n(x) = \frac{-1}{11}(x-1)+1.$$

A graph of f is given in Figure 9.2 along with its tangent and normal lines at $x = 1$. Note that in this figure these lines do not seem to perpendicular to one another, but this is a mere consequence of the chose aspect ration of the plot window.

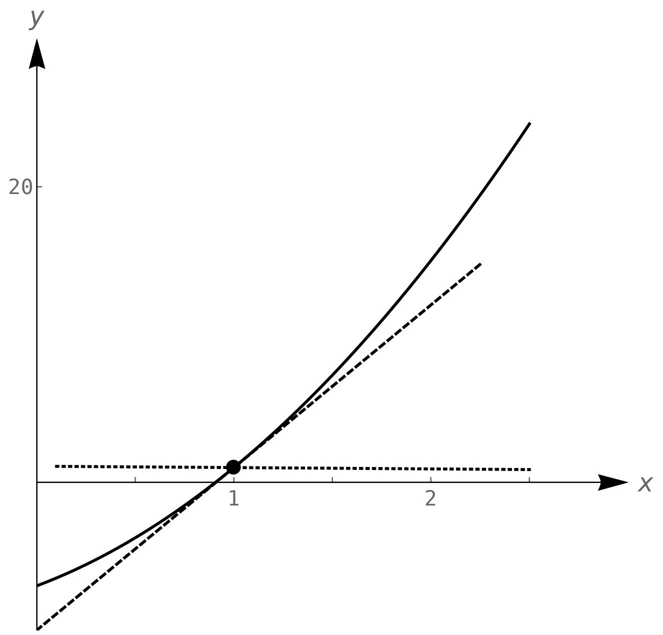


Figure 9.2: A graph of $f(x) = 3x^2 + 5x - 7$ and its tangent (dashed) and normal (dotted) lines at $x = 1$.

Linear functions are easy to work with; many functions that arise in the course of solving real problems, however, are not easy to work with. A common practice in mathematical problem solving is to approximate difficult functions with not–so–difficult functions. Lines are a common choice. It turns out that at any given point on the graph of a differentiable function f, the best linear approximation to f is its tangent line. That is one reason we will spend considerable time finding tangent lines to functions.

From Example 9.1, it is clear that we would have to evaluate a limit for every point c at which we want to find the derivative of f. Yet, instead of doing this repeatedly for different values of c, let us do it just once for the variable x. We then take a limit just once. The process now looks like:

input variable x ⟶ $\boxed{\text{do something to } f \text{ and } x}$ ⟶ return function $f'(x)$

The output is the derivative function, $f'(x)$. The $f'(x)$ function will take a number c as input and return the derivative of f at c. This gives rise to the following definition.

Definitie 9.4 (Derivative function)
Let f be a differentiable function on an open interval I. The function

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

is the derivative of f.

Note that the following notations all represent the derivative of a function f, if defined as $y = f(x)$:

$$f'(x) = y' = \frac{dy}{dx} = \frac{df}{dx} = \frac{d}{dx}(f) = \frac{d}{dx}(y).$$

Example 9.2

Find the derivative of the following functions:

1. $f(x) = 3x^2 + 5x - 7$
2. $f(x) = \sin(x)$

Solution

1. We apply Definition 9.4.

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{3(x+h)^2+5(x+h)-7-(3x^2+5x-7)}{h} \\ &= \lim_{h\to 0} \frac{3h^2+6xh+5h}{h} \\ &= \lim_{h\to 0}(3h+6x+5) \\ &= 6x+5 \end{aligned}$$

So $f'(x) = 6x+5$. Recall earlier we found that $f'(1) = 11$, which is affirmed by our new computation of $f'(x)$. Moreover, we can verify the correctness of our computation using Mathematica. More precisely, we can compute derivatives in Mathematica using the built-in command **D** as follows.

```
In[11]:= D[3*x^2+5*x-7,x]

Out[11]= 5+6x
```

The second argument of the command **D** indicates the variable with respect to which the derivative is computed.

2. We again apply Definition 9.4,

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{\sin(x+h)-\sin(x)}{h} \\ & \quad \left(\begin{array}{l}\text{Use trig identity } \sin(x+h) = \\ \sin(x)\cos(h)+\cos(x)\sin(h)\end{array}\right) \\ &= \lim_{h\to 0} \frac{\sin(x)\cos(h)+\cos(x)\sin(h)-\sin(x)}{h} && \text{(Sine of sum.)} \\ &= \lim_{h\to 0} \frac{\sin(x)(\cos(h)-1)+\cos(x)\sin(h)}{h} && \text{(Regroup.)} \\ &= \lim_{h\to 0} \left(\frac{\sin(x)(\cos(h)-1)}{h} + \frac{\cos(x)\sin(h)}{h}\right) && \text{(Split into two fractions.)} \\ & \quad \left(\text{use } \lim_{h\to 0}\frac{\cos(h)-1}{h} = 0 \text{ and } \lim_{h\to 0}\frac{\sin(h)}{h} = 1\right) \\ &= \sin(x)\cdot 0 + \cos(x)\cdot 1 && \text{(Special limits.)} \\ &= \cos(x) \end{aligned}$$

We have found that when $f(x) = \sin(x)$, $f'(x) = \cos(x)$. This is not entirely surprising. The sine function is periodic – it repeats itself on regular intervals. Therefore its rate of change also repeats itself on the same regular intervals.

The next example illustrates that the derivative of a function may not always exist.

Example 9.3

Find the derivative of the absolute value function:

$$f(x) = |x| = \begin{cases} -x, & \text{if} \quad x < 0, \\ x, & \text{if} \quad x \geq 0. \end{cases}$$

Its graph is shown in Figures 9.3(a).

Solution

We need to evaluate $\lim\limits_{h\to 0} \frac{f(x+h)-f(x)}{h}$. As f is piecewise–defined, we need to consider separately the limits when $x < 0$, $x > 0$ and $x = 0$.

1. When $x < 0$:

$$\begin{aligned} \frac{d}{dx}\big(-x\big) &= \lim_{h\to 0} \frac{-(x+h)-(-x)}{h} \\ &= \lim_{h\to 0} \frac{-h}{h} \\ &= \lim_{h\to 0}(-1) = -1\,. \end{aligned}$$

2. When $x > 0$, a similar computation shows that $\dfrac{d}{dx}\big(x\big) = 1$.

3. We need to also find the derivative at $x = 0$. By the definition of the derivative at a point, we have

$$f'(0) = \lim_{h\to 0} \frac{f(0+h)-f(0)}{h}\,.$$

Since $x = 0$ is the point where our function's definition switches from one piece to the other, we need to consider left- and right-hand limits. Consider the following, where we compute the left- and right-hand limits side by side.

$$\begin{aligned} &\lim_{\substack{h\to 0 \\ <}} \frac{f(0+h)-f(0)}{h} \\ &= \lim_{\substack{h\to 0 \\ <}} \frac{-h-0}{h} \\ &= \lim_{\substack{h\to 0 \\ <}} -1 = -1 \end{aligned}$$

$$\begin{aligned} &\lim_{\substack{h\to 0 \\ >}} \frac{f(0+h)-f(0)}{h} \\ &= \lim_{\substack{h\to 0 \\ >}} \frac{h-0}{h} \\ &= \lim_{\substack{h\to 0 \\ >}} 1 = 1 \end{aligned}$$

Clearly, the left- and right-hand limits are not equal. Therefore the limit does not exist at 0, and f is not differentiable at 0.

Summarising, we have

$$f'(x) = \begin{cases} -1, & \text{if} \quad x < 0, \\ 1, & \text{if} \quad x > 0. \end{cases}$$

At $x = 0$, $f'(x)$ does not exist; there is a discontinuity at 0 (Figure 9.3(b)). So $f(x) = |x|$ is differentiable everywhere except at 0.

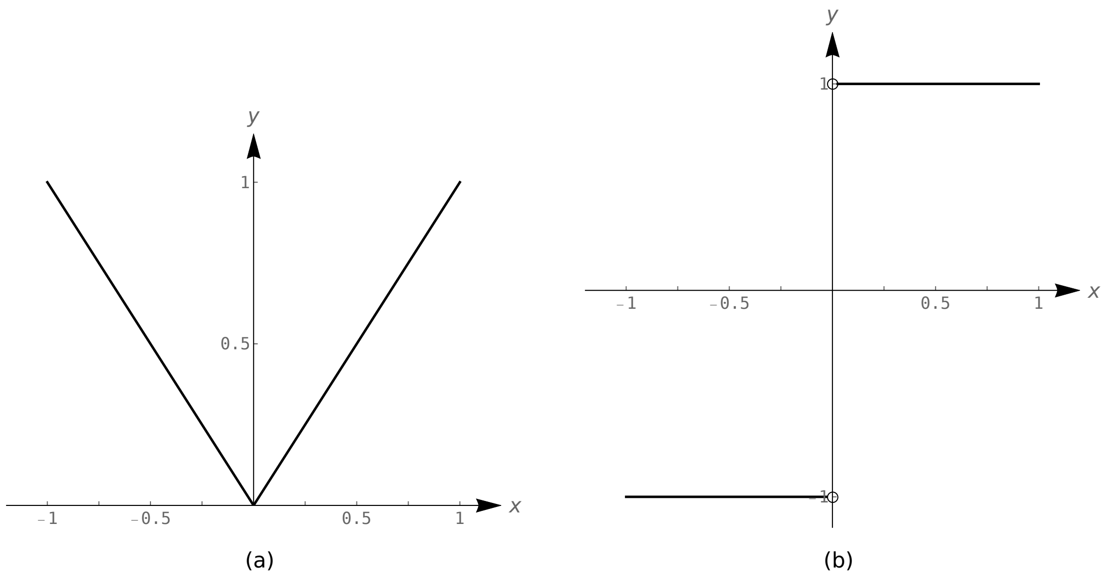


Figure 9.3: The graph of $f(x) = |x|$ (a) and its derivative (b).

The point of non-differentiability came where the piecewise defined function switched from one piece to the other.

Our next example shows that this does, however, not always cause trouble.

Example 9.4

Find the derivative of $f(x)$, given by

$$f(x) = \begin{cases} \sin(x), & x \le \frac{\pi}{2} \\ 1, & x > \frac{\pi}{2}. \end{cases}$$

Its graph is shown in Figure 9.4(a).

Solution

From Example 9.2, we know that when $x < \pi/2$, $f'(x) = \cos(x)$. It is easy to verify that when $x > \pi/2$, $f'(x) = 0$; consider:

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{1-1}{h} = \lim_{h \to 0} 0 = 0.$$

So far, we have

$$f'(x) = \begin{cases} \cos(x) & x < \frac{\pi}{2} \\ 0 & x > \frac{\pi}{2}. \end{cases}$$

We still need to find $f'(\pi/2)$. Notice at $x = \pi/2$ that both pieces of f' are 0, meaning we can state that $f'(\pi/2) = 0$.

Being more rigorous, we can again evaluate the difference quotient limit at $x = \pi/2$, utilizing again left- and right–hand limits:

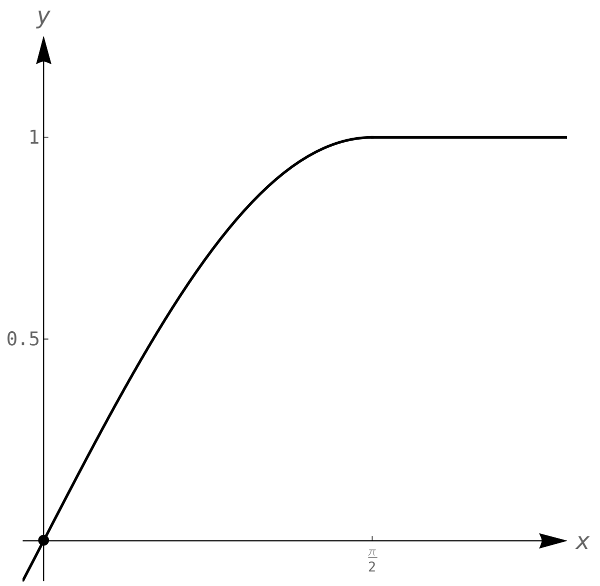
$$\begin{aligned}
&\lim_{\substack{h\to 0\\ <}} \frac{f\left(\frac{\pi}{2}+h\right)-f\left(\frac{\pi}{2}\right)}{h}\\
&= \lim_{\substack{h\to 0\\ <}} \frac{\sin\left(\frac{\pi}{2}+h\right)-\sin\left(\frac{\pi}{2}\right)}{h}\\
&= \lim_{\substack{h\to 0\\ <}} \frac{\sin\left(\frac{\pi}{2}\right)\cos(h)+\sin(h)\cos\left(\frac{\pi}{2}\right)-\sin\left(\frac{\pi}{2}\right)}{h}\\
&= \lim_{\substack{h\to 0\\ <}} \frac{1\cdot\cos(h)+\sin(h)\cdot 0-1}{h}\\
&= 0\,.
\end{aligned}$$

$$\begin{aligned}
&\lim_{\substack{h\to 0\\ >}} \frac{f\left(\frac{\pi}{2}+h\right)-f\left(\frac{\pi}{2}\right)}{h}\\
&= \lim_{\substack{h\to 0\\ >}} \frac{1-1}{h}\\
&= \lim_{\substack{h\to 0\\ >}} \frac{0}{h}\\
&= 0\,.
\end{aligned}$$

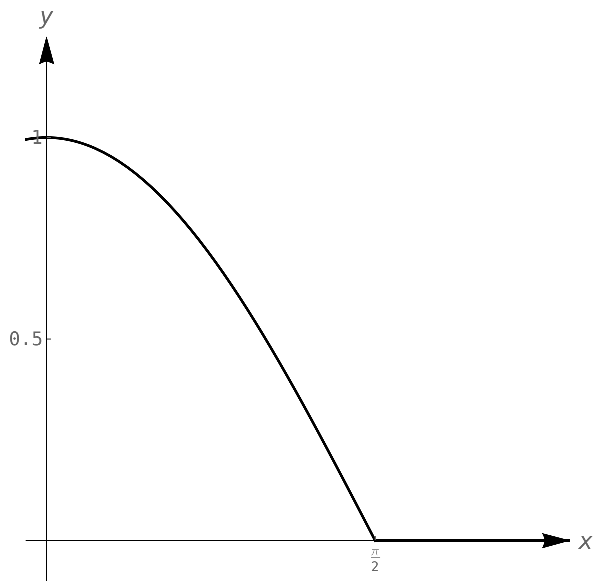
Since both are 0 at $x=\pi/2$, the limit exists and $f'(\pi/2)$ exists (and is 0). Therefore we can fully write f' as

$$f'(x)=\begin{cases}\cos(x), & x\leq\frac{\pi}{2}\\ 0, & x>\frac{\pi}{2}.\end{cases}$$

See Figure 9.4(b) for a graph of this function.



(a)



(b)

Figure 9.4: The graph of $f(x)$ as defined in Example 9.4 (a) and its derivative (b).

Loosely speaking, we defined a continuous function in Chapter 8 as one in which we could sketch its graph without lifting our pencil. Likewise, it can be understood that a function is differentiable if it is a continuous function that does not have any sharp corners. One such sharp corner is shown in Figure 9.3(a). On the other hand, even though the function f in Example 9.4 is piecewise–defined, the transition is smooth hence it is differentiable.

9.1.3 Differentiability on closed intervals

When we defined the derivative at a point in Definition 9.1, we specified that the interval I over which a function f was defined needed to be an open interval. Open intervals are required so that we can take a limit at any point c in I, meaning we want to approach c from both the left and right.

Recall we also required open intervals in Definition 8.3 when we defined what it meant for a function to be continuous. Later, we used one-sided limits to extend continuity to closed intervals. We now extend differentiability to closed intervals by again considering one-sided limits.

Our motivation for doing this is three-fold. First, we consider common sense. In Example 9.2 we found that when $f(x) = 3x^2 + 5x - 7$, $f'(x) = 6x + 5$, and this derivative is defined for all real numbers, hence f is differentiable everywhere. It seems appropriate to also conclude that f is differentiable on closed intervals, like $[0,1]$, as well. After all, $f'(x)$ is defined at both $x = 0$ and $x = 1$. Secondly, consider $f(x) = \sqrt{x}$. The domain of f is $\mathbb{R}^+$. It is natural to ask ourselves whether f is differentiable on its domain – specifically, is f differentiable at 0? Thirdly, having the derivative defined on closed intervals will prove useful throughout the remainder of this chapter.

Below we give a formal definition of differentiability on a closed interval.

Definitie 9.5 (Differentiability on a closed interval)
Let f be continuous on $[a,b]$ and differentiable on $]a,b[$, and let the one-sided limits

$$\lim_{\substack{h \to 0 \\ >}} \frac{f(a+h) - f(a)}{h} \quad \text{and} \quad \lim_{\substack{h \to 0 \\ <}} \frac{f(b+h) - f(b)}{h}$$

exist. Then we say f is **differentiable** on $[a,b]$.

Given the notation of Definition 9.5 and in line with the terminology introduced for one-sided limits, we say that the function f is **right differentiable** (*rechts afleidbaar*) in a and **left differentiable** (*links afleidbaar*) in b. Moreover, the one-sided limit

$$\lim_{\substack{h \to 0 \\ >}} \frac{f(a+h) - f(a)}{h}$$

is called the **right-derivative** (*rechter afgeleide*) of f in a, while the one-sided limit

$$\lim_{\substack{h \to 0 \\ <}} \frac{f(b+h) - f(b)}{h}$$

is referred to as the **left derivative** (*linker afgeleide*) of f in b. Using this terminology, we may say that a function f is differentiable in a if and only if it is left and right differentiable in a and if its left and right derivatives in a are equal.

Example 9.5

Consider the functions $f(x) = \sqrt{x}$ and $g(x) = \sqrt{x^3}$. The domain of these functions is $\mathbb{R}^+$ and it is easy to see that they are differentiable on $\mathbb{R}_0^+$. Determine the differentiability of each at $x = 0$.

Solution

We start by considering f and take the right-hand limit of the difference quotient:

$$\begin{aligned}
\lim_{\substack{h \to 0 \\ >}} \frac{f(a+h) - f(a)}{h} &= \lim_{\substack{h \to 0 \\ >}} \frac{\sqrt{0+h} - \sqrt{0}}{h} \\
&= \lim_{\substack{h \to 0 \\ >}} \frac{\sqrt{h}}{h} \\
&= \lim_{\substack{h \to 0 \\ >}} \frac{1}{\sqrt{h}} = +\infty.
\end{aligned}$$

The one-sided limit of the difference quotient does not exist at $x = 0$ for f; therefore f is differentiable on $\mathbb{R}_0^+$ and not differentiable on $\mathbb{R}^+$.

Now consider g:

$$\begin{aligned}\lim_{\substack{h\to 0\\ >}} \frac{g(a+h)-g(a)}{h} &= \lim_{\substack{h\to 0\\ >}} \frac{\sqrt{(0+h)^3}-\sqrt{0}}{h} \\ &= \lim_{\substack{h\to 0\\ >}} \frac{h^{3/2}}{h} \\ &= \lim_{\substack{h\to 0\\ >}} h^{1/2} = 0.\end{aligned}$$

As the one-sided limit exists at $x = 0$, we conclude g is differentiable on its domain of $\mathbb{R}^+$.

The two functions are graphed in Figure 9.5. Note how $f(x) = \sqrt{x}$ seems to go vertical as x approaches 0, implying the slopes of its tangent lines are growing toward infinity. Also note how the slopes of the tangent lines to $g(x) = \sqrt{x^3}$ approach 0 as x approaches 0.

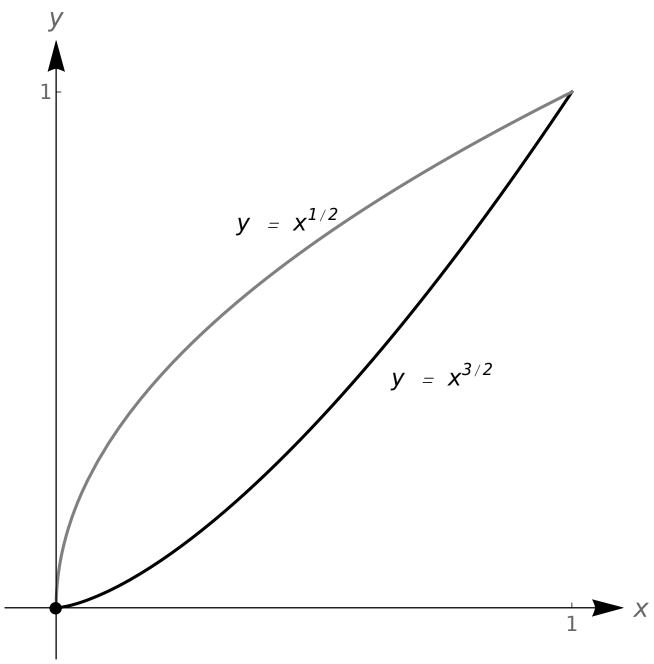


Figure 9.5: A graph of $y = x^{1/2}$ and $y = x^{3/2}$.

9.1.4 Interpretations of the derivative

We offer two interconnected interpretations of the derivative, hopefully explaining why we care about it and why it is worthy of study.

9.1.4.1 Instantaneous rate of change

If f is a function of x, then $f'(x)$ measures the instantaneous rate of change of f with respect to x. It is useful to recognize the units of the derivative function. If y is a function of x, i.e., $y = f(x)$ for some function f, and y is measured in metres and x in seconds, then the units of $y' = f'$ are metres per second. In general, if y is measured in units P and x is measured in units Q, then y' will be measured in units P per Q.

Referring back to the falling amusement–park ride, knowing that at $t = 2$ the velocity was -64 m/s, we could reasonably assume that 1 second later the riders' height would have dropped by about 64 metres. Knowing that the riders were accelerating as they fell would inform us that this is an under–approximation. If all we knew was that $f(2) = 86$ and $f'(2) = -64$, we'd know that we'd have to stop the riders quickly otherwise they would hit the ground.

Example 9.6

Let $P(t)$ represent the world population t minutes after 12:00 a.m., January 1, 2012. It is fairly accurate to say that $P(0) = 7028734178$ (`www.prb.org`). It is also fairly accurate to state that $P'(0) = 156$; that is, at midnight on January 1, 2012, the population of the world was growing by about 156 people per minute. Twenty days later (or, 28,800 minutes later) we could reasonably assume the population grew by about $28800 \cdot 156 = 4492800$ people.

In this example we made use of the important approximation the rate of change was constant. Notationally, we would say that

$$f(c+h) \approx f(c) + f'(c) \cdot h.$$

This approximation is best when h is small. Small is a relative term; when dealing with the world population, $h = 22$ days = 28,800 minutes is small in comparison to years.

One of the most fundamental applications of the derivative is the study of motion. Let $s(t)$ be a position function, where t is time and $s(t)$ is distance. For instance, s could measure the height of a projectile or the distance an object has travelled. Then $s'(t)$ has units metres per second, it measures the instantaneous rate of distance change – it measures **velocity** (*snelheid*).

Now consider $v(t)$, a velocity function. That is, at time t, $v(t)$ gives the velocity of an object. The derivative of v, $v'(t)$, gives the instantaneous rate of velocity change – **acceleration** (*versnelling*). If velocity is measured in metres per second, and time is measured in seconds, then the units of acceleration are metres per second per second, or (m/s)/s. We often shorten this to metres per second squared, but this tends to obscure the meaning of the units.

9.1.4.2 The slope of the tangent line

Given a function $y = f(x)$, the difference quotient

$$\frac{f(c+h) - f(c)}{h}$$

gives a change in y-values divided by a change in x-values; i.e., it is a measure of the slope of the line that goes through two points on the graph of f: $\big(c, f(c)\big)$ and $\big(c+h, f(c+h)\big)$. As h shrinks to 0, these two points come close together; in the limit we find $f'(c)$, the slope of a special line called the tangent line that intersects f only once near $x = c$. Lines have a constant rate of change, their slope. Nonlinear functions do not have a constant rate of change, but we can measure their instantaneous rate of change at a given x value c by computing $f'(c)$. We can get an idea of how f is behaving by looking at the slopes of its tangent lines.

If we know $f(c)$ and $f'(c)$ for some value $x = c$, then computing the tangent line at $(c, f(c))$ is easy:

$$y = \ell(x) = f'(c)(x - c) + f(c).$$

It can then be used to approximate a value of f. More specifically, Let us use the tangent line at $x = c$ to approximate a value of f near $x = c$; i.e., compute $\ell(c+h)$ to approximate $f(c+h)$, assuming again that h is small. We get

$$y = \ell(c+h) = f'(c)\big((c+h)-c\big)+f(c) = f'(c)h+f(c).$$

9.1.5 Higher-order derivatives

The derivative of a function f is itself a function, therefore we can take its derivative. The following definition gives a name to this concept and introduces its notation.

Definitie 9.6 (Higher-order derivatives)
Let $y = f(x)$ be a differentiable function on I. The following are defined, provided the corresponding limits exist.

1. The **second derivative** (*tweede afgeleide*) of f is:

$$f''(x) = \frac{d}{dx}\Big(f'(x)\Big) = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2y}{dx^2} = y''.$$

2. The **third derivative** (*derde afgeleide*) of f is:

$$f'''(x) = \frac{d}{dx}\Big(f''(x)\Big) = \frac{d}{dx}\left(\frac{d^2y}{dx^2}\right) = \frac{d^3y}{dx^3} = y'''.$$

3. The **nth derivative** (*n-de afgeleide*) of f is:

$$f^{(n)}(x) = \frac{d}{dx}\Big(f^{(n-1)}(x)\Big) = \frac{d}{dx}\left(\frac{d^{n-1}y}{dx^{n-1}}\right) = \frac{d^ny}{dx^n} = y^{(n)}.$$

In general, when finding the fourth derivative and so on, we resort to the $f^{(4)}(x)$ notation, not $f''''(x)$; because after a while, too many ticks is confusing. Moreover, the second derivative notation could be written as

$$\frac{d^2y}{dx^2} = \frac{d^2y}{(dx)^2} = \frac{d^2}{(dx)^2}(y).$$

That is, we take the derivative of y twice (hence d^2), both times with respect to x (hence $(dx)^2 = dx^2$). Also higher-order derivatives can be computed in Mathematica using the built-in command **D**. For instance, the second derivative of $y = x^2$ can be computed as follows.

```
In[12]:= D[x^2,{x,2}]

Out[12]= 2
```

The second argument of the function **D** is now a list containing the focal variable and the order of the derivative.

But what do higher order derivatives mean? What is the practical interpretation? Our first answer is

> The second derivative of a function f is the rate of change of the rate of change of f.

One way to grasp this concept is to let f describe a position function. Then, f' describes the rate of position change: velocity. We now consider f'', which describes the rate of velocity change. Its derivative describes the rate of change of the rate of position change, which we know as acceleration.

It can, however, be difficult to consider the meaning of the third, and higher order, derivatives. The third derivative is the rate of change of the rate of change of the rate of change of f. That is essentially meaningless to the uninitiated. In the context of our position/velocity/acceleration example, the third derivative is the rate of change of acceleration, commonly referred to as jerk.

Make no mistake: higher-order derivatives have great importance even if their practical interpretations are hard to understand. The mathematical topic of series makes extensive use of higher-order derivatives (Section 9.8.1).

9.1.6 Smoothness

Some of the graphs we encountered so far have sharp corners, or **cusps**, where the corresponding functions are not differentiable. This leads us to a definition.

Definitie 9.7 (Smoothness)
A **smooth function** (*gladde functie*) is a function that has continuous derivatives up to some desired order over some domain. A function can therefore be said to be smooth over a restricted interval I. Moreover, a function f is **piecewise smooth** (*stuksgewijs gladde functie*) on I if I can be partitioned into subintervals where f is smooth on each subinterval.

The number of continuous derivatives necessary for a function to be considered smooth depends on the problem at hand, and may vary from two to infinity.

Example 9.7

Determine whether or not the following functions are smooth.

1. $y = x^3$
2. $y = x\,|x|$

Solution

1. We observe that the first derivative is given by $y' = 3x^2$, the second derivative by $y'' = 6x$, the third derivative by 6, while the fourth and higher-derivatives are all 0. Clearly, all these derivatives are continuous, so $y = x^3$ is a smooth function.
2. For the sake of understanding, let us rewrite the equation as

$$y = \begin{cases} x^2, & \text{if } x \geq 0, \\ -x^2, & \text{if } x < 0\,. \end{cases}$$

Its first derivative is given by

$$y' = \begin{cases} 2x, & \text{if } x \geq 0, \\ -2x, & \text{if } x < 0\,, \end{cases}$$

and is continuous everywhere. Yet, its second derivative,

$$y'' = \begin{cases} 2, & \text{if } x \geq 0, \\ -2, & \text{if } x < 0 \end{cases}$$

is not continuous everywhere, as there is a discontinuity at $x = 0$. Consequently, this function is not smooth.

9.2 Basic differentiation rules

The derivative is a powerful tool but is admittedly awkward given its reliance on limits. Fortunately, one thing mathematicians are good at is abstraction.

9.2.1 Derivatives of algebraic and transcendental functions

Let us consider a linear function, $y = mx + b$. What is y'? Without limits, recognize that linear function are characterized by being functions with a constant rate of change (the slope). The derivative, y', gives the instantaneous rate of change; with a linear function, this is constant, m. Thus $y' = m$.

Let us abstract once more. Let us find the derivative of the general quadratic function,

$$f(x) = ax^2 + bx + c\,.$$

Using the definition of the derivative, we have:

$$\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{a(x+h)^2 + b(x+h) + c - (ax^2 + bx + c)}{h} \\
&= \lim_{h\to 0} \frac{ah^2 + 2ahx + bh}{h} \\
&= \lim_{h\to 0} (ah + 2ax + b) \\
&= 2ax + b.
\end{aligned}$$

So if $y = 6x^2 + 11x - 13$, we can immediately compute $y' = 12x + 11$.

In a similar way, using Definition 9.4 we can easily find the derivatives of the algebraic and transcendental functions we studied in Chapter 4 and 5, respectively. We find, for the constant function $f(x) = c$

$$\frac{d}{dx}(c) = 0\,,$$

where $c \in \mathbb{R}$. This indicates the logical fact that constant functions have no rate of change. For the other algebraic functions we encountered we have

- $\dfrac{d}{dx}\big(x^n\big) = nx^{n-1}$, where $n \in \mathbb{Z}$,
- $\dfrac{d}{dx}\big(x^a\big) = ax^{a-1}$, where $a \in \mathbb{R}_0$ and $x > 0$.

Hence, we immediately get

$$\frac{d}{dx}\big(\sqrt{x}\big) = \frac{1}{2\sqrt{x}},$$

where $x > 0$.

For what concerns the exponential and logarithmic functions, we get the following derivative functions:

- $\dfrac{d}{dx}(e^x) = e^x$,
- $\dfrac{d}{dx}(a^x) = a^x \ln(a)$, where $a > 0$,
- $\dfrac{d}{dx}\big(\ln(x)\big) = \dfrac{1}{x}$,
- $\dfrac{d}{dx}\big(\log_a(x)\big) = \dfrac{1}{x\ln(a)}$, where $a > 0, a \neq 1$,

while for the trigonometric functions we get:

- $\dfrac{d}{dx}\big(\sin(x)\big) = \cos(x)$,
- $\dfrac{d}{dx}\big(\cos(x)\big) = -\sin(x)$,
- $\dfrac{d}{dx}\big(\tan(x)\big) = \dfrac{1}{\cos^2(x)} = \sec^2(x)$,
- $\dfrac{d}{dx}\big(\cot(x)\big) = \dfrac{-1}{\sin^2(x)} = -\csc^2(x)$.

Example 9.8

Let $f(x) = x^3$.

1. Find $f'(x)$.
2. Find the equation of the line tangent to the graph of f at $x = -1$.
3. Use the tangent line to approximate $(-1.1)^3$.
4. Sketch f, f' and the found tangent line at $x = -1$ on the same axis.

Solution

1. Using the rule for differentiating a power function, we directly get $f'(x) = 3x^2$.
2. To find the equation of the line tangent to the graph of f at $x = -1$, we need a point and the slope. The point is $(-1, f(-1)) = (-1, -1)$. The slope is $f'(-1) = 3$. Thus the tangent line has equation $y = 3(x - (-1)) + (-1) = 3x + 2$.
3. We can use the tangent line to approximate $(-1.1)^3$ as -1.1 is close to -1. We have

$$(-1.1)^3 \approx 3(-1.1) + 2 = -1.3.$$

   We can easily find the actual answer; $(-1.1)^3 = -1.331$.
4. See Figure 9.6.

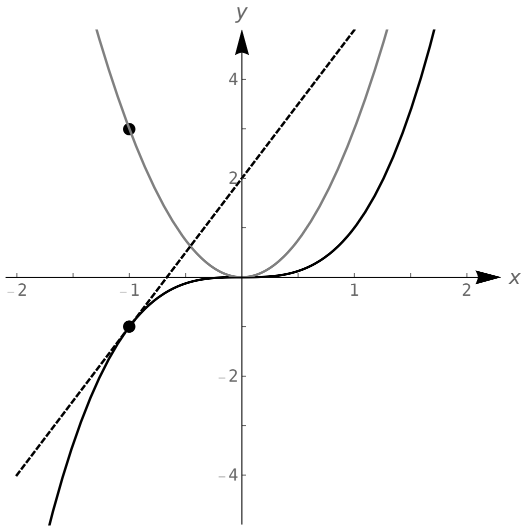


Figure 9.6: A graph of $f(x) = x^3$, along with its derivative $f'(x) = 3x^2$ (gray) and its tangent line at $x = -1$ (dashed).

9.2.2 Properties of the derivative

Using the derivatives of the basic algebraic and transcendental functions, we can easily find the derivative of $y = x^3$, but we cannot compute the derivative of $y = 2x^3$, $y = x^3 + \sin(x)$ nor $y = x^3 \sin(x)$. The following theorem helps with the first two of these examples.

Theorem 9.1 (Properties of the derivative)
Let f and g be differentiable on an open interval I and let c be a real number. Then the following properties hold:

1. ***Sum/Difference rule**:*

$$\frac{d}{dx}\Big(f(x) \pm g(x)\Big) = \frac{d}{dx}\Big(f(x)\Big) \pm \frac{d}{dx}\Big(g(x)\Big) = f'(x) \pm g'(x). \tag{9.1}$$

2. ***Constant multiple rule***:

$$\frac{d}{dx}\Big(c \cdot f(x)\Big) = c \cdot \frac{d}{dx}\Big(f(x)\Big) = c \cdot f'(x). \tag{9.2}$$

Theorem 9.1 allows us to find the derivatives of a wide variety of functions. It can be used in conjunction with the power rule to find the derivatives of any polynomial. Recall in Example 9.2 that we found, using the limit definition, the derivative of $f(x) = 3x^2 + 5x - 7$. We can now find its derivative without expressly using limits:

$$\begin{aligned}\frac{d}{dx}\Big(3x^2 + 5x + 7\Big) &= 3\frac{d}{dx}\Big(x^2\Big) + 5\frac{d}{dx}\Big(x\Big) + \frac{d}{dx}\Big(7\Big) \\ &= 3 \cdot 2x + 5 \cdot 1 + 0 \\ &= 6x + 5.\end{aligned}$$

Example 9.9

Find the first four derivatives of the following functions:

1. $f(x) = 4x^2$
2. $g(x) = \sin(x)$
3. $h(x) = 5e^x$

Solution

1. Using Theorem 9.1, we have: $f'(x) = 8x$. Continuing on, we have

$$f''(x) = \frac{d}{dx}\big(8x\big) = 8; \qquad f'''(x) = 0; \qquad f^{(4)}(x) = 0.$$

   Notice how all successive derivatives will also be 0.

2. Resorting repeatedly to the list of derivatives of elementary functions, we get the following:.

$$g(x) = \cos x; \qquad g'(x) = -\sin x; \qquad g''(x) = -\cos x; \qquad g^{(4)}(x) = \sin x.$$

   Note how we have come right back to $g(x)$ again.

3. Again resorting to the list of derivatives of elementary functions, we can see that

$$h(x) = h'(x) = h''(x) = h^{(4)}(x) = 5e^x.$$

When having to compute the derivative a product of functions, we may turn to the product rule.

Theorem 9.2 (Product rule)
Let f and g be differentiable functions on an open interval I. Then fg is a differentiable function on I, and

$$\frac{d}{dx}\Big(f(x)g(x)\Big) = f(x)g'(x) + f'(x)g(x).$$

Example 9.10

Find the derivatives of the following:

1. $y = 5x^2 \sin(x)$
2. $y = x^3 \ln(x)\cos(x)$

Solution

1. To make our use of the product rule explicit, let us set $f(x) = 5x^2$ and $g(x) = \sin(x)$. We easily compute/recall that $f'(x) = 10x$ and $g'(x) = \cos(x)$. Employing the product rule, we have
$$\frac{d}{dx}\Big(5x^2 \sin(x)\Big) = 5x^2 \cos(x) + 10x \sin(x).$$
2. We have a product of three functions, so, we get
$$y' = 3x^2 \ln(x)\cos(x) + x^3 \frac{1}{x}\cos(x) + x^3 \ln(x)(-\sin(x)).$$

We have learned how to compute the derivatives of sums, differences, and products of functions. We now learn how to find the derivative of a quotient of functions.

Theorem 9.3 (Quotient rule)
Let f and g be differentiable functions defined on an open interval I, where $g(x) \neq 0$ on I. Then f/g is differentiable on I, and

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

Example 9.11

Find the derivatives of the following:

1. $y = \dfrac{5x^2}{\sin(x)}$
2. $y = \tan(x)$

Solution

1. Directly applying the quotient rule gives:
$$y' = \frac{10x\,\sin(x) - 5x^2\,\cos(x)}{\sin^2(x)}.$$
2. Though we could resort to the list of derivatives of elementary functions, we can proceed as well by recalling that $\tan(x) = \sin(x)/\cos(x)$, so we can apply the quotient rule.
$$\begin{aligned}\frac{d}{dx}\Big(\tan(x)\Big) &= \frac{d}{dx}\left(\frac{\sin(x)}{\cos(x)}\right)\\ &= \frac{\cos(x)\cos(x) - \sin(x)(-\sin(x))}{\cos^2(x)}\\ &= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)}\end{aligned}$$

$$= \frac{1}{\cos^2(x)}$$

$$= \sec^2(x).$$

The derivatives of the cotangent, cosecant and secant functions can all be computed using the known derivatives of the cosine and sine function together with the quotient rule, so there is no need to learn those by heart.

Taking the derivative of many functions is relatively straightforward. It is clear what rules apply and in what order they should be applied. Other functions present multiple paths; different rules may be applied depending on how the function is treated. One of the beautiful things about calculus is that there is not the right way; each path, when applied correctly, leads to the same result, the derivative.

9.3 The chain rule

We have covered almost all of the derivative rules that deal with combinations of two (or more) functions. The operations of addition, subtraction, multiplication (including by a constant) and division led to the sum, difference, constant multiple, power rule, product and quotient rules. To complete the list of differentiation rules, we look at the last way two (or more) functions can be combined: the process of composition.

One example of a composition of functions is $f(x) = \cos(x^2)$. We currently do not know how to compute this derivative. If forced to guess, one would likely guess $f'(x) = -\sin(2x)$, where we recognize $-\sin(x)$ as the derivative of $\cos(x)$ and $2x$ as the derivative of x^2. However, this is not the case; $f'(x) \neq -\sin(2x)$.

Before we define this new rule, recall the notation for composition of functions. We write $(f \circ g)(x)$ or $f(g(x))$, read as f of g of x, to denote composing f with g. In shorthand, we simply write $f \circ g$ or $f(g)$ and read it as f of g. When composing functions, we need to make sure that the new function is actually defined. For instance, consider $f(x) = \sqrt{x}$ and $g(x) = -x^2 - 1$. The domain of f excludes all negative numbers, but the range of g is only negative numbers. Therefore the composition $f\big(g(x)\big) = \sqrt{-x^2 - 1}$ is not defined for any x, and hence is not differentiable.

The following theorem of the **chain rule** (*kettingregel*) takes care to ensure this problem does not arise. We'll focus more on the derivative result than on the domain/range conditions.

Theorem 9.4 (The chain rule)
Let g be a differentiable function on an interval I, let the range of g be a subset of the interval J, and let f be a differentiable function on J. Then $y = f(g(x))$ is a differentiable function on I, and

$$y' = f'(g(x))\, g'(x).$$

Acknowledging the chain rule, we can immediately state the so-called generalized power rule.

Theorem 9.5 (Generalized power rule)
Let $g(x)$ be a differentiable function and let $n \neq 0$ be an integer. Then

$$\frac{d}{dx}\Big(g(x)^n\Big) = n\big(g(x)\big)^{n-1}g'(x).$$

It is instructive to understand what the chain rule looks like using $\frac{dy}{dx}$ notation instead of y' notation. Suppose that $y = f(u)$ is a function of u, where $u = g(x)$ is a function of x, as stated in Theorem 9.4. Then, through the composition $f \circ g$, we can think of y as a function of x, as $y = f(g(x))$. Thus the derivative of y with respect to x makes sense; we can talk about $\frac{dy}{dx}$. This leads to an interesting progression of notation:

$$\begin{aligned} y' &= f'(g(x))g'(x) \\ \frac{dy}{dx} &= y'(u)u'(x) && \text{(Since } y = f(u) \text{ and } u = g(x).\text{)} \\ \frac{dy}{dx} &= \frac{dy}{du}\frac{du}{dx} && \text{(Using fractional notation for the derivative.)} \end{aligned}$$

It might seem as though the du terms cancel out, but it is important to realize that we are not cancelling these terms; the derivative notation of $\frac{dy}{du}$ is one symbol. It is equally important to realize that this notation was chosen precisely because of this behaviour. It makes applying the chain rule easy with multiple variables and/or with multiple functions. For instance, if we consider three functions $y = f(u)$, $u = h(v)$ en $v = g(x)$, then we may consider the function composition $y = f(h(g(x)))$. The derivative of y with respect to x is then given by

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dv}\frac{dv}{dx}$$

or equivalently

$$\frac{dy}{dx} = \frac{df(u)}{du}\frac{dh(v)}{dv}\frac{dg(x)}{dx}.$$

We now consider some examples that employ the chain rule.

Example 9.12

Find the derivatives of the following functions:

1. $y = \sin(2x)$
2. $y = \ln(4x^3 - 2x^2)$
3. $y = e^{-x^2}$

Solution

1. Consider $y = \sin(2x)$. Recognize that this is a composition of functions, where $f(u) = \sin(u)$ and $u = g(x) = 2x$. Thus

$$y' = y'(u)u'(x) = \cos(u)2 = 2\cos(2x).$$

2. Recognize that $y = \ln(4x^3 - 2x^2)$ is the composition of $f(u) = \ln(u)$ and $u = g(x) = 4x^3 - 2x^2$.

This leads us to:

$$y' = \frac{1}{4x^3-2x^2}(12x^2-4x) = \frac{12x^2-4x}{4x^3-2x^2} = \frac{4x(3x-1)}{2x(2x^2-x)} = \frac{2(3x-1)}{2x^2-x}.$$

3. Recognize that $y = e^{-x^2}$ is the composition of $f(u) = e^u$ and $u = g(x) = -x^2$. Remembering that $f'(x) = e^x$, we have
$$y' = e^{-x^2}(-2x) = (-2x)e^{-x^2}.$$

Of course, the chain rule can be applied in conjunction with any of the other rules we have already learned.

Example 9.13

Find the derivatives of the following:

1. $y = x^5 \sin(2x^3)$
2. $y = \tan^5(6x^3 - 7x)$
3. $y = \dfrac{x\cos(x^{-2}) - \sin^2(e^{4x})}{\ln(x^2+5x^4)}$.

Solution

1. We must use the product and chain rules and proceed step–by–step.
$$y' = x^5\big(6x^2\cos(2x^3)\big) + 5x^4\big(\sin(2x^3)\big) = 6x^7\cos(2x^3) + 5x^4\sin(2x^3).$$

2. Recognize that we have the $g(x) = \tan(6x^3-7x)$ function inside the $f(x) = x^5$ function. We begin using the generalized power rule; in this first step, we do not fully compute the derivative. Rather, we are approaching this step–by–step.
$$y' = 5\tan^4(6x^3-7x)g'(x).$$
We now find $g'(x)$. We again need the chain rule;
$$g'(x) = \sec^2(6x^3-7x)(18x^2-7).$$
Combine this with what we found above to give
$$\begin{aligned} y' &= 5\tan^4(6x^3-7x)\sec^2(6x^3-7x)(18x^2-7) \\ &= (90x^2-35)\sec^2(6x^3-7x)\tan^4(6x^3-7x). \end{aligned}$$

3. Using the quotient, product and chain rules we get the following answer without simplification:
$$y' = \frac{\ln(x^2+5x^4)\cdot\Big[\big(x(-\sin(x^{-2}))(-2x^{-3}) + 1\cos(x^{-2})\big) - 2\sin(e^{4x})\cos(e^{4x})(4e^{4x})\Big]}{\big(\ln(x^2+5x^4)\big)^2} - \frac{\Big(x\cos(x^{-2}) - \sin^2(e^{4x})\Big)\cdot\frac{2x+20x^3}{x^2+5x^4}}{\big(\ln(x^2+5x^4)\big)^2}$$

This example demonstrates that derivatives can be computed systematically, no matter how arbitrarily complicated the function is. A key to correctly working the considered problems is to break the problem down into smaller, more manageable pieces. For instance, when using the product and chain rules together, consider the first part of the product rule at first: $f(x)g'(x)$. Just rewrite $f(x)$, then find $g'(x)$. Then move on to the $f'(x)g(x)$ part. Do not attempt to figure out both parts at once. Likewise, using the quotient rule, approach the numerator in two steps and handle the denominator after completing that. Only simplify afterwards.

The chain rule also has theoretic value. That is, it can be used to find the derivatives of certain functions.

Example 9.14

Use the chain rule to find the derivative of $y = 2^x$.

Solution

We only know how to find the derivative of one exponential function, $y = e^x$. We can accomplish our goal by rewriting 2 in terms of e. Recalling that e^x and $\ln x$ are inverse functions, so we can write

$$2 = e^{\ln(2)} \quad \text{and so} \quad y = 2^x = \left(e^{\ln(2)}\right)^x = e^{x(\ln(2))}.$$

The function is now the composition $y = f(g(x))$, with $f(u) = e^u$ and $g(x) = x(\ln(2))$. Since $f'(u) = e^u$ and $g'(x) = \ln(2)$, the chain rule gives

$$y' = e^{x(\ln(2))}\ln(2).$$

Recall that the $e^{x(\ln(2))}$ term on the right hand side is just 2^x, our original function. Thus, the derivative contains the original function itself. We have

$$y' = y\ln(2) = 2^x\ln(2).$$

We can extend this process to use any base a, where $a > 0$ and $a \neq 1$. All we need to do is replace each “2” in our work with “a.” In this way, the chain rule, coupled with the derivative rule of e^x, allows us to find the derivatives of all exponential functions.

The comment at the end of previous example is important. Let $f(x) = a^x$, for $a > 0, a \neq 1$. Then f is differentiable everywhere and

$$f'(x) = \ln(a)a^x.$$

Likewise, it can be shown that

$$\frac{d}{dx}\big(\log_a(x)\big) = \frac{1}{\ln(a)}\frac{1}{x}.$$

In the next section, we use the chain rule to justify another differentiation technique. There are many curves that we can draw in the plane that fail the vertical line test. See for instance Section 4.4, where we studied amongst other things the equation $x^2 + y^2 = 1$, which describes the unit circle. We may still be interested in finding slopes of tangent lines to the circle at various points. The next section shows how we can find $\frac{dy}{dx}$ without first solving for y. While we can in this instance, in many other instances solving for y is impossible. In these situations, implicit differentiation is indispensable.

9.4 Implicit differentiation

9.4.1 First derivative

In the previous sections we learned to find the derivative when y is given explicitly as a function of x. That is, if we know $y = f(x)$ for some function f, we can find y'. Sometimes the relationship between y and x is not explicit; rather, it is implicit. For instance, we might know that $x^2 - y = 4$. Can we still find y'? In this case, sure; we solve for y to get $y = x^2 - 4$ and then differentiate to get $y' = 2x$. Sometimes, however, the implicit relationship between x and y is complicated. Suppose we are given $\sin(y) + y^3 = 6 - x^3$. In this case there is absolutely no way to solve for y in terms of elementary functions. The surprising thing is, however, that we can still find y' via implicit differentiation.

Implicit differentiation is a technique based on the chain rule that is used to find a derivative when the relationship between the variables is given implicitly rather than explicitly.

Let f and g be functions of x. Then we have according to the chain rule

$$\frac{d}{dx}\Big(f(g(x))\Big) = f'(g(x))\,g'(x).$$

Suppose now that $y = g(x)$. We can rewrite the above as

$$\frac{d}{dx}\Big(f(y)\Big) = f'(y)y', \quad \text{or} \quad \frac{d}{dx}\Big(f(y)\Big) = f'(y)\frac{dy}{dx}. \tag{9.3}$$

These equations look strange; the key concept to learn here is that we can find y' even if we do not exactly know how y and x relate.

Example 9.15

Given that

$$\sin(y) + y^3 = 6 - x^3,$$

find y' and the equation of the tangent line at the point $(\sqrt[3]{6}, 0)$.

Solution

We start by taking the derivative of both sides, which maintains the equality. We have :

$$\frac{d}{dx}\Big(\sin(y) + y^3\Big) = \frac{d}{dx}\Big(6 - x^3\Big).$$

The right-hand side is easy; it returns $-3x^2$.

The left-hand side requires more consideration. We take the derivative term–by–term. Using Equation (9.3), we can see that

$$\frac{d}{dx}\Big(\sin(y)\Big) = \cos(y)\,y'.$$

We apply the same process to the y^3 term.

$$\frac{d}{dx}\Big(y^3\Big) = \frac{d}{dx}\Big((y)^3\Big) = 3(y)^2\,y'.$$

Similarly, the derivative of y^3 is $3y^2y'$. Putting all this together with the right hand side, we have

$$\cos(y)y' + 3y^2y' = -3x^2.$$

Now solve for y'.

$$\begin{aligned} & \cos(y)y' + 3y^2y' = -3x^2 \\ \Leftrightarrow \quad & \big(\cos(y) + 3y^2\big)y' = -3x^2 \\ \Leftrightarrow \quad & y' = \frac{-3x^2}{\cos(y) + 3y^2} \end{aligned}$$

We can now find the slope of the tangent line at the point $(\sqrt[3]{6}, 0)$ by substituting $\sqrt[3]{6}$ for x and 0 for y. Thus at the point $(\sqrt[3]{6}, 0)$, we have the slope as

$$y' = \frac{-3(\sqrt[3]{6})^2}{\cos(0) + 3 \cdot 0^2} = \frac{-3\sqrt[3]{36}}{1} \approx -9.91.$$

Therefore the equation of the tangent line to the implicitly defined function $\sin y + y^3 = 6 - x^3$ at the point $(\sqrt[3]{6}, 0)$ is

$$y = -3\sqrt[3]{36}(x - \sqrt[3]{6}) + 0 \approx -9.91x + 18.$$

The curve and this tangent line are shown in Figure 9.7.

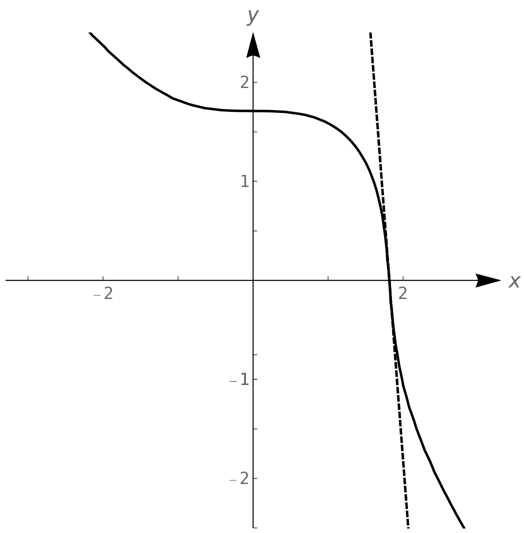


Figure 9.7: The function $\sin y + y^3 = 6 - x^3$ and its tangent line (dashed) at the point $(\sqrt[3]{6}, 0)$.

This example suggests a general method for implicit differentiation. For the steps below assume y is a function of x.

1. Take the derivative of each term in the equation. Treat the x terms like normal. When taking the derivatives of y terms, the usual rules apply except that, because of the chain rule, we need to multiply each term by y'.
2. Get all the y' terms on one side of the equal sign and put the remaining terms on the other side.
3. Factor out y'; solve for y' by dividing.

Example 9.16

Given the implicitly defined function

$$\sin(x^2y^2) + y^3 = x + y,$$

find y'.

Solution

Differentiating term by term, we find the most difficulty in the first term. It requires both the chain and product rules.

$$\begin{aligned}\frac{d}{dx}\Big(\sin(x^2y^2)\Big) &= \cos(x^2y^2)\,\frac{d}{dx}\Big(x^2y^2\Big)\\ &= \cos(x^2y^2)\big(x^2(2yy') + 2xy^2\big)\\ &= 2(x^2yy' + xy^2)\cos(x^2y^2).\end{aligned}$$

We leave the derivatives of the other terms to the reader. After taking the derivatives of both sides, we have

$$2(x^2yy' + xy^2)\cos(x^2y^2) + 3y^2y' = 1 + y'.$$

We now have to be careful to properly solve for y', particularly because of the product on the left. It is best to multiply out the product. Doing this, we get

$$2x^2y\cos(x^2y^2)y' + 2xy^2\cos(x^2y^2) + 3y^2y' = 1 + y'.$$

From here we can safely move around terms to get the following:

$$2x^2y\cos(x^2y^2)y' + 3y^2y' - y' = 1 - 2xy^2\cos(x^2y^2).$$

Then we can solve for y' to get

$$y' = \frac{1 - 2xy^2\cos(x^2y^2)}{2x^2y\cos(x^2y^2) + 3y^2 - 1}.$$

A graph of this implicit function is given in Figure 9.8. It is easy to verify that the points $(0,1)$ and $(0,-1)$ all lie on the graph. We can find the slopes of the tangent lines at each of these points using our formula for y':

- at $(0,1)$, the slope is $1/2$;
- at $(0,-1)$, the slope is also $1/2$.

The tangent lines have been added to the graph of the function in Figure 9.8.

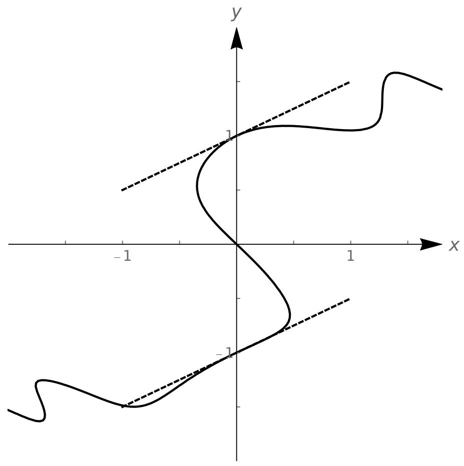


Figure 9.8: A graph of the implicitly defined function $\sin(x^2y^2)+y^3 = x+y$ and tangent lines at $(0,1)$ and $(0,-1)$.

We may also use Mathematica to check our answer for what concerns y'. We just have to be careful to explicitly mention the dependence of y on x.

```
In[13]:= D[Sin[x^2 y[x]^2] + y[x]^3 == x + y[x],x]

Out[13]= 3 y[x]^2 y'[x]+Cos[x^2 y[x]^2] (2 x y[x]^2+2 x^2 y[x] y'[x])==1+y'[x]
```

Implicit functions are generally harder to deal with than explicit functions. With an explicit function, given an x value, we have an explicit formula for computing the corresponding y value. With an implicit function, one often has to find x and y values at the same time that satisfy the equation. It is much easier to demonstrate that a given point satisfies the equation than to actually find such a point.

9.4.2 Higher-order derivatives

We can use implicit differentiation to find higher-order derivatives as well. In theory, this is simple: first find $\frac{dy}{dx}$, then take its derivative with respect to x. In practice, it is not hard, but it often requires a bit of algebra. We demonstrate this in an example.

Example 9.17

Given $x^2+y^2=1$, find y''.

Solution

Taking derivatives, we get $2x+2yy'=0$. Solving for y' gives:

$$y' = \frac{-x}{y}.$$

To find y'', we apply implicit differentiation to y'.

$$\begin{aligned} y'' &= \frac{d}{dx}(y') \\ &= \frac{d}{dx}\left(-\frac{x}{y}\right) && \text{(Use the quotient rule.)} \end{aligned}$$

$$= -\frac{y \cdot 1 - x(y')}{y^2}$$

replace y' with $-x/y$:

$$= -\frac{y - x(-x/y)}{y^2}$$
$$= -\frac{y + x^2/y}{y^2}.$$

While this is not a particularly simple expression, it is usable. For instance, we can see that $y'' > 0$ when $y < 0$ and $y'' < 0$ when $y > 0$. In Section 10.4, we will see how this relates to the shape of the graph.

9.4.3 Logarithmic differentiation

Consider the function $y = x^x$; it is graphed in Figure 9.9. It is well defined for $x > 0$ and we might be interested in finding equations of lines tangent and normal to its graph. How do we take its derivative?

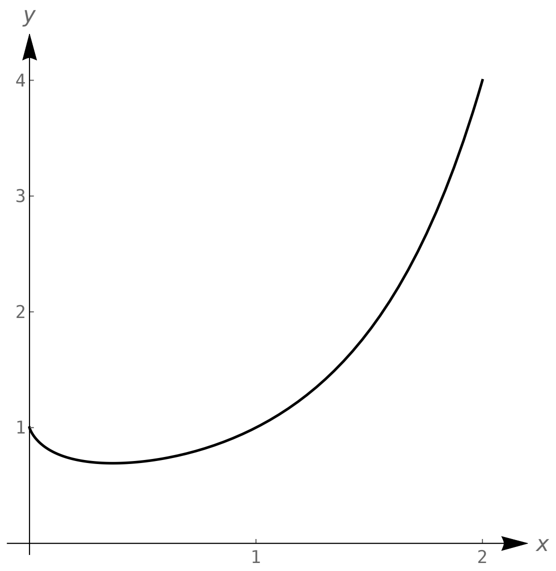


Figure 9.9: A plot of $y = x^x$.

The function is not a power function: it has a power of x, not a constant. It is not an exponential function: it has a base of x, not a constant. A differentiation technique known as **logarithmic differentiation** (*logaritmisch afleiden*) becomes useful here. The basic principle is this: take the natural log of both sides of an equation $y = f(x)$, then use implicit differentiation to find y'. We demonstrate this in the following example.

Example 9.18

Given $y = x^x$, use logarithmic differentiation to find y'.

Solution

We start by taking the natural log of both sides then applying implicit differentiation.

$$
\begin{array}{rrcll}
 & y & = & x^x & \\
\Leftrightarrow & \ln(y) & = & \ln(x^x) & \text{(Apply logarithm rule.)} \\
\Leftrightarrow & \ln(y) & = & x\ln(x) & \text{(Use implicit differentiation.)} \\
\Leftrightarrow & \dfrac{d}{dx}\Big(\ln(y)\Big) & = & \dfrac{d}{dx}\Big(x\ln(x)\Big) & \\
\Leftrightarrow & \dfrac{y'}{y} & = & \ln(x) + x\,\dfrac{1}{x} & \\
\Leftrightarrow & \dfrac{y'}{y} & = & \ln(x)+1 & \\
\Leftrightarrow & y' & = & y\big(\ln(x)+1\big) & (y = x^x\,.) \\
\Leftrightarrow & y' & = & x^x\big(\ln(x)+1\big). &
\end{array}
$$

9.5 Derivatives of inverse functions

Recall that a function $y = f(x)$ is said to be injective if it passes the horizontal line test; that is, for two different x values x_1 and x_2, we do not have $f(x_1) = f(x_2)$. In some cases the domain of f must be restricted so that it is injective. For instance, consider $f(x) = x^2$. Clearly, $f(-1) = f(1)$, so f is not one to one on its regular domain, but by restricting f to $\mathbb{R}_0^+$, f is one to one.

Now recall that injective functions have inverses. That is, if f is one to one, it has an inverse function, denoted by f^{-1}, such that if $f(a) = b$, then $f^{-1}(b) = a$. The domain of f^{-1} is the range of f, and vice-versa. For ease of notation, we set $g = f^{-1}$ and treat g as a function of x.

When the point (a,b) lies on the graph of f, the point (b,a) lies on the graph of g. This made us to discover in Section 3.4 that the graph of g is the reflection of f across the line $y = x$. Because of this relationship, whatever we know about f can quickly be transferred into knowledge about g.

For example, consider Figure 9.10 where the tangent line to f at the point (a,b) is drawn. That line has slope $f'(a)$. Through reflection across $y = x$, we can see that the tangent line to g at the point (b,a) should have slope $\frac{1}{f'(a)}$. This then tells us that $g'(b) = \frac{1}{f'(a)}$.

We have discovered a relationship between f' and g' in a mostly graphical way. We can realize this relationship analytically as well. Let $y = g(x)$, where again $g = f^{-1}$. We want to find y'. Since $y = g(x)$, we know that $f(y) = x$. Using the chain rule and implicit differentiation, take the derivative of both sides of this last equality:

$$
\begin{array}{rrcl}
 & \dfrac{d}{dx}\Big(f(y)\Big) & = & \dfrac{d}{dx}\Big(x\Big) \\
\Leftrightarrow & f'(y)\,y' & = & 1 \\
\Leftrightarrow & y' & = & \dfrac{1}{f'(y)} \\
\Leftrightarrow & y' & = & \dfrac{1}{f'(g(x))}.
\end{array}
$$

This leads us to the following theorem.

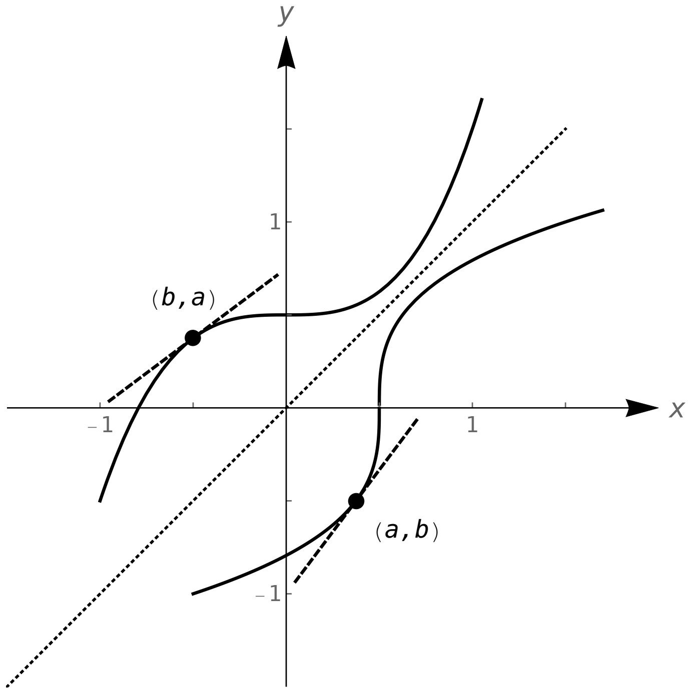


Figure 9.10: Corresponding tangent lines drawn to f and f^{-1}.

Theorem 9.6 (Derivatives of inverse functions)
Let f be differentiable and injective on an open interval I, where $f'(x) \neq 0$ for all x in I, let J be the range of f on I, let g be the inverse function of f, and let $f(a) = b$ for some a in I. Then g is a differentiable function on J, and in particular,

$$1.\ \left(f^{-1}\right)'(b) = g'(b) = \frac{1}{f'(a)} \quad \text{and} \quad 2.\ \left(f^{-1}\right)'(x) = g'(x) = \frac{1}{f'(g(x))}$$

The results of Theorem 9.6 are not trivial; the notation may seem confusing at first. Careful consideration, along with examples, should earn understanding.

In the next example we apply Theorem 9.6 to the arcsine function.

Example 9.19

Let $y = \arcsin(x)$. Find y'.

Solution

Adopting our previously defined notation, let $g(x) = \arcsin(x)$ and $f(x) = \sin(x)$. Consequently, $f'(x) = \cos(x)$. Applying the theorem, we have

$$\begin{aligned} g'(x) &= \frac{1}{f'(g(x))} \\ &= \frac{1}{\cos(\arcsin(x))}. \end{aligned}$$

This last expression is not immediately illuminating. Drawing a figure while assuming that $x >$ will help, as shown in Figure 9.11.

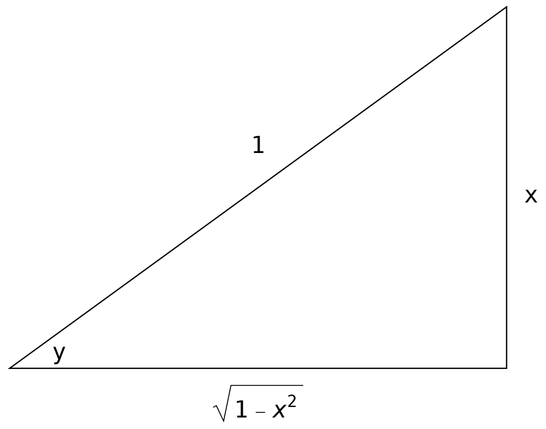


Figure 9.11: A right triangle defined by $y = \arcsin(x/1)$ with the length of the third leg found using the Pythagorean theorem.

Recall that the sine function can be viewed as taking in an angle and returning a ratio of sides of a right triangle, specifically, the ratio opposite over hypotenuse. This means that the arcsine function takes as input a ratio of sides and returns an angle. The equation $y = \arcsin(x)$ can be rewritten as $y = \arcsin(x/1)$; that is, consider a right triangle where the hypotenuse has length 1 and the side opposite of the angle with measure y has length x. This means the final side has length $\sqrt{1-x^2}$, using the Pythagorean theorem.

Therefore

$$\cos(\arcsin(x)) = \cos(y) = \frac{\sqrt{1-x^2}}{1} = \sqrt{1-x^2},$$

resulting in

$$\frac{d}{dx}\big(\arcsin(x)\big) = g'(x) = \frac{1}{\sqrt{1-x^2}}.$$

Remember that the input x of the arcsine function is a ratio of a side of a right triangle to its hypotenuse; the absolute value of this ratio will never be greater than 1. Therefore the inside of the square root will never be negative.

Using similar techniques as in Example 9.19, we can find the derivatives of all the inverse trigonometric

The inverse trigonometric functions are differentiable on all open sets contained in their domains (as listed in Table 5.5) and their derivatives are as follows:

- $\dfrac{d}{dx}\big(\arcsin(x)\big) = \dfrac{1}{\sqrt{1-x^2}},$
- $\dfrac{d}{dx}\big(\arccos(x)\big) = -\dfrac{1}{\sqrt{1-x^2}},$
- $\dfrac{d}{dx}\big(\arctan(x)\big) = \dfrac{1}{1+x^2}.$
- $\dfrac{d}{dx}\big(\operatorname{arccot}(x)\big) = -\dfrac{1}{1+x^2},$

In Section 9.2, we stated without proof or explanation that $\frac{d}{dx}\big(\ln(x)\big) = \frac{1}{x}$. We can justify that now using Theorem 9.6, as shown in the following example.

Example 9.20

Use Theorem 9.6 to compute

$$\frac{d}{dx}\big(\ln(x)\big).$$

Solution

View $y = \ln(x)$ as the inverse of $y = e^x$. Therefore, using our standard notation, let $f(x) = e^x$ and $g(x) = \ln(x)$. We wish to find $g'(x)$. Theorem 9.6 gives:

$$\begin{aligned} g'(x) &= \frac{1}{f'(g(x))} \\ &= \frac{1}{e^{\ln(x)}} \\ &= \frac{1}{x}. \end{aligned}$$

9.6 L'Hôpital's rule

Our treatment of limits in Chapter 8 exposed us to the notion of 0/0, an indeterminate form. If $\lim_{x\to c} f(x) = 0$ and $\lim_{x\to c} g(x) = 0$, we do not conclude that $\lim_{x\to c} f(x)/g(x)$ is $0/0$; rather, we use $0/0$ as notation to describe the fact that both the numerator and denominator approach 0. The expression 0/0 has no numeric value; other work must be done to evaluate the limit.

Other indeterminate forms exist; they are: ∞/∞, $0\cdot\infty$, $\infty-\infty$, 0^0, 1^∞ and ∞^0. Just as 0/0 does not mean divide 0 by 0, the expression ∞/∞ does not mean divide infinity by infinity. Instead, it means a quantity is growing without bound and is being divided by another quantity that is growing without bound. We cannot determine from such a statement what value, if any, results in the limit.

9.6.1 Indeterminate forms 0/0 and ∞/∞

Here, we introduce l'Hôpital's rule, a method of resolving limits that produce the indeterminate forms 0/0 and ∞/∞. We will also show how algebraic manipulation can be used to convert other indeterminate expressions into one of these two forms so that our new rule can be applied.

Theorem 9.7 (L'Hôpital's rule for 0/0)
Let $\lim_{x\to c} f(x) = 0$ and $\lim_{x\to c} g(x) = 0$, where f and g are differentiable functions on an open interval I containing c, and $g'(x) \neq 0$ on I except possibly at c. Then

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)}.$$

We demonstrate the use of l'Hôpital's rule (LHR) in the following examples.

Example 9.21

Evaluate the following limits.

1. $\lim_{x\to 0} \frac{\sin(x)}{x}$
2. $\lim_{x\to 1} \frac{\sqrt{x+3}-2}{1-x}$
3. $\lim_{x\to 0} \frac{x^2}{1-\cos(x)}$

Solution

1. We proved this limit is 1 in Example 8.6 using the squeeze theorem. Here we use l'Hôpital's rule to show its power.
$$\lim_{x\to 0}\frac{\sin(x)}{x} \overset{\text{by LHR}}{=} \lim_{x\to 0}\frac{\cos(x)}{1} = 1.$$
2. $$\lim_{x\to 1}\frac{\sqrt{x+3}-2}{1-x} \overset{\text{by LHR}}{=} \lim_{x\to 1}\frac{\frac{1}{2}(x+3)^{-1/2}}{-1} = -\frac{1}{4}.$$
3. $$\lim_{x\to 0}\frac{x^2}{1-\cos(x)} \overset{\text{by LHR}}{=} \lim_{x\to 0}\frac{2x}{\sin(x)}.$$
   This latter limit also evaluates to the 0/0 indeterminate form. To evaluate it, we apply l'Hôpital's rule again.
$$\lim_{x\to 0}\frac{2x}{\sin(x)} \overset{\text{by LHR}}{=} \lim_{x\to 0}\frac{2}{\cos(x)} = 2.$$
   Thus $\lim_{x\to 0}\dfrac{x^2}{1-\cos(x)} = 2.$

Note that at each step where l'Hôpital's rule was applied, it was needed: the initial limit returned the indeterminate form of $0/0$. If the initial limit returns, for example, 1/2, then l'Hôpital's rule does not apply.

The following theorem extends our initial version of l'Hôpital's rule in two ways. It allows the technique to be applied to the indeterminate form ∞/∞ and to limits where x approaches $\pm\infty$.

Theorem 9.8 (L'Hôpital's rule for ∞/∞)

1. *Let $\lim_{x\to a} f(x) = \pm\infty$ and $\lim_{x\to a} g(x) = \pm\infty$, where f and g are differentiable on an open interval I containing a. Then*
$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)}.$$
2. *Let f and g be differentiable functions on the open interval $]a,+\infty[$ for some value a, where $g'(x) \neq 0$ on $]a,+\infty[$ and $\lim_{x\to+\infty} f(x)/g(x)$ returns either "0/0" or "∞/∞". Then*
$$\lim_{x\to+\infty}\frac{f(x)}{g(x)} = \lim_{x\to+\infty}\frac{f'(x)}{g'(x)}.$$
   A similar statement can be made for limits where x approaches $-\infty$.

Example 9.22

Evaluate the following limits.

1. $\displaystyle\lim_{x\to+\infty}\frac{3x^2-100x+2}{4x^2+5x-1000}$
2. $\displaystyle\lim_{x\to+\infty}\frac{e^x}{x^3}$

Solution

1. We can evaluate this limit already using Theorem 8.8; the answer is 3/4. We apply l'Hôpital's

rule to demonstrate its applicability.

$$\lim_{x\to+\infty}\frac{3x^2-100x+2}{4x^2+5x-1000} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{6x-100}{8x+5} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{6}{8}=\frac{3}{4}.$$

2. We directly find that

$$\lim_{x\to+\infty}\frac{e^x}{x^3} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{e^x}{3x^2} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{e^x}{6x} \overset{\text{by LHR}}{=} \lim_{x\to+\infty}\frac{e^x}{6}=+\infty.$$

Recall that this means that the limit does not exist; as x approaches $+\infty$, the expression e^x/x^3 grows without bound. We can infer from this that e^x grows faster than x^3; as x gets large, e^x is far larger than x^3.

9.6.2 Indeterminate forms $0\cdot\infty$ and $\infty-\infty$

L'Hôpital's rule can only be applied to ratios of functions. When faced with an indeterminate form such as $0\cdot\infty$ or $\infty-\infty$, we can sometimes apply algebra to rewrite the limit so that l'Hôpital's rule can be applied. We demonstrate the general idea in the next example.

Example 9.23

Evaluate the following limits.

1. $\lim\limits_{\substack{x\to 0\\ >}}\left(x\,e^{1/x}\right)$
2. $\lim\limits_{\substack{x\to 0\\ <}}\left(x\,e^{1/x}\right)$
3. $\lim\limits_{x\to+\infty}\left(\ln(x+1)-\ln(x)\right)$

Solution

1. As $x \underset{>}{\to} 0$, $x\to 0$ and $e^{1/x}\to e^{+\infty}\to+\infty$. Thus we have the indeterminate form $0\cdot\infty$. We rewrite the expression $x\cdot e^{1/x}$ as
$$\frac{e^{1/x}}{1/x};$$
now, as $x \underset{>}{\to} 0$, we get the indeterminate form ∞/∞ to which l'Hôpital's rule can be applied.
$$\lim_{\substack{x\to 0\\ >}}\left(x\,e^{1/x}\right)=\lim_{\substack{x\to 0\\ >}}\frac{e^{1/x}}{1/x} \overset{\text{by LHR}}{=} \lim_{\substack{x\to 0\\ >}}\frac{(-1/x^2)e^{1/x}}{-1/x^2}=\lim_{\substack{x\to 0\\ >}}e^{1/x}=+\infty.$$
So, we may conclude that $e^{1/x}$ grows faster than x shrinks to zero, meaning their product grows without bound.

2. As $x \underset{<}{\to} 0$, $x\to 0$ and $e^{1/x}\to e^{-\infty}\to 0$. The the limit evaluates to $0\cdot 0$ which is not an indeterminate form. We conclude then that
$$\lim_{\substack{x\to 0\\ <}}\left(x\,e^{1/x}\right)=0.$$

3. This limit initially evaluates to the indeterminate form $\infty-\infty$. By applying a logarithmic rule,

we can rewrite the limit as

$$\lim_{x\to+\infty}\big(\ln(x+1)-\ln(x)\big)=\lim_{x\to+\infty}\ln\left(\frac{x+1}{x}\right).$$

As $x\to+\infty$, the argument of the ln term approaches ∞/∞, but

$$\lim_{x\to+\infty}\frac{x+1}{x}=\lim_{x\to+\infty}\frac{x(1+1/x)}{x}=1.$$

Since $x\to+\infty\Rightarrow\frac{x+1}{x}\to1$, it follows that $x\to+\infty$ implies

$$\ln\left(\frac{x+1}{x}\right)\to\ln 1=0.$$

Thus

$$\lim_{x\to+\infty}\big(\ln(x+1)-\ln(x)\big)=\lim_{x\to+\infty}\ln\left(\frac{x+1}{x}\right)=0.$$

Since this limit evaluates to 0, it means that for large x, there is essentially no difference between $\ln(x+1)$ and $\ln(x)$; their difference is essentially 0.

9.6.3 Indeterminate forms 0^0, 1^∞ and ∞^0

When faced with an indeterminate form that involves a power, it often helps to employ the natural logarithmic function. More precisely, we rely on the fact that if $\lim\limits_{x\to c}\ln\big(f(x)\big)=L$, then

$$\lim_{x\to c}f(x)=\lim_{x\to c}e^{\ln(f(x))}=e^L.$$

Example 9.24

Evaluate the following limits.

1. $\lim\limits_{x\to+\infty}\left(1+\dfrac{1}{x}\right)^x$

2. $\lim\limits_{\substack{x\to0\\>}}x^x$.

Solution

1. This is equivalent to a special limit given in Theorem 8.4. Note that the exponent approaches $+\infty$ while the base approaches 1, leading to the indeterminate form $1^{+\infty}$. Let $f(x)=(1+1/x)^x$; the problem asks to evaluate $\lim\limits_{x\to+\infty}f(x)$. Let us first evaluate $\lim\limits_{x\to+\infty}\ln\big(f(x)\big)$.

$$\begin{aligned}\lim_{x\to+\infty}\ln\big(f(x)\big)&=\lim_{x\to+\infty}\ln\left(1+\frac{1}{x}\right)^x\\&=\lim_{x\to+\infty}x\ln\left(1+\frac{1}{x}\right)\end{aligned}$$

$$= \lim_{x\to+\infty} \frac{\ln\left(1+\frac{1}{x}\right)}{1/x}$$

This produces the indeterminate form 0/0, so we apply l'Hôpital's rule.

$$\begin{aligned}
&= \lim_{x\to+\infty} \frac{\frac{1}{1+1/x}\cdot(-1/x^2)}{(-1/x^2)} \\
&= \lim_{x\to+\infty} \frac{1}{1+1/x} \\
&= 1.
\end{aligned}$$

Thus $\lim\limits_{x\to+\infty} \ln\big(f(x)\big) = 1$. We finally return to the original limit:

$$\lim_{x\to+\infty} \left(1+\frac{1}{x}\right)^x = \lim_{x\to+\infty} f(x) = \lim_{x\to+\infty} e^{\ln(f(x))} = e^1 = e.$$

2. This limit leads to the indeterminate form 0^0. Let $f(x) = x^x$ and consider first $\lim\limits_{x\underset{>}{\to}0} \ln\big(f(x)\big)$.

$$\begin{aligned}
\lim_{x\underset{>}{\to}0} \ln\big(f(x)\big) &= \lim_{x\underset{>}{\to}0} \ln\left(x^x\right) \\
&= \lim_{x\underset{>}{\to}0} x\ln(x) \\
&= \lim_{x\underset{>}{\to}0} \frac{\ln(x)}{1/x}.
\end{aligned}$$

This produces the indeterminate form $\frac{\infty}{\infty}$ so we apply l'Hôpital's rule.

$$\begin{aligned}
&= \lim_{x\underset{>}{\to}0} \frac{1/x}{-1/x^2} \\
&= \lim_{x\underset{>}{\to}0} -x \\
&= 0.
\end{aligned}$$

Thus $\lim\limits_{x\underset{>}{\to}0} \ln\big(f(x)\big) = 0$. We finally return to the original limit:

$$\lim_{x\underset{>}{\to}0} x^x = \lim_{x\underset{>}{\to}0} f(x) = \lim_{x\underset{>}{\to}0} e^{\ln(f(x))} = e^0 = 1.$$

This result is supported by the graph of $f(x) = x^x$ given in Figure 9.9.

9.7 Applications of the derivative

9.7.1 Newton's method

Solving equations is one of the most important things we do in mathematics, yet we are surprisingly limited in what we can solve analytically. For instance, equations as simple as $x^5+x+1 = 0$ or $\cos x = x$ cannot be solved by algebraic methods in terms of familiar functions. Fortunately, there are methods that can give us approximate solutions to equations like these. These methods can usually give an approximation correct to as many decimal places as we like. In Section 8.5 we learned about the bisection method. Here, we focus on another technique (which generally works faster), called **Newton's method** (*methode van Newton*) .

Newton's method is built around tangent lines. The main idea is that if x is sufficiently close to a root of $f(x)$, then the tangent line to the graph at $(x,f(x))$ will cross the x-axis at a point closer to the root than x.

We start Newton's method with an initial guess about roughly where the root is. Call this x_0 (Figure 9.12(a)). Draw the tangent line to the graph at $(x_0,f(x_0))$ and see where it meets the x-axis. Call this point x_1. Then repeat the process – draw the tangent line to the graph at $(x_1,f(x_1))$ and see where it meets the x-axis (Figure 9.12(b)). Call this point x_2. Repeat the process again to get x_3, x_4, etc. This sequence of points will often converge rather quickly to a root of f (Figure 9.12(c)).

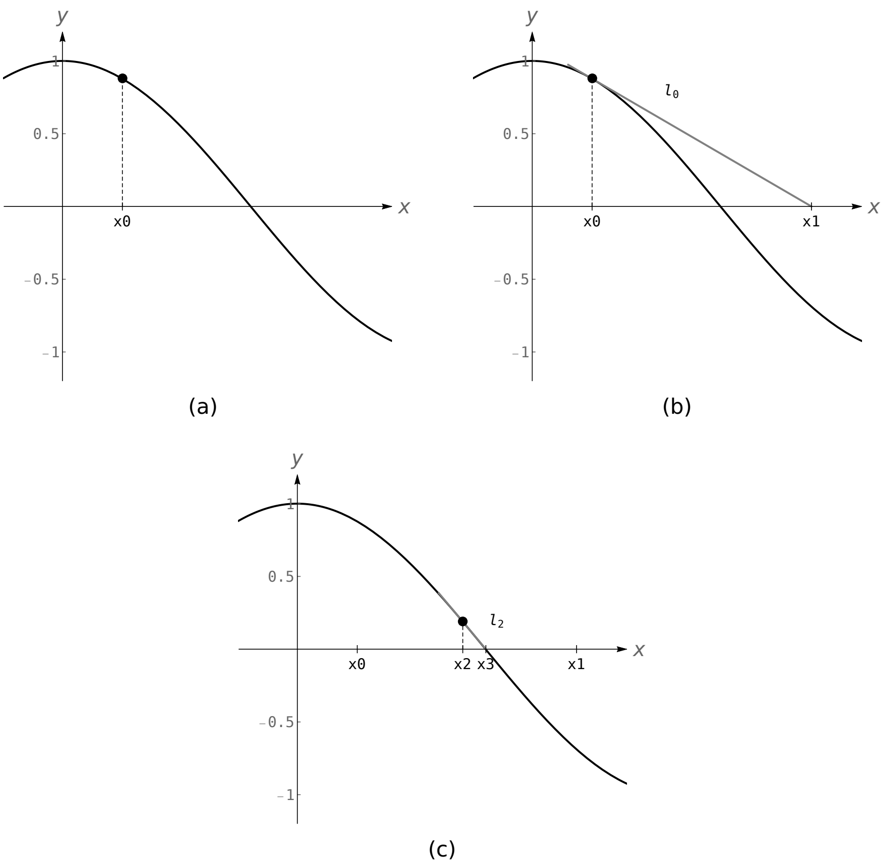


Figure 9.12: The geometric concept behind Newton's method. Note how x_3 is very close to a solution to $f(x) = 0$.

We can use this geometric process to create an algebraic process. Let us look at how we found x_1. We started with the tangent line to the graph at $(x_0,f(x_0))$. The slope of this tangent line is $f'(x_0)$ and the equation of the line is

$$y = f'(x_0)(x - x_0) + f(x_0).$$

This line crosses the x-axis when $y = 0$, and the x–value where it crosses is what we called x_1. So let $y = 0$ and replace x with x_1, giving the equation:

$$0 = f'(x_0)(x_1 - x_0) + f(x_0).$$

Now solve for x_1:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Since we repeat the same geometric process to find x_2 from x_1, we have

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

In general, given an approximation x_n, we can find the next approximation, x_{n+1} as follows:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

We summarize this process as follows for a function f that is differentiable on an interval I with a root in I. To approximate the value of the root, accurate to d decimal places:

1. Choose a value x_0 as an initial approximation of the root. This is often done by just looking at a graph of f.
2. Create successive approximations iteratively; given an approximation x_n, compute the next approximation x_{n+1} as
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$
3. Stop the iterations when successive approximations do not differ in the first d places after the decimal point.

Newton's method is not infallible. The sequence of approximate values may not converge, or it may converge so slowly that one is tricked into thinking a certain approximation is better than it actually is. Even though it is not (directly) a method for solving equations like $f(x) = g(x)$, this is not a problem; since we can rewrite the latter equation as $f(x) - g(x) = 0$ and then use Newton's method.

We can of course automate this process on a computer, but for now, let us see how Newton's method works using a concrete example on paper.

Example 9.25

Approximate the real root of $x^3 - x^2 - 1 = 0$, accurate to the first 3 places after the decimal, using Newton's method and an initial approximation of $x_0 = 1$.

Solution

To begin, we compute $f'(x) = 3x^2 - 2x$. Then we apply the Newton's method algorithm.

$$x_1 = 1 - \frac{f(1)}{f'(1)} = 1 - \frac{1^3 - 1^2 - 1}{3 \cdot 1^2 - 2 \cdot 1} = 2,$$

$$x_2 = 2 - \frac{f(2)}{f'(2)} = 2 - \frac{2^3 - 2^2 - 1}{3 \cdot 2^2 - 2 \cdot 2} = 1.625,$$

$$x_3 = 1.625 - \frac{f(1.625)}{f'(1.625)} = 1.625 - \frac{1.625^3 - 1.625^2 - 1}{3 \cdot 1.625^2 - 2 \cdot 1.625} \approx 1.48579.$$

$$x_4 = 1.48579 - \frac{f(1.48579)}{f'(1.48579)} \approx 1.46596$$

$$x_5 = 1.46596 - \frac{f(1.46596)}{f'(1.46596)} \approx 1.46557$$

We performed 5 iterations of Newton's method to find a root accurate to the first 3 places after the decimal; our final approximation is 1.465. The exact value of the root, to six decimal places, is 1.465571; It turns out that our x_5 is accurate to more than just 3 decimal places.

A graph of $f(x)$ is given in Figure 9.13. We can see from the graph that our initial approximation of $x_0 = 1$ was not particularly accurate; a closer guess would have been $x_0 = 1.5$. Our choice was based on ease of initial calculation, and shows that Newton's method can be robust enough that we do not have to make a very accurate initial approximation.

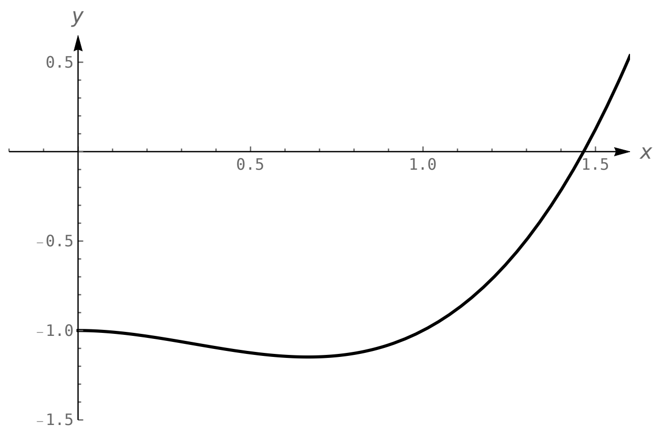


Figure 9.13: A graph of $f(x) = x^3 - x^2 - 1$ in Example 9.25.

While Newton's method does not always work, it does work most of the time, and it is generally very fast. Once the approximations get close to the root, Newton's method can as much as double the number of correct decimal places with each successive approximation.

9.7.2 Differentials

Recall that the derivative of a function f can be used to find the slopes of lines tangent to the graph of f. At $x = c$, the tangent line to the graph of f has equation

$$y = f'(c)(x - c) + f(c).$$

The tangent line can be used to find good approximations of $f(x)$ for values of x near c.

We now generalize this concept. Given $f(x)$ and an x value c, the tangent line is

$$y = \ell(x) = f'(c)(x - c) + f(c).$$

Clearly, $f(c) = \ell(c)$. Let Δx be a small number, representing a small change in x value. We assert that:

$$f(c + \Delta x) \approx \ell(c + \Delta x),$$

since the tangent line to a function approximates well the values of that function near $x = c$.

As the x-value changes from c to $c + \Delta x$, the y-value of f changes from $f(c)$ to $f(c + \Delta x)$. We call this change of y-value Δy. That is:

$$\Delta y = f(c + \Delta x) - f(c).$$

Replacing $f(c + \Delta x)$ with its tangent line approximation, we have

$$\begin{aligned} \Delta y &\approx \ell(c + \Delta x) - f(c) \\ &= f'(c)\big((c + \Delta x) - c\big) + f(c) - f(c) \\ &= f'(c)\Delta x. \end{aligned} \tag{9.4}$$

This final equation is important; it becomes the basis of the upcoming definition. In short, it says that when the x-value changes from c to $c + \Delta x$, the y value of a function f changes by about $f'(c)\Delta x$.

We now introduce two new variables, dx and dy in the context of a formal definition.

Definitie 9.8 (Differentials of x and y)
Let $y = f(x)$ be differentiable. The **differential** (*differentiaal*) of x, denoted dx, is any nonzero real number (usually taken to be a small number). The differential of y, denoted dy, is

$$dy = f'(x)dx.$$

We can solve for $f'(x)$ in the above equation: $f'(x) = dy/dx$. This states that the derivative of f with respect to x is the differential of y divided by the differential of x; this is not the alternate notation for the derivative, dy/dx. This latter notation was chosen because of the fraction–like qualities of the derivative, but again, it is one symbol and not a fraction.

In general, if $y = f(x)$ is a differentiable function, we have the following.

1. Let Δx represent a small, nonzero change in x-value.
2. Let dx represent a small, nonzero change in x-value (i.e., $\Delta x = dx$).
3. Let Δy be the change in y value as x changes by Δx; hence

$$\Delta y = f(x + \Delta x) - f(x).$$

4. Let $dy = f'(x)dx$ which, by Equation (9.4), is an approximation of the change in y-value as x changes by Δx; $dy \approx \Delta y$.

Differentials provide both practical and theoretical benefits. We explore both here.

Example 9.26

Consider $f(x) = x^2$. Knowing $f(3) = 9$, approximate $f(3.1)$.

Solution

The x-value is changing from $x = 3$ to $x = 3.1$; therefore, we see that $dx = 0.1$. If we know how much the y value changes from $f(3)$ to $f(3.1)$ (i.e., if we know Δy), we will know exactly what $f(3.1)$ is since we already know $f(3)$. We can approximate Δy with dy.

$$\begin{aligned} \Delta y &\approx dy \\ &= f'(3)dx \\ &= 2 \cdot 3 \cdot 0.1 = 0.6. \end{aligned}$$

We expect the y value to change by about 0.6, so we approximate $f(3.1) \approx 9.6$.

Of course, it is easy to compute the actual answer: $3.1^2 = 9.61$. So why bother? In most real life situations, we do not know the function that describes a particular behaviour. Instead, we can only take measurements of how things change – measurements of the derivative.

Imagine water flowing down a winding channel. It is easy to measure the speed and direction (i.e., the velocity) of water at any location. It is very hard to create a function that describes the overall flow, hence it is hard to predict where a floating object placed at the beginning of the channel will end up. However, we can approximate the path of an object using differentials. Over small intervals, the path taken by a floating object is essentially linear. Differentials allow us to approximate the true path by piecing together lots of short, linear paths. This technique is called **Euler's method** (*methode van Euler*).

We use differentials once more to approximate the value of a function. Even though calculators are very accessible, it is neat to see how these techniques can sometimes be used to easily compute something that looks rather hard.

Differentials will turn out to be important when we discuss integration (Chapter 12) and proper handling of integrals comes with proper handling of differentials. In light of that, we practice finding differentials in general.

Example 9.27

In each of the following, find the differential dy.

1. $y = \sin(x)$
2. $y = e^x(x^2+2)$
3. $y = \sqrt{x^2+3x-1}$

Solution

1. As $f(x) = \sin(x)$, $f'(x) = \cos(x)$. Thus
$$dy = \cos(x)dx.$$
2. Let $f(x) = e^x(x^2+2)$. We need $f'(x)$, requiring the product rule. We have
$$f'(x) = e^x(x^2+2) + 2xe^x,$$
so
$$dy = \big(e^x(x^2+2) + 2xe^x\big)dx.$$
3. Let $f(x) = \sqrt{x^2+3x-1}$; we need $f'(x)$, requiring the chain rule. We have
$$f'(x) = \frac{1}{2}(x^2+3x-1)^{-\frac{1}{2}}(2x+3) = \frac{2x+3}{2\sqrt{x^2+3x-1}}.$$
Thus
$$dy = \frac{(2x+3)dx}{2\sqrt{x^2+3x-1}}.$$

Finding the differential dy of $y = f(x)$ is really no harder than finding the derivative of f; we just multiply $f'(x)$ by dx. It is important to remember that we are not simply adding the symbol "dx" at the end.

We have seen a practical use of differentials as they offer a good method of making certain approximations. Another use is **error propagation** (*foutenpropagatie*). Suppose a length is measured to be x, although the actual value is $x+\Delta x$ (where Δx is the error, which we hope is small). This measurement of x may be used to compute some other value; we can think of this latter value as $f(x)$ for some function f. As the true length is $x+\Delta x$, one really should have computed $f(x+\Delta x)$. The difference between $f(x)$ and $f(x+\Delta x)$ is the propagated error. How close are $f(x)$ and $f(x+\Delta x)$? This is a difference in y-values:

$$f(x+\Delta x)-f(x)=\Delta y \approx dy.$$

We can approximate the propagated error using differentials.

Example 9.28

A steel ball bearing is to be manufactured with a diameter of 2cm. The manufacturing process has a tolerance of ±0.1mm in the diameter. Given that the density of steel is about 7.85g/cm^3, estimate the propagated error in the mass of the ball bearing.

Solution

The mass of a ball bearing is found using the equation mass = volume × density. In this situation the mass function is a product of the radius of the ball bearing, hence it is $m=7.85\frac{4}{3}\pi r^3$. The differential of the mass is

$$dm = 31.4\pi r^2 dr.$$

The radius is to be 1cm; the manufacturing tolerance in the radius is ±0.05mm, or ±0.005cm. The propagated error is approximately:

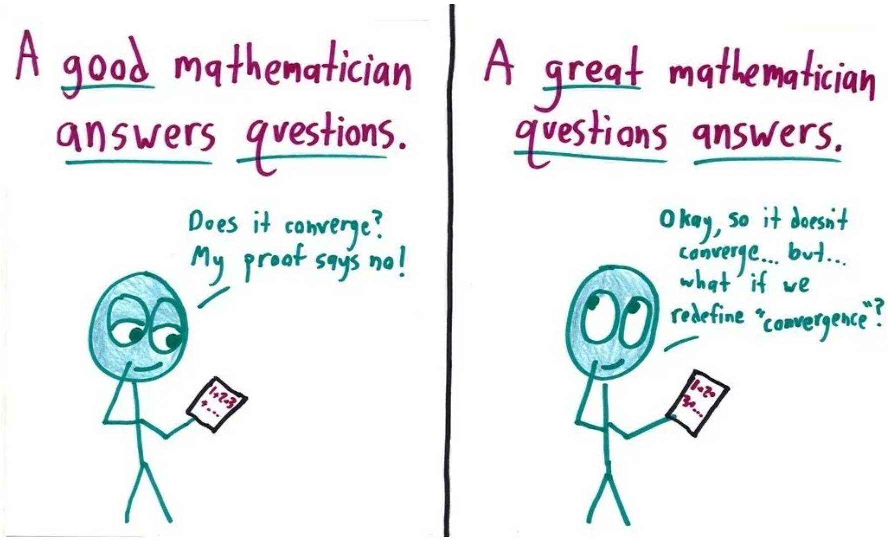
$$\begin{aligned}\Delta m &\approx dm\\ &= 31.4\pi(1)^2(\pm 0.005)\\ &= \pm 0.493\text{g}\end{aligned}$$

Is this error significant? It certainly depends on the application, but we can get an idea by computing the **relative error** (*relatieve fout*). The ratio between amount of error to the total mass is

$$\begin{aligned}\frac{dm}{m} &= \pm\frac{0.493}{7.85\frac{4}{3}\pi}\\ &= \pm\frac{0.493}{32.88}\\ &= \pm 0.015,\end{aligned}$$

or ±1.5%.

If the diameter of the ball was supposed to be 10cm, the same manufacturing tolerance would give a propagated error in mass of ±12.33g, which corresponds to a percent error of ±0.188%. While the amount of error is much greater (12.33 > 0.493), the percent error is much lower.



From *Math with Bad Drawings*, used by permission of Ben Orlin.

We conclude this chapter with one more very important engineering application of the derivative.

9.8 Taylor series

9.8.1 Taylor polynomials

Consider a function $y = f(x)$ and a point $\big(c, f(c)\big)$. The derivative, $f'(c)$, gives the instantaneous rate of change of f at $x = c$. Of all lines that pass through the point $\big(c, f(c)\big)$, the line that best approximates f at this point is the tangent line; that is, the line whose slope (rate of change) is $f'(c)$.

In Figure 9.14(a), we see a function $y = f(x)$ graphed, while its derivatives at $x = 0$ are given in table 9.1:

Table 9.1: Derivatives of a function $f(x)$ evaluated at $x = 0$.

$f(0) = 2$	$f'''(0) = -1$
$f'(0) = 1$	$f^{(4)}(0) = -12$
$f''(0) = 2$	$f^{(5)}(0) = -19$

This table shows that $f(0) = 2$ and $f'(0) = 1$; therefore, the tangent line to f at $x = 0$ is $p_1(x) = 1(x-0)+2 = x+2$. The tangent line is also given in the figure. Note that near $x = 0$, $p_1(x) \approx f(x)$; that is, the tangent line approximates f well. One shortcoming of this approximation is that the tangent line only matches the slope of f; it does not, for instance, match the concavity of f. We can find a polynomial, $p_2(x)$, that does match the concavity without much difficulty, though. The table of derivatives gives the following information:

$$f(0) = 2 \qquad f'(0) = 1 \qquad f''(0) = 2.$$

Therefore, we want our polynomial $p_2(x)$ to have these same properties. That is, we need

$$p_2(0) = 2 \qquad p_2'(0) = 1 \qquad p_2''(0) = 2.$$

We can solve this as follows. To keep $p_2(x)$ as simple as possible, we will assume that not only $p_2''(0) = 2$, but that $p_2''(x) = 2$. That is, the second derivative of p_2 is constant. If $p_2''(x) = 2$, then $p_2'(x) = 2x + C$

for some constant C. Since we have determined that $p_2'(0) = 1$, we find that $C = 1$ and so $p_2'(x) = 2x+1$. Finally, we can compute $p_2(x) = x^2+x+C$. Using our initial values, we know $p_2(0) = 2$ so $C = 2$. We conclude that $p_2(x) = x^2+x+2$. This function is plotted with f in Figure 9.14(b).

We can repeat this approximation process by creating polynomials of higher degree that match more of the derivatives of f at $x = 0$. In general, a polynomial of degree n can be created to match the first n derivatives of f. Figure 9.14(b) also shows $p_4(x) = -x^4/2-x^3/6+x^2+x+2$, whose first four derivatives at 0 match those of f.

As we use more and more derivatives, our polynomial approximation to f gets better and better. In this example, the interval on which the approximation is good gets bigger and bigger. Figure 9.14(c) shows $p_{13}(x)$; we can visually affirm that this polynomial approximates f very well on $[-2,3]$. Note, however, that the polynomial $p_{13}(x)$ is not particularly nice and that we determined the coefficients of the higher order terms with Mathematica. It is

$$p_{13}(x) = \frac{16901x^{13}}{6227020800}+\frac{13x^{12}}{1209600}-\frac{1321x^{11}}{39916800}-\frac{779x^{10}}{1814400}-\frac{359x^9}{362880}+ \frac{x^8}{240}+\frac{139x^7}{5040}+\frac{11x^6}{360}-\frac{19x^5}{120}-\frac{x^4}{2}-\frac{x^3}{6}+x^2+x+2.$$

The polynomials we have created are examples of **Taylor polynomials** (*Taylor-veelterm*), named after the British mathematician Brook Taylor who made important discoveries about such functions. It can be shown that Taylor polynomials follow a general pattern that make their formation much more direct. This is described in the following definition.

Definitie 9.9 (Taylor and MacLaurin polynomials)
Let f be a function whose first n derivatives exist at $x = x_0$.

1. The **Taylor polynomial of degree n of f at $x = x_0$** is

$$p_n(x) = f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\frac{f'''(x_0)}{3!}(x-x_0)^3+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n.$$

2. A special case of the Taylor polynomial is the Maclaurin polynomial, where $x_0 = 0$. That is, the **Maclaurin polynomial of degree n of f** is

$$p_n(x) = f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\cdots+\frac{f^{(n)}(0)}{n!}x^n.$$

9.8.2 Taylor's theorem

Taylor polynomials are used to approximate functions $f(x)$ in mainly two situations:

1. When $f(x)$ is known, but perhaps hard to compute directly. For instance, we can define $y = \cos(x)$ as either the ratio of sides of a right triangle or with the unit circle. However, neither of these provides a convenient way of computing $\cos(2)$. A Taylor polynomial of sufficiently high degree can provide a reasonable method of computing such values using only operations usually hard–wired into a computer ($+$, $-$, $\times$ and $\div$).

2. When $f(x)$ is not known, but information about its derivatives is known. This occurs more often than one might think, especially in the study of differential equations.

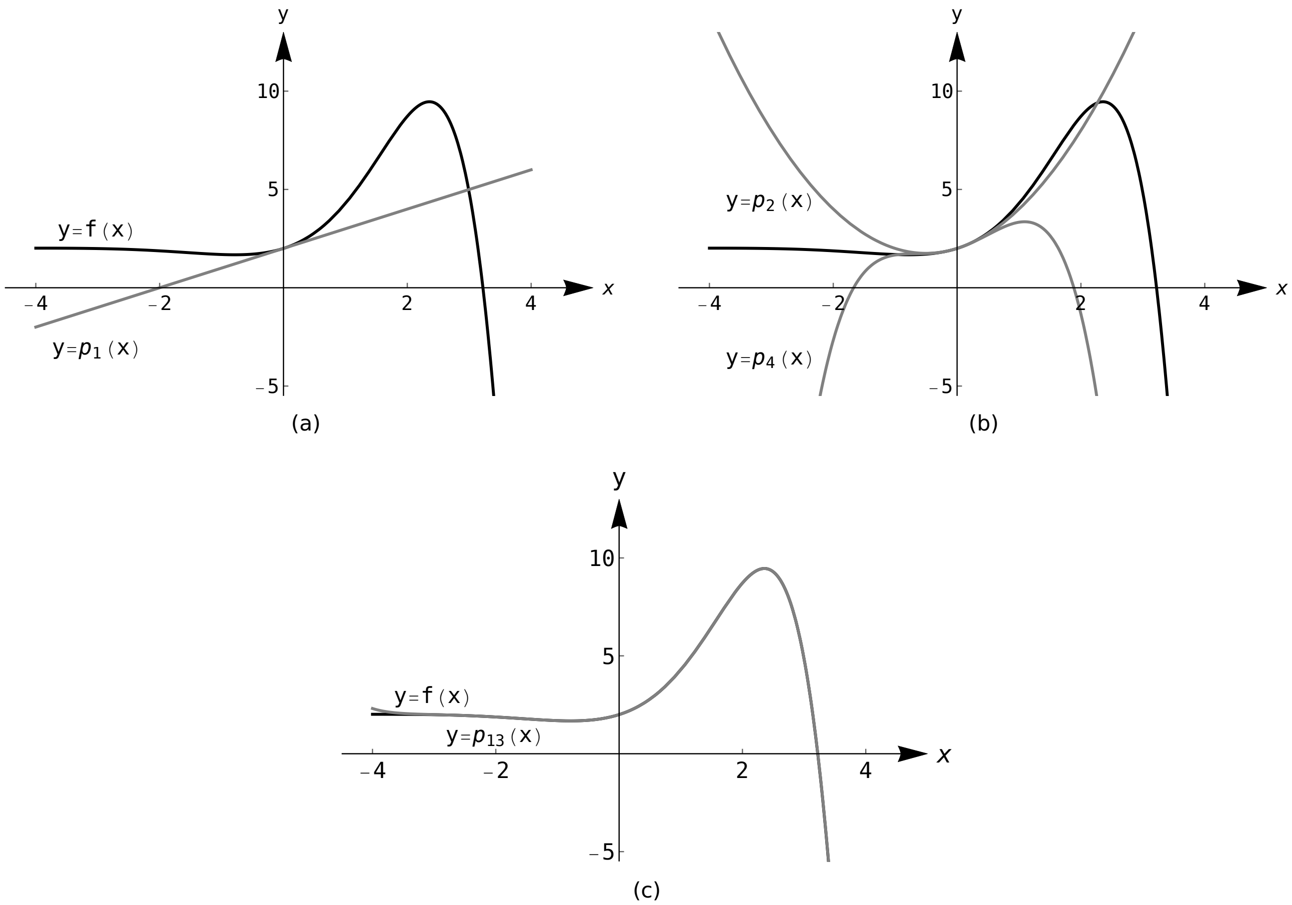


Figure 9.14: Plotting f and the tangent line at $x=0$ (a) , f, p_2 and p_4 (b), and f and p_{13} (c).

In both situations, it is crucial to know how good the approximation is. If we use a Taylor polynomial to compute $\cos(2)$, how do we know how accurate the approximation is?

The following theorem provides this kind of information for Taylor (and hence Maclaurin) polynomials.

Theorem 9.9 (Taylor's theorem)

1. *Let f be a function whose $(n+1)^{th}$ derivative exists on an interval I and let x_0 be in I. Then, for each x in I, there exists θ_x between x and x_0 such that*

$$\begin{aligned} f(x) &= \sum_{i=0}^{n}\frac{f^{(i)}(x_0)}{i!}(x-x_0)^i+\frac{f^{(n+1)}(\theta_x)}{(n+1)!}(x-x_0)^{n+1} \\ &= f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+R_n(x), \end{aligned}$$

where $R_n(x)=\dfrac{f^{(n+1)}(\theta_x)}{(n+1)!}(x-x_0)^{(n+1)}$ is the remainder term.

2. $\left|R_n(x)\right| \le \dfrac{\max\left|f^{(n+1)}(\theta)\right|}{(n+1)!}\left|(x-x_0)^{(n+1)}\right|$, *where θ is in I.*

Basically, the first part of Taylor's theorem states that $f(x)=p_n(x)+R_n(x)$, where $p_n(x)$ is the n^{th} order Taylor polynomial and $R_n(x)$ is the remainder, or error, in the Taylor approximation. The second part gives bounds on how big that error can be. If the $(n+1)^{\text{th}}$ derivative is large on I, the error may be large; if x is far from x_0, the error may also be large. However, the $(n+1)!$ term in the denominator tends to ensure that the error gets smaller as n increases.

We may also use Taylor's theorem to find n that guarantees our approximation is within a certain amount.

Example 9.29

Find n such that the n^{th} Taylor polynomial of $f(x) = \cos(x)$ at $x = 0$ approximates $\cos(2)$ to within 0.001 of the actual answer. What is $p_n(2)$?

Solution

Following Taylor's theorem, we need bounds on the size of the derivatives of $f(x) = \cos(x)$. In the case of this trigonometric function, this is easy. All derivatives of cosine are $\pm\sin(x)$ or $\pm\cos(x)$. In all cases, these functions are never greater than 1 in absolute value. We want the error to be less than 0.001. To find the appropriate n, consider the following inequalities:

$$\begin{aligned} \frac{\max\left|f^{(n+1)}(\theta)\right|}{(n+1)!}\left|(2-0)^{(n+1)}\right| &\le 0.001 \\ \frac{1}{(n+1)!}\cdot 2^{(n+1)} &\le 0.001 \end{aligned}$$

We find an n that satisfies this last inequality with trial–and–error. When $n = 8$, we have $\frac{2^{8+1}}{(8+1)!} \approx 0.0014$; when $n = 9$, we have $\frac{2^{9+1}}{(9+1)!} \approx 0.000282 < 0.001$. Thus we want to approximate $\cos(2)$ with $p_9(2)$.

We now set out to compute $p_9(x)$. We again need a table of the derivatives of $f(x) = \cos(x)$ evaluated at $x = 0$ (Table 9.2).

Table 9.2: The derivatives of $f(x) = \cos(x)$ evaluated at $x = 0$.

Derivative function	derivative at $x = 0$
$f(x) = \cos(x)$	$f(0) = 1$
$f'(x) = -\sin(x)$	$f'(0) = 0$
$f''(x) = -\cos(x)$	$f''(0) = -1$
$f'''(x) = \sin(x)$	$f'''(0) = 0$
$f^{(4)}(x) = \cos(x)$	$f^{(4)}(0) = 1$
$f^{(5)}(x) = -\sin(x)$	$f^{(5)}(0) = 0$
$f^{(6)}(x) = -\cos(x)$	$f^{(6)}(0) = -1$
$f^{(7)}(x) = \sin(x)$	$f^{(7)}(0) = 0$
$f^{(8)}(x) = \cos(x)$	$f^{(8)}(0) = 1$
$f^{(9)}(x) = -\sin(x)$	$f^{(9)}(0) = 0$

Notice how the derivatives, evaluated at $x = 0$, follow a certain pattern. All the odd powers of x in the Taylor polynomial will disappear as their coefficient is 0. While our error bounds state that we need $p_9(x)$, our work shows that this will be the same as $p_8(x)$.

Since we are forming our polynomial at $x = 0$, we are creating a Maclaurin polynomial, and:

$$\begin{aligned} p_8(x) &= f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(8)}(0)}{8!}x^8 \\ &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \frac{1}{8!}x^8 \end{aligned}$$

We finally approximate $\cos(2)$:

$$\cos(2) \approx p_8(2) = -\frac{131}{315} \approx -0.41587.$$

Our error bound guarantee that this approximation is within 0.001 of the correct answer. Technology shows us that our approximation is actually within about 0.0003 of the correct answer. Figure 9.15 shows a graph of $y = p_8(x)$ and $y = \cos(x)$. Note how well the two functions agree on about $]-\pi, \pi[$.

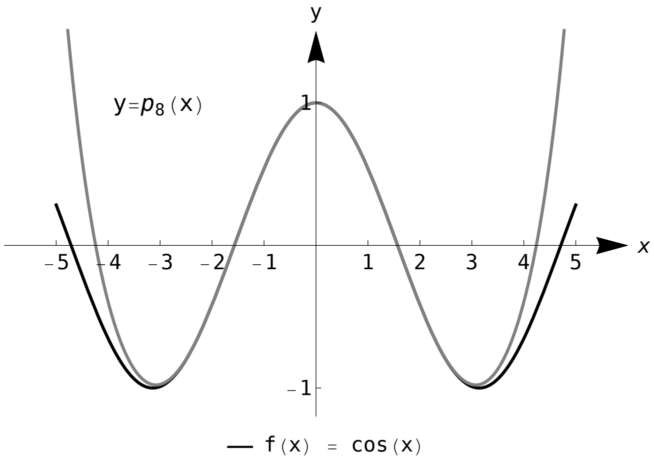


Figure 9.15: A graph of $f(x) = \cos(x)$ (black) and its degree 8 Maclaurin polynomial (gray).

9.8.3 Taylor series

In Section 9.8.1, we showed how we can approximate functions with polynomials, given that enough derivative information is available. In this section we combine these concepts: if a function $f(x)$ is infinitely differentiable, we show how to represent it with a power series function.

Definitie 9.10 (Taylor and Maclaurin series)
Let $f(x)$ have derivatives of all orders at $x = x_0$.

1. The **Taylor series of** $f(x)$ (*Taylor-reeks*), centred at x_0 is

$$\sum_{n=0}^{+\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

2. Setting $x_0 = 0$ gives the **Maclaurin series of** $f(x)$ (*Maclaurin-reeks*):

$$\sum_{n=0}^{+\infty} \frac{f^{(n)}(0)}{n!} x^n.$$

Note that the order of a Taylor series is determined by the highest-order derivative that appears in it. So the third-order Taylor series expansion of a function f contains terms up to those containing x^3. If $p_n(x)$ is the n^{th} degree Taylor polynomial for $f(x)$ centred at $x = x_0$, we saw how $f(x)$ is approximately equal to $p_n(x)$ near $x = x_0$. We also saw how increasing the degree of the polynomial generally reduced the error. We are now considering series, where we sum an infinite set of terms. Our ultimate hope is to see the error vanish and claim a function is equal to its Taylor series. The attentive reader now probably wonders whether the sums in Definition 9.10 will not always give rise to $\pm\infty$, because one intuitively would expect this. It can, however, be shown that the result will be a real number for the functions that we consider throughout this course.

When creating the Taylor polynomial of degree n for a function $f(x)$ at $x = x_0$, we needed to evaluate f, and the first n derivatives of f, at $x = x_0$. When creating the Taylor series of f, it helps to find a pattern that describes the n^{th} derivative of f at $x = x_0$. We demonstrate this in the next example.

Example 9.30

Find the Taylor series of $f(x) = \ln(x)$ centred at $x = 1$.

Solution

Table 9.3 shows the n^{th} derivative of $\ln(x)$ evaluated at $x = 1$ for $n = 0, \ldots, 5$, along with an expression for the n^{th} term:

$$f^{(n)}(1) = (-1)^{n+1}(n-1)! \quad \text{for } n \geq 1.$$

Remember that this is what distinguishes Taylor series from Taylor polynomials; we are very interested in finding a pattern for the n^{th} term, not just finding a finite set of coefficients for a polynomial.

Table 9.3: The derivatives of $\ln(x)$ evaluated at $x = 1$.

Derivative function	derivative at $x = 1$
$f(x) = \ln(x)$	$f(1) = 0$
$f'(x) = 1/x$	$f'(1) = 1$
$f''(x) = -1/x^2$	$f''(1) = -1$
$f'''(x) = 2/x^3$	$f'''(1) = 2$
$f^{(4)}(x) = -6/x^4$	$f^{(4)}(1) = -6$
$f^{(5)}(x) = 24/x^5$	$f^{(5)}(1) = 24$
$\vdots$	$\vdots$
$f^{(n)}(x) = \dfrac{(-1)^{n+1}(n-1)!}{x^n}$	$f^{(n)}(1) = (-1)^{n+1}(n-1)!$

Since $f(1) = \ln(1) = 0$, we skip the first term and start the summation with $n = 1$, giving the Taylor series for $\ln(x)$, centred at $x = 1$, as

$$\sum_{n=1}^{+\infty}(-1)^{n+1}(n-1)!\frac{1}{n!}(x-1)^n = \sum_{n=1}^{+\infty}(-1)^{n+1}\frac{(x-1)^n}{n}.$$

Note that we cannot (yet) say that $\ln(x)$ is equal to this Taylor series on $]0,2]$.

Also in Mathematica it is possible to determine the **series expansion** (*reeks-ontwikkeling*) of a function. For instance, to get the Taylor series of $f(x) = \ln(x)$ centred at $x = 1$, we can proceed as follows with the command **Series**.

```
In[14]:= Series[Log[x],{x,1,5}]

Out[14]= (x-1)-1/2 (x-1)^2+1/3 (x-1)^3-1/4 (x-1)^4+1/5 (x-1)^5+O[x-1]^6
```

The general syntax of this command is

```
In[15]:= Series[f[x],{x,x0,n},]
```

where `f[x]` is the function at stake, `x` the variable, `x0` the point at which the series is centred and `n` the order.

It is important to note that Definition 9.10 defines a Taylor series given a function $f(x)$, but makes no claim about their equality. We will find that most of the time they are equal, but we need to consider the conditions that allow us to conclude this.

Theorem 9.9 states that the error between a function $f(x)$ and its n^{th}–degree Taylor polynomial $p_n(x)$ is $R_n(x)$, where

$$\left|R_n(x)\right| \leq \frac{\max\left|f^{(n+1)}(\theta)\right|}{(n+1)!}\left|(x-x_0)^{(n+1)}\right|.$$

If $R_n(x)$ goes to 0 for each x in an interval I as n approaches infinity, we conclude that the function is equal to its Taylor series expansion. This formalized in the following theorem.

Theorem 9.10 (Function and Taylor series equality)
Let $f(x)$ have derivatives of all orders at $x = x_0$ i.e. $f(x)$ is a smooth function, let $R_n(x)$ be as stated in Theorem 9.9, and let I be an interval. If $\lim_{n\to+\infty} R_n(x) = 0$ for all x in I, then

$$f(x) = \sum_{n=0}^{+\infty}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n \text{ on } I.$$

It is natural to assume that a function is equal to its Taylor series, but this is not always the case. In order to properly establish equality, one must use Theorem 9.10. This is a bit disappointing, as proving that $R_n(x) \to 0$ can be difficult. For instance, it is not a simple task to show that $\ln(x)$ equals its Taylor series on $]0,2]$ as found in Example 9.30.

Fourier series

A Fourier series is another kind of series that is used to represent a function as the sum of simple sine waves. Essentially, it decomposes any periodic function into the sum of a (possibly infinite) set of simple oscillating functions, namely sines and cosines. The Fourier series has many applications in electrical engineering, vibration analysis, acoustics, optics, signal processing, image processing, quantum mechanics, econometrics, and so on. Figure 9.16 shows the Fourier series

Fourier series

approximation of a square wave using 5 and 15 terms.

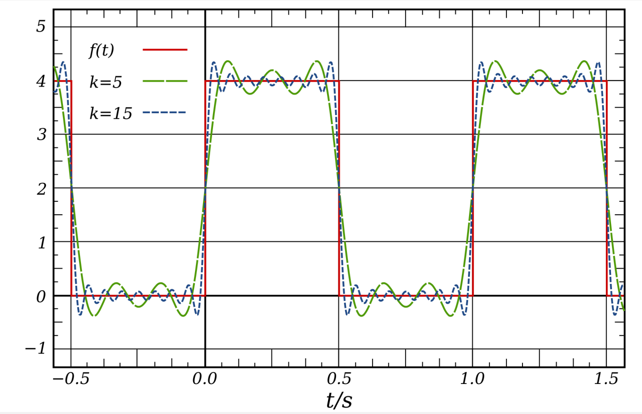


Figure 9.16: Fourier series approximation of a square wave using 5 and 15 terms.

A function $f(x)$ that is equal to its Taylor series, centered at any point of the domain of $f(x)$, is said to be an **analytic function** (*analytische functie*), and most functions that we will encounter are analytic functions. Generally speaking, any function that one creates with elementary functions (polynomials, exponentials, trigonometric functions, etc.) that is not piecewise defined is probably analytic. For most functions, we may assume the function is equal to its Taylor series and only use Theorem 9.10 when we suspect something may not work as expected.

We develop the Taylor series for one more important function, then give a table of the Taylor series for a number of common functions.

Example 9.31

Find the Maclaurin series of $f(x) = (1+x)^k$ with $k \neq 0$.

Solution

When k is a positive integer, the Maclaurin series is finite. For instance, when $k = 4$, we have

$$f(x) = (1+x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4.$$

The coefficients of x when k is a positive integer are known as the binomial coefficients, giving the series we are developing its name. When $k = 1/2$, we have $f(x) = \sqrt{1+x}$. Knowing a series representation of this function would give a useful way of approximating $\sqrt{1.3}$, for instance. To develop the Maclaurin series for $f(x) = (1+x)^k$ for any value of $k \neq 0$, we consider the derivatives of f evaluated at $x = 0$ as in Table 9.4

Table 9.4: The derivatives of $f(x) = (1+x)^k$ evaluated at $x = 0$.

Derivative function	derivative at $x = 0$
$f(x) = (1+x)^k$	$f(0) = 1$
$f'(x) = k(1+x)^{k-1}$	$f'(0) = k$
$f''(x) = k(k-1)(1+x)^{k-2}$	$f''(0) = k(k-1)$
$f'''(x) = k(k-1)(k-2)(1+x)^{k-3}$	$f'''(0) = k(k-1)(k-2)$
$\vdots$	$\vdots$
$f^{(n)}(x) = k(k-1)\cdots\big(k-(n-1)\big)(1+x)^{k-n}$	$f^{(n)}(0) = k(k-1)\cdots\big(k-(n-1)\big)$

Thus the Maclaurin series for $f(x) = (1+x)^k$ is

$$1 + kx + \frac{k(k-1)}{2!}x^2 + \frac{k(k-1)(k-2)}{3!}x^3 + \ldots + \frac{k(k-1)\cdots\big(k-(n-1)\big)}{n!}x^n + \ldots .$$

We learned that Taylor polynomials offer a way of approximating a difficult to compute function with a polynomial. Taylor series offer a way of exactly representing a function with a series. One probably can see the use of a good approximation; is there any use of representing a function exactly as a series? Yes, amongst other things they provide a valuable tool for solving a variety of problems, including problems relating to integration and differential equations.

In Table 9.5 we give the Taylor series of a number of common functions.

We also give a theorem about the algebra of power series, that is, how we can combine power series to create power series of new functions. This allows us to find the Taylor series of functions like $f(x) = e^x\cos(x)$ by knowing the Taylor series of e^x and $\cos(x)$.

Theorem 9.11 (Algebra of power series)

Let $f(x) = \sum_{n=0}^{+\infty} a_n x^n$ *and* $g(x) = \sum_{n=0}^{+\infty} b_n x^n$, *and let* $h(x)$ *be continuous.*

1. $f(x) \pm g(x) = \sum_{n=0}^{+\infty} (a_n \pm b_n)x^n$.
2. $f(x)g(x) = \left(\sum_{n=0}^{+\infty} a_n x^n\right)\left(\sum_{n=0}^{+\infty} b_n x^n\right) = \sum_{n=0}^{+\infty} \big(a_0 b_n + a_1 b_{n-1} + \ldots a_n b_0\big)x^n$.
3. $f\big(h(x)\big) = \sum_{n=0}^{+\infty} a_n \big(h(x)\big)^n$.

One can also apply calculus techniques to Taylor series; in particular, we can find derivatives and antiderivatives.

Theorem 9.12 (Derivatives and indefinite integrals of Taylor series)

Let $f(x) = \sum_{n=0}^{+\infty} a_n (x-x_0)^n$ *be a function defined by a power series. Then the following hold:*

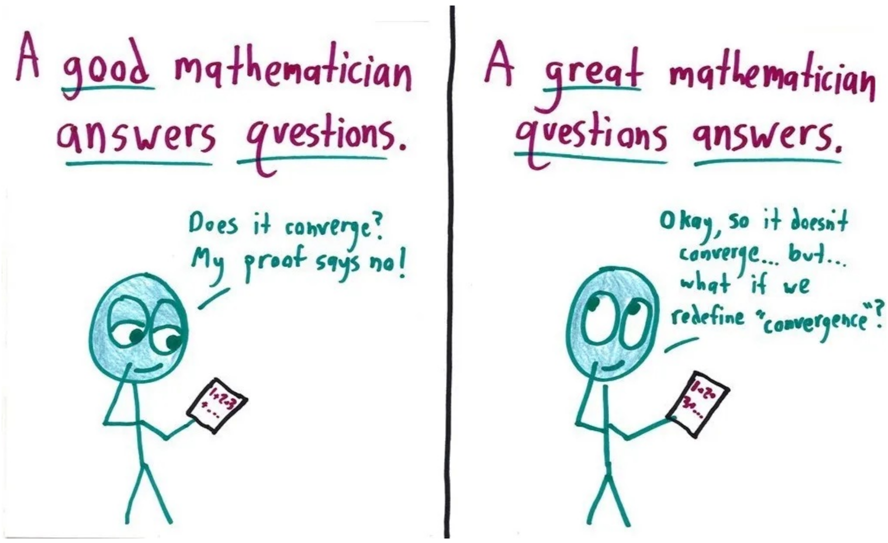
Table 9.5: Important Taylor series expansions.

Function and Series	First few terms
$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$	$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$
$\sin(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$	$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$
$\cos(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n}}{(2n)!}$	$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$
$\ln(x) = \sum_{n=1}^{+\infty} (-1)^{n+1} \frac{(x-1)^n}{n}$	$(x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots$
$\frac{1}{1-x} = \sum_{n=0}^{+\infty} x^n$	$1 + x + x^2 + x^3 + \cdots$
$(1+x)^k = \sum_{n=0}^{+\infty} \frac{k(k-1)\cdots(k-(n-1))}{n!} x^n$	$1 + kx + \frac{k(k-1)}{2!} x^2 + \cdots$
$\arctan(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$	$x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots$

1. $f'(x) = \sum_{n=1}^{+\infty} a_n \cdot n \cdot (x - x_0)^{n-1}$.

2. $\int f(x)\, dx = x_0 + \sum_{n=0}^{+\infty} a_n \frac{(x-x_0)^{n+1}}{n+1}$.

Note that differentiation and integration are simply calculated term–by–term using the power rules.



9.9 Exercises

9.9.1 Analytical exercises

Definition

Assignment 9.1 — Draw the graph of the function f and determine the continuity and differentiability of f in the given point.

(a) $f(x) = |x|, \qquad x = 0$

(b) $f(x) = |x^2 - 1|, \qquad x = 1$

(c) $f(x) = |\sin(x)|, \qquad x = 0$

(d) $f(x) = \sqrt{x}, \qquad x = 0$

(e) $f(x) = \sqrt{1 - x^2}, \qquad x = -1$

Assignment 9.2 — Determine the equation of the tangent and normal to graphs of the functions below at the given point.

(a) $f(x) = x^3 - 3x^2 + 2, \qquad x = 0$

(b) $f(x) = \dfrac{x}{2+x}, \qquad x = -1$

(c) $f(x) = \sin(x), \qquad x = \dfrac{\pi}{4}$

(d) $f(x) = \tan(x), \qquad x = \dfrac{\pi}{4}$

(e) $f(x) = e^x\left(x^2 + 2\right), \qquad x = 0$

(f) $f(x) = \dfrac{x^2}{x-1}, \qquad x = 2$

Assignment 9.3 — Determine the equation(s) of the tangent line(s) at the point with abscissa $x = 0$ to the graph of

(a) $f(x) = |\sin(x)|$

(b) $f(x) = x^{2/3}$

(c) $f(x) = x^{4/3}$

(d) $f(x) = x^3$

Assignment 9.4 — There exist two non-coinciding, intersecting lines that go through $(1, -3)$ and are tangent to $y = x^2$. Determine their equations.

The chain rule

Assignment 9.5 — Find the first derivative of the functions below.

(a) $f(x) = 2x^2 + 3x + 4$

(b) $f(x) = \dfrac{2}{x^2}$

(c) $f(x) = \dfrac{2x^2 - x}{3x+1}$

(d) $f(x) = x^5 - 4x^4 + 3x^2 - 6x + 1$

(e) $f(x) = x\sqrt{x^2-1}$

(f) $f(x) = \dfrac{3x+2}{2x-1}$

(g) $f(x) = \sqrt{\dfrac{x-1}{x+1}}$

(h) $f(x) = \dfrac{1}{(7x^2+3x-6)^2}$

(i) $f(x) = \sqrt[3]{(2x+3)^2}$

(j) $f(x) = x\sqrt[3]{x+1}$

(k) $f(x) = \dfrac{x}{2x + \dfrac{1}{3x+1}}$

(l) $f(x) = \left(1 + \sqrt{\dfrac{x-2}{3}}\right)^4$

Assignment 9.6 — Find the first derivative of the functions below.

(a) $f(x) = \cos(5x)$

(b) $f(x) = \dfrac{\cos(x)}{\sin(x)}$

(c) $f(x) = x\sin^2(x)$

(d) $f(x) = x^3\cos(x)$

(e) $f(x) = e^{-x}$

(f) $f(x) = x^5\left(\sec(x) + e^x\right)$

(g) $f(x) = \dfrac{e^{-2x}}{x^2}$

(h) $f(x) = \ln\left(\left(\dfrac{e^x+1}{e^x-1}\right)^{1/2}\right)$

(i) $f(x) = e^{-\arcsin(x)}$

(j) $f(x) = \tan(e^x)$

(k) $f(x) = \ln\left(\sin(2x) + \sin^2(x)\right)$

(l) $f(x) = \log_{10}\left(\sqrt{9-x^2}\right)$

(m) $f(x) = \log_2\left((x^2+x+2)^4\right)$

(n) $f(x) = \log_3\left(\sqrt{\dfrac{2x-1}{2x+1}}\right)$

(o) $f(x) = \arctan\left(x^3\right)$

(p) $f(x) = \arccos\left(\dfrac{1-x}{1+x}\right)$

(q) $f(x) = \arcsin\left(\dfrac{x}{x-1}\right)$

(r) $f(x) = \dfrac{\sin\left(\sqrt{x}\right)}{1+\cos\left(\sqrt{x}\right)}$

(s) $f(x) = \arccos\left(\dfrac{x-b}{a}\right)$

(t) $f(x) = \left(\arcsin(x^2)\right)^{1/2}$

(u) $f(x) = \sqrt{a^2-x^2} + a\arcsin\left(\dfrac{x}{a}\right)$

Assignment 9.7 — Find the first derivative of the functions below, then simplify the result.

(a) $f(x) = \ln\left|\dfrac{1}{\cos(x)} + \tan(x)\right|$

(b) $f(x) = \dfrac{1}{2}\dfrac{\tan(x)}{\cos(x)} + \dfrac{1}{2}\ln\left|\dfrac{1}{\cos(x)} + \tan(x)\right|$

(c) $f(x) = -\dfrac{\sqrt{x^2-a^2}}{x} + \ln\left|x+\sqrt{x^2-a^2}\right|$

(d) $f(x) = \dfrac{x}{8}\left(5-2x^2\right)\sqrt{1-x^2} + \dfrac{3}{8}\arcsin(x)$

(e) $f(x) = \ln\left|\dfrac{\sqrt{x+1}-1}{\sqrt{x+1}+1}\right|$

(f) $f(x) = \dfrac{1}{a^2}\left(\ln|ax+b| + \dfrac{b}{ax+b}\right)$

(g) $f(x) = x\left(2x^2-a^2\right)\sqrt{a^2-x^2} + a^4\arcsin\left(\dfrac{x}{a}\right)$

(h) $f(x) = \dfrac{(x+1)\left(9-2x-x^2\right)}{4}\sqrt{3-2x-x^2} + 6\arcsin\left(\dfrac{x+1}{2}\right)$

Implicit differentiation

Assignment 9.8 — Consider the implicitly defined functions below. Define y' as a function of x and y.

(a) $x^2e^2 + 2^y = 5$

(b) $\left(3x^2+2y^3\right)^4 = 2$

(c) $xy - x + 2y = 1$

(d) $x^3y + xy^5 = 2$

(e) $x^2 + 4(y-1)^2 = 4$

(f) $\dfrac{x-y}{x+y} = \dfrac{x^2}{y} + 1$

(g) $\dfrac{\sin(x)+y}{\cos(y)+x} = 1$

(h) $\ln\left(x^2+xy+y^2\right) = 1$

Assignment 9.9 — Determine an equation of the tangent to the given curve at the given point.

(a) $\dfrac{x}{y} + \left(\dfrac{y}{x}\right)^3 = 2, \qquad (-1,-1)$

(b) $x\sin(xy-y^2) = x^2-1, \qquad (1,1)$

(c) $\left(x^2+y^2+x\right)^2 = x^2+y^2, \qquad (0,1)$

Assignment 9.10 — We consider $x^2+4y^2=4$. Determine y'' as a function of x and y.

Assignment 9.11 — Determine y' for the functions below.

(a) $y = \dfrac{x^x}{x+1}$

(b) $y = x^{\sin(x)+2}$

(c) $y = \left(\sin(x)\right)^{\ln(x)}$ with $x \in]0,\pi[$

(d) $y = \left(\dfrac{1}{x}\right)^{\ln(x)}$

(e) $y = (\cos(x))^x - x^{\cos(x)}$

(f) $y = \dfrac{x^{\ln(x)}(\sin(x))^x}{x^x\ln(x)}$

Derivatives of inverse functions

Assignment 9.12 — We consider an injective function $f(x)$ for which it holds that $f'(x) = \dfrac{1}{1+x}$. Determine $(f^{-1})'(x)$.

Assignment 9.13 — Consider the functions below and find the required derivative.

(a) $f(x) = 1 + 2x^3, \qquad (f^{-1})'(x)$

(b) $f(x) = x\sqrt{3+x^2}, \qquad (f^{-1})'(-2)$

(c) $f(x) = \dfrac{4x^3}{x^2+1}, \qquad (f^{-1})'(2)$

(d) $f(x) = x^3 + x, \qquad (f^{-1})'(10)$

(e) $f(x) = \sin(2x), \qquad (f^{-1})'(\sqrt{3}/2)$

(f) $f(x) = 6e^{3x}, \qquad (f^{-1})'(6)$

L'Hôpital's rule

Assignment 9.14 — Find the limits below.

(a) $\displaystyle\lim_{x\to 0} \frac{2x - \sin(x)}{x^3}$

(b) $\displaystyle\lim_{x\to 1} \frac{3\ln(x)}{x^5 - 1}$

(c) $\displaystyle\lim_{x\to +\infty} \frac{x^n}{e^x}$

(d) $\displaystyle\lim_{x\to 0} \frac{\tan(x) - \sin(x)}{x^3}$

(e) $\displaystyle\lim_{x\to +\infty} \left(x\tan\left(\frac{1}{x}\right)\right)$

(f) $\displaystyle\lim_{x\to 0} \left(\frac{1}{x} - \frac{1}{\sin(x)}\right)$

(g) $\displaystyle\lim_{x\to 1} \left(\frac{1}{\ln(x)} - \frac{x}{x-1}\right)$

(h) $\displaystyle\lim_{x\to 0} \frac{\arctan(2x) - 2\arcsin(x)}{x^2\sin(x)}$

(i) $\displaystyle\lim_{x \underset{<}{\to} \frac{\pi}{2}} (\cos(x))^{\frac{\pi}{2} - x}$

(j) $\displaystyle\lim_{x \underset{>}{\to} 0} x^{\frac{3}{x+\ln(x)}}$

(k) $\displaystyle\lim_{x \underset{<}{\to} \frac{\pi}{2}} (\tan(x))^{\cos(x)}$

(l) $\displaystyle\lim_{x \underset{>}{\to} 0} (\csc(x))^{\sin^2(x)}$

(m) $\displaystyle\lim_{x\to +\infty} \left(\cos\left(\sqrt{\frac{3}{x}}\right)\right)^x$

(n) $\displaystyle\lim_{x\to 0} (\cos(x))^{1/x^2}$

(o) $\displaystyle\lim_{x\to 0} x\cot(x)$

Applications of the derivative

Assignment 9.15 — A cube has ribs of length 20 cm. By how much must the length of the ribs decrease so that the volume of the cube decreases with 12 cm^3?

Assignment 9.16 — A spherical balloon is inflated causing the radius to increase in one minute from 20 cm tot 20.2 cm. By how much is the volume going to increase in one minute?

Taylor polynomials

Assignment 9.17 — Determine the requested series expansion for the functions below.

(a) $f(x)=\dfrac{x^3}{1-2x^2}$ in powers of x

(b) $f(x)=\dfrac{1-x}{1+x}$ in powers of x

(c) $f(x)=\dfrac{1}{(2-x)^2}$ in powers of x

(d) $f(x)=\ln(2-x)$ in powers of x

(e) $f(x)=\ln(x)$ in powers of $x-4$

Assignment 9.18 — Determine the MacLaurin series expansion for the functions below.

(a) $f(x)=x^2\sin\left(\dfrac{x}{3}\right)$

(b) $f(x)=\cos^2\left(\dfrac{x}{2}\right)$

(c) $f(x)=\dfrac{e^{2x^2}-1}{x^2}$

(d) $f(x)=(1+x)^{\frac{1}{2}}\cos(x)$

Assignment 9.19 — Determine the Taylor series expansion for the functions below.

(a) $f(x)=\dfrac{1}{x}$ at $x=1$

(b) $f(x)=\dfrac{1}{x^2}$ using powers of $x+2$

(c) $f(x)=\sin(x)-\cos(x)$ at $x=\dfrac{\pi}{4}$

(d) $f(x)=x\ln(x)$ using powers of $x-1$

(e) $f(x)=xe^x$ using powers of $x+2$

(f) $f(x)=\ln(2+x)$ using powers of $x-2$

(g) $f(x)=\cos^2(x)$ at $x=\dfrac{\pi}{8}$

9.9.2 Root finding algorithms

Two methods for finding roots numerically were discussed: the **Bisection method** (*halveringsmethode*) (Section 8.5.2) and **Newton's method** (*methode van Newton*) (Section 9.7.1).

Here, we will implement both methods in Python. For those who have little or no programming experience, a Python Tutorial is provided in Appendix C. Both the code below and the Python Tutorial are available as Jupyter Notebooks.

9.9.2.1 Bisection method

Below is an implementation of the bisection method in Python. First, see how this method translates to executable Python code, then answer the questions below.

```
def bisection(f, interval, eps=10**-6, max_it=100):
        '''
         Bisection method for approximating the root
        of the function f on a given interval [a,b],
        where f(a) and f(b) have a different sign
        Inputs:
```

```
    - f: function whose roots should be found
    - interval: interval [a,b]
    - eps: maximum approximation error (default: 10^-6)
    - max_it: maximum number of iterations

    Output:
    - root: approximated root
    '''
    print("Bisection method")
    print("-----------------")
    # extract the values of a and b from the interval and calculate the corresponding
function values
    a = interval[0]
    b = interval[1]
    f_a = f(a)
    f_b = f(b)

    # display error message if the sign of a and b does not differ
    if np.sign(f_a)==np.sign(f_b):
            print("The sign of f(a) and f(b) does not differ!\n")
            return None

    # determine the midpoint of the interval
    m = (a+b)/2
    f_m = f(m)

    # initialize the iterator
    it = 0

    while abs(f_m)>=eps:
            # evaluate sign of m
            if np.sign(f_a)==np.sign(f_m):
                    a=m
                    f_a=f_m
            else:
                    b=m
                    f_b = f_m
                    m = (a+b)/2
                    f_m = f(m)

            # update the iterator
            it=it+1
            #stop when the maximum number of iterations is reached
            if it==max_it:
                    print("Maximum number of iterations reached!")
                    break
    root = m
    print("Approximated root {}\nwas reached after {} iterations\n".format(root,it))
    print("================================================")
    return root
```

Question 1.a How does the method in the implementation above differ from that of Example 8.13?

Question 1.b Use the function **bisection** to approximate the root(s) of the following functions with a maximum approximation error of $\epsilon = 10^{-6}$.

- $g_1(x) = \sin(x)$ on the interval $\left[\frac{3\pi}{4}, \frac{3\pi}{2}\right]$
- $g_2(x) = 2x^2 - 2$ on the interval $\left[-\frac{3}{2}, 0\right]$
- $g_3(x) = 2x^2 + 2$ on the interval $[-5, 5]$
- $g_4(x) = \frac{x^5}{2} - 3x^4 + 5x^3 + 6x^2 - 9x - 5$ on the interval $[-1, 1]$

These functions are available in the file `teachingtools`, from which they can be imported as follows.

```
from teachingtools import g_1, g_2, g_3
```

This way, for g_1 we obtain:

```
>>> bisection(g_1, [3*np.pi/4, 3*np.pi/2], eps=10**-6, max_it=100);
        Bisection method
        -----------------
        Approximated root 3.1415919045757357
        was reached after 19 iterations

        ==================================================
```

9.9.2.2 Newton's method

A faster alternative to the bisection method is Newton's method, which was discussed in Section 9.7.1.

Question 2.a Implement Newton's method by completing the code below where you find "...". You can find all functions and techniques necessary for this implementation in the Python Tutorial.

```
def Newton(f, df, x0, eps= 10**-6, max_it=100):
    '''
    Newton's method for the approximation of the root
   of the function f, starting from an initial estimate x0
    Inputs:
    - f: function whose root should be found
    - df: derivative of function whose root should be found
    - x0: initial approximation of the root
    - eps: maximum approximation error (default: 10^-6)
    - maxI: maximum number of iterations (default: 100)

    Output:
    - root: approximated root
    '''
    print("Newton's method")
    print("-------------------------")
    it = 0
    x = x0

    while ...: # to be completed
          ...
          ...
```

```
        if ...:  #stop when maximum number of iterations is reached
                 # to be completed

           print("Maximum number of iterations reached before convergence criterion
            was satisfied!")
           print("Verify if there is a root.")
           return None
    root = x
    print("Approximated root {}\nwas reached after {} iterations\n".format(root,it))
    print("================================================")
    return root
```

Question 2.b Find the derivative of the functions from Question 1 and implement them as the functions dg_i (for i going from 1 to 4).

```
def dg_1(x):
        return ...
def dg_2(x):
        return ...
def dg_3(x):
        return ...
```

Question 2.c Use the function **`Newton`** to find the root of the functions g_1 through g_4 from Question 1 (with a maximum approximation error of $\epsilon = 10^{-6}$). Choose a value at the border of the specified intervals as starting point $x0$ and compare your result with that of the bisection method. What stands out?

Question 3.a Consider the functions

$$k(x) = -\frac{5(-1+2x-5x^2+2x^5)}{3(4-3x^2+5x^3+x^4)} \quad \text{and} \quad l(x) = 2-\frac{e^x}{20}.$$

These functions are available in the file `teachingtools`, from where they can be imported.

QUestion 3.b Find the intersection point(s) of the graphs of $k(x)$ and $l(x)$ for $x \in [-0.8, 8]$, if any exist. To that end, first find the derivatives of $k(x)$ and $l(x)$ and implement them as $dk(x)$ and $dl(x)$.

Nothing takes place in the world whose meaning is not that of some maximum or minimum.

— Leonhard Euler —

10

The graphical behaviour of functions

Our study of limits led to continuous functions, a certain class of functions that behave in a particularly nice way. Limits then gave us an even nicer class of functions, functions that are differentiable.

This chapter explores many of the ways we can take advantage of the information that continuous and differentiable functions provide.

10.1 Extreme values

Given any quantity described by a function, we are often interested in the largest and/or smallest values that quantity attains. For instance, if a function describes the speed of an object, it seems reasonable to want to know the fastest/slowest the object traveled. If a function describes the value of a stock, we might want to know the highest/lowest values the stock attained over the past year. We call such values **extreme values** (*extrema*).

Definitie 10.1 (Extreme values)
Let f be defined on an interval I containing c.

1. $f(c)$ is the **minimum** (also, absolute minimum) of f on I if $f(c) \leq f(x)$ for all x in I.
2. $f(c)$ is the **maximum** (also, absolute maximum) of f on I if $f(c) \geq f(x)$ for all x in I.

The maximum and minimum values are the extreme values, or extrema, of f on I.

We can also define relative minima and maxima, which may be understood as the smallest and largest y-value nearby, respectively. We can make this intuitive understanding more formal as follows.

Definitie 10.2 (Relative minimum and relative maximum)
Let f be defined on an interval I containing c.

1. If there is a $\delta > 0$ such that $f(c) \leq f(x)$ for all x in I where $|x-c| < \delta$, then $f(c)$ is a **relative minimum** (*lokaal minimum*) of f. We also say that f has a relative minimum at $(c, f(c))$.
2. If there is a $\delta > 0$ such that $f(c) \geq f(x)$ for all x in I where $|x-c| < \delta$, then $f(c)$ is a **relative maximum** (*lokaal maximum*) of f. We also say that f has a relative maximum at $(c, f(c))$.

The relative maximum and minimum values comprise the **relative extrema** (*lokaal extremum*) of f.

The function displayed in Figure 10.1(a) has a maximum, but no minimum, as the interval over which the function is defined is open. In Figure 10.1(b), the function has a minimum, but no maximum; there is a discontinuity in the natural place for the maximum to occur. Finally, the function shown in Figure 10.1(c) has both a maximum and a minimum; note that the function is continuous and the interval on which it is defined is closed.

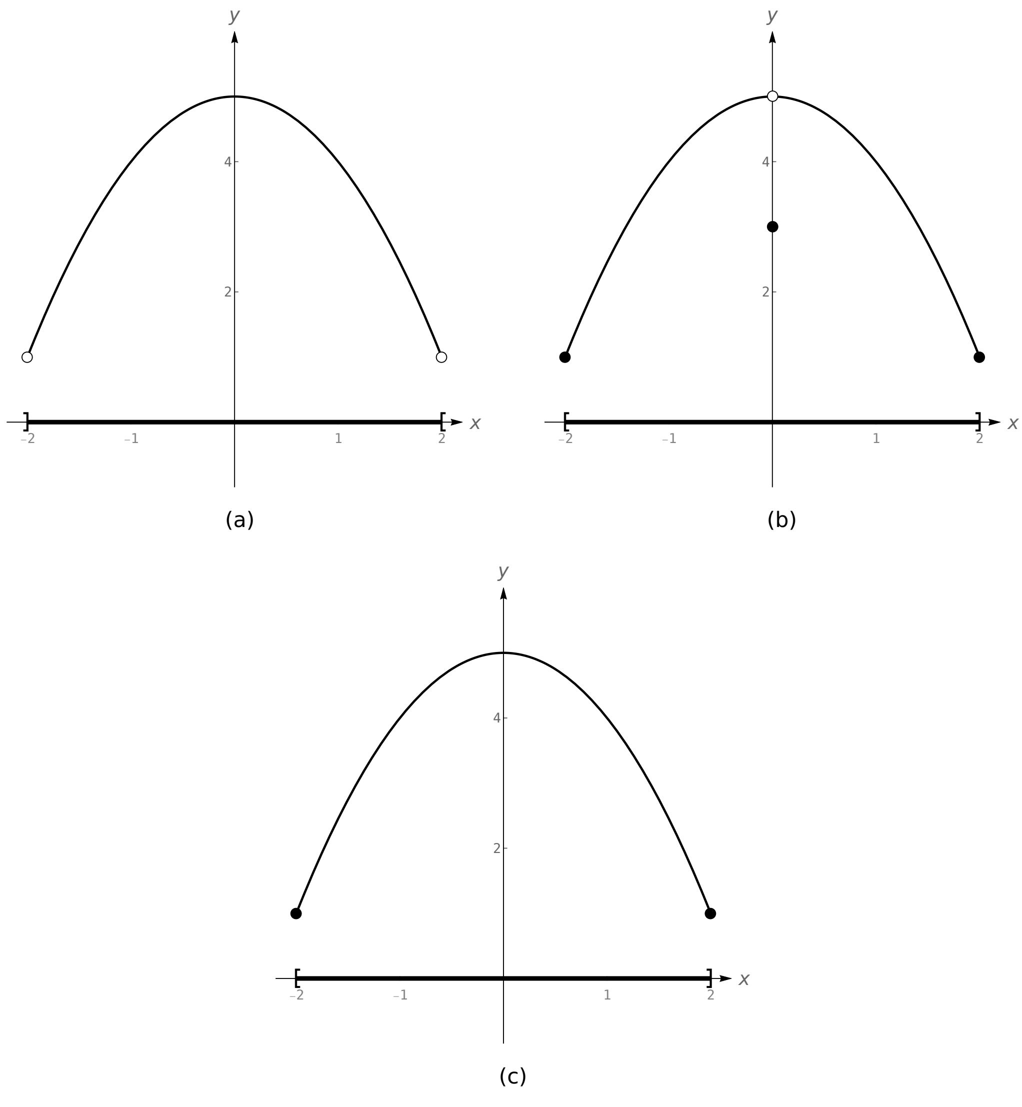


Figure 10.1: Graphs of functions with and without extreme values.

It is possible for discontinuous functions defined on an open interval to have both a maximum and minimum value, but we have just seen examples where they did not. On the other hand, continuous functions on a closed interval always have a maximum and minimum value. This is formalized in the following theorem.

Theorem 10.1 (The extreme value theorem)
Let f be a continuous function defined on a closed interval I. Then f has both a maximum and minimum value on I.

This theorem states that f has extreme values, but it does not offer any advice about how/where to find these values. The process can seem to be fairly easy, as the next example illustrates. After the example, we will draw on lessons learned to form a more general and powerful method for finding extreme values.

Example 10.1

Consider the functions

1. $f(x) = \dfrac{3x^4 - 4x^3 - 12x^2 + 5}{5}$,
2. $g(x) = (x-1)^{2/3} + 2$,

as shown in Figure 10.2(a) and 10.2(b), respectively. Approximate the relative extrema of these functions. At each of these points, evaluate the corresponding first derivative.

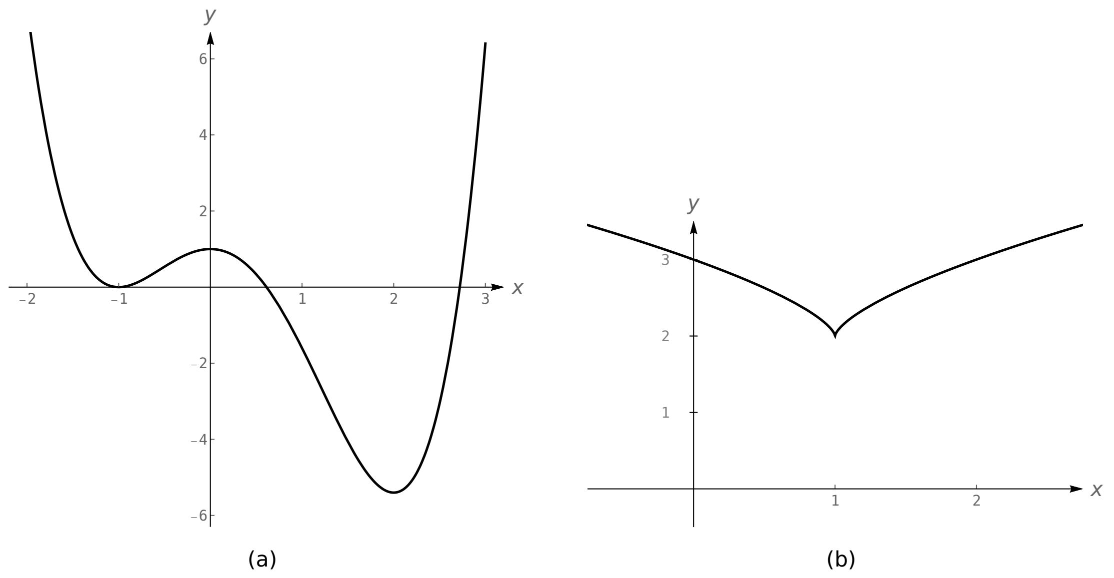


Figure 10.2: A graph of $f(x) = (3x^4 - 4x^3 - 12x^2 + 5)/5$ (a) and $g(x) = (x-1)^{2/3} + 2$ (b).

Solution

We do not yet have the tools to exactly find the relative extrema, but the graphs do allow us to make reasonable approximations.

1. It seems f has relative minima at $x = -1$ and $x = 2$, with values of $f(-1) = 0$ and $f(2) = -5.4$. It also seems that f has a relative maximum at the point $(0, 1)$. We approximate the relative minima to be 0 and -5.4; we approximate the relative maximum to be 1. It is straightforward to evaluate $f'(x) = (12x^3 - 12x^2 - 24x)/5$ at $x = 0, 1$ and 2. In each case, $f'(x) = 0$.

2. Figure 10.2(b) implies that g does not have any relative maxima, but has a relative minimum at $(1,2)$. The graph suggests that not only is this point a relative minimum, $y = g(1) = 2$ is the absolute minimum value of the function. We compute $g'(x) = 2/3(x-1)^{-1/3}$ note that when $x = 1$, g' is undefined.

What can we learn from the previous two examples? We were able to visually approximate relative extrema, and at each such point, the derivative was either 0 or it was not defined. This observation holds for all functions, leading to a definition and a theorem.

Definitie 10.3 (Critical numbers and critical points)
Let f be defined at c. The value c is a **critical number** (or **critical value** (*kritische waarde*)) of f if $f'(c) = 0$.
If c is a critical number of f, then the point $(c, f(c))$ is a **critical point** (*kritisch punt*) of f.

Definitie 10.4 (Singularities and singular points)
Let f be defined at c. The function has a **singularity** (*singulariteit*) at $x = c$ if $f'(c)$ is not defined.
The point $(c, f(c))$ is called the **singular point** (*singulier punt*).

Theorem 10.2 (Relative extrema and critical points)
Let a function f be defined on an open interval I containing c, and let f have a relative extremum at the point $(c, f(c))$. Then $(c, f(c))$ is a critical or singular point of f.

In case the function f is also differentiable on an open interval I containing c, we restate Theorem 10.2 as follows.

Theorem 10.3 (Fermat's theorem)
Let a function f be defined and differentiable on an open interval I containing c, and let f have a relative extremum at the point $(c, f(c))$. Then $f'(c) = 0$.

Be careful to understand that Theorem 10.2 states that relative extrema on open intervals occur at critical or singular points. It does not say that all such points produce relative extrema. For instance, consider the function $f(x) = x^3$. Since $f'(x) = 3x^2$, it is straightforward to determine that $x = 0$ is a critical number of f. However, f has no relative extrema, as illustrated in Figure 10.3.

Theorem 10.1 states that a continuous function on a closed interval will have both an absolute maximum and an absolute minimum. Common sense tells us extrema occur either at the endpoints or somewhere in between. It is easy to check for extrema at endpoints, but there are infinitely many points to check that are in between. Our theory tells us we need only to check at the critical and singular points that are in between the endpoints. We combine these concepts to offer a strategy for finding extrema of a continuous function f defined on a closed interval $[a, b]$.

1. Evaluate f at the endpoints a and b of the interval.
2. Find the critical numbers and singularities of f in $[a, b]$.
3. Evaluate f at each critical number and singularity.
4. The absolute maximum of f is the largest of these values, and the absolute minimum of f is the least of these values.

We practice these ideas in the next examples.

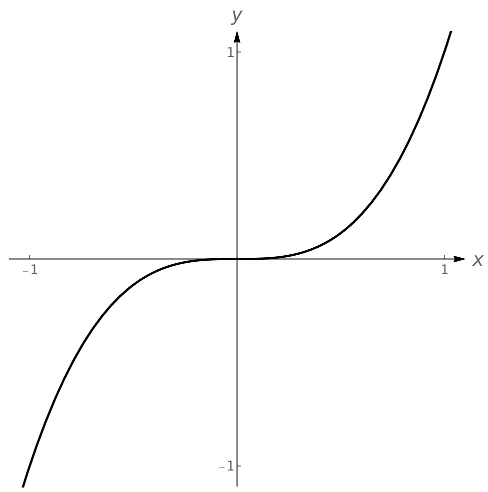


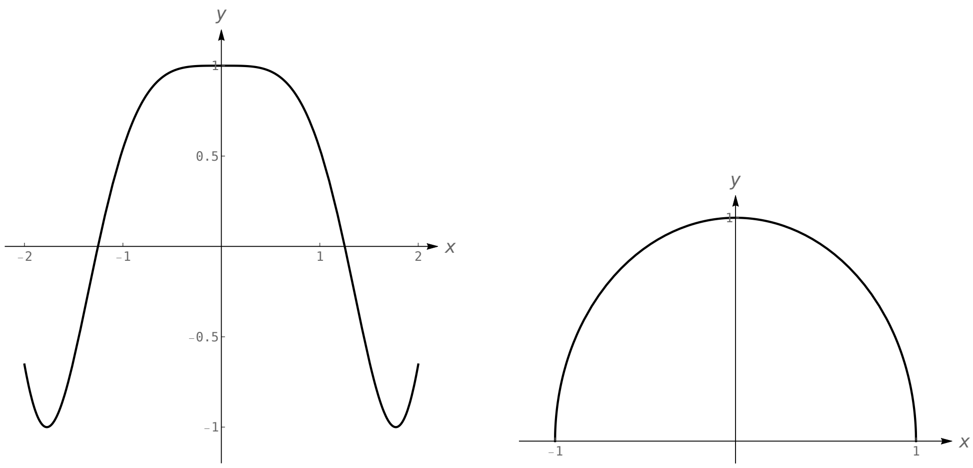
Figure 10.3: A graph of $f(x) = x^3$ which has a critical value of $x = 0$, but no relative extrema.

Example 10.2

Find the extrema of the following functions

1. $f(x) = \cos(x^2)$ on $[-2, 2]$,
2. $g(x) = \sqrt{1 - x^2}$,

which are graphed in Figure 10.4(a) and 10.4(b), respectively.



(a) (b)

Figure 10.4: A graph of $f(x) = \cos(x^2)$ on $[-2, 2]$ (a) and $g(x) = \sqrt{1 - x^2}$ (b).

Solution

1. Evaluating f at the endpoints of the interval gives: $f(-2) = f(2) = \cos(4) \approx -0.6536$. We now find the critical values of f. Applying the chain rule, we find $f'(x) = -2x\sin(x^2)$. Set $f'(x) = 0$ and solve for x to find the critical values of f. We do not have to bother about singularities because f' is defined everywhere.

We have $f'(x) = 0$ when $x = 0$ and when $\sin(x^2) = 0$. In general,

$$\sin(t) = 0 \quad \Leftrightarrow \quad t = \ldots -2\pi, -\pi, 0, \pi, \ldots .$$

Thus $\sin(x^2) = 0$ when $x^2 = 0, \pi, 2\pi, \ldots$ (x^2 is always positive so we ignore $-\pi$, etc.) So $\sin(x^2) = 0$ when $x = 0, \pm\sqrt{\pi}, \pm\sqrt{2\pi}$, etc. The only values to fall in the given interval of $[-2, 2]$ are 0 and $\pm\sqrt{\pi}$, where $\sqrt{\pi} \approx 1.77$. We construct a table for the 5 important values: $x = 0, \pm 2, \pm\sqrt{\pi}$:

x	-2	$-\sqrt{\pi}$	0	$\sqrt{\pi}$	2
$f(x)$	-0.65	-1	1	-1	-0.65

From this table it is clear that the maximum value of f on $[-2, 2]$ is 1 and occurs at $x = 0$; the minimum value is -1 and occurs at $x = \pm\sqrt{\pi}$. The graph of f in Figure 10.4(a) confirms our results.

2. A closed interval is not given, so we find the extreme values of g on its domain. g is defined whenever $1 - x^2 \geq 0$; thus the domain of g is $[-1, 1]$. Evaluating g at either endpoint returns 0. Using the chain rule, we find

$$g'(x) = \frac{-x}{\sqrt{1-x^2}}.$$

The critical points of g are found when $g'(x) = 0$, and its singularities when g' is undefined. It is straightforward to find that $g'(x) = 0$ when $x = 0$, and g' is undefined when $x = \pm 1$, the endpoints of the interval. We get the following table of important values:

x	-1	0	1
$g(x)$	0	1	0

The maximum value is 1 and occurs at $x = 0$. The minimum value is 0 and occurs at $x = \pm 1$.

We can also find extrema of piecewise-defined functions as illustrated in the following example.

Example 10.3

Find the maximum and minimum values of f on $[-4, 2]$, where

$$f(x) = \begin{cases} (x-1)^2, & x \leq 0 \\ x+1, & x > 0. \end{cases}$$

Solution

Here f is piecewise–defined, but we can still apply the same approach as before since it is continuous on $[-4, 2]$, i.e. $\lim_{x\to 0} f(x) = f(0)$. Evaluating f at the endpoints gives:

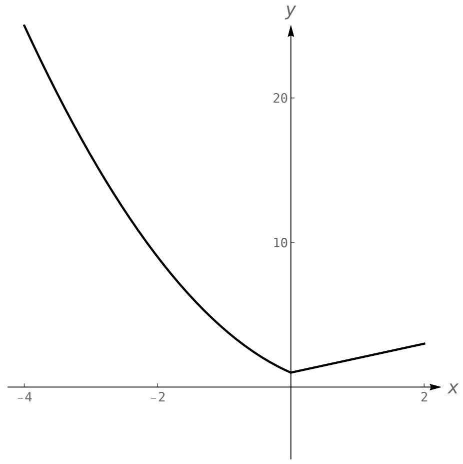
$$f(-4) = 25 \quad \text{and} \quad f(2) = 3.$$

We now find the critical numbers and/or singularities of f. We have to define f' in a piecewise manner; it is

$$f'(x) = \begin{cases} 2(x-1), & x < 0 \\ 1. & x > 0. \end{cases}$$

Note that while f is defined for all of $[-4, 2]$, f' is not, as the derivative of f does not exist when $x = 0$. From the left, the derivative approaches -2; from the right the derivative is 1. Thus f has a singularity at $x = 0$.

We now set $f'(x) = 0$. When $x > 0$, $f'(x)$ is never 0. When $x < 0$, $f'(x)$ is also never 0, so we find no critical values from setting $f'(x) = 0$. So we have three important x values to consider: $x = -4, 2$ and 0. Evaluating f at each gives, respectively, 25, 3 and 1, shown in Table 10.5(b). Thus the absolute minimum of f is 1, the absolute maximum of f is 25, confirmed by the graph of f in Figure 10.5(a).



(a)

x	$f(x)$
-4	25
0	1
2	3

(b)

Figure 10.5: A graph of $f(x)$ on $[-4, 2]$ as in Example 10.3 (a) and finding the extrema of f (b).

In the next section, we further our study of the information we can glean from nice functions with the mean value theorem. On a closed interval, we can find the average rate of change of a function. We will see that differentiable functions always have a point at which their instantaneous rate of change is same as the average rate of change. This is surprisingly useful, as we will see.

10.2 The mean value theorem

10.2.1 Introduction

We motivate this section with the following question: Suppose you leave your house and drive to your friend's house in a city 100 kilometres away, completing the trip in two hours. At any point during the trip do you necessarily have to be going 50 kilomteres per hour?

In answering this question, it is clear that the average speed for the entire trip is 50 km/hr, but the question is whether or not your instantaneous speed is ever exactly 50 km/hr. The answer, under some very reasonable assumptions, is yes.

Let us now see why this situation is in a calculus text by translating it into mathematical symbols.

First assume that the function $y = f(t)$ gives the distance (in kilometres) travelled from your home at time t (in hours) where $0 \leq t \leq 2$. In particular, this gives $f(0) = 0$ and $f(2) = 100$. The slope of the secant line connecting the starting and ending points $(0, f(0))$ and $(2, f(2))$ is therefore

$$\frac{\Delta f}{\Delta t} = \frac{f(2) - f(0)}{2 - 0} = \frac{100 - 0}{2} = 50\,\text{km/hr}.$$

The slope at any point on the graph itself is given by the derivative $f'(t)$. So, since the answer to the question above is yes, this means that at some time during the trip, the derivative takes on the value of 50 km/hr. Symbolically,

$$f'(c) = \frac{f(2)-f(0)}{2-0} = 50$$

for some time $0 \le c \le 2$.

How about more generally? Given any function $y = f(x)$ and a range $a \le x \le b$ does the value of the derivative at some point between a and b have to match the slope of the secant line connecting the points $(a, f(a))$ and $(b, f(b))$? Or equivalently, does the equation

$$f'(c) = \frac{f(b)-f(a)}{b-a}$$

have to hold for some $a < c < b$?

Consider, for instance, the functions

$$f_1(x) = \frac{1}{x^2} \quad \text{and} \quad f_2(x) = |x|$$

with $a = -1$ and $b = 1$ as shown in Figure 10.6 (a) and (b), respectively. Both functions have a value of 1 at a and b. Therefore the slope of the secant line connecting the end points is 0 in each case. But if you look at the plots of each, you can see that there are no points on either graph where the tangent lines have slope zero. Therefore we have found that there is no c in $[-1, 1]$ such that

$$f'(c) = \frac{f(1)-f(-1)}{1-(-1)} = 0.$$

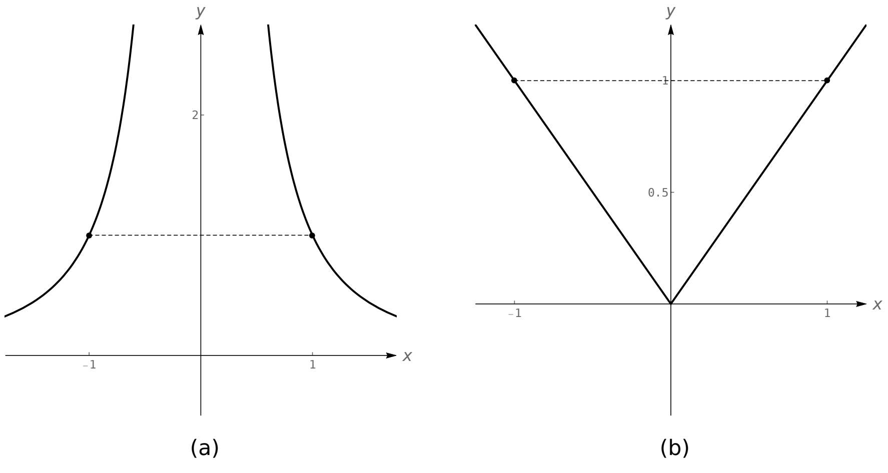


Figure 10.6: A graph of $f_1(x) = 1/x^2$ (a) and $f_2(x) = |x|$ (b).

10.2.2 The theorems

So what went wrong? It may not be surprising to find that the discontinuity of f_1 and the corner of f_2 play a role. If our functions had been continuous and differentiable, would we have been able to find that special value c? This is our motivation for the following theorem, which is sometimes also referred to as Lagrange's theorem.

Theorem 10.4 (The mean value theorem of differentiation)
Let $y = f(x)$ be a continuous function on the closed interval $[a, b]$ and differentiable on the open interval $]a, b[$. There exists a value c, with $a < c < b$, such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

That is, there is a value c in $]a, b[$ where the instantaneous rate of change of f at c is equal to the average rate of change of f on $[a, b]$.

Note that the reasons that the functions graphed in Figures 10.6(a) and 10.6(b) fail are indeed that f_1 has a discontinuity on the interval $[-1, 1]$ and f_2 is not differentiable at the origin.

The proof of the mean value theorem relies on the Rolle's theorem, which we also include to increase the understanding of the mean value theorem.

Theorem 10.5 (Rolle's theorem)
Let f be continuous on $[a, b]$ and differentiable on $]a, b[$, where $f(a) = f(b)$. There is some c in $]a, b[$ such that $f'(c) = 0$.

Consider Figure 10.7 where the graph of a function f is given, where $f(a) = f(b)$. It should make intuitive sense that if f is differentiable (and hence, continuous) that there would be a value c in $]a, b[$ where $f'(c) = 0$; that is, there would be a relative maximum or minimum of f in $]a, b[$. Rolle's theorem guarantees at least one; there may be more.

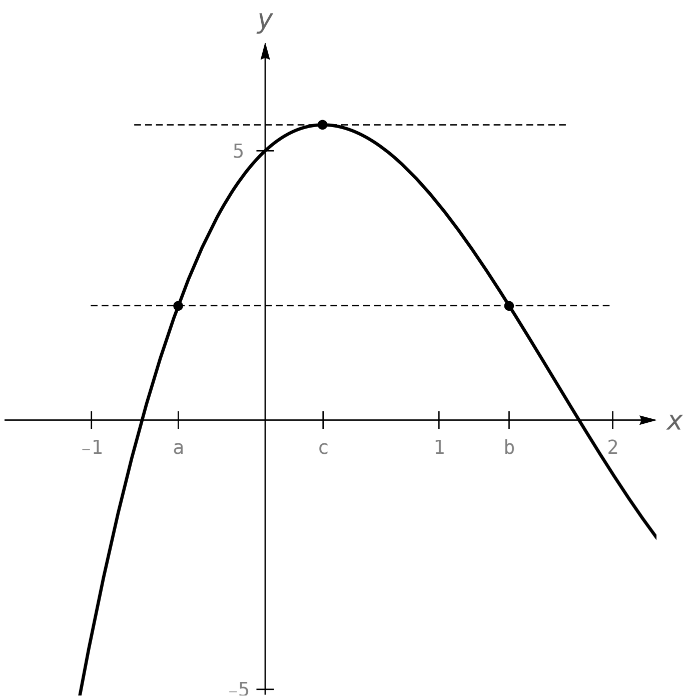


Figure 10.7: A graph of $f(x) = x^3 - 5x^2 + 3x + 5$, where $f(a) = f(b)$. Note the existence of c, where $a < c < b$, where $f'(c) = 0$.

Rolle's theorem is really just a special case of the mean value theorem. If $f(a) = f(b)$, then the average rate of change on $]a, b[$ is 0, and the theorem guarantees some c where $f'(c) = 0$.

Going back to the very beginning of the section, we see that the only assumption we would need about our distance function $f(t)$ is that it be continuous and differentiable for t from 0 to 2 hours (both reasonable assumptions). By the mean value theorem, we are guaranteed a time during the trip where our instantaneous speed is 50 km/hr. This fact is used in practice. Some law enforcement agencies monitor traffic speeds while in aircraft. They do not measure speed with radar, but rather by timing individual cars as they pass over lines painted on the highway whose distances apart are known. The officer is able to measure the average speed of a car between the painted lines; if that average speed is greater than the posted speed limit, the officer is assured that the driver exceeded the speed limit at some time.

Finally, note that the man value theorem is an **existence theorem** (*existentiestelling*). It states that a special value c exists, but it does not give any indication about how to find it. It turns out that when we need the mean value theorem, existence is all we need.

Example 10.4

Consider $f(x) = x^3 + 5x + 5$ on $[-3, 3]$. Find c in $]-3, 3[$ that satisfies the mean value theorem.

Solution

The average rate of change of f on $[-3, 3]$ is:

$$\frac{f(3) - f(-3)}{3 - (-3)} = \frac{84}{6} = 14.$$

We want to find c such that $f'(c) = 14$. We find $f'(x) = 3x^2 + 5$. We set this equal to 14 and solve for x.

$$\begin{array}{rrcl} & f'(x) & = & 14 \\ \Rightarrow & 3x^2 + 5 & = & 14 \\ \Leftrightarrow & x^2 & = & 3 \\ \Leftrightarrow & x & = & \pm\sqrt{3} \approx \pm 1.732 \end{array}$$

We have found 2 values c in $[-3, 3]$ where the instantaneous rate of change is equal to the average rate of change; the mean value theorem guaranteed at least one. In Figure 10.8, f is graphed with a dashed line representing the average rate of change; the lines tangent to f at $x = \pm\sqrt{3}$ are also given. Note how these lines are parallel (i.e., have the same slope) with the dashed line.

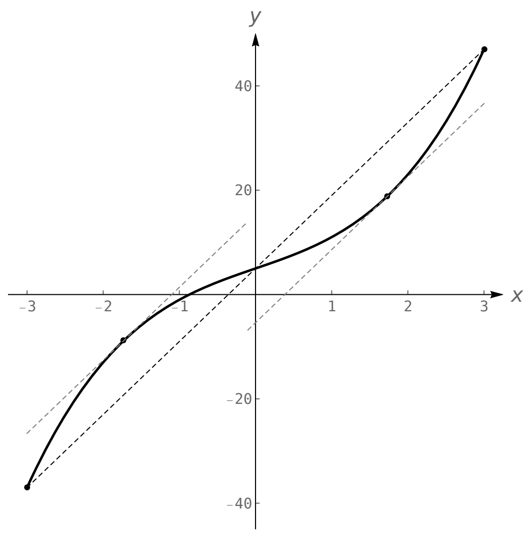


Figure 10.8: Demonstrating the mean value theorem in Example 10.4.

While the mean value theorem has practical use, such as the speed monitoring application mentioned before, it is mostly used to advance other theory. We will use it in the next section to relate the shape of a graph to its derivative.

10.3 Increasing and decreasing functions

Our study of nice functions f in this chapter has so far focused on individual points: points where f is maximal/minimal, points where $f'(x) = 0$ or f' does not exist, and points c where $f'(c)$ is the average rate of change of f on some interval.

In this section we begin to study how functions behave between special points; we begin studying in more detail the shape of their graphs by recalling the following definition from Chapter 3.

Definitie 10.5 (Increasing and decreasing functions)
Let f be a function defined on an interval I.

1. f is **increasing** (*stijgend*) on I if for every $a < b$ in I, $f(a) \le f(b)$.
2. f is **decreasing** (*dalend*) on I if for every $a < b$ in I, $f(a) \ge f(b)$.
3. f is **constant** (*constant*) on I if for every $a < b$ in I, $f(a) = f(b)$.

Informally, a function is increasing if as x gets larger (i.e., looking left to right) $f(x)$ gets larger, i.e. it does not decrease. Also recall that if the order $\le$ in the definition of an increasing function is replaced by $<$, we say that f is **strictly increasing** (*strikt stijgend*) on the interval I, and likewise for a **strictly decreasing** (*strikt dalend*) function.

Our interest lies in finding intervals in the domain of f on which f is either increasing or decreasing. Such information should seem useful. For instance, if f describes the speed of an object, we might want to know when the speed was increasing or decreasing (i.e., when the object was accelerating vs. decelerating). If f describes the population of a city, we should be interested in when the population is growing or declining.

To find such intervals, we again consider secant lines. Let f be a strictly increasing, differentiable function on an open interval I, such as the one shown in Figure 10.9, and let $a < b$ be given in I. The secant line on the graph of f from $x = a$ to $x = b$ is drawn; it has a slope of $(f(b) - f(a))/(b-a)$. But note:

$$\frac{f(b)-f(a)}{b-a} \quad \Rightarrow \quad \frac{\text{numerator } > 0}{\text{denominator } > 0} \quad \Rightarrow \quad \begin{array}{c}\text{slope of the}\\ \text{secant line}\\ > 0\end{array} \quad \Rightarrow \quad \begin{array}{c}\text{Average rate}\\ \text{of change of}\\ f \text{ on } [a,b] \text{ is}\\ > 0.\end{array}$$

We have shown mathematically what may have already been obvious: when f is strictly increasing, its secant lines will have a positive slope. Now recall the mean value theorem guarantees that there is a number c, where $a < c < b$, such that

$$f'(c) = \frac{f(b)-f(a)}{b-a} > 0.$$

By considering all such secant lines in I, we strongly imply that $f'(x) > 0$ on I. A similar statement can be made for strictly decreasing functions.

Our above logic can be summarized as If f is strictly increasing, then f' is probably positive. Theorem 10.6 turns this around by stating If f' is positive, then f is strictly increasing. This leads us to a method for finding when functions are strictly increasing and decreasing.

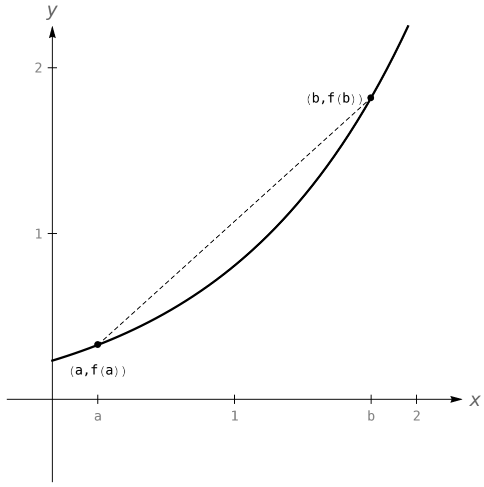


Figure 10.9: Examining the secant line of an increasing function.

Theorem 10.6 (Test for increasing/decreasing functions)
Let f be a continuous function on $[a, b]$ and differentiable on $]a, b[$.

1. *If $f'(c) > 0$ for all c in $]a, b[$, then f is strictly increasing on $[a, b]$.*
2. *If $f'(c) < 0$ for all c in $]a, b[$, then f is strictly decreasing on $[a, b]$.*
3. *If $f'(c) = 0$ for all c in $]a, b[$, then f is constant on $[a, b]$.*

Let f be differentiable on an interval I and let a and b be in I where $f'(a) > 0$ and $f'(b) < 0$. If f' is continuous on $[a, b]$, it follows from the intermediate value theorem (Theorem 8.7) that there must be some value c between a and b where $f'(c) = 0$. It turns out that this is still true even if f' is not continuous on $[a, b]$. This leads us to the following method for finding intervals on which a function is strictly increasing or decreasing.

Let f be a differentiable function on an interval I. To find intervals on which f is increasing and decreasing:

1. Find the critical values and singular points of f. That is, find all c in I where $f'(c) = 0$ or f' is not defined.
2. Use the critical values and singular points to divide I into subintervals.
3. Pick any point p in each subinterval, and find the sign of $f'(p)$.
   (a) If $f'(p) > 0$, then f is strictly increasing on that subinterval.
   (b) If $f'(p) < 0$, then f is strictly decreasing on that subinterval.

Note that parts 1 & 2 of Theorem 10.6 also hold if $f'(c) = 0$ for a finite number of values of c in I. Hence, acknowledging the difference between increasing and strictly increasing functions, we may say that

1. if $f'(p) \geq 0$, then f is increasing.
2. if $f'(p) \leq 0$, then f is decreasing.

We demonstrate using this process in the following example.

Example 10.5

Let $f(x) = x^3 + x^2 - x + 1$. Find intervals on which f is strictly increasing or decreasing.

Solution

Following the method outlined above, we first find the critical values of f. Hence, we have $f'(x) = 3x^2 + 2x - 1 = (3x-1)(x+1)$, so $f'(x) = 0$ when $x = -1$ and when $x = 1/3$. f' is never undefined.

Since an interval was not specified for us to consider, we consider the entire domain of f which is $\mathbb{R}$. We thus break the whole real line into three intervals based on the two critical values we just found: $]-\infty, -1[$, $]-1, 1/3[$ and $]1/3, +\infty[$.

We now pick a value p in each interval and find the sign of $f'(p)$. All we care about is the sign, so we do not actually have to fully compute $f'(p)$; pick nice values that make this simple.

Interval 1: $]-\infty, -1[$

We (arbitrarily) pick $p = -2$. We can compute $f'(-2)$ directly:

$$f'(-2) = 3(-2)^2 + 2(-2) - 1 = 7 > 0.$$

We conclude that f is increasing on $]-\infty, -1[$.

Note we can arrive at the same conclusion without computation. For instance, we could choose $p = -100$. The first term in $f'(-100)$, i.e., $3(-100)^2$ is clearly positive and very large. The other terms are small in comparison, so we know $f'(-100) > 0$. All we need is the sign.

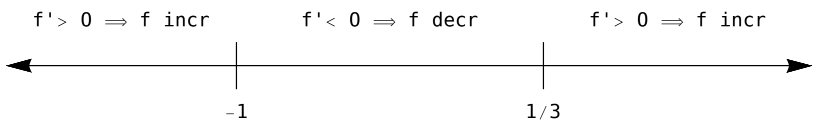
Interval 2: $]-1, 1/3[$

We pick $p = 0$ since that value seems easy to deal with. $f'(0) = -1 < 0$. We conclude f is decreasing on $]-1, 1/3[$.

Interval 3: $]1/3, +\infty[$

Pick an arbitrarily large value for $p > 1/3$ and note that $f'(p) = 3p^2 + 2p - 1 > 0$. We conclude that f is increasing on $]1/3, \infty[$.

In summary, we find:



We can verify our calculations by considering Figure 10.10, where f is graphed. The graph also presents f'; note how $f' > 0$ when f is strictly increasing and $f' < 0$ when f is strictly decreasing.

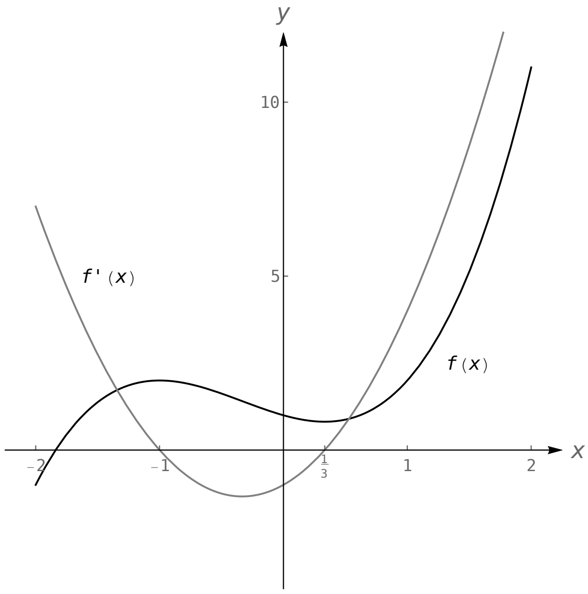


Figure 10.10: A graph of $f(x)$ (black) and $f'(x)$ (gray) in Example 10.5, showing where f is increasing and decreasing.

In Section 10.1 we learned the definition of relative maxima and minima and found that they occur at critical points. We are now learning from Example 10.5 that functions can switch from increasing to decreasing (and vice–versa) at critical points. This new understanding of increasing and decreasing creates a great method of determining whether a critical point corresponds to a maximum, minimum, or neither. Imagine a function increasing until a critical point at $x = c$, after which it decreases. A quick sketch helps confirm that $f(c)$ must be a relative maximum. A similar statement can be made for relative minimums. We formalize this concept in a theorem.

Theorem 10.7 (First derivative test)
Let f be differentiable on an interval I and let c be a critical number in I.

1. *If the sign of f' switches from positive to negative at c, then $f(c)$ is a relative maximum of f.*
2. *If the sign of f' switches from negative to positive at c, then $f(c)$ is a relative minimum of f.*
3. *If f' is positive (or, negative) before and after c, then $f(c)$ is not a relative extremum of f.*

Example 10.6

Find the intervals on which f is increasing and decreasing, and determine the relative extrema of f, where

$$f(x) = \frac{x^2+3}{x-1}.$$

Solution

We start by noting the domain of f: $]-\infty, 1[\cup]1, +\infty[$. Since the domain of f in this example is the union of two intervals, we apply Theorem 10.7 to both intervals of the domain of f.

Since f is not defined at $x = 1$, the increasing/decreasing nature of f could switch at this value. At this point f manifests a singularity, so we should keep track of it.

Using the quotient rule, we find
$$f'(x) = \frac{x^2-2x-3}{(x-1)^2}.$$

We can now find the critical values and possible further singular points of f; we want to know when $f'(x) = 0$ and when f' is not defined. That latter is straightforward: when the denominator of $f'(x)$ is 0, f' is undefined. That occurs when $x = 1$, which we have already recognized as an important value.

$f'(x) = 0$ when the numerator of $f'(x)$ is 0. That occurs when $x^2-2x-3 = (x-3)(x+1) = 0$; i.e., when $x = -1, 3$.

We have found that f has two critical numbers, $x = -1, 3$, and at $x = 1$ something important might also happen. These three numbers divide the real number line into 4 subintervals:
$$]-\infty, -1[, \quad]-1, 1[, \quad]1, 3[\quad \text{and} \quad]3, +\infty[.$$

Pick a number p from each subinterval and test the sign of f' at p to determine whether f is increasing or decreasing on that interval. Again, we do well to avoid complicated computations; notice that the denominator of f' is always positive so we can ignore it during our work.

Interval 1: $]-\infty, -1[$

Choosing a very small number (i.e., a negative number with a large magnitude) p returns p^2-2p-3 in the numerator of f'; that will be positive. Hence f is increasing on $]-\infty, -1[$.

Interval 2: $]-1, 1[$

Choosing 0 seems simple: $f'(0) = -3 < 0$. We conclude f is decreasing on $]-1, 1[$.

Interval 3: $]1, 3[$

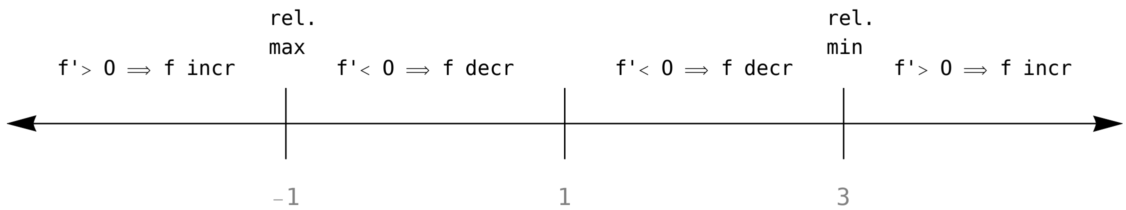
Choosing 2 seems simple: $f'(2) = -3 < 0$. Again, f is decreasing.

Interval 4: $]3, +\infty[$

Choosing an very large number p from this subinterval will give a positive numerator and (of course) a positive denominator. So f is increasing on $]3, +\infty[$.

In summary, f is increasing on the intervals $]-\infty, -1[$ and $]3, +\infty[$ and is decreasing on the intervals $]-1, 1[$ and $]1, 3[$. Since at $x = -1$, the sign of f' switched from positive to negative, Theorem 10.7 states that $f(-1)$ is a relative maximum of f. At $x = 3$, the sign of f' switched from negative to positive, meaning $f(3)$ is a relative minimum. At $x = 1$, f is not defined, so there is no relative extremum at $x = 1$.

This is summarized in the number line below. Also, Figure 10.11 shows a graph of f, confirming our calculations. This figure also shows f', again demonstrating that f is increasing when $f' > 0$ and decreasing when $f' < 0$.



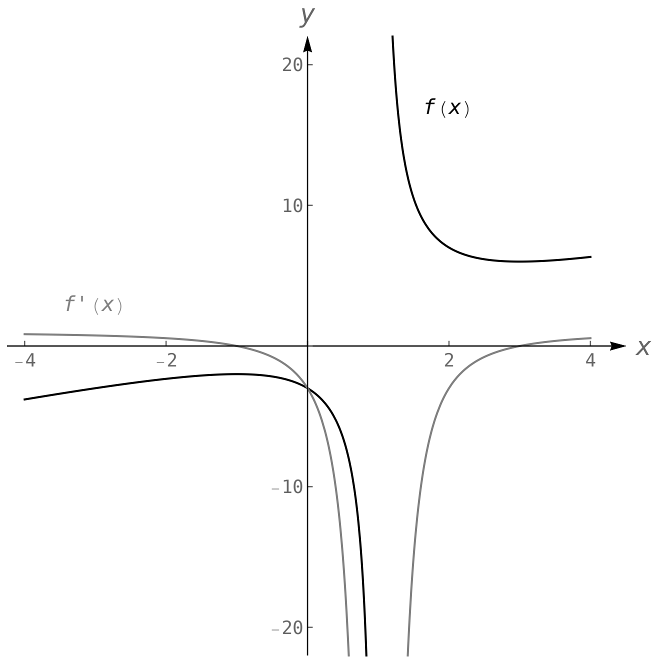


Figure 10.11: A graph of $f(x)$ and $f'(x)$ in Example 10.6, showing where f is increasing and decreasing.

We examine one example.

Example 10.7

Find the intervals on which $f(x) = x^{8/3} - 4x^{2/3}$ is increasing and decreasing and identify the relative extrema.

Solution

We start with taking a derivative. Since we know we want to solve $f'(x) = 0$, we will do some algebra after taking the derivative.

$$\begin{aligned} f(x) &= x^{\frac{8}{3}} - 4x^{\frac{2}{3}} \\ \Rightarrow \quad f'(x) &= \frac{8}{3}x^{\frac{5}{3}} - \frac{8}{3}x^{-\frac{1}{3}} \\ &= \frac{8}{3}x^{-\frac{1}{3}}\left(x^{\frac{6}{3}} - 1\right) \\ &= \frac{8}{3}x^{-\frac{1}{3}}(x^2 - 1) \\ &= \frac{8}{3}x^{-\frac{1}{3}}(x-1)(x+1). \end{aligned}$$

This derivation of f' shows that $f'(x) = 0$ when $x = \pm 1$ and f' is not defined when $x = 0$. Thus we have 2 critical values and one singular point, breaking the number line into 4 subintervals.

Interval 1: $]-\infty, -1[$

We choose $p = -2$; we can easily verify that $f'(-2) < 0$. So f is decreasing on $]-\infty, -1[$.

Interval 2: $]-1, 0[$

Choose $p = -1/2$. We can once more find the sign of $f'(p)$ without computing an actual value. We have $f'(p) = (8/3)p^{-1/3}(p-1)(p+1)$; find the sign of each of the three terms.

$$f'(p) = \frac{8}{3} \cdot \underbrace{p^{-\frac{1}{3}}}_{<0} \cdot \underbrace{(p-1)}_{<0}\underbrace{(p+1)}_{>0}.$$

Consequently, f is increasing on $]-1,0[$.

Interval 3: $]0,1[$

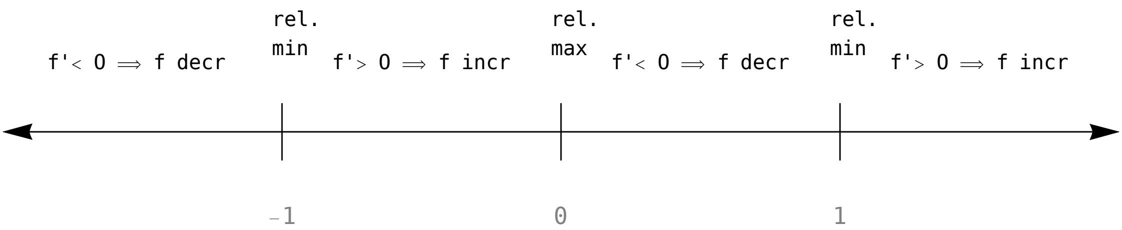
We do a similar sign analysis as before, using p in $]0,1[$.

$$f'(p) = \frac{8}{3} \cdot \underbrace{p^{-\frac{1}{3}}}_{>0} \cdot \underbrace{(p-1)}_{<0}\underbrace{(p+1)}_{>0}.$$

We have 2 positive factors and one negative factor; $f'(p) < 0$ and so f is decreasing on $]0,1[$.

Interval 4: $]1,+\infty[$ Similar work to that done for the other three intervals shows that $f'(x) > 0$ on $]1,+\infty[$, so f is increasing on this interval.

Finally, we have:



Consequently, we conclude by stating that f is increasing on the intervals $]-1,0[$ and $]1,+\infty[$ and decreasing on the intervals $]-\infty,-1[$ and $]0,1[$. The sign of f' changes from negative to positive around $x=-1$ and $x=1$, meaning by Theorem 10.7 that $f(-1)$ and $f(1)$ are relative minima of f. As the sign of f' changes from positive to negative at $x=0$, we have a relative maximum at $f(0)$. Figure 10.12 shows a graph of f, confirming our result. We also graph f', highlighting once more that f is increasing when $f'>0$ and is decreasing when $f'<0$.

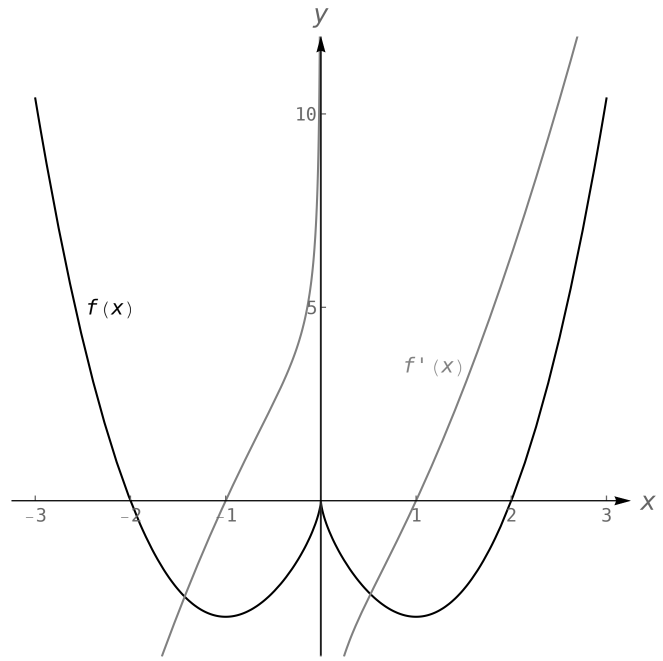


Figure 10.12: A graph of $f(x)$ (black) and $f'(x)$ (gray) in Example 10.7, showing where f is increasing and decreasing.

We have seen how the first derivative of a function helps determine when the function is going up or down. In the next section, we will see how the second derivative helps determine how the graph of a function curves.

10.4 Concavity and the second derivative

Our study of nice functions continues. The previous section showed how the first derivative of a function, f', can relay important information about f. We now apply the same technique to f' itself, and learn what this tells us about f.

The key to studying f' is to consider its derivative, namely f'', which is the second derivative of f. When $f'' \geq 0$, f' is increasing. When $f'' \leq 0$, f' is decreasing. f' has relative maxima and minima where $f'' = 0$ or is undefined.

This section explores how knowing information about f'' gives information about f.

10.4.1 Concavity

We begin with a definition, then explore its meaning.

Definitie 10.6 (Concave up and concave down)
Let f be differentiable on an interval I.

1. The graph of f is **concave up** (*convex*) on I if f' is increasing.
2. The graph of f is **concave down** (*concaaf*) on I if f' is decreasing.
3. If f' is constant then the graph of f is said to have no **concavity** (*concaviteit*).

Note that we often state that f is concave up instead of the graph of f is concave up for simplicity. Besides, in agreement with the terminology used for increasing and decreasing functions (Definition 10.5), we call a function f strictly concave up or down if f' is strictly increasing or decreasing, respectively.

The graph of a function f is concave up when f' is increasing. That means as one looks at a concave up graph from left to right, the slopes of the tangent lines will be increasing. Consider Figure 10.13(a), where a concave up graph is shown along with some tangent lines. Notice how the tangent line on the left is steep, downward, corresponding to a small value of f'. On the right, the tangent line is steep, upward, corresponding to a large value of f'. If a function is decreasing and concave up, then its rate of decrease is slowing; it is levelling off. If the function is increasing and concave up, then the rate of increase is increasing.

Now consider a function which is concave down. We essentially repeat the above paragraphs with slight variation. The graph of a function f is concave down when f' is decreasing. That means as one looks at a concave down graph from left to right, the slopes of the tangent lines will be decreasing. Consider Figure 10.13(b), where a concave down graph is shown along with some tangent lines. Notice how the tangent line on the left is steep, upward, corresponding to a large value of f'. On the right, the tangent line is steep, downward, corresponding to a small value of f'. If a function is increasing and concave down, then its rate of increase is slowing; it is levelling off. If the function is decreasing and concave down, then the rate of decrease is decreasing. The function is decreasing at a faster and faster rate. Geometrically speaking it is clear that a function is concave up if its graph lies above its tangent lines. A function is concave down if its graph lies below its tangent lines.

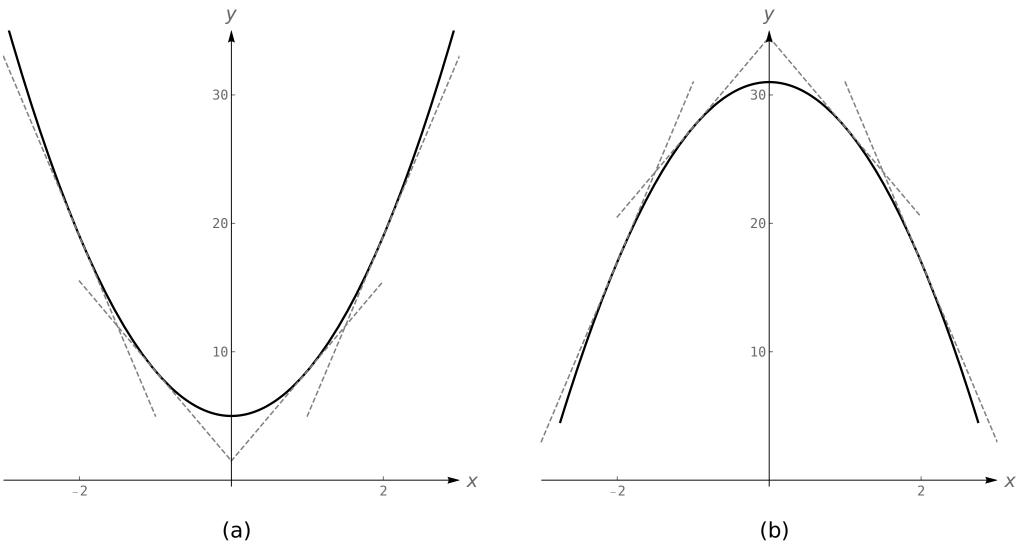


Figure 10.13: A function f with a concave up (a) and concave down (b) graph together with some tangent lines (dashed).

Our definition of concave up and concave down is given in terms of when the first derivative is increasing or decreasing. We can apply the results of Section 10.3 to find intervals on which a graph is concave up or down. That is, we recognize that f' is increasing when $f'' \geq 0$, etc.

Theorem 10.8 (Test for concavity)
Let f be twice differentiable on an interval I. The graph of f is concave up if $f'' \geq 0$ on I, and is concave down if $f'' \leq 0$ on I.

Figure 10.14 demonstrates the four ways that concavity interacts with increasing/decreasing, along with the relationships with the first and second derivatives.

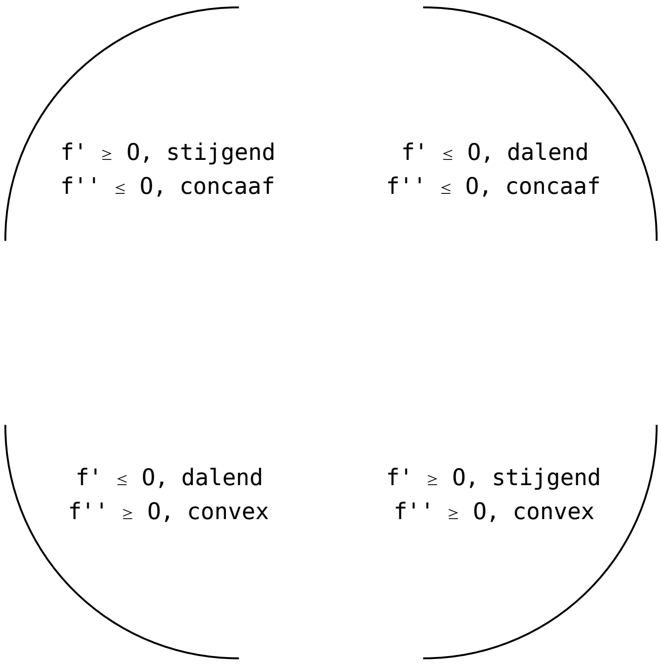


Figure 10.14: Demonstrating the 4 ways that concavity interacts with increasing/decreasing, along with the relationships with the first and second derivatives.

If knowing where a graph is concave up/down is important, it makes sense that the places where the graph changes from one to the other is also important. This leads us to a definition.

Definitie 10.7 (Point of inflection)
A **point of inflection** (*buigpunt*) is a point on the graph of f at which the concavity of f changes.

If the concavity of f changes at a point $(c, f(c))$, then f' is changing from increasing to decreasing (or, decreasing to increasing) at $x = c$. That means that the sign of f'' is changing from positive to negative (or, negative to positive) at $x = c$. This leads to the following theorem.

Theorem 10.9 (Points of inflection)
If $(c, f(c))$ is a point of inflection on the graph of f, then either $f''(c) = 0$ or f'' is not defined at c.

We have identified the concepts of concavity and points of inflection. It is now time to practice using these concepts; given a function, we should be able to find its points of inflection and identify intervals on which it is concave up or down. We do so in the following example.

Example 10.8

Let

$$f(x) = \frac{x}{x^2 - 1}.$$

Find the inflection points of f and the intervals on which it is concave up/down.

Solution

We need to find f' and f''. Using the quotient rule and simplifying, we find

$$f'(x) = \frac{-(1+x^2)}{(x^2-1)^2} \quad \text{and} \quad f''(x) = \frac{2x(x^2+3)}{(x^2-1)^3}.$$

To find the possible points of inflection, we seek to find where $f''(x) = 0$ and where f'' is not defined. Solving $f''(x) = 0$ reduces to solving $2x(x^2+3) = 0$; we find $x = 0$. We find that f'' is not defined when $x = \pm 1$, for then the denominator of f'' is 0. We also note that f itself is not defined at $x = \pm 1$, having a domain of $]-\infty,-1[\,\cup\,]-1,1[\,\cup\,]1,+\infty[$. Since the domain of f is the union of three intervals, it makes sense that the concavity of f could switch across intervals. We technically cannot say that f has a point of inflection at $x = \pm 1$ as they are not part of the domain, but we must still consider these x-values to be important and will include them in our number line.

The important x-values at which concavity might switch are $x = -1$, $x = 0$ and $x = 1$, which split the number line into four intervals.

We determine the concavity on each. Keep in mind that all we are concerned with is the sign of f'' on the interval.

Interval 1: $]-\infty,-1[$

Select a number c in this interval with a large magnitude (for instance, $c = -100$). The denominator of $f''(x)$ will be positive. In the numerator, the (c^2+3) will be positive and the $2c$ term will be negative. Thus the numerator is negative and $f''(c)$ is negative. We conclude f is concave down on $]-\infty,-1[$.

Interval 2: $]-1,0[$

For any number c in this interval, the term $2c$ in the numerator will be negative, the term (c^2+3) in the numerator will be positive, and the term $(c^2-1)^3$ in the denominator will be negative. Thus $f''(c) > 0$ and f is concave up on this interval.

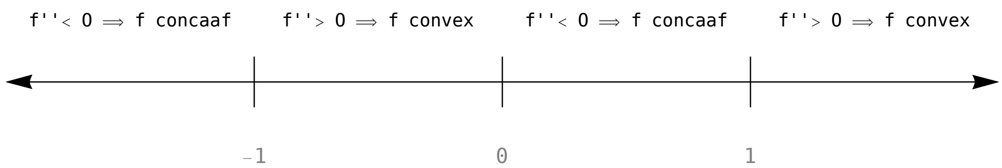
Interval 3: $]0,1[$

Any number c in this interval will be positive and small. Thus the numerator is positive while the denominator is negative. Thus $f''(c) < 0$ and f is concave down on this interval.

Interval 4: $]1,+\infty[$

Choose a large value for c. It is evident that $f''(c) > 0$, so we conclude that f is concave up on $]1,+\infty[$.

Since, we get



we conclude that f is concave up on $]-1,0[$ and $]1,+\infty[$ and concave down on $]-\infty,-1[$ and $]0,1[$. There is only one point of inflection, $]0,0[$, as f is not defined at $x=\pm 1$. Our work is confirmed by the graph of f in Figure 10.15.

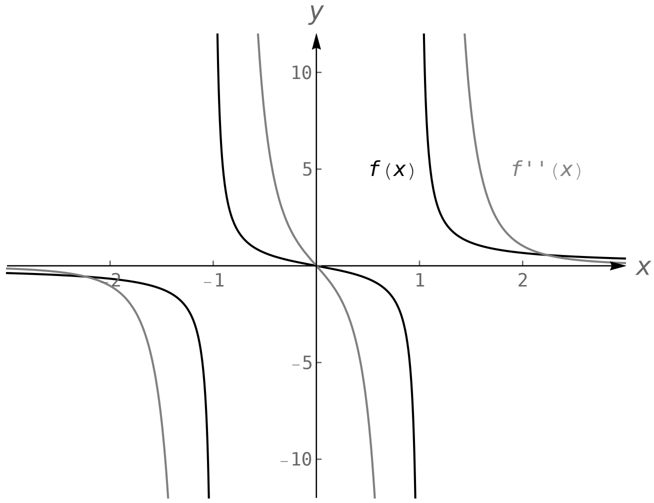


Figure 10.15: A graph of $f(x)$ (black) and $f''(x)$ (gray) in Example 10.8.

Recall that relative maxima and minima of f are found at critical points of f; that is, they are found when $f'(x)=0$ or when f' is undefined. Likewise, the relative maxima and minima of f' are found when $f''(x)=0$ or when f'' is undefined; note that these are the inflection points of f.

What does a relative maximum of f' mean? The derivative measures the rate of change of f; maximizing f' means finding where f is increasing the most – where f has the steepest tangent line. A similar statement can be made for minimizing f'; it corresponds to where f has the steepest negatively–sloped tangent line.

We utilize this concept in the next example.

Example 10.9

The sales of a certain product over a three-year span are modelled by $S(t)=t^4-8t^2+20$, where t is the time in years. Over the first two years, sales are decreasing. Find the point at which sales are decreasing at their greatest rate.

Solution

We want to maximize the rate of decrease, which is to say, we want to find where S' has a minimum. To do this, we find where S'' is 0. We find $S'(t)=4t^3-16t$ and $S''(t)=12t^2-16$. Setting $S''(t)=0$ and solving, we get $t=\sqrt{4/3}\approx 1.16$. Note that we ignore the negative value of t since it does not lie in the domain of our function S.

This is both the inflection point and the point of maximum decrease. This is the point at which things first start looking up for the company. After the inflection point, it will still take some time before sales start to increase, but at least sales are not decreasing quite as quickly as they had been.

A graph of $S(t)$ and $S'(t)$ is given in Figure 10.16. When $S'(t)<0$, sales are decreasing; note how at $t\approx 1.16$, $S'(t)$ is minimized. That is, sales are decreasing at the fastest rate at $t\approx 1.16$. On the interval of $(1.16,2)$, S is decreasing but concave up, so the decline in sales is levelling off.

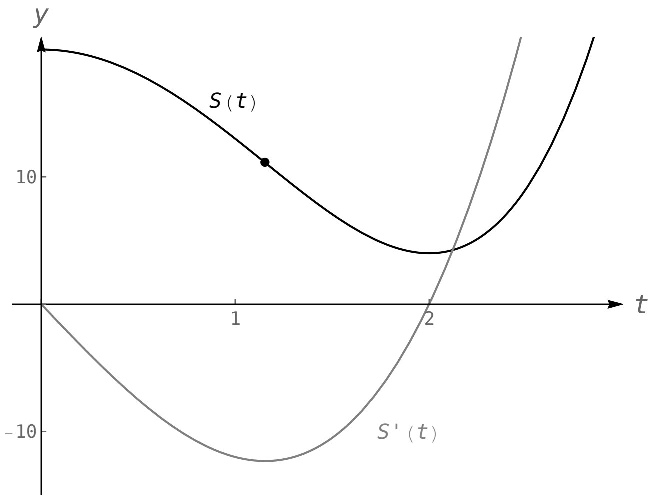


Figure 10.16: A graph of $S(t)$ (black) in Example 10.9 along with $S'(t)$ (gray).

Not every critical point corresponds to a relative extrema; $f(x) = x^3$ has a critical point at $(0,0)$ but no relative maximum or minimum (Figure 10.3). Likewise, just because $f''(x) = 0$ we cannot conclude concavity changes at that point. We were careful before to use terminology possible point of inflection since we needed to check to see if the concavity changed. The canonical example of $f''(x) = 0$ without concavity changing is $f(x) = x^4$. At $x = 0$, $f''(x) = 0$ but f is always concave up, as shown in Figure 10.17.

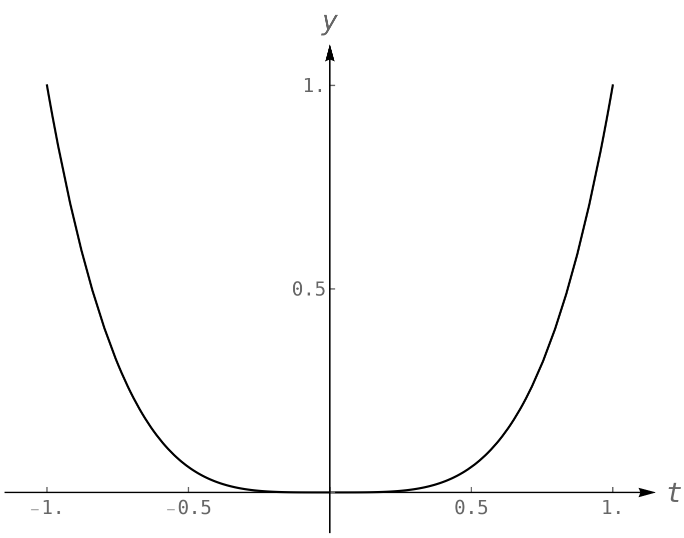


Figure 10.17: A graph of $f(x) = x^4$.

10.4.2 The second derivative test

The first derivative of a function gave us a test to find if a critical value corresponded to a relative maximum, minimum, or neither. The second derivative gives us another way to test if a critical point is a local maximum or minimum. The following theorem states something that is intuitive: if a critical value occurs in a region where a function f is concave up, then that critical value must correspond to a relative minimum of f, etc (Figure 10.18).

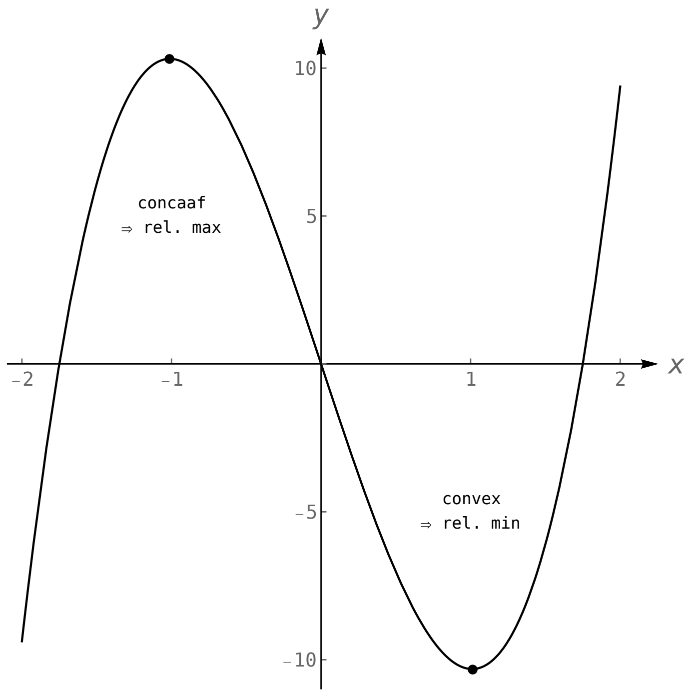


Figure 10.18: Demonstrating the second derivative test.

Theorem 10.10 (The second derivative test)
Let c be a critical value of f where $f''(c)$ is defined.

1. *If $f''(c) < 0$, then f has a local maximum at $(c, f(c))$.*
2. *If $f''(c) > 0$, then f has a local minimum at $(c, f(c))$.*
3. *If $f''(c) = 0$ then $x = c$ can be a local maximum, relative minimum or neither.*

The second derivative test relates to the first derivative test in the following way. If $f''(c) > 0$, then the graph is concave up at a critical point c and f' itself is growing. Since $f'(c) = 0$ and f' is growing at c, then it must go from negative to positive at c. This means the function goes from decreasing to increasing, indicating a local minimum at c.

Example 10.10

Let
$$f(x) = \frac{100}{x} + x.$$
Find the critical points of f and label them as relative maxima or minima.

Solution

We find
$$f'(x) = -\frac{100}{x^2} + 1$$
and
$$f''(x) = \frac{200}{x^3}.$$
We set $f'(x) = 0$ and solve for x to find the critical values. Note that f' is not defined at $x = 0$, but neither is f so this is not a critical value. We find the critical values are $x = \pm 10$. Evaluating f''

at $x = 10$ gives $0.2 > 0$, so there is a local minimum at $x = 10$. Evaluating $f''(-10) = -0.2 < 0$, determining a relative maximum at $x = -10$. These results are confirmed in Figure 10.19.

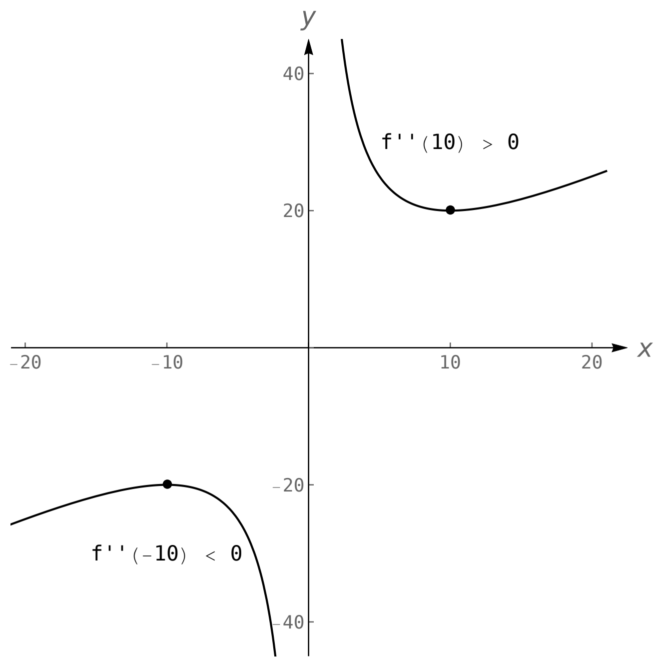


Figure 10.19: A graph of $f(x) = 100/x + x$ in Example 10.10.

We have been learning how the first and second derivatives of a function relate information about the graph of that function. We have found intervals of increasing and decreasing, intervals where the graph is concave up and down, along with the locations of relative extrema and inflection points. In Chapter 8 we saw how limits explained asymptotic behaviour. In the next section we combine all of this information to produce accurate sketches of functions.

10.5 Curve sketching

We have been learning how we can understand the behaviour of a function based on its first and second derivatives. While we have been treating the properties of a function separately, we combine them here to produce an accurate graph of the function without plotting lots of extraneous points. Why bother? Graphing utilities are very accessible, whether on a computer, a hand–held calculator, or a smartphone. These resources are usually very fast and accurate. For instance, remember that we can plot an explicitly defined function in Mathematica using the built-in command `Plot`, while `ContourPlot` may be used to plot implicitly defined functions (see Chapter 3). We will see that our method is not particularly fast – it will require time. We are attempting to understand the behavior of a function f based on the information given by its derivatives. While all of a function's derivatives relay information about it, it turns out that most of the behaviour we care about is explained by f' and f''. Understanding the interactions between the graph of f and f' and f'' is important. To gain this understanding, one might argue that all that is needed is to look at lots of graphs. This is true to a point, but is somewhat similar to stating that one understands how an engine works after looking only at pictures. It is true that the basic ideas will be conveyed, but hands–on access increases understanding.

To produce an accurate sketch of a given function f, take the following steps.

1. Find the domain of f. Generally, we assume that the domain is the entire real line then find restrictions, such as where a denominator is 0 or where negatives appear under the radical.

2. Find symmetries and intercepts.
3. Find the location of any asymptotes of f:
   (a) vertical
   (b) horizontal
   (c) slant asymptotes
4. Find the critical and singular points of f.
5. Find the possible points of inflection of f.
6. Create a number line that includes all critical points, possible points of inflection, and locations of vertical asymptotes. For each interval created, determine whether f is increasing or decreasing, concave up or down.
7. Evaluate f at each critical point and possible point of inflection. Plot these points on a set of axes. Connect these points with curves exhibiting the proper concavity. Sketch asymptotes and x- and y-intercepts where applicable.

Example 10.11

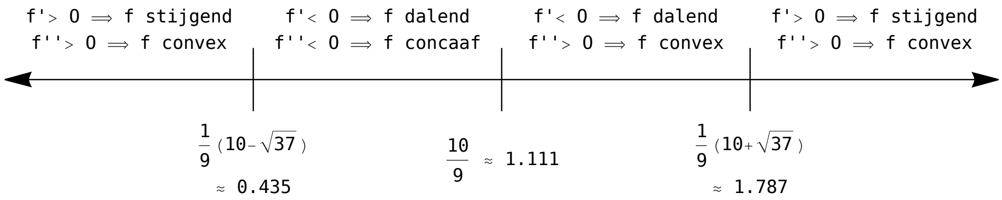
Sketch $f(x) = 3x^3 - 10x^2 + 7x + 5$.

Solution

We follow the steps outlined above.

1. The domain of f is the entire real line; there are no values x for which $f(x)$ is not defined.
2. It can be verified easily that the function is neither even nor odd. Besides, the x-intercept is about $x = -0.424$ while the y-intercept is $y = 5$.
3. (a) There are no vertical asymptotes.
   (b) We determine the end behaviour using limits as x approaches $\pm\infty$.
   $$\lim_{x\to-\infty} f(x) = -\infty \qquad \text{and} \qquad \lim_{x\to+\infty} f(x) = +\infty$$
   So, we do not have any horizontal asymptotes.
   (c) There is no slant asymptote since
   $$a = \lim_{x\to\pm\infty} \frac{f(x)}{x} = \lim_{x\to\pm\infty} \left(3x^2 - 10x + 7 + \frac{5}{x}\right) = +\infty.$$
4. Find the critical values of f. We compute $f'(x) = 9x^2 - 20x + 7$. Use the quadratic formula to find the roots of f':
   $$x = \frac{20 \pm \sqrt{(-20)^2 - 4(9)(7)}}{2(9)} = \frac{1}{9}\left(10 \pm \sqrt{37}\right),$$
   so we have $x \approx 0.435$ or $x \approx 1.787$.
5. Find the possible points of inflection of f. Compute $f''(x) = 18x - 20$. We have
   $$f''(x) = 0 \quad \Rightarrow \quad x = \frac{10}{9} \approx 1.111.$$

6. We place the values $x = (10 \pm \sqrt{37})/9$ and $x = 10/9$ on a number line and we mark each interval as increasing or decreasing, concave up or down:



7. We now plot the appropriate points and connect the points in such a way that the proper concavity is demonstrated. Our curve crosses the y-axis at $y = 5$ and crosses the x-axis near $x = -0.424$ (Figure 10.20).

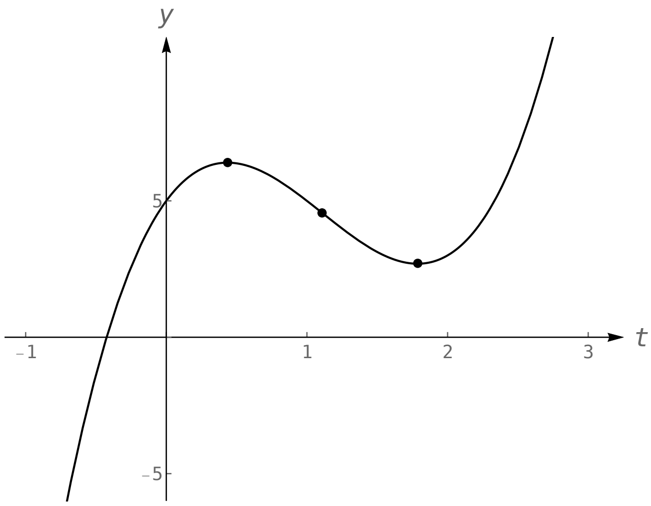


Figure 10.20: A sketch of $f(x) = 3x^3 - 10x^2 + 7x + 5$ in Example 10.11.

Example 10.12

Sketch

$$f(x) = \frac{x^2 - x - 2}{x^2 - x - 6}.$$

Solution

We again follow the steps outlined above.

1. In determining the domain, we assume it is all real numbers and look for restrictions. We find that at $x = -2$ and $x = 3$, $f(x)$ is not defined because the denominator of $f(x)$ is 0 at those points. So,

$$\text{dom} f = \{\text{real numbers } x \mid x \neq -2, 3\}.$$

2. It can be verified easily that the function is neither even nor odd. Besides, the x-intercepts are $x = -1$ and $x = 2$ while the y-intercept is $y = 1/3$.

3. (a) The vertical asymptotes of f are at $x = -2$ and $x = 3$, the places where f is undefined and the numerator of $f(x)$ is not zero.

(b) There is a horizontal asymptote of $y = 1$, as $\lim_{x\to-\infty} f(x) = 1$ and $\lim_{x\to+\infty} f(x) = 1$.

(c) There are no slant asymptotes because there are already horizontal ones.

4. To find the critical values of f, we first find $f'(x)$. Using the quotient rule, we find

$$f'(x) = \frac{-8x+4}{(x^2+x-6)^2} = \frac{-8x+4}{(x-3)^2(x+2)^2}.$$

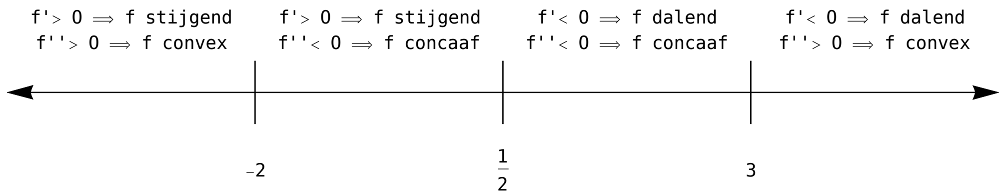
$f'(x) = 0$ when $x = 1/2$, and f' is undefined when $x = -2$ or $x = 3$. Since f' is undefined only when f is, these are not singular values. The only critical value is $x = 1/2$.

5. To find the possible points of inflection, we find $f''(x)$, again employing the Quotient Rule:

$$f''(x) = \frac{24x^2-24x+56}{(x-3)^3(x+2)^3}.$$

We find that $f''(x)$ is never 0 (setting the numerator equal to 0 and solving for x, we find the only roots to this quadratic are complex) and f'' is undefined when $x = -2$ or $x = 3$. Thus concavity will possibly only change at $x = -2$ and $x = 3$.

6. We place the values $x = 1/2$, $x = -2$ and $x = 3$ on a number line and we mark in each interval whether f is increasing or decreasing, concave up or down:



We see that f has a relative maximum at $x = 1/2$; concavity changes only at the vertical asymptotes.

7. In Figure 10.21, we plot the points from the number line on a set of axes and connect them in such a way that we get the appropriate concavity. We also show f crossing the x-axis at $x = -1$ and $x = 2$.

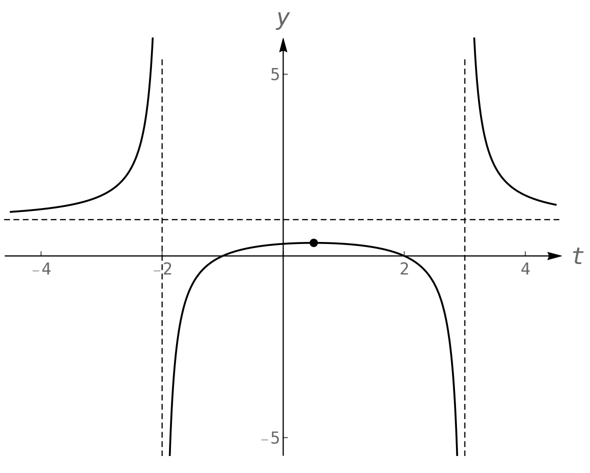


Figure 10.21: A sketch of $f(x) = \frac{x^2-x-2}{x^2-x-6}$ in Example 10.12.

Now why are computer graphics so good at curve sketching? It is not because computers are smarter than we are. Rather, it is largely because computers are much faster at computing than we are. In general, computers graph functions plot equally spaced points, then connect the dots using lines. By using lots of points, the connecting lines are short and the graph looks smooth. This does a fine job of graphing in most cases. However, in regions where the graph is very curvy, this can generate noticeable sharp edges on the graph unless a large number of points are used. High quality computer algebra systems, such as Mathematica, use special algorithms to plot lots of points only where the graph is curvy.

In Figure 10.22, a graph of $y = \sin(x)$ is given, generated by Mathematica using the Mathematica-function **`Plot`**. The small points represent each of the places Mathematica sampled the function. Notice how at the bends of $\sin(x)$, lots of points are used; where $\sin(x)$ is relatively straight, fewer points are used. Moreover, many points are also used at the endpoints to ensure the end behavior is accurate.

How does Mathematica know where the graph is curvy? Calculus. When we study curvature in a later chapter, we will see how the first and second derivatives of a function work together to provide a measurement of curviness. Mathematica employs algorithms to determine regions of high curvature and plots extra points there. Again, the goal of this section is to understand that the shape of the graph of a function is largely determined by understanding the behaviour of the function at a few key places. For instance, in Example 10.12, we were able to accurately sketch a complicated graph using only a few points and knowledge of asymptotes!

Computer algebra systems

A computer algebra system is any mathematical software with the ability to manipulate mathematical expressions in a way similar to the traditional manual computations of mathematicians and scientists. The development of such systems started in the second half of the previous century.

The first popular computer algebra systems were muMATH, Reduce, Derive, and Macsyma. Today, the most popular commercial systems are Mathematica and Maple, which are commonly used by research mathematicians, scientists, and engineers. Freely available alternatives include SageMath[a] and SymPy.

[a] http://www.sagemath.org/

10.6 Optimization

In Section 10.1 we learned about extreme values – the largest and smallest values a function attains on an interval. We motivated our interest in such values by discussing how it made sense to want

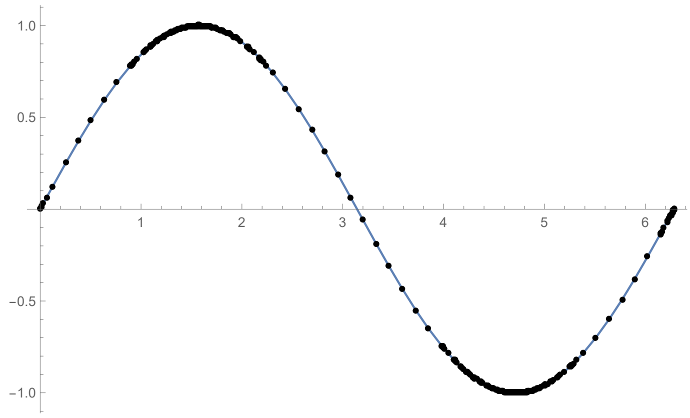


Figure 10.22: A graph of $y = \sin(x)$ generated by Mathematica.

to know the highest/lowest values of a stock, or the fastest/slowest an object was moving. Here, we apply the concepts of extreme values to solve problems stated in terms of situations that require us to create the appropriate mathematical framework in which to solve the problem.

We start with a classic example which is followed by a discussion of the topic of optimization.

Example 10.13

A man has 100 meters of fencing, a large yard, and a small dog. He wants to create a rectangular enclosure for his dog with the fencing that provides the maximal area. What dimensions provide the maximal area?

Solution

Drawing a rectangle forces us to realize that we need to know the dimensions of this rectangle so we can create an area function – after all, we are trying to maximize the area. We let x and y denote the lengths of the sides of the rectangle. Clearly, $\text{Area} = xy$.

We know more about the situation: the man has 100 metres of fencing. By knowing the perimeter of the rectangle must be 100, we can create another equation:

$$\text{Perimeter} = 100 = 2x + 2y.$$

We now have 2 equations and 2 unknowns. In the latter equation, we solve for y:

$$y = 50 - x.$$

Now substitute this expression for y in the area equation:

$$\text{Area} = A(x) = x(50 - x).$$

Note we now have an equation of one variable; we can truly call the Area a function of x.

This function only makes sense when $0 \le x \le 50$, otherwise we get negative values of area. So we find the extreme values of $A(x)$ on the interval $[0, 50]$.

To find the critical points, we take the derivative of $A(x)$ and set it equal to 0, then solve for x.

$$\begin{aligned} A(x) &= x(50 - x) \\ &= 50x - x^2 \\ A'(x) &= 50 - 2x \end{aligned}$$

We solve $50 - 2x = 0$ to find $x = 25$; this is the only critical point. We evaluate $A(x)$ at the endpoints of our interval and at this critical point to find the extreme values; in this case, all we care about is the maximum.

Clearly $A(0) = 0$ and $A(50) = 0$, whereas $A(25) = 625\text{m}^2$. This is the maximum. Since we earlier found $y = 50 - x$, we find that y is also 25. Thus the dimensions of the rectangular enclosure with perimeter of 100 m. with maximum area is a square, with sides of length 25 m.

This example is very simplistic and a bit contrived. (After all, most people create a design then buy fencing to meet their needs, and not buy fencing and plan later.) But it models well the necessary process: create equations that describe a situation, reduce an equation to a single variable, then find the needed extreme value.

"In real life," problems are much more complex. The equations are often *not* reducible to a single variable (hence multi–variable calculus is needed) and the equations themselves may be difficult to form. Understanding the principles here will provide a good foundation for the mathematics you will likely encounter later.

We outline here the basic process of solving these optimization problems.

1. Understand the problem. Clearly identify what quantity is to be maximized or minimized. Make a sketch if helpful.
2. Create equations relevant to the context of the problem, using the information given.
3. If the fundamental equation defines the quantity to be optimized as a function of more than one variable, reduce it to a single variable function using substitutions derived from the other equations.
4. Identify the domain of this function, keeping in mind the context of the problem.
5. Find the extreme values of this function on the determined domain.
6. Identify the values of all relevant quantities of the problem.

We will now follow these steps in the next example.

Example 10.14

Here is another classic calculus problem: A woman has a 100 m of fencing, a small dog, and a large yard that contains a stream (that is mostly straight). She wants to create a rectangular enclosure with maximal area that uses the stream as one side. What dimensions provide the maximal area?

Solution

We will follow the steps outlined earlier.

1. We are maximizing *area*. A sketch of the region will help; Figure 10.23 gives two sketches of the proposed enclosed area. A key feature of the sketches is to acknowledge that one side is not fenced.

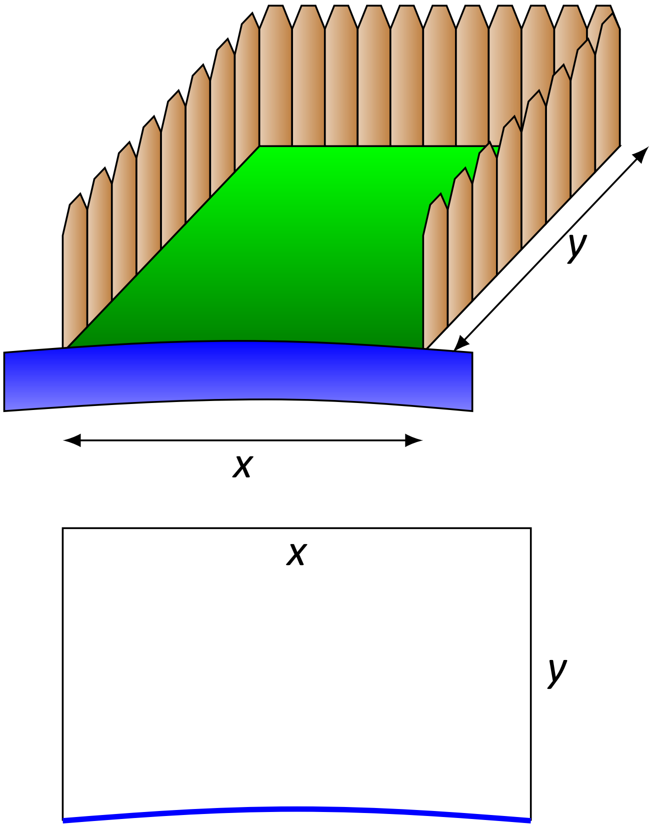


Figure 10.23: A sketch of the enclosure in Example 10.14.

2. We want to maximize the area. As in Example 10.13, we need another equation to reduce it to one variable.

   We again appeal to the perimeter; here the perimeter is

   $$\text{Perimeter} = 100 = x + 2y.$$

   Note how this is different than in our previous example.

3. We now reduce the fundamental equation to a single variable. In the perimeter equation, solve for y: $y = 50 - x/2$. We can now write Area as

   $$\text{Area} = A(x) = x(50 - x/2) = 50x - \frac{1}{2}x^2.$$

   Area is now defined as a function of one variable.

4. We want the area to be nonnegative. Since $A(x) = x(50 - x/2)$, we want $x \geq 0$ and $50 - x/2 \geq 0$. The latter inequality implies that $x \leq 100$, so $0 \leq x \leq 100$.

5. We now find the extreme values. At the endpoints, the minimum is found, giving an area of 0.

   Find the critical points. We have $A'(x) = 50 - x$; setting this equal to 0 and solving for x returns $x = 50$. This gives an area of

   $$A(50) = 50(25) = 1250.$$

6. We earlier set $y = 50 - x/2$; thus $y = 25$. Thus our rectangle will have two sides of length 25

and one side of length 50, with a total area of 1250 m^2.

Keep in mind as we do these problems that we are practicing a process; that is, we are learning to turn a situation into a system of equations. These equations allow us to write a certain quantity as a function of one variable, which we then optimize.

Example 10.15 is another classic calculus example, where we focus on minimizing costs.

Example 10.15

A power line needs to be run from a power station located on the beach to an offshore facility. Figure 10.24 shows the distances between the power station to the facility.

It costs €50/m. to run a power line along the land, and €130/m. to run a power line under water. How much of the power line should be run along the land to minimize the overall cost? What is the minimal cost?

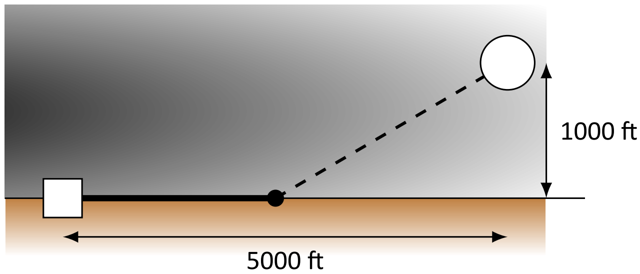


Figure 10.24: Running a power line from the power station to an offshore facility with minimal cost in Example 10.15.

Solution

There are two immediate solutions that we could consider, each of which we will reject through common sense. First, we could minimize the distance by directly connecting the two locations with a straight line. However, this requires that all the wire be laid underwater, the most costly option. Second, we could minimize the underwater length by running a wire all 5000 m. along the beach, directly across from the offshore facility. This has the undesired effect of having the longest distance of all, probably ensuring a non–minimal cost.

The optimal solution likely has the line being run along the ground for a while, then underwater, as the figure implies. We need to label our unknown distances – the distance run along the ground and the distance run underwater. Recognizing that the underwater distance can be measured as the hypotenuse of a right triangle, we choose to label the distance run along the ground as $5000-x$, so that the hypotenuse of the right triangle becomes $\sqrt{x^2+1000^2}$. We now create the cost function.

$$\begin{aligned}
\text{Cost} &= \text{land cost} &&+ \text{water cost}\\
&\text{€}50\times\text{land distance} &&+ \text{€}130\times\text{water distance}\\
&50(5000-x) &&+ 130\sqrt{x^2+1000^2}.
\end{aligned}$$

So we have $c(x) = 50(5000-x)+130\sqrt{x^2+1000^2}$. This function only makes sense on the interval $[0,5000]$. While we are fairly certain the endpoints will not give a minimal cost, we still evaluate $c(x)$ at each to verify.

$$c(0) = 380{,}000 \qquad c(5000) \approx 662{,}873.$$

We now find the critical values of $c(x)$. We compute $c'(x)$ as

$$c'(x) = -50 + \frac{130x}{\sqrt{x^2 + 1000^2}}.$$

Recognize that this is never undefined. Setting $c'(x) = 0$ and solving for x, we ultimate find that $x = 1250/3 \approx 416.67$. Evaluating $c(x)$ at $x = 416.67$ gives a cost of about €370,000. The distance the power line is laid along land is $5000 - 416.67 = 4583.33$ m., and the underwater distance is $\sqrt{416.67^2 + 1000^2} \approx 1083$ m.

In the exercises you will see a variety of situations that require you to combine problem–solving skills with calculus. Focus on learning how to form equations from situations that can be manipulated into what you need. Eschew memorizing how to do this kind of problem as opposed to that kind of problem. Learning a process will benefit one far longer than memorizing a specific technique.

In the next chapters, we will consider the reverse problem to computing the derivative: given a function f, can we find a function whose derivative is f? Being able to do so opens up an incredible world of mathematics and applications.

10.7 Exercises

Extreme values

Assignment 10.1 — Find the extremum of the function $y = x^{(x^2)}$. Prove that it is a minimum.

Assignment 10.2 — The Maxwell-Bolzmann distribution describes the distribution of velocities of gas molecules in an ideal gas. The probability that a molecule with mass m in a gas at temperature T, has velocity v is

$$f(v) = 4\pi\left(\frac{m}{2\pi kT}\right)^{3/2} v^2\, e^{-\frac{mv^2}{2kT}},$$

with k a constant. Determine the velocity v for which $f(v)$ is maximal.

The mean value theorem

Assignment 10.3 — For the functions listed below, verify whether the mean value theorem (theorem 10.4) can be applied to the given interval. Determine, if possible, a $c \in]a,b[$ that is guaranteed by the theorem.

(a) $f(x) = x^2 + 3x - 1, \qquad [-2, 2]$

(b) $f(x) = \sqrt{9 - x^2}, \qquad [0, 3]$

(c) $f(x) = \dfrac{x^2 - 9}{x^2 - 1}, \qquad [0, 2]$

Assignment 10.4 — For the functions listed below, verify that Rolle's theorem (stelling 10.5) can be applied to the given interval. Determine, if possible, a $c \in]a,b[$ such that $f'(c) = 0$.

(a) $f(x) = x^2 + x - 6, \qquad [-3, 2]$

(b) $f(x) = x^2 + x, \qquad [-2, 2]$

(c) $f(x) = \cos(x), \qquad [0, \pi]$

Curve sketching

Assignment 10.5 — Sketch the graph of the following functions.

(a) $f(x)=\sqrt{x^2-1}$

(b) $f(x)=\sqrt{x^2-4x+3}$

(c) $f(x)=\dfrac{x^3}{3-x^2}$

(d) $f(x)=\dfrac{x+7}{\sqrt{x^2-3}}$

(e) $f(x)=e^{-\frac{x^2}{2}}$

(f) $f(x)=x^3\,e^{-x}$

(g) $f(x)=\dfrac{e^{-x}}{x^3}$

(h) $f(x)=\ln(2^x-1)$

(i) $f(x)=\dfrac{\ln(x)}{x^2}$

(j) $f(x)=\ln\left(\sqrt{e^x+e^{-x}}\right)$

(k) $f(x)=\dfrac{2\ln(x)}{1-\ln(x)}$

(l) $f(x)=\ln\big(\cos(x)\big)$

(m) $f(x)=\arctan(\ln(x))$

(n) $f(x)=\arctan\left(\dfrac{1}{x}\right)$

(o) $f(x)=x^x$

(p) $f(x)=(x^2)^x$

(q) $f(x)=x-2\sin(x)$

(r) $f(x)=e^{-x}\sin(x),\quad (x\geq 0)$

(s) $f(x)=x+\sin(x)$

(t) $f(x)=\dfrac{|1+x|-1}{x}$

(u) $f(x)=\left|2-\sqrt{2x+4}\right|$

(v) $f(x)=\dfrac{x^2}{x|x|+1}$

(w) $f(x)=\left|(x-2)^2-4\right|$

Optimization

Assignment 10.6 — The town must construct a pipeline between point A and point B on both banks of a river (Figure 10.25). However, the cost of laying the pipeline underwater is three times that of laying it on land. Determine the point C where the pipeline must arrive at the opposite shore for the cost to be minimal.

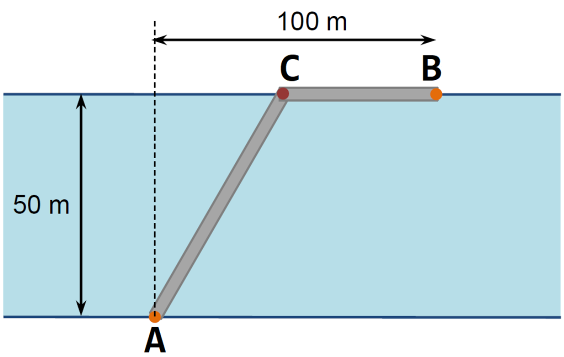


Figure 10.25: Schematic representation of the pipeline from Exercise 10.6.

Assignment 10.7 — A lead pipe is transported horizontally through a corridor that is 3 m wide (Figure 10.26). At the end of the corridor, there is a 90-degree turn and the corridor narrows to 2 m. What is the length of the largest tube that can be carried horizontally around the corner?

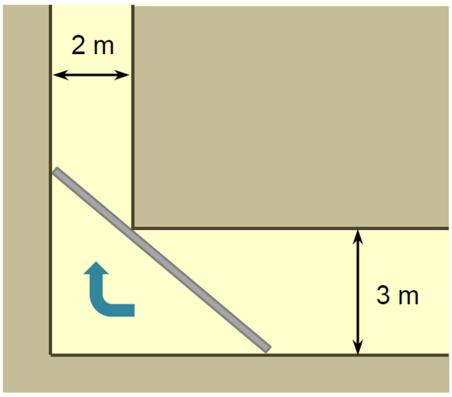


Figure 10.26: Schematic representation of the lead pipe from Exercise 10.7.

Assignment 10.8 — A cardboard box of capacity 0.5 l is formed from a piece of plasticized cardboard as shown in Figure 10.27. The base is a square with side B, the dotted lines indicate the fold lines and the gray strips are for gluing. Determine the height H and the base B to obtain 0.5 l as a volume with a minimum amount of cardboard.

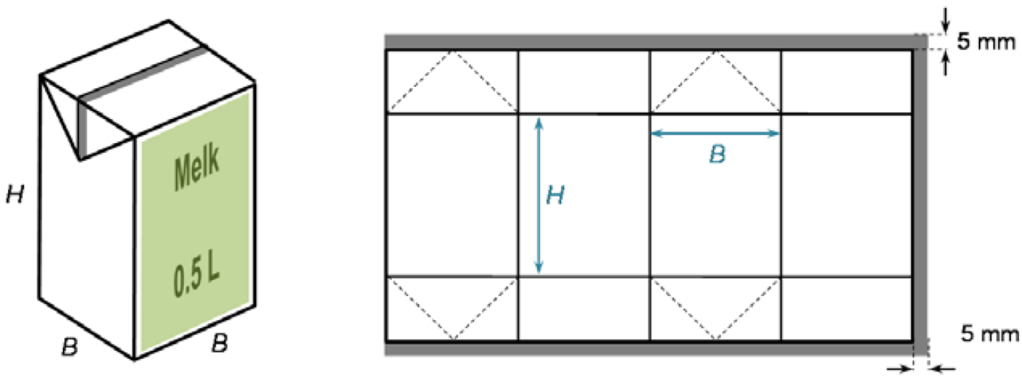


Figure 10.27: Schematic representation of the milk box from Exercise 10.8.

Assignment 10.9 — A conical solid of revolution lies completely inside a sphere of radius R cm. The top of the cone lies on the sphere and its axis is a diameter of the sphere. The volume of a cone is given by

$$V = \frac{\pi r^2 h}{3},$$

where r is the radius of the base and h is the height of the cone.

Calculate r and h such that the volume of the cone is maximal.

Review exercises

Assignment 10.10 — Figure 10.28 shows the graph of a function f, its derivative functions f' and f'' and a function g. Which graph corresponds to which function?

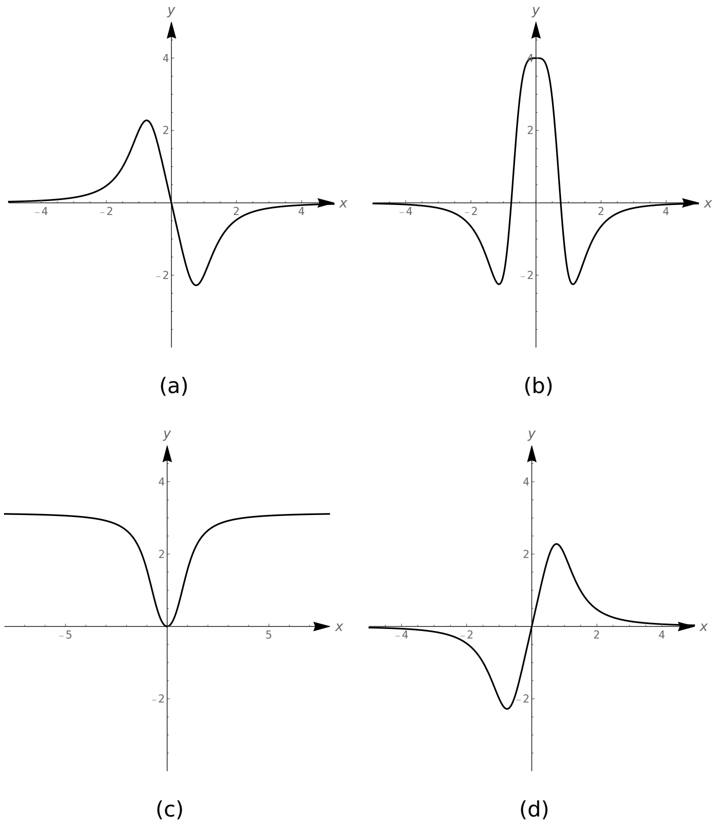


Figure 10.28: Figure belonging to Exercise 10.10

Assignment 10.11 — Figure 10.29 shows the graphs of four functions:

$$f(x) = \frac{x}{1-x^2}, \qquad g(x) = \frac{x^3}{1-x^4}, \qquad h(x) = \frac{x^3-x}{\sqrt{x^6+1}} \qquad \text{and} \qquad k(x) = \frac{x^3}{\sqrt{|x^4-1|}}.$$

Which graph corresponds to which function?

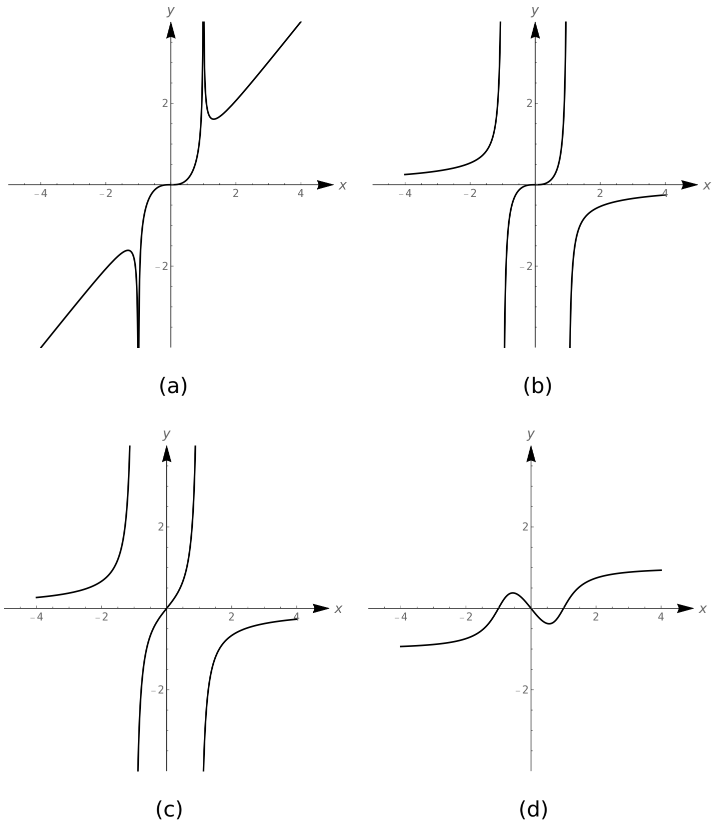


Figure 10.29: Figure corresponding to exercise 10.11.

Assignment 10.12 — For the functions below, if possible, determine the local maxima and/or minima and any inflection points. Also determine the intervals where the function is convex or concave.

(a) $f(x) = 2x^3 - 3x^2 + 9x + 5$

(b) $f(x) = (x-1)^2(x+2)$

(c) $f(x) = \dfrac{1}{x^2+1}$

(d) $f(x) = \sin(x) + \cos(x), \qquad x \in]-\pi, \pi[$

(e) $f(x) = x^2 \ln(x)$

At its heart, engineering is about using science to find creative, practical solutions. It is a noble profession.

— Queen Elizabeth II —

11

Polar coordinates and parametric equations

11.1 Polar coordinates

11.1.1 Definition

In Section 3.1, we introduced the Cartesian coordinates of a point in the plane as a means of assigning ordered pairs of numbers to points in the plane. We defined the Cartesian coordinate plane using two number lines – one horizontal and one vertical – which intersect at right angles at a point we called the origin. For this reason, the Cartesian coordinates of a point are often called **rectangular coordinates** (*rechthoekige coördinaten*) . In this section, we introduce a new system for assigning coordinates to points in the plane – **polar coordinates** (*poolcoördinaten*). We start with an origin point, called the **pole** (*pool*), and a ray called the **polar axis** (*poolas*). We then locate a point P using two coordinates, (r, θ), where r represents a directed distance from the pole and θ is a measure of rotation from the polar axis (Figure 11.1). Roughly speaking, the polar coordinates (r, θ) of a point measure how far out the point is from the pole (that is r), and how far to rotate from the polar axis, (that is θ).

For example, if we wish to plot the point P with polar coordinates $\left(4, \frac{5\pi}{6}\right)$, we would start at the pole, move out along the polar axis 4 units, then rotate $\frac{5\pi}{6}$ radians counter-clockwise. We may also consider this process by thinking of the rotation first. To plot $P\left(4, \frac{5\pi}{6}\right)$ this way, we rotate $\frac{5\pi}{6}$ counter-clockwise from the polar axis, then move outwards from the pole 4 units.

If $r < 0$, we begin by moving in the opposite direction on the polar axis from the pole and as you may have guessed, $\theta < 0$ means the rotation away from the polar axis is clockwise instead of counter-clockwise. Furthermore, it should come as no surprise that any given point expressed in polar coordinates has infinitely many other representations in polar coordinates. More formally, suppose (r, θ) and $\left(\tilde{r}, \tilde{\theta}\right)$ are polar coordinates where $r \neq 0$, $\tilde{r} \neq 0$ and the angles are measured in radians. Then (r, θ) and $\left(\tilde{r}, \tilde{\theta}\right)$ determine the same point P if and only if one of the following is true:

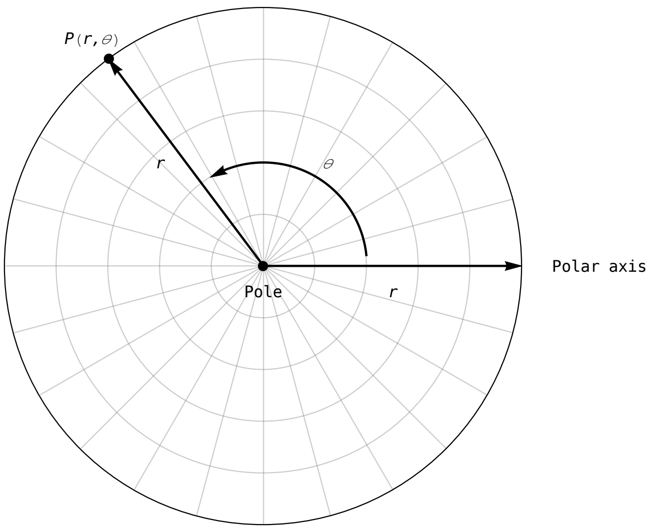


Figure 11.1: Polar coordinate system.

- $\tilde{r} = r$ and $\tilde{\theta} = \theta + 2\pi k$ for some integer k,
- $\tilde{r} = -r$ and $\tilde{\theta} = \theta + (2k+1)\pi$ for some integer k.

Moreover, all polar coordinates of the form $(0, \theta)$ represent the pole regardless of the value of θ.

Polar coordinates in aviation

Aircraft use a slightly modified version of the polar coordinates for navigation. In this system, the one generally used for any sort of navigation, the zero-degree ray is generally called heading 360, and the angles continue in a clockwise direction, rather than counterclockwise, as in the mathematical system. Heading 360 corresponds to magnetic north, while headings 90, 180, and 270 correspond to magnetic east, south, and west, respectively. Thus, an aircraft traveling 5 nautical miles due east will be traveling 5 units at heading 90.

11.1.2 Linking polar and rectangular coordinates

To marry the polar coordinate system with the Cartesian (rectangular) coordinate system we identify the pole and polar axis in the polar system with the origin and positive x-axis, respectively, in the rectangular system. We get the following result.

Theorem 11.1 (Conversion between rectangular and polar coordinates)
Suppose a point P is represented in rectangular coordinates as (x, y) and in polar coordinates as (r, θ). Then

- $x = r\cos(\theta)$ *and* $y = r\sin(\theta)$;
- $x^2 + y^2 = r^2$ *and* $\tan(\theta) = \dfrac{y}{x}$ *(provided $x \neq 0$)*.

In the case $r > 0$, Theorem 11.1 is an immediate consequence of Theorem 5.5 along with the definition of the tangent. If $r < 0$, then we know an alternate representation for (r, θ) is $(-r, \theta + \pi)$. Since we have that $\cos(\theta+\pi) = -\cos(\theta)$ and $\sin(\theta+\pi) = -\sin(\theta)$, applying the theorem to $(-r, \theta+\pi)$ gives

$$\begin{cases} x &= (-r)\cos(\theta+\pi) = (-r)(-\cos(\theta)) = r\cos(\theta) \\ y &= (-r)\sin(\theta+\pi) = (-r)(-\sin(\theta)) = r\sin(\theta). \end{cases}$$

Moreover, $x^2+y^2 = (-r)^2 = r^2$, and $\frac{y}{x} = \tan(\theta+\pi) = \tan(\theta)$, so the theorem is true in this case, too. The remaining case is $r = 0$, in which case $(r,\theta) = (0,\theta)$ is the pole. Since the pole is identified with the origin $(0,0)$ in rectangular coordinates, the theorem in this case amounts to checking $0 = 0$. The following example puts Theorem 11.1 to good use.

Example 11.1

Convert each point in rectangular coordinates given below into polar coordinates with $r \geq 0$ and $0 \leq \theta < 2\pi$.

1. $P\left(2, -2\sqrt{3}\right)$
2. $R(0, -3)$

Solution

1. The point $P\left(2, -2\sqrt{3}\right)$ lies in Quadrant IV. With $x = 2$ and $y = -2\sqrt{3}$, we get

$$r^2 = x^2+y^2 = (2)^2 + \left(-2\sqrt{3}\right)^2 = 4+12 = 16\,,$$

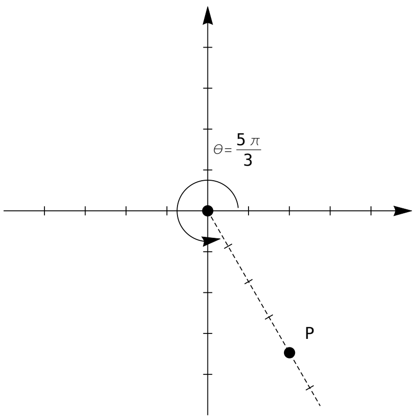
so $r = \pm 4$. Since we are asked for $r \geq 0$, we choose $r = 4$. To find θ, we have that

$$\tan(\theta) = \frac{y}{x} = \frac{-2\sqrt{3}}{2} = -\sqrt{3}\,.$$

This tells us θ has a reference angle of $-\frac{\pi}{3}$, and since P lies in Quadrant IV, we know θ is a Quadrant IV angle. We are asked to have $0 \leq \theta < 2\pi$, so we choose $\theta = \frac{5\pi}{3}$. Hence, our answer is $\left(4, \frac{5\pi}{3}\right)$ (Figure 11.2(a)).

2. The point $Q(0, -3)$ lies along the negative y-axis. While we could go through the usual computations to find the polar form of R, in this case we can find the polar coordinates of Q using the definition. Since the pole is identified with the origin, we can easily tell the point Q is 3 units from the pole, which means in the polar representation (r,θ) of Q we know $r = \pm 3$. Since we require $r \geq 0$, we choose $r = 3$. Concerning θ, the angle $\theta = \frac{3\pi}{2}$ satisfies $0 \leq \theta < 2\pi$ with its terminal side along the negative y-axis, so our answer is $\left(3, \frac{3\pi}{2}\right)$ (Figure 11.2(b)).

From the previous example, it is clear that it is important to know in which quadrant the point under investigation lies in order to infer the corresponding θ. Instead of using the arctan-function for that purpose and then figure out the correct angle, it is often more useful to use the atan2-function (2-argument tangent). The atan2 is defined as the angle in the Euclidean plane, given in radians, between the positive x-axis and the ray to the point $(x,y) \neq (0,0)$. The angles are signed, with counter-clockwise angles being positive, and clockwise ones being negative. In other words, $\mathrm{atan2}(y,x)$ is in the interval



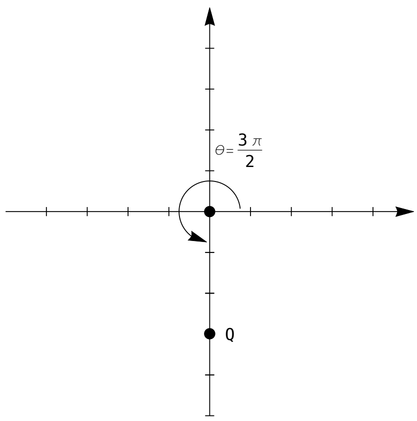


Figure 11.2: The location of the point P with rectangular coordinates $\left(2,-2\sqrt{3}\right)$ and polar coordinates $\left(4, \frac{5\pi}{3}\right)$ (a) and the point Q with rectangular coordinates $(0,-3)$ and polar coordinates $\left(3\sqrt{2}, \frac{5\pi}{4}\right)$ (b).

$[0,\pi]$ when $y > 0$ and in $]-\pi,0[$ when $y < 0$. The function is defined as:

$$\operatorname{atan2}(y,x) = \begin{cases} \arctan\left(\dfrac{y}{x}\right) & , \quad \text{if } x > 0 \\ \arctan\left(\dfrac{y}{x}\right) + \pi & , \quad \text{if } x < 0 \wedge y \geq 0 \\ \arctan\left(\dfrac{y}{x}\right) - \pi & , \quad \text{if } x < 0 \wedge y < 0 \\ \dfrac{\pi}{2} & , \quad \text{if } x = 0 \wedge y > 0 \\ -\dfrac{\pi}{2} & , \quad \text{if } x = 0 \wedge y < 0 \\ \text{undefined} & , \quad \text{if } x = 0 \wedge y = 0 \end{cases} \tag{11.1}$$

Of course, we do not have to restrict to points when converting from the rectangular to the polar coordinate system. We can do the same with equations using Theorem 11.1. In polar coordinates, we will end up with equations in the variables r and θ. The obvious strategy to convert an equation from rectangular to polar coordinates is to replace every occurrence of x with $r\cos(\theta)$ and every occurrence of y with $r\sin(\theta)$ and use identities to simplify. On the other hand, converting equations from polar to rectangular coordinates is not as straightforward. We could solve $r^2 = x^2 + y^2$ for r to get $r = \pm\sqrt{x^2+y^2}$ and solving $\tan(\theta) = \frac{y}{x}$ requires the arctangent function to get $\theta = \arctan\left(\frac{y}{x}\right) + \pi k$ for integers k. Still, since neither of these expressions for r and θ are especially user-friendly, we might resort to a second strategy involving the rearrangement of the given polar equation so that the expressions $r^2 = x^2 + y^2$, $r\cos(\theta) = x$, $r\sin(\theta) = y$ and/or $\tan(\theta) = \frac{y}{x}$ present themselves.

Example 11.2

1. Convert each equation in rectangular coordinates into an equation in polar coordinates.

   (a) $(x-3)^2 + y^2 = 9$

   (b) $y = -x$

2. Convert each equation in polar coordinates into an equation in rectangular coordinates.

   (a) $r = -3$

(b) $r = 1 - \cos(\theta)$

Solution

1. (a) We start by substituting $x = r\cos(\theta)$ and $y = r\sin(\theta)$ into $(x-3)^2 + y^2 = 9$ and then simplify. With no real direction in which to proceed, we follow our mathematical instincts and see where they take us.

$$\begin{array}{rrcll}
 & (r\cos(\theta)-3)^2 + (r\sin(\theta))^2 & = & 9 & \\
\Leftrightarrow & r^2\cos^2(\theta) - 6r\cos(\theta) + 9 + r^2\sin^2(\theta) & = & 9 & \\
\Leftrightarrow & r^2\left(\cos^2(\theta)+\sin^2(\theta)\right) - 6r\cos(\theta) & = & 0 & \text{(Subtract 9 from both sides.)} \\
\Leftrightarrow & r^2 - 6r\cos(\theta) & = & 0 & \text{(Since } \cos^2(\theta)+\sin^2(\theta)=1.\text{)} \\
\Leftrightarrow & r(r - 6\cos(\theta)) & = & 0 & \text{(Factor.)}
\end{array}$$

We get $r = 0$ or $r = 6\cos(\theta)$. From Section 4.4 we know that the equation $(x-3)^2 + y^2 = 9$ describes a circle, and since $r = 0$ describes just a point (namely the pole/origin), we choose $r = 6\cos(\theta)$ for our final answer.

(b) Substituting $x = r\cos(\theta)$ and $y = r\sin(\theta)$ into $y = -x$ gives $r\sin(\theta) = -r\cos(\theta)$. Rearranging, we get $r\cos(\theta) + r\sin(\theta) = 0$ or $r(\cos(\theta) + \sin(\theta)) = 0$. This gives $r = 0$ or $\cos(\theta) + \sin(\theta) = 0$. Solving the latter equation for θ, we get $\theta = -\frac{\pi}{4} + \pi k$ for integers k. We know $y = -x$ describes a line through the origin. As before, $r = 0$ describes the origin, but nothing else. Consider the equation $\theta = -\frac{\pi}{4}$. In this equation, the variable r is free, meaning it can assume any and all values including $r = 0$. If we imagine plotting points $(r, -\frac{\pi}{4})$ for all conceivable values of r (positive, negative and zero), we are essentially drawing the line containing the terminal side of $\theta = -\frac{\pi}{4}$ which is none other than $y = -x$. Hence, we can take as our final answer $\theta = -\frac{\pi}{4}$ here.

2. (a) Starting with $r = -3$, we can square both sides to get $r^2 = (-3)^2$ or $r^2 = 9$. We may now substitute $r^2 = x^2 + y^2$ to get the equation $x^2 + y^2 = 9$. As we have seen, squaring an equation does not, in general, produce an equivalent equation. The concern here is that the equation $r^2 = 9$ might be satisfied by more points than $r = -3$. On the surface, this appears to be the case since $r^2 = 9$ is equivalent to $r = \pm 3$, and not just $r = -3$. However, any point with polar coordinates $(3, \theta)$ can be represented as $(-3, \theta + \pi)$, which means any point (r, θ) whose polar coordinates satisfy the relation $r = \pm 3$ has an equivalent representation – meaning that they represent the same point in the plane – which satisfies $r = -3$.

(b) Once again, we need to manipulate $r = 1 - \cos(\theta)$ a bit before using the conversion formulas given in Theorem 11.1. We could square both sides of this equation to obtain an r^2 on the left hand side, but that does not result in something helpful for the right hand side. Instead, we multiply both sides by r to obtain $r^2 = r - r\cos(\theta)$. We now have an r^2 and an $r\cos(\theta)$ in the equation, which we can easily handle, but we also have another r to deal with. Rewriting the equation as $r = r^2 + r\cos(\theta)$ and squaring both sides yields $r^2 = \left(r^2 + r\cos(\theta)\right)^2$. Substituting $r^2 = x^2 + y^2$ and $r\cos(\theta) = x$ gives $x^2 + y^2 = \left(x^2 + y^2 + x\right)^2$. Once again, we have performed some algebraic manoeuvres which may have altered the set of points described by the original equation. First, we multiplied both sides by r. This means that now $r = 0$ is a viable solution to the equation. In the original equation, $r = 1 - \cos(\theta)$, we see that $\theta = 0$ gives $r = 0$, so the multiplication by r does not introduce any new points.

The squaring of both sides of this equation is also a reason to pause. Are there points with coordinates (r,θ) which satisfy $r^2 = \left(r^2 + r\cos(\theta)\right)^2$ but do not satisfy $r = r^2 + r\cos(\theta)$? Suppose $\left(r', \theta'\right)$ satisfies $r^2 = \left(r^2 + r\cos(\theta)\right)^2$. Then $r' = \pm\left((r')^2 + r'\cos(\theta')\right)$. If we have that $r' = (r')^2 + r'\cos(\theta')$, we are done. What if $r' = -\left((r')^2 + r'\cos(\theta')\right) = -(r')^2 - r'\cos(\theta')$? We claim that the coordinates $(-r', \theta' + \pi)$, which determine the same point as (r', θ'), satisfy $r = r^2 + r\cos(\theta)$. We substitute $r = -r'$ and $\theta = \theta' + \pi$ into $r = r^2 + r\cos(\theta)$ to see if we get a true statement.

$$\begin{array}{rrcll}
& -r' & \stackrel{?}{=} & \left(-r'\right)^2 + \left(-r'\cos(\theta' + \pi)\right) & \\
\Leftrightarrow & -\left(-(r')^2 - r'\cos(\theta')\right) & \stackrel{?}{=} & (r')^2 - r'\cos(\theta' + \pi) & \text{(Since } r' = -(r')^2 - r'\cos(\theta').\text{)} \\
\Leftrightarrow & (r')^2 + r'\cos(\theta') & \stackrel{?}{=} & (r')^2 - r'(-\cos(\theta')) & \text{(Since } \cos(\theta' + \pi) = -\cos(\theta').\text{)} \\
\Leftrightarrow & (r')^2 + r'\cos(\theta') & \stackrel{\checkmark}{=} & (r')^2 + r'\cos(\theta') &
\end{array}$$

Since both sides worked out to be equal, $(-r', \theta' + \pi)$ satisfies $r = r^2 + r\cos(\theta)$, which means that any point (r,θ) that satisfies $r^2 = \left(r^2 + r\cos(\theta)\right)^2$ has a representation which satisfies $r = r^2 + r\cos(\theta)$, and we are done.

In practice, much of the pedantic verification of the equivalence of equations in Example 11.2 is left unsaid. Indeed, in most textbooks, squaring equations like $r = -3$ to arrive at $r^2 = 9$ happens without a second thought. If you take anything away from Example 11.2, it should be that relatively nice things in rectangular coordinates, such as $y = x^2$, can turn ugly in polar coordinates, and vice-versa.

11.1.3 Graphs of polar equations

Having introduced polar coordinates and equations expressed therein, we now discuss how to graph equations in polar coordinates on the rectangular coordinate plane. Since any given point in the plane has infinitely many different representations in polar coordinates, we have the following fundamental graphing principle.

Definitie 11.1 (The fundamental graphing principle for polar equations)
The graph of an equation in polar coordinates is the set of points that satisfy the equation. That is, a point $P(r,\theta)$ is on the graph of an equation if and only if there is a representation of P, say $\left(\tilde{r}, \tilde{\theta}\right)$, such that $\tilde{r}$ and $\tilde{\theta}$ satisfy the equation.

Our first example focuses on some of the more structurally simple polar equations.

Example 11.3

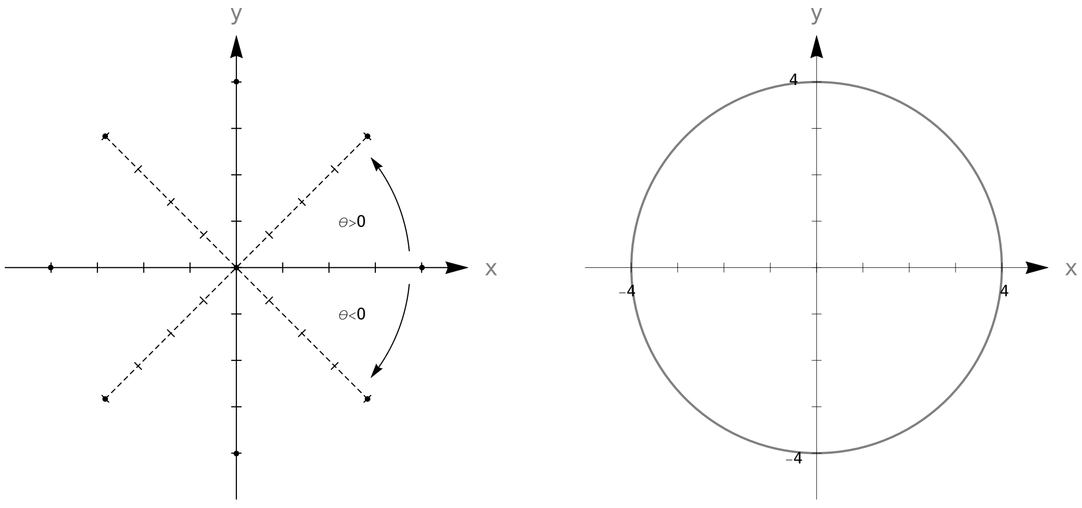
Graph the following polar equations.

1. $r = 4$
2. $\theta = -\dfrac{3\pi}{2}$

Solution

1. In the r equation, θ is free. Its graph is, therefore, all points which have a polar coordinate representation $(4,\theta)$, for any choice of θ (Figure 11.3(a)). Graphically this translates into

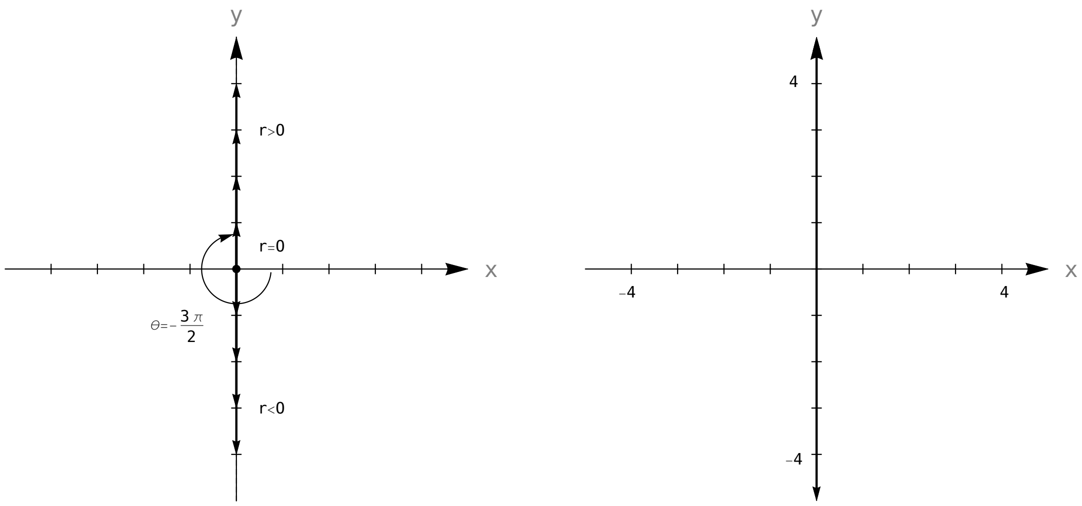
tracing out all of the points 4 units away from the origin. This is exactly the definition of circle, centred at the origin, with a radius of 4 (Figure 11.3(b)).



(a) In $r = 4$, θ is free. (b) The graph of $r = 4$.

Figure 11.3: Constructing the graph of $r = 4$.

2. Here, the variable r is free (Figure 11.4(a)). Plotting $\left(r, -\frac{3\pi}{2}\right)$ for various values of r shows us that we are tracing out the y-axis (Figure 11.4(b)).



(a) In $\theta = -\frac{3\pi}{2}$, r is free. (b) The graph of $\theta = -\frac{3\pi}{2}$.

Figure 11.4: Constructing the graph of $\theta = -\frac{3\pi}{2}$.

Our experience in Example 11.3 makes the following clear.

- The graph of the polar equation $r = \alpha$ on the Cartesian plane is a circle centred at the origin of radius $|\alpha|$.
- The graph of the polar equation $\theta = \alpha$ on the Cartesian plane is the line containing the terminal side of α when plotted in standard position.

Since it gets way more involved to construct the graphs of generic polar equations, we will resort to Mathematica for that purpose. This programme also allows us to check analytically, for instance, intersection of the graphs of multiple polar equations. More specifically, we can rely on the built-in function `PolarPlot` to construct the graph of polar equations.

Example 11.4

Graph the following polar equations.

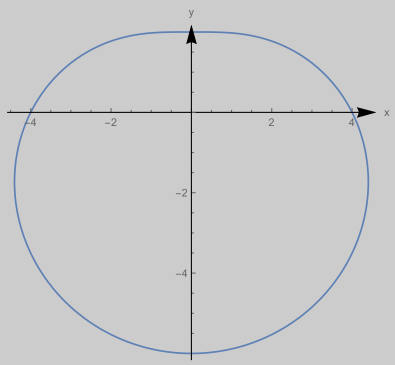
1. $r = 4 - 2\sin(\theta)$
2. $r^2 = 16\cos(2\theta)$

Solution

1. We proceed in Mathematica using the following instruction, where we chose to indicate the axis labels and directions.

```
In[16]:= PolarPlot[4-2 Sin[theta],{theta,0,2*Pi}, AxesLabel→{"x","y"},
AxesStyle→Arrowheads[{0,0.05}]]
```

Out[16]=



Note that we chose θ to vary between 0 and 2π because the period of the concerned polar equation is 2π.

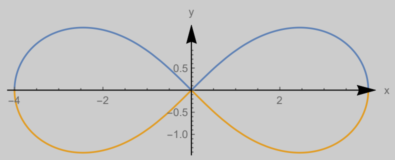
2. Graphing $r^2 = 16\cos(2\theta)$ is complicated by the r^2, so we solve to get

$$r = \pm\sqrt{16\cos(2\theta)} = \pm 4\sqrt{\cos(2\theta)}.$$

Since the period of the involved cosine is π, we plot both functions together using `PolarPlot` for $\theta \in [0, \pi]$.

```
In[17]:= PolarPlot[{Sqrt[16 Cos[2* theta]],- Sqrt[16 Cos[2*theta]]},{theta,0,Pi},
AxesLabel→{"x","y"},AxesStyle→Arrowheads[{0,0.05}]]
```

Out[17]=



Also note that we may plot values of θ outside of the interval $[0, \pi]$, but then we will find ourselves retracing parts of the curve we already had obtained.

The previous example makes us appreciate the symmetry that is a common occurrence when graphing polar equations. Indeed, it can be verified easily that $r = f(\theta)$ is symmetric about the x-axis if f is even because then we have that $f(\theta) = f(-\theta)$ (e.g. Example 11.4.2), symmetric about the y-axis if $f(\pi - \theta) = f(\theta)$ (e.g. Example 11.4.1 and 2) and symmetric about the origin if f is odd because then we have that $f(-\theta) = -f(\theta)$ (e.g. Example 11.4.2). In addition these usual kinds of symmetry, it is possible to talk about rotational symmetry. More specifically, if $f(\theta - \alpha) = f(\theta)$ it will be rotationally symmetric by α clockwise and counter-clockwise about the pole.

In our next example, we are given the task of finding the intersection points of polar curves.

Example 11.5

Find the points of intersection of the graphs of the following polar equations.

1. $r = 2\sin(\theta)$ and $r = 2 - 2\sin(\theta)$
2. $r = 3$ and $r = 6\cos(2\theta)$

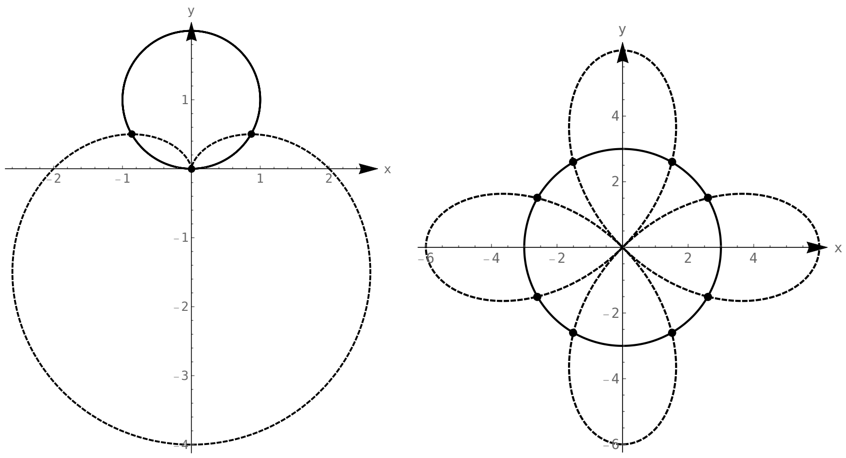
Solution

1. We first try to see if we can find any points which have a single representation $P(r,\theta)$ that satisfies both equations. Assuming such a pair (r,θ) exists, then equating the expressions for r gives

$$\begin{aligned} 2\sin(\theta) = 2 - 2\sin(\theta) \quad &\Leftrightarrow \quad \sin(\theta) = \frac{1}{2} \\ &\Leftrightarrow \quad \theta = \frac{\pi}{6} + 2\pi k \text{ or } \theta = \frac{5\pi}{6} + 2\pi k \end{aligned}$$

   for integers k. Plugging $\theta = \frac{\pi}{6}$ into $r = 2\sin(\theta)$, we get $r = 2\sin\left(\frac{\pi}{6}\right) = 2\left(\frac{1}{2}\right) = 1$, which is also the value we obtain when we substitute it into $r = 2 - 2\sin(\theta)$. Hence, $\left(1, \frac{\pi}{6}\right)$ is one representation for the point of intersection in quadrant I. For the point of intersection in Quadrant II, we try $\theta = \frac{5\pi}{6}$. Both equations give us the point $\left(1, \frac{5\pi}{6}\right)$, so this is our answer here. Now, let us check graphically whether we indeed found all intersection points of the graphs of the involved polar equations. (Figure 11.5(a)).

   From this graph it appears that there are three intersection points: one in Quadrant I, one in Quadrant II, and the origin. So, how to find the latter algebraically? We know that the pole may be represented as $(0,\theta)$ for any angle θ. On the graph of $r = 2\sin(\theta)$, we start at the origin when $\theta = 0$ and return to it at $\theta = \pi$. Actually, we are at the origin exactly when $\theta = \pi k$ for integers k. On the curve $r = 2 - 2\sin(\theta)$, however, we reach the origin when $\theta = \frac{\pi}{2}$, and more generally, when $\theta = \frac{\pi}{2} + 2\pi k$ for integers k. There is no integer value of k for which $\pi k = \frac{\pi}{2} + 2\pi k$, which means while the origin is on both graphs, the point is never reached simultaneously. In any case, we have determined the three points of intersection to be $\left(1, \frac{\pi}{6}\right)$, $\left(1, \frac{5\pi}{6}\right)$ and the origin.



(a) $r = 2\sin(\theta)$ (solid) and $r = 2 - 2\sin(\theta)$ (dashed).

(b) $r = 3$ (solid) and $r = 6\cos(2\theta)$ (dashed).

Figure 11.5: Points of intersection of the graphs of two polar equations.

2. Let us graph the equations to get an idea of how many intersection points to expect and where they lie. The graph of $r = 3$ is a circle centred at the origin with a radius of 3 and the graph of $r = 6\cos(2\theta)$ is a four-leafed rose (Figure 11.5(b)).

   It appears as if there are eight points of intersection - two in each quadrant. We first look to see if there any points $P(r, \theta)$ with a representation that satisfies both equations. For these points,

$$\begin{aligned} 6\cos(2\theta) = 3 \quad &\Leftrightarrow \quad \cos(2\theta) = \frac{1}{2} \\ &\Leftrightarrow \quad \theta = \frac{\pi}{6} + \pi k \text{ or } \theta = \frac{5\pi}{6} + \pi k \end{aligned}$$

   for integers k. Out of all of these solutions, we obtain just four distinct points represented by $\left(3, \frac{\pi}{6}\right)$, $\left(3, \frac{5\pi}{6}\right)$, $\left(3, \frac{7\pi}{6}\right)$ and $\left(3, \frac{11\pi}{6}\right)$. To determine the coordinates of the remaining four points, we have to consider how the representations of the points of intersection can differ. We know from the beginning of this section that if (r, θ) and (r', θ') represent the same point and $r \neq 0$, then either $r = r'$ or $r = -r'$. If $r = r'$, then $\theta' = \theta + 2\pi k$, so one possibility is that an intersection point P has a representation (r, θ) which satisfies $r = 3$ and another representation $(r, \theta + 2\pi k)$ for some integer k, which satisfies $r = 6\cos(2\theta)$. At this point, if we replace every occurrence of θ in the equation $r = 6\cos(2\theta)$ with $(\theta + 2\pi k)$ and then see if, by equating the resulting expressions for r, we get any more solutions for θ. Since $\cos(2(\theta + 2\pi k)) = \cos(2\theta + 4\pi k) = \cos(2\theta)$ for every integer k, however, the equation $r = 6\cos(2(\theta + 2\pi k))$ reduces to the same equation we had before, $r = 6\cos(2\theta)$, which means we get no additional solutions.
   Moving on to the case where $r = -r'$, we have that $\theta' = \theta + (2k+1)\pi$ for integers k. We look to see if we can find points P which have a representation (r, θ) that satisfies $r = 3$ and another, $(-r, \theta + (2k+1)\pi)$, that satisfies $r = 6\cos(2\theta)$. To do this, we substitute $(-r)$ for r and $(\theta + (2k+1)\pi)$ for θ in the equation $r = 6\cos(2\theta)$ and get $-r = 6\cos(2(\theta + (2k+1)\pi))$. Since $\cos(2(\theta + (2k+1)\pi)) = \cos(2\theta + (2k+1)(2\pi)) = \cos(2\theta)$ for all integers k, the equation $-r = 6\cos(2(\theta + (2k+1)\pi))$ reduces to $-r = 6\cos(2\theta)$, or $r = -6\cos(2\theta)$. Coupling this equation with $r = 3$ gives

$$-6\cos(2\theta) = 3 \quad \Leftrightarrow \quad \cos(2\theta) = -\frac{1}{2}$$

$$\Leftrightarrow \quad \theta = \frac{\pi}{3} + \pi k \text{ or } \theta = \frac{2\pi}{3} + \pi k$$

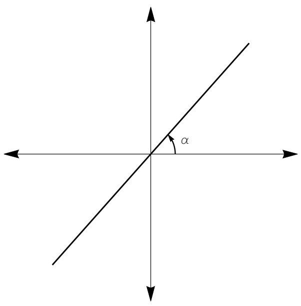
From these solutions, we obtain the remaining four intersection points with representations $\left(-3, \frac{\pi}{3}\right)$, $\left(-3, \frac{2\pi}{3}\right)$, $\left(-3, \frac{4\pi}{3}\right)$ and $\left(-3, \frac{5\pi}{3}\right)$

There are a number of basic and classic polar curves, famous for their beauty and/or applicability in science. For that reason, this section ends with a small gallery of some of these graphs.

Lines

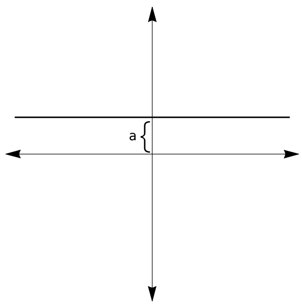
Through the origin:

$\theta = \alpha$



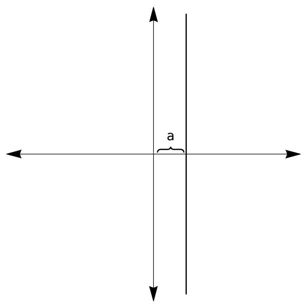
Horizontal line:

$r = a\csc(\theta)$



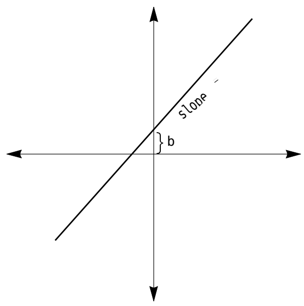
Vertical line:

$r = a\sec(\theta)$



Not through origin:

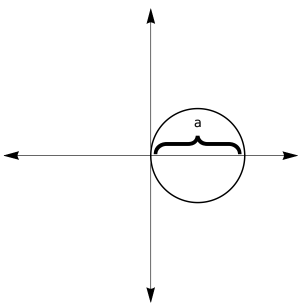
$$r = \frac{b}{\sin(\theta) - m\cos(\theta)}$$



Circles

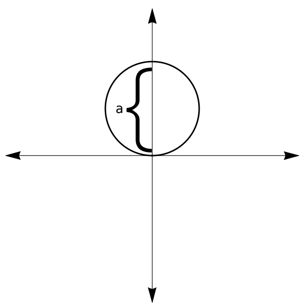
Centred on x-axis:

$r = a\cos(\theta)$



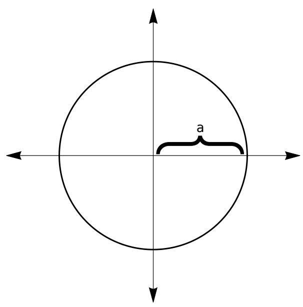
Centred on y-axis:

$r = a\sin(\theta)$



Centred on origin:

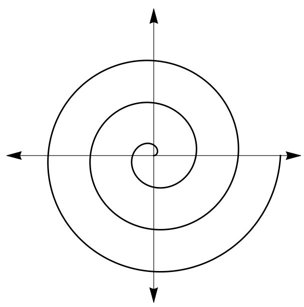
$r = a$



Spiral

Archimedean spiral

$r = \theta$

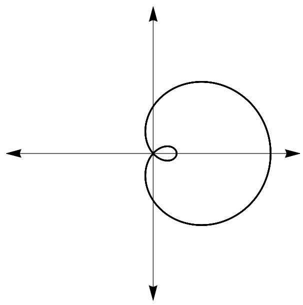


Limaçons

Symmetric about x-axis: $r = a \pm b\cos(\theta)$; Symmetric about y-axis: $r = a \pm b\sin(\theta)$; $a, b > 0$

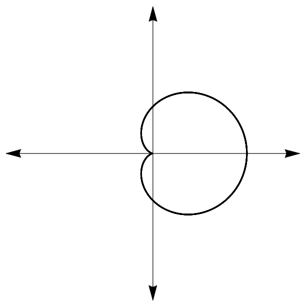
With inner loop:

$\frac{a}{b} < 1$



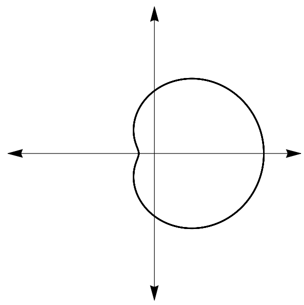
Cardioid:

$\frac{a}{b} = 1$



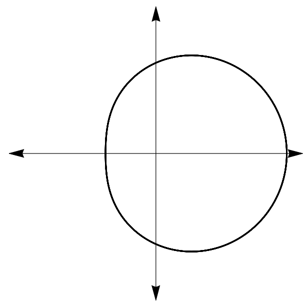
Dimpled:

$1 < \frac{a}{b} < 2$



Convex:

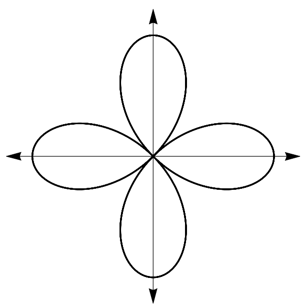
$\frac{a}{b} > 2$



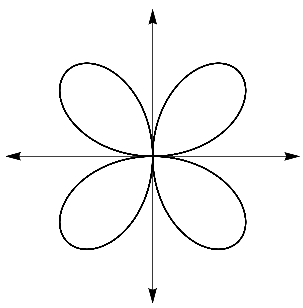
Rose Curves

Symmetric about x-axis: $r = a\cos(n\theta)$; Symmetric about y-axis: $r = a\sin(n\theta)$
Curve contains $2n$ petals when n is even and n petals when n is odd.

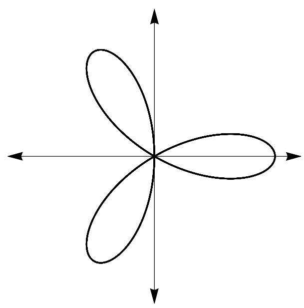
$r = a\cos(2\theta)$



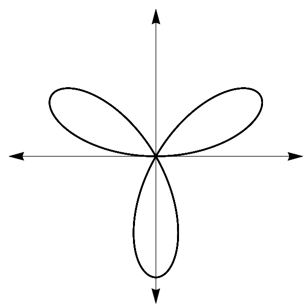
$r = a\sin(2\theta)$



$r = a\cos(3\theta)$



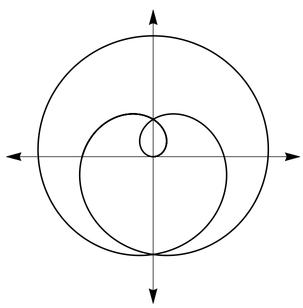
$r = a\sin(3\theta)$



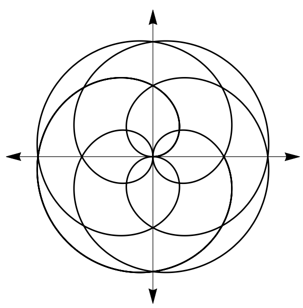
Special Curves

Rose curves

$r = a\sin(\theta/5)$

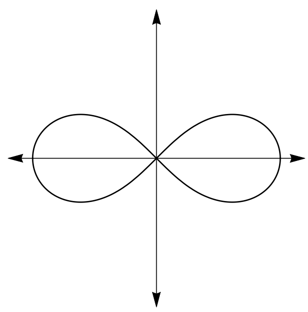


$r = a\sin(2\theta/5)$



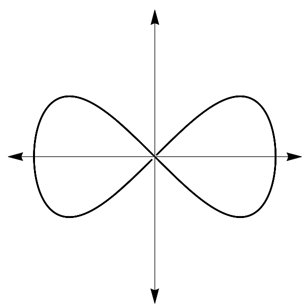
Lemniscate:

$r^2 = a^2\cos(2\theta)$



Eight Curve:

$r^2 = a^2\sec^4(\theta)\cos(2\theta)$



11.2 Parametric equations

11.2.1 Definition

As we have seen in Section 11.1 there are interesting curves which, when plotted in the xy-plane, neither represent y as a function of x nor x as a function of y. Here, we present a new concept which allows us to use functions to study these kinds of curves. To motivate the idea, we imagine a bug crawling across a table top starting at the point O and tracing out a curve C in the plane, as shown in Figure 11.6.

The curve C does not represent y as a function of x because it fails the vertical line test and it does not represent x as a function of y because it fails the horizontal line test. However, since the bug can be in only one place $P(x, y)$ at any given time t, we can define the x-coordinate of P as a function of t and the y-coordinate of P as a different function of t. The independent variable t in this case is called a **parameter** (*parameter*) and the system of equations

$$\begin{cases} x = f(t) \\ y = g(t) \end{cases}$$

is called a **system of parametric equations** (*stelsel parametervergelijkingen*) or a **parametrization** (*parametervoorstelling*) of the curve C.

The parametrization of C endows it with an **orientation** (*zin*) and the arrows on C indicate motion in the direction of increasing values of t. In this case, our bug starts at the point O, travels upwards to the left, then loops back around to cross its path at the point Q and finally heads off into the first quadrant. It is important to note that the curve itself is a set of points and as such is devoid of any orientation. It is the parametrization that determines the orientation and as we shall see, different parametrizations can determine different orientations. Actually, the system of equations

$$\begin{cases} x = \cos(t) \\ y = \sin(t) \end{cases}$$

parametrizes the unit circle, giving it a counter-clockwise orientation. More generally, the equations of circular motion (Theorem 5.5)

$$\begin{cases} x = r\cos(\omega t) \\ y = r\sin(\omega t) \end{cases}$$

are parametric equations that trace out a circle of radius r centred at the origin. If $\omega > 0$, the orientation is counter-clockwise; if $\omega < 0$, the orientation is clockwise. The angular frequency ω determines how fast the object moves around the circle.

11.2.2 Graphing parametric equations

Graphing parametric equations is pretty straightforward in the sense that we just choose some friendly values of t, plot the corresponding points and connect the results in a pleasing fashion. For instance, consider the following system of parametric equations:

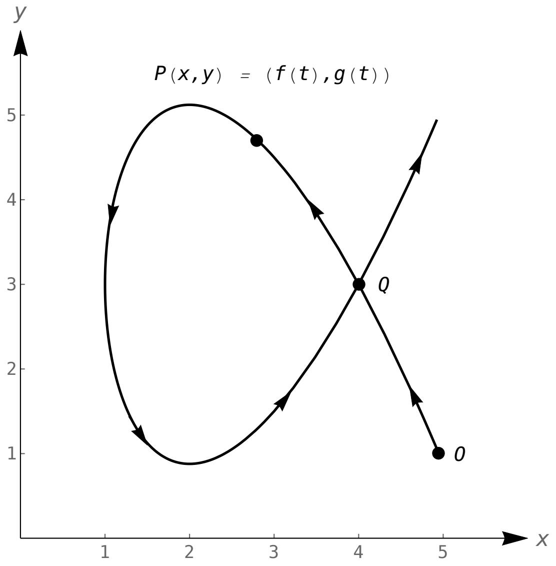


Figure 11.6: A bug crawling tracing out a curve C in the plane.

$$\begin{cases} x = t^2 - 3 \\ y = 2t - 1 \end{cases} \tag{11.2}$$

for $t \geq -2$.

Since we are told $t \geq -2$, we start there and evaluate $x(t)$ and $y(t)$, yielding the following values:

t	$x(t)$	$y(t)$
-2	1	-5
-1	-2	-3
0	-3	-1
1	-2	1
2	1	3
3	6	5

Then we plot the successive points in Figure 11.7 and we draw an arrow to indicate the direction of the path for increasing values of t. The curve looks like a parabola. To verify this we may eliminate the parameter t from Equation (11.2). To do so, we choose to solve the equation $y = 2t - 1$ for t to get $t = \frac{y+1}{2}$. Substituting this into the equation $x = t^2 - 3$ yields

$$x = \left(\frac{y+1}{2}\right)^2 - 3$$

or, after some rearrangement,

$$(y+1)^2 = 4(x+3)\,. \tag{11.3}$$

The graph of this equation is a parabola with vertex $(-3,-1)$ which opens to the right. Technically speaking, Equation (11.3) describes the entire parabola, while the parametric equations (Equation (11.2)) for $t \geq -2$ describe only a portion of the parabola. In this case, we can remedy this situation by restricting the bounds on y. Since the portion of the parabola we want is exactly the part where $y \geq -5$, Equation (11.3) coupled with the restriction $y \geq -5$ describes the same curve as the given parametric equations. The one piece of information we can never recover after eliminating the parameter is the orientation of the curve.

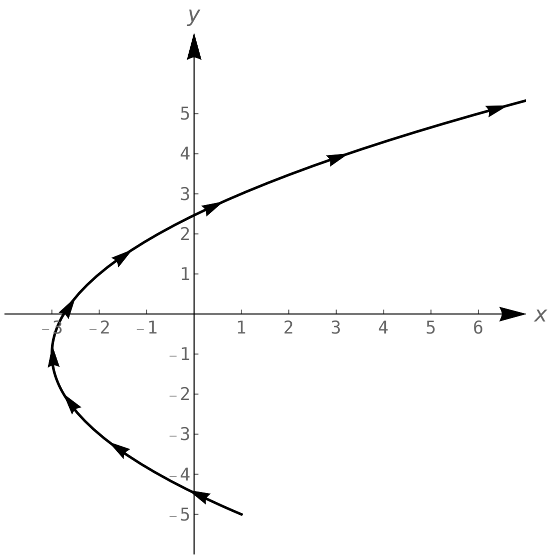


Figure 11.7: The curve described by Equation (11.2).

In Mathematica, we can use the built-in function **ParametricPlot** to construct the graph of a parametric equation. For instance, for Equation (11.2) this can be achieved as follows.

```
In[18]:= ParametricPlot[{t^2-3,2* t-1},{t,-2,3.2},AxesLabel→{"x","y"},
         AxesStyle→Arrowheads[{0,0.05}]]
```

Example 11.6

Sketch the curves described by the following parametric equations.

1. For $-1 \leq t \leq 1$
$$\begin{cases} x &= t^3 \\ y &= 2t^2, \end{cases}$$

2. For $0 \leq t \leq \dfrac{3\pi}{2}$
$$\begin{cases} x &= 1 + 3\cos(t) \\ y &= 2\sin(t), \end{cases}$$

Solution

1. We first sketch the graphs of $x = t^3$ and $y = 2t^2$ over the interval $[-1, 1]$ (Figures 11.8(a) and 11.8(b)). We note that as t takes on values in $[-1, 1]$, $x = t^3$ ranges between -1 and 1, and $y = 2t^2$ ranges between 0 and 2. This means that all of the action is happening on a portion of the plane, namely $\{(x, y) \mid -1 \leq x \leq 1,\ 0 \leq y \leq 2\}$. Next, we plot a few points. Certainly, $t = -1$ and $t = 1$ are good values to pick since these are the extreme values of t. We also choose $t = 0$ as this is a relative minimum on the graph of $y = 2t^2$. Plugging in $t = -1$ gives the point $(-1, 2)$, $t = 0$ gives $(0, 0)$ and $t = 1$ gives $(1, 2)$. More generally, we see that $x = t^3$ is increasing over the entire interval $[-1, 1]$ whereas $y = 2t^2$ is decreasing over the interval $[-1, 0]$ and then increasing over $[0, 1]$. Geometrically, we start at $(-1, 2)$ (where $t = -1$), then move to the right (since x is increasing) and down (since y is decreasing) to

$(0,0)$ (where $t = 0$). We continue to move to the right (since x is still increasing) but now move upwards (since y is now increasing) until we reach $(1,2)$ (where $t = 1$). Finally, we eliminate the parameter. Solving $x = t^3$ for t, we get $t = \sqrt[3]{x}$. Substituting this into $y = 2t^2$ gives $y = 2(\sqrt[3]{x})^2 = 2x^{2/3}$. The final graph is shown in Figure 11.8(c).

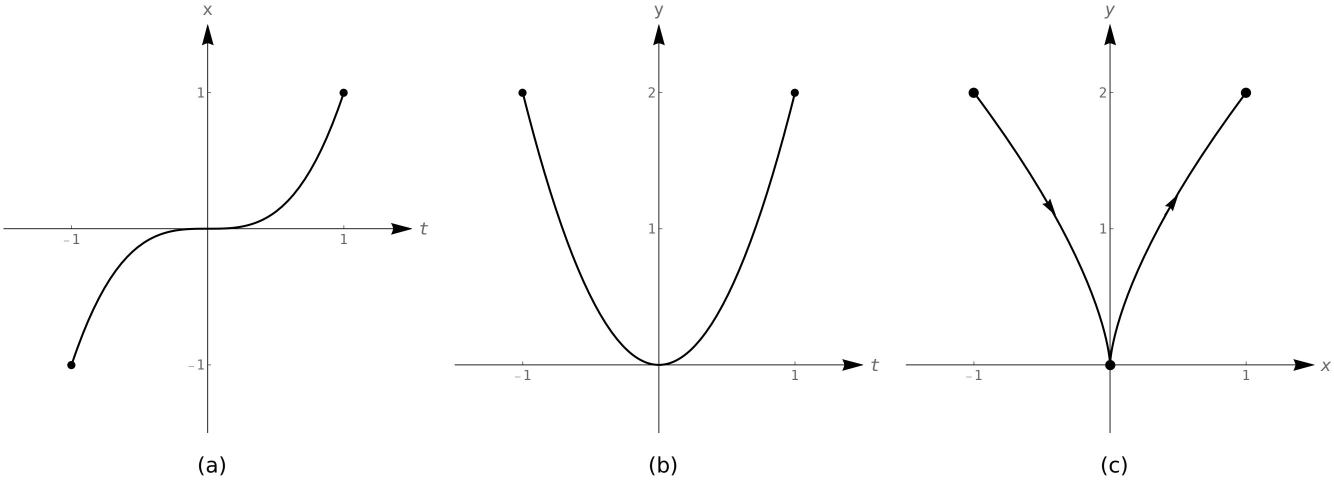


Figure 11.8: The graph of $x = t^3$ (a), $y = 2t^2$ (b) and $x = t^3$, $y = 2t^2$ (c) for $-1 \le t \le 1$.

2. Proceeding as above, we set about graphing this system by first graphing $x = 1 + 3\cos(t)$ and $y = 2\sin(t)$ on the interval $\left[0, \frac{3\pi}{2}\right]$ (Figures 11.9(a) and 11.9(b)). We see that x ranges from -2 to 4 and y ranges from -2 to 2. Plugging in $t = 0$, $\frac{\pi}{2}$, π and $\frac{3\pi}{2}$ gives the points $(4,0)$, $(1,2)$, $(-2,0)$ and $(1,-2)$, respectively. As t ranges from 0 to $\frac{\pi}{2}$, $x = 1 + 3\cos(t)$ is decreasing, while $y = 2\sin(t)$ is increasing. This means that we start tracing out our answer at $(4,0)$ and continue moving to the left and upwards towards $(1,2)$. For $\frac{\pi}{2} \le t \le \pi$, x is decreasing, as is y, so the motion is still right to left, but now is downwards from $(1,2)$ to $(-2,0)$. On the interval $\left[\pi, \frac{3\pi}{2}\right]$, x begins to increase, while y continues to decrease. Hence, the motion becomes left to right but continues downwards, connecting $(-2,0)$ to $(1,-2)$. To eliminate the parameter here, we use the Pythagorean identity. Hence, we solve $x = 1 + 3\cos(t)$ for $\cos(t)$ to get $\cos(t) = \frac{x-1}{3}$, and we solve $y = 2\sin(t)$ for $\sin(t)$ to get $\sin(t) = \frac{y}{2}$. Substituting these expressions into $\cos^2(t) + \sin^2(t) = 1$ gives

$$\left(\frac{x-1}{3}\right)^2 + \left(\frac{y}{2}\right)^2 = 1,$$

or

$$\frac{(x-1)^2}{9} + \frac{y^2}{4} = 1.$$

The graph of this equation is an ellipse centred at $(1,0)$ with vertices at $(-2,0)$ and $(4,0)$. Our parametric equations here are tracing out three-quarters of this ellipse, in a counter-clockwise direction (Figure 11.9(c)).

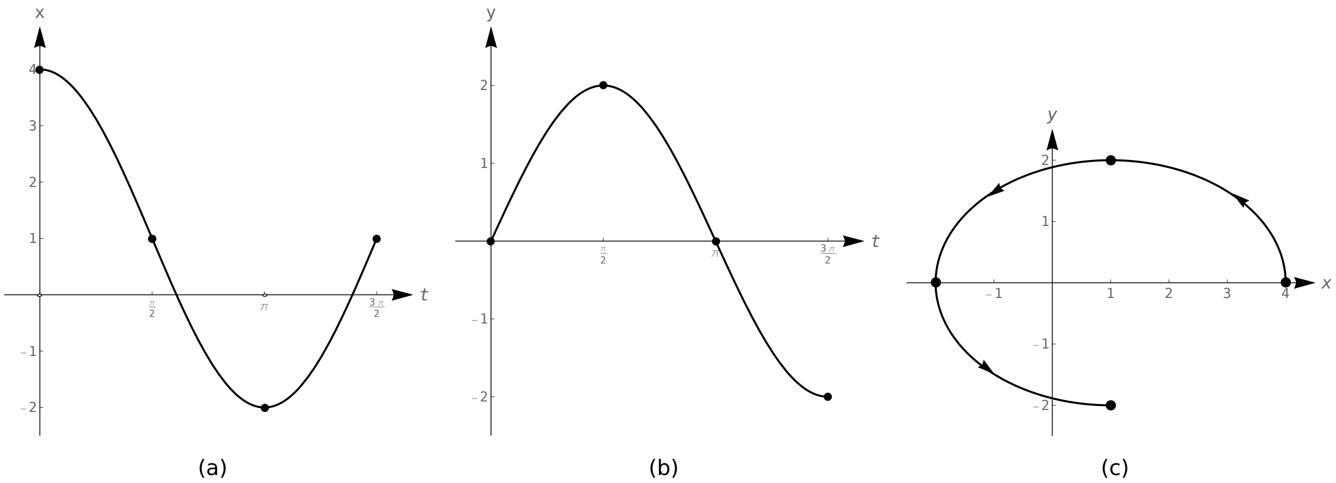


Figure 11.9: The graph of $x = 1 + 3\cos(t)$ (a), $y = 2\sin(t)$ (b) and $x = 1 + 3\cos(t)$, $y = 2\sin(t)$ (c) for $0 \le t \le \frac{3\pi}{2}$.

11.2.3 Parametrising curves

Now that we have had some good practice sketching the graphs of parametric equations, we turn to the problem of finding parametric representations of curves. For that purpose, we have the following guidelines.

- To parametrize $y = f(x)$ as x runs through some interval I, let $x = t$ and $y = f(t)$ and let t run through I.
- To parametrize $x = g(y)$ as y runs through some interval I, let $y = t$ and $x = g(t)$ and let t run through I.
- To parametrize a directed line segment with initial point (x_0, y_0) and terminal point (x_1, y_1), let $x = x_0 + (x_1 - x_0)t$ and $y = y_0 + (y_1 - y_0)t$ for $0 \le t \le 1$.
- To parametrize $\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 1$ where $a, b > 0$, let $x = x_0 + a\cos(t)$ and $y = y_0 + b\sin(t)$ for $0 \le t < 2\pi$. This will impart a counter-clockwise orientation.

Example 11.7

Find a parametrization for each of the following curves.

1. $y = x^2$ from $x = -3$ to $x = 2$.
2. The line segment which starts at $(2, -3)$ and ends at $(1, 5)$.
3. The circle $x^2 + 2x + y^2 - 4y = 4$.
4. The left half of the ellipse
$$\frac{x^2}{4} + \frac{y^2}{9} = 1.$$

Solution

1. Since $y = x^2$ is written in the form $y = f(x)$, we let $x = t$ and $y = f(t) = t^2$. Since $x = t$, the

bounds on t match precisely the bounds on x so we get

$$\begin{cases} x &= t \\ y &= t^2, \end{cases}$$

for $-3 \leq t \leq 2$.

2. To find the equation for x, we have that the line segment starts at $x = 2$ and ends at $x = 1$. This means the displacement in the x-direction is $(1-2) = -1$. Hence, the equation for x is $x = 2 + (-1)t = 2 - t$. For y, we note that the line segment starts at $y = -3$ and ends at $y = 5$. Hence, the displacement in the y-direction is $(5 - (-3)) = 8$, so we get $y = -3 + 8t$. Our final answer is

$$\begin{cases} x &= 2 - t \\ y &= -3 + 8t, \end{cases}$$

for $0 \leq t \leq 1$.

3. In order to use the formulas above to parametrize the circle $x^2 + 2x + y^2 - 4y = 4$, we first need to put it into the correct form. We complete the square and get $(x+1)^2 + (y-2)^2 = 9$, or

$$\frac{(x+1)^2}{9} + \frac{(y-2)^2}{9} = 1.$$

In this equation, we identify $\cos(t) = \frac{x+1}{3}$ and $\sin(t) = \frac{y-2}{3}$. Rearranging these last two equations, we get $x = -1 + 3\cos(t)$ and $y = 2 + 3\sin(t)$. In order to complete one revolution around the circle, we let t range through the interval $[0, 2\pi[$. We get as our final answer

$$\begin{cases} x &= -1 + 3\cos(t) \\ y &= 2 + 3\sin(t), \end{cases}$$

for $0 \leq t < 2\pi$.

4. We immediately get $x = 2\cos(t)$ and $y = 3\sin(t)$. The normal range on the parameter in this case is $0 \leq t < 2\pi$, but since we are interested in only the left half of the ellipse, we restrict t to the values which correspond to Quadrant II and Quadrant III angles, namely $\frac{\pi}{2} \leq t \leq \frac{3\pi}{2}$. Our final answer is

$$\begin{cases} x &= 2\cos(t) \\ y &= 3\sin(t), \end{cases}$$

for $\frac{\pi}{2} \leq t \leq \frac{3\pi}{2}$.

We note that the parametrisation approach offers only one of literally infinitely many ways to parametrize the concerning curves. Essentially, there are two easy ways to alter parametrizations.

- **Reversing Orientation:** Replacing every occurrence of t with $-t$ in a parametric description for a curve reverses the orientation of the curve.
- **Shift of Parameter:** Replacing every occurrence of t with $(t - c)$ in a parametric description for a curve shifts the start of the parameter t ahead by c units.

These techniques are illustrated in the following example.

Example 11.8

Find a parametrization for the unit circle, oriented clockwise, with $t = 0$ corresponding to $(0, -1)$.

Solution

We know that

$$\begin{cases} x &= \cos(t) \\ y &= \sin(t), \end{cases}$$

for $0 \le t < 2\pi$ gives a counter-clockwise parametrization of the unit circle with $t = 0$ corresponding to $(1, 0)$, so the first order of business is to reverse the orientation. Replacing t with $-t$ gives

$$\begin{cases} x &= \cos(-t) \\ y &= \sin(-t), \end{cases}$$

for $0 \le t < 2\pi$ which simplifies to

$$\begin{cases} x &= \cos(t) \\ y &= -\sin(t), \end{cases}$$

for $0 \le t < 2\pi$.

This parametrization gives a clockwise orientation, but $t = 0$ still corresponds to the point $(1, 0)$; the point $(0, -1)$ is reached when $t = -\frac{3\pi}{2}$. Our strategy is to first get the parametrization to start at the point $(0, -1)$ and then shift the parameter accordingly so the start coincides with $t = 0$. We know that any interval of length 2π will parametrize the entire circle, so we start the parameter t at $-\frac{3\pi}{2}$, and find the upper bound by adding 2π so $-\frac{3\pi}{2} \le t < \frac{\pi}{2}$. We now shift the parameter by introducing a time delay of $\frac{3\pi}{2}$ units by replacing every occurrence of t with $\left(t - \frac{3\pi}{2}\right)$, i.e.

$$\begin{cases} x &= \cos\left(t - \frac{3\pi}{2}\right) \\ y &= -\sin\left(t - \frac{3\pi}{2}\right), \end{cases}$$

for $-\frac{3\pi}{2} \le t - \frac{3\pi}{2} < \frac{\pi}{2}$. This simplifies to

$$\begin{cases} x &= -\sin(t) \\ y &= -\cos(t), \end{cases} \tag{11.4}$$

for $0 \le t < 2\pi$, as required.

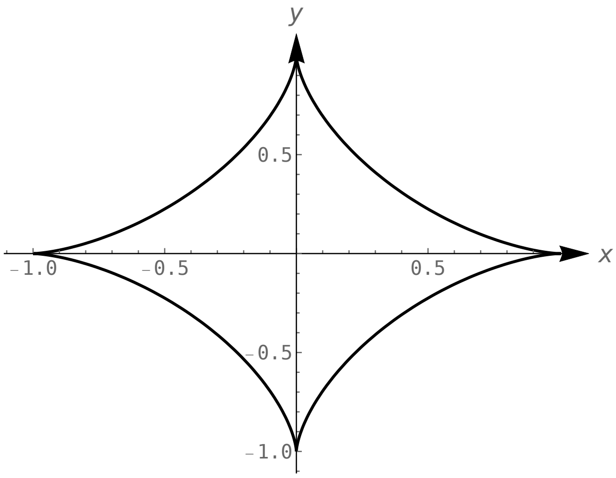
Figure 11.10 gives a small gallery of interesting and famous curves along with the parametric equations that produce them.

11.2.4 Conic sections continued even further

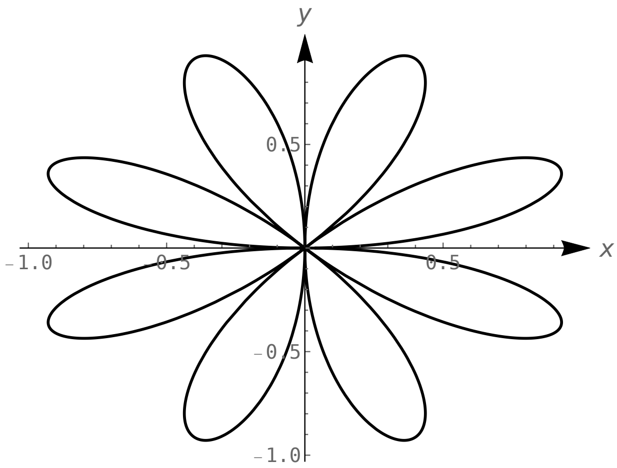
For completeness, we conclude this chapter by listing the parametric representations of the conic sections we introduced in Section 4.4.

The parametric representation of a circle with centre at the origin and radius r is given by

$$\begin{cases} x = r\cos(t) \\ y = r\sin(t), \end{cases} \tag{11.5}$$



Astroid
$x = \cos^3(t)$
$y = \sin^3(t)$



Rose Curve
$x = \cos(t)\sin(4t)$
$y = \sin(t)\sin(4t)$

Figure 11.10: A gallery of interesting planar curves.

for $0 \le t \le 2\pi$. Likewise, the parametric representation of an ellipse centred at the origin and with semi-major and conjugate axis of a and b respectively ($a > b$), is given by

$$\begin{cases} x = a\cos(t) \\ y = b\sin(t), \end{cases} \tag{11.6}$$

for $0 \le t \le 2\pi$.

11.3 Derivatives and parametric and polar equations

11.3.1 Parametric equations

Here we will exemplify the techniques of calculus to study curves given by a set of parametric equations. Amongst other things, we are interested in lines tangent to points on such a curve. They describe how the y-values are changing with respect to the x-values, they are useful in making approximations, and they indicate instantaneous direction of travel.

The slope of the tangent line is still $\frac{dy}{dx}$, and the chain rule allows us to calculate this in the context of parametric equations. If $x = f(t)$ and $y = g(t)$, the chain rule states that

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}.$$

Solving for $\frac{dy}{dx}$, we get

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{g'(t)}{f'(t)}, \tag{11.7}$$

provided that $f'(t) \neq 0$, and we also assume that f and g are differentiable on some open interval I.

These pieces of information allow us to define the tangent and normal lines to a curve C.

Definitie 11.2 (Tangent and normal lines)
Let a curve C be parametrized by $x = f(t)$ and $y = g(t)$, where f and g are differentiable functions on some interval I containing $t = t_0$. The **tangent line** to C at $t = t_0$ is the line through $\big(f(t_0), g(t_0)\big)$ with slope

$$m = \frac{g'(t_0)}{f'(t_0)},$$

provided $f'(t_0) \neq 0$.

The **normal line** to C at $t = t_0$ is the line through $\big(f(t_0), g(t_0)\big)$ with slope $m = -f'(t_0)/g'(t_0)$, provided $g'(t_0) \neq 0$.

This definition leaves two special cases to consider. When the tangent line is horizontal, the normal line is undefined by the above definition as $g'(t_0) = 0$. Likewise, when the normal line is horizontal, the tangent line is undefined. It seems reasonable that these lines be defined, so we add the following to the above definition.

1. If the tangent line at $t = t_0$ has a slope of 0, the normal line to C at $t = t_0$ is the line $x = f(t_0)$.
2. If the normal line at $t = t_0$ has a slope of 0, the tangent line to C at $t = t_0$ is the line $x = f(t_0)$.

Example 11.9

Let $x = 5t^2 - 6t + 4$ and $y = t^2 + 6t - 1$, and let C be the curve defined by these equations.

1. Find the equations of the tangent and normal lines to C at $t = 3$.
2. Find where C has vertical and horizontal tangent lines.

Solution

1. We start by computing $f'(t) = 10t - 6$ and $g'(t) = 2t + 6$. Thus

$$\frac{dy}{dx} = \frac{2t+6}{10t-6}.$$

Make note of something that might seem unusual: $\frac{dy}{dx}$ is a function of t, not x. Just as points on the curve are found in terms of t, so are the slopes of the tangent lines.

The point on C at $t = 3$ is $(31, 26)$. The slope of the tangent line is $m = 1/2$ and the slope of the normal line is $m = -2$. Thus,

- the equation of the tangent line is $y = \frac{1}{2}(x - 31) + 26$, and
- the equation of the normal line is $y = -2(x - 31) + 26$.

This is illustrated in Figure 11.11.

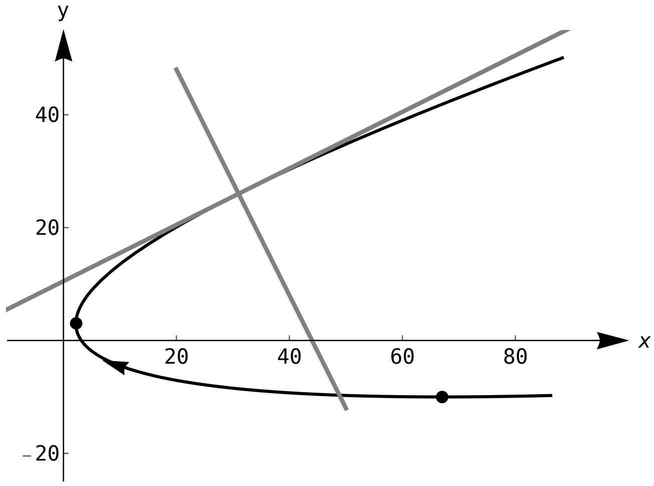


Figure 11.11: Graphing tangent and normal lines in Example 11.9.

2. To find where C has a horizontal tangent line, we set $\frac{dy}{dx}=0$ and solve for t. In this case, this amounts to setting $g'(t)=0$ and solving for t (and making sure that $f'(t)\neq 0$):

$$g'(t)=0 \quad \Rightarrow \quad 2t+6=0 \quad \Leftrightarrow \quad t=-3.$$

The point on C corresponding to $t=-3$ is $(67,-10)$; the tangent line at that point is horizontal (hence with equation $y=-10$).

To find where C has a vertical tangent line, we find where it has a horizontal normal line, and set $-\frac{f'(t)}{g'(t)}=0$. This amounts to setting $f'(t)=0$ and solving for t and making sure that $g'(t)\neq 0$.

$$f'(t)=0 \quad \Rightarrow \quad 10t-6=0 \quad \Leftrightarrow \quad t=0.6.$$

The point on C corresponding to $t=0.6$ is $(2.2, 2.96)$. The tangent line at that point is $x=2.2$.

Example 11.10

Find the equation of the tangent line to the astroid $x=\cos^3(t)$, $y=\sin^3(t)$ at $t=0$ shown in Figure 11.10.

Solution

We start by finding $x'(t)$ and $y'(t)$:

$$x'(t)=-3\sin(t)\cos^2(t), \qquad \text{and} \qquad y'(t)=3\cos(t)\sin^2(t).$$

Note that both of these are 0 at $t=0$; the curve is not smooth at $t=0$ forming a cusp on the graph. Evaluating $\frac{dy}{dx}$ at this point returns the indeterminate form of 0/0. $\frac{dy}{dx}$:

$$\frac{dy}{dx}=\frac{3\cos(t)\sin^2(t)}{-3\sin(t)\cos^2(t)}=-\frac{\sin(t)}{\cos(t)},$$

as long as $\cos(t)\neq 0$ and $\sin(t)\neq 0$. When $t=0$, it is tempting to declare that

$$\frac{dy}{dx}=-\frac{\sin(0)}{\cos(0)}=0,$$

but this overlooks the fact that we cancelled earlier with the stipulation that $\sin(t) \neq 0$. In fact, the graph of the curve has a cusp at $t = 0$, as both $x' = 0$ and $y' = 0$.

We can, however, examine the slopes of tangent lines near $t = 0$, and take the limit as $t \to 0$.

$$\begin{aligned} \lim_{t\to 0} \frac{y'(t)}{x'(t)} &= \lim_{t\to 0} \frac{3\cos(t)\sin^2(t)}{-3\sin(t)\cos^2(t)} && \text{(We can cancel as } t \neq 0.\text{)} \\ &= \lim_{t\to 0}\left(-\frac{\sin(t)}{\cos(t)}\right) \\ &= 0\,. \end{aligned}$$

We have accomplished something significant. When the derivative $\frac{dy}{dx}$ returns an indeterminate form at $t = t_0$, we can define its value by setting it to be $\lim\limits_{t\to t_0} \frac{dy}{dx}$, if that limit exists. This allows us to find slopes of tangent lines at cusps, which can be very beneficial.

We found the slope of the tangent line at $t = 0$ to be 0; therefore the tangent line is $y = 0$, the x-axis.

11.3.2 Polar equations

A basis for much of what is done in this section is the ability to turn a polar function $r = f(\theta)$ into a set of parametric equations. Using the identities $x = r\cos(\theta)$ and $y = r\sin(\theta)$, we can create the parametric equations $x = f(\theta)\cos(\theta)$, $y = f(\theta)\sin(\theta)$ and continue our work with those.

For instance, if we are asked to construct the tangent line to a curve described by $r = f(\theta)$, we will use $x = f(\theta)\cos(\theta)$, $y = f(\theta)\sin(\theta)$ to compute $\frac{dy}{dx}$. Using Equation (11.7) we have

$$\frac{dy}{dx} = \frac{\dfrac{dy}{d\theta}}{\dfrac{dx}{d\theta}}\,.$$

Each of the two derivatives on the right hand side of the equality requires the use of the product rule to arrive at

$$\frac{dy}{dx} = \frac{f'(\theta)\sin(\theta) + f(\theta)\cos(\theta)}{f'(\theta)\cos(\theta) - f(\theta)\sin(\theta)}. \tag{11.8}$$

Example 11.11

Consider the limaçon $r = 1 + 2\sin(\theta)$ on $[0, 2\pi]$. Find the equations of the tangent and normal lines to the graph at $\theta = \pi/4$.

Solution

We start by computing $\frac{dy}{dx}$. With $f'(\theta) = 2\cos(\theta)$, we have

$$\begin{aligned}\frac{dy}{dx} &= \frac{2\cos(\theta)\sin(\theta)+\cos(\theta)(1+2\sin(\theta))}{2\cos^2(\theta)-\sin(\theta)(1+2\sin(\theta))}\\ &= \frac{\cos(\theta)(4\sin(\theta)+1)}{2(\cos^2(\theta)-\sin^2(\theta))-\sin(\theta)}\,.\end{aligned}$$

When $\theta = \pi/4$, $\frac{dy}{dx} = -2\sqrt{2}-1$. In rectangular coordinates, the point on the graph at $\theta = \pi/4$ is $(1+\sqrt{2}/2, 1+\sqrt{2}/2)$. Thus the rectangular equation of the line tangent to the limaçon at $\theta = \pi/4$ is

$$y = (-2\sqrt{2}-1)\left(x-\left(1+\frac{\sqrt{2}}{2}\right)\right)+1+\frac{\sqrt{2}}{2} \approx -3.83x+8.24\,.$$

The limaçon and the tangent line are graphed in Figure 11.12.

The normal line has the opposite–reciprocal slope as the tangent line, so its equation is

$$y \approx \frac{1}{3.83}x + 1.26.$$

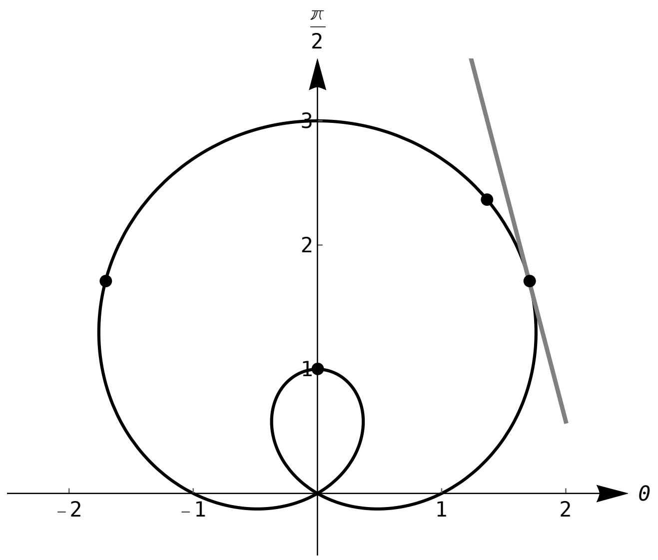


Figure 11.12: The limaçon in Example 11.11 with its tangent line at $\theta = \pi/4$ and points of vertical and horizontal tangency.

11.3.3 Smoothness

For what concerns the smoothness of parametric curves, we have the following – stricter – definition in order to arrive at curves without corners.

Definitie 11.3 (Smoothness of a parametric curve)
A curve C defined by $x = f(t)$, $y = g(t)$ is **smooth** (*glad*) on an interval I if f' and g' are continuous on I and not simultaneously 0 (except possibly at the endpoints of I). A curve is **piecewise**

smooth (*stuksgewijs glad*) on I if I can be partitioned into subintervals where C is smooth on each subinterval.

The continuity condition is in agreement with Definition 9.7 and relates to parameterizations that could fail to be differentiable at a point. The second condition, however, relates to parameterizations that could slow to a stop, and then start up again in a completely different direction. Indeed, if a curve is not smooth at $t = t_0$, it means that $x'(t_0) = y'(t_0) = 0$ as defined. This, in turn, means that rate of change of x (and y) is 0; that is, at that instant, neither x nor y is changing. If the parametric equations describe the path of some object, this means the object is at rest at t_0. An object at rest can make a sharp change in direction, whereas moving objects tend to change direction in a smooth fashion.

Consider the astroid, given by $x = \cos^3(t)$, $y = \sin^3(t)$ (Figure 11.10). Taking derivatives, we have:

$$x' = -3\cos^2(t)\sin(t) \quad \text{and} \quad y' = 3\sin^2(t)\cos(t).$$

It is clear that each is 0 when $t = 0,\ \pi/2,\ \pi, \ldots$. Thus the astroid is not smooth at these points, corresponding to the cusps seen in Figure 11.10.

Example 11.12

Let a curve C be defined by the parametric equations $x = t^3 - 12t + 17$ and $y = t^2 - 4t + 8$. Determine the points, if any, where it is not smooth.

Solution

We begin by taking derivatives.

$$x' = 3t^2 - 12, \text{ and } y' = 2t - 4.$$

We set each equal to 0. It follows that

$$x' = 0 \quad \Leftrightarrow \quad 3t^2 - 12 = 0 \quad \Leftrightarrow \quad t = \pm 2,$$

and

$$y' = 0 \quad \Leftrightarrow \quad 2t - 4 = 0 \quad \Leftrightarrow \quad t = 2.$$

We see that at $t = 2$ both x' and y' are 0; thus C is not smooth at $t = 2$, corresponding to the point $(1, 4)$. The curve is graphed in Figure 11.13, illustrating the cusp at $(1, 4)$.

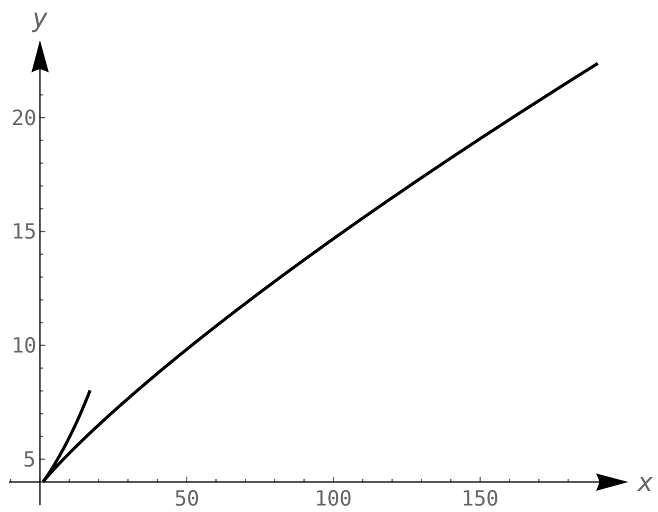


Figure 11.13: Graphing the curve in Example 11.12; note it is not smooth at $(1, 4)$.

One should be careful to note that a sharp corner does not have to occur when a curve is not smooth. For instance, one can verify that $x = t^3$ and $y = t^6$ produce the familiar $y = x^2$ parabola. However, in this parametrization, the curve is not smooth. A particle travelling along the parabola according to the given parametric equations comes to rest at $t = 0$, though no sharp point is created.

11.3.4 Concavity

For what concerns curves in the plane described by means of parametric equations, we may also consider their concavity; that is, we are interested in $\frac{d^2y}{dx^2}$. To find this, we need to find the derivative of $\frac{dy}{dx}$ with respect to x; that is,

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left[\frac{dy}{dx}\right],$$

but recall that $\frac{dy}{dx}$ is a function of t, not x, making this computation not straightforward.

Let now $h(t) = \frac{dy}{dx}$. We want $\frac{d}{dx}[h(t)]$, which follows from the chain rule. Indeed, we have

$$\frac{dh}{dt} = \frac{dh}{dx} \cdot \frac{dx}{dt} \quad \Rightarrow \quad \frac{dh}{dx} = \frac{\frac{dh}{dt}}{\frac{dx}{dt}}.$$

Hence, this leads to

$$\frac{d^2y}{dx^2} = \frac{\frac{d}{dt}\left[\frac{dy}{dx}\right]}{\frac{dx}{dt}} = \frac{\frac{d}{dt}\left[\frac{dy}{dx}\right]}{f'(t)}. \tag{11.9}$$

An example will help us understand this.

Example 11.13

Let $x = 5t^2 - 6t + 4$ and $y = t^2 + 6t - 1$ as in Example 11.9. Determine the t-intervals on which the graph is concave up/down.

Solution

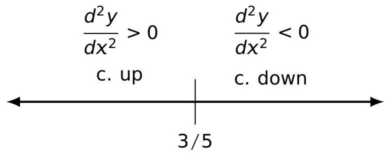
Concavity is determined by the second derivative of y with respect to x, $\frac{d^2y}{dx^2}$, so we compute that here following Equation (11.9).

In Example 11.9, we found $\frac{dy}{dx} = \frac{2t+6}{10t-6}$ and $f'(t) = 10t - 6$. So:

$$\begin{aligned}
\frac{d^2y}{dx^2} &= \frac{\frac{d}{dt}\left[\frac{2t+6}{10t-6}\right]}{10t-6} \\
&= -\frac{\frac{72}{(10t-6)^2}}{10t-6} \\
&= -\frac{72}{(10t-6)^3} \\
&= -\frac{9}{(5t-3)^3}.
\end{aligned}$$

The graph of the parametric functions is concave up when $\frac{d^2y}{dx^2} > 0$ and concave down when $\frac{d^2y}{dx^2} < 0$. We determine the intervals when the second derivative is greater/less than 0 by first finding when it is 0 or undefined.

As the numerator of $-\frac{9}{(5t-3)^3}$ is never 0, $\frac{d^2y}{dx^2} \neq 0$ for all t. It is undefined when $5t-3=0$; that is, when $t = 3/5$. Following the work established in Section 10.4, we look at values of t greater/less than $3/5$ on a number line:



Reviewing Example 11.9, we see that when $t = 3/5 = 0.6$, the graph of the parametric equations has a vertical tangent line. This point is also a point of inflection for the graph, illustrated in Figure 11.14.

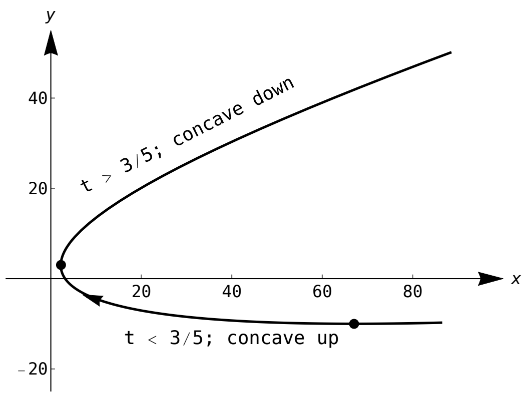


Figure 11.14: Graphing the parametric equations in Example 11.13 to demonstrate concavity.

11.4 Exercises

Polar coordinates

Assignment 11.1 — Which of the following pairs of polar coordinates represents the same point?

(a) $(3,0) = (-3,\pi)$

(b) $(-3,0) = (-3,2\pi)$

(c) $\left(2,\frac{2\pi}{3}\right) = \left(-2,-\frac{\pi}{3}\right)$

(d) $\left(2,\frac{7\pi}{3}\right) = \left(2,\frac{\pi}{3}\right) = \left(2,\frac{13\pi}{3}\right)$

Assignment 11.2 — Determine the Cartesian coordinates for the following points given in polar coordinates.

(a) $\left(\sqrt{2},\frac{\pi}{4}\right)$

(b) $(1,0)$

(c) $(0,\pi)$

(d) $\left(-\sqrt{2},\frac{\pi}{4}\right)$

(e) $\left(2\sqrt{3},\frac{2\pi}{3}\right)$

Assignment 11.3 — Convert the given polar equation into a Cartesian equation and name the curve.

(a) $\theta = \frac{\pi}{4}$

(b) $r = \frac{7}{5\sin(\theta) - 2\cos(\theta)}$

(c) $r = 2\cos(\theta)$

(d) $r = -4\sin(\theta)$

(e) $r = \sin(\theta) + \cos(\theta)$

(f) $r = \frac{2}{\sqrt{\cos^2(\theta) + 4\sin^2(\theta)}}$

(g) $r = \frac{1}{1-\cos(\theta)}$

(h) $r = \frac{1}{1-2\sin(\theta)}$

Assignment 11.4 — Sketch the graph of the curves below.

(a) $r = 2 \qquad \left(0 \le \theta \le \dfrac{\pi}{2}\right)$

(b) $r = 2 - \sin(\theta)$

(c) $r = \dfrac{3}{2\cos(\theta) - \sin(\theta)}$

(d) $r = 2 + 4\cos(\theta)$

(e) $r = 5\sin(2\theta)$

(f) $r = 2\sin(2\theta)$

(g) $r = \cos\left(\dfrac{2\theta}{3}\right) \qquad (0 \le \theta \le 6\pi)$

(h) $r = 3\sin(\theta) \qquad (0 \le \theta \le \pi)$

(i) $r = 3\csc(\theta) \qquad (0 < \theta < \pi)$

(j) $r^2 = 4\sin(2\theta)$

(k) $r^2 = 4\cos(3\theta)$

(l) $r^2 = \sin(3\theta)$

Assignment 11.5 — Determine the intersection(s) of the graphs represented by the polar equations below.

(a) $r = 3\cos(\theta), \quad r = 1 + \cos(\theta)$

(b) $r = \sec^2\left(\dfrac{\theta}{2}\right), \quad r = 3\csc^2\left(\dfrac{\theta}{2}\right)$

(c) $r = \sin(\theta), \quad r = 1 - \sin(\theta)$

(d) $r = \sqrt{3}\cos(\theta), \quad r = \sin(\theta)$

(e) $r^2 = 2\cos(2\theta), \quad r = 1$

(f) $r = \sin(3\theta), \quad r = \cos(3\theta), \quad [0, \pi]$

(g) $r = 1 - \cos(\theta), \quad r = 1 + \sin(\theta), \ [0, 2\pi]$

Parametric equations

Assignment 11.6 — Determine the Cartesian equation of the given parameter representation and draw the corresponding curve.

(a) $\begin{cases} x = 2 - t \\ y = t + 1 \end{cases} \qquad (0 < t < +\infty)$

(b) $\begin{cases} x = \dfrac{1}{t} \\ y = t - 1 \end{cases} \qquad (0 < t < 4)$

(c) $\begin{cases} x = \dfrac{1}{1+t^2} \\ y = \dfrac{t}{1+t^2} \end{cases} \qquad (t \in \mathbb{R})$

(d) $\begin{cases} x = 3\sin(2t) \\ y = 3\cos(2t) \end{cases} \qquad \left(0 \le t \le \dfrac{\pi}{3}\right)$

(e) $\begin{cases} x = 1 - \sqrt{4 - t^2} \\ y = 2 + t \end{cases} \qquad (-2 \le t \le 2)$

(f) $\begin{cases} x = \sec(t) \\ y = \tan(t) \end{cases} \qquad (t \in \mathbb{R})$

(g) $\begin{cases} x = e^t \\ y = e^{3t} - 3 \end{cases} \qquad (t \in \mathbb{R})$

Assignment 11.7 — Determine a parametrization of the curves below.

(a) The lower half of the parabola $y^2 = x - 1$.

(b) $x^{2/3} + y^{2/3} = 6^{2/3}$

Assignment 11.8 — Use $t = y$ to paramterize the intersection of the planes $y = 2x - 4$ and $z = 3x + 1$ between $(2, 0, 7)$ and $(3, 2, 10)$.

Assignment 11.9 — The curve of intersection of the plane $x + y = 1$ with the paraboloid $z = x^2 + y^2$ is a parabola. Parameterize this parabola using $t = x$ as a parameter. Can you use $t = y$ as well? What about $t = z$?

Assignment 11.10 — Parameterize the curve that defines the intersection between the given curves.

(a) $x^2 + y^2 = 9$ and $z = x + y$

(b) $z = \sqrt{1 - x^2 - y^2}$ and $x + y = 1$

(c) $z = x^2 + y^2$ and $2x - 4y - z - 1 = 0$

Assignment 11.11 — Sketch the graph of the curves given by the parameter representations below.

(a) $\begin{cases} x = t^2 \\ y = 2 \end{cases} \quad (-2 \le t \le 2)$

(b) $\begin{cases} x = t-1 \\ y = 2t+3 \end{cases} \quad (-\infty < t < +\infty)$

(c) $\begin{cases} x = t^2 \\ y = t-3 \end{cases} \quad (-\infty < t < +\infty)$

(d) $\begin{cases} x = \sin(t) \\ y = \cos^2(t) \end{cases} \quad \left(0 \le t \le \frac{3\pi}{2}\right)$

(e) $\begin{cases} x = -2\sin(t) \\ y = 3\cos(t) \end{cases} \quad (0 \le t \le 3\pi)$

(f) $\begin{cases} x = 2t \\ y = t^3 + 4 \end{cases} \quad (-2 \le t \le 2)$

Assignment 11.12 — Determine which graphs of $x = f(t)$ and $y = f(t)$ in Figure 11.19 belong to the graphs of the parametric curves in Figure 11.20.

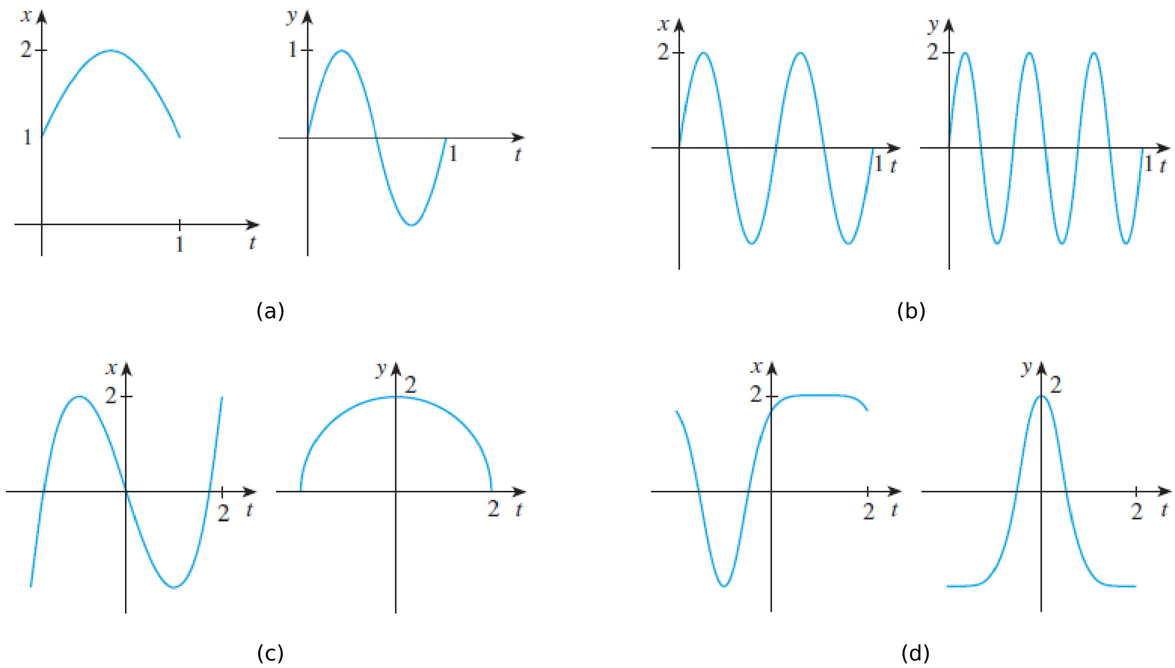


Figure 11.19: Graphs of the parametric equations from Exercise 12.

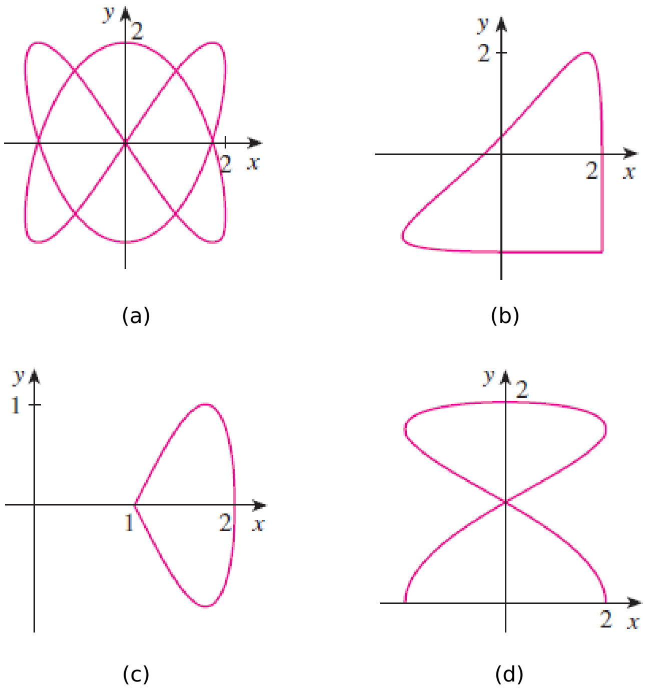


Figure 11.20: Graphs of the parameter curves from Exercise 12.

Derivatives and parametric and polar equations

Assignment 11.13 — Determine y' and y'' Of the following curves.

(a) $\begin{cases} x = t^2 - 1 \\ y = \dfrac{2}{t} \end{cases}$

(b) $\begin{cases} x = t^2 - 2t \\ y = t^2 + 2t \end{cases}$

(c) $\begin{cases} x = 3\sin(t) \\ y = 4\cos(t) \end{cases}$

(d) $\begin{cases} x = \dfrac{3}{\cos(t)} \\ y = 4\tan(t) \end{cases}$

Assignment 11.14 — Determine the slope of the tangent to the given curve at the given point.

(a) $r = 1 - 3\cos(\theta)$ in $\theta = \dfrac{3\pi}{4}$

(b) $r = \sin(4\theta)$ in $\theta = \dfrac{\pi}{3}$

(c) $x = t^3 + t, \quad y = 1 - t^3$ in $t = 1$

(d) $x = e^{2t}, \quad y = te^{2t}$ in $t = -2$

Assignment 11.15 — Determine an equation of the tangent and normal at the given point to the given curve.

(a) $r = 1 + \sin(\theta)$ in $\theta = \dfrac{\pi}{6}$

(b) $r = \dfrac{1}{\sin(\theta) - \cos(\theta)}$ in $\theta = \pi$

(c) $x = t^2 - t, \quad y = t^2 + t \quad$ in $t = 1$

(d) $x = \cos(t), \quad y = \sin(2t) \quad (t \in [0, 2\pi]) \quad$ in $t = \pi/4$

(e) $x = e^{t/10}\cos(t), \quad y = e^{t/10}\sin(t) \quad$ in $t = \pi/2$

Assignment 11.16 — Determine the coordinates of the points where the given curve has (a) a horizontal and (b) a vertical tangent.

(a) $x = t^3 - 3t, \quad y = 2t^3 + 3t^2$

(b) $x = \sin(t), \quad y = \sin(t) - t\cos(t)$

(c) $x = \dfrac{3t}{1+t^3}, \quad y = \dfrac{3t^2}{1+t^3}$

(d) $x = \sec(t), \quad y = \tan(t) \quad \left(-\dfrac{\pi}{2} < t < \dfrac{\pi}{2}\right)$

(e) $x = \cos(t)\sin(2t), \quad y = \sin(t)\sin(2t)$

(f) $r = 1 + \cos(\theta)$

(g) $r^2 = \cos(2\theta)$

(h) $r = 2(1 - \sin(\theta))$

Assignment 11.17 — The curve with the parametric equations

$$\begin{cases} x = t - 2\sin(t) \\ y = 1 - 2\cos(t) \end{cases}$$

with $t \in [0, 2\pi]$ is named a *trochoid*. This is the red curve in Figure 11.22.

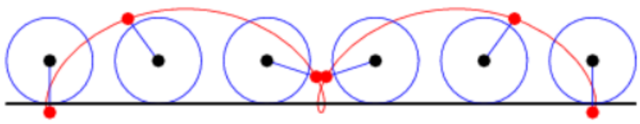


Figure 11.22: Figure from Exercise 11.17.

(a) Determine y' and y''.

(b) Determine the points in which the given curve has a horizontal or a vertical tangent.

(c) Determine the points in which the direction co-efficient of the tangent line is 1 or -1 is.

Assignment 11.18 — Determine the values of t for which the given curve is not smooth.

(a) $x = t^2 - 4t, \quad y = t^3 - 2t^2 - 4t$

(b) $x = t\sin(t), \quad y = t^3$

(c) $x = 2\cos(t) - \cos(2t), \quad y = 2\sin(t) - \sin(2t)$

(d) $x = \dfrac{1}{t^2+1}, \quad y = t^3$

(e) $x = t^3 - 3t^2 + 3t - 1, \quad y = t^2 - 2t + 1$

(f) $x = \cos^2(t), \quad y = 1 - \sin^2(t)$

Assignment 11.19 — Sketch the graph of the given curve based on the first and second derivatives.

(a) $x = t^2 - 2t, \quad y = t^2 - 4t$

(b) $x = t^3 - 3t, \quad y = \dfrac{2}{1+t^2}$

(c) $x = \cos(t) + t\sin(t), \quad y = \sin(t) - t\cos(t) \qquad (t \geq 0)$

(d) $x = t^2 + t, \quad y = 1 - t^2 \qquad (-3 \leq t \leq 3)$

Nature laughs at the difficulties of integration.

— Pierre-Simon Laplace —

12

Integration

We have spent considerable time considering the derivatives of a function and their applications. In the following chapters, we are going to starting thinking in the other direction. That is, given a function $f(x)$, we are going to consider functions $F(x)$ such that $F'(x) = f(x)$. These functions will help us compute area, volume, mass, force, pressure, work, and much more.

12.1 Antiderivatives and (in)definite integration

12.1.1 Antiderivatives and indefinite integration

Given a function $y = f(x)$, a **differential equation** (*differentiaalvergelijking*) is one that incorporates y, x, and the derivatives of y. For instance, a simple differential equation is:

$$y' = 2x.$$

Solving a differential equation amounts to finding a function y that satisfies the given equation. Take a moment and consider that equation; can you find a function y such that $y' = 2x$?

Hopefully one was able to come up with at least one solution: $y = x^2$. Finding another may have seemed impossible until one realizes that a function like $y = x^2 + 1$ also has a derivative of $2x$. Once that discovery is made, finding yet another is not difficult; the function $y = x^2 + 123\,456\,789$ also has a derivative of $2x$. The differential equation $y' = 2x$ has many solutions. This leads us to some definitions.

Definitie 12.1 (Antiderivatives and indefinite integrals)
Let a function $f(x)$ be given. An **antiderivative** (*primitieve functie*) of $f(x)$ is a function $F(x)$ such that $F'(x) = f(x)$.

The set of all antiderivatives of $f(x)$ is the **indefinite integral** (*onbepaalde integraal*) of f, denoted by

$$\int f(x)\ dx.$$

Note that we refer to an antiderivative of f, as opposed to the antiderivative of f, since there is always an infinite number of them. We often use upper-case letters to denote antiderivatives. Besides, knowing one antiderivative of f allows us to find infinitely more, simply by adding a constant. Not only does this give us more antiderivatives, it gives us all of them.

Theorem 12.1 (Antiderivative forms)
Let $F(x)$ and $G(x)$ be antiderivatives of $f(x)$ on an interval I. Then there exists a constant C such that, on I,

$$G(x) = F(x) + C.$$

Given a function f defined on an interval I and one of its antiderivatives F, we know all antiderivatives of f on I have the form $F(x)+C$ for some constant C. Using Definition 12.1, we can say that

$$\int f(x)\ dx = F(x) + C.$$

The integration symbol, $\int$, is in reality an elongated S, representing summing. We will later see how sums and antiderivatives are related. The function we want to find an antiderivative of is called the **integrand** (*integrand*). It contains the differential of the variable we are integrating with respect to.

Let us now use our notice to evaluate

$$\int \sin(x)\ dx.$$

Essentially, this means that we should find all functions $F(x)$ such that $F'(x) = \sin(x)$. Of course, some thought leads us to one solution: $F(x) = -\cos(x)$, because $\frac{d}{dx}(-\cos(x)) = \sin(x)$. The indefinite integral of $\sin(x)$ is thus $-\cos(x)$, plus a constant of integration C. So:

$$\int \sin(x)\ dx = -\cos(x) + C.$$

To fully understand what is happening, it is important to realise that the process of antidifferentiation is really solving a differential question. The integral

$$\int \sin(x)\ dx$$

presents us with a differential, $dy = \sin(x)\ dx$. It is asking: What is y? We found lots of solutions, all of the form $y = -\cos(x) + C$.

Letting $dy = \sin(x)\ dx$, rewrite

$$\int \sin(x)\ dx \quad \text{as} \quad \int dy.$$

This is asking: “What functions have a differential of the form dy?” The answer is “Functions of the form $y + C$, where C is a constant.” What is y? We have lots of choices, all differing by a constant; the simplest choice is $y = -\cos(x)$.

In Mathematica, we can use the command **Integrate** to evaluate an indefinite integral. For instance,

$$\int (3x^2 + 4x + 5)\, dx.$$

can be evaluated as follows.

```
In[19]:=  Integrate[3*x^2+4*x+5, x]

Out[19]=  5x +2x^2 +x^3
```

We can also see that taking the derivative of our answer returns the function in the integrand. Thus we can say that:

$$\frac{d}{dx}\left(\int f(x)\, dx\right) = f(x).$$

Differentiation undoes the work done by antidifferentiation.

Taking into account the lists of derivatives of algebraic and transcendental functions presented in Chapter 9, we may now state some important antiderivatives. We easily see that

$$\int 0\, dx = C\,,$$

and

$$\int 1\, dx = \int dx = x + C\,,$$

from which we can infer the following more general integral rule:

$$\int x^n\, dx = \frac{1}{n+1}x^{n+1} + C,$$

for $n \neq -1$.

For what concerns the exponential and logarithmic functions, we get the following derivative functions:

- $\int e^x\, dx = e^x + C,$
- $\int a^x\, dx = \frac{1}{\ln(a)}a^x + C,$
- $\int \frac{1}{x}\, dx = \ln|x| + C,$

while for the trigonometric functions we get:

- $\int \sin(x)\, dx = -\cos(x) + C$
- $\int \cos(x)\, dx = \sin(x) + C$
- $\int \sec^2(x)\, dx = \tan(x) + C$
- $\int \csc^2(x)\, dx = -\cot(x) + C$

Besides, we have the following properties, which are completely in line with those for derivatives (Theorem 9.1)

Theorem 12.2 (Properties of the antiderivative)
Let f and g be differentiable on an open interval I and let k be a real number. Then:

1. *Sum/Difference rule:*
$$\int \big(f(x) \pm g(x)\big)\, dx = \int f(x)\, dx \pm \int g(x)\, dx\,. \tag{12.1}$$

2. *Constant multiple rule:*
$$\int k f(x)\, dx = k \int f(x)\, dx\,. \tag{12.2}$$

In Section 9.1.4 we saw that the derivative of a position function gave a velocity function, and the derivative of a velocity function describes acceleration. We can now go the other way: the antiderivative of an acceleration function gives a velocity function, etc. While there is just one derivative of a given function, there are infinitely many antiderivatives. Therefore we cannot ask "What is the velocity of an object whose acceleration is -32m/s^2?", since there is more than one answer.

We can find the answer if we provide more information with the question, as done in the following example. Often the additional information comes in the form of an initial value, a value of the function that one knows beforehand.

Example 12.1

The acceleration due to gravity of a falling object is -9 m/s^2. At time $t = 3$, a falling object had a velocity of -10 m/s. Find the equation of the object's velocity.

Solution

We want to know a velocity function, $v(t)$. We know two things:

- The acceleration, i.e., $v'(t) = -9$, and
- the velocity at a specific time, i.e., $v(3) = -10$.

Using the first piece of information, we know that $v(t)$ is an antiderivative of $v'(t) = -9$. So we begin by finding the indefinite integral of -9:

$$\int v'(t)\, dt = \int (-9)\, dt = -9t + C = v(t).$$

Now we use the fact that $v(3) = -10$ by plugging in this point in the equation we just got for $v(t)$:

$$-9 \cdot (3) + C = -10\,,$$

for which it directly follows that $C = 17$.

Thus $v(t) = -9t + 17$. We can use this equation to understand the motion of the object: when $t = 0$, the object had a velocity of $v(0) = 17$ m/s. Since the velocity is positive, the object was moving upward.

In the remainder of this section, we will see how position and velocity are unexpectedly related by the areas of certain regions on a graph of the velocity function.

12.1.2 The definite integral

We start with an easy problem. An object travels in a straight line at a constant velocity of 5m/s for 10 seconds. How far away from its starting point is the object?

Since, we have that Distance = Rate × Time, it follows that this distance is 50 metres. This solution can be represented graphically. Consider Figure 12.1(a), where the constant velocity of 5m/s is graphed on the axes. Shading the area under the line from $t = 0$ to $t = 10$ gives a rectangle with an area of 50 square units; when one considers the units of the axes, we can say this area represents 50 m.

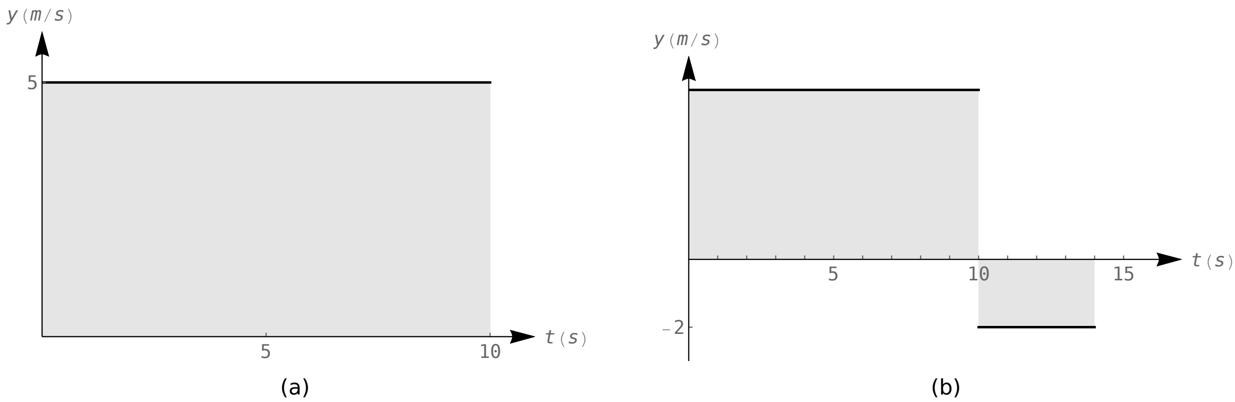


Figure 12.1: The total displacement of an object travelling in a straight line at a constant velocity of 5m/s for 10 seconds (a) and an object travelling a straight line with a constant velocity of 5m/s for 10 seconds, and then instantly reversing course at a rate of 2m/s for 4 seconds (b).

Now consider a slightly harder situation (and not particularly realistic): an object travels in a straight line with a constant velocity of 5m/s for 10 seconds, then instantly reverses course at a rate of 2m/s for 4 seconds. How far away from the starting point is the object – what is its displacement?

Here, we get:

$$\text{Distance} \;\; = 5 \cdot 10 + (-2) \cdot 4 = 42 \text{ m}.$$

Hence the object is 42 metres from its starting location.

We can again depict this situation graphically. In Figure 12.1(b) we have the velocities graphed as straight lines on $[0, 10]$ and $[10, 14]$, respectively. The displacement of the object is given by

$$\text{Area above the } t\text{–axis} \quad - \quad \text{Area below the } t\text{–axis},$$

which is easy to calculate as $50 - 8 = 42$ metres.

These examples do not prove a relationship between area under a velocity function and displacement, but it does imply a relationship exists. Section 12.3 will fully establish fact that the area under a velocity function is displacement.

Anyhow, given a graph of a function $y = f(x)$, we will find that there is great use in computing the area between the curve $y = f(x)$ and the x-axis. Because of this, we need to define some terms.

Definitie 12.2 (The definite integral, total signed area)
Let $y = f(x)$ be defined on a closed interval $[a, b]$. The total signed area from $x = a$ to $x = b$ between f and the x-axis is:

$$(\text{area under } f \text{ and above the } x\text{–axis on } [a,b]) - (\text{area above } f \text{ and under the } x\text{–axis on } [a,b]).$$

The **definite integral** (*bepaalde integraal*) of f on $[a, b]$ is the total signed area of f on $[a, b]$, denoted

$$\int_a^b f(x)\, dx,$$

where a and b are the bounds of integration.

By our definition, the definite integral gives the signed area under f. We usually drop the word signed when talking about the definite integral, and simply say the definite integral gives the area under f or, more commonly, the area under the curve. The indefinite integral and definite integral are very much related, as we will see in Section 12.3.

Let us now practice this definition.

Example 12.2

Consider the function f given in Figure 12.2.

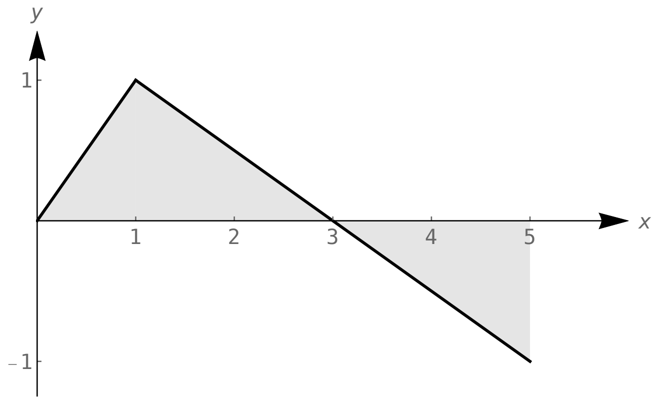


Figure 12.2: A graph of $f(x)$ in Example 12.2.

Find:

1. $\int_0^3 f(x)\, dx$
2. $\int_3^5 f(x)\, dx$
3. $\int_0^5 f(x)\, dx$
4. $\int_0^3 5f(x)\, dx$
5. $\int_1^1 f(x)\, dx$

Solution

1. This definite integral is the area under f on the interval $[0,3]$. This region is a triangle, so the area is

$$\int_0^3 f(x)\, dx = \frac{1}{2}(3)(1) = 1.5.$$

2. This definite integral represents the area of the triangle found under the x–axis on $[3,5]$. The

area is $1/2(2)(1) = 1$; since it is found under the x–axis, this is negative area. So,

$$\int_3^5 f(x)\, dx = -1.$$

3. This definite integral is the total signed area under f on $[0,5]$. This is $1.5 + (-1) = 0.5$.
4. This definite integral is the area under $5f$ on $[0,3]$. Again, the region is a triangle, with height 5 times that of the height of the original triangle. Thus the area is

$$\int_0^3 5f(x)\, dx = \frac{1}{2}(15)(1) = 7.5.$$

5. This definite integral is the area under f on the interval $[1,1]$. This describes a line segment, not a region; it has no width. Therefore the area is 0.

This example illustrates some of the properties of the definite integral, listed in the following theorem.

Theorem 12.3 (Properties of the definite integral)
Let f and g be defined on a closed interval I that contains the values a, b and c, and let k be a constant. The following hold:

1. $\displaystyle\int_a^a f(x)\, dx = 0,$
2. $\displaystyle\int_a^b f(x)\, dx + \int_b^c f(x)\, dx = \int_a^c f(x)\, dx,$
3. $\displaystyle\int_a^b f(x)\, dx = -\int_b^a f(x)\, dx,$
4. $\displaystyle\int_a^b \big(f(x) \pm g(x)\big)\, dx = \int_a^b f(x)\, dx \pm \int_a^b g(x)\, dx,$
5. $\displaystyle\int_a^b kf(x)\, dx = k \cdot \int_a^b f(x)\, dx.$

The area definition of the definite integral allows us to use geometry to compute the definite integral of some simple functions.

Example 12.3

Evaluate the following definite integrals:

1. $\displaystyle\int_{-2}^{5} (2x-4)\ dx$

2. $\displaystyle\int_{-3}^{3} \sqrt{9-x^2}\ dx.$

Solution

1. It is useful to sketch the function in the integrand, as shown in Figure 12.3(a). We see we need to compute the areas of two regions, which we have labelled R_1 and R_2. Both are triangles, so the area computation is straightforward:

$$R_1: \ \frac{1}{2}(4)(8) = 16 \qquad R_2: \ \frac{1}{2}(3)6 = 9.$$

Region R_1 lies under the x–axis, hence it is counted as negative area, so

$$\int_{-2}^{5} (2x-4)\ dx = -16+9 = -7.$$

We may check this answer in Mathematica as follows

```
In[20]:= Integrate[2*x-4, x,-2,5]

Out[20]= -7
```

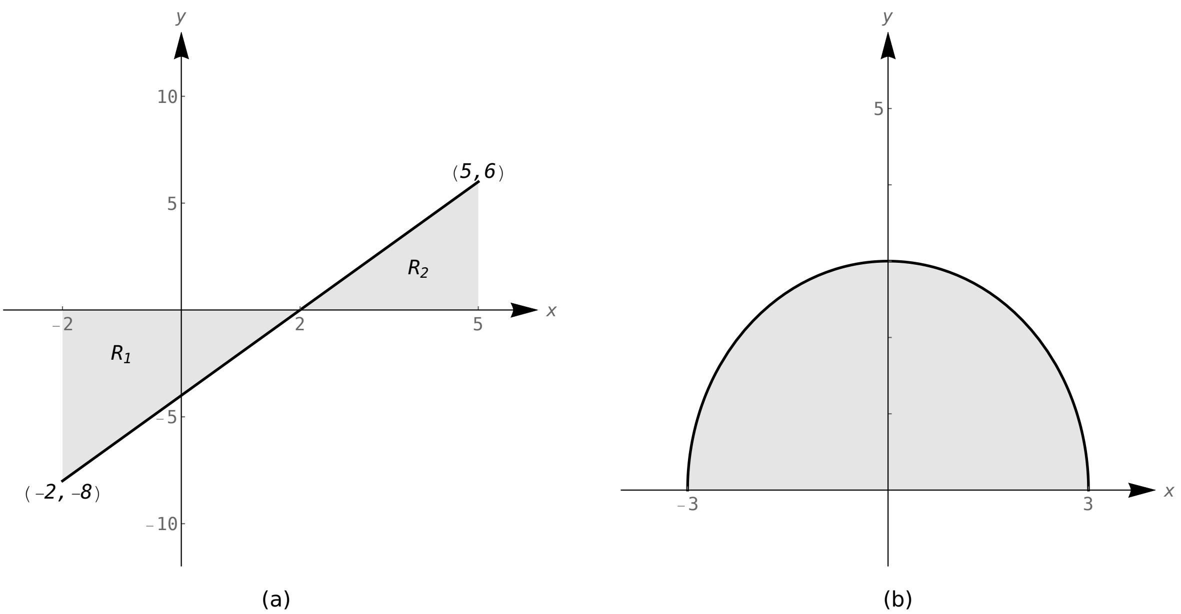


Figure 12.3: A graph of $f(x) = 2x-4$ in (a) and $f(x) = \sqrt{9-x^2}$ in (b), from Example 12.3.

2. Recognize that the integrand of this definite integral describes a half circle, as sketched in Figure 12.3(b), with radius 3. Thus the area is:

$$\int_{-3}^{3} \sqrt{9-x^2}\ dx = \frac{1}{2}\pi r^2 = \frac{9}{2}\pi.$$

12.2 Riemann sums

In our previous examples, we have either found the areas of regions that have nice geometric shapes or the areas were given to us. But what is, for instance, the area of a region below $y = x^2$? The function $y = x^2$ is relatively simple, yet the shape it defines has an area that is not simple to find geometrically. In this section we will explore how to find the areas of such regions.

12.2.1 Approximating areas

Consider the region given in Figure 12.4, which is the area under $y = 4x - x^2$ on $[0, 4]$. What is the signed area of this region – i.e., what is $\int_0^4 (4x - x^2)\, dx$? We start by approximating. We can surround the region with a rectangle with height and width of 4 and find the area is approximately 16 square units. This is obviously an over–approximation; we are including area in the rectangle that is not under the parabola.

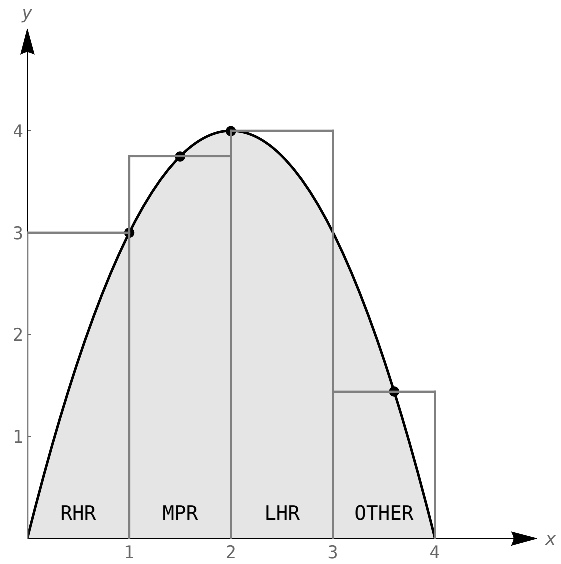


Figure 12.4: A graph of $f(x) = 4x - x^2$ and approximating $\int_0^4 (4x - x^2)\, dx$ using rectangles.

We have an approximation of the area, using one rectangle. How can we refine our approximation to make it better? The key to this section is this answer: use more rectangles. Let us use 4 rectangles with an equal width of 1. This partitions the interval $[0, 4]$ into 4 subintervals, $[0, 1]$, $[1, 2]$, $[2, 3]$ and $[3, 4]$. On each subinterval we will draw a rectangle.

There are three common ways to determine the height of these rectangles: the **left hand rule** (*linkerhand regel*), the **right hand rule** (*rechterhand regel*), and the **midpoint rule** (*midpoint regel*). The left hand rule says to evaluate the function at the left–hand endpoint of the subinterval and make the rectangle that height. In Figure 12.4, the rectangle drawn on the interval $[2, 3]$ has height determined by the left hand rule (LHR); it has a height of $f(2)$.

The right hand rule (RHR) says the opposite: on each subinterval, evaluate the function at the right endpoint and make the rectangle that height. In Figure 12.4, the rectangle drawn on $[0, 1]$ is drawn using $f(1)$ as its height. The midpoint rule (MPR) says to evaluate the function at the midpoint of each subinterval, and to make the rectangle that height. The rectangle drawn on $[1, 2]$ was made using the midpoint rule, with a height of $f(1.5)$.

These are the three most common rules for determining the heights of approximating rectangles, but one is not forced to use one of these three methods. The rectangle on $[3, 4]$ has a height of approximately $f(3.53)$, very close to the midpoint rule. It was chosen so that the area of the rectangle is exactly the area of the region under f on $[3,4]$.

It is hard to tell at this moment which is a better approximation. We can continue to refine our approximation by using more rectangles.

12.2.2 Riemann sums

Consider again $\int_0^4 (4x-x^2)\ dx$. We divide or partition the number line of $[0,4]$ into 16 equally spaced subintervals. We denote 0 as x_1 , so in general, we have

$$x_i = x_1 + (i-1)\Delta x\,,$$

where $i = 1,2,\ldots,16$ For the sake of simplicity, we will often write $\Delta x = \Delta x_i$, where Δx_i is the width of the i^{th} subinterval, whenever the width of the subintervals is the same.

Given any subdivision of $[0,4]$, the first subinterval is $[x_1,x_2]$; the second is $[x_2,x_3]$; the i^{th} subinterval is $[x_i,x_{i+1}]$. Hence, when using the left hand rule, the height of the i^{th} rectangle will be $f(x_i)$. When using the right hand rule, the height of the i^{th} rectangle will be $f(x_{i+1})$, and finally, when using the midpoint rule, the height of the i^{th} rectangle will be

$$f\left(\frac{x_i + x_{i+1}}{2}\right).$$

We illustrate this in the next example.

Example 12.4

Approximate

$$\int_0^4 (4x-x^2)\ dx$$

using the right hand rule with 16 and 1000 equally spaced intervals.

Solution

Using 16 equally spaced intervals and the right hand rule, we can approximate the definite integral as

$$\sum_{i=1}^{16} f(x_{i+1})\Delta x,$$

where we have $\Delta x = 4/16 = 0.25$. Moreover, since $x_1 = 0$, we have

$$\begin{aligned} x_{i+1} &= 0 + \big((i+1)-1\big)\Delta x \\ &= i\Delta x\,. \end{aligned}$$

Using summation formulas, we may now consider:

$$\int_0^4 (4x-x^2)\ dx \approx \sum_{i=1}^{16} f(x_{i+1})\Delta x = \sum_{i=1}^{16} f(i\Delta x)\Delta x$$

$$= \sum_{i=1}^{16}\big(4i\Delta x - (i\Delta x)^2\big)\Delta x = \sum_{i=1}^{16}(4i\Delta x^2 - i^2\Delta x^3)$$

$$= (4\Delta x^2)\sum_{i=1}^{16} i - \Delta x^3 \sum_{i=1}^{16} i^2 \tag{12.3}$$

$$= (4\Delta x^2)\frac{16\cdot 17}{2} - \Delta x^3 \frac{16(17)(33)}{6} = 10.625 \qquad (\Delta x = 0.25) \tag{12.4}$$

We were able to sum up the areas of 16 rectangles with very little computation. In Figure 12.5 the function and the 16 rectangles are graphed. While some rectangles over–approximate the area, other under–approximate the area by about the same amount. Thus our approximate area of 10.625 is likely a fairly good approximation.

For what concerns the approximation based on 1000 equally spaced, we can just use Equation (12.3); after replacing the 16's to 1000's and appropriately changing the value of Δx.

We do so here, skipping from the original summand to the equivalent of Equation (12.3) to save space. Note that $\Delta x = 4/1000 = 0.004$.

$$\begin{aligned}\int_0^4 (4x - x^2)\,dx &\approx \sum_{i=1}^{1000} f(x_{i+1})\Delta x \\ &= (4\Delta x^2)\sum_{i=1}^{1000} i - \Delta x^3 \sum_{i=1}^{1000} i^2 \\ &= (4\Delta x^2)\frac{1000\cdot 1001}{2} - \Delta x^3 \frac{1000(1001)(2001)}{6} \\ &= 10.666656\end{aligned}$$

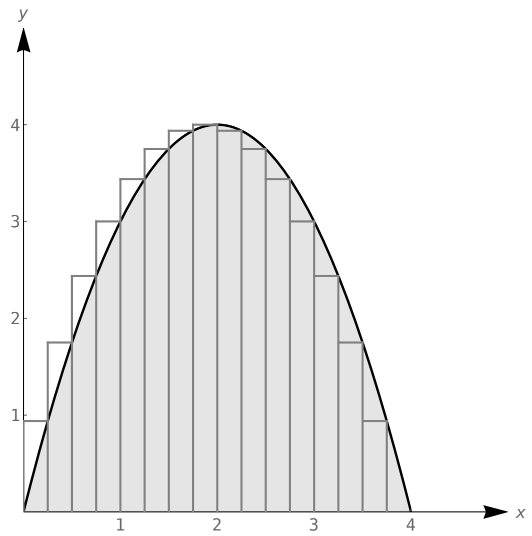


Figure 12.5: Approximating $\int_0^4 (4x - x^2)\,dx$ with the right hand rule and 16 evenly spaced subintervals.

Using many, many rectangles, we have a likely good approximation of

$\int_0^4 (4x - x^2)\Delta x$. That is,

$$\int_0^4 (4x - x^2)\, dx \approx 10.666656.$$

Instead of approximating a definite integral using rectangles of the same width and height determined by evaluating f at a particular point in each consecutive subinterval, we could partition an interval $[a, b]$ with subintervals that do not have the same size. We refer to the length of the i^{th} subinterval as Δx_i. Also, one could determine each rectangle's height by evaluating f at any point c_i in the i^{th} subinterval. Thus the height of the i^{th} subinterval would be $f(c_i)$, and the area of the i^{th} rectangle would then be $f(c_i)\Delta x_i$.

These ideas are formally defined below.

Definitie 12.3 (Partition)
A **partition** (*partitie*) of a closed interval $[a, b]$ is a set of numbers $x_1, x_2, \ldots x_{n+1}$ where

$$a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b.$$

The length of the i^{th} subinterval, $[x_i, x_{i+1}]$, is $\Delta x_i = x_{i+1} - x_i$. If $[a, b]$ is partitioned into subintervals of equal length, we let Δx_i represent the length of each subinterval.

The size of the partition, denoted $\mathcal{L}$, is the length of the largest subinterval of the partition, i.e. $\mathcal{L} = \max_i (\Delta x_i)$.

Summations of rectangles with area $f(c_i)\Delta x_i$ are named after mathematician Georg Friedrich Bernhard Riemann, as given in the following definition.

Definitie 12.4 (Riemann sum)
Let f be defined on a closed interval $[a, b]$, let $\{x_1, x_2, \ldots, x_{n+1}\}$ be a partition of $[a, b]$

and let c_i denote any value in the i^{th} subinterval.

The sum

$$\sum_{i=1}^{n} f(c_i)\Delta x_i$$

is a **Riemann sum** (*Riemann som*) of f on $[a, b]$.

Usually Riemann sums are calculated using one of the three methods we have introduced. The uniformity of construction makes computations easier. So

$$\int_a^b f(x)\, dx$$

is typically approximated by means of the following Riemann sum

$$\sum_{i=1}^{n} f(c_i)\Delta x_i\,,$$

for which we take the following steps.

1. Divide the interval $[a,b]$ in n subintervals have equal length, such that

$$\Delta x_i = \Delta x = \frac{b-a}{n}$$

and the i^{th} term of the equally spaced partition is

$$x_i = a + (i-1)\Delta x\,.$$

Thus $x_1 = a$ and $x_{n+1} = b$.

2. Evaluate one of the following summations:

(a) using the left hand rule we get the so-called **left Riemann sum** (*linker Riemann som*):

$$\sum_{i=1}^{n} f(x_i)\Delta x,$$

(b) using the right hand rule we get the so-called **right Riemann sum** (*rechter Riemann som*):

$$\sum_{i=1}^{n} f(x_{i+1})\Delta x,$$

(c) and using the midpoint rule we get the **middle Riemann sum** (*midden Riemann som*):

$$\sum_{i=1}^{n} f\left(\frac{x_i + x_{i+1}}{2}\right)\Delta x\,.$$

12.2.3 Limits of Riemann sums

We have used limits to evaluate given definite integrals. Will this always work? We will show, given not–very–restrictive conditions, that yes, it will always work.

The previous example has shown us how we can think of a summation as a function of n. More precisely, given a definite integral $\int_a^b f(x)\,dx$, we let:

- $S_L(n) = \sum_{i=1}^{n} f(x_i)\Delta x$, be the left Riemann sum,
- $S_R(n) = \sum_{i=1}^{n} f(x_{i+1})\Delta x$, be the right Riemann sum,
- $S_M(n) = \sum_{i=1}^{n} f\left(\frac{x_i + x_{i+1}}{2}\right)\Delta x$, be the sum of equally spaced rectangles formed using the midpoint rule,

and likewise for the lower and upper Riemann sums. Now, recall that the definition of the limit $\lim_{n\to+\infty} S_L(n) = K$ implies that given any $\epsilon > 0$, there exists $N > 0$ such that

$$\left|S_L(n) - K\right| < \epsilon,$$

when $n \geq N$.

The following theorem states that we can use any of our three rules to find the exact value of a definite integral.

Theorem 12.4 (Definite integrals and the limit of Riemann sums)
Let f be continuous on the closed interval $[a,b]$ and let $S_L(n)$, $S_R(n)$, $S_M(n)$, Δx, Δx_i and c_i be defined as before. Then:

1. $\lim\limits_{n\to+\infty} S_L(n) = \lim\limits_{n\to+\infty} S_R(n) = \lim\limits_{n\to+\infty} S_M(n) = \lim\limits_{n\to+\infty} \sum\limits_{i=1}^{n} f(c_i)\Delta x$,

2. $\lim\limits_{n\to+\infty} \sum\limits_{i=1}^{n} f(c_i)\Delta x = \int\limits_a^b f(x)\ dx$, and

3. $\lim\limits_{\mathcal{L}\to 0} \sum\limits_{i=1}^{n} f(c_i)\Delta x_i = \int\limits_a^b f(x)\ dx$.

This theorem also goes two steps further. It states that the height of each rectangle does not have to be determined following a specific rule, but could be $f(c_i)$, where c_i is any point in the i^{th} subinterval. Furthermore, it goes on to state that the rectangles do not need to be of the same width.

Let $\mathcal{L}$ represent the length of the largest subinterval in the partition: that is, $\mathcal{L}$ is the largest of all the Δx_i's. If $\mathcal{L}$ is small, then $[a,b]$ must be partitioned into many subintervals, since all subintervals must have small lengths. Taking the limit as $\mathcal{L}$ goes to zero implies that the number n of subintervals in the partition is growing to infinity, as the largest subinterval length is becoming arbitrarily small. We then interpret the expression

$$\lim_{\mathcal{L}\to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i$$

as the limit of the sum of the areas of rectangles, where the width of each rectangle can be different but getting small, and the height of each rectangle is not necessarily determined by a particular rule. The theorem states that this Riemann sum also gives the value of the definite integral of f over $[a,b]$.

We now know of a way to evaluate a definite integral using limits; in the next section we will see how the fundamental theorem of calculus makes the process simpler. The key feature of this theorem is its connection between the indefinite integral and the definite integral.

12.3 The fundamental theorem of calculus

12.3.1 Mean value theorem for definite integrals

Consider the graph of a function f in Figure 12.6(a) and the area defined by $\int_1^4 f(x)\ dx$. In Figure 12.6(b), the height of the rectangle is greater than f on $[1,4]$, hence the area of this rectangle is greater than $\int_1^4 f(x)\ dx$. In Figure 12.6(c), the height of the rectangle is smaller than f on $[1,4]$, hence the area of this rectangle is less than $\int_1^4 f(x)\ dx$. Finally, in Figure 12.6(d) the height of the rectangle is such that the area of the rectangle is exactly that of $\int_1^4 f(x)\ dx$. Since rectangles that are too big, as in Figure 12.6(b), and rectangles that are too little, as in Figure 12.6(c), give areas greater/lesser than $\int_1^4 f(x)\ dx$, it makes sense that there is a rectangle, whose top intersects $f(x)$ somewhere on $[1,4]$, whose area is exactly that of the definite integral.

We state this idea formally in a theorem.

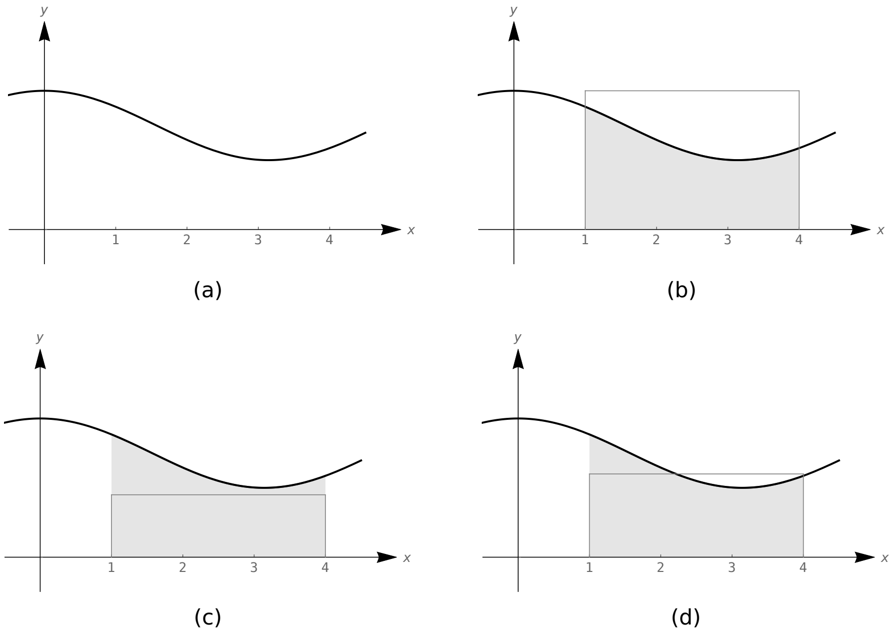


Figure 12.6: The graph of a function f (a) and differently sized rectangles give upper and lower bounds on $\int_1^4 f(x)\,dx$ (b-c).

Theorem 12.5 (The mean value theorem of integration)
Let f be continuous on $[a,b]$. There exists a value c in $]a,b[$ such that

$$\int_a^b f(x)\,dx = f(c)(b-a).$$

This is an existential statement; c exists, but we do not provide a method of finding it. Theorem 12.5 is directly connected to the mean value theorem of differentiation (Theorem 10.4).

12.3.2 Main theorems

Let $f(t)$ be a continuous function defined on $[a,b]$. The definite integral $\int_a^b f(x)\,dx$ is the area under f on $[a,b]$. We can turn this concept into a function by letting the upper (or lower) bound vary.

Let $F(x) = \int_a^x f(t)\,dt$. It computes the area under f on $[a,x]$ as illustrated in Figure 12.7. We can study this function using our knowledge of the definite integral.

We can also apply calculus ideas to $F(x)$; in particular, we can compute its derivative. While this may seem like an innocuous thing to do, it has far–reaching implications, as demonstrated by the fact that the result is given as an important theorem.

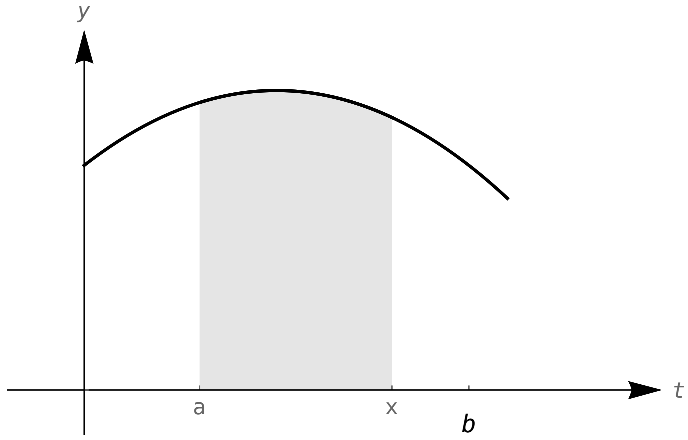


Figure 12.7: The area of the shaded region is $F(x) = \int_a^x f(t)\,dt$.

Theorem 12.6 (The fundamental theorem of calculus, Part 1)
Let f be continuous on $[a,b]$ and let $F(x) = \int_a^x f(t)\,dt$. Then F is a differentiable function on $]a,b[$, and

$$F'(x) = f(x).$$

To illustrate this theorem, let us consider

$$F(x) = \int_{-5}^{x} (t^2 + \sin(t))\,dt$$

and try to find $F'(x)$.

Using Theorem 12.6, we immediately have $F'(x) = x^2 + \sin(x)$. This simple example reveals that $F(x)$ is an antiderivative of $x^2 + \sin(x)$! Therefore, $F(x) = x^3/3 - \cos(x) + C$ for some value of C. We have done more, however, than found a complicated way of computing an antiderivative. Consider a function f defined on an open interval containing a, b and c. Suppose we want to compute $\int_a^b f(t)\,dt$. First, let

$$F(x) = \int_c^x f(t)\,dt.$$

Using the properties of the definite integral (Theorem 12.3), we know

$$\begin{aligned}
\int_a^b f(t)\,dt &= \int_a^c f(t)\,dt + \int_c^b f(t)\,dt \\
&= -\int_c^a f(t)\,dt + \int_c^b f(t)\,dt \\
&= -F(a) + F(b) \\
&= F(b) - F(a).
\end{aligned}$$

We now see how indefinite integrals and definite integrals are related: we can evaluate a definite integral using antiderivatives. This is the second part of the fundamental theorem of calculus.

Theorem 12.7 (The fundamental theorem of calculus, Part 2)
Let f be continuous on $[a,b]$ and let F be any antiderivative of f. Then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Example 12.5

We spent a great deal of time in the previous section studying $\int_0^4 (4x - x^2)\,dx$. Using the fundamental theorem of calculus, evaluate this definite integral.

Solution

We need an antiderivative of $f(x) = 4x - x^2$. All antiderivatives of f have the form

$$F(x) = 2x^2 - \frac{1}{3}x^3 + C;$$

for simplicity, choose $C = 0$.

The fundamental theorem of calculus states

$$\int_0^4 (4x - x^2)\,dx = F(4) - F(0) = \left(2(4)^2 - \frac{1}{3}4^3 - (0-0)\right) = 32 - \frac{64}{3} = \frac{32}{3}.$$

This is the same answer we obtained using limits in the previous section, just with much less work.

A special notation is often used in the process of evaluating definite integrals using the fundamental theorem of calculus. Instead of explicitly writing $F(b) - F(a)$, the notation $F(x)\Big|_a^b$ is used. Also note that any antiderivative $F(x)$ can be chosen when using the fundamental theorem of calculus to evaluate a definite integral, meaning any value of C can be picked. The constant always cancels out of the expression when evaluating $F(b) - F(a)$, so it does not matter what value is picked. This being the case, we might as well let $C = 0$.

Example 12.6

Evaluate the following definite integrals.

1. $\displaystyle\int_{-2}^{2} x^3\,dx$
2. $\displaystyle\int_0^{\pi} \sin(x)\,dx$
3. $\displaystyle\int_0^5 e^t\,dt$
4. $\displaystyle\int_4^9 \sqrt{u}\,du$
5. $\displaystyle\int_1^5 2\,dx$

Solution

1. $\displaystyle\int_{-2}^{2} x^3\,dx = \frac{1}{4}x^4\Big|_{-2}^{2} = \left(\frac{1}{4}2^4\right) - \left(\frac{1}{4}(-2)^4\right) = 0.$

2. $\displaystyle\int_0^{\pi} \sin(x)\,dx = -\cos(x)\Big|_0^{\pi} = -\cos(\pi) - (-\cos 0) = 1 + 1 = 2.$ So, the area under one hump of

a sine curve is 2.

3. $\displaystyle\int_0^5 e^t\ dt = e^t\Big|_0^5 = e^5 - e^0 = e^5 - 1 \approx 147.41.$

4. $\displaystyle\int_4^9 \sqrt{u}\ du = \int_4^9 u^{\frac{1}{2}}\ du = \frac{2}{3}u^{\frac{3}{2}}\Big|_4^9 = \frac{2}{3}\left(9^{\frac{3}{2}} - 4^{\frac{3}{2}}\right) = \frac{2}{3}(27-8) = \frac{38}{3}.$

5. $\displaystyle\int_1^5 2\ dx = 2x\Big|_1^5 = 2(5) - 2 = 2(5-1) = 8.$

This last integral in Example 12.6 is interesting; the integrand is a constant function, hence we are finding the area of a rectangle with width $(5-1) = 4$ and height 2. Notice how the evaluation of the definite integral led to $2(4) = 8$.

In general, if c is a constant, then

$$\int_a^b c\ dx = c(b-a).$$

12.3.3 Motion and the fundamental theorem of calculus

We established, starting in Section 9.1.4, that the derivative of a position function is a velocity function, and the derivative of a velocity function is an acceleration function. Now consider definite integrals of velocity and acceleration functions. Specifically, if $v(t)$ is a velocity function, what does $\int_a^b v(t)\ dt$ mean?

The fundamental theorem of calculus states that

$$\int_a^b v(t)\ dt = V(b) - V(a),$$

where $V(t)$ is any antiderivative of $v(t)$. Since $v(t)$ is a velocity function, $V(t)$ must be a position function, and $V(b) - V(a)$ measures a change in position, or **displacement** (*verplaatsing*).

Example 12.7

A ball is thrown straight up with velocity given by $v(t) = -32t + 20$m/s, where t is measured in seconds. Find, and interpret, $\int_0^1 v(t)\ dt$.

Solution

Using the fundamental theorem of calculus, we have

$$\begin{aligned}\int_0^1 v(t)\ dt &= \int_0^1 (-32t+20)\ dt \\ &= -16t^2 + 20t\Big|_0^1 \\ &= 4.\end{aligned}$$

Thus if a ball is thrown straight up into the air with velocity $v(t) = -32t + 20$, the height of the ball, 1 second later, will be 4 metres above the initial height.

Integrating a rate of change function gives total change. Velocity is the rate of position change; integrating velocity gives the total change of position, i.e., displacement.

Integrating a speed function gives a similar, though different, result. Speed is also the rate of position change, but does not account for direction. So integrating a speed function gives total change of position, without the possibility of negative position change. Hence the integral of a speed function gives **distance travelled** (*afgelegde afstand*).

12.3.4 The fundamental theorem of calculus and the chain rule

Using other notation, we may write Part 1 of the fundamental theorem of calculus as

$$\frac{d}{dx}\big(F(x)\big) = f(x).$$

While we have just practised evaluating definite integrals, sometimes finding antiderivatives is impossible and we need to rely on other techniques to approximate the value of a definite integral. Functions written as

$$F(x) = \int_a^x f(t)\ dt$$

are useful in such situations.

It may be of further use to compose such a function with another. As an example, we may compose $F(x)$ with $g(x)$ to get

$$F\big(g(x)\big) = \int_a^{g(x)} f(t)\ dt.$$

What is the derivative of such a function? The chain rule can be employed to find

$$\frac{d}{dx}\Big(F\big(g(x)\big)\Big) = F'\big(g(x)\big)g'(x) = f\big(g(x)\big)g'(x).$$

An example will help us understand this.

Example 12.8

Find the derivative of

1. $F(x) = \displaystyle\int_2^{x^2} \ln(t)\ dt$

2. $F(x) = \displaystyle\int_{\cos(x)}^{5} t^3\ dt.$

Solution

1. We can view $F(x)$ as being the function $G(x) = \int_2^x \ln(t)\,dt$ composed with $h(x) = x^2$; that is, $F(x) = G\big(h(x)\big)$. The fundamental theorem of calculus states that $G'(x) = \ln(x)$. The chain rule gives us

$$\begin{aligned} F'(x) &= G'\big(h(x)\big)h'(x) \\ &= \ln(h(x))h'(x) \\ &= \ln(x^2)2x \\ &= 2x\ln(x^2). \end{aligned}$$

Normally, of course, the steps defining $G(x)$ and $h(x)$ are skipped.

2. Note that $F(x) = -\int_5^{\cos(x)} t^3\,dt$. Viewed this way, the derivative of F is straightforward:

$$F'(x) = \sin(x)\cos^3(x).$$

12.3.5 Average value

Recognize that the mean value theorem can be rewritten as

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\,dx,$$

for some value of c in $[a,b]$. Next, partition the interval $[a,b]$ into n equally spaced subintervals, $a = x_1 < x_2 < \ldots < x_{n+1} = b$ and choose any c_i in $[x_i, x_{i+1}]$. The average of the numbers $f(c_1)$, $f(c_2)$, $\ldots$, $f(c_n)$ is:

$$\frac{1}{n}\Big(f(c_1)+f(c_2)+\ldots+f(c_n)\Big) = \frac{1}{n}\sum_{i=1}^n f(c_i).$$

Multiply this last expression by 1 in the form of $\frac{(b-a)}{(b-a)}$:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n f(c_i) &= \sum_{i=1}^n f(c_i)\frac{1}{n}\frac{(b-a)}{(b-a)} \\ &= \frac{1}{b-a}\sum_{i=1}^n f(c_i)\frac{b-a}{n} \\ &= \frac{1}{b-a}\sum_{i=1}^n f(c_i)\Delta x \end{aligned}$$

where $\Delta x = (b-a)/n$. Now take the limit as $n \to +\infty$:

$$\lim_{n\to+\infty}\frac{1}{b-a}\sum_{i=1}^n f(c_i)\Delta x = \frac{1}{b-a}\int_a^b f(x)\,dx = f(c).$$

This tells us this: when we evaluate f at n (somewhat) equally spaced points in $[a,b]$, the average value of these samples is $f(c)$ as $n \to +\infty$.

This leads us to a definition.

Definitie 12.5 (The average value of *f* on *[a,b]*)
Let f be continuous on $[a,b]$. The **average value of** f (*gemiddelde functiewaarde*) on $[a,b]$ is $f(c)$, where c is a value in $[a,b]$ guaranteed by the mean value theorem. I.e.,

$$\text{Average Value of } f \text{ on } [a,b] = \frac{1}{b-a}\int_a^b f(x)\, dx.$$

An application of this definition is given in the following example.

Example 12.9

An object moves back and forth along a straight line with a velocity given by $v(t) = (t-1)^2$ on $[0,3]$, where t is measured in seconds and $v(t)$ is measured in m/s.

1. What is the average velocity of the object?
2. What was the displacement of the object?

Solution

1. By Definition 12.5, the average velocity is:

$$\frac{1}{3-0}\int_0^3 (t-1)^2\, dt = \frac{1}{3}\int_0^3 \left(t^2-2t+1\right) dt = \frac{1}{3}\left(\frac{1}{3}t^3 - t^2 + t\right)\Big|_0^3 = 1 \text{ m/s}.$$

2. We calculate this by integrating its velocity function: $\int_0^3 (t-1)^2\, dt = 3$ m. Its final position was 3 meter from its initial position after 3 seconds: its average velocity was 1 m/s.

This section has laid the groundwork for a lot of great mathematics to follow. The most important lesson is this: definite integrals can be evaluated using antiderivatives. Since the previous section established that definite integrals are the limit of Riemann sums, we can later create Riemann sums to approximate values other than area under the curve, convert the sums to definite integrals, then evaluate these using the fundamental theorem of calculus. This will allow us to compute the work done by a variable force, the volume of certain solids, the arc length of curves, and more.

The downside is this: generally speaking, computing antiderivatives is much more difficult than computing derivatives. For that reason, we will see in Section 12.6.2 how to approximate the value of definite integrals beyond using the left hand, right hand and midpoint rules. These techniques are invaluable when antiderivatives cannot be computed, or when the actual function f is unknown and all we know is the value of f at certain x-values. But first, we will study techniques of finding antiderivatives analytically so that a wide variety of definite integrals can be evaluated.

12.4 Techniques of antidifferentiation

This chapter is devoted to exploring techniques of antidifferentiation. While not every function has an antiderivative in terms of elementary functions like polynomial, exponential or trigonometric functions, we can still find antiderivatives of a wide variety of functions. Nowadays there are also several websites that allow you to calculate integrals with steps[1].

[1] https://www.integral-calculator.com/

12.4.1 Substitution

12.4.1.1 Rationale

Essentially, integration by **substitution** (*substitutie*) allows us to undo the chain rule. Its underlying principle is to rewrite a complicated integral of the form $\int f(x)\ dx$ as a not–so–complicated integral $\int h(u)\ du$.

For instance, consider

$$\int (20x+30)(x^2+3x-5)^9\ dx.$$

Arguably the most complicated part of the integrand is $(x^2+3x-5)^9$. We wish to make this simpler; we do so through a substitution. Let $u = x^2+3x-5$. Thus

$$(x^2+3x-5)^9 = u^9.$$

We have established u as a function of x, so now consider the differential of u:

$$du = (2x+3)dx.$$

Let us return now to the original integral and do some substitutions through algebra:

$$\begin{aligned}
\int (20x+30)(x^2+3x-5)^9\ dx &= \int 10(2x+3)(x^2+3x-5)^9\ dx \\
&= \int 10(\underbrace{x^2+3x-5}_{u})^9\ \underbrace{(2x+3)\ dx}_{du} \\
&= \int 10u^9\ du \\
&= u^{10}+C \qquad \text{(Replace } u \text{ with } x^2+3x-5.\text{)} \\
&= (x^2+3x-5)^{10}+C
\end{aligned}$$

In general, let $F(x)$ and $g(x)$ be differentiable functions and consider the derivative of their composition:

$$\frac{d}{dx}\Big(F(g(x))\Big) = F'(g(x))g'(x).$$

Thus

$$\int F'(g(x))g'(x)\ dx = F(g(x))+C.$$

Integration by substitution works by recognizing the inside function $g(x)$ and replacing it with a variable. By setting $u = g(x)$, we can rewrite the derivative as

$$\frac{d}{dx}\Big(F(u)\Big) = F'(u)u'.$$

Since $du = g'(x)dx$, we can rewrite the above integral as

$$\int F'(g(x))g'(x)\ dx = \int F'(u)du = F(u)+C = F(g(x))+C.$$

The point of substitution is to make the integration step easy. Indeed, the step $\int F'(u)\ du = F(u)+C$ looks easy, as the antiderivative of the derivative of F is just F, plus a constant. The work involved is making the proper substitution. There is not a step–by–step process that one can memorize; rather, experience will be one's guide. To gain experience, we now embark on some examples.

Example 12.10

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{7}{-3x+1}\,dx,$ 2. $\displaystyle\int x\sin(x^2+5)\,dx,$ 3. $\displaystyle\int x\sqrt{x+3}\,dx.$

Solution

1. View the integrand as the composition of functions $f(g(x))$, where $f(x)=7/x$ and $g(x)=-3x+1$. Then, we let $u=-3x+1$, the inside function. Thus $du=-3dx$. The integrand lacks a -3; hence divide the previous equation by -3 to obtain $-du/3=dx$. We can now evaluate the integral through substitution.

$$\begin{aligned}\int \frac{7}{-3x+1}\,dx &= \int \frac{7}{u}\frac{du}{(-3)}\\ &= \frac{-7}{3}\int \frac{du}{u}\\ &= \frac{-7}{3}\ln|u|+C\\ &= -\frac{7}{3}\ln|-3x+1|+C.\end{aligned}$$

2. We choose to let u be the inside function of $\sin(x^2+5)$. So, let $u=x^2+5$, hence $du=2x\,dx$. The integrand has an $x\,dx$ term, but not a $2x\,dx$ term. We can divide both sides of the du expression by 2:

$$du=2x\,dx \quad\Rightarrow\quad \frac{1}{2}du=x\,dx.$$

We can now substitute.

$$\begin{aligned}\int x\sin(x^2+5)\,dx &= \int \sin(\underbrace{x^2+5}_{u})\underbrace{x\,dx}_{\frac{1}{2}du}\\ &= \int \frac{1}{2}\sin(u)\,du\\ &= -\frac{1}{2}\cos(u)+C \qquad \text{(Now replace } u \text{ with } x^2+5.\text{)}\\ &= -\frac{1}{2}\cos(x^2+5)+C.\end{aligned}$$

Thus

$$\int x\sin(x^2+5)\,dx = -\frac{1}{2}\cos(x^2+5)+C.$$

3. Recognizing the composition of functions, set $u=x+3$. Then $du=dx$, giving what seems

initially to be a simple substitution. But at this stage, we have:

$$\int x\sqrt{x+3}\ dx = \int x\sqrt{u}\ du.$$

We cannot evaluate an integral that has both an x and an u in it. We need to convert the x to an expression involving just u.

Since we set $u = x+3$, we can also state that $u-3 = x$. Thus we can replace x in the integrand with $u-3$. It will also be helpful to rewrite $\sqrt{u}$ as $u^{\frac{1}{2}}$.

$$\begin{aligned}
\int x\sqrt{x+3}\ dx &= \int (u-3)u^{\frac{1}{2}}\ du \\
&= \int \left(u^{\frac{3}{2}} - 3u^{\frac{1}{2}}\right) du \\
&= \frac{2}{5}u^{\frac{5}{2}} - 2u^{\frac{3}{2}} + C \\
&= \frac{2}{5}(x+3)^{\frac{5}{2}} - 2(x+3)^{\frac{3}{2}} + C
\end{aligned}$$

12.4.1.2 Integrals involving trigonometric functions

Integration by substitution can also be used to unveil the antiderivatives of the tangent, cotangent, secant and cosecant. For instance, consider the following example concerning the former function.

Example 12.11

Evaluate

$$\int \tan(x)\ dx.$$

Solution

Rewrite $\tan(x)$ as $\sin(x)/\cos(x)$. While the presence of a composition of functions may not be immediately obvious, recognize that $\cos(x)$ is inside the $1/x$ function. Therefore, we see if setting $u = \cos(x)$ returns usable results. We have that $du = -\sin(x)\ dx$, hence $-du = \sin(x)\ dx$. We can integrate:

$$\begin{aligned}
\int \tan(x)\ dx &= \int \frac{\sin(x)}{\cos(x)}\ dx \\
&= \int \underbrace{\frac{1}{\cos(x)}}_{1/u}\underbrace{\sin(x)\ dx}_{-du} \\
&= \int \frac{-1}{u}\ du \\
&= -\ln|u| + C \\
&= -\ln|\cos(x)| + C.
\end{aligned}$$

This can be simplified even further by bringing the −1 inside the logarithm as a power of $\cos(x)$, as in:

$$\begin{aligned} -\ln|\cos(x)|+C &= \ln|(\cos(x))^{-1}|+C \\ &= \ln\left|\frac{1}{\cos(x)}\right|+C \\ &= \ln|\sec(x)|+C. \end{aligned}$$

Thus the result they give is $\int \tan(x)\,dx = \ln|\sec(x)|+C$.

We can use similar techniques in Example 12.11 to find antiderivatives of the other trigonometric functions. In this way, one finds:

1. $\displaystyle\int \tan(x)\,dx = -\ln|\cos(x)|+C$
2. $\displaystyle\int \cot(x)\,dx = \ln|\sin(x)|+C$
3. $\displaystyle\int \sec(x)\,dx = \ln|\sec(x)+\tan(x)|+C$
4. $\displaystyle\int \csc(x)\,dx = -\ln|\csc(x)+\cot(x)|+C$

Using the power-reducing formulas we have seen in Chapter 5 (Theorem 5.12), we can also tackle integrals involving powers of trigonometric functions.

Example 12.12

Evaluate

$$\int \cos^2(x)\,dx.$$

Solution

We have a composition of functions as $\cos^2(x) = \big(\cos(x)\big)^2$. However, setting $u = \cos(x)$ means $du = -\sin(x)\,dx$, which we do not have in the integral. So, let us use Theorem 5.12, which states

$$\cos^2(x) = \frac{1+\cos(2x)}{2}.$$

The right hand side of this equation is not difficult to integrate. We have:

$$\begin{aligned} \int \cos^2(x)\,dx &= \int \frac{1+\cos(2x)}{2}\,dx \\ &= \int \left(\frac{1}{2}+\frac{1}{2}\cos(2x)\right)dx. \end{aligned}$$

So, we easily find:

$$\begin{aligned} &= \frac{1}{2}x+\frac{1}{2}\frac{\sin(2x)}{2}+C \\ &= \frac{1}{2}x+\frac{\sin(2x)}{4}+C. \end{aligned}$$

We will make significant use of this power–reducing technique in future sections.

12.4.1.3 Integrals leading to inverse trigonometric functions

When studying derivatives of inverse functions, we learned that

$$\frac{d}{dx}\big(\arctan(x)\big) = \frac{1}{1+x^2}.$$

Applying the chain rule to this is not difficult. For instance, in general, we have

$$\frac{d}{dx}\big(\arctan(ax)\big) = \frac{a}{1+a^2x^2}.$$

This result can be used to evaluate

$$\int \frac{1}{a^2+x^2}\ dx$$

.

For that purpose, we rewrite this integral as

$$\frac{1}{a^2}\int \frac{1}{1+\left(\frac{x}{a}\right)^2}\ dx.$$

This can now be integrated using substitution. Set $u = x/a$, hence $du = dx/a$ or $dx = a du$. Thus

$$\begin{aligned}\int \frac{1}{a^2+x^2}\ dx &= \frac{1}{a}\int \frac{1}{1+u^2}\ du \\ &= \frac{1}{a}\arctan(u) + C \\ &= \frac{1}{a}\arctan\left(\frac{x}{a}\right) + C\end{aligned}$$

This demonstrates a general technique that can be applied to other integrands that result in inverse trigonometric functions. More specifically, for $a > 0$, we have

$$\int \frac{1}{a^2+x^2}\ dx = \frac{1}{a}\arctan\left(\frac{x}{a}\right) + C \tag{12.5}$$

$$\int \frac{1}{\sqrt{a^2-x^2}}\ dx = \arcsin\left(\frac{x}{a}\right) + C \tag{12.6}$$

Example 12.13

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{1}{x^2-4x+13}\ dx,$
2. $\displaystyle\int \frac{4-x}{\sqrt{16-x^2}}\ dx,$

Solution

1. We start by completing the square in the denominator, i.e.

$$\frac{1}{x^2-4x+13} = \frac{1}{(x-2)^2+9}$$

We can now integrate, to arrive at

$$\int \frac{1}{x^2-4x+13}\ dx = \int \frac{1}{(x-2)^2+9}\ dx = \frac{1}{3}\arctan\left(\frac{x-2}{3}\right)+C.$$

2. This integral requires two different methods to evaluate it. We get to those methods by splitting up the integral:

$$\int \frac{4-x}{\sqrt{16-x^2}}\ dx = \int \frac{4}{\sqrt{16-x^2}}\ dx - \int \frac{x}{\sqrt{16-x^2}}\ dx.$$

The first integral is easy to handle; the second integral is handled by substitution, with $u = 16-x^2$. We handle each separately.

$\displaystyle\int \frac{4}{\sqrt{16-x^2}}\ dx = 4\arcsin\left(\frac{x}{4}\right)+C.$

$\displaystyle\int \frac{x}{\sqrt{16-x^2}}\ dx$: Set $u = 16-x^2$, so $du = -2x dx$ and $x dx = -du/2$. We have

$$\begin{aligned}\int \frac{x}{\sqrt{16-x^2}}\ dx &= \int \frac{-du/2}{\sqrt{u}} \\ &= -\frac{1}{2}\int \frac{1}{\sqrt{u}}\ du \\ &= -\sqrt{u}+C \\ &= -\sqrt{16-x^2}+C.\end{aligned}$$

Combining these together, we have

$$\int \frac{4-x}{\sqrt{16-x^2}}\ dx = 4\arcsin\left(\frac{x}{4}\right)+\sqrt{16-x^2}+C.$$

12.4.1.4 Substitution and definite integration

Definite integrals that require substitution can be calculated using the following workflow:

1. Start with a definite integral $\int_a^b f(x)\ dx$ that requires substitution.
2. Ignore the bounds; use substitution to evaluate $\int f(x)\ dx$ and find an antiderivative $F(x)$.
3. Evaluate $F(x)$ at the bounds; that is, evaluate $F(x)\Big|_a^b = F(b)-F(a)$.

This workflow works fine, but substitution offers an alternative that is powerful and time saving. Since substitution converts integrals of the form $\int F'(g(x))g'(x)\,dx$ into an integral of the form $\int F'(u)\,du$ with the substitution of $u = g(x)$, we just have to appropriately change the bounds of a definite integral, i.e.

$$\int_a^b F'\big(g(x)\big)g'(x)\ dx = \int_{g(a)}^{g(b)} F'(u)\ du.$$

This indicates that once you convert to integrating with respect to u, you do not need to switch back to evaluating with respect to x.

Example 12.14

Evaluate

$$\int_0^{\pi/2} \sin(x)\cos(x)\ dx.$$

Solution

Let $u = g(x) = \cos(x)$, giving $du = -\sin(x)\ dx$ and hence $\sin(x)\ dx = -du$. The new upper bound is $g(\pi/2) = 0$; the new lower bound is $g(0) = 1$. Note how the lower bound is actually larger than the upper bound now. We have

$$\begin{aligned}\int_0^{\pi/2} \sin(x)\cos(x)\ dx &= \int_1^0 -u\ du\\ &= \int_0^1 u\ du\\ &= \frac{1}{2}u^2\Big|_0^1 = \frac{1}{2}.\end{aligned}$$

In Figure 12.8 we have graphed the two regions defined by our definite integrals. They bear no resemblance to each other, but they have the same area.

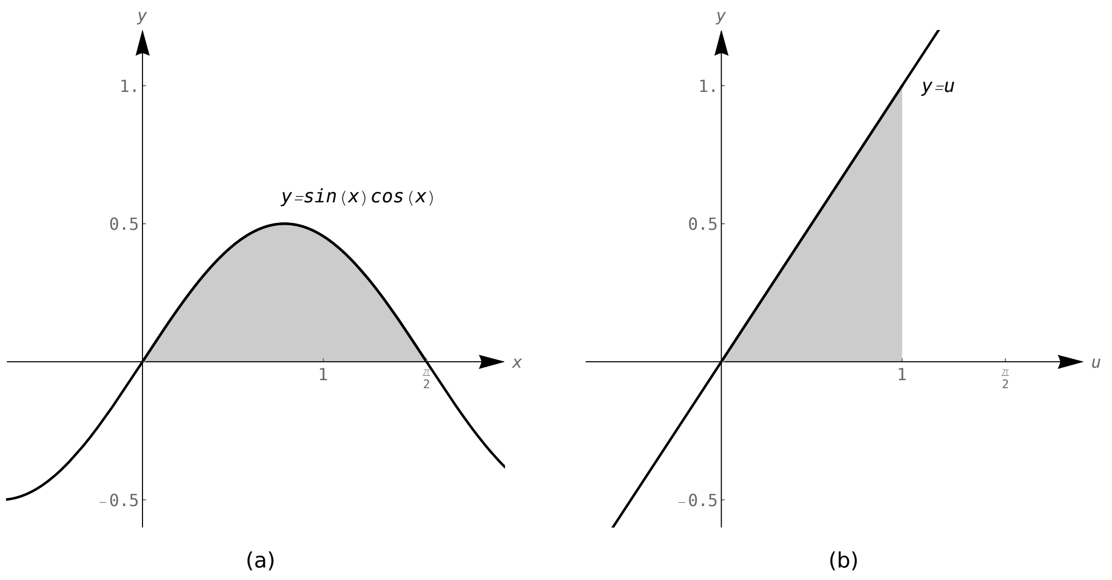


Figure 12.8: Graphing the areas defined by the definite integrals of Example 12.14.

12.4.2 Integration by parts

Here is a simple integral that we can not yet evaluate:

$$\int x\cos(x)\,dx.$$

It's a simple matter to take the derivative of the integrand using the product rule, but there is no such rule for integrals. However, this section introduces **integration by parts** (*partiële integratie*), a method of integration that is based on the product rule for derivatives. It will enable us to evaluate this integral.

The product rule says that if u and v are functions of x, then $(uv)' = u'v + uv'$. For simplicity, we have written u for $u(x)$ and v for $v(x)$. Suppose we integrate both sides with respect to x. This gives

$$\int (uv)'\,dx = \int (u'v + uv')\,dx.$$

By the fundamental theorem of calculus, the left side integrates to uv. The right side can be broken up into two integrals, and we have

$$uv = \int u'v\,dx + \int uv'\,dx.$$

Solving for the second integral we have

$$\int uv'\,dx = uv - \int u'v\,dx.$$

Using differential notation, we can write $du = u'(x)dx$ and $dv = v'(x)dx$ and the expression above can be written as follows:

$$\int u\,dv = uv - \int v\,du. \tag{12.7}$$

If our problem concerns a definite integral, we likewise arrive at

$$\int_{x=a}^{x=b} u\;dv = uv\Big|_a^b - \int_{x=a}^{x=b} v\;du.$$

Typically, we try to identify u and dv in the integral we are given, and the key is that we usually want to choose u and dv so that du is simpler than u and v is hopefully not too much more complicated than dv. This will mean that the integral on the right side of the integration by parts formula, $\int v\,du$ will be simpler to integrate than the original integral $\int u\,dv$.

Example 12.15

Evaluate the following indefinite integrals:

1. $\int x^2\cos(x)\,dx$
2. $\int e^x\cos(x)\,dx$
3. $\int \arctan(x)\,dx$
4. $\int \cos(\ln(x))\;dx$

Solution

1. Let $u = x^2$ so that $dv = \cos(x)\,dx$. Then, it follows that $du = 2x\,dx$ and $v = \sin(x)$. Equation (12.7) leads to
$$\int x^2\cos(x)\,dx = x^2\sin(x) - \int 2x\sin(x)\,dx.$$
At this point, the integral on the right is indeed simpler than the one we started with, but to evaluate it, we need to do integration by parts again. Here we choose $u = 2x$ and $dv = \sin x$, so that $du = 2dx$ and $v = -\cos(x)$. Through Equation (12.7) this yields:
$$\int x^2\cos(x)\,dx = x^2\sin(x) - \left(-2x\cos(x) - \int -2\cos(x)\,dx\right).$$
The integral all the way on the right is now something we can evaluate. It evaluates to $-2\sin x$. Then going through and simplifying, being careful to keep all the signs straight, our answer is
$$\int x^2\cos(x)\;dx = x^2\sin(x) + 2x\cos(x) - 2\sin(x) + C.$$

2. This is a classic problem. In this particular example, one can let u be either $\cos(x)$ or e^x; we choose $u = e^x$ and hence $dv = \cos(x)\,dx$. Then $du = e^x\,dx$ and $v = \sin(x)$ as shown below. Using Equation (12.7) yields
$$\int e^x\cos(x)\;dx = e^x\sin(x) - \int e^x\sin(x)\,dx.$$
The integral on the right is not much different than the one we started with, so it seems like we have gotten nowhere. Let us nonetheless keep working and apply integration by parts to the new integral, using $u = e^x$ and $dv = \sin(x)\,dx$. Then we get $du = e^x\,dx$ and $v = -\cos(x)$ and this leads us to the following:
$$\begin{aligned}\int e^x\cos(x)\,dx &= e^x\sin(x) - \left(-e^x\cos(x) - \int -e^x\cos(x)\,dx\right)\\ &= e^x\sin(x) + e^x\cos(x) - \int e^x\cos(x)\;dx.\end{aligned}$$
It seems we are back right where we started, as the right hand side contains $\int e^x\cos(x)\,dx$. But this is actually a good thing.

Add $\int e^x\cos(x)\;dx$ to both sides. This gives
$$2\int e^x\cos(x)\;dx = e^x\sin(x) + e^x\cos(x)$$
Now divide both sides by 2:
$$\int e^x\cos(x)\;dx = \frac{1}{2}\big(e^x\sin(x) + e^x\cos(x)\big).$$
Simplifying a little and adding the constant of integration, our answer is thus
$$\int e^x\cos(x)\;dx = \frac{1}{2}e^x\left(\sin(x) + \cos(x)\right) + C.$$

3. Let $u = \arctan(x)$ and $dv = dx$. Then $du = 1/(1+x^2)\,dx$ and $v = x$. Using Equation (12.7) yields
$$\int \arctan(x)\,dx = x\arctan(x) - \int \frac{x}{1+x^2}\,dx.$$
The integral on the right can be solved by substitution. Taking $u = 1+x^2$, we get $du = 2x\,dx$. The integral then becomes
$$\int \arctan(x)\,dx = x\arctan(x) - \frac{1}{2}\int \frac{1}{u}\,du.$$
The integral on the right evaluates to $\ln|u| + C$, which becomes $\ln(1+x^2)+C$, as we may drop the absolute values as $1+x^2$ is always positive. Therefore, the answer is
$$\int \arctan(x)\ dx = x\arctan(x) - \frac{1}{2}\ln(1+x^2) + C.$$

4. The integrand contains a composition of functions, leading us to think integration by parts would be beneficial. Letting $u = \cos\big(\ln(x)\big)$, we have $du = -\sin\big(\ln(x)\big)/x\ dx$, and consequently $dv = dx$ and $v = x$. We then have
$$\begin{aligned}\int \cos(\ln(x))\ dx &= x\cos(\ln(x)) + \int \sin(\ln(x))\ dx \\ &= x\cos(\ln(x)) + x\sin(\ln(x)) - \int \cos(\ln(x))\ dx\,.\end{aligned}$$
So, we see that
$$\int \cos(\ln(x))\ dx = \frac{1}{2}x\big(\sin(\ln(x)) + \cos(\ln(x))\big) + C.$$

In general, integration by parts is useful for integrating certain products of functions, like $\int xe^x\,dx$ or $\int x^3\sin(x)\,dx$. It is also useful for integrals involving logarithms and inverse trigonometric functions.

As stated before, integration is generally more difficult than differentiation. We are developing tools for handling a large array of integrals, and experience will tell us when one tool is preferable/necessary over another. For instance, consider the three similar–looking integrals
$$\int xe^x\,dx, \qquad \int xe^{x^2}\,dx \qquad \text{and} \qquad \int xe^{x^3}\,dx.$$
While the first is calculated easily with integration by parts, the second is best approached with substitution. Taking things one step further, the third integral has no answer in terms of elementary functions, so none of the methods we learn in calculus will get us the exact answer.

12.4.3 Trigonometric integrals

Functions involving trigonometric functions are useful as they are good at describing periodic behavior. Here, we describe several techniques for finding antiderivatives of certain combinations of trigonometric functions.

12.4.3.1 Integrals of the form $\int \sin^m(x)\cos^n(x)\,dx$

We consider integrals of the form

$$\int \sin^m(x)\cos^n(x)\,dx,$$

where m, n are nonnegative integers. Our strategy for evaluating these integrals is to use the identity $\cos^2(x)+\sin^2(x)=1$ to convert high powers of one trigonometric function into the other, leaving a single sine or cosine term in the integrand. This is summarized below.

1. If m is odd, then $m = 2k+1$ for some integer k. Rewrite

   $$\sin^m(x) = \sin^{2k+1}(x) = \sin^{2k}(x)\sin(x) = (\sin^2(x))^k\sin(x) = (1-\cos^2(x))^k\sin(x).$$

   Then

   $$\int \sin^m(x)\cos^n(x)\,dx = \int (1-\cos^2(x))^k\sin(x)\cos^n(x)\,dx = -\int (1-u^2)^k u^n\,du,$$

   where $u=\cos(x)$ and $du=-\sin(x)\,dx$.

2. If n is odd, then using substitutions similar to that outlined above we have

   $$\int \sin^m(x)\cos^n(x)\,dx = \int u^m(1-u^2)^k\,du,$$

   where $u=\sin(x)$ and $du=\cos(x)\,dx$.

3. If both m and n are even, use Theorem 5.12 to reduce the degree of the integrand. Expand the result and apply (1)-(3) again.

Let us check out how this all works in the following examples.

Example 12.16

Evaluate

$$\int \sin^5(x)\cos^9(x)\,dx.$$

Solution

The powers of both the sine and cosine terms are odd, therefore we can apply the techniques above to either power. We choose to work with the power of the cosine term.

We rewrite $\cos^9(x)$ as

$$\begin{aligned}\cos^9(x) &= \cos^8(x)\cos(x)\\ &= (\cos^2(x))^4\cos(x)\\ &= (1-\sin^2(x))^4\cos(x)\\ &= (1-4\sin^2(x)+6\sin^4(x)-4\sin^6(x)+\sin^8(x))\cos x.\end{aligned}$$

We rewrite the integral as

$$\int \sin^5(x)\cos^9(x)\,dx = \int \sin^5(x)\big(1-4\sin^2(x)+6\sin^4(x)-4\sin^6(x)+\sin^8(x)\big)\cos(x)\,dx.$$

Now substitute using $u = \sin(x)$ and $du = \cos(x)\ dx$ to arrive at the following integral

$$\int u^5(1-4u^2+6u^4-4u^6+u^8)\ du\,,$$

which can then be integrated:

$$\begin{aligned}\int u^5(1-4u^2+6u^4-4u^6+u^8)\ du &= \int \left(u^5-4u^7+6u^9-4u^{11}+u^{13}\right)\ du\\ &= \frac{1}{6}u^6-\frac{1}{2}u^8+\frac{3}{5}u^{10}-\frac{1}{3}u^{12}+\frac{1}{14}u^{14}+C\\ &= \frac{1}{6}\sin^6(x)-\frac{1}{2}\sin^8(x)+\frac{3}{5}\sin^{10}(x)+\ldots\\ &\quad -\frac{1}{3}\sin^{12}(x)+\frac{1}{14}\sin^{14}(x)+C.\end{aligned}$$

The work we are doing here can be a bit tedious, but the skills developed (problem solving, algebraic manipulation, etc.) are important. Nowadays problems of this sort are often solved using a computer algebra system. Mathematica, for instance, integrates $\int \sin^5(x)\cos^9(x)\ dx$ as

$$g(x) \;=\; -\frac{45\cos(2x)}{16384}-\frac{5\cos(4x)}{8192}+\frac{19\cos(6x)}{49152}+\frac{\cos(8x)}{4096}-\frac{\cos(10x)}{81920}-\frac{\cos(12x)}{24576}-\frac{\cos(14x)}{114688},$$

which clearly has a different form than our answer in Example 12.16, which we now refer to as $f(x)$. Figure 12.9 shows a graph of f and g; they are clearly not equal, but they differ only by a constant. That is $g(x) = f(x)+C$ for some constant C. So we have two different antiderivatives of the same function, meaning both answers are correct.

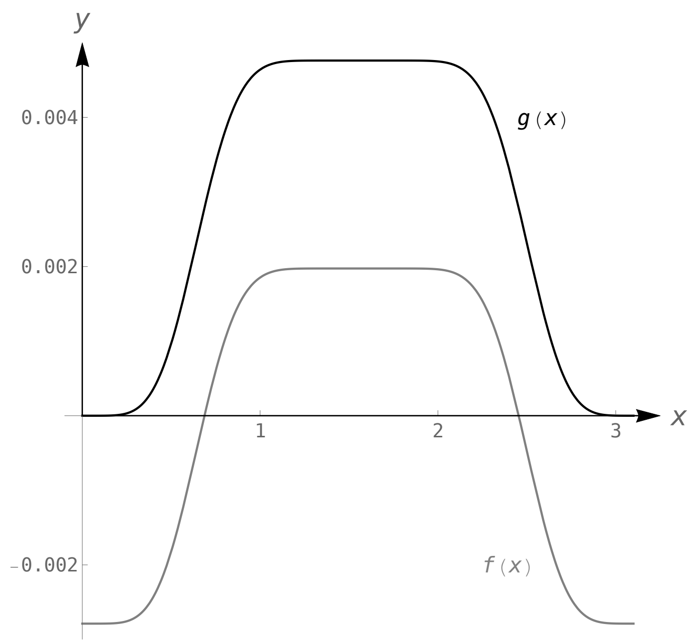


Figure 12.9: A plot of $f(x)$ and $g(x)$ from Example 12.16.

Example 12.17

Evaluate

$$\int \cos^4(x)\sin^2(x)\,dx.$$

Solution

The powers of sine and cosine are both even, so we employ the power–reducing formulas and algebra as follows.

$$\begin{aligned}\int \cos^4(x)\sin^2(x)\,dx &= \int \left(\frac{1+\cos(2x)}{2}\right)^2\left(\frac{1-\cos(2x)}{2}\right)dx\\ &= \int \frac{1+2\cos(2x)+\cos^2(2x)}{4}\cdot\frac{1-\cos(2x)}{2}\,dx\\ &= \int \frac{1}{8}\big(1+\cos(2x)-\cos^2(2x)-\cos^3(2x)\big)\,dx\end{aligned}$$

The $\cos(2x)$ term is easy to integrate. The $\cos^2(2x)$ term is another trigonometric integral with an even power, requiring the power–reducing formula again. The $\cos^3(2x)$ term is a cosine function with an odd power, requiring a substitution as done before. We integrate each in turn below.

$$\int \cos(2x)\,dx = \frac{1}{2}\sin(2x)+C.$$

$$\int \cos^2(2x)\,dx = \int \frac{1+\cos(4x)}{2}\,dx = \frac{1}{2}\left(x+\frac{1}{4}\sin(4x)\right)+C.$$

Finally, we rewrite $\cos^3(2x)$ as

$$\cos^3(2x) = \cos^2(2x)\cos(2x) = \big(1-\sin^2(2x)\big)\cos(2x).$$

Letting $u=\sin(2x)$, we have $du = 2\cos(2x)\,dx$, hence

$$\begin{aligned}\int \cos^3(2x)\,dx &= \int \big(1-\sin^2(2x)\big)\cos(2x)\,dx\\ &= \int \frac{1}{2}(1-u^2)\,du\\ &= \frac{1}{2}\left(u-\frac{1}{3}u^3\right)+C\\ &= \frac{1}{2}\left(\sin(2x)-\frac{1}{3}\sin^3(2x)\right)+C\,.\end{aligned}$$

Putting all the pieces together, we have

$$\begin{aligned}\int \cos^4(x)\sin^2(x)\,dx &= \int \frac{1}{8}\big(1+\cos(2x)-\cos^2(2x)-\cos^3(2x)\big)\,dx\\ &= \frac{1}{8}\left[x+\frac{1}{2}\sin(2x)-\frac{1}{2}\left(x+\frac{1}{4}\sin(4x)\right)-\frac{1}{2}\left(\sin(2x)-\frac{1}{3}\sin^3(2x)\right)\right]+C\end{aligned}$$

$$= \frac{1}{8}\left[\frac{1}{2}x - \frac{1}{8}\sin(4x) + \frac{1}{6}\sin^3(2x)\right] + C.$$

12.4.3.2 Integrals of products of sines and cosines of differing period

Integrals of the form

$$\int \sin(mx)\sin(nx)\ dx, \quad \int \cos(mx)\cos(nx)\ dx \quad \text{and} \quad \int \sin(mx)\cos(nx)\ dx$$

are best approached by first applying the product to sum formulas (Theorem 5.13).

Example 12.18

Evaluate

$$\int \sin(5x)\cos(2x)\ dx.$$

Solution

The application of the appropriate Simpson formula and subsequent integration are straightforward:

$$\begin{aligned}\int \sin(5x)\cos(2x)\ dx &= \int \frac{1}{2}\Big[\sin(3x) + \sin(7x)\Big]\ dx \\ &= -\frac{1}{6}\cos(3x) - \frac{1}{14}\cos(7x) + C\end{aligned}$$

12.4.3.3 Integrals of the form $\int \tan^m(x)\sec^n(x)\ dx$.

When evaluating integrals of the form $\int \sin^m(x)\cos^n(x)\ dx$, the Pythagorean theorem allowed us to convert even powers of sine into even powers of cosine, and vise–versa. The same basic strategy applies to integrals of the form $\int \tan^m(x)\sec^n(x)\ dx$, albeit a bit more nuanced.

Basically, if the integrand can be manipulated to separate a $\sec^2(x)$ term with the remaining secant power even, or if a $\sec(x)\tan(x)$ term can be separated with the remaining $\tan(x)$ power even, the Pythagorean theorem can be employed, leading to a simple substitution. This strategy is outlined below.

1. If n is even, then $n = 2k$ for some integer k. Rewrite $\sec^n x$ as

   $$\sec^n(x) = \sec^{2k}(x) = \sec^{2k-2}(x)\sec^2(x) = (1+\tan^2(x))^{k-1}\sec^2(x).$$

   Then

   $$\int \tan^m(x)\sec^n(x)\ dx = \int \tan^m(x)(1+\tan^2(x))^{k-1}\sec^2(x)\ dx = \int u^m(1+u^2)^{k-1}\ du,$$

   where $u = \tan(x)$ and $du = \sec^2(x)\ dx$.

2. If m is odd, then $m = 2k+1$ for some integer k. Rewrite $\tan^m(x)\sec^n(x)$ as

   $$\tan^m(x)\sec^n(x) \quad = \quad \tan^{2k+1}(x)\sec^n(x) = \tan^{2k}(x)\sec^{n-1}(x)\sec(x)\tan(x)$$

$$= \quad (\sec^2(x)-1)^k \sec^{n-1}(x)\sec(x)\tan(x).$$

Then

$$\int \tan^m(x)\sec^n(x)\ dx = \int (\sec^2(x)-1)^k \sec^{n-1}(x)\sec(x)\tan(x)\ dx = \int (u^2-1)^k u^{n-1}\ du,$$

where $u = \sec(x)$ and $du = \sec(x)\tan(x)\ dx$.

3. If n is odd and m is even, then $m = 2k$ for some integer k. Convert $\tan^m(x)$ to $(\sec^2(x)-1)^k$. Expand the new integrand and use Integration By Parts, with $dv = \sec^2(x)\ dx$.

4. If m is even and $n = 0$, rewrite $\tan^m(x)$ as

$$\tan^m(x) = \tan^{m-2}(x)\tan^2(x) = \tan^{m-2}(x)(\sec^2(x)-1) = \tan^{m-2}(x)\sec^2(x) - \tan^{m-2}(x).$$

So

$$\int \tan^m(x)\ dx = \underbrace{\int \tan^{m-2}(x)\sec^2(x)\ dx}_{\text{apply rule \#1}} - \underbrace{\int \tan^{m-2}(x)\ dx}_{\text{apply rule \#4 again}}.$$

Example 12.19

Evaluate the following indefinite integrals:

1. $\displaystyle\int \tan^2(x)\sec^6(x)\ dx$,

2. $\displaystyle\int \tan^6(x)\ dx$.

Solution

1. Since the power of secant is even, we use rule #1 above and pull out a $\sec^2(x)$ in the integrand. We convert the remaining powers of secant into powers of tangent.

$$\begin{aligned}\int \tan^2(x)\sec^6(x)\ dx &= \int \tan^2(x)\sec^4(x)\sec^2(x)\ dx \\ &= \int \tan^2(x)\big(1+\tan^2(x)\big)^2 \sec^2(x)\ dx\end{aligned}$$

Now substitute, with $u = \tan(x)$, with $du = \sec^2(x)\ dx$:

$$= \int u^2\big(1+u^2\big)^2\ du.$$

We leave the integration and subsequent substitution to the reader. The final answer is

$$= \frac{1}{3}\tan^3(x) + \frac{2}{5}\tan^5(x) + \frac{1}{7}\tan^7(x) + C.$$

2. We employ rule #4 of the workflow outlined above.

$$\int \tan^6(x)\ dx = \int \tan^4(x)\tan^2(x)\ dx$$

$$= \int \tan^4(x)\big(\sec^2(x)-1\big)\,dx$$

$$= \int \tan^4(x)\sec^2(x)\,dx - \int \tan^4(x)\,dx$$

Integrate the first integral with substitution, $u=\tan(x)$; integrate the second by employing rule #4 again.

$$= \frac{1}{5}\tan^5(x) - \int \tan^2(x)\tan^2(x)\,dx$$

$$= \frac{1}{5}\tan^5(x) - \int \tan^2(x)\big(\sec^2(x)-1\big)\,dx$$

$$= \frac{1}{5}\tan^5(x) - \int \tan^2(x)\sec^2(x)\,dx + \int \tan^2(x)\,dx$$

Again, use substitution for the first integral and rule #4 for the second.

$$= \frac{1}{5}\tan^5(x) - \frac{1}{3}\tan^3(x) + \int \big(\sec^2(x)-1\big)\,dx$$

$$= \frac{1}{5}\tan^5(x) - \frac{1}{3}\tan^3(x) + \tan(x) - x + C$$

These latter examples were admittedly long, with repeated applications of the same rule. Try to not be overwhelmed by the length of the problem, but rather admire how robust this solution method is. A trigonometric function of a high power can be systematically reduced to trigonometric functions of lower powers until all antiderivatives can be computed.

12.4.4 Trigonometric substitution

We have since learned a number of integration techniques, yet we are still unable to evaluate an integral like

$$\int_{-3}^{3} \sqrt{9-x^2}\,dx. \tag{12.8}$$

without resorting to a geometric interpretation. This section introduces **trigonometric substitution** (*goniometrische substitutie*), a method of integration that fills this gap in our integration skill. This technique works on the same principle as substitution, by setting set $x=f(\theta)$, where f is a trigonometric function, and then replacing x with $f(\theta)$.

For what concerns the integral given by Equation (12.8), we begin by noting that $9\sin^2(\theta)+9\cos^2(\theta)=9$, and hence $9\cos^2(\theta)=9-9\sin^2(\theta)$. If we let $x=3\sin(\theta)$, then $9-x^2=9-9\sin^2(\theta)=9\cos^2(\theta)$.

Setting $x=3\sin(\theta)$ gives $dx=3\cos(\theta)\,d\theta$. We are almost ready to substitute. We also wish to change our bounds of integration. The bound $x=-3$ corresponds to $\theta=-\pi/2$. Likewise, the bound of $x=3$ is replaced by the bound $\theta=\pi/2$. Thus

$$\int_{-3}^{3} \sqrt{9-x^2}\,dx = \int_{-\pi/2}^{\pi/2} \sqrt{9-9\sin^2(\theta)}\,(3\cos(\theta))\,d\theta$$

$$= \int_{-\pi/2}^{\pi/2} 3\sqrt{9\cos^2(\theta)}\cos(\theta)\ d\theta$$

$$= \int_{-\pi/2}^{\pi/2} 3\left|3\cos(\theta)\right|\cos(\theta)\ d\theta.$$

On $[-\pi/2, \pi/2]$, $\cos\theta$ is always positive, so we can drop the absolute value bars, then employ a power–reducing formula:

$$= \int_{-\pi/2}^{\pi/2} 9\cos^2(\theta)\ d\theta$$

$$= \int_{-\pi/2}^{\pi/2} \frac{9}{2}\big(1+\cos(2\theta)\big)\ d\theta$$

$$= \frac{9}{2}\left(\theta + \frac{1}{2}\sin(2\theta)\right)\Bigg|_{-\pi/2}^{\pi/2} = \frac{9}{2}\pi.$$

This matches our answer in Example 12.3.

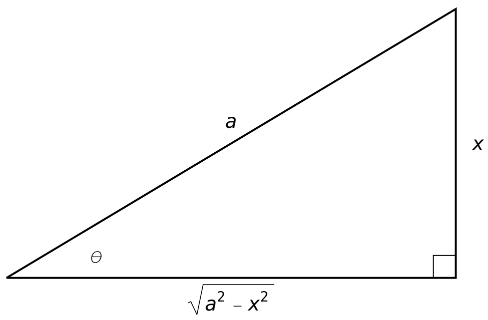
Trigonometric substitution excels when dealing with integrands that contain $\sqrt{a^2-x^2}$, $\sqrt{x^2-a^2}$ and $\sqrt{x^2+a^2}$. The following outlines the procedure for each case. Each right triangle acts as a reference to help us understand the relationships between x and θ.

(a) For integrands containing $\sqrt{a^2-x^2}$:

Let $x = a\sin(\theta)$, then $dx = a\cos(\theta)\ d\theta$.

Thus $\theta = \arcsin(x/a)$, for $-\pi/2 \le \theta \le \pi/2$.

On this interval, $\cos(\theta) \ge 0$, so $\sqrt{a^2-x^2} = a\cos(\theta)$.

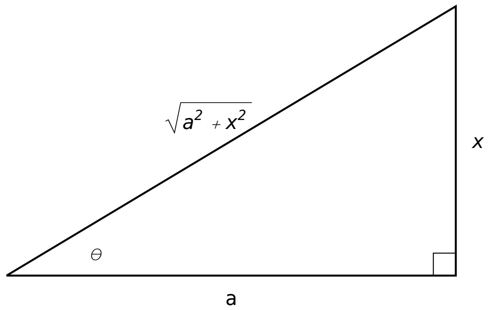


(b) For integrands containing $\sqrt{x^2+a^2}$:

Let $x = a\tan(\theta)$, then $dx = a\sec^2(\theta)\ d\theta$.

Thus $\theta = \arctan(x/a)$, for $-\pi/2 < \theta < \pi/2$.

On this interval, $\sec(\theta) > 0$, so $\sqrt{x^2+a^2} = a\sec(\theta)$.

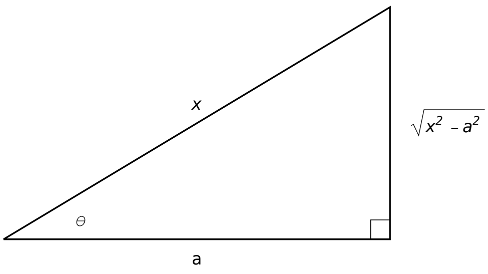


(c) For integrands containing $\sqrt{x^2-a^2}$:

Let $x = a\sec(\theta)$, then $dx = a\sec(\theta)\tan(\theta)\ d\theta$.

Thus $\theta = \operatorname{arcsec}(x/a)$. If $x/a \geq 1$, then $0 \leq \theta < \pi/2$; if $x/a \leq -1$, then $\pi/2 < \theta \leq \pi$.

We restrict our work to where $x \geq a$, so $x/a \geq 1$, and $0 \leq \theta < \pi/2$. On this interval, $\tan\theta \geq 0$, so $\sqrt{x^2-a^2} = a\tan(\theta)$.



Example 12.20

Evaluate

$$\int \sqrt{4x^2-1}\ dx.$$

Solution

We start by rewriting the integrand so that it looks like $\sqrt{x^2-a^2}$ for some value of a:

$$\begin{aligned}\sqrt{4x^2-1} &= \sqrt{4\left(x^2-\frac{1}{4}\right)}\\ &= 2\sqrt{x^2-\left(\frac{1}{2}\right)^2}.\end{aligned}$$

So we have $a = 1/2$, and following rule (c) from the above workflow, we set $x = \sec(\theta)/2$, and hence $dx = \sec(\theta)\tan(\theta)/2\ d\theta$. We now rewrite the integral with these substitutions:

$$\begin{aligned}\int \sqrt{4x^2-1}\ dx &= \int 2\sqrt{x^2-\left(\frac{1}{2}\right)^2}\ dx\\ &= \int 2\sqrt{\frac{1}{4}\sec^2(\theta)-\frac{1}{4}}\left(\frac{1}{2}\sec(\theta)\tan(\theta)\right)\ d\theta\\ &= \int \sqrt{\frac{1}{4}(\sec^2(\theta)-1)}\Big(\sec(\theta)\tan(\theta)\Big)\ d\theta\\ &= \int \sqrt{\frac{1}{4}\tan^2(\theta)}\Big(\sec\theta\tan(\theta)\Big)\ d\theta\\ &= \int \frac{1}{2}\tan^2(\theta)\sec(\theta)\ d\theta\\ &= \frac{1}{2}\int \Big(\sec^2(\theta)-1\Big)\sec(\theta)\ d\theta\\ &= \frac{1}{2}\int \big(\sec^3(\theta)-\sec(\theta)\big)\ d\theta.\end{aligned}$$

We can now integrate $\sec^3(\theta)$ using integration by parts with $dv = \sec^2(\theta)$ and $u = \sec(\theta)$, finding its antiderivatives to be

$$\int \sec^3(\theta)\ d\theta = \frac{1}{2}\Big(\sec(\theta)\tan(\theta) + \ln\big|\sec(\theta)+\tan(\theta)\big| \Big) + C.$$

Thus

$$\begin{aligned}
\int \sqrt{4x^2-1}\ dx &= \frac{1}{2}\int \big(\sec^3(\theta) - \sec(\theta)\big)\ d\theta \\
&= \frac{1}{2}\left(\frac{1}{2}\Big(\sec(\theta)\tan(\theta) + \ln\big|\sec(\theta)+\tan(\theta)\big| \Big) - \ln\big|\sec(\theta)+\tan(\theta)\big|\right) + C \\
&= \frac{1}{4}\big(\sec(\theta)\tan(\theta) - \ln\big|\sec(\theta)+\tan(\theta)\big|\big) + C.
\end{aligned}$$

We are not yet done. Our original integral is given in terms of x, whereas our final answer, as given, is in terms of θ. We need to rewrite our answer in terms of x. With $a = 1/2$, and $x = \sec(\theta)/2$, the reference triangle in rule (c) of the above workflow shows that

$$\tan\theta = \sqrt{x^2-\frac{1}{4}}\Big/\frac{1}{2} = 2\sqrt{x^2-\frac{1}{4}} \quad \text{and} \quad \sec(\theta) = 2x.$$

Thus

$$\frac{1}{4}\Big(\sec(\theta)\tan(\theta) - \ln\big|\sec(\theta)+\tan(\theta)\big| \Big) + C$$

becomes

$$\frac{1}{4}\left(2x\,2\sqrt{x^2-\frac{1}{4}} - \ln\left|2x + 2\sqrt{x^2-\frac{1}{4}}\right|\right) + C.$$

The final answer hence is:

$$\int \sqrt{4x^2-1}\ dx = \frac{1}{4}\left(4x\sqrt{x^2-\frac{1}{4}} - \ln\left|2x + 2\sqrt{x^2-\frac{1}{4}}\right|\right) + C.$$

It is important to realize that trigonometric substitution can be applied in many situations, even those not of the form $\sqrt{a^2-x^2}$, $\sqrt{x^2-a^2}$ or $\sqrt{x^2+a^2}$. This is illustrated in the following example.

Example 12.21

Evaluate

$$\int \frac{1}{(x^2+6x+10)^2}\ dx.$$

Solution

We start by completing the square, then make the substitution $u = x+3$, followed by the trigonometric substitution of $u = \tan(\theta)$:

$$\int \frac{1}{(x^2+6x+10)^2}\ dx = \int \frac{1}{\big((x+3)^2+1\big)^2}\ dx = \int \frac{1}{(u^2+1)^2}\ du.$$

Now make the substitution $u = \tan(\theta)$, $du = \sec^2(\theta)\ d\theta$:

$$\begin{aligned}
\int \frac{1}{(u^2+1)^2}\ du &= \int \frac{1}{(\tan^2(\theta)+1)^2}\sec^2(\theta)\ d\theta \\
&= \int \frac{1}{(\sec^2(\theta))^2}\sec^2(\theta)\ d\theta \\
&= \int \cos^2(\theta)\ d\theta.
\end{aligned}$$

Applying a power reducing formula, we have

$$\begin{aligned}
\int \cos^2(\theta)\ d\theta &= \int \left(\frac{1}{2}+\frac{1}{2}\cos(2\theta)\right)\ d\theta \\
&= \frac{1}{2}\theta + \frac{1}{4}\sin(2\theta) + C. \qquad (12.9)
\end{aligned}$$

We need to return to the variable x. As $u = \tan(\theta)$, $\theta = \arctan(u)$. Using the identity $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$ and using the reference triangle found in rule (b) of the workflow above, we have

$$\frac{1}{4}\sin(2\theta) = \frac{1}{2}\frac{u}{\sqrt{u^2+1}}\frac{1}{\sqrt{u^2+1}} = \frac{1}{2}\frac{u}{u^2+1}.$$

Finally, we return to x with the substitution $u = x+3$. We start with the expression in Equation (12.9):

$$\begin{aligned}
\frac{1}{2}\theta + \frac{1}{4}\sin(2\theta) + C &= \frac{1}{2}\arctan(u) + \frac{1}{2}\frac{u}{u^2+1} + C \\
&= \frac{1}{2}\arctan(x+3) + \frac{x+3}{2(x^2+6x+10)} + C.
\end{aligned}$$

Stating our final result in one line:

$$\int \frac{1}{(x^2+6x+10)^2}\ dx = \frac{1}{2}\arctan(x+3) + \frac{x+3}{2(x^2+6x+10)} + C.$$

Finally, it should be mentioned that given a definite integral that can be evaluated using trigonometric substitution, we could first evaluate the corresponding indefinite integral and then evaluate using the original bounds. It is much more straightforward, though, to change the bounds as we substitute.

12.4.5 Partial fraction decomposition

Here we investigate the antiderivatives of rational functions. Recall that rational functions are functions of the form $f(x) = \frac{p(x)}{q(x)}$, where $p(x)$ and $q(x)$ are polynomials and $q(x) \neq 0$.

Consider the integral

$$\int \frac{1}{x^2-1}\ dx.$$

We do not have a simple formula for this. It can be evaluated using trigonometric substitution, but note how the integral is easy to evaluate once we realize:

$$\frac{1}{x^2-1} = \frac{1/2}{x-1} - \frac{1/2}{x+1}.$$

Thus

$$\int \frac{1}{x^2-1}\,dx = \int \frac{1/2}{x-1}\,dx - \int \frac{1/2}{x+1}\,dx$$
$$= \frac{1}{2}\ln|x-1| - \frac{1}{2}\ln|x+1| + C.$$

Here, we will learn how to decompose fractions like

$$\frac{1}{x^2-1}.$$

We start with a rational function

$$f(x) = \frac{p(x)}{q(x)},$$

where p and q do not have any common factors and the degree of p is less than the degree of q. It can be shown that any polynomial, and hence q, can be factored into a product of real linear and irreducible quadratic terms. The following workflow states how to **decompose a rational function into partial fractions** (*splitsing in partieelbreuken*) as a sum of rational functions whose denominators are all of lower degree than q.

1. **Linear Terms:** Let $(x-a)$ divide $q(x)$, where $(x-a)^n$ is the highest power of $(x-a)$ that divides $q(x)$. Then the decomposition of $f(x)$ will contain the sum

$$\frac{A_1}{(x-a)} + \frac{A_2}{(x-a)^2} + \cdots + \frac{A_n}{(x-a)^n}.$$

2. **Quadratic Terms:** Let (x^2+bx+c) divide $q(x)$, where $(x^2+bx+c)^n$ is the highest power of (x^2+bx+c) that divides $q(x)$. Then the decomposition of $f(x)$ will contain the sum

$$\frac{B_1x+C_1}{x^2+bx+c} + \frac{B_2x+C_2}{(x^2+bx+c)^2} + \cdots + \frac{B_nx+C_n}{(x^2+bx+c)^n}.$$

To find the coefficients A_i, B_i and C_i:

1. Multiply all fractions by $q(x)$, clearing the denominators. Collect like terms.
2. Equate the resulting coefficients of the powers of x and solve the resulting system of linear equations.

Example 12.22

Perform the partial fraction decomposition of

$$\frac{1}{x^2-1}.$$

Solution

The denominator factors into two linear terms: $x^2-1 = (x-1)(x+1)$. Thus

$$\frac{1}{x^2-1} = \frac{A}{x-1} + \frac{B}{x+1}.$$

To solve for A and B, first multiply through by $x^2-1=(x-1)(x+1)$:

$$\begin{aligned} 1 &= \frac{A(x-1)(x+1)}{x-1}+\frac{B(x-1)(x+1)}{x+1} \\ &= A(x+1)+B(x-1) \\ &= Ax+A+Bx-B \\ &= (A+B)x+(A-B). \end{aligned}$$

The next step is key. Note the equality we have:

$$1=(A+B)x+(A-B).$$

For clarity's sake, rewrite the left hand side as

$$0x+1=(A+B)x+(A-B).$$

On the left, the coefficient of the x term is 0; on the right, it is $(A+B)$. Since both sides are equal, we must have that $0=A+B$.

Likewise, on the left, we have a constant term of 1; on the right, the constant term is $(A-B)$. Therefore we have $1=A-B$.

We have two linear equations with two unknowns. This one is easy to solve by hand, leading to

$$\begin{cases} A+B=0 \\ A-B=1 \end{cases} \quad \Rightarrow \quad \begin{cases} A=1/2 \\ B=-1/2. \end{cases}$$

Thus

$$\frac{1}{x^2-1}=\frac{1/2}{x-1}-\frac{1/2}{x+1}.$$

Clearly, it can become rather tedious to do a a partial fraction decomposition by hand if one is confronted with a more complex rational fraction. Luckily, we can resort in such cases to Mathematica, which can accomplish this with the command **`Apart`**. For instance, for what concerns the rational function in Example (12.22), we should proceed as follows.

```
In[21]:= Apart[1/(x^2 - 1), x]
```

The second argument of the command **`Apart`** is nothing but the variable at stake.

Out[21]= $\frac{1}{2\ (-1+x)}-\frac{1}{2\ (1+x)}$

Example 12.23

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{1}{(x-1)(x+2)^2}\,dx,$

2. $\displaystyle\int \frac{x^3}{(x-5)(x+3)}\,dx.$

Solution

1. We decompose the integrand as follows:

$$\frac{1}{(x-1)(x+2)^2} = \frac{A}{x-1} + \frac{B}{x+2} + \frac{C}{(x+2)^2}.$$

To solve for A, B and C, we multiply both sides by $(x-1)(x+2)^2$ and collect like terms:

$$\begin{aligned} 1 &= A(x+2)^2 + B(x-1)(x+2) + C(x-1) \\ &= Ax^2 + 4Ax + 4A + Bx^2 + Bx - 2B + Cx - C \\ &= (A+B)x^2 + (4A+B+C)x + (4A-2B-C). \end{aligned} \tag{12.10}$$

We have

$$0x^2 + 0x + 1 = (A+B)x^2 + (4A+B+C)x + (4A-2B-C),$$

leading to the equations

$$\left\{ \begin{array}{rcl} A+B &=& 0 \\ 4A+B+C &=& 0 \\ 4A-2B-C &=& 1 \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{rcl} A &=& \frac{1}{9} \\ B &=& -\frac{1}{9} \\ C &=& -\frac{1}{3} \end{array} \right.$$

Thus

$$\int \frac{1}{(x-1)(x+2)^2}\, dx = \int \frac{1/9}{x-1}\, dx + \int \frac{-1/9}{x+2}\, dx + \int \frac{-1/3}{(x+2)^2}\, dx.$$

Each can be integrated with a simple substitution with $u = x-1$ or $u = x+2$. The end result is

$$\int \frac{1}{(x-1)(x+2)^2}\, dx = \frac{1}{9}\ln|x-1| - \frac{1}{9}\ln|x+2| + \frac{1}{3(x+2)} + C.$$

2. Since the degree of the numerator is now higher than the one of the denominator, we begin by using polynomial division to reduce the degree of the numerator (see Section 4.1). Doing so, we arrive at

$$\frac{x^3}{(x-5)(x+3)} = x + 2 + \frac{19x+30}{(x-5)(x+3)}.$$

Consequently, we can rewrite the new rational function as:

$$\frac{19x+30}{(x-5)(x+3)} = \frac{A}{x-5} + \frac{B}{x+3},$$

for appropriate values of A and B. Clearing denominators, we have

$$\begin{aligned} 19x + 30 &= A(x+3) + B(x-5) \\ &= (A+B)x + (3A-5B). \end{aligned}$$

This implies that:

$$\begin{cases} 19 = A + B \\ 30 = 3A - 5B. \end{cases} \qquad \Leftrightarrow \qquad \begin{cases} A = \dfrac{125}{8} \\ B = \dfrac{27}{8}. \end{cases}$$

We can now integrate:

$$\begin{aligned} \int \frac{x^3}{(x-5)(x+3)}\, dx &= \int \left(x + 2 + \frac{125/8}{x-5} + \frac{27/8}{x+3} \right) dx \\ &= \frac{x^2}{2} + 2x + \frac{125}{8} \ln|x-5| + \frac{27}{8} \ln|x+3| + C. \end{aligned}$$

3. We observe that we are confronted with a rational function of trigonometric functions, so we first of all resort to the Weierstrass substitution. This leads to the following integral

$$2 \int \frac{t^2+t+1}{(t^2+2)(t^2+1)}\, dt\,,$$

which can be finished off using partial fraction decomposition. In this way, we get

$$2 \int \frac{t^2+t+1}{(t^2+2)(t^2+1)}\, dt = 2 \int \frac{1-t}{t^2+2}\, dt + \int \frac{t}{t^2+1}\, dt\,.$$

Hence, we arrive at

$$2 \int \frac{t^2+t+1}{(t^2+2)(t^2+1)}\, dt = \frac{1}{\sqrt{2}} \arctan\left(\frac{t}{\sqrt{2}} \right) - \frac{1}{2} \ln\left(t^2+2\right) + \frac{1}{2} \ln\left(t^2+1\right) + C\,,$$

where $t = \tan(x/2)$.

We conclude our discussion of partial fraction decomposition with a final example that combines several of the techniques we encountered earlier in this section.

Example 12.24

Evaluate

$$\int \frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\, dx.$$

Solution

The degree of the numerator is less than the degree of the denominator, so we have:

$$\frac{7x^2+31x+54}{(x+1)(x^2+6x+11)} = \frac{A}{x+1} + \frac{Bx+C}{x^2+6x+11}.$$

Now clear the denominators.

$$\begin{aligned} 7x^2+31x+54 &= A(x^2+6x+11) + (Bx+C)(x+1) \\ &= (A+B)x^2 + (6A+B+C)x + (11A+C). \end{aligned}$$

This implies that:

$$\begin{cases} 7 = A + B \\ 31 = 6A + B + C \\ 54 = 11A + C. \end{cases} \quad \Leftrightarrow \quad \begin{cases} A = 5 \\ B = 2 \\ C = -1. \end{cases}$$

Thus

$$\int \frac{7x^2 + 31x + 54}{(x+1)(x^2 + 6x + 11)}\, dx = \int \left(\frac{5}{x+1} + \frac{2x-1}{x^2 + 6x + 11} \right) dx.$$

The first term of this new integrand is easy to evaluate; it leads to a $5\ln|x+1|$ term. The second term is not hard, but takes several steps and uses substitution techniques.

The integrand $\dfrac{2x-1}{x^2+6x+11}$ has a quadratic in the denominator and a linear term in the numerator. This leads us to try substitution. Let $u = x^2 + 6x + 11$, so $du = (2x+6)\, dx$. The numerator is $2x-1$, not $2x+6$, but we can get a $2x+6$ term in the numerator by adding 0 in the form of "$7-7$."

$$\begin{aligned} \frac{2x-1}{x^2+6x+11} &= \frac{2x-1+7-7}{x^2+6x+11} \\ &= \frac{2x+6}{x^2+6x+11} - \frac{7}{x^2+6x+11}. \end{aligned}$$

We can now integrate the first term with substitution, leading to a $\ln|x^2+6x+11|$ term. The final term can be integrated using arctangent. First, complete the square in the denominator:

$$\frac{7}{x^2+6x+11} = \frac{7}{(x+3)^2+2}.$$

An antiderivative of the latter term can be found using Equation (12.5) and substitution:

$$\int \frac{7}{x^2+6x+11}\, dx = \frac{7}{\sqrt{2}} \arctan\left(\frac{x+3}{\sqrt{2}}\right) + C.$$

Let's start at the beginning and put all of the steps together.

$$\begin{aligned} \int \frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\, dx &= \int \left(\frac{5}{x+1} + \frac{2x-1}{x^2+6x+11} \right) dx \\ &= \int \frac{5}{x+1}\, dx + \int \frac{2x+6}{x^2+6x+11}\, dx - \int \frac{7}{x^2+6x+11}\, dx \\ &= 5\ln|x+1| + \ln\left|x^2+6x+11\right| - \frac{7}{\sqrt{2}} \arctan\left(\frac{x+3}{\sqrt{2}}\right) + C \end{aligned}$$

It is important to remember that one is not expected to see the final answer immediately after seeing the problem. Rather, given the initial problem, we break it down into smaller problems that are easier to solve. The final answer is a combination of the answers of the smaller problems.

Partial fraction decomposition is an important tool when dealing with rational functions. Note that at its heart, it is a technique of algebra, not calculus, as we are rewriting a fraction in a new form. Still, it is very useful in the realm of calculus as it lets us evaluate a certain set of complicated integrals.

Integral equations

In Chapter 9, we encountered differential equations, which are equations that relate some function with its derivatives. Likewise, we can formulate integral equations, which are equations in which an unknown function appears under an integral sign. Consider, for instance, the following integral equation:

$$f(x) = e^{-x} - \frac{1}{2} + \frac{1}{2}e^{-x-1} + \frac{1}{2}\int_0^1 (x+1)e^{-xy} f(y)\,dy.$$

Its solution is $f(x) = e^{-x}$, which can verified easily.

Just are the differential equations, integral equations are omnipresent in physics and engineering. For instance, Maxwell's equations of electromagnetism can be formulated in integral form.

12.5 Improper Integration

Consider the following definite integrals:

- $\displaystyle\int_0^{100} \frac{1}{1+x^2}\,dx \approx 1.5608,$
- $\displaystyle\int_0^{1000} \frac{1}{1+x^2}\,dx \approx 1.5698,$
- $\displaystyle\int_0^{10,000} \frac{1}{1+x^2}\,dx \approx 1.5707.$

Notice how the integrand is $1/(1+x^2)$ in each integral. It is sketched in Figure 12.10. As the upper bound gets larger, one would expect the area under the curve would also grow. While the definite integrals do increase in value as the upper bound grows, they are not increasing by much. In fact, consider:

$$\int_0^b \frac{1}{1+x^2}\,dx = \arctan(x)\Big|_0^b = \arctan(b) - \arctan(0) = \arctan(b).$$

As $b \to +\infty$, $\arctan(b) \to \pi/2$. Therefore it seems that as the upper bound b grows, the value of the concerned definite integral approaches $\pi/2 \approx 1.5708$. This should strike the reader as being a bit amazing: even though the curve extends to infinity, it has a finite amount of area underneath it.

When we defined the definite integral $\int_a^b f(x)\,dx$ in Definition 12.2, we made two stipulations:

1. The interval over which we integrated, $[a,b]$, was a finite interval, and
2. The function $f(x)$ was continuous on $[a,b]$ (ensuring that the range of f was finite).

In this section we consider integrals where one or both of the above conditions do not hold. Such integrals are called **improper integrals** (*oneigenlijke integraal*)

12.5.1 Improper integrals with infinite bounds

We start with a definition of Improper integrals with infinite bounds.

Definitie 12.6 (Improper integrals with infinite bounds)

1. Let f be a continuous function on $[a, +\infty[$. Define

$$\int_a^{+\infty} f(x)\,dx \quad \text{to be} \quad \lim_{b\to+\infty}\int_a^b f(x)\,dx.$$

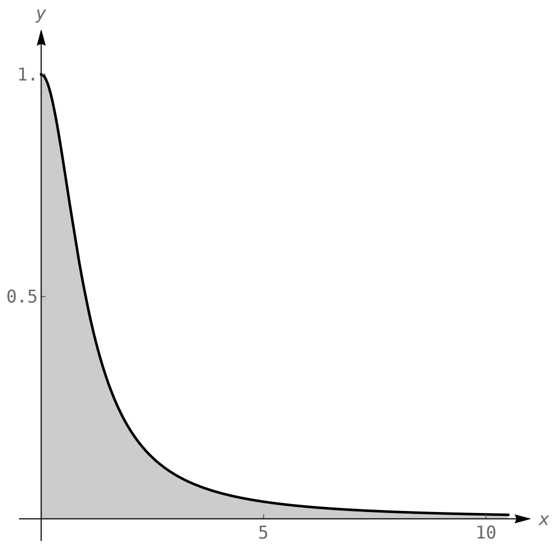


Figure 12.10: Graphing $f(x) = \dfrac{1}{1+x^2}$.

2. Let f be a continuous function on $]-\infty, b]$. Define

$$\int_{-\infty}^{b} f(x)\,dx \quad \text{to be} \quad \lim_{a\to-\infty} \int_{a}^{b} f(x)\,dx.$$

3. Let f be a continuous function on $\mathbb{R}$. Let c be any real number; define

$$\int_{-\infty}^{+\infty} f(x)\,dx \quad \text{to be} \quad \lim_{a\to-\infty} \int_{a}^{c} f(x)\,dx \;+\; \lim_{b\to+\infty} \int_{c}^{b} f(x)\,dx.$$

An improper integral is said to converge if its corresponding limit exists (is finite); otherwise, it diverges. The improper integral in part 3 converges if and only if both of its limits exist.

Example 12.25

Evaluate the following improper integrals:

1. $\displaystyle\int_{1}^{+\infty} \frac{1}{x^2}\,dx$, 2. $\displaystyle\int_{1}^{+\infty} \frac{1}{x}\,dx$, 3. $\displaystyle\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\,dx$.

Solution

1.

$$[t] \int_{1}^{+\infty} \frac{1}{x^2}\,dx = \lim_{b\to+\infty} \int_{1}^{b} \frac{1}{x^2}\,dx = \lim_{b\to+\infty} \left.\frac{-1}{x}\right|_{1}^{b} \tag{12.11}$$

$$= \lim_{b\to+\infty} \frac{-1}{b} + 1 = 1. \tag{12.12}$$

A graph of the area defined by this integral is given in Figure 12.11(a). In Mathematica, this result can be checked as follows:

```
In[22]:= Integrate[1/x^2, x, 1, +Infinity]

Out[22]= 1
```

2.
$$\begin{aligned}\int_1^\infty \frac{1}{x}\,dx &= \lim_{b\to+\infty}\int_1^b \frac{1}{x}\,dx\\ &= \lim_{b\to+\infty} \ln|x|\Big|_1^b\\ &= \lim_{b\to+\infty} \ln(b)\\ &= +\infty.\end{aligned}$$

The limit does not exist, hence the concerned improper integral diverges. Compare the graphs in Figures 12.11(a) and 12.11(b); notice how the graph of $f(x) = 1/x$ is noticeably larger. This difference is enough to cause the improper integral to diverge.

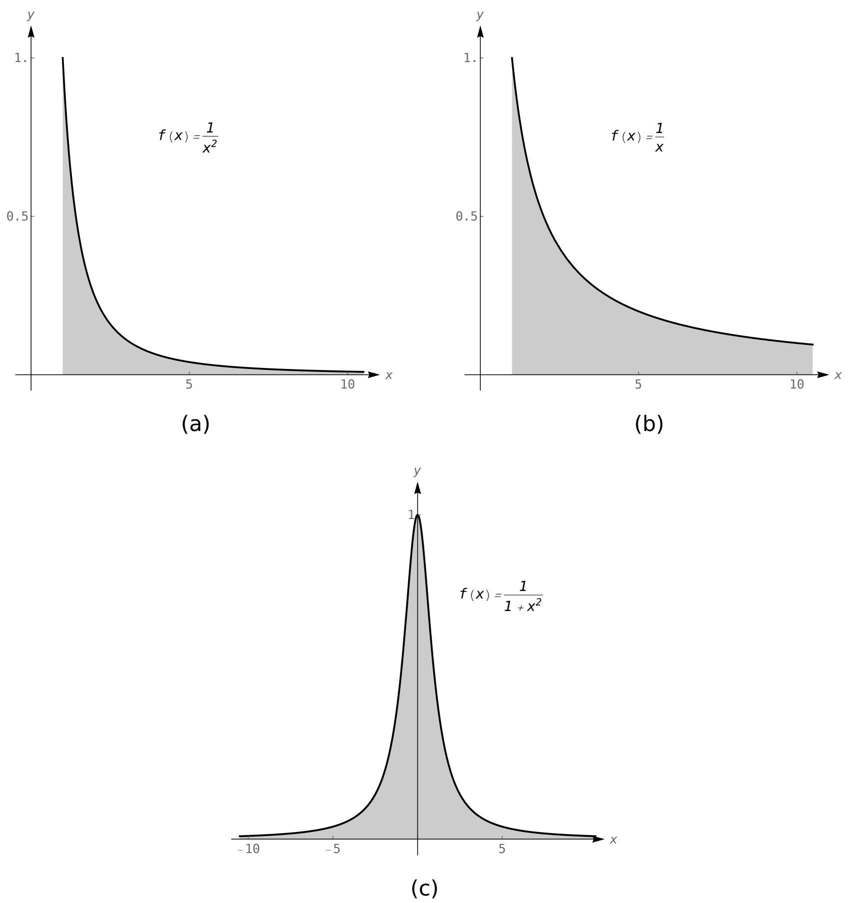


Figure 12.11: A graph of $f(x) = \frac{1}{x^2}$ (a), $f(x) = \frac{1}{x}$ (b) and $f(x) = \frac{1}{1+x^2}$ (c) in Example 12.25.

3. We will need to break this into two improper integrals and choose a value of c as in part 3 of Definition 12.6. Any value of c is fine; we choose $c = 0$.

$$\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\,dx = \lim_{a\to-\infty}\int_a^0 \frac{1}{1+x^2}\,dx + \lim_{b\to+\infty}\int_0^b \frac{1}{1+x^2}\,dx$$
$$= \lim_{a\to-\infty} \arctan(x)\Big|_a^0 + \lim_{b\to+\infty} \arctan(x)\Big|_0^b$$
$$= \lim_{a\to-\infty}\big(\arctan(0)-\arctan(a)\big) + \lim_{b\to+\infty}\big(\arctan(b)-\arctan(0)\big)$$
$$= \left(0-\frac{-\pi}{2}\right)+\left(\frac{\pi}{2}-0\right)$$

Each limit exists, hence the original integral converges and has value:

$$\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\,dx = \pi.$$

A graph of the area defined by this integral is given in Figure 12.11(c).

Note that it is not uncommon for the limits resulting from improper integrals to need l'Hôpital's rule.

12.5.2 Improper integrals with infinite range

We have just considered definite integrals where the interval of integration was infinite. We now consider another type of improper integration, where the range of the integrand is infinite.

Definitie 12.7 (Improper integrals with infinite range)
Let $f(x)$ be a continuous function on $[a,b]$ except at c, $a \le c \le b$, where $x = c$ is a vertical asymptote of f. Define

$$\int_a^b f(x)\,dx = \lim_{\substack{t\to c \\ <}} \int_a^t f(x)\,dx + \lim_{\substack{t\to c \\ >}} \int_t^b f(x)\,dx.$$

Example 12.26

Evaluate the following improper integrals:

1. $\displaystyle\int_0^1 \frac{1}{\sqrt{x}}\,dx$,
2. $\displaystyle\int_{-1}^1 \frac{1}{x^2}\,dx$.

Solution

1. A graph of $f(x) = 1/\sqrt{x}$ is given in Figure 12.12(a). Notice that f has a vertical asymptote at $x = 0$; in some sense, we are trying to compute the area of a region that has no top. Could

this have a finite value?

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx = \lim_{\substack{a\to 0\\ >}} \int_a^1 \frac{1}{\sqrt{x}}\,dx$$

$$= \lim_{\substack{a\to 0\\ >}} 2\sqrt{x}\Big|_a^1$$

$$= \lim_{\substack{a\to 0\\ >}} 2\left(\sqrt{1}-\sqrt{a}\right)$$

$$= 2$$

It turns out that the region does have a finite area even though it has no upper bound.

2. The function $f(x) = 1/x^2$ has a vertical asymptote at $x = 0$, as shown in Figure 12.12(b), so this integral is an improper integral. Let's eschew using limits for a moment and proceed without recognizing the improper nature of the integral. This leads to:

$$\int_{-1}^1 \frac{1}{x^2}\,dx = -\frac{1}{x}\Big|_{-1}^1$$

$$= -1-(1)$$

$$= -2.\ (!)$$

Clearly the area in question is above the x-axis, yet the area is supposedly negative! Why does our answer not match our intuition? To answer this, evaluate the integral using Definition 12.7.

$$\int_{-1}^1 \frac{1}{x^2}\,dx = \lim_{\substack{t\to 0\\ <}} \int_{-1}^t \frac{1}{x^2}\,dx + \lim_{\substack{t\to 0\\ >}} \int_t^1 \frac{1}{x^2}\,dx$$

$$= \lim_{\substack{t\to 0\\ <}} \left(-\frac{1}{x}\right)\Big|_{-1}^t + \lim_{\substack{t\to 0\\ >}} \left(-\frac{1}{x}\right)\Big|_t^1$$

$$= \lim_{\substack{t\to 0\\ <}} \left(-\frac{1}{t}-1\right) + \lim_{\substack{t\to 0\\ >}} \left(-1+\frac{1}{t}\right)$$

$$= \left(+\infty-1\right) + \left(-1+\infty\right)$$

Neither limit converges hence the original improper integral diverges. The nonsensical answer we obtained by ignoring the improper nature of the integral is just that: nonsensical.

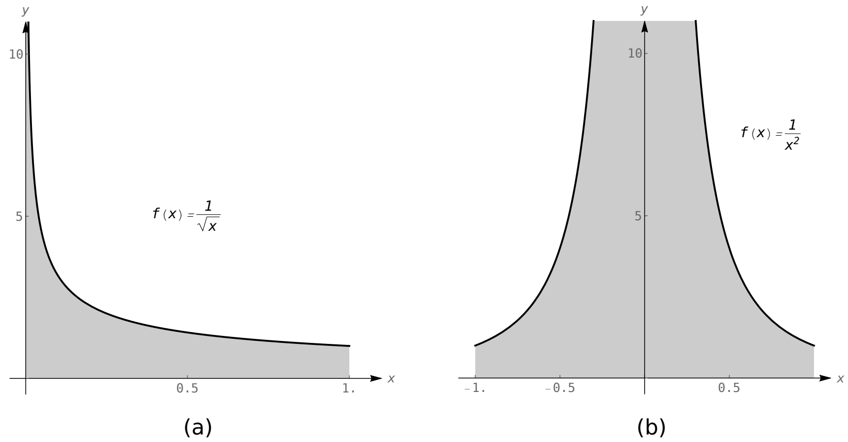


Figure 12.12: A graph of $f(x) = \frac{1}{\sqrt{x}}$ (a) and $f(x) = \frac{1}{x^2}$ (b) in Example 12.26.

This chapter has explored many integration techniques. All of them effectively have one goal in common: rewrite the integrand in a new way so that the integration step is easier to see and implement. As stated before, integration is, in general, hard. It is easy to write a function whose antiderivative is impossible to write in terms of elementary functions, and even when a function does have an antiderivative expressible by elementary functions, it may be really hard to discover what it is. Mathematica, for instance, has approximately 1,000 pages of code dedicated to integration. Do not let this difficulty discourage you. There is great value in learning integration techniques, as they allow one to manipulate an integral in ways that can illuminate a concept for greater understanding. There is also great value in understanding the need for good numerical techniques. The next chapter stresses the uses of integration. We generally do not find antiderivatives for antiderivative's sake, but rather because they provide the solution to some type of problem. The following chapter introduces us to a number of different problems whose solution is provided by integration.

12.6 Exercises

12.6.1 Analytical exercises

Antiderivatives and (in)definite integration

Assignments 12.1 — Determine the area between the x-axis and the graph of the function:

$$f(x) = \begin{cases} x, & \text{als} \quad 0 < x \le 1\,, \\ -2x+3, & \text{als} \quad 1 < x \le 2\,, \\ -1, & \text{als} \quad 2 < x \le 3\,, \\ 0, & \text{als} \quad x \le 0 \lor x > 3\,. \end{cases}$$

The fundamental theorem of calculus

Assignments 12.2 — Find the derivative of the functions below.

(a) $F(x) = \displaystyle\int_2^{x^3+x} \frac{1}{t}\,dt$

(b) $F(x) = \displaystyle\int_{x^3}^{0} t^3\,dt$

(c) $F(t) = \displaystyle\int_{-\pi}^{t} \frac{\cos(y)}{1+y^2}\,dy$

(d) $F(t) = \displaystyle\int_{t}^{3} \frac{\sin(x)}{x}\,dx$

(e) $F(x) = x^2 \displaystyle\int_0^{x^2} \frac{\sin(u)}{u}\,du$

(f) $F(\theta) = \displaystyle\int_{\sin(\theta)}^{\cos(\theta)} \frac{dx}{1-x^2}$

(g) $F(x) = 3x \displaystyle\int_4^{x^2} e^{-\sqrt{t}}\,dt$

(h) $F(x) = \displaystyle\int_x^{x^2} (t+2)\,dt$

(i) $F(x) = \displaystyle\int_{\ln(x)}^{e^x} \sin(t)\,dt$

Techniques of antidifferentiation

Assignments 12.3 — Determine the integrals below

(a) $\displaystyle\int (1-x)\sqrt{x}\,dx$

(b) $\displaystyle\int \left(4x^3-7x+5\right) dx$

(c) $\displaystyle\int \left(3x^3+2\sin(x)\right) dx$

(d) $\displaystyle\int \left(2x^3+3x-1\right)^{1/3}\left(2x^2+1\right) dx$

(e) $\displaystyle\int \sin(x)\cos(3x)\,dx$

(f) $\displaystyle\int \sin(6x)\sin(4x)\,dx$

(g) $\displaystyle\int 2x\ln(x+1)\,dx$

(h) $\displaystyle\int \frac{1}{x^2(x+1)}\,dx$

(i) $\displaystyle\int \frac{dx}{\sqrt{-x^2+2x+3}}$

(j) $\displaystyle\int \frac{2x+1}{-x^2+3x+2}\,dx$

(k) $\displaystyle\int \frac{\sin(x)}{1-\cos(x)}\,dx$

(l) $\displaystyle\int \frac{dx}{4+x^2}$

(m) $\displaystyle\int x\sin(x)\,dx$

(n) $\displaystyle\int (x+1)^2\cos(2x)\,dx$

(o) $\displaystyle\int e^{-x}\sin(2x)\,dx$

(p) $\displaystyle\int x^n\ln(x)\,dx$

(q) $\displaystyle\int \frac{x}{x^2+4x+5}\,dx$

Assignments 12.4 — Find the following indefinite integrals.

(a) $\displaystyle\int \frac{e^x\sqrt{1-x^2}-1}{\sqrt{1-x^2}}\,dx$

(b) $\displaystyle\int \frac{2x+1}{4x^2+4x+3}\,dx$

(c) $\displaystyle\int \frac{\sin(x)}{\cos^6(x)}\,dx$

(d) $\displaystyle\int \cos^5(x)\,dx$

(e) $\displaystyle\int \frac{\sin(x)-\cos(x)}{\sin(x)+\cos(x)}\,dx$

(f) $\displaystyle\int \frac{dx}{\cos^2(x)\sqrt{1-4\tan^2(x)}}$

(g) $\displaystyle\int \frac{dx}{\left(\cos(x)+\sin(x)\right)^2}$

(h) $\displaystyle\int \ln\left(x+\sqrt{x^2+5}\right) dx$

(i) $\displaystyle\int \frac{2x-1}{x^2+x-6}\,dx$

(j) $\displaystyle\int \left(\frac{x-1}{x^2-5x+6}\right)^2 dx$

(k) $\displaystyle\int \frac{x^2+1}{x^2+2x+2}\,dx$

(l) $\displaystyle\int \frac{x+1}{\left(x^2+1\right)^{3/2}}\,dx$

(m) $\displaystyle\int \frac{dx}{\sqrt{4x-x^2}}$

(n) $\displaystyle\int e^{2x}\sin(4x)\,dx$

(o) $\displaystyle\int \sin^4(x)\cos^2(x)\,dx$

(p) $\displaystyle\int \frac{\cos(x)}{2\cos^2(x)+\sin(x)-1}\,dx$

(q) $\displaystyle\int \frac{dx}{\sin^2(x)\cos^4(x)}$

Assignments 12.5 — Find the following indefinite integrals.

(a) $\displaystyle\int \frac{3x^2-4}{x^2+1}\,dx$

(b) $\displaystyle\int \frac{x^4}{x^3-8}\,dx$

(c) $\displaystyle\int \frac{dx}{x^4\sqrt{x^2-1}}$

(d) $\displaystyle\int \frac{3-4x}{\left(1-2\sqrt{x}\right)^2}\,dx$

(e) $\displaystyle\int \frac{dx}{\left(\tan(x)+1\right)\sin^2(x)}$

(f) $\displaystyle\int x\,e^{2x}\,dx$

(g) $\displaystyle\int \frac{5x}{\sqrt{x^4+1}}\,dx$

(h) $\displaystyle\int \sin\left(\frac{\pi}{4}-x\right)\sin\left(\frac{\pi}{4}+x\right)dx$

(i) $\displaystyle\int \frac{dx}{\sqrt[4]{5-x}+\sqrt{5-x}}$

Assignments 12.6 — Find the following indefinite integrals.

(a) $\displaystyle\int x^2\ln\left(\sqrt{1-x}\right)dx$

(b) $\displaystyle\int \frac{2x-1}{2x+3}\,dx$

(c) $\displaystyle\int \sin(2x)\cos(2x)\,dx$

(d) $\displaystyle\int \frac{dx}{e^x+1}$

(e) $\displaystyle\int \frac{dx}{x^2+x+1}$

(f) $\displaystyle\int \frac{2x+3}{\left(x^2+x+1\right)^2}\,dx$

(g) $\displaystyle\int \sqrt{\frac{a+x}{a-x}}\,dx$

(h) $\displaystyle\int \frac{x-2\sqrt{x-1}}{1+\sqrt[4]{x-1}}\,dx$

(i) $\displaystyle\int \arctan(\sqrt{x})\,dx$

(j) $\displaystyle\int x^5\left(1+x^3\right)^{1/2}\,dx$

(k) $\displaystyle\int \frac{dx}{\sin^3(x)\,\cos^5(x)}$

(l) $\displaystyle\int \frac{dx}{\sin^6(x)}$

(m) $\displaystyle\int \frac{dx}{\sqrt{1+e^x}}$

(n) $\displaystyle\int \sin^4(x)\,dx$

(o) $\displaystyle\int \frac{dx}{1+\cos(x)+\sin(x)}$

Improper Integration

Assignments 12.7 — Calculate the following improper integrals.

(a) $\displaystyle\int_0^{\frac{\pi}{2}} \sec(x)\, dx$

(b) $\displaystyle\int_0^{+\infty} e^{-x}\sin(x)\, dx$

(c) $\displaystyle\int_0^{+\infty} \frac{dx}{x^2+1}$

(d) $\displaystyle\int_{-1}^{1} \frac{dx}{x^2}$

(e) $\displaystyle\int_{-1}^{8} x^{-\frac{2}{3}}\, dx$

(f) $\displaystyle\int_{-2}^{1} \frac{1}{\sqrt{|x|}}\, dx$

(g) $\displaystyle\int_0^{+\infty} xe^{x^2}\, dx$

(h) $\displaystyle\int_{-1}^{1} \frac{1}{x^4}\, dx$

(i) $\displaystyle\int_0^{e^2} \big(1+\ln(x)\big)\, dx$

Assignments 12.8 — The shape of the spectral lines in magnetic resonance spectroscopy is often described by the Lorentz function

$$g(\omega) = \frac{1}{\pi}\cdot\frac{T}{1+T^2(\omega-\omega_0)^2},$$

with T and ω_0 constants. Evaluate

$$\int_{\omega_0}^{+\infty} g(\omega)\, d\omega\,.$$

Review exercises

Assignments 12.9 — Evaluate the definite integrals below.

(a) $\displaystyle\int_2^{5} \left(x^2+\frac{1}{x^2}\right) dx$

(b) $\displaystyle\int_{-2}^{2} \frac{x}{\sqrt{x^2+5}}\, dx$

(c) $\displaystyle\int_0^{\frac{\pi}{2}} \cos^2(3x)\, dx$

(d) $\displaystyle\int_{-1}^{1} x^2\cos(\pi x)\, dx$

(e) $\displaystyle\int_4^{9} \frac{1-\sqrt{x}}{1+\sqrt{x}}\, dx$

(f) $\displaystyle\int_0^{2} \sqrt{4-x^2}\,\frac{|x-1|}{x-1}\, dx$

Assignments 12.10 — Find the MacLaurin series of the functions below.

(a) $\displaystyle\int_0^{\sqrt{\pi}} \sin\left(x^2\right)\, dx$

(b) $\displaystyle\int_0^{\pi^2/4} \cos\left(\sqrt{x}\right)\, dx$

12.6.2 Numerical integration

Finding an antiderivative is far from obvious. Although Sections 12.4 and 12.5 provided many integration techniques, still, in many cases these techniques are not useful. Because, for example, the antiderivative cannot be expressed in terms of elementary function(s). It becomes even more difficult if we do not have a closed-form function in the integrand, something that occurs constantly in practice. What do we do in these cases? We approximate the (definite) integral as a sum of computable areas (Section 12.2.1). This method is conceptually very simple, but it is tedious if we want to obtain an approximation with an acceptable accuracy. Therefore, here we will implement and study some numerical integration methods in Python.

12.6.2.1 The midpoint method

We can approximate a definite integral by summing the areas of a series of rectangles. For the definite integral

$$S = \int_a^b f(x)\, dx\,,$$

we obtain these n rectangles as follows:

- Subdivide the integration interval $[a, b]$ into a partition $\{x_1, x_2, \ldots, x_n, x_{n+1}\}$, where

$$a = x_1 < x_2 < \cdots < x_n < x_{n+1} = b\,;$$

- the length of the i-th subinterval $[x_i, x_{i+1}]$, denoted by Δx, is the base of the i-th rectangle;
- the height of the i-th rectangle is determined by the left, right, or midpoint rule.

If Δx is positive and we apply the midpoint rule, we call the resulting method the **midpoint method**. The approximation $\hat{\boldsymbol{S}}$ of an integral S then follows from

$$\begin{aligned} S = \int_a^b f(x)\, dx \quad &\approx \quad \Delta x\, f\left(\frac{x_1 + x_2}{2}\right) + \Delta x\, f\left(\frac{x_2 + x_3}{2}\right) + \cdots + \Delta x\, f\left(\frac{x_n + x_{n+1}}{2}\right) \\ &\approx \quad \Delta x \sum_{i=1}^{n} f\left(\frac{x_i + x_{i+1}}{2}\right) = \hat{\boldsymbol{S}}\,. \end{aligned} \tag{12.13}$$

If we rewrite Equation (12.13) as

$$\hat{\boldsymbol{S}} = \Delta x \sum_{i=1}^{n} f\left(m_i\right) = \Delta x \sum_{i=1}^{n} f\left(a + \frac{\Delta x}{2} + (i-1)\,\Delta x\right),$$

where $m_i = \frac{x_i + x_{i+1}}{2}$ is the midpoint of interval $[x_i, x_{i+1}]$, we can translate the midpoint method to executable Python code as follows.

```
def midpoint(f, interval, n=20, plotf = False):
        '''
        Midpoint method for the approximation of the integral
        of f on a given interval [a,b]
        Inputs:
                - f: integrand
                - interval: integration interval [a,b], given as a list
                - n: number of subintervals (default: 20)
                - plotf (optional): indicates whether the integrand should be plotted
 (default: False)

        Output(s):
                - S_h: the approximated integral on [a,b] using the midpoint method
                * m_list: list with the midpoints of the n subintervals
                * fm_list: list with the function values of the midpoints of the n subintervals
                * deltax: width of the subintervals
        [*] these outputs are only returned if the function is plotted (plotf=True)
        '''
    # check if n is a strictly positive integer. If not, display an error message.
        if not (isinstance(n, int) and n>0):
        print('Error: n must be a strictly positive integer')
        return None

        # extract the values of a and b from the interval
        a = interval[0]
        b = interval[1]

        # calculate the width of the subintervals
        deltax = (b-a)/n

        # make a list with the midpoints of the n subintervals [m_1,...,m_n]
        # and a list with the function values [f(m_1),...,f(m_n)]
        m_list =  [a+deltax*i-deltax/2 for i in range(1,n+1)]
        fm_list = [f(m_i) for m_i in m_list]

        # calculate the approximated integral
        S_h = deltax*sum(fm_list)

        if plotf:
                return S_h, m_list, fm_list, deltax
        else:
                return S_h
```

The function **`plot_numerical_integration(method, f, interval, n=[1,100],interactive=True)`** makes a static or interactive plot of f and the approximation of the integral on the interval $[a, b]$ with a specified numerical integration method. In the interactive plot, the number of subintervals can be determined by means of a slider. The inputs of this function are defined as follows:

- method: numerical integration method (`'midpoint'` or `'trapezium'`)
- f: integrand
- interval: integration interval $[a, b]$

- n: number of subintervals
  - `interactive = True` : range of values $[n_{min}, n_{max}]$ (default = $[1, 100]$)
  - `interactive = False`: one value for n

Question 1.a Test the function **midpoint** for the definite integral from Example 12.4.

$$S_1 = \int_{a_1}^{b_1} f_1(x)\,dx = \int_0^4 (4x - x^2)\,dx.$$

```
def f_1(x):
    return 4*x-x**2

    # call to the function midpoint
    ...
```

Question 1.b Use the function **plot_numerical_integration** to plot $f_1(x)$ and the approximation of the corresponding integral on the interval $[0.4]$ with $n = 10$.

Now make an interactive plot of $f_1(x)$ and the approximation of the corresponding integral over the interval $[0, 4]$ with $n \in [1, 20]$.

Question 1.c Using the instruction(s) below, check the computing time for $n = 100, 10^4, 10^6, \ldots$. What is the influence of n?

```
%timeit midpoint(f=f_1, interval=[0,4], n=100)
```

12.6.2.2 Approximation error

To quantitatively check the accuracy of the numerical integration method, we can use the relative approximation error ϵ:

$$\epsilon = \frac{|S - \hat{\boldsymbol{S}}|}{|S|}.$$

To investigate the effect of the number of subintervals n on ϵ, we use the function **plot_error(S, method, f, interval, n_range)**. It plots the relative approximation error ϵ of a specified numerical integration method as a function of the number of subintervals n. The inputs of this function are defined as follows:

- `S`: exact value of the integral
- method: numerical integration method (`'midpoint'` or `'trapezium'`)
- `f`: integrand
- `interval`: integration interval $[a, b]$
- `n\_range`: range of values for n $[n_{min}, n_{max}]\big(default = [1, 20]\big)$

```
from teachingtools import plot_error
```

Question 2.a We want to test the function **plot_error** for the approximation of S_1 for n ranging from 1 to 20. For this, we need S_1. First calculate the exact value of S_1 and enter it below.

```
S_1 = ... # to be completed
```

Question 2.b Now plot with the function `plot_error` the relative approximation error (ϵ) of the integral of $f_1(x)$ for n ranging from 1 to 20.

```
# to be completed with call to the function plot_error
...
```

Question 2.c Implement the integrand of the following definite integrals and calculate their exact value (if possible):

- $S_2 = \int_{a_2}^{b_2} f_2(x)\,dx = \int_0^2 \sin(2x)\cos(2x)\,dx\,,$
- $S_3 = \int_{a_3}^{b_3} f_3(x)\,dx = \int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin\left(x^3\right)\,dx\,.$

What do you notice?

```
def f_2(x):
        return ... # to be completed
S_2 = ... # to be completed with the exact value of the integral of f2 on [0,2]


def f_3(x):
        return ... # to be completed
S_3 = ... # to be completed with the exact value of the integral of f3 on [-pi/4,pi/2]
```

Question 2.d For the integral of $f_2(x)$ and $f_3(x)$, calculate the approximation with the midpoint method, and plot the relative approximation error as a function of n (for n ranging from 1 to 20) with **plot_error**.

```
# numerical approximations for S_2 (no plot)

# errors for S_2

# numerical approximations for S_3 (no plot)

# errors for S_3 (if possible)
```

12.6.2.3 The trapezium method

The midpoint method approximates a definite integral as a sum of rectangular areas. The integrand in each subinterval $]x_i, x_{i+1}[$ is approximated by a constant function $f(m_i)$. However, this approximation is not accurate in subintervals where the function value changes significantly. This can be clearly seen in the plot of f_3.

An alternative to the midpoint method is the **trapezium method**, where the areas in the subintervals are approximated by - surprise - trapezia. This method approximates the integrand over the subinterval $[x_i, x_{i+1}]$ as a straight line going from $\left(x_i, f(x_i)\right)$ to $\left(x_{i+1}, f(x_{i+1})\right)$. The definite integral is then approximated as follows:

$$S = \int_a^b f(x)\,dx \quad \approx \quad \Delta x \sum_{i=1}^{n} \frac{f(x_i) + f(x_{i+1})}{2} = \hat{\boldsymbol{S}}\,. \tag{12.14}$$

Question 3.a Implement the trapezium method by completing the function below where you find "...".

```
def trapezium(f, interval, n = 20, plotf = False):
    '''
    Trapezium method for the approximation of the integral
    of the function f on a given interval [a,b]
    Inputs:
        - f: integrand
        - interval: integration interval [a,b], given as a list
        - n: number of subintervals (default: 20)
        - plotf (optional): indicates whether the function
                            should be plotted (default: False)

    Output:
        - S_h: the approximated integral on [a,b] using the trapezium method
        * x_list: list with the boundaries of the n subintervals [x_1,...,x_n+1]
        * f_list: list with the function values of the boundaries of
                  the n subintervals [f(x_1),...,f(x_n+1)]
        * deltax: width of the subintervals
        * these outputs are only returned if the function
         is plotted (plotf = True)
    '''
    # Check if n is a strictly positive integer.
    # If not, display an error message.
    if not (isinstance(n, int) and n > 0):
        print('Error: n must be a strictly positive integer')
        return None

    # 1) Extract the values of a and b from the interval
    ...
    ...

    # 2) Calculate the width of the subintervals
    ...
    ...

    # 3) Create a vector with the (n+1) values of x [x_0, ..., x_n]
     # in the interval [a,b]
    ...

    # 4) Create a vector with the n+1 values of f [f(x_0), ..., f(x_n)]
     # of the n+1 x values in the interval [a, b]
```

```
    ...

    # 5) Calculate the approximated integral
    ...

    if plotf == True:
        return S_h, x_list, f_list, deltax
    else:
        return S_h
```

Question 3.b Compare the results of the midpoint and trapezium methods for f_1, f_2 and f_3. Do you notice any differences? Which method do you prefer and why?

```
% % approximation for S1 with midpoint

% % approximation for S1 with trapezium

% % approximation for S2 with midpoint

% % approximation for S2 with trapezium

% % approximation for S3 with midpoint

% % approximation for S3 with trapezium
```

Question 3.c Consider the following definite integral:

$$S_{tot} = \int_a^b f_t(x)\,dx = \int_a^b \big(f_s(x) + f_r(x)\big)\,dx.$$

The integrand $f_t(x)$ here is a sum of two functions, i.e. a signal function $f_s(x)$ and noise function $f_r(x)$, of which we do not know the mathematical form. We constantly encounter this in practice, for example when we use a measuring device on which dirt is deposited (after a while). In this case, x is the time, $f_s(x)$ is the quantity to be measured as a function of time and $f_r(x)$ is the disturbance of the signal due to dirt deposition as a function of time.

Import the functions f_s, f_r and f_t and check for both numerical methods the effect of the noise on the approximated integral of $f_t(x)$ on the interval $[0.20]$. Would you prefer one method over another?

12.6.2.4 Simpson's rule

The trapezium rule locally replaces the function with a straight line so that the areas in the subintervals can be approximated by trapezia. Intuitively, you expect the approximation to get better if the function is locally approximated by a nonlinear function that is still easy to integrate.

Simpson's method, also called Simpson's 1/3 method, approximates the integrand $f(x)$ by means of a quadratic curve or parabola

$$f(x) \mapsto P_2(x) = a_0 + a_1 x + a_2 x^2. \tag{12.15}$$

On the one hand, we now need three conditions to determine the three coefficients a_0, a_1 and a_2 in Eq. (12.15). On the other hand, we do not immediately know the formula for the area under a parabola.

However, we can easily find it. To do this, we consider three points at a distance Δx from each other: $(0, y_i)$, $(\Delta x, y_{i+1})$ and $(2\Delta x, y_{i+2})$. The parabola $P_2(x)$ can then be identified by requiring that these three points should lie on it:

$$\begin{aligned} x = 0 \quad &\rightarrow \quad P_2(0) = a_0 \overset{\text{require}}{=} y_i, \\ x = \Delta x \quad &\rightarrow \quad P_2(\Delta x) = a_0 + a_1 \Delta x + a_2 \Delta x^2 \overset{\text{require}}{=} y_{i+1}, \\ x = 2\Delta x \quad &\rightarrow \quad P_2(2\Delta x) = a_0 + 2a_1 \Delta x + 4a_2 \Delta x^2 \overset{\text{require}}{=} y_{i+2}. \end{aligned}$$

If we substitute $a_0 = y_i$ in the second and third equation, we obtain a linear 2 × 2 system:

$$\begin{aligned} \Delta x a_1 + \Delta x^2 a_2 &= y_{i+1} - y_i, \\ 2\Delta x a_1 + 4\Delta x^2 a_2 &= y_{i+2} - y_i, \end{aligned}$$

with solution

$$\begin{aligned} a_1 &= -\frac{1}{2\Delta x} \left(3y_i - 4y_{i+1} + y_{i+2}\right), \\ a_2 &= \frac{1}{2\Delta x^2} \left(y_i - 2y_{i+1} + y_{i+2}\right). \end{aligned}$$

This allows us to unambiguously determine the coefficients in Eq. (12.15).

Question 4.a Show analytically that the area under the parabola passing through the three considered points is given by

$$S_i = \int_0^{2h} P_2(x)\, dx = \frac{\Delta x}{3} \left(y_i + 4y_{i+1} + y_{i+2}\right).$$

Note that the area S_i only depends on the heights y_i, y_{i+1} and y_{i+2} and on Δx.

Simpson's method divides the integration interval in n subintervals of equal width $\Delta x = (b-a)/n$. We require that n is even and define $x_1 = a$ and $x_{i+1} = x_i + \Delta x$ for $i = 1, \ldots, n$, so that $x_{n+1} = b$.
The integral of the function $f(x)$ on the interval $[a, b]$ can now be approximated by

$$\begin{aligned} S = \int_a^b f(x)\, dx \quad &\approx \quad \sum_{i=0}^{\frac{n}{2}-1} \left(\int_{x_{2i}}^{x_{2i+2}} P_2(x)\, dx \right) \\ &= \sum_{i=0}^{\frac{n}{2}-1} S_{2i} \\ &= \sum_{i=0}^{\frac{n}{2}-1} \frac{\Delta x}{3} \left(y_i + 4y_{i+1} + y_{i+2}\right) = \hat{\boldsymbol{S}} \end{aligned}$$

The expanded form of the latter is easier to remember:

$$\hat{\boldsymbol{S}} = \frac{\Delta x}{3} \left(f(x_0) + 4 \sum_{\substack{i=1 \\ i \text{ odd}}}^{n-1} f(x_i) + 2 \sum_{\substack{i=2 \\ i \text{ even}}}^{n-2} f(x_i) + f(x_n) \right).$$

For example, if $n = 8$, the approximation is given by:

$$\hat{S} = \frac{\Delta x}{3}\left(f(x_0) + 4\left(f(x_1) + f(x_3) + f(x_5) + f(x_7)\right) + 2\left(f(x_2) + f(x_4) + f(x_6)\right) + f(x_8)\right).$$

Question 4.b Implement Simpson's method by completing the function below where you find "...".

```
def simpson(f, interval, n):
    """Simpson's method for approximating the integral of f on an interval [a,b]
    Inputs:
    - f: integrand (function handle)
    - interval: integration interval [a, b] given as a 1x2 row vector
    - n: the number of subintervals
    Outputs:
    - Sh: the approximation of the integral on [a,b] using Simpson's method
    - If the value of n is odd, there will be an error message

      'The value of n should be even!' and the function will stop"""

    if n%2 != 0:
        raise ValueError("...")
    # Extract the start and end value a and b from the interval
    ...
    ...

    # Get the interval width h
    ...

    # calculate the approximated integral Sh based on the start and end value of the interval
    Sh = ..

    # calculate the approximation of the integral Sh using all start and end values from the
      partial intervals

    for i in range(...):
        if n%2 == 0:
            Sh += ...
        else:
            Sh += ...
    return Sh
```

Question 4.c Use the function **`simpson`** to approximate

$$\int_{\frac{\pi}{10}}^{\frac{\pi}{2}} \frac{\sin(x)}{x}\,dx$$

using 10 subintervals. Give your result up to 5 digits after the decimal point.

Engineering is achieving function while avoiding failure.

— Henry Petroski —

13

Applications of integration

This chapter employs the following technique to a variety of applications. Suppose the value Q of a quantity is to be calculated. We first approximate the value of Q using a Riemann sum, then find the exact value via a definite integral. This goes as follows.

1. Divide the quantity into n smaller subquantities of value Q_i.
2. Identify a variable x and function $f(x)$ such that each subquantity can be approximated with the product $f(c_i)\Delta x$, where Δx represents a small change in x. Thus $Q_i \approx f(c_i)\Delta x$. A sample approximation $f(c_i)\Delta x$ of Q_i is called a **differential element**.
3. Recognize that
$$Q = \sum_{i=1}^{n} Q_i \approx \sum_{i=1}^{n} f(c_i)\Delta x,$$
which is a Riemann Sum.
4. Taking the appropriate limit gives $Q = \int_a^b f(x)\,dx$.

13.1 Area between curves

13.1.1 Rectangular coordinates

Let Q be the area of a region bounded by continuous functions f and g. If we break the region into many subregions, we have an obvious equation:

Total area = sum of the areas of the subregions.

The issue to address next is how to systematically break a region into subregions. Consider Figure 13.1(a) where a region between two curves is shaded. While there are many ways to break this into subregions, one particularly efficient way is to slice it vertically, as shown in Figure 13.1(b), into n equally spaced slices.

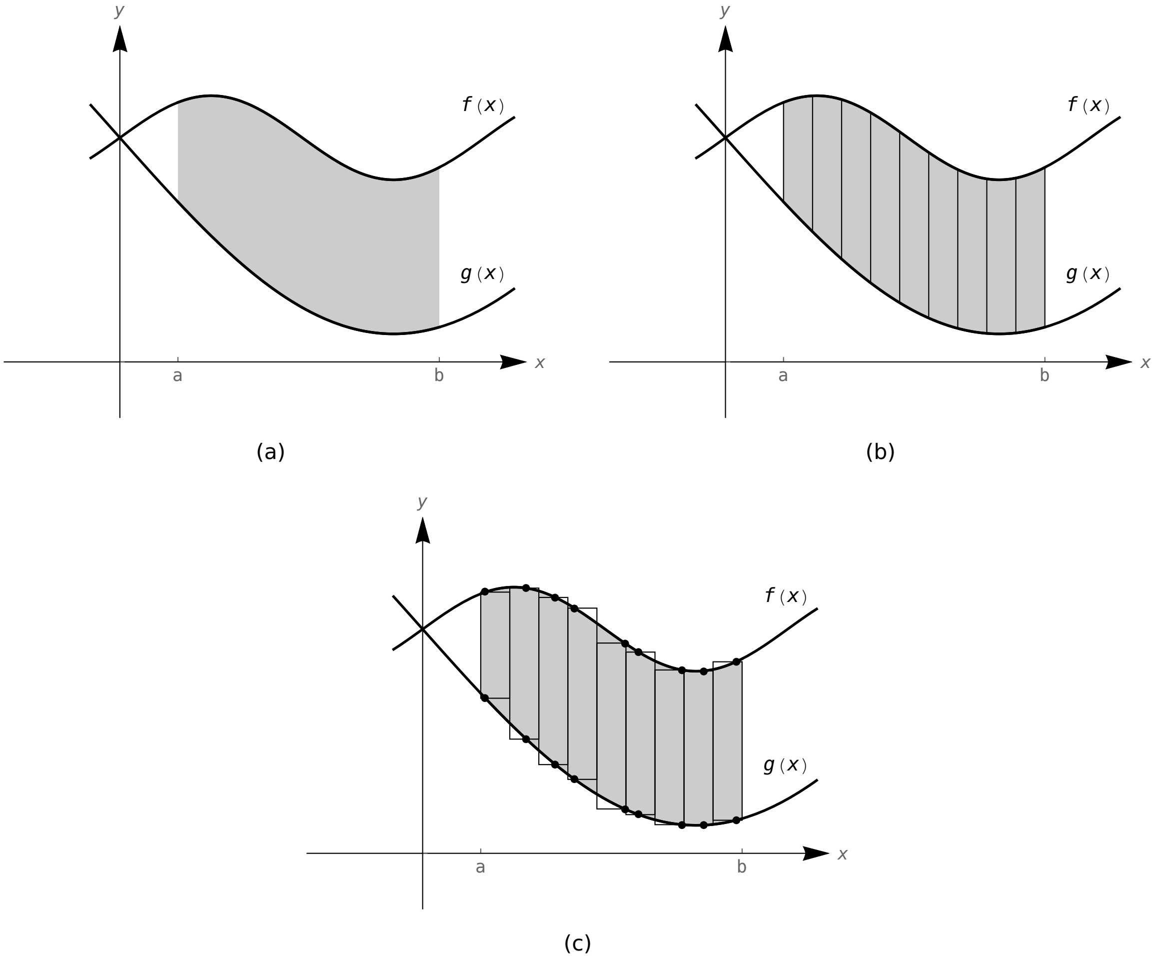


Figure 13.1: Subdividing a region into vertical slices and approximating the areas with rectangles.

We now approximate the area of a slice. Again, we have many options, but using a rectangle seems simplest. Picking any x-value c_i in the i^{th} slice, we set the height of the rectangle to be $f(c_i) - g(c_i)$, the difference of the corresponding y-values. The width of the rectangle is a small difference in x-values, which we represent with Δx. Figure 13.1(c) shows sample points c_i chosen in each subinterval and appropriate rectangles drawn. Each of these rectangles represents a differential element. Each slice has an area approximately equal to $\big(f(c_i) - g(c_i)\big)\Delta x$; hence, the total area is approximately the Riemann sum

$$Q = \sum_{i=1}^{n} \big(f(c_i) - g(c_i)\big)\Delta x.$$

Taking the limit as $n \to +\infty$ gives the exact area A as

$$\int_a^b \big(f(x) - g(x)\big)\, dx. \tag{13.1}$$

Example 13.1

Find the total area of the region enclosed by the functions $f(x) = -2x + 5$ and $g(x) = x^3 - 7x^2 + 12x - 3$ as shown in Figure 13.2.

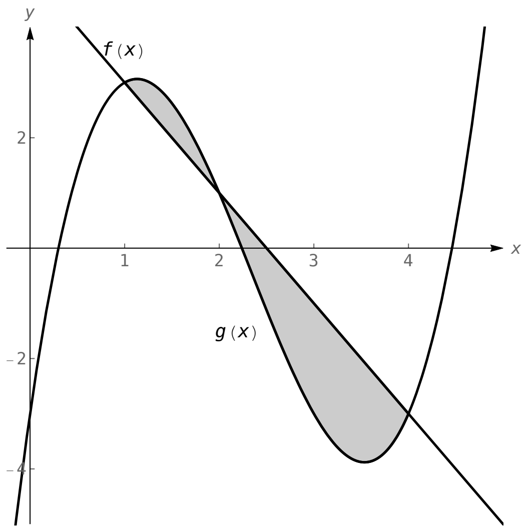


Figure 13.2: Graphing a region enclosed by two functions in Example 13.1.

Solution

A quick calculation shows that $f = g$ at $x = 1, 2$ and 4. One can proceed thoughtlessly by computing $\int_1^4 \big(f(x) - g(x)\big)\,dx$, but this ignores the fact that on $[1,2]$, $g(x) > f(x)$. Thus we compute the total area by breaking the interval $[1,4]$ into two subintervals, $[1,2]$ and $[2,4]$ and using the proper integrand in each.

$$\begin{aligned}
\text{Total Area} &= \int_1^2 \big(g(x) - f(x)\big)\,dx + \int_2^4 \big(f(x) - g(x)\big)\,dx \\
&= \int_1^2 \big(x^3 - 7x^2 + 14x - 8\big)\,dx + \int_2^4 \big(-x^3 + 7x^2 - 14x + 8\big)\,dx \\
&= \frac{5}{12} + \frac{8}{3} \\
&= \frac{37}{12} = 3.083 \text{ units}^2.
\end{aligned}$$

The previous example makes note that we are expecting area to be positive. When first learning about the definite integral, we interpreted it as signed area under the curve, allowing for negative area. That does not apply here; area is to be positive.

The previous example also demonstrates that we often have to break a given region into subregions. The following example shows another situation where this is applicable.

Example 13.2

Find the area of the region enclosed by the functions $y = \sqrt{x} + 2$, $y = -(x-1)^2 + 3$ and $y = 2$, as shown in Figure 13.3(a).

Solution

We give two approaches to this problem. In the first approach, we notice that the region's top is defined by two different curves. On $[0,1]$, the top function is $y=\sqrt{x}+2$; on $[1,2]$, the top function is $y=-(x-1)^2+3$.

Thus we compute the area as the sum of two integrals:

$$\begin{aligned} A &= \int_0^1 \Big((\sqrt{x}+2)-2 \Big)\,dx + \int_1^2 \Big((-(x-1)^2+3)-2 \Big)\,dx \\ &= \frac{2}{3}+\frac{2}{3} \\ &= \frac{4}{3}. \end{aligned}$$

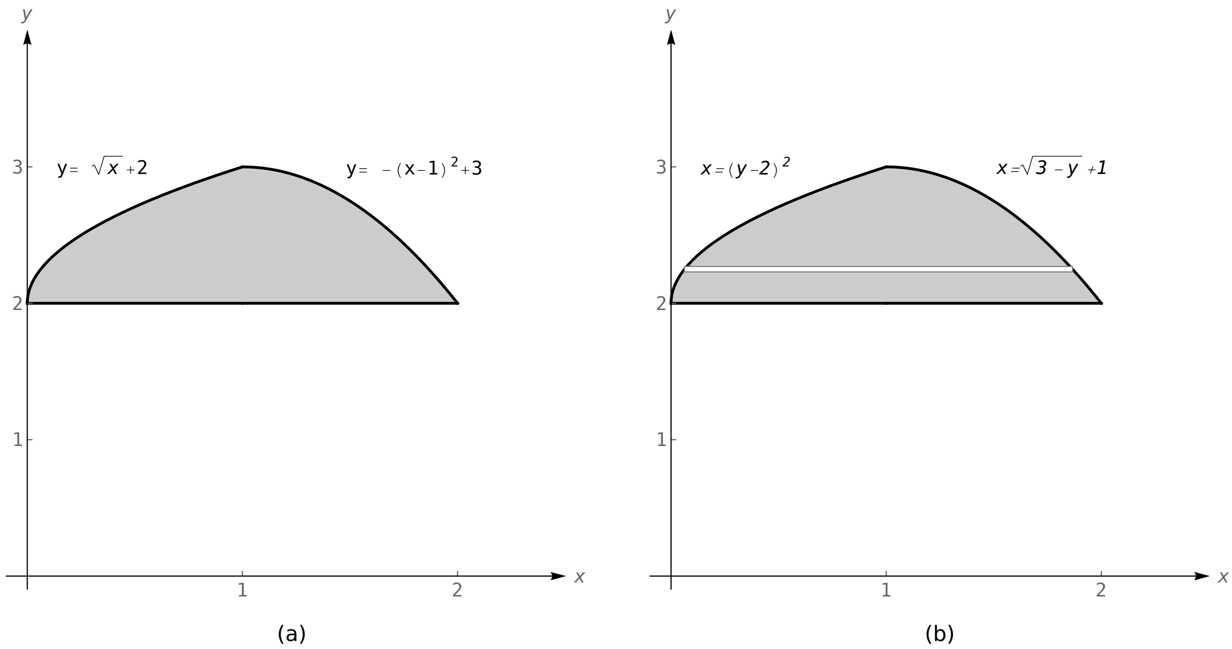


Figure 13.3: Graphing a region for Example 13.2 (a) and the region with boundaries relabelled as functions of y (b).

The second approach is clever and very useful in certain situations. We are used to viewing curves as functions of x; we input an x-value and a y-value is returned. Some curves can also be described as functions of y: input a y-value and an x-value is returned. We can rewrite the equations describing the boundary by solving for x:

$$y=\sqrt{x}+2 \quad \Rightarrow \quad x=(y-2)^2$$

$$y=-(x-1)^2+3 \quad \Rightarrow \quad x=\sqrt{3-y}+1.$$

Figure 13.3(b) shows the region with the boundaries relabelled. A differential element, a horizontal rectangle, is also pictured. The width of the rectangle is a small change in y: Δy. The height of the rectangle is a difference in x-values. The top x-value is the largest value, i.e., the rightmost.

The bottom x-value is the smaller, i.e., the leftmost. Therefore the height of the rectangle is

$$\left(\sqrt{3-y}+1\right)-(y-2)^2.$$

The area is found by integrating the above function with respect to y with the appropriate bounds. We determine these by considering the y-values the region occupies. It is bounded below by $y=2$, and bounded above by $y=3$. That is, both the top and bottom functions exist on the y interval $[2,3]$. Thus

$$\begin{aligned}
A &= \int_2^3 \left(\sqrt{3-y}+1-(y-2)^2\right) dy \\
&= \left(-\frac{2}{3}(3-y)^{3/2}+y-\frac{1}{3}(y-2)^3\right)\Big|_2^3 \\
&= \frac{4}{3}.
\end{aligned}$$

While we have focused on producing exact answers, we are also able to make approximations. The integrand in Equation (13.1) is a distance (top minus bottom); integrating this distance function gives an area. By taking discrete measurements of distance, we can approximate an area using numerical integration techniques developed in Chapter 12.

In the next section we apply our applications–of–integration techniques to finding the volumes of certain solids.

13.1.2 Polar coordinates

When using polar coordinates, the equations $\theta=\alpha$ form lines through the origin and $r=c$ form circles centred at the origin, respectively, and combinations of these curves form sectors of circles. It is then somewhat natural to calculate the area of regions defined by polar functions by first approximating with sectors of circles.

Consider Figure 13.4(a) where a region defined by $r=f(\theta)$ on $[\alpha,\beta]$ is given. Note how the sides of the region are the lines $\theta=\alpha$ and $\theta=\beta$, whereas in rectangular coordinates the sides of regions were often the vertical lines $x=a$ and $x=b$.

Partition the interval $[\alpha,\beta]$ into n equally spaced subintervals as $\alpha=\theta_1<\theta_2<\cdots<\theta_{n+1}=\beta$. The length of each subinterval is $\Delta\theta=(\beta-\alpha)/n$, representing a small change in angle. The area of the region defined by the i^{th} subinterval $[\theta_i,\theta_{i+1}]$ can be approximated with a sector of a circle with radius $f(c_i)$, for some c_i in $[\theta_i,\theta_{i+1}]$. The area of this sector is $\frac{1}{2}f(c_i)^2\Delta\theta$. This is shown in Figure 13.4(b), where $[\alpha,\beta]$ has been divided into 4 subintervals. We approximate the area of the whole region by summing the areas of all sectors:

$$\text{Area} \approx \sum_{i=1}^{n}\frac{1}{2}f(c_i)^2\Delta\theta.$$

This is a Riemann sum. By taking the limit of the sum as $n\to+\infty$, we find the exact area of the region in the form of a definite integral:

$$A = \frac{1}{2}\int_\alpha^\beta f(\theta)^2\,d\theta = \frac{1}{2}\int_\alpha^\beta r^2\,d\theta. \tag{13.2}$$

By having $0 \leq \beta - \alpha \leq 2\pi$, we ensure that the region does not overlap itself, which would give a result that does not correspond directly to the area.

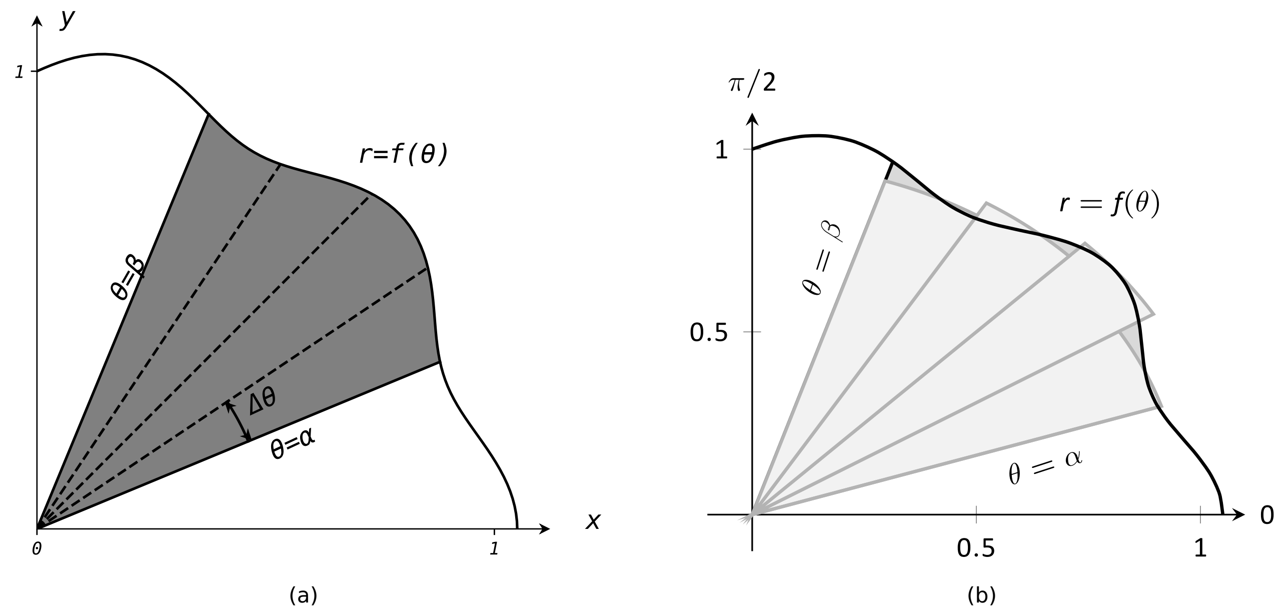


Figure 13.4: Computing the area of a polar region.

Example 13.3

Find the area of the cardioid $r = 1 + \cos(\theta)$ bound between $\theta = \pi/6$ and $\theta = \pi/3$, as shown in Figure 13.5.

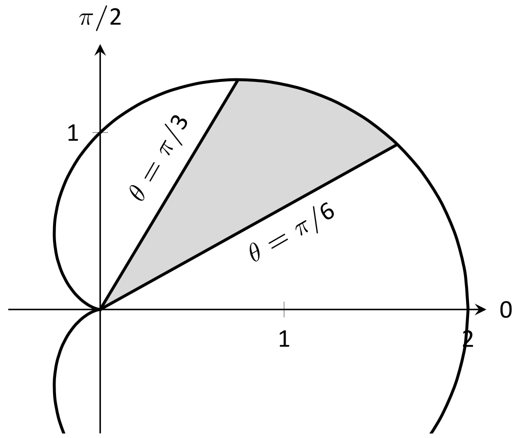


Figure 13.5: Finding the area of the shaded region of a cardioid in Example 13.3.

Solution

This is a direct application of Equation (13.2).

$$\text{Area} = \frac{1}{2}\int_{\pi/6}^{\pi/3} (1 + \cos(\theta))^2 \, d\theta$$

$$= \frac{1}{2}\int_{\pi/6}^{\pi/3} \left(1 + 2\cos(\theta) + \cos^2(\theta)\right)\, d\theta$$

$$= \frac{1}{2}\left(\theta + 2\sin(\theta) + \frac{1}{2}\theta + \frac{1}{4}\sin(2\theta)\right)\Bigg|_{\pi/6}^{\pi/3}$$

$$= \frac{1}{8}\left(\pi + 4\sqrt{3} - 4\right) \approx 0.7587$$

We may of course also determine the region enclosed between two polar curves. Consider for that purpose the shaded region shown in Figure 13.6. We can find the area of this region by computing the area bounded by $r_2 = f_2(\theta)$ and subtracting the area bounded by $r_1 = f_1(\theta)$ on $[\alpha, \beta]$. Thus

$$\text{Area} \;=\; \frac{1}{2}\int_{\alpha}^{\beta} r_2^2\, d\theta - \frac{1}{2}\int_{\alpha}^{\beta} r_1^2\, d\theta = \frac{1}{2}\int_{\alpha}^{\beta} \left(r_2^2 - r_1^2\right) d\theta. \tag{13.3}$$

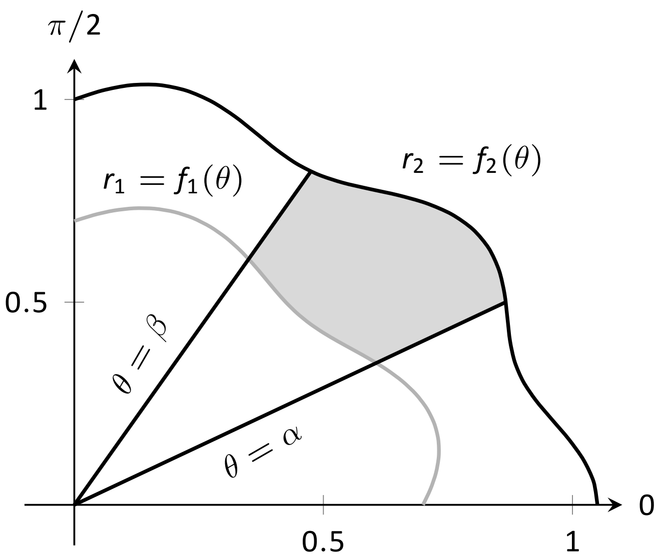


Figure 13.6: Illustrating area bound between two polar curves.

Example 13.4

Find the area bounded between the polar curves $r = 1$ and $r = 2\cos(2\theta)$, as shown in Figure 13.7(a).

Solution

We need to find the point of intersection between the two curves. Setting the two functions equal to each other, we have

$$2\cos(2\theta) = 1 \quad \Leftrightarrow \quad \cos(2\theta) = \frac{1}{2} \quad \Rightarrow \quad 2\theta = \frac{\pi}{3} \quad \Rightarrow \quad \theta = \frac{\pi}{6}.$$

In Figure 13.7(b), we zoom in on the region and note that it is not really bounded between two polar curves, but rather by two polar curves, along with $\theta = 0$. The dashed line breaks the region into its component parts. Below the dashed line, the region is defined by $r = 1$, $\theta = 0$ and $\theta = \pi/6$. Above the dashed line the region is bounded by $r = 2\cos(2\theta)$ and $\theta = \pi/6$. Since we have two separate regions, we find the area using two separate integrals.

Call the area below the dashed line A_1 and the area above the dashed line A_2. They are determined by the following integrals:

$$A_1 = \frac{1}{2}\int_0^{\pi/6} (1)^2\, d\theta \qquad \text{and} \qquad A_2 = \frac{1}{2}\int_{\pi/6}^{\pi/4} \big(2\cos(2\theta)\big)^2\, d\theta.$$

The upper bound of the integral for A_2 is $\pi/4$ as $r = 2\cos(2\theta)$ is at the pole when $\theta = \pi/4$. We omit the integration details and let the reader verify that $A_1 = \pi/12$ and $A_2 = \pi/12 - \sqrt{3}/8$; the total area is $A = \pi/6 - \sqrt{3}/8$.

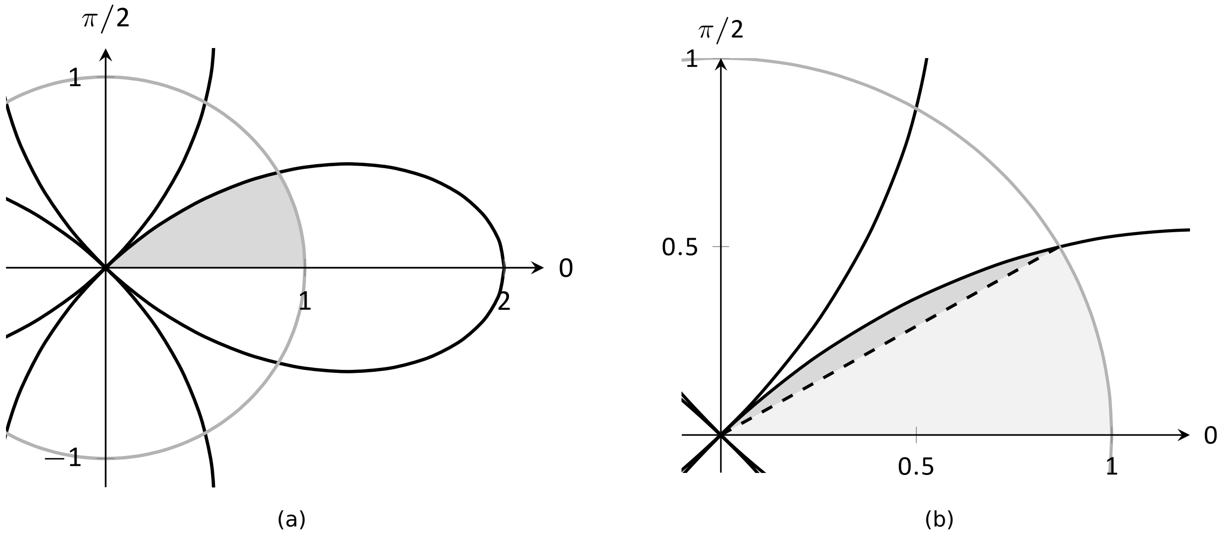


Figure 13.7: Graphing the region bounded by the functions in Example 13.4.

The error function

The error function is an example of a non-elementary function that contains an integral in its definition. More precisely, it is defined as:

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2}\, dt = \frac{2}{\sqrt{\pi}}\int_0^{x} e^{-t^2}\, dt.$$

It is of sigmoid shape and occurs in probability and statistic (Figure 13.8). There, for nonnegative values of x, it has the following interpretation: for a random variable Y that is normally distributed with mean 0 and variance 1/2, $\operatorname{erf}(x)$ describes the probability of Y falling in the range $[-x, x]$.

The error function

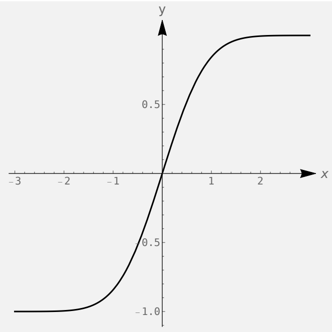


Figure 13.8: A graph of the error function.

13.2 Volume by cross-sectional area

13.2.1 Volumes by slicing

The volume of a general right cylinder, as shown in Figure 13.9, is

$$\text{Area of the base} \times \text{height}.$$

We can use this fact as the building block in finding volumes of a variety of shapes.

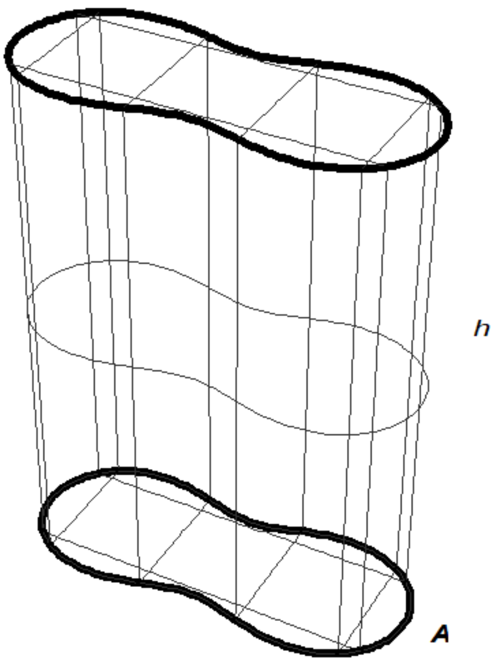


Figure 13.9: The volume of a general right cylinder.

Given an arbitrary solid, we can approximate its volume by cutting it into n thin slices. When the slices are thin, each slice can be approximated well by a general right cylinder. Thus the volume of each slice is approximately its cross-sectional area × thickness. These slices are the differential elements.

By orienting a solid along the x-axis, we can let $A(x_i)$ represent the cross-sectional area of the i^{th} slice, and let Δx_i represent the thickness of this slice. The total volume of the solid is approximately:

$$\begin{aligned} \text{Volume} &\approx \sum_{i=1}^{n} \Big[\text{Area} \times \text{thickness} \Big] \\ &= \sum_{i=1}^{n} A(x_i)\Delta x_i. \end{aligned}$$

Recognize that this is a Riemann sum. By taking a limit as the thickness of the slices goes to 0 we can find the volume exactly.

Theorem 13.1 (Volume by cross-sectional area)
The volume V of a solid, oriented along the x-axis with cross-sectional area $A(x)$ from $x = a$ to $x = b$, is

$$V = \int_a^b A(x)\, dx.$$

Example 13.5

Find the volume of a pyramid with a square base of side length 10 cm and a height of 5 cm.

Solution

There are many ways to orient the pyramid along the x-axis; Figure 13.10(a) gives one such way, with the pointed top of the pyramid at the origin and the x-axis going through the centre of the base.

Each cross section of the pyramid is a square; this is a sample differential element. To determine its area $A(x)$, we need to determine the side lengths of the square. When $x = 5$, the square has side length 10; when $x = 0$, the square has side length 0. Since the edges of the pyramid are lines, it is easy to figure that each cross-sectional square has side length $2x$, giving $A(x) = (2x)^2 = 4x^2$.

If one were to cut a slice out of the pyramid at $x = 3$, as shown in Figure 13.10(b), one would have a shape with square bottom and top with sloped sides. If the slice were thin, both the bottom and top squares would have sides lengths of about 6, and thus the cross–sectional area of the bottom and top would be about 36cm^2. Letting Δx_i represent the thickness of the slice, the volume of this slice would then be about $36\Delta x_i$ cm^3.

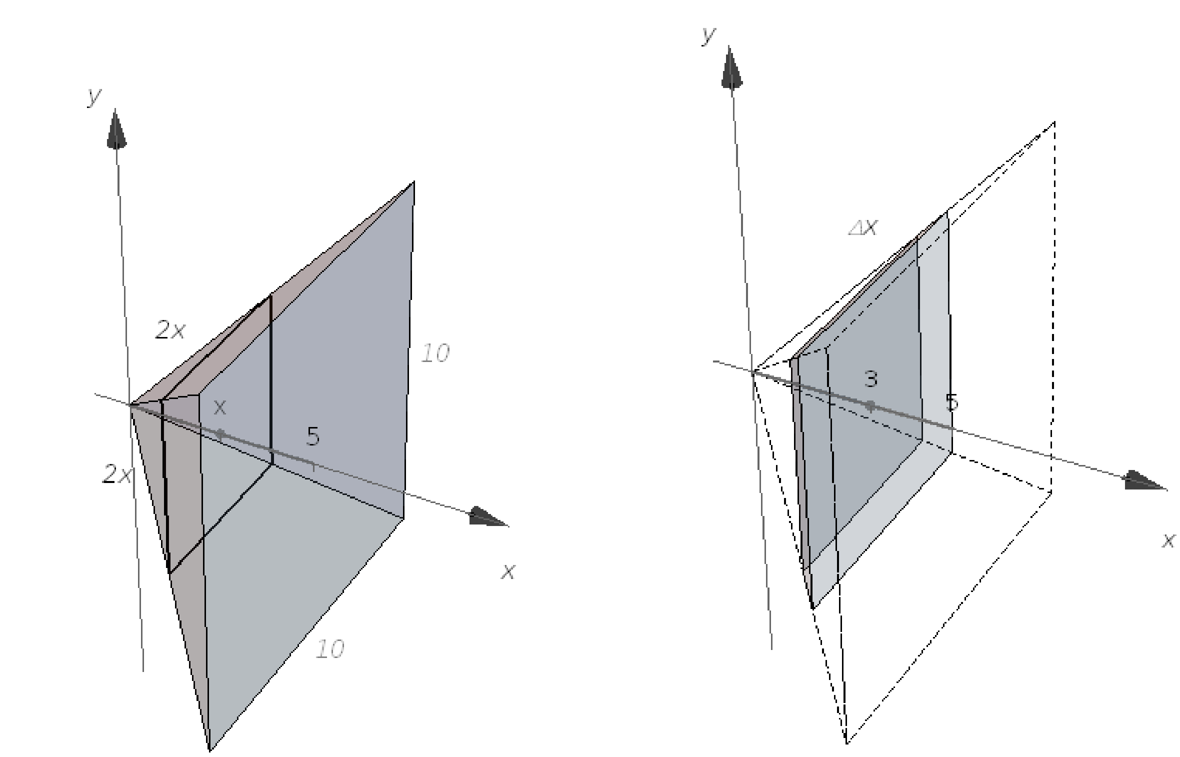
Cutting the pyramid into n slices divides the total volume into n equally–spaced smaller pieces, each with volume $(2x_i)^2\Delta x$, where x_i is the approximate location of the slice along the x-axis and Δx represents the thickness of each slice. One can approximate the total volume of the pyramid by summing up the volumes of these slices:

$$\text{Approximate volume } = \sum_{i=1}^{n} (2x_i)^2 \Delta x.$$

Taking the limit as $n \to +\infty$ gives the actual volume of the pyramid; recognizing this sum as a Riemann sum allows us to find the exact answer using a definite integral, matching the definite integral given by Theorem 13.1.

We have

$$\begin{aligned} V &= \lim_{n\to+\infty} \sum_{i=1}^{n} (2x_i)^2 \Delta x \\ &= \int_0^5 4x^2\, dx \\ &= \frac{4}{3}x^3\Big|_0^5 \\ &= \frac{500}{3} \text{ in}^3 \approx 166.67 \text{ cm}^3. \end{aligned}$$



(a) (b)

Figure 13.10: Orienting a pyramid along the x-axis (a) and cutting a slice in it at $x = 3$ (b) in Example 13.5.

13.2.2 Solids of revolution

An important special case of Theorem 13.1 is when the solid is a **solid of revolution** (*omwentelingslichaam*), that is, when the solid is formed by rotating a shape about an axis.

Start with a function $y = f(x)$ from $x = a$ to $x = b$. Revolving this curve about a horizontal axis creates a three-dimensional solid whose cross sections are disks (thin circles). Let $R(x)$ represent the radius of the cross-sectional disk at x; the area of this disk is $\pi R(x)^2$. Applying Theorem 13.1 gives the disk method.

More precisely, let a solid be formed by revolving the curve $y = f(x)$ from $x = a$ to $x = b$ about a horizontal axis, and let $R(x)$ be the radius of the cross-sectional disk at x. The volume of the resulting solid is

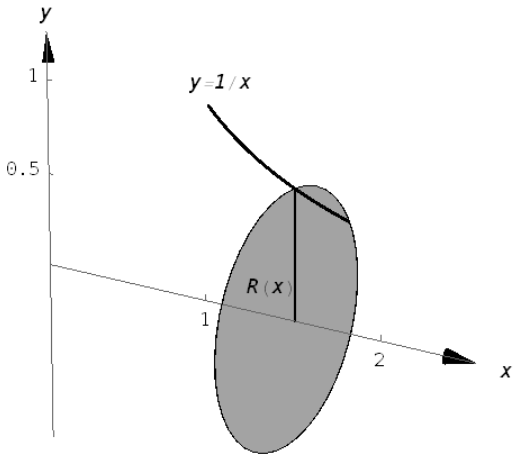
$$V = \pi \int_a^b R(x)^2 \, dx. \tag{13.4}$$

Example 13.6

Find the volume of the solid formed by revolving the curve $y = 1/x$, from $x = 1$ to $x = 2$, about the x-axis.

Solution

A sketch can help us understand this problem. In Figure 13.11(a) the curve $y = 1/x$ is sketched along with the differential element – a disk – at x with radius $R(x) = 1/x$. In Figure 13.11(b) the whole solid is pictured, along with the differential element.



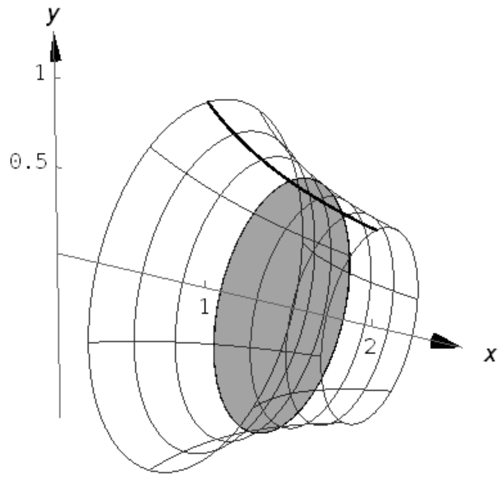


Figure 13.11: Sketching a solid in Example 13.6.

The volume of the differential element shown in part (a) of the figure is approximately $\pi R(x_i)^2\Delta x$, where $R(x_i)$ is the radius of the disk shown and Δx is the thickness of that slice. The radius $R(x_i)$ is the distance from the x-axis to the curve, hence $R(x_i) = 1/x_i$.

Slicing the solid into n equally–spaced slices, we can approximate the total volume by adding up the approximate volume of each slice:

$$\text{Approximate volume } = \sum_{i=1}^{n} \pi \left(\frac{1}{x_i}\right)^2 \Delta x.$$

Taking the limit of the above sum as $n \to +\infty$ gives the actual volume; recognizing this sum as a Riemann sum allows us to evaluate the limit with a definite integral, which matches Equation (13.4):

$$\begin{aligned} V &= \lim_{n\to+\infty} \sum_{i=1}^{n} \pi \left(\frac{1}{x_i}\right)^2 \Delta x \\ &= \pi \int_1^2 \left(\frac{1}{x}\right)^2 \, dx \\ &= \pi \int_1^2 \frac{1}{x^2} \, dx \\ &= \pi \left[-\frac{1}{x}\right]\Bigg|_1^2 \\ &= \pi \left[-\frac{1}{2} - (-1)\right] \end{aligned}$$

$$= \frac{\pi}{2} \text{ units}^3.$$

While Equation (13.4) is given in terms of functions of x, the principle involved can be as well applied to functions of y when the axis of rotation is vertical, not horizontal. We demonstrate this in the next example.

Example 13.7

Find the volume of the solid formed by revolving the curve $y = 1/x$, from $x = 1$ to $x = 2$, about the y-axis.

Solution

Since the axis of rotation is vertical, we need to convert the function into a function of y and convert the x-bounds to y-bounds. Since $y = 1/x$ defines the curve, we rewrite it as $x = 1/y$. The bound $x = 1$ corresponds to the y-bound $y = 1$, and the bound $x = 2$ corresponds to the y-bound $y = 1/2$. Thus we are rotating the curve $x = 1/y$, from $y = 1/2$ to $y = 1$ about the y-axis to form a solid. The curve and sample differential element are sketched in Figure 13.12(a), with a full sketch of the solid in Figure 13.12(b).

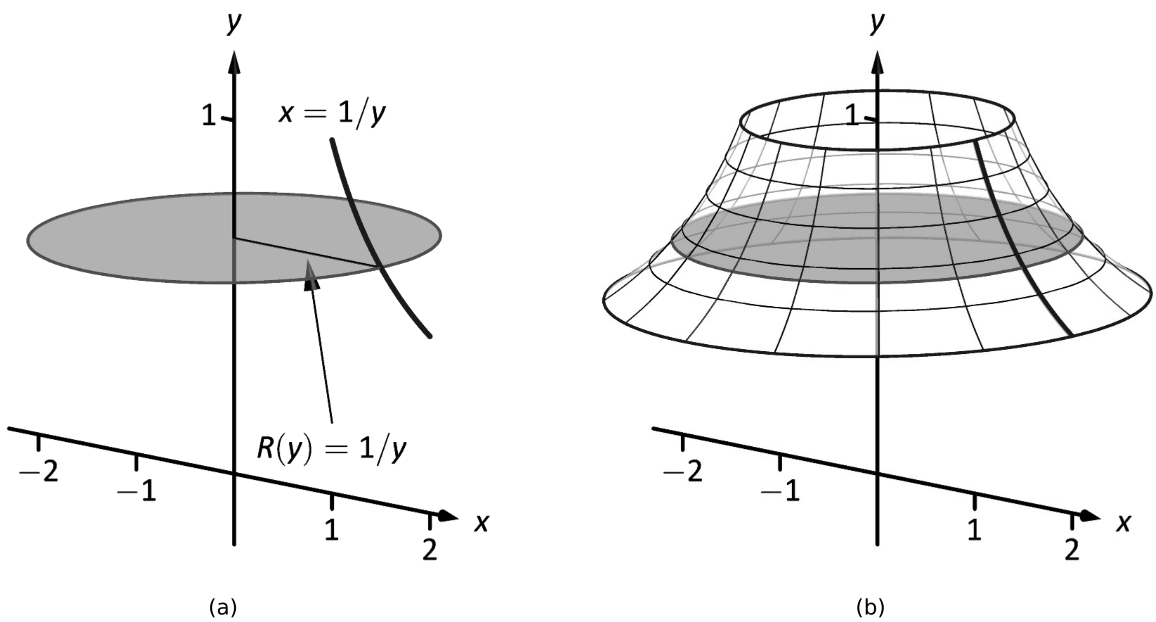


Figure 13.12: Sketching a solid in Example 13.7.

We integrate to find the volume:

$$V = \pi \int_{1/2}^{1} \frac{1}{y^2}\, dy$$

$$= -\frac{\pi}{y}\Big|_{1/2}^{1}$$

$$= \pi \text{ units}^3.$$

We can also compute the volume of solids of revolution that have a hole in the centre. The general principle is simple: compute the volume of the solid irrespective of the hole, then subtract the volume of the hole. If the outside radius of the solid is $R(x)$ and the inside radius (defining the hole) is $r(x)$, then the volume is

$$V = \pi \int_a^b R(x)^2 \, dx - \pi \int_a^b r(x)^2 \, dx = \pi \int_a^b \left(R(x)^2 - r(x)^2 \right) \, dx.$$

One can generate a solid of revolution with a hole in the middle by revolving a region about an axis. Consider Figure 13.13(a), where a region is sketched along with a dashed, horizontal axis of rotation. By rotating the region about the axis, a solid is formed as sketched in Figure 13.13(b). The outside of the solid has radius $R(x)$, whereas the inside has radius $r(x)$. Each cross section of this solid will be a washer (a disk with a hole in the center) as sketched in Figure 13.13(c). This leads us to the washer method.

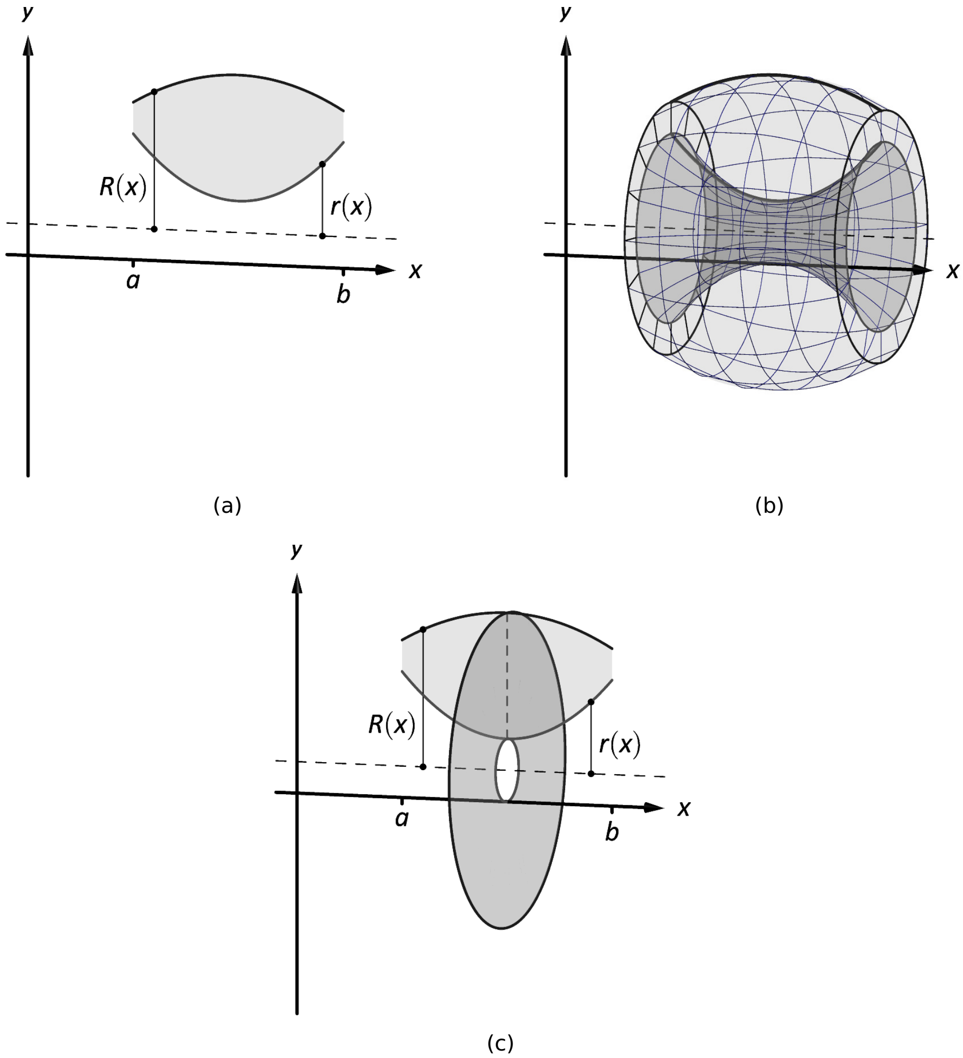


Figure 13.13: Establishing the washer method.

Let a region bounded by $y = f(x)$, $y = g(x)$, $x = a$ and $x = b$ be rotated about a horizontal axis that does not intersect the region, forming a solid. Each cross section at x will be a washer with outside radius $R(x)$ and inside radius $r(x)$. The volume of the solid is

$$V = \pi \int_a^b \left(R(x)^2 - r(x)^2 \right) dx. \tag{13.5}$$

Obviously, the disk method is just a special case of the washer method with an inside radius of $r(x) = 0$.

Example 13.8

Find the volume of the solid formed by rotating the region bounded by $y = x^2 - 2x + 2$ and $y = 2x - 1$ about the x-axis.

Solution

A sketch of the region will help, as given in Figure 13.10(a). Rotating about the x-axis will produce cross sections in the shape of washers, as shown in Figure 13.14(b); the complete solid is shown in Figure 13.14(c). The outside radius of this washer is $R(x) = 2x - 1$; the inside radius is $r(x) = x^2 - 2x + 2$. As the region is bounded from $x = 1$ to $x = 3$, we integrate as follows to compute the volume.

$$\begin{aligned} V &= \pi \int_1^3 \left((2x-1)^2 - (x^2 - 2x + 2)^2 \right) dx \\ &= \pi \int_1^3 \left(-x^4 + 4x^3 - 4x^2 + 4x - 3 \right) dx \\ &= \pi \left[-\frac{1}{5}x^5 + x^4 - \frac{4}{3}x^3 + 2x^2 - 3x \right]\Bigg|_1^3 \\ &= \frac{104}{15}\pi \text{ units}^3 \approx 21.78 \text{ units}^3. \end{aligned}$$

When rotating about a vertical axis, the outside and inside radius functions must be functions of y.

Example 13.9

Find the volume of the solid formed by rotating the triangular region with vertices at $(1,1)$, $(2,1)$ and $(2,3)$ about the y-axis.

Solution

The triangular region is sketched in Figure 13.15(a); the differential element is sketched in Figure 13.15(b) and the full solid is drawn in Figure 13.15(c). They help us establish the outside and inside radii. Since the axis of rotation is vertical, each radius is a function of y.

The outside radius $R(y)$ is formed by the line connecting $(2,1)$ and $(2,3)$; it is a constant function, because $R(y) = 2$. The inside radius is formed by the line connecting $(1,1)$ and $(2,3)$. The equation of this line is $y = 2x - 1$, but we need to refer to it as a function of y. Solving for x gives $r(y) = \frac{1}{2}(y+1)$.

We integrate over the y-bounds of $y = 1$ to $y = 3$. Thus the volume is

$$V = \pi \int_1^3 \left(2^2 - \left(\frac{1}{2}(y+1) \right)^2 \right) dy$$

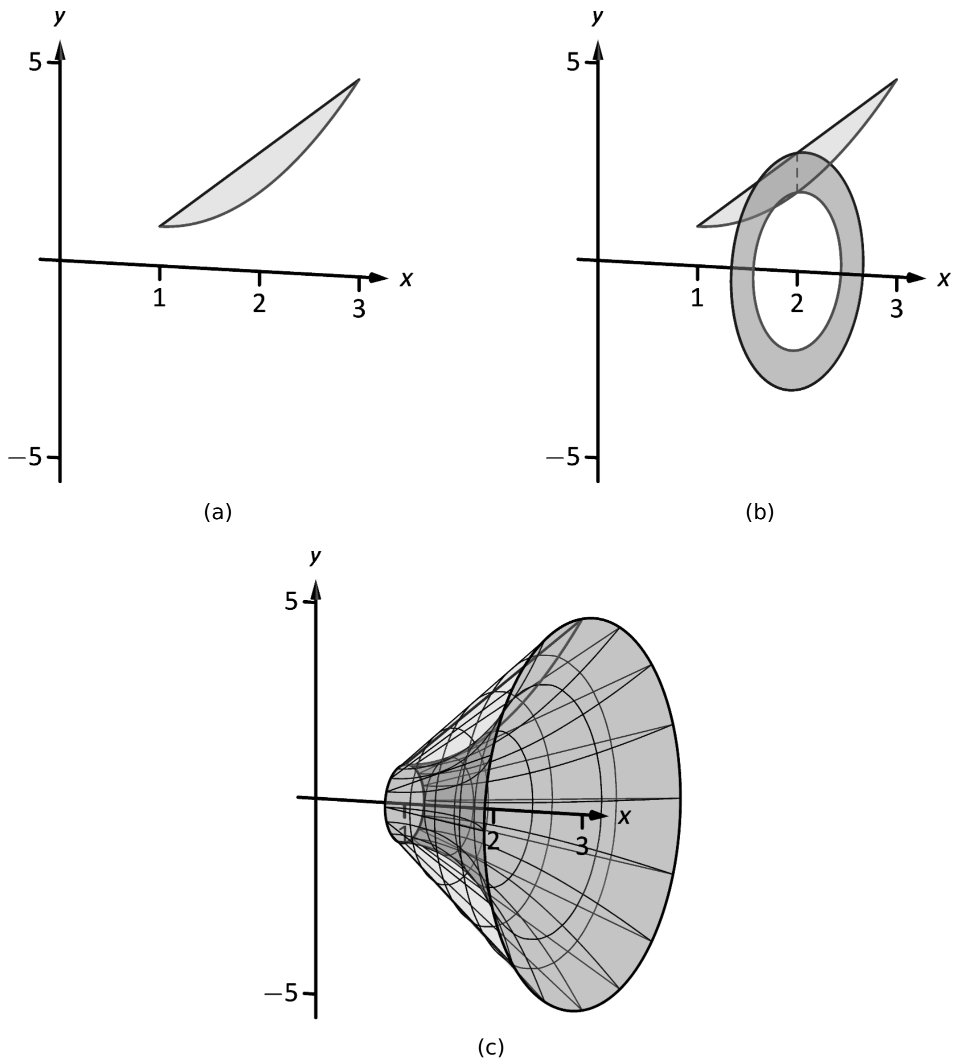


Figure 13.14: Sketching the differential element and solid in Example 13.8.

$$= \pi \int_1^3 \left(-\frac{1}{4}y^2 - \frac{1}{2}y + \frac{15}{4} \right) dy$$

$$= \pi \left[-\frac{1}{12}y^3 - \frac{1}{4}y^2 + \frac{15}{4}y \right]\Bigg|_1^3$$

$$= \frac{10}{3}\pi \text{units}^3 \approx 10.47 \text{ units}^3.$$

The ultimate goal of this section is not to compute volumes of solids. That can be useful, but what is more useful is the understanding of this basic principle of integral calculus: to find the exact value of some quantity,

- we start with an approximation (in this section, slice the solid and approximate the volume of each slice),
- then make the approximation better by refining our original approximation (i.e., use more slices),
- then use limits to establish a definite integral which gives the exact value.

We practice this principle in the next section where we find volumes by slicing solids in a different way.

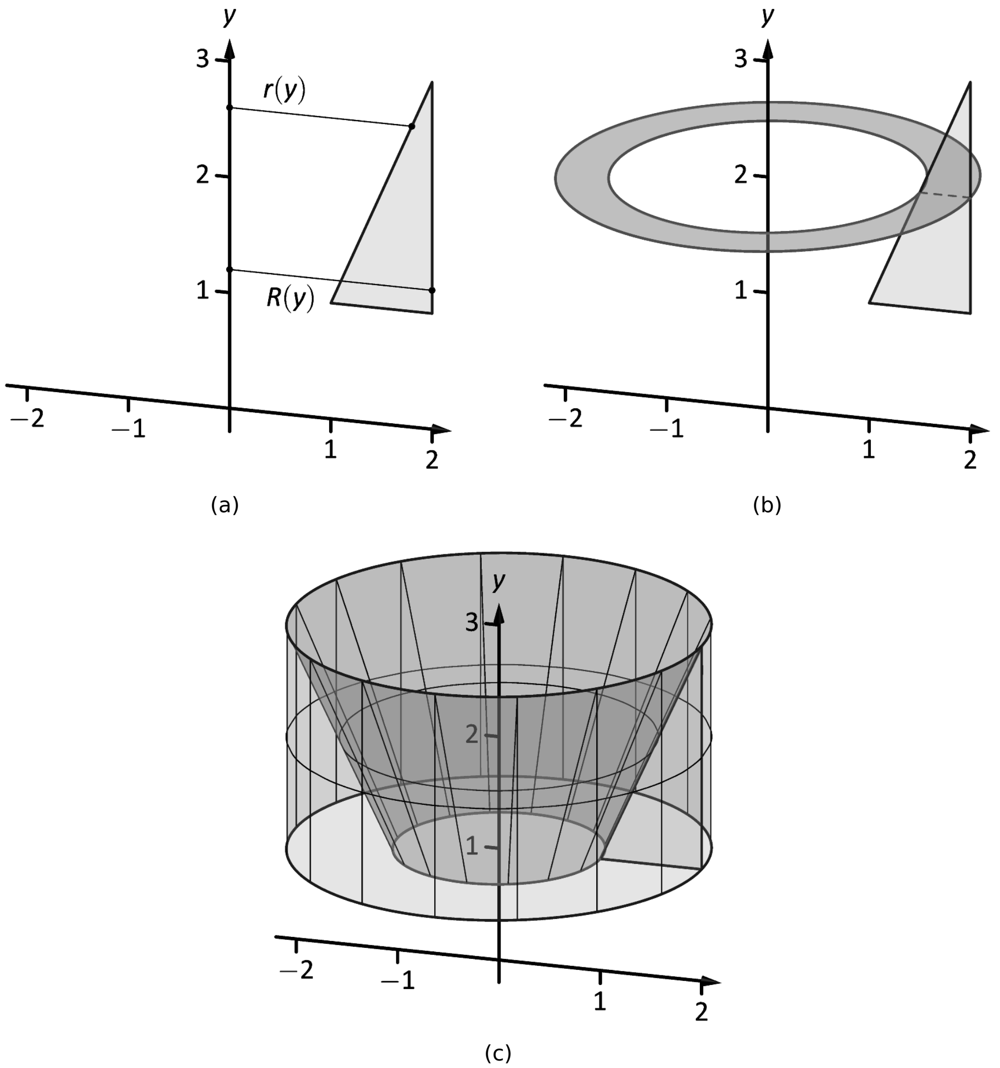


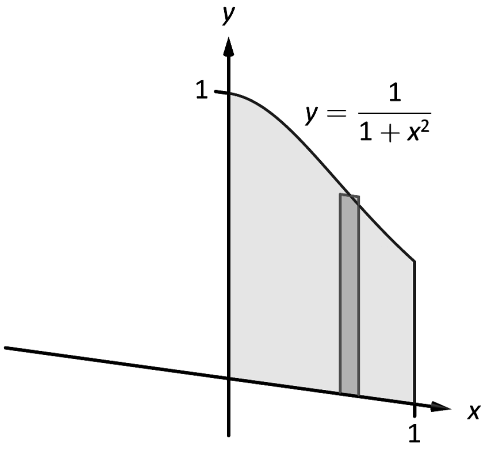
Figure 13.15: Sketching the differential element and solid in Example 13.9.

13.3 The shell method

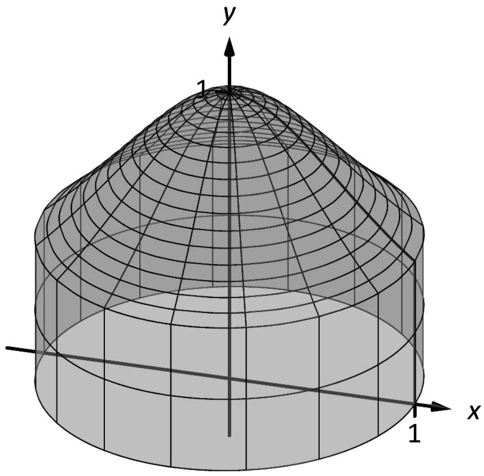
This section develops another method of computing volume, the **shell method.** Instead of slicing the solid perpendicular to the axis of rotation creating cross-sections, we now slice it parallel to the axis of rotation, creating shells.

Consider Figure 13.16(a), where the region is rotated about the y-axis forming the solid shown in Figure 13.16(b). A small slice of the region is drawn in Figure 13.16(a), parallel to the axis of rotation. When the region is rotated, this thin slice forms a cylindrical shell, as pictured in Figure 13.16(c). The previous section approximated a solid with lots of thin disks (or washers); we now approximate a solid with many thin cylindrical shells.

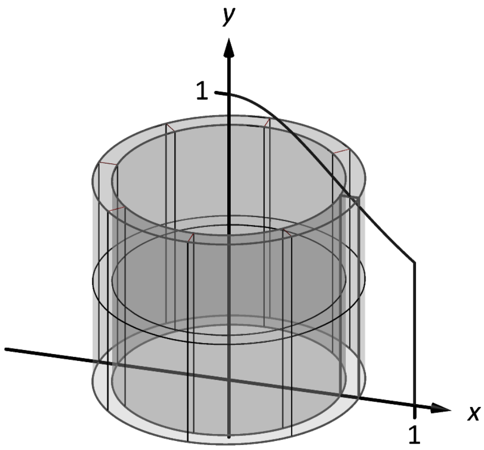
To compute the volume of one shell, first consider the paper label on a soup can with radius r and height h. What is the area of this label? A simple way of determining this is to cut the label and lay it out flat, forming a rectangle with height h and length $2\pi r$. Thus the area is $A = 2\pi rh$; see Figure 13.17(a). Do a similar process with a cylindrical shell, with height h, thickness Δx, and approximate radius r. Cutting the shell and laying it flat forms a rectangular solid with length $2\pi r$, height h and depth Δx. Thus the volume is $V \approx 2\pi rh\Delta x$; see Figure 13.17(b). We say approximately since our radius was an approximation.



(a)



(b)



(c)

Figure 13.16: The shell method.

By breaking the solid into n cylindrical shells, we can approximate the volume of the solid as

$$V \approx \sum_{i=1}^{n} 2\pi r_i h_i \Delta x_i,$$

where r_i, h_i and Δx_i are the radius, height and thickness of the i^{th} shell, respectively. This is a Riemann sum. Taking a limit as the thickness of the shells approaches 0 leads to a definite integral. So we arrive at the following.

Let a solid be formed by revolving a region R, bounded by $x = a$ and $x = b$, about a vertical axis. Let $r(x)$ represent the distance from the axis of rotation to x (i.e., the radius of a sample shell) and let $h(x)$ represent the height of the solid at x (i.e., the height of the shell). The volume of the solid is

$$V = 2\pi \int_a^b r(x)h(x)\ dx. \tag{13.6}$$

There are two special cases:

1. When the region R is bounded above by $y = f(x)$ and below by $y = g(x)$, then $h(x) = f(x) - g(x)$.
2. When the axis of rotation is the y-axis (i.e., $x = 0$) then $r(x) = x$.

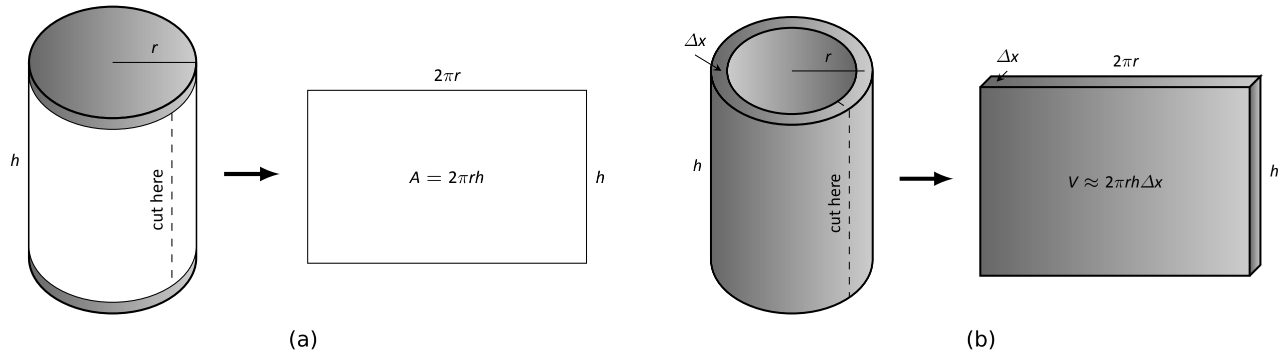


Figure 13.17: Determining the volume of a thin cylindrical shell.

Let us practice using this method.

Example 13.10

Find the volume of the solid formed by rotating the region bounded by $y=0$, $y=1/(1+x^2)$, $x=0$ and $x=1$ about the y-axis.

Solution

This is the region used to introduce the shell method in Figure 13.16(a), but is sketched again in Figure 13.18 for closer reference. A line is drawn in the region parallel to the axis of rotation representing a shell that will be carved out as the region is rotated about the y-axis. This is the differential element.

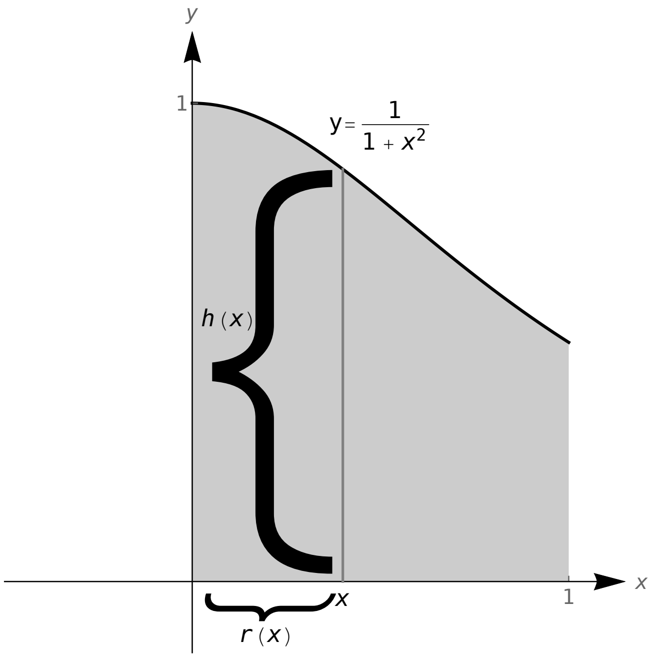


Figure 13.18: Graphing a region in Example 13.10.

The distance this line is from the axis of rotation determines $r(x)$; as the distance from x to the y-axis is x, we have $r(x)=x$. The height of this line determines $h(x)$; the top of the line is at $y=1/(1+x^2)$, whereas the bottom of the line is at $y=0$. Thus $h(x)=1/(1+x^2)-0=1/(1+x^2)$.

The region is bounded from $x=0$ to $x=1$, so the volume is

$$V = 2\pi \int_0^1 \frac{x}{1+x^2}\ dx.$$

This requires substitution. Let $u = 1+x^2$, so $du = 2x\ dx$. We also change the bounds: $u(0)=1$ and $u(1)=2$. Thus we have:

$$\begin{aligned} &= \pi \int_1^2 \frac{1}{u}\ du \\ &= \pi \ln(u)\Big|_1^2 \\ &= \pi \ln(2) \approx 2.178 \text{ units}^3. \end{aligned}$$

Note that in order to find this volume using the disk method, two integrals would be needed to account for the regions above and below $y = 1/2$.

With the shell method, nothing special needs to be accounted for to compute the volume of a solid that has a hole in the middle, as demonstrated next.

Example 13.11

Find the volume of the solid formed by rotating the triangular region determined by the points $(0,1)$, $(1,1)$ and $(1,3)$ about the line $x=3$.

Solution

The region is sketched in Figure 13.19(a) along with the differential element, a line within the region parallel to the axis of rotation. In Figure 13.19(b), we see the shell traced out by the differential element, and in Figure 13.19(c) the whole solid is shown.

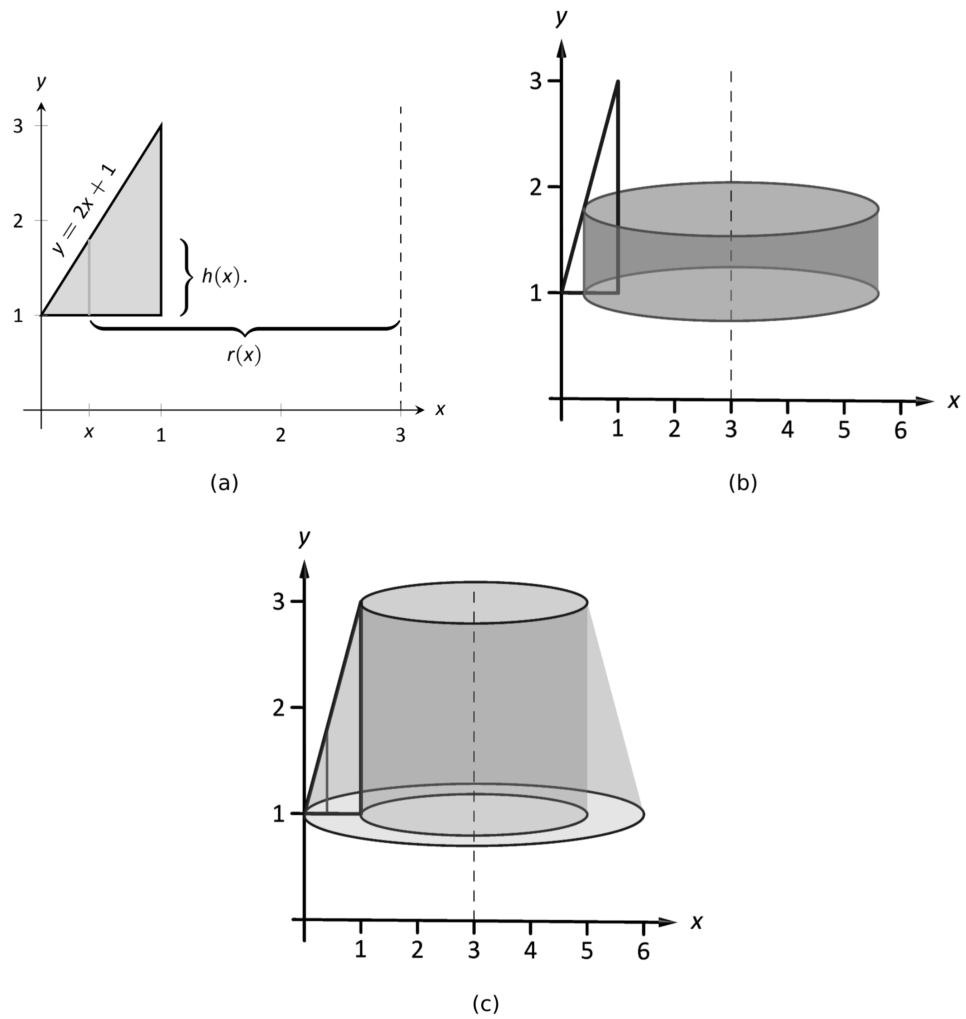


Figure 13.19: Graphing a region in Example 13.11.

The height of the differential element is the distance from $y = 1$ to $y = 2x+1$, the line that connects the points $(0,1)$ and $(1,3)$. Thus $h(x) = 2x+1-1 = 2x$. The radius of the shell formed by the differential element is the distance from x to $x = 3$; that is, it is $r(x) = 3-x$. The x-bounds of the region are $x = 0$ to $x = 1$, giving

$$V = 2\pi \int_0^1 (3-x)(2x)\,dx$$

$$= 2\pi \int_0^1 \left(6x-2x^2\right)\,dx$$

$$= 2\pi \left[3x^2 - \frac{2}{3}x^3\right]\Big|_0^1$$

$$= \frac{14}{3}\pi \text{ units}^3 \approx 14.66 \text{ units}^3.$$

When revolving a region about a horizontal axis, we must consider the radius and height functions in terms of y, not x.

Example 13.12

Find the volume of the solid formed by rotating the region given in Example 13.11 about the x-axis.

Solution

The region is sketched in Figure 13.20(a) with a sample differential element. In Figure 13.20(b) the shell formed by the differential element is drawn, and the solid is sketched in Figure 13.20(c).

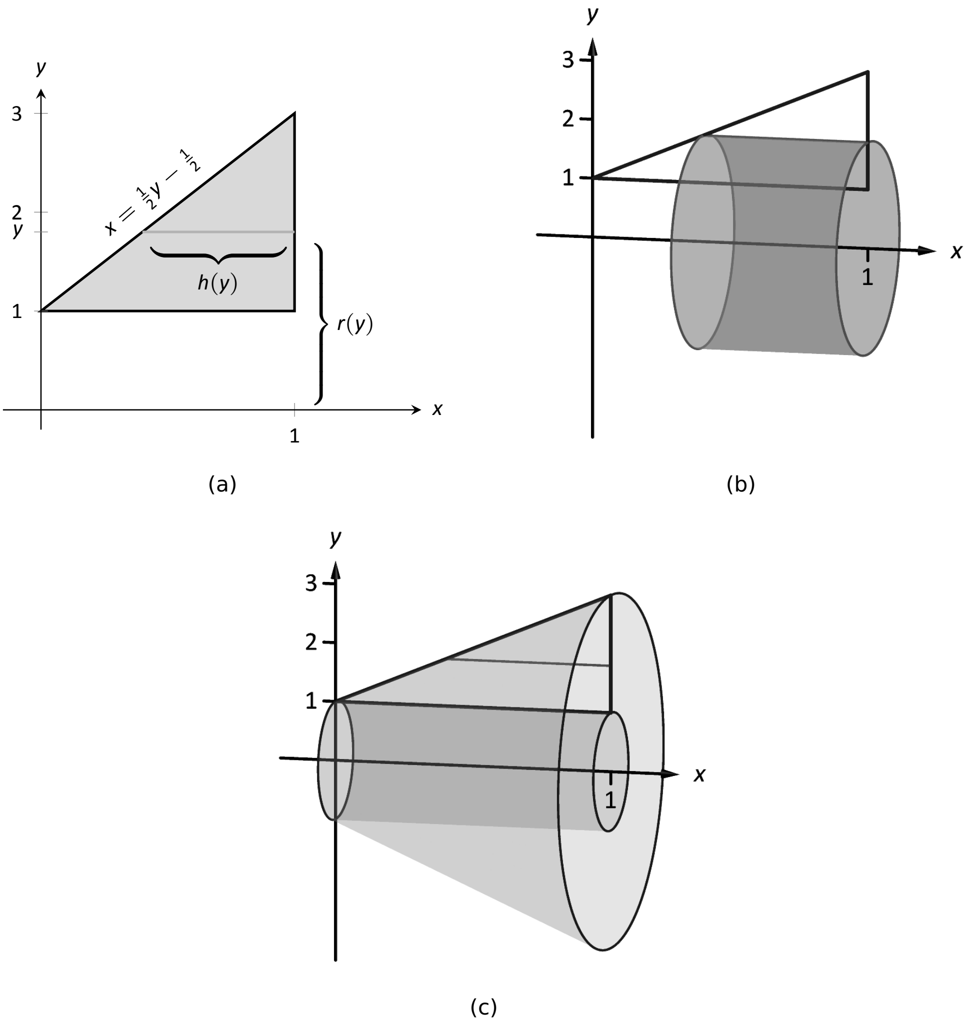


Figure 13.20: Graphing a region in Example 13.12.

The height of the differential element is an x-distance, between $x = y/2 - 1/2$ and $x = 1$. Thus

$$h(y) = 1 - \left(\frac{1}{2}y - \frac{1}{2}\right) = -\frac{1}{2}y + \frac{3}{2}.$$

The radius is the distance from y to the x-axis, so $r(y) = y$. The y bounds of the region are $y = 1$ and $y = 3$, leading to the integral

$$V = 2\pi \int_1^3 \left[y\left(-\frac{1}{2}y + \frac{3}{2}\right)\right] \, dy$$

$$= 2\pi \int_1^3 \left[-\frac{1}{2}y^2 + \frac{3}{2}y \right] dy$$

$$= 2\pi \left[-\frac{1}{6}y^3 + \frac{3}{4}y^2 \right] \Bigg|_1^3$$

$$= 2\pi \left[\frac{9}{4} - \frac{7}{12} \right]$$

$$= \frac{10}{3}\pi \text{ units}^3 \approx 10.472 \text{ units}^3.$$

We end this section with a table summarizing the usage of the washer and shell Methods.

	Washer method	Shell method
Horizontal axis	$\pi \int_a^b \left(R(x)^2 - r(x)^2\right) dx$	$2\pi \int_c^d r(y)h(y)\, dy$
Vertical axis	$\pi \int_c^d \left(R(y)^2 - r(y)^2\right) dy$	$2\pi \int_a^b r(x)h(x)\, dx$

In this section, we approximate the volume of a solid by cutting it into thin cylindrical shells. By summing up the volumes of each shell, we get an approximation of the volume. By taking a limit as the number of equally spaced shells goes to infinity, our summation can be evaluated as a definite integral, giving the exact value.

We use this same principle again in the next section, where we find the length of curves in the plane.

13.4 Arc length

13.4.1 Rectangular coordinates

In this section, we address the question: Given a curve, what is its length? This is often referred to as **arc length** (*booglengte*).

Consider the graph of $y = \sin(x)$ on $[0, \pi]$ given in Figure 13.21(a). How long is this curve? That is, if we were to use a piece of string to exactly match the shape of this curve, how long would the string be?

As we have done in the past, we start by approximating; later, we will refine our answer using limits to get an exact solution.

The length of straight–line segments is easy to compute using the distance formula. We can approximate the length of the given curve by approximating the curve with straight lines and measuring their lengths.

In Figure 13.21(b), the curve $y = \sin(x)$ has been approximated with 4 line segments, i.e. the interval $[0, \pi]$ has been divided into 4 equally–lengthed subintervals. It is clear that these four line segments approximate $y = \sin(x)$ very well on the first and last subinterval, though not so well in the middle.

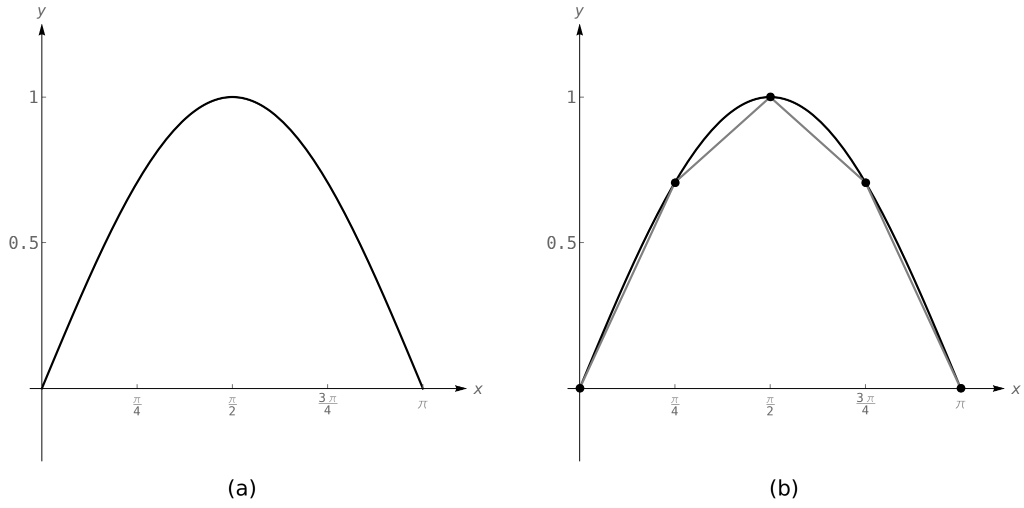


Figure 13.21: Graphing $y = \sin(x)$ on $[0, \pi]$ (a) and approximating the curve with line segments (b).

Regardless, the sum of the lengths of the line segments is 3.79, so we approximate the arc length of $y = \sin(x)$ on $[0, \pi]$ to be 3.79.

In general, we can approximate the arc length of $y = f(x)$ on $[a, b]$ in the following manner. Let $a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b$ be a partition of $[a, b]$ into n subintervals. Let Δx_i represent the length of the i^{th} subinterval $[x_i, x_{i+1}]$. Figure 13.22 zooms in on the i^{th} subinterval where $y = f(x)$ is approximated by a straight line segment. The dashed lines show that we can view this line segment as the hypotenuse of a right triangle whose sides have length Δx_i and Δy_i. Using the Pythagorean theorem, the length of this line segment is $\sqrt{\Delta x_i^2 + \Delta y_i^2}$. Summing over all subintervals gives an arc length approximation

$$L \approx \sum_{i=1}^{n} \sqrt{\Delta x_i^2 + \Delta y_i^2}.$$

As shown here, this is not a Riemann sum. While we could conclude that taking a limit as the subinterval length goes to zero gives the exact arc length, we would not be able to compute the answer with a definite integral. We need first to do a little algebra.

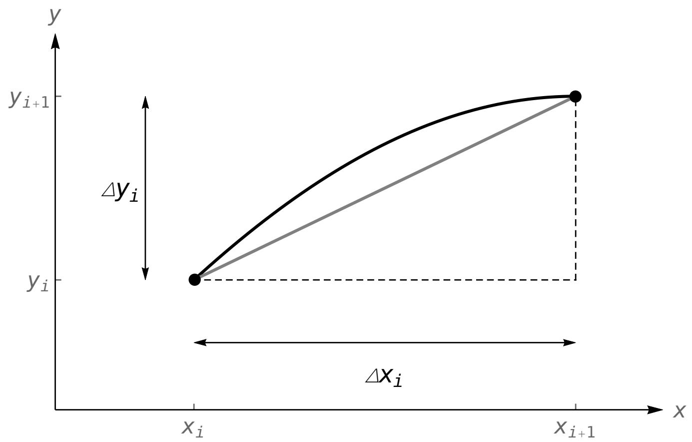


Figure 13.22: Zooming in on the i^{th} subinterval $[x_i, x_{i+1}]$ of a partition of $[a, b]$.

In the above expression factor out a Δx_i^2 term:

$$\sum_{i=1}^{n}\sqrt{\Delta x_i^2+\Delta y_i^2}=\sum_{i=1}^{n}\sqrt{\Delta x_i^2\left(1+\frac{\Delta y_i^2}{\Delta x_i^2}\right)}.$$

Now pull the Δx_i^2 term out of the square root:

$$\sum_{i=1}^{n}\sqrt{\Delta x_i^2+\Delta y_i^2}=\sum_{i=1}^{n}\sqrt{1+\frac{\Delta y_i^2}{\Delta x_i^2}}\ \Delta x_i.$$

This is nearly a Riemann sum. Consider the $\Delta y_i^2/\Delta x_i^2$ term. The expression $\Delta y_i/\Delta x_i$ measures the (change in y)/(change in x) of f on the i^{th} subinterval. The mean value theorem of differentiation (Theorem 10.4) states that there is a c_i in the i^{th} subinterval where $f'(c_i)=\Delta y_i/\Delta x_i$. Thus we can rewrite our above expression as:

$$\sum_{i=1}^{n}\sqrt{\Delta x_i^2+\Delta y_i^2}=\sum_{i=1}^{n}\sqrt{1+f'(c_i)^2}\ \Delta x_i.$$

This is a Riemann sum. As long as f' is continuous, we can invoke Theorem 12.4 and conclude

$$L=\int_a^b\sqrt{1+f'(x)^2}\ dx.$$

This result is summarized in the following theorem.

Theorem 13.2 (Arc length)
Let f be differentiable on $[a,b]$, where f' is also continuous on $[a,b]$. Then the arc length of f from $x=a$ to $x=b$ is

$$L=\int_a^b\sqrt{1+f'(x)^2}\ dx. \tag{13.7}$$

The theorem also requires that f' is continuous on $[a,b]$; while examples are arcane, it is possible for f to be differentiable yet f' is not continuous.

As the integrand contains a square root, it is often difficult to use Equation (13.7) to find the length exactly. When exact answers are difficult to come by, we resort to using numerical methods.

Example 13.13

Find the arc length of $f(x)=x^{3/2}$ from $x=0$ to $x=4$.

Solution

We find $f'(x)=3x^{1/2}/2$; note that on $[0,4]$, f is differentiable and f' is also continuous. Using Equation (13.7), we find the arc length L as

$$L=\int_0^4\sqrt{1+\left(\frac{3}{2}x^{1/2}\right)^2}\ dx$$

$$= \int_0^4 \sqrt{1 + \frac{9}{4}x}\; dx$$

$$= \frac{2}{3} \cdot \frac{4}{9} \cdot \left(1 + \frac{9}{4}x\right)^{3/2} \Bigg|_0^4$$

$$= \frac{8}{27}\left(10^{3/2} - 1\right) \approx 9.07\text{units.}$$

A graph of f is given in Figure 13.23.

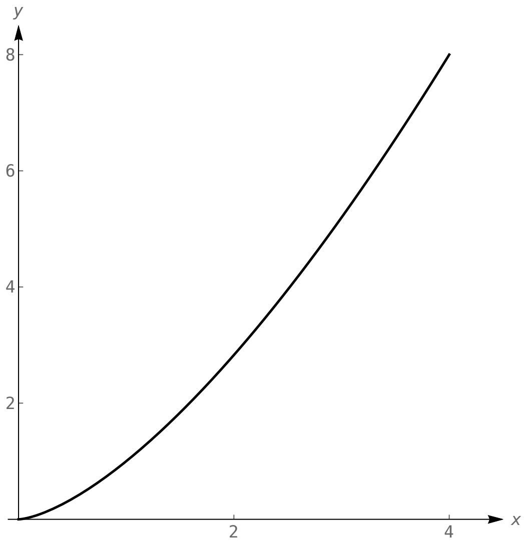


Figure 13.23: A graph of $f(x) = x^{3/2}$ from Example 13.13.

We conclude with one example where it is not possible to find an exact answer.

Example 13.14

Find the length of the sine curve from $x = 0$ to $x = \pi$.

Solution

The setup is straightforward: $f(x) = \sin(x)$ and $f'(x) = \cos(x)$. Thus

$$L = \int_0^\pi \sqrt{1 + \cos^2(x)}\; dx.$$

This integral cannot be evaluated in terms of elementary functions so we have to approximate it with one of the methods studied in Chapter 12. Doing this leads us to $L \approx 3.8202$.

13.4.2 Parametric and polar equations

When we are faced with a curve described by parametric equations, we can convert Equation (13.7) to such a context. Letting $x = f(t)$ and $y = g(t)$, we know that $dy/dx = g'(t)/f'(t)$. It will also be useful to calculate the differential of x:

$$dx = f'(t)dt \qquad \Rightarrow \qquad dt = \frac{1}{f'(t)}\,dx.$$

Starting with the arc length formula given by equation (13.7), consider:

$$\begin{aligned}
L &= \int_a^b \sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx \\
&= \int_a^b \sqrt{1+\frac{g'(t)^2}{f'(t)^2}}\,dx. \\
&= \int_a^b \sqrt{f'(t)^2+g'(t)^2}\,\underbrace{\frac{1}{f'(t)}\,dx}_{=dt} \\
&= \int_{t_1}^{t_2} \sqrt{f'(t)^2+g'(t)^2}\,dt.
\end{aligned} \tag{13.8}$$

Note the new bounds. They are found by solving $a = f(t)$ and $b = f(t)$ for t, and subsequently choosing t_1 such that $t_1 = \min(t_a, t_b)$, where t_a and t_b are the solutions of $a = f(t)$ and $b = f(t)$, respectively. Likewise, t_2 should be chosen such that $t_2 = \max(t_a, t_b)$.

Example 13.15

Find the arc length of the circle parametrized by $x = 3\cos(t)$, $y = 3\sin(t)$ on $[0, 3\pi/2]$.

Solution

By direct application of Equation (13.8), we have

$$L = \int_0^{3\pi/2} \sqrt{(-3\sin(t))^2+(3\cos(t))^2}\,dt.$$

Then apply the Pythagorean theorem:

$$= \int_0^{3\pi/2} 3\,dt = \frac{9\pi}{2}.$$

This should make sense; we know from geometry that the circumference of a circle with radius 3 is 6π; since we are finding the arc length of $3/4$ of a circle, the arc length is $3/4 \cdot 6\pi = 9\pi/2$.

As mentioned above, care should be taken when setting the limits of integration for curves defined by means of parametric equations.

Example 13.16

Find the arc length of the astroid parametrized by $x = c\cos^3(t)$, $y = c\sin^3(t)$ (Figure 13.24).

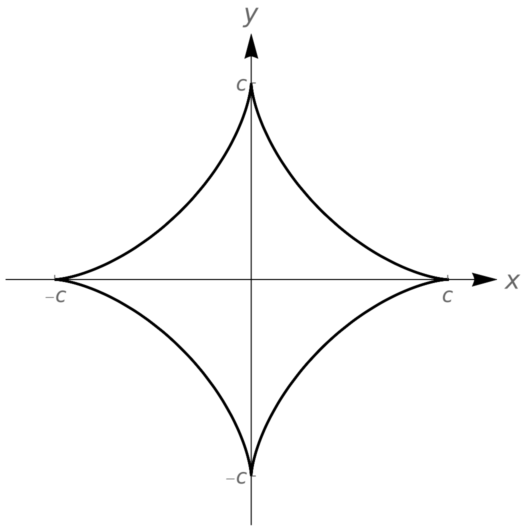


Figure 13.24: A graph of the astroid from Example 13.16.

Solution

By direct application of Equation (13.8), we have

$$\begin{aligned}
L &= \int_{t_1}^{t_2} \sqrt{\left(-3c\cos^2(t)\sin(t)\right)^2 + \left(3c\sin^2(t)\cos(t)\right)^2}\ dt \\
&= \int_{t_1}^{t_2} 3c\sqrt{\sin^2(t)\cos^2(t)\left(\sin^2(t)+\cos^2(t)\right)}\ dt \\
&= \int_{t_1}^{t_2} \left|3c\sin(t)\cos(t)\right|\ dt \\
&= \int_{t_1}^{t_2} \frac{3c}{2}\left|\sin(2t)\right|.
\end{aligned}$$

To find the limits of integration, we note that the arc length we are looking for is four times the length of one arc of the astroid. For what concerns the arc in the first quadrant, we would, when working with cartesian coordinates, vary x from $a = 0$ to $b = c$, where t is $\pi/2$ and 0, respectively. So, when integrating with respect to t the lower limit of integration should become $t_1 = 0$ and the upper limit $t_2 = \pi/2$. Taking into account these details, we get the following

$$L = 4\int_0^{\pi/2} \frac{3c}{2}\sin(2t) = 6c.$$

As we have already considered the arc length of curves defined by rectangular and parametric equations, we now consider it as well in the context of polar equations. Recall that the arc length L of the graph defined by the parametric equations $x = f(t)$, $y = g(t)$ on $[a,b]$ is

$$L = \int_a^b \sqrt{f'(t)^2 + g'(t)^2}\, dt = \int_a^b \sqrt{x'(t)^2 + y'(t)^2}\, dt. \tag{13.9}$$

Now consider the polar function $r = f(\theta)$. We again use the identities $x = f(\theta)\cos(\theta)$ and $y = f(\theta)\sin(\theta)$ to create parametric equations based on the polar function. We compute $x'(\theta)$ and $y'(\theta)$ as done before when computing $\frac{dy}{dx}$, then apply Equation (13.9).

The expression $x'(\theta)^2 + y'(\theta)^2$ can be simplified a great deal; we leave this as an exercise and state that

$$x'(\theta)^2 + y'(\theta)^2 = f'(\theta)^2 + f(\theta)^2.$$

This leads us to the arc length formula.

So, let $r = f(\theta)$ be a polar function with f' continuous on $[\alpha,\beta]$, on which the graph traces itself only once. The arc length L of the graph on $[\alpha,\beta]$ is

$$L = \int_\alpha^\beta \sqrt{f'(\theta)^2 + f(\theta)^2}\, d\theta = \int_\alpha^\beta \sqrt{(r')^2 + r^2}\, d\theta. \tag{13.10}$$

Again, care should be taken when setting the limits of integration for curves defined by means of polar equations.

13.5 Surface area

13.5.1 Rectangular coordinates

We have already seen how a curve $y = f(x)$ on $[a,b]$ can be revolved about an axis to form a solid. Instead of computing its volume, we now consider its surface area.

We begin as we have in the previous sections: we partition the interval $[a,b]$ with n subintervals, where the i^{th} subinterval is $[x_i, x_{i+1}]$. On each subinterval, we can approximate the curve $y = f(x)$ with a straight line that connects $f(x_i)$ and $f(x_{i+1})$ as shown in Figure 13.25(a). Revolving this line segment about the x-axis creates part of a cone (called a frustum of a cone) as shown in Figure 13.25(b). The surface area of a frustum of a cone is

$$2\pi \cdot \text{ length } \cdot \text{average of the two radii } R \text{ and } r.$$

The length is given by L_i. More precisely, we use the material just covered by arc length to state that

$$L \approx \sqrt{1 + f'(c_i)^2}\Delta x_i$$

for some c_i in the i^{th} subinterval. The radii are just the function evaluated at the endpoints of the interval. That is,

$$R_i = f(x_{i+1}) \quad \text{and} \quad r_i = f(x_i).$$

Thus the surface area of this sample frustum of the cone is approximately

$$2\pi \frac{f(x_i) + f(x_{i+1})}{2} \sqrt{1 + f'(c_i)^2}\Delta x_i.$$

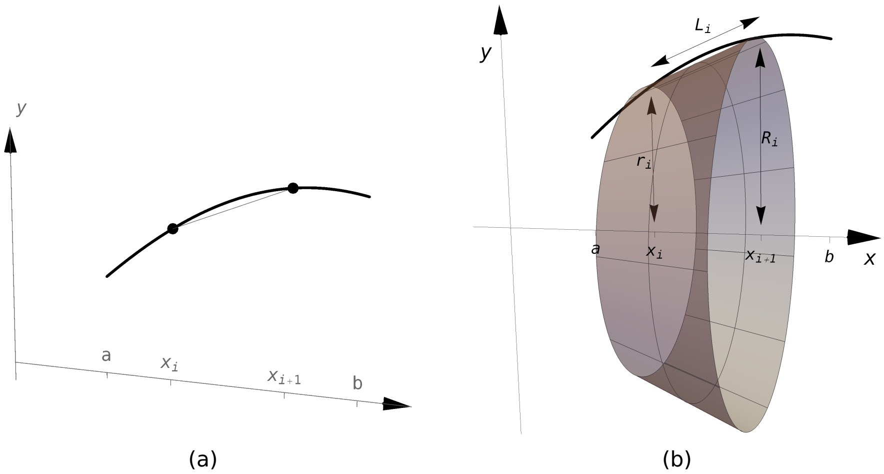


Figure 13.25: Establishing the formula for surface area.

Since f is a continuous function, the intermediate value theorem states there is some d_i in $[x_i, x_{i+1}]$ such that $f(d_i) = \frac{f(x_i)+f(x_{i+1})}{2}$; we can use this to rewrite the above equation as

$$2\pi f(d_i)\sqrt{1+f'(c_i)^2}\Delta x_i.$$

Summing over all the subintervals we get the total surface area to be approximately

$$SA \approx \sum_{i=1}^{n} 2\pi f(d_i)\sqrt{1+f'(c_i)^2}\Delta x_i,$$

which is a Riemann sum. Taking the limit as the subinterval lengths go to zero gives us the exact surface area, given in the following theorem.

Theorem 13.3 (Surface area of a solid of revolution using rectangular coordinates)
Let f be differentiable on $[a,b]$, where f' is also continuous on $[a,b]$.

1. *The surface area of the solid formed by revolving the graph of $y = f(x)$, where $f(x) \geq 0$, about the x-axis is*
$$SA = 2\pi \int_a^b f(x)\sqrt{1+f'(x)^2}\,dx.$$
2. *The surface area of the solid formed by revolving the graph of $y = f(x)$ about the y-axis, where $a, b \geq 0$, is*
$$SA = 2\pi \int_a^b x\sqrt{1+f'(x)^2}\,dx.$$

When revolving $y = f(x)$ about the y-axis, the radii of the resulting frustum are x_i and x_{i+1}; their average value is simply the midpoint of the interval. In the limit, this midpoint is just x. This gives the second part of Theorem 13.3.

We conclude this section with a famous mathematical paradox.

Example 13.17

Consider the solid formed by revolving $y = 1/x$ about the x-axis on $[1, +\infty[$. Find the volume and surface area of this solid. This shape, as graphed in Figure 13.26, is known as "Gabriel's Horn" since it looks like a very long horn that only a supernatural person, such as an angel, could play.

Solution

To compute the volume it is natural to use the disk method. We have:

$$\begin{aligned} V &= \pi \int_1^{+\infty} \frac{1}{x^2}\, dx \\ &= \lim_{b\to+\infty} \pi \int_1^{b} \frac{1}{x^2}\, dx \\ &= \lim_{b\to+\infty} \pi \left(1 - \frac{1}{b}\right) \\ &= \pi \text{ units}^3. \end{aligned}$$

Gabriel's Horn has a finite volume of π cubic units. Since we have already seen that regions with infinite length can have a finite area, this is not too difficult to accept.

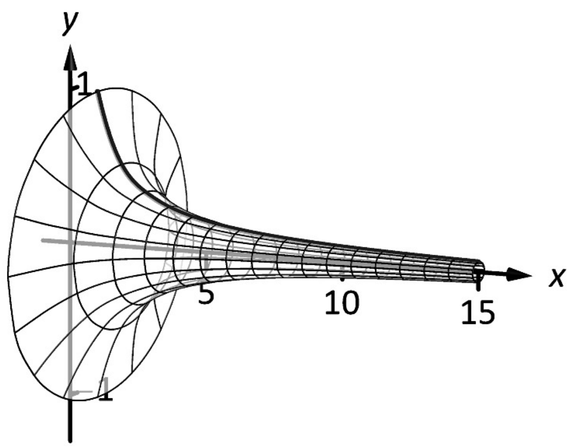


Figure 13.26: A graph of Gabriel's Horn.

We now consider its surface area. The integral is straightforward to setup:

$$SA = 2\pi \int_1^{+\infty} \frac{1}{x}\sqrt{1 + \frac{1}{x^4}}\, dx.$$

Integrating this expression is not trivial, but it can be shown that this improper integral diverges, meaning Gabriel's Horn has infinite surface area. Hence the paradox: we can fill Gabriel's Horn with a finite amount of paint, but since it has infinite surface area, we can never paint it.

13.5.2 Parametric and polar equations

When dealing with a plane curve described by parametric equations, we can adapt the formula found in Theorem 13.3 in a similar way as done to produce the formula for arc length done before.

Theorem 13.4 (Surface area of a solid of revolution using parametric equations)
Consider the graph of the parametric equations $x = f(t)$ and $y = g(t)$, where f' and g' are continuous on an open interval I containing t_1 and t_2 on which the graph does not cross itself.

1. *The surface area of the solid formed by revolving the graph about the x-axis is (where $g(t) \geq 0$ on $[t_1, t_2]$):*
$$SA = 2\pi \int_{t_1}^{t_2} g(t)\sqrt{f'(t)^2 + g'(t)^2}\, dt.$$
2. *The surface area of the solid formed by revolving the graph about the y-axis is (where $f(t) \geq 0$ on $[t_1, t_2]$):*
$$SA = 2\pi \int_{t_1}^{t_2} f(t)\sqrt{f'(t)^2 + g'(t)^2}\, dt.$$

Example 13.18

Consider the teardrop shape formed by the parametric equations $x = t(t^2 - 1)$, $y = t^2 - 1$. Find the surface area if this shape is rotated about the x-axis, as shown in Figure 13.27.

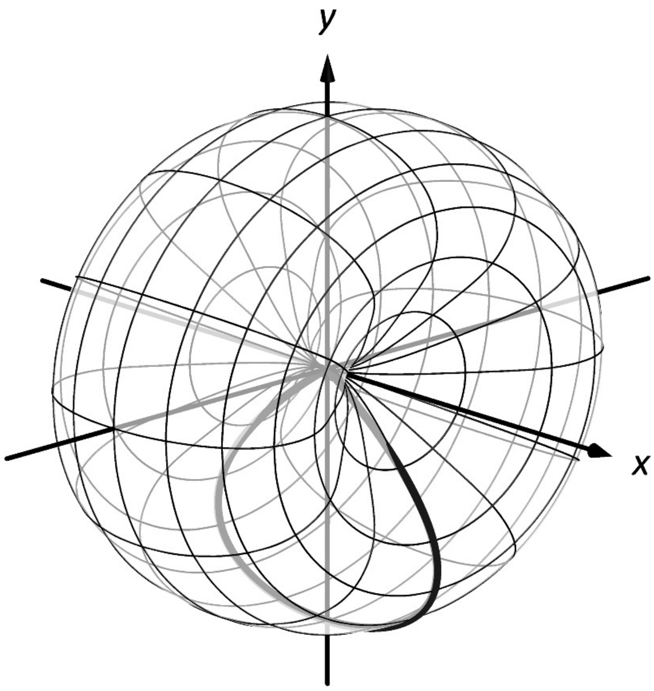


Figure 13.27: Rotating a teardrop shape about the x-axis in Example 13.18.

Solution

The teardrop shape is formed between $t = -1$ and $t = 1$. Using Theorem 13.4, we see we need for $g(t) \geq 0$ on $[-1, 1]$, and this is not the case. To fix this, we simplify replace $g(t)$ with $-g(t)$, which flips the whole graph about the x-axis. The surface area is:

$$\begin{aligned} SA &= 2\pi \int_{-1}^{1} (1-t^2)\sqrt{(3t^2-1)^2+(2t)^2}\, dt \\ &= 2\pi \int_{-1}^{1} (1-t^2)\sqrt{9t^4-2t^2+1}\, dt. \end{aligned}$$

Once again we arrive at an integral that we cannot compute in terms of elementary functions. Using the midpoint rule with $n = 20$, we find the area to be approximately $S = 9.44$.

When dealing with polar equations, we may resort to the following theorem to find surface areas of solids of revolution.

Theorem 13.5 (Surface area of a solid of revolution using polar equations)
Consider the graph of the polar equation $r = f(\theta)$, where f' is continuous on $[\alpha, \beta]$, on which the graph does not cross itself.

1. *The surface area of the solid formed by revolving the graph about the initial ray ($\theta = 0$) is:*

$$SA = 2\pi \int_{\alpha}^{\beta} f(\theta)\sin(\theta)\sqrt{f'(\theta)^2+f(\theta)^2}\, d\theta.$$

2. *The surface area of the solid formed by revolving the graph about the line $\theta = \pi/2$ is:*

$$SA = 2\pi \int_{\alpha}^{\beta} f(\theta)\cos(\theta)\sqrt{f'(\theta)^2+f(\theta)^2}\, d\theta.$$

Example 13.19

Find the surface area formed by revolving one petal of the rose curve $r = \cos(2\theta)$ about its central axis (see Figure 13.28(a)).

Solution

We choose, as implied by the figure, to revolve the portion of the curve that lies on $[0, \pi/4]$ about the initial ray. Using Theorem 13.5 and the fact that $f'(\theta) = -2\sin(2\theta)$, we have

$$\begin{aligned} SA &= 2\pi \int_{0}^{\pi/4} \cos(2\theta)\sin(\theta)\sqrt{\big(-2\sin(2\theta)\big)^2+\big(\cos(2\theta)\big)^2}\, d\theta \\ &\approx 1.36707. \end{aligned}$$

The integral is another that cannot be evaluated in terms of elementary functions. The midpoint's rule, with $n = 4$, approximates the value at 1.37.

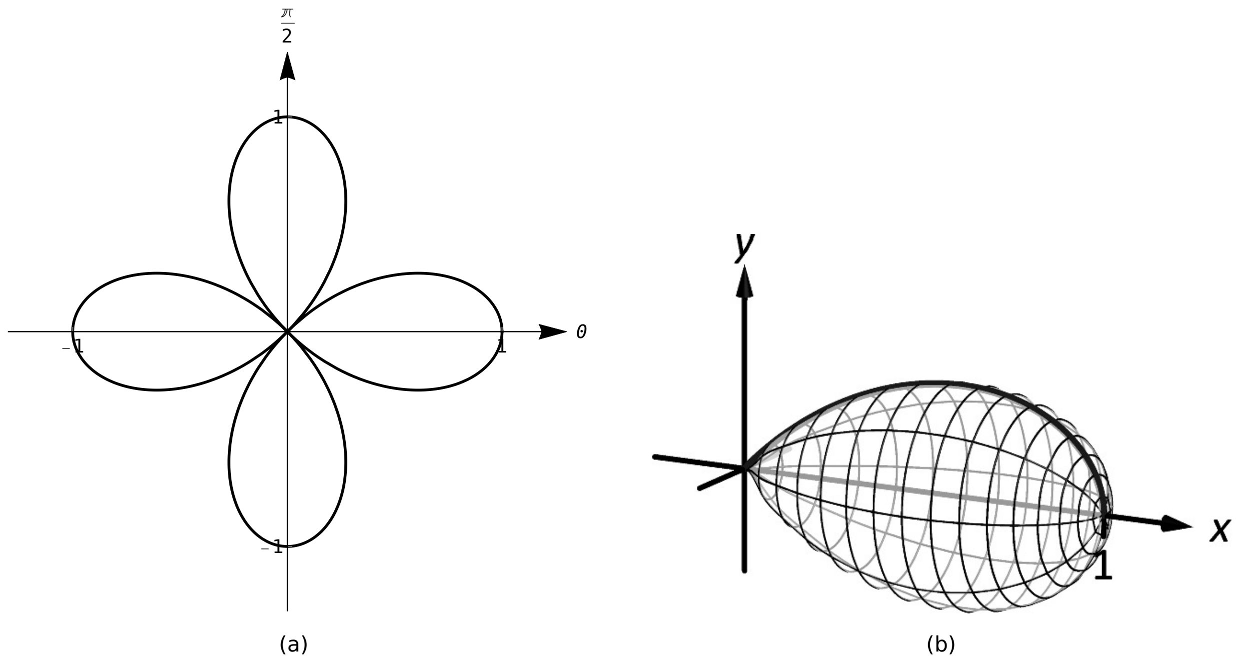


Figure 13.28: Finding the surface area of a rose–curve petal that is revolved about its central axis.

13.6 Exercises

Area between curves

Assignment 13.1 — Sketch the regions below and find their area.

(a) the area bounded by $y^2 = 4x$ and $y = 2x - 4$

(b) the area bounded by $x = 4 - y^2$ and the y-axis

(c) the smallest part within $x^2 + y^2 = 25$, cut off by $x = 3$

(d) the region enclosed between $y = 4x - x^2$, $y = 4 - x$ and the y-axis

(e) the region enclosed between $y = 6x - x^2$ and $y = x^2 - 2x$

(f) the region enclosed between $x^2 + y^2 = 12$ and $y^2 = x$

(g) the region enclosed between $y = 0$ and $y = \cos^2(x)$ for $x \in [0, 2\pi]$

(h) the region enclosed between $y = \ln(2x)$ and $y = \ln(x)$ for $x \in [1, e]$

(i) the region enclosed between $y = \sin(x)$, $x + y + \dfrac{\pi}{2} + 1 = 0$ and the x-axis

Assignment 13.2 — Find the areas of the regions below.

(a) a sector of a circle of radius R about an angle α

(b) the region enclosed between the parabola $y^2 = 4a(a - x)$, with $a > 0$, and the y-axis

(c) the region enclosed between $y = \dfrac{x^2}{2}$, $y = \sqrt{2x}$ and $y = \sqrt{2}x$

(d) the area inside the loop of the curve $\begin{cases} x = t^2 \\ y = t - \dfrac{t^3}{9} \end{cases}$

(e) The region enclosed by the cardioid $r(\theta) = a(1 + \cos(\theta))$

(f) The region enclosed by one loop of the curve $r(\theta) = 4\cos(2\theta)$

(g) the area region enclosed by the astroid

$$\begin{cases} x = c\cos^3(t) \\ y = c\sin^3(t). \end{cases}$$

(h) The area region enclosed by the curve $\begin{cases} x = 3 + \cos(\theta) \\ y = 4\sin(\theta) \end{cases}$

(i) the area region enclosed by the curve $\begin{cases} x = 3\sin(2t) \\ y = 2\cos(t) \end{cases}$

(j) The region outside the circle $r = 2$ and enclosed by the cardioid $r = 2\big(\cos(\theta)+1\big)$

(k) The region enclosed by the circle $r = 2$ and within the cardioid $r = 2\big(\cos(\theta)-1\big)$

(l) the region outside of the curve $r = a$ and enclosed by $r(\theta) = 2a\sin(3\theta)$

(m) the common region enclosed by the curves $\sqrt{3}r(\theta) = 1+\sin(\theta)$ en $r(\theta) = \cos(\theta)$

(n) the region enclosed by the first loop of the logarithmic spiral $r(\theta) = 3e^{2\theta}$.

Volume by cross-sectional area and The shell method

Assignment 13.3 — Using the most efficient method to find the volume of the body of revolution obtained by rotating the given region about the given axis.

(a) the region in the first quadrant bounded by $y^2 = 8x$ en $x = 2$ about the x-axis

(b) the region bounded by $y^2 = 8x$ and $x = 2$ about the y-axis

(c) the region bounded by $y = x^2$, $y = \sqrt{x}$, $x = 0$ and $x = 1$ about the x-axis

(d) the region bounded by $y^2 = 8x$ and $x = 2$ about the line $x = 2$

(e) the region bounded by $y = x$ and $x = 4y - y^2$ about (a) the x-axis and (b) the y-axis

(f) the region inside $y = 4x - x^2$, cut off by the x-axis about $y = 6$

Assignment 13.4 — Find the volume of the body of revolution obtained by rotating the given region about the given axis.

(a) $y = 2x$ about $x = 3$ for $y \in [0, 6]$

(b) the region bounded by $x^2 = 6 - y$ and the x-axis about the $y$axis

(c) the region enclosed by $y^2 = x^2\sqrt{1-x^2}$ about the x-axis

(d) the region enclosed by $y^2 = x^2(1-x^2)$ about (a) the x-axis and (b) the y-axis

(e) the area above $x - 2y + 5 = 0$ and within $x^2 + y^2 = 25$ about the x-axis

(f) the area between the first loop of the cycloid $\begin{cases} x = \theta - \sin(\theta) \\ y = 1 - \cos(\theta) \end{cases}$ and the x-as about (a) the y-axis and (b) the line $y = 2$

(g) the region enclosed by $r(\theta) = 4\cos^2(\theta)$ about the polar axis ($\theta = 0$)

(h) the region for which $0 \le y \le 1 - x^2$ about the line $y = 1$

(i) a circular disk about one of the tangents

Arc length

Assignment 13.5 — Find the arc length of (the part of) the given curve.

(a) an arc of a circle of radius R about an angle α

(b) the curve $y = \ln\left(1-x^2\right)$ from $x = -\dfrac{1}{2}$ to $x = \dfrac{1}{2}$

(c) the astroid $x^{2/3} + y^{2/3} = a^{2/3}$

(d) the cardioid $r(\theta) = a(1+\cos(\theta))$ with $0 \leq \theta \leq 2\pi$

(e) the part of $x = \ln\left(\dfrac{1}{\cos(y)}\right)$ between $y = 0$ and $y = \dfrac{\pi}{3}$

(f) the curve $r(\theta) = \dfrac{1}{\theta}$ with $\theta \in \left[\dfrac{1}{2}, 2\right]$

(g) the curve $\begin{cases} x = 2\cos(t) - \cos(2t) \\ y = 2\sin(t) - \sin(2t) \end{cases}$

(h) the part of $y^2 = x^3$ between the origin and the point with abscissa 4

(i) the closed part of $9y^2 = x(x-3)^2$

Surface area

Assignment 13.6 — Find the surface area of the body of revolution obtained by rotating the given region about the given axis.

(a) the area under $y = \sin(x)$ about the x-axis for $x \in [0, \pi]$

(b) the ellipsoid $\dfrac{x^2}{16} + \dfrac{y^2}{4} = 1$ about the x-axis

(c) the line $y = 2x$ about the line $x = 3$ for $x \in \mathbb{R}^+$.

(d) the region between $x = y^3$, $y = 0$ and $y = 1$ about (a) the y-axis and (b) the line $x = 1$ (do not evaluate the integral)

(e) one loop of $8y^2 = x^2 - x^4$ about the x-axis

(f) the region enclosed by the astroid $\begin{cases} x = \cos^3(t) \\ y = \sin^3(t) \end{cases}$ and the x-axis about (a) the x-axis and (b) the line $y = -1$

(g) $r^2 = a^2\cos(2\theta)$ about the polar axis

(h) the first loop of the cycloid $\begin{cases} x = a(\theta - \sin(\theta)) \\ y = a(1-\cos(\theta)) \end{cases}$ about (a) the x-axis and (b) the line $x = a\pi$

Review exercises

Assignment 13.7 — A cherry floats in a cocktail glass. The glass has the shape of a sphere with diameter 8 cm. The idealized cherry is spherical and has a diameter of 2 cm. The glass is filled to $3/2$ cm from its border with Kir and the top of the cherry is located 1 cm from the rim of the glass.

(a) How much Kir does the glass contain? Tip: the glass is created by rotating $x = f(y)$ about the y-axis. You can model the cherry by rotating $x = g(y)$ about the y-axis.

(b) Find the area of the part of the cherry that extends above the liquid surface.

Mathematics is the queen of science

— Carl Friedrich Gauss —

14

Vector–valued functions

In Chapter 6, we learned about vectors and we were introduced to the power of vectors within mathematics. In this chapter, we will build on this foundation to define functions whose input is a real number and whose output is a vector.

14.1 Vector–valued functions

14.1.1 Definition

We are very familiar with real–valued functions, that is, functions whose output is a real number. This section introduces vector–valued functions – functions whose output is a vector.

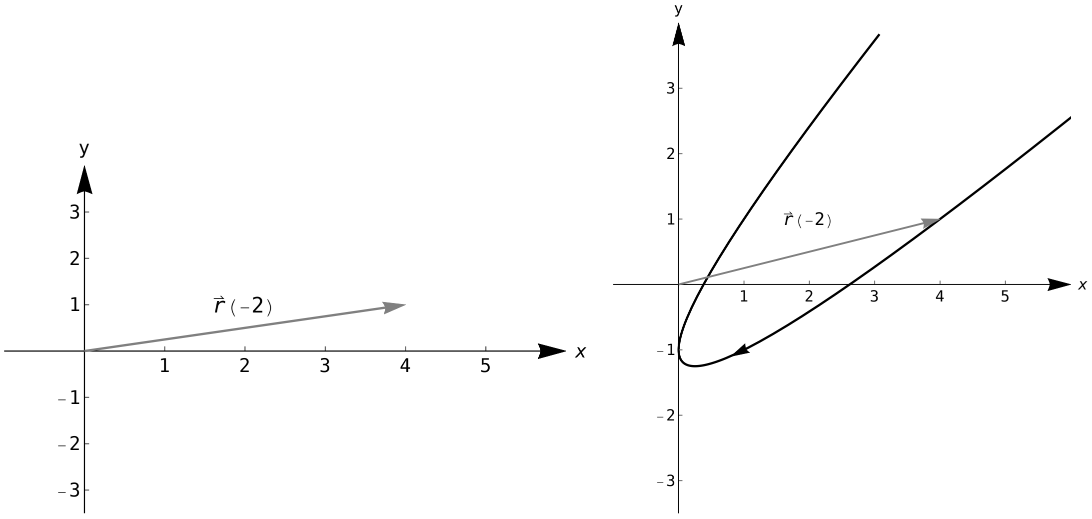
Definitie 14.1 (Vector–valued function)
A **vector–valued function** (*vectorfunctie*) is a function of the form

$$\vec{\boldsymbol{r}}(t) = \big(f(t), g(t)\big) \quad \text{or} \quad \vec{\boldsymbol{r}}(t) = \big(f(t), g(t), h(t)\big),$$

where f, g and h are real–valued functions, and are called the **component functions**.

The domain of $\vec{\boldsymbol{r}}$ is the set of all values of t for which $\vec{\boldsymbol{r}}(t)$ is defined. The range of $\vec{\boldsymbol{r}}$ is the set of all possible output vectors $\vec{\boldsymbol{r}}(t)$.

Evaluating a vector–valued function at a specific value of t is straightforward; simply evaluate each component function at that value of t. For instance, if $\vec{\boldsymbol{r}}(t) = \left(t^2, t^2 + t - 1\right)$, then $\vec{\boldsymbol{r}}(-2) = (4, 1)$. We can sketch this vector, as is done in Figure 14.1(a). Plotting lots of vectors is cumbersome, though, so generally we do not sketch the whole vector but just the terminal point. The graph of a vector–valued function is the set of all terminal points of $\vec{r}(t)$, where the initial point of each vector is always the origin. In Figure 14.1(b) we sketch the graph of $\vec{r}$; we can indicate individual points on the graph with their respective vector, as shown.



(a) (b)

Figure 14.1: Sketching the graph of a vector–valued function.

Vector–valued functions are closely related to parametric equations of graphs. While in both methods we plot points $\big(x(t), y(t)\big)$ or $\big(x(t), y(t), z(t)\big)$ to produce a graph, in the context of vector–valued functions each such point represents a vector. The implications of this will be more fully realized in the next section as we apply calculus ideas to these functions.

Example 14.1

Graph $\vec{r}(t) = \big(\cos(t), \sin(t), t\big)$ for $0 \leq t \leq 4\pi$.

Solution

We can plot points, but careful consideration of this function is very revealing. Momentarily ignoring the third component, we see that the x- and y- components trace out a circle of radius 1 centred at the origin. Noticing that the z component is t, we see that as the graph winds around the z-axis, it is also increasing at a constant rate in the positive z direction, forming a spiral. This is graphed in Figure 14.2. In the graph, $\vec{r}(7\pi/4) \approx (0.707, -0.707, 5.498)$ is highlighted to help us understand the graph.

14.1.2 Algebra of vector–valued functions

Let $\vec{r}_1(t) = \big(f_1(t), g_1(t)\big)$ and $\vec{r}_2(t) = \big(f_2(t), g_2(t)\big)$ be vector–valued functions in $\mathbb{R}^2$ and let c be a scalar. Then:

1. $\vec{r}_1(t) \pm \vec{r}_2(t) = \big(f_1(t) \pm f_2(t), g_1(t) \pm g_2(t)\big)$,
2. $c\vec{r}_1(t) = \big(cf_1(t), cg_1(t)\big)$.

A similar definition holds for vector–valued functions in $\mathbb{R}^3$.

This shows that we add, subtract and scale vector-valued functions component–wise. Combining vector–valued functions in this way can be very useful.

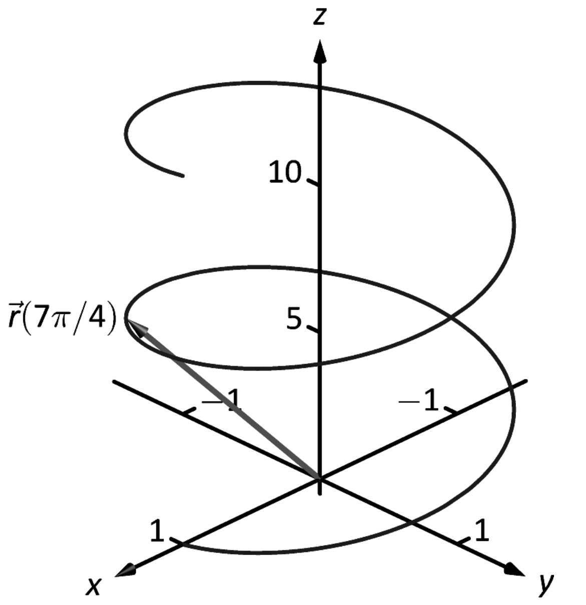


Figure 14.2: The graph of $\vec{r}(t)$ in Example 14.1.

Example 14.2

Let $\vec{r}_1(t) = (0.2t, 0.3t)$, $\vec{r}_2(t) = \big(\cos(t), \sin(t)\big)$ and $\vec{r}(t) = \vec{r}_1(t) + \vec{r}_2(t)$. Graph $\vec{r}_1(t)$, $\vec{r}_2(t)$, $\vec{r}(t)$ and $5\vec{r}(t)$ for $-10 \le t \le 10$.

Solution

We can graph $\vec{r}_1$ and $\vec{r}_2$ easily by plotting points. Let us think about each for a moment to better understand how vector–valued functions work.

We can rewrite $\vec{r}_1(t) = (0.2t, 0.3t)$ as $\vec{r}_1(t) = t\,(0.2, 0.3)$. That is, the function $\vec{r}_1$ scales the vector $(0.2, 0.3)$ by t. This scaling of a vector produces a line in the direction of $(0.2, 0.3)$. We are familiar with $\vec{r}_2(t) = \big(\cos(t), \sin(t)\big)$; it traces out a circle, centered at the origin, of radius 1. Figure 14.3(a) graphs $\vec{r}_1(t)$ and $\vec{r}_2(t)$.

Adding $\vec{r}_1(t)$ to $\vec{r}_2(t)$ produces $\vec{r}(t) = \big(\cos(t) + 0.2t, \sin(t) + 0.3t\big)$, graphed in Figure 14.3(b). The linear movement of the line combines with the circle to create loops that move in the direction of $(0.2, 0.3)$.

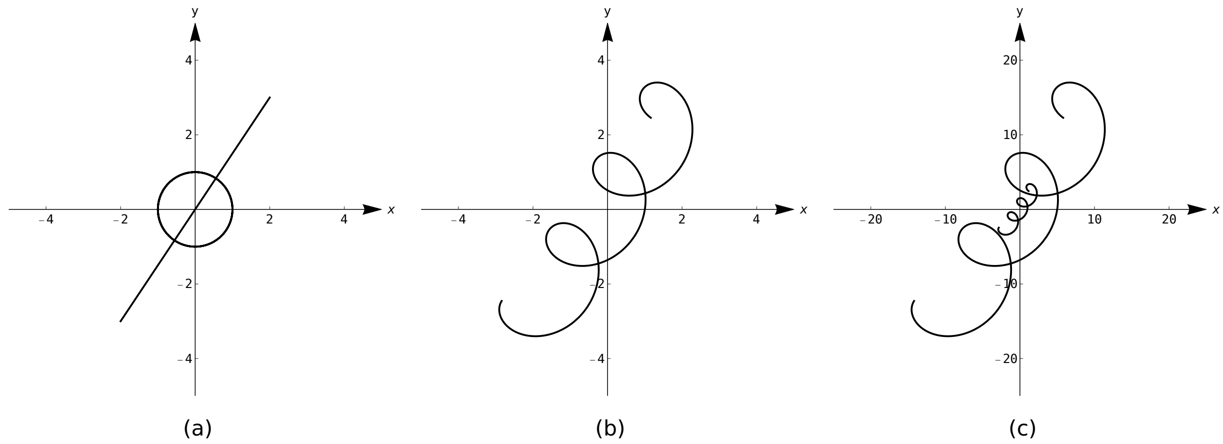


Figure 14.3: Graphing the functions in Example 14.2.

Multiplying $\vec{r}(t)$ by 5 scales the function by 5, producing $5\vec{r}(t) = \big(5\cos(t)+1, 5\sin(t)+1.5\big)$, which is graphed in Figure 14.3(c) along with $\vec{r}(t)$. The new function is 5 times bigger than $\vec{r}(t)$. Note how the graph of $5\vec{r}(t)$ in (c) looks identical to the graph of $\vec{r}(t)$ in (b). This is due to the fact that the x- and y- bounds of the plot in (c) are exactly 5 times larger than the bounds in (b).

A vector–valued function $\vec{r}(t)$ is often used to describe the position of a moving object at time t. At $t = t_0$, the object is at $\vec{r}(t_0)$; at $t = t_1$, the object is at $\vec{r}(t_1)$. Knowing the locations $\vec{r}(t_0)$ and $\vec{r}(t_1)$ gives no indication of the path taken between them, but often we only care about the difference of the locations, $\vec{r}(t_1) - \vec{r}(t_0)$, the **displacement** (*verplaatsing*).

Definitie 14.2 (Displacement)
Let $\vec{r}(t)$ be a vector–valued function and let $t_0 < t_1$ be values in the domain. The **displacement** $\vec{d}$ of $\vec{r}$, from $t = t_0$ to $t = t_1$, is

$$\vec{d} = \vec{r}(t_1) - \vec{r}(t_0).$$

When the displacement vector is drawn with initial point at $\vec{r}(t_0)$, its terminal point is $\vec{r}(t_1)$. We think of it as the vector which points from a starting position to an ending position.

Example 14.3

Let $\vec{r}(t) = \Big(\cos\big(\frac{\pi}{2}t\big), \sin\big(\frac{\pi}{2}t\big)\Big)$. Graph $\vec{r}(t)$ on $-1 \le t \le 1$, and find the displacement of $\vec{r}(t)$ on this interval.

Solution

The function $\vec{r}(t)$ traces out the unit circle, though at a different rate than the usual $\big(\cos(t), \sin(t)\big)$ parametrization. At $t_0 = -1$, we have $\vec{r}(t_0) = (0, -1)$; at $t_1 = 1$, we have $\vec{r}(t_1) = (0, 1)$. The displacement of $\vec{r}(t)$ on $[-1, 1]$ is thus $\vec{d} = (0, 1) - (0, -1) = (0, 2)$. A graph of $\vec{r}(t)$ on $[-1, 1]$ is given in Figure 14.4, along with the displacement vector $\vec{d}$ on this interval.

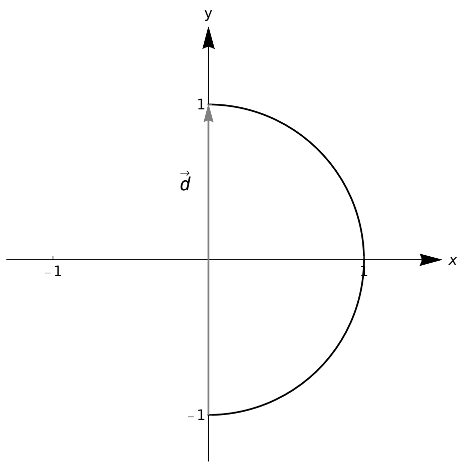


Figure 14.4: Graphing the displacement of a position function in Example 14.3.

Measuring displacement makes us contemplate related, yet very different, concepts. Considering the semi–circular path the object in Example 14.3 took, we can quickly verify that the object ended up a distance of 2 units from its initial location. That is, we can compute $\left\|\vec{d}\right\| = 2$. However, measuring distance from the starting point is different from measuring distance travelled. Being a semi–circle, we can measure the distance traveled by this object as $\pi \approx 3.14$ units. Knowing distance from the starting point allows us to compute average rate of change.

Definitie 14.3 (Average rate of change)
Let $\vec{r}(t)$ be a vector–valued function, where each of its component functions is continuous on its domain, and let $t_0 < t_1$. The **average rate of change of $\vec{r}(t)$** on $[t_0, t_1]$ is

$$\text{average rate of change} = \frac{\vec{r}(t_1) - \vec{r}(t_0)}{t_1 - t_0}.$$

Example 14.4

Let $\vec{r}(t) = \left(\cos\left(\frac{\pi}{2}t\right), \sin\left(\frac{\pi}{2}t\right)\right)$ as in Example 14.3. Find the average rate of change of $\vec{r}(t)$ on $[-1,1]$ and on $[-1,5]$.

Solution

We computed in Example 14.3 that the displacement of $\vec{r}(t)$ on $[-1,1]$ was $\vec{d} = (0,2)$. Thus the average rate of change of $\vec{r}(t)$ on $[-1,1]$ is:

$$\frac{\vec{r}(1) - \vec{r}(-1)}{1-(-1)} = \frac{(0,2)}{2} = (0,1).$$

We interpret this as follows: the object followed a semi–circular path, meaning it moved towards the right then moved back to the left, while climbing slowly, then quickly, then slowly again. On average, however, it progressed straight up at a constant rate of $(0,1)$ per unit of time.

We considered average rates of change in Sections 8.1 and 9.1 as we studied limits and derivatives. The same is true here; in the following section we apply calculus concepts to vector–valued functions as we find limits, derivatives, and integrals. Understanding the average rate of change will give us an understanding of the derivative; displacement gives us one application of integration.

14.2 Calculus and vector–valued functions

The previous section introduced us to a new mathematical object, the vector–valued function. We now apply calculus concepts to these functions. We start with the limit, then work our way through derivatives to integrals.

14.2.1 Limits of vector-valued functions

The initial definition of the limit of a vector–valued function is a bit intimidating, as was the definition of the limit in Definition 8.1. The theorem following the definition shows that in practice, taking limits of vector–valued functions is no more difficult than taking limits of real–valued functions. Of course, we can define one-sided limits in a manner very similar to Definition 14.4.

Definitie 14.4 (Limits of vector–valued functions)
Let I be an open interval containing c, and let $\vec{r}(t)$ be a vector–valued function defined on I, except possibly at c. The **limit of $\vec{r}(t)$** (*limiet*), as t approaches c, is $\vec{L}$, expressed as

$$\lim_{t\to c} \vec{r}(t) = \vec{L},$$

means that given any $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $t \neq c$, if $|t-c| < \delta$, we have $\left\|\vec{\boldsymbol{r}}(t)-\vec{\boldsymbol{L}}\right\| < \varepsilon$.

Note how the measurement of distance between real numbers is the absolute value of their difference; the measure of distance between vectors is the vector norm, or magnitude, of their difference.

The following theorem affirms that we can compute limits of vector–valued functions component–wise.

Theorem 14.1 (Limits of vector-valued functions)

1. *Let $\vec{\boldsymbol{r}}(t) = \big(f(t), g(t)\big)$ be a vector–valued function in $\mathbb{R}^2$ defined on an open interval I containing c, except possibly at c. Then*

$$\lim_{t\to c}\vec{\boldsymbol{r}}(t) = \left(\lim_{t\to c} f(t)\,,\ \lim_{t\to c} g(t)\right).$$

2. *Let $\vec{\boldsymbol{r}}(t) = \big(f(t), g(t), h(t)\big)$ be a vector–valued function in $\mathbb{R}^3$ defined on an open interval I containing c, except possibly at c. Then*

$$\lim_{t\to c}\vec{\boldsymbol{r}}(t) = \left(\lim_{t\to c} f(t)\,,\ \lim_{t\to c} g(t)\,,\ \lim_{t\to c} h(t)\right).$$

So, for instance, if

$$\vec{\boldsymbol{r}}(t) = \left(\frac{\sin(t)}{t},\ t^2-3t+3,\ \cos(t)\right). \tag{14.1}$$

Then, $\lim_{t\to 0}\vec{\boldsymbol{r}}(t)$ is given by:

$$\lim_{t\to 0}\vec{\boldsymbol{r}}(t) = \left(\lim_{t\to 0}\frac{\sin(t)}{t}\,,\ \lim_{t\to 0}(t^2-3t+3)\,,\ \lim_{t\to 0}\cos(t)\right) = (1,3,1)\,.$$

14.2.2 Continuity

Having defined limits of vector–valued functions, it makes sense to explore the continuity of such functions.

Definitie 14.5 (Continuity of vector–valued functions)
Let $\vec{\boldsymbol{r}}(t)$ be a vector–valued function defined on an open interval I containing c.

1. $\vec{\boldsymbol{r}}(t)$ is **continuous** at c if $\lim_{t\to c}\vec{\boldsymbol{r}}(t) = \vec{\boldsymbol{r}}(c)$.
2. If $\vec{\boldsymbol{r}}(t)$ is continuous at all c in I, then $\vec{\boldsymbol{r}}(t)$ is continuous on I.

Using one-sided limits, we can of course also define continuity on closed intervals as done before. Moreover, we again have a theorem that lets us evaluate continuity component–wise.

Theorem 14.2 (Continuity of vector–valued functions)
Let $\vec{\boldsymbol{r}}(t)$ be a vector–valued function defined on an open interval I containing c. Then $\vec{\boldsymbol{r}}(t)$ is continuous at c if, and only if, each of its component functions is continuous at c.

In the case of the vector–valued function defined by Equation (14.1), for instance, $\vec{\boldsymbol{r}}(t)$ is continuous at $t=1$ because each of the component functions is continuous at $t=1$. On the other hand, at $t=0$, the second and third components of $\vec{\boldsymbol{r}}(t)$ are defined, but the first component, $(\sin(t))/t$, is not. Since the first component is not even defined at $t=0$, $\vec{r}(t)$ is not defined at $t=0$, and hence it is not continuous at $t=0$.

14.2.3 Derivatives

14.2.3.1 Definition and properties

Consider a vector–valued function $\vec{r}$ defined on an open interval I containing t_0 and t_1. We can compute the displacement of $\vec{r}$ on $[t_0,t_1]$, as shown in Figure 14.5(a). Recall that dividing the displacement vector by t_1-t_0 gives the average rate of change on $[t_0,t_1]$, as shown in Figure 14.5(b).

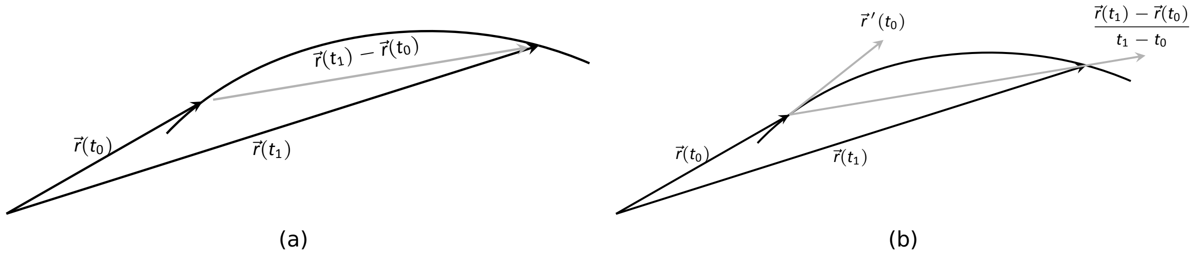


Figure 14.5: Illustrating displacement, leading to an understanding of the derivative of vector–valued functions.

The **derivative** (*afgeleide*) of a vector–valued function is a measure of the instantaneous rate of change, measured by taking the limit as the length of $[t_0,t_1]$ goes to 0. Instead of thinking of an interval as $[t_0,t_1]$, we think of it as $[c,c+h]$ for some value of h (hence the interval has length h). The average rate of change is

$$\frac{\vec{r}(c+h)-\vec{r}(c)}{h}$$

for any value of $h\neq 0$. We take the limit as $h\to 0$ to measure the instantaneous rate of change; this is the derivative of $\vec{r}$.

Definitie 14.6 (Derivative of a vector–valued function)
Let $\vec{r}(t)$ be continuous on an open interval I containing c.

1. The **derivative of $\vec{r}$ at** $t=c$ is

$$\vec{r}\,'(c)=\lim_{h\to 0}\frac{\vec{r}(c+h)-\vec{r}(c)}{h}.$$

2. The **derivative of $\vec{r}$** is

$$\vec{r}\,'(t)=\lim_{h\to 0}\frac{\vec{r}(t+h)-\vec{r}(t)}{h}.$$

Alternate notations for the derivative of $\vec{r}$ include:

$$\vec{r}\,'(t)=\frac{d}{dt}\big(\vec{r}(t)\big)=\frac{d\vec{r}}{dt}.$$

If a vector–valued function has a derivative for all c in an open interval I, we say that $\vec{r}(t)$ is **differentiable** (*afleidbaar*) on I. Again, of course, using one-sided limits, we can define differentiability on closed intervals. We might view Definition 14.6 as intimidating, but recall that we can evaluate limits component–wise. The following theorem verifies that this means we can compute derivatives component–wise as well, making the task not too difficult.

Theorem 14.3 (Derivative of a vector–valued function)

1. *Let* $\vec{r}(t) = \big(f(t), g(t)\big)$*. Then*
$$\vec{r}\,'(t) = \big(f\,'(t), g'(t)\big).$$
2. *Let* $\vec{r}(t) = \big(f(t), g(t), h(t)\big)$*. Then*
$$\vec{r}\,'(t) = \big(f\,'(t), g'(t), h'(t)\big).$$

Example 14.5

Let $\vec{r}(t) = \big(t^2, t\big)$.

1. Sketch $\vec{r}(t)$ and $\vec{r}\,'(t)$ on the same axes.
2. Compute $\vec{r}\,'(1)$ and sketch this vector with its initial point at the origin and at $\vec{r}(1)$.

Solution

1. Theorem 14.3 allows us to compute derivatives component–wise, so
$$\vec{r}\,'(t) = (2t, 1).$$
$\vec{r}(t)$ and $\vec{r}\,'(t)$ are graphed together in Figure 14.6(a). Note how plotting the two of these together, in this way, is not very illuminating. When dealing with real–valued functions, plotting $f(x)$ with $f\,'(x)$ gave us useful information as we were able to compare f and $f\,'$ at the same x-values. When dealing with vector–valued functions, it is hard to tell which points on the graph of $\vec{r}\,'$ correspond to which points on the graph of $\vec{r}$.
2. We easily compute $\vec{r}\,'(1) = (2, 1)$, which is drawn in Figure 14.6(b) with its initial point at the origin, as well as at $\vec{r}(1) = (1, 1)$. These are sketched in Figure 14.6(b).

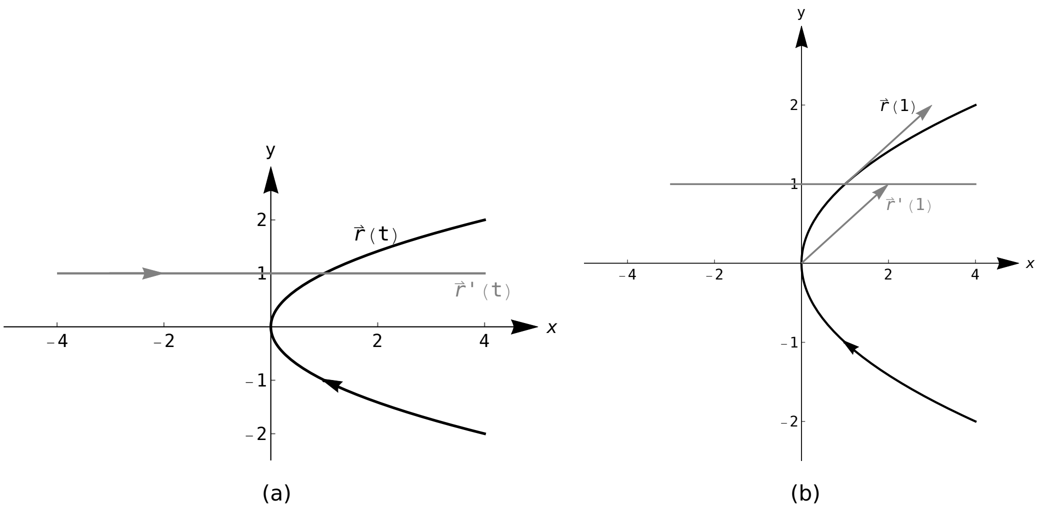


Figure 14.6: Graphing the derivative of a vector–valued function in Example 14.5.

Having established derivatives of vector–valued functions, we now explore the relationships between the derivative and other vector operations. The following properties show how the derivative interacts with vector addition and the various vector products. For that purpose, let $\vec{r}$ and $\vec{s}$ be differentiable vector–valued functions, let f be a differentiable real–valued function, and let c be a real number. Then the following properties hold.

- $\frac{d}{dt}\Big(\vec{r}(t) \pm \vec{s}(t)\Big) = \vec{r}\,'(t) \pm \vec{s}\,'(t)$
- $\frac{d}{dt}\Big(c\vec{r}(t)\Big) = c\vec{r}\,'(t)$
- $\frac{d}{dt}\Big(f(t)\vec{r}(t)\Big) = f\,'(t)\vec{r}(t) + f(t)\vec{r}\,'(t)$
- $\frac{d}{dt}\Big(\vec{r}(t)\cdot\vec{s}(t)\Big) = \vec{r}\,'(t)\cdot\vec{s}(t) + \vec{r}(t)\cdot\vec{s}\,'(t)$
- $\frac{d}{dt}\Big(\vec{r}(t)\times\vec{s}(t)\Big) = \vec{r}\,'(t)\times\vec{s}(t) + \vec{r}(t)\times\vec{s}\,'(t)$
- $\frac{d}{dt}\Big(\vec{r}\big(f(t)\big)\Big) = \vec{r}\,'\big(f(t)\big)f\,'(t)$

Example 14.6

Let $\vec{r}(t) = \big(t, t^2 - 1\big)$ and let $\vec{u}(t)$ be the unit vector that points in the direction of $\vec{r}(t)$.

1. Graph $\vec{r}(t)$ and $\vec{u}(t)$ on the same axes, on $[-2, 2]$.
2. Find $\vec{u}\,'(t)$ and sketch $\vec{u}\,'(-2)$, $\vec{u}\,'(-1)$ and $\vec{u}\,'(0)$. Sketch each with initial point the corresponding point on the graph of $\vec{u}$.

Solution

1. To form the unit vector that points in the direction of $\vec{r}$, we need to divide $\vec{r}(t)$ by its magnitude.
$$\big\|\vec{r}(t)\big\| = \sqrt{t^2 + (t^2-1)^2} \quad \Rightarrow \quad \vec{u}(t) = \frac{1}{\sqrt{t^2+(t^2-1)^2}}\vec{r}(t)$$
$\vec{r}(t)$ and $\vec{u}(t)$ are graphed in Figure 14.7(a). Note how the graph of $\vec{u}(t)$ forms part of a circle; this must be the case, as the length of $\vec{u}(t)$ is 1 for all t.

2. To compute $\vec{u}\,'(t)$, we rely on the above properties and write
$$\vec{u}(t) = f(t)\vec{r}(t), \quad \text{where} \quad f(t) = \frac{1}{\sqrt{t^2+(t^2-1)^2}} = \big(t^2 + (t^2-1)^2\big)^{-1/2}.$$
We find $f\,'(t)$ using the chain rule:
$$\begin{aligned} f\,'(t) &= -\frac{1}{2}\big(t^2+(t^2-1)^2\big)^{-3/2}\big(2t + 2(t^2-1)(2t)\big) \\ &= -\frac{2t(2t^2-1)}{2\big(\sqrt{t^2+(t^2-1)^2}\,\big)^3}. \end{aligned}$$
We now find $\vec{u}\,'(t)$:
$$\begin{aligned} \vec{u}\,'(t) &= f\,'(t)\vec{u}(t) + f(t)\vec{u}\,'(t) \\ &= -\frac{2t(2t^2-1)}{2\big(\sqrt{t^2+(t^2-1)^2}\,\big)^3}\Big(t, t^2-1\Big) + \frac{1}{\sqrt{t^2+(t^2-1)^2}}\,(1, 2t)\,. \end{aligned}$$
We can use this formula to compute $\vec{u}\,'(-2)$, $\vec{u}\,'(-1)$ and $\vec{u}\,'(0)$:
$$\vec{u}\,'(-2) = \left(-\frac{15}{13\sqrt{13}}, -\frac{10}{13\sqrt{13}}\right) \approx (-0.320, -0.213)\,,$$

$$\vec{\mathbf{u}}'(-1) = (0, -2),$$
$$\vec{\mathbf{u}}'(0) = (1, 0).$$

Each of these is sketched in Figure 14.7(b). Note how the length of the vector gives an indication of how quickly the circle is being traced at that point. When $t = -2$, the circle is being drawn relatively slow; when $t = -1$, the circle is being traced much more quickly.

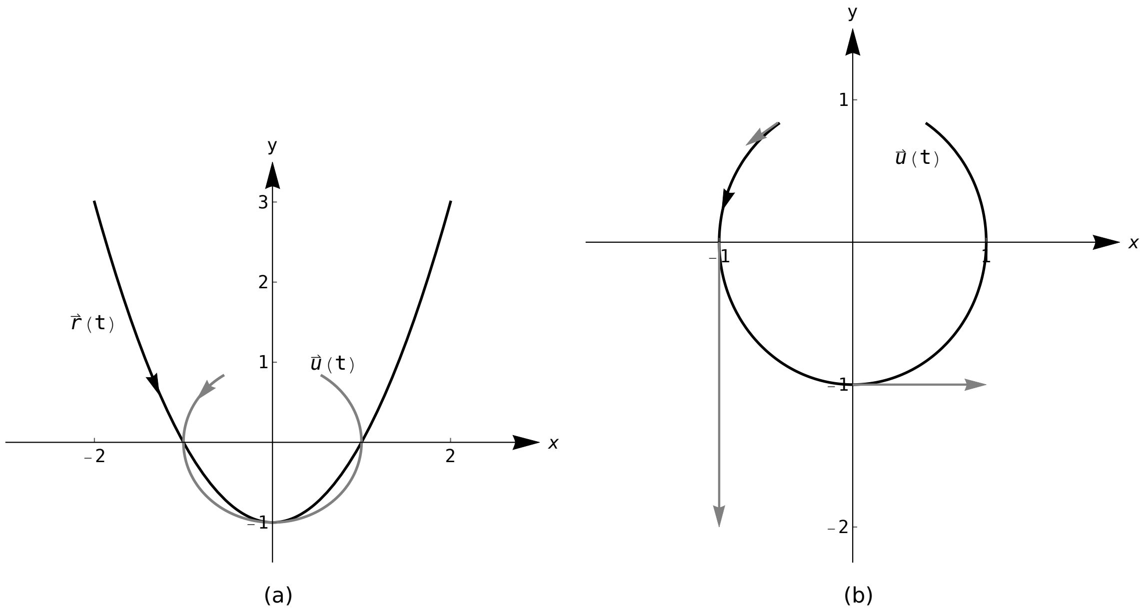


Figure 14.7: Graphing $\vec{\mathbf{r}}(t)$ and $\vec{\mathbf{u}}(t)$ (a) and some of the derivatives of $\vec{\mathbf{u}}(t)$ (b) in Example 14.6.

14.2.3.2 Tangent vector and lines

In Example 14.5, sketching a particular derivative with its initial point at the origin did not seem to reveal anything significant. However, when we sketched the vector with its initial point on the corresponding point on the graph, we did see something significant: the vector appeared to be tangent to the graph. We have not yet defined what tangent means in terms of curves in space; in fact, we use the derivative to define this term.

Definitie 14.7 (Tangent vector and line)
Let $\vec{\mathbf{r}}(t)$ be a differentiable vector–valued function on an open interval I containing c, where $\vec{\mathbf{r}}'(c) \neq \vec{\mathbf{0}}$.

1. A vector $\vec{\mathbf{v}}$ is **tangent** (*rakend*) to the graph of $\vec{\mathbf{r}}(t)$ at $t = c$ if $\vec{\mathbf{v}}$ is parallel to $\vec{\mathbf{r}}'(c)$.
2. The tangent line to the graph of $\vec{\mathbf{r}}(t)$ at $t = c$ is the line through $\vec{\mathbf{r}}(c)$ with direction parallel to $\vec{\mathbf{r}}'(c)$. An equation of the **tangent line** (*raaklijn*) is

$$\vec{\mathbf{y}} = \vec{\boldsymbol{\ell}}(t) = \vec{\mathbf{r}}(c) + t\,\vec{\mathbf{r}}'(c).$$

Example 14.7

Find the equations of the lines tangent to $\vec{\mathbf{r}}(t) = \left(t^3, t^2\right)$ at $t = -1$ and $t = 0$.

Solution

We find that $\vec{r}\,'(t) = \left(3t^2, 2t\right)$. At $t = -1$, we have

$$\vec{r}(-1) = (-1, 1) \quad \text{and} \quad \vec{r}\,'(-1) = (3, -2),$$

so the equation of the line tangent to the graph of $\vec{r}(t)$ at $t = -1$ is

$$\vec{\ell}(t) = (-1, 1) + t\,(3, -2).$$

This line is graphed with $\vec{r}(t)$ in Figure 14.8.

At $t = 0$, we have $\vec{r}\,'(0) = (0, 0) = \vec{0}$! This implies that the tangent line has no direction. We cannot apply Definition 14.7, hence cannot find the equation of the tangent line.

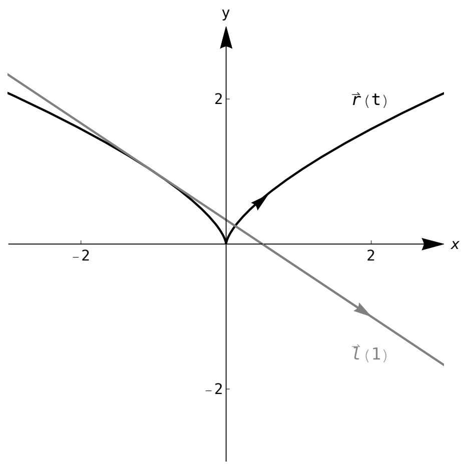


Figure 14.8: Graphing $\vec{r}(t)$ and its tangent line in Example 14.7.

14.2.3.3 Smoothness

We were unable to compute the equation of the tangent line to $\vec{r}(t) = \left(t^3, t^2\right)$ at $t = 0$ because $\vec{r}\,'(0) = \vec{0}$. The graph in Figure 14.8 shows that there is a cusp at this point. This leads us to another definition of **smooth** (*glad*), previously defined by Definition 9.7.

Definitie 14.8 (Smooth vector–valued function)
Let $\vec{r}(t)$ be a differentiable vector–valued function on an open interval I where $\vec{r}\,'(t)$ is continuous on I. $\vec{r}(t)$ is **smooth** on I if $\vec{r}\,'(t) \neq \vec{0}$ on I.

It is a basic geometric fact that a line tangent to a circle at a point P is perpendicular to the line passing through the center of the circle and P. This is illustrated in Figure 14.7(b); each tangent vector is perpendicular to the line that passes through its initial point and the centre of the circle. Since the centre of the circle is the origin, we can state this another way: $\vec{u}\,'(t)$ is orthogonal to $\vec{u}(t)$.

Recall that the dot product serves as a test for orthogonality: if $\vec{u} \cdot \vec{v} = 0$, then $\vec{u}$ is orthogonal to $\vec{v}$. Thus in the above example, $\vec{u}(t) \cdot \vec{u}\,'(t) = 0$. This is true for any vector–valued function that has a constant length, that is, that traces out part of a circle. It has important implications later on, so we state it as a theorem.

Theorem 14.4 (Vector–valued functions of constant length)
Let $\vec{\boldsymbol{r}}(t)$ be a vector–valued function of constant length that is differentiable on an open interval I. That is, $\|\vec{\boldsymbol{r}}(t)\| = c$ for all t in I. Then $\vec{\boldsymbol{r}}(t) \cdot \vec{\boldsymbol{r}}\,'(t) = 0$ for all t in I.

14.2.4 Integration

Before formally defining integrals of vector–valued functions, consider the following equation that our calculus experience tells us should be true:

$$\int_a^b \vec{\boldsymbol{r}}\,'(t)\;dt = \vec{\boldsymbol{r}}(b) - \vec{\boldsymbol{r}}(a).$$

That is, the integral of a rate of change function should give total change. In the context of vector–valued functions, this total change is displacement. The above equation is true; we now develop the theory to show why.

We can define antiderivatives and the indefinite integral of vector–valued functions in the same manner we defined indefinite integrals in Definition 12.1. However, we cannot define the definite integral of a vector–valued function as we did in Definition 12.2. That definition was based on the signed area between a function $y = f(x)$ and the x-axis. An area–based definition will not be useful in the context of vector–valued functions. Instead, we define the definite integral of a vector–valued function in a manner similar to that of Theorem 12.4, utilizing Riemann sums.

Definitie 14.9 (Antiderivatives, integrals of vector–valued functions)
Let $\vec{\boldsymbol{r}}(t)$ be a continuous vector–valued function on $[a,b]$. An **antiderivative of $\vec{\boldsymbol{r}}(t)$** (*primitieve functie*) is a function $\vec{\boldsymbol{R}}(t)$ such that $\vec{\boldsymbol{R}}'(t) = \vec{\boldsymbol{r}}(t)$.

The set of all antiderivatives of $\vec{\boldsymbol{r}}(t)$ is the **indefinite integral of $\vec{\boldsymbol{r}}(t)$** (*onbepaalde integraal*), denoted by

$$\int \vec{\boldsymbol{r}}(t)\;dt.$$

The **definite integral of $\vec{\boldsymbol{r}}(t)$** (*bepaalde integraal*) on $[a,b]$ is

$$\int_a^b \vec{\boldsymbol{r}}(t)\;dt = \lim_{\mathcal{T}\to 0}\sum_{i=1}^{n}\vec{\boldsymbol{r}}(c_i)\Delta t_i,$$

where Δt_i is the length of the i^{th} subinterval of a partition of $[a,b]$, $\mathcal{T}$ is the length of the largest subinterval in the partition, and c_i is any value in the i^{th} subinterval of the partition.

It is probably difficult to infer meaning from the definition of the definite integral. The important thing to realize from the definition is that it is built upon limits, which we can evaluate component–wise. The following theorem simplifies the computation of definite integrals.

Theorem 14.5 (Indefinite and definite integrals of vector–valued functions)
Let $\vec{\boldsymbol{r}}(t) = \big(f(t), g(t)\big)$ be a vector–valued function in $\mathbb{R}^2$ that are continuous on $[a,b]$.

1. $\displaystyle\int \vec{\boldsymbol{r}}(t)\ dt = \left(\int f(t)\ dt, \int g(t)\ dt\right)$

2. $\displaystyle\int_a^b \vec{\boldsymbol{r}}(t)\ dt = \left(\int_a^b f(t)\ dt, \int_a^b g(t)\ dt\right)$

A similar statement holds for vector–valued functions in $\mathbb{R}^3$.

Example 14.8

Let $\vec{\boldsymbol{r}}(t) = \left(e^{2t}, \sin(t)\right)$. Evaluate

$$\int_0^1 \vec{\boldsymbol{r}}(t)\ dt.$$

Solution

We follow Theorem 14.5.

$$\begin{aligned}
\int_0^1 \vec{\boldsymbol{r}}(t)\ dt &= \int_0^1 \left(e^{2t}, \sin(t)\right)\ dt \\
&= \left(\int_0^1 e^{2t}\ dt\ , \int_0^1 \sin(t)\ dt\right) \\
&= \left(\frac{1}{2}e^{2t}\Big|_0^1\ , (-\cos(t))\Big|_0^1\right) \\
&= \left(\frac{1}{2}(e^2-1)\ , -\cos(1)+1\right) \\
&\approx (3.19, 0.460)
\end{aligned}$$

What does the integration of a vector–valued function mean? There are many applications, but none as direct as the area under the curve that we used in understanding the integral of a real–valued function. A key understanding for us comes from considering the integral of a derivative:

$$\int_a^b \vec{\boldsymbol{r}}\,'(t)\ dt = \vec{\boldsymbol{r}}(t)\Big|_a^b = \vec{\boldsymbol{r}}(b) - \vec{\boldsymbol{r}}(a).$$

This indicates integrating a rate of change function gives displacement.

Noting that vector–valued functions are closely related to parametric equations, we can describe the arc length of the graph of a vector–valued function as an integral. Given parametric equations $x = f(t)$,

$y = g(t)$, the arc length on $[a, b]$ of the graph is

$$L = \int_a^b \sqrt{f'(t)^2 + g'(t)^2}\, dt,$$

as stated in Theorem 13.8. If $\vec{r}(t) = \big(f(t), g(t)\big)$, note that $\sqrt{f'(t)^2 + g'(t)^2} = \left\|\vec{r}\,'(t)\right\|$. Therefore we can express the arc length of the graph of a vector–valued function as an integral of the magnitude of its derivative.

Theorem 14.6 (Arc length of a vector–valued function)
Let $\vec{r}(t)$ be a vector–valued function where $\vec{r}\,'(t)$ is continuous on $[a, b]$. The ***arc length*** *(booglengte) L of the graph of $\vec{r}(t)$ is*

$$L = \int_a^b \left\|\vec{r}\,'(t)\right\|\, dt.$$

Note that we are actually integrating a scalar function here, not a vector–valued function.

The remainder of this section takes what we have established thus far and applies it to objects in motion.

14.2.5 The calculus of motion

A common use of vector–valued functions is to describe the motion of an object in the plane or in space. A position function $\vec{r}(t)$ gives the position of an object at time t. This section explores how derivatives and integrals are used to study the motion described by such a function.

Definitie 14.10 (Velocity, speed and acceleration)
Let $\vec{r}(t)$ be a position function in $\mathbb{R}^2$ or $\mathbb{R}^3$.

1. **Velocity**, denoted $\vec{v}(t)$, is the instantaneous rate of position change; that is, $\vec{v}(t) = \vec{r}\,'(t)$.
2. **Speed** is the magnitude of velocity, $\left\|\vec{v}(t)\right\|$.
3. **Acceleration**, denoted $\vec{a}(t)$, is the instantaneous rate of velocity change; that is, $\vec{a}(t) = \vec{v}\,'(t) = \vec{r}\,''(t)$.

Example 14.9

An object is moving with position function $\vec{r}(t) = \left(t^2 - t, t^2 + t\right)$, $-3 \le t \le 3$, where distances are measured in metres and time is measured in seconds.

1. Find $\vec{v}(t)$ and $\vec{a}(t)$.
2. Sketch $\vec{r}(t)$; plot $\vec{v}(-1)$, $\vec{a}(-1)$, $\vec{v}(1)$ and $\vec{a}(1)$, each with their initial point at their corresponding point on the graph of $\vec{r}(t)$.
3. When is the object's speed minimized?

Solution

1. Taking derivatives, we find

$$\vec{v}(t) = \vec{r}\,'(t) = \langle 2t-1, 2t+1\rangle \quad \text{and} \quad \vec{a}(t) = \vec{r}\,''(t) = \langle 2, 2\rangle.$$

   Note that the acceleration is constant.

2. $\vec{v}(-1) = \langle -3, -1\rangle$, $\vec{a}(-1) = \langle 2, 2\rangle$; $\vec{v}(1) = \langle 1, 3\rangle$, $\vec{a}(1) = \langle 2, 2\rangle$. These are plotted with $\vec{r}(t)$ in Figure 14.9(a).

   We can think of acceleration as pulling the velocity vector in a certain direction. At $t = -1$, the velocity vector points down and to the left; at $t = 1$, the velocity vector has been pulled in the $\langle 2, 2\rangle$ direction and is now pointing up and to the right. In Figure 14.9(b) we plot more velocity/acceleration vectors, making more clear the effect acceleration has on velocity.

   Since $\vec{a}(t)$ is constant in this example, as t grows large $\vec{v}(t)$ becomes almost parallel to $\vec{a}(t)$. For instance, when $t = 10$, $\vec{v}(10) = \langle 19, 21\rangle$, which is nearly parallel to $\langle 2, 2\rangle$.

3. The object's speed is given by

$$\|\vec{v}(t)\| = \sqrt{(2t-1)^2 + (2t+1)^2} = \sqrt{8t^2+2}.$$

   To find the minimal speed, we could apply calculus techniques (such as set the derivative equal to 0 and solve for t, etc.) but we can find it by inspection. Inside the square root we have a quadratic which is minimized when $t = 0$. Thus the speed is minimized at $t = 0$, with a speed of $\sqrt{2}$ m/s.

   The graph in Figure 14.9(b) also implies speed is minimized here. The filled dots on the graph are located at integer values of t between -3 and 3. Dots that are far apart imply the object travelled a far distance in 1 second, indicating high speed; dots that are close together imply the object did not travel far in 1 second, indicating a low speed. The dots are closest together near $t = 0$, implying the speed is minimized near that value.

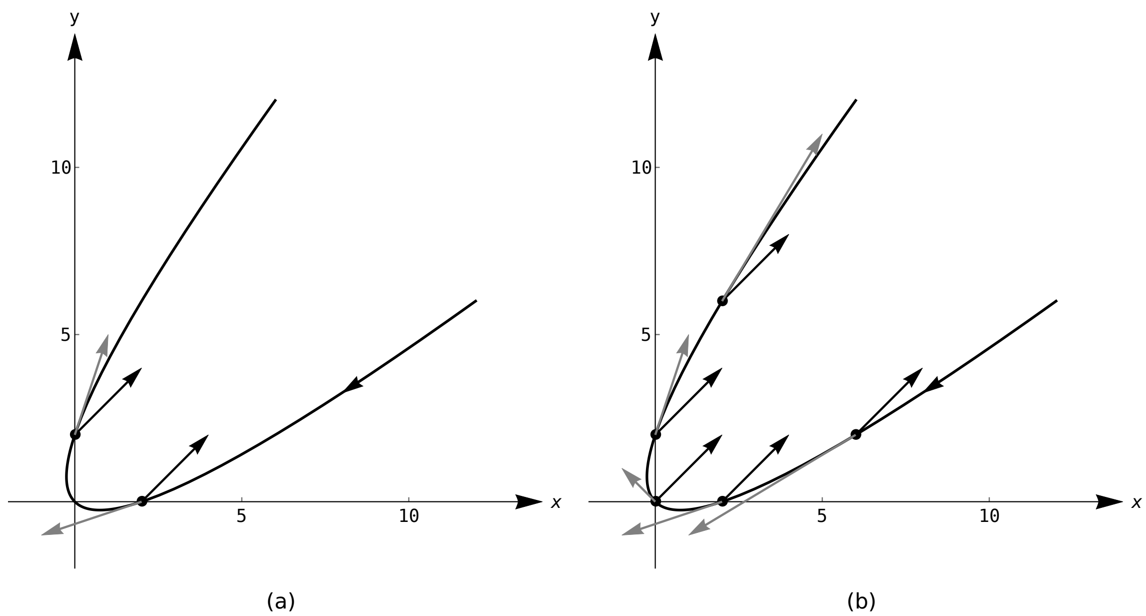


Figure 14.9: Graphing the position (curve), velocity (gray) and acceleration (black) of an object in Example 14.9.

If an object travels at a constant speed, we have $\left\|\vec{\boldsymbol{v}}(t)\right\| = c$ for some constant c. Recall Theorem 14.4, which states that if a vector–valued function $\vec{\boldsymbol{r}}(t)$ has constant length, then $\vec{\boldsymbol{r}}(t)$ is perpendicular to its derivative: $\vec{\boldsymbol{r}}(t) \cdot \vec{\boldsymbol{r}}\,'(t) = 0$. So, the corresponding velocity function has constant length, therefore we can conclude that the velocity is perpendicular to the acceleration: $\vec{\boldsymbol{v}}(t) \cdot \vec{\boldsymbol{a}}(t) = 0$.

An important application of vector–valued position functions is projectile motion: the motion of objects under only the influence of gravity. We will measure time in seconds, and distances will either be in meters or feet. We will show that we can completely describe the path of such an object knowing its initial position and initial velocity.

Suppose an object has initial position $\vec{\boldsymbol{r}}(0) = (x_0, y_0)$ and initial velocity $\vec{\boldsymbol{v}}(0) = \big(v_x, v_y\big)$. It is customary to rewrite $\vec{\boldsymbol{v}}(0)$ in terms of its speed v_0 and direction $\vec{\boldsymbol{u}}$, where $\vec{\boldsymbol{u}}$ is a unit vector. Recall all unit vectors in $\mathbb{R}^2$ can be written as $\big(\cos(\theta), \sin(\theta)\big)$, where θ is an angle measure counter–clockwise from the x-axis. We refer to θ as the angle of elevation. Thus $\vec{\boldsymbol{v}}(0) = v_0\big(\cos(\theta), \sin(\theta)\big)$.

Since the acceleration of the object is known, namely $\vec{\boldsymbol{a}}(t) = (0, -g)$, where g is the gravitational constant, we can find $\vec{\boldsymbol{r}}(t)$ knowing our two initial conditions. We first find $\vec{\boldsymbol{v}}(t)$:

$$\begin{aligned} \vec{\boldsymbol{v}}(t) &= \int \vec{\boldsymbol{a}}(t)\ dt \\ \Rightarrow\ \vec{\boldsymbol{v}}(t) &= \int (0, -g)\ dt \\ \Leftrightarrow\ \vec{\boldsymbol{v}}(t) &= (0, -gt) + \vec{\boldsymbol{C}}. \end{aligned}$$

Knowing $\vec{\boldsymbol{v}}(0) = v_0\big(\cos(\theta), \sin(\theta)\big)$, we have $\vec{\boldsymbol{C}} = v_0\big(\cos(\theta), \sin(\theta)\big)$ and so

$$\vec{\boldsymbol{v}}(t) = \big(v_0\cos(\theta), -gt + v_0\sin(\theta)\big).$$

We integrate once more to find $\vec{\boldsymbol{r}}(t)$:

$$\begin{aligned} \vec{\boldsymbol{r}}(t) &= \int \vec{\boldsymbol{v}}(t)\ dt \\ \vec{\boldsymbol{r}}(t) &= \int \big(v_0\cos(\theta), -gt + v_0\sin(\theta)\big)\ dt \\ \vec{\boldsymbol{r}}(t) &= \left(\big(v_0\cos(\theta)\big)t, -\frac{1}{2}gt^2 + \big(v_0\sin(\theta)\big)t\right) + \vec{\boldsymbol{C}}. \end{aligned}$$

Knowing $\vec{\boldsymbol{r}}(0) = (x_0, y_0)$, we conclude $\vec{\boldsymbol{C}} = (x_0, y_0)$ and

$$\vec{\boldsymbol{r}}(t) = \left(\big(v_0\cos(\theta)\big)t + x_0\ , -\frac{1}{2}gt^2 + \big(v_0\sin(\theta)\big)t + y_0\ \right).$$

This is the position function of a projectile propelled from an initial position of $\vec{\boldsymbol{r}}_0 = (x_0, y_0)$, with initial speed v_0, with angle of elevation θ and neglecting all accelerations but gravity.

We can also rely on vector–valued functions to compute the distance travelled. For instance, consider a driver who sets her cruise–control to 60 km/h, and travels at this speed for an hour. We can ask:

1. How far did the driver travel?
2. How far from her starting position is the driver?

The first is easy to answer: she travelled 60 kilometres. The second is impossible to answer with the given information. We do not know if she travelled in a straight line, on an oval racetrack, or along a slowly–winding highway.

This highlights an important fact: to compute distance travelled, we need only to know the speed, given by $\|\vec{\boldsymbol{v}}(t)\|$.

Definitie 14.11 (Distance travelled)
Let $\vec{\boldsymbol{v}}(t)$ be a velocity function for a moving object. The **distance travelled** (*afgelegde afstand*) by the object on $[a, b]$ is:

$$\text{distance travelled} = \int_a^b \|\vec{\boldsymbol{v}}(t)\| \; dt.$$

Note that this is just a restatement of Theorem 14.6: arc length is the same as distance travelled, just viewed in a different context.

Example 14.10

A particle moves in space with position function $\vec{\boldsymbol{r}}(t) = \left(t, t^2, \sin(\pi t)\right)$ on $[-2, 2]$, where t is measured in seconds and distances are in meters. Find:

1. The distance travelled by the particle on $[-2, 2]$.
2. The displacement of the particle on $[-2, 2]$.
3. The particle's average speed.

Solution

1. We use Definition 14.11 to establish the integral:

$$\begin{aligned}\text{distance travelled} &= \int_{-2}^{2} \|\vec{\boldsymbol{v}}(t)\| \; dt \\ &= \int_{-2}^{2} \sqrt{1 + (2t)^2 + \pi^2\cos^2(\pi t)} \; dt.\end{aligned}$$

   This cannot be solved in terms of elementary functions so we turn to numerical integration, finding the distance to be 12.88m.

2. The displacement is the vector

$$\vec{\boldsymbol{r}}(2) - \vec{\boldsymbol{r}}(-2) = (2, 4, 0) - (-2, 4, 0) = (4, 0, 0).$$

   That is, the particle ends with an x-value increased by 4 and with y- and z-values the same.

3. We found above that the particle travelled 12.88m over 4 seconds. We can compute average speed by dividing: 12.88/4 = 3.22 m/s. We should also consider Definition 12.5 to compute the average value of the speed as

$$\text{average speed} = \frac{1}{2 - (-2)} \int_{-2}^{2} \|\vec{\boldsymbol{v}}(t)\| \; dt \approx \frac{1}{4} 12.88 = 3.22\text{m/s}.$$

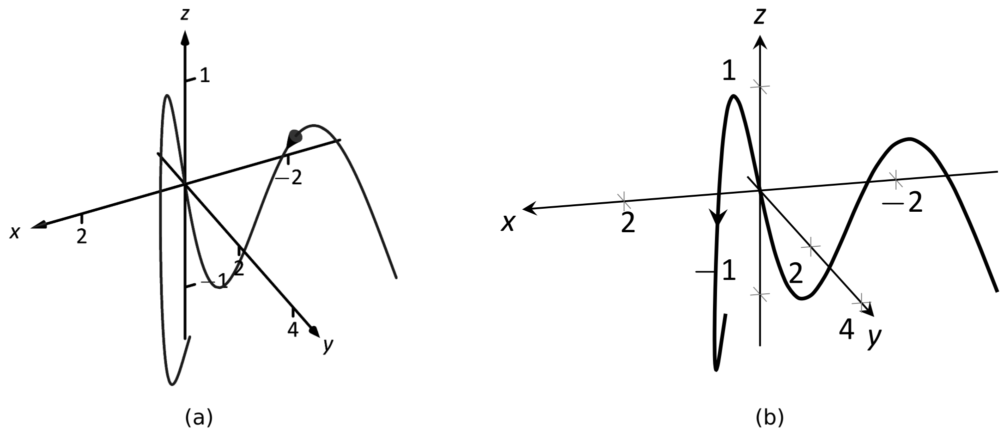


Figure 14.10: The path of the particle, from two perspectives, in Example 14.10.

Note how in Example 14.10 we computed the average speed as the average value of $\|\vec{\boldsymbol{v}}(t)\|$ on $[-2,2]$. Likewise, given the position function $\vec{\boldsymbol{r}}(t)$, the average velocity on $[a,b]$ is

$$\frac{\text{displacement}}{\text{travel time}} = \frac{1}{b-a}\int_a^b \vec{\boldsymbol{r}}\,'(t)\ dt = \frac{\vec{\boldsymbol{r}}(b)-\vec{\boldsymbol{r}}(a)}{b-a},$$

that is, it is the average value of $\vec{\boldsymbol{r}}\,'(t)$, or $\vec{\boldsymbol{v}}(t)$, on $[a,b]$.

14.3 Unit tangent and normal vectors

Given a smooth vector–valued function $\vec{\boldsymbol{r}}(t)$, we defined in Definition 14.7 that any vector parallel to $\vec{\boldsymbol{r}}\,'(t_0)$ is tangent to the graph of $\vec{\boldsymbol{r}}(t)$ at $t = t_0$. It is often useful to consider just the direction of $\vec{\boldsymbol{r}}\,'(t)$ and not its magnitude. Therefore we are interested in the unit vector in the direction of $\vec{\boldsymbol{r}}\,'(t)$. This leads to a definition.

Definitie 14.12 (Unit tangent vector)
Let $\vec{\boldsymbol{r}}(t)$ be a smooth function on an open interval I. The **unit tangent vector** $\widehat{\boldsymbol{T}}(t)$ (*eenheids-raakvector*) is

$$\widehat{\boldsymbol{T}}(t) = \frac{1}{\|\vec{\boldsymbol{r}}\,'(t)\|}\vec{\boldsymbol{r}}\,'(t).$$

Just as knowing the direction tangent to a path is important, knowing a direction orthogonal to a path is important. When dealing with real-valued functions, we defined the normal line at a point to be the line through the point that was perpendicular to the tangent line at that point. We can do a similar thing with vector–valued functions. Given $\vec{\boldsymbol{r}}(t)$ in $\mathbb{R}^2$, we have 2 directions perpendicular to the tangent vector, as shown in Figure 14.11.

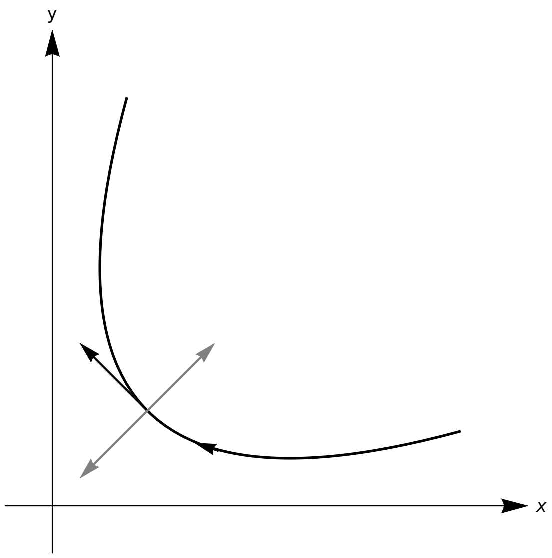


Figure 14.11: Given a direction in the plane, there are always two directions orthogonal to it.

Given $\vec{r}(t)$ in $\mathbb{R}^3$, there are infinitely many vectors orthogonal to the tangent vector at a given point. We might wonder whether one of these infinite choices preferable over the others.

The answer in both $\mathbb{R}^2$ and $\mathbb{R}^3$ is "Yes, there is one vector that is not only preferable, it is the right one to choose. Recall that if $\vec{r}(t)$ has constant length, then $\vec{r}(t)$ is orthogonal to $\vec{r}\,'(t)$ for all t. We know $\widehat{T}(t)$, the unit tangent vector, has constant length. Therefore $\widehat{T}(t)$ is orthogonal to $\widehat{T}\,'(t)$.

We will see that $\widehat{T}\,'(t)$ is more than just a convenient choice of vector that is orthogonal to $\vec{r}\,'(t)$; rather, it is the right choice. Since all we care about is the direction, we define this newly found vector to be a unit vector. Note that if $\widehat{T}(t)$ is a unit vector, this does not imply that $\widehat{T}\,'(t)$ is also a unit vector.

Definitie 14.13 (Unit normal vector)
Let $\vec{r}(t)$ be a vector–valued function where the unit tangent vector, $\widehat{T}(t)$, is smooth on an open interval I. The **unit normal vector** $\widehat{N}(t)$ (*eenheidsnormaalvector*) is

$$\widehat{N}(t) = \frac{1}{\left\|\widehat{T}\,'(t)\right\|}\widehat{T}\,'(t).$$

Example 14.11

Let

$$\vec{r}(t) = \left(t^2 - t, t^2 + t\right).$$

1. Find $\widehat{T}(t)$ and compute $\widehat{T}(0)$ and $\widehat{T}(1)$.
2. Find $\widehat{N}(t)$ and sketch $\vec{r}(t)$ with the unit tangent and normal vectors at $t = -1, 0$ and 1.

Solution

1. We find $\vec{r}\,'(t) = (2t - 1, 2t + 1)$, and

$$\left\|\vec{r}\,'(t)\right\| = \sqrt{(2t-1)^2 + (2t+1)^2} = \sqrt{8t^2 + 2}.$$

Therefore

$$\widehat{\boldsymbol{T}}(t)=\frac{1}{\left\|\vec{r}\,'(t)\right\|}\vec{r}\,'(t)=\frac{1}{\sqrt{8t^2+2}}\left(2t-1,2t+1\right)=\left(\frac{2t-1}{\sqrt{8t^2+2}},\frac{2t+1}{\sqrt{8t^2+2}}\right).$$

When $t=0$, we have $\widehat{\boldsymbol{T}}(0)=\left(-1/\sqrt{2},1/\sqrt{2}\right)$; when $t=1$, we have $\widehat{\boldsymbol{T}}(1)=\left(1/\sqrt{10},3/\sqrt{10}\right)$. We leave it to the reader to verify each of these is a unit vector.

2. Given $\widehat{\boldsymbol{T}}(t)$, finding $\widehat{\boldsymbol{T}}\,'(t)$ requires two applications of the quotient rule:

$$\begin{aligned}T\,'(t)&=\left(\frac{\sqrt{8t^2+2}(2)-(2t-1)\left(\frac{1}{2}(8t^2+2)^{-1/2}(16t)\right)}{8t^2+2},\right.\\&\qquad\left.\frac{\sqrt{8t^2+2}(2)-(2t+1)\left(\frac{1}{2}(8t^2+2)^{-1/2}(16t)\right)}{8t^2+2}\right)\\&=\left(\frac{4(2t+1)}{\left(8t^2+2\right)^{3/2}},\frac{4(1-2t)}{\left(8t^2+2\right)^{3/2}}\right).\end{aligned}$$

This is not a unit vector; to find $\widehat{\boldsymbol{N}}(t)$, we need to divide $\widehat{\boldsymbol{T}}\,'(t)$ by its magnitude:

$$\begin{aligned}\left\|\widehat{\boldsymbol{T}}\,'(t)\right\|&=\sqrt{\frac{16(2t+1)^2}{(8t^2+2)^3}+\frac{16(1-2t)^2}{(8t^2+2)^3}}\\&=\sqrt{\frac{16(8t^2+2)}{(8t^2+2)^3}}\\&=\frac{4}{8t^2+2}.\end{aligned}$$

Finally,

$$\begin{aligned}\widehat{\boldsymbol{N}}(t)=\frac{1}{\left\|\widehat{\boldsymbol{T}}'(t)\right\|}\widehat{\boldsymbol{T}}'(t)&=\frac{1}{4/(8t^2+2)}\left(\frac{4(2t+1)}{\left(8t^2+2\right)^{3/2}},\frac{4(1-2t)}{\left(8t^2+2\right)^{3/2}}\right)\\&=\left(\frac{2t+1}{\sqrt{8t^2+2}},-\frac{2t-1}{\sqrt{8t^2+2}}\right).\end{aligned}$$

Using this formula for $\widehat{\boldsymbol{N}}(t)$, we compute the unit tangent and normal vectors for $t=-1,0$ and 1 and sketch them in Figure 14.12.

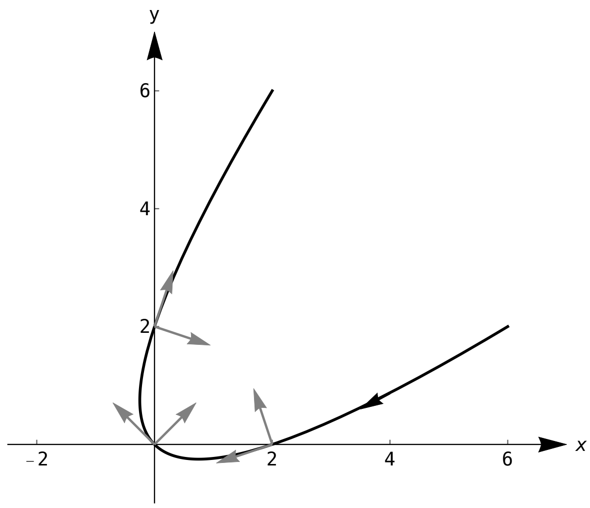


Figure 14.12: Unit tangent and normal vectors from Example 14.11.

The final result for $\widehat{\boldsymbol{N}}(t)$ in Example 14.11 is suspiciously similar to $\widehat{\boldsymbol{T}}(t)$. There is a clear reason for this. If $\vec{\boldsymbol{u}} = (u_1, u_2)$ is a unit vector in $\mathbb{R}^2$, then the only unit vectors orthogonal to $\vec{\boldsymbol{u}}$ are $(-u_2, u_1)$ and $(u_2, -u_1)$. Given $\widehat{\boldsymbol{T}}(t)$, we can quickly determine $\widehat{\boldsymbol{N}}(t)$ if we know which term to multiply by (-1). Consider again Figure 14.12, where we have plotted some unit tangent and normal vectors. Note how $\widehat{\boldsymbol{N}}(t)$ always points inside the curve, or to the concave side of the curve. This is not a coincidence; this is true in general. Knowing the direction that $\vec{\boldsymbol{r}}(t)$ turns allows us to quickly find $\widehat{\boldsymbol{N}}(t)$.

Theorem 14.7 (Unit normal vectors in $\mathbb{R}^2$)
Let $\vec{\boldsymbol{r}}(t)$ be a vector–valued function in $\mathbb{R}^2$ where $\widehat{\boldsymbol{T}}'(t)$ is smooth on an open interval I. Let t_0 be in I and $\widehat{\boldsymbol{T}}(t_0) = (t_1, t_2)$ Then $\widehat{\boldsymbol{N}}(t_0)$ is either

$$\widehat{\boldsymbol{N}}(t_0) = (-t_2, t_1) \quad \text{or} \quad \widehat{\boldsymbol{N}}(t_0) = (t_2, -t_1),$$

whichever is the vector that points to the concave side of the graph of $\vec{\boldsymbol{r}}$.

14.4 Arc length and curvature

14.4.1 Arc length

Currently, our vector–valued functions have defined points with a parameter t, which we often take to represent time. Consider Figure 14.13(a), where $\vec{\boldsymbol{r}}(t) = \left(t^2 - t, t^2 + t\right)$ is graphed and the points corresponding to $t = 0$, 1 and 2 are shown. Note how the arc length between $t = 0$ and $t = 1$ is smaller than the arc length between $t = 1$ and $t = 2$; if the parameter t is time and $\vec{\boldsymbol{r}}$ is position, we can say that the particle travelled faster on $[1, 2]$ than on $[0, 1]$.

Now consider Figure 14.13(b), where the same graph is parametrized by a different variable s. Points corresponding to $s = 0$ through $s = 6$ are plotted. The arc length of the graph between each adjacent pair of points is 1. We can view this parameter s as distance; that is, the arc length of the graph from $s = 0$ to $s = 3$ is 3, the arc length from $s = 2$ to $s = 6$ is 4, etc. If one wants to find the point 2.5 units from an initial location (i.e., $s = 0$), one would compute $\vec{\boldsymbol{r}}(2.5)$. This parameter s is very useful, and is called the **arc length parameter**.

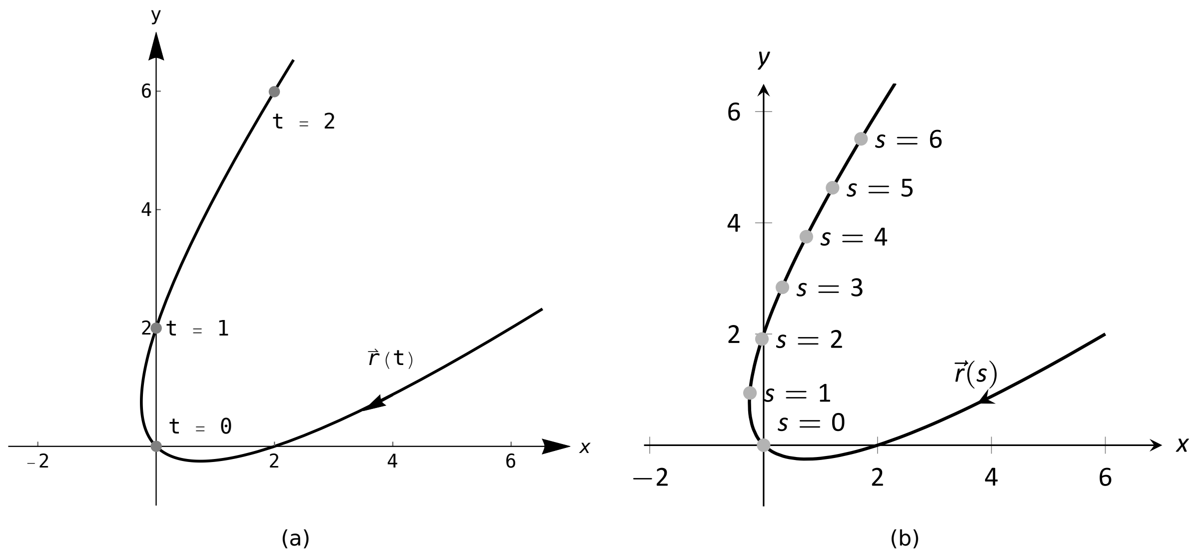


Figure 14.13: Introducing the arc length parameter.

How do we find the arc length parameter?

Start with any parametrization of $\vec{\boldsymbol{r}}$. We can compute the arc length of the graph of $\vec{\boldsymbol{r}}$ on the interval $[0, t]$ with

$$\text{arc length } = \int_0^t \left\|\vec{\boldsymbol{r}}\,'(u)\right\|\ du.$$

We can turn this into a function: as t varies, we find the arc length s from 0 to t. This function is

$$s(t) = \int_0^t \left\|\vec{\boldsymbol{r}}\,'(u)\right\|\ du. \tag{14.2}$$

This establishes a relationship between s and t. Knowing this relationship explicitly, we can rewrite $\vec{\boldsymbol{r}}(t)$ as a function of s: $\vec{\boldsymbol{r}}(s)$. We demonstrate this in an example.

Example 14.12

Let $\vec{\boldsymbol{r}}(t) = (3t - 1, 4t + 2)$. Parametrize $\vec{\boldsymbol{r}}$ with the arc length parameter s.

Solution

Using Equation (14.2), we write

$$s(t) = \int_0^t \left\|\vec{\boldsymbol{r}}\,'(u)\right\|\ du.$$

We can integrate this, explicitly finding a relationship between s and t:

$$\begin{aligned} s(t) &= \int_0^t \left\|\vec{\boldsymbol{r}}\,'(u)\right\|\ du \\ &= \int_0^t \sqrt{3^2 + 4^2}\ du \end{aligned}$$

$$= \int_0^t 5\, du$$

$$= 5t.$$

Since $s = 5t$, we can write $t = s/5$ and replace t in $\vec{r}(t)$ with $s/5$:

$$\vec{r}(s) = \left(3\left(\frac{s}{5}\right) - 1, 4\left(\frac{s}{5}\right) + 2\right) = \left(\frac{3}{5}s - 1, \frac{4}{5}s + 2\right).$$

Clearly, as shown in Figure 14.14, the graph of $\vec{r}$ is a line, where $t = 0$ corresponds to the point $(-1, 2)$. What point on the line is 2 units away from this initial point? We find it with $\vec{r}(2) = (1/5, 18/5)$.

Is the point $(1/5, 18/5)$ really 2 units away from $(-1, 2)$? We use the distance formula to check:

$$d = \sqrt{\left(\frac{1}{5} - (-1)\right)^2 + \left(\frac{18}{5} - 2\right)^2} = \sqrt{\frac{36}{25} + \frac{64}{25}} = \sqrt{4} = 2.$$

Yes, $\vec{r}(2)$ is indeed 2 units away, in the direction of travel, from the initial point.

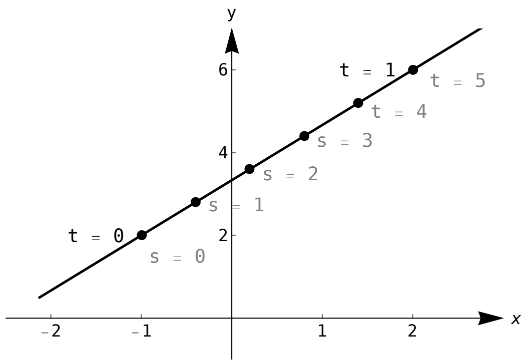


Figure 14.14: Graphing $\vec{r}$ in Example 14.12 with parameters t and s.

Things worked out very nicely in Example 14.12; we were able to establish directly that $s = 5t$. Usually, the arc length parameter is much more difficult to describe in terms of t, a result of integrating a square–root. There are a number of things that we can learn about the arc length parameter from Equation (14.2), though, that are useful.

First, take the derivative of s with respect to t. The fundamental theorem of calculus (see Theorem 12.6) states that

$$\frac{ds}{dt} = s'(t) = \left\|\vec{r}\,'(t)\right\|. \tag{14.3}$$

Letting t represent time and $\vec{r}(t)$ represent position, we see that the rate of change of s with respect to t is speed; that is, the rate of change of distance travelled is speed, which should match our intuition.

The chain rule states that

$$\frac{d\vec{r}}{dt} = \frac{d\vec{r}}{ds}\frac{ds}{dt}$$
$$\vec{r}\,'(t) = \vec{r}\,'(s)\left\|\vec{r}\,'(t)\right\|.$$

Solving for $\vec{r}\,'(s)$, we have

$$\vec{r}\,'(s) = \frac{\vec{r}\,'(t)}{\|\vec{r}\,'(t)\|} = \widehat{\mathbf{T}}(t), \tag{14.4}$$

where $\widehat{\mathbf{T}}(t)$ is the unit tangent vector. Equation (14.4) is often misinterpreted, as one is tempted to think it states $\vec{r}\,'(t) = \widehat{\mathbf{T}}(t)$, but there is a big difference between $\vec{r}\,'(s)$ and $\vec{r}\,'(t)$. The key to take from it is that $\vec{r}\,'(s)$ is a unit vector. In fact, the following definition states that this characterizes the arc length parameter.

Definitie 14.14 (Arc length parameter)
Let $\vec{r}(s)$ be a vector–valued function. The parameter s is the **arc length parameter** if, and only if, $\|\vec{r}\,'(s)\| = 1$.

14.4.2 Curvature

Consider points A and B on the curve graphed in Figure 14.15(a). One can readily argue that the curve curves more sharply at A than at B. It is useful to use a number to describe how sharply the curve bends; that number is the **curvature** (*kromming*) of the curve.

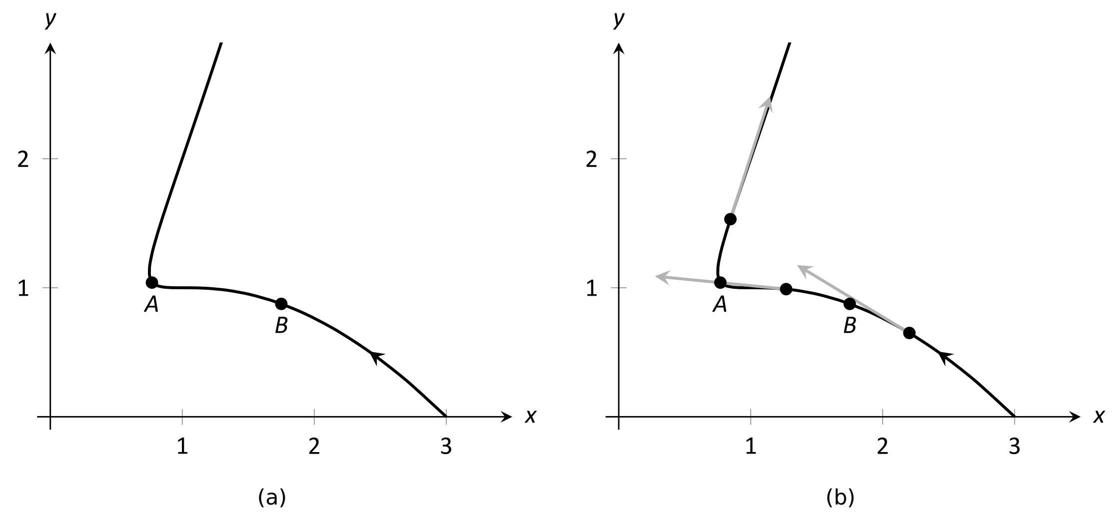


Figure 14.15: Establishing the concept of curvature.

We derive this number in the following way. Consider Figure 14.15(b), where unit tangent vectors are graphed around points A and B. Notice how the direction of the unit tangent vector changes quite a bit near A, whereas it does not change as much around B. This leads to an important concept: measuring the rate of change of the unit tangent vector with respect to arc length gives us a measurement of curvature.

Definitie 14.15 (Curvature)
Let $\vec{r}(s)$ be a vector–valued function where s is the arc length parameter. The **curvature** κ of the graph of $\vec{r}(s)$ is

$$\kappa(t) = \left\|\frac{d\widehat{\mathbf{T}}}{ds}\right\| = \left\|\widehat{\mathbf{T}}\,'(s)\right\|.$$

If $\vec{r}(s)$ is parametrized by the arc length parameter, then

$$\widehat{T}(s) = \frac{\vec{r}\,'(s)}{\|\vec{r}\,'(s)\|} \quad \text{and} \quad \widehat{N}(s) = \frac{\widehat{T}\,'(s)}{\left\|\widehat{T}\,'(s)\right\|}.$$

Having defined $\left\|\widehat{T}\,'(s)\right\| = \kappa$, we can rewrite the second equation as

$$\widehat{T}\,'(s) = \kappa\widehat{N}(s). \tag{14.5}$$

We already knew that $\widehat{T}\,'(s)$ is in the same direction as $\widehat{N}(s)$; that is, we can think of $\widehat{T}(s)$ as being pulled in the direction of $\widehat{N}(s)$. How hard is it being pulled? By a factor of κ. When the curvature is large, $\widehat{T}(s)$ is being pulled hard and the direction of $\widehat{T}(s)$ changes rapidly. When κ is small, $T(s)$ is not being pulled hard and hence its direction is not changing rapidly.

Example 14.13

Find the curvature of $\vec{r}(t) = (3t-1, 4t+2)$.

Solution

In Example 14.12, we found that the arc length parameter was defined by $s = 5t$, so $\vec{r}(s) = (3s/5-1, 4s/5+2)$ parametrized $\vec{r}$ with the arc length parameter. To find κ, we need to find $\widehat{T}\,'(s)$.

$$\begin{aligned} \widehat{T}(s) &= \vec{r}\,'(s) \qquad \text{(recall this is a unit vector)} \\ &= \left(\frac{3}{5}, \frac{4}{5}\right). \end{aligned}$$

Therefore

$$\widehat{T}\,'(s) = (0, 0)$$

and

$$\kappa = \left\| \widehat{T}\,'(s) \right\| = 0.$$

It probably comes as no surprise that the curvature of a line is 0.

While the definition of curvature is a beautiful mathematical concept, it is nearly impossible to use most of the time; writing $\vec{r}$ in terms of the arc length parameter is generally very hard. Fortunately, there are other methods of calculating this value that are much easier. There is a trade-off: the definition is easy to understand though hard to compute, whereas these other formulas are easy to compute though it may be hard to understand why they work.

Theorem 14.8 (Formulas for curvature)

Let C be a smooth curve in the plane or in space.

1. *If C is defined in space by a vector–valued function $\vec{r}(t)$, then*

$$\kappa(t) = \frac{\left\|\widehat{T}\,'(t)\right\|}{\left\|\vec{r}\,'(t)\right\|} = \frac{\left\|\vec{r}\,'(t) \times \vec{r}\,''(t)\right\|}{\left\|\vec{r}\,'(t)\right\|^3} = \frac{\vec{a}(t) \cdot \widehat{N}(t)}{\left\|\vec{v}(t)\right\|^2}.$$

2. *If C is defined by $y = f(x)$, then*

$$\kappa = \frac{|f''(x)|}{\Big(1+\big(f'(x)\big)^2\Big)^{3/2}}.$$

3. *If C is defined as a vector–valued function in the plane, $\vec{r}(t) = \big(x(t), y(t)\big)$, then*

$$\kappa = \frac{|x'y''-x''y'|}{\big((x')^2+(y')^2\big)^{3/2}}.$$

Example 14.14

Find the curvature of a circle with radius r, defined by $\vec{c}(t) = \big(r\cos(t), r\sin(t)\big)$.

Solution

Before we start, we should expect the curvature of a circle to be constant, and not dependent on t.

We compute κ using the second part of Theorem 14.8:

$$\begin{aligned}
\kappa &= \frac{\Big|\big(-r\sin(t)\big)\big(-r\sin(t)\big)-\big(-r\cos(t)\big)\big(r\cos(t)\big)\Big|}{\Big(\big(-r\sin(t)\big)^2+\big(r\cos(t)\big)^2\Big)^{3/2}} \\
&= \frac{r^2\big(\sin^2(t)+\cos^2(t)\big)}{\Big(r^2\big(\sin^2(t)+\cos^2(t)\big)\Big)^{3/2}} \\
&= \frac{r^2}{r^3} = \frac{1}{r}.
\end{aligned}$$

We have found that a circle with radius r has curvature $\kappa = 1/r$.

Example 14.14 gives a great result. Before this example, if we were told "The curve has a curvature of 5 at point A," we would have no idea what this really meant. Is 5 big – does is correspond to a really sharp turn, or a not-so-sharp turn? Now we can think of 5 in terms of a circle with radius 1/5. Knowing the units allows us to determine how sharply the curve is curving.

Let a point P on a smooth curve C be given, and let κ be the curvature of the curve at P. A circle that:

- passes through P,
- lies on the concave side of C,
- has a common tangent line as C at P, and
- has radius $r = 1/\kappa$ (hence has curvature κ)

is the **osculating circle** (*osculatiecirkel*), or **circle of curvature**, to C at P, and r is the **radius of curvature** (*kromtestraal*). Figure 14.16 shows the graph of the curve seen earlier in Figure 14.15(a) and its osculating circles at A and B. A sharp turn corresponds to a circle with a small radius; a gradual turn corresponds to a circle with a large radius. Being able to think of curvature in terms of the radius of a circle is very useful. The word osculating comes from a Latin word related to kissing; an osculating circle kisses the graph at a particular point. Many beautiful ideas in mathematics have come from studying the osculating circles to a curve.

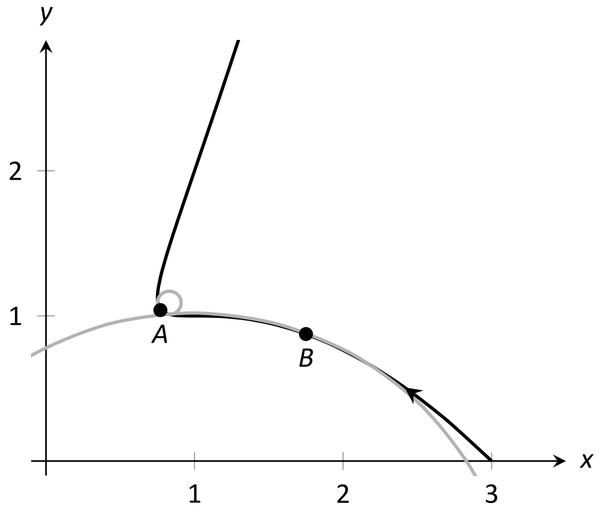


Figure 14.16: Illustrating the osculating circles for the curve seen in Figure 14.15(a).

Example 14.15

Find where the curvature of $\vec{r}(t) = \left\langle t, t^2, 2t^3 \right\rangle$ is maximized.

Solution

We use the third formula in Theorem 14.8 as $\vec{r}(t)$ is defined in space. We leave it to the reader to verify that

$$\vec{r}\,'(t) = \left\langle 1, 2t, 6t^2 \right\rangle, \quad \vec{r}\,''(t) = \langle 0, 2, 12t \rangle, \quad \text{and} \quad \vec{r}\,'(t) \times \vec{r}\,''(t) = \left\langle 12t^2, -12t, 2 \right\rangle.$$

Thus

$$\begin{aligned}\kappa(t) &= \frac{\left\lVert \vec{r}\,'(t) \times \vec{r}\,''(t) \right\rVert}{\left\lVert \vec{r}\,'(t) \right\rVert^3} \\ &= \frac{\left\lVert \left\langle 12t^2, -12t, 2 \right\rangle \right\rVert}{\left\lVert \left\langle 1, 2t, 6t^2 \right\rangle \right\rVert^3} \\ &= \frac{\sqrt{144t^4 + 144t^2 + 4}}{\left(\sqrt{1 + 4t^2 + 36t^4}\right)^3}.\end{aligned}$$

While this is not a particularly nice formula, it does explicitly tell us what the curvature is at a given t value. To maximize $\kappa(t)$, we should solve $\kappa'(t) = 0$ for t. This is doable, but time consuming. Instead, consider the graph of $\kappa(t)$ as given in Figure 14.17(a). We see that κ is maximized at two t values; using a numerical solver, we find these values are $t \approx \pm 0.189$. In Figure 14.17(b) we graph $\vec{r}(t)$ and indicate the points where curvature is maximized.

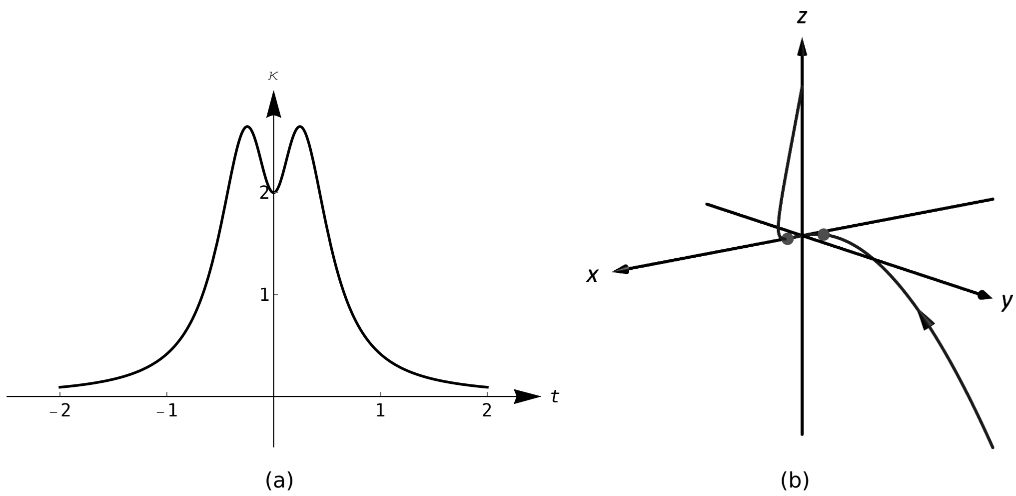


Figure 14.17: Understanding the curvature of a curve in space.

We started this chapter with vector–valued functions, which may have seemed at the time to be just another way of writing parametric equations. However, we have seen that the vector perspective has given us great insight into the behaviour of functions and the study of motion. Vector–valued position functions convey displacement, distance travelled, speed, velocity, acceleration and curvature information, each of which has great importance in science and engineering.

14.5 Exercices

Algebra of vector–valued functions

Assignment 14.1 — Show that the vector functions below all represent the same curve. Which curve is being described here?

$$\vec{r}_1(t) = \big(\sin(t), \cos(t)\big) \qquad \text{with} \quad -\frac{\pi}{2} \le t \le \frac{\pi}{2}$$

$$\vec{r}_2(t) = \left(t-1, \sqrt{2t-t^2}\right) \qquad \text{with} \quad 0 \le t \le 2$$

$$\vec{r}_3(t) = \left(t\sqrt{2-t^2}, 1-t^2\right) \qquad \text{with} \quad -1 \le t \le 1$$

Calculus and vector–valued functions

Assignment 14.2 — Find the limits below.

(a) $\displaystyle\lim_{t\to 5}\big(2t+1, 3t^2-1, \sin(t)\big)$

(b) $\displaystyle\lim_{t\to 3}\left(e^t, \frac{t^2-9}{t+3}\right)$

(c) $\displaystyle\lim_{t\to 0}\left(\frac{t}{\sin(t)}, (1+t)^{\frac{1}{t}}\right)$

(d) $\displaystyle\lim_{h\to 0}\frac{\vec{r}\,(t+h)-\vec{r}\,(t)}{h} \qquad \text{with} \quad \vec{r}(t) = \left(t^2, t, 1\right)$

Assignment 14.3 — Find the derivative of the vector functions below.

(a) $\vec{r}(t) = \left(\dfrac{1}{t}, \dfrac{2t-1}{3t+1}, \tan(t)\right)$

(b) $\vec{r}(t) = \left(t^2\right)\big(\sin(t), 2t+5\big)$

(c) $\vec{r}(t) = \left(t^2+1, t-1\right)\cdot\big(\sin(t), 2t+5\big)$

(d) $\vec{r}(t) = \left(t^2+1, t-1, 1\right)\times\big(\sin(t), 2t+5, 1\big)$

Assignment 14.4 — Determine the values of t at which $\vec{r}(t)$ is not smooth.

(a) $\vec{r}(t) = \big(\cos(t), \sin(t)-t\big)$

(b) $\vec{r}(t) = \left(t^2-2t+1, t^3+t^2-5t+3\right)$

(c) $\vec{r}(t) = \big(\cos(t)-\sin(t), \sin(t)-\cos(t), \cos(4t)\big)$

(d) $\vec{r}(t) = \left(t^3 - 3t + 2, -\cos(\pi t), \sin^2(\pi t)\right)$

Assignment 14.5 — Evaluate the integrals below.

(a) $\displaystyle\int \left(t^3, \cos(t), te^t\right)\, dt$

(b) $\displaystyle\int \left(\frac{1}{1+t^2}, \sec^2(t)\right)\, dt$

(c) $\displaystyle\int \left(\cos(t)e^{\sin(t)}, t\sin^2(t), -1\right)\, dt$

(d) $\displaystyle\int_0^{\pi} \left(\sin^2(t)\cos(t), \cos^2(t)\sin(t)\right)\, dt$

(e) $\displaystyle\int_0^{1} \left(\frac{1}{2}e^{-\frac{t}{2}}, \frac{1}{2}e^{\frac{t}{2}}, e^t\right)\, dt$

The calculus of motion

Assignment 14.6 — Determine the location vector of an object if the acceleration and initial velocity and position are given.

(a) $\vec{a}(t) = (2,3)$, $\vec{v}(1) = (1,2)$, $\vec{r}(1) = (5,-2)$

(b) $\vec{a}(t) = \left(\cos(t), -\sin(t)\right)$ $\vec{v}(0) = (0,1)$, $\vec{r}(0) = (0,0)$

(c) $\vec{a}(t) = (0,-32)$, $\vec{v}(0) = (10,50)$, $\vec{r}(0) = (0,0)$

Assignment 14.7 — Determine the velocity vector $\vec{v}(t)$, the velocity $\left\|\vec{v}(t)\right\|$ and the acceleration $\vec{a}(t)$ at time t of the object with position function $\vec{r}(t)$. Also describe the path followed by the object.

(a) $\vec{r}(t) = (1,t)$

(b) $\vec{r}(t) = \left(0, t^2, t\right)$

(c) $\vec{r}(t) = (1,t,t)$

(d) $\vec{r}(t) = \left(t^2, -t^2, 1\right)$

(e) $\vec{r}(t) = \left(3\cos(t), 4\cos(t), 5\sin(t)\right)$

(f) $\vec{r}(t) = \left(3\cos(t), 4\sin(t), t\right)$

Assignment 14.8 — An object moves at a constant speed of 5 to the right along the parabola $y = x^2$. Determine the velocity vector $\vec{v}(t)$ and the acceleration $\vec{a}(t)$ of the object at $(1,1)$.

Unit tangent and normal vectors

Assignment 14.9 — Find the unit tangent vector $\widehat{\boldsymbol{T}}(t)$ of the curves below.

(a) $\vec{\boldsymbol{r}}(t) = \left(2t^2, t^2 - t\right)$

(b) $\vec{\boldsymbol{r}}(t) = \left(t, -2t^2, 3t^3\right)$

(c) $\vec{\boldsymbol{r}}(t) = \left(t, \dfrac{t^2}{2}, \dfrac{t^3}{3}\right)$

(d) $\vec{\boldsymbol{r}}(t) = \left(\cos(t)\sin(t), \sin^2(t), \cos(t)\right)$

(e) $\vec{\boldsymbol{r}}(t) = \left(\dfrac{\cos^3(t)}{3}, \dfrac{\sin^3(t)}{3}\right)$ in $t = \dfrac{\pi}{6}$

Assignment 14.10 — Find the unite normal vector $\widehat{\boldsymbol{N}}(t)$ of the curves below.

(a) $\vec{\boldsymbol{r}}(t) = \left(\dfrac{t^3}{3} - t, t^2\right)$ in $t = 3$

(b) $\vec{\boldsymbol{r}}(t) = \left(3\cos(t), 3\sin(t)\right)$

(c) $\vec{\boldsymbol{r}}(t) = \left(e^t, e^{-t}\right)$

(d) $\vec{\boldsymbol{r}}(t) = \left(4t, 2\sin(t), 2\cos(t)\right)$

(e) $\vec{\boldsymbol{r}}(t) = \left(\dfrac{\cos^3(t)}{3}, \dfrac{\sin^3(t)}{3}\right)$ in $t = \dfrac{\pi}{6}$

Arc length and curvature

Assignment 14.11 — Determine the arc length of the curves below between the indicated points.

(a) $\vec{\boldsymbol{r}}(t) = \left(t^2, t^2, t^3\right)$ between $t = 0$ and $t = 1$

(b) $\vec{\boldsymbol{r}}(t) = \left(e^t\cos(t), e^t\sin(t), t\right)$ between $t = 0$ and $t = 2\pi$

(c) $\vec{\boldsymbol{r}}(t) = \left(t^3, t^2\right)$ between $t = -1$ and $t = 2$

Assignment 14.12 — Parameterize the vector functions below with the arc length parameter s starting from the point where $t = 0$.

(a) $\vec{\boldsymbol{r}}(t) = (2t, t, -2t)$

(b) $\vec{\boldsymbol{r}}(t) = \left(7\cos(t), 7\sin(t)\right)$

(c) $\vec{\boldsymbol{r}}(t) = \left(3\cos(t), 3\sin(t), 2t\right)$

(d) $\vec{\boldsymbol{r}}(t) = \left(e^t, \sqrt{2}t, -e^{-t}\right)$

Assignment 14.13 — Determine the radius of curvature r of the curves below at the indicated points.

(a) $y = x^2$ at $x = 0$ and $x = \sqrt{2}$

(b) $y = \cos(x)$ at $x = 0$ en $x = \pi/2$

(c) $y = \tan(x)$ at $x = \pi/4$

(d) $\vec{\boldsymbol{r}}(t) = \left(2t, \dfrac{1}{t}, -2t\right)$ at $(2, 1, -2)$

(e) $\vec{\boldsymbol{r}}(t) = \left(t^3, t^2, t\right)$ at $t = 1$

(f) $16y^2 = 4x^4 - x^6$ at $x = 2$

(g) $\vec{\boldsymbol{r}}(t) = \left(3t^2, 3t - t^3\right)$ at $t = 1$

(h) $\vec{\boldsymbol{r}}(t) = \left(\cos(t), \sin(3t)\right)$ at $t = 0$

Assignment 14.14 — Determine thecurvature κ and the radius of curvature r at a generic point on the given curve.

(a) $y = \dfrac{1}{x^2+1}$

(b) $y = \sqrt{1-x^2}$

(c) $x(t) = 2+\sqrt{2}\cos(t), \quad y(t) = 1-\sin(t), \quad z(t) = 3+\sin(t)$

(d) $y = e^x$

(e) $r(\theta) = a(1-\cos(\theta))$

(f) $\vec{r}(t) = \big(2\cos(t), \sin(t)\big)$

(g) $\vec{r}(t) = \big(t, \ln(\sin(t))\big) \qquad$ with $\quad 0 < t < \pi$

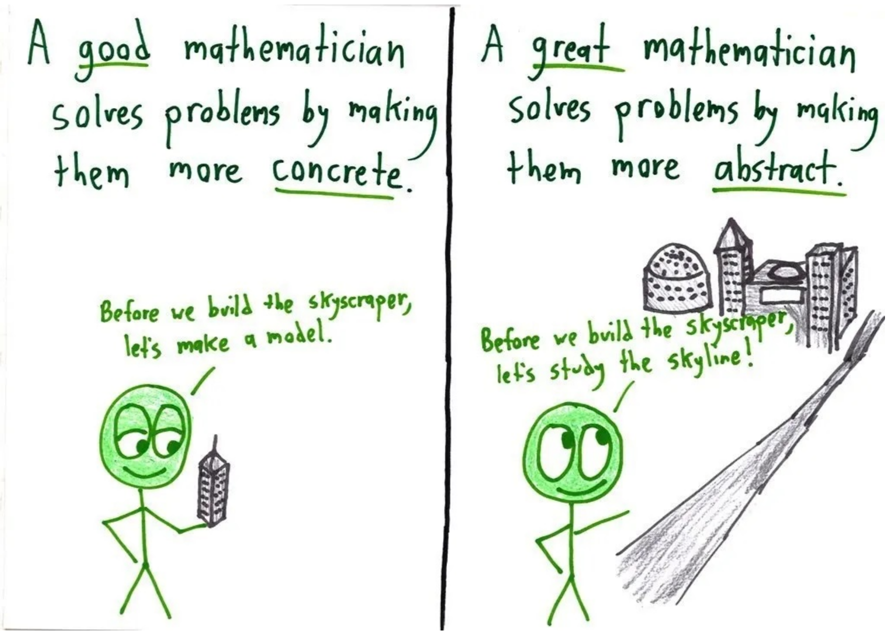
Review exercises

Assignment 14.15 — Determine the requested parametrization of the circle $x^2+y^2=a^2$ in the first quadrant.

(a) in terms of the y-coordinate, counterclockwise orientation

(b) in terms of the angle between the tangent at a point (x, y) and the positive x-axis, counterclockwise orientation

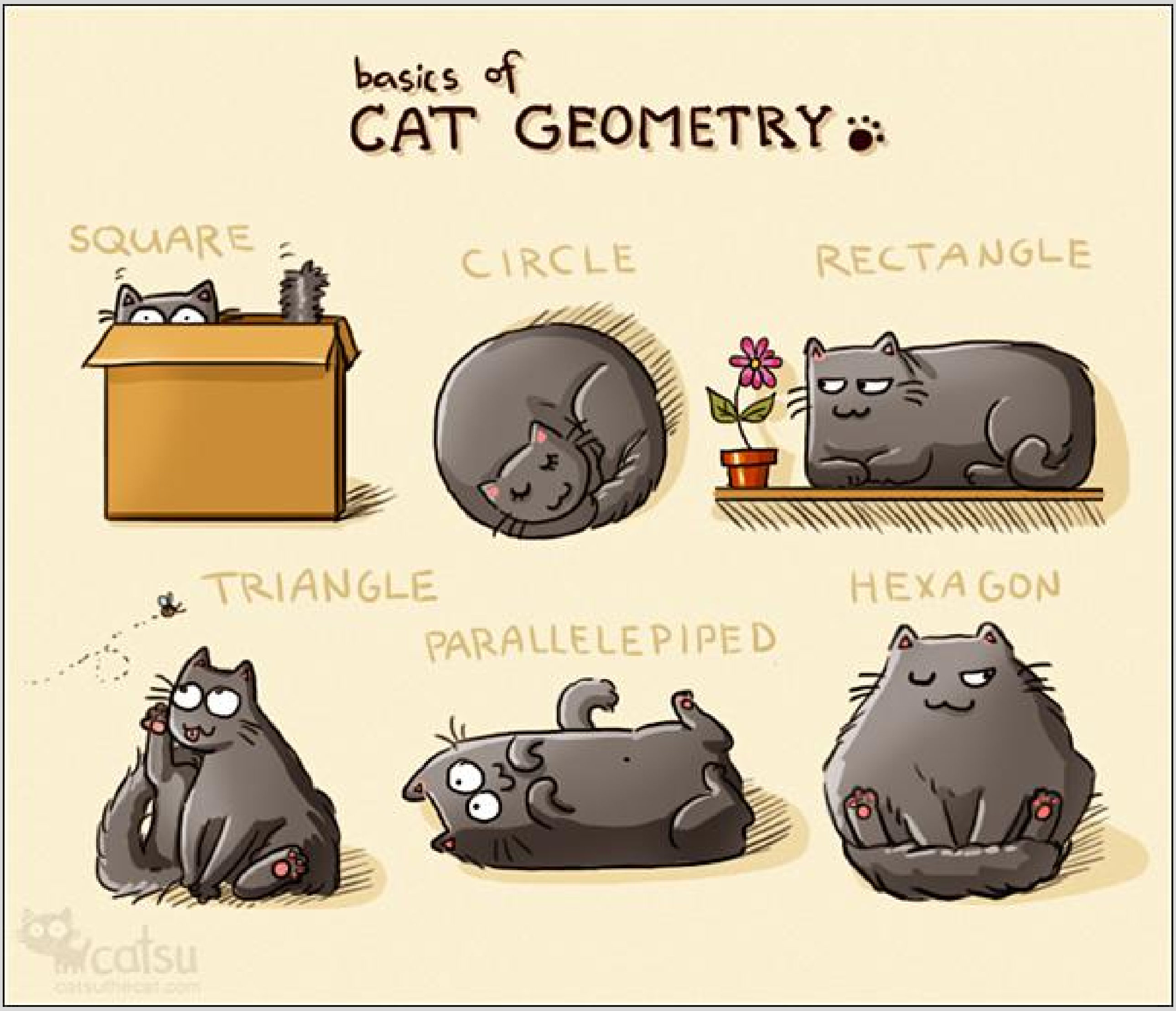
(c) in terms of the arc length measured from $(0, a)$, clockwise orientation



From *Math with Bad Drawings*, used by permission of Ben Orlin.

PART III

MULTIVARIABLE CALCULUS



As you will find in multivariable calculus, there is often a number of solutions for any given problem.

— John Nash —

15

Functions of several variables

A function of the form $y = f(x)$ is a function of a single variable; given a value of x, we can find a value y. Even the vector–valued functions of Chapter 14 are single–variable functions; the input is a single variable though the output is a vector.

There are many situations where a desired quantity is a function of two or more variables. For instance, wind chill is measured by knowing the temperature and wind speed; the volume of a gas can be computed knowing the pressure and temperature of the gas; to compute a baseball player's batting average, one needs to know the number of hits and the number of at–bats.

This chapter studies multivariable functions, that is, functions with more than one input.

15.1 Introduction to multivariable functions

15.1.1 Functions of two variables

We start with a definition of a function of two variables.

Definitie 15.1 (Function of two variables)
Let D be a subset of $\mathbb{R}^2$. A **function f of two variables** (*functie van twee veranderlijken*) is a rule that assigns each pair (x, y) in D a value $z = f(x, y)$ in $\mathbb{R}$. D is the domain of f; the set of all outputs of f is the range.

Example 15.1
Let

$$f(x,y) = \sqrt{1 - \frac{x^2}{9} - \frac{y^2}{4}}.$$

Find the domain and range of f.

Solution

The domain is all pairs (x,y) allowable as input in f. Because of the square root, we need (x,y) such that:

$$\begin{aligned} & 1-\frac{x^2}{9}-\frac{y^2}{4} \geq 0 \\ \Leftrightarrow \quad & \frac{x^2}{9}+\frac{y^2}{4} \leq 1 \end{aligned}$$

The above equation describes an ellipse and its interior. We can represent the domain D in set notation as

$$D=\left\{(x,y) \;\middle|\; \frac{x^2}{9}+\frac{y^2}{4} \leq 1\right\}.$$

The range is the set of all possible output values. The square root ensures that all output is positive. Since the x and y terms are squared, then subtracted, inside the square root, the largest output value comes at $x=0$, $y=0$: $f(0,0)=1$. Thus the range R is the interval $[0,1]$.

Definitie 15.2 (Graph of a function of two variables)
The **graph** of a function f of two variables is the set of all points $\big(x,y,f(x,y)\big)$ where (x,y) is in the domain of f. This creates a **surface** (*oppervlak*) in space.

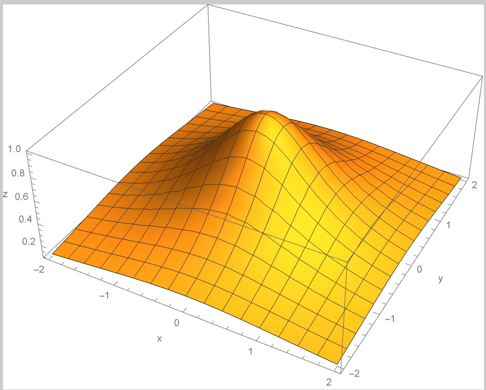
One can begin sketching a graph by plotting points, but this has limitations. Consider Figure 15.1(a) where 25 points have been plotted of

$$f(x,y)=\frac{1}{x^2+y^2+1}.$$

More points have been plotted than one would reasonably want to do by hand, yet it is not clear at all what the graph of the function looks like. Technology allows us to plot lots of points, connect adjacent points with lines and add shading to create a graph like Figure 15.1(b) which does a far better job of illustrating the behaviour of f. More specifically, in Mathematica, a function of two variables can be plotted using the command **Plot3D** as follows

```
In[23]:= Plot3D[{1/(x^2 + y^2 + 1)}, {x, -2, 2}, {y, -2, 2},  AxesLabel → {"x", "y", "z"}]
```

Out[23]=



Of course, many options are available to format such graphs according to one's preference. These can be checked in the Documentation Center.

While technology is readily available to help us graph functions of two variables, there is still a paper–and–pencil approach that is useful to understand and master as it, combined with high–quality graph-

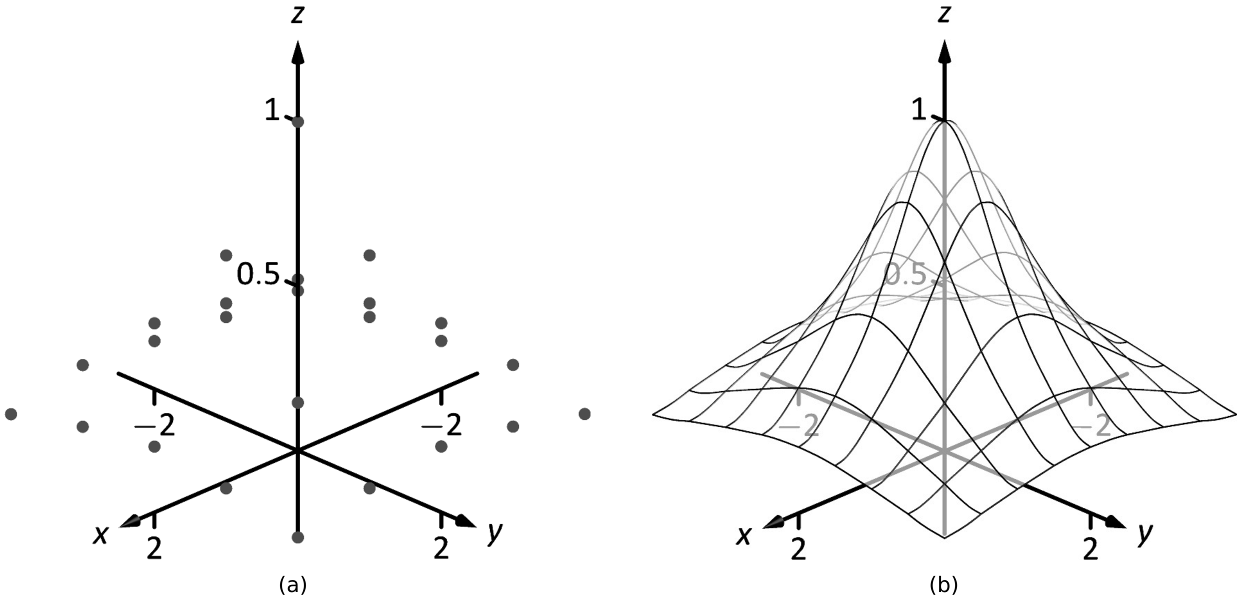


Figure 15.1: Graphing a function of two variables.

ics, gives one great insight into the behaviour of a function. This technique is known as sketching **level curves** (*niveaukromme*).

It may be surprising to find that the problem of representing a three dimensional surface on paper is familiar to most people. Topographical maps, like the one of Dinant shown in Figure 15.2, represent the surface of Earth by indicating points with the same elevation with **contour lines** (*countourlijn*). The elevations marked are equally spaced; in this example, each thin line indicates an elevation change in 10m increments and each thick line indicates a change of 50m. When lines are drawn close together, elevation changes rapidly. When lines are far apart, elevation changes more gradually as one has to walk farther to rise 10m.

Given a function $z = f(x, y)$, we can draw a topographical map of f by drawing **level curves** (or, contour lines). A level curve at $z = c$ is a curve in the xy-plane such that for all points (x, y) on the curve, $f(x, y) = c$. When drawing level curves, it is important that the c-values are spaced equally apart as that gives the best insight to how quickly the elevation is changing.

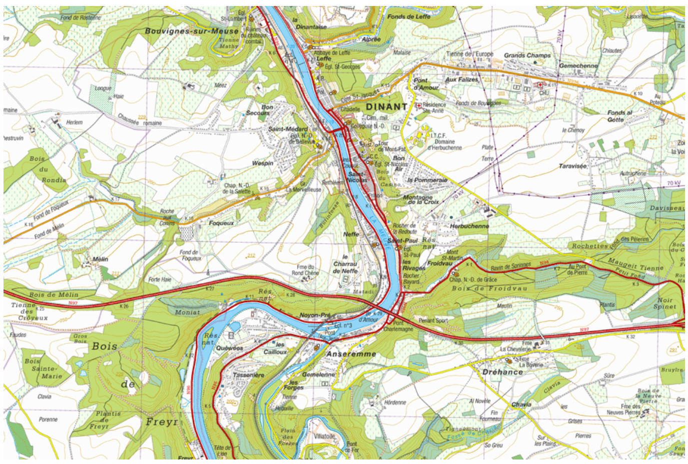


Figure 15.2: The topographical map of Dinant displays elevation by drawing contour lines, along which the elevation is constant.

Example 15.2

Let

$$f(x,y) = \frac{x+y}{x^2+y^2+1}.$$

Find level curves.

Solution

We begin by setting $f(x,y) = c$ for an arbitrary c and seeing if algebraic manipulation of the equation reveals anything significant.

$$\frac{x+y}{x^2+y^2+1} = c \qquad \Leftrightarrow \qquad x^2 - \frac{1}{c}x + y^2 - \frac{1}{c}y = -1.$$

We recognize this as a circle, though the centre and radius are not yet clear. By completing the square, we obtain:

$$\left(x - \frac{1}{2c}\right)^2 + \left(y - \frac{1}{2c}\right)^2 = \frac{1}{2c^2} - 1$$

a circle centred at $\big(1/(2c), 1/(2c)\big)$ with radius $\sqrt{1/(2c^2)-1}$, where $|c| < 1/\sqrt{2}$. The level curves for $c = \pm 0.2$, ± 0.4 and ± 0.6 are sketched in Figure 15.3(a). To help illustrate elevation, we use dashed lines where $c < 0$. There is one special level curve, when $c = 0$. The level curve in this situation is $x + y = 0$, the line $y = -x$.

In Figure 15.3(b) we see a graph of the surface. Note how the y-axis is pointing away from the viewer to more closely resemble the orientation of the level curves in Figure 15.3(a). Seeing the level curves helps us understand the graph. For instance, the graph does not make it clear that one can walk along the line $y = -x$ without elevation change, though the level curve does.

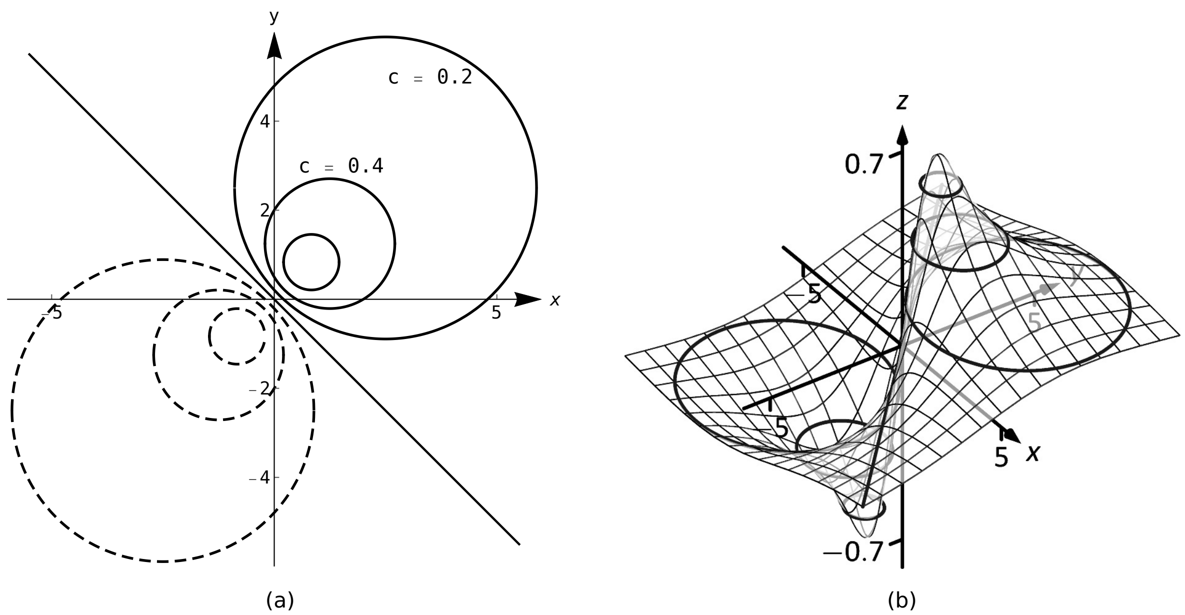


Figure 15.3: Graphing the level curves in Example 15.2.

The contour lines are established as intersection between the surface defined by $z = f(x,y)$ with the horizontal plane $z = c$. However, there are two more special types of planes with which we can intersect the surface and which improve our understanding of functions of two variables, namely $x = x_0$ and $y = y_0$. These are planes perpendicular to the x- or y-axis. The curves of intersection that we thus obtain are nothing more than the graphs of the so-called partial functions to y or x.

Definitie 15.3 (Partial functions)
Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables, then

- the partial function of f with regard to x, or the first partial function is given by:

$$f_1 : \mathbb{R} \to \mathbb{R} : x \to z = f_1(x) = f(x, y_0)\,,$$

  with y_0 constant;

- the partial function of f with regard to y, or the second partial function is given by:

$$f_2 : \mathbb{R} \to \mathbb{R} : x \to z = f_2(y) = f(x_0, y)\,,$$

  with x_0 constant.

15.1.2 Functions of three variables

We extend our study of multivariable functions to functions of three variables. One can make a function of as many variables as one likes; we limit our study to three variables.

Definitie 15.4 (Function of three variables)
Let D be a subset of $\mathbb{R}^3$. A **function f of three variables** (*functie van drie veranderlijken*) is a rule that assigns each triple (x, y, z) in D a value $w = f(x, y, z)$ in $\mathbb{R}$. D is the domain of f; the set of all outputs of f is the range.

It is very difficult to produce a meaningful graph of a function of three variables. A function of one variable is a curve drawn in 2 dimensions; a function of two variables is a surface drawn in 3 dimensions; a function of three variables is a **hypersurface** (*hyperoppervlak*) drawn in 4 dimensions.

There are a few techniques one can employ to try to picture a graph of three variables. One is an analogue of level curves: **level surfaces** (*niveau-oppervlak*). Given $w = f(x, y, z)$, the level surface at $w = c$ is the surface in space formed by all points (x, y, z) where $f(x, y, z) = c$.

Example 15.3

If a point source S is radiating energy, the intensity I at a given point P in space is inversely proportional to the square of the distance between S and P. That is, when $S = (0, 0, 0)$,

$$I(x, y, z) = \frac{k}{x^2 + y^2 + z^2}$$

for some constant k.

Let $k = 1$; find the level surfaces of I.

Solution

We can answer this question using common sense. If energy is emanating from the origin, its intensity will be the same at all points equidistant from the origin. That is, at any point on the surface of a sphere centred at the origin, the intensity should be the same. Therefore, the level surfaces are spheres. We now find this mathematically. The level surface at $I = c$ for $c > 0$ is defined by

$$c = \frac{1}{x^2+y^2+z^2}.$$

A small amount of algebra reveals

$$x^2+y^2+z^2 = \frac{1}{c}.$$

Given an intensity c, the level surface $I = c$ is a sphere of radius $1/\sqrt{c}$, centred at the origin. Table 15.1 gives the radii of the spheres for given c-values. Normally one would use equally spaced c-values, but these values have been chosen purposefully. At a distance of 0.25 from the point source, the intensity is 16; to move to a point of half that intensity, one just moves out 0.1 to 0.35 – not much at all. To again halve the intensity, one moves 0.15, a little more than before.

Table 15.1: A table of c-values and corresponding radius r of the spheres of constant value in Example 15.3.

c	16.	8.	4.	2.	1.	0.5	0.25	0.125	0.0625
r	0.25	0.35	0.5	0.71	1	1.41	2	2.83	4

Note how each time the intensity is halved, the distance required to move away grows. We conclude that the closer one is to the source, the more rapidly the intensity changes.

15.2 Limits and continuity of multivariable functions

This section investigates what it means for multivariable functions to be continuous.

15.2.1 Introductory concepts and definitions

We begin with a series of definitions. We are used to open and closed intervals. We need analogous definitions for open and closed sets in the xy-plane.

Definitie 15.5 (Points and sets)
An **open disk** (*open schijf*) B in $\mathbb{R}^2$ centred at (x_0, y_0) with radius r is the set of all points (x, y) such that $\sqrt{(x-x_0)^2+(y-y_0)^2} < r$.

Let S be a set of points in $\mathbb{R}^2$. A point P in $\mathbb{R}^2$ is a **boundary point** (*randpunt*) of S if all open disks centred at P contain both points in S and points not in S.

A point P in S is an **interior point** (*inwendig punt*) of S if there is an open disk centred at P that contains only points in S.

A set S is **open** (*open*) if every point in S is an interior point.

A set S is **closed** (*gesloten*) if it contains all of its boundary points.

A set S is **bounded** (*begrensd*) if there is an $M > 0$ such that the open disk, centred at the origin with radius M, contains S. A set that is not bounded is unbounded.

A set S is **convex** (*convex*) if, given any two points, it contains the whole line segment that joins. A set that is not convex is called non-convex.

Figure 15.4 shows several sets in the xy-plane. In each set, point P_1 lies on the boundary of the set as all open disks centred there contain both points in, and not in, the set. In contrast, point P_2 is an interior point for there is an open disk centred there that lies entirely within the set. The set depicted in Figure 15.4(a) is a closed set as it contains all of its boundary points. The set in Figure 15.4(b) is open, for all of its points are interior points (or, equivalently, it does not contain any of its boundary points). The set in Figure 15.4(c) is neither open nor closed as it contains some of its boundary points. Finally, it should be clear that all sets shown in Figure 15.4 are non-convex because we can easily find pairs of points that can only be connected by a straight line that is not completely contained in the considered sets.

Example 15.4

Determine if the domain of the functions

1. $f(x,y) = \sqrt{1 - \frac{x^2}{9} - \frac{y^2}{4}}$
2. $g(x,y) = \frac{1}{x-y}$

is open, closed, or neither, and if it is bounded.

Solution

1. The domain of this function was found in Example 15.1 to be

$$D = \left\{ (x,y) \;\middle|\; \frac{x^2}{9} + \frac{y^2}{4} \leq 1 \right\},$$

   the region bounded by the ellipse $\frac{x^2}{9} + \frac{y^2}{4} = 1$. Since the region includes the boundary, the set contains all of its boundary points and hence is closed. The region is bounded as a disk of radius 4, centred at the origin, contains D.

2. As we cannot divide by 0, we find the domain to be $D = \{(x,y) \mid x-y \neq 0\}$. In other words, the domain is the set of all points (x,y) not on the line $y = x$. The domain is sketched in Figure 15.5. Note how we can draw an open disk around any point in the domain that lies entirely inside the domain, and also note how the only boundary points of the domain are the points on the line $y = x$. We conclude the domain is an open set. The set is unbounded.

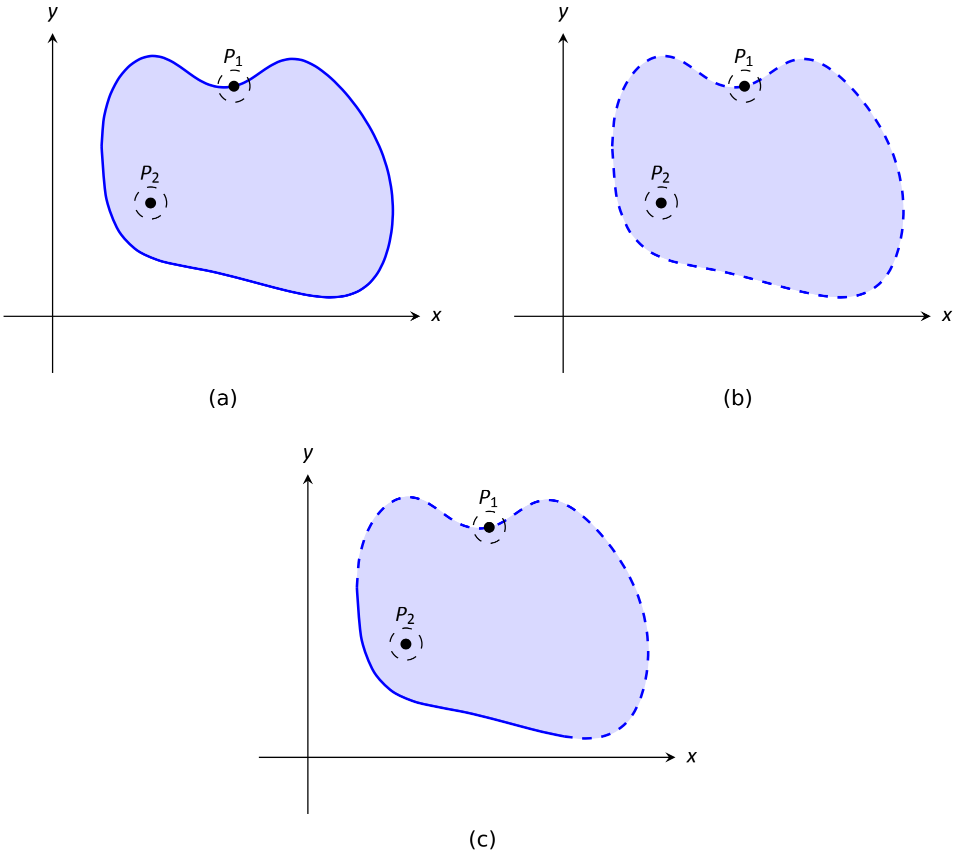


Figure 15.4: Illustrating open and closed sets in the xy-plane.

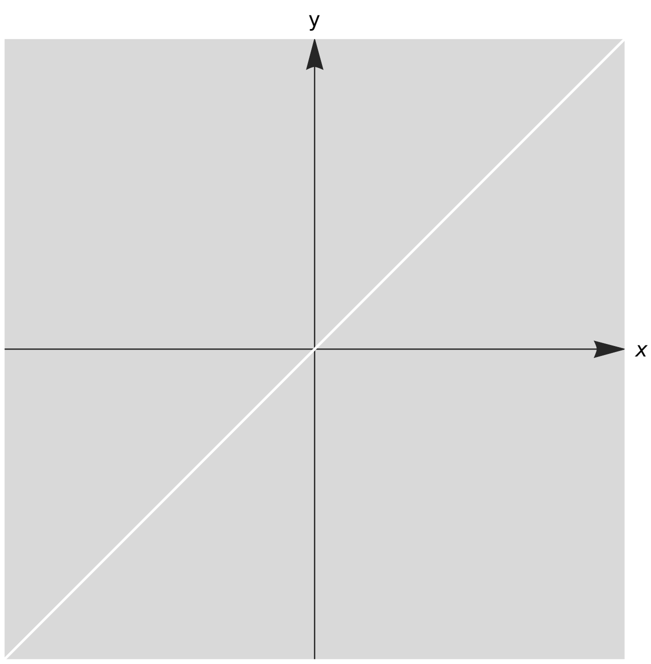


Figure 15.5: Sketching the domain of the function in Example 15.4.2.

15.2.2 Limits

We will say that

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L$$

means if the point (x,y) is really close to the point (x_0,y_0), then $f(x,y)$ is really close to L. The formal definition is given below.

Definitie 15.6 (Limit of a function of two variables)
Let S be a set containing $P = (x_0,y_0)$ where every open disk centred at P contains points in S other than P, i.e. P is a limit point, let f be a function of two variables defined on S, except possibly at P, and let L be a real number. The **limit of** $f(x,y)$ **as** (x,y) **approaches** (x_0,y_0) is L, denoted

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L,$$

means that given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all (x,y) in S, where $(x,y) \neq (x_0,y_0)$, if (x,y) is in the open disk centred at (x_0,y_0) with radius δ, then $|f(x,y) - L| < \varepsilon$.

Note that we now define limits over a set S in the plane (where S does not have to be open). As planar sets can be far more complicated than intervals, our definition adds the restriction "... where every open disk centred at P contains points in S other than P." This means that P should be a so-called **limit point** (*ophopingspunt*) of the set S. This in contrast to a so-called **isolated point** (*gëisoleerd punt*) Q of S for which there exists a neighbourhood of Q which does not contain any other points of S.

The concept behind Definition 15.6 is sketched in Figure 15.6. Given $\varepsilon > 0$, find $\delta > 0$ such that if (x,y) is any point in the open disk centred at (x_0,y_0) in the xy-plane with radius δ, then $f(x,y)$ should be within ε of L.

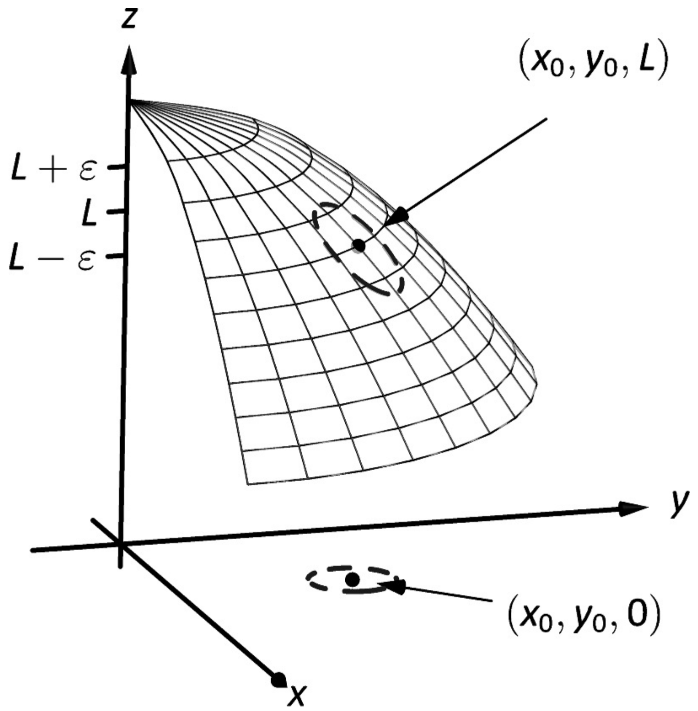


Figure 15.6: Illustrating the definition of a limit of a function of two variables.

Computing limits using this definition is rather cumbersome. The following properties allow us to evaluate limits much more easily. For that purpose, let b, x_0, y_0, L and K be real numbers, let n be a positive integer, and let f and g be functions with the following limits:

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L \quad \text{and} \quad \lim_{(x,y)\to(x_0,y_0)} g(x,y) = K.$$

The following limits hold.

- **Constants:**
$$\lim_{(x,y)\to(x_0,y_0)} b = b$$
- **Identity**
$$\lim_{(x,y)\to(x_0,y_0)} x = x_0; \qquad \lim_{(x,y)\to(x_0,y_0)} y = y_0$$
- **Sums/Differences:**
$$\lim_{(x,y)\to(x_0,y_0)} \big(f(x,y) \pm g(x,y)\big) = L \pm K$$
- **Scalar Multiples:**
$$\lim_{(x,y)\to(x_0,y_0)} b \cdot f(x,y) = bL$$
- **Products:**
$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) \cdot g(x,y) = LK$$
- **Quotients:**
$$\lim_{(x,y)\to(x_0,y_0)} f(x,y)/g(x,y) = L/K, \qquad (K \neq 0)$$
- **Powers:**
$$\lim_{(x,y)\to(x_0,y_0)} \big[f(x,y)\big]^n = L^n$$

These properties, combined with the ones we introduced in Chapter 8, allow us to evaluate many limits. For instance, we can easily evaluate

$$\lim_{(x,y)\to(1,\pi)} \left(\frac{y}{x} + \cos(xy)\right) = \frac{\pi}{1} + \cos(\pi) = \pi - 1.$$

This limit may as well be evaluated in Mathematica with a nested application of the command **`Limit`**.

```
In[24]:= Limit[Limit[y/x + Cos[x*y], x → 1], y → Pi]

Out[24]= -1+π
```

When dealing with functions of a single variable we also considered one–sided limits and stated

$$\lim_{x\to c} f(x) = L$$

if, and only if

$$\lim_{\substack{x\to c\\ >}} f(x) = L \quad \text{and} \quad \lim_{\substack{x\to c\\ <}} f(x) = L.$$

That is, the limit is L if and only if $f(x)$ approaches L when x approaches c from either direction, the left or the right.

In the plane, there are infinitely many directions from which (x,y) might approach (x_0,y_0). In fact, we do not have to restrict ourselves to approaching (x_0,y_0) from a particular direction, but rather we can approach that point along a path that is not a straight line. It is possible to arrive at different limiting values by approaching (x_0,y_0) along different paths. If this happens, we say that $\lim_{(x,y)\to(x_0,y_0)} f(x,y)$ does not exist. This is analogous to the left and right hand limits of single variable functions not being equal.

Our theorems tell us that we can evaluate most limits quite simply, without worrying about paths. When indeterminate forms arise, the limit may or may not exist. If it does exist, it can be difficult to prove this as we need to show the same limiting value is obtained regardless of the path chosen. The case where the limit does not exist is often easier to deal with, for we can often pick two paths along which the limit is different.

Example 15.5

1. Show
$$\lim_{(x,y)\to(0,0)} \frac{3xy}{x^2+y^2}$$
does not exist by finding the limits along the lines $y = mx$.

2. Show
$$\lim_{(x,y)\to(0,0)} \frac{\sin(xy)}{x+y}$$
does not exist by finding the limit along the path $y = -\sin(x)$.

Solution

1. Evaluating this limit along the lines $y = mx$ means replace all y's with mx and evaluating the resulting limit:

$$\begin{aligned}\lim_{(x,mx)\to(0,0)} \frac{3x(mx)}{x^2+(mx)^2} &= \lim_{x\to 0} \frac{3mx^2}{x^2(m^2+1)}\\ &= \lim_{x\to 0} \frac{3m}{m^2+1}\\ &= \frac{3m}{m^2+1}.\end{aligned}$$

While the limit exists for each choice of m, we get a different limit for each choice of m. That is, along different lines we get differing limiting values, meaning the limit does not exist.

2. We are to show that $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist by finding the limit along the path $y = -\sin(x)$. First, however, consider the limits found along the lines $y = mx$ as done above.

$$\begin{aligned}\lim_{(x,mx)\to(0,0)} \frac{\sin\big(x(mx)\big)}{x+mx} &= \lim_{x\to 0} \frac{\sin(mx^2)}{x(m+1)}\\ &= \lim_{x\to 0} \frac{\sin(mx^2)}{x}\cdot\frac{1}{m+1}.\end{aligned}$$

By applying L'Hôpital's rule, we can show this limit is 0 except when $m = -1$, that is, along the line $y = -x$. This line is not in the domain of f, so we have found the following fact: along every line $y = mx$ in the domain of f,

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0.$$

Now consider the limit along the path $y = -\sin(x)$:

$$\lim_{(x,-\sin(x))\to(0,0)} \frac{\sin\big(-x\sin x\big)}{x-\sin(x)} = \lim_{x\to 0} \frac{\sin\big(-x\sin x\big)}{x-\sin(x)}.$$

Now apply L'Hôpital's rule twice:

$$\begin{aligned}&= \lim_{x\to 0} \frac{\cos\big(-x\sin(x)\big)(-\sin(x)-x\cos(x))}{1-\cos(x)} \qquad \left(=\frac{0}{0}\right)\\ &= \lim_{x\to 0} \frac{-\sin\big(-x\sin(x)\big)(-\sin(x)-x\cos(x))^2+\cos\big(-x\sin(x)\big)(-2\cos(x)+x\sin(x))}{\sin(x)}\\ &= \frac{-2}{0}.\end{aligned}$$

It follows that the limit does not exist. Step back and consider what we have just discovered. Along any line $y = mx$ in the domain of the $f(x,y)$, the limit is 0. However, along the path $y = -\sin(x)$, which lies in the domain of $f(x,y)$ for all $x \neq 0$, the limit does not exist. Since the limit is not the same along every path to $(0,0)$, we say that the studied limit does not exist.

15.2.3 Continuity

Definition 8.3 defines what it means for a function of one variable to be continuous. In brief, it meant that the graph of the function did not have breaks, holes, jumps, etc. We define continuity for functions of two variables in a similar way as we did for functions of one variable.

Definitie 15.7 (Continuity)
Let a function $f(x,y)$ be defined on a set S containing the point (x_0,y_0).

1. f is continuous at (x_0,y_0) if $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y) = f(x_0,y_0)$.
2. f is **continuous on** S (*continu over*) if f is continuous at all points in S. If f is continuous at all points in $\mathbb{R}^2$, we say that f is continuous everywhere.

Example 15.6

Let

$$f(x,y) = \begin{cases} \dfrac{\cos(y)\sin(x)}{x}, & x \neq 0 \\ \cos(y), & x = 0. \end{cases}$$

Is f continuous at $(0,0)$? Is f continuous everywhere?

Solution

To determine if f is continuous at $(0,0)$, we need to compare $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ to $f(0,0)$. Applying the definition of f, we see that $f(0,0) = \cos(0) = 1$.

We now consider the limit $\lim\limits_{(x,y)\to(0,0)} f(x,y)$. Substituting 0 for x and y in $f(x,y)$ returns the indeterminate form "0/0", so we need to do more work to evaluate this limit.

Consider two related limits:

$$\lim_{(x,y)\to(0,0)} \cos(y) \qquad \text{and} \qquad \lim_{(x,y)\to(0,0)} \frac{\sin(x)}{x}.$$

The first limit does not contain x, and since $\cos(y)$ is continuous,

$$\lim_{(x,y)\to(0,0)} \cos(y) = \lim_{y\to 0} \cos(y) = \cos(0) = 1.$$

The second limit does not contain y. By Theorem 8.4 we can say

$$\lim_{(x,y)\to(0,0)} \frac{\sin(x)}{x} = \lim_{x\to 0} \frac{\sin(x)}{x} = 1.$$

Finally, following the properties of limits we can combine these two limits as follows:

$$\begin{aligned} \lim_{(x,y)\to(0,0)} \frac{\cos(y)\sin(x)}{x} &= \lim_{(x,y)\to(0,0)} \left(\cos(y)\left(\frac{\sin(x)}{x}\right)\right) \\ &= \left(\lim_{(x,y)\to(0,0)} \cos(y)\right)\left(\lim_{(x,y)\to(0,0)} \frac{\sin(x)}{x}\right) \\ &= (1)(1) = 1. \end{aligned}$$

We have found that

$$\lim_{(x,y)\to(0,0)} \frac{\cos(y)\sin(x)}{x} = f(0,0),$$

so f is continuous at $(0,0)$.

A similar analysis shows that f is continuous at all points in $\mathbb{R}^2$. As long as $x \neq 0$, we can evaluate the limit directly; when $x = 0$, a similar analysis shows that the limit is $\cos(y)$. Thus we can say that f is continuous everywhere. A graph of f is given in Figure 15.7. Notice how it has no breaks, jumps, etc.

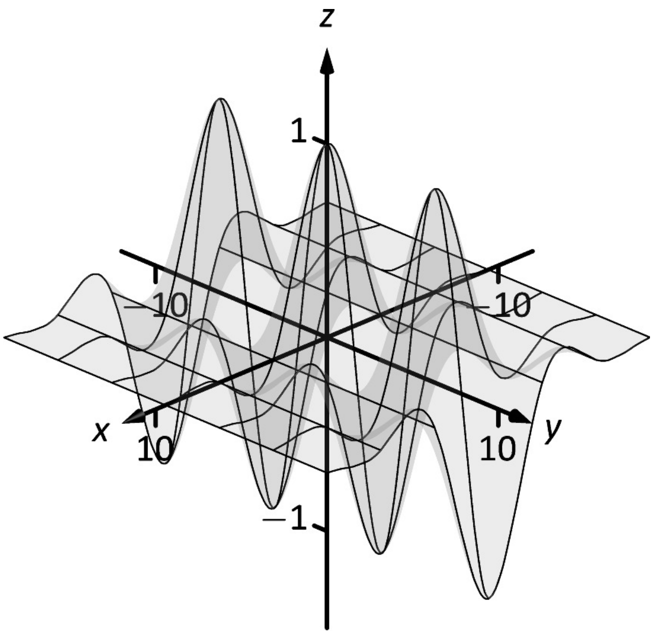


Figure 15.7: A graph of $f(x,y)$ in Example 15.6.

Of course, as with functions of one variable, we may combine continuous functions to create other continuous functions. More specifically, let f and g be continuous on a set S, let c be a real number, and let n be a positive integer. Then, the following functions are continuous on S.

- **Sums/Differences:** $f \pm g$
- **Constant multiples:** $c \cdot f$
- **Products:** $f \cdot g$
- **Quotients:** f/g (if $g \neq 0$ on S)
- **Powers:** f^n

For roots of a continuous function, we have that $\sqrt[n]{f}$ is continuous provided that $f \geq 0$ on S if n is even, whereas, if n is odd, this is true for all values of f on S. For what concerns function compositions, we let f be continuous on S, where the range of f on S is J, and let g be a single variable function that is continuous on J. Then $g \circ f$, i.e., $g(f(x,y))$, is continuous on S.

15.2.4 Functions of three variables

The definitions and theorems given in this section can be extended in a natural way to definitions and theorems about functions of three variables. We cover the key concepts here; some terms from Definitions 15.5 and 15.7 are not redefined but their analogous meanings should be clear to the reader.

Definitie 15.8 (Open balls, limit and continuity)

1. An **open ball** (*open bal*) in $\mathbb{R}^3$ centred at (x_0, y_0, z_0) with radius r is the set of all points (x, y, z) such that $\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2} = r$.

2. Let D be a set in $\mathbb{R}^3$ containing (x_0, y_0, z_0) where every open ball centred at (x_0, y_0, z_0) contains points of D other than (x_0, y_0, z_0), and let $f(x, y, z)$ be a function of three variables defined on D, except possibly at (x_0, y_0, z_0). The **limit of** $f(x, y, z)$ **as** (x, y, z) **approaches** (x_0, y_0, z_0) is L, denoted
$$\lim_{(x,y,z)\to(x_0,y_0,z_0)} f(x, y, z) = L,$$
means that given any $\varepsilon > 0$, there is a $\delta > 0$ such that for all (x, y, z) in D, $(x, y, z) \neq (x_0, y_0, z_0)$, if (x, y, z) is in the open ball centred at (x_0, y_0, z_0) with radius δ, then $|f(x, y, z) - L| < \varepsilon$.

3. Let $f(x, y, z)$ be defined on a set D containing (x_0, y_0, z_0). f is **continuous at** (x_0, y_0, z_0) if
$$\lim_{(x,y,z)\to(x_0,y_0,z_0)} f(x, y, z) = f(x_0, y_0, z_0)$$
; if f is continuous at all points in D, we say f is continuous on D.

When considering single variable functions, we studied limits, then continuity, then the derivative. In our current study of multivariable functions, we have studied limits and continuity. In the next section we study derivation, which takes on a slight twist as we are in a multivarible context.

15.3 Partial derivatives

15.3.1 First partial derivatives

Let y be a function of x. We have studied in great detail the derivative of y with respect to x, that is, which measures the rate at which y changes with respect to x. Consider now $z = f(x, y)$. It makes sense to want to know how z changes with respect to x and/or y. This section begins our investigation into these rates of change.

Consider the function $z = f(x, y) = x^2 + 2y^2$, as graphed in Figure 15.8(a). By fixing $y = 2$, we focus our attention to all points on the surface where the y-value is 2, shown in Figures 15.8(a) and 15.8(b). These points form a curve in space: $z = f(x, 2) = x^2 + 8$ which is a function of just one variable. We can take the derivative of z with respect to x along this curve and find equations of tangent lines, etc.

The key notion to extract from this example is: by treating y as constant (it does not vary) we can consider how z changes with respect to x. In a similar fashion, we can hold x constant and consider how z changes with respect to y. This is the underlying principle of **partial derivatives** (*partiële afgeleide*). We state the formal, limit–based definition first, then show how to compute these partial derivatives without directly taking limits.

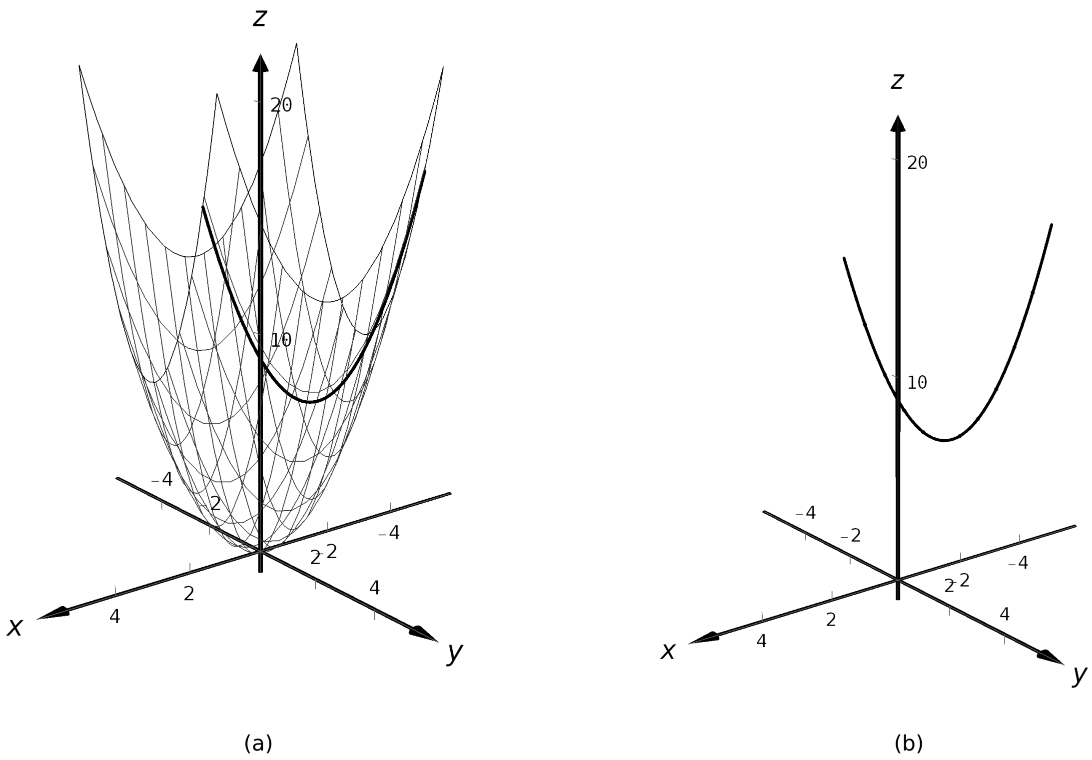


Figure 15.8: By fixing $y = 2$, the surface $f(x,y) = x^2 + 2y^2$ is a curve in space.

Definitie 15.9 (Partial derivative)
Let $z = f(x,y)$ be a continuous function on a set S in $\mathbb{R}^2$.

1. The **partial derivative of f with respect to x** is:

$$f_x(x,y) = \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h}.$$

2. The **partial derivative of f with respect to y** is:

$$f_y(x,y) = \lim_{h\to 0} \frac{f(x,y+h) - f(x,y)}{h}.$$

Alternate notations for $f_x(x,y)$ include:

$$\frac{\partial}{\partial x} f(x,y), \quad \frac{\partial f}{\partial x}, \quad \frac{\partial z}{\partial x}, \quad \text{and } z_x,$$

with similar notations for $f_y(x,y)$. For ease of notation, $f_x(x,y)$ is often abbreviated as f_x.

Using limits to compute partial derivatives is not necessary, as we can rely on our previous knowledge of derivatives to compute partial derivatives easily. When computing $f_x(x,y)$, we hold y fixed – it does not vary. Therefore we can compute the derivative with respect to x by treating y as a constant or coefficient.

Example 15.7

Find $f_x(x,y)$ and $f_y(x,y)$ for each of the following functions.

1. $f(x,y) = \cos(xy^2) + \sin(x)$
2. $f(x,y) = e^{x^2y^3}\sqrt{x^2+1}$

Solution

1. Begin with $f_x(x,y)$. We need to apply the chain rule with the cosine term; y^2 is the coefficient of the x-term inside the cosine function.

$$f_x(x,y) = -\sin(xy^2)(y^2) + \cos(x) = -y^2\sin(xy^2) + \cos(x).$$

To find $f_y(x,y)$, note that x is the coefficient of the y^2-term inside of the cosine term; also note that since x is fixed, $\sin(x)$ is also fixed, and we treat it as a constant.

$$f_y(x,y) = -\sin(xy^2)(2xy) = -2xy\sin(xy^2).$$

We may check our answer for what concerns f_x in Mathematica as follows:

```
In[25]:= D[Cos[x*y^2] + Sin[x], x]

Out[25]= Cos[x]-y^2 Sin[x y^2]
```

And likewise for what concerns f_y:

```
In[26]:= D[Cos[x*y^2] + Sin[x], y]

Out[26]= -2 x y Sin[x y^2]
```

2. Beginning with $f_x(x,y)$, note how we need to apply the product rule.

$$\begin{aligned} f_x(x,y) &= e^{x^2y^3}(2xy^3)\sqrt{x^2+1} + e^{x^2y^3}\frac{1}{2}\left(x^2+1\right)^{-1/2}(2x) \\ &= 2xy^3e^{x^2y^3}\sqrt{x^2+1} + \frac{xe^{x^2y^3}}{\sqrt{x^2+1}}. \end{aligned}$$

Note that when finding $f_y(x,y)$ we do not have to apply the product rule; since $\sqrt{x^2+1}$ does not contain y, we treat it as fixed and hence becomes a coefficient of the $e^{x^2y^3}$-term.

$$f_y(x,y) = e^{x^2y^3}(3x^2y^2)\sqrt{x^2+1} = 3x^2y^2e^{x^2y^3}\sqrt{x^2+1}.$$

We have shown how to compute a partial derivative, but it may still not be clear what a partial derivative means. Given $z = f(x,y)$, $f_x(x,y)$ measures the rate at which z changes as only x varies: y is held constant.

Imagine standing in a rolling meadow, then beginning to walk due east. Depending on your location, you might walk up, sharply down, or perhaps not change elevation at all. This is similar to measuring z_x: you are moving only east (in the x-direction) and not north/south at all. Going back to your original location, imagine now walking due north (in the y-direction). Perhaps walking due north does not change your elevation at all. This is analogous to $z_y = 0$: z does not change with respect to y. We can see that z_x and z_y do not have to be the same, or even similar, as it is easy to imagine circumstances where walking east means you walk downhill, though walking north makes you walk uphill.

15.3.2 Second partial derivatives

Let $z = f(x,y)$. We have learned to find the partial derivatives $f_x(x,y)$ and $f_y(x,y)$, which are each functions of x and y. Therefore we can take partial derivatives of them, each with respect to x and y. We define these second partials along with the notation, give examples, then discuss their meaning.

Definitie 15.10 (Second and mixed partial derivative)
Let $z = f(x,y)$ be continuous on a set S.

1. The **second partial derivative of f with respect to x then x** (*tweede partiële afgeleide van f naar x*) is

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = \left(f_x\right)_x = f_{xx}.$$

2. The **second partial derivative of f with respect to x then y is** (*tweede partiële afgeleide van f naar x en y*) is

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} = \left(f_x\right)_y = f_{xy}.$$

Similar definitions hold for $\dfrac{\partial^2 f}{\partial y^2} = f_{yy}$ and $\dfrac{\partial^2 f}{\partial x \partial y} = f_{yx}$. The second partial derivatives f_{xy} and f_{yx} are **mixed partial derivatives** (*gemengde partiële afgeleide*).

The terms in Definition 15.10 all depend on limits, so each definition comes with the caveat where the limit exists.

Example 15.8

For each of the following functions, find all 6 first and second partial derivatives. That is, find

$$f_x, \quad f_y, \quad f_{xx}, \quad f_{yy}, \quad f_{xy} \quad \text{and} \quad f_{yx}.$$

1. $f(x,y) = \dfrac{x^3}{y^2}$
2. $f(x,y) = e^x \sin(x^2 y)$

Solution

In each, we give f_x and f_y immediately and then spend time deriving the second partial derivatives.

1. $f(x,y) = \dfrac{x^3}{y^2} = x^3 y^{-2}$

$$f_x(x,y) = \frac{3x^2}{y^2}$$

$$f_y(x,y) = -\frac{2x^3}{y^3}$$

$$f_{xx}(x,y) = \frac{\partial}{\partial x}(f_x) = \frac{\partial}{\partial x}\left(\frac{3x^2}{y^2}\right) = \frac{6x}{y^2}$$

$$f_{yy}(x,y) = \frac{\partial}{\partial y}(f_y) = \frac{\partial}{\partial y}\left(-\frac{2x^3}{y^3}\right) = \frac{6x^3}{y^4}$$

$$f_{xy}(x,y) = \frac{\partial}{\partial y}(f_x) = \frac{\partial}{\partial y}\left(\frac{3x^2}{y^2}\right) = -\frac{6x^2}{y^3}$$

$$f_{yx}(x,y) = \frac{\partial}{\partial x}(f_y) = \frac{\partial}{\partial x}\left(-\frac{2x^3}{y^3}\right) = -\frac{6x^2}{y^3}$$

2. $f(x,y) = e^x \sin(x^2 y)$

   Because the following partial derivatives get rather long, we omit the extra notation and just give the results. In several cases, multiple applications of the product and chain rules will be necessary, followed by some basic combination of like terms.

   $f_x(x,y) = e^x \sin(x^2y) + 2xye^x \cos(x^2y)$

   $f_y(x,y) = x^2e^x \cos(x^2y)$

   $f_{xx}(x,y) = e^x \sin(x^2y) + 4xye^x \cos(x^2y) + 2ye^x \cos(x^2y) - 4x^2y^2e^x \sin(x^2y)$

   $f_{yy}(x,y) = -x^4e^x \sin(x^2y)$

   $f_{xy}(x,y) = x^2e^x \cos(x^2y) + 2xe^x \cos(x^2y) - 2x^3ye^x \sin(x^2y)$

   $f_{yx}(x,y) = x^2e^x \cos(x^2y) + 2xe^x \cos(x^2y) - 2x^3ye^x \sin(x^2y)$

Higher-order partial derivatives can also be computed in Mathematica. For instance, given $f(x,y) = x^3/y^2$ from Example 15.8, we can find f_{xy} as follows:

```
In[27]:= D[x^3/y^2, x, y]

           6 x^2
Out[27]= - -----
            y^3
```

Notice how for each of the two functions in Example 15.8, $f_{xy} = f_{yx}$. Due to the complexity of the examples, this likely is not a coincidence. The following theorem states that it is not. It is also known as **Schwarz's theorem**, **Clairaut's theorem**, or **Young's theorem**.

Theorem 15.1 (Symmetry of second derivatives)
Let f be defined such that f_{xy} and f_{yx} are continuous on a set S. Then for each point (x,y) in S, $f_{xy}(x,y) = f_{yx}(x,y)$.

Now that we know how to find second partials, we investigate what they tell us.

Again we refer back to a function $y = f(x)$ of a single variable. The second derivative of f is "the derivative of the derivative," or "the rate of change of the rate of change." The second derivative measures how much the derivative is changing. If $f''(x) < 0$, then the derivative is getting smaller (so the graph of f is concave down); if $f''(x) > 0$, then the derivative is growing, making the graph of f concave up.

Now consider $z = f(x,y)$. Similar statements can be made about f_{xx} and f_{yy} as could be made about $f''(x)$ above. When taking derivatives with respect to x twice, we measure how much f_x changes with respect to x. If $f_{xx}(x,y) < 0$, it means that as x increases, f_x decreases, and the graph of f will be concave down in the x-direction. Using the analogy of standing in the rolling meadow used earlier in this section, f_{xx} measures whether one's path is concave up/down when walking due east. Similarly, f_{yy} measures the concavity in the y-direction. If $f_{yy}(x,y) > 0$, then f_y is increasing with respect to y and the graph of f will be concave up in the y-direction. Appealing to the rolling meadow analogy again, f_{yy} measures whether one's path is concave up/down when walking due north.

We now consider the mixed partials f_{xy} and f_{yx}. The mixed partial f_{xy} measures how much f_x changes with respect to y. Once again using the rolling meadow analogy, f_x measures the slope if one walks due east. Looking east, begin walking north (side–stepping). Is the path towards the east getting steeper? If so, $f_{xy} > 0$. Is the path towards the east not changing in steepness? If so, then $f_{xy} = 0$. A similar thing can be said about f_{yx}: consider the steepness of paths heading north while side–stepping to the east.

The following example examines these ideas with concrete numbers and graphs.

Example 15.9

Let $z = x^2 - y^2 + xy$. Evaluate the 6 first and second partial derivatives at $(-1/2, 1/2)$ and interpret what each of these numbers mean.

Solution

We find that:

$f_x(x,y) = 2x + y$, $f_y(x,y) = -2y + x$, $f_{xx}(x,y) = 2$, $f_{yy}(x,y) = -2$ and $f_{xy}(x,y) = f_{yx}(x,y) = 1$. Thus at $(-1/2, 1/2)$ we have

$$f_x\left(-\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2}, \qquad f_y\left(-\frac{1}{2}, \frac{1}{2}\right) = -\frac{3}{2}.$$

The slope of the tangent line at $(-1/2, 1/2, -1/4)$ in the direction of x is $-1/2$: if one moves from that point parallel to the x-axis, the instantaneous rate of change will be $-1/2$. The slope of the tangent line at this point in the direction of y is $-3/2$: if one moves from this point parallel to the y-axis, the instantaneous rate of change will be $-3/2$. These tangents lines are graphed in Figure 15.9(a) and 15.9(b), together with the curve where $x = -1/2$ and $y = 1/2$, respectively, where the tangent lines are drawn in a solid line.

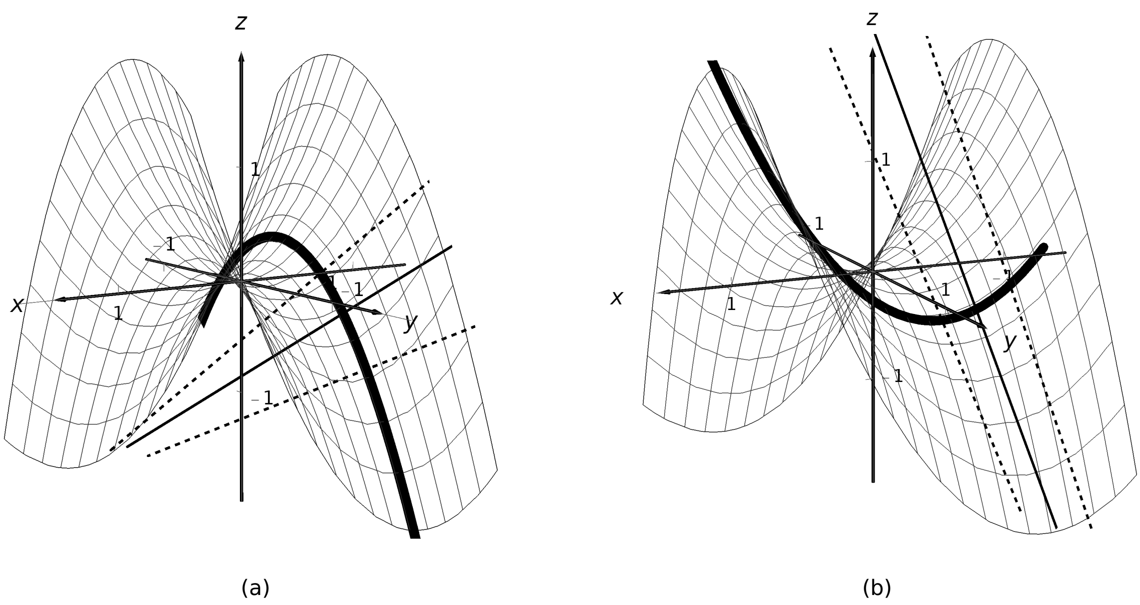


Figure 15.9: Understanding the second partial derivatives in Example 15.9.

Now consider only Figure 15.9(a). Three directed tangent lines are drawn (two are dashed), each in the direction of x; that is, each has a slope determined by f_x. Note how as y increases, the slope of these lines get closer to 0. Since the slopes are all negative, getting closer to 0 means the slopes are increasing. The slopes given by f_x are increasing as y increases, meaning f_{xy} must be positive.

Since $f_{xy} = f_{yx}$, we also expect f_y to increase as x increases. Consider Figure 15.9(b) where again three directed tangent lines are drawn, this time each in the direction of y with slopes determined by f_y. As x increases, the slopes become less steep (closer to 0). Since these are negative slopes, this means the slopes are increasing.

Thus far we have a visual understanding of f_x, f_y, and $f_{xy} = f_{yx}$. We now interpret f_{xx} and f_{yy}. In Figure 15.9(a), we see a curve drawn where x is held constant at $x = -1/2$: only y varies. This curve is clearly concave down, corresponding to the fact that $f_{yy} < 0$. In Figure 15.9(b), we see a similar curve where y is constant and only x varies. This curve is concave up, corresponding to the fact that $f_{xx} > 0$.

15.3.3 Higher-order partial derivatives

Essentially, we can continue taking partial derivatives of partial derivatives of partial derivatives of . . . ; we do not have to stop with second partial derivatives. These higher order partial derivatives do not have a tidy graphical interpretation; nevertheless they are not hard to compute and worthy of some practice.

We do not formally define each higher order derivative, but rather give just a few examples of the notation. For instance,

$$f_{xyx}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\right) \quad \text{and} \quad f_{xxz}(x,y) = \frac{\partial}{\partial z}\left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\right).$$

Example 15.10

Let

$$f(x,y) = x^2y^2 + \sin(xy).$$

Find f_{xxy} and f_{yxx}.

Solution

To find f_{xxy}, we first find f_x, then f_{xx}, then f_{xxy}:

$$f_x = 2xy^2 + y\cos(xy)$$
$$f_{xx} = 2y^2 - y^2\sin(xy)$$
$$f_{xxy} = 4y - 2y\sin(xy) - xy^2\cos(xy).$$

To find f_{yxx}, we first find f_y, then f_{yx}, then f_{yxx}:

$$\begin{aligned} f_y &= 2x^2y + x\cos(xy) \\ f_{yx} &= 4xy + \cos(xy) - xy\sin(xy) \\ f_{yxx} &= 4y - y\sin(xy) - \big(y\sin(xy) + xy^2\cos(xy)\big) \\ &= 4y - 2y\sin(xy) - xy^2\cos(xy). \end{aligned}$$

Note how $f_{xxy} = f_{yxx}$.

In the previous example we saw that $f_{xxy} = f_{yxx}$; this is not a coincidence. While we do not state this as a formal theorem, as long as each partial derivative is continuous, it does not matter the order in which the partial derivatives are taken. For instance, $f_{xxy} = f_{xyx} = f_{yxx}$.

With $z = f(x,y)$, the partial derivatives f_x and f_y measure the instantaneous rate of change of z when moving parallel to the x- and y-axes, respectively. How do we measure the rate of change at a point when we do not move parallel to one of these axes? What if we move in the direction given by the vector $\langle 2,1\rangle$? Can we measure that rate of change? The answer is, of course, yes, we can. This is the topic of Section 15.6. First, we need to define what it means for a function of two variables to be differentiable.

15.3.4 Functions of three variables

The concepts underlying partial derivatives can be easily extend to more than two variables. We give some definitions and examples in the case of three variables.

Definitie 15.11 (Partial derivative with three variables)
Let $w = f(x,y,z)$ be a continuous function on a set D in $\mathbb{R}^3$.

The **partial derivative of f with respect to x** is:

$$f_x(x,y,z) = \lim_{h\to 0} \frac{f(x+h,y,z) - f(x,y,z)}{h}.$$

Similar definitions hold for $f_y(x,y,z)$ and $f_z(x,y,z)$.

By taking partial derivatives of partial derivatives, we can find second partial derivatives of f with respect to z then y, for instance, just as before.

Example 15.11
For each of the following functions, find f_x, f_y, f_z, f_{xz}, f_{yz}, and f_{zz}.

1. $f(x,y,z) = x^2y^3z^4 + x^2y^2 + x^3z^3 + y^4z^4$
2. $f(x,y,z) = x\sin(yz)$

Solution

1. $f_x = 2xy^3z^4 + 2xy^2 + 3x^2z^3$
   $f_y = 3x^2y^2z^4 + 2x^2y + 4y^3z^4$
   $f_z = 4x^2y^3z^3 + 3x^3z^2 + 4y^4z^3$
   $f_{xz} = 8xy^3z^3 + 9x^2z^2$
   $f_{yz} = 12x^2y^2z^3 + 16y^3z^3$
   $f_{zz} = 12x^2y^3z^2 + 6x^3z + 12y^4z^2$

2. $f_x = \sin(yz)$
   $f_y = xz\cos(yz)$
   $f_z = xy\cos(yz)$
   $f_{xz} = y\cos(yz)$
   $f_{yz} = x\cos(yz) - xyz\sin(yz)$
   $f_{zz} = -xy^2\sin(xy)$

15.4 Total differential and differentiability

15.4.1 Total differential

We studied differentials in Section 9.7.2, where Definition 9.8 states that if $y = f(x)$ and f is differentiable, then $dy = f'(x)dx$. One important use of this differential is in integration by substitution. Another important application is approximation. Let $\Delta x = dx$ represent a change in x. When dx is small, $dy \approx \Delta y$, the change in y resulting from the change in x. So, as dx goes to 0, the error in approximating Δy with dy goes to 0.

We extend this idea to functions of two variables. Let $z = f(x,y)$, and let $\Delta x = dx$ and $\Delta y = dy$ represent changes in x and y, respectively (Figure 15.10). Let $\Delta z = f(x+dx, y+dy) - f(x,y)$ be the change in z over the change in x and y. Recalling that f_x and f_y give the instantaneous rates of z-change in the x- and y-directions, respectively, we can approximate Δz with $dz = f_x dx + f_y dy$; in words, the total change in z is approximately the change caused by changing x plus the change caused by changing y. In a moment we give an indication of whether or not this approximation is any good. First we give a name to dz.

Definitie 15.12 (Total differential)
Let $z = f(x,y)$ be continuous on a set S. Let dx and dy represent changes in x and y, respectively. Where the partial derivatives f_x and f_y exist, the **total differential of** z (*totale differentiaal van* z) is

$$dz = f_x(x,y)dx + f_y(x,y)dy.$$

Note that from Definition 15.12, we can as well use vector notation:

$$dz = \left(f_x, f_y\right) \cdot (dx, dy).$$

15.4.2 Differentiability

15.4.2.1 Definition

We can approximate Δz with dz, but as with all approximations, there is error involved. A good approximation is one in which the error is small. At a given point (x_0, y_0), let E_x and E_y be functions of dx and dy such that $E_x dx + E_y dy$ describes this error. Then

$$\begin{aligned}\Delta z &= dz + E_x dx + E_y dy\\ &= f_x(x_0,y_0)dx + f_y(x_0,y_0)dy + E_x dx + E_y dy.\end{aligned}$$

If the approximation of Δz by dz is good, then as dx and dy get small, so does $E_x dx + E_y dy$. The approximation of Δz by dz is even better if, as dx and dy go to 0, so do E_x and E_y. This leads us to our definition of differentiability.

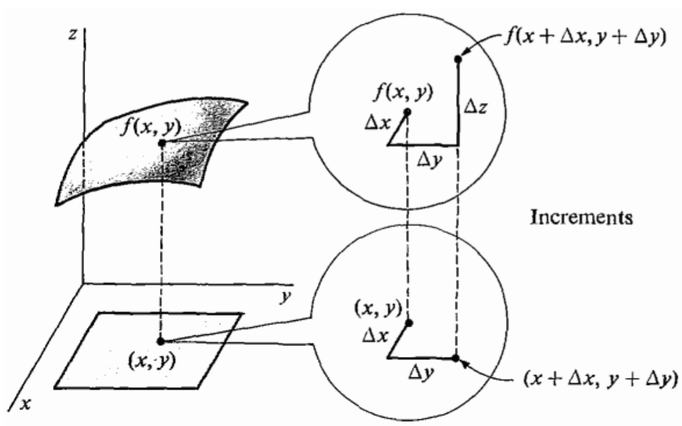


Figure 15.10: Understanding the total differential of a function of two variables.

Definitie 15.13 (Multivariable differentiability)
Let $z = f(x,y)$ be defined on a set S containing (x_0, y_0) where $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$ exist. Let dz be the total differential of z at (x_0,y_0), let $\Delta z = f(x_0+dx, y_0+dy) - f(x_0,y_0)$, and let E_x and E_y be functions of dx and dy such that

$$\Delta z = dz + E_x dx + E_y dy.$$

1. We say f **is differentiable at** (x_0,y_0) (*afleidbaar*) if, given $\varepsilon > 0$, there is a $\delta > 0$ such that if $||(dx,dy)|| < \delta$, then $||\big(E_x,E_y\big)|| < \varepsilon$. That is, as dx and dy go to 0, so do E_x and E_y.
2. We say f **is differentiable on** S (*afleidbaar over S*) if f is differentiable at every point in S. If f is differentiable on $\mathbb{R}^2$, we say that f is differentiable everywhere.

Example 15.12

Show $f(x,y) = xy + 3y^2$ is differentiable.

Solution

We begin by finding $f(x+dx, y+dy)$, Δz, f_x and f_y.

$$\begin{aligned} f(x+dx,y+dy) &= (x+dx)(y+dy)+3(y+dy)^2 \\ &= xy + xdy + ydx + dxdy + 3y^2 + 6ydy + 3dy^2. \end{aligned}$$

$\Delta z = f(x+dx, y+dy) - f(x,y)$, so

$$\Delta z = xdy + ydx + dxdy + 6ydy + 3dy^2.$$

It is straightforward to compute $f_x = y$ and $f_y = x + 6y$. Consider once more Δz:

$$\begin{aligned} \Delta z &= xdy + ydx + dxdy + 6ydy + 3dy^2 \qquad \text{(Now reorder.)} \\ &= ydx + xdy + 6ydy + dxdy + 3dy^2 \\ &= \underbrace{(y)}_{f_x} dx + \underbrace{(x+6y)}_{f_y} dy + \underbrace{(dy)}_{E_x} dx + \underbrace{(3dy)}_{E_y} dy \\ &= f_x dx + f_y dy + E_x dx + E_y dy. \end{aligned}$$

With $E_x = dy$ and $E_y = 3dy$, it is clear that as dx and dy go to 0, E_x and E_y also go to 0. Since this did not depend on a specific point (x_0,y_0), we can say that $f(x,y)$ is differentiable for all pairs (x,y) in $\mathbb{R}^2$, or, equivalently, that f is differentiable everywhere.

Our intuitive understanding of differentiability of functions $y = f(x)$ of one variable was that the graph of f was **smooth** (*glad*). A similar intuitive understanding of functions $z = f(x,y)$ of two variables is that the surface defined by f is also smooth, not containing cusps, edges, breaks, etc. The following theorem states that differentiable functions are continuous, followed by another theorem that provides a more tangible way of determining whether a great number of functions are differentiable or not.

Theorem 15.2 (Continuity and differentiability of multivariable functions)
Let $z = f(x,y)$ be defined on a set S containing (x_0,y_0). If f is differentiable at (x_0,y_0), then f is continuous at (x_0,y_0).

> **Theorem 15.3 (Differentiability of multivariable functions)**
> *Let $z = f(x,y)$ be defined on a set S. If f_x and f_y are both continuous on S, then f is differentiable on S.*

These theorems assure us that essentially all functions that we see in the course of our studies here are differentiable (and hence continuous) on their natural domains. There is a difference between Definition 15.13 and Theorem 15.3, though: it is possible for a function f to be differentiable yet f_x and/or f_y is not continuous. So when f_x and f_y exist at a point but are not continuous at that point, we need to use other methods to determine whether or not f is differentiable at that point.

For instance, consider the function

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0), \end{cases}$$

We can find $f_x(0,0)$ and $f_y(0,0)$ using Definition 15.9:

$$\begin{aligned} f_x(0,0) &= \lim_{h\to 0} \frac{f(0+h,0)-f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0}{h^2} = 0; \\ f_y(0,0) &= \lim_{h\to 0} \frac{f(0,0+h)-f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0}{h^2} = 0. \end{aligned}$$

Both f_x and f_y exist at $(0,0)$, but they are not continuous at $(0,0)$, as

$$f_x(x,y) = \frac{y(y^2-x^2)}{(x^2+y^2)^2} \qquad \text{and} \qquad f_y(x,y) = \frac{x(x^2-y^2)}{(x^2+y^2)^2}$$

are not continuous at $(0,0)$. Take the limit of f_x as $(x,y) \to (0,0)$ along the x- and y-axes; they give different results. So even though f_x and f_y exist at every point in the xy-plane, they are not continuous. Therefore it is possible, by Theorem 15.3, for f to not be differentiable.

Indeed, it is not. One can show that f is not continuous at $(0,0)$ (see Example 15.5), and by Theorem 15.2, this means f is not differentiable at $(0,0)$.

15.4.2.2 Approximating with differentials

By the definition, when f is differentiable dz is a good approximation for Δz when dx and dy are small. We give a simple example of how this is used here. We can use this to approximate error propagation; that is, if the input is a little off from what it should be, how far from correct will the output be? We demonstrate this in an example.

> ## Example 15.13
> A cylindrical steel storage tank is to be built that is 10m tall and 4m across in diameter. It is known that the steel will expand/contract with temperature changes; is the overall volume of the tank more sensitive to changes in the diameter or in the height of the tank?

Solution

A cylindrical solid with height h and radius r has volume $V = \pi r^2 h$. We can view V as a function of two variables, r and h. We can compute partial derivatives of V:

$$\frac{\partial V}{\partial r} = V_r(r,h) = 2\pi rh \qquad \text{and} \qquad \frac{\partial V}{\partial h} = V_h(r,h) = \pi r^2.$$

The total differential is $dV = (2\pi rh)dr + (\pi r^2)dh$. When $h = 10$ and $r = 2$, we have $dV = 40\pi dr + 4\pi dh$. Note that the coefficient of dr is $40\pi \approx 125.7$; the coefficient of dh is a tenth of that, approximately 12.57. A small change in radius will be multiplied by 125.7, whereas a small change in height will be multiplied by 12.57. Thus the volume of the tank is more sensitive to changes in radius than in height.

The previous example showed that the volume of a particular tank was more sensitive to changes in radius than in height. Keep in mind that this analysis only applies to a tank of those dimensions. A tank with a height of 0.3 m and radius of 5 m would be more sensitive to changes in height than in radius. One could make a chart of small changes in radius and height and find exact changes in volume given specific changes.

15.4.3 Functions of three variables

The definition of differentiability for functions of three variables is very similar to that of functions of two variables. We again start with the total differential.

Definitie 15.14 (Total differential)
Let $w = f(x,y,z)$ be continuous on a set D. Let dx, dy and dz represent changes in x, y and z, respectively. Where the partial derivatives f_x, f_y and f_z exist, the **total differential of** w is

$$dw = f_x(x,y,z)dx + f_y(x,y,z)dy + f_z(x,y,z)dz.$$

This differential can be a good approximation of the change in w when $w = f(x,y,z)$ is differentiable.

Definitie 15.15 (Multivariable differentiability)
Let $w = f(x,y,z)$ be defined on a set D containing (x_0,y_0,z_0) where $f_x(x_0,y_0,z_0)$, $f_y(x_0,y_0,z_0)$ and $f_z(x_0,y_0,z_0)$ exist. Let dw be the total differential of w at (x_0,y_0,z_0), let $\Delta w = f(x_0+dx, y_0+dy, z_0+dz) - f(x_0,y_0,z_0)$, and let E_x, E_y and E_z be functions of dx, dy and dz such that

$$\Delta w = dw + E_x dx + E_y dy + E_z dz.$$

1. We say f is differentiable at (x_0,y_0,z_0) if, given $\varepsilon > 0$, there is a $\delta > 0$ such that if $||(dx,dy,dz)|| < \delta$, then $||(E_x,E_y,E_z)|| < \varepsilon$.
2. We say f is differentiable on B if f is differentiable at every point in B. If f is differentiable on $\mathbb{R}^3$, we say that f is differentiable everywhere.

Just as before, this definition gives a rigorous statement about what it means to be differentiable that is not very intuitive. We follow it with a theorem similar to Theorem 15.3.

Theorem 15.4 (Continuity and differentiability of functions of three variables)
Let $w = f(x,y,z)$ be defined on a set D containing (x_0,y_0,z_0).

1. *If f is differentiable at (x_0,y_0,z_0), then f is continuous at (x_0,y_0,z_0).*
2. *If f_x, f_y and f_z are continuous on D, then f is differentiable on D.*

15.5 The multivariable chain rule

15.5.1 Rationale

Consider driving an off-road vehicle along a dirt road. As you drive, your elevation likely changes. What factors determine how quickly your elevation rises and falls? After some thought, generally one recognizes that one's velocity (speed and direction) and the terrain influence your rise and fall.

One can represent the terrain as the surface defined by a multivariable function $z = f(x,y)$; one can represent the path of the off-road vehicle, as seen from above, with a vector–valued function $\vec{r}(t) = \big(x(t), y(t)\big)$; the velocity of the vehicle is thus $\vec{r}\,'(t) = \big(x'(t), y'(t)\big)$.

Consider Figure 15.11 in which a surface $z = f(x,y)$ is drawn, along with a dashed curve in the xy-plane. Restricting f to just the points on this circle gives the curve shown on the surface (i.e., the path of the vehicle.) The derivative $\frac{df}{dt}$ gives the instantaneous rate of change of f with respect to t. If we consider an object travelling along this path, $\frac{df}{dt} = \frac{dz}{dt}$ gives the rate at which the object rises/falls. Conceptually, the multivariable chain rule combines terrain and velocity information properly to compute this rate of elevation change.

Abstractly, let z be a function of x and y; that is, $z = f(x,y)$ for some function f, and let x and y each be functions of t. By choosing a t-value, x- and y-values are determined, which in turn determine z: this defines z as a function of t. The multivariable chain rule gives a method of computing $\frac{dz}{dt}$.

Theorem 15.5 (Multivariable chain rule, Part I)
Let $z = f(x,y)$, $x = g(t)$ and $y = h(t)$, where f, g and h are differentiable functions. Then $z = f(x,y) = f\big(g(t),h(t)\big)$ is a function of t, and

$$\begin{aligned}\frac{dz}{dt} = \frac{df}{dt} &= f_x(x,y)\frac{dx}{dt} + f_y(x,y)\frac{dy}{dt} \\ &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \\ &= \big(f_x, f_y\big)\cdot\big(x', y'\big).\end{aligned} \tag{15.1}$$

Notice, the third line of equations in Theorem 15.5. The vector $\big(f_x, f_y\big)$ contains information about the surface (terrain); the vector $\big(x', y'\big)$ can represent velocity. In the context measuring the rate of elevation change of the off-road vehicle, the multivariable chain rule states it can be found through a product of terrain and velocity information.

We now practice applying the multivariable chain rule.

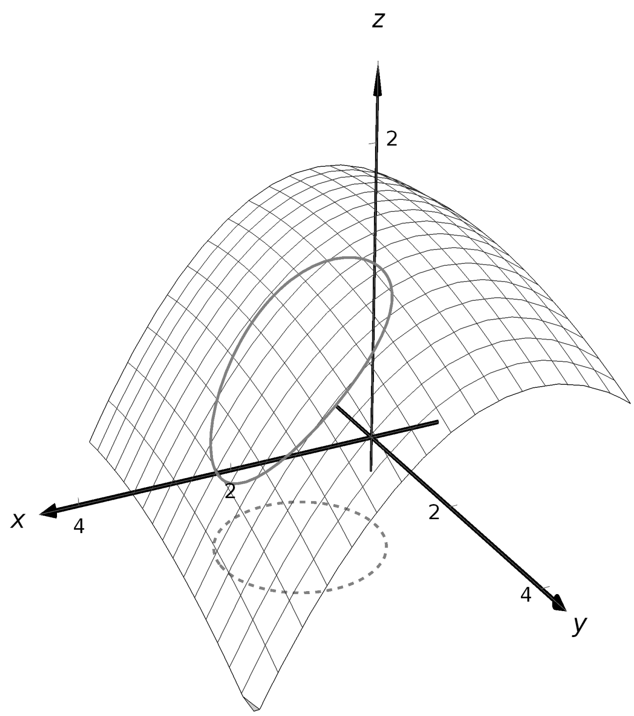


Figure 15.11: Understanding the application of the multivariable chain rule.

Example 15.14

Let $z = f(x, y) = x^2y + x$, where $x = \sin(t)$ and $y = e^{5t}$. Find $\frac{dz}{dt}$ using the chain rule.

Solution

Following Theorem 15.5, we first find

$$f_x(x, y) = 2xy + 1, \qquad f_y(x, y) = x^2, \qquad \frac{dx}{dt} = \cos(t), \qquad \frac{dy}{dt} = 5e^{5t}.$$

Applying the theorem, we have

$$\frac{dz}{dt} = (2xy + 1)\cos(t) + 5x^2e^{5t}.$$

This may look odd, as it seems that $\frac{dz}{dt}$ is a function of x, y and t. Since x and y are functions of t, $\frac{dz}{dt}$ is really just a function of t, and we can replace x with $\sin(t)$ and y with e^{5t} to arrive of:

$$\frac{dz}{dt} = (2\sin(t)e^{5t} + 1)\cos(t) + 5e^{5t}\sin^2(t).$$

The previous example can make us wonder: if we substituted for x and y at the end to show that $\frac{dz}{dt}$ is really just a function of t, why not substitute before differentiating, showing clearly that z is a function of t?

That is, $z = x^2y + x = \sin^2(t)e^{5t} + \sin(t)$. Applying the chain and product rules, we have

$$\frac{dz}{dt} = 2\sin(t)\cos(t)\,e^{5t} + 5\sin^2(t)\,e^{5t} + \cos(t),$$

which matches the result from the example.

This may now make one wonder "What's the point? If we could already find the derivative, why learn another way of finding it?" In some cases, applying this rule makes deriving simpler, but this is hardly the power of the chain rule. Rather, in the case where $z = f(x,y)$, $x = g(t)$ and $y = h(t)$, the chain rule is extremely powerful when we do not know what f, g and/or h are. We demonstrate this in the next example.

Example 15.15

An object travels along a path on a surface. The exact path and surface are not known, but at time $t = t_0$ it is known that :

$$\frac{\partial z}{\partial x} = 5, \qquad \frac{\partial z}{\partial y} = -2, \qquad \frac{dx}{dt} = 3 \qquad \text{and} \qquad \frac{dy}{dt} = 7.$$

Find $\frac{dz}{dt}$ at time t_0.

Solution

The multivariable chain rule states that

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}\\ &= 5(3) + (-2)(7)\\ &= 1.\end{aligned}$$

By knowing certain rates–of–change information about the surface and about the path of the particle in the xy-plane, we can determine how quickly the object is rising/falling.

We can also extend the chain rule to include the situation where z is a function of more than one variable, and each of these variables is also a function of more than one variable. The basic case of this is where $z = f(x,y)$, and x and y are functions of two variables, say s and t.

Theorem 15.6 (Multivariable chain rule, Part II)

1. *Let $z = f(x,y)$, $x = g(s,t)$ and $y = h(s,t)$, where f, g and h are differentiable functions. Then z is a function of s and t, and*

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} \qquad\qquad \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$$

2. *Let $w = f(x,y,z)$ be a differentiable function of three variables, where x, y and z are differentiable functions of the variables $t_1, t_2, \ldots, t_n$. Then w is a function of the t_i, and*

$$\frac{\partial z}{\partial t_i} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t_i} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t_i} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial t_i}.$$

Example 15.16

Let $z = f(x,y) = x^2y + x$, $x = s^2 + 3t$ and $y = 2s - t$. Find $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$, and evaluate each when $s = 1$ and $t = 2$.

Solution

Following Theorem 15.6, we compute the following partial derivatives:

$$\frac{\partial f}{\partial x} = 2xy + 1 \qquad \frac{\partial f}{\partial y} = x^2,$$

$$\frac{\partial x}{\partial s} = 2s \qquad \frac{\partial x}{\partial t} = 3 \qquad \frac{\partial y}{\partial s} = 2 \qquad \frac{\partial y}{\partial t} = -1.$$

Thus

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} = (2xy+1)(2s) + (x^2)(2) = 4xys + 2s + 2x^2,$$

and

$$\frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = (2xy+1)(3) + (x^2)(-1) = 6xy - x^2 + 3.$$

When $s = 1$ and $t = 2$, $x = 7$ and $y = 0$, so

$$\frac{\partial z}{\partial s} = 100 \qquad \text{and} \qquad \frac{\partial z}{\partial t} = -46.$$

Example 15.17

Let $w = xy + z^2$, where $x = t^2 e^s$, $y = t\cos(s)$, and $z = s\sin(t)$. Find $\frac{\partial w}{\partial t}$ when $s = 0$ and $t = \pi$.

Solution

Following Theorem 15.6, we compute the following partial derivatives:

$$\frac{\partial f}{\partial x} = y \qquad \frac{\partial f}{\partial y} = x \qquad \frac{\partial f}{\partial z} = 2z,$$

$$\frac{\partial x}{\partial t} = 2te^s \qquad \frac{\partial y}{\partial t} = \cos(s) \qquad \frac{\partial z}{\partial t} = s\cos(t).$$

Thus

$$\frac{\partial w}{\partial t} = y(2te^s) + x(\cos(s)) + 2z(s\cos(t)).$$

When $s = 0$ and $t = \pi$, we have $x = \pi^2$, $y = \pi$ and $z = 0$. Thus

$$\frac{\partial w}{\partial t} = \pi(2\pi) + \pi^2 = 3\pi^2.$$

15.5.2 Implicit functions

We studied finding $\frac{dy}{dx}$ when y is given as an implicit function of x in detail in Section 9.4. We find here that the multivariable chain rule gives a simpler method of finding $\frac{dy}{dx}$.

For instance, consider the implicit function $x^2y - xy^3 = 3$. We learned to use the following steps to find $\frac{dy}{dx}$:

$$\begin{aligned} \frac{d}{dx}\left(x^2y - xy^3\right) &= \frac{d}{dx}\left(3\right) \\ \Rightarrow \quad 2xy + x^2\frac{dy}{dx} - y^3 - 3xy^2\frac{dy}{dx} &= 0 \end{aligned}$$

$$\Leftrightarrow \quad \frac{dy}{dx} = -\frac{2xy - y^3}{x^2 - 3xy^2}. \tag{15.2}$$

Instead of using this method, consider $z = x^2y - xy^3$. The implicit function above describes the level curve $z = 3$. Considering x and y as functions of x, the multivariable chain rule states that

$$\frac{dz}{dx} = \frac{\partial z}{\partial x}\frac{dx}{dx} + \frac{\partial z}{\partial y}\frac{dy}{dx}. \tag{15.3}$$

Since z is constant (in our example, $z = 3$) it holds that, $\frac{dz}{dx} = 0$. We also know $\frac{dx}{dx} = 1$. Consequently, equation (15.3) becomes

$$0 = \frac{\partial z}{\partial x}(1) + \frac{\partial z}{\partial y}\frac{dy}{dx}.$$

Consequently,

$$\frac{dy}{dx} = -\frac{\partial z}{\partial x}\bigg/\frac{\partial z}{\partial y} = -\frac{f_x}{f_y} \tag{15.4}$$

Note how our solution for $\frac{dy}{dx}$ in Equation (15.2) is just the partial derivative of z with respect to x, divided by the partial derivative of z with respect to y, all multiplied by (-1).

Actually, Equation (15.4) is a consequence of the powerful implicit function theorem, which is, however, beyond the scope of this course.

We illustrate this approach for a problem from Section 9.4.

Example 15.18

Given the implicitly defined function $\sin(x^2y^2) + y^3 = x + y$, find y'. Note that this is the same problem as given in Example 9.16 of Section 9.4, where the solution took about a full page to find.

Solution

Let $f(x,y) = \sin(x^2y^2) + y^3 - x - y$; the implicitly defined function above is equivalent to $f(x,y) = 0$. We find $\frac{dy}{dx}$ by applying Theorem 15.4. We find

$$f_x(x,y) = 2xy^2\cos(x^2y^2) - 1 \qquad \text{and} \qquad f_y(x,y) = 2x^2y\cos(x^2y^2) + 3y^2 - 1,$$

so

$$\frac{dy}{dx} = -\frac{2xy^2\cos(x^2y^2) - 1}{2x^2y\cos(x^2y^2) + 3y^2 - 1},$$

which matches our solution from Example 9.16.

In Section 15.3 we learned how partial derivatives give certain instantaneous rate of change information about a function $z = f(x,y)$. In that section, we measured the rate of change of f by holding one variable constant and letting the other vary. We can visualize this change by considering the surface defined by f at a point and moving parallel to the x-axis.

What if we want to move in a direction that is not parallel to a coordinate axis? Can we still measure instantaneous rates of change? Yes; we find out how in the next section. In doing so, we will see how the multivariable chain rule informs our understanding of these directional derivatives.

15.6 Directional derivatives

15.6.1 Definition

Partial derivatives give us an understanding of how a surface changes when we move in the x- and y-directions. We made the comparison to standing in a rolling meadow and heading due east: the amount of rise/fall in doing so is comparable to f_x. Likewise, the rise/fall in moving due north is comparable to f_y. The steeper the slope, the greater in magnitude f_y.

But what if we did not move due north or east? What if we needed to move northeast and wanted to measure the amount of rise/fall? Partial derivatives alone cannot measure this. This section investigates **directional derivatives** (*richtingsafgeleide*), which do measure this rate of change.

We begin with a definition.

Definitie 15.16 (Directional derivative)
Let $z = f(x,y)$ be continuous on a set S and let $\boldsymbol{\hat{u}} = (u_1, u_2)$ be a unit vector. For all points (x,y), the **directional derivative of** f **at** (x,y) **in the direction of** $\boldsymbol{\hat{u}}$ is

$$D_{\boldsymbol{\hat{u}}} f(x,y) = \lim_{h \to 0} \frac{f(x+hu_1, y+hu_2) - f(x,y)}{h}.$$

The partial derivatives f_x and f_y are defined with similar limits, but only x or y varies with h, not both. Here both x and y vary with a weighted h, determined by a particular unit vector $\boldsymbol{\hat{u}}$. In practice it can be a very difficult limit to evaluate, so we need an easier way of taking directional derivatives.

For that purpose, let us define a new function of a single variable,

$$g(z) = f(x_0 + az, y_0 + bz),$$

where x_0, y_0, a, and b are some fixed numbers. Then, by the definition of the derivative for functions of a single variable we have,

$$g'(z) = \lim_{h \to 0} \frac{g(z+h) - g(z)}{h},$$

while the derivative at $z = 0$ is given by,

$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h}.$$

If we now substitute our expression for $g(z)$ we get,

$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h} = \lim_{h \to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0)}{h} = D_{\boldsymbol{\hat{u}}} f(x_0, y_0). \tag{15.5}$$

Now, let us look at this from another perspective and rewrite $g(z)$ as follows,

$$g(z) = f(x,y),$$

where $x = x_0 + az$ and $y = y_0 + bz$. We can now use the chain rule to compute,

$$g'(z) = \frac{dg}{dz} = \frac{\partial f}{\partial x}\frac{dx}{dz} + \frac{\partial f}{\partial y}\frac{dy}{dz} = f_x(x,y)a + f_y(x,y)b. \tag{15.6}$$

If we now take $z = 0$ we will get that $x = x_0$ and $y = y_0$ and plug these into Equation (15.6) we get

$$g'(0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b. \tag{15.7}$$

Now, simply equate Equations (15.5) and (15.7) to get that

$$D_{\hat{\mathbf{u}}} f(x_0, y_0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b.$$

If we now go back to allowing x and y to be any number we get the following formula for computing directional derivatives:

$$D_{\hat{\mathbf{u}}} f(x, y) = f_x(x, y)a + f_y(x, y)b.$$

This leads to the following theorem.

Theorem 15.7 (Directional derivatives)
Let $z = f(x, y)$ be differentiable on a set S containing (x_0, y_0), and let $\hat{\mathbf{u}} = (u_1, u_2)$ be a unit vector. The directional derivative of f at (x_0, y_0) in the direction of $\hat{\mathbf{u}}$ is

$$D_{\hat{\mathbf{u}}} f(x_0, y_0) = \big(f_x, f_y\big) \cdot (u_1, u_2) = f_x(x_0, y_0)u_1 + f_y(x_0, y_0)u_2.$$

Example 15.19

Let $z = 14 - x^2 - y^2$ and let $P = (1, 2)$. Find the directional derivative of f, at P, in the following directions:

1. toward the point $Q = (3, 4)$,
2. in the direction of $(2, -1)$, and
3. toward the origin.

The surface is plotted in Figure 15.12, where the point $P = (1, 2)$ is indicated in the xy-plane as well as the point $(1, 2, 9)$ which lies on the surface of f.

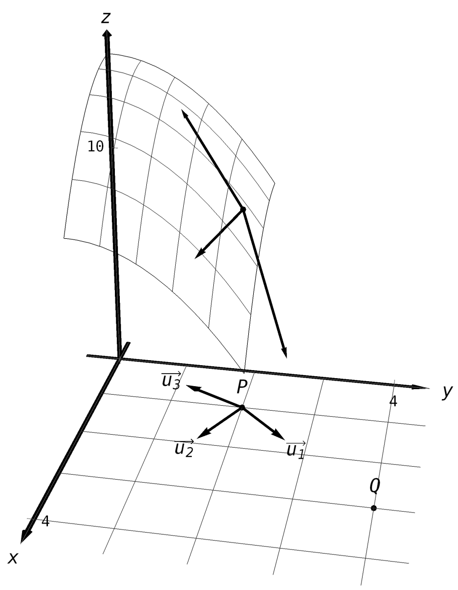


Figure 15.12: Understanding the directional derivative in Example 15.19.

Solution

We find that $f_x(x,y) = -2x$ and $f_x(1,2) = -2$; $f_y(x,y) = -2y$ and $f_y(1,2) = -4$.

1. Let $\hat{\boldsymbol{u}}_1$ be the unit vector that points from the point $P = (1,2)$ to the point $Q = (3,4)$, as shown in the figure. The vector $\overrightarrow{PQ} = (2,2)$; the unit vector in this direction is $\hat{\boldsymbol{u}}_1 = \left(1/\sqrt{2}, 1/\sqrt{2}\right)$. Thus the directional derivative of f at $(1,2)$ in the direction of $\hat{\boldsymbol{u}}_1$ is
$$D_{\hat{\boldsymbol{u}}_1} f(1,2) = -2\left(\frac{1}{\sqrt{2}}\right) + (-4)\left(\frac{1}{\sqrt{2}}\right) = -\frac{6}{\sqrt{2}} \approx -4.24.$$
Thus the instantaneous rate of change in moving from the point $(1,2,9)$ on the surface in the direction of $\hat{\boldsymbol{u}}_1$ (which points toward the point Q) is about -4.24. Moving in this direction moves one steeply downward.

2. We seek the directional derivative in the direction of $(2,-1)$. The unit vector in this direction is $\hat{\boldsymbol{u}}_2 = \left(2/\sqrt{5}, -1/\sqrt{5}\right)$. Thus the directional derivative of f at $(1,2)$ in the direction of $\hat{\boldsymbol{u}}_2$ is
$$D_{\hat{\boldsymbol{u}}_2} f(1,2) = -2\left(\frac{2}{\sqrt{5}}\right) + (-4)\left(-\frac{1}{\sqrt{5}}\right) = 0.$$
Starting on the surface of f at $(1,2)$ and moving in the direction of $(2,-1)$ (or $\hat{\boldsymbol{u}}_2$) results in no instantaneous change in z-value. This is analogous to standing on the side of a hill and choosing a direction to walk that does not change the elevation. One neither walks up nor down, rather just along the side of the hill.

3. At $P = (1,2)$, the direction towards the origin is given by the vector $(-1,-2)$; the unit vector in this direction is $\hat{\boldsymbol{u}}_3 = \left(-1/\sqrt{5}, -2/\sqrt{5}\right)$. The directional derivative of f at P in the direction of the origin is
$$D_{\hat{\boldsymbol{u}}_3} f(1,2) = -2\left(-\frac{1}{\sqrt{5}}\right) + (-4)\left(-\frac{2}{\sqrt{5}}\right) = \frac{10}{\sqrt{5}} \approx 4.47.$$
Moving towards the origin means walking uphill quite steeply, with an initial slope of about 4.47.

15.6.2 The gradient

As we study directional derivatives, it will help to make an important connection between the unit vector $\hat{\boldsymbol{u}} = (u_1, u_2)$ that describes the direction and the partial derivatives f_x and f_y. We start with a definition.

Definitie 15.17 (Gradient)
Let $z = f(x,y)$ be differentiable on a set S that contains the point (x_0, y_0).

1. The **gradient** (*gradiënt*) of f is
$$\vec{\nabla} f(x,y) = \left(f_x(x,y), f_y(x,y)\right).$$

2. The **gradient** of f at (x_0, y_0) is
$$\vec{\nabla} f(x_0,y_0) = \left(f_x(x_0,y_0), f_y(x_0,y_0)\right).$$

The symbol $\vec{\nabla}$ is named **nabla**, derived from the Greek name of a Jewish harp. Oddly enough, in mathematics the expression $\vec{\nabla} f$ is pronounced del f.

To simplify notation, we often express the gradient as $\vec{\nabla} f = (f_x, f_y)$. The gradient allows us to compute directional derivatives in terms of a dot product:

$$D_{\vec{u}} f = \vec{\nabla} f \cdot \vec{u}. \tag{15.8}$$

The properties of the dot product studied in Chapter 6 allow us to investigate the properties of the directional derivative. Given that the directional derivative gives the instantaneous rate of change of z when moving in the direction of $\vec{u}$, three questions naturally arise:

1. In what direction(s) is the change in z the greatest (i.e., the steepest uphill)?
2. In what direction(s) is the change in z the least (i.e., the steepest downhill)?
3. In what direction(s) is there no change in z?

Relying on the geometric interpretation of the dot product (Theorem 6.2), we have

$$\vec{\nabla} f \cdot \vec{u} = \left\|\vec{\nabla} f\right\| \|\vec{u}\| \cos(\theta) = \left\|\vec{\nabla} f\right\| \cos(\theta), \tag{15.9}$$

where θ is the angle between the gradient and $\vec{u}$. (Since $\vec{u}$ is a unit vector, $\|\vec{u}\| = 1$.) This equation allows us to answer the three questions stated previously.

1. Equation (15.9) is maximized when $\cos(\theta) = 1$, i.e., when the gradient and $\vec{u}$ have the same direction. We conclude the gradient points in the direction of greatest z change.
2. Equation (15.9) is minimized when $\cos(\theta) = -1$, i.e., when the gradient and $\vec{u}$ have opposite directions. We conclude the gradient points in the opposite direction of the least z change.
3. Equation (15.9) is 0 when $\cos(\theta) = 0$, i.e., when the gradient and $\vec{u}$ are orthogonal to each other. We conclude the gradient is orthogonal to directions of no z change.

This result is rather amazing. Once again imagine standing in a rolling meadow and face the direction that leads you steepest uphill. Then the direction that leads steepest downhill is directly behind you, and side–stepping either left or right (i.e., moving perpendicularly to the direction you face) does not change your elevation at all.

Recall that a level curve is defined as a curve in the xy-plane along which the z-values of a function do not change. Let a surface $z = f(x, y)$ be given, and let us represent one such level curve as a vector–valued function, $\vec{r}(t) = (x(t), y(t))$. As the output of f does not change along this curve, $f(x(t), y(t)) = c$ for all t, for some constant c.

Since f is constant for all t, $\frac{df}{dt} = 0$. By the multivariable chain rule, we also know

$$\begin{aligned} \frac{df}{dt} &= f_x(x, y)x'(t) + f_y(x, y)y'(t) \\ &= (f_x(x, y), f_y(x, y)) \cdot (x'(t), y'(t)) \\ &= \vec{\nabla} f \cdot \vec{r}\,'(t) \\ &= 0. \end{aligned}$$

This last equality states $\vec{\nabla} f \cdot \vec{r}\,'(t) = 0$: the gradient is orthogonal to the derivative of $\vec{r}$, meaning the gradient is orthogonal to the graph of $\vec{r}$. Our conclusion: at any point on a surface, the gradient at that point is orthogonal to the level curve that passes through that point.

We restate these ideas in a theorem, then use them in an example.

Theorem 15.8 (The gradient and directional derivatives)
Let $z = f(x,y)$ be differentiable on a set S with gradient $\vec{\nabla} f$, let $P = (x_0, y_0)$ be a point in S and let $\hat{\boldsymbol{u}}$ be a unit vector.

1. *The maximum value of $D_{\hat{\boldsymbol{u}}} f(x_0,y_0)$ is $\left\|\vec{\nabla} f(x_0,y_0)\right\|$; the direction of maximal z increase is $\vec{\nabla} f(x_0,y_0)$.*
2. *The minimum value of $D_{\hat{\boldsymbol{u}}} f(x_0,y_0)$ is $-\left\|\vec{\nabla} f(x_0,y_0)\right\|$; the direction of maximal z decrease is $-\vec{\nabla} f(x_0,y_0)$.*
3. *At P, $\vec{\nabla} f(x_0,y_0)$ is orthogonal to the level curve passing through $\big(x_0, y_0, f(x_0,y_0)\big)$.*

We now illustrate how to find the directions of maximal increase and decrease.

Example 15.20

Let $f(x,y) = \sin(x)\cos(y)$ and let $P = (\pi/3, \pi/3)$. Find the directions of maximal increase and decrease, and find a direction where the instantaneous rate of z change is 0.

Solution

We begin by finding the gradient. We have that $f_x = \cos(x)\cos(y)$ and $f_y = -\sin(x)\sin(y)$, thus

$$\vec{\nabla} f = \big(\cos(x)\cos(y), -\sin(x)\sin(y)\big)$$

and, at P

$$\vec{\nabla} f\left(\frac{\pi}{3}, \frac{\pi}{3}\right) = \left(\frac{1}{4}, -\frac{3}{4}\right).$$

Thus the direction of maximal increase is $(1/4, -3/4)$. In this direction, the instantaneous rate of z change is $||(1/4, -3/4)|| = \sqrt{10}/4 \approx 0.79$.

Figure 15.13 shows the surface. The gradient is drawn at P with a dashed line (because of the nature of this surface, the gradient points into the surface). Let $\hat{\boldsymbol{u}} = (u_1, u_2)$ be the unit vector in the direction of $\vec{\nabla} f$ at P. The graph also contains the vector $\big(u_1, u_2, ||\vec{\nabla} f||\big)$. This vector has a run of 1 (because in the xy-plane it moves 1 unit) and a rise of $||\vec{\nabla} f||$, hence we can think of it as a vector with slope of $||\vec{\nabla} f||$ in the direction of $\vec{\nabla} f$, helping us visualize how steep the surface is in its steepest direction.

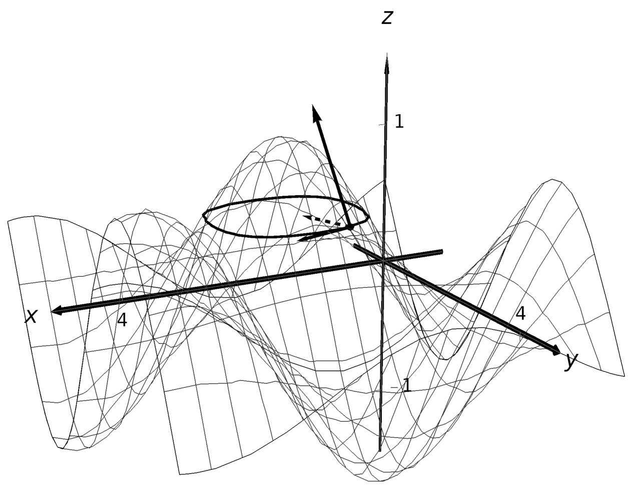


Figure 15.13: Graphing the surface and important directions in Example 15.20.

The direction of maximal decrease is $\langle -1/4, 3/4\rangle$; in this direction the instantaneous rate of z change is $-\sqrt{10}/4 \approx -0.79$.

Any direction orthogonal to $\vec{\nabla} f$ is a direction of no z change. We have two choices: the direction of $\langle 3, 1\rangle$ and the direction of $\langle -3, -1\rangle$. The unit vector in the direction of $\langle 3, 1\rangle$ is shown in the graph as well. The level curve at $z = \sqrt{3}/4$ is drawn: recall that along this curve the z-values do not change. Since $\langle 3, 1\rangle$ is a direction of no z-change, this vector is tangent to the level curve at P.

It is as well important to figure out when $\vec{\nabla} f = 0$.

Example 15.21

Let $f(x,y) = -x^2 + 2x - y^2 + 2y + 1$. Find the directional derivative of f in any direction at $P = (1, 1)$.

Solution

We find $\vec{\nabla} f = \langle -2x + 2, -2y + 2\rangle$. At P, we have $\vec{\nabla} f(1,1) = \langle 0, 0\rangle$. According to Theorem 15.8, this is the direction of maximal increase. However, $\langle 0, 0\rangle$ is directionless; it has no displacement. And regardless of the unit vector $\hat{\boldsymbol{u}}$ chosen, $D_{\hat{\boldsymbol{u}}} f = 0$.

Figure 15.14 helps us understand what this means. We can see that P lies at the top of a paraboloid. In all directions, the instantaneous rate of change is 0. So what is the direction of maximal increase? It is fine to give an answer of $\vec{\mathbf{0}} = \langle 0, 0\rangle$, as this indicates that all directional derivatives are 0.

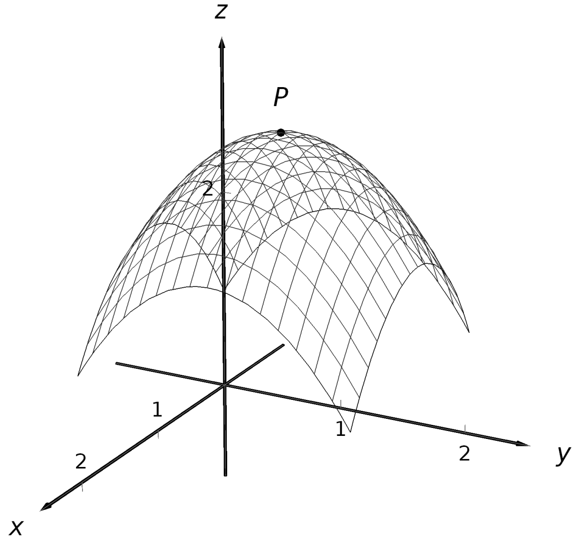


Figure 15.14: At the top of a paraboloid, all directional derivatives are 0.

In Mathematica, we could have computed the gradient in Example 15.21 using the command `Grad` as follows

```
In[28]:= Grad[-x^2 + 2*x - y^2 + 2*y + 1, {x, y}]

Out[28]= {2-2 x,2-2 y}
```

The fact that the gradient of a surface always points in the direction of steepest increase/decrease is very useful, as illustrated in the following example.

Example 15.22

Consider the surface given by $f(x,y) = 20 - x^2 - 2y^2$. Water is poured on the surface at $(1, 1/4)$. What path does it take as it flows downhill?

Solution

Let $\vec{r}(t) = \big(x(t), y(t)\big)$ be the vector–valued function describing the path of the water in the xy-plane; we seek $x(t)$ and $y(t)$. We know that water will always flow downhill in the steepest direction; therefore, at any point on its path, it will be moving in the direction of $-\vec{\nabla} f$. We ignore the physical effects of momentum on the water. Thus $\vec{r}\,'(t)$ will be parallel to $\vec{\nabla} f$, and there is some constant c such that $c\vec{\nabla} f = \vec{r}\,'(t) = \big(x'(t), y'(t)\big)$.

We find $\vec{\nabla} f = (-2x, -4y)$ and write $x'(t)$ as $\frac{dx}{dt}$ and $y'(t)$ as $\frac{dy}{dt}$. Then

$$\begin{aligned} & c\vec{\nabla} f = \big(x'(t), y'(t)\big) \\ \Leftrightarrow \qquad & (-2cx, -4cy) = \left(\frac{dx}{dt}, \frac{dy}{dt}\right). \end{aligned}$$

This implies

$$-2cx = \frac{dx}{dt} \qquad \text{and} \qquad -4cy = \frac{dy}{dt},$$

i.e.

$$c = -\frac{1}{2x}\frac{dx}{dt} \qquad \text{and} \qquad c = -\frac{1}{4y}\frac{dy}{dt}.$$

As c equals both expressions, we have

$$\frac{1}{2x}\frac{dx}{dt} = \frac{1}{4y}\frac{dy}{dt}.$$

To find an explicit relationship between x and y, we can integrate both sides with respect to t. Recall from our study of differentials that $\frac{dx}{dt}dt = dx$. Thus:

$$\begin{aligned}
\int \frac{1}{2x}\frac{dx}{dt}\,dt &= \int \frac{1}{4y}\frac{dy}{dt}\,dt \\
\Leftrightarrow \quad \int \frac{1}{2x}\,dx &= \int \frac{1}{4y}\,dy \\
\Leftrightarrow \quad \frac{1}{2}\ln|x| &= \frac{1}{4}\ln|y| + C_1 \\
\Leftrightarrow \quad 2\ln|x| &= \ln|y| + 4C_1 \\
\Leftrightarrow \quad \ln(x^2) &= \ln|y| + 4C_1.
\end{aligned}$$

Now raise both sides as a power of e:

$$\begin{aligned}
x^2 &= e^{\ln|y|+4C_1} \\
\Leftrightarrow \quad x^2 &= e^{\ln|y|}e^{4C_1}.
\end{aligned}$$

From which it follows that

$$x^2 = yC_2,$$

where $C_2 = \pm e^{4C_1}$, or alternatively

$$Cx^2 = y,$$

where $C = 1/C_2$. As the water started at the point $(1, 1/4)$, we can solve for C:

$$C(1)^2 = \frac{1}{4} \quad \Leftrightarrow \quad C = \frac{1}{4}.$$

Thus the water follows the curve $y = x^2/4$ in the xy-plane. The surface and the path of the water is graphed in Figure 15.15(a). In Figure 15.15(b), the level curves of the surface are plotted in the xy-plane, along with the curve $y = x^2/4$. Notice how the path intersects the level curves at right angles. As the path follows the gradient downhill, this reinforces the fact that the gradient is orthogonal to level curves.

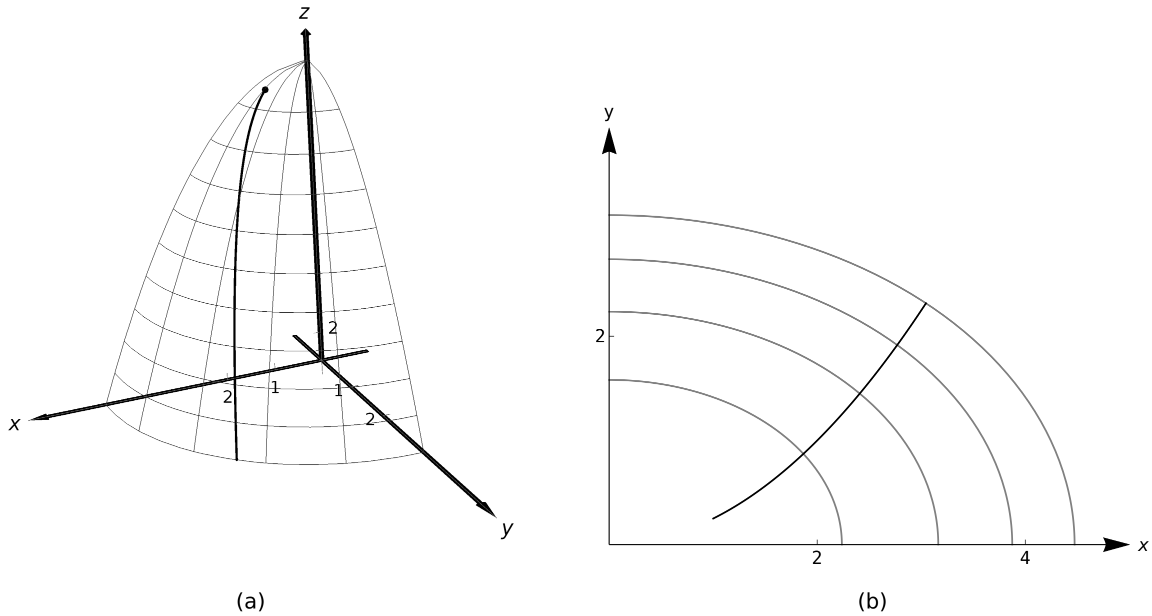


Figure 15.15: A graph of the surface described in Example 15.22 along with the path in the xy-plane with the level curves.

15.6.3 Functions of three variables

The concepts of directional derivatives and the gradient are easily extended to three (and more) variables. We combine the concepts behind Definitions 15.16 and 15.17 and Theorem 15.7 into one set of definitions.

Definitie 15.18 (Directional derivatives and gradient with three variables)
Let $w = f(x, y, z)$ be differentiable on a set D and let $\hat{\boldsymbol{u}}$ be a unit vector in $\mathbb{R}^3$.

1. The **gradient of** f is $\vec{\nabla} f = (f_x, f_y, f_z)$.
2. The **directional derivative of** f **in the direction of** $\hat{\boldsymbol{u}}$ is

$$D_{\hat{\boldsymbol{u}}} f = \vec{\nabla} f \cdot \hat{\boldsymbol{u}}.$$

The same properties of the gradient given in Theorem 15.8, when f is a function of two variables, hold for f, a function of three variables.

Theorem 15.9 (The gradient and directional derivatives with three variables)
Let $w = f(x, y, z)$ be differentiable on a set D, let $\vec{\nabla} f$ be the gradient of f, and let $\hat{\boldsymbol{u}}$ be a unit vector.

1. *The maximum value of $D_{\hat{\boldsymbol{u}}} f$ is $\|\vec{\nabla} f\|$, obtained when the angle between $\vec{\nabla} f$ and $\hat{\boldsymbol{u}}$ is 0, i.e., the direction of maximal increase is $\vec{\nabla} f$.*
2. *The minimum value of $D_{\hat{\boldsymbol{u}}} f$ is $-\|\vec{\nabla} f\|$, obtained when the angle between $\vec{\nabla} f$ and $\hat{\boldsymbol{u}}$ is π, i.e., the direction of maximal decrease is $-\vec{\nabla} f$.*
3. *$D_{\hat{\boldsymbol{u}}} f = 0$ when $\vec{\nabla} f$ and $\hat{\boldsymbol{u}}$ are orthogonal.*

We interpret the third statement of the theorem as the gradient is orthogonal to level surfaces, the three–variable analogue to level curves.

Example 15.23

If a point source S is radiating energy, the intensity I at a given point P in space is inversely proportional to the square of the distance between S and P. That is, when $S = (0,0,0)$, it holds that

$$I(x,y,z) = \frac{k}{x^2+y^2+z^2}$$

for some constant k.

Let $k=1$, let $\hat{\boldsymbol{u}} = (2/3,2/3,1/3)$ be a unit vector, and let $P = (2,5,3)$. Measure distances in centimetres. Find the directional derivative of I at P in the direction of $\hat{\boldsymbol{u}}$, and find the direction of greatest intensity increase at P.

Solution

We need the gradient $\vec{\nabla} I$, so we compute I_x, I_y and I_z:

$$\begin{aligned}
\vec{\nabla} I &= \left(\frac{-2x}{(x^2+y^2+z^2)^2}, \frac{-2y}{(x^2+y^2+z^2)^2}, \frac{-2z}{(x^2+y^2+z^2)^2}\right) \\
\Rightarrow \quad \vec{\nabla} I(2,5,3) &= \left(\frac{-4}{1444}, \frac{-10}{1444}, \frac{-6}{1444}\right) \approx (-0.003, -0.007, -0.004) \\
\Rightarrow \quad D_{\hat{\boldsymbol{u}}} I &= \vec{\nabla} I(2,5,3) \cdot \hat{\boldsymbol{u}} \\
&= -\frac{17}{2166} \approx -0.0078.
\end{aligned}$$

The directional derivative tells us that moving in the direction of $\hat{\boldsymbol{u}}$ from P results in a decrease in intensity of about -0.008 units per centimetre. The intensity is decreasing as $\hat{\boldsymbol{u}}$ moves one farther from the origin than P.

The gradient gives the direction of greatest intensity increase. Notice that

$$\begin{aligned}
\vec{\nabla} I(2,5,3) &= \left(\frac{-4}{1444}, \frac{-10}{1444}, \frac{-6}{1444}\right) \\
&= \frac{2}{1444}(-2,-5,-3).
\end{aligned}$$

That is, the gradient at $(2,5,3)$ is pointing in the direction of $(-2,-5,-3)$, that is, towards the origin. That should make intuitive sense: the greatest increase in intensity is found by moving towards to source of the energy.

The directional derivative allows us to find the instantaneous rate of z change in any direction at a point. We can use these instantaneous rates of change to define lines and planes that are tangent to a surface at a point, which is the topic of the next section.

15.7 Tangent lines, normal lines, and tangent planes

15.7.1 Tangent and normal lines

Derivatives and tangent lines go hand–in–hand. Given $y = f(x)$, the line tangent to the graph of f at $x = x_0$ is the line through $\big(x_0, f(x_0)\big)$ with slope $f'(x_0)$; that is, the slope of the tangent line is the instantaneous rate of change of f at x_0. When dealing with functions of two variables, the graph is no longer a curve but a surface. At a given point on the surface, it seems there are many lines that fit our intuition of being tangent to the surface.

In Figure 15.16 we see lines that are tangent to curves in space. Since each curve lies on a surface, it makes sense to say that the lines are also tangent to the surface. The next definition formally defines what it means to be tangent to a surface.

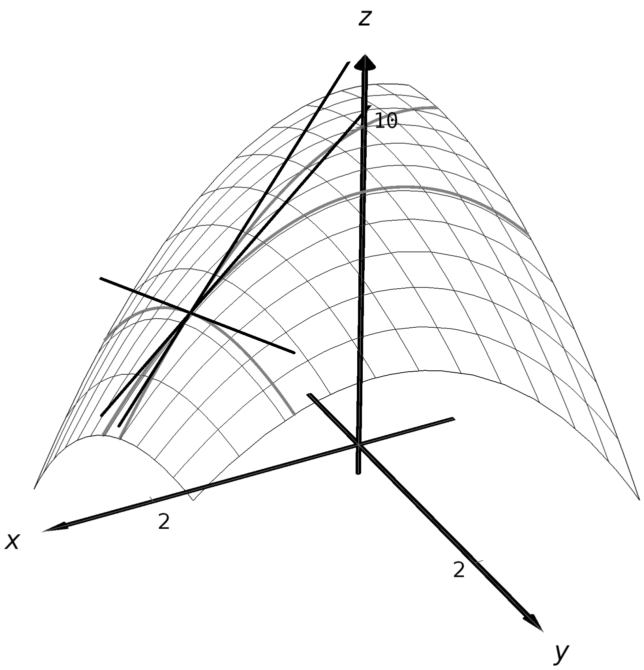


Figure 15.16: Showing various lines tangent to a surface.

Definitie 15.19 (Directional tangent line)
Let $z = f(x,y)$ be differentiable on a set S containing (x_0, y_0) and let $\hat{\boldsymbol{u}} = (u_1, u_2)$ be a unit vector.

1. The line ℓ_x through $\big(x_0, y_0, f(x_0,y_0)\big)$ parallel to $\big(1, 0, f_x(x_0,y_0)\big)$ is **the tangent line to f in the direction of x at** (x_0, y_0).
2. The line ℓ_y through $\big(x_0, y_0, f(x_0,y_0)\big)$ parallel to $\big(0, 1, f_y(x_0,y_0)\big)$ is **the tangent line to f in the direction of y at** (x_0, y_0).
3. The line $\ell_{\hat{\boldsymbol{u}}}$ through $\big(x_0, y_0, f(x_0,y_0)\big)$ parallel to $\big(u_1, u_2, D_{\hat{\boldsymbol{u}}} f(x_0,y_0)\big)$ is **the tangent line to f in the direction of $\hat{\boldsymbol{u}}$ at** (x_0, y_0).

It is instructive to consider each of three directions given in the definition in terms of slope. The direction of ℓ_x is $\big(1, 0, f_x(x_0,y_0)\big)$; that is, the "run" is one unit in the x-direction and the rise is $f_x(x_0,y_0)$ units in the z-direction. Note how the slope is just the partial derivative with respect to x. A similar statement can be made for ℓ_y. The direction of $\ell_{\hat{\boldsymbol{u}}}$ is $\big(u_1, u_2, D_{\hat{\boldsymbol{u}}} f(x_0,y_0)\big)$; the run is one unit in the $\hat{\boldsymbol{u}}$ direction (where $\hat{\boldsymbol{u}}$ is a unit vector) and the rise is the directional derivative of z in that direction.

Definition 15.19 leads to the following parametric equations of directional tangent lines:

$$\ell_x(t) = \begin{cases} x &= x_0 + t \\ y &= y_0 \\ z &= z_0 + f_x(x_0,y_0)t, \end{cases} \qquad \ell_y(t) = \begin{cases} x &= x_0 \\ y &= y_0 + t \\ z &= z_0 + f_y(x_0,y_0)t \end{cases} \qquad \text{and } \ell_{\hat{\boldsymbol{u}}}(t) = \begin{cases} x &= x_0 + u_1 t \\ y &= y_0 + u_2 t \\ z &= z_0 + D_{\hat{\boldsymbol{u}}} f(x_0,y_0)t. \end{cases}$$

where $z_0 = f(x_0, y_0)$.

Example 15.24

Find the lines tangent to the surface $z = \sin(x)\cos(y)$ at $(\pi/2, \pi/2)$ in the x- and y- directions and also in the direction of $\vec{v} = \langle -1, 1 \rangle$.

Solution

The partial derivatives with respect to x and y are:

$$f_x(x,y) = \cos(x)\cos(y)$$
$$f_y(x,y) = -\sin(x)\sin(y).$$

from which it follows that $f_x(\pi/2, \pi/2) = 0$ and $f_y(\pi/2, \pi/2) = -1$. At $(\pi/2, \pi/2)$, the z-value is 0.

Thus the parametric equations of the line tangent to f at $(\pi/2, \pi/2)$ in the directions of x and y are:

$$\ell_x(t) = \begin{cases} x = \pi/2 + t \\ y = \pi/2 \\ z = 0 \end{cases} \quad \text{and} \quad \ell_y(t) = \begin{cases} x = \pi/2 \\ y = \pi/2 + t \\ z = -t. \end{cases}$$

The two lines are shown with the surface in Figure 15.17(a).

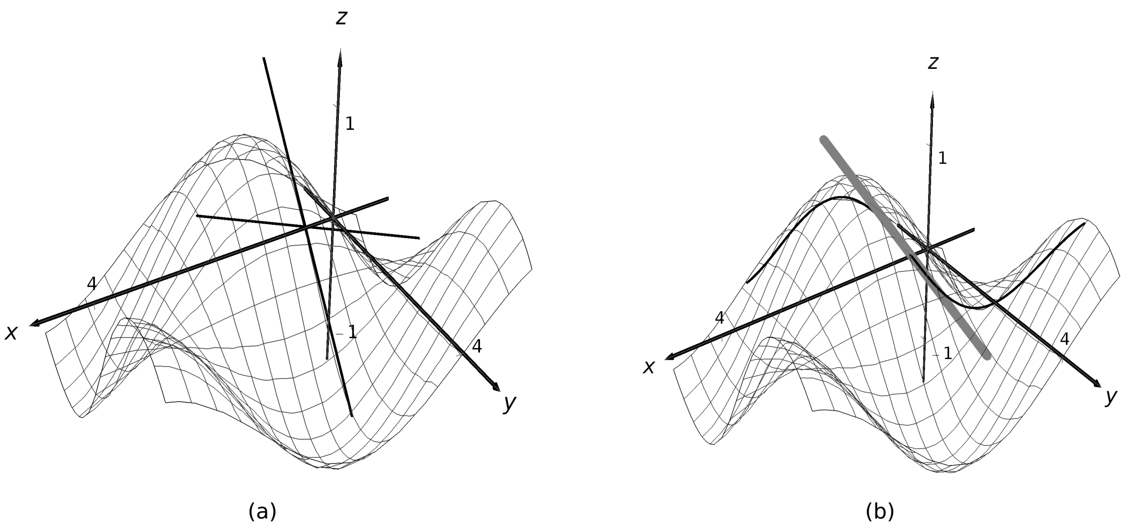


Figure 15.17: A surface and directional tangent lines in Example 15.24.

To find the equation of the tangent line in the direction of $\vec{v}$, we first find the unit vector in the direction of $\vec{v}$: $\vec{u} = \left\langle -1/\sqrt{2}, 1/\sqrt{2} \right\rangle$. The directional derivative at $(\pi/2, \pi/2)$ in the direction of $\vec{u}$ is

$$D_{\vec{u}} f\left(\frac{\pi}{2}, \frac{\pi}{2}\right) = \langle 0, -1 \rangle \cdot \left\langle -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle = -\frac{1}{\sqrt{2}}.$$

Thus the directional tangent line is

$$\ell_{\vec{u}}(t) = \begin{cases} x = \dfrac{\pi}{2} - \dfrac{t}{\sqrt{2}} \\ y = \dfrac{\pi}{2} + \dfrac{t}{\sqrt{2}} \\ z = -\dfrac{t}{\sqrt{2}} \end{cases}.$$

The curve through $(\pi/2,\pi/2,0)$ in the direction of $\vec{\mathbf{v}}$ is shown in Figure 15.17(b) along with $\ell_{\hat{\mathbf{u}}}(t)$.

The following example shows that the points on surfaces where all tangent lines have a slope of 0 can give us some information about the extrema of functions of several variables.

Example 15.25

Let $f(x,y)=4xy-x^4-y^4$. Find the equations of all directional tangent lines to f at $(1,1)$.

Solution

First note that $f(1,1)=2$. We need to compute directional derivatives, so we need $\vec{\nabla}f$. We begin by computing partial derivatives.

$$f_x=4y-4x^3 \quad\Rightarrow\quad f_x(1,1)=0; \qquad f_y=4x-4y^3 \quad\Rightarrow\quad f_y(1,1)=0.$$

Thus $\vec{\nabla}f(1,1)=(0,0)$. Let $\hat{\mathbf{u}}=(u_1,u_2)$ be any unit vector. The directional derivative of f at $(1,1)$ will be $D_{\hat{\mathbf{u}}}f(1,1)=(0,0)\cdot(u_1,u_2)=0$. It does not matter what direction we choose; the directional derivative is always 0. Therefore

$$\ell_{\hat{\mathbf{u}}}(t)=\begin{cases} x=1+u_1t \\ y=1+u_2t \\ z=2. \end{cases}$$

Figure 15.18 shows a graph of f and the point $(1,1,2)$. Note that this point comes at the top of a hill, and therefore every tangent line through this point will have a slope of 0.

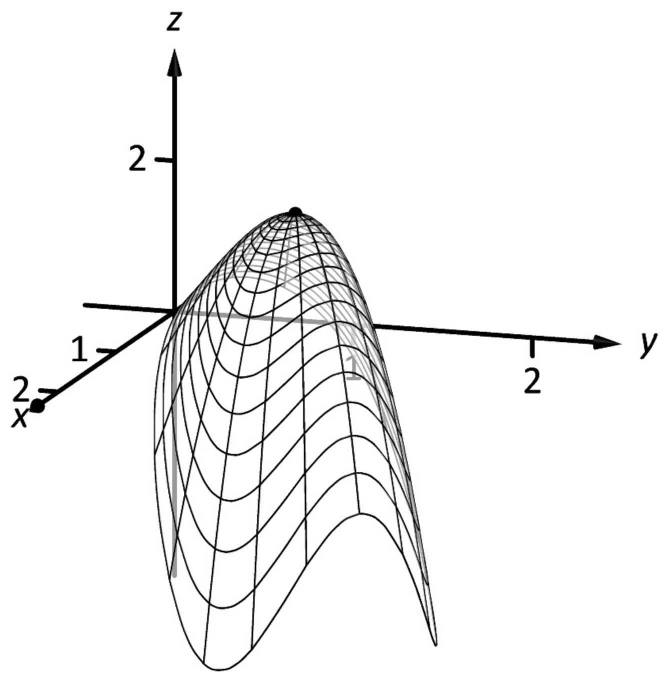


Figure 15.18: Graphing f in Example 15.25.

That is, consider any curve on the surface that goes through this point. Each curve will have a relative maximum at this point, hence its tangent line will have a slope of 0. The following section investigates the points on surfaces where all tangent lines have a slope of 0.

When dealing with a function $y=f(x)$ of one variable, we stated that a line through $(c,f(c))$ was tangent to f if the line had a slope of $f'(c)$ and was normal to f if it had a slope of $-1/f'(c)$. We extend the concept of normal, or orthogonal, to functions of two variables.

Let $z=f(x,y)$ be a differentiable function of two variables. By Definition 15.19, at (x_0,y_0), $\ell_x(t)$ is a line parallel to the vector $\vec{\boldsymbol{d}}_x = \big(1,0,f_x(x_0,y_0)\big)$ and $\ell_y(t)$ is a line parallel to $\vec{\boldsymbol{d}}_y = \big(0,1,f_y(x_0,y_0)\big)$. Since lines in these directions through $\big(x_0,y_0,f(x_0,y_0)\big)$ are tangent to the surface, a line through this point and orthogonal to these directions would be orthogonal, or normal, to the surface. We can use this direction to create a normal line.

The direction of the normal line is orthogonal to $\vec{\boldsymbol{d}}_x$ and $\vec{\boldsymbol{d}}_y$, hence the direction is parallel to $\vec{\boldsymbol{d}}_n = \vec{\boldsymbol{d}}_x \times \vec{\boldsymbol{d}}_y$. It turns out this cross product has a very simple form:

$$\vec{\boldsymbol{d}}_x \times \vec{\boldsymbol{d}}_y = (1,0,f_x) \times \big(0,1,f_y\big) = \big(-f_x,-f_y,1\big).$$

It is often more convenient to refer to the opposite of this direction, namely $\big(f_x,f_y,-1\big)$. This leads to a definition.

Definitie 15.20 (Normal line)
Let $z = f(x,y)$ be differentiable on a set S containing (x_0,y_0) where

$$a = f_x(x_0,y_0) \quad \text{and} \quad b = f_y(x_0,y_0)$$

are defined.

1. A nonzero vector parallel to $\vec{\boldsymbol{n}} = (a,b,-1)$ is orthogonal to f at $P = \big(x_0,y_0,f(x_0,y_0)\big)$.
2. The line ℓ_n through P with direction parallel to $\vec{\boldsymbol{n}}$ is the **normal line** (*normaal*) to f at P.

Thus the parametric equations of the normal line to a surface f at $\big(x_0,y_0,f(x_0,y_0)\big)$ are:

$$\ell_n(t) = \begin{cases} x = x_0 + at \\ y = y_0 + bt \\ z = f(x_0,y_0) - t. \end{cases}$$

The direction of the normal line has many uses, one of which is the definition of the **tangent plane** (*raakvlak*) which we define shortly. Another use is in measuring distances from the surface to a point. Given a point Q in space, it is a general geometric concept to define the distance from Q to the surface as being the length of the shortest line segment $\overrightarrow{PQ}$ over all points P on the surface. This, in turn, implies that $\overrightarrow{PQ}$ will be orthogonal to the surface at P. Therefore we can measure the distance from Q to the surface f by finding a point P on the surface such that $\overrightarrow{PQ}$ is parallel to the normal line to f at P.

Example 15.26

Let $f(x,y) = 2 - x^2 - y^2$ and let $Q = (2,2,2)$. Find the distance from Q to the surface defined by f.

Solution

From Definition 15.20, we know that at (x,y) the direction of the normal line will be $\vec{\boldsymbol{d}}_n = (-2x,-2y,-1)$. A point P on the surface will have coordinates $(x,y,2-x^2-y^2)$, so we have $\overrightarrow{PQ} = \big(2-x,2-y,x^2+y^2\big)$. To find where $\overrightarrow{PQ}$ is parallel to $\vec{\boldsymbol{d}}_n$, we need to find x, y and c such that $c\overrightarrow{PQ} = \vec{\boldsymbol{d}}_n$.

$$\begin{aligned} c\overrightarrow{PQ} &= \vec{\boldsymbol{d}}_n \\ \Rightarrow \quad c\big(2-x,2-y,x^2+y^2\big) &= (-2x,-2y,-1) \end{aligned}$$

This implies

$$c(2-x) = -2x,$$

$$c(2-y) = -2y,$$
$$c(x^2+y^2) = -1.$$

In each equation, we can solve for c:

$$c = \frac{-2x}{2-x} = \frac{-2y}{2-y} = \frac{-1}{x^2+y^2}.$$

The first two fractions imply $x = y$, and so the last fraction can be rewritten as $c = -1/(2x^2)$. Then

$$\begin{array}{rrcl} & \dfrac{-2x}{2-x} & = & \dfrac{-1}{2x^2} \\ \Leftrightarrow & -2x(2x^2) & = & -1(2-x) \\ \Leftrightarrow & 4x^3 & = & 2-x \\ \Leftrightarrow & 4x^3+x-2 & = & 0. \end{array}$$

This last equation is a cubic, which is not difficult to solve with a numeric solver. We find that $x \approx 0.689$, hence $P = (0.689, 0.689, 1.051)$. We find the distance from Q to the surface of f is

$$\left\|\overrightarrow{PQ}\right\| = \sqrt{(2-0.689)^2+(2-0.689)^2+(2-1.051)^2} = 2.083.$$

We can of course take the concept of measuring the distance from a point to a surface to find a point Q a particular distance from a surface at a given point P on the surface.

15.7.2 Tangent plane

We can use the direction of the normal line to define a plane. With $a = f_x(x_0,y_0)$, $b = f_y(x_0,y_0)$ and $P = \big(x_0,y_0,f(x_0,y_0)\big)$, the vector $\vec{n} = (a,b,-1)$ is orthogonal to f at P. The plane through P with normal vector $\vec{n}$ is therefore tangent to f at P.

Definitie 15.21 (Tangent plane)
Let $z = f(x,y)$ be differentiable on a set S containing (x_0,y_0), where $a = f_x(x_0,y_0)$, $b = f_y(x_0,y_0)$, $\vec{n} = (a,b,-1)$ and $P = \big(x_0,y_0,f(x_0,y_0)\big)$.

The plane through P with normal vector $\vec{n}$ is the **tangent plane to** f **at** P (*raakvlak aan f in P*). The standard form of this plane is

$$\begin{array}{rrcl} & \vec{n}\cdot\big((x-x_0),(y-y_0),(z-f(x_0,y_0))\big) & = & 0 \\ \Leftrightarrow & a(x-x_0)+b(y-y_0)-\big(z-f(x_0,y_0)\big) & = & 0. \end{array}$$

Example 15.27

Find the equation of the normal line and tangent plane to $z = -x^2 - y^2 + 2$ at $(0,1)$.

Solution

We find $z_x(x,y) = -2x$ and $z_y(x,y) = -2y$; at $(0,1)$, we have $z_x = 0$ and $z_y = -2$. We take the direction of the normal line, following Definition 15.20, to be $\vec{n} = (0,-2,-1)$. The line with this direction going through the point $(0,1,1)$ is

$$\ell_n(t) = \begin{cases} x = 0 \\ y = 1 - 2t \\ z = 1 - t \end{cases} \quad \text{or} \quad \ell_n(t) = (0,1,1) + (0,-2,-1)\,t.$$

The surface $z = -x^2 - y^2 + 2$, along with the found normal line, is graphed in Figure 15.19(a).

Since we have that $\vec{n} = (0,-2,-1)$ and $P = (0,1,1)$, the equation of the tangent plane is

$$-2(y-1) - (z-1) = 0.$$

The surface $z = -x^2 - y^2 + 2$ and tangent plane are graphed in Figure 15.19(b).

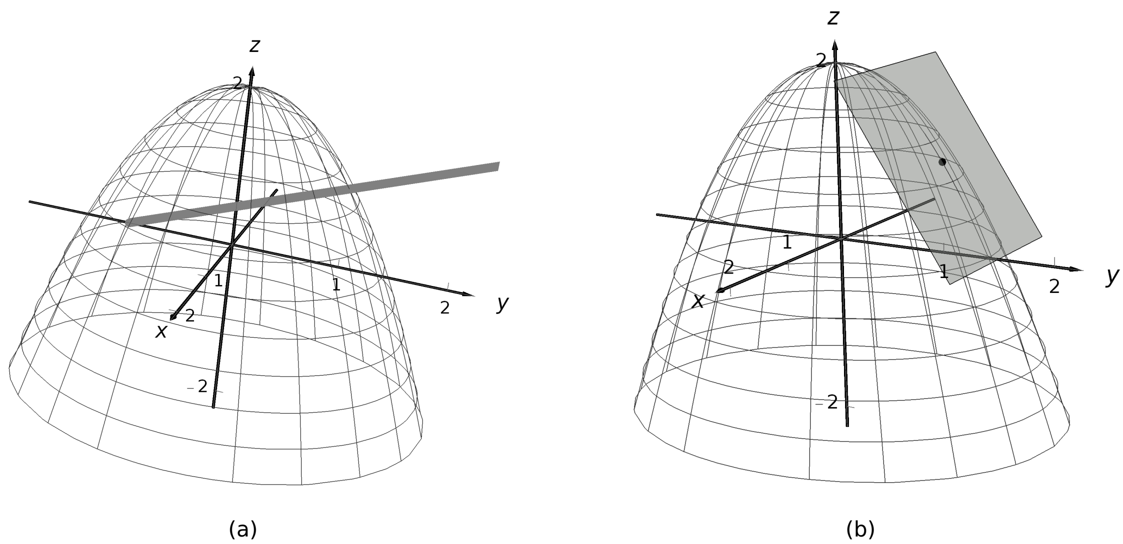


Figure 15.19: Graphing a surface with normal line and tangent plane from Example 15.27.

Just as tangent lines can be used to approximate function values of functions of one variable, tangent planes can be used to achieve this for functions of two variables.

Example 15.28

The point $(3,-1,4)$ lies on the surface of an unknown differentiable function f where $f_x(3,-1) = 2$ and $f_y(3,-1) = -1/2$. Find the equation of the tangent plane to f at P, and use this to approximate the value of $f(2.9,-0.8)$.

Solution

Knowing the partial derivatives at $(3,-1)$ allows us to form the normal vector to the tangent plane, $\vec{n} = (2, -1/2, -1)$. Thus the equation of the tangent line to f at P is:

$$2(x-3) - \frac{1}{2}(y+1) - (z-4) = 0 \quad \Leftrightarrow \quad z = 2(x-3) - \frac{1}{2}(y+1) + 4. \tag{15.10}$$

Just as tangent lines provide excellent approximations of curves near their point of intersection, tangent planes provide excellent approximations of surfaces near their point of intersection. So $f(2.9, -0.8) \approx z(2.9, -0.8) = 3.7$.

This is not a new method of approximation. Compare the right hand expression for z in Equation (15.10) to the total differential:

$$dz = f_x dx + f_y dy \quad \text{and} \quad z = \underbrace{\underbrace{2}_{f_x}\underbrace{(x-3)}_{dx} + \underbrace{-1/2}_{f_y}\underbrace{(y+1)}_{dy}}_{dz} + 4.$$

Thus the new z-value is the sum of the change in z (i.e., dz) and the old z-value (4). As mentioned when studying the total differential, it is not uncommon to know partial derivative information about an unknown function f, and tangent planes are used to give accurate approximations of f.

Recall that when $w = f(x, y, z)$, the gradient $\vec{\nabla} f = \left(f_x, f_y, f_z\right)$ is orthogonal to level surfaces of f. Given a point (x_0, y_0, z_0), let $c = f(x_0, y_0, z_0)$. Then $f(x, y, z) = c$ is a level surface that contains the point (x_0, y_0, z_0) and $\vec{\nabla} f(x_0, y_0, z_0)$ is orthogonal to this level surface. So, the gradient at a point gives a vector orthogonal to the surface at that point. This direction can be used to find tangent planes and normal lines.

Example 15.29

Find the equation of the plane tangent to the ellipsoid

$$\frac{x^2}{12} + \frac{y^2}{6} + \frac{z^2}{4} = 1$$

at $P = (1, 2, 1)$.

Solution

We consider the equation of the ellipsoid as a level surface of a function F of three variables, where $F(x, y, z) = \frac{x^2}{12} + \frac{y^2}{6} + \frac{z^2}{4}$. The gradient is:

$$\begin{aligned} \vec{\nabla} F(x, y, z) &= \left(F_x, F_y, F_z\right) \\ &= \left(\frac{x}{6}, \frac{y}{3}, \frac{z}{2}\right). \end{aligned}$$

At P, the gradient is $\vec{\nabla} F(1, 2, 1) = (1/6, 2/3, 1/2)$. Thus the equation of the plane tangent to the ellipsoid at P is

$$\frac{1}{6}(x-1) + \frac{2}{3}(y-2) + \frac{1}{2}(z-1) = 0.$$

The ellipsoid and tangent plane are graphed in Figure 15.20.

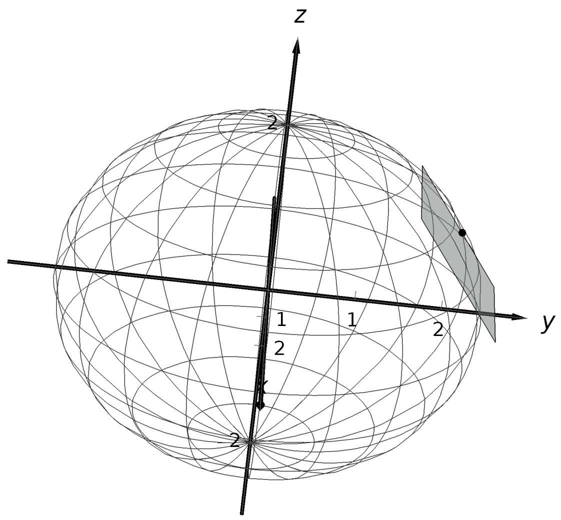


Figure 15.20: An ellipsoid and its tangent plane at a point.

Tangent lines and planes to surfaces have many uses, including the study of instantaneous rates of changes and making approximations. Normal lines also have many uses. In this section we focused on using them to measure distances from a surface. Another interesting application is in computer graphics, where the effects of light on a surface are determined using normal vectors.

The next section investigates another use of partial derivatives: Taylor series expansions of functions of several variables.

15.8 Taylor series expansions

Recall that we found in Section 9.8.3 a way of rewriting a continuous function of one variable as a series by relying on Taylor's theorem. Having introduced all mathematical tools that are needed to analyse functions of several variables, we are now ready to introduce Taylor's theorem for a function of two variables.

Theorem 15.10 (Taylor's theorem for a function of two variables)
Let f be a C^{n+1} function on a set D containing (x_0,y_0). Then, for each (x,y) in D, there exists (θ_x,θ_y) between (x,y) and (x_0,y_0) such that

$$f(x,y)=\sum_{i=0}^{n}\frac{1}{i!}\left(\frac{\partial}{\partial x}(x-x_0)+\frac{\partial}{\partial y}(y-y_0)\right)^i f(x_0,y_0)+R_n(x,y),$$

where

$$f(x,y)=f(x_0,y_0)+\frac{1}{1!}f_x(x_0,y_0)(x-x_0)+\frac{1}{1!}f_y(x_0,y_0)(y-y_0)+\frac{1}{2!}f_{xx}(x_0,y_0)(x-x_0)^2+$$
$$\frac{2}{2!}f_{xy}(x_0,y_0)(x-x_0)(y-y_0)+\frac{1}{2!}f_{yy}(x_0,y_0)(y-y_0)^2+\dots$$

and the remainder term is given by

$$R_n(x,y) = \frac{1}{(n+1)!}\left(\frac{\partial}{\partial x}(x-x_0)+\frac{\partial}{\partial y}(y-y_0)\right)^{n+1} f(\theta_x,\theta_y).$$

As for functions of one variable, the Talor polynomial of degree n provides the best n-th degree polynomial approximation of $f(x,y)$ near a point (x_0,y_0). For instance, letting $n=1$, we obtain

$$f(x,y) \approx f(x_0,y_0)+f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0).$$

Example 15.30

Find a second-degree polynomial approximation to the function

$$f(x,y) = \sqrt{x^2+y^3}$$

near the point $(1,2)$ and use it to estimate the value of $\sqrt{1.02^2+1.97^3}$.

Solution

For a second-degree approximation we need the values of the partial derivatives of f up to the second order at the point $(1,2)$. We have

Derivative function			Derivative at $(1,2)$		
$f(x,y)$	$=$	$\sqrt{x^2+y^3}$	$f(1,2)$	$=$	3
$f_x(x,y)$	$=$	$\frac{x}{\sqrt{x^2+y^3}}$	$f_x(1,2)$	$=$	$\frac{1}{3}$
$f_y(x,y)$	$=$	$\frac{3y^2}{2\sqrt{x^2+y^3}}$	$f_x(1,2)$	$=$	2
$f_{xx}(x,y)$	$=$	$\frac{y^3}{(x^2+y^3)^{3/2}}$	$f_{xx}(1,2)$	$=$	$\frac{8}{27}$
$f_{xy}(x,y)$	$=$	$\frac{-3xy^2}{2(x^2+y^3)^{3/2}}$	$f_{xy}(1,2)$	$=$	$-\frac{2}{9}$
$f_{yy}(x,y)$	$=$	$\frac{12x^2y+3y^4}{4(x^2+y^3)^{3/2}}$	$f_{yy}(1,2)$	$=$	$\frac{2}{3}$.

Thus, we get after evaluating the partial derivatives in $(1,2)$

$$f(x,y) \approx 3+\frac{1}{3}(x-1)+2(y-2)+\frac{4}{27}(x-1)^2-\frac{2}{9}(x-1)(y-2)+\frac{1}{3}(y-2)^2.$$

This is the required second-degree Taylor polynomial for f near $(1,2)$. Therefore,

$$\begin{aligned}
\sqrt{1.02^2+1.97^3} &= f(1+0.02,2-0.03) \\
&\approx 3+\frac{1}{3}(0.02)+2(-0.03)+\frac{4}{27}(0.02)^2-\frac{2}{9}(0.02)(-0.03)+\frac{1}{3}(-0.03)^2 \\
&\approx 2.9471593.
\end{aligned}$$

Clearly, in line with what we devised for functions of one variable, the Taylor series expansion of a function of two variables is given by

$$f(x,y) = \sum_{i=0}^{+\infty} \frac{1}{i!} \left(\frac{\partial}{\partial x}(x - x_0) + \frac{\partial}{\partial y}(y - y_0) \right)^i f(x_0, y_0),$$

and likewise a Maclaurin series expansion can be formulated.

Example 15.31

Find a second-order Taylor series expansion of the function

$$f(x,y) = e^x \ln(1+y),$$

around the point $(0,0)$.

Solution

In order to compute a second-order Taylor series expansion we first compute the necessary partial derivatives and evaluate these derivatives at the origin:

Derivative function			Derivative at $(0,0)$		
$f_x(x,y)$	$=$	$e^x \ln(1+y)$	$f_x(0,0)$	$=$	0
$f_y(x,y)$	$=$	$\dfrac{e^x}{1+y}$	$f_y(0,0)$	$=$	1
$f_{xx}(x,y)$	$=$	$e^x \ln(1+y)$	$f_{xx}(0,0)$	$=$	0
$f_{xy}(x,y)$	$=$	$f_{yx} = \dfrac{e^x}{1+y}$	$f_{xy}(0,0)$	$=$	1
$f_{yy}(x,y)$	$=$	$-\dfrac{e^x}{(1+y)^2}$	$f_{yy}(0,0)$	$=$	-1

Relying on Taylor's theorem, this leads to

$$\begin{aligned} f(x,y) &= 0 + 0(x-0) + 1(y-0) + \frac{1}{2}\Big(0(x-0)^2 + 2(x-0)(y-0) + (-1)(y-0)^2\Big) + \cdots \\ &= y + xy - \frac{y^2}{2} + \cdots \end{aligned}.$$

Finally, we have

$$e^x \ln(1+y) = y + xy - \frac{y^2}{2} + \cdots$$

for $y > -1$.

Of course, Taylor series expansions may be extended to functions of 3 variables.

15.9 Extreme values

In this section we discuss the basic concepts for analyzing extrema of functions of several variables. This will provide us a number of tools that are extremely useful for solving practical engineering problems, such as cost reduction in economics, yield optimization in synthetic biology or learning from data in artificial intelligence, to name just a few. For simplicity we will first discuss functions of two variables, for which concepts can be graphically visualized, before moving to functions of three variables, for which things get a bit more abstract and less intuitive. We start by analyzing necessary conditions for extrema (i.e. conditions that must hold for a point to be a candidate for an extremum), before moving to sufficient conditions (i.e. if such conditions hold, then we are sure that we have an extremum).

15.9.1 Necessary conditions for extrema

Figure 15.21 shows two functions of two variables. The function $f(x,y) = x^2 + y^2$ obviously has a minimum value of 0 (Figure 15.21(a)). This minimum value is found in the point $(x = 0, y = 0)$. Conversely, the function $g(x,y) = 1 - x^2 - y^2$ has a maximum value at $(x = 0, y = 0)$ (Figure 15.21(b)). At this point $g(0,0) = 1$, while g has a smaller value for any other point of its domain. What techniques could be used to find such minima and maxima, if they were not immediately visible from a plot?

Unfortunately, the techniques for finding the minima and maxima of functions of several variables are substantially more complicated than those for functions of one variable.

First of all, we formalise the notion relative and absolute extrema of a function of two variables in agreement with the definition given for functions of one variable (Definitions 10.1 and 10.2).

Definitie 15.22 (Relative and absolute extrema)
Let $z = f(x,y)$ be defined on a set S containing the point $P = (x_0, y_0)$.

1. If $f(x_0,y_0) \geq f(x,y)$ for all (x,y) in S, then f has an **absolute maximum** (*globaal maximum*) at P.

   If $f(x_0,y_0) \leq f(x,y)$ for all (x,y) in S, then f has an **absolute minimum** (*globaal minimum*) at P.

2. If there is an open disk D containing P such that $f(x_0,y_0) \geq f(x,y)$ for all points (x,y) that are in both D and S, then f has a **relative maximum** (*lokaal maximum*) at P.

   If there is an open disk D containing P such that $f(x_0,y_0) \leq f(x,y)$ for all points (x,y) that are in both D and S, then f has a **relative minimum** (*lokaal minimum*) at P.

3. If f has an absolute maximum or minimum at P, then f has an **absolute extremum** at P.

   If f has a relative maximum or minimum at P, then f has a **relative extremum** at P.

Similarly to functions of one variable, $z = f(x,y)$ can have an extremum in critical and singular points and in the end points of the domain of f, in situations where the domain is not $\mathbb{R}$. Critical points of functions of two variables are defined as follows.

Definitie 15.23 (Critical point)
Let $z = f(x,y)$ be continuous on a set S. A **critical point** (*kritisch punt*) $P = (x_0, y_0)$ of f is a point in S such that, at P,

$$f_x(x_0,y_0) = 0 \qquad \text{and} \qquad f_y(x_0,y_0) = 0\,..$$

Besides, when $f_x(x_0,y_0)$ and/or $f_y(x_0,y_0)$ is undefined, we call $P = (x_0,y_0)$ a **singular point** (*singulier punt*), just as we did within the framework of functions of one variable (Definition 10.4).

In what follows we will present methods to find local minima or maxima. The determination of global minima and maxima becomes a trivial task once the local minima and maxima are found. The global minima are simply the points with minimal value among the local mimina (usually we have a single

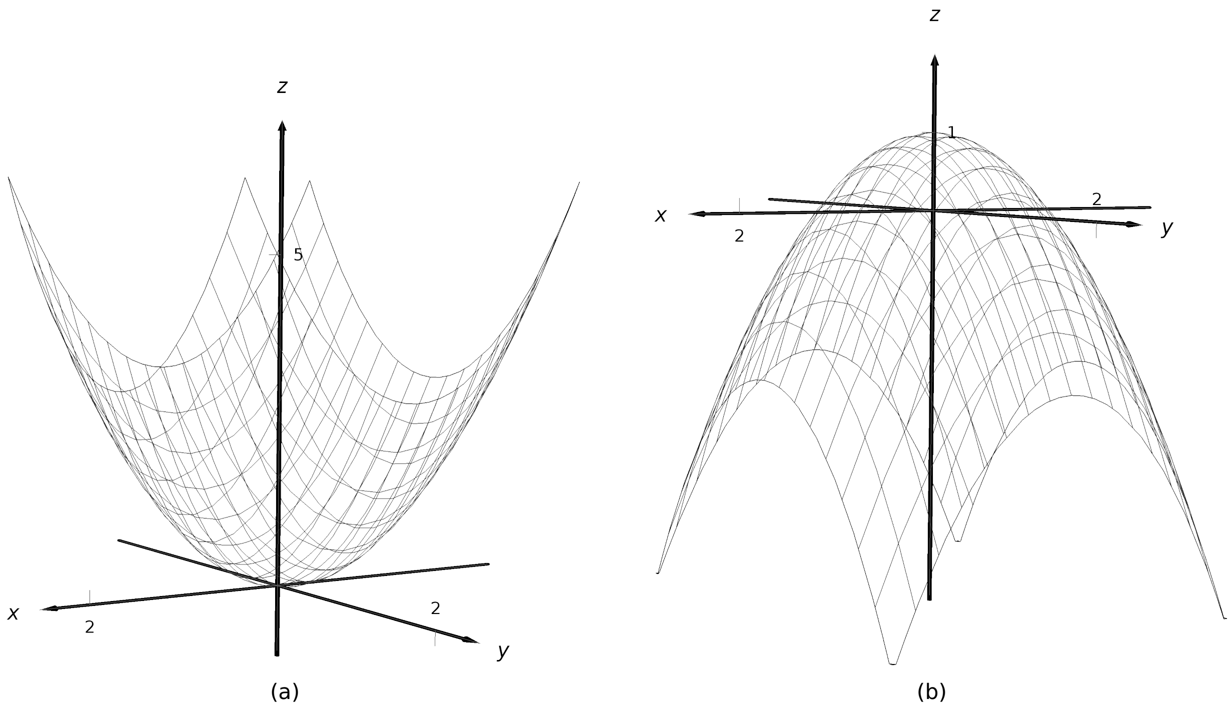


Figure 15.21: A plot of two simple functions that have a single minimum (a) and maximum (b).

global minimum, the exception is the case where two or more points have exactly the same minimal value). Likewise, the global maxima are the points with maximal value among the local maxima. Figure 15.22 shows a plot of a function of two variables where local and global extrema are distinguished.

If f has a relative or absolute maximum at $P = (x_0, y_0)$, it means that every curve on the surface of f through P will also have a relative or absolute maximum at P. Recalling what we learned in Section 10.1, the slopes of the tangent lines to these curves at P must be 0 or undefined. Since directional derivatives are computed using f_x and f_y, we are led to the following theorem.

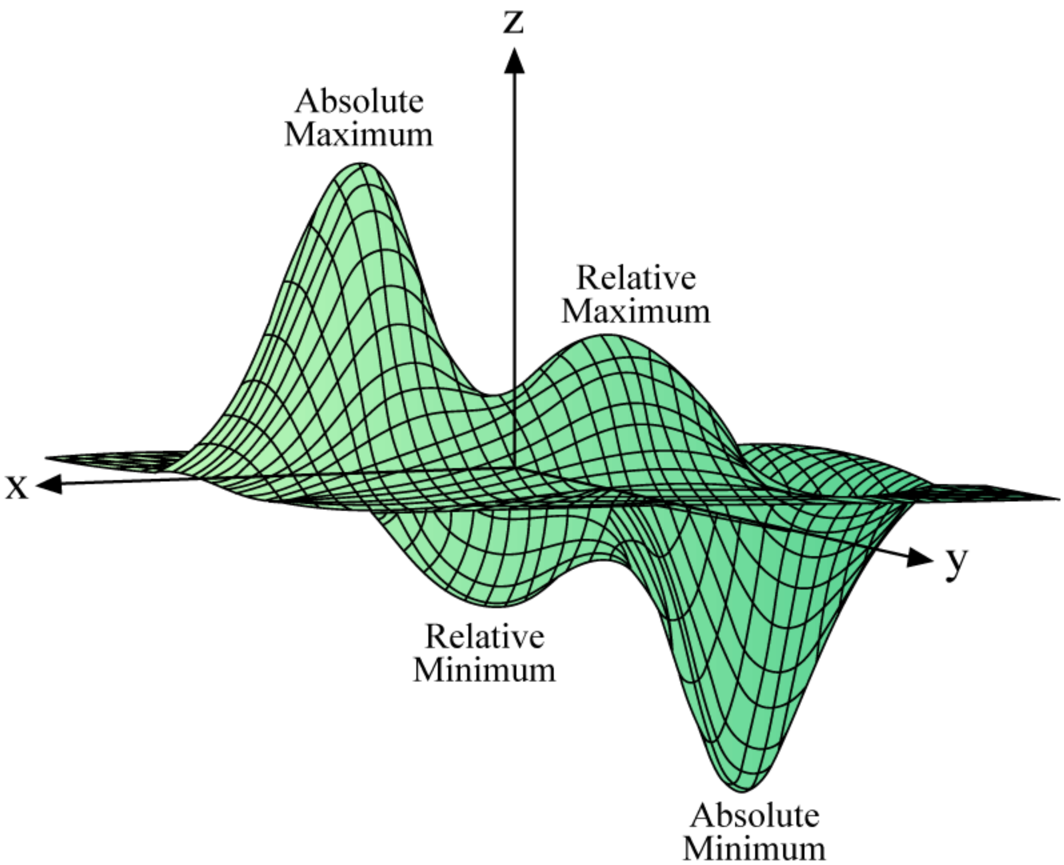


Figure 15.22: A plot of a function with more than one local minimum and maximum. The global minimum and maximum are indicated.

Theorem 15.11 (Necessary Conditions for Extreme Values)
A function $f(x,y)$ can have a local extremum in a point (a,b) of its domain only if (a,b) is one of the following:

1. *a critical point, i.e., a point where $\vec{\nabla} f(a,b) = \vec{\mathbf{0}}$;*
2. *a singular point, i.e., a point where $\vec{\nabla} f(a,b)$ does not exist;*
3. *an end point of the domain of f.*

In this section we will only consider functions with an unrestricted domain. Therefore, to find local extrema, we find the critical and singular points of f and determine which correspond to local maxima, local minima, or neither. The examples below will demonstrate this process. For functions with a restricted domain, the analysis will be require specific techniques, which are discussed in the next chapter. Also, remark that Theorem 15.11 remains valid for functions of more than two variables, with identical arguments as the ones given in its proof.

Example 15.32

Let $f(x,y) = -\sqrt{x^2+y^2}+2$. Find the relative extrema of f. The surface of f is graphed in Figure 15.23 along with the point $(0,0,2)$.

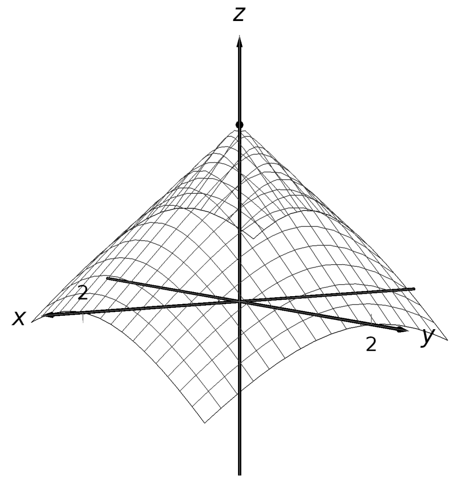


Figure 15.23: The surface in Example 15.32 with its absolute maximum indicated.

Solution

We start by computing the partial derivatives of f:

$$f_x(x,y) = \frac{-x}{\sqrt{x^2+y^2}} \qquad \text{and} \qquad f_y(x,y) = \frac{-y}{\sqrt{x^2+y^2}}.$$

It is clear that $f_x = 0$ when $x = 0$ and $y \neq 0$, and that $f_y = 0$ when $y = 0$ and $x \neq 0$. At $(0,0)$, both f_x and f_y are not 0, but undefined. The point $(0,0)$ is hence a singular point, though. The graph in Figure 15.23 shows that this point is the absolute maximum of f.

In this example we found a critical point of f and then determined whether or not it was a local (or global) maximum or minimum by graphing. It would be nice to be able to determine whether a critical point corresponded to a max or a min without a graph. Before we develop such a test, we do two more examples that shed more light on the issues our test needs to consider.

Example 15.33

Let $f(x,y) = x^3 - 3x - y^2 + 4y$. Find the relative extrema of f.

Solution

Once again we start by finding the partial derivatives of f:

$$f_x(x,y) = 3x^2 - 3 \quad \text{and} \quad f_y(x,y) = -2y + 4.$$

Each is always defined. Setting each equal to 0 and solving for x and y, we find

$$\begin{aligned} f_x(x,y) = 0 \quad &\Leftrightarrow \quad x = \pm 1\,, \\ f_y(x,y) = 0 \quad &\Leftrightarrow \quad y = 2. \end{aligned}$$

We have two critical points: $(-1,2)$ and $(1,2)$, while there are no singular points. To determine if they correspond to a relative maximum or minimum, we consider the graph of f in Figure 15.24.

The critical point $(-1,2)$ clearly corresponds to a relative maximum. However, the critical point at $(1,2)$ is neither a maximum nor a minimum, displaying a different, interesting characteristic.

If one walks parallel to the y-axis towards this critical point, then this point becomes a relative maximum along this path. But if one walks towards this point parallel to the x-axis, this point becomes a relative minimum along this path. A point that seems to act as both a maximum and a minimum is a saddle point. A formal definition follows.

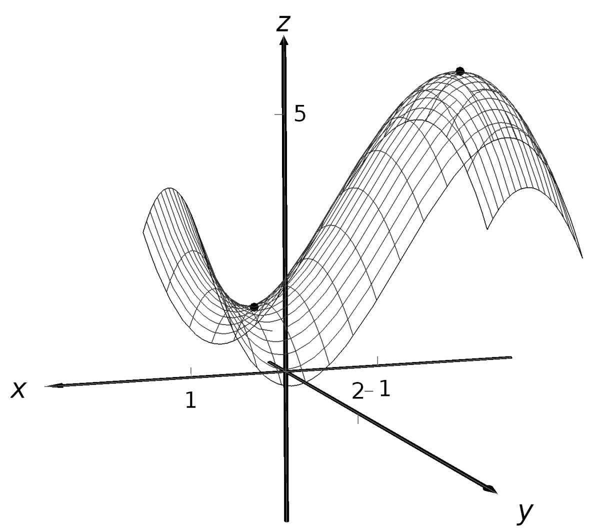


Figure 15.24: The surface in Example 15.33 with both critical points marked.

Definitie 15.24 (Saddle point)
Let $P = (x_0, y_0)$ be in the domain of f where $f_x = 0$ and $f_y = 0$ at P. We say P is a **saddle point** (*zadelpunt*) of f if, for every open disk D containing P, there are points (x_1, y_1) and (x_2, y_2) in D such that $f(x_0, y_0) > f(x_1, y_1)$ and $f(x_0, y_0) < f(x_2, y_2)$.

At a saddle point, the instantaneous rate of change in all directions is 0 and there are points nearby with z-values both less than and greater than the z-value of the saddle point.

Example 15.34

Let us try to classify the critical points of $f(x,y) = 2x^3 - 6xy + 3y^2$. As in the previous examples we start by setting the partial derivatives of f to zero, resulting in the following nonlinear system:

$$\begin{aligned} f_x(x,y) &= 6x^2 - 6y = 0\,, \\ f_y(x,y) &= -6x + 6y = 0\,. \end{aligned}$$

These partial derivatives are defined everywhere, so we have no singular points. To analyze the critical points, let us first put f_y to zero, which yields $x = y$. Setting f_x to zero and utilizing this observation then gives

$$6x^2 - 6x = 0 \qquad \Leftrightarrow \qquad 6x(x-1) = 0\,.$$

As a result, x must be 0 or 1. In both cases we also find a solution for y when putting f_y to zero. We obtain in this way two critical points: $(0,0)$ and $(1,1)$.

Let us start by analyzing the point $(0,0)$. If this point is a minimum, then the value of the function has to increase when we move away from $(0,0)$ with a small step, irrespective of the direction. Similarly, if the point is a maximum, then the function must decrease if we move in any direction. So, let us assume that we move to the point (h,k) from $(0,0)$ and let us analyze the difference in function values.

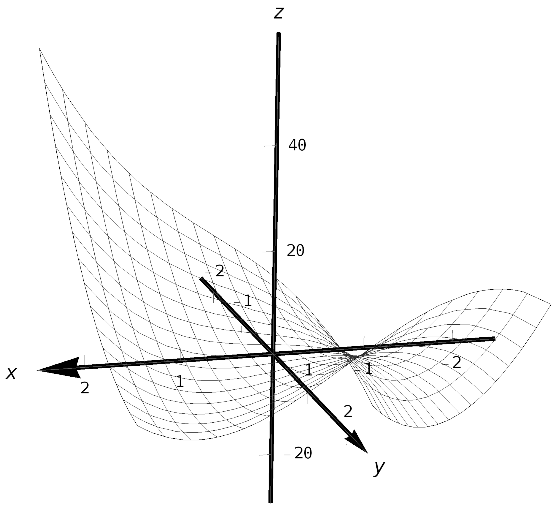
$$\Delta f = f(h,k) - f(0,0) = 2h^3 - 6hk + 3k^2$$

This difference Δf does not have a fixed sign. It is sometimes negative, e.g. for $(h = -0.1, k = 0)$, and sometimes positive, e.g. for $(h = 0, k = 0.1)$. So the point $(0,0)$ can neither be a local maximum, nor a local minimum. The point is a saddle point because the function increases when we move in a certain direction, while it decreases when we move in another direction.

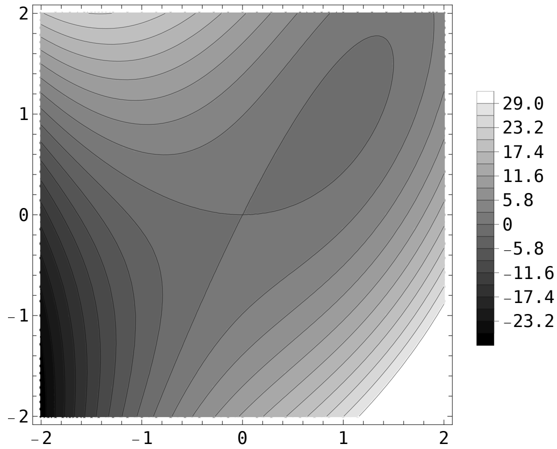
In a similar way let us analyze the point $(1,1)$. Again, we analyze the difference between the function in $(1,1)$ and a point close to $(1,1)$, namely the point $(1+h, 1+k)$ in which h and k take positive or negative values that are close to zero (both values have to be close to zero because we are investigating local extrema, so we want to analyze the neighborhood of a critical point). A few manipulations will yield the following:

$$\begin{aligned} \Delta f &= f(1+h, 1+k) - f(1,1) \\ &= 2(1+h)^3 - 6(1+h)(1+k) + 3(1+k)^2 - (-1) \\ &= \cdots \\ &= 3(h-k)^2 + h^2(3+2h). \end{aligned}$$

For h and k close to zero this Δf has a fixed sign. The first term is a square, and hence always positive. The second term is positive when $3 + 2h > 0$, so when $h > -3/2$. Δf is always positive for h and k close to zero, so $f(1+h, 1+k) \geq f(1,1)$ in that case. We are sure that $(1,1)$ is a local minimum. Figure 15.25 plots the function f, and one can see that $(1,1)$ indeed is a point where f reaches a local minimum.



(a)



(b)

Figure 15.25: A plot of the function $f(x,y) = 2x^3 - 6xy + 3y^2$: graph of $z = f(x,y)$ (a) and level curves (b).

15.9.2 Sufficient conditions for extrema

15.9.2.1 Functions of two variables

Before Example 15.33 we mentioned the need for a test to differentiate between relative maxima and minima. We now recognize that our test also needs to account for saddle points. To do so, we consider the second partial derivatives of f. Recall that with single variable functions, such as $y = f(x)$, if $f''(c) > 0$, then if f is concave up at c, and if $f'(c) = 0$, then f has a relative minimum at $x = c$. Note that at a saddle point, it seems the graph is both concave up and concave down, depending on which direction you are considering.

It would be nice if the following were true:

$$\begin{aligned} f_{xx} \text{ and } f_{yy} > 0 \quad &\Rightarrow \quad \text{relative minimum,} \\ f_{xx} \text{ and } f_{yy} < 0 \quad &\Rightarrow \quad \text{relative maximum,} \\ f_{xx} \text{ and } f_{yy} \text{ have opposite signs} \quad &\Rightarrow \quad \text{saddle point.} \end{aligned}$$

However, this is not the case. Functions f exist where f_{xx} and f_{yy} are both positive but a saddle point still exists. In such a case, while the concavity in the x-direction is up (i.e., $f_{xx} > 0$) and the concavity in the y-direction is also up (i.e., $f_{yy} > 0$), the concavity switches somewhere in between the x- and y-directions.

In Example 15.34 we presented a simple procedure to determine whether a critical point corresponds to a local minimum, local maximum or a saddle point. This procedure was rather ad-hoc, because some arithmatic manipulations were needed to bring Δf in a shape, so that we could analyze whether Δf has a fixed sign. In addition to being ad-hoc, the procedure is also not very scalable, because the complexity of those manipulations increases when moving to functions of three variables. We will need a more systematic test to classify critical points. Remember from Theorem 15.11 that we need to investigate for that purpose the critical points, singular points and end points of the domain. To analyze the critical and singular points, the gradient $\vec{\nabla} f$ has to be analyzed and we have to set both partial derivatives to zero, which leads to a system of two equations. The solutions of this system will be the critical points, which will need further investigation. It is at this stage that so-called Hessian matrices come into play.

Definitie 15.25 (Hessian Matrix (of a function of two variables))
Suppose that $z = f(x, y)$ is a function that is twice differentiable w.r.t. all variables x and y Then, **the Hessian matrix** $\mathscr{H}(x, y)$ (*Hessiaan (matrix)*) of f in the point (x, y) is defined as

$$\mathscr{H}(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}.$$

This Hessian matrix is the foundation for the so-called second derivative test that allows to classify critical points.

Theorem 15.12 (Second derivative test)
Suppose that $P = (x_0, y_0) \in \mathbb{R}^2$ is a critical point of the function $f(x, y)$ and lying in the interior of the domain of f. Also, suppose that all the second partial derivatives are continuous throughout a neighborhood of P so that the Hessian matrix is also continuous in that neighborhood.

1. *If $\mathscr{H}(x_0, y_0)$ is positive definite, then f has a local minimum at P;*
2. *If $\mathscr{H}(x_0, y_0)$ is negative definite, then f has a local maximum at P;*
3. *If $\mathscr{H}(x_0, y_0)$ is indefinite, then f has a saddle point at P;*
4. *If $\mathscr{H}(x_0, y_0)$ is neither positive definite, nor negative definite, nor indefinite, then this test gives no information.*

Note that the continuity of the partials guarantees that $\mathscr{H}(x_0, y_0)$ is a symmetric matrix.

Clearly, it is crucial to know the nature of $\mathscr{H}(x_0, y_0)$ in terms of it definiteness. Fortunately, we can rely on the following theorem for this purpose.

Theorem 15.13 (Definiteness of symmetric matrices)
Let A be a real, symmetric $m \times m$ matrix. Let us consider the sequence $D_1, D_2, \ldots, D_m$ of determinants defined by

$$D_1 = \begin{vmatrix} a_{11} \end{vmatrix}, \quad D_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad D_i = \begin{vmatrix} a_{11} & \ldots & a_{1i} \\ \vdots & & \vdots \\ a_{i1} & \ldots & a_{ii} \end{vmatrix},$$

where $1 \le i \le m$. Then the following hold:

- *If $D_i > 0$ for $1 \le i \le m$, then A is positive definite;*
- *If $D_i > 0$ for all even i and $D_i < 0$ for all odd i, then A is negative definite;*
- *If $D_m = \det(A) \neq 0$ and the above situations do not apply, then A is indefinite;*
- *If $D_m = \det(A) = 0$, then A is positive or negative semi-definite.*

We now first practice using this test with some of the functions in the previous examples, where we visually determined whether a critical point corresponded to a local minimum, local maximum or saddle point.

Example 15.35

Let $f(x, y) = x^3 - 3x - y^2 + 4y$ be as in Example 15.33 and let us investigate whether the function has a local minimum, local maximum, or saddle point at each critical point. We determined previously that the critical points of f are $(-1,2)$ and $(1,2)$. To use the second derivative test, we must find the second partial derivatives of f:

$$f_{xx}(x,y) = 6x, \qquad f_{yy}(x,y) = -2, \qquad f_{xy}(x,y) = 0.$$

Organizing them in the Hessian matrix gives

$$\mathscr{H}(x,y) = \begin{bmatrix} 6x & 0 \\ 0 & -2 \end{bmatrix}.$$

For the critical point $(-1,2)$ one obtains the following matrix:

$$\mathscr{H}(-1,2) = \begin{bmatrix} -6 & 0 \\ 0 & -2 \end{bmatrix}.$$

Applying Theorem 15.13 gives $D_1 = -6 < 0$ and $D_2 = 12 > 0$, so the matrix is negative definite and we have a local maximum. For the critical point $(1,2)$ one obtains

$$\mathscr{H}(1,2) = \begin{bmatrix} 6 & 0 \\ 0 & -2 \end{bmatrix}.$$

Applying Theorem 15.13 gives in this case $D_1 = 6 > 0$ and $D_2 = -12 < 0$, so the matrix is indefinite and we have a saddle point. The second derivative test confirmed what we determined visually.

Example 15.36

Let $f(x,y) = 2x^3 - 6xy + 3y^2$ be as in Example 15.34. We found two critical points $(0,0)$ and $(1,1)$. Let us first construct the Hessian matrix for an arbitrary point (x,y).

$$\mathscr{H}(x,y) = \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix} = \begin{bmatrix} 12x & -6 \\ -6 & 6 \end{bmatrix}$$

For the critical point $(0,0)$ one obtains the following matrix:

$$\mathscr{H}(0,0) = \begin{bmatrix} 0 & -6 \\ -6 & 6 \end{bmatrix}.$$

Applying Theorem 15.13 gives $D_1 = 0$ and $D_2 = -36 \neq 0$, so the matrix is indefinite and we have a saddle point. For the critical point $(1,1)$ one obtains

$$\mathscr{H}(1,1) = \begin{bmatrix} 12 & -6 \\ -6 & 6 \end{bmatrix}.$$

Applying Theorem 15.13 gives in this case $D_1 = 12 > 0$ and $D_2 = 36 > 0$, so the matrix is positive definite and we have a local minimum.

15.9.2.2 Functions of three variables

To determine the extrema of a function of three variables $w = f(x,y,z)$, we proceed in the same way as before upon intuitively extending Theorem 15.11 the definition of the Hessian to such functions.

Definitie 15.26 (Hessian Matrix (of a function of three variables))
Suppose that $w = f(x,y,z)$ is a function that is twice differentiable w.r.t. all variables x, y and z. Then, **the Hessian matrix** $\mathscr{H}(x,y,z)$ (*Hessiaan (matrix)*) of f in the point (x,y,z) is defined as

$$\mathscr{H}(x,y,z) = \begin{bmatrix} f_{xx}(x,y,z) & f_{xy}(x,y,z) & f_{xz}(x,y,z) \\ f_{yx}(x,y,z) & f_{yy}(x,y,z) & f_{yz}(x,y,z) \\ f_{zx}(x,y,z) & f_{zy}(x,y,z) & f_{zz}(x,y,z) \end{bmatrix}.$$

Combining this definition with the intuitive extension of Theorem 15.12 to functions of three variables allows us to classify the critical points of such functions.

Example 15.37

Determine the nature of the extrema of $f(x,y,z) = x^2y + y^2z + z^2 - 2x$.

Solution

We start by computing the partial derivatives and equating them to zero, which results in the following system:

$$\begin{aligned} f_x(x,y,z) &= 2xy - 2 = 0\,, \\ f_y(x,y,z) &= x^2 + 2yz = 0\,, \\ f_z(x,y,z) &= y^2 + 2z = 0\,. \end{aligned}$$

From the third equation we learn that $z = -y^2/2$. Substituting this in the second equation gives $x^2 - y^3 = 0$. Substituting this last result in the first equation gives $y^{5/2} = 1$. As a result we have $y = 1$ and we find one critical point $(1,1,-1/2)$. Let us compute the Hessian in that point:

$$\mathscr{H}(x,y,z) = \begin{bmatrix} 2y & 2x & 0 \\ 2x & 2z & 2y \\ 0 & 2y & 2 \end{bmatrix} \quad \text{and hence} \quad \mathscr{H}\left(1,1,-\frac{1}{2}\right) = \begin{bmatrix} 2 & 2 & 0 \\ 2 & -1 & 2 \\ 0 & 2 & 2 \end{bmatrix}.$$

Theorem 15.13 results in $D_1 = 2$, $D_2 = -6$ and $D_3 = -20$. The matrix is indefinite, and the critical point is a saddle point.

15.10 Constrained optimisation

In the previous chapter we have analyzed extrema of functions up to three variables without making any restrictions on the domain. In this section we further specialize in functions with a restricted domain. This portion of mathematics is often entitled **constrained optimization** (*gebonden extremumproblemen*) because we want to optimize a function (i.e. find its maximum and/or minimum values) subject to a constraint – some limit to what values the function can attain.

Solving constrained optimization problems is a very important topic in applied mathematics, and we will discuss some applications in this chapter. In Section 15.10.1 we first present some general techniques for finding extrema in a restricted domain, before discussing the specific but widely-applicable case of linear functions in Section 15.10.2. We will mainly focus on functions of two variables to keep things simple, but the techniques developed here are the basis for solving larger problems, where more than two variables are involved.

15.10.1 Extrema in a restricted domain

When optimizing functions of one variable such as $y = f(x)$, you made use of the extreme value theorem (Theorem 10.1), which said that over a closed interval I, a continuous function has both a maximum and minimum value.

A similar theorem applies to functions of several variables over a closed set (see Definition 15.5). We can find these values by evaluating the function at the critical and singular points in the set and over the boundary of the set. After formally stating this extreme value theorem for functions of three variables, we give examples.

Theorem 15.14 (Extreme Value Theorem)
Let $w = f(x, y, z)$ be a continuous function defined on a closed, bounded set S. Then f has a global maximum and minimum value on S.

Example 15.38

Let $f(x, y) = x^2 - y^2 + 5$ and let S be the triangle with vertices $(-1, -2)$, $(0, 1)$ and $(2, -2)$. Determine the maximum and minimum values of f on S.

Solution

It can help to see a graph of f along with the set S. In Figure 15.26(a) the triangle defining S is shown on the surface of f, as we are only concerned with the portion of f enclosed by the triangle on its surface.

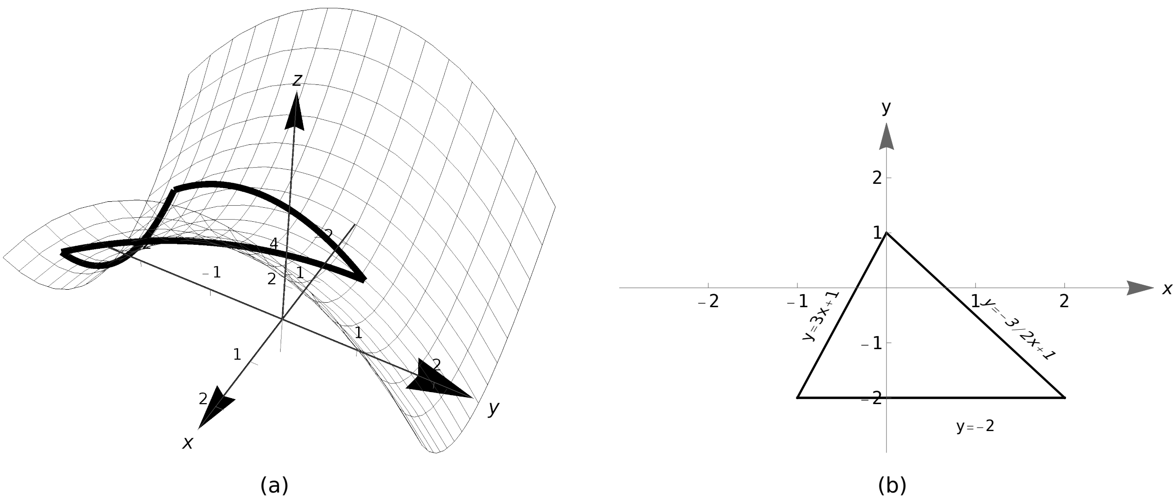


Figure 15.26: Plot of the surface of f in Example 15.38 (a), along with the restricted domain S (b).

We begin by finding the critical points of f. With $f_x = 2x$ and $f_y = -2y$, we find only one critical point, at $(0, 0)$, which is situated in S.

We now find the maximum and minimum values that f attains along the boundary of S, that is, along the edges of the triangle. In Figure 15.26(b) we see the triangle sketched in the plane with the equations of the lines forming its edges labeled. Start with the bottom edge, along the line $y = -2$. If $y = -2$, then on the surface, we are considering points $f(x, -2)$; that is, our function reduces to $f(x, -2) = x^2 - (-2)^2 + 5 = x^2 + 1 = f_1(x)$. We want to maximize/minimize $f_1(x) = x^2 + 1$ on the interval $[-1, 2]$. To do so, we evaluate $f_1(x)$ at its critical points and at the endpoints. The critical points of f_1 are found by setting its derivative equal to 0:

$$f_1'(x) = 2x = 0 \qquad \Leftrightarrow \qquad x = 0.$$

Evaluating f_1 at this critical point, and at the endpoints of $[-1, 2]$ gives:

$$\begin{aligned} f_1(-1) = 2 \qquad &\Rightarrow \qquad f(-1,-2) = 2, \\ f_1(0) = 1 \qquad &\Rightarrow \qquad f(0,-2) = 1, \\ f_1(2) = 5 \qquad &\Rightarrow \qquad f(2,-2) = 5. \end{aligned}$$

Notice how evaluating f_1 at a point is the same as evaluating f at its corresponding point.

We need to do this process twice more, for the other two edges of the triangle. Along the left edge, along the line $y = 3x + 1$, we substitute $3x + 1$ in for y in $f(x, y)$:

$$f(x,y) = f(x,3x+1) = x^2 - (3x+1)^2 + 5 = -8x^2 - 6x + 4 = f_2(x).$$

We want the maximum and minimum values of f_2 on the interval $[-1, 0]$, so we evaluate f_2 at its critical points and the endpoints of the interval. We find the critical points:

$$f_2'(x) = -16x - 6 = 0 \qquad \Leftrightarrow \qquad x = -3/8.$$

Evaluate f_2 at its critical point and the endpoints of $[-1, 0]$:

$$\begin{aligned} f_2(-1) = 2 \qquad &\Rightarrow \qquad f(-1,-2) = 2, \\ f_2(-3/8) = 5.125 \qquad &\Rightarrow \qquad f(-3/8,-1/8) = 5.125, \\ f_2(0) = 1 \qquad &\Rightarrow \qquad f(0,1) = 4. \end{aligned}$$

Finally, we evaluate f along the right edge of the triangle, where $y = -\frac{3}{2}x + 1$.

$$f(x,y) = f\left(x, -\frac{3}{2}x + 1\right) = x^2 - \left(-\frac{3}{2}x + 1\right)^2 + 5 = -\frac{5}{4}x^2 + 3x + 4 = f_3(x).$$

The critical points of $f_3(x)$ are:

$$f_3'(x) = -\frac{5}{2}x + 3 = 0 \qquad \Leftrightarrow \qquad x = \frac{6}{5} = 1.2.$$

We evaluate f_3 at this critical point and at the endpoints of the interval $[0, 2]$:

$$\begin{aligned} f_3(0) = 4 \qquad &\Rightarrow \qquad f(0,1) = 4, \\ f_3(1.2) = 5.8 \qquad &\Rightarrow \qquad f(1.2,-0.8) = 5.8, \\ f_3(2) = 5 \qquad &\Rightarrow \qquad f(2,-2) = 5. \end{aligned}$$

One last point to test: the critical point of f, $(0, 0)$. We find $f(0, 0) = 5$.

We have evaluated f at a total of 7 different places. We checked each vertex of the triangle twice, as each showed up as the endpoint of an interval twice. Of all the z-values found, the maximum is 5.8, found at $(1.2, -0.8)$; the minimum is 1, found at $(0, -2)$.

We illustrate the technique once more with a classic problem that also emphasizes the practical usefulness of constrained optimization.

Example 15.39

Post NL states that the girth pluslength of a standard post package must not exceed 130. Given a rectangular box, the length is the longest side, and the so-called girth is twice the width+height. Given a rectangular box where the width and height are equal, what are the dimensions of the box that give the maximum volume subject to the constraint of the size of a standard post package?

Solution

Let w, h and ℓ denote the width, height and length of a rectangular box; we assume here that $w = h$. The girth is then $2(w+h) = 4w$. The volume of the box is $V(w,\ell) = wh\ell = w^2\ell$. We want to maximize this volume subject to the constraint $4w + \ell \leq 130$, or $\ell \leq 130 - 4w$. Common sense also indicates that $\ell \geq 0, w \geq 0$. Moreover, for $\ell = 0$ or $w = 0$ one cannot obtain a maximum, so this part of the boundary does not need further investigation.

We begin by finding the critical values of V. We find that $V_w = 2w\ell$ and $V_\ell = w^2$; these are simultaneously 0 only at $(0,0)$. This gives a volume of 0, so we can ignore this critical point.

We now consider the volume along the constraint $\ell = 130 - 4w$. Along this line, we have:

$$V(w,\ell) = V(w, 130-4w) = w^2(130-4w) = 130w^2 - 4w^3 = V_1(w).$$

The constraint is applicable on the w-interval $[0, 130/4]$ as indicated in the figure. Thus we want to maximize V_1 on $[0, 32.5]$. Finding the critical values of V_1, we take the derivative and set it equal to 0:

$$V_1'(w) = 260w - 12w^2 = 0 \quad \Leftrightarrow \quad w(260-12w) = 0 \quad \Leftrightarrow \quad w = 0 \vee \frac{260}{12} \approx 21.67.$$

We found two critical values: when $w = 0$ and when $w = 21.67$. We again ignore the $w = 0$ solution, because then $V = 0$. The maximum volume, subject to the constraint, comes at $w = h = 21.67$, $\ell = 130 - 4(21.67) = 43.33$. This gives a volume of $V(21.67, 43.33) \approx 20.343\text{in}^3$.

The volume function $V(w,\ell)$ is shown in Figure 15.27 along with the constraint $\ell = 130 - 4w$. As done previously, the constraint is drawn dashed in the xy-plane and also along the surface of the function. The point where the volume is maximized is indicated.

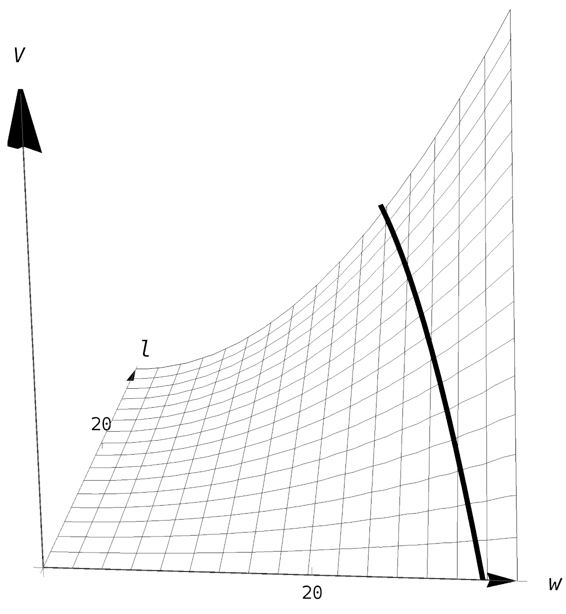


Figure 15.27: Graphing the volume of a box with girth $4w$ and length ℓ, subject to a size constraint.

In the previous two examples we have deployed the same procedure to find the global maximum and minimum in a restricted domain. In essence this procedure consists of three steps.

1. Find any critical or singular points in the interior of the domain S.
2. Find any points on the boundary of S where f might have extreme values. To do so, you can parameterize the whole boundary, or parts of it, and express f as a function of the parameters. If you break the boundary into pieces, you must analyze the end points of those pieces.
3. Evaluate f on all the points found in steps 1 and 2.

15.10.2 Linear programming

It is hard to overemphasize the importance of optimization. In “the real world”, we routinely seek to make something better. By expressing the something as a mathematical function, “making something better” means “optimize some function”. In this section we further specialize in methods to achieve the best outcome (such as maximum profit or lowest cost) in a mathematical model whose requirements are represented by linear relationships.

More formally, linear programming is a technique for the optimization of a linear objective function, subject to linear equality and linear inequality constraints. Its feasible region is a convex polytope, which is a special type of convex set, defined by one or several linear inequalities. Its objective function is a real-valued linear function defined on this convex polytope. A linear programming algorithm finds a point in the convex polytope where this function has the smallest (or largest) value if such a point exists. Linear programs are problems that can be expressed in canonical form as

$$\begin{array}{ll} \text{Maximize} & \vec{\boldsymbol{c}}^T\vec{\boldsymbol{x}}, \\ \text{subject to} & A\vec{\boldsymbol{x}} \leq \vec{\boldsymbol{b}}, \\ \text{and} & \vec{\boldsymbol{x}} \geq \vec{\boldsymbol{0}}. \end{array}$$

where $\vec{\boldsymbol{x}}$ represents the vector of variables (to be determined), $\vec{\boldsymbol{c}}$ and $\vec{\boldsymbol{b}}$ are vectors of (known) coefficients, A is a (known) matrix of coefficients. The expression to be maximized or minimized is called the objective function ($\vec{\boldsymbol{c}}^T\vec{\boldsymbol{x}}$ in this case). The inequalities $A\vec{\boldsymbol{x}} \leq \vec{\boldsymbol{b}}$ and $\vec{\boldsymbol{x}} \geq \vec{\boldsymbol{0}}$ are the constraints which specify a convex polytope over which the objective function is to be optimized. In this context, two vectors are comparable when they have the same dimensions. If every entry in the first is less-than or equal-to the corresponding entry in the second, then it can be said that the first vector is less-than or equal-to the second vector.

Linear programming can be applied to various fields of study. It is widely used in mathematics, and to a lesser extent in business, economics, and for some engineering problems. Industries that use linear programming models include transportation, energy, telecommunications, and manufacturing. It has proven useful in modeling diverse types of problems in planning, routing, scheduling, assignment, and design.

Example 15.40

A store has requested a manufacturer to produce pants and sports jackets. For materials, the manufacturer has 750 m^2 of cotton textile and 1000 m^2 of polyester. Every pair of pants (1 unit) needs 1m^2 of cotton and 2 m^2 of polyester. Every jacket needs 1.5m^2 of cotton and 1 m^2 of polyester. The price of the pants is fixed at €50 and the jacket at €40. What is the number of pants and jackets that the manufacturer must give to the stores so that these items obtain a maximum sale?

Solution

To solve this problem, let us start with defining the unknowns.

$$\begin{aligned} x &= \text{number of pants} \\ y &= \text{number of jackets} \end{aligned}$$

In the second step the objective function needs to be constructed. The manufacturer wants to have a maximum sale, so the following objective needs to be maximized.

$$f(x,y) = 50x + 40y$$

We have to maximize this function subject to a number of constraints. With the unknowns that we have defined, those constraints become

$$\begin{aligned} x + 1.5y &\leq 750\,, \\ 2x + y &\leq 1000\,. \end{aligned}$$

Simplifying the first constraints, and observing that the number of pants and jackets are natural numbers, we end up with the following set of four constraints:

$$\begin{aligned} 2x + 3y &\leq 1500\,, \\ 2x + y &\leq 1000\,, \\ x &\geq 0\,, \\ y &\geq 0\,. \end{aligned}$$

Those four constraints define the restricted domain of f, in which a maximum needs to be found.

In the next step, let us represent the restricted domain in a two-dimensional graph. As $x \geq 0$ and $y \geq 0$, we work in the first quadrant. Next, we plot the other two constraints as straight lines by replacing the inequality by an equality. To draw these lines, determine their points of intersection with the axes, as shown in Figure 15.28. Together the four constraints define a feasible region that represents the restricted domain of the function f.

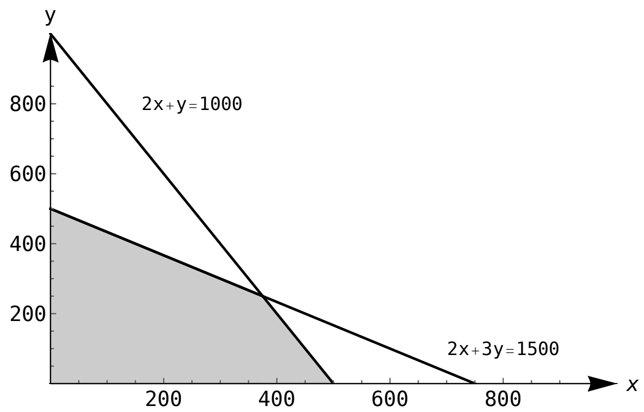


Figure 15.28: A plot of the restricted domain of the objective function in Example 15.40. The lines correspond to the constraints of the linear program, and the grey region represents the feasible region.

To find the maximum, we have to analyze the interior and the boundary of this restricted domain.

The objective function f is a linear function, and its gradient is the vector

$$\vec{\nabla} f(x,y) = \begin{bmatrix} 50 \\ 40 \end{bmatrix}.$$

Hence the function f does not have any critical points or singular points in the interior.

It suffices to analyze the boundary, and the gradient can give us useful information on which part of the boundary needs further investigation. In this case the gradient points in a direction that makes more or less an angle of $\pi/4$ with the horizontal axis. The optimal solution, if unique, must be in a vertex, and the origin can be excluded because $f(0,0) = 0$. We therefore analyze the other three vertices of the feasible region. These are the solutions to the systems:

$$\begin{aligned} 2x + 3y = 1500; x = 0 &\quad\Rightarrow\quad (x,y) = (0,500) \\ 2x + y = 1000; y = 0 &\quad\Rightarrow\quad (x,y) = (500,0) \\ 2x + 3y = 1500;\ 2x + y = 1000 &\quad\Rightarrow\quad (x,y) = (375,250) \end{aligned}$$

We have to calculate the value of the objective function at each of the vertices to determine which of them has the maximum or minimum values. The possible non-existence of a solution must be taken into account, if the compound is not bounded:

$$\begin{aligned} f(0,500) &= 50 \times 0 + 40 \times 500 = €20000, \\ f(500,0) &= 50 \times 500 + 40 \times 0 = €25000, \\ f(375,250) &= 50 \times 375 + 40 \times 250 = €28750. \end{aligned}$$

The optimum solution is to make 375 pants and 250 jackets to obtain a benefit of €28750.

Let us remark that in the previous example we did not make use of vectors notations, as we were analyzing a linear program in two variables. In vector notation the objective function $\vec{\boldsymbol{c}}^T\vec{\boldsymbol{x}}$ would have been the vectors

$$\vec{\boldsymbol{c}} = \begin{bmatrix} 50 \\ 40 \end{bmatrix} \quad \text{and} \quad \vec{\boldsymbol{x}} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

The first two constraints could be written as $A\vec{\boldsymbol{x}} \leq \vec{\boldsymbol{b}}$ with

$$A = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad \vec{\boldsymbol{b}} = \begin{bmatrix} 1500 \\ 1000 \end{bmatrix}.$$

The last two constraints could simply be written as $\vec{\boldsymbol{x}} \geq 0$.

The previous example has provided a couple of interesting properties of linear programs. When the objective function is linear, no critical points or singular points can occur in the interior of the feasible region, because the gradient will always differ from zero. In linear programs it will suffice to analyze the boundary, and because the feasible region is a convex polytope, this analysis will also be straightforward. For linear programs in two or three variables one can often determine the maxima visually. For more than three variables, there exist computer algorithms that find the maxima efficiently. A method called the simplex algorithm is commonly used. In this course we will only consider pen-and-paper techniques for objective functions with two or three variables.

In the previous example a unique solution was found. However, the solution is not always unique, as shown in the next and final example.

Example 15.41

If the objective function of Example 15.40 is modified to

$$f(x,y) = 20x + 30y\,,$$

then we find

$$\begin{aligned} f(0,500) &= 20 \times 0 + 30 \times 500 = €15000 \\ f(500,0) &= 20 \times 500 + 30 \times 0 = €10000 \\ f(375,250) &= 20 \times 375 + 30 \times 250 = €15000 \end{aligned}$$

In this case, we will find more than one maximum. The objective function will have a maximum at all (x,y) points with integer solutions on the segment that connects $(0,500)$ and $(375,250)$. This line segment is drawn with dashes Figure 15.29.

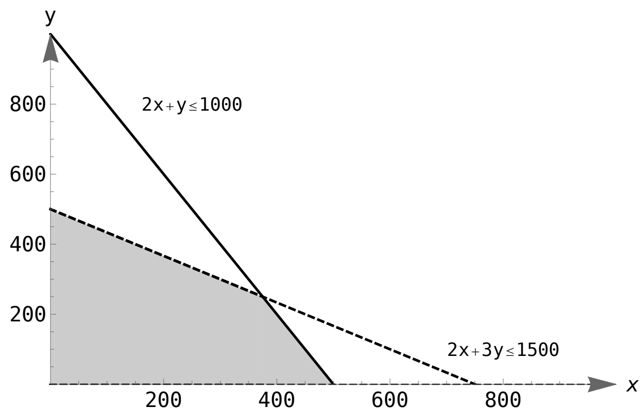


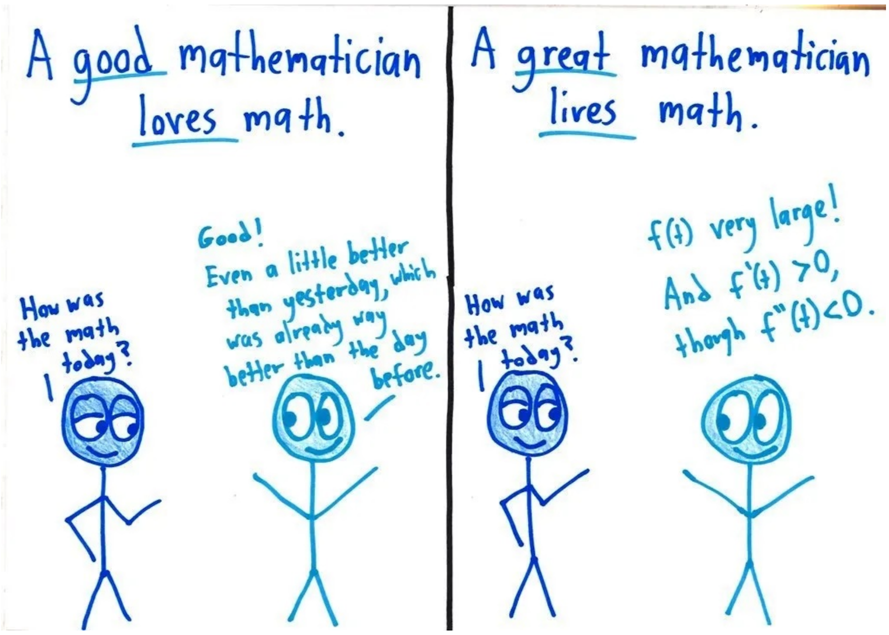
Figure 15.29: A plot of the restricted domain of the objective function in Example 15.41. In this case the maximum is not unique. The objective function has a maximum at value at all points on the black line.

So, when exactly will a linear program have more than one solution? Observe that for this modified objective function the gradient is

$$\vec{\nabla} f(x,y) = \begin{bmatrix} 20 \\ 30 \end{bmatrix}.$$

The gradient points in a direction orthogonal to the dashed line segment. The line segment coincides with a level curve of the function f.

More generally, if a part of the boundary coincides with a level curve of the objective function, then a linear program can have more than one solution. Moreover, because the objective is linear, the level curves will always be straight lines. In economics these lines are also known as iso-profit lines.



From *Math with Bad Drawings*, used by permission of Ben Orlin.

15.11 Exercices

Introduction to multivariable functions

Assignment 15.1 — Find the domain and image of the functions below.

(a) $f(x,y) = \dfrac{1}{\sqrt{4-x^2-y^2}}$

(b) $f(x,y) = \dfrac{\ln(x)}{\sin(y)}$

(c) $f(x,y) = \dfrac{2+\arcsin(y)}{\ln(2x)}$

(d) $f(x,y) = \sqrt{1-x^2-y^2}$

(e) $f(x,y) = \sin(xy)$

(f) $f(x,y) = |x|-|y|$

(g) $f(x,y) = \dfrac{1}{x-y}$

(h) $f(x,y) = \ln^{-1}(x^2+y^2-3)$

(i) $f(x,y) = \pi - \arcsin(x^2+2y^2)$

Assignment 15.2 — Sketch some level curves for the functions below

(a) $f(x,y) = \dfrac{x^2}{y}$

(b) $f(x,y) = \dfrac{y}{x^2+y^2}$

(c) $f(x,y) = \sqrt{\dfrac{1}{y}-x^2}$

Partial derivatives

Assignment 15.3 — Calculate the first-order partial derivatives of the functions below.

(a) $f(x,y) = x^2+2y^2-3x+2y+3$

(b) $f(x,y) = 2x^3y^2+2y+4x$

(c) $f(x,y) = \sin\left(2x^3-y^3\right)$

(d) $f(x,y) = e^{x^2}\cos(xy)$

(e) $f(x,y) = y\sin\left(x^2+y^2\right)$

(f) $f(x,y) = x^4\sin\left(xy^3\right)$

(g) $f(x,y) = e^{xy}\sin\left(4y^2\right)$

(h) $f(x,y) = \dfrac{x+y}{x-y}$

(i) $f(x,y) = y^{-3/2}\arctan\left(\dfrac{x}{y}\right)$

(j) $f(x,y) = \sqrt{x^2+4y^2}$

(k) $f(x,y,z) = x^2 - 2y^2 + 3z^2 + 4xy + 5xz + 6yz$

(l) $f(x,y,z) = x^2y^4z^3 + xy + z^2 + 1$

(m) $f(x,y,z) = \sqrt{x^2+y^2+z^2}$

(n) $f(x,y,z) = x^2 y \cos(z)$

(o) $f(x,y,z) = z^2 \ln(x^2 y)$

Assignment 15.4 — Find the partial derivatives of the first and second order of the functions below, and at the specified point.

(a) $f(x,y,z) = \arctan(x+y+z)$ $\quad (1,0,0)$

(b) $f(x,y,z) = x^2 + 3y^2 + 6z^2 - 2xy + 6xz + 7yz + 4x - 3y + 7$ $\quad (0,0,0)$

(c) $f(x,y,z) = \sqrt{xy+z^2}$ $\quad (1,1,1)$

(d) $f(x,y,z) = xe^{xy+z}$ $\quad (1,0,0)$

(e) $f(x,y,z) = e^{x+y^2+z^3}$ $\quad (0,0,0)$

(f) $f(x,y,z) = x\sin(y) + y\ln(z)$ $\quad (1,\pi,1)$

(g) $f(x,y,z) = (xy)^z + z^{xy}$ $\quad (1,1,1)$

Assignment 15.5 — Determine what's requested.

(a) $\dfrac{\partial z}{\partial x}$ if $z = e^{xy(x^2+y^2)}$

(b) $\dfrac{\partial u}{\partial x}, \dfrac{\partial u}{\partial y}, \dfrac{\partial u}{\partial z}$ if $u = 2y\sqrt{x} + 3y^2\sqrt[3]{z^2}$

(c) dz if $z = \arctan\left(\dfrac{x+y}{x-y}\right)$

Assignment 15.6 — Let $z = \ln\left(\sqrt{x^2+y^2}\right)$. Then, prove

$$x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} = 1.$$

Assignment 15.7 — Let

$$z = e^{\frac{x}{y}}\sin\left(\frac{x}{y}\right) + e^{\frac{y}{x}}\cos\left(\frac{y}{x}\right).$$

Then, prove

$$x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} = 0.$$

Assignment 15.8 — Consider the Van der Waals equation

$$f(p,V,n,T) = \left(p + \frac{n^2 a}{V^2}\right)(V - nb) - nRT = 0,$$

with a, b and R constant. Find

(a) $\left(\frac{\partial V}{\partial T}\right)_{p,n}$

(b) $\left(\frac{\partial V}{\partial p}\right)_{T,n}$

(c) $\left(\frac{\partial p}{\partial T}\right)_{V,n}$

(d) $\left(\frac{\partial p}{\partial V}\right)_{T,n}$

The notation $\left(\frac{\partial f}{\partial x}\right)_{y,z}$ indicates that y and z are considered constants when differentiating with respect to x.

Total differential and differentiability

Assignment 15.9 — Find the total differential of the first order of the function $f(x,y) = (xy)^3$.

Assignment 15.10 — Find the partial derivatives of the first and second order of the functions below, and at the specified point. Also, find the corresponding total differentials of first and second order.

(a) $f(x,y) = x^2 + 2xy + y^2 - 2x + 3y - 7$ $(1,2)$

(b) $f(x,y) = x^2y^5 + xy^2 + x^3y$ $(3,1)$

(c) $f(x,y) = \dfrac{xy}{x^2+y^2}$ $(1,1)$

(d) $f(x,y) = x^y$ $(1,1)$

(e) $f(x,y) = \ln(2x - 3y)$ $(2,1)$

(f) $f(x,y) = \dfrac{e^y}{x}$ $(1,1)$

(g) $f(x,y) = e^{\frac{y}{x}}$ $(1,1)$

(h) $f(x,y) = \cos(3x + 2y)$ $(0,\pi)$

(i) $f(x,y) = \arctan(x + y)$ $(1,0)$

(j) $f(x,y) = (2x + y)^{x+3y}$ $(0,1)$

Assignment 15.11 — The sides of a rectangular box are measured to the nearest 1% of their length. What is the estimated maximum error rate of

(a) the volume of the box,

(b) the area of the sides of the box,

(c) the length of the diagonal of the box?

The multivariable chain rule

Assignment 15.12 — Let $z = xy$ with $x = \dfrac{1}{t}$ and $y = t^2$. Find $\frac{dz}{dt}$.

Assignment 15.13 — Find u_t if $u = \sqrt{x^2 + y^2}$ with $x = e^{st}$ and $y = 1 + s^2\cos(t)$.

Assignment 15.14 — Let $z = f(u, v)$ with $u = u(x, y)$ and $v = v(x, y)$. Find z_x and z_y.

(a) $z = \ln\left(u^2 + v^2\right)$ with $\begin{cases} u = x + 2y + 1 \\ v = 3x - y - 1 \end{cases}$

(b) $z = 8u^2v - 2u + 3v$ with $\begin{cases} u = xy \\ v = x - y \end{cases}$

(c) $z = \dfrac{v}{u}$ with $\begin{cases} u = \sin\left(x^2 - y^2\right) \\ v = e^{xy} \end{cases}$

(d) $z = ve^{uv}$ with $\begin{cases} u = x^2y \\ v = xy^2 \end{cases}$

(e) $z = u^v$ with $\begin{cases} u = x^2 + y^2 \\ v = xy \end{cases}$

(f) $z = \cos(u)\sin(v)$ with $\begin{cases} u = x - y \\ v = x^2 + y^2 \end{cases}$

(g) $z = \arctan\left(\dfrac{u}{v}\right)$ with $\begin{cases} u = 2x + y \\ v = 3x - y \end{cases}$

(h) $z = u^2 + 3uv + v^2$ with $\begin{cases} u = \sin(x) + \cos(y) \\ v = \sin(x) - \cos(y) \end{cases}$

Assignment 15.15 — We consider $w = f(x, y, z) = xy + xz + yz$, with $x = e^t$, $y = e^{-t}$ and $z = e^t + e^{-t}$. Determine $\frac{dw}{dt}$.

Assignment 15.16 — Let $z = f(x, y)$ with $x = 2s + 3t$ and $y = 3s - 2t$. Find

$$\frac{\partial^2 z}{\partial s^2}, \quad \frac{\partial^2 z}{\partial s \partial t} \quad \text{and} \quad \frac{\partial^2 z}{\partial t^2}.$$

Assignment 15.17 — Consider the function $w = f(x, y, z)$ with $x = g(s)$, $y = h(s, t)$ en $z = k(t)$. Find an expression for $\frac{\partial w}{\partial t}$.

Assignment 15.18 — Consider the function $z = g(x, y)$ with $y = f(x)$ and $x = h(u, v)$. Find an expression for $\frac{\partial z}{\partial u}$.

Directional derivatives

Assignment 15.19 — Determine the degree of change for the function $f(x,y) = x^2y$ at $(-1,-1)$ in the direction of $\vec{\boldsymbol{v}} = (1,2)$.

Assignment 15.20 — Let $f(x,y) = x + y^2 - 3xy + 5y - 1$. In which direction at$(1,1)$ does the function value change the most?

Assignment 15.21 — Determine the directional derivative $D_{\hat{\boldsymbol{u}}}f(P)$ if $f(x,y,z) = xy^2z^3$, $P = (1,1,1)$ and $\hat{\boldsymbol{u}}$ are perpendicular to the plane $x^4 + 2y^4 + 2z^4 = 5$ in P.

Tangent lines, normal lines, and tangent planes

Assignment 15.22 — Find the equations of the tangent plane and the normal to the graph of the given function at the given point.

(a) $f(x,y) = x^2 - 2xy + y^2 - x + 2y \quad$ in $\quad (1,-1)$

(b) $f(x,y) = 4x^3y^2 + 2y \quad$ in $\quad (1,-2)$

(c) $x^2 + y^2 + z^2 = 25 \quad$ in $\quad (-3,0,4)$

(d) $2x^2 + 4yz - 5z^2 = -10 \quad$ in $\quad (3,-1,2)$

(e) $f(x,y) = \dfrac{2xy}{x^2+y^2} \quad$ in $\quad (0,2)$

(f) $f(x,y) = 2\sin(x)\cos(y) \quad$ in $\quad \left(\dfrac{\pi}{4}, \dfrac{\pi}{4}\right)$

(g) $f(x,y) = \ln\left(\sqrt{x^2+y^2}\right) \quad$ in $\quad (-1,0)$

Assignment 15.23 — Determine the distance from the point $(1,1,0)$ to the circular paraboloid $z = x^2 + y^2$.

Taylor series expansions

Assignment 15.24 — For the functions below, determine a Taylor series of second order at the specified point.

(a) $f(x,y) = xe^{xy+y}$ $\quad (1,1)$

(b) $f(x,y) = x\ln(y)$ $\quad (0,1)$

(c) $f(x,y) = xy + \ln(xy)$ $\quad (1,1)$

(d) $f(x,y) = x\sin(y)$ $\quad (1,0)$

(e) $f(x,y) = xy\cos(x+y)$ $\left(0, \dfrac{\pi}{2}\right)$

Extreme values

Assignment 15.25 — Determine and classify the critical points of the given functions.

(a) $f(x,y) = x^2 + xy + y^2 - 3x$

(b) $f(x,y) = xy + \dfrac{2}{x} + \dfrac{4}{y}$

(c) $f(x,y) = x^2y - 6y^2 - 3x^2$

(d) $f(x,y) = x^2 + y^2 + \dfrac{2}{xy}$

Assignment 15.26 — Determine and classify the critical points of the given functions.

(a) $f(x,y) = x^2 + y^2 - 3xy$

(b) $f(x,y) = xy$

(c) $f(x,y) = y\sqrt{x} - xy + y^2$

(d) $f(x,y) = (2x^2 - y)(2 - y)$

(e) $f(x,y) = x^4 + y^4 - 2x^2 + 4xy - 2y^2$

(f) $f(x,y) = x^3 - 3xy^2$

(g) $f(x,y) = (x-y)^4 + (y-1)^4$

(h) $f(x,y) = x + y\sin(x)$

(i) $f(x,y) = (x+y)e^{-xy}$

(j) $f(x,y) = x^2 + y^2$

(k) $f(x,y) = \cos(x+y)$

(l) $f(x,y) = x^2 + 2y^2 - 4x + 4y$

(m) $f(x,y) = xy - x + y$

(n) $f(x,y) = x^3 + y^3 - 3xy$

(o) $f(x,y) = x^4 + y^4 - 4xy$

(p) $f(x,y) = \dfrac{x}{y} + \dfrac{8}{x} - y$

(q) $f(x,y) = x\sin(y)$

(r) $f(x,y) = \cos(x) + \cos(y)$

(s) $f(x,y) = x^2ye^{-(x^2+y^2)}$

(t) $f(x,y) = \dfrac{xy}{2 + x^4 + y^4}$

(u) $f(x,y) = xe^{-x^3+y^3}$

(v) $f(x,y) = \dfrac{x^2}{x^2+y^2}$

(w) $f(x,y) = \dfrac{xy}{x^2+y^2}$

(x) $f(x,y,z) = xy + x^2z - x^2 - y - z^2$

Assignment 15.27 — Determine the maximum and minimum of the function $f(x,y,z) = x + y^2z$ under the conditions $y^2 + z^2 = 2$ and $z = x$.

Assignment 15.28 — The strength of a wooden beam with rectangular cross-section and given length is directly proportional to the product of its width x and the square of its height y (take as a proportionality factor 1). Find the dimensions (x and y) of the strongest beam that can be sawn from a tree trunk, of circular cross-section and diameter $\sqrt{3}$.

Assignment 15.29 — Find the dimensions of an open (no top surface) beam-shaped box that has a volume of 32 dm^3, but a minimum sheath area.

Assignment 15.30 — Find the positive numbers a, b and c for which the sum is 30 and ab^2c^3 is maximal.

Assignment 15.31 — The material to make the bottom of a beam-shaped box is twice as expensive per unit area as the material used for the lid and side walls. Find the dimensions of a box of volume 12 m^2 for which the material cost is minimal.

Assignment 15.32 — A metal surface whose shape can be described by $4x^2+y^2+4z^2=16$ is heated. After one hour the temperature at (x,y,z) equals:

$$T(x,y,z)=8x^2+4yz-16z+600.$$

Find the warmest point on the surface if we restrict ourselves to $x>0$.

Assignment 15.33 — An open beam-shaped cardboard box is reinforced at the ribs of the bottom and the sides with tape. One has 96 cm of tape available. What are the dimensions of the box with the largest possible volume?

Constrained optimisation

Assignment 15.34 — Find the extreme value(s) of the functions below constrained by the specified region.

(a) $f(x,y)=x-x^2+y^2$ constrained by the rectangle $0\le x\le 2, 0\le y\le 1$

(b) $f(x,y)=xy-2x$ constrained by the rectangle $-1\le x\le 1, 0\le y\le 1$

(c) $f(x,y)=xy-x^3y^2$ constrained by the square $0\le x\le 1, 0\le y\le 1$

(d) $f(x,y)=x^2y+xy^3+y$ constrained by the square $-1\le x\le 1, -1\le y\le 1$

(e) $f(x,y)=x^3-3x+y^2+2y$ constrained by the triangle with vertices $(1,0)$, $(3,-2)$ and $(-1,-2)$

(f) $f(x,y)=xy(1-x-y)$ constrained by the triangle with vertices $(0,0)$, $(0,1)$ and $(1,0)$

(g) $f(x,y)=\sin x\cos y$ constrained by the triangle with vertices $(0,0)$, $(0,2\pi)$ and $(2\pi,0)$

(h) $f(x,y)=x^2ye^{-(x+y)}$ constrained by the triangle with vertices $(0,0)$, $(0,4)$ and $(4,0)$

(i) $f(x,y)=x^2+x+3y^2$ constrained by the square formed by $y=x+1$, $y=x-1$, $y=-x+1$ and $y=-x-1$

Assignment 15.35 — Find the maximum value of the function f constrained by the specified area.

(a) $f(x,y)=2x+7y$ constrained by the region where $x+2y\le 6$, $2x+y\le 6$, $x\ge 0$ and $y\ge 0$

(b) $f(x,y)=2x+3y$ constrained by the region where $x+2y\le 12$, $4x+y\le 12$, $y\le 5$, $x\ge 0$ and $y\ge 0$

⚘⚘ **Assignment 15.36 —** Find the minimal value of $f(x,y,z) = 2x+3y+4z$ constrained by the region where $x+y \geq 2$, $y+z \geq 2$, $x+z = 2$, $x \geq 0$, $y \geq 0$ and $z \geq 0$.

⚘⚘ **Assignment 15.37 —** A textile manufacturer produces two types of fabric made from wool, cotton and polyester. The luxury fabric consists of 20 percent wool, 50 percent cotton, and 30 percent polyester and is sold at $3 per kilogram. The standard fabric consists of 10 percent wool, 40 percent cotton and 50 percent polyester and is sold at $2 per kilogram. There is 2000 kg of wool in stock, 6000 kg of cotton and also 6000 kg of polyester. How many kilograms of fabric should the manufacturer produce to maximize its profit?

⚘ **Assignment 15.38 —** A circuit board manufacturer has a stock of 200 resistors, 120 transistors, and 150 capacitors. The manufacturer is asked to produce two types of circuits A and B. Type A requires 20 resistors, 10 transistors, and 10 capacitors, while type B requires 10 resistors, 20 transistors, and 30 capacitors. The profit on the former is 5 €and on the latter 12 €. How many circuits of each type should be produced to have maximum profit? What is that profit then?

⚘⚘ **Assignment 15.39 —** A radio station conducted a survey on the ratings for 3 types of radio programs: pop music, oldies and information programs. Broadcasting pop music costs €6 per hour and is rated 10, whereas broadcasting oldies costs €3 per hour and is rated 15. Information programs cost €2 per hour and are rated 5. The budget per day is at most €60, for which the station should broadcast exactly 20 hours per day with at least 4 hours of pop music. Find the ratio of different programs that maximizes the rating.

⚘⚘ **Assignment 15.40 —** A plot of 10-hectare is divided into zones in which 6 detached houses per hectare, 8 semi-detached buildings per hectare or 12 apartments per hectare can be built. The entire plot should be built on. The plot's owner can make a profit of €40,000 per detached house, €20,000 per semi-detached house and €16,000 per apartment. In total, he would like to have at least as many apartments as houses (detached and half open together). How many buildings of each type should be built to maximize profit?

⚘⚘ **Assignment 15.41 —** Karl Lagerfeld has 230 m of fabric that he plans to use to its full potential. He wants to use it to make a maximum of 20 suits, 30 jackets and 40 pants. For a suit he needs 6 m of fabric, for a coat 3 m and for pants 2 m. For each suit, coat and trouser he has a respective profit of €20,€14 and €12. How many pieces of each product must Karl make to maximize his profit?

⚘⚘ **Assignment 15.42 —** A businesswoman wants to buy coffee, tea and sugar together for no more than €340. Coffee costs €3 per kg, tea €3.5 per kg and sugar €4 per kg. The businesswoman can buy no more than 75 kg of coffee, 75 kg of tea and 90 kg of sugar per day. She wants to buy exactly 90 kg all together per day, which is the maximum carrying capacity of her van. Suppose the businesswoman makes a profit of €1 per kg of coffee, €2 per kg of tea and €1.5 per kg of sugar. For which combination of coffee, tea and sugar does she makes most profit?

⚘⚘ **Assignment 15.43 —** A tennis ball manufacturer produces three types of tennis balls: Silver, Yellow and Gold. To practise at Roland Garros, the organization wants to order exactly 1800 tennis balls, of which at least 200 must be Silver balls, 300 Yellow balls, and at most 1200 Gold balls. The number of Silver balls may at most be double of the number of Yellow balls. The manufacturer sells the Silver balls at €2 each, the Yellow balls cost €2.20 each and the Gold balls cost only €1 each. Based on these conditions and prices, the Roland Garros organization would like to place an order at the lowest possible cost. How many balls of each type should the organization order? What will the organization spend in total?

Revieuw exercises

Assignment 15.44 — Let $f(x, y) = \ln(x^2 + y^2)$ and $P = (1, -2)$. Find

(a) the gradient of this function at P,

(b) an equation of the tangent plane to the graph of f at P,

(c) an equation of the tangent at P to the level curve of f through P.

Assignment 15.45 — The temperature $T(x, y)$ across the xy-plane is given by $T(x, y) = x^2 - 2y^2$.

(a) Draw a contour plot for T.

(b) In which direction should an ant move in $(2, -1)$ to cool down as quickly as possible?

(c) Along which curve through $(2, -1)$ should the ant move to experience maximum cooling?

God does not care about our mathematical difficulties. He integrates empirically.

— Albert Einstein —

16

Double and line integrals

The previous chapter introduced multivariable functions and we applied concepts of differential calculus to these functions. We learned how we can view a function of two variables as a surface in space, and learned how partial derivatives convey information about how the surface is changing in any direction.

In this chapter we apply techniques of integral calculus to multivariable functions. In Chapter 12 we learned how the definite integral of a single variable function gave us area under the curve. In this chapter we will see that integration applied to a multivariable function gives us volume under a surface. And just as we learned applications of integration beyond finding areas, we will find applications of integration in this chapter beyond finding volume.

16.1 Iterated integrals and area

16.1.1 Iterated integrals

In Chapter 15 we found that it was useful to differentiate functions of several variables with respect to one variable, while treating all the other variables as constants or coefficients. We can integrate functions of several variables in a similar way. For instance, if we are told that $f_x(x,y) = 2xy$, we can treat y as staying constant and integrate to obtain $f(x,y)$:

$$\begin{aligned} f(x,y) &= \int f_x(x,y)\ dx \\ &= \int 2xy\ dx \\ &= x^2y + C. \end{aligned}$$

Make a careful note about the constant of integration, C. This "constant" is something with a derivative of 0 with respect to x, so it could be any expression that contains only constants and functions of y. For instance, if $f(x,y) = x^2y + \sin(y) + y^3 + 17$, then $f_x(x,y) = 2xy$. To signify that C is actually a function of y, we write:

$$f(x,y) = \int f_x(x,y)\ dx = x^2y + C(y).$$

Using this process we can even evaluate definite integrals. For instance, to evaluate the integral

$$\int_1^{2y} 2xy\ dx.$$

We find the indefinite integral as before, then apply the fundamental theorem of calculus to evaluate the definite integral:

$$\begin{aligned}\int_1^{2y} 2xy\ dx &= x^2y\Big|_1^{2y} \\ &= (2y)^2y - (1)^2y \\ &= 4y^3 - y.\end{aligned}$$

We can also integrate with respect to y. In general,

$$\int_{h_1(y)}^{h_2(y)} f_x(x,y)\ dx = f(x,y)\Big|_{h_1(y)}^{h_2(y)} = f\big(h_2(y),y\big) - f\big(h_1(y),y\big),$$

and

$$\int_{g_1(x)}^{g_2(x)} f_y(x,y)\ dy = f(x,y)\Big|_{g_1(x)}^{g_2(x)} = f\big(x,g_2(x)\big) - f\big(x,g_1(x)\big).$$

Note that when integrating with respect to x, the bounds are functions of y (of the form $x = h_1(y)$ and $x = h_2(y)$) and the final result is also a function of y. When integrating with respect to y, the bounds are functions of x (of the form $y = g_1(x)$ and $y = g_2(x)$) and the final result is a function of x.

When evaluating $\int_1^{2y} 2xy\ dx$, we integrated a function with respect to x and ended up with a function of y. We can integrate this as well. This process is known as iterated integration, or **double integration** (*dubbelintegratie*). Of course, when considering

$$\int_{x_1}^{x_2}\int_{g_1(x)}^{g_2(x)} f(x,y)dy\ dx$$

x should have constant bounds, whereas y may have variable ones, and vice versa when considering

$$\int_{y_1}^{y_2}\int_{h_1(y)}^{h_2(y)} f(x,y)dx\ dy$$

Example 16.1

Evaluate

$$\int_1^2 \left(\int_1^x \left(5x^3y^{-3} + 6y^2\right) dy \right) dx.$$

Solution

We follow a standard order of operations and perform the operations inside parentheses first.

$$\begin{aligned}
\int_1^2 \left(\int_1^x \left(5x^3y^{-3} + 6y^2\right) dy \right) dx &= \int_1^2 \left(\frac{5x^3y^{-2}}{-2} + \frac{6y^3}{3} \right)\Bigg|_1^x dx \\
&= \int_1^2 \left(\frac{5x^3}{-2x^2} + \frac{6x^3}{3} - \frac{5x^3}{-2} - \frac{6}{3} \right) dx \\
&= \int_1^2 \left(-\frac{5x}{2} + 2x^3 + \frac{5x^3}{2} - 2 \right) dx \\
&= \int_1^2 \left(\frac{9x^3}{2} - \frac{5x}{2} - 2 \right) dx \\
&= \left(\frac{9}{8}x^4 - \frac{5}{4}x^2 - 2x \right)\Bigg|_1^2 \\
&= \frac{89}{8}
\end{aligned}$$

Note how the bounds of x were $x = 1$ to $x = 2$ and the final result was a number.

The previous example showed how we could perform something called an iterated integral; we do not yet know why we would be interested in doing so nor what the result, such as the number $89/8$, means. We will now investigate that.

16.1.2 Area of a plane region

Consider the plane region R bounded by $a \le x \le b$ and $g_1(x) \le y \le g_2(x)$, shown in Figure 16.1(a). We learned in Section 13.1 that the area of R is given by

$$\int_a^b \left(g_2(x) - g_1(x)\right) dx.$$

We can view the expression $\left(g_2(x) - g_1(x)\right)$ as

$$\left(g_2(x) - g_1(x)\right) = \int_{g_1(x)}^{g_2(x)} 1 \, dy = \int_{g_1(x)}^{g_2(x)} dy,$$

meaning we can express the area of R as an iterated integral:

$$\text{area} = \int_a^b \big(g_2(x) - g_1(x)\big)\,dx = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} dy\right) dx = \int_a^b \int_{g_1(x)}^{g_2(x)} dy\,dx.$$

In short: a certain iterated integral can be viewed as giving the area of a plane region.

A region R could also be defined by $c \le y \le d$ and $h_1(y) \le x \le h_2(y)$, as shown in Figure 16.1(b). Using a process similar to that above, we have

$$\int_c^d \int_{h_1(y)}^{h_2(y)} dx\,dy.$$

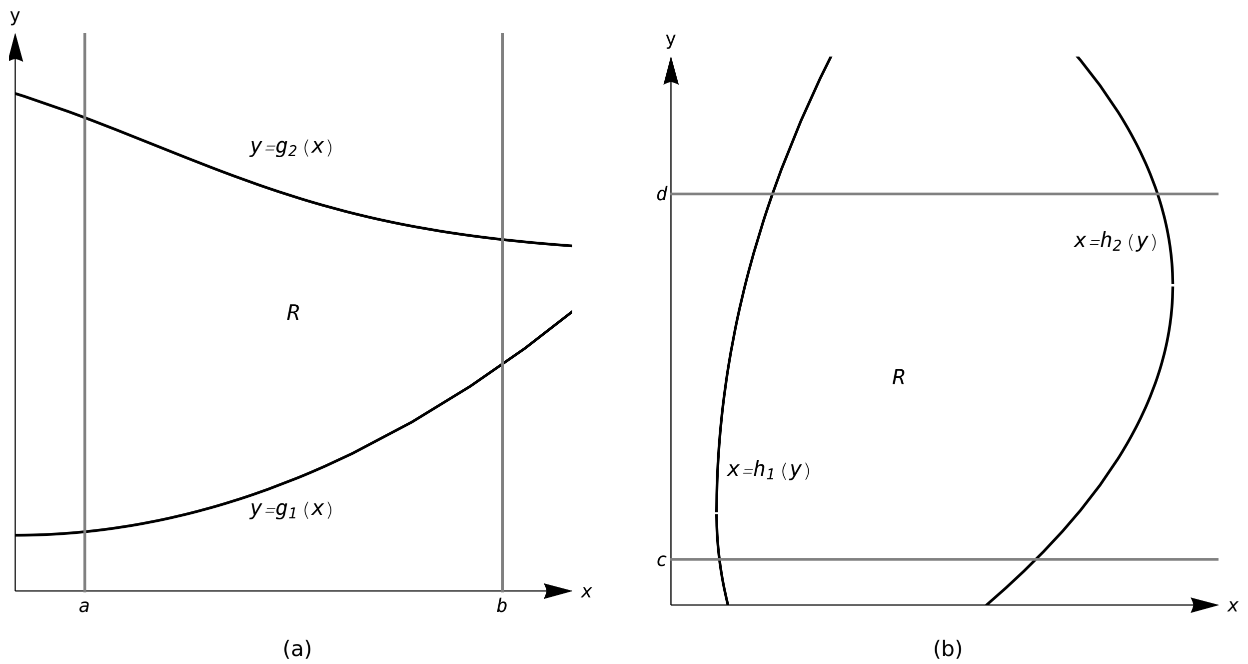


Figure 16.1: Calculating the area of a plane region R with iterated integrals.

We state this formally in a theorem.

Theorem 16.1 (Area of a plane region)

1. *Let R be a plane region bounded by $a \le x \le b$ and $g_1(x) \le y \le g_2(x)$, where g_1 and g_2 are continuous functions on $[a,b]$. The* ***area*** *A* ***of*** *R is*

$$A = \int_a^b \int_{g_1(x)}^{g_2(x)} dy\,dx.$$

2. *Let R be a plane region bounded by $c \le y \le d$ and $h_1(y) \le x \le h_2(y)$, where h_1 and h_2 are*

continuous functions on $[c,d]$. *The* ***area*** A ***of*** R *is*

$$A = \int_c^d \int_{h_1(y)}^{h_2(y)} dx\, dy.$$

The following examples should help us understand this theorem.

Example 16.2

Find the area A of the triangle with vertices at $(1,1)$, $(3,1)$ and $(5,5)$, as shown in Figure 16.2(a).

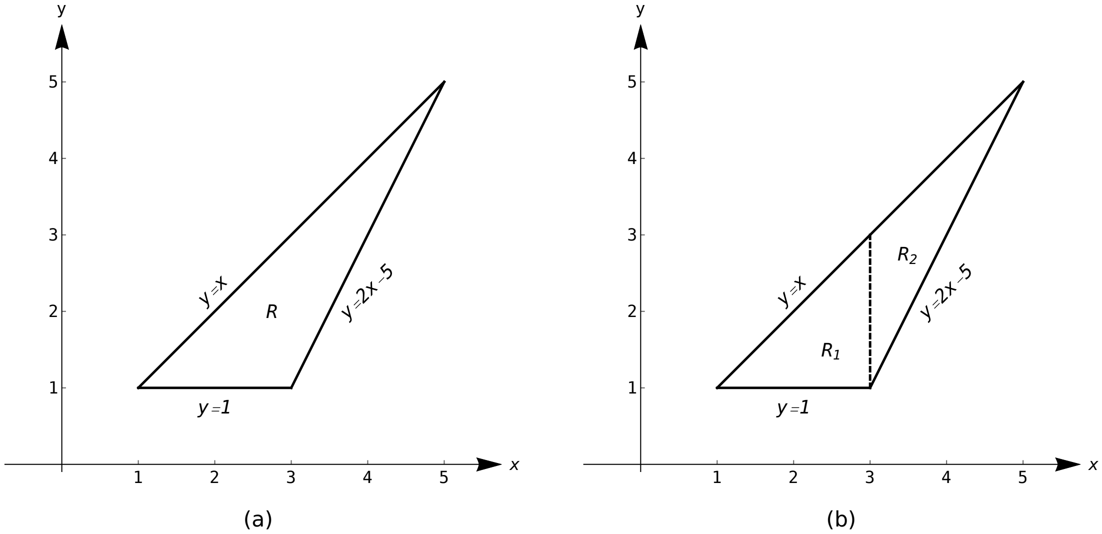


Figure 16.2: Calculating the area of a triangle with iterated integrals in Example 16.2 by using constant bounds on y (a) and x (b)

Solution

The triangle is bounded by the lines as shown in Figure 16.2(a). Choosing to integrate with respect to x first gives that x is bounded by $x = y$ to $x = \frac{y+5}{2}$, while y is bounded by $y = 1$ to $y = 5$, i.e. the bounds with respect to y are fixed. Recall that since x-values increase from left to right, the leftmost curve, $x = y$, is the lower bound and the rightmost curve, $x = (y+5)/2$, is the upper bound. The area is

$$\begin{aligned} A &= \int_1^5 \int_y^{\frac{y+5}{2}} dx\, dy \\ &= \int_1^5 (x) \Big|_y^{\frac{y+5}{2}} dy \\ &= \int_1^5 \left(-\frac{1}{2}y + \frac{5}{2}\right) dy \\ &= \left(-\frac{1}{4}y^2 + \frac{5}{2}y\right)\Bigg|_1^5 \\ &= 4. \end{aligned}$$

We can also find the area by integrating with respect to y first, which implies that the bounds on y are variable and dependent on x, whereas the one on x are fixed. In this situation, though, there is not one function that defines the lower bound on the interval $[1,5]$. Essentially, we have two functions that act as the lower bound for the region R, $y=1$ and $y=2x-5$. In this way, the region R is split into two smaller regions, namely R_1 and R_2 (Figure 16.2(b)). This requires us to use two iterated integrals. Note how the x-bounds are different for each integral:

$$\begin{aligned} A &= \int_1^3 \int_1^x 1\,dy\,dx + \int_3^5 \int_{2x-5}^x 1\,dy\,dx \\ &= \int_1^3 y\Big|_1^x\,dx + \int_3^5 y\Big|_{2x-5}^x\,dx \\ &= \int_1^3 \big(x-1\big)\,dx + \int_3^5 \big(-x+5\big)\,dx \\ &= 2+2 \\ &= 4. \end{aligned}$$

As expected, we get the same answer both ways.

Example 16.3

Find the area of the region enclosed by $y=2x$ and $y=x^2$, as shown in Figure 16.3.

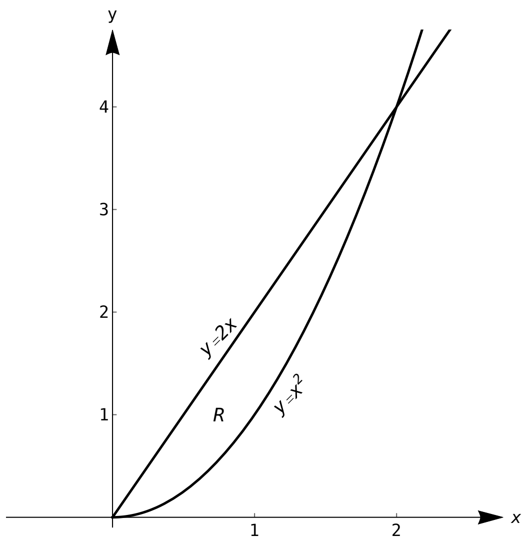


Figure 16.3: Calculating the area of a plane region with iterated integrals in Example 16.3.

Solution

Once again we will find the area of the region using both orders of integration. We can approach this problem in two ways, either by choosing fixed bounds for x and variable ones - depending on x - for y or by choosing fixed bounds for y and variable ones for x, which then depend on y. So, in the former case, we first integrate with respect to y and then with respect to x:

$$\int_0^2 \int_{x^2}^{2x} 1\, dy\, dx = \int_0^2 (2x - x^2)\, dx = \left(x^2 - \frac{1}{3}x^3\right)\Big|_0^2 = \frac{4}{3}.$$

In the latter case, we, however, do exactly the opposite:

$$\int_0^4 \int_{y/2}^{\sqrt{y}} 1\, dx\, dy = \int_0^4 \left(\sqrt{y} - \frac{y}{2}\right) dy = \left(\frac{2}{3}y^{3/2} - \frac{1}{4}y^2\right)\Big|_0^4 = \frac{4}{3}.$$

In each of the previous examples, we have been given a region R and found the bounds needed to find the area of R using both orders of integration. We integrated using both orders of integration to demonstrate their equality.

We now approach the skill of describing a region using both orders of integration from a different perspective. Instead of starting with a region and creating iterated integrals, we will start with an iterated integral and rewrite it in the other integration order. To do so, we will need to understand the region over which we are integrating.

The simplest of all cases is when both integrals are bound by constants. The region described by these bounds is a rectangle, and so:

$$\int_a^b \int_c^d 1\, dy\, dx = \int_c^d \int_a^b 1\, dx\, dy.$$

When the inner integral's bounds are not constants, it is generally very useful to sketch the bounds to determine what the region we are integrating over looks like. From the sketch we can then rewrite the integral with the other order of integration.

Example 16.4

Change the order of integration of

$$\int_0^4 \int_{y^2/4}^{(y+4)/2} 1\, dx\, dy.$$

Solution

We sketch the region described by the bounds to help us change the integration order. x is bounded below and above (i.e., to the left and right) by $x = y^2/4$ and $x = (y+4)/2$ respectively, and y is bounded between 0 and 4. Graphing the previous curves, we find the region R to be that shown in Figure 16.4(a).

To change the order of integration, we need to give x fixed bounds. The figure makes it clear that there are two lower bounds for y: $y = 0$ on $0 \le x \le 2$, and $y = 2x - 4$ on $2 \le x \le 4$, thereby splitting R in two smaller regions R_1 and R_2 (Figure 16.4(b)). Thus we need two double integrals. The upper bound for each is $y = 2\sqrt{x}$. Thus we have

$$\int_0^4 \int_{y^2/4}^{(y+4)/2} 1\, dx\, dy = \int_0^2 \int_0^{2\sqrt{x}} 1\, dy\, dx + \int_2^4 \int_{2x-4}^{2\sqrt{x}} 1\, dy\, dx.$$

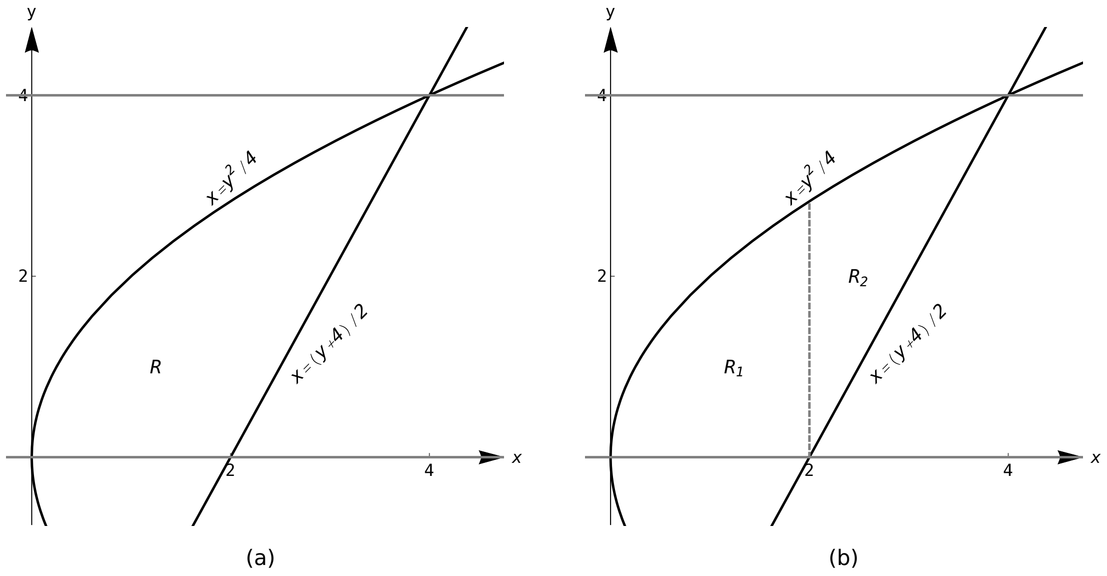


Figure 16.4: Drawing the region determined by the bounds of integration in Example 16.4.

This section has introduced a new concept, the iterated integral. We developed one application for iterated integration: area between curves. However, this is not new, for we already know how to find areas bounded by curves. In the next section we apply iterated integration to solve problems we currently do not know how to handle.

16.2 Double integration and volume

16.2.1 Definition

The definite integral of f over $[a,b]$, $\int_a^b f(x)\ dx$, was introduced as the signed area under the curve. We approximated the value of this area by first subdividing $[a,b]$ into n subintervals, where the i^{th} subinterval has length Δx_i, and letting c_i be any value in the i^{th} subinterval. We formed rectangles that approximated part of the region under the curve with width Δx_i, height $f(c_i)$, and hence with area $f(c_i)\Delta x_i$. Summing all the rectangle's areas gave an approximation of the definite integral, and Theorem 12.4 stated that

$$\int_a^b f(x)\ dx = \lim_{\mathcal{L}\to 0}\sum f(c_i)\Delta x_i,$$

connecting the area under the curve with sums of the areas of rectangles.

We use a similar approach in this section to find volume under a surface. Let R be a closed, bounded region in the xy-plane and let $z=f(x,y)$ be a continuous function defined on R. We wish to find the signed volume under the surface of f over R. We use the term signed volume to denote that space above the xy-plane, under f, will have a positive volume; space above f and under the xy-plane will have a negative volume, similar to the notion of signed area used before.

We start by partitioning R into n rectangular subregions as shown in Figure 16.5(a). For simplicity's sake, we let all widths be Δx and all heights be Δy. Note that the sum of the areas of the rectangles is not equal to the area of R, but rather is a close approximation. Arbitrarily number the rectangles 1 through n, and pick a point (x_i, y_i) in the i^{th} subregion.

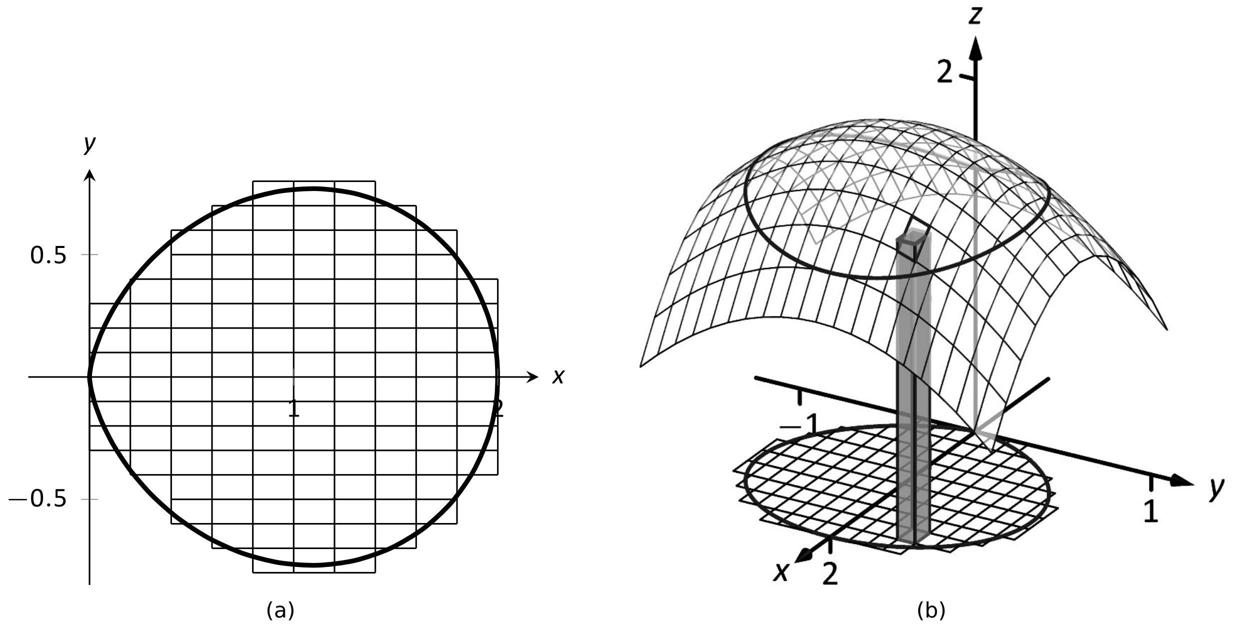


Figure 16.5: Developing a method for finding signed volume under a surface.

The volume of the rectangular solid whose base is the i^{th} subregion and whose height is $f(x_i, y_i)$ is $V_i = f(x_i, y_i)\Delta x \Delta y$. Such a solid is shown in Figure 16.5(b). Note how this rectangular solid only approximates the true volume under the surface; part of the solid is above the surface and part is below.

For each subregion R_i used to approximate R, create the rectangular solid with base area $\Delta x \Delta y$ and height $f(x_i, y_i)$. The sum of all rectangular solids is

$$\sum_{i=1}^{n} f(x_i, y_i)\Delta x \Delta y.$$

This approximates the signed volume under f over R. As we have done before, to get a better approximation we can use more rectangles to approximate the region R.

In general, each rectangle could have a different width Δx_j and height Δy_k, giving the i^{th} rectangle an area $\Delta A_i = \Delta x_j \Delta y_k$ and the i^{th} rectangular solid a volume of $f(x_i, y_i)\Delta A_i$. Let now $\mathcal{A}$ denote the length of the longest diagonal of all rectangles in the subdivision of R; $\mathcal{A} \to 0$ means each rectangle's width and height are both approaching 0. If f is a continuous function, as $\mathcal{A}$ shrinks (and hence $n \to +\infty$) the summation $\displaystyle\sum_{i=1}^{n} f(x_i, y_i)\Delta A_i$ approximates the signed volume better and better.

When adding up the volumes of rectangular solids over a partition of a region R, as done in Figure 16.5(b), one could first add up the volumes across each row (one type of sum), then add these totals together (another sum), as in

$$\sum_{j=1}^{n}\sum_{i=1}^{m} f(x_i, y_j)\Delta x_i \Delta y_j.$$

One can rewrite this as

$$\sum_{j=1}^{n}\left(\sum_{i=1}^{m} f(x_i, y_j)\Delta x_i\right)\Delta y_j.$$

The summation inside the parenthesis indicates the sum of (heights × widths), which gives an area; multiplying these areas by the thickness Δy_j gives a volume. The illustration in Figure 16.5(b) relates to this understanding.

This all leads us to a definition.

Definitie 16.1 (Double integral and signed volume)
Let $z = f(x,y)$ be a continuous function defined over a closed, bounded region R in the xy-plane. The **signed volume** V **under** f **over** R is denoted by the **double integral** (*dubbelintegraal*)

$$V = \iint_R f(x,y)\; dA.$$

Alternate notations for the double integral are

$$\iint_R f(x,y)\; dA = \iint_R f(x,y)\; dx\; dy = \iint_R f(x,y)\; dy\; dx.$$

We can find the signed volume by considering increasingly smaller, so more, rectangles; that is by considering the limit

$$\lim_{\mathcal{A}\to 0} \sum_{i=1}^{n} f(x_i, y_i)\Delta A_i.$$

In this limiting situation, it holds that

$$V = \iint_R f(x,y)\; dA = \lim_{\mathcal{A}\to 0} \sum_{i=1}^{n} f(x_i, y_i)\Delta A_i. \tag{16.1}$$

Note that this equation does not specify the partition of the region R, so any partitioning where the diagonal of each rectangle shrinks to 0 results in the same answer. This does not offer a very satisfying way of computing volume, though. Our experience has shown that evaluating the limits of sums can be tedious. We seek a more direct method.

Recall Theorem 13.1. This stated that if $A(x)$ gives the cross-sectional area of a solid at x, then $\int_a^b A(x)\; dx$ gave the volume of that solid over $[a,b]$. Consider Figure 16.6, where a surface $z = f(x,y)$ is drawn over a region R. Fixing a particular x-value, we can consider the area under f over R where x has that fixed value. That area can be found with a definite integral, namely

$$A(x) = \int_{g_1(x)}^{g_2(x)} f(x,y)\; dy.$$

Remember that though the integrand contains x, we are viewing x as fixed. Also note that the bounds of integration are functions of x: the bounds depend on the value of x. As $A(x)$ is a cross-sectional area function, we can find the signed volume V under f by integrating it:

$$V = \int_a^b A(x)\; dx = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} f(x,y)\; dy \right) dx = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\; dy\; dx.$$

This gives a concrete method for finding signed volume under a surface. We could do a similar procedure where we started with y fixed, resulting in an iterated integral with the order of integration $dx\; dy$. The following theorem states that both methods give the same result, which is the value of the double integral. It is such an important theorem it has a name associated with it.

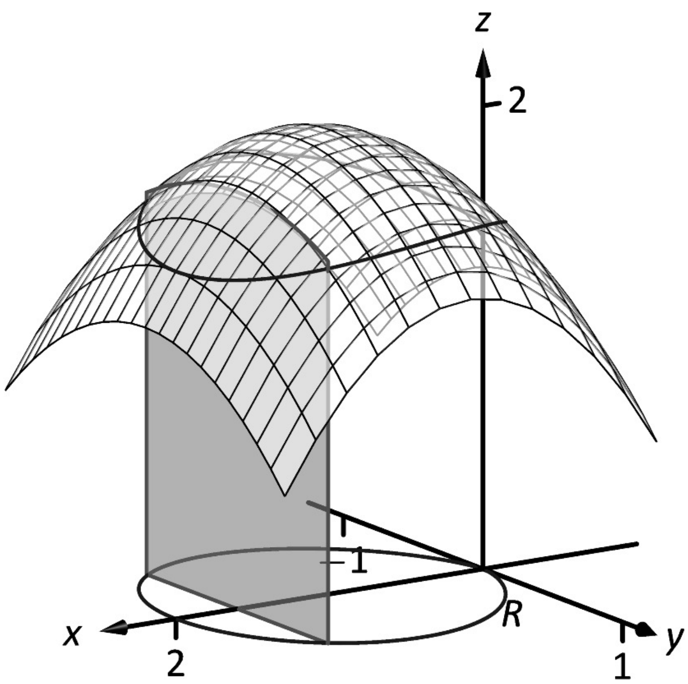


Figure 16.6: Finding volume under a surface by sweeping out a cross–sectional area.

Theorem 16.2 (Fubini's theorem)
Let R be a closed, bounded region in the xy-plane and let $z = f(x, y)$ be a continuous function on R.

1. *If R is bounded by $a \le x \le b$ and $g_1(x) \le y \le g_2(x)$, where g_1 and g_2 are continuous functions on $[a, b]$, then*
$$\iint_R f(x,y)\,dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx.$$
2. *If R is bounded by $c \le y \le d$ and $h_1(y) \le x \le h_2(y)$, where h_1 and h_2 are continuous functions on $[c, d]$, then*
$$\iint_R f(x,y)\,dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)\,dx\,dy.$$

Example 16.5

Let $f(x, y) = xy + e^y$. Find the signed volume under f on the region R, which is the rectangle with corners $(3, 1)$ and $(4, 2)$ pictured in Figure 16.7, using both orders of integration.

Solution

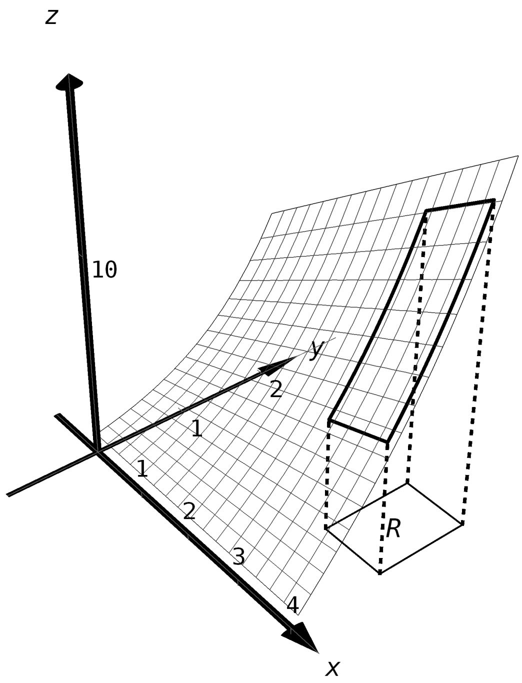


Figure 16.7: Finding the signed volume under a surface in Example 16.5.

We wish to evaluate $\iint_R \big(xy + e^y\big)\ dA$. As R is a rectangle, the bounds are easily described as $3 \leq x \leq 4$ and $1 \leq y \leq 2$.

Using the order $dy\ dx$:

$$\begin{aligned}
\iint_R \big(xy+e^y\big)\ dA &= \int_3^4 \int_1^2 \big(xy+e^y\big)\ dy\ dx \\
&= \int_3^4 \left(\frac{1}{2}xy^2+e^y\right)\Bigg|_1^2\ dx \\
&= \int_3^4 \left(\frac{3}{2}x+e^2-e\right)\ dx \\
&= \left(\frac{3}{4}x^2+\big(e^2-e\big)x\right)\Bigg|_3^4 \\
&= \frac{21}{4}+e^2-e \approx 9.92.
\end{aligned}$$

Now we check the validity of Fubini's theorem by using the order $dx\ dy$:

$$\iint_R \big(xy+e^y\big)\ dA = \int_1^2 \int_3^4 \big(xy+e^y\big)\ dx\ dy$$

$$= \int_1^2 \left(\frac{1}{2}x^2y + xe^y\right)\Bigg|_3^4 \, dy$$
$$= \int_1^2 \left(\frac{7}{2}y + e^y\right) dy$$
$$= \left(\frac{7}{4}y^2 + e^y\right)\Bigg|_1^2$$
$$= \frac{21}{4} + e^2 - e \approx 9.92.$$

Both orders of integration return the same result, as expected.

Example 16.6

Evaluate

$$\iint_R \left(3xy - x^2 - y^2 + 6\right) dA,$$

where R is the triangle bounded by $x = 0$, $y = 0$ and $x/2 + y = 1$, as shown in Figure 16.8.

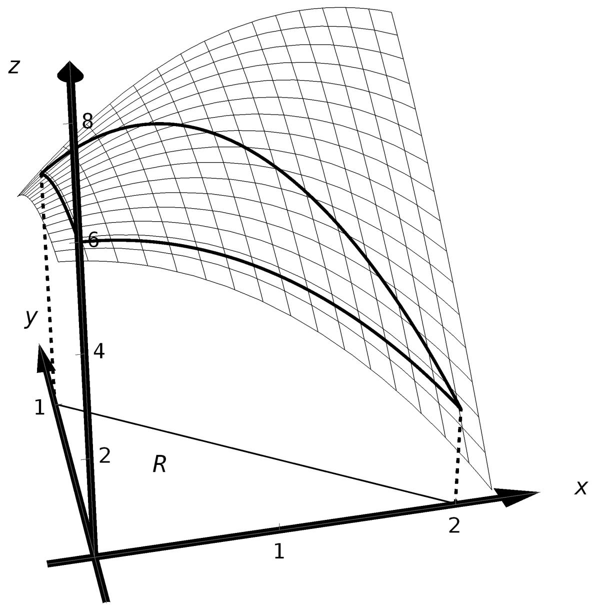


Figure 16.8: Finding the signed volume under a surface in Example 16.6.

Solution

While it is not specified which order we are to use, we will evaluate the double integral using both orders to help drive home the point that it does not matter which order we use.

Using the order $dy\ dx$: The bounds on y go from curve to curve, i.e., $0 \le y \le 1 - x/2$, and the bounds on x go from point to point, i.e., $0 \leq x \leq 2$.

$$\begin{aligned}
\iint_R \left(3xy - x^2 - y^2 + 6\right) dA &= \int_0^2 \int_0^{-\frac{x}{2}+1} \left(3xy - x^2 - y^2 + 6\right) dy\, dx \\
&= \int_0^2 \left(\frac{3}{2}xy^2 - x^2y - \frac{1}{3}y^3 + 6y\right)\Bigg|_0^{-\frac{x}{2}+1} dx \\
&= \int_0^2 \left(\frac{11}{12}x^3 - \frac{11}{4}x^2 - x + \frac{17}{3}\right) dx \\
&= \left(\frac{11}{48}x^4 - \frac{11}{12}x^3 - \frac{1}{2}x^2 + \frac{17}{3}x\right)\Bigg|_0^2 \\
&= \frac{17}{3} \approx 5.6.
\end{aligned}$$

Now lets consider the order $dx\, dy$. Here x goes from curve to curve, $0 \leq x \leq 2-2y$, and y goes from point to point, $0 \leq y \leq 1$:

$$\begin{aligned}
\iint_R \left(3xy - x^2 - y^2 + 6\right) dA &= \int_0^1 \int_0^{2-2y} \left(3xy - x^2 - y^2 + 6\right) dx\, dy \\
&= \int_0^1 \left(\frac{3}{2}x^2y - \frac{1}{3}x^3 - xy^2 + 6x\right)\Bigg|_0^{2-2y} dy \\
&= \int_0^1 \left(\frac{32}{3}y^3 - 22y^2 + 2y + \frac{28}{3}\right) dy \\
&= \left(\frac{8}{3}y^4 - \frac{22}{3}y^3 + y^2 + \frac{28}{3}y\right)\Bigg|_0^1 \\
&= \frac{17}{3} \approx 5.6.
\end{aligned}$$

We obtained the same result using both orders of integration.

16.2.2 Properties

Note how in these two examples that the bounds of integration depend only on R; the bounds of integration have nothing to do with $f(x,y)$. Moreover, let f and g be continuous functions over a closed, bounded plane region R, and let c be a constant, then we have the following properties, in line with the ones of single integrals.

- Constant multiple rule:
$$\iint_R cf(x,y)\, dA = c\iint_R f(x,y)\, dA.$$
- Sum/Difference rule:
$$\iint_R \big(f(x,y) \pm g(x,y)\big)\, dA = \iint_R f(x,y)\, dA \pm \iint_R g(x,y)\, dA$$

- If $f(x,y) \geq 0$ on R, then
$$\iint_R f(x,y)\ dA \geq 0.$$
- If $f(x,y) \geq g(x,y)$ on R, then
$$\iint_R f(x,y)\ dA \geq \iint_R g(x,y)\ dA.$$
- Let R be the union of two nonoverlapping regions, $R = R_1 \cup R_2$. Then
$$\iint_R f(x,y)\ dA = \iint_{R_1} f(x,y)\ dA + \iint_{R_2} f(x,y)\ dA.$$

Actually, since this property is intuitively clear, we relied already in Examples 16.2 and 16.4. Of course, this property generalizes to n nonoverlapping regions.

Example 16.7

Let $f(x,y) = \sin(x)\cos(y)$ and R be the triangle with vertices $(-1,0)$, $(1,0)$ and $(0,1)$ (see Figure 16.9). Evaluate the double integral $\iint_R f(x,y)\ dA$.

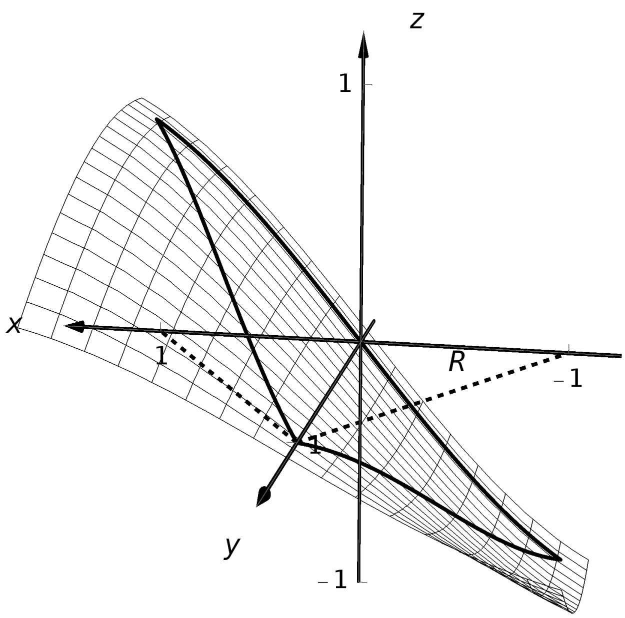


Figure 16.9: Finding the signed volume under a surface in Example 16.7.

Solution

If we attempt to integrate using an iterated integral with the order $dy\ dx$, note how there are two upper bounds on R meaning we will need to use two iterated integrals. We would need to split the triangle into two regions along the y-axis.

Instead, let us use the order $dx\ dy$. The curves bounding x are $y-1 \leq x \leq 1-y$; the bounds on y are $0 \leq y \leq 1$. This gives us:

$$\iint_R f(x,y)\ dA = \int_0^1 \int_{y-1}^{1-y} \sin(x)\cos(y)\ dx\ dy$$

$$= \int_0^1 \Big(-\cos(x)\cos(y) \Big)\Big|_{y-1}^{1-y} dy$$

$$= \int_0^1 \cos(y)\Big(-\cos(1-y) + \cos(y-1) \Big)\, dy.$$

Recall that the cosine function is an even function; that is, $\cos(x) = \cos(-x)$. Therefore, from the last integral above, we have $\cos(y-1) = \cos(1-y)$. Thus the integrand simplifies to 0, and we have

$$\iint_R f(x,y)\, dA = \int_0^1 0\, dy = 0.$$

It turns out that over R, there is just as much volume above the xy-plane as below (look again at Figure 16.9), giving a final signed volume of 0.

In the previous section we practised changing the order of integration of a given iterated integral, where the region R was not explicitly given. Changing the bounds of an integral is more than just an test of understanding. Rather, there are cases where integrating in one order is really hard, if not impossible, whereas integrating with the other order is feasible.

Example 16.8

Rewrite the iterated integral

$$\int_0^3 \int_y^3 e^{-x^2}\, dx\, dy$$

with the order $dy\, dx$. Comment on the feasibility to evaluate each integral.

Solution

Once again we make a sketch of the region over which we are integrating to facilitate changing the order. The bounds on x are from $x = y$ to $x = 3$; the bounds on y are from $y = 0$ to $y = 3$. These curves are sketched in Figure 16.10(a), enclosing the region R.

To change the bounds, note that the curves bounding y are $y = 0$ up to $y = x$; the triangle is enclosed between $x = 0$ and $x = 3$. Thus the new bounds of integration are $0 \leq y \leq x$ and $0 \leq x \leq 3$, giving the iterated integral

$$\int_0^3 \int_0^x e^{-x^2}\, dy\, dx.$$

How easy is it to evaluate each iterated integral? Consider the order of integrating $dx\, dy$, as given in the original problem. The first indefinite integral we need to evaluate is $\int e^{-x^2}\, dx$; we have stated before that this integral cannot be evaluated in terms of elementary functions. We are stuck.

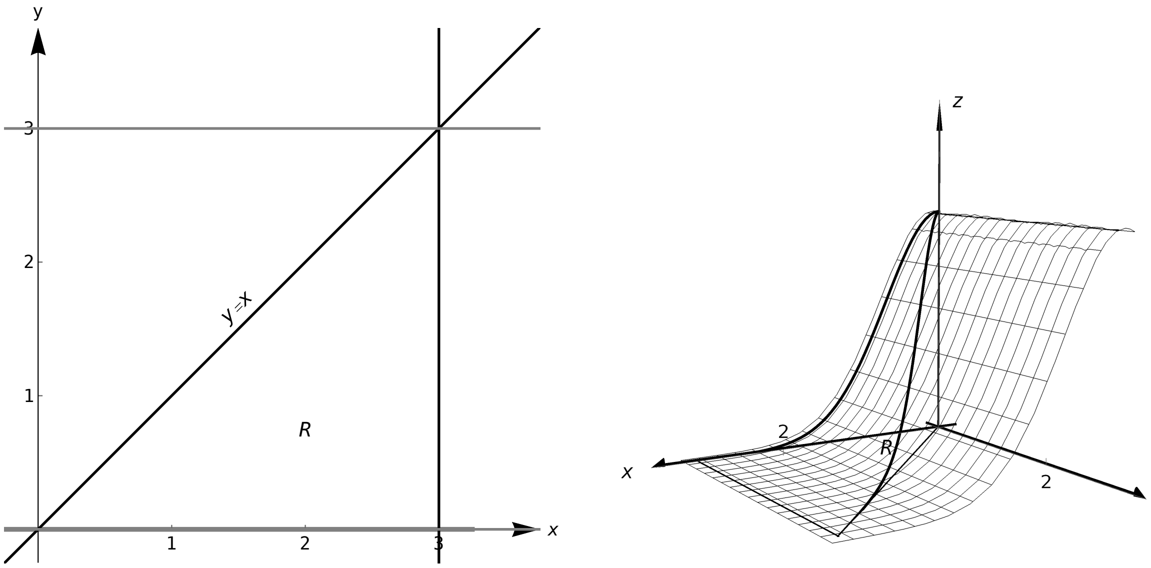
Changing the order of integration makes a big difference here. In the second iterated integral, we are faced with $\int e^{-x^2}\, dy$; integrating with respect to y gives us $ye^{-x^2} + C$, and the first definite integral evaluates to

$$\int_0^x e^{-x^2}\, dy = xe^{-x^2}.$$

Thus

$$\int_0^3 \int_0^x e^{-x^2}\, dy\, dx = \int_0^3 x e^{-x^2}\, dx.$$

This last integral is easy to evaluate with substitution, giving a final answer of $(1-e^{-9})/2 \approx 0.5$. Figure 16.10(b) shows the surface over R. In short, evaluating one iterated integral is impossible; the other iterated integral is relatively simple.



(a) (b)

Figure 16.10: Determining the region R determined by the bounds of integration (a) and the surface f over its region R in Example 16.8.

Using double integrals we can also determine the average value of $z = f(x,y)$ over a region R. This is nothing but the volume under f over R divided by the area of R; that is

$$\text{average value of } f \text{ on } R = \frac{\iint_R f(x,y)\, dA}{\iint_R dA}. \tag{16.2}$$

Example 16.9

Find the average value of $f(x,y) = 4-y$ over the region R, which is bounded by the parabolas $y^2 = 4x$ and $x^2 = 4y$ (see Figure 16.11).

Solution

Graphing each curve can help us find their points of intersection. Solving analytically, the second equation tells us that $y = x^2/4$. Substituting this value in for y in the first equation gives us $x^4/16 = 4x$. Solving for x:

$$\begin{aligned} & \frac{x^4}{16} = 4x \\ \Leftrightarrow \quad & x^4 - 64x = 0 \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow \quad x(x^3-64)=0 \\ &\Leftrightarrow \quad x=0 \ \vee \ x=4. \end{aligned}$$

Thus we have found analytically what was easy to approximate graphically: the regions intersect at $(0,0)$ and $(4,4)$, as shown in Figure 16.11.

We now choose an order of integration: $dy\,dx$ or $dx\,dy$? Either order works; since the integrand does not contain x, choosing $dx\,dy$ might be simpler – at least, the first integral is very simple.

Thus we have the following curve to curve, point to point bounds: $y^2/4 \le x \le 2\sqrt{y}$, and $0 \le y \le 4$.

$$\begin{aligned}
\iint_R (4-y)\,dA &= \int_0^4 \int_{y^2/4}^{2\sqrt{y}} (4-y)\,dx\,dy \\
&= \int_0^4 (4-y) \int_{y^2/4}^{2\sqrt{y}} dx\,dy \\
&= \int_0^4 \left(x(4-y)\right)\Big|_{y^2/4}^{2\sqrt{y}} dy \\
&= \int_0^4 \left[\left(2\sqrt{y}-\frac{y^2}{4}\right)(4-y)\right] dy = \int_0^4 \left(\frac{y^3}{4}-y^2-2y^{3/2}+8y^{1/2}\right) dy \\
&= \left(\frac{y^4}{16}-\frac{y^3}{3}-\frac{4y^{5/2}}{5}+\frac{16y^{3/2}}{3}\right)\Bigg|_0^4 \\
&= \frac{176}{15} \approx 11.73.
\end{aligned}$$

We now find the area of R by computing $\iint_R dA$:

$$\iint_R dA = \int_0^4 \int_{y^2/4}^{2\sqrt{y}} dx\,dy = \frac{16}{3}.$$

Dividing the volume under the surface by the area gives the average value:

$$\text{average value of } f \text{ on } R = \frac{176/15}{16/3} = \frac{11}{5} = 2.2.$$

While the surface covers z-values from $z=0$ to $z=4$, the average z-value on R is 2.2.

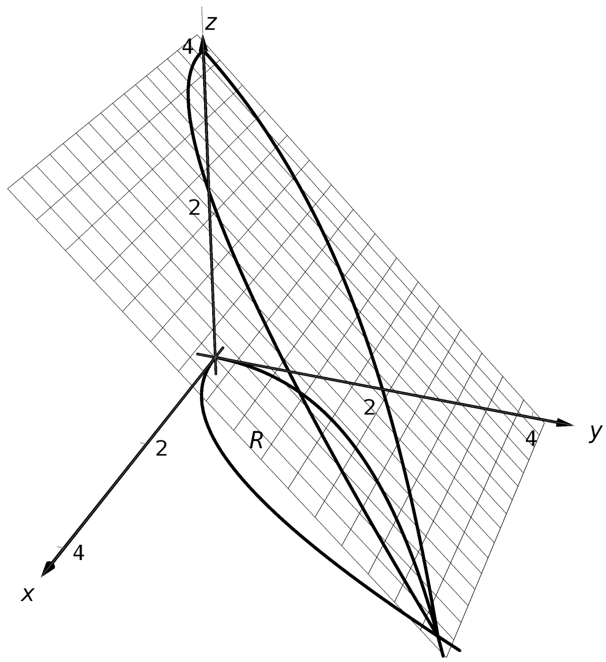


Figure 16.11: Finding the signed volume under a surface in Example 16.9.

Our new understanding of double integrals allows us to revisit what we did in the previous section. Given a region R in the plane, we computed $\iint_R 1\ dA$; again. Our understanding at the time was that we were finding the area of R. However, we can now view the function $z = 1$ as a surface, a flat surface with constant z-value of 1. The double integral $\iint_R 1\ dA$ finds the volume, under $z = 1$, over R. We were actually computing the volume of a solid, though we interpreted the number as an area.

16.2.3 Double integration with polar coordinates

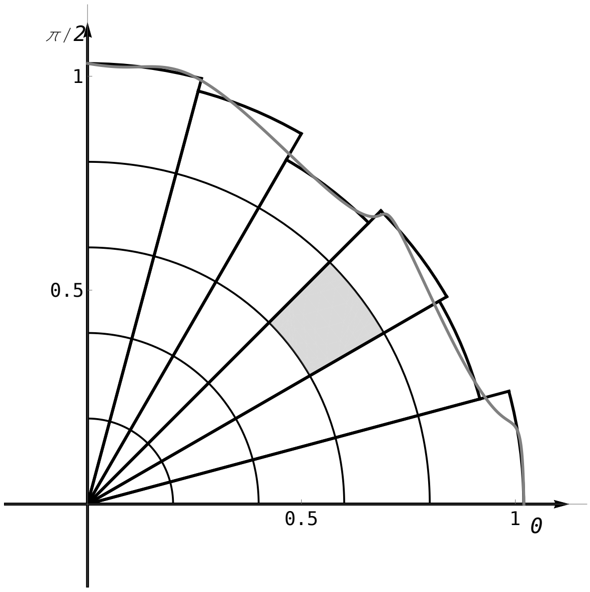
Some regions R are easy to describe using rectangular coordinates – that is, with equations of the form $y = f(x)$, $x = a$, etc. However, some regions are easier to handle if we represent their boundaries with polar equations of the form $r = f(\theta)$, $\theta = \alpha$, etc.

The basic form of the double integral is $\iint_R f(x,y)\ dA$. We interpret this integral as follows: over the region R, sum up lots of products of heights (given by $f(x_i, y_i)$) and areas (given by ΔA_i). That is, dA represents a little bit of area. In rectangular coordinates, we can describe a small rectangle as having area $dx\ dy$ or $dy\ dx$ – the area of a rectangle is simply (length×width) – a small change in x times a small change in y. Thus we replace dA in the double integral with $dx\ dy$ or $dy\ dx$.

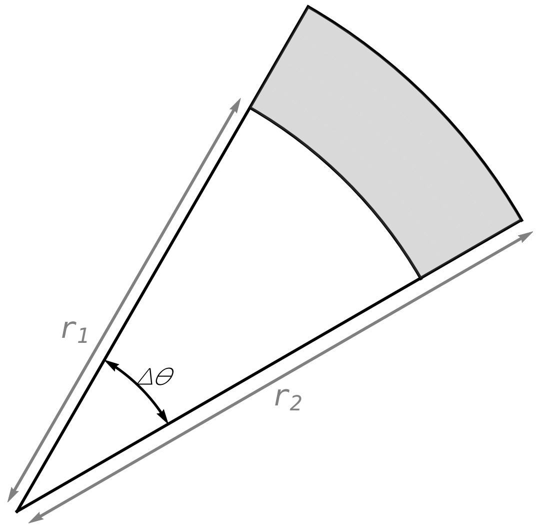
Now consider representing a region R with polar coordinates. Consider Figure 16.12(a). Let R be the region in the first quadrant bounded by the curve. We can approximate this region using the natural shape of polar coordinates: portions of sectors of circles. In the figure, one such region is shaded, shown again in Figure 16.12(b).

As the area of a sector of a circle with radius r, subtended by an angle θ, is $A = \frac{1}{2}r^2\theta$, we can find the area of the shaded region. The whole sector has area $\frac{1}{2}r_2^2\Delta\theta$, whereas the smaller, unshaded sector has area $\frac{1}{2}r_1^2\Delta\theta$. The area of the shaded region is the difference of these areas:

$$\Delta A_i = \frac{1}{2}r_2^2\Delta\theta - \frac{1}{2}r_1^2\Delta\theta = \frac{1}{2}\left(r_2^2 - r_1^2\right)\Delta\theta = \frac{r_2 + r_1}{2}(r_2 - r_1)\Delta\theta.$$



(a)



(b)

Figure 16.12: Approximating a region R with portions of sectors of circles.

Note that $(r_2+r_1)/2$ is just the average of the two radii.

To approximate the region R, we use many such subregions; doing so shrinks the difference r_2-r_1 between radii to 0 and shrinks the change in angle $\Delta\theta$ also to 0. We represent these infinitesimal changes in radius and angle as dr and $d\theta$, respectively. Finally, as dr is small, $r_2 \approx r_1$, and so $(r_2+r_1)/2 \approx r_1$. Thus, when dr and $d\theta$ are small,

$$\Delta A_i \approx r_i\ dr\ d\theta.$$

Taking a limit, where the number of subregions goes to infinity and both r_2-r_1 and $\Delta\theta$ go to 0, we get

$$dA = r\ dr\ d\theta.$$

So to evaluate $\iint_R f(x,y)\ dA$, replace dA with $r\ dr\ d\theta$. Convert the function $z=f(x,y)$ to a function with polar coordinates with the substitutions $x=r\cos(\theta)$, $y=r\sin(\theta)$. Finally, find bounds $g_1(\theta) \le r \le g_2(\theta)$ and $\alpha \le \theta \le \beta$ that describe R. Consequently, if $z=f(x,y)$ is a continuous function defined over a closed, bounded region R in the xy-plane, where R is bounded by the polar equations $\alpha \le \theta \le \beta$ and $g_1(\theta) \le r \le g_2(\theta)$, then

$$\iint_R f(x,y)\ dA = \int_\alpha^\beta \int_{g_1(\theta)}^{g_2(\theta)} f\big(r\cos(\theta), r\sin(\theta)\big)\ r\ dr\ d\theta. \tag{16.3}$$

Examples will help us understand this.

Example 16.10

Find the signed volume under the plane $z=4-x-2y$ over the disk bounded by the circle with equation $x^2+y^2=1$.

Solution

The bounds of the integral are determined solely by the region R over which we are integrating.

In this case, it is a disk with boundary $x^2 + y^2 = 1$. We need to find polar bounds for this region. Bounds for this disk are $0 \le r \le 1$ and $0 \le \theta \le 2\pi$.

We replace $f(x, y)$ with $f\big(r\cos(\theta), r\sin(\theta)\big)$. That means we make the following substitutions:

$$4 - x - 2y \quad \Rightarrow \quad 4 - r\cos(\theta) - 2r\sin(\theta).$$

Finally, we replace dA in the double integral with $r\ dr\ d\theta$. This gives the final iterated integral, which we evaluate:

$$\begin{aligned}
\iint_R f(x,y)\ dA &= \int_0^{2\pi}\int_0^1 \big(4 - r\cos(\theta) - 2r\sin(\theta)\big) r\ dr\ d\theta \\
&= \int_0^{2\pi}\int_0^1 \Big(4r - r^2\big(\cos(\theta) - 2\sin(\theta)\big)\Big)\ dr\ d\theta \\
&= \int_0^{2\pi} \left(2r^2 - \frac{1}{3}r^3\big(\cos(\theta) - 2\sin(\theta)\big)\right)\Bigg|_0^1\ d\theta \\
&= \int_0^{2\pi} \left(2 - \frac{1}{3}\big(\cos(\theta) - 2\sin(\theta)\big)\right)\ d\theta \\
&= \left(2\theta - \frac{1}{3}\big(\sin(\theta) + 2\cos(\theta)\big)\right)\Bigg|_0^{2\pi} \\
&= 4\pi \approx 12.566.
\end{aligned}$$

The surface and region R are shown in Figure 16.13.

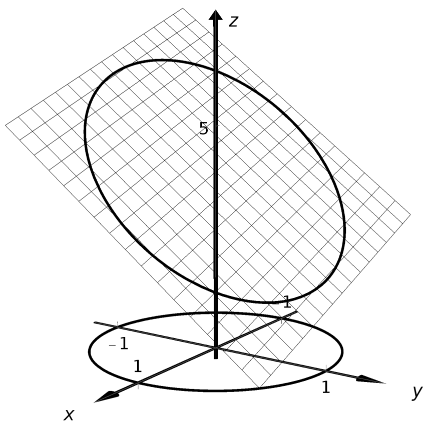


Figure 16.13: Evaluating a double integral with polar coordinates in Example 16.10.

Example 16.11

Find the volume under the paraboloid $z = 4 - (x-2)^2 - y^2$ over the region bounded by the circles $(x-1)^2 + y^2 = 1$ and $(x-2)^2 + y^2 = 4$.

Solution

At first glance, this seems like a very hard volume to compute as the region R (shown in Figure 16.14(a)) has a hole in it, cutting out a strange portion of the surface, as shown in Figure 16.14(b). However, by describing R in terms of polar equations, the volume is not very difficult to compute.

The circle $(x-1)^2+y^2=1$ has polar equation $r=2\cos(\theta)$, while the circle $(x-2)^2+y^2=4$ has polar equation $r=4\cos(\theta)$. We may trace out semicircles on the interval $0\le\theta\le\pi/2$. The bounds on r are $2\cos(\theta)\le r\le 4\cos(\theta)$. Replacing x with $r\cos(\theta)$ in the integrand, along with replacing y with $r\sin(\theta)$ and noting that we should add a factor 2 to account for the entire volume, prepares us to evaluate the double integral $\iint_R f(x,y)\,dA$:

$$\begin{aligned}
\iint_R f(x,y)\,dA &= 2\int_0^{\pi/2}\int_{2\cos(\theta)}^{4\cos(\theta)}\Big(4-\big(r\cos(\theta)-2\big)^2-\big(r\sin(\theta)\big)^2\Big)r\,dr\,d\theta\\
&= 2\int_0^{\pi/2}\int_{2\cos(\theta)}^{4\cos(\theta)}\big(-r^3+4r^2\cos(\theta)\big)\,dr\,d\theta\\
&= 2\int_0^{\pi/2}\left(-\frac{1}{4}r^4+\frac{4}{3}r^3\cos(\theta)\right)\Bigg|_{2\cos(\theta)}^{4\cos(\theta)}d\theta\\
&= 2\int_0^{\pi/2}\Bigg(\left[-\frac{1}{4}\Big(256\cos^4(\theta)\Big)+\frac{4}{3}\Big(64\cos^4(\theta)\Big)\right]-\\
&\qquad\left[-\frac{1}{4}\Big(16\cos^4(\theta)\Big)+\frac{4}{3}\Big(8\cos^4(\theta)\Big)\right]\Bigg)\,d\theta\\
&= 2\int_0^{\pi/2}\frac{44}{3}\cos^4(\theta)\,d\theta.
\end{aligned}$$

To integrate $\cos^4(\theta)$, rewrite it as $\cos^2(\theta)\cos^2(\theta)$ and employ the power-reducing formula twice:

$$\begin{aligned}
\cos^4(\theta) &= \cos^2(\theta)\cos^2(\theta)\\
&= \frac{1}{2}\big(1+\cos(2\theta)\big)\frac{1}{2}\big(1+\cos(2\theta)\big)\\
&= \frac{1}{4}\big(1+2\cos(2\theta)+\cos^2(2\theta)\big)\\
&= \frac{1}{4}\Big(1+2\cos(2\theta)+\frac{1}{2}\big(1+\cos(4\theta)\big)\Big)\\
&= \frac{3}{8}+\frac{1}{2}\cos(2\theta)+\frac{1}{8}\cos(4\theta).
\end{aligned}$$

Picking up from where we left off above, we have

$$\iint_R f(x,y)\,dA = 2\int_0^{\pi/2}\frac{44}{3}\cos^4(\theta)\,d\theta$$

$$= 2\int_0^{\pi/2} \frac{44}{3}\left(\frac{3}{8} + \frac{1}{2}\cos(2\theta) + \frac{1}{8}\cos(4\theta)\right) d\theta$$

$$= \frac{88}{3}\left(\frac{3}{8}\theta + \frac{1}{4}\sin(2\theta) + \frac{1}{32}\sin(4\theta)\right)\Bigg|_0^{\pi/2}$$

$$= \frac{11}{2}\pi \approx 17.279.$$

Note that the double integral would have been much harder to evaluate had we used rectangular coordinates.

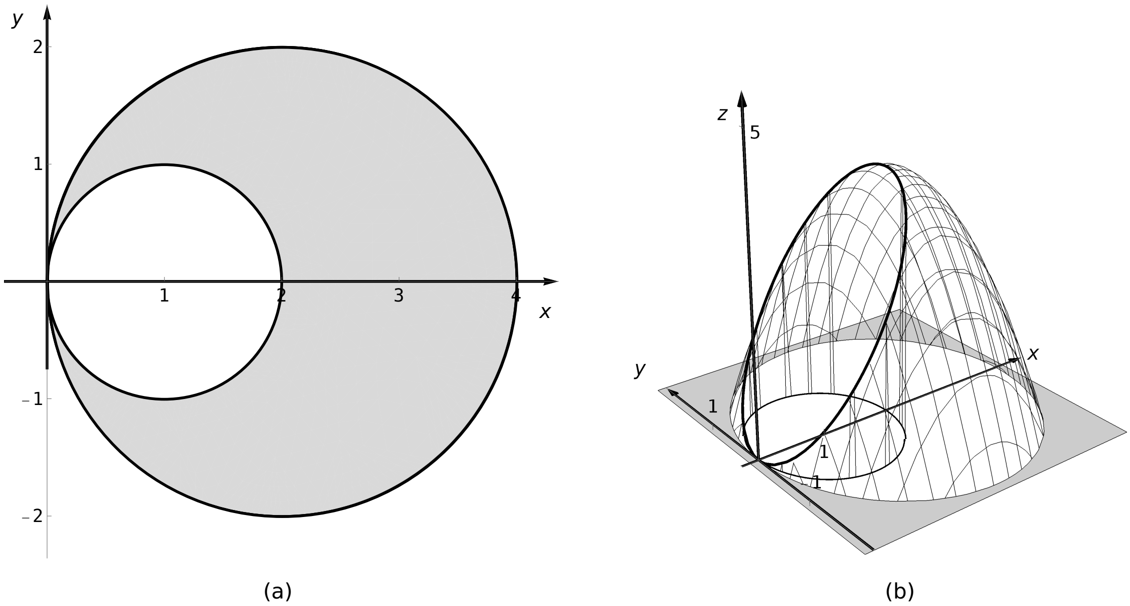


Figure 16.14: Showing the region R (a) and surface (b) used in Example 16.11

Example 16.12

Find the volume of a sphere with radius a.

Solution

The sphere of radius a, centred at the origin, has equation $x^2 + y^2 + z^2 = a^2$; solving for z, we have

$$z = \pm\sqrt{a^2 - x^2 - y^2}.$$

The half solution $z > 0$ gives the upper half of a sphere. We wish to find the volume under this top half, then double it to find the total volume.

The region we need to integrate over is the disk of radius a, centred at the origin. Polar bounds for this equation are $0 \le r \le a$, $0 \le \theta \le 2\pi$.

All together, the volume of a sphere with radius a is:

$$2\iint_R \sqrt{a^2 - x^2 - y^2}\, dA = 2\int_0^{2\pi}\int_0^{a} \sqrt{a^2 - \Big(r\cos(\theta)\Big)^2 - \Big(r\sin(\theta)\Big)^2}\, r\, dr\, d\theta$$

$$= 2\int_0^{2\pi}\int_0^{a} r\sqrt{a^2-r^2}\,dr\,d\theta.$$

We can evaluate this inner integral with substitution. With $u = a^2 - r^2$, $du = -2r\ dr$. The new bounds of integration are $u(0) = a^2$ to $u(a) = 0$. Thus we have:

$$\begin{aligned}
&= \int_0^{2\pi}\int_{a^2}^{0} \left(-u^{1/2}\right) du\,d\theta \\
&= \int_0^{2\pi} \left(-\frac{2}{3}u^{3/2}\right)\Bigg|_{a^2}^{0} d\theta \\
&= \int_0^{2\pi} \left(\frac{2}{3}a^3\right) d\theta \\
&= \left(\frac{2}{3}a^3\theta\right)\Bigg|_0^{2\pi} \\
&= \frac{4}{3}\pi a^3.
\end{aligned}$$

Generally, the formula for the volume of a sphere with radius r is given as $4\pi r^3/3$; we have justified this formula with our calculation.

We have used iterated integrals to find areas of plane regions and volumes under surfaces. Just as a single integral can be used to compute much more than area under the curve, iterated integrals can be used to compute much more than we have thus far seen. The next two sections show two, among many, applications of iterated integrals.

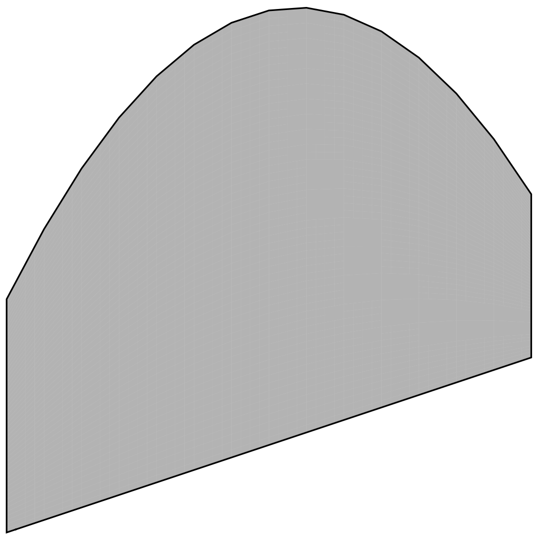
16.3 Centre of mass

We have used iterated integrals to find areas of plane regions and signed volumes under surfaces. Here, we will apply iterated integrals to compute the **mass** and **centre of mass** (*massamiddelpunt*) of planar regions.

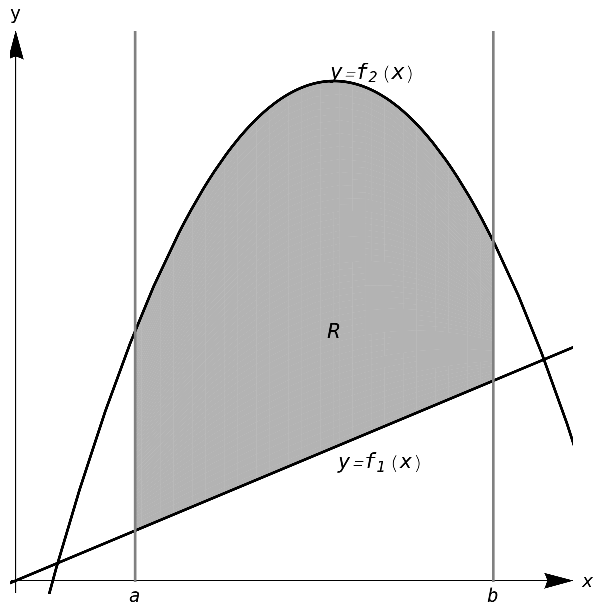
16.3.1 Mass and weight

Consider a thin sheet of material with constant thickness and finite area. Mathematicians (and physicists and engineers) call such a sheet a lamina. So consider a lamina, as shown in Figure 16.15(a), with the shape of some planar region R, as shown in Figure 16.15(b).

We can write a simple double integral that represents the mass of the lamina: $\iint_R\ dm$, where dm means a little mass. That is, the double integral states the total mass of the lamina can be found by summing up lots of little masses over R. To evaluate this double integral, partition R into n subregions as we have done in the past. The i^{th} subregion has area ΔA_i. A fundamental property of mass is that "mass=density×area". If the lamina has a constant density δ, then the mass of this i^{th} subregion is $\Delta m_i = \delta\Delta A_i$. That is, we can compute a small amount of mass by multiplying a small amount of area by the density.



(a)



(b)

Figure 16.15: Illustrating the concept of a lamina.

If density is variable, with density function $\delta = \delta(x,y)$, then we can approximate the mass of the i^{th} subregion of R by multiplying ΔA_i by $\delta(x_i, y_i)$, where (x_i, y_i) is a point in that subregion. That is, for a small enough subregion of R, the density across that region is almost constant.

Note that mass and weight are different measures. Since they are scalar multiples of each other, it is often easy to treat them as the same measure. Here, we effectively treat them as the same, as our technique for finding mass is the same as for finding weight. The density functions used will simply have different units.

The total mass M of the lamina is approximately the sum of approximate masses of subregions:

$$M \approx \sum_{i=1}^{n} \Delta m_i = \sum_{i=1}^{n} \delta(x_i, y_i)\Delta A_i.$$

Taking the limit as the size of the subregions shrinks to 0 gives us the actual mass; that is, integrating $\delta(x,y)$ over R gives the mass of the lamina:

$$M = \iint_R dm = \iint_R \delta(x,y)\, dA. \tag{16.4}$$

Example 16.13

Find the mass of a square lamina, represented by the unit square with lower lefthand corner at the origin, with variable density $\delta(x,y) = (x+y+2)\text{g/cm}^2$.

Solution

The variable density δ, in this example, is very uniform, giving a density of 3 in the centre of the square and changing linearly. A graph of $\delta(x,y)$ can be seen in Figure 16.16; notice how same amount of density is above $z = 3$ as below. We'll comment on the significance of this momentarily.

The mass M is found by integrating $\delta(x,y)$ over R. The order of integration is not important; we choose $dx\ dy$ arbitrarily. Thus:

$$\begin{aligned}
M = \iint_R (x+y+2)\ dA &= \int_0^1 \int_0^1 (x+y+2)\ dx\ dy \\
&= \int_0^1 \left(\frac{1}{2}x^2 + x(y+2) \right)\Bigg|_0^1\ dy \\
&= \int_0^1 \left(\frac{5}{2} + y \right)\ dy \\
&= \left(\frac{5}{2}y + \frac{1}{2}y^2 \right)\Bigg|_0^1 = 3\text{g}.
\end{aligned}$$

It turns out that since the density of the lamina is so uniformly distributed above and below $z = 3$ that the mass of the lamina is the same as if it had a constant density of 3. The density function $\delta = 3\text{g/cm}^2$ and the one from this example are graphed in Figure 16.16, which illustrates this concept.

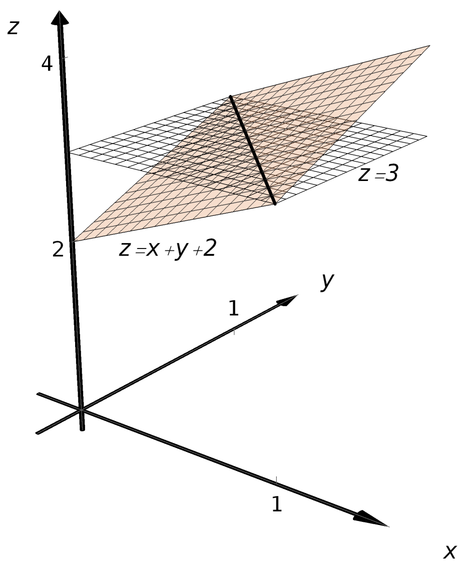


Figure 16.16: Graphing the density functions $\delta = 3\text{g/cm}^2$ and $\delta(x,y) = (x+y+2)\text{g/cm}^2$.

Example 16.14

Find the weight of the lamina represented by the disk with radius 2cm, centred at the origin, with density function $\delta(x,y) = (x^2+y^2+1)\text{g/cm}^2$. Compare this to the weight of the lamina with the same shape and density $\delta(x,y) = (2\sqrt{x^2+y^2}+1)\text{g/cm}^2$.

Solution

A direct application of Equation (16.4) states that the weight of the lamina is $\iint_R \delta(x,y)\ dA$. Since our lamina is in the shape of a circle, it makes sense to approach the double integral using polar coordinates.

The density function $\delta(x,y) = x^2 + y^2 + 1$ becomes

$$\delta(r,\theta) = (r\cos(\theta))^2 + (r\sin(\theta))^2 + 1 = r^2 + 1.$$

The circle is bounded by $0 \le r \le 2$ and $0 \le \theta \le 2\pi$. Thus the weight W is:

$$\begin{aligned} W &= \int_0^{2\pi}\int_0^2 (r^2+1)r\ dr\ d\theta \\ &= \int_0^{2\pi} \left(\frac{1}{4}r^4 + \frac{1}{2}r^2\right)\Bigg|_0^2\ d\theta \\ &= \int_0^{2\pi} 6\ d\theta \\ &= 12\pi \approx 37.70\text{g}. \end{aligned}$$

Now compare this with the density function $\delta(x,y) = 2\sqrt{x^2+y^2}+1$. Converting this to polar coordinates gives

$$\delta(r,\theta) = 2\sqrt{\Big(r\cos(\theta)\Big)^2 + \Big(r\sin(\theta)\Big)^2} + 1 = 2r + 1.$$

Thus the weight W is:

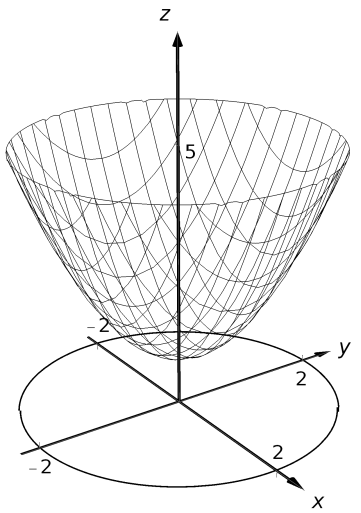
$$\begin{aligned} W &= \int_0^{2\pi}\int_0^2 (2r+1)r\ dr\ d\theta \\ &= \int_0^{2\pi} \left(\frac{2}{3}r^3 + \frac{1}{2}r^2\right)\Bigg|_0^2 d\theta \\ &= \int_0^{2\pi} \left(\frac{22}{3}\right)\ d\theta \\ &= \frac{44}{3}\pi \approx 46.08\text{g}. \end{aligned}$$

One would expect different density functions to return different weights, as we have here. The density functions were chosen, though, to be similar: each gives a density of 1 at the origin and a density of 5 at the outside edge of the circle, as seen in Figure 16.17.

Notice how $x^2 + y^2 + 1 \le 2\sqrt{x^2+y^2}+1$ over the circle; this results in less weight.

Plotting the density functions can be useful as our understanding of mass can be related to our understanding of volume under a surface. We interpreted $\iint_R f(x,y)\ dA$ as giving the volume under f over R; we can understand $\iint_R \delta(x,y)\ dA$ in the same way. The volume under δ over R is actually mass; by compressing the volume under δ onto the xy-plane, we get more mass in some areas than others – i.e., areas of greater density.

Knowing the mass of a lamina is one of several important measures. Another is the centre of mass, which we discuss next.



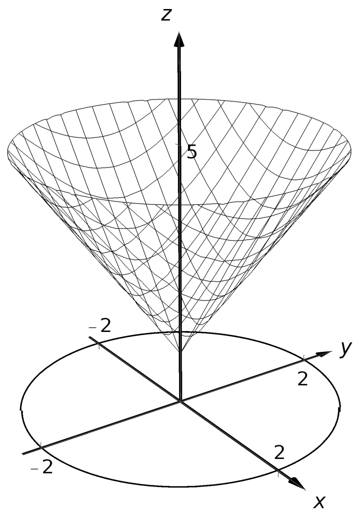


Figure 16.17: Graphing the density functions $\delta(x,y) = x^2 + y^2 + 1$ (a) and $\delta(x,y) = 2\sqrt{x^2+y^2}+1$ (b).

16.3.2 Centre of mass

Consider a disk of radius 1 with uniform density. It is common knowledge that the disk will balance on a point if the point is placed at the centre of the disk. What if the disk does not have a uniform density? Through trial-and-error, we should still be able to find a spot on the disk at which the disk will balance on a point. This balance point is referred to as the **centre of mass** (*massamiddelpunt*), or **centre of gravity** (*zwaartepunt*). It is though all the mass is centred there. In fact, if the disk has a mass of 3kg, the disk will behave physically as though it were a point mass of 3kg located at its centre of mass. For instance, the disk will naturally spin with an axis through its centre of mass.

We find the centre of mass based on the principle of a weighted average. Consider a college class in which your homework average is 90%, your test average is 73%, and your final exam grade is an 85%. Experience tells us that our final grade is not the *average* of these three grades: that is, it is not:

$$\frac{0.9+0.73+0.85}{3} \approx 0.837 = 83.7\%.$$

That is, you are probably not pulling a B in the course. Rather, your grades are weighted. Let us say the homework is worth 10% of the grade, tests are 60% and the exam is 30%. Then your final grade is:

$$(0.1)(0.9)+(0.6)(0.73)+(0.3)(0.85) = 0.783 = 78.3\%.$$

Each grade is multiplied by a weight.

In general, given values $x_1, x_2, \ldots, x_n$ and weights $w_1, w_2, \ldots, w_n$, the weighted average of the n-values is

$$\frac{\displaystyle\sum_{i=1}^{n} w_i x_i}{\displaystyle\sum_{i=1}^{n} w_i}.$$

How this relates to centre of mass is given in the following definition.

Definitie 16.2 (Centre of mass of a discrete linear system)
Let point masses $m_1, m_2, \ldots, m_n$ be distributed along the x-axis at locations $x_1, x_2, \ldots, x_n$, respectively. The **centre of mass** $\overline{x}$ of the system is located at

$$\overline{x} = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}.$$

In a discrete system (i.e., mass is located at individual points, not along a continuum) we find the centre of mass by dividing the mass into a **moment** (*moment*) of the system. In general, a moment is a weighted measure of distance from a particular point or line. In the case described by Definition 16.2, we are finding a weighted measure of distances from the y-axis, so we refer to this as **the moment about the y-axis** (*moment om de y-as*), represented by M_y. Letting M be the total mass of the system, we have $\overline{x} = M_y / M$.

We can extend the concept of the centre of mass of discrete points along a line to the centre of mass of discrete points in the plane rather easily. To do so, we define some terms then give a theorem.

Definitie 16.3 (Moments about the x- and y- axes)
Let point masses m_1, $m_2, \ldots, m_n$ be located at points (x_1, y_1), $(x_2, y_2) \ldots, (x_n, y_n)$, respectively, in the xy-plane.

1. The **moment about the x-axis**, M_x, is

$$M_x = \sum_{i=1}^{n} m_i y_i.$$

2. The **moment about the y-axis**, M_y, is

$$M_y = \sum_{i=1}^{n} m_i x_i.$$

We now define the centre of mass of discrete points in the plane.

Definitie 16.4 (Centre of mass of a discrete planar system)
Let point masses m_1, $m_2, \ldots, m_n$ be located at points (x_1, y_1), $(x_2, y_2) \ldots, (x_n, y_n)$, respectively, in the xy-plane, and let $M = \sum_{i=1}^{n} m_i$.

The **centre of mass** of the system is at $(\overline{x}, \overline{y})$, where

$$\overline{x} = \frac{M_y}{M} \quad \text{and} \quad \overline{y} = \frac{M_x}{M}.$$

Example 16.15

Let point masses of 1kg, 2kg and 5kg be located at points $(2,0)$, $(1,1)$ and $(3,1)$, respectively, and are connected by thin rods of negligible weight. Find the centre of mass of the system.

Solution

We follow Definitions 16.4 and 16.3 to find M, M_x and M_y:

$M = 1 + 2 + 5 = 8$kg.

$$M_x = \sum_{i=1}^{n} m_i y_i = 1(0) + 2(1) + 5(1) = 7. \qquad M_y = \sum_{i=1}^{n} m_i x_i = 1(2) + 2(1) + 5(3) = 19.$$

Thus the centre of mass is

$$(\overline{x}, \overline{y}) = \left(\frac{M_y}{M}, \frac{M_x}{M}\right) = \left(\frac{19}{8}, \frac{7}{8}\right) = (2.375, 0.875),$$

illustrated in Figure 16.18.

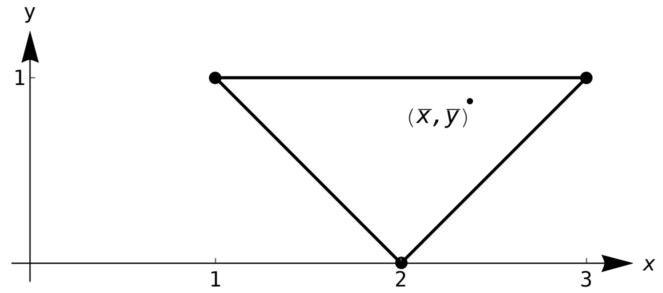


Figure 16.18: Illustrating the centre of mass of a discrete planar system in Example 16.15.

We finally arrive at our true goal of this section: finding the centre of mass of a lamina with variable density. While the above measurement of centre of mass is interesting, it does not directly answer more realistic situations where we need to find the centre of mass of a contiguous region. However, understanding the discrete case allows us to approximate the centre of mass of a planar lamina; using calculus, we can refine the approximation to an exact value.

We begin by representing a planar lamina with a region R in the xy-plane with density function $\delta(x, y)$. Partition R into n subdivisions, each with area ΔA_i. As done before, we can approximate the mass of the i^{th} subregion with $\delta(x_i, y_i)\Delta A_i$, where (x_i, y_i) is a point inside the i^{th} subregion. We can approximate the moment of this subregion about the y-axis with $x_i\delta(x_i, y_i)\Delta A_i$ – that is, by multiplying the approximate mass of the region by its approximate distance from the y-axis. Similarly, we can approximate the moment about the x-axis with $y_i\delta(x_i, y_i)\Delta A_i$. By summing over all subregions, we have:

$$\text{mass: } M \approx \sum_{i=1}^{n} \delta(x_i, y_i)\Delta A_i\,, \qquad \text{(as seen before)}$$

$$\text{moment about the } x\text{-axis: } M_x \approx \sum_{i=1}^{n} y_i\delta(x_i, y_i)\Delta A_i\,,$$

$$\text{moment about the } y\text{-axis: } M_y \approx \sum_{i=1}^{n} x_i\delta(x_i, y_i)\Delta A_i\,.$$

By taking limits, where size of each subregion shrinks to 0 in both the x- and y- directions, we arrive at the double integrals given in the following definition.

Definitie 16.5 (Centre of mass of a planar lamina)
Let a planar lamina be represented by a closed, bounded region R in the xy-plane with density function $\delta(x,y)$. Then we can infer the following information about the lamina:

1. The **mass** (*massa*) of a planar lamina is $M = \iint_R \delta(x,y)\ dA$.
2. The **moment about the x-axis** is $M_x = \iint_R y\delta(x,y)\ dA$.
3. The **moment about the y-axis** is $M_y = \iint_R x\delta(x,y)\ dA$.
4. The **centre of mass** (*massamiddelpunt*) of the object is

$$(\overline{x},\overline{y}) = \left(\frac{M_y}{M}, \frac{M_x}{M}\right).$$

We practice finding centres of mass by revisiting some of the lamina used previously in this section when finding mass. We will just set up the integrals needed to compute M, M_x and M_y and leave the details of the integration to the reader.

Example 16.16

Find the centre of mass of a square lamina, represented by the unit square with lower lefthand corner at the origin (see Figure 16.16), with variable density $\delta(x,y) = (x+y+2)\text{g/cm}^2$. This is the lamina from Example 16.13.

Solution

We follow Theorem 16.5, to find M, M_x and M_y:

$$M = \iint_R (x+y+2)\ dA = \int_0^1 \int_0^1 (x+y+2)\ dx\ dy = 3\text{g}.$$

$$M_x = \iint_R y(x+y+2)\ dA = \int_0^1 \int_0^1 y(x+y+2)\ dx\ dy = \frac{19}{12}.$$

$$M_y = \iint_R x(x+y+2)\ dA = \int_0^1 \int_0^1 x(x+y+2)\ dx\ dy = \frac{19}{12}.$$

Thus the centre of mass is

$$(\overline{x},\overline{y}) = \left(\frac{M_y}{M}, \frac{M_x}{M}\right) = \left(\frac{19}{36}, \frac{19}{36}\right) \approx (0.528, 0.528).$$

While the mass of this lamina is the same as the lamina in the previous example, the greater density found with greater x- and y-values pulls the centre of mass from the centre slightly towards the upper righthand corner.

Example 16.17

Find the centre of mass of the lamina represented by the circle with radius 2cm, centred at the origin, with density function $\delta(x,y) = (x^2+y^2+1)\text{g/cm}^2$. This is one of the lamina used in Example

16.14.

Solution

As done in Example 16.14, it is best to describe R using polar coordinates. Thus when we compute M_y, we will integrate not $x\delta(x,y) = x(x^2+y^2+1)$, but rather $\big(r\cos(\theta)\big)\delta(r\cos(\theta), r\sin(\theta)) = \big(r\cos(\theta)\big)\big(r^2+1\big)$. We compute M, M_x and M_y:

$$M = \int_0^{2\pi}\int_0^2 (r^2+1)r\,dr\,d\theta = 12\pi \approx 37.7\text{g}\,,$$

$$M_x = \int_0^{2\pi}\int_0^2 (r\sin(\theta))(r^2+1)r\,dr\,d\theta = 0\,,$$

$$M_y = \int_0^{2\pi}\int_0^2 (r\cos(\theta))(r^2+1)r\,dr\,d\theta = 0\,.$$

Since R and the density of R are both symmetric about the x- and y-axes, it should come as no big surprise that the moments about each axis is 0. Thus the centre of mass is $(\overline{x},\overline{y}) = (0,0)$.

16.4 Surface area

In Section 13.4 we used definite integrals to compute the arc length of plane curves of the form $y = f(x)$. We later extended these ideas to compute the arc length of plane curves defined by parametric or polar equations.

The natural extension of the concept of arc length over an interval to surfaces is surface area over a region. For that purpose, consider the surface $z = f(x,y)$ over a region R in the xy-plane, shown in Figure 16.19(a). Because of the domed shape of the surface, the surface area will be greater than that of the area of the region R. We can find this area using the same basic technique we have used over and over: we'll make an approximation, then using limits, we'll refine the approximation to the exact value.

As done to find the volume under a surface or the mass of a lamina, we subdivide R into n subregions. Here we subdivide R into rectangles, as shown in the figure. One such subregion is outlined in the figure, where the rectangle has dimensions Δx_i and Δy_i, along with its corresponding region on the surface.

In Figure 16.19(b), we zoom in on this portion of the surface. When Δx_i and Δy_i are small, the function is approximated well by the tangent plane at any point (x_i, y_i) in this subregion, which is graphed in part (b). In fact, the tangent plane approximates the function so well that in this figure, it is virtually indistinguishable from the surface itself! Therefore we can approximate the surface area S_i of this region of the surface with the area T_i of the corresponding portion of the tangent plane.

This portion of the tangent plane is a parallelogram, defined by sides $\vec{u}$ and $\vec{v}$, as shown. One of the applications of the cross product from Section 6.6 is that the area of this parallelogram is $\|\vec{u}\times\vec{v}\|$. So, once we can determine $\vec{u}$ and $\vec{v}$, we can determine the area.

$\vec{u}$ is tangent to the surface in the direction of x, therefore, from Section 15.7, $\vec{u}$ is parallel to $\big(1,0,f_x(x_i,y_i)\big)$. The x-displacement of $\vec{u}$ is Δx_i, so we know that $\vec{u} = \Delta x_i\big(1,0,f_x(x_i,y_i)\big)$. Similar logic shows that

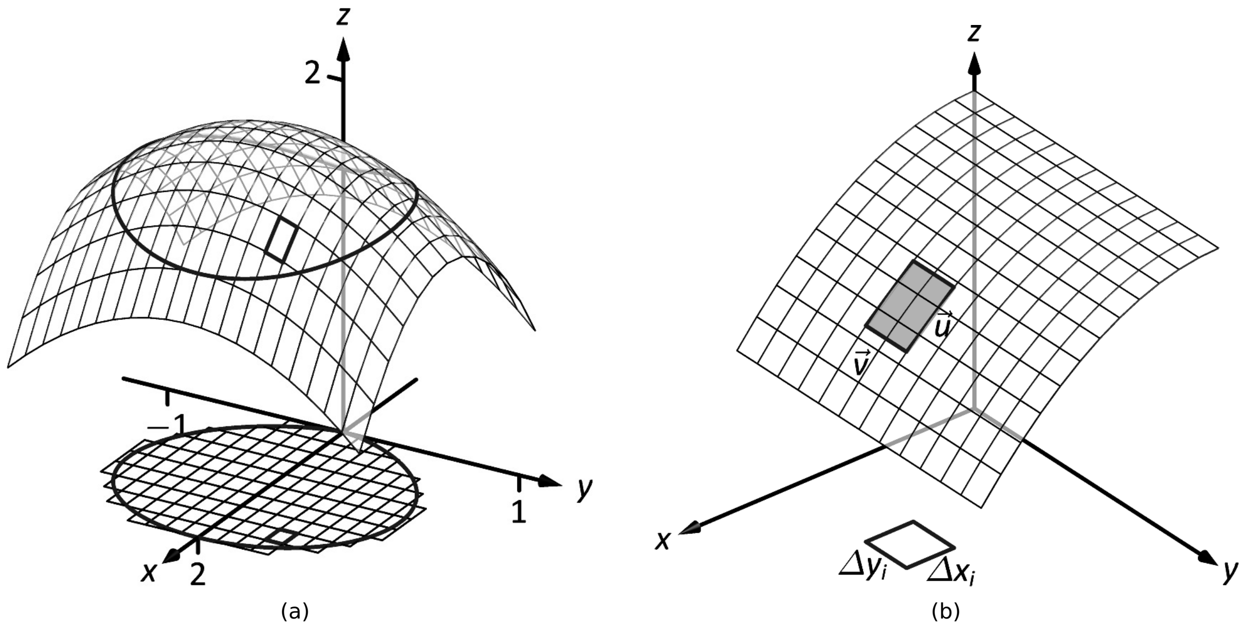


Figure 16.19: Developing a method of computing surface area.

$\vec{\mathbf{v}} = \Delta y_i\big(0, 1, f_y(x_i, y_i)\big)$.Thus:

$$\begin{aligned}\text{surface area } S_i \approx \text{area of } T_i\\ &= \|\vec{\mathbf{u}} \times \vec{\mathbf{v}}\|\\ &= \Big|\Big|\Delta x_i\big(1, 0, f_x(x_i, y_i)\big) \times \Delta y_i\big(0, 1, f_y(x_i, y_i)\big)\Big|\Big|\\ &= \sqrt{1 + f_x(x_i, y_i)^2 + f_y(x_i, y_i)^2}\Delta x_i \Delta y_i.\end{aligned}$$

Note that $\Delta x_i \Delta y_i = \Delta A_i$, the area of the i^{th} subregion.

Summing up all n of the approximations to the surface area gives

$$\text{surface area over } R \approx \sum_{i=1}^{n} \sqrt{1 + f_x(x_i, y_i)^2 + f_y(x_i, y_i)^2}\Delta A_i.$$

Once again take a limit as all of the Δx_i and Δy_i shrink to 0; this leads to a double integral:

$$SA = \iint_R \sqrt{1 + f_x(x, y)^2 + f_y(x, y)^2}\ dA. \tag{16.5}$$

We use this definition to compute surface areas of known surfaces.

Example 16.18

Find the surface area of the sphere with radius a centred at the origin, whose top hemisphere has equation $f(x, y) = \sqrt{a^2 - x^2 - y^2}$.

Solution

We start by computing partial derivatives and find

$$f_x(x,y)=\frac{-x}{\sqrt{a^2-x^2-y^2}} \quad \text{and} \quad f_y(x,y)=\frac{-y}{\sqrt{a^2-x^2-y^2}}.$$

As our function f only defines the top upper hemisphere of the sphere, we double our surface area result to get the total area:

$$\begin{aligned} SA &= 2\iint_R \sqrt{1+f_x(x,y)^2+f_y(x,y)^2}\ dA \\ &= 2\iint_R \sqrt{1+\frac{x^2+y^2}{a^2-x^2-y^2}}\ dA. \end{aligned}$$

The region R that we are integrating over is bounded by the circle, centered at the origin, with radius a: $x^2+y^2=a^2$. Because of this region, we are likely to have greater success with our integration by converting to polar coordinates. Using the substitutions $x=r\cos(\theta)$, $y=r\sin(\theta)$, $dA=r\ dr\ d\theta$ and bounds $0\le\theta\le 2\pi$ and $0\le r\le a$, we have:

$$\begin{aligned} SA &= 2\int_0^{2\pi}\int_0^{a}\sqrt{1+\frac{r^2\cos^2(\theta)+r^2\sin^2(\theta)}{a^2-r^2\cos^2(\theta)-r^2\sin^2(\theta)}}\ r\ dr\ d\theta \\ &= 2\int_0^{2\pi}\int_0^{a} r\sqrt{1+\frac{r^2}{a^2-r^2}}\ dr\ d\theta \\ &= 2\int_0^{2\pi}\int_0^{a} r\sqrt{\frac{a^2}{a^2-r^2}}\ dr\ d\theta. \end{aligned} \tag{16.6}$$

Apply substitution $u=a^2-r^2$ and integrate the inner integral, giving

$$\begin{aligned} &= 2\int_0^{2\pi} a^2\ d\theta \\ &= 4\pi a^2. \end{aligned}$$

Our work confirms the known formula.

Note that the inner integral in Equation (16.6) is an improper integral, as its is not defined at $r=a$. To properly evaluate this integral, one must use the techniques of Section 12.5. Since the resulting improper integral does converge, the surface area is accurately computed.

Example 16.19

Find the area of the surface $f(x,y)=x^2-3y+3$ over the region R bounded by $-x\le y\le x$, $0\le x\le 4$, as pictured in Figure 16.20.

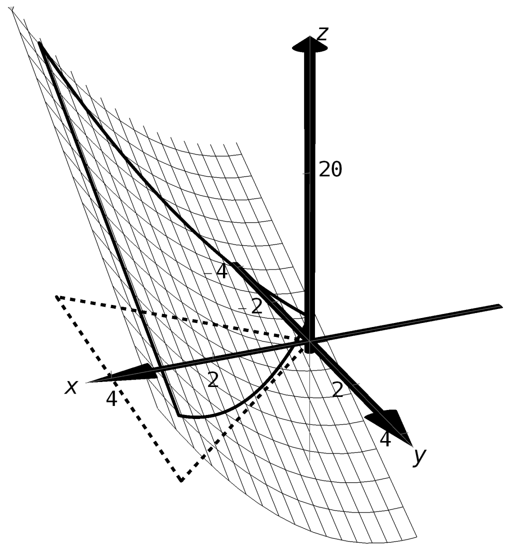


Figure 16.20: Graphing the surface in Example 16.19.

Solution

It is straightforward to compute $f_x(x,y) = 2x$ and $f_y(x,y) = -3$. Thus the surface area is described by the double integral

$$\iint_R \sqrt{1+(2x)^2+(-3)^2}\,dA = \iint_R \sqrt{10+4x^2}\,dA.$$

As with integrals describing arc length, double integrals describing surface area are in general hard to evaluate directly because of the square root. This particular integral can be easily evaluated, though, with judicious choice of our order of integration.

Integrating with order $dx\,dy$ requires us to evaluate $\int \sqrt{10+4x^2}\,dx$. This can be done, though it involves the goniometric substitution $2x = \sqrt{10}\tan(t)$. Integrating with order $dy\,dx$ has as its first integral $\int \sqrt{10+4x^2}\,dy$, which is easy to evaluate: it is simply $y\sqrt{10+4x^2}+C$. So we proceed with the order $dy\,dx$.

$$\begin{aligned}
SA &= \iint_R \sqrt{10+4x^2}\,dA \\
&= \int_0^4 \int_{-x}^{x} \sqrt{10+4x^2}\,dy\,dx \\
&= \int_0^4 \left(y\sqrt{10+4x^2}\right)\Big|_{-x}^{x}\,dx \\
&= \int_0^4 2x\sqrt{10+4x^2}\,dx
\end{aligned}$$

Apply substitution with $u = 10 + 4x^2$:

$$SA = \left(\frac{1}{6}\big(10+4x^2\big)^{3/2}\right)\Bigg|_0^4$$
$$= \frac{1}{3}\big(37\sqrt{74} - 5\sqrt{10}\big) \approx 100.825 \text{ units}^2.$$

So while the region R over which we integrate has an area of 16 units^2, the surface has a much greater area as its z-values change dramatically over R.

In practice, technology helps greatly in the evaluation of such integrals. High powered computer algebra systems can compute integrals that are difficult, or at least time consuming, by hand, and can at least produce very accurate approximations with numerical methods. In general, just knowing how to set up the proper integrals brings one very close to being able to compute the needed value. Most of the work is actually done in just describing the region R in terms of polar or rectangular coordinates. Once this is done, technology can usually provide a good answer.

16.5 Line integrals over a scalar field

This section explores completely different relationships between vectors and integration. These relationships will enable us to compute the work done by a magnetic field in moving an object along a path and find how much air moves through an oddly–shaped screen in space, among other things.

16.5.1 Definition

Consider the surface and curve shown in Figure 16.21(a). The surface is given by

$$f(x,y) = 1 - \cos(x)\sin(y).$$

The dashed curve lies in the xy-plane and is the familiar $y = x^2$ parabola from $-1 \le x \le 1$; we will call this curve C. The curve drawn with a solid line in the graph is the curve in space that lies on our surface with x- and y- values that lie on C.

The question we want to answer is this: what is the area that lies below the curve drawn with the solid line? In other words, what is the area of the region above C and under the the surface f? This region is shown in Figure 16.21(b). We suspect the answer can be found using an integral, but before trying to figure out what that integral is, let us first try to approximate its value.

In Figure 16.21(c), four rectangles have been drawn over the curve C. The bottom corners of each rectangle lie on C, and each rectangle has a height given by the function $f(x,y)$ for some (x,y) pair along C between the rectangle's bottom corners. As we know how to find the area of each rectangle, we are able to approximate the area above C and under f. Clearly, our approximation will be an approximation. The heights of the rectangles do not match exactly with the surface f, nor does the base of each rectangle follow perfectly the path of C.

In typical calculus fashion, our approximation can be improved by using more rectangles. The sum of the areas of these rectangles gives an approximate value of the true area above C and under f. As the area of each rectangle is height × width, we assert that the

$$\text{area above } C \approx \sum(\text{heights} \times \text{widths}).$$

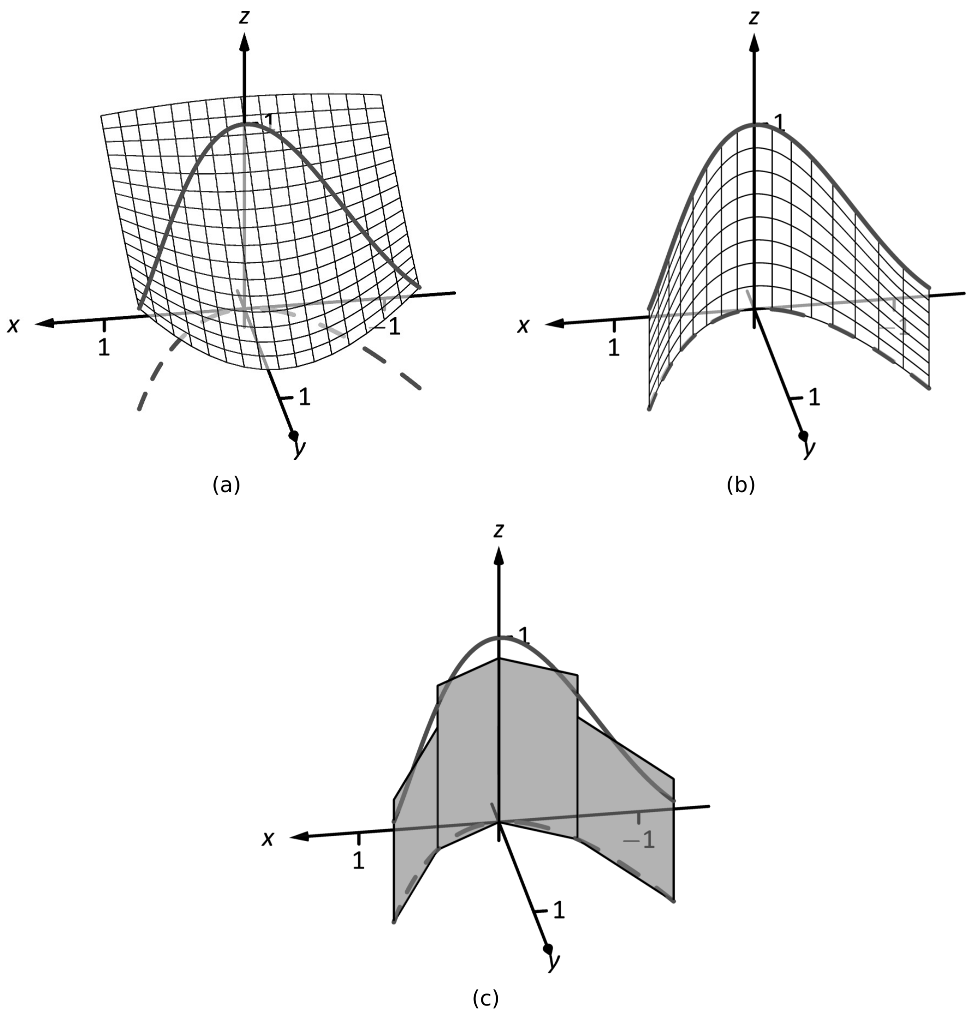


Figure 16.21: Finding area under a curve in space.

When first learning of the integral, and approximating areas with (heights × widths), the width was a small change in x: dx. That will not suffice in this context. Rather, each width of a rectangle is actually approximating the arc length of a small portion of C. In Section 14.4, we used s to represent the arc length parameter of a curve. Hence, a small amount of arc length will thus be represented by ds.

The height of each rectangle will be determined in some way by the surface f. If we parametrize C by s, an s-value corresponds to an (x, y) pair that lies on the parabola C. Since f is a function of x and y, and x and y are functions of s, we can say that f is a function of s. Given a value s, we can compute $f(s)$ and find a height. Thus

$$\begin{aligned}\text{area under } f \text{ and above } C &\approx \sum(\text{heights} \times \text{widths});\\ \text{area under } f \text{ and above } C &= \lim_{\mathcal{L}\to 0}\sum f(c_i)\Delta s_i\\ &= \int_C f(s)\, ds. \qquad (16.7)\end{aligned}$$

Here we have introduce a new notation, the integral symbol with a subscript of C. It is reminiscent of our usage of $\int\int_R$. Using the train of thought found in the Integration Review preceding this section, we interpret $\int_C f(s)\, ds$ as meaning sum up, along a curve C, function values $f(s)$ × small arc lengths. It is understood here that s represents the arc length parameter.

All this leads us to a definition. The integral found in Equation 16.7 is called a **line integral**. We formally define it below.

Definitie 16.6 (Line integral over a scalar field)
Let C be a smooth curve parametrized by s, the arc length parameter, and let f be a continuous function of s. A **line integral** (*lijnintegraal*) is an integral of the form

$$\int_C f(s)\ ds = \lim_{\mathcal{L}\to 0}\sum_{i=1}^{n} f(c_i)\Delta s_i,$$

where $s_1 < s_2 < \cdots < s_n$ is any partition of the s-interval over which C is defined, c_i is any value in the i^{th} subinterval, Δs_i is the width of the i^{th} subinterval, and $\mathcal{L}$ is the length of the longest subinterval in the partition.

Note that Definition 16.6 uses the term scalar field which has not yet been defined. Its meaning is discussed in when it is compared to a vector field. Besides, when C is a closed curve, i.e., a curve that ends at the same point at which it starts, we use

$$\oint_C f(s)\ ds$$

instead of

$$\int_C f(s)\ ds.$$

The definition of the line integral does not specify whether C is a curve in the plane or space (or hyperspace), as the definition holds regardless. For now, however, we will assume C lies in the xy-plane.

Actually, this definition of the line integral does not really say anything new. If C is a curve and s is the arc length parameter of C on $a \le s \le b$, then

$$\int_C f(s)\ ds = \int_a^b f(s)\ ds.$$

The real difference with this integral from the standard $\int_a^b f(x)\ dx$ we used in the past is that of context. Our previous integrals naturally summed up values over an interval on the x-axis, whereas now we are summing up values over a curve. If we can parametrize the curve with the arc length parameter, we can evaluate the line integral just as before. Unfortunately, parametrizing a curve in terms of the arc length parameter is usually very difficult, so we must develop a method of evaluating line integrals using a different parametrization.

Given a curve C, find any parametrization of C: $x = g(t)$ and $y = h(t)$, for continuous functions g and h, where $a \le t \le b$. We can represent this parametrization with a vector–valued function, $\vec{\boldsymbol{r}}(t) = \big(g(t), h(t)\big)$.

In Section 14.4, we defined the arc length parameter as

$$s(t) = \int_0^t \left\|\vec{\boldsymbol{r}}\,'(u)\right\|\ du.$$

By the fundamental theorem of calculus, $ds = \left\|\vec{\boldsymbol{r}}\,'(t)\right\|\ dt$. We can substitute the right hand side of this equation for ds in the line integral definition. Moreover, we can view f as being a function of x and y since it is a function of s. Thus $f(s) = f(x,y) = f\big(g(t), h(t)\big)$. This gives us a concrete way to evaluate a line integral:

$$\int_C f(s)\,ds = \int_a^b f\big(g(t),h(t)\big)\,\|\vec{r}\,'(t)\|\,dt.$$

We restate this as a theorem for its 3–dimensional analogue.

Theorem 16.3 (Evaluating a line integral over a scalar field)
Let C be a curve parametrized by $\vec{r}(t) = \big(g_1(t), g_2(t), g_3(t)\big)$, $a \le t \le b$, where g_i is continuously differentiable, and let $z = f(\mathbf{x})$, where f is continuous over C. Then

$$\int_C f(s)\,ds = \int_a^b f\big(g_1(t), g_2(t), g_3(t)\big)\,\|\vec{r}\,'(t)\|\,dt.$$

To be clear, the first point of Theorem 16.3 can be used to find the area under a surface $z = f(x,y)$ and above a curve C. We will later give an understanding of the line integral when C is a curve in space.

Let us do an example where we actually compute an area.

Example 16.20

Find the area under the surface $f(x,y) = \cos(x) + \sin(y) + 2$ over the curve C, which is the segment of the line $y = 2x + 1$ on $-1 \le x \le 1$, as shown in Figure 16.22.

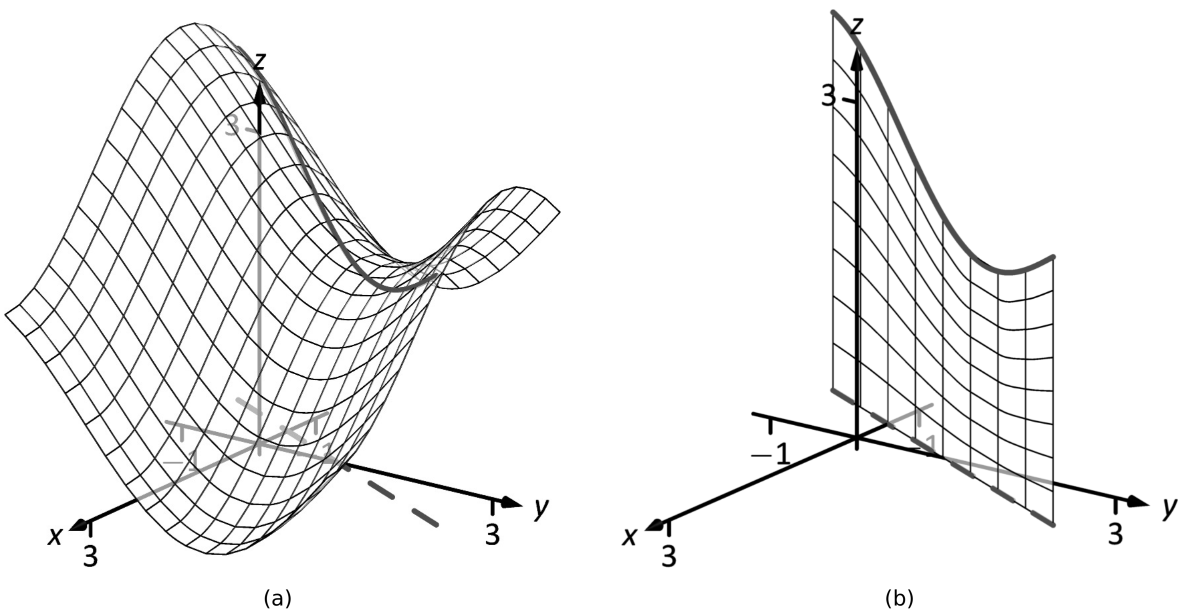


Figure 16.22: Finding area under a curve in Example 16.20.

Solution

Our first step is to represent C with a vector–valued function. Since C is a simple line, and we have a explicit relationship between y and x (namely, that $y = 2x+1$), we can let $x = t$, $y = 2t+1$, and write $\vec{r}(t) = (t, 2t+1)$ for $-1 \le t \le 1$.

We find the values of f over C as

$$f(x,y) = f(t, 2t+1) = \cos(t) + \sin(2t+1) + 2.$$

We also need $\left\|\vec{r}\,'(t)\right\|$; with $\vec{r}\,'(t) = (1,2)$, we have $\left\|\vec{r}\,'(t)\right\| = \sqrt{5}$. Thus $ds = \sqrt{5}\ dt$.

The area we seek is

$$\begin{aligned}\int_C f(s)\ ds &= \int_{-1}^{1} \big(\cos(t) + \sin(2t+1) + 2\big)\sqrt{5}\ dt \\ &= \sqrt{5}\left(\sin(t) - \frac{1}{2}\cos(2t+1) + 2t\right)\Bigg|_{-1}^{1} \\ &\approx 14.418 \text{ units}^2.\end{aligned}$$

We now consider the example that introduced this section.

Example 16.21

Find the area under $f(x,y) = 1 - \cos(x)\sin(y)$ and over the parabola $y = x^2$, from $-1 \le x \le 1$.

Solution

We parametrize our curve C as $\vec{r}(t) = \left(t, t^2\right)$ for $-1 \le t \le 1$; we find $\left\|\vec{r}\,'(t)\right\| = \sqrt{1+4t^2}$, so $ds = \sqrt{1+4t^2}\ dt$.

Replacing x and y with their respective functions of t, we have

$$f(x,y) = f(t, t^2) = 1 - \cos(t)\sin\left(t^2\right).$$

Thus the area under f and over C is found to be

$$\int_C f(s)\ ds = \int_{-1}^{1} \left(1 - \cos(t)\sin\left(t^2\right)\right)\sqrt{1+4t^2}\ dt.$$

This integral is impossible to evaluate using the techniques developed in this text. We resort to a numerical approximation; accurate to two places after the decimal, we find the area is 2.17.

Note how in each of the previous examples we are effectively finding area under a curve, just as we did when first learning of integration. We have used the phrase area over a curve C and under a surface, but that is because of the important role C plays in the integral. The figures show how the curve C defines another curve on the surface $z = f(x,y)$, and we are finding the area under that curve.

16.5.2 Properties

Many properties of line integrals can be inferred from general integration properties. For instance, if k is a scalar, then

$$\int_C kf(s)ds = k\int_C f(s)ds,$$

and similarly

$$\int_C \big(f(s)+g(s)\big)\ ds = \int_C f(s)\ ds + \int_C g(s)\ ds,$$

where f and g are continuous functions of s.

One property in particular of line integrals is worth noting. If C is a curve composed of subcurves C_1 and C_2, where they share only one point in common (see Figure 16.23(a)), then the line integral over C is the sum of the line integrals over C_1 and C_2:

$$\int_C f(s)\ ds = \int_{C_1} f(s)\ ds + \int_{C_2} f(s)\ ds.$$

This property allows us to evaluate line integrals over some curves C that are not smooth. Note how in Figure 16.23(b) the curve is not smooth at D, so by our definition of the line integral we cannot evaluate $\int_C f(s)ds$. However, one can evaluate line integrals over C_1 and C_2 and their sum will be the desired quantity. A curve C that is composed of two or more smooth curves is said to be piecewise smooth. In this section, any statement that is made about smooth curves also holds for piecewise smooth curves.

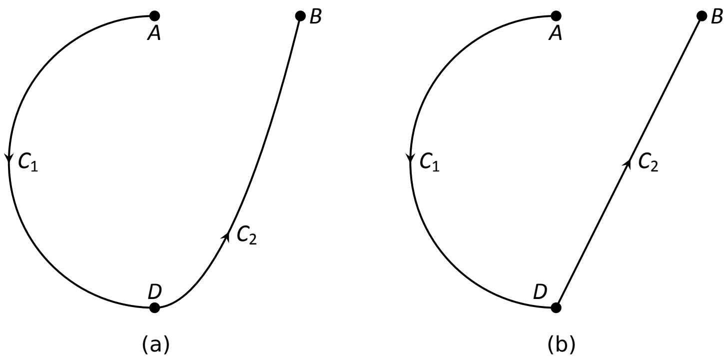


Figure 16.23: Illustrating properties of line integrals.

16.5.3 Centre of mass

Let a curve C (either in the plane or in space) represent a thin wire with variable density $\delta(s)$. We can approximate the mass of the wire by dividing the wire (i.e., the curve) into small segments of length Δs_i and assume the density is constant across these small segments. The mass of each segment is density of the segment × its length; by summing up the approximate mass of each segment we can approximate the total mass:

$$\text{total mass of wire } \approx \sum_i \delta(s_i)\Delta s_i.$$

By taking the limit as the length of the segments approaches 0, we have the definition of the line integral as seen in Definition 16.6. When learning of the line integral, we let $f(s)$ represent a height; now we let $f(s) = \delta(s)$ represent a density. We can extend this understanding of computing mass to also compute the centre of mass of a thin wire. We give the relevant formulas in the next definition, followed by an example.

Definitie 16.7 (Mass and centre of mass of a thin wire)
Let a thin wire lie along a smooth curve C with continuous density function $\delta(s)$, where s is the arc length parameter.

1. The **mass** (*massa*) of the thin wire is $M = \displaystyle\int_C \delta(s)\, ds$.
2. The **moment about the** yz**-plane** is $M_{yz} = \displaystyle\int_C x\delta(s)\, ds$.
3. The **moment about the** xz**-plane** is $M_{xz} = \displaystyle\int_C y\delta(s)\, ds$.
4. The **moment about the** xy**-plane** is $M_{xy} = \displaystyle\int_C z\delta(s)\, ds$.
5. The **centre of mass** (*massamiddelpunt*) of the wire is

$$(\overline{x}, \overline{y}, \overline{z}) = \left(\frac{M_{yz}}{M}, \frac{M_{xz}}{M}, \frac{M_{xy}}{M}\right).$$

Example 16.22

A thin wire follows the path $\vec{r}(t) = \big(1+\cos(t), 1+\sin(t), 1+\sin(2t)\big)$, $0 \le t \le 2\pi$. The density of the wire is determined by its position in space: $\delta(x,y,z) = y+z$ g/cm. The wire is shown in Figure 16.24, where a light colour indicates low density and a dark colour represents high density. Find the mass and centre of mass of the wire.

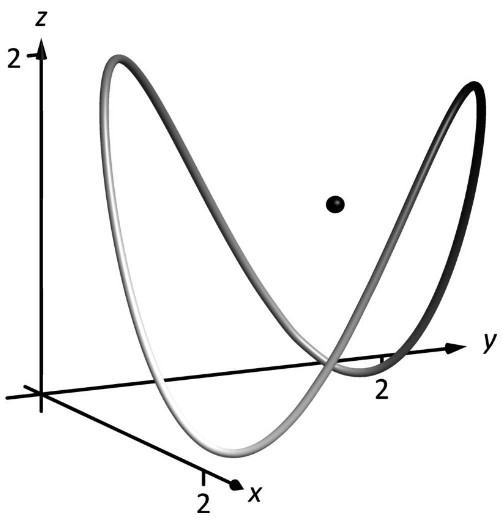


Figure 16.24: Finding the mass of a thin wire in Example 16.22.

Solution

We compute the density of the wire as

$$\delta(x,y,z) = \delta\big(1+\cos(t), 1+\sin(t), 1+\sin(2t)\big) = 2+\sin(t)+\sin(2t).$$

We compute ds as

$$ds = \left\|\vec{r}\,'(t)\right\|\, dt = \sqrt{\sin^2(t)+\cos^2(t)+4\cos^2(2t)}\, dt = \sqrt{1+4\cos^2(2t)}\, dt.$$

Thus the mass is

$$M = \oint_C \delta(s)\, ds = \int_0^{2\pi} \big(2+\sin(t)+\sin(2t)\big)\sqrt{1+4\cos^2(2t)}\; dt \approx 21.08 \text{ g} .$$

We compute the moments about the coordinate planes:

$$M_{yz} = \oint_C x\delta(s)\, ds = \int_0^{2\pi} \big(1+\cos(t)\big)\big(2+\sin(t)+\sin(2t)\big)\sqrt{1+4\cos^2(2t)}\; dt \approx 21.08\,,$$

$$M_{xz} = \oint_C y\delta(s)\, ds = \int_0^{2\pi} \big(1+\sin(t)\big)\big(2+\sin(t)+\sin(2t)\big)\sqrt{1+4\cos^2(2t)}\; dt \approx 26.35\,,$$

$$M_{xy} = \oint_C z\delta(s)\, ds = \int_0^{2\pi} \big(1+\sin(2t)\big)\big(2+\sin(t)+\sin(2t)\big)\sqrt{1+4\cos^2(2t)}\; dt \approx 25.40.$$

Thus the center of mass of the wire is located at

$$(\overline{x},\overline{y},\overline{z}) = \left(\frac{M_{yz}}{M}, \frac{M_{xz}}{M}, \frac{M_{xy}}{M}\right) \approx (1, 1.25, 1.20),$$

as indicated by the dot in Figure 16.24. Note how in this example, the curve C is "centered" about the point $(1,1,1)$, though the variable density of the wire pulls the center of mass out along the y- and z-axes.

In the following, we investigate a new mathematical object, the vector field, after which we increase our understanding of integration in the context of vector fields.

16.6 Vector fields

16.6.1 Definition

We have studied functions, where the input of such functions is a point and the output is a number. We could also create functions where the input is a point, but the output is a vector. For instance, we could create the following function: $\vec{\boldsymbol{F}}(x,y) = (x+y, x-y)$, where $\vec{\boldsymbol{F}}(2,3) = (5,-1)$. We are to think of $\vec{\boldsymbol{F}}$ assigning the vector $(5,-1)$ to the point $(2,3)$; in some sense, the vector $(5,-1)$ lies at the point $(2,3)$.

Such functions are extremely useful in any context where magnitude and direction are important. For instance, we could create a function $\vec{\boldsymbol{F}}$ that represents the electromagnetic force exerted at a point by a electromagnetic field, or the velocity of air as it moves across an airfoil.

Because these functions are so important, we need to formally define them.

Definitie 16.8 (Vector field)

1. A **vector field in the plane** (*vectorveld in het vlak*) is a function $\vec{\boldsymbol{F}}(x,y)$ whose domain is a subset of $\mathbb{R}^2$ and whose output is a two–dimensional vector:

$$\vec{\boldsymbol{F}}(x,y) = \big(M(x,y), N(x,y)\big).$$

2. A **vector field in** 3**-dimensional space** (*vectorveld in de* 3*-dimensionale ruimte*) is a function $\vec{\boldsymbol{F}}(\mathbf{x})$ whose domain is a subset of $\mathbb{R}^3$ and whose output is a 3–dimensional vector:

$$\vec{\boldsymbol{F}}(\mathbf{x}) = \big(M_1(\mathbf{x}), M_2(\mathbf{x}), M_3(\mathbf{x})\big).$$

This definition may seem odd at first, as a special type of function is called a field. However, as the function determines a field of vectors, we can say the field is defined by the function, and thus the field is a function.

When graphing a vector field in the plane, the general idea is to draw the vector $\vec{\boldsymbol{F}}(x,y)$ at the point (x,y). For instance, using $\vec{\boldsymbol{F}}(x,y) = (x+y, x-y)$ as before, at $(1,1)$ we would draw $(2,0)$.

In Figure 16.25(a), one can see that the vector $(2,0)$ is drawn starting from the point $(1,1)$. A total of 8 vectors are drawn, with the x- and y-values of $-1,0,1$. In many ways, the resulting graph is a mess. In Figure 16.25(b), the same field is redrawn with each vector $\vec{\boldsymbol{F}}(x,y)$ drawn centred on the point (x,y). This makes for a better looking image, though when one vector intersects another, the image looks cluttered. A common way to address this problem is limit the length of each arrow, and represent long vectors with thick arrows, as done in Figure 16.25(c). Usually we do not use a graph of a vector field to determine exactly the magnitude of a particular vector. Rather, we are more concerned with the relative magnitudes of vectors: which are bigger than others? Thus limiting the length of the vectors is not problematic. Mathematica obviously allows us to plot many vectors in a vector field nicely; in Figure 16.25(d), we see the same vector field drawn with using the Mathematica command `VectorPlot`, and finally get a clear picture of how this vector field behaves. If this vector field represented the velocity of air moving across a flat surface, we could see that the air tends to move either to the upper–right or lower–left, and moves very slowly near the origin. We can similarly plot vector fields in space, though the plots get very busy very quickly.

16.6.2 The del operator

Often, we will drop the of x, y and z portions of the notation in Definition 16.8 and refer to vector fields in the plane and in space as

$$\vec{\boldsymbol{F}} = (M,N) \quad \text{and} \quad \vec{\boldsymbol{F}} = (M,N,P),$$

respectively, as this shorthand is quite convenient.

Another item of notation will become useful: the del operator. Recall in Section 15.6 how we used the symbol $\vec{\nabla}$ to represent the gradient of a function of two variables. We now define $\vec{\nabla}$ to be the del operator. It is a vector whose components are partial derivative operations.

In the plane,

$$\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right);$$

in space,

$$\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right).$$

With this definition of $\vec{\nabla}$, we can better understand the gradient $\vec{\nabla} f$. As f returns a scalar, the properties of scalar and vector multiplication gives

$$\vec{\nabla} f = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) f = \left(\frac{\partial}{\partial x} f, \frac{\partial}{\partial y} f\right) = \big(f_x, f_y\big).$$

Now apply the del operator $\vec{\nabla}$ to vector fields. Let $\vec{\boldsymbol{F}} = \big(x+\sin(y), y^2+z, x^2\big)$. We can use vector

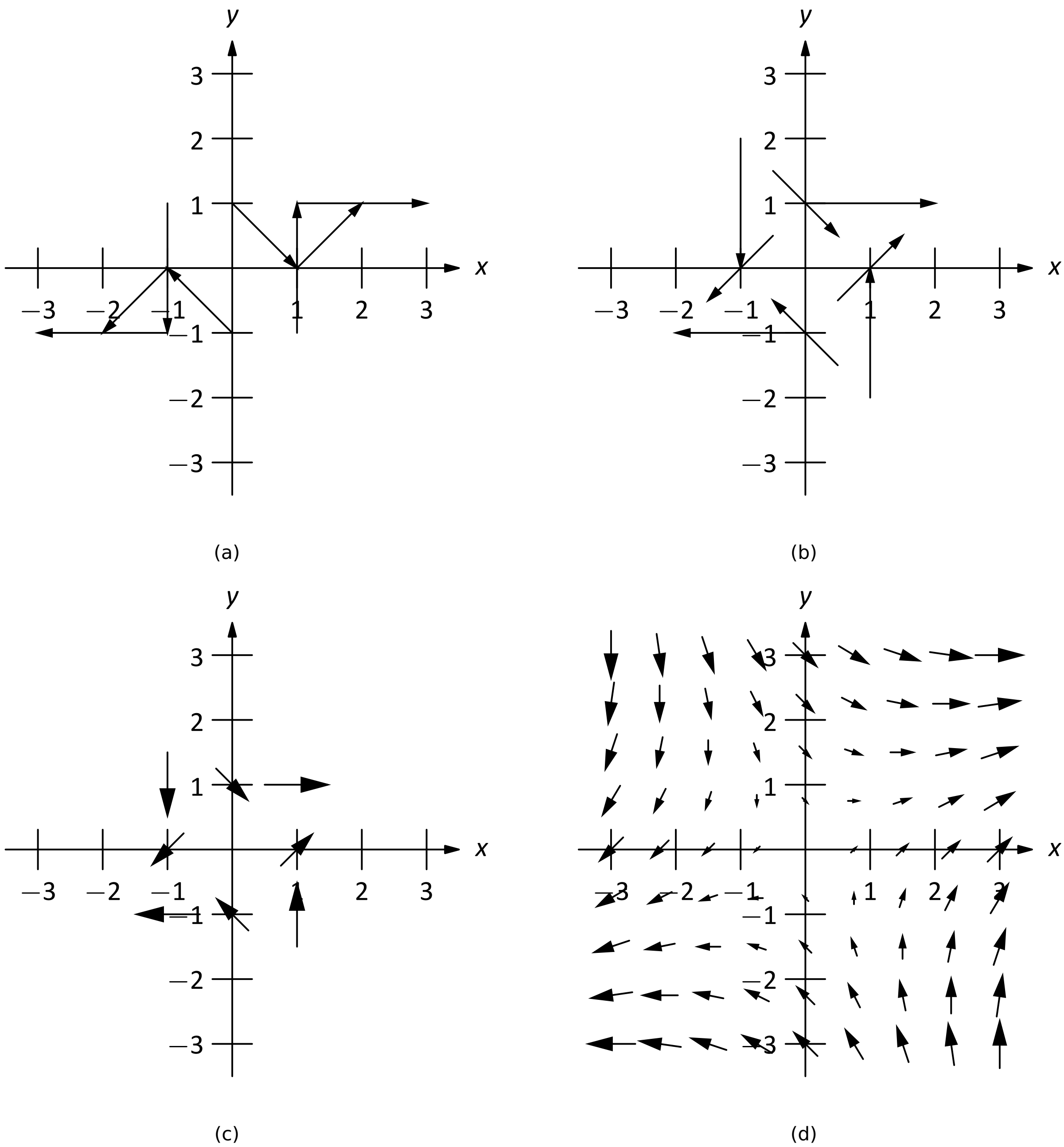


Figure 16.25: Demonstrating methods of graphing vector fields.

operations and find the dot product of $\vec{\nabla}$ and $\vec{F}$:

$$\begin{aligned}\vec{\nabla}\cdot\vec{F} &= \left(\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z}\right)\cdot\left(x+\sin(y),y^2+z,x^2\right)\\ &= \frac{\partial}{\partial x}\left(x+\sin(y)\right)+\frac{\partial}{\partial y}\left(y^2+z\right)+\frac{\partial}{\partial z}\left(x^2\right)\\ &= 1+2y.\end{aligned}$$

We can also compute their cross products:

$$\begin{aligned}\vec{\nabla}\times\vec{F} &= \left(\frac{\partial}{\partial y}\left(x^2\right)-\frac{\partial}{\partial z}\left(y^2+z\right),\frac{\partial}{\partial z}\left(x+\sin(y)\right)-\frac{\partial}{\partial x}\left(x^2\right),\frac{\partial}{\partial x}\left(y^2+z\right)-\frac{\partial}{\partial y}\left(x+\sin(y)\right)\right)\\ &= \left(-1,-2x,-\cos(y)\right).\end{aligned}$$

As we next learn about properties of vector fields, we will see how these dot and cross products with the del operator are quite useful.

16.6.3 Divergence and curl

Two properties of vector fields will prove themselves to be very important: divergence and curl. Each is a special "derivative" of a vector field; that is, each measures an instantaneous rate of change of a vector field.

If the vector field represents the velocity of a fluid or gas, then the divergence of the field is a measure of the compressibility of the fluid. If the divergence is negative at a point, it means that the fluid is compressing: more fluid is going into the point than is going out. If the divergence is positive, it means the fluid is expanding: more fluid is going out at that point than going in. A divergence of zero means the same amount of fluid is going in as is going out. If the divergence is zero at all points, we say the field is incompressible.

It turns out that the proper measure of divergence is simply $\vec{\nabla}\cdot\vec{\boldsymbol{F}}$, as stated in the following definition.

Definitie 16.9 (Divergence)
The **divergence of a vector field $\vec{\boldsymbol{F}}$** (*divergentie van een vectorveld*) is

$$\operatorname{div}\vec{\boldsymbol{F}} = \vec{\nabla}\cdot\vec{\boldsymbol{F}}.$$

- In the plane, with $\vec{\boldsymbol{F}} = (M,N)$: $\operatorname{div}\vec{\boldsymbol{F}} = M_x + N_y$.
- In space, with $\vec{\boldsymbol{F}} = (M,N,P)$: $\operatorname{div}\vec{\boldsymbol{F}} = M_x + N_y + P_z$.

Curl is a measure of the spinning action of the field. Let $\vec{\boldsymbol{F}}$ represent the flow of water over a flat surface. If a small round cork were held in place at a point in the water, would the water cause the cork to spin? No spin corresponds to zero curl; counterclockwise spin corresponds to positive curl and clockwise spin corresponds to negative curl.

In space, things are a bit more complicated. Again let $\vec{\boldsymbol{F}}$ represent the flow of water, and imagine suspending a tennis ball in one location in this flow. The water may cause the ball to spin along an axis. If so, the curl of the vector field is a vector (not a scalar, as before), parallel to the axis of rotation, following a right hand rule: when the thumb of one's right hand points in the direction of the curl, the ball will spin in the direction of the curling fingers of the hand.

In space, it turns out the proper measure of curl is $\vec{\nabla}\times\vec{\boldsymbol{F}}$, as stated in the following definition. To find the curl of a planar vector field $\vec{\boldsymbol{F}} = (M,N)$, embed it into space as $\vec{\boldsymbol{F}} = (M,N,0)$ and apply the cross product definition. Since M and N are functions of just x and y (and not z), all partial derivatives with respect to z become 0 and the result is simply $\left(0,0,N_x - M_y\right)$. The third component is the measure of curl of a planar vector field.

Definitie 16.10 (Curl)

- Let $\vec{\boldsymbol{F}} = (M,N)$ be a vector field in the plane. The **curl of $\vec{\boldsymbol{F}}$** (*rotatie of rotor van $\vec{\boldsymbol{F}}$*) is

$$\operatorname{curl}\vec{\boldsymbol{F}} = N_x - M_y.$$

- Let $\vec{F} = (M, N, P)$ be a vector field in space. The **curl of $\vec{F}$** (*rotatie of rotor van* $\vec{F}$) is

$$\text{curl}\,\vec{F} = \vec{\nabla} \times \vec{F} = \left(P_y - N_z,\, M_z - P_x,\, N_x - M_y\right).$$

We adopt the convention of referring to curl as $\vec{\nabla} \times \vec{F}$, regardless of whether $\vec{F}$ is a vector field in two or three dimensions.

We now practice computing these quantities.

Example 16.23

For each of the planar vector fields given below, view its graph and try to visually determine if its divergence and curl are 0. Then compute the divergence and curl.

1. $\vec{F} = (y, 0)$ (see Figure 16.26(a))
2. $\vec{F} = (-y, x)$ (see Figure 16.26(b))
3. $\vec{F} = (x, y)$ (see Figure 16.26(c))
4. $\vec{F} = \left(\cos(y), \sin(x)\right)$ (see Figure 16.26(d))

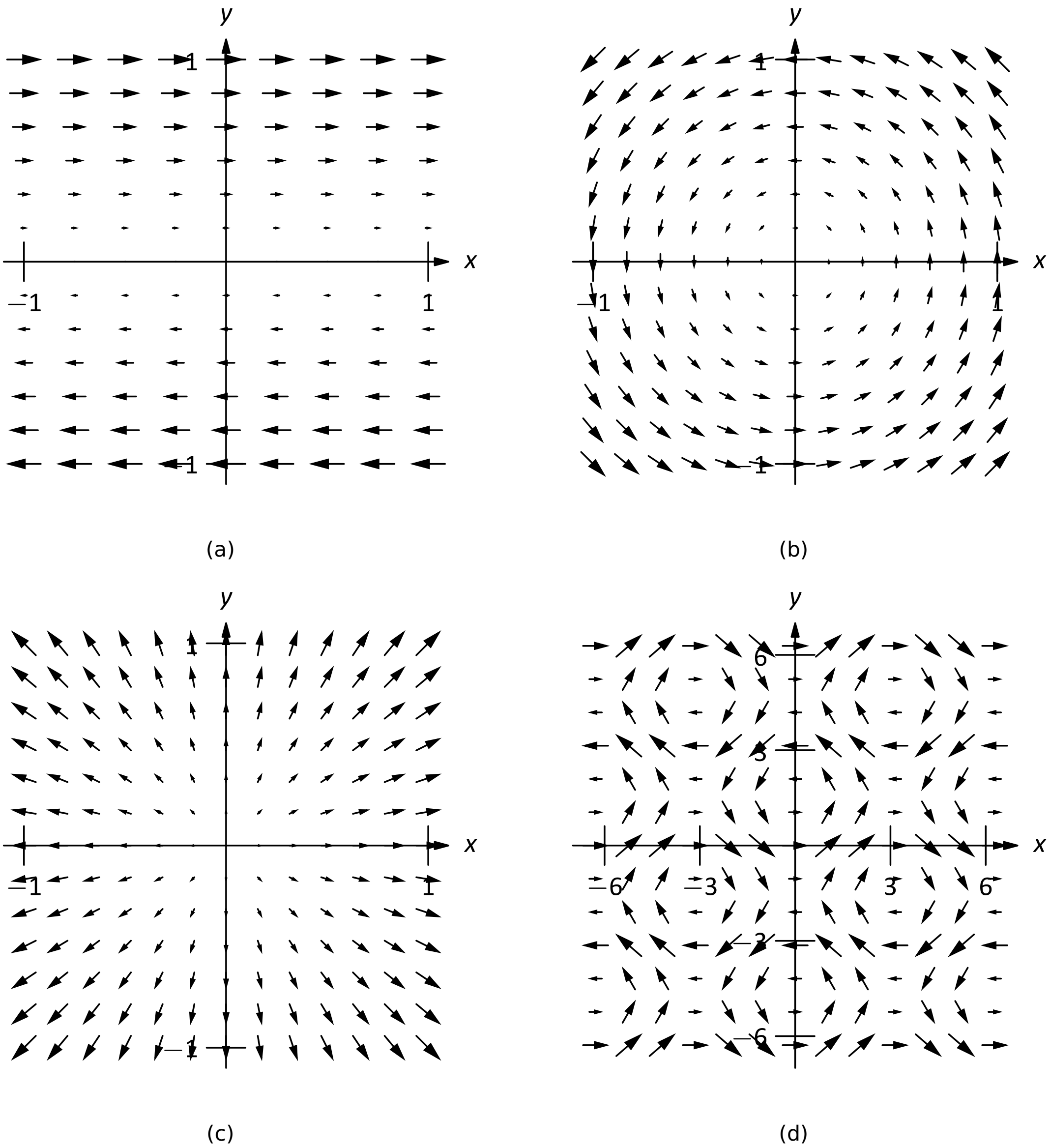


Figure 16.26: The vector fields in parts 1 (a), 2 (b), 3 (c) and 4 (d) in Example 16.23.

Solution

1. The arrow sizes are constant along any horizontal line, so if one were to draw a small box anywhere on the graph, it would seem that the same amount of fluid would enter the box as exit. Therefore it seems the divergence is zero; it is, as

$$\operatorname{div}\vec{\boldsymbol{F}} = \vec{\nabla}\cdot\vec{\boldsymbol{F}} = M_x + N_y = \frac{\partial}{\partial x}(y) + \frac{\partial}{\partial y}(0) = 0.$$

At any point on the x-axis, arrows above it move to the right and arrows below it move to the left, indicating that a cork placed on the axis would spin clockwise. A cork placed anywhere above the x-axis would have water above it moving to the right faster than the water below it, also creating a clockwise spin. A clockwise spin also appears to be created at points below the x-axis. Thus it seems the curl should be negative (and not zero). Indeed, it is:

$$\operatorname{curl}\vec{\boldsymbol{F}} = \vec{\nabla}\times\vec{\boldsymbol{F}} = N_x - M_y = \frac{\partial}{\partial x}(0) - \frac{\partial}{\partial y}(y) = -1.$$

2. It appears that all vectors that lie on a circle of radius r, centered at the origin, have the same length (and indeed this is true). That implies that the divergence should be zero: draw any box on the graph, and any fluid coming in will lie along a circle that takes the same amount of fluid out. Indeed, the divergence is zero, as

$$\operatorname{div}\vec{\boldsymbol{F}} = \vec{\nabla}\cdot\vec{\boldsymbol{F}} = M_x + N_y = \frac{\partial}{\partial x}(-y) + \frac{\partial}{\partial y}(x) = 0.$$

Clearly this field moves objects in a circle, but would it induce a cork to spin? It appears that yes, it would: place a cork anywhere in the flow, and the point of the cork closest to the origin would feel less flow than the point on the cork farthest from the origin, which would induce a counterclockwise flow. Indeed, the curl is positive:

$$\operatorname{curl}\vec{\boldsymbol{F}} = \vec{\nabla}\times\vec{\boldsymbol{F}} = N_x - M_y = \frac{\partial}{\partial x}(x) - \frac{\partial}{\partial y}(-y) = 1-(-1) = 2.$$

Since the curl is constant, we conclude the induced spin is the same no matter where one is in this field.

3. At the origin, there are many arrows pointing out but no arrows pointing in. We conclude that at the origin, the divergence must be positive (and not zero). If one were to draw a box anywhere in the field, the edges farther from the origin would have larger arrows passing through them than the edges close to the origin, indicating that more is going from a point than going in. This indicates a positive (and not zero) divergence. This is correct:

$$\operatorname{div}\vec{\boldsymbol{F}} = \vec{\nabla}\cdot\vec{\boldsymbol{F}} = M_x + N_y = \frac{\partial}{\partial x}(x) + \frac{\partial}{\partial y}(y) = 1+1 = 2.$$

One may find this curl to be harder to determine visually than previous examples. One might note that any arrow that induces a clockwise spin on a cork will have an equally sized arrow inducing a counterclockwise spin on the other side, indicating no spin and no curl. This is correct, as

$$\operatorname{curl}\vec{\boldsymbol{F}} = \vec{\nabla}\times\vec{\boldsymbol{F}} = N_x - M_y = \frac{\partial}{\partial x}(y) - \frac{\partial}{\partial y}(x) = 0.$$

4. One might find this divergence hard to determine visually as large arrows appear in close proximity to small arrows, each pointing in different directions. Instead of trying to rationalize a guess, we compute the divergence:

$$\operatorname{div}\vec{F} = \vec{\nabla}\cdot\vec{F} = M_x + N_y = \frac{\partial}{\partial x}\big(\cos(y)\big) + \frac{\partial}{\partial y}\big(\sin(x)\big) = 0.$$

Perhaps surprisingly, the divergence is 0.

With all the loops of different directions in the field, one is apt to reason the curl is variable. Indeed, it is:

$$\operatorname{curl}\vec{F} = \vec{\nabla}\times\vec{F} = N_x - M_y = \frac{\partial}{\partial x}\big(\sin(x)\big) - \frac{\partial}{\partial y}\big(\cos(y)\big) = \cos(x) + \sin(y).$$

Depending on the values of x and y, the curl may be positive, negative, or zero.

Example 16.24

The force of gravity between two objects is inversely proportional to the square of the distance between the objects. Locate a point mass at the origin. Create a vector field $\vec{F}$ that represents the gravitational pull of the point mass at any point (x, y, z). Find the divergence and curl of this field.

Solution

The point mass pulls toward the origin, so at (x, y, z), the force will pull in the direction of $(-x, -y, -z)$. To get the proper magnitude, it will be useful to find the unit vector in this direction. Dividing by its magnitude, we have

$$\vec{u} = \left(\frac{-x}{\sqrt{x^2+y^2+z^2}}, \frac{-y}{\sqrt{x^2+y^2+z^2}}, \frac{-z}{\sqrt{x^2+y^2+z^2}}\right).$$

The magnitude of the force is inversely proportional to the square of the distance between the two points. Letting k be the constant of proportionality, we have the magnitude as

$$\frac{k}{x^2+y^2+z^2}.$$

Multiplying this magnitude by the unit vector above, we have the desired vector field:

$$\vec{F} = \left(\frac{-kx}{(x^2+y^2+z^2)^{3/2}}, \frac{-ky}{(x^2+y^2+z^2)^{3/2}}, \frac{-kz}{(x^2+y^2+z^2)^{3/2}}\right).$$

We leave it to the reader to confirm that $\operatorname{div}\vec{F} = 0$ and $\operatorname{curl}\vec{F} = \vec{0}$.

The analogous planar vector field is given in Figure 16.27. Note how all arrows point to the origin, and the magnitude gets very small when far from the origin.

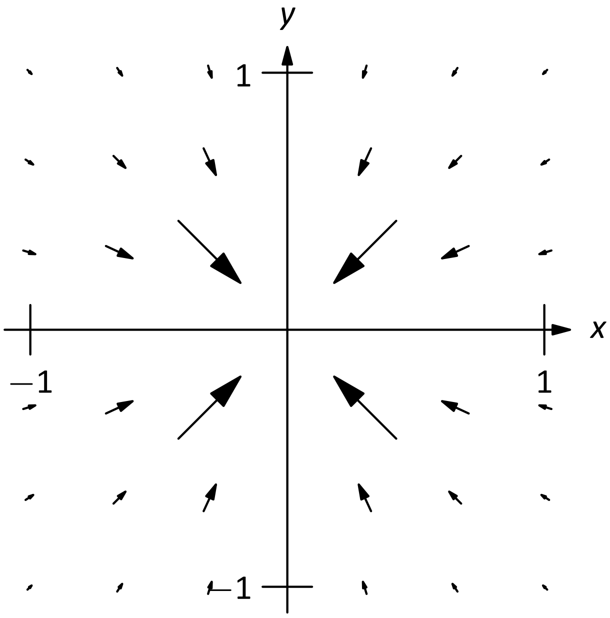


Figure 16.27: A vector field representing a planar gravitational force.

A function $z = f(x,y)$ naturally induces a vector field, $\vec{\boldsymbol{F}} = \vec{\nabla} f = \left(f_x, f_y\right)$. Given what we learned of the gradient in Section 15.6, we know that the vectors of $\vec{\boldsymbol{F}}$ point in the direction of greatest increase of f. Because of this, f is said to be the **potential function of $\vec{\boldsymbol{F}}$**. Vector fields that are the gradient of potential functions will play an important role in the remainder of this section.

The last part of this section applies calculus to vector fields. A common application is this: let $\vec{\boldsymbol{F}}$ be a vector field representing a force (hence it is called a force field) and let a particle move along a curve C under the influence of this force. What work is performed by the field on this particle? The solution lies in correctly applying the concepts of line integrals in the context of vector fields.

16.7 Line Integrals over vector fields

16.7.1 Definition

Suppose a particle moves along a curve C under the influence of an electromagnetic force described by a vector field $\vec{\boldsymbol{F}}$. Since a force is inducing motion, work is performed. How can we calculate how much work is performed?

Recall that when moving in a straight line, if $\vec{\boldsymbol{F}}$ represents a constant force and $\vec{\boldsymbol{d}}$ represents the direction and length of travel, then work is simply $W = \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{d}}$. However, we generally want to be able to calculate work even if $\vec{\boldsymbol{F}}$ is not constant and C is not a straight line.

As we have practised many times before, we can calculate work by first approximating, then refining our approximation through a limit that leads to integration.

Assume as we did at the beginning of this section, C can be parametrized by the arc length parameter s. Over a short piece of the curve with length ds, the curve is approximately straight and our force is approximately constant. The straight–line direction of this short length of curve is given by $\vec{\boldsymbol{T}}$, the unit tangent vector; let $\vec{\boldsymbol{d}} = \vec{\boldsymbol{T}}\,ds$, which gives the direction and magnitude of a small section of C. Thus work over this small section of C is $\vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{d}} = \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{T}}\,ds$.

Summing up all the work over these small segments gives an approximation of the work performed. By taking the limit as ds goes to zero, and hence the number of segments approaches infinity, we can obtain the exact amount of work. Hence, we see that

$$W = \int_C \vec{F} \cdot \vec{T}\, ds,$$

is a line integral.

This line integral is beautiful in its simplicity, yet is not so useful in making actual computations (largely because the arc length parameter is so difficult to work with). To compute actual work, we need to parametrize C with another parameter t via a vector–valued function $\vec{r}(t)$. Since $ds = \left\|\vec{r}\,'(t)\right\|\, dt$ and $\vec{T} = \vec{r}\,'(t)/\left\|\vec{r}\,'(t)\right\|$, we get

$$W = \int_C \vec{F} \cdot \vec{T}\; ds = \int_C \vec{F} \cdot \frac{\vec{r}\,'(t)}{\left\|\vec{r}\,'(t)\right\|}\; \left\|\vec{r}\,'(t)\right\|\; dt = \int_C \vec{F} \cdot \vec{r}\,'(t)\; dt = \int_C \vec{F} \cdot d\vec{r}, \tag{16.8}$$

where the final integral uses the differential $d\vec{r}$ for $\vec{r}\,'(t)\, dt$.

These integrals are known as line integrals over vector fields. By contrast, the line integrals we dealt with earlier are sometimes referred to as line integrals over scalar fields. Just as a vector field is defined by a function that returns a vector, a scalar field is a function that returns a scalar, such as $z = f(x, y)$.

We formally define this line integral, then give examples and applications.

Definitie 16.11 (Line integral over a vector field)
Let $\vec{F}$ be a vector field with continuous components defined on a smooth curve C, parametrized by $\vec{r}(t)$, and let $\vec{T}$ be the unit tangent vector of $\vec{r}(t)$. The **line integral over $\vec{F}$** along C is

$$\int_C \vec{F} \cdot d\vec{r} = \int_C \vec{F} \cdot \vec{T}\; ds.$$

In Definition 16.11, note how the dot product $\vec{F} \cdot \vec{T}$ is just a scalar. Therefore, this new line integral is really just a special kind of line integral. Indeed, letting $f(s) = \vec{F}(s) \cdot \vec{T}(s)$, the right–hand side simply becomes $\int_C f(s)\; ds$, and we can use the corresponding techniques to evaluate the integral. This is summarized in the following theorem.

Theorem 16.4 (Evaluating a line integral over a vector field)
Let $\vec{F}$ be a vector field with continuous components defined on a smooth curve C, parametrized by $\vec{r}(t)$, $a \le t \le b$, where $\vec{r}$ is continuously differentiable. Then

$$\int_C \vec{F} \cdot \vec{T}\; ds = \int_C \vec{F} \cdot d\vec{r} = \int_a^b \vec{F}\big(\vec{r}(t)\big) \cdot \vec{r}\,'(t)\; dt.$$

This theorem indicates that we can use any continuously differentiable parametrization $\vec{r}(t)$ of C that preserves the orientation of C: there isn't a right one. In practice, choose one that seems easy to work with.

Note that the above definition and theorem implicitly evaluate $\vec{F}$ along the curve C, which is parametrized by $\vec{r}(t)$. For instance, if $\vec{F} = (x+y, x-y)$ and $\vec{r}(t) = \big(t^2, \cos(t)\big)$, then evaluating $\vec{F}$ along C means substituting the x- and y-components of $\vec{r}(t)$ in for x and y, respectively, in $\vec{F}$. Therefore, along C, $\vec{F} = (x+y, x-y) = \big(t^2 + \cos(t), t^2 - \cos(t)\big)$. Since we are substituting the output of $\vec{r}(t)$ for the input of $\vec{F}$, we write this as $\vec{F}\big(\vec{r}(t)\big)$. This is a slight abuse of notation as technically the input of $\vec{F}$ is to be a point, not a vector, but this shorthand is useful.

Example 16.25

Two particles move from $(0,0)$ to $(1,1)$ under the influence of the force field $\vec{F} = \langle x, x+y \rangle$. One particle follows C_1, the line $y = x$; the other follows C_2, the curve $y = x^4$, as shown in Figure 16.28. Force is measured in Newtons and distance is measured in meters. Find the work performed by each particle.

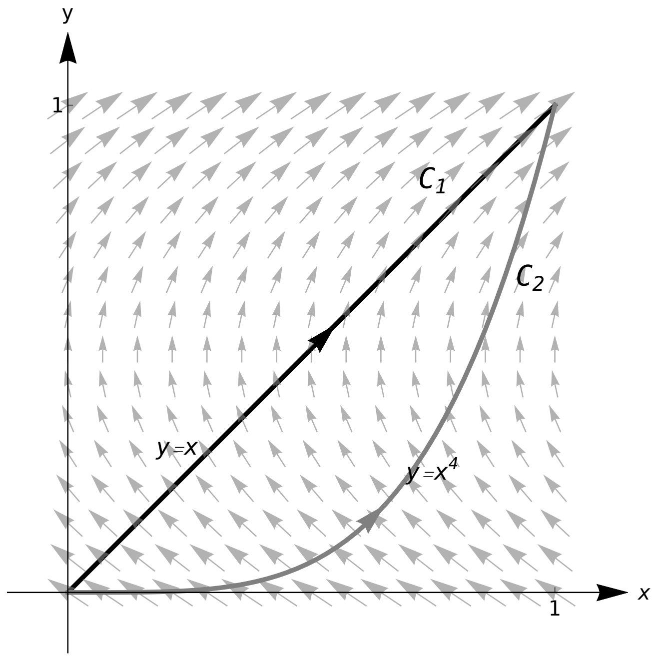


Figure 16.28: Paths through a vector field in Example 16.25.

Solution

To compute work, we need to parametrize each path. We use $\vec{r}_1(t) = \langle t, t \rangle$ to parametrize $y = x$, and let $\vec{r}_2(t) = \langle t, t^4 \rangle$ parametrize $y = x^4$; for each, $0 \le t \le 1$.

Along the straight–line path, $\vec{F}\big(\vec{r}_1(t)\big) = \langle x, x+y \rangle = \langle t, t+t \rangle = \langle t, 2t \rangle$. We find $\vec{r}\,'_1(t) = \langle 1, 1 \rangle$. The integral that computes work is:

$$\begin{aligned}
\int_{C_1} \vec{F} \cdot d\vec{r} &= \int_0^1 \langle t, 2t \rangle \cdot \langle 1, 1 \rangle \ dt \\
&= \int_0^1 3t \ dt \\
&= \left(\frac{3}{2}t^2\right)\Bigg|_0^1 = 1.5 \text{ joules.}
\end{aligned}$$

Along the curve $y = x^4$, we have that

$$\vec{F}\big(\vec{r}_2(t)\big) = \langle x, x+y \rangle = \left\langle t, t+t^4 \right\rangle.$$

We find $\vec{r}\,'_2(t) = \left(1, 4t^3\right)$. The work performed along this path is

$$\begin{aligned}\int_{C_2} \vec{F}\cdot d\vec{r} &= \int_0^1 \left(t, t+t^4\right)\cdot\left(1, 4t^3\right)\, dt\\ &= \int_0^1 \left(t+4t^4+4t^7\right) dt\\ &= \left(\frac{1}{2}t^2+\frac{4}{5}t^5+\frac{1}{2}t^8\right)\Bigg|_0^1 = 1.8 \text{ joules.}\end{aligned}$$

Note how differing amounts of work are performed along the different paths. This should not be too surprising: the force is variable, one path is longer than the other, etc.

Example 16.26

Two particles move from $(-1, 1)$ to $(1, 1)$ under the influence of a force field $\vec{F} = (y, x)$. One moves along the curve C_1, the parabola defined by $y = 2x^2 - 1$. The other particle moves along the curve C_2, the bottom half of the circle defined by $x^2 + (y-1)^2 = 1$, as shown in Figure 16.29. Force is measured in Newton and distances are measured in meters. Find the work performed by moving each particle along its path.

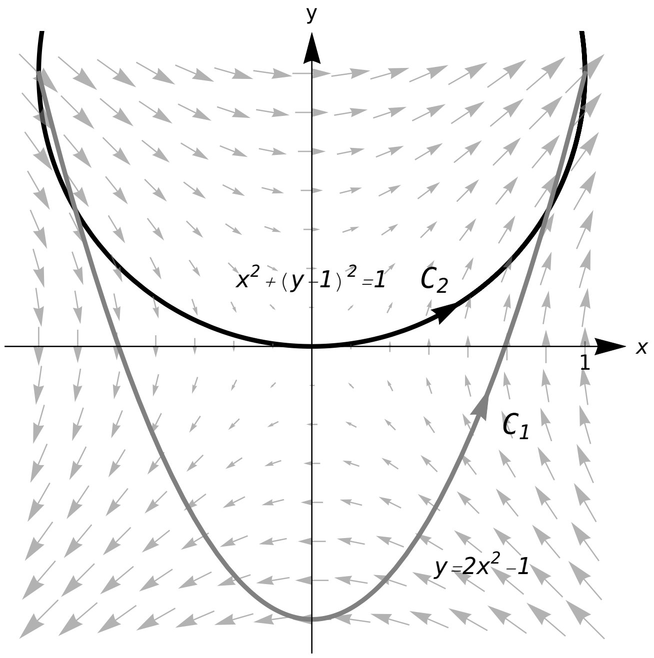


Figure 16.29: Paths through a vector field in Example 16.26.

Solution

We start by parametrizing C_1: the parametrization $\vec{r}_1(t) = \left(t, 2t^2-1\right)$ is straightforward, giving $\vec{r}\,'_1 = (1, 4t)$. On C_1, $\vec{F}\big(\vec{r}_1(t)\big) = (y, x) = \left(2t^2-1, t\right)$.

Computing the work along C_1, we have:

$$\begin{aligned}\int_{C_1} \vec{F}\cdot d\vec{r}_1 &= \int_{-1}^{1} \left(2t^2-1,t\right)\cdot(1,4t)\ dt\\ &= \int_{-1}^{1} \left(2t^2-1+4t^2\right)dt = 2 \text{ joules.}\end{aligned}$$

For C_2, it is probably simplest to parametrize the half circle using sine and cosine. Recall that $\vec{r}(t) = \big(\cos(t),\sin(t)\big)$ is a parametrization of the unit circle on $0 \le t \le 2\pi$; we add 1 to the second component to shift the circle up one unit, then restrict the domain to $\pi \le t \le 2\pi$ to obtain only the lower half, giving $\vec{r}_2(t) = \big(\cos(t),\sin(t)+1\big)$, $\pi \le t \le 2\pi$, and hence $\vec{r}\,'_2(t) = \big(-\sin(t),\cos(t)\big)$ and $\vec{F}\big(\vec{r}_2(t)\big) = (y,x) = \big(\sin(t)+1,\cos(t)\big)$.

Computing the work along C_2, we have:

$$\begin{aligned}\int_{C_2} \vec{F}\cdot d\vec{r}_2 &= \int_{\pi}^{2\pi} \big(\sin(t)+1,\cos(t)\big)\cdot\big(-\sin(t),\cos(t)\big)\ dt\\ &= \int_{\pi}^{2\pi} \big(-\sin^2(t)-\sin(t)+\cos^2(t)\big)\,dt = 2 \text{ joules.}\end{aligned}$$

Note how the work along C_1 and C_2 in this example is the same.

16.7.2 Properties

Line integrals over vector fields share the same properties as line integrals over scalar fields, with one important distinction. The orientation of the curve C matters with line integrals over vector fields, whereas it did not matter with line integrals over scalar fields.

It is relatively easy to see why. Let C be the unit circle. The area under a surface over C is the same whether we traverse the circle in a clockwise or counterclockwise fashion, hence the line integral over a scalar field on C is the same irrespective of orientation. On the other hand, if we are computing work done by a force field, direction of travel definitely matters. Opposite directions create opposite signs when computing dot products, so traversing the circle in opposite directions will create line integrals that differ by a factor of -1.

In summary, we have the following properties of line integrals over vector fields.

1. Let $\vec{F}$ and $\vec{G}$ be vector fields with continuous components defined on a smooth curve C, parametrized by $\vec{r}(t)$, and let k_1 and k_2 be scalars. Then
$$\int_C \big(k_1\vec{F}+k_2\vec{G}\big)\cdot d\vec{r} = k_1\int_C \vec{F}\cdot d\vec{r} + k_2\int_C \vec{G}\cdot d\vec{r}.$$
2. Let C be piecewise smooth, composed of smooth components C_1 and C_2. Then
$$\int_C \vec{F}\cdot d\vec{r} = \int_{C_1}\vec{F}\cdot d\vec{r} + \int_{C_2}\vec{F}\cdot d\vec{r}.$$

3. Let C^* be the curve C with opposite orientation, parametrized by $\vec{r}^*$. Then

$$\int_C \vec{F}\cdot d\vec{r} = -\int_{C^*} \vec{F}\cdot d\vec{r}^*.$$

We demonstrate using these properties in the following example.

Example 16.27

Let $\vec{F} = \big(3(y-1/2), 1\big)$ and let C be the path that starts at $(0,0)$, goes to $(1,1)$ along the curve $y = x^3$, then returns to $(0,0)$ along the line $y = x$, as shown in Figure 16.30. Evaluate $\oint_C \vec{F}\cdot d\vec{r}$.

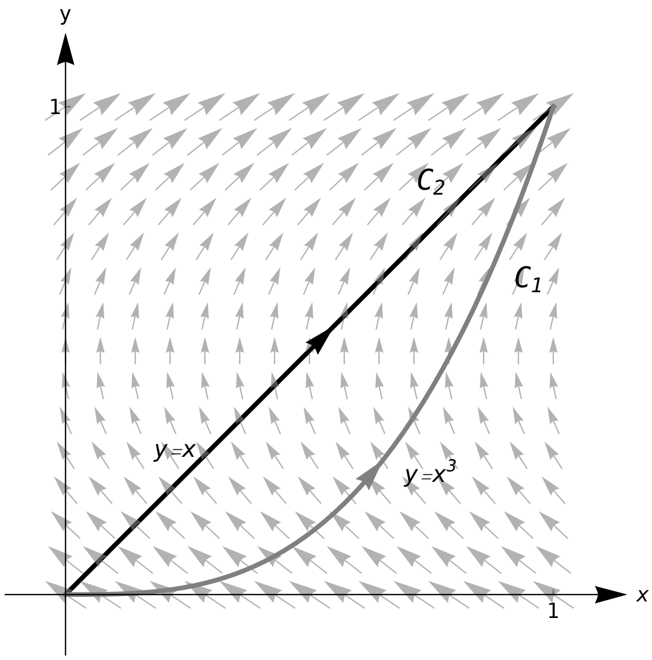


Figure 16.30: The vector field and curve in Example 16.27.

Solution

As C is piecewise smooth, we break it into two components C_1 and C_2, where C_1 follows the curve $y = x^3$ and C_2 follows the curve $y = x$.

We parametrize C_1 with $\vec{r}_1(t) = \big(t, t^3\big)$ on $0 \le t \le 1$, with $\vec{r}\,'_1(t) = \big(1, 3t^2\big)$. We will use $\vec{F}\big(\vec{r}_1(t)\big) = \big(3(t^3-1/2), 1\big)$.

While we always have unlimited ways in which to parametrize a curve, there are two direct methods to choose from when parametrizing C_2. The parametrization $\vec{r}_2(t) = (t, t)$, $0 \le t \le 1$ traces the correct line segment but with the wrong orientation. Relying on property (3) of line integrals over vector fields, we can use this parametrization and negate the result.

Another choice is to use the techniques of Section 7.1 to create the line with the orientation we desire. We wish to start at $(1,1)$ and travel in the $\vec{d} = (-1,-1)$ direction for one length of $\vec{d}$, giving equation $\vec{\ell}(t) = (1,1) + t(-1,-1) = (1-t, 1-t)$ on $0 \le t \le 1$.

Either choice is fine; we choose $\vec{r}_2(t)$ to practice using line integral properties. We find $\vec{r}\,'_2(t) = (1,1)$ and $\vec{F}\big(\vec{r}_2(t)\big) = \big(3(t-1/2), 1\big)$.

Evaluating the line integral:

$$\begin{aligned}
\oint_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} &= \int_{C_1} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}_1 - \int_{C_2} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}_2 \\
&= \int_0^1 \left(3\left(t^3 - \frac{1}{2}\right), 1\right) \cdot \left(1, 3t^2\right) dt - \int_0^1 \left(3\left(t - \frac{1}{2}\right), 1\right) \cdot (1,1)\, dt \\
&= \int_0^1 \left(3t^3 + 3t^2 - \frac{3}{2}\right) dt - \int_0^1 \left(3t - \frac{1}{2}\right) dt \\
&= \frac{1}{4} - 1 = -\frac{3}{4}.
\end{aligned}$$

If we interpret this integral as computing work, the negative work implies that the motion is mostly against the direction of the force, which seems plausible when we look at Figure 16.30.

16.7.3 The fundamental theorem of line integrals

We are preparing to make important statements about the value of certain line integrals over special vector fields. Before we can do that, we need to define some terms that describe the domains over which a vector field is defined.

A region in the plane is **connected** (*samenhangend*) if any two points in the region can be joined by a piecewise smooth curve that lies entirely in the region. In Figure 16.31(a), sets R_1 and R_2 are connected; set R_3 is not connected, though it is composed of two connected subregions.

A region is **simply connected** (*enkelvoudig samenhangend*) if every simple closed curve that lies entirely in the region can be continuously deformed (shrunk) to a single point without leaving the region. (A curve is **simple** if it does not cross itself.) In Figure 16.31(a), only set R_1 is simply connected. Region R_2 is not simply connected as any closed curve that goes around the "hole" in R_2 cannot be continously shrunk to a single point. As R_3 is not even connected, it cannot be simply connected, though again it consists of two simply connected subregions.

We have applied these terms to regions of the plane, but they can be extended intuitively to domains in space (and hyperspace). In Figure 16.31(b), the domain bounded by the sphere (at left) and the domain with a subsphere removed (at right) are both simply connected. Any simple closed path that lies entirely within these domains can be continuously deformed into a single point. In Figure 16.31(c), neither domain is simply connected. At left, the ball has a hole that extends its length and the pictured closed path cannot be deformed to a point. At right, two paths are illustrated on the torus that cannot be shrunk to a point.

We will use the terms connected and simply connected in subsequent definitions and theorems.

Recall how in Example 16.26 particles moved from $A = (-1,1)$ to $B = (1,1)$ along two different paths, wherein the same amount of work was performed along each path. It turns out that regardless of the choice of path from A to B, the amount of work performed under the field $\vec{\boldsymbol{F}} = (y, x)$ is the same. Since our expectation is that differing amounts of work are performed along different paths, we give such special fields a name.

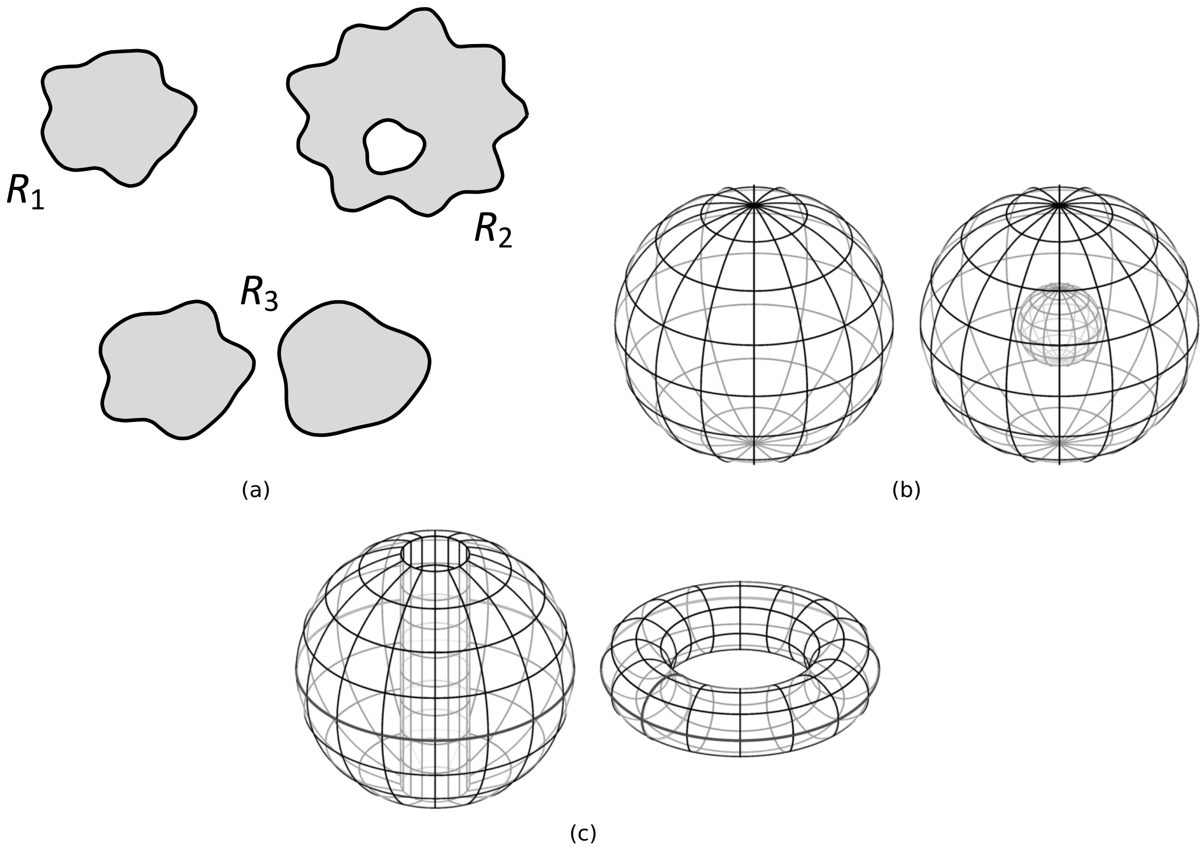


Figure 16.31: Different types of regions (a): R_1 is simply connected; R_2 is connected, but not simply connected; R_3 is not connected; simply connected domains (b) and not simply connected domains in (c).

Definitie 16.12 (Conservative field)
Let $\vec{\boldsymbol{F}}$ be a vector field defined on an open, connected domain D containing points A and B. If the line integral $\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$ has the same value for all choices of paths C starting at A and ending at B, and parametrized by $\vec{\boldsymbol{r}}(t)$, then

- $\vec{\boldsymbol{F}}$ is a **conservative field** (*conservatief vectorveld*) and
- The line integral $\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$ is path independent and can be written as

$$\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \int_A^B \vec{\boldsymbol{F}} \cdot \, d\vec{\boldsymbol{r}}.$$

How can we tell if a field is conservative? To show a field $\vec{\boldsymbol{F}}$ is conservative using the definition, we need to show that all line integrals from points A to B have the same value. It is equivalent to show that all line integrals over closed paths C are 0. Each of these tasks are generally nontrivial.

There is, however, a simpler method. Consider the surface defined by $z = f(x,y) = xy$. We can compute the gradient of this function: $\vec{\nabla} f = \left(f_x, f_y\right) = (y, x)$. Note that this is the field from Example 16.26, which we have claimed is conservative. We will soon give a theorem that states that a field $\vec{\boldsymbol{F}}$ is conservative if, and only if, it is the gradient of some scalar function f. To show $\vec{\boldsymbol{F}}$ is conservative, we need to determine whether or not $\vec{\boldsymbol{F}} = \vec{\nabla} f$ for some function f. To recognize the special relationship between $\vec{\boldsymbol{F}}$ and f in this situation, f is given a name.

Definitie 16.13 (Potential function)
Let f be a differentiable function defined on a domain D (e.g., $z = f(x,y)$ or $w = f(x,y,z)$) and let $\vec{\boldsymbol{F}} = \vec{\nabla} f$, the gradient of f. Then f is a **potential function** (*potentiaalfunctie*) of $\vec{\boldsymbol{F}}$.

We now state the fundamental theorem of line integrals, also known as the gradient theorem, which connects conservative fields and path independence to fields with potential functions.

Theorem 16.5 (Fundamental theorem of line integrals)
Let $\vec{\boldsymbol{F}}$ be a vector field whose components are continuous on a connected domain D, let A and B be any points in D, and let C be any path in D starting at A at $t = a$, ending at B at $t = b$ and parametrized by $\vec{\boldsymbol{r}}(t)$ such that $\vec{\boldsymbol{r}}(a) = A$ and $\vec{\boldsymbol{r}}(b) = B$.

1. *$\vec{\boldsymbol{F}}$ is conservative if and only if there exists a differentiable function f such that $\vec{\boldsymbol{F}} = \vec{\nabla} f$.*
2. *If $\vec{\boldsymbol{F}}$ is conservative, then*

$$\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \int_A^B \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = f\big(\vec{\boldsymbol{r}}\,(b)\big) - f\big(\vec{\boldsymbol{r}}\,(a)\big) = f(B) - f(A).$$

Note that we did not specify the number of variables for the function since it is really immaterial to the theorem. The theorem will hold regardless of the number of variables in the function.

Once again considering Example 16.26, we have $A = (-1,1)$, $B = (1,1)$ and $\vec{\boldsymbol{F}} = (y,x)$. In that example, we evaluated two line integrals from A to B and found the value of each was 2. Note that $f(x,y) = xy$ is a potential function for $\vec{\boldsymbol{F}}$. Following the fundamental theorem of line integrals, consider $f(B) - f(A)$:

$$f(B) - f(A) = f(1,1) - f(-1,1) = 1 - (-1) = 2,$$

the same value given by the line integrals.

We practice using this theorem again in the next example.

Example 16.28

Let $\vec{\boldsymbol{F}} = \big(3x^2y + 2x, x^3 + 1\big)$, $A = (0,1)$ and $B = (1,4)$. Use the first part of the fundamental theorem of line integrals to show that $\vec{\boldsymbol{F}}$ is conservative, then choose any path from A to B and confirm the second part of the theorem.

Solution

To show $\vec{\boldsymbol{F}}$ is conservative, we need to find $z = f(x,y)$ such that $\vec{\boldsymbol{F}} = \vec{\nabla} f = \big(f_x, f_y\big)$. That is, we need to find f such that $f_x = 3x^2y + 2x$ and $f_y = x^3 + 1$. As all we know about f are its partial derivatives, we recover f by integration:

$$\int \frac{\partial f}{\partial x}\, dx = f(x,y) + K_1(y).$$

Note how the constant of integration $K_1(y)$ is more than just a constant: it is anything that acts as a constant when taking a derivative with respect to x. Any function that is a function of y (containing no x's) acts as a constant when deriving with respect to x.

Integrating f_x in this example gives:

$$\int \frac{\partial f}{\partial x}\, dx = \int (3x^2y + 2x)\, dx = x^3y + x^2 + K_1(y).$$

Likewise, integrating f_y with respect to y gives:

$$\int \frac{\partial f}{\partial y}\, dy = \int (x^3+1)\, dy = x^3y + y + K_2(x).$$

These two results should be equal with appropriate choices of $K_2(x)$ and $K_1(y)$:

$$x^3y + x^2 + K_1(y) = x^3y + y + K_2(x) \quad \Rightarrow \quad K_2(x) = x^2 \quad \text{and} \quad K_1(y) = y.$$

We find $f(x,y) = x^3y + x^2 + y$, a potential function of $\vec{F}$.

By the fundamental theorem of line integrals, regardless of the path from A to B,

$$\begin{aligned}\int_A^B \vec{F}\cdot d\vec{r} &= f(B) - f(A)\\ &= f(1,4) - f(0,1)\\ &= 9 - 1 = 8.\end{aligned}$$

To illustrate the validity of the Fundamental Theorem, we pick a path from A to B. The line between these two points would be simple to construct; we choose a slightly more complicated path by choosing the parabola $y = x^2 + 2x + 1$. This leads to the parametrization $\vec{r}(t) = \left(t, t^2 + 2t + 1\right)$, $0 \le t \le 1$, with $\vec{r}\,'(t) = (1, 2t+2)$. Thus

$$\begin{aligned}\int_C \vec{F}\cdot d\vec{r} &= \int_C \vec{F}\big(\vec{r}(t)\big)\cdot \vec{r}\,'(t)\ dt\\ &= \int_0^1 \left(3t^2(t^2+2t+1) + 2t, t^3+1\right)\cdot(1, 2t+2)\ dt\\ &= \int_0^1 \left(5t^4 + 8t^3 + 3t^2 + 4t + 2\right)\ dt\\ &= \left(t^5 + 2t^4 + t^3 + 2t^2 + 2t\right)\Big|_0^1\\ &= 8,\end{aligned}$$

which matches our previous result.

The fundamental theorem of line integrals states that we can determine whether or not $\vec{F}$ is conservative by determining whether or not $\vec{F}$ has a potential function. This can be difficult. A simpler method exists if the domain of $\vec{F}$ is simply connected, which is a reasonable requirement. We state this simpler method as a theorem.

Theorem 16.6 (Curl of conservative fields)
Let $\vec{F}$ be a vector field whose components are continuous on a simply connected domain D in the plane or in space. Then $\vec{F}$ is conservative if and only if $\operatorname{curl}\vec{F} = 0$ *or* $\vec{0}$, *respectively.*

In Example 16.28, we showed that $\vec{F} = \left(3x^2y + 2x, x^3 + 1\right)$ is conservative by finding a potential function for $\vec{F}$. Using the above theorem, we can show that $\vec{F}(M,N)$ is conservative much more easily by computing its curl:

$$\operatorname{curl}\vec{F} = N_x - M_y = 3x^2 - 3x^2 = 0.$$

16.8 Exercises

Iterated integrals and area

Assignment 16.1 — Evaluate the given double integrals.

(a) $\displaystyle\int_0^2 \int_x^{2x} \left(x^2+y^2\right)\, dy\, dx$

(b) $\displaystyle\int_1^2 \int_y^{y^3} e^{\frac{x}{y}}\, dx\, dy$

(c) $\displaystyle\int_1^2 \int_y^{3y} (x+y)\, dx\, dy$

(d) $\displaystyle\int_{-1}^2 \int_{2x^2-2}^{x^2+x} x\, dy\, dx$

(e) $\displaystyle\int_0^{\pi} \int_0^{\cos(\theta)} r\sin(\theta)\, dr\, d\theta$

(f) $\displaystyle\int_0^{\frac{\pi}{2}} \int_2^{4\cos(\theta)} r^3\, dr\, d\theta$

(g) $\displaystyle\int_0^1 \int_0^1 \frac{x^2}{1+y^2}\, dy\, dx$

(h) $\displaystyle\int_{-1}^2 \int_0^1 (xy^2+x^2y)\, dx\, dy$

(i) $\displaystyle\int_0^1 \int_{x^2}^{2-x^2} \sqrt{x}y\, dy\, dx$

(j) $\displaystyle\int_0^{\pi/4} \int_{\sin(x)}^{\cos(x)} xy\, dy\, dx$

(k) $\displaystyle\int_0^1 \int_{\sqrt[3]{y}}^1 e^{x^2}\, dx\, dy$

Assignment 16.2 — Find the integration boundaries for the given double integrals and evaluate.

(a) $\displaystyle\iint_R x^2y\, dA$ with R the triangle OAB with $A(3,0)$ and $B(3,2)$.

(b) $\displaystyle\iint_R dA$ with R the region in the first quadrant between $y^2=x^3$ and $y=x$

(c) $\displaystyle\iint_R x^2\, dA$ with R the region in the first quadrant between $xy=16$, $y=x$, $y=0$ and $x=8$

(d) $\displaystyle\iint_R y\, dA$ with R the region enclosed by $y=x^2$ and $y=x^3$

(e) $\displaystyle\iint_R \frac{1}{\sqrt{2y-y^2}}\, dA$ with R the region enclosed by $x^2=4-2y$ in the first quadrant

(f) $\displaystyle\iint_R e^{\frac{x}{y}}\, dA$ with R the region bounded by $y^2=x$, $x=0$ en $y=1$

(g) $\displaystyle\iint_R (y-x)\, dA$ with R the region enclosed by $(x-1)^2+y^2=1$

Assignment 16.3 — Reverse the order of integration in the integrals below.

(a) $\displaystyle\int_0^3 \int_1^{\sqrt{4-y}} f(x,y)\, dx\, dy$

(b) $\displaystyle\int_0^1 \int_{\arccos(y)}^{\frac{\pi}{2}} f(x,y)\, dx\, dy$

(c) $\displaystyle\int_{-6}^{2} \int_{\frac{x^2}{4}}^{3-x} f(x,y)\, dy\, dx$

Assignment 16.4 — Using a double integral, find the area of the following regions.

(a) the region bounded by $y^2 = 10x + 25$ and $y^2 = -6x + 9$

(b) the region bounded by $y = x^3$, $x + y = 2$ and the x-axis

(c) the region bounded by $x^2 + y^2 = 2x$, $x^2 + y^2 = 4x$, $y = x$ and $y = 0$

(d) the region bounded by $r\cos(\theta) = 1$ and $r = 2$ (area that does not contain $r = 0$)

(e) region bounded by $r = 1 + \cos(\theta)$ and $r = \cos(\theta)$

Double integration and volume

Assignment 16.5 — Find the volume of the given regions using Cartesian coordinates.

(a) the region enclosed by $x = 0$, $x = 4$, $y = 0$, $y = 4$, $z = 0$ en $xy = 4z$

(b) the region enclosed by the three coordinate planes, together with the planes $x^2 + 4y^2 = 4$ and $y^2 + 2z = 4$

(c) the region enclosed by the three coordinate planes, together with the planes $z = 1 - x^2 - y^2$ en $x + y \leq 1$

(d) the region enclosed by the planes $y = x^2$, $x = y^2$, $z = 0$ en $z = 12 + y - x^2$

(e) the region formed by a truncated prism with upright side faces $x = 0$, $x = 2$, $y = 0$, $y = 1$, as ground plane $z = 0$ and as upper plane $-x + y + 2z = 4$

(f) the region enclosed by $x^2 + y^2 = 1$ and $x^2 + z^2 = 1$

Assignment 16.6 — Set up the double integral to find the volume of the given region in Cartesian coordinates and in polar coordinates.

(a) the region enclosed by $z = 0$, $z = x + y + 2$ and $x^2 + y^2 = 16$ in the first quadrant

(b) the region enclosed by $z = 0$, $z = 2xy$, $x^2 + y^2 + x = 0$, $x < 0$ and $y < 0$

Assignment 16.7 — Find the volume of the given region using polar coordinates.

(a) the region enclosed by $x^2 + y^2 = 2$, $z = 0$ and $z = 2$ and where $z \geq x^2 + y^2$

(b) the region enclosed by $x^2+y^2=2$, $z=0$ and $z=2$ and where $z \leq x^2+y^2$

(c) the region enclosed by $x^2+y^2=1$, cut off by $x^2+y^2+z^2=4$

(d) the region enclosed by $x^2+y^2=2z$, cut off by $x^2+y^2+z^2=3$

(e) the region enclosed by $x^2+y^2+z^2=a^2$ en $x^2+y^2=ax$

(f) the region enclosed by $x^2+y^2=2y$ and $z^2=y$

Centre of mass

Assignment 16.8 — Find the center of mass of the regions below with the given mass density.

(a) the flat region enclosed by $y=\sin(x)$ and $y=0$ $(0 \leq x \leq \pi)$ with mass density $\delta(x,y)=ky$.

(b) the flat region enclosed by $y^2=4x+4$ en $y^2=-2x+4$ with mass density $\delta=1$

(c) a triangular plate enclosed by the x- and y-axis and the $2x+3y=12$ with mass density $\delta=1$

Surface area

Assignment 16.9 — Find the requested surface area.

(a) the part of $x^2+y^2=3z^2$, located above the xy-plane and inside $x^2+y^2=4y$

(b) the part of $x^2+z^2=16$, inside $x^2+y^2=16$

(c) the part of $z^2=4x$, inside $y^2=4x$ and $x \leq 1$

(d) the part of $z=1-x^2-y^2$, inside $x^2+y^2=1$

(e) the part of $x^2+y^2+z^2=1$ that is cut off by $x^2+4y^2=1$.

Line integrals over a scalar field

Assignment 16.10 — Evaluate the line integral of the scalar functions below along the given curve.

(a) $\displaystyle\int_C x^2\, ds$ along the intersection of $x-y+z=0$ en $x+y+2z=0$ from $(0,0,0)$ to $(3,1,-2)$

(b) $\displaystyle\int_C y\, ds$ from $x=3$ to $x=24$ along the curve $C: y=2\sqrt{x}$

(c) $\displaystyle\int_C (x+y)\,ds$ along the right loop of $r^2 = 2\cos(2\theta)$

(d) $\displaystyle\int_C \frac{ds}{x^2+y^2+z^2}$ along the first loop of $x(t) = 8\cos(t),\ y(t) = 8\sin(t),\ z(t) = t$

(e) $\displaystyle\int_C \sqrt{2y^2+z^2}\,ds$ along the curve

$$C : \begin{cases} x^2+y^2+z^2 = 4 \\ y = x \end{cases}$$

(f) $\displaystyle\int_C e^z\,ds$ along the curve $x(t) = e^t\cos(t),\ y(t) = e^t\sin(t),\ z(t) = t$

(g) $\displaystyle\int_C \sqrt{1+4x^2z^2}\,ds$ along the curve

$$C : \begin{cases} x^2+z^2 = 1 \\ y = x^2 \end{cases}$$

(h) $\displaystyle\int_C x\,ds$ along the curve

$$C : \begin{cases} x^2+y^2 = a^2 \\ z = x \end{cases}$$

in the first octant

(i) $\displaystyle\int_C z\,ds$ along the curve

$$C : \begin{cases} x^2+y^2+z^2 = 1 \\ x+y = 1 \end{cases} \qquad \text{met } z \geq 0$$

Vector fields

Assignment 16.11 — Which vector field belongs to which graph in Figure 16.32?

(a) $\vec{\boldsymbol{F}}_1 = \dfrac{\vec{\boldsymbol{r}}}{\|\vec{\boldsymbol{r}}\|}$

(b) $\vec{\boldsymbol{F}}_2 = \vec{\boldsymbol{r}}$

(c) $\vec{\boldsymbol{F}}_3 = y\hat{\boldsymbol{\imath}} - x\hat{\boldsymbol{\jmath}}$

(d) $\vec{\boldsymbol{F}}_4 = x\hat{\boldsymbol{\jmath}}$

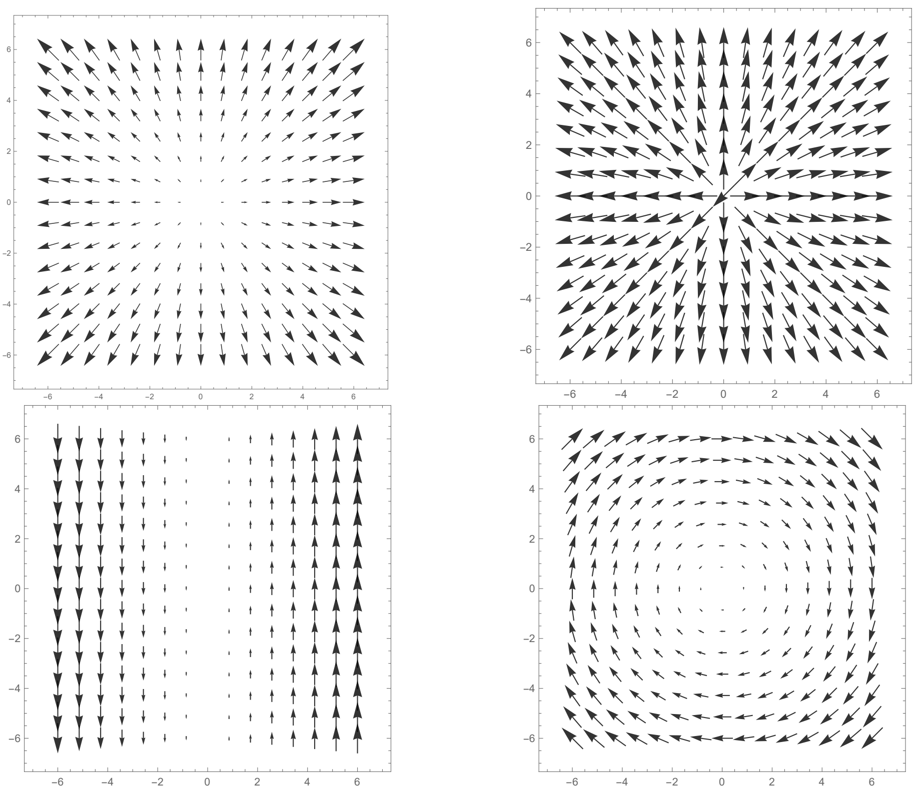


Figure 16.32: Vector fields from exercise 16.11

Assignment 16.12 — Consider a scalar function $f(x,y,z) = x^2yz^3$ and a vector field $\vec{F} = (xz, -y^2, 2x^2y)$. Find $\vec{\nabla} f$, $\vec{\nabla} \cdot \vec{F}$ and $\vec{\nabla} \times \vec{F}$.

Assignment 16.13 — Consider a scalar function $f(x,y,z) = xy + yz + xz$ and a vector field $\vec{F} = (x^2y, -y^2z, z^2x)$. Find $\vec{\nabla} f$, $\vec{\nabla} \cdot \vec{F}$, $\vec{\nabla} \times \vec{F}$ en $(\vec{\nabla} f) \times \vec{F}$ at $(3,-1,2)$.

Assignment 16.14 — Find all points in space where the direction of the given vector field does not change.

(a) $\vec{F} = (xy^2, xyz, z - 2x)$

(b) $\vec{F} = (xy^3, xyz, z - x^2)$

Assignment 16.15 — Consider the vector field $\vec{F} = \left(\frac{3x}{z}, 2x, 7yz\right)$ and the curve C described by $\vec{r}(t) = (\cos^2(t), \sin(t), -\cos(t))$ with $0 \le t < 2\pi$. At which points of C has the rotor of $\vec{F}$ a min/max length?

Assignment 16.16 — The coulomb potential V in a vacuum at a point $P(x,y,z)$ originating from a point charge q placed at the origin is given by

$$V(x,y,z) = \frac{q}{4\pi\epsilon_0\sqrt{x^2+y^2+z^2}},$$

with ϵ_0 a constant. Find the corresponding electric field given by $\vec{E} = -\vec{\nabla} V$.

Line Integrals over vector fields

Assignment 16.17 — Evaluate the line integral of the given vector field along the given curve(s).

(a) $\vec{\boldsymbol{F}} = \big(\cos(x), y\big)$ along $y = \sin(x)$ from $(0,0)$ to $(\pi, 0)$

(b) $\vec{\boldsymbol{F}} = (xy, y - x)$ from $(0,0)$ to $(1,1)$ along the curves $C_1 : y = x$, $C_2 : y = x^2$ and $C_3 : y^2 = x$

(c) $\vec{\boldsymbol{F}} = (2xy, x^2)$ from $(0,0)$ to $(1,2)$ along the curves $C_1 : y = 2x$, $C_2 : y = 2x^2$, $C_3 : y^2 = 4x$ and $C_4 : y = 2x^3$

(d) $\vec{\boldsymbol{F}} = \left(\dfrac{1}{\sqrt{xy}}, -\dfrac{1}{\sqrt{xy}}\right)$ along the curve $C : y = 1 - x$ from $x = 0$ to $x = 1$

(e) $\vec{\boldsymbol{F}} = (x^2 y, xy^2)$ along the curve

$$C : \begin{cases} x = \dfrac{t}{2} \\ y = \sqrt{2t} \end{cases}$$

from $(0,0)$ to $(1,2)$

(f) $\vec{\boldsymbol{F}} = \left(\sqrt{b^2 - y^2}, \sqrt{a^2 - x^2}\right)$ along the curve

$$C : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

with $x \geq 0, y \geq 0$ travelled in a counterclockwise fashion

(g) $\vec{\boldsymbol{F}} = (-3y, 2x)$ along the curve $C : abca$ with $a(1,2)$, $b(3,1)$ and $c(3,2)$

(h) $\vec{\boldsymbol{F}} = \left(3x^2 + 2xy, x^2 + y^2\right)$ from $(1,1)$ to $(2,2)$ along the curve $C_1 : y = x$ and along the path given by $C_2 : y = 1$ and $C_3 : x = 2$

(i) $\vec{\boldsymbol{F}} = (y^2, z^2, x^2)$ along the curve

$$C : \begin{cases} y = 1 \\ x^2 + y^2 + z^2 = 5 \end{cases}$$

with $x \geq 0$, travelled in the direction of increasing z-values

Assignment 16.18 — Do the following sets of points represent a region, a connected region or simply connected region?

(a) $R = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x > 0,\ y \geq 0\right\}$

(b) $R = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x = 0,\ y \geq 0\right\}$

(c) $R = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x \neq 0,\ y > 0\right\}$

(d) $R = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 > 1\right\}$

(e) $R = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 > 1\right\}$

(f) $R = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 + z^2 > 1\right\}$

Assignment 16.19 — Verify that the given vector field is conservative and, if possible, find a potential function.

(a) $\vec{\boldsymbol{F}} = \left(xy, \frac{1}{2}x^2 - y^2\right)$

(b) $\vec{\boldsymbol{F}} = (y, x, -2z)$

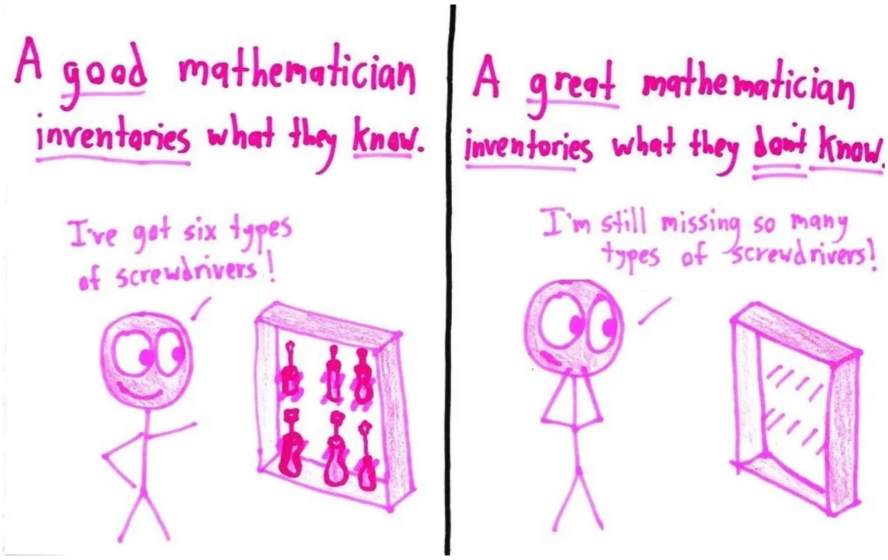
(c) $\vec{\boldsymbol{F}} = \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$

(d) $\vec{\boldsymbol{F}} = \left(\frac{x}{x^2+y^2}, -\frac{y}{x^2+y^2}\right)$

(e) $\vec{\boldsymbol{F}} = \left(2xy - z^2, 2yz + x^2, -2zx + y^2\right)$

(f) $\vec{\boldsymbol{F}} = e^{x^2+y^2+z^2}(xz, yz, xy)$

(g) $\vec{\boldsymbol{F}} = \left(xy - \sin(z), \frac{1}{2}x^2 - \frac{e^y}{z}, \frac{e^y}{z^2} - x\cos(z)\right)$



From *Math with Bad Drawings*, used by permission of Ben Orlin.

PART IV

DIFFERENTIAL EQUATIONS

Science is a differential equation and religion is a boundary condition.

— Alan Turing —

17

Introduction

17.1 Mathematical modelling and simulation

17.1.1 Mathematical models

In order to describe a biological, natural or physical process mathematically, we must formulate it in mathematical terms; that is, we must construct a mathematical model of the process. Once in place, it can be used to map one or more inputs related to the process at stake to one or multiple outputs as illustrated in Figure 17.1. Below we give a more formal definition of a mathematical model.

Definitie 17.1 (Mathematical model)
A **mathematical model** (*wiskundig model*) is an abstract, simplified description of a part of reality that is created for a particular purpose.

From this definition we infer several important aspects of mathematical models. Firstly, they are abstract because they are based on mathematical formalism, i.e. mathematical equations. Secondly, they constitute a simplified representation of the studied process(es), which implies that their development typically involves some simplifying assumptions, and hence that the modelling results only approximate the studied real-world processes up to some level. Thirdly, even though many other processes might affect the focal process, some of these are neglected, again possibly leading to discrepancies between the modelling results and the dynamics of real-world processes. Fourthly, and lastly, it is only meaningful to build a model if it is clear what question(s) it should be able to answer. A good mathematical model has two important properties:

- It is sufficiently simple so that the problem can be solved.
- It represents the actual situation sufficiently well so that the modelling results agree with observations to within a useful degree of accuracy.

Clearly, if the modelling results do not agree with observations, the underlying assumptions of the model must be revised until satisfactory agreement is obtained. When sufficient agreement between the modelling results and observations is reached, i.e. the model got validated, it can be used to run simulations.

Definitie 17.2 (Simulation)
Simulation (*simulatie*) is the imitation of the operation of a real-world process by means of a mathematical model.

Increasingly often such simulations are run on computers or computing clusters, and are therefore also referred to as *in silico* experiments.

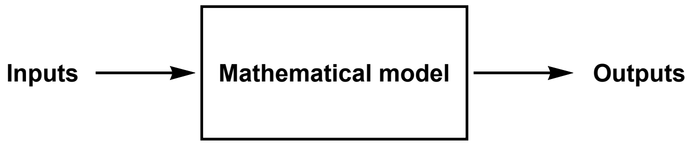


Figure 17.1: Schematic representation of the functioning of a mathematical model.

In contrast to general thinking, everyone is nowadays on a daily basis often confronted with mathematical models or results obtained therewith. For instance, the weather forecast is based on dedicated weather models that describe atmospheric processes such as radiation, turbulence, evaporation, and so on. Indeed, for compiling its weather forecasts Belgium's Royal Meteorological Institute uses the weather model ALARO. Similarly, the Flemish Environment Agency (VMM) uses hydrological models to forecast the water level in Flemish rivers, and, in the case of inundations, to forecast their extent and the water height above the ground level in real time [1].

17.1.2 Model development

Many physical problems concern relationships between changing quantities. Since rates of change are represented mathematically by derivatives, mathematical models often involve equations relating an unknown function and one or more of its derivatives.

Consider for instance the rate of change of the velocity of a free-falling object with mass m [M] in a vacuum, and suppose that we want to know the velocity v [$\mathrm{L\,T^{-1}}$] of this object over time [T]; this the goal of our mathematical model. In that case, there is only one force exerted on this object, namely the gravitational force $m\,g$ [$\mathrm{M\,L\,T^{-2}}$], where g is the acceleration due to gravity. If we assume that the velocity v of the object is positive in the downward direction, the velocity and gravitational vectors point in the same direction. So, we know from Newton's second law that the instantaneous acceleration a [$\mathrm{L\,T^{-2}}$] of an object with constant mass m is related to this force as:

$$m\,a = m\,g\,, \tag{17.1}$$

or equivalently

$$a = g\,. \tag{17.2}$$

As the acceleration of an object a is nothing but the rate of change of its velocity, we can rewrite Equation (17.2) using the derivative of $v(t)$ as:

$$\frac{dv}{dt} = g\,. \tag{17.3}$$

[1] https://www.waterinfo.be/

The solution of Equation (17.3) we are looking for is not just a scalar value like in the case of algebraic equations, but rather the function $v(t)$. Since this equation contains the derivative of this unknown function, it is referred to as a **differential equation** (*differentiaalvergelijking*).

The mathematical model based on Equation (17.3) has the initial velocity as input, while its output will be the velocity of the falling object over time. Its practical value, however, will be limited because it is based on the simplifying assumption that the object is falling in a vacuum. Fortunately it can be brought closer to reality by accounting for the drag force that acts on the object as it is moving downwards. This drag force acts in the direction opposite to the one of the gravitational force, and is often assumed to be proportional to the velocity of the falling object, so the more realistic counterpart of Equation (17.3) becomes

$$m\frac{dv}{dt} = m\,g - \mu\,v\,, \tag{17.4}$$

where μ [$\mathrm{M\,T^{-1}}$] is the drag coefficient. Figure 17.2 shows different solutions of Equation (17.4) for $g = 9.81$ m s^{-2}, $m = 75$ kg, $\mu = 5.28$ kg s^{-1} and an initial velocity v_0 varying between 0 and 240 m s^{-1}. It is clear that the initial velocity has only an impact on how fast the terminal velocity is reached, but not on the terminal velocity itself.

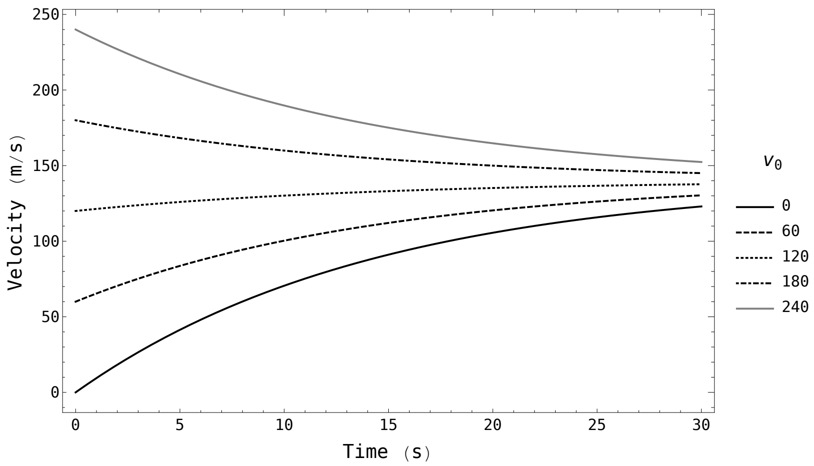


Figure 17.2: Solutions of Equation (17.4) for $g = 9.81$ m s^{-2}, $m = 75$ kg, $\mu = 5.28$ kg s^{-1} and initial velocity v_0 varying between 0 and 240 m s^{-1}.

In addition to processes that involve a continuous change over time, there are also many processes that bring along a change only after a fixed amount of time has passed, such as reproductive events of many animals like fish (e.g. pike) and butterflies (e.g. the box moth). In these cases it becomes more meaningful to track, for instance, the species population size at discrete points in time, so we shall consider populations with a fixed time period between generations. Thus, we shall describe population size by a sequence N_k, with N_0 [–] denoting the initial population size, N_1 the population size of the first generation (at time step 1), N_2 the population size of the second generation (at time step 2), and so on. In the case of the box moth (*Cydalima perspectalis*), the population changes only through eggs hatching and caterpillars entering the pupa stage, which happens at rates e [–] and p [–], respectively. Then, the size of the box moth population at a time step $k+1$ is given by

$$N_{k+1} = (e - p)\,N_k\,, \tag{17.5}$$

or upon introducing the net growth $r = e - p$ [–]:

$$N_{k+1} = r\,N_k\,. \tag{17.6}$$

The mathematical model based on Equation (17.6) is referred to as the model of Malthus, and the type of equation it uses is called a **difference equation** (*differentievergelijking*). The long-term population size when starting from an initial population of size N_0 can be forecast by iteratively replacing N_k by $r\,N_{k-1}$ in the right-hand side of Equation (17.6):

$$\begin{aligned} N_{k+1} &= r\,N_k \\ &= r^2\,N_{k-1} = r^3\,N_{k-2} \\ &= r^{k+1}\,N_0\,. \end{aligned}$$

We infer that the population will grow exponentially if $r > 1$ and evolve to 0 if $r < 1$. In reality, the exponential growth dictated by this model is of course very unrealistic because the more individuals that are already present in a region and use its resources, the more difficult it will become to sustain yet another individual. Suppose the environment of a given region can support K [–] individuals, i.e. the carrying capacity of the environment, then Equation (17.6) can be reformulated as

$$N_{k+1} = r\left(1 - \frac{N_k}{K}\right)N_k\,. \tag{17.7}$$

The mathematical model based on this equation is widely known as the model of Verhulst and is an example of a logistic population model. Figure 17.3 visualizes for different net growth rates the population size over time of a box moth population occupying an environment that can support up to 200 individuals ($K = 200$) starting from an initial population with two individuals. As opposed to the impact of the drag coefficient μ on the terminal velocity (Figure 17.2), the net growth rate r does not only affect the road towards the terminal population, but also the terminal population size itself. Moreover, it seems that for some net growth rates (e.g. $r = 3.5$) the population size does not converge to a constant value but keeps fluctuating.

From the two examples given in this section, it should be clear that mathematical models are very useful for gaining insights into long-term behaviour, while they also allow to investigate what might happen in different scenarios. This is, however, possible only when one has good understanding of the equations underlying those models.

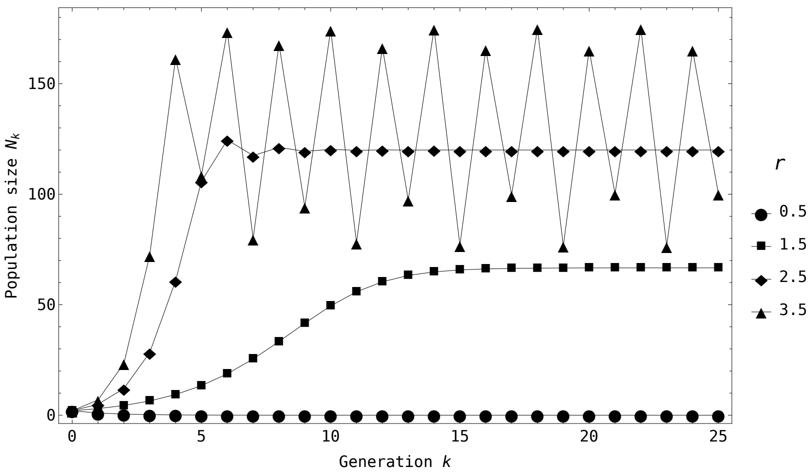


Figure 17.3: Population size of a box moth population over time as given by Equation (17.7) for a carrying capacity $K = 200$ and different net growth rates r.

17.2 Differential and difference equations

Here, we will introduce the formal definitions of both differential and difference equations, and also have a look at their established nomenclature.

17.2.1 Differential equations

17.2.1.1 Nomenclature

Definitie 17.3 (Differential equation)
A **differential equation** (*differentiaalvergelijking*) is an equation that contains one or more derivatives of an unknown function $y = g(t)$, or mathematically, a differential equation is any equation that can be written as:

$$F\left(t, y, \frac{dy}{dt}, \frac{d^2y}{dt^2}, \ldots, \frac{d^ny}{dt^n}\right) = 0\,. \tag{17.8}$$

The **order** (*orde*) of a differential equation corresponds to the order of the highest-order derivative that appears in the differential equation, while the **degree** (*graad*) of a differential equation is the power to which the highest order derivative is raised. So, Equation (17.8) may be referred to as an n-th order differential equation of degree one. Similarly, the differential equation we encountered in Section 17.1 for describing the velocity of a free-falling object (Equation (17.4)) was a first-order differential equation of degree one. Equation (17.8) gives the implicit form of a differential equation. More commonly, we write differential equations in some explicit way, for instance, by isolating the highest order derivative:

$$\frac{d^ny}{dt^n} = f\left(t, y, \frac{dy}{dt}, \frac{d^2y}{dt^2}, \ldots, \frac{d^{n-1}y}{dt^{n-1}}\right),$$

or upon introducing the shorthand notation $y^{(n)} = \frac{d^ny}{dt^n}$:

$$y^{(n)} = f\left(t, y, y', y'', \ldots, y^{(n-1)}\right). \tag{17.9}$$

Alternatively, we might separate the unknown function and its derivatives from the terms that involve the independent variable only,

$$G\left(t, y, y', y'', \ldots, y^{(n)}\right) = q(t)\,, \tag{17.10}$$

where G and q are just functions. When the latter function $q(t)$ is identically zero, Equation (17.10) is called a **homogeneous differential equation** (*homogene differentiaalvergelijking*). Otherwise, it is a non-homogeneous differential equation, and $q(t)$ is referred to as the **non-homogeneous term** (*niet-homogene term*).

Definitie 17.4 (Linear differential equation)
An n-th order differential equation is called **linear** (*lineair*) if it is linear in the terms involving the unknown function and its derivatives, and hence can be written as

$$F\left(t, y, y', y'', \ldots, y^{(n)}\right) = a_n(t)\,y^{(n)} + a_{n-1}(t)\,y^{(n-1)} + \cdots + a_2(t)\,y'' + a_1(t)\,y' + a_0(t)\,y + b(t) = 0\,.$$

Otherwise, it is called **non-linear** (*niet-lineair*).

The differential equation we proposed in Section 17.1 to describe the velocity of a free-falling object (Equation (17.4)) was a linear first-order differential equation. In Chapters 19, **??** and **??** we will review the most important analytical and numerical solution methods for such differential equations and systems thereof.

For many processes, the rate of change of the studied quantity depends only on the system, described by $y(t)$, and not on external factors. Describing such processes can be done using so-called autonomous differential equations.

Definitie 17.5 (Autonomous differential equation)
A differential equation is called **autonomous** (*autonoom*) if none of its terms depend explicitly on the independent variable t, so mathematically

$$H\left(y, y', y'', \ldots, y^{(n)}\right) = 0\,.$$

A differential equation is an **ordinary differential equation** (*gewone differentiaalvergelijking*) if it involves an unknown function of only one variable, or a **partial differential equation** (*partiële differentiaal vergelijking*) if it involves partial derivatives of a function of more than one variable. For instance, let $T(x,y)$ denote the temperature in some point (x,y) of a room, then the partial differential equation that describes the temperature distribution of the room is given by

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0\,.$$

In the remainder of this course, the focus will, however, be on ordinary differential equations, from now on referred to as differential equations for short.

17.2.1.2 Solutions of differential equations

A solution of a differential equation is a function that satisfies the differential equation on some open t interval; thus, $y(t)$ is a solution of Equation (17.9) if 1) $y(t)$ is n times differentiable and 2)

$$y^{(n)}(t) = f(t, y(t), y'(t), \ldots, y^{(n-1)}(t))$$

for all t in some open interval $]a,b[$. From this we infer that it is in general not possible to define a priori the open interval $]a,b[$, though in the subsequent chapters we will see that this can be done for linear differential equations.

The graph of a solution of a differential equation is called a **solution curve** (*oplossingsgrafiek*), so the curves in Figure 17.2 were nothing but solution curves of Equation (17.4) for different initial velocities.

Example 17.1

Show that if C_1 and C_2 are constants then

$$y(t) = (C_1 + C_2 t)e^{-t} + 2t - 4 \tag{17.11}$$

is a solution of the following linear second-order differential equation

$$y'' + 2y' + y = 2t \tag{17.12}$$

on the open interval $]-\infty, +\infty[$.

Solution

Differentiating Equation (17.11) twice yields

$$y'(t) = -(C_1 + C_2 t)\,e^{-t} + C_2 e^{-t} + 2$$

and

$$y''(t) = (C_1 + C_2 t)\,e^{-t} - 2C_2 e^{-t}\,.$$

Substituting these expressions into the left-hand side of Equation (17.12) one gets

$$(C_1 + C_2 t)\,e^{-t} - 2C_2 e^{-t} + 2\big[-(C_1 + C_2 t)\,e^{-t} + C_2 e^{-t} + 2\big] + (C_1 + C_2 t)\,e^{-t} + 2t - 4\,,$$

and after some algebra:

$$(1 - 2 + 1)\,(C_1 + C_2 t)\,e^{-t} + (-2 + 2)\,C_2 e^{-t} + 4 + 2t - 4 = 2t\,,$$

for all values of t. Therefore $y(t)$ is a solution of Equation (17.12) on the open interval $]-\infty, +\infty[$.

In the next example, we will see that the open interval $]a, b[$ does not necessarily stretch from $-\infty$ to $+\infty$, and that there might be points where the solution is not defined.

Example 17.2

Determine the open interval(s) $]a, b[$ where

$$y(t) = \frac{t^2}{3} + \frac{1}{t} \tag{17.13}$$

is a solution of

$$ty' + y = t^2\,. \tag{17.14}$$

Solution

Substituting Equation (17.13) and its first-order derivative

$$y'(t) = \frac{2t}{3} - \frac{1}{t^2}$$

into the left-hand side of Equation (17.14) yields

$$t\left(\frac{2t}{3} - \frac{1}{t^2}\right) + \left(\frac{t^2}{3} + \frac{1}{t}\right).$$

This equals t^2 only if $t \neq 0$. So, $y(t)$ is a solution of Equation (17.14) on $]-\infty, 0[$ and $]0, +\infty[$. However, $y(t)$ is not a solution of the differential equation on any open interval that contains $t = 0$, since $y(t)$ is not defined at $t = 0$.

Figure 17.4 shows the graph of Equation (17.13). The part of the graph of Equation (17.13) on $]0, +\infty[$ is a solution curve of Equation (17.14), as is the part of the graph on $]-\infty, 0[$.

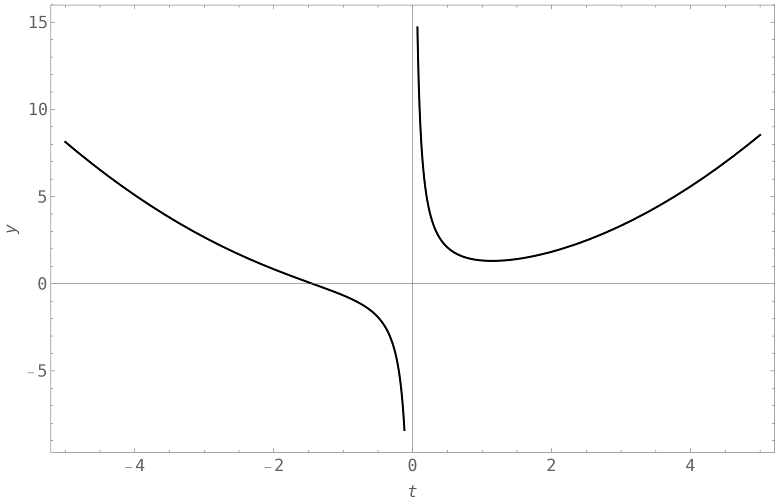


Figure 17.4: Solution curves of Equation (17.14).

In Example 17.1 we saw that the differential equation $y'' + 2y' + y = 2t$ has an infinite family of solutions that depend upon 2 arbitrary constants C_1 and C_2. We refer to such a solution still involving constants as the **general solution** (*algemene oplossing*) of a differential equation. In the absence of additional conditions, there is no reason to prefer one solution over another. However, often we will be interested in finding a solution of a differential equation that satisfies one or more specific conditions. Suppose, for instance, that we know the initial state of the system described by means of Equation (17.12). More precisely, let us assume that $y(0) = 1$ and $y'(0) = 0$, i.e. we know one point on the solution curve of Equation (17.12) and one point on the curve describing the first-order derivative of $y(t)$. Plugging these two points in Equation (17.11) and its first-order derivative, respectively, we arrive at the following system of equations:

$$\begin{aligned} 1 &= C_1 - 4, \\ 0 &= -C_1 + C_2 + 2. \end{aligned}$$

From this we immediately see that $C_1 = 5$ and $C_2 = 3$, so that the solution of Equation (17.12) satisfying the imposed conditions becomes

$$y(t) = (5 + 3t)e^{-t} + 2t - 4\,.$$

The solution of Equation (17.12) we obtained by assigning particular values to the arbitrary constants in its general solution (Equation (17.11)) is referred to as a **particular solution** (*particuliere oplossing*) of the differential equation.

Problems that involve both a differential equation and information of the initial state of the studied system are of interest in many disciplines, and are referred to as initial value problems. A more formal definition is given below.

Definitie 17.6 (Initial value problem)
An **initial value problem** (*beginwaardeprobleem*) consists of an n-th order differential equation

$$F\left(t, y, y', y'', \ldots, y^{(n)}\right) = 0\,,$$

defined for t in an open interval $]a,b[$ and a set of n initial conditions

$$\begin{aligned} y(t_0) &= y_0, \\ y'(t_0) &= y_1, \\ y''(t_0) &= y_2, \\ &\vdots \\ y^{(n-1)}(t_0) &= y_{n-1}, \end{aligned}$$

where t_0 lies in $]a,b[$, and $y_0, y_1, \ldots, y_{n-1}$ are given constants.

From this definition it is clear that we will need n initial conditions in order to be able to solve an initial value problem involving an n-th order differential equation. Consistent with the definition of a solution of the n-th order differential equation given by Equation (17.9), we say that $y(t)$ is a solution of the initial value problem stated in Definition 17.6 if 1) $y(t)$ is n times differentiable and 2)

$$y^{(n)}(t) = f(t, y(t), y'(t), \ldots, y^{(n-1)}(t))$$

for all t in some open interval $]a,b[$ that contains t_0, and 3) $y(t)$ satisfies the imposed initial conditions. In the remainder of this course, we will use the notation $y_{(t_0,y_0)}(t)$ to denote the solution of a differential equation that satisfies $y(t_0) = y_0$.

The largest open interval that contains t_0 on which $y(t)$ is defined and satisfies the differential equation is called the **interval of existence** (*bestaansinterval*). Let us now see what this all gives for Equation (17.4).

Example 17.3

Determine with Mathematica the solution of

$$m\frac{dv}{dt} = m\,g - \mu\,v$$

that satisfies the initial condition $v(0) = v_0$.

Solution

Let us first of all compute the general solution of Equation (17.4). In Mathematica, this is possible using the function **DSolve** as follows:

```
In[29]:= DSolve[m*D[v[t],t] == m*g-mu*v[t], v[t], t]

Out[29]= {{v[t]→(g m)/mu + e^(-(t mu)/m) C[1]}}
```

Note the double = in the input line to express equality! Furthermore, note that curly brackets in Mathematica indicate a list, so our output is a list within a list. This is due to the fact that **DSolve** can lead to multiple solutions. These are given as replacement rules (denoted by arrows) within a list, that are all again collected in one list. Lastly, as can be inferred from the output, Mathematica uses C[1] to denote the arbitrary constant in the general solution. The particular solution satisfying $v(0) = 0$ can also be computed easily in Mathematica by providing it with the appropriate initial condition:

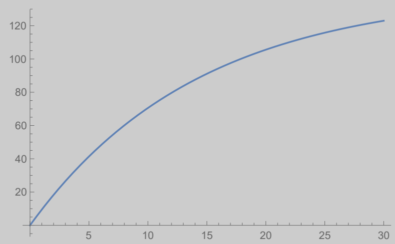
In[30]:=
```
sol=DSolve[{m*D[v[t],t] == m*g-mu*v[t], v[0] == v0}, v[t], t]
```

Out[30]= $\{\{v[t] \to \frac{e^{-\frac{t\,mu}{m}}\left(-g\;m+e^{\frac{t\,mu}{m}}\;g\;m+v0\;mu\right)}{mu}\}\}$

Note that the particular solution was stored in the variable `sol`. Upon choosing $v_0 = 0\ \mathrm{m\,s^{-1}}$, $g = 9.81\ \mathrm{m\,s^{-2}}$, $m = 75$ kg and $\mu = 5.25\ \mathrm{kg\,s^{-1}}$, we can plot the particular solution over the time interval $[0, 30]$ using the Mathematica function **`Plot`**. For this, we can make use of the **`ReplaceAll`** function, denoted for brevity by `/.`, and the replacement rule of the previous output (`sol`).

In[31]:=
```
v0=0;
g=9.81;
m=75;
μ=5.25;
Plot[v[t]/.sol,{t,0,30}]
```

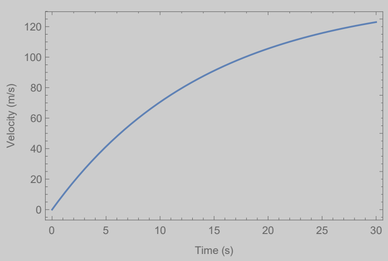
Out[31]=



For the sake of clarity, it is even better to use a frame and add frame labels. Note that the arrows in the Mathematica syntax are obtained by typing ->.

In[32]:=
```
Plot[v[t]/.sol,{t,0,30},Frame→True,FrameLabel→{"Time (s)","Velocity (m/s)"}]
```

Out[32]=



So far, we only encountered differential equations whose solution can be written explicitly as a function of the independent variable t. In many cases, however, and especially when dealing with non-linear differential equations, we can only obtain an **implicit solution** (*impliciete oplossing*) $S(t, y) = 0$. Consider, for instance, the following non-linear first-order differential equation, which does not look complicated at all.

Example 17.4

Determine and plot the solution of the differential equation

$$y\,y' - y + t = 0\,.$$

that satisfies $y(1) = 1$ using Mathematica.

Solution

The appropriate syntax for determining the general solution of this differential equation is:

```
In[33]:= sol2 = DSolve[{y[t]*D[y[t],t]-y[t]+t==0,y[1]==1},y[t],t]
```

Somehow surprisingly in the light of the simplicity of this differential equation, Mathematica returns only an implicit solution.

Out[33]= $\text{Solve}\left[\frac{\text{ArcTan}\left[\frac{-1+\frac{2y[t]}{t}}{\sqrt{3}}\right]}{\sqrt{3}}+\frac{1}{2}\text{Log}\left[1-\frac{y[t]}{t}+\frac{y[t]^2}{t^2}\right]==\frac{\pi}{6\sqrt{3}}-\text{Log}[t],y[t]\right]$

Such an implicit solution can be visualized using the Mathematica function **ContourPlot**. For this, we first have to extract the implicit function from the obtained solution, using the Mathematica function **First**, and replace the `y[t]` in the implicit solution by a variable. Then we simply copy and paste the result into **ContourPlot**.

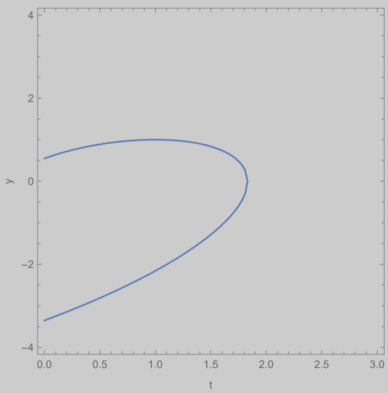
```
In[34]:= First[sol2]/.{y[t]→y}
```

Out[34]= $\frac{\text{ArcTan}\left[\frac{-1+\frac{2y}{t}}{\sqrt{3}}\right]}{\sqrt{3}}+\frac{1}{2}\text{Log}\left[1-\frac{y}{t}+\frac{y^2}{t^2}\right]==\frac{\pi}{6\sqrt{3}}-\text{Log}[t]$

In[35]:= $\text{ContourPlot}\left[\frac{\text{ArcTan}\left[\frac{-1+\frac{2y}{t}}{\sqrt{3}}\right]}{\sqrt{3}}+\frac{1}{2}\text{Log}\left[1-\frac{y}{t}+\frac{y^2}{t^2}\right]==\frac{\pi}{6\sqrt{3}}-\text{Log}[t],\right.$

```
{t,0,3},{y,-4,4},FrameLabel→{"t","y"}]
```

Out[35]=



17.2.2 Difference equations

Though difference equations are not the main focus of this course, they are closely related to differential equations, with the main difference being that the independent variable is discrete. Hence, they can be viewed either as a discrete analogue of differential equations. Much of the terminology that was introduced for differential equations can also be used for difference equations, but let us first of all state a formal definition.

Definitie 17.7 (Difference equation)
A **difference equation** (*differentievergelijking*) is an equation that contains unknown quantities

y_i, $i = 0, 1, \ldots,$

$$y_{n+k} = \mathcal{F}\left(y_{n+k-1}, y_{n+k-2}, \ldots, y_n\right), \tag{17.15}$$

where $\mathcal{F}$ is some function and k is the order of the difference equation.

Essentially, a difference equation recursively defines a sequence of values y_i, $i = 0, 1, \ldots$, once one or more initial terms are given: each further term of the sequence is defined as a function of the preceding terms. Hence, a difference equation is also sometimes referred to as a **recurrence relation** (*recursiebetrekking*).

Just as with differential equations, we can also define linear difference equations.

Definitie 17.8 (Linear difference equation)
An k-th order difference equation is called **linear** (*lineair*) if it can be written as

$$\mathcal{F}\left(y_{n+k-1}, y_{n+k-2}, \ldots, y_n\right) = a_k\, y_{n+k} + a_{k-1}\, y_{n+k-1} + \cdots + a_1\, y_{n+1} + a_0\, y_n + b = 0\,. \tag{17.16}$$

Otherwise, it is called **non-linear** (*niet-lineair*).

A linear difference equation like Equation (17.16) is **homogeneous** (*homogeen*) if $b = 0$. Otherwise, it is a **non-homogeneous** (*niet-homogeen*) difference equation. The Malthus model that describes exponential growth and is based on Equation (17.6) is an example of a linear first-order difference equation, whereas the more realistic Verhulst model is based on a non-linear first-order difference equation (Equation (17.7)).

Again similar to the terminology used in the framework of differential equations, one either arrives at the **general solution** (*algemene oplossing*) of the difference equation at stake, or one obtains a **particular solution** (*particuliere oplossing*) that satisfies some additional conditions. In general, k such conditions need to be specified in order to be able to find a particular solution of an k-th order difference equation.

Example 17.5

Let us try to find a general solution of Equation (17.6), i.e.

$$N_{k+1} = r\, N_k$$

using Mathematica.

In Mathematica, the function **`RSolve`** solves difference equations. However, since **`N`** is a built-in function of Mathematica, we replace N by y when implementing the difference equation, so:

```
In[36]:= RSolve[{y[k+1]==r*y[k]},y[k],k]

Out[36]= {{y[k]→r^(-1+k) C[1]}}
```

Or, supposing that the initial population size is given by N_0:

```
In[37]:= sol3=RSolve[{y[k+1]==r*y[k],y[0]==y0},y[k],k]

Out[37]= {{y[k]→r^k y0}}
```

So, we get

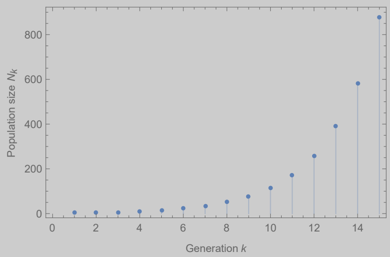
$$N_k = r^k\, N_0\,,$$

which agrees with what we inferred intuitively in Section 17.1. Finally, using this solution we can easily plot the population size up to generation 15. For that purpose, we first use the Mathematica function **`Table`** to generate a list containing the population sizes of the 15 first generations and then the function **`ListPlot`**, which is suited to plot discrete data points. Note that, in order to make **`Table`** work properly, we have to extract the replacement rule from the solution list, again using **`First`**.

In[38]:=
```
r=1.5;
y0=2;
ListPlot[Table[y[k]/.First[sol2],{k,1,15}], Filling→Axis,Frame→True,

FrameLabel→{"Generation k","Population size Nk"}]
```

Out[38]=



17.3 Exercises

Assignment 17.1 — Determine the order and degree of the following differential equations..

(a) $dy + \big(ty - \cos(t)\big)\, dt = 0$

(b) $\big(y'''\big)^2 + ty'' + 2y\big(y'\big)^3 + ty = 0$

(c) $y''\, y' + t\, y'^3 + y = 0$

(d) $y' + t = \big(y - ty'\big)^{-3}$

Assignment 17.2 — Show that following families of curves can be written with one constant.

(a) $y = C_1\, e^{t+C_2}$

(b) $y = C_1 + \ln|C_2 t|$

Assignment 17.3 — Show that the function $y(t)$ is a solution of the given differential equation. Then find the particular solution based on the initial conditions.

	Differentiaalvergelijking	Oplossing	Beginvoorwaarde(n)
(a)	$y + y' = t^3 + 3t^2$	$y(t) = t^3 + Ce^{-t}$	$y(0) = 3$
(b)	$ty' - y = t^2 e^t$	$y(t) = Ct + te^t$	$y(1) = e$
(c)	$y - ty' - y'^2 = 0$	$y(t) = Ct + C^2$	$y(2) = -1$
(d)	$y''' + y'' = 0$	$y(t) = C_1 e^{-t} + C_2 t + C_3$	$y(0) = 2$, $y'(0) = 1$, $y''(0) = 3$
(e)	$y' - y = 2(1-t)$	$y(t) = 2t + Ce^t$	$y(0) = 3$

Assignment 17.4 — Write the differential equations that describe the problems below.

(a) It has been empirically established that the rate at which the stomach contents of a predatory fish decreases is directly proportional to the square root of this volume.

(b) During a chemical reaction, substance A is converted to substance B at a rate that is directly proportional to the square of the amount of substance A. Call $A(t)$ the amount of unreacted substance A at time t.

(c) Newton's law of cooling states that the temperature change of an object is directly proportional to the difference between the (variable) temperature $T(t)$ of the object itself and the (constant) temperature R of the environment.

(d) When a patient is infused with a constant amount of painkiller per hour, the body breaks down this drug at a rate that is directly proportional to the amount present. We call that amount $M(t)$, the constant supply a and the degradation rate v (with $0 < v < 1$), the relative amount of degraded substance per unit time.

Assignment 17.5 — Find and plot the solution of the following differential equations using Mathematica.

(a) $y' = 1 + 2ty$, if $y(2) = -1$

(b) $y' + 2ty = 4t$, if $y(0) = 5$

(c) $(2 + \sin(y))y' + t = 0$, if $y(2) = 0$

(d) $x^2\, y\, y' - e^y = 0$, if $y\,(1) = 0$

(e) $y'' + 2y' + 5y = 0, \quad$ if $\; y(0) = 2 \;$ and $\; y'(0) = 2$

(f) $ty'' + \big(\cos(t)\big)y' + t^2 y = t, \quad$ if $\; y(-1) = -1 \;$ and $\; y'(-1) = 2$

(g) $y^{(4)} - 4y''' - 5y'' + 36y' - 36\,y = 0, \quad$ if $\; y(0) = 1,\; y'(0) = 2,\; y''(0) = 3$ and $y'''(0) = 4$

Assignment 17.6 — Consider Verhulst's logistic population model (Equation (17.7), Section 1.1.2):

$$N_{k+1} = r\left(1 - \frac{N_k}{K}\right)N_k\,,$$

with r the growth factor [–], K the carrying capacity of the environment [–], en k [–] the number of generations. N_0 [–] represents the initial number of individuals.

(a) Write a function `verhulst` with inputs r, K, $N0$, and k an which calculates the amount of individuals N_k after each generation. The outputs of your function are a vector N with $k+1$ elements ($N0$ as the first element), and a vector T. The last one consists of the numbers 0 until k.

(b) Assume $r = 1.5$, $K = 200$, $N_0 = 2$, $k = 25$, and calculate the population size N_k after each generation.

(c) Make a figure in which you plot N as a function of T. You should get Figure 17.5.

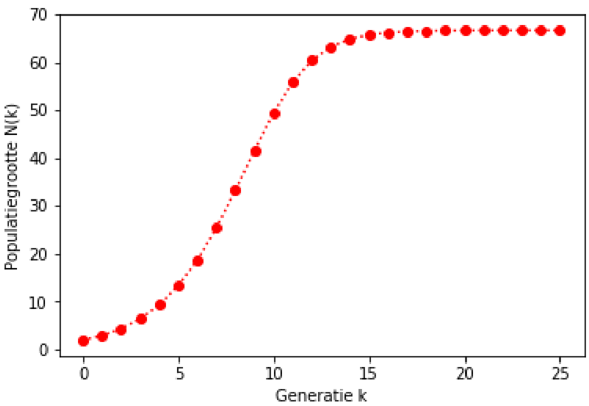


Figure 17.5: Evolution of the population size N as a function of the generation k.

To what value does the number of individuals converge? N?

(d) Discuss the influence of N_0 on the final number of individuals by repeating the calculations with $N_0 = 5$ and $N_0 = 25$.

(e) Assume $K = 200$, $N_0 = 2$, $k = 25$, and let r variate from 0.5 to 3.5 in steps of 0.5. Determine for each r a plot of N as a function of K. Describe what you see.

Assignment 17.7 — Consider the Verhulst model from the previous exercise. Suppose the individuals are harvested at a rate h [T^{-1}] and the number of individuals harvested is proportional to the number of individuals:

$$N_{k+1} = r\left(1 - \frac{N_k}{K}\right)N_k - h\,N_k\,.$$

Assume $h = 0.125$ s^{-1}.

(a) Write a function `verhulst2` with inputs r, K, $N0$, and k and which calculates the number of individuals N_k after each generation. The outputs of your function are a vector N with $k+1$ elements ($N0$ as the first element), and a vector T. The latter contains the numbers 0 to k.

(b) Assume $r = 1.5$, $K = 200$, $N_0 = 2$, $k = 25$, and calculate the population size N_k after each generation.

(c) Make a figure in which you plot N as a function of T. You should get Figure 17.6.

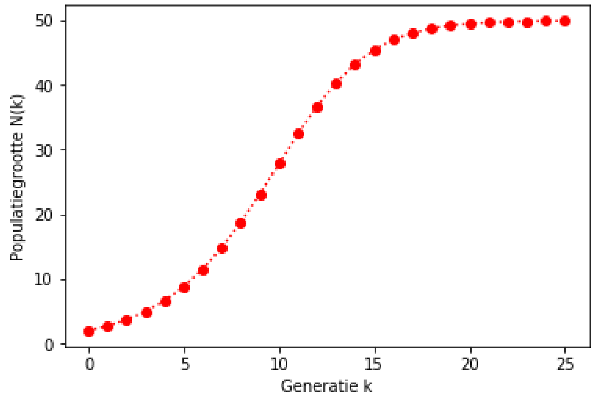


Figure 17.6: Evolution of the population size N as a function of the generation k.

To what value does the number of individuals N converge?

(d) Discuss the influence of the parameter h on the final number of individuals by repeating the calculations with $h = 0.25\ \text{s}^{-1}$ and $h = 0.75\ \text{s}^{-1}$.

(e) Assume $K = 200$, $N_0 = 2$, $k = 25$,and let r vary from 0.5 to 3.5 in 0.5 increments. For each r, make a plot of N as a function of K. Describe what you see.

Many who have had an opportunity of knowing any more about mathematics confuse it with arithmetic, and consider it an arid science. In reality, however, it is a science which requires a great amount of imagination.

— Sofia Kovalevskaya —

18

Qualitative analysis of first-order differential equations

In this chapter we will focus on a qualitative analysis of first-order differential equations that can be written as

$$y' = f(t, y) \ . \tag{18.1}$$

This means that we will not yet try to solve such equations analytically, but rather that we will resort to graphical or numerical methods to get some idea of how the solutions of the given equation behave. This is motivated because it is often impossible to find explicit formulas for solutions of differential equations. Moreover, even if there are such formulas, they may be so complicated that they are useless. Quantitative methods of solutions for first-order differential equations will be presented in Chapter 19.

18.1 Direction and vector fields

Let us assume that Equation (18.1) has solutions. Recall that a solution of Equation (18.1) is a function $y = y(t)$ such that

$$y'(t) = f(t, y(t))$$

for all values of t in some open interval $]\,a, b\,[$. Though we cannot yet solve such a differential equation analytically, we can get some insight into the shape of its solution curves by realizing that Equation (18.1) represents the derivative of $y(t)$ at any point (t, y) in the (t, y)-plane, provided f is continuous everywhere on this plane. This means that Equation (18.1) gives us a means to assess the slope of the solution curve in any point (t, y) in the (t, y)-plane, or put in other words, it yields the slope of the tangent line to the solution curve passing through (t, y) for any (t, y) in the (t, y)-plane.

In practice, we evaluate the right-hand side of Equation (18.1) in a finite number of equidistantly spaced points (t_i, y_i) in some rectangular region R in the (t, y)-plane, and we draw short line segments of equal length through these points with slope $f(t_i, y_i)$. Moreover, we add an arrowhead to these segments pointing in the direction of increasing t as we move in that direction along the solution curves with increasing t. The plot we obtain in this way is referred to as the **direction field** (*richtingsveld*) or **slope field** of the differential equation. Since we endowed the line segments with a direction, we may as well regard them as vectors connecting the points (t_i, y_i) and $(t_i + \Delta t, y_i + \Delta y_i) = \left(t_i + \Delta t, y_i + \Delta t f(t_i, y_i)\right)$. In that way, we do not only visualize the direction but also magnitude of the change at a given set of points. Such a plot is typically referred to as the **vector field** (*vectorveld*) of a differential equation.

Clearly, it is impracticable to construct such a direction field by hand, but fortunately this functionality is available in both Mathematica and Python. Note that the initial condition is not needed to construct the direction field.

Example 18.1

Draw the direction field of Equation (17.4), i.e.

$$m \frac{dv}{dt} = m g - \mu v,$$

using Mathematica for $g = 9.81\ \mathrm{m\,s^{-2}}$, $m = 75$ kg and $\mu = 5.25\ \mathrm{kg\,s^{-1}}$.

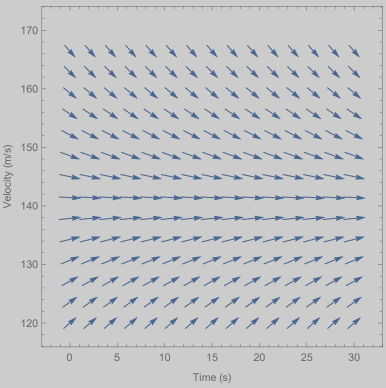
Solution

In Mathematica, this can be achieved using the function `VectorPlot`:

```
In[41]:= g=9.81;
         m=75;
         mu=5.25;
         df=VectorPlot[{1,g-mu*v/m},{t,0,30},{v,120,170},VectorScale→{Tiny,Small,None},

         FrameLabel→{"Time (s)", "Velocity (m/s)"}]
```

Out[41]=

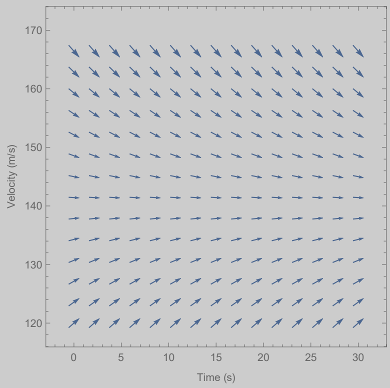


The option `VectorScale` is used to tune the layout of the arrows in the plot. Its third argument (**`None`**) dictates that the size of the line segments does not depend on the magnitude of the derivative. In other words, all plotted arrows should have the same size and we obtain the direction field. On the other hand, the vector field of the differential equation at stake can be plotted by using **`Automatic`** as third argument of the option `VectorScale` as follows

```
In[42]:= g=9.81;
         m=75;
         mu=5.25;
         df=VectorPlot[{1,g-mu*v/m},{t,0,30},{v,120,170},VectorScale→{Tiny,Small,Automatic},

         FrameLabel→{"Time (s)", "Velocity (m/s)"}]
```

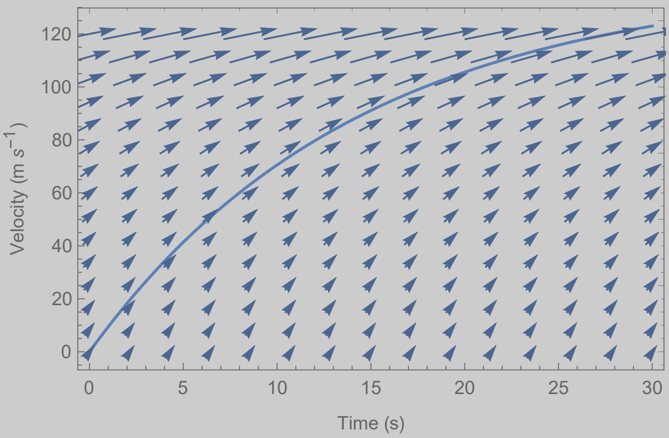
Out[42]=



For completeness, let us now add the particular solution of Equation (17.4) satisfying $v(0) = 0$ to the same plot. In Mathematica, plots can be combined using the function **Show**:

```
In[43]:= v0=0;
         sol=DSolve[{m*D[v[t],t]==m*g-mu*v[t],v[0]==v0},v[t],t];
         df=VectorPlot[{1,g-mu*v/m},{t,0,30},{v,0,120},VectorScale →{Tiny,Tiny,None},
         FrameLabel→{"Time (s)","Velocity (m/s)"}];
         solplot=Plot[v[t]/.sol,{t,0,30},Frame→True,FrameLabel→{"Time (s)","Velocity (m s^-1)"}
         Show[solplot,df]
```

Out[43]=



From this example, it is clear that we can get a fairly accurate idea of the shape of the solution curve of a first-order differential equation given an initial condition. Essentially, it seems that we can qualitatively construct the solution curve passing through a point (t_0, y_0) by following the direction indicated by the directed line segments in the direction field. This becomes even clearer if we plot the solution curves of Equation (17.4) for $g = 9.81$ m s^{-2}, $m = 75$ kg, $\mu = 5.28$ kg s^{-1} and an initial velocity v_0 varying between 0 and 240 m s^{-1} on top of the direction field of the governing differential equation (Figure 18.1). Irrespective of the initial velocity, the direction allows us to gain qualitative insight into the shape of the solution curve.

The same conclusion can be drawn by inspecting an overlay of the solution curves of Equation (17.14), whose interval of existence we determined in Example 17.2, and the corresponding direction field

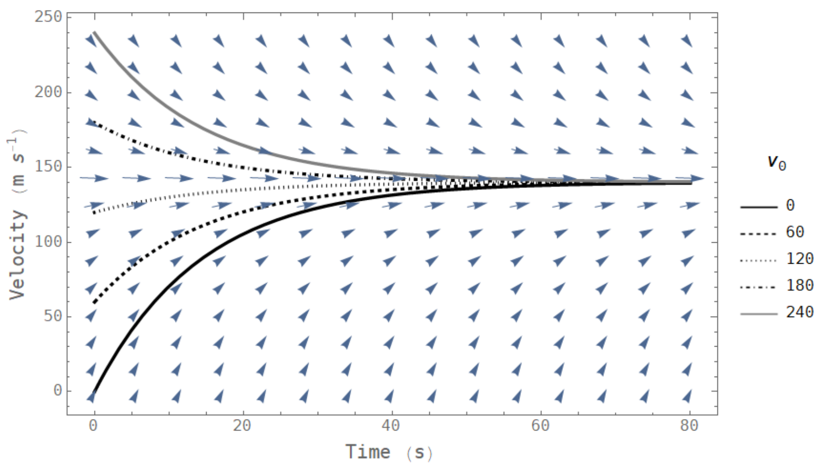


Figure 18.1: Solutions of Equation (17.4) for $g = 9.81\ \mathrm{m\,s^{-2}}$, $m = 75$ kg , $\mu = 5.28\ \mathrm{kg\,s^{-1}}$ and initial velocity v_0 varying between 0 and 240 $\mathrm{m\,s^{-1}}$ superimposed on the direction field of Equation (17.4).

(Figure 18.2).

18.2 Equilibria and stability

18.2.1 Equilibrium points

A peculiarity about the solution curves of Equation (17.4) is that they seem to converge to the same constant value irrespective of the initial velocity. Besides, in Figure 18.1 we can also see that the line segments near that value are oriented more or less horizontally, which implies locally that

$$m\,v' = m\,g - \mu\,v \approx 0\,.$$

Clearly, there exists a velocity v_e for which it holds that $v' = 0$. It follows from

$$m\,v' = m\,g - \mu\,v_e = 0\,,$$

which yields

$$v_e = \frac{m\,g}{\mu}\,.$$

For the situation considered in Figure 18.1, we have $g = 9.81\ \mathrm{m\,s^{-2}}$, $m = 75$ kg , $\mu = 5.28\ \mathrm{m\,s^{-1}}$, so that we find that $v_e \approx 139\ \mathrm{kg\,s^{-1}}$, which corresponds to the terminal velocity the solution curves in Figure 18.1 are converging to. Once this velocity is reached, any further change of the velocity over time is impossible because $v' = 0$ in v_e. In general, we will refer to such points as equilibrium points.

Definitie 18.1 (Equilibrium point)
An **equilibrium point** (*evenwichtspunt*) of the first-order differential equation

$$y' = f\,(t, y)$$

is any point y_e for which it holds that $f\,(t, y_e) = 0$ for all $t \geq \tau$, with τ arbitrary.

This definition also indicates that $y(t) = y_e$ is a solution of the governing differential equation for all $t \geq \tau$. This solution is commonly known as an **equilibrium solution** (*evenwichtsoplossing*) of

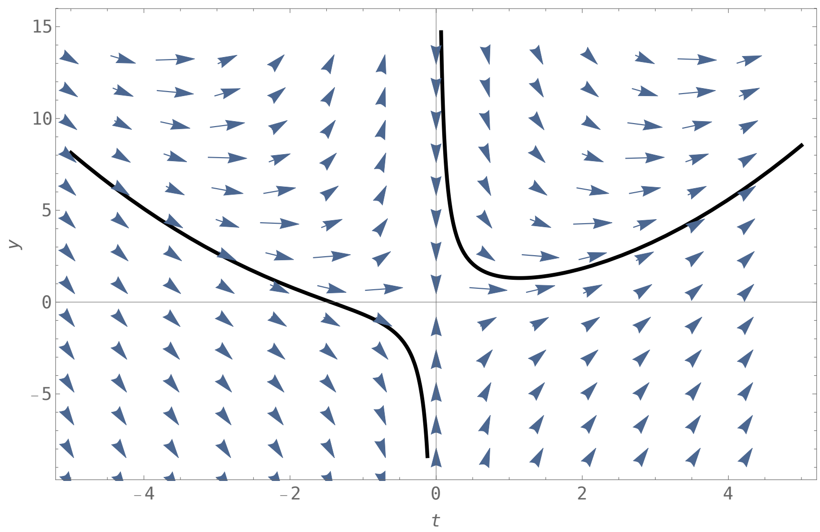


Figure 18.2: Solution curves of Equation (17.14) imposed on its direction field.

the differential equation. For an autonomous differential equation $y' = f(y)$ like Equation (17.4), the equilibrium points can be retrieved easily be solving the equation $f(y_e) = 0$. For most non-autonomous differential equations, however, there exist no equilibrium points. Indeed, in Figure 18.2 visualizing the direction field of the non-autonomous differential equation (17.14) no such equilibrium points can be observed. Still, there are specific non-autonomous differential equations for which one or more equilibrium points do exist.

18.2.2 Stability

In the case of our free-falling object, it is clear that all solutions will in the end converge to the equilibrium point v_e (Figure 18.1). So, we may say that this equilibrium point attracts solutions in the neighbourhood of v_e, and hence acts as an **attractor** (*aantrekker*). It is, however, not always that an equilibrium point attracts nearby solutions. Let us consider the following example to clarify this.

Example 18.2

Recall the Verhulst model we derived in Section 17.1 (Equation (17.7)), and which is based on a difference equation. Under the same assumptions, we can also derive its counterpart in continuous time. It is given by

$$P' = r\,P\left(1 - \frac{P}{K}\right), \tag{18.2}$$

with $P(t)$ the population size at time t, r [T^{-1}] the growth rate in the absence of resource limitation, and K [–] the carrying capacity of the environment.

The equilibrium points of Equation (18.2) follow from solving $P' = 0$ for P_e

$$r\,P_e\left(1 - \frac{P_e}{K}\right) = 0\,.$$

Immediately, we get the two equilibrium points, namely $P_e = 0$ and $P_e = K$. The obvious question now is towards which one, if any, of these two points the solution curves possibly converge for growing t. For that purpose, let us consider the direction field of Equation (18.2) for $K = 200$ and $r = 0.15$ month^{-1} depicted in Figure 18.3. Both equilibrium points can be located easily as the line segments near $P_e = 0$ and $P_e = 200$ are oriented almost horizontally. Suppose that we start from a strictly positive population size $N_0 < K$, then we may anticipate that the solution curves will converge to the carrying capacity K by studying the direction field of this differential equation. Likewise, we may anticipate that solution curves starting in points $N_0 > K$ will converge to K. This is confirmed when we look at the general solution of Equation (18.2), which is given by

$$P(t) = \frac{K\,e^{C_1 K + rt}}{e^{C_1 K + rt} - 1}\,,$$

and whose limit for $t \to +\infty$ is K.

Essentially, the fact that the solution curves converge towards $P_e = K$ does not come as a complete surprise because it is one of the equilibrium points, but $P_e = 0$ is an equilibrium point as well and does not seem to attract any nearby solutions at all. Rather, it seems to be repelling them. For that reason, we will refer to a repelling equilibrium point like $P_e = 0$ as an unstable equilibrium point, whereas $P_e = K$ will be referred to as an asymptotically stable equilibrium point. Besides, there exist also stable equilibrium that neither repel nor attract the solution curves.

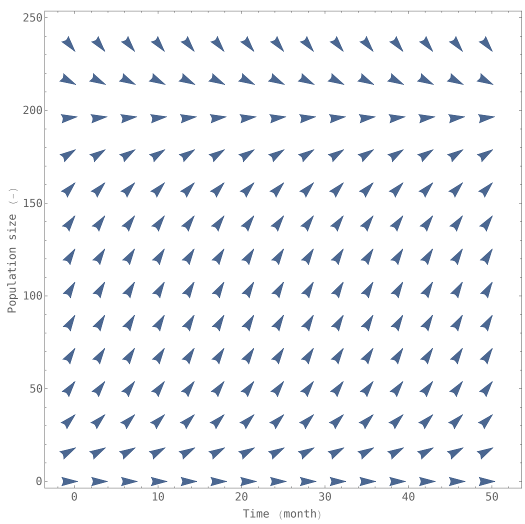


Figure 18.3: Direction field of Equation (18.2) underlying the continuous-time Verhulst model.

Below we give more rigorous definitions of stable and asymptotically stable equilibrium points.

Definitie 18.2 (Stable equilibrium point)
An equilibrium point y_e of the first-order differential equation

$$y' = f(t, y)$$

is called **stable** (*stabiel*) if, given $\epsilon > 0$, there is a $\delta > 0$ such that $\left|y_{(t_0,y_0)}(t) - y_e\right| < \epsilon$ for all $t > 0$ and for all y_0 such that $\left|y_0 - y_e\right| < \delta$.

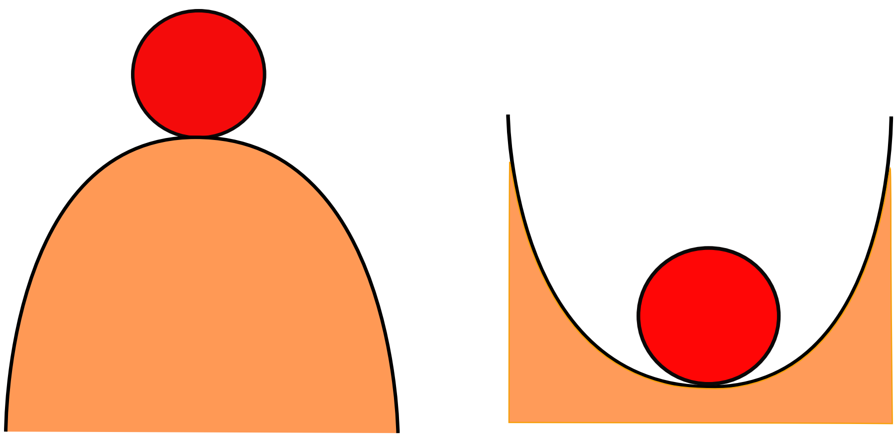
Definitie 18.3 (Asymptotically stable equilibrium point)
An equilibrium point y_e of the first-order differential equation

$$y' = f(t, y)$$

is called **asymptotically stable** (*asymptotisch stabiel*) if it is stable and if $y_{(t_0,y_0)}(t) \to y_e$ as $t \to +\infty$.

An equilibrium point that is neither stable nor asymptotically stable is called **unstable** (*onstabiel*). In the case of the free-falling object, it is clear that $v_e = \frac{mg}{\mu}$ is an asymptotically stable equilibrium point as the directed line segments in the corresponding direction field (Figure 18.1) point towards this velocity.

Stability of an equilibrium point can be illustrated as well intuitively by imagining a ball at rest on a hill or in a valley (Figure 18.4). When the ball is at rest on top of a hill, the slightest push will cause the ball to roll away (Figure 18.4(a)), which represents the behaviour of solutions near an unstable equilibrium point. Conversely, a ball in a valley will, when pushed, return to its resting position (Figure 18.4(b)).



(a) (b)

Figure 18.4: Visual representation of an unstable (a) and stable (b) equilibrium.

Recall that the right-hand side $f(t^*, y^*)$ of a differential equation in a point (t^*, y^*) tells us whether the solution curve described by $y(t)$ and passing through (t^*, y^*) is increasing ($f(t^*, y^*) > 0$) or decreasing ($f(t^*, y^*) < 0$) in that point. In the case of an autonomous differential equation $y' = f(y)$, this is even completely determined by y^* only because t does not appear explicitly in its right-hand side. Hence, given y^* one can verify whether the solution through that point will evolve towards values greater or smaller than y^*, and consequently whether the solution curve will converge to some equilibrium point, remains close to it, or is repelled by it. So, by studying the graph of $f(y)$ it should be possible to infer the stability of the possible equilibrium points of $y' = f(y)$. For the sake of clarity, we adopt the convention that asymptotically stable equilibrium points on such graphs are marked with filled circles and unstable equilibrium points with open circles.

Example 18.3

Consider the graphs of the functions corresponding to the right-hand sides of Equations (17.4) and (18.2) in Figure 18.5(a) and 18.5(b), respectively. In these figures the red arrows indicate the direction to which solutions evolve over time.

From the former we may conclude that v' is positive for all velocities $v < v_e$, while it is negative for all $v > v_e$. This implies that $v_{(t_0,v_0)}(t)$ increases monotonically up to the point where the solution curve reaches the terminal velocity v_e for all $v_0 < v_e$, whereas the opposite is true if $v_0 > v_e$. Essentially, this corresponds to our physical understanding of this process: a free-falling object will accelerate up to a certain terminal velocity when starting its fall with an initial velocity that was lower than the terminal one, whereas it will decelerate when it started its fall at $v_0 > v_e$. Once the terminal velocity is reached, there is an equilibrium between the gravitational and frictional forces.

A similar reasoning allows us to understand that $P_e = K$ is an asymptotically stable equilibrium point of Equation (18.2), whereas $P_e = 0$ is an unstable equilibrium point. Indeed, $P' > 0$ for all $P < K$, so the population size will gradually increase when it was initially lower than the carrying capacity K, and vice versa when the population was initially greater than the carrying capacity K. Again, this corresponds to our intuition of the process. Note by the way that the function $f(P)$ is decreasing in $P_e = K$, whereas it is increasing in $P_e = 0$. This intuitive insight into the link between the stability of an equilibrium point and the derivative of the right-hand side of the differential equation at the equilibrium point is formalized in Theorem 18.1.

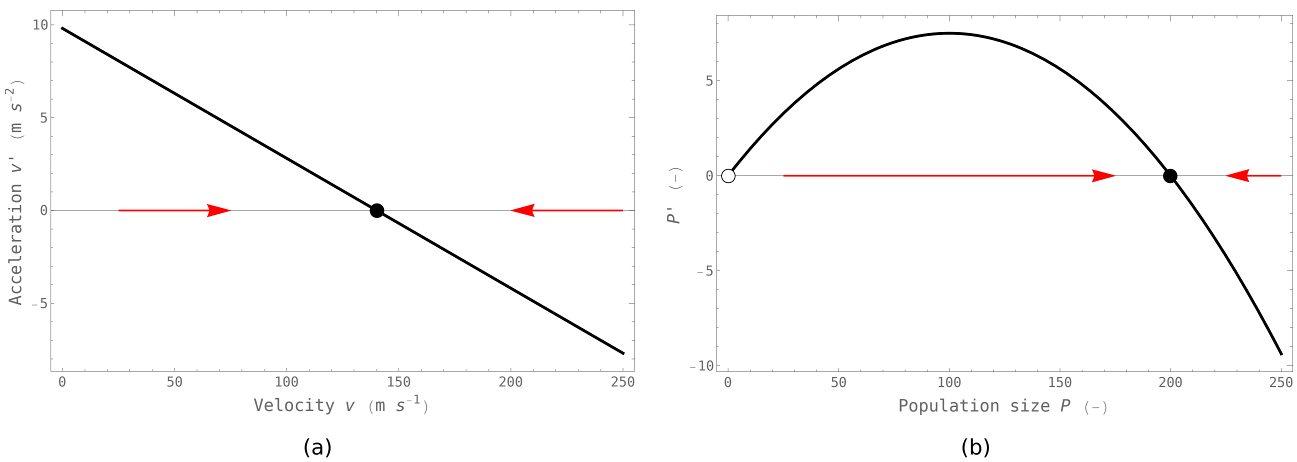


Figure 18.5: Graphs of the functions corresponding to the right-hand sides of Equations (17.4) (a) and (18.2) (b), with their asymptotically stable and unstable equilibrium points (filled and open circles, respectively) and the direction to which the solutions evolve over time (red arrows).

Theorem 18.1 (First derivative test)
Suppose that y_e is an equilibrium point of the autonomous differential equation

$$y' = f(y) \,,$$

where f is a differentiable function.

1. *If $f'(y_e) < 0$, then y_e is asymptotically stable.*
2. *If $f'(y_e) > 0$, then y_e is unstable.*
3. *If $f'(y_e) = 0$, then no conclusion can be drawn solely on the basis of $f'(y_e)$.*

Using this theorem one can in most cases easily determine the stability of the equilibrium point of an autonomous differential equation. Yet, if $f'(y_e) = 0$ different situations are possible, one of which is illustrated in Figure 18.6. Since $y' = 0$ for all y in the interval $[2,4]$, there are infinitely many equilibrium points, namely $[2,4]$. For all of them, it holds that $f'(y_e) = 0$, so Theorem 18.1 is of no help to determine their stability. Yet, since y' is positive for all $y < 2$, it is obvious that for any $y_0 < 2$ we will have $y_{(t_0,y_0)}(t)$ converges to 2. Likewise, for any $y_0 > 4$ we will have $y_{(t_0,y_0)}(t)$ converges to 4. On the other hand, solution curves originating from a point y_0 lying in $[2,4]$ will stay at that point, and will not be attracted by other points. Hence, these solutions will be given by $y(t) = y_0$ because $y' = 0$ in any of these points. Consequently, in the light of Definition 18.2, the infinitely many equilibrium points in $[2,4]$ are stable. Sometimes the points at the boundaries of the interval $[2,4]$, i.e. $y_e = 2$ and $y_e = 4$, are referred to as **semi-asymptotically stable equilibrium points** (*semi-asymptotisch evenwichtspunt*) because solution curves originating from points to left of the former converge to $y_e = 2$, whereas there is no such convergence for solution curves originating from y_0 in $[2,4]$, and likewise for $y_e = 4$. Moreover, in cases where solutions on one side of an equilibrium point move towards the equilibrium point and on the other side of the equilibrium point move away from it, we call the equilibrium point **semi-stable** (*semi-stabiel*). Such equilibrium points will be indicated by means of triangular markers.

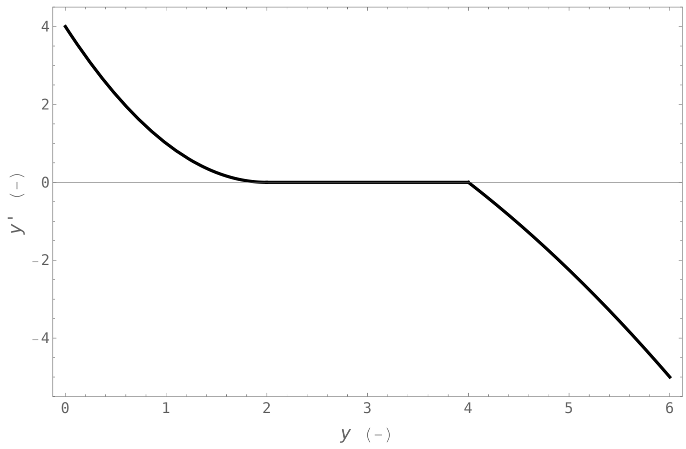


Figure 18.6: Graph of $f(y)$ for an autonomous differential equation $y' = f(y)$, with $f'(y_e) = 0$.

Example 18.4

Determine the equilibrium points of the following differential equation

$$y' = y^3 - 2\,y^2 - y + 2\,, \tag{18.3}$$

as well as their stability.

Solution

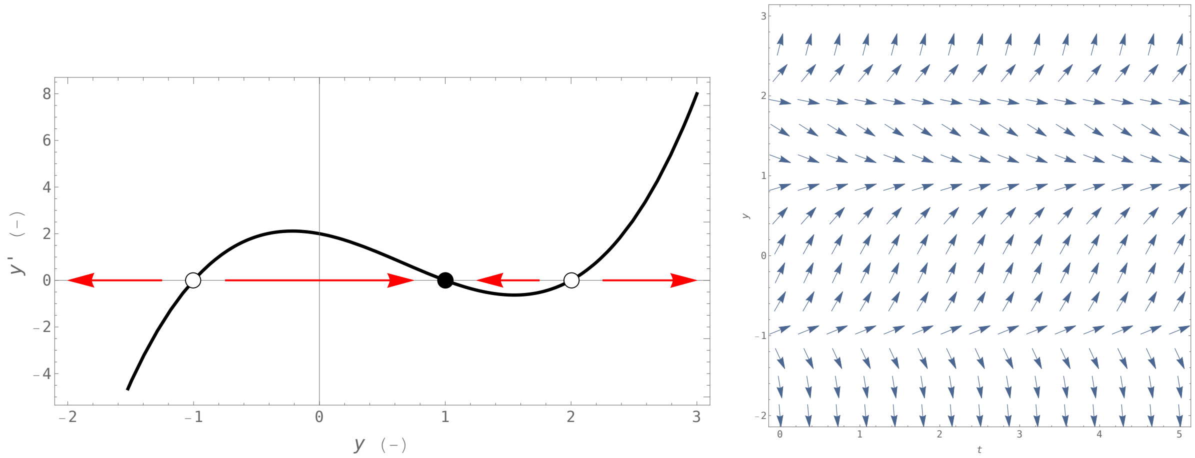
The equilibrium points of Equation (18.3) follow from:

$$\begin{aligned} 0 &= y_e^3 - 2\,y_e^2 - y_e + 2\,,\\ &= \left(y_e^2 - 1\right)(y_e - 2)\,. \end{aligned}$$

So, the equilibrium points of Equation (18.3) are $y_e = -1$, $y_e = 1$ and $y_e = 2$. To determine their stability, we consider the first-order derivative of the right-hand side of Equation (18.3) with respect to y:

$$y'' = f'(y) = 3\,y^2 - 4\,y - 1\,.$$

Evaluating $f'(y)$ in each of the equilibrium points, yields $f'(-1) = 6$, $f'(1) = -2$ and $f'(2) = 3$, so by relying on Theorem 18.1, we may conclude that both $y_e = -1$ and $y_e = 2$ are unstable equilibrium points, whereas $y_e = 1$ is an asymptotically stable equilibrium point. This is confirmed by Figure 18.7 that shows both the graph of $f(y)$ and the direction field of Equation (18.3).



(a) (b)

Figure 18.7: Graph of the right-hand side (a) and direction field (b) of Equation (18.3). On the former the asymptotically stable equilibrium points are marked with filled circles and unstable equilibrium points with open circles. The direction to which the solutions evolve over time is indicated by red arrows.

18.3 Bifurcation

In Example 18.2 we already observed that the value of an equilibrium point might depend on a model parameter; that is the carrying capacity K in the case of the Verhulst model. Moreover, we encountered an example where the specific value depends even on more than one parameter. Indeed, the equilibrium velocity of a free-falling object that is subject to gravitational and frictional forces is given by $v_e = \frac{m\,g}{\mu}$, which depends on three parameters.

Not only can model parameters affect the value of an equilibrium point, they might as well have a decisive impact on its stability and even affect the number of equilibrium points. For instance, it might happen than for a certain range of parameter values an equilibrium point y_e is asymptotically stable, whilst it is unstable for parameter values outside this range, or vice versa. This phenomenon is known as **bifurcation** (*bifurcatie*) and plays an important role in the analysis of many mathematical models. A parameter p that gives rise to bifurcation is referred to as a **bifurcation parameter** (*bifurcatiepa-rameter*). The effect of a bifurcation parameter on the stability and/or number of equilibrium points is typically summarized in a so-called **bifurcation diagram** (*bifurcatiediagram*), which visualizes the equilibrium points and their stability as a function the bifurcation parameter p. For what concerns the bifurcation diagram, we adopt the convention that asymptotically equilibrium points will be indicated by means of a solid line, while a dashed line will be used to indicate unstable equilibrium points.

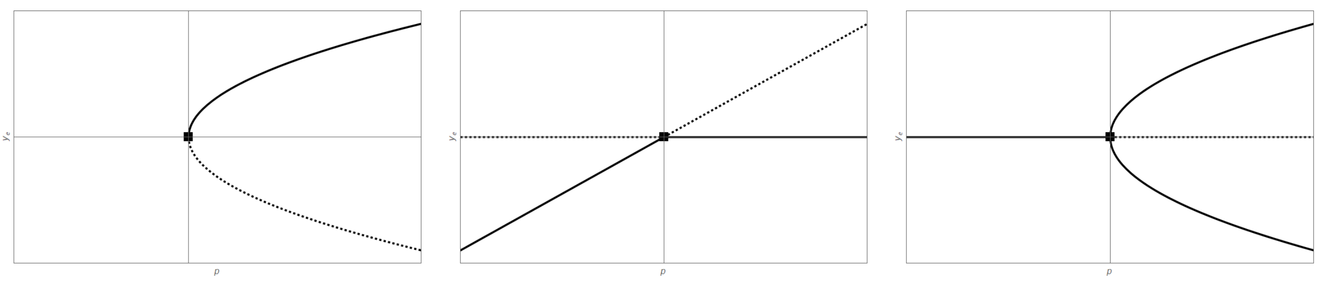
Definitie 18.4 (Bifurcation point)
A **bifurcation point** (*bifurcatiepunt*) of the autonomous differential equation

$$y' = f_p(y) \ ,$$

where p is a real bifurcation parameter, is the parameter value p^* where there is a change in the number and/or stability of equilibrium points.

Depending on the transition that happens when crossing from one side of a bifurcation point to the other side along the parameter axis, we distinguish three main types of bifurcation when there is one parameter, which are illustrated in Figure 18.8.

- **Saddle-node bifurcation** (*zadel-knoop bifurcatie*): this manifests itself when there is one equilibrium point in the bifurcation point p^*, two at one side of p^* and none at the other side (Figure 18.8(a)).
- **Transcritical bifurcation** (*transkritische bifurcatie*): this happens when there are always two equilibrium points, irrespective of the value of the bifurcation parameter, but these points exchange their stability when we move across the bifurcation point p^* (Figure 18.8(b)).
- **Pitchfork bifurcation** (*hooivorkbifurcatie*): this kind of bifurcation can be observed when there is one equilibrium point at one side of the bifurcation point p^*, while there are three at the other side (Figure 18.8(c)).



(a) (b) (c)

Figure 18.8: Possible bifurcations: saddle-node (a), transcritical (b) and pitchfork (c). The bifurcation point is indicated by means of a square marker.

Example 18.5

Consider the autonomous differential equation

$$y' = p + y^2\,, \tag{18.4}$$

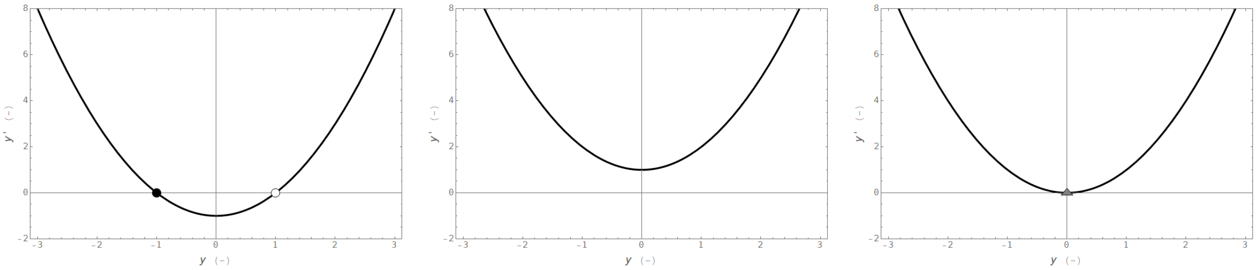
where p is a real parameter. Determine the equilibrium point(s) of this differential equation and its (their) stability as a function of the parameter p.

Solution

The equilibrium point(s) of Equation (18.4) is (are) given by

$$y_e = \pm\sqrt{-p}\,.$$

From this we already infer that there are two equilibrium points if $p < 0$, none if p is positive and one if $p = 0$, so $p = 0$ is the bifurcation point, and we are faced with a saddle-node bifurcation. Figure 18.9 shows a typical plot of the right-side of Equation (18.4) for each of the three parameter settings. From Figure 18.9(a), we may conclude that for $p < 0$ the equilibrium points $y_e = +\sqrt{-p}$ are unstable, while the equilibrium points $y_e = -\sqrt{-p}$ are asymptotically stable. Figure 18.9(b) confirms that there are no equilibrium points if $p > 0$, while Figure 18.9(c) demonstrates that $y_e = 0$ is a semi-stable equilibrium point if $p = 0$. All this information is summarized in the bifurcation diagram depicted in Figure 18.10.



(a) $p = -1$ (b) $p = 1$ (c) $p = 0$

Figure 18.9: Plots of the right-hand side of Equation (18.4) for $p < 0$ (a), $p > 0$ (b) and $p = 0$ (c).

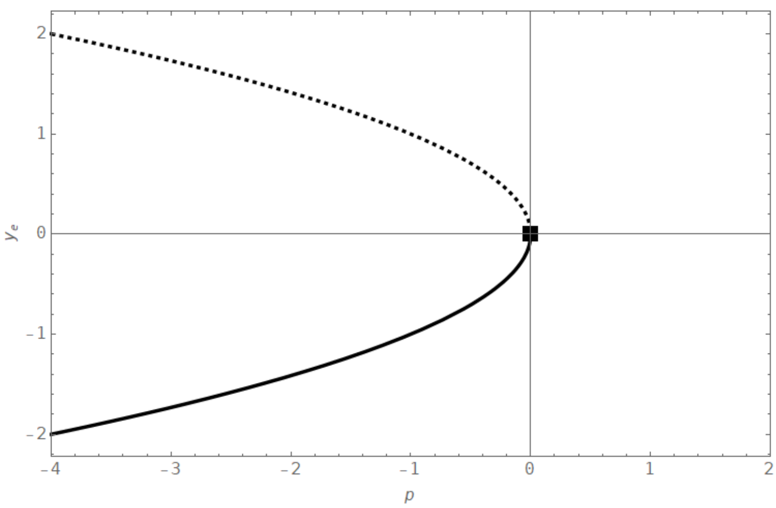


Figure 18.10: Bifurcation diagram of Equation (18.4).

To conclude, let us investigate whether bifurcation plays a role in a modified version of the Verhulst model that accounts for harvesting of the population.

Example 18.6

We can easily extend Equation (18.2) to account for the harvesting of individuals at a certain rate h [T^{-1}], assuming that the harvested number of individuals is proportional to the number of individuals in the population, i.e.

$$P' = r\,P\left(1 - \frac{P}{K}\right) - h\,P\,. \tag{18.5}$$

The equilibrium points of Equation (18.5) follow from solving

$$0 = r\,P_e\left(1 - \frac{P_e}{K}\right) - h\,P_e$$

for P_e, which yields $P_e = 0$ and $P_e = K\left(1 - \frac{h}{r}\right)$. As in Example 18.2, let us assume that $K = 200$ and $r = 0.15$ month^{-1}, so that the latter equilibrium point becomes

$$P_e = 200\left(1 - \frac{h}{0.15}\right).$$

Consequently, there are two equilibrium points irrespective of the harvest rate h. In order to verify whether the harvest rate h can affect the stability of these equilibrium, we compute the derivative of the right-hand side of Equation (18.5) with respect to P, which yields

$$f'(P) = r - \frac{2\,r\,P}{K} - h = \frac{3}{20} - \frac{3\,P}{2000} - h\,,$$

so that we can apply the first-derivative test (Theorem 18.1) at both equilibrium points. At $P_e = 0$ we get $f'(0) = 0.15 - h$, whereas we obtain

$$f'\left(200\left(1 - \frac{h}{0.15}\right)\right) = h - 0.15\,.$$

It is clear that the sign of these expressions depends on the harvest rate h. More precisely, we see that $P_e = 0$ is asymptotically stable for all $h > 0.15$, whereas it is unstable for $h < 0.15$, and the first-derivative test is inconclusive for $h = 0.15$. Likewise, we see that $P_e = 200\left(1 - \frac{h}{0.15}\right)$ is asymptotically stable for all $h < 0.15$, whereas it is unstable for $h > 0.15$, and the first-derivative test is inconclusive for $h = 0.15$. We may conclude that Equation (18.5) has a bifurcation point at $h = 0.15$, where the equilibrium points exchange their stability. Consequently, Equation (18.5) gives rise to a transcritical bifurcation. Figure 18.11 shows the corresponding bifurcation diagram. It is obvious that $P_e = 0$ is the only ecologically relevant equilibrium point if $h > 0.15$.

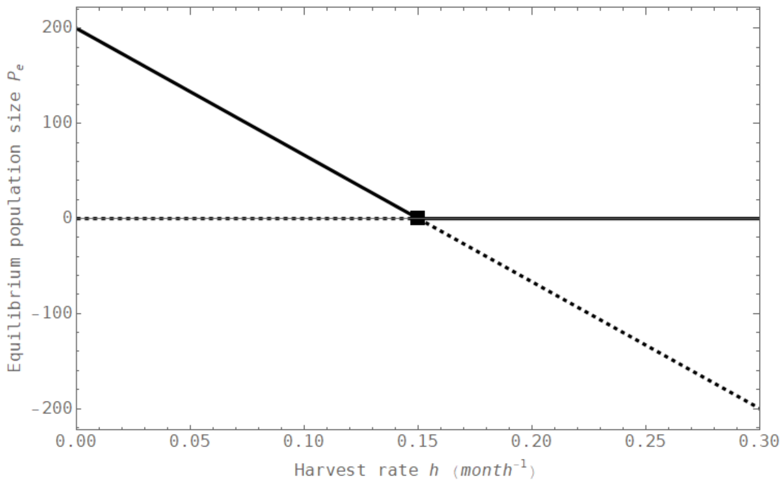


Figure 18.11: Bifurcation diagram of Equation (18.5).

The outcome of the analysis in Example 18.6 also corresponds with our ecological intuition: if individuals are harvested faster than they can reproduce the populations must go extinct in the long run. It is argued that such a tipping point might exist in many ecosystems, in the sense that its recovery becomes impossible if the disturbance rate/frequency exceeds a certain threshold. For that reason, it is important to be able to study the governing models from a qualitative point of view because one gets insight into the possible equilibrium states of ecosystems and the drivers that might bring it to a collapse.

18.4 Exercises

18.4.1 Direction and vector fields

Direction fields can also be plotted in Python. For this purpose the function `directionfield` is available in the script `directionfield.py`, the implementation of which is given below:

```
def richtingsveld(f, tinterval, yinterval, n = 25):
    tmin = tinterval[0]
    tmax = tinterval[1]
    ymin = yinterval[0]
    ymax = yinterval[1]
    T = numpy.linspace(tmin, tmax, n)
    y = numpy.linspace(ymin, ymax, n)
    #Vector field
    [X, Y] = numpy.meshgrid(T, y)
    U = 1
    V = f(T, Y)
    #Normalize arrows
    N = numpy.sqrt(U**2 + V**2)
    U = U/N
    V = V/N
    plt.quiver(X, Y, U, V)
    plt.xlim(tinterval)
    plt.ylim(yinterval)
    plt.xlabel('$t$')
    plt.ylabel('$y$')
    plt.show()
```

The function `directionfield` has the following inputs:

1. `f`:the right-hand side of the differential equation,
2. `tinterval`: the considered time interval,
3. `yinterval`: the considered y interval, and
4. n: precision of the direction field, default is $n = 25$.

To illustrate, consider the following differential equation describing the dynamics of a population of size P:

$$P' = r\,P\left(1 - \frac{P}{K}\right) - h\,P\,, \qquad P(0) = 5\,, \tag{18.6}$$

with h [T^{-1}] the rate at which the population is harvested (cf. Equation (18.5) from Example 18.6). We first implement the right-hand side in Python as the function `exampleRL` with $r = 1.5$, $K = 2$, and $h = 1$ s^{-1}:

```
def voorbeeldRL(t, P):
    r = 1.5
    K = 2
    h = 1
    ydot = r*P*(1 - P/K) - h*P
    return ydot
```

Note that, although the right-hand side does not explicitly depend on t, it is always given as the first input.

The corresponding direction field can now be easily plotted by entering the following instruction in the instruction window:

```
richtingsveld(voorbeeldRL, [0, 10], [0, 5], 25)
```

The result is shown in Figure 18.12.

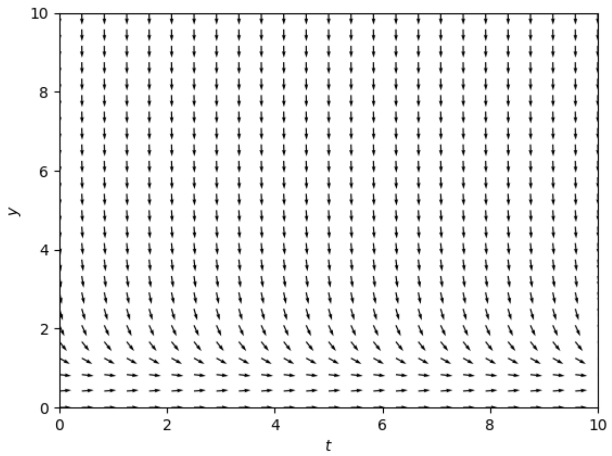


Figure 18.12: The direction field of Equation (18.6) with $r = 1.5$, $K = 2$, and $h = 1$.

Assignment 18.1 — Consider the first-order differential equation

$$y' = y\left(-2\,t + \frac{1}{t}\right) \tag{18.7}$$

over the t-interval $[0.2,\, 3]$, with initial condition $y\,(0.2) = 0.2$.

(a) Determine the exact solution of differential equation (18.7) using Mathematica.

(b) Plot the direction field of differential equation (18.7) over the time interval $[0.2, 3]$ and the y-interval $[0, 1]$. First, implement the right-hand side in Python. To do this, complete the script *function1.py*:

Listing 18.1: functie1

```
def functie1(t, y):
    ydot =
    return ydot
```

(c) Plot on top of the obtained direction field the exact solution you obtained with Mathematica.

Assignment 18.2 — Consider the following first-order differential equation over the t interval $[-2,\, 2]$:

$$y' = y + 3\,t - t^3.$$

(a) Implement the right-hand side of this differential equation in Python as `function2` in the script `function2.py`.

(b) Plot the direction field of this differential equation. Set the y-axis to $[-2,\, 2]$.

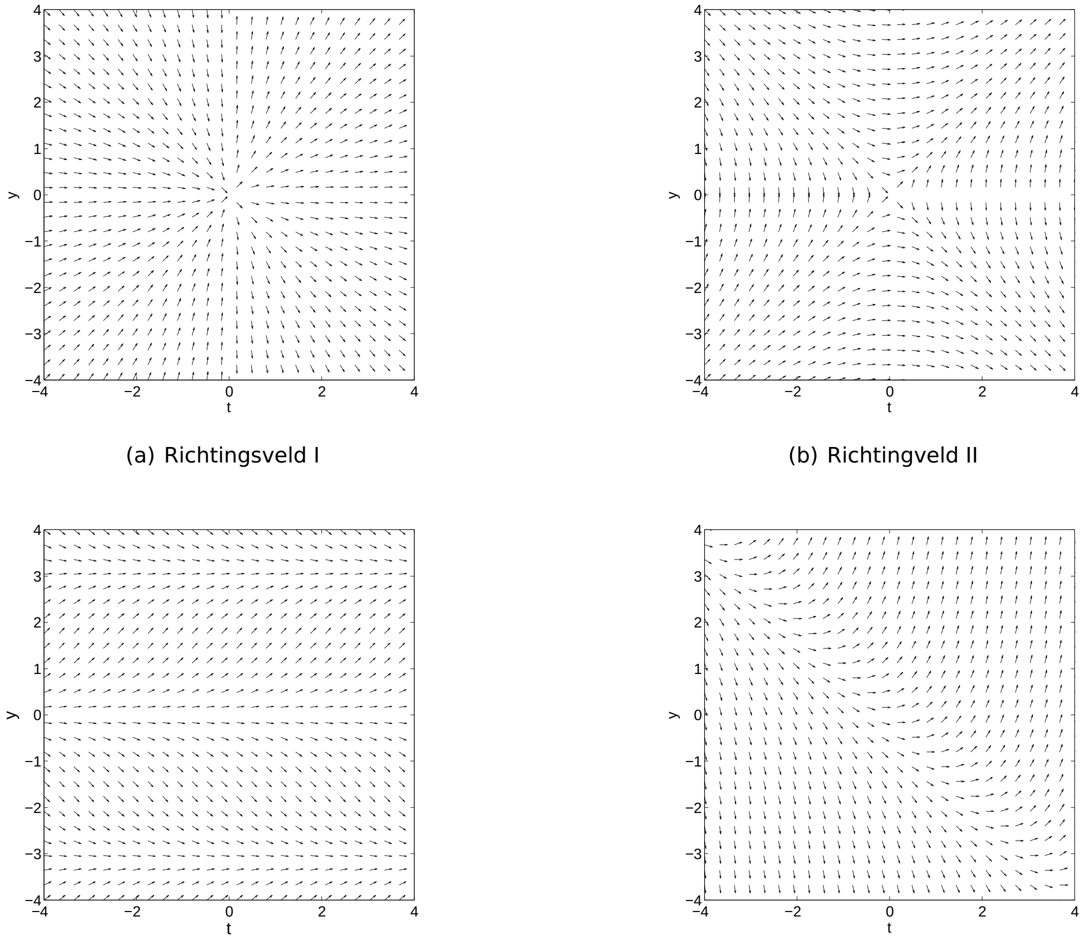
(c) Describe what happens when you increase the precision n. For example, set n equal to 10, 20, 50, and 100.

(d) Can you predict the global course of the exact solution through the point $(-2, 0.5)$ using this plot?

(e) Determine the general solution of this differential equation using Mathematica.

(f) Use Mathematica to determine the exact solution of the differential equation through the point $(-2, 0.5)$ and implement this solution in Python as `exact2`. Optionally use `Simplify` to simplify the obtained solution. Plot the exact solution through $(-2, 0.5)$ on top of the direction field. Does the result match the trend predicted in (d)?

Assignment 18.3 — Consider the four direction fields depicted in Figure 18.15.



(a) Richtingsveld I

(b) Richtingveld II

(c) Richtingsveld III

(d) Richtingsveld IV

Figure 18.15: Direction fields from Exercise 18.3.

Which differential equation corresponds to which direction field? Justify your answer.

1. $y' = \sin(y)$
2. $y' = \dfrac{y}{t}$
3. $y' = \dfrac{t}{y}$
4. $y' = t + y$

18.4.2 Equilibria and stability

Assignment 18.4 — Determine analytically the equilibrium points of the following differential equations. Check your results with Mathematica.

(a) $y' = (y-4)(y+1)$

(b) $y' = y^2(6-y)$

(c) $y' = y^2 - y - 6$

(d) $y' = y^3 - 4y$

(e) $y' = 2\sqrt{y}$

(f) $N' = rN\left(1-\dfrac{N}{K}\right) - pN$ with $r = 2.5$, $K = 250$, and $p = 0.2$

(g) $y' = h(t-1)$, where $h(t) = 1$ if $t \geq 0$ and $h(t) = 0$ else

Assignment 18.5 —

(a) Determine the stability of the equilibrium points you found in the previous exercise. Check your results by plotting the corresponding direction fields.

Assignment 18.6 —

(a) Consider the differential equation describing the velocity of an object in free fall with air resistance (cfr. Equation (17.4)):
$$m\frac{dv}{dt} = mg - \mu v\,.$$

(b) Determine the equilibrium point(s) as a function of m and μ.

$v_e =$

(c) Determine the stability of the equilibrium point(s).

Assignment 18.7 — In a chemical reaction, one molecule of the substance P and one molecule of the substance Q together form one molecule of the substance R. At the start of the reaction, p [–] molecules P and q [–] molecules Q are present. The number of molecules R at the instant t [T] is x. The rate at which R is formed is at any time during the reaction directly proportional to the product of the remaining numbers of molecules P and Q:

$$\frac{dx}{dt} = k(p-x)(q-x),$$

with $k > 0$.

(a) Deteremine the equilibrium point(s) as a function of k, p and q.

$x_e =$

(b) Determine the stability of the equilibrium point(s).

Assignment 18.8 — Consider a wastewater treatment plant where wastewater circulates in cylindrical tanks so that microorganisms can break down the organic material present (see Figure 18.16). At the top of such a tank with volume V (l) waste water enters at a flow rate q [$L^3\,T^{-1}$]. The concentration of organic material in the inflow is known and equal to C_{in} [$M\,L^{-3}$]. At the bottom of the tank waste water leaves the tank at the same flow rate q [$L^3\,T^{-1}$] so that the volume of waste water in the tank remains constant.

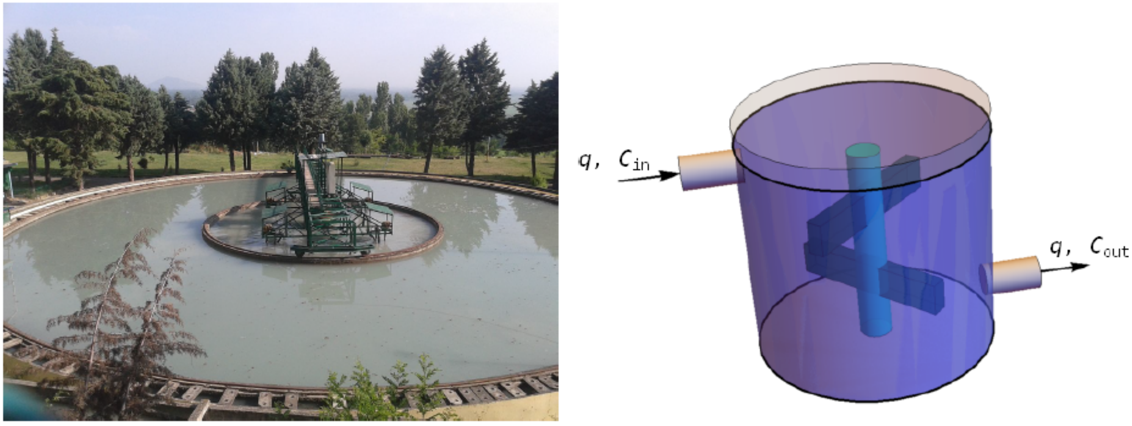


Figure 18.16: Wastewater treatment plant (left) and schematic representation (right).

The concentration of organic material in the effluent is also known and equal to C_{out} [$M\,L^{-3}$]. The microorganisms in the tank break down the organic material at a rate r [T^{-1}]. In the middle of the tank is a mixing system that ensures that the waste water is well mixed. This means that concentration in the outflow is equal to the concentration in the tank: $C_{out} = C$. The differential equation describing the change in the concentration $C(t)$ [$M\,L^{-3}$] of the organic material is given by:

$$C' + \left(\frac{q}{V} + r\right) C = \frac{q\,C_{in}}{V}\,.$$

(a) Determine the equilibrium point(s).

$C_e =$

(b) Determine the stability of the equilibrium point(s).

18.4.3 Bifurcation

Assignment 18.9 — Consider the following alternative model for Equation (18.5):

$$P' = r\left(1 - \frac{P}{K}\right)P - h\frac{P}{1+P}\,, \tag{18.8}$$

with $r = 2$, $K = 200$ and h [T^{-1}] the rate at which the population is harvested.

(a) Determine the equilibrium point(s) as a function of r, K, and h.

$P_e =$

(b) Determine the stability of the equilibrium point(s).

(c) Sketch the bifurcation diagram. What type of bifurcation do we have here?

Assignment 18.10 — Some populations are in danger of extinction when their size falls below a critical value. For example, if the population size becomes too small, individuals will have a hard time finding a suitable mate for reproduction. This is called the Allee effect. A simple extension of the logistic model that takes this effect into account is given by the following differential equation:

$$P' = r\,(P - a)\left(1 - \frac{P}{K}\right)P\,,$$

with $0 < a < K$.

(a) Determine the equilibirium point(s) as a function of a, r, and K.

$P_e =$

(b) Determine the stability of the equilibrium point(s) if $r = 1.5$ and $K = 500$.

(c) Sketch the bifurcation diagram. What type of bifurcation do we have here?

Assignment 18.11 — Suppose a free falling object experiences air resistance proportional to the square of the object's velocity:

$$m\frac{dv}{dt} = m\,g - k\,v^2\,,$$

with k [M L^{-1}] the resistance coefficient.

(a) Determine the equilibrium point(s) as a function of k.

$v_e =$

(b) Determine the stability of the equilibrium point(s).

(c) Sketch the bifurcation diagram. What type of bifurcation do we have here?

The shortest path between two truths in the real domain passes through the complex domain.

— Jacques Hadamard —

19 Methods of solution for first-order differential equations

In this chapter we will study first-order differential equations for which there are general analytical methods of solution. We will start our investigation with linear first-order differential equations because these are the easiest to treat mathematically. Analytical solutions for their non-linear counterparts fall beyond the scope of this course. Instead, we will develop a few of the most important numerical methods for solving first-order differential equations.

19.1 Linear first-order differential equations

19.1.1 Existence and uniqueness of solutions

A first-order differential equation is said to be linear if it can be written in standard form as

$$y' + p(t)y = g(t)\,. \tag{19.1}$$

A first-order differential equation that cannot be written like this is non-linear. In line with the terminology introduced in Section 17.2.1 we say that Equation (19.1) is homogeneous if $g \equiv 0$; otherwise it is non-homogeneous. Since $y(t) = 0$ is obviously a solution of the homogeneous differential equation

$$y' + p(t)y = 0\,,$$

we call it the **trivial solution** (*triviale oplossing*). Any other solution is non-trivial.

Before trying to find a solution of Equation (19.1), it would be good to be assured that there exists a general solution, and given an initial condition $y(t_0) = y_0$, that there exists a unique solution of the initial value problem. For that purpose, the following theorem definitely helps.

Theorem 19.1 (Existence and uniqueness theorem for linear first-order differential equations) *Let the functions $p(t)$ and $g(t)$ be continuous on an open t-interval $]\,a,b\,[$ containing the point $t=t_0$, then there exists a unique function $y(t)$ that satisfies the differential equation*

$$y'+p(t)y=g(t),$$

for every t in $]\,a,b\,[$, and that also satisfies the initial condition $y\left(t_0\right)=y_0$.

In order to find the largest possible t-interval where Theorem 19.1 guarantees that $y_{(t_0,y_0)}(t)$ is a unique solution of Equation (19.1) satisfying the initial condition $y\left(t_0\right)=y_0$, we merely have to look for the points t where $p(t)$ and/or $g(t)$ are not continuous and choose the largest t-interval without any singularities that contains t_0.

Example 19.1

Determine the largest possible t-interval where Theorem 19.1 guarantees that $y_{(t_0,y_0)}(t)$ is a unique solution of the following differential equations:

1. $2\,y'+t\,y=\sin(t)$,
2. $y'+\dfrac{1}{t}\,y=0$,
3. $\cos(t)\,y'+y=\sin(t)$,

for $y(1)=y_0$.

Solution

1. The functions $p(t)=t/2$ and $g(t)=\sin(t)/2$ are continuous everywhere, so this differential equation has a unique solution $y_{(1,y_0)}(t)$ on $]-\infty,+\infty\,[$.
2. The function $p(t)=1/t$ is not continuous in $t=0$, while $g(t)=0$ is continuous everywhere $y'+\dfrac{1}{t}\,y=0$, so this differential equation has a unique solution $y_{(1,y_0)}(t)$ on $]\,0,+\infty\,[$.
3. The functions $p(t)=1/\cos(t)$ and $g(t)=\sin(t)/\cos(t)=\tan(t)$ are not continuous in $t=\pi/2+k\,\pi$, with $k\in\mathbb{Z}$, so this differential equation has a unique solution $y_{(1,y_0)}(t)$ on $]-\pi/2,\pi/2\,[$.

19.1.2 Homogeneous linear first-order differential equations

Let us now try to find the general solution of the homogeneous linear first-order equation

$$y'+p(t)y=0$$

on $]\,a,b\,[$. We can rewrite this differential equation as

$$\frac{y'}{y}=-p(t) \tag{19.2}$$

for t in $]\,a,b\,[$. Integrating both sides of Equation (19.2) yields

$$\ln|y| = -\int p(t)\,dt + C', \tag{19.3}$$

where C' is an integration constant. Upon introducing the antiderivative of $p(t)$, i.e.

$$P(t) = \int p(t)\,dt,$$

we can rewrite Equation (19.3) as

$$\ln|y(t)| = -P(t) + C'.$$

This implies that

$$|y(t)| = e^{C'}\,e^{-P(t)}.$$

Since a function that satisfies this equation cannot change sign on either $]-\infty,0\,[$ or $]\,0,+\infty\,[$ we can rewrite it as

$$y(t) = C\,e^{-P(t)}, \tag{19.4}$$

where

$$C = \begin{cases} e^{C'}, & \text{if} \quad y > 0 \text{ on }]a,b[, \\ -e^{C'}, & \text{if} \quad y < 0 \text{ on }]a,b[. \end{cases}$$

Equation (19.4) is the general solution of a homogeneous linear first-order differential equation.

Example 19.2

Find the general solution of

$$t\,y' + y = 0. \tag{19.5}$$

Then, determine the solution that satisfies $y(1) = 3$ and the largest t-interval for which Theorem 19.1 guarantees its uniqueness.

Solution

First, we rewrite Equation (19.5) in standard form as

$$y' + \frac{1}{t}y = 0. \tag{19.6}$$

Hence, the function $p(t)$ is given by $1/t$, whose antiderivative equals

$$\begin{aligned} P(t) &= \int p(t)\,dt \\ &= \int \frac{1}{t}\,dt \\ &= \ln|t|. \end{aligned}$$

So, substituting this expression for $P(t)$ in Equation (19.4), we find the following general solution

$$y(t) = C\,e^{-\ln|t|} = \frac{C}{|t|}. \tag{19.7}$$

Figure 19.1 shows some solution curves given by Equation (19.7) for varying values of the constant C, imposed on the direction field of Equation (19.5).

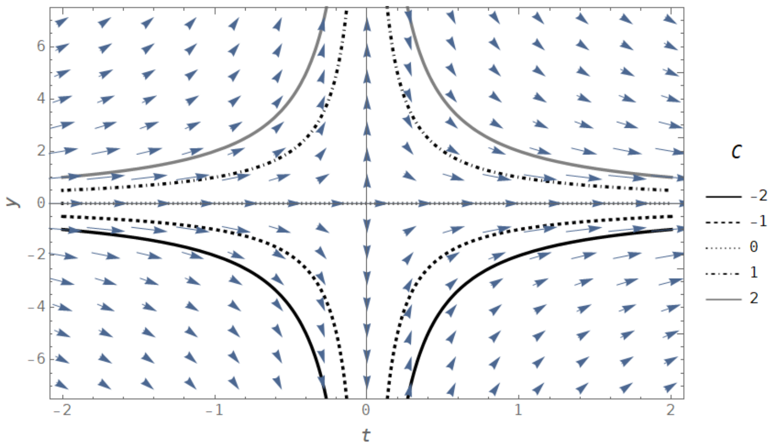


Figure 19.1: Solution curves given by Equation (19.7) for varying values of the constant C, imposed on the direction field of Equation (19.5).

Imposing the initial condition $y(1) = 3$ on Equation (19.7) yields $C = 3$. Therefore, the solution of the initial value problem defined by Equation (19.6) and $y(1) = 3$ is

$$y = \frac{3}{t}.$$

Given the singularity of $p(t)$ in $t = 0$, Theorem 19.1 guarantees its uniqueness on $]0, +\infty[$.

19.1.3 Non-homogeneous linear first-order differential equations

Here, we want to find the solution of the non-homogeneous linear first-order differential equation

$$y' + p(t)y = g(t). \tag{19.8}$$

Equation (19.4) is the general solution of its homogeneous counterpart, also referred to as the **complementary solution** (*complementaire oplossing*) of the non-homogeneous differential equation $y_c(t)$.

For solving Equation (19.8) let us assume that its general solution is the product of two functions, being an unknown function, say $u(t)$, and the complimentary solution $y_c(t)$. Essentially, we assume the following form for $y(t)$:

$$y(t) = u(t)\, y_c(t).$$

Since $y = u\, y_c$ is assumed to be a solution of Equation (19.8), the latter should still hold upon substituting $y = u\, y_c$ and its derivative, given by

$$y' = u'\, y_c + u\, y_c'.$$

Substituting these expressions for y and y' into Equation (19.8) yields

$$u'\, y_c + u\,(y_c' + p(t)\, y_c) = g(t),$$

which reduces to

$$u'\, y_c = g(t), \tag{19.9}$$

since $y_c(t)$ is a solution of the homogeneous equation; that is,

$$y_c' + p(t) y_c = 0.$$

Given that $y_c = e^{-P(t)}$ has no zeros on an interval where $p(t)$ is continuous, we can divide Equation (19.9) by $y_c(t)$ to obtain

$$u' = \frac{g(t)}{y_c(t)} = g(t)\, e^{P(t)}\,.$$

Now, we can integrate this:

$$u(t) = \int g(t)\, e^{P(t)}\, dt + C\,,$$

where C is an integration constant. Finally, we multiply the result by $y_c(t)$ to get the general solution of Equation (19.8):

$$\begin{aligned} y(t) &= u(t)\, y_c(t) \\ &= y_c(t) \int g(t)\, e^{P(t)}\, dt + C\, y_c(t) \\ &= e^{-P(t)} \int g(t)\, e^{P(t)}\, dt + C\, e^{-P(t)}\,. \end{aligned} \tag{19.10}$$

This result is summarized in the following theorem.

Theorem 19.2
Suppose $p(t)$ and $g(t)$ are continuous on an open interval $]\,a, b\,[$. Then, the general solution of the non-homogeneous equation

$$y' + p(t)\, y = g(t) \tag{19.11}$$

on $]\,a, b\,[$ is

$$y(t) = e^{-P(t)} \int g(t)\, e^{P(t)}\, dt + C\, e^{-P(t)}\,, \tag{19.12}$$

where $P(t)$ is the antiderivative of $p(t)$.

Of course, we must not know Equation (19.12) by heart to find the general solution of linear differential equation. Instead, we can just we can just follow the procedure leading to Theorem 19.2.

Example 19.3

Find the general solution of

$$y' + 2y = t^3 e^{-2t}. \tag{19.13}$$

Solution

It is easy to see that $y_c(t) = C\, e^{-2t}$ is the complimentary solution of Equation (19.13). Therefore we seek solutions of this differential equation in the form $y(t) = u(t)\, e^{-2t}$. Since $y' = u' e^{-2t} - 2\, u\, e^{-2t}$, we get after substituting y and y' in the left-hand side of Equation (19.13):

$$\begin{aligned} y' + 2\, y &= u' e^{-2t} - 2u e^{-2t} + 2u e^{-2t} \\ &= u'\, e^{-2t}\,. \end{aligned}$$

Therefore $y(t)$ is a solution of Equation (19.13) if and only if

$$u' e^{-2t} = t^3 e^{-2t}\,,$$

or equivalently $u' = t^3$. Therefore

$$u(t) = \frac{t^4}{4} + C\,,$$

and
$$y(t) = u(t)\, e^{-2t} = e^{-2t} \left(\frac{t^4}{4} + C \right) \tag{19.14}$$
is the general solution of Equation (19.13).

Figure 19.2 shows some solution curves given by Equation (19.14) for varying values of the constant C imposed on the direction field of Equation (19.13).

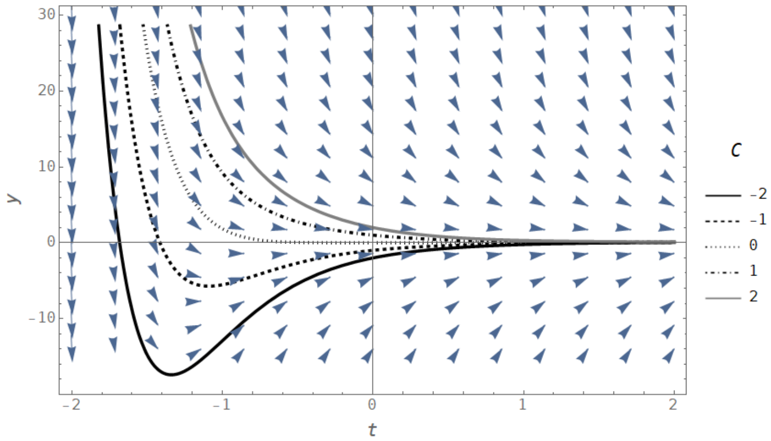


Figure 19.2: Solution curves given by Equation (19.14) for varying values of the constant C imposed on the direction field of Equation (19.13).

Example 19.4

To conclude, we will try to find the general solution of the differential equation describing the velocity of a free-falling object (Equation (17.4)):
$$m\, v' = m\, g - \mu\, v\, .$$
We start by rewriting it in standard form, i.e.
$$v' + \frac{\mu\, v}{m} = g\, ,$$
from which we infer that both $p(t) = \mu/m$ and $g(t) = g$ are constant. So, $P(t) = \frac{\mu}{m} t$ and the complimentary solution is given by $y_c(t) = e^{-\frac{\mu}{m} t}$. Direct application of Equation (19.12) yields
$$v(t) = e^{-\frac{\mu}{m} t} \int g\, e^{\frac{\mu}{m} t} dt + C\, e^{-\frac{\mu}{m} t}\, ,$$
which after evaluating the integral can be simplified to
$$v(t) = \frac{g\, m}{\mu} + C\, e^{-\frac{\mu}{m} t}\, .$$
This general solution clearly shows that
$$\lim_{t \to +\infty} v(t) = \frac{g\, m}{\mu}\, ,$$
which corroborates our findings in Section 18.2.1, for what concerns the terminal velocity of the free-falling object. Some solution curves are illustrated in Figure 18.1, together with the corre-

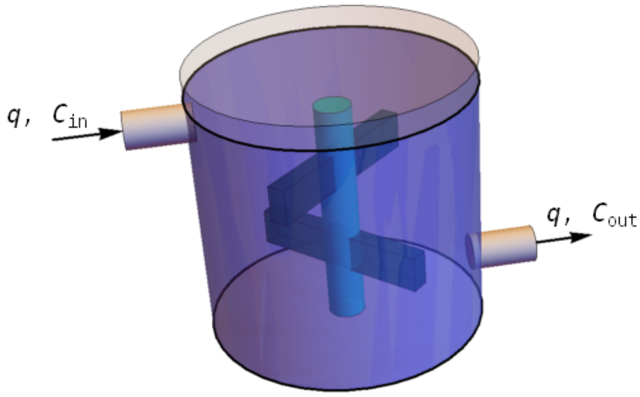
sponding direction field.

Even though linear differential equations are intrinsically simple, they are useful to describe phenomena governed by a first-order kinetics, where the rate by which a certain compound reacts is proportional to the mass or concentration of that component. In the next example, we will study one such a process in more detail.

Example 19.5

Consider a waste water treatment plant where waste water is circulated in huge aerated cylindrical tanks so that microorganisms are stimulated to decompose to organic material present in the waste water. Figure 19.3(a) shows such a typical tank, while its schematic representation is depicted in Figure 19.3(b). At the top of the tank with volume V [L^3] waste water enters the tank with a constant flow rate q [$\mathrm{L}^3\,\mathrm{T}^{-1}$]. The known concentration of the organic material in this influent is C_{in} [$\mathrm{M\,L}^{-3}$]. Water leaves the tank at the bottom with a flow rate q [$\mathrm{L}^3\,\mathrm{T}^{-1}$] and concentration C_{out} [$\mathrm{M\,L}^{-3}$]. Microorganisms decompose the organic material in the tank with a rate constant r [T^{-1}]. In the centre of the tank there is a paddle that ensures that the fluid is well mixed.

(a)



(b)

Figure 19.3: A typical aeration tank of a waste water treatment plant (a) and its schematic representation (b).

In operation phase, we normally want to know how long it takes before the concentration of organic material in the tank gets lower than a certain threshold, as this guarantees the quality of the effluent. So, we want to find how the concentration $C(t)$ in the tank changes over time. For that purpose, let us try to formulate a mass balance equation for the organic material (OM) in the tank, based on the law of mass conservation. Since OM cannot just disappear, it should hold that:

$$\begin{array}{c}\text{Change of mass}\\ \text{OM in tank}\\ \text{during } \Delta t\end{array} = \begin{array}{c}\text{Mass OM}\\ \text{entering tank}\\ \text{during } \Delta t\end{array} - \begin{array}{c}\text{Mass OM}\\ \text{leaving tank}\\ \text{during } \Delta t\end{array} - \begin{array}{c}\text{Mass OM decomposed}\\ \text{by microorganisms}\\ \text{during } \Delta t\end{array}$$

The mass M entering the tank during Δt is given by

$$q\,C_{in}\,\Delta t\,,$$

and similarly the mass leaving the tank during Δt is given by

$$q\,C_{out}\,\Delta t\,.$$

Now, we face the problem that C_{out} is not known. Yet, at this point we can rely on the fact that the tank is well mixed, meaning that the concentration in tank is spatially homogeneous. Consequently, the concentration in the effluent is the same as the concentration in tank, i.e.

$$q\,C_{out}\,\Delta t = q\,C\,\Delta t\,.$$

The last term that we need to formalize in the mass balance equation is the one relating to the decomposition of the organic material by the micro-organisms. If we assume that this decomposition is proportional to the concentration of organic material in the tank (first-order kinetics), we get

$$r\,C\,V\Delta t\,.$$

Plugging these expression into the mass balance equation, we obtain:

$$\Delta M = q\,C_{in}\,\Delta t - q\,C\,\Delta t - r\,C\,V\Delta t\,,$$

or since $C = M/V$, in terms of the concentration of organic material:

$$\Delta C = \frac{q\,C_{in}\,\Delta t}{V} - \frac{q\,C\,\Delta t}{V} - r\,C\,\Delta t\,.$$

We can also express the concentration change per unit of time by dividing both sides by Δt:

$$\frac{\Delta C}{\Delta t} = \frac{q\,C_{in}}{V} - \frac{q\,C}{V} - r\,C\,. \tag{19.15}$$

Finally, we assume that Δt becomes infinitesimally small, so ΔM too, and we arrive at the continuum limit of Equation (19.15):

$$C' = \frac{q\,C_{in}}{V} - \frac{q\,C}{V} - r\,C\,.$$

This is a linear first-order differential equation, which reads in standard form:

$$C' + \left(\frac{q}{V} + r\right)C = \frac{q\,C_{in}}{V}\,. \tag{19.16}$$

We leave a detailed analyse of this equation and the behaviour of its solutions to the exercises.

19.2 Solutions of non-linear first-order differential equations

Before turning to the method for solving first-order differential equations, we will first state the existence and uniqueness theorem for non-linear first-order differential equations.

Theorem 19.3 (Existence and uniqueness theorem for non-linear first-order differential equat

1. *If f is a continuous function on an open rectangle*

$$R:\{a<t<b, c<y<d\}$$

that contains (t_0, y_0), then the initial value problem

$$y' = f(t,y), \quad y(t_0) = y_0 \tag{19.17}$$

has at least one solution on some open subinterval of $]a,b[$ that contains t_0.

2. *If both f and $f_y = \frac{\partial f}{\partial y}$ are continuous on R, then the initial value problem* (19.17) *has a unique solution on some open subinterval of $]a,b[$ that contains t_0.*

It is important to understand exactly what Theorem 19.3 means. Its first part guarantees that a solution exists on some open t-interval that contains t_0, but provides no information on how to find the solution, or to determine the open t-interval on which it exists. Moreover, it provides no information on the number of solutions that the initial value problem (19.17) may have. It leaves open the possibility that it has two or more solutions that differ for values of t arbitrarily close to t_0. The second part of Theorem 19.3 guarantees that the initial value problem (19.17) has a unique solution on some open subinterval of $]a,b[$ that contains t_0.

Example 19.6

Consider the differential equation

$$y' = \frac{t^2-y^2}{t^2+y^2}\,, \tag{19.18}$$

subject to $y(t_0) = y_0$.

The right-hand side of Equation (19.18) is given by

$$f(t,y) = \frac{t^2-y^2}{t^2+y^2}\,,$$

and its derivative with respect to y by

$$f_y(t,y) = -\frac{4\,t^2\,y}{(t^2+y^2)^2}\,.$$

Both are continuous everywhere, except at $(0,0)$. If $(t_0,y_0)\neq(0,0)$, there is an open rectangle R that contains (t_0,y_0) that does not contain $(0,0)$ and where f is continuous. This already implies that Equation (19.23) has at least one solution on some open t-interval that contains t_0. Moreover, since f and f_y are continuous on R, Theorem 19.3 implies that if $(t_0,y_0)\neq(0,0)$, then the initial value problem has a unique solution on some open t-interval $]a,b[$ that contains t_0.

Example 19.7

Consider the initial value problem

$$y' = \frac{10}{3} t y^{2/5}, \quad y(t_0) = y_0. \tag{19.19}$$

1. For what points (t_0, y_0) does Theorem 19.3 imply that it has a solution?
2. For what points (t_0, y_0) does Theorem 19.3 imply that it has a unique solution on some open t-interval that contains t_0?

Solution

1. Since
$$f(t, y) = \frac{10}{3} t y^{2/5}$$
is continuous for all (t, y), Theorem 19.3 implies that the initial value problem (19.19) has at least one solution for every (t_0, y_0).
2. Here, we have that
$$f_y(t, y) = \frac{4}{3} t y^{-3/5}$$
is continuous for all (t, y), except where $y_0 = 0$. Therefore, if $y_0 \neq 0$ there is an open rectangle R on which both f and f_y are continuous, and Theorem 19.3 implies that the initial value problem (19.19) has a unique solution on some open interval that contains t_0. If $y = 0$ then $f_y(t, y)$ is discontinuous; hence, Theorem 19.3 does not apply to the initial value problem (19.19) if $y_0 = 0$.

At this point, it is important to realize the difference between linear and non-linear first-order differential equations in terms of the existence and uniqueness guaranteed by Theorem 19.1 and 19.3 respectively. The former states that if $p(t)$ and $g(t)$ are continuous on $]a, b[$ then the solution of

$$y' + p(t) y = g(t)$$

satisfying $y(t_0) = y_0$ is guaranteed to exist on the entire open interval $]a, b[$ and is the only one on $]a, b[$. This is, however, not true for non-linear first-order differential equations. Indeed, Theorem 19.3 merely assures that there is at least one solution on some open subinterval of $]a, b[$ if f is continuous on some open rectangle R containing (t_0, y_0). Besides, it only assures its uniqueness on this subinterval if also f_y is continuous on R. Still, both theorems imply that two solutions cannot intersect in a point where their premises are fulfilled.

19.3 Separable differential equations

A first-order differential equation is **separable** (*scheidbaar*) if it can be written as

$$h(y) y' = g(t), \tag{19.20}$$

where the left side is a product of y' and a function of y and the right side is a function of t only. Rewriting a separable differential equation in this form is called **separation of variables** (*scheiding van veranderlijken*). Essentially, without really giving it a name we already used separation of variables to solve homogeneous linear differential equations in Section 19.1.2. Here we will apply this method to non-linear differential equations.

To see how to solve Equation (19.20), let us first assume that $y(t)$ is a solution. Moreover, let $H(t)$ and $G(y)$ be antiderivatives of $h(y)$ and $g(t)$, respectively; that is,

$$H(y) = \int h(y)\,dy \quad \text{and} \quad G(t) = \int g(t)\,dt\,. \tag{19.21}$$

Then, using the fact that $y' = dy/dt$, we may rewrite Equation (19.20) as

$$h(y)\,dy = g(t)\,dt\,,$$

so that we can integrate both sides

$$\int h(y)\,dy = \int g(t)\,dt$$

to obtain

$$H(y(t)) = G(t) + C\,, \tag{19.22}$$

where C is an integration constant.

Although we derived this equation on the assumption that $y(t)$ is a solution of Equation (19.20), we can now view it differently. Any differentiable function y that satisfies Equation (19.22) for some constant C is a solution of Equation (19.20). To see this, we differentiate both sides of Equation (19.22) with respect to t, using the chain rule on the left, to obtain

$$H'(y(t))y'(t) = G'(t)\,,$$

which is equivalent to

$$h(y(t))y'(t) = g(t)$$

because of Equations (19.21).

In conclusion, to solve Equation (19.20) it suffices to find functions $G = G(t)$ and $H = H(y)$ that satisfy Equations (19.21). Then any differentiable function $y = y(t)$ that satisfies Equation (19.22) is a solution of Equation (19.20).

Example 19.8

Solve the following first-order differential equation:

$$y' = t\left(1 + y^2\right).$$

Solution

Separating variables yields

$$\frac{y'}{1+y^2} = t\,.$$

Integrating yields

$$\arctan(y) = \frac{t^2}{2} + C.$$

Therefore

$$y(t) = \tan\left(\frac{t^2}{2} + C\right).$$

There are many situations where one ends up with an implicit solution of the differential equation when using the method of separation of variables, though in some cases it happens that it falls apart into two or more explicit solutions. When this happens when solving an initial value problem for which the premises of Theorem 19.3 are satisfied for some initial condition (t_0, y_0), only one of these explicit solutions will satisfy the initial condition (t_0, y_0). This is illustrated in the following example.

Example 19.9

Find the general solution of the following first-order differential equation:

$$y' = -\frac{t}{y}\,. \tag{19.23}$$

Then, solve the initial value problems involving this equation and

1. $y(1) = 1$,
2. $y(1) = -2$.

Solution

Separating variables in Equation (19.23) yields

$$yy' = -t\,.$$

Integrating both sides gives

$$\frac{y^2}{2} = -\frac{t^2}{2} + C\,.$$

This is the equation of a circle centred at the origin, and embodies two explicit solutions. Indeed, recasting this equation and introducing $a^2 = 2C$ yields

$$t^2 + y^2 = a^2. \tag{19.24}$$

The two explicit solutions are:

$$y = \sqrt{a^2 - t^2}\,, \tag{19.25}$$

and

$$y = -\sqrt{a^2 - t^2}\,, \tag{19.26}$$

for $-a < t < a$, implying that the interval of existence is $]-a, a[$. The solution curves defined by Equation (19.25) are semicircles lying completely above the t-axis, while those defined by Equation (19.26) are semicircles lying completely below the t-axis.

1. For the solution satisfying $y(1) = 1$ it must hold that it is positive when $t = 1$. Consequently, it must be of the form given by Equation (19.25). Substituting $t = 1$ and $y = 1$ into Equation (19.24) gives $a^2 = 2$. Hence, the solution of this initial value problem is

   $$y = \sqrt{2 - t^2}\,,$$

   for $-\sqrt{2} < t < \sqrt{2}$.

2. For the solution satisfying $y(1) = -2$ it must hold that it is negative when $t = 1$. So, it must be of the form given by Equation (19.26). Substituting $t = 1$ and $y = -2$ into Equation (19.24) gives $a^2 = 5$. Hence, the solution of this intial value problem is

   $$y = -\sqrt{5 - t^2}\,,$$

for $-\sqrt{5} < t < \sqrt{5}$.

So, for each of the initial value problems it turns out that there is only one valid explicit solution. This may of course, be understood by acknowledging that the premises of Theorem 19.3 are met in both cases, so we are guaranteed that there exists a unique – so only one – solution on some t-interval containing the point $t_0 = 1$. Figure 19.4 shows the solutions of both initial value problems imposed on the direction field of Equation (19.23). Given the discontinuity of the right-hand side of Equation (19.23) if $y = 0$, Theorem 19.3 is of no use if $y_0 = 0$, but given that the general solution represents a circle we can anticipate that there will be two solutions in those cases.

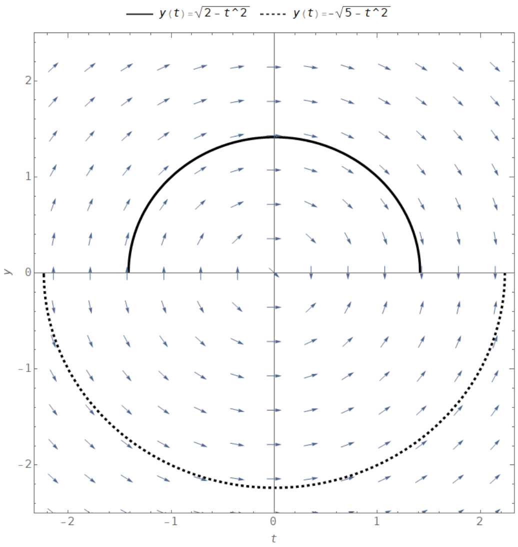


Figure 19.4: Solutions of the initial value problems given by Equation (19.23) and a) $y(1) = 1$ and b) $y(1) = -2$, imposed on the direction field of the corresponding differential equation.

Unfortunately, it is not always possible to extract the explicit solutions contained in an implicit one. This is illustrated in the following example.

Example 19.10

Consider the following differential equation:

$$y' = \frac{2t+1}{5y^4+1}. \tag{19.27}$$

Separating variables yields

$$\left(5y^4+1\right)y' = 2t+1,$$

so that we can integrate both sides of this equation, which yields

$$y^5 + y = t^2 + t + C, \tag{19.28}$$

where C is an integration constant. Clearly, this solution is an implicit one and there is no way to solve it for $y(t)$ as we did in Example 19.9.

Suppose that we are looking for the solution satisfying the initial condition $y(2) = 1$. Since both $f(t,y)$ and $f_y(t,y)$, given by

$$-\frac{20(2t+1)y^3}{\left(5y^4+1\right)^2},$$

are continuous everywhere in $\mathbb{R}^2$, Theorem 19.3 guarantees the existence of a unique solution of this initial value problem on some t- subinterval of $]-\infty,+\infty[$ containing $t_0 = 2$. Now, plugging $y(2) = 1$ into Equation (19.28) yields $1+1 = 4+2+C$, so that $C = -4$. Therefore,

$$y^5 + y = t^2 + t - 4$$

is the unique solution of the initial value problem given by Equation (19.27) and the initial condition $y(2) = 1$. Figure 19.5 shows this unique solution imposed on the direction field of Equation (19.27).

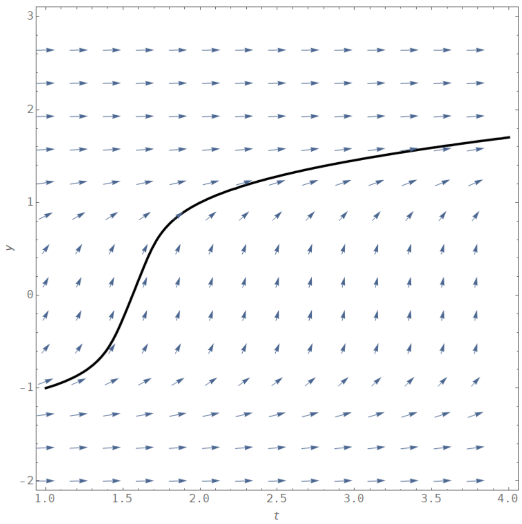


Figure 19.5: Unique solution of Equation (19.27) satisfying $y(2) = 1$ imposed on the direction field of this differential equation.

19.4 Exact differential equations

Here, it is convenient to write first-order differential equations in the form

$$M(t,y)\,dt + N(t,y)\,dy = 0\,. \tag{19.29}$$

This equation can be interpreted as

$$M(t,y) + N(t,y)\,\frac{dy}{dt} = 0\,, \tag{19.30}$$

where t is the independent variable and y is the dependent variable, or as

$$M(t,y)\,\frac{dt}{dy} + N(t,y) = 0\,, \tag{19.31}$$

where y is the independent variable and t is the dependent variable. Often the solutions of Equations (19.30) and (19.31) will be implicit; that is $F(t,y) = C$.

Since the total differential of $F(t,y)$ is given by

$$dF = \frac{\partial F}{\partial t}\,dt + \frac{\partial F}{\partial y}\,dy\,,$$

it immediately follows that $F(t,y) = C$ is a solution of the differential equation

$$F_t(t,y)\,dt + F_y(t,y)\,dy = 0\,. \tag{19.32}$$

Consequently, if we are able to find a function $F(t,y)$ for which

$$F_t(t,y) = M(t,y) \quad \text{and} \quad F_y(t,y) = N(t,y) \tag{19.33}$$

for all (t,y) in some open rectangle R where F_t and F_y are continuous, we can straightforwardly solve Equation (19.29), which is then called an **exact differential equation** (*exacte differentiaalvergelijking*). The obvious question now of course is when a first-order differential equation is exact, i.e. when can we expect to find such an implicit solution $F(t,y) = C$?

Well, if there exists such a function $F(t,y)$ for which Equations (19.33) hold, it must also hold that

$$F_{ty} = F_{yt}$$

because second-order derivatives are symmetric (Theorem of Schwarz). This implies that

$$F_{ty} = M_y = F_{yt} = N_t\,. \tag{19.34}$$

So, a necessary condition for exactness is that $M_y = N_t$. This formalized in the following theorem.

Theorem 19.4 (Exactness condition)
Suppose M and N are continuous and have continuous partial derivatives M_y and N_t on an open rectangle R. Then

$$M(t,y)\,dt + N(t,y)\,dy = 0$$

is exact on R if and only if

$$M_y(t,y) = N_t(t,y) \tag{19.35}$$

for all (t,y) in R.

Example 19.11

Find the general solution of the following differential equation

$$\left(4t^3y^3 + 3t^2\right)dt + \left(3t^4y^2 + 6y^2\right)dy = 0\,. \tag{19.36}$$

Solution

Here we have

$$M(t,y) = 4t^3y^3 + 3t^2 \quad \text{and} \quad N(t,y) = 3t^4y^2 + 6y^2\,,$$

so we see that

$$M_y(t,y) = N_t(t,y) = 12\,t^3y^2$$

for all (t,y). Hence, Theorem 19.4 implies that Equation (19.36) is exact from which it follows that there is a function F such that

$$F_t(t,y) = M(t,y) = 4\,t^3y^3 + 3\,t^2 \tag{19.37}$$

and

$$F_y(t,y) = N(t,y) = 3\,t^4y^2 + 6\,y^2 \tag{19.38}$$

for all (t,y). To find F, we integrate, for instance Equation (19.37) with respect to t to obtain

$$F(t,y) = t^4\,y^3 + t^3 + \phi(y), \tag{19.39}$$

where ϕ is a function that is independent of t, the variable of integration. To determine this function ϕ in such a way that F also satisfies Equation (19.38), we assume that ϕ is differentiable and differentiate F, given by Equation (19.39), with respect to y. This yields

$$F_y(t,y) = 3\,t^4\,y^2 + \phi'(y)\,.$$

Its right-hand side should be identical to the one of Equation (19.38), so it should hold that

$$\phi'(y) = 6y^2\,.$$

Clearly, ϕ can be found by integrating both sides of this equation with respect to y. Moreover, we may take the constant of integration to be zero because we are interested only in finding some function F that satisfies Equations (19.37) and (19.38). This yields

$$\phi(y) = 2\,y^3.$$

Finally, we substitute this expression into Equation (19.39) to obtain

$$F(t,y) = t^4\,y^3 + t^3 + 2y^3. \tag{19.40}$$

Consequently,

$$t^4\,y^3 + t^3 + 2y^3 = C$$

is the general solution of Equation (19.36) in an implicit form. This ultimately leads to the following explicit solution

$$y(t) = \left(\frac{C - t^3}{2 + t^4}\right)^{\frac{1}{3}}.$$

The choice of first integrating Equation (19.37) with respect to t was rather arbitrary and we may as well begin by integrating Equation (19.38) with respect to y to obtain

$$F(t,y) = t^4\,y^3 + 2\,y^3 + \psi(t)\,, \tag{19.41}$$

where ψ is an function of t. To determine it, we assume that it is differentiable and differentiate F with respect to t, which yields

$$F_t(t,y) = 4\,t^3\,y^3 + \psi'(t)\,.$$

Comparing this with Equation (19.37) shows that the following must hold

$$\psi'(t) = 3\,t^2.$$

Integrating this and again taking the constant of integration to be zero yields

$$\psi(t) = t^3.$$

Substituting this expression into Equation (19.41) yields Equation (19.40). Figure 19.6 shows some solutions curves of Equation (19.36), imposed on its direction field.

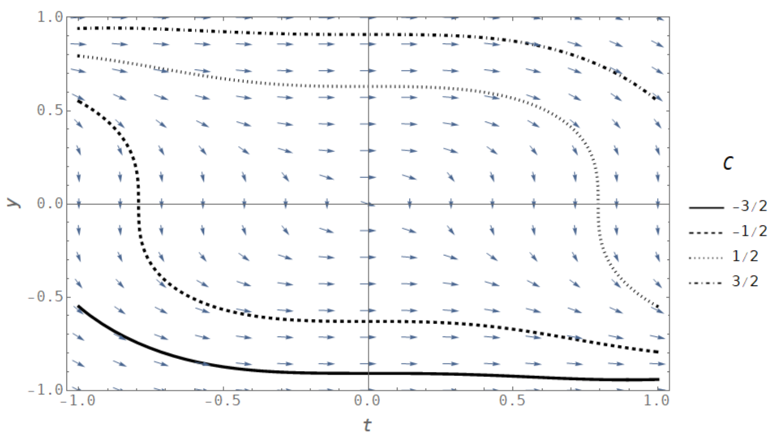


Figure 19.6: Solution curves of Equation (19.36) imposed on its direction field.

19.5 Change of variables

There are two main reasons why we might want to consider a change of the variables in a differential equation. Firstly such a change of the dependent and/or independent variable might ease the algebraic manipulations needed to solve the differential equation analytically. In this case we typically rescale the equation's variables by means of one or more of its parameters. For instance, rescaling the dependent variable in this way, we put

$$\tilde{y}(t) = p\,y\,,$$

where p is a arbitrary model parameter. Secondly, there are situations where a change of variables allows us to transform an non-separable to a separable equation or even a linear equation. In such cases, we will introduce

$$y(t) = u(t)\,\phi(t)\,,$$

where $\phi(t)$ is some known function and $u(t)$ satisfies a separable equation.

Below we give two examples to illustrate each of these settings.

Example 19.12

Consider the Verhulst model based on Equation (18.2) (Example 18.2):

$$P' = r\,P\left(1 - \frac{P}{K}\right).$$

This is a separable equation, though as it involves two parameters, the algebraic manipulations might become cumbersome. Since we know from our analysis in Example 18.2 that the population size will always converge to $P_e = K$, it makes sense to express it relative to the carrying capacity K. For that reason, we put

$$\widetilde{P} = \frac{P}{K}\,,$$

so

$$\widetilde{P}' = \frac{P'}{K}\,.$$

Substituting these expressions for P and $\widetilde{P}$ in Equation (18.2) we obtain after dividing both sides by K

$$\widetilde{P}' = r\,\widetilde{P}\left(1-\widetilde{P}\right),$$

which contains only one parameter, namely r. That parameter too can be eliminated by choosing $\tilde{t} = r\,t$ so that $d\tilde{t} = r\,dt$. In this way, we arrive at

$$d\widetilde{P} = \widetilde{P}\left(1-\widetilde{P}\right)d\tilde{t}\,.$$

Separating variables yields

$$\frac{d\widetilde{P}}{\widetilde{P}\left(1-\widetilde{P}\right)} = d\tilde{t}\,,$$

whose left-hand side can be split in partial fractions:

$$\left(\frac{1}{\widetilde{P}} - \frac{1}{\widetilde{P}-1}\right)d\widetilde{P} = d\tilde{t}\,.$$

Integrating both sides yields

$$\ln\left|\widetilde{P}\right| - \ln\left|\widetilde{P}-1\right| = \tilde{t} + C\,,$$

where C is an integration constant. This is the general solution of Equation (18.2), though in implicit form. To arrive at an explicit solution, we first take the exponential of both sides:

$$\exp\left(\ln\left|\widetilde{P}\right| - \ln\left|\widetilde{P}-1\right|\right) = e^{\tilde{t}+C}\,,$$

which can be simplified to

$$\frac{\left|\widetilde{P}\right|}{\left|\widetilde{P}-1\right|} = \widetilde{C}\,e^{\tilde{t}}\,,$$

where $\widetilde{C} = e^{C}$, which is always greater than zero. From a biological viewpoint we know that $\widetilde{P} > 0$, so $\left|\widetilde{P}\right| = \widetilde{P}$, and we can introduce

$$\widetilde{\widetilde{C}} = \begin{cases} \widetilde{C} & \text{if } \widetilde{P}-1 > 0\,, \\ -\widetilde{C} & \text{if } \widetilde{P}-1 < 0\,, \end{cases}$$

to arrive at

$$\frac{\widetilde{P}}{\widetilde{P}-1} = \widetilde{\widetilde{C}}\,e^{\tilde{t}}\,.$$

Solving the last equation for $\widetilde{P}$, we finally arrive at the explicit solution of Equation (18.2):

$$\widetilde{P}(t) = \frac{\widetilde{\widetilde{C}}\,e^{\tilde{t}}}{\widetilde{\widetilde{C}}\,e^{\tilde{t}}-1}\,,$$

or in terms of the original variables and upon dropping the tildes for the sake of readability

$$P(t) = K\,\frac{C\,e^{rt}}{C\,e^{rt}-1}\,. \tag{19.42}$$

Example 19.13

Find the general solution of

$$y' - y = t\,y^2 \tag{19.43}$$

by introducing the change of variables $y(t) = u(t)\,e^{t}$.

Solution

Using the chain rule we find

$$y' = u'\,e^t + u\,e^t\,.$$

Substituting the expressions for y and y' in Equation (19.43), we arrive at

$$u'\,e^t + u\,e^t - u\,e^t = t\,u^2\,e^{2t}\,,$$

or equivalently

$$u' = t\,u^2\,e^t\,.$$

Separating variables yields

$$\frac{u'}{u^2} = t\,e^t\,,$$

and integrating both sides gives

$$-\frac{1}{u} = (t-1)\,e^t + C\,.$$

Note that integration by parts was used to integrate the right-hand side. Hence,

$$u(t) = -\frac{1}{(t-1)\,e^t + C}\,,$$

and equivalently in terms of the original variables:

$$y(t) = -\frac{1}{t-1+C\,e^{-t}}.$$

Figure 19.7 shows some solutions curves of Equation (19.43), imposed on its direction field.

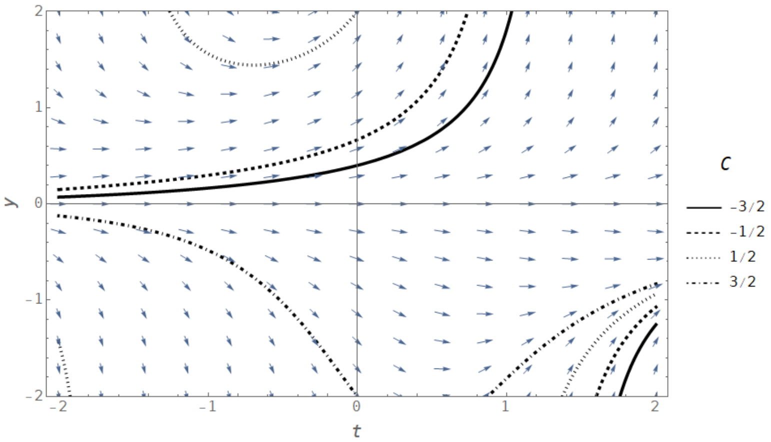


Figure 19.7: Solution curves of Equation (19.43) imposed on its direction field.

Example 19.14

First, find the general solution of

$$t^2\,y' = y^2 + t\,y - t^2\,. \tag{19.44}$$

on $]\,0,+\infty\,[$ by introducing $y(t) = u(t)\,t$. Then, solve the initial value problem defined by this differential equation and $y(1) = 2$.

Solution

Rewriting Equation (19.44) in standard form and substituting both $y = u\,t$ and $y' = u'\,t + u$, we obtain

$$u'\,t + u = \frac{(u\,t)^2 + t\,(u\,t) - t^2}{t^2} = u^2 + u - 1\,.$$

So

$$u'\,t = u^2 - 1. \tag{19.45}$$

By inspecting this equation we see that it has the constant solutions $u \equiv 1$ and $u \equiv -1$. Therefore $y(t) = t$ and $y(t) = -t$ are solutions of Equation (19.44). Other solutions can be found by separating variables

$$\frac{u'}{u^2 - 1} = \frac{1}{t}\,,$$

or, after a partial fraction expansion,

$$\frac{1}{2}\left[\frac{1}{u-1} - \frac{1}{u+1}\right]u' = \frac{1}{t}\,.$$

Multiplying by 2 and integrating yields

$$\ln\left|\frac{u-1}{u+1}\right| = 2\ln|t| + C,$$

or

$$\left|\frac{u-1}{u+1}\right| = e^C\,t^2\,.$$

Consequently, we finally have

$$\frac{u-1}{u+1} = \widetilde{C}\,t^2\,. \tag{19.46}$$

Solving this expression for u yields

$$u(t) = \frac{1 + \widetilde{C}\,t^2}{1 - \widetilde{C}\,t^2}.$$

Finally, we obtain that

$$y(t) = u(t)\,t = t\,\frac{1 + \widetilde{C}\,t^2}{1 - \widetilde{C}\,t^2} \tag{19.47}$$

is a solution of Equation (19.44) for any choice of the constant $\widetilde{C}$. Setting $\widetilde{C} = 0$ in Equation (19.47) yields the solution $y(t) = t$. However, the other solution that we obtained by inspecting the differential equation, namely $y(t) = -t$, cannot be obtained from Equation (19.47). Thus, the solutions of (19.44) on $]\,0, +\infty\,[$ are $y(t) = -t$ and functions of the form given by Equation (19.47).

Before trying to find the constant $\widetilde{C}$ such that Equation (19.47) satisfies the initial value problem, let us see what Theorem 19.3 tells us about the existence and uniqueness of the studied initial value problem. Writing Equation (19.44) in standard from we get

$$y' = \frac{y^2 + t\,y - t^2}{t^2}\,,$$

from which we conclude that f and f_y are continuous everywhere except along the line $t = 0$. In our case we have $(t_0, y_0) = (1, 2)$, so we can define an open rectangle containing this point where both f and f_y are continuous. Consequently, there must exist a unique solution of the considered initial value problem on some t-interval in $]\,0, +\infty\,[$ containing t_0.

We could obtain $\widetilde{C}$ by imposing the initial condition $y(1) = 2$ in Equation (19.47), and then solve for $\widetilde{C}$. However, it's easier to use Equation (19.46). Since $u = y/t$, the initial condition $y(1) = 2$ implies that $u(1) = 2$. Substituting this into Equation (19.46) yields $\widetilde{C} = 1/3$. Hence, the solution of the initial value problem is

$$y = \frac{t\left(1+\frac{t^2}{3}\right)}{1-\frac{t^2}{3}}.$$

The interval of existence of this solution is $\left]-\sqrt{3}, \sqrt{3}\right[$. However, the largest interval on which the initial value problem has a unique solution is $]\,0, \sqrt{3}\,[$.

Figure 19.8 shows this unique solution imposed on its direction field.

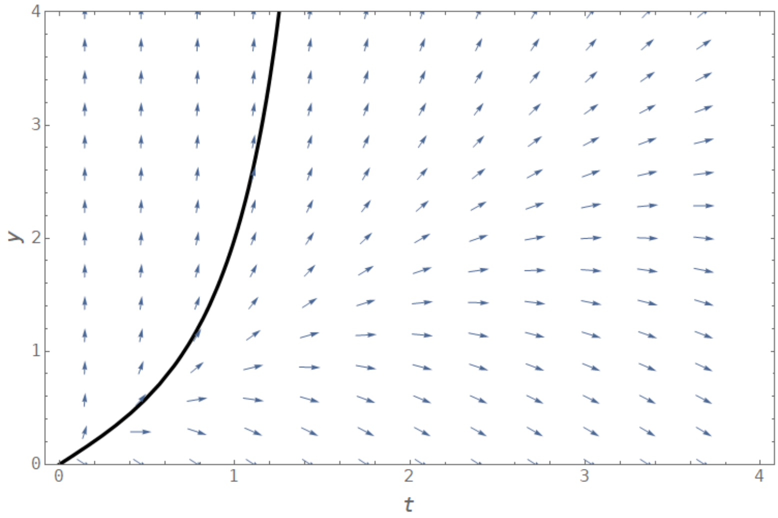


Figure 19.8: Unique solution of Equation (19.44) satisfying $y(1) = 2$ imposed on its direction field.

19.6 Euler's method

19.6.1 Motivation and rationale

If the initial value problem

$$y' = f(t, y), \quad y(t_0) = y_0 \tag{19.48}$$

cannot be solved analytically, we need numerical methods to obtain useful approximations to a solution of this initial value problem. In the remainder of this chapter, we will consider such methods. The methods that we will develop share the common feature that they lead to approximate values of the solution of the initial value problem at equally spaced points t_0, t_1, ..., t_n in an interval $[t_0, t_n]$. Hence, these points follow from

$$t_i = t_0 + i\,\Delta t, \quad i = 0, 1, \ldots, n,$$

where

$$\Delta t = \frac{t_n - t_0}{n}.$$

We will denote the approximate values of the solution at these points by y_0, y_1, ..., y_n; thus, y_i is an approximation to $y(t_i)$.

The main rationale behind the different methods that we will develop here is that the tangent line to the solution curve is known at any point (t,y) in the (t,y)-plane, and hence this information should be somehow of use to approximate for instance $y(t_1)$, provided $y(t_0)$ is known. Essentially, we will assume that (t_1,y_1) lies on a straight line passing through (t_0,y_0), whose equation is given by

$$y - y_0 = m\ (t - t_0)\ ,$$

where m is the line's slope whose definition depends on the method. We will first turn to the simple Euler's method. Yet, this method is so crude that it is seldom used in practice. Still, its simplicity makes it useful for illustrative purposes. Then, we will introduce the Runge-Kutta method, perhaps the most widely used method for the numerical solution of differential equations.

19.6.2 The method

Euler's method (*methode van Euler*), illustrated in Figure 19.9, is based on the assumption that the tangent line to the solution curve at $(t_i, y(t_i))$ approximates the solution curve itself over the interval $[t_i, t_{i+1}]$. Since the slope of the solution curve of the differential equation

$$y' = f(t,y)$$

at $(t_i, y(t_i))$ is known and equals $y'(t_i) = f(t_i, y(t_i))$, the equation of the tangent line to the solution curve at $(t_i, y(t_i))$ becomes

$$y = y(t_i) + f(t_i, y(t_i))(t - t_i)\,. \tag{19.49}$$

Suppose that (t_i, y_i) is known, then Equation (19.49) can be used to determine y_{i+1}. Indeed, setting $t = t_{i+1} = t_i + \Delta t$ in Equation (19.49) yields

$$y_{i+1} = y(t_i) + \Delta t f(t_i, y(t_i)) \tag{19.50}$$

as an approximation to $y(t_{i+1})$. Since we are focusing on initial value problems we know that $y(t_0) = y_0$, so that we can use Equation (19.50) with $i = 0$ to compute

$$y_1 = y_0 + \Delta t f(t_0, y_0)\,.$$

This is illustrated in Figure 19.9(b). Now, let us try find an approximation of $y(t_2)$, so we set $i = 1$ in Equation (19.50) which yields

$$y_2 = y(t_1) + \Delta t f(t_1, y(t_1))\,,$$

which we cannot evaluate since we do not know $y(t_1)$. Yet, we can replace $y(t_1)$ by its approximate value y_1 and redefine

$$y_2 = y_1 + \Delta t f(t_1, y_1)\,.$$

Having computed y_2, we can compute y_3 in a similar fashion, i.e.

$$y_3 = y_2 + \Delta t f(t_2, y_2)\,.$$

In general, Euler's method starts with the known value $y(t_0) = y_0$ and computes y_1, y_2, ..., y_n successively using the following recursion equation

$$y_{i+1} = y_i + \Delta t f(t_i, y_i)\,, \tag{19.51}$$

for $0 \le i \le n-1$.

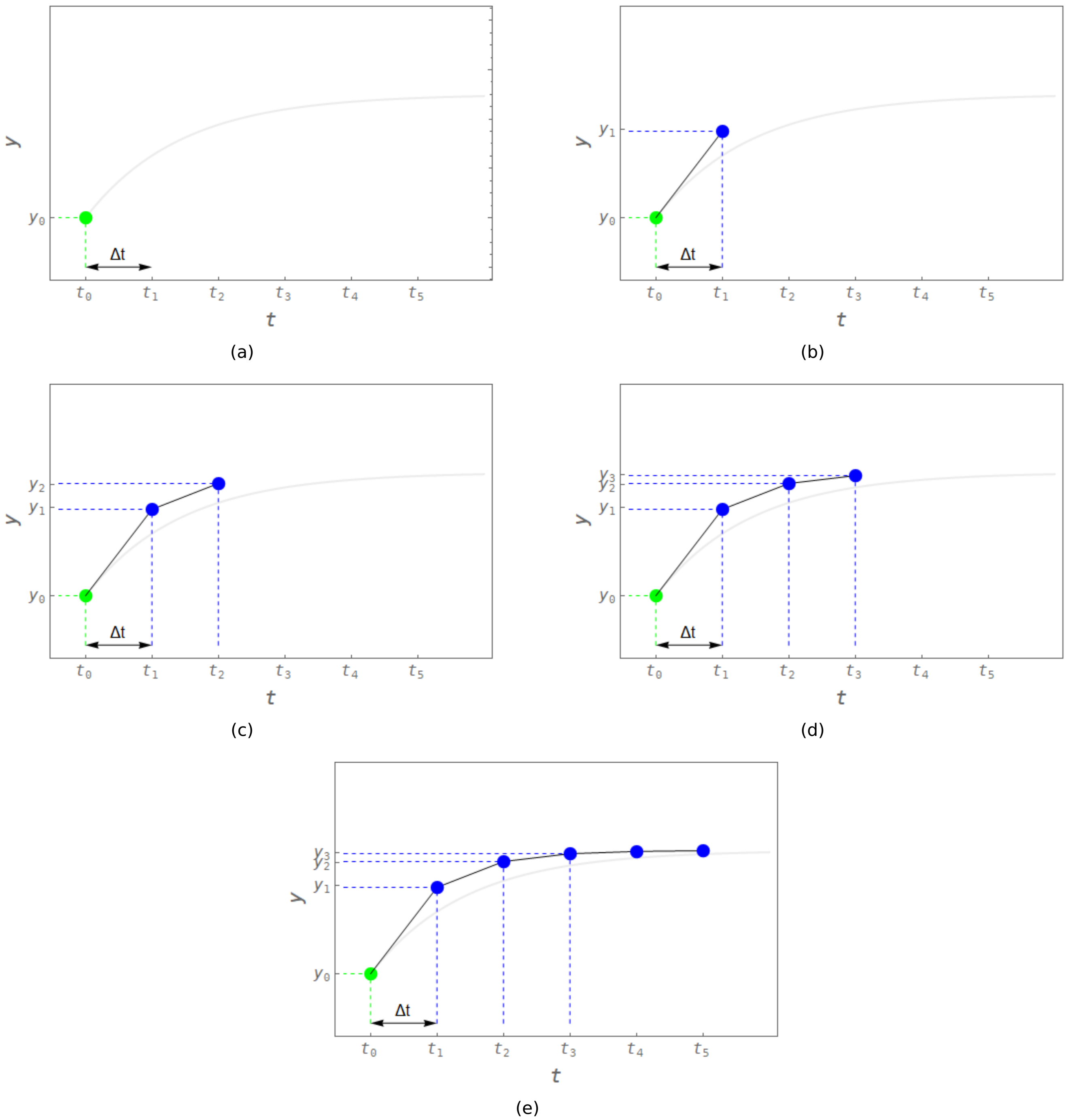


Figure 19.9: Euler's method illustrated for a generic first-order differential equation $y' = f(t, y)$: initial condition (t_0, y_0) and corresponding unique solution (a), approximation of $y(t_1)$ at t_1 (b), approximation of $y(t_2)$ at t_2(c), approximation of $y(t_3)$ at t_3 (d) and complete numerical solution on $[t_0, t_5]$.

19.6.3 Error analysis

We call

$$e_i = y(t_i) - y_i$$

the **error** (*fout*) at the i-th step. Because of the initial condition $y(t_0) = y_0$, we will always have $e_0 = 0$. However, in general $e_i \neq 0$ if $i > 0$.

We encounter two sources of error in applying a numerical method to solve an initial value problem:

- The formulas defining the method are based on some sort of approximation. Errors due to the inaccuracy of the approximation are called **truncation errors** (*benaderingsfout*).
- Computers do arithmetic with a fixed number of digits, and therefore make errors in evaluating the formulas defining the numerical methods. Errors due to the computer's inability to do exact arithmetic are called **roundoff errors** (*afrondingsfout*).

Since a careful analysis of roundoff error is beyond the scope of this course, we will consider only truncation errors.

There are two sources of truncation error in Euler's method:

1. The error committed in approximating the solution curve by the tangent line over the interval $[t_i, t_{i+1}]$.
2. The error committed in replacing $y(t_i)$ by y_i in Equation (19.49) and using Equation (19.51) rather than Equation (19.49) to compute y_{i+1}.

The **local truncation error** (*lokale benaderingsfout*) at the i-th step T_i is given by

$$T_i = y(t_{i+1}) - y(t_i) - \Delta t f(t_i, y(t_i)). \tag{19.52}$$

Since we have $t_{i+1} = t_i + \Delta t$, Taylor's theorem implies that

$$y(t_{i+1}) = y(t_i) + \Delta t\, y'(t_i) + \frac{\Delta t^2}{2} y''(\tilde{t}_i),$$

where $\tilde{t}_i$ is some number between t_i and t_{i+1}. Since $y'(t_i) = f(t_i, y(t_i))$ this can be written as

$$y(t_{i+1}) = y(t_i) + \Delta t f(t_i, y(t_i)) + \frac{\Delta t^2}{2} y''(\tilde{t}_i),$$

or, equivalently,

$$y(t_{i+1}) - y(t_i) - \Delta t f(t_i, y(t_i)) = \frac{\Delta t^2}{2} y''(\tilde{t}_i).$$

Comparing this with our definition of the local truncation error (Equation (19.52)) shows that

$$T_i = \frac{\Delta t^2}{2} y''(\tilde{t}_i).$$

Assuming that f and its partial derivatives are bounded, also its second-order derivative is bounded, so we can establish the bound

$$|T_i| \leq \frac{M\,\Delta t^2}{2}. \tag{19.53}$$

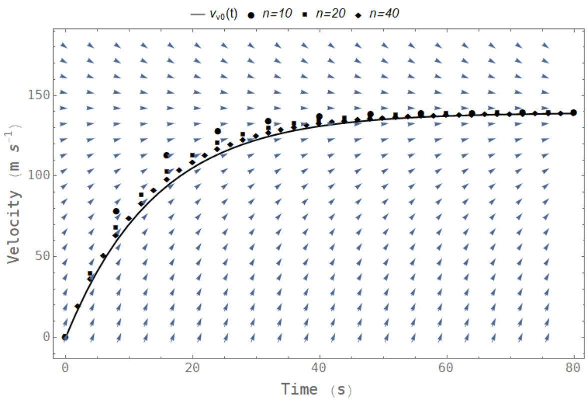
for $0 \leq i \leq n-1$. The most important take-home message from this analysis is that the local truncation error of Euler's method is of **order** (*orde*) Δt^2, which we sometimes write as $O\left(\Delta t^2\right)$. Note that the magnitude of the local truncation error in Euler's method is determined by the second derivative of the solution of the initial value problem. Therefore, the local truncation error will be larger where $|y''(t)|$ is large, or smaller where $|y''(t)|$ is small.

Example 19.15

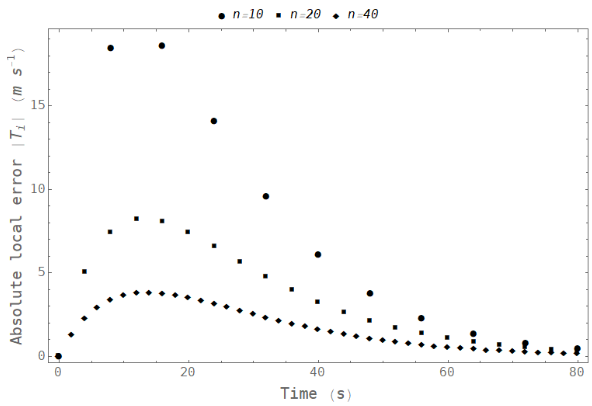
Consider once more the differential equation describing the velocity of a free-falling object (Equation (17.4)):

$$m\frac{dv}{dt} = mg - \mu v\,.$$

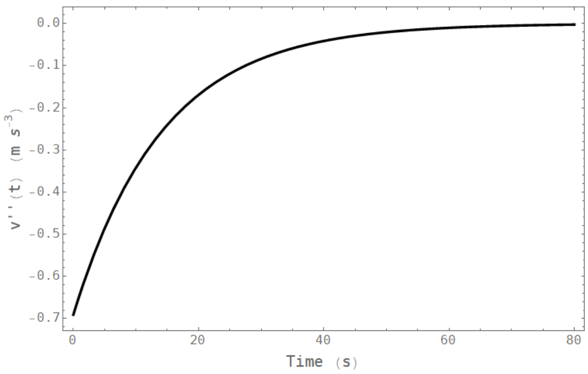
Let us take $g = 9.81\ \mathrm{m\,s^{-2}}$, $m = 75$ kg, $\mu = 5.28\ \mathrm{kg\,s^{-1}}$, and assume that the initial velocity of the object was $0\ \mathrm{m\,s^{-1}}$, so $v_0 = 0$ at $t = 0$ s. Figure 19.10 shows the analytical and corresponding numerical solutions for 10, 20 and 40 steps, together with the corresponding local absolute truncation errors and a plot of the second derivative of $v(t)$. From these plots it is clear that the local absolute truncation error decreases as the number of steps increases, while there is also a clear correlation between the magnitude of the second derivative of $v(t)$ and the absolute local truncation error.



(a)



(b)



(c)

Figure 19.10: Analytical solution and corresponding numerical solutions of Equation (17.4) satisfying $v(0) = 0$ obtained with Euler's method for a different number of steps n (a), together with the corresponding local truncation errors (b) and a plot of the second derivative of $v(t)$ (c).

In addition to the local truncation error there also exists a **global truncation error** (*globale benaderingsfout*) that is the accumulation of the local truncation errors over all iterations. A detailed analysis of the global truncation error of Euler's method is beyond the scope of this course, but it can be proofed that it is of order Δt.

19.7 Runge-Kutta methods

19.7.1 Improving the Euler method

In the previous section we saw that the global truncation error of Euler's method is $O(\Delta t)$, which would seem to imply that we can achieve arbitrarily accurate results with Euler's method by simply choosing the step size sufficiently small. However, this is not a good idea, for two reasons. First, after a certain point decreasing the step size will increase roundoff errors to the point where the accuracy will deteriorate rather than improve. The second and more important reason is that in most applications of numerical methods the expensive part of the computation is the evaluation of the right-hand side of the differential equation. Therefore we want methods that give good results for a given number of such evaluations.

An obvious way to improve Euler's method is by computing the slope of the line through (t_i, y_i) on which we can find the approximation y_{i+1} of $y(t_{i+1})$ in a more meaningful way. In Euler's method this slope is completely determined by the slope of the tangent to the solution curve passing through (t_i, y_i). Yet, if $f(t, y)$ changes rapidly as t increases, this slope will change rapidly as we move away from (t_i, y_i), so it might make sense to determine the slope of the approximating line on the basis of two or more points near (t_i, y_i).

The improved Euler method, the **Midpoint method** (*Midpointmethode*) does this by considering a line through $(t_i, y(t_i))$ with slope

$$m_i = \frac{f(t_i, y(t_i)) + f(t_{i+1}, y(t_{i+1}))}{2}$$

that is, m_i is the average of the slopes of the tangent lines to the solution curve at the endpoints of $[t_i, t_{i+1}]$. The equation of the approximating line therefore becomes

$$y = y(t_i) + \frac{f(t_i, y(t_i)) + f(t_{i+1}, y(t_{i+1}))}{2}(t - t_i). \tag{19.54}$$

Setting $t = t_{i+1} = t_i + \Delta t$ in Equation (19.54) yields

$$y_{i+1} = y(t_i) + \frac{\Delta t}{2}\big(f(t_i, y(t_i)) + f(t_{i+1}, y(t_{i+1}))\big) \tag{19.55}$$

as an approximation to $y(t_{i+1})$. As in our derivation of Euler's method, we have to replace $y(t_i)$ in this equation by its approximate value y_i because the former is unknown if $i > 0$. In this way, Equation (19.55) becomes

$$y_{i+1} = y_i + \frac{\Delta t}{2}\big(f(t_i, y_i) + f(t_{i+1}, y(t_{i+1})\big).$$

However, this still will not work, because we do not know $y(t_{i+1})$, which appears in the right-hand side of this equation. We overcome this by replacing $y(t_{i+1})$ by $y_i + \Delta t f(t_i, y_i)$, the value that the Euler method would assign to y_{i+1} when applied to the point (t_i, y_i). Thus, the improved Euler method starts with the known value $y(t_0) = y_0$ and computes y_1, y_2, ..., y_n successively with the following recursive equation:

$$y_{i+1} = y_i + \frac{\Delta t}{2}\big(f(t_i, y_i) + f(t_{i+1}, y_i + \Delta t f(t_i, y_i))\big). \tag{19.56}$$

The computation indicated here can be conveniently organized as follows: given y_i, compute

$$\begin{aligned} m_{1i} &= f(t_i, y_i), \\ m_{2i} &= f\left(t_i + \Delta t, y_i + \Delta t\, m_{1i}\right). \end{aligned}$$

Equation (19.56) can be rewritten as

$$y_{i+1} = y_i + \frac{\Delta t}{2}(m_{1i} + m_{2i}) \tag{19.57}$$

The improved Euler method requires two evaluations of $f(t, y)$ per step, while Euler's method requires only one. However, the local truncation error with the improved Euler method is $O\left(\Delta t^3\right)$, rather than $O\left(\Delta t^2\right)$ as with Euler's method. Therefore the global truncation error with the improved Euler method is $O\left(\Delta t^2\right)$, which makes this a second-order method.

We note that the magnitude of the local truncation error in the improved Euler method is determined by the third derivative of the solution of the initial value problem. Therefore the local truncation error will be larger where $|y'''(t)|$ is large, or smaller where $|y'''(t)|$ is small.

19.7.2 The methods of Runge and Kutta

One of the most widely used methods for solving differential equations numerically is the so-called **Runge-Kutta method** (*Runge-Kutta methode*). For this method it can be shown that the magnitude of the local truncation error is determined by the fifth derivative of the solution of the initial value problem. Therefore the local truncation error will be larger where $|y^{(5)}(t)|$ is large, or smaller where $|y^{(5)}(t)|$ is small. The Runge-Kutta method is sufficiently accurate for most applications.

The Runge-Kutta method computes approximate values $y_1, y_2, \ldots, y_n$ of the solution of the initial value problem given by Equation (19.48) at t_0, $t_0 + \Delta t$, ..., $t_0 + n\,\Delta t$ as follows. First, given y_i, compute the slope of the tangent line to the solution curve at four points, i.e.

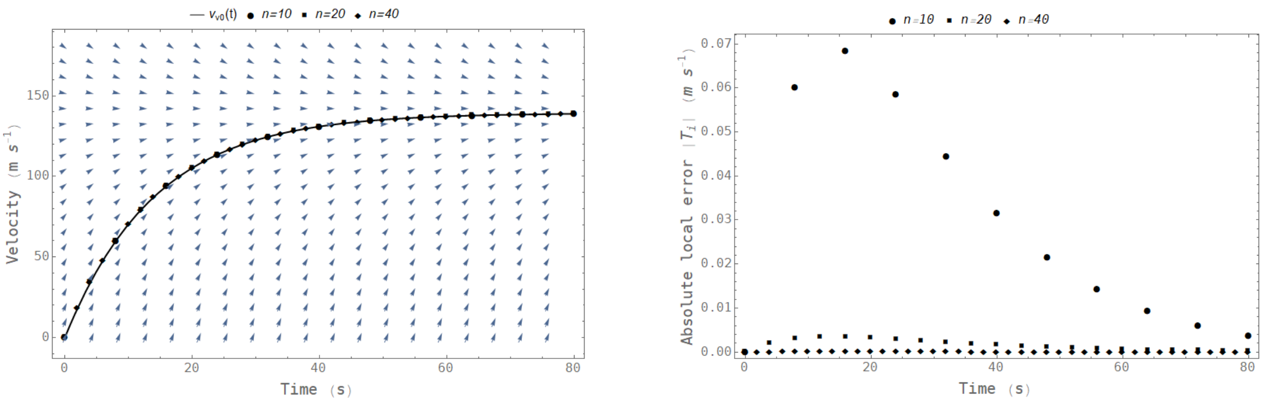
$$\begin{aligned}
m_{1i} &= f(t_i, y_i), \\
m_{2i} &= f\left(t_i + \frac{\Delta t}{2}, y_i + \frac{\Delta t}{2} m_{1i}\right), \\
m_{3i} &= f\left(t_i + \frac{\Delta t}{2}, y_i + \frac{\Delta t}{2} m_{2i}\right), \\
m_{4i} &= f(t_i + \Delta t, y_i + \Delta t m_{3i}).
\end{aligned}$$

Finally, compute y_{i+1} on the basis of a weighted average of these four slopes:

$$y_{i+1} = y_i + \frac{\Delta t}{6}\left(m_{1i} + 2\,m_{2i} + 2\,m_{3i} + m_{4i}\right). \tag{19.58}$$

Example 19.16

For what concerns Equation (17.4), Figure 19.11 shows its analytical solution and corresponding numerical solutions for 10, 20 and 40 steps if $v_0 = 0$, together with the corresponding local absolute truncation errors. From these plots it is again clear that the local absolute truncation error decreases as the number of steps increases. Comparing this figure with the one obtained when using Euler's method, it is obvious that the Runge-Kutta method is much more accurate because the magnitude of the local truncation is always much lower than in the case of Euler's method, irrespective of the number of steps and time.



(a) (b)

Figure 19.11: Analytical solution and corresponding numerical solutions of Equation (17.4) satisfying $v(0) = 0$ obtained with the Runge-Kutta method for a different number of steps n (a), together with the corresponding local truncation errors (b).

19.8 Exercises

Assignment 19.1 — Solve the following differential equations by separating the variables. Determine the largest possible solution interval in each case.

(a) $-t^2\,dy = (y+3)\,dt$

(b) $y' = 1+y^2$

(c) $t^3\,dy + y^3\,dt = 0$

(d) $e^t y' = e^y$

(e) $t^2 y y' = e^y$

To determine the largest possible solution interval, we distinguish between linear and nonlinear first order differential equations. In the case of a linear first order differential equation, we put the differential equation in the form $y' + p(t)y = q(t)$ and examine the continuity of $p(t)$ and $q(t)$. For a nonlinear differential equation, we put it in the form $y' = f(t,y)$ and examine both the continuity of $f(t,y)$ and $\dfrac{\partial f(t,y)}{\partial y}$.

Assignment 19.2 — Solve the following exact differential equations.

(a) $\left(2ty+3t^2y^2\right)\,dt + \left(t^2+2t^3y\right)\,dy = 0$

(b) $(t+3y)\,dt + 3\,(t-y)\,dy = 0$

(c) $y'\left(t-\sin(y)\right)+y = 0$

(d) $\left(1+y\,e^{ty}\right)\,dt + \left(1+t\,e^{ty}\right)\,dy = 0$

Assignment 19.3 — Consider the differential equation

$$2y^{a-1}\left(1+\sin(2t)\right)y' + 3t^2 + y^a\cos(2t) = 0.$$

(a) For which values of a is the differential equation exact?

(b) Determine the general solution of this differential equation for the value of a found in a).

Assignment 19.4 — Consider the differential equation

$$\frac{y'}{y} = \frac{e^{2a}+\sin(t)+1}{\cos(t)-2\,e^{a}\,t-a}\,.$$

(a) For what value of a is the differential equation exact?

(b) Then determine the unique solution corresponding to initial value $y(0) = -1$.

(c) What is the largest open rectangle in the (t,y)-plane within which the conditions of Theorem 3.3 are satisfied?

Assignment 19.5 — Solve the linear differential equations below.

(a) $y' + 2y = 3t+1$

(b) $\left(1-t^2\right)y' + ty = t^3 \qquad \left(t \in]-1,1[\right)$

(c) $ty' - y = (t-1)\,e^t$

(d) $y'\sqrt{1-t^2} + y = 1$

Assignment 19.6 — Determine the general solution of the differential equations below.

(a) $(1+t)\,y\,dt + (1-y)t\,dy = 0$

(b) $(t+\sin(y))\,dt + (t\cos(y) - 2y)\,dy = 0$

(c) $\left(1+e^{3t}\right)dy + 3ye^{3t}\,dt = 0$

(d) $y' - y = 2(t-1)\cos(t)$

(e) $y' - ty = t^3$

(f) $y\,y' = t^2 + 2$

(g) $\left(y^2-1\right)y' = 2$

(h) $(2y+3)\,dt = (t+1)\,dy$

(i) $2ty\,dt + \left(3t^2+1\right)dy = 0$

(j) $\left(y - \dfrac{1}{1+t^2}\right)dt + t\,dy = 0$

(k) $\sin(t)\cos(t)\,dy = \left(\sin(t) + y\cos^2(t)\right)dt$

(l) $\left(1+e^{-t}\right)y'\sin(y) + \cos(y) = 0$

(m) $(y - \cos(t))\,dt + (t+1)\,dy = 0$

Assignment 19.7 — Consider the differential equation

$$y' + (\cot(t))\,y = 2\cos(t)\,.$$

(a) Determine the intervals over which Theorem 3.1 guarantees a unique solution.

(b) Determine the unique solution of the differential equation if $y(\pi/2) = 2$. Also, point out in which interval this is guaranteed to be the unique solution of the initial value problem by Theorem 3.1.

Assignment 19.8 — Consider the differential equation

$$y' + (\tan(t))\,y = \cos^2(t)\,, \quad y(0) = -1\,.$$

(a) Determine the largest t-interval over which Theorem 3.1 guarantees the existence of a unique solution.

(b) Solve the initial value problem.

(c) Determine the equilibrium points of the differential equation and their stability.

Assignment 19.9 — Determine the unique solution of the initial value problem

$$y' = \begin{cases} \cos^2(t), & \text{if } t > 2, \\ \dfrac{y^2}{3}, & \text{else,} \end{cases}$$

where $y(0) = 1$.

Assignment 19.10 —

(a) Solve Equation (19.16) (mass balance).

Assignment 19.11 —

(a) In a chemical reaction, one molecule of substance P and one molecule of substance Q together form one molecule of the substance R. There are p [–] molecules P and q [–]

molecules Q present at the start of the reaction. The number of molecules R at time t [T] is x. The rate at which R is formed at any time during the reaction is directly proportional to the product of the remaining numbers of molecules P and Q:

$$\frac{dx}{dt} = k(p-x)(q-x),$$

met $k > 0$. Determine x with respect to t if $x = 0$ for $t = 0$.

Assignment 19.12 —

(a) The disintegration rate $\frac{dN}{dt}$ $[T^{-1}]$ of a radioactive substance is directly proportional to N [–], the number of particles. In 1600 years, radiation converts half of what was originally present into converted. How long does it take to convert 5% of it?

Assignment 19.13 —

(a) A rocket with mass M (fuel included) is shot up vertically and undergoes frictional resistance directly proportional to the speed v achieved. Per unit of time, fuel with mass m is consumed. The products of combustion are emitted at a constant velocity $-v_0$ with respect to the rocket. The differential equation describing the change in velocity is given by:

$$\frac{dv}{dt} + \frac{K}{M-mt}v = -g + \frac{mv_0}{M-mt} \qquad (K \text{ is constant}).$$

Determine the function that expresses v as a function of t if the rocket departs from rest.

PART V

APPENDICES

A

Proofing Techniques

In this appendix we gather information to help you read, understand and construct definitions, theorems and proofs.

A.1 Definitions

A definition is a term conceived by humans and used as a shortcut for a complicated idea. For example, we say that an integer is even as a shortcut to say that if we divide this number by two, we get a remainder of zero. With a precise definition we can answer certain questions unambiguously. For example, have you ever wondered if zero was an even number? Now the answer should be clear as we have a precise definition of what we mean by the term even.

A single term can have multiple definitions. For example, they could say that the integer n is even if there is another integer k such that $n = 2k$. We call this an equivalent definition, because it categorizes even numbers in the same way as our first definition. Definitions are like two-way streets — we can use a definition to replace something rather complicated with its definition (if it fits) and we can replace a definition with its more complicated description. A definition is usually written as some form of implication, such as "If something-nice-happens, then party.' However, this also means that "If party, then something-nice-happens," even though this may not be formally stated. This is what we mean when we say a definition is a two-way street — it's actually two implications going in opposite "directions". Anyone (including you) can come up with a definition, as long as it's unambiguous, but the real test of a definition's usefulness is whether or not it's useful for describing interesting or common situations.

A.2 Theorems

Advanced mathematics is about understanding theorems. Reading them, understanding them, applying them, proving them. Each theorem is a shortcut — we prove something in general and then when we encounter a specific case that falls under the theorem, we can immediately say that we know something else about the situation by applying the theorem. In many cases, this new information can be obtained with much less effort than when we would not have the theorem at our disposal. The first step to arrive at an understanding of a theorem is realising that the statement of any theorem can be rewritten with statements of the form “If something-happens, then something-else-happens.” The “something-happens” part is the **hypothesis** or **assumption** (*hypothese*) and the “something-else-happening” is the **conclusion** or **decision** (*conclusie*). To understand a theorem, it helps to rewrite its statements using this construction. To apply a theorem, we verify that in a certain case “something-happens” and immediately conclude that “something-else-happens”. To prove a theorem, we must argue based on the assumption that the hypothesis is true and via the process of logic obtain that the conclusion must then be true as well.

B

Calculus in Mathematica and Wolfram|Alpha

The methods given in this course can be used to solve mathematical problems with pen and paper, but in practice this only works for relatively simple problems that require little computation, such as those covered in the board exercise sessions. Today, however, we usually leave the repetitive work to computers, which are specially designed to perform a huge number of calculations in a very short time.

Here we use Wolfram Mathematica (or Mathematica for short) and Wolfram|Alpha. Mathematica is a so-called computer algebra system capable of performing symbolic mathematical calculations on the computer. Wolfram|Alpha is freely accessible online at the following url: `https://www.wolframalpha.com/` and uses the same syntax as Mathematica, but is limited in available computation time. You can use both Mathematica and and Wolfram|Alpha to solve complex problems, but also to support and check your calculations when solving problems with pen and paper. First we give an introduction to using Mathematica, then we discuss in detail how Mathematica can be used to (help) solve calculus problems. For more extensive documentation, we refer you to the Mathematica Documentation Centre (you can find it under Help > Documentation).

B.1 Mathematica Notebooks

Mathematica uses so-called Notebooks, denoted by the .nb extension, an interactive document containing both formatted text and code. This document is structured in what are known as **cells**, indicated by straight brackets on the right side of the notebook. To create a new cell, move your cursor below/above one or between two existing cell(s) and start typing. Each cell has a particular style, which defines its properties and formatting. This style can be changed by right-clicking on the cell and then under Styles selecting the desired style. We give a brief overview of the most commonly used cell styles.

B.1.1 Input cell

When we create a new cell in the Mathematica notebook, the default cell style is **Input**. In these cells we can enter mathematical operations, which can be performed with Shift + Enter . If the computation time gets too high (which often indicates a bug in the code), the evaluation can be aborted by pressing keystrokeAlt + keystroke. or in the menu bar via Evaluation > Abort Evaluation. The result of the evaluated Input cell is written out to a linked Output cell, which appears below it.

```
In[45]:= 1+1

Out[45]= 2
```

B.1.2 Text and layout

Text cells contain formatted text and are used to provide explanations for the notebook. In addition, there are numerous cell types that help structure the notebook. Cell types such as **Title**, **Chapter**, **Section**. . . are arranged according to a clear hierarchy, where cell types higher up the ladder (e.g. Title), include all subordinate cell types (e.g. Section, Text and Input). This group of cells is indicated by a straight bracket on the right side of the document and can be closed with a double-click so that only the cell with the highest rank is shown.

B.2 Mathematica for dummies

This section goes over the basics of the Mathematica language. Feel free to modify the following examples or experiment on your own! If something is unclear, you can consult the Documentation Centre via the Help menu in the taskbar at the top of the document or via a right-click at the level of a specific command.

B.2.1 Operations, evaluations and lists

B.2.1.1 Mathematical and realtional operations

$+$, $-$, $*$, $/$, $\wedge$	Basic mathematical operators: sum, difference, multiplication, division, exponentiation
$==, != , >, >=, <, =<$	relational operators: (un)equality, bigger than (or equal to), less than (or equal to)
()	brackets to group operations

Below we illustrate the use of some of these operators.

```
In[46]:= (10-5)*2^3

Out[46]= 40

In[47]:= 1 < 2 ≤ 3

Out[47]= True
```

```
In[48]:= 5/2 == 10/3

Out[48]= False
```

B.2.1.2 Evaluation of expressions

expr	the exact result is shown in an Output cell below after executing the expression *expr*
expr//N or N[*expr*]	the numerical (approximate) result is given after executing the expression expr
N[*expr*,*n*]	the numerical (approximate) result is given after executing the expression expr; With a precision of *n* significant numbers in the Output cell
expr;	the result is not displayed after executing the expression *expr*

B.2.1.3 Assignements

x = *value*	direct assignement (*x* will from now on always be replaced by *value*)
x=*y*=*value*	Assume *x* and *y* both equal to *value*
x=. or Clear[x]	deletes the value assigned to *x*
expr /. *x* -> *value*	replace all *x*'s in the expression *expr* by value

It should be noted that *value* need not be a scalar value, but can equally be a symbolic expression. In the last expression, the order to assign is indicated by /. and *x* -> *value* is the rule that determines what should be replaced (the so-called. **replacement rule**).

We illustrate all this in the example below.

Example B.1

Save the expression $xy^2 - 2x(1+y^{-1})$ in the variable `expr`:

```
In[49]:= expr =  x*y^2-2*x(1+1/y)
```

Out[49]= $-2x\left(1+\frac{1}{y}\right) + xy^2$

Exact and numerical evaluation of `expr`, voor $x = y = 1/3$, by direct assignment:

```
In[50]:= x = y = 1/3;
         expr

Out[50]= -71/27

In[51]:= expr // N

Out[51]= -2.62963

In[52]:= N[expr,2]

Out[52]= -2.6
```

Evaluation of `expr`, for $x = 3, y = 2$ and $x = uv, y = u^2$, by exchange:

```
In[53]:= Clear[x, y];
         expr/.{x→3, y→2}
         expr/.{x→u*v, y→u^2}
         expr

Out[53]= 3

Out[54]= -2( 1+1/u^2 )uv + u^5 v

Out[55]= -2x( 1+1/y ) + xy^2
```

B.2.1.4 Lijsten

Often we want to work with several objects (values, functions ...) at the same time. This can be done by using lists:

List[e1, e2, ...] of {e1, e2, . . . } ordered sequence of elements e1, e2 . . .

Note that these elements can also be Lists themselves. Lists of lists are called **nested lists**. Lists can be used to pass multiple arguments to functions (see below), but can also be considered vectors.We can select elements from a list. For example, consider the vector v, then

$v[[k]]$ element k in v

Example B.2

Create the vector $\vec{\mathbf{v}}_1 = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T$ and select the first element.

```
In[56]:= v1 = {1, 2, 3};
         First[v1]

Out[56]= 1
```

B.2.2 Functions

Functions with arguments x, y, etc. are called as follows:

f[x,y,. . .]

and entered:

f[x_,y_,. . .] := *expr*,

where, *expr* is an expression with the variables x, y . . . .

Implicit functions can also be defined, but this requires specifying the dependent variables in *expr* (see example).

Piecewise functions are implemented as follows:

f [*x*_, *y*_, ...] := Piecewise[*expr1*, *cond1*, *expr2*, *cond2*, ..., Indeterminate]

where *expr1, expr2,. . .* are expressions with variables *x, y,. . .* and *cond1, cond2,. . .* the conditions they must satisfy. Indeterminate indicates that in the regions where the variables do not satisfy any of the conditions, f is undefined.

In addition to self-defined functions, Mathematica has numerous built-in functions:

Log[x]	$\ln(x)$
Log[x,b]	$\log_b(x)$
Exp[x]	e^x
Sin[x], Cos[x], Sec[x], Csc[x],Tan[x], Cot[x]	trigonometric functions
ArcSin[x], ArcCos[x], ArcSec[x], ArcCsc[x]	inverse trigonometric functions
ArcTan[x], ArcCot[x]	

Note that Mathematica can work both symbolically and numerically. It always works with exact values, unless it is explicitly stated that it should work numerically. The latter can be done with `//N` or `N[expr]`, or by using a decimal.

Finally, we note that there are numerous other functions available in Mathematica, such as the previously cited **`Piecewise`** or the function `Manipulate`, with which we can generate interactive ouput.

Manipulate[*expr*, {k, k_0, k_1}]	gives an interactive result of expr in which we can adjust the value of k through a controller.

The values we can get k to assume lie between k_0 and k_1.

Example B.3

Calculate $e^{\ln(x)} = x$:

```
In[57]:= Exp[Log[x]]

Out[57]= x
```

Exact and numerical evaluation of $f(x) = \sin(x)^2/5$ for $x = 5$:

```
In[58]:= f[x_] := 2/10*Sin[x]^2
         f[5]
         f[5] // N

          Sin[5]^2
Out[58]= ---------
             5

Out[59]= 0.183907
```

Once more the function: $f(x) = \sin(x)^2/5$, but numerically defined:

```
In[60]:= f[x_] := 0.2*Sin[x]^2
         f[5]

Out[60]= 0.183907
```

Implicitly defined function $\sin(y) + y^3 = 6 - x^3$:

```
In[61]:= fImpl[x_] := Sin[y[x]] + y[x]^3 == 6 -x^3
```

Remark that x is the only independent variable.

```
In[62]:= fImpl[x]
         fImpl[2]

Out[62]= Sin[y[2]] + y[x]^3 == 6-x^3

         Sin[y[2]] + y[2]^3 == 6-2^3
```

A piecewise defined function:

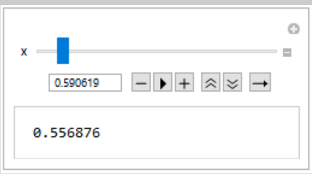
```
In[63]:= fpw[x_] := Piecewise[{{x+1, x<0},{-x^2+1, x>0}},Indeterminate]
         fpw[-.5]
         fpw[0]
         fpw[.5]

Out[63]= 0.5
         Indeterminate
         0.75
```

Interactive solution of Sin[x] for $x \in [0, 2\pi]$:

```
In[64]:= Manipulate[Sin[x], {x, 0, 2*Pi}]
```

Out[64]=



B.2.3 Solving (un)equality's

Solve[*expr*, *vars*]	tries to find solutions of the equation, inequality, or system of equations/inequalities in *expr* as a function of the variable *vars*
NSolve[*expr*, *vars*]	tries to find a numerical approximation of the solutions of the equation, inequality, or system of equations/inequalities in *expr* to the variables in *vars*
Simplify[expr]	ries to find a simplified form of the expression in expr
Apart[expr]	splits the expression in expr into partial fractions

The results of **Solve** and **NSolve** are given as lists of replacement rules.

Example B.4

Find the exact and approximate values of the zero points of $x^2 - 2$:

```
In[65]:= ZerosExact = Solve[x^2 - 2 == 0, x]
         ZerosNumeriek = NSolve[x^2 - 2 == 0, x]

Out[65]= {{x→-√2},{x→√2}}

         {{x→-1.41421,{x→1.41421}}
```

Retrieve the replacement rule of the first zero point from the list of solutions :

```
In[66]:= nulpuntenExact[[1, 1]]

Out[66]= x→-√2
```

The output of an unsolvable equation or system is an empty list.

```
In[67]:= Solve[{x^2-2 == 0, 2x-5 == 0}, x]

Out[67]= {}
```

Split the fraction $\frac{x^2-2}{x+4}$ in partial fractions:

```
In[68]:= Apart[(x^2 - 2)/(x + 4)]

                          14
Out[68]= -4 + x + -------
                        4 + x
```

B.2.4 Visualisatie

Finally, we go over some of the functions that serve to create plots.

Plot[$f[x]$,{x,a,b}]	Create a plot of the function f over the interval $[a,b]$
Plot[{$f_1[x]$,$f_2[x]$,...},{x,a,b}]	Creat a plot of the functions f_1 , f_2, ... over the interval $[a,b]$.
Plot3D[$g[x,y]$, {x,a,b}, {y,c,d}]	Create a 3D plot of g over the range $[a,b]\times[c,d]$.
ListPlot[{{x_1,y_1},{x_2,y_2},...}]	Create a plot for the points (x_i,y_i).
ContourPlot[$g[x,y]$, {x,a,b}, {y,c,d}]	plot the contour(s) of g over the range $[a,b]\times[c,d]$ (g can also be implicitly defined)
RegionPlot[*cond*, {x,a,b}, {y,c,d}]	plot the subarea of $[a,b]\times[c,d]$ where the conditions in *cond* are met
ParametricPlot[{$k[u],l[u]$}, {u,u_{min},u_{max}}]	plot the parametric equations $x=k(u)$ and $y=l(u)$ for $u\in[u_{min},u_{max}]$

There are numerous options for customizing plot formatting. Below we give an overview of the most important ones:

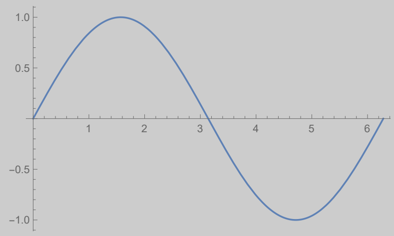
PlotLabel -> "title"	gives a title to the plot
AxesStyle -> Arrowheads[*s*]	places arrows on the axes of the plot (pointing in an increasing sense). *s* is a number that determines the size of the arrows.
AxesLabel -> {"x","y"}	labels the axes
PlotRange -> {*ymin, ymax*}	specifies the y-range of the plot
ImageSize -> grootte	specifies the size of the plot
PlotStyles -> {*stijl1*, *stijl2*, ...}	plots the first function in *stijl1*, the second in *stijl2* ...
PlotLegend -> {*name1*, *name2*, ...}	creates a legend with the names of the plotted function

Example B.5

Plot $\sin(x)$ the default layout:

```
In[69]:= Plot[Sin[x],{x,0,2*Pi}]
```

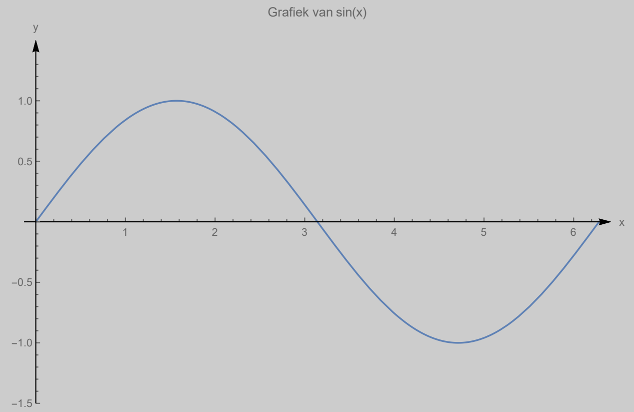
Out[69]=



Plot $\sin(x)$ with custom layout:

```
In[70]:= Plot[Sin[x],{x,0,2*Pi},
         PlotRange → {-1.5, 1.5},
         PlotLabel → "Grafiek van sin(x)",
         AxesStyle → Arrowheads[0.02],
         AxesLabel → "x", "y",
         ImageSize → Large ]
```

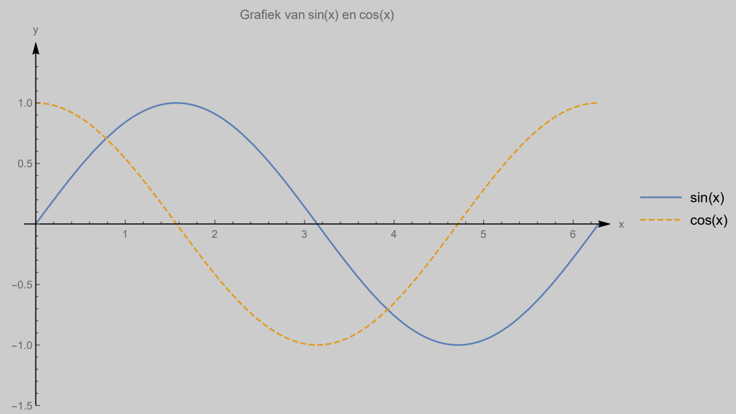
Out[70]=



Plot sin(x) en $cos(x)$ with custom layout, where $cos(x)$ is drawn as a dotted line:

```
In[71]:= Plot[{Sin[x], Cos[x]}, {x, 0, 2*Pi},
         PlotRange → {-1.5, 1.5},
         PlotLabel → "Graph of sin(x) and cos(x)",
         AxesStyle → Arrowheads[0.02],
         AxesLabel → {"x", "y"},
         ImageSize → Large,
         PlotStyle → {Line, Dashed},
         PlotLegends → {"sin(x)", "cos(x)"}]
```

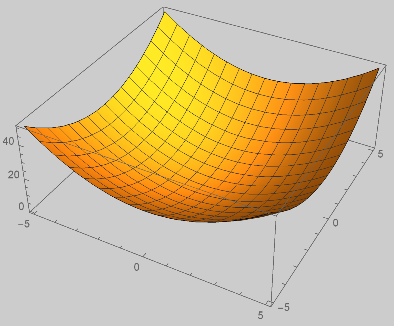
Out[71]=



Consider the function $g(x, y) = x^2 + y^2$ over $[-5, 5] \times [-5, 5]$. Create a 3D and contour plot:

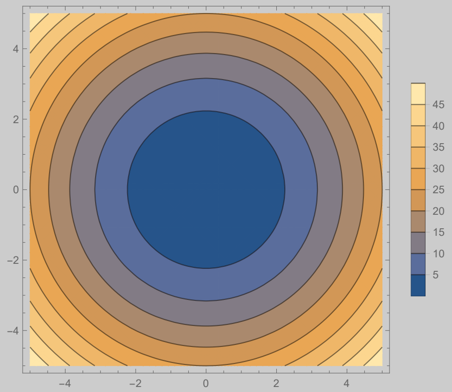
In[72]:=
```
Plot3D[x^2 + y^2, {x, -5, 5}, {y, -5, 5}]
```

Out[72]=



In[73]:=
```
ContourPlot[x^2 + y^2, {x, -5, 5}, {y, -5, 5}]
```

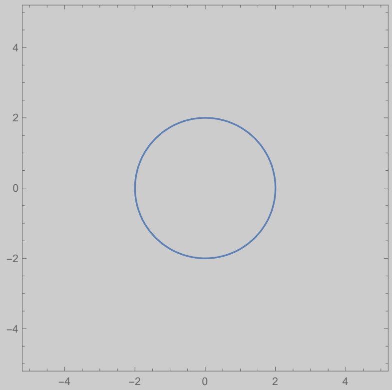
Out[73]=



Plot the circle of radius 2, using the implicit equation and the parametric equation:

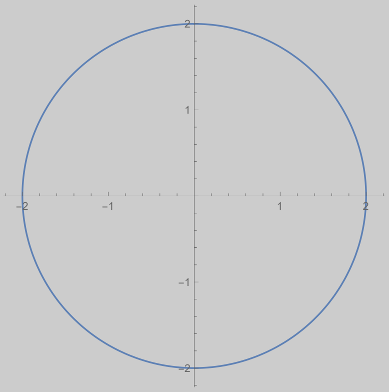
```
In[74]:= ContourPlot[x^2 + y^2 == 4, {x, -5, 5}, {y, -5, 5}]
```

Out[74]=



```
In[75]:= ParametricPlot[{2*Cos[u],2*Sin[u]}, {u, 0, 2*Pi}]
```

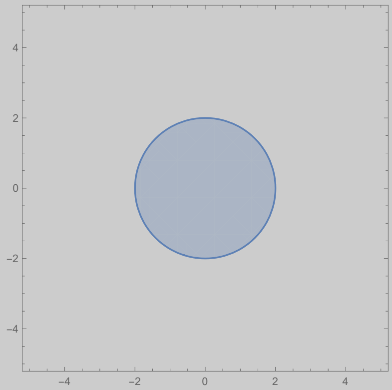
Out[75]=



Plot the subarea of $[-5,5] \times [-5,5]$, delimited by the circle of radius 2:

```
In[76]:= RegionPlot[x^2 + y^2 ≤ 4, {x, -5, 5}, {y, -5, 5}]
```

Out[76]=



B.3 Calculus specific intstructions

B.3.1 Limits

In Mathematica we can calculate limits with the function **`Limit`**.

Limit[$f[x]$, x -> c]	determine the limit of f in c
Limit[$f[x]$, x -> c, Direction -> "FromAbove"]	determine the right limit of f in c
Limit[$f[x]$, x -> c, Direction -> "FromBelow"]	determine the right limit of f in c

Example B.6

Determine the total, right and left limit of $1/x$ in 0:

```
In[77]:= Limit[ 1/x, x → 0]

Out[77]= Indeterminate

In[78]:= Limit[ 1/x, x → 0, Direction → "FromAbove"]

Out[78]= ∞

In[79]:= Limit[ 1/x, x → 0, Direction → "FromBelow"]

Out[79]= -∞
```

Determine the limit of $1/x$ for $x \to +\infty$:

```
In[80]:= Limit[ 1/x, x → Infinity]

Out[80]= 0
```

B.3.2 Derivatives

Derivatives of functions with one variable can be calculated using **`Derivative`**:

Derivative$[n][f][x]$	n-th order derivative of the function $f(x)$ towards x

Usually we will be using the short notation of **`Derivative`**:

$f'[x]$	first order derivative of $f(x)$
$f''[x]$	second order derivative of $f(x)$
D$[expr, x]$	first order partial derivative of *expr* to x
D$[expr, \{x, n\}]$	n-th order partial derivative of *expr* to x

The difference between these notations is subtle and often they can be used interchangeably. However, we recommend to use $f'[x]$ when possible. Remark that $f(x)$ can be implicitly defined as well.

Example B.7

Determine the first and second derivatives of $x^2 + x$:

```
In[81]:= f[x_] := x^2 + x;
         f'[x]
         f''[x]

Out[81]= 1 + 2x
         2
```

We can keep the derivative as a function:

```
In[82]:= df[x_] := f'[x]
         df[x]

Out[82]= 1 + 2 x
```

Determine the derivative of the implicit function $y^3 + y\sin = 6 - x^3$:

```
In[83]:= fImpl[x_] := Sin[y[x]] + y[x]^3 == 6 - x^3
         fImpl'[x]

Out[83]= Cos[y[x]] y'[x] + 3 y[x]^2 y'[x] == -3 x^2
```

Try to rewrite the outcome in explicit form:

```
In[84]:= Solve[fImpl'[x], y'[x]]

Out[84]= {{y'[x]→-3x^2/(Cos[y[x]]+3y[x]^2)}}
```

With **`Derivative`** we can also determine derivatives of functions of multiple variables:

Derivative$[n_1, n_2, \ldots][f][x_1, x_2, \ldots]$	derivative of the function $f(x_1, x_2, \ldots)$ that is derived n_i times with respect to the variable x_i.

The short notation $f'[x_1, x_2, \ldots]$ can no longer be used for functions of multiple variables since it does not indicate to which variable(s) it should be derived. **`D`** can be used:

D$[expr, \{x_1, n_1\}, \{x_2, n_2\}, \ldots]$	derivative of *expr*, that is n_i derived with respect to the variable x_i,

where *expr* may again be explicitly or implicitly defined. For functions with multiple variables, we can also compute the gradient:

Grad$[f, \{x_1, x_2, \ldots\}]$	gradient of f

Example B.8

Determine $f_x(x, y)$ and $f_{xy}(x, y)$ if $f(x, y) = x^2 + xy + y^2$.

```
In[85]:= D[x^2 + x y + y^2, {x, 1}]

Out[85]= 2x + y

In[86]:= D[x^2 + x y + y^2, {x, 1},{y, 1}]

Out[86]= 1
```

Determine the gradient of $f(x,y) = x^2 + xy + y^2$ in the point $(1,2)$.

```
In[87]:= gGrad = Grad[x^2 + x y + y^2, {x, y}]

Out[87]= {2x+y, x+2y}

In[88]:= gGrad /. {x→1, y→2}

Out[88]= {4, 5}
```

B.3.3 Integrals

Mathematica has two functions for integrating functions, being **`Integrate`** and **`NIntegrate`**. The former calculates an integral analytically, while the latter provides a numerical approximation.

Integrate$[f, x]$	determines the indefinite integral $\int f(x)dx$,
Integrate$[f, \{x, x_{min}, x_{max}\}]$	determines the definite integral $s\int_{x_{min}}^{x_{max}} f(x)dx$,
NIntegrate$[f, \{x, x_{min}, x_{max}\}]$	determines the numerical approximation of $\int_{x_{min}}^{x_{max}} f(x)dx$,

Example B.9

Determine the following integral

$$\int \left(4x - x^2\right) dx.$$

```
In[89]:= Integrate[4x-x^2,x]

Out[89]= 2x^2 - x^3/3
```

Determine

$$\int_0^{-\infty} e^x dx.$$

```
In[90]:=  Integrate[Exp[x],{x,0,-Infinity}]

Out[90]=  -1
```

Determine the numerical value of the definite integral

$$\int_0^1 \frac{\sin(x)}{x} dx.$$

```
In[91]:=  NIntegrate[Sin[x]/x,{x,0,1}]

Out[91]=  0.946083
```

B.3.4 Series

We use following Mathematica functions to determine the Taylor series expansion of a function:

Series[$f,\{x,x_0,n\}$]	gives the taylor series expansion of $f(x)$ around x_0 with terms up to and including the n-the order (+ the error term of the $n+1$th order)
Normal[s]	Returns round the Taylor development s to the nth order

Example B.10

Determine the Taylor series expansion of $\ln(x)$ in the area of $x=1$ up to the 2—the order term.

```
In[92]:=  s = Series[Log[x], {x, 1, 2}]

Out[92]=  (x-1) - 1/2 (x-1)^2 +O[x-1]^3

In[93]:=  Normal[s]

Out[93]=  -1 - 1/2 (-1+x)^2 +x
```

C

Python Tutorial

This tutorial is available as a Jupyter notebook (extension .ipynb). Note that this tutorial is made for **Python 3.x** and is not compatible with older Python versions. Make sure you are working with an appropriate version!

C.1 What are Python and Jupyter Notebooks?

Python is a programming language used to write computer programs. These programs consist of lines of code that can be interpreted and executed by a computer one by one. Like all (programming) languages, Python has its own vocabulary and grammar. This is called the **syntax**, i.e. the rules that the code must follow in order for computers to be able to read it.

Here we use Python in a **Jupyter Notebook**, available through a web browser (Jupyter.org). Jupyter Notebooks are structured in so-called cells. We distinguish two types of cells:

1. **Markdown cells** contain text with titles, explanations, assignments,.... This text is entered as code, which is converted to formatted text when executed.
2. **Input cells** contain Python code that can be executed. In front of such cells you see `In [x]:`, where x keeps track of the number of cells executed. When the cell is executed, the output(s) appear below the input cell. These are then preceded by `Out [x]:`.

Two modes are possible for a cell:

1. **Command Mode** (left bar is blue): in this mode, operations can be done on the entire cell (e.g. changing cell type, cutting and pasting cells, inserting cells,...).
2. **Edit Mode** (left bar is green): in this mode, the text/code in the cell can be edited.

Press Enter to switch to edit mode and press Esc to switch back to command mode. Use Ctrl + Enter to enter a cell, and Shift + Enter to select the cell below after entering. An overview of commands for both modes is listed under Help > Keyboard Shortcuts.

Question 1 What should you enter to create a new Input cell under an existing cell? And in what mode should you enter this? Enter your answer below and convert this cell back to formatted text.

> Answer: . . .

Question 2 Create an Input cell below with the command

```
print("Hello, world!")
```

and enter this cell.

C.2 Python for dummies

C.2.1 Objects

Python is a so-called object-oriented programming language, which means that we perform operations on objects. The operations covered here are mainly mathematical ones. Objects are data stored in computer memory. Each object has a particular class, which determines which operations can be performed on it. The object class can be checked with the command `type`.

```
>>> type(1)
int
>>> type(1.5)
float
```

The examples above are both numbers, yet belong to different object classes! `1` is an integer, whereas `1.5` is a is a so-called floating point number. Note that Python distinguishes these two classes from each other by the presence or absence of a decimal sign.

```
>>> type(1.)
float
```

Besides number objects, other types exist as well. A first example is a so-called `list`. This datatype stores a set of objects together in an ordered sequence, indicated by square brackets `[]`. A second example is the data type `string`, which is used to store text. Strings are indicated by single (`'`) or double (`"`) quotation marks. Text not between quotation marks is interpreted as a command.

```
>>> [1,1.5,2]
[1,1.5,2]

>>> type([1,1.5,2])
list

>>> type("This is a string.")
str
```

```
>>> type('This is a string as well.')
str

>>> This is not a string and will thus result in an error.
        File "<ipython-input-5-ec99100478b7>", line 1
          This is not a string and will thus result in an error.
                           ^
     SyntaxError: invalid syntax
```

C.2.2 Variables

Often we want to store a certain object (temporarily), in order to use it further. To do this, we use variables. A variable is a name we assign to a particular object in computer memory. This is done as follows: `variable_name = object`.

```
>>> a=1
    a
1
```

In the cell above, a value of 1 is assigned to the variable a on the first line. This operation gives no output. When we call a again, we get a value of 1 as output. Keep in mind the order! When assigning, the name of the variable must always be to the left and that of the object to the right of "=". Otherwise we get an error message.

```
>>> 1 = a
        File "<ipython-input-14-7596acd8e627>", line 1
         1 = a
              ^
      SyntaxError: can't assign to literal
```

Variables can also be overwritten, which means that we assign a new object to the name of the variable.

```
>>> a = 2
    a
2
```

Moreover, to determine the new object in the variable, reference can be made to the current object linked to the variable.

```
>>> a = a+1
    a
3
```

We can also delete a variable.

```
>>> del a
    a
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-22-ef9d13752aff> in <module>()
----> 1 del a
2 a

NameError: name 'a' is not defined
```

After closing the Python notebook, all variables are deleted from the memory. So, when restarting the notebook, we will have to enter all the (necessary) input cells again before we can continue working.

C.2.3 Mathematical operations

When we want to perform mathematical operations on number objects we can use following symbols:

$+$, $-$	addition and subtraction
$*$, $/$	multiplication and division
$**$	exponentiation
$()$	grouping of elements

The classical order of operations applies for parenthesis, power-law, multiplication/division, addition/-subtraction.

```
>>> a = 1
    a + 2
3
>>> b = 2
    c = a+b
    c
3
>>> b**(a+c)
16
```

C.2.4 Logical operations

The following symbols allow logical operations:

$<$, $>$	greater and less than
$<=$, $>=$	greater and less than or equal to
$==$, $!=$	equal to or not equal to

The result of such an operation is an object of the class `bool`, the Boolean numbers. These can take on only two values, viz. **`True`** (1) and **`False`** (0). Boolean numbers are widely used to control programs with so-called Control statements (see Section C.2.7).

```
>>> a == 1
True
>>> 1 == a
True
>>> d = a>b
    print(d)
    type(d)
False
bool
```

C.2.5 List operations

The object type `list` comes with a number of class specific operations. The most important is the so-called indexing, which allows us to retrieve an element from the list based on its rank. This is done using brackets.

```
>>> l = [1,2,3,4,5]
    l[0]
1
>>> l[2]
3
```

Note that indexing in Python always starts at 0!

Moreover, the function `sum` can be used to sum the elements in a list.

```
>>> sum(l)
15
```

C.2.6 Python functions

In addition to the (symbolic) operators, there are numerous built-in Python functions to perform operations with. These are called using:

`functiename(arguments)`.

If the function takes multiple (n) arguments, it can be called in two ways:

1. `functionname(argument_1, ... ,argument_n)`,
2. `functionname(argument_1 = argument_1, ... ,argument_n = argument_n)`,

where `argument_i` is the name of argument i. We can also define functions ourselves.

```
    def functionname(argument_1, argument_2, ..., argument_n):
        (operations with arguments)
        return outputs
```

The operations to be performed by the function may span several lines of code. To indicate which expressions belong to the function, expressions that belong together are aligned in the same way using (*tabs*).

It is often necessary to include comments in our code. This way, we make the code not only readable for others, but also for ourselves when we look back after a while. In Python, we can use # at the beginning of a line to indicate that it should not be interpreted as code. When comments span multiple lines, they are between """.

It is important to note that functions work with local memory, which is separate from the notebook's global memory. The variables in local memory are included as arguments or defined in the function itself. Once is executed, everything following `return` is returned to the notebook, after which the local function memory is closed.

This is made clear by the following example.

```
>>> a = 2       # a is defined in the notebook memory

        def f(x,y):
                """
                This function calculates the sum of the squares of x and y.
```

```
            Inputs: x, y
            """
            z = x**2+y**2 # local variable z is the sum of the squares of x and y
            a = 10        # variable a is defined in local function memory
            return z      # just the value of z (and thus not the variable)
                          # is returned to the notebook memory

# we can now call the function in two ways for e.g. x=2 and y=3
>>> f(2,3)
13

>>> f(x=2,y=3)
13


>>> x   # x was only defined in local memory
        # and thus does not exist in the Notebook memory.
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-34-401b30e3b8b5> in <module>()
----> 1 x

NameError: name 'x' is not defined


>>> z   # z was only defined in local memory
# and therefore does not exist in notebook memory
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-35-a8a78d0ff555> in <module>()
----> 1 z

NameError: name 'z' is not defined


>>> a   # a was already defined in Notebook memory,
        # but has thus remained unchanged after the function evaluation
2
```

Finally, we mention that while defining a function, we can already pass so-called *default* values with the arguments. This makes these arguments optional, since the function has a default value for them.

```
#example:
def f(x,y=2):
        return x*y

>>> f(2)
4

>>> f(2,3)
6
```

Question 3 Write a function to find the weighted sum of two numbers. Where p is the weight of the first number and $(1-p)$ is the weight of the second. If no value p is given, the average of the two numbers should be returned.

C.2.7 Control statements

Lastly, we introduce two more so-called **control statements** (the **while** and **for** loop) and a **decision statement** (**if**), which are frequently used in programming. These statements define the so-called **flow of the program**, i.e. the way code is interpreted. By default this is done sequentially, line by line, but control statements can ensure that certain parts of the code are repeated or skipped.

C.2.7.1 While-loop

First we have the **while-loop**. This performs a number of operations as long as a certain condition is met and thus yields True as the result. The Python syntax is as follows:

```
        while condition:
              operations
```

The expressions within the while loop are aligned in the same way as functions. The control flow is illustrated in Figure C.1.

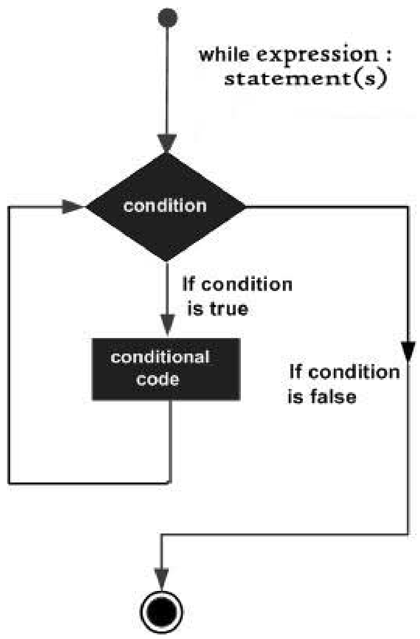


Figure C.1: The control flow of a while loop

Below we provide a simple example.

```
#Example
>>> x = 10       # assign an initial value to x
    print(x)     # prints the initial value
    while x > 1: # as long as x is bigger than 1:
        x=x/2    #       - divide x by 2
        print(x) #       - print the new value of x

 10
 5.0
 2.5
```

```
1.25
0.625
```

Note that if the condition is always met, the program will be stuck in a while loop and must be aborted manually. This can be done via **Kernel > Interrupt**.

C.2.7.2 For-loop

A second control statement is the **for-loop**. This repeats, like the while loop, a number of operations. However, the number of iterations here is predefined by the so-called **iterator**. This iterator iterates through the elements of a specified set of data. In Python, we write a for loop as:

```
        for element in iterator:
            operations
```

The expressions within the for-loop are also grouped together. The control flow is illustrated in Figure C.2.

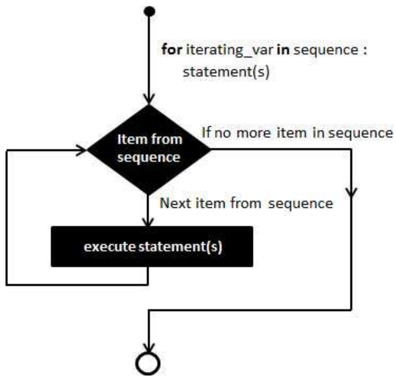


Figure C.2: Controlflow of a for-loop.

```
#example
>>> for value in [1,2,3,4]:
        print(value)

 1
 2
 3
 4
```

A for-loop can be used to quickly create lists.

```
        new_list = [operation with element for element in iterator]
```

```
#example
>>> l_1 = [1,2,3,4]
    l_2 = [value**2 for value in l_1] #squares the values in list l_1
    l_2

[1, 4, 9, 16]
```

C.2.7.3 If, else-control statement

Finally, we discuss the decision structure **if**. This imposes a certain condition that must be met before a sequence of operations may be performed. In Python, we enter this as follows.

```
if condition == True:
    operations
```

The control flow is illustrated in Figure C.3.

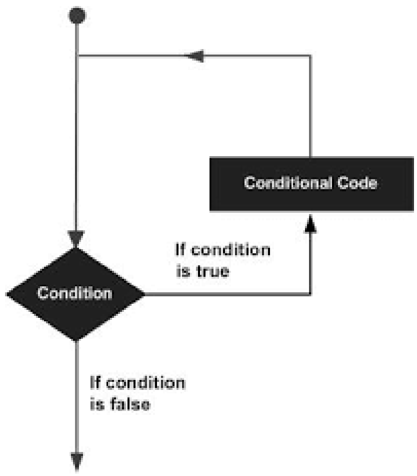


Figure C.3: Control flow of an if-statement.

In order to perform different operations depending on whether a condition is met or not, we can also use **else**. Then, we specify the operations to be performed if the condition is not met.

```
if condition == True:
    operations # are executed if condition was met
else:
    operations # are executed if condition was not met
```

The expressions within if and else are aligned in the same way.

```
#Example
>>> cd = True
    if cd:
        print("The condition was met")
    else:
        print("The condition was not met")

    The condition was met
```

C.3 Packages

In addition to the basic syntax, Python has numerous additional functionalities. These are collected in so-called **packages**, which can be loaded via:

```
import package as packagename
```

This makes that all the definitions in the package are stored in the variable `packagename`. Next, when we want to use a function from the imported package, we do so as follows:

```
packagename.functionname()
```

We can also import the function straight from the package using `from`:

```
from package import functiename
```

after which we can call the function directly:

```
functiename()
```

The input cell below loads a number of packets.

```
import numpy as np
import matplotlib.pyplot as plt
from ipywidgets import interact, fixed
```

C.3.1 Numpy

Numpy is the package that provides the foundation for scientific programming in Python. It contains numerous mathematical functions and numbers, such as:

- trigonometric functions: `np.sin(), np.cos() ...`
- exponential and logarithmic functions: `np.exp(), np.log()`
- the number π : `np.pi`

C.3.2 Sympy

C.3.3 Other Packages

In addition to Numpy, a number of other packages can be loaded:

- `matplotlib.pyplot` to make figures
- `ipywidgets` to make these figures interactive

D

Answers to the exercises

Chapter 2

Assignment 2.1 —

(a) $\forall\, x \in \mathbb{Z},\, \exists\, y \in \mathbb{Z} : \, x \geq y$

(b) $\exists\, y \in \mathbb{Z},\, \forall\, x \in \mathbb{Z} : \, x \geq y$

(c) $\exists\, x \in \mathbb{Z},\, \forall\, y \in \mathbb{Z} : \, x + y \neq 0$

(d) $\forall\, y \in \mathbb{Z},\, \exists\, x \in \mathbb{Z} : \, x + y \neq 0$

Assignment 2.2 —

(a) The set A contains the natural numbers between 2 and 6. $A = \{3, 4, 5\}$

(b) The set B contains the positive fractions that are solutions of $2x^2 + x - 6 = 0$. $B = \left\{\frac{3}{2}\right\}$

(c) The set C contains the positive integers that are solutions of $x^2 - 5 = 0$. $C = \varnothing$

Assignment 2.3 —

(a) $A = \left\{a \;\middle|\; \frac{a}{2} \in \mathbb{N} \;\wedge\; a > 100\right\}$

(b) $A = \left\{(a, b) \;\middle|\; a, b \in \mathbb{Z} \;\wedge\; \frac{a}{2} \in \mathbb{Z} \;\wedge\; \frac{b+1}{2} \in \mathbb{Z}\right\}$

(c) $A = \left\{a \;\middle|\; a \in \mathbb{Z}_0 \;\wedge\; \frac{a}{3} \in \mathbb{Z}_0\right\}$

(d) $A = \left\{a \;\middle|\; a \in \mathbb{Q}^+ \;\wedge\; \sqrt{a} > 3\right\}$

(e) $A = \left\{a \;\middle|\; a \in \mathbb{R} \;\wedge\; \frac{a-2}{6} \in \mathbb{Z}\right\}$

Assignment 2.4 —

(a) $\{1,3,5,7,9,11,...\} = \{x \in \mathbb{N} | \, x \text{ is an odd number}\}$

(b) $\{x | \, x \text{ is a rose}\} \subset \{x | \, x \text{ is a flower}\}$

(c) $\{1,3,5,7,9\} \not\ni 2$

(d) $\{1\} \subset \{1,3,5,7,9\}$

(e) $\{1,3\} \not\subset \{\{1\},\{3\},\{5\},\{7\},\{9\}\}$ also $\{1,3\} \notin \{\{1\},\{3\},\{5\},\{7\},\{9\}\}$

(f) $\{1,3\} \subset \{1,3,5,7,9\}$

(g) $\{1,3,5,7,9\} \neq \varnothing$

(h) $\{1\} \in \{\{1\},\{3\},\{5\},\{7\},\{9\}\}$

(i) $\{1,3,\{5,7,9\}\} \not\ni 5$

Assignment 2.5 —

(a) $1 \in A$: True

(b) $\{1\} \in A$: True

(c) $\{1\} \subseteq A$: True

(d) $\{\{1\}\} \subseteq A$: True

(e) $2 \in A$: False

(f) $\{\{2\}\} \subseteq A$: True

(g) $\{\{2\}\} \subset A$: True

(h) $\{2\} \subseteq A$: False

Assignment 2.6 —

(a) $(A \cup B) \cap C = \{3,5\}$

(b) $A \cup (B \cap C) = \{1,2,3,4,5\} = A$

(c) $\overline{C} \cup \overline{D} = U$

(d) $\overline{C \cap D} = \overline{C} \cup \overline{D} = U$ (rules of De Morgan)

(e) $(A \cup B) \backslash C = \{1,2,4\}$

(f) $A \cup (B \backslash C) = \{1,2,3,4,5\} = A$

(g) $(B \backslash C) \backslash D = \{1\}$

(h) $B \backslash (C \backslash D) = \{1,2,4\}$

(i) $(A \cup B) \backslash (C \cap D) = \{1,2,3,4,5\} = A$

Assignment 2.7 —

(a) $A \cap (B \backslash A) = \varnothing$

(b) $(A \backslash B) \cup (A \cap B) = A$

(c) $\overline{A} \cup \overline{B} \cup \left(A \cap B \cap \overline{C}\right) = \overline{A \cap B \cap C}$

(d) $(A \cap B) \cup \left(A \cap B \cap \overline{C} \cap D\right) \cup \left(\overline{A} \cap B\right) = B$

(e) $(A \cap B) \cup \left(B \cap \left((C \cap D) \cup \left(C \cap \overline{D}\right)\right)\right) = (A \cap B) \cup (B \cap C)$

Assignment 2.8 — The numbering of the areas is done from left to right.

(a) $A \cup (B \cap C)$

(b) $((B \cap C) \backslash A) \cup ((A \cap C) \backslash B)$

(c) $A \backslash (B \cup C)$ of $(A \backslash B) \cap (A \backslash C)$

(d) $(A \cap C) \backslash B$

(e) $((A \cap C) \backslash B) \cup (B \cap D)$

(f) $(D \backslash (A \cup C)) \cup ((C \cap A) \backslash B)$

Assignment 2.9 —

(a) rational

(b) rational

(c) irrational

(d) irrational

(e) rational

(f) rational

Assignment 2.10 —

(a) $\sum_{j=1}^{99} x^j$

(b) $\sum_{j=1}^{25} \sqrt{2j+1}$

(c) $\prod_{j=1}^{13} \frac{j^2}{a+j}$

Assignment 2.11 —

(a) 15

(b) 14

(c) 65

(d) $-\frac{7}{12}$

(e) 165

(f) -2368450

(g) -482295

(h) 165

Assignment 2.12 —

(a) $\frac{n(3n-1)}{2}$

(b) $\frac{3n-7}{2n}$

(c) $\frac{n}{2}(6n^2-15n+11)$

(d) $\frac{(n+1)(2n-11)+24}{6n^2}$

(e) $(n!)^3$

(f) $\frac{n+1}{2n}$

Assignment 2.13 —

(a) $2a^m b^{2n} c^{3p}$

(b) $1+\sqrt{x}$

(c) $6-2\sqrt{2}-2\sqrt{3}+2\sqrt{2}\sqrt{3}$

(d) $\frac{19\sqrt{a}}{2b}$

(e) $\sqrt[6]{\frac{x+1}{x-1}}$

(f) $a+\frac{b}{2}$

(g) a^2+b^2

(h) $\frac{x^2-1}{x}$

(i) $x^{\frac{5}{2}}$

(j) $\frac{81}{256}a^{\frac{7}{4}}b^{-\frac{9}{2}}$

Assignment 2.14 —

	$z+w$	zw	z^2	z^{-1}	$\frac{z}{w}$	$\frac{w}{z}$	$\bar{z}$	$z\bar{z}$	$(\bar{z})^2$
(a)	$2+7i$	$8i$	$-5+12i$	$\frac{2-3i}{13}$	$\frac{3-2i}{4}$	$\frac{12+8i}{13}$	$2-3i$	13	$-5-12i$
(b)	1	$1-i$	$2i$	$\frac{1-i}{2}$	$-1+i$	$\frac{-1-i}{2}$	$1-i$	2	$-2i$
(c)	$5+2i$	$41+11i$	$-16-30i$	$\frac{3+5i}{34}$	$\frac{-29-31i}{53}$	$\frac{-29+31i}{34}$	$3+5i$	34	$-16+30i$
(d)	$2\sqrt{2}$	4	$-4i$	$\frac{\sqrt{2}+\sqrt{2}i}{4}$	$-i$	i	$\sqrt{2}+\sqrt{2}i$	4	$4i$
(e)	$-2\sqrt{3}i$	-4	$-2-2\sqrt{3}i$	$\frac{1+\sqrt{3}i}{4}$	$\frac{1+\sqrt{3}i}{2}$	$\frac{1-\sqrt{3}i}{2}$	$1+\sqrt{3}i$	4	$-2+2\sqrt{3}i$
(f)	$-\sqrt{2}$	1	$-i$	$\frac{-\sqrt{2}-\sqrt{2}i}{2}$	$-i$	i	$\frac{-\sqrt{2}-\sqrt{2}i}{2}$	1	i

Assignment 2.15 —

(a) $19-4i$

(b) $-4+11i$

(c) $-3+4i$

(d) $3+4i$

(e) 61

(f) $\frac{5-2i}{29}$

(g) $\frac{5-i}{13}$

(h) $\frac{-2}{5}$

Chapter 3

Assignment 3.1 —

(a) no function

(b) function

(c) no function

(d) no function

Assignment 3.2 —

(a) $f(x)=x^3-1$ and $g(x)=\frac{x+1}{x-1}$

a) $(f+g)(x)=\frac{x^4-x^3+2}{x-1}$ with $\operatorname{dom}(f+g)=\mathbb{R}\setminus\{1\}$

b) $(f-g)(x)=\frac{x(x^3-x^2-2)}{x-1}$ and $\operatorname{dom}(f-g)=\mathbb{R}\setminus\{1\}$

c) $(fg)(x)=(x^2+x+1)(x+1)$ with $\operatorname{dom}(fg)=\mathbb{R}\setminus\{1\}$

d) $\left(\frac{f}{g}\right)(x)=\frac{(x^3-1)(x-1)}{(x+1)}$ met $\operatorname{dom}\left(\frac{f}{g}\right)=\mathbb{R}\setminus\{-1,1\}$

(b) $f(x)=\frac{x}{2}$ and $g(x)=\frac{2}{x}$

a) $(f+g)(x)=\frac{x^2+4}{2x}$ with $\operatorname{dom}(f+g)=\mathbb{R}_0$

b) $(f-g)(x)=\dfrac{x^2-4}{2x}$ with dom $(f-g)=\mathbb{R}_0$

c) $(fg)(x)=1$ with dom $(fg)=\mathbb{R}_0$

d) $\left(\dfrac{f}{g}\right)(x)=\dfrac{x^2}{4}$ with dom $\left(\dfrac{f}{g}\right)=\mathbb{R}_0$

(c) $f(x)=x$ and $g(x)=\sqrt{x+1}$

a) $(f+g)(x)=x+\sqrt{x+1}$ with dom $(f+g)=[-1,+\infty[$

b) $(f-g)(x)=x-\sqrt{x+1}$ met dom $(f-g)=[-1,+\infty[$

c) $(fg)(x)=x\sqrt{x+1}$ with dom $(fg)=[-1,+\infty[$

d) $\left(\dfrac{f}{g}\right)(x)=\dfrac{x}{\sqrt{x+1}}$ with dom $\left(\dfrac{f}{g}\right)=]-1,+\infty[$

Assignment 3.3 —

(a) $f\circ g(0)=2$

(b) $g(f(0))=22$

(c) $f\circ f(-5)=5$

(d) $g(g(2))=-2$

Assignment 3.4 —

(a) $(g\circ f)(x)=\dfrac{1}{5x-2}$ with dom $(g\circ f)=\mathbb{R}\setminus\left\{\dfrac{2}{5}\right\}$

$(f\circ g)(x)=\dfrac{5}{x-2}$ with dom $(f\circ g)=\mathbb{R}\setminus\{2\}$

(b) $(g\circ f)(x)=|x|$ with dom $(g\circ f)=\mathbb{R}$

$(f\circ g)(x)=x$ with dom $(f\circ g)=\mathbb{R}^+$

(c) $(g\circ f)(x)=\sqrt[3]{1-x^3}$ with dom $(g\circ f)=\mathbb{R}$

$(f\circ g)(x)=1-x$ with dom $(f\circ g)=\mathbb{R}$

(d) $(g\circ f)(x)=|x^2-5|$ with dom $(g\circ f)=\mathbb{R}$

$(f\circ g)(x)=x^2-2x-3$ with dom $(f\circ g)=\mathbb{R}$

(e) $(g\circ f)(x)=\sqrt{4-|x|}$ with dom $(g\circ f)=[-4,4]$

$(f\circ g)(x)=\sqrt{4-x}$ with dom $(f\circ g)=]-\infty,4]$

Assignment 3.5 — $(g\circ g)(x)=x$ with dom $(g\circ g)=\mathbb{R}\setminus\{-1\}$.

Assignment 3.6 — Consider the graphs below.

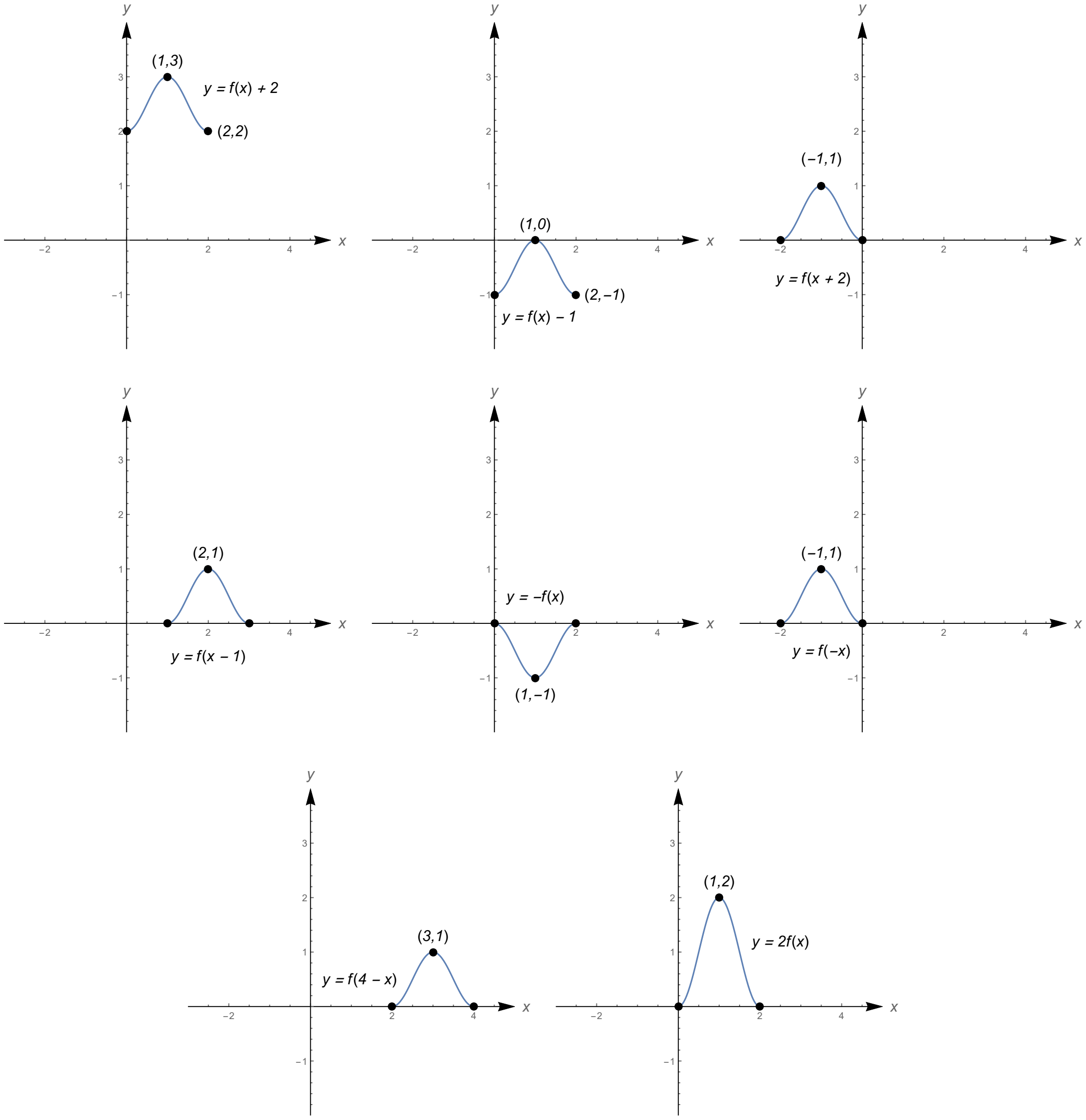


Figure 3.28: Graphs from the transformations from Exercise 3.6 (part 1).

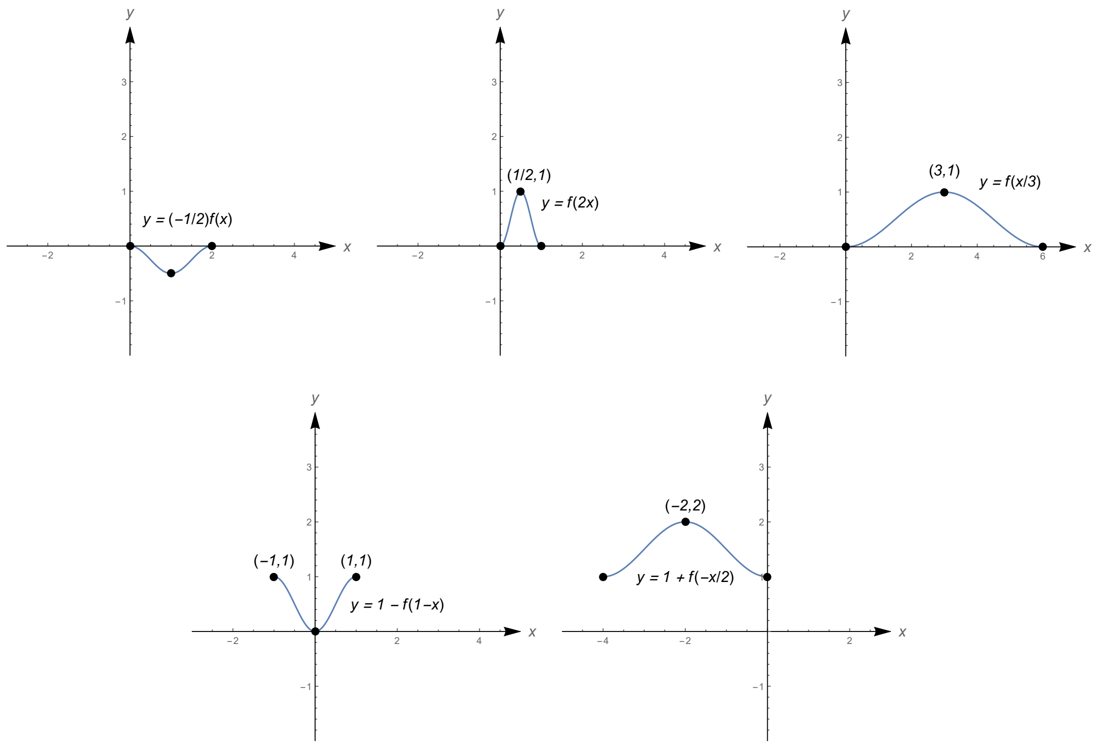


Figure 3.29: Graphs for the transformations in Exercise 3.6 (part 2).

Assignment 3.7 — Consider the graphs below.

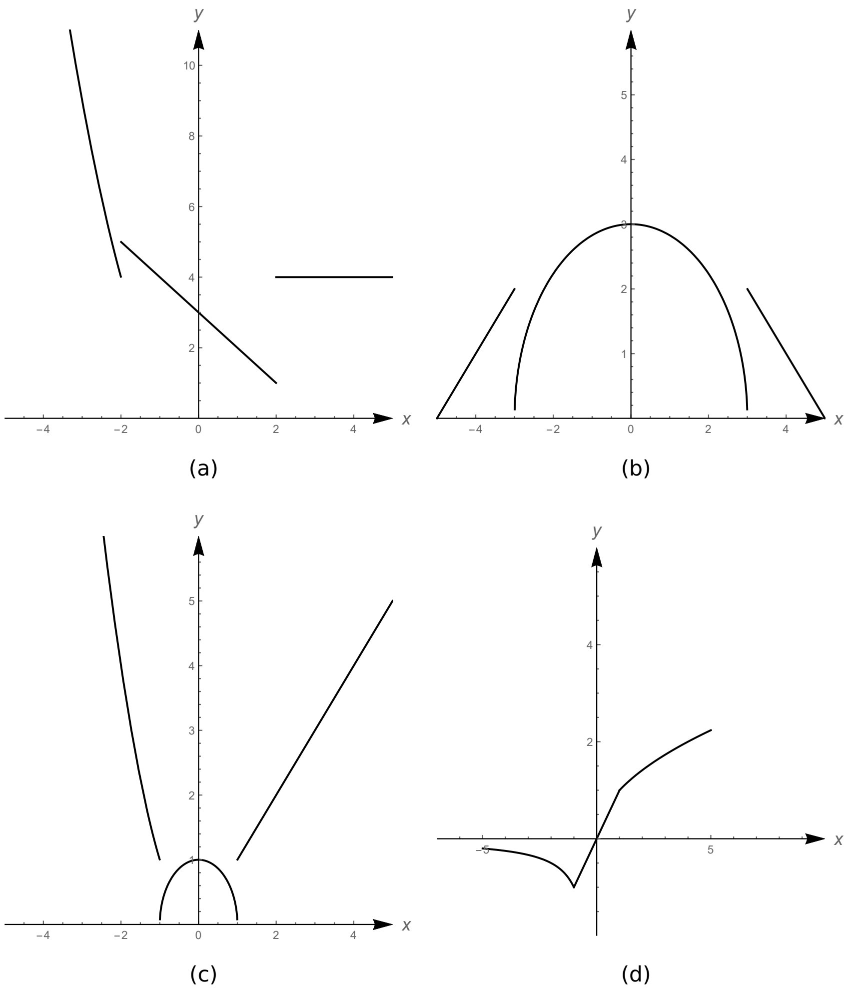


Figure 3.30: Graphs of the piecewise functions in Exercise 3.7.

Assignment 3.8 — $f(x) = \begin{cases} |x+1|-1, & \text{if} \quad x < 2, \\ 2, & \text{if} \quad x \geq 2. \end{cases}$ or: $f(x) = \begin{cases} -(x+2), & \text{if} \quad x < -1, \\ x, & \text{if} \quad -1 \leq x \leq 2, \\ 2, & \text{if} \quad x > 2. \end{cases}$

Assignment 3.9 —

(a) true

(b) false, assume $x = 1$

(c) false, assume $x = 1$

(d) true

(e) true

(f) false, because $\frac{1}{x} > \frac{1}{6}$

(g) false, because (h) is true

(h) true

Assignment 3.10 —

(a) $x = -\dfrac{13}{8} \vee x = \dfrac{53}{8}$

(b) $x = -\dfrac{3}{10}$

(c) $x = \dfrac{-4}{3} \vee x = 2$

(d) $x = -1 \vee x = 2$

(e) No solution

(f) No solution

(g) $x \in \left]\dfrac{5}{3}, 3\right[$

(h) $x \in [0, 4]$

(i) $x = -5$

(j) $x = -\dfrac{49}{18} \vee x = \dfrac{17}{6}$

(k) $x \geq 2$

(l) $x > 1$

(m) $x < -3 \vee x > 1$

(n) $x < 2$

(o) $2 < x < 9$

(p) $x < -10 \vee -4 < x < 0 \vee x > 6$

(q) $\dfrac{3}{4} < x < 1 \vee x > 1$

(r) $x \in]-1, +\infty[$

(s) No solution

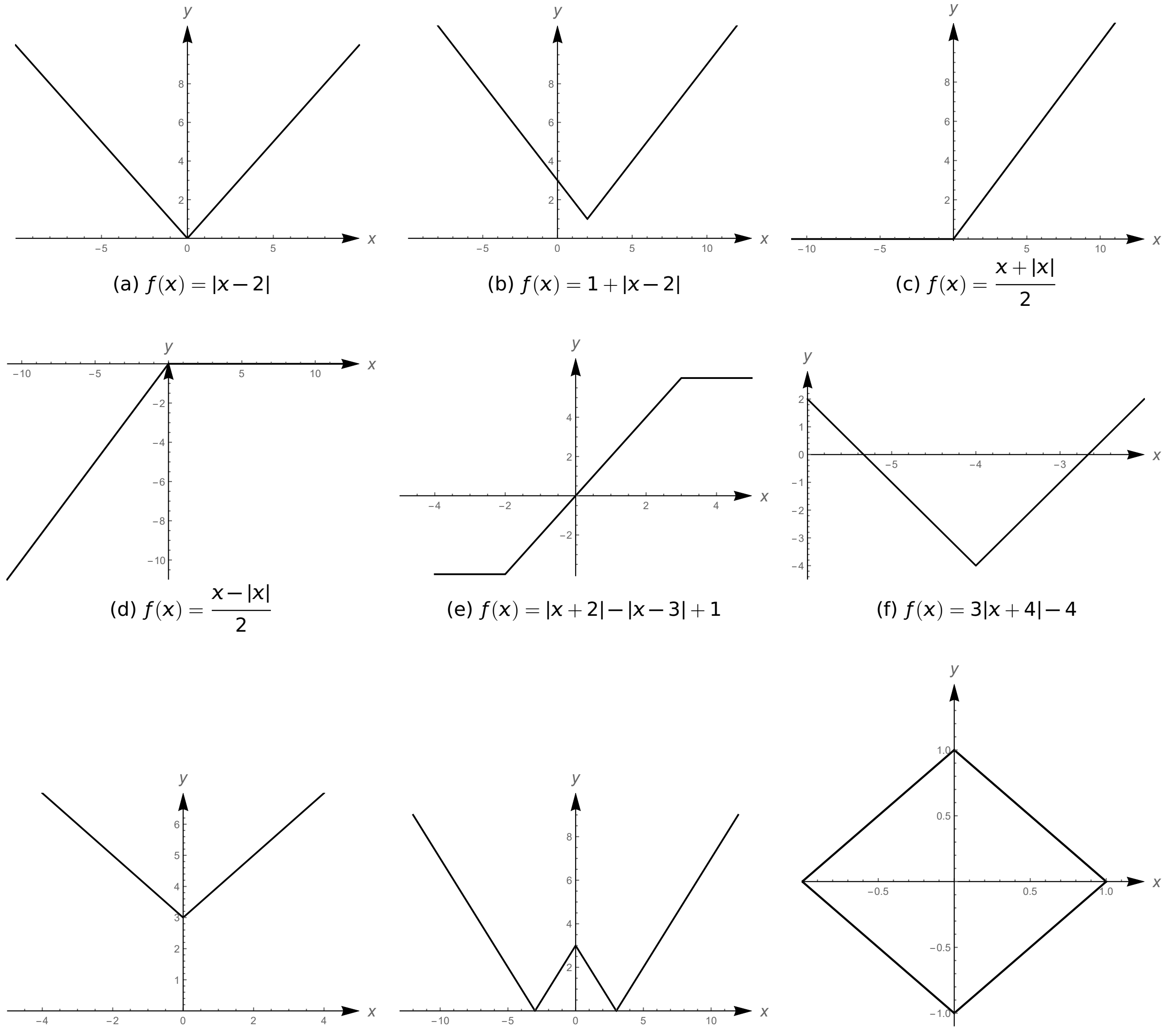
(t) $x = 1 \vee x = 2 - \sqrt{3} \vee x = -2 - \sqrt{7}$

(u) $-1 \leq x \leq \dfrac{3}{5} \vee x \geq 3$

(v) $x \in \mathbb{R}$

(w) $x \geq 1$

Assignment 3.11 — Consider the graphs below.



(a) $f(x) = |x-2|$ (b) $f(x) = 1+|x-2|$ (c) $f(x) = \dfrac{x+|x|}{2}$

(d) $f(x) = \dfrac{x-|x|}{2}$ (e) $f(x) = |x+2|-|x-3|+1$ (f) $f(x) = 3|x+4|-4$

(g) $f(x) = ||x|+3|$ (h) $f(x) = ||x|-3|$ (i) $|x|+|y| = 1$

Figure 3.32: Graphs of the absolute value functions in Exercise 3.11.

Assignment 3.12 —

(a) $f^{-1}(x) = \dfrac{x-8}{2}$ Function

(b) $f^{-1}(x) = -\dfrac{5}{3}x + \dfrac{1}{3}$ Function

(c) $f^{-1}(x) = \dfrac{6-5x}{x-1}$ Function

(d) $f^{-1}(x) = \sqrt[3]{x-1}$ Function

(e) $f^{-1}(x) = -\dfrac{1}{2} \pm \sqrt{x+\dfrac{1}{4}} \quad \wedge \quad x \geq -\dfrac{1}{4}$ No Function

(f) $f^{-1}(x) = \pm\sqrt{1-x^2} \quad \wedge \quad x \in [0,1]$ No Function

(g) $f^{-1}(x) = \dfrac{1}{9}(x+4)^2 + 1 \quad \wedge \quad x \geq -4$ Function

(h) $f^{-1}(x) = 3 \pm \sqrt{x+4} \quad \wedge \quad x \geq -4$ No Function

Assignment 3.13 —

	dom f	codom f	im f	inj	surj	bij	periodic	even/odd	mon.	lok. max/min
(a)	$[-2,2]$	$\mathbb{R}^+$	$[0,2]$	no	no	no	no	even	no	$(0,2),(\pm 2,0)$
(b)	$\mathbb{R}$	$\mathbb{R}^+$	$[2,+\infty[$	no	no	no	no	even	no	$(0,2)$
(c)	$\mathbb{R}\setminus\{-1,1\}$	$\mathbb{R}$	$\mathbb{R}\setminus[0,1[$	no	no	no	no	even	no	$(0,1)$
(d)	$\mathbb{R}$	$\mathbb{R}^+$	$]0,1]$	no	no	no	no	even	no	$(0,1)$
(e)	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	yes	yes	yes	no	no	yes	$(0,1)$
(f)	$\mathbb{R}$	$\mathbb{R}^+$	$[1,+\infty[$	no	no	no	no	even	no	$(0,1)$
(g)	$]-\infty,1]$	$\mathbb{R}$	$\mathbb{R}^+$	yes	no	no	no	no	yes	$(1,0)$
(h)	$\mathbb{R}$	$\mathbb{R}^+$	$\mathbb{R}^+$	no	yes	no	no	even	no	$(0,0)$
(i)	$\mathbb{R}\setminus\{-1\}$	$\mathbb{R}\setminus\{1\}$	$\mathbb{R}\setminus\{1\}$	yes	yes	yes	no	no	yes	none
(j)	$\mathbb{R}_0$	$\mathbb{R}_0$	$\mathbb{R}_0$	yes	yes	yes	no	odd	no	none
(k)	$[-2,+\infty[$	$\mathbb{R}$	$]-\infty,3]$	yes	no	no	no	no	yes	$(-2,3)$
(l)	$\mathbb{R}$	$\mathbb{R}^+$	$\mathbb{R}^+$	no	yes	no	no	no	no	$(-2,0)$

Chapter 4

Assignment 4.1 —

(a) All points located in the half plane under the line $y = x - 1$, including the points on the line.

(b) All points located in the area between the lines $y = -x - 4$, $y = x - 4$ and $y = 2 - x$, which includes the origin.

(c) All points located in the region understood between the line $y = x + 2$ and the parabola $y = x^2$, including points on the parabola.

(d) All points located within the circle with center $(-1,0)$ and radius 3.

(e) All points located within the circle with center $(1,0)$ and radius 1 ánd within the circle with center $(0,1)$ and radius 1.

(f) All points located outside the circle with center $(2,-1)$ and radius 3 ánd to the right of the line $x + y = 1$.

Assignment 4.2 —

(a) $2(x-2)\left(x^2+2x+4\right)(x+2)\left(x^2-2x+4\right)$

(b) $(2x+1)^3$

(c) $\left(x^2+1\right)(2x+3)$

(d) $(x-2)^3(x-1)$

(e) $(x-2)(x+2)\left(x^2+x+2\right)$

(f) $(2x-1)(x+1)(x-2)$

(g) $(2a-5b)^3$

(h) $(a-1-b-2c)(a-1+b+2c)$

(i) $(x-y)(x^2-3xy+4y^2)$

(j) $9(x+1)^2(x^2+4x+1)$

Assignment 4.3 —

(a) $16x^4-8x^2+1=(2x+1)^2(2x-1)^2=(4x^2-1)^2$

(b) $x^4-1=(x^2+1)(x+1)(x-1)$

(c) $x^5-x^4-16x+16=(x-2)(x+2)(x^2+4)(x-1)$

(d) $x^5+x^3+8x^2+8=(x^2+1)(x+2)(x^2-2x+4)$

(e) $x^4+6x^3+9x^2=x^2(x+3)^2$

(f) $x^6-3x^4+3x^2-1=(x-1)^3(x+1)^3$

(g) $x^9-4x^7-x^6+4x^4=x^4(x-2)(x-1)(x+2)(x^2+x+1)$

Assignment 4.4 —

(a) $x^3-3x^2+20=(x+2)(x^2-5x+10)=(x+2)\left(x-\dfrac{5+\sqrt{15}i}{2}\right)\left(x-\dfrac{5-\sqrt{15}i}{2}\right)$
real zeros: $x=-2$, complex zeros: $x=\dfrac{5\pm\sqrt{15}i}{2}$

(b) $2x^3-4x^2-10x+12=2(x-3)(x-1)(x+2)$
real zeros: $x=3$, $x=1$ and $x=-2$

(c) $x^6-16x^3+64=(x^3-8)^2=(x-2)^2(x^2+2x+4)^2=(x-2)^2\left(x+1-\sqrt{3}i\right)^2\left(x+1+\sqrt{3}i\right)^2$
real zeros: $x=2$ (2x), complex zeros: $x=-1\pm\sqrt{3}i$ (2x)

(d) $8x^4-20x^3+18x^2-7x+1=(x-1)(2x-1)^3$
real zeros: $x=\dfrac{1}{2}$ (3x) and $x=1$

(e) $x^3-16x^2+48x+72=(x-6)\left(x-5-\sqrt{37}\right)\left(x-5+\sqrt{37}\right)$
real zeros: $x=6$, $x=5+\sqrt{37}$ and $x=5-\sqrt{37}$

(f) $4x^3-14x^2+8x+8=4(x-2)^2\left(x+\dfrac{1}{2}\right)$
real zeros: $x=2$ (2x) and $x=-\dfrac{1}{2}$

(g) $x^5+6x^4+x^3-26x^2-32=(x^3+6x^2-32)(x^2+1)=(x-2)(x+4)^2(x-i)(x+i)$
real zeros: $x=2$ and $x=-4$ (2x), complex zeros: $x=\pm i$

(h) $-2x^6-10x^5-16x^4-8x^3=-2x^3(x+1)(x+2)^2$
real zeros: $x=0$ (3x), $x=-1$ and $x=-2$ (2x)

Assignment 4.5 —

(a) $\left]-\infty, \frac{1}{2}\right[\cup]4, 5[$

(b) $\{-2\} \cup [1, 3]$

(c) $]-\infty, -1[\cup]-1, 0[\cup]2, +\infty[$

(d) $]-\infty, -2[\cup]-\sqrt{2}, \sqrt{2}[$

(e) $]-\infty, -\sqrt{3}[\cup]\sqrt{3}, +\infty[$

Assignment 4.6 —

(a) $-4x^4 + x^3 + x^2 + x + 1$

(b) $x + \frac{2x-1}{x^2-2}$

(c) $1 - \frac{5x+3}{x^2+5x+1}$

(d) $x - 2 + \frac{x+6}{x^2+2x+3}$

(e) $2x + 9 + \frac{44x-68}{x^2-6x+7}$

Assignment 4.7 —

(a) $\varnothing$

(b) $]-1, 0[\cup]1, +\infty[$

(c) $[0, +\infty[$

(d) $\left]-3, -\frac{1}{3}\right[\cup]2, 3[$

(e) $[-3, 0[\cup]0, 4[\cup [5, +\infty[$

(f) $\left]-1, -\frac{1}{2}\right] \cup]1, +\infty[$

Assignment 4.8 —

(a) dom f: $\mathbb{R}$, VA: none, HA: $y = 0$

(b) dom f: $\mathbb{R}\backslash\{-1, 0, 1\}$, VA: $x = -1$, $x = 0$, and $x = 1$, HA: $y = 0$

(c) dom f: $\mathbb{R}\backslash\{-1, 0\}$, VA: $x = -1$ and $x = 0$, HA: $y = 0$

(d) dom f: $\mathbb{R}\backslash\left\{\frac{-1 \pm \sqrt{5}}{2}\right\}$, VA: $x = \frac{-1 \pm \sqrt{5}}{2}$, HA: none

(e) dom f: $\mathbb{R}\backslash\{-3, 2\}$, VA: $x = 2$, HA: $y = 1$

(f) dom f: $\mathbb{R}\backslash\{-1\}$, VA: $x = -1$, HA: none

Assignment 4.9 —

(a) No solution

(b) $x = 10 + 4\sqrt{5}$

(c) $x = 2$

(d) $x > 0$

(e) $x = 8$

(f) $\frac{1}{10}\left(\sqrt{5} - 5\right) \leq x \leq 0$

Assignment 4.10 — $v = \frac{\sqrt{3}c}{2}$

Assignment 4.11 —

(a)
- $\operatorname{dom} f = [-4,+\infty[$,
- zero(s): $(17/2,0)$,
- intersection(s) x-axis: $(17/2,0)$, intersection(s) y-axis: $(0,5-2\sqrt{2})$,
- asymptotes: none.

(b)
- $\operatorname{dom} f =]-3,0] \cup]3,+\infty[$,
- zero(s): $(0,0)$,
- intersection(s) x-axis: $(0,0)$, intersection(s) y-axis: $(0,0)$,
- asymptotes: $x = \pm 3$, $y = 0$.

(c)
- $\operatorname{dom} f = \mathbb{R}\setminus\{-2\}$,
- zero(s): $(0,0)$,
- intersection(s) x-axis: $(0,0)$, intersection(s) y-axis: $(0,0)$,
- asymptotes: $x = -2$, $y = 5$.

(d)
- $\operatorname{dom} f = \mathbb{R}$,
- zero(s): $(0,0)$, $(7,0)$
- intersection(s) x-axis: $(0,0)$, $(7,0)$, intersection(s) y-axis: none,
- asymptotes: none.

(e)
- $\operatorname{dom} f =]-\infty,-2[\cup]2,+\infty[$,
- zero(s): none,
- intersection(s) x-axis: none, intersection(s) y-axis: none,
- asymptotes: $x = \pm 2$, $y = 0$.

(f)
- $\operatorname{dom} f = [3,17[\cup]17,+\infty[$,
- zero(s): $(3,0)$,
- intersection(s) x-axis: $(3,0)$, intersection(s) y-axis: none,
- asymptotes: $x = 17$.

(g)
- $\operatorname{dom} f =]-\infty,-1/2] \cup]2,+\infty[$,
- zero(s): $(-1/2,0)$,
- intersection(s) x-axis: $(-1/2,0)$, intersection(s) y-axis: none,
- asymptotes: $x = 2$

Assignment 4.12 —

(a) Graph (c)

(b) Graph (a)

(c) Graph (b)

(d) Graph (d)

Assignment 4.13 —

(a) $(x-2)^2 = y+1$

- top V: $(2,-1)$,

- focal point F: $(2,-3/4)$,
- axis of symmetry: $x=2$,
- directrix d: $y=-5/4$,
- intersection(s) x-axis: $(1,0)$ en $(3,0)$, intersection(s) y-axis: $(0,3)$.

(b) $(x-1)^2=y+1$

- top V: $(1,-1)$,
- focal point F: $(1,-3/4)$,
- axis of symmetry: $x=1$,
- directrix d: $y=-5/4$,
- intersection(s) x-axis: $(0,0)$ en $(2,0)$, intersection(s) y-axis: $(0,0)$.

(c) $(y+1)^2=-2(x-1/2)$

(d) top V: $(1/2,-1)$,

- focal point F: $(0,-1)$,
- axis of symmetry: $y=-1$,
- directrix d: $x=1$,
- intersection(s) x-axis: $(0,0)$, intersection(s) y-axis: $(0,0)$ en $(0,-2)$.

(e) $(x+1/2)^2=-(y-1/4)$

- top V: $(-1/2,1/4)$,
- focal point F: $(-1/2,0)$,
- axis of symmetry: $x=-1/2$,
- directrix d: $y=1/2$,
- intersection(s) x-axis: $(0,0)$ en $(-1,0)$, intersection(s) y-axis: $(0,0)$.

(f) $\left(x-\frac{5}{6}\right)^2=-\frac{1}{3}\left(y-\frac{73}{12}\right)$

- top V: $\left(\frac{5}{6},\frac{73}{12}\right)$,
- focal point F: $\left(\frac{5}{6},6\right)$,
- axis of symmetry: $x=5/6$,
- directrix d: $y=37/6$,
- intersection(s) x-axis: $\left(\frac{5\pm\sqrt{73}}{6},0\right)$, intersection(s) y-axis: $(0,4)$.

Assignment 4.14 —

(a) $y=x^2-3$

(b) $y=(x-4)^2$

(c) $y=(x-3)^2+3$

(d) $y=(x-4)^2-2$

Assignment 4.15 —

(a) $x^2 + \dfrac{y^2}{2/3} = 1$: ellipse

(b) $y^2 = -\dfrac{3}{2}x$: parabola

(c) $-\dfrac{x^2}{3} + \dfrac{y^2}{3/2} = 1$: hyperbole

(d) $x^2 = \dfrac{4}{3}y$: parabola

(e) $\dfrac{x^2}{3/2} + \dfrac{y^2}{3/4} = 1$: ellipse

(f) $x^2 = -\dfrac{3}{2}y$: parabola

(g) $y^2 = 3x$: parabola

(h) $\dfrac{x^2}{3} - \dfrac{y^2}{3} = 1$: hyperbole

(i) $\dfrac{x^2}{3/2} - y^2 = 1$: hyperbole

Assignment 4.16 —

(a) $\dfrac{x^2}{5} + \dfrac{y^2}{9} = 1$

(b) $\dfrac{(x-2)^2}{16} + \dfrac{(y-1)^2}{12} = 1$

(c) $\dfrac{x^2}{12} + \dfrac{y^2}{4} = 1$

(d) $(x-2)^2 = -4(y-4)$

(e) $x^2 = -4y$

(f) $y^2 = -8x$

(g) $-\dfrac{x^2}{17/8} + \dfrac{y^2}{17} = 1$

(h) $-\dfrac{x^2}{3} + y^2 = 1$

(i) $\dfrac{x^2}{25/2} - \dfrac{(y-1)^2}{25/2} = 1$

Assignment 4.17 —

(a) $(x+1)^2 + \left(y - \dfrac{3}{2}\right)^2 = \dfrac{13}{4}$: circle

(b) $3\left(x + \dfrac{1}{3}\right)^2 + 3\left(y + \dfrac{7}{6}\right)^2 = \dfrac{89}{12}$: circle

(c) $-\dfrac{(x+1/2)^2}{5/12} + \dfrac{(y-1)^2}{5/8} = 1$: hyperbola

(d) $\dfrac{(x-1)^2}{2} + \dfrac{(y-2)^2}{4/3} = 1$: ellipse

(e) $(y-1)^2 = -\dfrac{3}{2}x$: parabola

(f) $\left(x - \dfrac{1}{4}\right)^2 = \dfrac{1}{2}\left(y - \dfrac{55}{8}\right)$: parabola

(g) $\dfrac{x^2}{2/3} + \dfrac{(y-1)^2}{2/7} = 1$: ellipse

(h) $-\dfrac{(x-1)^2}{1/2} + y^2 = 1$: hyperbola

(i) $-x^2 + \dfrac{(y+2)^2}{4} = 1$: hyperbola

(j) $\dfrac{(x-1)^2}{4} + \dfrac{(y+1)^2}{9} = 1$: ellipse

Assignment 4.18 — $\dfrac{x^2}{225} - \dfrac{7y^2}{3600} = 1$, with $x \geq 15$

Assignment 4.19 — $\dfrac{x^2}{14400} - \dfrac{13y^2}{921600} = 1$, <ith $x \geq 120$

Assignment 4.20 —

(a) $x \in \left[\frac{1}{2}, 1\right] \cup \left[2, \frac{5}{2}\right]$

(b) $x \in \left]-\infty, -2\right[\cup \left]\frac{4}{3}, +\infty\right[$

(c) $x \in \mathbb{R}$

(d) $x \in \left]-\infty, 1\right[\cup \left]\frac{7}{3}, +\infty\right[$

(e) $x \in \left]-\infty, \frac{5-\sqrt{73}}{6}\right] \cup \left[\frac{5+\sqrt{73}}{6}, +\infty\right[$

(f) $x \in \left]-3\sqrt{2}, -\sqrt{11}\right] \cup \left[-\sqrt{7}, 0\right[\cup \left]0, \sqrt{7}\right] \cup \left[\sqrt{11}, 3\sqrt{2}\right[$

(g) $x \in \left[-2-\sqrt{7}, \sqrt{7}-2\right] \cup [1, 3]$

(h) $x \in [-6, -3] \cup [-2, +\infty[$

(i) $x \in \left]-\infty, 1\right[\cup \left]2, \frac{3+\sqrt{17}}{2}\right[$

(j) $x \in \left]-1-\sqrt{2}, -1\right[\cup \left]0, 1\right[\cup \left]1, +\infty\right[$

Chapter 5

Assignment 5.1 —

(a) $\frac{-3}{2}$

(b) $\frac{3}{2}$

(c) 27

(d) $-2x$

(e) $1 + \log_x(2)$

(f) 2

(g) -2

(h) $\ln\left(x^2(x-2)^5\right)$

Assignment 5.2 — Self demonstration.

Assignment 5.3 —

(a) $x = -\frac{1}{3}$

(b) $x = 2$

(c) $x = \frac{1}{2}$

(d) $x = 81$

(e) $x = -\log_4\left(\frac{512}{5}\right)$

(f) $x = 1$ of $x = \frac{1}{e}$

(g) $x = \frac{15}{22}$

(h) $x = 3$ of $x = 3 + \frac{\ln(2)}{\ln(5)}$

(i) $\varnothing$

(j) $x = -2$

(k) $x = 1 \vee x = 2$

(l) $x = 0 \vee x = 1$

(m) $x = \frac{1}{81} \vee x = 729$

(n) $x = 4$

(o) $x = 16$

Assignment 5.4 —

(a) $\sqrt{2} < x \leq 2$

(b) $x > -1$

(c) $0 < x < 2 \ \vee \ 8 < x < 10$

(d) $1 < x < \frac{9}{5}$

(e) $x > e$

(f) $-\sqrt{3} < x < \sqrt{3}$

Assignment 5.5 —

(a)
- $\operatorname{dom} f =]-2,-1[\cup]1,+\infty[$,
- intersections(s) x-axis: $x = \frac{1 \pm \sqrt{13}}{2}$,
- intersections(s) y-axis: none.

(b)
- $\operatorname{dom} f =]5,+\infty[$,
- intersections(s) x-axis: $x \approx 5$,
- intersections(s) y-axis: none.

(c)
- $\operatorname{dom} f =]13,+\infty[$,
- intersections(s) x-axis: $x = 20$,
- intersections(s) y-axis: none.

(d)
- $\operatorname{dom} f = \emptyset$,
- intersections(s) x-axis: none,
- intersections(s) y-axis: none.

(e)
- $\operatorname{dom} f =]-1,1[$,
- intersections(s) x-axis: $x = 0$,
- intersections(s) y-axis: $y = 0$.

(f)
- $\operatorname{dom} f =]0,1[$,
- intersections(s) x-axis: $x = 0.2$,
- intersections(s)) y-axis: none.

Assignment 5.6 — $h = -\frac{RT}{Mg} \ln\left(\frac{p}{p_0}\right) = \frac{RT}{Mg} \ln\left(\frac{p_0}{p}\right)$

Assignment 5.7 — The culture will contain 1012 cells after another 12 hours.

Assignment 5.8 — It takes 182 years to reduce radioactivity by 10%.

Assignment 5.9 —

(a) After one week there will be approximately 390 million bacteria.

(b) After 125 hours there will be 5 million bacteria.

Assignment 5.10 — After 5 minutes the thermometer will indicate $22.35°C$.

Assignment 5.11 — It takes 92.88 minutes for the object to cool to $0°C$.

Assignment 5.12 —

(a) $-\frac{\sqrt{2}}{2}$

(b) 1

(c) $\frac{\sqrt{3}}{2}$

(d) $\frac{\sqrt{2}}{2}\left(\frac{\sqrt{3}}{2}+\frac{1}{2}\right)$

(e) $\frac{\sqrt{2}}{2}\left(\frac{\sqrt{3}}{2}-\frac{1}{2}\right)$

(f) $\frac{\sqrt{2}}{2}\left(\frac{\sqrt{3}}{2}-\frac{1}{2}\right)$

(g) $2+\sqrt{3}$

(h) $-\sqrt{2}\left(1-\sqrt{3}\right)$

(i) $\frac{-2\sqrt{3}}{3}$

(j) $\sqrt{3}$

Assignment 5.13 —

(a) $\cos(\theta)=-\frac{4}{5}, \quad \csc(\theta)=\frac{5}{3}, \quad \sec(\theta)=-\frac{5}{4}, \quad \tan(\theta)=-\frac{3}{4}, \quad \cot(\theta)=-\frac{4}{3}$

(b) $\cos(\theta)=\frac{\sqrt{5}}{5}, \quad \sin(\theta)=\frac{2\sqrt{5}}{5}, \quad \csc(\theta)=\frac{\sqrt{5}}{2}, \quad \sec(\theta)=\sqrt{5}, \quad \cot(\theta)=\frac{1}{2}$

(c) $\cos(\theta)=\frac{1}{3}, \quad \sin(\theta)=-\frac{2\sqrt{2}}{3}, \quad \csc(\theta)=-\frac{3\sqrt{2}}{4}, \quad \tan(\theta)=-2\sqrt{2}, \quad \cot(\theta)=-\frac{\sqrt{2}}{4}$

(d) $\sin(\theta)=\frac{12}{13}, \quad \csc(\theta)=\frac{13}{12}, \quad \sec(\theta)=-\frac{13}{5}, \quad \tan(\theta)=-\frac{12}{5}, \quad \cot(\theta)=-\frac{5}{12}$

(e) $\cos(\theta)=-\frac{\sqrt{3}}{2}, \quad \sin(\theta)=-\frac{1}{2}, \quad \sec(\theta)=-\frac{2\sqrt{3}}{3}, \quad \tan(\theta)=\frac{\sqrt{3}}{3}, \quad \cot(\theta)=\sqrt{3}$

(f) $\cos(\theta)=-\frac{2\sqrt{5}}{5}, \quad \sin(\theta)=-\frac{\sqrt{5}}{5}, \quad \csc(\theta)=-\sqrt{5}, \quad \sec(\theta)=-\frac{\sqrt{5}}{2}, \quad \cot(\theta)=2$

Assignment 5.14 — Prove yourself.

Assignment 5.15 —

(a) $f(x)=3\sin\left(\frac{x}{2\pi}\right)$

- Period: $4\pi^2$
- Amplitude: 3
- Phase shift: none
- Vertical shift: none
- $\operatorname{dom} f=\mathbb{R}$
- $\operatorname{im} f=[-3,3]$
- zeros: $x=2\pi^2 k,\ k\in\mathbb{Z}$

(b) $f(x) = \frac{2}{3}\sin\left(\frac{2}{3}\left(x - \frac{\pi}{4}\right)\right) - 11$

- Period: 3π
- Amplitude: $\frac{2}{3}$
- Phase shift: $\frac{\pi}{4}$
- Vertical shift: -11
- $\operatorname{dom} f = \mathbb{R}$
- $\operatorname{im} f = \left[-\frac{35}{3}, -\frac{31}{3}\right]$
- zeros: none

(c) $f(x) = \sin\left(10\pi\left(x + \frac{1}{2}\right)\right) + 3$

- Period: $\frac{1}{5}$
- Amplitude: 1
- Phase shift: $-\frac{1}{2}$
- Vertical shift: 3
- $\operatorname{dom} f = \mathbb{R}$
- $\operatorname{im} f = [2, 4]$
- zeros: none

(d) $f(x) = 2\sin(3x - 2) + 1$

- Period: $\frac{2\pi}{3}$
- Amplitude: 2
- Phase shift: $\frac{2}{3}$
- Vertical shift: 1
- $\operatorname{dom} f = \mathbb{R}$
- $\operatorname{im} f = [-1, 3]$
- zeros: $x = \frac{2}{3} - \frac{\pi}{18}(12k + 5),\ k \in \mathbb{Z}, \quad x = \frac{2}{3} - \frac{\pi}{18}(12k + 1),\ k \in \mathbb{Z}$

Assignment 5.16 —

(a) $x = \frac{\pi}{8} + \frac{k\pi}{2} \quad \vee \quad x = \frac{\pi}{20} + \frac{k\pi}{5} \qquad (k \in \mathbb{Z})$

(b) $x = -\frac{\pi}{2} + k\pi \quad \vee \quad x = \frac{3\pi}{2} + k2\pi \qquad (k \in \mathbb{Z})$

(c) $x = \frac{\pi}{2} + k\pi \quad \vee \quad x = \arctan\left(\frac{1}{3}\right) + k\pi \qquad (k \in \mathbb{Z})$

(d) $x = \arcsin\left(\frac{1}{3}\right) + k2\pi \quad \vee \quad x = \pi - \arcsin\left(\frac{1}{3}\right) + k2\pi \qquad (k \in \mathbb{Z})$

(e) $x = -\dfrac{\pi}{3} + k2\pi \quad \vee \quad x = \dfrac{\pi}{3} + k2\pi \qquad (k \in \mathbb{Z})$

(f) $x = k\pi \quad \vee \quad x = \arctan\left(\pm\sqrt{\dfrac{3}{5}}\right) + k\pi \qquad (k \in \mathbb{Z})$

(g) $x = \pm\dfrac{\pi}{4} + k\pi \quad \vee \quad x = \pm\dfrac{\pi}{3} + k2\pi \qquad (k \in \mathbb{Z})$

(h) $x = -\dfrac{5\pi}{6} + k\pi \quad \vee \quad x = -\dfrac{\pi}{6} + k\pi \qquad (k \in \mathbb{Z})$

(i) $x = \pm\dfrac{\pi}{4} + k\pi \quad \vee \quad x = \pm\dfrac{4\pi}{3} + k\pi \qquad (k \in \mathbb{Z})$

(j) $x = \dfrac{3\pi}{4} + k\pi \quad \vee \quad x = \dfrac{\pi}{4} + k\pi \qquad (k \in \mathbb{Z})$

Assignment 5.17 —

(a) $-\dfrac{\pi}{4} + k\dfrac{\pi}{2} < x < 0,1608\ldots + k\dfrac{\pi}{2} \qquad (k \in \mathbb{Z})$

(b) $\dfrac{3\pi}{4} + k\pi < x < \dfrac{13\pi}{12} + k\pi \qquad (k \in \mathbb{Z})$

(c) $-\dfrac{14\pi}{45} + k\dfrac{2\pi}{3} < x < -\dfrac{4\pi}{45} + k\dfrac{2\pi}{3} \qquad (k \in \mathbb{Z})$

(d) $\dfrac{\pi}{6} + k\dfrac{\pi}{2} < x < \dfrac{7\pi}{12} + k\dfrac{\pi}{2} \qquad (k \in \mathbb{Z})$

(e) $\dfrac{5\pi}{12} + k\pi < x < \dfrac{\pi}{2} + k\pi \quad \vee \quad \dfrac{\pi}{2} + k\pi < x < \dfrac{7\pi}{12} + k\pi \qquad (k \in \mathbb{Z})$

(f) $-\dfrac{\pi}{6} + 2k\pi < x < \arcsin\left(\dfrac{-1+\sqrt{3}}{2}\right) + 2k\pi$

$\vee \quad \pi - \arcsin\left(\dfrac{-1+\sqrt{3}}{2}\right) + 2k\pi < x < \dfrac{7\pi}{6} + 2k\pi$

$\vee \quad \dfrac{\pi}{6} + 2k\pi < x < \dfrac{5\pi}{6} + 2k\pi \qquad (k \in \mathbb{Z})$

Assignment 5.18 —

(a) $\mathbb{R}$

(b) $]-\infty, -1] \cup [1, +\infty[$

(c) $\left[-2, -\sqrt{2}\right] \cup \left[\sqrt{2}, 2\right]$

(d) $\left[-\dfrac{1}{2}, \dfrac{1}{2}\right]$

(e) $\left[\dfrac{\sqrt{2}}{4}, \dfrac{1}{2}\right]$

Assignment 5.19 —

(a) $-\frac{\pi}{4}$

(b) $\frac{\pi}{6}$

(c) $\frac{\pi}{6}$

(d) $\frac{1}{2}$

(e) -1

(f) $\frac{\sqrt{5}}{3}$

(g) $\frac{3}{4}$

(h) $\frac{56\sqrt{2}}{17}$

(i) 1

(j) $\frac{\sqrt{2}}{2}$

(k) -1

(l) $\frac{17}{25}$

(m) $\frac{3\pi}{4}$

Assignment 5.20 —

(a) $a = 0,\ b = 0,\ c = 9 \quad \Rightarrow b^2 - 4ac = 0 \quad \Rightarrow$ parabole

$9\left(y + \frac{2}{3}\right)^2 = 0$ (falls apart into 2 parallel lines after square root).

(b) $a = 3,\ b = 0,\ c = -1 \quad \Rightarrow b^2 - 4ac = 12 > 0 \quad \Rightarrow$ hyperbole.

$\frac{(x-2)^2}{5/3} - \frac{(y+1)^2}{5} = 1$

Chapter 6

Assignment 6.1 —

(a) $\vec{a} = (3, -2) \quad \Rightarrow \quad \|\vec{a}\| = \sqrt{13}$ and $\theta = \arctan\left(-\frac{2}{3}\right)$

(b) $\vec{b} = \left(\pi, -\frac{4}{3}\right) \quad \Rightarrow \quad \left\|\vec{b}\right\| = \frac{4}{3}$ and $\theta = \pi$

(c) $\vec{c} = (0, 2) \quad \Rightarrow \quad \|\vec{c}\| = 2$ and $\theta = \frac{\pi}{2}$

(d) $\vec{d} = (1, -1) \quad \Rightarrow \quad \left\|\vec{d}\right\| = \sqrt{2}$ and $\theta = -\frac{\pi}{4}$

Assignment 6.2 —

(a) $\vec{a} + \vec{x} + \vec{b} = \vec{o} \quad \Leftrightarrow \quad \vec{x} = -\vec{a} - \vec{b} \quad \Rightarrow \quad \vec{x} = (-4, -6)$

(b) $\vec{a} - \vec{b} = 2\vec{b} + \vec{x} - \vec{a} \quad \Leftrightarrow \quad \vec{x} = 2\vec{a} - 3\vec{b} \quad \Rightarrow \quad \vec{x} = (3, -8)$

(c) $3(\vec{x} - \vec{a}) = \vec{x} - \vec{b} \quad \Leftrightarrow \quad \vec{x} = \frac{3\vec{a} - \vec{b}}{2} \quad \Rightarrow \quad \vec{x} = (4, 1)$

(d) $2(\vec{x} - \vec{a}) = 3(\vec{x} - \vec{b}) \quad \Leftrightarrow \quad \vec{x} = -2\vec{a} + 3\vec{b} \quad \Rightarrow \quad \vec{x} = (-3, 8)$

Assignment 6.3 —

(a) $\overrightarrow{AB} = 3\hat{\boldsymbol{\imath}} - 2\hat{\boldsymbol{\jmath}}$

(b) $\overrightarrow{BA} = -3\hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}}$

(c) $\overrightarrow{AC} = 2\hat{\boldsymbol{\imath}} - 5\hat{\boldsymbol{\jmath}}$

(d) $\overrightarrow{AB} - \overrightarrow{BC} = 4\hat{\boldsymbol{\imath}} + \hat{\boldsymbol{\jmath}}$

(e) $\overrightarrow{AC} - 2\overrightarrow{AB} + 3\overrightarrow{CD} = -7\hat{\boldsymbol{\imath}} + 20\hat{\boldsymbol{\jmath}}$

(f) $\dfrac{\overrightarrow{AB} + \overrightarrow{AC} + \overrightarrow{AD}}{3} = 2\hat{\boldsymbol{\imath}} - \dfrac{5}{3}\hat{\boldsymbol{\jmath}}$

Assignment 6.4 — $x = -4$

Assignment 6.5 — $\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}} = 4$ and $\cos\left(\vec{\boldsymbol{a}}, \vec{\boldsymbol{b}}\right) = \dfrac{4}{5}$

Assignment 6.6 — $\left(\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}}\right) \vec{\boldsymbol{c}} = (a_1 b_1 c_1 + a_2 b_2 c_1, a_1 b_1 c_2 + a_2 b_2 c_2)$
$\vec{\boldsymbol{a}} \left(\vec{\boldsymbol{b}} \cdot \vec{\boldsymbol{c}}\right) = (a_1 b_1 c_1 + a_1 b_2 c_2, a_2 b_1 c_1 + a_2 b_2 c_2)$

Assignment 6.7 — Vertices of a square:

- $\overrightarrow{AB} \perp \overrightarrow{BC} \quad \Leftrightarrow \quad \overrightarrow{AB} \cdot \overrightarrow{BC} = 0$
- $d(A,B) = d(B,C) = \sqrt{17}$

The fourth vertex is $(-2,-2)$.

Assignment 6.8 —

(a) $\vec{\boldsymbol{v}}(-1,2,3) \perp \vec{\boldsymbol{w}}(1,2,h) \ \Leftrightarrow \ h = -1$

(b) $\vec{\boldsymbol{a}}(\sqrt{3}, h, 8) \perp \vec{\boldsymbol{b}}(h,-4,2) \ \Leftrightarrow \ h = \dfrac{-16}{\sqrt{3}-4}$

Assignment 6.9 —

(a) $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}} = (5, 13, 7)$

(b) $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}} = (3, -2, 1)$

Assignment 6.10 — $\vec{\boldsymbol{u}} \times (\vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}}) = \vec{\boldsymbol{u}} \times (3\hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}} + 2\hat{\boldsymbol{k}}) = -2\hat{\boldsymbol{\imath}} + 7\hat{\boldsymbol{\jmath}} - 4\hat{\boldsymbol{k}}$

$\rightarrow \vec{\boldsymbol{u}} \times (\vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}})$ lies in the plane of $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$

$(\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}) \times \vec{\boldsymbol{w}} = (9\hat{\boldsymbol{\imath}} + 6\hat{\boldsymbol{\jmath}} + -7\hat{\boldsymbol{k}}) \times \vec{\boldsymbol{w}} = \hat{\boldsymbol{\imath}} + 9\hat{\boldsymbol{\jmath}} + 9\hat{\boldsymbol{k}}$

$\rightarrow (\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}) \times \vec{\boldsymbol{w}}$ lies in the plane of $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$

Assignment 6.11 —

(a) $\vec{\boldsymbol{u}} = \hat{\boldsymbol{\imath}} - \hat{\boldsymbol{\jmath}}$ en $\vec{\boldsymbol{v}} = \hat{\boldsymbol{\jmath}} + 2\hat{\boldsymbol{k}}$

a) $\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}} = \hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{k}}, \quad \vec{\boldsymbol{u}} - \vec{\boldsymbol{v}} = \hat{\boldsymbol{\imath}} - 2\hat{\boldsymbol{\jmath}} - 2\hat{\boldsymbol{k}}, \quad 2\vec{\boldsymbol{u}} - 3\vec{\boldsymbol{v}} = 2\hat{\boldsymbol{\imath}} - 5\hat{\boldsymbol{\jmath}} - 6\hat{\boldsymbol{k}}$

b) $\|\vec{\boldsymbol{u}}\| = \sqrt{2}, \quad \|\vec{\boldsymbol{v}}\| = \sqrt{5}$

c) $\hat{\boldsymbol{u}} = \dfrac{1}{\sqrt{2}}(\hat{\boldsymbol{\imath}} - \hat{\boldsymbol{\jmath}}), \quad \hat{\boldsymbol{v}} = \dfrac{1}{\sqrt{5}}\left(\hat{\boldsymbol{\jmath}} + 2\hat{\boldsymbol{k}}\right)$

d) $\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}} = -1$

e) $\theta = \arccos\left(\frac{-1}{\sqrt{10}}\right)$

(b) $\vec{\boldsymbol{u}} = 3\hat{\boldsymbol{\imath}} + 4\hat{\boldsymbol{\jmath}} - 5\hat{\boldsymbol{k}}$ en $\vec{\boldsymbol{v}} = 3\hat{\boldsymbol{\imath}} - 4\hat{\boldsymbol{\jmath}} - 5\hat{\boldsymbol{k}}$

a) $\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}} = 6\hat{\boldsymbol{\imath}} - 10\hat{\boldsymbol{k}}$, $\quad \vec{\boldsymbol{u}} - \vec{\boldsymbol{v}} = 8\hat{\boldsymbol{\jmath}}$, $\quad 2\vec{\boldsymbol{u}} - 3\vec{\boldsymbol{v}} = -3\hat{\boldsymbol{\imath}} + 20\hat{\boldsymbol{\jmath}} + 5\hat{\boldsymbol{k}}$

b) $\|\vec{\boldsymbol{u}}\| = 5\sqrt{2}$, $\quad \|\vec{\boldsymbol{v}}\| = 5\sqrt{2}$

c) $\hat{\boldsymbol{u}} = \frac{1}{5\sqrt{2}}\left(3\hat{\boldsymbol{\imath}} + 4\hat{\boldsymbol{\jmath}} - 5\hat{\boldsymbol{k}}\right)$, $\quad \hat{\boldsymbol{v}} = \frac{1}{5\sqrt{2}}\left(3\hat{\boldsymbol{\imath}} - 4\hat{\boldsymbol{\jmath}} - 5\hat{\boldsymbol{k}}\right)$

d) $\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}} = 18$

e) $\theta = \arccos\left(\frac{9}{25}\right)$

Chapter 7

Assignment 7.1 — Line ℓ through A and B: $\quad \frac{x-1}{1} = \frac{y-2}{1} = \frac{z-3}{1}$.

$C(3,4,5) \in \ell: \quad 3-1 = 4-2 = 5-3 \quad \to$ ok

Assignment 7.2 — $\ell: \begin{cases} x = 8/7 - t \\ y = 4/7 + 10t \\ z = 7t \end{cases}$

Assignment 7.3 —

(a)
- Cartesian equation: $\frac{x}{1} = \frac{y}{2} = \frac{z}{3}$
- Vector equation: $\vec{\boldsymbol{l}}(t) = (0,0,0) + t(1,2,3)$
- Parameter equations: $\ell: \begin{cases} x = t \\ y = 2t \\ z = 3t \end{cases}$

(b)
- Cartesian equation: $\begin{cases} \frac{x-3}{-2} = \frac{z-1}{-1} \\ y = 4 \end{cases}$
- Vector equation: $\vec{\boldsymbol{l}}(t) = (3,4,1) + t(-2,0,-1)$
- Parameter equations: $\ell: \begin{cases} x = 3 - 2t \\ y = 4 \\ z = 1 - t \end{cases}$

(c)
- Cartesian equation: $\frac{x-1}{3/2} = \frac{y-2}{2} = \frac{z}{1/2}$

- Vector equation: $\vec{\boldsymbol{l}}(t) = (1,2,0) + t(3/2,2,1/2)$

- Parameter equations: $\ell : \begin{cases} x = 1 + 3/2t \\ y = 2 + 2t \\ z = 1/2t \end{cases}$

(d)
- Cartesian equation: $\dfrac{x-1}{2} = \dfrac{y-2}{-3} = \dfrac{z-3}{-4}$

- Vector equation: $\vec{\boldsymbol{l}}(t) = (1,2,3) + t(2,-3,-4)$

- Parameter equations: $\ell : \begin{cases} x = 1 + 2t \\ y = 2 - 3t \\ z = 3 - 4t \end{cases}$

(e)
- Cartesian equation: $\dfrac{x+1}{2} = \dfrac{y}{-1} = \dfrac{z-1}{7}$

- Vector equation: $\vec{\boldsymbol{l}}(t) = (-1,0,1) + t(2,-1,7)$

- Parameter equations: $\ell : \begin{cases} x = -1 + 2t \\ y = -t \\ z = 1 + 7t \end{cases}$

(f)
- Cartesian equation: $\dfrac{x}{7} = \dfrac{y}{-6} = \dfrac{z}{-5}$

- Vector equation: $\vec{\boldsymbol{l}}(t) = (0,0,0) + t(7,-6,-5)$

- Parameter equations: $\begin{cases} x = 7t \\ y = -6t \\ z = -5t \end{cases}$

(g)
- Cartesian equation: $\dfrac{x-2}{1} = \dfrac{y+1}{-1} = \dfrac{z+1}{-1}$

- Vector equation: $\vec{\boldsymbol{l}}(t) = (2,-1,-1) + t(1,-1,-1)$

- Parameter equations: $\begin{cases} x = 2 + t \\ y = -1 - t \\ z = -1 - t \end{cases}$

Assignment 7.4 —

(a) $\vec{\boldsymbol{d}}_1 = (3,4,2)$, $\vec{\boldsymbol{d}}_2 = (6,8,4) \quad \Rightarrow \quad \vec{\boldsymbol{d}}_2 = 2\vec{\boldsymbol{d}}_1 \quad \Rightarrow \quad \ell_1 \parallel \ell_2$

(b) ℓ_1 and ℓ_2 are skew.

(c) $\vec{\boldsymbol{d}}_1 = (8,-7,-5)$, $\vec{\boldsymbol{d}}_2 = (8,-7,-5) \quad \Rightarrow \quad \vec{\boldsymbol{d}}_1 = \vec{\boldsymbol{d}}_2 \quad \Rightarrow \quad \ell_1 \parallel \ell_2$

(d) ℓ_1 and ℓ_2 are skew.

Assignment 7.5 —

(a) $3x+y-4z+1=0$

(b) $-2x+z+1=0$

(c) $x-2y+3z+12=0$

(d) $x+6y-4z+4=0$

(e) $7x-3y-5z+18=0$

(f) $x-5y-3z+7=0$

(g) $5x-9y+5z+15=0$

Assignment 7.6 —

(a) The planes p_1 and p_2 are perpendicular if and only if $\vec{\boldsymbol{n}}_{p_1} \cdot \vec{\boldsymbol{n}}_{p_2} = 0$.

(b) The planes p_1 and p_3 are parallel if and only if $\vec{\boldsymbol{n}}_{p_1}$ and $\vec{\boldsymbol{n}}_{p_3}$ are parallel.

(c) The planes p_2 and p_3 are perpendicular to each other.

Assignment 7.7 — The angle between the planes p_1 and p_2 is $\arccos\left(\dfrac{-4}{\sqrt{91}}\right)$.

Assignment 7.8 —

(a) $3x^2-2y^2+z^2+3=0$: 2-leaf hyperboloid.

(b) $-4x^2+2z^2-3=0$: hyperbolic cylinder parallel to the y-axis.

(c) $3x^2-y^2=z^2$: cone with top in the origin and located around the x-axis.

(d) $(x-1)^2+(y-2)^2=(z-4)^2$: cone with top in $(1,2,4)$.

(e) $2x^2+3z^2=1$: elliptical cylinder parallel to the y-axis.

(f) $y^2+2z^2=x$: elliptical paraboloid.

(g) $x^2+4y^2+9z^2=36$: ellipsoid.

(h) $\dfrac{25}{9}x^2-25y^2+z^2=25$: 1-leaf hyperboloid.

(i) $y=4z^2-x^2$: hyperbolic paraboloid.

(j) $x^2+y^2+4z^2+2x=0$: ellipsoid with center in $(-1,0,0)$.

(k) $9x^2+y^2-4z^2+2y=0$: 1-leaf hyperboloid with center in $(0,-1,0)$.

(l) $x-2z^2=0$: parabolic cylinder parallel to the y-axis.

Assignment 7.9 — $\dfrac{x-2}{3}=\dfrac{y+1}{2}=\dfrac{z-5}{-2}$

Assignment 8.1 —

(a) 12
(b) 0
(c) $\frac{2}{3}$
(d) $\frac{1}{4}$
(e) 1
(f) 2
(g) 0
(h) $-\frac{1}{16}$
(i) -2
(j) $\frac{7}{12}$
(k) $\frac{7}{3}$
(l) -6
(m) 0
(n) $\frac{1}{4}$
(o) 8
(p) $\frac{8}{3}$

Assignment 8.2 — $\lim_{x\to 0}\left(x^2\sin\left(\frac{1}{x}\right)\right) = 0$

Use the fact that $-1 \leq \sin\left(\frac{1}{x}\right) \leq 1$, multiply both parts with x^2 and take the limit for $x \to 0$.

Assignment 8.3 —

(a) 1
(b) $+\infty$
(c) 1
(d) 2
(e) $-\infty$
(f) $+\infty$
(g) 2
(h) 0
(i) -1
(j) 0
(k) 1

Assignment 8.4 —

(a) -2
(b) 2
(c) π^2
(d) 1

Assignment 8.5 —

(a) yes, between $0,2$ en $0,25$
(b) yes, slightly smaller than -8
(c) no
(d) yes, between $\frac{\pi}{2}$ and $\frac{3\pi}{4}$
(e) yes, one between -4 en -3, one between -3 and 1 and one between 1 and 4

Assignment 8.6 —

(a) $-\infty$

(b) $+\infty : +\infty \quad \wedge \quad -\infty : -\infty$

(c) $+\infty : 1 \quad \wedge \quad -\infty : \frac{1}{3}$

(d) $+\infty : 0 \quad \wedge \quad -\infty : 0$

(e) $+\infty : 3 \quad \wedge \quad -\infty : 1$

(f) 2

(g) $+\infty : -1 \quad \wedge \quad -\infty : -1$

(h) $+\infty : 0 \quad \wedge \quad -\infty : +\infty$

(i) $+\infty : -\frac{5}{4} \quad \wedge \quad -\infty : -\infty$

(j) $-\frac{3}{5}$

(k) $+\infty : \frac{2}{\sqrt{3}} \quad \wedge \quad -\infty : -\frac{2}{\sqrt{3}}$

(l) $-\frac{2}{3}$

(m) $+\infty : +\infty \quad \wedge \quad -\infty : 2$

Assignment 8.7 —

	VA	HA	SA
(a)	$x = \frac{1}{2}$	$y = -\frac{1}{2}$	none
(b)	$x = 3$	none	$y = x + 2$
(c)	none	$y = 3$	none
(d)	none	none	$y = 2x$
(e)	$x = 0$	none	$y = x$
(f)	$x = -2$ $x = 0$ $x = 2$	none	$y = \frac{1}{3}x$
(g)	none	none	$y = x - 3\ (+\infty)$ $y = -x - 3\ (-\infty)$
(h)	none	none	$y = x + 1\ (+\infty)$ $y = -x - 1\ (-\infty)$
(i)	none	$y = 0\ (+\infty)$	$y = 2x\ (-\infty)$
(j)	$x = -2$ $x = 3$	none	$y = x + \frac{3}{2}\ (+\infty)$ $y = -x - \frac{3}{2}\ (-\infty)$

Assignment 8.8 —

(a) 0

(b) $-\infty$

(c) 0

(d) 3

(e) 1

(f) $\frac{2}{5}$

(g) $\frac{1}{2}$

(h) $\frac{49}{4}$

(i) 8

(j) $\frac{1}{2}$

Assignment 8.9 —

(a) $\sqrt{e}$

(b) e^3

(c) -1

(d) a

Assignment 8.10 —

(a) $f(x)$ is continuous over $\mathbb{R}$.

(b) $f(x)$ is continuous over $]-\infty,-3[$, in $]-3,3[$ and in $]3,+\infty[$.

(c) $f(x)$ is continuous over $\mathbb{R}$.

(d) $f(x)$ is continuous over $]-\infty,-1]$ and in $[1,+\infty[$.

(e) $f(x)$ is continuous over $\mathbb{R}$.

(f) $f(x)$ is continuous over $\mathbb{R}_0^-$ and in $\mathbb{R}_0^+$.

Assignment 8.11 — $f(x)$ is discontinuous in $x=1, x=2, x=3, x=4$ and $x=5$. f is left continuous in $x=4$ en right continuous in $x=2$ and $x=5$.

Assignment 8.12 —

(a) $a=-1$

(b) $a=2$

Assignment 8.13 —

(a) perforation in $x=-3$

(b) 2 vertical asymptotes in $x=-3$ and $x=2$

(c) 1 vertical asymptote in $x=-1$ and 1 perforation in $x=3$

Chapter 9

Assignment 9.1 —

(a) $f(x)=|x|$ is continuous in $x=0$, but not differentiable because the left and right derivative are not equal.

(b) $f(x)=|x^2-1|$ is continuous in $x=1$, but not differentiable because the left and right derivative are not equal.

(c) $f(x)=|\sin(x)|$ is continuous in $x=0$, but not differentiable because the left and right derivative are not equal.

(d) $f(x)=\sqrt{x}$ is right-continuous in $x=0$, but not differentiable because the derivative does not exist at $x=0$. The function is differentiable over $\mathbb{R}_0^+$.

(e) $f(x) = \sqrt{1-x^2}$ is right-continuous in $x = -1$, but not differentiable because the derivative does not exist at $x = -1$.

Assignment 9.2 —

(a) tangent: $y = 2$, normal: $x = 0$

(b) tangent: $y = 2x+1$, normal: $y = -\frac{1}{2}x - \frac{3}{2}$

(c) tangent: $y = \frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}\left(1-\frac{\pi}{4}\right)$, normal: $y = -\sqrt{2}\,x + \frac{\sqrt{2}}{2}\left(1+\frac{\pi}{2}\right)$

(d) tangent: $y = 2x+1-\frac{\pi}{2}$, normal: $y = -\frac{1}{2}x+1+\frac{\pi}{8}$

(e) tangent: $y = 2x+2$, normal: $y = -\frac{1}{2}x+2$

(f) tangent: $y = 4$, normal: $x = 2$

Assignment 9.3 —

(a) $0^+ :\ y = x \quad \wedge \quad 0^- :\ y = -x$

(b) $0^+ :\ x = 0 \quad \wedge \quad 0^- :\ x = 0$

(c) $y = 0$

(d) $y = 0$

Assignment 9.4 — $y = 6x-9$ and $y = -2x-1$

Assignment 9.5 —

(a) $f'(x) = 4x+3$

(b) $f'(x) = -\frac{4}{x^3}$

(c) $f'(x) = \frac{6x^2+4x-1}{(3x+1)^2}$

(d) $f'(x) = 5x^4-16x^3+6x-6$

(e) $f'(x) = \sqrt{x^2-1} + \frac{x^2}{\sqrt{x^2-1}}$

(f) $f'(x) = -\frac{7}{(2x-1)^2}$

(g) $f'(x) = \frac{1}{(x+1)^2}\sqrt{\frac{x+1}{x-1}}$

(h) $f'(x) = -\frac{2(14x+3)}{(7x^2+3x-6)^3}$

(i) $f'(x) = \frac{4}{3\sqrt[3]{2x+3}}$

(j) $f'(x) = \frac{4x+3}{3\sqrt[3]{(x+1)^2}}$

(k) $f'(x) = \frac{1+6x}{(1+2x+6x^2)^2}$

(l) $f'(x) = \frac{2(\sqrt{x-2}+\sqrt{3})^3}{9\sqrt{x-2}}$

Assignment 9.6 —

(a) $f'(x) = -5\sin(5x)$

(b) $f'(x) = -\dfrac{1}{\sin^2(x)}$

(c) $f'(x) = \sin^2(x) + x\sin(2x)$

(d) $f'(x) = x^2(3\cos(x) - x\sin(x))$

(e) $f'(x) = -e^{-x}$

(f) $f'(x) = x^4\left(e^x(x+5) + \left(x\tan(x)+5\right)\sec(x)\right)$

(g) $f'(x) = \dfrac{-2e^{-2x}}{x^3}(x+1)$

(h) $f'(x) = \dfrac{-e^x}{e^{2x}-1}$

(i) $f'(x) = -\dfrac{e^{-\arcsin(x)}}{\sqrt{1-x^2}}$

(j) $f'(x) = \sec^2(e^x)e^x$

(k) $f'(x) = \dfrac{2\cos(2x)+\sin(2x)}{\sin(2x)+\sin^2(x)}$

(l) $f'(x) = \dfrac{x}{\ln(10)\left(x^2-9\right)}$

(m) $f'(x) = \dfrac{4(2x+1)}{\ln(2)\left(x^2+x+2\right)}$

(n) $f'(x) = \dfrac{2}{\ln(3)\left(4x^2-1\right)}$

(o) $f'(x) = \dfrac{3x^2}{1+x^6}$

(p) $f'(x) = \dfrac{1}{(1+x)\sqrt{x}}$

(q) $f'(x) = \dfrac{\pm 1}{(x-1)\sqrt{1-2x}}$

(r) $f'(x) = \dfrac{1}{2\sqrt{x}\left(1+\cos\left(\sqrt{x}\right)\right)}$

(s) $f'(x) = -\dfrac{1}{\sqrt{a^2-(x-b)^2}}$

(t) $f'(x) = \dfrac{x}{\sqrt{1-x^4}\sqrt{\arcsin(x^2)}}$

(u) $f'(x) = \sqrt{\dfrac{a-x}{a+x}}$

Assignment 9.7 —

(a) $f'(x) = \dfrac{1}{\cos(x)}$

(b) $f'(x) = \dfrac{1}{\cos^3(x)}$

(c) $f'(x) = \dfrac{\sqrt{x^2-a^2}}{x^2}$

(d) $f'(x) = \left(1-x^2\right)^{\frac{3}{2}}$

(e) $f'(x) = \dfrac{1}{x\sqrt{x+1}}$

(f) $f'(x) = \dfrac{x}{(ax+b)^2}$

(g) $f'(x) = 8x^2\sqrt{a^2-x^2}$

(h) $f'(x) = \left(3-2x-x^2\right)^{\frac{3}{2}}$

Assignment 9.8 —

(a) $y' = -\dfrac{2e^2x}{2^y\ln(2)}$

(b) $y' = -\dfrac{x}{y^2}$

(c) $y' = \dfrac{1-y}{2+x}$

(d) $y' = \dfrac{-3x^2y-y^5}{x^3+5xy^4}$

(e) $y' = \dfrac{x}{4(1-y)}$

(f) $y' = -\dfrac{3x^2+2xy}{x^2+4y}$

(g) $y' = \dfrac{1-\cos(x)}{1+\sin(y)}$

(h) $y' = -\dfrac{2x+y}{2y+x}$

Assignment 9.9 —

(a) $y = x$

(b) $y = -x + 2$

(c) $y = -x + 1$

Assignment 9.10 — $y' = \dfrac{-x}{4y}, \quad y'' = -\dfrac{1}{4y^3}$

Assignment 9.11 —

(a) $y' = \dfrac{x^x\left(x + (x+1)\ln(x)\right)}{(x+1)^2}$

(b) $y' = x^{\sin(x)+2}\left(\cos(x)\ln(x) + \dfrac{\sin(x)+2}{x}\right)$

(c) $y' = (\sin(x))^{\ln(x)}\left(\dfrac{\ln(\sin(x))}{x} + \ln(x)\cot(x)\right)$

(d) $y' = -\dfrac{2\ln(x)}{x}\left(\dfrac{1}{x}\right)^{\ln(x)}$

(e) $y' = (\cos x)^x(\ln(\cos(x)) - x\tan(x)) - x^{\cos(x)}\left(-\sin(x)\ln(x) + \dfrac{1}{x}\cos(x)\right)$

(f) $y' = \dfrac{x^{\ln(x)}(\sin(x))^x}{x^x\ln(x)}\left(\dfrac{2\ln(x)}{x} + \ln(\sin(x)) + x\cot(x) - \ln(x) - 1 - \dfrac{1}{x\ln(x)}\right)$

Assignment 9.12 — $(f^{-1})'(x) = 1 + f^{-1}(x)$

Assignment 9.13 —

(a) $(f^{-1})'(x) = \dfrac{1}{6y^2} = \dfrac{1}{6(f^{-1}(x))^2}$

(b) $(f^{-1})'(-2) = \dfrac{1}{f'(-1)} = \dfrac{2}{5}$

(c) $(f^{-1})'(2) = \dfrac{1}{f'(1)} = \dfrac{1}{4}$

(d) $(f^{-1})'(10) = \dfrac{1}{f'(2)} = \dfrac{1}{13}$

(e) $(f^{-1})'(\sqrt{3}/2) = \dfrac{1}{f'(\pi/6)} = 1$

(f) $(f^{-1})'(6) = \dfrac{1}{f'(0)} = \dfrac{1}{18}$

Assignment 9.14 —

(a) $+\infty$

(b) $\dfrac{3}{5}$

(c) 0

(d) $\dfrac{1}{2}$

(e) 1

(f) 0

(g) $-\dfrac{1}{2}$

(h) -3

(i) 1

(j) e^3

(k) 1

(l) 1

(m) $e^{-\frac{3}{2}}$

(n) $\dfrac{1}{\sqrt{e}}$

(o) 1

Assignment 9.15 — 0.01 cm

Assignment 9.16 — 1005 cm^3

Assignment 9.17 —

(a) $\dfrac{x^3}{1-2x^2} = \displaystyle\sum_{n=0}^{+\infty} 2^n x^{2n+3}$

(b) $\dfrac{1-x}{1+x} = 1 + 2\displaystyle\sum_{n=1}^{+\infty} (-x)^n$

(c) $\dfrac{1}{(2-x)^2} = \displaystyle\sum_{n=0}^{+\infty} \dfrac{(n+1)x^n}{2^{n+2}}$

(d) $\ln(2-x) = \ln(2) - \displaystyle\sum_{n=1}^{+\infty} \dfrac{x^n}{n\,2^n}$

(e) $\ln(x) = \ln(4) - \displaystyle\sum_{n=1}^{+\infty} \dfrac{(-1)^n (x-4)^n}{n\,4^n}$

Assignment 9.18 —

(a) $x^2 \sin\left(\dfrac{x}{3}\right) = \displaystyle\sum_{n=0}^{+\infty} \dfrac{(-1)^n x^{2n+3}}{3^{2n+1}(2n+1)!}$

(b) $\cos^2\left(\dfrac{x}{2}\right) = 1 + \dfrac{1}{2}\displaystyle\sum_{n=1}^{+\infty} (-1)^n \dfrac{x^{2n}}{(2n)!}$

(c) $\dfrac{e^{2x^2}-1}{x^2} = \displaystyle\sum_{n=0}^{+\infty} \dfrac{2^{n+1}}{(n+1)!} x^{2n}$

(d) $(1+x)^{\frac{1}{2}} \cos(x) = 1 + \dfrac{x}{2} - \dfrac{5x^2}{8} - \dfrac{3x^3}{16} + \dfrac{25x^4}{384} + \dfrac{13x^5}{768} + \dots$

Assignment 9.19 —

(a) $\dfrac{1}{x} = \displaystyle\sum_{n=1}^{+\infty} (-1)^{n-1}(x-1)^{n-1}$

(b) $\dfrac{1}{x^2} = \dfrac{1}{4}\displaystyle\sum_{n=1}^{+\infty} \dfrac{n(x+2)^{n-1}}{2^{n-1}}$

(c) $\sin(x) - \cos(x) = \sqrt{2}\displaystyle\sum_{n=0}^{+\infty} \dfrac{(-1)^n}{(2n+1)!}\left(x - \dfrac{\pi}{4}\right)^{2n+1}$

(d) $x\ln(x) = (x-1) + \displaystyle\sum_{n=2}^{+\infty} \dfrac{(-1)^n}{n(n-1)}(x-1)^n$

(e) $xe^x = -\dfrac{2}{e^2} + \dfrac{1}{e^2}\displaystyle\sum_{n=1}^{+\infty} \dfrac{n-2}{n!}(x+2)^n$

(f) $\ln(2+x) = \ln 4 + \displaystyle\sum_{n=1}^{+\infty} (-1)^{n-1} \dfrac{(x-2)^n}{n4^n}$

(g) $\cos^2(x) = \dfrac{1}{2} + \dfrac{1}{2\sqrt{2}} + \dfrac{1}{2\sqrt{2}}\displaystyle\sum_{n=1}^{+\infty} (-1)^n \left(\dfrac{2^{2n-1}}{(2n-1)!}\left(x - \dfrac{\pi}{8}\right)^{2n-1} + \dfrac{2^{2n}}{(2n)!}\left(x - \dfrac{\pi}{8}\right)^{2n}\right)$

Chapter 10

Assignment 10.1 — $f'(x) = x^{(x^2)}x(2\ln(x)+1) = 0 \quad \Leftrightarrow \quad x = e^{-1/2}$

$f''(x) = x^{(x^2)}x^2(2\ln(x)+1)^2 + x^{(x^2)}(2\ln(x)+3) \quad \Rightarrow f''\left(e^{-1/2}\right) = 2\left(e^{-1/2}\right)^{e^{-1}} > 0$

$\Rightarrow$ $x = e^{-1/2}$ is a minimum

Assignment 10.2 — $f'(v) = \dfrac{\sqrt{2/\pi}\, m^{3/2}\ v\ e^{-mv^2/(2kT)}\left(2kT - mv^2\right)}{(kT)^{5/2}}$. $f(v)$ is at a maximum if $v = \sqrt{\dfrac{2kT}{m}}$.

Assignment 10.3 —

(a) $c = 0$

(b) $c = 3/\sqrt{2}$

(c) The mean value theorem is not applicable.

Assignment 10.4 —

(a) $c = -1/2$

(b) Rolle's theorem is not applicable.

(c) Rolle's theorem is not applicable.

Assignment 10.5 —

(a) Asymptotes: $y = x$ and $y = -x$

Derivatives: $f'(x) = \dfrac{x}{\sqrt{x^2-1}}$ en $f''(x) = -\dfrac{1}{(x^2-1)^{\frac{3}{2}}}$

x	$-\infty$		-1		1		$+\infty$
$f'(x)$	$-$	$-$	|	///	|	$+$	$+$
$f''(x)$	$-$	$-$	|	///	|	$-$	$-$
$f(x)$	$(+\infty)$	$\searrow$	0	///	0	$\nearrow$	$(+\infty)$
		$\cap$				$\cap$	

(b) Asymptotes: $y = x-2$ and $y = -x+2$

Derivatives: $f'(x) = \dfrac{x-2}{\sqrt{x^2-4x+3}}$ and $f''(x) = -\dfrac{1}{(x^2-4x+3)^{\frac{3}{2}}}$

x	$-\infty$		1		3		$+\infty$
$f'(x)$	$-$	$-$	|	///	|	$+$	$+$
$f''(x)$	$-$	$-$	|	///	|	$-$	$-$
$f(x)$	$(+\infty)$	$\searrow$	0	///	0	$\nearrow$	$(+\infty)$
	S.A.	$\cap$				$\cap$	S.A.

(c) Asymptotes: $x=-\sqrt{3}$, $x=\sqrt{3}$, $y=-x$

Derivatives: $f'(x)=\dfrac{x^2(9-x^2)}{(3-x^2)^2}$ and $f''(x)=\dfrac{6x(9+x^2)}{(3-x^2)^3}$

x	$-\infty$		-3		$-\sqrt{3}$		0		$\sqrt{3}$		3		$+\infty$
$f'(x)$	$-$	$-$	0	$+$	$\vert$	$+$	0	$+$	$\vert$	$+$	0	$-$	$-$
$f''(x)$	$+$	$+$		$+$	$\vert$	$-$	0	$+$	$\vert$	$-$		$-$	$-$
$f(x)$	$(+\infty)$	$\searrow$	$9/2$	$\nearrow$	$\vert$	$\nearrow$	0	$\nearrow$	$\vert$	$\nearrow$	$-9/2$	$\searrow$	$(-\infty)$
	S.A.	$\cup$		$\cup$	V.A.	$\cap$		$\cup$	V.A.	$\cap$		$\cap$	S.A.

(d) Asymptotes: $x=-\sqrt{3}$, $x=\sqrt{3}$, $y=-1$ and $y=1$

Derivatives: $f'(x)=-\dfrac{7x+3}{(x^2-3)^{\frac{3}{2}}}$ and $f''(x)=\dfrac{14x^2+9x+21}{(x^2-3)^{\frac{5}{2}}}$

x	$-\infty$		-7		$-\sqrt{3}$		$\sqrt{3}$		$+\infty$
$f'(x)$	$+$	$+$	$+$	$+$	$\vert$	///	$\vert$	$-$	$-$
$f''(x)$	$+$	$+$	$+$	$+$	$\vert$	///	$\vert$	$+$	$+$
$f(x)$	(-1)	$\nearrow$	0	$\nearrow$	$\vert$	///	$\vert$	$\searrow$	(1)
	H.A.	$\cup$	$\cup$	$\cup$	V.A.		V.A.	$\cup$	H.A.

(e) Asymptotes: $y=0$

Derivatives: $f'(x)=-xe^{-\frac{x^2}{2}}$ and $f''(x)=\left(x^2-1\right)e^{-\frac{x^2}{2}}$

x	$-\infty$		-1		0		1		$+\infty$
$f'(x)$	$+$	$+$	$+$	$+$	0	$-$	$-$	$-$	$-$
$f''(x)$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$
$f(x)$	(0)	$\nearrow$	$\nearrow$	$\nearrow$	1	$\searrow$	$\searrow$	$\searrow$	(0)
	H.A.	$\cup$	inf.p.	$\cap$	max.	$\cap$	inf.p.	$\cup$	H.A.

(f) Asymptotes: $y=0$

Derivatives: $f'(x)=e^{-x}\left(-x^3+3x^2\right)$ and $f''(x)=e^{-x}\left(x^3-6x^2+6x\right)$

x	$-\infty$		0		$3-\sqrt{3}$		3		$3+\sqrt{3}$		$+\infty$
$f'(x)$	$+$	$+$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	$-$
$f''(x)$	$-$	$-$	0	$+$	0	$-$	$-$	$-$	0	$+$	$+$
$f(x)$	$(-\infty)$	$\nearrow$	0	$\nearrow$	$(3-\sqrt{3})e^{-3+\sqrt{3}}$	$\nearrow$	$27e^{-3}$	$\searrow$	$(3+\sqrt{3})e^{-3-\sqrt{3}}$	$\searrow$	(0)
		$\cap$	inf.p.	$\cup$	inf.p.	$\cap$	max.	$\cap$	inf.p.	$\cup$	H.A.

(g) Asymptotes: $x=0$ and $y=0$

Derivatives: $f'(x) = -\dfrac{e^{-x}(x+3)}{x^4}$ and $f''(x) = \dfrac{e^{-x}\left(x^2+6x+12\right)}{x^5}$

x	$-\infty$		-3		0		$+\infty$
$f'(x)$	$+$	$+$	0	$-$	$\vert$	$-$	$-$
$f''(x)$	$-$	$-$	$-$	$-$	$\vert$	$+$	$+$
$f(x)$	$(-\infty)$	$\nearrow$	$-\dfrac{e^3}{27}$	$\searrow$	$\vert$	$\searrow$	(0)
		$\cap$	max.	$\cap$	V.A.	$\cup$	H.A.

(h) Asymptotes: $x=0$ and $y=\ln(2)x$

Derivatives: $f'(x) = \dfrac{2^x\ln(2)}{2^x-1}$ and $f''(x) = -\dfrac{2^x\ln^2(2)}{(2^x-1)^2}$

x		0		1		$+\infty$
$f'(x)$	$///$	$\vert$	$+$	$+$	$+$	$+$
$f''(x)$	$///$	$\vert$	$-$	$-$	$-$	$-$
$f(x)$	$///$	$\vert$	$\nearrow$	0	$\nearrow$	$(+\infty)$
		V.A.	$\cap$	$\cap$	$\cap$	S.A.

(i) Asymptotes: $x=0$ and $y=0$

Derivatives: $f'(x) = \dfrac{1-2\ln(x)}{x^3}$ and $f''(x) = \dfrac{6\ln(x)-5}{x^4}$

x		0		1		$\sqrt{e}$		$e^{\frac{5}{6}}$		$+\infty$
$f'(x)$	$///$	$\vert$	$+$	$+$	$+$	0	$-$	$-$	$-$	$-$
$f''(x)$	$///$	$\vert$	$-$	$-$	$-$	$-$	$-$	0	$+$	$+$
$f(x)$	$///$	$\vert$	$\nearrow$	$\nearrow$	$\nearrow$	$\dfrac{1}{2e}$	$\searrow$	$\searrow$	$\searrow$	(0)
		V.A.	$\cap$	$\cap$	$\cap$	max.	$\cap$	inf.p.	$\cup$	H.A.

(j) Asymptotes: $y=\dfrac{x}{2}$ and $y=-\dfrac{x}{2}$

Derivatives: $f'(x) = \dfrac{e^x-e^{-x}}{2(e^x+e^{-x})}$ and $f''(x) = \dfrac{2}{(e^x+e^{-x})^2}$

x	$-\infty$		0		$+\infty$
$f'(x)$	$-$	$-$	0	$+$	$+$
$f''(x)$	$+$	$+$	$+$	$+$	$+$
$f(x)$	$(+\infty)$	$\searrow$	$\dfrac{\ln(2)}{2}$	$\nearrow$	$(+\infty)$
	S.A.	$\cup$	min.	$\cup$	S.A.

(k) Asymptotes: $x=e$ and $y=-2$

Derivatives: $f'(x) = \dfrac{2}{(\ln(x)-1)^2\,x}$ and $f''(x) = -\dfrac{2(\ln(x)+1)}{(\ln(x)-1)^3\,x^2}$

x		0		e^{-1}		e		$+\infty$
$f'(x)$	///	|	$+$	$+$	$+$	|	$+$	$+$
$f''(x)$	///	|	$-$	0	$+$	|	$-$	$-$
$f(x)$	///	(-2)	$\nearrow$	$\nearrow$	$\nearrow$	|	$\nearrow$	(-2)
			$\cap$	inf.p.	$\cup$	V.A.	$\cap$	H.A.

(l) Asymptotes: $x = \dfrac{\pi}{2} + k\pi,\ k \in \mathbb{Z}$

Derivatives: $f'(x) = -\tan(x)$ and $f''(x) = -\dfrac{1}{\cos^2(x)}$

x	$\cdots$		$-\frac{\pi}{2}$		0		$\frac{\pi}{2}$		$\frac{3\pi}{2}$		2π		$\frac{5\pi}{2}$		$\cdots$
$f'(x)$	$\cdots$	///	|	$+$	0	$-$	|	///	|	$+$	0	$-$	|	///	$\cdots$
$f''(x)$	$\cdots$	///	|	$-$	$-$	$-$	|	///	|	$-$	$-$	$-$	|	///	$\cdots$
$f(x)$	$\cdots$	///	|	$\nearrow$	max.	$\searrow$	|	///	|	$\nearrow$	max.	$\searrow$	|	///	$\cdots$
			V.A.	$\cap$	$\cap$	$\cap$	V.A.		V.A.	$\cap$	$\cap$	$\cap$	V.A.		

There are an infinite number of maxima at $x = 2k\pi$, with $k \in \mathbb{Z}$.

(m) Asymptotes: $y = \dfrac{\pi}{2}$

Derivatives: $f'(x) = \dfrac{1}{x(1+\ln^2(x))}$ and $f''(x) = -\dfrac{(1+\ln(x))^2}{x^2(1+\ln^2(x))^2}$

x		0		e^{-1}		1		$+\infty$
$f'(x)$	///	|	$+$	$+$	$+$	$+$	$+$	$+$
$f''(x)$	///	|	$-$	0	$-$	$-$	$-$	$-$
$f(x)$	///	$\left(-\frac{\pi}{2}\right)$	$\nearrow$	$\nearrow$	$\nearrow$	0	$\nearrow$	$\left(\frac{\pi}{2}\right)$
			$\cap$	$\cap$	$\cap$	$\cap$	$\cap$	H.A.

(n) Asymptotes: $y = 0$

Derivatives: $f'(x) = -\dfrac{1}{1+x^2}$ and $f''(x) = \dfrac{2x}{\left(1+x^2\right)^2}$

x	$-\infty$		0		$+\infty$
$f'(x)$	$-$	$-$	|	$-$	$-$
$f''(x)$	$-$	$-$	|	$+$	$+$
$f(x)$	(0)	$\searrow$	$\left(-\frac{\pi}{2} \,\middle|\, \frac{\pi}{2}\right)$	$\searrow$	(0)
	H.A.	$\cap$	V.A.	$\cup$	H.A.

(o) Asymptotes: none

Derivatives: $f'(x) = x^x(\ln(x)+1)$ and $f''(x) = x^x\left((\ln(x)+1)^2 + \dfrac{1}{x}\right)$

x		0		e^{-1}		$+\infty$
$f'(x)$	///	|	$-$	0	$+$	$+$
$f''(x)$	///	|	$+$	$+$	$+$	$+$
$f(x)$	///	(1)	$\searrow$	$e^{-e^{-1}}$	$\nearrow$	$(+\infty)$
			$\cup$	min.	$\cup$	

(p) Asymptotes: $y = 0$

Derivatives: $f'(x) = (x^2)^x(\ln(x^2)+2)$ and $f''(x) = (x^2)^x\left(\left(\ln(x^2)+2\right)^2 + \dfrac{2}{x}\right)$

x	$-\infty$		$-0,8$		$-\dfrac{1}{e}$		0		$\dfrac{1}{e}$		$+\infty$
$f'(x)$	$+$	$+$	$+$	$+$	0	$-$	|	$-$	0	$+$	$+$
$f''(x)$	$+$	$+$	0	$-$	$-$	$-$	|	$+$	$+$	$+$	$+$
$f(x)$	(0)	$\nearrow$	$\nearrow$	$\nearrow$	$e^{2e^{-1}}$	$\searrow$	(1)	$\searrow$	$e^{-2e^{-1}}$	$\nearrow$	$(+\infty)$
	H.A.	$\cup$	inf.p.	$\cap$	max.	$\cap$		$\cup$	min.	$\cup$	$\cup$

(q) Asymptotes: none

Derivatives: $f'(x) = 1 - 2\cos(x)$ and $f''(x) = 2\sin(x)$

x	$\cdots$	0		$\dfrac{\pi}{3}$		π		$\dfrac{5\pi}{3}$		2π		$\dfrac{7\pi}{3}$		3π	$\cdots$
$f'(x)$	$\cdots$	$-$	$-$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$	$\cdots$
$f''(x)$	$\cdots$	0	$+$	$+$	$+$	0	$-$	$-$	$-$	0	$+$	$+$	$+$	0	$\cdots$
$f(x)$	$(-\infty)$	0	$\searrow$	$f\left(\dfrac{\pi}{3}\right)$	$\nearrow$	$f(\pi)$	$\nearrow$	$f\left(\dfrac{5\pi}{3}\right)$	$\searrow$	$f(2\pi)$	$\searrow$	$f\left(\dfrac{7\pi}{3}\right)$	$\nearrow$	$f(3\pi)$	$(+\infty)$
	$\cdots$	inf.p.	$\cup$	min.	$\cup$	inf.p.	$\cap$	max.	$\cap$	inf.p.	$\cup$	min.	$\cup$	inf.p.	$\cdots$

(r) Asymptotes: y=0

Derivatives: $f'(x) = e^{-x}(-\sin(x)+\cos(x))$ and $f''(x) = -2e^{-x}\cos(x)$

x	0		$\dfrac{\pi}{4}$		$\dfrac{\pi}{2}$		π		$\dfrac{5\pi}{4}$		$\dfrac{3\pi}{2}$		2π		$\dfrac{9\pi}{4}$		$\dfrac{5\pi}{2}$	$\cdots$
$f'(x)$	$+$	$+$	0	$-$	$-$	$-$	$-$	$-$	0	$+$	$+$	$+$	$+$	$+$	0	$-$	$-$	$\cdots$
$f''(x)$	$-$	$-$	$-$	$-$	0	$+$	$+$	$+$	$+$	$+$	0	$-$	$-$	$-$	$-$	$-$	0	$\cdots$
$f(x)$	0	$\nearrow$	$f\left(\dfrac{\pi}{4}\right)$	$\searrow$	$f\left(\dfrac{\pi}{2}\right)$	$\searrow$	0	$\searrow$	$f\left(\dfrac{5\pi}{4}\right)$	$\nearrow$	$f\left(\dfrac{3\pi}{2}\right)$	$\nearrow$	0	$\nearrow$	$f\left(\dfrac{9\pi}{4}\right)$	$\searrow$	$f\left(\dfrac{5\pi}{2}\right)$	(0)
		$\cap$	max.	$\cap$	inf.p.	$\cup$		$\cup$	min.	$\cup$	inf.p.	$\cap$		$\cap$	max.	$\cap$	inf.p.	$\cdots$

(s) Asymptotes: none

Derivatives: $f'(x) = 1 + \cos(x)$ and $f''(x) = -\sin(x)$

x	$\cdots$	$\cdots$	-2π		$-\pi$		0		π		2π		3π	$\cdots$	$\cdots$
$f'(x)$	$\cdots$	$\cdots$	$+$	$+$	0	$+$	$+$	$+$	0	$+$	$+$	$+$	0	$\cdots$	$\cdots$
$f''(x)$	$\cdots$	$+$	0	$-$	0	$+$	0	$-$	0	$+$	0	$-$	0	$+$	$\cdots$
$f(x)$	$(-\infty)$	$\nearrow$	$f(-2\pi)$	$\nearrow$	$f(-\pi)$	$\nearrow$	$f(0)$	$\nearrow$	$f(\pi)$	$\nearrow$	$f(2\pi)$	$\nearrow$	$f(3\pi)$	$\nearrow$	$(+\infty)$
	$\cdots$	$\cup$	inf.p.	$\cap$	inf.p.	$\cup$	inf.p.	$\cap$	inf.p.	$\cup$	inf.p.	$\cap$	inf.p.	$\cup$	$\cdots$

(t) $f(x) = \dfrac{|1+x|-1}{x} = \begin{cases} 1, & \text{if } x \geq -1, \\ -\dfrac{x+2}{x} = -1 - \dfrac{2}{x}, & \text{if } x < -1. \end{cases}$

Asymptotes: $y = -1$ (for $x \to -\infty$)

Derivatives: $f'(x) = \begin{cases} 0, & \text{if } x \geq -1, \\ \dfrac{2}{x^2}, & \text{if } x < -1. \end{cases}$ and $f''(x) = \begin{cases} 0, & \text{if } x \geq -1, \\ \dfrac{-4}{x^3}, & \text{if } x < -1. \end{cases}$

x	$-\infty$		-2		-1		0		$+\infty$
$f'(x)$	$+$	$+$	$+$	$+$	2	0	0	0	0
$f''(x)$	$+$	$+$	$+$	$+$	4	0	0	0	0
$f(x)$	(-1)	$\nearrow$	0	$\nearrow$	1	1	(1)	1	1
	H.A.	$\cup$		$\cup$		$-$		$-$	

(u) $f(x) = \left|2 - \sqrt{2x+4}\right| \quad \Rightarrow \quad \operatorname{dom} f(x) = [-2, +\infty[, \quad f(x) = 0 \iff x = 0$

$\Rightarrow f(x) = \begin{cases} 2 - \sqrt{2x+4}, & \text{if } -2 \leq x < 0, \\ -2 + \sqrt{2x+4}, & \text{if } x > 0. \end{cases}$

- $f(x) = 2 - \sqrt{2x+4}$ with $-2 \leq x < 0$

  Asymptotes: none

  Derivatives: $f'(x) = -\dfrac{1}{\sqrt{2x+4}}$ and $f''(x) = \dfrac{1}{\sqrt{(2x+4)^3}}$

- $f(x) = -2 + \sqrt{2x+4}$ with $x > 0$

  Asymptotes: none

  Derivatives: $f'(x) = \dfrac{1}{\sqrt{2x+4}}$ and $f''(x) = -\dfrac{1}{\sqrt{(2x+4)^3}}$

x	-2		0			$+\infty$
$f'(x)$	$\mid$	$-$	$\left(-\frac{1}{2} \middle\vert \frac{1}{2}\right)$	$+$	$+$	$+$
$f''(x)$	$\mid$	$+$	$\left(\frac{1}{8} \middle\vert -\frac{1}{8}\right)$	$-$	$-$	$-$
$f(x)$	2	$\searrow$	0	$\nearrow$	$\nearrow$	$(+\infty)$
		$\cup$	inf.p.	$\cap$	$\cap$	$\cap$

(v) $f(x) = \dfrac{x^2}{x|x|+1} \quad \Rightarrow \quad f(x) = 0 \iff x = 0$

$\Rightarrow f(x) = \begin{cases} \dfrac{x^2}{-x^2+1}, & \text{if } x < 0, \\ \dfrac{x^2}{x^2+1}, & \text{if } x > 0. \end{cases}$

- $f(x) = \dfrac{x^2}{-x^2+1}$ with $x < 0$

Asymptotes: $x = -1$ and $y = -1$

Derivatives: $f'(x) = \dfrac{2x}{(-x^2+1)^2}$ $(x \neq -1)$ and $f''(x) = \dfrac{2(3x^2+1)}{(-x^2+1)^3}$ $(x \neq -1)$

- $f(x) = \dfrac{x^2}{x^2+1}$ with $x > 0$

Asymptotes: $y = 1$

Derivatives: $f'(x) = \dfrac{2x}{(x^2+1)^2}$ and $f''(x) = \dfrac{2(-3x^2+1)}{(x^2+1)^3}$

x	$-\infty$		-1		0		$\frac{\sqrt{3}}{3}$		$+\infty$
$f'(x)$	$-$	$-$	|	$-$	0	$+$	$+$	$+$	$+$
$f''(x)$	$-$	$-$	|	$+$	$+$	$+$	0	$-$	$-$
$f(x)$	(-1)	$\searrow$	|	$\searrow$	0	$\nearrow$	$f\left(\frac{\sqrt{3}}{3}\right)$	$\nearrow$	(1)
		$\cap$	V.A	$\cup$	min.	$\cup$	inf.p.	$\cap$	

(w) $f(x) = \left|(x-2)^2 - 4\right| \quad \Rightarrow \quad f(x) = 0 \quad \Leftrightarrow \quad x = 0 \quad \vee \quad x = 4$

$$\Rightarrow f(x) = \begin{cases} (x-2)^2 - 4, & \text{if } x < 0, \\ -(x-2)^2 + 4, & \text{if } 0 < x < 4, \\ (x-2)^2 - 4, & \text{if } x > 4. \end{cases}$$

Asymptotes: none

Derivatives: $f'(x) = 2(x-2)$ and $f''(x) = 2$ if $x < 0$ or $x > 4$

$f'(x) = -2(x-2)$ and $f''(x) = -2$ if $0 < x < 4$

x	$-\infty$		0		2		4		$+\infty$
$f'(x)$	$-$	$-$	$(-4\,\vert\,4)$	$+$	0	$-$	$(-4\,\vert\,4)$	$+$	$+$
$f''(x)$	$+$	$+$	$+$	$-$	$-$	$-$	$+$	$+$	$+$
$f(x)$	$(+\infty)$	$\searrow$	0	$\nearrow$	4	$\searrow$	0	$\nearrow$	$(+\infty)$
		$\cup$	inf.p.	$\cap$	max.	$\cap$	inf.p.	$\cup$	

Assignment 10.6 — The minimum cost is achieved for a distance of 83.32 m between the point B and the point C.

Assignment 10.7 — The length of the largest tube that can be carried horizontally around the corner is 7.02 m.

Assignment 10.8 — A minimal amount of cardboard is consumed if $B = 6.2$ cm and $H = 13$ cm.

Assignment 10.9 — The volume of the cone is maximal for $h = 4/3R$ and $r = 2\sqrt{2}/3R$.

Assignment 10.10 — Graph (c) is the graph of f, (d) is the graph of f', (b) is the graph of f'' and (a) is that of the function g.

Assignment 10.11 — Graph (c) is the graph of $f(x) = \dfrac{x}{1-x^2}$, graph (b) is the graph of $g(x) = \dfrac{x^3}{1-x^4}$, graph (d) is the graph of $h(x) = \dfrac{x^3-x}{\sqrt{1+x^6}}$ graph (a) is the graph of $k(x) = \dfrac{x^3}{\sqrt{|x^4-1|}}$.

Assignment 10.12 —

(a)
- local maxima/minima: none,
- inflection points: $x = 1/2$,
- concave over $]-\infty, 1/2[$, convex over $]1/2, +\infty[$.

(b)
- local maxima/minima: local max. in $(-1, 4)$, local min. in $(1, 0)$,
- inflection points: $x = 0$,
- concave over $]-\infty, 0[$, convex over $]0, +\infty[$.

(c)
- local maxima/minima: global max. in $(0, 1)$,
- inflection points: $x = \pm\sqrt{3}/3$,
- concave over $]-\sqrt{3}/3, \sqrt{3}/3[$, convex over $]-\infty, -\sqrt{3}/3[\cup]\sqrt{3}/3, +\infty[$.

(d)
- local maxima/minima: global max. in $(\pi/4, \sqrt{2})$, global min. in $(-3\pi/4, -\sqrt{2})$,
- inflection points: $x = -\pi/4, 3\pi/4$,
- concave over $]-\pi/4, 3\pi/4[$, convex over $]-\pi, -\pi/4[\cup]3\pi/4, \pi[$.

(e)
- local maxima/minima: global min. in $(e^{-1/2}, -e/2)$,
- inflection points: $x = 1/e^{3/2}$,
- concave over $]0, 1/e^{3/2}[$, convex over $]1/e^{3/2}, +\infty[$.

Chapter 11

Assignment 11.1 —

(a) $(3, 0)$

(b) $(-3, 0)$

(c) $\left(2, \dfrac{2\pi}{3}\right)$

(d) $\left(2, \dfrac{7\pi}{3}\right)$

(e) $(-3, \pi)$

(f) $\left(2, \dfrac{\pi}{3}\right)$

(g) $(-3, 2\pi)$

(h) $\left(-2, -\dfrac{\pi}{3}\right)$

(i) $\left(2, \dfrac{13\pi}{3}\right)$

Assignment 11.2 —

(a) $(1,1)$

(b) $(1,0)$

(c) $(0,0)$

(d) $(-1,-1)$

(e) $\left(-\sqrt{3},3\right)$

Assignment 11.3 —

(a) $y=x$ → line

(b) $y=\frac{2}{5}x+\frac{7}{5}$ → line

(c) $(x-1)^2+y^2=1$ → circle

(d) $x^2+(y+2)^2=4$ → circle

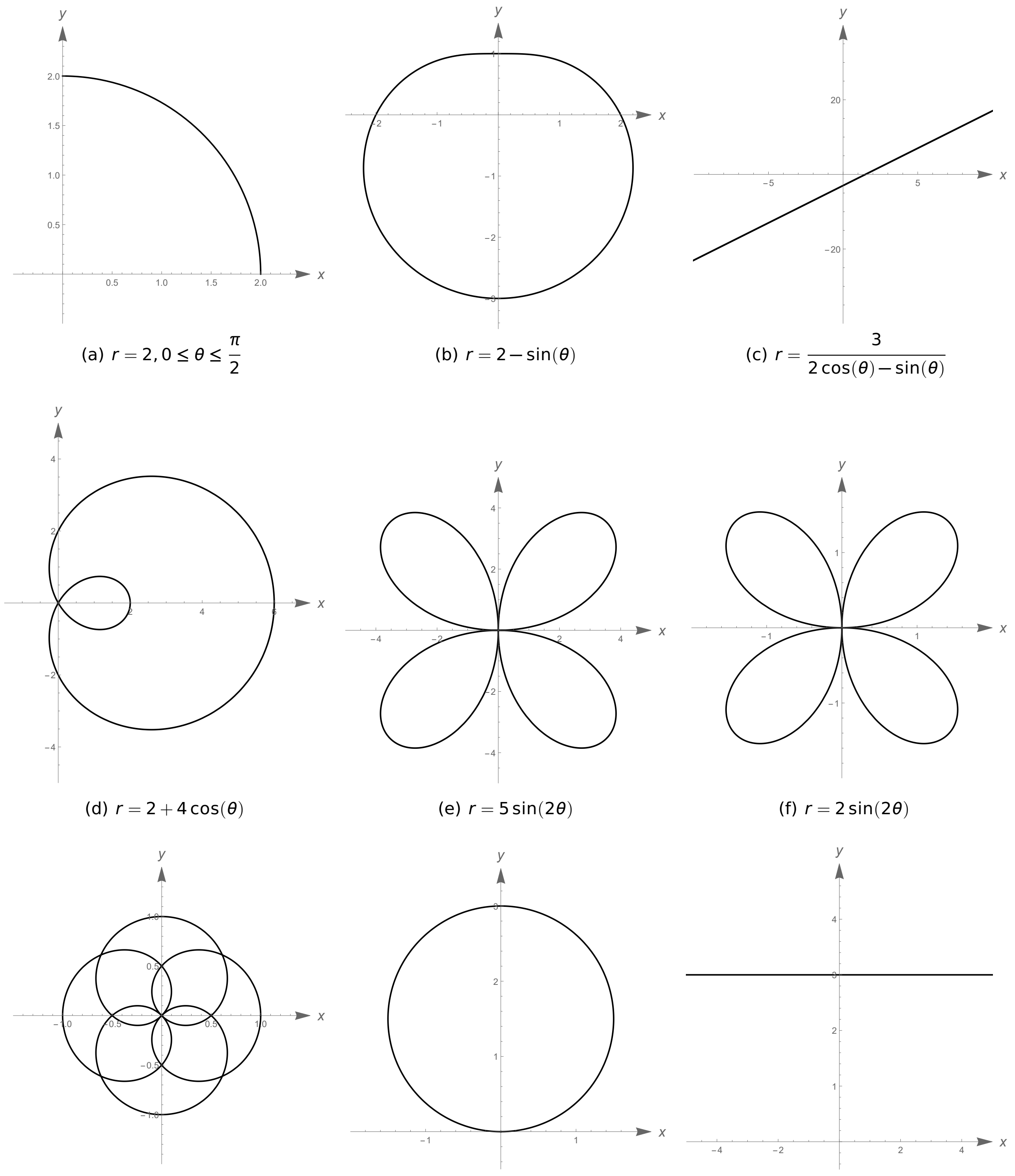
(e) $\left(x-\frac{1}{2}\right)^2+\left(y-\frac{1}{2}\right)^2=\frac{1}{2}$ → circle

(f) $x^2+4y^2=4$ → ellipse

(g) $y^2=1+2x$ → parabola

(h) $-3x^2+9\left(y+\frac{2}{3}\right)^2 = 1$ → hyperbola

Assignment 11.4 — Consider the graphs below.



(a) $r = 2, 0 \leq \theta \leq \dfrac{\pi}{2}$

(b) $r = 2 - \sin(\theta)$

(c) $r = \dfrac{3}{2\cos(\theta) - \sin(\theta)}$

(d) $r = 2 + 4\cos(\theta)$

(e) $r = 5\sin(2\theta)$

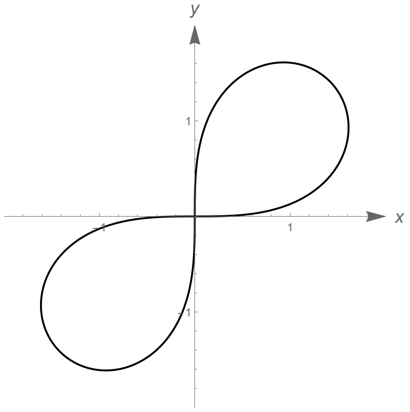
(f) $r = 2\sin(2\theta)$

(g) $r = \cos(2\theta/3), 0 \leq \theta \leq 6\pi$

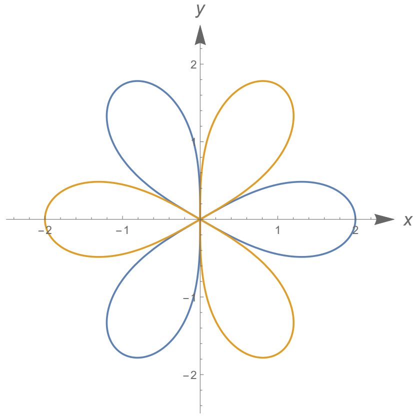
(h) $r = 3\sin(\theta), 0 \leq \theta \leq \pi$

(i) $r = 3\csc(\theta), 0 < \theta < \pi$

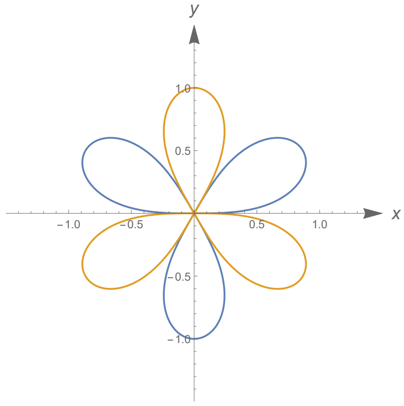
Figure 11.15: Graphs of the polar curves in Ex. 11.4 (deel 1).



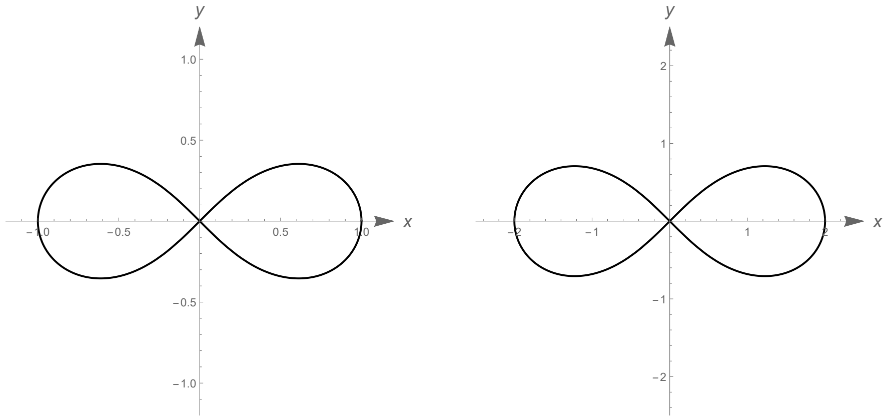
(j) $r^2 = 4\sin(2\theta)$



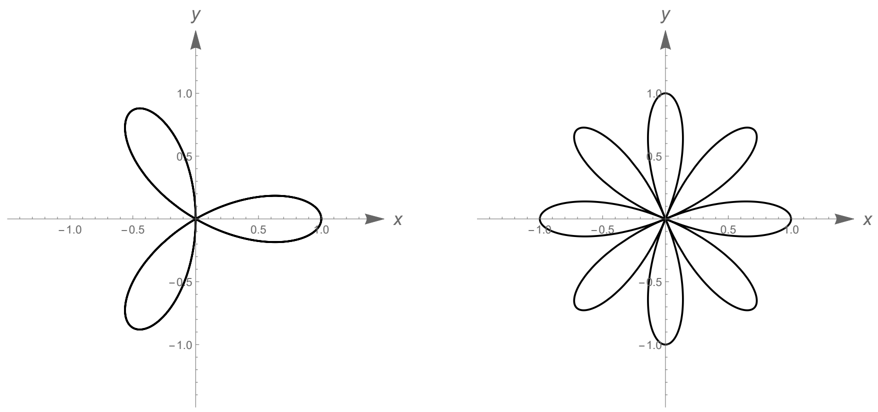
(k) $r^2 = 4\cos(3\theta)$



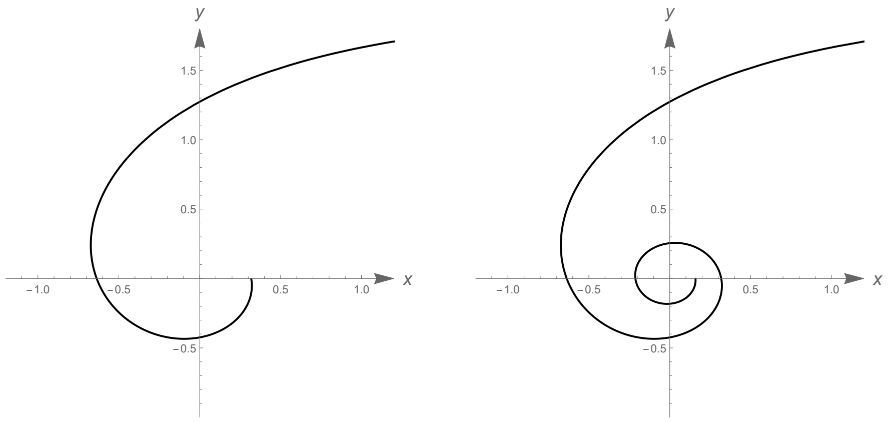
(l) $r^2 = \sin(3\theta)$



(m) $r = a\sqrt{\cos(2\theta)}$, $a > 0$. Links $a = 1$, right $a = 2$



(n) $r = a\cos(n\theta)$, $a > 0$. If n is odd, the curve has n 'leaves'. If n is even, the curve has $2n$ 'leaves'. The curve was plotted for $n = 3$ (left) and $n = 4$ (right).



(o) $r = a/\theta$. The curve was plotted for $a = 2$ with two different intervals for θ. Left: $0 \leq \theta \leq 2\pi$, right: $0 \leq \theta \leq 4\pi$.

Figure 11.16: Graphs of the polar curves in Exercise 11.4 (part 2).

Assignment 11.5 —

(a) The origin and $\left(\frac{3}{2}, \pm\frac{\pi}{3}\right)$

(b) $\left(4, \pm\frac{2\pi}{3}\right)$

(c) The origin, $\left(\frac{1}{2}, \frac{\pi}{6}\right)$ and $\left(\frac{1}{2}, \frac{5\pi}{6}\right)$

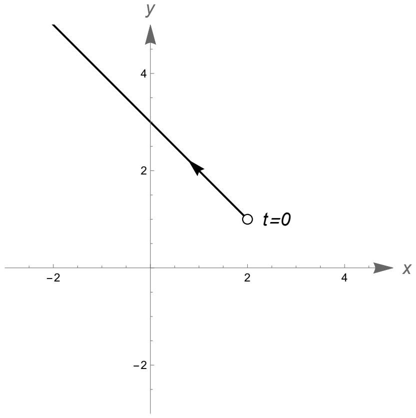
(d) The origin and $\left(\frac{\sqrt{3}}{2}, \frac{\pi}{3}\right)$

(e) $\left(1, \pm\frac{\pi}{6}\right)$ and $\left(1, \pm\frac{5\pi}{6}\right)$

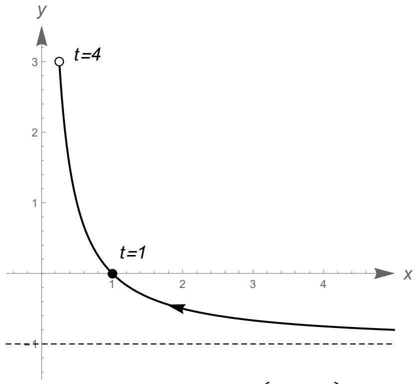
(f) The origin, $\left(\frac{\sqrt{2}}{2}, \frac{\pi}{12}\right)$, $\left(-\frac{\sqrt{2}}{2}, \frac{5\pi}{12}\right)$ and $\left(\frac{\sqrt{2}}{2}, \frac{3\pi}{4}\right)$

(g) The origin, $\left(1+\frac{\sqrt{2}}{2}, \frac{3\pi}{4}\right)$ and $\left(1-\frac{\sqrt{2}}{2}, \frac{7\pi}{4}\right)$

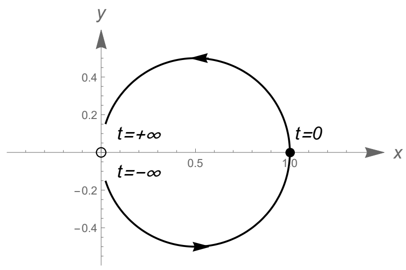
Assignment 11.6 — Consider the graphs below.



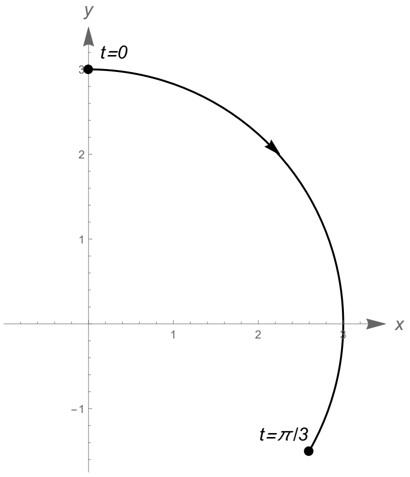
(a) $y = 3 - x \qquad (-\infty < x < 2)$



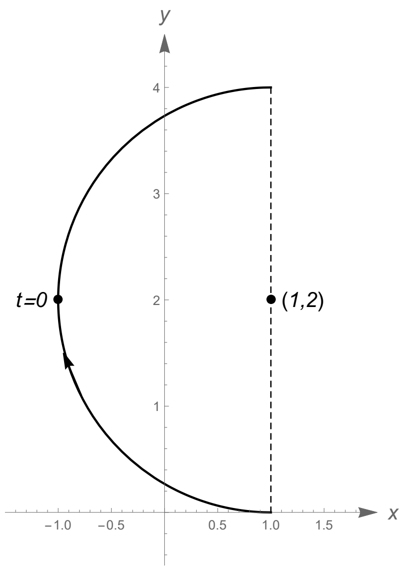
(b) $y = \dfrac{1}{x} - 1 \qquad \left(x > \dfrac{1}{4}\right)$



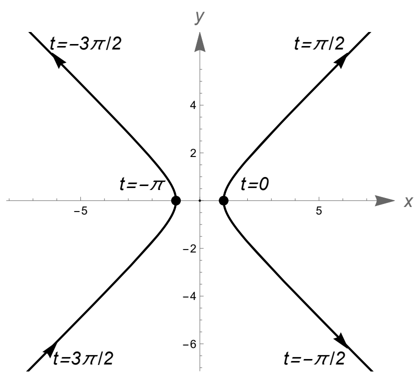
(c) $\left(x - \dfrac{1}{2}\right)^2 + y^2 = \dfrac{1}{4}$



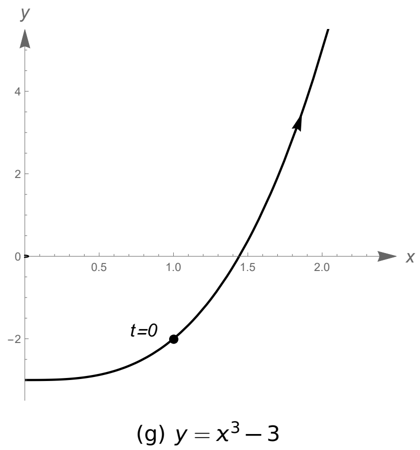
(d) $x^2 + y^2 = 9$ (part of the circle indicated on the figure)



(e) $(x-1)^2 + (y-2)^2 = 4 \qquad (x \le 1)$



(f) $x^2 - y^2 = 1$



(g) $y = x^3 - 3$

Figure 11.17: Graphs of the parametric curves in Exercise 11.6.

Assignment 11.7 —

(a) $\begin{cases} x = t \\ y = -\sqrt{t-1} \end{cases} \qquad (t > 1)$

(b) $\begin{cases} x = 6\cos^3(t) \\ y = 6\sin^3(t) \end{cases} \qquad (0 < t < 2\pi)$

Assignment 11.8 — $\vec{r}(t) = \left(\dfrac{t+4}{2}, t, \left(\dfrac{3}{2}t + 7\right)\right)$ with $0 \le t \le 2$

Assignment 11.9 —
parameter $t = x$: $\vec{r}(t) = \left(t, 1-t, 1-2t+2t^2\right)$ with $-\infty < t < +\infty$
parameter $t = y$: $\vec{r}(t) = \left(1-t, t, 1-2t+2t^2\right)$ with $-\infty < t < +\infty$
parameter $t = z$: We have to solve the system of equations $\begin{cases} x+y=1 \\ x^2+y^2=t \end{cases}$ for x and y. There are two possible solutions each corresponding to one half of the parabola starting at the lowest point $(1/2, 1/2, 1/2)$ because there are two points on the parabola at every $z > 1/2$. The entire parabola can not be parameterized by using z as a parameter.

Assignment 11.10 —

(a) $x^2 + y^2 = 9$ and $z = x + y$

A possible parameterization is $\vec{r}(t) = (3\cos(t), 3\sin(t), 3(\cos(t) + \sin(t)))$.

(b) $z = \sqrt{1 - x^2 - y^2}$ and $x + y = 1$

A possible parameterization is $\vec{r}(t) = \left(t, 1-t, \sqrt{2(t-t^2)}\right)$.

(c) $z = x^2 + y^2$ and $2x - 4y - z - 1 = 0$

A possible parameterization is $\vec{r}(t) = (1 + 2\cos(t), -2(1 - \sin(t)), 9 + 4\cos(t) - 8\sin(t))$.

Assignment 11.11 — Consider the graphs below.

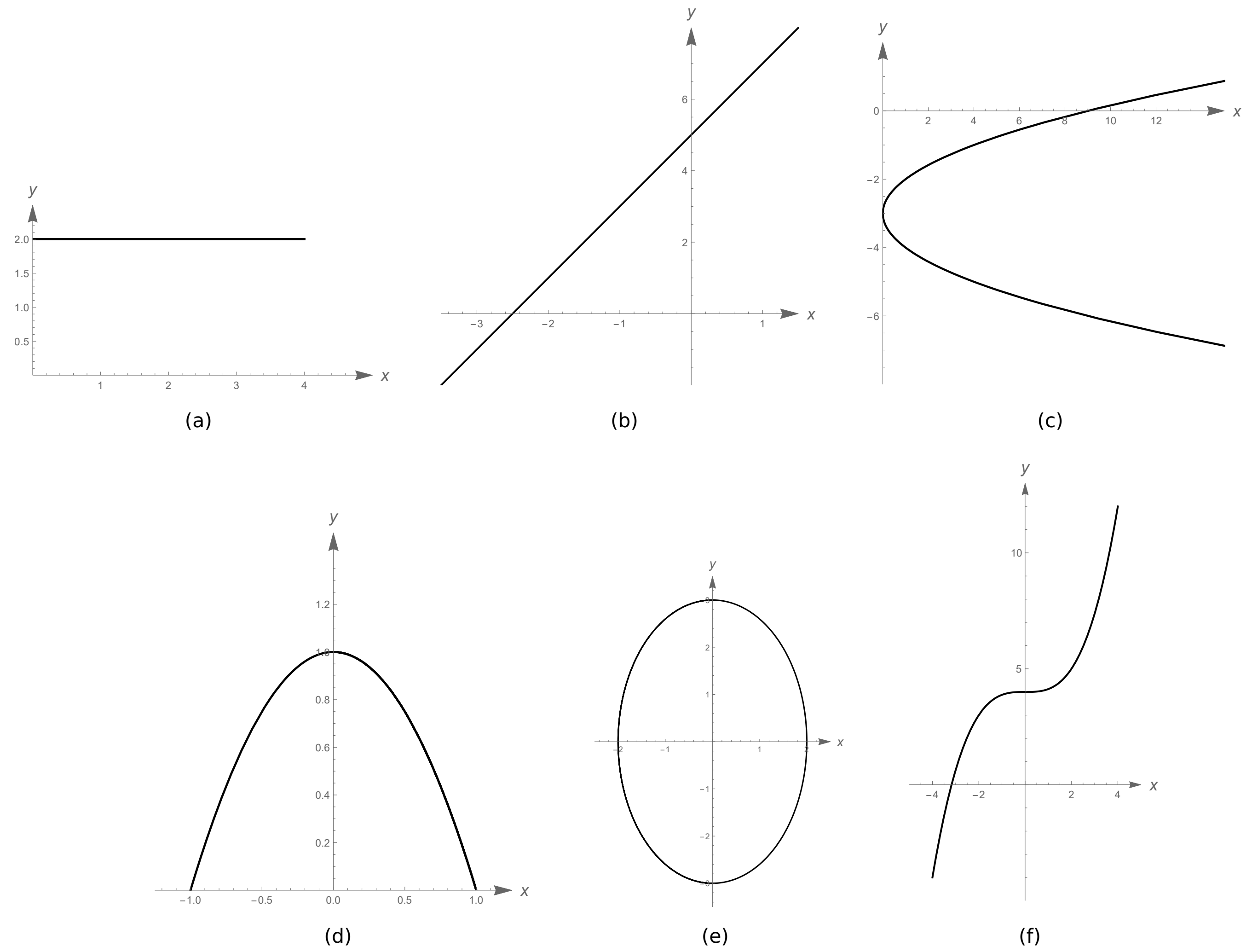


Figure 11.18: Grafieken van de parameterkrommen in Oefening 11.11.

Assignment 11.12 — Consider the figure below for the graphs of the parametric equations.

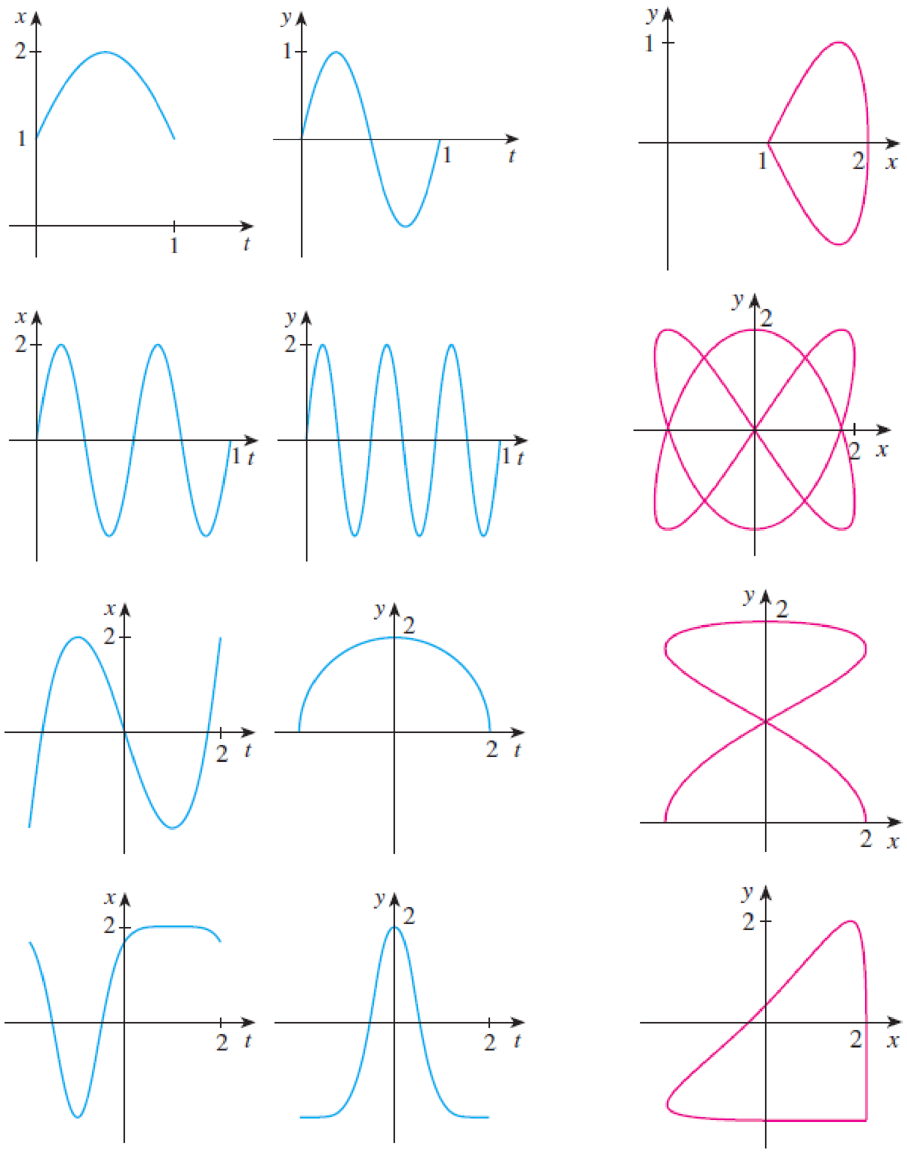


Figure 11.21: Graphs of the parametric equations from Exercise 11.12.

Assignment 11.13 —

(a) $y' = -\dfrac{1}{t^3}, \quad y'' = \dfrac{3}{2t^5}$

(b) $y' = 1 + \dfrac{2}{t-1}, \quad y'' = -\dfrac{1}{(t-1)^3}$

(c) $y' = -\dfrac{4}{3}\tan(t), \quad y'' = -\dfrac{4}{9\cos^3(t)}$

(d) $y' = \dfrac{4}{3\sin(t)}, \quad y'' = -\dfrac{4}{9\cot^3(t)}$

Assignment 11.14 —

(a) $y' = \dfrac{\sqrt{2}}{6+\sqrt{2}}$

(b) $y' = 5\sqrt{3}$

(c) $y' = -\dfrac{3}{4}$

(d) $y' = -\dfrac{3}{2}$

Assignment 11.15 —

(a) Tangent: $x = \dfrac{3\sqrt{3}}{4}$, normal: $y = \dfrac{3}{4}$

(b) Tangent: $y = x+1$, normal: $y = -x-1$

(c) Tangent: $y = 3x+2$, normal: $y = -\dfrac{1}{3}x+2$

(d) Tangent: $y = 1$, normal: $x = \dfrac{\sqrt{2}}{2}$

(e) Tangent: $y = -\dfrac{x}{10} + e^{\pi/20}$, normal: $y = 10x + e^{\pi/20}$

Assignment 11.16 —

(a) Horizontal tangent at $t = 0$, this is in $(0,0)$. Vertical tangent at $t = 1$, in $(-2,5)$.

(b) Horizontal tangent at $t = n\pi$, this is in $(0, -(-1)^n n\pi)(n \in \mathbb{Z})$. Vertical tangent at $t = \left(n + \dfrac{1}{2}\right)\pi$, this is in $(1,1)$ and $(-1,-1)$.

(c) Horizontal tangents at $t = 0$ and $t = 2^{1/3}$, this is in $(0,0)$ en $(2^{1/3}, 2^{2/3})$. Vertical tangent at $t = 2^{-1/3}$, this is in $(2^{2/3}, 2^{1/3})$. The curve approximates $(0,0)$ vertically if $t \to \pm\infty$.

(d) There are no horizontal tangents. Vertical tangent at $t = 0$, this is in $(1,0)$.

(e) Horizontal tangents at $t = k\pi$ and $t = \pm\arctan\left(\sqrt{2}\right) + k\pi$. Vertical tangents at $t = \dfrac{\pi}{2} + k\pi$ and $t = \pm\arctan\left(\dfrac{\sqrt{2}}{2}\right) + k\pi$.

(f) Horizontal tangent at $\left(\dfrac{3}{2}, \pm\dfrac{\pi}{3}\right)$. Vertical tangents at $(2,0)$ and $\left(\dfrac{1}{2}, \pm\dfrac{2\pi}{3}\right)$.

(g) Horizontal tangents at $\left(\dfrac{1}{\sqrt{2}}, \pm\dfrac{\pi}{6}\right)$ and $\left(\dfrac{1}{\sqrt{2}}, \pm\dfrac{5\pi}{6}\right)$. Vertical tangents in $(1,0)$ and $(1,\pi)$.

(h) Horizontal tangents at $\left(4, -\dfrac{\pi}{2}\right)$, $\left(1, \dfrac{\pi}{6}\right)$ and $\left(1, \dfrac{5\pi}{6}\right)$. Vertical tangents at $\left(3, -\dfrac{\pi}{6}\right)$ and $\left(3, -\dfrac{5\pi}{6}\right)$.

Assignment 11.17 —

(a) $y' = \dfrac{2\sin(t)}{1-2\cos(t)}$, $y'' = \dfrac{2\cos(t) - 4\cos^2(t) - 4\sin^2(t)}{(1-2\cos(t))^3}$

(b) Horizontal tangents in $t = k\pi$ $(k \in \mathbb{Z})$, this is in $(k\pi, -1)$ for even values of k and in $(k\pi, 3)$ for odd values of k. Vertical tangents in $t = \dfrac{\pi}{3} + 2k\pi$ $(k \in \mathbb{Z})$ and in $t = \dfrac{5\pi}{3} + 2k\pi$ $(k \in \mathbb{Z})$, this is in $\left(\dfrac{\pi}{3} + 2k\pi - \sqrt{3}, 0\right)$ and in $\left(\dfrac{5\pi}{3} + 2k\pi + \sqrt{3}, 0\right)$.

(c) $y' = 1 \quad \Leftrightarrow \quad t = -\dfrac{\pi}{4} + \arcsin\left(\dfrac{\sqrt{2}}{4}\right) + 2k\pi \; (k \in \mathbb{Z}) \;\vee\; t = \dfrac{3\pi}{4} - \arcsin\left(\dfrac{\sqrt{2}}{4}\right) + 2k\pi \; (k \in \mathbb{Z})$

$y' = -1 \quad \Leftrightarrow \quad t = -\dfrac{\pi}{4} + \arccos\left(\dfrac{\sqrt{2}}{4}\right) + 2k\pi \; (k \in \mathbb{Z}) \;\vee\; t = -\dfrac{\pi}{4} - \arccos\left(\dfrac{\sqrt{2}}{4}\right) + 2k\pi \; (k \in \mathbb{Z})$

Assignment 11.18 —

(a) $t=2$

(b) $t=0$

(c) $t=2k\pi, \quad k\in\mathbb{Z}$

(d) $t=0$

(e) $t=1$

(f) $t=k\pi \quad$ or $\quad t=\dfrac{\pi}{2}+k\pi, \quad k\in\mathbb{Z}$

Assignment 11.19 — See the graphs below.

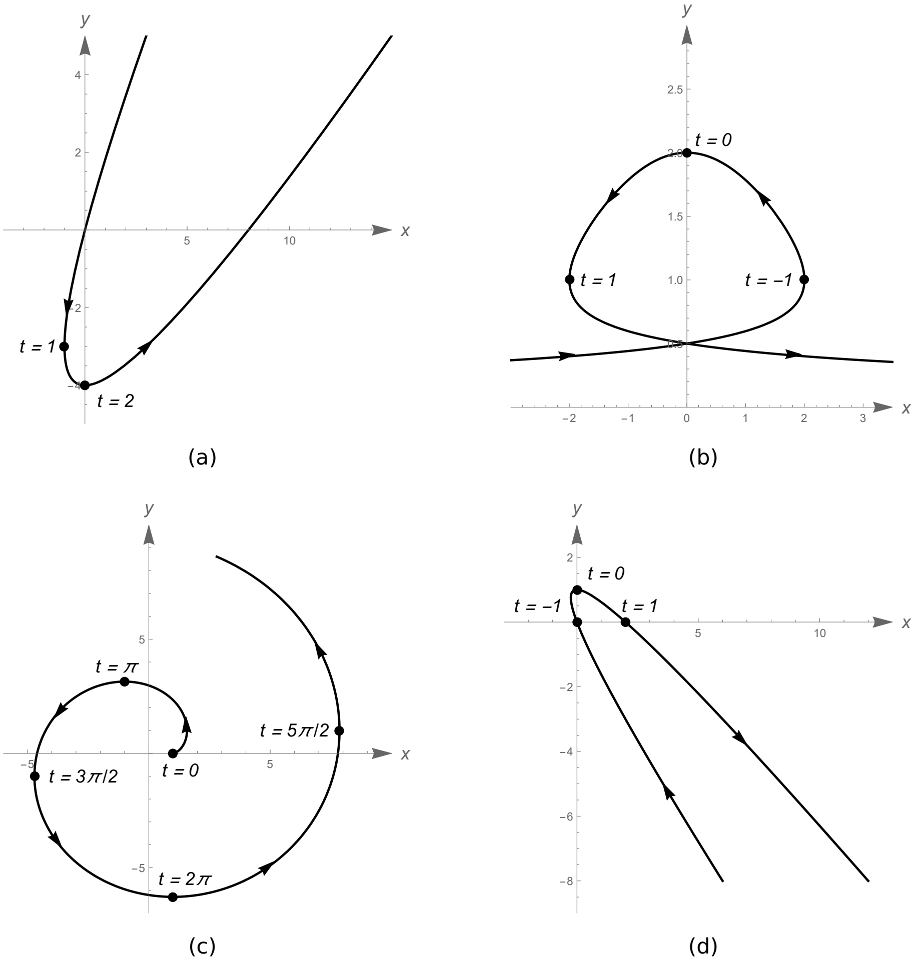


Figure 11.23: Graphs of the curves in Exercise 11.19.

Chapter 12

Assignments 12.1 — The area is 2.

Assignments 12.2 —

(a) $F'(x) = \dfrac{3x^2+1}{x^3+x}$

(b) $F'(x) = -3x^{11}$

(c) $F'(t) = \dfrac{\cos(t)}{1+t^2}$

(d) $F'(t) = -\dfrac{\sin(t)}{t}$

(e) $F'(x) = 2x \displaystyle\int_0^{x^2} \frac{\sin(u)}{u}\, du + 2x\sin(x^2)$

(f) $F'(\theta) = -\dfrac{1}{\sin(\theta)} - \dfrac{1}{\cos(\theta)}$

(g) $F'(x) = 3\displaystyle\int_4^{x^2} e^{-\sqrt{t}}\, dt + 6x^2 e^{-|x|}$

(h) $F'(x) = 2x^3+3x-2$

(i) $F'(x) = e^x \sin(e^x) - \dfrac{\sin\big(\ln(x)\big)}{x}$

Assignments 12.3 —

(a) $I = \dfrac{2}{3}x^{3/2} - \dfrac{2}{5}x^{5/2} + C$

Can be integrated directly after rewriting $\sqrt{x}$ as $x^{1/2}$ and splitting the integral into the difference of two integrals.

(b) $I = x^4 - \dfrac{7}{2}x^2 + 5x + C$

(c) $I = 3\dfrac{x^4}{4} - 2\cos(x) + C$

(d) $I = \dfrac{1}{4}\left(2x^3+3x-1\right)^{4/3} + C$

Perform the substitution $u = 2x^3+3x-1$.

(e) $I = -\dfrac{1}{8}\cos(4x) + \dfrac{1}{4}\cos(2x) + C$

The integrand can be rewritten as $\sin(x)\cos(3x) = \dfrac{1}{2}(\sin(4x) - \sin(2x))$.

(f) $I = \dfrac{1}{4}\sin(2x) - \dfrac{1}{20}\sin(10x) + C$

The integrand can be rewritten as $\sin(6x)\sin(4x) = \dfrac{1}{2}(\cos(2x) - \cos(10x))$.

(g) $I = x^2\ln(x+1) - \dfrac{1}{2}x^2 + x - \ln|x+1| + C$

Perform integration by parts. You obtain an integral of a rational function. Perform Euclidean division and obtain two basic integrals.

(h) $I = -\ln|x| - \dfrac{1}{x} + \ln|x+1| + C$

Split the integrand into partial fractions to obtain three basic integrals.

(i) $I = \arcsin\left(\dfrac{x-1}{2}\right) + C$

Write the argument of the square root as a perfect square and then set $x-1$ equal to t.

(j) $I = -\ln|x^2-3x-2| - \dfrac{4}{\sqrt{17}}\ln\left|\dfrac{2x-3-\sqrt{17}}{2x-3+\sqrt{17}}\right| + C$

Rewrite the numerator as the derivative of the denominator. After this you obtain a basic integral and an integral of a rational function. Rewrite the denominator as a perfect square and again obtain a basic integral.

(k) $I = \ln(1 - \cos(x)) + C$

Perform the substitution $u = 1 - \cos(x)$.

(l) $I = \dfrac{1}{2}\arctan\left(\dfrac{x}{2}\right)$

Perform the substitution $x = 2\tan(t)$ or rewrite the denominator as $4(1 + (x/2)^2)$.

(m) $I = -x\cos(x) + \sin(x) + C$

TUse integration by parts with $u(x) = x$ and $dv(x) = \sin(x)\,dx$.

(n) $I = \dfrac{1}{2}(x+1)^2\sin(2x) + \dfrac{1}{2}(x+1)\cos(2x) - \dfrac{1}{4}\sin(2x) + C$

Use integration by parts twice, with respectively $u(x) = (x+1)^2$ and $u(x) = x + 1$ and with respectively $dv(x) = \cos(2x)\,dx$ and $dv(x) = \sin(2x)\,dx$.

(o) $I = -\dfrac{1}{5}e^{-x}\sin(2x) - \dfrac{2}{5}e^{-x}\cos(2x) + C$

Use integration by parts twice with $u(x) = e^(-x)$ and form the equality to determine the integral.

(p) $I = \dfrac{x^{n+1}\ln(x)}{n+1} - \dfrac{x^{n+1}}{(n+1)^2} + C$

Perform integration by parts with $u(x) = \ln(x)$ and $dv(x) = x^n\,dx$.

(q) $I = \dfrac{1}{2}\ln\left|x^2 + 4x + 5\right| - 2\arctan(x+2) + C$

Rewrite the statement as follows:

$$\int \frac{x}{x^2 + 4x + 5}\,dx = \int \frac{x+2}{x^2 + 4x + 5}\,dx - 2\int \frac{1}{x^2 + 4x + 5}\,dx.$$

For the first integral, use the substitution $u = x^2 + 4x + 5$ and rewrite the denominator as $x^2 + 4x + 5 = (x+2)^2 + 1$ in the second integral.

Assignments 12.4 —

(a) $I = e^x - \arcsin(x) + C$

Split the integral to obtain two basic integrals.

(b) $I = \dfrac{1}{4}\ln\left(4x^2 + 4x + 3\right) + C$

Substitute $4x^2 + 4x + 3$ by t.

(c) $I = \dfrac{1}{5\cos^5(x)} + C$

Substitute $\cos(x)$ by t.

(d) $I = \sin(x) - \dfrac{2}{3}\sin^3(x) + \dfrac{1}{5}\sin^5(x) + C$

Rewrite the integrand as $\left(\cos^2(x)\right)^2\cos(x)$ and then use the Pythagorean identity $\cos^2(x) = 1 - \sin^2(x)$. Assume $t = \sin(x)$, expand the product and split the integral.

(e) $I = -\ln|\sin(x)+\cos(x)| + C$

Use $t = \sin(x)+\cos(x)$.

(f) $I = \frac{1}{2}\arcsin\big(2\tan(x)\big) + C$

Use $t = 2\tan(x)$.

(g) $I = -\frac{1}{1+\tan(x)} + C$

Rewrite the denominator by using $(\cos(x)+\sin(x))^2 = \cos^2(x)\big(1+\tan(x)\big)^2$ and then use $t = 1+\tan(x)$.

(h) $I = x\ln\left(x+\sqrt{x^2+5}\right) - \sqrt{x^2+5} + C$

Use integration by parts with $u(x) = \ln\left(x+\sqrt{x^2+5}\right)$ and $dv(x) = dx$. Then, use $t^2 = x^2+5$.

(i) $I = \frac{3}{5}\ln|x-2| + \frac{7}{5}\ln|x+3| + C$

Split the integrand into partial fractions.

(j) $I = 4\ln|x-2| - \frac{1}{x-2} - 4\ln|x-3| - \frac{4}{x-3} + C$

There are two ways to start. Either you first take the square of the integrand (square of the numerator over square of the denominator) and split the expression into partial fractions, or you first split the $\frac{x-1}{x^2-5x+6}$ into partial fractions and then square the result.

(k) $I = x - \ln(x^2+2x+2) + \arctan(x+1) + C$

First, perform Euclidean division. Rewrite the numerator of the integral as $2x+2-1$ and split it into two integrals. One integral can be found using $x^2+2x+2 = t$. The denominator of the second integral should be written as a perfect square to obtain $\frac{1}{1+u^2}$ as integrand.

(l) $I = \frac{x-1}{\sqrt{1+x^2}} + C$

The denominator contains the square root of the sum of two squares, therefore use $x = \tan(t)$.

(m) $I = \arcsin\left(\frac{x-2}{2}\right) + C$

Write the denominator as the square root of the difference of two squares by completing to a perfect square. Then work toward the integral with integrand $\frac{1}{\sqrt{1-u^2}}$ or use $x-2 = 2\sin(t)$.

(n) $I = -\frac{e^{2x}\cos(4x)}{5} + \frac{e^{2x}\sin(4x)}{10} + C$

By applying integration by parts twice with $u(x) = e^{2x}$ (choice) you arrive at the original integral I. Solve the resulting integral equation for I.

(o) $I = \frac{1}{16}\left(x - \frac{\sin(4x)}{4} - \frac{\sin^3(2x)}{3}\right) + C$

Use the angle doubling formula for $\cos(2x)$:

$$\sin^2(x) = \frac{1-\cos(2x)}{2} \qquad \text{en} \qquad \cos^2(x) = \frac{1+\cos(2x)}{2}.$$

Then expand the powers in the integrand. For integrands with an even power of $\cos(x)$ the doubling formula can be used again for $\cos(2x)$. For integrands with odd powers of $\cos(x)$ you can use the Pythagorean identity.
Remark: If you use a different trigonometric formula, you will arrive at a different formula for I.

(p) $I = \frac{1}{3}\ln\left|\frac{2\sin(x)+1}{\sin(x)-1}\right| + C$

Rewrite $\cos^2(x)$ by using the Pythagorean identity $1-\sin^2(x)$ and perform the substitution $\sin(x) = t$. Afterwards, split the integrand in partial fractions.

(q) $I = \frac{1}{3}\tan^3(x) - \cot(x) + 2\tan(x) + C$

Write $\sin(x)$ as $\tan(x)\cos(x)$ and perform the substitution $\tan(x) = u$ together with the trigonometric formula $1+\tan^2(x) = \sec^2(x)$.

Assignments 12.5 —

(a) $I = 3x - 7\arctan(x) + C$

Rewrite the numerator as $3\left(x^2+1\right)-7$ and split the integral into two integrals or perform Euclidean division. We then obtain two basic integrals.

(b) $I = \frac{x^2}{2} + \frac{4}{3}\ln|x-2| - \frac{2}{3}\ln|x^2+2x+4| + \frac{4\sqrt{3}}{3}\arctan\left(\frac{x+1}{\sqrt{3}}\right) + C$

Perform a Euclidean division and split $\frac{8x}{x^3-8}$ in partial fractions. Then, the integral $\int \frac{x-2}{x^2+2x+4}\,d$ has to be found. Rewrite the numerator as $x+1-3$ and split the integral into two integrals.
To find integral $\int \frac{1}{x^2+2x+4}\,dx$ rewrite the denominator as a perfect square.

(c) $I = \frac{\sqrt{x^2-1}\left(2x^2+1\right)}{3x^3} + C$

let $x = \sec(t)$ and then rewrite $\frac{1}{\cos^2(t)} - 1$ as $\tan^2(t)$. Next, write $\cos^3(t)$ as $(1-\sin^2(t))\cos(t)$ and perform the substitution $u = \sin(t)$.

(d) $I = -x - 2\sqrt{x} + \frac{1}{1-2\sqrt{x}} + C$

Let $t^2 = x$ and then perform a Euclidean division.

(e) $I = -\cot(x) + \ln\left|1+\cot(x)\right| + C$

Let $t = \cot(x)$.

(f) $I = \frac{x\,e^{2x}}{2} - \frac{e^{2x}}{4} + C$

Use integration by parts with $u(x) = x$ and $dv(x) = e^{2x}\,dx$.

(g) $I = \frac{5}{2}\ln\left|x^2 + \sqrt{x^4+1}\right| + C$

Let $x^2 = \tan(t)$ or let $x^2 = t$.

(h) $I = \frac{\sin(2x)}{4} + C$

Use Simpsons formula $2\sin(\alpha)\sin(\beta) = \cos(\alpha-\beta) - \cos(\alpha+\beta)$.

(i) $I = -2\sqrt{5-x} + 4\sqrt[4]{5-x} - 4\ln|\sqrt[4]{5-x}+1| + C$

Let $t^4 = 5-x$ and then perform a Euclidean division.

Assignments 12.6 —

(a) $I = \dfrac{x^3\ln\left(\sqrt{1-x}\right)}{3} - \dfrac{x^3}{18} - \dfrac{x^2}{12} - \dfrac{x}{6} - \dfrac{\ln|1-x|}{6} + C$

Use integration by parts with $u(x) = \ln(\sqrt{1-x})$ and $dv(x) = x^2\,dx$ and then perform a Euclidean division.

(b) $I = x - 2\ln|2x+3| + C$

Rewrite the numerator as $2x+3-4$ ans split the integral in two basic integrals.

(c) $I = \dfrac{\sin^2(2x)}{4} + C$

Use $t = \sin(2x)$.

(d) $I = -\ln(1+e^x) + x + C$

Use $t = e^x + 1$ and split the integrand in partial fractions.

(e) $I = \dfrac{2\sqrt{3}}{3}\arctan\left(\dfrac{2x+1}{\sqrt{3}}\right) + C$

Rewrite the denominator as $1+t^2$ with $t = \dfrac{2x+1}{\sqrt{3}}$.

(f) $I = -\dfrac{1}{x^2+x+1} + \dfrac{2}{3}\dfrac{2x+1}{x^2+x+1} + \dfrac{8\sqrt{3}}{9}\arctan\left(\dfrac{2x+1}{\sqrt{3}}\right) + C$

Rewrite the numerator as $2x+1+2$ and split the integrals in two integrals. The first integral $\displaystyle\int \frac{2x+1}{(x^2+x+1)^2}\,dx$ can be found using $t = x^2+x+1$. In the second integral $\displaystyle 2\int \frac{dx}{(x^2+x+1)^2}$ the denominator should be rewritten as a perfect square. Then let $u = \dfrac{2x+1}{\sqrt{3}}$. This integral can be calculated using $u = \tan(\theta)$.

(g) $I = 2a\arctan\left(\sqrt{\dfrac{a+x}{a-x}}\right) - (a-x)\sqrt{\dfrac{a+x}{a-x}} + C$

Let $t^2 = \dfrac{a+x}{a-x}$. Rewrite the obtained numerator as t^2+1-1 and split the integral in two integrals, then use $t = \tan(\theta)$.

(h) $I = \dfrac{4}{7}(x-1)^{7/4} - \dfrac{2}{3}(x-1)^{3/2} - \dfrac{4}{5}(x-1)^{5/4} + x - 1 + C$

Use $t^4 = x-1$.

(i) $I = x\arctan(\sqrt{x}) - \sqrt{x} + \arctan(\sqrt{x}) + C$

Use integration by parts with $u(x) = \arctan(\sqrt{x})$ and $dv(x) = dx$. Then, let $t^2 = x$ and rewrite the numerator of the integrand as t^2+1-1.

(j) $I = \dfrac{2}{15}(1+x^3)^{5/2} - \dfrac{2}{9}(1+x^3)^{3/2} + C$

Let $1+x^3 = t^2$ and rewrite x^5 as $x^3.x^2$.

(k) $I = \dfrac{\tan^4(x)}{4} + \dfrac{3}{2}\tan^2(x) + 3\ln|\tan(x)| - \dfrac{1}{2\tan^2(x)} + C$

Use $t = \tan(x)$ and write $\sin(x)$ and $\cos(x)$ as a function of t.

(l) $I = -\dfrac{1}{5\tan^5(x)} - \dfrac{2}{3\tan^3(x)} - \dfrac{1}{\tan(x)} + C = -\dfrac{\cot^5(x)}{5} - \dfrac{2\cot^3(x)}{3} - \cot(x) + C$

Use $t = \tan(x)$ or $t = \cot(x)$.

(m) $I = \ln\left(\sqrt{1+e^x} - 1\right) - \ln\left(\sqrt{1+e^x} + 1\right) + C$

Let $t^2 = 1 + e^x$.

(n) $I = \dfrac{1}{32}(12x - 8\sin(2x) + \sin(4x)) + C$

Use the formula $\sin^2(x) = \dfrac{1-\cos(2x)}{2}$, then $\cos^2(2x) = \dfrac{1+\cos(4x)}{2}$.

(o) $I = \ln\left|1 + \tan\left(\frac{x}{2}\right)\right| + C$

Rewrite the denominator first by using $\cos(x) = 1 - 2\sin^2\left(\frac{x}{2}\right)$ en $\sin(x) = 2\sin\left(\frac{x}{2}\right)\cos\left(\frac{x}{2}\right)$. Then, multiply both numerator and denominator with $\sec^2\left(\frac{x}{2}\right)$ and perform the substitution $u = \tan\left(\frac{x}{2}\right) + 1$.

Assignments 12.7 —

(a) $I = +\infty \quad \Rightarrow \quad I$ is divergent.

(b) $I = \dfrac{1}{2} \quad \Rightarrow \quad I$ is convergent.

(c) $I = \dfrac{\pi}{2} \quad \Rightarrow \quad I$ is convergent.

(d) $I = +\infty \quad \Rightarrow \quad I$ is divergent.

(e) $I = 9 \quad \Rightarrow \quad I$ is convergent.

(f) $I = 2\left(1+\sqrt{2}\right) \quad \Rightarrow \quad I$ is convergent.

(g) $I = +\infty \quad \Rightarrow \quad I$ is divergent.

(h) $I = +\infty \quad \Rightarrow \quad I$ is divergent.

(i) $I = 2e^2 \quad \Rightarrow \quad I$ is convergent.

Assignments 12.8 — $\displaystyle\int_{\omega_0}^{+\infty} g(\omega)\,d\omega = \lim_{b\to+\infty}\int_{\omega_0}^{b} g(\omega)\,d\omega = \lim_{b\to+\infty} \frac{1}{\pi}\arctan[T(\omega-\omega_0)]\Big|_{\omega_0}^{b} = \frac{1}{2}$

Assignments 12.9 —

(a) $I = \dfrac{393}{10}$

(b) $I = 0$

(c) $I = \dfrac{\pi}{4}$

(d) $I = -\dfrac{4}{\pi^2}$

(e) $I = -1 + 4\ln\dfrac{3}{4}$

(f) $I = \dfrac{\pi}{3} - \sqrt{3}$

Assignments 12.10 —

(a) $$\int_0^{\sqrt{\pi}} \sin\left(x^2\right)\,dx = \int_0^{\sqrt{\pi}} \left(x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \ldots\right)\,dx = \frac{\pi^{3/2}}{3} - \frac{\pi^{7/2}}{7.3!} + \frac{\pi^{11/2}}{11.5!} - \frac{\pi^{15/2}}{15.7!} + \ldots$$

(b) $$\int_0^{\pi^2/4} \cos\left(\sqrt{x}\right)\,dx = \int_0^{\pi^2/4} \left(1 - \frac{x}{2!} + \frac{x^2}{4!} - \frac{x^3}{6!} + \ldots\right)\,dx = \frac{\pi^2}{4} - \frac{\pi^4}{4^2.2.2!} + \frac{\pi^6}{4^3.3.4!} - \frac{\pi^8}{4^4.4.6!} + \ldots$$

Chapter 13

Assignment 13.1 —

(a) $$A = \int_{-2}^{4} \left(\frac{y}{2} + 2 - \frac{y^2}{4}\right) dy = 9 \quad \text{or} \quad A = 4\int_0^1 \sqrt{x}\,dx + \int_1^4 \left(2\sqrt{x} - 2x + 4\right) dx = 9$$

(b) $$A = 2\int_0^2 (4 - y^2)\,dy = \frac{32}{3} \quad \text{or} \quad A = 2\int_0^4 \sqrt{4-x}\,dx = \frac{32}{3}$$

(c) $$A = 2\int_3^5 \sqrt{25 - x^2}\,dx = -12 + \frac{25\pi}{2} - 25\arcsin\left(\frac{3}{5}\right) = 25\arccos\left(\frac{3}{5}\right) - 12$$

or $$A = 2\int_0^4 \left(\sqrt{25 - y^2} - 3\right) dy = 25\arccos\left(\frac{3}{5}\right) - 12$$

(d) $$A = \int_0^1 \left((4-x) - \left(4x - x^2\right)\right) dx + \int_1^4 \left(\left(4x - x^2\right) - (4-x)\right) dx = \frac{19}{3}$$

This exercise is not suited to use y as variable of integration.

(e) $$A = \int_0^4 \left(6x - x^2 - \left(x^2 - 2x\right)\right) dx = \int_0^4 \left(8x - 2x^2\right) dx = \frac{64}{3}$$

This exercise is not suited to use y as variable of integration.

(f) $$A = 2\int_0^3 \sqrt{x}\,dx + 2\int_3^{2\sqrt{3}} \sqrt{12 - x^2}\,dx = 2\left(\frac{\sqrt{3}}{2} + \pi\right)$$

or $$A = 2\int_0^{\sqrt{3}} \left(\sqrt{12 - y^2} - y^2\right) dy = 2\left(\frac{\sqrt{3}}{2} + \pi\right)$$

(g) $$A = 2\int_{\pi/2}^{3\pi/2} \cos^2(x)\,dx = \pi \quad \text{of} \quad A = 4\int_0^1 \arccos\left(\sqrt{y}\right)\,dy = \pi$$

(h) $A = \int\limits_1^e \big(\ln(2x) - \ln(x)\big)\, dx = (e-1)\ln(2)$

or $\quad A = \int\limits_0^{\ln(2)} \left(e^y - 1\right)\, dy + \int\limits_{\ln(2)}^{1} \left(e^y - \frac{e^y}{2}\right)\, dy + \int\limits_1^{\ln(2e)} \left(e - \frac{e^y}{2}\right)\, dy = (e-1)\ln(2)$

(i) $A = \int\limits_{-\pi/2-1}^{-\pi/2} \left(x + \frac{\pi}{2} + 1\right) dx + \int\limits_{-\pi/2}^{0} \left(-\sin(x)\right) dx = \frac{3}{2}$

or $\quad A = \int\limits_{-1}^{0} \left(1 + y + \frac{\pi}{2} + \arcsin(y)\right) dy = \frac{3}{2}$

Assignment 13.2 —

(a) $A = \frac{1}{2}\int_0^{\alpha} R^2\, d\theta = \frac{\alpha R^2}{2}$

(b) $A = 2\int_0^{2a} \frac{4a^2 - y^2}{4a}\, dy = 2\int_0^{a} \sqrt{4a(a-x)}\, dx = \frac{8}{3}a^2$

(c) $A = \int\limits_0^1 \left(\sqrt{2}x - \frac{x^2}{2}\right) dx + \int\limits_1^2 \left(\sqrt{2x} - \frac{x^2}{2}\right) dx = \frac{4}{3} - \frac{\sqrt{2}}{6}$

or $\quad A = \int\limits_0^{\sqrt{2}} \left(\sqrt{2y} - \frac{y}{\sqrt{2}}\right) dy + \int\limits_{\sqrt{2}}^{2} \left(\sqrt{2y} - \frac{y^2}{2}\right) dy = \frac{4}{3} - \frac{\sqrt{2}}{6}$

(d) $A = 2\int\limits_0^3 \left(t - \frac{t^3}{9}\right) 2t\, dt = 4\int\limits_0^3 \left(t^2 - \frac{t^4}{9}\right) dt = \frac{72}{5}$

(e) $A = \int\limits_0^{\pi} \big(a(1+\cos(\theta))\big)^2\, d\theta = a^2 \int\limits_0^{\pi} \left(1 + 2\cos(\theta) + \cos^2(\theta)\right) d\theta = \frac{3\pi a^2}{2}$

(f) $A = \int\limits_0^{\pi/4} \big(4\cos(2\theta)\big)^2\, d\theta = 16 \int\limits_0^{\pi/4} \cos^2(2\theta)\, d\theta = 2\pi$

(g) $A = 4\int\limits_{\pi/2}^{0} c\sin^3(t)\, 3c\cos^2(t) \left(-\sin(t)\right) dt = 12c^2 \int\limits_0^{\pi/2} \sin^4(t)\cos^2(t)\, dt = \frac{3c^2\pi}{8}$

(h) $A = 4\int\limits_{\pi/2}^{0} |4\sin(\theta)| \left(-\sin(\theta)\right) d\theta = 16 \int\limits_0^{\pi/2} \sin^2(\theta)\, d\theta = 4\pi$

(i) $A = 4\int\limits_{\pi/2}^{0} |3\sin(2t)|\, 2\left(-\sin(t)\right) dt = 48 \int\limits_0^{\pi/2} \sin^2(t)\cos(t)\, dt = 16$

(j) $A = \int_0^{\pi/2} \left(2\left(\cos(\theta)+1\right)\right)^2 \, d\theta - \int_0^{\pi/2} 4 \, d\theta = \pi + 8$

(k) $A = \int_0^{\pi/2} 4 \, d\theta + \int_{\pi/2}^{\pi} \left(2\left(\cos(\theta)-1\right)\right)^2 \, d\theta = 5\pi + 8$

(l) $A = \frac{3}{2} \int_{\pi/18}^{5\pi/18} \left(\left(2a\sin(3\theta)\right)^2 - a^2\right) d\theta = 3a^2 \int_{\pi/18}^{5\pi/18} \left(4\sin^2(3\theta) - 1\right) d\theta = \frac{(3\sqrt{3}+2\pi)a^2}{6}$

(m) $A = \frac{1}{2} \int_{-\pi/2}^{\pi/6} \left(\frac{1+\sin(\theta)}{\sqrt{3}}\right)^2 d\theta + \frac{1}{2} \int_{\pi/6}^{\pi/2} \cos^2(\theta) \, d\theta$

$= \frac{1}{6} \int_{-\pi/2}^{\pi/6} \left(1 + 2\sin(\theta) + \sin^2(\theta)\right) d\theta + \frac{1}{2} \int_{\pi/6}^{\pi/2} \cos^2(\theta) \, d\theta = \frac{\pi - \sqrt{3}}{4}$

(n) $A = \frac{1}{2} \int_0^{2\pi} \left(3e^{2\theta}\right)^2 d\theta = \frac{9}{2} \int_0^{2\pi} e^{4\theta} \, d\theta = \frac{9}{8}\left(e^{8\pi} - 1\right)$

Assignment 13.3 —

(a) disk method: $V = \pi \int_0^2 8x \, dx = 16\pi$ (most efficient)

shell method: $V = 2\pi \int_0^4 \left(2 - \frac{y^2}{8}\right) y \, dy = 16\pi$

(b) disk method: $V = 2\pi \int_0^4 2^2 \, dy - 2\pi \int_0^4 \left(\frac{y^2}{8}\right)^2 dy = 32\pi - \frac{\pi}{32} \int_0^4 y^4 \, dy = \frac{128\pi}{5}$

shell method: $V = 4\pi \int_0^2 x\sqrt{8x} \, dx = 8\sqrt{2}\pi \int_0^2 x\sqrt{x} \, dx = \frac{128\pi}{5}$ (most efficient)

(c) disk method: $V = \pi \int_0^1 \left(\left(\sqrt{x}\right)^2 - \left(x^2\right)^2\right) dx = \pi \int_0^1 \left(x - x^4\right) dx = \frac{3\pi}{10}$

shell method: $V = 2\pi \int_0^1 y\left(\sqrt{y} - y^2\right) dy = \frac{3\pi}{10}$ (most efficient)

(d) disk method: $V = 2\pi \int_0^4 \left(2 - \frac{y^2}{8}\right)^2 dy = 2\pi \int_0^4 \left(\frac{y^4}{64} - \frac{y^2}{2} + 4\right) dy = \frac{256\pi}{15}$

shell method: $V = 4\pi \int_0^2 (2-x)\sqrt{8x} \, dx = 8\sqrt{2}\pi \int_0^2 (2-x)\sqrt{x} \, dx = \frac{256\pi}{15}$ (most efficient)

(e) a) disk method: $V = \pi \int_0^3 \left(x^2 - \left(2 - \sqrt{-x+4}\right)^2\right) dx + \pi \int_3^4 \left(\left(2 + \sqrt{-x+4}\right)^2 - \left(2 - \sqrt{-x+4}\right)^2\right) dx$

shell method: $V = 2\pi \int_0^3 y(4y - y^2 - y)\, dy = 2\pi \int_0^3 y(3y - y^2)\, dy = \frac{27\pi}{2}$ (most efficient)

b) disk method: $V = \pi \int_0^3 \left(\left(4y - y^2\right)^2 - y^2\right) dy = \pi \int_0^3 \left(y^4 - 8y^3 + 15y^2\right) dy = \frac{108\pi}{5}$

(most efficient)

shell method: $V = 2\pi \int_0^3 x\left(x - 2 + \sqrt{-x+4}\right) dx + 2\pi \int_3^4 x\left(2\sqrt{-x+4}\right) dx$

(f) disk method: $V = \pi \int_0^4 6^2\, dx - \pi \int_0^4 \left(6 - (4x - x^2)\right)^2 dx = 144\pi - \pi \int_0^4 \left(6 - 4x + x^2\right)^2 dx = \frac{1408\pi}{15}$

shell method: $V = 4\pi \int_0^4 \sqrt{4-y}(6-y)\, dy = \frac{1408\pi}{15}$ (most efficient)

Assignment 13.4 —

(a) $V = 2\pi \int_0^3 (3-x)(2x)\, dx = 18\pi$

(b) $V = \pi \int_0^6 (6-y)\, dy = 18\pi$

(c) $V = \pi \int_{-1}^1 x^2\sqrt{1-x^2}\, dx = \frac{\pi^2}{8}$

(d) a) about the x-axis: $V = 2\pi \int_0^1 x^2\left(1 - x^2\right) dx = \frac{4\pi}{15}$

b) about the y-axis: $V = 4\pi \int_0^1 x^2\sqrt{1-x^2}\, dx = \frac{\pi^2}{4}$

(e) $V = \pi \int_{-5}^3 \left(\sqrt{25-x^2}\right)^2 dx - \pi \int_{-5}^3 \left(\frac{x+5}{2}\right)^2 dx = \pi \int_{-5}^3 \left(\frac{75}{4} - \frac{5x^2}{4} - \frac{5x}{2}\right) dx = \frac{320\pi}{3}$

(f) a) about the y-axis: $V = 2\pi \int_0^{2\pi} (\theta - \sin(\theta))(1 - \cos(\theta))(1 - \cos(\theta))\, d\theta = 6\pi^3$

b) about $y = 2$: $V = \pi \int_0^{2\pi} 2^2\, d\theta - \pi \int_0^{2\pi} \left(2 - \left(1 - \cos(\theta)\right)\right)^2 \left(1 - \cos(\theta)\right) d\theta$

$$= 8\pi^2 - 2\pi \int_0^{\pi} \sin^2(\theta)\big(1+\cos(\theta)\big)\, d\theta = 7\pi^2$$

(g) $$V = \pi \int_{\pi}^{0} 16\cos^4(\theta)\sin^2(\theta)\Big(-8\cos^2(\theta)\sin(\theta) - 4\cos^2(\theta)\sin(\theta)\Big)\, d\theta$$

$$= 192\pi \int_0^{\pi} \cos^6(\theta)\sin^3(\theta)\, d\theta = \frac{256\pi}{21}$$

Remarks:
If a curve K is given by $y = f(x)$, the volume when rotated about the x-axis is given by

$$V = \pi \int_{x_a}^{x_b} \big(y(x)\big)^2\, dx.$$

This expression can for a polar curve $r = r(\theta)$ be transformed by changing to a parameter representation, being:

$$\begin{cases} x = r(\theta)\cos(\theta) \\ y = r(\theta)\sin(\theta). \end{cases}$$

The volume when rotated about the x-axis is given by

$$V = \pi \int_{\alpha}^{\beta} r^2 \sin^2(\theta)\big(r'\cos(\theta) - r\sin(\theta)\big)\, d\theta.$$

(h) $$V = \pi \int_{-1}^{1} 1\, dx - \pi \int_{-1}^{1} x^4\, dx = \frac{8\pi}{5}$$

(i) We consider for simplicity a circle with center $(a,0)$ on the x-axis and radius a. We rotate this circle about the y-axis (a tangent).

disk method: $V = 2\pi \int_0^{a} \left(a + \sqrt{a^2-y^2}\right)^2 dy - 2\pi \int_0^{a} \left(a - \sqrt{a^2-y^2}\right)^2 dy = 2\pi \int_0^{a} 4a\sqrt{a^2-y^2}\, dy =$ $2\pi^2 a^3$

shell method: $V = 2\pi \int_0^{2a} 2x\sqrt{a^2-(x-a)^2}\, dx = 2\pi^2 a^3$

Assignment 13.5 —

(a) $$L = \int_0^{\alpha} R\, d\theta = \alpha R$$

(b) $$L = 2\int_0^{1/2} \frac{1+x^2}{1-x^2}\, dx = 2\ln(3) - 1$$

(c) $$L = 4\int_0^{a} \sqrt{1 + \frac{a^{2/3}-x^{2/3}}{x^{2/3}}}\, dx = 4a^{1/3} \int_0^{a} x^{-1/3}\, dx = 6a$$

(d) $L = \int\limits_0^{2\pi} \sqrt{a^2 \sin^2(\theta) + a^2 \left(1+\cos(\theta)\right)^2}\, d\theta = \sqrt{2}a \int\limits_0^{2\pi} \sqrt{1+\cos(\theta)}\, d\theta = 8a$

(e) $L = \int\limits_0^{\pi/3} \frac{1}{\cos(y)}\, dy = \ln(\sqrt{3}+2)$

(f) $L = \int\limits_{1/2}^{2} \sqrt{\frac{1}{\theta^2} + \left(-\frac{1}{\theta^2}\right)^2}\, d\theta = \int\limits_{1/2}^{2} \frac{\sqrt{1+\theta^2}}{\theta^2}\, d\theta = \frac{\sqrt{5}}{2} + \ln\left(\frac{3+\sqrt{5}}{2}\right)$

(g) $L = 2\int\limits_0^{\pi} \sqrt{4\left(\sin(2t)-\sin(t)\right)^2 + 4\left(\cos(t)-\cos(2t)\right)^2}\, dt = 4\sqrt{2}\int\limits_0^{\pi} \sqrt{1-\cos(t)}\, dt = 16$

(h) $L = 2\int\limits_0^{4} \sqrt{1+\left(\frac{3\sqrt{x}}{2}\right)^2}\, dx = \int\limits_0^{4} \sqrt{4+9x}\, dx = \frac{16}{27}\left(10\sqrt{10}-1\right)$

(i) $L = 2\int\limits_0^{3} \sqrt{1+\frac{1}{4}\left(\frac{x-1}{\sqrt{x}}\right)^2}\, dx = \int\limits_0^{3} \left(\sqrt{x}+\frac{1}{\sqrt{x}}\right) dx = 4\sqrt{3}$

Assignment 13.6 —

(a) $SA = 2\pi \int\limits_0^{\pi} \sin(x)\sqrt{1+\cos^2(x)}\, dx = 2\pi\left(\sqrt{2}-\ln\left(\sqrt{2}-1\right)\right)$

(b) $SA = 4\pi \int\limits_0^{4} \sqrt{\frac{16-x^2}{4}}\sqrt{1+\frac{x^2}{4(16-x^2)}}\, dx = \pi\int\limits_0^{4} \sqrt{64-3x^2}\, dx = \frac{8\pi}{9}\left(9+4\sqrt{3}\pi\right)$

(c) $SA = 2\pi \int\limits_0^{6} \left(3-\frac{y}{2}\right)\sqrt{1+\frac{1}{4}}\, dy = 9\pi\sqrt{5}$

(d) a) about the y-axis: $SA = 2\pi \int\limits_0^{1} y^3\sqrt{1+9y^4}\, dy = \frac{1}{27}(10\sqrt{10}-1)\pi$

b) the line $x=1$: $SA = 2\pi \int\limits_0^{1} (1-y^3)\sqrt{1+9y^4}\, dy$

(e) $SA = 2\pi \int\limits_0^{1} \sqrt{\frac{x^2-x^4}{8}}\sqrt{1+\frac{\left(1-2x^2\right)^2}{8(1-x^2)}}\, dx = \frac{\pi}{4}\int\limits_0^{1} (3x-2x^3)\, dx = \frac{\pi}{4}$

(f) a) about the x-axis:

$$SA = 4\pi \int\limits_0^{\pi/2} \sin^3(t)3\cos(t)\sin(t)\, dt = 12\pi\int\limits_0^{\pi/2} \sin^4(t)\cos(t)\, dt = \frac{12\pi}{5}$$

b) about the line $y=-1$:

$$SA = 4\pi \int_0^{\pi/2} \left(1+|\sin^3(t)|\right) 3\cos(t)\sin(t)\,dt + 4\pi \int_{-\pi/2}^{0} \left(1-|\sin^3(t)|\right) 3\cos(t)\sin(t)\,dt$$

$$= 12\pi \int_{-\pi/2}^{\pi/2} \left(1+\sin^3(t)\right) |\sin(t)|\cos(t)\,dt = 12\pi$$

(g) $$SA = 4\pi \int_0^{\pi/4} a\sqrt{\cos(2\theta)}\sin(\theta)\sqrt{a^2\cos(2\theta)+\frac{a^2\sin^2(2\theta)}{\cos(2\theta)}}\,d\theta = 4a^2\pi \int_0^{\pi/4} \sin(\theta)\,d\theta$$

$$= 2(2-\sqrt{2})\pi a^2$$

(h) a) about the x-axis:

$$SA = 2\pi \int_0^{2\pi} a\left(1-\cos(\theta)\right)\sqrt{2}\,a\sqrt{1-\cos(\theta)}\,d\theta = 2\sqrt{2}\pi a^2 \int_0^{2\pi} \left(1-\cos(\theta)\right)^{3/2} d\theta = \frac{64\pi a^2}{3}$$

b) about the line $x = a\pi$:

$$SA = 2\pi \int_0^{\pi} \left(a\pi - a\left(\theta-\sin(\theta)\right)\right)\sqrt{2}\,a\sqrt{1-\cos(\theta)}\,d\theta$$

$$= 2\sqrt{2}\pi a^2 \int_0^{\pi} \left(\pi-\left(\theta-\sin(\theta)\right)\right)\sqrt{1-\cos(\theta)}\,d\theta = 8\pi a^2\left(\pi-\frac{4}{3}\right)$$

Assignment 13.7 —

(a) $$V_{kir} = \pi \int_0^{5/2} \left(16-(y-4)^2\right)dy - \pi \int_1^{5/2} \left(1-(y-2)^2\right)dy = \frac{56}{3}\pi$$

(b) $$SA = 2\pi \int_{5/2}^{3} \sqrt{1-(y-2)^2}\,\frac{dy}{\sqrt{1-(y-2)^2}} = 2\pi \int_{5/2}^{3} dy = \pi$$

Chapter 14

Assignment 14.1 — The three functions $\vec{r}_i(t)$ are on the x-axis $(y \geq 0)$. The points on the 3 curves are at distance 1 from the origin. Calculate the norm of the vector functions to verify this. The vector functions $\vec{r}_i(t)$ correspond to the semicircle $y = \sqrt{1-x^2}$, travelled from left to right. To do this, calculate the values at start and end points.

Chapter 14

Assignment 14.2 —

(a) $\big(11, 74, \sin(5)\big)$

(b) $\big(e^3, 0\big)$

(c) $(1, e)$

(d) $(2t, 1, 0)$

Chapter 14

Assignment 14.3 —

(a) $\left(-\frac{1}{t^2}, \frac{5}{(3t+1)^2}, \frac{1}{\cos^2(t)}\right)$

(b) $\big(2t\sin(t) + t^2\cos(t), 6t^2 + 10t\big)$

(c) $2t\sin(t) + \big(t^2+1\big)\cos(t) + 4t + 3$

(d) $\big(-1, -2t + \cos(t), 6t^2 + 10t + 2 - (t-1)\cos(t) - \sin(t)\big)$

Chapter 14

Assignment 14.4 —

(a) $t = 2k\pi, \quad k \in \mathbb{Z}$

(b) $t = 1$

(c) $t = \frac{3\pi}{4} + k\pi, \quad k \in \mathbb{Z}$

(d) $t = \pm 1$

Chapter 14

Assignment 14.5 —

(a) $\left(\frac{t^4}{4}, \sin(t), e^t(t-1)\right) + \vec{\boldsymbol{c}}$

(b) $\big(\arctan(t), \tan(t)\big) + \vec{\boldsymbol{c}}$

(c) $\left(e^{\sin(t)}, \frac{t^2}{4} - \frac{t\sin(2t)}{4} - \frac{\cos(2t)}{8}, -t\right) + \vec{\boldsymbol{c}}$

(d) $\left(0, \frac{2}{3}\right)$

(e) $\left(1-\dfrac{1}{\sqrt{e}}, \sqrt{e}-1, e-1\right)$

Chapter 14

Assignment 14.6 —

(a) $\vec{r}(t) = \left(t^2-t+5, \dfrac{3}{2}t^2-t-\dfrac{5}{2}\right)$

(b) $\vec{r}(t) = \big(1-\cos(t), \sin(t)\big)$

(c) $\vec{r}(t) = \big(10t, -16t^2+50t\big)$

Chapter 14

Assignment 14.7 —

(a) $\vec{r}(t) = (1,t)$, $\quad \vec{v}(t) = (0,1)$, $\quad \|\vec{v}(t)\| = 1$, $\quad \vec{a}(t) = (0,0)$

path: $x = 1$ in the xy-plane

(b) $\vec{r}(t) = \big(0,t^2,t\big)$, $\quad \vec{v}(t) = (0,2t,1)$, $\quad \|\vec{v}(t)\| = \sqrt{4t^2+1}$, $\quad \vec{a}(t) = (0,2,0)$

path: $y = z^2$ in the plane $x = 0$

(c) $\vec{r}(t) = (1,t,t)$, $\quad \vec{v}(t) = (0,1,1)$, $\quad \|\vec{v}(t)\| = \sqrt{2}$, $\quad \vec{a}(t) = (0,0,0)$

path: the intersection of the planes $x = 1$ and $y = z$

(d) $\vec{r}(t) = \big(t^2,-t^2,1\big)$, $\quad \vec{v}(t) = (2t,-2t,0)$, $\quad \|\vec{v}(t)\| = 2\sqrt{2}t$, $\quad \vec{a}(t) = (2,-2,0)$

path: the ray $\begin{cases} x = -y \geq 0 \\ z = 1 \end{cases}$

(e) $\vec{r}(t) = \big(3\cos(t), 4\cos(t), 5\sin(t)\big)$, $\quad \vec{v}(t) = \big(-3\sin(t), -4\sin(t), 5\cos(t)\big)$, $\quad \|\vec{v}(t)\| = 5$,

$\vec{a}(t) = \big(-3\cos(t), -4\cos(t), -5\sin(t)\big)$

path: circle that is the cross section of the sphere $x^2+y^2+z^2 = 25$ and the plane $4x = 3y$

(f) $\vec{r}(t) = \big(3\cos(t), 4\sin(t), t\big)$, $\quad \vec{v}(t) = \big(-3\sin(t), 4\cos(t), 1\big)$, $\quad \|\vec{v}(t)\| = \sqrt{10+7\cos^2(t)}$,

$\vec{a}(t) = \big(-3\cos(t), -4\sin(t), 0\big)$

path: a spiral around the elliptical cylinder $\dfrac{x^2}{9}+\dfrac{y^2}{16} = 1$

Assignment 14.8 —

(a) Parameterization C: $\vec{r}(t) = \big(x(t), x^2(t)\big) = x(t)\hat{\imath} + x^2(t)\hat{\jmath}$

(b) $\vec{v}(t) = \dfrac{dx(t)}{dt}\,(\hat{\imath}+2x\hat{\jmath})$

(c) $\vec{\boldsymbol{a}}(t) = \dfrac{d^2x(t)}{dt^2}(\hat{\boldsymbol{\imath}} + 2x\hat{\boldsymbol{\jmath}}) + 2\left(\dfrac{dx}{dt}\right)^2 \hat{\boldsymbol{\jmath}}$

(d) $\left\|\vec{\boldsymbol{v}}(t)\right\| = \left|\dfrac{dx}{dt}\right|\sqrt{1+4x^2} = 5 \;\Rightarrow\; \dfrac{dx}{dt} = \dfrac{5}{\sqrt{1+4x^2}} \overset{x=1}{\Rightarrow} \dfrac{dx}{dt} = \sqrt{5}$

(e) $\dfrac{d^2x(t)}{dt^2} = -\dfrac{100x}{(1+4x^2)^2} \overset{x=1}{\Rightarrow} \dfrac{d^2x(t)}{dt^2} = -4$

(f) $\vec{\boldsymbol{v}}(t) \overset{x=1}{=} \sqrt{5}\hat{\boldsymbol{\imath}} + 2\sqrt{5}\hat{\boldsymbol{\jmath}} = \left(\sqrt{5}, 2\sqrt{5}\right)$

(g) $\vec{\boldsymbol{a}}(t) \overset{x=1}{=} -4\hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}} = (-4, 2)$

Assignment 14.9 —

(a) $\widehat{\boldsymbol{T}}(t) = \dfrac{1}{\sqrt{20t^2-4t+1}}(4t, 2t-1)$

(b) $\widehat{\boldsymbol{T}}(t) = \dfrac{1}{\sqrt{1+16t^2+81t^4}}\left(1, -4t, 9t^2\right)$

(c) $\widehat{\boldsymbol{T}}(t) = \dfrac{1}{\sqrt{1+t^2+t^4}}\left(1, t, t^2\right)$

(d) $\widehat{\boldsymbol{T}}(t) = \dfrac{1}{\sqrt{1+\sin^2(t)}}\left(\cos(2t), \sin(2t), -\sin(t)\right)$

(e) $\widehat{\boldsymbol{T}}(t) = \left(-\cos(t), \sin(t)\right) \quad\Rightarrow\quad \widehat{\boldsymbol{T}}\left(\dfrac{\pi}{6}\right) = \left(-\dfrac{\sqrt{3}}{2}, \dfrac{1}{2}\right)$

Assignment 14.10 —

(a) $\widehat{\boldsymbol{N}}(t) = \left(\dfrac{3}{5}, -\dfrac{4}{5}\right)$

(b) $\widehat{\boldsymbol{N}}(t) = \left(-\cos(t), -\sin(t)\right)$

(c) $\widehat{\boldsymbol{N}}(t) = \left(\dfrac{e^{-t}}{\left(e^{2t}+e^{-2t}\right)^2}, \dfrac{e^{t}}{\left(e^{2t}+e^{-2t}\right)^2}\right)$

(d) $\widehat{\boldsymbol{N}}(t) = \left(0, -\sin(t), -\cos(t)\right)$

(e) $\widehat{\boldsymbol{N}}(t) = \left(\sin(t), \cos(t)\right) \quad\Rightarrow\quad \widehat{\boldsymbol{N}}\left(\dfrac{\pi}{6}\right) = \left(\dfrac{1}{2}, \dfrac{\sqrt{3}}{2}\right)$

Assignment 14.11 —

(a) $L = \displaystyle\int_0^1 t\sqrt{8+9t^2}\,dt = \dfrac{17\sqrt{17}-16\sqrt{2}}{27}$ (assume $8+9t^2 = u^2$)

(b) $L = \displaystyle\int_0^{2\pi} \sqrt{2e^{2t}+1}\,dt = \sqrt{2e^{4\pi}+1} - \sqrt{3} + \ln\left(\sqrt{2e^{4\pi}+1}-1\right) - 2\pi - \ln\left(\sqrt{3}-1\right)$

(assume $2e^{2t}+1 = u^2$)

(c) $L = \int_{-1}^{0} (-t)\sqrt{9t^2+4}\, dt + \int_{0}^{2} t\sqrt{9t^2+4}\, dt = \frac{1}{27}\left(13^{3/2} + 40^{3/2} - 16\right)$ (stel $9t^2+4 = u^2$)

Assignment 14.12 —

(a) $\vec{r}(s) = \left(\frac{2s}{3}, \frac{s}{3}, -\frac{2s}{3}\right)$

(b) $\vec{r}(s) = \left(7\cos\left(\frac{s}{7}\right), 7\sin\left(\frac{s}{7}\right)\right)$

(c) $\vec{r}(s) = \left(3\cos\left(\frac{s}{\sqrt{13}}\right), 3\sin\left(\frac{s}{\sqrt{13}}\right), \frac{2s}{\sqrt{13}}\right)$

(d) $\vec{r}(s) = \left(\frac{s+\sqrt{s^2+4}}{2}, \sqrt{2}\ln\left(\frac{s+\sqrt{s^2+4}}{2}\right), \frac{-2}{s+\sqrt{s^2+4}}\right)$ met $s(t) = e^t - e^{-t}$

Assignment 14.13 —

(a) $\kappa(x) = \frac{2}{(1+4x^2)^{3/2}} \Rightarrow \kappa(0) = 2$ and $\kappa(\sqrt{2}) = 2/27$

The radii of curvature at $x = 0$ and $x = \sqrt{2}$ are $1/2$ and $27/2$.

(b) $\kappa(x) = \frac{|\cos(x)|}{(1+\sin^2(x))^{3/2}} \Rightarrow \kappa(0) = 1$ and $\kappa(\pi/2) = 0$

The radii of curvature at $x = 0$ and $x = \pi/2$ are 1 and infinite.

(c) $\kappa(x) = \frac{\left|2\frac{\tan(x)}{\cos^2(x)}\right|}{(1+\cos^{-4}(x))^{3/2}} \Rightarrow \kappa(\pi/4) = 4/5\sqrt{5}$

The radius of curvature at $x = \pi/4$ is $5\sqrt{5}/4$.

(d) $\kappa(t) = \frac{\left\|\left(\frac{4}{t^3}, 0, \frac{4}{t^3}\right)\right\|}{\left\|\left(2, -\frac{1}{t^2}, -2\right)\right\|^3} = \frac{4\sqrt{2}t^3}{\left(8t^4+1\right)^{3/2}}$. In $(2,1,-2)$ is $t = 1$. $\Rightarrow \kappa(1) = \frac{4\sqrt{2}}{27}$

The radius of curvature at $t = 1$ is $\frac{27}{4\sqrt{2}}$.

(e) $\kappa(t) = \frac{\left\|(-2, 6t, -6t^2)\right\|}{\left\|(3t^2, 2t, 1)\right\|^3} = \frac{\sqrt{4+36t^2+36t^4}}{(9t^4+4t^2+1)^{3/2}} \Rightarrow \kappa(1) = \frac{2\sqrt{19}}{14^{3/2}}$

The radius of curvature at $t = 1$ is $\frac{14^{3/2}}{2\sqrt{19}}$.

(f) The given curve has a vertical tangent ($y' = +\infty$) in $(\pm 2, 0)$, which makes $\kappa(x) = 0$. The radius of curvature at $x = 2$ equals the distance between the origin and $(2,0)$, being 2.

(g) $\kappa(t) = \frac{|-18(t^2+1)|}{(3t^2+3)^3} \Rightarrow \kappa(1) = 1/6$

The radius of curvature at $t = 1$ is 6.

(h) $\kappa(t) = \dfrac{\left|9\sin(t)\sin(3t) + 3\cos(t)\cos(3t)\right|}{\left(\sin^2(t) + 9\cos^2(t)\right)^{3/2}} \quad \Rightarrow \kappa(0) = 1/9$

The radius of curvature at $t = 0$ is 9.

Assignment 14.14 —

(a) $\kappa(x) = \dfrac{\left|2(x^2+1)^4\left(4x^2(x^2+1)^{-1} - 1\right)\right|}{\left((x^2+1)^4 + 4x^2\right)^{3/2}}$.

The radius of curvature is $\dfrac{\left((x^2+1)^4 + 4x^2\right)^{3/2}}{\left|2(x^2+1)^4\left(4x^2(x^2+1)^{-1} - 1\right)\right|}$.

(b) $\kappa(x) = 1$. The radius of curvature is 1.

(c) $\kappa(t) = \dfrac{\left\|\left(-\cos(t), \dfrac{\sin(t)}{\sqrt{2}}, \dfrac{-\sin(t)}{\sqrt{2}}\right)\right\|}{\left\|\left(-\sqrt{2}\sin(t), -\cos(t), \cos(t)\right)\right\|} = \dfrac{\sqrt{2}}{2}$. The radius of curvature is $\sqrt{2}$.

(d) $\kappa(x) = \dfrac{e^x}{(1+e^{2x})^{3/2}}$. The radius of curvature is $\dfrac{(1+e^{2x})^{3/2}}{e^x}$.

(e) $\kappa(\theta) = \dfrac{3a^2(1-\cos(\theta))}{\left(2a^2(1-\cos(\theta))\right)^{3/2}} = \dfrac{3}{2\sqrt{2ar}}$. The radius of curvature is $\dfrac{2\sqrt{2ar}}{3}$.

(f) $\kappa(t) = \dfrac{2}{\left(3\sin^2(t) + 1\right)^{3/2}}$.

The radius of curvature is $\dfrac{\left(3\sin^2(t) + 1\right)^{3/2}}{2}$.

(g) $\kappa(t) = \dfrac{\left|-\frac{1}{\sin^2(t)}\right|}{(1+\cot^2(t))^{3/2}} = \sin(t)$. The radius of curvature is $\csc(t)$.

Assignment 14.15 —

(a) $\vec{\boldsymbol{r}}(y) = \left(\sqrt{a^2 - y^2}, y\right)$ with $0 \le y \le a$

(b) $\vec{\boldsymbol{r}}(\phi) = \left(a\sin(\phi), -a\cos(\phi)\right)$ with $\dfrac{\pi}{2} \le \phi \le \pi$

(c) $\vec{\boldsymbol{r}}(s) = \left(a\sin\left(\dfrac{s}{a}\right), a\cos\left(\dfrac{s}{a}\right)\right)$ with $0 \le s \le \dfrac{a\pi}{2}$

Chapter 15

Assignment 15.1 —

(a) $\operatorname{dom} f = \left\{(x,y) \mid x^2 + y^2 < 4\right\}$ and $\operatorname{im} f = \mathbb{R}_0^+$

(b) $\operatorname{dom} f = \left\{(x,y) |\ x \in \mathbb{R}_0^+ \ \wedge\ y \in \mathbb{R} \backslash \{k\pi, k \in \mathbb{Z}\}\right\}$ and $\operatorname{im} f = \mathbb{R}$

(c) $\operatorname{dom} f = \left\{(x,y) |\ x \in \mathbb{R}^+ \backslash \{0, 1/2\} \ \wedge\ y \in [-1,1]\right\}$ and $\operatorname{im} f = \mathbb{R}_0$

(d) $\operatorname{dom} f = \left\{(x,y) |\ x^2 + y^2 \leq 1\right\}$ and $\operatorname{im} f = [0,1]$

(e) $\operatorname{dom} f = \left\{(x,y) |\ x \in \mathbb{R} \ \wedge\ y \in \mathbb{R}\right\}$ and $\operatorname{im} f = [-1,1]$

(f) $\operatorname{dom} f = \left\{(x,y) |\ x \in \mathbb{R} \ \wedge\ y \in \mathbb{R}\right\}$ and $\operatorname{im} f = \mathbb{R}$

(g) $\operatorname{dom} f = \left\{(x,y) |\ x \neq y\right\}$ and $\operatorname{im} f = \mathbb{R}_0$

(h) $\operatorname{dom} f = \left\{(x,y) |\ x^2 + y^2 > 3; \wedge\ x^2 + y^2 \neq 4\right\}$ and $\operatorname{im} f = \mathbb{R}_0$

(i) $\operatorname{dom} f = \left\{(x,y) |\ x^2 + 2y^2 \leq 1\right\}$ and $\operatorname{im} f = \left[\dfrac{\pi}{2}, \pi\right]$

Assignment 15.2 —

(a) The level curves corresponding to $c \neq 0$ are parabolas with their vertex at the origin and the y-axis as axis of symmetry. The level curves corresponding to $c = 0$ is al ine (the y-axis).

(b) The level curves corresponding to $c \neq 0$ are circles with the center at the y-axis. The level curves corresponding to $c = 0$ are lines (the x-axis).

(c) The level curves corresponding to $c \neq 0$ are bell-shaped curves with the y-axis as the axis of symmetry. The level curve for $c = 0$ is a curve with the y-axis as its vertical asymptote.

Assignment 15.3 —

(a) $f_x = 2x - 3$ and $f_y = 4y + 2$

(b) $f_x = 6x^2y^2 + 4$ and $f_y = 4x^3y + 2$

(c) $f_x = 6x^2 \cos\left(2x^3 - y^3\right)$ and $f_y = -3y^2 \cos\left(2x^3 - y^3\right)$

(d) $f_x = e^{x^2}\left(2x\cos(xy) - y\sin(xy)\right)$ and $f_y = -x\, e^{x^2} \sin(xy)$

(e) $f_x = 2xy\cos\left(x^2 + y^2\right)$ and $f_y = \sin\left(x^2 + y^2\right) + 2y^2\cos\left(x^2 + y^2\right)$

(f) $f_x = 4x^3 \sin\left(xy^3\right) + x^4y^3\cos\left(xy^3\right)$ and $f_y = 3x^5y^2\cos\left(xy^3\right)$

(g) $f_x = y\, e^{xy} \sin\left(4y^2\right)$ en $f_y = x\, e^{xy}\sin\left(4y^2\right) + 8y\, e^{xy}\cos\left(4y^2\right)$

(h) $f_x = -\dfrac{2y}{(x-y)^2}$ and $f_y = \dfrac{2x}{(x-y)^2}$

(i) $f_x = \dfrac{1}{\sqrt{y}\left(x^2 + y^2\right)}$ and $f_y = -\dfrac{x}{y^{3/2}\left(x^2 + y^2\right)} - \dfrac{3}{2} y^{-5/2} \arctan\left(\dfrac{x}{y}\right)$

(j) $f_x = \dfrac{x}{\sqrt{x^2 + 4y^2}}$ and $f_y = \dfrac{4y}{\sqrt{x^2 + 4y^2}}$

(k) $f_x = 2x + 4y + 5z$ and $f_y = -4y + 4x + 6z$ and $f_z = 6z + 5x + 6y$

(l) $f_x = 2xy^4z^3 + y$ and $f_y = 4x^2y^3z^3 + x$ and $f_z = 3x^2y^4z^2 + 2z$

(m) $f_x = \dfrac{x}{\sqrt{x^2 + y^2 + z^2}}$ and $f_y = \dfrac{y}{\sqrt{x^2 + y^2 + z^2}}$ and $f_z = \dfrac{z}{\sqrt{x^2 + y^2 + z^2}}$

(n) $f_x = 2xy\cos(z)$ and $f_y = x^2\cos(z)$ and $f_z = -x^2y\sin(z)$

(o) $f_x = \dfrac{2z^2}{x}$ and $f_y = \dfrac{z^2}{y}$ and $f_z = 2z\ln(x^2y)$

Assignment 15.4 —

	f_x	in a	f_y	in a	f_z	in a
(a)	$\dfrac{1}{1+(x+y+z)^2}$	$\dfrac{1}{2}$	$\dfrac{1}{1+(x+y+z)^2}$	$\dfrac{1}{2}$	$\dfrac{1}{1+(x+y+z)^2}$	$\dfrac{1}{2}$
(b)	$2x-2y+6z+4$	4	$6y-2x+7z-3$	-3	$12z+6x+7y$	0
(c)	$\dfrac{y}{2\sqrt{xy+z^2}}$	$\dfrac{\sqrt{2}}{4}$	$\dfrac{x}{2\sqrt{xy+z^2}}$	$\dfrac{\sqrt{2}}{4}$	$\dfrac{z}{\sqrt{xy+z^2}}$	$\dfrac{\sqrt{2}}{2}$
(d)	$(1+xy)\,e^{xy+z}$	1	$x^2\,e^{xy+z}$	1	$x\,e^{xy+z}$	1
(e)	$e^{x+y^2+z^3}$	1	$2y\,e^{x+y^2+z^3}$	0	$3z^2\,e^{x+y^2+z^3}$	0
(f)	$\sin(y)$	0	$x\cos(y)+\ln(z)$	-1	$\dfrac{y}{z}$	π
(g)	$\dfrac{(xy)^z z}{x} + z^{xy}\,y\ln(z)$	1	$\dfrac{(xy)^z z}{y} + z^{xy}\,x\ln(z)$	1	$(xy)^z\ln(xy) + \dfrac{z^{xy}\,xy}{z}$	1

	f_{xx}	in a	f_{xy}	in a	f_{yy}	in a
(a)	$-\dfrac{2(x+y+z)}{(1+(x+y+z)^2)^2}$	$-\dfrac{1}{2}$	$=f_{xx}$	$-\dfrac{1}{2}$	$=f_{xx}$	$-\dfrac{1}{2}$
(b)	2	2	-2	-2	6	6
(c)	$-\dfrac{y^2}{4(xy+z^2)^{\frac{3}{2}}}$	$-\dfrac{\sqrt{2}}{16}$	$\dfrac{xy+2z^2}{4(xy+z^2)^{\frac{3}{2}}}$	$\dfrac{3\sqrt{2}}{16}$	$-\dfrac{x^2}{4(xy+z^2)^{\frac{3}{2}}}$	$-\dfrac{\sqrt{2}}{16}$
(d)	$ye^{xy+z}(xy+2)$	0	$xe^{xy+z}(xy+2)$	2	x^3e^{xy+z}	1
(e)	$e^{x+y^2+z^3}$	1	$2ye^{x+y^2+z^3}$	0	$2\left(2y^2+1\right)e^{x+y^2+z^3}$	2
(f)	0	0	$\cos(y)$	-1	$-x\sin(y)$	0
(g)	(*)	0	(*)	1	(*)	0

(*): These expressions are too long to fit in the table

	f_{xz}	in a	f_{yz}	in a	f_{zz}	in a
(a)	$=f_{xx}$	$-\frac{1}{2}$	$=f_{xx}$	$-\frac{1}{2}$	$=f_{xx}$	$-\frac{1}{2}$
(b)	6	6	7	7	12	12
(c)	$-\frac{yz}{2(xy+z^2)^{\frac{3}{2}}}$	$-\frac{\sqrt{2}}{8}$	$-\frac{xz}{2(xy+z^2)^{\frac{3}{2}}}$	$-\frac{\sqrt{2}}{8}$	$\frac{xy}{(xy+z^2)^{\frac{3}{2}}}$	$\frac{\sqrt{2}}{4}$
(d)	$e^{xy+z}(xy+1)$	1	x^2e^{xy+z}	1	xe^{xy+z}	1
(e)	$3z^2e^{x+y^2+z^3}$	0	$6yz^2e^{x+y^2+z^3}$	0	$3z\left(3z^3+2\right)e^{x+y^2+z^3}$	0
(f)	0	0	$\frac{1}{z}$	1	$-\frac{y}{z^2}$	$-\pi$
(g)	(*)	2	(*)	2	(*)	0

(*): These expressions are too long to fit in the table

Assignment 15.5 —

(a) $\frac{\partial z}{\partial x}=e^{xy(x^2+y^2)}(3x^2y+y^3)$

(b) $\frac{\partial u}{\partial x}=\frac{y}{\sqrt{x}},\quad \frac{\partial u}{\partial y}=2\sqrt{x}+6y\sqrt[3]{z^2},\quad \frac{\partial u}{\partial z}=\frac{2y^2}{\sqrt[3]{z}}$

(c) $dz=\frac{xdy-ydx}{x^2+y^2}$

Assignment 15.6 —

$$\frac{\partial z}{\partial x}=\frac{x}{x^2+y^2},\quad \frac{\partial z}{\partial y}=\frac{y}{x^2+y^2}$$

Assignment 15.7 —

$$\frac{\partial z}{\partial x}=\frac{e^{\frac{x}{y}}}{y}\left(\sin\left(\frac{x}{y}\right)+\cos\left(\frac{x}{y}\right)\right)-\frac{ye^{\frac{y}{x}}}{x^2}\left(\cos\left(\frac{y}{x}\right)-\sin\left(\frac{y}{x}\right)\right)$$

$$\frac{\partial z}{\partial y}=-\frac{xe^{\frac{x}{y}}}{y^2}\left(\sin\left(\frac{x}{y}\right)+\cos\left(\frac{x}{y}\right)\right)+\frac{e^{\frac{y}{x}}}{x}\left(\cos\left(\frac{y}{x}\right)-\sin\left(\frac{y}{x}\right)\right)$$

Assignment 15.8 —

(a) $$\left(\frac{\partial f}{\partial T}\right)_{p,n}=\left(-\frac{2n^2a}{V^3}\right)\left(\frac{\partial V}{\partial T}\right)_{p,n}(V-nb)+\left(p+\frac{n^2a}{V^2}\right)\left(\frac{\partial V}{\partial T}\right)_{p,n}-nR=0$$

$$\Leftrightarrow\quad \left(\frac{\partial V}{\partial T}\right)_{p,n}=\frac{nR}{p-\frac{n^2a}{V^2}+\frac{2n^3ab}{V^3}}$$

(b) $$\left(\frac{\partial f}{\partial p}\right)_{T,n}=\left(1-\frac{2n^2a}{V^3}\left(\frac{\partial V}{\partial p}\right)_{T,n}\right)(V-nb)+\left(p+\frac{n^2a}{V^2}\right)\left(\frac{\partial V}{\partial p}\right)_{T,n}=0$$

$$\Leftrightarrow \quad \left(\frac{\partial V}{\partial p}\right)_{T,n} = \frac{nb-V}{p-\frac{n^2a}{V^2}+\frac{2n^3ab}{V^3}}$$

(c) $$\left(\frac{\partial f}{\partial T}\right)_{V,n} = \left(\frac{\partial p}{\partial T}\right)_{V,n}(V-nb)-nR=0 \quad \Leftrightarrow \quad \left(\frac{\partial p}{\partial T}\right)_{V,n} = \frac{nR}{V-nb}$$

(d) $$\left(\frac{\partial f}{\partial V}\right)_{T,n} = \left(\left(\frac{\partial p}{\partial V}\right)_{T,n} - \frac{2n^2a}{V^3}\right)(V-nb)+\left(p+\frac{n^2a}{V^2}\right)=0 \quad \Leftrightarrow \quad \left(\frac{\partial p}{\partial V}\right)_{T,n} = \frac{2n^2a}{V^3} - \frac{p+\frac{n^2a}{V^2}}{V-nb}$$

Assignment 15.9 — $df = 3y(xy)^2\,dx + 3x(xy)^2\,dy$

Assignment 15.10 —

	f_x	in a	f_y	in a
(a)	$2x+2y-2$	4	$2x+2y+3$	9
(b)	$2xy^5+3x^2y+y^2$	34	$5x^2y^4+x^3+2xy$	78
(c)	$-\frac{y\left(x^2-y^2\right)}{\left(x^2+y^2\right)^2}$	0	$\frac{x\left(x^2-y^2\right)}{\left(x^2+y^2\right)^2}$	0
(d)	$\frac{x^y y}{x}$	1	$x^y \ln(x)$	0
(e)	$\frac{2}{2x-3y}$	2	$-\frac{3}{2x-3y}$	-3
(f)	$-\frac{e^y}{x^2}$	$-e$	$\frac{e^y}{x}$	e
(g)	$-\frac{ye^{\frac{y}{x}}}{x^2}$	$-e$	$\frac{e^{\frac{y}{x}}}{x}$	e
(h)	$-3\sin(3x+2y)$	0	$-2\sin(3x+2y)$	0
(i)	$\frac{1}{1+x^2+2xy+y^2}$	$\frac{1}{2}$	$\frac{1}{1+x^2+2xy+y^2}$	$\frac{1}{2}$
(j)	$f(x,y)\left(\ln(2x+y)+\frac{2(x+3y)}{2x+y}\right)$	6	$f(x,y)\left(3\ln(2x+y)+\frac{x+3y}{2x+y}\right)$	3

	f_{xx}	in a	f_{xy}	in a	f_{yy}	in a
(a)	2	2	2	2	2	2
(b)	$2y^5+6xy$	20	$10xy^4+3x^2+2y$	59	$20x^2y^3+2x$	186
(c)	$\frac{2xy\left(x^2-3y^2\right)}{\left(x^2+y^2\right)^3}$	$-\frac{1}{2}$	$\frac{6x^2y^2-x^4-y^4}{\left(x^2+y^2\right)^3}$	$\frac{1}{2}$	$-\frac{2xy\left(3x^2-y^2\right)}{\left(x^2+y^2\right)^3}$	$-\frac{1}{2}$
(d)	$\frac{x^y y(y-1)}{x^2}$	0	$\frac{x^y(y\ln(x)+1)}{x}$	1	$x^y\ln^2(x)$	0
(e)	$-\frac{4}{(2x-3y)^2}$	-4	$\frac{6}{(2x-3y)^2}$	6	$-\frac{9}{(2x-3y)^2}$	-9
(f)	$\frac{2e^y}{x^3}$	$2e$	$-\frac{e^y}{x^2}$	$-e$	$\frac{e^y}{x}$	e
(g)	$\frac{ye^{\frac{y}{x}}(2x+y)}{x^4}$	$3e$	$-\frac{e^{\frac{y}{x}}(x+y)}{x^3}$	$-2e$	$\frac{e^{\frac{y}{x}}}{x^2}$	e
(h)	$-9\cos(3x+2y)$	-9	$-6\cos(3x+2y)$	-6	$-4\cos(3x+2y)$	-4
(i)	$-\frac{2(x+y)}{(1+(x+y)^2)^2}$	$-\frac{1}{2}$	$-\frac{2(x+y)}{(1+(x+y)^2)^2}$	$-\frac{1}{2}$	$-\frac{2(x+y)}{(1+(x+y)^2)^2}$	$-\frac{1}{2}$
(j)	(*)	28	(*)	19	(*)	12

(*): These expressions are too long to fit in the table

Total differentials ine the given points (first order) :

(a) $df=(2x+2y-2)\,dx+(2x+2y+3)\,dy \quad \Rightarrow df(1,2)=4dx+9dy$

(b) $df=(2xy^5+3x^2y+y^2)\,dx+(5x^2y^4+x^3+2xy)\,dy \quad \Rightarrow df(3,1)=34dx+78dy$

(c) $df=-\frac{y(x^2-y^2)}{(x^2+y^2)^2}\,dx+\frac{x(x^2-y^2)}{(x^2+y^2)^2}\,dy \quad \Rightarrow df(1,1)=0$

(d) $df=\frac{x^y y}{x}\,dx+x^y\ln(x)\,dy \quad \Rightarrow df(1,1)=dx$

(e) $df=\frac{2}{2x-3y}\,dx-\frac{3}{2x-3y}\,dy \quad \Rightarrow df(2,1)=2dx-3dy$

(f) $df=-\frac{e^y}{x^2}\,dx+\frac{e^y}{x}\,dy \quad \Rightarrow df(1,1)=-e\,dx+e\,dy$

(g) $df=-\frac{ye^{\frac{y}{x}}}{x^2}\,dx+\frac{e^{\frac{y}{x}}}{x}\,dy \quad \Rightarrow df(1,1)=-e\,dx+e\,dy$

(h) $df=-3\sin(3x+2y)\,dx-2\sin(3x+2y)\,dy \quad \Rightarrow df(0,\pi)=0$

(i) $df=\frac{1}{1+(x+y)^2}(dx+dy) \quad \Rightarrow df(1,0)=\frac{1}{2}(dx+dy)$

(j) $df=f(x,y)\left(\ln(2x+y)+\frac{2(x+3y)}{2x+y}\right)dx+f(x,y)\left(3\ln(2x+y)+\frac{x+3y}{2x+y}\right)$

$$\Rightarrow df(0,1) = 6dx + 3dy$$

Assignment 15.11 —

(a) 3% (b) 2% (c) 1%

Assignment 15.12 —

$$\frac{dz}{dt} = 1$$

Assignment 15.13 —

$$u_t = \frac{\partial u}{\partial t} = \frac{xse^{st} - ys^2 \sin(t)}{\sqrt{x^2+y^2}}$$

Assignment 15.14 —

(a) $f_x = \dfrac{20x - 4y - 5)}{(x+2y+1)^2 + (3x-y-1)^2}$

$f_y = \dfrac{-2(x-5y-3)}{(x+2y+1)^2 + (3x-y-1)^2}$

(b) $f_x = [16xy(x-y) - 2]y + (8x^2y^2+3) = 24x^2y^2 - 16xy^3 - 2y + 3$

$f_y = [16xy(x-y) - 2]x - (8x^2y^2+3) = 16x^3y - 24x^2y^2 - 2x - 3$

(c) $f_x = \dfrac{e^{xy}}{\sin^2\left(x^2-y^2\right)}\left(y\sin\left(x^2-y^2\right) - 2x\cos\left(x^2-y^2\right)\right)$

$f_y = \dfrac{e^{xy}}{\sin^2\left(x^2-y^2\right)}\left(x\sin\left(x^2-y^2\right) + 2y\cos\left(x^2-y^2\right)\right)$

(d) $f_x = y^2 e^{x^3y^3}\left(3x^3y^3+1\right)$

$f_y = xye^{x^3y^3}\left(3x^3y^3+2\right)$

(e) $f_x = \left(x^2+y^2\right)^{xy}\left(\dfrac{2x^2y}{x^2+y^2} + y\ln\left(x^2+y^2\right)\right) f_y = \left(x^2+y^2\right)^{xy}\left(\dfrac{2xy^2}{x^2+y^2} + x\ln\left(x^2+y^2\right)\right)$

(f) $f_x = -\sin(x-y)\sin(x^2+y^2) + 2x\cos(x-y)\cos(x^2+y^2)$

$f_y = \sin(x-y)\sin(x^2+y^2) + 2y\cos(x-y)\cos(x^2+y^2)$

(g) $f_x = \dfrac{2(3x-y) - 3(2x+y)}{(2x+y)^2 + (3x-y)^2} = \dfrac{-5y}{(2x+y)^2 + (3x-y)^2}$

$f_y = \dfrac{(3x-y) - (2x+y)}{(2x+y)^2 + (3x-y)^2} = \dfrac{x-2y}{(2x+y)^2 + (3x-y)^2}$

(h) $f_x = 5[\sin(x) + \cos(y) + \sin(x) - \cos(y)]\cos(x) = 10\sin(x)\cos(x)$

$f_y = [\sin(x) + \cos(y) - \sin(x) + \cos(y)]\sin(y) = 2\sin(y)\cos(y)$

Assignment 15.15 — $\dfrac{dw}{dt} = (e^t + 2e^{-t})e^t - (2e^t + e^{-t})e^{-t} + (e^t + e^{-t})(e^t - e^{-t}) = 2e^{2t} - 2e^{-2t}$

Assignment 15.16 —

$$\frac{\partial^2 z}{\partial s^2} = 4f_{xx}(x,y) + 12f_{xy}(x,y) + 9f_{yy}(x,y)$$
$$\frac{\partial^2 z}{\partial s \partial t} = 6f_{xx}(x,y) + 5f_{xy}(x,y) - 6f_{yy}(x,y)$$
$$\frac{\partial^2 z}{\partial t^2} = 9f_{xx}(x,y) - 12f_{xy}(x,y) + 4f_{yy}(x,y)$$

Assignment 15.17 —

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial w}{\partial z}\frac{dz}{dt}$$

Assignment 15.18 —

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{dy}{dx}\frac{\partial x}{\partial u}$$

Assignment 15.19 — $D_{\hat{\boldsymbol{u}}}f(-1,-1) = \dfrac{4}{\sqrt{5}}$ with $\hat{\boldsymbol{u}}$ the unit vector in the direction of $\vec{\boldsymbol{v}}$

Assignment 15.20 —

$\nabla f(1,1) = (-2,4) = -2\hat{\boldsymbol{\imath}} + 4\hat{\boldsymbol{\jmath}}$

Assignment 15.21 —

$$D_{\hat{\boldsymbol{u}}}f(P) = \frac{11}{3}$$

Assignment 15.22 —

(a) Tangent plane: $3x - 2y - z = 4$. Normal: $\dfrac{x-1}{3} = \dfrac{y+1}{-2} = \dfrac{z-1}{-1}$

(b) Tangent plane: $48x - 14y - z = 64$. Normal: $\dfrac{x-1}{48} = \dfrac{y+2}{-14} = \dfrac{z-12}{-1}$

(c) Tangent plane: $-3x + 4z = 25$. Normal: $\begin{cases} y = 0 \\ \dfrac{x+3}{3} = \dfrac{z-4}{-4}. \end{cases}$

(d) Tangent plane: $3x + 2y - 6z = -5$. Normal: $\dfrac{x-3}{3} = \dfrac{y+1}{2} = \dfrac{z-2}{-6}$

(e) Tangent plane: $z = x$. Normal: $\begin{cases} y = 2 \\ x = -z. \end{cases}$

(f) Tangent plane: $x - y - z + 1 = 0$. Normal: $4x - \pi = \pi - 4y = 4 - 4z$

(g) Tangent plane: $x + z = -1$. Normal: $\begin{cases} y = 0 \\ x + 1 = z. \end{cases}$

Assignment 15.23 — The distance form the point $(1,1,0)$ to the paraboloid is $\dfrac{\sqrt{3}}{2}$.

Assignment 15.24 —

(a) $xe^{xy+y} = e^2\left(1+2(x+y-2)+\frac{3}{2}(x-1)^2+5(x-1)(y-1)+2(y-1)^2+\cdots\right)$

(b) $x\ln(y) = x(y-1)+\cdots$

(c) $xy+\ln(xy) = 1+2(x+y-2)-\frac{1}{2}\,(x-y)^2+\cdots$

(d) $x\sin(y) = xy+\cdots$

(e) $xy\cos(x+y) = \frac{\pi^2}{4}x-\frac{\pi}{2}x^2-\frac{\pi}{2}xy+\ldots$

Assignment 15.25 —

(a) Local minimum in $(2,-1)$.

(b) Local minimum in $(1,2)$.

(c) Local maximum in $(0,0)$.

(d) Local minima in $(1,1)$ en $(-1,-1)$.

Assignment 15.26 —

(a) no extrema, saddle point at $(0,0)$

(b) no extrema, saddle point at $(0,0)$

(c) minimum at $\left(\frac{1}{4}, -\frac{1}{8}\right)$, saddle point at $(1,0)$

(d) minimum at $(0,1)$, saddle points at $(\pm 1, 2)$

(e) minima at $\left(\sqrt{2}, -\sqrt{2}\right)$ and $\left(-\sqrt{2}, \sqrt{2}\right)$, no conclusion for $(0,0)$

(f) no extrema, no conclusion for $(0,0)$

(g) minimum at $(1,1)$

(h) saddle points at $\left(k\pi, (-1)^{k+1}\right)$ with $k \in \mathbb{Z}$

(i) saddle points at $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ and $\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$

(j) no conclusion for $(x,0)$ with $x \in \mathbb{R}$ and $(0,y)$ with $y \in \mathbb{R}$

(k) no conclusion for $(x, k\pi - x)$ with $x \in \mathbb{R}$ and $k \in \mathbb{Z}$

(l) Local minimum at $(2,-1)$.

(m) Saddle point at $(-1,1)$.

(n) Saddle point at $(0,0)$ and Local minimum at $(1,1)$.

(o) Saddle point at $(0,0)$ and Local minima at $(-1,-1)$ and $(1,1)$.

(p) Local maximum at $(-4,2)$.

(q) Saddle point at $(0, n\pi)$, with $n \in \mathbb{N}$.

(r) Saddle points at $(m\pi, n\pi)$ if $m+n$ is odd, Local minima at $(m\pi, n\pi)$ if m and n are odd, Local maxima at $(m\pi, n\pi)$ if m and n are even.

(s) Local minima at $(0,y)$ if $y > 0$ and at $\left(\pm 1, -1/\sqrt{2}\right)$, Local maxima at $(0,y)$ if $y < 0$ and at $\left(\pm 1, 1/\sqrt{2}\right)$, Saddle point at $(0,0)$.

(t) Saddle point at $(0,0)$, Local maxima at $(1,1)$ and $(-1,-1)$, Local minima at $(1,-1)$ and $(-1,1)$.

(u) Saddle point at $(3^{-1/3}, 0)$.

(v) Local minima at $(0,y)$ if $y \neq 0$ and Local maxima at $(x,0)$ if $x \neq 0$.

(w) Local minima at $(x,-x)$ if $x \neq 0$ and Local maxima at (x,x) if $x \neq 0$.

(x) Saddle point at $(1,1,1/2)$.

Assignment 15.27 — Maxima at $(1, \pm 1, 1)$ and minima at $(-1, \pm 1, -1)$.

Assignment 15.28 — $x = 1$ and $y = \sqrt{2}$

Assignment 15.29 — The box has a minimum area if the base of the box has dimensions 4 dm and 4 dm and the box has a height 2 dm.

Assignment 15.30 — The numbers for which ab^2c^3 is maximal, are $a = 5$, $b = 10$ and $c = 15$

Assignment 15.31 — The dimensions of the base and cover should be 2 m and 2 m, while the height of the box should be 3 m.

Assignment 15.32 —

$$\left(\frac{4}{3}, -\frac{4}{3}, -\frac{4}{3}\right)$$

Assignment 15.33 — The box has a maximum volume if the base of the box has dimensions 16 cm and 16 cm and the box has a height of 8 cm.

Assignment 15.34 —

(a) Global maximum: $f\left(\frac{1}{2},1\right)=\frac{5}{4}$. Global minimum: $f(2,0)=-2$.

(b) Global maximum: $f(-1,0)=2$. Global minimum: $f(1,0)=-2$.

(c) Global maximum: $f\left(\frac{\sqrt{3}}{3},1\right)=\frac{2\sqrt{3}}{9}$. Global minima: $f(x,0)=f(0,y)=f(1,1)=0$, with $0\leq x,y\leq 1$.

(d) Global maximum: $f(1,1)=3$. Global minimum: $f(1,-1)=-3$.

(e) Global maximum: $f(3,-2)=18$. Global minimum: $f(1,-1)=-3$.

(f) Global maximum: $f\left(\frac{1}{3},\frac{1}{3}\right)=\frac{1}{27}$. Global minima: $f(x,0)=f(0,y)=f(x,1-x)=0$, with $0\leq x,y\leq 1$.

(g) Global maximum: $f\left(\frac{\pi}{2},0\right)=1$. Global minima: $f\left(\frac{3\pi}{2},0\right)=f\left(\frac{\pi}{2},\pi\right)=-1$.

(h) Global maximum: $f(2,1)=\frac{4}{e^3}$. Global minima: $f(x,0)=f(0,y)=0$, with $0\leq x,y\leq 4$.

(i) Global maxima: $f\left(\frac{5}{8},\frac{3}{8}\right)=f\left(\frac{5}{8},-\frac{3}{8}\right)=\frac{23}{16}$. Global minimum: $f\left(-\frac{1}{2},0\right)=-\frac{1}{4}$.

Assignment 15.35 —

(a) Maximum: $f(0,3)=21$.

(b) Maximum: $f\left(\frac{7}{4},5\right)=\frac{37}{2}$.

Assignment 15.36 — Minimum: $f(1,1,1)=9$.

Assignment 15.37 — The manufacturer should produce 20000/3 kg ≈ 6667 kg of each textile for a maximum profit of \$100 000/3 ≈ \$33 333.

Assignment 15.38 — The profit is maximal (€66) with 6 circuits of type A and three circuits of type B.

Assignment 15.39 — There should be 4 hours of pop music, 4 hours of oldies and 12 hours of information programs to maximize the ratin, which then will be 160.

Assignment 15.40 — There should be 40 detached houses, no half open buildings and 40 apartments. The profit will then be 2 €240 000.

Assignment 15.41 — The profit is maximum for 10 suits, 30 jackets and 40 pants.

Assignment 15.42 — The profit is maximum for 75 kg of tea and 15 kg of sugar.

Assignment 15.43 — The cost is minimum for 300 Silver balls, 300 Yellow balls and 1200 Gold balls. The cost is then €2460.

Assignment 15.44 —

(a) $\nabla f(1,-2) = \left(\frac{2}{5}, -\frac{4}{5}\right) = \frac{2}{5}(\hat{\boldsymbol{\imath}} - 2\hat{\boldsymbol{\jmath}})$

(b) $2x - 4y - 5z + 5\ln 5 - 10 = 0$

(c) $x - 2y - 5 = 0$

Assignment 15.45 —

(a) The isotherms that are found by $T(x,y) = c > 0$ are hyperbolas with their real vertices on the x-axis. The isotherms found by $T(x,y) = c < 0$ are hyperbolas with their real vertices on the y-axis. The ones belonging to $T(x,y) = c = 0$ are the lines $x = \pm\sqrt{2}y$.

(b) $-\nabla T(2,-1) = (-4,-4) = -4\hat{\boldsymbol{\imath}} - 4\hat{\boldsymbol{\jmath}} \qquad \to$ direction $(-1,-1)$

(c) $yx^2 = -4$

Chapter 16

Assignment 16.1 —

(a) $\displaystyle\int_0^2 \int_x^{2x} \left(x^2 + y^2\right)\, dy\, dx = \frac{40}{3}$

(b) $\displaystyle\int_1^2 \int_y^{y^3} e^{\frac{x}{y}}\, dx\, dy = \frac{e}{2}(e^3 - 4)$

(c) $\displaystyle\int_1^2 \int_y^{3y} (x + y)\, dx\, dy = 14$

(d) $\displaystyle\int_{-1}^2 \int_{2x^2-2}^{x^2+x} x\, dy\, dx = \frac{9}{4}$

(e) $\displaystyle\int_0^{\pi} \int_0^{\cos(\theta)} r\sin(\theta)\, dr\, d\theta = \frac{1}{3}$

(f) $\displaystyle\int_0^{\frac{\pi}{2}} \int_2^{4\cos(\theta)} r^3\, dr\, d\theta = 10\pi$

(g) $\displaystyle\int_0^1 \int_0^1 \frac{x^2}{1 + y^2}\, dy\, dx = \frac{\pi}{12}$

(h) $\displaystyle\int_{-1}^2 \int_0^1 (xy^2 + x^2y)\, dx\, dy = 2$

(i) $\displaystyle\int_0^1 \int_{x^2}^{2-x^2} \sqrt{x}y\, dy\, dx = \frac{16}{21}$

(j) $\displaystyle\int_0^{\pi/4} \int_{\sin(x)}^{\cos(x)} xy\, dy\, dx = \frac{\pi - 2}{16}$

(k) $\displaystyle\int_0^1 \int_{\sqrt[3]{y}}^1 e^{x^2}\, dx\, dy = \int_0^1 \int_0^{x^3} e^{x^2}\, dy\, dx = \frac{1}{2}$

Assignment 16.2 —

(a) $$\iint_R x^2 y \, dA = \int_0^3 \int_0^{\frac{2}{3}x} x^2 y \, dy \, dx = \int_0^2 \int_{\frac{3}{2}y}^3 x^2 y \, dx \, dy = \frac{54}{5}$$

(b) $$\iint_R dA = \int_0^1 \int_{\sqrt{x^3}}^{x} dy \, dx = \int_0^1 \int_y^{\sqrt[3]{x^2}} dx \, dy = \frac{1}{10}$$

(c) $$\iint_R x^2 \, dA = \int_0^4 \int_0^x x^2 \, dy \, dx + \int_4^8 \int_0^{\frac{16}{x}} x^2 \, dy \, dx = \int_0^2 \int_y^8 x^2 \, dx \, dy + \int_2^4 \int_y^{\frac{16}{y}} x^2 \, dx \, dy = 448$$

(d) $$\iint_R y \, dA = \int_0^1 \int_{x^3}^{x^2} y \, dy \, dx = \int_0^1 \int_{\sqrt{y}}^{\sqrt[3]{y}} y \, dx \, dy = \frac{1}{35}$$

(e) $$\iint_R \frac{1}{\sqrt{2y-y^2}} \, dA = \int_0^2 \int_0^{\frac{4-x^2}{2}} \frac{1}{\sqrt{2y-y^2}} \, dy \, dx = \int_0^2 \int_0^{\sqrt{4-2y}} \frac{1}{\sqrt{2y-y^2}} \, dx \, dy = 4$$

(f) $$\iint_R e^{\frac{x}{y}} \, dA = \int_0^1 \int_{\sqrt{x}}^1 e^{\frac{x}{y}} \, dy \, dx = \int_0^1 \int_0^{y^2} e^{\frac{x}{y}} \, dx \, dy = \frac{1}{2}$$

(g) $$\iint_R (y-x) \, dA = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\cos(\theta)} r^2(\sin(\theta) - \cos(\theta)) \, dr \, d\theta = -\pi$$

Assignment 16.3 —

(a) $$\int_0^3 \int_1^{\sqrt{4-y}} f(x,y) \, dx \, dy = \int_1^2 \int_0^{4-x^2} f(x,y) \, dy \, dx$$

(b) $$\int_0^1 \int_{\arccos(y)}^{\frac{\pi}{2}} f(x,y) \, dx \, dy = \int_0^{\frac{\pi}{2}} \int_{\cos(x)}^1 f(x,y) \, dy \, dx$$

(c) $$\int_{-6}^2 \int_{\frac{x^2}{4}}^{3-x} f(x,y) \, dy \, dx = \int_0^1 \int_{-2\sqrt{y}}^{2\sqrt{y}} f(x,y) \, dx \, dy + \int_1^9 \int_{-2\sqrt{y}}^{3-y} f(x,y) \, dx \, dy$$

Assignment 16.4 —

(a) $$A = 22 \int_0^{\sqrt{15}} \int_{\frac{25-y^2}{-10}}^{\frac{y^2-9}{-6}} dx \, dy = \frac{16\sqrt{15}}{3}$$

(b) $A = \int_0^1 \int_{y^{\frac{1}{3}}}^{2-y} dx\ dy = \int_0^1 \int_0^{x^3} dy\ dx + \int_1^2 \int_0^{2-x} dy\ dx = \frac{3}{4}$

(c) $A = \int_1^2 \int_{\sqrt{1-(x-1)^2}}^{x} dy\ dx + \int_2^4 \int_0^{\sqrt{4-(x-2)^2}} dy\ dx = \int_0^1 \int_{1+\sqrt{1-y^2}}^{2+\sqrt{4-y^2}} dx\ dy + \int_1^2 \int_y^{2+\sqrt{4-y^2}} dx\ dy$

$= \int_0^{\frac{\pi}{4}} \int_{2\cos(\theta)}^{4\cos(\theta)} r\ dr\ d\theta = \frac{3\pi}{4} + \frac{3}{2}$

(d) $A = 2\int_0^{\frac{\pi}{3}} \int_{\frac{1}{\cos(\theta)}}^{2} r\ dr\ d\theta = \frac{4\pi}{3} - \sqrt{3}$

(e) $A = 2\int_0^{\pi} \int_0^{1+\cos(\theta)} r\ dr\ d\theta - 2\int_0^{\frac{\pi}{2}} \int_0^{\cos(\theta)} r\ dr\ d\theta = \frac{5\pi}{4}$

Assignment 16.5 —

(a) $V = \int_0^4 \int_0^4 \frac{xy}{4}\ dy\ dx = 16$

(b) $V = \int_0^2 \int_0^{\frac{\sqrt{4-x^2}}{2}} \frac{4-y^2}{2}\ dy\ dx = \int_0^1 \int_0^{2\sqrt{1-y^2}} \frac{4-y^2}{2}\ dx\ dy = \frac{15\pi}{16}$

(c) $V = \int_0^1 \int_0^{1-x} (1-x^2-y^2)\ dy\ dx = \int_0^1 \int_0^{1-y} (1-x^2-y^2)\ dx\ dy = \frac{1}{3}$

(d) $V = \int_0^1 \int_{x^2}^{\sqrt{x}} (12+y-x^2)\ dy\ dx = \int_0^1 \int_{y^2}^{\sqrt{y}} (12+y-x^2)\ dx\ dy = \frac{569}{140}$

(e) $V = \int_0^2 \int_0^1 \frac{4+x-y}{2}\ dy\ dx = \int_0^1 \int_0^2 \frac{4+x-y}{2}\ dx\ dy = \frac{9}{2}$

(f) $V = 8\int_0^1 \int_0^{\sqrt{1-x^2}} \sqrt{1-x^2}\ dy\ dx = 8\int_0^1 \int_0^{\sqrt{1-y^2}} \sqrt{1-x^2}\ dx\ dy = \frac{16}{3}$

Assignment 16.6 —

(a) $\int_0^4 \int_0^{\sqrt{16-x^2}} (x+y+2)\ dy\ dx = \int_0^{\frac{\pi}{2}} \int_0^4 (r\cos(\theta) + r\sin(\theta) + 2)\, r\ dr\ d\theta$

(b) $\displaystyle\int_{-1}^{0}\int_{-\sqrt{-x^2-x}}^{0} 2xy\;dy\;dx = \int_{\pi}^{\frac{3\pi}{2}}\int_{0}^{-\cos(\theta)} 2r^3\cos(\theta)\sin(\theta)\;dr\;d\theta$

Assignment 16.7 —

(a) $\displaystyle V = \int_0^{2\pi}\int_0^{\sqrt{2}} \left(2-r^2\right) r\;dr\;d\theta = 2\pi$

(b) $\displaystyle V = \int_0^{2\pi}\int_0^{\sqrt{2}} r^3\;dr\;d\theta = 2\pi$

(c) $\displaystyle V = 8\int_0^{\frac{\pi}{2}}\int_0^{1} \sqrt{4-r^2}\,r\;dr\;d\theta = \frac{4\pi}{3}\left(8-3\sqrt{3}\right)$

(d) $\displaystyle V = \int_0^{\sqrt{2}}\int_0^{2\pi} \left(\sqrt{3-r^2}-\frac{r^2}{2}\right) r\;d\theta\;dr = \frac{\pi}{3}\left(6\sqrt{3}-5\right)$

(e) $\displaystyle V = 4\int_0^{\frac{\pi}{2}}\int_0^{a\cos(\theta)} \sqrt{a^2-r^2}\,r\;dr\;d\theta = \frac{4}{3}a^3\left(-\frac{2}{3}+\frac{\pi}{2}\right)$

(f) $\displaystyle V = 4\int_0^{\frac{\pi}{2}}\int_0^{2\sin(\theta)} \sqrt{r\sin(\theta)}\,r\;dr\;d\theta = \frac{64\sqrt{2}}{15}$

Assignment 16.8 —

(a)
- $\displaystyle M = \int_0^{\pi}\int_0^{\sin(x)} ky\;dy\;dx = \frac{k\pi}{4}$
- $\displaystyle M_y = k\int_0^{\pi}\int_0^{\sin(x)} xy\;dy\;dx = \frac{k\pi^2}{8} \quad\Rightarrow \overline{x} = \frac{\pi}{2}$
- $\displaystyle M_x = k\int_0^{\pi}\int_0^{\sin(x)} y^2\;dy\;dx = \frac{4k}{9} \quad\Rightarrow \overline{y} = \frac{16}{9\pi}$

Mass density: $\displaystyle\left(\frac{\pi}{2},\frac{16}{9\pi}\right)$

(b)
- $\displaystyle M = 2\int_0^{2}\int_{\frac{y^2-4}{4}}^{\frac{y^2-4}{-2}} 1\;dx\;dy = 8$
- $\displaystyle M_y = 2\int_0^{2}\int_{\frac{y^2-4}{4}}^{\frac{y^2-4}{-2}} x\;dx\;dy = \frac{16}{5} \quad\Rightarrow \overline{x} = \frac{2}{5}$

- $M_x = 2\int_0^2 \int_{\frac{y^2-4}{4}}^{\frac{y^2-4}{-2}} y\,dx\,dy = 0 \quad \Rightarrow \overline{y} = 0$

Mass density: $\left(\frac{2}{5}, 0\right)$

(c)
- $M = \int_0^4 \int_0^{\frac{12-3y}{2}} 1\,dx\,dy = 12$
- $M_y = \int_0^4 \int_0^{\frac{12-3y}{2}} x\,dx\,dy = 24 \quad \Rightarrow \overline{x} = 2$
- $M_x = \int_0^4 \int_0^{\frac{12-3y}{2}} y\,dx\,dy = 16 \quad \Rightarrow \overline{y} = \frac{4}{3}$

Mass density: $\left(2, \frac{4}{3}\right)$

Assignment 16.9 —

(a) $SA = 2\int_0^{\frac{\pi}{2}} \int_0^{4\sin(\theta)} \frac{2}{\sqrt{3}} r\,dr\,d\theta = \frac{8\pi}{\sqrt{3}}$

(b) $SA = 8\int_0^4 \int_0^{\sqrt{16-x^2}} \sqrt{\frac{16}{16-x^2}}\,dy\,dx = 128$

(c) $SA = 4\int_0^1 \int_0^{2\sqrt{x}} \sqrt{\frac{x+1}{x}}\,dy\,dx = \frac{16}{3}\left(2\sqrt{2}-1\right)$

(d) $SA = \int_0^{2\pi} \int_0^1 \sqrt{1+4r^2}\,r\,dr\,d\theta = \frac{\pi}{6}\left(5\sqrt{5}-1\right)$

(e) $SA = 8\int_0^1 \int_0^{\frac{\sqrt{1-x^2}}{2}} \sqrt{\frac{1}{1-x^2-y^2}}\,dy\,dx = \frac{4\pi}{3}$

Assignment 16.10 —

(a) $\int_C x^2\,ds = 3\sqrt{14}$

(b) $\int_C y\,ds = 156$

(c) $\displaystyle\int_C (x+y)\, ds = 2\sqrt{2}$

(d) $\displaystyle\int_C \frac{ds}{x^2+y^2+z^2} = \frac{\sqrt{65}}{8} \arctan\left(\frac{\pi}{4}\right)$

(e) $\displaystyle\int_C \sqrt{2y^2+z^2}\, ds = 8\pi$

(f) $\displaystyle\int_C e^z\, ds = \frac{e^{2\pi}\sqrt{1+2e^{4\pi}} - \sqrt{3}}{2} + \frac{1}{2\sqrt{2}} \ln\left(\frac{\sqrt{2}e^{2\pi} + \sqrt{1+2e^{4\pi}}}{\sqrt{2}+\sqrt{3}}\right)$

(g) $\displaystyle\int_C \sqrt{1+4x^2z^2}\, ds = 3\pi$

(h) $\displaystyle\int_C x\, ds = \frac{a^2}{2}\left(\sqrt{2} + \ln\left(1+\sqrt{2}\right)\right)$

(i) $\displaystyle\int_C z\, ds = 1$

Assignment 16.11 —

(a) $\vec{\boldsymbol{F}}_1 = \dfrac{\vec{\boldsymbol{r}}}{\|\vec{\boldsymbol{r}}\|}$ $\quad\to$ graph *II*

(b) $\vec{\boldsymbol{F}}_2 = \vec{\boldsymbol{r}}$ $\quad\to$ graph *I*

(c) $\vec{\boldsymbol{F}}_3 = y\hat{\boldsymbol{\imath}} - x\hat{\boldsymbol{\jmath}}$ $\quad\to$ graph *IV*

(d) $\vec{\boldsymbol{F}}_4 = x\hat{\boldsymbol{\jmath}}$ $\quad\to$ graph *III*

Assignment 16.12 — $\nabla f = (2xyz^3, x^2z^3, 3x^2yz^2)$, $\nabla\cdot\vec{\boldsymbol{F}} = z-2y$ and $\nabla\times\vec{\boldsymbol{F}} = (2x^2, -4xy+x, 0)$

Assignment 16.13 —

$\nabla f = (y+z, x+z, y+x) \quad\Rightarrow\quad \nabla f(3,-1,2) = (1,5,2),\ \ \nabla\cdot\vec{\boldsymbol{F}} = 0,$

$\nabla\times\vec{\boldsymbol{F}} = (y^2, -z^2, -x^2) \quad\Rightarrow\quad (\nabla\times\vec{\boldsymbol{F}})(3,-1,2) = (1,-4,-9)$ and

$(\vec{\nabla} f)\times\vec{\boldsymbol{F}} = (x^2z^2+xz^3+y^3z+xy^2z, -xyz^2-xz^3+x^2y^2+x^3y, -y^3z-y^2z^2-x^3y-x^2yz)$

$\Rightarrow\quad ((\vec{\nabla} f)\times\vec{\boldsymbol{F}})(3,-1,2) = (64,-30,43)$

Assignment 16.14 —

(a) The direction of the vector field changes in all points.

(b) In points on the y-axis or the z-axis, the direction of the vector field does not change.

Assignment 16.15 —

$\nabla\times\vec{\boldsymbol{F}} = \left(7z, -\dfrac{3x}{z^2}, 2\right) \quad\Rightarrow$ on the curve C: $\nabla\times\vec{\boldsymbol{F}} = \big(-7\cos(t), -3, 2\big)$

The length of the rotor is maximal if $\cos(t) = \pm 1$. This corresponds to the points $(1,0,\pm 1)$ on C. The length of the rotor is minimal if $\cos(t) = 0$. This corresponds to the points $(0,\pm 1,0)$ on C.

Assignment 16.16 —

Imagine $\overrightarrow{OP} = (x, y, z)$ with $r = \sqrt{x^2 + y^2 + z^2}$, then $V = \dfrac{q}{4\pi\epsilon_0\sqrt{x^2+y^2+z^2}}$. We then obtain that

$$\vec{\boldsymbol{E}} = -\nabla V = -\frac{q}{4\pi\epsilon_0\left(\sqrt{x^2+y^2+z^2}\right)^3}(x, y, z) = \frac{q\vec{\boldsymbol{r}}}{4\pi\epsilon_0\, r^3}$$

Assignment 16.17 —

(a) $\displaystyle\int_C \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = 0$

(b) $\displaystyle\int_{C_1} \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = \frac{1}{3}, \quad \int_{C_2} \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = \frac{1}{12}, \quad \int_{C_3} \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = \frac{17}{30}$

(c) $\displaystyle\int_{C_i} \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = 2 \text{ for } i = 1\dots 4$

(d) $\displaystyle\int_C \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = 2\pi$

(e) $\displaystyle\int_C \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = \frac{76}{35}$

(f) $\displaystyle\int_C \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = 0$

(g) $\displaystyle\int_C \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = 5$

(h) $\displaystyle\int_{C_1} \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = \frac{49}{3}, \quad \int_{C_2} \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = \frac{49}{3}, \quad \int_{C_3} \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = 0$

(i) $\displaystyle\int_C \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{r}} = \frac{32}{3}$

Assignment 16.18 —

(a) $R = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x > 0,\ y \geq 0\right\}$ is a simply connected region

(b) $R = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x = 0,\ y \geq 0\right\}$ is not a region.

(c) $R = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x \neq 0,\ y > 0\right\}$ is a region but not connected. There is no path in R from $(-1, 1)$ to $(1, 1)$.

(d) $R = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 > 1\right\}$ is a region but not connected. There is no path in R from $(-2, 0, 0)$ to $(2, 0, 0)$.

(e) $R = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 > 1\right\}$ is a connected region. The circle $x^2 + y^2 = 2, z = 0$ lies in R, but cannot be shrunk to a point while staying the region.

(f) $R = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x^2 + y^2 + z^2 > 1\right\}$ is a single coherent area.

Assignment 16.19 —

(a) $\vec{\boldsymbol{F}} = \left(xy, \dfrac{1}{2}x^2 - y^2\right)$ is conservative. A potential function is $f(x,y) = \dfrac{x^2 y}{2} - \dfrac{y^3}{3}$.

(b) $\vec{\boldsymbol{F}} = (y, x, -2z)$ is conservative. A potential function is $f(x,y,z) = xy - z^2$.

(c) $\vec{\boldsymbol{F}} = \left(\dfrac{x}{x^2+y^2}, \dfrac{y}{x^2+y^2}\right)$ is conservative. A potential function is $f(x,y) = \dfrac{\ln\left(x^2+y^2\right)}{2}$.

(d) $\vec{\boldsymbol{F}} = \left(\dfrac{x}{x^2+y^2}, -\dfrac{y}{x^2+y^2}\right)$ is not conservative.

(e) $\vec{\boldsymbol{F}} = \left(2xy - z^2, 2yz + x^2, -2zx + y^2\right)$ is conservative. A potential function is $f(x,y,z) = x^2y - xz^2 + y^2 z$.

(f) $\vec{\boldsymbol{F}} = e^{x^2+y^2+z^2}\,(xz, yz, xy)$ is not conservative.

(g) $\vec{\boldsymbol{F}} = \left(xy - \sin(z), \dfrac{1}{2}x^2 - \dfrac{e^y}{z}, \dfrac{e^y}{z^2} - x\cos(z)\right)$ is conservative. A potential function is $f(x,y,z) = \dfrac{x^2 y}{2} - \dfrac{e^y}{z} - x\sin(z)$.

Chapter 17

Assignment 17.1 — By rewriting the differential equations, we conclude:

(a) first order, first degree

(b) third order, second degree

(c) second order, first degree

(d) first order, fourth degree

Assignment 17.2 — The families of curves can be rewritten as follows:

(a) $y = C_3 e^t$ with $C_3 = C_1 e^{C_2}$

(b) $y = C_3 + \ln|t|$ with $C_3 = C_1 + \ln|C_2|$

Assignment 17.3 — To show that $y(t)$ is a solution of the given differential equation, it suffices to differentiate $y(t)$, put it in the differential equation, and note that this yields zero. To determine the particular solution on the basis of the initial conditions, it suffices to insert it into the differential equation.

(a) $y(t) = t^3 + 3e^{-t}$

(b) $y(t) = te^t$

(c) $y(t) = 1 - t$

(d) $y(t) = 3e^{-t} + 4t - 1$

(e) $y(t) = 2t + 3e^t$

Assignment 17.4 — The differential equations describing the problems are given by:

(a) $V'(t) = k\sqrt{V(t)}$,

(b) $A'(t) = kA(t)^2$,

(c) $T'(t) = k\big(T(t) - R\big)$,

(d) $M'(t) = a - vM(t)$.

Assignment 17.5 — We only give the instructions for e).

(a) $y'' + 2y' + 5y = 0, \quad$ if $\quad y(0) = 2 \quad$ en $\quad y'(0) = 2.$

Determine solution with mathematica:

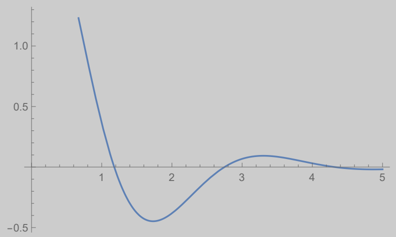
```
In[39]:= opl = DSolve[{y''[t] + 2*y'[t] + 5*y[t] == 0, y[0] == 2, y'[0] == 2},
         y[t], t]

Out[39]= {{v[t]→ 2e^-t(Cos[2t] + Sin[2t])}}
```

Plot solution with mathematica:

```
In[40]:= Plot[y[t]/.opl, {t, 0, 5}]
```

Out[40]=



Assignment 17.6 —

(a) If the function `verhulst` was used correctly, you get the vectors below N and T.

```
>>> N
array([ 2.        ,  2.97      ,  4.38884325,  6.43880029,  9.34726431,
       13.36561134, 18.70862026, 25.43783685, 33.3036287 , 41.63695542,
       49.4531627 , 55.8376293 , 60.3726376 , 63.22254112, 64.85563889,
       65.73655412, 66.19512207, 66.42922671, 66.54752386, 66.6069888 ,
       66.63680102, 66.65172715, 66.65919524, 66.66293053, 66.6647985 ,
       66.66573255])

>>> T
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
       17, 18, 19, 20, 21, 22, 23, 24, 25])
```

(b) The number of individuals N converges to about 67.

(c) $N0$ has **no** influence on the final number of individuals.

(d) If $r = 0.5$ the population becomes extinct. If r increases, the final number of individuals increases. From $r = 2.5$, the gradient shows oscillations that increase as r increases. With

$r = 3.0$ and $r = 3.5$ it is therefore not clear how large the final number of individuals will be. Figure **??** summarises this.

Assignment 17.7 —

(a) THe function `verhulst2` is a simple extension of the function `verhulst`.

(b) If the function was written correctly, you get the following vectors N and T.

```
>>> N
array([ 2.        ,  2.72      ,  3.684512  ,  4.96438678,  6.64119331,
        8.80084993, 11.52025646, 14.84498032, 18.75904713, 23.15442594,
       27.81637986, 32.4443899 , 36.71624784, 40.37421936, 43.28896971,
       45.46782161, 47.0133337 , 48.06643224, 48.76348002, 49.21570765,
       49.50520392, 49.68891628, 49.80484688, 49.87774366, 49.92347769,
       49.95212964])

>>> T
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
       17, 18, 19, 20, 21, 22, 23, 24, 25])
```

(c) The number of individuals N converges to about 50.

(d) If $h = 0.25$ the final number of individuals drops to about 32. If $h = 0.75$ the population dies out over time.

(e) The observations are similar to those from the previous assignment. Figure **??** summarises this.

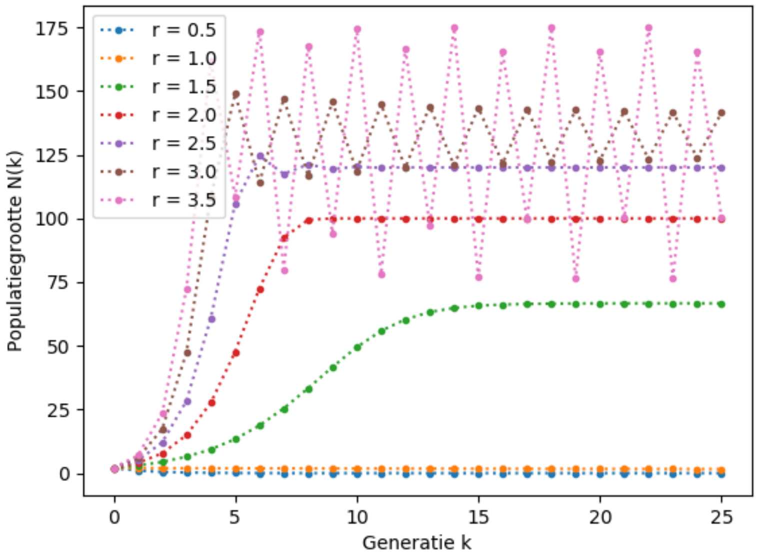


Figure 17.7: Evolution of the population size N as a function of r.

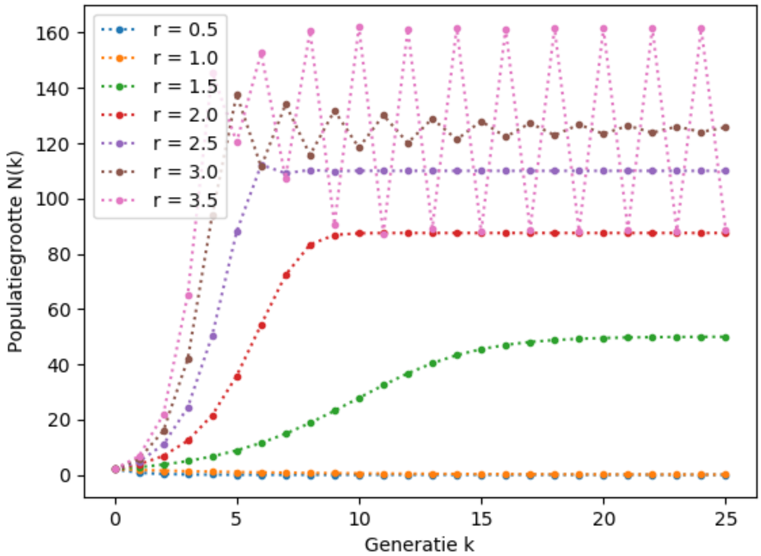


Figure 17.8: Evolution of the population size N as a function of r.

Chapter 18

Assignment 18.1 —

(a) The exact solution we find with Mathematica is $y(t) = 1.04081e^{-t^2}t$. If however you use 1/5 instead of 0.2, you get: $y(t) = e^{1/25-t^2}t$.

(b) The directional field of the given differential equation can be seen in **??**

(c) With the instructions below we get Figure **??**.

```
direction field(function1, [0.2, 3], [0, 1], 25)
t = arange(0.2, 3, 0.1)
y = t*exp(1/25 - t**2)
plt.plot(t, y, 'r.')
```

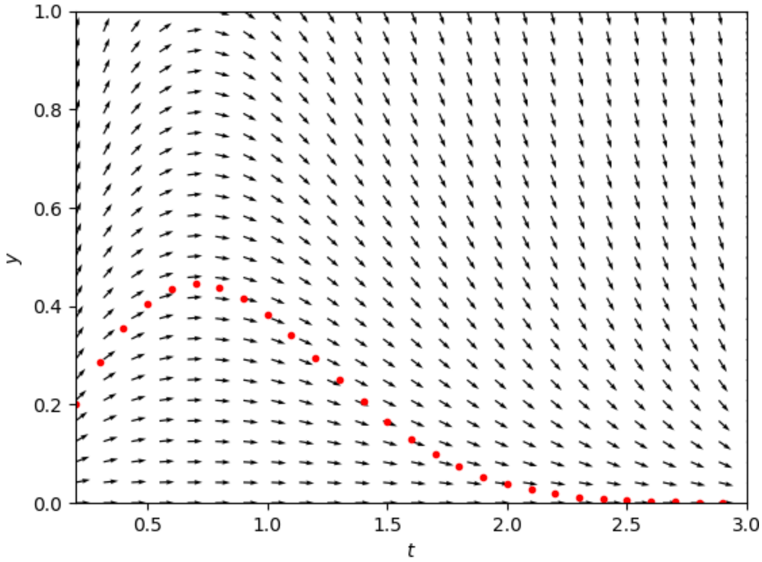


Figure 18.13: Direction field and exact solution of $y' = y(-2t + 1/t)$.

Assignment 18.2 —

(a) /

(b) The direction field of the given differential equation can be seen in **??**.

(c) When the precision n increases, more direction vectors are shown in the direction field.

(d) You can predict the global course of the exact solution by following the direction vectors from the point $(-2, 0.5)$.

(e) The general solution we determine with Mathematica, is $y(t) = 3+3t+3t^2+t^3+e^tC[1]$

(f) The exact solution through the point $(-2, 0.5)$ is (after simplification with `Simplify`) $y(t) = 3 - \frac{e^{2+t}}{2} + 3t + 3t^2 + t^3$. After implementation in Python as `exact2` and plotting on top of the directional field, you get Figure **??**.

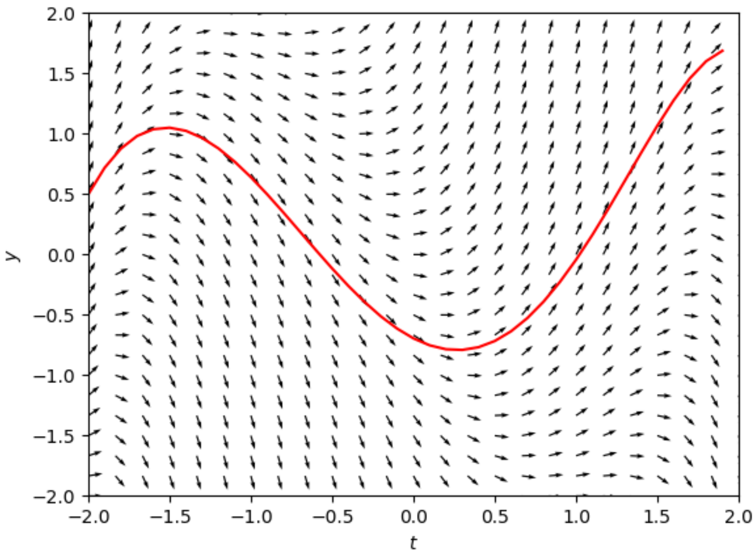


Figure 18.14: Direction field and exact solution of $y' = y + 3t - t^3$.

Assignment 18.3 —

(a) III

(b) I

(c) II

(d) IV

Assignment 18.4 —

(a) $y_e = -1$ and $y_e = 4$

(b) $y_e = 0$ and $y_e = 6$

(c) $y_e = -2$ and $y_e = 3$

(d) $y_e = 0$ and $y_e = \pm 2$

(e) $y_e = 0$

(f) $N_e = 0$ and $N_e = (p-r)\frac{K}{r}$

(g) no equilibrium points

Checking the results for the third differentail equation:

```
In[44]:=   Solve[{0 == y^2 - y - 6 }, y]

Out[44]=   {{y → -2}, {y → 3}}
```

Assignment 18.5 — In the following, AS means asymptotically stable and US means unstable.

(a) $y_e = -1 \rightarrow$ AS and $y_e = 4 \rightarrow$ US

(b) $y_e = 0 \rightarrow$ US and $y_e = 6 \rightarrow$ AS

(c) $y_e = -2 \rightarrow$ AS and $y_e = 3 \rightarrow$ US

(d) $y_e = 0 \rightarrow$ AS and $y_e = \pm 2 \rightarrow$ US

(e) $y_e = 0 \rightarrow$ US

(f) $N_e = 0 \rightarrow$ US and $N_e = (p-r)\dfrac{K}{r} \rightarrow$ US

(g) no equilibrium points

Assignment 18.6 —

(a) $v_e = \dfrac{mg}{\mu}$

(b) AS

Assignment 18.7 —

(a) $v_e = \dfrac{mg}{\mu}$

(b) To determine the stability, we apply the first derivative test. Deriving the right-hand side yields $2kx - k(p+q)$. If we use $x_e = p$ and $x_e = q$, we achieve $k(p-q)$ and $k(q-p)$. We now distinguish two cases.

1) If $p > q$, then: $x_e = p \rightarrow$ US and $x_e = q \rightarrow$ AS.
2) If $p < q$, then: $x_e = p \rightarrow$ AS and $x_e = q \rightarrow$ US.

Assignment 18.8 —

(a) $C_e = \dfrac{qVC_{in}}{q + rV}$

(b) Om de stabiliteit te bepalen, passen we de eerste afgeleide test toe. Het afleiden van het rechterlid levert $-(q/V + r)$ op. Aangezien q, r en V steeds positief zijn, is deze uitdrukking steeds negatief. Bijgevolg is het evenwichtspunt AS.

Assignment 18.9 —

(a) With $r = 2$ and $K = 200$, we get $P_e = 0$, $P_e = \frac{1}{2}(199 - \sqrt{40401 - 400h})$ and $P_e = \frac{1}{2}(199 + \sqrt{40401 - 400h})$.

(b) If $h = \frac{40401}{400}$ (≈ 101) there are two equilibrium points: $P_e = 0$ and $P_e = 99.5$. If we plot the direction field with $r = 2$, $K = 200$ and $h = 40401/400$, we conclude that $P_e = 0$ is AS and $P_e = 99.5$ is US.
If h is bigger than 40401/400, there is one real equilibrium point ($P_e = 0$ AS) and two complex conjugate numbers (which we will not take further into account).
If h was smaller than 40401/400, for example $h = 1$, there are three equilibrium points, being: $P_e = 0$, $P_e = -0.5013$ and $P_e = 199.501$. Now $P_e = 0$ turns out to be US, the other two are AS.

(c) The bifurcation point is $h = \frac{40401}{400}$. The bifurcation diagram is given in **??**. A full line means AS, a dotted line means US.

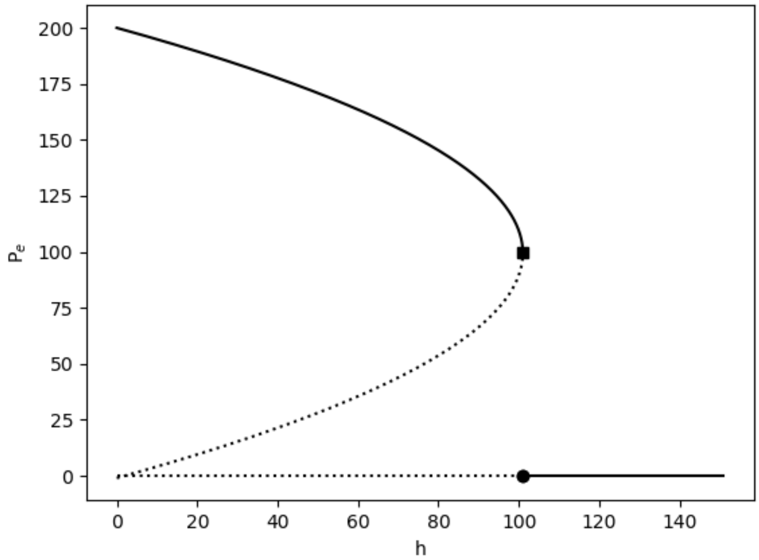


Figure 18.17: Stability of the equilibrium points of $P' = r\left(1 - \frac{P}{K}\right)P - h\frac{P}{1+P}$.

We have a pitchfork bifurcation since to the right of $h = \dfrac{40401}{400}$ there is but one real equilibrium point and to the left, there are three.

Assignment 18.10 —

(a) With $r = 1.5$ and $K = 500$, we achieve $P_e = 0$, $P_e = 500$ and $P_e = \alpha$ with $0 < \alpha < K$. There are thus three equilibrium points.

(b) The equilibrium points $P_e = 0$ and $P_e = 500$ are AS. The equilibrium points $P_e = \alpha$ is for all $0 < \alpha < K$ US (Figure **??**).

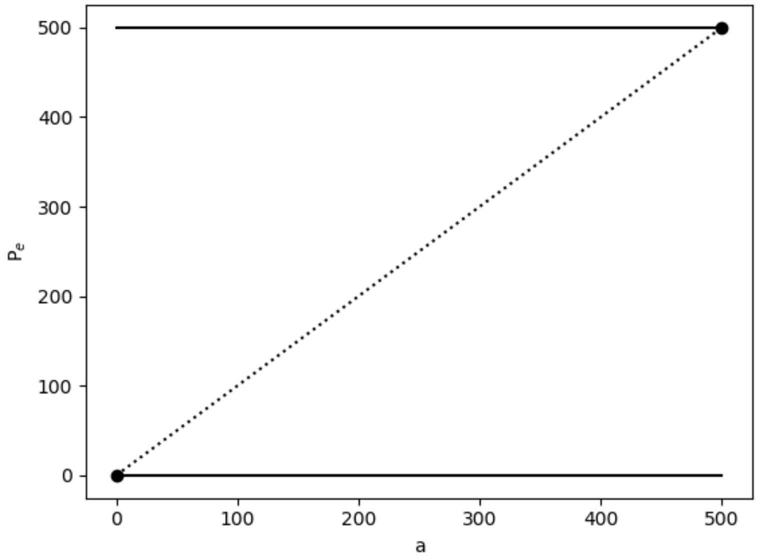


Figure 18.18: Stability of the equilibrium points of $P' = r\,(P-\alpha)\left(1-\dfrac{P}{K}\right)P$.

(c) It is not possible to determine the type of bifurcation since there is no bifurcation point. The number and/or stability of the equilibrium points does not change or depend on the parameter α (Figure **??**).

Assignment 18.11 —

(a) The two equilibrium points are given by: $v_e = \pm \dfrac{27.125}{\sqrt{k}}$. From this it follows that $k > 0$.

(b) With the first derivative test we can show that $P_e = +\sqrt{\dfrac{mg}{k}}$ is AS and $P_e = -\sqrt{\dfrac{mg}{k}}$ is US.

(c) It is not possible to determine the type of bifurcation since there is no bifurcation point. The number and/or stability of the equilibrium points does not change or depend on the parameter k (Figure **??**).

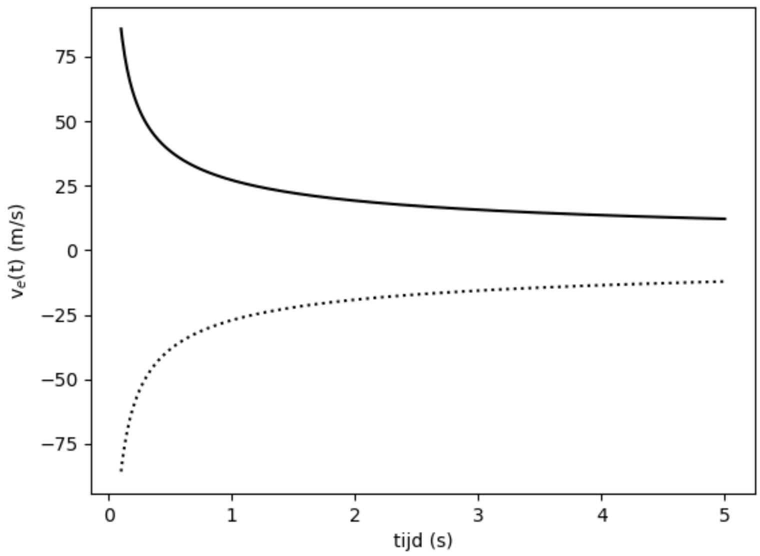


Figure 18.19: Stability of the equilibrium points of $m\dfrac{dv}{dt} = m\,g - k\,v^2$.

Chapter 19

Assignment 19.1 —

(a) $y(t) = -3 + C\,e^{\frac{1}{t}}$, solution intervals: $t \in]-\infty, 0[$ of $t \in]0, +\infty[$

(b) $y(t) = \tan(t + C)$, solution interval: $t \in]-\infty, +\infty[$ en $y \in]-\infty, +\infty[$

(c) $y(t) = \pm \dfrac{t}{\sqrt{Ct^2 - 1}}$, solution intervals: $t \in]-\infty, 0[$ of $t \in]0, +\infty[$ en $y \in]-\infty, +\infty[$

(d) $y(t) = -\ln\left|e^{-t} - C\right|$, solution interval: $t \in]-\infty, +\infty[$ en $y \in]-\infty, +\infty[$

(e) $\dfrac{1+y}{e^y} = \dfrac{1}{t} + C$, solution intervals: $t \in]-\infty, 0[$ of $t \in]0, +\infty[$ en $y \in]-\infty, 0[$ of $y \in]0, +\infty[$

Assignment 19.2 —

(a) $t^2 y + t^3 y^2 = C$

(b) $\dfrac{t^2}{2} + 3yt - 3\dfrac{y^2}{2} = C$

(c) $ty + \cos(y) = C$

(d) $t + y + e^{ty} = C$

Assignment 19.3 —

(a) $a = 4$

(b) $y(t) = \left(\dfrac{2\left(C - t^3\right)}{1 + \sin(2t)}\right)^{1/4}$

Assignment 19.4 —

(a) $a = 0$

(b) $y(t) = \dfrac{1}{-\cos(t) + 2t}$

(c) $]-\infty, 0.450184[\times]-\infty, +\infty[$ with 0.450184 the solution of $\cos(t) = 2t$ (determined in Mathematica).

Assignment 19.5 —

(a) $y(t) = -\dfrac{1}{4} + \dfrac{3t}{2} + Ce^{-2t}$

(b) $y(t) = 2 - t^2 + C\sqrt{1 - t^2}$

(c) $y(t) = C|t| - e^t$

(d) $y(t) = 1 + Ce^{-\arcsin(t)}$

Assignment 19.6 —

(a) $\ln|y(t)| - y = -\ln|t| - t + C$

(b) $t\sin(y(t)) - y(t)^2 + \frac{t^2}{2} = C$

(c) $y(t) = \frac{C}{1+e^{3t}}$

(d) $y(t) = Ce^t + \cos(t) - t\cos(t) + t\sin(t)$

(e) $y(t) = Ce^{\frac{t^2}{2}} - 2 - t^2$

(f) $y(t) = \pm\sqrt{\frac{2}{3}t^3 + 4t + C}$

(g) $y(t)^3 - 3y(t) = 6t + C$

(h) $y(t) = -\frac{3}{2} + C(1+t)^2$

(i) $y(t) = \frac{C}{|1+3t^2|^{1/3}}$

(j) $y(t) = \frac{\arctan(t)}{t} + \frac{C}{t}$

(k) $y(t) = C|\sin(t)| + \sin(t)\Big(\ln\big|\sin(t)\big| - \ln\big|\cos(t)\big|\Big)$

(l) $y(t) = \arccos\Big(C\big(1+e^t\big)\Big)$

(m) $y(t) = \frac{C+\sin(t)}{1+t}$

Assignment 19.7 —

(a) Solution intervals: $\mathbb{R} \setminus \{k\pi, k \in \mathbb{Z}\}$

(b) $y(t) = \sin(t) + \frac{1}{\sin(t)}$

Assignment 19.8 —

(a) solution interval: $t \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$

(b) $y(t) = \cos(t)\big(\sin(t) - 1\big)$

(c) $y' = 0 \Leftrightarrow y_e = \frac{\cos^3(t)}{\sin(t)}$. This is not a constant value for y_e. There are no equilibrium points.

Assignment 19.9 — First, we solve the differential equation $y' = \frac{y^2}{3}$ for $t \leq 2$ with $y(0) = 1$. We arrive at the exact solution $y(t) = \frac{3}{3-t}$ if $t \leq 2$. If we use $t = 2$, then we get $y(2) = 3$. We now use this as the initial condition for the differential equation $y'(t) = \cos^2(t)$. We arrive at the exact solution $y(t) = \frac{1}{4}\Big(8 + 2t - \sin(4) + \sin(2t)\Big)$ if $t > 2$.

Concluded:

$$y(t) = \begin{cases} \frac{3}{3-t}, & \text{if } t \leq 2, \\ \frac{1}{4}\Big(8 + 2t - \sin(4) + \sin(2t)\Big), & \text{als } t > 2. \end{cases}$$

Assignment 19.10 —

(a) $C(t) = \frac{qC_{in}}{q+rV} + Ce^{-t(r+\frac{q}{V})}$

Assignment 19.11 —

(a) $C(t) = \dfrac{qC_{in}}{q+rV} + Ce^{-t(r+\frac{q}{V})}$

Assignment 19.12 —

(a) The differential equation is $N' = -\lambda N$ with $N(0) = N_0$. The general solution is $N(t) = N_0 e^{-\lambda t}$. If $t = 1600$ is $N(1600) = N_0/2$. From this it follows that $\lambda = \dfrac{\ln(1/2)}{-1600}$ or approximately 0.0004332.
To calculate how long it takes for 5% to convert, we need to solve the equation $N(t_5) = N_0 e^{-\lambda t_5} = 0.95N_0$. We get:

$$t_5 = \frac{\ln(0.95)}{-0.0004332} \approx 118.37 \text{ years.}$$

Assignment 19.13 —

(a) $v(t) = \dfrac{g(M-mt)}{m-K} + \dfrac{mv_0}{K} - \dfrac{mv_0 M^{\frac{-K}{m}}}{K}(M-mt)^{\frac{K}{m}} - \dfrac{gM^{1-\frac{K}{m}}}{m-K}(M-mt)^{\frac{K}{m}}$

E

Review exercises

E.1 Exam January 2019

Assignment E.1 — For each statement, indicate whether it is True, False, or not verifiable based on the data provided. Also provide a brief rationale for your answer.

(a) The function

$$f(x) = \frac{\sin^2(x) + \cos(x)}{\tan(x)\ \csc(x)}$$

is not even.

(b) It holds true that

$$\cos\big(2\arcsin(x)\big) = 1 - 2x^2 .$$

(c) If $x = f(t)$ and $y = g(t)$, then we can find a function h for which it holds that $y = h(x)$.

(d) For the circumference O of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 ,$$

where $b < a$, it holds that $2\pi b < O < 2\pi a$.

(e) For any two random vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ it holds that

$$|\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}}| \leq \|\vec{\boldsymbol{u}}\| \|\vec{\boldsymbol{v}}\| .$$

(f) All intersections of the graphs of $r = f(\theta)$ and $r = g(\theta)$ can be found by solving $f(\theta) = g(\theta)$.

(g) If the radius of a spherical balloon being inflated increases at a rate of 3 mm per second, then the volume of the balloon increases at a rate of 27 mm^3 per second.

(h) For an increasing function, the left Riemann sum is always smaller than the right Riemann sum.

(i) If $f(x, y) = 4x + 4y$, then it holds true that

$$\left|D_{\vec{\mathbf{u}}}f(x, y)\right| \leq 4\,.$$

Assignment E.2 — Determine which equation belongs to which graph in Figure E.1.

(1) $r(\theta) = 1 + \dfrac{\sin(\theta)}{2}$

(2) $\vec{\boldsymbol{r}}(t) = \big(3\sin(2t), t, t\big)$

(3) $x(t) = \sin\big(t + \cos(t)\big)$ en
$y(t) = \cos\big(t + \sin(t)\big)$

(4) $r(\theta) = 1 + \dfrac{\cos(\theta)}{2}$

(5) $\vec{\boldsymbol{r}}(t) = \big(t\cos(t), t\sin(t), t\big)$

(6) $\vec{\boldsymbol{r}}(t) = \big(\cos(t), \ln(t), \sin(t)\big)$

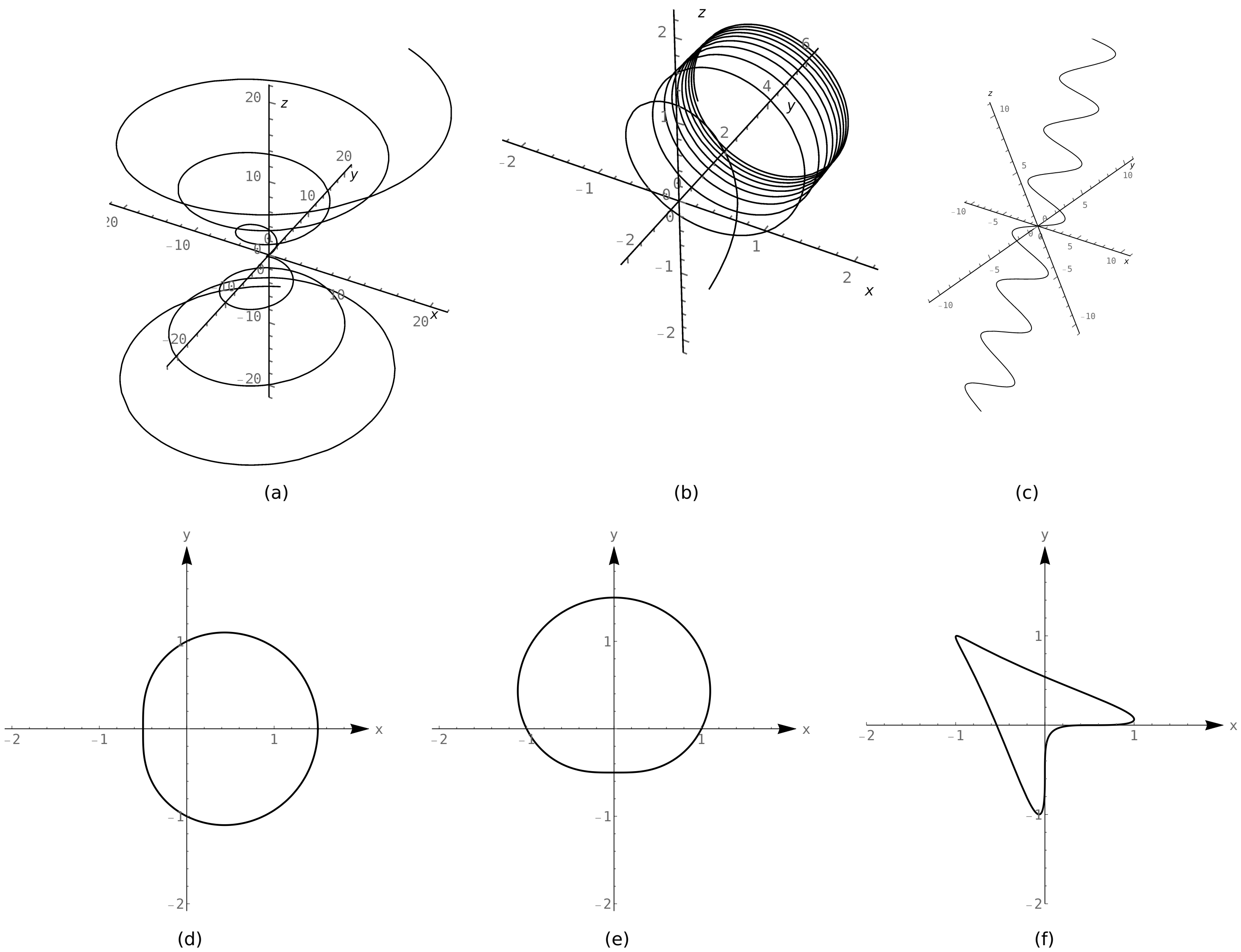
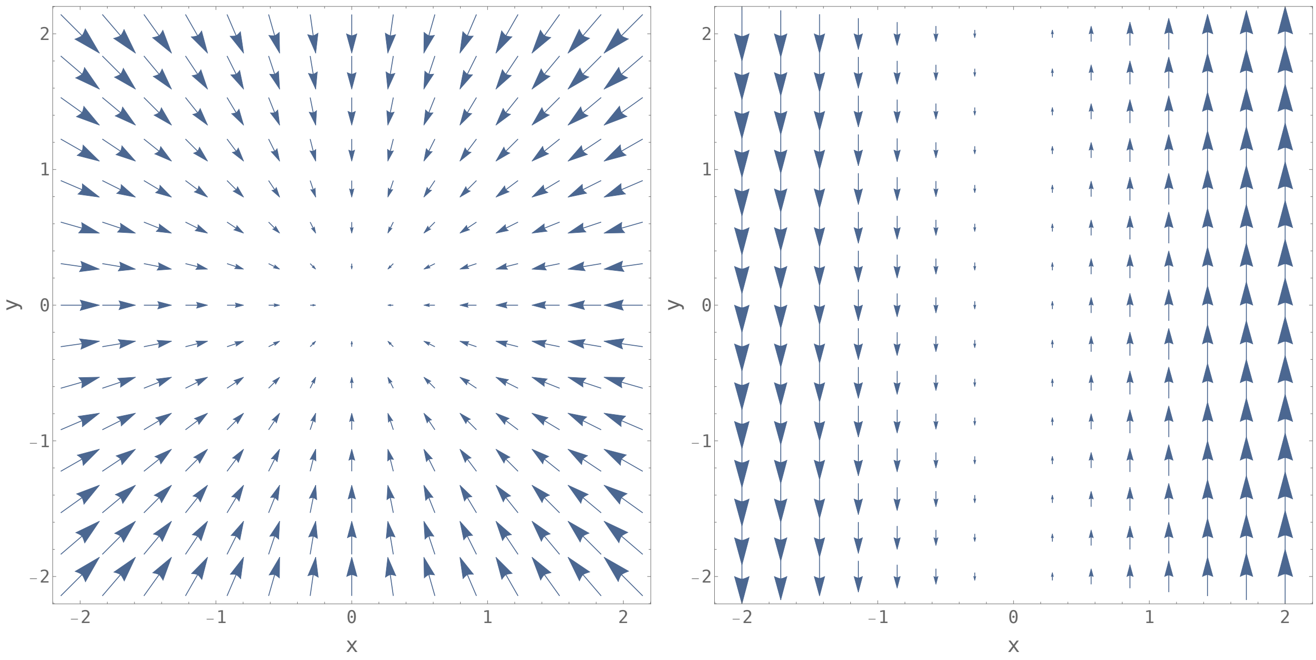


Figure E.1

Assignment E.3 — Determine whether the divergence and rotation of the vector fields in Figure E.2 is positive, negative or zero.



(a) (b)

Figure E.2

Assignment E.4 — Consider the curve defined by $y = x^2 - 1$ and suppose that we want to find the zero points of $x^2 - 1 = 0$ using Newtons method.

(a) What will happen if we start this method from the point $x_0 = 0$?

(b) Argue whether we best start this method from the point $x_0 = 0.2$ or $x_0 = 10$.

Assignment E.5 — Write down the integral that allows you to calculate the volume of the body of revolution created by rotating the area bounded by the x-axis, the y-axis and the graph of the function $f(x) = -x^2 + 2x + 3$ around the

(a) y-axis

(b) the line $y = -1$.

Assignment E.6 — Determine the surface area of rotation of the body created by rotating the curve described by $x = t$ and $y = t^3$ around the x-axis for $0 \leq x \leq 1$.

Assignment E.7 — If $w = x^2 y + z^2$ with $x = \rho\cos(\theta)\sin(\phi)$, $y = \rho\sin(\theta)\sin(\phi)$ and $z = \rho\cos(\phi)$, determine then $\frac{\partial w}{\partial \theta}$ and $\frac{\partial w}{\partial \phi}$.

Assignment E.8 — Determine the volume located under $2x^2 + 2y^2 + z^2 = 18$, above $z = 0$ and within $x^2 + y^2 = 4$.

Assignment E.9 — Swap the integration order of

$$\int_0^1 \int_0^{\arccos(y)} f(x,y)\, dx\, dy\,.$$

Assignment E.10 — A particle moves through a helicoidal orbit given by

$$\vec{\boldsymbol{r}}(t) = 6\cos(\pi t)\hat{\boldsymbol{\imath}} + 6\sin(\pi t)\hat{\boldsymbol{\jmath}} + 2t\hat{\boldsymbol{k}}\,,$$

with $t \geq 0$, towards a spherical tumor cell with cartesian equation given by

$$x^2 + y^2 + z^2 = 100\,.$$

(a) Determine when the particle arrives at the wall of the tumor cell.

(b) Determine where the particle enters the tumor cell.

(c) Determine the angle at which the particle enters the tumor cell.

Assignment E.11 — The so-called error function is used, among other things, to describe groundwater flow and is given by

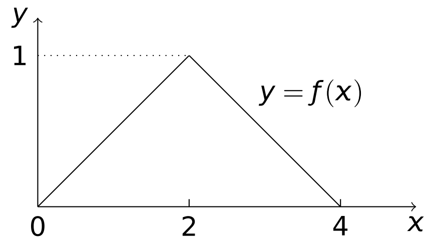
$$f(x) = \int_0^x e^{-u^2}\,du\,.$$

However, the integral cannot be calculated analytically and numerical integration is not evident here either. However, the function can be approximated using a Taylor series development. Determine the MacLaurin series development of this function up to fourth order terms.

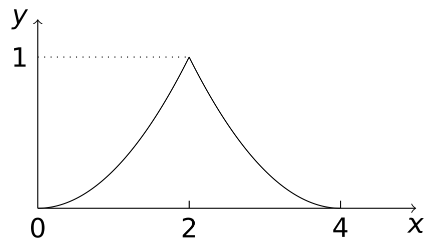
E.2 Exam August 2019

Assignment E.12 — For each statement, indicate whether it is True or False. Also, give a brief motivation for your answer.

(a) Given the function $f : \mathbb{R} \to \mathbb{R}$ with the graph below being the figure.



Furthermore, the function g is given by $g : \mathbb{R} \to \mathbb{R} : x \mapsto g(x) = x/2$. Then the figure below shows the graph of the product p of these functions
$p : \mathbb{R} \to \mathbb{R} : x \mapsto p(x) = f(x) \cdot g(x)$.



(b) The equation of the function whose graph is a parabola with top in $(3,2)$ and second derivative -4 is given by $f(x) = -2x^2 + 12x - 16$.

(c) $\arcsin\left(\sin\left(\frac{3\pi}{4}\right)\right) = \frac{3\pi}{4}$

(d) If $f(x)$ and $g(x)$ are polynomial functions, then the rational function has $h(x) = \frac{f(x)}{g(x)}$ a vertical asymptote at the height of all x for which it holds that $g(x) = 0$.

(e) The curvature of the intersection of $x^2 + y^2 + z^2 = 1$ and $ax + by + cz = 0$ is constant and equal to 1.

(f) The tangent to the curve $\vec{\boldsymbol{r}}(t) = (t, 2t, 3t^2)$ at the origin is perpendicular to the vector $\vec{\boldsymbol{u}} = (2, -1, 3)$.

(g) The contour lines of the graph of $f(x,y) = y^2 + (x-2)^2$ are all circles.

(h) The function $f(x,y,z) = x + y^2 + z^3$ descends the quickest in the point $(0,1,1)$ in the direction of $\hat{\boldsymbol{\imath}} + 2\hat{\boldsymbol{\jmath}} + 3\hat{\boldsymbol{k}}$.

(i) Consider a point P on a curve C for which there are two different parametrisations available, being $\vec{\boldsymbol{r}}_1(t)$ and $\vec{\boldsymbol{r}}_2(t)$. Then the tangent vector according to $\vec{\boldsymbol{r}}_1(t)$ in P the same as these according to $\vec{\boldsymbol{r}}_2(t)$.

(j) Consider the vector field $\vec{\boldsymbol{F}}(x,y,z) = (x + \sin(y), y - \sin(z), z)$. The divergence of this vector field is equal to $(1,1,1)$.

Assignment E.13 — For each graph in Figure E.3 determine the appropriate equation.

(1) $y = \arccos(2\cos(x))$	(4) $y = \cos(2\arccos(x))$
(2) $y = \arccos(2\cos(\lvert x\rvert))$	(5) $y = \cos(2\arccos(\lvert x\rvert))$
(3) $y = \arccos(2\lvert\cos(x)\rvert)$	(6) $y = \cos(2\lvert\arccos(x)\rvert)$

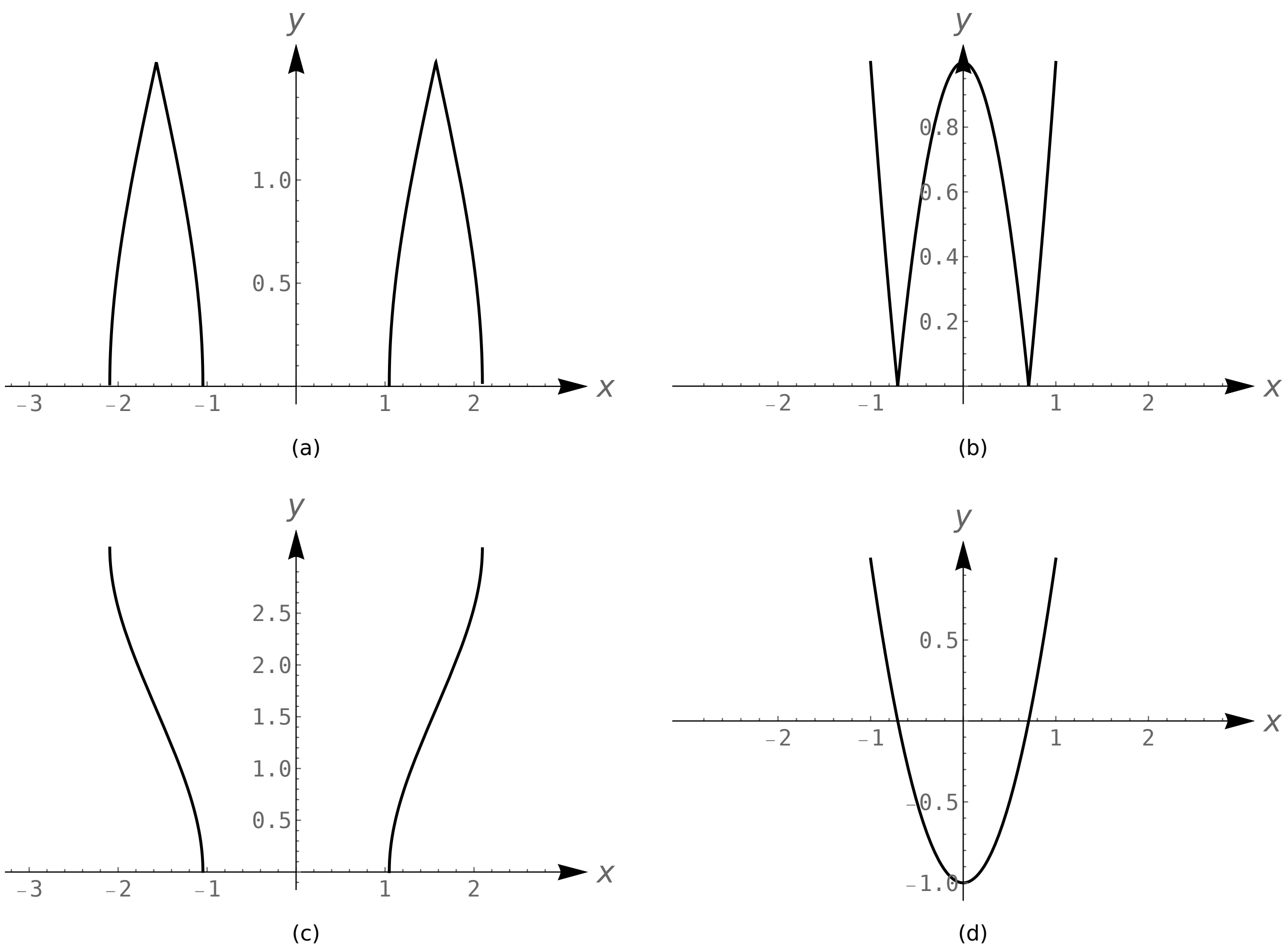


Figure E.3

Assignment E.14 — Consider the trapezoidal method over the interval $[a,b]$, which we divide into n subintervals for calculating

$$\int_a^b f(x)dx\,.$$

For the absolute total error E on this method, i.e., the total deviation between the determined integral and its numerical approximation, it is true that

$$E \leq \frac{\max|f''(\xi)|(b-a)^3}{12n^2}\,,$$

with $\xi \in [a,b]$.

(a) For which function family does this method work flawlessly?

(b) Give two examples of functions for which we are assured that $0 < E \leq b^3/(12n^2)$ if we are working over $[0,b]$.

Assignment E.15 — Calculate the limit below.

$$\lim_{x\to 2}\left(\frac{1}{x-2}-\frac{1}{|x-2|}\right)$$

Assignment E.16 — For each of the statements below, compose the integral that allows you to compute the demand. Write the integrand as simply as possible.

(a) The surface of the area within $r = -4\cos(\theta)$ and $r = -4\sin(\theta)$;

(b) the distance travelled by a particle that moves between $t = 0$ en $t = 1$ along the curve C parameterized by $\vec{r}(t) = \left(e^t \sin(t), e^t \cos(t)\right)$;

(c) the work performed by a particle moving along the curve C parameterized by $\vec{r}(t) = \left(1+t^2, 1+\sin(\pi t)\right)$ going from $t = 0$ to $t = 1$ and thereby subject to the field $\vec{F} = \left(\frac{1}{y}, 1 - \frac{x}{y^2}\right)$.

Assignment E.17 — A water reservoir has a parabolic shape with the equation of the underlying parabola being $y = ax^2$. The maximum depth of water in this reservoir is 8 m and when the reservoir is filled to a height of 5 m, the water surface has a diameter of 20 m. What is the maximum volume of water that this reservoir can hold?

Assignment E.18 — Consider the curve

$$\arctan(x+y) = x^2 + \frac{\pi}{4},$$

where we assume that $y = f(x)$.

(a) Determine the equation of the tangent to the given curve at the point with $x = 0$.

(b) Determine the concavity of the curve at this point.

Assignment E.19 — Examine the convergence of

$$\int_0^{+\infty} e^{-x} \sin(x)\, dx.$$

Assignment E.20 — Determine the Maclaurin series expansion of the function

$$f(x) = 2x^3 \cos\left(4x^5\right).$$

Assignment E.21 — If $W(s,t) = f(u(s), v(s,t))$.

(a) Determine $\frac{\partial W}{\partial s}$ and $\frac{\partial W}{\partial t}$.

(b) Assume that $u'(1) = 1$, $\frac{\partial v}{\partial s}(1,3) = 1$, $\frac{\partial v}{\partial t}(1,3) = 4$, $u(1) = 0$, $v(1,3) = 4$, $f_u(0,4) = 1$ and $f_v(0,4) = -2$. Determine than $\frac{\partial W}{\partial s}(1,3)$.

Assignment E.22 — Calculate

$$\int_0^{\sqrt{3}} \int_{\frac{y}{\sqrt{3}}}^{\sqrt{4-y^2}} e^{-x^2-y^2}\, dx\, dy.$$

Assignment E.23 — Calculate the volume of the body bounded by the cylinders $x^2+y^2-y=0$, $x^2+y^2-2y=0$, $z=2+x$ and the xy-plane.

E.3 Exam January 2020

Assignment E.24 — For each statement, indicate whether it is True or False. Also, give a brief motivation for your answer.

(a) The tangents to the graphs of the functions $f(x) = x^3$ and $g(x) = \frac{1}{3x}$ are in every point $x \neq 0$ perpendicular to each other.

(b) The function

$$F(x) = \int_a^x f(u)\ du$$

is convex where $f(x)$ inclines and concave where $f(x)$ declines.

(c) The MacLaurin series expansion of $f(x) = x^4 e^{-3x^2}$ is given by

$$\sum_{n=0}^{+\infty} \frac{(-3)^n x^{2n+4}}{n!}.$$

(d) The contour lines of the function $f(x,y) = \alpha \arcsin\left(\sqrt{x^2+y^2}\right)$ are not all circles.

(e) The calculation of $\lim\limits_{(x,y)\to(0,0)} f(x,y)$, where

$$f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4}, & \text{als} \quad (x,y) \neq (0,0), \\ 0, & \text{als} \quad (x,y) = (0,0), \end{cases}$$

along $y = x$ and $x = y^2$ allows us to conclude that f is not continuous in $(0,0)$.

(f) A cylindrical tank made of steel has a height of 10m and a diameter of 4m. The steel jacket of the tank expands and contracts under the influence of temperature changes. The tank volume is more sensitive to changes in height than to changes in diameter.

Assignment E.25 — Consider the function

$$f(x) = \begin{cases} 2x^2 + x^2 \sin\left(\dfrac{1}{x}\right), & \text{als} \quad x \neq 0, \\ 0, & \text{als} \quad x = 0. \end{cases}$$

Discuss the continuity and derivability of the function. f in $[0, +\infty[$.

Assignment E.26 — Figure E.4 shows the graphs of four functions, while Figure E.5 shows the graphs of some of the derived functions. For each of the functions in Figure, fill in the table below E.4 with the letter(s) of the graph(s) of the corresponding derivative function(s).

Graph function	Graph derivative function	Graph function	Graph derivative function
a		c	
b		d	

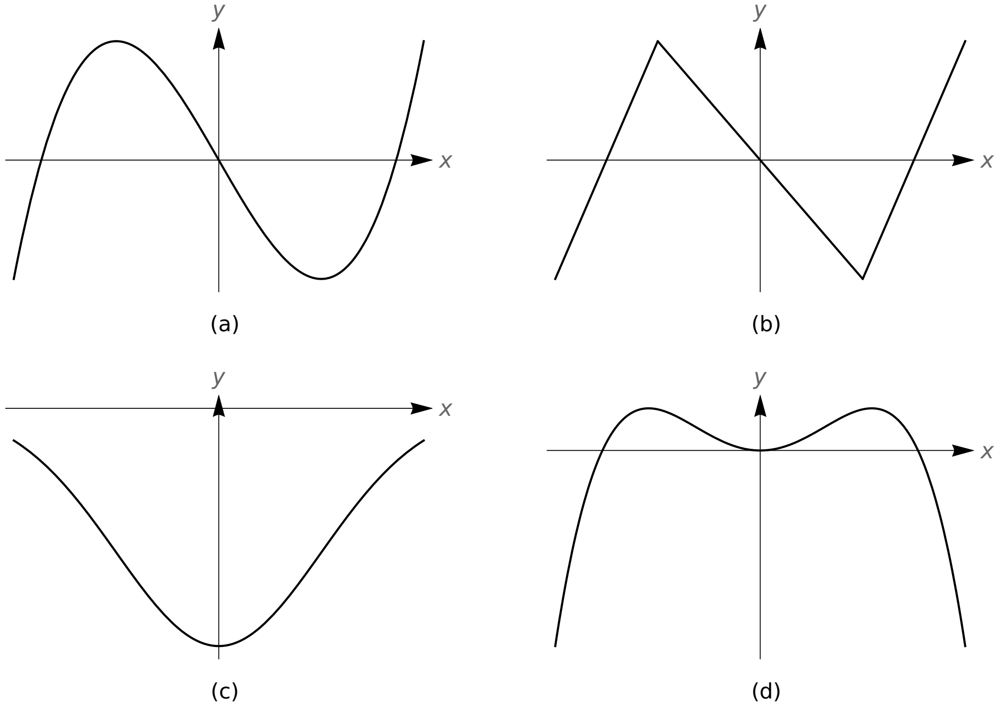


Figure E.4

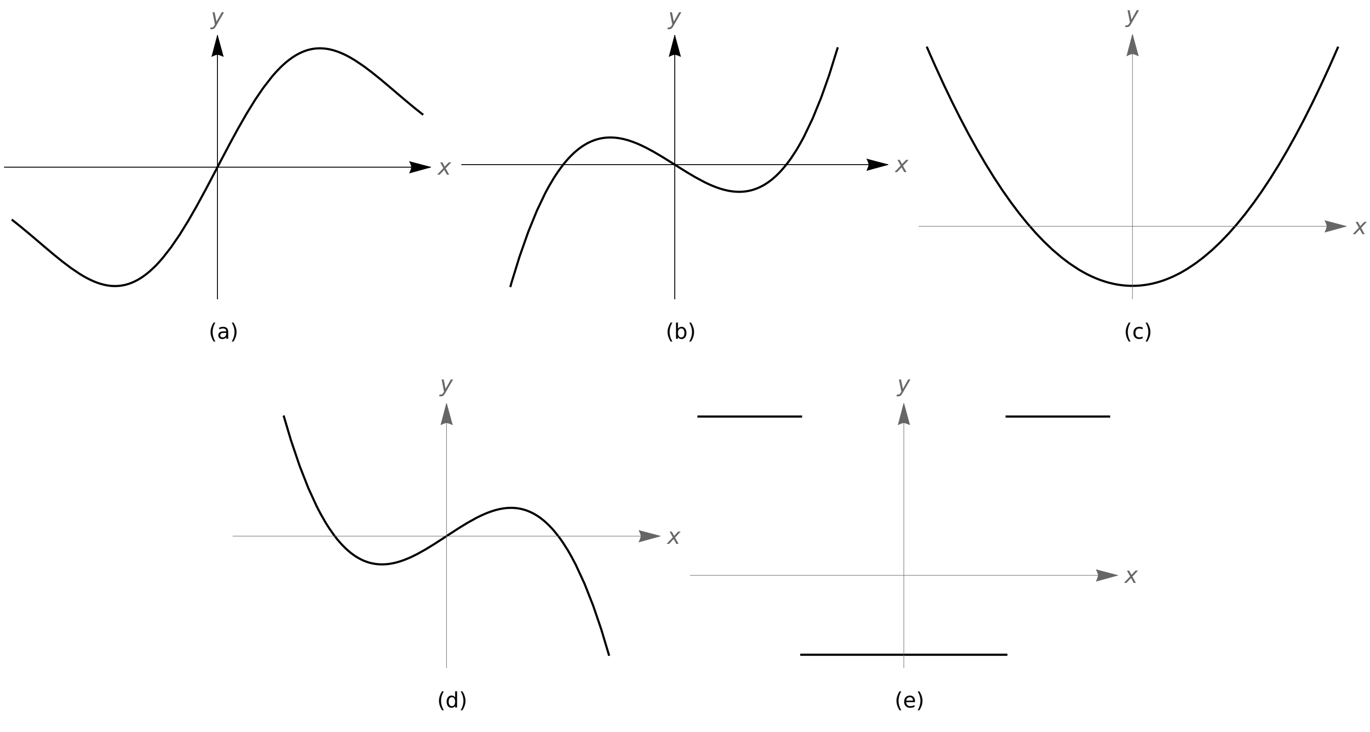


Figure E.5

Assignment E.27 — Calculate

$$\int_0^{\frac{\pi}{4}} \frac{x\sin(x)}{\cos^3(x)}\,dx\,.$$

Assignment E.28 — Consider the curve

$$x = \sin(\theta), \quad y = \sin(\theta)\cos(\theta),$$

graphically represented in Figure E.6.

(a) Determine the value(s) of the parameter θ for which the curve has a horizontal tangent. Also determine the corresponding points in the xy-plane.

(b) Compose the integral to determine the volume of the body of revolution created by the rotation of this curve about the y-axis.

(c) Compose the integral to determine the volume of the body of revolution created by the rotation of this curve about the x-axis.

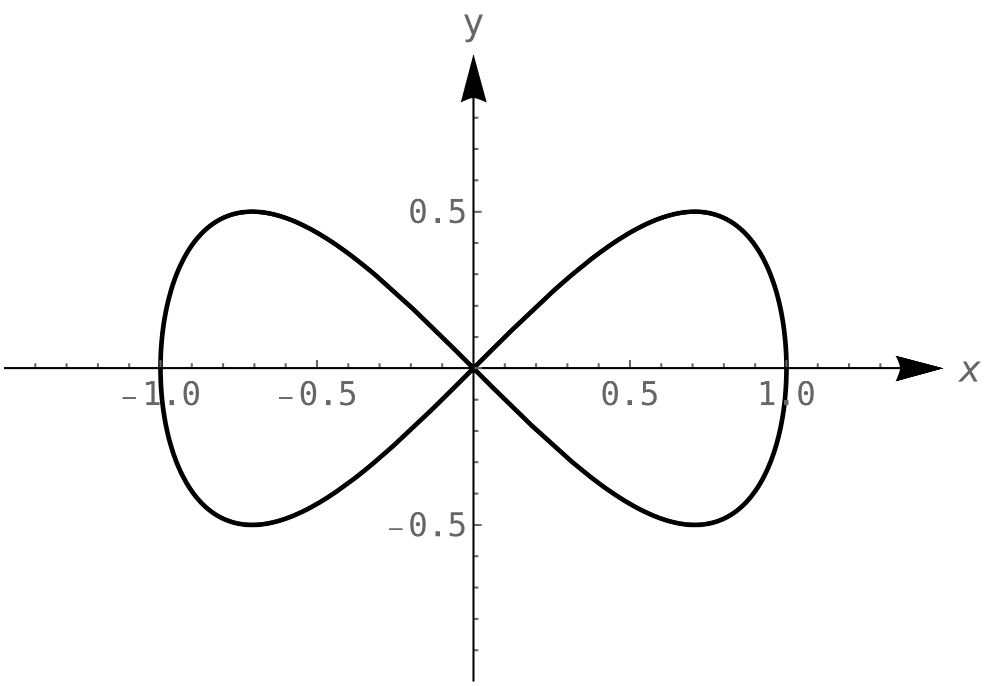


Figure E.6

Assignment E.29 — Determine the area of the region enclosed by the curves $y = \sqrt{x}$, $y = -\frac{x}{2} + 1$, $x = 1$ and $x = 4$.

Assignment E.30 — The cat Kamiel moves according to $\vec{r}(t) = \left(t, t^2, t^3\right)$ near a wooden plate that can be described by the equation

$$2x + y - 2z = 1.$$

(a) Determine all the times t when Kamiel is moving parallel to the wooden plate.

(b) Determine all the times t when Kamiel moves perpendicular to the wooden plate.

(c) The wooden plate contains some holes through which Kamiel can move. Find the locations of these holes.

Assignment E.31 — Consider the function

$$f(x,y) = \ln\left(\sqrt{x^2+y^2}\right).$$

(a) Determine the domain of f.

(b) Prove that

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0.$$

(c) Determine the equation of the tangent plane to f in the point (x_0, y_0, z_0) with $x_0 = 0$ and $y_0 = 1$.

Assignment E.32 — Swap the integration order in

$$\int_0^1 \int_{-\sqrt{1-y^2}}^{1-y} f(x,y)\, dx\, dy.$$

Assignment E.33 — Consider a wire of negligible thickness with mass density

$$\delta(x,y,z) = y^2 + z^2.$$

The wire is formed by a piecewise smooth curve formed by 1) the line between $(1,0,0)$ and $(1,2,0)$, 2) the line from $(1,2,0)$ to $(0,2,0)$ and finally 3) half a circle yz-plane from $(0,2,0)$ to $(0,-2,0)$.

(a) Determine the mass of the wire.

(b) How many moments do you need to calculate to find the center of mass?

E.4 Exam August 2020

Assignment E.34 — Consider the real-valued function f:

$$f(x) = \begin{cases} \dfrac{x^3-27}{x-3} & , \quad x < 3, \\ 2x^2 + bx & , \quad x \geq 3, \end{cases}$$

with b a parameter $\in \mathbb{R}$

(a) If possible, determine b such that f is continuous over $\mathbb{R}$.

(b) If possible, determine b such that f is derivable over $\mathbb{R}$.

Assignment E.35 — Consider the real-valued function f:

$$f(x) = \ln(x)^2 + \ln(x^2).$$

(a) Determine the domain and image of f.

(b) Examine the symmetry of f.

(c) Determine all zeros of f.

(d) Determine the asymptotes of f.

(e) Determine the extrema and inflection points of f.

(f) Examine the behavior of f and sketch its graph.

Assignment E.36 — Determine the angle α between the curve with polar equation $r = e^{\theta}$, $0 \leq \theta \leq \pi$, and the y-axis.

Assignment E.37 — Calculate

$$\int x \cos\left(x^3\right) dx$$

using the MacLaurin series expansion of function $f(x) = \cos(x)$.

Assignment E.38 — Sketch some level curves of the function $f(x, y) = x^3 - y$.

Assignment E.39 — Consider

$$\int\int_R \left(\left(x^2 + y^2\right)^{-3/2} + 3\right) dA\,,$$

with

$$R = \left\{(x, y) \mid x^2 + y^2 \leq 1 \wedge |x| + |y| \geq 1\right\}\,.$$

(a) Sketch the area R.

(b) Determine appropriate integration boundaries.

(c) Calculate the integral.

E.5 Examination January/March 2021

Assignment E.40 — To contain the coronavirus, numerous measures were taken in 2020 to reduce the reproductive rate R. Researchers demonstrated in 2020 that the evolution of the reproductive number can be described as:

$$R(t) = R_0 - \frac{1}{2}\left(1 + \tanh\left(\frac{t - t^*}{T}\right)\right)(R_0 - R_t)\,,$$

where R_0 represents the reproduction number at the start of the corona epidemic, R_t the reproduction number achieved by taking measures, t^* the time it takes before measures are followed, and T the transition time.

Sketch the graph of $R(t)$ iff $R_0 = 4$ and $R_t = 2$, given the graph of $f(t) = \tanh(t) = \frac{e^{2t}-1}{e^{2t}+1}$ in Figure E.10. Show the graphs of all transformations required for this. In each case, indicate the asymptotes and the point of intersection with x-axis.

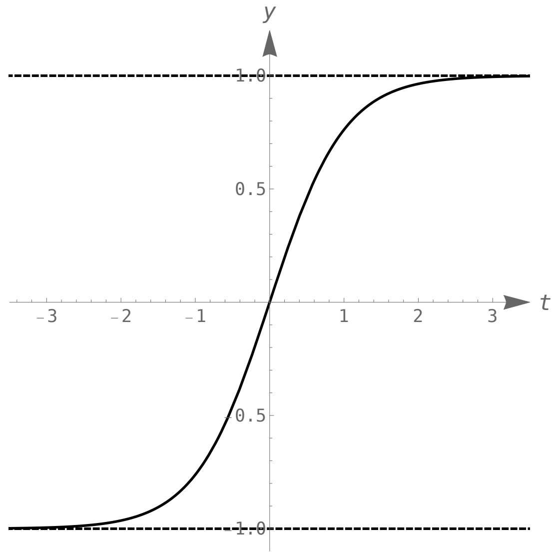


Figure E.10: Graph of the function $f(t) = \tanh(t)$ (full line) and her asymptotes (dotted line).

Assignment E.41 — Calculate the following limit:

$$\lim_{x\to+\infty} \ln(1+x)\,\ln\left(1+\frac{1}{x}\right).$$

Assignment E.42 — Consider the function $f(x) = \cos(3\arcsin(x))$ over $[-1,1]$.

(a) Prove that

$$\cos(3\arcsin(x)) = \left(1-4x^2\right)\sqrt{1-x^2}\,.$$

(b) Calculate all local and global extremes and indicate for each extremum whether it is a minimum or maximum.

(c) Calculate the points of infliction.

Assignment E.43 — Calculate

$$\lim_{x\to 0} \frac{\left(e^{2x}-1\right)\ln\left(1+x^3\right)}{\left(1-\cos(3x)\right)^2}$$

using the following Maclaurin series developments (and not l'Hôpital's rule):

$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}\,, \qquad \sin(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}\,, \qquad \text{en } \ln(x) = \sum_{n=1}^{+\infty} (-1)^{n+1} \frac{(x-1)^n}{n}\,.$$

Assignment E.44 — Calculate the integrals below:

(a)

$$\int \frac{dx}{x^2\sqrt{x^2+16}}$$

(b)

$$\int \frac{dx}{e^{2x} - 4e^{x} + 4}$$

Assignment E.45 — On Figure E.12 you can see the lemniscate of Bernouilli, given by the parametric equations

$$x(\theta) = 3\sin(\theta), \quad y(\theta) = 2\sin(2\theta), \quad \text{met } 0 \le \theta \le 2\pi.$$

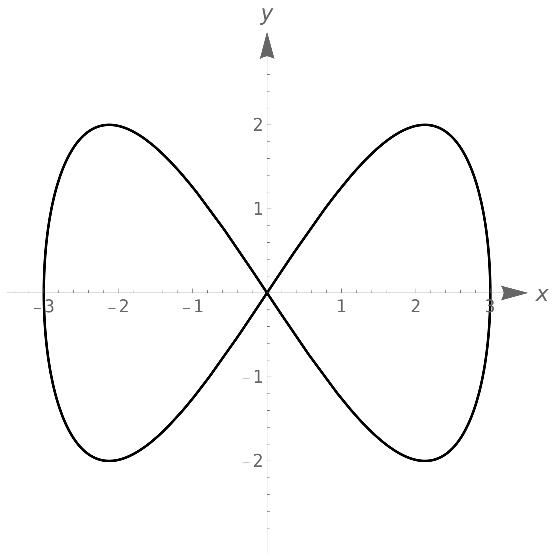


Figure E.12: Graph of the lemniscate of Bernouilli.

(a) Show that the cartesian equation of the curve is equal to

$$81y^2 = 16x^2(9 - x^2).$$

(b) Determine the volume of the body that you obtain by rotating the curve about the x-axis.

(c) Determine the integral that allows you to calculate the volume of the body you obtain by rotating the curve around the y-axis.

Assignment E.46 — Consider the vector-valued function $\vec{r}(t) = \big(4\cos(t), 4\sin(t), 3t\big)$.

(a) Determine the unit tangent vector and unit normal vector for $t = \pi$.

(b) Determine an equation of the tangent plane to the curve for $t = \pi$.

(c) Parameterize the curve with the arc length parameter.

(d) Calculate the curvature in $t = \frac{4\pi}{3}$.

Assignment E.47 — The temperature in a room can be described in three dimensions using the following function:

$$T = f(x, y, z) = 10\frac{\cos^2(xy + z)}{\sin\left(\dfrac{y}{x}\right)}.$$

(a) Determine the domain and image of f.

(b) Determine the directional derivative in the point $P(1,\pi/2,\pi/4)$ in the direction of the vector $\vec{\boldsymbol{v}} = (-1,-1,-1)$.

(c) Determine the direction in which we must move from the point P to continue to experience maximum cooling.

(d) Determine the tangent plane to the graph of f at the point P.

(e) Determine f_s and f_t if $x = g(s,t)$, $y = h(s,t,v)$ and $z = i(s)$. You should not calculate the occurring partial derivatives explicitly.

Assignment E.48 — Determine the surface of the part of $x^2+y^2+z^2 = 4z$ that is located inside $z = x^2+y^2$.

E.6 Exam August 2021

Assignment E.49 — Consider the function

$$f(x) = \alpha x+\beta+\ln\left(e^{-2x}-4\right),$$

with α, $\beta \in \mathbb{R}_0$.

(a) Determine the domain of the function f.

(b) Determine α and β such that the line $y = 2x-1$ is a skewed asymptote for f.

(c) For $\alpha = 4$ and $\beta = -1$

(i) Determine all asymptotes.

(ii) Determine all (local and global) extremes and discuss where f inclines/declines.

(iii) Examine the concavity of f.

Assignment E.50 — Determine the integral below by using a Maclaurin series development for the function in the integrandum.

$$\int \ln\left(1+x^2\right)\, dx$$

Assignment E.51 — Determine

$$\int \tan(x)\ln\left(\cos(x)\right) dx\,.$$

Assignment E.52 — Determine the arc length of the curve described by

$$\vec{\boldsymbol{r}}(t) = \left(\cos(t),\sin(t),\ln(\cos(t))\right)$$

voor $t \in [0,\pi/4]$.

Assignment E.53 — We consider a balloon moving in the direction of the maximum wind force. We know from meteorology that the wind always blows from a high-pressure area to a low-pressure area, consequently the wind force is always directed in the direction of maximum air pressure decrease. Now suppose the air pressure is given by following function:

$$P(x,y) = x^2 + \frac{y^2}{2}$$

and let us assume for simplicity that it does not change through time and is independent of height.

(a) Determine the direction in which the balloon will move from the point $(4,1)$.

(b) Along which curve through the point $(4,1)$ is the balloon going to move?

(c) At what point will the balloon end up after waiting a sufficiently long time (no calculation required)?

Assignment E.54 — Consider the double integral

$$\int\int_R f(x,y)\,dA\,,$$

where A is the area in the first quadrant enclosed between
$y = 2/x$, $y = 6/x$, $y = x-1$ and $y = x+1$. Determine the boundaries of this double integral if

(a) $dA = dx\,dy$

(b) $dA = dy\,dx$

Assignment E.55 — Consider the vector field

$$\vec{\boldsymbol{F}} = \left(2xe^{-y}, 2y - x^2e^{-y}\right)$$

(a) Show that this vector field is conservative.

(b) Calculate the line integral along the polar curve $r(\theta) = \theta$ between $\theta = 0$ and $\theta = \pi/2$.

E.7 Oplossingen

Assignment E.1 —

(a) False. The function is even, because $f(-x) = f(x)$

(b) True. This can be verified with the formula $\cos(2\alpha) = \cos^2(\alpha) - \sin^2(\alpha)$.

(c) False. For polar curves, this is not possible.

(d) True. It can be reasoned that a circle with radius a has a larger circumference, because it is completely around the ellipse, and a circle with radius b has a smaller circumference, because it is completely inside the ellipse.

(e) True. Can be shown via definition of scalar product.

(f) False. Sometimes the origin is not found, e.g. when $r_1 = \sqrt{3}\cos(\theta)$ and $r_2 = \sin(\theta)$.

(g) False. $V_t = 12\pi r^2$.

(h) True. To be seen on figures with rectangles.

(i) False. It is true that $\left|D_{\vec{u}} f(x,y)\right| \leq 4\sqrt{3}$

Assignment E.2 — (1) e; (2) c; (3) f; (4) d; (5) a; (6) b

Assignment E.3 — (a) Negative divergence and rotation 0; (b) Divergence 0 and positive rotation.

Assignment E.4 —

(a) The method cannot be performed because there are no intersections with the x-axis.

(b) Best to start from $x_0 = 10$, because the tangent will have a steeper slope. In the case $x_0 = 0.2$ you will end up far from zero.

Assignment E.5 —

(a) $V = 2\pi \displaystyle\int_0^3 x(-x^2 + 2x + 3)\,dx$

(b) $V = \pi \displaystyle\int_0^3 \left[(-x^2 + 2x + 4)^2 - 1\right] dx$

Assignment E.6 — $SA = 2\pi \displaystyle\int_0^1 x^3\sqrt{1 + 9x^4}\,dx = \frac{\pi}{9}\left(\frac{10\sqrt{10}}{3} - \frac{1}{3}\right)$

Assignment E.7 —

(a) $\dfrac{\partial w}{\partial \theta} = \dfrac{\partial w}{\partial x}\dfrac{\partial x}{\partial \theta} + \dfrac{\partial w}{\partial y}\dfrac{\partial y}{\partial \theta} = -2xy\rho\cos(\theta)\sin(\phi) + x^2\rho\cos(\theta)\sin(\phi)$

(b) $\dfrac{\partial w}{\partial \phi} = \dfrac{\partial w}{\partial x}\dfrac{\partial x}{\partial \phi} + \dfrac{\partial w}{\partial y}\dfrac{\partial y}{\partial \phi} + \dfrac{\partial w}{\partial z}\dfrac{\partial z}{\partial \phi} = 2xy\rho\cos(\theta)\cos(\phi) + x^2\rho\sin(\theta)\sin(\phi) - 2z\rho\sin(\phi)$

Assignment E.8 — $V = \displaystyle\int_0^{2\pi}\int_0^2 \sqrt{18 - 2r^2}\,r\,dr\,d\theta = \pi\left(18\sqrt{2} - \frac{10\sqrt{10}}{3}\right)$

Assignment E.9 — $I = \int_0^{\pi/2} \int_0^{\cos(x)} f(x,y)\,dydx$

Assignment E.10 —

(a) $t = 4$

(b) $\vec{\boldsymbol{r}}(4) = (6,0,8)$

(c) $\alpha = \frac{\pi}{2} - \arccos\left(\frac{-4}{5\sqrt{9\pi^2+1}}\right)$

Assignment E.11 — $f(x) = \int\limits_0^x e^{-u^2}\,du = \int\limits_0^x \left(1 - u^2 + \frac{u^4}{2!} - \frac{u^6}{3!} + \ldots\right)\,du = x - \frac{x^3}{3.1!} + \frac{x^5}{5.2!} - \frac{x^7}{7.3!} + \ldots$

Assignment E.12 —

(a) False. The part between $x = 2$ and $x = 4$ describes part of a parabola that opens downwards and therefore not a parabola that opens upwards.

(b) True. From the first derivative, the top of the parabola easily follows and the second derivative confirms that this is the top of a parabola that opens downwards.

(c) False. $\frac{3\pi}{4} \notin \text{im } \arcsin = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

(d) False. If $f(x)$ and $g(x)$ are simultaneously 0, then a perforation point and not a VA occurs.

(e) True. The cross section is a circle centered at the origin and radius 1.

(f) True. The scalar product of $\vec{\boldsymbol{r}}'(0)$ and $\vec{\boldsymbol{u}}$ equals 0.

(g) False. For negative values of C we get nothing, because $y^2 + (x-2)^2 \geq 0$.

(h) False. It's the direction of the fastest increase.

(i) False. Consider e.g. the circle with equation $x^2 + y^2 = 4$ and parametrisations $\vec{\boldsymbol{r}}_1(t) = (2\cos(t), 2\sin(t))$ and $\vec{\boldsymbol{r}}_2(t) = \left(t, \pm\frac{t}{\sqrt{4-t^2}}\right)$. Then it is clear from the first derivations that the tangent vectors are different.

(j) False. The divergence is $1 + 1 + 1 = 3$.

Assignment E.13 — (1) c; (2) c; (3) a; (4) d; (5) d; (6) b

Assignment E.14 —

(a) Constant and linear functions.

(b) $f(x) = \sin(x)$ en $f(x) = \cos(x)$

Assignment E.15 — $\lim\limits_{x\to 2}\left(\frac{1}{x-2} - \frac{1}{|x-2|}\right)$ does not exist, but $\lim\limits_{x \overset{>}{\to} 2}\left(\frac{1}{x-2} - \frac{1}{|x-2|}\right) = 0$ and $\lim\limits_{x \overset{<}{\to} 2}\left(\frac{1}{x-2} - \frac{1}{|x-2|}\right) = -\infty$

Assignment E.16 —

(a) $O = \dfrac{1}{2}\displaystyle\int_{\pi}^{5\pi/4} 16\sin^2(\theta)\,d\theta + \dfrac{1}{2}\displaystyle\int_{5\pi/4}^{3\pi/2} 16\cos^2(\theta)\,d\theta$

(b) $s = \sqrt{2}\displaystyle\int_0^1 e^t\,dt$

(c) $A = \displaystyle\int_0^1 \left(\dfrac{2t}{1+\sin(t\pi)}, \pi\cos(t\pi) - \dfrac{\pi(1+t^2)\cos(t\pi)}{(1+\sin(t\pi))^2}\right) dt$

Assignment E.17 — The maximum water volume is 640π m^3.

Assignment E.18 —

(a) $y = -x+1$

(b) $y''(0,1) = 4$

Assignment E.19 — Convergent.

Assignment E.20 — $f(x) = 2x^3\cos\left(4x^5\right) = 2\displaystyle\sum_{n=0}^{+\infty} \dfrac{(-1)^n\, 4^{2n}\, x^{10n+3}}{(2n)!}$

Assignment E.21 —

(a) $\dfrac{\partial W}{\partial s} = \dfrac{\partial f}{\partial u}\dfrac{du}{ds} + \dfrac{\partial f}{\partial v}\dfrac{\partial v}{\partial s}$ and $\dfrac{\partial W}{\partial t} = \dfrac{\partial f}{\partial v}\dfrac{\partial v}{\partial t}$

(b) $\dfrac{\partial W}{\partial s}(1,3) = -1$

Assignment E.22 — $\displaystyle\int_0^{\sqrt{3}} \int_{\frac{y}{\sqrt{3}}}^{\sqrt{4-y^2}} e^{-x^2-y^2}\,dx\,dy = \dfrac{\pi}{6}\left(1-e^{-4}\right)$

Assignment E.23 — $V = \displaystyle\int_0^{\pi}\int_{\sin(\theta)}^{2\sin(\theta)}\int_0^{2+r\cos(\theta)} r\,dz\,dr\,d\theta = \dfrac{3\pi}{2}$

Assignment E.24 —

(a) True. The product of the corresponding ricos (first derivatives) is always -1.

(b) True. Use that $F'(x) = f(x)$ and $F''(x) = f'(x)$.

(c) True. Transform the MacLaurin series developement of e^x.

(d) True. If $C = 0$ we have a point.

(e) True. The limit along $y = x$ is 0 and the limit along $x = y^2$ is $1/2$.

(f) False. $dV = 4\pi\,dH + 20\pi\,dD$ and $20\pi > 4\pi$.

Assignment E.25 — The function f is continuous everywhere in $[0,+\infty[$. The function f is derivable if $x \neq 0$ and not derivable if $x = 0$.

Assignment E.26 —

Graph function	Graph derivative function	Graph function	Graph derivative function
a	c	c	a
b	e	d	d

Assignment E.27 — $\displaystyle\int_0^{\frac{\pi}{4}} \frac{x\sin(x)}{\cos^3(x)}\,dx = \frac{\pi}{4} - \frac{1}{2}$

Assignment E.28 —

(a) Horizontal tangent for $\theta = \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$. The corresponding points in the xy-plane are $\left(\frac{\sqrt{2}}{2}, \frac{1}{2}\right), \left(\frac{\sqrt{2}}{2}, -\frac{1}{2}\right), \left(-\frac{\sqrt{2}}{2}, \frac{1}{2}\right), \left(-\frac{\sqrt{2}}{2}, -\frac{1}{2}\right)$.

(b) $V = 4\pi \displaystyle\int_0^{\pi/2} \sin^2(\theta)\cos^2(\theta)\,d\theta$

(c) $V = 2\pi \displaystyle\int_0^{\pi/2} \sin^2(\theta)\cos^3(\theta)\,d\theta$

Assignment E.29 — $A = \dfrac{65}{12}$

Assignment E.30 —

(a) $t = \dfrac{1 \pm \sqrt{13}}{6}$

(b) There are no such times.

(c) Holes: $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}\right)$, $(-1,1,-1)$ and $(1,1,1)$.

Assignment E.31 —

(a) $\operatorname{dom} f = \left\{(x,y) \in \mathbb{R}^2 \,|\, x \neq 0 \ \wedge \ y \neq 0\right\}$

(b) Prove yourself.

(c) $y - z = 1$

Assignment E.32 — $I = \displaystyle\int_{-1}^{0}\int_0^{\sqrt{1-x^2}} f(x,y)\,dydx + \int_0^1\int_0^{1-x} f(x,y)\,dydx$

Assignment E.33 —

(a) $M = \dfrac{20}{3} + 8\pi$

(b) Three moments.

Assignment E.34 —

(a) $b = 3$

(b) No value of b.

Assignment E.35 —

(a) domain $f = \mathbb{R}_0^+$ and image $f = [-1, +\infty[$

(b) No symmetry.

(c) $x = 1$ and $x = e^{-2}$

(d) VA: $x = 0$

(e) Minimum: $f(e^{-1}) = -1$ and Infliction point: $f(1) = 0$

(f) See graph below.

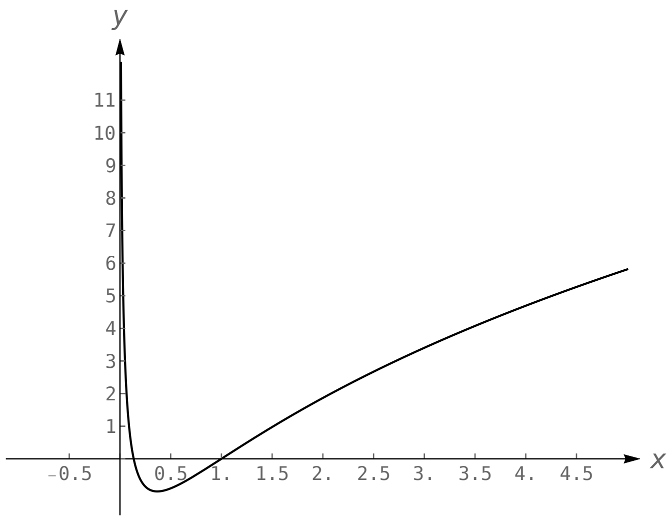


Figure E.7: Figure from Exercise E.35.

Assignment E.36 — $\alpha = \dfrac{\pi}{4}$

Assignment E.37 — $\displaystyle\int x \cos\left(x^3\right) dx = \sum_{n=0}^{+\infty} \frac{(-1)^n x^{6n+2}}{(2n)!\,(6n+2)} + C$

Assignment E.38 — See the figure below.

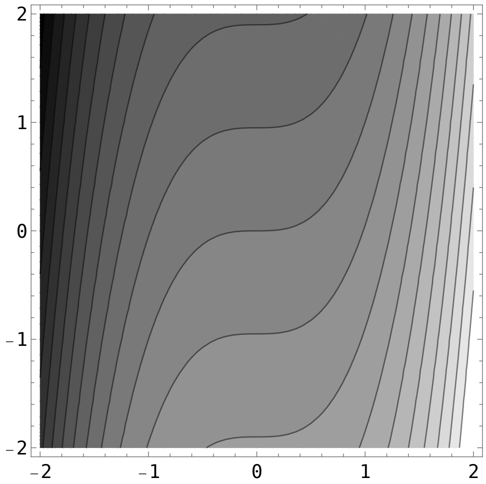


Figure E.8: Figure from Exercise E.38.

Assignment E.39 —

(a) See the figure below.

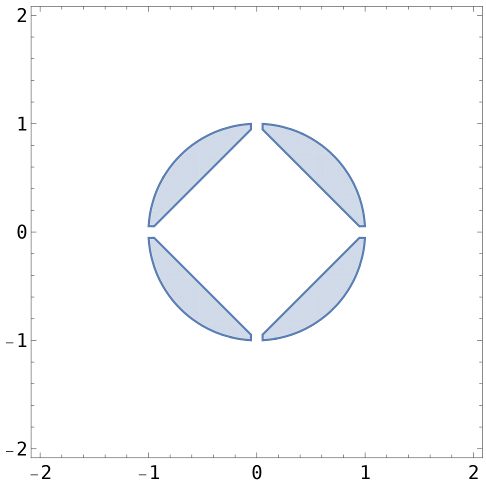


Figure E.9: Figure form exercise E.39.

(b) $I = 4\int_0^{\pi/2}\int_{1/(\cos(\theta)+r\theta)}^{1}\left(r^{-2}+3r\right)drd\theta$

(c) $I = 2+\pi$

Assignment E.40 — See the figure below for $T = 5$ and $t^* = 5$.

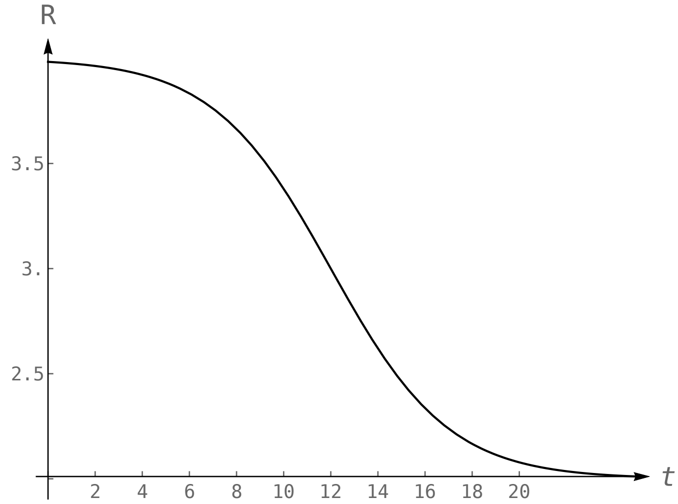


Figure E.11: Figure from Exercise E.40.

Assignment E.41 — $\lim_{x\to+\infty} \ln(1+x)\ln\left(1+\frac{1}{x}\right) = 0$

Assignment E.42 —

(a) Prove by saying $y = \arcsin(x)$ and using the trigonometric formulas $\cos(3y) = \cos(y+2y) = \cos(2y)\cos(y) - \sin(2y)\sin(y)$ en $\sin^2(y) + \cos^2(y) = 1$.

(b) Global minima: $f\left(\pm\frac{\sqrt{3}}{2}\right) = -1$; local maxima: $f(\pm 1) = 0$; global maximum: $f(0) = 1$.

(c) Infliction points if $x = \pm\frac{\sqrt{3-\sqrt{3}}}{2}$.

Assignment E.43 — $\lim_{x\to 0} \frac{\left(e^{2x}-1\right)\ln\left(1+x^3\right)}{\left(1-\cos(3x)\right)^2} = \frac{8}{81}$

Assignment E.44 —

(a) $\int \frac{dx}{x^2\sqrt{x^2+16}} = -\frac{1}{16}\frac{\sqrt{16+x^2}}{x} + C$

(b) $\int \frac{dx}{e^{2x}-4e^x+4} = \frac{\ln(e^x)}{4} - \frac{\ln|e^x-2|}{4} - \frac{1}{2(e^x-2)} + C$

Assignment E.45 —

(a) Prove yourself.

(b) $V = 2\pi\int_0^3 y^2\, dx = \frac{64\pi}{5}$

(c) $V = 36\pi\int_0^{\pi/2} \sin^2(2\theta)\, d\theta$

Assignment E.46 —

(a) $\vec{\boldsymbol{T}}(\pi) = \left(0, -\frac{4}{5}, \frac{3}{5}\right)$ en $\vec{\boldsymbol{N}}(\pi) = (1, 0, 0)$

(b) $x = -4$

(c) $\vec{\boldsymbol{r}}(s) = \left(4\cos\left(\frac{s}{5}\right), 4\sin\left(\frac{s}{5}\right), \frac{3s}{5}\right)$

(d) $\frac{4}{25}$

Assignment E.47 —

(a) domain $f = \left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, x \neq 0 \,\wedge\, \sin\left(\frac{y}{x}\right) \neq 0\right\}$

(b) Image $f = [-10, 10]$

(c) $D_{\vec{\boldsymbol{v}}} f\left(1, \frac{\pi}{2}, \frac{\pi}{4}\right) = -\frac{5\pi + 20}{\sqrt{3}}$

(d) $-\vec{\nabla} f\left(1, \frac{\pi}{2}, \frac{\pi}{4}\right) = -(5\pi, 10, 10)$

(e) $\pi(x-1) + 2\left(y - \frac{\pi}{2}\right) + 2\left(z - \frac{\pi}{4}\right) = 0$

(f) $f_s = \frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial f}{\partial z}\frac{dz}{ds}$

(g) $f_t = \frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$

Assignment E.48 — $SA = 4\pi$

Assignment E.49 —

(a) domain $f =]-\infty, -\ln(2)[$

(b) $\alpha = 4$ and $\beta = -1$

(c) (i) SA: $y = 2x - 1$ en VA: $x = -\ln(2)$

(ii) There is a maximum if $x = \ln(8^{-1/2})$. The function f inclines over $]-\infty, \ln(8^{-1/2})[$ and declines over $]\ln(8^{-1/2}), -\ln(2)[$.

(iii) The function f is concave over $]-\infty, -\ln(2)[$.

Assignment E.50 — $\displaystyle\int \ln\left(1 + x^2\right)\, dx = \sum_{n=1}^{+\infty} \frac{(-1)^{n+1} x^{2n+1}}{n(2n+1)} + C$

Assignment E.51 — $\displaystyle\int \tan(x)\ln\left(\cos(x)\right) dx = -\frac{1}{2}\ln^2(\cos(x)) + C$

Assignment E.52 — $s = \frac{1}{2}\ln\left(\left|\frac{2+\sqrt{2}}{2-\sqrt{2}}\right|\right)$

Assignment E.53 —

(a) $-\vec{\nabla}p(4,1) = (-8,-1)$

(b) $y = \frac{\sqrt{x}}{2}$

(c) The origin.

Assignment E.54 —

(a) $\int_1^2 \int_{2/y}^{y+1} f(x,y)\,dxdy + \int_2^3 \int_{y-1}^{6/y} f(x,y)\,dxdy$

(b) $\int_1^2 \int_{2/x}^{x+1} f(x,y)\,dydx + \int_2^3 \int_{x-1}^{6/x} f(x,y)\,dydx$

Assignment E.55 —

(a) Prove for yourself that $\text{rot}\,\vec{\boldsymbol{F}} = 0$.

(b) $\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \frac{\pi^2}{4}$

F

List with definitions

2.1	Set	6
2.2	Complex number	19
3.1	Relation	28
3.2	Fundamental graphing principle	29
3.3	Function	32
3.4	Fundamental graphing principle for functions	34
3.5	Composite function	37
3.6	Injective, surjective and bijective	40
3.7	Even/odd functions	42
3.8	Periodic function	43
3.9	Function behaviour	44
3.10	Local and global extrema	45
3.11	Absolute value	54
3.12	Inverse function	57
4.1	Linear function	69
4.2	Quadratic function	70
4.3	Polynomial function	74
4.4	Polynomial division	75
4.5	Rational function	78
4.6	Circle	85

4.7 Unit circle . . . 86
4.8 Ellipse . . . 87
4.9 Eccentricity . . . 91
4.10 Parabola . . . 92
4.11 Hyperbola . . . 96
5.1 Exponential function . . . 108
5.2 Logarithmic function . . . 111
6.1 Component form of a vector . . . 170
6.2 Vector addition . . . 172
6.3 Scalar multiplication . . . 174
6.4 Magnitude and direction of a vector . . . 176
6.5 Unit vector . . . 177
6.6 Dot product . . . 178
6.7 Component form of a vector in three dimensions . . . 182
6.8 Vector addition in three dimensions . . . 182
6.9 Scalar multiplication in three dimensions . . . 183
6.10 Dot product in three dimensions . . . 183
6.11 Cross product . . . 184
7.1 Equations of lines in space . . . 193
7.2 Equations of a plane . . . 196
7.3 Standard equation of a sphere . . . 201
7.4 Cylinder . . . 201
7.5 Quadric surface . . . 203
8.1 The limit of a function f . . . 223
8.2 One-sided limits . . . 232
8.3 Continuous function . . . 237
8.4 Continuity on closed intervals . . . 238
8.5 Limit of infinity . . . 242
8.6 One-sided limits of infinity . . . 243
8.7 Vertical asymptote . . . 245
8.8 Limits at infinity and horizontal asymptotes . . . 246
8.9 Slant asymptotes . . . 250
9.1 Derivative at a point . . . 261
9.2 Tangent line . . . 261
9.3 Normal line . . . 261
9.4 Derivative function . . . 263

9.5 Differentiability on a closed interval . . . 268
9.6 Higher-order derivatives . . . 271
9.7 Smoothness . . . 272
9.8 Differentials of *x* and *y* . . . 298
9.9 Taylor and MacLaurin polynomials . . . 302
9.10 Taylor and Maclaurin series . . . 305
10.1 Extreme values . . . 319
10.2 Relative minimum and relative maximum . . . 320
10.3 Critical numbers and critical points . . . 322
10.4 Singularities and singular points . . . 322
10.5 Increasing and decreasing functions . . . 329
10.6 Concave up and concave down . . . 336
10.7 Point of inflection . . . 338
11.1 The fundamental graphing principle for polar equations . . . 364
11.2 Tangent and normal lines . . . 379
11.3 Smoothness of a parametric curve . . . 382
12.1 Antiderivatives and indefinite integrals . . . 393
12.2 The definite integral, total signed area . . . 397
12.3 Partition . . . 404
12.4 Riemann sum . . . 404
12.5 The average value of *f* on *[a,b]* . . . 413
12.6 Improper integrals with infinite bounds . . . 439
12.7 Improper integrals with infinite range . . . 442
14.1 Vector–valued function . . . 497
14.2 Displacement . . . 500
14.3 Average rate of change . . . 501
14.4 Limits of vector–valued functions . . . 501
14.5 Continuity of vector–valued functions . . . 502
14.6 Derivative of a vector–valued function . . . 503
14.7 Tangent vector and line . . . 506
14.8 Smooth vector–valued function . . . 507
14.9 Antiderivatives, integrals of vector–valued functions . . . 508
14.10 Velocity, speed and acceleration . . . 510
14.11 Distance travelled . . . 513
14.12 Unit tangent vector . . . 514
14.13 Unit normal vector . . . 515

14.14 Arc length parameter 520
14.15 Curvature 520
15.1 Function of two variables 531
15.2 Graph of a function of two variables 532
15.3 Partial functions 535
15.4 Function of three variables 535
15.5 Points and sets 536
15.6 Limit of a function of two variables 539
15.7 Continuity 543
15.8 Open balls, limit and continuity 545
15.9 Partial derivative 546
15.10 Second and mixed partial derivative 548
15.11 Partial derivative with three variables 552
15.12 Total differential 553
15.13 Multivariable differentiability 554
15.14 Total differential 556
15.15 Multivariable differentiability 556
15.16 Directional derivative 562
15.17 Gradient 564
15.18 Directional derivatives and gradient with three variables 570
15.19 Directional tangent line 572
15.20 Normal line 575
15.21 Tangent plane 576
15.22 Relative and absolute extrema 582
15.23 Critical point 582
15.24 Saddle point 586
15.25 Hessian Matrix (of a function of two variables) 589
15.26 Hessian Matrix (of a function of three variables) 591
16.1 Double integral and signed volume 618
16.2 Centre of mass of a discrete linear system 637
16.3 Moments about the x- and y- axes 637
16.4 Centre of mass of a discrete planar system 637
16.5 Centre of mass of a planar lamina 639
16.6 Line integral over a scalar field 646
16.7 Mass and centre of mass of a thin wire 650
16.8 Vector field 651

16.9 Divergence . . . 654
16.10 Curl . . . 654
16.11 Line integral over a vector field . . . 659
16.12 Conservative field . . . 665
16.13 Potential function . . . 666
17.1 Mathematical model . . . 677
17.2 Simulation . . . 678
17.3 Differential equation . . . 681
17.4 Linear differential equation . . . 681
17.5 Autonomous differential equation . . . 682
17.6 Initial value problem . . . 684
17.7 Difference equation . . . 687
17.8 Linear difference equation . . . 688
18.1 Equilibrium point . . . 696
18.2 Stable equilibrium point . . . 699
18.3 Asymptotically stable equilibrium point . . . 699
18.4 Bifurcation point . . . 703

G

List with theorems

3.1 Vertical line test . . . 33
4.1 The factor theorem . . . 75
4.2 The remainder theorem . . . 75
4.3 Zeros and multiplicity . . . 77
4.4 Conjugate pairs theorem . . . 77
4.5 Real factorization theorem . . . 77
5.1 Algebraic properties of exponential functions . . . 114
5.2 Algebraic properties of logarithmic functions . . . 114
5.3 Change of base formulas . . . 115
5.4 The Pythagorean identity . . . 126
5.5 Sine and cosine in the unit circle . . . 128
5.6 Even/Odd identities . . . 132
5.7 Sum and difference identities for cosine . . . 132
5.8 Cofunction identities . . . 133
5.9 Sum and difference identities for sine . . . 134
5.10 Sum and difference identities for tangent . . . 135
5.11 Double angle identities . . . 135
5.12 Power reduction formulas . . . 135
5.13 Product to sum formulas . . . 136
5.14 Sum to product formulas . . . 136

5.15 Classification of conic sections . . . 163
6.1 Principal vector decomposition theorem . . . 178
6.2 Geometric interpretation of dot product . . . 179
6.3 Angle between vectors . . . 180
6.4 Orthogonality of vectors . . . 181
6.5 The cross product and angles . . . 186
8.1 Limits of polynomials and rational functions . . . 226
8.2 Limits of irrational functions . . . 226
8.3 Squeeze theorem . . . 227
8.4 Special limits . . . 229
8.5 Equality of functions and limits . . . 230
8.6 Limits and one sided limits . . . 234
8.7 Intermediate value theorem . . . 240
8.8 Limits of rational functions at infinity . . . 247
9.1 Properties of the derivative . . . 275
9.2 Product rule . . . 276
9.3 Quotient rule . . . 277
9.4 The chain rule . . . 278
9.5 Generalized power rule . . . 278
9.6 Derivatives of inverse functions . . . 288
9.7 L'Hôpital's rule for 0/0 . . . 290
9.8 L'Hôpital's rule for ∞/∞ . . . 291
9.9 Taylor's theorem . . . 303
9.10 Function and Taylor series equality . . . 307
9.11 Algebra of power series . . . 309
9.12 Derivatives and indefinite integrals of Taylor series . . . 309
10.1 The extreme value theorem . . . 321
10.2 Relative extrema and critical points . . . 322
10.3 Fermat's theorem . . . 322
10.4 The mean value theorem of differentiation . . . 327
10.5 Rolle's theorem . . . 327
10.6 Test for increasing/decreasing functions . . . 330
10.7 First derivative test . . . 332
10.8 Test for concavity . . . 338
10.9 Points of inflection . . . 338
10.10 The second derivative test . . . 342

11.1 Conversion between rectangular and polar coordinates 360
12.1 Antiderivative forms 394
12.2 Properties of the antiderivative 396
12.3 Properties of the definite integral 399
12.4 Definite integrals and the limit of Riemann sums 406
12.5 The mean value theorem of integration 407
12.6 The fundamental theorem of calculus, Part 1 408
12.7 The fundamental theorem of calculus, Part 2 409
13.1 Volume by cross-sectional area 467
13.2 Arc length 482
13.3 Surface area of a solid of revolution using rectangular coordinates 487
13.4 Surface area of a solid of revolution using parametric equations 489
13.5 Surface area of a solid of revolution using polar equations 490
14.1 Limits of vector-valued functions 502
14.2 Continuity of vector–valued functions 502
14.3 Derivative of a vector–valued function 504
14.4 Vector–valued functions of constant length 508
14.5 Indefinite and definite integrals of vector–valued functions 509
14.6 Arc length of a vector–valued function 510
14.7 Unit normal vectors in $\mathbb{R}^2$ 517
14.8 Formulas for curvature 521
15.1 Symmetry of second derivatives 549
15.2 Continuity and differentiability of multivariable functions 554
15.3 Differentiability of multivariable functions 555
15.4 Continuity and differentiability of functions of three variables 557
15.5 Multivariable chain rule, Part I 557
15.6 Multivariable chain rule, Part II 559
15.7 Directional derivatives 563
15.8 The gradient and directional derivatives 566
15.9 The gradient and directional derivatives with three variables 570
15.10 Taylor’s theorem for a function of two variables 579
15.11 Necessary Conditions for Extreme Values 585
15.12 Second derivative test 589
15.13 Definiteness of symmetric matrices 589
15.14 Extreme Value Theorem 592
16.1 Area of a plane region 612

16.2 Fubini's theorem 619
16.3 Evaluating a line integral over a scalar field 647
16.4 Evaluating a line integral over a vector field 659
16.5 Fundamental theorem of line integrals 666
16.6 Curl of conservative fields 667
18.1 First derivative test 701
19.1 Existence and uniqueness theorem for linear first-order differential equations 716
19.2 719
19.3 Existence and uniqueness theorem for non-linear first-order differential equations 723
19.4 Exactness condition 729