John Vince

Foundation Mathematics for Computer Science

A Visual Approach

Second Edition

Springer

Foundation Mathematics for Computer Science

John Vince

Foundation Mathematics for Computer Science

A Visual Approach

Second Edition

Springer

John Vince
Bournemouth University
Breinton, Hereford, UK

ISBN 978-3-030-42077-2 ISBN 978-3-030-42078-9 (eBook)
https://doi.org/10.1007/978-3-030-42078-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to my wife and best friend, Heidi.

Preface

Computer science is a very large subject, and graduates will pursue a wide variety of careers, including programming, systems design, cryptography, website design, real-time systems, computer animation, computer games, data visualisation, etc. Consequently, it is virtually impossible to write a mathematics book that caters for all of these potential career paths. Nevertheless, I have attempted to describe a range of mathematical topics that I believe are relevant, and have helped me during my own career in computer science. The book's subtitle "A Visual Approach" reflects the importance I place on coloured illustrations and function graphs, of which there are over 160. Each chapter contains a variety of worked examples.

This second edition remains an introductory text, and is aimed at students studying for an undergraduate degree in computer science. There are four extra chapters on combinatorics, probability, modular arithmetic and complex numbers, which together with the original twelve chapters should provide readers with a solid foundation, upon which more advanced topics of mathematics can be studied.

Throughout the book I have referenced the key people behind the various mathematical discoveries covered, which I hope adds a human dimension to the subject. I have found it very interesting and entertaining to discover how some mathematicians ridiculed their fellow peers, when they could not comprehend the significance of a new invention—Cantor's Set Theory, being an excellent example.

There is no way I could have written this book without the assistance of the Internet and my books previously published by Springer Verlag. In particular, I would like to acknowledge Wikipedia and Richard Elwes' excellent book *Maths 1001*. I prepared this book on an Apple iMac, using LaTeX 2_ε, Pages and the Grapher package, and would recommend this combination to anyone considering writing a book on mathematics. I do hope you enjoy reading this book, and that you are tempted to study mathematics to a deeper level.

Breinton, Herefordshire, UK — John Vince
March 2020

Contents

1 Visual Mathematics 1
1.1 Visual Brains Versus Analytic Brains 1
1.2 Learning Mathematics 2
1.3 What Makes Mathematics Difficult? 2
1.4 Does Mathematics Exist Outside Our Brains? 3
1.5 Symbols and Notation 3

2 Numbers 5
2.1 Introduction 5
2.2 Counting 5
2.3 Sets of Numbers 6
2.4 Zero 7
2.5 Negative Numbers 8
2.5.1 The Arithmetic of Positive and Negative Numbers 9
2.6 Observations and Axioms 10
2.6.1 Commutative Law 10
2.6.2 Associative Law 10
2.6.3 Distributive Law 11
2.7 The Base of a Number System 11
2.7.1 Background 11
2.7.2 Octal Numbers 12
2.7.3 Binary Numbers 13
2.7.4 Hexadecimal Numbers 13
2.7.5 Adding Binary Numbers 17
2.7.6 Subtracting Binary Numbers 18
2.8 Types of Numbers 19
2.8.1 Natural Numbers 19
2.8.2 Integers 19
2.8.3 Rational Numbers 20

2.8.4 Irrational Numbers . . . 20
2.8.5 Real Numbers . . . 20
2.8.6 Algebraic and Transcendental Numbers . . . 20
2.8.7 Imaginary Numbers . . . 21
2.8.8 Complex Numbers . . . 22
2.8.9 Quaternions and Octonions . . . 23
2.8.10 Transcendental and Algebraic Numbers . . . 24
2.9 Prime Numbers . . . 25
2.9.1 The Fundamental Theorem of Arithmetic . . . 26
2.9.2 Is 1 a Prime? . . . 27
2.9.3 Prime Number Distribution . . . 27
2.9.4 Infinity of Primes . . . 28
2.9.5 Perfect Numbers . . . 29
2.9.6 Mersenne Numbers . . . 30
2.10 Infinity . . . 30
2.11 Worked Examples . . . 31
2.11.1 Algebraic Expansion . . . 31
2.11.2 Binary Subtraction . . . 31
2.11.3 Complex Numbers . . . 32
2.11.4 Complex Rotation . . . 32
2.11.5 Quaternions . . . 33
References . . . 33

3 Algebra . . . 35
3.1 Introduction . . . 35
3.2 Background . . . 35
3.3 Notation . . . 36
3.3.1 Solving the Roots of a Quadratic Equation . . . 38
3.4 Indices . . . 41
3.4.1 Laws of Indices . . . 42
3.5 Logarithms . . . 42
3.6 Further Notation . . . 44
3.7 Functions . . . 44
3.7.1 Explicit and Implicit Equations . . . 45
3.7.2 Function Notation . . . 45
3.7.3 Intervals . . . 46
3.7.4 Function Domains and Ranges . . . 47
3.7.5 Odd and Even Functions . . . 48
3.7.6 Power Functions . . . 49
3.8 Worked Examples . . . 50
3.8.1 Algebraic Manipulation . . . 50
3.8.2 Solving a Quadratic Equation . . . 51
3.8.3 Factorising . . . 53

4 Logic 55
4.1 Introduction 55
4.2 Background 55
4.3 Truth Tables 56
4.3.1 Logical Connectives 56
4.4 Logical Premises 57
4.4.1 Material Equivalence 57
4.4.2 Implication 58
4.4.3 Negation 59
4.4.4 Conjunction 59
4.4.5 Inclusive Disjunction 59
4.4.6 Exclusive Disjunction 59
4.4.7 Idempotence 60
4.4.8 Commutativity 61
4.4.9 Associativity 62
4.4.10 Distributivity 63
4.4.11 de Morgan's Laws 63
4.4.12 Simplification 64
4.4.13 Excluded Middle 65
4.4.14 Contradiction 65
4.4.15 Double Negation 66
4.4.16 Implication and Equivalence 66
4.4.17 Exportation 66
4.4.18 Contrapositive 66
4.4.19 Reductio Ad Absurdum 67
4.4.20 Modus Ponens 68
4.4.21 Proof by Cases 69
4.5 Set Theory 70
4.5.1 Empty Set 71
4.5.2 Membership and Cardinality of a Set 71
4.5.3 Subsets, Supersets and the Universal Set 72
4.5.4 Set Building 72
4.5.5 Union 73
4.5.6 Intersection 74
4.5.7 Relative Complement 74
4.5.8 Absolute Complement 75
4.5.9 Power Set 76
4.6 Worked Examples 76
4.6.1 Truth Tables 76
4.6.2 Set Building 76
4.6.3 Sets 78
4.6.4 Power Set 78

5 Combinatorics 79
5.1 Introduction 79
5.2 Permutations 79
5.3 Permutations of Multisets 82
5.4 Combinations 83
5.5 Worked Examples 85
5.5.1 Eight-Permutations of a Multiset 85
5.5.2 Eight-Permutations of a Multiset 86
5.5.3 Number of Permutations 87
5.5.4 Number of Five-Card Hands 87
5.5.5 Hand Shakes with 100 People 87
5.5.6 Permutations of MISSISSIPPI 88

6 Probability 89
6.1 Introduction 89
6.2 Definition and Notation 89
6.2.1 Independent Events 91
6.2.2 Dependent Events 91
6.2.3 Mutually Exclusive Events 92
6.2.4 Inclusive Events 93
6.2.5 Probability Using Combinations 93
6.3 Worked Examples 95
6.3.1 Product of Probabilities 95
6.3.2 Book Arrangements 96
6.3.3 Winning a Lottery 96
6.3.4 Rolling Two Dice 96
6.3.5 Two Dice Sum to 7 96
6.3.6 Two Dice Sum to 4 97
6.3.7 Dealing a Red Ace 97
6.3.8 Selecting Four Aces in Succession 97
6.3.9 Selecting Cards 97
6.3.10 Selecting Four Balls from a Bag 98
6.3.11 Forming Teams 98
6.3.12 Dealing Five Cards 99

7 Modular Arithmetic 101
7.1 Introduction 101
7.2 Informal Definition 101
7.3 Notation 102
7.4 Congruence 102
7.5 Negative Numbers 103
7.6 Arithmetic Operations 103
7.6.1 Sums of Numbers 104
7.6.2 Products 105

7.6.3 Multiplying by a Constant . . . 105
7.6.4 Congruent Pairs . . . 106
7.6.5 Multiplicative Inverse . . . 106
7.6.6 Modulo a Prime . . . 108
7.6.7 Fermat's Little Theorem . . . 109
7.7 Applications of Modular Arithmetic . . . 110
7.7.1 ISBN Parity Check . . . 110
7.7.2 IBAN Check Digits . . . 113
7.8 Worked Examples . . . 115
7.8.1 Negative Numbers . . . 115
7.8.2 Sums of Numbers . . . 115
7.8.3 Remainders of Products . . . 116
7.8.4 Multiplicative Inverse . . . 116
7.8.5 Product Table for Modulo 13 . . . 117
7.8.6 ISBN Check Digit . . . 117
References . . . 118

8 Trigonometry . . . 119
8.1 Introduction . . . 119
8.2 Background . . . 119
8.3 Units of Angular Measurement . . . 119
8.4 The Trigonometric Ratios . . . 120
8.4.1 Domains and Ranges . . . 123
8.5 Inverse Trigonometric Ratios . . . 123
8.6 Trigonometric Identities . . . 125
8.7 The Sine Rule . . . 126
8.8 The Cosine Rule . . . 126
8.9 Compound-Angle Identities . . . 127
8.9.1 Double-Angle Identities . . . 128
8.9.2 Multiple-Angle Identities . . . 129
8.9.3 Half-Angle Identities . . . 130
8.10 Perimeter Relationships . . . 130

9 Coordinate Systems . . . 133
9.1 Introduction . . . 133
9.2 Background . . . 133
9.3 The Cartesian Plane . . . 134
9.4 Function Graphs . . . 134
9.5 Shape Representation . . . 135
9.5.1 2D Polygons . . . 135
9.5.2 Areas of Shapes . . . 136
9.6 Theorem of Pythagoras in 2D . . . 137

9.7 3D Cartesian Coordinates . . . 137
9.7.1 Theorem of Pythagoras in 3D . . . 138
9.8 Polar Coordinates . . . 139
9.9 Spherical Polar Coordinates . . . 139
9.10 Cylindrical Coordinates . . . 140
9.11 Barycentric Coordinates . . . 141
9.12 Homogeneous Coordinates . . . 142
9.13 Worked Examples . . . 142
9.13.1 Area of a Shape . . . 142
9.13.2 Distance Between Two Points . . . 143
9.13.3 Polar Coordinates . . . 143
9.13.4 Spherical Polar Coordinates . . . 144
9.13.5 Cylindrical Coordinates . . . 144
9.13.6 Barycentric Coordinates . . . 145
Reference . . . 145

10 Determinants . . . 147
10.1 Introduction . . . 147
10.2 Background . . . 147
10.3 Linear Equations with Two Variables . . . 148
10.4 Linear Equations with Three Variables . . . 152
10.4.1 Sarrus's Rule . . . 158
10.5 Mathematical Notation . . . 159
10.5.1 Matrix . . . 159
10.5.2 Order of a Determinant . . . 159
10.5.3 Value of a Determinant . . . 159
10.5.4 Properties of Determinants . . . 161
10.6 Worked Examples . . . 162
10.6.1 Determinant Expansion . . . 162
10.6.2 Complex Determinant . . . 162
10.6.3 Simple Expansion . . . 163
10.6.4 Simultaneous Equations . . . 163

11 Vectors . . . 165
11.1 Introduction . . . 165
11.2 Background . . . 165
11.3 2D Vectors . . . 166
11.3.1 Vector Notation . . . 166
11.3.2 Graphical Representation of Vectors . . . 167
11.3.3 Magnitude of a Vector . . . 168
11.4 3D Vectors . . . 169
11.4.1 Vector Manipulation . . . 170
11.4.2 Scaling a Vector . . . 170
11.4.3 Vector Addition and Subtraction . . . 171

11.4.4 Position Vectors . . . 172
11.4.5 Unit Vectors . . . 173
11.4.6 Cartesian Vectors . . . 173
11.4.7 Products . . . 174
11.4.8 Scalar Product . . . 174
11.4.9 The Vector Product . . . 176
11.4.10 The Right-Hand Rule . . . 181
11.5 Deriving a Unit Normal Vector for a Triangle . . . 181
11.6 Surface Areas . . . 182
11.6.1 Calculating 2D Areas . . . 183
11.7 Worked Examples . . . 184
11.7.1 Position Vector . . . 184
11.7.2 Unit Vector . . . 184
11.7.3 Vector Magnitude . . . 184
11.7.4 Angle Between Two Vectors . . . 185
11.7.5 Vector Product . . . 185
Reference . . . 186

12 Complex Numbers . . . 187
12.1 Introduction . . . 187
12.2 Representing Complex Numbers . . . 187
12.2.1 Complex Numbers . . . 187
12.2.2 Real and Imaginary Parts . . . 188
12.2.3 The Complex Plane . . . 188
12.3 Complex Algebra . . . 188
12.3.1 Algebraic Laws . . . 188
12.3.2 Complex Conjugate . . . 190
12.3.3 Complex Division . . . 192
12.3.4 Powers of i . . . 193
12.3.5 Rotational Qualities of i . . . 194
12.3.6 Modulus and Argument . . . 196
12.3.7 Complex Norm . . . 198
12.3.8 Complex Inverse . . . 199
12.3.9 Complex Exponentials . . . 200
12.3.10 de Moivre's Theorem . . . 204
12.3.11 nth Root of Unity . . . 206
12.3.12 nth Roots of a Complex Number . . . 207
12.3.13 Logarithm of a Complex Number . . . 208
12.3.14 Raising a Complex Number to a Complex Power . . . 209
12.3.15 Visualising Simple Complex Functions . . . 212
12.3.16 The Hyperbolic Functions . . . 215
12.4 Summary . . . 216
12.5 Worked Examples . . . 217

12.5.1 Complex Addition . . . 217
12.5.2 Complex Products . . . 217
12.5.3 Complex Division . . . 217
12.5.4 Complex Rotation . . . 218
12.5.5 Polar Notation . . . 218
12.5.6 Real and Imaginary Parts . . . 219
12.5.7 Magnitude of a Complex Number . . . 219
12.5.8 Complex Norm . . . 220
12.5.9 Complex Inverse . . . 220
12.5.10 de Moivre's Theorem . . . 220
12.5.11 *n*th Root of Unity . . . 222
12.5.12 Roots of a Complex Number . . . 222
12.5.13 Logarithm of a Complex Number . . . 223
12.5.14 Raising a Number to a Complex Power . . . 223
References . . . 224

13 Matrices . . . 225
13.1 Introduction . . . 225
13.2 Geometric Transforms . . . 225
13.3 Transforms and Matrices . . . 227
13.4 Matrix Notation . . . 230
13.4.1 Matrix Dimension or Order . . . 230
13.4.2 Square Matrix . . . 230
13.4.3 Column Vector . . . 231
13.4.4 Row Vector . . . 231
13.4.5 Null Matrix . . . 231
13.4.6 Unit Matrix . . . 231
13.4.7 Trace . . . 232
13.4.8 Determinant of a Matrix . . . 233
13.4.9 Transpose . . . 233
13.4.10 Symmetric Matrix . . . 234
13.4.11 Antisymmetric Matrix . . . 236
13.5 Matrix Addition and Subtraction . . . 238
13.5.1 Scalar Multiplication . . . 238
13.6 Matrix Products . . . 239
13.6.1 Row and Column Vectors . . . 239
13.6.2 Row Vector and a Matrix . . . 240
13.6.3 Matrix and a Column Vector . . . 241
13.6.4 Square Matrices . . . 241
13.6.5 Rectangular Matrices . . . 242
13.7 Inverse Matrix . . . 243
13.7.1 Inverting a Pair of Matrices . . . 249
13.8 Orthogonal Matrix . . . 250
13.9 Diagonal Matrix . . . 251

13.10 Worked Examples . . . 251
13.10.1 Matrix Inversion . . . 251
13.10.2 Identity Matrix . . . 252
13.10.3 Solving Two Equations Using Matrices . . . 253
13.10.4 Solving Three Equations Using Matrices . . . 254
13.10.5 Solving Two Complex Equations . . . 255
13.10.6 Solving Three Complex Equations . . . 255
13.10.7 Solving Two Complex Equations . . . 256
13.10.8 Solving Three Complex Equations . . . 257

14 Geometric Matrix Transforms . . . 259
14.1 Introduction . . . 259
14.2 Matrix Transforms . . . 259
14.2.1 2D Translation . . . 260
14.2.2 2D Scaling . . . 261
14.2.3 2D Reflections . . . 263
14.2.4 2D Shearing . . . 264
14.2.5 2D Rotation . . . 265
14.2.6 2D Scaling . . . 268
14.2.7 2D Reflection . . . 268
14.2.8 2D Rotation About an Arbitrary Point . . . 269
14.3 3D Transforms . . . 270
14.3.1 3D Translation . . . 270
14.3.2 3D Scaling . . . 271
14.3.3 3D Rotation . . . 271
14.3.4 Rotating About an Axis . . . 274
14.3.5 3D Reflections . . . 276
14.4 Rotating a Point About an Arbitrary Axis . . . 276
14.4.1 Matrices . . . 276
14.5 Determinant of a Transform . . . 279
14.6 Perspective Projection . . . 280
14.7 Worked Examples . . . 282
14.7.1 2D Scale and Translate . . . 282
14.7.2 2D Rotation . . . 283
14.7.3 Determinant of the Rotate Transform . . . 284
14.7.4 Determinant of the Shear Transform . . . 284
14.7.5 Yaw, Pitch and Roll Transforms . . . 285
14.7.6 Rotation About an Arbitrary Axis . . . 285
14.7.7 3D Rotation Transform Matrix . . . 286
14.7.8 Perspective Projection . . . 287

15 Calculus: Derivatives . . . 289
15.1 Introduction . . . 289
15.2 Background . . . 289

15.3 Small Numerical Quantities 290
15.4 Equations and Limits 291
15.4.1 Quadratic Function 291
15.4.2 Cubic Equation 293
15.4.3 Functions and Limits 294
15.4.4 Graphical Interpretation of the Derivative 296
15.4.5 Derivatives and Differentials 297
15.4.6 Integration and Antiderivatives 298
15.5 Function Types 299
15.6 Differentiating Groups of Functions 300
15.6.1 Sums of Functions 300
15.6.2 Function of a Function 302
15.6.3 Function Products 306
15.6.4 Function Quotients 309
15.7 Differentiating Implicit Functions 311
15.8 Differentiating Exponential and Logarithmic Functions 314
15.8.1 Exponential Functions 314
15.8.2 Logarithmic Functions 317
15.9 Differentiating Trigonometric Functions 318
15.9.1 Differentiating tan 318
15.9.2 Differentiating csc 320
15.9.3 Differentiating sec 321
15.9.4 Differentiating cot 322
15.9.5 Differentiating arcsin, arccos and arctan 323
15.9.6 Differentiating arccsc, arcsec and arccot 324
15.10 Differentiating Hyperbolic Functions 324
15.10.1 Differentiating sinh, cosh and tanh 326
15.11 Higher Derivatives 327
15.12 Higher Derivatives of a Polynomial 328
15.13 Identifying a Local Maximum or Minimum 330
15.14 Partial Derivatives 332
15.14.1 Visualising Partial Derivatives 335
15.14.2 Mixed Partial Derivatives 336
15.15 Chain Rule 338
15.16 Total Derivative 340
15.17 Power Series 342
15.18 Worked Examples 344
15.18.1 Antiderivative 1 344
15.18.2 Antiderivative 2 345
15.18.3 Differentiating Sums of Functions 345
15.18.4 Differentiating a Function Product 345
15.18.5 Differentiating an Implicit Function 346
15.18.6 Differentiating a General Implicit Function 346

15.18.7 Local Maximum or Minimum . . . 347
15.18.8 Partial Derivatives . . . 348
15.18.9 Mixed Partial Derivative 1 . . . 348
15.18.10 Mixed Partial Derivative 2 . . . 349
15.18.11 Total Derivative . . . 349

16 Calculus: Integration . . . 351
16.1 Introduction . . . 351
16.2 Indefinite Integral . . . 351
16.3 Integration Techniques . . . 352
16.3.1 Continuous Functions . . . 352
16.3.2 Difficult Functions . . . 353
16.4 Trigonometric Identities . . . 354
16.4.1 Exponent Notation . . . 356
16.4.2 Completing the Square . . . 357
16.4.3 The Integrand Contains a Derivative . . . 359
16.4.4 Converting the Integrand into a Series of Fractions . . . 360
16.4.5 Integration by Parts . . . 361
16.4.6 Integration by Substitution . . . 366
16.4.7 Partial Fractions . . . 368
16.5 Area Under a Graph . . . 370
16.6 Calculating Areas . . . 370
16.7 Positive and Negative Areas . . . 378
16.8 Area Between Two Functions . . . 380
16.9 Areas with the y-Axis . . . 382
16.10 Area with Parametric Functions . . . 383
16.11 The Riemann Sum . . . 385
16.12 Worked Examples . . . 386
16.12.1 Integrating a Function Containing Its Own Derivative . . . 386
16.12.2 Dividing an Integral into Several Integrals . . . 387
16.12.3 Integrating by Parts 1 . . . 388
16.12.4 Integrating by Parts 2 . . . 388
16.12.5 Integrating by Substitution 1 . . . 390
16.12.6 Integrating by Substitution 2 . . . 390
16.12.7 Integrating by Substitution 3 . . . 392
16.12.8 Integrating with Partial Fractions . . . 392

Appendix A: Limit of $(\sin\theta)/\theta$ . . . 395

Appendix B: Integrating $\cos^n\theta$ . . . 399

Index . . . 401

Chapter 1
Visual Mathematics

1.1 Visual Brains Versus Analytic Brains

I consider myself a *visual* person, as pictures help me understand complex problems. I also don't find it too difficult to visualise objects from different view points. I remember learning about electrons, neutrons and protons for the first time, where our planetary system provided a simple model to visualise the hidden structure of matter. My mental image of electrons was one of small orange spheres, spinning around a small, central nucleus containing blue protons and grey neutrons. And although this visual model was seriously flawed, it provided a first step towards understanding the structure of matter.

As my knowledge of mathematics grew, this, too, was image based. Equations were curves and surfaces, simultaneous equations were intersecting or parallel lines, etc., and when I embarked upon computer science, I found a natural application for mathematics. For me, mathematics is a visual science, although I do appreciate that many professional mathematicians need only a formal, symbolic notation for constructing their world. Such people do not require visual scaffolding—they seem to be able to manipulate abstract mathematical concepts at a symbolic level. Their books do not require illustrations or diagrams—Greek symbols, upside-down and back-to-front Latin fonts are sufficient to annotate their ideas.

Today, when reading popular science books on quantum theory, I still try to form images of 3D fields of energy and probability oscillating in space—to no avail—and I have accepted that human knowledge of such phenomena is best left to a mathematical description. Nevertheless, mathematicians, such as Sir Roger Penrose, know the importance of visual models in communicating complex mathematical ideas. His book *The Road to Reality: A Complete Guide to the Laws of the Universe* is decorated with beautiful, informative, hand-drawn illustrations, which help readers understand the mathematics of science. In this book I rely heavily on images to communicate an idea. They are simple and are the first step on a ladder towards understanding a difficult idea. Eventually, when that *Eureka* moment arrives, that moment when

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_1

"I understand what you are saying," the image becomes closely associated with the mathematical notation.

1.2 Learning Mathematics

I was fortunate in my studies in that I was taught by people interested in mathematics, and their interest rubbed off on me. I feel sorry for children who have given up on mathematics, simply because they are being taught by teachers whose primary subject is not mathematics. I was never too concerned about the uses of mathematics, although applied mathematics is of special interest.

One of the problems with mathematics is its incredible breadth and depth. It embraces everything from 2D geometry, calculus, topology, statistics, complex functions to number theory and propositional calculus. All of these subjects can be studied superficially or to a mind-numbing complexity. Fortunately, no one is required to understand everything, which is why mathematicians tend to specialise in one or two areas and develop a specialist knowledge.

1.3 What Makes Mathematics Difficult?

"What makes mathematics difficult?" is also a difficult question to answer, but one that has to be asked and answered. There are many answers to this question, and I believe that problems begin with mathematical notation and how to read it; how to analyse a problem and express a solution using mathematical statements. Unlike learning a foreign language—which I find very difficult—mathematics is a language that needs to be learned by discovering facts and building upon them to discover new facts. Consequently, a good memory is always an advantage, as well as a sense of logic.

Mathematics can be difficult for anyone, including mathematicians. For example, when the idea of $\sqrt{-1}$ was originally proposed, it was criticised and looked down upon by mathematicians, mainly because its purpose was not fully understood. Eventually, it transformed the entire mathematical landscape, including physics. Similarly, when the German mathematician Georg Cantor (1845–1919), published his papers on set theory and transfinite sets, some mathematicians hounded him in a disgraceful manner. The German mathematician Leopold Kronecker (1823–1891), called Cantor a "scientific charlatan", a "renegade", and a "corrupter of youth", and did everything to hinder Cantor's academic career. Similarly, the French mathematician and physicist Henri Poincaré (1854–1912), called Cantor's ideas a "grave disease", whilst the Austrian-British philosopher and logician Ludwig Wittgenstein (1889–1951) complained that mathematics is "ridden through and through with the pernicious idioms of set theory." How wrong they all were. Today, set theory is a major branch of mathematics and has found its way into every math curriculum. So don't be surprised to

discover that some mathematical ideas are initially difficult to understand—you are in good company.

1.4 Does Mathematics Exist Outside Our Brains?

Many people have considered the question "What is mathematics?" Some mathematicians and philosophers argue that numbers and mathematical formulae have some sort of external existence and are waiting to be discovered by us. Personally, I don't accept this idea. I believe that we enjoy searching for patterns and structure in anything that finds its way into our brains, which is why we love poetry, music, storytelling, art, singing, architecture, science, as well as mathematics. The piano, for example, is an instrument for playing music using different patterns of notes. When the piano was invented—a few hundred years ago—the music of Chopin, Liszt and Rachmaninoff did not exist in any form—it had to be composed by them. Similarly, by building a system for counting using numbers, we have an amazing tool for composing mathematical systems that help us measure quantity, structure, space and change. Such systems have been applied to topics such as fluid dynamics, optimisation, statistics, cryptography, game theory probability theory, and many more. I will attempt to develop this same idea by showing how the concept of number, and the visual representation of number reveals all sorts of patterns, that give rise to number systems, algebra, trigonometry, geometry, analytic geometry and calculus. The universe does not need any of these mathematical ideas to run its machinery, but we need these ideas to understand its operation.

1.5 Symbols and Notation

One of the reasons why many people find mathematics inaccessible is due to its symbols and notation. Let's look at symbols first. The English alphabet possesses a reasonable range of familiar character shapes:

a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

which find their way into every branch of mathematics and physics, and permit us to write equations such as

$$E = mc^2$$

and

$$A = \pi r^2.$$

It is important that when we see an equation, we are able to read it as part of the text. In the case of $E = mc^2$, this is read as "E equals m, c squared", where E stands for energy, m for mass, and c the speed of light. In the case of $A = \pi r^2$, this is read as "A equals pi, r squared", where A stands for area, π the ratio of a circle's circumference to its diameter, and r the circle's radius. Greek symbols, which happen to look nice and impressive, have also found their way into many equations, and often disrupt the flow of reading, simply because we don't know their English names. For example, the English theoretical physicist Paul Dirac (1902–1984) derived an equation for a moving electron using the symbols α_i and β, which are 4×4 matrices, where

$$\alpha_i \beta + \beta \alpha_i = 0$$

and is read as

"the sum of the products alpha-i beta, and beta alpha-i, equals zero."

Although we will not come across moving electrons in this book, we will have to be familiar with the following Greek symbols:

α	alpha	ν	nu
β	beta	ξ	xi
γ	gamma	o	o
δ	delta	π	pi
ϵ	epsilon	ρ	rho
ζ	zeta	σ	sigma
η	eta	τ	tau
θ	theta	υ	upsilon
ι	iota	ϕ	phi
κ	kappa	χ	chi
λ	lambda	ψ	psi
μ	mu	ω	omega

and some upper-case symbols:

Γ	Gamma	Σ	Sigma
Δ	Delta	Υ	Upsilon
Θ	Theta	Φ	Phi
Λ	Lambda	Ψ	Psi
Ξ	Xi	Ω	Omega
Π	Pi.		

Being able to read an equation does not mean that we understand it—but we are a little closer than just being able to stare at a jumble of symbols! Therefore, in future, when I introduce a new mathematical object, I will tell you how it should be read.

Chapter 2
Numbers

2.1 Introduction

This chapter revises the sets of numbers employed in mathematics such as natural, integer, rational, irrational, real, algebraic, transcendental, imaginary, complex, quaternions and octonions. It also describes how these numbers behave in the context of three laws: commutative law, associative law and the distributive law. Apart from the every-day base of 10, the three important bases in computer science are covered: binary, octal and hexadecimal.

As prime numbers find their way into all aspects of cryptography, the chapter introduces the fundamental theorem of arithmetic, prime number distribution, perfect numbers and Mersenne numbers. The chapter concludes with the concept of infinity and some worked examples.

2.2 Counting

Our brain's visual cortex possesses some incredible image processing features. For example, children know instinctively when they are given less sweets than another child, and adults know instinctively when they are short-changed by a Parisian taxi driver, or driven around the Arc de Triumph several times, on the way to the airport! Intuitively, we can assess how many donkeys are in a field without counting them, and generally, we seem to know within a second or two, whether there are just a few, dozens, or hundreds of something. But when accuracy is required, one can't beat counting. But what is counting?

Well normally, we are taught to count by our parents by memorising first, the counting words *one, two, three, four, five, six, seven, eight, nine, ten, ..* and second, associating them with our fingers, so that when asked to count the number of donkeys in a picture book, each donkey is associated with a counting word. When each donkey has been identified, the number of donkeys equals the last word mentioned.

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_2

However, this still assumes that we know the meaning of *one, two, three, four, ..* etc. Memorising these counting words is only part of the problem—getting them in the correct sequence is the real challenge. The incorrect sequence *one, two, five, three, nine, four, ..* etc., introduces an element of randomness into any calculation, but practice makes perfect, and it's useful to master the correct sequence before going to university!

2.3 Sets of Numbers

A *set* is a collection of distinct objects called its *elements* or *members*. For example, each system of number belongs to a set with given a name, such as $\mathbb{N}$ for the natural numbers, $\mathbb{R}$ for real numbers, and $\mathbb{Q}$ for rational numbers. When we want to indicate that something is whole, real or rational, etc., we use the notation

$$n \in \mathbb{N}$$

which reads "n is a member of ($\in$) the set $\mathbb{N}$", i.e. n is a whole number. Similarly,

$$x \in \mathbb{R}$$

stands for "x is a real number."

A *well-ordered set* possesses a unique order, such as the natural numbers $\mathbb{N}$. Therefore, if P is the well-ordered set of prime numbers and $\mathbb{N}$ is the well-ordered set of natural numbers, we can write

$$P = \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, \ldots\}$$
$$\mathbb{N} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, \ldots\}.$$

By pairing the prime numbers in P with the numbers in $\mathbb{N}$, we have

$$\{2, 1\}, \{3, 2\}, \{5, 3\}, \{7, 4\}, \{11, 5\}, \{13, 6\}, \{17, 7\}, \{19, 8\}, \{23, 9\}, \ldots$$

and we can reason that 2 is the 1st prime, and 3 is the 2nd prime, etc. However, we still have to declare what we mean by $1, 2, 3, 4, 5, \ldots$ etc., and without getting too philosophical, I like the idea of defining them as follows. The word *one*, represented by 1, stands for one-ness of anything: one finger, one house, one tree, one donkey, etc. The word *two*, represented by 2, is "one more than one". The word *three*, represented by 3, is "one more than two", and so on.

We are now in a position to associate some mathematical notation with our numbers by introducing the $+$ and $=$ signs. We know that $+$ means *add*, but it also can stand for *more*. We also know that $=$ means *equal*, and it can also stand for *is the same as*. Thus the statement

$$2 = 1 + 1$$

is read as "two is the same as one more than one."

We can also write

$$3 = 1 + 2$$

which is read as "three is the same as one more than two." But as we already have a definition for 2, we can write

$$\begin{aligned} 3 &= 1 + 2 \\ &= 1 + 1 + 1. \end{aligned}$$

Developing this idea, and including some extra combinations, we have

$$\begin{aligned} 2 &= 1 + 1 \\ 3 &= 1 + 2 \\ 4 &= 1 + 3 = 2 + 2 \\ 5 &= 1 + 4 = 2 + 3 \\ 6 &= 1 + 5 = 2 + 4 = 3 + 3 \\ 7 &= 1 + 6 = 2 + 5 = 3 + 4 \\ &\text{etc.} \end{aligned}$$

and can be continued without limit. The numbers, 1, 2, 3, 4, 5, 6, etc., are called *natural numbers*, and are the set $\mathbb{N}$.

2.4 Zero

The concept of zero has a well-documented history, which shows that it has been used by different cultures over a period of two-thousand years or more. It was the Indian mathematician and astronomer Brahmagupta (598-c.–670) who argued that zero was just as valid as any natural number, with the definition: *the result of subtracting any number from itself*. However, even today, there is no universal agreement as to whether zero belongs to the set $\mathbb{N}$, consequently, the set $\mathbb{N}^0$ stands for the set of natural numbers including zero.

In today's positional decimal system, which is a *place value system*, the digit 0 is a placeholder. For example, 203 stands for: two hundreds, no tens and three units. Although $0 \in \mathbb{N}^0$, it does have special properties that distinguish it from other members of the set, and Brahmagupta also gave rules showing this interaction.

If $x \in \mathbb{N}^0$, then the following rules apply:

$$\begin{aligned} \text{addition:} \quad & x + 0 = x \\ \text{subtraction:} \quad & x - 0 = x \\ \text{multiplication:} \quad & x \times 0 = 0 \times x = 0 \\ \text{division:} \quad & 0/x = 0 \\ \text{undefined division:} \quad & x/0. \end{aligned}$$

The expression $0/0$ is called an *indeterminate form*, as it is possible to show that under different conditions, especially limiting conditions, it can equal anything. So for the moment, we will avoid using it until we cover calculus.

2.5 Negative Numbers

When negative numbers were first proposed, they were not accepted with open arms, as it was difficult to visualise -5 of something. For instance, if there are 5 donkeys in a field, and they are all stolen to make salami, the field is now empty, and there is nothing we can do in the arithmetic of donkeys to create a field of -5 donkeys. However, in applied mathematics, numbers have to represent all sorts of quantities such as temperature, displacement, angular rotation, speed, acceleration, etc., and we also need to incorporate ideas such as left and right, up and down, before and after, forwards and backwards, etc. Fortunately, negative numbers are perfect for representing all of the above quantities and ideas.

Consider the expression $4 - x$, where $x \in \mathbb{N}^0$. When x takes on certain values, we have

$$\begin{aligned} 4 - 1 &= 3 \\ 4 - 2 &= 2 \\ 4 - 3 &= 1 \\ 4 - 4 &= 0 \end{aligned}$$

and unless we introduce negative numbers, we are unable to express the result of $4 - 5$. Consequently, negative numbers are visualised as shown in Fig. 2.1, where the *number line* shows negative numbers to the left of the natural numbers, which are *positive*, although the $+$ sign is omitted for clarity.

Fig. 2.1 The number line showing negative and positive numbers

Moving from left to right, the number line provides a numerical continuum from large negative numbers, through zero, towards large positive numbers. In any calculation we could agree that angles above the horizon are positive, and angles below the horizon, negative. Similarly, a movement forwards is positive, and a movement backwards is negative. So now we are able to write

$$4-5=-1$$
$$4-6=-2$$
$$4-7=-3$$
$$\text{etc.},$$

without worrying about creating impossible conditions.

2.5.1 The Arithmetic of Positive and Negative Numbers

Once again, Brahmagupta compiled all the rules, Tables 2.1 and 2.2, supporting the addition, subtraction, multiplication and division of positive and negative numbers. The real fly in the ointment, being negative numbers, which cause problems for children, math teachers and occasional accidents for mathematicians. Perhaps, the one rule we all remember from our school days is that "two negatives make a positive".

Another problem with negative numbers arises when we employ the square-root function. As the product of two positive or negative numbers results in a positive result, the square-root of a positive number gives rise to a positive **and** a negative answer. For example, $\sqrt{4}=\pm 2$. This means that the square-root function only applies to positive numbers. Nevertheless, it did not stop the invention of the *imaginary* unit i, where $i^2=-1$. However, i is not a number, but an operator, which is described later.

Table 2.1 Rules for adding and subtracting positive and negative numbers

+	b	$-b$	−	b	$-b$
a	$a+b$	$a-b$	a	$a-b$	$a+b$
$-a$	$b-a$	$-(a+b)$	$-a$	$-(a+b)$	$b-a$

Table 2.2 Rules for multiplying and dividing positive and negative numbers

×	b	$-b$	/	b	$-b$
a	ab	$-ab$	a	a/b	$-a/b$
$-a$	$-ab$	ab	$-a$	$-a/b$	a/b

2.6 Observations and Axioms

The following *axioms* or laws provide a formal basis for mathematics, and in the descriptions a *binary operation* is an arithmetic operation such as $+, -, \times, /$ which operates on two operands.

2.6.1 Commutative Law

The *commutative law* in algebra states that when two elements are linked through some binary operation, the result is independent of the order of the elements. The commutative law of addition is

$$a + b = b + a$$
$$\text{e.g. } 1 + 2 = 2 + 1.$$

The commutative law of multiplication is

$$a \times b = b \times a$$
$$\text{e.g. } 1 \times 2 = 2 \times 1.$$

Note that subtraction is not commutative

$$a - b \neq b - a$$
$$\text{e.g. } 1 - 2 \neq 2 - 1.$$

2.6.2 Associative Law

The *associative law* in algebra states that when three or more elements are linked together through a binary operation, the result is independent of how each pair of elements is grouped. The associative law of addition is

$$a + (b + c) = (a + b) + c$$
$$\text{e.g. } 1 + (2 + 3) = (1 + 2) + 3.$$

The associative law of multiplication is

$$a \times (b \times c) = (a \times b) \times c$$
$$\text{e.g. } 1 \times (2 \times 3) = (1 \times 2) \times 3.$$

However, note that subtraction is not associative

$$a - (b - c) \neq (a - b) - c$$
$$\text{e.g. } 1 - (2 - 3) \neq (1 - 2) - 3.$$

which may seem surprising, but at the same time confirms the need for clear axioms.

2.6.3 Distributive Law

The *distributive law* in algebra describes an operation which when performed on a combination of elements is the same as performing the operation on the individual elements. The distributive law does not work in all cases of arithmetic. For example, multiplication over addition holds

$$a(b + c) = ab + ac$$
$$\text{e.g. } 2(3 + 4) = 6 + 8,$$

whereas addition over multiplication does not:

$$a + (b \times c) \neq (a + b) \times (a + c)$$
$$\text{e.g. } 3 + (4 \times 5) \neq (3 + 4) \times (3 + 5).$$

Although these laws are natural for numbers, they do not necessarily apply to all mathematical objects. For instance, the vector product, which multiplies two vectors together, is not commutative. The same applies for matrix multiplication.

2.7 The Base of a Number System

2.7.1 Background

Over recent millennia, mankind has invented and discarded many systems for representing number. People have counted on their fingers and toes, used pictures (hieroglyphics), cut marks on clay tablets (cuneiform symbols), employed Greek symbols (Ionic system) and struggled with, and abandoned Roman numerals (I, V, X, L, C, D, M, etc.), until we reach today's decimal place system, which has Hindu-Arabic and Chinese origins. And since the invention of computers, we have witnessed the emergence of binary, octal and hexadecimal number systems, where 2, 8 and 16 respectively, replace the 10 in our decimal system.

The decimal number 23 means "two tens and three units", and in English is written "twenty-three", in French "vingt-trois" (twenty-three), and in German "dreiundzwanzig" (three and twenty). Let's investigate the algebra behind the decimal system and see how it can be used to represent numbers to any base. The expression

$$a \times 1000 + b \times 100 + c \times 10 + d \times 1$$

where a, b, c, d take on any value between 0 and 9, describes any whole number between 0 and 9999. By including

$$e \times 0.1 + f \times 0.01 + g \times 0.001 + h \times 0.0001$$

where e, f, g, h take on any value between 0 and 9, any decimal number between 0 and 9999.9999 can be represented.

Indices bring the notation alive and reveal the true underlying pattern:

$$\ldots a10^3 + b10^2 + c10^1 + d10^0 + e10^{-1} + f10^{-2} + g10^{-3} + h10^{-4} \ldots.$$

Remember that any number raised to the power 0 equals 1. By adding extra terms both left and right, any number can be accommodated.

In this example, 10 is the base, which means that the values of a to h range between 0 and 9, 1 less than the base. Therefore, by substituting B for the base we have

$$\ldots aB^3 + bB^2 + cB^1 + dB^0 + eB^{-1} + fB^{-2} + gB^{-3} + hB^{-4} \ldots$$

where the values of a to h range between 0 and $B - 1$.

2.7.2 Octal Numbers

The octal number system has $B = 8$, and a to h range between 0 and 7

$$\ldots a8^3 + b8^2 + c8^1 + d8^0 + e8^{-1} + f8^{-2} + g8^{-3} + h8^{-4} \ldots$$

and the first 17 octal numbers are

$$1_8, 2_8, 3_8, 4_8, 5_8, 6_8, 7_8, 10_8, 11_8, 12_8, 13_8, 14_8, 15_8, 16_8, 17_8, 20_8, 21_8.$$

The subscript 8, reminds us that although we may continue to use the words "twenty-one", it is an octal number, and not a decimal. But what is 14_8 in decimal? Well, it stands for

$$1 \times 8^1 + 4 \times 8^0 = 12.$$

Thus 356.4_8 in decimal, equals

$$\begin{gathered}(3 \times 8^2) + (5 \times 8^1) + (6 \times 8^0) + (4 \times 8^{-1}) \\ (3 \times 64) + (5 \times 8) + (6 \times 1) + (4 \times 0.125) \\ (192 + 40 + 6) + (0.5) \\ 238.5.\end{gathered}$$

Counting in octal appears difficult, simply because we have never been exposed to it, like the decimal system. If we had evolved with 8 fingers, instead of 10, we would be counting in octal!

2.7.3 Binary Numbers

The binary number system has $B = 2$, and a to h are 0 or 1

$$\ldots a2^3 + b2^2 + c2^1 + d2^0 + e2^{-1} + f2^{-2} + g2^{-3} + h2^{-4} \ldots$$

and the first 13 binary numbers are

$$1_2, 10_2, 11_2, 100_2, 101_2, 110_2, 111_2, 1000_2, 1001_2, 1010_2, 1011_2, 1100_2, 1101_2.$$

Thus 11011.11_2 in decimal, equals

$$\begin{gathered}(1 \times 2^4) + (1 \times 2^3) + (0 \times 2^2) + (1 \times 2^1) + (1 \times 2^0) + (1 \times 2^{-1}) + (1 \times 2^{-2}) \\ (1 \times 16) + (1 \times 8) + (0 \times 4) + (1 \times 2) + (1 \times 0.5) + (1 \times 0.25) \\ (16 + 8 + 2) + (0.5 + 0.25) \\ 26.75.\end{gathered}$$

The reason why computers work with binary numbers—rather than decimal—is due to the difficulty of designing electrical circuits that can store decimal numbers in a stable fashion. A switch, where the open state represents 0, and the closed state represents 1, is the simplest electrical component to emulate. No matter how often it is used, or how old it becomes, it will always behave like a switch. The main advantage of electrical circuits is that they can be switched on and off trillions of times a second, and the only disadvantage is that the encoded binary numbers and characters contain a large number of bits, and humans are not familiar with binary.

2.7.4 Hexadecimal Numbers

The hexadecimal number system has $B = 16$, and a to h can be 0 to 15, which presents a slight problem, as we don't have 15 different numerical characters. Consequently,

we use 0 to 9, and the letters A, B, C, D, E, F to represent 10, 11, 12, 13, 14, 15 respectively

$$\ldots a16^3 + b16^2 + c16^1 + d16^0 + e16^{-1} + f16^{-2} + g16^{-3} + h16^{-4} \ldots$$

and the first 17 hexadecimal numbers are

$$1_{16}, 2_{16}, 3_{16}, 4_{16}, 5_{16}, 6_{16}, 7_{16}, 8_{16}, 9_{16}, A_{16}, B_{16}, C_{16}, D_{16}, E_{16}, F_{16}, 10_{16}, 11_{16}.$$

Thus $1E.8_{16}$ in decimal, equals

$$\begin{gathered}(1 \times 16) + (E \times 1) + (8 \times 16^{-1}) \\ (16 + 14) + (8/16) \\ 30.5.\end{gathered}$$

Although it is not obvious, binary, octal and hexadecimal numbers are closely related, which is why they are part of a programmer's toolkit. Even though computers work with binary, it's the last thing a programmer wants to use. So to simplify the man-machine interface, binary is converted into octal or hexadecimal. To illustrate this, let's convert the 16-bit binary code 1101011000110001 into octal.

Using the following general binary integer

$$a2^8 + b2^7 + c2^6 + d2^5 + e2^4 + f2^3 + g2^2 + h2^1 + i2^0$$

we group the terms into threes, starting from the right, because $2^3 = 8$

$$(a2^8 + b2^7 + c2^6) + (d2^5 + e2^4 + f2^3) + (g2^2 + h2^1 + i2^0).$$

Simplifying

$$\begin{gathered}2^6(a2^2 + b2^1 + c2^0) + 2^3(d2^2 + e2^1 + f2^0) + 2^0(g2^2 + h2^1 + i2^0) \\ 8^2(a2^2 + b2^1 + c2^1) + 8^1(d2^2 + e2^1 + f2^0) + 8^0(g2^2 + h2^1 + i2^0) \\ 8^2R + 8^1S + 8^0T\end{gathered}$$

where

$$\begin{aligned}R &= a2^2 + b2^1 + c \\ S &= d2^2 + e2^1 + f \\ T &= g2^2 + h2^1 + i\end{aligned}$$

and the values of R, S, T vary between 0 and 7. Therefore, given 1101011000110001, we divide the binary code into groups of three, starting at the right, and adding two leading zeros

$$(001)(101)(011)(000)(110)(001).$$

For each group, multiply the zeros and ones by 4, 2, 1, right to left

$$(0+0+1)(4+0+1)(0+2+1)(0+0+0)(4+2+0)(0+0+1)$$
$$(1)(5)(3)(0)(6)(1)$$
$$153061_8.$$

Therefore, $1101011000110001_2 \equiv 153061_8$, ($\equiv$ stands for "equivalent to") which is much more compact. The secret of this technique is to memorise the patterns

$$000_2 \equiv 0_8$$
$$001_2 \equiv 1_8$$
$$010_2 \equiv 2_8$$
$$011_2 \equiv 3_8$$
$$100_2 \equiv 4_8$$
$$101_2 \equiv 5_8$$
$$110_2 \equiv 6_8$$
$$111_2 \equiv 7_8.$$

Here are a few more examples, with the binary digits grouped in threes:

$$111_2 \equiv 7_8$$
$$101\ 101_2 \equiv 55_8$$
$$100\ 000_2 \equiv 40_8$$
$$111\ 000\ 111\ 000\ 111_2 \equiv 70707_8.$$

It's just as easy to reverse the process, and convert octal into binary. Here are some examples:

$$567_8 \equiv 101\ 110\ 111_2$$
$$23_8 \equiv 010\ 011_2$$
$$1741_8 \equiv 001\ 111\ 100\ 001_2.$$

A similar technique is used to convert binary to hexadecimal, but this time we divide the binary code into groups of four, because $2^4 = 16$, starting at the right, and adding leading zeros, if necessary. To illustrate this, let's convert the 16-bit binary code 1101 0110 0011 0001 into hexadecimal.

Using the following general binary integer number

$$a2^{11} + b2^{10} + c2^9 + d2^8 + e2^7 + f2^6 + g2^5 + h2^4 + i2^3 + j2^2 + k2^1 + l2^0$$

from the right, we divide the binary code into groups of four:

$$(a2^{11} + b2^{10} + c2^9 + d2^8) + (e2^7 + f2^6 + g2^5 + h2^4) + (i2^3 + j2^2 + k2^1 + l2^0).$$

Simplifying

$$2^8(a2^3 + b2^2 + c2^1 + d2^0) + 2^4(e2^3 + f2^2 + g2^1 + h2^0) + 2^0(i2^3 + j2^2 + k2^1 + l2^0)$$
$$16^2(a2^3 + b2^2 + c2^1 + d) + 16^1(e2^3 + f2^2 + g2^1 + h) + 16^0(i2^3 + j2^2 + k2^1 + l)$$
$$16^2R + 16^1S + 16^0T$$

where

$$R = a2^3 + b2^2 + c2^1 + d$$
$$S = e2^3 + f2^2 + g2^1 + h$$
$$T = i2^3 + j2^2 + k2^1 + l$$

and the values of R,S,T vary between 0 and 15. Therefore, given 1101011000110001_2, we divide the binary code into groups of fours, starting at the right:

$$(1101)(0110)(0011)(0001).$$

For each group, multiply the zeros and ones by 8, 4, 2, 1 respectively, right to left:

$$(8 + 4 + 0 + 1)(0 + 4 + 2 + 0)(0 + 0 + 2 + 1)(0 + 0 + 0 + 1)$$
$$(13)(6)(3)(1)$$
$$D631_{16}.$$

Therefore, $1101\ 0110\ 0011\ 0001_2 \equiv D631_{16}$, which is even more compact than its octal value 153061_8.

I have deliberately used whole numbers in the above examples, but they can all be extended to include a fractional part. For example, when converting a binary number such as 11.1101_2 to octal, the groups are formed about the binary point:

$$(011).(110)(100) \equiv 3.64_8.$$

Similarly, when converting a binary number such as 101010.100110_2 to hexadecimal, the groups are also formed about the binary point:

$$(0010)(1010).(1001)(1000) \equiv 2A.98_{16}.$$

Table 2.3 shows the first twenty decimal, binary, octal and hexadecimal numbers.

Table 2.3 The first twenty decimal, binary, octal, and hexadecimal numbers

decimal	binary	octal	hex	decimal	binary	octal	hex
1	1	1	1	11	1011	13	B
2	10	2	2	12	1100	14	C
3	11	3	3	13	1101	15	D
4	100	4	4	14	1110	16	E
5	101	5	5	15	1111	17	F
6	110	6	6	16	10000	20	10
7	111	7	7	17	10001	21	11
8	1000	10	8	18	10010	22	12
9	1001	11	9	19	10011	23	13
10	1010	12	A	20	10100	24	14

2.7.5 Adding Binary Numbers

When we are first taught the addition of integers containing several digits, we are advised to solve the problem digit by digit, working from right to left. For example, to add 254 to 561 we write:

$$\begin{array}{r} 561 \\ 254 \\ \hline 815 \end{array}$$

where $4+1=5$, $5+6=1$ with a $carry=1$, $2+5+carry=8$.

Table 2.4 shows all the arrangements for adding two digits with the $carry$ shown as ${}^{carry}n$. However, when adding binary numbers, the possible arrangements collapse to the four shown in Table 2.5, which greatly simplifies the process.

For example, to add 124 to 188 as two 16-bit binary integers, we write, showing the status of the $carry$ bit:

$$\begin{array}{l} 0000000011111000 \ carry \\ 0000000010111100 = 188 \\ 0000000001111100 = 124 \\ \hline 0000000100111000 = 312 \end{array}$$

Such addition is easily undertaken by digital electronic circuits, and instead of having separate circuitry for subtraction, it is possible to perform subtraction using the technique of *two's complement*.

Table 2.4 Addition of two decimal integers showing the *carry*

+	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9	$^{1}0$
2	2	3	4	5	6	7	8	9	$^{1}0$	$^{1}1$
3	3	4	5	6	7	8	9	$^{1}0$	$^{1}1$	$^{1}2$
4	4	5	6	7	8	9	$^{1}0$	$^{1}1$	$^{1}2$	$^{1}3$
5	5	6	7	8	9	$^{1}0$	$^{1}1$	$^{1}2$	$^{1}3$	$^{1}4$
6	6	7	8	9	$^{1}0$	$^{1}1$	$^{1}2$	$^{1}3$	$^{1}4$	$^{1}5$
7	7	8	9	$^{1}0$	$^{1}1$	$^{1}2$	$^{1}3$	$^{1}4$	$^{1}5$	$^{1}6$
8	8	9	$^{1}0$	$^{1}1$	$^{1}2$	$^{1}3$	$^{1}4$	$^{1}5$	$^{1}6$	$^{1}7$
9	9	$^{1}0$	$^{1}1$	$^{1}2$	$^{1}3$	$^{1}4$	$^{1}5$	$^{1}6$	$^{1}7$	$^{1}8$

Table 2.5 Addition of two binary integers showing the *carry*

+	0	1
0	0	1
1	1	$^{1}0$

2.7.6 Subtracting Binary Numbers

Two's complement is a technique for converting a binary number into a form such that when it is added to another binary number, it results in a subtraction. There are two stages to the conversion: inversion, followed by the addition of 1. For example, 24 in binary is 0000000000110000, and is inverted by switching every 1 to 0, and *vice versa*: 1111111111100111. Next , we add 1: 1111111111101000, which now represents -24. If this is added to binary 36: 0000000000100100, we have

$$\begin{array}{r} 0000000000100100 = +36 \\ \underline{1111111111101000 = -24} \\ 0000000000001100 = +12 \end{array}$$

Note that the last high-order addition creates a *carry* of 1, which is ignored. Here is another example, $100 - 30$:

	$0000000000011110 = +30$
Inversion	1111111111100001
Add 1	0000000000000001
	$1111111111100010 = -30$
Add 100	$0000000001100100 = +100$
	$0000000001000110 = +70$

2.8 Types of Numbers

As mathematics evolved, mathematicians introduced different types of numbers to help classify equations and simplify the language employed to describe their work. These are the various types and their set names.

2.8.1 Natural Numbers

The *natural numbers* $\{1,\ 2,\ 3,\ 4, \ldots\}$ are used for counting, ordering and labelling and represented by the set $\mathbb{N}$. When zero is included, $\mathbb{N}^0$ or $\mathbb{N}_0$ is used:

$$\mathbb{N}^0 = \mathbb{N}_0 = \{0, 1, 2, \ldots\}.$$

Note that negative numbers are not included. Natural numbers are used to subscript a quantity to distinguish one element from another, e.g. $x_1,\ x_2,\ x_3,\ x_4, \ldots$.

2.8.2 Integers

Integer numbers include the natural numbers, both positive and negative, and zero, and are represented by the set $\mathbb{Z}$:

$$\mathbb{Z} = \{\ldots,\ -2,\ -1,\ 0,\ 1,\ 2,\ 3, \ldots\}.$$

The reason for using $\mathbb{Z}$ is because the German for whole number is *ganzen Zahlen*. Leopold Kronecker apparently criticised Georg Cantor for his work on set theory with the jibe: "*Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk*", which translates: "*God made the integers, and all the rest is man's work*", implying that the rest are artificial. However, Cantor's work on set theory and transfinite numbers proved to be far from artificial.

2.8.3 Rational Numbers

Any number that equals the quotient of one integer divided by another non-zero integer, is a *rational number*, and represented by the set $\mathbb{Q}$. For example, 2, $\sqrt{16}$, 0.25 are rational numbers because

$$\begin{aligned} 2 &= 4/2 \\ \sqrt{16} &= 4 = 8/2 \\ 0.25 &= 1/4. \end{aligned}$$

Some rational numbers can be stored accurately inside a computer, but many others can only be stored approximately. For example, $4/3$ produces an infinite sequence of threes $1.333333\ldots$ and is truncated when stored as a binary number.

2.8.4 Irrational Numbers

An *irrational number* cannot be expressed as the quotient of two integers. Irrational numbers never terminate, nor contain repeated sequences of digits, consequently, they are always subject to a small error when stored within a computer. Examples are:

$$\begin{aligned} \sqrt{2} &= 1.41421356\ldots \\ \phi &= 1.61803398\ldots \text{(golden section)} \\ \mathrm{e} &= 2.71828182\ldots \\ \pi &= 3.14159265\ldots \end{aligned}$$

2.8.5 Real Numbers

Rational and irrational numbers comprise the set of *real numbers* $\mathbb{R}$. Examples are 1.5, 0.004, 12.999 and 23.0.

2.8.6 Algebraic and Transcendental Numbers

Polynomial equations with rational coefficients have the form:

$$f(x) = ax^n + bx^{n-1} + cx^{n-2} \cdots + C$$

such as

$$y = 3x^2 + 2x - 1$$

and their roots belong to the set of *algebraic numbers* $\mathbb{A}$. A consequence of this definition implies that all rational numbers are algebraic, since if

$$x = \frac{p}{q}$$

then

$$qx - p = 0$$

which is a polynomial. Numbers that are not roots to polynomial equations are *transcendental numbers* and include most irrational numbers, but not $\sqrt{2}$, since if

$$x = \sqrt{2}$$

then

$$x^2 - 2 = 0$$

which is a polynomial.

2.8.7 Imaginary Numbers

Imaginary numbers employ the symbol i to represent the impossible operation $\sqrt{-1}$. When combined with a real number they form a *complex number* which possesses vector-like properties. An imaginary number such as bi is defined as

$$b \in \mathbb{R}, \quad i^2 = -1.$$

Imaginary numbers obey all the axioms associated with real numbers: they can be added, subtracted, multiplied and divided. For example, given

$$x = -6i$$
$$y = 3i$$

then

$$\begin{aligned} x + y &= -6i + 3i = -3i \\ x - y &= -6i - 3i = -9i \\ xy &= (-6i)(3i) = -18i^2 = 18 \\ \frac{x}{y} &= \frac{-6i}{3i} = -2. \end{aligned}$$

2.8.8 Complex Numbers

A *complex number* has a real and imaginary part: $z = a + ib$, and represented by the set $\mathbb{C}$:

$$z = a + bi, \qquad z \in \mathbb{C}, \quad a, b \in \mathbb{R}, \quad i^2 = -1.$$

Examples are

$$\begin{aligned} z &= 1 + i \\ z &= 3 - 2i \\ z &= -23 + \sqrt{23}i. \end{aligned}$$

Complex numbers obey all the axioms associated with real numbers. For example, if we multiply $a + bi$ by $c + di$ we have

$$(a + bi)(c + di) = ac + adi + bci + bdi^2.$$

Collecting up like terms and substituting -1 for i^2 we get

$$(a + bi)(c + di) = ac + (ad + bc)i - bd$$

which simplifies to

$$(a + bi)(c + di) = ac - bd + (ad + bc)i$$

which is another complex number.

For example, given

$$\begin{aligned} x &= 2 + 3i \\ y &= 3 + 4i \end{aligned}$$

then

$$\begin{aligned} x + y &= (2 + 3i) + (3 + 4i) = 5 + 7i \\ x - y &= (2 + 3i) - (3 + 4i) = -1 - i \\ xy &= (2 + 3i)(3 + 4i) = 6 + 8i + 9i + 12i^2 = -6 + 17i. \end{aligned}$$

Something interesting happens when we multiply a complex number by its *complex conjugate*, which is the same complex number but with the sign of the imaginary part reversed:

$$(a + bi)(a - bi) = a^2 - abi + bai - b^2 i^2.$$

Collecting up like terms and simplifying we obtain

$$(a + bi)(a - bi) = a^2 + b^2$$

which is a real number, as the imaginary part has been cancelled out by the action of the complex conjugate. Given a complex number y, its complex conjugate is represented by $\bar{y}$. This permits us to divide one complex number by another as follows:

$$\begin{aligned}
x &= 2 + 3i \\
y &= 3 + 4i \\
\bar{y} &= 3 - 4i \\
\frac{x}{y} &= \frac{x}{y}\frac{\bar{y}}{\bar{y}} = \frac{(2+3i)}{(3+4i)}\frac{(3-4i)}{(3-4i)} = \frac{6 - 8i + 9i + 12}{9 + 16} = \frac{18 + i}{25} = \tfrac{18}{25} + \tfrac{1}{25}i.
\end{aligned}$$

Chapter 12 provides more information on complex numbers.

2.8.9 *Quaternions and Octonions*

In 1843, the brilliant Irish mathematician and physicist Sir William Rowan Hamilton (1805–1865) invented *quaternions*, represented by the set $\mathbb{H}$, in honour of its inventor, which were the first non-commutative algebra:

$$q = a + bi + cj + dk$$

where

$$q \in \mathbb{H}, \quad a, b, c, d \in \mathbb{R}, \quad i^2 = j^2 = k^2 = -1,$$

$$ij = k, \quad ji = -k, \quad jk = i, \quad kj = -i, \quad ki = j, \quad ik = -j, \quad ijk = -1.$$

The imaginary products are shown in Table 2.6.

Given two quaternions:

$$\begin{aligned}
x &= a + bi + cj + dk \\
y &= e + fi + gj + hk
\end{aligned}$$

their product xy equals

$$\begin{aligned}
xy = {} & (ae - bf - cg - dh) + (af + be + ch - dg)i \\
& + (ag + ce + df - bh)j + (ah + de + bg - cf)k.
\end{aligned}$$

Table 2.6 The quaternion's imaginary products

$\times$	i	j	k
i	-1	k	$-j$
j	$-k$	-1	i
k	j	$-i$	-1

The American mathematician Josiah Willard Gibbs (1839–1903), realised that a quaternion's imaginary part could be isolated and represent quantities with magnitude and direction, and 3D vectors were born:

$$\mathbf{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

Almost immediately quaternions were invented, the hunt began for the next complex algebra, which was discovered simultaneously in 1843 by a colleague of Hamilton, John Thomas Graves (1806–1870), who called them *octaves*, and by the young English mathematician Arthur Cayley (1821–1895), who called them *Cayley Numbers*:

$$z = a + bi + cj + dk + ep + fq + gr + hs$$

$$a, b, c, d, e, f, g, h \in \mathbb{R}, \quad i^2, j^2, k^2, p^2, q^2, r^2, s^2 = -1.$$

They are now called *octonions*, and are not only non-commutative, but non-associative, which means that in general, given three octonions A, B, C, then $(AB)C \neq A(BC)$. In 1898, the German mathematician Adolf Hurwitz (1859–1919), proved that there are only four algebras where it is possible to multiply and divide in the accepted sense: $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, $\mathbb{O}$. Figure 2.2 shows the sets of numbers diagrammatically.

2.8.10 Transcendental and Algebraic Numbers

Given a polynomial built from integers, for example

$$y = 3x^3 - 4x^2 + x + 23,$$

if the result is an integer, it is called an *algebraic number*, otherwise it is a *transcendental number*. Familiar examples of the latter being $\pi = 3.141\ 592\ 653\ldots$, and $\mathrm{e} = 2.718\ 281\ 828\ldots$, which can be represented as various continued fractions:

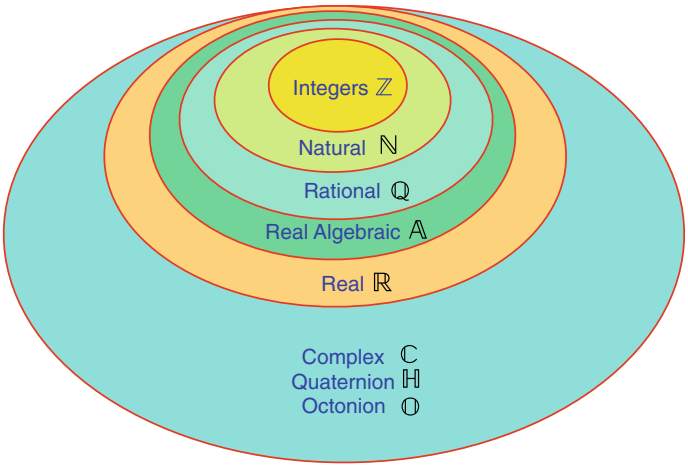

Fig. 2.2 The nested sets of numbers

$$\pi = \cfrac{4}{1+\cfrac{1^2}{2+\cfrac{3^2}{2+\cfrac{5^2}{2+\cfrac{7^2}{2+\cdots}}}}}$$

$$\mathrm{e} = 2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{4+\cdots}}}}}$$

2.9 Prime Numbers

A *prime number* is defined as a positive integer that can only be divided by 1 and itself, without leaving a remainder. The first five prime numbers are 2, 3, 5, 7, 11. We can prove that *any* positive integer must either be a prime, or the product of two or more primes, using the following reasoning.

The set of natural numbers comprises two sets: primes and non-primes. A prime, by definition, has no factors, apart from 1 and itself. A non-prime has factors and is called *composite*. However, these factors are natural numbers, which must either be

Table 2.7 The prime factors for the first 30 numbers

number	factors	number	factors	number	factors
1		11	11	21	3×7
2	2	12	2^2×3	22	2×11
3	3	13	13	23	23
4	2^2	14	2×7	24	$2^3 \times 3$
5	5	15	3×5	25	5^2
6	2×3	16	2^4	26	2×13
7	7	17	17	27	3^3
8	2^3	18	2×3^2	28	$2^2 \times 7$
9	3^2	19	19	29	29
10	2×5	20	$2^2 \times 5$	30	$2 \times 3 \times 5$

prime or non-prime. Eventually, the composite factors *must* decompose into composite primes.

For example, $72 = 8 \times 9$, but $8 = 2^3$ and $9 = 3^2$, therefore, $72 = 2^3 \times 3^2$. Even starting with $72 = 6 \times 12$, but $6 = 2 \times 3$ and $12 = 2^2 \times 3$, therefore, $72 = 2^3 \times 3^2$. Table 2.7 shows the prime factors for the first 30 numbers.

2.9.1 The Fundamental Theorem of Arithmetic

Original work by the Greek mathematician Euclid (Mid-4th to mid-3rd century BC), revealed the *Fundamental Theorem of Arithmetic* (FTAr), also called the *Unique Factorisation Theorem*, which states that every integer greater than 1, is either prime or the unique product of primes, and is expressed symbolically as follows. Let $p_1, p_2, p_3, \ldots, p_k$ be prime numbers, and $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_k$ be their associated positive integer powers: $p_1^{\alpha_1}, p_2^{\alpha_2}, p_3^{\alpha_3}, \ldots, p_k^{\alpha_k}$. We now use the product function Pi, $\prod$ to create the product: $p_1^{\alpha_1} p_2^{\alpha_2} p_3^{\alpha_3}, \ldots, p_k^{\alpha_k}$, and introduce the variable i with a range of 1 to k, which permits the FTAr to be written as

$$n = p_1^{\alpha_1} p_2^{\alpha_2} p_3^{\alpha_3}, \ldots, p_k^{\alpha_k} = \prod_{i=1}^{k} p_i^{\alpha_i}$$

where $\prod$ is shorthand for "multiply together the associated terms".

For example, 2250 equals the unique product: $2^1 3^2 5^3$, and $245 = 5^1 7^2$. To prove that these prime products are unique, let's first assume that they are not, and show that this leads to a contradiction.

Let $n > 1$ and equals the product of two prime numbers: $n = p_1 p_2$. Now let's assume that n also equals the product of two other prime numbers: $q_1 q_2$. Therefore,

$$p_1 p_2 = q_1 q_2$$

and

$$p_1 = \frac{q_1 q_2}{p_2}$$

which implies that either q_1/p_2 or q_2/p_2 factorises. However, this is impossible as q_1 and q_2 are prime, therefore, the original assumption was incorrect. The same reasoning may be generalised to any number of prime factors.

2.9.2 Is 1 a Prime?

You may notice in Table 2.7 showing factors that 1 is not a prime, which has not always been the case. The reason is due to maintaining the logical integrity of the FTAr, which emphasises the uniqueness of the product of primes. If 1 were a prime, we would have the following non-unique products

$$24 = 2^3 \times 3$$
$$24 = 1 \times 2^3 \times 3$$

and it doesn't seem satisfying to make 1 a prime, and then qualify the FTAr with the rider: "This only applies for primes greater than 1."

In 1742, the German mathematician Christian Goldbach (1690–1764) conjectured that every even integer greater than 2 could be written as the sum of two primes:

$$4 = 2 + 2$$
$$14 = 11 + 3$$
$$18 = 11 + 7, \text{ etc.}$$

No one has ever found an exception to this conjecture, and no one has ever confirmed it.

2.9.3 Prime Number Distribution

As one moves higher through the set of natural numbers, new primes are uncovered. But every prime discovered increases the possibility for more composite numbers, which overall, creates a falling distribution for primes. Table 2.7 shows that there are

10 primes in the first 30 numbers, and further analysis reveals 25 primes in the first 100 numbers, after which, they slowly decline, but never disappear.

The German mathematician Carl Gauss (1777–1855), proved, at the age of fourteen, that as $x \to \infty$, (x moves towards infinity), the function $\pi(x)$, which estimates the number of primes up to x, is given by

$$\pi(x) \sim \frac{x}{\ln x}$$

(where $\sim$ stands for "similar to".)

Testing this for $x = 100$:

$$\pi(100) \sim \frac{100}{\ln 100} \approx \frac{100}{4.60517} \approx 22.$$

which is lower than the actual value of 25. However, the French mathematician Adrien-Marie Legendre (1752–1833), conjectured the following relationship:

$$\pi(x) \sim \frac{x}{\ln x - B}$$

where $B = 1.08366$. But it appears that the best result is when $B = 1$. Testing this for $x = 100$:

$$\pi(100) \sim \frac{100}{\ln 100 - 1} \approx \frac{100}{3.605} \approx 28.$$

which is higher than the actual value of 25.

2.9.4 Infinity of Primes

Euclid also showed that there are an infinite number of primes. As we know that the number of primes is either finite or infinite, we begin by assuming that the number is finite, and proving that the assumption is contradicted by an example. We begin by declaring that there are n primes: $p_1, p_2, p_3, \ldots p_{n-1}, p_n$. Next, we form the operation $N = p_1 \times p_2 \times p_3 \cdots p_{n-1} \times p_n + 1$, which can also be written using the product operation:

$$N = \prod_{i=1}^{n} p_i + 1.$$

Now, N must be prime or have factors:
1: If N is prime, then p_n is not the largest prime, as assumed.
2: If N has factors, it must be divisible by some prime factor. But this prime factor cannot include any of the original primes as there is a remainder of 1. Therefore, p_n is not the largest prime, as assumed. Either way, the original assumption is incorrect; therefore, there must be an infinite number of primes. Table 2.8 lists some examples of these primes and prime factors.

Table 2.8 Examples of primes and prime factors

2	3	5	7	11	13	17	19	23	29	31	N
2											3
2	3										7
2	3	5									31
2	3	5	7								211
2	3	5	7	11							2,311
2	3	5	7	11	13						$30,031 = 59 \times 509$
2	3	5	7	11	13	17					$510,511 = 19 \times 97 \times 277$
2	3	5	7	11	13	17	19				$9,699,691 = 347 \times 27,953$
2	3	5	7	11	13	17	19	23			$223,092,871 = 317 \times 703,763$
2	3	5	7	11	13	17	19	23	29		$6,469,693,231 = 331 \times 571 \times 34,231$
2	3	5	7	11	13	17	19	23	29	31	200,560,490,131

See www.compoasso.free.fr for an amazing list of prime numbers and related features. Also, readers interested in learning more about prime numbers should investigate Prime Numbers (2006).

2.9.5 Perfect Numbers

A *perfect number* equals the sum of its factors. For example, the factors of 6 are 1, 2 and 3, whose sum is also 6. One would imagine that there would be a large quantity of small perfect numbers, but the first five are: 6, 28, 496, 8128 and 33,550,336, which are all even. And as the search continues to discover higher values, using computers, no odd perfect number has emerged. Euclid proved that if m is prime, and of the form $2^k - 1$, then $m(m+1)/2$ is an even perfect number. For example, 3 is prime and

$$3 = 2^2 - 1 \quad \text{and} \quad \frac{3 \times 4}{2} = 6.$$

Similarly, 7 is prime and

$$7 = 2^3 - 1 \quad \text{and} \quad \frac{7 \times 8}{2} = 28.$$

2.9.6 Mersenne Numbers

Numbers of the form $2^k - 1$ are called *Mersenne numbers*, some of which, are also prime. The French theologian and mathematician Marin Mersenne (1588–1648) became interested in them towards the end of his life, and today they are known as *Mersenne primes*.

By the end of the 16th-century, the highest Mersenne prime was 524,287 which equals $2^{19} - 1$. At the start of the 21st-century, $2^{43,112,609} - 1$ was the highest, containing approximately 13 million digits!

Apart from the fact that prime numbers are so mysterious, they are very important in public key cryptography, which is central to today's Internet security systems.

2.10 Infinity

The term *infinity* is used to describe the size of unbounded systems. For example, prime numbers are infinite, so too are the sets of other numbers. Consequently, no matter how we try, it is impossible to visualise the size of infinity. Nevertheless, this did not stop Georg Cantor from showing that one infinite set could be infinitely larger than another.

Cantor distinguished between those infinite number sets that could be "counted", and those that could not. For Cantor, counting meant the one-to-one correspondence of a natural number with the members of another infinite set. If there was a clear correspondence, without leaving any gaps, then the two sets shared a common infinite size, called its *cardinality* using the first letter of the Hebrew alphabet aleph: $\aleph$. The cardinality of the natural numbers $\mathbb{N}$ is $\aleph_0$, called aleph-zero.

Cantor discovered a way of representing the rational numbers as a grid, which is traversed diagonally, back and forth, as shown in Fig. 2.3. Some ratios appear several times, such as $\frac{2}{2}$, $\frac{3}{3}$ etc., which are not counted. Nevertheless, the one-to-one correspondence with the natural numbers means that the cardinality of rational numbers is also $\aleph_0$.

A real surprise was that there are infinitely more transcendental numbers than natural numbers. Furthermore, there are an infinite number of cardinalities rising to $\aleph_\aleph$. Cantor had been alone working in this esoteric area, and as he published his results, he shook the very foundations of mathematics, which is why he was treated so badly by his fellow mathematicians.

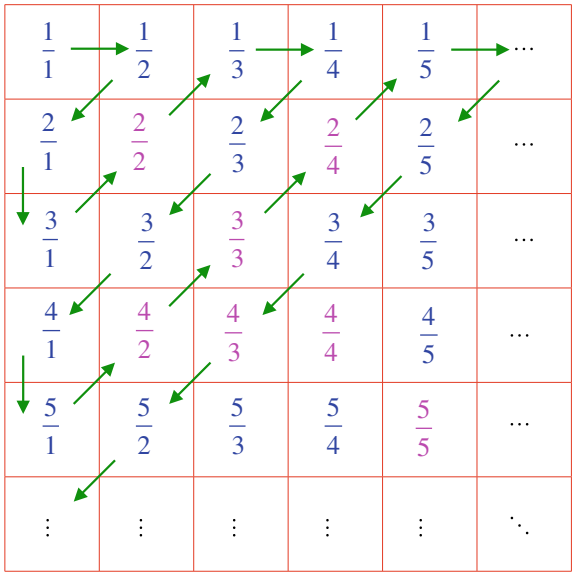

Fig. 2.3 Rational number grid

2.11 Worked Examples

2.11.1 Algebraic Expansion

Expand $(a+b)(c+d)$, $(a-b)(c+d)$, and $(a-b)(c-d)$.
Solution:

$$\begin{aligned}(a+b)(c+d) &= a(c+d)+b(c+d)\\ &= ac+ad+bc+bd.\\ (a-b)(c+d) &= a(c+d)-b(c+d)\\ &= ac+ad-bc-bd.\\ (a-b)(c-d) &= a(c-d)-b(c-d)\\ &= ac-ad-bc+bd.\end{aligned}$$

2.11.2 Binary Subtraction

Using two's complement, subtract 12 from 50.
Solution:

	0000000000001100 = +12
Inversion	1111111111110011
Add 1	0000000000000001
	1111111111110100 = −12
Add 50	0000000000110010 = +50
	0000000000100110 = +38

2.11.3 Complex Numbers

Compute $(3 + 2\mathrm{i}) + (2 + 2\mathrm{i}) + (5 - 3\mathrm{i})$ and $(3 + 2\mathrm{i})(2 + 2\mathrm{i})(5 - 3\mathrm{i})$.
Solution:

$$(3 + 2\mathrm{i}) + (2 + 2\mathrm{i}) + (5 - 3\mathrm{i}) = 10 + \mathrm{i}.$$

$$\begin{aligned}(3 + 2\mathrm{i})(2 + 2\mathrm{i})(5 - 3\mathrm{i}) &= (3 + 2\mathrm{i})(10 - 6\mathrm{i} + 10\mathrm{i} + 6)\\ &= (3 + 2\mathrm{i})(16 + 4\mathrm{i})\\ &= 48 + 12\mathrm{i} + 32\mathrm{i} - 8\\ &= 40 + 44\mathrm{i}.\end{aligned}$$

2.11.4 Complex Rotation

Rotate the complex point $3 + 2\mathrm{i}$ by $\pm 90°$ and $\pm 180°$.
Solution:
To rotate $+90°$ (anticlockwise) multiply by i.

$$\mathrm{i}(3 + 2\mathrm{i}) = 3\mathrm{i} - 2 = -2 + 3\mathrm{i}.$$

To rotate $-90°$ (clockwise) multiply by $-\mathrm{i}$.

$$-\mathrm{i}(3 + 2\mathrm{i}) = -3\mathrm{i} + 2 = 2 - 3\mathrm{i}.$$

To rotate $+180°$ (anticlockwise) multiply by -1.

$$-1(3 + 2\mathrm{i}) = -3 - 2\mathrm{i}.$$

To rotate $-180°$ (clockwise) multiply by -1.

$$-1(3 + 2\mathrm{i}) = -3 - 2\mathrm{i}.$$

2.11.5 Quaternions

Compute $(2 + 3i + 4j + k) + (6 + 2i + j + 2k)$ and $(2 + 3i + 4j + k)(6 + 2i + j + 2k)$.
Solution:

$$(2 + 3i + 4j + k) + (6 + 2i + j + 2k) = 8 + 5i + 5j + 3k.$$

$$\begin{aligned}(2 + 3i + 4j + k)(6 + 2i + j + 2k) &= 12 + 4i + 2j + 4k + 18i + 6i^2 + 3ij + 6ik \\ &\quad + 24j + 8ji + 4j^2 + 8jk + 6k + 2ki + kj + 2k^2 \\ &= 12 + 4i + 2j + 4k + 18i - 6 + 3k - 6j \\ &\quad + 24j - 8k - 4 + 8i + 6k + 2j - i - 2 \\ &= 0 + 29i + 22j + 5k.\end{aligned}$$

References

www.compoasso.free.fr
Crandall R, Pomerance C (2006) Prime numbers: a computational perspective, 2nd edn

Chapter 3
Algebra

3.1 Introduction

This chapter revises the elements of algebra such as notation, rules, indices, logarithms, explicit and implicit functions, intervals, function domains and ranges, odd and even functions and power series. The chapter concludes with some worked examples.

3.2 Background

Some people, including me, find learning a foreign language a real challenge; one of the reasons being the inconsistent rules associated with its syntax. For example, why is a table feminine in French, "la table", and a bed masculine, "le lit"? They both have four legs! The rules governing natural language are continuously being changed by each generation, whereas mathematics appears to be logical and consistent. The reason for this consistency is due to the rules associated with numbers and the way they are combined together and manipulated at an abstract level. Such rules, or *axioms*, generally make our life easy, however, as we saw with the invention of negative numbers, extra rules have to be introduced, such as "two negatives make a positive", which is easily remembered. However, as we explore mathematics, we discover all sorts of inconsistencies, such as there is no real value associated with the square-root of a negative number. It's forbidden to divide a number by zero. Zero divided by zero gives inconsistent results. Nevertheless, such conditions are easy to recognise and avoided. At least in mathematics, we don't have to worry about masculine and feminine numbers!

As a student, I discovered *Principia Mathematica*, a three-volume work written by the British philosopher, logician, mathematician and historian Bertrand Russell (1872–1970), and the British mathematician and philosopher Alfred North Whitehead (1861–1947), in which the authors attempted to deduce all of mathematics

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_3

using the axiomatic system developed by the Italian mathematician Giuseppe Peano (1858–1932). The first volume established type theory, the second was devoted to numbers, and the third to higher mathematics. The authors did intend a fourth volume on geometry, but it was too much effort to complete. It made extremely intense reading. In fact, I never managed to get pass the first page! It took the authors almost 100 pages of deep logical analysis in the second volume to prove that $1 + 1 = 2$!

Russell wrote in his *Principles of Mathematics* (1903):

> "The fact that all Mathematics is Symbolic Logic
> is one of the greatest discoveries of our age;
> and when this fact has been established,
> the remainder of the principles of mathematics
> consists in the analysis of Symbolic Logic itself."

Unfortunately, this dream cannot be realised, for in 1931, the Austrian-born, and later American logician and mathematician Kurt Gödel (1906–1978), showed that even though mathematics is based upon a formal set of axioms, there will always be statements involving natural numbers that cannot be proved or disproved. Furthermore, a consistent axiomatic system cannot demonstrate its own consistency. These theorems are known as Gödel's *incompleteness theorems*.

Even though we start off with some simple axioms, it does not mean that everything discovered in mathematics is provable, which does not mean that we cannot continue our every-day studies using algebra to solve problems. So let's examine the basic rules of algebra and prepare ourselves for the following chapters.

3.3 Notation

Modern algebraic notation has evolved over thousands of years where different civilisations developed ways of annotating mathematical and logical problems. The word "algebra" comes from the Arabic "*al-jabr w'al-muqabal*" meaning "restoration and reduction". In retrospect, it does seem strange that centuries passed before the "equals" sign (=) was invented, and concepts such as "zero" (CE 876) were introduced, especially as they now seem so important. But we are not at the end of this evolution, because new forms of annotation and manipulation will continue to emerge as new mathematical objects are invented.

One fundamental concept of algebra is the idea of giving a name to an unknown quantity. For example, m is often used to represent the slope of a 2D line, and c is the line's y-coordinate where it intersects the y-axis. René Descartes formalised the idea of using letters from the beginning of the alphabet ($a,\ b,\ c, \ldots$) to represent arbitrary quantities, and letters at the end of the alphabet ($p,\ q,\ r,\ s,\ t, \ldots,\ x,\ y,\ z$) to represent quantities such as pressure (p), time (t) and coordinates ($x,\ y,\ z$).

With the aid of the basic arithmetic operators: $+, -, \times, /$ we can develop expressions that describe the behaviour of a physical process or a logical computation. For

example, the expression $ax + by - d$ equals zero for a straight line. The variables x and y are the coordinates of any point on the line and the values of a, b and d determine the position and orientation of the line. The $=$ sign permits the line equation to be expressed as a self-evident statement:

$$0 = ax + by - d.$$

Such a statement implies that the expressions on the left- and right-hand sides of the $=$ sign are "equal" or "balanced", and in order to maintain equality or balance, whatever is done to one side, must also be done to the other. For example, adding d to both sides, the straight-line equation becomes

$$d = ax + by.$$

Similarly, we could double or treble both expressions, divide them by 4, or add 6, without disturbing the underlying relationship. When we are first taught algebra, we are often given the task of rearranging a statement to make different variables the subject. For example, (3.1) can be rearranged such that x is the subject:

$$\begin{aligned} y &= \frac{x+4}{2-\frac{1}{z}} \\ y\left(2-\frac{1}{z}\right) &= x+4 \\ x &= y\left(2-\frac{1}{z}\right)-4. \end{aligned} \tag{3.1}$$

Making z the subject requires more effort:

$$\begin{aligned} y &= \frac{x+4}{2-\frac{1}{z}} \\ y\left(2-\frac{1}{z}\right) &= x+4 \\ 2y-\frac{y}{z} &= x+4 \\ 2y-x-4 &= \frac{y}{z} \\ z &= \frac{y}{2y-x-4}. \end{aligned}$$

Parentheses are used to isolate part of an expression in order to select a subexpression that is manipulated in a particular way. For example, the parentheses in $c(a+b)+d$ ensure that the variables a and b are added together before being multiplied by c, and finally added to d.

3.3.1 Solving the Roots of a Quadratic Equation

Problem solving is greatly simplified if one has solved it before, and having a good memory is always an advantage. In mathematics, we keep coming across problems that have been encountered before, apart from different numbers. For example, $(a+b)(a-b)$ always equals a^2-b^2, therefore factorising the following is a trivial exercise:

$$\begin{aligned} a^2-16 &= (a+4)(a-4) \\ x^2-49 &= (x+7)(x-7) \\ x^2-2 &= (x+\sqrt{2})(x-\sqrt{2}). \end{aligned}$$

A perfect square has the form:

$$a^2+2ab+b^2=(a+b)^2.$$

Consequently, factorising the following is also a trivial exercise:

$$\begin{aligned} a^2+4ab+4b^2 &= (a+2b)^2 \\ x^2+14x+49 &= (x+7)^2 \\ x^2-20x+100 &= (x-10)^2. \end{aligned}$$

Now let's solve the roots of the quadratic equation $ax^2+bx+c=0$, i.e. those values of x that make the equation equal zero. As the equation involves an x^2 term, we will exploit any opportunity to factorise it. We begin with the quadratic where $a \neq 0$:

$$ax^2+bx+c=0.$$

Step 1: Subtract c from both sides to begin the process of creating a perfect square:

$$ax^2+bx=-c.$$

Step 2: Divide both sides by a to create an x^2 term:

$$x^2+\frac{b}{a}x=-\frac{c}{a}.$$

Step 3: Add $b^2/4a^2$ to both sides to create a perfect square on the left side:

$$x^2+\frac{b}{a}x+\frac{b^2}{4a^2}=\frac{b^2}{4a^2}-\frac{c}{a}.$$

Step 4: Factorise the left side:

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2}{4a^2} - \frac{c}{a}.$$

Step 5: Make $4a^2$ the common denominator for the right side:

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}.$$

Step 6: Take the square root of both sides:

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

Step 7: Subtract $b/2a$ from both sides:

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

Step 8: Rearrange the right side:

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

which provides the roots for any quadratic equation.

The discriminant $\sqrt{b^2 - 4ac}$ may be positive, negative or zero. A positive value reveals two real roots:

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \tag{3.2}$$

A negative value reveals two complex roots:

$$x_1 = \frac{-b + i\sqrt{|b^2 - 4ac|}}{2a}, \quad x_2 = \frac{-b - i\sqrt{|b^2 - 4ac|}}{2a}.$$

And a zero value reveals a single root:

$$x = \frac{-b}{2a}.$$

For example, Fig. 3.1 shows the graph of $y = x^2 + x - 2$, where we can see that $y = 0$ at two points: $x = -2$ and $x = 1$. In this equation

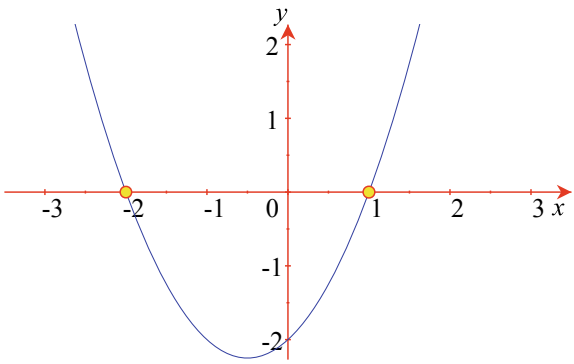

Fig. 3.1 Graph of $y = x^2 + x - 2$

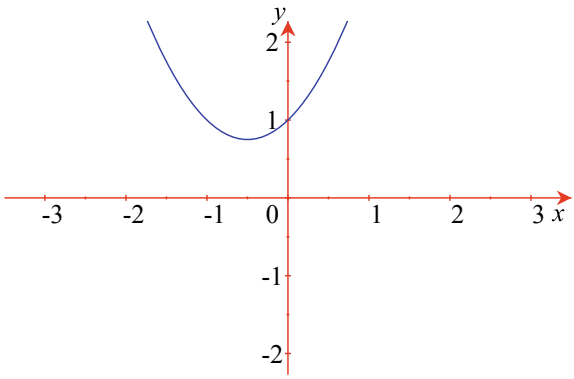

Fig. 3.2 Graph of $y = x^2 + x + 1$

$$
\begin{aligned}
a &= 1 \\
b &= 1 \\
c &= -2
\end{aligned}
$$

which when plugged into (3.2) confirms the graph:

$$
\begin{aligned}
x_1 &= \frac{-1 + \sqrt{1+8}}{2} = 1 \\
x_2 &= \frac{-1 - \sqrt{1+8}}{2} = -2.
\end{aligned}
$$

Figure 3.2 shows the graph of $y = x^2 + x + 1$, where at no point does $y = 0$. In this equation

$$
\begin{aligned}
a &= 1 \\
b &= 1 \\
c &= 1
\end{aligned}
$$

which when plugged into (3.2) confirms the graph by giving complex roots:

$$x_1 = \frac{-1+\sqrt{1-4}}{2} = -\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2}$$
$$x_2 = \frac{-1-\sqrt{1-4}}{2} = -\tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}.$$

Let's show that x_1 satisfies the original equation:

$$\begin{aligned} y &= x_1^2 + x_1 + 1 \\ &= \left(-\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2}\right)^2 - \tfrac{1}{2} + i\tfrac{\sqrt{3}}{2} + 1 \\ &= \tfrac{1}{4} - i\tfrac{\sqrt{3}}{2} - \tfrac{3}{4} - \tfrac{1}{2} + i\tfrac{\sqrt{3}}{2} + 1 \\ &= 0. \end{aligned}$$

x_2 also satisfies the same equation.

Algebraic expressions also contain a wide variety of functions, such as

$$\begin{aligned} \sqrt{x} &= \text{square root of } x \\ \sqrt[n]{x} &= n\text{-th root of } x \\ x^n &= x \text{ to the power } n \\ n! &= \text{factorial } n \\ \sin x &= \text{sine of } x \\ \cos x &= \text{cosine of } x \\ \tan x &= \text{tangent of } x \\ \log x &= \text{logarithm of } x \\ \ln x &= \text{natural logarithm of } x. \end{aligned}$$

Trigonometric functions are factorised as follows:

$$\begin{aligned} \sin^2 x - \cos^2 x &= (\sin x + \cos x)(\sin x - \cos x) \\ \sin^2 x - \tan^2 x &= (\sin x + \tan x)(\sin x - \tan x) \\ \sin^2 x + 4\sin x \cos x + 4\cos^2 x &= (\sin x + 2\cos x)^2 \\ \sin^2 x - 6\sin x \cos x + 9\cos^2 x &= (\sin x - 3\cos x)^2. \end{aligned}$$

3.4 Indices

Indices are used to imply repeated multiplication and create a variety of situations where laws are required to explain how the result is to be computed.

3.4.1 Laws of Indices

The laws of indices are expressed as follows:

$$a^m \times a^n = a^{m+n}$$
$$\frac{a^m}{a^n} = a^{m-n}$$
$$(a^m)^n = a^{mn}$$

and are verified using some simple examples:

$$2^3 \times 2^2 = 2^5 = 32$$
$$\frac{2^4}{2^2} = 2^2 = 4$$
$$(2^2)^3 = 2^6 = 64.$$

From the above laws, it is evident that

$$a^0 = 1$$
$$a^{-p} = \frac{1}{a^p}$$
$$a^{\frac{1}{q}} = \sqrt[q]{a}$$
$$a^{\frac{p}{q}} = \sqrt[q]{a^p}.$$

3.5 Logarithms

Two people are associated with the invention of logarithms: the Scottish theologian and mathematician John Napier (1550–1617) and the Swiss clockmaker and mathematician Joost Bürgi (1552–1632). Both men were frustrated by the time they spent multiplying numbers together, and both realised that multiplication could be replaced by addition using logarithms. Logarithms exploit the addition and subtraction of indices shown above, and are always associated with a base. For example, if $a^x = n$, then $\log_a n = x$, where a is the base. Where no base is indicated, it is assumed to be 10. Two examples bring the idea to life:

$$10^2 = 100 \quad \text{then} \quad \log 100 = 2$$
$$10^3 = 1000 \quad \text{then} \quad \log 1000 = 3$$

which is interpreted as "10 has to be raised to the power (index) 2 to equal 100." The log operation finds the power of the base for a given number. Thus a multiplication

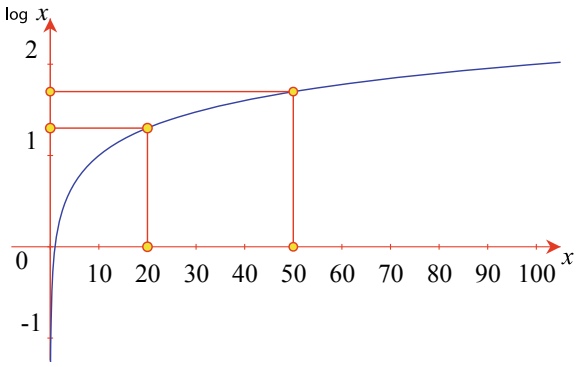

Fig. 3.3 Graph of $\log x$

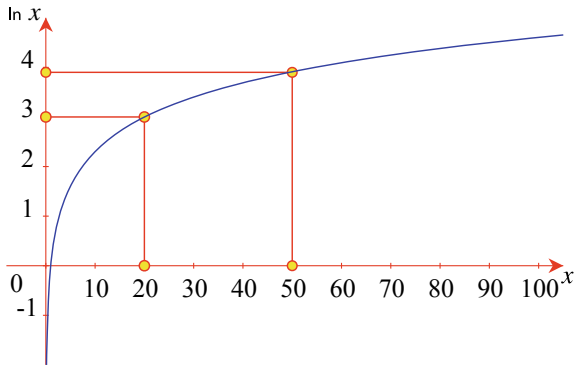

Fig. 3.4 Graph of $\ln x$

is translated into an addition using logs. Figure 3.3 shows the graph of $\log x$, up to $x = 100$, where we see that $\log 20 \approx 1.3$ and $\log 50 \approx 1.7$. Therefore, given suitable software, logarithm tables, or a calculator with a log function, we can compute the product 20×50 as follows:

$$20 \times 50 = \log 20 + \log 50 \approx 1.3 + 1.7 = 3$$
$$10^3 = 1000.$$

In general, the two bases used in calculators and software are 10 and $\mathrm{e} = 2.718\,281\,846\ldots$. To distinguish one type of logarithm from the other, a logarithm to the base 10 is written as log, and a natural logarithm to the base e is written ln.

Figure 3.4 shows the graph of $\ln x$, up to $x = 100$, where we see that $\ln 20 \approx 3$ and $\ln 50 \approx 3.9$. Therefore, given suitable software, a set of natural logarithm tables or a calculator with a ln function, we can compute the product 20×50 as follows:

$$20 \times 50 = \ln 20 + \ln 50 \approx 3 + 3.9 = 6.9$$
$$\mathrm{e}^{6.9} \approx 1000.$$

From the above notation, it is evident that

$$\log(ab) = \log a + \log b$$
$$\log\left(\frac{a}{b}\right) = \log a - \log b$$
$$\log(a^n) = n \log a.$$

3.6 Further Notation

All sorts of symbols are used to stand in for natural language expressions; here are some examples:

$<$ less than
$>$ greater than
$\leq$ less than or equal to
$\geq$ greater than or equal to
$\approx$ approximately equal to
$\equiv$ equivalent to
$\neq$ not equal to
$|x|$ absolute value of x.

For example, $0 \leq t \leq 1$ is interpreted as: t is greater than or equal to 0, and is less than or equal to 1. Basically, this means t varies between 0 and 1.

3.7 Functions

The theory of *functions* is a large subject, and at this point in the book, I will only touch upon some introductory ideas that will help you understand the following chapters.

The German mathematician Gottfried von Leibniz (1646–1716) is credited with an early definition of a function, based upon the slope of a graph. However, it was the Swiss mathematician Leonhard Euler (1707–1783) who provided a definition along the lines: "A function is a variable quantity, whose value depends upon one or more independent variables." Other mathematicians have introduced more rigorous definitions, which are examined later on in the chapter on calculus.

3.7.1 Explicit and Implicit Equations

The equation

$$y = 3x^2 + 2x + 4$$

associates the value of y with different values of x. The directness of the equation: "$y =$", is why it is called an *explicit equation*, and their explicit nature is extremely useful. However, simply by rearranging the terms, creates an *implicit equation*:

$$4 = y - 3x^2 - 2x$$

which implies that certain values of x and y combine to produce the result 4. Another implicit form is

$$0 = y - 3x^2 - 2x - 4$$

which means the same thing, but expresses the relationship in a slightly different way.

An implicit equation can be turned into an explicit equation using algebra. For example, the implicit equation

$$4x + 2y = 12$$

has the explicit form:

$$y = 6 - 2x$$

where it is clear what y equals.

3.7.2 Function Notation

The explicit equation

$$y = 3x^2 + 2x + 4$$

tells us that the value of y depends on the value of x, and not the other way around. For example, when $x = 1$, $y = 9$; and when $x = 2$, $y = 20$. As y depends upon the value of x, it is called the *dependent variable*; and as x is independent of y, it is called the *independent variable*.

We can also say that y is a function of x, which can be written as

$$y = f(x)$$

where the letter "f" is the name of the function, and the independent variable is enclosed in brackets. We could have also written $y = g(x)$, $y = h(x)$, etc.

Eventually, we have to identify the nature of the function, which in this case is

$$f(x) = 3x^2 + 2x + 4.$$

Nothing prevents us from writing

$$y = f(x) = 3x^2 + 2x + 4$$

which means: y equals the value of the function $f(x)$, which is determined by the independent variable x using the expression $3x^2 + 2x + 4$.

An equation may involve more than one independent variable, such as the volume of a cylinder:

$$V = \pi r^2 h$$

where r is the radius, and h, the height, and is written:

$$V(r, h) = \pi r^2 h.$$

3.7.3 Intervals

An *interval* is a continuous range of numerical values associated with a variable, which can include or exclude the upper and lower values. For example, a variable such as x is often subject to inequalities like $x \geq a$ and $x \leq b$, which can also be written as

$$a \leq x \leq b$$

and implies that x is located in the *closed interval* $[a, b]$, where the square brackets indicate that the interval includes a and b. For example,

$$1 \leq x \leq 10$$

means that x is located in the closed interval [1, 10], which includes 1 and 10.

When the boundaries of the interval are not included, then we would state $x > a$ and $x < b$, which is written

$$a < x < b$$

and means that x is located in the *open interval* $]a, b[$, where the reverse square brackets indicate that the interval excludes a and b. For example,

$$1 < x < 10$$

means that x is located in the open interval]1, 10[, which excludes 1 and 10.

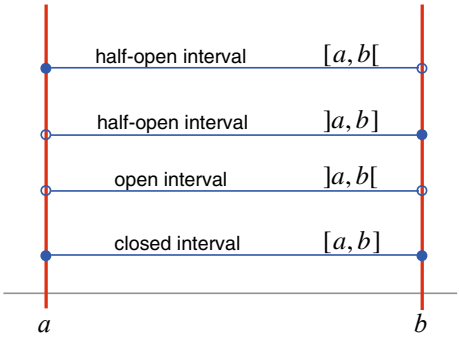

Fig. 3.5 Closed, open and half-open intervals. The filled circles indicate that a or b are included in the interval

Closed and open intervals may be combined as follows. If $x \geq a$ and $x < b$ then

$$a \leq x < b$$

and means that x is located in the *half-open interval* $[a, b[$. For example,

$$1 \leq x < 10$$

means that x is located in the half-open interval $[1, 10[$, which includes 1, but not 10.

Similarly, if

$$1 < x \leq b$$

means that x is located in the half-open interval $]1, 10]$, which includes 10, but not 1.

An alternative notation employs parentheses instead of reversed brackets:

$$]a, b[= (a, b)$$
$$[a, b[= [a, b)$$
$$]a, b] = (a, b].$$

Figure 3.5 shows open, closed and half-open intervals diagrammatically.

3.7.4 Function Domains and Ranges

The following descriptions of domains and ranges only apply to functions with one independent variable: $f(x)$.

Returning to the above function:

$$y = f(x) = 3x^2 + 2x + 4$$

the independent variable x, can take on any value from $-\infty$ to $+\infty$, which is called the *domain* of the function. In this case, the domain of $f(x)$ is the set of real numbers $\mathbb{R}$. The notation used for intervals, is also used for domains, which in this case is

$$]-\infty, +\infty[$$

and is open, as there are no precise values for $-\infty$ and $+\infty$.

As the independent variable takes on different values from its domain, so the dependent variable, y or $f(x)$, takes on different values from its *range*. Therefore, the range of $y = f(x) = 3x^2 + 2x + 4$ is also the set of real numbers $\mathbb{R}$.

The domain of $\log x$ is

$$]0, +\infty[$$

which is open, because $x \neq 0$. Whereas, the range of $\log x$ is

$$]-\infty, +\infty[.$$

The domain of $\sqrt{x}$ is

$$[0, +\infty[$$

which is half-open, because $\sqrt{0} = 0$, and $+\infty$ has no precise value. Similarly, the range of $\sqrt{x}$ is

$$[0, +\infty[.$$

Sometimes, a function is sensitive to one specific number. For example, in the function

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

when $x = 1$, there is a divide by zero, which is meaningless. Consequently, the domain of $f(x)$ is the set of real numbers $\mathbb{R}$, apart from 1.

3.7.5 Odd and Even Functions

An *odd function* satisfies the condition:

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

where x is located in a valid domain. Consequently, the graph of an odd function is symmetrical relative to the x-axis, relative to the origin. For example, $\sin\alpha$ is odd because

$$\sin(-\alpha) = -\sin\alpha$$

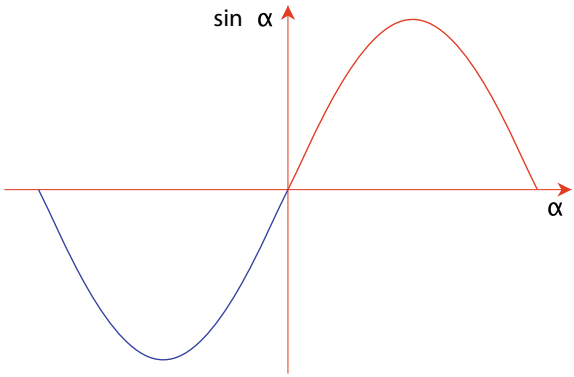

Fig. 3.6 The sine function is an odd function

as illustrated in Fig. 3.6. Other odd functions include:

$$f(x) = ax$$
$$f(x) = ax^3.$$

An *even function* satisfies the condition:

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

where x is located in a valid domain. Consequently, the graph of an even function is symmetrical relative to the $f(x)$ axis. For example, $\cos\alpha$ is even because

$$\cos(-\alpha) = \cos\alpha$$

as illustrated in Fig. 3.7. Other even functions include:

$$f(x) = ax^2$$
$$f(x) = ax^4.$$

3.7.6 Power Functions

Functions of the form $f(x) = x^n$ are called *power functions of degree n* and are either odd or even. If n is an odd natural number, then the power function is odd, else if n is an even natural number, then the power function is even.

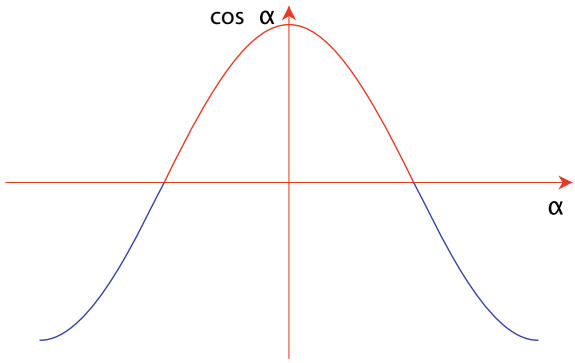

Fig. 3.7 The cosine function is an even function

3.8 Worked Examples

3.8.1 Algebraic Manipulation

Rearrange the following equations to make y the subject.

$$7 = \frac{x+4}{3-y}, \quad 23 = \frac{x+68}{3+\frac{1}{e^y}}, \quad 23 = \frac{x+68}{3-\sin y}.$$

Solution:

$$\begin{aligned} 7 &= \frac{x+4}{3-y} \\ 3-y &= \frac{x+4}{7} \\ y &= 3 - \frac{x+4}{7} = \frac{17-x}{7}. \end{aligned}$$

Solution:

$$\begin{aligned} 23 &= \frac{x+68}{3+\frac{1}{e^y}} \\ 3+\frac{1}{e^y} &= \frac{x+68}{23} \\ \frac{1}{e^y} &= \frac{x+68}{23} - 3 \\ &= \frac{x-1}{23} \end{aligned}$$

$$\begin{aligned} \mathrm{e}^y &= \frac{23}{x-1} \\ y &= \ln\left(\frac{23}{x-1}\right). \end{aligned}$$

Solution:

$$\begin{aligned} 23 &= \frac{x+68}{3-\sin y} \\ 3-\sin y &= \frac{x+68}{23} \\ \sin y &= 3-\frac{x+68}{23} \\ &= \frac{1-x}{23} \\ y &= \arcsin\left(\frac{1-x}{23}\right). \end{aligned}$$

3.8.2 Solving a Quadratic Equation

Solve the following quadratic equations, and test the answers.

$$0 = x^2+4x+1, \quad 0 = 2x^2+4x+2, \quad 0 = 2x^2+4x+4.$$

Solution: $0 = x^2+4x+1$

$$\begin{aligned} x &= \frac{-b \pm \sqrt{b^2-4ac}}{2a} \\ &= \frac{-4 \pm \sqrt{16-4}}{2} \\ &= \frac{-4 \pm \sqrt{12}}{2} \\ &= -2 \pm \sqrt{3}. \end{aligned}$$

Test with $x = -2+\sqrt{3}$.

$$\begin{aligned} x^2+4x+1 &= (-2+\sqrt{3})^2 + 4(-2+\sqrt{3}) + 1 \\ &= 4-4\sqrt{3}+3-8+4\sqrt{3}+1 \\ &= 0. \end{aligned}$$

Test with $x = -2 - \sqrt{3}$.

$$\begin{aligned} x^2 + 4x + 1 &= (-2 - \sqrt{3})^2 + 4(-2 - \sqrt{3}) + 1 \\ &= 4 + 4\sqrt{3} + 3 - 8 - 4\sqrt{3} + 1 \\ &= 0. \end{aligned}$$

Solution: $0 = 2x^2 + 4x + 2$

$$\begin{aligned} x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\ &= \frac{-4 \pm \sqrt{16 - 16}}{4} \\ &= \frac{-4}{4} \\ &= -1. \end{aligned}$$

Test with $x = -1$.

$$\begin{aligned} 2x^2 + 4x + 2 &= 2 - 4 + 2 \\ &= 0. \end{aligned}$$

Solution: $0 = 2x^2 + 4x + 4$

$$\begin{aligned} x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\ &= \frac{-4 \pm \sqrt{16 - 32}}{4} \\ &= \frac{-4 \pm \sqrt{-16}}{4} \\ &= -1 \pm \sqrt{-1} \\ &= -1 \pm i. \end{aligned}$$

Test with $x = -1 + i$.

$$\begin{aligned} 2x^2 + 4x + 4 &= 2(-1 + i)^2 + 4(-1 + i) + 4 \\ &= 2(1 - 2i - 1) - 4 + 4i + 4 \\ &= -4i + 4i \\ &= 0. \end{aligned}$$

Test with $x = -1 - i$.

$$\begin{aligned} 2x^2 + 4x + 4 &= 2(-1 - i)^2 + 4(-1 - i) + 4 \\ &= 2(1 + 2i - 1) - 4 - 4i + 4 \\ &= 4i - 4i \\ &= 0. \end{aligned}$$

3.8.3 Factorising

Factorise the following equations:

$$\begin{aligned} &4\sin^2 x - 4\cos^2 x \\ &9\sin^2 x + 6\sin x \cdot \cos x + \cos^2 x \\ &25\sin^2 x + 10\sin x \cdot \cos x + \cos^2 x. \end{aligned}$$

Solution:

$$\begin{aligned} 4\sin^2 x - 4\cos^2 x &= (2\sin x + 2\cos x)(2\sin x - 2\cos x) \\ 9\sin^2 x + 6\sin x \cdot \cos x + \cos^2 x &= (3\sin x + \cos x)^2 \\ 25\sin^2 x + 10\sin x \cdot \cos x + \cos^2 x &= (5\sin x + \cos x)^2. \end{aligned}$$

Chapter 4
Logic

4.1 Introduction

This chapter divides into three sections: truth tables, logical premises and set theory. In truth tables the reader is introduced to a tabular form of expressing logical outcomes. The section on logical premises introduces the idea of logical reasoning using constructs such as material equivalence, implication and reductio ad absurdum. The third section is an introduction to set theory including set building, empty sets, subsets, supersets and universal sets, etc. The chapter concludes with some worked examples.

4.2 Background

The English mathematician George Boole (1815–1864) is regarded as the "father" of symbolic logic, which is why it bears his name. He did not associate logic with mathematics, but wanted to devise a logical framework for expressing and analysing logical statements. A logical statement contains one or more premises (or propositions), that form the basis of an argument. However, not all premises are true, and starting from an incorrect premise is not a good strategy for winning an argument, therefore one must anticipate the existence of valid and invalid premises. Complex arguments often combine individual premises using the logical connectives negation (NOT), conjunction (AND), inclusive disjunction (OR) and the exclusive disjunction (XOR). Today, logic is considered to play a central role in mathematics, and Russell believed that mathematics could be derived entirely from logic.

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_4

4.3 Truth Tables

In algebra, variables can assume an infinite range of numerical values, and equations often require careful analysis to discover their roots. However, in logic, variables possess two states: true (T) or false (F), or the binary states: 1 and 0, respectively. Consequently, when two or more logical variables are combined, the number of possible combinations is finite, and often restricted to a dozen or so. These combinations are easily summarised in the form of a *truth table*, which automatically reveals the logical "roots" of the statement being investigated. For example, a single variable **p** can only assume two possible logic states T or F, or using binary 1 or 0, as shown in Table 4.1. With two variables **p** and **q**, the number of combinations increases to $4 = 2^2$, as shown in Table 4.2. With three variables **p**, **q** and **r**, the number of combinations increases to $8 = 2^3$, as shown in Table 4.3.

4.3.1 Logical Connectives

Over the past century, logicians have developed a framework for describing and analysing logical statements using propositions and logical connectives. Unfortunately, the logical connectives developed by logicians are not found on a computer keyboard, and the computer science community has replaced them by other characters. Table 4.4 shows the correspondence between the logical connectives used by logicians and computer scientists. It also shows the priority of each connective to control the sequence of evaluation.

Table 4.1 Truth table for one logic variable

p
T
F

Table 4.2 Truth table for two logic variables

p	q
T	T
T	F
F	T
F	F

Table 4.3 Truth table for three logic variables

p	q	r
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

Table 4.4 Correspondence of logical connectives

Operation	Meaning	Logic	Computer Science	Priority
negation	NOT	$\neg$	! or ~	1
conjunction	AND	$\wedge$	&	2
inclusive disjunction	OR	$\vee$	—	3
exclusive disjunction	XOR	$\oplus$	X	3
implication	IMPLIES	$\Rightarrow$	$\rightarrow$	4
equivalence	EQUIVALENCE	$\Leftrightarrow$	$\leftrightarrow$	4

4.4 Logical Premises

4.4.1 Material Equivalence

The premises "Descartes is human" and "Descartes is mortal" are both true, but they are not equivalent. For example, you don't have to be human in order to be mortal, but you *do* have to be mortal in order to be human. Therefore, being mortal is a *necessary* condition for being human, but is not *sufficient*, because Descartes could be the pet name for a donkey, which although mortal, is not human.

Now consider the premises "Descartes is human" and "Descartes is a man", where once again they are true, but not equivalent. For example, you don't have to be a man to be human, but you *do* have to be human to be a man. Therefore, being a man is a *sufficient* condition for being human, but is not *necessary*, because Descartes could refer to Madame Descartes, who is not a man, but still human.

Specifically, if **p** and **q** are premises such that **p** implies **q**, then **p** is a *sufficient* condition for **q**, and **q** is a *necessary* condition for **p**. If **p** is both a necessary and sufficient condition for **q** to be true, then **p** and **q** are logically equivalent. Such a

Table 4.5 Equivalence: $\mathbf{p} \leftrightarrow \mathbf{q}$

p	q	$\mathbf{p} \leftrightarrow \mathbf{q}$
T	T	T
T	F	F
F	T	F
F	F	T

condition is usually expressed using "*if and only if*". For example, "Descartes is married *if and only if* he has a wife". This is also expressed as "Descartes is married *iff* he has a wife". Therefore **p** and **q** are equivalent if **p** is true and **q** is true, and **p** is false and **q** is false. This material equivalence is written $\mathbf{p} \leftrightarrow \mathbf{q}$ as shown in Table 4.5.

4.4.2 Implication

One observation in logical reasoning is that a false premise can lead to a false or true conclusion. Furthermore, it is false that a true premise leads to a false conclusion. However, a true premise must always lead to a true conclusion. Such a relationship is called *implication*. For example, if $n \in \mathbb{N}$ then $n^2 \in \mathbb{N}$, and if $n \notin \mathbb{N}$ then $n^2 \notin \mathbb{N}$. However, starting with a false value of n, such as $n = 2 + 2i$ then it is false that $n^2 \in \mathbb{N}$. But starting with $n = -i$, which is also false, leads to $n^2 = 1$ which satisfies $n^2 \in \mathbb{N}$. Therefore, given two premises **p** and **q**, where **p** implies **q**, this is written $\mathbf{p} \rightarrow \mathbf{q}$, as shown in Table 4.6.

Let's examine the action of the other connectives using truth tables. In some cases, different combinations of T and F produce a true or false result, whilst others always produce a true result. This latter condition is called a *tautology*.

Table 4.6 Implication $\mathbf{p} \rightarrow \mathbf{q}$

p	q	$\mathbf{p} \rightarrow \mathbf{q}$
T	T	T
T	F	F
F	T	T
F	F	T

4.4.3 Negation

Negation is the act of reversing a logical state. For example, if **p** is false, then $\neg$**p** is true, and vice versa, as shown in Table 4.7.

4.4.4 Conjunction

Conjunction is the linking together of two or more premises using the $\wedge$ connective. Table 4.8 shows the action of conjunction with two premises **p** and **q**, where **p** $\wedge$ **q** is true only when both **p** *and* **q** are true, otherwise it is false. Table 4.9 shows the result with three premises.

4.4.5 Inclusive Disjunction

Table 4.10 shows the action of *inclusive disjunction*, where **p** $\vee$ **q** is true when either **p** *or* **q**, *or* both are true, otherwise it is false.

4.4.6 Exclusive Disjunction

Table 4.11 shows the action of *exclusive disjunction*, where **p** $\oplus$ **q** is true when either **p** *or* **q** is true, but not both, otherwise it is false.

Table 4.7 Negation: $\neg$**p**

p	$\neg$p
T	F
F	T

Table 4.8 Conjunction: **p** $\wedge$ **q**

p	q	p $\wedge$ q
T	T	T
T	F	F
F	T	F
F	F	F

Table 4.9 Conjunction: $\mathbf{p} \wedge \mathbf{q} \wedge \mathbf{r}$

p	q	r	$\mathbf{p} \wedge \mathbf{q} \wedge \mathbf{r}$
T	T	T	T
T	T	F	F
T	F	T	F
T	F	F	F
F	T	T	F
F	T	F	F
F	F	T	F
F	F	F	F

Table 4.10 Inclusive disjunction: $\mathbf{p} \vee \mathbf{q}$

p	q	$\mathbf{p} \vee \mathbf{q}$
T	T	T
T	F	T
F	T	T
F	F	F

Table 4.11 Exclusive disjunction: $\mathbf{p} \oplus \mathbf{q}$

p	q	$\mathbf{p} \oplus \mathbf{q}$
T	T	F
T	F	T
F	T	T
F	F	F

4.4.7 Idempotence

Tables 4.12 and 4.13 contain the unusual term: *idempotence*, which was introduced by the American mathematician Benjamin Peirce to clarify certain mathematical or logical operations, and literally means "having the same power" or is "not affected by". For example, in algebra, $1 \times 1 = 1$, where the number 1 is idempotent (not affected by) multiplication by itself. In logic, $\mathbf{p} \vee \mathbf{p} = \mathbf{p}$, where $\vee$ is idempotent when associated with two equal premises, and gives the same result $\mathbf{p}$. Table 4.12 shows the idempotence of $\vee$, and Table 4.13 shows the idempotence of $\wedge$.

Table 4.12 Idempotence of $\vee$: $\mathbf{p} \leftrightarrow \mathbf{p} \vee \mathbf{p}$

p	p ∨ p	p ↔ p ∨ p
T	T	T
F	F	T

Table 4.13 Idempotence of $\wedge$: $\mathbf{p} \leftrightarrow \mathbf{p} \wedge \mathbf{p}$

p	p ∧ p	p ↔ p ∧ p
T	T	T
F	F	T

4.4.8 Commutativity

In simple arithmetic statements, the order of elements does not affect the numerical result. For example, $4 + 6 = 6 + 4$ and $xy = yx$. This is called commutativity. In logic, something similar exists when two premises are linked together with $\vee$ or $\wedge$. For example, $\mathbf{p} \vee \mathbf{q}$ is identical to $\mathbf{q} \vee \mathbf{p}$, and $\mathbf{p} \wedge \mathbf{q}$ is identical to $\mathbf{q} \wedge \mathbf{p}$. Table 4.14 shows the commutativity of $\vee$, and Table 4.15 shows the commutativity of $\wedge$.

Table 4.14 Commutativity of $\vee$: $\mathbf{p} \vee \mathbf{q} \leftrightarrow \mathbf{q} \vee \mathbf{p}$

p	q	p ∨ q	q ∨ p	p ∨ q ↔ q ∨ p
T	T	T	T	T
T	F	T	T	T
F	T	T	T	T
F	F	F	F	T

Table 4.15 Commutativity of $\wedge$: $\mathbf{p} \wedge \mathbf{q} \leftrightarrow \mathbf{q} \wedge \mathbf{p}$

p	q	p ∧ q	q ∧ p	p ∧ q ↔ q ∧ p
T	T	T	T	T
T	F	F	F	T
T	T	F	F	T
F	F	F	F	T

4.4.9 *Associativity*

In simple arithmetic statements, the grouping of elements does not affect the numerical result. For example, $2 + (3 + 4) = (2 + 3) + 4$ and $x(yz) = (xy)z$. This is called associativity. In logic, something similar exists when two or more premises are linked together with $\vee$ or $\wedge$. For example, $(\mathbf{p} \vee \mathbf{q}) \vee \mathbf{r}$ is identical to $\mathbf{p} \vee (\mathbf{q} \vee \mathbf{r})$, and $(\mathbf{p} \wedge \mathbf{q}) \wedge \mathbf{r}$ is identical to $\mathbf{p} \wedge (\mathbf{q} \wedge \mathbf{r})$. Table 4.16 shows the associativity of $\vee$, and Table 4.17 shows the associativity of $\wedge$.

Table 4.16 Associativity of $\vee$: $(\mathbf{p} \vee \mathbf{q}) \vee \mathbf{r} \leftrightarrow \mathbf{p} \vee (\mathbf{q} \vee \mathbf{r})$

p	q	r	$(\mathbf{p} \vee \mathbf{q}) \vee \mathbf{r}$	$\mathbf{p} \vee (\mathbf{q} \vee \mathbf{r})$	$(\mathbf{p} \vee \mathbf{q}) \vee \mathbf{r} \leftrightarrow \mathbf{p} \vee (\mathbf{q} \vee \mathbf{r})$
T	T	T	T	T	T
T	T	F	T	T	T
T	F	T	T	T	T
T	F	F	T	T	T
F	T	T	T	T	T
F	T	F	T	T	T
F	F	T	T	T	T
F	F	F	F	F	T

Table 4.17 Associativity of $\wedge$: $(\mathbf{p} \wedge \mathbf{q}) \wedge \mathbf{r} \leftrightarrow \mathbf{p} \wedge (\mathbf{q} \wedge \mathbf{r})$

p	q	r	$(\mathbf{p} \wedge \mathbf{q}) \wedge \mathbf{r}$	$\mathbf{p} \wedge (\mathbf{q} \wedge \mathbf{r})$	$(\mathbf{p} \wedge \mathbf{q}) \wedge \mathbf{r} \leftrightarrow \mathbf{p} \wedge (\mathbf{q} \wedge \mathbf{r})$
T	T	T	T	T	T
T	T	F	F	F	T
T	F	T	F	F	T
T	F	F	F	F	T
F	T	T	F	F	T
F	T	F	F	F	T
F	F	T	F	F	T
F	F	F	F	F	T

4.4.10 Distributivity

The distributive law of algebra permits us to expand $x(y+z)$ into $xy+xz$. Similarly, in logic, the laws of distributivity permit us to write $\mathbf{p} \wedge (\mathbf{q} \vee \mathbf{r}) \leftrightarrow (\mathbf{p} \wedge \mathbf{q}) \vee (\mathbf{p} \wedge \mathbf{r})$ and $\mathbf{p} \vee (\mathbf{q} \wedge \mathbf{r}) \leftrightarrow (\mathbf{p} \vee \mathbf{q}) \wedge (\mathbf{p} \vee \mathbf{r})$, as shown in Tables 4.18 and 4.19.

4.4.11 de Morgan's Laws

The British mathematician and logician Augustus de Morgan (1806–1871) formulated what are now known as de Morgan's Laws, as shown in Tables 4.20 and 4.21.

Table 4.18 Distributivity of $\wedge$ over $\vee$: $\mathbf{p} \wedge (\mathbf{q} \vee \mathbf{r}) \leftrightarrow (\mathbf{p} \wedge \mathbf{q}) \vee (\mathbf{p} \wedge \mathbf{r})$

p	q	r	$\mathbf{p} \wedge (\mathbf{q} \vee \mathbf{r})$	$(\mathbf{p} \wedge \mathbf{q}) \vee (\mathbf{p} \wedge \mathbf{r})$	$\mathbf{p} \wedge (\mathbf{q} \vee \mathbf{r}) \leftrightarrow (\mathbf{p} \wedge \mathbf{q}) \vee (\mathbf{p} \wedge \mathbf{r})$
T	T	T	T	T	T
T	T	F	T	T	T
T	F	T	T	T	T
T	F	F	F	F	T
F	T	T	F	F	T
F	T	F	F	F	T
F	F	T	F	F	T
F	F	F	F	F	T

Table 4.19 Distributivity of $\vee$ over $\wedge$: $\mathbf{p} \vee (\mathbf{q} \wedge \mathbf{r}) \leftrightarrow (\mathbf{p} \vee \mathbf{q}) \wedge (\mathbf{p} \vee \mathbf{r})$

p	q	r	$\mathbf{p} \vee (\mathbf{q} \wedge \mathbf{r})$	$(\mathbf{p} \vee \mathbf{q}) \wedge (\mathbf{p} \vee \mathbf{r})$	$\mathbf{p} \vee (\mathbf{q} \wedge \mathbf{r}) \leftrightarrow (\mathbf{p} \vee \mathbf{q}) \wedge (\mathbf{p} \vee \mathbf{r})$
T	T	T	T	T	T
T	T	F	T	T	T
T	F	T	T	T	T
T	F	F	T	T	T
F	T	T	F	F	T
F	T	F	F	F	T
F	F	T	F	F	T
F	F	F	F	F	T

Table 4.20 de Morgan's Law: $\neg(\mathbf{p} \vee \mathbf{q}) \leftrightarrow \neg\mathbf{p} \wedge \neg\mathbf{q}$

p	q	$\neg(\mathbf{p} \vee \mathbf{q})$	$\neg\mathbf{p} \wedge \neg\mathbf{q}$	$\neg(\mathbf{p} \vee \mathbf{q}) \leftrightarrow \neg\mathbf{p} \wedge \neg\mathbf{q}$
T	T	F	F	T
T	F	F	F	T
F	T	F	F	T
F	F	T	T	T

Table 4.21 de Morgan's Law: $\neg(\mathbf{p} \wedge \mathbf{q}) \leftrightarrow \neg\mathbf{p} \vee \neg\mathbf{q}$

p	q	$\neg(\mathbf{p} \wedge \mathbf{q})$	$\neg\mathbf{p} \vee \neg\mathbf{q}$	$\neg(\mathbf{p} \wedge \mathbf{q}) \leftrightarrow \neg\mathbf{p} \vee \neg\mathbf{q}$
T	T	F	F	T
T	F	T	T	T
F	T	T	T	T
F	F	T	T	T

4.4.12 Simplification

Some statements are so obvious it seems unnecessary to consider them. Nevertheless, their existence should be recognised, as they can help simplify complex logical statements. For example, $\mathbf{p} \vee \mathrm{T}$ must always be true, irrespective of $\mathbf{p}$. Similarly, $\mathbf{p} \wedge \mathrm{F}$ must always be false. Table 4.22 shows the equivalence $\mathbf{p} \vee \mathrm{T} \leftrightarrow \mathrm{T}$, and Table 4.23 shows the equivalence $\mathbf{p} \wedge \mathrm{F} \leftrightarrow \mathrm{F}$. Another form of simplification arises with $\mathbf{p} \vee \mathrm{F}$ and $\mathbf{p} \wedge \mathrm{T}$, which both equal $\mathbf{p}$, as shown in Tables 4.24 and 4.25.

Table 4.22 Simplification: $\mathbf{p} \vee \mathrm{T} \leftrightarrow \mathrm{T}$

p	$\mathbf{p} \vee \mathrm{T}$	$\mathbf{p} \vee \mathrm{T} \leftrightarrow \mathrm{T}$
T	T	T
F	T	T

Table 4.23 Simplification: $\mathbf{p} \wedge \mathrm{F} \leftrightarrow \mathrm{F}$

p	$\mathbf{p} \wedge \mathrm{F}$	$\mathbf{p} \wedge \mathrm{F} \leftrightarrow \mathrm{F}$
T	F	T
F	F	T

Table 4.24 Simplification: $\mathbf{p} \vee \mathrm{F} \leftrightarrow \mathbf{p}$

p	p∨F	p∨F ↔ p
T	T	T
F	F	T

Table 4.25 Simplification: $\mathbf{p} \wedge \mathrm{T} \leftrightarrow \mathbf{p}$

p	p∧T	(p∧T) ↔ p
T	T	T
F	F	T

4.4.13 Excluded Middle

A condition known as the *excluded middle* arises with the choice $\mathbf{p} \vee \neg\mathbf{p}$, which, after a little thought, must always be true. For when **p** is true, its negation is false, and vice versa, which guarantees either scenario being true, as shown in Table 4.26.

4.4.14 Contradiction

A condition known as *contradiction* arises with the combination $\mathbf{p} \wedge \neg\mathbf{p}$, which after a little more thought, must always be false. For **p** and $\neg\mathbf{p}$ can never be equivalent, making a conjunction impossible, as shown in Table 4.27.

Table 4.26 Excluded middle: $(\mathbf{p} \vee \neg\mathbf{p}) \leftrightarrow \mathrm{T}$

p	¬p	p ∨ ¬p
T	F	T
F	T	T

Table 4.27 Contradiction: $(\mathbf{p} \wedge \neg\mathbf{p}) \leftrightarrow \mathrm{F}$

p	¬p	p ∧ ¬p
T	F	F
F	T	F

4.4.15 Double Negation

Knowing that two negatives make a positive, it will come as no surprise that $\neg(\neg\mathbf{p}) \leftrightarrow \mathbf{p}$, as shown in Table 4.28.

4.4.16 Implication and Equivalence

The truth table for implication has already been covered in Table 4.6, however, implication is also expressed by $\mathbf{p} \rightarrow \mathbf{q} \leftrightarrow \neg\mathbf{p} \vee \mathbf{q}$, as shown in Table 4.29. Similarly, the truth table for equivalence (Table 4.5) is also expressed by $(\mathbf{p} \leftrightarrow \mathbf{q}) \leftrightarrow (\mathbf{p} \rightarrow \mathbf{q}) \wedge (\mathbf{q} \rightarrow \mathbf{p})$, as shown in Table 4.30.

4.4.17 Exportation

Exportation covers the equivalence $(\mathbf{p} \wedge \mathbf{q}) \rightarrow \mathbf{r} \leftrightarrow \mathbf{p} \rightarrow (\mathbf{q} \rightarrow \mathbf{r})$, as shown in Table 4.31.

4.4.18 Contrapositive

The law of *contrapositive*, is also known as *modus tollens*, and acknowledges the negative form of $\mathbf{p} \rightarrow \mathbf{q}$, i.e. $\neg\mathbf{q} \rightarrow \neg\mathbf{p}$, as shown in Table 4.32.

Table 4.28 Double negation: $\neg(\neg\mathbf{p}) \leftrightarrow \mathbf{p}$

p	¬(¬p)	¬(¬p) ↔ p
T	T	T
F	F	T

Table 4.29 Implication: $\mathbf{p} \rightarrow \mathbf{q} \leftrightarrow \neg\mathbf{p} \vee \mathbf{q}$

p	q	p → q	¬p ∨ q	p → q ↔ ¬p ∨ q
T	T	T	T	T
T	F	F	F	T
F	T	T	T	T
F	F	T	T	T

Table 4.30 Equivalence: $(\mathbf{p} \leftrightarrow \mathbf{q}) \leftrightarrow (\mathbf{p} \rightarrow \mathbf{q}) \wedge (\mathbf{q} \rightarrow \mathbf{p})$

p	q	$p \leftrightarrow q$	$p \rightarrow q$	$q \rightarrow p$	$A \equiv (p \rightarrow q) \wedge (q \rightarrow p)$	$(p \leftrightarrow q) \leftrightarrow A$
T	T	T	T	T	T	T
T	F	F	F	T	F	T
F	T	F	T	F	F	T
F	F	T	T	T	T	T

Table 4.31 Exportation: $(\mathbf{p} \wedge \mathbf{q}) \rightarrow \mathbf{r} \leftrightarrow \mathbf{p} \rightarrow (\mathbf{q} \rightarrow \mathbf{r})$

p	q	r	$p \wedge q$	$A \equiv (p \wedge q) \rightarrow r$	$q \rightarrow r$	$B \equiv p \rightarrow (q \rightarrow r)$	$A \leftrightarrow B$
T	T	T	T	T	T	T	T
T	T	F	T	F	F	F	T
T	F	T	F	T	T	T	T
T	F	F	F	T	T	T	T
F	T	T	F	T	T	T	T
F	T	F	F	T	F	T	T
F	F	T	F	T	T	T	T
F	F	F	F	T	T	T	T

Table 4.32 Contrapositive: $\mathbf{p} \rightarrow \mathbf{q} \leftrightarrow \neg\mathbf{q} \rightarrow \neg\mathbf{p}$

p	q	$\neg p$	$\neg q$	$p \rightarrow q$	$\neg q \rightarrow \neg p$	$p \rightarrow q \leftrightarrow \neg q \rightarrow \neg p$
T	T	F	F	T	T	T
T	F	F	T	F	F	T
F	T	T	F	T	T	T
F	F	T	T	T	T	T

4.4.19 Reductio Ad Absurdum

Reductio Ad Absurdum is Latin for "reduction to absurdity" and describes a form of argument that proposes a statement is true, by claiming that its opposite leads to an impossible or absurd result. For example, "men must have knees, otherwise they wouldn't be able to kneel down and propose!" There is a wonderful story concerning Bertrand Russell, who, during a lecture on logic, mentioned that in the sense of material implication, a false proposition implies any proposition. A bright student raised his hand and said "In that case, given that $1 = 0$, prove that you are the Pope".

Table 4.33 Reductio ad absurdum: $(\mathbf{p} \to \mathbf{q}) \wedge (\mathbf{p} \to \neg\mathbf{q}) \to \neg\mathbf{p}$

p	q	$p \to q$	$p \to \neg q$	$(p \to q) \wedge (p \to \neg q)$	$(p \to q) \wedge (p \to \neg q) \to \neg p$
T	T	F	F	F	T
T	F	F	T	F	T
F	T	T	T	T	T
F	F	T	T	T	T

Table 4.34 Reductio ad absurdum: $(\neg\mathbf{p} \to \mathbf{q}) \wedge (\neg\mathbf{p} \to \neg\mathbf{q}) \to \mathbf{p}$

p	q	$\neg p \to q$	$\neg p \to \neg q$	$(\neg p \to q) \wedge (\neg p \to \neg q)$	$(\neg p \to q) \wedge (\neg p \to \neg q) \to p$
T	T	T	T	T	T
T	F	T	T	T	T
F	T	F	F	F	T
F	F	F	T	F	T

Russell immediately replied, "Add 1 to both sides of the equation: then we have $2 = 1$. The set containing just me and the Pope has 2 members. But $2 = 1$, so it has only 1 member; therefore, I am the Pope."

We can express this fallacious form of argument by examining the conjunction: $(\mathbf{p} \to \mathbf{q}) \wedge (\mathbf{p} \to \neg\mathbf{q})$, which posits that $\mathbf{p}$ implies $\mathbf{q}$ and $\neg\mathbf{q}$, which is absurd. Table 4.33 reveals that $(\mathbf{p} \to \mathbf{q}) \wedge (\mathbf{p} \to \neg\mathbf{q}) \to \neg\mathbf{p}$, which is a useless result, and Table 4.34 shows its negative form.

4.4.20 Modus Ponens

Modus ponens is Latin for *affirming mode* and describes an argument containing three parts: a major premise, a minor premise, and a conclusion. For example, "If an integer is positive, then it is a natural number", is the major proposition. "23 is a positive integer", is the minor proposition. From which, we conclude that 23 is a natural number. By letting $\mathbf{p}$ stand for a positive integer, and $\mathbf{q}$ stand for a natural number, the major proposition takes the form $\mathbf{p} \to \mathbf{q}$. The minor proposition is simply $\mathbf{p}$, which makes $(\mathbf{p} \to \mathbf{q}) \wedge \mathbf{p}$, which in turn, implies $\mathbf{q}$, as shown in Table 4.35.

Table 4.35 Modus Ponens: $(\mathbf{p} \rightarrow \mathbf{q}) \wedge \mathbf{p} \rightarrow \mathbf{q}$

p	q	$p \rightarrow q$	$(p \rightarrow q) \wedge p$	$(p \rightarrow q) \wedge p \rightarrow q$
T	T	T	T	T
T	F	F	T	T
F	T	T	F	T
F	F	T	F	T

4.4.21 Proof by Cases

Table 4.36 shows the principle known as *proof by cases* where if at least one of **p** or **q** is true, and each implies **r**, then **r** must be true as well: $[(\mathbf{p} \vee \mathbf{q}) \wedge (\mathbf{p} \rightarrow \mathbf{r}) \wedge (\mathbf{q} \rightarrow \mathbf{r})] \rightarrow \mathbf{r}$.

Truth tables are extremely useful in the design of microelectronic logic gates, whose elements contain AND, OR, NOT, NAND, NOR, XOR and XNOR. They also play a role in clarifying logical outcomes in programming, but their principal weakness is their size, which is proportional to 2^n, where n is the number of logical premises. The above truth tables are summarised in Table 4.37.

Table 4.36 Proof by Cases: $[(\mathbf{p} \vee \mathbf{q}) \wedge (\mathbf{p} \rightarrow \mathbf{r}) \wedge (\mathbf{q} \rightarrow \mathbf{r})] \rightarrow \mathbf{r}$

p	q	r	$p \vee q$	$p \rightarrow r$	$q \rightarrow r$	$A \equiv [(p \vee q) \wedge (p \rightarrow r) \wedge (q \rightarrow r)]$	$A \rightarrow r$
T	T	T	T	T	T	T	T
T	T	F	T	F	F	F	T
T	F	T	T	T	T	T	T
T	F	F	T	F	T	F	T
F	T	T	T	T	T	T	T
F	T	F	T	T	F	F	T
F	F	T	F	T	T	F	T
F	F	F	F	T	T	F	T

Table 4.37 Logical identities

Identity	Law
$\neg(\neg p) \leftrightarrow p$	Double negation
$p \vee T \leftrightarrow T$	Simplification
$p \wedge F \leftrightarrow F$	Simplification
$p \vee F \leftrightarrow p$	Simplification
$p \wedge T \leftrightarrow p$	Simplification
$p \vee \neg p \leftrightarrow T$	Excluded middle
$p \wedge \neg p \leftrightarrow F$	Contradiction
$p \vee p \leftrightarrow p$	Idempotence of $\vee$
$p \wedge p \leftrightarrow p$	Idempotence of $\wedge$
$p \vee q \leftrightarrow q \vee p$	Commutativity of $\vee$
$p \wedge q \leftrightarrow q \wedge p$	Commutativity of $\wedge$
$(p \vee q) \vee r \leftrightarrow p \vee (q \vee r)$	Associativity of $\vee$
$(p \wedge q) \wedge r \leftrightarrow p \wedge (q \wedge r)$	Associativity of $\wedge$
$\neg(p \vee q) \leftrightarrow \neg p \wedge \neg q$	de Morgan's Law
$\neg(p \wedge q) \leftrightarrow \neg p \vee \neg q$	de Morgan's Law
$p \wedge (q \vee r) \leftrightarrow (p \wedge q) \vee (p \wedge r)$	Distributivity of $\wedge$ over $\vee$
$p \vee (q \wedge r) \leftrightarrow (p \vee q) \wedge (p \vee r)$	Distributivity of $\vee$ over $\wedge$
$p \rightarrow q \leftrightarrow \neg p \vee q$	Implication
$p \leftrightarrow q \leftrightarrow (p \rightarrow q) \wedge (q \rightarrow p)$	Equivalence
$(p \wedge q) \rightarrow r \leftrightarrow p \rightarrow (q \rightarrow r)$	Exportation
$p \rightarrow q \leftrightarrow \neg q \rightarrow \neg p$	Contrapositive
$(p \rightarrow q) \wedge (p \rightarrow \neg q) \rightarrow \neg p$	Reductio ad absurdum
$(\neg p \rightarrow q) \wedge (\neg p \rightarrow \neg q) \rightarrow p$	Reductio ad absurdum
$(p \rightarrow q) \wedge q \rightarrow q$	Modus ponens
$(p \vee q) \wedge (p \rightarrow r) \wedge (q \rightarrow r) \rightarrow r$	Proof by cases

4.5 Set Theory

We have already covered some of the ideas behind set theory, especially Cantor's work in classifying infinite sets. Bertrand Russell's paradox showed that the set of all sets could never be a set referenced by another set, which kept alive the search for an alternative system. In 1902, the German logician and mathematician Ernst Zermelo (1871–1953) published a paper on adding transfinite cardinals, and in 1908 published another paper on Zermelo Set Theory. This was developed by the Norwegian math-

ematician Thoralf Skolem (1887–1963) and the German-born Israeli mathematician Abraham Fraenkel (1891–1965), which resulted in the Zermelo-Fraenkel set theory, abbreviated to ZF . This system builds upon the empty set and Cantor's power set, but does not accept that all collections of sets constitute a set, which prevents recursive scenarios.

In the late 19th-century, the English logician and philosopher John Venn (1834–1923) introduced a graphical technique using circles to represent sets, whose relationships are reflected in the nesting or intersection of the circles. He referred to them as *Euler diagrams*, as Euler had previously used them. However, Venn popularised their usage, which is why now they bear his name: *Venn Diagrams*.

Developing the definition that a set is a collection of objects, let's examine empty sets, set building, combining sets, and Cantor's power set, using Venn diagrams.

4.5.1 Empty Set

An *empty set* is a set with no members. For example, if a farmer owns a set comprising a field of five donkeys {Betty, George, Albert, Descartes, Mary} and one night they are all stolen, the farmer now owns an empty set in the form of an empty field, represented by $\emptyset$, or {}. One is tempted to question whether the field really belongs to the set of donkeys, to which the answer is no, but it does provide a mental device for visualising the concept of emptiness. Although the concept of an empty set is fundamental to ZF, and can be manipulated constructively, it is an abstract idea and remains a topic for continued discussion.

4.5.2 Membership and Cardinality of a Set

We have already discovered that the symbol $\in$ means "member of" or "belongs to" which permits us to write:

$$\text{if} \quad S = \{a, e, i, o, u\} \quad \text{then} \quad a \in S,\ e \in S,\ i \in S,\ o \in S,\ u \in S.$$

In this example, the set S contains 5 elements, written $|S| = 5$, and is called the *cardinality* of S.

A set element may also be another set. For example,

$$\text{if} \quad S = \{a, b, c, \{d, e\}\} \quad \text{then} \quad a \in S,\ b \in S,\ c \in S,\ \{d, e\} \in S.$$

Therefore, the set's cardinality is $|S| = 4$.

4.5.3 Subsets, Supersets and the Universal Set

The set of natural numbers $\mathbb{N}$ can be divided into various subsets. For example, if E is the set of all even natural numbers, and O is the set of all odd natural numbers, then E and O are *subsets* of $\mathbb{N}$, written $E \subset \mathbb{N}$, and $O \subset \mathbb{N}$. (The symbol $\subseteq$ is also used in place of $\subset$.) For any problem associated with sets, there is an associated *universal set* to which the sets belong. In this case, $\mathbb{N}$ is the universal set, as shown in the Venn diagram in Fig. 4.1.

Conversely, if $E \subset \mathbb{N}$ and $O \subset \mathbb{N}$, then $\mathbb{N}$ is a *superset* of E and O, written $\mathbb{N} \supset E$ and $\mathbb{N} \supset O$. (The symbol $\supseteq$ is also used in place of $\supset$.)

The two sets E and O can also be divided into two subsets P_O for odd primes, and P_E for even primes, as shown in Fig. 4.2. We could have also added the subset of primes P, as shown in Fig. 4.3.

4.5.4 Set Building

A set is constructed using the notation {<variable>|<predicate>} or {<variable>: <predicate>}, where <variable> is an arbitrary name given to the set's elements,

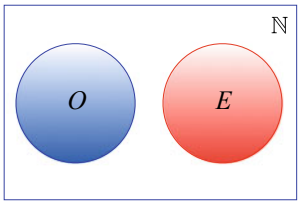

Fig. 4.1 Odd and even numbers as subsets of $\mathbb{N}$

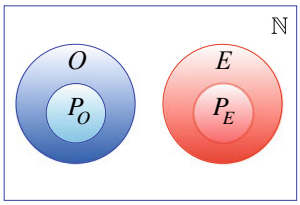

Fig. 4.2 P_O and P_E as subsets of O and E

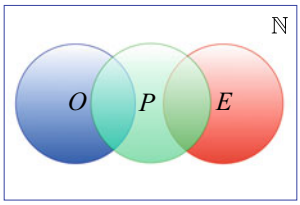

Fig. 4.3 P as a subset of both O and E

and <predicate> is a logical filter associated with the variable. For example, $S = \{n \mid n \in \mathbb{N} \wedge n \geq 23\}$, reads "The elements of set S comprise n, such that n is a natural number and is greater than, or equal to 23", i.e. $S = \{23, 24, 25, 26, 27, \dots\}$.

Here are some more examples, where the *existential quantifier* $\exists$ is introduced, which stands for "there exists". For example, $\exists m \in \mathbb{N}\ n = 2m$, reads "there exists an m belonging to $\mathbb{N}$ where $n = 2m$".

$$\begin{aligned} S &= \{n \mid n \in \mathbb{N} \wedge 1 \leq n \leq 5\} &&= \{1, 2, 3, 4, 5\} \\ S &= \{n \mid \exists m \in \mathbb{N}\ n = 2m\} &&= \{2, 4, 6, 8, 10, \dots\} \\ S &= \{n \mid \exists m \in \mathbb{N}\ n = 2m - 1\} &&= \{1, 3, 5, 7, 9, \dots\} \\ S &= \{n \mid \exists m \in \mathbb{N}\ n = m^2\} &&= \{1, 4, 9, 16, 25, \dots\}. \end{aligned}$$

4.5.5 Union

The *union* of two sets A and B, is another set combining their respective elements without duplications, and is written $A \cup B$, which reads "A union B", or "the union of A and B".

Figure 4.4 shows the Venn diagram for two sets A and B, which are subsets of the universal set U, and their union is represented by the complete shaded area. For example, if

$$\begin{aligned} A &= \{\text{Euler, Newton, Russell, Cantor}\} \\ B &= \{\text{Gauss, Peano, Russell, Cantor}\} \\ A \cup B &= \{\text{Euler, Newton, Russell, Cantor, Gauss, Peano}\}. \end{aligned}$$

Similarly,

$$\begin{aligned} A &= \{1, 4, 6, 8, \{23, 41\}\} \\ B &= \{\{23, 41\}, 3, 5, 8\} \\ A \cup B &= \{\{23, 41\}, 1, 3, 4, 5, 6, 8\}. \end{aligned}$$

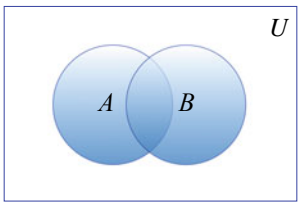

Fig. 4.4 The union of A and B: $A \cup B$

4.5.6 Intersection

The *intersection* of two sets A and B, is another set containing their common elements, and is written $A \cap B$, which reads "A intersection B", or "the intersection of A and B".

Figure 4.5 shows the Venn diagram for two sets A and B, which are subsets of the universal set U, and their intersection is represented by the shaded area. For example,

$$\begin{aligned} A &= \{\text{Euler, Newton, Russell, Cantor}\} \\ B &= \{\text{Gauss, Peano, Russell, Cantor}\} \\ A \cap B &= \{\text{Russell, Cantor}\}. \end{aligned}$$

4.5.7 Relative Complement

The *relative complement* between two sets is the set of elements belonging to one, but not the other. For example, the relative complement of B in A is written $A \setminus B$, and represents all the elements belonging to A, and not B. For example,

$$\begin{aligned} A &= \{\text{Euler, Newton, Russell, Cantor}\} \\ B &= \{\text{Gauss, Peano, Russell, Cantor}\} \\ A \setminus B &= \{\text{Euler, Newton}\} \\ B \setminus A &= \{\text{Gauss, Peano}\}. \end{aligned}$$

Figure 4.6 shows the relationship $A \setminus B$, and Fig. 4.7 shows the relationship $B \setminus A$.

As this can be a difficult relationship to grasp, let's give a formal definition of A in B, and B in A:

$B \setminus A = \{x \mid x \in B \wedge x \notin A\}$ which reads "an element x is in B but not A".
$A \setminus B = \{x \mid x \in A \wedge x \notin B\}$ which reads "an element x is in A but not B".

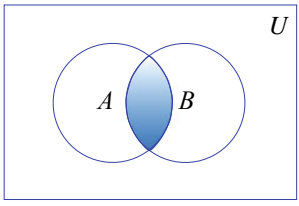

Fig. 4.5 The intersection of A and B: $A \cap B$

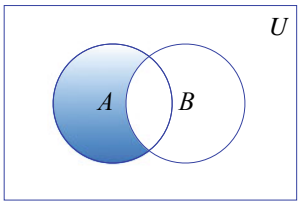

Fig. 4.6 The relative complement of B in A: $A \setminus B$

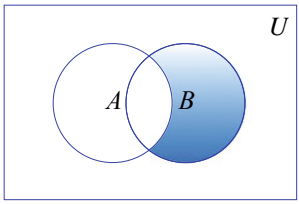

Fig. 4.7 The relative complement of A in B: $B \setminus A$

Examples:

$$\{a, b, c, d\} \setminus \{b, c, d, e\} = \{a\}$$
$$\{\text{john, heidi, edwin, marie}\} \setminus \{\text{edwin, marie}\} = \{\text{john, heidi}\}$$
$$\{23, 5, 41, 27, 3, 29, 2\} \setminus \{23, 5, 19, 41, 44, 29, 2\} = \{27, 3\}.$$

4.5.8 Absolute Complement

The *absolute complement* (or *complement*) of a set A is the difference between the associated universal set U and A, and written $A^c = U \setminus A$, as illustrated in Fig. 4.8. For example, if the universal set is $\mathbb{N}$, and $A = \{n \mid n \in \mathbb{N} \wedge n > 1\} = \{1\}$, then $A^c = \mathbb{N} \setminus A\} = 1$. It follows that $U^c = \emptyset$ and $\emptyset^c = U$.

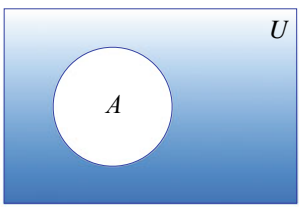

Fig. 4.8 The absolute complement $A^c = U \setminus A$

Table 4.38 Power sets for different sets

S	$\mathcal{P}(S)$	$\|S\|$
{a}	$\{\emptyset, a\}$	2
{a,b}	$\{\emptyset, a, b, \{a, b\}\}$	4
{a,b,c}	$\{\emptyset, a, b, c, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$	8

4.5.9 Power Set

Given a set S, Cantor's *power set* is the set of all subsets of S, which includes S and the empty set $\emptyset$, and is written $\mathcal{P}(S)$. Therefore, if $S = \{a\}$, then $\mathcal{P}(S) = \{\emptyset, a\}$. Table 4.38 shows the power sets for sets with 1, 2 and 3 elements, where it become clear that if a set has n elements, its power set contains 2^n elements. Cantor used the power set and the idea of one-to-one correspondence to reveal an infinite hierarchy of infinities.

4.6 Worked Examples

4.6.1 Truth Tables

Design the truth tables for:

$$(\mathbf{p} \wedge \mathbf{q}) \vee \neg \mathbf{q}$$
$$(\mathbf{p} \wedge \mathbf{q}) \vee \neg \mathbf{p}$$
$$(\mathbf{p} \vee \mathbf{q}) \wedge \neg \mathbf{q}$$
$$(\mathbf{p} \vee \mathbf{q}) \wedge \neg \mathbf{p}.$$

Solutions: See Tables 4.39, 4.40, 4.41 and 4.42.

4.6.2 Set Building

State the sets or the positive real numbers, and the positive integers greater than 100. Solutions:

$$S = \{x \mid x \in \mathbb{R} \wedge x > 0\}$$
$$S = \{n \mid n \in \mathbb{Z} \wedge n > 100\} = \{101, 102, 103, 104, \dots\}.$$

Table 4.39 Truth table for $(\mathbf{p} \wedge \mathbf{q}) \vee \neg\mathbf{q}$

p	q	$p \wedge q$	$\neg q$	$(p \wedge q) \vee \neg q$
T	T	T	F	T
T	F	F	T	T
F	T	F	F	F
F	F	F	T	T

Table 4.40 Truth table for $(\mathbf{p} \wedge \mathbf{q}) \vee \neg\mathbf{p}$

p	q	$p \wedge q$	$\neg p$	$(p \wedge q) \vee \neg p$
T	T	T	F	T
T	F	F	F	F
F	T	F	T	T
F	F	F	T	T

Table 4.41 Truth table for $(\mathbf{p} \vee \mathbf{q}) \wedge \neg\mathbf{q}$

p	q	$p \vee q$	$\neg q$	$(p \vee q) \wedge \neg q$
T	T	T	F	F
T	F	T	T	T
F	T	T	F	F
F	F	F	T	F

Table 4.42 Truth table for $(\mathbf{p} \vee \mathbf{q}) \wedge \neg\mathbf{p}$

p	q	$p \vee q$	$\neg p$	$(p \vee q) \wedge \neg p$
T	T	T	F	F
T	F	T	F	F
F	T	T	T	T
F	F	F	T	F

4.6.3 Sets

Given

$$A = \{1, 2, 3, 4, 5\}$$
$$B = \{2, 4, 6\}$$

find $A \cup B$, $A \cap B$, $A \setminus B$, $B \setminus A$.
Solution:

$$A \cup B = \{1, 2, 3, 4, 5, 6\}$$
$$A \cap B = \{2, 4\}$$
$$A \setminus B = \{1, 3, 5\}$$
$$B \setminus A = \{6\}.$$

4.6.4 Power Set

Specify the power set for $S = \{1, 2, 3, 4\}$.
Solution:

$$\mathcal{P}(S) = \{\emptyset, 1, 2, 3, 4, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\}.$$

Chapter 5
Combinatorics

5.1 Introduction

Combinatorics includes permutations and combinations, which are introduced in this chapter and formulae derived and illustrated with examples. The chapter concludes with a variety of worked examples.

5.2 Permutations

Given a set of a finite number of distinct elements, a *permutation* is an ordered arrangement of all or part of these elements, and is sensitive to their order. *Braces* are used for unordered sets, and *parentheses* for ordered sets.

As it is possible to select a group of elements, rather than the entire set, the following notation is often used to represent the number of permutations: $P(n, k)$, where n is the number of elements in the set, and k is the number of elements taken at a time. Let's illustrate this with sets of two, three and four letters.

The set of two letters {A, B} can be ordered in two ways taking all the elements:

$$P(2, 2) = 2, \quad (\text{A, B}) \quad (\text{B, A}).$$

The set of three letters {A, B, C} can be ordered in six ways taking all the elements:

$$P(3, 3) = 6, \quad (\text{A, B, C}) \quad (\text{A, C, B}) \quad (\text{B, A, C}) \quad (\text{B, C, A}) \quad (\text{C, A, B}) \quad (\text{C, B, A}).$$

The set of four letters {A, B, C, D} can be ordered in twenty-four ways taking all the elements, $P(4, 4) = 24$, as shown in Table 5.1.

Using the last example, the first letter has four possibilities; the second letter has three possibilities; the third letter has two possibilities; and the last letter has one possibility. Therefore, the number of permutations is $P(4, 4) = 4 \times 3 \times 2 \times 1 = 24$,

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_5

Table 5.1 Twenty-four ways the letters {A, B, C, D} can be ordered

(A, B, C, D)	(A, B, D, C)	(A, C, B, D)	(A, C, D, B)	(A, D, B, C)	(A, D, C, B)
(B, A, C, D)	(B, A, D, C)	(B, C, A, D)	(B, C, D, A)	(B, D, A, C)	(B, D, C, A)
(C, A, B, D)	(C, A, D, B)	(C, B, A, D)	(C, B, D, A)	(C, D, A, B)	(C, D, B, A)
(D, A, B, C)	(D, A, C, B)	(D, B, A, C)	(D, B, C, A)	(D, C, A, B)	(D, C, B, A)

and $P(n, n) = n!$. For example, six different coloured tulip bulbs can be planted in a row, in $P(6, 6) = 6 \times 5 \times 4 \times 3 \times 2 = 720$ permutations.

Now let's take ordered pairs from sets of two, three and four letters.
The set of two letters {A, B} can be ordered in two ways taking two elements at a time:

$$P(2, 2) = 2, \quad (\text{A, B}) \quad (\text{B, A}).$$

The set of three letters {A, B, C} can be ordered in six ways taking two elements at a time:

$$P(3, 2) = 6, \quad (\text{A, B}) \quad (\text{A, C}) \quad (\text{B, A}) \quad (\text{B, C}) \quad (\text{C, A}) \quad (\text{C, B}).$$

The set of four letters {A, B, C, D} can be ordered in twelve ways taking two elements at a time:

$$P(4, 2) = 12, (\text{A, B}) \quad (\text{A, C}) \quad (\text{A, D}) \quad (\text{B, A}) \quad (\text{B, C}) \quad (\text{B, D})$$
$$(\text{C, A}) \quad (\text{C, B}) \quad (\text{C, D}) \quad (\text{D, A}) \quad (\text{D, B}) \quad (\text{D, C}).$$

A formula that satisfies these three examples is

$$P(n, 2) = n(n - 1), \quad P(2, 2) = 2, \quad P(3, 2) = 6, \quad P(4, 2) = 12.$$

Let's derive a formula for selecting any number of elements from a set of distinct elements.

Given a set of n elements, we can construct **all** possible permutations by selecting the elements one at a time. The first element can be chosen in n different ways. The second element in $(n - 1)$ ways, the third in $(n - 2)$ ways, etc. Thus, the number of all permutations that are possible with n elements is

$$P(n, n) = n(n - 1)(n - 2) \times \cdots \times 3 \times 2 \times 1 = n!.$$

If we start with n elements and construct the permutations with k positions, the first element can be chosen in n different ways. There are $n - 1$ elements left for the next $k - 1$ positions. The second position has $n - 1$ elements, the third $n - 2$, and so on;

for the kth position there are $n - (k - 1)$ elements.

$$\begin{aligned} P(n,k) &= n(n-1)(n-2) \times \cdots \times [n-(k-1)] \\ &= n(n-1)(n-2) \times \cdots \times (n-k+1) \\ &= \frac{n(n-1)(n-2) \times \cdots \times (n-k+1)(n-k)!}{(n-k)!} \\ &= \frac{n(n-1)(n-2) \times \cdots \times (n-k+1)(n-k) \times \cdots \times 3 \times 2 \times 1}{(n-k)!} \\ &= \frac{n!}{(n-k)!} \end{aligned}$$

where $n, k \in \mathbb{N}, \quad n \geq k$.

As a simple test, when $n = 4$ and $k = 2$,

$$P(4,2) = \frac{4!}{2!} = 12$$

which agrees with our original example.

To compute the number of ordered sets, or permutations, taking two letters at a time from the unordered set {C, O, M, P, U, T, E, R} we have $n = 8$ and $k = 2$:

$$\begin{aligned} P(8,2) &= \frac{8!}{(8-2)!} \\ &= \frac{8!}{6!} \\ &= 8 \times 7 \\ &= 56. \end{aligned}$$

Table 5.2 lists the 56 two-letter sets for {C, O, M, P, U, T, E, R}.

Table 5.2 Two-letter permutations of the word COMPUTER

(C, E)	(C, O)	(C, M)	(C, P)	(C, R)	(C, T)	(C, U)
(O, C)	(O, E)	(O, M)	(O, P)	(O, R)	(O, T)	(O, U)
(M, C)	(M, E)	(M, O)	(M, P)	(M, R)	(M, T)	(M, U)
(P, C)	(P, E)	(P, O)	(P, M)	(P, R)	(P, T)	(P, U)
(U, C)	(U, E)	(U, O)	(U, M)	(U, P)	(U, R)	(U, T)
(T, C)	(T, E)	(T, O)	(T, M)	(T, P)	(T, R)	(T, U)
(E, C)	(E, O)	(E, M)	(E, P)	(E, R)	(E, T)	(E, U)
(R, C)	(R, E)	(R, O)	(R, M)	(R, P)	(R, T)	(R, U)

5.3 Permutations of Multisets

A set **must** contain a collection of unique elements, whereas a *multiset* contains repetitions of one or more of its elements. The number of repetitions is called the element's *multiplicity*. For example, in the multiset {M, A, N, H, A, T, T, A, N},

the element M has a multiplicity of 1
the element A has a multiplicity of 3
the element N has a multiplicity of 2
the element H has a multiplicity of 1
the element T has a multiplicity of 2

and can be written {1M, 3A, 2N, 1H, 2T}.

Given a multiset M with n elements, an r-permutation of M is an ordered arrangements of r elements of M. However, an n-permutation of M is called a permutation of M.

In order to calculate the number of permutations of a multiset M containing n elements, we let n_i be the multiplicity for each element, where $i = 1, k$. Therefore, $n_1 + n_2 + n_3 + \cdots + n_k = n$.

In order to calculate the number of permutations of M, we begin with n free places for n_1 elements: $\binom{n}{n_1}$.

There now remain $n - n_1$ free places for n_2 elements: $\binom{n-n_1}{n_2}$.

There now remain $n - n_1 - n_2$ free places for n_3 elements: $\binom{n-n_1-n_2}{n_3}$.

And so on, until the kth elements: $\binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k}$.

The total number of permutations of M is given by

$$\begin{aligned}
&\binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \cdots \times \binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k} \\
&= \frac{n!}{n_1!(n-n_1)!} \times \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \times \cdots \times \frac{n-n_1-n_2-n_3-\cdots-n_{k-1}}{n_k!(n-n_1-n_2-n_3-\cdots-n_k)!} \\
&= \frac{n!}{n_1!n_2!n_3!\cdots n_k!}.
\end{aligned}$$

For example, the multiset {M, A, N, H, A, T, T, A, N} = {1M, 3A, 2N, 1H, 2T}. where $n = 9, \quad n_1 = 1, \quad n_2 = 3, \quad n_3 = 2, \quad n_4 = 1, \quad n_5 = 2$. Therefore, the number of permutations of the word is given by

$$\begin{aligned}\frac{n!}{n_1!n_2!n_3!n_4!n_5!} &= \frac{9!}{1!3!2!1!2!}\\ &= \frac{9\times 8\times 7\times 6\times 5\times 4\times 3\times 2}{3\times 2\times 2\times 2}\\ &= 9\times 8\times 7\times 6\times 5\\ &= 15,120.\end{aligned}$$

Similarly, the word REDDER is the multiset {2R, 2E, 2D}, where $n = 6, \quad n_1 = 2, n_2 = 2, \quad n_3 = 2.$

Therefore, the number of permutations of the word is given by

$$\begin{aligned}\frac{n!}{n_1!n_2!n_3!} &= \frac{6!}{2!2!2!}\\ &= \frac{6\times 5\times 4\times 3\times 2}{2\times 2\times 2}\\ &= 6\times 5\times 3\\ &= 90.\end{aligned}$$

In order to visualise the permutations of a multiset, let's consider a shorter word. The word NOON is the multiset {2N, 2O}, where $n = 4, \quad n_1 = 2, \quad n_2 = 2.$

Therefore, the number of permutations of the word is given by

$$\begin{aligned}\frac{n!}{n_1!n_2!} &= \frac{4!}{2!2!}\\ &= \frac{4\times 3\times 2}{2\times 2}\\ &= 3\times 2\\ &= 6.\end{aligned}$$

The permutations are NNOO, NOON, NONO, ONNO, ONON, OONN.

5.4 Combinations

Unlike permutations, combinations disregard the order of elements created from a set. Thus (B, A) is the same as (A, B), and only one is selected. The disregard for order conflicts with the use of the word combination in the context of locks, where a combination lock depends upon the correct order of digits to open the lock. However, we are where we are, and we must pay special attention to the mathematical usage of the word.

When selecting two elements from a set of three elements {A, B, C} creates six permutations:

$$P(3,2) = 6, \quad (\text{A, B}) \quad (\text{B, A}) \quad (\text{A, C}) \quad (\text{C, A}) \quad (\text{B, C}) \quad (\text{C,B}).$$

However, the number of combinations is reduced to three:

$$\binom{3}{2} = (\text{A, B}) \quad (\text{A, C}) \quad (\text{B, C}),$$

and in this context, the relationship between combinations and permutations is

$$\binom{3}{2} = \frac{P(3,2)}{2}.$$

Therefore, given a set of n distinct elements, and taking k elements at a time, there are $\binom{n}{k}$ combinations. We now require a general formula for the number of combinations.

Table 5.1 shows the twenty-four permutations of the set {A, B, C, D} taking all the elements at a time. However, if we disregard those permutations where the elements are ordered differently, we end up with only four combinations:

$$(\text{A, B, C}) \quad (\text{A, B, D}) \quad (\text{A, C, D}) \quad (\text{B, C, D}).$$

As each combination creates $3! = 6$ ordered triples, we can generalise the following relationship:

$$P(n,k) = \binom{n}{k} k! \quad \text{or} \quad \binom{n}{k} = \frac{P(n,k)}{k!} = \frac{n!}{(n-k)!k!}.$$

Thus (5.1) gives the number of combinations of a set of n distinct elements taken k at a time.

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}, \quad n, k \in \mathbb{N}, \quad n \geq k. \tag{5.1}$$

Substituting some values for n and k into (5.1) we get

$$\begin{aligned}
&\binom{2}{1} = \frac{2!}{1!1!} = 2, \binom{2}{2} = \frac{2!}{0!2!} = 1 \\
&\binom{3}{1} = \frac{3!}{2!1!} = 3, \binom{3}{2} = \frac{3!}{1!2!} = 3, \binom{3}{3} = \frac{3!}{0!3!} = 1 \\
&\binom{4}{1} = \frac{4!}{3!1!} = 4, \binom{4}{2} = \frac{4!}{2!2!} = 6, \binom{4}{3} = \frac{4!}{1!3!} = 4, \binom{4}{4} = \frac{4!}{0!4!} = 1
\end{aligned}$$

which are identical to the *binomial coefficients* for $x^1, x^2, \cdots, x^n$ in the expansion $(1+x)^n$:

Table 5.3 Two-letter combinations of the word COMPUTER

(C, E)	(C, O)	(C, M)	(C, P)	(C, R)	(C, T)	(C, U)
(O, E)	(O, M)	(O, P)	(O, R)	(O, T)	(O, U)	
(M, E)	(M, P)	(M, R)	(M, T)	(M, U)		
(P, E)	(P, R)	(P, T)	(P, U)			
(U, E)	(U, R)	(U, T)				
(T, E)	(T, R)					
(E, R)						

$$(1+x)^2 = 1 + 2x + 1x^2$$
$$(1+x)^3 = 1 + 3x + 3x^2 + 1x^3$$
$$(1+x)^4 = 1 + 4x + 6x^2 + 4x^3 + 1x^4.$$

Combinations are used in problems where we are interested in selecting groups of things, irrespective of order, from some collection. For example, let's find the number of combinations in the set of letters {C, O, M, P, U, T, E, R} taking two letters at a time. As there are eight letters, $n = 8$, and $k = 2$. Therefore

$$\binom{8}{2} = \frac{8!}{(8-2)!2!} = \frac{8 \times 7}{2} = 28.$$

The combinations, which have to be derived manually, are shown in Table 5.3

5.5 Worked Examples

5.5.1 Eight-Permutations of a Multiset

Find the number of 8-permutations of the multiset

$$M = \{\text{A, A, B, B, B, B, C, C, C}\} = \{\text{2A, 4B, 3C}\}.$$

Solution: We observe that M contains 9 elements, but have to limit our permutations to 8 elements. The extra element can be removed either from the 2As, the 4Bs, or the 3Cs, which means that we must sum the permutations contributed by each possibility.

Removing it from the 2As, the number of 8-permutations of $\{A, 4B, 3C\}$: $\frac{8!}{1!4!3!}$.

Removing it from the 4Bs, the number of 8-permutations of $\{2A, 3B, 3C\}$: $\frac{8!}{2!3!3!}$.

Removing it from the 3Cs, the number of 8-permutations of $\{2A, 4B, 2C\}$: $\frac{8!}{2!4!2!}$.
Computing the three contributions:

$$\begin{aligned}\frac{8!}{1!4!3!} &= \frac{8\times 7\times 6\times 5\times 4\times 3\times 2}{4\times 3\times 2\times 3\times 2}\\ &= 8\times 7\times 5 = 280\\ \frac{8!}{2!3!3!} &= \frac{8\times 7\times 6\times 5\times 4\times 3\times 2}{2\times 3\times 2\times 3\times 2}\\ &= 8\times 7\times 5\times 2 = 560\\ \frac{8!}{2!4!2!} &= \frac{8\times 7\times 6\times 5\times 4\times 3\times 2}{2\times 4\times 3\times 2\times 2}\\ &= 2\times 7\times 6\times 5 = 420.\end{aligned}$$

The total number of 8-permutations is $280 + 560 + 420 = 1,260$.

5.5.2 Eight-Permutations of a Multiset

Find the number of 8-permutations of the multiset

$$M = \{\text{A, A, A, B, B, B, C, C, C}\} = \{\text{3A, 3B, 3C}\}.$$

Solution: We observe that M contains 9 elements, but have to limit our permutations to 8 elements. The extra element can either be removed from the 3As, the 3Bs, or the 3Cs, which means that we must sum the permutations contributed by each possibility.
Removing it from the 3As, the number of 8-permutations of $\{2A, 3B, 3C\}$: $\frac{8!}{2!3!3!}$.
Removing it from the 3Bs, the number of 8-permutations of $\{3A, 2B, 3C\}$: $\frac{8!}{3!2!3!}$.
Removing it from the 3Cs, the number of 8-permutations of $\{3A, 3B, 2C\}$: $\frac{8!}{3!3!2!}$.
Each possibility gives the same result, therefore the total number of permutations is $3\times\frac{8!}{3!3!2!}$.

$$\begin{aligned}3\times\frac{8!}{3!3!2!} &= 3\times\frac{8\times 7\times 6\times 5\times 4\times 3\times 2}{3\times 2\times 3\times 2\times 2}\\ &= 3\times 8\times 7\times 5\times 2 = 1,680.\end{aligned}$$

The total number of 8-permutations is 1,680.

Table 5.4 The 24 permutations of M

(2,3,4,5)	(2,3,5,4)	(2,4,3,5)	(2,4,5,3)	(2,5,3,4)	(2,5,4,3)	(3,2,4,5)	(3,2,5,4)
(3,4,2,5)	(3,4,5,2)	(3,5,2,4)	(3,5,4,2)	(4,2,3,5)	(4,2,5,3)	(4,3,2,5)	(4,3,5,2)
(4,5,2,3)	(4,5,3,2)	(5,2,3,4)	(5,2,4,3)	(5,3,2,4)	(5,3,4,2)	(5,4,2,3)	(5,4,3,2)

5.5.3 Number of Permutations

Given a set M={2, 3, 4, 5}, calculate the number of permutations of M, and write down their values.
Solution: The number of permutations of M is 4!, and the numbers range between 2,345 and 5,432.

$$4! = 4 \times 3 \times 2 = 24.$$

Table 5.4 shows the permutations of M.

5.5.4 Number of Five-Card Hands

How many different 5-card hands can be dealt from a deck of 52 cards?
Solution: As the order of the cards is not asked for, we are dealing with a problem of combinations.

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

The number of combinations taking 5 cards from 52 is

$$\binom{52}{5} = \frac{52!}{(52-5)!5!} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2} = \frac{311,875,200}{120} = 2,598,960.$$

5.5.5 Hand Shakes with 100 People

A group of 100 people shake hands with each other. How many handshakes take place?
Solution: As it is not sensitive to order, we require the number of combinations.

$$\binom{100}{2} = \frac{100!}{(100-2)!2!} = \frac{100 \times 99}{2} = 4,950.$$

5.5.6 Permutations of MISSISSIPPI

Find the number of permutations of the word MISSISSIPPI.
Solution: The word MISSISSIPPI is the multiset {M, 4I, 4S, 2P}, where $n = 11, \quad n_1 = 1, \quad n_2 = 4, \quad n_3 = 4, \quad n_4 = 2$. Therefore, the number of permutations of the word is given by

$$\begin{aligned}\frac{n!}{n_1!n_2!n_3!n_4!} &= \frac{11!}{1!4!4!2!} \\ &= \frac{11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2}{4 \times 3 \times 2 \times 4 \times 3 \times 2 \times 2} \\ &= 11 \times 10 \times 9 \times 7 \times 5 \\ &= 34,650.\end{aligned}$$

Chapter 6
Probability

6.1 Introduction

This chapter introduces the subject of probability using examples to create various scenarios, including dependent and independent events, mutually exclusive and inclusive events, and the use of combinations. The chapter concludes with a section on worked examples.

6.2 Definition and Notation

Probability is concerned with quantifying the likelihood of an event occurring. The closed interval of probability is [0, 1], where 0 corresponds to an event never happening, and 1 to an event always happening. This requires dividing the number of ways of securing a successful outcome by the total number of possible outcomes:

$$\text{probability} = \frac{\text{number of ways of securing a successful outcome}}{\text{total number of possible outcomes}}.$$

Thus probability does not predict what will happen, but the likelihood of something happening based upon the available outcomes.

It is assumed in the following examples that coins and dice are unbiased, and that decks of playing cards are randomly shuffled. For example, when a coin is tossed, it will finish in one of two possible states: *heads* or *tails*. The possibility of a coin landing on its edge is so remote that it is ignored. Therefore, the probability of securing a head or tail is $1/2 = 0.5$, and the sum of the two probabilities is 1.

The notation is relatively simple: the probability of an event A occurring is expressed as

$$P(A), \quad \text{or} \quad p(A), \quad \text{or} \quad Pr(A).$$

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_6

Table 6.1 Possible number outcomes of rolling a red and blue dice

1, 1	1, 2	1, 3	1, 4	1, 5	1, 6
2, 1	2, 2	2, 3	2, 4	2, 5	2, 6
3, 1	3, 2	3, 3	3, 4	3, 5	3, 6
4, 1	4, 2	4, 3	4, 4	4, 5	4, 6
5, 1	5, 2	5, 3	5, 4	5, 5	5, 6
6, 1	6, 2	6, 3	6, 4	6, 5	6, 6

Consider a bag containing 10 balls, with 4 red, 3 green, 2 blue and 1 white. If we randomly select a single ball from the bag, the probabilities are:

$$P(\text{red}) = \tfrac{4}{10} = 0.4, \quad P(\text{green}) = \tfrac{3}{10} = 0.3, \quad P(\text{blue}) = \tfrac{2}{10} = 0.2, \quad P(\text{white}) = \tfrac{1}{10} = 0.1.$$

Similarly, the probability of getting a 6 when rolling a dice is $P(6) = \frac{1}{6} \approx 0.167$, and is the same for all other numbers. When a red and blue dice are rolled simultaneously, there are 36 possibilities, and the probability of rolling two 6s is the product of the individual probabilities:

$$P(\text{two 6s}) = \tfrac{1}{6} \times \tfrac{1}{6} = \tfrac{1}{36}.$$

Table 6.1 shows all possible number outcomes. Naturally, all pairs of identical numbers have a probability of $\frac{1}{36}$, however, the other number combinations are repeated twice, and their probabilities are doubled: $\frac{1}{36} + \frac{1}{36} = \frac{1}{18}$.

Therefore, when two dice are rolled together, there are 21 possibilities: 6 pairs of identical numbers, each with a probability of $\frac{1}{36} \approx 2.778\%$, and 15 pairs of different numbers, each with a probability of $\frac{1}{18} \approx 5.556\%$.

We can express this in various ways: The probability of getting a pair of identical numbers is $6 \times \frac{1}{36} = \frac{1}{6} \approx 16.667\%$, and the probability of rolling a pair of different numbers is $15 \times \frac{1}{18} = \frac{15}{18} \approx 83.333\%$. If the probability of rolling a pair of identical numbers is ≈ 0.167, then the probability for all other number combinations must be $1 - 0.167 \approx 0.833$.

We are now in a position to declare formally that the probability of event A is within the interval $[0, 1]$:

$$P(A) \in [0, 1]$$

and the probability of A not happening as

$$P(\overline{A}) = 1 - P(A)$$

where $\overline{A}$ is the complement of the set A.

6.2.1 Independent Events

Given two events A and B that are independent of one another, with individual probabilities $P(A)$ and $P(B)$, their combined probability $P(A \cap B)$ is given by

$$P(A \cap B) = P(A) \times P(B).$$

For example, tossing a coin twice are two independent events. The probability of getting a head or tail on the first or second toss is 0.5. Even if the coin is tossed 100 times and lands repeatedly in a head position, the probability for a head or tail the next time remains 0.5. I would however, be very suspicious if this occurred! Therefore, if we toss a coin twice, the probability of getting two heads is the product of the individual probabilities: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Naturally, this is the same for tails. The other possibilities are heads followed by tails, or tails followed by heads. As both of these probabilities are $\frac{1}{4}$, their combined probability is their sum: $\frac{1}{2}$. We can also say that if the probability for getting both heads is $\frac{1}{4}$, or both tails is $\frac{1}{4}$, then the probability of getting both a head and tail must be $1 - \left(\frac{1}{4} + \frac{1}{4}\right) = \frac{1}{2}$.

Tossing a coin 3 times creates the following $2^3 = 8$ outcomes:

$$\text{HHH}, \quad \text{HHT}, \quad \text{HTH}, \quad \text{HTT}, \quad \text{THH}, \quad \text{THT}, \quad \text{TTH}, \quad \text{TTT}.$$

We can see that the probability of getting all heads or all tails is $\frac{1}{8}$; the probability of getting two heads and one tail $\frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$; and the probability of getting alternating heads or tails is $\frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

6.2.2 Dependent Events

Two *dependent events* happen when the first event influences the probability of the second event occurring. For example, if a bag contains 10 balls with 3 red, 4 green and 3 blue. We pose the question: "What is the probability of selecting a red ball on two successive selections, without returning the first ball to the bag?"

As there are 10 balls, with 3 red, the probability of selecting a red ball on the first selection is $P(\text{red}_1) = \frac{3}{10}$. There are now 9 balls left, and the probability of selecting a second red ball is $P(\text{red}_2) = \frac{2}{9}$. Therefore, the total probability is their product:

$$P(\text{red}_{1,2}) = \frac{3}{10} \times \frac{2}{9} = \frac{6}{90} \approx 0.0667.$$

Similarly, the probability of selecting 2 successive green balls is for the first selection $P(\text{green}_1) = \frac{4}{10}$, and for the second selection $P(\text{green}_2) = \frac{3}{9}$, making a combined probability of

$$P(\text{green}_{1,2}) = \frac{4}{10} \times \frac{3}{9} = \frac{12}{90} \approx 0.133.$$

The possibilities are:

$$\begin{aligned}
P(\text{red}_1, \text{red}_2) &= \tfrac{3}{10} \times \tfrac{2}{9} = \tfrac{6}{90} \approx 0.0667 \\
P(\text{red}_1, \text{blue}_2) &= \tfrac{3}{10} \times \tfrac{3}{9} = \tfrac{9}{90} = 0.1 \\
P(\text{red}_1, \text{green}_2) &= \tfrac{3}{10} \times \tfrac{4}{9} = \tfrac{12}{90} \approx 0.133 \\
P(\text{green}_1, \text{red}_2) &= \tfrac{4}{10} \times \tfrac{3}{9} = \tfrac{12}{90} \approx 0.133 \\
P(\text{green}_1, \text{blue}_2) &= \tfrac{4}{10} \times \tfrac{3}{9} = \tfrac{12}{90} \approx 0.133 \\
P(\text{green}_1, \text{green}_2) &= \tfrac{4}{10} \times \tfrac{3}{9} = \tfrac{12}{90} \approx 0.133 \\
P(\text{blue}_1, \text{red}_2) &= \tfrac{3}{10} \times \tfrac{3}{9} = \tfrac{9}{90} = 0.1 \\
P(\text{blue}_1, \text{green}_2) &= \tfrac{3}{10} \times \tfrac{4}{9} = \tfrac{12}{90} \approx 0.133 \\
P(\text{blue}_1, \text{blue}_2) &= \tfrac{3}{10} \times \tfrac{2}{9} = \tfrac{6}{90} \approx 0.0667.
\end{aligned}$$

Note that the probabilities sum to 1.

Dependent events also occur in card games. For example, the probability of selecting a black card from a deck of playing cards is $\frac{26}{52}$. The probability of selecting a second black card, without returning the first to the deck is $\frac{25}{51}$. Therefore, the probability of selecting two successive black cards is

$$P(\text{2 blacks}) = \tfrac{26}{52} \times \tfrac{25}{51} \approx 0.245 = 24.5\%.$$

6.2.3 *Mutually Exclusive Events*

Two events are *mutually exclusive* when they cannot occur at the same time. Therefore, given two mutually exclusive events A and B, the probability of both happening is zero: $P(A \cup B) = 0$.

The probability of either event occurring is the sum of the probabilities of the individual events:

$$P(A \cap B) = P(A) + P(B).$$

For example, let's calculate the probability of rolling a dice such that the number is a 6 or a prime number. The probability of rolling a six is $\frac{1}{6}$, and the probability of rolling a prime number (2, 3, 5) is $\frac{3}{6}$. Therefore, the probability of rolling a six, and a prime number is

$$P(\text{six or prime}) = \tfrac{1}{6} + \tfrac{3}{6} = \tfrac{4}{6} \approx 0.667.$$

6.2.4 Inclusive Events

Inclusive events occur at the same time and their outcomes overlap. For example, given two events A and B with individual probabilities $P(A)$ and $P(B)$, their combined probability $P(A \cup B)$ is given by

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Returning to the example above with a single dice, let's calculate the probability of rolling an even number or a prime number. The probability of rolling an even number $(2, 4, 6)$ is $\frac{3}{6}$, and the probability of rolling a prime number $(2, 3, 5)$ is also $\frac{3}{6}$. But we can see that both events share the number 2, and the probability of rolling a 2 is $\frac{1}{6}$, therefore, the answer is

$$\frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6} \approx 0.833.$$

Similarly, the probability of rolling an odd number $(1, 3, 5)$ is $\frac{3}{6}$, and the probability of rolling a prime number $(2, 3, 5)$ is also $\frac{3}{6}$. But we can see that both events share the numbers 3 and 5, and the probability of rolling a 3 or 5 is $\frac{2}{6}$, therefore, the answer is

$$\frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6} = \frac{2}{3} \approx 0.667.$$

6.2.5 Probability Using Combinations

Some problems lend themselves to the use of combinations to calculate probabilities. This occurs especially when order is not important. For example, consider the problem of dealing 5 playing cards and calculating the probability of finding 2 queens, 2 kings and an ace. As order is not important combinations can be used.

The number of all possible outcomes is the number of arrangements of 5 cards from a pack of 52: $\binom{52}{5}$.

The number of outcomes of choosing 2 queens from a total of 4 is $\binom{4}{2}$.

The number of outcomes of choosing 2 kings from a total of 4 is $\binom{4}{2}$.

The number of outcomes of choosing 1 ace from a total of 4 is $\binom{4}{1}$.

The probability of finding 2 queens, 2 kings and an ace is

$$\frac{\binom{4}{2} \times \binom{4}{2} \times \binom{4}{1}}{\binom{52}{5}}$$

$$\binom{52}{5} = \frac{52!}{5!47!} = 2,598,960$$

$$\binom{4}{1} = \frac{4!}{1!3!} = 4$$

$$\binom{4}{2} = \frac{4!}{2!2!} = 6$$

$$\frac{\binom{4}{2} \times \binom{4}{2} \times \binom{4}{1}}{\binom{52}{5}} = \frac{6 \times 6 \times 4}{2,598,960} \approx 0.0000554.$$

The probability of finding 2 queens, 2 kings and an ace is ≈0.0000554.

Let's chose another example. Four men and 6 women have been shortlisted for 4 scholarships. If the selection process is random, what is the probability that 2 will be male and 2 will be female?

As the question does not mention order, we can use combinations to calculate the probability.

The number of outcomes of choosing 2 males from a total of 4 is $\binom{4}{2}$.

The number of outcomes of choosing 2 females from a total of 6 is $\binom{6}{2}$.

The number of all possible outcomes of choosing 4 people from a total of 10 is $\binom{10}{4}$.

Therefore,

$$P(\text{2 male and 2 female}) = \frac{\binom{4}{2} \times \binom{6}{2}}{\binom{10}{4}}$$

$$\binom{4}{2} = \frac{4!}{(4-2)!2!}$$

$$= \frac{4 \times 3 \times 2}{2 \times 2} = 6$$

$$\binom{6}{2} = \frac{6!}{(6-2)!2!}$$

$$= \frac{6 \times 5 \times 4 \times 3 \times 2}{4 \times 3 \times 2 \times 2} = 15$$

$$\binom{10}{4} = \frac{10!}{(10-4)!4!}$$

$$= \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2}{6 \times 5 \times 4 \times 3 \times 2 \times 4 \times 3 \times 2} = 210$$

$$P(\text{2 male and 2 female}) = \frac{6 \times 15}{210} \approx 0.429.$$

The probability that 2 will be male and 2 will be female is ≈0.429.

Table 6.2 Possible outcomes and probabilities

Male	Female	Outcomes	Probability
0	4	1	≈ 0.0143
1	3	16	≈ 0.229
2	2	36	≈ 0.514
3	1	16	≈ 0.229
4	0	1	≈ 0.0143

If there had been 4 women instead of 6, we would intuitively expect the probability to be 0.5, but it is not:

$$
\begin{aligned}
P(\text{2 male and 2 female}) &= \frac{\binom{4}{2} \times \binom{4}{2}}{\binom{8}{4}} \\
\binom{4}{2} &= \frac{4!}{(4-2)!2!} \\
&= \frac{4 \times 3 \times 2}{2 \times 2} = 6 \\
\binom{8}{4} &= \frac{8!}{(8-4)!4!} \\
&= \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2}{4 \times 3 \times 2 \times 4 \times 3 \times 2} = 70 \\
P(\text{2 male and 2 female}) &= \frac{6 \times 6}{70} \approx 0.514.
\end{aligned}
$$

This result is due to the small sample and the way probabilities are calculated. Table 6.2 shows how probability is divided for different outcomes.

6.3 Worked Examples

6.3.1 Product of Probabilities

How many outcomes are possible with four independent events: tossing a coin, using a light switch, rolling a dice, and selecting a ball from a bag containing four coloured balls?

Solution: There are 2 outcomes when tossing a coin: $O(\text{coin}) = 2$. There are 2 outcomes when switching a light: $O(\text{switch}) = 2$. There are 6 outcomes when rolling a dice: $O(\text{dice}) = 6$. And there are 4 outcomes when selecting a ball: $O(\text{ball}) = 4$.

Therefore, the total number of outcomes is

$$O(\text{coin}) \times O(\text{switch}) \times O(\text{dice}) \times O(\text{ball}) = 2 \times 2 \times 6 \times 4 = 96.$$

6.3.2 Book Arrangements

I have 3 books on LaTeX, 2 books on C, 1 book on Java, and 2 books on Python. How many ways can these different books be arranged in my bookcase?
Solution: There are 8 books in total; the first book can be one of 8, the second book can be one of 7, etc., therefore the number of arrangements is $8! = 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 = 40,320$.

6.3.3 Winning a Lottery

Calculate the probability of winning a lottery where the successful winner has a six-digit number that has no leading zeros.
Solution: The first digit can be 1 to 9, and the remaining five digits can be 0 to 9, therefore the probability of winning is

$$\frac{1}{900,000} \approx 0.00000111.$$

6.3.4 Rolling Two Dice

Calculate the probability of rolling a 2 and a 3 with two dice.
Solution: The two dice are independent of one another. The probability of rolling a 2 and 3 is

$$P(2 \text{ and } 3) = \tfrac{1}{6} \times \tfrac{1}{6} = \tfrac{1}{36} \approx 0.0278.$$

Similarly, the probability of rolling a 3 and 2 is also $\frac{1}{36}$. Therefore, the total probability is $\frac{1}{36} + \frac{1}{36} = \frac{1}{18} \approx 0.0556$.

6.3.5 Two Dice Sum to 7

Calculate the probability of rolling two dice that sum to 7.
Solution: There are 6 ways the two dice can total 7:

$$1,6, \quad 2,5, \quad 3,4, \quad 4,3, \quad 5,2, \quad 6,1.$$

The probability of any one of the above outcomes is $\frac{1}{36}$, therefore the total probability is $\frac{6}{36} = \frac{1}{6} \approx 0.1667$.

6.3.6 Two Dice Sum to 4

Calculate the probability of rolling two dice that sum to 4.
Solution: There are 3 ways the two dice can total 4:

$$1,3, \quad 2,2, \quad 3,1.$$

The probability of any one of the above outcomes is $\frac{1}{36}$, therefore the total probability is $\frac{3}{36} = \frac{1}{12} \approx 0.0833$.

6.3.7 Dealing a Red Ace

Calculate the probability of dealing a red ace from a deck of playing cards.
Solution: There are 52 cards in a full deck, and 2 red aces, therefore, the probability of dealing a red ace is $\frac{2}{52} = \frac{1}{26} \approx 0.0385$.

6.3.8 Selecting Four Aces in Succession

Calculate the probability of selecting four aces in succession from a deck of playing cards without replacement.
Solution: There are 52 cards in a full deck, and 4 aces. The probability of selecting the first ace is $\frac{4}{52}$. The probability of selecting a second ace is $\frac{3}{51}$. The probability of selecting a third ace is $\frac{2}{50}$. The probability of selecting a forth ace is $\frac{1}{49}$. The combined probability is

$$P(\text{4 aces}) = \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} = \frac{24}{6{,}497{,}400} \approx 0.00037\%$$

6.3.9 Selecting Cards

Calculate the probability of selecting an ace or a king from a deck of playing cards without replacement.

Solution: There are 52 cards in a full deck, with 4 aces and 4 kings. The probability of selecting an ace or a king is $\frac{8}{52} = \frac{2}{13} \approx 0.154$.

6.3.10 Selecting Four Balls from a Bag

A bag contains 14 balls with 5 red, 6 green and 3 blue. Four balls are taken from the bag in one go; calculate the probability of selecting 2 red balls, and 2 non-red balls.
Solution: As order is not important, use combinations to compute the different possibilities.

There are $\binom{5}{2}$ ways of selecting 2 balls from 5 red balls.
There are $\binom{9}{2}$ ways of selecting 2 balls from the 9 non-red balls.
Therefore there are $\binom{5}{2} \cdot \binom{9}{2}$ ways of selecting 2 red balls and 2 non-red balls.
There are $\binom{14}{4}$ ways of selecting 4 balls from 14 balls.
Therefore, the probability of selecting 2 red balls and 2 non-red balls is

$$\frac{\binom{5}{2} \times \binom{9}{2}}{\binom{14}{4}}$$

$$\binom{5}{2} = \frac{5!}{2!3!} = 10$$

$$\binom{9}{2} = \frac{9!}{2!7!} = 36$$

$$\binom{14}{4} = \frac{14!}{4!10!} = 1001$$

$$\frac{\binom{5}{2} \times \binom{9}{2}}{\binom{14}{4}} = \frac{10 \times 36}{1001} \approx 0.36.$$

6.3.11 Forming Teams

Teams of 5 are formed randomly from 8 female and 7 male students. What is the probability that a team contains 2 male and 3 female students?
Solution: As order is not important, use combinations to compute the probability.

There are $\binom{15}{5}$ ways of forming the teams.
There are $\binom{7}{2}$ ways of selecting 2 males from a total of 7.
There are $\binom{8}{3}$ ways of selecting 3 females from a total of 8.
The probability that a team contains 2 male and 3 female students is

$$\frac{\binom{7}{2} \times \binom{8}{3}}{\binom{15}{5}}$$

$$\binom{7}{2} = \frac{7!}{2!5!} = 21$$

$$\binom{8}{3} = \frac{8!}{3!5!} = 56$$

$$\binom{15}{5} = \frac{15!}{5!10!} = 3003$$

$$\frac{\binom{7}{2} \times \binom{8}{3}}{\binom{15}{5}} = \frac{21 \times 56}{3003} \approx 0.3916.$$

6.3.12 Dealing Five Cards

Five cards are taken from a deck of playing cards. What is the probability that the five cards contain 2 clubs and 3 diamonds?

Solution: As order is not important use combinations to compute the probability.

There are $\binom{52}{5}$ ways of selecting 5 cards from 52 cards.

There are $\binom{13}{2}$ ways of selecting 2 cards from 13 club cards.

There are $\binom{13}{3}$ ways of selecting 3 cards from 13 diamond cards.

The probability that the five cards contain 2 clubs and 3 diamonds is

$$\frac{\binom{13}{2} \times \binom{13}{3}}{\binom{52}{5}}$$

$$\binom{52}{5} = \frac{52!}{5!47!} = 2,598,960$$

$$\binom{13}{2} = \frac{13!}{2!11!} = 78$$

$$\binom{13}{3} = \frac{13!}{3!10!} = 286$$

$$\frac{\binom{13}{2} \times \binom{13}{3}}{\binom{52}{5}} = \frac{78 \times 286}{2,598,960} \approx 0.00858.$$

The probability that the five cards contain 2 clubs and 3 diamonds is $\approx$0.858%.

Chapter 7
Modular Arithmetic

7.1 Introduction

This chapter introduces modular arithmetic and its notation. It also shows how modular arithmetic is used in practice with worked examples. The author acknowledges the following references in researching this chapter, Gullberg (1997), https://www.doc.ic.ac.uk/~mrh/330tutor/ch03.html, www.en.wikipedia.org/modular_arithmetic, www.vocalink.com.

7.2 Informal Definition

Modular arithmetic is concerned with integers, especially when one is divided by another called a *modulus* giving a remainder. This results in a cyclic remainder sequence for successive integers. For example, Table 7.1 shows the remainders, 1, 2, 0, 1, 2, 0, 1, 2, 0, 1, when the integers 1 to 10 are divided by the modulus 3, and Table 7.2 shows the remainders, 1, 2, 3, 0, 1, 2, 3, 0, 1, 2, when dividing by 4.

An every-day example of modular arithmetic is the 12-h clock where the modulus is 12. The 24 h become the sequence:

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 0.$$

Table 7.1 The remainder dividing by modulus 3

n	1	2	3	4	5	6	7	8	9	10
$n/3$	0	0	1	1	1	2	2	2	3	3
remainder	1	2	0	1	2	0	1	2	0	1

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_7

Table 7.2 The remainder dividing by modulus 4

n	1	2	3	4	5	6	7	8	9	10
$n/4$	0	0	0	1	1	1	1	2	2	2
remainder	1	2	3	0	1	2	3	0	1	2

7.3 Notation

The notation of modular arithmetic is very simple and takes the form:

$$a \pmod{n} = r, \quad a \in \mathbb{Z}, \quad r \in \mathbb{N}^0, \quad n \in \mathbb{N},$$

where a is some integer, n is the modulus, and r is the remainder. An example being $23 \pmod{5} = 3$.

7.4 Congruence

In mathematics *congruence* means "similar to", and in the context of modular arithmetic, two numbers are congruent when they have the same remainder with a common modulus. Although this is an informal definition, it remains mathematically correct. Formally, congruence is expressed without any mention of remainders as:

Let $a, b \in \mathbb{Z}, n \in \mathbb{N}$, then a and b are congruent modulo n, written $a \equiv b \pmod{n}$, iff (if and only if) $a - b$ is divisible by n, written $n \mid a - b$, and implies that there is an integer k such that $a - b = k \times n$.

There is no mention of remainders, however the definitions are identical as a consequence of the definition: $a \equiv b \pmod{n}$.

For example, given $18 \pmod{4}$ and $14 \pmod{4}$, then $18 - 14 = 1 \times 4$ therefore, $14 \equiv 18 \pmod{4}$. They also both have remainders 2, which is the useful feature of modular arithmetic. Remember that the integers a and b can be positive or negative, but the modulus $n > 0$. Here are some further examples:

$$\begin{aligned} 3 &\equiv 6 \pmod{3} = 0 \\ 23 &\equiv 34 \pmod{11} = 1 \\ 31 &\equiv 11 \pmod{4} = 3 \\ 35 &\equiv 5 \pmod{6} = 5 \\ 121 &\equiv 49 \pmod{8} = 1. \end{aligned}$$

We can also write $a \equiv r \pmod{n}$ which is stating that an integer a is congruent with its remainder r, modulus n. For example, $14 \equiv 4 \pmod{5}$.

Table 7.3 The continuum between −5 and +5

a	−5	−4	−3	−2	−1	0	1	2	3	4	5
$a \pmod 3$	1	2	0	1	2	0	1	2	0	1	2

7.5 Negative Numbers

Although modular arithmetic applies to positive and negative integers, the remainder is always positive, which means that we need to be careful when dealing with negative numbers. The formula associated with modular arithmetic is:

$$a = k \times n + r. \tag{7.1}$$

Using (7.1), for any positive number a and modulus n, there will exist a value of k such that $r < n$. For example, when $a = 11$ and $n = 3$, then $k = 3$ and a remainder $r = 2$.

Now when a is negative, k must also take on a negative value, such that $r < n$, and remains positive. For example, if $a = -11$ and $n = 3$, then $k = -4$:

$$-11 = -4 \times 3 + 1$$

therefore, $-11 \pmod 3 = 1$.

Let's take another example where $a = -7$ and $n = 3$, then $k = -3$:

$$-7 = -3 \times 3 + 2$$

therefore, $-7 \pmod 3 = 2$.

The reason for the above rule is that a continuum must exist between negative and positive numbers. Table 7.3 shows an example where $n = 3$ and a varies between −5 and +5.

One last observation concerning negative numbers; if $a \equiv b \pmod n$ then $-a \equiv -b \pmod n$. To illustrate this given $13 \equiv 18 \pmod 5 = 3$, then $-13 \equiv -18 \pmod 5 = 2$. Observe the change in remainder.

7.6 Arithmetic Operations

In this section we explore some of the relationships between numbers and number pairs when sharing the same modulus.

7.6.1 Sums of Numbers

Let's see what happens to the remainders of five numbers when they are summed together. Let the numbers be $5, 6, 7, 8, 9$ with a sum of 35, and a modulus of 3. Therefore,

$$\begin{aligned} 5 \pmod 3 &= 2 \\ 6 \pmod 3 &= 0 \\ 7 \pmod 3 &= 1 \\ 8 \pmod 3 &= 2 \\ 9 \pmod 3 &= 0 \\ 35 \pmod 3 &= 2. \end{aligned}$$

It is easy to add the individual numbers and find the remainder from the sum, but with much larger numbers, we can sum the individual remainders and subject this to the same modular arithmetic. In the case of the above five numbers, the remainders sum to $2 + 0 + 1 + 2 + 0 = 5$, and $5 \pmod 3 = 2$.

Let's take two relatively large numbers $3,412$ and $5,354$ modulo 17. Therefore,

$$3,412 \pmod{17} = 12 \quad \text{and} \quad 5,354 \pmod{17} = 16$$

the sum of the two numbers is 8,766, and the sum of their two remainders is $12 + 16 = 28$. Therefore, the remainder of 8,766 is $28 \pmod{17} = 11$.

The general formula for this observation is

$$\text{if } a + b = c, \quad a \pmod n + b \pmod n \equiv c \pmod n.$$

Furthermore, if a and b are increased by an integer i, then the following relationship holds:

$$\text{if } a \equiv b \pmod n \text{ then } a + i \equiv b + i \pmod n.$$

For example, if $a = 13$ and $b = 20$, then $13 \equiv 20 \pmod 7$. Therefore, when $i = 5$, then

$$18 \equiv 25 \pmod 7 = 4.$$

Another relationship concerns pairs of congruent numbers.

$$\text{if } a \equiv b \pmod n \quad \text{and} \quad c \equiv d \pmod n \text{ then } a + c \equiv b + d \pmod n.$$

For example, if $a = 21,\ b = 31,\ c = 18,\ d = 33,\ n = 5$ then

$$21 \equiv 31 \pmod 5 = 1$$
$$18 \equiv 33 \pmod 5 = 3$$
$$21 + 18 \equiv 31 + 33 \pmod 5 = 4.$$

7.6.2 Products

When we find the product $a \times b = c$, then $a \pmod n \times b \pmod n \equiv c \pmod n$. For example, $12 \times 14 = 168$, then:

$$12 \pmod 5 = 2$$
$$14 \pmod 5 = 4$$
$$168 \pmod 5 = 3.$$

Therefore, $2 \times 4 \pmod 5 \equiv 3 \pmod 5 = 3$.

Let's try another example: $11 \times 12 = 132$, then:

$$11 \pmod 3 = 2$$
$$12 \pmod 3 = 0$$
$$132 \pmod 3 = 0.$$

Therefore, $2 \times 0 \pmod 3 \equiv 0 \pmod 3 = 0$.

7.6.3 Multiplying by a Constant

Given $a \equiv b \pmod n$, then multiplying a and b by an integer k does not alter the congruence relationship: $k \times a \equiv k \times b \pmod n$. For example,

$$23 \equiv 33 \pmod 5 = 3,$$

therefore multiplying 23 and 33 by 2 gives $46 \equiv 66 \pmod 5 = 1$. Note that the remainder changes from 3 to 1, which is determined by $3 \times 2 \pmod 5 = 1$.

Let's try another example,

$$11 \equiv 17 \pmod 6 = 5,$$

therefore multiplying 11 and 17 by 3 gives $33 \equiv 51 \pmod 6 = 3$. This time the remainder changes from 5 to 3, which is determined by $5 \times 3 \pmod 6 = 3$.

7.6.4 Congruent Pairs

Given $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $a \times c \equiv b \times d \pmod{n}$.

For example,

$$3 \equiv 7 \pmod{4} = 3$$
$$11 \equiv 15 \pmod{4} = 3$$

then

$$3 \times 11 \equiv 7 \times 15 \pmod{4} = 1$$
$$33 \equiv 105 \pmod{4} = 1.$$

Note that the remainder has changed from 3 to 1.

7.6.5 Multiplicative Inverse

In normal arithmetic, we must avoid accidentally dividing by zero. So too, in modular arithmetic, as zeros often arise simply by changing a modulus. The following are not permitted:

$$2/5 \pmod{5}$$
$$3/10 \pmod{5}$$
$$5/15 \pmod{5}$$
$$\text{etc.}$$

because 5, 10, 15, etc, are all congruent with 0 (mod 5), and create a divide by zero condition.

To avoid a potential divide by zero, a *multiplicative inverse* is employed, such that division is replaced by multiplication.

For example, in normal arithmetic, $10/5$ can be replaced by a product as follows:

$$10/5 = 10 \times i$$
$$\text{where } 5 \times i = 1.$$

In this case, the answer is easy: $i = 0.2$, which means that $10/5 = 10 \times 0.2$.

In modular arithmetic, the operation

$$a/b \pmod{n}$$

Table 7.4 Products (mod 6)

×	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

is replaced by

$$a \times i \pmod{n},$$

where i is the multiplicative inverse of b, if one exists, such that

$$b \times i \equiv 1 \pmod{n}.$$

Say we need to compute $4/5 \pmod 6$, then we require the multiplicative inverse of 5 (mod 6). To simplify this search, Table 7.4 shows the products of the numbers 0 to 5 (mod 6), and is referenced (row, column). The table shows that there are 2 entries containing 1: (1, 1), (5, 5), highlighted in red.

The value of i for the divisor 5 (mod 6) is determined as follows:

$$\begin{aligned} 4/5 &\equiv 4 \times i \pmod 6 \\ \text{and } 5 \times i &\equiv 1 \pmod 6. \end{aligned}$$

Table 7.4 shows that the entry (5, 5) contains a 1, therefore $i = 5$, and means that

$$4/5 \equiv 4 \times 5 \pmod 6.$$

Table 7.4 shows that the product $4 \times 5 \equiv 2 \pmod 6$, therefore

$$4/5 \equiv 2 \pmod 6.$$

Next, say we need to compute $3/2 \pmod 6$, then we have

$$\begin{aligned} 3/2 &\equiv 3 \times i \pmod 6 \\ \text{and } 3 \times i &\equiv 1 \pmod 6. \end{aligned}$$

But the product $3 \times i$ in Table 7.4 does not contain an entry with a 1, therefore the operation 3/2 (mod 6) is not permitted.

7.6.6 Modulo a Prime

Table 7.5 shows the products for the numbers 0 to 6 (mod 7), where one notices that every row, apart from zero, contains the numbers 0 to 6 in different order, and only a single 1. This is because 7 is a prime number, and the product of any two numbers, apart from zero, will never introduce any congruent results. This results in a theorem which states that:

When the modulus n is a prime number, then it is possible to divide by any non-zero number. Consequently, for any $b \in \{1, 2, 3, \ldots, n-1\}$, there is only one number $i \in \{1, 2, 3, \ldots, n-1\}$ such that

$$b \times i \equiv 1 \pmod{n} \quad \text{and} \quad \frac{1}{b} \equiv i \pmod{n}.$$

From Table 7.5 we have $(1, 1) = (2, 4) = (3, 5) = (4, 2) = (5, 3) = (6, 6) = 1$, therefore,

$$\frac{1}{1} = 1, \quad \frac{1}{2} = 4, \quad \frac{1}{3} = 5, \quad \frac{1}{4} = 2, \quad \frac{1}{5} = 3, \quad \frac{1}{6} = 6.$$

Generally, we can write

$$b \times i \equiv 1 \pmod{n} \quad \text{and} \quad \frac{a}{b} \equiv a \times i \pmod{n}.$$

Therefore, using modulo 7:

Table 7.5 Products (mod 7)

×	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

Table 7.6 Product pairs that equal 1 with different modulo

mod	1 ×	2 ×	3 ×	4 ×	5 ×	6 ×	7 ×	8 ×	9 ×	10 ×
2	1 × 1									
3	1 × 1	2 × 2								
4	1 × 1		3 × 3							
5	1 × 1	2 × 3	3 × 2	4 × 4						
6	1 × 1				5 × 1					
7	1 × 1	2 × 4	3 × 5	4 × 2	5 × 3	6 × 6				
8	1 × 1		3 × 3		5 × 5		7 × 7			
9	1 × 1	2 × 5		4 × 7	5 × 2		7 × 4	8 × 8		
10	1 × 1		3 × 7				7 × 3		9 × 9	
11	1 × 1	2 × 6	3 × 4	4 × 3	5 × 9	6 × 2	7 × 8	8 × 7	9 × 5	10 × 10

Table 7.7 Products (mod 2)

×	0	1
0	0	0
1	0	1

$$\frac{3}{1} = 3, \quad \frac{3}{2} = 5, \quad \frac{3}{3} = 1, \quad \frac{3}{4} = 6, \quad \frac{3}{5} = 2, \quad \frac{3}{6} = 4.$$

Table 7.6 shows the products that equal 1 for a modulus between 2 and 11. The table shows that when the modulus n is a prime number, i.e. 2, 3, 5, 7 and 11, there are $n - 1$ products. These are highlighted in yellow.

For completeness, I have included the following product tables: Table 7.7 modulo 2, Table 7.8 modulo 3,Table 7.9 modulo 4, Table 7.10 modulo 5, Table 7.11 modulo 9, Table 7.12 modulo 10 and Table 7.13 modulo 11.

7.6.7 Fermat's Little Theorem

The French lawyer and mathematician Pierre de Fermat (1601–1665) wrote to a friend in 1640 that he had discovered the following connection: If p is a prime number, then for any integer a, the number $a^p - a$ is an integer multiple of p. In modular arithmetic this is written:

Table 7.8 Products (mod 3)

×	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Table 7.9 Products (mod 4)

×	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Table 7.10 Products (mod 5)

×	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

$$a^p \equiv a \pmod{p}, \quad a \in \mathbb{Z}, \quad p \text{ is a prime number.}$$

Table 7.14 shows some examples of the theorem.

7.7 Applications of Modular Arithmetic

7.7.1 ISBN Parity Check

When an ISBN (International Standard Book Number) is assigned to a text-based monographic publication, it now always consists of 13 digits. Each ISBN consists of 5 elements with each section being separated by spaces or hyphens. Three of the five elements may be of varying length.

Table 7.11 Products (mod 9)

×	0	1	2	3	4	5	6	7	8
0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8
2	0	2	4	6	8	1	3	5	7
3	0	3	6	0	3	6	0	3	6
4	0	4	8	3	7	2	6	1	5
5	0	5	1	6	2	7	3	8	4
6	0	6	3	0	6	3	0	6	3
7	0	7	5	3	1	8	6	4	2
8	0	8	7	6	5	4	3	2	1

Table 7.12 Products (mod 10)

×	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9
2	0	2	4	6	8	0	2	4	6	8
3	0	3	6	9	2	5	8	1	4	7
4	0	4	8	2	6	0	4	8	2	6
5	0	5	0	5	0	5	0	5	0	5
6	0	6	2	8	4	0	6	2	8	4
7	0	7	4	1	8	5	2	9	6	3
8	0	8	6	4	2	0	8	6	4	2
9	0	9	8	7	6	5	4	3	2	1

For example, the ISBN for the first edition of *Foundation Mathematics for Computer Science* is 978-3-319-21436-8. The first 3 digits can either be 978 or 979; 3 identifies the country, geographical region, or language area; 319 identifies the particular publisher or imprint; 21436 identifies the particular edition and format of a specific title; 8 is a check digit that validates the rest of the number using modulo 10 with alternate weights of 1 and 3. If the remainder isn't zero, it is subtracted from 10, giving the check digit.

Table 7.13 Products (mod 11)

×	0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10
2	0	2	4	6	8	10	1	3	5	7	9
3	0	3	6	9	1	4	7	10	2	5	8
4	0	4	8	1	5	9	2	6	10	3	7
5	0	5	10	4	9	3	8	2	7	1	6
6	0	6	1	7	2	8	3	9	4	10	5
7	0	7	3	10	6	2	9	5	1	8	4
8	0	8	5	2	10	7	4	1	9	6	3
9	0	9	7	5	3	1	10	8	6	4	2
10	0	10	9	8	7	6	5	4	3	2	1

Table 7.14 Examples of Fermat's Little Theorem

a	p	a^p	$a^p - a$	$a^p \pmod p$	$a \pmod p$
2	2	4	2	0	0
2	3	8	6	2	2
2	5	32	30	2	2
2	7	128	126	2	2
3	2	9	6	1	1
3	3	27	24	0	0
3	5	243	240	3	3
3	7	2,187	2,184	3	3
5	2	25	20	1	1
5	3	125	120	2	2
5	5	3,125	3,120	0	0
5	7	78,125	78,120	5	5
7	2	49	42	1	1
7	3	343	336	1	1
7	5	16,807	16,800	2	2
7	7	823,543	823,536	0	0

$$
\begin{aligned}
&[(9 \times 1) + (7 \times 3) + (8 \times 1) + (3 \times 3) + (3 \times 1) + (1 \times 3) \\
&+ (9 \times 1) + (2 \times 3) + (1 \times 1) + (4 \times 3) + (3 \times 1) + (6 \times 3)] \pmod{10} \\
&= 9 + 21 + 8 + 9 + 3 + 3 + 9 + 6 + 1 + 12 + 3 + 18 \pmod{10} \\
&= 102 \pmod{10} \\
&= 2.
\end{aligned}
$$

As 2 isn't zero, it is subtracted from 10 giving the check digit 8, as used in the ISBN.

7.7.2 IBAN Check Digits

IBAN (International Bank Account Number) validation through control digits is used as an effective way of reducing failed transactions when processing international and domestic payments. Three types of modulus checks are performed on UK bank accounts:

Mod 10 - Standard 10 modulus check
Mod 11 - Standard 11 modulus check
DBlAl - Double alternate modulus check.

Table 7.15 shows the notation used to define the specific digits within sort codes and account numbers.

For the standard (10 and 11) modulus check process, each digit of the combined sort code and account number is multiplied by a weight. These values are summed and divided by the specified modulus. If the result has no remainder, the sort code and account number are valid. For example, using the sort code and account number: 000000 58177632, Table 7.16 shows the calculation, where each digit of the IBAN number is multiplied by its weight and summed. This results in:

$$0 + 0 + 0 + 0 + 0 + 0 + 35 + 40 + 8 + 21 + 28 + 36 + 6 + 2 = 176.$$

The IBAN system specifies that the sum 176 is working with a modulus of 11, which gives a remainder of zero, and validates the sort code and account number.

Table 7.15 IBAN Notation

	Sort Code						Account number							
Digit	1	2	3	4	5	6	1	2	3	4	5	6	7	8
Weight	u	v	w	x	y	z	a	b	c	d	e	f	g	h

Table 7.16 IBAN Standard modulus check process

	Sort Code						Account number							
Digit	0	0	0	0	0	0	5	8	1	7	7	6	3	2
Weight	0	0	0	0	0	0	7	5	8	3	4	6	2	1
Weighted digits	0	0	0	0	0	0	35	40	8	21	28	36	6	2

Table 7.17 IBAN Double alternate modulus check process

	Sort Code						Account number							
Digit	4	9	9	2	7	3	1	2	3	4	5	6	7	8
Weight	2	1	2	1	2	1	2	1	2	1	2	1	2	1
Weighted digits	8	9	18	2	14	3	2	2	6	4	10	6	14	8

For the double alternate modulus check process, each digit of the combined sort code and account number is multiplied by an alternating weight of 2 and 1. This time the single digits are summed and divided by 10. If the result has no remainder, the sort code and account number are valid. For example, using the sort code and account number: 499273 12345678, Table 7.17 shows the calculation, where each digit of the IBAN number is multiplied by its weight and summed. This results in:

$$8+9+1+8+2+1+4+3+2+2+6+4+1+0+6+1+4+8=70.$$

The IBAN system specifies that the sum 70 is working with a modulus of 10, which gives a remainder of zero, and validates the sort code and account number.

The weights and modulus are stored in a large table and vary according to the sort code range. For example, for sort codes between 090190 and 090196, the modulus is 10 and the weights are 0, 0, 3, 7, 1, 3, 7, 1, 3, 7, 1, 3, 7, 1. Whereas for sort codes between 090720 and 090726, the modulus is 11 and the weights are 0, 0, 0, 0, 0, 0, 9, 8, 7, 6, 5, 4, 3, 2, 1.

www.vocalink.com contains a complete description of the modulus-based checking system.

7.8 Worked Examples

7.8.1 Negative Numbers

Calculate the remainders of the following negative numbers:
$-20 \pmod 4$, $-13 \pmod 7$, $-6 \pmod 9$, $-23 \pmod 5$.
Solution: Use the the formula $a = k \times n + r$.

$$\begin{aligned} -20 \pmod 4 &= -5 \times 4 + 0 = 0 \\ -13 \pmod 7 &= -2 \times 7 + 1 = 1 \\ -6 \pmod 9 &= -1 \times 9 + 3 = 3 \\ -23 \pmod 5 &= -5 \times 5 + 2 = 2. \end{aligned}$$

7.8.2 Sums of Numbers

Calculate the remainder of the following summations using the individual remainders.

$$23 + 5 + 19 + 41 + 29 + 2 + 19 + 44 \pmod 7$$

$$123 + 345 + 678 + 910 \pmod{11}$$

Solution: Find the remainder of the sum the individual remainders.

$$\begin{aligned} 23 \pmod 7 &= 2 \\ 5 \pmod 7 &= 5 \\ 19 \pmod 7 &= 5 \\ 41 \pmod 7 &= 6 \\ 29 \pmod 7 &= 1 \\ 2 \pmod 7 &= 2 \\ 19 \pmod 7 &= 5 \\ 49 \pmod 7 &= 0 \end{aligned}$$

$$(2 + 5 + 5 + 6 + 1 + 2 + 5) \pmod 7 = 5.$$

$$123 \pmod{11} = 2$$
$$345 \pmod{11} = 4$$
$$678 \pmod{11} = 7$$
$$910 \pmod{11} = 8$$

$$(2 + 4 + 7 + 8) \pmod{11} = 10.$$

7.8.3 Remainders of Products

Calculate the remainders of the following products.

$$123, 456 \times 789, 101 \pmod{11}$$

$$101, 101 \times 202, 202 \pmod{13}$$

Solution: Find the remainder of the product of the individual remainders.

$$123, 456 \pmod{11} = 3$$
$$789, 101 \pmod{11} = 5$$
$$3 \times 5 \pmod{11} = 4.$$

$$101, 102 \pmod{13} = 1$$
$$202, 203 \pmod{13} = 1$$
$$1 \times 1 \pmod{13} = 1.$$

7.8.4 Multiplicative Inverse

Using Table 7.5 find the multiplicative inverse of the following using modulo 7:

$$\frac{1}{4}, \quad \frac{2}{4}, \quad \frac{3}{5}, \quad \frac{5}{3}.$$

Solution: Look for entries in Table 7.5 containing a 1, otherwise there is no multiplicative inverse.

$\frac{1}{4}$: $4 \times 2 \pmod{7} \equiv 1$, therefore $\frac{1}{4} \equiv 1 \times 2 \pmod{7} = 2$
$\frac{2}{4}$: $4 \times 2 \pmod{7} \equiv 1$, therefore $\frac{2}{4} \equiv 2 \times 2 \pmod{7} = 4$

Table 7.18 Products (mod 13)

×	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10	11	12
2	0	2	4	6	8	10	12	1	3	5	7	9	11
3	0	3	6	9	12	2	5	8	11	1	4	7	10
4	0	4	8	12	3	7	11	2	6	10	1	5	9
5	0	5	10	2	7	12	4	9	1	6	11	3	8
6	0	6	12	5	11	4	10	3	9	2	8	1	7
7	0	7	1	8	2	9	3	10	4	11	5	12	6
8	0	8	3	11	6	1	9	4	12	7	2	10	5
9	0	9	5	1	10	6	2	11	7	3	12	8	4
10	0	10	7	4	1	11	8	5	2	12	9	6	3
11	0	11	9	7	5	3	1	12	10	8	6	4	2
12	0	12	11	10	9	8	7	6	5	4	3	2	1

$\frac{3}{5}$: $5 \times 3 \pmod 7 \equiv 1$, therefore $\frac{3}{5} \equiv 3 \times 3 \pmod 7 = 2$
$\frac{5}{3}$: $3 \times 5 \pmod 7 \equiv 1$, therefore $\frac{5}{3} \equiv 5 \times 5 \pmod 7 = 4$.

7.8.5 *Product Table for Modulo 13*

Construct the product table for modulo 13 and show the entries that permit a multiplicative inverse.
Solution: Table 7.18 shows the products for modulo 13 with the entries equal to 1, highlighted in red.

7.8.6 *ISBN Check Digit*

The author's book *Imaginary Mathematics for Computer Science* has the ISBN 978-3-319-94636-8. Using the algorithm described above, confirm the value of the check digit.

Solution:

$$
\begin{aligned}
&[(9 \times 1) + (7 \times 3) + (8 \times 1) + (3 \times 3) + (3 \times 1) + (1 \times 3) \\
&+ (9 \times 1) + (9 \times 3) + (4 \times 1) + (6 \times 3) + (3 \times 1) + (6 \times 3)] \pmod{10} \\
&= 9 + 21 + 8 + 9 + 3 + 3 + 9 + 27 + 4 + 18 + 3 + 18 \pmod{10} \\
&= 132 \pmod{10} \\
&= 2.
\end{aligned}
$$

As 2 isn't zero, it is subtracted from 10 giving the check digit 8, as used in the ISBN.

References

Gullberg J (1997) Mathematics: from the birth of numbers. W W Norton & Co
https://www.doc.ic.ac.uk/~mrh/330tutor/ch03.html
www.en.wikipedia.org/modular_arithmetic
www.vocalink.com

Chapter 8
Trigonometry

8.1 Introduction

This chapter covers some basic features of trigonometry such as angular measure, trigonometric ratios, inverse ratios, trigonometric identities and various rules, with which the reader should be familiar.

8.2 Background

The word "trigonometry" divides into three parts: "tri", "gon", "metry", which means the measurement of three-sided polygons, i.e. triangles. It is an ancient subject and is used across all branches of mathematics.

8.3 Units of Angular Measurement

The measurement of angles is at the heart of trigonometry, and today two units of angular measurement have survived into modern usage: *degrees* and *radians*. The degree (or sexagesimal) unit of measure derives from defining one complete rotation as 360°. Each degree divides into 60 min, and each minute divides into 60 s. The number 60 has survived from Mesopotamian days and is rather incongruous when used alongside today's decimal system—which is why the radian has secured a strong foothold in modern mathematics.

The radian of angular measure does not depend upon any arbitrary constant—it is the angle created by a circular arc whose length is equal to the circle's radius. And because the perimeter of a circle is $2\pi r$, 2π radians correspond to one complete rotation. As 360° correspond to 2π radians, 1 radian equals $180°/\pi$, which is approximately 57.3°. The following relationships between radians and degrees are

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_8

worth remembering:

$$\frac{\pi}{2}\ [\text{rad}] \equiv 90^\circ, \qquad \pi\ [\text{rad}] \equiv 180^\circ$$
$$\frac{3\pi}{2}\ [\text{rad}] \equiv 270^\circ, \quad 2\pi\ [\text{rad}] \equiv 360^\circ.$$

To convert x° to radians:

$$\frac{\pi x^\circ}{180}\ [\text{rad}].$$

To convert x [rad] to degrees:

$$\frac{180x}{\pi}\ [\text{degrees}].$$

8.4 The Trigonometric Ratios

Ancient civilisations knew that triangles—whatever their size—possessed some inherent properties, especially the ratios of sides and their associated angles. This means that if these ratios are known in advance, problems involving triangles with unknown lengths and angles, can be discovered using these ratios.

Figure 8.1 shows a point P with coordinates $(base,\ height)$, on a unit-radius circle rotated through an angle θ. As P is rotated, it moves into the 2nd quadrant, 3rd quadrant, 4th quadrant and returns back to the first quadrant. During the rotation, the sign of $height$ and $base$ change as follows:

1st	quadrant:	$height$ (+),	$base$ (+)
2nd	quadrant:	$height$ (+),	$base$ (−)
3rd	quadrant:	$height$ (−),	$base$ (−)
4th	quadrant:	$height$ (−),	$base$ (+).

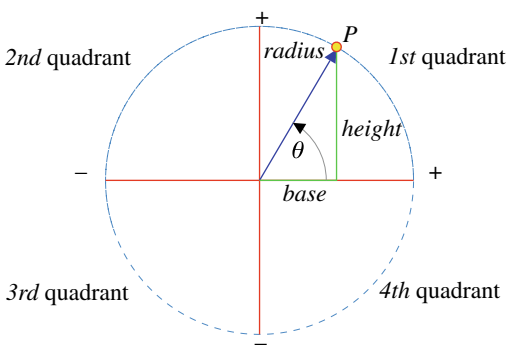

Fig. 8.1 The four quadrants for the trigonometric ratios

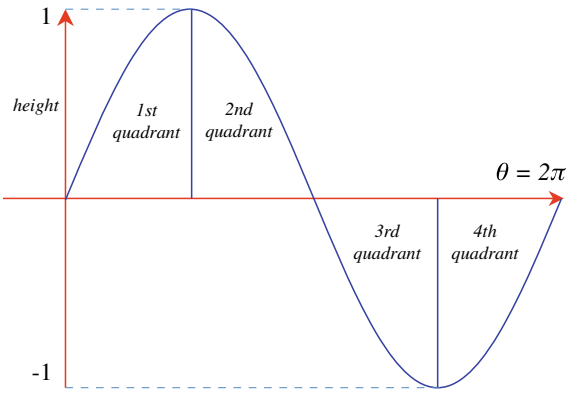

Fig. 8.2 The graph of *height* over the four quadrants

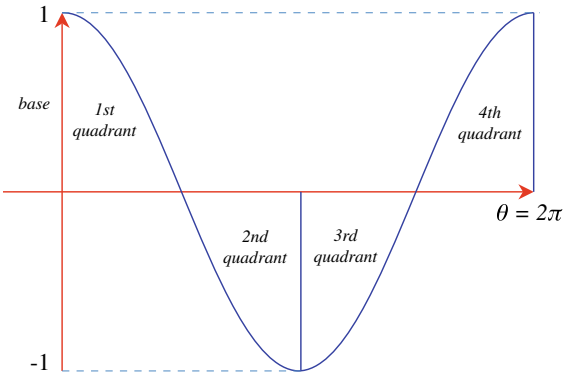

Fig. 8.3 The graph of *base* over the four quadrants

Figures 8.2 and 8.3 plot the changing values of *height* and *base* over the four quadrants, respectively. When $radius = 1$, the curves vary between 1 and -1. In the context of triangles, the sides are labelled as follows:

$$\begin{aligned} hypotenuse &= \text{radius} \\ opposite &= \text{height} \\ adjacent &= \text{base}. \end{aligned}$$

Thus, using the right-angle triangle shown in Fig. 8.4, the trigonometric ratios: sine, cosine and tangent are defined as

$$\sin\theta = \frac{opposite}{hypotenuse}, \quad \cos\theta = \frac{adjacent}{hypotenuse}, \quad \tan\theta = \frac{opposite}{adjacent}.$$

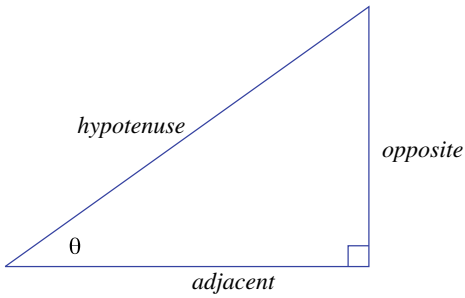

Fig. 8.4 Sides of a right-angle triangle

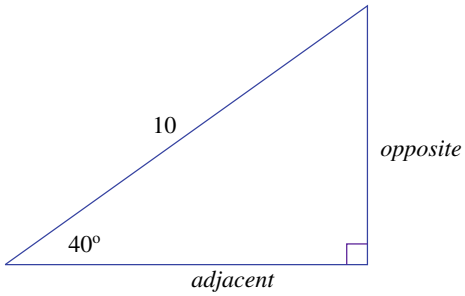

Fig. 8.5 A right-angle triangle with two unknown sides

The reciprocals of these functions, cosecant, secant and cotangent are also useful:

$$\csc\theta = \frac{1}{\sin\theta}, \quad \sec\theta = \frac{1}{\cos\theta}, \quad \cot\theta = \frac{1}{\tan\theta}.$$

As an example, Fig. 8.5 shows a triangle where the hypotenuse and an angle are known. The other sides are calculated as follows:

$$\begin{aligned}
\frac{opposite}{10} &= \sin 40^\circ \\
opposite &= 10\sin 40^\circ \approx 10 \times 0.64278 = 6.4278 \\
\frac{adjacent}{10} &= \cos 40^\circ \\
adjacent &= 10\cos 40^\circ \approx 10 \times 0.7660 = 7.660.
\end{aligned}$$

The theorem of Pythagoras confirms that these lengths are correct:

$$6.4278^2 + 7.660^2 \approx 10^2.$$

Figure 8.6 shows the graph of the tangent function, which, like the sine and cosine functions, is periodic, but with only a period of π radians.

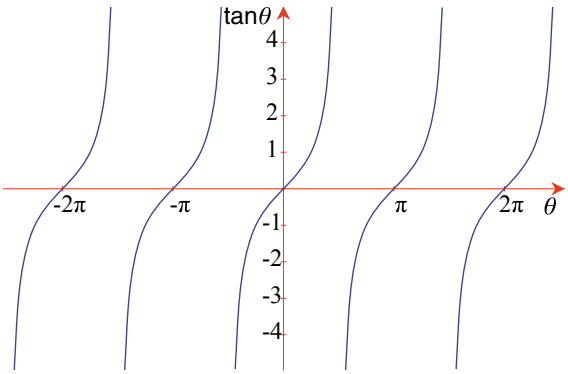

Fig. 8.6 Graph of the tangent function

8.4.1 Domains and Ranges

The periodic nature of $\sin\theta$, $\cos\theta$ and $\tan\theta$, means that their domains are infinitely large. Consequently, it is customary to confine the domain of $\sin\theta$ to

$$\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

and $\cos\theta$ to

$$[0, \pi].$$

The range for both $\sin\theta$ and $\cos\theta$ is

$$[-1, 1].$$

The domain for $\tan\theta$ is the open interval

$$\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$$

and its range is the open interval:

$$]-\infty, \infty[.$$

8.5 Inverse Trigonometric Ratios

The functions $\sin\theta$, $\cos\theta$, $\tan\theta$, $\csc\theta$, $\sec\theta$ and $\cot\theta$ provide different ratios for the angle θ, and the inverse trigonometric functions convert a ratio back into an angle.

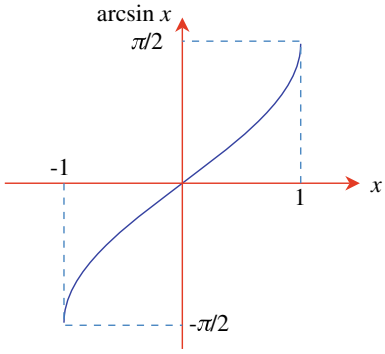

Fig. 8.7 Graph of the arcsin function

These are arcsin, arccos, arctan, arccsc, arcsec and arccot, and are sometimes written as $\sin^{-1}$, $\cos^{-1}$, $\tan^{-1}$, $\csc^{-1}$, $\sec^{-1}$ and $\cot^{-1}$. For example, $\sin 30° = 0.5$, therefore, $\arcsin(0.5) = 30°$. Consequently, the domain for arcsin is the range for sin:

$$[-1, 1]$$

and the range for arcsin is the domain for sin:

$$\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

as shown in Fig. 8.7. Similarly, the domain for arccos is the range for cos:

$$[-1, 1]$$

and the range for arccos is the domain for cos:

$$[0, \pi]$$

as shown in Fig. 8.8. The domain for arctan is the range for tan:

$$]-\infty, \infty[$$

and the range for arctan is the domain for tan:

$$\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$$

as shown in Fig. 8.9.

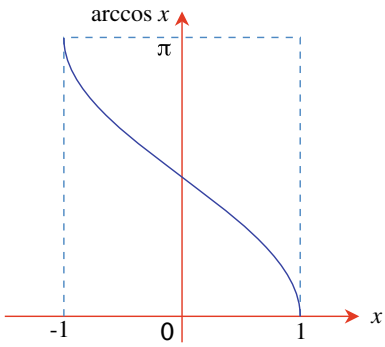

Fig. 8.8 Graph of the arccos function

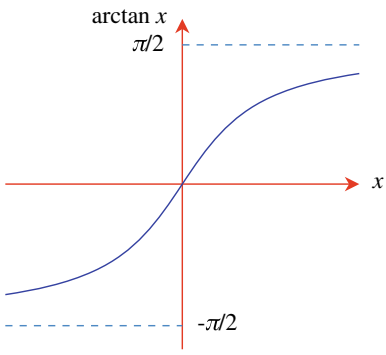

Fig. 8.9 Graph of the arctan function

Various programming languages include the atan2 function, which is an arctan function with two arguments: atan2$(y,\ x)$. The signs of x and y provide sufficient information to locate the quadrant containing the angle, and gives the atan2 function a range of $[0,\ 2\pi]$.

8.6 Trigonometric Identities

The sin and cos curves are identical, apart from being displaced by 90°, and are related by

$$\cos\theta = \sin(\theta + \pi/2).$$

Also, simple algebra and the theorem of Pythagoras can be used to derive other formulae such as

$$\begin{aligned} \frac{\sin\theta}{\cos\theta} &= \tan\theta \\ \sin^2\theta + \cos^2\theta &= 1 \\ 1 + \tan^2\theta &= \sec^2\theta \\ 1 + \cot^2\theta &= \csc^2\theta. \end{aligned}$$

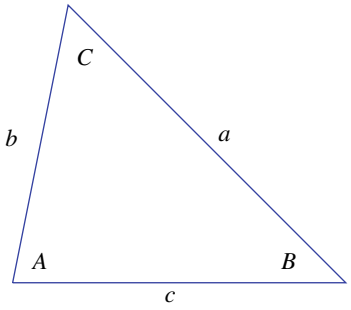

Fig. 8.10 An arbitrary triangle

8.7 The Sine Rule

Figure 8.10 shows a triangle labeled such that side a is opposite angle A, side b is opposite angle B, etc. The sine rule states:

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

which can be used to compute the length of an unknown length or angle. For example, if $A = 60°$, $B = 40°$, $C = 80°$, and $b = 10$, then

$$\frac{a}{\sin 60°} = \frac{10}{\sin 40°}$$

rearranging, we have

$$a = \frac{10 \sin 60°}{\sin 40°} \approx 13.47.$$

Similarly:

$$\frac{c}{\sin 80°} = \frac{10}{\sin 40°}$$

therefore

$$c = \frac{10 \sin 80°}{\sin 40°} \approx 15.32.$$

8.8 The Cosine Rule

The cosine rule expresses the $\sin^2 \theta + \cos^2 \theta = 1$ identity for the arbitrary triangle shown in Fig. 8.10. In fact, there are three versions:

$$a^2 = b^2 + c^2 - 2bc\cos A$$
$$b^2 = c^2 + a^2 - 2ca\cos B$$
$$c^2 = a^2 + b^2 - 2ab\cos C.$$

Three further relationships also hold:

$$a = b\cos C + c\cos B$$
$$b = c\cos A + a\cos C$$
$$c = a\cos B + b\cos A.$$

8.9 Compound-Angle Identities

Trigonometric identities are useful for solving various mathematical problems, but apart from this, their proof often contains a strategy that can be used else where. In the first example, watch out for the technique of multiplying by 1 in the form of a ratio, and swapping denominators. The technique is rather elegant and suggests that the result was known in advance, which probably was the case. Let's begin by finding a way of representing $\sin(\alpha + \beta)$ in terms of $\sin\alpha$, $\cos\alpha$, $\sin\beta$, $\cos\beta$.

With reference to Fig. 8.11:

$$\begin{aligned}\sin(\alpha+\beta) &= \frac{FD}{AD} = \frac{BC + ED}{AD}\\ &= \frac{BC}{AD}\frac{AC}{AC} + \frac{ED}{AD}\frac{CD}{CD}\\ &= \frac{BC}{AC}\frac{AC}{AD} + \frac{ED}{CD}\frac{CD}{AD}\\ \sin(\alpha+\beta) &= \sin\alpha\cos\beta + \cos\alpha\sin\beta. \qquad (8.1)\end{aligned}$$

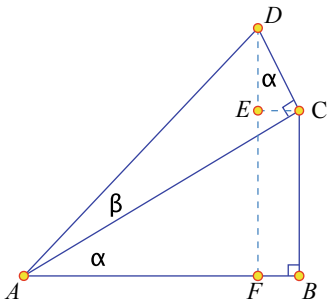

Fig. 8.11 The geometry to expand $\sin(\alpha + \beta)$

To find $\sin(\alpha - \beta)$, reverse the sign of β in (8.1):

$$\sin(\alpha - \beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta. \tag{8.2}$$

Now let's expand $\cos(\alpha + \beta)$ with reference to Fig. 8.11:

$$\begin{aligned}\cos(\alpha+\beta) &= \frac{AE}{AD} = \frac{AB - EC}{AD} \\ &= \frac{AB}{AD}\frac{AC}{AC} - \frac{EC}{AD}\frac{CD}{CD} \\ &= \frac{AB}{AC}\frac{AC}{AD} - \frac{EC}{CD}\frac{CD}{AD} \\ \cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta. \end{aligned} \tag{8.3}$$

To find $\cos(\alpha - \beta)$, reverse the sign of β in (8.3):

$$\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta.$$

To expand $\tan(\alpha + \beta)$, divide (8.1) by (8.3):

$$\begin{aligned}\frac{\sin(\alpha+\beta)}{\cos(\alpha+\beta)} &= \frac{\sin\alpha\cos\beta + \cos\alpha\sin\beta}{\cos\alpha\cos\beta - \sin\alpha\sin\beta} \\ &= \frac{\dfrac{\sin\alpha\cos\beta}{\cos\alpha\cos\beta} + \dfrac{\cos\alpha\sin\beta}{\cos\alpha\cos\beta}}{\dfrac{\cos\alpha\cos\beta}{\cos\alpha\cos\beta} - \dfrac{\sin\alpha\sin\beta}{\cos\alpha\cos\beta}} \\ \tan(\alpha+\beta) &= \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta}. \end{aligned} \tag{8.4}$$

To find $\tan(\alpha - \beta)$, reverse the sign of β in (8.4):

$$\tan(\alpha - \beta) = \frac{\tan\alpha - \tan\beta}{1 + \tan\alpha\tan\beta}.$$

8.9.1 Double-Angle Identities

By making $\beta = \alpha$, the three compound-angle identities

$$\begin{aligned}\sin(\alpha \pm \beta) &= \sin\alpha\cos\beta \pm \cos\alpha\sin\beta \\ \cos(\alpha \pm \beta) &= \cos\alpha\cos\beta \mp \sin\alpha\sin\beta \\ \tan(\alpha \pm \beta) &= \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}\end{aligned}$$

provide the starting point for deriving three corresponding double-angle identities:

$$\begin{aligned}\sin(\alpha \pm \alpha) &= \sin\alpha\cos\alpha \pm \cos\alpha\sin\alpha \\ \sin(2\alpha) &= 2\sin\alpha\cos\alpha.\end{aligned}$$

Similarly,

$$\begin{aligned}\cos(\alpha \pm \alpha) &= \cos\alpha\cos\alpha \mp \sin\alpha\sin\alpha \\ \cos(2\alpha) &= \cos^2\alpha - \sin^2\alpha\end{aligned}$$

which can be further simplified using $\sin^2\alpha + \cos^2\alpha = 1$:

$$\begin{aligned}\cos(2\alpha) &= \cos^2\alpha - \sin^2\alpha \\ \cos(2\alpha) &= 2\cos^2\alpha - 1 \\ \cos(2\alpha) &= 1 - 2\sin^2\alpha.\end{aligned}$$

And for $\tan(2\alpha)$, we have:

$$\begin{aligned}\tan(\alpha + \alpha) &= \frac{\tan\alpha + \tan\alpha}{1 - \tan\alpha\tan\alpha} \\ \tan(2\alpha) &= \frac{2\tan\alpha}{1 - \tan^2\alpha}.\end{aligned}$$

8.9.2 Multiple-Angle Identities

In Chap. 12, Euler's trigonometric formula shows how the following multiple-angle identities are computed:

$$\begin{aligned}\sin(3\alpha) &= 3\sin\alpha - 4\sin^3\alpha \\ \cos(3\alpha) &= 4\cos^3\alpha - 3\cos\alpha \\ \tan(3\alpha) &= \frac{3\tan\alpha - \tan^3\alpha}{1 - 3\tan^2\alpha} \\ \sin(4\alpha) &= 4\sin\alpha\cos\alpha - 8\sin^3\alpha\cos\alpha \\ \cos(4\alpha) &= 8\cos^4\alpha - 8\cos^2\alpha + 1 \\ \tan(4\alpha) &= \frac{4\tan\alpha - 4\tan^3\alpha}{1 - 6\tan^2\alpha + \tan^4\alpha} \\ \sin(5\alpha) &= 16\sin^5\alpha - 20\sin^3\alpha + 5\sin\alpha \\ \cos(5\alpha) &= 16\cos^5\alpha - 20\cos^3\alpha + 5\cos\alpha \\ \tan(5\alpha) &= \frac{5\tan\alpha - 10\tan^3\alpha + \tan^5\alpha}{1 - 10\tan^2\alpha + 5\tan^4\alpha}.\end{aligned}$$

8.9.3 Half-Angle Identities

Every now and then, it is necessary to compute the sine, cosine or tangent of a half-angle from the corresponding whole-angle functions. To do this, we rearrange the double-angle identities as follows.

$$\begin{aligned}\cos(2\alpha) &= 1 - 2\sin^2\alpha \\ \sin^2\alpha &= \frac{1-\cos(2\alpha)}{2} \\ \sin^2(\alpha/2) &= \frac{1-\cos\alpha}{2} \\ \sin(\alpha/2) &= \pm\sqrt{\frac{1-\cos\alpha}{2}}.\end{aligned} \tag{8.5}$$

Similarly,

$$\begin{aligned}\cos^2\alpha &= \frac{1+\cos(2\alpha)}{2} \\ \cos^2(\alpha/2) &= \frac{1+\cos\alpha}{2} \\ \cos(\alpha/2) &= \pm\sqrt{\frac{1+\cos\alpha}{2}}.\end{aligned} \tag{8.6}$$

Dividing (8.5) by (8.6) we have

$$\tan(\alpha/2) = \sqrt{\frac{1-\cos\alpha}{1+\cos\alpha}}.$$

8.10 Perimeter Relationships

Finally, with reference to Fig. 8.10, we come to the relationships that integrate angles with the perimeter of a triangle:

$$\begin{aligned}s &= \tfrac{1}{2}(a+b+c) \\ \sin(A/2) &= \sqrt{\frac{(s-b)(s-c)}{bc}} \\ \sin(B/2) &= \sqrt{\frac{(s-c)(s-a)}{ca}} \\ \sin(C/2) &= \sqrt{\frac{(s-a)(s-b)}{ab}}\end{aligned}$$

$$\cos(A/2) = \sqrt{\frac{s(s-a)}{bc}}$$
$$\cos(B/2) = \sqrt{\frac{s(s-b)}{ca}}$$
$$\cos(C/2) = \sqrt{\frac{s(s-c)}{ab}}$$
$$\sin A = \frac{2}{bc}\sqrt{s(s-a)(s-b)(s-c)}$$
$$\sin B = \frac{2}{ca}\sqrt{s(s-a)(s-b)(s-c)}$$
$$\sin C = \frac{2}{ab}\sqrt{s(s-a)(s-b)(s-c)}.$$

Chapter 9
Coordinate Systems

9.1 Introduction

In this chapter we revise Cartesian coordinates, axial systems, the distance between two points in space, and the area of simple 2D shapes. It also covers polar, spherical polar and cylindrical coordinate systems.

9.2 Background

René Descartes is often credited with the invention of the xy-plane, but Pierre de Fermat was probably the first inventor. In 1636 Fermat was working on a treatise titled *Ad locus planos et solidos isagoge*, which outlined what we now call "analytic geometry". Unfortunately, Fermat never published his treatise, although he shared his ideas with other mathematicians such as Blaise Pascal (1623–1662). At the same time, Descartes devised his own system of analytic geometry and in 1637 published his results in the prestigious journal *Géométrie*. In the eyes of the scientific world, the publication date of a technical paper determines when a new idea or invention is released into the public domain. Consequently, ever since this publication Descartes has been associated with the xy-plane, which is why it is called the *Cartesian plane*.

The Cartesian plane is such a simple idea that it is strange that it took so long to be discovered. However, although it is true that René Descartes showed how an orthogonal coordinate system could be used for graphs and coordinate geometry, coordinates had been used by ancient Egyptians, almost 2000 years earlier! If Fermat had been more efficient in publishing his research results, the xy-plane could have been called the Fermatian plane! (Merzbach and Boyer 2011).

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_9

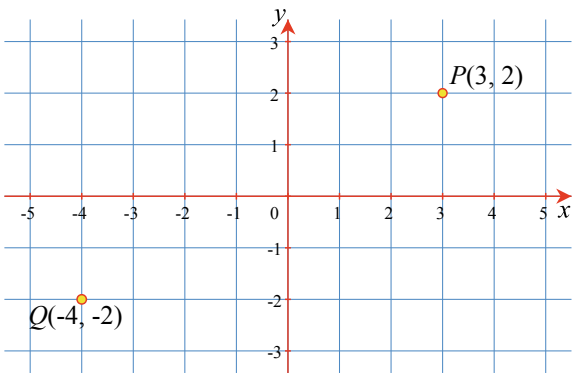

Fig. 9.1 The Cartesian plane

9.3 The Cartesian Plane

The Cartesian plane provides a mechanism for locating points with a unique, ordered pair of numbers (x, y) as shown in Fig. 9.1, where P has coordinates $(3, 2)$ and Q has coordinates $(-4, -2)$. The point $(0, 0)$ is called the *origin*. As previously mentioned, Descartes suggested that the letters x and y should be used to represent variables, and letters at the other end of the alphabet should stand for numbers. Which is why equations such as $y = ax^2 + bx + c$, are written this way.

The axes are said to be *oriented* as the x-axis rotates anticlockwise towards the y-axis. They could have been oriented in the opposite sense, with the y-axis rotating anticlockwise towards the x-axis.

9.4 Function Graphs

When functions such as

$$\begin{aligned} \text{linear:} \quad & y = mx + c, \\ \text{quadratic:} \quad & y = ax^2 + bx + c, \\ \text{cubic:} \quad & y = ax^3 + bx^2 + cx + d, \\ \text{trigonometric:} \quad & y = a \sin x, \end{aligned}$$

are drawn as graphs, they create familiar shapes that permit the function to be easily identified. Linear functions are straight lines; quadratics are parabolas; cubics have an "S" shape; and trigonometric functions often possess a wave-like trace. Figure 9.2 shows examples of each type of function.

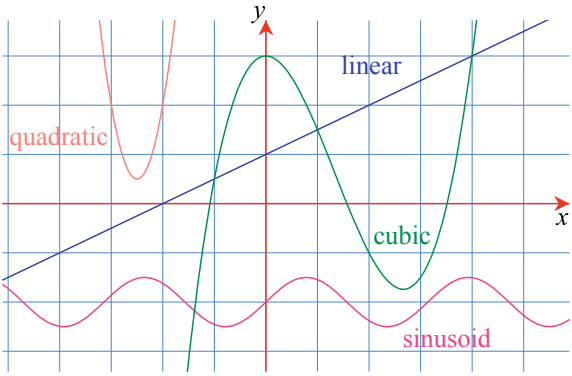

Fig. 9.2 Graphs of four function types

9.5 Shape Representation

The Cartesian plane also provides a way to represent 2D shapes numerically, which permits them to be manipulated mathematically. Let's begin with 2D polygons and show how their internal area can be calculated.

9.5.1 2D Polygons

A polygon is formed from a chain of *vertices* (points) as shown in Fig. 9.3. A straight line is assumed to connect each pair of neighbouring vertices; intermediate points on the line are not explicitly stored. There is no convention for starting a chain of vertices, but software will often dictate whether polygons have a clockwise or anticlockwise vertex sequence.

We can now subject this list of coordinates to a variety of arithmetic and mathematical operations. For example, if we double the values of x and y and redraw the vertices, we discover that the shape's geometric integrity is preserved, but its size

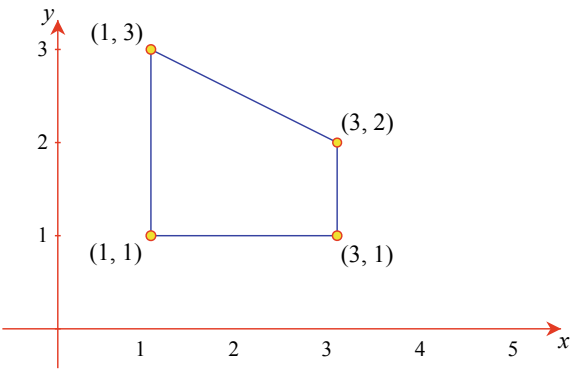

Fig. 9.3 A simple polygon created by a chain of vertices

is doubled relative to the origin. Similarly, if we divide the values of x and y by 2, the shape is still preserved, but its size is halved relative to the origin. On the other hand, if we add 1 to every x-coordinate, and 2 to every y-coordinate, and redraw the vertices, the shape's size remains the same but is displaced 1 unit horizontally and 2 units vertically.

9.5.2 Areas of Shapes

The area of a polygonal shape is readily calculated from its list of coordinates. For example, using the list of coordinates shown in Table 9.1: the area is computed by

$$area = \tfrac{1}{2}[(x_0y_1 - x_1y_0) + (x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + (x_3y_0 - x_0y_3)].$$

You will observe that the calculation sums the results of multiplying an x by the next y, minus the next x by the previous y. When the last vertex is selected, it is paired with the first vertex to complete the process. The result is then halved to reveal the area. As a simple test, let's apply this formula to the shape described in Fig. 9.3:

$$area = \tfrac{1}{2}[(1 \times 1 - 3 \times 1) + (3 \times 2 - 3 \times 1) + (3 \times 3 - 1 \times 2) + (1 \times 1 - 1 \times 3)]$$
$$area = \tfrac{1}{2}[-2 + 3 + 7 - 2] = 3.$$

which, by inspection, is the true area. The beauty of this technique is that it works with any number of vertices and any arbitrary shape.

Another feature of the technique is that if the set of coordinates is clockwise, the area is negative, which means that the calculation computes vertex orientation as well as area. To illustrate this feature, the original vertices are reversed to a clockwise sequence as follows:

$$area = \tfrac{1}{2}[(1 \times 3 - 1 \times 1) + (1 \times 2 - 3 \times 3) + (3 \times 1 - 3 \times 2) + (3 \times 1 - 1 \times 1)]$$
$$area = \tfrac{1}{2}[2 - 7 - 3 + 2] = -3.$$

The minus sign confirms that the vertices are in a clockwise sequence.

Table 9.1 A polygon's coordinates

x	y
x_0	y_0
x_1	y_1
x_2	y_2
x_3	y_3

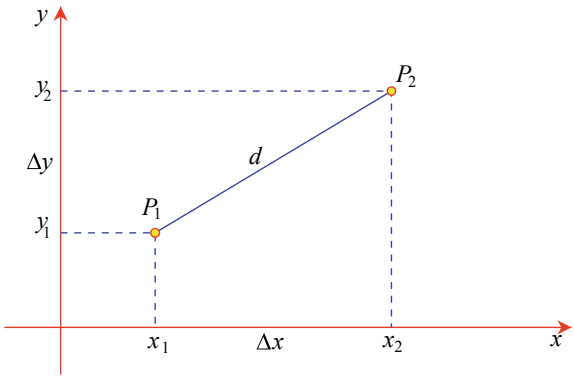

Fig. 9.4 Calculating the distance between two points

9.6 Theorem of Pythagoras in 2D

The theorem of Pythagoras is used to calculate the distance between two points. Figure 9.4 shows two arbitrary points $P_1(x_1,\ y_1)$ and $P_2(x_2,\ y_2)$. The distance $\Delta x = x_2 - x_1$ and $\Delta y = y_2 - y_1$. Therefore, the distance d between P_1 and P_2 is given by

$$d = \sqrt{(\Delta x)^2 + (\Delta y)^2}.$$

For example, given $P_1(1, 1)$, $P_2(4, 5)$, then $d = \sqrt{3^2 + 4^2} = 5$.

9.7 3D Cartesian Coordinates

Two coordinates are required to locate a point on the 2D Cartesian plane, and three coordinates are required for 3D space. The corresponding axial system requires three mutually perpendicular axes; however, there are two ways to add the extra z-axis. Figure 9.5 shows the two orientations, which are described as *left*- and *right-handed* axial systems. The left-handed system permits us to align our left hand with the axes such that the thumb aligns with the x-axis, the first finger aligns with the y-axis, and the middle finger aligns with the z-axis. The right-handed system permits the same system of alignment, but using our right hand. The choice between these axial systems is arbitrary, but one should be aware of the system employed by commercial computer graphics packages. The main problem arises when projecting 3D points onto a 2D plane, which has an oriented axial system. A right-handed system is employed throughout this book, as shown in Fig. 9.6, which also shows a point P with its coordinates. It also worth noting that handedness has no meaning in spaces with 4 dimensions or more.

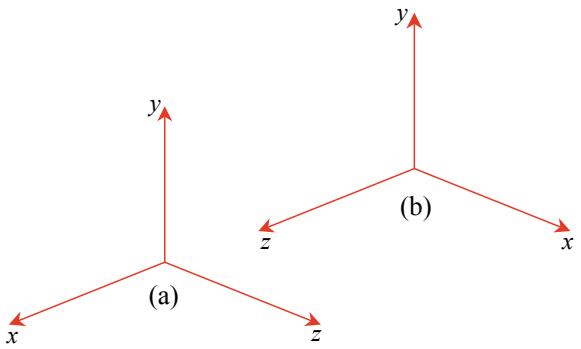

Fig. 9.5 **a** A left-handed system. **b** A right-handed system

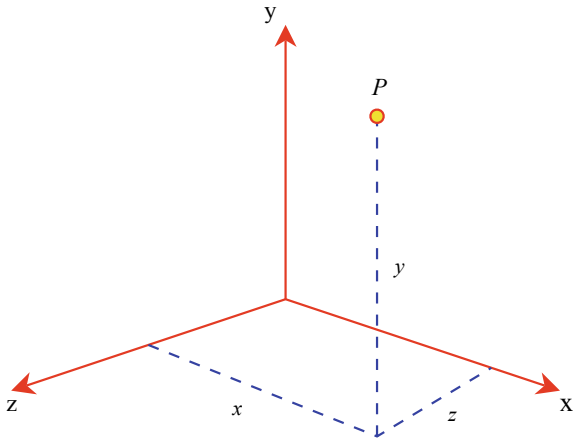

Fig. 9.6 A right-handed axial system showing the coordinates of a point P

9.7.1 Theorem of Pythagoras in 3D

The theorem of Pythagoras in 3D is a natural extension of the 2D rule. In fact, it even works in higher dimensions. Given two arbitrary points $P_1(x_1,\ y_1,\ z_1)$ and $P_2(x_2,\ y_2,\ z_2)$, we compute $\Delta x = x_2 - x_1$, $\Delta y = y_2 - y_1$ and $\Delta z = z_2 - z_1$, from which the distance d between P_1 and P_2 is given by

$$d = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$$

and the distance from the origin to a point $P(x, y, z)$ is simply

$$d = \sqrt{x^2 + y^2 + z^2}.$$

Therefore, the point $(3, 4, 5)$ is $\sqrt{3^2 + 4^2 + 5^2} \approx 7.07$ from the origin.

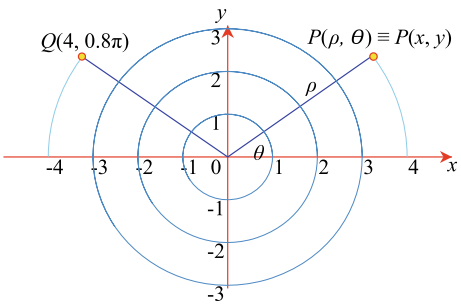

Fig. 9.7 2D polar coordinates

9.8 Polar Coordinates

Polar coordinates are used for handling data containing angles, rather than linear offsets. Figure 9.7 shows the convention used for 2D polar coordinates, where the point $P(x,\ y)$ has equivalent polar coordinates $P(\rho,\ \theta)$, where:

$$
\begin{aligned}
x &= \rho \cos\theta \\
y &= \rho \sin\theta \\
\rho &= \sqrt{x^2 + y^2} \\
\theta &= \arctan(y/x).
\end{aligned}
$$

For example, the point $Q(4,\ 0.8\pi)$ in Fig. 9.7 has Cartesian coordinates:

$$
\begin{aligned}
x &= 4\cos(0.8\pi) \approx -3.24 \\
y &= 4\sin(0.8\pi) \approx 2.35
\end{aligned}
$$

and the point $(3, 4)$ has polar coordinates:

$$
\begin{aligned}
\rho &= \sqrt{3^2 + 4^2} = 5 \\
\theta &= \arctan(4/3) \approx 53.13°.
\end{aligned}
$$

These conversion formulae work only for the first quadrant. The atan2 function should be used in a software environment, as it works with all four quadrants.

9.9 Spherical Polar Coordinates

Figure 9.8 shows one convention used for spherical polar coordinates, where the point $P(x,\ y,\ z)$ has equivalent polar coordinates $P(\rho,\ \phi,\ \theta)$, where:

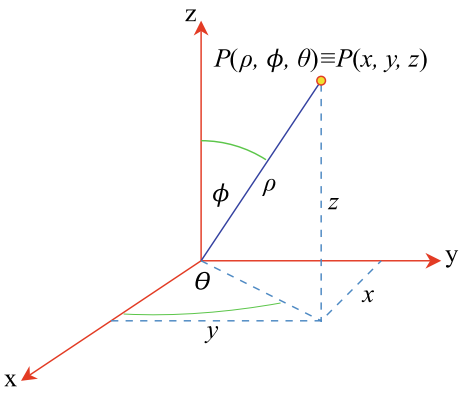

Fig. 9.8 Spherical polar coordinates

$$
\begin{aligned}
x &= \rho \sin\phi \cos\theta \\
y &= \rho \sin\phi \sin\theta \\
z &= \rho \cos\phi \\
\rho &= \sqrt{x^2 + y^2 + z^2} \\
\phi &= \arccos(z/\rho) \\
\theta &= \arctan(y/x).
\end{aligned}
$$

For example, the point $(3,\ 4,\ 0)$ has spherical polar coordinates $(5,\ 90°,\ 53.13°)$:

$$
\begin{aligned}
\rho &= \sqrt{3^2 + 4^2 + 0^2} = 5 \\
\phi &= \arccos(0/5) = 90° \\
\theta &= \arctan(4/3) \approx 53.13°.
\end{aligned}
$$

Take great care when using spherical coordinates, as authors often swap ϕ with θ, as well as the alignment of the Cartesian axes; not to mention using a left-handed axial system in preference to a right-handed system!

9.10 Cylindrical Coordinates

Figure 9.9 shows one convention used for cylindrical coordinates, where the point $P(x,\ y,\ z)$ has equivalent cylindrical coordinates $P(\rho,\ \theta,\ z)$, where

$$
\begin{aligned}
x &= \rho \cos\theta \\
y &= \rho \sin\theta \\
z &= z \\
\rho &= \sqrt{x^2 + y^2} \\
\theta &= \arctan(y/x).
\end{aligned}
$$

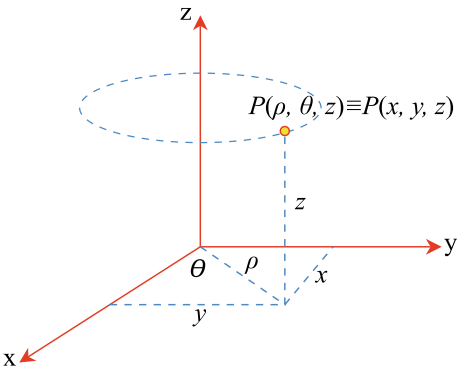

Fig. 9.9 Cylindrical coordinates

For example, the point $(3,\ 4,\ 6)$ has cylindrical coordinates $(5,\ 53.13°,\ 6)$:

$$\begin{aligned}\rho &= \sqrt{3^2 + 4^2} = 5\\ \theta &= \arctan(4/3) \approx 53.13°\\ z &= 6.\end{aligned}$$

Again, be careful when using cylindrical coordinates to ensure compatibility.

9.11 Barycentric Coordinates

Barycentric coordinates locate a point in space relative to existing points, rather than to an origin, and are also known as *local coordinates*. The German mathematician August Möbius (1790–1868) is credited with their invention.

Given two points in 2D space $A(x_a, y_a)$ and $B(x_b, y_b)$, an intermediate point $P(x_p, y_p)$ has barycentric coordinates:

$$\begin{aligned}x_p &= sx_a + tx_b\\ y_p &= sy_a + ty_b\\ 1 &= s + t.\end{aligned}$$

For example, when $s = t = 0.5$, P is halfway between A and B. In 3D, we simply add a z-coordinate.

Given three points in 3D space: $A(x_a, y_a, z_a)$, $B(x_b, y_b, z_b)$ and $C(x_c, y_c, z_c)$, then any point $P(x_p, y_p, z_p)$ inside the triangle ABC has barycentric coordinates:

$$
\begin{aligned}
x_p &= rx_a + sx_b + tx_c \\
y_p &= ry_a + sy_b + ty_c \\
z_p &= rz_a + sz_b + tz_c \\
1 &= r + s + t.
\end{aligned}
$$

For example, when $r = 1,\ s = t = 0$, then $P = A$. Similarly, when $s = 1,\ r = t = 0$, then $P = B$. When $r = s = t = 1/3$, then P is located at the triangle's centroid.

9.12 Homogeneous Coordinates

Homogeneous coordinates surfaced in the early 19th century where they were independently proposed by Möbius, Feuerbach, Bobillier, and Plücker. Homogeneous coordinates, define a point in a plane using three coordinates instead of two. This means that for a point $(x,\ y)$ there exists a homogeneous point $(xt,\ yt,\ t)$ where t is an arbitrary number. For example, the point $(3, 4)$ has homogeneous coordinates $(6,\ 8,\ 2)$, because $3 = 6/2$ and $4 = 8/2$. But the homogeneous point $(6,\ 8,\ 2)$ is not unique to $(3,\ 4)$; $(12,\ 16,\ 4)$, $(15,\ 20,\ 5)$ and $(300,\ 400,\ 100)$ are all possible homogeneous coordinates for $(3,\ 4)$.

The reason why this coordinate system is called "homogeneous" is because it is possible to transform functions such as $f(x,\ y)$ into the form $f(x/t,\ y/t)$ without disturbing the degree of the curve. To the non-mathematician this may not seem anything to get excited about, but in the field of projective geometry it is a very powerful concept.

In 3D, a point $(x,\ y,\ z)$ becomes $(xt,\ yt,\ zt,\ t)$ and for may applications $t = 1$, which seems a futile operation, but in matrix theory it is very useful, as we will discover.

9.13 Worked Examples

9.13.1 Area of a Shape

Compute the area and orientation of the shape defined by the coordinates in Table 9.2.
Solution:

$$
\begin{aligned}
area &= \tfrac{1}{2}[(2 \times 2 - 0 \times 2) + (2 \times 2 - 2 \times 1) + (1 \times 1 - 2 \times 1) + (1 \times 1 - 1 \times 0)] \\
&= \tfrac{1}{2}(4 + 2 - 1 + 1) \\
&= 3.
\end{aligned}
$$

The shape is oriented anticlockwise, as the area is positive.

Table 9.2 Coordinates of the shape

x	0	2	2	1	1	0
y	0	0	2	2	1	1

9.13.2 Distance Between Two Points

Find the distance d_{12} between $P_1(1, 1)$ and $P_2(6, 7)$, and d_{34} between $P_3(1, 1, 1)$ and $P_4(7, 8, 9)$.
Solution:

$$d_{12} = \sqrt{(6-1)^2 + (7-1)^2} = \sqrt{61} \approx 7.81$$
$$d_{34} = \sqrt{(7-1)^2 + (8-1)^2 + (9-1)^2} = \sqrt{149} \approx 12.21.$$

9.13.3 Polar Coordinates

Convert the 2D polar coordinates $(3, \pi/2)$ to Cartesian form, and the point $(4, 5)$ to polar form.
Solution:

$$\rho = 3$$
$$\theta = \pi/2$$
$$x = \rho \cos\theta = 3\cos(\pi/2) = 0$$
$$y = \rho \sin\theta = 3\sin(\pi/2) = 3$$

therefore, $(3,\ \pi/2) \equiv (0, 3)$.

$$x = 4$$
$$y = 5$$
$$\rho = \sqrt{x^2 + y^2} = \sqrt{4^2 + 5^2} \approx 6.4$$
$$\theta = \arctan(y/x) = \arctan(5/4) \approx 51.34^\circ$$

therefore, $(4, 5) \approx (6.4,\ 51.34^\circ)$.

9.13.4 Spherical Polar Coordinates

Convert the spherical polar coordinates $(10,\ \pi/2,\ 45°)$ to Cartesian form, and the point $(3, 4, 5)$ to spherical form.
Solution:

$$\begin{aligned}
\rho &= 10\\
\phi &= \pi/2\\
\theta &= 45°\\
x &= \rho\sin\phi\cos\theta = 10\sin(\pi/2)\cos 45° = 10\sqrt{2}/2 \approx 7.07\\
y &= \rho\sin\phi\sin\theta = 10\sin(\pi/2)\sin 45° = 10\sqrt{2}/2 \approx 7.07\\
z &= \rho\cos\phi = 10\cos(\pi/2) = 0
\end{aligned}$$

therefore, $(10,\ \pi/2,\ 45°) \approx (7.07, 7.07, 0)$.

$$\begin{aligned}
x &= 3\\
y &= 4\\
z &= 5\\
\rho &= \sqrt{x^2+y^2+z^2} = \sqrt{3^2+4^2+5^2} \approx 7.07\\
\phi &= \arccos(z/\rho) \approx \arccos(5/7.07) = 45°\\
\theta &= \arctan(y/x) = \arctan(4/3) \approx 53.13°
\end{aligned}$$

therefore, $(3, 4, 5) \approx (7.07,\ 45°,\ 53.13°)$.

9.13.5 Cylindrical Coordinates

Convert the 3D cylindrical coordinates $(10, \pi/2, 5)$ to Cartesian form, and the point $(3, 4, 5)$ to cylindrical form.
Solution:

$$\begin{aligned}
\rho &= 10\\
\theta &= \pi/2\\
z &= 5\\
x &= \rho\cos\theta = 10\cos(\pi/2) = 0\\
y &= \rho\sin\theta = 10\sin(\pi/2) = 10\\
z &= 5
\end{aligned}$$

therefore, $(10, \pi/2, 5) \equiv (0, 10, 5)$.

Table 9.3 Barycentric coordinates

Point	r	s	t
A	1	0	0
B	0	1	0
C	0	0	1
P_{ab}	0.5	0.5	0
P_{bc}	0	0.5	0.5
P_{ca}	0.5	0	0.5
P_c	1/3	1/3	1/3

9.13.6 Barycentric Coordinates

Given $A(x_a, y_a, z_a)$, $B(x_b, y_b, z_b)$ and $C(x_c, y_c, z_c)$, state the barycentric coordinates for A, B, C the points P_{ab}, P_{bc} and P_{ca} mid-way between AB, BC and CA respectively, and the centroid P_c.
Solution: Table 9.3 shows the barycentric coordinates for ABC.

Reference

Merzbach UC, Boyer CB (2011) A history of mathematics. ISBN: 978-0470525487

Chapter 10
Determinants

10.1 Introduction

This chapter introduces the determinant as a mathematical construct that simplifies the solution of groups of simultaneous equations. The chapter begins by tracing the determinant's historical development, and the reader is shown how to evaluate the determinant's magnitude for real and complex values.

10.2 Background

When patterns of numbers or symbols occur over and over again, mathematicians often devise a way to simplify their description and assign a name to them. For example,

$$\prod_{i=1}^{4} p_i^{\alpha_i}$$

is shorthand for

$$p_1^{\alpha_1} p_2^{\alpha_2} p_3^{\alpha_3} p_4^{\alpha_4}$$

and

$$\sum_{i=1}^{4} p_i^{\alpha_i}$$

is shorthand for

$$p_1^{\alpha_1} + p_2^{\alpha_2} + p_3^{\alpha_3} + p_4^{\alpha_4}.$$

A *determinant* is another example of this process, and is a value derived from a square matrix of terms, often associated with sets of equations. Such problems were

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_10

studied by the Babylonians around 300 BC and by the Chinese, between 200 BC and 100 BC. Since then many mathematicians have been associated with the evolution of determinants and matrices, including Girolamo Cardano (1501–1576), Jan de Witt (1625–1672), Takakazu Seki (1642–1708), Gottfried von Leibniz, Guillaume de L'Hôpital (1661–1704), Augustin-Louis Cauchy (1789–1857), Pierre Laplace (1749–1827) and Arthur Cayley (1821–1895). To understand the rules used to compute a determinant's value, we need to understand their origin, which is in the solution of sets of linear equations.

10.3 Linear Equations with Two Variables

Consider the following linear equations where we want to find values of x and y that satisfy both equations:

$$7 = 3x + 2y \tag{10.1}$$
$$10 = 2x + 4y. \tag{10.2}$$

A standard way to resolve this problem is to multiply (10.1) by 2 and subtract (10.2) from (10.1), which removes the y-terms:

$$\begin{aligned} 14 &= 6x + 2y \\ 10 &= 2x + 4y \\ 4 &= 4x \\ x &= 1. \end{aligned}$$

Substituting $x = 1$ in (10.1) reveals the value of y:

$$\begin{aligned} 7 &= 3 + 2y \\ 4 &= 2y \\ y &= 2. \end{aligned}$$

Therefore, $x = 1$ and $y = 2$, solves (10.1) and (10.2).

The equations must be linearly independent, otherwise we only have one equation. For example, starting with

$$\begin{aligned} 7 &= 3x + 2y \\ 14 &= 6x + 4y \end{aligned}$$

is a futile exercise, as the second equation is double the first, and does not provide any extra information.

To find a general solution to this problem, we start with

$$d_1 = a_1x + b_1y \tag{10.3}$$
$$d_2 = a_2x + b_2y. \tag{10.4}$$

Multiply (10.3) by b_2 and (10.4) by b_1:

$$d_1b_2 = a_1b_2x + b_1b_2y \tag{10.5}$$
$$b_1d_2 = b_1a_2x + b_1b_2y. \tag{10.6}$$

Subtract (10.6) from (10.5):

$$\begin{aligned} d_1b_2 - b_1d_2 &= a_1b_2x - b_1a_2x \\ &= (a_1b_2 - b_1a_2)x \\ x &= \frac{d_1b_2 - b_1d_2}{a_1b_2 - b_1a_2}. \end{aligned} \tag{10.7}$$

To find y, multiply (10.3) by a_2 and (10.4) by a_1:

$$d_1a_2 = a_2a_1x + b_1a_2y \tag{10.8}$$
$$a_1d_2 = a_2a_1x + a_1b_2y. \tag{10.9}$$

Subtract (10.8) from (10.9):

$$\begin{aligned} a_1d_2 - d_1a_2 &= a_1b_2y - b_1a_2y \\ &= (a_1b_2 - b_1a_2)y \\ y &= \frac{a_1d_2 - d_1a_2}{a_1b_2 - b_1a_2}. \end{aligned} \tag{10.10}$$

Observe that both (10.7) and (10.10) share the common denominator: $a_1b_2 - b_1a_2$. Furthermore, note the positions of a_1, b_1, a_2 and b_2 in the original equations:

$$\begin{matrix} a_1 & b_1 \\ a_2 & b_2 \end{matrix}$$

and the denominator is formed by cross-multiplying the diagonal terms a_1b_2 and subtracting the other cross-multiplied terms b_1a_2. Placing the four terms between two vertical lines creates a *second-order determinant* whose value equals:

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - b_1a_2.$$

Although the name was originally given by Johann Gauss, it was the French mathematician Augustin-Louis Cauchy who clarified its current modern identity.

If the original equations were linearly related by a factor λ, the determinant equals zero:

$$\begin{vmatrix} a_1 & b_1 \\ \lambda a_1 & \lambda b_1 \end{vmatrix} = a_1 \lambda b_1 - b_1 \lambda a_1 = 0.$$

Observe that the numerators of (10.7) and (10.10) are also second-order determinants:

$$\begin{vmatrix} d_1 & b_1 \\ d_2 & b_2 \end{vmatrix} = d_1 b_2 - b_1 d_2$$

and

$$\begin{vmatrix} a_1 & d_1 \\ a_2 & d_2 \end{vmatrix} = a_1 d_2 - d_1 a_2$$

which means that Eqs. (10.7) and (10.10) can be written using determinants:

$$x = \frac{\begin{vmatrix} d_1 & b_1 \\ d_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}, \qquad y = \frac{\begin{vmatrix} a_1 & d_1 \\ a_2 & d_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}.$$

And one final piece of algebra permits the solution to be written as

$$\frac{x}{\begin{vmatrix} d_1 & b_1 \\ d_2 & b_2 \end{vmatrix}} = \frac{y}{\begin{vmatrix} a_1 & d_1 \\ a_2 & d_2 \end{vmatrix}} = \frac{1}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}. \tag{10.11}$$

Observe another pattern in (10.11) where the determinant is

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

but the d-terms replace the x-coefficients:

$$\begin{vmatrix} d_1 & b_1 \\ d_2 & b_2 \end{vmatrix}$$

and then the y-coefficients

$$\begin{vmatrix} a_1 & d_1 \\ a_2 & d_2 \end{vmatrix}.$$

Returning to the original equations:

$$7 = 3x + 2y$$
$$10 = 2x + 4y$$

and substituting the constants in (10.11), we have

$$\frac{x}{\begin{vmatrix} 7 & 2 \\ 10 & 4 \end{vmatrix}} = \frac{y}{\begin{vmatrix} 3 & 7 \\ 2 & 10 \end{vmatrix}} = \frac{1}{\begin{vmatrix} 3 & 2 \\ 2 & 4 \end{vmatrix}}$$

which, when expanded reveals

$$\frac{x}{28 - 20} = \frac{y}{30 - 14} = \frac{1}{12 - 4}$$

$$\frac{x}{8} = \frac{y}{16} = \frac{1}{8}$$

making $x = 1$ and $y = 2$.

Let's try another example:

$$11 = 4x + y$$
$$5 = x + y$$

and substituting the constants in (10.11), we have

$$\frac{x}{\begin{vmatrix} 11 & 1 \\ 5 & 1 \end{vmatrix}} = \frac{y}{\begin{vmatrix} 4 & 11 \\ 1 & 5 \end{vmatrix}} = \frac{1}{\begin{vmatrix} 4 & 1 \\ 1 & 1 \end{vmatrix}}$$

which, when expanded reveals

$$\frac{x}{11 - 5} = \frac{y}{20 - 11} = \frac{1}{4 - 1}$$

$$\frac{x}{6} = \frac{y}{9} = \frac{1}{3}$$

making $x = 2$ and $y = 3$.

Now let's see how a *third-order* determinant arises from the coefficients of three equations in three unknowns.

10.4 Linear Equations with Three Variables

Consider the following set of three linear equations:

$$13 = 3x + 2y + 2z \tag{10.12}$$
$$20 = 2x + 3y + 4z \tag{10.13}$$
$$7 = 2x + y + z. \tag{10.14}$$

A standard way to resolve this problem is to multiply (10.12) by 2 and subtract (10.13), which removes the z-terms:

$$\begin{aligned} 26 &= 6x + 4y + 4z \\ 20 &= 2x + 3y + 4z \\ 6 &= 4x + y \end{aligned} \tag{10.15}$$

leaving (10.15) with two unknowns.

Next, we take (10.13) and (10.14) and remove the z-term by multiplying (10.14) by 4 and subtract (10.13):

$$\begin{aligned} 28 &= 8x + 4y + 4z \\ 20 &= 2x + 3y + 4z \\ 8 &= 6x + y \end{aligned} \tag{10.16}$$

leaving (10.16) with two unknowns. We are now left with (10.15) and (10.16):

$$\begin{aligned} 6 &= 4x + y \\ 8 &= 6x + y \end{aligned}$$

which can be solved using (10.11):

$$\frac{x}{\begin{vmatrix} 6 & 1 \\ 8 & 1 \end{vmatrix}} = \frac{y}{\begin{vmatrix} 4 & 6 \\ 6 & 8 \end{vmatrix}} = \frac{1}{\begin{vmatrix} 4 & 1 \\ 6 & 1 \end{vmatrix}}$$

therefore,

$$\begin{aligned} x &= \frac{6-8}{4-6} = 1 \\ y &= \frac{32-36}{4-6} = 2. \end{aligned}$$

Substituting $x = 1$ and $y = 2$ in (10.12) reveals that $z = 3$.

We can generalise (10.11) for three equations using third-order determinants:

$$\frac{x}{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}} = \frac{y}{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}} = \frac{z}{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}} = \frac{1}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}. \tag{10.17}$$

Once again, there is an important pattern in (10.17) where the underlying determinant is

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

but the d-terms replace the x-coefficients:

$$\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}$$

the d-terms replace the y-coefficients:

$$\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}$$

and the d-terms replace the z-coefficients:

$$\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}.$$

We must now find a way of computing the value of a third-order determinant, which requires the following algebraic analysis of three equations in three unknowns. We start with three linear equations:

$$d_1 = a_1x + b_1y + c_1z \tag{10.18}$$

$$d_2 = a_2x + b_2y + c_2z \tag{10.19}$$

$$d_3 = a_3x + b_3y + c_3z \tag{10.20}$$

and derive one equation in two unknowns from (10.18) and (10.19), and another from (10.19) and (10.20).

We multiply (10.18) by c_2, (10.19) by c_1 and subtract them:

$$\begin{aligned} c_2 d_1 &= a_1 c_2 x + b_1 c_2 y + c_1 c_2 z \\ c_1 d_2 &= c_1 a_2 x + b_2 c_1 y + c_1 c_2 z \\ c_2 d_1 - c_1 d_2 &= (a_1 c_2 - c_1 a_2) x + (b_1 c_2 - b_2 c_1) y. \end{aligned} \tag{10.21}$$

Next, we multiply (10.19) by c_3, (10.20) by c_2 and subtract them:

$$\begin{aligned} c_3 d_2 &= a_2 c_3 x + b_2 c_3 y + c_2 c_3 z \\ c_2 d_3 &= a_3 c_2 x + b_3 c_2 y + c_2 c_3 z \\ c_3 d_2 - c_2 d_3 &= (a_2 c_3 - a_3 c_2) x + (b_2 c_3 - b_3 c_2) y. \end{aligned} \tag{10.22}$$

Simplify (10.21) by letting

$$\begin{aligned} e_1 &= c_2 d_1 - c_1 d_2 \\ f_1 &= a_1 c_2 - c_1 a_2 \\ g_1 &= b_1 c_2 - b_2 c_1 \end{aligned}$$

therefore,

$$e_1 = f_1 x + g_1 y. \tag{10.23}$$

Simplify (10.22) by letting

$$\begin{aligned} e_2 &= c_3 d_2 - c_2 d_3 \\ f_2 &= a_2 c_3 - a_3 c_2 \\ g_2 &= b_2 c_3 - b_3 c_2 \end{aligned}$$

therefore,

$$e_2 = f_2 x + g_2 y. \tag{10.24}$$

Now we have two equations in two unknowns:

$$\begin{aligned} e_1 &= f_1 x + g_1 y \\ e_2 &= f_2 x + g_2 y \end{aligned}$$

which are solved using

$$\frac{x}{A} = \frac{y}{B} = \frac{1}{C} \tag{10.25}$$

where

$$A = \begin{vmatrix} e_1 & g_1 \\ e_2 & g_2 \end{vmatrix} = \begin{vmatrix} c_2 d_1 - c_1 d_2 & b_1 c_2 - b_2 c_1 \\ c_3 d_2 - c_2 d_3 & b_2 c_3 - b_3 c_2 \end{vmatrix} \tag{10.26}$$

$$B = \begin{vmatrix} f_1 & e_1 \\ f_2 & e_2 \end{vmatrix} = \begin{vmatrix} a_1c_2 - c_1a_2 & c_2d_1 - c_1d_2 \\ a_2c_3 - a_3c_2 & c_3d_2 - c_2d_3 \end{vmatrix} \tag{10.27}$$

$$C = \begin{vmatrix} f_1 & g_1 \\ f_2 & g_2 \end{vmatrix} = \begin{vmatrix} a_1c_2 - c_1a_2 & b_1c_2 - b_2c_1 \\ a_2c_3 - a_3c_2 & b_2c_3 - b_3c_2 \end{vmatrix} \tag{10.28}$$

We first compute A, from which we can derive B, because the only difference between (10.26) and (10.27) is that d_1, d_2, d_3 become a_1, a_2, a_3 respectively, and b_1, b_2, b_3 become d_1, d_2, d_3 respectively.

We can derive C from A, as the only difference between (10.26) and (10.28) is that d_1, d_2, d_3 become a_1, a_2, a_3 respectively. Starting with A:

$$\begin{aligned} A &= (c_2d_1 - c_1d_2)(b_2c_3 - b_3c_2) - (b_1c_2 - b_2c_1)(c_3d_2 - c_2d_3) \\ &= b_2c_2c_3d_1 - b_3c_2^2d_1 - b_2c_1c_3d_2 + b_3c_1c_2d_2 \\ &\quad - b_1c_2c_3d_2 + b_1c_2^2d_3 + b_2c_1c_3d_2 - b_2c_1c_2d_3 \\ &= b_2c_2c_3d_1 - b_3c_2^2d_1 + b_3c_1c_2d_2 - b_1c_2c_3d_2 + b_1c_2^2d_3 - b_2c_1c_2d_3 \\ &= c_2(b_2c_3d_1 - b_3c_2d_1 + b_3c_1d_2 - b_1c_3d_2 + b_1c_2d_3 - b_2c_1d_3) \\ A &= c_2\Big(d_1(b_2c_3 - c_2b_3) - b_1(d_2c_3 - c_2d_3) + c_1(d_2b_3 - b_2d_3)\Big). \end{aligned} \tag{10.29}$$

Using the substitutions described above we can derive B and C from (10.29):

$$B = c_2\Big(a_1(d_2c_3 - c_2d_3) - b_1(a_2c_3 - c_2a_3) + c_1(a_2d_3 - d_2a_3)\Big) \tag{10.30}$$

$$C = c_2\Big(a_1(b_2c_3 - c_2b_3) - b_1(a_2c_3 - c_2a_3) + c_1(a_2b_3 - b_2a_3)\Big). \tag{10.31}$$

We can now rewrite (10.29), (10.30) and (10.31) using determinant notation. At the same time, we can drop the c_2 terms as they cancel out when computing x, y and z:

$$A = d_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1 \begin{vmatrix} d_2 & c_2 \\ d_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} d_2 & b_2 \\ d_3 & b_3 \end{vmatrix} \tag{10.32}$$

$$B = a_1 \begin{vmatrix} d_2 & c_2 \\ d_3 & c_3 \end{vmatrix} - d_1 \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} a_2 & d_2 \\ a_3 & d_3 \end{vmatrix} \tag{10.33}$$

$$C = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1 \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}. \tag{10.34}$$

As (10.17) and (10.25) refer to the same x and y, then

$$\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix} = d_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1 \begin{vmatrix} d_2 & c_2 \\ d_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} d_2 & b_2 \\ d_3 & b_3 \end{vmatrix} \tag{10.35}$$

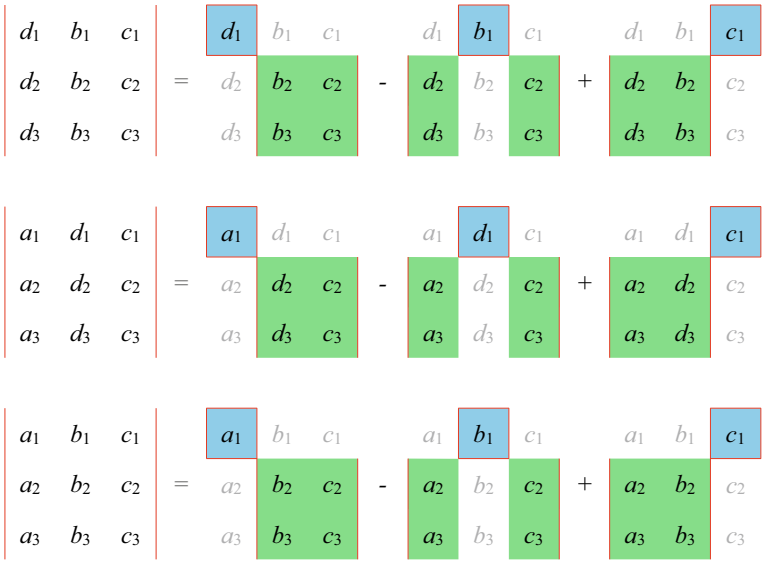

Fig. 10.1 Evaluating the determinants shown in (10.35)–(10.37)

$$\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} d_2 & c_2 \\ d_3 & c_3 \end{vmatrix} - d_1 \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} a_2 & d_2 \\ a_3 & d_3 \end{vmatrix} \tag{10.36}$$

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1 \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}. \tag{10.37}$$

As a consistent algebraic analysis has been pursued to derive (10.35), (10.36) and (10.37), a consistent pattern has surfaced in Fig. 10.1 which shows how the three determinants are evaluated. This pattern comprises taking each entry in the top row, called a *cofactor*, and multiplying it by the determinant of entries in rows 2 and 3, whilst ignoring the column containing the original term, called a *first minor*. Observe that the second term of the top row is switched negative, called an *inversion correction factor*.

Let's repeat (10.31) again without the c_2 term, as it has nothing to do with the calculation of the determinant.

$$C = a_1(b_2c_3 - c_2b_3) - b_1(a_2c_3 - c_2a_3) + c_1(a_2b_3 - b_2a_3). \tag{10.38}$$

It is possible to arrange the terms of (10.38) as a square matrix such that each row and column sums to C:

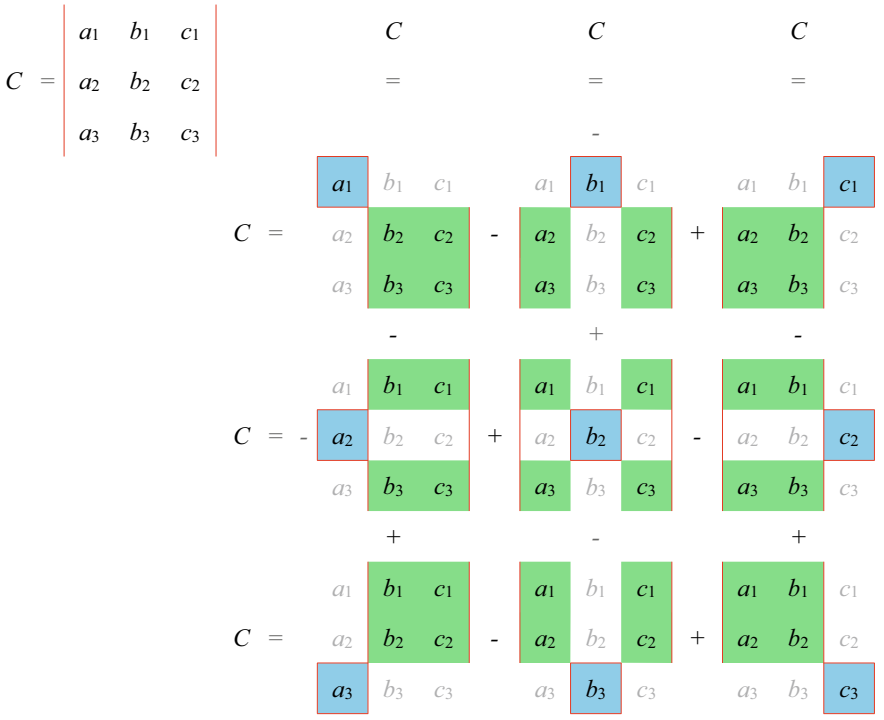

Fig. 10.2 The patterns of multipliers with their respective second-order determinants

$$
\begin{aligned}
a_1(b_2c_3 - c_2b_3) - b_1 \quad & (a_2c_3 - c_2a_3) + c_1(a_2b_3 - b_2a_3) \\
-a_2(b_1c_3 - c_1b_3) + b_2 \quad & (a_1c_3 - c_1a_3) - c_2(a_1b_3 - b_1a_3) \\
a_3(b_1c_2 - c_1b_2) - b_3 \quad & (a_1c_2 - c_1a_2) + c_3(a_1b_2 - b_1a_2)
\end{aligned}
$$

which means that there are six ways to evaluate the determinant C: summing the rows, or summing the columns. Figure 10.2 shows this arrangement with the cofactors in blue, and the first minor determinants in green. Observe how the signs alternate between the terms.

Having discovered the origins of these patterns, let's evaluate the original equations declared at the start of this section using (10.11)

$$
\begin{aligned}
13 &= 3x + 2y + 2z \\
20 &= 2x + 3y + 4z \\
7 &= 2x + y + z.
\end{aligned}
$$

$$\frac{x}{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}} = \frac{y}{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}} = \frac{z}{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}} = \frac{1}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}$$

therefore,

$$\frac{x}{\begin{vmatrix} 13 & 2 & 2 \\ 20 & 3 & 4 \\ 7 & 1 & 1 \end{vmatrix}} = \frac{y}{\begin{vmatrix} 3 & 13 & 2 \\ 2 & 20 & 4 \\ 2 & 7 & 1 \end{vmatrix}} = \frac{z}{\begin{vmatrix} 3 & 2 & 13 \\ 2 & 3 & 20 \\ 2 & 1 & 7 \end{vmatrix}} = \frac{1}{\begin{vmatrix} 3 & 2 & 2 \\ 2 & 3 & 4 \\ 2 & 1 & 1 \end{vmatrix}}$$

computing the determinants using the top row entries as cofactors:

$$\frac{x}{-13+16-2} = \frac{y}{-24+78-52} = \frac{z}{3+52-52} = \frac{1}{-3+12-8}$$

$$\frac{x}{1} = \frac{y}{2} = \frac{z}{3} = \frac{1}{1}$$

therefore, $x = 1$, $y = 2$ and $z = 3$.

10.4.1 Sarrus's Rule

The French mathematician Pierre Sarrus (1798–1861) discovered another way to compute the value of a third-order determinant, that arises from (10.38):

$$\begin{aligned} C &= a_1(b_2c_3 - c_2b_3) - b_1(a_2c_3 - c_2a_3) + c_1(a_2b_3 - b_2a_3) \\ &= a_1b_2c_3 - a_1c_2b_3 - b_1a_2c_3 + b_1c_2a_3 + c_1a_2b_3 - c_1b_2a_3 \\ &= a_1b_2c_3 + b_1c_2a_3 + c_1a_2b_3 - a_1c_2b_3 - b_1a_2c_3 - c_1b_2a_3. \end{aligned} \tag{10.39}$$

The pattern in (10.39) becomes clear in Fig. 10.3, where the first two columns of the matrix are repeated, and comprises two diagonal sets of terms: on the left in blue, we have the products $a_1b_2c_3$, $b_1c_2a_3$, $c_1a_2b_3$, and on the right in red and orange, the products $a_1c_2b_3$, $b_1a_2c_3$, $c_1b_2a_3$. These diagonal patterns provide a useful *aide-mémoire* when computing the determinant. Unfortunately, this rule only applies to third-order determinants.

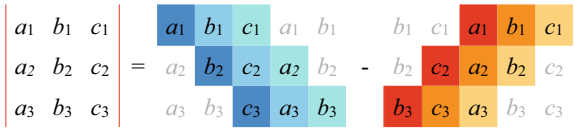

Fig. 10.3 The pattern behind Sarrus's rule

10.5 Mathematical Notation

Having discovered the background of determinants, now let's explore a formal description of their structure and characteristics.

10.5.1 Matrix

In the following definitions, a *matrix* is a square array of entries, with an equal number of rows and columns. The entries may be numbers, vectors, complex numbers or even partial differentials, in the case of a Jacobian. In general, each entry is identified by two subscripts $row\ col$:

$$a_{row\ col}.$$

A matrix with n rows and m columns has the following entries:

$$\begin{matrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{matrix}$$

The entries lying on the two diagonals are identified as follows: a_{11} and a_{nm} lie on the *main diagonal*, and a_{1m} and a_{n1} lie on the *secondary diagonal*.

10.5.2 Order of a Determinant

The *order* of a square determinant equals the number of rows or columns. For example, a first-order determinant contains a single entry; a second-order determinant has two rows and two columns; and a third-order determinant has three rows and three columns.

10.5.3 Value of a Determinant

A determinant posses a unique, single value derived from its entries. The algorithms used to compute this value must respect the algebra associated with solving sets of linear equations, as discussed above.

The French mathematician, astronomer, and physicist Pierre-Simon Laplace (1749–1827) developed a way to expand the determinant of any order. The Laplace expansion is the idea described above and shown in Fig. 10.1, where cofactors and

first minors or *principal minors* are used. For example, starting with a fourth-order determinant, when any row **and** column are removed, the remaining entries create a third-order determinant, called the *first minor* of the original determinant.

The following equation is used to control the sign of each cofactor:

$$(-1)^{row+col}$$

which, for a fourth-order determinant creates:

$$\begin{vmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{vmatrix}.$$

The Laplace expansion begins by choosing a convenient row or column as the source of cofactors. Any zeros are particularly useful, as they cancel out any contribution by the first minor determinant. It then sums the products of every cofactor in the chosen row or column, with its associated first minor, including an appropriate inversion correction factor to adjust the sign changes. The final result is the determinant's value.

A first-order determinant:

$$\left| a_{11} \right| = a_{11}.$$

A second-order determinant:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

A third-order determinant using the Laplace expansion with cofactors from the first row:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

A fourth-order determinant using the Laplace expansion with cofactors from the first row:

$$\begin{aligned} a_{11} \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix} & + \\ a_{13} \begin{vmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} - a_{14} \begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{vmatrix} &= \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} \end{aligned}$$

Sarrus's rule is useful to compute a third-order determinant:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{13}a_{22}a_{31}$$

The Laplace expansion works with higher-order determinants, as any first minor can itself be expanded using the same expansion.

10.5.4 Properties of Determinants

If a determinant contains a row or column of zeros, the Laplace expansion implies that the value of the determinant is zero.

$$\begin{vmatrix} 3 & 0 & 2 \\ 2 & 0 & 4 \\ 2 & 0 & 1 \end{vmatrix} = 0.$$

If a determinant's rows and columns are interchanged, the Laplace expansion also implies that the value of the determinant is unchanged.

$$\begin{vmatrix} 3 & 12 & 2 \\ 2 & 10 & 4 \\ 2 & 8 & 1 \end{vmatrix} = \begin{vmatrix} 3 & 2 & 2 \\ 12 & 10 & 8 \\ 2 & 4 & 1 \end{vmatrix} = -2.$$

If any two rows, or columns, are interchanged, without changing the order of their entries, the determinant's numerical value is unchanged, but its sign is reversed.

$$\begin{vmatrix} 3 & 12 & 2 \\ 2 & 10 & 4 \\ 2 & 8 & 1 \end{vmatrix} = -2$$

$$\begin{vmatrix} 12 & 3 & 2 \\ 10 & 2 & 4 \\ 8 & 2 & 1 \end{vmatrix} = 2.$$

If the entries of a row or column share a common factor, the entries may be adjusted, and the factor placed outside.

$$\begin{vmatrix} 3 & 12 & 2 \\ 2 & 10 & 4 \\ 2 & 8 & 1 \end{vmatrix} = 2 \begin{vmatrix} 3 & 6 & 2 \\ 2 & 5 & 4 \\ 2 & 4 & 1 \end{vmatrix} = -2.$$

10.6 Worked Examples

10.6.1 Determinant Expansion

Evaluate this determinant using the Laplace expansion and Sarrus's rule.

$$\begin{vmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{vmatrix}.$$

Solution: Using the Laplace expansion:

$$\begin{aligned} \begin{vmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{vmatrix} &= 1 \begin{vmatrix} 5 & 8 \\ 6 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 7 \\ 6 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 7 \\ 5 & 8 \end{vmatrix} \\ &= 1(45 - 48) - 2(36 - 42) + 3(32 - 35) \\ &= -3 + 12 - 9 \\ &= 0. \end{aligned}$$

Using Sarrus's rule:

$$\begin{aligned} \begin{vmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{vmatrix} &= 1 \times 5 \times 9 + 4 \times 8 \times 3 + 7 \times 2 \times 6 - 7 \times 5 \times 3 - 1 \times 8 \times 6 - 4 \times 2 \times 9 \\ &= 45 + 96 + 84 - 105 - 48 - 72 \\ &= 0. \end{aligned}$$

10.6.2 Complex Determinant

Evaluate the complex determinant

$$\begin{vmatrix} 4 + i2 & 1 + i \\ 2 - i3 & 3 + i3 \end{vmatrix}.$$

Solution: Using the Laplace expansion:

$$\begin{aligned} \begin{vmatrix} 4 + i2 & 1 + i \\ 2 - i3 & 3 + i3 \end{vmatrix} &= (4 + i2)(3 + i3) - (1 + i)(2 - i3) \\ &= (12 + i18 - 6) - (2 - i + 3) \\ &= 6 + i18 - 5 + i \\ &= 1 + i19. \end{aligned}$$

10.6.3 Simple Expansion

Write down the simplest expansion of this determinant with its value:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 0 \\ 6 & 7 & 0 \end{vmatrix}.$$

Solution: Using the Laplace expansion with cofactors from the third column:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 0 \\ 6 & 7 & 0 \end{vmatrix} = 3 \begin{vmatrix} 4 & 5 \\ 6 & 7 \end{vmatrix} = -6.$$

10.6.4 Simultaneous Equations

Solve the following equations using determinants:

$$\begin{aligned} 3 &= 2x + y - z \\ 12 &= x + 2y + z \\ 8 &= 3x - 2y + 2z. \end{aligned}$$

Solution: Using (10.17):

$$\frac{x}{\begin{vmatrix} 3 & 1 & -1 \\ 12 & 2 & 1 \\ 8 & -2 & 2 \end{vmatrix}} = \frac{y}{\begin{vmatrix} 2 & 3 & -1 \\ 1 & 12 & 1 \\ 3 & 8 & 2 \end{vmatrix}} = \frac{z}{\begin{vmatrix} 2 & 1 & 3 \\ 1 & 2 & 12 \\ 3 & -2 & 8 \end{vmatrix}} = \frac{1}{\begin{vmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ 3 & -2 & 2 \end{vmatrix}}.$$

Therefore,

$$x = \frac{\begin{vmatrix} 3 & 1 & -1 \\ 12 & 2 & 1 \\ 8 & -2 & 2 \end{vmatrix}}{\begin{vmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ 3 & -2 & 2 \end{vmatrix}} = \frac{18 - 16 + 40}{12 + 1 + 8} = \frac{42}{21} = 2,$$

$$y = \frac{\begin{vmatrix} 2 & 3 & -1 \\ 1 & 12 & 1 \\ 3 & 8 & 2 \end{vmatrix}}{\begin{vmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ 3 & -2 & 2 \end{vmatrix}} = \frac{32 + 3 + 28}{12 + 1 + 8} = \frac{63}{21} = 3,$$

$$z = \frac{\begin{vmatrix} 2 & 1 & 3 \\ 1 & 2 & 12 \\ 3 & -2 & 8 \end{vmatrix}}{\begin{vmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ 3 & -2 & 2 \end{vmatrix}} = \frac{80 + 28 - 24}{24 + 1 + 8} = \frac{84}{21} = 4.$$

Chapter 11
Vectors

11.1 Introduction

This chapter introduces the historical development of vectors and shows how a graphical interpretation increases one's understanding of its mathematical manipulation. The chapter includes position vectors and Cartesian vectors, their addition, subtraction and products, and includes several worked examples. Readers interested in the history of mathematics are recommended to read Michael Crowe's excellent book: *A History of Vector Analysis* (Crowe 2011).

11.2 Background

Vectors are a relative new invention in the world of mathematics, dating only from the 19th century. They enable us to solve complex geometric problems, the dynamics of moving objects, and problems involving forces and fields.

We often only require a single number to represent quantities used in our daily lives such as height, age, shoe size, waist and chest measurement. The magnitude of these numbers depends on our age and whether we use metric or imperial units. Such quantities are called *scalars*. On the other hand, there are some things that require more than one number to represent them: wind, force, weight, velocity and sound are just a few examples. For example, any sailor knows that wind has a magnitude and a direction. The force we use to lift an object also has a value *and* a direction. Similarly, the velocity of a moving object is measured in terms of its speed (e.g. miles per hour), and a direction such as north-west. Sound, too, has intensity and a direction. Such quantities are called *vectors*.

Complex numbers seemed to be a likely candidate for representing forces, and were being investigated by the Norwegian-Danish mathematician Caspar Wessel (1745–1818), the French amateur mathematician Jean-Robert Argand (1768–1822) and the English mathematician John Warren (1796–1852). At the time, complex

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_11

numbers were two-dimensional, and their 3D form was being investigated by Sir William Rowan Hamilton who invented them in 1843, calling them *quaternions*. In 1853, Hamilton published his book *Lectures on Quaternions* in which he described terms such as "*vector*", "*transvector*" and "*provector*". Hamilton's work was not widely accepted until in 1881, when Josiah Gibbs published his treatise *Vector Analysis*, describing modern *vector analysis*.

Gibbs was not a fan of the imaginary quantities associated with Hamilton's quaternions, but saw the potential of creating a vectorial system from the imaginary i, j and k into the unit basis vectors **i**, **j** and **k**, which is what we use today.

Some mathematicians were not happy with the direction mathematics had taken. The German mathematician Hermann Gunther Grassmann (1809–1877), believed that his own *geometric calculus* was far superior to Hamilton's quaternions, but he died without managing to convince any of his fellow mathematicians. Fortunately, the English mathematician and philosopher William Kingdon Clifford (1845–1879) recognised the brilliance of Grassmann's ideas, and formalised what today has become known as *geometric algebra*.

With the success of Gibbs' vector analysis, quaternions faded into obscurity, only to be rediscovered in the 1970s when they were employed by the flight simulation community to control the dynamic behaviour of a simulator's motion platform. A decade later they found their way into computer graphics where they are used for rotations about an arbitrary axis.

Now this does not mean that vector analysis is dead – far from it. Vast quantities of scientific software depends upon the vector mathematics developed over a century ago, and will continue to employ it for many years to come. Nevertheless, geometric algebra is destined to emerge as a powerful mathematical framework that could eventually replace vector analysis one day.

11.3 2D Vectors

11.3.1 Vector Notation

A scalar such as x represents a single numeric quantity. However, as a vector contains two or more numbers, its symbolic name is printed using a **bold** font to distinguish it from a scalar variable. Examples being **n**, **i** and **q**.

When a scalar variable is assigned a value, we use the standard algebraic notation:

$$x = 3.$$

However, a vector has one or more numbers enclosed in brackets, written as a column or as a row – in this text *column vectors* are used:

$$\mathbf{n} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

The numbers 3 and 4 are the *components* of **n**, and their sequence within the brackets is important. A *row vector* places the components horizontally:

$$\mathbf{n} = [3 \quad 4].$$

The difference between the two, is only appreciated in the context of matrices. Sometimes it is convenient – for presentation purposes – to write a column vector as a row vector, in which case, it is written

$$\mathbf{n} = [3 \quad 4]^T,$$

where the superscript ^{T} reminds us that **n** is really a *transposed* column vector.

11.3.2 Graphical Representation of Vectors

An arrow is used to represent a vector as it possesses length and direction, as shown in Fig. 11.1. By assigning coordinates to the arrow it is possible to translate the arrow's length and direction into two numbers. For example, in Fig. 11.2 the vector **r** has its tail defined by $(x_1,\ y_1) = (1,\ 2)$, and its head by $(x_2,\ y_2) = (3,\ 4)$. Vector **s** has its tail defined by $(x_3,\ y_3) = (5,\ 3)$, and its head by $(x_4,\ y_4) = (3,\ 1)$. The x- and y-components for **r** are computed as follows

$$x_r = x_2 - x_1 = 3 - 1 = 2$$
$$y_r = y_2 - y_1 = 4 - 2 = 2$$

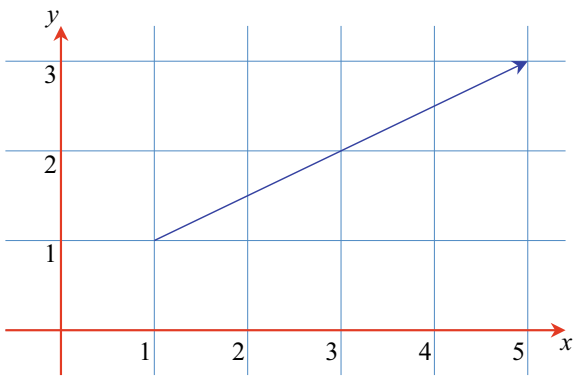

Fig. 11.1 An arrow with magnitude and direction

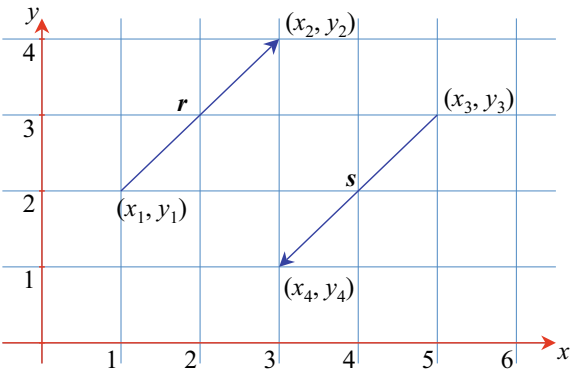

Fig. 11.2 Two vectors **r** and **s** have the same magnitude but opposite directions

and the components for **s** are computed as follows

$$x_s = x_4 - x_3 = 3 - 5 = -2$$
$$y_s = y_4 - y_3 = 1 - 3 = -2.$$

It is the negative value of x_s and y_s that encode the vector's direction. In general, if the coordinates of a vector's head and tail are $(x_h,\ y_h)$ and $(x_t,\ y_t)$ respectively, its components Δx and Δy are given by

$$\Delta x = x_h - x_t$$
$$\Delta y = y_h - y_t.$$

One can readily see from this notation that a vector does not have an absolute position. It does not matter where we place a vector, so long as we preserve its length and orientation, its components are unaltered.

11.3.3 Magnitude of a Vector

The *magnitude* or length of a vector **r** is written $|\mathbf{r}|$ and computed using the theorem of Pythagoras:

$$|\mathbf{r}| = \sqrt{(\Delta x)^2 + (\Delta y)^2}$$

and used as follows. Consider a vector defined by

$$(x_h,\ y_h) = (4,\ 5)$$
$$(x_t,\ y_t) = (1,\ 1)$$

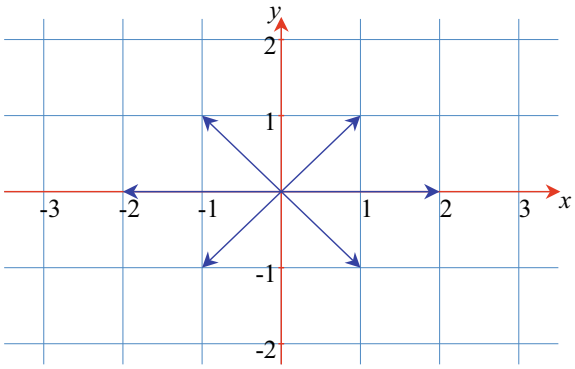

Fig. 11.3 Eight vectors whose coordinates are shown in Table 11.1

Table 11.1 Values associated with the eight vectors in Fig. 11.3

x_h	y_h	x_t	y_t	Δx	Δy	**|vector|**
2	0	0	0	2	0	2
0	2	0	0	0	2	2
−2	0	0	0	−2	0	2
0	−2	0	0	0	−2	2
1	1	0	0	1	1	$\sqrt{2}$
−1	1	0	0	−1	1	$\sqrt{2}$
−1	−1	0	0	−1	−1	$\sqrt{2}$
1	−1	0	0	1	−1	$\sqrt{2}$

where the x- and y-components are 3 and 4 respectively. Therefore its magnitude equals $\sqrt{3^2 + 4^2} = 5$.

Figure 11.3 shows eight vectors, and their geometric properties are listed in Table 11.1.

11.4 3D Vectors

The above vector examples are in 2D, but it is easy to extend this notation to embrace an extra dimension. Figure 11.4 shows a 3D vector **r** with its head, tail, components and magnitude annotated. The vector, its components and magnitude are given by

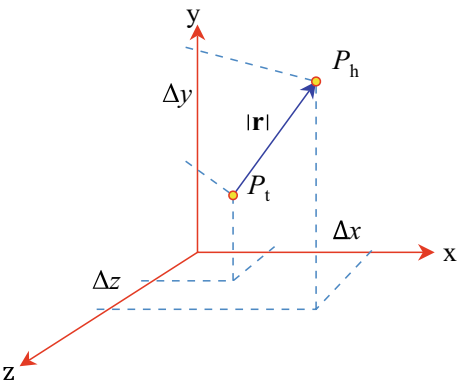

Fig. 11.4 The vector **r** has components Δx, Δy, Δz

$$
\begin{aligned}
\mathbf{r} &= [\Delta x \quad \Delta y \quad \Delta z]^T \\
\Delta x &= x_h - x_t \\
\Delta y &= y_h - y_t \\
\Delta z &= z_h - z_t \\
|\mathbf{r}| &= \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}.
\end{aligned}
$$

All future examples are three-dimensional.

11.4.1 Vector Manipulation

As vectors are different to scalars, there are rules to control how the two mathematical entities interact with one another. For instance, we need to consider vector addition, subtraction and products, and how a vector is scaled.

11.4.2 Scaling a Vector

Given a vector **n**, 2**n** means that the vectors components are scaled by a factor of 2. For example, given

$$
\mathbf{n} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \quad \text{then} \quad 2\mathbf{n} = \begin{bmatrix} 6 \\ 8 \\ 10 \end{bmatrix}
$$

which seems logical. Similarly, if we divide **n** by 2, its components are halved. Note that the vector's direction remains unchanged – only its magnitude changes.

In general, given

$$\mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}, \quad \text{then} \quad \lambda\mathbf{n} = \begin{bmatrix} \lambda n_1 \\ \lambda n_2 \\ \lambda n_3 \end{bmatrix}, \text{where} \quad \lambda \in \mathrm{R}.$$

There is no obvious way we can resolve the expression $2 + \mathbf{n}$, for it is not clear which component of $\mathbf{n}$ is to be increased by 2. However, if we can add a scalar to an imaginary (e.g. $2 + 3i$), why can't we add a scalar to a vector (e.g. $2 + \mathbf{n}$)? Well, the answer to this question is two-fold: First, if we change the meaning of "add" to mean "associated with", then there is nothing to stop us from "associating" a scalar with a vector, like complex numbers. Second, the axioms controlling our algebra must be clear on this matter. Unfortunately, the axioms of traditional vector analysis do not support the "association" of scalars with vectors in this way. However, geometric algebra does! Furthermore, geometric algebra even permits division by a vector, which does sound strange. Consequently, whilst reading the rest of this chapter keep an open mind about what is permitted, and what is not permitted. At the end of the day, virtually anything is possible, so long as we have a well-behaved axiomatic system.

11.4.3 Vector Addition and Subtraction

Given vectors $\mathbf{r}$ and $\mathbf{s}$, $\mathbf{r} \pm \mathbf{s}$ is defined as

$$\mathbf{r} = \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix}, \quad \text{then} \quad \mathbf{r} \pm \mathbf{s} = \begin{bmatrix} x_r \pm x_s \\ y_r \pm y_s \\ z_r \pm z_s \end{bmatrix}.$$

Vector addition is commutative:

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$$

$$\text{e.g.} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

However, like scalar subtraction, vector subtraction is not commutative:

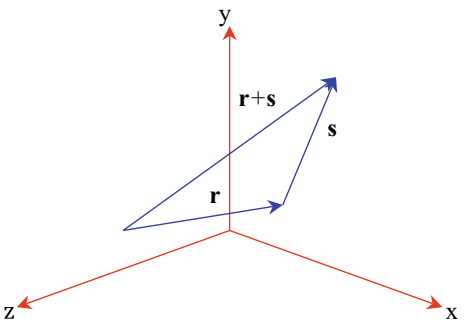

Fig. 11.5 Vector addition $\mathbf{r} + \mathbf{s}$

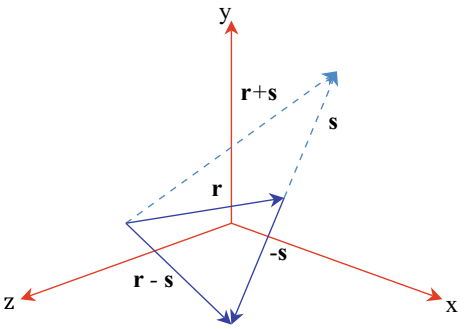

Fig. 11.6 Vector subtraction $\mathbf{r} - \mathbf{s}$

$$\mathbf{a} - \mathbf{b} \neq \mathbf{b} - \mathbf{a}$$

$$\text{e.g.} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

Let's illustrate vector addition and subtraction with two examples. Figure 11.5 shows the graphical interpretation of adding two vectors $\mathbf{r}$ and $\mathbf{s}$. Note that the tail of vector $\mathbf{s}$ is attached to the head of vector $\mathbf{r}$. The resultant vector $\mathbf{t} = \mathbf{r} + \mathbf{s}$ is defined by adding the corresponding components of $\mathbf{r}$ and $\mathbf{s}$ together. Figure 11.6 shows a graphical interpretation for $\mathbf{r} - \mathbf{s}$. This time, the components of vector $\mathbf{s}$ are reversed to produce an equal and opposite vector. Then it is attached to $\mathbf{r}$ and added as described above.

11.4.4 Position Vectors

Given any point $P(x,\ y,\ z)$, a *position vector* $\mathbf{p}$ is created by assuming that P is the vector's head and the origin is its tail. As the tail coordinates are $(0,\ 0,\ 0)$ the

vector's components are x, y, z. Consequently, the vector's magnitude $|\mathbf{p}|$ equals $\sqrt{x^2 + y^2 + z^2}$.

11.4.5 Unit Vectors

By definition, a *unit vector* has a magnitude of 1. A simple example is $\mathbf{i}$, where

$$\mathbf{i} = [1 \quad 0 \quad 0]^T, \quad \text{where} \quad |\mathbf{i}| = 1.$$

Unit vectors are extremely useful in the product of two vectors, where their magnitudes are required; and if these are unit vectors, the computation is greatly simplified.

Converting a vector into a unit form is called *normalising*, and is achieved by dividing its components by the vector's magnitude. To formalise this process, consider a vector $\mathbf{r} = [x \quad y \quad z]^T$, with magnitude $|\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$. The unit form of $\mathbf{r}$ is given by

$$\hat{\mathbf{r}} = \frac{1}{|\mathbf{r}|}[x \quad y \quad z]^T$$

This is confirmed by showing that the magnitude of $\hat{\mathbf{r}}$ is 1:

$$\begin{aligned} |\hat{\mathbf{r}}| &= \sqrt{\left(\frac{x}{|\hat{\mathbf{r}}|}\right)^2 + \left(\frac{y}{|\hat{\mathbf{r}}|}\right)^2 + \left(\frac{z}{|\hat{\mathbf{r}}|}\right)^2} \\ &= \frac{1}{|\hat{\mathbf{r}}|}\sqrt{x^2 + y^2 + z^2} \\ |\hat{\mathbf{r}}| &= 1. \end{aligned}$$

11.4.6 Cartesian Vectors

A *Cartesian vector* is constructed from three unit vectors: $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, aligned with the x-, y- and z-axis, respectively:

$$\mathbf{i} = [1 \quad 0 \quad 0]^T, \qquad \mathbf{j} = [0 \quad 1 \quad 0]^T, \qquad \mathbf{k} = [0 \quad 0 \quad 1]^T.$$

Therefore, any vector aligned with the x-, y- or z-axis is a scalar multiple of the associated unit vector. For example, $10\mathbf{i}$ is aligned with the x-axis, with a magnitude of 10. $20\mathbf{k}$ is aligned with the z-axis, with a magnitude of 20. By employing the rules of vector addition and subtraction, we can compose a vector $\mathbf{r}$ by summing three scaled Cartesian unit vectors as follows

$$\mathbf{r} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

which is equivalent to

$$\mathbf{r} = [a \quad b \quad c]^T$$

where the magnitude of $\mathbf{r}$ is

$$|\mathbf{r}| = \sqrt{a^2 + b^2 + c^2}.$$

Any pair of Cartesian vectors, such as $\mathbf{r}$ and $\mathbf{s}$, can be combined as follows

$$\begin{aligned} \mathbf{r} &= a\mathbf{i} + b\mathbf{j} + c\mathbf{k} \\ \mathbf{s} &= d\mathbf{i} + e\mathbf{j} + f\mathbf{k} \\ \mathbf{r} \pm \mathbf{s} &= (a \pm d)\mathbf{i} + (b \pm e)\mathbf{j} + (c \pm f)\mathbf{k}. \end{aligned}$$

11.4.7 Products

The product of two scalars is very familiar: for example, 6×7 or $7 \times 6 = 42$. We often visualise this operation as a rectangular area, where 6 and 7 are the dimensions of a rectangle's sides, and 42 is the area. However, a vector's qualities are its length and orientation, which means that any product must include them in any calculation. The length is easily calculated, but we must know the angle between the two vectors as this reflects their relative orientation. Although the angle can be incorporated within the product in various ways, two particular ways lead to useful results. For example, the product of $\mathbf{r}$ and $\mathbf{s}$, separated by an angle θ could be $|\mathbf{r}||\mathbf{s}| \cos\theta$ or $|\mathbf{r}||\mathbf{s}| \sin\theta$. It just so happens that $\cos\theta$ forces the product to result in a scalar quantity, and $\sin\theta$ creates a vector. Consequently, there are two products to consider: the *scalar* product, and the *vector* product, which are written as $\mathbf{r} \cdot \mathbf{s}$ and $\mathbf{r} \times \mathbf{s}$ respectively.

11.4.8 Scalar Product

Figure 11.7 shows two vectors $\mathbf{r}$ and $\mathbf{s}$ that have been drawn, for convenience, with their tails touching. Taking $\mathbf{s}$ as the reference vector – which is an arbitrary choice – we compute the projection of $\mathbf{r}$ on $\mathbf{s}$, which takes into account their relative orientation. The length of $\mathbf{r}$ on $\mathbf{s}$ is $|\mathbf{r}| \cos\theta$. The product of $|\mathbf{s}|$ and $|\mathbf{r}| \cos\theta$, becomes the scalar product, which is the product of the two vector lengths projected onto either vector. As cos is an even function, $\cos\theta = \cos(-\theta)$, the projection can be on either vector. This scalar product is written

$$\mathbf{r} \cdot \mathbf{s} = |\mathbf{r}||\mathbf{s}| \cos\theta. \tag{11.1}$$

Because of the dot symbol "$\cdot$", the scalar product is also called the *dot* product.

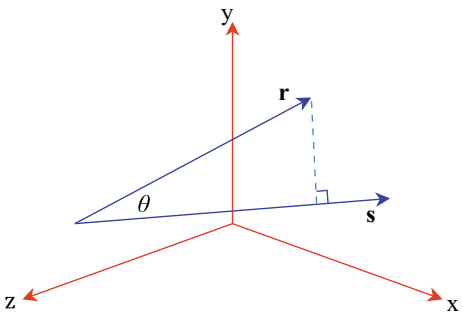

Fig. 11.7 The projection of **r** on **s**

Fortunately, everything is in place to perform this task. To begin with, we define two Cartesian vectors **r** and **s**, and proceed to multiply them together using (11.1):

$$\begin{aligned}
\mathbf{r} &= a\mathbf{i} + b\mathbf{j} + c\mathbf{k} \\
\mathbf{s} &= d\mathbf{i} + e\mathbf{j} + f\mathbf{k} \\
\mathbf{r} \cdot \mathbf{s} &= (a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) \cdot (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\
&= a\mathbf{i} \cdot (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\
&\quad + b\mathbf{j} \cdot (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\
&\quad + c\mathbf{k} \cdot (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\
&= ad\mathbf{i} \cdot \mathbf{i} + ae\mathbf{i} \cdot \mathbf{j} + af\mathbf{i} \cdot \mathbf{k} \\
&\quad + bd\mathbf{j} \cdot \mathbf{i} + be\mathbf{j} \cdot \mathbf{j} + bf\mathbf{j} \cdot \mathbf{k} \\
&\quad + cd\mathbf{k} \cdot \mathbf{i} + ce\mathbf{k} \cdot \mathbf{j} + cf\mathbf{k} \cdot \mathbf{k}.
\end{aligned}$$

Before we proceed any further, we can see that we have created various dot product terms such as $\mathbf{i} \cdot \mathbf{i}$, $\mathbf{i} \cdot \mathbf{j}$, $\mathbf{i} \cdot \mathbf{k}$, etc. These terms can be divided into two groups: those that reference the same unit vector, and those that reference different unit vectors.

Using the definition of the dot product (11.1), terms such as $\mathbf{i} \cdot \mathbf{i}$, $\mathbf{j} \cdot \mathbf{j}$ and $\mathbf{k} \cdot \mathbf{k} = 1$, because the angle between **i** and **i**, **j** and **j**, or **k** and **k**, is 0°; and $\cos 0° = 1$. But as the other vector combinations are separated by 90°, and $\cos 90° = 0$, all remaining terms collapse to zero, and we are left with

$$\mathbf{r} \cdot \mathbf{s} = ad\mathbf{i} \cdot \mathbf{i} + ae\mathbf{i} \cdot \mathbf{j} + af\mathbf{i} \cdot \mathbf{k}.$$

But as the the magnitude of a unit vector is 1, we can write

$$\mathbf{r} \cdot \mathbf{s} = |\mathbf{r}||\mathbf{s}| \cos\theta = ad + be + cf$$

which confirms that the dot product is indeed a scalar quantity.

It is worth pointing out that the angle returned by the dot product ranges between 0° and 180°. This is because, as the angle between two vectors increases beyond 180° the returned angle θ is always the smallest angle associated with the geometry.

11.4.9 The Vector Product

As mentioned above, the vector product $\mathbf{r} \times \mathbf{s}$ creates a third vector whose magnitude equals $|\mathbf{r}||\mathbf{s}| \sin\theta$, where θ is the angle between the original vectors. Figure 11.8 reminds us that the area of a parallelogram formed by $\mathbf{r}$ and $\mathbf{s}$ equals $|\mathbf{r}||\mathbf{s}| \sin\theta$. Because of the cross symbol "×", the vector product is also called the *cross* product.

$$\begin{aligned} \mathbf{r} \times \mathbf{s} &= \mathbf{t} \\ |\mathbf{t}| &= |\mathbf{r}||\mathbf{s}| \sin\theta. \end{aligned} \tag{11.2}$$

We will discover that the vector $\mathbf{t}$ is normal (90°) to the plane containing the vectors $\mathbf{r}$ and $\mathbf{s}$, as shown in Fig. 11.9, which makes it an ideal way of computing the vector normal to a surface. Once again, let's define two vectors and this time multiply them together using (11.2):

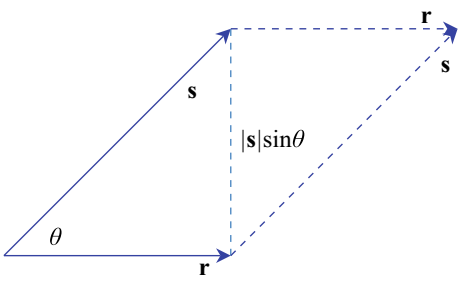

Fig. 11.8 The area of the parallelogram formed by $\mathbf{r}$ and $\mathbf{s}$

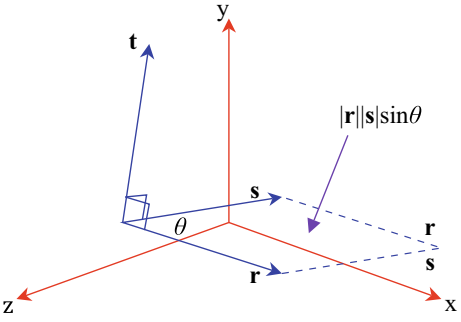

Fig. 11.9 The vector product

$$
\begin{aligned}
\mathbf{r} &= a\mathbf{i} + b\mathbf{j} + c\mathbf{k} \\
\mathbf{s} &= d\mathbf{i} + e\mathbf{j} + f\mathbf{k} \\
\mathbf{r} \times \mathbf{s} &= (a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) \times (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\
&= a\mathbf{i} \times (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\
&\quad + b\mathbf{j} \times (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\
&\quad + c\mathbf{k} \times (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\
&= ad\mathbf{i} \times \mathbf{i} + ae\mathbf{i} \times \mathbf{j} + af\mathbf{i} \times \mathbf{k} \\
&\quad + bd\mathbf{j} \times \mathbf{i} + be\mathbf{j} \times \mathbf{j} + bf\mathbf{j} \times \mathbf{k} \\
&\quad + cd\mathbf{k} \times \mathbf{i} + ce\mathbf{k} \times \mathbf{j} + cf\mathbf{k} \times \mathbf{k}.
\end{aligned}
$$

As we found with the dot product, there are two groups of vector terms: those that reference the same unit vector, and those that reference different unit vectors.

Using the definition for the cross product (11.2), operations such as $\mathbf{i} \times \mathbf{i}$, $\mathbf{j} \times \mathbf{j}$ and $\mathbf{k} \times \mathbf{k}$ result in a vector whose magnitude is 0. This is because the angle between the vectors is 0°, and $\sin 0° = 0$. Consequently these terms disappear and we are left with

$$\mathbf{r} \times \mathbf{s} = ae\mathbf{i} \times \mathbf{j} + af\mathbf{i} \times \mathbf{k} + bd\mathbf{j} \times \mathbf{i} + bf\mathbf{j} \times \mathbf{k} + cd\mathbf{k} \times \mathbf{i} + ce\mathbf{k} \times \mathbf{j}. \quad (11.3)$$

Sir William Rowan Hamilton struggled for many years when working on quaternions to resolve the meaning of a similar result. At the time, he was not using vectors, as they had yet to be defined, but the imaginary terms i, j and k. Hamilton's problem was to resolve the products ij, jk, ki and their opposites ji, kj and ik. What did the products mean? He reasoned that $ij = k$, $jk = i$ and $ki = j$, but could not resolve their opposites. One day in 1843, when he was out walking, thinking about this problem, he thought the impossible: $ij = k$, but $ji = -k$, $jk = i$, but $kj = -i$, and $ki = j$, but $ik = -j$. To his surprise, this worked, but it contradicted the commutative multiplication law of scalars where $6 \times 7 = 7 \times 6$. We now accept that the commutative multiplication law is there to be broken!

Although Hamilton had invented 3D complex numbers, to which he gave the name *quaternions*, they were not popular with everyone. And as mentioned earlier, Josiah Gibbs saw that converting the imaginary i, j and k terms into the unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ created a stable algebra for manipulating vectors, and for over a century we have been using Gibbs' vector notation.

The question we must ask is "Was Gibbs right?" to which the answer is probably "no!" The reason for this is that although the scalar product works in space of any number of dimensions, the vector (cross) product does not. It obviously does not work in 2D as there is no direction for the resultant vector. It obviously works in 3D, but in 4D and above there is no automatic spatial direction for the resultant vector. So, the vector product is possibly a special condition of some other structure. Hermann Grassmann knew this but did not have the mathematical reputation to convince his fellow mathematicians.

Let's continue with Hamilton's rules and reduce the cross product terms of (11.3) to

$$\mathbf{r} \times \mathbf{s} = ae\mathbf{k} - af\mathbf{j} - bd\mathbf{k} + bf\mathbf{i} + cd\mathbf{j} - ce\mathbf{i}. \tag{11.4}$$

Equation (11.4) can be tidied up to bring like terms together:

$$\mathbf{r} \times \mathbf{s} = (bf - ce)\mathbf{i} + (cd - af)\mathbf{j} + (ae - bd)\mathbf{k}. \tag{11.5}$$

Now let's repeat the original vector equations to see how Eq. (11.5) is computed:

$$\begin{aligned} \mathbf{r} &= a\mathbf{i} + b\mathbf{j} + c\mathbf{k} \\ \mathbf{s} &= d\mathbf{i} + e\mathbf{j} + f\mathbf{k} \\ \mathbf{r} \times \mathbf{s} &= (bf - ce)\mathbf{i} + (cd - af)\mathbf{j} + (ae - bd)\mathbf{k}. \end{aligned} \tag{11.6}$$

To compute the **i** scalar term we consider the scalars associated with the other two unit vectors, i.e. b, c, e, and f, and cross-multiply and subtract them to form $(bf - ce)$.

To compute the **j** scalar term we consider the scalars associated with the other two unit vectors, i.e. a, c, d, and f, and cross-multiply and subtract them to form $(cd - af)$.

To compute the **k** scalar term we consider the scalars associated with the other two unit vectors, i.e. a, b, d, and e, and cross-multiply and subtract them to form $(ae - bd)$.

The middle operation seems out of step with the other two, but in fact it preserves a cyclic symmetry often found in mathematics. Nevertheless, some authors reverse the sign of the **j** scalar term and cross-multiply and subtract the terms to produce $-(af - cd)$ which maintains a visual pattern for remembering the cross-multiplication. Equation (11.6) now becomes

$$\mathbf{r} \times \mathbf{s} = (bf - ce)\mathbf{i} - (af - cd)\mathbf{j} + (ae - bd)\mathbf{k}. \tag{11.7}$$

However, we now have to remember to introduce a negative sign for the **j** scalar term!

We can write (11.7) using determinants as follows:

$$\mathbf{r} \times \mathbf{s} = \begin{vmatrix} b & c \\ e & f \end{vmatrix} \mathbf{i} - \begin{vmatrix} a & c \\ d & f \end{vmatrix} \mathbf{j} + \begin{vmatrix} a & b \\ d & e \end{vmatrix} \mathbf{k}.$$

or

$$\mathbf{r} \times \mathbf{s} = \begin{vmatrix} b & c \\ e & f \end{vmatrix} \mathbf{i} + \begin{vmatrix} c & a \\ f & d \end{vmatrix} \mathbf{j} + \begin{vmatrix} a & b \\ d & e \end{vmatrix} \mathbf{k}.$$

Therefore, to derive the cross product of two vectors we first write the vectors in the correct sequence. Remembering that $\mathbf{r} \times \mathbf{s}$ does not equal $\mathbf{s} \times \mathbf{r}$. Second, we compute the three scalar terms and form the resultant vector, which is perpendicular to the plane containing the original vectors.

So far, we have assumed that

$$\begin{aligned} \mathbf{r} \times \mathbf{s} &= \mathbf{t} \\ |\mathbf{t}| &= |\mathbf{r}||\mathbf{s}| \sin\theta \end{aligned}$$

where θ is the angle between $\mathbf{r}$ and $\mathbf{s}$, and $\mathbf{t}$ is perpendicular to the plane containing $\mathbf{r}$ and $\mathbf{s}$. Now let's prove that this is the case:

$$\begin{aligned} \mathbf{r} \cdot \mathbf{s} &= |\mathbf{r}||\mathbf{s}| \cos\theta = x_r x_s + y_r y_s + z_r z_s \\ \cos^2\theta &= \frac{(x_r x_s + y_r y_s + z_r z_s)^2}{|\mathbf{r}|^2 |\mathbf{s}|^2} \\ |\mathbf{t}| &= |\mathbf{r}||\mathbf{s}| \sin\theta \\ |\mathbf{t}|^2 &= |\mathbf{r}|^2 |\mathbf{s}|^2 \sin^2\theta \\ &= |\mathbf{r}|^2 |\mathbf{s}|^2 (1 - \cos^2\theta) \\ &= |\mathbf{r}|^2 |\mathbf{s}|^2 \Big(1 - \frac{(x_r x_s + y_r y_s + z_r z_s)^2}{|\mathbf{r}|^2 |\mathbf{s}|^2}\Big) \\ &= |\mathbf{r}|^2 |\mathbf{s}|^2 - (x_r x_s + y_r y_s + z_r z_s)^2 \\ &= (x_r^2 + y_r^2 + z_r^2)(x_s^2 + y_s^2 + z_s^2) - (x_r x_s + y_r y_s + z_r z_s)^2 \\ &= x_r^2 (y_s^2 + z_s^2) + y_r^2 (x_s^2 + z_s^2) + z_r^2 (x_s^2 + y_s^2) \\ &\quad - 2x_r x_s y_r y_s - 2x_r x_s z_r z_s - 2y_r y_s z_r z_s \\ &= x_r^2 y_s^2 + x_r^2 z_s^2 + y_r^2 x_s^2 + y_r^2 z_s^2 + z_r^2 x_s^2 + z_r^2 y_s^2 \\ &\quad - 2x_r x_s y_r y_s - 2x_r x_s z_r z_s - 2y_r y_s z_r z_s \\ &= (y_r z_s - z_r y_s)^2 + (z_r x_s - x_r z_s)^2 + (x_r y_s - y_r x_s)^2 \end{aligned}$$

which in determinant form is

$$|\mathbf{t}|^2 = \begin{vmatrix} y_r & z_r \\ y_s & z_s \end{vmatrix}^2 + \begin{vmatrix} z_r & x_r \\ z_s & x_s \end{vmatrix}^2 + \begin{vmatrix} x_r & y_r \\ x_s & y_s \end{vmatrix}^2$$

and confirms that $\mathbf{t}$ could be the vector

$$\mathbf{t} = \begin{vmatrix} y_r & z_r \\ y_s & z_s \end{vmatrix} \mathbf{i} + \begin{vmatrix} z_r & x_r \\ z_s & x_s \end{vmatrix} \mathbf{j} + \begin{vmatrix} x_r & y_r \\ x_s & y_s \end{vmatrix} \mathbf{k}.$$

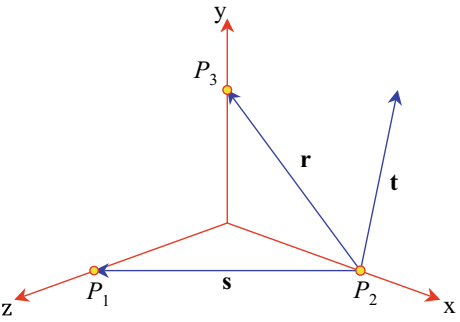

Fig. 11.10 Vector **t** is normal to the vectors **r** and **s**

Table 11.2 Coordinates of the vertices used in Fig. 11.10

Vertex	x	y	z
P_1	0	0	1
P_2	1	0	0
P_3	0	1	0

All that remains is to prove that **t** is orthogonal (perpendicular) to **r** and **s**, which is achieved by showing that $\mathbf{r} \cdot \mathbf{t} = \mathbf{s} \cdot \mathbf{t} = 0$:

$$
\begin{aligned}
\mathbf{r} &= x_r\mathbf{i} + y_r\mathbf{j} + z_r\mathbf{k} \\
\mathbf{s} &= x_s\mathbf{i} + y_s\mathbf{j} + z_s\mathbf{k} \\
\mathbf{t} &= (y_r z_s - z_r y_s)\mathbf{i} + (z_r x_s - x_r z_s)\mathbf{j} + (x_r y_s - y_r x_s)\mathbf{k} \\
\mathbf{r} \cdot \mathbf{t} &= x_r(y_r z_s - z_r y_s) + y_r(z_r x_s - x_r z_s) + z_r(x_r y_s - y_r x_s) \\
&= x_r y_r z_s - x_r y_s z_r + x_s y_r z_r - x_r y_r z_s + x_r y_s z_r - x_s y_r z_r = 0 \\
\mathbf{s} \cdot \mathbf{t} &= x_s(y_r z_s - z_r y_s) + y_s(z_r x_s - x_r z_s) + z_s(x_r y_s - y_r x_s) \\
&= x_s y_r z_s - x_s y_s z_r + x_s y_s z_r - x_r y_s z_s + x_r y_s z_s - x_s y_r z_s = 0
\end{aligned}
$$

and we have proved that $\mathbf{r} \times \mathbf{s} = \mathbf{t}$, where $|\mathbf{t}| = |\mathbf{r}||\mathbf{s}| \sin\theta$ and **t** is orthogonal to the plane containing **r** and **s**.

Let's now consider two vectors **r** and **s** and compute the normal vector **t**. The vectors are chosen so that we can anticipate approximately the answer. For the sake of clarity, the vector equations include the scalar multipliers 0 and 1. Normally, these are omitted. Figure 11.10 shows the vectors **r** and **s** and the normal vector **t**, and Table 11.2 contains the coordinates of the vertices forming the two vectors which confirms what we expected from Fig. 11.10.

$$
\begin{aligned}
\mathbf{r} &= [(x_3 - x_2) \quad (y_3 - y_2) \quad (z_3 - z_2)]^T \\
\mathbf{s} &= [(x_1 - x_2) \quad (y_1 - y_2) \quad (z_1 - z_2)]^T \\
P_1 &= (0, 0, 1) \\
P_2 &= (1, 0, 0) \\
P_3 &= (0, 1, 0) \\
\mathbf{r} &= -1\mathbf{i} + 1\mathbf{j} + 0\mathbf{k} \\
\mathbf{s} &= -1\mathbf{i} + 0\mathbf{j} + 1\mathbf{k} \\
\mathbf{r} \times \mathbf{s} &= [1 \times 1 - 0 \times 0]\mathbf{i} \\
&\quad - [-1 \times 1 - (-1) \times 0]\mathbf{j} \\
&\quad + [-1 \times 0 - (-1) \times 1]\mathbf{k} \\
\mathbf{t} &= \mathbf{i} + \mathbf{j} + \mathbf{k}
\end{aligned}
$$

Now let's reverse the vectors to illustrate the importance of vector sequence.

$$
\begin{aligned}
\mathbf{s} &= -1\mathbf{i} + 0\mathbf{j} + 1\mathbf{k} \\
\mathbf{r} &= -1\mathbf{i} + 1\mathbf{j} + 0\mathbf{k} \\
\mathbf{s} \times \mathbf{r} &= [0 \times 0 - 1 \times 1]\mathbf{i} \\
&\quad - [-1 \times 0 - (-1) \times 1]\mathbf{j} \\
&\quad + [-1 \times 1 - (-1) \times 0]\mathbf{k} \\
\mathbf{t} &= -\mathbf{i} - \mathbf{j} - \mathbf{k}
\end{aligned}
$$

which is in the opposite direction to $\mathbf{r} \times \mathbf{s}$ and confirms that the vector product is non-commutative.

11.4.10 The Right-Hand Rule

The *right-hand rule* is an *aide mémoire* for working out the orientation of the cross product vector. Given the operation $\mathbf{r} \times \mathbf{s}$, if the right-hand thumb is aligned with $\mathbf{r}$, the first finger with $\mathbf{s}$, and the middle finger points in the direction of $\mathbf{t}$. However, we must remember that this only holds in 3D. In 4D and above, it makes no sense.

11.5 Deriving a Unit Normal Vector for a Triangle

Figure 11.11 shows a triangle with vertices defined in an anticlockwise sequence from its visible side. This is the side from which we want the surface normal to point. Using the following information we will compute the surface normal using the cross product and then convert it to a unit normal vector.

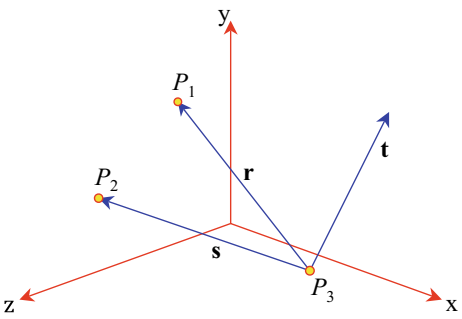

Fig. 11.11 The normal vector **t** is derived from the cross product $\mathbf{r} \times \mathbf{s}$

Create vector **r** between P_3 and P_1, and vector **s** between P_3 and P_2:

$$\begin{aligned}
\mathbf{r} &= -1\mathbf{i} + 1\mathbf{j} + 0\mathbf{k} \\
\mathbf{s} &= -1\mathbf{i} + 0\mathbf{j} + 2\mathbf{k} \\
\mathbf{r} \times \mathbf{s} &= (1 \times 2 - 0 \times 0)\mathbf{i} \\
&\quad - (-1 \times 2 - 0 \times -1)\mathbf{j} \\
&\quad + (-1 \times 0 - 1 \times -1)\mathbf{k} \\
\mathbf{t} &= 2\mathbf{i} + 2\mathbf{j} + \mathbf{k} \\
|\mathbf{t}| &= \sqrt{2^2 + 2^2 + 1^2} = 3 \\
\hat{\mathbf{t}}_u &= \tfrac{2}{3}\mathbf{i} + \tfrac{2}{3}\mathbf{j} + \tfrac{1}{3}\mathbf{k}.
\end{aligned}$$

The unit vector $\hat{\mathbf{t}}_u$ can now be used for illumination calculations in computer graphics, and as it has unit length, dot product calculations are simplified.

11.6 Surface Areas

Figure 11.12 shows two vectors **r** and **s**, where the height $h = |\mathbf{s}| \sin\theta$. Therefore the area of the associated parallelogram is

$$\text{area} = |\mathbf{r}|\ h = |\mathbf{r}||\mathbf{s}| \sin\theta.$$

But this is the magnitude of the cross product vector **t**. Thus when we calculate $\mathbf{r} \times \mathbf{s}$, the length of the normal vector **t** equals the area of the parallelogram formed by **r** and **s**; which means that the triangle formed by halving the parallelogram is half the area.

$$\begin{aligned}
\text{area of parallelogram} &= |\mathbf{t}| \\
\text{area of triangle} &= \tfrac{1}{2}|\mathbf{t}|.
\end{aligned}$$

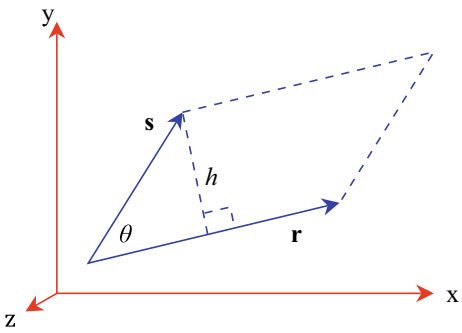

Fig. 11.12 The area of the parallelogram formed by two vectors **r** and **s**

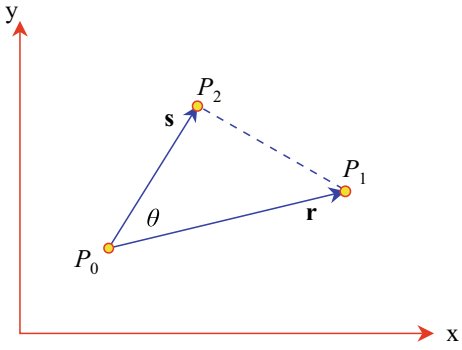

Fig. 11.13 The area of the triangle formed by the vectors **r** and **s**

This makes it relatively easy to calculate the surface area of an object constructed from triangles or parallelograms. In the case of a triangulated surface, we simply sum the magnitudes of the normals and halve the result.

11.6.1 Calculating 2D Areas

Figure 11.13 shows a triangle with vertices $P_0(x_0,\ y_0)$, $P_1(x_1,\ y_1)$ and $P_2(x_2,\ y_2)$ formed in an anti-clockwise sequence. The vectors **r** and **s** are computed as follows:

$$
\begin{aligned}
\mathbf{r} &= (x_1 - x_0)\mathbf{i} + (y_1 - y_0)\mathbf{j} \\
\mathbf{s} &= (x_2 - x_0)\mathbf{i} + (y_2 - y_0)\mathbf{j} \\
|\mathbf{r} \times \mathbf{s}| &= (x_1 - x_0)(y_2 - y_0) - (x_2 - x_0)(y_1 - y_0) \\
&= x_1(y_2 - y_0) - x_0(y_2 - y_0) - x_2(y_1 - y_0) + x_0(y_1 - y_0) \\
&= x_1y_2 - x_1y_0 - x_0y_2 + x_0y_0 - x_2y_1 + x_2y_0 + x_0y_1 - x_0y_0 \\
&= x_1y_2 - x_1y_0 - x_0y_2 - x_2y_1 + x_2y_0 + x_0y_1 \\
&= (x_0y_1 - x_1y_0) + (x_1y_2 - x_2y_1) + (x_2y_0 - x_0y_2).
\end{aligned}
$$

But the area of the triangle formed by the three vertices is $\frac{1}{2}|\mathbf{r} \times \mathbf{s}|$. Therefore

$$\text{area} = \tfrac{1}{2}[(x_0y_1 - x_1y_0) + (x_1y_2 - x_2y_1) + (x_2y_0 - x_0y_2)]$$

which is the formula disclosed in Chap. 9!

11.7 Worked Examples

11.7.1 Position Vector

Calculate the magnitude of the position vector $\mathbf{p}$, for the point $P(4,\ 5,\ 6)$.
Solution:

$$\mathbf{p} = [4 \quad 5 \quad 6]^T, \quad \text{therefore,} \quad |\mathbf{p}| = \sqrt{4^2 + 5^2 + 6^2} \approx 20.88.$$

11.7.2 Unit Vector

Convert $\mathbf{r} = [1 \quad 2 \quad 3]^T$ to a unit vector.
Solution:

$$\begin{aligned}
\mathbf{r} &= [1 \quad 2 \quad 3]^T \\
|\mathbf{r}| &= \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14} \\
\hat{\mathbf{r}} &= \frac{1}{\sqrt{14}}[1 \quad 2 \quad 3]^T \approx [0.267 \quad 0.535 \quad 0.802]^T.
\end{aligned}$$

11.7.3 Vector Magnitude

Given

$$\begin{aligned}
\mathbf{r} &= 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k} \\
\mathbf{s} &= 5\mathbf{i} + 6\mathbf{j} + 7\mathbf{k}
\end{aligned}$$

compute the magnitude of $\mathbf{r} + \mathbf{s}$.
Solution:

$$\begin{aligned}
\mathbf{r} + \mathbf{s} &= 7\mathbf{i} + 9\mathbf{j} + 11\mathbf{k} \\
|\mathbf{r} + \mathbf{s}| &= \sqrt{7^2 + 9^2 + 11^2} \approx 15.84.
\end{aligned}$$

11.7.4 Angle Between Two Vectors

Given

$$\mathbf{r} = [2 \quad 0 \quad 4]^T$$
$$\mathbf{s} = [5 \quad 6 \quad 10]^T$$

find the angle between $\mathbf{r}$ and $\mathbf{s}$.
Solution:

$$|\mathbf{r}| = \sqrt{2^2 + 0^2 + 4^2} \approx 4.472$$
$$|\mathbf{s}| = \sqrt{5^2 + 6^2 + 10^2} \approx 12.689.$$

Therefore,

$$|\mathbf{r}||\mathbf{s}| \cos\theta = 2 \times 5 + 0 \times 6 + 4 \times 10 = 50$$
$$12.689 \times 4.472 \times \cos\theta = 50$$
$$\cos\theta = \frac{50}{12.689 \times 4.472} \approx 0.8811$$
$$\theta = \arccos 0.8811 \approx 28.22^\circ.$$

The angle between the two vectors is approximately 28.22°.

11.7.5 Vector Product

Show that the vector product works with the unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$.
Solution: We start with

$$\mathbf{r} = 1\mathbf{i} + 0\mathbf{j} + 0\mathbf{k}$$
$$\mathbf{s} = 0\mathbf{i} + 1\mathbf{j} + 0\mathbf{k}$$

and then compute (11.7):

$$\mathbf{r} \times \mathbf{s} = (0 \times 0 - 0 \times 1)\mathbf{i} - (1 \times 0 - 0 \times 0)\mathbf{j} + (1 \times 1 - 0 \times 0)\mathbf{k}.$$

The $\mathbf{i}$ scalar and $\mathbf{j}$ scalar terms are both zero, but the $\mathbf{k}$ scalar term is 1, which makes $\mathbf{i} \times \mathbf{j} = \mathbf{k}$.

Let's see what happens when we reverse the vectors. This time we start with

$$\mathbf{r} = 0\mathbf{i} + 1\mathbf{j} + 0\mathbf{k}$$
$$\mathbf{s} = 1\mathbf{i} + 0\mathbf{j} + 0\mathbf{k}$$

and then compute (11.7)

$$\mathbf{r} \times \mathbf{s} = (1 \times 0 - 0 \times 0)\mathbf{i} - (0 \times 0 - 0 \times 1)\mathbf{j} + (0 \times 0 - 1 \times 1)\mathbf{k}.$$

The **i** scalar and **j** scalar terms are both zero, but the **k** scalar term is -1, which makes $\mathbf{j} \times \mathbf{i} = -\mathbf{k}$. So we see that the vector product is *antisymmetric*, i.e. there is a sign reversal when the vectors are reversed. Similarly, it can be shown that

$$\begin{aligned}
\mathbf{j} \times \mathbf{k} &= \mathbf{i} \\
\mathbf{k} \times \mathbf{i} &= \mathbf{j} \\
\mathbf{k} \times \mathbf{j} &= -\mathbf{i} \\
\mathbf{i} \times \mathbf{k} &= -\mathbf{j}.
\end{aligned}$$

Reference

Crowe MJ, A history of vector analysis. Dover, Illinois. ISBN 9780486679105

Chapter 12
Complex Numbers

12.1 Introduction

In this chapter I review complex numbers. The complex plane is described as a way of visualising complex numbers and various algebraic operations, and two functions for isolating the real and imaginary parts of a complex expression. The section on Complex Algebra examines topics such as the complex conjugate, powers of i, complex exponentials, logarithms of a complex number, and the hyperbolic functions. Finally, there are a dozen worked examples.

Readers interested in the history of complex numbers are invited to read the author's book *Imaginary Mathematics for Computer Science*, Vince (2018) where they will discover the problems associated with the choice of the words *imaginary* and *complex*. For this chapter, the reader should forget the every-day meaning of imaginary, and regard $i = \sqrt{-1}$ as a mechanism for dividing a mathematical expression into two parts.

12.2 Representing Complex Numbers

This section explores various ways of representing complex numbers numerically and graphically.

12.2.1 Complex Numbers

Numbers such as $a + bi$ form the set of complex numbers $\mathbb{C}$, where

$$a, b \in \mathbb{R}, \quad i^2 = -1.$$

Examples are $2i, \quad 23 - 12i, \quad 3 + x^2 i.$

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_12

12.2.2 Real and Imaginary Parts

The real and imaginary parts of a complex number z are isolated by the Re and Im functions. For example:

$$\begin{aligned} z &= a + bi \\ a &= \mathrm{Re}(z) \\ b &= \mathrm{Im}(z). \end{aligned}$$

These two functions permit one to construct formal algebraic definitions such as defining one complex number being equal to another. In words, one would say "two complex numbers are equal iff (*if and only if*) they have identical real **and** imaginary parts". e.g. given $z_1 = x_1 + y_1 i$ and $z_2 = x_2 + y_2 i$, then $z_1 = z_2$ iff $x_1 = x_2$ **and** $y_1 = y_2$. Using Re and Im, we can write:

$$z_1 = z_2 \quad \leftrightarrow \quad [\mathrm{Re}(z_1) = \mathrm{Re}(z_2) \wedge \mathrm{Im}(z_1) = \mathrm{Im}(z_2)]\,.$$

12.2.3 The Complex Plane

When the real number line is combined with a vertical imaginary axis, it creates the *complex plane*, as shown in Fig. 12.1, where any complex number has a unique position. Figure 12.2 shows the positions of the following four complex numbers

$$P = 4 + 3i, \quad Q = -3 + 2i, \quad R = -3 - 3i, \quad S = 4 - 5i.$$

12.3 Complex Algebra

This section reviews the axioms associated with complex algebra, the complex conjugate, complex division, powers of i, the rotational properties of i, polar notation, the complex norm, the complex inverse, complex exponentials, the roots and logarithms of a complex number, hyperbolic functions, and the role of the complex plane in visualising complex functions.

12.3.1 Algebraic Laws

Complex numbers obey the axioms associated with real numbers. But for clarity, examples are included to show how imaginary terms are resolved.

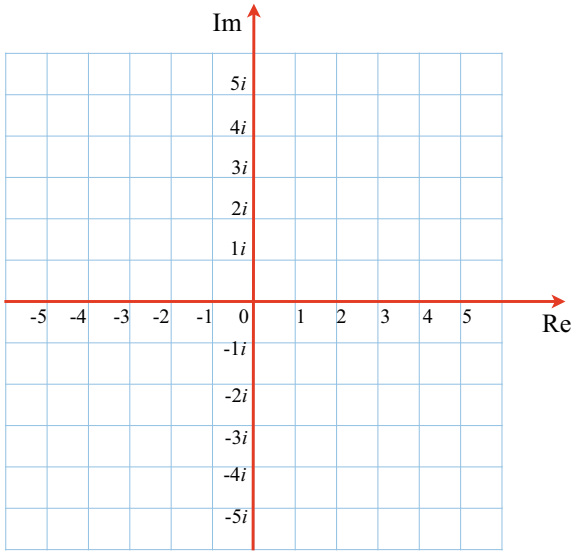

Fig. 12.1 The complex plane

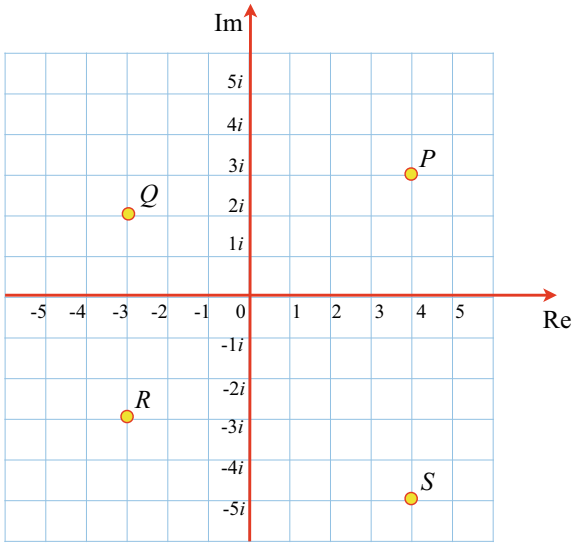

Fig. 12.2 Four complex numbers

Given

$$z_1 = x_1 + y_1 i$$
$$z_2 = x_2 + y_2 i$$
$$z_3 = x_3 + y_3 i.$$

The commutative law of addition:

$$z_1 + z_2 = z_2 + z_1 = (x_1 + x_2) + (y_1 + y_2)i$$
$$\text{e.g. } (2 + 3i) + (4 + 5i) = 6 + 8i.$$

The commutative law of multiplication:

$$z_1 z_2 = z_2 z_1$$
$$\begin{aligned}\text{e.g. } (2 + 3i)(4 + 5i) &= 8 + 10i + 12i + 15i^2 \\ &= -7 + 22i.\end{aligned}$$

The associative law of addition:

$$z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3 = z_1 + z_2 + z_3$$
$$\text{e.g. } (2 + 3i) + (4 + 5i) + (6 + 7i) = 12 + 15i.$$

The associative law of multiplication:

$$z_1(z_2 z_3) = (z_1 z_2)z_3 = z_1 z_2 z_3$$
$$\begin{aligned}\text{e.g. } (2 + 3i)(4 + 5i)(6 + 7i) &= (8 + 22i + 15i^2)(6 + 7i) \\ &= (-7 + 22i)(6 + 7i) \\ &= -42 + 132i - 49i + 154i^2 \\ &= -196 + 83i.\end{aligned}$$

The distributive law of multiplication:

$$z_1(z_2 + z_3) = z_1 z_2 + z_1 z_3$$
$$\begin{aligned}\text{e.g. } (2 + 3i)\,[(4 + 5i) + (6 + 7i)] &= (2 + 3i)(10 + 12i) \\ &= 20 + 30i + 24i + 36i^2 \\ &= -16 + 54i.\end{aligned}$$

From the above, one can see that the addition of complex numbers is identical to the addition of vectors. Figure 12.3 illustrates the addition of $z_1 = -2 + 3i$ and $z_2 = 3 + i$.

12.3.2 Complex Conjugate

The *complex conjugate* is a useful algebraic construct and is denoted by $\bar{z}$ or z^*. To avoid confusion, I will use $\bar{z}$.

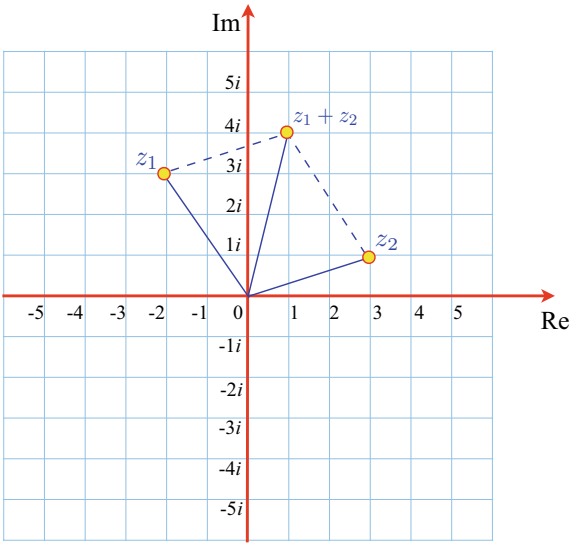

Fig. 12.3 The addition of $z_1 + z_2$

Given $z = a + bi$, then $\bar{z} = a - bi$. Also

$$\bar{z} = \text{Re}(z) - \text{Im}(z)i.$$

The product $z\bar{z}$ is extremely useful, as it is a real quantity. Generally,

$$\begin{aligned} z &= a + bi \\ \bar{z} &= a - bi \\ z\bar{z} &= (a + bi)(a - bi) \\ &= a^2 - abi + abi - b^2 i^2 \\ &= a^2 + b^2 \end{aligned}$$

which is real. For example,

$$\begin{aligned} z &= 3 + 4i \\ \bar{z} &= 3 - 4i \\ z\bar{z} &= 25. \end{aligned}$$

Figure 12.4 shows z and $\bar{z}$.

Let's prove that $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$. Given

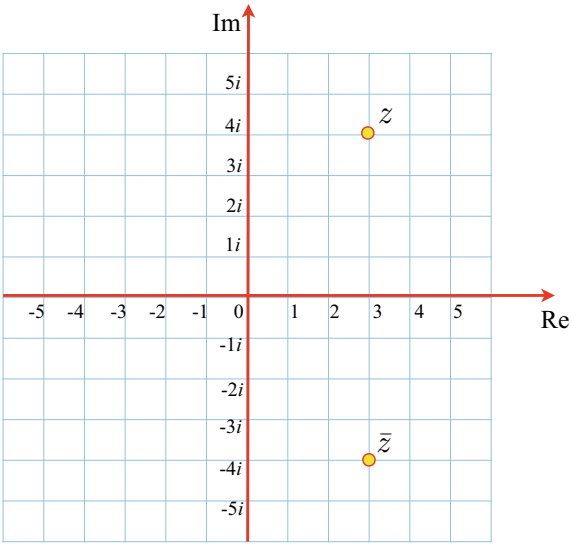

Fig. 12.4 A complex number and its complex conjugate

$$
\begin{aligned}
z_1 &= a_1 + b_1 i \\
z_2 &= a_2 + b_2 i \\
\overline{z_1 z_2} &= \overline{(a_1 + b_1 i)(a_2 + b_2 i)} \\
&= a_1 \overline{(a_2 + b_2 i)} + \overline{b_1 i (a_2 + b_2 i)} \\
&= a_1 (a_2 - b_2 i) - b_1 i (a_2 - b_2 i) \\
&= (a_1 - b_1 i)(a_2 - b_2 i) \\
&= \overline{z}_1 \overline{z}_2 .
\end{aligned}
$$

12.3.3 Complex Division

The complex conjugate is useful in resolving the quotient of two complex numbers; for if we multiply the numerator and the denominator by the complex conjugate of the denominator, the denominator becomes a real quantity and simplifies the division. For example, we evaluate this quotient as follows

$$
\begin{aligned}
z &= \frac{10 + 5i}{1 + 2i} \\
&= \frac{(10 + 5i)}{(1 + 2i)} \frac{(1 - 2i)}{(1 - 2i)} \\
&= \frac{(10 + 5i)(1 - 2i)}{5}
\end{aligned}
$$

$$
\begin{aligned}
&= (2+i)(1-2i) \\
&= 2 - 4i + i - 2i^2 \\
&= 4 - 3i.
\end{aligned}
$$

12.3.4 Powers of i

As $i^2 = -1$, it must be possible to raise i to other powers. For example,

$$i^4 = i^2 i^2 = 1$$

and

$$i^5 = i i^4 = i.$$

Table 12.1 shows the sequence up to i^6.

This cyclic pattern is quite striking, and reminds one of:

$$x,\ y,\ -x,\ -y,\ x,\ \ldots$$

that arises when rotating around the Cartesian axes in a anti-clockwise direction. The above sequence is summarised as

$$
\left.
\begin{aligned}
i^{4n} &= 1 \\
i^{4n+1} &= i \\
i^{4n+2} &= -1 \\
i^{4n+3} &= -i
\end{aligned}
\right\} \quad \text{where } n \in \mathbb{N}^0.
$$

But what about negative powers? Well they too, are also possible. Consider i^{-1}, which is evaluated as follows

$$i^{-1} = \frac{1}{i} = \frac{1(-i)}{i(-i)} = \frac{-i}{1} = -i.$$

Similarly,

Table 12.1 Increasing powers of i

i^0	i^1	i^2	i^3	i^4	i^5	i^6
1	i	-1	$-i$	1	i	-1

Table 12.2 Decreasing powers of i

i^0	i^{-1}	i^{-2}	i^{-3}	i^{-4}	i^{-5}	i^{-6}
1	$-i$	-1	i	1	$-i$	-1

$$i^{-2} = \frac{1}{i^2} = \frac{1}{-1} = -1$$

and

$$i^{-3} = i^{-1}i^{-2} = -i(-1) = i.$$

Table 12.2 shows the sequence down to i^{-6}. This time the cyclic pattern is reversed and is similar to

$$x,\ -y,\ -x,\ y,\ x,\ \ldots$$

that arises when rotating around the Cartesian axes in a clockwise direction.

12.3.5 Rotational Qualities of i

Now let's investigate how a real number behaves when it is repeatedly multiplied by i. Starting with the number 3, we have,

$$\begin{aligned} i \times 3 &= 3i \\ i \times 3i &= -3 \\ i \times (-3) &= -3i \\ i \times (-3)i &= 3. \end{aligned}$$

The cycle is $3,\ 3i,\ -3,\ -3i,\ 3,\ 3i,\ -3,\ -3i,\ 3,\ \ldots$ which has four steps, as shown in Fig. 12.5.

If we multiply a complex number by i, it is also rotated 90°. For example, the complex number $P = 4 + 3i$ in Fig. 12.6 is rotated 90° to Q by multiplying it by i,

$$\begin{aligned} i(4 + 3i) &= 4i + 3i^2 \\ &= 4i - 3 \\ &= -3 + 4i. \end{aligned}$$

The point $Q = -3 + 4i$ is rotated 90° to R by multiplying it by i,

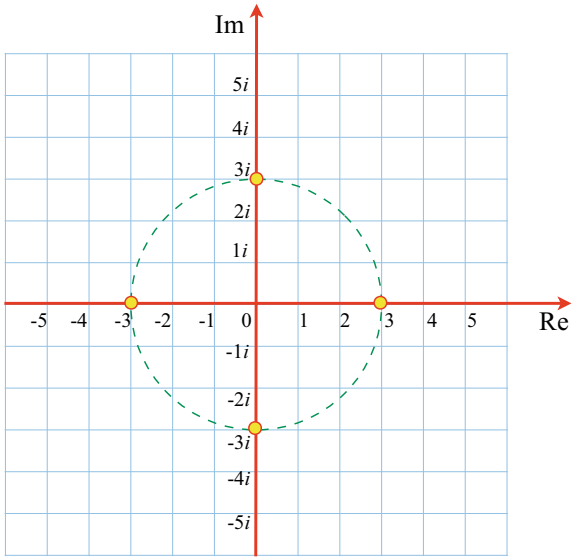

Fig. 12.5 The cycle of points created by repeatedly multiplying 3 by i

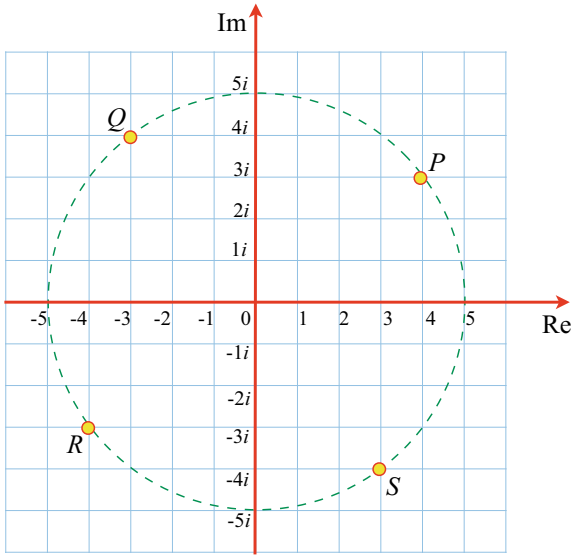

Fig. 12.6 Multiplying a complex number by i

$$\begin{aligned} i(-3+4i) &= -3i + 4i^2 \\ &= -3i - 4 \\ &= -4 - 3i. \end{aligned}$$

The point $R = -4 - 3i$ is rotated 90° to S by multiplying it by i,

$$\begin{aligned} i(-4-3i) &= -4i - 3i^2 \\ &= -4i + 3 \\ &= 3 - 4i. \end{aligned}$$

Finally, the point $S = 3 - 4i$ is rotated 90° back to P by multiplying it by i,

$$\begin{aligned} i(3-4i) &= 3i - 4i^2 \\ &= 3i + 4 \\ &= 4 + 3i. \end{aligned}$$

As you can see, complex numbers are intimately related to Cartesian coordinates, in that the ordered pair $(x,\ y) \equiv a + bi$.

12.3.6 Modulus and Argument

As a complex number has a unique position on the complex plane, and is always relative to the origin of the real and imaginary axes, it can be visualised as a position vector and assigned a *modulus* or *magnitude*, which is the distance r of the complex point to the origin; consequently, $z = a + bi$ has a modulus $r = \sqrt{a^2 + b^2}$ and notated as $|z| = \sqrt{a^2 + b^2}$. This can also be expressed as

$$|z|^2 = a^2 + b^2 = [\mathrm{Re}(z)]^2 + [\mathrm{Im}(z)]^2 .$$

Here are some useful relationships:

$$\begin{aligned} z\bar{z} &= (a + bi)(a - bi) = a^2 + b^2 \\ z\bar{z} &= |z|^2 \\ |z| &= |\bar{z}| \\ |-z| &= |z| \\ |z_1 z_2| &= |z_1||z_2| \\ |z_1 z_2|^2 &= |z_1|^2 |z_2|^2 \end{aligned}$$

Pursuing the similarity between complex numbers and position vectors, the straight line from the origin to a complex number $z = a + bi$, subtends with the real axis an angle θ, called the *argument*; consequently, $\theta = \tan^{-1}(b/a)$, and is notated as $\arg(z) = \tan^{-1}(b/a)$. Figure 12.7 shows the complex number $z = 3 + 4i$ with a modulus $r = \sqrt{3^2 + 4^2} = 5$ and an argument $\theta = \tan^{-1}(4/3) \approx 53.125°$.

From Fig. 12.7 we can generalise that a complex number $z = a + bi$ has real and imaginary components,

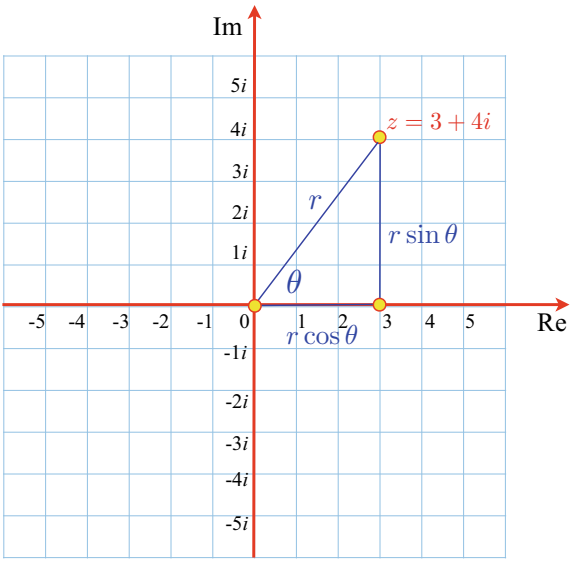

Fig. 12.7 The argument θ and modulus r of a complex number

$$a = r\cos\theta$$
$$b = r\sin\theta$$

which permits us to state

$$z = r\cos\theta + ir\sin\theta$$

and when $r = 1$,

$$z = \cos\theta + i\sin\theta.$$

For example, given two complex numbers,

$$z_1 = a_1 + b_1 i$$
$$z_2 = a_2 + b_2 i$$

where

$$a_1 = r_1\cos\theta_1, \quad b_1 = r_1\sin\theta_1$$
$$a_2 = r_2\cos\theta_2, \quad b_2 = r_2\sin\theta_2$$

then

$$\begin{aligned} z_1 z_2 &= (a_1 + b_1 i)(a_2 + b_2 i) \\ &= (a_1 a_2 - b_1 b_2) + (a_1 b_2 + b_1 a_2)i \\ &= (r_1\cos\theta_1 \cdot r_2\cos\theta_2 - r_1\sin\theta_1 \cdot r_2\sin\theta_2) \end{aligned} \tag{12.1}$$

$$
\begin{aligned}
&+ (r_1 \cos\theta_1 \cdot r_2 \sin\theta_2 + r_1 \sin\theta_1 \cdot r_2 \cos\theta_2)i \\
&= r_1 r_2(\cos\theta_1 \cdot \cos\theta_2 \cdot \sin\theta_1 \cdot \sin\theta_2) + i r_1 r_2(\cos\theta_1 \cdot \sin\theta_2 + \sin\theta_1 \cdot \cos\theta_2) \\
&= r_1 r_2 \cos(\theta_1 + \theta_2) + i r_1 r_2 \sin(\theta_1 + \theta_2) \qquad (12.2)
\end{aligned}
$$

which shows that to compute the product of two complex numbers, we multiply their moduli and add their arguments. Let's illustrate this operation with an example. We'll start by computing the product using (12.1).

Given

$$
\begin{aligned}
z_1 &= \tfrac{1}{2} + \tfrac{\sqrt{3}}{2}i \\
z_2 &= -\tfrac{1}{2} + \tfrac{\sqrt{3}}{2}i \\
z_1 z_2 &= \tfrac{1}{2}\left(-\tfrac{1}{2}\right) - \tfrac{\sqrt{3}}{2}\tfrac{\sqrt{3}}{2} + \left[\tfrac{1}{2}\tfrac{\sqrt{3}}{2} + \tfrac{\sqrt{3}}{2}\left(-\tfrac{1}{2}\right)\right] i \\
&= -\tfrac{1}{4} - \tfrac{3}{4} + \left(\tfrac{\sqrt{3}}{4} - \tfrac{\sqrt{3}}{4}\right) i \\
&= -1.
\end{aligned}
$$

Next using (12.2). But first, we need to compute the moduli and arguments:

$$
\begin{aligned}
r_1 &= \sqrt{\left(\tfrac{1}{2}\right)^2 + \left(\tfrac{\sqrt{3}}{2}\right)^2} = 1 \\
r_2 &= \sqrt{\left(-\tfrac{1}{2}\right)^2 + \left(\tfrac{\sqrt{3}}{2}\right)^2} = 1 \\
\theta_1 &= \tan^{-1}\left(\tfrac{\sqrt{3}}{2}\tfrac{2}{1}\right) = 60^\circ \\
\theta_2 &= \tan^{-1}\left(-\tfrac{\sqrt{3}}{2}\tfrac{2}{1}\right) = 120^\circ \\
z_1 z_2 &= \cos(60^\circ + 120^\circ) + i \sin(60^\circ + 120^\circ) \\
&= -1.
\end{aligned}
$$

Naturally, the results are the same.

12.3.7 Complex Norm

The term *norm* causes a lot of confusion, simply because there are so many, and each one requires a precise definition. For our purposes, norms are associated with vector spaces, where the norm of a vector is a function that returns some numerical property of the vector. The Euclidean norm of vector $\mathbf{v}$, is generally written

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

which is the square-root of the inner product of the vector with itself. For example, if vector $\mathbf{v} = [3 \quad 4]$, then $\|\mathbf{v}\| = \sqrt{3^2 + 4^2} = 5$, and represents the Euclidean length of the vector.

The absolute value of a signed number $\pm x$ is written $|x|$. For example, if $x = +23$, $|x| = 23$, and if $x = -23$, $|x| = 23$. The absolute-value norm $\|x\|$, equals the absolute value, i.e. $\|x\| = |x|$.

The Euclidean norm of a complex number $z = a + bi$ is given by

$$\|z\| = |z| = \sqrt{a^2 + b^2}$$

which is the modulus of z.

The modulus or Euclidean norm of a complex number measures an abstract distance corresponding to the length of the complex point to the origin on the complex plane, and is normally expressed:

$$\|z\| = \sqrt{z\overline{z}} = \sqrt{(a + bi)(a - bi)} = \sqrt{a^2 + b^2}.$$

Let's prove that the Euclidean norm of the product of two complex numbers, equals the product of the individual Euclidean norms.

$$\begin{aligned} z_1 &= a_1 + b_1 i = r_1, \theta_1 \\ z_2 &= a_2 + b_2 i = r_2, \theta_2 \\ \|z_1 z_2\| &= |z_1 z_2| \\ &= |z_1| \cdot |z_2| \\ &= \|z_1\| \cdot \|z_2\|. \end{aligned}$$

12.3.8 Complex Inverse

We have already seen that to divide a complex number x by another z, we multiply the numerator and denominator by the conjugate of the denominator:

$$\frac{x}{z} = \frac{a + bi}{c + di} \frac{c - di}{c - di}$$

which can be written as

$$xz^{-1} = x\frac{\overline{z}}{|z|^2}$$

thus the inverse of a complex number is

$$z^{-1} = \frac{\overline{z}}{|z|^2}.$$

For example, the inverse of $3 + 4i$ is

$$(3 + 4i)^{-1} = \tfrac{1}{25}(3 - 4i).$$

12.3.9 Complex Exponentials

In order to describe complex exponentials we require three power series. We start with the power series for e^{θ}, $\sin\theta$ and $\cos\theta$,

$$\begin{aligned}
e^{\theta} &= 1 + \frac{\theta^1}{1!} + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{\theta^5}{5!} + \frac{\theta^6}{6!} + \frac{\theta^7}{7!} + \frac{\theta^8}{8!} + \frac{\theta^9}{9!} + \cdots \\
\sin\theta &= \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \frac{\theta^9}{9!} + \cdots \\
\cos\theta &= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \frac{\theta^8}{8!} + \cdots .
\end{aligned}$$

These series explain why the radian of angular measurement is used, for when $\theta = \pi$, $\sin\theta = 0$, and $\cos\theta = -1$. Figure 12.8 shows curves of the sine power series for 3, 5, 7 and 9 terms.

Euler discovered that by making θ imaginary: $e^{i\theta}$, we have

$$\begin{aligned}
e^{i\theta} &= 1 + \frac{i\theta^1}{1!} - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} - \frac{\theta^6}{6!} - \frac{i\theta^7}{7!} + \frac{\theta^8}{8!} + \frac{i\theta^9}{9!} \cdots \\
&= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \frac{\theta^8}{8!} + \cdots + \frac{i\theta^1}{1!} - \frac{i\theta^3}{3!} + \frac{i\theta^5}{5!} - \frac{i\theta^7}{7!} + \frac{i\theta^9}{9!} + \cdots
\end{aligned}$$

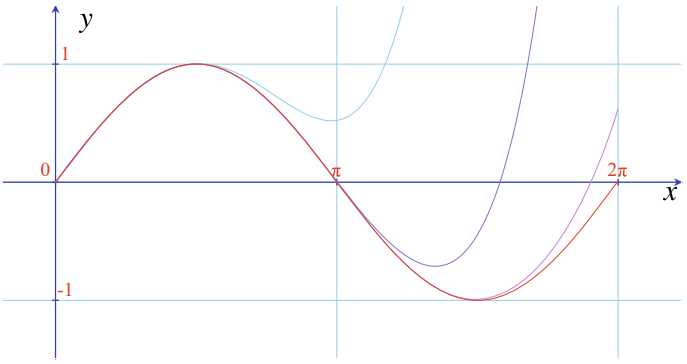

Fig. 12.8 The sine power series for different number of terms

$$
\begin{aligned}
&= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \frac{\theta^8}{8!} + \cdots + i\left(\frac{\theta^1}{1!} - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \frac{\theta^9}{9!} + \cdots\right) \\
&= \cos\theta + i\sin\theta
\end{aligned}
$$

which is *Euler's trigonometric formula*. If we now reverse the sign of $i\theta$ to $-i\theta$, we have

$$
\begin{aligned}
\mathrm{e}^{-i\theta} &= 1 - \frac{i\theta^1}{1!} - \frac{\theta^2}{2!} + \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} - \frac{i\theta^5}{5!} - \frac{\theta^6}{6!} + \frac{i\theta^7}{7!} + \frac{\theta^8}{8!} - \frac{i\theta^9}{9!}\cdots \\
&= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \frac{\theta^8}{8!} + \cdots - \frac{i\theta^1}{1!} + \frac{i\theta^3}{3!} - \frac{i\theta^5}{5!} + \frac{i\theta^7}{7!} - \frac{i\theta^9}{9!} + \cdots \\
&= 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \frac{\theta^8}{8!} + \cdots - i\left(\frac{\theta^1}{1!} - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \frac{\theta^9}{9!} + \cdots\right) \\
&= \cos\theta - i\sin\theta
\end{aligned}
$$

thus we have

$$
\begin{aligned}
\mathrm{e}^{i\theta} &= \cos\theta + i\sin\theta \\
\mathrm{e}^{-i\theta} &= \cos\theta - i\sin\theta
\end{aligned}
$$

from which we obtain

$$
\begin{aligned}
\cos\theta &= \frac{\mathrm{e}^{i\theta} + \mathrm{e}^{-i\theta}}{2} \\
\sin\theta &= \frac{\mathrm{e}^{i\theta} - \mathrm{e}^{-i\theta}}{2i}.
\end{aligned}
$$

Given $\mathrm{e}^{i\theta} = \cos\theta + i\sin\theta$, when $\theta = \pi$, we have $\mathrm{e}^{i\pi} = -1$, or $\mathrm{e}^{i\pi} + 1 = 0$, which is Euler's famous equation. The American physicist Richard Feynman (1918–1988) referred to the equation as "our jewel" and "the most remarkable formula in mathematics (Feynman 1977)."

Another strange formula emerges as follows:

$$
\begin{aligned}
\cos\theta + i\sin\theta &= \mathrm{e}^{i\theta} \\
\cos\left(\tfrac{\pi}{2}\right) + i\sin\left(\tfrac{\pi}{2}\right) &= \mathrm{e}^{i\pi/2} \\
i &= \mathrm{e}^{i\pi/2} \\
i^i &= \left(\mathrm{e}^{i\pi/2}\right)^i \\
&= \mathrm{e}^{i^2\pi/2} \\
&= \mathrm{e}^{-\pi/2} \\
i^i &= 0.207\ 879\ 576\ldots
\end{aligned}
$$

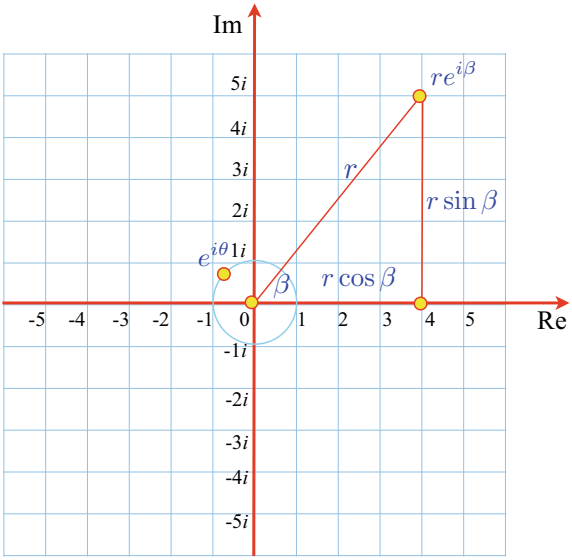

Fig. 12.9 The unit circle and $re^{i\beta}$

which reveals that i^i is a real number, even though i is not a number, as we know it!

Geometrically, $e^{i\theta}$ is a point on the unit circle, on the complex plane. Consequently, $re^{i\beta}$ is another point, radius r from the origin, with real and imaginary coordinates $x = r\cos\beta$ and $y = r\sin\beta$, respectively, as shown in Fig. 12.9. This is the polar form of a complex number.

Let's return to the product of two complex numbers, and see how the product can be visualised using polar notation.

Equation (12.2) shows that

$$\begin{aligned} z_1 &= r_1(\cos\theta_1 + i\sin\theta_1) \\ z_2 &= r_2(\cos\theta_2 + i\sin\theta_2) \\ z_1 z_2 &= r_1 r_2\left(\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)\right). \end{aligned}$$

Using polar notation:

$$\begin{aligned} z_1 &= r_1 e^{i\theta_1} \\ z_2 &= r_2 e^{i\theta_2} \\ z_1 z_2 &= r_1 e^{i\theta_1} r_2 e^{i\theta_2} \\ &= r_1 r_2 e^{i(\theta_1+\theta_2)}. \end{aligned}$$

Now let's show three ways of computing the product of the following three complex numbers shown in Fig. 12.10:

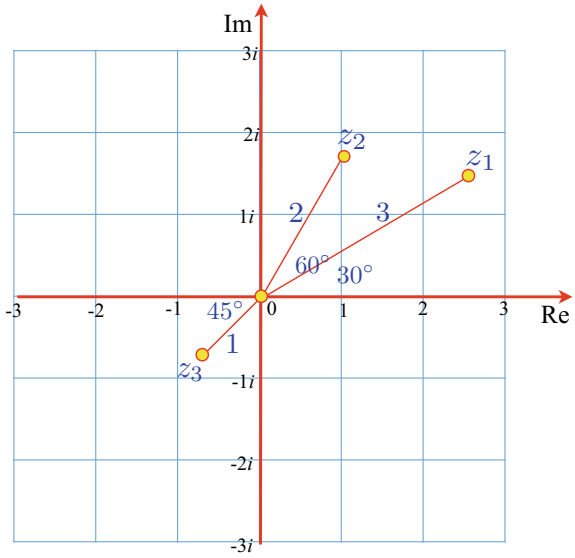

Fig. 12.10 Three complex numbers

$$z_1 = \tfrac{3\sqrt{3}}{2} + \tfrac{3}{2}i$$
$$z_2 = 1 + \sqrt{3}i$$
$$z_3 = -\tfrac{\sqrt{2}}{2} - \tfrac{\sqrt{2}}{2}i.$$

z_1: the argument is 30° and a modulus of 3.
z_2: the argument is 60° and a modulus of 2.
z_3: the argument is 225° and a modulus of 1.

Let's compute the product $z_1 z_2 z_3$

$$\begin{aligned} z_1 z_2 z_3 &= \left(\tfrac{3\sqrt{3}}{2} + \tfrac{3}{2}i\right)\left(1 + \sqrt{3}i\right)\left(-\tfrac{\sqrt{2}}{2} - \tfrac{\sqrt{2}}{2}i\right) \\ &= \left(\tfrac{3\sqrt{3}}{2} + \tfrac{9}{2}i + \tfrac{3}{2}i - \tfrac{3\sqrt{3}}{2}\right)\left(-\tfrac{\sqrt{2}}{2} - \tfrac{\sqrt{2}}{2}i\right) \\ &= 6i\left(-\tfrac{\sqrt{2}}{2} - \tfrac{\sqrt{2}}{2}i\right) \\ &= 3\sqrt{2} - 3\sqrt{2}i \end{aligned}$$

which confirms that the product $z_1 z_2 z_3$ rotates any complex number 315°, and scales its modulus by 6.

Now let's compute the product using cosines and sines.

$$z_1 = 3(\cos 30° + i \sin 30°)$$
$$z_2 = 2(\cos 60° + i \sin 60°)$$

$$\begin{aligned} z_3 &= \cos 225^\circ + i \sin 225^\circ \\ z_1 z_2 z_3 &= 3(\cos 30^\circ + i \sin 30^\circ) 2(\cos 60^\circ + i \sin 60^\circ)(\cos 225^\circ + i \sin 225^\circ) \\ &= 6(\cos 315^\circ + i \sin 315^\circ) \\ &= 3\sqrt{2} - 3\sqrt{2}i \end{aligned}$$

which is much simpler. Finally, let's define the complex numbers in polar form, with angles in degrees, for clarity.

$$\begin{aligned} z_1 &= 3\mathrm{e}^{i30^\circ} \\ z_2 &= 2\mathrm{e}^{i60^\circ} \\ z_3 &= \mathrm{e}^{i225^\circ} \\ z_1 z_2 z_3 &= 3\mathrm{e}^{i30^\circ} 2\mathrm{e}^{i60^\circ} \mathrm{e}^{i225^\circ} \\ &= 6\mathrm{e}^{315^\circ} \\ &= 6(\cos 315^\circ + i \sin 315^\circ) \\ &= 3\sqrt{2} - 3\sqrt{2}i \end{aligned}$$

which is even neater!

12.3.10 de Moivre's Theorem

Euler's trigonometric formula can be developed as follows.

$$\begin{aligned} \cos\theta + i \sin\theta &= \mathrm{e}^{i\theta} \\ (\cos\theta + i \sin\theta)^n &= \left(\mathrm{e}^{i\theta}\right)^n = \mathrm{e}^{in\theta} \\ (\cos\theta + i \sin\theta)^n &= \cos(n\theta) + i \sin(n\theta) \end{aligned} \tag{12.3}$$

where (12.3) is known as *de Moivre's theorem*, after Abraham de Moivre.

Substituting $n = 2$ in (12.3) we obtain

$$\begin{aligned} \cos(2\theta) + i \sin(2\theta) &= (\cos\theta + i \sin\theta)^2 \\ &= \cos^2\theta - \sin^2\theta + 2i \cos\theta \cdot \sin\theta. \end{aligned}$$

Therefore,

$$\begin{aligned} \cos(2\theta) &= \mathrm{Re}\left(\cos^2\theta - \sin^2\theta + 2i \cos\theta \cdot \sin\theta\right) \\ &= \cos^2\theta - \sin^2\theta \\ \sin(2\theta) &= \mathrm{Im}\left(\cos^2\theta - \sin^2\theta + 2i \cos\theta \cdot \sin\theta\right) \\ &= 2 \cos\theta \cdot \sin\theta. \end{aligned}$$

de Moivre's theorem can be used for similar identities by substituting other values of n. Let's try $n = 3$:

$$\begin{aligned}\cos(3\theta) + i\sin(3\theta) &= (\cos\theta + i\sin\theta)^3 \\ &= (\cos\theta + i\sin\theta)(\cos^2\theta - \sin^2\theta + 2i\cos\theta\cdot\sin\theta).\end{aligned}$$

Therefore,

$$\begin{aligned}\cos(3\theta) &= \operatorname{Re}\left[(\cos\theta + i\sin\theta)(\cos^2\theta - \sin^2\theta + 2i\cos\theta\cdot\sin\theta)\right] \\ &= \cos^3\theta - \cos\theta\cdot\sin^2\theta - 2\cos\theta\cdot\sin^2\theta \\ &= \cos^3\theta - 3\cos\theta\cdot\sin^2\theta \\ &= \cos^3\theta - 3\cos\theta(1 - \cos^2\theta) \\ &= 4\cos^3\theta - 3\cos\theta. \\ \sin(3\theta) &= \operatorname{Im}\left[(\cos\theta + i\sin\theta)(\cos^2\theta - \sin^2\theta + 2i\cos\theta\cdot\sin\theta)\right] \\ &= \cos^2\theta\cdot\sin\theta - \sin^3\theta + 2\cos^2\theta\cdot\sin\theta \\ &= 3\cos^2\theta\cdot\sin\theta - \sin^3\theta \\ &= 3\sin\theta(1 - \sin^2\theta) - \sin^3\theta \\ &= 3\sin\theta - 4\sin^3\theta.\end{aligned}$$

Let's test these identities with $\theta = 30°$:

$$\begin{aligned}\cos(3\theta) &= 4\cos^3\theta - 3\cos\theta \\ \cos 90° &= 4\cos^3 30° - 3\cos 30° \\ &= 4\left(\tfrac{\sqrt{3}}{2}\right)^3 - 3\tfrac{\sqrt{3}}{2} \\ &= \tfrac{3}{2}\sqrt{3} - \tfrac{3}{2}\sqrt{3} \\ &= 0. \\ \sin(3\theta) &= 3\sin\theta - 4\sin^3\theta \\ \sin 90° &= 3\sin 30° - 4\sin^3 30° \\ &= \tfrac{3}{2} - 4\left(\tfrac{1}{2}\right)^3 \\ &= 1.\end{aligned}$$

Given $z = \cos\theta + i\sin\theta$, we can define $\cos\theta$ and $\sin\theta$ in terms of z as follows.

$$z = \cos\theta + i\sin\theta \tag{12.4}$$
$$= e^{i\theta}$$
$$\tfrac{1}{z} = e^{-i\theta}$$
$$= \cos\theta - i\sin\theta \tag{12.5}$$

adding and subtracting (12.4) and (12.5):

$$z + \tfrac{1}{z} = 2\cos\theta$$
$$z - \tfrac{1}{z} = 2i\sin\theta$$
$$\cos\theta = \tfrac{1}{2}\left(z + \tfrac{1}{z}\right) \tag{12.6}$$
$$\sin\theta = -i\tfrac{1}{2}\left(z - \tfrac{1}{z}\right). \tag{12.7}$$

Let's use de Moivre's formula to show that

$$z^n + \frac{1}{z^n} = 2\cos(n\theta).$$

Proof:

$$z^n = \cos(n\theta) + i\sin(n\theta)$$
$$z^{-n} = \cos(-n\theta) + i\sin(-n\theta)$$
$$= \cos(n\theta) - i\sin(n\theta)$$
$$z^n + z^{-n} = 2\cos(n\theta).$$

12.3.11 nth Root of Unity

The real roots of 1 can only be ± 1, but complex numbers introduce the concept that unity possesses an infinite number of complex roots. The complex roots of 1 satisfy the equation $z^n = 1$, where n is a positive integer. Such roots are employed in different branches of mathematics, such as number theory and discrete Fourier transforms. (See https://en.wikipedia.org/wiki/Root_of_unity).

If the nth root of 1 is z, then $z^n = 1$. Therefore, using de Moivre's theorem:

$$1^{1/n} = e^{i2\pi k/n}, \quad k = 0, 1, 2, \ldots, n-1$$
$$= \cos\left(\tfrac{2\pi k}{n}\right) + i\sin\left(\tfrac{2\pi k}{n}\right).$$

For example, when $n = 3$:

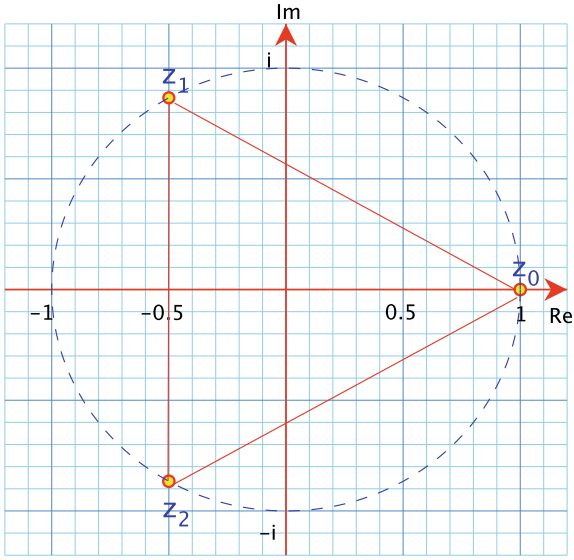

Fig. 12.11 Three roots of unity

$$
\begin{aligned}
[k=0] \quad & z_0 = \cos\left(\tfrac{0}{3}\right) + i\sin\left(\tfrac{0}{3}\right) = 1 \\
[k=1] \quad & z_1 = \cos\left(\tfrac{2\pi}{3}\right) + i\sin\left(\tfrac{2\pi}{3}\right) = -\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2} \\
[k=2] \quad & z_1 = \cos\left(\tfrac{4\pi}{3}\right) + i\sin\left(\tfrac{4\pi}{3}\right) = -\tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}.
\end{aligned}
$$

Let's confirm these results:

$$
\begin{aligned}
z_1^3 &= \left(-\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2}\right)\left(-\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2}\right)\left(-\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2}\right) \\
&= \left(-\tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}\right)\left(-\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2}\right) = 1 \\
z_2^3 &= \left(-\tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}\right)\left(-\tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}\right)\left(-\tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}\right) \\
&= \left(-\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2}\right)\left(-\tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}\right) = 1.
\end{aligned}
$$

These roots are located on the unit-radius complex circle, as shown in Fig. 12.11. Connecting the points together creates a regular polygon.

12.3.12 nth Roots of a Complex Number

Given

$$z = r(\cos\theta + i\sin\theta)$$

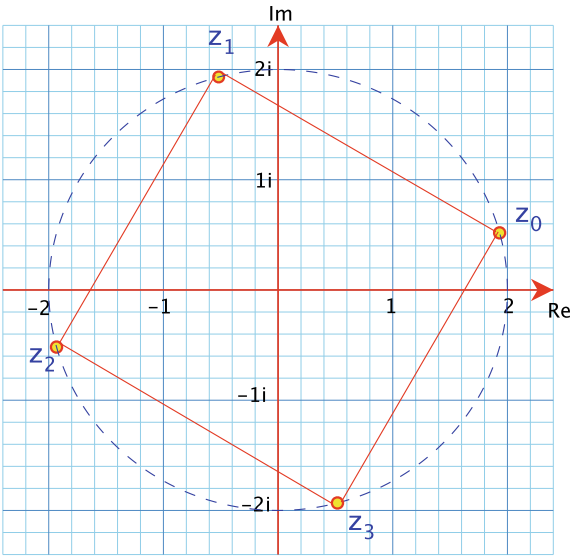

Fig. 12.12 4th roots of $16(\cos(\pi/3) + i\sin(\pi/3)$

then

$$\sqrt[n]{z} = \sqrt[n]{r}\left[\cos\left(\frac{\theta + k2\pi}{n}\right) + i\sin\left(\frac{\theta + k2\pi}{n}\right)\right], \quad 0 \le k \le (n-1).$$

For example, let's find $\left(8 + i8\sqrt{3}\right)^{1/4}$.

To begin, we convert it into polar form: $z = 16e^{i\pi/3}$.

$$\begin{aligned}
z &= 16\left[\cos\left(\tfrac{\pi}{3}\right) + i\sin\left(\tfrac{\pi}{3}\right)\right] \\
\sqrt[4]{z} &= \sqrt[4]{16}\left[\cos\left(\tfrac{\pi/3+k2\pi}{4}\right) + i\sin\left(\tfrac{\pi/3+k2\pi}{4}\right)\right] \\
[k=0] \quad z_0 &= 2\left[\cos\left(\tfrac{\pi}{12}\right) + i\sin\left(\tfrac{\pi}{12}\right)\right] \approx 1.932 + i0.518 \\
[k=1] \quad z_1 &= 2\left[\cos\left(\tfrac{\pi}{12} + \tfrac{2\pi}{4}\right) + i\sin\left(\tfrac{\pi}{12} + \tfrac{2\pi}{4}\right)\right] \approx -0.518 + i1.932 \\
[k=2] \quad z_2 &= 2\left[\cos\left(\tfrac{\pi}{12} + \tfrac{4\pi}{4}\right) + i\sin\left(\tfrac{\pi}{12} + \tfrac{4\pi}{4}\right)\right] \approx -1.932 - i0.518 \\
[k=3] \quad z_3 &= 2\left[\cos\left(\tfrac{\pi}{12} + \tfrac{6\pi}{4}\right) + i\sin\left(\tfrac{\pi}{12} + \tfrac{6\pi}{4}\right)\right] \approx 0.518 - i1.932.
\end{aligned}$$

Figure 12.12 shows these roots.

12.3.13 Logarithm of a Complex Number

In order to take the natural logarithm of a complex number, we use the exponential form. Consequently, if we are given $a + bi$, this must be converted to $re^{i\theta}$. Therefore, given

$$z = re^{i\theta}$$

then

$$\ln z = \ln r + i\theta$$

or

$$\ln z = \ln |z| + i \arg(z).$$

As exponential functions can have multiple values, the imaginary component is restricted to the interval $-\pi < \theta \leq \pi$. For example, -1 is represented by $e^{i\pi}$, $e^{3i\pi}$, $e^{5i\pi}$ etc., but to satisfy the agreed interval constraint, $-1 = e^{i\pi}$.

For example, given $z = -2 + 2i$, then

$$\begin{aligned} -2 + 2i &= \sqrt{(-2)^2 + 2^2} \cdot e^{i \tan^{-1}(2/-2)} \\ &= \sqrt{8} e^{i0.75\pi} \\ \ln(-2 + 2i) &= \ln\left(\sqrt{8} e^{i0.75\pi}\right) \\ &= \ln \sqrt{8} + 0.75\pi i \\ &\approx 1.039721 + 2.356194i. \end{aligned}$$

Similarly, given $z = 3 - 4i$, then

$$\begin{aligned} 3 - 4i &= \sqrt{3^2 + (-4)^2} \cdot e^{i \tan^{-1}(-4/3)} \\ &= 5e^{-i0.927295} \\ \ln(3 - 4i) &= \ln\left(5e^{-i0.927295}\right) \\ &= \ln 5 - 0.927295i \\ &\approx 1.609438 - 0.927295i. \end{aligned}$$

Logarithms of other complex numbers are shown in Table 12.3.

12.3.14 Raising a Complex Number to a Complex Power

Now that we have seen how to take a logarithm of a complex number, the way is open to raise a complex number to a complex power. For example, given

$$z = e^y \tag{12.8}$$

then

$$y = \ln z \tag{12.9}$$

Table 12.3 Logarithms of complex numbers

z	e form	$\ln z$
1	e^{i0}	0
-1	$e^{i\pi}$	πi
i	$e^{i\pi/2}$	$\frac{\pi}{2}i$
$-i$	$e^{-i\pi/2}$	$-\frac{\pi}{2}i$
5	$5e^{i0}$	1.609438
-5	$5e^{i\pi}$	$1.609438+\pi i$
$5i$	$5e^{i\pi/2}$	$1.609438+\frac{\pi}{2}i$
$-5i$	$5e^{-i\pi/2}$	$1.609438-\frac{\pi}{2}i$
$5+5i$	$\sqrt{50}e^{i\pi/4}$	$1.956012+\frac{\pi}{4}i$
$5-5i$	$\sqrt{50}e^{-i\pi/4}$	$1.956012-\frac{\pi}{4}i$
$-5+5i$	$\sqrt{50}e^{i3\pi/4}$	$1.956012+\frac{3\pi}{4}i$
$-5-5i$	$\sqrt{50}e^{-i3\pi/4}$	$1.956012-\frac{3\pi}{4}i$
0.5	$0.5e^{i0}$	-0.693147
-0.5	$0.5e^{i\pi}$	$-0.693147+\pi i$
$0.5i$	$0.5e^{\pi/2}$	$-0.693147+\frac{\pi}{2}i$
$-0.5i$	$0.5e^{-i\pi/2}$	$-0.693147-\frac{\pi}{2}i$

and substituting (12.9) in (12.8), we obtain

$$z = e^{\ln z}. \tag{12.10}$$

Raising both sides of (12.10) to some power w,

$$z^w = \left(e^{\ln z}\right)^w = e^{w \ln z}. \tag{12.11}$$

Equation (12.11) applies to both real and complex numbers, so first, let's begin with

$$z = 2$$
$$w = 1 + i$$

which requires raising e to the product of $1 + i$ and the natural logarithm of 2.

$$\begin{aligned}
\ln 2 &\approx 0.693147 \\
(1+i)0.693147 &= 0.693147 + 0.693147i \\
e^{(0.693147+0.693147i)} &= e^{0.693147}e^{0.693147i} \\
&= 2(\cos 0.693147 + i \sin 0.693147) \\
&\approx 2(0.769239 + 0.638961i) \\
&\approx 1.538478 + 1.277922i
\end{aligned}$$

therefore,

$$2^{1+i} \approx 1.538478 + 1.277922i.$$

Now let's use

$$\begin{aligned}
z &= 2 + 2i \\
w &= 1 + i
\end{aligned}$$

then

$$z^w = (2+2i)^{1+i} = e^{(1+i)\ln(2+2i)}$$

which requires raising e to the product of $1+i$ and the natural logarithm of $2+2i$. Not very nice, but let's have a go!

$$\begin{aligned}
2 + 2i &= \sqrt{2^2 + 2^2}e^{i\pi/4} \\
&= \sqrt{8}e^{i\pi/4} \\
\ln(2+2i) &= \ln 2.828427 + \frac{i\pi}{4} \\
&\approx 1.039721 + 0.785398i \\
(1+i)(1.039721 + 0.785398i) &= 1.039721 + 1.039721i + 0.785398i - 0.785398 \\
&= 0.254323 + 1.825119i \\
e^{(0.254323+1.825119i)} &= e^{0.254323}e^{1.825119i} \\
&= 1.289588(\cos 1.825119 + i \sin 1.825119) \\
&\approx 1.289588(-0.25159 + 0.967834i) \\
&\approx -0.324447 + 1.248107i
\end{aligned}$$

therefore,

$$(2+2i)^{1+i} \approx -0.324447 + 1.248107i.$$

12.3.15 Visualising Simple Complex Functions

We are aware of how real functions such as $f(x) = 2x^2 + 3x + 5$ behave, as it is possible to draw a graph relating $f(x)$ to x. But when it comes to complex functions, such as $f(z) = (a + bi)^2$, we require two dimensions to represent the original real and imaginary terms, and two further dimensions to represent the function, which is difficult in our three-dimensional world. However, in order to get a feel for what is happening between a complex variable and function, we can plot how individual numbers behave when subject to a function. To illustrate this, Fig. 12.13 illustrates how nine complex numbers in the first quadrant, behave when they are subject to a square function. For example, $(1 + 3i)^2$ moves to $-8 + 6i$, and $(3 + 3i)^2$ moves to $18i$. The dashed lines show the trajectory as the exponent increases from 1 to 2. Note that the squaring function imposes an anti-clockwise rotation on the trajectories, with the end complex numbers in the same or second quadrant. The solid blue lines connect the transformed points together to emphasise the distortion caused by the squaring transformation.

The functions used to draw the blue lines are

$$\left.\begin{aligned} f(z) &= [(1+i)(1-t) + (1+3i)t]^2 \\ f(z) &= [(2+i)(1-t) + (2+3i)t]^2 \\ f(z) &= [(3+i)(1-t) + (3+3i)t]^2 \\ f(z) &= [(1+i)(1-t) + (3+i)t]^2 \\ f(z) &= [(1+2i)(1-t) + (3+2i)t]^2 \\ f(z) &= [(1+3i)(1-t) + (3+3i)t]^2 \end{aligned}\right\} \quad 0 \leq t \leq 1.$$

Figure 12.14 shows the trajectories for nine similar complex numbers in the second quadrant, where the squaring function imposes an anti-clockwise rotation on the trajectories, with the end complex numbers in the third or fourth quadrant.

Figure 12.15 shows the trajectories for nine complex numbers in the third quadrant, where the squaring function imposes a clockwise rotation on the trajectories, with the end complex numbers in the first or second quadrant.

Figure 12.16 shows the trajectories for nine complex numbers in the fourth quadrant, where the squaring function imposes a clockwise rotation on the trajectories, with the end complex numbers in the third or fourth quadrant.

The only remaining numbers to consider are on the real and imaginary axes. The real axis is simple, as the square of any real number is another real number. Figure 12.17 shows the anti-clockwise trajectories of four positive imaginary numbers, and the clockwise trajectories of four negative imaginary numbers.

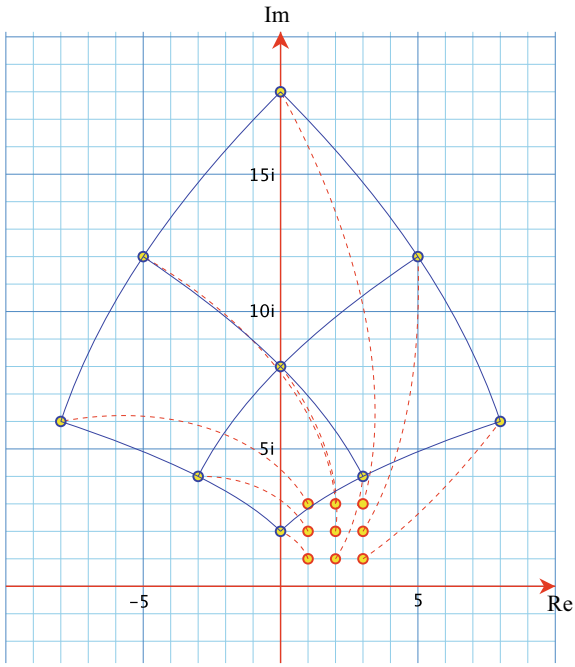

Fig. 12.13 The trajectories of nine complex numbers in the first quadrant, when squared

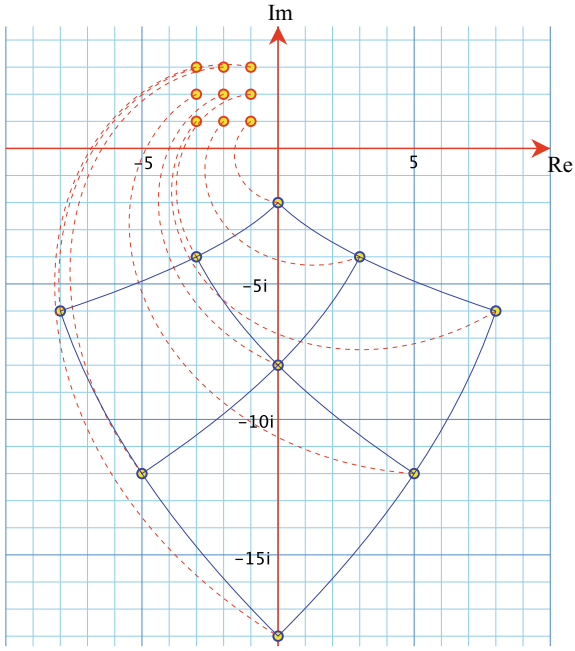

Fig. 12.14 The trajectories of nine complex numbers in the second quadrant, when squared

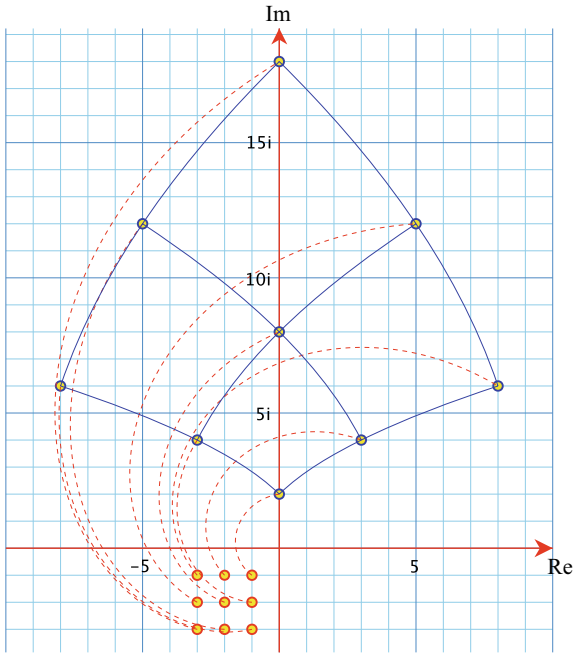

Fig. 12.15 The trajectories of nine complex numbers in the third quadrant, when squared

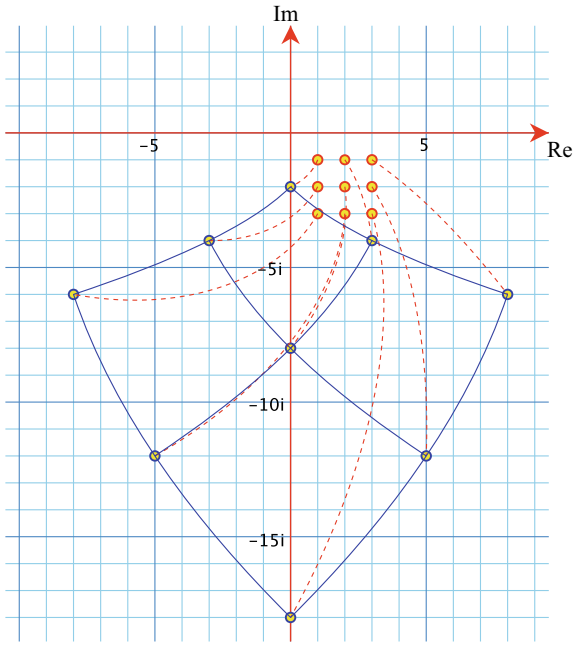

Fig. 12.16 The trajectories of nine complex numbers in the fourth quadrant, when squared

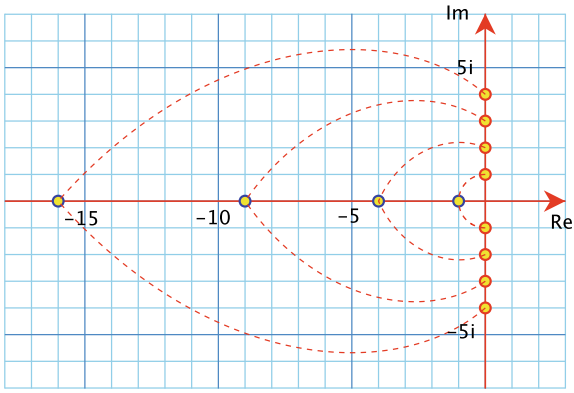

Fig. 12.17 The trajectories of eight, squared imaginary numbers

12.3.16 The Hyperbolic Functions

The trigonometric functions derive from the geometry of the circle $x^2 + y^2 = 1$, whereas the *hyperbolic functions* are associated with the geometry of the hyperbola $x^2 - y^2 = 1$. However, they are all related to e as we will see.

Given

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}$$

then

$$\cos(i\theta) = \frac{e^{i(i\theta)} + e^{-i(i\theta)}}{2} = \frac{e^{-\theta} + e^{\theta}}{2} = \cosh\theta.$$

Similarly, given

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

then

$$\sin(i\theta) = \frac{e^{ii\theta} - e^{-ii\theta}}{2i} = \frac{e^{-\theta} - e^{\theta}}{2i} = \frac{i(e^{\theta} - e^{-\theta})}{2} = i\sinh\theta.$$

By definition:

$$\tanh\theta = \frac{\sinh\theta}{\cosh\theta} = \frac{e^{\theta} - e^{-\theta}}{e^{\theta} + e^{-\theta}}.$$

Therefore,

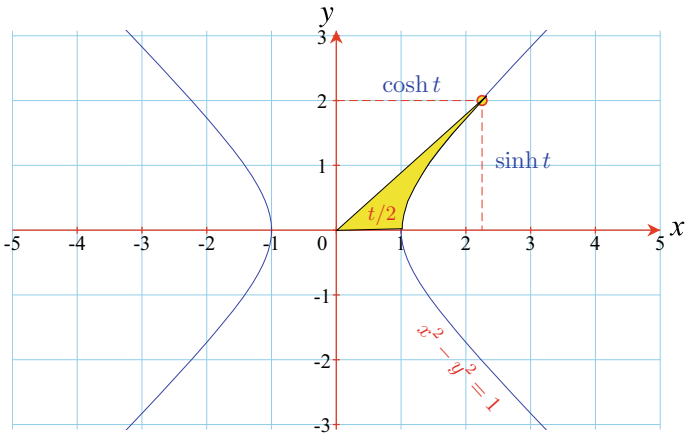

Fig. 12.18 The hyperbolic functions

$$\cosh\theta = \frac{e^{\theta} + e^{-\theta}}{2} = \cos(i\theta)$$
$$\sinh\theta = \frac{e^{\theta} - e^{-\theta}}{2} = -i\sin(i\theta)$$
$$\tanh\theta = \frac{e^{\theta} - e^{-\theta}}{e^{\theta} + e^{-\theta}} = -i\tan(i\theta)$$
$$\cosh\theta + \sinh\theta = e^{\theta}$$
$$\cosh\theta - \sinh\theta = e^{-\theta}$$
$$\cosh^2\theta - \sinh^2\theta = 1.$$

Figure 12.18 shows how sinh and cosh relate to the hyperbola.

12.4 Summary

Hopefully, this chapter has established $i = \sqrt{-1}$ as an incredible invention. Even though it does not belong to the traditional number systems, it is a valid mathematical object and reveals hidden numerical relationships between various constants and functions. Perhaps the two outstanding examples being $e^{i\pi} + 1 = 0$ and $i^i = 0.207\,879\ldots$.

The complex plane provides a simple way of visualising complex numbers, and illustrates their connection with vectors.

Euler's proof for $e^{i\theta} = \cos\theta + i\sin\theta$ opens the door for associating complex numbers with wave phenomena, which include acoustic waves, sea waves, electronics and quantum mechanics.

12.5 Worked Examples

12.5.1 Complex Addition

Compute $(3+2i)+(2+2i)+(5-3i)$.
Solution: Collect up like terms.

$$(3+2i)+(2+2i)+(5-3i) = 10+i.$$

12.5.2 Complex Products

Compute $(3+2i)(2+2i)(5-3i)$.

Solution: Expand algebraically and simplify.

$$\begin{aligned}(3+2i)(2+2i)(5-3i) &= (3+2i)(10-6i+10i+6)\\ &= (3+2i)(16+4i)\\ &= 48+12i+32i-8\\ &= 40+44i.\end{aligned}$$

12.5.3 Complex Division

Compute $\frac{1}{(2+3i)(4-5i)}$.

Solution: Expand the denominator, then multiply top and bottom by the denominator's conjugate.

$$\begin{aligned}\frac{1}{(2+3i)(4-5i)} &= \frac{1}{23+2i}\\ &= \frac{1}{(23+2i)}\frac{(23-2i)}{(23-2i)}\\ &= \frac{23-2i}{529+4}\\ &= \tfrac{1}{533}(23-2i).\end{aligned}$$

12.5.4 Complex Rotation

Rotate the complex point $3 + 2i$ by $\pm 90°$ and $\pm 180°$.
Solution: Multiply by $\pm i$ and -1.
To rotate $+90°$ (anti-clockwise) multiply by i.

$$i(3 + 2i) = 3i - 2 = -2 + 3i.$$

To rotate $-90°$ (clockwise) multiply by $-i$.

$$-i(3 + 2i) = -3i + 2 = 2 - 3i.$$

To rotate $+180°$ (anti-clockwise) multiply by -1.

$$-1(3 + 2i) = -3 - 2i.$$

To rotate $-180°$ (clockwise) multiply by -1.

$$-1(3 + 2i) = -3 - 2i.$$

12.5.5 Polar Notation

Given $z_1 = \frac{1}{\sqrt{2}} + \frac{\sqrt{2}}{2}i$ and $z_2 = -\frac{1}{\sqrt{2}} + \frac{\sqrt{2}}{2}i$, compute their product using standard complex number format, and polar notation.

Standard complex number format.
Solution: Expand algebraically.

$$\begin{aligned} z_1 z_2 &= \left(\tfrac{1}{\sqrt{2}} + \tfrac{\sqrt{2}}{2}i\right)\left(-\tfrac{1}{\sqrt{2}} + \tfrac{\sqrt{2}}{2}i\right) \\ &= -\tfrac{1}{2} - \tfrac{1}{2} \\ &= -1. \end{aligned}$$

Polar notation.
Solution: Compute the amplitude and argument for z_1 and z_2; multiply the amplitudes, and add the arguments.

$$r_1 = \sqrt{\left(\tfrac{1}{\sqrt{2}}\right)^2 + \left(\tfrac{\sqrt{2}}{2}\right)^2} = 1$$

$$r_2 = \sqrt{\left(-\tfrac{1}{\sqrt{2}}\right)^2 + \left(\tfrac{\sqrt{2}}{2}\right)^2} = 1$$

$$\begin{aligned}
\theta_1 &= \tan^{-1}\left(\tfrac{\sqrt{2}}{2}\tfrac{\sqrt{2}}{1}\right) = 45^\circ \\
\theta_2 &= \tan^{-1}\left(-\tfrac{\sqrt{2}}{2}\tfrac{\sqrt{2}}{1}\right) = 135^\circ \\
z_1 &= (1, 45^\circ) \\
z_2 &= (1, 135^\circ) \\
z_1 z_2 &= (1, 180^\circ) = -1.
\end{aligned}$$

12.5.6 Real and Imaginary Parts

Find the real and imaginary parts of $1/(1 + \mathrm{e}^{i2\theta})$.
Solution: Multiply top and bottom by the conjugate, expand and isolate the real and imaginary parts.

$$\begin{aligned}
\frac{1}{1+\mathrm{e}^{i2\theta}} &= \frac{1+\mathrm{e}^{-i2\theta}}{(1+\mathrm{e}^{i2\theta})(1+\mathrm{e}^{-i2\theta})} \\
&= \frac{1+\cos(2\theta) - i\sin(2\theta)}{2+2\cos(2\theta)} \\
&= \frac{1+\cos(2\theta)}{2+2\cos(2\theta)} - i\frac{\sin(2\theta)}{2+2\cos(2\theta)} \\
\mathrm{Re}\left(\frac{1}{1+\mathrm{e}^{i2\theta}}\right) &= \tfrac{1}{2} \\
\mathrm{Im}\left(\frac{1}{1+\mathrm{e}^{i2\theta}}\right) &= -\tfrac{1}{2}\left(\frac{\sin(2\theta)}{1+\cos(2\theta)}\right).
\end{aligned}$$

12.5.7 Magnitude of a Complex Number

Find $\left|\frac{1}{1+\mathrm{e}^{i2\theta}}\right|$.
Solution: Use $z\bar{z} = |z|^2$ and expand.

$$\begin{aligned}
\left|\frac{1}{1+\mathrm{e}^{i2\theta}}\right|^2 &= \frac{1}{(1+\mathrm{e}^{i2\theta})(1+\mathrm{e}^{-i2\theta})} \\
&= \frac{1}{1+\mathrm{e}^{-i2\theta}+\mathrm{e}^{i2\theta}+1} \\
&= \frac{1}{2+2\cos(2\theta)} \\
\left|\frac{1}{1+\mathrm{e}^{i2\theta}}\right| &= [2+2\cos(2\theta)]^{-1/2}.
\end{aligned}$$

12.5.8 Complex Norm

Find the norm of $z = 5 + 12i$.
Solution: Use $\|z\| = \sqrt{a^2 + b^2}$.

$$\|z\| = |z| = \sqrt{5^2 + 12^2} = 13.$$

12.5.9 Complex Inverse

Find the inverse of $1 + i$.
Solution: Multiply top and bottom by the conjugate and expand.

$$(1+i)^{-1} = \frac{(1-i)}{(1-i)(1+i)} = \tfrac{1}{2}(1-i).$$

12.5.10 de Moivre's Theorem

Express $\cos(5\theta)$ in terms of $\cos\theta$, and $\sin(5\theta)$ in terms of $\sin\theta$.
Solution: Use $(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta)$ and simplify.

$$\begin{aligned}
\cos(5\theta) + i\sin(5\theta) &= (\cos\theta + i\sin\theta)^5 \\
&= \cos^5\theta + 5i\cos^4\theta\cdot\sin\theta + 10i^2\cos^3\theta\cdot\sin^2\theta \\
&\quad + 10i^3\cos^2\theta\cdot\sin^3\theta + 5i^4\cos\theta\cdot\sin^4\theta + i^5\sin^5\theta \\
&= \cos^5\theta - 10\cos^3\theta\cdot\sin^2\theta + 5\cos\theta\cdot\sin^4\theta \\
&\quad + i(5\cos^4\theta\cdot\sin\theta - 10\cos^2\theta\cdot\sin^3\theta + \sin^5\theta) \\
\cos(5\theta) &= \mathrm{Re}[(\cos\theta + i\sin\theta)^5] \\
&= \cos^5\theta - 10\cos^3\theta\cdot\sin^2\theta + 5\cos\theta\cdot\sin^4\theta
\end{aligned}$$

but $\sin^2\theta = 1 - \cos^2\theta$

$$\begin{aligned}
&= \cos^5\theta - 10\cos^3\theta(1-\cos^2\theta) + 5\cos\theta(1-\cos^2\theta)^2 \\
&= \cos^5\theta - 10\cos^3\theta + 10\cos^5\theta + 5\cos\theta(1 - 2\cos^2\theta + \cos^4\theta) \\
&= 11\cos^5\theta - 10\cos^3\theta + 5\cos\theta - 10\cos^3\theta + 5\cos^5\theta \\
\cos(5\theta) &= 16\cos^5\theta - 20\cos^3\theta + 5\cos\theta.
\end{aligned}$$

$$\begin{aligned}
\sin(5\theta) &= \mathrm{Im}[(\cos\theta + i\sin\theta)^5] \\
&= 5\cos^4\theta\cdot\sin\theta - 10\cos^2\theta\cdot\sin^3\theta + \sin^5\theta
\end{aligned}$$

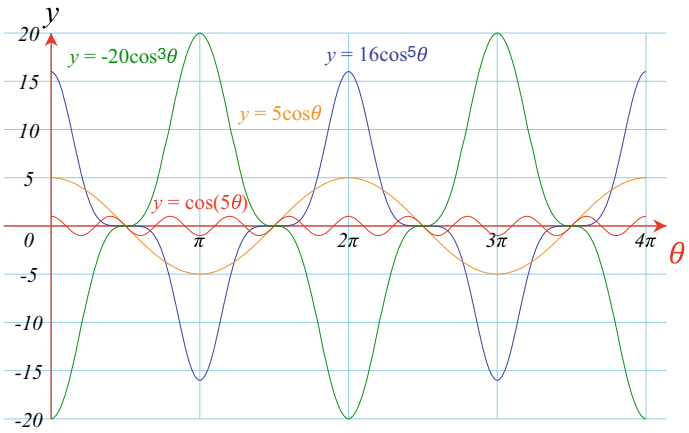

Fig. 12.19 $\cos(5\theta) = 16\cos^5\theta - 20\cos^3\theta + 5\cos\theta$

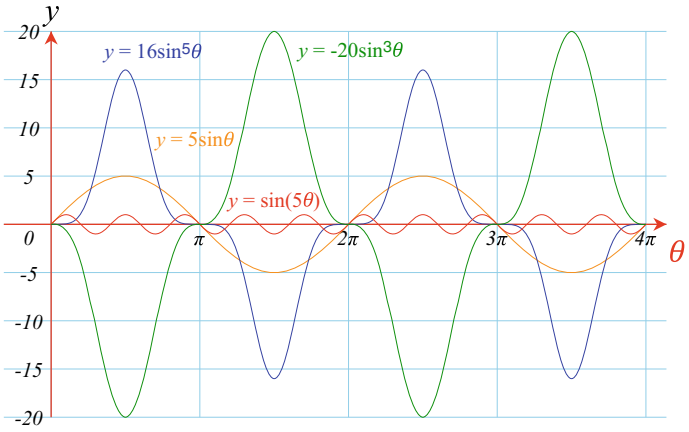

Fig. 12.20 $\sin(5\theta) = 16\sin^5\theta - 20\sin^3\theta + 5\sin\theta$

but $\cos^2\theta = 1 - \sin^2\theta$

$$
\begin{aligned}
&= 5\sin\theta(1-\sin^2\theta)^2 - 10\sin^3\theta(1-\sin^2\theta) + \sin^5\theta \\
&= 5\sin\theta - 10\sin^3\theta + 5\sin^5\theta - 10\sin^3\theta + 10\sin^5\theta + \sin^5\theta \\
\sin(5\theta) &= 16\sin^5\theta - 20\sin^3\theta + 5\sin\theta.
\end{aligned}
$$

Figure 12.19 shows the individual waveforms contributing towards $\cos(5\theta)$, and Fig. 12.20, the individual waveforms contributing towards $\sin(5\theta)$.

12.5.11 nth Root of Unity

Find the 4th and 6th roots of 1.
Solution: Use

$$\begin{aligned} z &= \mathrm{e}^{i2k\pi/n}, \quad k = 0, 1, 2, \ldots, n-1 \\ &= \cos\left(\tfrac{2k\pi}{n}\right) + i \sin\left(\tfrac{2k\pi}{n}\right). \end{aligned}$$

substituting different values for n and k.
When $n = 4$:

$$\begin{aligned} [k=0] \quad & z_0 = \cos\left(\tfrac{0}{4}\right) + i \sin\left(\tfrac{0}{4}\right) = 1 \\ [k=1] \quad & z_1 = \cos\left(\tfrac{2\pi}{4}\right) + i \sin\left(\tfrac{2\pi}{4}\right) = i \\ [k=2] \quad & z_1 = \cos\left(\tfrac{4\pi}{4}\right) + i \sin\left(\tfrac{4\pi}{4}\right) = -1 \\ [k=3] \quad & z_1 = \cos\left(\tfrac{6\pi}{4}\right) + i \sin\left(\tfrac{6\pi}{4}\right) = -i. \end{aligned}$$

When $n = 6$:

$$\begin{aligned} [k=0] \quad & z_0 = \cos\left(\tfrac{0}{6}\right) + i \sin\left(\tfrac{0}{6}\right) = 1 \\ [k=1] \quad & z_1 = \cos\left(\tfrac{2\pi}{6}\right) + i \sin\left(\tfrac{2\pi}{6}\right) = \tfrac{1}{2} + i\tfrac{\sqrt{3}}{2} \\ [k=2] \quad & z_1 = \cos\left(\tfrac{4\pi}{6}\right) + i \sin\left(\tfrac{4\pi}{6}\right) = -\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2} \\ [k=3] \quad & z_1 = \cos\left(\tfrac{6\pi}{6}\right) + i \sin\left(\tfrac{6\pi}{6}\right) = -1 \\ [k=4] \quad & z_1 = \cos\left(\tfrac{8\pi}{6}\right) + i \sin\left(\tfrac{8\pi}{6}\right) = -\tfrac{1}{2} - i\tfrac{\sqrt{3}}{2} \\ [k=5] \quad & z_1 = \cos\left(\tfrac{10\pi}{6}\right) + i \sin\left(\tfrac{10\pi}{6}\right) = \tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}. \end{aligned}$$

12.5.12 Roots of a Complex Number

Find $\sqrt[3]{i}$.
Solution: Convert i to polar form and use

$$\sqrt[n]{z} = \sqrt[n]{r}\left[\cos\left(\tfrac{\theta+k2\pi}{n}\right) + i \sin\left(\tfrac{\theta+k2\pi}{n}\right)\right], \quad 0 \le k \le n-1.$$

$$\begin{aligned} z &= 0 + i = (r, \theta) \\ r &= 1 \\ \theta &= \pi/2 \end{aligned}$$

$$\begin{aligned}
\sqrt[3]{i} &= \cos\left(\tfrac{\theta+k2\pi}{3}\right) + i\sin\left(\tfrac{\theta+k2\pi}{3}\right), \quad 0 \le k \le 2 \\
&= \cos\left(\tfrac{\pi}{6} + \tfrac{k2\pi}{3}\right) + i\sin\left(\tfrac{\pi}{6} + \tfrac{k2\pi}{3}\right) \\
z_0 &= \cos\left(\tfrac{\pi}{6}\right) + i\sin\left(\tfrac{\pi}{6}\right) = \tfrac{\sqrt{3}}{2} + i\,\tfrac{1}{2} \\
z_1 &= \cos\left(\tfrac{\pi}{6} + \tfrac{2\pi}{3}\right) + i\sin\left(\tfrac{\pi}{6} + \tfrac{2\pi}{3}\right) = -\tfrac{\sqrt{3}}{2} + i\,\tfrac{1}{2} \\
z_2 &= \cos\left(\tfrac{\pi}{6} + \tfrac{4\pi}{3}\right) + i\sin\left(\tfrac{\pi}{6} + \tfrac{4\pi}{3}\right) = -i.
\end{aligned}$$

12.5.13 Logarithm of a Complex Number

Compute the natural logarithm of $z = 5 - 12i$.
Solution: Convert z to polar form and use $\ln z = \ln|z| + i\arg(z)$.

$$\begin{aligned}
5 - 12i &= \sqrt{5^2 + (-12)^2}\,\mathrm{e}^{i\tan^{-1}(-12/5)} \\
&\approx 13\mathrm{e}^{-i1.176} \\
\ln(5 - 12i) &\approx \ln\left(13\mathrm{e}^{-i1.176}\right) \\
&\approx \ln 13 - 1.176i \\
&\approx 2.565 - 1.176i.
\end{aligned}$$

12.5.14 Raising a Number to a Complex Power

Compute 3^{1+i}.
Solution: Use $z^w = \mathrm{e}^{w\ln z}$.

$$\begin{aligned}
z &= 3 \\
w &= 1 + i \\
z^w &= \mathrm{e}^{w\ln z} \\
\ln 3 &\approx 1.0986 \\
(1+i)\ln 3 &\approx 1.0986 + 1.0986i \\
3^{1+i} &\approx \mathrm{e}^{(1.0986+1.0986i)} \\
&\approx \mathrm{e}^{1.0986}\mathrm{e}^{1.0986i} \\
&\approx 3(\cos 1.0986 + i\sin 1.0986) \\
&\approx 3(0.4548 + 0.8906i) \\
&\approx 1.3644 + 2.6718i
\end{aligned}$$

References

Feynman RP (1977) The Feynman lectures on physics, vol 1. Addison-Wesley, Reading, p 22-10
Vince J (2018) Imaginary mathematics for computer science. Springer, Berlin. ISBN 978-3-319-94636-8

Chapter 13
Matrices

13.1 Introduction

Matrices, like determinants, have their background in algebra and offer another way to represent and manipulate equations. Matrices can be added, subtracted and multiplied together, and even inverted, however, they must give the same result obtained through traditional algebraic techniques. A useful way to introduce the subject is via geometric transforms, which we examine first.

13.2 Geometric Transforms

Let $P(x, y)$ be a vertex on a 2D shape, then we can devise a *geometric transform* where $P(x, y)$ becomes $P'(x', y')$ on a second shape. For example, when the following transform is applied to every point on a shape, it is halved in size, relative to the origin:

$$x' = 0.5x$$
$$y' = 0.5y$$

and this transform translates a shape horizontally by 4 units:

$$x' = x + 4$$
$$y' = y.$$

Figure 13.1 illustrates two successive transforms applied to the large green star centred at the origin. The first transform scales the star by a factor of 0.5 creating the smaller yellow star, which in turn is subjected to a horizontal translation of 4 units, creating the blue star.

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_13

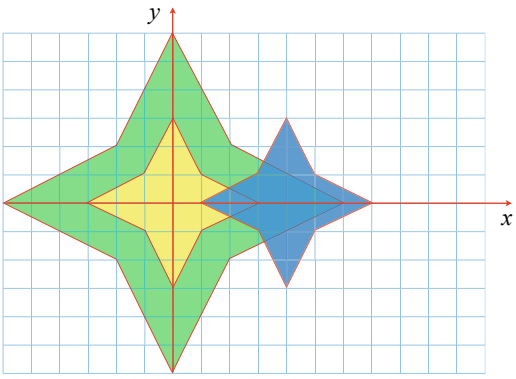

Fig. 13.1 A scale transform followed by a translate transform

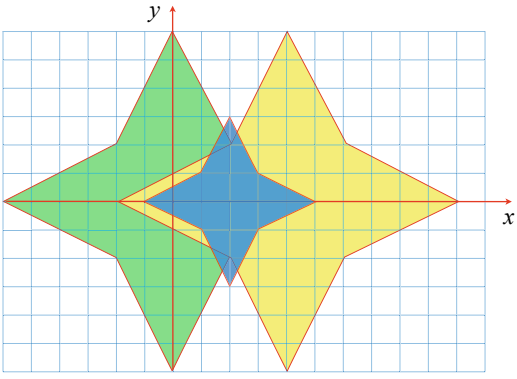

Fig. 13.2 A translate transform followed by a scale transform

Figure 13.2 starts with the same green star, but this time it is translated before being scaled. The final blue star ends up in a different position to the one shown in Fig. 13.1, which demonstrates the importance of transform order.

The algebra supporting the transforms in Fig. 13.1 comprises:

$$
\begin{aligned}
x' &= 0.5x \\
y' &= 0.5y \\
x'' &= x' + 4 \\
y'' &= y'
\end{aligned}
$$

which simplifies to

$$
\begin{aligned}
x'' &= 0.5x + 4 \\
y'' &= 0.5y
\end{aligned}
$$

whereas, the algebra supporting the transforms in Fig. 13.2 comprises:

$$\begin{aligned} x' &= x + 4 \\ y' &= y \\ x'' &= 0.5x' \\ y'' &= 0.5y' \end{aligned}$$

which simplifies to

$$\begin{aligned} x'' &= 0.5(x + 4) \\ y'' &= 0.5y \end{aligned}$$

and reveals the difference between the two transform sequences.

13.3 Transforms and Matrices

Matrix notation was researched by the British mathematician Arthur Cayley around 1858. Cayley formalised matrix algebra, along with the American mathematicians Charles Peirce (1839–1914) and his father, Benjamin Peirce (1809–1880). Previously, Carl Gauss had shown that transforms were not commutative, i.e. $\mathbf{T}_1\mathbf{T}_2 \neq \mathbf{T}_2\mathbf{T}_1$, (where $\mathbf{T}_1$ and $\mathbf{T}_2$ are transforms) and matrix notation clarified such observations.

Consider the transform $\mathbf{T}_1$, where x and y are transformed into x' and y' respectively:

$$\mathbf{T}_1 = \begin{cases} x' = ax + by \\ y' = cx + dy \end{cases} \tag{13.1}$$

and a second transform $\mathbf{T}_2$, where x' and y' are transformed into x'' and y'' respectively:

$$\mathbf{T}_2 = \begin{cases} x'' = Ax' + By' \\ y'' = Cx' + Dy' \end{cases}. \tag{13.2}$$

Substituting (13.1) in (13.2) we get

$$\mathbf{T_3} = \begin{cases} x'' = A(ax + by) + B(cx + dy) \\ y'' = C(ax + by) + D(cx + dy) \end{cases}$$

which simplifies to

$$\mathbf{T_3} = \begin{cases} x'' = (Aa + Bc)x + (Ab + Bd)y \\ y'' = (Ca + Dc)x + (Cb + Dd)y \end{cases}. \tag{13.3}$$

Having derived the algebra for $\mathbf{T_3}$, let's examine matrix notation, where constants are separated from the variables. For example, the transform (13.4)

$$\begin{aligned} x' &= ax + by \\ y' &= cx + dy \end{aligned} \tag{13.4}$$

can be written in matrix form as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{13.5}$$

where (13.5) contains two different structures: two single-column matrices or column vectors

$$\begin{bmatrix} x' \\ y' \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x \\ y \end{bmatrix},$$

and a 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Algebraically, (13.4) and (13.5) are identical, which dictates the way (13.5) is converted to (13.4). Therefore, using (13.5) we have x' followed by the "=" sign, and the sum of the products of the top row of constants a and b with the x and y in the last column vector:

$$x' = ax + by.$$

Next, we have y' followed by the "=" sign, and the sum of the products of the bottom row of constants c and d with the x and y in the last column vector:

$$y' = cx + dy.$$

As an example,

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

is equivalent to

$$\begin{aligned} x' &= 3x + 4y \\ y' &= 5x + 6y. \end{aligned}$$

We can now write $\mathbf{T}_1$ and $\mathbf{T}_2$ using matrix notation:

$$\mathbf{T}_1 = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{13.6}$$

$$\mathbf{T}_2 = \begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} \tag{13.7}$$

and substituting (13.6) in (13.7) we have

$$\mathbf{T}_3 = \begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{13.8}$$

But we have already computed $\mathbf{T_3}$ (13.3), which in matrix form is:

$$\mathbf{T_3} = \begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} Aa + Bc & Ab + Bd \\ Ca + Dc & Cb + Dd \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{13.9}$$

which implies that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} Aa + Bc & Ab + Bd \\ Ca + Dc & Cb + Dd \end{bmatrix}$$

and demonstrates how matrices must be multiplied. Here are the rules for matrix multiplication:

$$\begin{bmatrix} A & B \\ \cdots & \cdots \end{bmatrix} \begin{bmatrix} a & \cdots \\ c & \cdots \end{bmatrix} = \begin{bmatrix} Aa + Bc & \cdots \\ \cdots & \cdots \end{bmatrix}.$$

1: The top left-hand corner element $Aa + Bc$ is the product of the top row of the first matrix by the left column of the second matrix.

$$\begin{bmatrix} A & B \\ \cdots & \cdots \end{bmatrix} \begin{bmatrix} \cdots & b \\ \cdots & d \end{bmatrix} = \begin{bmatrix} \cdots & Ab + Bd \\ \cdots & \cdots \end{bmatrix}.$$

2: The top right-hand element $Ab + Bd$ is the product of the top row of the first matrix by the right column of the second matrix.

$$\begin{bmatrix} \cdots & \cdots \\ C & D \end{bmatrix} \begin{bmatrix} a & \cdots \\ c & \cdots \end{bmatrix} = \begin{bmatrix} \cdots & \cdots \\ Ca + Dc & \cdots \end{bmatrix}.$$

3: The bottom left-hand element $Ca + Dc$ is the product of the bottom row of the first matrix by the left column of the second matrix.

$$\begin{bmatrix} \cdots & \cdots \\ C & D \end{bmatrix} \begin{bmatrix} \cdots & b \\ \cdots & d \end{bmatrix} = \begin{bmatrix} \cdots & \cdots \\ \cdots & Cb + Dd \end{bmatrix}.$$

4: The bottom right-hand element $Cb + Dd$ is the product of the bottom row of the first matrix by the right column of the second matrix.

Let's multiply the following matrices together:

$$\begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 7 & 9 \end{bmatrix} = \begin{bmatrix} (2 \times 3 + 4 \times 7) & (2 \times 5 + 4 \times 9) \\ (6 \times 3 + 8 \times 7) & (6 \times 5 + 8 \times 9) \end{bmatrix} = \begin{bmatrix} 34 & 46 \\ 74 & 102 \end{bmatrix}.$$

13.4 Matrix Notation

Having examined the background to matrices, we can now formalise their notation.

A matrix is an array of numbers (real, imaginary, complex, etc.) organised in m rows and n columns, where each entry a_{ij} belongs to the ith row and jth column:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}.$$

It is also convenient to express the above definition as

$$\mathbf{A} = [a_{ij}]_{m\,n}.$$

13.4.1 Matrix Dimension or Order

The *dimension* or *order* of a matrix is the expression $m \times n$ where m is the number of rows, and n is the number of columns.

13.4.2 Square Matrix

A *square matrix* has the same number of rows as columns:

$$\mathbf{A} = [a_{ij}]_{n\,n} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}, \quad \text{e.g.,} \quad \begin{bmatrix} 1 & -2 & 4 \\ 6 & 5 & 7 \\ 4 & 3 & 1 \end{bmatrix}.$$

13.4.3 Column Vector

A *column vector* is a matrix with a single column:

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad \text{e.g.,} \quad \begin{bmatrix} 2 \\ 3 \\ 23 \end{bmatrix}.$$

13.4.4 Row Vector

A *row vector* is a matrix with a single row:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{bmatrix}, \quad \text{e.g.,} \quad \begin{bmatrix} 2 & 3 & 5 \end{bmatrix}.$$

13.4.5 Null Matrix

A *null matrix* has all its elements equal to zero:

$$\theta_n = [a_{ij}]_{n\,n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \text{e.g.,} \quad \theta_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The null matrix behaves like zero when used with numbers, where we have, $0 + n = n + 0 = n$ and $0 \times n = n \times 0 = 0$, and similarly, $\theta + \mathbf{A} = \mathbf{A} + \theta = \mathbf{A}$ and $\theta\mathbf{A} = \mathbf{A}\theta = \theta$. For example,

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

13.4.6 Unit Matrix

A *unit matrix* $\mathbf{I}_n$, is a square matrix with the elements on its diagonal a_{11} to a_{nn} equal to 1:

$$\mathbf{I}_n = [a_{ij}]_{n\,n} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \quad \text{e.g.,} \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The unit matrix behaves like the number 1 in a conventional product, where we have, $1 \times n = n \times 1 = n$, and similarly, $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$. For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

13.4.7 Trace

The *trace* of a square matrix is the sum of the elements on its diagonal a_{11} to a_{nn}:

$$\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}.$$

For example, given

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad \text{then} \quad \mathrm{Tr}(\mathbf{A}) = 1 + 5 + 9 = 15.$$

The trace of a rotation matrix can be used to compute the angle of rotation. For example, the matrix to rotate a point about the origin is

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

where

$$\mathrm{Tr}(\mathbf{A}) = 2\cos\theta$$

which means that

$$\theta = \arccos\left(\frac{\mathrm{Tr}(\mathbf{A})}{2}\right).$$

The three matrices for rotating points about the x-, y- and z-axis are respectively:

$$\mathbf{R}_{\alpha,x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$

$$\mathbf{R}_{\alpha,y} = \begin{bmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{bmatrix}$$

$$\mathbf{R}_{\alpha,z} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and it is clear that

$$\mathrm{Tr}(\mathbf{R}_{\alpha,x}) = \mathrm{Tr}(\mathbf{R}_{\alpha,y}) = \mathrm{Tr}(\mathbf{R}_{\alpha,z}) = 1 + 2\cos\alpha$$

therefore,

$$\alpha = \arccos\left(\frac{\mathrm{Tr}(\mathbf{R}_{\alpha,x}) - 1}{2}\right).$$

13.4.8 Determinant of a Matrix

The *determinant* of a matrix is a scalar value computed from the elements of the matrix. The different methods for computing the determinant are described in Chap. 10. For example, using Sarrus's rule:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \text{then,} \quad \det\mathbf{A} = 45 + 84 + 96 - 105 - 48 - 72 = 0.$$

13.4.9 Transpose

The *transpose* of a matrix exchanges all row elements for column elements. The transposition is indicated by the letter 'T' outside the right-hand bracket.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}.$$

For example,

$$\begin{bmatrix} 1 & 2 & 4 \\ 6 & 5 & 7 \\ 4 & 3 & 1 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 & 6 & 4 \\ 2 & 5 & 3 \\ 4 & 7 & 1 \end{bmatrix},$$

and

$$\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 2 & 3 & 5 \end{bmatrix}.$$

To prove that $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$, we could develop a general proof using $n \times n$ matrices, but for simplicity, let's employ 3×3 matrices and assume the result generalises to higher dimensions. Given

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

and

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}, \quad \mathbf{B}^{\mathrm{T}} = \begin{bmatrix} b_{11} & b_{21} & b_{31} \\ b_{12} & b_{22} & b_{32} \\ b_{13} & b_{23} & b_{33} \end{bmatrix}$$

then,

$$\mathbf{AB} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} & a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} & a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} & a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{bmatrix}$$

$$(\mathbf{AB})^{\mathrm{T}} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} \\ a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} & a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} \\ a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} & a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} & a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{bmatrix}$$

and

$$\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} b_{11}a_{11} + b_{21}a_{12} + b_{31}a_{13} & b_{11}a_{21} + b_{21}a_{22} + b_{31}a_{23} & b_{11}a_{31} + b_{21}a_{32} + b_{31}a_{33} \\ b_{12}a_{11} + b_{22}a_{12} + b_{32}a_{13} & b_{12}a_{21} + b_{22}a_{22} + b_{32}a_{23} & b_{12}a_{31} + b_{22}a_{32} + b_{32}a_{33} \\ b_{13}a_{11} + b_{23}a_{12} + b_{33}a_{13} & b_{13}a_{21} + b_{23}a_{22} + b_{33}a_{23} & b_{13}a_{31} + b_{23}a_{32} + b_{33}a_{33} \end{bmatrix}$$

which confirms that $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$.

13.4.10 Symmetric Matrix

A *symmetric matrix* is a square matrix that equals its transpose: i.e., $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$. For example, $\mathbf{A}$ is a symmetric matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 5 & 3 \\ 4 & 3 & 6 \end{bmatrix}^{\mathrm{T}}.$$

In general, a square matrix $\mathbf{A} = \mathbf{S} + \mathbf{Q}$, where $\mathbf{S}$ is a symmetric matrix, and $\mathbf{Q}$ is an antisymmetric matrix. The symmetric matrix is computed as follows. Given a matrix $\mathbf{A}$ and its transpose $\mathbf{A}^{\mathrm{T}}$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{bmatrix}$$

their sum is

$$\mathbf{A} + \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 2a_{11} & a_{12} + a_{21} & \dots & a_{1n} + a_{n1} \\ a_{12} + a_{21} & 2a_{22} & \dots & a_{2n} + a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} + a_{n1} & a_{2n} + a_{n2} & \dots & 2a_{nn} \end{bmatrix}.$$

By inspection, $\mathbf{A} + \mathbf{A}^{\mathrm{T}}$ is symmetric, and if we divide throughout by 2 we have

$$\mathbf{S} = \tfrac{1}{2}\left(\mathbf{A} + \mathbf{A}^{\mathrm{T}}\right)$$

which is defined as the symmetric part of $\mathbf{A}$. For example, given

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

then

$$\begin{aligned} \mathbf{S} &= \tfrac{1}{2}\left(\mathbf{A} + \mathbf{A}^{\mathrm{T}}\right) \\ &= \begin{bmatrix} a_{11} & (a_{12} + a_{21})/2 & (a_{13} + a_{31})/2 \\ (a_{12} + a_{21})/2 & a_{22} & a_{23} + a_{32} \\ (a_{13} + a_{31})/2 & (a_{23} + a_{32})/2 & a_{33} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & s_3/2 & s_2/2 \\ s_3/2 & a_{22} & s_1/2 \\ s_2/2 & s_1/2 & a_{33} \end{bmatrix} \end{aligned}$$

where

$$s_1 = a_{23} + a_{32}$$
$$s_2 = a_{13} + a_{31}$$
$$s_3 = a_{12} + a_{21}.$$

Using a real example:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 4 \\ 3 & 1 & 4 \\ 4 & 2 & 6 \end{bmatrix}, \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 0 & 3 & 4 \\ 1 & 1 & 2 \\ 4 & 4 & 6 \end{bmatrix}$$
$$\mathbf{S} = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 1 & 3 \\ 4 & 3 & 6 \end{bmatrix}$$

which equals its own transpose.

13.4.11 Antisymmetric Matrix

An *antisymmetric matrix* is a matrix whose transpose is its own negative:

$$\mathbf{A}^{\mathrm{T}} = -\mathbf{A}$$

and is also known as a *skew-symmetric matrix*.

As the elements of $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}$ are related by

$$a_{row,col} = -a_{col,row}.$$

When $k = row = col$:

$$a_{k,k} = -a_{k,k}$$

which implies that the diagonal elements must be zero. For example, this is an antisymmetric matrix

$$\mathbf{A} = \begin{bmatrix} 0 & -2 & 4 \\ 2 & 0 & -3 \\ -4 & 3 & 0 \end{bmatrix} = -\begin{bmatrix} 0 & -2 & 4 \\ 2 & 0 & -3 \\ -4 & 3 & 0 \end{bmatrix}^{\mathrm{T}}.$$

The antisymmetric part is computed as follows. Given a matrix $\mathbf{A}$ and its transpose $\mathbf{A}^{\mathrm{T}}$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{bmatrix}$$

their difference is

$$\mathbf{A} - \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 0 & a_{12} - a_{21} & \dots & a_{1n} - a_{n1} \\ -(a_{12} - a_{21}) & 0 & \dots & a_{2n} - a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ -(a_{1n} - a_{n1}) & -(a_{2n} - a_{n2}) & \dots & 0 \end{bmatrix}.$$

It is clear that $\mathbf{A} - \mathbf{A}^{\mathrm{T}}$ is antisymmetric, and if we divide throughout by 2 we have

$$\mathbf{Q} = \tfrac{1}{2}\left(\mathbf{A} - \mathbf{A}^{\mathrm{T}}\right).$$

For example:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} 0 & (a_{12} - a_{21})/2 & (a_{13} - a_{31})/2 \\ (a_{21} - a_{12})/2 & 0 & (a_{23} - a_{32})/2 \\ (a_{31} - a_{13})/2 & (a_{32} - a_{23})/2 & 0 \end{bmatrix}$$

and if we maintain some symmetry with the subscripts, we have

$$\begin{aligned} \mathbf{Q} &= \begin{bmatrix} 0 & (a_{12} - a_{21})/2 & -(a_{31} - a_{13})/2 \\ -(a_{12} - a_{21})/2 & 0 & (a_{23} - a_{32})/2 \\ (a_{31} - a_{13})/2 & -(a_{23} - a_{32})/2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & q_3/2 & -q_2/2 \\ -q_3/2 & 0 & q_1/2 \\ q_2/2 & -q_1/2 & 0 \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} q_1 &= a_{23} - a_{32} \\ q_2 &= a_{31} - a_{13} \\ q_3 &= a_{12} - a_{21}. \end{aligned}$$

Using a real example:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 4 \\ 3 & 1 & 4 \\ 4 & 2 & 6 \end{bmatrix}, \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 0 & 3 & 4 \\ 1 & 1 & 2 \\ 4 & 4 & 6 \end{bmatrix}$$
$$\mathbf{Q} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

Furthermore, we have already computed

$$\mathbf{S} = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 1 & 3 \\ 4 & 3 & 6 \end{bmatrix}$$

and

$$\mathbf{S} + \mathbf{Q} = \begin{bmatrix} 0 & 1 & 4 \\ 3 & 1 & 4 \\ 4 & 2 & 6 \end{bmatrix} = \mathbf{A}.$$

13.5 Matrix Addition and Subtraction

As equations can be added and subtracted together, it follows that matrices can also be added and subtracted, as long as they have the same dimension. For example, given

$$\mathbf{A} = \begin{bmatrix} 11 & 22 \\ 14 & -15 \\ 27 & 28 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & 1 \\ -4 & 5 \\ 1 & 8 \end{bmatrix}$$

then

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 13 & 23 \\ 10 & -10 \\ 28 & 36 \end{bmatrix}, \quad \mathbf{A} - \mathbf{B} = \begin{bmatrix} 9 & 21 \\ 18 & -20 \\ 26 & 20 \end{bmatrix}.$$

13.5.1 Scalar Multiplication

As equations can be scaled and factorised, it follows that matrixes can also be scaled and factorised.

$$\lambda\mathbf{A} = \lambda \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} \lambda a_{11} & \lambda a_{12} & \dots & \lambda a_{13} \\ \lambda a_{21} & \lambda a_{22} & \dots & \lambda a_{23} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \dots & \lambda a_{mn} \end{bmatrix}.$$

For example,

$$2\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}.$$

13.6 Matrix Products

We have already seen that matrices can be multiplied together employing rules that maintain the algebraic integrity of the equations they represent. And as matrices may be vectors, rectangular or square, we need to examine the products that are permitted. To keep the notation simple, the definitions and examples are restricted to a dimension of 3 or 3×3.

We begin with row and column vectors.

13.6.1 Row and Column Vectors

Given

$$\mathbf{A} = \begin{bmatrix} a & b & c \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}$$

then

$$\mathbf{AB} = \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = a\alpha + b\beta + c\gamma$$

which is a scalar and equivalent to the dot or scalar product of two vectors.

For example, given

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 10 \\ 30 \\ 20 \end{bmatrix}$$

then

$$\mathbf{AB} = \begin{bmatrix} 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 10 \\ 30 \\ 20 \end{bmatrix} = 20 + 90 + 80 = 190.$$

Whereas,

$$\mathbf{BA} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} = \begin{bmatrix} b_{11}a_{11} & b_{11}a_{12} & b_{11}a_{13} \\ b_{21}a_{11} & b_{21}a_{12} & b_{21}a_{13} \\ b_{31}a_{11} & b_{31}a_{12} & b_{31}a_{13} \end{bmatrix}.$$

For example,

$$\mathbf{BA} = \begin{bmatrix} 10 \\ 30 \\ 20 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 20 & 30 & 40 \\ 60 & 90 & 120 \\ 40 & 60 & 80 \end{bmatrix}.$$

The products **AA** and **BB** are not permitted.

13.6.2 Row Vector and a Matrix

Given

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{m1} & b_{m2} & b_{33} \end{bmatrix}$$

then

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{m1} & b_{m2} & b_{33} \end{bmatrix} \\ &= \Big[(a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}) \; (a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}) \; (a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33}) \Big]. \end{aligned}$$

The product **BA** is not permitted.

For example, given

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}$$

then

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix} \\ &= \begin{bmatrix} (2+9+16) & (4+12+20) & (6+15+24) \end{bmatrix} \\ &= \begin{bmatrix} 27 & 36 & 45 \end{bmatrix}. \end{aligned}$$

13.6.3 Matrix and a Column Vector

Given

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}$$

then

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} \end{bmatrix}.$$

The product **BA** is not permitted.

For example, given

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}, \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

then

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2+6+12 \\ 6+12+20 \\ 8+15+24 \end{bmatrix} = \begin{bmatrix} 20 \\ 38 \\ 47 \end{bmatrix}.$$

13.6.4 Square Matrices

To clarify the products, lower-case Greek symbols are used with lower-case letters. Here are their names:

$$\begin{array}{lll} \alpha = \text{alpha}, & \beta = \text{beta}, & \gamma = \text{gamma}, \\ \lambda = \text{lambda}, & \mu = \text{mu}, & \nu = \text{nu}, \\ \rho = \text{rho}, & \sigma = \text{sigma}, & \tau = \text{tau}. \end{array}$$

Given

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ p & q & r \\ u & v & w \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \alpha & \beta & \gamma \\ \lambda & \mu & \nu \\ \rho & \sigma & \tau \end{bmatrix}$$

then

$$\mathbf{AB} = \begin{bmatrix} a & b & c \\ p & q & r \\ u & v & w \end{bmatrix} \begin{bmatrix} \alpha & \beta & \gamma \\ \lambda & \mu & \nu \\ \rho & \sigma & \tau \end{bmatrix} = \begin{bmatrix} a\alpha + b\lambda + c\rho & a\beta + b\mu + c\sigma & a\gamma + b\nu + c\tau \\ p\alpha + q\lambda + r\rho & p\beta + q\mu + r\sigma & p\gamma + q\nu + r\tau \\ u\alpha + v\lambda + w\rho & u\beta + v\mu + w\sigma & u\gamma + v\nu + w\tau \end{bmatrix}$$

and

$$\mathbf{BA} = \begin{bmatrix} \alpha & \beta & \gamma \\ \lambda & \mu & \nu \\ \rho & \sigma & \tau \end{bmatrix} \begin{bmatrix} a & b & c \\ p & q & r \\ u & v & w \end{bmatrix} = \begin{bmatrix} \alpha a + \beta p + \gamma u & \alpha b + \beta q + \gamma v & \alpha c + \beta r + \gamma w \\ \lambda a + \mu p + \nu u & \lambda b + \mu q + \nu v & \lambda c + \mu r + \nu w \\ \rho a + \sigma p + \tau u & \rho b + \sigma q + \tau v & \rho c + \sigma r + \tau w \end{bmatrix}.$$

For example, given

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 6 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 5 & 6 \\ 6 & 7 & 8 \end{bmatrix}$$

then

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 6 & 7 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 \\ 4 & 5 & 6 \\ 6 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 28 & 34 & 40 \\ 52 & 64 & 76 \\ 76 & 92 & 112 \end{bmatrix}$$

and

$$\mathbf{BA} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 5 & 6 \\ 6 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 6 & 7 \end{bmatrix} = \begin{bmatrix} 31 & 40 & 49 \\ 49 & 64 & 89 \\ 67 & 88 & 109 \end{bmatrix}.$$

13.6.5 Rectangular Matrices

Given two rectangular matrices **A** and **B**, where **A** has a dimension $m \times n$, the product **AB** is permitted, if and only if, **B** has a dimension $n \times p$. The resulting matrix has a dimension $m \times p$. For example, given

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \end{bmatrix}$$

then

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \end{bmatrix} \\ &= \begin{bmatrix} (a_{11}b_{11} + a_{12}b_{21}) & (a_{11}b_{12} + a_{12}b_{22}) & (a_{11}b_{13} + a_{12}b_{23}) & (a_{11}b_{14} + a_{12}b_{24}) \\ (a_{21}b_{11} + a_{22}b_{21}) & (a_{21}b_{12} + a_{22}b_{22}) & (a_{21}b_{13} + a_{22}b_{23}) & (a_{21}b_{14} + a_{22}b_{24}) \\ (a_{31}b_{11} + a_{32}b_{21}) & (a_{31}b_{12} + a_{32}b_{22}) & (a_{31}b_{13} + a_{32}b_{23}) & (a_{31}b_{14} + a_{32}b_{24}) \end{bmatrix}. \end{aligned}$$

13.7 Inverse Matrix

A square matrix $\mathbf{A}_{nn}$ that is *invertible* satisfies the condition:

$$\mathbf{A}_{nn}\mathbf{A}_{nn}^{-1} = \mathbf{A}_{nn}^{-1}\mathbf{A}_{nn} = \mathbf{I}_n,$$

where $\mathbf{A}_{nn}^{-1}$ is unique, and is the *inverse matrix* of $\mathbf{A}_{nn}$. For example, given

$$\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 5 & 4 \end{bmatrix}$$

then

$$\mathbf{A}^{-1} = \begin{bmatrix} 4 & -3 \\ -5 & 4 \end{bmatrix}$$

because

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 4 & 3 \\ 5 & 4 \end{bmatrix}\begin{bmatrix} 4 & -3 \\ -5 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

A square matrix whose determinant is 0, cannot have an inverse, and is known as a *singular matrix*.

We now require a way to compute $\mathbf{A}^{-1}$, which is rather easy.

Consider two linear equations:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}. \tag{13.10}$$

Let the inverse of

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

be

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

therefore,

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{13.11}$$

From (13.11) we have

$$ae + cf = 1 \tag{13.12}$$
$$be + df = 0 \tag{13.13}$$
$$ag + ch = 0 \tag{13.14}$$
$$bg + dh = 1. \tag{13.15}$$

Multiply (13.12) by d and (13.13) by c, and subtract:

$$\begin{aligned} ade + cdf &= d \\ bce + cdf &= 0 \\ ade - bce &= d \end{aligned}$$

therefore,

$$e = \frac{d}{ad - bc}.$$

Multiply (13.12) by b and (13.13) by a, and subtract:

$$\begin{aligned} abe + bcf &= b \\ abe + adf &= 0 \\ adf - bcf &= -b \end{aligned}$$

therefore,

$$f = \frac{-b}{ad - bc}.$$

Multiply (13.14) by d and (13.15) by c, and subtract:

$$\begin{aligned} adg + cdh &= 0 \\ bcg + cdh &= c \\ adg - bcg &= -c \end{aligned}$$

therefore,

$$g = \frac{-c}{ad - bc}.$$

Multiply (13.14) by b and (13.15) by a, and subtract:

$$\begin{aligned} abg + bch &= 0 \\ abg + adh &= a \\ adh - bch &= a \end{aligned}$$

therefore,

$$h = \frac{a}{ad - bc}.$$

We now have values for e, f, g and h, which are the elements of the inverse matrix. Consequently, given

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \begin{bmatrix} e & f \\ g & h \end{bmatrix},$$

then

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

The inverse matrix permits us to solve a pair or linear equations as follows. Starting with

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{A} \begin{bmatrix} x \\ y \end{bmatrix}$$

multiply both sides by the inverse matrix:

$$\begin{aligned} \mathbf{A}^{-1} \begin{bmatrix} x' \\ y' \end{bmatrix} &= \mathbf{A}^{-1}\mathbf{A} \begin{bmatrix} x \\ y \end{bmatrix} \\ \mathbf{A}^{-1} \begin{bmatrix} x' \\ y' \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \\ \begin{bmatrix} x \\ y \end{bmatrix} &= \mathbf{A}^{-1} \begin{bmatrix} x' \\ y' \end{bmatrix} \\ \begin{bmatrix} x \\ y \end{bmatrix} &= \frac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}. \end{aligned}$$

Although the elements of $\mathbf{A}^{-1}$ come from $\mathbf{A}$, the relationship is not obvious. However, if $\mathbf{A}$ is transposed, a pattern is revealed. Given

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{then} \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

and placing $\mathbf{A}^{-1}$ alongside $\mathbf{A}^{\mathrm{T}}$, we have

$$\mathbf{A}^{-1} = \begin{bmatrix} e & f \\ g & h \end{bmatrix} \quad \text{and} \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

The elements of $\mathbf{A}^{-1}$ share a common denominator (det $\mathbf{A}$), which is placed outside the matrix, therefore, the matrix elements are taken from $\mathbf{A}^{\mathrm{T}}$ as follows. For any entry a_{ij} in $\mathbf{A}^{-1}$, mask out the ith row and jth column in $\mathbf{A}^{\mathrm{T}}$, and the remaining entry is copied to the ijth entry in $\mathbf{A}^{-1}$. In the case of e, it is d. For f, it is b, with a sign reversal. For g, it is c, with a sign reversal, and for h, it is a. The sign change is computed by the same formula used with determinants:

$$(-1)^{i+j}.$$

which generates this pattern:

$$\begin{bmatrix} + & - \\ - & + \end{bmatrix}.$$

You may be wondering what happens when a 3×3 matrix is inverted. Well, the same technique is used, but when the ith row and jth column in $\mathbf{A}^{\mathrm{T}}$ is masked out, it leaves behind a 2×2 determinant, whose value is copied to the ijth entry in $\mathbf{A}^{-1}$, with the appropriate sign change. We investigate this later on.

Let's illustrate this with an example. Given

$$\begin{aligned} 42 &= 6x + 2y \\ 28 &= 2x + 3y \end{aligned}$$

let

$$\mathbf{A} = \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}$$

then $\det \mathbf{A} = 14$, therefore,

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &= \frac{1}{14} \begin{bmatrix} 3 & -2 \\ -2 & 6 \end{bmatrix} \begin{bmatrix} 42 \\ 28 \end{bmatrix} \\ &= \frac{1}{14} \begin{bmatrix} 70 \\ 84 \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ 6 \end{bmatrix}. \end{aligned}$$

which is the solution.

Now let's investigate how to invert a 3×3 matrix. Given three simultaneous equations in three unknowns:

$$\begin{aligned} x' &= ax + by + cz \\ y' &= dx + ey + fz \\ z' &= gx + hy + jz \end{aligned}$$

they can be written using matrices as follows:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & j \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{A} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Let

$$\mathbf{A}^{-1} = \begin{bmatrix} l & m & n \\ p & q & r \\ s & t & u \end{bmatrix}$$

therefore,

$$\begin{bmatrix} l & m & n \\ p & q & r \\ s & t & u \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & j \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{13.16}$$

From (13.16) we can write:

$$la + md + ng = 1 \tag{13.17}$$
$$lb + me + nh = 0 \tag{13.18}$$
$$lc + mf + nj = 0. \tag{13.19}$$

Multiply (13.17) by e and (13.18) by d, and subtract:

$$\begin{aligned} ael + dem + egn &= e \\ bdl + dem + dhn &= 0 \\ ael - bdl + egn - dhn &= e \\ l(ae - bd) + n(eg - dh) &= e. \end{aligned} \tag{13.20}$$

Multiply (13.18) by f and (13.19) by e, and subtract:

$$\begin{aligned} bfl + efm + fhn &= 0 \\ cel + efm + ejn &= 0 \\ bfl - cel + fhn - ejn &= 0 \\ l(bf - ce) + n(fh - ej) &= 0. \end{aligned} \tag{13.21}$$

Multiply (13.20) by $(fh - ej)$ and (13.21) by $(eg - dh)$, and subtract:

$$\begin{aligned} l(ae - bd)(fh - ej) + n(eg - dh)(fh - ej) &= e(fh - ej) \\ l(bf - ce)(eg - dh) + n(eg - dh)(fh - ej) &= 0 \\ l(ae - bd)(fh - ej) - l(bf - ce)(eg - dh) &= efh - e^2 j \\ l(aefh - ae^2 j - bdfh + bdej - befg + bdfh + ce^2 g - cdeh) &= efh - e^2 j \\ l(aefh - ae^2 j + bdej - befg + ce^2 g - cdeh) &= efh - e^2 j \\ l(afh + bdj + ceg - aej - cdh - bfg) &= fh - ej \\ l(aej + bfg + cdh - afh - bdj - ceg) &= ej - fh \end{aligned}$$

but $(aej + bfg + cdh - afh - bdj - ceg)$ is the Sarrus expansion for $\det \mathbf{A}$, therefore

$$l = \frac{ej - fh}{\det \mathbf{A}}.$$

An exhaustive algebraic analysis reveals:

$$l = \frac{ej - fh}{\det \mathbf{A}}, \qquad m = -\frac{bj - ch}{\det \mathbf{A}}, \qquad n = \frac{bf - ce}{\det \mathbf{A}},$$
$$p = -\frac{dj - gf}{\det \mathbf{A}}, \qquad q = \frac{aj - gc}{\det \mathbf{A}}, \qquad r = -\frac{af - dc}{\det \mathbf{A}},$$
$$s = \frac{dh - ge}{\det \mathbf{A}}, \qquad t = -\frac{ah - gb}{\det \mathbf{A}}, \qquad u = \frac{ae - bd}{\det \mathbf{A}},$$

where

$$\mathbf{A}^{-1} = \begin{bmatrix} l & m & n \\ p & q & r \\ s & t & u \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & j \end{bmatrix}.$$

However, there does not appear to be an obvious way of deriving $\mathbf{A}^{-1}$ from $\mathbf{A}$. But, as we discovered with the 2×2 matrix, the transpose $\mathbf{A}^{\mathrm{T}}$ resolves the problem:

$$\mathbf{A}^{-1} = \begin{bmatrix} l & m & n \\ p & q & r \\ s & t & u \end{bmatrix}, \quad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & j \end{bmatrix}.$$

The elements for $\mathbf{A}^{-1}$ share a common denominator ($\det \mathbf{A}$), which is placed outside the matrix, therefore, the matrix elements are taken from $\mathbf{A}^{\mathrm{T}}$ as follows. For any entry a_{ij} in $\mathbf{A}^{-1}$, mask out the ith row and jth column in $\mathbf{A}^{\mathrm{T}}$, and the remaining elements, in the form of a 2×2 determinant, is copied to the ijth entry in $\mathbf{A}^{-1}$. In the case of l, it is $(ej - hf)$. For m, it is $(bj - hc)$, with a sign reversal, and for n, it is $(bf - ec)$. The sign change is computed by the same formula used with determinants:

$$(-1)^{i+j},$$

which generates the pattern:

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}.$$

With the above *aide-mémoire*, it is easy to write down the inverse matrix:

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} (ej - fh) & -(bj - ch) & (bf - ce) \\ -(dj - gf) & (aj - gc) & -(af - dc) \\ (dh - ge) & -(ah - gb) & (ae - bd) \end{bmatrix}.$$

This technique is known as the *Laplacian expansion* or the *cofactor expansion*, after Pierre–Simon Laplace. The matrix of minor determinants is called the *cofactor matrix* of $\mathbf{A}$, which permits the inverse matrix to be written as:

$$\mathbf{A}^{-1} = \frac{(\text{cofactor matrix of } \mathbf{A})^{\mathrm{T}}}{\det \mathbf{A}}.$$

Let's illustrate this solution with an example. Given

$$\begin{aligned} 18 &= 2x + 2y + 2z \\ 20 &= x + 2y + 3z \\ 7 &= y + z \end{aligned}$$

therefore,

$$\begin{bmatrix} 18 \\ 20 \\ 7 \end{bmatrix} = \begin{bmatrix} 2\ 2\ 2 \\ 1\ 2\ 3 \\ 0\ 1\ 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{A} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

and

$$\det \mathbf{A} = 4 + 2 - 2 - 6 = -2$$

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 2\ 1\ 0 \\ 2\ 2\ 1 \\ 2\ 3\ 1 \end{bmatrix}$$

therefore,

$$\mathbf{A}^{-1} = -\tfrac{1}{2} \begin{bmatrix} -1 & 0 & 2 \\ -1 & 2 & -4 \\ 1 & -2 & 2 \end{bmatrix}$$

and

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = -\tfrac{1}{2} \begin{bmatrix} -1 & 0 & 2 \\ -1 & 2 & -4 \\ 1 & -2 & 2 \end{bmatrix} \begin{bmatrix} 18 \\ 20 \\ 7 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

which is the solution.

13.7.1 Inverting a Pair of Matrices

Having seen how to invert a single matrix, let's investigate how to invert of a pair of matrices.

Given two matrices $\mathbf{T}$ and $\mathbf{R}$, the product $\mathbf{TR}$ and its inverse $(\mathbf{TR})^{-1}$ must equal the identity matrix $\mathbf{I}$:

$$(\mathbf{TR})(\mathbf{TR})^{-1} = \mathbf{I}$$

and multiplying throughout by $\mathbf{T}^{-1}$ we have

$$\begin{aligned} \mathbf{T}^{-1}\mathbf{TR}(\mathbf{TR})^{-1} &= \mathbf{T}^{-1} \\ \mathbf{R}(\mathbf{TR})^{-1} &= \mathbf{T}^{-1}. \end{aligned}$$

Multiplying throughout by $\mathbf{R}^{-1}$ we have

$$\begin{aligned}\mathbf{R}^{-1}\mathbf{R}(\mathbf{TR})^{-1} &= \mathbf{R}^{-1}\mathbf{T}^{-1}\\ (\mathbf{TR})^{-1} &= \mathbf{R}^{-1}\mathbf{T}^{-1}.\end{aligned}$$

Therefore, if $\mathbf{T}$ and $\mathbf{R}$ are invertible, then

$$(\mathbf{TR})^{-1} = \mathbf{R}^{-1}\mathbf{T}^{-1}.$$

Generalising this result to a triple product such as $\mathbf{STR}$ we can reason that

$$(\mathbf{STR})^{-1} = \mathbf{R}^{-1}\mathbf{T}^{-1}\mathbf{S}^{-1}.$$

13.8 Orthogonal Matrix

A matrix is *orthogonal* if its transpose is also its inverse, i.e., matrix $\mathbf{A}$ is orthogonal if

$$\mathbf{A}^{\mathrm{T}} = \mathbf{A}^{-1}.$$

For example,

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

and

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

and

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

which implies that $\mathbf{A}^{\mathrm{T}} = \mathbf{A}^{-1}$.

The following matrix is also orthogonal

$$\mathbf{A} = \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix}$$

because

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix}$$

and

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Orthogonal matrices play an important role in rotations because they leave the origin fixed and preserve all angles and distances. Consequently, an object's geometric integrity is maintained after a rotation, which is why an orthogonal transform is known as a *rigid motion* transform.

13.9 Diagonal Matrix

A *diagonal matrix* is a square matrix whose elements are zero, apart from its diagonal:

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix}.$$

The determinant of a diagonal matrix must be

$$\det \mathbf{A} = a_{11} \times a_{22} \times \cdots a_{nn}.$$

Here is a diagonal matrix with its determinant

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$
$$\det \mathbf{A} = 2 \times 3 \times 4 = 24.$$

The identity matrix $\mathbf{I}$ is a diagonal matrix with a determinant of 1.

13.10 Worked Examples

13.10.1 Matrix Inversion

Invert $\mathbf{A}$ and show that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_2$.

$$\mathbf{A} = \begin{bmatrix} 3 & 5 \\ 2 & 4 \end{bmatrix}.$$

Solution: Using

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

then $\det \mathbf{A} = 2$, and

$$\mathbf{A}^{-1} = \tfrac{1}{2} \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}.$$

Calculating $\mathbf{A}\mathbf{A}^{-1}$:

$$\mathbf{A}\mathbf{A}^{-1} = \tfrac{1}{2} \begin{bmatrix} 3 & 5 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix} = \tfrac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

13.10.2 Identity Matrix

Invert $\mathbf{A}$ and show that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_3$.

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 1 \\ 5 & 6 & 7 \end{bmatrix}.$$

Solution: Using Sarrus's rule for $\det \mathbf{A}$:

$$\det \mathbf{A} = 28 + 15 + 24 - 40 - 12 - 21 = -6.$$

Therefore,

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 2 & 1 & 5 \\ 3 & 2 & 6 \\ 4 & 1 & 7 \end{bmatrix}$$

$$\mathbf{A}^{-1} = -\tfrac{1}{6} \begin{bmatrix} (14-6) & -(21-24) & (3-8) \\ -(7-5) & (14-20) & -(2-4) \\ (6-10) & -(12-15) & (4-3) \end{bmatrix}$$

$$= -\tfrac{1}{6} \begin{bmatrix} 8 & 3 & -5 \\ -2 & -6 & 2 \\ -4 & 3 & 1 \end{bmatrix}$$

and

$$\mathbf{A}\mathbf{A}^{-1} = -\tfrac{1}{6}\begin{bmatrix} 2\,3\,4 \\ 1\,2\,1. \\ 5\,6\,7 \end{bmatrix}\begin{bmatrix} 8 & 3 & -5 \\ -2 & -6 & 2 \\ -4 & 3 & 1 \end{bmatrix}$$
$$= -\tfrac{1}{6}\begin{bmatrix} -6 & 0 & 0 \\ 0 & -6 & 0 \\ 0 & 0 & -6 \end{bmatrix} = \begin{bmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \end{bmatrix}.$$

13.10.3 Solving Two Equations Using Matrices

Solve the following equations using matrices.

$$20 = 2x + 3y$$
$$36 = 7x + 2y.$$

Solution: Let

$$\mathbf{A} = \begin{bmatrix} 2\,3 \\ 7\,2 \end{bmatrix}$$

therefore, $\det \mathbf{A} = -17$, and

$$\mathbf{A}^{-1} = -\tfrac{1}{17}\begin{bmatrix} 2 & -3 \\ -7 & 2 \end{bmatrix}$$

therefore,

$$\begin{bmatrix} x \\ y \end{bmatrix} = -\tfrac{1}{17}\begin{bmatrix} 2 & -3 \\ -7 & 2 \end{bmatrix}\begin{bmatrix} 20 \\ 36 \end{bmatrix}$$
$$= -\tfrac{1}{17}\begin{bmatrix} 40 - 108 \\ -140 + 72 \end{bmatrix}$$
$$= -\tfrac{1}{17}\begin{bmatrix} -68 \\ -68 \end{bmatrix}$$
$$= \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

therefore, $x = y = 4$.

13.10.4 Solving Three Equations Using Matrices

Solve the following equations using matrices.

$$\begin{aligned} 10 &= 2x + y - z \\ 13 &= -x - y + z \\ 28 &= -x + 2y + z. \end{aligned}$$

Solution: Let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}.$$

Using Sarrus's rule for det $\mathbf{A}$:

$$\det \mathbf{A} = -2 - 1 + 2 + 1 - 4 + 1 = -3.$$

Therefore,

$$\begin{aligned} \mathbf{A}^{\mathrm{T}} &= \begin{bmatrix} 2 & -1 & -1 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} \\ \mathbf{A}^{-1} &= -\tfrac{1}{3} \begin{bmatrix} (-1-2) & -(1+2) & (1-1) \\ -(-1+1) & (2-1) & -(2-1) \\ (-2-1) & -(4+1) & (-2+1) \end{bmatrix} \\ &= -\tfrac{1}{3} \begin{bmatrix} -3 & -3 & 0 \\ 0 & 1 & -1 \\ -3 & -5 & -1 \end{bmatrix} \end{aligned}$$

therefore,

$$\begin{aligned} \begin{bmatrix} x \\ y \\ z \end{bmatrix} &= -\tfrac{1}{3} \begin{bmatrix} -3 & -3 & 0 \\ 0 & 1 & -1 \\ -3 & -5 & -1 \end{bmatrix} \begin{bmatrix} 10 \\ 13 \\ 28 \end{bmatrix} \\ &= -\tfrac{1}{3} \begin{bmatrix} -30 - 39 \\ 13 - 28 \\ -30 - 65 - 28 \end{bmatrix} \\ &= -\tfrac{1}{3} \begin{bmatrix} -69 \\ -15 \\ -123 \end{bmatrix} \\ &= \begin{bmatrix} 23 \\ 5 \\ 41 \end{bmatrix} \end{aligned}$$

therefore, $x = 23, \quad y = 5, \quad z = 41.$

13.10.5 Solving Two Complex Equations

Solve the following complex equations using matrices.

$$\begin{aligned} 7 + i8 &= 2x + y \\ -4 - i &= x - 2y. \end{aligned}$$

Solution: Let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix}$$

therefore, $\det \mathbf{A} = -5$, and

$$\begin{aligned} \mathbf{A}^{\mathrm{T}} &= \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} \\ \mathbf{A}^{-1} &= -\tfrac{1}{5}\begin{bmatrix} -2 & -1 \\ -1 & 2 \end{bmatrix} \end{aligned}$$

therefore,

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &= -\tfrac{1}{5}\begin{bmatrix} -2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 7 + i8 \\ -4 - i \end{bmatrix} \\ &= -\tfrac{1}{5}\begin{bmatrix} -14 - i16 + 4 + i \\ -7 - i8 - 8 - i2 \end{bmatrix} \\ &= -\tfrac{1}{5}\begin{bmatrix} -10 - i15 \\ -15 - i10 \end{bmatrix} \\ &= \begin{bmatrix} 2 + i3 \\ 3 + i2 \end{bmatrix} \end{aligned}$$

therefore, $x = 2 + i3, \quad y = 3 + i2.$

13.10.6 Solving Three Complex Equations

Solve the following complex equations using matrices.

$$\begin{aligned} 0 &= x + y - z \\ 3 + i3 &= 2x - y + z \\ -5 - i5 &= -x + y - 2z. \end{aligned}$$

Solution: Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 \\ 2 & -1 & 1 \\ -1 & 1 & -2 \end{bmatrix}$$

therefore, $\det \mathbf{A} = 2 - 1 - 2 + 1 - 1 + 4 = 3$, and

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 1 & 2 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -2 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \tfrac{1}{3} \begin{bmatrix} (2-1) & -(-2+1) & 0 \\ -(-4+1) & (-2-1) & -(1+2) \\ (2-1) & -(1+1) & (-1-2) \end{bmatrix}$$

therefore,

$$\begin{aligned} \begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \tfrac{1}{3} \begin{bmatrix} 1 & 1 & 0 \\ 3 & -3 & -3 \\ 1 & -2 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ 3 + i3 \\ -5 - i5 \end{bmatrix} \\ &= \tfrac{1}{3} \begin{bmatrix} 3 + i3 \\ -9 - i9 + 15 + i15 \\ -6 - i6 + 15 + i15 \end{bmatrix} \\ &= \begin{bmatrix} 1 + i \\ 2 + i2 \\ 3 + i3 \end{bmatrix} \end{aligned}$$

therefore, $x = 1 + i, \quad y = 2 + i2, \quad z = 3 + i3.$

13.10.7 Solving Two Complex Equations

Solve the following complex equations using matrices.

$$\begin{aligned} 3 + i5 &= ix + 2y \\ 5 + i &= 3x - iy. \end{aligned}$$

Solution: Let

$$\mathbf{A} = \begin{bmatrix} i & 2 \\ 3 & -i \end{bmatrix}$$

therefore, set $\mathbf{A} = 1 - 6 = -5$, and

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} i & 3 \\ 2 & -i \end{bmatrix}$$
$$\mathbf{A}^{-1} = -\tfrac{1}{5}\begin{bmatrix} -i & -2 \\ -3 & i \end{bmatrix}$$

therefore,

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &= -\tfrac{1}{5}\begin{bmatrix} -i & -2 \\ -3 & i \end{bmatrix}\begin{bmatrix} 3+i5 \\ 5+i \end{bmatrix} \\ &= -\tfrac{1}{5}\begin{bmatrix} -i3+5-10-i2 \\ -9-i15+i5-1 \end{bmatrix} \\ &= -\tfrac{1}{5}\begin{bmatrix} -5-i5 \\ -10-i10 \end{bmatrix} \\ &= \begin{bmatrix} 1+i \\ 2+i2 \end{bmatrix} \end{aligned}$$

therefore, $x = 1 + i, \quad y = 2 + i2$.

13.10.8 Solving Three Complex Equations

Solve the following complex equations using matrices.

$$\begin{aligned} 6 + i2 &= ix + 2y - iz \\ -2 + i6 &= 2x - iy + i2z \\ 2 + i10 &= i2x + iy + 2z. \end{aligned}$$

Solution: Let

$$\mathbf{A} = \begin{bmatrix} i & 2 & -i \\ 2 & -i & i2 \\ i2 & i & 2 \end{bmatrix}$$

therefore, $\det \mathbf{A} = 2 - 8 + 2 + i2 + i2 - 8 = -12 + i4$, and

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} i & 2 & i2 \\ 2 & -i & i \\ -i & i2 & 2 \end{bmatrix}$$
$$\mathbf{A}^{-1} = \frac{1}{-12+i4}\begin{bmatrix} (-i2+2) & -(4-1) & (i4+1) \\ -(4+4) & (i2-2) & -(-2+i2) \\ (i2-2) & -(-1-i4) & (1-4) \end{bmatrix}$$

$$= \frac{1}{-12+i4}\begin{bmatrix} 2-i2 & -3 & 1+i4 \\ -8 & -2+i2 & 2-i2 \\ -2+i2 & 1+i4 & -3 \end{bmatrix}$$

therefore,

$$\begin{aligned}
\begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \frac{1}{-12+i4}\begin{bmatrix} 2-i2 & -3 & 1+i4 \\ -8 & -2+i2 & 2-i2 \\ -2+i2 & 1+i4 & -3 \end{bmatrix}\begin{bmatrix} 6+i2 \\ -2+i6 \\ 2+i10 \end{bmatrix} \\
&= \frac{1}{-12+i4}\begin{bmatrix} (2-i2)(6+i2)-3(-2+i6)+(1+i4)(2+i10) \\ -8(6+i2)+(-2+i2)(-2+i6)+(2-i2)(2+i10) \\ (-2+i2)(6+i2)+(1+i4)(-2+i6)-3(2+i10) \end{bmatrix} \\
&= \frac{1}{-12+i4}\begin{bmatrix} 12+i4-i12+4+6-i18+2+i10+i8-40 \\ -48-i16+4-i12-i4-12+4+i20-i4+20 \\ -12-i4+i12-4-2+i6-i8-24-6-i30 \end{bmatrix} \\
&= \frac{1}{-12+i4}\begin{bmatrix} -16-i8 \\ -32-i16 \\ -48-i24 \end{bmatrix}
\end{aligned}$$

multiply by the conjugate of $-12+i4$:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{-12-i4}{160}\begin{bmatrix} -16-i8 \\ -32-i16 \\ -48-i24 \end{bmatrix}$$

therefore,

$$\begin{aligned}
x &= \tfrac{1}{160}(-12-i4)(-16-i8) \\
&= \tfrac{1}{160}(192+i64+i96-32) \\
&= \tfrac{1}{160}(160+i160) = 1+i \\
y &= \tfrac{1}{160}(-12-i4)(-32-i16) \\
&= \tfrac{1}{160}(384+i128+i192-64) \\
&= \tfrac{1}{160}(320+i320) = 2+i2 \\
z &= \tfrac{1}{160}(-12-i4)(-48-i24) \\
&= \tfrac{1}{160}(576+i192+i288-96) \\
&= \tfrac{1}{160}(480+i480) = 3+i3
\end{aligned}$$

therefore, $x = 1+i, \quad y = 2+i2, \quad z = 3+i3.$

Chapter 14
Geometric Matrix Transforms

14.1 Introduction

Geometric matrix transforms are an intuitive way of defining and building geometric operations such as scale, translate, reflect, shear and rotate. In 2D, such operations are generally associated with images and text, and widely used in internet browsers, image-processing software, smart phones and watches. In 3D, they are used in computer games, computer animation, film special effects, virtual reality and scientific visualisation. They have proved so useful that they are incorporated in hardware to provide the highest possible execution speeds and real-time performance.

In this chapter, we build upon the ideas of matrices described in the previous chapter, and provide a coherent framework for describing transforms in two and three dimensions.

14.2 Matrix Transforms

The general 2D transform is

$$\begin{aligned} x' &= ax + by \\ y' &= cx + dy \end{aligned} \tag{14.1}$$

or in matrix form:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

where the values of a, b, c and d determine the type of transform. Let's examine 2D transforms and generalise their application to 3D, and start with the translate transform, as this reveals a fundamental problem with matrices.

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_14

14.2.1 2D Translation

The translate transform is described by

$$x' = x + t_x$$
$$y' = y + t_y$$

where the point (x, y) is translated by (t_x, t_y). Modifying (14.1), this becomes

$$x' = ax + by + t_x$$
$$y' = cx + dy + t_y$$

where $a = d = 1$, and $b = c = 0$. However, this does not appear to have a single matrix representation, due to the addition of t_x and t_y. Fortunately, homogeneous coordinates come to the rescue, and support any type of transform incorporating addition or subtraction. The idea is to solve a 2D problem in 3D, where any point (x, y) becomes $(x, y, 1)$, i.e. the z-coordinate equals 1. Rewriting (14.1) in 3D, we have

$$x' = ax + by + t_x$$
$$y' = cx + dy + t_y$$
$$1 = 0x + 0y + 1$$

or in matrix form:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & t_x \\ c & d & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

The 2D translation transform is

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

Figure 14.1 shows a shape translated by (5, 2) using this matrix:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

As we are only interested in (x', y'), the z-coordinate is ignored.

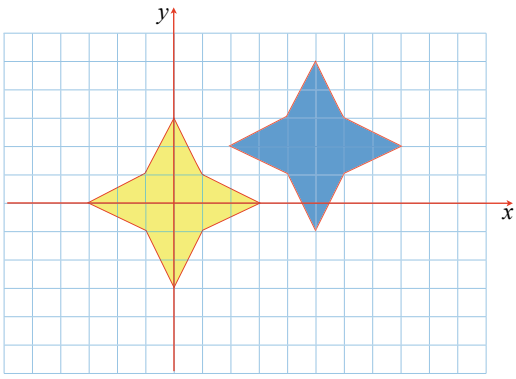

Fig. 14.1 The blue shape is the translated yellow shape

14.2.2 2D Scaling

2D scaling is achieved using

$$x' = s_x x$$
$$y' = s_y y$$

or as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

The homogeneous form is maintained as we often have to combine different matrices. Figure 14.2 shows the effect of scaling a shape by $\times 2$ horizontally, and $\times 1.5$ vertically.

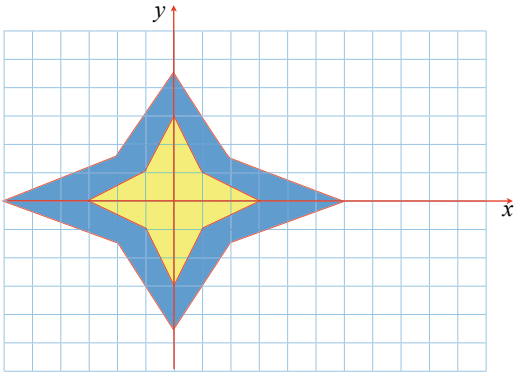

Fig. 14.2 Asymmetric scaling relative to the origin

The scaling action is relative to the origin, i.e. the point (0, 0) remains unchanged. All other points move away from the origin when the scale factor exceeds 1, or towards the origin when it is less than 1. To scale relative to another point $(p_x,\ p_y)$ we first subtract $(p_x,\ p_y)$ from $(x,\ y)$. This effectively makes the reference point $(p_x,\ p_y)$ the new origin. Second, we perform the scaling operation relative to the new origin, and third, add $(p_x,\ p_y)$ back to the new $(x,\ y)$ to compensate for the original subtraction. Algebraically this is

$$\begin{aligned} x' &= s_x(x - p_x) + p_x \\ y' &= s_y(y - p_y) + p_y \end{aligned}$$

which simplifies to

$$\begin{aligned} x' &= s_x x + p_x(1 - s_x) \\ y' &= s_y y + p_y(1 - s_y) \end{aligned}$$

or as a homogeneous matrix

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & p_x(1 - s_x) \\ 0 & s_y & p_y(1 - s_y) \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{14.2}$$

Figure 14.3 shows a scale of $\times 2$ horizontally, and $\times 1.5$ vertically relative to the point $(-3, 0)$ using this matrix:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 1.5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

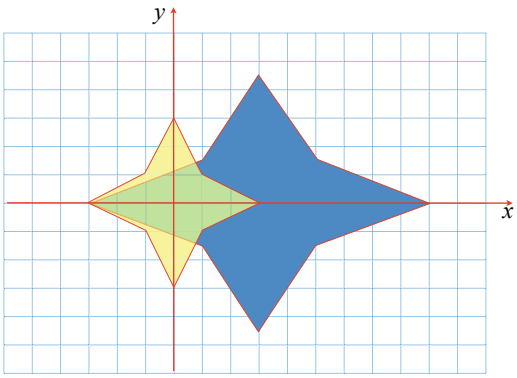

Fig. 14.3 Asymmetric scaling relative to $(-3, 0)$

14.2.3 2D Reflections

The matrix transform for reflecting about the y-axis is

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{14.3}$$

or about the x-axis

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{14.4}$$

where (14.3) reverses the sign of the x-coordinate, and (14.4) reverses the sign of the y-coordinate. However, to make a reflection about an arbitrary vertical or horizontal axis we need to introduce some more algebraic deception.

To make a reflection about a vertical axis $x = a_x$, we first subtract a_x from the x-coordinate. This effectively makes the line $x = a_x$ coincident with the major y-axis. Next we perform the reflection by reversing the sign of the modified x-coordinate, and finally, we add a_x to the reflected coordinate to compensate for the original subtraction. Algebraically, the three steps are

$$\begin{aligned} x_1 &= x - a_x \\ x_2 &= -(x - a_x) \\ x' &= -(x - a_x) + a_x \end{aligned}$$

which simplifies to

$$\begin{aligned} x' &= -x + 2a_x \\ y' &= y \end{aligned}$$

or as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 2a_x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{14.5}$$

Figure 14.4 shows a shape reflected about the line $x = 3$ using (14.6)

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{14.6}$$

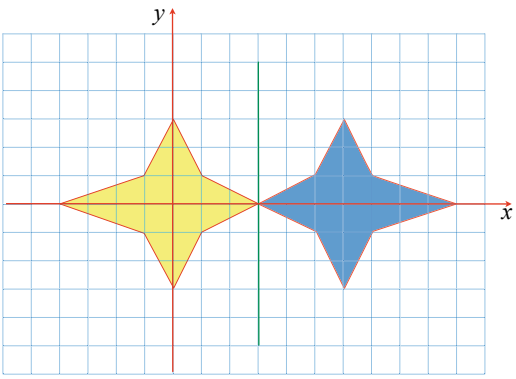

Fig. 14.4 Reflecting a shape about the line $x = 3$

To reflect a point about the line $y = a_y$, the following transform is required:

$$\begin{aligned} x' &= x \\ y' &= -(y - a_y) + a_y \quad = -y + 2a_y \end{aligned}$$

or as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 2a_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

14.2.4 2D Shearing

A shape is sheared by leaning it over at an angle β. Figure 14.5 illustrates the geometry, where we see that the y-coordinates remain unchanged but the x-coordinates

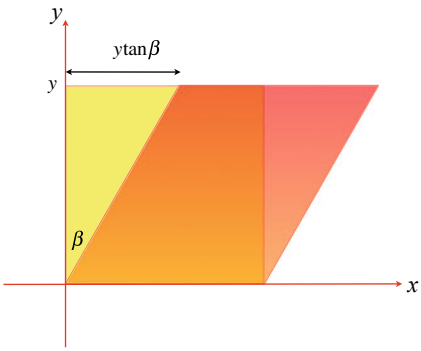

Fig. 14.5 Shearing a shape by β

are a function of y and $\tan\beta$.

$$x' = x + y\tan\beta$$
$$y' = y$$

or as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & \tan\beta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

In this example, the angle β is assumed positive when rotating from the y-axis towards the x-axis.

14.2.5 2D Rotation

Figure 14.6 shows a point $P(x,\ y)$ rotated by an angle β about the origin to $P'(x',\ y')$. From the figure:

$$x' = R\cos(\theta + \beta)$$
$$y' = R\sin(\theta + \beta)$$

and substituting the identities for $\cos(\theta + \beta)$ and $\sin(\theta + \beta)$ we have

$$x' = R(\cos\theta \cdot \cos\beta - \sin\theta \cdot \sin\beta)$$
$$y' = R(\sin\theta \cdot \cos\beta + \cos\theta \cdot \sin\beta)$$
$$x' = R\left(\frac{x}{R}\cos\beta - \frac{y}{R}\sin\beta\right)$$
$$y' = R\left(\frac{y}{R}\cos\beta + \frac{x}{R}\sin\beta\right)$$
$$x' = x\cos\beta - y\sin\beta$$
$$y' = x\sin\beta + y\cos\beta$$

or as a homogeneous matrix

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{14.7}$$

and is the general transform for rotating a point about the origin.

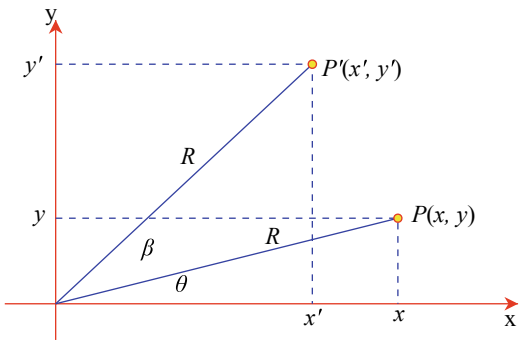

Fig. 14.6 Rotating a point through an angle β

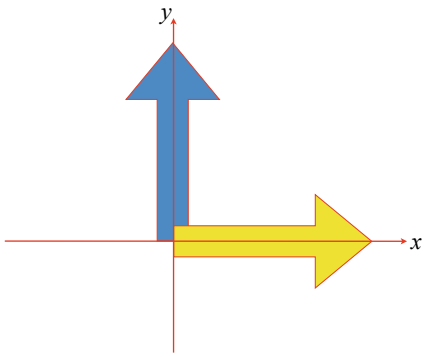

Fig. 14.7 The yellow shape is rotated 90°

Figure 14.7 shows the effect of rotating the yellow arrow by 90° using this matrix:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

When $\beta = 360°$ the matrix becomes the identity matrix, and has a null effect:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

To rotate a point $(x,\ y)$ about an arbitrary point $(p_x,\ p_y)$ we first, subtract $(p_x,\ p_y)$ from $(x,\ y)$. This enables us to perform the rotation about the origin. Second, we perform the rotation, and third, we add $(p_x,\ p_y)$ to compensate for the original subtraction. Here are the steps:

1. Subtract $(p_x,\ p_y)$:

$$x_1 = x - p_x$$
$$y_1 = y - p_y.$$

2. Rotate β about the origin:

$$x_2 = x_1 \cos\beta - y_1 \sin\beta$$
$$y_2 = x_1 \sin\beta + y_1 \cos\beta.$$

3. Add $(p_x,\ p_y)$:

$$x' = x_1 \cos\beta - y_1 \sin\beta + p_x$$
$$y' = x_1 \sin\beta + y_1 \cos\beta + p_y.$$

Simplifying,

$$x' = x\cos\beta - y\sin\beta + p_x(1 - \cos\beta) + p_y \sin\beta$$
$$y' = x\sin\beta + y\cos\beta + p_y(1 - \cos\beta) - p_x \sin\beta$$

and as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & p_x(1-\cos\beta) + p_y\sin\beta \\ \sin\beta & \cos\beta & p_y(1-\cos\beta) - p_x\sin\beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{14.8}$$

To rotate a point 90° about the point $(1,\ 1)$ (14.8) becomes

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 2 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

A simple test is to substitute the point $(2,\ 1)$ for $(x,\ y)$, which is transformed correctly to $(1,\ 2)$.

The algebraic approach in deriving the above transforms is relatively easy. However, it is also possible to use matrices to derive compound transforms, such as a reflection relative to an arbitrary line and scaling and rotation relative to an arbitrary point. These transforms are called *affine*, as parallel lines remain parallel after being transformed. Furthermore, the word "affine" is used to imply that there is a strong geometric *affinity* between the original and transformed shape. One can not always guarantee that angles and lengths are preserved, as the scaling transform can alter these when different x and y scaling factors are used. For completeness, these transforms are repeated from a matrix perspective.

14.2.6 2D Scaling

The strategy used to scale a point $(x,\ y)$ relative to some arbitrary point $(p_x,\ p_y)$ is to first, translate $(-p_x,\ -p_y)$; second, perform the scaling; and third translate $(p_x,\ p_y)$. These three transforms are represented in matrix form as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \big[\,\text{translate}(p_x,\ p_y)\,\big]\big[\,\text{scale}(s_x,\ s_y)\,\big]\big[\,\text{translate}(-p_x,\ -p_y)\,\big]\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

which expands to

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & p_x \\ 0 & 1 & p_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & -p_x \\ 0 & 1 & -p_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

Note the sequence of the transforms, as this often causes confusion. The first transform acting on the point $(x,\ y,\ 1)$ is translate $(-p_x,\ -p_y)$, followed by scale $(s_x,\ s_y)$, followed by translate $(p_x,\ p_y)$. If they are placed in any other sequence, you will discover, like Gauss, that transforms are not commutative!

Now we concatenate these matrices into a single matrix by multiplying them together. This can be done in any sequence, so long as we preserve the original order. Let's start with scale $(s_x,\ s_y)$ and translate $(-p_x,\ -p_y)$ matrices. This produces

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & p_x \\ 0 & 1 & p_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} s_x & 0 & -s_x p_x \\ 0 & s_y & -s_y p_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

and finally

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & p_x(1-s_x) \\ 0 & s_y & p_y(1-s_y) \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

which is the same as the previous transform (14.2).

14.2.7 2D Reflection

A reflection about the y-axis is given by

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

Therefore, using matrices, we can reason that a reflection transform about an arbitrary line $x = a_x$, parallel with the y-axis, is given by

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \big[\,\text{translate}(a_x,\ 0)\,\big]\big[\,\text{reflection}\,\big]\big[\,\text{translate}(-a_x,\ 0)\,\big]\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

which expands to

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & a_x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & -a_x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

Now we concatenate these matrices into a single matrix by multiplying them together. Let's begin by multiplying the reflection and the translate $(-a_x,\ 0)$ matrices together. This produces

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & a_x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 & 0 & a_x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

and finally

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 2a_x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

which is the same as the previous transform (14.5).

14.2.8 2D Rotation About an Arbitrary Point

A rotation about the origin is given by

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

Therefore, using matrices, we can develop a rotation about an arbitrary point $(p_x,\ p_y)$ as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \big[\,\text{translate}(p_x,\ p_y)\,\big]\big[\,\text{rotate}\beta\,\big]\big[\,\text{translate}(-p_x,\ -p_y)\,\big]\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

which expands to

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & p_x \\ 0 & 1 & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -p_x \\ 0 & 1 & -p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

Now we concatenate these matrices into a single matrix by multiplying them together. Let's begin by multiplying the rotate β and the translate $(-p_x, -p_y)$ matrices together. This produces

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & p_x \\ 0 & 1 & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & -\sin\beta & -p_x\cos\beta + p_y\sin\beta \\ \sin\beta & \cos\beta & -p_x\sin\beta - p_y\cos\beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & p_x(1-\cos\beta) + p_y\sin\beta \\ \sin\beta & \cos\beta & p_y(1-\cos\beta) - p_x\sin\beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

which is the same as the previous transform (14.8).

I hope it is now clear to the reader that one can derive all sorts of transforms either algebraically, or by using matrices—it is just a question of convenience.

14.3 3D Transforms

Now we come to transforms in three dimensions, where we apply the same reasoning as in two dimensions. However, translation remains a problem, unless we move the problem to a four-dimensional homogeneous space, which means turning (x, y, z) into $(x, y, z, 1)$. Scaling and translation are basically the same, but in 2D, where we rotated a shape about a point, in 3D, we rotate an object about an axis.

14.3.1 3D Translation

The algebra is so simple for 3D translation, that we can write the homogeneous matrix directly:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

14.3.2 3D Scaling

The algebra for 3D scaling is

$$\begin{aligned} x' &= s_x x \\ y' &= s_y y \\ z' &= s_z z \end{aligned}$$

and as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

The scaling is relative to the origin, but we can arrange for it to be relative to an arbitrary point $(p_x,\ p_y,\ p_z)$ using the following algebra:

$$\begin{aligned} x' &= s_x(x - p_x) + p_x \\ y' &= s_y(y - p_y) + p_y \\ z' &= s_z(z - p_z) + p_z \end{aligned}$$

and as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 & p_x(1 - s_x) \\ 0 & s_y & 0 & p_y(1 - s_y) \\ 0 & 0 & s_z & p_z(1 - s_z) \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

14.3.3 3D Rotation

In two dimensions a shape is rotated about a point, whether it be the origin or some other position. In three dimensions an object is rotated about an axis, whether it be the x-, y- or z-axis, or some arbitrary axis. To begin with, let's look at rotating a vertex about one of the three orthogonal axes; such rotations are called *Euler rotations* after Leonhard Euler.

Recall that a general 2D rotation transform is given by

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

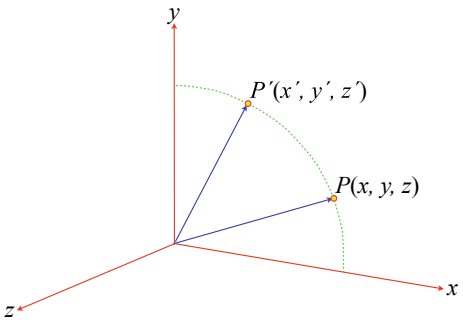

Fig. 14.8 Rotating the point P about the z-axis

which in 3D can be visualised as rotating a point $P(x,\ y,\ z)$ on a plane parallel with the xy-plane as shown in Fig. 14.8. In algebraic terms this is written as

$$\begin{aligned} x' &= x\cos\beta - y\sin\beta \\ y' &= x\sin\beta + y\cos\beta \\ z' &= z \end{aligned}$$

and as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 & 0 \\ \sin\beta & \cos\beta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

which rotates a point about the z-axis.

When rotating about the x-axis, the x-coordinates remain constant whilst the y- and z-coordinates are changed. Algebraically, this is

$$\begin{aligned} x' &= x \\ y' &= y\cos\beta - z\sin\beta \\ z' &= y\sin\beta + z\cos\beta \end{aligned}$$

and as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta & 0 \\ 0 & \sin\beta & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

When rotating about the y-axis, the y-coordinate remains constant whilst the x- and z-coordinates are changed. Algebraically, this is

$$\begin{aligned} x' &= z\sin\beta + x\cos\beta \\ y' &= y \\ z' &= z\cos\beta - x\sin\beta \end{aligned}$$

and as a homogeneous matrix:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & 0 & \sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

Note that the matrix terms do not appear to share the symmetry seen in the previous two matrices. Nothing really has gone wrong, it is just the way the axes are paired together to rotate the coordinates.

The above rotations are also known as *yaw*, *pitch* and *roll*, and great care should be taken with these angles when referring to other books and technical papers. Sometimes a left-handed system of axes is used rather than a right-handed set, and the vertical axis may be the y-axis or the z-axis. Consequently, the matrices representing the rotations can vary greatly. In this book, all Cartesian coordinate systems are right-handed, and the vertical axis is generally the y-axis.

The roll, pitch and yaw angles are defined as follows:

- $roll$ is the angle of rotation about the z-axis,
- $pitch$ is the angle of rotation about the x-axis,
- yaw is the angle of rotation about the y-axis.

Figure 14.9 illustrates these rotations and the sign convention. The homogeneous matrices representing these rotations are as follows:

- rotate $roll$ about the z-axis:

$$\begin{bmatrix} \cos roll & -\sin roll & 0 & 0 \\ \sin roll & \cos roll & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- rotate $pitch$ about the x-axis:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos pitch & -\sin pitch & 0 \\ 0 & \sin pitch & \cos pitch & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

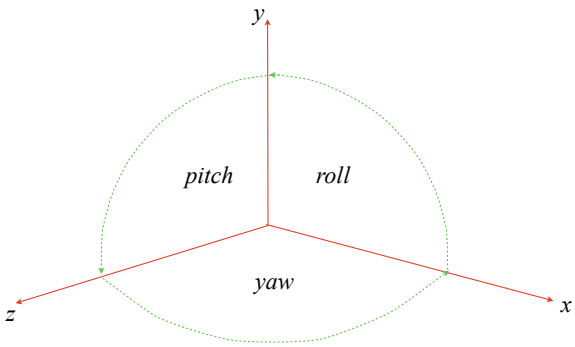

Fig. 14.9 Rotating the point P about the z-axis

- rotate yaw about the y-axis:

$$\begin{bmatrix} \cos yaw & 0 & \sin yaw & 0 \\ 0 & 1 & 0 & 0 \\ -\sin yaw & 0 & \cos yaw & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

A common sequence for applying these rotations is $roll$, $pitch$, yaw, as seen in the following transform:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = [\,yaw\,][\,pitch\,][\,roll\,] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

and if a translation is involved,

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = [\,translate\,][\,yaw\,][\,pitch\,][\,roll\,] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

14.3.4 Rotating About an Axis

The above rotations are relative to the x-, y-, z-axis. Now let's consider rotations about an axis parallel to one of these axes. To begin with, we will rotate about an axis parallel with the z-axis, as shown in Fig. 14.10. The scenario is very reminiscent of the 2D case for rotating a point about an arbitrary point, and the general transform is given by

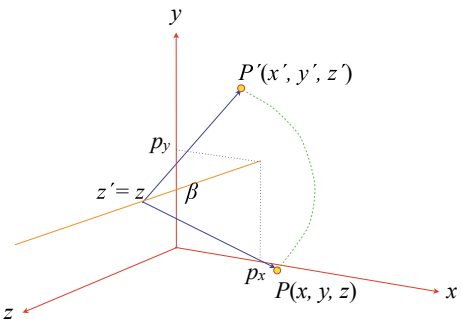

Fig. 14.10 Rotating the point P about an arbitrary axis

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \left[\,\text{translate}(p_x,\ p_y, 0)\,\right]\left[\,\text{rotate}\beta\,\right]\left[\,\text{translate}(-p_x,\ -p_y, 0)\,\right]\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

and the matrix is

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 & p_x(1-\cos\beta) + p_y \sin\beta \\ \sin\beta & \cos\beta & 0 & p_y(1-\cos\beta) - p_x \sin\beta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

Hopefully, you can see the similarity between rotating in 3D and 2D: the x- and y-coordinates are updated while the z-coordinate is held constant. We can now state the other two matrices for rotating about an axis parallel with the x-axis and parallel with the y-axis:

- rotating about an axis parallel with the x-axis:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta & p_y(1-\cos\beta) + p_z \sin\beta \\ 0 & \sin\beta & \cos\beta & p_z(1-\cos\beta) - p_y \sin\beta \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

- rotating about an axis parallel with the y-axis:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & 0 & \sin\beta & p_x(1-\cos\beta) - p_z \sin\beta \\ 0 & 1 & 0 & 0 \\ -\sin\beta & 0 & \cos\beta & p_z(1-\cos\beta) + p_x \sin\beta \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

14.3.5 3D Reflections

Reflections in 3D occur with respect to a plane, rather than an axis. The homogeneous matrix giving the reflection relative to the yz-plane is

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

and the reflection relative to a plane parallel to, and a_x units from the yz-plane is

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 2a_x \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}.$$

14.4 Rotating a Point About an Arbitrary Axis

14.4.1 Matrices

Rotating a point about an arbitrary axis is achieved in various ways. We can employ vectors, analytic geometry, matrices or quaternions. In this example, vectors are used, and Fig. 14.11 shows a view of the geometry associated with the task at hand. For clarification, Fig. 14.12 shows a cross-section and a plan view of the geometry.

The axis of rotation is given by the unit vector:

$$\hat{\mathbf{n}} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

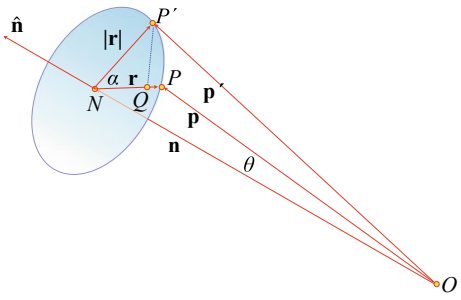

Fig. 14.11 The geometry associated with rotating a point about an arbitrary axis

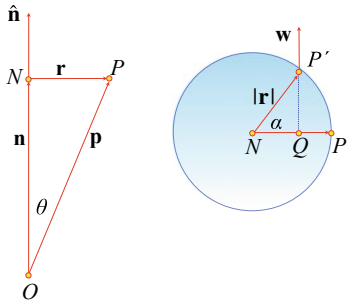

Fig. 14.12 A cross-section and plan view of the geometry

$P(x_p,\ y_p\ z_p)$ is the point to be rotated by angle α to $P'(x'_p,\ y'_p,\ z'_p)$. O is the origin, whilst $\mathbf{p}$ and $\mathbf{p}'$ are position vectors for P and P' respectively. From Figs. 14.11 to 14.12:

$$\mathbf{p}' = \overrightarrow{ON} + \overrightarrow{NQ} + \overrightarrow{QP'}.$$

To find $\overrightarrow{ON}$:

$$\begin{aligned} |\mathbf{n}| &= |\mathbf{p}| \cos\theta \\ \hat{\mathbf{n}} \cdot \mathbf{p} &= |\mathbf{p}| \cos\theta \\ |\mathbf{n}| &= \hat{\mathbf{n}} \cdot \mathbf{p} \\ \mathbf{n} &= \hat{\mathbf{n}}|\mathbf{n}| \\ \mathbf{n} &= \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \mathbf{p}) \end{aligned}$$

therefore,

$$\overrightarrow{ON} = \mathbf{n} = \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \mathbf{p}).$$

To find $\overrightarrow{NQ}$:

$$\overrightarrow{NQ} = \frac{NQ}{NP}\mathbf{r} = \frac{NQ}{NP'}\mathbf{r} = \cos\alpha\ \mathbf{r}$$

but

$$\mathbf{p} = \mathbf{n} + \mathbf{r} = \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \mathbf{p}) + \mathbf{r}$$

therefore,

$$\mathbf{r} = \mathbf{p} - \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \mathbf{p})$$

and

$$\overrightarrow{NQ} = [\mathbf{p} - \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \mathbf{p})] \cos\alpha.$$

To find $\overrightarrow{QP'}$:
Let

$$\hat{\mathbf{n}} \times \mathbf{p} = \mathbf{w}$$

where

$$|\mathbf{w}| = |\hat{\mathbf{n}}||\mathbf{p}| \sin\theta = |\mathbf{p}| \sin\theta$$

but

$$|\mathbf{r}| = |\mathbf{p}| \sin\theta$$

therefore,

$$|\mathbf{w}| = |\mathbf{r}|.$$

Now

$$\frac{QP'}{NP'} = \frac{QP'}{|\mathbf{r}|} = \frac{QP'}{|\mathbf{w}|} = \sin\alpha$$

therefore,

$$\overrightarrow{QP'} = \mathbf{w} \sin\alpha = (\hat{\mathbf{n}} \times \mathbf{p}) \sin\alpha$$

then

$$\mathbf{p}' = \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \mathbf{p}) + [\mathbf{p} - \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \mathbf{p}] \cos\alpha + (\hat{\mathbf{n}} \times \mathbf{p}) \sin\alpha$$

and

$$\mathbf{p}' = \mathbf{p} \cos\alpha + \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \mathbf{p})(1 - \cos\alpha) + (\hat{\mathbf{n}} \times \mathbf{p}) \sin\alpha.$$

Let

$$K = 1 - \cos\alpha$$

then

$$\mathbf{p}' = \mathbf{p} \cos\alpha + \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \mathbf{p})K + (\hat{\mathbf{n}} \times \mathbf{p}) \sin\alpha$$

and

$$\begin{aligned}
\mathbf{p}' &= (x_p\mathbf{i} + y_p\mathbf{j} + z_p\mathbf{k}) \cos\alpha + (a\mathbf{i} + b\mathbf{j} + c\mathbf{k})(ax_p + by_p + cz_p)K \\
&\quad + [(bz_p - cy_p)\mathbf{i} + (cx_p - az_p)\mathbf{j} + (ay_p - bx_p)\mathbf{k}] \sin\alpha \\
\mathbf{p}' &= [x_p \cos\alpha + a(ax_p + by_p + cz_p)K + (bz_p - cy_p) \sin\alpha]\mathbf{i} \\
&\quad + [y_p \cos\alpha + b(ax_p + by_p + cz_p)K + (cx_p - az_p) \sin\alpha]\mathbf{j} \\
&\quad + [z_p \cos\alpha + c(ax_p + by_p + cz_p)K + (ay_p - bx_p) \sin\alpha]\mathbf{k} \\
\mathbf{p}' &= [x_p(a^2K + \cos\alpha) + y_p(abK - c \sin\alpha) + z_p(acK + b \sin\alpha)]\mathbf{i} \\
&\quad + [x_p(abK + c \sin\alpha) + y_p(b^2K + \cos\alpha) + z_p(bcK - a \sin\alpha)]\mathbf{j} \\
&\quad + [x_p(acK - b \sin\alpha) + y_p(bcK + a \sin\alpha) + z_p(c^2K + \cos\alpha)]\mathbf{k}
\end{aligned}$$

and the transform is:

$$\begin{bmatrix} x'_p \\ y'_p \\ z'_p \\ 1 \end{bmatrix} = \begin{bmatrix} a^2K + \cos\alpha & abK - c\sin\alpha & acK + b\sin\alpha & 0 \\ abK + c\sin\alpha & b^2K + \cos\alpha & bcK - a\sin\alpha & 0 \\ acK - b\sin\alpha & bcK + a\sin\alpha & c^2K + \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_p \\ y_p \\ z_p \\ 1 \end{bmatrix}$$

where

$$K = 1 - \cos\alpha.$$

The Worked Examples at the end of this chapter illustrate how this matrix is used.

14.5 Determinant of a Transform

The determinant of the transform (14.9) is $ad - bc$.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{14.9}$$

If we subject the vertices of a unit-square to this transform, we create the situation shown in Fig. 14.13. The vertices of the unit-square are transformed as follows:

$$\begin{aligned} (0,\ 0) &\Rightarrow (0,\ 0) \\ (1,\ 0) &\Rightarrow (a,\ c) \\ (1,\ 1) &\Rightarrow (a+b, c+d) \\ (0,\ 1) &\Rightarrow (b,\ d). \end{aligned}$$

From Fig. 14.13 it can be seen that the area of the transformed unit-square A is given by

$$\begin{aligned} area &= (a+b)(c+d) - B - C - D - E - F - G \\ &= ac + ad + bc + bd - \frac{bd}{2} - bc - \frac{ac}{2} - \frac{bd}{2} - bc - \frac{ac}{2} \\ &= ad - bc \end{aligned}$$

which is the determinant of the transform. But as the area of the original unit-square is 1, the determinant of the transform controls the scaling factor applied to the transformed shape.

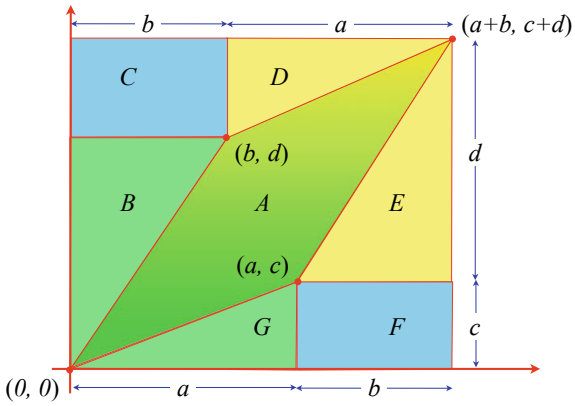

Fig. 14.13 The inner parallelogram is the transformed unit square

Let's examine the determinants of two transforms: The first 2D transform encodes a scaling of 2, and results in an overall area scaling of 4:

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

whose determinant is:

$$\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4.$$

The second 2D transform encodes a scaling of 3 and a translation of (3, 3), and results in an overall area scaling of 9:

$$\begin{bmatrix} 3 & 0 & 3 \\ 0 & 3 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

whose determinant is:

$$3\begin{vmatrix} 3 & 3 \\ 0 & 1 \end{vmatrix} - 0\begin{vmatrix} 0 & 3 \\ 0 & 1 \end{vmatrix} + 0\begin{vmatrix} 0 & 3 \\ 3 & 3 \end{vmatrix} = 9.$$

These two examples demonstrate the extra role played by the elements of a matrix.

14.6 Perspective Projection

In any 3D computer graphic application a database stores a collection of virtual objects in the form of Cartesian coordinates, or other permitted formats. A virtual camera is then located within this *world space* with position and direction using

a compound transform $\mathbf{T}_c$. To capture a perspective view, each point (x, y, z) is transformed to the camera's coordinate system (x_c, y_c, z_c) using the inverse of the compound transform $\mathbf{T}_c^{-1}$.

A virtual camera is directed along its z-axis as shown in Fig. 14.14. Positioned d units along the z-axis is a virtual projection screen, which is used to capture the perspective projection. Figure 14.15 shows that any point $(x_c,\ y_c,\ z_c)$ is transformed to $(x_p,\ y_p,\ d)$. It also shows that the screen's x-axis is pointing in the opposite direction to the camera's x-axis, which can be compensated for by reversing the sign of x_p when it is computed.

Figure 14.15 shows a plan view of the scenario depicted in Figs. 14.14, and 14.16 a side view, which permits us to inspect the geometry and make the following observations:

$$\frac{x_c}{z_c} = \frac{-x_p}{d}$$
$$x_p = \frac{-x_c}{z_c/d}$$

and

$$\frac{y_c}{z_c} = \frac{y_p}{d}$$
$$y_p = \frac{y_c}{z_c/d}.$$

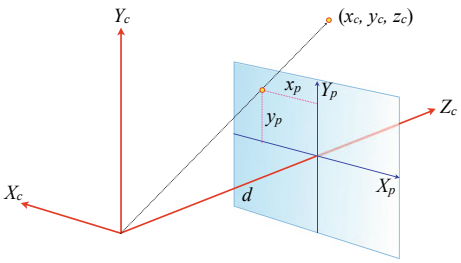

Fig. 14.14 The axial system used for the perspective projection

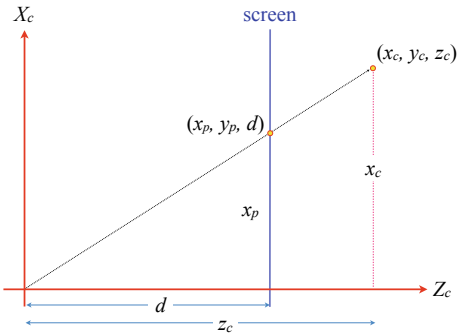

Fig. 14.15 A plan view of the camera's axial system

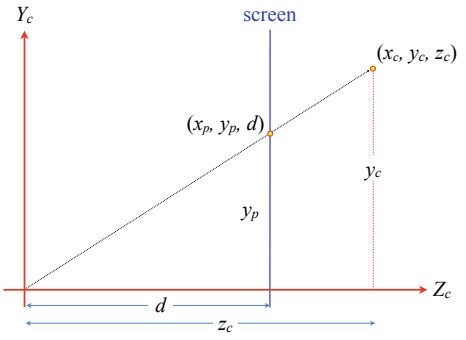

Fig. 14.16 A side view of the camera's axial system

This is expressed in matrix form as

$$\begin{bmatrix} x_p \\ y_p \\ z_p \\ w \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/d & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}.$$

At first, the transform seems strange, but if we multiply this out we get

$$[x_p \quad y_p \quad z_p \quad w]^T = [-x_c \quad y_c \quad z_c \quad z_c/d]^T$$

and if we remember the idea behind homogeneous coordinates, we must divide the terms x_p, y_p, z_p by w to get the scaled terms, which produces

$$\begin{aligned} x_p &= \frac{-x_c}{z_c/d} \\ y_p &= \frac{y_c}{z_c/d} \\ z_p &= \frac{z_c}{z_c/d} = d \end{aligned}$$

which, after all, is rather elegant. The value of d controls the size of the image, and acts like a zoom control. Notice that this transform takes into account the sign change that occurs with the x-coordinate. Some books will leave this sign reversal until the mapping is made to the hardware display coordinates.

14.7 Worked Examples

14.7.1 2D Scale and Translate

$\mathbf{T}_1$ and $\mathbf{T}_2$ translate and scale a 2D point (x, y) to (x', y'). Concatenate them in two possible ways and show that the point $(1, 1)$ is transformed to two different places.

Solution:

$$\mathbf{T}_1 = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 2\ 0\ 0 \\ 0\ 2\ 0 \\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\mathbf{T}_2 = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1\ 0\ 2 \\ 0\ 1\ 2 \\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\mathbf{T}_1\mathbf{T}_2 = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 2\ 0\ 0 \\ 0\ 2\ 0 \\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} 1\ 0\ 2 \\ 0\ 1\ 2 \\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2\ 0\ 4 \\ 0\ 2\ 4 \\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

and the point (1, 1) is transformed to (6, 6).

$$\mathbf{T}_2\mathbf{T}_1 = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1\ 0\ 2 \\ 0\ 1\ 2 \\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} 2\ 0\ 0 \\ 0\ 2\ 0 \\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2\ 0\ 2 \\ 0\ 2\ 2 \\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

and the point (1, 1) is transformed to (4, 4).

14.7.2 2D Rotation

Compute the coordinates of the unit square in Table 14.1 after a rotation of 90°.
Solution: The points are rotated as follows:

Table 14.1 Original and rotated coordinates of the unit square

x	y	x'	y'
0	0	0	0
1	0	0	1
1	1	-1	1
0	1	-1	0

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

14.7.3 Determinant of the Rotate Transform

Using determinants, show that the rotate transform preserves area.
Solution: The determinant of a 2D matrix transform reflects the area change produced by the transform. Therefore, if area is preserved, the determinant must equal 1. Using Sarrus's rule:

$$\left| \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \right| = \cos^2\beta + \sin^2\beta = 1$$

which confirms the role of the determinant.

14.7.4 Determinant of the Shear Transform

Using determinants, show that the shear transform preserves area.
Solution: The determinant of a 2D matrix transform reflects the area change produced by the transform. Therefore, if area is preserved, the determinant must equal 1. Using Sarrus's rule:

$$\left|\begin{bmatrix} 1 & \tan\beta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right| = 1$$

which confirms the role of the determinant.

14.7.5 Yaw, Pitch and Roll Transforms

Using the yaw and pitch transforms in the sequence $yaw \times pitch$, compute how the point (1, 1, 1) is transformed with $yaw = pitch = 90°$.
Solution:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos yaw & 0 & \sin yaw & 0 \\ 0 & 1 & 0 & 0 \\ -\sin yaw & 0 & \cos yaw & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos pitch & -\sin pitch & 0 \\ 0 & \sin pitch & \cos pitch & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

therefore, (1, 1, 1) is transformed to (1, −1, −1).

14.7.6 Rotation About an Arbitrary Axis

Rotate the point (3, 0, 0), 180° about the axis defined by the vector $\mathbf{n} = \mathbf{i} + \mathbf{j} + \mathbf{k}$.
Solution:

$$\begin{bmatrix} x'_p \\ y'_p \\ z'_p \\ 1 \end{bmatrix} = \begin{bmatrix} a^2K + \cos\alpha & abK - c\sin\alpha & acK + b\sin\alpha & 0 \\ abK + c\sin\alpha & b^2K + \cos\alpha & bcK - a\sin\alpha & 0 \\ acK - b\sin\alpha & bcK + a\sin\alpha & c^2K + \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_p \\ y_p \\ z_p \\ 1 \end{bmatrix}$$

where

$$K = 1 - \cos\alpha.$$

Given $\alpha = 180°$, then $K = 1 + 1 = 2$, and $\hat{\mathbf{n}} = \frac{1}{\sqrt{3}}\mathbf{i} + \frac{1}{\sqrt{3}}\mathbf{j} + \frac{1}{\sqrt{3}}\mathbf{k}$. Therefore,

$$\begin{bmatrix} x'_p \\ y'_p \\ z'_p \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} - 1 & \frac{2}{3} & \frac{2}{3} & 0 \\ \frac{2}{3} & \frac{2}{3} - 1 & \frac{2}{3} & 0 \\ \frac{2}{3} & \frac{2}{3} & \frac{2}{3} - 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 \\ 2 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} & 0 \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} & 0 \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

and the point $(3, 0, 0)$ is rotated to $(-1, 2, 2)$.

14.7.7 3D Rotation Transform Matrix

Show that the matrix for rotating a point about an arbitrary axis corresponds to the three matrices for rotating about the x-, y- and z-axis.
Solution:

$$\begin{bmatrix} a^2K + \cos\alpha & abK - c\sin\alpha & acK + b\sin\alpha & 0 \\ abK + c\sin\alpha & b^2K + \cos\alpha & bcK - a\sin\alpha & 0 \\ acK - b\sin\alpha & bcK + a\sin\alpha & c^2K + \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Pitch about the x-axis: $\hat{\mathbf{n}} = \mathbf{i}$, where $a = 1$ and $b = c = 0$; $K = 1 - \cos\alpha$.

$$\text{pitch} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha & 0 \\ 0 & \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Yaw about the y-axis: $\hat{\mathbf{n}} = \mathbf{j}$, where $b = 1$ and $a = c = 0$; $K = 1 - \cos\alpha$.

$$\text{yaw} = \begin{bmatrix} \cos\alpha & 0 & \sin\alpha & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\alpha & 0 & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Roll about the z-axis: $\hat{\mathbf{n}} = \mathbf{k}$, where $c = 1$ and $a = b = 0$; $K = 1 - \cos\alpha$.

$$\text{roll} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 & 0 \\ \sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Table 14.2 Coordinates of a 3D cube

vertex	x_c	y_c	z_c	x_p	y_p
1	0	0	10	0	0
2	10	0	10	20	0
3	10	10	10	20	20
4	0	10	10	0	20
5	0	0	20	0	0
6	10	0	20	10	0
7	10	10	20	10	10
8	0	10	20	0	10

14.7.8 Perspective Projection

Compute the perspective coordinates of a 3D cube stored in Table 14.2 with the projection screen distance $d = 20$. Sketch the result.
Solution: Using the perspective transform:

$$\begin{bmatrix} x_p \\ y_p \\ z_p \\ w \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/d & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}.$$

the perspective coordinates are stored in Table 14.2, and Fig. 14.17 shows a sketch of the result.

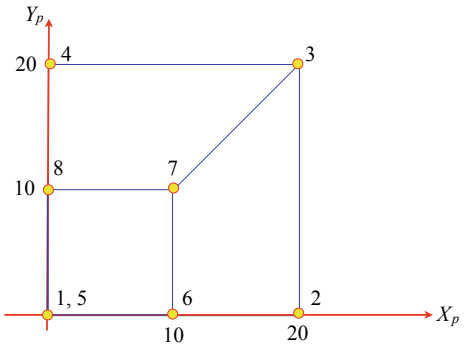

Fig. 14.17 A perspective sketch of a 3D cube

Chapter 15
Calculus: Derivatives

15.1 Introduction

This chapter provides an introduction to differential calculus. It begins with a short description of calculus' origins, and towards the idea of a limiting process. The chapter continues by showing how the derivative is calculated for different functions, and concludes with several worked examples.

15.2 Background

Some quantities, such as the area of a circle or an ellipse, cannot be written precisely, as they incorporate π, which is transcendental. However, an approximate value can be obtained by devising a definition that includes a parameter that is made infinitesimally small. The techniques of limits and infinitesimals have been used in mathematics for over two-thousand years, and paved the way towards today's calculus.

Although the principles of integral calculus were being used by Archimedes (287–212 B.C.) to compute areas, volumes and centres of gravity, it was the English astronomer, physicist and mathematician Isaac Newton (1643–1727) and Gottfried Leibniz who are regarded as the true inventors of modern calculus. Leibniz published his results in 1684, followed by Newton in 1704. However, Newton had been using his *calculus of fluxions* as early as 1665. Since then, calculus has evolved conceptually and in notation.

Up until recently, calculus was described using *infinitesimals*, which are numbers so small, they can be ignored in certain products. However, infinitesimals, no matter how small they are, do not belong to an axiomatic mathematical system, and eventually the French mathematician Augustin-Louis Cauchy, and the German mathematician Karl Weierstrass (1815–1897), showed how they could be replaced by limits.

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_15

15.3 Small Numerical Quantities

The adjective *small* is a relative term, and requires clarification in the context of numbers. For example, if numbers are in the hundreds, and also contain some decimal component, then it seems reasonable to ignore digits after the 3rd decimal place for any quick calculation. For instance,

$$100.000003 \times 200.000006 \approx 20,000$$

and ignoring the decimal part has no significant impact on the general accuracy of the answer, which is measured in tens of thousands.

To develop an algebraic basis for this argument let's divide a number into two parts: a primary part x, and some very small secondary part δx (pronounced *delta x*). In one of the above numbers, $x = 100$ and $\delta x = 0.000003$. Given two such numbers, x_1 and y_1, their product is given by

$$\begin{aligned} x_1 &= x + \delta x \\ y_1 &= y + \delta y \\ x_1 y_1 &= (x + \delta x)(y + \delta y) \\ &= xy + x \cdot \delta y + y \cdot \delta x + \delta x \cdot \delta y. \end{aligned}$$

Using $x_1 = 100.000003$ and $y_1 = 200.000006$ we have

$$\begin{aligned} x_1 y_1 &= 100 \times 200 + 100 \times 0.000006 + 200 \times 0.000003 + 0.000003 \times 0.000006 \\ &= 20,000 + 0.0006 + 0.0006 + 0.00000000018 \\ &= 20,000 + 0.0012 + 0.00000000018 \\ &= 20,000.00120000018 \end{aligned}$$

where it is clear that the products $x \cdot \delta y$, $y \cdot \delta x$ and $\delta x \cdot \delta y$ contribute very little to the result. Furthermore, the smaller we make δx and δy, their contribution becomes even more insignificant. Just imagine if we reduce δx and δy to the level of quantum phenomenon, e.g. 10^{-34}, then their products play no part in every-day numbers. But there is no need to stop there, we can make δx and δy as small as we like, e.g. $10^{-100,000,000,000}$. Later on we employ the device of reducing a number towards zero, such that any products involving them can be dropped from any calculation.

Even though the product of two numbers less than zero is an even smaller number, care must be taken with their quotients. For example, in the above scenario, where $\delta y = 0.000006$ and $\delta x = 0.000003$,

$$\frac{\delta y}{\delta x} = \frac{0.000006}{0.000003} = 2$$

so we must watch out for such quotients.

Differential calculus is concerned with the rate at which a function changes relative to one of its independent variables, and employs the ratio $\delta y/\delta x$ to compute this value. The limiting value of this ratio is called the function's *derivative*, and we will explore two ways of computing it, and provide a graphical interpretation of the process. The first method uses simple algebraic equations, and the second uses a functional representation. Needless to say, they both give the same result.

15.4 Equations and Limits

15.4.1 Quadratic Function

Here is a simple algebraic approach using limits to compute the derivative of a quadratic function. Starting with the function $y = x^2$, let x change by δx, and let δy be the corresponding change in y. We then have

$$\begin{aligned} y &= x^2 \\ y + \delta y &= (x + \delta x)^2 \\ &= x^2 + 2x \cdot \delta x + (\delta x)^2 \\ \delta y &= 2x \cdot \delta x + (\delta x)^2. \end{aligned}$$

Dividing throughout by δx we have

$$\frac{\delta y}{\delta x} = 2x + \delta x.$$

The ratio $\delta y/\delta x$ provides a measure of how fast y changes relative to x, in increments of δx. For example, when $x = 10$

$$\frac{\delta y}{\delta x} = 20 + \delta x,$$

and if $\delta x = 1$, then $\delta y/\delta x = 21$. Equally, if $\delta x = 0.001$, then $\delta y/\delta x = 20.001$. By making δx smaller and smaller, δy becomes equally smaller, and their ratio converges towards a limiting value of 20.

In this case, as δx approaches zero, $\delta y/\delta x$ approaches $2x$, which is written

$$\lim_{\delta x \to 0} \frac{\delta y}{\delta x} = 2x.$$

Thus in the limit, when $\delta x = 0$, we create a condition where δy is divided by zero—which is a meaningless operation. However, if we hold onto the idea of a limit, as $\delta x \to 0$, it is obvious that the quotient $\frac{\delta y}{\delta x}$ is converging towards $2x$. The subterfuge employed to avoid dividing by zero is to substitute $\frac{dy}{dx}$ to stand for the limiting condition:

$$\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x} = 2x.$$

$\frac{dy}{dx}$ (pronounced *dee y dee x*) is the derivative of $y = x^2$, i.e. $2x$. For instance, when $x = 0$, $\frac{dy}{dx} = 0$, and when $x = 3$, $\frac{dy}{dx} = 6$. The derivative $\frac{dy}{dx}$, is the instantaneous rate at which y changes relative to x.

If we had represented this equation as a function:

$$f(x) = x^2$$
$$f'(x) = 2x$$

where $f'(x)$ is another way of writing $\frac{dy}{dx}$.

Now let's introduce two constants into the original quadratic equation to see what effect, if any, they have on the derivative. We begin with

$$y = ax^2 + b$$

and increment x and y:

$$\begin{aligned} y + \delta y &= a(x + \delta x)^2 + b \\ &= a\left(x^2 + 2x \cdot \delta x + (\delta x)^2\right) + b \\ \delta y &= a\left(2x \cdot \delta x + (\delta x)^2\right). \end{aligned}$$

Dividing throughout by δx:

$$\frac{\delta y}{\delta x} = a(2x + \delta x)$$

and the derivative is

$$\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x} = 2ax.$$

Thus the added constant b disappears (i.e. because it does not change), whilst the multiplied constant a is transmitted through to the derivative.

15.4.2 Cubic Equation

Now let's repeat the above analysis for $y = x^3$:

$$\begin{aligned} y &= x^3 \\ y + \delta y &= (x + \delta x)^3 \\ &= x^3 + 3x^2 \cdot \delta x + 3x(\delta x)^2 + (\delta x)^3 \\ \delta y &= 3x^2 \cdot \delta x + 3x(\delta x)^2 + (\delta x)^3. \end{aligned}$$

Dividing throughout by δx:

$$\frac{\delta y}{\delta x} = 3x^2 + 3x \cdot \delta x + (\delta x)^2.$$

Using limits, we have

$$\lim_{\delta x \to 0} \frac{\delta y}{\delta x} = 3x^2$$

or

$$\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x} = 3x^2.$$

We could also show that if $y = ax^3 + b$ then

$$\frac{dy}{dx} = 3ax^2.$$

This incremental technique can be used to compute the derivative of all sorts of functions.

If we continue computing the derivatives of higher-order polynomials, we discover the following pattern:

$$\begin{aligned} y &= x^2, \quad & \frac{dy}{dx} &= 2x \\ y &= x^3, \quad & \frac{dy}{dx} &= 3x^2 \\ y &= x^4, \quad & \frac{dy}{dx} &= 4x^3 \\ y &= x^5, \quad & \frac{dy}{dx} &= 5x^4. \end{aligned}$$

Clearly, the rule is

$$y = x^n, \quad \frac{dy}{dx} = nx^{n-1}$$

but we need to prove why this is so. The solution is found in the binomial expansion for $(x + \delta x)^n$, which can be divided into three components:

1. Decreasing terms of x.
2. Increasing terms of δx.
3. The terms of Pascal's triangle.

For example, the individual terms of $(x + \delta x)^4$ are:

$$\begin{array}{rccccc} \text{Decreasing terms of } x: & x^4 & x^3 & x^2 & x^1 & x^0 \\ \text{Increasing terms of } \delta x: & (\delta x)^0 & (\delta x)^1 & (\delta x)^2 & (\delta x)^3 & (\delta x)^4 \\ \text{The terms of Pascal's triangle:} & 1 & 4 & 6 & 4 & 1 \end{array}$$

which combined, produce

$$x^4 + 4x^3(\delta x) + 6x^2(\delta x)^2 + 4x(\delta x)^3 + (\delta x)^4.$$

Thus when we begin an incremental analysis:

$$\begin{aligned} y &= x^4 \\ y + \delta y &= (x + \delta x)^4 \\ &= x^4 + 4x^3(\delta x) + 6x^2(\delta x)^2 + 4x(\delta x)^3 + (\delta x)^4 \\ \delta y &= 4x^3(\delta x) + 6x^2(\delta x)^2 + 4x(\delta x)^3 + (\delta x)^4. \end{aligned}$$

Dividing throughout by δx:

$$\frac{\delta y}{\delta x} = 4x^3 + 6x^2(\delta x)^1 + 4x(\delta x)^2 + (\delta x)^3.$$

In the limit, as δx slides to zero, only the second term of the original binomial expansion remains:

$$4x^3.$$

The second term of the binomial expansion $(1 + \delta x)^n$ is always of the form

$$nx^{n-1}$$

which is the proof we require.

15.4.3 Functions and Limits

In order to generalise the above findings, let's approach the above analysis using a function of the form $y = f(x)$. We begin by noting some arbitrary value of its independent variable and note the function's value. In general terms, this is x and

$f(x)$ respectively. We then increase x by a small amount δx, to give $x + \delta x$, and measure the function's value again: $f(x + \delta x)$. The function's change in value is $f(x + \delta x) - f(x)$, whilst the change in the independent variable is δx. The quotient of these two quantities approximates to the function's rate of change at x:

$$\frac{f(x + \delta x) - f(x)}{\delta x}. \tag{15.1}$$

By making δx smaller and smaller towards zero, (15.1) converges towards a limiting value expressed as

$$\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{f(x + \delta x) - f(x)}{\delta x} \tag{15.2}$$

which can be used to compute all sorts of functions. For example, to compute the derivative of $\sin x$ we proceed as follows:

$$\begin{aligned} y &= \sin x \\ y + \delta y &= \sin(x + \delta x). \end{aligned}$$

Using the identity $\sin(A + B) = \sin A \cdot \cos B + \cos A \cdot \sin B$, we have

$$\begin{aligned} y + \delta y &= \sin x \cdot \cos(\delta x) + \cos x \cdot \sin(\delta x) \\ \delta y &= \sin x \cdot \cos(\delta x) + \cos x \cdot \sin(\delta x) - \sin x \\ &= \sin x(\cos(\delta x) - 1) + \cos x \cdot \sin(\delta x). \end{aligned}$$

Dividing throughout by δx we have

$$\frac{\delta y}{\delta x} = \frac{\sin x}{\delta x}(\cos(\delta x) - 1) + \frac{\sin(\delta x)}{\delta x}\cos x.$$

In the limit as $\delta x \to 0$, $(\cos(\delta x) - 1) \to 0$ and $\sin(\delta x)/\delta x = 1$, (See Appendix A) and

$$\frac{dy}{dx} = \frac{d(\sin x)}{dx} = \cos x.$$

Before moving on, let's compute the derivative of $\cos x$.

$$\begin{aligned} y &= \cos x \\ y + \delta y &= \cos(x + \delta x). \end{aligned}$$

Using the identity $\cos(A + B) = \cos A \cdot \cos B - \sin A \cdot \sin B$, we have

$$\begin{aligned} y + \delta y &= \cos x \cdot \cos(\delta x) - \sin x \cdot \sin(\delta x) \\ \delta y &= \cos x \cdot \cos(\delta x) - \sin x \cdot \sin(\delta x) - \cos x \\ &= \cos x(\cos(\delta x) - 1) - \sin x \cdot \sin(\delta x). \end{aligned}$$

Dividing throughout by δx we have

$$\frac{\delta y}{\delta x} = \frac{\cos x}{\delta x}(\cos(\delta x) - 1) - \frac{\sin(\delta x)}{\delta x}\sin x.$$

In the limit as $\delta x \to 0$, $(\cos(\delta x) - 1) \to 0$ and $\sin(\delta x)/\delta x = 1$ (See Appendix A), and

$$\frac{dy}{dx} = \frac{d(\cos x)}{dx} = -\sin x.$$

We will continue to employ this strategy to compute the derivatives of other functions later on.

15.4.4 Graphical Interpretation of the Derivative

To illustrate this limiting process graphically, consider the scenario in Fig. 15.1 where the sample point is P. In this case the function is $f(x) = x^2$ and P's coordinates are $(x,\ x^2)$. We identify another point R, displaced δx to the right of P, with coordinates $(x + \delta x,\ x^2)$. The point Q on the curve, vertically above R, has coordinates $\left(x + \delta x,\ (x + \delta x)^2\right)$. When δx is relatively small, the slope of the line PQ approximates to the function's rate of change at P, which is the graph's slope. This is given by

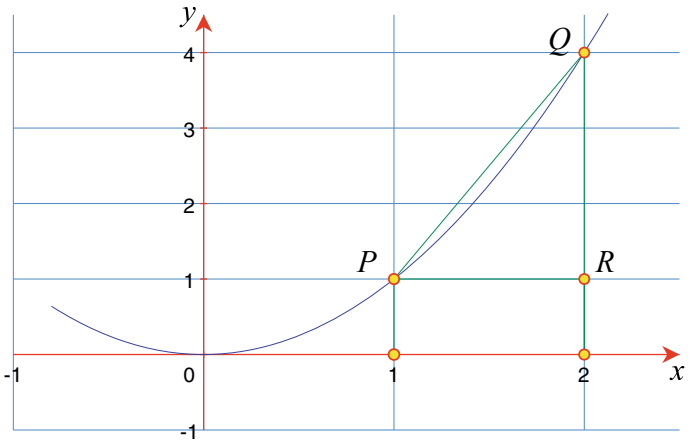

Fig. 15.1 Sketch of $f(x) = x^2$

$$\begin{aligned} slope &= \frac{QR}{PR} = \frac{(x+\delta x)^2 - x^2}{\delta x} \\ &= \frac{x^2 + 2x(\delta x) + (\delta x)^2 - x^2}{\delta x} \\ &= \frac{2x(\delta x) + (\delta x)^2}{\delta x} \\ &= 2x + \delta x. \end{aligned}$$

We can now reason that as δx is made smaller and smaller, Q approaches P, and *slope* becomes the graph's slope at P. This is the *limiting* condition:

$$\frac{dy}{dx} = \lim_{\delta x \to 0} (2x + \delta x) = 2x.$$

Thus, for any point with coordinates $(x,\ x^2)$, the slope is given by $2x$. For example, when $x = 0$, the slope is 0, and when $x = 4$, the slope is 8, etc.

15.4.5 Derivatives and Differentials

Given a function $f(x)$, $\frac{df}{dx}$ represents the instantaneous change of f for some x, and is called the *first derivative* of $f(x)$. For linear functions, this is constant, for other functions, the derivative's value changes with x and is represented by a function.

The elements df, dy and dx are called *differentials*, and historically, the derivative used to be called the *differential coefficient*, but has now been dropped in favour of *derivative*. One can see how the idea of a differential coefficient arose if we write, for example:

$$\frac{dy}{dx} = 3x$$

as

$$dy = 3x\ dx.$$

In this case, $3x$ acts like a coefficient of dx. However, today, the derivative is regarded as an operator, and even though it is written

$$\frac{dy}{dx}$$

it really means

$$\frac{d}{dx}[y]$$

where the operator $\frac{d}{dx}$ acts on y.

15.4.6 Integration and Antiderivatives

If it is possible to differentiate a function, it seems reasonable to assume the existence of an inverse operator which turns the derivative back to the original function. Fortunately, this is the case, but there are some limitations. This inverse process is called *integration* and reveals the *antiderivative* of a function. Many functions can be paired together in the form of a derivative and an antiderivative, such as $2x$ with x^2, and $\cos x$ with $\sin x$. However, there are many functions where it is impossible to derive its antiderivative in a precise form. For example, there is no simple, finite functional antiderivative for $\sin x^2$ or $(\sin x)/x$. To understand integration, let's begin with a simple derivative.

If we are given

$$\frac{dy}{dx} = 18x^2 - 8x + 8$$

it is not too difficult to reason that the original function could have been

$$y = 6x^3 - 4x^2 + 8x.$$

However, it could have also been

$$y = 6x^3 - 4x^2 + 8x + 2$$

or

$$y = 6x^3 - 4x^2 + 8x + 20$$

or with any other constant. Consequently, the integration process has to include an arbitrary constant:

$$y = 6x^3 - 4x^2 + 8x + C.$$

The value of C is not always required, but it can be determined if we are given some extra information, such as $y = 10$ when $x = 0$, then $C = 10$.

Given a function

$$y = 6x^3 - 4x^2 + 8x + 10$$

its derivative is written

$$\frac{d}{dx}[y] = 18x^2 - 8x + 8.$$

The antiderivative of $18x^2 - 8x + 8$ reveals the original function, and is written

$$y = \int (18x^2 - 8x + 8)\, dx$$

although brackets are not always used:

$$y = \int 18x^2 - 8x + 8 \, dx.$$

This equation reads: "*y equals the integral of* $18x^2 - 8x + 8$ *dee x*." The dx reminds us that x is the independent variable. In this case we can write the answer:

$$\begin{aligned} y &= \int 18x^2 - 8x + 8 \, dx \\ &= 6x^3 - 4x^2 + 8x + C \end{aligned}$$

where C is some constant.

The antiderivatives for the sine and cosine functions are written:

$$\int \sin x \, dx = -\cos x + C$$

$$\int \cos x \, dx = \sin x + C$$

which you may think obvious, as we have just computed their derivatives. However, the reason for introducing integration alongside differentiation, is to make you familiar with the notation, and memorise the two distinct processes, as well as lay the foundations for the next chapter.

15.5 Function Types

Mathematical functions come in all sorts of shapes and sizes. Sometimes they are described explicitly where y equals some function of its independent variable(s), such as

$$y = x \sin x$$

or implicitly where y, and its independent variable(s) are part of an equation, such as

$$x^2 + y^2 = 10.$$

A function may reference other functions, such as

$$y = \sin(\cos^2 x)$$

or

$$y = x^{\sin x}.$$

There is no limit to the way functions can be combined, which makes it impossible to cover every eventuality. Nevertheless, we will explore some useful combinations that prepare us for any future surprises.

First, we examine how to differentiate different types of functions, that include sums, products and quotients, which are employed later on to differentiate specific functions such as trigonometric, logarithmic and hyperbolic. Where relevant, I include the appropriate antiderivative to complement its derivative.

15.6 Differentiating Groups of Functions

So far, we have only considered simple individual functions, which unfortunately, do not represent the equations found in mathematics, science, physics or even computer science. In general, the functions we have to differentiate include sums of functions, functions of functions, function products and function quotients. Let's explore these four scenarios.

15.6.1 Sums of Functions

A function normally computes a numerical value from its independent variable(s), and if it can be differentiated, its derivative generates another function with the same independent variable. Consequently, if a function contains two functions of x, such as u and v, where

$$y = u(x) + v(x)$$

which can be abbreviated to

$$y = u + v$$

then

$$\frac{dy}{dx} = \frac{du}{dx} + \frac{dv}{dx}$$

where we just sum their individual derivatives. For example: find $\frac{dy}{dx}$, given

$$\begin{aligned} u &= 2x^6 \\ v &= 3x^5 \\ y &= u + v \\ y &= 2x^6 + 3x^5. \end{aligned}$$

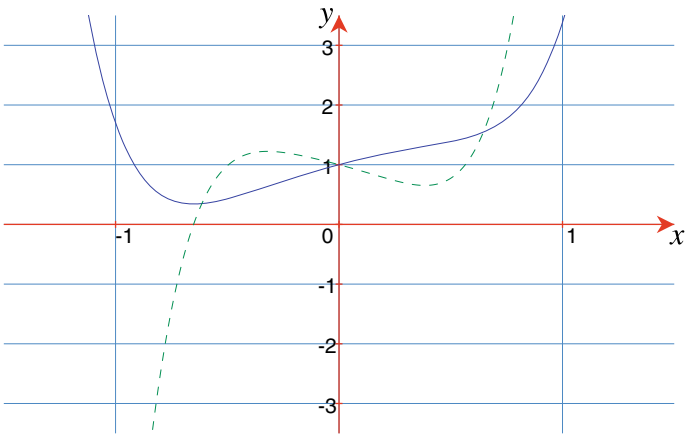

Fig. 15.2 Graph of $y = 2x^6 + \sin x + \cos x$ and its derivative, $\frac{dy}{dx} = 12x^5 + \cos x - \sin x$ (dashed)

Differentiating y:

$$\frac{dy}{dx} = 12x^5 + 15x^4.$$

Figure 15.2 shows a graph of $y = 2x^6 + \sin x + \cos x$ and its derivative, $\frac{dy}{dx} = 12x^5 + \cos x - \sin x$. Differentiating such functions is relatively easy, so too, is integrating. Given

$$\frac{dy}{dx} = \frac{du}{dx} + \frac{dv}{dx}$$

then

$$\begin{aligned} y &= \int u \, dx + \int v \, dx \\ &= \int (u + v) \, dx. \end{aligned}$$

For example, given

$$\frac{dy}{dx} = 12x^5 + \cos x - \sin x$$

we find y by integrating:

$$\begin{aligned} y &= \int 12x^5 \, dx + \int \cos x \, dx - \int \sin x \, dx \\ &= 2x^6 + \sin x + \cos x + C. \end{aligned}$$

15.6.2 Function of a Function

One of the advantages of modern mathematical notation is that it lends itself to unlimited elaboration without introducing any new symbols. For example, the polynomial $3x^2 + 2x$ is easily raised to some power by adding brackets and an appropriate index: $(3x^2 + 2x)^2$. Such an object is a *function of a function*, because the function $3x^2 + 2x$ is subjected to a further squaring function. The question now is: how are such functions differentiated? Well, the answer is relatively easy, but does introduce some new ideas.

Imagine that person A swims twice as fast as person B, who in turn, swims three times as fast as person C. It should be obvious that person A swims six (2×3) times faster than person C. This product rule, also applies to derivatives, because if y changes twice as fast as u, i.e. $\frac{dy}{du} = 2$, and u changes three times as fast as x, i.e. $\frac{du}{dx} = 3$, then y changes six times as fast as x:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

To differentiate

$$y = (3x^2 + 2x)^2$$

we substitute

$$u = 3x^2 + 2x$$

then

$$y = u^2$$

and

$$\begin{aligned} \frac{dy}{du} &= 2u \\ &= 2(3x^2 + 2x) \\ &= 6x^2 + 4x. \end{aligned}$$

Next, we require $\frac{du}{dx}$:

$$\begin{aligned} u &= 3x^2 + 2x \\ \frac{du}{dx} &= 6x + 2 \end{aligned}$$

therefore, we can write

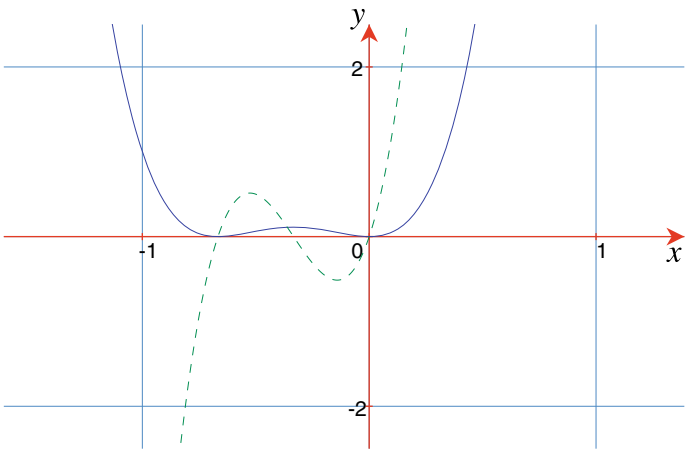

Fig. 15.3 Graph of $y = (3x^2 + 2x)^2$ and its derivative, $\frac{dy}{dx} = 36x^3 + 36x^2 + 8x$ (dashed)

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= (6x^2 + 4x)(6x + 2) \\ &= 36x^3 + 36x^2 + 8x.\end{aligned}$$

This result is easily verified by expanding the original polynomial and differentiating:

$$\begin{aligned}y &= (3x^2 + 2x)^2 \\ &= (3x^2 + 2x)(3x^2 + 2x) \\ &= 9x^4 + 12x^3 + 4x^2 \\ \frac{dy}{dx} &= 36x^3 + 36x^2 + 8x.\end{aligned}$$

Figure 15.3 shows a graph of $y = (3x^2 + 2x)^2$ and its derivative, $\frac{dy}{dx} = 36x^3 + 36x^2 + 8x$.

$y = \sin(ax)$ is a function of a function, and is differentiated as follows:

$$y = \sin(ax).$$

Substitute u for ax:

$$\begin{aligned}y &= \sin u \\ \frac{dy}{du} &= \cos u \\ &= \cos(ax).\end{aligned}$$

Next, we require $\frac{du}{dx}$:

$$\begin{aligned} u &= ax \\ \frac{du}{dx} &= a \end{aligned}$$

therefore, we can write

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= \cos(ax) \cdot a \\ &= a\cos(ax). \end{aligned}$$

Consequently, given

$$\frac{dy}{dx} = \cos(ax)$$

then

$$\begin{aligned} y &= \int \cos(ax)\, dx \\ &= \frac{1}{a}\sin(ax) + C. \end{aligned}$$

Similarly, given

$$\frac{dy}{dx} = \sin(ax)$$

then

$$\begin{aligned} y &= \int \sin(ax)\, dx \\ &= -\frac{1}{a}\cos(ax) + C. \end{aligned}$$

The equation $y = \sin(x^2)$ is also a function of a function, and is differentiated as follows:

$$y = \sin(x^2).$$

Substitute u for x^2:

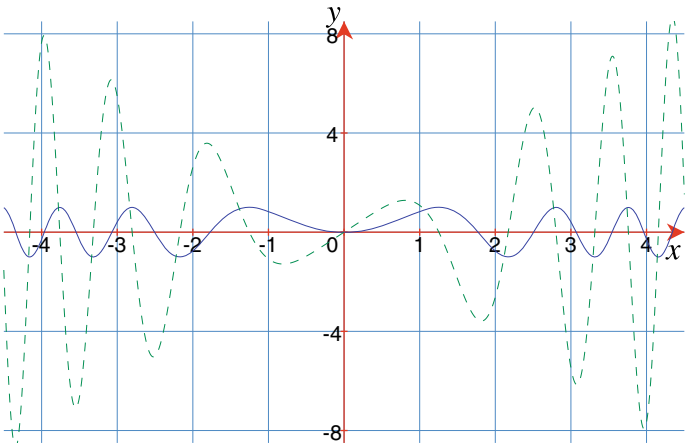

Fig. 15.4 Graph of $y = \sin x^2$ and its derivative, $\frac{dy}{dx} = 2x \cos x^2$ (dashed)

$$\begin{aligned} y &= \sin u \\ \frac{dy}{du} &= \cos u \\ &= \cos(x^2). \end{aligned}$$

Next, we require $\frac{du}{dx}$:

$$\begin{aligned} u &= x^2 \\ \frac{du}{dx} &= 2x \end{aligned}$$

therefore, we can write

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= \cos(x^2) \cdot 2x \\ &= 2x \cos(x^2). \end{aligned}$$

Figure 15.4 shows a graph of $y = \sin(x^2)$ and its derivative, $\frac{dy}{dx} = 2x \cos(x^2)$. In general, there can be any depth of functions within a function, which permits us to write the *chain rule* for derivatives:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dv} \cdot \frac{dv}{dw} \cdot \frac{dw}{dx}$$

15.6.3 Function Products

Function products occur frequently in every-day mathematics, and involve the product of two, or more functions. Here are three simple examples:

$$y = (3x^2 + 2x)(2x^2 + 3x)$$
$$y = \sin x \cdot \cos x$$
$$y = x^2 \sin x.$$

When it comes to differentiating function products of the form

$$y = uv$$

it seems natural to assume that

$$\frac{dy}{dx} = \frac{du}{dx} \cdot \frac{dv}{dx} \tag{15.3}$$

which unfortunately, is incorrect. For example, in the case of

$$y = (3x^2 + 2x)(2x^2 + 3x)$$

differentiating using (15.3) produces

$$\begin{aligned} \frac{dy}{dx} &= (6x + 2)(4x + 3) \\ &= 24x^2 + 26x + 6. \end{aligned}$$

However, if we expand the original product and then differentiate, we obtain

$$\begin{aligned} y &= (3x^2 + 2x)(2x^2 + 3x) \\ &= 6x^4 + 13x^3 + 6x^2 \\ \frac{dy}{dx} &= 24x^3 + 39x^2 + 12x \end{aligned}$$

which is correct, but differs from the first result. Obviously, (15.3) must be wrong. So let's return to first principles and discover the correct rule.

So far, we have incremented the independent variable—normally x—by δx to discover the change in y-normally δy. Next, we see how the same notation can be used to increment functions.

Given the following functions of x, u and v, where

$$y = uv$$

if x increases by δx, then there will be corresponding changes of δu, δv and δy, in u, v and y respectively. Therefore,

$$\begin{aligned} y + \delta y &= (u + \delta u)(v + \delta v) \\ &= uv + u\delta v + v\delta u + \delta u\delta v \\ \delta y &= u\delta v + v\delta u + \delta u\delta v. \end{aligned}$$

Dividing throughout by δx we have

$$\frac{\delta y}{\delta x} = u\frac{\delta v}{\delta x} + v\frac{\delta u}{\delta x} + \delta u\frac{\delta v}{\delta x}.$$

In the limiting condition:

$$\frac{dy}{dx} = \lim_{\delta x \to 0}\left(u\frac{\delta v}{\delta x}\right) + \lim_{\delta x \to 0}\left(v\frac{\delta u}{\delta x}\right) + \lim_{\delta x \to 0}\left(\delta u\frac{\delta v}{\delta x}\right).$$

As $\delta x \to 0$, then $\delta u \to 0$ and $\left(\delta u\frac{\delta v}{\delta x}\right) \to 0$. Therefore,

$$\frac{dy}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}. \tag{15.4}$$

Using (15.4) for the original function product:

$$\begin{aligned} u &= 3x^2 + 2x \\ v &= 2x^2 + 3x \\ y &= uv \\ \frac{du}{dx} &= 6x + 2 \\ \frac{dv}{dx} &= 4x + 3 \\ \frac{dy}{dx} &= u\frac{dv}{dx} + v\frac{du}{dx} \\ &= (3x^2 + 2x)(4x + 3) + (2x^2 + 3x)(6x + 2) \\ &= (12x^3 + 17x^2 + 6x) + (12x^3 + 22x^2 + 6x) \\ &= 24x^3 + 39x^2 + 12x \end{aligned}$$

which agrees with our previous prediction. Figure 15.5 shows the graph of $y = (3x^2 + 2x)(2x^2 + 3x)$ and its derivative, $\frac{dy}{dx} = 24x^3 + 39x^2 + 12x$.

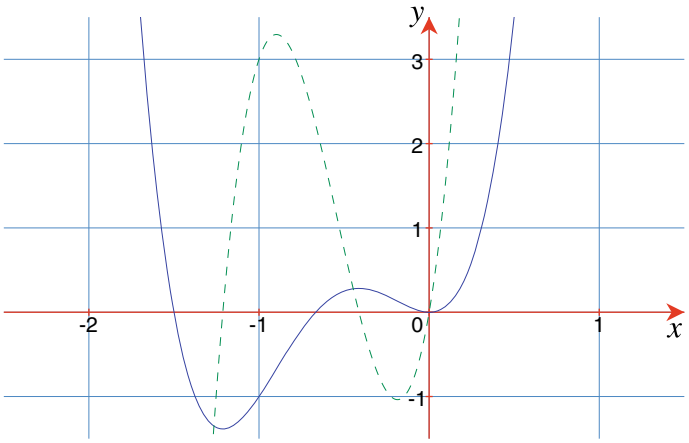

Fig. 15.5 Graph of $y = (3x^2 + 2x)(2x^2 + 3x)$ and its derivative, $\frac{dy}{dx} = 24x^3 + 39x^2 + 12x$ (dashed)

The equation $y = \sin x \cdot \cos x$ contains the product of two functions and is differentiated using (15.4) as follows:

$$\begin{aligned}
y &= \sin x \cdot \cos x \\
u &= \sin x \\
\frac{du}{dx} &= \cos x \\
v &= \cos x \\
\frac{dv}{dx} &= -\sin x \\
\frac{dy}{dx} &= u\frac{dv}{dx} + v\frac{du}{dx} \\
&= \sin x(-\sin x) + \cos x \cdot \cos x \\
&= \cos^2 x - \sin^2 x \\
&= \cos(2x).
\end{aligned}$$

Using the identity $\sin(2x) = 2\sin x \cdot \cos x$, we can rewrite the original function as

$$\begin{aligned}
y &= \sin x \cdot \cos x \\
\frac{dy}{dx} &= \tfrac{1}{2}\sin(2x) \\
&= \cos(2x)
\end{aligned}$$

which confirms the above derivative. Now let's consider the antiderivative of $\cos(2x)$.

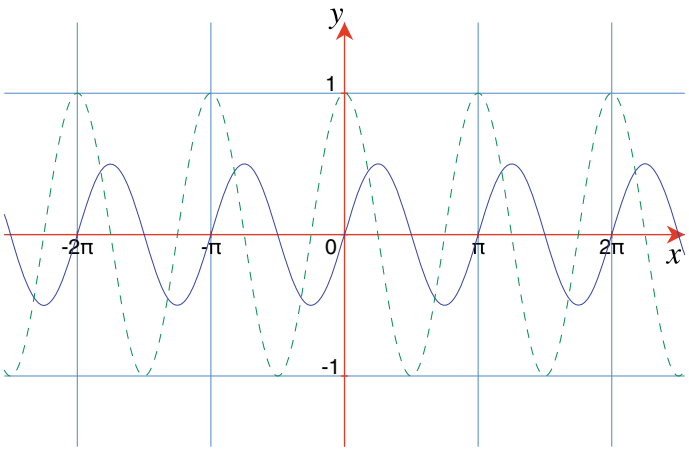

Fig. 15.6 Graph of $y = \sin x \cos x$ and its derivative, $\frac{dy}{dx} = \cos(2x)$ (dashed)

Given

$$\frac{dy}{dx} = \cos(2x)$$

then

$$\begin{aligned} y &= \int \cos(2x)\, dx \\ &= \tfrac{1}{2} \sin(2x) + C \\ &= \sin x \cdot \cos x + C. \end{aligned}$$

Figure 15.6 shows the graph of $y = \sin x \cos$ and its derivative, $\frac{dy}{dx} = \cos(2x)$.

15.6.4 Function Quotients

Next, we investigate how to differentiate the quotient of two functions. We begin with two functions of x, u and v, where

$$y = \frac{u}{v}$$

which makes y also a function of x.

We now increment x by δx and measure the change in u as δu, and the change in v as δv. Consequently, the change in y is δy:

$$\begin{aligned} y + \delta y &= \frac{u + \delta u}{v + \delta v} \\ \delta y &= \frac{u + \delta u}{v + \delta v} - \frac{u}{v} \\ &= \frac{v(u + \delta u) - u(v + \delta v)}{v(v + \delta v)} \\ &= \frac{v\delta u - u\delta v}{v(v + \delta v)}. \end{aligned}$$

Dividing throughout by δx we have

$$\frac{\delta y}{\delta x} = \frac{v\dfrac{\delta u}{\delta x} - u\dfrac{\delta v}{\delta x}}{v(v + \delta v)}.$$

As $\delta x \to 0$, δu, δv and δy also tend towards zero, and the limiting conditions are

$$\begin{aligned} \frac{dy}{dx} &= \lim_{\delta x \to 0} \frac{\delta y}{\delta x} \\ v\frac{du}{dx} &= \lim_{\delta x \to 0} v\frac{\delta u}{\delta x} \\ u\frac{dv}{dx} &= \lim_{\delta x \to 0} u\frac{\delta v}{\delta x} \\ v^2 &= \lim_{\delta x \to 0} v(v + \delta v) \end{aligned}$$

therefore,

$$\frac{dy}{dx} = \frac{v\dfrac{du}{dx} - u\dfrac{dv}{dx}}{v^2}.$$

As an example, let's differentiate

$$y = \frac{x^3 + 2x^2 + 3x + 6}{x^2 + 3}.$$

Substitute $u = x^3 + 2x^2 + 3x + 6$ and $v = x^2 + 3$, then

$$\begin{aligned} \frac{du}{dx} &= 3x^2 + 4x + 3 \\ \frac{dv}{dx} &= 2x \end{aligned}$$

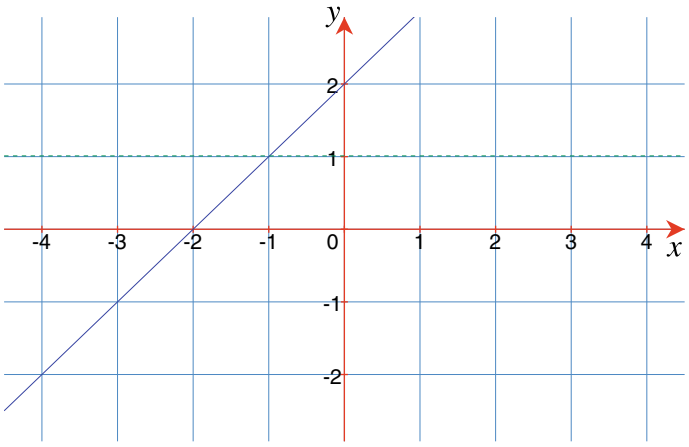

Fig. 15.7 Graph of $y = (x^2 + 3)(x + 2)/(x^2 + 3)$ and its derivative, $\frac{dy}{dx} = 1$ (dashed)

$$\begin{aligned}\frac{dy}{dx} &= \frac{(x^2+3)(3x^2+4x+3) - (x^3+2x^2+3x+6)(2x)}{(x^2+3)^2} \\ &= \frac{(3x^4+4x^3+3x^2+9x^2+12x+9) - (2x^4+4x^3+6x^2+12x)}{x^4+6x^2+9} \\ &= \frac{x^4+6x^2+9}{x^4+6x^2+9} \\ &= 1\end{aligned}$$

which is not a surprising result when one sees that the original function has the factors

$$y = \frac{(x^2+3)(x+2)}{x^2+3} = x + 2$$

whose derivative is 1. Figure 15.7 shows a graph of $y = (x^2 + 3)(x + 2)/(x^2 + 3)$ and its derivative, $\frac{dy}{dx} = 1$.

15.7 Differentiating Implicit Functions

Simple functions conveniently fall into two types: explicit and implicit. An explicit function, describes a function in terms of its independent variable(s), such as

$$y = a \sin x + b \cos x$$

where the value of y is determined by the values of a, b and x. On the other hand, an implicit function, such as

$$x^2 + y^2 = 25$$

combines the function's name with its definition. In this case, it is easy to untangle the explicit form:

$$y = \sqrt{25 - x^2}.$$

So far, we have only considered differentiating explicit functions, so now let's examine how to differentiate implicit functions. Let's begin with a simple explicit function and differentiate it as it is converted into its implicit form.

Let

$$y = 2x^2 + 3x + 4$$

then

$$\frac{dy}{dx} = 4x + 3.$$

Now let's start the conversion into the implicit form by bringing the constant 4 over to the left-hand side:

$$y - 4 = 2x^2 + 3x$$

differentiating both sides:

$$\frac{dy}{dx} = 4x + 3.$$

Bringing 4 and $3x$ across to the left-hand side:

$$y - 3x - 4 = 2x^2$$

differentiating both sides:

$$\frac{dy}{dx} - 3 = 4x$$

$$\frac{dy}{dx} = 4x + 3.$$

Finally, we have

$$y - 2x^2 - 3x - 4 = 0$$

differentiating both sides:

$$\frac{dy}{dx} - 4x - 3 = 0$$
$$\frac{dy}{dx} = 4x + 3$$

which seems straight forward. The reason for working through this example is to remind us that when y is differentiated we get $\frac{dy}{dx}$. Let's differentiate these two examples:

$$y + \sin x + 4x = 0 \quad \text{and} \quad y + x^2 - \cos x = 0.$$

Differentiating the individual terms:

$$y + \sin x + 4x = 0$$
$$\frac{dy}{dx} + \cos x + 4 = 0$$
$$\frac{dy}{dx} = -\cos x - 4.$$
$$y + x^2 - \cos x = 0$$
$$\frac{dy}{dx} + 2x + \sin x = 0$$
$$\frac{dy}{dx} = -2x - \sin x.$$

But how do we differentiate $y^2 + x^2 = r^2$? Well, the important difference between this implicit function and previous functions, is that it involves a function of a function. y is not only a function of x, but is squared, which means that we must employ the chain rule described earlier:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

Therefore, given

$$y^2 + x^2 = r^2$$
$$2y\frac{dy}{dx} + 2x = 0$$
$$\frac{dy}{dx} = \frac{-2x}{2y}$$
$$= \frac{-x}{\sqrt{r^2 - x^2}}.$$

This is readily confirmed by expressing the original function in its explicit form and differentiating:

$$y = (r^2 - x^2)^{\frac{1}{2}}$$

which is a function of a function.

Let $u = r^2 - x^2$, then

$$\frac{du}{dx} = -2x.$$

As $y = u^{\frac{1}{2}}$, then

$$\begin{aligned}\frac{dy}{du} &= \frac{1}{2}u^{-\frac{1}{2}} \\ &= \frac{1}{2u^{\frac{1}{2}}} \\ &= \frac{1}{2\sqrt{r^2 - x^2}}.\end{aligned}$$

However,

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= \frac{-2x}{2\sqrt{r^2 - x^2}} \\ &= \frac{-x}{\sqrt{r^2 - x^2}}\end{aligned}$$

which agrees with the implicit differentiated form.

15.8 Differentiating Exponential and Logarithmic Functions

15.8.1 Exponential Functions

Exponential functions have the form $y = a^x$, where the independent variable is the exponent. Such functions are used to describe various forms of growth or decay, from the compound interest law, to the rate at which a cup of tea cools down. One special value of a is 2.718 282 . . ., called e, where

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n.$$

Raising e to the power x:

$$e^x = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^{nx}$$

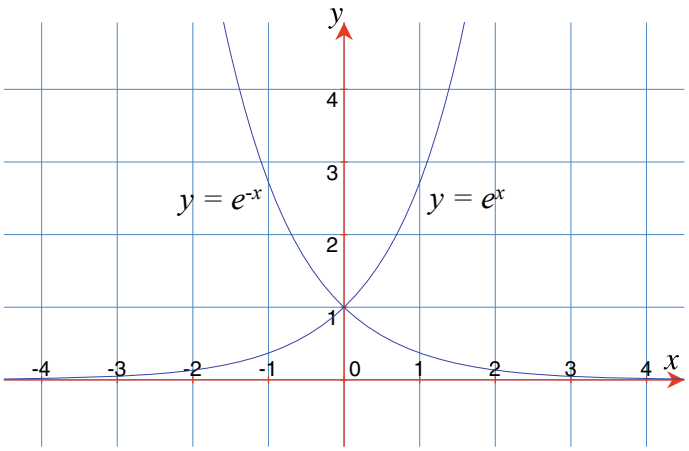

Fig. 15.8 Graphs of $y = e^x$ and $y = e^{-x}$

which, using the Binomial Theorem, is

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots .$$

If we let

$$\begin{aligned} y &= e^x \\ \frac{dy}{dx} &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \\ &= e^x . \end{aligned}$$

which is itself. Figure 15.8 shows graphs of $y = e^x$ and $y = e^{-x}$.

Now let's differentiate $y = a^x$. We know from the rules of logarithms that

$$\log x^n = n \log x$$

therefore, given

$$y = a^x$$

then

$$\ln y = \ln a^x = x \ln a$$

therefore

$$y = e^{x \ln a}$$

which means that

$$a^x = e^{x \ln a}.$$

Consequently,

$$\begin{aligned}\frac{d}{dx}[a^x] &= \frac{d}{dx}[e^{x \ln a}] \\ &= \ln a \; e^{x \ln a} \\ &= \ln a \; a^x.\end{aligned}$$

Similarly, it can be shown that

$$\begin{aligned} y &= e^{-x}, & \frac{dy}{dx} &= -e^{-x} \\ y &= e^{ax}, & \frac{dy}{dx} &= ae^{ax} \\ y &= e^{-ax}, & \frac{dy}{dx} &= -ae^{-ax} \\ y &= a^x, & \frac{dy}{dx} &= \ln a \; a^x \\ y &= a^{-x}, & \frac{dy}{dx} &= -\ln a \; a^{-x}. \end{aligned}$$

The exponential antiderivatives are written:

$$\begin{aligned} \int e^x \, dx &= e^x + C \\ \int e^{-x} \, dx &= -e^{-x} + C \\ \int e^{ax} \, dx &= \frac{1}{a} e^{ax} + C \\ \int e^{-ax} \, dx &= -\frac{1}{a} e^{ax} + C \\ \int a^x \, dx &= \frac{1}{\ln a} a^x + C \\ \int a^{-x} \, dx &= -\frac{1}{\ln a} a^{-x} + C. \end{aligned}$$

15.8.2 Logarithmic Functions

Given a function of the form

$$y = \ln x$$

then

$$x = e^y.$$

Therefore,

$$\begin{aligned} \frac{dx}{dy} &= e^y \\ &= x \\ \frac{dy}{dx} &= \frac{1}{x}. \end{aligned}$$

Thus

$$\frac{d}{dx}[\ln x] = \frac{1}{x}.$$

Figure 15.9 shows the graph of $y = \ln x$ and its derivative, $\frac{dy}{dx} = \frac{1}{x}$. Conversely,

$$\int \frac{1}{x}\, dx = \ln|x| + C.$$

When differentiating logarithms to a base a, we employ the conversion formula:

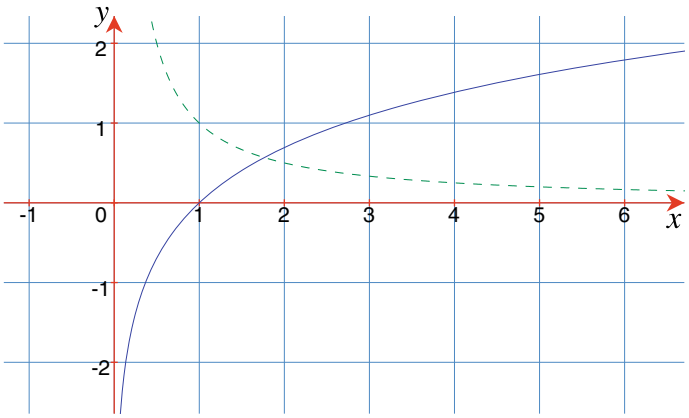

Fig. 15.9 Graph of $y = \ln x$ and its derivative, $\frac{dy}{dx} = \frac{1}{x}$ (dashed)

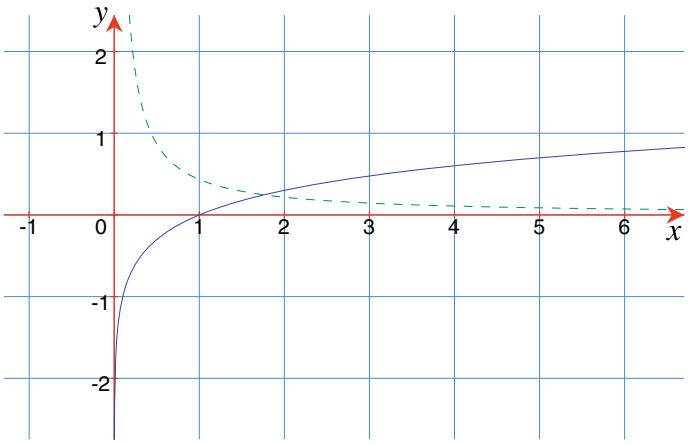

Fig. 15.10 Graph of $y = \log_{10} x$ and its derivative, $\frac{dy}{dx} = \frac{0.4343}{x}$ (dashed)

$$\begin{aligned} y &= \log_a x \\ &= (\ln x)(\log_a e) \end{aligned}$$

whose derivative is

$$\frac{dy}{dx} = \frac{1}{x} \log_a e.$$

When $a = 10$, then $\log_{10} e = 0.434\,3\ldots$ and

$$\frac{d}{dx}[\log_{10} x] = \frac{0.4343}{x}$$

Figure 15.10 shows the graph of $y = \log_{10} x$ and its derivative, $\frac{dy}{dx} = \frac{0.4343}{x}$.

15.9 Differentiating Trigonometric Functions

We have only differentiated two trigonometric functions: $\sin x$ and $\cos x$, so let's add $\tan x$, $\csc x$, $\sec x$ and $\cot x$ to the list, as well as their inverse forms.

15.9.1 Differentiating tan

Rather than return to first principles and start incrementing x by δx, we can employ the rules for differentiating different function combinations and various trigonometric identities. In the case of $\tan ax$, this can be written as

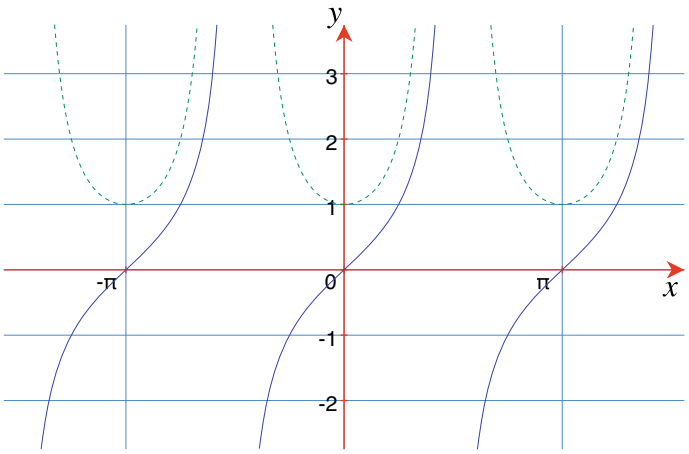

Fig. 15.11 Graph of $y = \tan x$ and its derivative, $\frac{dy}{dx} = 1 + \tan^2 x$ (dashed)

$$\tan ax = \frac{\sin(ax)}{\cos(ax)}$$

and employ the quotient rule:

$$\frac{dy}{dx} = \frac{v\dfrac{du}{dx} - u\dfrac{dv}{dx}}{v^2}.$$

Therefore, let $u = \sin(ax)$ and $v = \cos(ax)$, and

$$\begin{aligned}\frac{d}{dx}[\tan(ax)] &= \frac{a\cos(ax)\cdot\cos(ax) + a\sin(ax)\cdot\sin(ax)}{\cos^2(ax)} \\ &= \frac{a(\cos^2(ax) + \sin^2(ax))}{\cos^2(ax)} \\ &= \frac{a}{\cos^2(ax)} \\ &= a\sec^2(ax) \\ &= a(1 + \tan^2(ax)).\end{aligned}$$

Figure 15.11 shows a graph of $y = \tan x$ and its derivative, $\frac{dy}{dx} = 1 + \tan^2 x$.

It follows that

$$\int \sec^2(ax)\, dx = \frac{1}{a}\tan(ax) + C.$$

15.9.2 *Differentiating csc*

Using the quotient rule:

$$\begin{aligned} y &= \csc(ax) \\ &= \frac{1}{\sin(ax)} \\ \frac{d}{dx}[\csc(ax)] &= \frac{0 - a\cos(ax)}{\sin^2(ax)} \\ &= \frac{-a\cos(ax)}{\sin^2(ax)} \\ &= -\frac{a}{\sin(ax)} \cdot \frac{\cos(ax)}{\sin(ax)} \\ &= -a\csc(ax) \cdot \cot(ax). \end{aligned}$$

Figure 15.12 shows a graph of $y = \csc x$ and its derivative, $\frac{dy}{dx} = -\csc x \cot x$.

It follows that

$$\int \csc(ax) \cdot \cot(ax)\, dx = -\frac{1}{a}\csc(ax) + C.$$

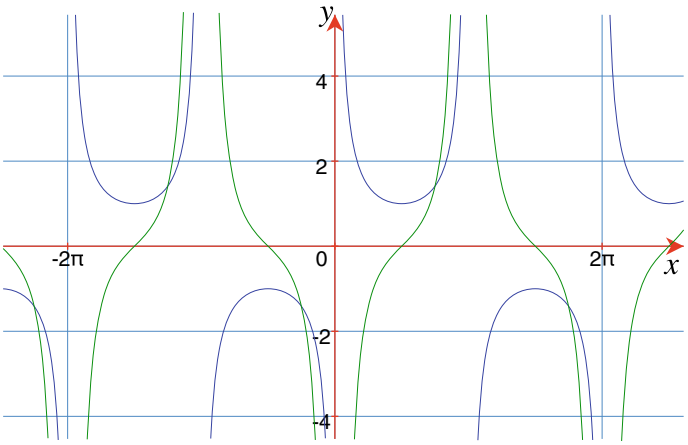

Fig. 15.12 Graph of $y = \csc x$ and its derivative, $\frac{dy}{dx} = -\csc x \cot x$ (dashed)

15.9.3 *Differentiating sec*

Using the quotient rule:

$$\begin{aligned} y &= \sec(ax) \\ &= \frac{1}{\cos(ax)} \\ \frac{d}{dx}[\sec(ax)] &= \frac{-(-a\sin(ax))}{\cos^2(ax)} \\ &= \frac{a\sin(ax)}{\cos^2(ax)} \\ &= \frac{a}{\cos(ax)} \cdot \frac{\sin(ax)}{\cos(ax)} \\ &= a\sec(ax) \cdot \tan(ax). \end{aligned}$$

Figure 15.13 shows a graph of $y = \sec x$ and its derivative, $\frac{dy}{dx} = \sec x \tan x$.

It follows that

$$\int \sec(ax) \cdot \tan(ax)\, dx = \frac{1}{a}\sec(ax) + C.$$

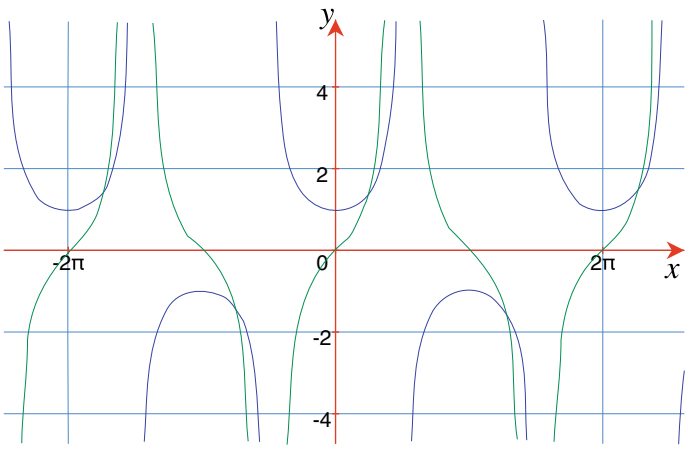

Fig. 15.13 Graph of $y = \sec x$ and its derivative, $\frac{dy}{dx} = \sec x \tan x$ (dashed)

15.9.4 *Differentiating cot*

Using the quotient rule:

$$\begin{aligned} y &= \cot(ax) \\ &= \frac{1}{\tan(ax)} \\ \frac{d}{dx}[\cot(ax)] &= \frac{-a\sec^2(ax)}{\tan^2(ax)} \\ &= -\frac{a}{\cos^2(ax)} \cdot \frac{\cos^2(ax)}{\sin^2(ax)} \\ &= -\frac{a}{\sin^2(ax)} \\ &= -a\csc^2(ax) \\ &= -a(1+\cot^2(ax)). \end{aligned}$$

Figure 15.14 shows a graph of $y = \cot x$ and its derivative, $\frac{dy}{dx} = -(1+\cot^2 x)$.

It follows that

$$\int 1+\cot^2(ax)\,dx = -\frac{1}{a}\cot(ax) + C.$$

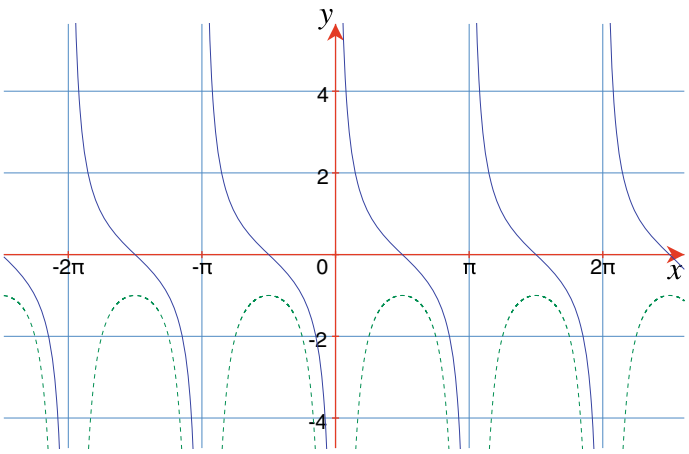

Fig. 15.14 Graph of $y = \cot x$ and its derivative, $\frac{dy}{dx} = -(1+\cot^2 x)$ (dashed)

15.9.5 Differentiating arcsin, arccos and arctan

These inverse functions are solved using a clever strategy.

Let

$$x = \sin y$$

then

$$y = \arcsin x.$$

Differentiating the first expression, we have

$$\frac{dx}{dy} = \cos y$$
$$\frac{dy}{dx} = \frac{1}{\cos y}$$

and as $\sin^2 y + \cos^2 y = 1$, then

$$\cos y = \sqrt{1 - \sin^2 y} = \sqrt{1 - x^2}$$

and

$$\frac{d}{dx}[\arcsin x] = \frac{1}{\sqrt{1 - x^2}}.$$

Using a similar technique, it can be shown that

$$\frac{d}{dx}[\arccos x] = -\frac{1}{\sqrt{1 - x^2}}$$
$$\frac{d}{dx}[\arctan x] = \frac{1}{1 + x^2}.$$

It follows that

$$\int \frac{dx}{\sqrt{1 - x^2}} = \arcsin x + C$$
$$\int \frac{-dx}{\sqrt{1 - x^2}} = \arccos x + C$$
$$\int \frac{dx}{1 + x^2} = \arctan x + C.$$

15.9.6 *Differentiating arccsc, arcsec and arccot*

Let

$$y = \operatorname{arccsc} x$$

then

$$\begin{aligned} x &= \csc y \\ &= \frac{1}{\sin y} \\ \frac{dx}{dy} &= \frac{-\cos y}{\sin^2 y} \\ \frac{dy}{dx} &= \frac{-\sin^2 y}{\cos y} \\ &= -\frac{1}{x^2}\frac{x}{\sqrt{x^2-1}} \\ \frac{d}{dx}[\operatorname{arccsc} x] &= -\frac{1}{x\sqrt{x^2-1}}. \end{aligned}$$

Similarly,

$$\begin{aligned} \frac{d}{dx}[\operatorname{arcsec} x] &= \frac{1}{x\sqrt{x^2-1}} \\ \frac{d}{dx}[\operatorname{arccot} x] &= -\frac{1}{x^2+1}. \end{aligned}$$

It follows:

$$\begin{aligned} \int \frac{dx}{x\sqrt{x^2-1}} &= \operatorname{arcsec}|x| + C \\ \int \frac{dx}{x^2+1} &= -\operatorname{arccot} x + C. \end{aligned}$$

15.10 Differentiating Hyperbolic Functions

Trigonometric functions are useful for parametric, circular motion, whereas, hyperbolic functions arise in equations for the absorption of light, mechanics and in integral calculus. Figure 15.15 shows graphs of the unit circle and a hyperbola whose respective equations are

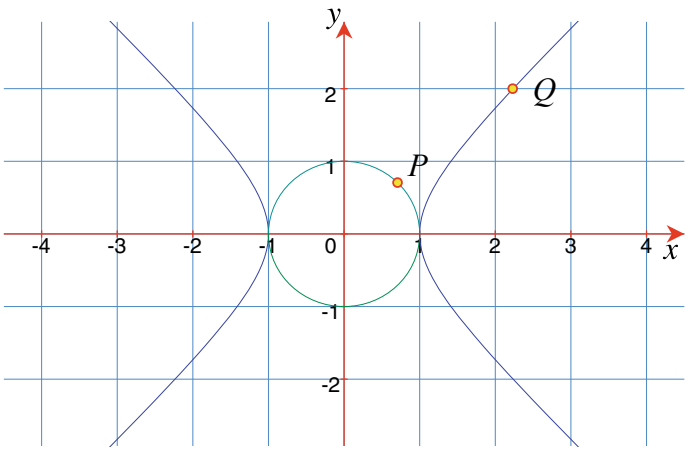

Fig. 15.15 Graphs of the unit circle $x^2 + y^2 = 1$ and the hyperbola $x^2 - y^2 = 1$

$$x^2 + y^2 = 1$$
$$x^2 - y^2 = 1$$

where the only difference between them is a sign. The parametric form for the trigonometric, or circular functions and the hyperbolic functions are respectively:

$$\sin^2\theta + \cos^2\theta = 1$$
$$\cosh^2 x - \sinh^2 x = 1.$$

The three hyperbolic functions have the following definitions:

$$\sinh x = \frac{e^x - e^{-x}}{2}$$
$$\cosh x = \frac{e^x + e^{-x}}{2}$$
$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

and their reciprocals are:

$$\operatorname{cosech} x = \frac{1}{\sinh x} = \frac{2}{e^x - e^{-x}}$$
$$\operatorname{sech} x = \frac{1}{\cosh x} = \frac{2}{e^x + e^{-x}}$$
$$\coth x = \frac{1}{\tanh x} = \frac{e^{2x} + 1}{e^{2x} - 1}.$$

Other useful identities include:

$$\operatorname{sech}^2 x = 1 - \tanh^2 x$$
$$\operatorname{cosech}^2 x = \coth^2 x - 1.$$

The coordinates of P and Q in Fig. 15.15 are given by $P(\cos\theta,\ \sin\theta)$ and $Q(\cosh x,\ \sinh x)$.

Table 15.1 shows the names of the three hyperbolic functions, their reciprocals and inverse forms. As these functions are based upon e^x and e^{-x}, they are relatively easy to differentiate.

15.10.1 Differentiating sinh, cosh and tanh

Table 15.2 gives the rules for differentiating hyperbolic functions, and Table 15.3 for inverse hyperbolic functions.

Table 15.4 gives the rules for integrating hyperbolic functions, and Table 15.5 for inverse hyperbolic functions.

Table 15.1 Hyperbolic function names

Function	Reciprocal	Inverse Function	Inverse Reciprocal
sinh	cosech	arsinh	arcsch
cosh	sech	arcosh	arsech
tanh	coth	artanh	arcoth

Table 15.2 Rules for differentiating hyperbolic functions

y	dy/dx
$\sinh x$	$\cosh x$
$\cosh x$	$\sinh x$
$\tanh x$	$\operatorname{sech}^2 x$
$\operatorname{cosech} x$	$-\operatorname{cosech} x \coth x$
$\operatorname{sech} x$	$-\operatorname{sech} x \tanh x$
$\coth x$	$-\operatorname{cosech}^2 x$

Table 15.3 Rules for differentiating inverse hyperbolic functions

y	dy/dx
$\operatorname{arsinh} x$	$\dfrac{1}{\sqrt{1+x^2}}$
$\operatorname{arcosh} x$	$\dfrac{1}{\sqrt{x^2-1}}$
$\operatorname{artanh} x$	$\dfrac{1}{1-x^2}$
$\operatorname{arcsch} x$	$-\dfrac{1}{x\sqrt{1+x^2}}$
$\operatorname{arsech} x$	$-\dfrac{1}{x\sqrt{1-x^2}}$
$\operatorname{arcoth} x$	$-\dfrac{1}{x^2-1}$

Table 15.4 Rules for integrating hyperbolic functions

$f(x)$	$\int f(x)\,dx$
$\sinh x$	$\cosh x + C$
$\cosh x$	$\sinh x + C$
$\operatorname{sech}^2 x$	$\tanh x + C$

Table 15.5 Rules for integrating inverse hyperbolic functions

$f(x)$	$\int f(x)\,dx$
$\dfrac{1}{\sqrt{1+x^2}}$	$\operatorname{arsinh} x + C$
$\dfrac{1}{\sqrt{x^2-1}}$	$\operatorname{arcosh} x + C$
$\dfrac{1}{1-x^2}$	$\operatorname{artanh} x + C.$

15.11 Higher Derivatives

There are three parts to this section: The first part shows what happens when a function is repeatedly differentiated; the second shows how these higher derivatives resolve local minimum and maximum conditions; and the third part provides a physical interpretation for these derivatives. Let's begin by finding the higher derivatives of simple polynomials.

15.12 Higher Derivatives of a Polynomial

We have previously seen that polynomials of the form

$$y = ax^r + bx^s + cx^t \ldots$$

are differentiated as follows:

$$\frac{dy}{dx} = rax^{r-1} + sbx^{s-1} + tcx^{t-1} \ldots .$$

For example, given

$$y = 3x^3 + 2x^2 - 5x$$

then

$$\frac{dy}{dx} = 9x^2 + 4x - 5$$

which describes how fast y changes relative to x.

Figure 15.16 shows the graph of $y = 3x^3 + 2x^2 - 5x$ and its derivative $\frac{dy}{dx} = 9x^2 + 4x - 5$, and we can see that when $x = -1$ there is a local maximum, where the function reaches a value of 4, then begins a downward journey to 0, where the slope is -5. Similarly, when $x \simeq 0.55$, there is a point where the function reaches a local minimum with a value of approximately -1.65. The slope is zero at both points, which is reflected in the graph of the derivative.

Having differentiated the function once, there is nothing to prevent us differentiating a second time, but first we require a way to annotate the process, which is

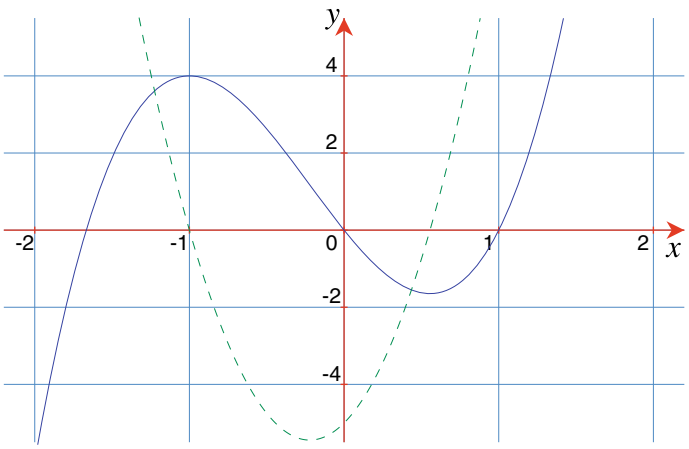

Fig. 15.16 Graph of $y = 3x^3 + 2x^2 - 5x$ and its derivative $\frac{dy}{dx} = 9x^2 + 4x - 5$ (dashed)

performed as follows. At a general level, let y be some function of x, then the first derivative is

$$\frac{d}{dx}[y].$$

The second derivative is found by differentiating the first derivative:

$$\frac{d}{dx}\left[\frac{d}{dx}[y]\right]$$

and is written:

$$\frac{d^2}{dx^2}[y] \quad \text{or} \quad \frac{d^2y}{dx^2}$$

Similarly, the third derivative is

$$\frac{d^3y}{dx^3}$$

and the nth derivative:

$$\frac{d^ny}{dx^n}.$$

When a function is expressed as $f(x)$, its derivative is written $f'(x)$. The second derivative is written $f''(x)$, and so on for higher derivatives.

Returning to the original function, the first and second derivatives are

$$\frac{dy}{dx} = 9x^2 + 4x - 5$$
$$\frac{d^2y}{dx^2} = 18x + 4$$

and the third and fourth derivatives are

$$\frac{d^3y}{dx^3} = 18$$
$$\frac{d^4y}{dx^4} = 0.$$

Figure 15.17 shows the original function and the first two derivatives. The graph of the first derivative shows the slope of the original function, whereas the graph of the second derivative shows the slope of the first derivative. These graphs help us identify a local maximum and minimum. By inspection of Fig. 15.17, when the first derivative equals zero, there is a local maximum or a local minimum. Algebraically, this is when

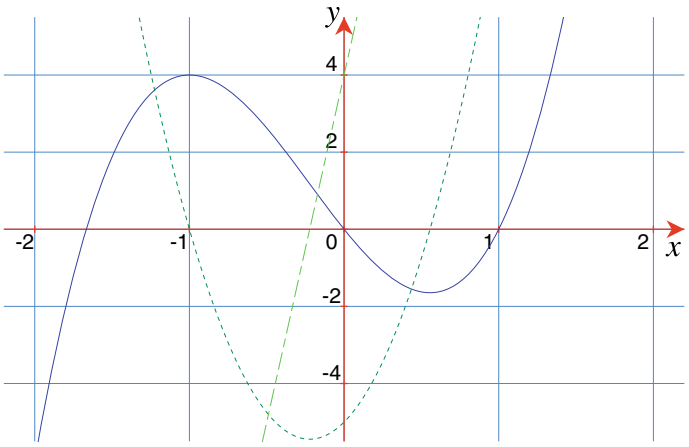

Fig. 15.17 Graph of $y = 3x^3 + 2x^2 - 5x$, its first derivative $\frac{dy}{dx} = 9x^2 + 4x - 5$ (short dashes) and its second derivative $\frac{d^2y}{dx^2} = 18x + 4$ (long dashes)

$$\frac{dy}{dx} = 0$$
$$9x^2 + 4x - 5 = 0.$$

Solving this quadratic in x we have

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

where $a = 9, \quad b = 4, \quad c = -5$:

$$x = \frac{-4 \pm \sqrt{16 + 180}}{18}$$
$$x_1 = -1, \quad x_2 = 0.555$$

which confirms our earlier analysis. However, what we don't know, without referring to the graphs, whether it is a minimum, or a maximum.

15.13 Identifying a Local Maximum or Minimum

Figure 15.18 shows a function containing a local maximum of 5 when $x = -1$. Note that as the independent variable x, increases from -2 towards 0, the slope of the graph changes from positive to negative, passing through zero at $x = -1$. This is shown in the function's first derivative, which is the straight line passing through the

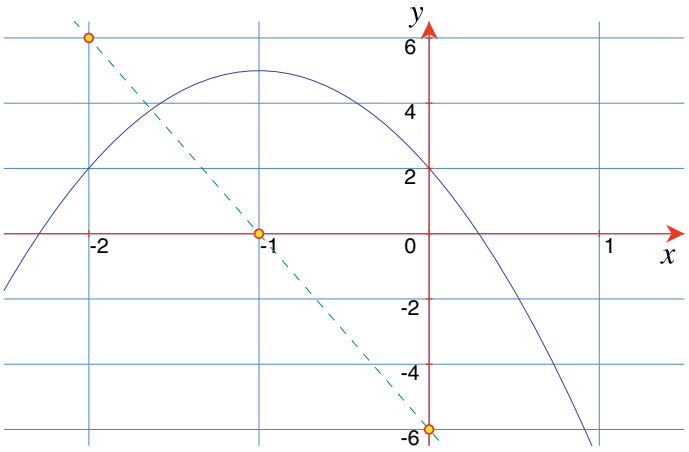

Fig. 15.18 A function containing a local maximum, and its first derivative (dashed)

points $(-2,\ 6)$, $(-1,\ 0)$ and $(0,\ -6)$. A natural consequence of these conditions implies that the slope of the first derivative must be negative:

$$\frac{d^2y}{dx^2} = -\text{ve}.$$

Figure 15.19 shows another function containing a local minimum of 5 when $x = -1$. Note that as the independent variable x, increases from -2 towards 0, the slope of the graph changes from negative to positive, passing through zero at $x = -1$. This is shown in the function's first derivative, which is the straight line passing through the points $(-2,\ -6)$, $(-1,\ 0)$ and $(0,\ 6)$. A natural consequence of these conditions implies that the slope of the first derivative must be positive:

$$\frac{d^2y}{dx^2} = +\text{ve}.$$

We can now apply this observation to the original function $y = 3x^3 + 2x^2 - 5x$ for the two values of x, $x_1 = -1, x_2 = 0.555$:

$$\begin{aligned} y &= 3x^3 + 2x^2 - 5x \\ \frac{dy}{dx} &= 9x^2 + 4x - 5 \\ \frac{d^2y}{dx^2} &= 18x + 4 \\ &= 18 \times (-1) = -18 \\ &= 18 \times (0.555) = +10. \end{aligned}$$

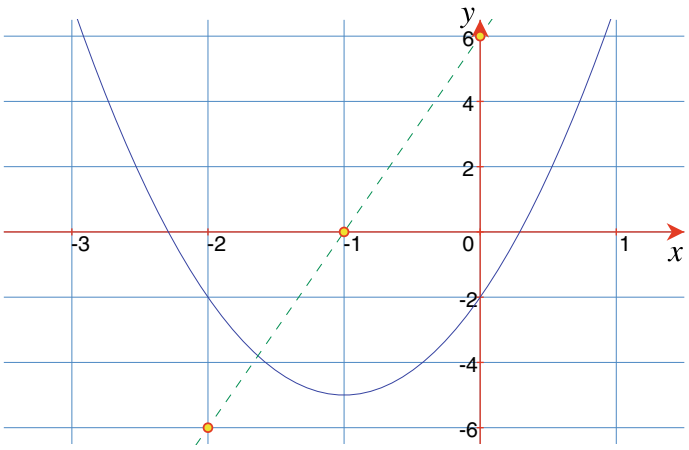

Fig. 15.19 A function containing a local minimum, and its first derivative (dashed)

Which confirms that when $x = -1$ there is a local maximum, and when $x = 0.555$, there is a local minimum, as shown in Fig. 15.16.

15.14 Partial Derivatives

Up to this point, we have used functions with one independent variable, such as $y = f(x)$. However, we must be able to compute derivatives of functions with more than one independent variable, such as $y = f(u, v, w)$. The technique employed is to assume that only one variable changes, whilst the other variables are held constant. This means that a function can possess several derivatives—one for each independent variable. Such derivatives are called *partial derivatives* and employ a new symbol ∂, which can be read as "*partial dee*".

Given a function $f(u, v, w)$, the three partial derivatives are defined as

$$\begin{aligned}
\frac{\partial f}{\partial u} &= \lim_{h \to 0} \frac{f(u+h, v, w) - f(u, v, w)}{h} \\
\frac{\partial f}{\partial v} &= \lim_{h \to 0} \frac{f(u, v+h, w) - f(u, v, w)}{h} \\
\frac{\partial f}{\partial w} &= \lim_{h \to 0} \frac{f(u, v, w+h) - f(u, v, w)}{h}.
\end{aligned}$$

For example, a function for the volume of a cylinder is

$$V(r, h) = \pi r^2 h$$

where r is the radius, and h is the height. Say we wish to compute the function's partial derivative with respect to r. First, the partial derivative is written

$$\frac{\partial V}{\partial r}.$$

Second, we hold h constant, whilst allowing r to change. This means that the function becomes

$$V(r, h) = kr^2 \tag{15.5}$$

where $k = \pi h$. Thus the partial derivative of (15.5) with respect to r is

$$\begin{aligned}\frac{\partial V}{\partial r} &= 2kr \\ &= 2\pi hr.\end{aligned}$$

Next, by holding r constant, and allowing h to change, we have

$$\frac{\partial V}{\partial h} = \pi r^2.$$

Sometimes, for purposes of clarification, the partial derivatives identify the constant variable(s):

$$\begin{aligned}\left(\frac{\partial V}{\partial r}\right)_h &= 2\pi hr \\ \left(\frac{\partial V}{\partial h}\right)_r &= \pi r^2.\end{aligned}$$

Partial differentiation is subject to the same rules for ordinary differentiation—we just to have to remember which independent variable changes, and those held constant. As with ordinary derivatives, we can compute higher-order partial derivatives. For example, let's find the second-order partial derivatives of $f(u, v)$, given

$$f(u, v) = u^4 + 2u^3v^2 - 4v^3.$$

The first partial derivatives are

$$\begin{aligned}\frac{\partial f}{\partial u} &= 4u^3 + 6u^2v^2 \\ \frac{\partial f}{\partial v} &= 4u^3v - 12v^2\end{aligned}$$

and the second-order partial derivatives are

$$\frac{\partial^2 f}{\partial u^2} = 12u^2 + 12uv^2$$
$$\frac{\partial^2 f}{\partial v^2} = 4u^3 - 24v.$$

In general, given $f(u, v) = uv$, then

$$\frac{\partial f}{\partial u} = v$$
$$\frac{\partial f}{\partial v} = u$$

and the second-order partial derivatives are

$$\frac{\partial^2 f}{\partial u^2} = 0$$
$$\frac{\partial^2 f}{\partial v^2} = 0.$$

Similarly, given $f(u, v) = u/v$, then

$$\frac{\partial f}{\partial u} = \frac{1}{v}$$
$$\frac{\partial f}{\partial v} = -\frac{u}{v^2}$$

and the second-order partial derivatives are

$$\frac{\partial^2 f}{\partial u^2} = 0$$
$$\frac{\partial^2 f}{\partial v^2} = \frac{2u}{v^3}.$$

Finally, given $f(u, v) = u^v$, then

$$\frac{\partial f}{\partial u} = vu^{v-1}$$

whereas, $\partial f/\partial v$ requires some explaining. First, given

$$f(u, v) = u^v$$

taking natural logs of both sides, we have

$$\ln f(u, v) = v \ln u$$

and

$$f(u, v) = e^{v \ln u}.$$

Therefore,

$$\begin{aligned}\frac{\partial f}{\partial v} &= e^{v \ln u} \ln u \\ &= u^v \ln u.\end{aligned}$$

The second-order partial derivatives are

$$\begin{aligned}\frac{\partial^2 f}{\partial u^2} &= v(v-1)u^{v-2} \\ \frac{\partial^2 f}{\partial v^2} &= u^v \ln^2 u.\end{aligned}$$

15.14.1 Visualising Partial Derivatives

Functions of the form $y = f(x)$ are represented by a 2D graph, and the function's derivative $f'(x)$ represents the graph's slope at any point x. Functions of the form $z = f(x, y)$ can be represented by a 3D surface, like the one shown in Fig. 15.20, which is $z(x, y) = 2.5x^2 - 2.5y^2$. The two partial derivatives are

$$\begin{aligned}\frac{\partial z}{\partial x} &= 5x \\ \frac{\partial z}{\partial y} &= -5y\end{aligned}$$

where $\frac{\partial z}{\partial x}$ is the slope of the surface in the x-direction, as shown in Fig. 15.21, and $\frac{\partial z}{\partial y}$ is the slope of the surface in the y-direction, as shown in Fig. 15.22.

The second-order partial derivatives are

$$\begin{aligned}\frac{\partial^2 z}{\partial x^2} &= 5 = +\text{ve} \\ \frac{\partial^2 z}{\partial y^2} &= -5 = -\text{ve}.\end{aligned}$$

As $\frac{\partial^2 z}{\partial x^2}$ is positive, there is a local minimum in the x-direction, and as $\frac{\partial^2 z}{\partial y^2}$ is negative, there is a local maximum in the y-direction, as confirmed by Figs. 15.21 and 15.22.

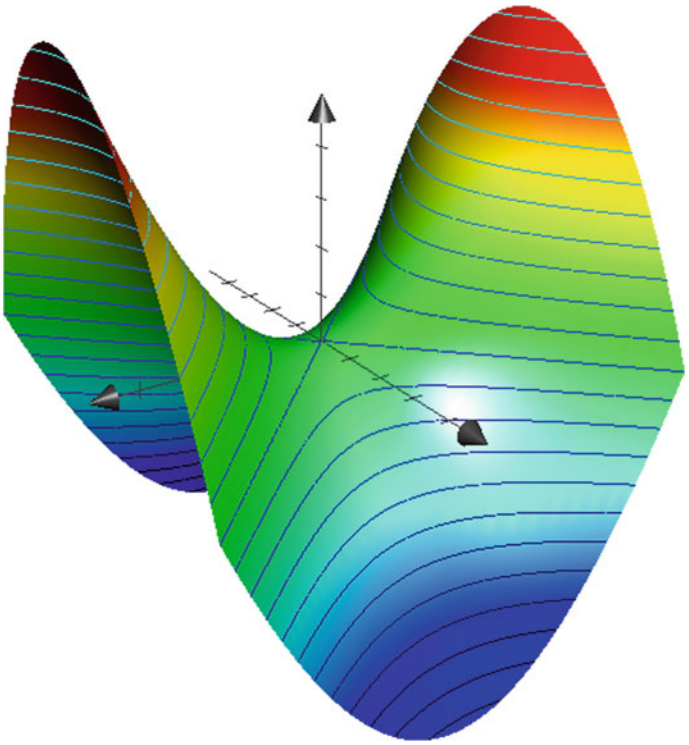

Fig. 15.20 Surface of $z = 2.5x^2 - 2.5y^2$ using a right-handed axial system with a vertical z-axis

15.14.2 Mixed Partial Derivatives

We have seen that, given a function of the form $f(u, v)$, the partial derivatives $\frac{\partial f}{\partial u}$ and $\frac{\partial f}{\partial v}$ provide the relative instantaneous changes in f and u, and f and v, respectively, whilst the second independent variable remains fixed. However, nothing prevents us from differentiating $\frac{\partial f}{\partial u}$ with respect to v, whilst keeping u constant:

$$\frac{\partial}{\partial v}\left(\frac{\partial f}{\partial u}\right)$$

which is also written as

$$\frac{\partial^2 f}{\partial v \partial u}$$

and is a *mixed partial derivative*. For example, to find the mixed partial derivative of f, given

$$f(u, v) = u^3 v^4$$

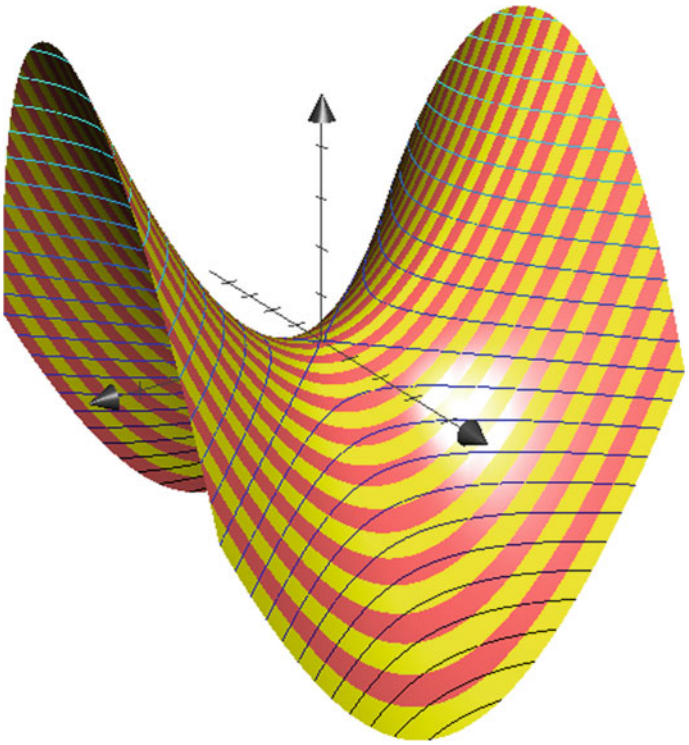

Fig. 15.21 $\frac{\partial z}{\partial x}$ describes the slopes of these contour lines

we have

$$\frac{\partial f}{\partial u} = 3u^2v^4$$

and

$$\frac{\partial^2 f}{\partial v \partial u} = 12u^2v^3.$$

It should be no surprise that reversing the differentiation gives the same result: Let

$$f(u, v) = u^3v^4$$

then

$$\frac{\partial f}{\partial v} = 4u^3v^3$$

and

$$\frac{\partial^2 f}{\partial u \partial v} = 12u^2v^3.$$

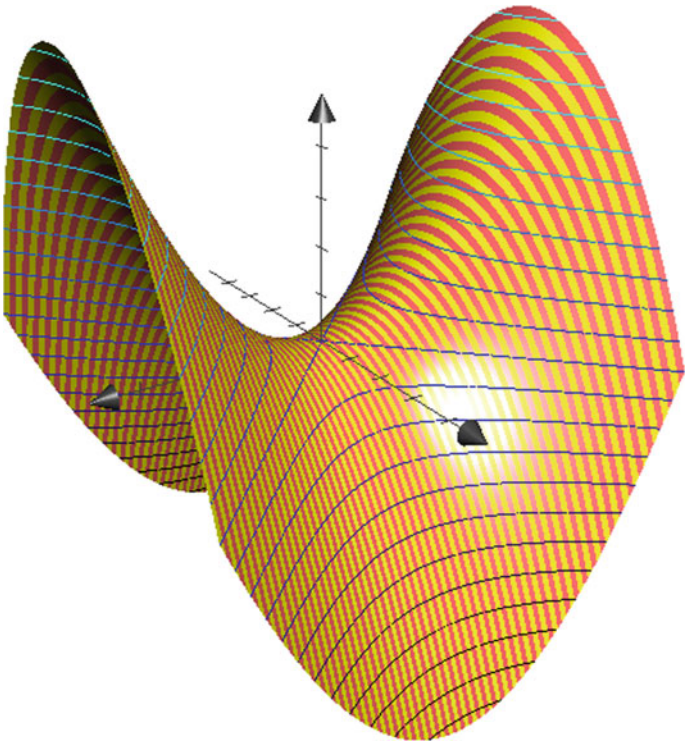

Fig. 15.22 $\frac{\partial z}{\partial y}$ describes the slopes of these contour lines

Generally, for continuous functions, we can write

$$\frac{\partial^2 f}{\partial u \partial v} = \frac{\partial^2 f}{\partial v \partial u}.$$

15.15 Chain Rule

Earlier, we came across the chain rule for computing the derivatives of functions of functions. For example, to compute the derivative of $y = \sin^2 x$ we substitute $u = x^2$, then

$$\begin{aligned} y &= u \\ \frac{dy}{du} &= \cos u \\ &= \cos(x^2). \end{aligned}$$

Next, we compute $\frac{du}{dx}$:

$$\begin{aligned} u &= x^2 \\ \frac{du}{dx} &= 2x \end{aligned}$$

and $\frac{dy}{dx}$ is the product of the two derivatives using the chain rule:

$$\begin{aligned} \frac{dy}{dx} &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= \cos(x^2)2x \\ &= 2x\cos(x^2). \end{aligned}$$

But say we have a function where w is a function of two variables x and y, which in turn, are a function of u and v. Then we have

$$\begin{aligned} w &= f(x, y) \\ x &= r(u, v) \\ y &= s(u, v). \end{aligned}$$

With such a scenario, we have the following partial derivatives:

$$\begin{array}{cc} \frac{\partial w}{\partial x}, & \frac{\partial w}{\partial y} \\ \frac{\partial w}{\partial u}, & \frac{\partial w}{\partial v} \\ \frac{\partial x}{\partial u}, & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u}, & \frac{\partial y}{\partial v}. \end{array}$$

These are chained together as follows

$$\frac{\partial w}{\partial u} = \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial u} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial u} \tag{15.6}$$

$$\frac{\partial w}{\partial v} = \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial v} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial v}. \tag{15.7}$$

For example, to find $\frac{\partial w}{\partial u}$ and $\frac{\partial w}{\partial v}$, given

$$w = f(2x + 3y)$$
$$x = r(u^2 + v^2)$$
$$y = s(u^2 - v^2)$$

we have

$$\frac{\partial w}{\partial x} = 2, \quad \frac{\partial w}{\partial y} = 3,$$
$$\frac{\partial x}{\partial u} = 2u, \quad \frac{\partial x}{\partial v} = 2v,$$
$$\frac{\partial y}{\partial u} = 2u, \quad \frac{\partial y}{\partial v} = -2v,$$

and plugging these into (15.6) and (15.7) we have

$$\begin{aligned}\frac{\partial w}{\partial u} &= \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial u} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial u} \\ &= 2 \times 2u + 3 \times 2u \\ &= 10u \\ \frac{\partial w}{\partial v} &= \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial v} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial v} \\ &= 2 \times 2v + 3 \times (-2v) \\ &= -2v.\end{aligned}$$

Thus, when $u = 2$ and $v = 1$

$$\frac{\partial w}{\partial u} = 20, \quad \text{and} \quad \frac{\partial w}{\partial v} = -2.$$

15.16 Total Derivative

Given a function with three independent variables, such as $w = f(x,\ y,\ t)$, where $x = g(t)$ and $y = h(t)$, there are three primary partial derivatives:

$$\frac{\partial w}{\partial x}, \quad \frac{\partial w}{\partial y}, \quad \frac{\partial w}{\partial t},$$

which show the differential change of w with x, y and t respectively. There are also three derivatives:

$$\frac{dx}{dt}, \quad \frac{dy}{dt}, \quad \frac{dt}{dt}$$

where $\frac{dt}{dt} = 1$. The partial and ordinary derivatives can be combined to create the *total derivative* which is written

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial t}.$$

$\frac{dw}{dt}$ measures the instantaneous change of w relative to t, when all three independent variables change. For example, to find $\frac{dw}{dt}$, given

$$\begin{aligned} w &= x^2 + xy + y^3 + t^2 \\ x &= 2t \\ y &= t - 1 \end{aligned}$$

we have,

$$\begin{aligned} \frac{dx}{dt} &= 2 \\ \frac{dy}{dt} &= 1 \\ \frac{\partial w}{\partial x} &= 2x + y = 4t + t - 1 = 5t - 1 \\ \frac{\partial w}{\partial y} &= x + 3y^2 = 2t + 3(t-1)^2 = 3t^2 - 4t + 3 \\ \frac{\partial w}{\partial t} &= 2t \\ \frac{dw}{dt} &= \frac{\partial w}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial w}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial w}{\partial t} \\ &= (5t-1)2 + (3t^2 - 4t + 3) + 2t = 3t^2 + 8t + 1 \end{aligned}$$

and the total derivative equals

$$\frac{dw}{dt} = 3t^2 + 8t + 1$$

and when $t = 1$, $dw/dt = 12$.

15.17 Power Series

A *power series* is an infinite string of nomial terms with increasing powers. A general form is written

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$

where a_n and x^n are generally real quantities. And because each term is individually simple, they are relatively easy to differentiate and integrate. For example, given

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots$$
$$\frac{d}{dx}[y] = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \cdots$$

An excellent example is found in the exponential function e^x:

$$e^x = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$
$$\frac{d}{dx}[e^x] = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

In particular:

$$e^1 = e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots$$

In 1715, the English mathematician Brook Taylor (1685–1731) published *Methods Incrementorum Directa et Inversa* which contained a theorem concerning power series. Today, this is known as *Taylor's theorem*, and the associated series: *Taylor's series*. Lagrange recognised its importance and called it "the main foundation of differential calculus".

Taylor proposed that any reasonable function, such as $\sin x$ and $\cos x$ can be written as a power series:

$$\sin x = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + a_6 x^6 + \cdots$$

To find the values of a_0, a_1, a_2, etc., we proceed as follows. When $x = 0$, $\sin 0 = 0$, which implies $a_0 = 0$. Therefore,

$$\sin x = a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + a_6 x^6 + \cdots \tag{15.8}$$

Differentiating (15.8) we get

$$\frac{d}{dx}[\sin x] = \cos x = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + 5a_5 x^4 + 6a_6 x^5 + \cdots$$

When $x = 0$, $\cos 0 = 1$, which implies $a_1 = 1$. Therefore,

$$\cos x = a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + 5a_5x^4 + 6a_6x^5 + \cdots \tag{15.9}$$

Differentiating (15.9) we get

$$\frac{d}{dx}[\cos x] = -\sin x = 2a_2 + 6a_3x + 12a_4x^2 + 20a_5x^3 + 30a_6x^4 + \cdots$$

When $x = 0$, $-\sin 0 = 0$, which implies $a_2 = 0$. Therefore,

$$-\sin x = 6a_3x + 12a_4x^2 + 20a_5x^3 + 30a_6x^4 + \cdots \tag{15.10}$$

Differentiating (15.10) we get

$$\frac{d}{dx}[-\sin x] = -\cos x = 6a_3 + 24a_4x + 60a_5x^2 + 120a_6x^3 + \cdots$$

When $x = 0$, $-\cos 0 = -1$, which implies $a_3 = -\frac{1}{6}$. Therefore,

$$-\cos x = 6a_3 + 24a_4x + 60a_5x^2 + 120a_6x^3 + \cdots \tag{15.11}$$

Differentiating (15.11) we get

$$\frac{d}{dx}[-\cos x] = \sin x = 24a_4 + 120a_5x + 360a_6x^2 + \cdots$$

When $x = 0$, $\sin 0 = 0$, which implies $a_4 = 0$. Therefore,

$$\sin x = 120a_5x + 360a_6x^3 + \cdots \tag{15.12}$$

Differentiating (15.12) we get

$$\frac{d}{dx}[\sin x] = \cos x = 120a_5 + 1080a_6x + \cdots$$

When $x = 0$, $\cos x = 1$, which implies $a_5 = \frac{1}{120}$.

We now have $a_0 = 0$, $a_1 = 1$, $a_2 = 0$, $a_3 = -\frac{1}{6}$, $a_4 = 0$, and $a_5 = \frac{1}{120}$, which means that the original $\sin x$ function comprises only the odd powers of x, with alternating signs. This permits us to write

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

Conversely, the $\cos x$ function comprises only the even powers of x, with alternating signs. This permits us to write

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

It is clear that $\sin x$ and $\cos x$ are closely related to e^x:

$$\begin{aligned} e^x &= 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \frac{x^6}{6!} + \frac{x^7}{7!} + \frac{x^8}{8!} + \frac{x^9}{9!} + \cdots \\ \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} + \cdots \\ \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} + \cdots \end{aligned}$$

and it was Euler who discovered that by making x imaginary: e^{ix}, we have

$$\begin{aligned} e^{ix} &= 1 + \frac{ix^1}{1!} - \frac{x^2}{2!} - \frac{ix^3}{3!} + \frac{x^4}{4!} + \frac{ix^5}{5!} - \frac{x^6}{6!} - \frac{ix^7}{7!} + \frac{x^8}{8!} + \frac{ix^9}{9!} \cdots \\ &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} + \cdots + \frac{ix^1}{1!} - \frac{ix^3}{3!} + \frac{ix^5}{5!} - \frac{ix^7}{7!} + \frac{ix^9}{9!} + \cdots \\ &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} + \cdots + i\left(\frac{x^1}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} + \cdots\right) \\ &= \cos x + i \sin x \end{aligned}$$

which is *Euler's trigonometric formula.*

15.18 Worked Examples

15.18.1 Antiderivative 1

Given $\frac{dy}{dx} = 1$, find y.
Solution: Integrating:

$$\begin{aligned} y &= \int 1 \, dx \\ &= x + C. \end{aligned}$$

15.18.2 Antiderivative 2

Given $\frac{dy}{dx} = 6x^2 + 10x$, find y.
Solution: Integrating:

$$\begin{aligned} y &= \int 6x^2 + 10x \; dx \\ &= 2x^3 + 5x^2 + C. \end{aligned}$$

15.18.3 Differentiating Sums of Functions

Differentiate $y = 2x^6 + \sin x + \cos x$.
Solution:

$$\frac{dy}{dx} = 12x^5 + \cos x - \sin x.$$

15.18.4 Differentiating a Function Product

Differentiate $y = x^2 \sin x$.
Solution:

$$\begin{aligned} y &= x^2 \sin x \\ u &= x^2 \\ \frac{du}{dx} &= 2x \\ v &= \sin x \\ \frac{dv}{dx} &= \cos x \\ \frac{dy}{dx} &= u\frac{dv}{dx} + v\frac{du}{dx} \\ &= x^2 \cos x + 2x \sin x. \end{aligned}$$

Figure 15.23 shows a graph of $y = x^2 \sin x$ and its derivative, $\frac{dy}{dx} = x^2 \cos x + 2x \sin x$.

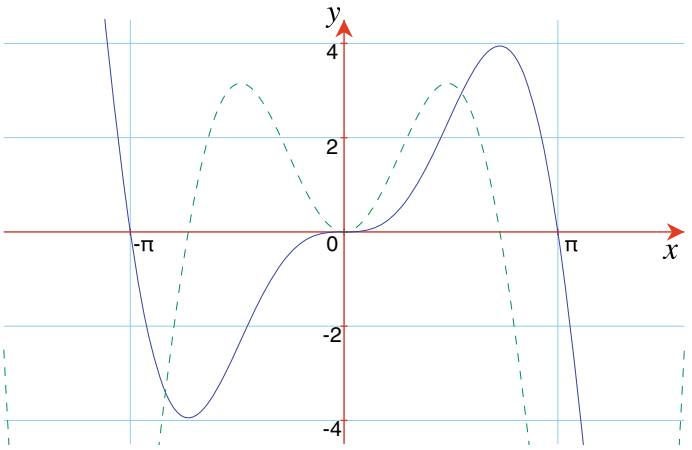

Fig. 15.23 Graph of $y = x^2 \sin x$ and its derivative, $\frac{dy}{dx} = x^2 \cos x + 2x \sin x$ (dashed)

15.18.5 Differentiating an Implicit Function

Differentiate $x^2 - y^2 + 4x = 6y$.
Solution:

$$2x - 2y\frac{dy}{dx} + 4 = 6\frac{dy}{dx}.$$

Rearranging the terms, we have

$$\begin{aligned} 2x + 4 &= 6\frac{dy}{dx} + 2y\frac{dy}{dx} \\ &= \frac{dy}{dx}(6 + 2y) \\ \frac{dy}{dx} &= \frac{2x + 4}{6 + 2y}. \end{aligned}$$

If we have to find the slope of $x^2 - y^2 + 4x = 6y$ at the point $(4,\ 3)$, then we simply substitute $x = 4$ and $y = 3$ in $\frac{dy}{dx}$ to obtain the answer 1.

15.18.6 Differentiating a General Implicit Function

Differentiate $x^n + y^n = a^n$.
Solution:

$$x^n + y^n = a^n$$

$$nx^{n-1} + ny^{n-1}\frac{dy}{dx} = 0$$
$$\frac{dy}{dx} = -\frac{nx^{n-1}}{ny^{n-1}}$$
$$= -\frac{x^{n-1}}{y^{n-1}}.$$

15.18.7 Local Maximum or Minimum

Given $y = -3x^3 + 9x$, find the local minimum and maximum for y.
Solution: The first derivative is

$$\frac{dy}{dx} = -9x^2 + 9$$

and second derivative

$$\frac{d^2y}{dx^2} = -18x$$

as shown in Fig. 15.24. For a local maximum or minimum, the first derivative equals zero:

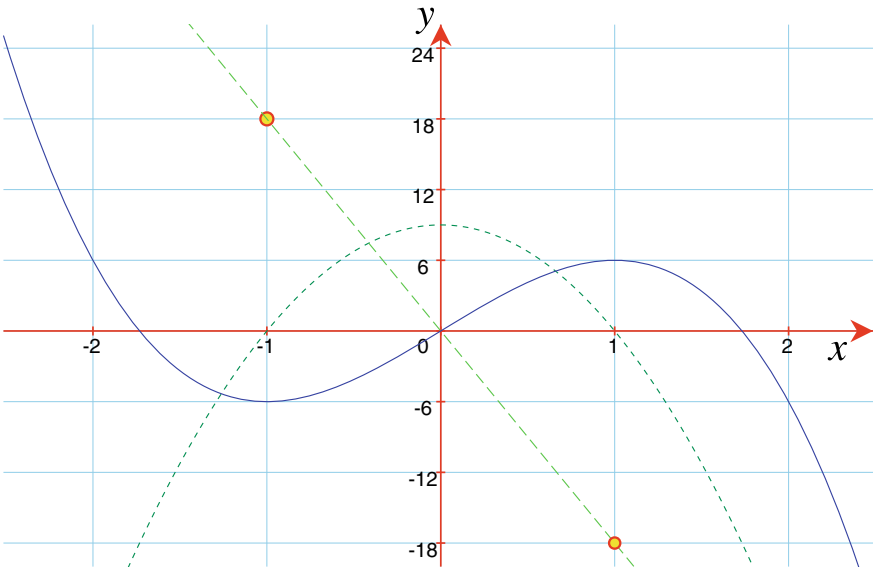

Fig. 15.24 Graph of $y = -3x^3 + 9x$, its first derivative, $\frac{dy}{dx} = -9x^2 + 9$ (short dashes) and its second derivative $y = -18x$ (long dashes)

$$-9x^2 + 9 = 0$$

which implies that $x = \pm 1$.

The sign of the second derivative determines whether there is a local minimum or maximum.

$$\begin{aligned}\frac{d^2y}{dx^2} &= -18x \\ &= -18 \times (-1) = +\text{ve} \\ &= -18 \times (+1) = -\text{ve}\end{aligned}$$

therefore, when $x = -1$ there is a local minimum, and when $x = +1$ there is a local maximum, as confirmed by Fig. 15.24.

15.18.8 Partial Derivatives

Find the second-order partial derivatives of f, given

$$f(u, v) = \sin(4u) \cdot \cos(5v).$$

Solution: the first partial derivatives are

$$\begin{aligned}\frac{\partial f}{\partial u} &= 4\cos(4u) \cdot \cos(5v) \\ \frac{\partial f}{\partial v} &= -5\sin(4u) \cdot \sin(5v)\end{aligned}$$

and the second-order partial derivatives are

$$\begin{aligned}\frac{\partial^2 f}{\partial u^2} &= -16\sin(4u) \cdot \cos(5v) \\ \frac{\partial^2 f}{\partial v^2} &= -25\sin(4u) \cdot \cos(5v).\end{aligned}$$

15.18.9 Mixed Partial Derivative 1

Given the formula for the volume of a cylinder is $V(r, h) = \pi r^2 h$, where r and h are the cylinder's radius and height respectively, compute the mixed partial derivative.
Solution:

$$
\begin{aligned}
V(r,h) &= \pi r^2 h \\
\frac{\partial V}{\partial r} &= 2\pi hr \\
\frac{\partial^2 V}{\partial h \partial r} &= 2\pi r
\end{aligned}
$$

or

$$
\begin{aligned}
V(r,h) &= \pi r^2 h \\
\frac{\partial V}{\partial h} &= \pi r^2 \\
\frac{\partial^2 V}{\partial r \partial h} &= 2\pi r.
\end{aligned}
$$

15.18.10 Mixed Partial Derivative 2

Given $f(u, v) = \sin(4u)\cos(3v)$, compute the mixed partial derivative.
Solution:

$$
\begin{aligned}
\frac{\partial f}{\partial u} &= 4\cos(4u) \cdot \cos(3v) \\
\frac{\partial^2 f}{\partial v \partial u} &= -12\cos(4u) \cdot \sin(3v)
\end{aligned}
$$

or

$$
\begin{aligned}
\frac{\partial f}{\partial v} &= -3\sin(4u) \cdot \sin(3v) \\
\frac{\partial^2 f}{\partial u \partial v} &= -12\cos(4u) \cdot \sin(3v).
\end{aligned}
$$

15.18.11 Total Derivative

Given

$$
\begin{aligned}
w &= x^2 + xy + y + t \\
x &= 2t \\
y &= t - 1
\end{aligned}
$$

compute the total derivative $\frac{dw}{dt}$.

Solution:

$$\begin{aligned}
\frac{dx}{dt} &= 2 \\
\frac{dy}{dt} &= 1 \\
\frac{\partial w}{\partial x} &= 2x + y = 4t + t - 1 = 5t - 1 \\
\frac{\partial w}{\partial y} &= x + 1 = 2t + 1 \\
\frac{\partial w}{\partial t} &= 1 \\
\frac{dw}{dt} &= \frac{\partial w}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial w}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial w}{\partial t} \\
&= (5t - 1)2 + (2t + 1) + 1
\end{aligned}$$

and the total derivative equals

$$\frac{dw}{dt} = 12t.$$

Chapter 16
Calculus: Integration

16.1 Introduction

In this chapter we develop the idea that integration is the inverse of differentiation, and explore the standard algebraic strategies for integrating functions, where the derivative is unknown; these include simple algebraic manipulation, trigonometric identities, integration by parts, integration by substitution and integration using partial fractions.

16.2 Indefinite Integral

In the previous chapter we have seen that given a simple function, such as

$$\begin{aligned} y &= \sin x + 23 \\ \frac{dy}{dx} &= \cos x \end{aligned}$$

and the constant term 23 disappears. Inverting the process, we have

$$\begin{aligned} y &= \int \cos x \; dx \\ &= \sin x + C. \end{aligned}$$

An integral of the form

$$\int f(x) \; dx$$

is known as an *indefinite integral*; and as we don't know whether the original function contains a constant term, a constant C has to be included. Its value remains undeter-

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9_16

mined unless we are told something about the original function. In this example, if we are told that when $x = \pi/2$, $y = 24$, then

$$\begin{aligned} 24 &= \sin \pi/2 + C \\ &= 1 + C \\ C &= 23. \end{aligned}$$

16.3 Integration Techniques

16.3.1 Continuous Functions

Functions come in all sorts of shapes and sizes, which is why we have to be very careful before they are differentiated or integrated. If a function contains any form of discontinuity, then it cannot be differentiated or integrated. For example, the square-wave function shown in Fig. 16.1 cannot be differentiated as it contains discontinuities. Consequently, to be very precise, we identify an *interval* $[a,\ b]$, over which a function is analysed, and stipulate that it must be continuous over this interval. For example, a and b define the upper and lower bounds of the interval such that

$$a \leq x \leq b$$

then we can say that for $f(x)$ to be continuous

$$\lim_{h \to 0} f(x + h) = f(x).$$

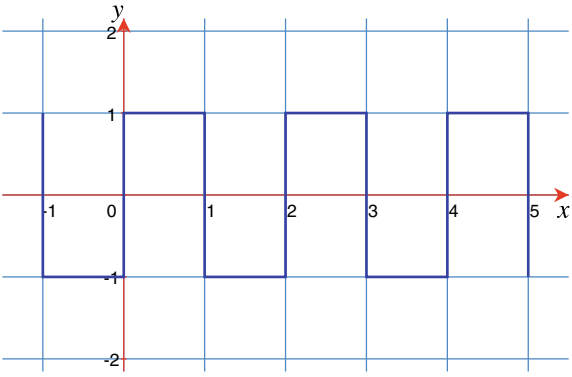

Fig. 16.1 A discontinuous square-wave function

Even this needs further clarification as h must not take x outside of the permitted interval. So, from now on, we assume that all functions are continuous and can be integrated without fear of singularities.

16.3.2 *Difficult Functions*

There are many functions that cannot be differentiated and represented by a finite collection of elementary functions. For example, the derivative $f'(x) = \sin x/x$ does not exist, which precludes the possibility of its integration. Figure 16.2 shows this function, and even though it is continuous, its derivative and integral can only be approximated. Similarly, the derivative $f'(x) = \sqrt{x}\sin x$ does not exist in a precise form, which precludes the possibility of finding a precise integral. Figure 16.3 shows this continuous function. So now let's examine how most functions have to be rearranged to secure their integration.

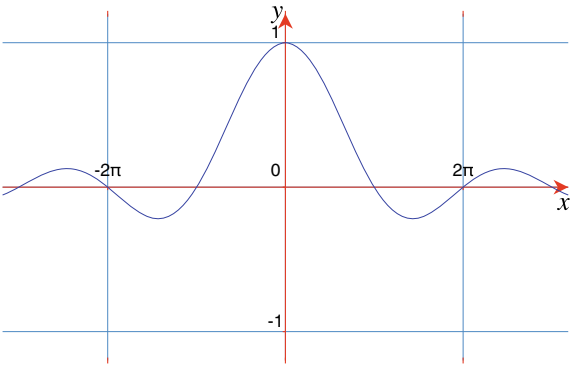

Fig. 16.2 Graph of $y = (\sin x)/x$

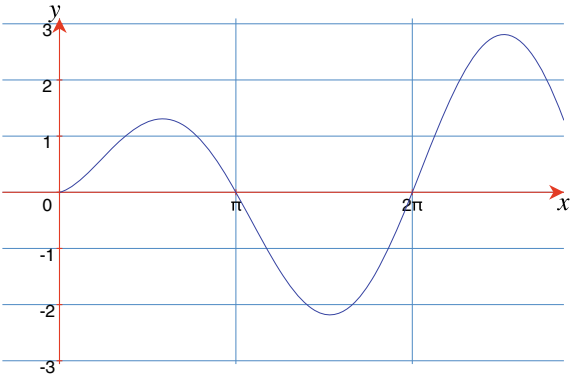

Fig. 16.3 Graph of $y = \sqrt{x}\sin x$

16.4 Trigonometric Identities

Sometimes it is possible to simplify the integrand by substituting a trigonometric identity. For example, let's evaluate $\int \sin^2 x\, dx$, $\int \cos^2 x\, dx$, $\int \tan^2 x\, dx$ and $\int \sin(3x)\cos x\, dx$.

The identity $\sin^2 x = \frac{1}{2}(1 - \cos(2x))$ converts $\sin^2 x$ into a double-angle form:

$$\begin{aligned}\int \sin^2 x\, dx &= \tfrac{1}{2}\int 1 - \cos(2x)\, dx \\ &= \tfrac{1}{2}\int dx - \tfrac{1}{2}\int \cos(2x)\, dx \\ &= \tfrac{1}{2}x - \tfrac{1}{4}\sin(2x) + C.\end{aligned}$$

Figure 16.4 shows the graphs of $y = \sin^2 x$ and $y = \frac{1}{2}x - \frac{1}{4}\sin(2x)$.

The identity $\cos^2 x = \frac{1}{2}(\cos(2x) + 1)$ converts $\cos^2 x$ into a double-angle form:

$$\begin{aligned}\int \cos^2 x\, dx &= \tfrac{1}{2}\int \cos(2x) + 1\, dx \\ &= \tfrac{1}{2}\int \cos(2x)\, dx + \tfrac{1}{2}\int dx \\ &= \tfrac{1}{4}\sin(2x) + \tfrac{1}{2}x + C.\end{aligned}$$

Figure 16.5 shows the graphs of $y = \cos^2 x$ and $y = \frac{1}{4}\sin(2x) + \frac{1}{2}x$.

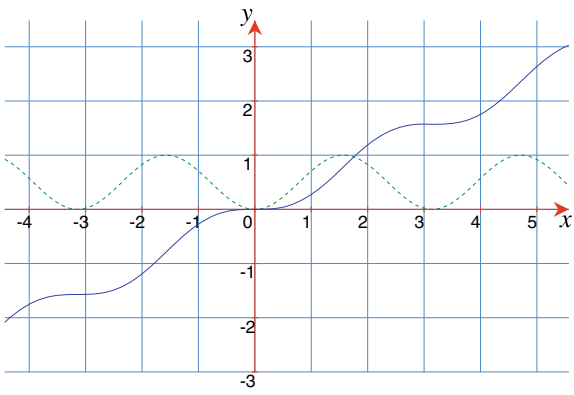

Fig. 16.4 The graphs of $y = \sin^2 x$ (dashed) and $y = \frac{1}{2}x - \frac{1}{4}\sin(2x)$

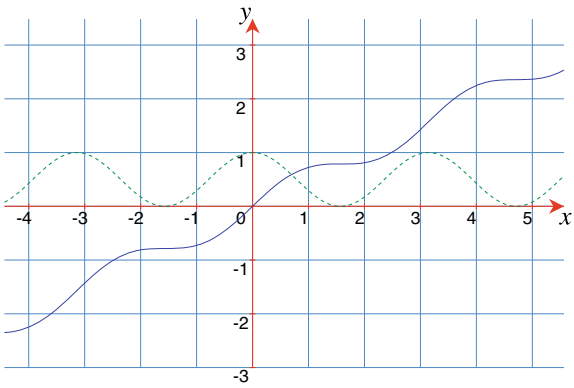

Fig. 16.5 The graphs of $y = \cos^2 x$ (dashed) and $y = \frac{1}{4}\sin(2x) + \frac{1}{2}x$

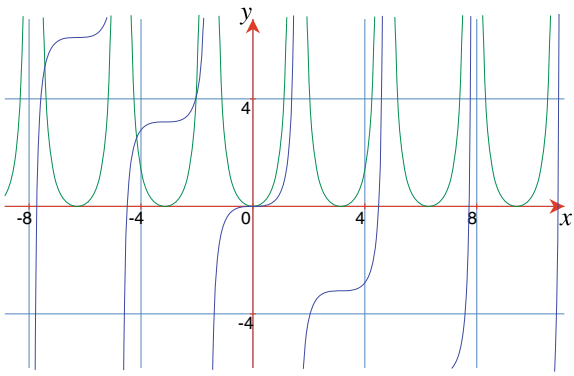

Fig. 16.6 The graphs of $y = \tan^2 x$ (dashed) and $y = \tan x - x$

The identity $\sec^2 x = 1 + \tan^2 x$, permits us to write

$$\begin{aligned}\int \tan^2 x \, dx &= \int \sec^2 x - 1 \, dx \\ &= \int \sec^2 x \, dx - \int dx \\ &= \tan x - x + C.\end{aligned}$$

Figure 16.6 shows the graphs of $y = \tan^2 x$ and $y = \tan x - x$.

Finally, to evaluate $\int \sin(3x) \cdot \cos x \, dx$ we use the identity

$$2 \sin a \cdot \cos b = \sin(a + b) + \sin(a - b)$$

which converts the integrand's product into the sum and difference of two angles:

$$\sin(3x) \cdot \cos x = \tfrac{1}{2}(\sin(4x) + \sin(2x))$$

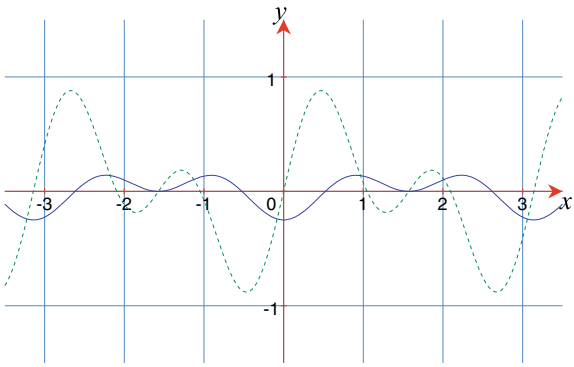

Fig. 16.7 The graphs of $y = \sin(3x) \cdot \cos x$ (dashed) and $y = -\frac{1}{8}\cos(4x) - \frac{1}{4}\cos(2x)$

$$\begin{aligned}\int \sin(3x) \cdot \cos x \, dx &= \tfrac{1}{2} \int \sin(4x) + \sin(2x) \, dx \\ &= \tfrac{1}{2} \int \sin(4x) \, dx + \tfrac{1}{2} \int \sin(2x) \, dx \\ &= -\tfrac{1}{8}\cos(4x) - \tfrac{1}{4}\cos(2x) + C.\end{aligned}$$

Figure 16.7 shows the graphs of $y = \sin(3x) \cdot \cos x$ and $y = -\frac{1}{8}\cos(4x) - \frac{1}{4}\cos(2x)$.

16.4.1 Exponent Notation

Radicals are best replaced by their equivalent exponent notation. For example, to evaluate

$$\int \frac{2}{\sqrt[4]{x}} \, dx$$

we proceed as follows:

The constant 2 is moved outside the integral, and the integrand is converted into an exponent form:

$$\begin{aligned}2 \int \frac{1}{\sqrt[4]{x}} \, dx &= 2 \int x^{-\frac{1}{4}} \\ &= 2 \left[\frac{x^{\frac{3}{4}}}{\frac{3}{4}} \right] + C \\ &= 2 \left[\frac{4}{3} x^{\frac{3}{4}} \right] + C \\ &= \tfrac{8}{3} x^{\frac{3}{4}} + C.\end{aligned}$$

Figure 16.8 shows the graphs of $y = 2/\sqrt[4]{x}$ and $y = 8x^{\frac{3}{4}}/3$.

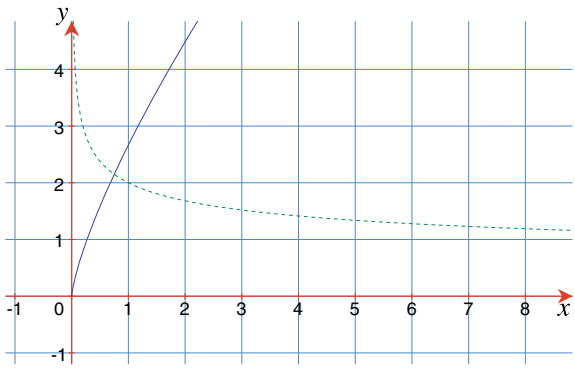

Fig. 16.8 The graphs of $y = 2/\sqrt[4]{x}$ (dashed) and $y = 8x^{\frac{3}{4}}/3$

16.4.2 Completing the Square

Where possible, see if an integrand can be simplified by completing the square. For example, to evaluate

$$\int \frac{1}{x^2 - 4x + 8}\, dx$$

we proceed as follows:

We have already seen that

$$\int \frac{1}{1 + x^2}\, dx = \arctan x + C$$

and it's not too difficult to prove that

$$\int \frac{1}{a^2 + x^2}\, dx = \frac{1}{a} \arctan\left(\frac{x}{a}\right) + C.$$

Therefore, if we can manipulate an integrand into this form, then the integral will reduce to an arctan result. The following needs no manipulation:

$$\int \frac{1}{4 + x^2}\, dx = \tfrac{1}{2} \arctan\left(\frac{x}{2}\right) + C.$$

However, the original integrand has $x^2 - 4x + 8$ as the denominator, which is resolved by completing the square:

$$x^2 - 4x + 8 = 4 + (x - 2)^2.$$

Therefore,

$$\int \frac{1}{x^2-4x+8}\,dx = \int \frac{1}{2^2+(x-2)^2}\,dx$$
$$= \tfrac{1}{2}\arctan\left(\frac{x-2}{2}\right)+C.$$

Figure 16.9 shows the graphs of $y = 1/(x^2-4x+8)$ and $y = \left(\arctan\frac{x-2}{2}\right)/2$.

To evaluate

$$\int \frac{1}{x^2+6x+10}\,dx.$$

we factorize the denominator:

$$\int \frac{1}{x^2+6x+10}\,dx = \int \frac{1}{1^2+(x+3)^2}\,dx$$
$$= \arctan(x+3)+C.$$

Figure 16.10 shows the graphs of $y = 1/(x^2+6x+10)$ and $y = \arctan(x+3)$.

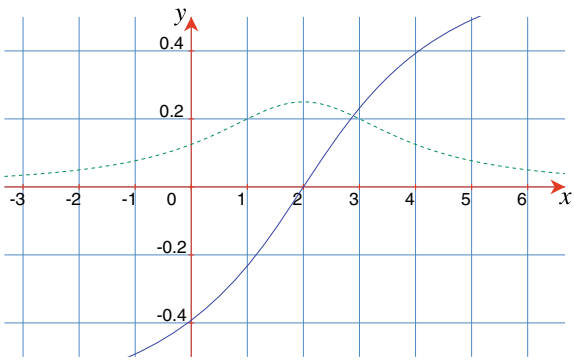

Fig. 16.9 The graphs of $y = 1/(x^2-4x+8)$ (dashed) and $y = \left(\arctan\frac{x-2}{2}\right)/2$

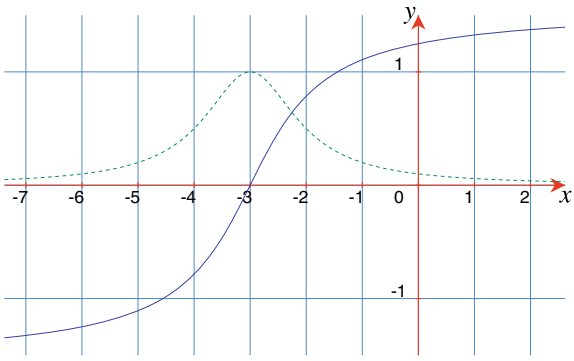

Fig. 16.10 The graphs of $y = 1/(x^2+6x+10)$ (dashed) and $y = \arctan(x+3)$

16.4.3 The Integrand Contains a Derivative

An integral of the form

$$\int \frac{f(x)}{f'(x)}\,dx$$

is relatively easy to integrate. For example, let's evaluate

$$\int \frac{\arctan x}{1+x^2}\,dx.$$

Knowing that

$$\frac{d}{dx}[\arctan x] = \frac{1}{1+x^2}$$

let $u = \arctan x$, then

$$\frac{du}{dx} = \frac{1}{1+x^2}$$

and

$$\begin{aligned}\int \frac{\arctan x}{1+x^2}\,dx &= \int u\,du \\ &= \frac{u^2}{2} + C \\ &= \tfrac{1}{2}(\arctan x)^2 + C.\end{aligned}$$

Figure 16.11 shows the graphs of $y = \arctan x/(1+x^2)$ and $y = \frac{1}{2}(\arctan x)^2$.

An integral of the form

$$\int \frac{f'(x)}{f(x)}\,dx$$

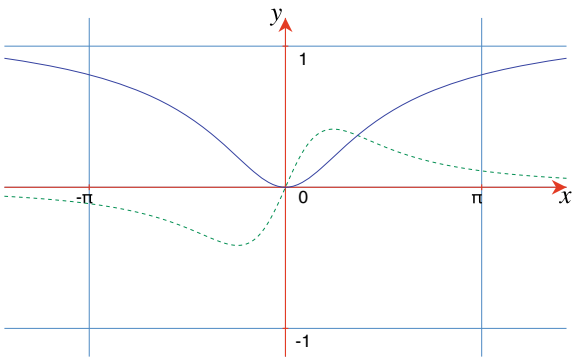

Fig. 16.11 The graphs of $y = \arctan x/(1+x^2)$ (dashed) and $y = \frac{1}{2}(\arctan x)^2$

is also relatively easy to integrate. For example, let's evaluate

$$\int \frac{\cos x}{\sin x}\, dx.$$

Knowing that

$$\frac{d}{dx}[\sin x] = \cos x$$

let $u = \sin x$, then

$$\frac{du}{dx} = \cos x$$

and

$$\begin{aligned}\int \frac{\cos x}{\sin x}\, dx &= \int \frac{1}{u}\, du \\ &= \ln|u| + C \\ &= \ln|\sin x| + C.\end{aligned}$$

Figure 16.12 shows the graphs of $y = \cos x / \sin x$ and $y = \ln|\sin x|$.

16.4.4 Converting the Integrand into a Series of Fractions

Integration is often made easier by converting an integrand into a series of fractions. For example, to integrate

$$\int \frac{4x^3 + x^2 - 8 + 12x\cos x}{4x}\, dx$$

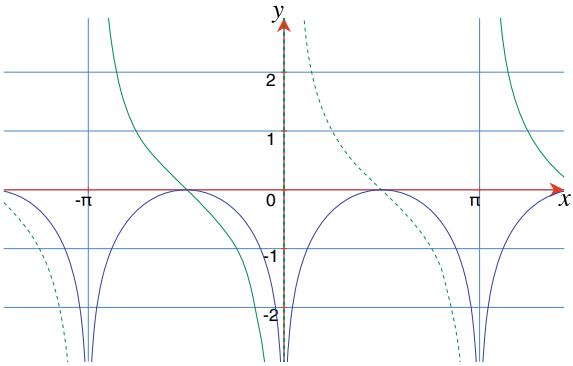

Fig. 16.12 The graphs of $y = \cos x / \sin x$ (dashed) and $y = \ln|\sin x|$

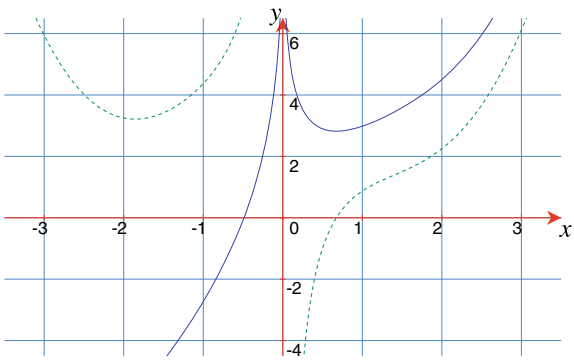

Fig. 16.13 The graphs of $y = (4x^3 + x^2 - 8 + 12x\cos x)/4x$ (dashed) and $y = x^3/3 + x^2/8 - 2\ln|x| + 3\sin x$

we divide the numerator by $4x$:

$$\int \frac{4x^3 + x^2 - 8 + 12x\cos x}{4x}\,dx = \int x^2\,dx + \int \frac{x}{4}\,dx - \int \frac{2}{x}\,dx + \int 3\cos x\,dx$$
$$= \frac{x^3}{3} + \frac{x^2}{8} - 2\ln|x| + 3\sin x + C.$$

Figure 16.13 shows the graphs of $y = (4x^3 + x^2 - 8 + 12x\cos x)/4x$ and $y = x^3/3 + x^2/8 - 2\ln|x| + 3\sin x$.

16.4.5 Integration by Parts

Integration by parts is based upon the rule for differentiating function products where

$$\frac{d}{dx}[uv] = u\frac{dv}{dx} + v\frac{du}{dx}$$

therefore,

$$uv = \int uv'\,dx + \int vu'\,dx$$

which rearranged, gives

$$\int uv'\,dx = uv - \int vu'\,dx.$$

Thus, if an integrand contains a product of two functions, we can attempt to integrate it by parts. For example, let's evaluate

$$\int x\sin x\,dx.$$

In this case, we try the following:

$$u = x \quad \text{and} \quad v' = \sin x$$

therefore

$$u' = 1 \quad \text{and} \quad v = C_1 - \cos x.$$

Integrating by parts:

$$\int uv' \, dx = uv - \int vu' \, dx$$

$$\begin{aligned} \int x \sin x \, dx &= x(C_1 - \cos x) - \int (C_1 - \cos x)(1) \, dx \\ &= C_1 x - x \cos x - C_1 x + \sin x + C \\ &= -x \cos x + \sin x + C. \end{aligned}$$

Figure 16.14 shows the graphs of $y = x \sin x$ and $y = -x \cos x + \sin x$.

Note the problems that arise if we make the wrong substitution:

$$u = \sin x \quad \text{and} \quad v' = x$$

therefore

$$u' = \cos x \quad \text{and} \quad v = \frac{x^2}{2} + C_1$$

Integrating by parts:

$$\int uv' \, dx = uv - \int vu' \, dx$$

$$\int x \sin x \, dx = \sin x \left(\frac{x^2}{2} + C_1 \right) - \int \left(\frac{x^2}{2} + C_1 \right) \cos x \, dx$$

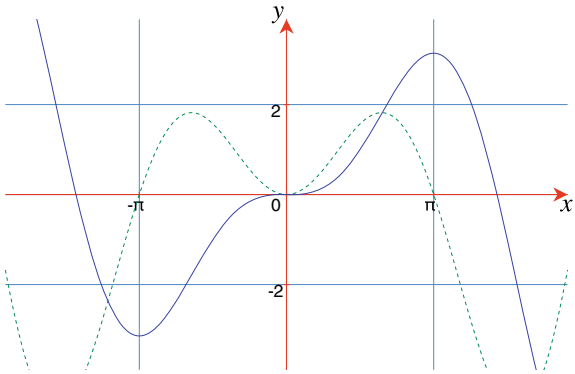

Fig. 16.14 The graphs of $y = x \sin x$ (dashed) and $y = -x \cos x + \sin x$

which requires to be integrated by parts, and is even more difficult, which suggests the substitution was not useful.

Now let's evaluate

$$\int x^2 \cos x \, dx.$$

In this case, we try the following:

$$u = x^2 \quad \text{and} \quad v' = \cos x$$

therefore

$$u' = 2x \quad \text{and} \quad v = \sin x + C_1.$$

Integrating by parts:

$$\begin{aligned}
\int uv' \, dx &= uv - \int vu' \, dx \\
\int x^2 \cos x \, dx &= x^2(\sin x + C_1) - 2\int (\sin x + C_1)(x) \, dx \\
&= x^2 \sin x + C_1 x^2 - 2C_1 \int x \, dx - 2\int x \sin x \, dx \\
&= x^2 \sin x + C_1 x^2 - 2C_1 \left(\frac{x^2}{2} + C_2 \right) - 2\int x \sin x \, dx \\
&= x^2 \sin x - C_3 - 2\int x \sin x \, dx.
\end{aligned}$$

At this point we come across $\int x \sin x \, dx$, which we have already solved:

$$\begin{aligned}
\int x^2 \cos x \, dx &= x^2 \sin x - C_3 - 2(-x \cos x + \sin x + C_4) \\
&= x^2 \sin x - C_3 + 2x \cos x - 2 \sin x - C_5 \\
&= x^2 \sin x + 2x \cos x - 2 \sin x + C
\end{aligned}$$

Figure 16.15 shows the graphs of $y = x^2 \cos x$ and $y = x^2 \sin x + 2x \cos x - 2 \sin x$.

Now let's evaluate

$$\int x \ln x \, dx.$$

In this case, we try the following:

$$u = \ln x \quad \text{and} \quad v' = x$$

therefore

$$u' = \frac{1}{x} \quad \text{and} \quad v = \tfrac{1}{2}x^2.$$

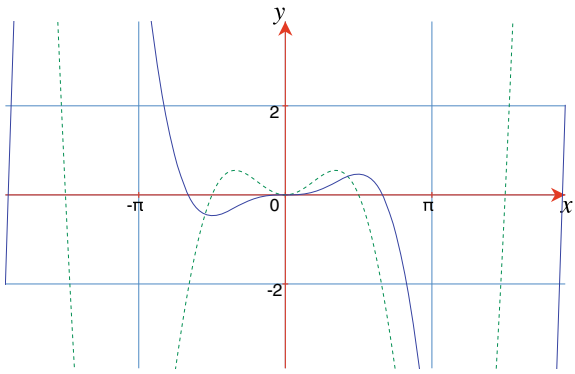

Fig. 16.15 The graphs of $y = x^2 \cos x$ (dashed) and $y = x^2 \sin x + 2x \cos x - 2 \sin x$

Integrating by parts:

$$\begin{aligned}
\int uv' \, dx &= uv - \int vu' \, dx \\
\int x \ln x \, dx &= \tfrac{1}{2}x^2 \ln x - \int \left(\tfrac{1}{2}x^2 \right) \frac{1}{x} \, dx \\
&= \tfrac{1}{2}x^2 \ln x - \tfrac{1}{2} \int x \, dx \\
&= \tfrac{1}{2}x^2 \ln x - \frac{x^2}{4} + C.
\end{aligned}$$

Figure 16.16 shows the graphs of $y = x \ln x$ and $y = \frac{1}{2}x^2 \ln x - x^2/4$.

Finally, let's evaluate

$$\int \sqrt{1 + x^2} \, dx.$$

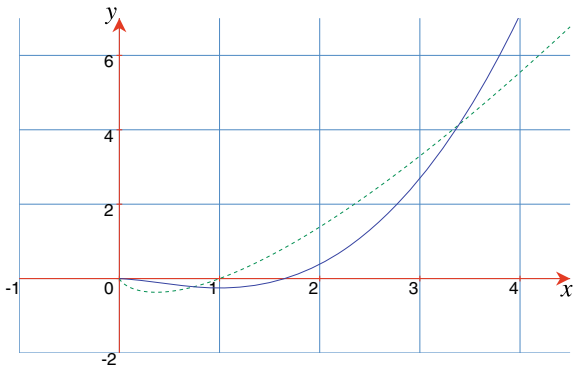

Fig. 16.16 The graphs of $y = x \ln x$ (dashed) and $y = \frac{1}{2}x^2 \ln x - x^2/4$

Although this integrand does not look as though it can be integrated by parts, if we rewrite it as

$$\int \sqrt{1+x^2}(1)\,dx.$$

then we can use the formula.

Let

$$u = \sqrt{1+x^2} \quad \text{and} \quad v' = 1$$

therefore

$$u' = \frac{x}{\sqrt{1+x^2}} \quad \text{and} \quad v = x.$$

Integrating by parts:

$$\int uv'\,dx = uv - \int vu'\,dx$$

$$\int \sqrt{1+x^2}\,dx = x\sqrt{1+x^2} - \int \frac{x^2}{\sqrt{1+x^2}}\,dx.$$

Now we simplify the right-hand integrand:

$$\begin{aligned}\int \sqrt{1+x^2}\,dx &= x\sqrt{1+x^2} - \int \frac{(1+x^2)-1}{\sqrt{1+x^2}}\,dx \\ &= x\sqrt{1+x^2} - \int \frac{1+x^2}{\sqrt{1+x^2}}\,dx + \int \frac{1}{\sqrt{1+x^2}}\,dx \\ &= x\sqrt{1+x^2} - \int \sqrt{1+x^2}\,dx + \operatorname{arsinh} x + C_1.\end{aligned}$$

Now we have the original integrand on the right-hand side, therefore

$$2\int \sqrt{1+x^2}\,dx = x\sqrt{1+x^2} + \operatorname{arsinh} x + C_1$$

$$\int \sqrt{1+x^2}\,dx = \tfrac{1}{2}x\sqrt{1+x^2} + \tfrac{1}{2}\operatorname{arsinh} x + C.$$

Figure 16.17 shows the graphs of $y = \sqrt{1+x^2}$ and $y = \frac{1}{2}x\sqrt{1+x^2} + \frac{1}{2}\operatorname{arsinh} x$.

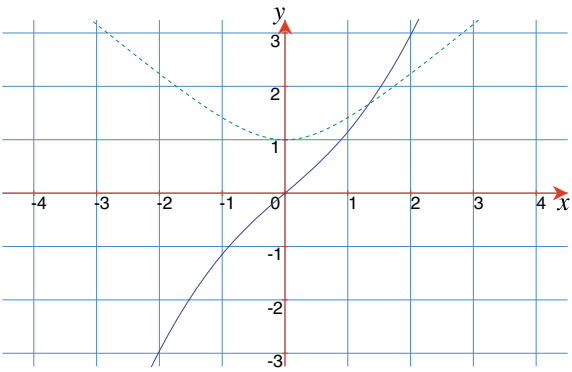

Fig. 16.17 The graphs of $y = \sqrt{1+x^2}$ (dashed) and $y = \frac{1}{2}x\sqrt{1+x^2} + \frac{1}{2}\text{arsinh}x$

16.4.6 Integration by Substitution

Integration by substitution is based upon the chain rule for differentiating a function of a function, which states that if y is a function of u, which in turn is a function of x, then

$$\frac{dy}{dx} = \frac{dy}{du}\,\frac{du}{dx}.$$

For example, let's evaluate

$$\int x^2\sqrt{x^3}\,dx.$$

This is easily solved by rewriting the integrand:

$$\begin{aligned}\int x^2\sqrt{x^3}\,dx &= \int x^{\frac{7}{2}}\,dx \\ &= \tfrac{2}{9}x^{\frac{9}{2}} + C.\end{aligned}$$

However, introducing a constant term within the square-root requires integration by substitution. For example,

$$\text{evaluate} \quad \int x^2\sqrt{x^3+1}\,dx.$$

First, we let $u = x^3 + 1$, then

$$\frac{du}{dx} = 3x^2 \quad \text{or} \quad dx = \frac{du}{3x^2}.$$

Substituting u and dx in the integrand gives

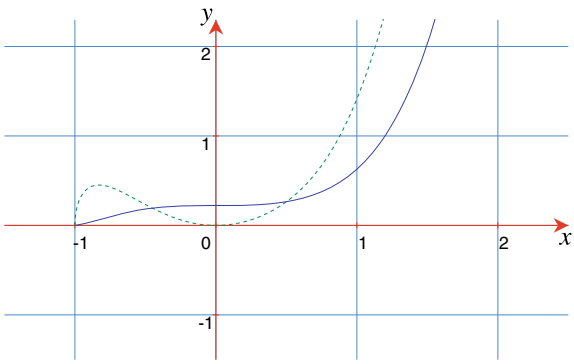

Fig. 16.18 The graphs of $y = x^2\sqrt{x^3+1}$ (dashed) and $y = \frac{2}{9}(x^3+1)^{\frac{3}{2}}$

$$\begin{aligned}\int x^2\sqrt{x^3+1}\,dx &= \int x^2\sqrt{u}\,\frac{du}{3x^2}\\ &= \tfrac{1}{3}\int \sqrt{u}\,du\\ &= \tfrac{1}{3}\int u^{\frac{1}{2}}\,du\\ &= \tfrac{1}{3}\cdot\tfrac{2}{3}u^{\frac{3}{2}}+C\\ &= \tfrac{2}{9}(x^3+1)^{\frac{3}{2}}+C.\end{aligned}$$

Figure 16.18 shows the graphs of $y = x^2\sqrt{x^3+1}$ and $y = \frac{2}{9}(x^3+1)^{\frac{3}{2}}$.

Now let's evaluate

$$\int 2\sin x\cdot\cos x\,dx.$$

Integrating by substitution we let $u = \sin x$, then

$$\frac{du}{dx} = \cos x \quad \text{or} \quad dx = \frac{du}{\cos x}.$$

Substituting u and dx in the integrand gives

$$\begin{aligned}\int 2\sin x\cdot\cos x\,dx &= 2\int u\cos x\,\frac{du}{\cos x}\\ &= 2\int u\,du\\ &= u^2+C_1\\ &= \sin^2 x+C.\end{aligned}$$

Figure 16.19 shows the graphs of $y = 2\sin x\cdot\cos x$ and $y = \sin^2 x$.

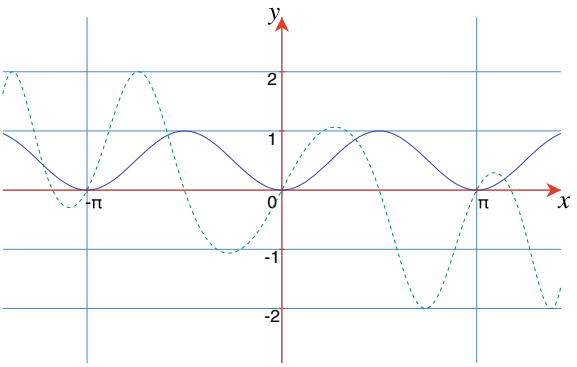

Fig. 16.19 The graphs of $y = 2\sin x \cdot \cos x$ (dashed) and $y = \sin^2 x$

16.4.7 Partial Fractions

Integration by *partial fractions* is used when an integrand's denominator contains a product that can be split into two fractions. For example, it should be possible to convert

$$\int \frac{3x+4}{(x+1)(x+2)}\,dx$$

into

$$\int \frac{A}{x+1}\,dx + \int \frac{B}{x+2}\,dx$$

which individually, are easy to integrate. Let's compute A and B:

$$\begin{aligned}\frac{3x+4}{(x+1)(x+2)} &= \frac{A}{x+1} + \frac{B}{x+2}\\ 3x+4 &= A(x+2) + B(x+1)\\ &= Ax + 2A + Bx + B.\end{aligned}$$

Equating constants and terms in x:

$$4 = 2A + B \tag{16.1}$$

$$3 = A + B \tag{16.2}$$

Subtracting (16.2) from (16.1), gives $A = 1$ and $B = 2$. Therefore,

$$\begin{aligned}\int \frac{3x+4}{(x+1)(x+2)}\,dx &= \int \frac{1}{x+1}\,dx + \int \frac{2}{x+2}\,dx\\ &= \ln(x+1) + 2\ln(x+2) + C.\end{aligned}$$

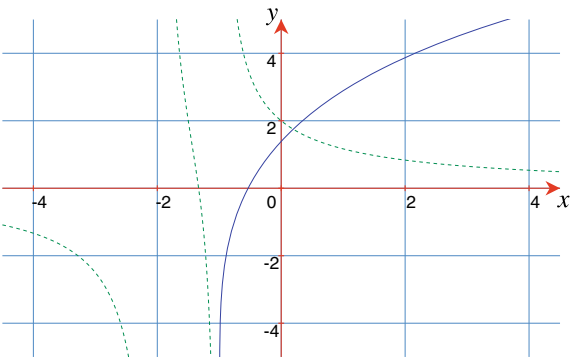

Fig. 16.20 The graphs of $y = (3x + 4)/((x + 1)(x + 2))$ (dashed) and $y = \ln(x + 1) + 2\ln(x + 2)$

Figure 16.20 shows the graphs of $y = (3x + 4)/((x + 1)(x + 2))$ and $y = \ln(x + 1) + 2\ln(x + 2)$.

Now let's evaluate

$$\int \frac{5x - 7}{(x - 1)(x - 2)}\, dx.$$

Integrating by partial fractions:

$$\begin{aligned} \frac{5x - 7}{(x - 1)(x - 2)} &= \frac{A}{x - 1} + \frac{B}{x - 2} \\ 5x - 7 &= A(x - 2) + B(x - 1) \\ &= Ax + Bx - 2A - B. \end{aligned}$$

Equating constants and terms in x:

$$-7 = -2A - B \tag{16.3}$$

$$5 = A + B \tag{16.4}$$

Subtracting (16.3) from (16.4), gives $A = 2$ and $B = 3$. Therefore,

$$\begin{aligned} \int \frac{3x + 4}{(x - 1)(x - 2)}\, dx &= \int \frac{2}{x - 1}\, dx + \int \frac{3}{x - 2}\, dx \\ &= 2\ln(x - 1) + 3\ln(x - 2) + C. \end{aligned}$$

Figure 16.21 shows the graphs of $y = (5x - 7)/((x - 1)(x - 2))$ and $y = 2\ln(x - 1) + 3\ln(x - 2)$.

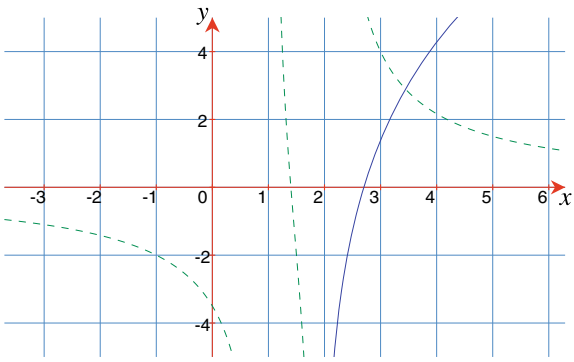

Fig. 16.21 The graphs of $y = (5x - 7)/((x - 1)(x - 2))$ (dashed) and $y = 2\ln(x - 1) + 3\ln(x - 2)$

16.5 Area Under a Graph

The ability to calculate the area under a graph is one of the most important discoveries of integral calculus. Prior to calculus, area was computed by dividing a zone into very small strips and summing the individual areas. The accuracy of the result is improved simply by making the strips smaller and smaller, taking the result towards some limiting value. In this section, we discover how integral calculus provides a way to compute the area between a function's graph and the x- and y-axis.

16.6 Calculating Areas

Before considering the relationship between area and integration, let's see how area is calculated using functions and simple geometry.

Figure 16.22 shows the graph of $y = 1$, where the area A of the shaded zone is

$$A = x, \quad x > 0.$$

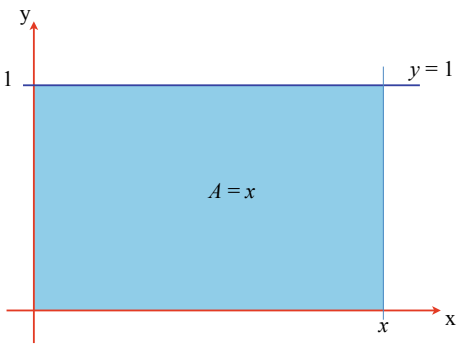

Fig. 16.22 Area of the shaded zone is $A = x$

For example, when $x = 4$, $A = 4$, and when $x = 10$, $A = 10$. An interesting observation is that the original function is the derivative of A:

$$\frac{dA}{dx} = 1 = y.$$

Figure 16.23 shows the graph of $y = 2x$. The area A of the shaded triangle is

$$\begin{aligned} A &= \tfrac{1}{2} base \times height \\ &= \tfrac{1}{2} x \times 2x \\ &= x^2. \end{aligned}$$

Thus, when $x = 4$, $A = 16$. Once again, the original function is the derivative of A:

$$\frac{dA}{dx} = 2x = y$$

which is no coincidence.

Finally, Fig. 16.24 shows a circle where $x^2 + y^2 = r^2$, and the curve of the first quadrant is described by the function

$$y = \sqrt{r^2 - x^2}, \quad 0 \geq x \geq r.$$

The total area of the shaded zones is the sum of the two parts A_1 and A_2. To simplify the calculations the function is defined in terms of the angle θ, such that

$$x = r \sin \theta$$

and

$$y = r \cos \theta.$$

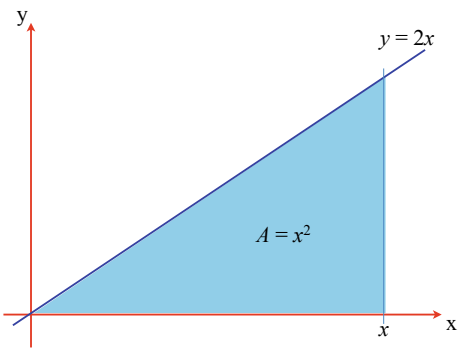

Fig. 16.23 Area of the shaded zone is $A = x^2$

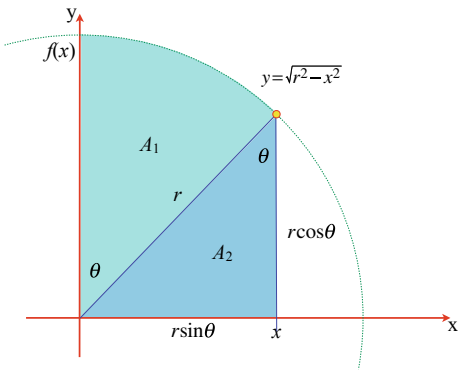

Fig. 16.24 Graph of $y = \sqrt{r^2 - x^2}$

Therefore,

$$\begin{aligned} A_1 &= \frac{r^2\theta}{2} \\ A_2 &= \tfrac{1}{2}(r\cos\theta)(r\sin\theta) = \frac{r^2}{4}\sin(2\theta) \\ A &= A_1 + A_2 \\ &= \frac{r^2}{2}\left(\theta + \frac{\sin(2\theta)}{2}\right). \end{aligned}$$

To show that the total area is related to the function's derivative, let's differentiate A with respect to θ:

$$\frac{dA}{d\theta} = \frac{r^2}{2}(1 + \cos(2\theta)) = r^2\cos^2\theta.$$

But we want the derivative $\frac{dA}{dx}$, which requires the chain rule

$$\frac{dA}{dx} = \frac{dA}{d\theta}\frac{d\theta}{dx}$$

where

$$\frac{dx}{d\theta} = r\cos\theta$$

or

$$\frac{d\theta}{dx} = \frac{1}{r\cos\theta}$$

therefore,

$$\frac{dA}{dx} = \frac{r^2\cos^2\theta}{r\cos\theta} = r\cos\theta = y$$

which is the equation for the quadrant.

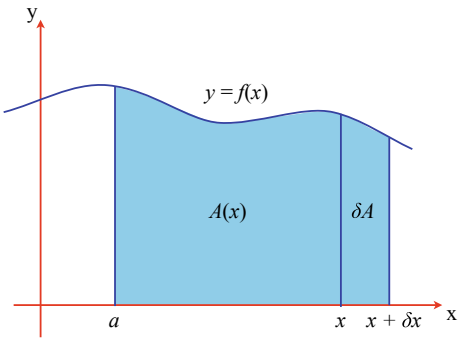

Fig. 16.25 Relationship between $y = f(x)$ and $A(x)$

Hopefully, these three examples provide strong evidence that the derivative of the function for the area under a graph, equals the graph's function:

$$\frac{dA}{dx} = f(x)$$

which implies that

$$A = \int f(x)\, dx.$$

Now let's prove this observation using Fig. 16.25, which shows a continuous function $y = f(x)$. Next, we define a function $A(x)$ to represent the area under the graph over the interval $[a,\ x]$. δA is the area increment between x and $x + \delta x$, and

$$\delta A \approx f(x) \cdot \delta x.$$

We can also reason that

$$\delta A = A(x + \delta x) - A(x) \approx f(x) \cdot \delta x$$

and the derivative $\frac{dA}{dx}$ is the limiting condition:

$$\frac{dA}{dx} = \lim_{\delta x \to 0} \frac{A(x + \delta x) - A(x)}{\delta x} = \lim_{\delta x \to 0} \frac{f(x) \cdot \delta x}{\delta x} = f(x)$$

thus,

$$\frac{dA}{dx} = f(x),$$

whose antiderivative is

$$A(x) = \int f(x)\, dx.$$

The function $A(x)$ computes the area over the interval $[a,\ b]$ and is represented by

$$A(x) = \int_a^b f(x)\ dx$$

which is called *the integral* or *definite integral*.

Let's assume that $A(b)$ is the area under the graph of $f(x)$ over the interval $[0,\ b]$, as shown in Fig. 16.26, and is written

$$A(b) = \int_0^b f(x)\ dx.$$

Similarly, let $A(a)$ be the area under the graph of $f(x)$ over the interval $[0,\ a]$, as shown in Fig. 16.27, and is written

$$A(a) = \int_0^a f(x)\ dx.$$

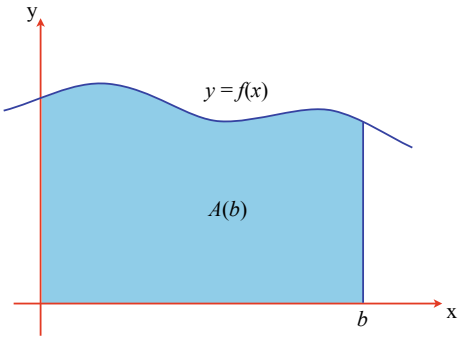

Fig. 16.26 $A(b)$ is the area under the graph $y = f(x)$, $0 \geq x \geq b$

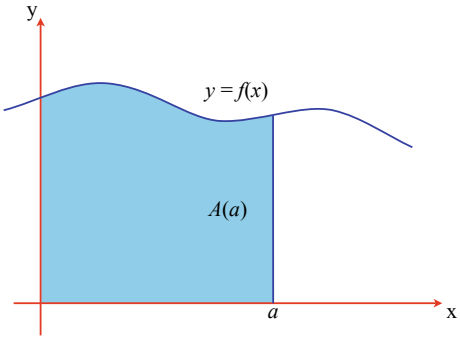

Fig. 16.27 $A(a)$ is the area under the graph $y = f(x)$, $0 \geq x \geq a$

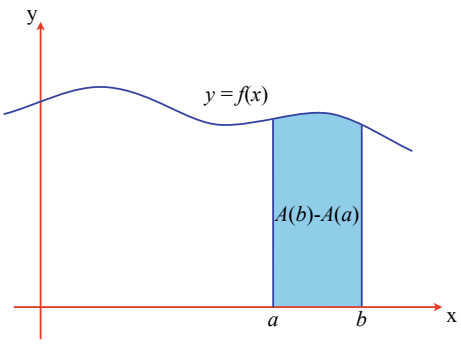

Fig. 16.28 $A(b) - A(a)$ is the area under the graph $y = f(x),\ \ a \geq x \geq b$

Figure 16.28 shows that the area of the shaded zone over the interval $[a,\ b]$ is calculated by

$$A = A(b) - A(a)$$

which is written

$$A = \int_0^b f(x)\, dx - \int_0^a f(x)\, dx$$

and is contracted to

$$A = \int_a^b f(x)\, dx. \tag{16.5}$$

The *fundamental theorem of calculus* states that the definite integral

$$\int_a^b f(x)\, dx = F(b) - F(a)$$

where

$$F(a) = \int f(x)\, dx, \quad x = a$$
$$F(b) = \int f(x)\, dx, \quad x = b.$$

In order to compute the area beneath a graph of $f(x)$ over the interval $[a,\ b]$, we first integrate the graph's function

$$F(x) = \int f(x)\, dx$$

and then calculate the area, which is the difference

$$A = F(b) - F(a).$$

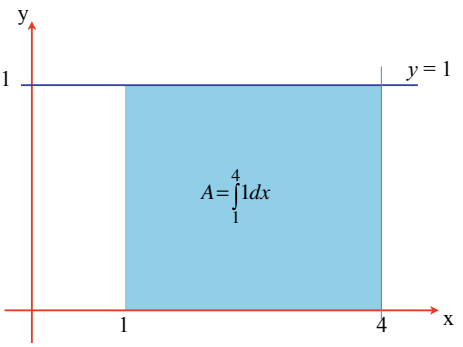

Fig. 16.29 Area under the graph is $\int_1^4 1\ dx$

To illustrate how (16.5) is used in the context of the earlier three examples, let's calculate the area over the interval $[1,\ 4]$ for $y = 1$, as shown in Fig. 16.29. We begin with

$$A = \int_1^4 1\ dx.$$

Next, we integrate the function, and transfer the interval bounds employing the *substitution symbol* $\Big|_1^4$, or square brackets $\Big[\ \Big]_1^4$. Using $\Big|_1^4$, we have

$$\begin{aligned} A &= \Big|_1^4\ x \\ &= 4 - 1 \\ &= 3 \end{aligned}$$

or using $\Big[\ \Big]_1^4$, we have

$$\begin{aligned} A &= \Big[\ x\ \Big]_1^4 \\ &= 4 - 1 \\ &= 3. \end{aligned}$$

I will continue with square brackets.

Now let's calculate the area over the interval $[1,\ 4]$ for $y = 2x$, as shown in Fig. 16.30. We begin with

$$A = \int_1^4 2x\ dx.$$

Next, we integrate the function and evaluate the area

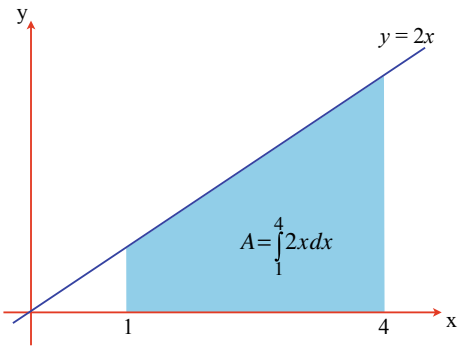

Fig. 16.30 Area under the graph is $\int_1^4 2x\ dx$

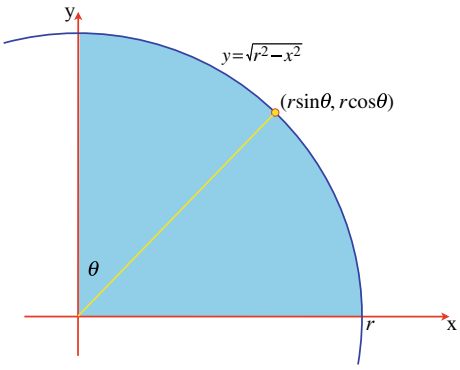

Fig. 16.31 Area under the graph is $\int_0^r \sqrt{r^2 - x^2}\ dx$

$$\begin{aligned} A &= \Big[\ x^2\ \Big]_1^4 \\ &= 16 - 1 \\ &= 15. \end{aligned}$$

Finally, let's calculate the area over the interval $[0,\ r]$ for $y = \sqrt{r^2 - x^2}$, which is the equation for a circle, as shown in Fig. 16.31. We begin with

$$A = \int_0^r \sqrt{r^2 - x^2}\ dx. \tag{16.6}$$

Unfortunately, (16.6) contains a function of a function, which is resolved by substituting another independent variable. In this case, the geometry of the circle suggests

$$x = r \sin\theta$$

therefore,

$$\sqrt{r^2 - x^2} = r \cos\theta$$

and

$$\frac{dx}{d\theta} = r\cos\theta. \tag{16.7}$$

However, changing the independent variable requires changing the interval for the integral. In this case, changing $0 \geq x \geq r$ into $\theta_1 \geq \theta \geq \theta_2$:

When $x = 0$, $r\sin\theta_1 = 0$, therefore $\theta_1 = 0$.

When $x = r$, $r\sin\theta_2 = r$, therefore $\theta_2 = \pi/2$.

Thus, the new interval is $[0, \pi/2]$.

Finally, the dx in (16.6) has to be changed into $d\theta$, which using (16.7) makes

$$dx = r\cos\theta \; d\theta.$$

Now we are in a position to rewrite the original integral using θ as the independent variable:

$$\begin{aligned} A &= \int_0^{\frac{\pi}{2}} (r\cos\theta)(r\cos\theta) \; d\theta \\ &= r^2 \int_0^{\frac{\pi}{2}} \cos^2\theta \; d\theta \\ &= \frac{r^2}{2} \int_0^{\frac{\pi}{2}} 1 + \cos(2\theta) \; d\theta \\ &= \frac{r^2}{2} \left[\theta + \tfrac{1}{2}\sin(2\theta) \right]_0^{\frac{\pi}{2}} \\ &= \frac{r^2}{2} \left[\frac{\pi}{2} \right] \\ &= \frac{\pi r^2}{4} \end{aligned}$$

which makes the area of a full circle πr^2.

16.7 Positive and Negative Areas

Area in the real world is always regarded as a positive quantity—no matter how it is measured. In mathematics, however, area is often a signed quantity, and is determined by the clockwise or anticlockwise direction of vertices. As we generally use a left-handed Cartesian axial system in calculus, areas above the x-axis are positive, whilst areas below the x-axis are negative. This can be illustrated by computing the area of the positive and negative parts of a sine wave.

Figure 16.32 shows a sketch of a sine wave over one cycle, where the area above the x-axis is labelled A_1, and the area below the x-axis is labelled A_2. These areas

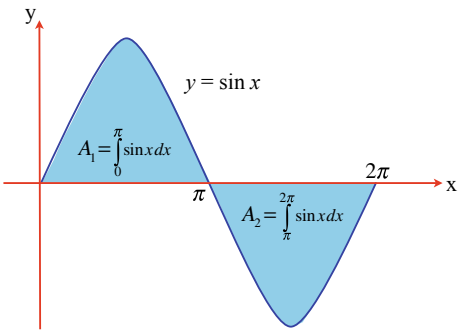

Fig. 16.32 The two areas associated with a sine wave

are computed as follows.

$$\begin{aligned} A_1 &= \int_0^{\pi} \sin x \; dx \\ &= \Big[-\cos x \Big]_0^{\pi} \\ &= \Big[1 + 1 \Big] \\ &= 2. \end{aligned}$$

However, A_2 gives a negative result:

$$\begin{aligned} A_2 &= \int_{\pi}^{2\pi} \sin x \; dx \\ &= \Big[-\cos x \Big]_{\pi}^{2\pi} \\ &= \Big[-1 - 1 \Big] \\ &= -2. \end{aligned}$$

This means that the area is zero over the bounds 0 to 2π, .

$$\begin{aligned} A_2 &= \int_0^{2\pi} \sin x \; dx \\ &= \Big[-\cos x \Big]_0^{2\pi} \\ &= \Big[-1 + 1 \Big] \\ &= 0. \end{aligned}$$

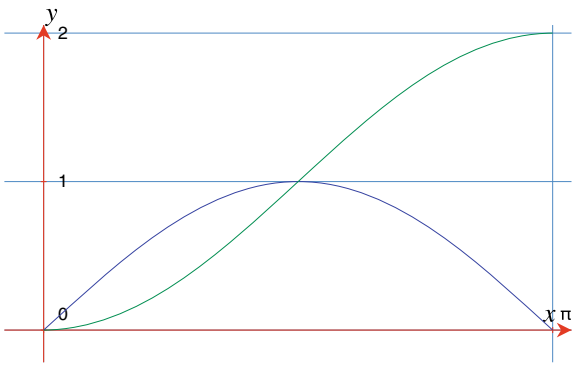

Fig. 16.33 The accumulated area of a sine wave

Consequently, one must be very careful using this technique for functions that are negative in the interval under investigation. Figure 16.33 shows a sine wave over the interval $[0,\ \pi]$ and its accumulated area.

16.8 Area Between Two Functions

Figure 16.34 shows the graphs of $y = x^2$ and $y = x^3$, with two areas labelled A_1 and A_2. A_1 is the area trapped between the two graphs over the interval $[-1,\ 0]$ and A_2 is the area trapped between the two graphs over the interval $[0,\ 1]$. These areas are calculated very easily: in the case of A_1 we sum the individual areas under the two graphs, remembering to reverse the sign for the area associated with $y = x^3$. For A_2 we subtract the individual areas under the two graphs.

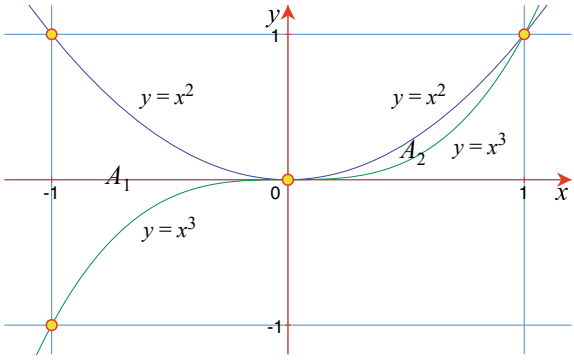

Fig. 16.34 Two areas between $y = x^2$ and $y = x^3$

$$
\begin{aligned}
A_1 &= \int_{-1}^{0} x^2\,dx - \int_{-1}^{0} x^3\,dx \\
&= \left[\frac{x^3}{3}\right]_{-1}^{0} - \left[\frac{x^4}{4}\right]_{-1}^{0} \\
&= \tfrac{1}{3} + \tfrac{1}{4} \\
&= \tfrac{7}{12}.
\end{aligned}
$$

$$
\begin{aligned}
A_2 &= \int_{0}^{1} x^2\,dx - \int_{0}^{1} x^3\,dx \\
&= \left[\frac{x^3}{3}\right]_{0}^{1} - \left[\frac{x^4}{4}\right]_{0}^{1} \\
&= \tfrac{1}{3} - \tfrac{1}{4} \\
&= \tfrac{1}{12}.
\end{aligned}
$$

Note, that in both cases the calculation is the same, which implies that when we employ

$$A = \int_a^b \Big[\, f(x) - g(x) \,\Big]\, dx$$

A is always the area trapped between $f(x)$ and $g(x)$ over the interval $[a,\ b]$.

Let's take another example, by computing the area A between $y = \sin x$ and the line $y = 0.5$, as shown in Fig. 16.35. The horizontal line intersects the sine curve at $x = 30°$ and $x = 150°$, marked in radians as 0.5236 and 2.618 respectively.

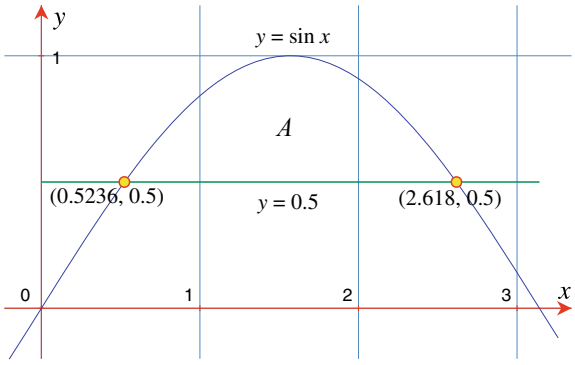

Fig. 16.35 The area between $y = \sin x$ and $y = 0.5$

$$
\begin{aligned}
A &= \int_{30^\circ}^{150^\circ} \sin x \; dx - \int_{\pi/6}^{5\pi/6} 0.5 \; dx \\
&= \Big[-\cos x \Big]_{30^\circ}^{150^\circ} - \tfrac{1}{2} \Big[x \Big]_{\pi/6}^{5\pi/6} \\
&= \Big[\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2} \Big] - \tfrac{1}{2} \Big[\frac{5\pi}{6} - \frac{\pi}{6} \Big] \\
&= \sqrt{3} - \frac{\pi}{3} \\
&\approx 0.685.
\end{aligned}
$$

16.9 Areas with the y-Axis

So far we have only calculated areas between a function and the x-axis. So let's compute the area between a function and the y-axis. Figure 16.36 shows the function $y = x^2$ over the interval $[0,\ 4]$, where A_1 is the area between the curve and the x-axis, and A_2 is the area between the curve and y-axis. The sum $A_1 + A_2$ must equal $4 \times 16 = 64$, which is a useful control. Let's compute A_1.

$$
\begin{aligned}
A_1 &= \int_0^4 x^2 \; dx \\
&= \Big[\frac{x^3}{3} \Big]_0^4 \\
&= \tfrac{64}{3} \\
&\approx 21.333
\end{aligned}
$$

which means that $A_2 \approx 42.666$. To compute A_2 we construct an integral relative to dy with a corresponding interval. If $y = x^2$ then $x = y^{\frac{1}{2}}$, and the interval is $[0,\ 16]$:

$$
\begin{aligned}
A_2 &= \int_0^{16} y^{\frac{1}{2}} \; dy \\
&= \Big[\tfrac{2}{3} y^{\frac{3}{2}} \Big]_0^{16} \\
&= \tfrac{2}{3} 64 \\
&\approx 42.666.
\end{aligned}
$$

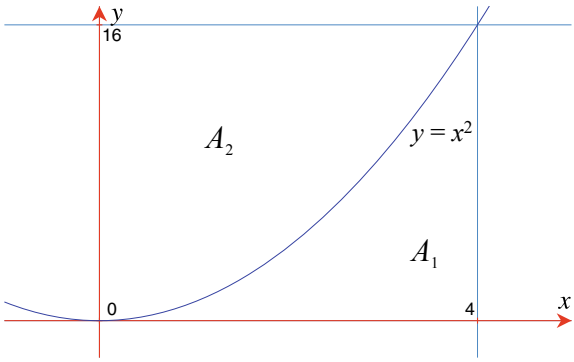

Fig. 16.36 The areas between the x-axis and the y-axis

16.10 Area with Parametric Functions

When working with functions of the form $y = f(x)$, the area under its curve and the x-axis over the interval $[a,\ b]$ is

$$A = \int_a^b f(x)\ dx.$$

However, if the curve has a parametric form where

$$x = f_x(t) \quad \text{and} \quad y = f_y(t)$$

then we can derive an equivalent integral as follows.

First: We need to establish equivalent limits $[\alpha, \beta]$ for t, such that

$$a = f_x(\alpha) \quad \text{and} \quad b = f_y(\beta).$$

Second: Any point on the curve has corresponding Cartesian and parametric coordinates:

$$x \quad \text{and} \quad f_x(t)$$

$$y = f(x) \quad \text{and} \quad f_y(t).$$

Third:

$$\begin{aligned} x &= f_x(t) \\ dx &= f_x'(t)dt \\ A &= \int_a^b f(x)\ dx \\ &= \int_\alpha^\beta f_y(t) f_x'(t)\ dt \end{aligned}$$

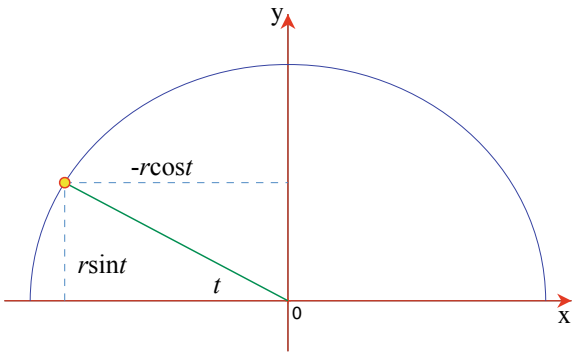

Fig. 16.37 The parametric functions for a circle

therefore

$$A = \int_{\alpha}^{\beta} f_y(t) f_x'(t)\, dt. \tag{16.8}$$

Let's apply (16.8) using the parametric equations for a circle

$$\begin{aligned} x &= -r\cos t \\ y &= r\sin t. \end{aligned}$$

as shown in Fig. 16.37. Remember that the Cartesian interval is $[a, b]$ left to right, and the polar interval $[\alpha,\ \beta]$, must also be left to right, which is why $x = -r\cos t$. Therefore,

$$\begin{aligned} f_x' t &= r\sin t \\ f_y(t) &= r\sin t \\ A &= \int_{\alpha}^{\beta} f_y(t) f_x'(t)\, dt \\ &= \int_0^{\pi} r\sin t \cdot r\sin(t)\, dt \\ &= r^2 \int_0^{\pi} \sin^2 t\, dt \\ &= \frac{r^2}{2} \int_0^{\pi} 1 - \cos(2t)\, dt \\ &= \frac{r^2}{2}\left[t + \tfrac{1}{2}\sin(2t)\right]_0^{\pi} \\ &= \frac{\pi r^2}{2} \end{aligned}$$

which makes the area of a full circle πr^2.

16.11 The Riemann Sum

The German mathematician Bernhard Riemann (1826–1866) (pronounced "Reeman") made major contributions to various areas of mathematics, including integral calculus, where his name is associated with a formal method for summing areas and volumes. Through the *Riemann Sum*, Riemann provides an elegant and consistent notation for describing single, double and triple integrals when calculating area and volume. Let's see how the Riemann sum explains why the area under a curve is the function's integral.

Figure 16.38 shows a function $f(x)$ divided into eight equal sub-intervals where

$$\Delta x = \frac{b-a}{8}$$

and

$$a = x_0 < x_1 < x_2 < \cdots < x_7 < x_8 = b.$$

In order to compute the area under the curve over the interval $[a,\ b]$, the interval is divided into some large number of sub-intervals. In this case, eight, which is not very large, but convenient to illustrate. Each sub-interval becomes a rectangle with a common width Δx and a different height. The area of the first rectangular sub-interval shown shaded, can be calculated in various ways. We can take the left-most height x_0 and form the product $x_0\Delta x$, or we can take the right-most height x_1 and form the product $x_1\Delta x$. On the other hand, we could take the mean of the two heights $(x_0 + x_1)/2$ and form the product $(x_0 + x_1)\Delta x/2$. A solution that shows no bias towards either left, right or centre, is to let x_i^* be anywhere in a specific sub-interval Δx_i, then the area of the rectangle associated with the sub-interval is $f(x_i^*)\Delta x_i$, and the sum of the rectangular areas is given by

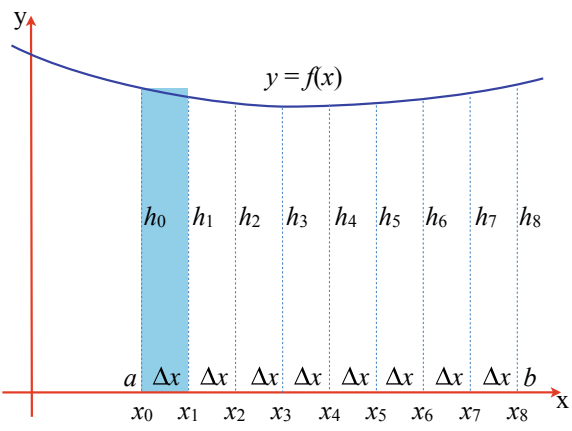

Fig. 16.38 The graph of function $f(x)$ over the interval $[a,\ b]$

$$A = \sum_{i=1}^{8} f(x_i^*)\Delta x_i.$$

Dividing the interval into eight equal sub-intervals will not generate a very accurate result for the area under the graph. But increasing it to eight-thousand or eight-million, will take us towards some limiting value. Rather than specify some specific large number, it is common practice to employ n, and let n tend towards infinity, which is written

$$A = \sum_{i=1}^{n} f(x_i^*)\Delta x_i. \tag{16.9}$$

The right-hand side of (16.9) is called a Riemann sum, of which there are many. For the above description, I have assumed that the sub-intervals are equal, which is not a necessary requirement.

If the number of sub-intervals is n, then

$$\Delta x = \frac{b-a}{n}$$

and the definite integral is defined as

$$\int_a^b f(x)\, dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x_i.$$

16.12 Worked Examples

16.12.1 Integrating a Function Containing Its Own Derivative

Evaluate

$$\int \frac{\sin x}{\cos x}\, dx.$$

Solution: Knowing that

$$\frac{d}{dx}[\cos x] = -\sin x$$

let $u = \cos x$, then

$$\frac{du}{dx} = -\sin x$$
$$du = -\sin x\, dx$$

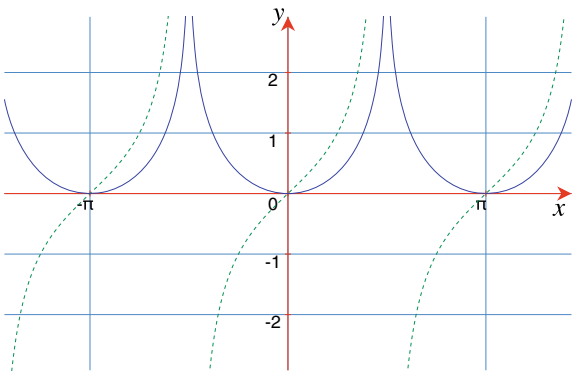

Fig. 16.39 The graphs of $y = \sin x / \cos x$ (dashed) and $y = \ln|\sec x|$

and

$$\begin{aligned}\int \frac{\sin x}{\cos x}\,dx &= \int \frac{1}{u}(-1)\,du \\ &= -\ln|u| + C \\ &= -\ln|\cos x| + C \\ &= \ln|\cos x|^{-1} + C \\ &= \ln|\sec x| + C.\end{aligned}$$

Figure 16.39 shows the graphs of $y = \sin x / \cos x$ and $y = \ln|\sec x|$.

16.12.2 Dividing an Integral into Several Integrals

Evaluate

$$\int \frac{2\sin x + \cos x + \sec x}{\cos x}\,dx.$$

Solution: Divide the numerator by $\cos x$:

$$\begin{aligned}\int \frac{2\sin x + \cos x + \sec x}{\cos x}\,dx &= 2\int \tan x\,dx + \int 1\,dx + \int \sec^2 x\,dx \\ &= 2\ln|\sec x| + x + \tan x + C.\end{aligned}$$

Figure 16.40 shows the graphs of $y = (2\sin x + \cos x + \sec x)/\cos x$ and $y = 2\ln|\sec x| + x + \tan x$.

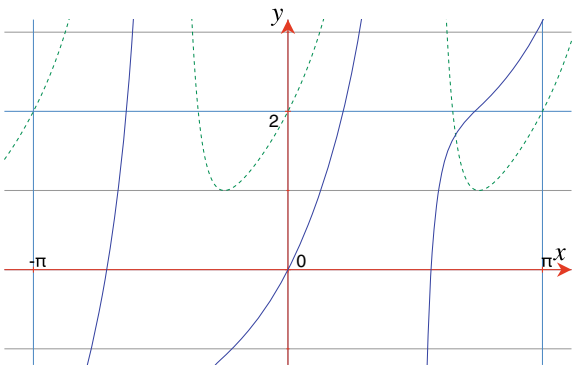

Fig. 16.40 The graphs of $y = (2\sin x + \cos x + \sec x)/\cos x$ (dashed) and $y = 2\ln|\sec x| + x + \tan x$

16.12.3 Integrating by Parts 1

Evaluate

$$\int x\cos x\ dx.$$

Solution: In this case, we try the following:

$$u = x \quad \text{and} \quad v' = \cos x$$

where

$$u' = 1 \quad \text{and} \quad v = \sin x + C_1.$$

Integrating by parts:

$$\begin{aligned}\int uv'\ dx &= uv - \int vu'\ dx\\ \int x\cos x\ dx &= x(\sin x + C_1) - \int (\sin x + C_1)(1)\ dx\\ &= x\sin x + C_1 x + \cos x - C_1 x + C\\ &= x\sin x + \cos x + C.\end{aligned}$$

Figure 16.41 shows the graphs of $y = x\cos x$ and $y = x\sin x + \cos x$.

16.12.4 Integrating by Parts 2

Evaluate

$$\int x^2\sin x\ dx.$$

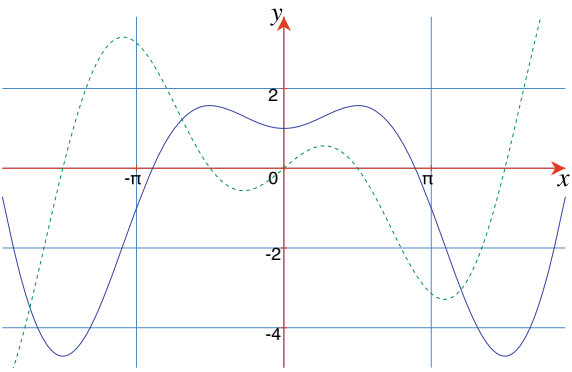

Fig. 16.41 The graphs of $y = x\cos x$ (dashed) and $y = x\sin x + \cos x$

Solution: In this case, we try the following:

$$u = x^2 \quad \text{and} \quad v' = \sin x$$

where

$$u' = 2x \quad \text{and} \quad v = -\cos x + C_1.$$

Integrating by parts:

$$\begin{aligned}
\int uv'\,dx &= uv - \int vu'\,dx \\
\int x^2 \sin x\,dx &= x^2(-\cos x + C_1) - 2\int(-\cos x + C_1)(x)\,dx \\
&= -x^2\cos x + C_1x^2 - 2C_1\int x\,dx + 2\int x\cos x\,dx \\
&= -x^2\cos x + C_1x^2 - 2C_1\left(\frac{x^2}{2} + C_2\right) + 2\int x\cos x\,dx \\
&= -x^2\cos x - C_3 + 2\int x\cos x\,dx.
\end{aligned}$$

At this point we come across $\int x\cos x\,dx$, which we have already solved:

$$\begin{aligned}
\int x^2\sin x\,dx &= -x^2\cos x - C_3 + 2(x\sin x + \cos x + C_4) \\
&= -x^2\cos x - C_3 + 2x\sin x + 2\cos x + C_5 \\
&= -x^2\cos x + 2x\sin x + 2\cos x + C
\end{aligned}$$

Figure 16.42 shows the graphs of $y = x^2\sin x$ and $y = -x^2\cos x + 2x\sin x + 2\cos x$.

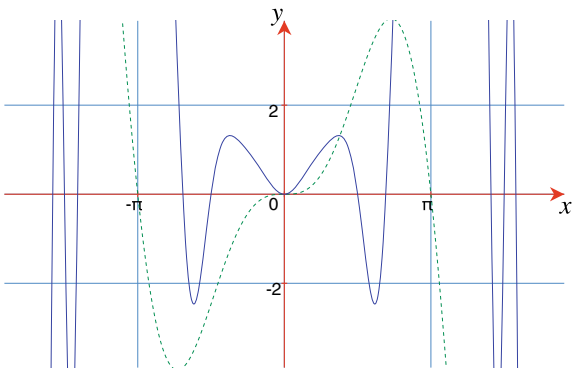

Fig. 16.42 The graphs of $y = x^2 \sin x$ (dashed) and $y = -x^2 \cos x + 2x \sin x + 2 \cos x$

16.12.5 Integrating by Substitution 1

Evaluate

$$\int 2e^{\cos(2x)} \sin x \cdot \cos x \, dx.$$

Solution: Integrating by substitution, let $u = \cos(2x)$, then

$$\frac{du}{dx} = -2\sin(2x) \quad \text{or} \quad dx = -\frac{du}{2\sin(2x)}.$$

Substituting a double-angle identity, u and du:

$$\begin{aligned}\int 2e^{\cos(2x)} \sin x \cdot \cos x \, dx &= -\int e^u \sin(2x) \frac{du}{2\sin(2x)} \\ &= -\tfrac{1}{2}\int e^u \, du \\ &= -\tfrac{1}{2}e^u + C \\ &= -\tfrac{1}{2}e^{\cos(2x)} + C.\end{aligned}$$

Figure 16.43 shows the graphs of $y = 2e^{\cos(2x)} \sin x \cdot \cos x$ and $y = -\frac{1}{2}e^{\cos(2x)}$.

16.12.6 Integrating by Substitution 2

Evaluate

$$\int \frac{\cos x}{(1 + \sin x)^3} \, dx.$$

Solution: Integrating by substitution, let $u = 1 + \sin x$, then

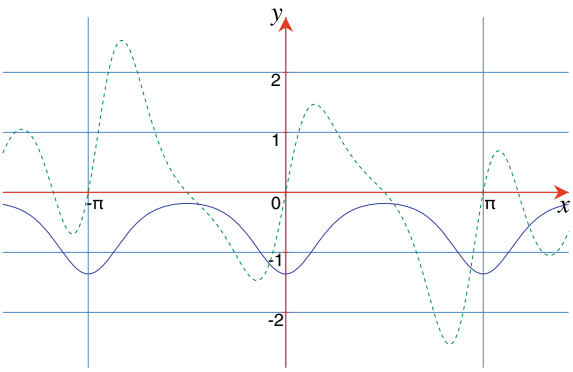

Fig. 16.43 The graphs of $y = 2e^{\cos 2x} \sin x \cos x$ (dashed) and $y = -\frac{1}{2}e^{\cos 2x}$

$$\frac{du}{dx} = \cos x \quad \text{or} \quad dx = \frac{du}{\cos x}.$$

$$\begin{aligned}
\int \frac{\cos x}{(1+\sin x)^3}\, dx &= \int \frac{\cos x}{u^3}\, \frac{du}{\cos x} \\
&= \int u^{-3}\, du \\
&= -\tfrac{1}{2}u^{-2} + C \\
&= -\tfrac{1}{2}(1+\sin x)^{-2} + C \\
&= -\tfrac{1}{2(1+\sin x)^2} + C.
\end{aligned}$$

Figure 16.44 shows the graphs of $y = \cos x/(1+\sin x)^3$ and $y = -1/2(1+\sin x)^2$.

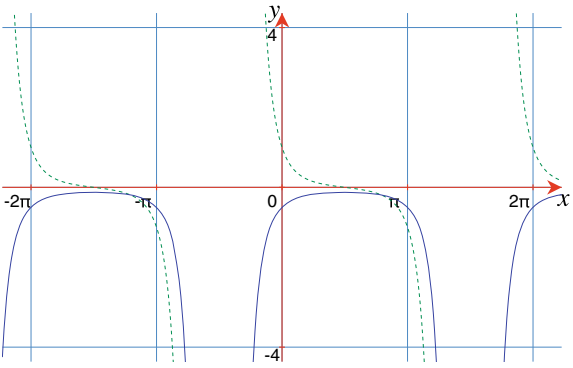

Fig. 16.44 The graphs of $y = \cos x/(1+\sin x)^3$ (dashed) and $y = -1/2(1+\sin x)^2$

16.12.7 Integrating by Substitution 3

Evaluate

$$\int \sin(2x)\, dx.$$

Solution: Integrating by substitution, let $u = 2x$, then

$$\frac{du}{dx} = 2 \quad \text{or} \quad dx = \frac{du}{2}.$$

$$\begin{aligned}\int \sin(2x)\, dx &= \tfrac{1}{2}\int \sin u\, du \\ &= -\tfrac{1}{2}\cos u + C \\ &= -\tfrac{1}{2}\cos(2x) + C\end{aligned}$$

Figure 16.45 shows the graphs of $y = \sin(2x)$ and $y = -\frac{1}{2}\cos(2x)$.

16.12.8 Integrating with Partial Fractions

Evaluate

$$\int \frac{6x^2 + 5x - 2}{x^3 + x^2 - 2x}\, dx.$$

Solution: Integrating by partial fractions:

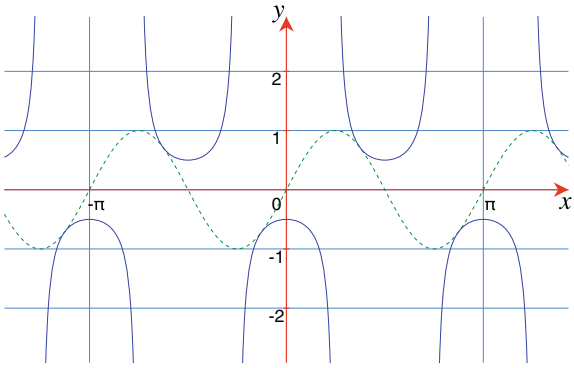

Fig. 16.45 The graphs of $y = \sin(2x)$ (dashed) and $y = -\frac{1}{2}\cos(2x)$

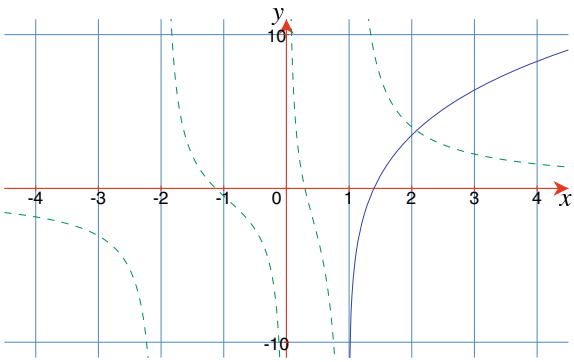

Fig. 16.46 The graphs of $y = (6x^2 + 5x - 2)/(x^3 + x^2 - 2x)$ (dashed) and $y = \ln x + 2\ln(x + 2) + 3\ln(x - 1)$

$$\begin{aligned}\frac{6x^2 + 5x - 2}{x^3 + x^2 - 2x} &= \frac{A}{x} + \frac{B}{x + 2} + \frac{C}{x - 1}\\ 6x^2 + 5x - 2 &= A(x + 2)(x - 1) + Bx(x - 1) + Cx(x + 2)\\ &= Ax^2 + Ax - 2A + Bx^2 - Bx + Cx^2 + 2Cx.\end{aligned}$$

Equating constants, terms in x and x^2:

$$-2 = -2A \tag{16.10}$$
$$5 = A - B + 2C \tag{16.11}$$
$$6 = A + B + C \tag{16.12}$$

Manipulating (16.10)–(16.12): $A = 1$, $B = 2$ and $C = 3$, therefore,

$$\begin{aligned}\int \frac{6x^2 + 5x - 2}{x^3 + x^2 - 2x}\,dx &= \int \frac{1}{x}\,dx + \int \frac{2}{x + 2}\,dx + \int \frac{3}{x - 1}\,dx\\ &= \ln x + 2\ln(x + 2) + 3\ln(x - 1) + C.\end{aligned}$$

Figure 16.46 shows the graphs of $y = (6x^2 + 5x - 2)/(x^3 + x^2 - 2x)$ and $y = \ln x + 2\ln(x + 2) + 3\ln(x - 1)$.

Appendix A
Limit of $(\sin\theta)/\theta$

This appendix proves that

$$\lim_{\theta \to 0} \frac{\sin\theta}{\theta} = 1, \quad \text{where } \theta \text{ is in radians.}$$

From high-school mathematics we know that $\sin\theta \approx \theta$, for small values of θ. For example:

$$\begin{aligned} \sin 0.1 &\approx 0.099833 \\ \sin 0.05 &\approx 0.04998 \\ \sin 0.01 &\approx 0.0099998 \end{aligned}$$

and

$$\begin{aligned} \frac{\sin 0.1}{0.1} &\approx 0.99833 \\ \frac{\sin 0.05}{0.05} &\approx 0.99958 \\ \frac{\sin 0.01}{0.01} &\approx 0.99998. \end{aligned}$$

Therefore, we can reason that in the limit, as $\theta \to 0$:

$$\lim_{\theta \to 0} \frac{\sin\theta}{\theta} = 1.$$

Figure A.1 shows a graph of $(\sin\theta)/\theta$, which confirms this result. However, this is an observation, rather than a proof. So, let's pursue a geometric line of reasoning.

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9

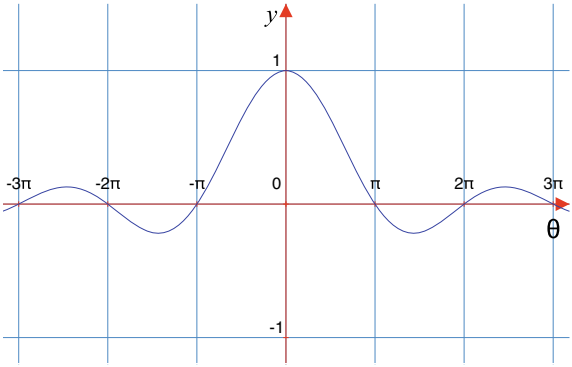

Fig. A.1 Graph of $(\sin\theta)/\theta$

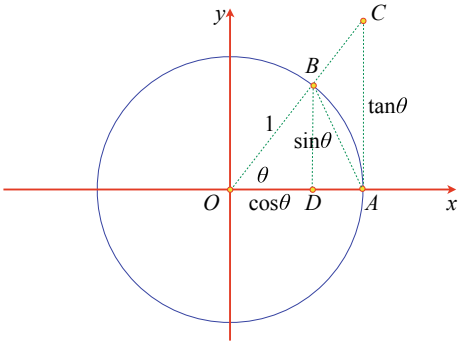

Fig. A.2 Unit radius circle with trigonometric ratios

From Fig. A.2 we see as the circle's radius is unity, $OA = OB = 1$, and $AC = \tan\theta$. As part of the strategy, we need to calculate the area of the triangle $\triangle OAB$, the sector OAB and the $\triangle OAC$:

$$\begin{aligned}
\text{Area of } \triangle OAB &= \triangle ODB + \triangle DAB \\
&= \tfrac{1}{2}\cos\theta \cdot \sin\theta + \tfrac{1}{2}(1-\cos\theta)\sin\theta \\
&= \tfrac{1}{2}\cos\theta \cdot \sin\theta + \tfrac{1}{2}\sin\theta - \tfrac{1}{2}\cos\theta \cdot \sin\theta \\
&= \frac{\sin\theta}{2}.
\end{aligned}$$

$$\text{Area of sector } OAB = \frac{\theta}{2\pi}\pi(1)^2 = \frac{\theta}{2}.$$

$$\text{Area of } \triangle OAC = \tfrac{1}{2}(1)\tan\theta = \frac{\tan\theta}{2}.$$

From the geometry of a circle, we know that

$$\begin{aligned}\frac{\sin\theta}{2} &< \frac{\theta}{2} < \frac{\tan\theta}{2}\\ \sin\theta &< \theta < \frac{\sin\theta}{\cos\theta}\\ 1 &< \frac{\theta}{\sin\theta} < \frac{1}{\cos\theta}\\ 1 &> \frac{\sin\theta}{\theta} > \cos\theta\end{aligned}$$

and as $\theta \to 0$, $\cos\theta \to 1$ and $\dfrac{\sin\theta}{\theta} \to 1$. This holds, even for negative values of θ, because

$$\frac{\sin(-\theta)}{-\theta} = \frac{-\sin\theta}{-\theta} = \frac{\sin\theta}{\theta}.$$

Therefore,

$$\lim_{\theta\to 0}\frac{\sin\theta}{\theta} = 1.$$

Appendix B
Integrating $\cos^n \theta$

We start with

$$\int \cos^n x \; dx = \int \cos x \cdot \cos^{n-1} x \; dx.$$

Let $u = \cos^{n-1} x$ and $v' = \cos x$, then

$$u' = -(n-1)\cos^{n-2} x \cdot \sin x$$

and

$$v = \sin x.$$

Integrating by parts:

$$\begin{aligned}
\int uv' \; dx &= uv - \int v \, u' \; dx + C \\
\int \cos^{n-1} x \cdot \cos x \; dx &= \cos^{n-1} x \cdot \sin x + \int \sin x \cdot (n-1)\cos^{n-2} x \cdot \sin x \; dx + C \\
&= \sin x \cdot \cos^{n-1} x + (n-1) \int \sin^2 x \cdot \cos^{n-2} x \; dx + C \\
&= \sin x \cdot \cos^{n-1} x + (n-1) \int (1 - \cos^2 x) \cdot \cos^{n-2} x \; dx + C \\
&= \sin x \cdot \cos^{n-1} x + (n-1) \int \cos^{n-2} \; dx - (n-1) \int \cos^n x \; dx + C \\
n \int \cos^n x \; dx &= \sin x \cdot \cos^{n-1} x + (n-1) \int \cos^{n-2} \; dx + C \\
\int \cos^n x \; dx &= \frac{\sin x \cdot \cos^{n-1} x}{n} + \frac{n-1}{n} \int \cos^{n-2} \; dx + C
\end{aligned}$$

where n is an integer, $\neq 0$.

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9

Similarly,

$$\int \sin^n x \, dx = -\frac{\cos x \cdot \sin^{n-1} x}{n} + \frac{n-1}{n} \int \sin^{n-2} \, dx + C.$$

For example,

$$\int \cos^3 x \, dx = \tfrac{1}{3} \sin x \cdot \cos^2 x + \tfrac{2}{3} \sin x + C.$$

Index

A
Absolute complement, 75
Aleph-zero, 30
Algebra, 35
Algebraic number, 21, 24
AND, 57, 69
Angle
  compound, 127
Anticlockwise, 135
Antiderivative, 298
Antisymmetric matrix, 236
Area, 182
  between two functions, 380
  circle, 371
  negative, 378
  of a shape, 136, 142
  parametric function, 383
  positive, 378
  under a graph, 370
  with the y-axis, 382
Argand, Jean-Robert, 165
Argument, 196
Associative law, 10, 190
Associativity, 62
  of $\vee$, 62
  of $\wedge$, 62
Atan2, 125
Axial systems
  left-handed, 137
  right-handed, 137
Axioms, 10

B
Barycentric coordinates, 141, 145
Binary
  addition, 17
  negative number, 18
  number, 13
  operation, 10
  subtraction, 18
Binomial coefficient, 84
Binomial expansion, 294
Boolean logic, 55
Boole, George, 55
Braces, 79
Bürgi, Joost, 42

C
Calculus, 289
Cantor, Georg, 30
Cardinality, 30, 71
Cartesian
  coordinates, 133, 134
  plane, 134
  vector, 173
Cauchy, Augustin-Louis, 289
Cayley, Arthur, 24, 227
Cayley numbers, 24
Chain rule, 338
Clockwise, 135
Closed interval, 46
Cofactor expansion, 248
Column vector, 166, 228, 231, 239
Combinations, 83
Combinatorics, 79
Commutative law, 10, 190
Commutativity, 61
  of $\vee$, 61
  of $\wedge$, 61
Complex
  conjugate, 22, 190
  exponentials, 200

J. Vince, *Foundation Mathematics for Computer Science*,
https://doi.org/10.1007/978-3-030-42078-9

norm, 198
number, 22
plane, 188
Complex number
inverse, 199
logarithm, 208
n^{th} roots, 207
norm, 198
to a complex power, 209
Composite number, 25
Compound angles, 127
Congruence, 102
Congruent pairs, 106
Conjugate
complex number, 190
Conjunction, 57, 59
Continuity, 289
Continuous functions, 352
Contradiction, 65
Contrapositive, 66
Coordinates
barycentric, 141, 145
Cartesian, 133
cylindrical, 140, 144
homogeneous, 142
local, 141
polar, 139, 143
spherical polar, 139, 144
Cosecant, 122
Cosine, 122
rule, 126
Cotangent, 122
Cross product, 176
Cubic
equation, 293
function, 134
Cylindrical coordinates, 140, 144

D

2D
polygons, 135
scaling transform, 268
vector, 166
3D
complex numbers, 177
coordinates, 137
rotation transform, 271
transforms, 270
vector, 169
Decimal
number, 12
system, 7
Definite integral, 374
Degree, 119
de Moivre, Abraham, 204
de Moivre's theorem, 204
de Morgan, Augustus, 63
de Morgan's Laws, 60, 64
Dependent events, 91
Dependent variable, 45
Derivative, 289, 297
graphical interpretation, 296
total, 341
Descartes, René, 36, 133
Determinant, 147, 233, 279
Diagonal matrix, 251
Differential, 297
Differentiating, 300
arccos function, 323
arccot function, 324
arccsc function, 324
arcsec function, 324
arcsin function, 323
arctan function, 323
cosh function, 326
cot function, 322
csc function, 320
exponential functions, 314
function of a function, 302
function products, 306
function quotients, 309
hyperbolic functions, 324
implicit functions, 311
logarithmic functions, 317
sec function, 321
sine function, 304
sinh function, 326
sums of functions, 300
tan function, 318
tanh function, 326
trigonometric functions, 318
Differentiation
partial, 333
Disjunction
exclusive, 57
inclusive, 57
Distance between two points, 137, 138
Distributive law, 11, 190
Distributivity, 63
of $\vee$ over $\wedge$, 63
of $\wedge$ over $\vee$, 63
Domain, 47, 123
Dot product, 174
Double-angle identities, 129
Double negation, 66

E
Element, 6
Empty set, 71
Equation
  explicit, 45
  implicit, 45
Equivalence, 58, 67
Euclid, 26, 28, 29
Euler
  diagram, 71
  rotations, 271
Euler, Leonhard, 44
Even function, 48
Events
  dependent, 91
  inclusive, 93
  independent, 91
  mutually exclusive, 92
Excluded middle, 65
Exclusive disjunction, 57, 59
Explicit equation, 45
Exportation, 66

F
Fermat, Pierre de, 109, 133
Fermat's little theorem, 109
Feynman, Richard, 201
Fraenkel, Abraham, 71
Function, 44, 294
  continuous, 352
  cubic, 134, 293
  domain, 47
  even, 49
  graph, 134
  linear, 134
  notation, 45
  odd, 48
  power, 49
  quadratic, 134, 291
  range, 47
  second derivative, 348
  trigonometric, 134
Function of a function
  differentiating, 302
Fundamental theorem of arithmetic, 26
Fundamental theorem of calculus, 375

G
Gauss, Carl, 28, 227
Geometric transforms, 225
Gibbs, Josiah, 24
Gödel, Kurt, 36
Goldbach, Christian, 27
Graves, John, 24

H
Half-angle identities, 130
Half-open interval, 47
Hamilton, Sir William Rowan, 23
Hexadecimal number, 13
Higher derivatives, 327
Homogeneous coordinates, 142
Hyperbolic functions, 216

I
IBAN check digits, 113
Idempotence, 60
  of $\vee$, 61
  of $\wedge$, 61
Identity matrix, 266
Iff, 58
Imaginary number, 21
Im function, 188
Implication, 57, 58, 66
Implicit equation, 45
Inclusive disjunction, 57, 59
Indefinite integral, 351
Independent events, 91
Independent variable, 45
Indeterminate form, 8
Indices, 41
  laws of, 42
Infinitesimals, 289
Infinity, 30
Infinity of primes, 28
Integer, 19
  number, 6
Integral
  definite, 374
Integrating
  arccos function, 323
  arccot function, 324
  arccsc function, 324
  arcsec function, 324
  arcsin function, 323
  arctan function, 323
  cot function, 322
  csc function, 320
  exponential function, 316
  logarithmic function, 317
  sec function, 321
  tan function, 318

Integration, 298
by parts, 361
by substitution, 366
completing the square, 357
difficult functions, 353
integrand contains a derivative, 359
partial fractions, 368
radicals, 356
techniques, 352
trigonometric identities, 354
Intersection, 74
Interval, 46
closed, 46
half-open, 46
open, 46
Inverse
complex number, 199
matrix, 243
trigonometric function, 123
Irrational number, 20
ISBN check digit, 117
ISBN parity check, 110

K
Kronecker, Leopold, 19

L
Laplace, Pierre–Simon, 248
Laplacian expansion, 248
Legendre, Adrien-Marie, 28
Leibniz, Gottfried, 44
Limits, 289, 295
Linear function, 134
Local coordinates, 141
Logarithms, 42
Logic, 55
Logical
connectives, 56
premise, 57
logic identities, 60

M
Magnitude, 196
Material equivalence, 57
Matrices, 225
Matrix
addition, 238
antisymmetric, 236
determinant, 233
diagonal, 251
dimension, 230
identity, 266
inverse, 243
multiplication, 229
notation, 230
null, 231
order, 230
orthogonal, 250
products, 239
rectangular, 242
scalar multiplication, 238
singular, 243
skew-symmetric, 236
square, 230, 241
subtraction, 238
symmetric, 234
trace, 232
transforms, 259
transpose, 233
unit, 231
Maxima, 330
Member, 6
Mersenne, Marin, 30
Mersenne prime, 30
Minima, 330
Mixed partial derivative, 336
Möbius, August, 141
Modular arithmetic, 101
Modulo a prime, 108
Modulus, 101, 196
Modus
ponens, 68
tollens, 66
Multiple-angle identities, 129
Multiplicative inverse, 106, 116
Multiplicity, 82
Multiset, 82

N
NAND, 69
Napier, John, 42
Natural number, 19
Negation, 57, 59
Negative number, 8
Non-associative algebra, 24
Non-commutative algebra, 23, 24
NOR, 69
Norm
complex number, 198
NOT, 57, 69
Notation, 3
Null matrix, 231
Number

algebraic, 21, 24
arithmetic, 9
binary, 13
Cayley, 24
complex, 22
composite, 25
hexadecimal, 14
imaginary, 21
integer, 6, 19
line, 8
Mersenne, 30
natural, 19
negative, 8
octal, 12
perfect, 29
positive, 8
prime, 25
rational, 6, 20
real, 6, 20
transcendental, 21, 24

O
Octal number, 12
Odd function, 48
One-to-one correspondence, 30
Open interval, 46
OR, 57, 69
Oriented axes, 134
Origin, 134
Orthogonal matrix, 250

P
Parentheses, 79
Partial derivative, 332
chain rule, 338
first, 333
mixed, 336
second, 348
visualising, 335
Pascal, Blaise, 133
Pascal's triangle, 294
Peirce, Benjamin, 227
Peirce, Charles, 227
Perimeter relationships, 130
Permutations, 79
Perspective projection, 281
Pitch, 273
Placeholder, 7
Polar
coordinates, 139, 143
notation, 202
Polynomial equation, 20
Position vector, 172
Power functions, 49
Power series, 200, 342
Prime number distribution, 27
Prime numbers, 25
Probability, 89
definition, 89
notation, 89
Proof by cases, 69
Proposition, 57
necessary, 57
sufficient, 57

Q
Quadratic
equation, 38
function, 134, 291
Quaternion, 23

R
Radian, 119, 200
Range, 47, 123
Rational
coefficients, 20
number, 6, 20
Real number, 6, 20
Rectangular matrix, 242
Reductio ad absurdum, 67, 68
Re function, 188
Relative complement, 74
Remainder, 101
Riemann, Bernhard, 385
Riemann sum, 385
Right-hand rule, 181
Roll, 273
Roots of 1, 206
Rotating about an axis, 274, 276
Row vector, 167, 231, 239
Russell, Bertrand, 35, 70

S
Scalar product, 174
Secant, 122
Second derivative, 348
Series
cosine, 203
power, 200, 342
sine, 200
Taylor's, 342
Set, 6, 71

absolute complement, 75
building, 72
cardinality, 71
empty, 71
intersection, 74
member, 71
ordered, 79
relative complement, 74
union, 73
universal, 72
unordered, 79
Simplification, 64
Sine, 122
differentiating, 303
rule, 126
Singular matrix, 243
Skew-symmetric matrix, 236
Skolem, Thoralf, 71
Spherical polar coordinates, 139, 144
Square matrix, 230, 241
Subset, 72
Sufficient proposition, 57
Superset, 72
Symbolic logic, 55
Symmetric matrix, 234

T
Tangent, 122
Tautology, 58
Taylor, Brook, 342
Taylor's series, 342
Theorem of
Pythagoras, 137, 138
Total derivative, 341
Trace, 232
Transcendental number, 21, 24
Transform
affine, 267
2D reflection, 263, 268
2D rotation about a point, 269
2D scaling, 261
2D shearing, 264
2D translation, 260
3D reflection, 276
3D scaling, 271
3D translation, 270
geometric, 225
scale, 225
translate, 225
Transpose matrix, 233
Trigonometric
function, 120, 134
identities, 125
ratios, 120
Trigonometric function
inverse, 123
Trigonometry, 119
Truth table, 56
Two's complement, 18

U
Union, 73
Unit
matrix, 231
normal vector, 181
vector, 173
Universal set, 72

V
Vector
addition, 171
Cartesian, 173
column, 166, 228, 231, 239
2D, 166
3D, 169
magnitude, 168
normalising, 173
position, 172
product, 174, 176
row, 166, 231, 239
scaling, 170
subtraction, 171
unit, 173
Venn diagram, 71
Venn, John, 71
Vertices, 135

W
Warren, John, 165
Weierstrass, Karl, 289
Wessel, Caspar, 165
Whitehead, Alfred North, 35

X
XNOR, 69
XOR, 57, 69
Xy-plane, 133

Y
Yaw, 273

Z
Zermelo, Ernst, 70
Zermelo-Fraenkel set theory, 71
Zero, 7
ZF, 71