Essential Mathematics

for Economic Analysis

Knut Sydsæter, Peter Hammond,
Arne Strøm & Andrés Carvajal

Pearson

Sixth Edition

ESSENTIAL MATHEMATICS FOR

ECONOMIC ANALYSIS

ESSENTIAL MATHEMATICS FOR

ECONOMIC ANALYSIS

SIXTH EDITION

Knut Sydsæter, Peter Hammond,

Arne Strøm and Andrés Carvajal

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • São Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

PEARSON EDUCATION LIMITED
KAO Two
KAO Park
Harlow CM17 9NA
United Kingdom
Tel: +44 (0)1279 623623
Web: www.pearson.com/uk

First published by Prentice Hall, Inc. 1995 (print)
Second edition published 2006 (print)
Third edition published 2008 (print)
Fourth edition published by Pearson Education Limited 2012 (print)
Fifth edition published 2016 (print and electronic)

ISBN: 978-1-292-35928-1 (print)
978-1-292-35929-8 (PDF)
978-1-292-35932-8 (ePub)

British Library Cataloguing-in-Publication Data
A catalogue record for the print edition is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Sydsæter, Knut, author. | Hammond, Peter J., 1945- author.
Title: Essential mathematics for economic analysis / Knut Sydsæter, Peter Hammond, Arne Strøm and Andrés Carvajal.
Description: Sixth edition. | Hoboken , NJ: Pearson, 2021. | Includes bibliographical references and index. | Summary: "The subject matter that modern economics students are expected to master makes significant mathematical demands. This is true even of the less technical "applied" literature that students will be expected to read for courses in fields such as public finance, industrial organization, and labour economics, amongst several others. Indeed, the most relevant literature typically presumes familiarity with several important mathematical tools, especially calculus for functions of one and several variables, as well as a basic understanding of multivariable optimization problems with or without constraints. Linear algebra is also used to some extent in economic theory, and a great deal more in econometrics"– Provided by publisher.
Identifiers: LCCN 2021006079 (print) | LCCN 2021006080 (ebook) | ISBN 9781292359281 (paperback) | ISBN 9781292359298 (pdf) | ISBN 9781292359328 (epub)
Subjects: LCSH: Economics, Mathematical.
Classification: LCC HB135 .S886 2021 (print) | LCC HB135 (ebook) | DDC 330.01/51–dc23
LC record available at https://lccn.loc.gov/2021006079
LC ebook record available at https://lccn.loc.gov/2021006080

10 9 8 7 6 5 4 3 2 1
25 24 23 22 21

Cover design by Michelle Morgan, At the Pop Ltd.

Front cover image © yewkeo/iStock//Getty Images Plus

Print edition typeset in 10/13pt TimesLTPro by SPi Global
Printed and bound by L.E.G.O. S.p.A., Italy

NOTE THAT ANY PAGE CROSS REFERENCES REFER TO THE PRINT EDITION

To Knut Sydsæter (1937–2012), an inspiring mathematics teacher, as well as wonderful friend and colleague, whose vision, hard work, high professional standards, and sense of humour were all essential in creating this book.

—Arne, Peter and Andrés

To Else, my loving and patient wife.

—Arne

To the memory of my parents Elsie (1916–2007) and Fred (1916–2008), my first teachers of Mathematics, basic Economics, and many more important things.

—Peter

To Yeye and Tata, my best ever students of "matemáquinas", who wanted this book to start with "Once upon a time …". E para a Pipoca, com amor infinito à infinito.

—Andrés

CONTENTS

Supporting resources

Visit **go.pearson.com/uk/he/resources** to find valuable online resources

For students

- A new Student's Manual provides more detailed solutions to the problems marked (SM) in the book

For instructors

- The fully updated Instructor's Manual provides instructors with a collection of problems that can be used for tutorials and exams

In MyLab Maths you will find:

- A study plan, which creates a personalised learning path through resources, based on diagnostic tests;
- A wealth of questions and activities, including algorithmic, graphing and multiple-choice questions and exercises;
- Pearson eText, an online version of your textbook which you can take with you anywhere and which allows you to add notes, highlights, bookmarks and easily search the text;
- If assigned by your lecturer, additional homework, quizzes and tests.

Lecturers additionally will find:

- A powerful gradebook, which tracks student performance and can highlight areas where the whole class needs more support;
- Additional questions and exercises, which can be assigned to students as homework, quizzes and tests, freeing up the time in class to take learning further;
- Lecturer resources, which support the delivery of the module;
- Communication tools, such as discussion boards, to drive student engagement.

PREFACE

Once upon a time there was a sensible straight line who was hopelessly in love with a dot. 'You're the beginning and the end, the hub, the core and the quintessence,' he told her tenderly, but the frivolous dot wasn't a bit interested, for she only had eyes for a wild and unkempt squiggle who never seemed to have anything on his mind at all. All of the line's romantic dreams were in vain, until he discovered . . . angles! Now, with newfound self-expression, he can be anything he wants to be—a square, a triangle, a parallelogram . . . And that's just the beginning!
—Norton Juster (*The Dot and the Line: A Romance in Lower Mathematics* 1963)

I came to the position that mathematical analysis is not one of many ways of doing economic theory: It is the only way. Economic theory is mathematical analysis. Everything else is just pictures and talk.
—R. E. Lucas, Jr. (2001)

Purpose

The subject matter that modern economics students are expected to master makes significant mathematical demands. This is true even of the less technical "applied" literature that students will be expected to read for courses in fields such as public finance, industrial organization, and labour economics, amongst several others. Indeed, the most relevant literature typically presumes familiarity with several important mathematical tools, especially calculus for functions of one and several variables, as well as a basic understanding of multivariable optimization problems with or without constraints. Linear algebra is also used to some extent in economic theory, and a great deal more in econometrics.

The purpose of *Essential Mathematics for Economic Analysis*, therefore, is to help economics students acquire enough mathematical skill to access the literature that is most relevant to their undergraduate study. This should include what some students will need to conduct successfully an undergraduate research project or honours thesis.

As the title suggests, this is a book on *mathematics*, whose material is arranged to allow progressive learning of mathematical topics. That said, we do frequently emphasize economic applications, many of which are listed on the inside front cover. These not only help motivate particular mathematical topics; we also want to help prospective economists acquire mutually reinforcing intuition in both mathematics and economics. Indeed, as the list of examples on the inside front cover suggests, a considerable number of economic concepts and ideas receive some attention.

We emphasize, however, that this is not a book about economics or even about mathematical economics. Students should learn economic theory systematically from other courses, which use other textbooks. We will have succeeded if they can concentrate on the economics in these courses, having already thoroughly mastered the relevant mathematical tools this book presents.

Special Features and Accompanying Material

Virtually all sections of the book conclude with exercises, often quite numerous. There are also many review exercises at the end of each chapter. Solutions to almost all these exercises are provided at the end of the book, sometimes with several steps of the answer laid out.

There are two main sources of supplementary material. The first, for both students and their instructors, is via MyLab. Students who have arranged access to this web site for our book will be able to generate a practically unlimited number of additional problems which test how well some of the key ideas presented in the text have been understood. More explanation of this system is offered after this preface. The same web page also has a "student resources" tab with access to a *Student's Manual* with more extensive answers (or, in the case of a few of the most theoretical or difficult problems in the book, the only answers) to problems marked with the special symbol SM.

The second source, for instructors who adopt the book for their course, is an *Instructor's Manual* that may be downloaded from the publisher's Instructor Resource Centre.

In addition, for courses with special needs, there is a brief online appendix on trigonometric functions and complex numbers. This is also available via MyLab.

Prerequisites

Experience suggests that it is quite difficult to start a book like this at a level that is really too elementary.[1] These days, in many parts of the world, students who enter college or university and specialize in economics have an enormous range of mathematical backgrounds and aptitudes. These range from, at the low end, a rather shaky command of elementary algebra, up to real facility in the calculus of functions of one variable. Furthermore, for many economics students, it may be some years since their last formal mathematics course. Accordingly, as mathematics becomes increasingly essential for specialist studies in economics, we feel obliged to provide as much quite elementary material as is reasonably possible. Our aim here is to give those with weaker mathematical backgrounds the chance to get started, and even to acquire a little confidence with some easy problems they can really solve on their own.

[1] In a recent test for 120 first-year students intending to take an elementary economics course, there were 35 different answers to the problem of expanding $(a + 2b)^2$.

To help instructors judge how much of the elementary material students really know before starting a course, the *Instructor's Manual* provides some diagnostic test material. Although each instructor will obviously want to adjust the starting point and pace of a course to match the students' abilities, it is perhaps even more important that each individual student appreciates his or her own strengths and weaknesses, and receives some help and guidance in overcoming any of the latter. This makes it quite likely that weaker students will benefit significantly from the opportunity to work through the early more elementary chapters, even if they may not be part of the course itself.

As for our economic discussions, students should find it easier to understand them if they already have a certain very rudimentary background in economics. Nevertheless, the text has often been used to teach mathematics for economics to students who are studying elementary economics at the same time. Nor do we see any reason why this material cannot be mastered by students interested in economics before they have begun studying the subject in a formal university course.

Topics Covered

After the introductory material in Chapters 1 to 3, a fairly leisurely treatment of standard single variable differential calculus is contained in Chapters 4 to 7. This is followed by Chapter 8 on concave and convex functions, by Chapter 9 on optimization, Chapter 10 on integration, and then by some basic financial models as well as difference and differential equations in Chapter 11. This may be as far as some elementary courses will go. Students who already have a thorough grounding in single variable calculus, however, may only need to go fairly quickly over some special topics in these chapters such as elasticity and conditions for global optimization that are often not thoroughly covered in standard calculus courses.

We have already suggested the importance for budding economists of the algebra of matrices and determinants (Chapters 12 and 13), of multivariable calculus (Chapters 14–16), and of optimization theory with and without constraints (Chapters 17–20). These last nine chapters in some sense represent the heart of the book, on which students with a thorough grounding in single variable calculus can probably afford to concentrate.

Satisfying Diverse Requirements

The less ambitious student can concentrate on learning the key concepts and techniques of each chapter. Often, these appear boxed and/or in colour, in order to emphasize their importance. Problems are essential to the learning process, and the easier ones should definitely be attempted. These basics should provide enough mathematical background for the student to be able to understand much of the economic theory that is embodied in applied work at the advanced undergraduate level.

Students who are more ambitious, or who are led on by more demanding teachers, can try the more difficult problems. They can also study the more technical material which is intended to encourage students to ask why a result is true, or why a problem should be tackled in a particular way. If more readers gain at least a little additional mathematical insight from working through these more challenging parts of our book, so much the better.

The most able students, especially those intending to undertake postgraduate study in economics or some related subject, will benefit from a fuller explanation of some topics than we have been able to provide here. On a few occasions, therefore, we take the liberty of referring to our more advanced companion volume, *Further Mathematics for Economic Analysis* (usually abbreviated to FMEA). This is written jointly with our colleague Atle Seierstad in Oslo. In particular, FMEA offers a proper treatment of topics like systems of difference and differential equations, as well as dynamic optimization, that we think go rather beyond what is really “essential” for all economics students.

Changes in the Fourth Edition

We have been gratified by the number of students and their instructors from many parts of the world who appear to have found the first three editions useful.[2] We have accordingly been encouraged to revise the text thoroughly once again. There are numerous minor changes and improvements, including the following in particular:

1. The main new feature is MyMathLab Global,[3] explained on the page after this preface, as well as on the back cover.
2. New exercises have been added for each chapter.
3. Some of the figures have been improved.

Changes in the Fifth Edition

The most significant change in this edition is that, tragically, we have lost the main author and instigator of this project. Our good friend and colleague Knut Sydsæter died suddenly on 29th September 2012, while on holiday in Spain with his wife Malinka Staneva, a few days before his 75th birthday. An obituary written by Jens Stoltenberg, at that time the Prime Minister of Norway, includes this tribute to Knut’s skills as one of his teachers:

> With a small sheet of paper as his manuscript he introduced me and generations of other economics students to mathematics as a tool in the subject of economics. With professional weight, commitment, and humour, he was both a demanding and an inspiring lecturer. He opened the door into the world of mathematics. He showed that mathematics is a language that makes it possible to explain complicated relationships in a simple manner.

At a web page that hosts a copy of this obituary one can also find other tributes to Knut, including some recollections of how previous editions of this book came to be written.[4]

Despite losing Knut as its main author, it was clear that this book needed to be kept alive, following desires that Knut himself had often expressed while he was still with us.

---

[2] Different English versions of this book have been translated into Albanian, French, German, Hungarian, Italian, Portuguese, Spanish, and Turkish.

[3] Superseded by MyLab for this sixth edition.

[4] See https://web.stanford.edu/~hammond/sydsaeter.html

Fortunately, it had already been agreed that the team of co-authors should be joined by Andrés Carvajal, a former colleague of Peter's at Warwick who, at the time of preparing the Fifth Edition, had just joined the University of California at Davis. Andrés had already produced a new Spanish version of the previous edition of this book; he has now become a co-author of this latest English version. It is largely on his initiative that we have taken the important step of extensively rearranging the material in the first three chapters in a more logical order, with set theory now coming first.

The other main change is one that we hope is invisible to the reader. Previous editions had been produced using the "plain TEX" typesetting system that dates back to the 1980s, along with some ingenious macros that Arne had devised in collaboration with Arve Michaelsen of the Norwegian typesetting firm Matematisk Sats. For technical reasons we decided that the new edition had to be produced using the enrichment of plain TEX called LATEX that has by now become the accepted international standard for typesetting mathematical material. We have therefore attempted to adapt and extend some standard LATEX packages in order to preserve as many good features as possible of our previous editions.

Changes in the Sixth Edition

For this sixth edition, the surviving authors decided to rearrange the chapters considerably. Recent previous editions included a chapter on linear programming, which was deferred until after the two chapters on matrix algebra. Yet the key idea of complementary slackness had arisen previously in an earlier chapter on nonlinear programming. So we have moved matrix algebra much further forward, so that it precedes multivariate calculus. This allows new tools to be used in our treatment of multivariate calculus, and subsequently in the last four chapters that are now devoted exclusively to optimization.

Not only have the existing chapters been rearranged, however. We have increased their number from 17 to 20. This is partly because the chapter on constrained optimization has been split into two. The first part dealing with equality constraints now comes in Chapter 18, before Chapter 19 on linear programming, including its discussion of complementary slackness. The last part of the earlier chapter on inequality constraints is now the separate Chapter 20.

The other two extra chapters are new. Chapter 8 considers concave and convex functions of one variable, including results on supergradients of concave functions and subgradients of convex functions that play a key role in the theory of optimization. Later chapters extend some of these results to functions of 2 and then n variables. There is also a brief chapter (16) on multiple integrals.

Finally, we mention significant additions to Chapter 13 that consider eigenvalues and quadratic forms. These additions allow a more extensive treatment, based on the Hessian matrix, of second-order conditions for, in Chapter 15, a function of several variables to be concave, and in Chapter 17, for a critical point to be a maximum or minimum. As a result, we can provide a somewhat better discussion in Chapter 20 of how, for the case of concave programming problems, the Karush–Kuhn–Tucker conditions provide sufficient conditions for an optimal point.

Other Acknowledgements

Over the years we have received help from so many colleagues, lecturers at other institutions, and students, that it is impractical to mention them all.

Andrés Carvajal is indebted to: Yiqian Zhao and Xinhui Yang, for all their great work in the revision of the material for this edition; Professor Janine Wilson for encouraging him in the idea that the more economic applications the book contains, the better is the mathematical explanation; Professor Jim Wiseman, for his feedback on the previous edition and for sharing his views on how it could be improved; and to the following UC Davis students who patiently went over different chapters, fishing for mistakes and making sure that all was well: Xinghe Bai, Veronica Contreras, Nathan Gee, Anjali Khalasi, Yannan Li, Daniel Scates, Kelly Stangl, and Yiping Su.

As in previous editions of this book, we are very happy to acknowledge with gratitude the encouragement and assistance of our contacts at Pearson. For this sixth edition, these include Catherine Yates (Product Manager) and Melanie Carter (Senior Content Producer). We were also glad to be able to work successfully with Vivek Khandelwal of SPi Global, who was in charge of the typesetting, and Lou Attwood of SpacedEns Editorial Services, who assisted us with proof-reading. All were very helpful and attentive in answering our frequent e-mails in a friendly and encouraging way, while making sure that this new edition really is getting into print in a timely manner.

On the more academic side, very special thanks go to Prof. Dr Fred Böker at the University of Göttingen. He is not only responsible for translating several previous editions of this book into German, but has also shown exceptional diligence in paying close attention to the mathematical details of what he was translating. We appreciate the resulting large number of valuable suggestions for improvements and corrections that he has continued to provide, sometimes at the instigation of Dr Egle Tafenau, who was also using the German version of our textbook in her teaching.

We are also grateful to Kenneth Judd of the Hoover Institution at Stanford for taking the trouble to persuade us that we should follow what has become the standard practice of attaching the name of William Karush, along with those of Harold Kuhn and Albert Tucker, to the key "KKT conditions" presented in Chapter 20 for solving a nonlinear programming problem with inequality constraints.

Thanks too, to Dr Mauro Bambi at Durham University for creating and curating question content for MyLab Maths, and to Professor Carsten Berthram Haahr Andersen at Aarhus University, Denmark for his feedback on the MyLab.

To these and all the many unnamed persons and institutions who have helped us make this text possible, including some whose anonymous comments on earlier editions were forwarded to us by the publisher, we would like to express our deep appreciation and gratitude. We hope that all those who have assisted us may find the resulting product of benefit to their students. This, we can surely agree, is all that really matters in the end.

Andrés Carvajal, Peter Hammond, and Arne Strøm

Davis, Coventry, and Oslo, January 2021

I

PRELIMINARIES

1 ESSENTIALS OF LOGIC AND SET THEORY

It is clear that economics, if it is to be a science at all, must be a mathematical science.
—William Stanley Jevons[1]

Arguments in mathematics require tight logical reasoning, and arguments in modern economic analysis are no exception to this rule. It is useful for us, then, to present some basic concepts from logic, as well as a brief section on mathematical proofs.

We precede this with a short introduction to set theory. This is useful not just for its importance in mathematics, but also because of a key role that sets play in economics: in most economic models, it is assumed that economic agents pursue some specific goal like profit, and make an optimal choice from a specified feasible set of alternatives.

The chapter winds up with a discussion of mathematical induction. Occasionally, this method is used directly in economic arguments; more often, it is needed to understand mathematical results which economists use.

1.1 Essentials of Set Theory

In daily life, we constantly group together objects of the same kind. For instance, the faculty of a university signifies all the members of its academic staff. A garden refers to all the plants that are growing in it. An economist may talk about all Scottish firms with over 300 employees, or all taxpayers in Germany who earned between €50 000 and €100 000 in 2019. Or suppose a student who is planning what combination of laptop and smartphone to buy for use in college. The student may consider all combinations whose total price does not exceed what she can afford. In all these cases, we have a collection of objects that we may want to view as a whole. In mathematics, such a collection is called a *set*, and the objects that belong to the set are called its *elements*, or its *members*.

[1] *The Theory of Political Economy* (1871)

The simplest way to specify a set is to list its members, in any order, between the opening brace { and the closing brace }. An example is the set whose members are the first three letters in the English alphabet, $S = \{a, b, c\}$. Or it might be a set consisting of three members represented by the letters a, b, and c. For example, if $a = 0$, $b = 1$, and $c = 2$, then $S = \{0, 1, 2\}$. Also, $S = \{a, b, c\}$ denotes the set of roots of the cubic equation $(x - a)(x - b)(x - c) = 0$ in the unknown x, where a, b, and c are any three real numbers. Verbally, the braces are read as "the set consisting of".

Since a set is fully specified by listing all its members, two sets A and B are considered *equal* if they contain exactly the same elements: each element of A is an element of B; conversely, each element of B is an element of A. In this case, we write $A = B$. Consequently, $\{1, 2, 3\} = \{3, 2, 1\}$, because the order in which the elements are listed has no significance; and $\{1, 1, 2, 3\} = \{1, 2, 3\}$, because a set is not changed if some elements are listed more than once.

The symbol "$\varnothing$" denotes the set that has no elements. It is called the *empty set*. Note that it is *the*, and not *an*, empty set. This is so, following the principle that a set is completely defined by listing all its members: there can only be one set that contains no elements. The empty set is the same, whether it is being studied by a child in elementary school who thinks about cows that can jump over the moon, or by a physicist at CERN who thinks about subatomic particles that move faster than the speed of light—or, indeed, by an economics student reading this book!

Specifying a Property

Not every set can be defined by listing all its members, however. For one thing, some sets are infinite—that is, they contain infinitely many members. Such infinite sets are rather common in economics. Take, for instance, the budget set that arises in consumer theory. Suppose there are two goods with quantities denoted by x and y. Suppose these two goods can be bought at prices per unit that equal p and q, respectively. A consumption bundle is a pair of quantities of the two goods, (x, y). Its value at prices p and q is $px + qy$. Suppose that a consumer has an amount m to spend on the two goods. Then the *budget constraint* is $px + qy \le m$, assuming that the consumer is free to underspend. If one also accepts that the quantity consumed of each good must be nonnegative, then the *budget set*, which will be denoted by B, consists of all those consumption bundles (x, y) satisfying the three inequalities $px + qy \le m$, $x \ge 0$, and $y \ge 0$. This set is illustrated in Fig. 4.4.12. Standard notation for it is

$$B = \{(x, y) : px + qy \le m,\ x \ge 0,\ y \ge 0\} \tag{1.1.1}$$

The two braces { and } are still used to denote "the set consisting of". However, instead of listing all the members, which is impossible for the infinite set of points in the triangular budget set B, it is specified in two parts. First, before the colon, (x, y) is used to denote the typical member of B, here a consumption bundle that is specified by listing the respective quantities of the two goods. The colon is read as "such that".[2] Second, after the colon, the three properties that these typical members must satisfy are all listed.

[2] Alternative notation for "such that" is |.

This completes the specification of B. Indeed, Eq. (1.1.1) is an example of the general specification:

$$S = \{\text{typical member : defining properties}\}$$

Note that it is not just infinite sets that can be specified by properties like this—finite sets can too. Indeed, some finite sets almost *have* to be specified in this way, such as the set of all human beings currently alive.

Set Membership

As we stated earlier, sets contain members or elements. Some convenient standard notation is used to express the relation between a set and its members. First,

$$x \in S$$

indicates that x *is an element of* S. Note the special "belongs to" symbol $\in$ (which is a variant of the Greek letter ε, or "epsilon").

To express the fact that x is *not* a member of S, we write $x \notin S$. For example, $d \notin \{a, b, c\}$ says that d is not an element of the set $\{a, b, c\}$.

To see how set membership notation can be applied, consider again the example of a first-year college student who *must* buy both a laptop and a smartphone. Suppose that there are two types of each device, "cheap" and "expensive". Suppose too that the student cannot afford to combine the expensive smartphone with the expensive laptop. Then the set of three combinations that the student can afford is {cheap laptop and cheap smartphone, expensive laptop and cheap smartphone, cheap laptop and expensive smartphone}. Thus, the student is restricted to choosing one of the three combinations in this set. If we denote the choice by s and the affordable set by B, we can say that the student's choice is constrained by the requirement that $s \in B$. If we denote by t the unaffordable combination of an expensive laptop and an expensive smartphone, we can express this unaffordability by writing $t \notin B$.

Let A and B be any two sets. Set A is a *subset* of B if it is true that every member of A is also a member of B. When that is the case, we write $A \subseteq B$. In particular, $A \subseteq A$ and $\varnothing \subseteq A$. Recall that two sets are *equal* if they contain the same elements. From the definitions, we see that $A = B$ when, and only when, both $A \subseteq B$ and $B \subseteq A$.

To continue the previous example, suppose that the student can make do with a cheap smartphone, so she chooses not to buy an expensive one. Having made this choice, she only needs to decide which laptop to buy in addition to the cheap smartphone. Let A denote the set {cheap laptop and cheap smartphone, expensive laptop and cheap smartphone} of options the student has not ruled out. Then we have $A \subseteq B$.

Set Operations

Sets can be combined in many different ways. Especially important are three operations: the *union*, *intersection*, and the *difference* of any two sets A and B, as shown in Table 1.1.1.

Table 1.1.1 Elementary set operations

Notation	Name	The set that consists of:
$A \cup B$	*A union B*	all elements belonging to at least one of the sets A and B
$A \cap B$	*A intersection B*	all elements belonging to both A and B
$A \setminus B$	*A minus B*	all elements belonging to set A, but not to B

In symbols:

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

$$A \setminus B = \{x : x \in A \text{ and } x \notin B\}$$

It is important to notice that the word "or" in mathematics is *inclusive*, in the sense that the statement "$x \in A$ or $x \in B$" allows for the possibility that $x \in A$ and $x \in B$ are *both* true.

EXAMPLE 1.1.1 Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{3, 6\}$. Find $A \cup B$, $A \cap B$, $A \setminus B$, and $B \setminus A$.[3]

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

As an economic example, considering everybody who worked in California during the year 2019. Let A denote the set of all those workers who have an income of at least \$35 000 for the year; let B denote the set of all who have a net worth of at least \$200 000. Then $A \cup B$ would be those workers who earned at least \$35 000 or who had a net worth of at least \$200 000, whereas $A \cap B$ are those workers who earned at least \$35 000 and who also had a net worth of at least \$200 000. Finally, $A \setminus B$ would be those who earned at least \$35 000 but whose net worth was less than \$200 000.

If two sets A and B have no elements in common, they are said to be *disjoint*. Thus, the sets A and B are disjoint if $A \cap B = \varnothing$.

A collection of sets is often referred to as a *family* of sets. When considering a certain family of sets, it is often natural to think of each set in the family as a subset of one particular fixed set $\mathcal{U}$, hereafter called the *universal set*. In the previous example, the set of all residents of California in 2019 would be an obvious choice for a universal set.

If A is a subset of the universal set $\mathcal{U}$, then according to the definition of difference, $\mathcal{U} \setminus A$ is the set of elements of $\mathcal{U}$ that are not in A. This set is called the *complement* of A in $\mathcal{U}$ and is denoted by A^c.[4] When finding the complement of a set, it is *very* important to be clear about which universal set is being used.

EXAMPLE 1.1.2 Let the universal set $\mathcal{U}$ be the set of all students at a particular university. Among these, let F denote the set of female students, M the set of all mathematics students, C the set of students in the university choir, B the set of all biology students, and T the set of all tennis

[3] Here and throughout the book, we often write the examples in the form of exercises. We strongly suggest that you first attempt to solve the problem, while covering the solution, and then gradually reveal the proposed solution to see if you are right.

[4] Other ways of denoting the complement of A include $\complement A$ and $\tilde{A}$.

players. Describe the members of the following sets: $\mathcal{U} \setminus M$, $M \cup C$, $F \cap T$, $M \setminus (B \cap T)$, and $(M \setminus B) \cup (M \setminus T)$.

Solution: $\mathcal{U} \setminus M$ consists of those students who are not studying mathematics, $M \cup C$ of those students who study mathematics and/or are in the choir. The set $F \cap T$ consists of those female students who play tennis. The set $M \setminus (B \cap T)$ has those mathematics students who do not both study biology and play tennis. Finally, the last set $(M \setminus B) \cup (M \setminus T)$ has those students who either are mathematics students not studying biology or mathematics students who do not play tennis. Can you see that the last two sets must be equal?[5]

Venn Diagrams

When considering how different sets may be related, it is often both instructive and extremely helpful to represent each set by a region in a plane. Diagrams constructed in this manner are called *Venn diagrams*.[6]

For pairs of sets, the definitions discussed in the previous section can be illustrated as in Fig. 1.1.1. By using the definitions directly, or by illustrating sets with Venn diagrams, one can derive formulas that are universally valid regardless of which sets are being considered. For example, the formula $A \cap B = B \cap A$ follows immediately from the definition of the intersection between two sets.

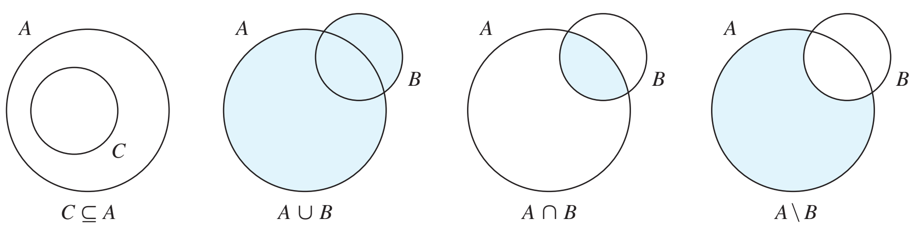

Figure 1.1.1 Four Venn diagrams

When dealing with three general sets A, B, and C, it is important to draw the Venn diagram so that all possible relations between an element and each of the three sets are represented. In other words, as in Fig. 1.1.3, the following eight different regions should all be nonempty:[7]

1. $(A \cap B) \setminus C$
2. $(B \cap C) \setminus A$
3. $(C \cap A) \setminus B$
4. $A \setminus (B \cup C)$
5. $B \setminus (C \cup A)$
6. $C \setminus (A \cup B)$
7. $(A \cap B) \cap C$
8. $((A \cup B) \cup C)^c$

[5] For arbitrary sets M, B, and T, it is true that $(M \setminus B) \cup (M \setminus T) = M \setminus (B \cap T)$. It should become easier to verify this equality after you have studied the following discussion of Venn diagrams.

[6] Named after the English mathematician John Venn (1834–1923), who was the first to use them extensively.

[7] That is, all should be nonempty unless something more is known about the relation between the three sets. For example, one might have specified that the sets must be disjoint, meaning that $A \cap B \cap C = \varnothing$. In this case region (7) in Fig. 1.1.3 disappears.

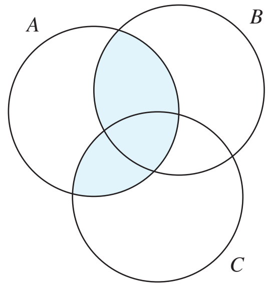

Figure 1.1.2 Venn diagram for $A \cap (B \cup C)$

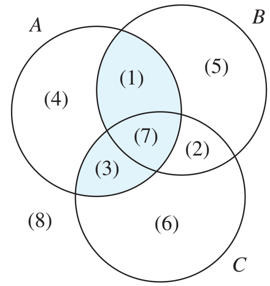

Figure 1.1.3 Venn diagram for three sets

Venn diagrams are particularly useful when limited to no more than three sets. For instance, consider the following possible relationship between the three sets A, B, and C:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{1.1.2}$$

Using only the definitions in Table 1.1.1, it is somewhat difficult to verify that Eq. (1.1.2) holds for all sets A, B, C. Using a Venn diagram, however, it is easily seen that the two sets on the left- and right-hand sides of (1.1.2) are both represented by the region made up of the three regions that are shaded in both Fig. 1.1.2 and Fig. 1.1.3. This confirms Eq. (1.1.2). Similar reasoning allows one to prove that

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{1.1.3}$$

Using either the definition of intersection and union or appropriate Venn diagrams, one can see that $A \cup (B \cup C) = (A \cup B) \cup C$ and that $A \cap (B \cap C) = (A \cap B) \cap C$. Consequently, in such cases it does not matter where the parentheses are placed, so they can be dropped and the expressions written as $A \cup B \cup C$ and $A \cap B \cap C$. That said, note that the parentheses cannot generally be removed in the two expressions on the left-hand sides of Eqs (1.1.2) and (1.1.3). This is because $A \cap (B \cup C)$ is generally not equal to $(A \cap B) \cup C$, and $A \cup (B \cap C)$ is generally not equal to $(A \cup B) \cap C$.[8]

Notice, however, that this way of representing sets in the plane becomes unmanageable if four or more sets are involved. This is because a Venn diagram with, for example, four sets would have to contain $2^4 = 16$ regions.[9]

Georg Cantor

The founder of set theory is Georg Cantor (1845–1918), who was born in Saint Petersburg but moved to Germany at the age of eleven. He is regarded as one of history's great mathematicians. This is not because of his contributions to the development of the useful, but relatively trivial, aspects of set theory outlined above. Rather, Cantor is remembered for his profound study of infinite sets. Below we try to give just a hint of his theory's implications.

[8] For practice, demonstrate this fact by considering the case where $A = \{1, 2, 3\}$, $B = \{2, 3\}$, and $C = \{4, 5\}$, or by using a Venn diagram.

[9] One can show that a Venn diagram with n sets would have to contain 2^n regions.

A collection of individuals are gathering in a room that has a certain number of chairs. How can we find out if there are exactly as many individuals as chairs? One method would be to count the chairs and count the individuals, and then see if they total the same number. Alternatively, we could ask all the individuals to sit down. If they all have a seat to themselves and there are no chairs unoccupied, then there are exactly as many individuals as chairs. In that case each chair corresponds to an individual and each individual corresponds to a chair—i.e., there is a "one-to-one correspondence" between individuals and chairs.

Generally mathematicians say that two sets of elements have the same *cardinality*, if there is a one-to-one correspondence between the sets. This definition is also valid for sets with an infinite number of elements. Cantor struggled for three years to prove a surprising implication of this definition—that there are as many points in a square as there are points on one of its edges of the square, in the sense that the two sets have the same cardinality.[10]

EXERCISES FOR SECTION 1.1

1. Let $A = \{2, 3, 4\}$, $B = \{2, 5, 6\}$, $C = \{5, 6, 2\}$, and $D = \{6\}$.

(a) Determine which of the following six statements are true: $4 \in C$; $5 \in C$; $A \subseteq B$; $D \subseteq C$; $B = C$; and $A = B$.

(b) List all members of each of the following eight sets: $A \cap B$; $A \cup B$; $A \setminus B$; $B \setminus A$; $(A \cup B) \setminus (A \cap B)$; $A \cup B \cup C \cup D$; $A \cap B \cap C$; and $A \cap B \cap C \cap D$.

2. Let F, M, C, B, and T be the sets in Example 1.1.2.

(a) Describe the following sets: $F \cap B \cap C$, $M \cap F$, and $((M \cap B) \setminus C) \setminus T$.

(b) Write the following statements in set terminology:

(i) All biology students are mathematics students.

(ii) There are female biology students in the university choir.

(iii) No tennis player studies biology.

(iv) Those female students who neither play tennis nor belong to the university choir all study biology.

3. A survey revealed that 50 people liked coffee and 40 liked tea. Both these figures include 35 who liked both coffee and tea. Finally, ten did not like either coffee or tea. How many people in all responded to the survey?

4. Make a complete list of all the different subsets of the set $\{a, b, c\}$. How many are there if the empty set and the set itself are included? Do the same for the set $\{a, b, c, d\}$.

5. Determine which of the following formulas are true. If any formula is false, find a counter example to demonstrate this, using a Venn diagram if you find it helpful.

---

[10] In 1877, in a letter to German mathematician Richard Dedekind (1831–1916), Cantor wrote of this result: "I see it, but I do not believe it."

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

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

(c) $A \cup (B \cap C) \subseteq (A \cup B) \cap C$

(d) $A \setminus (B \setminus C) = (A \setminus B) \setminus C$

6. Use Venn diagrams to prove that: (a) $(A \cup B)^c = A^c \cap B^c$; and (b) $(A \cap B)^c = A^c \cup B^c$

7. If A is a set with a finite number of distinct elements, let $n(A)$ denote its *cardinality*, defined as the number of elements in A. If A and B are arbitrary finite sets, prove the following:

(a) $n(A \cup B) = n(A) + n(B) - n(A \cap B)$

(b) $n(A \setminus B) = n(A) - n(A \cap B)$

8. A thousand people took part in a survey to reveal which newspaper, A, B, or C, they had read on a certain day. The responses showed that 420 had read A, 316 had read B, and 160 had read C. These figures include 116 who had read both A and B, 100 who had read A and C, and 30 who had read B and C. Finally, all these figures include 16 who had read all three papers.

(a) How many had read A, but not B?

(b) How many had read C, but neither A nor B?

(c) How many had read neither A, B, nor C?

(d) Denote the complete set of all people in the survey by $\mathcal{U}$. Applying the notation in Exercise 7, we have $n(A) = 420$ and $n(A \cap B \cap C) = 16$, for example. Describe the numbers given in the previous answers using the same notation. Why is

$$n(\mathcal{U} \setminus (A \cup B \cup C)) = n(\mathcal{U}) - n(A \cup B \cup C)?$$

SM 9. [HARDER] The equalities proved in Exercise 6 are particular cases of the *De Morgan's laws*. State and prove these two laws:

(a) The complement of the union of any family of sets equals the intersection of all the sets' complements.

(b) The complement of the intersection of any family of sets equals the union of all the sets' complements.

1.2 Essentials of Logic

Mathematical models play a critical role in the empirical sciences, including modern economics. This has been a useful development, but demands that practitioners exercise great care. Otherwise errors in mathematical reasoning, which are all too easy to make, can easily lead to nonsensical conclusions.

Here is a typical example of how faulty logic can lead to an incorrect answer. The example involves square roots, which are briefly discussed after the example.

EXAMPLE 1.2.1 Suppose that we want to find *all* the values of x for which the following equality is true: $x + 2 = \sqrt{4 - x}$.

Squaring each side of the equation gives $(x+2)^2 = (\sqrt{4-x})^2$. Expanding the left-hand side while using the definition of square root yields $x^2 + 4x + 4 = 4 - x$. Rearranging this last equation gives $x^2 + 5x = 0$. Cancelling x results in $x + 5 = 0$, and therefore $x = -5$.

According to this reasoning, the answer should be $x = -5$. Let us check this. For $x = -5$, we have $x + 2 = -3$. Yet $\sqrt{4-x} = \sqrt{9} = 3$, so this answer is incorrect.[11]

This example highlights the dangers of routine calculation without adequate thought. It may be easier to avoid similar mistakes after studying the structure of logical reasoning, which we discuss in the rest of this subsection.

A Reminder about Square Roots

The square root $\sqrt{a}$ of a nonnegative number a is the (unique) nonnegative number x such that $x^2 = a$. Thus, in particular, $\sqrt{9} = 3$, because 3 is the only nonnegative number whose square is 9.

Some older mathematical writing may claim that a positive number has two square roots, one positive and the other negative. For instance, $\sqrt{64}$ could be either 8 or -8. Allowing two values like this has now become obsolete, as it easily leads to confusion. For example, $\sqrt{49} + \sqrt{25}$ could mean any of the four numbers $7+5$, $7-5$, $-7+5$, and $-7-5$. To avoid such confusion, instead of $\sqrt{a}$ we use the more explicit notation $\pm\sqrt{a}$ to denote the set $\{\sqrt{a}, -\sqrt{a}\}$ consisting, in case $a > 0$, of the two distinct solutions to $x^2 = a$.

So remember that $\sqrt{a}$ always denotes the unique *nonnegative* solution of the equation $x^2 = a$. Of course, if a is negative, then it has no square root at all (as long as we insist that the square root must be a real number).

Propositions

Assertions that are either true or false are called statements, or *propositions*. Most of the propositions in this book are mathematical ones, but other kinds may arise in daily life. “All individuals who breathe are alive” is an example of a true proposition, whereas the assertion “all individuals who breathe are healthy” is a false proposition. Note that if the words used to express such an assertion lack precise meaning, it will often be difficult to tell whether it is true or false. For example, the assertion “67 is a large number” is neither true nor false without a precise definition of “large number”.

The assertion “$x^2 - 1 = 0$” includes the variable x. For an assertion like this, by substituting various real numbers for the variable x, we can generate many different propositions, some true and some false. For this reason we say that the assertion is an “open proposition”. In fact, the particular proposition $x^2 - 1 = 0$ happens to be true if $x = 1$ or -1, but not otherwise. Thus, an open proposition is not simply true or false. Instead, it is neither true nor false until we choose a particular value for the variable. Or for several variables in case of assertions like $x^2 + y^2 = 1$.

---

[11] Note the wisdom of checking your answer whenever you think you have solved an equation. In Example 1.2.4, below, we explain how the error arose.

Implications

In order to keep track of each step in a chain of logical reasoning, it often helps to use *implication arrows*. Suppose P and Q are two propositions such that whenever P is true, then Q is necessarily true. In this case, we usually write

$$P \implies Q$$

This can be read as "P implies Q", but it can also be read as "if P, then Q"; or "Q is a consequence of P"; or "Q if P". Furthermore, since in this case Q cannot be false while P is true, the implication can also be read as "P only if Q". The symbol $\Rightarrow$ is an *implication arrow*, which points in the direction of the logical implication. Thus $P \Rightarrow Q$ can also be written as $Q \Leftarrow P$.

EXAMPLE 1.2.2 Here are some examples of correct implications:[12]

(a) $x > 2 \Rightarrow x^2 > 4$

(b) $xy = 0 \Rightarrow$ (either $x = 0$ or $y = 0$)

(c) S is a square $\Rightarrow$ S is a rectangle

(d) She lives in Paris $\Rightarrow$ She lives in France

In certain cases where the implication $P \Rightarrow Q$ is valid, it may also be possible to draw a logical conclusion in the other direction: $Q \Rightarrow P$ (or $P \Leftarrow Q$). In such cases, we can write both implications together in a single *logical equivalence*:

$$P \iff Q$$

We then say that "P is equivalent to Q". Because both "P if Q" and "P only if Q" are true, we also say that "P if and only if Q", which is often written as "P iff Q" for short. Unsurprisingly, the symbol $\Leftrightarrow$ is called an *equivalence arrow*.

In Example 1.2.2, we see that the implication arrow in (b) could be replaced with the equivalence arrow, because it is also true that $x = 0$ or $y = 0$ implies $xy = 0$. Note, however, that no other implication in Example 1.2.2 can be replaced by the equivalence arrow, because:

(a) even if x^2 is larger than 4, it is not necessarily true that x is larger than 2 (for instance, x might be -3);

(c) a rectangle is not necessarily a square;

(d) there are millions of people who live in France but not in Paris.

EXAMPLE 1.2.3 Here are three examples of correct equivalences:

(a) $(x < -2 \text{ or } x > 2) \iff x^2 > 4$

(b) $xy = 0 \iff (x = 0 \text{ or } y = 0)$

(c) $A \subseteq B \iff (B^c \subseteq A^c)$

[12] Of course, in part (d) we are talking about Paris, France, rather than Paris, Texas, or Paris, Ontario.

The Contrapositive Principle

Suppose P and Q are propositions such that the implication $P \Rightarrow Q$ is valid. This means that *if* P is true, then Q must also be true. From this it follows that if Q is false, then P is also false. Therefore we have the implication (not $Q \Rightarrow$ not P).

We have just shown that $P \Rightarrow Q$ implies that (not $Q \Rightarrow$ not P). Suppose we replace P by not Q and Q by not P in this implication. The new implication that results is (not $Q \Rightarrow$ not P) implies (not not $P \Rightarrow$ not not Q). But (not not P) is true if and only if (not P) is false, in other words if and only if P is true. In the same way we can see that (not not Q) is the same as Q.

Thus we have shown that $P \Rightarrow Q$ implies (not $Q \Rightarrow$ not P), and that (not $Q \Rightarrow$ not P) implies $P \Rightarrow Q$. We formalize this result as follows:

THE CONTRAPOSITIVE PRINCIPLE

The statement $P \Rightarrow Q$ is logically equivalent to the statement

$$\text{not } Q \;\Rightarrow\; \text{not } P$$

This principle is often useful when proving mathematical results.

Necessity and Sufficiency

There are other commonly used ways of expressing the statement that proposition P implies proposition Q, or the alternative statement that P is equivalent to Q. Thus, if proposition P implies proposition Q, we say that P is a "sufficient condition" for Q; after all, for Q to be true, it is sufficient that P be true. Accordingly, we know that if P is satisfied, then it is certain that Q is also satisfied. In this case, because Q must necessarily be true if P is true, we say that Q is a "necessary condition" for P. Hence:

NECESSARY AND SUFFICIENT CONDITIONS

(a) $P \Rightarrow Q$ means both that P is a *sufficient condition* for Q and, equivalently, that Q is a *necessary condition* for P.

(b) The corresponding verbal expression for $P \Leftrightarrow Q$ is that P is a *necessary and sufficient condition* for Q.

It is worth noting how important it is to distinguish between the three propositions "P is a necessary condition for Q", "P is a sufficient condition for Q", and "P is a necessary and sufficient condition for Q". To emphasize this point, consider the propositions:

Living in France is a necessary condition for a person to live in Paris.

and

Living in Paris is a necessary condition for a person to live in France.

The first proposition is clearly true. But the second is false,[13] because it is possible to live in France, but outside Paris. What is true, though, is that

Living in Paris is a sufficient condition for a person to live in France.

In the following pages, we shall repeatedly refer to necessary conditions, to sufficient conditions, as well as to necessary and sufficient conditions—i.e., conditions that are both necessary and sufficient. Understanding these three, and the differences between them, is a necessary condition for understanding much of economic analysis. It is not a sufficient condition, alas!

EXAMPLE 1.2.4 In solving Example 1.2.1, why did we need to check that the values we found were actually solutions? To answer this, we must analyse the logical structure of our analysis. Using implication arrows marked by letters, we can express the "solution" proposed there as follows:

$$
\begin{aligned}
x+2=\sqrt{4-x} &\overset{(a)}{\Longrightarrow} (x+2)^2 = 4-x \\
&\overset{(b)}{\Longrightarrow} x^2+4x+4 = 4-x \\
&\overset{(c)}{\Longrightarrow} x^2+5x = 0 \\
&\overset{(d)}{\Longrightarrow} x(x+5) = 0 \\
&\overset{(e)}{\Longrightarrow} [x=0 \text{ or } x=-5]
\end{aligned}
$$

Implication (a) is true, because $a=b \Rightarrow a^2=b^2$ and $(\sqrt{a})^2=a$. It is important to note, however, that the implication cannot be replaced by an equivalence: if $a^2=b^2$, then either $a=b$ or $a=-b$; it need not be true that $a=b$. Implications (b), (c), (d), and (e) are also all true; moreover, all could have been written as equivalences, though this is not necessary in order to find the solution. In the end, therefore, we have obtained a chain of implications that leads from the equation $x+2=\sqrt{4-x}$ to the proposition "$x=0$ or $x=-5$".

Because the implication (a) cannot be reversed, there is no corresponding chain of implications going in the opposite direction. All we have done is verify that if the number x satisfies $x+2=\sqrt{4-x}$, then x must be either 0 or -5; no other value can satisfy the given equation. However, we have not yet shown that either 0 or -5 really satisfies the equation. Only after we try inserting 0 and -5 into the equation do we see that $x=0$ is the only solution.

Looking back at Example 1.2.4, we can now see that two errors were committed. First, the implication $x^2+5x=0 \Rightarrow x+5=0$ is wrong, because $x=0$ is also a solution of $x^2+5x=0$. Second, it is logically necessary to check if 0 or -5 really satisfies the equation.

[13] As is the proposition *Living in France is equivalent to living in Paris.*

EXERCISES FOR SECTION 1.2

1. There are many other ways to express implications and equivalences, apart from those already mentioned. Use appropriate implication or equivalence arrows to represent the following propositions:

   (a) The equation $2x - 4 = 2$ is fulfilled only when $x = 3$.

   (b) If $x = 3$, then $2x - 4 = 2$.

   (c) The equation $x^2 - 2x + 1 = 0$ is satisfied if $x = 1$.

   (d) If $x^2 > 4$, then $|x| > 2$, and conversely.

2. Determine which of the following formulas are true. If any formula is false, find a counter example to demonstrate this, using a Venn diagram if you find it helpful.

   (a) $A \subseteq B \Leftrightarrow A \cup B = B$

   (b) $A \subseteq B \Leftrightarrow A \cap B = A$

   (c) $A \cap B = A \cap C \Rightarrow B = C$

   (d) $A \cup B = A \cup C \Rightarrow B = C$

   (e) $A = B \Leftrightarrow (x \in A \Leftrightarrow x \in B)$

3. In each of the following implications, where x, y, and z are numbers, decide: (i) if the implication is true; and (ii) if the converse implication is true.

   (a) $x = \sqrt{4} \Rightarrow x = 2$

   (b) $(x = 2 \text{ and } y = 5) \Rightarrow x + y = 7$

   (c) $(x - 1)(x - 2)(x - 3) = 0 \Rightarrow x = 1$

   (d) $x^2 + y^2 = 0 \Rightarrow x = 0 \text{ or } y = 0$

   (e) $(x = 0 \text{ and } y = 0) \Rightarrow x^2 + y^2 = 0$

   (f) $xy = xz \Rightarrow y = z$

4. Consider the proposition $2x + 5 \geq 13$.

   (a) Is the condition $x \geq 0$ necessary, or sufficient, or both necessary and sufficient for the inequality to be satisfied?

   (b) Answer the same question when $x \geq 0$ is replaced by $x \geq 50$.

   (c) Answer the same question when $x \geq 0$ is replaced by $x \geq 4$.

SM 5. [HARDER] If P is a statement, its *negation* is that statement which is true when P is false, and false when P is true. For example, the negation of the statement $2x + 3y \leq 8$ is $2x + 3y > 8$. For each of the following six propositions, state the negation as simply as possible.

   (a) $x \geq 0$ and $y \geq 0$.

   (b) All x satisfy $x \geq a$.

   (c) Neither x nor y is less than 5.

   (d) For each $\varepsilon > 0$, there exists a $\delta > 0$ such that B is satisfied.

   (e) No one can help liking cats.

   (f) Everyone loves somebody some of the time.

1.3 Mathematical Proofs

In mathematics, the most important results are called *theorems*. Constructing logically valid proofs for these results often can be very complicated.[14] In this book, we often omit formal proofs of theorems. Instead, the emphasis is on providing a good intuitive grasp of what the theorems tell us. That said, it is still useful to understand something about the different types of proof that are used in mathematics.

Every mathematical theorem can be formulated as one or more implications of the form

$$P \implies Q \tag{$*$}$$

where P represents a proposition, or a series of propositions, called *premises* ("what we already know"), and Q represents a proposition or a series of propositions that are called the *conclusions* ("what we want to know").

Usually, it is most natural to prove a result of the type ($*$) by starting with the premises P and successively working forward to the conclusions Q; we call this a *direct proof*. Sometimes, however, it is more convenient to prove the implication $P \Rightarrow Q$ by a *contrapositive* or *indirect proof*. In this case, we begin by supposing that Q is not true, and on that basis demonstrate that P cannot be true either. This is completely legitimate, because of the contrapositive principle set out in Section 1.2.

The method of indirect proof is closely related to an alternative one known as *proof by contradiction* or *reductio ad absurdum*. According to this method, in order to prove that $P \Rightarrow Q$, one assumes that P is true and Q is not, and develops an argument that leads to something that *cannot* be true. So, since P and the negation of Q lead to something absurd, it must be that whenever P holds, so does Q.

EXAMPLE 1.3.1 Show that $-x^2 + 5x - 4 > 0 \Rightarrow x > 0$.

Solution: We can use any of the three methods of proof:

(a) *Direct proof*: Suppose $-x^2 + 5x - 4 > 0$. Adding $x^2 + 4$ to each side of the inequality gives $5x > x^2 + 4$. Because $x^2 + 4 \geq 4$, for all x, we have $5x > 4$, and so $x > 4/5$. In particular, $x > 0$.

(b) *Contrapositive proof*: Suppose $x \leq 0$. Then $5x \leq 0$ and so $-x^2 + 5x - 4$, as a sum of three nonpositive terms, is itself nonpositive.

(c) *Proof by contradiction*: Assume that $-x^2 + 5x - 4 > 0$ and $x \leq 0$ are true simultaneously. Then, as in the first step of the direct proof, we have $5x > x^2 + 4$. But since $5x \leq 0$, as in the first step of the contrapositive proof, we are forced to conclude that $0 >$

[14] For example, the "four-colour theorem" considers any map that divides a plane into several regions, and the problem of colouring these regions in order that no two adjacent regions have the same colour. As its name suggests, the theorem states that at most four colours are needed. The result was conjectured in the mid 19th century. Yet proving this involved checking hundreds of thousands of different cases. Not until the 1980s did a sophisticated computer program make possible a proof that mathematicians now generally accept as correct.

$x^2 + 4$. Since the latter cannot possibly be true, we have proved that $-x^2 + 5x - 4 > 0$ and $x \leq 0$ cannot be both true, so that $-x^2 + 5x - 4 > 0 \Rightarrow x > 0$, as desired.

Deductive and Inductive Reasoning

The methods of proof just outlined are all examples of *deductive reasoning*—that is, reasoning based on consistent rules of logic. In contrast, many branches of science use *inductive reasoning*. This process draws general conclusions based only on a few (or even many) observations. For example, the statement that "the price level has increased every year for the last n years; therefore, it will surely increase next year too" demonstrates inductive reasoning. This inductive approach is of fundamental importance in the experimental and empirical sciences, despite the fact that conclusions based upon it never can be absolutely certain. Indeed, in economics, such examples of inductive reasoning (or the implied predictions) often turn out to be false, with hindsight.

In mathematics, inductive reasoning is not recognized as a form of proof. Suppose, for instance, that students in a geometry course are asked to show that the sum of the angles of a triangle is always 180 degrees. Suppose they painstakingly measure as accurately as possible, say, one thousand different triangles, and demonstrate that in every case the sum of the angles is 180 degrees. This would not prove the assertion. At best, it would represent a very good indication that the proposition is true, yet it is not a mathematical proof. Similarly, in business economics, the fact that a particular company's profits have risen for each of the past 20 years is no guarantee that they will rise once again this year.

EXERCISES FOR SECTION 1.3

1. Which of the following statements are equivalent to the (dubious) statement: "If inflation increases, then unemployment decreases"?

   (a) For unemployment to decrease, inflation must increase.

   (b) A sufficient condition for unemployment to decrease is that inflation increases.

   (c) Unemployment can only decrease if inflation increases.

   (d) If unemployment does not decrease, then inflation does not increase.

   (e) A necessary condition for inflation to increase is that unemployment decreases.

2. Analyse the following epitaph, using logic:

   Those who knew him, loved him. Those who loved him not, knew him not.

   Might this be a case where poetry is better than logic?

3. Use the contrapositive principle to show that if x and y are integers and xy is an odd number, then x and y are both odd.

1.4 Mathematical Induction

Unlike inductive reasoning, mathematical induction is a form of argument that relies entirely on logic. It sees widespread use in proving formulas and even theorems that involve natural numbers. Consider, for example, the sum of the first n odd numbers. A little calculation shows that for $n = 1, 2, 3, 4, 5$ one has

$$\begin{aligned}1 &= 1 = 1^2\\1+3 &= 4 = 2^2\\1+3+5 &= 9 = 3^2\\1+3+5+7 &= 16 = 4^2\\1+3+5+7+9 &= 25 = 5^2\end{aligned}$$

This suggests a general pattern, with the sum of the first n odd numbers equal to n^2:

$$P(n): \quad 1 + 3 + 5 + \cdots + (2n-1) = n^2 \tag{$*$}$$

We call Eq. $(*)$ the *induction hypothesis*, a proposition that we denote by $P(n)$. To prove that $P(n)$ really is valid for general n, we can proceed as follows. First, start with the *base case*, denoted by $P(1)$, which states that the formula $(*)$ is correct when n equals 1.

Next, the key *induction step* involves showing that, for any $k \geq 1$, if $P(k)$ is true, then it follows that $P(k+1)$ is true. In other words, one proves that $P(k) \Rightarrow P(k+1)$. To do this, simply add the $(k+1)$th odd number, which is $2k+1$, to each side of $(*)$. This gives

$$1 + 3 + 5 + \cdots + (2k-1) + (2k+1) = k^2 + (2k+1) = (k+1)^2$$

But this is precisely $P(k+1)$ in which: (i) the left-hand side of formula $(*)$ ends, not with the kth odd number $2k-1$, but with the $(k+1)$th odd number $2k+1$; (ii) the right-hand side of $(*)$ has been "stepped up" from k^2 to $(k+1)^2$. This completes the proof of the "induction step" showing that, if $P(k)$ holds because the sum of the first k odd numbers really is k^2, then $P(k+1)$ holds because the sum of the first $k+1$ odd numbers equals $(k+1)^2$.

Given the base case stating that formula $(*)$ is valid for $n = 1$, this "induction step" *implies* that $(*)$ is valid for general n. This is because, if $(*)$ holds for $n = 1$, the induction step we have just shown implies that it holds also for $n = 2$; that if it holds for $n = 2$, then it also holds for $n = 3$; ...; that if it holds for n, then it holds also for $n+1$; and so on.

A proof of this type is called a *proof by induction*.[15] It requires showing: (i) first that the formula is indeed valid in the base case when $n = 1$; (ii) second that, *if* the formula is valid when $n = k$, then it is also valid when $n = k+1$, which is the induction step. It follows by induction that the formula is valid for all natural numbers n.

[15] Arguments by induction can be traced as far back as ancient Greek philosophers and mathematicians, including Plato and Euclid.

EXAMPLE 1.4.1 Prove by induction that, for all positive integers n,

$$3 + 3^2 + 3^3 + 3^4 + \cdots + 3^n = \tfrac{1}{2}(3^{n+1} - 3) \qquad (**)$$

Solution: In the base case when $n = 1$, both sides are 3. For the induction step, suppose that $(**)$ is true for $n = k$. Then adding the next term 3^{k+1} to each side of $(**)$ gives

$$3 + 3^2 + 3^3 + 3^4 + \cdots + 3^k + 3^{k+1} = \tfrac{1}{2}(3^{k+1} - 3) + 3^{k+1} = \tfrac{1}{2}(3^{k+2} - 3)$$

But this is precisely $(**)$ restated for $n = k + 1$ instead of k. So, by induction, $(**)$ is true for all n.

Following these examples, the general structure of an induction proof can be explained as follows. The aim is to prove that a logical statement, for instance a mathematical formula $P(n)$ that depends on n, is true for all natural numbers n. In the two previous examples, the respective statements $P(n)$ were

$$P(n): \quad 1 + 3 + 5 + \cdots + (2n - 1) = n^2$$

and

$$P(n): \quad 3 + 3^2 + 3^3 + 3^4 + \cdots + 3^n = \tfrac{1}{2}(3^{n+1} - 3)$$

The steps required in each proof are as follows. First, as the base case, verify that $P(1)$ is valid, which means that the formula is correct for $n = 1$. Second, prove that for each natural number k, *if* $P(k)$ is true, *then* it follows that $P(k + 1)$ must be true. Here, the fact that $P(k)$ is true is called *the induction hypothesis*, and the move from $P(k)$ to $P(k + 1)$ is called *the induction step* of the proof. When $P(1)$ is true and the induction step has been proved for an arbitrary natural number k, we can conclude, by induction, that statement $P(n)$ is true for all n.

The general principle can be formulated as follows:

THE PRINCIPLE OF MATHEMATICAL INDUCTION

For each natural number n, let $P(n)$ denote a statement that depends on n. Suppose that:

(a) $P(1)$ is true; and

(b) for each natural number k, *if* $P(k)$ is true *then* $P(k + 1)$ is true.

It follows that $P(n)$ is true for all natural numbers n.

The principle of mathematical induction seems intuitively obvious. If the truth of $P(k)$ for each k implies the truth of $P(k + 1)$, then because $P(1)$ is true, it follows that $P(2)$ must be true, which, in turn, means that $P(3)$ is true, and so on indefinitely.[16] That is, we have established the infinite chain of implications

$$P(1) \Rightarrow P(2) \Rightarrow P(3) \Rightarrow \cdots \Rightarrow P(k) \Rightarrow P(k+1) \Rightarrow \cdots$$

The principle of mathematical induction can easily be extended to the case when a statement $P(n)$ is true for each integer greater than or equal to an arbitrary integer n_0. Indeed, suppose we can prove that $P(n_0)$ is valid and moreover that, for each $k \geq n_0$, if $P(k)$ is true, then $P(k+1)$ is true. It follows that $P(n)$ is true for all $n \geq n_0$.

EXERCISES FOR SECTION 1.4

1. Prove by induction that for all natural numbers n,

$$1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1) \tag{$*$}$$

2. Prove by induction that

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \cdots + \frac{1}{n(n+1)} = \frac{n}{n+1} \tag{$**$}$$

3. After noting that the sum $1^3 + 2^3 + 3^3 = 36$ is divisible by 9, prove by induction that the sum $n^3 + (n+1)^3 + (n+2)^3$ of three consecutive cubes is always divisible by 9.

4. Let $P(n)$ be the statement:

   Any collection of n people in one room all have the same income.

   Find what is wrong with the following "induction argument":

   $P(1)$ is obviously true. Suppose $P(n)$ is true for some natural number n. We will then prove that $P(n+1)$ is true. So take any collection of $n+1$ people in one room and send one of them outside. The remaining n people all have the same income by the induction hypothesis. Bring the person back inside and send another outside instead. Again the remaining people will have the same income. But then all the $n+1$ people will have the same income. By induction, this proves that all n people have the same income.

REVIEW EXERCISES

1. Let $A = \{1, 3, 4\}$, $B = \{1, 4, 6\}$, $C = \{2, 4, 3\}$, and $D = \{1, 5\}$. Find $A \cap B$; $A \cup B$; $A \setminus B$; $B \setminus A$; $(A \cup B) \setminus (A \cap B)$; $A \cup B \cup C \cup D$; $A \cap B \cap C$; and $A \cap B \cap C \cap D$.

[16] Here is an analogy: Consider climbing a ladder with an infinite number of steps. Suppose you can climb the first step and suppose, moreover, that after each step, you can always climb the next. Then you are able to climb up to any step. Literally, induction arguments prove general statements *step by step*!

2. Let the universal set be $\Omega = \{1, 2, 3, 4, \ldots, 11\}$, and define $A = \{1, 4, 6\}$ and $B = \{2, 11\}$. Find $A \cap B$; $A \cup B$; $\Omega \setminus B$; A^c.

SM 3. A liberal arts college has one thousand students. The numbers studying various languages are: English 780; French 220; and Spanish 52. These figures include 110 who study English and French, 32 who study English and Spanish, 15 who study French and Spanish. Finally, all these figures include ten students taking all three languages.

(a) How many study English and French, but not Spanish?

(b) How many study English, but not French?

(c) How many study no languages?

SM 4. Let x and y be real numbers. Consider the following implications and decide in each case: (i) if the implication is true; and (ii) if the converse implication is true.

(a) $x = 5$ and $y = -3 \Rightarrow x + y = 2$

(b) $x^2 = 16 \Rightarrow x = 4$

(c) $(x - 3)^2(y + 2)$ is a positive number $\Rightarrow y$ is greater than -2

(d) $x^3 = 8 \Rightarrow x = 2$

5. [Harder] (If you are not yet familiar with inequalities and with nth powers, see Sections 2.2 and 2.6 in Chapter 2.) Prove the following result, known as *Bernoulli's inequality*:[17] for every real number $x \geq -1$ and every natural number n, one has $(1 + x)^n \geq 1 + nx$.

[17] Named after Jacob Bernoulli (1654–1705), one of a large extended family of prominent mathematicians and other scholars.

2

ALGEBRA

God made the integers, all else is the work of Man.
—Leopold Kronecker[1]

The main topic covered in this chapter is elementary algebra. Nevertheless, we also briefly consider a few other topics that you might find the need to review. Indeed, tests reveal that even students with a good background in mathematics often benefit from a brief review of what they had learned in the past. These students should browse through the material in this chapter, and do some of the less simple exercises. Students with a weaker background in mathematics, however, or those who have been away from mathematics for a long time, should read the text carefully and then do most of the exercises. Finally, those students who have considerable difficulties with this chapter should turn to a more elementary book on algebra.

2.1 The Real Numbers

We start by reviewing some important facts and concepts concerning numbers. The basic numbers are the *natural numbers*:

$$1,\ 2,\ 3,\ 4,\ \dots$$

also called *positive integers*. Of these, the multiples of 2 are the *even numbers* 2, 4, 6, 8, …, whereas 1, 3, 5, 7, … are the *odd numbers*. Though familiar, such numbers are in reality rather abstract and advanced concepts. Civilization crossed a significant threshold when it grasped the idea that a flock of four sheep and a collection of four stones have something in common, namely "fourness". This idea came to be represented by symbols such as the primitive :: (still used on dominoes, dice, and playing cards), the Roman numeral IV, and eventually the modern 4. Most children grasp this key notion

[1] Attributed; circa 1886.

even while quite young, and then continually refine it as they develop their mathematical skills.

The positive integers, together with 0 and the *negative integers* $-1, -2, -3, -4, \ldots$, make up the integers, which are

$$0, \pm 1, \pm 2, \pm 3, \pm 4, \ldots$$

They can be represented on a *number line* like the one shown in Fig. 2.1.1, where the arrow gives the direction in which the numbers increase.

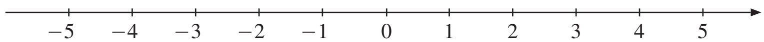

Figure 2.1.1 The number line

The *rational numbers* are those like $3/5$ that can be written in the form a/b, where a and b are both integers and $b \neq 0$. Any integer n is also a rational number, because $n = n/1$. Other examples of rational numbers are

$$\frac{1}{2}, \quad \frac{11}{70}, \quad \frac{125}{7}, \quad -\frac{10}{11}, \quad 0 = \frac{0}{1}, \quad -19, \quad -1.26 = -\frac{126}{100}$$

The rational numbers can also be represented on the number line. Suppose that we first mark $1/2$ and all the multiples of $1/2$. Next we mark $1/3$ and all the multiples of $1/3$, then all the multiples of $1/4$, and so forth. You can be excused for thinking that "finally" there will be no gaps left in which one can find other points on the line. But in fact this is quite wrong. The ancient Greeks already understood that "holes" would remain in the number line even after all the rational numbers had been marked off. For instance, there are no integers p and q such that $\sqrt{2} = p/q$. Hence, $\sqrt{2}$ is not a rational number.[2]

The rational numbers are therefore insufficient for measuring all possible lengths, let alone areas and volumes. This deficiency can be remedied by extending the concept of numbers to allow for the so-called *irrational numbers*. As explained below, this extension can be carried out rather naturally by using decimal notation.

The way most people write numbers today is called the *decimal*, or *base 10 system*. It is a positional system with 10 as the base number. Every natural number can be written using only the symbols, 0, 1, 2, ..., 9, which are called (decimal) *digits*.[3] The positional decimal system uses a list of decimal digits to represent a sum of powers of 10. For example,

$$1\,984 = 1 \cdot 10^3 + 9 \cdot 10^2 + 8 \cdot 10^1 + 4 \cdot 10^0$$

Each natural number can be uniquely expressed in this manner. With the use of the signs $+$ and $-$, all integers, positive or negative, can be written in the same way. Decimal points also enable us to express rational numbers that are not natural numbers. For example,

$$3.1415 = 3 + \frac{1}{10^1} + \frac{4}{10^2} + \frac{1}{10^3} + \frac{5}{10^4}$$

---

[2] The first full proof that has survived appeared in Euclid's *Elements*, dating from around the year 300 BCE.

[3] You may recall that a digit is either a finger or a thumb, and that most humans and indeed other primates have ten digits.

Rational numbers that can be written exactly using only a finite number of decimal places are called *finite decimal fractions*.

Each finite decimal fraction is a rational number, but not every rational number can be written as a finite decimal fraction. We must also allow *infinite decimal fractions* such as

$$\frac{100}{3} = 33.333\ldots$$

where the three dots at the end indicate that the digit 3 is repeated indefinitely.

If the decimal fraction is a rational number, then it will always be *periodic* or *recurring* —that is, after a certain place in the decimal expansion, it either stops or continues by repeating indefinitely a finite sequence of digits. For example,

$$\frac{11}{70} = 0.1\underbrace{571428}\underbrace{571428}\ldots$$

using notation which indicates that the sequence 571428 of six successive digits is repeated infinitely often.

Our definition of a real number follows from the previous discussion. We define a *real number* as an arbitrary infinite decimal fraction. Hence, a real number is of the form $x = \pm m.\alpha_1\alpha_2\alpha_3\ldots$, where m is a nonnegative integer, and for each natural number n, the symbol α_n indicates a *decimal digit* that belongs to the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.[4]

We have already identified the periodic decimal fractions with the rational numbers. In addition, there are infinitely many new numbers given by nonperiodic decimal fractions. These new numbers are called *irrational numbers*. Examples include the numbers $\sqrt{2}$, $-\sqrt{5}$, π, $2^{\sqrt{2}}$, and $0.12112111211112\ldots$.[5]

We mentioned earlier that each rational number can be represented by a unique point on the number line. Even after all the rational numbers have been positioned on this line, there are still some "holes" which can be "filled up" with irrational numbers. Thus, an unbroken and endless straight line with an origin and a positive unit of length constitutes a satisfactory model for the real numbers. We frequently state that there is a *one-to-one correspondence* between the real numbers and the points on a number line. For this reason, one often speaks of the "real line" rather than the "number line".

Both the set of rational numbers as well as the complementary set of irrational numbers are said to be "dense" in the number line. This means that between any two different real numbers, no matter how close they are to each other, we can always find both a rational and an irrational number—in fact, we can always find infinitely many of each.

When applied to the real numbers, each of the four basic arithmetic operations of addition, subtraction, multiplication and division always results in a unique real number. The only exception is that we cannot divide by 0: in words usually attributed to the

---

4 It is worth noting that any finite decimal fraction (except 0) can also be written as an infinite decimal fraction with a tail entirely of repeated 9s. For instance, $5.347 = 5.346999\ldots$.

5 In general, mathematicians find it very difficult to show that a number which seems evidently irrational really is. For example, whereas it has been known since the year 1776 that π is irrational, it took until 1927 to determine that $2^{\sqrt{2}}$ is irrational. For many other numbers the challenge of proving their irrationality remains.

American stand-up comedian Steven Wright, "Black holes are where God divided by zero."

DIVISION BY ZERO

The ratio $a/0$ is *not* defined for any real number a.

This exception is very important; it should not be confused with the fact that $0/b = 0$ for all $b \neq 0$. Notice especially that $0/0$ is not defined as any real number. For example, if a car requires 60 litres of fuel to go 600 kilometres, then its fuel consumption is $60/600 = 10$ litres per 100 kilometres. However, if told that a car uses 0 litres of fuel to go 0 kilometres, we know nothing about its fuel consumption; $0/0$ is completely undefined.

EXERCISES FOR SECTION 2.1

1. Which of the following statements are true?

   (a) 1984 is a natural number.

   (b) -5 is to the right of -3 on the number line.

   (c) -13 is a natural number.

   (d) There is no natural number that is not rational.

   (e) 3.1415 is not rational.

   (f) The sum of two irrational numbers is irrational.

   (g) $-3/4$ is rational.

   (h) All rational numbers are real.

2. Explain why the infinite decimal expansion

$$1.01001000100001000001\ldots$$

   is not a rational number.

2.2 Integer Powers

You should recall that we often write 3^4 instead of the fourfold product $3 \cdot 3 \cdot 3 \cdot 3$. Furthermore, the number $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$ can be written as $\left(\frac{1}{2}\right)^5$, and $(-10)^3$ is the triple product $(-10)(-10)(-10) = -1000$. Indeed, if a is any real number and n is any natural number, then a^n is defined by

$$a^n = \underbrace{a \cdot a \cdot \ldots \cdot a}_{n \text{ factors}}$$

The expression a^n is called the *nth power* of a; here a is the *base*, and n is the *exponent*. For example, we have $a^2 = a \cdot a$, and $x^4 = x \cdot x \cdot x \cdot x$. When $a = p/q$ and $n = 5$ we have

$$\left(\frac{p}{q}\right)^5 = \frac{p}{q} \cdot \frac{p}{q} \cdot \frac{p}{q} \cdot \frac{p}{q} \cdot \frac{p}{q}$$

By convention, the first power $a^1 = a$ is a "product" with only one factor.

We usually drop the multiplication sign if this is unlikely to create misunderstanding. For example, we write abc instead of $a \cdot b \cdot c$, but it is safest to keep the multiplication sign in expressions with decimal points like $1.05^3 = 1.05 \cdot 1.05 \cdot 1.05$.

For any real number $a \neq 0$, we also define its "zeroth" power $a^0 = 1$. Thus, $5^0 = 1$; $(-16.2)^0 = 1$; and $(x \cdot y)^0 = 1$, if $x \cdot y \neq 0$. But in case $a = 0$, we do *not* assign a numerical value to a^0: the expression 0^0 is *undefined*.

We also need to define powers with negative exponents. What do we mean by 3^{-2}? It turns out that the sensible definition is to set 3^{-2} equal to $1/3^2 = 1/9$. In general,

$$a^{-n} = \frac{1}{a^n}$$

whenever n is a natural number and $a \neq 0$. In particular, $a^{-1} = 1/a$. In this way we have defined a^x for all integers x, regardless of whether x is positive, negative, or zero.

Properties of Powers

There are some rules for powers that you really must not only know by heart, but also understand why they are true. The two most important are:

PROPERTIES OF POWERS

For any real number a, and any integer numbers r and s:

$$a^r \cdot a^s = a^{r+s}$$

while

$$(a^r)^s = a^{rs}$$

Note carefully what these rules say. According to the first rule, powers with the same base are multiplied by *adding* the exponents. For example,

$$a^3 \cdot a^5 = \underbrace{a \cdot a \cdot a}_{3 \text{ factors}} \cdot \underbrace{a \cdot a \cdot a \cdot a \cdot a}_{5 \text{ factors}} = \underbrace{a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a}_{3+5=8 \text{ factors}} = a^8 = a^{3+5}$$

Here is an example of the second rule:

$$(a^2)^4 = \underbrace{\underbrace{a \cdot a}_{2 \text{ factors}} \cdot \underbrace{a \cdot a}_{2 \text{ factors}} \cdot \underbrace{a \cdot a}_{2 \text{ factors}} \cdot \underbrace{a \cdot a}_{2 \text{ factors}}}_{4 \text{ factors}} = \underbrace{a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a \cdot a}_{4 \cdot 2 = 8 \text{ factors}} = a^8 = a^{2 \cdot 4}$$

Division of two powers with the same nonzero base goes like this:

$$a^r \div a^s = \frac{a^r}{a^s} = a^r \frac{1}{a^s} = a^r \cdot a^{-s} = a^{r-s}$$

Thus we divide two powers with the same base by *subtracting* the exponent in the denominator from that in the numerator.[6] For example, $a^3 \div a^5 = a^{3-5} = a^{-2}$.

Finally, note that

$$(ab)^r = \underbrace{ab \cdot ab \cdot \ldots \cdot ab}_{r \text{ factors}} = \underbrace{a \cdot a \cdot \ldots \cdot a}_{r \text{ factors}} \cdot \underbrace{b \cdot b \cdot \ldots \cdot b}_{r \text{ factors}} = a^r b^r$$

and

$$\left(\frac{a}{b}\right)^r = \underbrace{\frac{a}{b} \cdot \frac{a}{b} \cdot \ldots \cdot \frac{a}{b}}_{r \text{ factors}} = \frac{\overbrace{a \cdot a \cdot \ldots \cdot a}^{r \text{ factors}}}{\underbrace{b \cdot b \cdot \ldots \cdot b}_{r \text{ factors}}} = \frac{a^r}{b^r} = a^r b^{-r}$$

These rules can be extended to cases where there are several factors. For instance,

$$(abcde)^r = a^r b^r c^r d^r e^r$$

We saw that $(ab)^r = a^r b^r$. What about $(a+b)^r$? One of the most common errors committed in elementary algebra is to equate this to $a^r + b^r$. For example, $(2+3)^3 = 5^3 = 125$, but $2^3 + 3^3 = 8 + 27 = 35$. Thus, in general, $(a+b)^r \neq a^r + b^r$.

EXAMPLE 2.2.1 Simplify the expressions:

(a) $x^p x^{2p}$, where p is an integer

(b) $t^s \div t^{s-1}$, where $t \neq 0$ and s is an integer

(c) $a^2 b^3 a^{-1} b^5$, where $a \neq 0$

(d) $\dfrac{t^p t^{q-1}}{t^r t^{s-1}}$, where $t \neq 0$ and p, q, r, s are integers

Solution:

(a) $x^p x^{2p} = x^{p+2p} = x^{3p}$

(b) $t^s \div t^{s-1} = t^{s-(s-1)} = t^{s-s+1} = t^1 = t$

(c) $a^2 b^3 a^{-1} b^5 = a^2 a^{-1} b^3 b^5 = a^{2-1} b^{3+5} = a^1 b^8 = ab^8$

(d) Finally,

$$\frac{t^p \cdot t^{q-1}}{t^r \cdot t^{s-1}} = \frac{t^{p+q-1}}{t^{r+s-1}} = t^{p+q-1-(r+s-1)} = t^{p+q-1-r-s+1} = t^{p+q-r-s}$$

EXAMPLE 2.2.2 If $x^{-2}y^3 = 5$, compute $x^{-4}y^6$, x^6y^{-9}, and $x^2y^{-3} + 2x^{-10}y^{15}$.

Solution: First, note that $x^{-2}y^3 = 5$ is only possible if $x \neq 0$ and $y \neq 0$. Now, in computing $x^{-4}y^6$, how can we make use of the assumption that $x^{-2}y^3 = 5$? A moment's reflection might lead you to see that $(x^{-2}y^3)^2 = x^{-4}y^6$, and hence $x^{-4}y^6 = 5^2 = 25$. Similarly,

$$x^6y^{-9} = (x^{-2}y^3)^{-3} = 5^{-3} = 1/125$$

[6] An important motivation for introducing the definitions $a^0 = 1$ and $a^{-n} = 1/a^n$ is that we want the properties of powers to be valid for negative and zero exponents as well as for positive ones. For example, we want $a^r \cdot a^s = a^{r+s}$ to be valid when $r = 5$ and $s = 0$. This requires that $a^5 \cdot a^0 = a^{5+0} = a^5$, so we must choose $a^0 = 1$. If $a^n \cdot a^m = a^{n+m}$ is to be valid when $m = -n$, we must have $a^n \cdot a^{-n} = a^{n+(-n)} = a^0 = 1$. Because $a^n \cdot (1/a^n) = 1$, we *must* define a^{-n} to be $1/a^n$.

and

$$x^2y^{-3} + 2x^{-10}y^{15} = (x^{-2}y^3)^{-1} + 2(x^{-2}y^3)^5 = 5^{-1} + 2 \cdot 5^5 = 6250.2$$

EXAMPLE 2.2.3 It is easy to make mistakes when dealing with powers. The following examples highlight some common sources of confusion.

(a) There is an important difference between $(-10)^2 = (-10)(-10) = 100$, and $-10^2 = -(10 \cdot 10) = -100$. The square of minus 10 is not equal to minus the square of 10.

(b) Note that $(2x)^{-1} = 1/(2x)$. Here the product $2x$ is raised to the power of -1. On the other hand, in $2x^{-1}$ only x is raised to the power -1, so $2x^{-1} = 2 \cdot (1/x) = 2/x$.

(c) The volume of a ball with radius r is $\frac{4}{3}\pi r^3$. What will the volume be if the radius is doubled? The new volume is

$$\frac{4}{3}\pi(2r)^3 = \frac{4}{3}\pi(2r)(2r)(2r) = \frac{4}{3}\pi 8r^3 = 8\left(\frac{4}{3}\pi r^3\right)$$

so the volume is 8 times the initial one. If we made the mistake of "simplifying" $(2r)^3$ to $2r^3$, the result would imply only a doubling of the volume; this should be contrary to common sense.

Percentages and Compound Interest

Powers are used in practically every branch of applied mathematics, including economics. To illustrate their use, recall how they are needed to calculate compound interest.

First, recall that the percentage 1% means one in a hundred, or 0.01. So 23%, for example, is $23 \cdot 0.01 = 0.23$. Then we can calculate 23% of 4000 as either $4000 \cdot 23/100 = 920$ or, equivalently, as $4000 \cdot 0.23 = 920$. It may be also be worth pointing out a subtlety of percentages that is often overlooked by those who know rather little mathematics. This is that, for example, it takes an increase of 100% to reverse an earlier decrease of 50%, and a decrease of 50% to offset an earlier increase of 100%.[7]

Now suppose you deposit \$1 000 in a bank account paying 8% interest at the end of each year. After one year you will have earned $\$1\,000 \cdot 0.08 = \$80$ in interest, so the total amount in your bank account will be \$1 080. This can be rewritten as

$$1000 + \frac{1000 \cdot 8}{100} = 1000\left(1 + \frac{8}{100}\right) = 1000 \cdot 1.08$$

Suppose this new amount of \$1 080 is left in the bank for another year at an interest rate of 8%. After a second year, the extra interest will be $\$1\,000 \cdot 1.08 \cdot 0.08$. So the total amount will have grown to

$$1000 \cdot 1.08 + (1000 \cdot 1.08) \cdot 0.08 = 1000 \cdot 1.08(1 + 0.08) = 1000 \cdot (1.08)^2$$

[7] Compare review exercise 20 at the end of the chapter.

Extending this argument in an obvious way, we see that each year the amount will increase by the factor 1.08, and that at the end of t years it will have grown to $\$1\,000 \cdot (1.08)^t$.

If the original amount is $\$K$ and the interest rate is $p\%$ per year, by the end of the first year, the amount will be $K + K \cdot p/100 = K(1 + p/100)$ dollars. The growth factor per year is thus $1 + p/100$. In general, after t (whole) years, the original investment of $\$K$ will have grown to an amount

$$K\left(1 + \frac{p}{100}\right)^t$$

when the interest rate is $p\%$ per year and interest is added to the capital every year—that is, there is compound interest.

This example illustrates a general principle:

EXPONENTIAL GROWTH

A quantity K which increases by $p\%$ per year will have increased after t years to

$$K\left(1 + \frac{p}{100}\right)^t$$

Here $1 + p/100$ is called the *growth factor* for a growth of $p\%$.

If you see an expression like $(1.08)^t$ you should immediately be able to recognize it as the amount to which \$1 has grown after t years when the interest rate is 8% per year. How should you interpret $(1.08)^0$? Suppose you deposit \$1 at 8% per year, and leave the amount in the account for 0 years. Then you still have only \$1, because there has been no time to accumulate any interest. This explains why $(1.08)^0$ must necessarily equal 1.[8]

EXAMPLE 2.2.4 A new car has been bought for \$30 000 and is assumed to decrease in value (depreciate) by 15% per year over a six-year period. What is its value after six years?

Solution: After one year its value is down to

$$30\,000 - \frac{30\,000 \cdot 15}{100} = 30\,000 \cdot \left(1 - \frac{15}{100}\right) = 30\,000 \cdot 0.85 = 25\,500$$

After two years its value is $\$30\,000 \cdot (0.85)^2 = \$21\,675$, and so on. After six years we realize that its value must be $\$30\,000 \cdot (0.85)^6 \approx \$11\,314$.

[8] Note that $1000 \cdot (1.08)^5$ is the amount you will have in your account after five years if you invest \$1 000 at 8% interest per year. Using a calculator shows that you will have approximately \$1 469.33. A rather common mistake is to put $1000 \cdot (1.08)^5 = (1000 \cdot 1.08)^5 = (1080)^5$. This is one trillion ($10^{12}$) times the right answer!

This example illustrates a general principle:

EXPONENTIAL DECLINE

A quantity K which decreases by $p\%$ per year will have shrunk after t years to

$$K\left(1-\frac{p}{100}\right)^t$$

Here $1 - p/100$ is the growth factor that results from a decline of $p\%$ per year. (Note that a growth factor that is less than 1 indicates shrinkage.)

Do We Really Need Negative Exponents?

How much money should you have deposited in a bank five years ago in order to have \$1 000 today, given that the interest rate has been 8% per year over this period? If we call this amount x, the requirement is that $x \cdot (1.08)^5$ must equal \$1 000, or that $x \cdot (1.08)^5 = 1000$. Dividing by 1.08^5 on both sides yields

$$x = \frac{1000}{(1.08)^5} = 1000 \cdot (1.08)^{-5}$$

which is approximately \$681. Thus, $\$(1.08)^{-5}$ is what you should have deposited five years ago in order to have \$1 today, given the constant interest rate of 8%.

In general, $\$P(1+p/100)^{-t}$ *is what you should have deposited t years ago in order to have* $\$P$ *today, if the interest rate has been* $p\%$ *every year.*

EXERCISES FOR SECTION 2.2

1. Compute the following numbers: (a) 10^3; (b) $(-0.3)^2$; (c) 4^{-2}; and (d) $(0.1)^{-1}$.

2. Write as powers of 2 the following numbers: (a) 4; (b) 1; (c) 64; and (d) 1/16.

3. Write as powers the following numbers:

(a) $15 \cdot 15 \cdot 15$ (b) $\left(-\frac{1}{3}\right)\left(-\frac{1}{3}\right)\left(-\frac{1}{3}\right)$ (c) $\frac{1}{10}$ (d) 0.0000001

(e) $t\,t\,t\,t\,t\,t$ (f) $(a-b)(a-b)(a-b)$ (g) $a\,a\,b\,b\,b\,b$ (h) $(-a)(-a)(-a)$

4. Expand and simplify the following expressions:

(a) $2^5 \cdot 2^5$ (b) $3^8 \cdot 3^{-2} \cdot 3^{-3}$ (c) $(2x)^3$ (d) $(-3xy^2)^3$

(e) $\dfrac{p^{24}p^3}{p^4 p}$ (f) $\dfrac{a^4 b^{-3}}{(a^2 b^{-3})^2}$ (g) $\dfrac{3^4(3^2)^6}{(-3)^{15}3^7}$ (h) $\dfrac{p^{\gamma}(pq)^{\sigma}}{p^{2\gamma+\sigma}q^{\sigma-2}}$