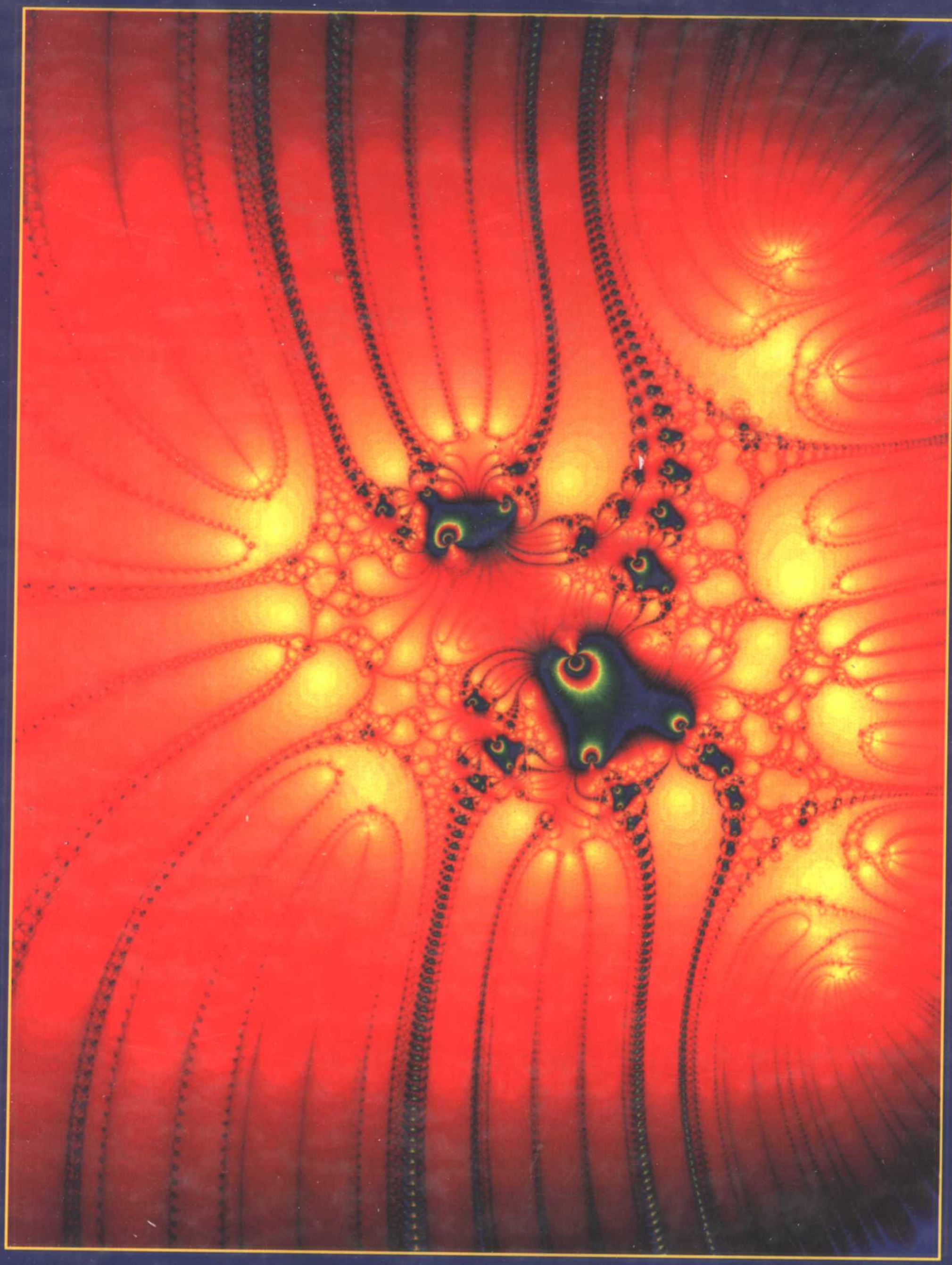
Vector Calculus, Linear Algebra, and Differential Forms

A Unified Approach



John H. Hubbard • Barbara Burke Hubbard

Vector Calculus, Linear Algebra, And Differential Forms

A Unified Approach

John Hamal Hubbard
Cornell University

Barbara Burke Hubbard

PRENTICE HALL
Upper Saddle River, New Jersey 07458

Preface

> *... The numerical interpretation ... is however necessary. ... So long as it is not obtained, the solutions may be said to remain incomplete and useless, and the truth which it is proposed to discover is no less hidden in the formulae of analysis than it was in the physical problem itself.*
>
> –Joseph Fourier, *The Analytic Theory of Heat*

This book covers most of the standard topics in multivariate calculus, and a substantial part of a standard first course in linear algebra. The teacher may find the organization rather less standard.

There are three guiding principles which led to our organizing the material as we did. One is that at this level linear algebra should be more a convenient setting and language for multivariate calculus than a subject in its own right. We begin most chapters with a treatment of a topic in linear algebra and then show how the methods apply to corresponding nonlinear problems. In each chapter, enough linear algebra is developed to provide the tools we need in teaching multivariate calculus (in fact, somewhat more: the spectral theorem for symmetric matrices is proved in Section 3.7). We discuss abstract vector spaces in Section 2.6, but the emphasis is on $\mathbb{R}^n$, as we believe that most students find it easiest to move from the concrete to the abstract.

Another guiding principle is that one should emphasize computationally effective algorithms, and prove theorems by showing that those algorithms really work: to marry theory and applications by using practical algorithms as theoretical tools. We feel this better reflects the way this mathematics is used today, in both applied and in pure mathematics. Moreover, it can be done with no loss of rigor.

For linear equations, row reduction (the practical algorithm) is the central tool from which everything else follows, and we use row reduction to prove all the standard results about dimension and rank. For nonlinear equations, the cornerstone is Newton's method, the best and most widely used method for solving nonlinear equations. We use Newton's method both as a computational tool and as the basis for proving the inverse and implicit function theorem, rather than basing those proofs on Picard iteration, which converges too slowly to be of practical interest.

In keeping with our emphasis on computations, we include a section on numerical methods of integration, and we encourage the use of computers to both to reduce tedious calculations (row reduction in particular) and as an aid in visualizing curves and surfaces. We have also included a section on probability and integrals, as this seems to us too important a use of integration to be ignored.

A third principle is that differential forms are the right way to approach the various forms of Stokes's theorem. We say this with some trepidation, especially after some of our most distinguished colleagues told us they had never really understood what differential forms were about. We believe that differential forms can be taught to freshmen and sophomores, if forms are presented geometrically, as integrands that take an oriented piece of a curve, surface, or manifold, and return a number. We are aware that students taking courses in other fields need to master the language of vector calculus, and we devote three sections of Chapter 6 to integrating the standard vector calculus into the language of forms.

The great conceptual simplifications gained by doing electromagnetism in the language of forms is a central motivation for using forms, and we will apply the language of forms to electromagnetism in a subsequent volume.

Although most long proofs have been put in Appendix A, we made an exception for the material in Section 1.6. These theorems in topology are often not taught, but we feel we would be doing the beginning student a disservice not to include them, particularly the mean value theorem and the theorems concerning convergent subsequences in compact sets and the existence of minima and maxima of functions. In our experience, students do not find this material particularly hard, and systematically avoiding it leaves them with an uneasy feeling that the foundations of the subject are shaky.

Jean Dieudonné, for many years a leader of Bourbaki, is the very personification of rigor in mathematics. In his book *Infinitesimal Calculus*, he put the harder proofs in small type, saying " ... a beginner will do well to accept plausible results without taxing his mind with subtle proofs ... ."

Following this philosophy, we have put many of the more difficult proofs in the appendix, and feel that for a first course, these proofs should be omitted. Students should learn how to drive before they learn how to take the car apart.

Different ways to use the book

This book can be used either as a textbook in multivariate calculus or as an accessible textbook for a course in analysis.

We see calculus as analogous to learning how to drive, while analysis is analogous to learning how and why a car works. To use this book to "learn how to drive," the proofs in Appendix A should be omitted. To use it to "learn how a car works," the emphasis should be on those proofs. For most students, this will be best attempted when they already have some familiarity with the material in the main text.

Students who have studied first year calculus only

(1) For a one-semester course taken by students have studied neither linear algebra nor multivariate calculus, we suggest covering only the first four chapters, omitting the sections marked "optional," which, in the analogy of learning

to drive rather than learning how a car is built, correspond rather to learning how to drive on ice. (These sections include the part of Section 2.8 concerning a stronger version of the Kantorovitch theorem, and Section 4.4 on measure 0). Other topics that can be omitted in a first course include the proof of the fundamental theorem of algebra in Section 1.6, the discussion of criteria for differentiability in Section 1.9, Section 3.2 on manifolds, and Section 3.8 on the geometry of curves and surfaces. (In our experience, beginning students do have trouble with the proof of the fundamental theorem of algebra, while manifolds do not pose much of a problem.)

(2) The entire book could also be used for a full year's course. This could be done at different levels of difficulty, depending on the students' sophistication and the pace of the class. Some students may need to review the material in Sections 0.3 and 0.5; others may be able to include some of the proofs in the appendix, such as those of the central limit theorem and the Kantorovitch theorem.

(3) With a year at one's disposal (and excluding the proofs in the appendix), one could also cover more than the present material, and a second volume is planned, covering

applications of differential forms;
abstract vector spaces, inner product spaces, and Fourier series;
electromagnetism;
differential equations;
eigenvalues, eigenvectors, and differential equations.

We favor this third approach; in particular, we feel that the last two topics above are of central importance. Indeed, we feel that three semesters would not be too much to devote to linear algebra, multivariate calculus, differential forms, differential equations, and an introduction to Fourier series and partial differential equations. This is more or less what the engineering and physics departments expect students to learn in second year calculus, although we feel this is unrealistic.

Students who have studied some linear algebra or multivariate calculus

The book can also be used for students who have some exposure to either linear algebra or multivariate calculus, but who are not ready for a course in analysis. We used an earlier version of this text with students who had taken a course in linear algebra, and feel they gained a great deal from seeing how linear algebra and multivariate calculus mesh. Such students could be expected to cover Chapters 1–6, possibly omitting some of the optional material discussed above. For a less fast-paced course, the book could also be covered in an entire year, possibly including some proofs from the appendix.

Students ready for a course in analysis

We view Chapter 0 primarily as a resource for students, rather than as part of the material to be covered in class. An exception is Section 0.4, which might well be covered in a class on analysis.

If the book is used as a text for an analysis course, then in one semester one could hope to cover all six chapters and some or most of the proofs in Appendix A. This could be done at varying levels of difficulty; students might be expected to follow the proofs, for example, or they might be expected to understand them well enough to construct similar proofs. Several exercises in Appendix A and in Section 0.4 are of this nature.

Numbering of theorems, examples, and equations

Mathematical notation is not always uniform. For example, $|A|$ can mean the length of a matrix A (the meaning in this book) or it can mean the determinant of A. Different notations for partial derivatives also exist. This should not pose a problem for readers who begin at the beginning and end at the end, but for those who are using only selected chapters, it could be confusing. Notations used in the book are listed on the front inside cover, along with an indication of where they are first introduced.

Theorems, lemmas, propositions, corollaries, and examples share the same numbering system. For example, Proposition 2.3.8 is not the eighth proposition of Section 2.3; it is the eighth numbered item of that section, and the first numbered item following Example 2.3.7. We often refer back to theorems, examples, and so on, and hope this numbering will make them easier to find.

Figures are numbered independently; Figure 3.2.3 is the third figure of Section 3.2. All displayed equations are numbered, with the numbers given at right; Equation 4.2.3 is the third equation of Section 4.2. When an equation is displayed a second time, it keeps its original number, but the number is in parentheses.

We use the symbol $\triangle$ to mark the end of an example or remark, and the symbol $\square$ to mark the end of a proof.

Exercises

Exercises are given at the end of each chapter, grouped by section. They range from very easy exercises intended to make the student familiar with vocabulary, to quite difficult exercises. The hardest exercises are marked with a star (or, in rare cases, two stars). On occasion, figures and equations are numbered in the exercises. In this case, they are given the number of the exercise to which they pertain.

In addition, there are occasional "mini-exercises" incorporated in the text, with answers given in footnotes. These are straightforward questions containing no tricks or subtleties, and are intended to let the student test his or her understanding (or be reassured that he or she has understood). We hope that even the student who finds them too easy will answer them; working with pen and paper helps vocabulary and techniques sink in.

Web page

Errata will be posted on the web page

http://math.cornell.edu/~hubbard/vectorcalculus.

The three programs given in Appendix B will also be available there. We plan to expand the web page, making the programs available on more platforms, and adding new programs and examples of their uses.

Readers are encouraged to write the authors at jhh8@cornell.edu to signal errors, or to suggest new exercises, which will then be shared with other readers via the web page.

Acknowledgments

Many people contributed to this book. We would in particular like to express our gratitude to Robert Terrell of Cornell University, for his many invaluable suggestions, including ideas for examples and exercises, and advice on notation; Adrien Douady of the University of Paris at Orsay, whose insights shaped our presentation of integrals in Chapter 4; and Régine Douady of the University of Paris-VII, who inspired us to distinguish between points and vectors. We would also like to thank Allen Back, Harriet Hubbard, Peter Papadopol, Birgit Speh, and Vladimir Veselov, for their many contributions.

Cornell undergraduates in Math 221, 223, and 224 showed exemplary patience in dealing with the inevitable shortcomings of an unfinished text in photocopied form. They pointed out numerous errors, and they made the course a pleasure to teach. We would like to thank in particular Allegra Angus, Daniel Bauer, Vadim Grinshpun, Michael Henderson, Tomohiko Ishigami, Victor Kam, Paul Kautz, Kevin Knox, Mikhail Kobyakov, Jie Li, Surachate Limkumnerd, Mon-Jed Liu, Karl Papadantonakis, Marc Ratkovic, Susan Rushmer, Samuel Scarano, Warren Scott, Timothy Slate, and Chan-ho Suh. Another Cornell student, Jon Rosenberger, produced the *Newton* program in Appendix B.1. Karl Papadantonakis helped produce the picture used on the cover.

For insights concerning the history of linear algebra, we are indebted to the essay by J.-L. Dorier in *L'Enseignement de l'algèbre linéaire en question.* Other books that were influential include *Infinitesimal Calculus* by Jean Dieudonné, *Advanced Calculus* by Lynn Loomis and Shlomo Sternberg, and *Calculus on Manifolds* by Michael Spivak.

Ben Salzberg of Blue Sky Research saved us from despair when a new computer refused to print files in Textures. Barbara Beeton of American Mathematical Society's Technical Support gave prompt and helpful answers to technical questions.

We would also like to thank George Lobell at Prentice Hall, who encouraged us to write this book; Nicholas Romanelli for his editing and advice, and

Gale Epps, as well as the mathematicians who served as reviewers for Prentice-Hall and made many helpful suggestions and criticisms: Robert Boyer, Drexel University; Ashwani Kapila, Rensselaer Polytechnic Institute; Krystyna Kuperberg, Auburn University; Ralph Oberste-Vorth, University of South Florida; and Ernest Stitzinger, North Carolina State University. We are of course responsible for any remaining errors, as well as for all our editorial choices.

We are indebted to our son, Alexander, for his suggestions, for writing numerous solutions, and for writing a program to help with the indexing. We thank our oldest daughter, Eleanor, for the goat figure of Section 3.8, and for earning part of her first-year college tuition by reading through the text and pointing out both places where it was not clear and passages she liked—the first invaluable for the text, the second for our morale. With her sisters, Judith and Diana, she also put food on the table when we were to busy to do so. Finally, we thank Diana for discovering errors in the page numbers in the table of contents.

John H. Hubbard
Barbara Burke Hubbard

Ithaca, N.Y.
jhh8@cornell.edu

John H. Hubbard is a professor of mathematics at Cornell University and the author of several books on differential equations. His research mainly concerns complex analysis, differential equations, and dynamical systems. He believes that mathematics research and teaching are activities that enrich each other and should not be separated.

Barbara Burke Hubbard is the author of *The World According to Wavelets*, which was awarded the prix d'Alembert by the French Mathematical Society in 1996.

Contents

0

Preliminaries

0.0 Introduction

This chapter is intended as a resource, providing some background for those who may need it. In Section 0.1 we share some guidelines that in our experience make reading mathematics easier, and discuss a few specific issues like sum notation. Section 0.2 analyzes the rather tricky business of negating mathematical statements. (To a mathematician, the statement "All seven-legged alligators are orange with blue spots" is an obviously true statement, not an obviously meaningless one.) Section 0.3 reviews set theory notation; Section 0.4 discusses the real numbers; Section 0.5 discusses countable and uncountable sets and Russell's paradox; and Section 0.6 discusses complex numbers.

0.1 Reading Mathematics

The most efficient logical order for a subject is usually different from the best psychological order in which to learn it. Much mathematical writing is based too closely on the logical order of deduction in a subject, with too many definitions without, or before, the examples which motivate them, and too many answers before, or without, the questions they address.— William Thurston

We recommend not spending much time on Chapter 0. In particular, if you are studying multivariate calculus for the first time you should definitely skip certain parts of Section 0.4 (Definition 0.4.4 and Proposition 0.4.6). However, Section 0.4 contains a discussion of sequences and series which you may wish to consult when we come to Section 1.5 about convergence and limits, if you find you don't remember the convergence criteria for sequences and series from first year calculus.

Reading mathematics is different from other reading. We think the following guidelines can make it easier. First, keep in mind that there are two parts to understanding a theorem: understanding the statement, and understanding the proof. *The first is more important than the second.*

What if you don't understand the statement? If there's a symbol in the formula you don't understand, perhaps a δ, look to see whether the next line continues, "where δ is such-and-such." In other words, read the whole sentence before you decide you can't understand it. In this book we have tried to define all terms before giving formulas, but we may not have succeeded everywhere.

If you're still having trouble, *skip ahead to examples.* This may contradict what you have been told—that mathematics is sequential, and that you must understand each sentence before going on to the next. In reality, although mathematical writing is necessarily sequential, mathematical understanding is not: you (and the experts) never understand perfectly up to some point and not at all beyond. The "beyond," where understanding is only partial, is an essential part of the motivation and the conceptual background of the "here and now." You may often (perhaps usually) find that when you return to something you left half-understood, it will have become clear in the light of the further things you have studied, even though the further things are themselves obscure.

Many students are very uncomfortable in this state of partial understanding, like a beginning rock climber who wants to be in stable equilibrium at all times. To learn effectively one must be willing to leave the cocoon of equilibrium. So *if you don't understand something perfectly, go on ahead and then circle back.*

In particular, an example will often be easier to follow than a general statement; you can then go back and reconstitute the meaning of the statement in light of the example. Even if you still have trouble with the general statement, you will be ahead of the game if you understand the examples. We feel so strongly about this that we have sometimes flouted mathematical tradition and given examples before the proper definition.

Read with pencil and paper in hand, making up little examples for yourself as you go on.

Some of the difficulty in reading mathematics is notational. A pianist who has to stop and think whether a given note on the staff is A or F will not be able to sight-read a Bach prelude or Schubert sonata. The temptation, when faced with a long, involved equation, may be to give up. You need to take the time to identify the "notes."

Learn the names of Greek letters—not just the obvious ones like alpha, beta, and pi, but the more obscure psi, xi, tau, omega. The authors know a mathematician who calls all Greek letters "xi," (ξ) except for omega (ω), which he calls "w." This leads to confusion. Learn not just to recognize these letters, but how to pronounce them. Even if you are not reading mathematics out loud, it is hard to think about formulas if $\xi, \psi, \tau, \omega, \varphi$ are all "squiggles" to you.

The Greek Alphabet

Greek letters that look like Roman letters are not used as mathematical symbols; for example, A is capital a, not capital α. The letter χ is pronounced "kye," to rhyme with "sky"; φ, ψ and ξ may rhyme with either "sky" or "tea."

α	A	alpha
β	B	beta
γ	Γ	gamma
δ	Δ	delta
ϵ	E	epsilon
ζ	Z	zeta
η	H	eta
θ	Θ	theta
ι	I	iota
κ	K	kappa
λ	Λ	lambda
μ	M	mu
ν	N	nu
ξ	Ξ	xi
o	O	omicron
π	Π	pi
ρ	P	rho
σ	Σ	sigma
τ	T	tau
υ	Υ	upsilon
ϕ, φ	Φ	phi
χ	X	chi
ψ	Ψ	psi
ω	Ω	omega

Sum and product notation

Sum notation can be confusing at first; we are accustomed to reading in one dimension, from left to right, but something like

$$\sum_{k=1}^{n} a_{i,k}b_{k,j} \qquad 0.1.1$$

requires what we might call two-dimensional (or even three-dimensional) thinking. It may help at first to translate a sum into a linear expression:

$$\sum_{i=0}^{\infty} 2^i = 2^0 + 2^1 + 2^2 \dots \qquad 0.1.2$$

or

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k}b_{k,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,n}b_{n,j}. \qquad 0.1.3$$

In Equation 0.1.3, the symbol $\sum_{k=1}^{n}$ says that the sum will have n terms. Since the expression being summed is $a_{i,k}b_{k,j}$, each of those n terms will have the form ab.

Two $\sum$ placed side by side do not denote the product of two sums; one sum is used to talk about one index, the other about another. The same thing could be written with one $\sum$, with information about both indices underneath. For example,

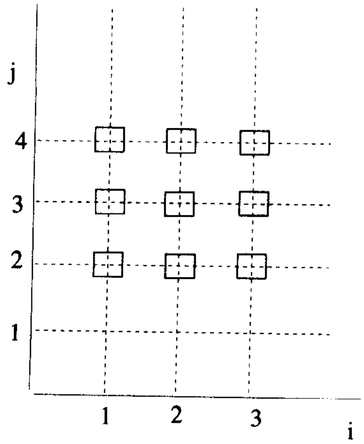


FIGURE 0.1.1.
In the double sum of Equation 0.1.4, each sum has three terms, so the double sum has nine terms.

$$\begin{aligned}\sum_{i=1}^{3}\sum_{j=2}^{4}(i+j) &= \sum_{\substack{i \text{ from } 1 \text{ to } 3,\\ j \text{ from } 2 \text{ to } 4}} (i+j)\\ &= \left(\sum_{j=2}^{4} 1+j\right) + \left(\sum_{j=2}^{4} 2+j\right) + \left(\sum_{j=2}^{4} 3+j\right) \qquad 0.1.4\\ &= \big((1+2)+(1+3)+(1+4)\big)\\ &\quad + \big((2+2)+(2+3)+(2+4)\big)\\ &\quad + \big((3+2)+(3+3)+(3+4)\big);\end{aligned}$$

this double sum is illustrated in Figure 0.1.1.

Rules for product notation are analogous to those for sum notation:

$$\prod_{i=1}^{n} a_i = a_1 \cdot a_2 \cdots a_n; \quad \text{for example,} \quad \prod_{i=1}^{n} i = n!.$$

Proofs

We said earlier that it is more important to understand a mathematical statement than to understand its proof. We have put some of the harder proofs in the appendix; these can safely be skipped by a student studying multivariate calculus for the first time. We urge you, however, to read the proofs in the main text. By reading many proofs you will learn what a proof is, so that (for one thing) you will know when you have proved something and when you have not.

When Jacobi complained that Gauss's proofs appeared unmotivated, Gauss is said to have answered, *You build the building and remove the scaffolding.* Our sympathy is with Jacobi's reply: he likened Gauss to *the fox who erases his tracks in the sand with his tail.*

In addition, a good proof doesn't just convince you that something is true; it tells you why it is true. You presumably don't lie awake at night worrying about the truth of the statements in this or any other math textbook. (This is known as "proof by eminent authority"; you assume the authors know what they are talking about.) But reading the proofs will help you understand the material.

If you get discouraged, keep in mind that the content of this book represents a cleaned-up version of many false starts. For example, John Hubbard started by trying to prove Fubini's theorem in the form presented in Equation 4.5.1. When he failed, he realized (something he had known and forgotten) that the statement was in fact false. He then went through a stack of scrap paper before coming up with a correct proof. *Other statements in the book represent the efforts of some of the world's best mathematicians over many years.*

0.2 How to negate mathematical statements

Even professional mathematicians have to be careful not to get confused when negating a complicated mathematical statement. The rules to follow are:

(1) The opposite of

$$\text{[For all } x, P(x) \text{ is true]}$$
$$\text{is [There exists } x \text{ for which } P(x) \text{ is not true].} \qquad 0.2.1$$

Above, P stands for "property." Symbolically the same sentence is written:

$$\text{The opposite of } \forall x, P(x) \text{ is } \exists x |\text{ not } P(x). \qquad 0.2.2$$

Instead of using the bar | to mean "such that" we could write the last line $(\exists x)(\text{not } P(x))$. Sometimes (not in this book) the symbols $\sim$ and $\neg$ are used to mean "not."

(2) The opposite of

$$\text{[There exists } x \text{ for which } R(x) \text{ is true]}$$
$$\text{is [For all } x, R(x) \text{ is not true].} \qquad 0.2.3$$

Symbolically the same sentence is written:

$$\text{The opposite of } (\exists x)(P(x)) \text{ is } (\forall x) \text{ not } P(x). \qquad 0.2.4$$

These rules may seem reasonable and simple. Clearly the opposite of the (false) statement, "All rational numbers equal 1," is the statement, "There exists a rational number that does not equal 1."

However, by the same rules, the statement, "All seven-legged alligators are orange with blue spots" is true, since if it were false, then there would exist a seven-legged alligator that is not orange with blue spots. The statement, "All seven-legged alligators are black with white stripes" is equally true.

Statements that to the ordinary mortal are false or meaningless are thus accepted as true by mathematicians; if you object, the mathematician will retort, "find me a counter-example."

In addition, mathematical statements are rarely as simple as "All rational numbers equal 1." Often there are many quantifiers and even the experts have to watch out. At a lecture attended by one of the authors, it was not clear to the audience in what order the lecturer was taking the quantifiers; when he was forced to write down a precise statement, he discovered that he didn't know what he meant and the lecture fell apart.

Here is an example where the order of quantifiers really counts: in the definitions of continuity and uniform continuity. A function f is *continuous* if for all x, and for all $\epsilon > 0$, there exists $\delta > 0$ such that for all y, if $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$. That is, f is continuous if

$$(\forall x)(\forall \epsilon > 0)(\exists \delta > 0)(\forall y)\left(|x - y| < \delta \implies |f(x) - f(y)| < \epsilon\right). \qquad 0.2.5$$

A function f is *uniformly continuous* if for all $\epsilon > 0$, there exists $\delta > 0$ for all x and all y such that if $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$. That is, f is uniformly continuous if

$$(\forall \epsilon > 0)(\exists \delta > 0)(\forall x)(\forall y)\left(|x - y| < \delta \implies |f(x) - f(y)| < \epsilon\right). \qquad 0.2.6$$

For the continuous function, we can choose *different* δ for different x; for the uniformly continuous function, we start with ϵ and have to find a *single* δ that works for all x.

For example, the function $f(x) = x^2$ is continuous but not uniformly continuous: as you choose bigger and bigger x, you will need a smaller δ if you want the statement $|x - y| < \delta$ to imply $|f(x) - f(y)| < \epsilon$, because the function keeps climbing more and more steeply. But $\sin x$ is uniformly continuous; you can find one δ that works for all x and all y.

0.3 Set Theory

At the level at which we are working, set theory is a language, with a vocabulary consisting of seven words. In the late 1960's and early 1970's, under the sway of the "New Math," they were a standard part of the elementary school curriculum, and set theory was taught as a subject in its own right. This was a resounding failure, possibly because many teachers, understandably not knowing why set theory was being taught at all, made mountains out of molehills. As a result the schools (elementary, middle, high) have often gone to the opposite extreme, and some have dropped the subject altogether.

The seven vocabulary words are

There is nothing new about the concept of "set" denoted by $\{a|p(a)\}$. Euclid spoke of geometric *loci*, a locus being the set of points defined by some property. (The Latin word *locus* means "place"; its plural is *loci*.)

$\in$	"is an element of"
$\{a \mid p(a)\}$	"the set of a such that $p(a)$ is true"
$\subset$	"is a subset of" (or equals, when $A \subset A$)
$\cap$	"intersect": $A \cap B$ is the set of elements of both A and B.
$\cup$	"union": $A \cup B$ is the set of elements of either A or B or both.
$\times$	"cross": $A \times B$ is the set of pairs (a, b) with $a \in A$ and $b \in B$.
$-$	"complement": $A - B$ is the set of elements in A that are not in B.

One set has a standard name: the empty set ϕ, which has no elements. There are also sets of numbers that have standard names; they are written in *black-board bold*, a font we use only for these sets. Throughout this book and most other mathematics books (with the exception of $\mathbb{N}$, as noted in the margin below), they have exactly the same meaning: