Second Edition

Handbook of Writing for the Mathematical Sciences

NICHOLAS J. HIGHAM

siam

Second Edition

Handbook of Writing for the Mathematical Sciences

NICHOLAS J. HIGHAM
University of Manchester
Manchester, England

SIAM Society for Industrial and Applied Mathematics
Philadelphia

10 9 8 7 6 5

Library of Congress Cataloging-in-Publication Data

Higham, Nicholas J., 1961-
Handbook of writing for the mathematical sciences / Nicholas J. Higham. -- 2nd ed.
Includes bibliographical references and indexes.
ISBN-10: 0-89871-420-6
ISBN-13: 978-0-898714-20-3
1. Mathematics--Authorship. 2. Technical writing. I. Title.
QA42.H54 1998
808'.06651--dc21 98-7284

Contents

Preface to the Second Edition

In the five years since the first edition of this book was published I have received numerous email messages and letters from readers commenting on the book and suggesting how it could be improved. I have also built up a large file of ideas based on my own experiences in reading, writing, and editing and in examining and supervising theses. With the aid of all this information I have completely revised the book. The most obvious changes in this second edition are the new chapters.

- *Writing and Defending a Thesis.* Since many of the readers of the book are graduate students, advice on how to write a thesis and how to handle the thesis defence was a natural addition.

- *Giving a Talk.* The revised chapter "Writing a Talk" from the first edition gives advice on preparing slides for a talk. The new chapter explains how to deliver a talk in front of an audience.

- *Preparing a Poster.* The poster is growing in popularity as a medium of communication at conferences and elsewhere, yet many of us have little experience of preparing posters.

- *TeX and LaTeX.* Since the first edition of this book was published, LaTeX 2_ε has become the official version of LaTeX, thereby solving many of the problems involving, for example, incompatible dialects of LaTeX, font handling, and inclusion of PostScript figures in a LaTeX document. I have moved the discussion of TeX, LaTeX, and their associated tools to a new chapter. Many more tips on the use of TeX and LaTeX for typesetting mathematics are now given, and the discussions of BibTeX and indexing have been expanded. The many mathematical symbols in the AMS fonts have been added to Appendix B ("Summary of TeX and LaTeX Symbols").

Among the new material in existing chapters, the section "How to Referee" in the chapter "Publishing a Paper" offers advice on this important aspect of the publication process, and in the chapter "Writing a Paper" suggested formats are given for referencing items on the World Wide Web.

The renamed chapter "Aids and Resources for Writing and Research" contains a new section "Library Classification Schemes", which should help readers to find their way around libraries. The material on the Internet in this chapter has been completely rewritten in the light of the World Wide Web (which was not mentioned in the first edition). I have minimized the number of URLs (Web addresses) given, for two reasons. First, URLs can and do change over time. Second, if you want to know more about almost any subject mentioned in the book, just choose appropriate key words (e.g., "mathematical writing", "Roget's Thesaurus", or "Society of Indexers") and invoke your favourite search engine from your Web browser. There is a good chance that you will find the information, or particular Web pages, that you are looking for.

The subject of mathematical writing can easily become dull and boring, so to liven it up I like to include anecdotes, unusual paper titles, humorous quotations, and so on. The new edition contains many more of these.

Separate author and subject indexes are now provided. The author index removes some clutter from the subject index, and you can use it to find where references in the bibliography are discussed.

The bibliography has been updated. Many new editions of books are referenced and over 70 new references have been added.

A Web page has been created for the book, at

`http://www.siam.org/books/ot63/`

It includes

- Updates relating to material in the book.
- Links to references in the bibliography that are available on the Web.
- Links to other Web pages related to mathematical writing, LaTeX, BibTeX, etc.
- Links to Web pages giving examples of posters.
- The bibliography for the book in BibTeX form, which is also available from Bibnet as `han-wri-mat-sci.bib`.

Several people helped with the second edition by reading and commenting on drafts:

David Abrahams, Henri Casanova, Bobby Cheng, Tony Cox, Des Higham, Doris Higham, Nil Mackey, Alicia Roca, Françoise Tisseur, Nick Trefethen, Joan Walsh, Barry White.

Working with the SIAM staff was once again a pleasure. I thank, in particular, my copy editor Beth Gallagher, Vickie Kearn and Mary Rose Muccie.

This book was typeset in LaTeX using the `book` document style and various LaTeX packages. The references were prepared in BibTeX and the index with MakeIndex. I used software from the emTeX distribution, running on a Pentium workstation. I used text editors The Semware Editor (Semware Corporation) and GNU Emacs (Free Software Foundation) and checked spelling with PC-Write (Quicksoft).

Manchester
January 1998

Nicholas J. Higham

Preface to the First Edition

In this book I aim to describe most of what a scientist needs to know about mathematical writing. Although the emphasis is strongly on *mathematical* writing, many of the points and issues I discuss are relevant to *scientific* writing in general. My main target audience is graduate students. They often have little experience or knowledge of technical writing and are daunted by the task of writing a report or thesis. The advice given here reflects what I have learned in the ten years since I wrote my first research report as a graduate student and describes what I would have liked to know as I started to write that first report. I hope that as well as being a valuable resource for graduate students, this book will also be of use to practising scientists.

The book has grown out of notes for a short lecture course on mathematical writing that I gave at the University of Manchester in May 1992. As I prepared the course I realized that, although several excellent articles and books on mathematical writing are available (notably those by Halmos (1970) [121], Gillman (1987) [104] and Knuth, Larrabee and Roberts (1989) [164]), none functions as a comprehensive handbook that can be both read sequentially and used as a reference when questions about mathematical writing arise. I have attempted to provide such a handbook. (I hope that the comment of one journal referee, "This paper fills a much needed gap in the literature", is not applicable to this book!)

As well as covering standard topics such as English usage, the anatomy of a research paper, and revising a draft, I examine in detail four topics that are usually discussed only briefly, if at all, in books on technical writing.

- The whole publication process, from submission of a manuscript to its appearance as a paper in a journal.
- Writing when English is a foreign language.
- How to write slides for a talk.
- The use of computers in writing and research. In particular, I discuss modern practices such as computerized typesetting in TeX, the use of computer tools for indexing and checking spelling and style, and electronic mail and ftp (file transfer protocol).

An important feature of the book is that many examples are given to illustrate the ideas and principles discussed. In particular, Chapter 7 contains a collection of extracts from the mathematics and computer science literature, with detailed comments on how each extract can be improved.

In writing the book I have been helped and influenced by many people. Several people read the entire manuscript at one or more of its various stages, offered constructive suggestions, encouragement and advice, and made sure I said what I meant and meant what I said. They are

> Ian Gladwell, Des Higham, Doris Higham, Nil Mackey, Fred Schneider, Pete (G. W.) Stewart and Nick Trefethen.

Other people who have read portions of the book and have given help, suggestions or advice are

> Carl de Boor, David Carlisle, Valérie Frayssé, Paul Halmos, Bo Kågström, Philip Knight, Sven Leyffer, Steve Mackey, Volker Mehrmann, June O'Brien, Pythagoras Papadimitriou, Beresford Parlett, Nigel Ray, Stephan Rudolfer, Bob Sandling, Zdenek Strakos, Gil Strang, Charlie Van Loan, Rossana Vermiglio, Joan Walsh, Barry White, and Yuanjing Xu.

I thank all these people, together with many others who have answered my questions and made suggestions. In researching the contents of the book I was inspired by many of the references listed in the bibliography, and learned a lot from them. I acknowledge the Nuffield Foundation for the support of a Nuffield Science Research Fellowship, during the tenure of which I wrote much of the book. Last, but not least, I thank the SIAM staff for their help and advice in the production of the book—in particular, Susan Ciambrano, Beth Gallagher and Tricia Manning. I would be happy to receive comments, notification of errors, and suggestions for improvement, which I will collect for inclusion in a possible second edition.

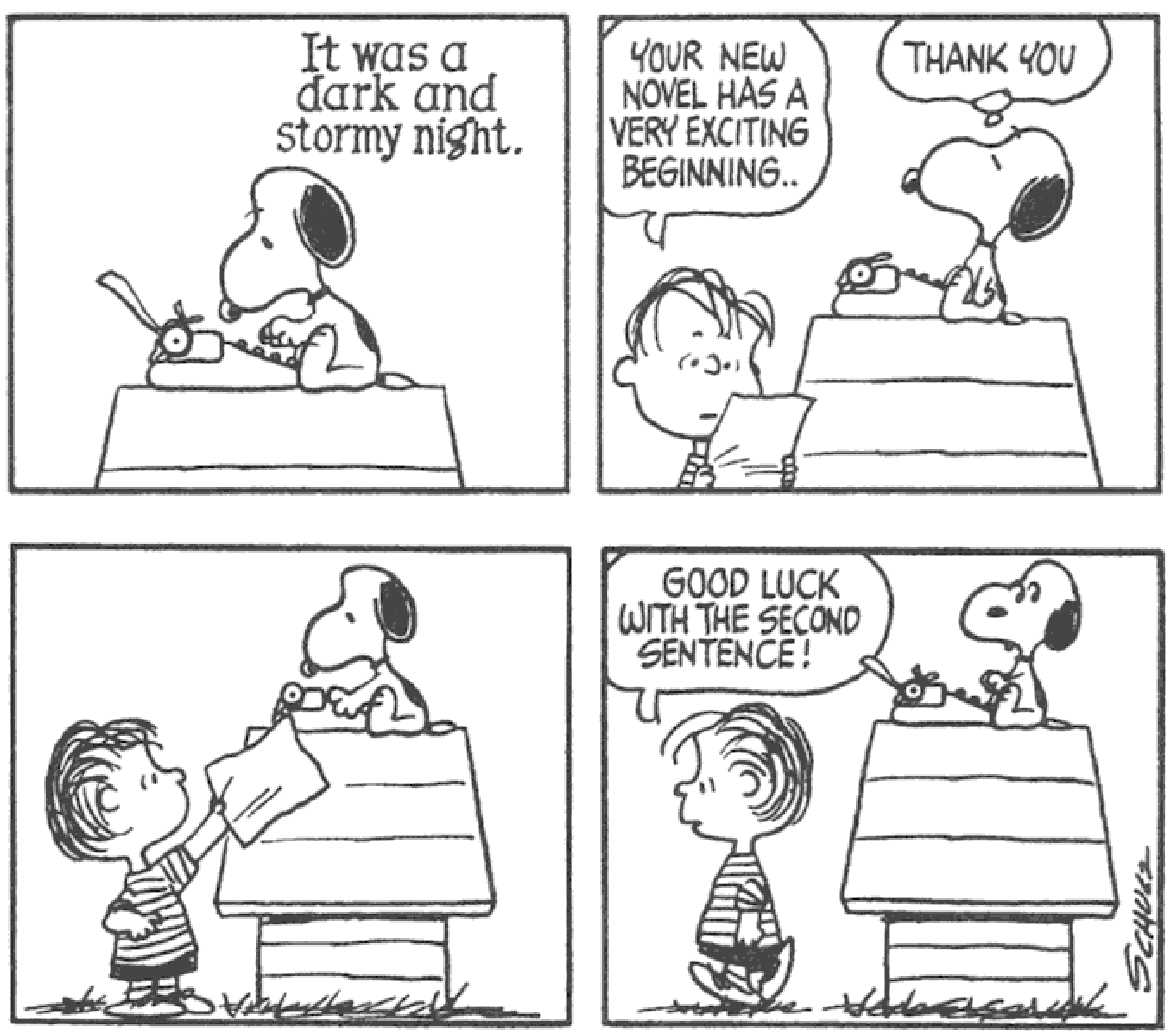
This book was typeset in LaTeX using the book document style with G. W. Stewart's jeep style option. I prepared the references with BibTeX and the index with MakeIndex. I used text editors Qedit (Semware) and GNU Emacs (Free Software Foundation), and checked spelling with PC-Write (Quicksoft).

Manchester
December 1992

Nicholas J. Higham

Chapter 1

General Principles



Peanuts reprinted by permission of United Feature Syndicate.

Good writing . . . is clear thinking made visible.
— AMBROSE BIERCE, *Write it Right: A Little Blacklist of Literary Faults* (1937)

A writer needs three qualities: creativity, originality, clarity and a good short term memory.
— DESMOND J. HIGHAM, *More Commandments of Good Writing* (1992)

Writing helps you to learn. Writing is not simply a task to be done once research or other preparation is completed—it can be an integral part of the work process. Baker explains it well [13]:

> In writing, you clarify your own thoughts, and strengthen your conviction. Indeed, you probably grasp your thoughts for the first time. Writing is a way of thinking. Writing actually creates thought, and generates your ability to think: you discover thoughts you hardly knew you had, and come to know what you know. You learn as you write.

Writing brings out gaps in your understanding, by forcing you to focus on steps in your thinking that you might otherwise skip. (In this respect, writing is similar to explaining your ideas to a colleague.) Writing down partial research results as your work progresses can reveal structure and suggest further areas of development. Zinsser has written a delightful book that explores the idea of writing to learn [303].

Good writing reflects clear thinking. It is very hard for a woolly thinker to produce clear prose. Clear thinking leads to good organization, a vital ingredient of technical writing. A well-organized piece is much easier to write than a badly organized one. If you find a particular piece difficult to write it may be because you have not found the right structure to express your ideas.

Writing is difficult! Zinsser [304] says "It's one of the hardest things that people do." It is often difficult to get started. One solution, which works for some, is to force yourself to write something, however clumsy it may be, for it is often easier to modify something you have written previously than to compose from scratch.

The most fundamental tenet of technical writing is to keep your prose simple and direct. Much of written English is unnecessarily complicated. In writing up your research you are aiming at a relatively small audience, so it is important not to alienate part of it with long-winded or imprecise text. English may not be the first language of many of your readers—they, particularly, will appreciate plain writing. Aim for economy of words. Early drafts can usually be reduced in length substantially with consequent improvements in readability (see Chapter 7).

Probably the best way to improve your writing skills is to receive, and learn from, constructive criticism. Ask a colleague to read and comment on your writing. Another reader can often find errors and suggest improvements that you miss because of your familiarity with the work. Criticism can be difficult to take, but it should be welcomed; it is a privilege to have someone else take the time to comment on your writing.

Another way to improve your writing is to read as much as you can, always with a critical eye. In particular, I recommend perusal of some of the following mathematical books. They are by excellent writers, each of whom has his own distinctive style (this selection is inevitably biased towards my own area of research):

▷ Forman S. Acton (1970), *Numerical Methods That Work* [3].

▷ Albert H. Beiler (1966), *Recreations in the Theory of Numbers* [19].

▷ David M. Burton (1980), *Elementary Number Theory* [44].

▷ Gene H. Golub and Charles F. Van Loan (1996), *Matrix Computations* [108].

▷ Paul R. Halmos (1982), *A Hilbert Space Problem Book* [125].

▷ Donald E. Knuth (1973–1981), *The Art of Computer Programming* [157]. (Knuth was awarded the 1986 Leroy P. Steele Prize by the AMS for these three volumes.)

▷ Beresford N. Parlett (1998), *The Symmetric Eigenvalue Problem* [217].

▷ G. W. Stewart (1973), *Introduction to Matrix Computations* [261].

▷ Gilbert Strang (1986), *Introduction to Applied Mathematics* [262].

Also worth studying are papers or books that have won prizes for expository writing in mathematics. Appendix E lists winners of the Chauvenet Prize, the Lester R. Ford Award, the George Polya Award, the Carl B. Allendoerfer Award, the Beckenbach Book Prize and the Merten M. Hasse Prize.

Chapter 2
Writer's Tools and Recommended Reading

I use three dictionaries almost every day.
— JAMES A. MICHENER, *Writer's Handbook* (1992)

The purpose of an ordinary dictionary is simply
to explain the meaning of the words ....
The object aimed at in the present undertaking is exactly the converse of this:
namely,—The idea being given, to find the word, or words,
by which that idea may be most fitly and aptly expressed.
— PETER MARK ROGET, *Thesaurus of English Words and Phrases* (1852)

The dictionary and thesaurus interruptions are usually
not about meaning in the gross sense
(what's the correct use of "oppugn"),
but about precision, and about finding the right word...
What did the examples that von Neumann and I constructed
do to the conjugacy conjecture for shifts...
did they contradict, contravene, gainsay, dispute,
disaffirm, disallow, abnegate, or repudiate it?...
Writing can stop for 10 or 15 minutes while I search and weigh.
— PAUL R. HALMOS, *I Want to be a Mathematician: An Automathography in Three Parts* (1985)

Mathematicean. *One that is skilled in Augurie, Geometrie, and Astronomie.*
— HENRY COCKERAM[1], *English Dictionarie* (1623)

[1]Quoted in [255].

2.1. Dictionaries and Thesauruses

Apart from pen, paper and keyboard, the most valuable tool for a writer in any subject is a dictionary. Writing or reading in the mathematical sciences you will come across questions such as the following:

1. What is the plural of modulus: moduli or moduluses?
2. Which of parameterize and parametrize is the correct spelling?
3. What is a gigaflop?
4. When was the mathematician Abel born and what was his nationality?
5. What is the meaning of mutatis mutandis?
6. Who was Procrustes (as in the "orthogonal Procrustes problem")?
7. When should you use "special" and when "especial"?
8. What are the differences between mind-bending, mind-blowing and mind-boggling?

All the answers can be found in general-purpose dictionaries (and are given at the end of this chapter). As these questions illustrate, dictionaries are invaluable for choosing a word with just the right shade of meaning, checking on spelling and usage, and even finding encyclopedic information. Furthermore, the information about a word's history provided in a dictionary etymology can make it easier to use the word precisely.

The most authoritative dictionary is the *Oxford English Dictionary* (OED) [215]. It was originally published in parts between 1884 and 1928, and a four volume supplement was produced from 1972–1986. A twenty volume second edition of the dictionary was published in 1989; it defines more than half a million words, using 2.4 million illustrative quotations. The OED traces the history of words from around 1150. In 1992 a compact disc (CD-ROM) version of the OED was published. It contains the full text of the printed version (at about a third of the price) and the accompanying software includes powerful search facilities. Other large dictionaries are *Webster's Third New International Dictionary* [294], which was published in the United States in 1961 and has had three supplements, *The American Heritage Dictionary of the English Language* [7], the *Random House Unabridged Dictionary* [233], and *The New Shorter Oxford English Dictionary*, in two volumes [214].

For everyday use the large dictionaries are too unwieldy and too thorough, so a more concise dictionary is needed. *The Concise Oxford Dictionary* (COD) [213] is now in its ninth edition (1995). It is the favourite of many, and is suitable for American use, as American spellings and usages are included. (The COD was my main dictionary of reference in writing this book.) Other dictionaries suitable for regular use by the writer include, from the United States:

- *The American Heritage College Dictionary* [6].
- *The Random House Webster's College Dictionary* [234].
- *Merriam-Webster's Collegiate Dictionary* [203]. Most main entries state the date of first recorded use of the word. Contains usage and synonym notes and appendices "Biographical Names" and "Geographical Names".
- *Webster's New World College Dictionary* [293].

From Britain:

- *The Chambers Dictionary* [54]. Renowned for its rich vocabulary, which includes literary terms, Scottish words and many archaic and obsolete words. Also contains some humorous entries: *éclair* is defined as "a cake, long in shape but short in duration ...".
- The *Collins English Dictionary* [60]. Contains extensive encyclopedic entries, both biographical and geographical, strong coverage of scientific and technical vocabulary, and usage notes.
- The *Longman Dictionary of the English Language* [182]. The same comments apply as for the Collins. Has an extensive collection of notes on usage, synonyms and word history.

The American dictionaries listed, but not the British ones, show allowable places to divide words when they must be broken and hyphenated at the end of a line.

To make good use of dictionaries, it helps to be aware of some of their characteristics.

Order of definitions. For words with several meanings, most dictionaries give the most common or current meanings first, but some give meanings in their historical sequence. The historical order is the one used by the *Oxford English Dictionary*, since its purpose is to trace the development of words from their first use to the present day. The *Merriam-Webster's Collegiate* also uses the historical order, but for a desk dictionary intended for quick reference this order can be disorienting. For example, under the headword *nice*, *Merriam-Webster's Collegiate* lists "showing fastidious or finicky tastes" before "pleasing, agreeable".

Etymologies. Etymologies vary in their location within an entry, in the style in which they are presented (for example, the symbol < may be used for "from"), and in their depth and amount of detail. Some words with interesting etymologies are *diploma*, *OK*, *shambles*, *symposium*, and *sine*.

Scientific and technical vocabulary. Since there are vastly more scientific and technical terms than any general dictionary can accommodate, there is much variation in the coverage provided by different dictionaries.

Up-to-date vocabulary. The constantly changing English language is monitored by lexicographers (Johnson's "harmless drudges"), who add new words and meanings to each new edition of their dictionaries. Coverage of modern vocabulary varies between dictionaries, depending on the year of publication and the compilers' tastes and citation files (which usually include material submitted by the general public).

British versus American spelling and usage. Since much mathematical science is written for an international audience it is useful to be able to check differences in British and American spelling and usage. Most British and American dictionaries are good in this respect.

General-purpose dictionaries do not always give correct definitions of mathematical terms. In a comparison of eight major British and American dictionaries I found errors in definitions of terms such as *determinant*, *eigenvector*[2], *polynomial*, and *power series* [141].

Annotated lists of dictionaries and usage guides are given by Stainton [253], [254]. Comparisons and analyses of dictionaries are also given by Quirk and Stein [232, Chap. 11] and Burchfield [43].

Specialized dictionaries can also be useful to the mathematical writer. There are many dictionaries of mathematics, one example being the Penguin dictionary [206], which is small and inexpensive yet surprisingly thorough. Schwartzman's *The Words of Mathematics* [247] explains the etymology of words used in mathematics (see also [248]).

The synonyms provided in a thesaurus can be helpful in your search for an elusive word or a word with the right connotation. *Roget's Thesaurus*, first published in 1852, is the classic one. The words in *Roget's Thesaurus* are traditionally arranged according to the ideas they express, instead of alphabetically, though versions are now available in dictionary form. The *Bloomsbury Thesaurus* [32] is arranged according to a new classification

[2]One dictionary offers this definition of eigenvector: a vector that in one dimension under a given rotational, reflectional, expanding, or shrinking operation becomes a number that is a multiple of itself.

designed to be more appropriate for modern English than that of Roget, and it has a very detailed index. Rodale's *The Synonym Finder* [236] is a large thesaurus arranged alphabetically. Thesauruses are produced by all the major publishers of dictionaries.

2.2. Usage and Style Guides

Every writer should own and read a guide to English usage. One of the most accessible is *The Elements of Style* by Strunk and White [263]. Zinsser [304] says this is "a book that every writer should read at least once a year", and, as if following this advice, Luey [185] says "I read it once a year without fail." An even shorter, but equally readable, guide is Lambuth et al.'s *The Golden Book on Writing* [170]. Fowler's *Dictionary of Modern English Usage* [83] is a much longer and more detailed work, as is its predecessor, *The King's English*, by the Fowler brothers [84]. A favourite of mine is the revision [298] by Flavell and Flavell of the 1962 *Current English Usage* by Wood. Gowers's influential *Complete Plain Words* [115] stems from his *Plain Words* of 1948, which was written to improve the use of English in the British civil service. Partridge's *Usage and Abusage* [218] is another valuable guide, this one in dictionary form.

Excellent advice on punctuation is given by Carey in *Mind the Stop* [52] and by Bernstein [28]. For a whimsical treatment, see *The New Well-Tempered Sentence* by Gordon [112].

Bryson's *Dictionary of Troublesome Words* [41] offers practical, witty advice on usage, while Safire [243] presents fifty "fumblerules" (mistakes that call attention to the rules) accompanied by pithy explanatory essays. The books *On Newspaper Style* and *English our English* by Waterhouse [287], [288] make fascinating and informative reading, though they are hard to use for reference since they lack an index; [287] is a standard handbook for journalists, but is of much wider interest. Baker's *The Practical Stylist* [13] is a widely used course text on writing; it has thorough discussions of usage, style and revision and gives many illustrative examples. Day's *Scientific English* [69] contains general advice on grammar and usage, with particular reference to English in scientific writing. Perry's *The Fine Art of Technical Writing* [221] offers selective, practical advice on the psychology, artistry and technique of technical writing, which the author defines as "all writing other than fiction". In *Miss Thistlebottom's Hobgoblins* [26] Bernstein provides an antidote for those brainwashed by over-prescriptive usage guides, in the form of letters to his (fictional) English schoolteacher. Two other books by Bernstein, *The Careful Writer* [25] and *Dos, Don'ts and Maybes of English Usage* [27], are also useful guides. Gordon's *The Transitive Vampire* [111] is a grammar guide in the same fanciful vein as [112].

The Chicago Manual of Style [58], first published in 1906, is a long and comprehensive guide to book production, style and printing. It is the standard reference for authors and editors in many organizations. It includes chapters on typesetting mathematics and preparing bibliographies and indexes. Turabian's *A Manual for Writers of Term Papers, Theses, and Dissertations* [278], first published in 1937, is based on the guidelines in *The Chicago Manual of Style* but its aim is more limited, as defined in the title, so it does not discuss bookmaking and copy editing. *Words into Type* [249] is another thorough guide for authors and editors, covering manuscript and index preparation, copy editing style, grammar, typographical style and the printing process. Other valuable references on editing, copy editing and proofreading are *Hart's Rules* [131], which describes the house style of Oxford University Press; Butcher's *Copy-Editing* [45], which is regarded as the standard British work on copy editing; Eisenberg's *Guide to Technical Editing* [77]; O'Connor's *How to Copyedit Scientific Books and Journals* [208]; Stainton's *The Fine Art of Copyediting* [254]; and Tarutz's *Technical Editing* [270].

Some interesting techniques for revising a sentence by analysing its structure are presented by Lanham in *Revising Prose* [175].

2.3. Technical Writing Guides

Several guides to mathematical writing are available. Halmos's essay "How to Write Mathematics" [121] is essential reading for every mathematician; it contains much sound advice not found elsewhere. Halmos's "automathography" [127] includes insight into mathematical writing, editing and refereeing; it begins with the sentence "I like words more than numbers, and I always did." Transcripts of a lecture course called "Mathematical Writing" that was given by Knuth in 1987 at Stanford are collected in *Mathematical Writing* [164], which I highly recommend. This manual contains many anecdotes and insights related by Knuth and his guest lecturers, including Knuth's battle with the copy editors at *Scientific American* and his experiences in writing the book *Concrete Mathematics* [116]. Other very useful guides are Flanders's article [80] for authors who write in the journal *American Mathematical Monthly*; Gillman's booklet *Writing Mathematics Well* [104] on preparing manuscripts for Mathematical Association of America journals; Steenrod's essay "How to Write Mathematics" [256]; Krantz's wide-ranging *A Primer of Mathematical Writing* [167]; and Swanson's guide *Mathematics into Type* [267] for mathematical copy editors and authors. Knuth's book on TeX [161] contains much general advice on how to typeset mathematics, and an old guide to this subject is *The Printing of Mathematics* [55].

Most books and papers on mathematical writing, including this one, are aimed primarily at graduate students and advanced undergraduate students. Maurer [197] gives advice on mathematical writing aimed specifically at undergraduate students, covering a number of basic issues omitted elsewhere.

Guides to writing in other scientific disciplines often contain much that is relevant to the mathematical writer; an example is the book by Pechenik [219], which is aimed at biology students. General guides to scientific writing that I recommend are those by Barrass [14], [15], Cooper [62], Ebel, Bliefert and Russey [76], Kirkman [153], O'Connor [209] (this is a revised and extended version of an earlier book by O'Connor and Woodford [210]), and Turk and Kirkman [280]. The book edited by Woodford [300] contains three examples of short papers in both original and revised forms, with detailed annotations. Particularly informative and pleasant to read are Booth's *Communicating in Science* [36] and Day's *How to Write and Publish a Scientific Paper* [68].

The journal *IEEE Transactions on Professional Communication* publishes papers on many aspects of technical communication, including how to write papers and give talks. A selection of 63 papers from this and other journals is collected in *Writing and Speaking in the Technology Professions: A Practical Guide* [18].

How to Do It [180] contains 47 chapters that give advice for medical doctors, but many of them are of general interest to scientists. Chapter titles include "Write a Paper", "Referee a Paper", "Attract the Reader", "Review a Book", "Use an Overhead Projector", and "Apply for a Research Grant". Many of the chapters originally appeared in the *British Medical Journal*.

Van Leunen's *A Handbook for Scholars* [283] is a unique and indispensable guide to the mechanics of scholarly writing, covering reference lists, quotations, citations, footnotes and style. This is the place to look if you want to know how to prepare a difficult reference or quotation (what date to list for a reprint of a work from a previous century, or how to punctuate a quotation placed in mid-sentence). There is also an appendix on how to prepare a CV. Luey's *Handbook for Academic Authors* [185] offers much useful advice to the writer of an academic book.

O'Connor has written a book about how to edit and manage scientific books and journals [207].

Thirty-one essays discussing how writing is being used to teach mathematics in undergraduate courses are contained in *Using Writing to Teach Mathematics* [259].

A useful source for examples of expository mathematical writing is the annotated bibliography of Gaffney and Steen [87], which contains more than 1100 entries.

Finally, Pemberton's book *How to Find Out in Mathematics* [220] tells you precisely what the title suggests. It includes information on mathematical dictionaries (including interlingual ones) and encyclopedias, mathematical histories and biographies, and mathematical societies, periodicals and abstracts. Although it appeared in 1969, the book is still worth consulting.

2.4. General Reading

The three books by Zinsser [302], [303], [304] are highly recommended; all are informative and beautifully written. In *Writing with a Word Processor* [302] Zinsser summarizes his experience in moving to a computer from his trusty typewriter. His book *Writing to Learn* contains chapters on "Writing Mathematics" and "Writing Physics and Chemistry"; they explain how writing can be used in the teaching of these subjects and give examples of good writing. Michener's *Writer's Handbook* [204] provides insight into how this prodigious writer worked. The reader is led through the development of parts of two of Michener's books (one fiction, one nonfiction), from early drafts to proofs to the published versions.

Mitchell [205] gives hints on writing with a computer, with good examples of how to revise drafts.

Valuable insight into the English language—its history, its eccentricities, and its uses—is provided by Bryson [42], Crystal [66] and Potter [229].

Answers to the Questions at the Start of the Chapter

1. The plural of modulus is *moduli.*

2. The *Concise Oxford Dictionary* gives only the spelling parametrize, but the *Longman Dictionary of the English Language*, *Merriam-Webster's Collegiate Dictionary* and *Oxford English Dictionary* give both parameterize and parametrize.

3. From the *Collins English Dictionary*: "**gigaflop**... *n. Computer technol.* a measure of processing speed, consisting of a thousand million floating-point operations a second. [C20 ...]".

4. From the entry for *Abelian group* in the *Collins English Dictionary*: "Niels Henrik *Abel* (1802–29), Norwegian mathematician".

5. Mutatis mutandis means "with necessary changes" (*The Chambers Dictionary*).

6. Procrustes was "a villainous son of Poseidon in Greek mythology who forces travelers to fit into his bed by stretching their bodies or cutting off their legs" (*Merriam-Webster's Collegiate Dictionary*).

7. From the *Collins English Dictionary* (usage note after *especial*): "*Especial* and *especially* have a more limited use than *special* and *specially*. *Special* is always used in preference to *especial* when the sense is one of being out of the ordinary ... Where an idea of pre-eminence or individuality is involved, either *especial* or *special* may be used."

8. From the *Longman Dictionary of the English Language*, all three words being labelled *adj, informal*:

   mind-bending means "at the limits of understanding or credibility",
   mind-blowing means "**1** of or causing a psychic state similar to that produced by a psychedelic drug **2** mentally or emotionally exhilarating; overwhelming",
   mind-boggling means "causing great surprise or wonder".

Chapter 3
Mathematical Writing

Suppose you want to teach the "cat" concept to a very young child.
Do you explain that a cat is a relatively small,
primarily carnivorous mammal with retractile claws,
a distinctive sonic output, etc.?
I'll bet not.
You probably show the kid a lot of different cats,
saying "kitty" each time, until it gets the idea.
To put it more generally,
generalizations are best made by abstraction from experience.

— RALPH P. BOAS, *Can We Make Mathematics Intelligible?* (1981)

A good notation should be unambiguous, pregnant, easy to remember;
it should avoid harmful second meanings, and
take advantage of useful second meanings;
the order and connection of signs should
suggest the order and connection of things.

— GEORGE POLYA, *How to Solve It* (1957)

We have not succeeded in finding or constructing
a definition which starts out
"A Bravais lattice is . . . ";
the sources we have looked at say
"That was a Bravais lattice."

— CHARLES KITTEL, *Introduction to Solid State Physics* (1971)

Notation is everything.

— CHARLES F. VAN LOAN, *FFTs and the Sparse Factorization Idea* (1992)

The mathematical writer needs to be aware of a number of matters specific to mathematical writing, ranging from general issues, such as choice of notation, to particular details, such as how to punctuate mathematical expressions. In this chapter I begin by discussing some of the general issues and then move on to specifics.

3.1. What Is a Theorem?

What are the differences between theorems, lemmas, and propositions? To some extent, the answer depends on the context in which a result appears. Generally, a *theorem* is a major result that is of independent interest. The proof of a theorem is usually nontrivial. A *lemma*[3] is an auxiliary result—a stepping stone towards a theorem. Its proof may be easy or difficult. A straightforward and independent result that is worth encapsulating but that does not merit the title of a theorem may also be called a lemma. Indeed, there are some famous lemmas, such as the Riemann–Lebesgue Lemma in the theory of Fourier series and Farkas's Lemma in the theory of constrained optimization. Whether a result should be stated formally as a lemma or simply mentioned in the text depends on the level at which you are writing. In a research paper in linear algebra it would be inappropriate to give a lemma stating that the eigenvalues of a symmetric positive definite matrix are positive, as this standard result is so well known; but in a textbook for undergraduates it would be sensible to formalize this result.

It is not advisable to label all your results theorems, because if you do so you miss the opportunity to emphasize the logical structure of your work and to direct attention to the most important results. If you are in doubt about whether to call a result a lemma or a theorem, call it a lemma.

The term *proposition* is less widely used than lemma and theorem and its meaning is less clear. It tends to be used as a way to denote a minor theorem. Lecturers and textbook authors might feel that the modest tone of its name makes a proposition appear less daunting to students than a theorem. However, a proposition is not, as one student thought, "a theorem that might not be true".

A *corollary* is a direct or easy consequence of a lemma, theorem or proposition. It is important to distinguish between a corollary, which does not imply the parent result from which it came, and an extension or generalization of a result. Be careful not to over-glorify a corollary by failing to label it as such, for this gives it false prominence and obscures the role of the parent result.

[3]The plural of lemma is lemmata, or, more commonly, lemmas.

How many results are formally stated as lemmas, theorems, propositions or corollaries is a matter of personal style. Some authors develop their ideas in a sequence of results and proofs interspersed with definitions and comments. At the other extreme, some authors state very few results formally. A good example of the latter style is the classic book *The Algebraic Eigenvalue Problem* [296] by Wilkinson, in which only four titled theorems are given in 662 pages. As Boas [33] notes, "A great deal can be accomplished with arguments that fall short of being formal proofs."

A fifth kind of statement used in mathematical writing is a *conjecture*—a statement that the author thinks may be true but has been unable to prove or disprove. The author will usually have some strong evidence for the veracity of the statement. A famous example of a conjecture is the Goldbach conjecture (1742), which states that every even number greater than 2 is the sum of two primes; this is still unproved. One computer scientist (let us call him Alpha) joked in a talk "This is the Alpha and Beta conjecture. If it turns out to be false I would like it to be known as Beta's conjecture." However, it is not necessarily a bad thing to make a conjecture that is later disproved: identifying the question that the conjecture aims to answer can be an important contribution.

A *hypothesis* is a statement that is taken as a basis for further reasoning, usually in a proof—for example, an induction hypothesis. Hypotheses that stand on their own are uncommon; two examples are the Riemann hypothesis and the continuum hypothesis.

3.2. Proofs

Readers are often not very interested in the details of a proof but want to know the outline and the key ideas. They hope to learn a technique or principle that can be applied in other situations. When readers do want to study the proof in detail they naturally want to understand it with the minimum of effort. To help readers in both circumstances, it is important to emphasize the structure of a proof, the ease or difficulty of each step, and the key ideas that make it work. Here are some examples of the sorts of phrases that can be used (most of these are culled from proofs by Parlett in [217]).

> The aim/idea is to
> Our first goal is to show that
> Now for the harder part.
> The trick of the proof is to find
> ... is the key relation.
> The only, but crucial use of ... is that

> To obtain ... a little manipulation is needed.
> The essential observation is that

When you omit part of a proof it is best to indicate the nature and length of the omission, via phrases such as the following.

> It is easy/simple/straightforward to show that
> Some tedious manipulation yields
> An easy/obvious induction gives
> After two applications of ... we find
> An argument similar to the one used in ... shows that

You should also strive to keep the reader informed of where you are in the proof and what remains to be done. Useful phrases include

> First, we establish that
> Our task is now to
> Our problem reduces to
> It remains to show that
> We are almost ready to invoke
> We are now in a position to
> Finally, we have to show that

The end of a proof is often marked by the halmos symbol □ (see the quote on page 24). Sometimes the abbreviation QED (Latin: quod erat demonstrandum = which was to be demonstrated) is used instead.

There is much more to be said about writing (and devising) proofs. References include Franklin and Daoud [85], Garnier and Taylor [101], Lamport [173], Leron [177] and Polya [228].

3.3. The Role of Examples

A pedagogical tactic that is applicable to all forms of technical writing (from teaching to research) is to discuss specific examples before the general case. It is tempting, particularly for mathematicians, to adopt the opposite approach, but beginning with examples is often the more effective way to explain (see Boas's article [33] and the quote from it at the beginning of this chapter, a quote that itself illustrates this principle!).

A good example of how to begin with a specific case is provided by Strang in Chapter 1 of *Introduction to Applied Mathematics* [262]:

> The simplest model in applied mathematics is a system of linear equations. It is also by far the most important, and we begin

this book with an extremely modest example:

$$\begin{aligned} 2x_1 + 4x_2 &= 2, \\ 4x_1 + 11x_2 &= 1. \end{aligned}$$

After some further introductory remarks, Strang goes on to study in detail both this 2×2 system and a particular 4×4 system. General $n \times n$ matrices appear only several pages later.

Another example is provided by Watkins's *Fundamentals of Matrix Computations* [289]. Whereas most linear algebra textbooks introduce Gaussian elimination for general matrices before discussing Cholesky factorization for symmetric positive definite matrices, Watkins reverses the order, giving the more specific but algorithmically more straightforward method first.

An exercise in a textbook is a form of example. I saw a telling criticism in one book review that complained "The first exercise in the book was pointless, so why do the others?" To avoid such criticism, it is important to choose exercises and examples that have a clear purpose and illustrate a point. The first few exercises and examples should be among the best, to gain the reader's confidence. The same reviewer complained of another book that "it hides information in exercises and contains exercises that are too difficult." Whether such criticism is valid depends on your opinion of what are the key issues to be transmitted to the reader and on the level of the readership. Again, it helps to bear such potential criticism in mind when you write.

3.4. Definitions

Three questions to be considered when formulating a definition are "why?", "where?" and "how?" First, ask yourself why you are making a definition: is it really necessary? Inappropriate definitions can complicate a presentation and too many can overwhelm a reader, so it is wise to imagine yourself being charged a large sum for each one. Instead of defining a square matrix A to be contractive with respect to a norm $\|\cdot\|$ if $\|A\| < 1$, which is not a standard definition, you could simply say "A with $\|A\| < 1$" whenever necessary. This is easy to do if the property is needed on only a few occasions, and saves the reader having to remember what "contractive" means. For notation that is standard in a given subject area, judgement is needed to decide whether the definition should be given. Potential confusion can often be avoided by using redundant words. For example, if $\rho(A)$ is not obviously the spectral radius of the matrix A you can say "the spectral radius $\rho(A)$".

The second question is "where?" The practice of giving a long sequence of definitions at the start of a work is not recommended. Ideally, a definition should be given in the place where the term being defined is first used. If it is given much earlier, the reader will have to refer back, with a possible loss of concentration (or worse, interest). Try to minimize the distance between a definition and its place of first use.

It is not uncommon for an author to forget to define a new term on its first occurrence. For example, Steenrod uses the term "grasshopper reader" on page 6 of his essay on mathematical writing [256], but does not define it until it occurs again on the next page.

To reinforce notation that has not been used for a few pages you may be able to use redundancy. For example, "The optimal steplength α^* can be found as follows." This implicit redefinition either reminds readers what α^* is, or reassures them that they have remembered it correctly.

Finally, how should a term be defined? There may be a unique definition or there may be several possibilities (a good example is the term M-matrix, which can be defined in at least fifty different ways [23]). You should aim for a definition that is short, expressed in terms of a fundamental property or idea, and consistent with related definitions. As an example, the standard definition of a normal matrix is a matrix $A \in \mathbb{C}^{n\times n}$ for which $A^*A = AA^*$ (where $*$ denotes the conjugate transpose). There are at least 70 different ways of characterizing normality [119], but none has the simplicity and ease of use of the condition $A^*A = AA^*$.

By convention, *if* means *if and only if* in definitions, so do not write "The graph G is connected if and only if there is a path from every node in G to every other node in G." Write "The graph G is connected if there is a path from every node in G to every other node in G" (and note that this definition can be rewritten to omit the symbol G). It is common practice to italicize the word that is being defined: "A graph is *connected* if there is a path from every node to every other node." This has the advantage of making it perfectly clear that a definition is being given, and not a result. This emphasis can also be imparted by writing "A graph is defined to be connected if ...", or "A graph is said to be connected if ...."

If you have not done so before, it is instructive to study the definitions in a good dictionary. They display many of the attributes of a good mathematical definition: they are concise, precise, consistent with other definitions, and easy to understand.

Definitions of symbols are usually made with a simple equality, perhaps preceded by the word "let" if they are in-line, as in "let $q(x) = ax^2 + bx + c$." Various other notations have been devised to give emphasis to a definition, including

$$\begin{aligned} q(x) &:= ax^2 + bx + c, \\ ax^2 + bx + c &=: q(x), \\ q(x) &\stackrel{\text{def}}{=} ax^2 + bx + c, \\ q(x) &\equiv ax^2 + bx + c, \\ q(x) &\triangleq ax^2 + bx + c. \end{aligned}$$

If you use one of these special notations you must use it consistently, otherwise the reader may not know whether a straightforward equality is meant to be a definition.

3.5. Notation

Consider the following extract.

> Let $\widehat{H_k} = Q_k^H \widetilde{H_k} Q_k$, partition $X = [X_1, \ X_2]$ and let $\mathcal{X} = \text{range}(X_1)$. Let U^* denote the nearest orthonormal matrix to X_1 in the 2-norm.

These two sentences are full of potentially confusing notation. The distinction between the hat and the tilde in $\widehat{H_k}$ and $\widetilde{H_k}$ is slight enough to make these symbols difficult to distinguish. The symbols $\mathcal{X}$ and X are also too similar for easy recognition. Given that $\mathcal{X}$ is used, it would be more consistent to give it a subscript 1. The name H_k is unfortunate, because H is being used to denote the conjugate transpose, and it might be necessary to refer to $\widetilde{H_k}^H$! Since A^* is a standard synonym for A^H, the use of a superscripted asterisk to denote optimality is confusing.

As this example shows, the choice of notation deserves careful thought. Good notation strikes a balance among the possibly conflicting aims of being readable, natural, conventional, concise, logical and aesthetically pleasing. As with definitions, the amount of notation should be minimized.

Although there are 26 letters in the alphabet and nearly as many again in the Greek alphabet, our choice diminishes rapidly when we consider existing connotations. Traditionally, ϵ and δ denote small quantities, i, j, k, m and n are integers (or i or j the imaginary unit), λ is an eigenvalue and π and e are fundamental constants; π is also used to denote a permutation. These conventions should be respected. But by modifying and combining eligible letters we widen our choice. Thus γ and A yield, for example, $\widehat{A}$, $\overline{A}$, $\widetilde{A}$, A', γ_A, A_γ, $\mathbf{A}$, $\mathcal{A}$, $\mathbb{A}$.

Particular areas of mathematics have their own notational conventions. For example, in numerical linear algebra lower case Greek letters represent scalars, lower case roman letters represent column vectors, and upper case Greek or roman letters represent matrices. This convention was introduced by Householder [143].

In his book on the symmetric eigenvalue problem [217], Parlett uses the symmetric letters A, H, M, T, U, V, W, X, Y to denote symmetric matrices and the symmetric Greek letters Λ, Θ, Φ, Δ to denote diagonal matrices. Actually, the roman letters printed above are not symmetric because they are slanted, but Parlett's book uses a sans serif mathematics font that yields the desired symmetry. Parlett uses this elegant, but restrictive, convention to good effect.

We can sometimes simplify an expression by giving a meaning to extreme cases of notation. Consider the display

$$\beta_{ij} = \begin{cases} 0, & i > j, \\ \dfrac{1}{u_j}, & i = j, \\ \dfrac{1}{u_j} \displaystyle\prod_{r=i}^{j-1} \left(\dfrac{-c_r}{u_r} \right), & i < j. \end{cases}$$

There are really only two cases: $i > j$ and $i \le j$. This structure is reflected and the display made more compact if we define the empty product to be 1, and write

$$\beta_{ij} = \begin{cases} 0, & \text{if } i > j, \\ \dfrac{1}{u_j} \displaystyle\prod_{r=i}^{j-1} \left(\dfrac{-c_r}{u_r} \right), & \text{if } i \le j. \end{cases}$$

(Here, I have put "if" before each condition, which is optional in this type of display.) Incidentally, note that in a matrix product the order of evaluation needs to be specified: $\prod_{i=1}^{n} A_i$ could mean $A_1 A_2 \dots A_n$ or $A_n A_{n-1} \dots A_1$.

Notation also plays a higher level role in affecting the way a method or proof is presented. For example, the $n \times n$ matrix multiplication $C = AB$ can be expressed in terms of scalars,

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \qquad 1 \le i, j \le n,$$

or at the matrix-vector level,

$$C = [\, Ab_1, \quad Ab_2, \quad \dots, \quad Ab_n \,],$$

where $B = [b_1, b_2, \dots, b_n]$ is a partition into columns. One of these two viewpoints may be superior, depending on the circumstances. A deeper example is provided by the fast Fourier transform (FFT). The discrete Fourier transform (DFT) is a product $y = F_n x$, where F_n is the unitary Vandermonde matrix with (r,s) element $\omega^{(r-1)(s-1)}$ $(1 \le r, s \le n)$, and $\omega = \exp(-2\pi i/n)$. The FFT is a way of forming this product in $O(n \log n)$ operations. It is traditionally expressed through equations such as the following (copied from a numerical methods textbook):

$$\sum_{j=0}^{n-1} e^{2\pi ijk/n} f_j = \sum_{j=0}^{n/2-1} e^{2\pi ikj/(n/2)} f_{2j} + \omega^k \sum_{j=0}^{n/2-1} e^{2\pi ikj/(n/2)} f_{2j+1}.$$

The language of matrix factorizations can be used to give a higher level description. If $n = 2m$, the matrix F_n can be factorized as

$$F_n \Pi_n = \begin{bmatrix} I_m & \Omega_m \\ I_m & -\Omega_m \end{bmatrix} \begin{bmatrix} F_m & 0 \\ 0 & F_m \end{bmatrix},$$

where Π_n is a permutation matrix and $\Omega_m = \mathrm{diag}(1, \omega, \dots, \omega^{m-1})$. This factorization shows that an n-point DFT can be computed from two $n/2$-point transforms, and this reduction is the gist of the radix-2 FFT. The book *Computational Frameworks for the Fast Fourier Transform* by Van Loan [284], from which this factorization is taken, shows how, by using matrix notation, the many variants of the FFT can be unified and made easier to understand.

An extended example of how notation can be improved is given by Gillman in the appendix titled "The Use of Symbols: A Case Study" of *Writing Mathematics Well* [104]. Gillman takes the proof of a theorem by Sierpinski (1933) and shows how simplifying the notation leads to a better proof. Knuth set his students the task of simplifying Gillman's version even further, and four solutions are given in [164, §21].

Mathematicians are always searching for better notation. Knuth [163] describes two notations that he and his students have been using for many years and that he thinks deserve widespread adoption. One is notation for the Stirling numbers. The other is the notation $\mathcal{S}$, where $\mathcal{S}$ is any true-or-false statement. The definition is

$$[\mathcal{S}] = \begin{cases} 1, & \text{if } \mathcal{S} \text{ is true,} \\ 0, & \text{if } \mathcal{S} \text{ is false.} \end{cases}$$

For example, the Kronecker delta can be expressed as $\delta_{ij} = [i = j]$. The square bracket notation will seem natural to those who program; indeed, Knuth adapted it from a similar notation in the 1962 book by Iverson that led to the programming language APL [144, p. 11]. The square bracket notation is used in the textbook *Concrete Mathematics* [116]; that book and Knuth's paper give a convincing demonstration of the usefulness of the notation.

Halmos has these words to say about two of his contributions to mathematical notation [127]:

> My most nearly immortal contributions to mathematics are an abbreviation and a typographical symbol. I invented "iff," for "if and only if"—but I could never believe that I was really its first inventor .... The symbol is definitely not my invention—it appeared in popular magazines (not mathematical ones) before I adopted it, but, once again, I seem to have introduced it into mathematics. It is the symbol that sometimes looks like $\Box$, and is used to indicate an end, usually the end of a proof. It is most frequently called the "tombstone," but at least one generous author referred to it as the "halmos".

Table 3.1 shows the date of first use in print of some standard symbols; some of them are not as old as you might expect. Not all these notations met with approval when they were introduced. In 1842 Augustus de Morgan complained (quoted by Cajori [49, p. 328 (Vol. II)]):

> Among the worst of barbarisms is that of introducing symbols which are quite new in mathematical, but perfectly understood in common, language. Writers have borrowed from the Germans the abbreviation $n!$ to signify $1.2.3.\ldots(n-1)n$, which gives their pages the appearance of expressing surprise and admiration that 2, 3, 4, etc., should be found in mathematical results.

3.6. Words versus Symbols

Mathematicians are supposed to like numbers and symbols, but I think many of us prefer words. If we had to choose between reading a paper dominated by symbols and one dominated by words then, all other things being equal, most of us would choose the wordy paper, because we would expect it to be easier to understand. One of the decisions constantly facing the mathematical writer is how to express ideas: in symbols, in words, or both. I suggest some guidelines.

- Use symbols if the idea would be too cumbersome to express in words, or if it is important to make a precise mathematical statement.

- Use words as long as they do not take up much more space than the corresponding symbols.

Table 3.1. First use in print of some symbols. Sources: [49], [116], [163].

Symbol	Name	Year of publication
∞	infinity	1655 (Wallis)
π	pi (3.14159...)	1706 (Jones)
e	e (2.71828...)	1736 (Euler)
i	imaginary unit ($\sqrt{-1}$)	1794 (Euler)
$\equiv$	congruence	1801 (Gauss)
$n!$	factorial	1808 (Kramp)
$\sum$	summation	1820 (Fourier)
$\binom{n}{k}$	binomial coefficient	1826 (von Ettinghausen)
$\prod$	product	1829 (Jacobi)
∇	nabla	1853 (Hamilton)
δ_{ij}	Kronecker delta	1868 (Kronecker)
$\|z\|$	absolute value	1876 (Weierstrass)
$O(f(n))$	big oh	1894 (Bachmann)
$\lfloor x \rfloor$, $\lceil x \rceil$	floor, ceiling	1962 (Iverson)

- Explain in words what the symbols mean if you think the reader might have difficulty grasping the meaning or essential feature.

Here are some examples.

> (1) Define $C \in \mathbb{R}^{n\times n}$ by the property that vec(C) is the eigenvector corresponding to the smallest eigenvalue in magnitude of A, where the vec operator stacks the columns of a matrix into one long vector.

To make this definition using equations takes much more space, and is not worthwhile unless the notation that needs to be introduced (in this case, a name for $\min\{\, |\lambda| : \lambda \text{ is an eigenvalue of } A\,\}$) is used elsewhere. A possible objection to the above wordy definition of vec is that it does not specify in which order the columns are stacked, but that can be overcome by appending "taking the columns in order from first to last".

> (2) Since $|g'(0)| > 1$, zero is a repelling fixed point, so x_k does not tend to zero as $k \to \infty$.

An alternative is

> Since $|g'(0)| > 1$, 0 is a repelling fixed point, so $x_k \not\to 0$ as $k \to \infty$.

This sentence is only slightly shorter than the original and is harder to read—the symbols are beginning to intrude on the grammatical structure of the sentence.

> (3) If $B \in \mathbb{R}^{n\times n}$ has a unique eigenvalue λ of largest modulus then $B^k \approx \lambda^k xy^T$, where $Bx = \lambda x$ and $y^T B = \lambda y^T$ with $y^T x = 1$.

The alternative of "where x and y are a right and left eigenvector corresponding to λ, respectively, and $y^T x = 1$" is cumbersome.

> (4) Under these conditions the perturbed least squares solution $x + \Delta x$ can be shown to satisfy
>
> $$\frac{\|\Delta x\|_2}{\|x\|_2} \le \epsilon\kappa_2(A)\left(1 + \frac{\|b\|_2}{\|A\|_2\|x\|_2}\right) + \epsilon\,\kappa_2(A)^2\frac{\|r\|_2}{\|A\|_2\|x\|_2} + O(\epsilon^2).$$
>
> Thus the sensitivity of x is measured by $\kappa_2(A)$ if the residual r is zero or small, and otherwise by $\kappa_2(A)^2$.

Here, we have a complicated bound that demands an explanation in words, lest the reader overlook the significant role played by the residual r.

> (5) If $y_1, y_2, \ldots, y_n$ are all $\neq 1$ then $g(y_1, y_2, \ldots, y_n) > 0$.

In the first sentence "all $\neq 1$" is a clumsy juxtaposition of word and equation and most writers would express the statement differently. Possibilities include

> If $y_i \neq 1$ for $i = 1, 2, \ldots, n$, then $g(y_1, y_2, \ldots, y_n) > 0$.
> If none of the y_i $(i = 1, 2, \ldots, n)$ equals 1, then $g(y_1, y_2, \ldots, y_n) > 0$.

If the condition were "$\neq 0$" instead of "$\neq 1$", then it could simply be replaced by the word "nonzero". In cases such as this, the choice between words and symbols in the text (as opposed to in displayed equations) is a matter of taste; good taste is acquired by reading a lot of well-written mathematics.

The symbols $\forall$ and $\exists$ are widely used in handwritten notes and are an intrinsic part of the language in logic. But generally, in equations that are in-line, they are better replaced by the equivalent words "for all" and "there exists". In displayed equations either the symbol or the phrase is acceptable, though I usually prefer the phrase. Compare

$$\sigma\bigl(G(t)\bigr) = \exp\bigl(t\sigma(A)\bigr) \quad \text{for all } t \geq 0$$

with

$$\sigma\bigl(G(t)\bigr) = \exp\bigl(t\sigma(A)\bigr) \quad \forall t \geq 0.$$

Similar comments apply to the symbols $\Rightarrow$ (implies) and $\iff$ (if and only if), though these symbols are more common in displayed formulas.

Of course, for some standard phrases that appear in displayed formulas, there is no equivalent symbol:

$$\text{minimize} \quad c^T x - \mu \sum_{i=1}^{n} \ln x_i \quad \text{subject to } Ax = b,$$

$$\varliminf \tfrac{1}{n} \log D_j^n < 0 \quad \text{almost surely},$$

$$z^T y = \|z\|_D \|y\| = 1, \quad \text{where} \quad \|z\|_D = \max_{v \neq 0} \frac{|z^T v|}{\|v\|}.$$

3.7. Displaying Equations

An equation is displayed when it needs to be numbered, when it would be hard to read if placed in-line, or when it merits special attention, perhaps because it contains the first occurrence of an important variable. The following extract gives an illustration of what and what not to display.

> Because $\delta(\overline{x}, \mu)$ is the smallest value of $\|\overline{X}z/\mu - e\|$ for all vectors y and z satisfying $A^T y + z = c$, we have
>
> $$\delta(\overline{x}, \mu) \leq \|\tfrac{1}{\mu}\overline{X}z - e\|.$$
>
> Using the relations $z = \mu X^{-1} s$ and $\overline{x}_i = 2x_i - x_i s_i$ gives
>
> $$\frac{1}{\mu}\overline{X}z = \overline{X}X^{-1}s = (2X - XS)X^{-1}s = 2s - S^2 e.$$
>
> Therefore, $\delta(\overline{x}, \mu) \leq \|2s - S^2 e - e\|$, which means that
>
> $$\delta(\overline{x}, \mu)^2 \leq \sum_{i=1}^{n} (2s_i - s_i^2 - 1)^2 = \sum_{i=1}^{n} (s_i - 1)^4 \leq \left(\sum_{i=1}^{n} (s_i - 1)^2 \right)^2 = \delta(x, \mu)^4.$$

> The condition $\delta(x,\mu) < 1$ thus ensures that the Newton iterates $\overline{x}$ converge quadratically.

The second and third displayed equations are too complicated to put in-line. The first $\delta(\overline{x},\mu)$ inequality is displayed because it is used in conjunction with the second display and it is helpful to the reader to display both these steps of the argument. The consequent inequality $\delta(\overline{x},\mu) \le \|2s - S^2e - e\|$ fits nicely in-line, and since it is used immediately it is not necessary to display it.

When a displayed formula is too long to fit on one line it should be broken before a binary operation. Example:

$$\begin{aligned} |e_{m+1}| \le |G^{m+1}e_0| + c_n u(1+\theta_x)\bigl\{c(A)|(I-G)^D M^{-1}| \\ + (m+1)|(I-E)M^{-1}|\bigr\}(|M|+|N|)|x|. \end{aligned}$$

The indentation on the second line should take the continuation expression past the beginning of the left operand of the binary operation at which the break occurred, though, as this example illustrates, this is not always possible for long expressions. A formula in the text should be broken after a relation symbol or binary operation symbol, not before.

3.8. Parallelism

Parallelism should be used, where appropriate, to aid readability and understanding. Consider this extract:

> The Cayley transform is defined by $C = (A - \theta_1 I)^{-1}(A - \theta_2 I)$. If λ is an eigenvalue of A then
>
> $$(\lambda - \theta_2)(\lambda - \theta_1)^{-1}$$
>
> is an eigenvalue of C.

The factors in the eigenvalue expression are presented in the reverse order to the factors in the expression for C. This may confuse the reader, who might, at first, think there is an error. The two expressions should be ordered in the same way.

Parallelism works at many levels, from equations and sentences to theorem statements and section headings. It should be borne in mind throughout the writing process. If one theorem is very similar to another, the statements should reflect that—the wording should not be changed just for the sake of variety (see *elegant variation*, §4.15). However, it is perfectly acceptable to economize on words by saying, in Theorem 2 (say) "Under the conditions of Theorem 1".

For a more subtle example, consider the sentence

It is easy to see that $f(x, y) > 0$ for $x > y$.

In words, this sentence is read as "It is easy to see that $f(x, y)$ is greater than zero for x greater than y." The first $>$ translates to "is greater than" and the second to "greater than", so there is a lack of parallelism, which the reader may find disturbing. A simple cure is to rewrite the sentence:

It is easy to see that $f(x, y) > 0$ when $x > y$.
It is easy to see that if $x > y$ then $f(x, y) > 0$.

3.9. Dos and Don'ts of Mathematical Writing

Punctuating Expressions

Mathematical expressions are part of the sentence and so should be punctuated. In the following display, all the punctuation marks are necessary. (The second displayed equation might be better moved in-line.)

The three most commonly used matrix norms in numerical analysis are particular cases of the Hölder p-norm

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}, \qquad A \in \mathbb{R}^{m \times n},$$

where $p \geq 1$ and

$$\|x\|_p = \Big(\sum_{i=1}^{n} |x_i|^p \Big)^{1/p}.$$

Otiose Symbols

Do not use mathematical symbols unless they serve a purpose. In the sentence "A symmetric positive definite matrix A has real eigenvalues" there is no need to name the matrix unless the name is used in a following sentence. Similarly, in the sentence "This algorithm has $t = \log_2 n$ stages", the "$t =$ " can be omitted unless t is defined in this sentence and used immediately. Watch out for unnecessary parentheses, as in the phrase "the matrix $(A - \lambda I)$ is singular."

Placement of Symbols

Avoid starting a sentence with a mathematical expression, particularly if a previous sentence ended with one, otherwise the reader may have difficulty parsing the sentence. For example, "A is an ill-conditioned matrix" (possible confusion with the word "A") can be changed to "The matrix A is ill-conditioned."

Separate mathematical symbols by punctuation marks or words, if possible, for the same reason.

Bad: If $x > 1$ $f(x) < 0$.
Fair: If $x > 1$, $f(x) < 0$.
Good: If $x > 1$ then $f(x) < 0$.

Bad: Since $p^{-1} + q^{-1} = 1$, $\|\cdot\|_p$ and $\|\cdot\|_q$ are dual norms.
Good: Since $p^{-1} + q^{-1} = 1$, the norms $\|\cdot\|_p$ and $\|\cdot\|_q$ are dual.

Bad: It suffices to show that $\|H\|_p = n^{1/p}$, $1 \le p \le 2$.
Good: It suffices to show that $\|H\|_p = n^{1/p}$ for $1 \le p \le 2$.
Good: It suffices to show that $\|H\|_p = n^{1/p}$ $(1 \le p \le 2)$.

Bad: For $n = r$ (2.2) holds with $\delta_r = 0$.
Good: For $n = r$, (2.2) holds with $\delta_r = 0$.
Good: For $n = r$, inequality (2.2) holds with $\delta_r = 0$.

"The" or "A"

In mathematical writing the use of the article "the" can be inappropriate when the object to which it refers is (potentially) not unique or does not exist. Rewording, or changing the article to "a", usually solves the problem.

Bad: Let the Schur decomposition of A be QTQ^*.
Good: Let a Schur decomposition of A be QTQ^*.
Good: Let A have the Schur decomposition QTQ^*.

Bad: Under what conditions does the iteration converge to the solution of $f(x) = 0$?
Good: Under what conditions does the iteration converge to a solution of $f(x) = 0$?

Notational Synonyms

Sometimes you have a choice of notational synonyms, one of which is preferable. In the following examples, the first of each pair is, to me, the more aesthetically pleasing or easier to read (a capital letter denotes a matrix).

$$\Bigl(\sum_{i,j}(a_{ij}-b_{ij})^2\Bigr)^{1/2}, \qquad \sqrt{\sum_{i,j}(a_{i,j}-b_{i,j})^2},$$

$$\exp\bigl(2\pi i(x^2+y^2)^{-1/2}\bigr), \qquad e^{\frac{2\pi i}{\sqrt{x^2+y^2}}},$$

$$(1-n\epsilon)^{-1}|L||U|, \qquad \frac{|L||U|}{1-n\epsilon},$$

$$X_{k+1}=\tfrac{1}{2}X_k(3I-X_k^2), \qquad X_{k+1}=\frac{X_k}{2}[3I-X_k^2],$$

$$\min\{\,\epsilon : |b-Ay|\le \epsilon|A||y|\,\}, \qquad \min\{\,\epsilon \mid |b-Ay|\le \epsilon|A||y|\,\},$$

$$\min\{\,\|A-UBP\| : U^TU=I,\ P \text{ a permutation}\,\},$$

$$\min_{\substack{U^TU=I\\ P \text{ a permutation}}} \|A-UBP\|.$$

In the next two examples, the first form is preferable because it saves space without a loss of readability.

$$x=[x_1,\ \ x_2,\ \ \ldots,\ \ x_n]^T, \qquad x=\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix},$$

$$\Lambda=\mathrm{diag}(\lambda_i), \qquad \Lambda=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}.$$

Of course, the diag$(\cdot)$ notation should be defined if it is not regarded as standard.

Referencing Equations

When you reference an earlier equation it helps the reader if you add a word or phrase describing the nature of that equation. The aim is to save the reader the trouble of turning back to look at the earlier equation. For example, "From the definition (6.2) of dual norm" is more helpful than "From (6.2)"; and "Combining the recurrence (3.14) with inequality (2.9)" is more helpful than "Combining (3.14) and (2.9)". Mermin [200] calls this advice the "Good Samaritan Rule". As in these examples, the word added should be something more informative than just "equation" (or the ugly abbreviation "Eq."), and inequalities, implications and lone expressions should not be referred to as equations.

Miscellaneous

When working with complex numbers it is best not to use "i" as a counting index, to avoid confusion with the imaginary unit. More generally, do not use a letter as a dummy variable if it is already being used for another purpose.

Note the difference between the Greek letter epsilon, ϵ, and the "belongs to" symbol $\in$, as in $\|x\| \le \epsilon$ and $x \in \mathbb{R}^n$. Another version of the Greek epsilon is ε. Note the distinction between the Greek letter π and the product symbol $\prod$.

By convention, standard mathematical functions such as sin, cos, arctan, max, gcd, trace and lim are set in roman type, as are multiple-letter variable names. It is a common mistake to set these in italic type, which is ambiguous. For example, is $tanx$ the product of four scalars or the tangent of x?

In bracketing multilayered expressions you have a choice of brackets for the layers and a choice of sizes, for example $\Big\{\Big[\Big(\big\{\big[\big($, this ordering being the one recommended by *The Chicago Manual of Style* [58]. Most authors try to avoid mixing different brackets in the same expression, as it leads to a rather muddled appearance.

Write "the kth term", not "the k^{th} term", "the k'th term" or "the k-th term." (It is interesting to note that nth is a genuine word that can be found in most dictionaries.)

A slashed exponent, as in $y^{1/2}$, is generally preferable to a stacked one, as in $y^{\frac{1}{2}}$.

The standard way to express that i is to take the values 1 to n in steps of 1 is to write

$$i = 1, \ldots, n \quad \text{or} \quad i = 1, 2, \ldots, n,$$

where all the commas are required. An alternative notation originating in programming languages such as Fortran 90 and MATLAB is $i = 1\colon n$. For counting down we can write $i = n, n-1, \ldots, 1$ or $i = n\colon -1\colon 1$, where the middle integer denotes the increment. This notation is particularly convenient when extended to describe submatrices: $A(i\colon j, p\colon q)$ denotes the submatrix formed from the intersection of rows i to j and columns p to q of the matrix A.

Avoid (or rewrite) tall in-line expressions, such as $\begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$, which can disrupt the line spacing.

There are two different kinds of ellipsis: vertically centred $(\cdots)$ and "ground level" or "baseline" $(\ldots)$. Generally, the former is used between operators such as $+$, $=$, and $\le$, and the latter is used between a list of

Glossary for Mathematical Writing

1. Without loss of generality = I have done an easy special case.
2. By a straightforward computation = I lost my notes.
3. The details are left to the reader = I can't do it.
4. The following alternative proof of X's result may be of interest = I cannot understand X.
5. It will be observed that = I hope you hadn't noticed that.
6. Correct to within an order of magnitude = wrong.

Adapted from [222].

symbols or to indicate a product. Examples:

$$x_1 + x_2 + \cdots + x_n, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n, \quad \lambda_1, \lambda_2, \ldots, \lambda_n, \quad A_1 A_2 \ldots A_n.$$

An operator or comma should be symmetrically placed around the ellipsis; thus $x_1 + x_2 + \cdots x_n$ and $\lambda_1, \lambda_2, \ldots \lambda_n$ are incorrect.

When an ellipsis falls at the end of a sentence there is the question of how the full stop (or period) is treated. Recommendations vary. *The Chicago Manual of Style* suggests typing the full stop before the three ellipsis points (so that there is no space between the first of the four dots and the preceding character). When the ellipsis is part of a mathematical formula it seems natural to put it before the full stop, but the two possibilities may be visually indistinguishable, as in the sentence

> The Mandelbrot set is defined in terms of the iteration $z_{k+1} = z_k^2 + c$, $k = 0, 1, 2, \ldots$.

A vertically centred dot is useful for denoting multiplication in expressions where terms need to be separated for clarity:

$$16046641 = 13 \cdot 37 \cdot 73 \cdot 457,$$

$$\operatorname{cond}(A, x) = \frac{\| \, |I - A^+A| \cdot |A^T| \cdot |A^{+T}x| \, \|}{\|x\|}.$$

Care is needed to avoid ambiguity in slashed fractions. For example, the expression $-(b-a)^3/12f''(\eta)$ is better written as $-\big((b-a)^3/12\big)f''(\eta)$ or $-f''(\eta)(b-a)^3/12$.

Chapter 4
English Usage

There's almost no more beautiful sight than a simple declarative sentence.
— WILLIAM ZINSSER, *Writing with a Word Processor* (1983)

Quite aside from format and style,
mathematical writing is supposed to say something.
Put another way: the number of ideas divided by the
number of pages is supposed to be positive.
— J. L. KELLEY, *Writing Mathematics* (1991)

Let us not deceive ourselves.
"There is no God but Allah" is a more gripping sentence than
Mohammed (also Mahomet, Muhammad; 570?–632) *asserted a doctrine of*
unqualified monotheism (*suras* 8, 22, 33–37, 89, 91, *Koran*).
— MARY-CLAIRE VAN LEUNEN, *A Handbook for Scholars* (1992)

I am about to—or I am going to—die;
either expression is used.
— DOMINIQUE BONHOURS[4] (on his deathbed)

One should not aim at being possible to understand,
but at being impossible to misunderstand.
— QUINTILLIAN

[4]Quoted in [42, p. 146].

In this chapter I discuss aspects of English usage that are particularly relevant to mathematical writing. You should keep three things in mind as you read this chapter. First, on many matters of English usage rules have exceptions, and, moreover, not all authorities agree on the rules. I have consulted several usage guides (those described in §2.2) and have tried to give a view that reflects usage in writing today. Second, about half the topics discussed here are not peculiar to the English language, but simply reflect common sense in writing. Third, many of the points mentioned are not vitally important when taken in isolation. But, as van Leunen explains (quoted in [164, p. 97]),

> Tone is important, and tone consists entirely of making these tiny, tiny choices. If you make enough of them wrong ... then you won't get your maximum readership. The reader who has to read the stuff will go on reading it, but with less attention, less commitment than you want. And the reader who doesn't have to read will stop.

4.1. A or An?

Whether *a* or *an* should precede a noun depends on how the first syllable is pronounced: *a* is used if the first syllable begins with a consonant sound and *an* if it begins with a vowel sound. For this rule, the initial "yew" sound in the words university and European is regarded as a consonant sound: thus "a university", "a European". For words beginning with h, *a* is used unless the h is not sounded. The only words in this last category are *heir*, *honest*, *honour* (US *honor*), *hour* and their derivatives.

The question "*a* or *an*?" most frequently arises with acronyms, abbreviations and proper nouns. An easy example is "an NP-complete problem". In the world of mathematical software, given that the usual pronunciation of LAPACK is l-a-pack, and that of NAG is nag, we would write "an LAPACK routine" but "a NAG library routine".

4.2. Abbreviations

One school of thought says that the use of the abbreviations *e.g.* and *i.e.* is bad style, and that *for example* and *that is* make for a smoother flowing sentence. In any case, *i.e.* and *that is* should usually be followed by a comma, and all four forms should be preceded by a comma. When using the abbreviations you should type *e.g.* and *i.e.*, not *eg.* or *ie.*, since the abbreviations represent two words (the Latin *exempli gratia* and *id est*). Note that in the following sentence *i.e.* should be deleted: "The most expensive method, i.e. Newton's method, converges quadratically."

The less frequently used *cf.* has only one full stop (a "period" in American English), because it is an abbreviation of a single word: the Latin *confer*, meaning compare. Often, cf. is used incorrectly in the sense of "see", as in "cf. [6] for a discussion". The abbreviation *et al.* is short for *et alia*, so it needs only one full stop.

The abbreviation *N.B.* (of the Latin *nota bene*) is not often used in technical writing, probably because it has to appear at the beginning of a sentence and is somewhat inelegant. You can usually find a better way to give the desired emphasis.

The abbreviation *iff*, although handy in notes, is usually spelled out as *if and only if.*

The normal practice when introducing a nonstandard abbreviation or acronym is to spell out the word or phrase in full on its first occurrence and place the abbreviation immediately after it in parentheses. Thereafter the abbreviation is used. Example:

> Gaussian elimination (GE) is a method for solving a system of n linear equations in n unknowns. GE has a long history; a variant of it for solving systems of three equations in three unknowns appears in the classic Chinese work "Chiu-Chang Suan-Shu", written around 250 B.C.

4.3. Absolute Words

Certain adjectives have an absolute meaning and cannot be qualified by words such as less, quite, rather and very. For example, it is wrong to write *most unique* (replace by *unique*, or perhaps *most unusual*), *absolutely essential*, *more ideal*, or *quite impossible*. However, *essentially unique* is an acceptable term in mathematical writing: it means unique up to some known transformations. Many other words are frequently used with an absolute sense but can be modified (although for some words this usage is open to criticism); example phrases are "convergence is *almost certain*", "a *very complete* survey", "the *most obvious* advantage", and "the function is differentiable *almost everywhere*."

4.4. Active versus Passive

Prefer the active to the passive voice (prefer "X did Y" to "Y was done by X"). The active voice adds life and movement to writing, whereas too much of the passive voice weakens the communication between writer and reader.

Passive: The answer was provided to sixteen decimal places by Gaussian elimination.
Active: Gaussian elimination gave the answer to sixteen decimal places.

Passive: The failure of Newton's method to converge is attributed to the fact that the Jacobian is singular at the solution.
Active: Newton's method fails to converge because the Jacobian is singular at the solution.

Passive: A numerical example is now given to illustrate the above result.
Active: We give a numerical example to illustrate this result, *or* The following numerical example illustrates this result.

The second example in the following trio illustrates a further degree of abstraction in which a verb is replaced by an abstract noun modified by another verb.

Passive: The optimality of y was verified by checking that the Hessian matrix was positive definite.
Passive and indirect: Verification of the optimality of y was achieved by checking that the Hessian matrix was positive definite.
Active: We verified the optimality of y by checking that the Hessian matrix was positive definite.

Other "was" phrases that can often be removed by rewriting are "was performed", "was experienced", "was carried out", "was conducted" and "was accomplished".

The passive voice has an important role to play, however. It adds variety, may be needed to put emphasis on a certain part of a sentence, and may be the only choice if specific information required for an active variant is unknown or inappropriate to mention. Examples where the passive voice allows the desired emphasis are "An ingenious proof of this conjecture was constructed by C. L. Ever", and (from the writings of Halmos [245, p. 96]) "The subjects just given honorable mention, as well as the three actually discussed in detail, have been receiving serious research attention in the course of the last twenty years." The passive voice is also useful for euphemistic effect, allowing the clumsy experimenter to say "The specimen was accidentally strained during mounting" instead of "I dropped the specimen on the floor."

The Ten Commandments of Good Writing

1. Each pronoun should agree with their antecedent.
2. Just between you and I, case is important.
3. A preposition is a poor word to end a sentence with.
4. Verbs has to agree with their subject.
5. Don't use no double negatives.
6. Remember to never split an infinitive.
7. When dangling, don't use participles.
8. Join clauses good, like a conjunction should.
9. Don't write a run-on sentence it is difficult when you got to punctuate it so it makes sense when the reader reads what you wrote.
10. About sentence fragments.

Reprinted, with permission, from *How to Write and Publish a Scientific Paper* [68].

4.5. Adjective and Adverb Abuse

Use an adjective or adverb only if it earns its place. The adjectives or adverbs *very*, *rather*, *quite*, *nice* and *interesting* should be used with caution in technical writing, as they are imprecise. For example, in "The proof is very easy" and "This inequality is quite important" the adverbs are best omitted. Examples of acceptable usage are "These results are very similar to those of Smith" and "This bound can be very weak." *Interesting* is an overworked adjective that can often be avoided. For example, in the sentence "It is interesting to re-prove this result using Laplace transforms", *instructive* is probably the intended word.

Try to avoid using nouns as adjectives. "The method for iteration parameter estimation" can be expressed more clearly as "The method for estimating iteration parameters." While proper nouns are often used as adjectives in speech ("this sequence is Cauchy", "that matrix is Toeplitz"), such usage in formal writing is best avoided (write "this is a Cauchy sequence", "that is a Toeplitz matrix"). Similarly, write "Euler's method is unstable" instead of "Euler is unstable."

An *adverb* that is overworked in mathematical writing is *essentially*. Dictionaries define it to mean necessarily or fundamentally, but it is often used with a vague sense meaning "almost, but not quite". Before using the word, consider whether you can be more precise.

> Bad: Beltrami (1873) essentially derived the singular value decomposition.
>
> Good: Beltrami (1873) derived the singular value decomposition for square, nonsingular matrices.

A valid use of *essentially* is in the expression "essentially the same as", which by convention in scientific writing means "the same, except for minor details".

4.6. -al and -age

Certain words that can be extended with an -al or -age ending are often misused in the extended form. The suffix tends to give a more abstract meaning, which makes it more difficult to use the word correctly. For example, *usage* means a manner of using, so correct *usage* is illustrated by "in the original usage the conjugate gradient method was not preconditioned" and "the use of Euler's method is not recommended for stiff differential equations."

An example where an -al ending is used incorrectly is "the most pragmatical opinion is the one expressed by the term's inventor", in which the third word should be "pragmatic".

4.7. Ambiguous "This" and "It"

A requirement of good writing is to make clear to the reader, at all times, what is the entity under discussion. *This* phrases such as "This is a consequence of Theorem 2" should be used with caution as they can force the reader to backtrack to find what "this" refers to. Often it helps to insert an appropriate noun after *this. It* can also be ambiguous: in the sentence "Condition 3 is not satisfied for the steepest descent method, which is why we do not consider it further" we cannot tell whether it is the condition or the method that is not being pursued.

4.8. British versus American Spelling

In my opinion (as a Briton) it makes little difference whether you use British or American (US) spellings, as long as you are consistent within a given piece of writing. For some journals, copy editors will convert a manuscript to one or the other form of spelling. Major dictionaries give both spellings. I find it natural to use the spelling of the country in which I am working at the time! See §5.9 for some examples of the different spellings.

4.9. Capitalization

Words that are derived from a person's name inherit the capitalization. Thus: Gaussian elimination, Hamiltonian system, Hermitian matrix, Jacobian matrix, Lagrangian function, Euler's method, and so on. The incorrect form "hermitian" is sometimes seen. The Lax Equivalence Theorem is quite different from a lax equivalence theorem! Some (but not all) dictionaries list "abelian" with a small "a", showing that eponymous adjectives can gradually become accepted in uncapitalized form.

There does not seem to be a standard rule for when to capitalize the word following a colon. Bernstein [28] and Knuth [164, p. 11] both suggest the useful convention of capitalizing when the phrase following the colon is a full sentence.

4.10. Common Misspellings or Confusions

The errors shown in Table 4.1 seem to be common in mathematical writing. The misspellings marked with an asterisk are genuine words, but have different meanings from the corresponding words in the left column.

One Web page I visited describes "seperate" as the most common misspelling on the Internet and lose/loose as the second most common. Using the Web search engine Alta Vista I found one occurrence of "seperate" for every 24 occurrences of "separate".

The journal *Physical Review Letters* started spelling *Lagrangian* as *Lagrangean* in mid 1985, a change which is incorrect according to most dictionaries. Mermin, a Cornell physicist, spotted the switch and wrote an article criticizing it [201]. The journal has now reverted to *Lagrangian*.

According to McIlroy [199], on most days at Bell Laboratories someone misspells the word *accommodate*, in one of seven incorrect ways.

4.11. Consistency

It is important to be consistent. Errors of consistency often go unnoticed, but can be puzzling to the reader. Don't refer to $\ker(A)$ as the null-space, or $\mathrm{null}(A)$ as the kernel—stick to matching synonyms. If you use the term "Cholesky factorization", don't say the "Cholesky decomposition" in the

Table 4.1. Common errors. Asterisk denotes a genuine word.

Correct/Intended	Misspelling
analogous	analagous
criterion	criteria*
dependent	dependant*
discrete	discreet*
Frobenius	Frobenious
idiosyncrasy	idiosyncracy
in practice	in practise
led (past tense of lead)	lead*
lose	loose* (very common)
phenomenon	phenomena* (plural of phenomenon)
preceding	preceeding
principle	principal*
propagation	propogation
referring	refering
Riccati	Ricatti
separate	seperate
supersede[a]	supercede
zeros	zeroes[b]

[a]The only English word ending in -sede.

[b]Some, but not all, dictionaries give the -oes ending as an alternative spelling for the plural noun.

same work. If words have alternative spellings, stick to one: don't use both orthogonalise and orthogonalize, and if you use orthogonalize also use, for example, optimize, not optimise. But note that not all -ise words can be spelt with -ize; examples are listed in §5.9.

4.12. Contractions

Contractions such as *it's*, *let's*, *can't* and *don't* are not used in formal works, but are acceptable in popular articles if used sparingly. Note the distinction between the contraction *it's* (short for *it is*) and the possessive *its*: "It's raining", "A matrix is singular if its determinant is zero." One editor comments that the two most frequent errors she encounters are the use of *it's* for *its* and incorrect punctuation surrounding *however* [286, p. 39].

4.13. Dangling Participle

What is wrong with the following sentence?

> Substituting (3) into (7), the integral becomes $\pi^2/9$.

This sentence suggests that the integral makes the substitution. The error is that the intended subject ("we") of the participle *substituting* is not present in the sentence. Rewritten and unambiguous versions are

> Substituting (3) into (7), we find that the integral is $\pi^2/9$.
> When (3) is substituted into (7), the integral becomes $\pi^2/9$.

A similar example that is less obviously wrong is

> When deriving parallel algorithms, the model of computation must be considered carefully.

Dangling participles are usually not ambiguous when read in context, but they can be distracting:

> A bug was found in the program using random test data.

Here is another example of a different type:

> Being stiff, the Runge-Kutta routine required a large amount of CPU time to solve the differential equation.

Here, the problem is that the noun immediately following "being" is not the one to which this participle refers. There are several ways to rewrite the sentence. One that preserves the emphasis is

> Because the differential equation is stiff, the Runge-Kutta routine required a large amount of CPU time to solve it.

Certain participial constructions are idiomatic and hence are regarded as acceptable:

> Assuming $G(x^*)$ is positive definite, x^* is a minimum point for F.
> Considering the large dimension of the problem, convergence was obtained in remarkably few iterations.
> Strictly speaking, the bound holds only for $n\epsilon < 1$.

4.14. Distinctions

Affect, Effect. *Affect* is a verb meaning to produce a change. *Effect* is a noun meaning the result of a change. Examples: "Multiple roots affect the convergence rate of Newton's method", "One effect of multiple roots is to reduce the convergence rate of Newton's method to linear." Effect is also a verb meaning to bring about (as in "to effect a change"), but in this form it is rarely needed in mathematical writing.

Alternate, Alternative. *Alternate* implies changing repeatedly from one thing to another. An *alternative* is one of several options. Compare "While writing his thesis the student alternated between elation and misery", with "The first attempt to prove the theorem failed, so an alternative method of proof was tried."

Compare with, Compare to. *Compare with* analyses similarities and differences between two things, whereas *compare to* states a resemblance between them. Examples: "We now compare Method A with Method B", "Shakespeare compared the world to a stage", "Shall I compare thee to a summer's day?" As Bryson [41] explains, "Unless you are writing poetry or love letters, *compare with* is usually the expression you want." *Compare and* is an alternative to *compare with*: "We now compare Method A and Method B" or, better, "We now compare Methods A and B."

Compose, Comprise, Constitute. *Compose* means to make up, *comprise* means to consist of. "Comprised of" is always incorrect. Thus, "the course is composed of three topics" or "the course comprises three topics", but not "the course is comprised of three topics." *Constitute* is a transitive verb used in the reverse sense: "these three topics constitute the course."

Due to, Owing to. These two expressions are not interchangeable, though writers frequently use *due to* in place of *owing to*. Use *due to* where you could use "caused by", or "attributable to"; use *owing to* where you could use "because of". Thus "The instability is due to a rank deficient submatrix" but "Owing to a rank deficient submatrix the computed result was inaccurate."

Fewer, Less. *Less* refers to quantity, amount or size, *fewer* to number. Thus "the zeros of $f(x)$ are less than those of $g(x)$" means that if x is a zero of f and y a zero of g then $x < y$, whereas "the zeros of $f(x)$ are fewer than those of $g(x)$" means that g has more zeros than f. Bryson [41] states the rule of thumb that *less* should be used with singular nouns and *fewer* with plural nouns: less research, less computation, fewer graduates, fewer papers.

Practice, Practise. In British English, *practice* is the noun and *practise* the verb (as with advice and advise). Thus "in practice", "practice session", "practise the technique", "practised speaker". But in American English both verb and noun are spelt *practice.*

Which, That. A "wicked which"[5] is an instance of the word *which* that should be *that* (example: replace the word before *should*, earlier in this sentence, by *which*). The rule is that *that* defines and restricts, whereas *which* informs and does not restrict.[6] (Mathematicians should be good at spotting definitions.) Note the difference between the following two examples.

"Consider the Pei matrix, which is positive definite." We are being told additional information about the Pei matrix: that it is positive definite.

"Consider the Pei matrix that is positive definite." Now we are being asked to focus on a particular Pei matrix from among several: the one that is positive definite.

A useful guide is that which-clauses are surrounded by commas, or preceded by a comma if at the end of a sentence. If you're not sure whether to use *which* or *that*, see whether your sentence looks right with commas around the relevant clause. Sometimes it pays to introduce a wicked which to avoid ugly repetition, as has been done in the sentence "This approach is similar to that which we used in our earlier paper" (though "the one we used" is better). A rule of thumb discussed in [164, pp. 96–97] is to replace *which* by *that* whenever it sounds right to do so.

4.15. Elegant Variation

Elegant variation is defined by the Fowlers [84] as "substitution of one word for another for the sake of variety". It is a tempting way to avoid repetition, but is often unnecessary and can introduce ambiguity. Consider the sentence "The eigenvalue estimate from Gershgorin's theorem is a crude bound, but it is easy to compute." Does Gershgorin's theorem yield an estimate or a strict bound? We cannot tell from the sentence. In fact, the answer is that it can yield either, depending on how you interpret the theorem. A rewrite of the sentence (with knowledge of the theorem) produces "The eigenvalue inclusion regions provided by Gershgorin's theorem are crude, but easy to evaluate", where *inclusion regions* can be replaced by *estimates* or *bounds*, depending on the desired emphasis.

The opposite of elegant variation is when the same word is repeated in different forms or with different meanings. Here are two examples.

The performance is impressive and gives the impression that

[5] A term coined by Knuth [164].

[6] Some authorities permit *which* to be used in a defining clause (e.g., Gowers [115]), but, as Bryson [41] puts it, "the practice is on the whole better avoided."

> the blocksize is nearly optimal. [impressive, impression]
> In the remainder of this chapter we examine the remainder in Euler's summation formula. [remainder, remainder]

Such echoing of words is distracting and is easily avoided by choosing a synonym for one of them.

4.16. Enumeration

Consider the extract

> The Basic Linear Algebra Subprograms (BLAS) have several advantages. They
>
> - Lead to shorter and clearer codes.
> - Improve modularity.
> - Machine dependent optimizations can be confined to the BLAS, aiding portability, and
> - Tuned BLAS have been provided by several manufacturers.

This explanation reads badly because the entries in the list are not grammatically parallel: the preceding "they" applies only to the first two entries of the list, and the third entry is not a complete sentence, unlike the others. This is an example of bastard enumeration, so-named by Fowler [83, p. 28], who explains that in an enumeration "there must be nothing common to two or more of the items without being common to all."

4.17. False If

The if–then construct is a vital tool in expressing technical arguments, but it is sometimes used incorrectly. Consider the sentence

> If we wish to compare the solutions of $f - \lambda k(f) = 0$ and $f_n - \lambda k_n(f_n) = 0$, then Jones shows that for a wide class of nonlinear $k(f)$, $\|f - f_n\| \le c(\lambda)\|k_n(f) - k(f)\|$.

Jones's demonstration is independent of whether or not we wish to compare solutions, so the *if* is misleading: it falsely heralds a logical condition. False ifs can always be removed by rewriting:

> To compare the solutions of $f - \lambda k(f) = 0$ and $f_n - \lambda k_n(f_n) = 0$, we can use Jones's result that for a wide class of nonlinear $k(f)$, $\|f - f_n\| \le c(\lambda)\|k_n(f) - k(f)\|$.

A more confusing example is

> If we assume that rational fractions behave like almost all real numbers, a theorem of Khintchine states that the sum of the first k partial quotients will be approximately $k \log_2 k$.

The *if* appears to be a false one, because the statement of Khintchine's theorem must be independent of what the writer assumes. In fact, with knowledge of the theorem, it can be seen that the main error is in the word "states". If we change "states" to "implies" and delete "we assume that", then the sentence is correct.

The next example is an unnecessary *if*, rather than a false *if*.

> We show that if $\widehat{x}$ is the computed solution to $Lx = b$ then $(L + \Delta L)\widehat{x} = b$, where $\|\Delta L\| \le \alpha(n)\epsilon\|L\|$.

This type of construction is acceptable if used sparingly. I prefer

> We show that the computed solution $\widehat{x}$ to $Lx = b$ satisfies $(L + \Delta L)\widehat{x} = b$, where $\|\Delta L\| \le \alpha(n)\epsilon\|L\|$.

4.18. Hyphenation

As Turabian [278, p. 44] notes, the trend is not to hyphenate compound words beginning with prefixes such as multi, pre, post, non, pseudo and semi. In mathematical writing it is common to write *nonsingular*, *semidefinite* (but *semi-infinite* to avoid a double i) and *pseudorandom*. However, a hyphen is retained before a proper noun, as in *non-Euclidean*. In deciding whether to hyphenate or to combine two words as one, it is worth bearing in mind that the hyphenated form tends to be easier to read because the prefix can be seen at a glance. Readers whose first language is not English may appreciate the hyphenated form.

Compound words involving *ill* and *well* occur frequently in mathematical writing and opinions differ about their hyphenation. *The Chicago Manual of Style* [58] recommends hyphenation when a compound with *ill*, *well*, *better*, *best*, *little*, *lesser*, etc., appears before a noun, unless the compound is itself modified. The purpose of this hyphenation rule is to avoid ambiguity. Examples:

> This is an ill-posed problem *but* This problem is ill posed.
> The well-known theorem *but* The theorem is well known.
> An ill-conditioned function *but* A very ill conditioned function.
> The second-order term has a constant 2 *but* This term is of second order.

The second example is widely accepted, but many writers always hyphenate compounds involving *ill*, such as ill-conditioned, and it is hard to argue against this practice. There are some common phrases that some writers hyphenate and others do not. An example is *floating point arithmetic* (*floating-point arithmetic*).

In the phrase "we use the 1, 2 and ∞-norms", a *suspensive hyphen* is required after "1" and "2" since they are prefixes to "norm" and we need to show that they are to be linked to this word. Thus the phrase should be rewritten "we use the 1-, 2- and ∞-norms".

Notice the hyphen in the title of Halmos's best-seller *Finite-Dimensional Vector Spaces* [122]. Halmos explains in [128, p. 146] that in the original 1942 edition the hyphen was omitted, but it was added for the 1958 edition.

4.19. Linking Words

If a piece of writing is to read well there must be no abrupt changes in mood or direction from sentence to sentence within a paragraph (unless such changes are used deliberately for effect). One way to achieve a smooth flow is to use linking words. Notice how the following paragraph would be improved by adding "In particular" to the start of the second sentence and "Furthermore" to the start of the third.

> Once we move from a convex program to a general nonlinear program, matters become far more complicated. Certain topological assumptions are required to avoid pathological cases. The results apply only in a neighbourhood of a constrained minimizer, and involve convergence of subsequences of global minimizers of the barrier function.

Of course, a sequence of sentences of the form "adverb, fact" quickly becomes tiresome, so linking words should not be overused.

Here is a list of linking words and phrases, arranged according to sense. For examples of use see §5.8.

combinations. also, and, as well as, besides, both, furthermore, in addition to, likewise, moreover, similarly.

implications or explanations. as, because, conversely, due to, for example, given, in other words, in particular, in view of, it follows that, otherwise, owing to, since, specifically, that is, thus, unlike.

modifications and restrictions. although, alternatively, but, despite, except, however, in contrast, in spite of, nevertheless, of course, on the contrary, on the other hand, though, unfortunately, whereas, yet.

emphasis. actually, certainly, clearly, in fact, indeed, obviously, surely.

consequences. accordingly, consequently, hence, therefore, thus.

4.20. Misused and Ambiguous Words

Both. A common misuse of *both* is illustrated by "In Gaussian elimination we can order the inner loops 'ij' or 'ji'. Both orderings are equivalent, mathematically." *Both* means "the two together" and is redundant when the sentence already carries this implication, as in this example. It would also be incorrect to say "Both orderings produce identical results." Correct versions are "These orderings are equivalent, mathematically", "Both orderings yield the same result", or "The two orderings produce identical results." Another common misuse of *both* is misplacement when it is used with prepositional phrases. For example,

> Incorrect: "Solutions are found both in the left and right quadrants." (*Both* is followed by a preposition, *in the left*, but *and* is followed by a noun.)
>
> Correct: "'Solutions are found both in the left and in the right quadrants." (Prepositional phrases follow *both* and *and*.)
>
> Correct: "'Solutions are found in both the left and the right quadrants." (Nouns follow *both* and *and*.)

Like. Consider the sentence

> Solving triangular systems is such a common operation that it has been standardized as a subroutine, along with many other common linear algebra operations like matrix multiplication.

The word *like* incorrectly limits the choice of linear algebra operations rather than serving as an example. Replacing *like* by *such as* conveys the intended meaning. The correct use of *like* is illustrated by "The Schulz iteration is quadratically convergent, like the Newton iteration."

Problem. An overused and, at times, ambiguous word in mathematical writing is *problem*, which can refer to both the focus of a piece of work and the difficulties encountered in carrying out the work. Sentences such as "In solving this problem we encountered a number of problems" can be avoided by using a synonym for the second occurrence of "problem", or rewriting. The sentence "We describe the special problems arising when solving stiff differential equations" is ambiguous: "problems" could refer to classes of sub-problems produced by the solution process (such as nonlinear equations), or particular difficulties faced when solving the differential equations. Again, a rewrite is necessary.

Reason. In the phrase "the reason ... is because" the word *because* is redundant, since it means "for the reason that". Therefore in the sentence "The reason for the slow convergence is because α is a double root" *because* should be replaced by *that*. Similarly, in the phrase "the reason why" the word *why* can often be omitted. Thus "The reason why this question is important is that" is better written as "The reason this question is important is that" or "This question is important because".

Significant. Be careful if you use the word *significant* in mathematical writing, because to some readers it is synonymous with *statistically significant*, which carries a precise statistical meaning.

Try and, try to. The expression *try and* is frequently used in spoken English, but it is colloquial and should be replaced by *try to* in written English. Thus in the sentence "We sum the numbers in ascending order to try and minimize the effect of rounding errors" *try and* should be replaced by *try to*.

4.21. Numbers

Small integers should be spelled out when used as adjectives ("The three lemmas"), but not when used as names or numbers ("The median age is 43" or "This follows from Theorem 3"). The number 1 is a special case, for often "one" or "unity" reads equally well or better: "z has modulus one".

4.22. Omit These Words?

Here are some words and phrases whose omission often improves a sentence:

actually, very, really, currently, in fact, thing, without doubt.

The phrase "we have" can often be omitted. "Hence we have $x = 0$" should be replaced by "Hence $x = 0$." "Hence we have the following theorem" should be deleted or replaced by a sentence conveying some information (e.g., "Hence we have proved the following theorem").

4.23. Paragraphs

A standard device for making text more appetizing is to break it into small paragraphs, as is done in newspapers. The short paragraph principle is worth following in technical writing, where the complexity of the ideas makes it more important than usual for the reader to be given frequent rests. Furthermore, long paragraphs tend to give a page a heavy image that can be a visual deterrent to the potential reader. A mix of different lengths is best. Ideally, each paragraph contains a main idea or thought that separates it from its neighbours. A long paragraph that is hard to break may be indicative of convoluted thinking.

The best writers occasionally slip in one-sentence paragraphs.

4.24. Punctuation

Much can be said about punctuation, and for thorough treatments of the topic I refer the reader to the references mentioned in Chapter 2. It is worth keeping in mind Carey's explanation [52] that "the main function of punctuation is *to make perfectly clear the construction* of the written words." A few common mistakes and difficulties deserve mention here.

- In "This result is well known, see [9]" the comma should be a semicolon, which conveys a slightly longer pause. Even better is to say "This result is well known [9]." A common mistake is the let-comma-then construction: "Let x^* be a local maximum of $F(x)$, then a Taylor series expansion gives ...." The comma should be a full stop.

  Similarly, the comma should be a semicolon, or even a full stop, in this sentence: "This bound has the disadvantage that it uses a norm of X, moreover the multiplicative constant can be large when X is not a normal matrix." These errors are called "comma splices" by Gordon [111].

- In this sentence the semicolon should be a comma: "The secant method can also be used; its lack of need for derivatives being an advantage." A rough guide is to use a semicolon only where you could also use a full stop.

- In the sentence "If we use iterative refinement solutions are computed to higher accuracy", a comma is needed after "refinement", otherwise the reader may take "iterative refinement" as modifying "solutions". Another example where a comma is needed to avoid ambiguity is the sentence "However, the singularity can be removed by a change of variable."

- In sentences such as

  > Fortran 77 contains the floating point data types real, complex and double precision.
  > The output can be rotated, stretched, reduced or magnified.

we have a choice of whether or not to place a comma (called a serial comma) before the *and* or *or* that precedes the final element of the list. In some sentences a serial comma is needed to avoid ambiguity, as in the sentence "A dictionary is used to check spelling, shades of meaning, and usage", where the absence of the comma makes "shades" modify usage. Opinion differs on whether a serial comma should always be used. It is a matter of personal preference. The house styles of some publishers require serial commas.

If a list contains commas within the items, ambiguity can be avoided by using a semicolon as the list separator. Example:

> The test collection includes matrices with known inverses or known eigenvalues; ill-conditioned or rank deficient matrices; and symmetric, positive definite, orthogonal, defective, involutary, and totally positive matrices.

- The exclamation mark should be used with extreme caution in technical writing. If you are tempted to exclaim, read "!" as "shriek"; nine times out of ten you will decide a full stop is adequate. An example of correct usage is, from [217, p. 46], "When A is tridiagonal the computation of $A^{-1}u$ costs little more than the computation of Au!" The exclamation mark could be omitted, but then the reader might not realize that this is a surprising fact. Another example is, from [159, p. 42], "The chi-square table tells us, in fact, that V_2 *is much too low*: the observed values are so close to the expected values, we cannot consider the result to be random!"

- In the US, standard practice is to surround quotes by double quotation marks, with single quotation marks being used for a quote within a quote. In the UK, the reverse practice is generally used. The placement of final punctuation marks also differs: in US usage, the final punctuation is placed inside the closing quotation marks (except for "!" and "?" when they are not part of the quotation), while in UK usage it goes outside (except for "!" and "?" when they are part of the quotation). In this book, for quotations that end sentences, the end of sentence period appears outside quotation marks unless the quotation is itself a valid sentence.

- An apostrophe denotes possession for nouns (*the proof's length*) but not for pronouns (*this book is yours*). An apostrophe is used with the plural of letters and of words when the words are used without regard to their meaning: "there are three l's in the word parallel", "his prose contains too many however's." For plurals of numbers the apostrophe can be omitted: "a random matrix of 0s and 1s". For plurals of mathematical symbols or expressions the apostrophe can again be omitted provided there is no ambiguity: "these fs are all continuous", "likewise for the Z_ks", "these δs are all of order 10^{-8}."

4.25. Say It Better, Think It Gooder

The title of this section combines the titles of two papers by George Piranian [227] and Paul Halmos [126] that appeared in *The Mathematical Intelligencer* in 1982. As Gillman explains in [105], "George said good English is important. Paul said, what do you mean, good English is important?—good mathematics is more important. They are both right." While correct English usage is important, it must not be allowed to deflect you from the language-independent tasks of planning and organizing your writing.

4.26. Saying What You Mean

In technical writing you need to take great care to say what you mean. A hastily constructed sentence can have a meaning very different from the one intended. In a book review in *SIAM Review* [vol. 34, 1992, pp. 330–331] the reviewer quotes the statement "According to Theorem 1.1, a single trajectory $X(t,x)$ passes through almost every point in phase space ...." The book's author meant to say that for every point in phase space there is a unique trajectory that goes through it.

4.27. Sentence Opening

Try not to begin a sentence with *there is* or *there are*. These forms of the verb *be* make a weak start to a sentence, because they delay the appearance of the main verb (see the quote by Dixon on page 77). Sometimes these phrases can simply be deleted, as in the sentence "There are several methods that are applicable" ("Several methods are applicable"). Also worth avoiding, if possible, are "It is" openers, such as "It is clear that" and "It is interesting to note that". If you can find alternative wordings, your writing will be more fresh and lively.

4.28. Simplification

Each word or phrase in the left column below can (or, if marked with an asterisk, should) be replaced by the corresponding one in the right column. This is not an exhaustive list (see [69, Appendix 4] or [14] for many