Accuracy and Stability of Numerical Algorithms

Higham, Nicholas J.

2002

MIMS EPrint: **2006.75**

ISSN 1749-9097

Preface to Second Edition

> *We dare not lengthen this book much more,*
> *lest it be out of moderation and should*
> *stir up men's apathy because of its size.*
> — AELFRIC, schoolteacher of Cerne Abbas,
> *later Abbot of Eynsham (c. 995–1020)*

In the nearly seven years since I finished writing the first edition of this book research on the accuracy and stability of numerical algorithms has continued to flourish and mature. Our understanding of algorithms has steadily improved, and in some areas new or improved algorithms have been derived.

Three developments during this period deserve particular note. First, the widespread adoption of electronic publication of journals and the increased practice of posting technical reports and preprints on the Web have both made research results more quickly available than before. Second, the inclusion of routines from state-of-the-art numerical software libraries such as LAPACK in packages such as MATLAB* and Maple† has brought the highest-quality algorithms to a very wide audience. Third, IEEE arithmetic is now ubiquitous—indeed, it is hard to find a computer whose arithmetic does not comply with the standard.

This new edition is a major revision of the book that brings it fully up to date, expands the coverage, and includes numerous improvements to the original material. The changes reflect my own experiences in using the book, as well as suggestions received from readers.

The changes to the book can be summarized as follows.

New Chapters

- *Symmetric Indefinite and Skew-Symmetric Systems* (Chapter 11). A greatly expanded treatment is given of symmetric indefinite systems (previously contained in the chapter *Cholesky Factorization*) and a new section treats skew-symmetric systems.

- *Nonlinear Systems and Newton's Method* (Chapter 25). Results on the limiting accuracy and limiting residual of Newton's method are given under general assumptions that permit the use of extended precision in calculating residuals. The conditioning of nonlinear systems, and termination criteria for iterative methods, are also investigated.

*MATLAB is a registered trademark of The MathWorks, Inc.

†Maple is a registered trademark of Waterloo Maple Software.

New Sections

- *Fused Multiply-Add Operation* (§2.6). The advantages of this operation, which is included in the Intel IA-64 architecture, are discussed, along with some subtle issues that it raises.

- *Elementary Functions* (§2.10). We explain why it is difficult to compute elementary functions in a way that satisfies all the natural requirements, and give pointers to relevant work.

- *Matrix Polynomials* (§5.4). How to evaluate three different matrix generalizations of a scalar polynomial is discussed.

- *More Error Bounds* (§9.7). Some additional backward and forward error bounds for Gaussian elimination (GE) without pivoting are given, leading to the new result that GE is row-wise backward stable for row diagonally dominant matrices.

- *Variants of Gaussian Elimination* (§9.9). Some lesser-known variants of GE with partial pivoting are described.

- *Rank-Revealing LU Factorizations* (§9.12). This section explains why LU factorization with an appropriate pivoting strategy leads to a factorization that is usually rank revealing.

- *Parallel Inversion Methods* (§14.5). Several methods for matrix inversion on parallel machines are described, including the Schulz iteration, which is of wider interest.

- *Block* 1*-Norm Estimator* (§15.4). An improved version of the LAPACK condition estimator, implemented in MATLAB's `condest` function, is outlined.

- *Pivoting and Row-Wise Stability* (§19.4). The behaviour of Householder QR factorization for matrices whose rows are poorly scaled is examined. The backward error result herein is the only one I know that requires a particular choice of sign when constructing Householder matrices.

- *Weighted Least Squares Problems* (§20.8). Building on §19.4, an overall row-wise backward error result is given for solution of the least squares problem by Householder QR factorization with column pivoting.

- *The Equality Constrained Least Squares Problem* (§20.9). This section treats the least squares problem subject to linear equality constraints. It gives a perturbation result and describes three classes of methods (the method of weighting, null space methods, and elimination methods) and their numerical stability.

- *Extended and Mixed Precision BLAS* (§27.10). A brief description is given of these important new aids to carrying out extended precision computations in a portable way.

Other Changes

In the error analysis of QR factorization in the first edition of the book, backward error bounds were given in normwise form and in a componentwise form that essentially provided columnwise bounds. I now give just columnwise bounds, as they are the natural result of the analysis and trivially imply both normwise and componentwise bounds. The basic lemma on construction of the Householder vector has been modified so that most of the ensuing results apply for either choice of sign in constructing the vector. These and other results are expressed using the error constant $\tilde{\gamma}_n$, which replaces the more clumsy notation γ_{cn} used in the first edition (see §3.4).

Rook pivoting is a pivoting strategy that is applicable to both GE for general matrices and block $\mathrm{LDL^T}$ factorization for symmetric indefinite matrices, and it is of pedagogical interest because it is intermediate between partial pivoting and complete pivoting in both cost and stability. Rook pivoting is described in detail and its merits for practical computation are explained. A thorough discussion is given of the choice of pivoting strategy for GE and of the effects on the method of scaling. Some new error bounds are included, as well as several other results that help to provide a comprehensive picture of current understanding of GE.

This new edition has a more thorough treatment of block $\mathrm{LDL^T}$ factorization for symmetric indefinite matrices, including recent error analysis, rook pivoting, and Bunch's pivoting strategy for tridiagonal matrices. Aasen's method and Bunch's block $\mathrm{LDL^T}$ factorization method for skew-symmetric matrices are also treated.

Strengthened error analysis includes results for Gauss–Jordan elimination (Theorem 14.5, Corollary 14.7), fast solution of Vandermonde systems (Corollary 22.7), the fast Fourier transform (FFT) (Theorem 24.2), and solution of circulant linear systems via the FFT (Theorem 24.3).

All the numerical experiments have been redone in the latest version, 6.1, of MATLAB. The figures have been regenerated and their design improved, where possible. Discussions of LAPACK reflect the current release, 3.0.

A major effort has gone into updating the bibliography, with the aim of referring to the most recent results and ensuring that the latest editions of books are referenced and papers are cited in their final published form. Over 190 works published since the first edition are cited. See page 587 for a histogram that shows the distribution of dates of publication of the works cited.

In revising the book I took the opportunity to rewrite and rearrange material, improve the index, and fine tune the typesetting (in particular, using ideas of Knuth [745, 1999, Chap. 33]). Several research problems from the first edition have been solved and are now incorporated into the text, and new research problems and general problems have been added.

In small ways the emphasis of the book has been changed. For example, when the first edition was written IEEE arithmetic was not so prevalent, so a number of results were stated with the proviso that a guard digit was present. Now it is implicitly assumed throughout that the arithmetic is "well behaved" and unfortunate consequences of lack of a guard digit are given less prominence.

A final change concerns the associated MATLAB toolbox. The Test Matrix Toolbox from the first edition is superseded by the new Matrix Computation Toolbox, described in Appendix D. The new toolbox includes functions implementing a number of algorithms in the book—in particular, GE with rook pivoting and block LDL^T factorization for symmetric and skew-symmetric matrices. The toolbox should be of use for both teaching and research.

I am grateful to Bill Gragg, Beresford Parlett, Colin Percival, Siegfried Rump, Françoise Tisseur, Nick Trefethen, and Tjalling Ypma for comments that influenced the second edition.

It has been a pleasure working once again with the SIAM publication staff, in particular Linda Thiel, Sara Triller Murphy, Marianne Will, and my copy editor, Beth Gallagher.

Research leading to this book has been supported by grants from the Engineering and Physical Sciences Research Council and by a Royal Society Leverhulme Trust Senior Research Fellowship.

The tools used to prepare the book were the same as for the first edition, except that for TEX-related tasks I used MikTEX (`http://www.miktex.org/`), including its excellent YAP previewer.

Manchester Nicholas J. Higham
February 2002