
Nicholas J. Higham

University of Manchester
Manchester, England...

....
....

Accuracy and Stability of Numerical Algorithms

siam

Society for Industrial and Applied Mathematics
Philadelphia

Contents

List of Figures	xvii
List of Tables	xix
Preface	xxi
About the Dedication	xxvii
1 Principles of Finite Precision Computation	1
1.1 Notation and Background	2
1.2 Relative Error and Significant Digits	4
1.3 Sources of Errors	5
1.4 Precision Versus Accuracy	7
1.5 Backward and Forward Errors	7
1.6 Conditioning	9
1.7 Cancellation	10
1.8 Solving a Quadratic Equation	11
1.9 Computing the Sample Variance	12
1.10 Solving Linear Equations	13
1.10.1 GEPP Versus Cramer's Rule	14
1.11 Accumulation of Rounding Errors	16
1.12 Instability Without Cancellation	17
1.12.1 The Need for Pivoting	17
1.12.2 An Innocuous Calculation?	17
1.12.3 An Infinite Sum	18
1.13 Increasing the Precision	19
1.14 Cancellation of Rounding Errors	21
1.14.1 Computing $(e^x - 1)/x$	22
1.14.2 QR Factorization	24
1.15 Rounding Errors Can Be Beneficial	26
1.16 Stability of an Algorithm Depends on the Problem	27
1.17 Rounding Errors Are Not Random	29
1.18 Designing Stable Algorithms	30
1.19 Misconceptions	31

1.20	Rounding Errors in Numerical Analysis	32
1.21	Notes and References	32
	Problems	36
2	Floating Point Arithmetic	39
2.1	Floating Point Number System	40
2.2	Model of Arithmetic	44
2.3	IEEE Arithmetic	45
2.4	Aberrant Arithmetics	48
2.5	Choice of Base and Distribution of Numbers	51
2.6	Statistical Distribution of Rounding Errors	52
2.7	Alternative Number Systems	53
2.8	Accuracy Tests	54
2.9	Notes and References	56
	Problems	62
3	Basics	67
3.1	Inner and Outer Products	68
3.2	The Purpose of Rounding Error Analysis	71
3.3	Running Error Analysis	72
3.4	Notation for Error Analysis	73
3.5	Matrix Multiplication	76
3.6	Complex Arithmetic	78
3.7	Miscellany	80
3.8	Error Analysis Demystified	82
3.9	Other Approaches	83
3.10	Notes and References	84
	Problems	84
4	Summation	87
4.1	Summation Methods	88
4.2	Error Analysis	89
4.3	Compensated Summation	92
4.4	Other Summation Methods	97
4.5	Statistical Estimates of Accuracy	98
4.6	Choice of Method	98
4.7	Notes and References	100
	Problems	100
5	Polynomials	103
5.1	Horner's Method	104
5.2	Evaluating Derivatives	106
5.3	The Newton Form and Polynomial Interpolation	109

5.4	Notes and References	113
	Problems	114
6	Norms	117
6.1	Vector Norms	118
6.2	Matrix Norms	120
6.3	The Matrix p -Norm	124
6.4	Notes and References	126
	Problems	127
7	Perturbation Theory for Linear Systems	131
7.1	Normwise Analysis	132
7.2	Componentwise Analysis	134
7.3	Scaling to Minimize the Condition Number	137
7.4	The Matrix Inverse	140
7.5	Extensions	140
7.6	Numerical Stability	141
7.7	Practical Error Bounds	142
7.8	Perturbation Theory by Calculus	144
7.9	Notes and References	145
	Problems	147
8	Triangular Systems	151
8.1	Backward Error Analysis	152
8.2	Forward Error Analysis	155
8.3	Bounds for the Inverse	159
8.4	A Parallel Fan-In Algorithm	162
8.5	Notes and References	164
	8.5.1 LAPACK	166
	Problems	166
9	LU Factorization and Linear Equations	169
9.1	Gaussian Elimination	170
9.2	Error Analysis	174
9.3	The Growth Factor	177
9.4	Special Matrices	181
9.5	Tridiagonal Matrices	183
9.6	Historical Perspective	186
9.7	Scaling	191
9.8	A Posteriori Stability Tests	192
9.9	Sensitivity of the LU Factorization	194
9.10	Notes and References	195
	9.10.1 LAPACK	198

Problems	199
10 Cholesky Factorization	203
10.1 Symmetric Positive Definite Matrices	204
10.1.1 Error Analysis	205
10.2 Sensitivity of the Cholesky Factorization	209
10.3 Positive Semidefinite Matrices	210
10.3.1 Perturbation Theory	211
10.3.2 Error Analysis	214
10.4 Symmetric Indefinite Matrices and Diagonal Pivoting Method	218
10.4.1 Complete Pivoting	219
10.4.2 Partial Pivoting	221
10.5 Nonsymmetric Positive Definite Matrices	223
10.6 Notes and References	224
10.6.1 LAPACK	226
Problems	227
11 Iterative Refinement	231
11.1 Convergence of Iterative Refinement	232
11.2 Iterative Refinement Implies Stability	235
11.3 Notes and References	241
11.3.1 LAPACK	243
Problems	243
12 Block LU Factorization	245
12.1 Block Versus Partitioned LU Factorization	246
12.2 Error Analysis of Partitioned LU Factorization	248
12.3 Error Analysis of Block LU Factorization	250
12.3.1 Block Diagonal Dominance	251
12.3.2 Symmetric Positive Definite Matrices	255
12.4 Notes and References	257
12.4.1 LAPACK	258
Problems	258
13 Matrix Inversion	261
13.1 Use and Abuse of the Matrix Inverse	262
13.2 Inverting a Triangular Matrix	265
13.2.1 Unblocked Methods	265
13.2.2 Block Methods	267
13.3 Inverting a Full Matrix by LU Factorization	270
13.3.1 Method A	270
13.3.2 Method B	271

13.3.3	Method C	272
13.3.4	Method D	273
13.3.5	Summary	275
13.4	Gauss-Jordan Elimination	275
13.5	The Determinant	281
13.5.1	Hyman's Method	282
13.6	Notes and References	283
13.6.1	LAPACK	285
	Problems	285
14	Condition Number Estimation	289
14.1	How to Estimate Componentwise Condition Numbers	290
14.2	The p -Norm Power Method	291
14.3	LAPACK 1-Norm Estimator	294
14.4	Other Condition Estimators	297
14.5	Condition Numbers of Tridiagonal Matrices	301
14.6	Notes and References	304
14.6.1	LAPACK	306
	Problems	306
15	The Sylvester Equation	309
15.1	Solving the Sylvester Equation	311
15.2	Backward Error	313
15.2.1	The Lyapunov Equation	316
15.3	Perturbation Result	318
15.4	Practical Error Bounds	320
15.5	Extensions	321
15.6	Notes and References	322
15.6.1	LAPACK	324
	Problems	324
16	Stationary Iterative Methods	325
16.1	Survey of Error Analysis	327
16.2	Forward Error Analysis	329
16.2.1	Jacobi's Method	332
16.2.2	Successive Overrelaxation	334
16.3	Backward Error Analysis	334
16.4	Singular Systems	336
16.4.1	Theoretical Background	336
16.4.2	Forward Error Analysis	338
16.5	Stopping an Iterative Method	341
16.6	Notes and References	343

Problems	343
17 Matrix Powers	345
17.1 Matrix Powers in Exact Arithmetic	346
17.2 Bounds for Finite Precision Arithmetic	353
17.3 Application to Stationary Iteration	358
17.4 Notes and References	358
Problems	359
18 QR Factorization	361
18.1 Householder Transformations	362
18.2 QR Factorization	363
18.3 Error Analysis of Householder Computations	364
18.4 Aggregated Householder Transformations	370
18.5 Givens Rotations	371
18.6 Iterative Refinement	375
18.7 Gram-Schmidt Orthogonalization	376
18.8 Sensitivity of the QR Factorization	381
18.9 Notes and References	383
18.9.1 LAPACK	386
Problems	387
19 The Least Squares Problem	391
19.1 Perturbation Theory	392
19.2 Solution by QR Factorization	395
19.3 Solution by the Modified Gram-Schmidt Method	396
19.4 The Normal Equations	397
19.5 Iterative Refinement	399
19.6 The Seminormal Equations	403
19.7 Backward Error	404
19.8 Proof of Wedin's Theorem	407
19.9 Notes and References	409
19.9.1 LAPACK	412
Problems	412
20 Underdetermined Systems	415
20.1 Solution Methods	416
20.2 Perturbation Theory	417
20.3 Error Analysis	419
20.4 Notes and References	422
20.4.1 LAPACK	423
Problems	423

21 Vandermonde Systems	425
21.1 Matrix Inversion	426
21.2 Primal and Dual Systems	428
21.3 Stability	434
21.3.1 Forward Error	435
21.3.2 Residual	437
21.3.3 Dealing with Instability	438
21.4 Notes and References	440
Problems	441
22 Fast Matrix Multiplication	445
22.1 Methods	446
22.2 Error Analysis	450
22.2.1 Winograd's Method	451
22.2.2 Strassen's Method	452
22.2.3 Bilinear Noncommutative Algorithms	456
22.2.4 The 3M Method	458
22.3 Notes and References	459
Problems	461
23 The Fast Fourier Transform and Applications	465
23.1 The Fast Fourier Transform	466
23.2 Circulant Linear Systems	468
23.3 Notes and References	470
Problems	471
24 Automatic Error Analysis	473
24.1 Exploiting Direct Search Optimization	474
24.2 Direct Search Methods	477
24.3 Examples of Direct Search	479
24.3.1 Condition Estimation	480
24.3.2 Fast Matrix Inversion	481
24.3.3 Solving a Cubic	483
24.4 Interval Analysis	485
24.5 Other Work	487
24.6 Notes and References	489
Problems	490
25 Software Issues in Floating Point Arithmetic	491
25.1 Exploiting IEEE Arithmetic	492
25.2 Subtleties of Floating Point Arithmetic	495
25.3 Cray Peculiarities	496
25.4 Compilers	497

25.5	Determining Properties of Floating Point Arithmetic	497
25.6	Testing a Floating Point Arithmetic	498
25.7	Portability	499
	25.7.1 Arithmetic Parameters	499
	25.7.2 2×2 Problems in LAPACK	500
	25.7.3 Numerical Constants	501
	25.7.4 Models of Floating Point Arithmetic	501
25.8	Avoiding Underflow and Overflow	502
25.9	Multiple Precision Arithmetic	504
25.10	Patriot Missile Software Problem	506
25.11	Notes and References	507
	Problems	508
26	A Gallery of Test Matrices	513
26.1	The Hilbert and Cauchy Matrices	514
26.2	Random Matrices	517
26.3	"Randsvd" Matrices	519
26.4	The Pascal Matrix	520
26.5	Tridiagonal Toeplitz Matrices	524
26.6	Companion Matrices	525
26.7	Notes and References	526
	26.7.1 LAPACK	527
	Problems	527
A	Solutions to Problems	529
B	Singular Value Decomposition, M-Matrices	579
B.1	Singular Value Decomposition	580
B.2	M -Matrices	580
C	Acquiring Software	581
C.1	Internet	582
C.2	Netlib	582
C.3	MATLAB	583
C.4	NAG Library and FTN90 Compiler	583
D	Program Libraries	585
D.1	Basic Linear Algebra Subprograms	586
D.2	EISPACK	587
D.3	LINPACK	587
D.4	LAPACK	587
	D.4.1 Structure of LAPACK	588
E	The Test Matrix Toolbox	591

Bibliography	595
Name Index	665
Subject Index	675