Componentwise Error Analysis for Stationary Iterative Methods*

Nicholas J. Higham† Philip A. Knight‡

June 1, 1992

Abstract

How small can a stationary iterative method for solving a linear system $Ax = b$ make the error and the residual in the presence of rounding errors? We give a componentwise error analysis that provides an answer to this question and we examine the implications for numerical stability. The Jacobi, Gauss-Seidel and successive over-relaxation methods are all found to be forward stable in a componentwise sense and backward stable in a normwise sense, provided certain conditions are satisfied that involve the matrix, its splitting, and the computed iterates. We show that the stronger property of componentwise backward stability can be achieved using one step of iterative refinement in fixed precision, under suitable assumptions.

Key words. stationary iteration, Jacobi method, Gauss-Seidel method, successive over-relaxation, error analysis, numerical stability.

AMS subject classifications. primary 65F10, 65G05

*An earlier version of this paper appears as: N. J. Higham and P. A. Knight. Componentwise error analysis for stationary iterative methods. In C. D. Meyer and R. J. Plemmons, editors, *Linear Algebra, Markov Chains, and Queueing Models*, volume 48 of *IMA Volumes in Mathematics and its Applications*, pages 29–46. Springer-Verlag, New York, 1993.

†Nuffield Science Research Fellow. Department of Mathematics, University of Manchester, Manchester, M13 9PL, England (na.nhigham@na-net.ornl.gov).

‡Supported by a SERC Research Studentship. Department of Mathematics, University of Manchester, Manchester, M13 9PL, England (na.pknight@na-net.ornl.gov).

1 Introduction

The effect of rounding errors on LU and QR factorization methods for solving linear systems is well understood. Various backward and forward error bounds are known that are informative and easy to interpret. Many of the bounds can be approximate equalities, to within factors depending on the problem dimension, and so they give a useful guide to the accuracy that can be attained in practice. (Note that it is still not well understood why the growth factor for LU factorization with partial pivoting is usually small, but progress towards explaining this phenomenon has been made recently [24].)

In contrast, there is little published error analysis for iterative methods. This is surprising, since iterative methods for computer solution of $Ax = b$ have at least as long a history as direct methods [28]. One reason for the paucity of error analysis may be that in many applications accuracy requirements are modest and are satisfied without difficulty. Nevertheless, we believe the following question is an important one, and in this work we attempt to answer it for a particular class of methods.

> How accurate a solution can we obtain using an iterative method in floating point arithmetic?

To be more precise, how small can we guarantee that the backward or forward error will be over all iterations $k = 1, 2, \ldots$? Without an answer to this question we cannot be sure that a convergence test of the form $\|b - A\widehat{x}_k\| \leq \epsilon$ (say) will ever be satisfied, for any given value of $\epsilon < \|b - Ax_0\|$!

As an indication of the potentially devastating effects of rounding errors we present an example constructed and discussed by Hammarling and Wilkinson [12]. Here, A is the 100×100 lower bidiagonal matrix with $a_{ii} \equiv 1.5$ and $a_{i,i-1} \equiv 1$, and $b_i \equiv 2.5$. The successive over-relaxation (SOR) method is applied in Matlab with parameter $\omega = 1.5$, starting with the rounded version of the exact solution x, given by $x_i = 1 - (-2/3)^i$. The forward errors $\|\widehat{x}_k - x\|_\infty / \|x\|_\infty$ and the normwise backward errors $\eta_\infty(\widehat{x}_k)$ (defined in (1)) are plotted in Figure 1. The SOR method converges in exact arithmetic, since the iteration matrix has spectral radius $1/2$, but in the presence of rounding errors it diverges. The iterate $\widehat{x}_{238}$ has a largest element of order 10^{13}, $\widehat{x}_{k+2} \equiv \widehat{x}_k$ for $k \geq 238$, and for $k > 100$, $\widehat{x}_k(60\colon 100) \approx (-1)^k \widehat{x}_{100}(60\colon 100)$. The divergence is not a result of ill-conditioning of A, since $\kappa_\infty(A) \approx 5$. The reason for the initial rapid growth of the errors in this example is that the iteration matrix is far from normal; this allows the norms of the powers to

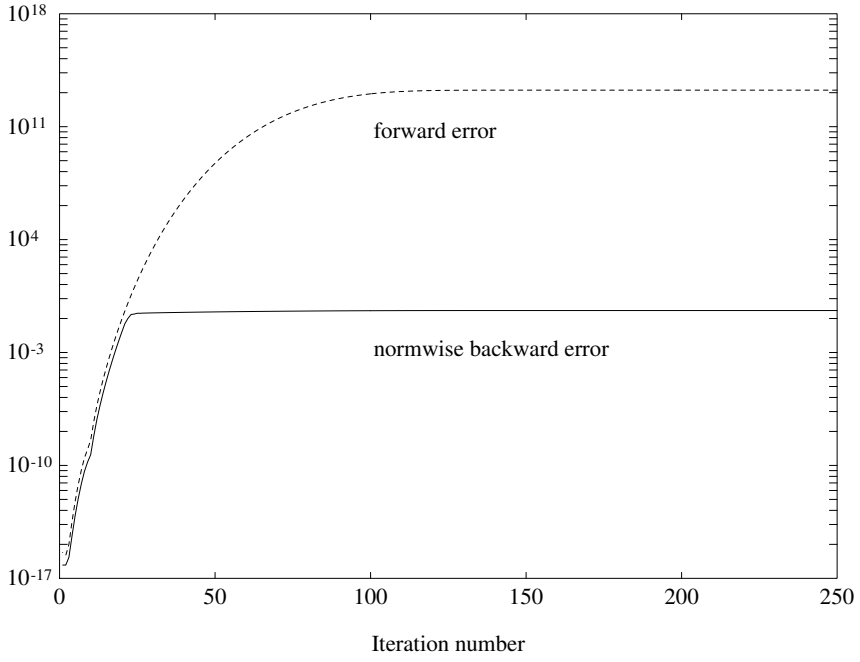

Figure 1: SOR iteration

become very large before they ultimately decay by a factor $\approx 1/2$ with each successive power. The effect of rounding errors in this example is to cause the forward error curve in Figure 1 to level off near $k = 100$, instead of decaying to zero as it would in exact arithmetic. More insight into the initial behaviour of the errors can be obtained using the notion of pseudo-eigenvalues [23].

To establish what we should try to prove, we review some normwise and componentwise backward error results and perturbation theory. If y is an approximate solution to $Ax = b$ then the *normwise (relative) backward error* is

$$\eta(y) = \min\{\epsilon : (A + \Delta A)y = b + \Delta b, \ \ \|\Delta A\| \le \epsilon\|A\|, \ \ \|\Delta b\| \le \epsilon\|b\|\}, \quad (1)$$

where $\|\cdot\|$ denotes any vector norm and the corresponding subordinate matrix norm. Rigal and Gaches [22] show that $\eta(y)$ is given by the explicit formula

$$\eta(y) = \frac{\|r\|}{\|A\|\|y\| + \|b\|}. \quad (2)$$

The forward error of y can be bounded using the standard perturbation result

$$\frac{\|x-y\|}{\|x\|} \le \kappa(A)\eta(y) + O(\eta(y)^2), \tag{3}$$

where $\kappa(A) = \|A\|\|A^{-1}\|$ is the matrix condition number.

The *componentwise (relative) backward error* is defined by

$$\omega(y) = \min\{\epsilon : (A+\Delta A)y = b + \Delta b, \;\; |\Delta A| \le \epsilon|A|, \;\; |\Delta b| \le \epsilon|b|\},$$

where the absolute values and inequalities are interpreted componentwise. This is a more stringent measure of backward error than the normwise measure, since each perturbation Δa_{ij} and Δb_i is measured relative to the entry it perturbs. The constraints ensure that $a_{ij} = 0 \Rightarrow \Delta a_{ij} = 0$, and similarly for b. This latter property is particularly attractive in the context of iterative solvers, where A is usually sparse and it may not be meaningful to perturb the zero entries of A [1]. Oettli and Prager [21] derive the convenient formula

$$\omega(y) = \max_i \frac{|b-Ay|_i}{(|A||y| + |b|)_i},$$

where $\xi/0$ is interpreted as zero if $\xi = 0$ and infinity otherwise.

A perturbation bound involving componentwise quantities is, for any monotonic norm[1],

$$\begin{aligned} \frac{\|y-x\|}{\|x\|} &\le \frac{\| \, |A^{-1}|(|A||x| + |b|) \, \|}{\|x\|}\, \omega(y) + O(\omega(y)^2) \\ &\le 2\mathrm{cond}(A,x)\omega(y) + O(\omega(y)^2), \end{aligned} \tag{4}$$

where $\mathrm{cond}(A,x) = \| \, |A^{-1}||A||x| \, \|/\|x\|$. The condition number $\mathrm{cond}(A,x)$ is independent of row scalings $A \to \mathrm{diag}(d_i)A$, and it satisfies $\mathrm{cond}_\infty(A,x) \le \kappa_\infty(A)$, so the bound (4) is potentially much smaller than (3).

Ideally a numerical method for solving $Ax = b$ will produce a computed solution $\widehat{x}$ that satisfies $\omega(\widehat{x}) = O(u)$, where u is the unit roundoff. Such a method is said to be *componentwise backward stable*. A method that is not componentwise backward stable may still satisfy a bound of the form

$$\frac{\|\widehat{x}-x\|}{\|x\|} \le c_n \mathrm{cond}(A,x)u + O(u^2), \tag{5}$$

[1] A monotonic norm is one for which $|x| \le |y| \Rightarrow \|x\| \le \|y\|$ [18, p. 285].

where c_n is a constant; this is the same type of forward error bound as holds for a method that is componentwise backward stable. We will call a method for which (5) holds *componentwise forward stable*. Similarly, if $\eta(\widehat{x}) = O(u)$ we call the method *normwise backward stable*, and if (5) holds with $\mathrm{cond}(A, x)$ replaced by $\kappa(A)$ we call the method *normwise forward stable*.

In this paper we analyse stationary iterative methods: those that iterate according to $Mx_{k+1} = Nx_k + b$, using the splitting $A = M - N$. Even though this is a relatively straightforward class of iterative methods (as regards determining the rate of convergence in exact arithmetic, say), our error analysis is not as concise as that which can be done for LU factorization (see, e.g., [9, Sec. 3.3], [15]), and our conclusions are less clear-cut. It seems inherently more difficult to obtain useful rounding error bounds for iterative methods than it is for direct methods. An important feature of our analysis is that we use sharp inequalities wherever possible, so as to obtain the best possible bounds. Our results should be of use in guiding the choice of stopping criterion for an iterative solver; cf. the discussion in [2].

In section 2 we derive the basic recurrences for the error and residual. Forward error bounds are developed in section 3, and these are specialised to the Jacobi, Gauss-Seidel and SOR methods in section 4. We find that all three methods are componentwise forward stable if a certain product $c(A)\theta_x$ is not too large (and ω is not too close to zero for the SOR iteration). Here, $c(A) \geq 1$ (defined in (23)) depends on the matrix and the splitting, and $\theta_x \geq 1$ (defined in (20)) describes how "well-behaved" the iterates are. For each method, $c(A) = 1$ if A is an M-matrix (assuming $0 \leq \omega \leq 1$ for the SOR iteration), and $c(A)$ can be expected to be of modest size in many applications. This forward stability result is quite strong, for apart from triangular matrices and certain classes of tridiagonal matrices [14], the only class of matrices we know for which LU factorization is guaranteed to be componentwise forward stable is the class of totally nonnegative matrices [6].

In section 5 we derive bounds for the residual. We show that any stationary iterative method is normwise backward stable under conditions which include the requirement that the spectral radius of NM^{-1} is not too close to 1. Unfortunately, it does not seem possible to prove that a small componentwise backward error will be obtained under any reasonable assumptions. However, we explain why one step of iterative refinement in fixed precision does lead to a small componentwise backward error provided certain assumptions are satisfied.

We briefly survey existing error analyses for iterative methods. For symmetric positive definite systems, Golub [8] derives both statistical and nonstatistical bounds for the forward error and residual of the Richardson method. Benschop and Ratz [4] give a statistical analysis of the effect of rounding errors on stationary iteration, under the assumption that the rounding errors are independent random variables with zero mean. Lynn [20] presents a statistical analysis for the SOR method with a symmetric positive definite matrix.

Hammarling and Wilkinson [12] give a normwise error analysis for the SOR method. With the aid of numerical examples, they emphasise that while it is the spectral radius of the iteration matrix $M^{-1}N$ that determines the asymptotic rate of convergence, it is the norms of the powers of this matrix that govern the behaviour of the iteration in the early stages. This point is also elucidated by Trefethen [23], using the tool of pseudospectra.

Dennis and Walker [7] obtain bounds for $\|x-\widehat{x}_{k+1}\|/\|x-\widehat{x}_k\|$ for stationary iteration as a special case of error analysis of quasi-Newton methods for nonlinear systems. The bounds in [7] do not readily yield information about normwise or componentwise forward stability.

Bollen [5] analyses the class of "descent methods" for solving $Ax = b$, where A is required to be symmetric positive definite; these are obtained by iteratively using exact line searches to minimize the quadratic function $F(x) = (A^{-1}b-x)^T A(A^{-1}b-x)$. The choice of search direction $p_k = b-Ax_k \equiv r_k$ yields the steepest descent method, while $p_k = e_j$ (unit vector), where $|r_k|_j = \|r_k\|_\infty$, gives the Gauss-Southwell method. Bollen shows that both these methods are normwise backward stable as long as a condition of the form $c_n\kappa(A)u < 1$ holds. If the p_k are cyclically chosen to be the unit vectors e_1, e_2,..., e_n then the Gauss-Seidel method results, but unfortunately no results specific to this method are given in [5].

Woźniakowski [25] shows that the Chebyshev semi-iterative method is normwise forward stable but not normwise backward stable. In [27] Woźniakowski analyses a class of conjugate gradient algorithms (which does not include the usual conjugate gradient method). He obtains a forward error bound proportional to $\kappa(A)^{3/2}$ and a residual bound proportional to $\kappa(A)$, from which neither backward nor forward normwise stability can be deduced. We note that as part of the analysis in [27] Woźniakowski obtains a residual bound for the steepest descent method that is proportional to $\kappa(A)$, and is therefore much weaker than the bound obtained by Bollen [5].

Zawilski [29] shows that the cyclic Richardson method for symmetric

positive definite systems is normwise forward stable provided the parameters are suitably ordered. He also derives a sharp bound for the residual that includes a factor $\kappa(A)$, and which therefore shows that the method is not normwise backward stable.

Arioli and Romani [3] give a statistical error analysis of stationary iterative methods. They investigate the relations between a statistically defined asymptotic stability factor, ill-conditioning of $M^{-1}A$, where $A = M - N$ is the splitting, and the rate of convergence.

Greenbaum [10] presents a detailed error analysis of the conjugate gradient method, but her concern is with the rate of convergence rather than the attainable accuracy. An excellent survey of work concerned with the effects of rounding error on the conjugate gradient method (and the Lanczos method) is given in the introduction of [11].

The work most closely related to ours is [26], wherein Woźniakowski gives a normwise error analysis of stationary iterative methods. Some of the assumptions in [26] are difficult to justify, as we explain in section 3.

Finally, we mention that extension of the analysis reported here to singular systems is described in [17].

2 Basic Equations

We are concerned with the iteration

$$Mx_{k+1} = Nx_k + b,$$

where $A = M - N \in \mathbb{R}^{n\times n}$ is nonsingular, and M is nonsingular. We assume throughout that the spectral radius $\rho(M^{-1}N) < 1$, so that in exact arithmetic the iteration converges for any starting vector. We assume that x_{k+1} is computed by forming $Nx_k + b$ and then solving a linear system with M. The computed vectors $\widehat{x}_k$ therefore satisfy

$$(M + \Delta M_{k+1})\widehat{x}_{k+1} = N\widehat{x}_k + b + f_k, \tag{6}$$

where f_k is the error in forming $fl(N\widehat{x}_k + b)$ and ΔM_{k+1} represents the error in solving the linear system. We will use the following model of floating point arithmetic, where u is the unit roundoff:

$$\begin{aligned} fl(x \pm y) &= x(1+\alpha) \pm y(1+\beta), \qquad |\alpha|, |\beta| \le u, \\ fl(x \operatorname{op} y) &= (x \operatorname{op} y)(1+\delta), \qquad |\delta| \le u, \quad \operatorname{op} = *, /. \end{aligned} \tag{7}$$

This model is valid for machines that do not use a guard digit in addition/subtraction. Under this model we have the standard result that

$$|f_k| \leq c_n u(|N||\widehat{x}_k| + |b|), \tag{8}$$

where c_n is a constant of order n. If N has at most m nonzeros per row then c_n can be taken to be of order m, but sparsity does not otherwise affect the analysis below.

We will assume that

$$|\Delta M_{k+1}| \leq c_n' u|M|. \tag{9}$$

This is valid if M is triangular (see, e.g., [13]), which is the case for the Jacobi, Gauss-Seidel and SOR iterations. It also holds if M is (a) totally nonnegative [6] or (b) tridiagonal and either symmetric positive definite, an M-matrix, or diagonally dominant [14], assuming in both cases (a) and (b) that Gaussian elimination without pivoting is used.

We now derive recurrences for the residuals $r_k = b - A\widehat{x}_k$ and errors $e_k = x - \widehat{x}_k$. Using (6),

$$\begin{aligned} r_{k+1} &= b - (M-N)\widehat{x}_{k+1} \\ &= \Delta M_{k+1}\widehat{x}_{k+1} - N(\widehat{x}_k - \widehat{x}_{k+1}) - f_k. \end{aligned} \tag{10}$$

From (6) again,

$$\begin{aligned} r_k &= b - (M-N)\widehat{x}_k \\ &= -M\widehat{x}_k + (M + \Delta M_{k+1})\widehat{x}_{k+1} - f_k, \end{aligned}$$

so that

$$\widehat{x}_k - \widehat{x}_{k+1} = M^{-1}(\Delta M_{k+1}\widehat{x}_{k+1} - f_k - r_k).$$

Substituting into (10) gives

$$r_{k+1} = NM^{-1}r_k + (I - NM^{-1})(\Delta M_{k+1}\widehat{x}_{k+1} - f_k), \tag{11}$$

which is our recurrence for the residuals. Since $e_k = A^{-1}r_k$, we have from (11)

$$e_{k+1} = A^{-1}NM^{-1}Ae_k + A^{-1}(I - NM^{-1})(\Delta M_{k+1}\widehat{x}_{k+1} - f_k).$$

Simple manipulation shows that $A^{-1}NM^{-1}A = M^{-1}N$ and $A^{-1}(I - NM^{-1}) = M^{-1}$, so the recurrence for the errors is

$$e_{k+1} = M^{-1}Ne_k + M^{-1}(\Delta M_{k+1}\widehat{x}_{k+1} - f_k). \tag{12}$$

The following lemmas are needed in the next section.

Lemma 2.1 *(a) If $|B| \le E \in \mathbb{R}^{n\times n}$ then $|Bx| \le E|x|$, with equality for some B with $|B| = E$.*

(b) If $B_j \in \mathbb{R}^{n\times n}$ and $|x_j| \le h_j$, $j = 0, \ldots, m$, then $\sum_{j=0}^m B_j x_j \le \sum_{j=0}^m |B_j| h_j$, and there is equality in the ith component for some $x_0, \ldots, x_m$ with $|x_j| = h_j$. In particular,

$$\|\sum_{j=0}^m B_j x_j\|_\infty \le \|\sum_{j=0}^m |B_j| h_j\|_\infty,$$

and equality is attainable.

Proof. (a) The inequality is straightforward. Equality is obtained when $b_{ij} = \mathrm{sign}(x_j) e_{ij}$. (b) The vector inequality is straightforward. Equality is obtained in the ith component for $x_j(k) = \mathrm{sign}(b_{ik}^{(j)}) h_j(k)$, where $B_j = (b_{ik}^{(j)})$. The norm results follows easily. ∎

Lemma 2.2 *If $B \in \mathbb{R}^{n\times n}$ and $\rho(B) < 1$, then the series $\sum_{k=0}^\infty |B^k|$ and $\sum_{k=0}^\infty \|B^k\|$ are both convergent, where $\|\cdot\|$ is any consistent norm.*

Proof. Since $\rho(B) < 1$, a standard result [18, Lemma 5.6.10] guarantees the existence of a norm $\|\cdot\|_\rho$ for which $\|B\|_\rho < 1$. The series $\sum_{k=0}^\infty \|B^k\|_\rho \le \sum_{k=0}^\infty \|B\|_\rho^k = (1 - \|B\|_\rho)^{-1}$ is clearly convergent, and so, by the equivalence of norms, $\sum_{k=0}^\infty \|B^k\|$ is convergent for any norm.

Since $(|B^k|)_{ij} \le \|B^k\|_\infty$, the convergence of $\sum_{k=0}^\infty \|B^k\|_\infty$ ensures that of $\sum_{k=0}^\infty |B^k|$. (The convergence of $\sum_{k=0}^\infty |B^k|$ can also be proved directly using the Jordan canonical form.) ∎

3 Forward Error Analysis

The basic equation from which we work is (12), which we write as

$$e_{k+1} = G e_k + M^{-1}\xi_k, \tag{13}$$

where $G = M^{-1}N$ and

$$\xi_k = \Delta M_{k+1} \widehat{x}_{k+1} - f_k. \tag{14}$$

A componentwise bound for ξ_k is, from (8) and (9),

$$|\xi_k| \le d_n u(|M||\widehat{x}_{k+1}| + |N||\widehat{x}_k| + |b|) \equiv \mu_k, \tag{15}$$

where $d_n = \max(c_n, c_n')$, and this bound is sharp modulo the multiplicative constant d_n, as we now explain. Note that each element of M, N and b takes part in at least one floating point operation. Using Lemma 2.1 (a) it is easy to see that if we associate with each element of M, N and b a single rounding error of maximal modulus and the appropriate sign, equality is attained in (8), (9) and (15), modulo the multiplicative constants c_n, c_n' and d_n. In other words, there exists a set of rounding errors for which there is equality in (15) if we ignore the constant d_n.

Now we return to the recurrence (13), which has the solution

$$e_{m+1} = G^{m+1}e_0 + \sum_{k=0}^{m} G^k M^{-1}\xi_{m-k}.$$

The first term, $G^{m+1}e_0$, is the error of the iteration in exact arithmetic. This term tends to zero as $m \to \infty$, since $\rho(G) < 1$, so Lemma 2.1 (b) shows that each component of the following inequality is sharp for large m:

$$|e_{m+1}| \le |G^{m+1}e_0| + \sum_{k=0}^{m} |G^k M^{-1}|\,\mu_{m-k}, \tag{16}$$

where μ_k is the sharp bound for ξ_k defined in (15). As $m \to \infty$ the accuracy that can be guaranteed by the analysis is determined by the last term in (16), and it is this term on which the rest of the analysis focuses.

At this point we can proceed by using further componentwise inequalities or by using norms. First, we consider the norm approach. In place of (16) we use the normwise bound

$$\|e_{m+1}\|_\infty \le \|G^{m+1}e_0\|_\infty + \|\sum_{k=0}^{m} |G^k M^{-1}|\,\mu_{m-k}\,\|_\infty, \tag{17}$$

which is sharp for large m, in view of Lemma 2.1 (b). Defining

$$\gamma_x = \sup_k \frac{\|\widehat{x}_k\|_\infty}{\|x\|_\infty},$$

we obtain from (17) and (15)

$$\begin{aligned}
\|e_{m+1}\|_\infty &\le \|G^{m+1}e_0\|_\infty + \max_{0\le k\le m} \|\mu_k\|_\infty \sum_{k=0}^{m} \|G^k M^{-1}\|_\infty \\
&\le \|G^{m+1}e_0\|_\infty + d_n u(1+\gamma_x)(\|M\|_\infty + \|N\|_\infty)\|x\|_\infty \sum_{k=0}^{\infty} \|G^k M^{-1}\|_\infty
\end{aligned} \tag{18}$$

where the existence of the sum is assured by Lemma 2.2, since $\rho(G) < 1$. This is similar to a result in the analysis of Woźniakowski [26] (see (3.11) and (3.12) therein). Woźniakowski's result has $\|x\|$ in place of $\sup_k \|\widehat{x}_k\|$ (i.e., $\gamma_x \equiv 1$), which we believe is erroneous, despite an extra "$+O(u^2)$" term. (Like (18), the bounds for the SOR iteration in [12] also contain a factor $\sup_k \|\widehat{x}_k\|$.) Also, Woźniakowski's bound contains $\|G^k\|$ rather than $\|G^k M^{-1}\|$; this is because he assumes $\widehat{x}_{k+1}$ is computed as

$$\widehat{x}_{k+1} = fl(H\widehat{x}_k+h) = (H+\Delta H)\widehat{x}_k+h+\Delta h, \quad \|\Delta H\| \le c_n u\|H\|, \quad \|\Delta h\| \le c_n' u\|h\|. \tag{19}$$

Comparison with (6)–(9) shows that (19) is not valid if the iteration is implemented in the natural way described in section 2. Woźniakowski assumes that the vectors ξ_k (defined slightly differently in his analysis) are arbitrary subject to a bound on $\|\xi_k\|$, and by taking the ξ_k to be eigenvectors of G he shows that his bound analogous to (18) is sharp when G is symmetric. However, under our assumptions on how x_{k+1} is computed, ξ_k satisfies (15), and the structure imposed by this inequality may preclude ξ_k from being an eigenvector of G. (Moreover, G is almost always unsymmetric for the SOR method). The bound for $\|e_{m+1}\|_\infty$ in (17) cannot be weakened without losing the sharpness.

If $\|G\|_\infty = \|M^{-1}N\|_\infty = q < 1$ then (18) yields

$$\|e_{m+1}\|_\infty \le \|G^{m+1}e_0\|_\infty + d_n u(1+\gamma_x)(\|M\|_\infty + \|N\|_\infty)\|x\|_\infty \frac{\|M^{-1}\|_\infty}{1-q}.$$

Thus if q is not too close to 1 ($q \le 0.9$, say), and γ_x and $\|M^{-1}\|_\infty$ are not too large, this bound guarantees a small forward error.

Of more interest for us is the following componentwise development of (16). Defining

$$\theta_x = \sup_k \max_{1\le i\le n} \left(\frac{|\widehat{x}_k|_i}{|x_i|} \right), \tag{20}$$

so that $|\widehat{x}_k| \le \theta_x |x|$ for all k, we have from (15),

$$|\mu_k| \le d_n u(1+\theta_x)(|M| + |N|)|x|. \tag{21}$$

Hence (16) yields

$$|e_{m+1}| \le |G^{m+1}e_0| + d_n u(1+\theta_x)(\sum_{k=0}^{\infty} |G^k M^{-1}|)(|M| + |N|)|x|. \tag{22}$$

Again, the existence of the sum is assured by Lemma 2.2, since $\rho(G) = \rho(M^{-1}N) < 1$. Since $A = M - N = M(I - M^{-1}N)$ we have

$$A^{-1} = (\sum_{k=0}^{\infty} (M^{-1}N)^k) M^{-1}.$$

The sum in (22) is clearly an upper bound for $|A^{-1}|$. Defining $c(A) \geq 1$ by

$$c(A) = \min\{\epsilon : \sum_{k=0}^{\infty} |(M^{-1}N)^k M^{-1}| \leq \epsilon \, |\sum_{k=0}^{\infty} (M^{-1}N)^k M^{-1}| = \epsilon \, |A^{-1}|\}, \quad (23)$$

we have our final bound

$$|e_{m+1}| \leq |G^{m+1} e_0| + d_n u (1 + \theta_x) c(A) |A^{-1}| (|M| + |N|) |x|. \quad (24)$$

Unlike (16), this bound is not sharp in general, but it is optimal in the sense that it is the smallest bound that can be used to assess the componentwise forward stability.

An interesting feature of stationary iteration methods is that if the elements of M and N are linear combinations of those of A, then any scaling of the form $Ax = b \rightarrow D_1 A D_2 \cdot D_2^{-1} x = D_1^{-1} b$ (D_i diagonal) leaves the eigenvalues of $M^{-1}N$ unchanged; hence the asymptotic convergence rate is independent of row and column scaling. This scale independence applies to the Jacobi and SOR iterations, but not, for example, to the stationary Richardson iteration, for which $M = I$. One of the benefits of doing a componentwise analysis is that under the above assumptions on M and N the bound (24) largely shares the scale independence. In (24) the scalar $c(A)$ is independent of the row and column scaling of A, and the term $|A^{-1}|(|M|+|N|)|x|$ scales in the same way as x. Furthermore, θ_x can be expected to depend only mildly on the row and column scaling, because the bounds in (8) and (9) for the rounding error terms have the correct scaling properties.

What can be said about $c(A)$? In general, it can be arbitrarily large. Indeed, $c(A)$ is infinite for the Jacobi and Gauss-Seidel iterations for any $n \geq 3$ if A is the symmetric positive definite matrix with $a_{ij} = \min(i, j)$, because A^{-1} is tridiagonal and $(M^{-1}N)^k M^{-1}$ is not.

If M^{-1} and $M^{-1}N$ both have nonnegative elements then $c(A) = 1$; as we will see in the next section, this condition holds in some important instances.

Some further insight into $c(A)$ can be obtained by examining the cases where $M^{-1}N$ has rank 1 or is diagonal. The rank 1 case is motivated by

the fact that if $B \in \mathbb{R}^{n\times n}$ has a unique eigenvalue λ of largest modulus then $B^k \approx \lambda^k x y^T$, where $Bx = \lambda x$ and $y^T B = \lambda y^T$ with $y^T x = 1$. If we set $(M^{-1}N)^k M^{-1} \equiv \lambda^k x y^T M^{-1} = \lambda^k x z^T$, where $|\lambda| < 1$, then

$$c(A) = \min\{\epsilon : \sum_{k=0}^{\infty} |\lambda^k x z^T| \le \epsilon \, |\sum_{k=0}^{\infty} \lambda^k x z^T|\},$$

or $\sum_{k=0}^{\infty} |\lambda^k| = c(A) |\sum_{k=0}^{\infty} \lambda^k|$, that is, $(1-|\lambda|)^{-1} = c(A)(1-\lambda)^{-1}$. Hence $c(A) = (1-\lambda)/(1-|\lambda|)$, and so $c(A)$ is of modest size unless $M^{-1}N$ has an eigenvalue close to -1. If $M^{-1}N \in \mathbf{C}^{n\times n}$ is diagonal with eigenvalues λ_i then it is easy to show that $c(A) = \max_i |1-\lambda_i|/(1-|\lambda_i|)$, so $c(A)$ can be large only if $\rho(M^{-1}N)$ is close to 1. Although $M^{-1}N$ cannot be diagonal for the Jacobi or Gauss-Seidel methods, this formula can be taken as being indicative of the size of $c(A)$ when $M^{-1}N$ is diagonalizable with a well-conditioned matrix of eigenvectors. These considerations suggest the heuristic inequality, for general A,

$$c(A) \ge \max_i \frac{|1-\lambda_i|}{1-|\lambda_i|}, \qquad \lambda_i = \lambda_i(M^{-1}N). \tag{25}$$

In practical problems where stationary iteration is used we would expect $c(A)$ to be of modest size ($O(n)$, say), for two reasons. First, to achieve a reasonable convergence rate $\rho(M^{-1}N)$ has to be safely less than 1, which implies that the heuristic lower bound (25) for $c(A)$ is not too large. Second, even if A is sparse, A^{-1} will usually be full, and so there are unlikely to be zeros on the right-hand side of (23). (Such zeros are dangerous because they can make $c(A)$ infinite.)

Note that in (24) the only terms that depend on the history of the iteration are $|G^{m+1}e_0|$ and θ_x. In using this bound we can redefine x_0 to be any iterate $\widehat{x}_k$, thereby possibly reducing θ_x. This is a circular argument if used to obtain a priori bounds, but it does suggest that the potentially large θ_x term will generally be innocuous. Note that if $x_i = 0$ for some i then θ_x is infinite unless $(\widehat{x}_k)_i = 0$ for all k. This difficulty with zero components of x can usually be overcome by redefining

$$\theta_x = \sup_k \max_{1\le i\le n} \frac{((|M| + |N|)|\widehat{x}_k|)_i}{((|M| + |N|)|x|)_i},$$

for which the above bounds remain valid if θ_x is replaced by $2\theta_x$.

Finally, we note that (24) implies

$$\|e_{m+1}\|_\infty \le \|G^{m+1}e_0\|_\infty + d_n u(1+\theta_x)c(A)\| \, |A^{-1}|(|M|+|N|)|x| \, \|_\infty. \quad (26)$$

If $\theta_x c(A) = O(1)$ and $|M| + |N| \le \alpha|A|$, with $\alpha = O(1)$, this bound is of the form (5) as $m \to \infty$ and we have componentwise forward stability.

4 Forward Error Bounds for Specific Methods

In this section we specialize the forward error bound (26) of the previous section to the Jacobi, Gauss-Seidel and SOR iterations.

4.1 Jacobi's Method

For the Jacobi iteration $M = D = \mathrm{diag}(A)$ and $N = \mathrm{diag}(A) - A$. Hence $|M| + |N| = |M - N| = |A|$, and so (26) yields

$$\|e_{m+1}\|_\infty \le \|G^{m+1}e_0\|_\infty + d_n u(1+\theta_x)c(A)\| \, |A^{-1}||A||x| \, \|_\infty. \quad (27)$$

If A is an M-matrix then $M^{-1} \ge 0$ and $M^{-1}N \ge 0$, so $c(A) = 1$. Hence in this case we have componentwise forward stability as $m \to \infty$ if θ_x is suitably bounded.

Woźniakowski [26, Example 4.1] cites the symmetric positive definite matrix

$$A = \begin{bmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{bmatrix}, \quad 0 < a < \frac{1}{2}, \quad \kappa_2(A) = \frac{1+2a}{1-a}, \quad \rho(M^{-1}N) = 2a,$$

as an example where the Jacobi method can be unstable, in the sense that there exist rounding errors such that no iterate has a relative error bounded by $c_n\kappa_\infty(A)u$. It is interesting to compare our bound (27) with the observed errors for this example. Straightforward manipulation shows that if $a = 1/2 - \epsilon$ ($\epsilon > 0$), then $c(A) \approx (3\epsilon)^{-1}$, so $c(A) \to \infty$ as $\epsilon \to 0$. (The heuristic lower bound (25) is $\approx 3(2\epsilon)^{-1}$ in this case.) Therefore (27) suggests that the Jacobi iteration can be unstable for this matrix. To confirm the instability we applied the Jacobi method to the problem with $x = (1, 1, \ldots, 1)^T$ and

Table 1: Jacobi method, $a = 1/2 - 8^{-j}$

	$\rho(M^{-1}N)$	Iters.	$\mathrm{cond}_\infty(A,x)$	$\min_k \phi_\infty(\widehat{x}_k)$	$\min_k \eta_\infty(\widehat{x}_k)$
$j=1$	0.75	90	3.40	2.22e-16	1.27e-16
$j=2$	0.97	352	4.76	1.78e-15	9.02e-16
$j=3$	0.996	1974	4.97	1.42e-14	7.12e-15
$j=4$	1.00	11226	5.00	1.14e-13	5.69e-14
$j=5$	1.00	55412	5.00	9.10e-13	4.55e-13

Table 2: Jacobi method, $a = -(1/2 - 8^{-j})$

	$\rho(M^{-1}N)$	Iters.	$\mathrm{cond}_\infty(A,x)$	$\min_k \phi_\infty(\widehat{x}_k)$	$\min_k \eta_\infty(\widehat{x}_k)$
$j=1$	0.75	39	7.00	4.44e-16	5.55e-17
$j=2$	0.97	273	6.30e1	4.88e-15	7.63e-17
$j=3$	0.996	1662	5.11e2	4.22e-14	8.24e-17
$j=4$	1.00	9051	4.09e3	3.41e-13	8.32e-17
$j=5$	1.00	38294	3.28e4	2.73e-12	8.33e-17

$a = 1/2 - 8^{-j}$, $j = 1\colon 5$. We used Matlab, for which the unit roundoff $u \approx 1.1 \times 10^{-16}$, and we took a random x_0 with $\|x - x_0\|_2 = 10^{-10}$. The iteration was terminated when there was no decrease in the norm of the residual for 50 consecutive iterations. Table 1 reports the smallest value of $\phi_\infty(\widehat{x}_k) = \|x - \widehat{x}_k\|_\infty / \|x\|_\infty$ over all iterations, for each j; the number of iterations is shown in the column "Iters."

The ratio $\min_k \phi_\infty(\widehat{x}_k)_{j+1} / \min_k \phi_\infty(\widehat{x}_k)_j$ takes the values 8.02, 7.98, 8.02, 7.98 for $j = 1\colon 4$, showing excellent agreement with the behaviour predicted by (27), since $c(A) \approx 8^j/3$. Moreover, $\theta_x \approx 1$ in these tests and setting $d_n \approx 1$ the bound (27) is at most a factor 13.3 larger than the observed error, for each j.

If $-1/2 < a < 0$ then A is an M-matrix and $c(A) = 1$. The bound (27) shows that if we set $a = -(1/2 - 8^{-j})$ and repeat the above experiment then the Jacobi method will perform in a componentwise forward stable manner (clearly, $\theta_x \approx 1$ is to be expected). We carried out the modified experiment, obtaining the results shown in Table 2. All the $\min_k \phi_\infty(\widehat{x}_k)_j$ values are less than $\mathrm{cond}_\infty(A,x)u$, so the Jacobi iteration is indeed componentwise

forward stable in this case. Note that since $\rho(M^{-1}N)$ and $\|M^{-1}N\|_2$ take the same values for a and $-a$, the usual rate of convergence measures cannot distinguish between these two examples.

4.2 Successive Over-Relaxation

The SOR method can be written in the form $Mx_{k+1} = Nx_k + b$, where

$$M = \frac{1}{\omega}(D + \omega L), \quad N = \frac{1}{\omega}((1-\omega)D - \omega U),$$

and where $A = D + L + U$, with L and U strictly lower triangular and upper triangular, respectively. The matrix $|M| + |N|$ agrees with $|A|$ everywhere except, possibly, on the diagonal, and the best possible componentwise inequality between these two matrices is

$$|M| + |N| \leq \frac{1 + |1-\omega|}{\omega}|A| \equiv f(\omega)|A|. \tag{28}$$

Note that $f(\omega) = 1$ for $1 \leq \omega \leq 2$, and $f(\omega) \to \infty$ as $\omega \to 0$. From (26) we have

$$\|e_{m+1}\|_\infty \leq \|G^{m+1}e_0\|_\infty + d_n u(1+\theta_x)c(A)f(\omega)\| \, |A^{-1}||A||x| \, \|_\infty.$$

If A is an M-matrix and $0 \leq \omega \leq 1$ then $M^{-1} \geq 0$ and $M^{-1}N \geq 0$, so $c(A) = 1$. The Gauss-Seidel method corresponds to $\omega = 1$, and it is interesting to note that for this method the above results have exactly the same form as those for the Jacobi method (though $c(A)$ and θ_x are, of course, different for the two methods).

5 Backward Error Analysis

In this section we obtain bounds for the residual vector $r_k = b - A\widehat{x}_k$. We write (11) as

$$r_{k+1} = Hr_k + (I - H)\xi_k,$$

where $H = NM^{-1}$ and ξ_k is defined and bounded in (14) and (15). This recurrence has the solution

$$r_{m+1} = H^{m+1}r_0 + \sum_{k=0}^{m} H^k(I - H)\xi_{m-k}. \tag{29}$$

Using the same reasoning as in the derivation of (17), we obtain the sharp bound

$$\|r_{m+1}\|_\infty \le \|H^{m+1}r_0\|_\infty + \|\sum_{k=0}^{m} |H^k(I-H)|\,\mu_{m-k}\,\|_\infty,$$

where μ_{m-k} is defined in (15). Taking norms in the second term gives, similarly to (18),

$$\|r_{m+1}\|_\infty \le \|H^{m+1}r_0\|_\infty + d_n u(1+\gamma_x)(\|M\|_\infty + \|N\|_\infty)\|\overline{H}\|_\infty\|x\|_\infty, \quad (30)$$

where

$$\overline{H} = \sum_{k=0}^{\infty} |H^k(I-H)|.$$

The following bound shows that $\|\overline{H}\|_\infty$ is small if $\|H\|_\infty = q < 1$, with q not too close to 1:

$$\|\overline{H}\|_\infty \le \|I-H\|_\infty \sum_{k=0}^{\infty} \|H\|_\infty^k = \frac{\|I-H\|_\infty}{1-q}.$$

A potentially much smaller bound can be obtained under the assumption that H is diagonalisable. If $H = XDX^{-1}$, with $D = \mathrm{diag}(\lambda_i)$, then

$$\begin{aligned}
\overline{H} &= \sum_{k=0}^{\infty} |X(D^k - D^{k+1})X^{-1}| \\
&\le |X|\Big(\sum_{k=0}^{\infty} \mathrm{diag}(|1-\lambda_i||\lambda_i^k|)\Big)|X^{-1}| \\
&= |X|\mathrm{diag}\Big(\frac{|1-\lambda_i|}{1-|\lambda_i|}\Big)|X^{-1}|.
\end{aligned}$$

Hence

$$\|\overline{H}\|_\infty \le \kappa_\infty(X)\max_i \frac{|1-\lambda_i|}{1-|\lambda_i|}. \quad (31)$$

Note that $\lambda_i = \lambda_i(H) = \lambda_i(NM^{-1}) = \lambda_i(M^{-1}N)$, so we see the reappearance of the term in the heuristic bound (25). The bound (31) is of modest size if the eigenproblem for H is well-conditioned ($\kappa_\infty(X)$ is small) and $\rho(H)$ is not too close to 1. Note that real eigenvalues of H near $+1$ do not affect the bound for $\|\overline{H}\|_\infty$, even though they may cause slow convergence.

To summarise, (30) and (2) show that for large m the normwise backward error $\eta_\infty(\widehat{x}_m)$ is certainly no larger than

$$d_n u(1+\gamma_x)\Big(\frac{\|M\|_\infty + \|N\|_\infty}{\|A\|_\infty}\Big)\|\overline{H}\|_\infty.$$

Note that $\|M\|_\infty + \|N\|_\infty \le 2\|A\|_\infty$ for the Jacobi and Gauss-Seidel methods, and also for the SOR method if $\omega \ge 1$.

To investigate the componentwise backward error we use the bound, from (29),

$$|r_{m+1}| \le |H^{m+1} r_0| + \sum_{k=0}^{m} |H^k(I-H)|\mu_{m-k}.$$

Lemma 2.1 (b) shows that there can be equality in any component of this inequality, modulo the term $|H^{m+1}r_0|$. With θ_x defined as in (20) we have, using (21),

$$|r_{m+1}| \le |H^{m+1} r_0| + d_n u(1+\theta_x)\overline{H}(|M| + |N|)|x|. \tag{32}$$

To bound the componentwise backward error we need an inequality of the form

$$|\overline{H}| \le \overline{c}(A)|\sum_{k=0}^{\infty} H^k(I-H)| = \overline{c}(A)I.$$

Unfortunately, no such bound holds, since $\overline{H}$ has nonzero off-diagonal elements unless $H = NM^{-1}$ is diagonal. It therefore does not seem possible to obtain a useful bound for the componentwise backward error. However, the bound (32) does have one interesting implication, as we now explain.

Consider any linear equation solver that provides a computed solution $\widehat{x}$ to $Ax = b$ that satisfies

$$|b - A\widehat{x}| \le uE|\widehat{x}|,$$

where E is a nonnegative matrix depending on any or all of A, b, n and u (with $E = O(1)$ as $u \to 0$). Suppose one step of iterative refinement is done in fixed precision; thus, $r = b - A\widehat{x}$ is formed (in the working precision), $Ad = r$ is solved, and the update $y = \widehat{x} + d$ is computed. Theorem 2.1 of [16] shows that $\widehat{y}$ satisfies

$$|b - A\widehat{y}| \le c_n(|A||\widehat{y}| + |b|) + O(u^2). \tag{33}$$

Theorem 2.2 of [16] shows further that if $E = \overline{E}|A|$, and $2\| \, |A||A^{-1}| \, \|_\infty \sigma(A, \widehat{y})(\|\overline{E}\|_\infty + n + 2)^2/(n+1) < u^{-1}$, where $\sigma(A, x) = \max_i(|A||x|)_i / \min_i(|A||x|)_i$, then

$$|b - A\widehat{y}| \leq 2(n+2)u|A||\widehat{y}|. \tag{34}$$

The bound (33) shows that iterative refinement relegates E to the second order term, yielding asymptotic componentwise stability; but, since it is only an asymptotic result no firm conclusion can be drawn about the stability. Under further assumptions (that A is not too ill-conditioned, and that the components of $|A||\widehat{y}|$ do not vary too much in magnitude) (34) shows that the componentwise backward error is indeed small.

If m is so large that $H^{m+1}r_0$ can ignored in (32), then we can apply these results on fixed precision iterative refinement with $E = d_n(1+\theta_x)\overline{H}(|M| + |N|)$ and $\overline{E} = \gamma d_n(1+\theta_x)|\overline{H}|$, where $\gamma = 1$ for the Jacobi and Gauss-Seidel methods, and $\gamma = f(\omega)$ in (28) for the SOR method. We conclude that even though the basic iterative methods are not guaranteed to produce a small componentwise backward error, one step of iterative refinement in fixed precision is enough to achieve this desirable property, provided the iteration and the problem are both sufficiently "well-behaved" for the given data A and b. When A and b are sparse $\sigma(A, x)$ can be very large; see [1] for a discussion of the performance of iterative refinement in this case. Of course, in the context of iterative solvers, iterative refinement may be unattractive because the solution of $Ad = r$ in the refinement step will in general be as expensive as the computation of the original solution $\widehat{x}$. We note that iterative refinement is identified as a means of improving the *normwise* backward error for iterative methods in [19].

To conclude, we return to our numerical examples. For the SOR example in section 1, $c(A) = O(10^{45})$ and $\|\overline{H}\|_\infty = O(10^{30})$, so our error bounds for this problem are all extremely large. In this problem $\max_i |1-\lambda_i|/(1-|\lambda_i|) = 3$, where $\lambda_i = \lambda_i(M^{-1}N)$, so (25) is very weak; (31) is not applicable since $M^{-1}N$ is defective.

For the first numerical example in section 4.1, Table 1 reports the minimum normwise backward errors $\eta_\infty(\widehat{x}_k)$. For this problem it is straightforward to show that $\|\overline{H}\|_\infty = (1-\epsilon)/\epsilon = 8^j(1-8^{-j})$. The ratios of backward errors for successive value of j are 7.10, 7.89, 7.99, 8.00, so we see excellent agreement with the behaviour predicted by the bounds. Table 2 reports the normwise backward errors for the second numerical example in section 4.1. The backward errors are all less than u, which again is close

to what the bounds predict, since it can be shown that $\|\overline{H}\|_\infty \leq 23/3$ for $-1/2 \leq a \leq 0$. In both the examples of section 4.1 the componentwise backward error $\omega(\widehat{x}_k) \approx \eta_\infty(\widehat{x}_k)$, and in our practical experience this behaviour is typical for the Jacobi and SOR iterations.

Acknowledgements

We thank Des Higham for his valuable comments on the manuscript, Sven Hammarling for providing us with a copy of [12], and Gene Golub for pointing out [20].

References

[1] M. Arioli, J. W. Demmel, and I. S. Duff. Solving sparse linear systems with sparse backward error. *SIAM J. Matrix Anal. Appl.*, 10(2):165–190, 1989.

[2] Mario Arioli, Iain S. Duff, and Daniel Ruiz. Stopping criteria for iterative solvers. *SIAM J. Matrix Anal. Appl.*, 13(1):138–144, 1992.

[3] Mario Arioli and Francesco Romani. Stability, convergence, and conditioning of stationary iterative methods of the form $x^{(i+1)} = Px^{(i)} + q$ for the solution of linear systems. *IMA J. Numer. Anal.*, 12:21–30, 1992.

[4] N. F. Benschop and H. C. Ratz. A mean square estimate of the generated roundoff error in constant matrix iterative processes. *J. Assoc. Comput. Mach.*, 18(1):48–62, 1971.

[5] Jo A. M. Bollen. Numerical stability of descent methods for solving linear equations. *Numer. Math.*, 43:361–377, 1984.

[6] Carl de Boor and Allan Pinkus. Backward error analysis for totally positive linear systems. *Numer. Math.*, 27:485–490, 1977.

[7] John E. Dennis, Jr. and Homer F. Walker. Inaccuracy in quasi-Newton methods: Local improvement theorems. *Math. Prog. Study*, 22:70–85, 1984.

[8] Gene H. Golub. Bounds for the round-off errors in the Richardson second order method. *BIT*, 2:212–223, 1962.

[9] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, USA, second edition, 1989.

[10] Anne Greenbaum. Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences. *Linear Algebra Appl.*, 113:7–63, 1989.

[11] Anne Greenbaum and Zdeněk Strakoš. Predicting the behavior of finite precision Lanczos and conjugate gradient computations. *SIAM J. Matrix Anal. Appl.*, 13(1):121–137, 1992.

[12] S. J. Hammarling and J. H. Wilkinson. The practical behaviour of linear iterative methods with particular reference to S.O.R. Report NAC 69, National Physical Laboratory, Teddington, UK, September 1976.

[13] Nicholas J. Higham. The accuracy of solutions to triangular systems. *SIAM J. Numer. Anal.*, 26(5):1252–1265, October 1989.

[14] Nicholas J. Higham. Bounding the error in Gaussian elimination for tridiagonal systems. *SIAM J. Matrix Anal. Appl.*, 11(4):521–530, October 1990.

[15] Nicholas J. Higham. How accurate is Gaussian elimination? In D. F. Griffiths and G. A. Watson, editors, *Numerical Analysis* 1989*, Proceedings of the* 13*th Dundee Conference*, volume 228 of *Pitman Research Notes in Mathematics*, pages 137–154. Longman Scientific and Technical, Essex, UK, 1990.

[16] Nicholas J. Higham. Iterative refinement enhances the stability of QR factorization methods for solving linear equations. *BIT*, 31:447–468, 1991.

[17] Nicholas J. Higham and Philip A. Knight. Finite precision behavior of stationary iteration for solving singular systems. *Linear Algebra Appl.*, 192:165–186, 1993.

[18] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.

[19] M. Jankowski and H. Woźniakowski. Iterative refinement implies numerical stability. *BIT*, 17:303–311, 1977.

[20] M. Stuart Lynn. On the round-off error in the method of successive over-relaxation. *Math. Comp.*, 18(85):36–49, 1964.

[21] W. Oettli and W. Prager. Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides. *Numer. Math.*, 6:405–409, 1964.

[22] J. L. Rigal and J. Gaches. On the compatibility of a given solution with the data of a linear system. *J. Assoc. Comput. Mach.*, 14(3):543–548, 1967.

[23] Lloyd N. Trefethen. Pseudospectra of matrices. In D. F. Griffiths and G. A. Watson, editors, *Numerical Analysis* 1991*, Proceedings of the* 14*th Dundee Conference*, volume 260 of *Pitman Research Notes in Mathematics*, pages 234–266. Longman Scientific and Technical, Essex, UK, 1992.

[24] Lloyd N. Trefethen and Robert S. Schreiber. Average-case stability of Gaussian elimination. *SIAM J. Matrix Anal. Appl.*, 11(3):335–360, 1990.

[25] H. Woźniakowski. Numerical stability of the Chebyshev method for the solution of large linear systems. *Numer. Math.*, 28:191–209, 1977.

[26] H. Woźniakowski. Roundoff-error analysis of iterations for large linear systems. *Numer. Math.*, 30:301–314, 1978.

[27] H. Woźniakowski. Roundoff-error analysis of a new class of conjugate-gradient algorithms. *Linear Algebra Appl.*, 29:507–529, 1980.

[28] David M. Young. A historical overview of iterative methods. *Comput. Phys. Comm.*, 53:1–17, 1989.

[29] Adam T. Zawilski. Numerical stability of the cyclic Richardson iteration. *Numer. Math.*, 60:251–290, 1991.