Robust Incremental Condition Estimation*

Christian H. Bischof
bischof@mcs.anl.gov

Ping Tak Peter Tang
tang@mcs.anl.gov

Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439-4801

December 16, 1991

Abstract. This paper presents an improved version of *incremental condition estimation*, a technique for tracking the extremal singular values of a triangular matrix as it is being constructed one column at a time. We present a new motivation for this estimation technique using orthogonal projections. The paper focuses on an implementation of this estimation scheme in an accurate and consistent fashion. In particular, we address the subtle numerical issues arising in the computation of the eigensystem of a symmetric rank-one perturbed diagonal 2×2 matrix. Experimental results show that the resulting scheme does a good job in estimating the extremal singular values of triangular matrices, independent of matrix size and matrix condition number, and that it performs qualitatively in the same fashion as some of the commonly used nonincremental condition estimation schemes.

AMS(MOS) subject classifications. 65F35, 65F05

Key words. Condition number, singular values, incremental condition estimation.

1 Introduction

Let $A = [a_1, \cdots, a_n]$ be an $m \times n$ matrix, and let $\sigma_1 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$ be the singular values of A. The smallest singular value

$$\sigma_{\min} \equiv \sigma_{\min(m,n)}$$

of A measures how close A is to a rank-deficient matrix [18, p. 19]. If we let $\sigma_{\max} \equiv \sigma_1$, the condition number

$$\kappa_2(A) \equiv \frac{\sigma_{\max}}{\sigma_{\min}},$$

which determines the sensitivity of equation systems involving A [18, 28], also depends crucially on $\sigma_{\min}$. For most practical purposes an order-of-magnitude estimate of $\sigma_{\min}$ or $\kappa_2(A)$ is sufficient. Most of the schemes for estimating $\sigma_{\min}$ and $\kappa_2(A)$ apply to triangular matrices, since in common applications A will be factored into a product of matrices involving a triangular matrix. A survey of those so-called condition estimation techniques for triangular matrices as well as their applications is given by Higham [20].

All of these condition estimators estimate the smallest singular value of a triangular matrix R in $O(n^2)$ time *after* it has been factored; and the entire condition estimation process has to be repeated if one wishes to estimate the condition number of a different matrix $\hat{R}$, even when $\hat{R}$ is closely related

*This work was supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U. S. Department of Energy, under Contract W-31-109-Eng-38.

to R. This issue has been addressed by recent work on so-called incremental and adaptive condition estimators.

"Incremental" condition estimation [5, 7] is an $O(n)$ scheme to arrive at an estimate for the condition number of $\hat{R}$ when

$$\hat{R} = \begin{bmatrix} R & w \\ & \gamma \end{bmatrix},$$

that is, $\hat{R}$ is R augmented by a column. This estimator is well suited for restricting column exchanges in rank-revealing orthogonal factorizations [3, 4, 6, 8].

"Adaptive" condition estimation schemes address the issue of rank-one updates of a triangular matrix R. Pierce and Plemmons [25, 24] suggest an $O(n)$ scheme and Ferng, Golub, and Plemmons [14] an $O(n^2)$ scheme for the situation where

$$\hat{R}^T \hat{R} = R^T R + uu^T.$$

These schemes are designed for recursive least-squares computations in signal processing. Shroff and Bischof [26] extend this work to the general rank-one update

$$\hat{R} = R + uv^T,$$

which appears for example in many optimization algorithms. The key difference between these two flavors of condition estimators is that incremental condition estimation obtains condition number estimates of a triangular factor that grows, whereas adaptive estimators maintain condition estimates when information is added or extracted from an already existing factorization.

In this paper we present an improved version and a robust implementation of the incremental condition estimator (ICE) originally suggested by Bischof [5]. We present a different motivation of this technique using orthogonal projections, and we address the subtle numerical issues involve in implementing this scheme in a numerically robust and consistent fashion. At the heart of our technique is the accurate computation of the eigensystem of a symmetric rank-one perturbed diagonal 2×2 matrix, and considerable care must be taken to compute this eigensystem accurately.

The paper is organized as follows. Section 2 derives the incremental condition estimation scheme, and Section 3 shows how we can ensure consistency in the sense of always producing an over(under)estimate for the smallest (largest) singular value, as the mathematical theory suggests. We then turn to the actual implementation of ICE; Section 4 discusses special cases, and Section 5 discusses the general case. In Section 6, we present numerical results that illustrate the reliability of our scheme and implementation; in particular, we show that the scheme behaves as reliably as nonincremental condition estimation schemes. Lastly, we summarize our contribution and discuss future work.

2 Incremental Condition Estimation

Let R be an $m \times m$ upper triangular complex matrix (in particular, R can be real), x be a complex m-vector, and τ be a real scalar such that

$$\|x^H R\| = \tau \text{ and } \begin{cases} \tau \approx \sigma_{\max}(R), & \text{or} \\ \tau \approx \sigma_{\min}(R). \end{cases}$$

Throughout this paper, $\|\cdot\|$ denotes the 2-norm, and $\sigma_j(R)$ denotes the j-th singular value of R,

$$\sigma_{\max}(R) \equiv \sigma_1(R) \geq \sigma_2(R) \geq \ldots \geq \sigma_m(R) \equiv \sigma_{\min}(R).$$

Clearly, having estimates for both $\sigma_{\max}(R)$ and $\sigma_{\min}(R)$ gives an estimate for the condition number of R in the 2-norm.

Given such a pair (x, τ), our goal is to obtain a new pair $(\hat{x}, \hat{\tau})$ for the augmented matrix

$$\hat{R} = \begin{bmatrix} R & w \\ & \gamma \end{bmatrix},$$

where w is a complex m-vector and γ a complex scalar.

Bischof [5] motivated ICE by exploiting the implication

$$Rx = d \implies \frac{1}{\sigma_{\min}(R)} = ||R^{-1}||_2 \geq \frac{||R^{-1}d||_2}{||d||_2} = \frac{||x||_2}{||d||_2},$$

which suggests generating a large norm solution x to a moderately sized right-hand side d and then using

$$\hat{\tau} := \frac{||d||_2}{||x||_2}$$

as an estimate for $\sigma_{\min}(R)$. This idea underlies many condition estimators [12, 13, 19]. The incremental characteristic of ICE was achieved by choosing the right-hand side d in a special way.

As it turns out, the same estimator can also be derived by considering the following well-known projection property of singular values. Let A be an $n \times n$ complex matrix and Y be an $n \times k$, $k \leq n$, complex matrix of orthonormal columns, that is, $Y^H Y = I$. Then,

$$\sigma_1(A) \geq \sigma_1(Y^H A), \quad \sigma_2(A) \geq \sigma_2(Y^H A), \quad \ldots, \quad \sigma_k(A) \geq \sigma_k(Y^H A),$$

and

$$\sigma_n(A) \leq \sigma_k(Y^H A), \quad \sigma_{n-1}(A) \leq \sigma_{k-1}(Y^H A), \quad \ldots, \quad \sigma_{n-k+1}(A) \leq \sigma_1(Y^H A).$$

We apply these inequalities to estimate the extremal singular values of $\hat{R}$ by letting $k = 2$,

$$Y = \begin{bmatrix} x & \\ & 1 \end{bmatrix} \in \mathbb{C}^{(m+1)\times 2}, \quad \text{and} \quad A = \hat{R}.$$

The left singular vectors of $Y^H A = Y^H \hat{R}$ are the eigenvectors of $M \equiv Y^H \hat{R}\hat{R}^H Y$, and the singular values are the square roots of M's eigenvalues. Denote M's eigenvalues by λ_1, λ_2, where $\lambda_1 \geq \lambda_2$, and denote the corresponding eigenvectors by z_1, z_2, respectively. The new estimates suggested naturally by the mathematics are

$$\left\{ \begin{array}{c} \sqrt{\lambda_1} \text{ and } Yz_1 \\ \sqrt{\lambda_2} \text{ and } Yz_2 \end{array} \right\} \text{ if } \left\{ \begin{array}{c} \tau \approx \sigma_1(R) \\ \tau \approx \sigma_n(R) \end{array} \right\}.$$

As a result of the choice of Y, M can be expressed in a particularly simple form:

$$\begin{aligned} M &= Y^H \hat{R}\hat{R}^H Y \\ &= \begin{bmatrix} x^H & \\ & 1 \end{bmatrix} \begin{bmatrix} R & w \\ & \gamma \end{bmatrix} \begin{bmatrix} R^H & \\ w^H & \bar{\gamma} \end{bmatrix} \begin{bmatrix} x & \\ & 1 \end{bmatrix} \\ &= \begin{bmatrix} x^H R R^H x & \\ & 0 \end{bmatrix} + \begin{bmatrix} x^H w \\ \gamma \end{bmatrix} [w^H x \quad \bar{\gamma}] \\ &= \begin{bmatrix} \tau^2 & \\ & 0 \end{bmatrix} + \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} [\bar{\alpha} \quad \bar{\gamma}], \quad \alpha = x^H w. \end{aligned}$$

The eigensystems of rank-one perturbed diagonal systems are well understood (see [10, 15, 16] for example).

The eigenvalues λ_1, λ_2 are the roots of the rational function

$$f(\lambda) = 1 - \frac{|\gamma|^2}{\lambda} + \frac{|\alpha|^2}{\tau^2 - \lambda},$$

and $\lambda_1 > \tau^2 > \lambda_2 > 0$. The corresponding eigenvectors are

$$\begin{bmatrix} \tau^2 - \lambda_1 & \\ & -\lambda_1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} \text{ and } \begin{bmatrix} \tau^2 - \lambda_2 & \\ & -\lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} \alpha \\ \gamma \end{bmatrix},$$

respectively.

3 Ensuring Consistency

Theoretically, the estimate $\hat{\tau}$ lies between the extreme singular values of $\hat{R}$. That is,

$$\sigma_{m+1}(\hat{R}) \leq \hat{\tau} \leq \sigma_1(\hat{R}).$$

It is desirable, therefore, for the computed estimate to also lie in this range. We call such an implementation *consistent*. By the basic properties of extremal singular values, the implementation will remain consistent as long as

$$\hat{\tau}_c^2 = z_c^H M z_c \left(1 + O(\varepsilon)\right),$$

where $\hat{\tau}_c$ is the computed new estimate and z_c is the computed eigenvector of M. That is, whatever the computed eigenvector z_c is, we would like the computed eigenvalue to be consistent with $z_c^H M z_c$. The following example shows that fulfilling this requirement is not as straightforward as it seems.

Consider the situation where $\tau = 2\varepsilon$, $\alpha = 1$, and $\gamma = 1 + \varepsilon$, where ε is the machine precision. Thus,

$$M = \begin{bmatrix} 4\varepsilon^2 & \\ & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1+\varepsilon \end{bmatrix} [1 \quad 1+\varepsilon].$$

The characteristic equation is

$$1 - \frac{(1+\varepsilon)^2}{\lambda} + \frac{1}{4\varepsilon^2 - \lambda} = 0$$

or, equivalently,

$$(4\varepsilon^2 - \lambda)\lambda - (1+\varepsilon)^2(4\varepsilon^2 - \lambda) + \lambda = 0.$$

Let λ_2 be the smaller root. Expressing $\lambda_2 = 2\varepsilon^2 + \Delta$, we have

$$(2\varepsilon^2 - \Delta)(2\varepsilon^2 + \Delta) - (1+\varepsilon)^2(2\varepsilon^2 - \Delta) + (2\varepsilon^2 + \Delta) = 0,$$

which is

$$\Delta^2 - \Delta(1 + (1+\varepsilon)^2) - 4\varepsilon^3 + 2\varepsilon^4 = 0.$$

Since the smaller root is $\Delta = -2\varepsilon^3 + O(\varepsilon^4)$, we have

$$\begin{aligned} \lambda_2 &= 2\varepsilon^2 - 2\varepsilon^3 + O(\varepsilon^4) \\ &= 2\varepsilon^2 + O(\varepsilon^3). \end{aligned}$$

Thus, $2\varepsilon^2$ is a fully accurate approximation to λ_2. Unfortunately, this fully accurate solution leads to a potentially inconsistent estimation, as the following shows. If we use $2\varepsilon^2$ as the computed λ_2 (that is, $\hat{\tau}_c^2 = 2\varepsilon^2$), the corresponding unnormalized eigenvector is

$$\left[\frac{1}{2\varepsilon^2}, \frac{-(1+\varepsilon)}{2\varepsilon^2}\right]^T,$$

which normalizes to $z_c = [1, -(1+\varepsilon)]^T$. But straightforward calculation shows

$$z_c^H M z_c = 4\varepsilon^2 + O(\varepsilon^2),$$

which is bigger than $\hat{\tau}_c^2$ by a factor of 2.

This example motivates the following analysis. Suppose $\hat{R}$'s smallest singular value is approximately ε, but its next singular value is approximately 1. Let (λ, z) be an eigenpair of M, and let (λ_c, z_c) be the corresponding computed quantities. Assume that (λ_c, z_c) are computed to full precision in the sense that

$$z_c = z + \tilde{z}, \quad \|\tilde{z}\| = \varepsilon_1, \text{ and } \lambda_c = \lambda(1+\varepsilon_2),$$

where $|\varepsilon_1|, |\varepsilon_2| \approx \varepsilon$. Now consider

$$z_c^H M z_c = z^H M z + 2\tilde{z}^H M z + \tilde{z}^H M \tilde{z}.$$

Thus, the relative error

$$\frac{z_c^H M z_c - \lambda_c}{\lambda} \approx (2\varepsilon_1 - \varepsilon_2) + \tilde{z}^H M \tilde{z}/\lambda.$$

The error $2\varepsilon_1 - \varepsilon_2$ is negligible. Next, denote $\tilde{z}^H M \tilde{z}/\lambda$ by δ. Note that $\delta \geq 0$. If we are estimating the largest singular value of $\hat{R}$, then $\lambda \approx \|M\|$, giving $\delta \approx \varepsilon^2$, which is obviously negligible.

But if are trying to estimate $\hat{R}$'s smallest singular value and it happens to be near ε, we will have

$$\|M\| = \sigma_1^2(Y^H \hat{R}) \geq \sigma_m^2(\hat{R}) \approx 1$$

(recall that the dimension of $\hat{R}$ is $m+1$), and hence $\|\tilde{z}^H M \tilde{z}\|$ can be as big as ε^2. Now if $\lambda \approx \sigma_{m+1}^2(\hat{R}) \approx \varepsilon^2$, then $\delta \approx 1$; that is, the first digit of λ_c is wrong in the direction that may lead to an inconsistent estimation. Fortunately, the following simple calculation offers a practical safeguard.

Let λ_c and z_c be the computed eigenpair. If we are estimating the largest singular value $\sigma_1(\hat{R})$, then proceed as usual:

$$\hat{\tau} := \sqrt{\lambda_c} \quad \text{and} \quad \hat{x} := Y \cdot z.$$

If we are estimating the smallest singular value, then

$$\hat{\tau} := \sqrt{\lambda_c + 4\varepsilon^2\|M\|} \quad \text{and} \quad \hat{x} := Y \cdot z.$$

For computational convenience, $\|M\|$ can be replaced by $\|M\|_\infty$. When the condition number of $\hat{R}$ is moderate, the compensation term $4\varepsilon^2\|M\|$ is so small that the quality of the estimation is unaffected. When $\hat{R}$ is ill conditioned, this compensation ensures $\hat{\tau} \geq \|\hat{x}^H \hat{R}\| \geq \sigma_{m+1}(\hat{R})$.

4 Special Cases

Mathematically, the eigensystem

$$M = \begin{bmatrix} \tau^2 & \\ & 0 \end{bmatrix} + \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} [\bar{\alpha} \quad \bar{\gamma}]$$

($\tau \geq 0$ real and α, γ complex) simplifies greatly if one (or both) of α and γ is zero. In this section, we will address the cases (1) $\tau = 0$, (2) $|\gamma| \leq \varepsilon\tau$ and $\tau > 0$, (3) $|\alpha| \leq \varepsilon\tau$ and $\tau > 0$, and (4) $0 < \tau \leq \varepsilon|\gamma|$ or $0 < \tau \leq \varepsilon|\alpha|$. Handling these cases separately allows the computations for the usual case to be free from possible spurious overflow.

4.1 Case 1. $\tau = 0$

Since $M = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} [\bar{\alpha} \quad \bar{\gamma}]$, the two eigenvalues of M are 0 and $|\alpha|^2 + |\gamma|^2$. If $|\alpha|^2 + |\gamma|^2 > 0$, the corresponding eigenvectors are $[-\bar{\gamma} \quad \bar{\alpha}]^T$ and $[\alpha \quad \gamma]^T$, respectively. If $\alpha = \gamma = 0$, then $[1 \quad 0]^T$ and $[0 \quad 1]^T$ are an appropriate pair of eigenvectors. By scaling properly, one can easily calculate the square roots of the eigenvalues and the corresponding normalized eigenvectors without spurious overflow.

4.2 Case 2. $|\gamma| \leq \varepsilon\tau$ and $\tau > 0$

Note that $\tau > 0$ is immediate if we first test for Case 1. Now,

$$M = \tau^2 \left\{ \begin{bmatrix} 1 & \\ & 0 \end{bmatrix} + \begin{bmatrix} \alpha/\tau \\ \gamma/\tau \end{bmatrix} [\bar{\alpha}/\tau \quad \bar{\gamma}/\tau] \right\},$$

where $|\gamma/\tau| \leq \varepsilon$ and $\|M/\tau^2\| \geq 1$. Thus, one can consider the two eigenvalues to be $|\gamma|^2$ and $\tau^2 + |\alpha|^2$. (Clearly, $|\gamma|^2$ is the smaller one.) The corresponding eigenvectors are $[0 \quad 1]^T$ and $[1 \quad 0]^T$, respectively.

4.3 Case 3. $|\alpha| \leq \varepsilon\tau$ and $\tau > 0$

By an analysis similar to that in Case 2, the two eigenvalues are $|\gamma|^2$ and τ^2. A comparison between $|\gamma|$ and τ is needed to determine the smaller and the larger eigenvalues. The corresponding eigenvectors are $[0 \quad 1]^T$ and $[1 \quad 0]^T$.

4.4 Case 4. $0 < \tau \leq \varepsilon|\gamma|$ or $0 < \tau \leq \varepsilon|\alpha|$

First, consider the smallest eigenvalue λ_2 of

$$M = \begin{bmatrix} \tau^2 & \\ & 0 \end{bmatrix} + \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} [\bar{\alpha} \quad \bar{\gamma}].$$

Now $0 < \lambda_2 < \tau^2$, and

$$1 - \frac{|\gamma|^2}{\lambda_2} + \frac{|\alpha|^2}{\tau^2 - \lambda_2} = 0.$$

But

$$\tau \leq \varepsilon|\gamma| \quad \text{or} \quad \tau \leq \varepsilon|\alpha|$$

implies

$$\frac{|\gamma|^2}{\lambda_2} \geq \varepsilon^{-2} \gg 1 \quad \text{or} \quad \frac{|\alpha|^2}{\tau^2 - \lambda_2} \geq \varepsilon^{-2} \gg 1.$$

Consequently, for all practical purposes,

$$-\frac{|\gamma|^2}{\lambda_2} + \frac{|\alpha|^2}{\tau^2 - \lambda_2} = 0,$$

giving

$$\lambda_2 = \tau^2 \frac{|\gamma|^2}{|\alpha|^2 + |\gamma|^2}.$$

The corresponding eigenvector is $[-\bar{\gamma} \quad \bar{\alpha}]^T$. By scaling properly, we can compute $\sqrt{\lambda_2}$ and the normalized eigenvector without spurious overflow.

Next, consider the largest eigenvalue λ_1 of M. Now $\lambda_1 > \tau^2$, and

$$1 - \frac{|\gamma|^2}{\lambda_1} + \frac{|\alpha|^2}{\tau^2 - \lambda_1} = 0.$$

Clearly,

$$0 < \frac{|\gamma|^2}{\lambda_1}, \frac{|\alpha|^2}{\lambda_1 - \tau^2} < 1.$$

Thus

$$0 < \tau \leq \varepsilon|\gamma| \quad \text{or} \quad 0 < \tau \leq \varepsilon|\alpha|$$

implies

$$\lambda_1 - \tau^2 = \lambda_1(1 + O(\varepsilon^2)).$$

Consequently, for all practical purposes,

$$1 - \frac{|\gamma|^2}{\lambda_1} - \frac{|\alpha|^2}{\lambda_1} = 0,$$

giving

$$\lambda_1 = |\alpha|^2 + |\gamma|^2.$$

The corresponding eigenvector is $[\alpha \quad \gamma]^T$. By scaling properly, we can compute $\sqrt{\lambda_1}$ and the normalized eigenvector without spurious overflow.

5 The Usual Case

The goal here is to compute the eigensystem of

$$M = \begin{bmatrix} \tau^2 & \\ & 0 \end{bmatrix} + \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} [\bar{\alpha} \quad \bar{\gamma}]$$

accurately. If M does not belong to any of the special cases described previously, we must have

$$\varepsilon \le |\alpha|/\tau \text{ and } |\gamma|/\tau \le 1/\varepsilon.$$

We therefore consider the eigensystem of

$$\tau^{-2} M = \begin{bmatrix} 1 & \\ & 0 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} [\bar{\zeta}_1 \quad \bar{\zeta}_2],$$

where $\zeta_1 = \alpha/\tau$ and $\zeta_2 = \gamma/\tau$. The advantage of this scaling is that subsequent computations involving $|\zeta_j|^2$ are extremely unlikely to overflow or underflow.

Denote $\tau^{-2}M$ by A, and let μ_1, μ_2, where $\mu_1 > 1 > \mu_2 > 0$, be A's eigenvalues. Clearly,

$$\sqrt{\lambda_j} = \tau\sqrt{\mu_j}, \quad j = 1, 2,$$

and the eigenvectors of A and M are identical. Recall that the eigenvectors are given by

$$\begin{bmatrix} 1-\mu_1 & \\ & -\mu_1 \end{bmatrix}^{-1} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1-\mu_2 & \\ & -\mu_2 \end{bmatrix}^{-1} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}.$$

Hence, in order for the eigenvectors to be accurate, the quantites μ_j *and* $1 - \mu_j$, $j = 1, 2$, have to be computed accurately. The implication is that, depending on the situation, computing the μ_j's accurately may not be sufficient. The following example illustrates the point.

Consider

$$A = \begin{bmatrix} 1 & \\ & 0 \end{bmatrix} + \begin{bmatrix} \sqrt{\varepsilon} \\ 1 \end{bmatrix} [\sqrt{\varepsilon} \quad 1].$$

Thus,

$$\begin{aligned} \mu_1 &= 1 + \frac{1}{2}\varepsilon + \sqrt{\varepsilon}(1 + \frac{1}{4}\varepsilon)^{1/2}, \\ \mu_2 &= 1 + \frac{1}{2}\varepsilon - \sqrt{\varepsilon}(1 + \frac{1}{4}\varepsilon)^{1/2}. \end{aligned}$$

Clearly,

$$\begin{aligned}\hat{\mu}_1 &= 1+\sqrt{\varepsilon}+\varepsilon \\ \hat{\mu}_2 &= 1-\sqrt{\varepsilon}\end{aligned}$$

are accurate approximations to μ_1, μ_2 (with only a rounding error in the last digit).

Using these computed values, we obtain for the eigenvectors

$$\begin{bmatrix} -(\sqrt{\varepsilon}+\varepsilon) & \\ & -(1+\sqrt{\varepsilon}+\varepsilon) \end{bmatrix}^{-1} \begin{bmatrix} \sqrt{\varepsilon} \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \Big/ 1+\sqrt{\varepsilon} \\ -1 \Big/ 1+\sqrt{\varepsilon}+\varepsilon \end{bmatrix}, \quad \text{and}$$

$$\begin{bmatrix} \sqrt{\varepsilon} & \\ & -(1-\sqrt{\varepsilon}) \end{bmatrix}^{-1} \begin{bmatrix} \sqrt{\varepsilon} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \Big/ 1+\sqrt{\varepsilon} \end{bmatrix}.$$

To normalize the vectors, we need only multiply $1/\sqrt{2}$ to each. The inner product of the normalized vectors is $\frac{1}{2}\sqrt{\varepsilon}+O(\varepsilon)$, instead of the desired $O(\varepsilon)$. The problem here is that the rounding errors in $\hat{\mu}_1$ and $\hat{\mu}_2$ are being magnified in the subtractions $1-\hat{\mu}_1$ and $1-\hat{\mu}_2$. The next two subsections show how M's eigensystem can be computed accurately.

5.1 Computations for the Larger Eigenpair

Since $\mu_1 > 1$, we can obtain μ_1 accurately by $1-(1-\mu_1)$, provided we can compute $1-\mu_1$ accurately. The rational function characterizing the eigenvalues of M is

$$f(\mu) = 1 - \frac{|\zeta_2|^2}{\mu} + \frac{|\zeta_1|^2}{1-\mu}.$$

We consider the translated equation

$$\begin{aligned} g(\eta) &= f(1+\eta) = 1 - \frac{|\zeta_2|^2}{1+\eta} - \frac{|\zeta_1|^2}{\eta} \\ &= \frac{\eta(1+\eta) - |\zeta_2|^2\eta - |\zeta_1|^2(1+\eta)}{\eta(1+\eta)}. \end{aligned}$$

We are interested in the larger root of

$$q(\eta) = \eta^2 + 2b\eta - c = 0,$$

where $b = (1-|\zeta_1|^2 - |\zeta_2|^2)/2$ and $c = |\zeta_1|^2$. The larger root η_1 is given by

$$\eta_1 = \begin{cases} -b+\sqrt{b^2+c}, \text{ or} \\ c/(b+\sqrt{b^2+c}). \end{cases}$$

To avoid cancellation, we use the first formula when $b \le 0$ and the second one when $b > 0$.

Since $\mu_1 = 1+\eta_1$, we have $\lambda_1 = \tau^2(1+\eta_1)$ or $\sqrt{\lambda_1} = \tau\sqrt{1+\eta_1}$. The corresponding eigenvector is

$$\begin{bmatrix} -\eta_1 & \\ & -(1+\eta_1) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}.$$

5.2 Computations for the Smaller Eigenpair

Here we are interested in the smaller root, μ_2, of

$$f(\mu) = 1 - \frac{|\zeta_2|^2}{\mu} + \frac{|\zeta_1|^2}{1-\mu}.$$

The objective is to be able to obtain both μ_2 and $1-\mu_2$ accurately. This objective can be achieved by first computing μ_2 or $1-\mu_2$ accurately, whichever has smaller magnitude, and then computing the other from the first. Since $0 < \mu_2 < 1$, $f'(\mu) > 0$ on $(0,1)$, $\lim_{x\to 0^+} f(x) = -\infty$, and $\lim_{x\to 1^-} f(x) = +\infty$, we know that $\mu_2 \leq 1-\mu_2$ if $f(1/2) \geq 0$; otherwise $\mu_2 > 1-\mu_2$. We therefore consider two cases.

5.2.1 Case a. $f(1/2) \geq 0$

Here we compute the smaller root of the untranslated equation

$$f(\mu) = 1 - \frac{|\zeta_2|^2}{\mu} + \frac{|\zeta_1|^2}{1-\mu},$$

which is the smaller root of the quadratic

$$q(\mu) = \mu^2 - 2b\mu + c,$$

where $b = (1+|\zeta_1|^2+|\zeta_2|^2)/2$ and $c = |\zeta_2|^2$. The smaller root $\mu_2 = c/(b+\sqrt{b^2-c})$. The corresponding eigenvector is

$$\begin{bmatrix} 1-\mu_2 & \\ & -\mu_2 \end{bmatrix}^{-1} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}.$$

5.2.2 Case b. $f(1/2) < 0$

Since $1-\mu_2 < \mu_2$, the objective is to calculate $1-\mu_2$ accurately. Thus, we consider the translated equation

$$\begin{aligned} g(\eta) &= f(1+\eta) = 1 - \frac{|\zeta_2|^2}{1+\eta} - \frac{|\zeta_1|^2}{\eta} \\ &= \frac{\eta(1+\eta) - |\zeta_2|^2\eta - |\zeta_1|^2(1+\eta)}{\eta(1+\eta)}. \end{aligned}$$

We are interested in the smaller root of

$$q(\eta) = \eta^2 + 2b\eta - c = 0,$$

where $b = (1-|\zeta_1|^2-|\zeta_2|^2)/2$ and $c = |\zeta_1|^2$. The smaller root

$$\eta_2 = \begin{cases} -b - \sqrt{b^2+c}, \text{ or} \\ c/(b - \sqrt{b^2+c}). \end{cases}$$

The first formula is preferable when $b \geq 0$, while the second one is better when $b < 0$. Since $\mu_2 = 1+\eta_2$, we have $\lambda_2 = \tau^2(1+\eta_2)$ or $\sqrt{\lambda_2} = \tau\sqrt{1+\eta_2}$. The corresponding eigenvector is

$$\begin{bmatrix} -\eta_2 & \\ & -(1+\eta_2) \end{bmatrix}^{-1} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}.$$

6 Numerical Results

The purpose of our experiments is threefold. First, we wish to establish that in practice our ICE scheme delivers reliable estimates (even though one can construct matrices where it performs arbitrarily badly [5]). Second, we wish to show that ICE performs qualitatively in the same way as some of the well-known condition estimators currently in use, in particular the Linpack condition estimator [13] and Higham's condition estimator [21, 22] which is being used in the LAPACK package [1, 2]. Third, we wish to demonstrate that ICE is more reliable in correctly identifying the rank of triangular matrices produced by the QR factorization with column pivoting [17] than is the heuristic that is typically employed. The experiments reported here were performed with real matrices.

6.1 The Accuracy of ICE

We performed two sets of test runs. In the first set, we chose n singular values $\sigma_1, \sigma_2, \ldots, \sigma_n$, (not necessarily in order) from $[0, 1]$ according to some specified distribution. Then, we employed Stewart's method [27] to generate random orthogonal matrices U and V. The upper-triangular matrix R used in testing was the R factor of the QR decomposition

$$QR = U \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) V^T.$$

For $n = 50, 100, 150$, and 200, we used four distributions of singular values and generated 200 test matrices in each distribution. The four distributions are as follows:

Random: the singular values are chosen randomly from the interval $[0, 1]$.

Sharp Break: one singular value is 10^{-10}; all the others are 1.

Exponential: the singular values are $1, r, r^2, \ldots, r^{n-1} = 10^{-10}$.

Cluster: five singular values are chosen randomly from the interval $[0.9 \times 10^{-10}, 1.1 \times 10^{-10}]$; the rest are chosen randomly from the interval $[10^{-7}, 1]$.

These experiments were performed using double precision on a Sun Sparcstation.

Figure 1 presents the results of our algorithm. The two histograms show by what factor we overestimate the smallest singular value and underestimate the condition number of R, having used our ICE scheme to estimate both the smallest and largest singular value of R. That is, we display

$$r_{\min} \equiv \frac{\tau_{\min}}{\sigma_{\min}} \text{ and } r_{\text{cond}} \equiv \frac{(\tau_{\max}/\tau_{\min})}{(\sigma_{\max}/\sigma_{\min})}.$$

So, for example, in 219 out of the total 800 cases, we overestimated the smallest singular value by a factor between 1 and 2, and there were only two occurrences where the overestimate was worse by a factor of more than 10. The situation for the condition number is much the same, and in all but 8 cases our estimates were within a factor of 10 of the true condition number.

Table 1 displays the median value and the worst observed value for $r_{\min}$, $r_{\max} \equiv \sigma_{\max}/\tau_{\max}$, and r_{cond}, grouped according to the different singular value distributions that we employed.

We see that, apart from the "sharp break" distribution, the singular value distribution does not have a noticeable influence on the performance of our estimator. We also did not notice a significant influence of the matrix size on the quality of the estimates produced. Except for rare occurrences, our ICE implementation delivers estimates for the smallest singular value that are within a factor of ten of the true smallest singular value of R. Moreover, all estimates for the largest singular value are within a factor of two of the true largest singular value of R.

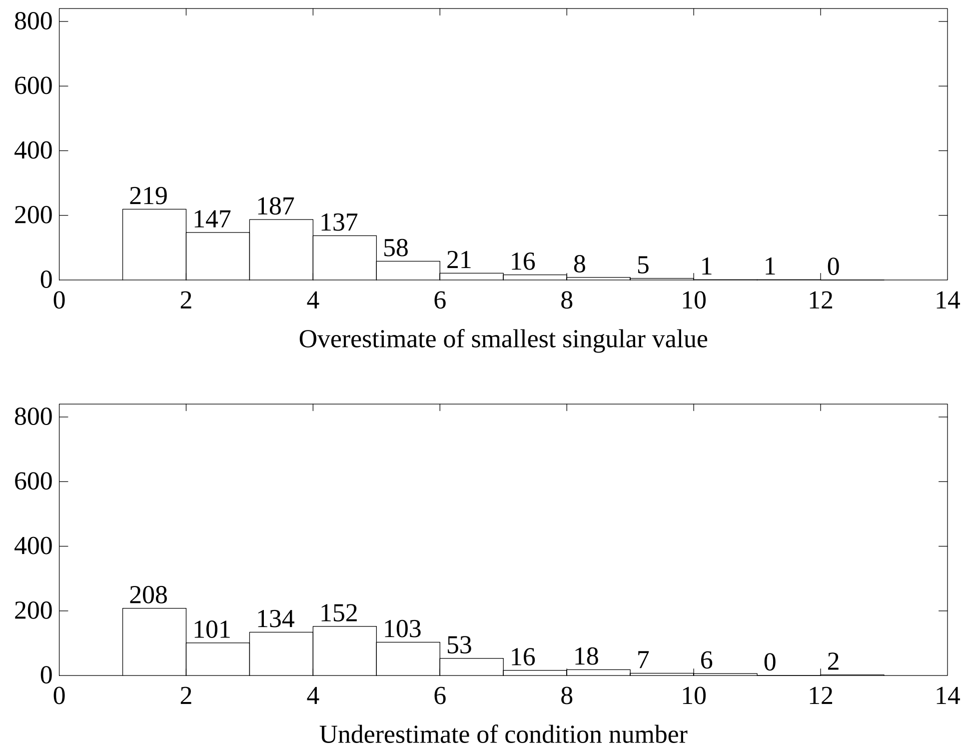


Figure 1: Accuracy of ICE

Table 1: Results of Double-Precision Tests

	$r_{\min}$		$r_{\max}$		r_{cond}	
Distribution	Median	Worst	Median	Worst	Median	Worst
Random	3.25	11.30	1.13	1.22	3.65	12.50
Sharp Break	1.00	1.00	1.00	1.00	1.00	1.00
Exponential	3.75	6.11	1.21	1.81	4.71	9.55
Cluster	3.94	9.54	1.15	1.32	4.53	10.85

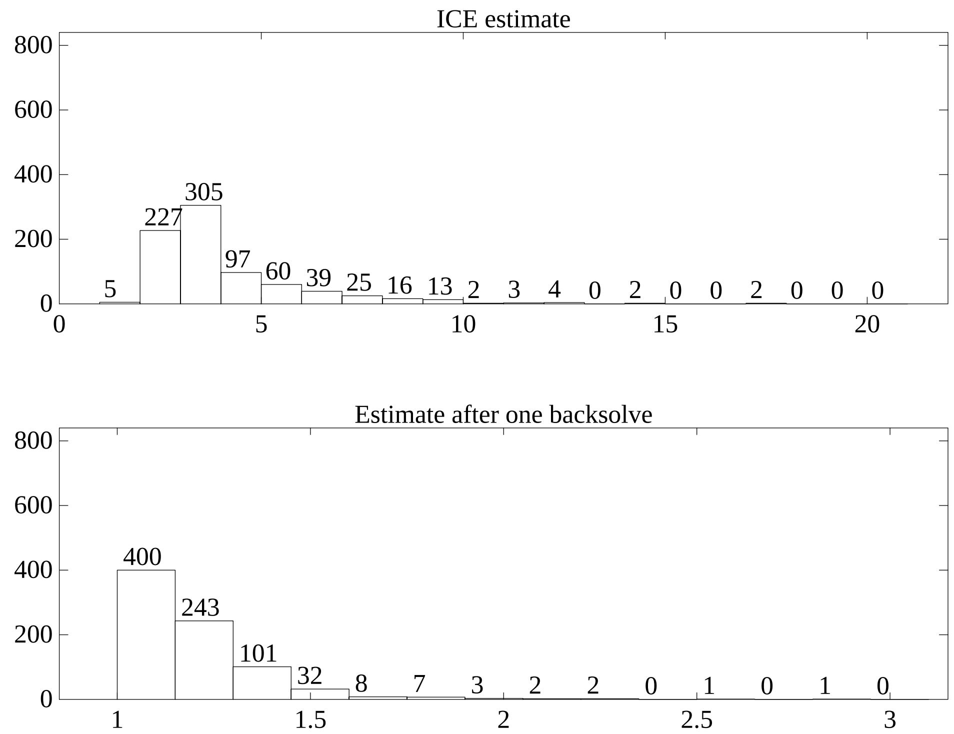


Figure 2: Overestimates of Smallest Singular Value by ICE

6.2 ICE in Comparison with Other Estimators

The second set of experiments was designed mainly to show that ICE shows the same type of qualitative behavior as the Linpack estimator STRCO [13] and the LAPACK estimator SGECON [21, 22]. For $n = 100$ and 200 we used four types of matrices:

Exponential: the singular values are $1, r, r^2, \ldots, r^{n-1} = 10^{-6}$.

Randomlog: the singular values are random numbers in the range $[10^{-6}, 1]$ such that their logarithms are uniformly distributed.

Cluster: ten singular values are chosen randomly from the interval $[\varepsilon, 4\varepsilon]$; the rest are chosen randomly from the interval $(\varepsilon, 1]$.

These test matrices were generated as were those in the preceding section. Lastly, we employed

RandomA: the elements of A were generated randomly using a uniform distribution on $(0, 1)$; R is the triangular factor from a QR factorization of A.

The upper plot in Figure 2 shows how much ICE overestimates the smallest singular value of A on this set of experiments. The behavior is much the same as in Figure 1; for example, in 305 of the 800 test cases, the smallest singular value was overestimated by a factor of between 3 and 4. The second plot shows how the estimate returned by ICE is improved through one backsolve. That is, given the approximate left nullvector x returned by ICE, we solve the triangular system

$$Rz = x$$

to generate an approximate right nullvector z, and we use

$$\tilde{\tau} \equiv 1/\|z\|_2$$

as our estimate of the smallest singular value of R. The histogram shows by what factor we overestimate the smallest singular value using this estimate, that is, $\tilde{\tau}/\sigma_{\min}(R)$. Note that while the bucket size in upper histogram is 1.0, it is 0.15 in the lower histogram. Note further that in 243 of the 800 test cases, the smallest singular value is now overestimated by a factor between 1.15 and 1.30. Of course, the greater is the gap between σ_n and σ_{n-1}, the more effective this improvement step is. Nevertheless, we did not artificially put pronounced gaps in our examples. These experiments show that the approximate nullvectors produced by ICE would be very good starting vectors for an inverse iteration process for computing exact singular values and vectors of R.

Figure 3 shows the condition number estimates returned by ICE and the Linpack and LAPACK condition estimators on these test matrices. The first 100 sample points correspond to the matrices of dimension 100; the second 100 sample points correspond to the matrices of dimension 200. The Linpack and LAPACK condition estimators both estimate the one-norm of R. To make them comparable to the two-norm estimates returned by ICE, we scaled them by a factor of $\sqrt{n}$. That is, if $\hat{\kappa}_1$ is the condition number estimate returned by the Linpack or LAPACK condition estimator for an $n \times n$ matrix, we display $\hat{\kappa}_1/\sqrt{n}$. As we can see, the three estimators show the same qualitative behavior in tracking the condition number of R. In particular, in the plots showing the "RandomA" and "Cluster" distributions, all estimators track "spikes" in the condition number correctly.

6.3 ICE and the QR Factorization with Column Pivoting

A well-known strategy for extracting a set of reasonably independent columns of a given matrix A and for computing an orthonormal basis for the span of A is the QR factorization with column pivoting [9, 11, 23]:

$$AP = QR.$$

Viewed geometrically [18, p. 168, P.6.4–5] this strategy chooses at every step that column of A that is farthest away (in the two-norm sense) from the subspace spanned by the columns that were selected before.

One approach to estimating the rank of A is first to compute a QR factorization with column pivoting of A and then to use $\sigma_{\min}(R(1:i,1:i))$, that is, the smallest singular value of the leading $i \times i$ submatrix, as an estimate for the ith singular value of A. In particular, A is considered to have rank k with respect to a condition number threshold ξ if

$$\frac{\sigma_{\max}(R(1:k,1:k))}{\sigma_{\min}(R(1:k,1:k))} \leq \xi \leq \frac{\sigma_{\max}(R(1:k+1,1:k+1))}{\sigma_{\min}(R(1:k+1,1:k+1))}.$$

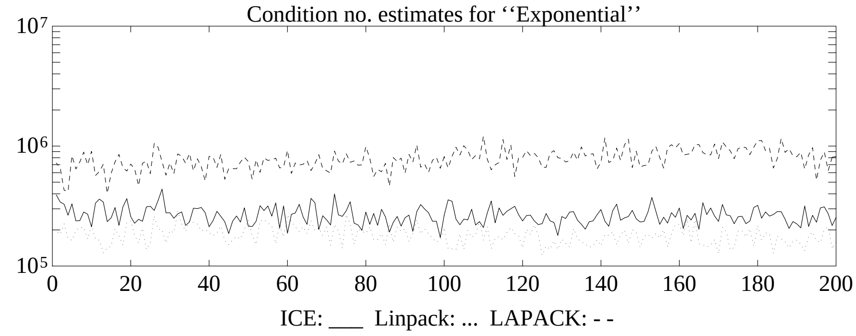
Since the matrix R produced by the QR factorization with column pivoting is graded, the moduli of the first and ith diagonal entry are heuristically good estimates for the extremal singular values of $R(1:i,1:i)$. Thus, we estimate

$$\sigma_{\max}(R(1:i,1:i)) \approx |r_{11}|$$

and

$$\sigma_{\min}(R(1:i,1:i)) \approx |r_{ii}|.$$

In Table 2 we compare these estimates with the ICE estimates. For $n = 100$ we generated the same full matrices as in the preceding section, but the triangular test matrices R were now the results of an orthogonal factorization *with column pivoting*. The column labeled "ICE" shows by what factor ICE underestimated the condition number of R; the column "Diagonal" shows by what factor the ratio $|r_{11}|/|r_{nn}|$ underestimated the condition number of R.



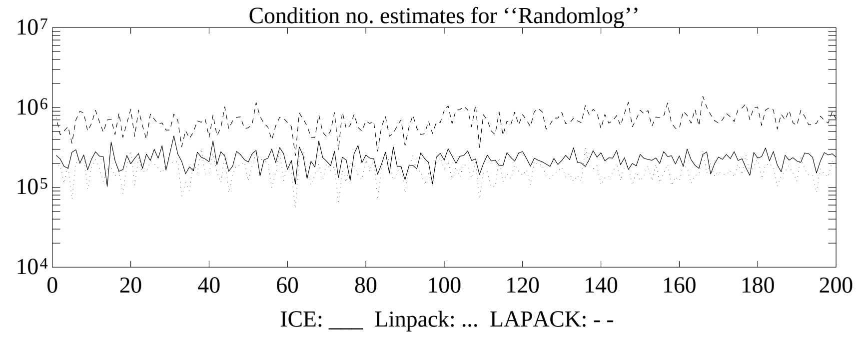
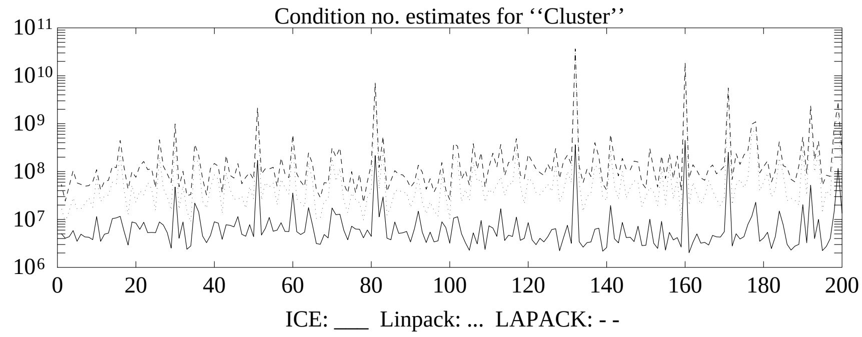


Figure 3: ICE in Comparison with Other Estimators



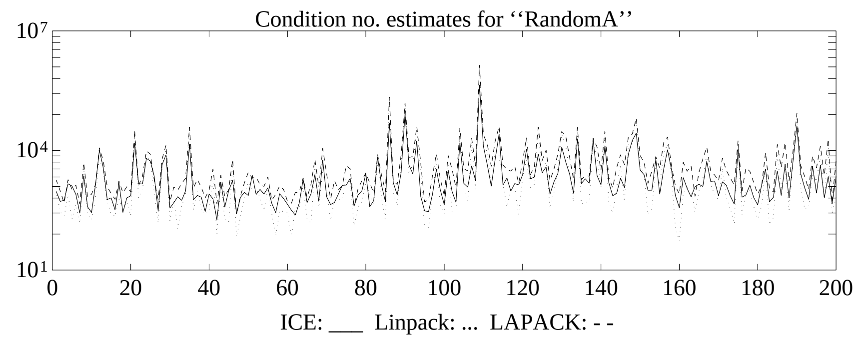


Table 2: ICE versus Diagonal Estimate

	Median		Worst	
Distribution	ICE	Diagonal	ICE	Diagonal
RandomA	3.58	66.1	14.1	1343.9
Randomlog	3.48	11.7	7.46	22.8
Exponential	3.78	12.9	5.84	23.3
Cluster	4.60	10.8	12.5	46.9

In particular, the random matrices show that using ICE yields much more reliable estimates than using the heuristics based on diagonal elements. For this reason, ICE has been incorporated in the LAPACK driver routine SGELSX, which computes the minimum-norm solution of a possibly rank-deficient least-squares problem by using an orthogonal factorization with column pivoting.

7 Concluding Remarks

We have presented an improved version of *incremental condition estimation*, a technique for tracking the extremal singular values of a triangular matrix as it is being constructed one column at a time. At the heart of our technique is the accurate computation of the eigensystem of a 2×2 symmetric rank-one perturbed diagonal matrix. This seemingly simple task requires great care when finite-precision arithmetic is used. The eigenvalue solver then leads to a robust and consistent implementation of incremental condition estimation.

Experimental results show that our scheme delivers good estimates of the extremal singular values and performs qualitatively as well as the one-norm estimators used in Linpack and LAPACK. The results also demonstrate the advantages of using incremental condition estimation over the usual heuristic in estimating the rank of a triangular matrix generated by the QR factorization with column pivoting.

The derivation of incremental condition estimation used in this paper suggests that one could design incremental condition estimators that estimate several extremal singular values at the same time (for example, the two smallest ones). We are currently investigating this approach.

References

[1] Edward Anderson, Zhaojun Bai, Christian Bischof, James Demmel, Jack Dongarra, Jeremy DuCroz, Anne Greenbaum, Sven Hammarling, Alan McKenney, and Danny Sorensen. LAPACK: A portable linear algebra library for high-performance computers. In Joanne Martin, editor, *SUPERCOMPUTING '90*, pages 2–10, New York, 1990. ACM Press. Also LAPACK working note # 20, CS-90-105.

[2] Christian Bischof, James Demmel, Jack Dongarra, Jeremy Du Croz, Anne Greenbaum, Sven Hammarling, and Danny Sorensen. LAPACK Working Note #5: Provisional contents. Technical Report ANL–88–38, Argonne National Laboratory, Mathematics and Computer Science Division, September 1988.

[3] Christian H. Bischof. A parallel QR factorization algorithm with controlled local pivoting. Technical Report ANL/MCS–P21–1088, Argonne National Laboratory, Mathematics and Computer Science Division, 1988.

[4] Christian H. Bischof. A block QR factorization algorithm using restricted pivoting. In *Proceedings SUPERCOMPUTING '89*, pages 248–256, Baltimore, MD, 1989. ACM Press.

[5] Christian H. Bischof. Incremental condition estimation. *SIAM Journal on Matrix Analysis and Applications*, 11(2):312–322, 1990.

[6] Christian H. Bischof and Per Christian Hansen. Structure-preserving and rank-revealing QR factorizations. Preprint MCS-P100-0989, Argonne National Laboratory, Mathematics and Computer Science Division, September 1989.

[7] Christian H. Bischof, Daniel J. Pierce, and John G. Lewis. Incremental condition estimation for sparse matrices. *SIAM Journal on Matrix Analysis and Applications*, 11(4):644–659, 1990.

[8] Christian H. Bischof and Gautam M. Shroff. On updating signal subspaces. Preprint MCS-P101-0989, Argonne National Laboratory, Mathematics and Computer Science Division, September 1989.

[9] Åke Björck. *Difference Methods – Solutions of Equations in R^n*, volume I of *Handbook of Numerical Analysis*, chapter Least Squares Methods. Elsevier Publishers, 1990.

[10] James R. Bunch, Christopher R. Nielsen, and Danny C. Sorensen. Rank-one modification of the symmetric eigenproblem. *Numerische Mathematik*, 31:31–48, 1978.

[11] P. A. Businger and G. H. Golub. Linear least squares solution by Householder transformation. *Numerische Mathematik*, 7:269–276, 1965.

[12] A. K. Cline, A. R. Conn, and C. F. Van Loan. *Generalizing the LINPACK Condition Estimator*, volume 909 of *Lecture Notes in Mathematics*, pages 73–83. Springer Verlag, 1982.

[13] A. K. Cline, C. B. Moler, G. W. Stewart, and J. H. Wilkinson. An estimate for the condition number of a matrix. *SIAM Journal on Numerical Analysis*, 16:368–375, 1979.

[14] William Ferng, Gene H. Golub, and Robert J. Plemmons. Adaptive Lanczos methods for recursive condition estimation. In *SPIE Volume 1348, Advanced Signal-Processing Algorithms, Architectures, and Implementations*, pages 326–337, Washington, D. C., 1990. The International Society for Optical Engineering.

[15] P. E. Gill and W. Murray. A numerically stable form of the simplex method. *Linear Algebra and Its Applications*, 7:99–138, 1973.

[16] P. E. Gill, G. H. Golub, W. Murray, and M. A. Saunders. Methods for modifying matrix factorizations. *Mathematics of Computation*, 28:505–535, 1974.

[17] Gene H. Golub. Numerical methods for solving linear least squares problems. *Numerische Mathematik*, 7:206–216, 1965.

[18] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, 1983, Baltimore.

[19] Nicholas J. Higham. Efficient algorithms for computing the condition number of a tridiagonal matrix. *SIAM Journal on Scientific and Statistical Computing*, 7:150–165, 1986.

[20] Nicholas J. Higham. A survey of condition number estimation for triangular matrices. *SIAM Review*, 29(4):575–596, 1987.

[21] Nicholas J. Higham. FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation. *ACM Transactions on Mathematical Software*, 14(4):381–396, 1988.

[22] Nicholas J. Higham. Experience with a matrix norm estimator. *SIAM Journal on Scientific and Statistical Computing*, 1990. to appear.

[23] Charles L. Lawson and Richard J. Hanson. *Solving Least Squares Problems.* Prentice-Hall, Englewood Cliffs, N.J., 1974.

[24] Daniel J. Pierce and Robert J. Plemmons. Fast adaptive condition estimation. Technical Report ECA-TR-146, Boeing Computer Services, Engineering and Scientific Services Division, October 1990.

[25] Daniel J. Pierce and Robert J. Plemmons. Tracking the condition number for RLS in signal processing. Technical Report ECA-TR-134, Boeing Computer Services, Engineering and Scientific Services Division, March 1990.

[26] Gautam M. Shroff and Christian H. Bischof. Adaptive condition estimation for rank-one updates of QR factorizations. Preprint MCS-P166-0790, Argonne National Laboratory, Mathematics and Computer Science Division, 1990.

[27] G. W. Stewart. The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM Journal on Numerical Analysis*, 17:403–409, 1980.

[28] James H. Wilkinson. *The Algebraic Eigenvalue Problem.* Clarendon Press, 1965.