Generalizing 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

Abstract. This paper presents a generalization of *incremental condition estimation*, a technique for tracking the extremal singular values of a triangular matrix. While the original approach allowed for the estimation of the largest or smallest singular value, the generalized scheme allows for the estimation of *any number* of extremal singular values. For example, we can derive estimates for the three smallest singular values and the corresponding singular vectors at the same time. When estimating k singular values at the same time, the cost of one step of our generalized scheme on an $n \times n$ matrix is $O(nk^2)$. Experimental results show that the resulting estimator does a good job of estimating the extremal singular values of triangular matrices and that, in particular, it leads to an inexpensive, yet very accurate and robust condition estimator.

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 [14, p. 19]. If one lets $\sigma_{\max} \equiv \sigma_1$, the condition number

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

determines the sensitivity of equation systems involving A [14, 21]. 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 [15].

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

*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.

Incremental condition estimation [3, 5] 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 [1, 2, 4, 6].

Adaptive condition estimation schemes address the issue of rank-one updates of a triangular matrix R. Pierce and Plemmons [17, 16] suggest an $O(n)$ scheme and Ferng, Golub, and Plemmons [11] 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 [18] 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 condition estimation maintains condition estimates when information is added or extracted from an already existing factorization.

In this paper, we generalize incremental condition estimation to estimate any number of extremal singular values (and the corresponding vectors) instead of just the largest or smallest one as the original incremental condition estimation scheme suggested. In Section 2 we show how we can estimate k extremal singular values by computing the eigensystem of a $(k+1) \times (k+1)$ symmetric rank-one perturbed diagonal matrix, and we give some examples of how this scheme could be used. In Section 3 we discuss the issues that arise in the accurate solution of this eigensystem in floating-point arithmetic. In Section 4 we present numerical experiments showing that the scheme delivers good singular value estimates and that, in particular, it delivers more reliable estimates for the smallest singular value than the original incremental condition estimation (ICE) scheme. Lastly, we present some concluding remarks and suggest some problems where it is useful to obtain estimates of several extremal singular values and vectors.

2 Derivation of the Generalized Incremental Condition Estimation Scheme

Assume that we are interested in estimating k extremal singular values $\sigma_{i_1}, \ldots, \sigma_{i_k}$ $(i_j < i_{j+1})$ of an $m \times m$ upper triangular matrix R. Our scheme will be able to estimate extremal singular values at both ends of the spectrum. Thus, for $k = 3$, for example, we have four choices: $(i_1, i_2, i_3) = (1,2,3), (1,2,m), (1,m-1,m),$ and $(m-2, m-1, m)$. Further assume that we have singular value estimates $\tau_1, \tau_2, \ldots, \tau_k$ and a corresponding set of orthonormal approximate left singular vectors x_j, that is,

$$\|x_j^H R\|_2 = \tau_j, \quad \tau_j \approx \sigma_{i_j}, \quad j = 1, \ldots, k,$$

and $X^H X = I$, where $X = [x_1, \ldots, x_k]$.

Our goal is to obtain updated singular value estimates $\hat{\tau}_i, \hat{\tau}_2, \ldots, \hat{\tau}_k$ and corresponding approximate singular vectors $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k$ for the augmented matrix

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

where w is an m-vector and γ a scalar. A scheme that accomplishes this task cheaply (that is, in less than $O(n^2)$ work), is called an ICE(k) estimator. We present the derivation for complex matrices, which is in particular applicable to real matrices.

As was mentioned in [7], incremental condition estimation can be motivated by the following well-known projection property of singular values. Let A be an $n \times n$ complex matrix and Y be an $n \times l$, $l \leq n$, complex matrix of orthonormal columns, that is, $Y^H Y = I$. Then,

$$\sigma_1(A) \geq \sigma_1(Y^H A), \quad \sigma_2(A) \geq \sigma_2(Y^H A), \quad \ldots, \quad \sigma_l(A) \geq \sigma_l(Y^H A) \tag{2.1}$$

and

$$\sigma_n(A) \leq \sigma_l(Y^H A), \quad \sigma_{n-1}(A) \leq \sigma_{l-1}(Y^H A), \quad \ldots, \quad \sigma_{n-l+1}(A) \leq \sigma_1(Y^H A). \tag{2.2}$$

The ICE scheme suggested in [3, 7] applies these inequalities to estimate either $\sigma_{\max}$ or $\sigma_{\min}$ as follows. We have

$$\|x_1^H R\|_2 = \tau_1, \quad \text{where } \tau_1 \approx \sigma_{\max} \text{ or } \tau_1 \approx \sigma_{\min}.$$

We let $k = 1$,

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

Inequalities 2.1 and 2.2 suggest that we calculate the extreme singular values and vectors of $Y^H A$ in order to estimate the extremal singular values of A. To this end, we consider the eigensystem of

$$\begin{aligned} M &= Y^H A A^H Y \\ &= Y^H \hat{R} \hat{R}^H Y \\ &= \begin{bmatrix} \tau_1^2 & \\ & 0 \end{bmatrix} + \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} [\bar{\alpha}, \bar{\gamma}], \quad \text{where } \alpha = x_1^H w. \end{aligned}$$

Denoting M's eigenvalues by λ_1, λ_2, where $\lambda_1 \geq \lambda_2$, and the corresponding eigenvectors by z_1 and z_2, we obtain the following ICE(1) estimates for $\hat{R}$:

$$\left\{ \begin{array}{l} \hat{\tau}_1 = \sqrt{\lambda_1} \text{ and } \hat{x}_1 = Y z_1 \\ \hat{\tau}_1 = \sqrt{\lambda_2} \text{ and } \hat{x}_1 = Y z_2 \end{array} \right\} \quad \text{if} \quad \left\{ \begin{array}{l} \hat{\tau}_1 \approx \sigma_{\max}(R) \\ \hat{\tau}_1 \approx \sigma_{\min}(R) \end{array} \right\}.$$

This scheme can be generalized as follows. Assume we have approximate singular vectors $x_j, j = 1, \ldots, k$ corresponding to the k extremal singular values $\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_k}$ of A. That is,

$$\|x_j^H R\|_2 = \tau_j, \quad \tau_j \approx \sigma_{i_j}, \quad j = 1, \ldots, k,$$

and $X^H X = I$, where $X = [x_1, \ldots, x_k]$. Now let

$$A = \hat{R} = \begin{bmatrix} R & w \\ & \gamma \end{bmatrix}, Y = \begin{bmatrix} x_1 & \cdots & x_k & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \in \mathbb{C}^{(m+1)\times k+1}, \quad \text{and} \quad l = k+1.$$

To calculate the extreme singular values of $Y^H A$, we compute the eigensystem of $M = Y^H A A^H Y$. Denote M's eigenvalues by $\lambda_1, \lambda_2, \ldots, \lambda_l$, where $\lambda_j \geq \lambda_{j+1}$, and denote the corresponding eigenvectors by $z_1, z_2, \ldots, z_{k+1}$. Suppose that the first ℓ estimates $\tau_1, \tau_2, \ldots, \tau_\ell$ approximate R's large singular values, $\tau_1 \approx \sigma_1(R), \tau_2 \approx \sigma_2(R), \ldots, \tau_\ell \approx \sigma_\ell(R)$, and suppose that the remaining $k - \ell$ estimates approximate R's small singular values, $\tau_k \approx \sigma_m(R), \tau_{k-1} \approx \sigma_{m-1}(R), \ldots, \tau_{\ell+1} \approx \sigma_{m-(k-\ell-1)}(R)$. We then use $\hat{\tau}_j$'s as new estimates for $\hat{R}$ defined by

$$\hat{\tau}_j = \sqrt{\lambda_j}, \quad \hat{x}_j = Y z_j, \quad j = 1, 2, \ldots, \ell,$$

and

$$\hat{\tau}_j = \sqrt{\lambda_{j+1}}, \quad \hat{x}_j = Yz_{j+1}, \quad j = \ell+1, \ell+2, \ldots, k.$$

We have, therefore,

$$\|\hat{x}_j^H \hat{R}\|_2 = \hat{\tau}_j, \quad j = 1, 2, \ldots, k,$$

and $\hat{X}^H \hat{X} = I$. The first ℓ of the $\hat{\tau}_j$'s approximate $\hat{R}$'s large singular values, and the remaining $k-\ell$ approximate $\hat{R}$'s small singular values. Moreover, from the Inequalities 2.1 and 2.2, we have

$$\hat{\tau}_1 \leq \sigma_1(\hat{R}), \hat{\tau}_2 \leq \sigma_2(\hat{R}), \ldots, \hat{\tau}_\ell \leq \sigma_\ell(\hat{R}),$$

and

$$\hat{\tau}_k \geq \sigma_m(\hat{R}), \hat{\tau}_{k-1} \geq \sigma_{m-1}(\hat{R}), \ldots, \hat{\tau}_{\ell+1} \geq \sigma_{m-(k-\ell-1)}(\hat{R}).$$

Hence, our scheme will underestimate the large singular values and overestimate the small ones.

To illustrate the scheme, let us consider $k = 4$ and $\ell = 2$, that is, $\tau_1 \approx \sigma_1(R), \tau_2 \approx \sigma_2(R), \tau_3 \approx \sigma_{m-1}(R)$, and $\tau_4 \approx \sigma_m(R)$. We let

$$Y = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $M = Y^H \hat{R}\hat{R}^H Y$, a 5×5 Hermitian matrix. Denote M's eigenvalues by $\lambda_1, \lambda_2, \ldots, \lambda_5$ and its eigenvectors by $z_1, z_2, \ldots, z_5$. We now take M's two extreme large and two extreme small singular pairs to construct our estimates for $\hat{R}$:

$$\hat{\tau}_1 = \sqrt{\lambda_1}, \hat{\tau}_2 = \sqrt{\lambda_2}, \hat{\tau}_3 = \sqrt{\lambda_4}, \hat{\tau}_4 = \sqrt{\lambda_5},$$

and

$$\hat{x}_1 = Yz_1, \hat{x}_2 = Yz_2, \hat{x}_3 = Yz_4, \hat{x}_4 = Yz_5.$$

The computational cost of our scheme involves the computation of M's eigensystem and k matrix-vector multiplications Yz_j. The matrix-vector multiplications cost $2mk^2$ flops total. This is the dominant cost, since our choice of Y gives M a simple structure:

$$\begin{aligned} M &= \begin{bmatrix} x_1 & \ldots & x_k & 0 \\ 0 & \ldots & 0 & 1 \end{bmatrix}^H \begin{bmatrix} R & w \\ & \gamma \end{bmatrix} \begin{bmatrix} R^H & \\ w^H & \bar{\gamma} \end{bmatrix} \begin{bmatrix} x_1 & \ldots & x_k & 0 \\ 0 & \ldots & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \tau_1^2 & & & & \\ & \tau_2^2 & & & \\ & & \ddots & & \\ & & & \tau_k^2 & \\ & & & & 0 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \\ \gamma \end{bmatrix} [\bar{\alpha_1} \quad \bar{\alpha_2} \quad \ldots \quad \bar{\alpha_k} \quad \bar{\gamma}], \quad \text{where } \alpha_j = x_j^H w. \end{aligned}$$

Thus, M is a symmetric rank-one perturbed diagonal matrix. The eigensystems of such matrices have been well studied theoretically (see [8, 12, 13] for example), and they can be computed considerably more efficiently than those of general $l \times l$ matrices. However, the computation of M's eigensystem using floating-point arithmetic was not resolved satisfactorily until the recent work of Sorensen and Tang [19]. We therefore devote the next section to discussing the computational issues of our condition estimation scheme.

3 Eigensystems of Symmetric Rank-One Perturbed Diagonal Matrices

The structure of M can be expressed as

$$\begin{aligned} M &= \begin{bmatrix} \delta_1 & & & \\ & \delta_2 & & \\ & & \ddots & \\ & & & \delta_l \end{bmatrix} + \rho \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_l \end{bmatrix} [\, \bar{\beta}_1 \quad \bar{\beta}_2 \quad \ldots \quad \bar{\beta}_l \,] \\ &= D + \rho b b^H, \quad \rho > 0 \text{ and } \|b\|_2 = 1. \end{aligned}$$

By applying the deflation techniques in [8, 19] if necessary, we can assume $\delta_1 > \delta_2 > \ldots \delta_l$ and $|\beta_j| > \varepsilon$ for $j = 1, 2, \ldots, l$, where ε is the machine precision. As the derivation of our generalized ICE scheme suggests, the mutual orthogonality of the approximate singular vectors of M is crucial. Consequently, it is important that the computed eigenvectors of M be numerically orthogonal. As it turns out, fulfilling this requirement is not a straightforward task.

With the assumptions that $\delta_1 > \delta_2 > \ldots > \delta_l$ and $|\beta_j| > \varepsilon$ (in particular, that $|\beta_j| > 0$), M's eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_l$ satisfy the well-known interlacing property

$$\lambda_1 > \delta_1 > \lambda_2 > \delta_2 > \ldots > \lambda_l > \delta_l.$$

An eigenvector corresponding to λ_j is given by

$$\begin{bmatrix} \delta_1 - \lambda_j & & & \\ & \delta_2 - \lambda_j & & \\ & & \ddots & \\ & & & \delta_l - \lambda_j \end{bmatrix}^{-1} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_l \end{bmatrix}.$$

Thus, provided that the differences $\Delta_{ij} \equiv \delta_i - \lambda_j$, $j = 1, 2, \ldots, l$, can be computed to full machine precision, each component of the computed eigenvectors will be fully accurate, yielding a set of numerically orthogonal eigenvectors.

A standard approach to determine the differences Δ_{ij} is to numerically solve for the roots of the secular equation (see [8] for example):

$$f(\lambda) = 1 + \rho \sum_{j=1}^{l} \frac{|\beta_j|^2}{\delta_j - \lambda}.$$

Although different root finders had been proposed [8, 10], none could always guarantee numerical orthogonality — all failed occasionally in a few examples, either contrived or natural. Only recently did the analysis in [19] explain that there is an intrinsic difficulty related to the accuracy in which f is evaluated. It was shown there that, for any given j, the magnitude

$$\frac{\varepsilon \max_i |\beta_i / (\delta_i - \lambda_j)|}{f'(\lambda_j)}$$

is a key indicator of the accuracy of the calculated differences Δ_{ij}. In particular, since $|\beta_j| \geq \varepsilon$ for $j = 1, 2, \ldots, l$, the resulting bounds obtained for those magnitudes imply that numerical orthogonality is guaranteed provided f is evaluated in a precision that doubles the working precision. An efficient algorithm fulfilling this requirement without the need of a double-precision data type is also given in [19]. We have employed this eigensystem solver throughout our condition estimation scheme.

4 Numerical Results

Our experiments were designed to answer several questions. First, we wished to see how much better the original scheme is in estimating the smallest singular value of a matrix, since this is the difficult part in obtaining a reliable two-norm condition number estimate. Second, we wished to get some idea of how the accuracy of the generalized schemes varied as we increased k. Lastly, we wished to get a feeling for the accuracy with which the generalized schemes estimate the other extremal singular values whose estimates are being maintained.

In our tests, we employed various triangular matrices. For the first class of matrices, 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 [20] to generate random orthogonal matrices U and V. The upper-triangular matrices R used in testing were the R factor of the QR decomposition,

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

Three distributions for the singular values were used:

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

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

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

Other matrices were

Random: R was the upper triangular factor of a QR factorization of a full matrix with elements uniformly random distributed in $(0,1)$.

Difficult: R was generated so as to be a "difficult" matrix for the ICE(1) scheme for estimating the smallest singular value and was generated as follows. Given an $m \times m$ triangular matrix $T(m)$, and an approximate smallest left singular vector of T as generated by the ICE(1) scheme, we augmented T by generating a random m-vector z, subtracting most of its contribution in the direction of x. We then adjusted the diagonal entry γ such that

$$T(m+1) = \begin{bmatrix} T(m) & z \\ & \gamma \end{bmatrix}$$

was not too ill conditioned. Starting with $m = 2$, we thus built up an $n \times n$ matrix. As was shown in [3], the ICE(1) scheme can produce arbitrarily bad estimates when z is orthogonal to x. While this is not the case here, these matrices should be more likely than others to cause the ICE(1) scheme to produce large overestimates of the smallest singular value.

4.1 Accuracy of ICE(k) Estimates for Estimating the Smallest Singular Value

Our first set of experiments was designed to show to what extent the generalized ICE scheme improves the estimate for the smallest singular value. Let

$$r_{\min}(k) \equiv \frac{\tau_{\min}(k)}{\sigma_{\min}},$$

where $\tau_{\min}(k)$ is the estimate for $\sigma_{\min}$ produced by the ICE(k) estimate with $\ell = 0$, that is, the ICE-vectors are approximations of the singular vectors corresponding to the k smallest singular values.

Table 1: ICE(2) versus ICE(1)

Distribution	$r_{\min}(1)$ Median	$r_{\min}(1)$ Worst	$r_{\min}(2)$ Median	$r_{\min}(2)$ Worst	$r_{\max}(1)$ Median	$r_{\max}(1)$ Worst	$r_{\max}(2)$ Median	$r_{\max}(2)$ Worst
Exponential	3.10	4.47	2.82	3.63	1.24	1.82	1.17	1.72
Randomlog	3.03	5.73	2.70	5.62	1.20	1.66	1.14	1.60
Cluster	4.75	17.00	3.89	9.92	1.15	1.26	1.13	1.21
Random	4.31	17.28	3.15	9.82	1.00	1.01	1.00	1.00
Difficult	1.06	2181.30	1.03	15.18	2.18	2.54	1.66	2.09

Thus, $r_{\min}(k)$ is an overestimate of the smallest singular value by an ICE(k) estimate. Figure 1 shows average and maximum values of $r_{\min}(k)$, where k ranges from 1 to 6. The experiments reflect 50 trials with 50×50 matrices having a "random", "exponential", and "cluster" distribution. For example, for the 'random' distribution, the worst-case value for $r_{\min}(1)$ is 16.7, whereas the worst-case value for $r_{\min}(2)$ is 6.4. In contrast, the average overestimate decreases from 3.8 to 2.9. These results (as well as the results in the other plots, though in a somewhat less spectacular way) suggest that the increase in accuracy for the smallest singular value is not very pronounced on average, but that (as expected) the generalized ICE schemes are more robust, in that the worst-case overestimate is reduced significantly. The plots also suggest that for the purposes of estimating only the smallest singular value, using $k = 2$, that is ICE(2), seems to be fully sufficient, and that the added computional cost for higher-degree ICE schemes is not rewarded by great improvements in the estimate for $\sigma_{\min}$.

4.2 ICE(2) Estimates for the Largest or Smallest Singular Values

The experiments of the preceding section prompted us to investigate the behavior of the ICE(2) estimator in more detail. Table 1 shows (for $k = 1, 2$) the median and worst-case values for $r_{\min}(k)$ and

$$r_{\max}(k) \equiv \frac{\sigma_{\max}}{\tau_{\max}(k)},$$

where $\tau_{\max}(k)$ is the estimate for the largest singular value of R using an ICE(k) estimator that approximates the k largest singular values. These experiments reflect 100 trials with matrices of size 100 and 200 each. As was suggested by the previous round of experiments, the generalized scheme does prevent "outliers" and increases the average accuracy of the condition estimation scheme. This point is made in a particularly impressive fashion by the "difficult" matrices. As mentioned before, they were designed to break the ICE(1) scheme, and they did. On one matrix, the ICE(1) scheme overestimates the smallest singular value by a factor of more than 2,000. Equally surprising to us, however, was the fact that the median overestimate for this distribution was only 1.06. Using the ICE(2) scheme, the large overestimates disappear. The improvement realized by the ICE(2) scheme in estimating $\sigma_{\max}$ was somewhat disappointing, largely in view of the already good performance of the ICE(1) scheme.

While our exploration of the generalized ICE scheme was mainly motivated by the desire to develop a more robust condition estimator, the ICE(2) schemes we employed also produce estimates for σ_{n-1} and σ_2. In Table 2 we display the median and worst-case values of

$$r_{n-1} \equiv \frac{\tau_{n-1}}{\sigma_{n-1}}$$

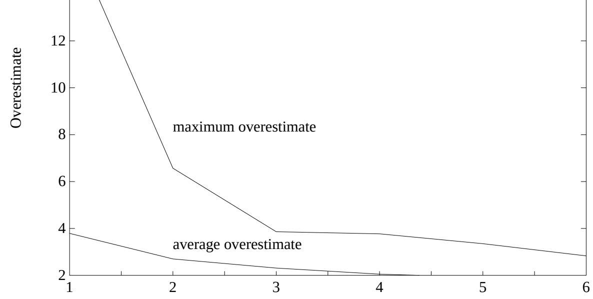
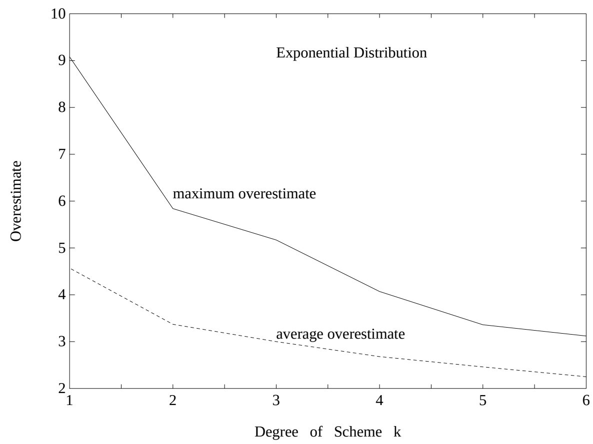


Figure 1: Overestimate of Smallest Singular Value by ICE(k)



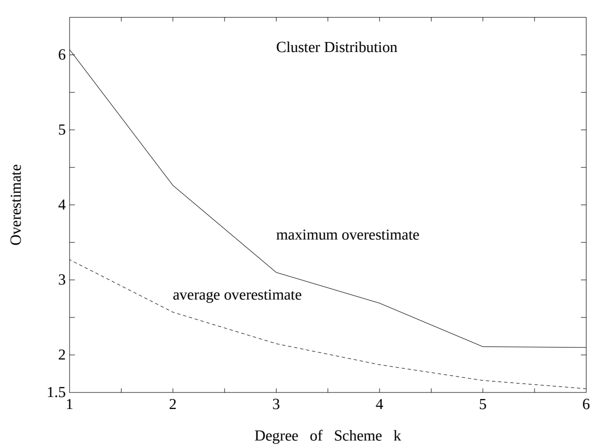


Table 2: Accuracy of ICE2 on σ_2 and σ_{n-1}

Distribution	r_{n-1} Median	r_{n-1} Worst	r_2 Median	r_2 Worst
Exponential	3.21	4.50	1.40	1.98
Randomlog	3.08	4.79	1.38	2.95
Cluster	3.92	11.42	1.17	1.33
Random	3.41	12.54	12.91	15.82
Difficult	1.37	24.83	3.73	5.02

and

$$r_2 \equiv \frac{\sigma_2}{\tau_2},$$

which are the factors by which the ICE(2) schemes overestimated the next-to-smallest singular value and underestimated the second singular value on the test matrices used for Table 1. The performance of the ICE(2) scheme for estimating σ_{n-1} is comparable to the performance of ICE(1) for estimating $\sigma_{\min}$. Similarly, the estimates of ICE(2) for σ_2 and the ICE(1) estimate of σ_1 are comparable, with the ICE(2) estimate for σ_2 being slightly worse. At any rate, however, the quality of the ICE estimates would be sufficient to gain a qualitative understanding of the separation of the extremal singular values.

Using the ICE(2) estimates for $\sigma_{\min}$ and $\sigma_{\max}$, we do obtain a condition estimator that is considerably more robust than the one derived from the original ICE(1) scheme. This becomes apparent when one considers

$$\frac{\hat{\kappa}(1)}{\hat{\kappa}(2)},$$

the ratio of the ICE(1) condition estimate

$$\hat{\kappa}(1) \equiv \frac{\tau_{\min}(1)}{\tau_{\max}(1)}$$

to the ICE(2) condition estimate

$$\hat{\kappa}(2) \equiv \frac{\tau_{\min}(2)}{\tau_{\max}(2)}.$$

Figure 2 shows a histogram of

$$\log_2(\frac{\hat{\kappa}(1)}{\hat{\kappa}(2)}).$$

In 859 out of the 1,000 test cases, the generalized scheme improved the quality of the estimates, by a factor between 1 and 2 in 785 cases, by a factor between 2 and 4 in 48 cases, and by a factor between 2^9 and 2^{10} in one case. In 141 cases the original ICE(1) scheme did produce better results than the generalized scheme, although the improvements were not pronounced. For those 141 cases, the median ratio for $\log_2(\frac{\hat{\kappa}(1)}{\hat{\kappa}(2)})$ was 0.96, with the smallest value being 0.53. This situation is also reflected in Figure 3, which shows in more detail how the ICE(2) condition estimate compared to the ICE(1) estimate. We only show the three distributions, where (according to Table 1) the ICE(2) scheme led to noticeable improvements, namely, the "difficult," "cluster," and "random" distribution. The plot shows $\log_2(\frac{\hat{\kappa}(1)}{\hat{\kappa}(2)})$, and the first 100 data points for each distribution correspond to the matrices of size 100, the seond 100 data points correspond to the matrices of size 200. The dotted line at 0 represents the cases where the ICE(1) and ICE(2) estimate produced the same

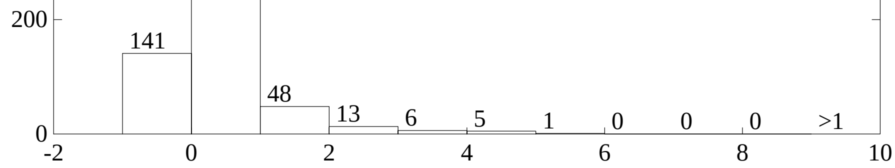


Figure 2: Improvement of Condition Estimate through Use of Generalized Scheme
Histogram of $\log_2(\hat{\kappa}(1)/\hat{\kappa}(2))$

results. When the plot is above this line, ICE(2) produces a better estimate, if it is below, ICE(1) is better. The "difficult' distribution again confirms the findings suggested by Table 1: ICE(2) is usually more reliable, and in particular it is much less likely to produce large overestimates than is the ICE(1) scheme.

4.3 ICE(2) Estimates for Both the Largest and Smallest Singular Values

Lastly, we used the ICE(2) scheme to estimate both the smallest and largest singular value at the same time (that is, using the notation of Section 2, $i_1 = 1$ and $i_2 = n$). Table 3 shows $r_{\min}$ and $r_{\max}$ when the smallest and largest singular values are estimated at the same time using ICE(2) and independently using the ICE(1) scheme. As we can see, there is not a tremendous difference

Table 3: Estimating $\sigma_{\min}$ and $\sigma_{\max}$ at the Same Time Using ICE2 and Independently Using ICE1

	$r_{\min}$(ICE1)		$r_{\min}$(ICE2)		$r_{\max}$(ICE1)		$r_{\max}$(ICE2)	
Distribution	Median	Worst	Median	Worst	Median	Worst	Median	Worst
Arithmetic	5.53	13.22	5.50	15.53	1.14	1.18	1.14	1.19
Exponential	3.13	3.65	1.09	3.09	1.22	1.39	1.19	1.34
Randomlog	3.16	4.34	1.57	3.23	1.26	1.67	1.21	1.43
Random	6.55	14.4	6.05	11.1	1.00	1.00	1.00	1.00
Difficult	1.96	489.83	1.88	242.90	2.40	2.67	2.22	2.59

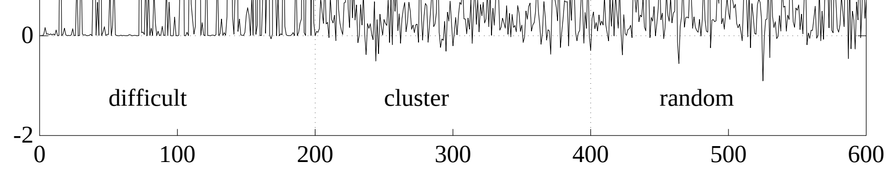


Figure 3: ICE(1) versus ICE(2) Estimate for Selected Distributions
Graphs of $\log_2(\hat{\kappa}(1)/\hat{\kappa}(2))$

between the two approaches, although the ICE(2) scheme produces more accurate results for the smallest singular value.

5 Concluding Remarks

We presented a generalization of incremental condition estimation, a technique for estimating the extremal singular values of a triangular matrix as it is being generated one column at a time. In contrast to previous schemes, which estimated either the smallest or largest singular value, our generalized scheme can estimate any number of extremal singular values, from either the small or large end of the spectrum. Usually, the estimation of k singular values requires $O(2nk^2)$ flops at every step, with the computationally subtle part being the accurate computation of the eigensystem of a symmetrically perturbed diagonal $(k+1) \times (k+1)$ matrix.

We performed a series of experiments showing that the generalized ICE scheme significantly improves the robustness of the estimate for the smallest singular value. The experiments also showed that other extremal singular values were approximated well. Computationally, the ICE(2) scheme (the simplest generalized scheme) seems the most promising for the purposes of condition estimation, and we are working on a streamlined implementation of this case.

While the main focus of our paper was to increase the robustness of ICE for the purposes of condition estimation, there are applications where it is advantageous to have estimates of several singular values and vectors. One is subspace iteration, where the generalized ICE scheme provides excellent starting vectors. Another use arises in rank estimation, where one can deduce both the rank and the relative gap between small and large singular values from the quantities provided by ICE, and the subsequent use of such a factorization in the solution of rank-deficient least squares systems [9].

References

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

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

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

[4] 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.

[5] 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.

[6] 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.

[7] Christian H. Bischof and Ping Tak Peter Tang. Robust incremental condition estimation. Preprint MCS-P225-0391, Argonne National Laboratory, Mathematics and Computer Science Division, 1991.

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

[9] Tony F. Chan and Per Christian Hansen. Computing truncated SVD least squares solutions by rank revealing QR factorizations. *SIAM Journal on Scientific and Statistical Computing*, 11(3):519–530, 1990.

[10] J. M. J. Cuppen. A divide and conquer method for the symmetric tridiagonal eigenproblem. *Numerische Mathematik*, 36:31–48, 1981.

[11] 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.

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

[13] Gene H. Golub. Some modified matrix eigenvalue problems. *SIAM Review*, 15(2):318–334, 1973.

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

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

[16] 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.

[17] 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.

[18] 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, July 1990.

[19] Danny Sorensen and Peter Tang. On the orthogonality of eigenvectors computed by a divide-and-conquer method. Preprint MCS-P152-0490, Argonne National Laboratory, Mathematics and Computer Science Division, April 1990.

[20] 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.

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