University of Maryland College Park

Institute for Advanced Computer Studies TR–95–93
Department of Computer Science TR–3535

On the Perturbation of LU and Cholesky Factors*

G. W. Stewart†

October, 1995

ABSTRACT

In a recent paper, Chang and Paige have shown that the usual perturbation bounds for Cholesky factors can systematically overestimate the errors. In this note we sharpen their results and extend them to the factors of the LU decomposition. The results are based on a new formula for the first order terms of the error in the factors.

*This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports`.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the National Science Foundation under grant CCR 95503126.

On the Perturbation of
LU and Cholesky Factors

G. W. Stewart

ABSTRACT

In a recent paper, Chang and Paige have shown that the usual perturbation bounds for Cholesky factors can systematically overestimate the errors. In this note we sharpen their results and extend them to the factors of the LU decomposition. The results are based on a new formula for the first order terms of the error in the factors.

1. Introduction

Let A be a positive definite matrix of order n. Then A has a unique Cholesky factorization of the form $A = R^{\mathrm{T}}R$, where R is upper triangular with positive diagonal elements.

Let $\tilde{A} = A + E$ be a perturbation of A in which E is symmetric. If E is sufficiently small, then $\tilde{A}$ also has a Cholesky factorization:

$$A + E = (R + F_R)^{\mathrm{T}}(R + F_R).$$

Several workers [1, 3, 4, 7] have given bounds on the matrix F_R. The common result is essentially that

$$\frac{\|F_R\|_{\mathrm{F}}}{\|R\|_2} \leq \frac{1}{\sqrt{2}}\kappa_2(A)\frac{\|E\|_{\mathrm{F}}}{\|A\|_2} + O(\|E\|_2^2). \tag{1.1}$$

Recently Chang and Paige [2] have shown that (1.1) can consistently overestimate the error in the Cholesky factor and have proposed new bounds. Following [3], they note that

$$E = R^{\mathrm{T}}F_R + F_R^{\mathrm{T}}R + O(\|F_R\|^2). \tag{1.2}$$

Consequently if one defines the linear operator $\mathbf{T}_R$ on the space of upper triangular matrices by

$$\mathbf{T}_R(F) = R^{\mathrm{T}}F + F^{\mathrm{T}}R,$$

then

$$F_R \cong \breve{F}_R \equiv \mathbf{T}_R^{-1}(E),$$

and

$$\|F_R\| \lesssim \|\mathbf{T}_R^{-1}\|\|E\|,$$

where $\|\cdot\|$ denotes a suitably chosen norm. By examining the matrix representation of $\mathbf{T}_R$, Chang and Paige were able to show that bounds based on $\|\mathbf{T}_R^{-1}\|$ are sharper than the conventional bounds. They also derive a lower bound for $\|\mathbf{T}_R^{-1}\|$, and show by example that the failure of (1.1) is somehow connected with pivoting in the computation of the decomposition.

The purpose of this note is to generalize and strengthen the results of Chang and Paige. We do so by exhibiting an explicit matrix representation of $\breve{F}_R$. The representation is invariant under a certain kind of diagonal scaling, and by adjusting the scaling we can improve the usual bound. Since the approach works for the more general LU decomposition, we will treat that case first and then specialize to the Cholesky decomposition.

Throughout this note $\|\cdot\|$ will denote an absolute norm, such as the 1-norm, the ∞-norm, or the Frobenius norm, which will also be denoted by $\|\cdot\|_{\mathrm{F}}$. The matrix 2-norm, which is not absolute, will be denoted by $\|\cdot\|_2$. For any nonsingular matrix X we will define

$$\kappa_\nu(X) = \|X\|_\nu \|X^{-1}\|_\nu.$$

For more on norms see [5].

2. The LU Decomposition

Let A be a matrix of order n whose leading principal submatrices are nonsingular. Then A can be written in the form

$$A = LU,$$

where L is lower triangular and U is upper triangular. The decomposition is not unique, but it can be made so by specifying the diagonal elements of L. (The conventional choice is to require them to be one.)

If E is sufficiently small, $A + E$ has an LU factorization:

$$A + E = (L + F_L)(U + F_U). \tag{2.1}$$

Again, the factorization is not unique, but it can be made so, say by requiring that the diagonals of L remain unaltered.

Multiplying out the right hand side of (2.1) and ignoring higher order terms, we obtain a linear matrix equation for first order approximations $\breve{F}_L$ and $\breve{F}_U$ to F_L and F_U:

$$L\breve{F}_U + U\breve{F}_L = E.$$

We shall show how to solve this equation in terms of two matrix operators.

Let $0 \le p \le 1$, and define $\mathcal{L}_p$ and $\mathcal{U}_p$ as illustrated below for a 3×3 matrix:

$$\mathcal{L}_p(X) = \begin{pmatrix} px_{11} & 0 & 0 \\ x_{21} & px_{22} & 0 \\ x_{31} & x_{32} & px_{33} \end{pmatrix} \quad \text{and} \quad \mathcal{U}_p(X) = \begin{pmatrix} px_{11} & x_{12} & x_{13} \\ 0 & px_{22} & x_{23} \\ 0 & 0 & px_{33} \end{pmatrix}.$$

It then follows that for any matrix X,

$$X = \mathcal{L}_p(X) + \mathcal{U}_{1-p}(X), \tag{2.2}$$

and

$$\|\mathcal{L}_p(X)\|, \|\mathcal{U}_p(X)\| \le \|X\|.$$

Finally, if X is symmetric

$$\|\mathcal{U}_{\frac{1}{2}}(X)\|_{\mathrm{F}} \le \frac{1}{\sqrt{2}}\|X\|_{\mathrm{F}}. \tag{2.3}$$

Our basic result is the following:

$$\breve{F}_L = L\mathcal{L}_p(L^{-1}EU^{-1}) \quad \text{and} \quad \breve{F}_U = \mathcal{U}_{1-p}(L^{-1}EU^{-1})U.$$

To see this, write

$$\begin{aligned} &L[\mathcal{U}_{1-p}(L^{-1}EU^{-1})U] + [L\mathcal{L}_p(L^{-1}EU^{-1})]U \\ &\quad = L[\mathcal{U}_{1-p}(L^{-1}EU^{-1}) + \mathcal{L}_p(L^{-1}EU^{-1})]U \\ &\quad = L(L^{-1}EU^{-1})U \qquad\qquad \text{by (2.2)} \\ &\quad = E. \end{aligned}$$

The number p is a normalizing parameter, controling how much of the perturbation is attached to the diagonals of L and U. If $p = 0$, the diagonal elements of L do not change. If $p = 1$, the diagonal elements of U do not change.

We can take norms in the expressions $\breve{F}_L$ and $\breve{F}_U$ to get first order perturbation bounds for the LU decomposition. But it is possible to introduce degrees of freedom in the expressions that can later be used to reduce the bounds. Specifically, for any nonsingular diagonal matrix D_L, we have

$$\breve{F}_L = LD_L\mathcal{L}_0(D_L^{-1}L^{-1}EU^{-1}) \equiv \hat{L}\mathcal{L}_0(\hat{L}^{-1}EU^{-1})$$

Consequently

$$\|\breve{F}_L\| \le \|\hat{L}\|\|\hat{L}^{-1}\|\|U^{-1}\|\|E\|,$$

or

$$\frac{\|\breve{F}_L\|}{\|L\|} \le \frac{\kappa(\hat{L})\kappa(U)\|E\|}{\|L\|\|U\|}. \tag{2.4}$$

Since $\|A\| \le \|L\|\|U\|$, we have

$$\frac{\|\breve{F}_L\|}{\|L\|} \le \kappa(\hat{L})\kappa(U)\frac{\|E\|}{\|A\|}. \tag{2.5}$$

Similarly, if D_U is a nonsingular diagonal matrix and we set

$$\hat{U} = D_U U,$$

then

$$\frac{\|\breve{F}_U\|}{\|U\|} \le \kappa(L)\kappa(\hat{U})\frac{\|E\|}{\|A\|}. \tag{2.6}$$

The bounds (2.5) and (2.6) differ from the usual bounds (e.g., see [4]) by the substitution of $\hat{L}$ or $\hat{U}$ for L or U. However, if the diagonal matrices D_L and D_U are chosen appropriately, $\kappa(\hat{L})$ and $\kappa(\hat{U})$ can be far less that $\kappa(L)$ or $\kappa(U)$. For example, if

$$U = \begin{pmatrix} 1 & \epsilon \\ 0 & \epsilon \end{pmatrix}, \tag{2.7}$$

then $\kappa_1(U) \cong 1/\epsilon$. But if we set $D_U = \mathrm{diag}(1, 1/\epsilon)$, then $\kappa(\hat{U}) \cong 1$.

Poorly scaled but essentially well-conditioned matrices like U in (2.7) occur naturally. If A is ill-conditioned and the LU decomposition of A is computed with pivoting, the ill-conditioning of A will usually reveal itself in the diagonal elements of U. In [6] the author has shown that such upper triangular matrices are artificially ill conditioned in the sense that they can be made well conditioned by scaling their rows.

If $\kappa(\hat{L}) = 1$ (it cannot be less), then the bound (2.4) reduces to

$$\|\breve{F}_L\| \le \|U^{-1}\|\|E\|. \tag{2.8}$$

It is reasonable to ask if there are problems for which we can replace $\|U^{-1}\|$ by an even smaller number and still have inequality for all E. The answer depends on p. For example, suppose that $\|\cdot\|$ is the ∞-norm. Let $\mathbf{e}_i$ be the unit coordinate vectors, and let k be such that $\|\mathbf{e}_k^{\mathrm{T}} U^{-1}\| = \|U^{-1}\|$. Let $E = \mathbf{e}_n \mathbf{e}_k^{\mathrm{T}}$, so that $\|E\| = 1$ Then it is easy to see that

$$\|\hat{L}\mathcal{L}_p(\hat{L}^{-1}EU^{-1})\| = \|\mathcal{L}_p(\mathbf{e}_n\mathbf{e}_k^{\mathrm{T}}U^{-1})\|$$

Hence

$$\|U^{-1}\|\|E\| \geq \|\breve{F}_L\| \geq p\|U^{-1}\|\|E\|. \tag{2.9}$$

Consequently, if p is near one, (2.8) is essentially the smallest bound that holds uniformly for all E.

The reason for the appearance of the factor p in (2.9) is that the error may concentrate in the last column of $L^{-1}EU^{-1}$, in which case it is reduced by a factor of at least p by the operator $\mathcal{L}_p$. This can happen, for example, when $L = I$ and $U = \mathrm{diag}(I_{n-1}, \epsilon)$ for ϵ small. However, if p si small, the perturbation will show up in $\breve{F}_U$, for which the factor is $1 - p$.

The bounds (2.5) and (2.6) suggest a strategy for estimating the condition of the LU factorizations. Van der Sluis [8] has shown that in the 2-norm, the condition number is approximately minimized when the rows or columns of the matrix are scaled to have norm one. Thus the strategy is to so scale $\hat{L}$ and $\hat{U}$ and use a condition estimator to estimate the condition of L, $\hat{L}$, U, $\hat{U}$.

In [4] it is shown how to obtain rigorous bounds for the errors in the first order approximations $\breve{F}_L$ and $\breve{F}_U$. Since the second order terms decay rapidly, the error bounds are less important than the condition that insures their existence: namely,

$$\|L^{-1}\|\|U^{-1}\|\|E\| \leq \frac{1}{4}.$$

3. The Cholesky Decomposition

We now return to the Cholesky decomposition. In analyzing the perturbation of the the Cholesky factor R it is natural to take $p = \frac{1}{2}$ so that symmetry is preserved. In this case the solution of the perturbation equation becomes

$$\breve{F}_R = \mathcal{U}_{\frac{1}{2}}(R^{-\mathrm{T}}ER^{-1})R.$$

Hence if $\hat{R}$ is defined in analogy with $\hat{L}$ and $\hat{U}$, it follows from (2.3) that

$$\frac{\|\breve{F}_R\|_{\mathrm{F}}}{\|R\|_2} \leq \frac{1}{\sqrt{2}}\kappa_2(\hat{R})\kappa_2(R)\frac{\|E\|_{\mathrm{F}}}{\|A\|_2}.$$

Moreover, by a variant of the argument that lead to (2.9), for any A, there is an E such that

$$\|R^{-1}\|_2\|E\|_{\mathrm{F}} \geq \|\breve{F}_R\|_{\mathrm{F}} \geq \frac{1}{2}\|R^{-1}\|_2\|E\|_{\mathrm{F}}$$

which shows that we cannot reduce the constant ρ in the bound

$$\|\breve{F}_R\|_{\mathrm{F}} \leq \rho\|E\|_{\mathrm{F}}$$

to less than $\frac{1}{2}\|R^{-1}\|$.

References

[1] A. Barrland. Perturbation bounds for the LDL^{H} and the LU factorizations. *BIT*, 31:358–363, 1991.

[2] X-W. Chang and C. C. Paige. A new perturbaton analysis for the Cholesky factorization. School of Computer Science, McGill University. To appear in the *IMA Journal of Numerical Analysis.*, 1995.

[3] G. W. Stewart. Perturbation bounds for the QR factorization of a matrix. *SIAM Journal on Numerical Analysis*, 1977:509–518, 1977.

[4] G. W. Stewart. On the perturbation of LU, Cholesky, and QR factorizations. *SIAM Journal on Matrix Analysis and Applications*, 14:1141–1146, 1993.

[5] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory.* Academic Press, Boston, 1990.

[6] G. W. Stwart. The triangular matrices of Gaussian elimination and related decompositions. Technical Report CS-TR-3533 UMIACS-TR-95-91, University of Maryland, Department of Computer Science, 1995.

[7] J.-G. Sun. Rounding-error and perturbation bounds for the Cholesky and LDL^{T} factorizations. *Linear Algebra and Its Applications*, 173:77–98, 1992.

[8] A. van der Sluis. Condition numbers and equilibration of matrices. *Numerische Mathematik*, 14:14–23, 1969.