BIT 36:2 (1996), 247–263.

PERTURBATION AND ERROR ANALYSES FOR BLOCK DOWNDATING OF A CHOLESKY DECOMPOSITION *

L. ELDÉN[1] and H. PARK[2] †

[1] *Department of Mathematics, University of Linköping*
S–581 83 Linköping, Sweden email: laeld@math.liu.se

[2] *Computer Science Department, University of Minnesota*
Minneapolis, MN 55455, U.S.A. email: hpark@cs.umn.edu

Abstract.

A new perturbation result is presented for the problem of block downdating a Cholesky decomposition $X^TX = R^TR$. Then, a condition number for block downdating is proposed and compared to other downdating condition numbers presented in literature recently. This new condition number is shown to give a tighter bound in many cases. Using the perturbation theory, an error analysis is presented for the block downdating algorithms based on the LINPACK downdating algorithm and stabilized hyperbolic transformations. An error analysis is also given for block downdating using Corrected Seminormal Equations (CSNE), and it is shown that for ill-conditioned downdates this method gives more accurate results than the algorithms based on the LINPACK downdating algorithm or hyperbolic transformations. We classify the problems for which the CSNE downdating method produces a downdated upper triangular matrix which is comparable in accuracy to the upper triangular factor obtained from the QR decomposition by Householder transformations on the data matrix with the row block deleted.

AMS subject classification: 65F20, 65F35.

Key words: Block downdating, Cholesky decomposition, condition number, error analysis, perturbation theory, seminormal equations.

1 Introduction.

Let

$$X = \begin{pmatrix} Z \\ \tilde{X} \end{pmatrix} \in \mathbf{R}^{m\times n}, \tag{1.1}$$

be a matrix with $\mathrm{rank}(X) = n$, where its first k rows are denoted as $Z \in \mathbf{R}^{k\times n}$, $m \geq k+n$. Frequently, given the upper triangular Cholesky factor R for X^TX,

$$X^TX = R^TR, \tag{1.2}$$

*Received March 1995. Revised November 1995.

†The work of the second author was supported in part by the National Science Foundation grant CCR-9209726.

one needs to find the Cholesky factor of $\tilde{X}^T\tilde{X}$ from R and Z. This problem is called the **block downdating** of the Cholesky decomposition. From (1.1) and (1.2), we have the relation

$$R^TR = \tilde{X}^T\tilde{X} + Z^TZ$$

and we are to find the upper triangular matrix U that satisfies

$$\tilde{X}^T\tilde{X} = R^TR - Z^TZ = U^TU.$$

Throughout this paper, we assume that the matrices for which we want to compute the Cholesky factor have full rank,

$$\text{rank}(\tilde{X}) = \text{rank}(X) = n.$$

The problem of block downdating the Cholesky decomposition of X^TX is equivalent to the problem of block downdating the *QR decomposition* of X

$$X = Q\begin{pmatrix} R \\ 0 \end{pmatrix},$$

when the orthogonal factor Q is not stored. The problem of downdating a single row (when $k = 1$) of a Cholesky decomposition has been extensively studied in literature [1, 4, 11, 12, 19, 22, 24]. Downdating a block of rows ($k > 1$) with variations of the Householder transformations has been considered in [6, 21]. The so called LINPACK downdating algorithm due to Saunders [22], [10, p. 531], [24], for $k = 1$ was generalized to block downdating in [8] (see also [16]). The amount of work for downdating a $k \times n$ block Z in these algorithms is $O(kn^2)$ as opposed to $O((m-k)n^2)$ for recomputing the decomposition completely from $\tilde{X}$. Therefore, when $k \ll m$, which is usually the case, the block downdating of Cholesky decomposition is much faster.

Perturbation results for the Cholesky downdating problem for $k = 1$ were presented in [24] and more complete analyses in [18, 25]. In [9], we analyzed the Cholesky block downdating problem using a recent perturbation result for the Cholesky factorization [26]. A measure of conditioning for the Cholesky block downdating problem was presented and compared to the conditioning for the single row downdating given in [18, 24].

The condition numbers for downdating which are presented in [18] and [9] involve the square of the condition number of R, $\kappa^2(R)$. Recently Sun [25] proposed condition numbers for downdating which are proportional to $\kappa(R)$ instead of $\kappa^2(R)$. In Section 2, we give a new perturbation result for block downdating. This leads to an alternative condition number for downdating, which is also proportional to $\kappa(R)$. We show that in many cases our new perturbation bound is tighter than other bounds presented in [9, 18, 25]. Another advantage of the new perturbation results is that they can be used to detect ill-conditioned downdates before the downdated matrix U is actually computed, and ill-conditioned downdates can be performed by more accurate methods, based on Corrected Seminormal Equations (CSNE) [4, 8].

The new perturbation theory is applied in Section 3 to obtain forward error estimates for the LINPACK downdating algorithm and its block variant, as well as for the downdating algorithm based on stabilized hyperbolic transformations [7, 23].

In [4, 8] we introduced downdating algorithms based on CSNE, which use the original matrix X in addition to R and Z. These algorithms are more accurate than those that use only R and Z as data, and the complexity is still only $O(kmn)$[1]. In Section 3, we analyze a simplified version of the CSNE downdating algorithm and show that when downdating does not involve removing outliers[2], then, under a certain condition, the forward error in the downdated matrix U produced by this algorithm is *comparable to the error in the upper triangular factor in the QR decomposition of* $\tilde{X}$ *computed with orthogonal transformations.* This shows that the downdating algorithms based on CSNE produce accurate results for a considerably larger class of problems than the LINPACK or stabilized hyperbolic transformation-based downdating algorithms. This is a theoretical confirmation of the numerical results presented in [4, 8, 16, 20].

In all these downdating algorithms, it is assumed that the upper triangular factor R from the QR decomposition of X is available but not the orthogonal factor Q. When the full orthogonal factor Q from the QR decomposition is available, there is a downdating algorithm which is *mixed stable* [17]. An algorithm is mixed stable, if the computed solution is close to the exact solution of a slightly perturbed problem [3].

Some notations used in this paper are as follows. The singular values of X are denoted as $\sigma_i(X)$, $i = 1, 2, \ldots, n$, and assumed to be numbered in descending order. The matrix norms are the spectral norm $\|A\|_2 = \sigma_1(A)$, and the Frobenius norm $\|A\|_F = (\sum_{i,j} a_{ij}^2)^{1/2}$. The condition number of X is denoted as $\kappa(X) = \sigma_1(X)/\sigma_n(X)$. The matrix I_k denotes the identity matrix of order k.

2 Perturbation analysis.

2.1 A new perturbation result for downdating.

In the block downdating of a Cholesky decomposition, the given data are the upper triangular matrix R and the submatrix Z, which is the block of rows to be deleted. In this section, we present a new, first order perturbation result that shows how perturbations in R and Z propagate to a perturbation in the triangular matrix U in downdating. In the following lemma, we summarize the mathematical relations to be used in the analysis.

LEMMA 2.1. *Assume that X is defined as in (1.1), $X^T X$ has the Cholesky factor R as in (1.2), and the Cholesky decomposition*

$$U^T U = R^T R - Z^T Z$$

[1]The algorithms based on CSNE are more costly than the LINPACK and hyperbolic transformation-based algorithms. However, we can monitor the downdating conditioning, and use the more expensive CSNE approach for ill-conditioned downdates only. For details on the hybrid algorithms, see [4, 8].

[2]In this paper, an *outlier* is defined as a matrix element in Z that makes $\|X\|_2 \gg \|\tilde{X}\|_2$.

exists. Then Z and U are the (1,2) and (2,2) blocks, respectively, of the upper triangular factor in the QR decomposition of the matrix

$$\begin{pmatrix} \Gamma & 0 \\ Q_1 & R \end{pmatrix} \in \mathbf{R}^{(n+k)\times(n+k)},$$

where $Q_1 = R^{-T}Z^T$ and $\Gamma^T\Gamma = I_k - Q_1^TQ_1$. In other words, there is an orthogonal matrix $W \in \mathbf{R}^{(n+k)\times(n+k)}$, such that

$$\begin{pmatrix} \Gamma & 0 \\ Q_1 & R \end{pmatrix} = W \begin{pmatrix} I_k & Z \\ 0 & U \end{pmatrix}. \tag{2.1}$$

Proof: Since X and $\tilde{X}$ are assumed to have full rank, and

$$U^TU = R^T(I_n - Q_1Q_1^T)R, \tag{2.2}$$

$I_n - Q_1Q_1^T$ is positive definite with all its eigenvalues in the half open interval $(0,1]$. Since all the eigenvalues of $I_k - Q_1^TQ_1$ are eigenvalues of $I_n - Q_1Q_1^T$, $I_k - Q_1^TQ_1$ is also positive definite, and its Cholesky factor Γ exists. The columns of $\begin{pmatrix} \Gamma \\ Q_1 \end{pmatrix}$ are orthonormal, and therefore it has the QR decomposition

$$\begin{pmatrix} \Gamma \\ Q_1 \end{pmatrix} = W \begin{pmatrix} I_k \\ 0 \end{pmatrix},$$

for some orthogonal matrix W, and the first k columns of W will be $\begin{pmatrix} \Gamma \\ Q_1 \end{pmatrix}$. Moreover, W can be chosen so that

$$W^T \begin{pmatrix} 0 \\ R \end{pmatrix} = \begin{pmatrix} Y \\ T \end{pmatrix}$$

for some matrices $Y \in \mathbf{R}^{k\times n}$ and $T \in \mathbf{R}^{n\times n}$ which is upper triangular. It is easy to show that $Y = Z$ and $T = U$ since $Q_1^TR = Z$ and $R^TR = Z^TZ + T^TT$. ∎

From the proof we see that all singular values of Γ satisfy $\sigma_i(\Gamma) < 1$. This will be used throughout the paper.

Now, assume that R and Z have been perturbed by E_R and E_Z, respectively. In the following theorem we will give a condition under which the decomposition

$$(R + E_R)^T(R + E_R) - (Z + E_Z)^T(Z + E_Z) = (U + E_U)^T(U + E_U), \tag{2.3}$$

exists, and we will relate the relative perturbation in U to those of R and Z.

Theorem 2.2. *If the perturbations E_R and E_Z of R and Z, are small so that*

$$\epsilon := \max\{\epsilon_R, \epsilon_Z\} = \max\left\{\frac{\|E_R\|_F}{\|R\|_2}, \frac{\|E_Z\|_F}{\|Z\|_2}\right\} \ll \frac{\sigma_k^2(\Gamma)}{8\kappa(R)}, \tag{2.4}$$

where $\sigma_k(\Gamma)$ is the smallest singular value of Γ, then the perturbed decomposition (2.3) exists, and for a small ϵ the perturbation of the downdated matrix U can be estimated as

$$\frac{\|E_U\|_F}{\|U\|_2} \leq \beta \left(\frac{\|E_R\|_F}{\|R\|_2} + \frac{\|E_Z\|_F}{\|Z\|_2} \right) + O(\epsilon^2), \tag{2.5}$$

where

$$\beta = K_1 \frac{\|R\|_2}{\|U\|_2} \frac{\kappa(R)\,\|\Gamma\|_2}{\sigma_k^2(\Gamma)}, \tag{2.6}$$

and $K_1 < 11$.

PROOF: From the assumption (2.4) we easily get

$$\|R^{-1}E_R\|_F < 1. \tag{2.7}$$

Then, according to the perturbation theory for linear systems of equations [12, p. 79], the perturbed solution matrix

$$\bar{Q}_1 \equiv Q_1 + E_Q$$

for the system

$$(R + E_R)^T (Q_1 + E_Q) = (Z + E_Z)^T,$$

satisfies the bound

$$\frac{\|E_Q\|_F}{\|Q_1\|_2} \leq \kappa(R)(\epsilon_R + \epsilon_Z) + O(\epsilon^2). \tag{2.8}$$

From (2.4) and (2.8), we have

$$\|Q_1^T E_Q + E_Q^T Q_1 + E_Q^T E_Q\|_F \leq \sigma_k^2(\Gamma)/2. \tag{2.9}$$

Therefore, the Cholesky decomposition of $I - \bar{Q}_1^T \bar{Q}_1$ exists as

$$I - \bar{Q}_1^T \bar{Q}_1 = (\Gamma + E_\Gamma)^T (\Gamma + E_\Gamma)$$

and we have the bound

$$\frac{\|E_\Gamma\|_F}{\|\Gamma\|_2} \leq \frac{2\sqrt{2}}{\sigma_k^2(\Gamma)} \|E_Q\|_F + O(\|E_Q\|_F^2) = \frac{2\sqrt{2}}{\sigma_k^2(\Gamma)} \|E_Q\|_F + O(\epsilon^2), \tag{2.10}$$

see the appendix. This is a special case of [9, Corollary 3.3]. For ill-conditioned downdating problems this estimate is better than those presented in [25, (2.40) and (2.41)], cf. Table 2.1.

Since

$$\tau = \left\| \begin{pmatrix} E_\Gamma \\ E_Q \end{pmatrix} \right\|_2 \tag{2.11}$$

can be estimated from

$$\begin{aligned} \tau &\leq \|E_Q\|_F + \|E_\Gamma\|_F \\ &\leq \left(1 + \frac{2\sqrt{2}\|\Gamma\|_2}{\sigma_k^2(\Gamma)} \right) \kappa(R)(\epsilon_R + \epsilon_Z) + O(\epsilon^2), \end{aligned} \tag{2.12}$$

we see that (2.4) also gives $\tau < 1$. The columns of the unperturbed matrix $\begin{pmatrix} \Gamma \\ Q_1 \end{pmatrix}$ are orthonormal, and therefore its condition number is equal to 1. The QR decomposition of the perturbed matrix

$$\begin{pmatrix} \bar{\Gamma} \\ \bar{Q}_1 \end{pmatrix} \equiv \begin{pmatrix} \Gamma + E_\Gamma \\ Q_1 + E_Q \end{pmatrix}$$

satisfies [26, Theorem 1.6]

$$\begin{pmatrix} \bar{\Gamma} \\ \bar{Q}_1 \end{pmatrix} = (W + E_W) \left(\begin{pmatrix} I \\ 0 \end{pmatrix} + \begin{pmatrix} E_I \\ 0 \end{pmatrix} \right),$$

where

$$\|E_W\|_F \le K \left\| \begin{pmatrix} E_\Gamma \\ E_Q \end{pmatrix} \right\|_F \le K(\|E_\Gamma\|_F + \|E_Q\|_F), \quad K = \frac{1+\sqrt{2}}{1-\tau}. \tag{2.13}$$

Next, we use the relation (2.1)

$$W^T \begin{pmatrix} 0 \\ R \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}^T \begin{pmatrix} 0 \\ R \end{pmatrix} = \begin{pmatrix} Z \\ U \end{pmatrix}$$

where W is partitioned so that

$$W_{22}^T R = U.$$

When E_W is partitioned like W,

$$E_W = \begin{pmatrix} E_{W_{11}} & E_{W_{12}} \\ E_{W_{21}} & E_{W_{22}} \end{pmatrix},$$

the perturbed result of applying $(W_{22} + E_{W_{22}})^T$ to $R + E_R$ will satisfy

$$(W_{22}^T + E_{W_{22}}^T)(R + E_R) = U + E_U,$$

where the bound on E_U is given as

$$\begin{aligned} \|E_U\|_F &\le \|E_{W_{22}}^T R\|_F + \|W_{22}^T E_R\|_F + O(\epsilon^2) \\ &\le \|R\|_2 \|E_{W_{22}}\|_F + \|W_{22}\|_2 \|E_R\|_F + O(\epsilon^2) \\ &\le \|R\|_2 \|E_W\|_F + \|E_R\|_F + O(\epsilon^2), \end{aligned} \tag{2.14}$$

since $\|W_{22}\|_2 \le \|W\|_2 = 1$ and $\|AB\|_F \le \|A\|_2 \|B\|_F$.

Combining (2.8)-(2.14), we obtain a bound for E_U in terms of norms of R and Z, and their perturbations E_R and E_Z, as

$$\begin{aligned} \|E_U\|_F &\le \|R\|_2 \|E_W\|_F + \|E_R\|_F + O(\epsilon^2) \\ &\le K\|R\|_2 \left(\|E_Q\|_F + \|E_\Gamma\|_F\right) + \|E_R\|_F + O(\epsilon^2) \end{aligned}$$

$$
\begin{aligned}
&\leq K\|R\|_2\left(\|E_Q\|_F+\frac{2\sqrt{2}\,\|\Gamma\|_2}{\sigma_k^2(\Gamma)}\|E_Q\|_F\right)+\|E_R\|_F+O(\epsilon^2)\\
&\leq K\|R\|_2\kappa(R)\|Q_1\|_2\left(1+\frac{2\sqrt{2}\,\|\Gamma\|_2}{\sigma_k^2(\Gamma)}\right)(\epsilon_R+\epsilon_Z) \qquad (2.15)\\
&\quad+\|E_R\|_F+O(\epsilon^2).
\end{aligned}
$$

Since $K=(1+\sqrt{2})(1+\tau+\tau^2+\cdots)$, using (2.12) and the inequalities $\|Q_1\|_2\leq 1$ and $\sigma_k(\Gamma)\leq\|\Gamma\|_2\leq 1$, and dividing (2.15) by $\|U\|_2$, we have

$$
\frac{\|E_U\|_F}{\|U\|_2}\leq C_1\frac{\|R\|_2\kappa(R)\|\Gamma\|_2}{\|U\|_2\,\sigma_k^2(\Gamma)}(\epsilon_R+\epsilon_Z)+\frac{\|R\|_2}{\|U\|_2}\epsilon_R+O(\epsilon^2),
$$

where $C_1=(1+\sqrt{2})(1+2\sqrt{2})<9.25$. After an obvious simplification, where we use the fact that $\kappa(R)\|\Gamma\|_2/\sigma_k^2(\Gamma)\geq 1$, we finally get (2.5)–(2.6). ☐

In the error analysis in the following sections, we will not be concerned with the precise values of constants like K_1. This is because the main purpose of error analysis of floating point computations is not to obtain the sharpest possible error estimate (since in practice most estimates will be gross overestimates), but to exhibit the stability properties of different algorithms. Therefore, in order to make the bound simpler, we made a few coarse estimates leading to $K_1<11$ in the last part of the proof of Theorem 2.2.

In the new perturbation bound (2.5)–(2.6), we have one factor $\kappa(R)$ less than in our previous analysis [9]. In most cases this perturbation analysis gives a tighter bound than those in [9, 18], which is shown in the following subsection.

$$
\beta=K_1\frac{\|R\|_2}{\|U\|_2}\frac{\kappa(R)\,\|\Gamma\|_2}{\sigma_k^2(\Gamma)},
$$

as a *condition number for block downdating*[3]. When $k=1$ (only one row is downdated), then the downdating condition number β is simplified to

$$
\beta_1=K_1\frac{\|R\|_2}{\|U\|_2}\frac{\kappa(R)}{\sigma_k(\Gamma)}=K_1\frac{\|R\|_2}{\|U\|_2}\frac{\kappa(R)}{\sqrt{1-\|Q_1\|_2^2}}. \qquad (2.16)
$$

Note that when $\|R\|_2\approx\|U\|_2$, then (2.16) is essentially the square root of the downdating condition number given in [18].

If we can determined the ill-conditioning of a downdating problem *before* the actual downdating is performed, we can design a hybrid algorithm which uses the more accurate but expensive algorithm, only for ill-conditioned downdating problems [4, 8]. From (2.5)-(2.6) we see that a block downdating problem is ill-conditioned if one or more of the following hold:

(i) $\kappa(R)$ is large

[3]Note that in order to distinguish between this condition number and ordinary matrix condition numbers, we will always refer to it in conjunction with "downdating".

(*ii*) $\sigma_k(\Gamma)$ is small

(*iii*) $\|R\|_2 \gg \|U\|_2$ (this can happen when an outlier is removed).

Now it may seem that we need to compute U to be able to detect ill-conditioning due to the above condition *(iii)*. However, generalizing a statement in [25], we have

$$\|U\|_2 \geq \sigma_k(\Gamma)\|R\|_2, \tag{2.17}$$

(see the appendix), and from this inequality we see that (*iii*) implies (*ii*). Thus, the possible ill-conditioning of the downdating problem can be inferred from $\kappa(R)$ and $\sigma_k(\Gamma)$. In the downdating algorithms that are generalizations of the LINPACK single row downdating algorithm, these quantities can be estimated before the downdating is actually performed. Then the ill-conditioned downdates can be computed by more accurate methods such as the method based on CSNE. For such hybrid methods, see [4, 8]. We do not propose to use (2.17) to simplify the expression (2.6) for the downdating condition number, since this would give a much worse bound when $\|R\|_2 \approx \|U\|_2$ and $\sigma_k(\Gamma)$ is small. But there is no loss of accuracy in using conditions *(i)* and *(ii)* only, for deciding which algorithm to use since this results in more lenient use of the more accurate algorithm in a hybrid method.

2.2 Comparisons of perturbation bounds.

Condition numbers for the Cholesky block downdating problem are given in [9] and [25], and these results generalize that in [18]. The analysis in [23] is also based on a condition number for Cholesky downdating, even if it is not stated explicitly. As it is not immediately obvious which condition number gives the tighter bound, we first present a few simple test problems, and then give the corresponding downdating condition numbers.

EXAMPLE 2.1. Let $\omega \gg 1$, and

$$X = \begin{pmatrix} 0 & \omega \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \tilde{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

then

$$\begin{aligned} \|R\|_2 &\approx \kappa(R) \approx \omega, \qquad \|U\|_2 = 1, \qquad \kappa(U) = 1, \\ Q_1 &= \begin{pmatrix} 0 \\ \omega/\sqrt{\omega^2+1} \end{pmatrix}, \qquad \Gamma = \left(1 - \frac{\omega^2}{\omega^2+1}\right)^{1/2} \approx \frac{1}{\omega}. \end{aligned}$$

This is an example of downdating of an outlier. The downdated matrix is well-conditioned.

EXAMPLE 2.2. Let $\eta \ll 1$, and

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & \eta \end{pmatrix}, \qquad \tilde{X} = \begin{pmatrix} 1 & 0 \\ 0 & \eta \end{pmatrix}.$$

Then

$$\begin{aligned} \|R\|_2 &\approx 1, \qquad \kappa(R) \approx 1, \qquad \|U\|_2 \approx 1, \qquad \kappa(U) = 1/\eta, \\ Q_1 &= \begin{pmatrix} 0 \\ 1/\sqrt{1+\eta^2} \end{pmatrix}, \qquad \Gamma = \left(1 - \frac{1}{1+\eta^2}\right)^{1/2} \approx \eta. \end{aligned}$$

This is an example of downdating from a well-conditioned matrix to an ill-conditioned one.

EXAMPLE 2.3. Let $\eta \ll 1$, and

$$X = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & \eta & 0 \\ 0 & 0 & \sqrt{\eta} \end{pmatrix}, \qquad \tilde{X} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \eta & 0 \\ 0 & 0 & \sqrt{\eta} \end{pmatrix}.$$

Then

$$\begin{aligned} \|R\|_2 &\approx 1, \qquad \kappa(R) \approx 1/\sqrt{\eta}, \qquad \|U\|_2 \approx 1, \qquad \kappa(U) = 1/\eta, \\ Q_1 &= \begin{pmatrix} 0 \\ 1/\sqrt{1+\eta^2} \\ 0 \end{pmatrix}, \qquad \Gamma = \left(1 - \frac{1}{1+\eta^2}\right)^{1/2} \approx \eta. \end{aligned}$$

We downdate from a moderately ill-conditioned R to an ill-conditioned U.

EXAMPLE 2.4. We combine Examples 2.1 and 2.2, and let

$$X = \begin{pmatrix} 0 & \omega & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \eta \end{pmatrix}, \qquad \tilde{X} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \eta \end{pmatrix},$$

where $\omega = 1/\eta$. We have

$$\|R\|_2 \approx \omega, \qquad \kappa(R) \approx \omega = 1/\eta, \qquad \|U\|_2 = 1, \qquad \kappa(U) = 1/\eta,$$

and

$$Q_1 = \begin{pmatrix} 0 & 0 \\ \omega/\sqrt{\omega^2+1} & 0 \\ 0 & 1/\sqrt{1+\eta^2} \end{pmatrix}, \qquad \Gamma^T\Gamma = \begin{pmatrix} 1/(\omega^2+1) & 0 \\ 0 & \eta^2/(1+\eta^2) \end{pmatrix},$$

which gives

$$\sigma_2(\Gamma) \approx \eta, \qquad \|\Gamma\|_2 \approx \eta.$$

Thus, we downdate an outlier and obtain an ill-conditioned matrix as a result.

In Table 2.1 we summarize the different downdating condition numbers and their results applied to the above examples. The examples were constructed to demonstrate that there are cases when (2.6) gives a tighter bound than other downdating condition numbers. But they also demonstrate four representative

Table 2.1: Downdating condition numbers for four test problems. For simplicity we have omitted constants depending on the dimension of the matrices, etc.

	Condition Number	Example			
		2.1	2.2	2.3	2.4
Stewart[23]	$\|U^{-1}\|_F^2 \, \|R\|_F^2$	ω^2	$1/\eta^2$	$1/\eta^2$	$1/\eta^4$
Sun[25](2.40)	$\dfrac{\kappa(R)}{\sigma_k^3(\Gamma)}$	ω^4	$1/\eta^3$	$1/\eta^{3.5}$	$1/\eta^4$
Sun[25](2.41)	$\dfrac{\kappa(U)}{\sigma_k^2(\Gamma)}$	ω^2	$1/\eta^3$	$1/\eta^3$	$1/\eta^3$
E-P[9]	$\dfrac{\kappa^2(R)}{\sigma_k^2(\Gamma)}$	ω^4	$1/\eta^2$	$1/\eta^3$	$1/\eta^4$
E-P (2.6)	$\dfrac{\|R\|_2}{\|U\|_2} \dfrac{\kappa(R)\|\Gamma\|_2}{\sigma_k^2(\Gamma)}$	ω^3	$1/\eta$	$1/\eta^{1.5}$	$1/\eta^3$

ill-conditioned downdating problems. It is easy to modify Examples 2.2-2.4 by increasing the dimension of Z, so that another row is downdated and $\|\Gamma\|_2 \approx 1$. Then (2.6) is comparable in magnitude to the best among them all, if not the best itself. On the other hand, we consider it as an important virtue of the downdating condition number (2.6) that in the case when $k = 1$ it reduces to (2.16) which has one less $\sigma_k(\Gamma)$ term than other downdating condition numbers for $k = 1$.

It is remarkable that the simplest downdating condition number, i.e., the one given by Stewart [23], performed well for these examples. However, it requires the solution of the downdating problem before the downdating condition number can be computed.

3 Error analysis.

Error analyses of numerical algorithms are often involved, and in order to avoid unnecessary details, it is customary to use C to denote a constant times a low degree polynomial in the dimensions of the problem, without specifying the constant or the polynomial. Note that each occurrence of C may have different values. As a consequence, we will not differentiate between Euclidean and Frobenius norms, since they satisfy $\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2$ [12, p. 57]. The matrix norms appearing in this section denote either the Euclidean or Frobenius norm.

In the rest of this paper we let μ denote the unit round off error [12, p. 62] of the floating point system. We will use $\hat{U}$ (for instance) to denote the computed value of U in floating point arithmetic.

3.1 Algorithms based on LINPACK and stabilized hyperbolic transformations.

We first consider the block downdating algorithm [8] based on the LINPACK algorithm for a single row downdating [24]. It is shown in [23] that the computed value $\hat{U}$ of the downdated matrix U satisfies

$$V^T \begin{pmatrix} 0 \\ R \end{pmatrix} = \begin{pmatrix} Z + F_Z \\ \hat{U} + F_U \end{pmatrix},$$

for some orthogonal matrix V, and that the norm of the error matrices F_U and F_Z can be estimated as

$$\left\| \begin{pmatrix} F_Z \\ F_U \end{pmatrix} \right\| \le C\mu \|R\|. \tag{3.1}$$

The bound is also valid for Chambers' downdating algorithm [7] and downdating based on stabilized hyperbolic transformations[1, 5, 23]. This shows that the LINPACK and the stabilized hyperbolic transformation-based downdating algorithms are mixed stable when $\|U\|$ and $\|R\|$ are about the same. When $\|U\| \ll \|R\|$, a more general concept of stability called *relational stability* is applicable [23].

We can consider $\hat{U} + F_U$ as the exact result of downdating the exact R with slightly perturbed data Z. If the perturbation F_Z of Z is so small that the assumption of Theorem 2.2 is satisfied, then from (3.1) and (2.5) we obtain the following error estimate for $\hat{U} + F_U$:

$$\frac{\|\hat{U} + F_U - U\|}{\|U\|} \le C\mu \frac{\|R\|^2 \|\Gamma\|}{\|U\| \|Z\|} \frac{\kappa(R)}{\sigma_k^2(\Gamma)} + O(\mu^2).$$

Then using the triangular inequality, we get the *forward error estimate*

$$\frac{\|\hat{U} - U\|}{\|U\|} \le \frac{\|\hat{U} + F_U - U\| + \|F_U\|}{\|U\|} \le C\mu \frac{\|R\|^2 \|\Gamma\|}{\|U\| \|Z\|} \frac{\kappa(R)}{\sigma_k^2(\Gamma)} + O(\mu^2). \tag{3.2}$$

Similarly, using one of the downdating condition numbers of Sun[25], we get

$$\frac{\|\hat{U} - U\|}{\|U\|} \le C\mu \frac{\|R\|}{\|Z\|} \frac{\kappa(U)}{\sigma_k^2(\Gamma)} + O(\mu^2). \tag{3.3}$$

If $\kappa(R) < \kappa(U)$ and $\|R\| \approx \|U\|$, then the error bound (3.2) is better than (3.3).

If $\hat{U}$ is obtained as the upper triangular factor in the QR decomposition of $\tilde{X}$ computed with orthogonal transformations such as Householder transformations, then by combining perturbation theory for the QR decomposition [26] and error analysis of Householder transformations [15] (see also [14]), we have

$$\frac{\|\hat{U} - U\|}{\|U\|} \le C\mu\kappa(U) + O(\mu^2). \tag{3.4}$$

For problems where $\sigma_k(\Gamma)$ is small, the bounds (3.2) and (3.3) can be much larger than (3.4). However, if instead of using *only* R and Z in the computation of Γ (as is done in the LINPACK and stabilized hyperbolic transformation-based algorithms), we also use the data matrix X, then the factor $\sigma_k^2(\Gamma)$ in the denominator can be eliminated and we can achieve higher accuracy in the computed downdated factor $\hat{U}$. This result is shown in the following subsection.

3.2 Algorithm based on corrected seminormal equations.

The CSNE method can be used to solve least squares problems, when the orthogonal matrix Q in the QR decomposition of the data matrix X is not available, but X and the upper triangular matrix R are. In [2] stability properties of the CSNE method for solving least squares problems are analyzed, and it is shown that for moderately ill-conditioned least squares problems, the accuracy of the CSNE method is comparable to the accuracy of methods based on the orthogonal decomposition. In [4, 8, 20, 16] the CSNE method is used to obtain better accuracy for ill-conditioned downdating problems, and numerical experiments show that CSNE downdating often gives remarkably good results. We now present an error analysis that corroborates the experimental evidence.

The following theorem shows how Γ can be computed from the data matrix X using the CSNE method [8].

THEOREM 3.1. *Let V be the solution to*

$$\min_V \|E_k - XV\|, \qquad E_k = \begin{pmatrix} I_k \\ 0 \end{pmatrix} \in \mathbf{R}^{m\times k}, \tag{3.5}$$

where $X \in \mathbf{R}^{m\times n}$, $m \geq n$. Then the $(n+k)\times(n+k)$ upper triangular factor in the QR decomposition of $(X\ E_k)$ is

$$\begin{pmatrix} R & Q_1 \\ 0 & \Gamma \end{pmatrix}, \tag{3.6}$$

where the upper triangular matrix $\Gamma \in \mathbf{R}^{k\times k}$ is the same as the $k\times k$ upper triangular factor in the QR decomposition of $E_k - XV$. The matrix V satisfies the equation $Q_1 = RV$.

Applying the CSNE method to (3.5), we obtain the following algorithm for downdating[4], where more accurate Γ (3.6) can be obtained as the upper triangular factor of the matrix T [4, 8] using the original data matrix X. A derivation of the block algorithm can be found in [8].

[4]This algorithm differs from the block downdating algorithm based on CSNE presented in [8] in that here the correction δQ_1 is not used to update Q_1. The correction δQ_1 is used only to obtain an accurate approximation of T. The matrix Q_1 updated with δQ_1 should be more accurate, but, unfortunately, so far we have not been able to give an error estimate using the updated Q_1.

Downdating Algorithm Based on CSNE for improved Γ:

Given $X = \begin{pmatrix} Z \\ \tilde{X} \end{pmatrix} \in \mathbf{R}^{m\times n}$ with $Z \in \mathbf{R}^{k\times n}$, and $R \in \mathbf{R}^{n\times n}$ which is the upper triangular factor in the QR decomposition of X, this algorithm computes the upper triangular factor U for $\tilde{X}$ by downdating.

1. Compute Q_1, V and T from

$$\text{(a)}\ R^T Q_1 = Z^T, \qquad \text{(b)}\ RV = Q_1, \qquad \text{(c)}\ T := E_k - XV.$$

2. Update T:

$$\text{(a)}\ R^T \delta Q_1 = X^T T, \qquad \text{(b)}\ R\delta V = \delta Q_1, \qquad \text{(c)}\ T := T - X\delta V.$$

3. Compute the upper triangular factor Γ in the QR decomposition of T.
4. Determine an orthogonal matrix $W_0 \in \mathbf{R}^{(k+n)\times(k+n)}$, such that

$$\begin{pmatrix} I_k & Z \\ 0 & U \end{pmatrix} := W_0^T \begin{pmatrix} \Gamma & 0 \\ Q_1 & R \end{pmatrix}.$$

For better efficiency, we compute the QR decomposition of T and obtain the matrix Γ in Step 3 [8]. However, this step can be combined with Step 4 and we may use T in Step 4 instead of Γ. Since this simplifies the error analysis somewhat, we will assume that U is computed from

$$\begin{pmatrix} 0 & 0 \\ I_k & Z \\ 0 & U \end{pmatrix} := W^T \begin{pmatrix} T & 0 \\ Q_1 & R \end{pmatrix}, \tag{3.7}$$

where $W \in \mathbf{R}^{(m+n)\times(m+n)}$ and orthogonal. Note that the algorithm can be organized so that W is completely determined by T and Q_1, see [8].

We will now derive a forward error estimate for this downdating algorithm. To obtain Q_1 we solve a triangular system with multiple right hand sides. The computed solution $\hat{Q}_1$ satisfies (see e.g. [13])

$$\|\hat{Q}_1 - Q_1\| \le C\mu\kappa(R) + O(\mu^2). \tag{3.8}$$

Assume that $\mu\kappa^2(R) \ll 1$. The forward error in the residual, computed using CSNE, can be estimated [2, Theorem 3.2] as

$$\|\hat{T} - T\| \le C\,\mu^2\kappa^2(R)\|X\|\,\|V\| + C\mu\kappa(R)\|T\| + C\mu\|I_k\|.$$

Since

$$\|V\| \le \|R^{-1}\|\,\|Q_1\| \le C\|R^{-1}\|, \qquad \|T\| \le \|I_k\| \le C,$$

we get

$$\|\hat{T} - T\| \le C\mu\kappa(R). \tag{3.9}$$

Combining (3.8) and (3.9), we have the estimate

$$\left\| \begin{pmatrix} \hat{T} - T \\ \hat{Q}_1 - Q_1 \end{pmatrix} \right\| \leq C\mu\kappa(R) + O(\mu^2). \tag{3.10}$$

Consider the computation of (3.7). From the error analysis of Householder or Givens transformations (see e. g. [15, Chap. 15], [12, p. 205], [14, Chap. 14]), we obtain

$$\hat{W}^T \left(\begin{pmatrix} \hat{T} \\ \hat{Q}_1 \end{pmatrix} + H_1 \right) = \begin{pmatrix} 0 \\ \hat{S} \\ 0 \end{pmatrix}, \tag{3.11}$$

where $\hat{W}$ is orthogonal, $\hat{S}$ is the computed upper triangular matrix, and

$$\|H_1\| \leq C\mu \left\| \begin{pmatrix} \hat{T} \\ \hat{Q}_1 \end{pmatrix} \right\| + O(\mu^2).$$

We can estimate $\|H_1\|$ further, using (3.10), as

$$\begin{aligned} \|H_1\| &\leq C\mu \left(\left\| \begin{pmatrix} T \\ Q_1 \end{pmatrix} \right\| + \left\| \begin{pmatrix} \hat{T} - T \\ \hat{Q}_1 - Q_1 \end{pmatrix} \right\| \right) + O(\mu^2) \\ &\leq C\mu(C + C\mu\kappa(R)) + O(\mu^2) = C\mu + O(\mu^2). \end{aligned}$$

Again using (3.10) we can rewrite (3.11) in the form

$$\hat{W}^T \left[\begin{pmatrix} T \\ Q_1 \end{pmatrix} + H_2 \right] = \begin{pmatrix} 0 \\ \hat{S} \\ 0 \end{pmatrix}, \tag{3.12}$$

where

$$\|H_2\| \leq C\mu\kappa(R) + \|H_1\| + O(\mu^2) = C\mu\kappa(R) + O(\mu^2).$$

We can assume that $\|H_2\| \ll 1$ (note that we have assumed that $\mu\kappa^2(R) \ll 1$). To estimate how much $\hat{W}$ in (3.12) deviates from W defined in (3.7), we use the perturbation theory for the QR decomposition given in [26], and get

$$\|W - \hat{W}\| \leq C\|H_2\| / \left\| \begin{pmatrix} T \\ Q_1 \end{pmatrix} \right\| \leq C\mu\kappa(R) + O(\mu^2). \tag{3.13}$$

The same orthogonal transformation is used to compute U (cf. (3.7)), and (from [15, Chap. 15], [12, p. 205], [14, Chap. 14]), the computed $\hat{U}$ satisfies

$$\hat{W}^T \left[\begin{pmatrix} 0 \\ R \end{pmatrix} + H_3 \right] = \begin{pmatrix} 0 \\ \hat{Z} \\ \hat{U} \end{pmatrix}, \qquad \|H_3\| \leq C\mu\|R\| + O(\mu^2).$$

With conformal partitionings of $\hat{W}$ and W,

$$\hat{W} = \begin{pmatrix} \hat{W}_{11} & \hat{W}_{12} \\ \hat{W}_{21} & \hat{W}_{22} \end{pmatrix}, \qquad W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix},$$

we have

$$\hat{U} = \hat{W}_{22}^T R + (\hat{W}_{12}^T\, \hat{W}_{22}^T) H_3, \quad U = W_{22}^T R,$$

and we can estimate

$$\begin{aligned} \|\hat{U} - U\| &\leq \|(\hat{W}_{22}^T - W_{22})^T R\| + \|H_3\| \\ &\leq \|\hat{W} - W\|\,\|R\| + C\mu\|R\| \leq C\mu\kappa(R)\|R\| + O(\mu^2), \end{aligned}$$

where we have used (3.13). Finally,

$$\frac{\|\hat{U} - U\|}{\|U\|} \leq C\mu \frac{\|R\|}{\|U\|}\,\kappa(R) + O(\mu^2). \tag{3.14}$$

Before discussing this result, we remind the reader of the assumption that $\mu\kappa^2(R) \ll 1$. An important aspect of (3.14) is that *the accuracy no longer depends on the quantity* $\sigma_k^2(\Gamma)$ (as in (3.2) and (3.3)), which means that *CSNE downdating can give good results even if the downdating is ill-conditioned with small* $\sigma_k(\Gamma)$.

Comparing (3.14) and (3.4), we see that when $\kappa(U) \ll \kappa(R)$ and $\|U\| \ll \|R\|$, then computing U from CSNE downdating is less accurate than that from the QR decomposition by orthogonal transformations on $\tilde{X}$. This is when an outlier is downdated from R. However, there are cases when the error estimate for CSNE downdating is at least as good as that for Householder or Givens QR decomposition on $\tilde{X}$, e.g., when $\kappa(U) \gg \kappa(R)$ and $\|U\| \approx \|R\|$, which is when we downdate from well-conditioned R to an ill-conditioned U.

Acknowledgements.

The second author would like to thank for the support and hospitality of the Department of Mathematics, Linköping University, Linköping, Sweden, where a part of this work was carried out during her visit in August 1994. Comments from two anonymous referees helped to improve the presentation.

REFERENCES

1. S. T. Alexander, C.-T. Pan, and R. J. Plemmons, *Analysis of a recursive least squares hyperbolic rotation algorithm for signal processing*, Linear Algebra Appl., 98 (1988), pp. 3–40.
2. Å. Björck, *Stability analysis of the method of semi-normal equations for least squares problems*, Linear Algebra Appl., 88/89 (1987), pp. 31–48.
3. Å. Björck, *Error analysis of least squares algorithms*, in Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms, G. H. Golub and P. V. Dooren, eds., NATO ASI Series, Berlin, 1991, Springer-Verlag, pp. 41–73.
4. Å. Björck, H. Park, and L. Eldén, *Accurate downdating of least squares solutions*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 549–568.
5. A. W. Bojanczyk, R. P. Brent, P. van Dooren, and F. R. de Hoog, *A note on downdating the Cholesky factorization*, SIAM J. Sci. Stat. Comput., 8 (1987), pp. 210–221.

6. A. W. Bojanczyk and A. O. Steinhardt, *Stabilized hyperbolic Householder transformations*, IEEE Trans. Acoust., Speech, Signal Processing, ASSP-37 (1989), pp. 1286–1288.
7. J. Chambers, *Regression updating*, J. Amer. Stat. Assoc., 66 (1971), pp. 744–748.
8. L. Eldén and H. Park, *Block downdating of least squares solutions*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 1018–1034.
9. L. Eldén and H. Park, *Perturbation analysis for block downdating of a Cholesky decomposition*, Numer. Math., 68 (1994), pp. 457–467.
10. P. E. Gill, G. H. Golub, W. Murray, and M. A. Saunders, *Methods for modifying matrix factorizations*, Math. Comp., 28 (1974), pp. 505–535.
11. G. H. Golub and G. P. Styan, *Numerical computations for univariate linear models*, J. Stat. Comp. Simul., 2 (1973), pp. 253–272.
12. G. H. Golub and C. F. Van Loan, *Matrix Computations. 2nd ed.*, Johns Hopkins Press, Baltimore, MD, 1989.
13. N. Higham, *The accuracy of solutions to triangular systems*, SIAM J. Numer. Anal., 26 (1989), pp. 1252–1265.
14. N. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, 1996.
15. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
16. S. J. Olszanskyj, J. L. Lebak, and A. W. Bojanczyk, *Rank-k modification methods for recursive least squares problems*, Numerical Algorithms, 7 (1994), pp. 325–354.
17. C. C. Paige, *Error analysis of some techniques for updating orthogonal decompositions*, Math. Comp., 34 (1980), pp. 465–471.
18. C.-T. Pan, *A perturbation analysis of the problem of downdating a Cholesky factorization*, Linear Algebra Appl., 183 (1993), pp. 103–116.
19. C.-T. Pan and R. J. Plemmons, *Least squares modifications with inverse factorizations: parallel implications*, J. Comp. Appl. Math., 27 (1989), pp. 109–127.
20. H. Park and L. Eldén, *Downdating the rank-revealing URV decomposition*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 138–155.
21. C. Rader and A. Steinhardt, *Hyperbolic Householder transformations*, IEEE Trans. Acoust., Speech, Signal Processing, ASSP-34 (1986), pp. 1589–1602.
22. M. A. Saunders, *Large-scale linear programming using the Cholesky factorization*, Technical Report CS252, Computer Science Department, Stanford University, 1972.
23. G. Stewart, *On the stability of sequential updates and downdates*, Report TR-3238, Department of Computer Science, University of Maryland, College Park, MD 20742, March 1994.
24. G. W. Stewart, *The effects of rounding error on an algorithm for downdating a Cholesky factorization*, Journal of the Institute for Mathematics and Applications, 23 (1979), pp. 203–213.
25. J. Sun, *Perturbation analysis of the Cholesky downdating and QR updating problems*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 760–775.
26. J.-G. Sun, *Perturbation bounds for the Cholesky and QR factorizations*, BIT, 31 (1991), pp. 341–352.

A Appendix

A.1 Proof of (2.10)

The lemma given here is a special case and slight modification of Corollary 3.3 in [9].

LEMMA A.1. *Assume that* $\|E_Q\|_F = \epsilon_1$, *and consider the perturbed Cholesky downdating problem*

$$I - (Q_1 + E_Q)^T (Q_1 + E_Q) = \Gamma^T \Gamma + H,$$

where $H = -(Q_1^T E_Q + E_Q^T Q_1 + E_Q^T E_Q)$. *If* $\|H\|_F \leq \sigma_{min}^2(\Gamma)/2$, *then the factorization*

$$I - (Q_1 + E_Q)^T (Q_1 + E_Q) = (\Gamma + E_\Gamma)^T (\Gamma + E_\Gamma),$$

exists and

$$\frac{\|E_\Gamma\|_F}{\|\Gamma\|_2} \leq \frac{2\sqrt{2}}{\sigma_{min}^2(\Gamma)} \|E_Q\|_F + O(\epsilon_1^2).$$

Proof: Applying Theorem 1.1 in [26] to

$$(\Gamma + E_\Gamma)^T (\Gamma + E_\Gamma) = \Gamma^T \Gamma + H,$$

we get

$$\frac{\|E_\Gamma\|_F}{\|\Gamma\|_2} \leq \sqrt{2} \|(\Gamma^T \Gamma)^{-1}\|_2 \|H\|_F = \frac{\sqrt{2}}{\sigma_{min}^2(\Gamma)} \|H\|_F.$$

Using the inequalities $\|AB\|_F \leq \|A\|_2 \|B\|_F$, and $\|Q_1\|_2 \leq 1$, we can estimate

$$\|H\|_F \leq 2\|Q_1^T E_Q\|_F + O(\epsilon_1^2) \leq 2\|Q_1\|_2 \|E_Q\|_F + O(\epsilon_1^2) \leq 2\|E_Q\|_F + O(\epsilon_1^2),$$

which gives the lemma. □

A.2 Derivation of (2.17)

This is a generalization of a statement in [25] for the case when $k = 1$. From (2.2) we have

$$(I - Q_1 Q_1^T)^{-1} = RU^{-1} (RU^{-1})^T,$$

and

$$\|(I - Q_1 Q_1^T)^{-1}\|_2 = \|RU^{-1}\|_2^2.$$

Since $Q_1 \in \mathbf{R}^{n \times k}$, where $k \leq n$, the eigenvalues of $I - Q_1 Q_1^T$ that are different from 1, are equal to the eigenvalues of $I - Q_1^T Q_1 = \Gamma^T \Gamma$. The eigenvalues of $I - Q_1 Q_1^T$ lie in the interval $(0, 1]$ (see the proof of Lemma 2.1), and therefore we get

$$\|RU^{-1}\|_2^2 = \|(I - Q_1^T Q_1)^{-1}\|_2 = \frac{1}{\sigma_k^2(\Gamma)}.$$

Consequently,

$$\|U\|_2 = \|(RU^{-1})^{-1} R\|_2 \geq \|RU^{-1}\|_2^{-1} \|R\|_2 = \sigma_k(\Gamma) \|R\|_2.$$