BACKWARD ERROR ANALYSIS OF FACTORIZATION ALGORITHMS FOR SYMMETRIC AND SYMMETRIC TRIADIC MATRICES

HAW-REN FANG*

Abstract.
We consider the LBL^T factorization of a symmetric matrix where L is unit lower triangular and B is block diagonal with diagonal blocks of order 1 or 2. This is a generalization of the Cholesky factorization, and pivoting is incorporated for stability. However, the reliability of the Bunch-Kaufman pivoting strategy and Bunch's pivoting method for symmetric tridiagonal matrices could be questioned, because they may result in unbounded L. In this paper, we give a condition under which LBL^T factorization will run to completion in inexact arithmetic with inertia preserved. In addition, we present a new proof of the componentwise backward stability of the factorization using the inner product formulation, giving a slight improvement of the bounds in Higham's proofs, which relied on the outer product formulation and normwise analysis.

We also analyze the stability of rank estimation of symmetric indefinite matrices by LBL^T factorization incorporated with the Bunch-Parlett pivoting strategy, generalizing results of Higham for the symmetric semidefinite case.

We call a matrix triadic if it has no more than two non-zero off-diagonal elements in any column. A symmetric tridiagonal matrix is a special case. In this paper, we display the improvement in stability bounds when the matrix is triadic.

1. Introduction. A symmetric matrix $A \in R^{n\times n}$ can be factored into LBL^T, where L is unit lower triangular and B is block diagonal with each block of order 1 or 2. The process is described as follows. Assuming A is non-zero, there exists a permutation matrix Π such that

$$\Pi A \Pi^T = \begin{array}{c} \\ s \\ n-s \end{array} \begin{array}{c} \begin{array}{cc} s & n-s \end{array} \\ \begin{bmatrix} A_{11} & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix} \end{array},$$

where A_{11} is nonsingular, and $s = 1$ or 2 denoting that A_{11} is a 1×1 or 2×2 pivot. The decomposition is

$$\begin{bmatrix} A_{11} & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_s & 0 \\ A_{21}A_{11}^{-1} & I_{n-s} \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & \bar{A} \end{bmatrix} \begin{bmatrix} I_s & A_{11}^{-T}A_{21}^T \\ 0 & I_{n-s} \end{bmatrix},$$

where $\bar{A} = A_{22} - A_{21}A_{11}^{-1}A_{21}^T \in R^{(n-s)\times(n-s)}$ is the Schur complement. Iteratively applying the reduction to the Schur complement, we obtain the factorization in the form $PAP^T = LBL^T$, where P is a permutation matrix, L is unit lower triangular, and B is block diagonal with each block of order 1 or 2.

Choosing the permutation matrix Π and pivot size s at each step is called *diagonal pivoting*. In the literature, there are four major pivoting methods for general symmetric matrices: Bunch-Parlett [4], Bunch-Kaufman [3], bounded Bunch-Parlett and fast Bunch-Kaufman [1] pivoting strategies. They correspond to the complete pivoting, partial pivoting, and rook pivoting for LU factorization, respectively. In addition, there is a pivoting strategy specifically for symmetric tridiagonal matrices by Bunch [2].

*Department of Computer Science, University of Maryland, A.V. Williams Building, College Park, Maryland 20742, USA; hrfang@cs.umd.edu. This work was supported by the National Science Foundation under Grant CCR 02-04084.

The Bunch-Kaufman pivoting strategy and Bunch's method may lead to unbounded L. Therefore, the reliability could be questioned. Nevertheless, Higham proved the stability of Bunch-Kaufman pivoting strategy [9] and Bunch's method [10]. His proofs consist of componentwise backward error analysis using an outer product formulation and normwise analysis. In this paper, we present a new proof of componentwise backward stability using an inner product formulation. In addition, we give a sufficient condition under which the LBL^T factorization of a symmetric matrix is guaranteed to run to completion numerically and preserve inertia.

With complete pivoting, LU factorization and Cholesky factorization can be applied to rank estimation. The stability is analyzed in [8]. Given $A \in R^{n\times n}$ of rank r, the LU factorization needs $2(3n-2r)r^2/3$ flops, whereas Cholesky factorization needs $(3n-2r)r^2/3$ flops but requires symmetric positive semidefiniteness. To estimate the rank of a symmetric indefinite matrix, we can use LBL^T factorization with Bunch-Parlett pivoting strategy (complete pivoting), which needs $(3n-2r)r^2/3$ flops. The stability is analyzed in this paper.

A matrix A is called *triadic* if the number of non-zero off-diagonal elements in each column is bounded by 2. Tridiagonal matrices are a special case of these. The triadic structure is preserved in LBL^T factorization, so the sparsity is sustained [7]. In this paper, we show the improvement in backward error bounds for matrices with triadic structure.

This paper is organized as follows. Section 2 and Section 3 give the componentwise backward error analysis of LBL^T factorization and the application to solve symmetric linear systems, respectively. In Section 4 we discuss the stability using normwise analysis. Section 5 analyzes rank estimation for symmetric indefinite matrices by LBL^T factorization with Bunch-Parlett pivoting, as well as rank estimation for positive definite or diagonally dominant matrices by LDL^T factorization with complete pivoting. Section 6 gives the concluding remarks.

Throughout the paper, without loss of generality, we assume the required interchanges for any diagonal pivoting are done prior to the factorization, so that $A := PAP^T$, where P is the permutation matrix for pivoting. We denote the identity matrix of dimension 2 by I_2.

2. Componentwise Analysis.

The stability of Cholesky factorization in LL^T form, which requires a positive definite or semidefinite matrix, is well studied in [8] and [11, Chapter 10]. In this paper, we focus on LDL^T factorization and LBL^T factorization. The improvement of the stability because of the triadic structure is also discussed. We begin with basics for rounding error analysis.

2.1. Basics.

We use $fl(\cdot)$ to denote the computed value of a given expression, and follow the standard model

$$fl(x\ op\ y) = (x\ op\ y)(1+\delta),\ \text{for } |\delta| \le u \text{ and } op = +, -, \times, /,$$

where u is the unit roundoff. This model holds in most computers, including those using IEEE standard arithmetic. Lemma 2.1 gives the basic tool for rounding error analysis [11, Lemma 3.1].

LEMMA 2.1. *If $|\delta_i| \le u$ and $\sigma_i = \pm 1$ for $i = 1, \ldots, k$ then if $ku < 1$,*

$$\prod_{i=1}^{k}(1+\delta_i)^{\sigma_i} = 1+\theta_k,\ |\theta_k| \le \epsilon_k,$$

where

$$\epsilon_k = \frac{ku}{1-ku} \text{ for } k > 0.$$

The function ϵ_k defined in Lemma 2.1 has two useful properties[1]:

$$\epsilon_m + \epsilon_n + 2\epsilon_m\epsilon_n \le \epsilon_{m+n} \text{ for } m, n \ge 0,$$

and

$$c\epsilon_n \le \epsilon_{cn} \text{ for } c \ge 1.$$

Since we assume $ku < 1$ for all practical k,

$$\epsilon_k = ku + ku\epsilon_k = ku + O(u^2).$$

These properties are used frequently to derive inequalities in this paper.

2.2. LDL^T Factorization. We now investigate the stability of LDL^T factorization for symmetric matrices. The factorization is denoted by $A = LDL^T \in R^{n\times n}$, where $D = \text{diag}(d_1, d_2, \ldots, d_n)$ and the (i,j) entries of A and L are a_{ij} and l_{ij}, respectively. Note that $a_{ij} = a_{ji}$ and $l_{ij} = 0$ for all $1 \le i < j \le n$, and $l_{ii} = 1$ for $1 \le i \le n$. Algorithm 1 is computationally equivalent to the LDL^T factorization in inner product form.

Algorithm 1 LDL^T factorization in inner product form

```
for i = 1, ..., n do
  for j = 1, ..., i − 1 do
    (*) l_ij = (a_ij − Σ_{k=1}^{j−1} d_k l_ik l_jk)/d_j
  end for
  (**) d_i = a_ii − Σ_{k=1}^{i−1} d_k l_ik^2
end for
```

For a positive definite symmetric matrix $A = LL^T \in R^{n\times n}$,

$$|A - \hat{L}\hat{L}^T| \le \epsilon_{n+1}|\hat{L}||\hat{L}^T|,$$

where $\hat{L}$ is the computed version of L [11, Theorem 10.3]. Here and throughout this paper, we use a hat to denote computed quantities, and inequality and absolute value for matrices are defined elementwise. We begin with developing a bound for LDL^T factorization given in Theorem 2.3 with proof via Lemma 2.2. The result is extended to LBL^T factorization in Subsection 2.3.

LEMMA 2.2. *Let $y = (s - \sum_{k=1}^{n-1} a_k b_k c_k)/d$. No matter what the order of evaluation, the computed $\hat{y}$ satisfies*

$$\hat{y}d + \sum_{k=1}^{n-1} a_k b_k c_k = s + \Delta s,$$

[1] The two properties were listed in [9] and [11, Lemma 3.3], but with "$2\epsilon_m\epsilon_n$" replaced by "$\epsilon_m\epsilon_n$". Here we give a slightly tighter inequality.

where

$$|\Delta s| \le \epsilon_n(|\hat{y}d| + \sum_{k=1}^{n-1} |a_k b_k c_k|).$$

Proof. The proof is analogous to that of [11, Lemma 8.4]. Using Lemma 2.1, one may obtain

$$\hat{y}d(1+\theta_n^{(0)}) = s - \sum_{k=1}^{n-1} a_k b_k c_k (1+\theta_n^{(k)}),$$

where $|\theta_n^{(k)}| \le \epsilon_n$ for $k = 0, 1, \ldots, n-1$. The result follows immediately. □

THEOREM 2.3. *If the LDL^T factorization of a symmetric matrix $A \in R^{n\times n}$ runs to completion, then the computed $\hat{L}\hat{D}\hat{L}^T$ satisfies*

$$|A - \hat{L}\hat{D}\hat{L}^T| \le \epsilon_n |\hat{L}||\hat{D}||\hat{L}^T|.$$

Proof. By Lemma 2.2, no matter what the order of evaluation in $(*)$ and $(**)$ in Algorithm 1,

$$|a_{ij} - \sum_{k=1}^{j} \hat{d}_k \hat{l}_{ik} \hat{l}_{jk}| \le \epsilon_j \sum_{k=1}^{j} |\hat{d}_k \hat{l}_{ik} \hat{l}_{jk}| \tag{2.1}$$

for $1 \le j \le i \le n$, where we define $\hat{l}_{ii} = 1$ for $i = 1, \ldots, n$ to simplify the notation. The rest of the proof is by collecting all (2.1) into one matrix presentation. □

Theorem 2.3 shows that an LDL^T factorization is stable if $|\hat{L}||\hat{D}||\hat{L}^T|$ is suitably bounded. However, even with pivoting, the LDL^T factorization of a given symmetric matrix may not exist, and $|\hat{L}||\hat{D}||\hat{L}^T|$ could be catastrophically large. See Subsection 4.1 for a sufficient condition for the stability of LDL^T factorization.

The LDL^T factorization of a symmetric triadic matrix has L triadic, but the last row in L can be full [7]. Therefore, the bounding coefficient ϵ_n in Theorem 2.3 cannot be reduced with the triadic structure. Instead, we write the bound as

$$|A - \hat{L}\hat{D}\hat{L}^T| \le C \circ (|\hat{L}||\hat{D}||\hat{L}^T|). \tag{2.2}$$

Here and throughout this paper, $\circ$ denotes *Hadamard* (elementwise) product. Let $\|C\|_S = \sum_{i,j} |c_{ij}|$, where c_{ij} denotes the (i,j) entry of C. To show the improvement of stability because of the triadic structure, we compare $\|C\|_S$ for a general symmetric matrix with that for a triadic symmetric matrix.

By (2.1), we obtain $c_{ij} = c_{ji} = \epsilon_j$ for $1 \le j \le i \le n$. Therefore,

$$|A - \hat{L}\hat{D}\hat{L}^T| \le \begin{bmatrix} \epsilon_1 & \epsilon_1 & \epsilon_1 & \cdots & \epsilon_1 \\ \epsilon_1 & \epsilon_2 & \epsilon_2 & \cdots & \epsilon_2 \\ \epsilon_1 & \epsilon_2 & \epsilon_3 & \cdots & \epsilon_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \epsilon_1 & \epsilon_2 & \epsilon_3 & \cdots & \epsilon_n \end{bmatrix} \circ (|\hat{L}||\hat{D}||\hat{L}^T|). \tag{2.3}$$

Then

$$\begin{aligned}\|C\|_S &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} = \sum_{i=0}^{n-1}(2i+1)\epsilon_{n-i} \\ &\le \sum_{i=0}^{n-1}\epsilon_{(2i+1)(n-i)} \le \epsilon_{\sum_{i=0}^{n-1}(2i+1)(n-i)} \\ &= \epsilon_{\frac{1}{6}n(n+1)(2n+1)} = \frac{1}{6}(2n^3+3n^2+n)u + O(u^2). \end{aligned} \tag{2.4}$$

Note that (2.3) is approximately tight allowing any order of evaluation in $(*)$ and $(**)$ in Algorithm 1. Before investigating the case of A as a triadic symmetric matrix, we introduce the following lemma.

LEMMA 2.4. *For any triadic and lower triangular matrix $L \in R^{n\times n}$, LL^T has at most $7n-14$ non-zero elements for $n \ge 4$. The bound, $7n-14$, is attained by $L = Z^3 + Z + I$, where $Z \in R^{n\times n}$ is the shift-down matrix.*

Proof. Let $L = \tilde{L} + \tilde{D}$, where $\tilde{L}$ and $\tilde{D}$ are the off-diagonal and the diagonal part of L, respectively. Then $LL^T = (\tilde{L}+\tilde{D})(\tilde{L}+\tilde{D})^T = \tilde{L}\tilde{L}^T + \tilde{D}\tilde{L}^T + \tilde{L}\tilde{D} + \tilde{D}^2$, in which $\tilde{L}\tilde{D}$ contributes at most $2n-3$ non-zero elements in the lower triangular part. Now we inspect $\tilde{L}\tilde{L}^T$, in which each column of $\tilde{L}$ multiplying $\tilde{L}^T$ may contribute one off-diagonal element in the lower triangular part, except the last two columns of $\tilde{L}$. Therefore, $\tilde{L}\tilde{L}^T$ contributes at most $n-2$ non-zero off-diagonal elements in the lower triangular part. There are at most $(2n-3)+(n-2) = 3n-5$ off-diagonal terms in the lower triangular part. However, the two non-zero off-diagonal elements contributed by the third to last and fourth to last columns in $\tilde{L}$ must be in the bottom-right most 3×3 block of LL^T, which have collisions if the bottom-right most 3×3 block of $\tilde{L}$ is full. As a result, there are at most $(3n-5)-2 = 3n-7$ non-zero off-diagonal elements in the lower triangular part for $n \ge 4$. Along with the n diagonal elements, there are at most $2(3n-7)+n = 7n-14$ non-zero elements for $n \ge 4$. Note that $\tilde{D}^2$ and $\tilde{L}\tilde{L}^T$ can contribute n and $2n-3$ non-zero terms to the diagonal of LL^T, respectively. Overall, there are at most $2(3n-5)+(3n-3) = 9n-13$ non-zero terms for $n \ge 4$. ∎

If $A \in R^{n\times n}$ is symmetric triadic, then A has at most $3n$ non-zero elements. and so does its factorization LDL^T (or LL^T). However, because of rounding errors, the computed $\hat{L}\hat{D}\hat{L}^T$ (or $\hat{L}\hat{L}^T$) may have more non-zero elements than A. Nevertheless, by Lemma 2.4, the number of non-zero elements in $\hat{L}\hat{L}^T$ or $\hat{L}\hat{D}\hat{L}^T$ is bounded by $7n-14$ for $n \ge 4$.

For $1 \le j \le i \le n$, c_{ij} depends on the number of non-zero terms $\hat{d}_k\hat{l}_{ik}\hat{l}_{jk}$ in (2.1). By the proof of Lemma 2.4, there are at most $9n-13$ non-zero terms in $\hat{L}\hat{D}\hat{L}^T$ for $n \ge 4$. Therefore,

$$\|C\|_S = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} \le \epsilon_{9n-13} = 9nu + O(u^2). \tag{2.5}$$

Comparing (2.5) with (2.4), we see the improvement of componentwise backward error because of the triadic structure. Note that the analysis is independent of the order of evaluation in $(*)$ and $(**)$ in Algorithm 1.

2.3. LBL^T Factorization. Now we analyze the LBL^T factorization of a symmetric indefinite $A \in R^{n\times n}$. The factorization is denoted by $A = LBL^T \in R^{n\times n}$,

where $B = \begin{bmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_m \end{bmatrix}$ and $L = \begin{bmatrix} L_{11} & & & \\ L_{21} & L_{22} & & \\ \vdots & & \ddots & \\ L_{m1} & L_{m2} & \cdots & L_{mm} \end{bmatrix}$ for $i = 1, \ldots, m$. Each B_i is a 1×1 or 2×2 block, with $L_{ii} = 1$ or $L_{ii} = I_2$, respectively. The rest of L is partitioned accordingly. Algorithm 2 is computationally equivalent to the LBL^T factorization in inner product form.

Algorithm 2 LBL^T factorization in inner product form

for $i = 1, \ldots, m$ **do**
  for $j = 1, \ldots, i-1$ **do**
    $(*)$ $L_{ij} = (A_{ij} - \sum_{k=1}^{j-1} L_{ik}B_kL_{jk}^T)B_j^{-1}$
  end for
  $(**)$ $B_i = A_{ii} - \sum_{k=1}^{i-1} L_{ik}B_kL_{ik}^T$
end for

In Algorithm 2, each multiplication by B_j^{-1} in $(*)$ with $B_j \in R^{2\times2}$ can be computed by solving a 2×2 linear system, denoted by $Ey = z$. We assume the linear system is solved successfully with computed $\hat{y}$ satisfying

$$|\Delta E| \le \epsilon_c |E|, \text{ where } (E + \Delta E)\hat{y} = z \tag{2.6}$$

for some constant ϵ_c. In [9], Higham showed that with Bunch-Kaufman pivoting with pivoting argument $\alpha = \frac{1+\sqrt{17}}{8} \approx 0.64$, if the system is solved by GEPP, then $\epsilon_c = \epsilon_{12}$; if it is solved by explicit inverse of E with scaling (as implemented in both LAPACK and LINPACK), $\epsilon_c = \epsilon_{180}$. In a similar vein, the assumption (2.6) also holds with the other suggested pivoting argument $\alpha = 0.5$ to minimize the elements in magnitude in L and $\alpha = \frac{\sqrt{5}-1}{2} \approx 0.62$ for triadic matrices [7]. Since Bunch-Parlett, bounded Bunch-Parlett and fast Bunch-Kaufman pivoting strategies satisfy stronger conditions than the Bunch-Kaufman, condition (2.6) still holds. In [10], Higham showed that with Bunch's pivoting strategy for symmetric tridiagonal matrices [2], if a 2×2 linear system is solved by GEPP, then $\epsilon_c = \epsilon_{6\sqrt{5}}$. A constant ϵ_c for the 2×2 linear system solved by explicit inverse with scaling can also be obtained. We conclude that all pivoting strategies in the literature [1, 2, 3, 4, 7] satisfy condition (2.6).

LEMMA 2.5. *Let* $Y = (S - \sum_{k=1}^{m-1} A_kB_kC_k)E^{-1}$, *where* E *and each* B_k *are either* 1×1 *or* 2×2, *such that the matrix operations are well-defined. If* E *is a* 2×2 *matrix, we assume condition (2.6) holds. Then*

$$|\Delta S| \le \max\{\epsilon_c, \epsilon_{4m-3}\}(|S| + |\hat{Y}||E| + \sum_{k=1}^{m-1} |A_k||B_k||C_k|),$$

where

$$\hat{Y}E + \sum_{k=1}^{m-1} A_kB_kC_k = S + \Delta S.$$

If E *is an identity, then* $\max\{\epsilon_c, \epsilon_{4m-3}\}$ *can be replaced by* ϵ_{4m-3}, *since* $\epsilon_c = 0$.

Proof. Let

$$Z = S - \sum_{k=1}^{m-1} A_kB_kC_k \text{ and therefore } Y = ZE^{-1}.$$

If B_k is a 2×2 matrix, then each element in $A_kB_kC_k$ can be represented in the form $\sum_{i=1}^{4} a_ib_ic_i$. Therefore, each element in $\sum_{k=1}^{m-1} A_kB_kC_k$ sums at most $4(m-1)$ terms. By Lemma 2.2,

$$|\Delta Z| \le \epsilon_{4m-3}(|S| + \sum_{k=1}^{m-1} |A_k||B_k||C_k|), \text{ where } \hat{Z} = Z + \Delta Z. \tag{2.7}$$

If E is a 2×2 matrix, then applying (2.6) by substituting $y = Y(i:i,1:2)^T$ and $z = \hat{Z}(i:i,1:2)^T$, we obtain $\hat{Y}(i:i,1:2)(E+\Delta E_i) = \hat{Z}(i:i,1:2)$ with $|\Delta E_i| \le \epsilon_c|E|$ for $i = 1$ and $i = 2$ (if any). To simplify the notation, we write

$$\hat{Y}(E + \Delta E) = \hat{Z}, \text{ where } |\Delta E| \le \epsilon_c|E|. \tag{2.8}$$

In the rest of the proof, keep in mind that ΔE can be ΔE_1 or ΔE_2 for the first or second rows of $\hat{Y}$ and $\hat{Z}$, respectively. By (2.7) and (2.8),

$$\begin{aligned}
|\Delta S| &= |\hat{Y}E - (S - \sum_{k=1}^{m-1} A_kB_kC_k)| = |\hat{Y}E - \hat{Z} + \hat{Z} - Z| \\
&= |-\hat{Y}\Delta E + \Delta Z| \le |\hat{Y}||\Delta E| + |\Delta Z| \\
&\le \epsilon_c|\hat{Y}||E| + \epsilon_{4m-3}(|S| + \sum_{k=1}^{m-1} |A_k||B_k||C_k|) \\
&\le \max\{\epsilon_c, \epsilon_{4m-3}\}(|S| + |\hat{Y}||E| + \sum_{k=1}^{m-1} |A_k||B_k||C_k|).
\end{aligned}$$

If E is a 1×1 matrix, then we apply Lemma 2.2 and obtain $|\Delta S| \le \epsilon_{4m-3}(|\hat{Y}||E| + \sum_{k=1}^{m-1} |A_k||B_k||C_k|)$. □

THEOREM 2.6. *If the LBL^T factorization of a symmetric matrix $A \in R^{n\times n}$ runs to completion, then the computed $\hat{L}\hat{B}\hat{L}^T$ satisfies*

$$|A - \hat{L}\hat{B}\hat{L}^T| \le \max\{\epsilon_c, \epsilon_{4m-3}\}(|A| + |\hat{L}||\hat{B}||\hat{L}^T|),$$

where we assume condition (2.6) holds for all linear systems involving 2×2 pivots, and m is the number of blocks in B, $m \le n$.

Proof. Applying Lemma 2.5 to $(*)$ and $(**)$ in Algorithm 2, we obtain

$$|A_{ij} - \sum_{k=1}^{j} \hat{L}_{ik}\hat{B}_k\hat{L}_{jk}^T| \le \max\{\epsilon_c, \epsilon_{4j-3}\}(|A_{ij}| + \sum_{k=1}^{i} |\hat{L}_{ik}||\hat{B}_k||\hat{L}_{jk}^T|) \tag{2.9}$$

for $1 \le j \le i \le m$, where $\hat{L}_{ii} = 1$ or I_2, depending on whether B_i is 1×1 or 2×2 for $i = 1, \ldots, m$. The result is obtained by collecting all (2.9) into one matrix presentation. □

Similar to the coefficient ϵ_n in Theorem 2.3 for LDL^T factorization, the bounding coefficient $\max\{\epsilon_c, \epsilon_{4m-3}\}$ in Theorem 2.6 for LBL^T factorization can hardly be reduced because of the triadic structure. Instead, we bound $\|C\|_S$, where

$$|A - \hat{L}\hat{B}\hat{L}^T| \le C \circ (|A| + |\hat{L}||\hat{B}||\hat{L}^T|). \tag{2.10}$$

Each c_{ij} depends on the number of blocks before itself, which is at most $j-1$ for $1 \le j \le i \le n$. By (2.9), $c_{ij} \le \max\{\epsilon_c, \epsilon_{4j-3}\}$ for $1 \le j \le i \le n$. Therefore,

$$\|C\|_S \le \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} = n^2\epsilon_c + \sum_{i=0}^{n-1}(2i+1)\epsilon_{4(n-i)-3} + O(u^2)$$

$$\le n^2\epsilon_c + \epsilon_{\frac{1}{3}n(4n^2-3n+2)} + O(u^2) = \frac{1}{3}(4n^3 + 3(c-1)n^2 + 2n)u + O(u^2). \tag{2.11}$$

Because of the rounding errors, the computed $\hat{L}\hat{B}\hat{L}^T$ of a symmetric triadic matrix may have more non-zero elements than LBL^T. The triadic structure is preserved in the L [7]. By Lemma 2.4, the number of non-zero blocks in $\hat{L}\hat{B}\hat{L}^T$ is bounded by $7m-14$, and the proof shows that there are at most $9m-13$ block terms for $m \ge 4$, where m is the number of blocks in B. By (2.9), for $m \ge 4$,

$$\|C\|_S \le 4((7m-14)\epsilon_c + \epsilon_{4(9m-13)}) \le 4(7c+36)nu + O(u^2). \tag{2.12}$$

Therefore, $\hat{L}\hat{B}\hat{L}^T$ will still be sparse, but not necessarily triadic.

Comparing (2.12) with (2.11), we see the improvement of componentwise backward error because of the triadic structure. Note that the analysis is independent of the order of evaluation in $(*)$ and $(**)$ in Algorithm 2.

3. Solving Symmetric Linear Systems. In this section we use LBL^T factorization to solve a symmetric linear system $Ax = b$. After a possible permutation, which is omitted for notational convenience, we obtain $LBL^Tx = b$. Then we may solve three simplified systems, $Ly = b$ for y, $Bz = y$ for z, and $L^Tx = z$ for x.

If A is triadic, then each column of L has at most two off-diagonal elements and we can solve $Ly = b$ and $L^Tx = z$, traversing columns of L.

3.1. LDL^T Factorization. The computed solution $\hat{x}$ to an $n \times n$ symmetric positive definite system $Ax = b$ using LL^T factorization satisfies [11, Theorem 10.4],

$$(A + \Delta A)\hat{x} = b,\ |\Delta A| \le \epsilon_{3n+1}|\hat{L}||\hat{L}^T|.$$

Theorem 3.3 gives this bound for LDL^T factorization, with proof via Lemmas 3.1 and 3.2. The result is extended to LBL^T factorization in Subsection 3.2.

LEMMA 3.1. *Let $\hat{y}$ be the computed solution to the lower triangular system $Ly = b$ by forward substitution with any ordering of arithmetic operations, where $L \in R^{n\times n}$ is nonsingular. Then*

$$(L + \Delta L)\hat{y} = b,\ |\Delta L| \le \epsilon_n|L|.$$

If L is unit lower triangular, then there is no division so $|\Delta L| \le \epsilon_{n-1}|L|$. The bounds for upper triangular systems are the same.

Proof. Similar to the derivation leading to [11, Theorem 8.5]. □

LEMMA 3.2. *$\forall m, n, k > 0$ with $m + n + k < 1/u$,*

$$\epsilon_m + \epsilon_n + \epsilon_k + \epsilon_m\epsilon_n + \epsilon_n\epsilon_k + \epsilon_m\epsilon_k + \epsilon_m\epsilon_n\epsilon_k \le \epsilon_{m+n+k}.$$

Proof. Without loss of generality, let $k \le m$. Then

$$\begin{aligned} &\epsilon_m + \epsilon_n + \epsilon_k + \epsilon_m\epsilon_n + \epsilon_n\epsilon_k + \epsilon_m\epsilon_k + \epsilon_m\epsilon_n\epsilon_k \\ &\le (\epsilon_m + \epsilon_n + 2\epsilon_m\epsilon_n) + \epsilon_k + \epsilon_m\epsilon_k + \epsilon_m\epsilon_n\epsilon_k \\ &\le \epsilon_{m+n} + \epsilon_k + \epsilon_{m+n}\epsilon_k + \epsilon_{m+n}\epsilon_k \le \epsilon_{m+n+k}. \end{aligned}$$

□

THEOREM 3.3. *Suppose the LDL^T factorization of a symmetric matrix $A \in R^{n\times n}$ runs to completion and produces a computed solution $\hat{x}$ to $Ax = b$. Then*

$$(A + \Delta A)\hat{x} = b,\ |\Delta A| \le \epsilon_{3n-1}|\hat{L}||\hat{D}||\hat{L}^T|.$$

Proof. By Theorem 2.3, $A + \Delta A_1 = \hat{L}\hat{D}\hat{L}^T$ with $|\Delta A_1| \le \epsilon_n|\hat{L}||\hat{D}||\hat{L}^T|$. By Lemma 3.1,

$$(\hat{L} + \Delta L)\hat{y} = b,\ |\Delta L| \le \epsilon_{n-1}|\hat{L}|,$$

$$(\hat{D} + \Delta D)\hat{z} = \hat{y},\ |\Delta D| \le \epsilon_1|\hat{D}|, \tag{3.1}$$

$$(\hat{L}^T + \Delta R)\hat{x} = \hat{z},\ |\Delta R| \le \epsilon_{n-1}|\hat{L}^T|. \tag{3.2}$$

Then

$$\begin{aligned} b &= (\hat{L} + \Delta L)(\hat{D} + \Delta D)(\hat{L}^T + \Delta R)\hat{x} \\ &= (\hat{L}\hat{D}\hat{L}^T + \Delta L\hat{D}\hat{L}^T + \hat{L}\Delta D\hat{L}^T + \hat{L}\hat{D}\Delta R + \\ &\quad + \hat{L}\Delta D\Delta R + \Delta L\hat{D}\Delta R + \Delta L\Delta D\hat{L}^T + \Delta L\Delta D\Delta R)\hat{x}. \end{aligned}$$

Since $\hat{L}\hat{D}\hat{L}^T = A + \Delta A_1$,

$$\begin{aligned} |\Delta A| &= |\Delta A_1 + \Delta L\hat{D}\hat{L}^T + \hat{L}\Delta D\hat{L}^T + \hat{L}\hat{D}\Delta R + \\ &\quad + \hat{L}\Delta D\Delta R + \Delta L\hat{D}\Delta R + \Delta L\Delta D\hat{L}^T + \Delta L\Delta D\Delta R| \\ &\le |\Delta A_1| + |\Delta L||\hat{D}||\hat{L}^T| + |\hat{L}||\Delta D||\hat{L}^T| + |\hat{L}||\hat{D}||\Delta R| \\ &\quad + |\hat{L}||\Delta D||\Delta R| + |\Delta L||\hat{D}||\Delta R| + |\Delta L||\Delta D||\hat{L}^T| + |\Delta L||\Delta D||\Delta R| \\ &\le (\epsilon_n + \epsilon_{n-1} + \epsilon_1 + \epsilon_{n-1} + 2\epsilon_1\epsilon_{n-1} + \epsilon_{n-1}\epsilon_{n-1} + \epsilon_1\epsilon_{n-1}\epsilon_{n-1})|\hat{L}||\hat{D}||\hat{L}^T| \\ &\le (\epsilon_n + \epsilon_{2n-1})|\hat{L}||\hat{D}||\hat{L}^T| \le \epsilon_{3n-1}|\hat{L}||\hat{D}||\hat{L}^T|. \end{aligned}$$

The second to last inequality is derived by invoking Lemma 3.2. □

Now we define C by

$$(A + \Delta A)\hat{x} = b,\ |\Delta A| \le C \circ |\hat{L}||\hat{D}||\hat{L}^T|. \tag{3.3}$$

We follow the notation in the proof of Theorem 3.3. By (2.2) and (2.4),

$$|\Delta A_1| \le C_1 \circ (|\hat{L}||\hat{D}||\hat{L}^T|),\ \|C_1\|_S \le \frac{1}{6}(2n^3 + 3n^2 + n)u + O(u^2).$$

In the unit lower triangular system $\hat{L}y = b$, $\hat{L}(1\!:\!k, 1\!:\!k)y(1\!:\!k) = b(1\!:\!k)$ is a $k\times k$ unit lower triangular system for $k = 1, \ldots, n$. Repeatedly applying Lemma 3.1, we obtain

$$(\hat{L} + \Delta L)\hat{y} = b,\ |\Delta L| \le \mathrm{diag}(\epsilon_0, \epsilon_1, \ldots, \epsilon_{n-1})|\hat{L}|.$$

Therefore,

$$\begin{aligned} |\Delta A| &= |\Delta A_1 + \Delta L\hat{D}\hat{L}^T + \hat{L}\Delta D\hat{L}^T + \hat{L}\hat{D}\Delta R| + O(u^2) \\ &\le |\Delta A_1| + |\Delta L||\hat{D}||\hat{L}^T| + |\hat{L}||\Delta D||\hat{L}^T| + |\hat{L}||\hat{D}||\Delta R| + O(u^2) \\ &\le (C_1 + \mathrm{diag}(\epsilon_0, \epsilon_1, \ldots, \epsilon_{n-1})ee^T + (\epsilon_1 + \epsilon_{n-1})ee^T) \circ (|\hat{L}||\hat{D}||\hat{L}^T|) + O(u^2). \end{aligned}$$

Finally,

$$\begin{aligned}\|C\|_S &\le \|C_1\|_S + \|\mathrm{diag}(\epsilon_0, \epsilon_1, \ldots, \epsilon_{n-1})ee^T\|_S + \epsilon_n\|ee^T\|_S + O(u^2)\\ &= \frac{1}{6}(11n^3 + n)u + O(u^2). \qquad (3.4)\end{aligned}$$

Now suppose that $A \in R^{n\times n}$ is symmetric triadic. Following the notation in the proof of Theorem 3.3, the bound (3.2) can be reduced to be $|\Delta R| \le \epsilon_2|\hat{L}^T|$, and therefore the bound in Theorem 3.3 becomes $(A + \Delta A)\hat{x} = b$ with $|\Delta A| \le \epsilon_{2n+2}|\hat{L}||\hat{D}||\hat{L}^T|$. The bound on $\|C\|_S$ in (3.3) can be tightened with the triadic structure as follows.

$$\begin{aligned}|\Delta A| &\le |\Delta A_1| + |\Delta L||\hat{D}||\hat{L}^T| + |\hat{L}||\Delta D||\hat{L}^T| + |\hat{L}||\hat{D}||\Delta R| + O(u^2)\\ &\le |\Delta A_1| + \epsilon_{n-1}|\hat{L}||\hat{D}||\hat{L}^T| + \epsilon_1|\hat{L}||\hat{D}||\hat{L}^T| + \epsilon_2|\hat{L}||\hat{D}||\hat{L}^T| + O(u^2)\\ &\le C_1 \circ |\hat{L}||\hat{D}||\hat{L}^T| + \epsilon_{n+2}|\hat{L}||\hat{D}||\hat{L}^T| + O(u^2).\end{aligned}$$

By (2.5), $\|C_1\|_S \le 9nu + O(u^2)$. By Lemma 2.4, there are at most $7n - 14$ non-zero elements in $\hat{L}\hat{D}\hat{L}^T$ for $n \ge 4$. Therefore,

$$\|C\|_S \le \|C_1\|_S + (7n - 14)\epsilon_{n+2} + O(u^2) \le (7n^2 + 9n)u + O(u^2) \qquad (3.5)$$

Comparing (3.5) with (3.4), we see the improvement of componentwise backward error because of the triadic structure. The analysis is independent of the order of evaluation in $(*)$ and $(**)$ in Algorithm 1 and the order of substitution to solve the unit triangular systems.

3.2. LBL^T **Factorization.** Now we extend Theorem 3.3 for LDL^T factorization to Theorem 3.4 for LBL^T factorization.

THEOREM 3.4. *Suppose the LBL^T factorization of a symmetric matrix $A \in R^{n\times n}$ runs to completion and produces a computed solution $\hat{x}$ to $Ax = b$. Then*

$$(A + \Delta A)\hat{x} = b,\ |\Delta A| \le (\max\{\epsilon_c, \epsilon_{4n-3}\} + \epsilon_{2n+c-2})(|A| + |\hat{L}||\hat{B}||\hat{L}^T|),$$

where we assume condition (2.6) holds for all linear systems involving 2×2 pivots.

Proof. The proof is analogous to that of Theorem 3.3 but with two differences. First, since condition (2.6) holds, (3.1) is replaced by

$$(\hat{B} + \Delta B)\hat{z} = \hat{y},\ |\Delta B| \le \epsilon_c|\hat{B}|.$$

Second, we invoke Theorem 2.6 instead of Theorem 2.3 and obtain

$$|\Delta A_1| \le \max\{\epsilon_c, \epsilon_{4n-3}\}(|A| + |\hat{L}||\hat{B}||\hat{L}^T|).$$

The result is not difficult to see after a little thought. ∎

Theorem 2.6 and Theorem 3.4 coincide with a theorem by Higham [9, Theorem 4.1][11, Theorem 11.3] described as follows.

THEOREM 3.5 (Higham). *Suppose the LBL^T factorization of symmetric $A \in R^{n\times n}$ runs to completion and produces a computed solution to $Ax = b$. Without loss of generality, we assume all the interchanges are done with Bunch-Kaufman pivoting strategy (i.e., $A := PAP^T$, where P is the permutation matrix for pivoting). Let $\hat{L}\hat{B}\hat{L}^T$ be the computed factorization and $\hat{x}$ be the computed solution. Assuming condition (2.6) holds for all linear systems involving 2×2 pivots, then*

$$(A + \Delta A_1) = \hat{L}\hat{B}\hat{L}^T \text{ and } (A + \Delta A_2)\hat{x} = b,$$

where

$$|\Delta A_i| \le p(n)u(|A| + |\hat{L}||\hat{B}||\hat{L}^T|) + O(u^2),\ i = 1, 2,$$

with $p(n)$ a linear polynomial.

Three remarks are in order. First, Higham's proof is via the LBL^T factorization in outer product form [9], whereas our proof uses inner product form. Second, we give a precise bounding coefficient. Third, Lemma 3.2 eliminates the $O(u^2)$ terms. The result is also true for the Bunch-Parlett, fast Bunch-Parlett, and bounded Bunch-Kaufman pivoting strategies, because they satisfy stronger conditions than the Bunch-Kaufman.

We may also bound $\|C\|_S$, where C defined by

$$(A + \Delta A)\hat{x} = b,\ |\Delta A| \le C \circ (|A| + |\hat{L}||\hat{B}||\hat{L}^T|). \tag{3.6}$$

By (2.11), $\|C_1\|_S \le \frac{1}{3}(4n^3 + 3(c-1)n^2 + 2n)u + O(u^2)$, where $|A - \hat{L}\hat{B}\hat{L}^T| \le C_1 \circ (|A| + |\hat{L}||\hat{B}||\hat{L}^T|)$. Similar to (3.4), we find that

$$\begin{aligned} \|C\|_S &\le \|C_1\|_S + \|\mathrm{diag}(\epsilon_0, \epsilon_1, \ldots, \epsilon_{n-1})ee^T\|_S + (\epsilon_c + \epsilon_{n-1})\|ee^T\|_S + O(u^2) \\ &= (\frac{17}{6}n^3 + \frac{4c-5}{2}n^2 + 2n)u + O(u^2). \end{aligned} \tag{3.7}$$

Now suppose that $A \in R^{n\times n}$ is symmetric triadic. By (2.12), $\|C_1\|_S = 4(7c + 36)nu + O(u^2)$. By Lemma 2.4, there are at most $7n - 14$ non-zero blocks in $\hat{L}\hat{D}\hat{L}^T$ for $n \ge 4$. Each block has at most 4 elements. In the similar vein as (3.5),

$$\begin{aligned} \|C\|_S &\le \|C_1\|_S + 4(7n - 14)\epsilon_{n+c+1} + O(u^2) \\ &\le (28n^2 + 56cn + 144n)u + O(u^2). \end{aligned} \tag{3.8}$$

Comparing (3.8) with (3.7), we see the improvement of componentwise backward error because of the triadic structure. Note that the analysis is independent of the order of evaluation in $(*)$ and $(**)$ in Algorithm 2 and the order of substitution for solving each unit triangular system.

4. Normwise Analysis. In this section, we focus on bounding $\||L||D||L^T|\|$ and $\||L||B||L^T|\|$ in terms of $\|A\|$ to analyze the stability of LDL^T factorization and LBL^T factorization, respectively. We also give a sufficient condition for the success of LBL^T factorization with inertia preserved.

4.1. LDL^T Factorization. Theorem 2.3 and Theorem 3.3 imply that the LDL^T factorization of a symmetric matrix $A \in R^{n\times n}$ and its application to solve $Ax = b$ are stable, if $\||\hat{L}||\hat{D}||\hat{L}^T|\|$ is suitably bounded relative to $\|A\|$. We begin with bounding $\||L||D||L^T|\|$ instead of $\||\hat{L}||\hat{D}||\hat{L}^T|\|$ for simplicity.

If $A \in R^{n\times n}$ is symmetric positive definite, its LL^T factorization shares the properties with LDL^T factorization, including

$$\||L||D||L^T|\|_2 = \||LD^{\frac{1}{2}}||D^{\frac{1}{2}}L^T|\|_2 = \||LD^{\frac{1}{2}}|\|_2^2 \le n\|LD^{\frac{1}{2}}\|_2^2 = n\|A\|_2. \tag{4.1}$$

If $A \in R^{n\times n}$ is symmetric and diagonally dominant, its LDL^T factorization inherits the properties of LU factorization of a diagonally dominant matrix. Diagonal dominance guarantees that $|L|$ is diagonally dominant by columns, which implies $\||L^{-T}||L^T|\|_\infty \le 2n - 1$ [11, Lemma 8.8]. Therefore,

$$\begin{aligned} \||L||D||L^T|\|_\infty &= \||LD||L^T|\|_\infty = \||AL^{-T}||L^T|\|_\infty \\ &\le \||A|\|_\infty\||L^{-T}||L^T|\|_\infty \le (2n-1)\|A\|_\infty. \end{aligned}$$

The derivation is adapted from that for LU factorization of a diagonally dominant matrix.

We conclude that if A is positive definite or diagonally dominant, its LDL^T factorization and the application to solve linear system are stable even without pivoting. A weaker condition for the stability of LDL^T factorization can be obtained by [11, Theorem 9.5] described as follows.

THEOREM 4.1. *The LU factorization of $A \in R^{n\times n}$ satisfies*

$$\||L||U|\|_\infty \le (1 + 2(n^2-n)\rho_n)\|A\|_\infty,$$

where ρ_n is the growth factor (the largest element in magnitude in all Schur complements divided by the largest element in $|A|$). This statement is independent of the pivoting strategy applied, if any.

Proof. The proof is similar to that of a theorem by Wilkinson. See the discussion of [11, Theorem 9.5] for details. □

When A is symmetric, DL^T in its LDL^T factorization plays the role of U in the LU factorization. Therefore, Theorem 4.1 is applicable to bound $\||L||D||L^T|\|_\infty$, because $|L||D||L^T| = |L||DL^T| = |LU|$. As a result, the stability of LDL^T factorization is achieved as long as the growth factor ρ_n is modest. By Theorems 2.3 and 4.1,

$$\||\hat{L}||\hat{D}||\hat{L}^T|\|_\infty \le \frac{1 + 2(n^2-n)\rho_n}{1 - (1 + 2(n^2-n)\rho_n)\epsilon_n}\|A\|_\infty.$$

From this point of view, both positive definiteness and diagonal dominance guarantee the stability, because their growth factors are bounded by 1 and 2, respectively. Unfortunately, $|L||D||L^T|$ could be catastrophically large for general matrices. For example, if $A = \begin{bmatrix} \epsilon & 1 \\ 1 & 0 \end{bmatrix}$, then the corresponding $|L||D||L^T| = \begin{bmatrix} \epsilon & 1 \\ 1 & 2/\epsilon \end{bmatrix}$ is unbounded as $\epsilon \to 0$. This is an illustration of the well-known fact that LDL^T factorization is not generally stable.

4.2. LBL^T Factorization. Theorems 2.6 and 3.4 imply that the LBL^T factorization of a symmetric matrix $A \in R^{n\times n}$ and its application to solving $Ax = b$ are stable, if all linear 2×2 systems are solved with condition (2.6) satisfied and $\||\hat{L}||\hat{B}||\hat{L}^T|\|$ is suitably bounded relative to $\|A\|$. For simplicity, we begin with bounding $\||L||B||L^T|\|$ instead of $\||\hat{L}||\hat{B}||\hat{L}^T|\|$. In other words, our objective is to find a modest c_n such that

$$\||L||B||L^T|\| \le c_n\|LBL^T\| \tag{4.2}$$

in some proper norm. Using the ∞-norm, we obtain

$$\||L||B||L^T|\|_\infty \le \||L|\|_\infty\||B|\|_\infty\||L^T|\|_\infty = \|L\|_\infty\|B\|_\infty\|L^T\|_\infty.$$

Therefore, if $\|L\|_\infty\|B\|_\infty\|L^T\|_\infty$ is modest relative to $\|A\|_\infty$ and condition (2.6) holds, then the corresponding LBL^T factorization is normwise backward stable. The same statement can also be obtained from a suitable modification of error analysis for block LU factorization in [5]. For Bunch-Parlett, fast Bunch-Parlett, and bounded Bunch-Kaufman pivoting strategies, the element growth of B is well-controlled and the elements in L are bounded, so they are normwise backward stable methods.

Bunch-Kaufman pivoting strategy for symmetric matrices [3] and Bunch's pivoting strategy for symmetric tridiagonal matrices [2] may result in unbounded L, so we

cannot prove the stability by bounding $\|L\|_\infty\|B\|_\infty\|L^T\|_\infty$. Both strategies do have growth factors well-controlled. However, unlike LU factorization and LDL^T factorization, $\||L||B||L^T|\|$ cannot be bounded in terms of the growth factor and $\|A\|$. For example, without pivoting, we have

$$A = \begin{bmatrix} 1 & 1 & \\ 1 & 1+\epsilon^2 & -\epsilon \\ & -\epsilon & 2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \\ \frac{1}{\epsilon} & -\frac{1}{\epsilon} & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \\ 1 & 1+\epsilon^2 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \frac{1}{\epsilon} \\ & 1 & -\frac{1}{\epsilon} \\ & & 1 \end{bmatrix} = LBL^T.$$

The factorization has modest $\|A\|$ for small $\epsilon \neq 0$ and no element growth, but $\||L||B||L^T|\|$ is unbounded when $\epsilon \to 0$.

Higham [9] showed that using Bunch-Kaufman pivoting strategy with the pivoting argument $\alpha = \frac{1+\sqrt{17}}{8} \approx 0.64$, the LBL^T factorization of symmetric $A \in R^{n\times n}$ satisfies

$$\begin{aligned} \||L||B||L^T|\|_M &\leq \max\{\frac{1}{\alpha}, \frac{(3+\alpha^2)(3+\alpha)}{(1-\alpha^2)^2}\} n\rho_n \|A\|_M \\ &\approx 35.674 n\rho_n \|A\|_M < 36 n\rho_n \|A\|_M, \end{aligned} \tag{4.3}$$

where ρ_n is the growth factor and $\|\cdot\|_M$ is the largest magnitude element in the given matrix. These bounds also hold for the other suggested pivoting arguments $\alpha = \frac{\sqrt{5}-1}{2} \approx 0.62$ (for triadic matrices) and $\alpha = 0.5$ (to minimize the element bound on L) [7], as well as a variant by Sorensen and Van Loan [6, section 5.3.2]. The Bunch-Parlett, fast Bunch-Parlett and bounded Bunch-Kaufman pivoting strategies satisfy a stronger condition than the Bunch-Kaufman, so they also satisfy (4.3). By Theorem 2.6 and (4.3),

$$\||\hat{L}||\hat{B}||\hat{L}^T|\|_M \leq 36 n\rho_n \frac{1+\max\{\epsilon_c, \epsilon_{4n-3}\}}{1-36n\rho_n \max\{\epsilon_c, \epsilon_{4n-3}\}} \|A\|_M. \tag{4.4}$$

Higham showed that with Bunch's pivoting strategy, $\||L||B||L^T|\|_M \leq 42\|A\|_M$, where A is symmetric tridiagonal [10]. So a bound on $\||\hat{L}||\hat{B}||\hat{L}^T|\|_M$ in terms of $\|A\|_M$ can be obtained similarly.

In summary, all pivoting strategies for LBL^T factorization in the literature [1, 2, 3, 4, 7] are stable methods. Wilkinson [14] showed that the Cholesky factorization of a symmetric positive definite matrix $A \in R^{n\times n}$ is guaranteed to run to completion if $20n^{3/2}\kappa_2(A)u \leq 1$. We give a sufficient condition for the success of LBL^T factorization with inertia preserved in Theorem 4.3, with proof invoking a theorem by Weyl.

THEOREM 4.2 (Weyl [13, Corollary 4.9]). *Let A, B be two $n \times n$ Hermitian matrices and $\lambda_k(A)$, $\lambda_k(B)$, $\lambda_k(A+B)$ be the eigenvalues of A, B, and $A+B$ arranged in increasing order for $k = 1, \ldots, n$. Then for $k = 1, \ldots, n$, we have*

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B).$$

THEOREM 4.3. *With the Bunch-Parlett, Bunch-Kaufman, bounded Bunch-Parlett or fast Bunch-Kaufman pivoting strategy, the LBL^T factorization of symmetric $A \in R^{n\times n}$ succeeds with inertia preserved if $f(n)\kappa_n(A) < 1$ (i.e., A is not too ill conditioned), where*

$$f(n) = 36n^2\rho_n \max\{\epsilon_c, \epsilon_{4n-3}\} \frac{1+\max\{\epsilon_c, \epsilon_{4n-3}\}}{1-36n\rho_n \max\{\epsilon_c, \epsilon_{4n-3}\}} = 144n^3\rho_n u + O(u^2).$$

Proof. The proof is by finite induction. Consider Algorithm 2. Let $A_k = A(1:k, 1:k)$, $\Delta A_k = \Delta A(1:k, 1:k)$ and $\hat{A}_k = A_k + \Delta A_k$, where $A + \Delta A = \hat{L}\hat{B}\hat{L}^T$. The process of LBL^T factorization is to iteratively factor A_k, increasing k from 1 to n. Obviously, the first stage succeeds. Suppose the factorization of A_k is successfully completed with inertia preserved (i.e., the inertia of $\hat{A}_k$ is the same as that of A_k). Let $s = 1$ or 2 denote whether the next pivot is 1×1 or 2×2. Since the inertia of $\hat{A}_k$ is preserved, all the pivots in $\hat{A}_k$ are full rank, so the factorization of A_{k+s} succeeds (i.e., with no division by zero). By Theorem 2.6, (4.3) and (4.4), the componentwise backward error satisfies

$$\|\Delta A_k\|_2 \le 36k^2\rho_k \max\{\epsilon_c, \epsilon_{4k-3}\}\frac{1+\max\{\epsilon_c, \epsilon_{4k-3}\}}{1-36k\rho_k\max\{\epsilon_c, \epsilon_{4k-3}\}}\|A_k\|_2 =: f(k)\|A_k\|_2$$

for all possible $1 \le k \le n$. Note that $f(k) = 144k^3\rho_k u + O(u^2)$. Let

$$\lambda_*(A_k) := \min_{1\le i\le k} |\lambda_i(A_k)|.$$

Assume $f(n)\kappa_2(A_n) < 1$. By Theorem 4.2, if $\lambda_i(A_k) > 0$,

$$\begin{aligned}\lambda_i(A_k + \Delta A_k) &\ge \lambda_i(A_k) - \|\Delta A_k\|_2 \ge \lambda_*(A_k) - f(k)\|A_k\|_2 \\ &\ge \lambda_*(A_k)(1 - f(k)\kappa_2(A_k)) \ge \lambda_*(A_n)(1 - f(n)\kappa_2(A_n)) > 0.\end{aligned}$$

Similarly, if $\lambda_i(A_k) < 0$,

$$\lambda_i(A_k + \Delta A_k) \le \lambda_i(A_k) + \|\Delta A_k\|_2 \le -\lambda_*(A_k) + f(k)\|A_k\|_2 < 0.$$

Therefore, $\lambda_i(A_k + \Delta A_k)$ and $\lambda_i(A_k)$ have the same sign for $i = 1, \ldots, k$. So the pivoting guarantees that the inertia of $\hat{A}_k$ is preserved. By induction, the factorization is guaranteed running to completion with inertia preserved. ∎

5. Rank Estimation. Cholesky factorization for symmetric positive semidefinite matrices with complete pivoting can be used for rank estimation. The stability is well studied in [8]. In this section, we discuss the stability of Bunch-Parlett pivoting strategy applied to the rank estimation of symmetric indefinite matrices.

5.1. LDL^T Factorization. Given a symmetric matrix $A \in R^{n\times n}$ of rank $r < n$, we assume the necessary interchanges are done so that $A(1:r, 1:r)$ is nonsingular. If A is positive semi-definite or diagonally dominant, such a permutation exists. The LDL^T factorization has $L = \begin{bmatrix} L_{11} \\ L_{21} \end{bmatrix}$, where $L_{11} \in R^{n\times n}$ is unit lower triangular, $L_{21} \in R^{(n-r)\times r}$, and $D \in R^{r\times r}$ is diagonal.The factorization is computationally equivalent to Algorithm 3.

Algorithm 3 LDL^T factorization of a symmetric matrix of rank $r < n$

```
for j = 1, ..., r do
  (*) d_j = a_jj − Σ_{k=1}^{j−1} d_k l_jk^2
  for i = j + 1, ..., n do
    (**) l_ij = (a_ij − Σ_{k=1}^{j−1} d_k l_ik l_jk)/d_j
  end for
end for
```

Assuming the factorization in Algorithm 3 runs to completion, we define the backward error ΔA by

$$A + \Delta A = \hat{L}\hat{D}\hat{L}^T + \hat{A}^{(r+1)},$$

where

$$\hat{A}^{(r+1)} = \begin{array}{c} \\ r \\ n-r \end{array} \begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \begin{bmatrix} 0 & 0 \\ 0 & \hat{S}_{r+1} \end{bmatrix} \end{array},$$

with $\hat{S}_{r+1}$ the computed Schur complement. Denote the (i,j) entry of $\hat{A}^{(r+1)}$ by $\hat{a}_{ij}^{(r+1)}$. To simplify the notation, we define $\hat{l}_{jj} = 1$ for $j = 1, \ldots, r$. By Lemma 2.2,

$$|a_{ij} - \sum_{k=1}^{j} \hat{d}_k \hat{l}_{jk} \hat{l}_{ik}| \le \epsilon_j \sum_{k=1}^{j} |\hat{d}_k \hat{l}_{ik} \hat{l}_{jk}| \tag{5.1}$$

for $j = 1, \ldots, r$ and $i = j, \ldots, n$.

$$|a_{ij} - \hat{a}_{ij}^{(r+1)} - \sum_{k=1}^{r} \hat{d}_k \hat{l}_{ik} \hat{l}_{jk}| \le \epsilon_{r+1}(|\hat{a}_{ij}^{(r+1)}| + \sum_{k=1}^{r} |\hat{d}_k \hat{l}_{ik} \hat{l}_{jk}|) \tag{5.2}$$

for $j = r+1, \ldots, n$ and $i = j, \ldots, n$.

Collecting all (5.1)–(5.2) into one matrix representation, we obtain

$$|\Delta A| \le C \circ (|\hat{L}||\hat{D}||\hat{L}^T| + |\hat{A}^{(r+1)}|), \tag{5.3}$$

where $c_{ij} = c_{ji} = \epsilon_{\min\{j,r+1\}}$ for $1 \le j \le i \le n$, or simply

$$|\Delta A| \le \epsilon_{r+1}(|\hat{L}||\hat{D}||\hat{L}^T| + |\hat{A}^{(r+1)}|). \tag{5.4}$$

Applying (2.4) to bound $\|C(1\!:\!r, 1\!:\!r)\|_S$,

$$\begin{aligned} \|C\|_S = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} &= \sum_{i=1}^{r}\sum_{i=1}^{r} c_{ij} + 2\sum_{i=r+1}^{n}\sum_{j=1}^{r} c_{ij} + \sum_{i=r+1}^{n}\sum_{j=r+1}^{n} c_{ij} \\ &\le \frac{1}{6}(2r^3 + 3r^2 + r)u + 2(n-r)\sum_{j=1}^{r}\epsilon_j + (n-r)^2\epsilon_{r+1} + O(u^2) \\ &\le (n(n-r)(r+1) + \frac{1}{3}r^3 + \frac{1}{2}r^2 + \frac{1}{6}r)u + O(u^2) \end{aligned} \tag{5.5}$$

LEMMA 5.1. *For any triadic matrix $T \in R^{n\times r}$, TT^T has at most $6r$ non-zero off-diagonal elements/terms and $\min\{n, 3r\}$ diagonal elements/terms. The bounds, $6r$ and $\max\{n, 3r\}$, are attained with $L = (Z^3 + Z + I)(1\!:\!n, 1\!:\!r)$ for $n - r \ge 3$, where $Z \in R^{n\times n}$ is the shift-down matrix.*

Proof. The proof is analogous to that of Lemma 2.4 and omitted here. □

Suppose $A \in R^{n\times n}$ is symmetric triadic of rank $r < n$ and has an LDL^T factorization. By Lemma 5.1, there are at most $9r$ terms in $\hat{L}\hat{D}\hat{L}^T$. By (5.1)–(5.2),

$$\|C\|_S = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} \le \epsilon_{2\cdot 9r} = 18ru + O(u^2). \tag{5.6}$$

Comparing (5.6) with (5.5), we see the improvement of componentwise backward error because of the triadic structure. Note that the analysis is independent of the order of evaluation in $(*)$ and $(**)$ in Algorithm 3.

5.2. LBL^T **Factorization.** Now we investigate the LBL^T factorization of symmetric $A \in R^{n\times n}$ of rank $r < n$. Assume that the necessary interchanges are done so that $A(1:r, 1:r)$ has rank r, as they would be with Bunch-Parlett pivoting. Denote the factorization by $A = LBL^T$, where $B = \mathrm{diag}(B_1, B_2, \ldots, B_{m-1}) \in R^{r\times r}$, and

$$L = \begin{bmatrix} L_{11} & & & \\ L_{21} & L_{22} & & \\ \vdots & & \ddots & \\ L_{m-1,1} & L_{m-1,2} & \cdots & L_{m-1,m-1} \\ L_{m1} & L_{m2} & \cdots & L_{m,m-1} \end{bmatrix} \in R^{n\times r}.$$

Each B_i is either 1×1 or 2×2 for $i = 1, \ldots, m-1$, with L partitioned accordingly. Each L_{mj} has $n-r$ rows for $j = 1, \ldots, m$. The factorization is computationally equivalent to Algorithm 4.

Algorithm 4 LBL^T factorization of a matrix of rank $r < n$

```
for j = 1, ..., m − 1 do
  (*) B_j = A_jj − Σ_{k=1}^{j−1} L_jk B_k L_jk^T
  for i = j + 1, ..., m do
    (**) L_ij = (A_ij − Σ_{k=1}^{j−1} L_ik B_k L_jk^T) B_j^{−1}
  end for
end for
```

Assuming the factorization in Algorithm 4 runs to completion, the backward error ΔA of LBL^T factorization is defined by

$$A + \Delta A = \hat{L}\hat{B}\hat{L}^T + \hat{A}^{(r+1)}, \tag{5.7}$$

where

$$\hat{A}^{(r+1)} = \begin{array}{c} \\ r \\ n-r \end{array} \begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \begin{bmatrix} 0 & 0 \\ 0 & \hat{S}_{r+1} \end{bmatrix} \end{array},$$

with $\hat{S}_{r+1}$ the computed Schur complement. To simplify the notation, we let $\hat{L}_{ii} = 1$ or I_2 depending on whether B_i is 1×1 or 2×2 for $i = 1, \ldots, m-1$. We assume condition (2.6) holds. By Lemma 2.5,

$$|A_{ij} - \sum_{k=1}^{j} \hat{L}_{ik}\hat{B}_k\hat{L}_{jk}^T| \le \max\{\epsilon_c, \epsilon_{4j-3}\}(|A_{ij}| + \sum_{k=1}^{j} |\hat{L}_{ik}||\hat{B}_k||\hat{L}_{jk}^T|), \tag{5.8}$$

for $j = 1, \ldots, m-1$ and $i = j, \ldots, m$. Note that though A_{mm} can be larger than 2×2, the error analysis in Lemma 2.5 for $\hat{S}_{r+1}$ is still valid. Therefore,

$$|A_{mm} - \hat{S}_{r+1} + \sum_{k=1}^{m-1} \hat{L}_{mk}\hat{B}_k\hat{L}_{mk}^T| \le \epsilon_{4m-3}(|A_{mm}| + |\hat{S}_{r+1}| + \sum_{k=1}^{m-1} |\hat{L}_{mk}||\hat{B}_k||\hat{L}_{mk}^T|). \tag{5.9}$$

Collecting all (5.8) and (5.9) into one matrix representation, we obtain

$$|\Delta A| \le C \circ (|A| + |\hat{L}||\hat{B}||\hat{L}^T| + |\hat{A}^{(r+1)}|), \tag{5.10}$$

where the elements in (i, j) block entry of C is $\max\{\epsilon_c, \epsilon_{4j-3}\}$ for $1 \le j \le i \le m$. Since $m \le r + 1$, we may simply write

$$|\Delta A| \le \max\{\epsilon_c, \epsilon_{4r+1}\}(|A| + |\hat{L}||\hat{B}||\hat{L}^T| + |\hat{A}^{(r+1)}|). \tag{5.11}$$

Applying (2.11) to bound $\|C(1:r,1:r)\|_S$,

$$\begin{aligned}\|C\|_S &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} = \sum_{i=1}^{r}\sum_{j=1}^{r} c_{ij} + 2\sum_{i=r+1}^{n}\sum_{j=1}^{r} c_{ij} + \sum_{i=r+1}^{n}\sum_{j=r+1}^{n} c_{ij} \\ &\le cn^2u + \frac{1}{3}(4r^3-3r^2+2r)u + 2(n-r)\sum_{j=1}^{r}\epsilon_{4j-3} + (n-r)^2\epsilon_{4r+1} + O(u^2) \\ &= (cn^2 + (4nr+n-3r)(n-r) + \frac{4}{3}r^3 - r^2 + \frac{2}{3}r)u + O(u^2). \qquad (5.12)\end{aligned}$$

By Lemma 5.1, if A is triadic of rank r, there are at most $9r$ block terms in the LBL^T factorization. Each has at most 4 elements. By (5.8) and (5.9),

$$\|C\|_S \le 4(9r\epsilon_c + \epsilon_{9(4r+1)}) \le 36(cr+4r+1)u + O(u^2). \qquad (5.13)$$

Comparing (5.13) with (5.12), we see the improvement of componentwise backward error because of the triadic structure. Note that the analysis is independent of the order of evaluation in $(*)$ and $(**)$ in Algorithm 4.

5.3. Normwise Analysis. In this subsection we bound $\|A-\hat{L}\hat{B}\hat{L}^T\|_F$ for analyzing the stability of Bunch-Parlett pivoting strategy applied to rank estimation for symmetric indefinite matrices. We also bound $\|A-\hat{L}\hat{D}\hat{L}^T\|_2$ and $\|A-\hat{L}\hat{D}\hat{L}^T\|_\infty$ for positive semidefinite matrices and diagonal dominance matrices, respectively.

THEOREM 5.2. *With Bunch-Parlett pivoting on a symmetric indefinite matrix A,*

$$\|A-\hat{L}\hat{B}\hat{L}^T\|_F \le \max\{c,4r+1\}(\tau(A)+1)((\|W\|_F+1)^2+1)u\|A\|_F + O(u^2),$$

where $W = A(1:r,1:r)^{-1}A(1:r,r+1:n)$, $\tau(A) = \||\hat{L}||\hat{B}||\hat{L}^T|\|_F/\|\hat{L}\hat{B}\hat{L}^T\|_F$ and c is from condition (2.6).

Proof. Since (2.6) holds, so does (5.11). The growth factor and $\tau(A)$ are well-controlled, so that $\Delta A = O(u)$. Note that $\hat{L}\hat{B}\hat{L}^T$ is the partial LBL^T factorization of $A+\Delta A$ with $\hat{A}^{(r+1)}$ the Schur complement. The perturbation theory in [7] gives

$$\|\hat{A}^{(r+1)}\|_F \le (\|W\|_F+1)^2\|\Delta A\|_F + O(u^2), \qquad (5.14)$$

where $W = A(1:r,1:r)^{-1}A(1:r,r+1:n)$. By (5.7), $\hat{L}\hat{B}\hat{L}^T = A + O(u)$. Therefore,

$$\begin{aligned}\|\Delta A\|_F &\le \max\{\epsilon_c,\epsilon_{4r+1}\}(\||\hat{L}||\hat{B}||\hat{L}^T|\|_F + \||A|\|_F + \||\hat{A}^{(r+1)}|\|_F) \\ &\le \max\{c,4r+1\}(\tau(A)+1)u\|A\|_F + O(u^2). \qquad (5.15)\end{aligned}$$

Substituting (5.15) into (5.14), we obtain

$$\|\hat{A}^{(r+1)}\|_F \le \max\{c,4r+1\}(\tau(A)+1)(\|W\|_F+1)^2u\|A\|_F + O(u^2). \qquad (5.16)$$

The result is concluded from (5.7), (5.15) and (5.16). ∎

Now we consider the Bunch-Parlett pivoting strategy incorporated into the LBL^T factorization. By Theorem 5.2, the bound on $\|A-\hat{L}\hat{B}\hat{L}^T\|_F/\|A\|_F$ is governed by $\|W\|_F$ and $\tau(A)$. For any general singular symmetric $A \in R^{n\times n}$,

$$\|W\|_{2,F} \le \sqrt{\frac{\gamma}{\gamma+2}(n-r)((1+\gamma)^{2r}-1)}, \qquad (5.17)$$

where $\gamma = \max\{\frac{1}{\alpha}, \frac{1}{1-\alpha}\}$ is the element bound of L [7]. With the suggested pivoting argument $\alpha = \frac{1+\sqrt{17}}{8} \approx 0.64$, $\gamma = \frac{1+\sqrt{17}}{4} \approx 2.56$. Applying the analysis for (4.3) to bound $\tau(A)$, we obtain

$$\tau(A) \leq 36n(r+1)\rho_{r+1} \tag{5.18}$$

for symmetric $A \in R^{n\times n}$ of rank $r < n$, where ρ_{r+1} is the growth factor.

If all the non-zero eigenvalues are positive, then the matrix is semidefinite. and the LBL^T factorization with Bunch-Parlett pivoting strategy is equivalent to the LDL^T factorization with complete pivoting. With an argument similar to that used in obtaining (4.1), the bound on $\tau(A)$ is reduced to $\tau(A) \leq r$, where we use the 2-norm instead of the Frobenius norm. Following the proof of Theorem 5.2 but using (5.4) instead of (5.11), we find that

$$\|A - \hat{L}\hat{D}\hat{L}^T\|_2 \leq r(r+1)((\|W\|_2+1)^2+1)u\|A\|_2 + O(u^2).$$

A comparable bound for Cholesky factorization of a positive semidefinite matrix was given in [8] by Higham. Note that the bound on $\|W\|_2$ is also reduced because $\gamma = 1$. A similar analysis for diagonal dominant matrices gives

$$\|A - \hat{L}\hat{D}\hat{L}^T\|_\infty \leq (2n-r)(r+1)((\|W\|_\infty+1)^2+1)u\|A\|_\infty + O(u^2).$$

Note that $W = L_{11}^{-T}L_{21}^T$, where $L_{11} = L(1:r, 1:r)$ and $L_{21} = L(r+1:n, 1:r)$. Diagonal dominance guarantees that L_{11}^{-T} has all elements bounded by 1. Therefore, $\|W\|_\infty = \|L_{11}^{-T}L_{21}^T e\|_\infty \leq \|L_{11}^{-T}e\|_\infty \leq r$, which implies the potential of high stability.

5.4. Experiments. For rank estimation, the important practical issue is when to stop the factorization. In (5.16), the bound on $\|\hat{A}^{(r+1)}\|_F/\|A\|_F$ is governed by $\|W\|_F$ and $\tau(A)$. However, both bounds (5.17) and (5.18) are pessimistic. To investigate the typical ranges of $\|W\|_F$ and $\tau(A)$ in practice, we used the random matrices described as follows.

Each indefinite matrix was constructed as $Q\Lambda Q^T \in R^{n\times n}$, where Q is a random unitary matrix generated by the method in [12] (different for each matrix) and $\Lambda = \text{diag}(\lambda_i)$ of rank r. The following three test sets were used.

$$|\lambda_1| = |\lambda_2| = \cdots = |\lambda_{r-1}| = 1,\ \lambda_r = \sigma,$$
$$|\lambda_1| = |\lambda_2| = \cdots = |\lambda_{r-1}| = \sigma,\ \lambda_r = 1,$$
$$|\lambda_i| = \beta^i,\ i = 1, \ldots, r-1,\ \lambda_r = 1,$$

where $0 < \sigma \leq 1$, and $\beta^{r-1} = \sigma$ for $r > 1$. We assign the sign of λ_i randomly for $i = 1, \ldots, r-1$, and let t denote the number of negative eigenvalues. For each test set, we experimented with all combinations of $n = 10, 20, \ldots, 100$, $r = 2, 3, \ldots, n$, $t = 1, 2, \ldots, r-1$, and $\sigma = 1, 10^{-3}, \ldots, 10^{-12}$, for a total of $94,875$ indefinite matrices for each set. Among all the indefinite matrices, the largest $\|W\|_F$ and $\tau(A)$ were 34.054 and 3.173, respectively. Both numbers are modest. Note that the bounds (5.17) and (5.18) depend on r more than n, as does (5.16). Experimenting with several different stopping criteria, we suggest

$$\|\hat{A}_{k+1}\|_F \leq (k+1)^{3/2}u\|A\|_F, \tag{5.19}$$

which gave the best accuracy of the estimated ranks, or the less expensive

$$\|\hat{B}_i\|_F \leq (k+1)^{3/2}u\|B_1\|_F, \tag{5.20}$$

where $\hat{B}_i$ is the computed ith block pivot. With the Bunch-Parlett pivoting strategy, $\|A\|_M \approx \|B_1\|_M$ and $\|\hat{A}_{k+1}\|_M \approx \|B_i\|_M$, where $\mathrm{diag}(B_1, B_2, \ldots, B_{i-1}) \in R^{k\times k}$. Therefore, (5.19) and (5.20) are related.

One potential problem is that continuing the factorization on $\hat{S}_{r+1}$ could be unstable for rank estimation. However, the element growth is well-controlled by pivoting. The dimensions of Schur complements are reduced, whereas the upper bounds in (5.19) and (5.20) are increased during factorization. These properties safeguard the stability of rank estimation.

Our experiments were on a laptop with a Intel Celeron 2.8GHz CPU using IEEE standard arithmetic with machine epsilon $2^{-52} \approx 2.22 \times 10^{-16}$. The estimated ranks were all correct. We further experimented with $\sigma = 10^{-15}$, and instability occurred in the second and third test sets, since the conditioning of the non-singular part of a matrix affects the stability of rank estimation.

Forcing all the non-zero eigenvalues to be positive in the three test sets, we also experimented with rank estimation of positive semidefinite matrices by LDL^T factorization. With stopping criteria (5.19) and (5.20), all the estimated ranks were accurate for $\sigma = 1, 10^{-3}, \ldots, 10^{-12}$, and similar instability occurred with $\sigma = 10^{-15}$. The criteria suggested in [8] were for rank estimation of positive semidefinite matrices by Cholesky factorization. They result in less accuracy than (5.19) and (5.20) for indefinite matrices.

Both stopping criteria (5.19) and (5.20) work very well in our experiments except when $\sigma = 10^{-15}$. However, they may not be the best for all matrices. *A priori* information about the matrix, such as the size of $\|W\|$, growth factor and distribution of non-zero eigenvalues, may help adjust the stopping criterion.

6. Concluding Remarks. Table 6.1 lists the highest order terms of the bounds on $\|C\|_S$ for symmetric matrices and symmetric triadic matrices, where n is the matrix size. The references to their definitions and bounds are also included. For LBL^T factorization, the constant c is from (2.6). For singular matrices, r denotes the rank. It shows the improvement of bounds on backward errors because of triadic structure.

TABLE 6.1
Bounds on $\|C\|_S$ for LDL^T and LBL^T factorizations

Bounds on $\|C\|_S$		Def.	General		Triadic	
LDL^T	Nonsingular	(2.2)	$\frac{1}{3}n^3u$	(2.4)	$9nu$	(2.5)
	Solving $Ax=b$	(3.3)	$\frac{11}{6}n^3u$	(3.4)	$7n^2u$	(3.5)
	Singular	(5.3)	$(nr(n-r)+\frac{1}{3}r^3)u$	(5.5)	$18ru$	(5.6)
LBL^T	Nonsingular	(2.10)	$\frac{4}{3}n^3u$	(2.11)	$4(7c+36)nu$	(2.12)
	Solving $Ax=b$	(3.6)	$\frac{17}{6}n^3u$	(3.7)	$28n^2u$	(3.8)
	Singular	(5.10)	$(4nr(n-r)+\frac{4}{3}r^3)u$	(5.12)	$36(cr+4r+1)u$	(5.13)

We have studied the componentwise backward error analysis and normwise analysis for LBL^T factorization and its applications to solving linear systems and rank estimation. Our concluding remarks are listed below.

1. LDL^T factorization and its application to solve linear systems are stable if the growth factor is modest. Both positive definiteness and diagonal dominance guarantee the stability, because the growth factors are bounded by 1 and 2, respectively. A modest growth factor does not guarantee a stable LBL^T factorization. Nevertheless, LBL^T factorization and its application to solve

symmetric linear systems are stable if conditions (2.6) and (4.2) hold. All the pivoting strategies in the literature [1, 2, 3, 4, 7] satisfy both conditions.

2. In [9] and [10], Higham proved the stability of the Bunch-Kaufman pivoting strategy [3] and Bunch's pivoting method [2], respectively. His componentwise backward error analysis is based on the LBL^T factorization in outer product form. In this paper, we presented a new proof of the componentwise backward stability using inner product formulation. We also gave a sufficient condition such that an LBL^T factorization is guaranteed to run to completion with inertia preserved in Theorem 4.3.
3. We also analyzed the rank estimation of symmetric indefinite matrices using LBL^T factorization with Bunch-Parlett pivoting. In our experiments, both stopping criteria (5.19) and (5.20) give accurate estimated ranks except when $\sigma = 10^{-15}$ (i.e., the non-singular part is ill-conditioned). We recommend (5.20) because the cost is negligible. Both criteria can be also used for rank estimation of symmetric semidefinite matrices by LDL^T factorization with complete pivoting, and give better results than the stopping criteria of [8] for indefinite matrices.
4. Due to the sparsity, the stability of LDL^T and LBL^T factorizations is improved for symmetric triadic matrices, as shown in Table 6.1.

Acknowledgment. The author is very grateful to his research advisor, Dianne P. O'Leary for helpful discussions, invaluable comments (e.g., pointing out the simpler proof of Lemma 3.2), and her help in editing this article. It is also a pleasure to thank Che-Rung Lee for a helpful discussion about eigenvalues.

REFERENCES

[1] C. Ashcraft, R. G. Grimes, and J. G. Lewis, *Accurate symmetric indefinite linear equation solvers*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 513–561.
[2] J. R. Bunch, *Partial pivoting strategies for symmetric matrices*, SIAM J. Numerical Analysis, 11 (1974), pp. 521–528.
[3] J. R. Bunch and L. Kaufman, *Some stable methods for calculating inertia and solving symmetric linear systems*, Math. Comp., 31 (1977), pp. 163–179.
[4] J. R. Bunch and B. N. Parlett, *Direct methods for solving symmetric indefinite systems of linear equations*, SIAM J. Numerical Analysis, 8 (1974), pp. 639–655.
[5] J. W. Demmel, N. J. Higham, and R. S. Schreiber, *Block LU factorization*, Numerical Linear Algebra with Applications, 2 (1995), pp. 173–190.
[6] J. J. Dongarra, I. S. Duff, D. C. Sorensen, and H. A. van der Vorst, *Solving Linear Systems on Vector and Shared Memory Computers*, SIAM, 1991.
[7] Haw-Ren Fang and Dianne P. O'Leary, *Stable factorizations of symmetric tridiagonal and triadic matrices*, Tech. Report CS-4733, Computer Science Department, Univ. of Maryland, College Park, MD, July, 2005.
[8] N. J. Higham, *Analysis of the Cholesky Decomposition of a Semi-Definite Matrix*, M. G. Cox and S. Hammarling, eds., Oxford University Press, 1990, ch. Reliable Numerical Computation, pp. 161–185.
[9] ———, *Stability of the diagonal pivoting method with partial pivoting*, SIAM J. Matrix Analy. Appl., 18 (1997), pp. 52–65.
[10] ———, *Stability of block LDL^T factorization of a symmetric tridiagonal matrix*, SIAM J. Matrix Anal. Appl., 287 (1999), pp. 181–189.
[11] ———, *Accuracy and Stability of Numerical Algorithms*, SIAM, 2002.
[12] G. W. Stewart, *The efficient generation of random orthogonal matrices with an application to condition estimation*, SIAM J. Numer. Anal., 17 (1980), pp. 403–409.
[13] G. W. Stewart and J.-g. Sun, *Matrix Perturbation Theory*, Academic Press, 1990.
[14] J. H. Wilkinson, *A priori error analysis of algebraic processes (cited in [11])*, In Proc. International Congress of Mathematicians, 25 (1968), pp. 629–640.