Series expansion of the LU and Cholesky factors of a matrix

Froilán M. Dopico and Juan M. Molera

Departamento de Matemáticas
Universidad Carlos III de Madrid
Spain

Setting the problem

Given a square matrix $A = LU$ having LU factorization and a perturbation $\widetilde{A} \equiv A + E = \widetilde{L}\widetilde{U}$, we look for two convergent series of matrices:

$$\widetilde{L} = \sum_{k=0}^{\infty} L_k \quad \text{and} \quad \widetilde{U} = \sum_{k=0}^{\infty} U_k$$

with $L_k = O(\|E\|^k)$, $U_k = O(\|E\|^k)$, $L_0 = L$, $U_0 = U$, as well as for domains and rates of convergence.

The same approach is valid for the Cholesky factor of positive definite Hermitian matrices.

Previous work on perturbations of LU

- Barrlund: strict normwise perturbation bounds.
- Sun: strict norm and componentwise perturbation bounds.
- Stewart: first-order perturbation expansion and bounds on the norm of the second-order terms.
- Chang and Paige: optimum first order bounds, i.e. condition numbers.

Series for $I + F$

If $I + L^{-1}EU^{-1} = \mathcal{L}\mathcal{U}$ then

$$\widetilde{A} = L(I + L^{-1}EU^{-1})U = (L\mathcal{L})(\mathcal{U}U) \equiv \widetilde{L}\widetilde{U}.$$

Therefore, we can focus on the series of the LU factors of a perturbation of the identity: $I + F = \mathcal{L}\mathcal{U}$.

We will use absolute, or monotone, consistent norms

$$|A| \leq |B| \Rightarrow \|A\| \leq \|B\| \quad \text{and} \quad \|AB\| \leq \|A\|\|B\|.$$

Let us introduce for simplicity a parameter z

$$I + zF = \mathcal{L}(z)\mathcal{U}(z).$$

For $\|zF\| < 1$, the LU factors of $I + zF$

- exist and are unique,
- their entries are rational functions of z with nonzero denominators, therefore

$$\mathcal{L}(z) = \sum_{k=0}^{\infty} z^k \mathcal{L}_k \quad \text{and} \quad \mathcal{U}(z) = \sum_{k=0}^{\infty} z^k \mathcal{U}_k.$$

with $\quad \mathcal{L}_0 = \mathcal{U}_0 = I$.

Recurrence relations for $\mathcal{L}_k$ and $\mathcal{U}_k$

$$I + zF = \mathcal{L}(z)\mathcal{U}(z) = \sum_{k=0}^{\infty} \left(\sum_{j=0}^{k} \mathcal{L}_j \mathcal{U}_{k-j} \right) z^k.$$

Consequently

$$\begin{aligned} F &= \mathcal{U}_1 + \mathcal{L}_1 \\ 0 &= \mathcal{U}_k + \mathcal{L}_1 \mathcal{U}_{k-1} + \cdots + \mathcal{L}_{k-1}\mathcal{U}_1 + \mathcal{L}_k \quad \text{for } k \geq 2. \end{aligned}$$

Setting $z = 1$, we can state:

Theorem. *Let F be an $n \times n$ matrix with $\|F\| < 1$ then:*

1. *$I + F$ has a unique LU factorization: $I + F = \mathcal{L}\mathcal{U}$.*
2. $$\mathcal{L} = \sum_{k=0}^{\infty} \mathcal{L}_k \quad \textit{and} \quad \mathcal{U} = \sum_{k=0}^{\infty} \mathcal{U}_k.$$
   where $\mathcal{L}_0 = I$, $\mathcal{U}_0 = I$, $\mathcal{L}_1 = F_L$, $\mathcal{U}_1 = F_U$ and for $k \geq 2$:
   $$\mathcal{L}_k = (-\mathcal{L}_1 \mathcal{U}_{k-1} - \mathcal{L}_2 \mathcal{U}_{k-2} \cdots - \mathcal{L}_{k-1} \mathcal{U}_1)_L,$$
   $$\mathcal{U}_k = (-\mathcal{L}_1 \mathcal{U}_{k-1} - \mathcal{L}_2 \mathcal{U}_{k-2} \cdots - \mathcal{L}_{k-1} \mathcal{U}_1)_U.$$
3. *$L_k = O(\|F\|^k)$ and $U_k = O(\|F\|^k)$.*

Here

- $(\cdot)_L$ stands for the strict lower triangular part.
- $(\cdot)_U$ stands for the upper triangular part.

A few terms:

$$
\begin{aligned}
\mathcal{L}_1 + \mathcal{U}_1 &= F \\
\mathcal{L}_2 + \mathcal{U}_2 &= -F_L F_U \\
\mathcal{L}_3 + \mathcal{U}_3 &= F_L (F_L F_U)_U + (F_L F_U)_L F_U \\
\mathcal{L}_4 + \mathcal{U}_4 &= -F_L (F_L (F_L F_U)_U)_U - F_L ((F_L F_U)_L F_U)_U \\
&\quad -(F_L F_U)_L (F_L F_U)_U - (F_L (F_L F_U)_U)_L F_U \\
&\quad -((F_L F_U)_L F_U)_L F_U
\end{aligned}
$$

Rate of convergence

Theorem: *Let F be an $n \times n$ matrix with $\|F\| < 1$, let $I + F = \mathcal{L}\mathcal{U}$ be the unique LU factorization. If*

$$\mathcal{L} = \sum_{k=0}^{\infty} \mathcal{L}_k \quad \text{and} \quad \mathcal{U} = \sum_{k=0}^{\infty} \mathcal{U}_k$$

then

$$|\mathcal{L}_k + \mathcal{U}_k| \leq |F|^k \quad \text{for} \quad k \geq 1.$$

Therefore $|\mathcal{L}_k| \leq \left(|F|^k\right)_L$ and $|\mathcal{U}_k| \leq \left(|F|^k\right)_U$.

Idea of the proof for $k = 4$:

sing $|C_L D_U| \leq |C|_L |D|_U$

$$
\begin{aligned}
|\mathcal{L}_4 + \mathcal{U}_4| &\leq |F|_L(|F|_L(|F|_L|F|_U)_U)_U + |F|_L((|F|_L|F|_U)_L|F|_U)_U \\
&\quad +(|F|_L|F|_U)_L(|F|_L|F|_U)_U + (|F|_L(|F|_L|F|_U)_U)_L|F|_U \\
&\quad +((|F|_L|F|_U)_L|F|_U)_L|F|_U \\
&\leq |F|_L|F|_L|F|_L|F|_U + |F|_L(|F|_L|F|_U)_L|F|_U \\
&\quad +|F|_L|F|_U|F|_L|F|_U + |F|_L(|F|_L|F|_U)_U|F|_U + |F|_L|F|_U|F|_U| \\
&= |F|_L|F|_L|F|_L|F|_U + |F|_L|F|_L|F|_U|F|_U \\
&\quad +|F|_L|F|_U|F|_L|F|_U + |F|_L|F|_U|F|_U|F|_U \\
&\leq (|F|_L + |F|_U)^4 \\
&= |F|^4
\end{aligned}
$$

Bounds on the remainders

Theorem: *Let F be an $n \times n$ matrix with $\|F\| < 1$, let $I + F = \mathcal{L}\mathcal{U}$ be the unique LU factorization, then*

$$\left|\mathcal{L} - \sum_{k=0}^{N} \mathcal{L}_k\right| \leq \left(|F|^{N+1}(I - |F|)^{-1}\right)_L,$$

$$\left|\mathcal{U} - \sum_{k=0}^{N} \mathcal{U}_k\right| \leq \left(|F|^{N+1}(I - |F|)^{-1}\right)_U.$$

Proof:

$$\left|\mathcal{L} - \sum_{k=0}^{N} \mathcal{L}_k\right| \leq \sum_{k=N+1}^{\infty} |\mathcal{L}_k| \leq \sum_{k=N+1}^{\infty} \left(|F|^k\right)_L$$

$$= \left(\sum_{k=N+1}^{\infty} |F|^k\right)_L = \left(|F|^{N+1} \sum_{k=0}^{\infty} |F|^k\right)_L$$

Normwise Bounds

$$\left\|\mathcal{L} - \sum_{k=0}^{N} \mathcal{L}_k\right\| \leq \frac{\|F\|^{N+1}}{1 - \|F\|}$$

Comparison with $N = 1$ by Stewart.

Three improvements

1. Smaller bounds.
2. Componentwise bounds.
3. Bounds valid for $\|F\| < 1$ (Stewart's $\|F\| \leq 1/4$).

Proving strict bounds (Sun 1992, Barrlund 1991

Given $A = LU$ and E such that $\|L^{-1}EU^{-1}\| < 1$, then $A + E = \widetilde{L}\widetilde{U}$ has a unique LU factorization. Denote $F = L^{-1}EU^{-1}$.

$$\widetilde{A} = L(I+F)U = (L\mathcal{L})(\mathcal{U}U) = \underbrace{\left(L\sum_{k=0}^{\infty}\mathcal{L}_k\right)}_{\widetilde{L}}\underbrace{\left(\sum_{k=0}^{\infty}\mathcal{U}_k\right)U}_{\widetilde{U}}.$$

Then

$$|\widetilde{U} - U| = \left|\left(\sum_{k=1}^{\infty}\mathcal{U}_k\right)U\right| \le \left(\sum_{k=1}^{\infty}|\mathcal{U}_k|\right)|U|$$

$$\le \left(\sum_{k=1}^{\infty}\left(|F|^k\right)_U\right)|U|$$

$$\le \left(|F|(I - |F|)^{-1}\right)_U |U|,$$

and

$$\|\widetilde{U} - U\| \le \frac{\|L^{-1}EU^{-1}\|\|U\|}{1 - \|L^{-1}EU^{-1}\|}$$

Similar results for L.