AMS526: Numerical Analysis I (Numerical Linear Algebra)

Lecture 15: Stability of LU Factorization; Cholesky Factorization

Xiangmin Jiao

SUNY Stony Brook

October 30, 2008

Outline

1 Stability of LU Factorization

2 Cholesky Factorization

Stability of LU without Pivoting

- For $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$ computed without pivoting

$$\tilde{\boldsymbol{L}}\tilde{\boldsymbol{U}} = \boldsymbol{A} + \delta\boldsymbol{A}, \qquad \frac{\|\delta\boldsymbol{A}\|}{\|\boldsymbol{L}\|\|\boldsymbol{U}\|} = O(\epsilon_{\text{machine}})$$

- This is close to backward stability, except that we have $\|\boldsymbol{L}\|\|\boldsymbol{U}\|$ instead of $\|\boldsymbol{A}\|$ in the denominator
- Unfortunately, $\|\boldsymbol{L}\|$ and $\|\boldsymbol{U}\|$ can be arbitrarily large (even for well-conditioned $\boldsymbol{A}$), e.g.,

$$\boldsymbol{A} = \begin{bmatrix} 10^{-20} & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 10^{20} & 1 \end{bmatrix} \begin{bmatrix} 10^{-20} & 1 \\ 0 & 1 - 10^{20} \end{bmatrix}$$

- Therefore, the algorithm is *unstable*

Stability of LU with Partial Pivoting

- With pivoting, all entries of $\boldsymbol{L}$ are in $[-1,1]$, so $\|L\| = O(1)$
- To measure growth in $\boldsymbol{U}$, we introduce the growth factor $\rho = \frac{\max_{i,j}|u_{ij}|}{\max_{i,j}|a_{ij}|}$, and hence $\|\boldsymbol{U}\| = O(\rho\|\boldsymbol{A}\|)$
- We then have $\boldsymbol{PA} = \boldsymbol{LU}$

$$\tilde{\boldsymbol{L}}\tilde{\boldsymbol{U}} = \boldsymbol{PA} + \delta\boldsymbol{A}, \qquad \frac{\|\delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|} = O(\rho\epsilon_{\text{machine}})$$

- If $\rho = O(1)$, then the algorithm is backward stable
- In fact, $\rho \leq 2^{m-1}$, so by definition ρ is a constant but can be very large

The Growth Factor

- **Bad news**: ρ can indeed be as large as 2^{m-1}. Consider matrix

$$\begin{bmatrix} 1 & & & & 1 \\ -1 & 1 & & & 1 \\ -1 & -1 & 1 & & 1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & & 0 \\ -1 & 1 & & & 0 \\ -1 & -1 & 1 & & 0 \\ -1 & -1 & -1 & 1 & 0 \\ -1 & -1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & 1 \\ & 1 & & & 2 \\ & & 1 & & 4 \\ & & & 1 & 8 \\ & & & & 16 \end{bmatrix}$$

  where growth factor $\rho = 16 = 2^{m-1}$
- **Good news**: Large ρ occurs only for very skewed matrices, so the probability of large ρ decreases exponentially in ρ
- In practice, ρ is no larger than $O(\sqrt{m})$. However, this behavior is not fully understood yet
- In conclusion,
  - Gaussian elimination with partial pivoting is backward stable
  - In theory, its error may grow exponentially in m
  - In practice, it is stable for matrices of practical interests

Outline

1. Stability of LU Factorization

2. Cholesky Factorization

Hermitian Positive-Definite Matrices

- Symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{m \times m}$ is *symmetric positive definite* (SPD) if $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0$ for $\boldsymbol{x} \in \mathbb{R}^m \backslash \{\mathbf{0}\}$
- Hermitian matrix $\boldsymbol{A} \in \mathbb{C}^{m \times m}$ is *Hermitian positive definite* (HPD) if $\boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x} > 0$ for $\boldsymbol{x} \in \mathbb{C}^m \backslash \{\mathbf{0}\}$
- If $\boldsymbol{A}$ is $m \times m$ HPD and $\boldsymbol{X} \in \mathbb{C}^{m \times n}$ has full column rank, then $\boldsymbol{X}^* \boldsymbol{A} \boldsymbol{X}$ is HPD
- Any principal submatrix (picking some rows and corresponding columns) of $\boldsymbol{A}$ is HPD and $a_{ii} > 0$
- HPD matrices have positive real eigenvalues and orthogonal eigenvectors
- Note: Most textbooks only talk about SPD or HPD matrices, but a positive-definite matrix does not need to be symmetric or Hermitian! A real matrix $\boldsymbol{A}$ is positive definite iff $\boldsymbol{A} + \boldsymbol{A}^T$ is SPD.

Cholesky Factorization

- Key idea: take advantage and preserve the properties of symmetry and positive-definiteness in factorization
- Eliminate below diagonal and to the right of diagonal

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \boldsymbol{w}^* \\ \boldsymbol{w} & \boldsymbol{K} \end{bmatrix} = \begin{bmatrix} \alpha & \boldsymbol{0} \\ \boldsymbol{w}/\alpha & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \alpha & \boldsymbol{w}^*/\alpha \\ \boldsymbol{0} & \boldsymbol{K} - \boldsymbol{w}\boldsymbol{w}^*/a_{11} \end{bmatrix}$$
$$= \begin{bmatrix} \alpha & \boldsymbol{0} \\ \boldsymbol{w}/\alpha & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \boldsymbol{0} & \boldsymbol{K} - \boldsymbol{w}\boldsymbol{w}^*/a_{11} \end{bmatrix} \begin{bmatrix} \alpha & \boldsymbol{w}^*/\alpha \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} = \boldsymbol{R}_1^* \boldsymbol{A}_1 \boldsymbol{R}_1$$

where $\alpha = \sqrt{a_{11}}$, where $a_{11} > 0$

- $\boldsymbol{K} - \boldsymbol{w}\boldsymbol{w}^*/a_{11}$ is principal submatrix of HPD $\boldsymbol{A}_1 = \boldsymbol{R}_1^{-*} \boldsymbol{A} \boldsymbol{R}_1^{-1}$ and therefore is HPD, with positive diagonal entries

Cholesky Factorization

- Apply recursively to obtain

$$A = (R_1^* R_2^* \cdots R_m^*)(R_m \cdots R_2 R_1) = R^* R, \qquad r_{jj} > 0$$

which is known as *Cholesky factorization*

- Question: Is R simply "union" of kth rows of R_k (or R^* "union" of kth columns of R_k^*)? Yes. Hint: Write R_k^* in a form similar to $L_k = I + \ell_k e_k^T$ in LU
- Existence and uniqueness: every HPD matrix has a unique Cholesky factorization
  - Exists because algorithm for Cholesky factorization always works for HPD matrices
  - Is unique since once $\alpha = \sqrt{a_{11}}$ is determined at each step, entire column w/α is determined
  - Question: How to check whether a Hermitian matrix is positive definite? Answer: Run Cholesky factorization and it would succeeds iff the matrix is positive definite.

Algorithm of Cholesky Factorization

- Factorize Hermitian positive definite matrix $\boldsymbol{A} \in \mathbb{C}^{m \times m}$ into $\boldsymbol{A} = \boldsymbol{R}^* \boldsymbol{R}$

```
Algorithm: Cholesky factorization
  R = A
  for k = 1 to m
      for j = k + 1 to m
          r_{j,j:m} ← r_{j,j:m} − r_{k,j:m} r_{kj} / r_{kk}
      r_{k,k:m} ← r_{k,k:m} / sqrt(r_{kk})
```

$$\boldsymbol{R} = \boldsymbol{A}$$
for $k = 1$ to m
    for $j = k + 1$ to m
$$\boldsymbol{r}_{j,j:m} \leftarrow \boldsymbol{r}_{j,j:m} - \boldsymbol{r}_{k,j:m} r_{kj} / r_{kk}$$
$$\boldsymbol{r}_{k,k:m} \leftarrow \boldsymbol{r}_{k,k:m} / \sqrt{r_{kk}}$$

- Operation count

$$\sum_{k=1}^{m} \sum_{j=k+1}^{m} 2(m-j) \sim 2 \sum_{k=1}^{m} \sum_{j=1}^{k} j \sim \sum_{k=1}^{m} k^2 \sim m^3/3$$

LDL^* Factorization

- Cholesky factorization is sometimes given by $\boldsymbol{A} = \boldsymbol{LDL}^*$ where $\boldsymbol{D}$ is diagonal matrix and $\boldsymbol{L}$ is unit lower triangular matrix
- It avoids computing square roots
  - Unlike Cholesky factorization, $\boldsymbol{LDL}^*$ factorization works for Hermitian matrices that are not positive definite
- Analogously, LU factorization can also be written as LDU, where $\boldsymbol{U}$ is then unit upper triangular
- Question: How is $\boldsymbol{R}$ in $\boldsymbol{A} = \boldsymbol{R}^*\boldsymbol{R}$ related to the $\boldsymbol{L}$ and $\boldsymbol{U}$ factors of $\boldsymbol{A} = \boldsymbol{LU}$?
  - $\boldsymbol{U} = \boldsymbol{DL}^* = \sqrt{\boldsymbol{D}}\boldsymbol{R}$, where $\sqrt{\boldsymbol{D}} \equiv \mathrm{diag}(\sqrt{d_{11}}, \sqrt{d_{22}}, \cdots, \sqrt{d_{mm}})$

Stability

Theorem

The computed Cholesky factor $\tilde{R}$ satisfies

$$\tilde{R}^*\tilde{R} = A + \delta A, \qquad \frac{\|\delta A\|}{\|A\|} = O(\epsilon_{machine}),$$

i.e., Cholesky factorization is backward stable

- Forward errors in $\tilde{R}$ is $\|\tilde{R} - R\|/\|R\| = O(\kappa(A)\epsilon_{\text{machine}})$, which may be large for ill-conditioned A
- Solve $Ax = b$ for positive definite A
  - Factorize $A = R^*R$; Solve $R^*y = b$; Solve $Rx = y$
  - Operation count is $\sim m^3/3$
  - Algorithm is backward stable:

$$(A + \Delta A)\tilde{x} = b, \qquad \frac{\|\Delta A\|}{\|A\|} = O(\epsilon_{\text{machine}})$$