Convex Optimization: Reading Notes 5

GKxx

May 11, 2022

1 Triangular factorizations

Definition 1.1 (LU factorization). *Let $A \in \mathbb{C}^{n\times n}$. A presentation $A = LU$, in which $L \in \mathbb{C}^{n\times n}$ is lower triangular and $U \in \mathbb{C}^{n\times n}$ is upper triangular, is called an* LU factorization *of A.*

Theorem 1.2 (Row inclusion). *Let $A \in \mathbb{C}^{n\times n}$ be given. A has an LU factorization in which L is nonsingular if and only if A has the* row inclusion property*: For each $i = 1, \cdots, n-1$, $A[\{i+1; 1, \cdots, i\}]$ is a linear combination of the rows of $A[\{1, \cdots, i\}]$.*

Here we use $A[\{i+1; 1, \cdots, i\}]$ to denote the vector $\begin{bmatrix} a_{(i+1)1} & a_{(i+1)2} & \cdots & a_{(i+1)i} \end{bmatrix}$, where $A = [a_{ij}]_{i,j=1}^n$.

Remark 1.3 (Leading principal submatrix). *For $A \in \mathbb{C}^{n\times n}$, the i-th* leading principal submatrix *of A, denoted $A[\{1, \cdots, i\}]$, is the submatrix obtained from A by deleting the last $n-i+1$ rows and columns.*

The proof of Theorem 1.2 is as follows.

Proof. Suppose $A = LU \in \mathbb{C}^{n\times n}$ is the LU factorization with L nonsingular. First we show that $A[\{n; 1, \cdots, n-1\}]$ is the linear combination of the rows of $A[\{1, \cdots, n-1\}]$. Partition A, L and U as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}, \quad U = \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix},$$

where $A_{11}, L_{11}, U_{11} \in \mathbb{C}^{(n-1)\times(n-1)}$. Since $A = LU$, we have $A_{21} = L_{21}U_{11} = L_{21}L_{11}^{-1}L_{11}U_{11}$, and $A_{11} = L_{11}U_{11}$. Therefore $A_{21} = \left(L_{21}L_{11}^{-1}\right)A_{11}$, which represents $A_{21} = A[\{n; 1, \cdots, n-1\}]$ as the linear combination of the rows of $A[\{1, \cdots, n-1\}]$. Moreover, since $A = LU$, for every $i = 1, \cdots, n$ we have

$$A[\{1, \cdots, i\}] = L[\{1, \cdots, i\}]\, U[\{1, \cdots, i\}].$$

Applying what we have obtained to this LU factorization of every leading principal submatrix of A verifies the row inclusion property.

Conversely, if A has the row inclusion property, we can construct an LU factorization inductively with L nonsingular. The cases $n = 1, 2$ are easily verified. Now suppose $A_{11} =$

$L_{11}U_{11}$, where L_{11} is nonsingular and $A_{11}, L_{11}, U_{11} \in \mathbb{C}^{(n-1)\times(n-1)}$. The row inclusion property gives that the row vector A_{21} is a linear combination of the rows of A_{11}, so there exists $y \in \mathbb{C}^{n-1}$ such that

$$A_{21} = y^T A_{11} = y^T L_{11} U_{11}.$$

From $A_{21} = L_{21}U_{11}$, we get $L_{21} = y^T L_{11}$. From $A_{12} = L_{11}U_{12}$ and that L_{11} is nonsingular, we obtain $U_{12} = L_{11}^{-1}A_{12}$. Let $L_{22} = 1$ and $U_{22} = A_{22} - L_{21}U_{12}$. In this way, we obtain an LU factorization of A, in which L is nonsingular, from the LU factorization of A_{11}. □

Remark 1.4. *Similarly we can define the* column inclusion property. *A has an LU factorization with nonsingular U if and only if the column inclusion property holds. This follows from considering the LU factorization of A^T.*

Corollary 1.5. *Suppose that $A \in \mathbb{C}^{n\times n}$ and* $\operatorname{rank} A = k$*. If $A[\{1, \cdots, j\}]$ is nonsingular for all $j = 1, \cdots, k$, then A has an LU factorization. Furthermore, either factor may be chosen to be unit triangular; both L and U are nonsingular if and only if $k = n$.*

Proof. For $A \in \mathbb{C}^{n\times n}$ with $\operatorname{rank} A = k$ such that $A[\{1, \cdots, j\}]$ is nonsingular for all $j = 1, \cdots, k$, we first verify that A has both the row inclusion and column inclusion properties. For $j = 1, \cdots, k$, since $A[\{1, \cdots, j\}]$ is nonsingular, the rows of $A[\{1, \cdots, j\}]$ are linearly independent, so they span $\mathbb{C}^j$ and $A[\{j+1; 1, \cdots, j\}]$ is certainly in it. For $j > k$, since $\operatorname{rank} A = k$, the rows of $A[\{1, \cdots, j\}]$ must span $\mathbb{C}^j$, and $A[\{j+1; 1, \cdots, j\}]$ is in it. This shows the row inclusion property, and similarly the column inclusion property also holds. From Theorem 1.2 and Remark 1.4 it follows that A has an LU factorization where either L or U may be nonsingular.

Suppose $A = LU$ and L is nonsingular, which means that L has nonzero diagonal elements $\ell_{11}, \ell_{22}, \cdots, \ell_{nn}$. Let $D = \operatorname{diag}(\ell_{11}, \cdots, \ell_{nn})$ which is nonsingular, and let $L = L'D$ so that L' is unit lower triangular. Note that $U' = DU$ is still upper triangular, so we obtain a new LU factorization of $A = L'U'$ in which the left factor is unit triangular. Similarly, the right factor could also be unit triangular.

For $k = n$, the matrix A is nonsingular, so the other factor must be nonsingular if either factor is nonsingular. Conversely, for both L and U nonsingular, $A = LU$ is nonsingular, and therefore has full rank. □

Corollary 1.6 (LDU factorization)**.** *Let $A = [a_{ij}] \in \mathbb{C}^{n\times n}$ be given. Suppose that the leading principal submatrix $A[\{1, \cdots, i\}]$ is nonsingular for all $i = 1, \cdots, n$. Then $A = LDU$, in which $L, D, U \in \mathbb{C}^{n\times n}$, L is unit lower triangular, U is unit upper triangular, $D = \operatorname{diag}(d_1, \cdots, d_n)$ is diagonal, $d_1 = a_{11}$, and*

$$d_i = \frac{\det A[\{1, \cdots, i\}]}{\det A[\{1, \cdots, i-1\}]}.$$

The factors L, D and U are uniquely determined.

Proof. From Corollary 1.5, A must have an LU factorization $A = L'U'$ where both L' and U' are nonsingular. It is easy to find unit triangular L and U such that $L' = LD_1$ and $D_2U = U'$, where D_1 and D_2 are diagonal. Then A has an LDU factorization $A = LDU, D = D_1 D_2$, and

$$\det A\left[\{1,\cdots,i\}\right] = \det D\left[\{1,\cdots,i\}\right] = \prod_{j=1}^{i} d_j, \quad i = 1,\cdots,n.$$

So $d_1 = a_{11}$, and

$$d_i = \frac{\det A\left[\{1,\cdots,i\}\right]}{\det A\left[\{1,\cdots,i-1\}\right]}, \quad i = 2,\cdots,n.$$

The uniqueness of L and U could be proved inductively. □

Lemma 1.7. *Let $A \in \mathbb{C}^{n\times n}$ be nonsingular. Then there is a permutation matrix $P \in \mathbb{C}^{n\times n}$ such that* $\det\left(\left(P^T A\right)\left[\{1,\cdots,j\}\right]\right) \neq 0$ *for $j = 1,\cdots,n$.*

Proof. Prove by induction on n. The cases $n = 1, 2$ are trivial. Suppose that such matrix exists for cases $1, 2, \cdots, n-1$. For a nonsingular $A \in \mathbb{C}^{n\times n}$, we delete its last column, and the remaining $n-1$ columns are linearly independent and hence they contain $n-1$ linearly independent rows. Permute these rows to be the first $n-1$ rows, and then apply the induction hypothesis to the nonsingular $(n-1)\times(n-1)$ leading principal submatrix. This determines a desired permutation P, and $P^T A$ is nonsingular. □

Theorem 1.8 (PLU factorization)**.** *For each $A \in \mathbb{C}^{n\times n}$ there is a permutation matrix $P \in \mathbb{C}^{n\times n}$, a unit lower triangular $L \in \mathbb{C}^{n\times n}$ and an upper triangular $U \in \mathbb{C}^{n\times n}$ such that $A = PLU$.*

Proof. It suffices to show that there exists a permutation matrix Q such that QA has the row inclusion property, and thus has an LU factorization $QA = LU$. Let $P = Q^T$ and we have $A = PLU$.

If A is nonsingular, the desired permutation matrix is guaranteed by Lemma 1.7. Suppose A is singular, with $\operatorname{rank} A = k < n$. First we permute the rows of A so that the first k rows are linearly independent. This gives that $A\left[\{i+1; 1,\cdots,i\}\right]$ is a linear combination of the rows of $A\left[\{1,\cdots,i\}\right]$ for $i = k,\cdots,n-1$. If $A\left[\{1,\cdots,k\}\right]$ is nonsingular, apply Lemma 1.7 again to $A\left[\{1,\cdots,k\}\right]$ so that the row inclusion property holds for $A\left[\{1,\cdots,k\}\right]$, and thus holds for A. If $A\left[\{1,\cdots,k\}\right]$ is singular, apply the same procedure recursively to $A\left[\{1,\cdots,k\}\right]$ until either the upper left block is 0, or it is nonsingular. Hence the desired permutation matrix exists. □

Theorem 1.9 (LPU factorization)**.** *For every $A \in \mathbb{C}^{n\times n}$, there exists a permutation matrix $P \in \mathbb{C}^{n\times n}$, a unit lower triangular matrix $L \in \mathbb{C}^{n\times n}$ and an upper triangular matrix $U \in \mathbb{C}^{n\times n}$ such that $A = LPU$. Moreover, the factor P is uniquely determined if A is nonsingular.*

Theorem 1.9 could be proved by induction. The proof is omitted here.

Theorem 1.10 ($LPDU$ factorization)**.** *For each nonsingular $A \in \mathbb{C}^{n\times n}$, there is a unique permutation matrix P, a unique nonsingular diagonal matrix D, a unit lower triangular matrix L, and a unit upper triangular matrix U such that $A = LPDU$.*

Proof. Theorem 1.9 guarantees the existence of a unique permutation matrix P, a unit lower triangular matrix L and a nonsingular upper triangular matrix U' such that $A = LPU'$. Let $D = \operatorname{diag}(\operatorname{diag}(U'))$ so that $U' = DU$, where U is unit upper triangular. This shows the existence.

To show the uniqueness of D, we assume there exists another diagonal matrix D_1, unit lower triangular L_1 and unit upper triangular U_1 such that $A = L_1PD_1U_1$. Since $A = LPDU$, we have $LPDU = L_1PD_1U_1$, and thus

$$P^T L_1^{-1} LPD = D_1 U_1 U^{-1}.$$

Sicne the unit lower/upper triangular matrices form a multiplicative group, the main diagonal entries of $L_1^{-1}L$ and U_1U^{-1} are all ones. Since P is a permutation matrix, it is clear that the main diagonal entries of $P^T(L_1^{-1}L)P$ are also all ones. Hence $D = D_1$. $\square$

2 Cholesky factorization

Theorem 2.1 (Cholesky factorization)**.** *Let $A \in \mathbb{S}^n$ be given. A is positive definite if and only if there exists a lower triangular matrix L with positive diagonal elements such that $A = LL^T$. Such L is unique and is called the* Cholesky factor *of A.*

Proof. $\Longleftarrow$: Suppose $A = LL^T$ in which L is lower triangular with positive diagonal elements. It follows immediately that A is positive semidefinite. Since the diagonal elements of L are all positive, L and L^T are nonsingular, so $A = LL^T$ is nonsingular, and thus positive definite.

$\Longrightarrow$: We will show the existence of the Cholesky factor L by induction on n. The case $n = 1$ is trivial. Suppose this is true up to $n - 1$, and for $A \in \mathbb{S}^n$, write

$$A = \begin{bmatrix} \alpha & b^T \\ b & A' \end{bmatrix}, \quad A' \in \mathbb{R}^{(n-1)\times(n-1)}.$$

Decompose this in the form

$$A = \begin{bmatrix} 1 & 0 \\ \frac{1}{\alpha}b & I_{n-1} \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & A' - \frac{bb^T}{\alpha} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{\alpha}b^T \\ 0 & I_{n-1} \end{bmatrix},$$

where $\Delta_A = A' - \frac{bb^T}{\alpha}$ is the *Schur complement* of A. Since A is positive definite, it follows that the Schur complement Δ_A is also positive definite. By induction hypothesis there exists a lower triangular matrix L_Δ with positive diagonal elements such that $\Delta_A = L_\Delta L_\Delta^T$. Therefore

$$\begin{aligned} A &= \begin{bmatrix} 1 & 0 \\ \frac{1}{\alpha}b & I_{n-1} \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & L_\Delta L_\Delta^T \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{\alpha}b^T \\ 0 & I_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ \frac{1}{\alpha}b & I_{n-1} \end{bmatrix} \begin{bmatrix} \sqrt{\alpha} & 0 \\ 0 & L_\Delta \end{bmatrix} \begin{bmatrix} \sqrt{\alpha} & 0 \\ 0 & L_\Delta^T \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{\alpha}b^T \\ 0 & I_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{\alpha} & 0 \\ \frac{1}{\sqrt{\alpha}}b & L_\Delta \end{bmatrix} \begin{bmatrix} \sqrt{\alpha} & \frac{1}{\sqrt{\alpha}}b^T \\ 0 & L_\Delta^T \end{bmatrix} = LL^T, \end{aligned}$$

where $L = \begin{bmatrix} \sqrt{\alpha} & 0 \\ \frac{1}{\sqrt{\alpha}} b & L_\Delta \end{bmatrix}$ is lower-triangular with positive diagonal elements. This shows the existence.

To prove the uniqueness, suppose $A = LL^T = KK^T$, where K is also a Cholesky factor of A. Then $K^{-1}L = K^T \left(L^T\right)^{-1}$. Since the nonsingular lower triangular matrices form a group, $K^{-1}L$ is lower triangular and $K^T \left(L^T\right)^{-1}$ is upper triangular. Therefore $K^{-1}L = D$ for some diagonal matrix D. For real matrices, take this into $A = LL^T = KK^T$ and we obtain $KDD^TK^T = KK^T$, which gives $DD^T = I$, so $D = I$. The uniqueness also holds for complex matrices, but the proof is omitted here. □

Remark 2.2. *From the proof above, we also obtain a Cholesky factorization algorithm. To find the Cholesky factor of* $A \in \mathbb{S}^n_{++}$, *we only need to compute* $\Delta_A = A' - \frac{1}{\alpha} bb^T$, *where*

$$A = \begin{bmatrix} \alpha & b^T \\ b & A' \end{bmatrix},$$

which involves n^2 *flops (for computing* bb^T*), and then recursively find the Cholesky factorization of* $\Delta_A \in \mathbb{S}^{n-1}_{++}$. *So the total number of flops is approximately*

$$\sum_n n^2 = \frac{1}{3} n^3.$$

3 Matrix inversion lemma

There are many different versions of the matrix inversion lemma. The *Sherman-Morrison formula* gives that

$$\left(A + uv^T\right)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1 + v^TA^{-1}u},$$

where A is invertible matrix, $u, v \in \mathbb{R}^n$, $1 + v^TA^{-1}u \neq 0$ and $A + uv^T$. The *Sherman-Morrison-Woodbury formula* deals with the case of two matrices and gives that

$$\left(D + VV^T\right)^{-1} = D^{-1} - D^{-1}V\left(I + V^TD^{-1}V\right)^{-1} V^TD^{-1}.$$

However, the most generalized version is as follows.

Theorem 3.1. *Suppose* A, B, C, D *are matrices, vectors or numbers, such that* $A + BCD$ *exists. Suppose* $A, C, C^{-1} + DA^{-1}B$ *are invertible. Then* $A + BCD$ *is invertible and*

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}.$$

The proof involves nothing but fundamental operations of matrices.

4 Condition number

The condition number associated with a linear equation $Ax = b$ gives a bound on how inaccurate the solution x will be after approximation. Let e be the error in b. Assuming that A is a nonsingular matrix, the error in the solution $A^{-1}b$ is $A^{-1}e$. Then the ratio of the relative error in the solution to the relative error in b is.

$$\frac{\left\|A^{-1}e\right\| / \left\|A^{-1}b\right\|}{\|e\| / \|b\|} = \frac{\left\|A^{-1}e\right\|}{\|e\|} \cdot \frac{\|b\|}{\|A^{-1}b\|}.$$

The maximum value of the above when $e, b \neq 0$ is

$$\begin{aligned}
\max_{e,b\neq 0} \frac{\left\|A^{-1}e\right\|}{\|e\|} \cdot \frac{\|b\|}{\|A^{-1}b\|} &= \max_{e\neq 0} \frac{\left\|A^{-1}e\right\|}{\|e\|} \max_{b\neq 0} \frac{\|b\|}{\|A^{-1}b\|} \\
&= \max_{e\neq 0} \frac{\left\|A^{-1}e\right\|}{\|e\|} \max_{x\neq 0} \frac{\|Ax\|}{\|x\|} \\
&= \left|\left|\left|A^{-1}\right|\right|\right| \, |||A|||,
\end{aligned}$$

where $|||\cdot|||$ is the matrix norm associated with the vector norm $\|\cdot\|$. If $\|\cdot\|$ is the Euclidean norm, then

$$\left|\left|\left|A^{-1}\right|\right|\right| \, |||A||| = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}.$$

Definition 4.1 (Condition number of matrices)**.** *The* condition number *of a matrix* A *is defined as*

$$\kappa(A) = \begin{cases} \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}, & A \text{ is nonsingular,} \\ \infty, & A \text{ is singular.} \end{cases}$$

Definition 4.2 (Width of convex sets)**.** *The* width *of a convex set* $C \subseteq \mathbb{R}^n$ *in the direction* q*, where* $\|q\|_2 = 1$*, is defined as*

$$W(C, q) = \sup_{z\in C} q^T z - \inf_{z\in C} q^T z.$$

Definition 4.3 (Minimum and maximum width)**.** *The* minimum width *and* maximum width *of a convex set* $C \subseteq \mathbb{R}^n$ *are given by*

$$W_{\min} = \inf_{\|q\|_2=1} W(C, q), \quad W_{\max} = \sup_{\|q\|_2=1} W(C, q).$$

Definition 4.4 (Condition number of convex sets)**.** *The* condition number *of a convex set* $C \subseteq \mathbb{R}^n$ *is defined as*

$$\mathbf{cond}(C) = \frac{W_{\max}^2}{W_{\min}^2}.$$

Let $C_\alpha = \{x \mid f(x) \leqslant \alpha\}$ be the α-sublevel set. Suppose there exist constants m and M such that $mI \preceq \nabla^2 f(x) \preceq MI$. Then

$$p^* + \frac{M}{2} \|y - x^*\|_2^2 \geqslant f(y) \geqslant p^* + \frac{m}{2} \|y - x^*\|_2^2.$$

This shows that $B_{\text{inner}} \subseteq C_\alpha \subseteq B_{\text{upper}}$, where

$$B_{\text{inner}} = \left\{ y \mid \|y - x^*\|_2 \leqslant \sqrt{2\left(\alpha - p^*\right)/M} \right\},$$

$$B_{\text{outer}} = \left\{ y \mid \|y - x^*\|_2 \leqslant \sqrt{2\left(\alpha - p^*\right)/m} \right\}.$$

Therefore

$$\mathbf{cond}\left(C_\alpha\right) \leqslant \frac{M}{m}.$$

From the Taylor's formula we can also see that

$$f(y) \approx p^* + \frac{1}{2}\left(y - x^*\right)^T \nabla^2 f(x^*)(y - x^*),$$

so for α close to p^*,

$$C_\alpha \approx \left\{ y \mid (y - x^*)^T \nabla^2 f(x^*)(y - x^*) \leqslant 2(\alpha - p^*) \right\},$$

which is an ellipsoid with center x^*. Therefore

$$\lim_{\alpha \to p^*} \mathbf{cond}(C_\alpha) = \kappa\left(\nabla^2 f(x^*)\right).$$