Matrix Factorization

CE40282-1: Linear Algebra

Hamid R. Rabiee and Maryam Ramezani
Sharif University of Technology

Schur Triangularization

- Suppose $A \in \mathcal{M}_n(\mathbb{C})$. There exists a unitary matrix $U \in \mathcal{M}_n(\mathbb{C})$ and an upper triangular matrix $T \in \mathcal{M}_n(\mathbb{C})$ such that

$$A = UTU^*.$$

  - Proof?

- Compute a Schur triangularization of the following matrices:

  a) $A = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix}$ b) $B = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 3 & -3 & 4 \end{bmatrix}$

Schur Triangularization

- Important Note:

matrix

$$A = \begin{bmatrix} 1 & -2 \\ 1 & -1 \end{bmatrix}$$

has no real eigenvalues and thus no real Schur triangularization (since the diagonal entries of its triangularization T necessarily have the same eigenvalues as A). However, it does have a complex Schur triangularization: $A = UTU^*$, where

$$U = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2}(1+i) & 1+i \\ \sqrt{2} & -2 \end{bmatrix} \quad \text{and} \quad T = \frac{1}{\sqrt{2}} \begin{bmatrix} i\sqrt{2} & 3-i \\ 0 & -i\sqrt{2} \end{bmatrix}.$$

Determinant and Trace in Terms of Eigenvalues

- Let $A \in \mathcal{M}_n(\mathbb{C})$ have eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (listed according to algebraic multiplicity). Then

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n \qquad \text{and} \qquad \operatorname{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

Spectral Decomposition (complex)

- Suppose $A \in \mathcal{M}_n(\mathbb{C})$. Then there exists a unitary matrix $U \in \mathcal{M}_n(\mathbb{C})$ and diagonal matrix $D \in \mathcal{M}_n(\mathbb{C})$ such that

$$A = UDU^*$$

if and only if A is normal (i.e., $A^*A = AA^*$).

Suppose $A \in \mathcal{M}_n(\mathbb{C})$ is normal. If $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$ are eigenvectors of A corresponding to different eigenvalues then $\mathbf{v} \cdot \mathbf{w} = 0$.

Spectral Decomposition (real)

- Suppose $A \in \mathcal{M}_n(\mathbb{R})$. Then there exists a unitary matrix $U \in \mathcal{M}_n(\mathbb{R})$ and diagonal matrix $D \in \mathcal{M}_n(\mathbb{R})$ such that

$$A = UDU^T$$

if and only if A is symmetric (i.e., $A = A^T$).

LU-factorization

- Review: Gaussian Elimination, row operations are used to change the coefficient matrix to an upper triangular matrix.
- LU Decomposition is very useful when we have large matrices n x n and if we use gauss-jordan or the other methods, we can get errors.

DEFINITION 1 A factorization of a square matrix A as

$$A = LU \tag{1}$$

where L is lower triangular and U is upper triangular, is called an ***LU-decomposition*** (or ***LU-factorization***) of A.

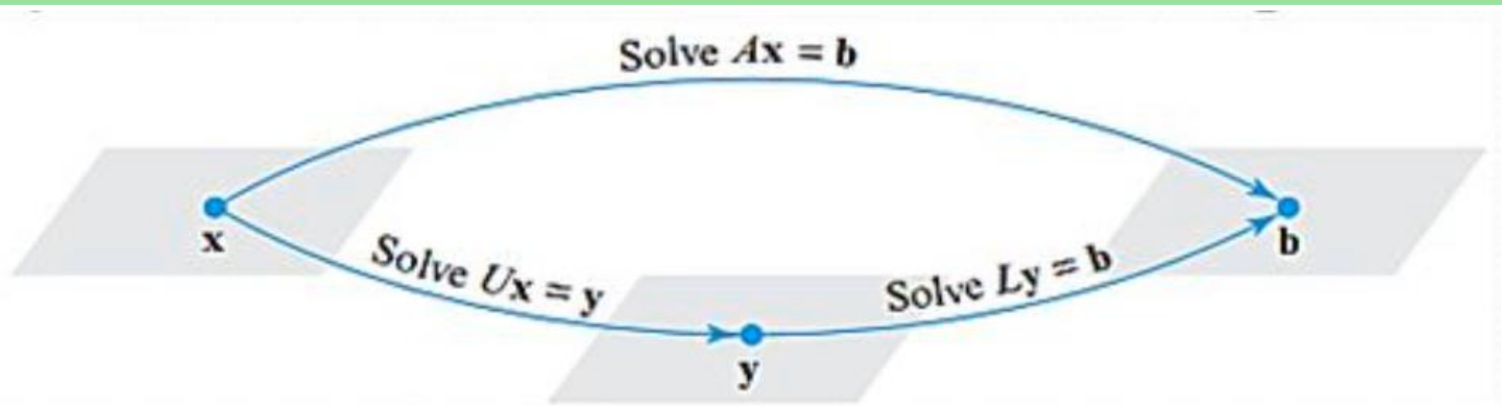
Method of LU Factorization

1) Rewrite the system $\boldsymbol{Ax = b}$ as $\boldsymbol{LUx = b}$

2) Define a new $\boldsymbol{n \times 1}$ matrix $\boldsymbol{y}$ by $\boldsymbol{Ux = y}$

3) Use $\boldsymbol{Ux = y}$ to rewrite $\boldsymbol{LUx = b}$ as $\boldsymbol{Ly = b}$ and solve the system for $\boldsymbol{y}$

4) Substitute $\boldsymbol{y}$ in $\boldsymbol{Ux = y}$ and solve for x



Constructing LU Factorization

1) Reduce A to a REF form U by Gaussian elinmination without row exchanges, keeping track of the multipliers used to introduce the leading $1s$ and multipliers used to introduce the zeros below the leading $1s$
2) In each position along the main diagonal of L place the reciprocal of the multiplier that introduced the leading 1 in that position in U
3) In each position below the main diagonal of L place negative of the multiplier used to introduce the zero in that position in U
4) Form the decomposition $A = LU$

Constructing LU Factorization

- Example

$$A = \begin{bmatrix} 6 & -2 & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix}$$

$$A = \begin{bmatrix} 6 & -2 & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix} \qquad \begin{bmatrix} \bullet & 0 & 0 \\ \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet \end{bmatrix}$$

$\bullet$ denotes an unknown entry of L.

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix} \begin{matrix} \longleftarrow \text{multiplier} = \frac{1}{6} \\ \\ \\ \end{matrix} \qquad \begin{bmatrix} 6 & 0 & 0 \\ \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 2 & 1 \\ 0 & 8 & 5 \end{bmatrix} \begin{matrix} \\ \longleftarrow \text{multiplier} = -9 \\ \longleftarrow \text{multiplier} = -3 \end{matrix} \qquad \begin{bmatrix} 6 & 0 & 0 \\ 9 & \bullet & 0 \\ 3 & \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 8 & 5 \end{bmatrix} \begin{matrix} \\ \longleftarrow \text{multiplier} = \frac{1}{2} \\ \\ \end{matrix} \qquad \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & \bullet & \bullet \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \longleftarrow \text{multiplier} = -8 \end{matrix} \qquad \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & 8 & \bullet \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \longleftarrow \text{multiplier} = 1 \end{matrix} \qquad L = \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & 8 & 1 \end{bmatrix}$$

No actual operation is performed here since there is already a leading 1 in the third row.

Thus, we have constructed the LU-decomposition

$$A = LU = \begin{bmatrix} 6 & 0 & 0 \\ 9 & 2 & 0 \\ 3 & 8 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{3} & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

LU Numerical notes

- The following operation counts apply to an $n \times n$ dense matrix A (with most entries nonzero) for n moderately large, say, $n \geq 30$.[1]

1. Computing an LU factorization of A takes about $2n^3/3$ flops (about the same as row reducing $[\,A \quad \mathbf{b}\,]$), whereas finding A^{-1} requires about $2n^3$ flops.

2. Solving $L\mathbf{y} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{y}$ requires about $2n^2$ flops, because any $n \times n$ triangular system can be solved in about n^2 flops.

3. Multiplication of $\mathbf{b}$ by A^{-1} also requires about $2n^2$ flops, but the result may not be as accurate as that obtained from L and U (because of roundoff error when computing both A^{-1} and $A^{-1}\mathbf{b}$).

4. If A is sparse (with mostly zero entries), then L and U may be sparse, too, whereas A^{-1} is likely to be dense. In this case, a solution of $A\mathbf{x} = \mathbf{b}$ with an LU factorization is *much* faster than using A^{-1}.

Some Notes

- Sometimes it is impossible to write a matrix in the form "lower triangular"×"upper triangular".
- An invertible matrix A has an LU decomposition provided that all upper left determinants are non-zero.

PLU Factorization

- if A is $n \times n$ and nonsingular, then it can be factored as

$$A = PLU$$

P is a permutation matrix, L is unit lower triangular, U is upper triangular

  - not unique; there may be several possible choices for P, L, U
  - interpretation: permute the rows of A and factor P^TA as $P^TA = LU$
  - also known as *Gaussian elimination with partial pivoting* (GEPP)

$$\begin{bmatrix} 0 & 5 & 5 \\ 2 & 9 & 0 \\ 6 & 8 & 8 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 0 & 15/19 & 1 \end{bmatrix} \begin{bmatrix} 6 & 8 & 8 \\ 0 & 19/3 & -8/3 \\ 0 & 0 & 135/19 \end{bmatrix}$$

- we will skip the details of calculating P, L, U

Cholesky Factorization

every positive definite matrix $A \in \mathbf{R}^{n\times n}$ can be factored as

$$A = R^T R$$

where R is upper triangular with positive diagonal elements

- complexity of computing R is $(1/3)n^3$ flops
- R is called the *Cholesky factor* of A
- can be interpreted as “square root” of a positive definite matrix
- gives a practical method for testing positive definiteness

Cholesky factorization algorithm

$$\begin{bmatrix} A_{11} & A_{1,2:n} \\ A_{2:n,1} & A_{2:n,2:n} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{1,2:n}^T & R_{2:n,2:n}^T \end{bmatrix} \begin{bmatrix} R_{11} & R_{1,2:n} \\ 0 & R_{2:n,2:n} \end{bmatrix}$$

$$= \begin{bmatrix} R_{11}^2 & R_{11}R_{1,2:n} \\ R_{11}R_{1,2:n}^T & R_{1,2:n}^T R_{1,2:n} + R_{2:n,2:n}^T R_{2:n,2:n} \end{bmatrix}$$

1. compute first row of R:

$$R_{11} = \sqrt{A_{11}}, \qquad R_{1,2:n} = \frac{1}{R_{11}} A_{1,2:n} \qquad \boxed{A_{11} > 0}$$

if A is positive definite

2. compute 2, 2 block $R_{2:n,2:n}$ from

$$A_{2:n,2:n} - R_{1,2:n}^T R_{1,2:n} = R_{2:n,2:n}^T R_{2:n,2:n} \boxed{= A_{2:n,2:n} - \frac{1}{A_{11}} A_{2:n,1} A_{2:n,1}^T}$$

this is a Cholesky factorization of order $n - 1$

Cholesky factorization algorithm

- Example

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 \\ R_{12} & R_{22} & 0 \\ R_{13} & R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

- first row of R

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & R_{22} & 0 \\ -1 & R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

- second row of R

$$\begin{bmatrix} 18 & 0 \\ 0 & 11 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} \begin{bmatrix} 3 & -1 \end{bmatrix} = \begin{bmatrix} R_{22} & 0 \\ R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix}$$

$$\begin{bmatrix} 9 & 3 \\ 3 & 10 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 1 & R_{33} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & R_{33} \end{bmatrix}$$

- third column of R: $10 - 1 = R_{33}^2$, *i.e.*, $R_{33} = 3$

Rank and matrix factorizations

Let $\mathcal{B} = \{b_1, \dots, b_r\} \subset \mathbb{R}^m$ with $r = \operatorname{rank}(A)$ be basis of $\operatorname{range}(A)$. Then each of the columns of $A = [a_1, a_2, \dots, a_n]$ can be expressed as linear combination of $\mathcal{B}$:

$$a_i = b_1 c_{i1} + b_2 c_{i2} + \cdots + b_r c_{ir} = [b_1, \dots, b_r] \begin{bmatrix} c_{i1} \\ \vdots \\ c_{ir} \end{bmatrix},$$

for some coefficients $c_{ij} \in \mathbb{R}$ with $i = 1, \dots, n$, $j = 1, \dots, r$.

Stacking these relations column by column $\rightsquigarrow$

$$[a_1, \dots, a_n] = [b_1, \dots, b_r] \begin{bmatrix} c_{11} & \cdots & c_{n1} \\ \vdots & & \vdots \\ c_{1r} & \cdots & c_{nr} \end{bmatrix}$$

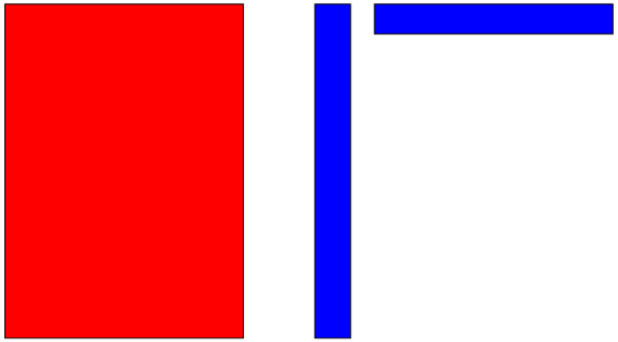
Rank and matrix factorizations

Lemma. A matrix $A \in \mathbb{R}^{m \times n}$ of rank r admits a factorization of the form

$$A = BC^T, \quad B \in \mathbb{R}^{m \times r}, \quad C \in \mathbb{R}^{n \times r}.$$

We say that A has low rank if $\text{rank}(A) \ll m, n$.

Illustration of low-rank factorization:



	A	BC^T
#entries	mn	$mr + nr$

- Generically (and in most applications), A has full rank, that is, $\text{rank}(A) = \min\{m, n\}$.
- Aim instead at approximating A by a low-rank matrix.