Linear Algebra

Matrix Factorization

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee rabiee@sharif.edu
Maryam Ramezani maryam.ramezani@sharif.edu

Table of contents

01

Introduction

Introduction

Matrix Multiplication as Composition of Transformations

$$= \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.6 \end{bmatrix} \begin{bmatrix} 2.5 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}$$

reflect around y-axis

scale y axis by 0.6

scale x-axis by 2.5

02

QR Decomposition

QR Decomposition (QU) (Factorization)

Theorem

if $A \epsilon \mathbb{R}^{m\times n}$ has linearly independent columns then it can be factored as

$$A = QR$$

Q-factor

- ❑ Q is $m \times n$ with orthonormal columns $(Q^T Q = I)$
- ❑ If A is square $(m = n)$, then Q is orthogonal $(Q^T Q = QQ^T = I)$

R-factor

- ❑ R is n$\times$ n, upper triangular, with nonzero diagonal elements
- ❑ R is nonsingular (diagonal elements are nonzero)

QR Decomposition

Example

$$A = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix}$$

$$q_1 = \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, q_2 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, q_3 = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \|\tilde{q}_1\| = 2, \|\tilde{q}_2\| = 2, \|\tilde{q}_3\| = 4$$

- QR :

$$\begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 2 & 8 \\ 0 & 0 & 4 \end{bmatrix}$$

QR Decomposition

- A QR decomposition can be created for any matrix — it need not be square and it need not have full rank.
- Every matrix has a QR-decomposition, though R may not always be invertible.

03

Schur Triangularization

Schur Triangularization

Theorem

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

$$A = UTU^*.$$

$$A = U \begin{bmatrix} \lambda_1 & x & \cdots & x \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & x \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} U^*.$$

Schur triangularization are highly non-unique

Example

Compute a Schur triangularization of the following matrices:

$a)\quad A = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix}$

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

Schur Triangularization

Important Note

Matrix

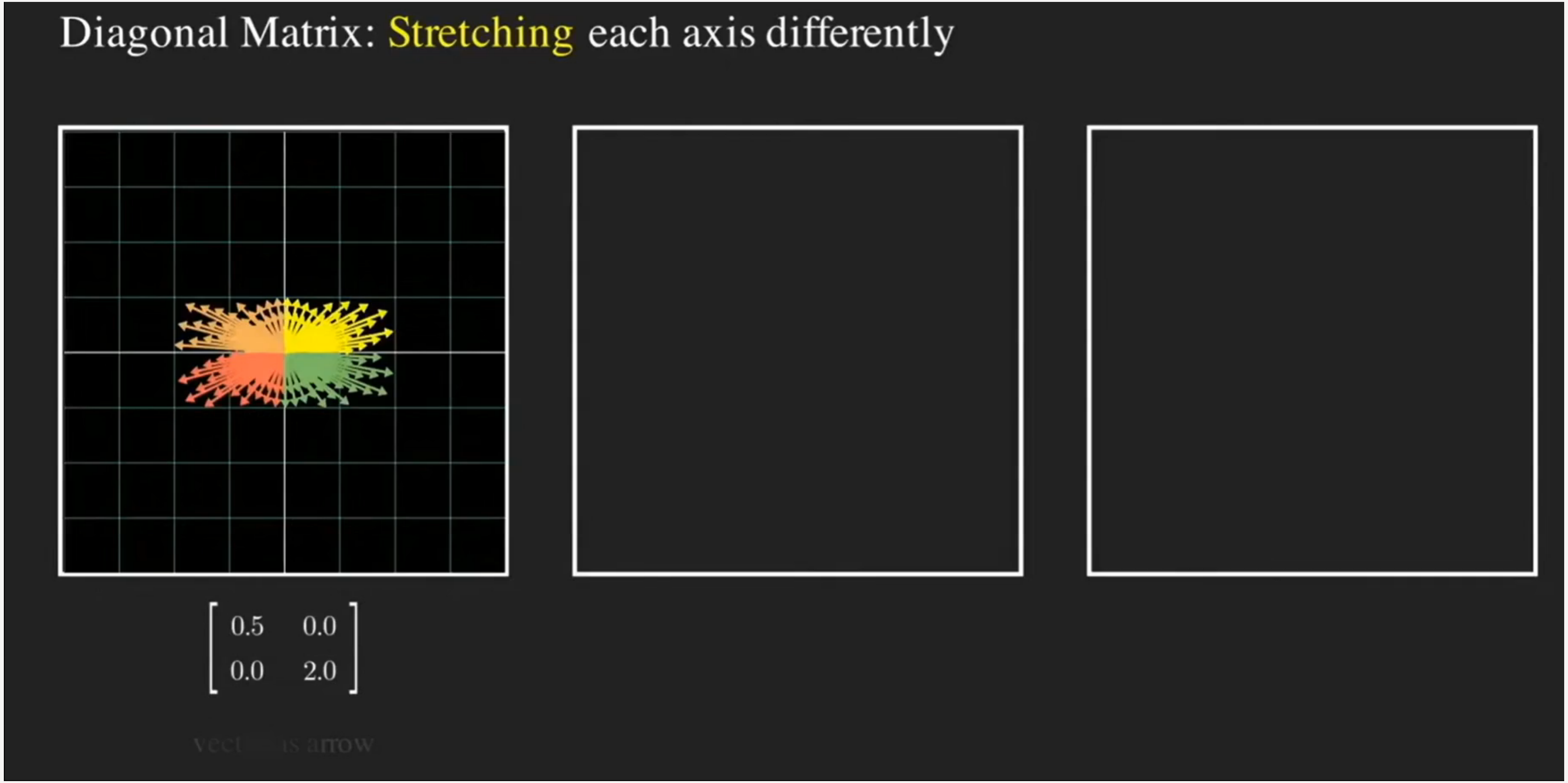
$$A = \begin{bmatrix} 1 & -2 \\ 1 & -1 \end{bmatrix}$$

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

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

Review


Diagonal Matrix: Stretching each axis differently
0.5 0.0
0.0 2.0

04

Spectral Decomposition (complex and real)

Spectral Decomposition (complex and real)

Theorem

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

$$A = UDU^*.$$

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

Theorem

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

$$A = UDU^T.$$

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

Spectral Decomposition (complex)

$$[T^*T]_{1,1} = \left[\begin{bmatrix} \overline{t_{1,1}} & 0 & \cdots & 0 \\ \overline{t_{1,2}} & \overline{t_{2,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \overline{t_{1,n}} & \overline{t_{2,n}} & \cdots & \overline{t_{n,n}} \end{bmatrix} \begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,n} \\ 0 & t_{2,2} & \cdots & t_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{n,n} \end{bmatrix}\right]_{1,1}$$

$$= |t_{1,1}|^2,$$

and

$$[TT^*]_{1,1} = \left[\begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,n} \\ 0 & t_{2,2} & \cdots & t_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{n,n} \end{bmatrix} \begin{bmatrix} \overline{t_{1,1}} & 0 & \cdots & 0 \\ \overline{t_{1,2}} & \overline{t_{2,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \overline{t_{1,n}} & \overline{t_{2,n}} & \cdots & \overline{t_{n,n}} \end{bmatrix}\right]_{1,1}$$

$$= |t_{1,1}|^2 + |t_{1,2}|^2 + \cdots + |t_{1,n}|^2.$$

$$[T^*T]_{2,2} = \left[\begin{bmatrix} \overline{t_{1,1}} & 0 & \cdots & 0 \\ 0 & \overline{t_{2,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \overline{t_{2,n}} & \cdots & \overline{t_{n,n}} \end{bmatrix} \begin{bmatrix} t_{1,1} & 0 & \cdots & 0 \\ 0 & t_{2,2} & \cdots & t_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{n,n} \end{bmatrix}\right]_{2,2}$$

$$= |t_{2,2}|^2,$$

$$[TT^*]_{2,2} = \left[\begin{bmatrix} t_{1,1} & 0 & \cdots & 0 \\ 0 & t_{2,2} & \cdots & t_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{n,n} \end{bmatrix} \begin{bmatrix} \overline{t_{1,1}} & 0 & \cdots & 0 \\ 0 & \overline{t_{2,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \overline{t_{2,n}} & \cdots & \overline{t_{n,n}} \end{bmatrix}\right]_{2,2}$$

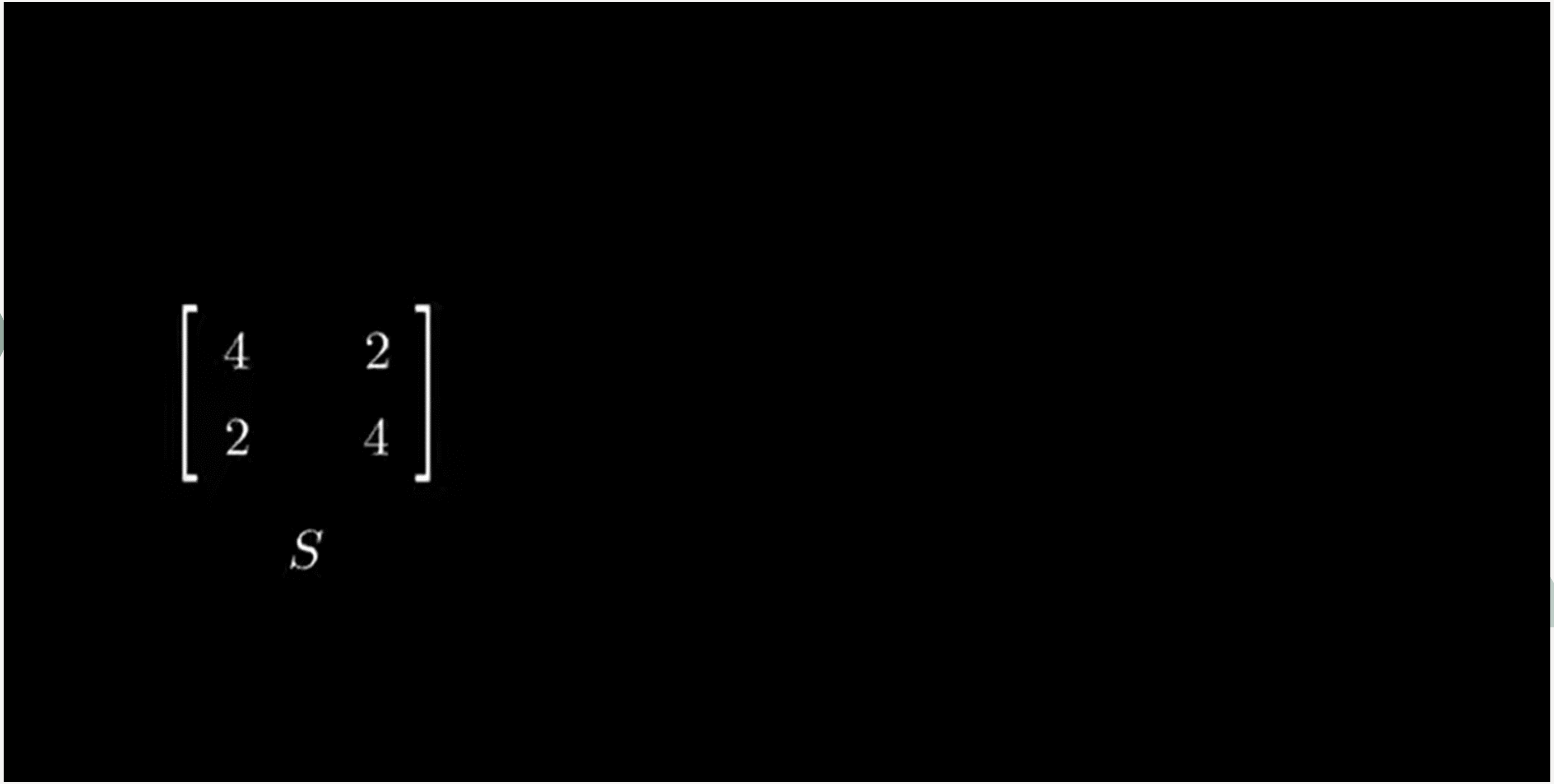
$$= |t_{2,2}|^2 + |t_{2,3}|^2 + \cdots + |t_{2,n}|^2,$$

Spectral Decomposition (real)

$$S = Q \Lambda Q^T$$

$$\begin{bmatrix} & & \\ & \underset{\text{3 by 3}}{S} & \\ & & \end{bmatrix} = \begin{bmatrix} | & | & | \\ \vec{e_1} & \vec{e_2} & \vec{e_3} \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} \begin{bmatrix} | & | & | \\ \vec{e_1} & \vec{e_2} & \vec{e_3} \\ | & | & | \end{bmatrix}^T$$

Visualization of Spectral Decomposition



Important Note

- Spectral Decomposition is nice and pretty, but with loss of generality:

  Real Field: For square and symmetric matrices!

  Complex Field: For square and normal matrices!

  For General?? SVD!!!

Think with spectral decomposition

Normal Matrices have Orthogonal Eigenspaces

Theorem

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

05

LU Factorization
PLU Factorization

LU-factorization for square matrix

- 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 $n \times n$ and if we use gauss-jordan or the other methods, we can get

Definition

A factorization of a square matrix A as

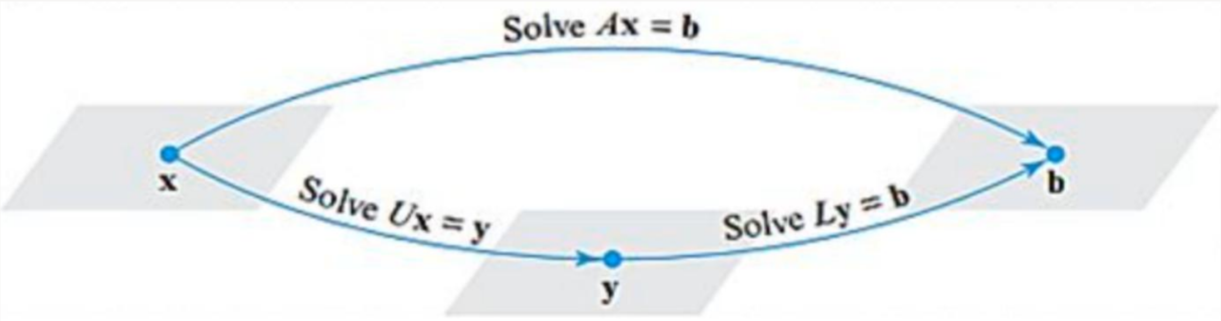
$$A = LU$$

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

Method of LU Factorization

Important

1) Rewrite the system $\boldsymbol{Ax} = \boldsymbol{b}$ as $\boldsymbol{LUx} = \boldsymbol{b}$
2) Define a new $\boldsymbol{n} \times \mathbf{1}$ matrix $\boldsymbol{y}$ by $\boldsymbol{Ux} = \boldsymbol{y}$
3) Use $\boldsymbol{Ux} = \boldsymbol{y}$ to rewrite $\boldsymbol{LUx} = \boldsymbol{b}$ as $\boldsymbol{Ly} = \boldsymbol{b}$ and solve the system for $\boldsymbol{y}$
4) Substitute $\boldsymbol{y}$ in $\boldsymbol{Ux} = \boldsymbol{y}$ and solve for $\boldsymbol{x}$.



Constructing LU Factorization

Important

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

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.

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

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

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

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

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

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

Thus, we have constructed $LU-$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-factorization for non-square matrix

$$\begin{matrix} 3 & 4 \\ -5 & 3 \\ 5 & 4 \end{matrix}$$

$$U = \begin{matrix} 3 & 4 \\ 0 & \frac{29}{3} \\ 0 & 0 \end{matrix}, \quad L = \begin{matrix} 1 & 0 & 0 \\ -\frac{5}{3} & 1 & 0 \\ \frac{5}{3} & \frac{8}{29} & 1 \end{matrix}$$

LU Numerical notes

Note

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. 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

Note

- 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 Why??

If A is invertible, then it admits an LU (or LDU) factorization if and only if all its leading principal minors are non-zero.

If A is a singular matrix of rank k, then it admits an LU factorization if the first k leading principal minors are non-zero

Some Notes

In general, any square matrix $A_{n\times n}$ could have one of the following:

1. a unique LU factorization (as mentioned above);
2. infinitely many LU factorizations if two or more of any first (n−1) columns are linearly dependent or any of the first (n−1) columns are 0;
3. no LU factorization if the first (n−1) columns are non-zero and linearly independent and at least one leading principal minor is zero.

In Case 3, one can approximate an LU factorization by changing a diagonal entry a_{jj} to $a_{jj} \pm \varepsilon$ to avoid a zero leading principal minor.[10]

PLU Factorization

Theorem

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 $P^T A$ as $P^T A = LU$
- also known as Gaussian elimination with partial pivoting (GEPP)
- Is it unique??

Example

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

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

06

Cholesky Factorization

Cholesky Factorization

Important

Every positive definite matrix $A \epsilon \mathbb{R}^{n \times n}$ can be factored as

$$A = \mathbb{R}^T \mathbb{R}$$

where $\mathbb{R}$ is upper triangular with positive diagonal elements

- ❑ complexity of computing $\mathbb{R}$ is $(1/3)n^3$ flops
- ❑ $\mathbb{R}$ 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

Example

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

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

1. compute first row of R:

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

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

Example

- Let $B = \{b_1, \dots, b_r\} \subset \mathbb{R}^m$ with $r = \text{rank}(A)$ be basis of range(A). Then each of the columns of $A = [a_1, a_2, \dots, a_n]$ can be expressed as linear combination of 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 $c_{ij} \epsilon\, \mathbb{R}$ with $i = 1, \dots, n, j = 1, \dots, r$.

Stacking these relations column by column →

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

Rank and matrix factorizations

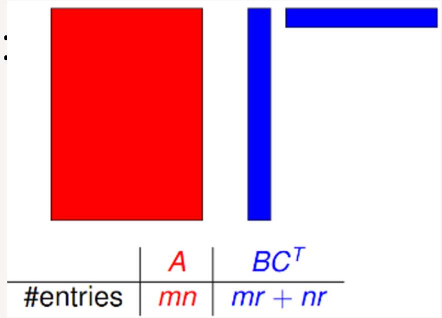
Lemma

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

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

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

Illustration of low-rank factorization:



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