Matrix Factorization

Linear Algebra

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee rabiee@sharif.edu

Maryam Ramezani maryam.ramezani@sharif.edu

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

Orthogonal Matrix

Note

- Columns of A are orthonormal $\leftrightarrow A^T A = I$

- Square matrix with orthonormal columns is a **orthogonal matrix**
  - Columns and rows are orthonormal vectors
  - $A^T A = AA^T = I$
  - is necessarily invertible with inverse $A^T = A^{-1}$

Orthogonal Matrix

Example

- Identity matrix $I^T I = I$

- Rotation matrix

$$R^T R = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} =$$

$$\begin{bmatrix} cos^2\theta + sin^2\theta & -cos\theta sin\theta + sin\theta cos\theta \\ -sin\theta cos\theta + cos\theta sin\theta & sin^2\theta + cos^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Orthogonal Matrix

Example

❑Reflection matrix

$$\begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}^T \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix} =$$

$$\begin{bmatrix} \cos^2(2\theta) + \sin^2(2\theta) & cos\,(2\theta)\sin(2\theta) - \sin(2\theta)\cos(2\theta) \\ \sin(2\theta)\cos(2\theta) - \cos(2\theta)\sin(2\theta) & \sin^2(2\theta) + \cos^2(2\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Lemma

All orthogonal matrices can be expressed as Rotation or Reflection

Note

If $A \in \mathbb{R}^{m\times n}$ has orthonormal columns, then the linear function $f(x) = Ax$

- Preserves inner product:

$$(Ax)^T(Ay) = x^T y$$

- Preserves norm:

$$||Ax|| = ||x||$$

- Preserves distances:

$$||Ax - Ay|| = ||x - y||$$

- Preserves angels:

$$\angle(Ax, Ay) = \arccos\left(\frac{(Ax)^T(Ay)}{||Ax||\,||Ay||}\right) = \arccos\left(\frac{x^T y}{||x||\,||y||}\right) = \angle(x, y)$$

This is a mapping with preserving properties of input

Gram–Schmidt in matrix notation

Important

Run Gram–Schmidt on columns $a_1, \dots, a_k$ of $n \times k$ matrix A:

$$\tilde{q}_1 = a_1, \qquad q_1 = \frac{\tilde{q}_1}{\|\tilde{q}_1\|}$$

$$\Longrightarrow a_1 = \|\tilde{q}_1\| q_1$$

$$\tilde{q}_2 = a_2 - (q_1^T a_2) q_1, \qquad q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|}$$

$$\Rightarrow a_2 = (q_1^T a_2) q_1 + \|\tilde{q}_2\| q_2$$

$$\vdots$$

$$\tilde{q}_i = a_i - (q_1^T a_i) q_1 - \cdots - \left(q_{i-1}^T a_i\right) q_{i-1}, \qquad q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$$

$$a_i = (q_1^T a_i) q_1 + \cdots + \left(q_{i-1}^T a_i\right) q_{i-1} + \|\tilde{q}_i\| q_i$$

Matrix–Matrix Multiplication

As a set of matrix–vector products.

$$C = AB = A\begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab_1 & Ab_2 & \cdots & Ab_p \\ | & | & & | \end{bmatrix}$$

Here the *i*th column of C is given by the matrix–vector product with the vector on the right, $c_i = Ab_i$. These matrix–vector products can in turn be interpreted using both viewpoints given in the previous subsection.

Matrix–Vector Multiplication

If we write A by columns, then we have:

$$y = Ax = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [a_1]x_1 + [a_2]x_2 + \cdots + [a_n]x_n \,.$$

- y is a linear combination of the columns A.

Gram–Schmidt in matrix notation

Important

$$a_1 = \|\tilde{q}_1\| q_1$$

$$a_2 = (q_1^T a_2) q_1 + \|\tilde{q}_2\| q_2$$
$$\vdots$$
$$a_k = (q_1^T a_k) q_1 + \cdots + \left(q_{k-1}^T a_k\right) q_{k-1} + \|\tilde{q}_k\| q_k$$

$$[a_1 \quad a_2 \quad \dots \quad a_k] = [q_1 \quad q_2 \quad \dots \quad q_k] \begin{bmatrix} \|\tilde{q}_1\| & q_1^T a_2 & \cdots & q_1^T a_k \\ 0 & \|\tilde{q}_2\| & \cdots & q_2^T a_k \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q_{k-1}^T a_k \\ 0 & 0 & \cdots & \|\tilde{q}_k\| \end{bmatrix}$$

$$A_{n\times k} = Q_{n\times k} \times R_{k\times k}$$

Gram–Schmidt in matrix notation

Important

1. Run Gram–Schmidt on columns $a_1, \dots, a_k$ of $\mathrm{n} \times k$ matrix A
2. If columns are linearly independent, get orthonormal $q_1, \dots, q_k$
3. Define $\mathrm{n} \times k$ matrix Q with columns $q_1, \dots, q_k$
4. $Q^T Q = I$
5. From Gram–Schmidt algorithm

$$a_i = (q_1^T a_i) q_1 + \cdots + \left(q_{i-1}^T a_i\right) q_{i-1} + + \|\tilde{q}_i\| q_i$$
$$= R_{1i} q_1 + \cdots + R_{ii} q_i$$

With $R_{1j} = q_i^T a_j$ for $i < j$ and $R_{ii} = \|\tilde{q}_i\|$

6. Defining $R_{ij} = 0$ for $i > j$ we have $A = QR$
7. R is upper triangular, with positive diagonal entries

Definition

A factorization of a matrix A as $A = QR$ where Factors satisfy $Q^TQ = I$, R upper triangular with positive diagonal entries, is called a **QR factorization** of A.

Suppose A is a square matrix with linearly independent columns. Then there exist unique matrices Q and R such that Q is unitary, R is upper triangular with only positive numbers on its diagonal, and

$$A = QR.$$

$R_{jk} = < a_k, q_j >$

Note

The QR factorization of a matrix :

- ❑ Can be computed using Gram–Schmidt algorithm (or some variations)
- ❑ Has a huge number of uses, which we'll see soon

Important

To find QR decomposition:

- Q: Use Gram-Schmidt to find orthonormal basis for column space of A
- Let $R = Q^T A$
- OR: $R_{jk} =< a_k, q_j >$

- If A is a square matrix, then Q is square and orthonormal (orthogonal)

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

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 :

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

$$A_{4\times 6} = [\underline{a_1 \quad a_2} \quad a_3 \quad \underline{a_4} \quad a_5 \quad a_6]$$

Linear Independent

$$\begin{cases} a_1 = a_{11}q_1 \\ a_2 = a_{21}q_1 + a_{22}q_2 \\ a_3 = a_{31}q_1 + a_{32}q_2 \\ a_4 = a_{41}q_1 + a_{42}q_2 + a_{43}q_3 \\ a_5 = a_{51}q_1 + a_{52}q_2 + a_{53}q_3 \\ a_6 = a_{61}q_1 + a_{62}q_2 + a_{63}q_3 \end{cases}$$

Block upper triangular matrix

$$[a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6] = [q_1 \quad q_2 \quad q_3]\begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} & a_{51} & a_{61} \\ 0 & a_{22} & a_{32} & a_{42} & a_{52} & a_{62} \\ 0 & 0 & 0 & a_{43} & a_{53} & a_{63} \end{bmatrix}$$

$$A_{4\times 6} = Q_{4\times \boxed{3}} \times R_{3\times 6}$$

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.

Note

suppose A is square and invertible :

- So its columns are linearly independent
- So Gram-Schmidt gives QR factorization
  - $A = QR$
  - Q is orthogonal $Q^T Q = I$
  - R is upper triangular with positive diagonal entries, hence invertible
- So we have

$$A^{-1} = (QR)^{-1} = R^{-1}Q^{-1} = R^{-1}Q^T$$

Algorithm: Computing Matrix Inverse

Input: $A_{n\times n}$ invertible

Output: $A^{-1}_{n\times n}$

Find QR factorization $A = QR$

$\begin{bmatrix} \bar{q}_1 & \cdots & \bar{q}_n \end{bmatrix} = Q^T$

for $i = 1, \dots, n$ **do**

  Solve $Rx_i = \bar{q}_i$ using back substituition

end

$A^{-1} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$

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

$$A = UTU^*.$$

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

Schur triangularization are highly non-unique

Example

Compute a Schur triangularization of the following matrices:

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

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

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

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

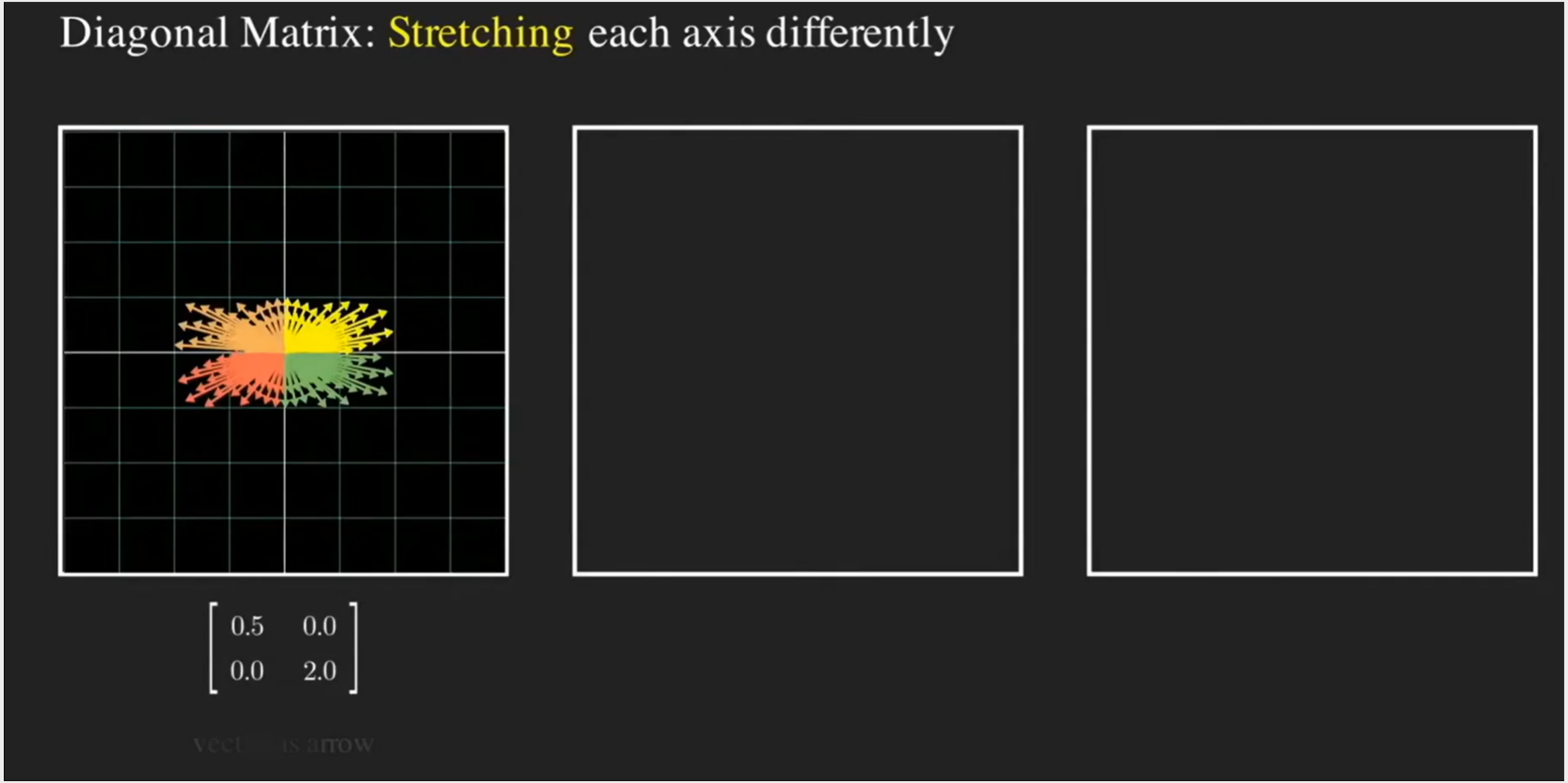
Determinant and Trace in Terms of Eigenvalues

Important

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

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

Review


Diagonal Matrix: Stretching each axis differently
0.5 0.0
0.0 2.0

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., $A = A^T$).

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

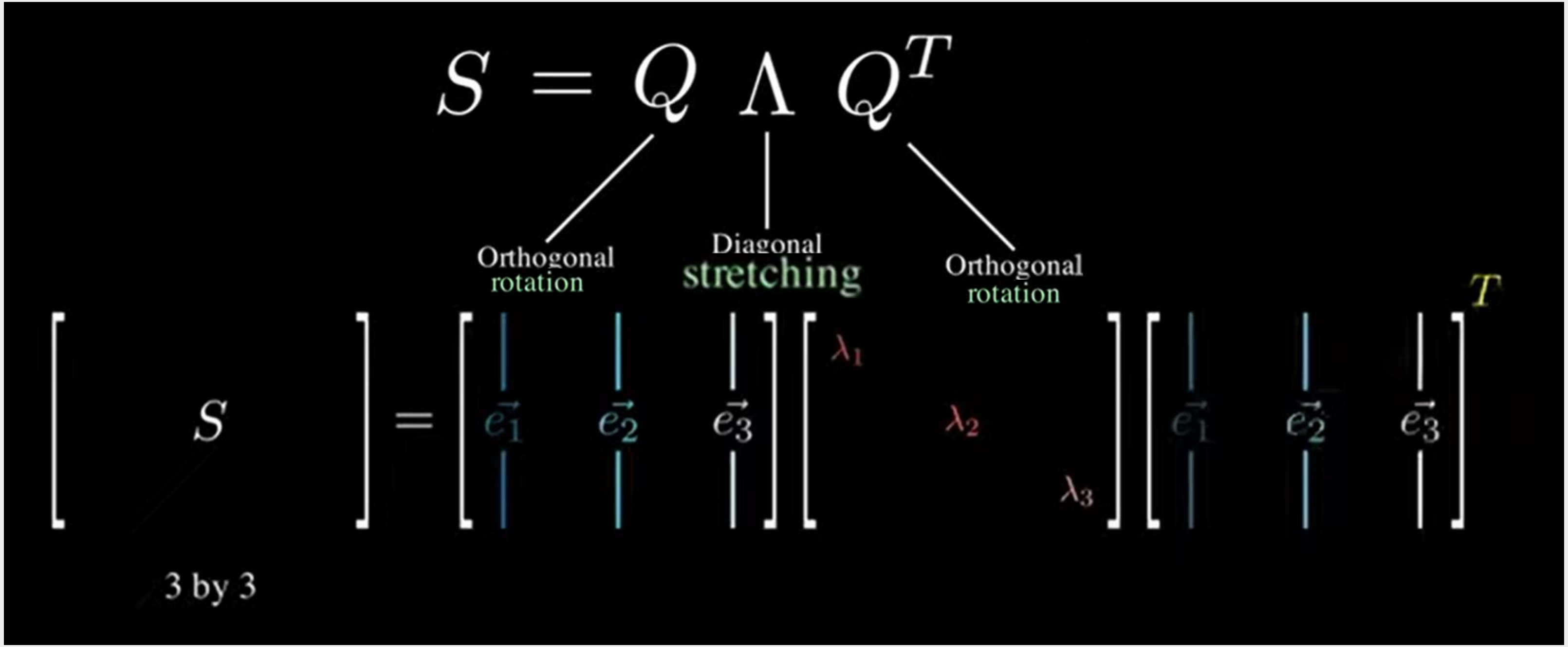
$$[T^*T]_{2,2} = \left[\begin{bmatrix} \overline{t_{1,1}} & 0 & \cdots & 0 \\ 0 & \overline{t_{2,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \overline{t_{2,n}} & \cdots & \overline{t_{n,n}} \end{bmatrix} \begin{bmatrix} t_{1,1} & 0 & \cdots & 0 \\ 0 & t_{2,2} & \cdots & t_{2,n} \\ \vdots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

and

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

$$[TT^*]_{2,2} = \left[\begin{bmatrix} t_{1,1} & 0 & \cdots & 0 \\ 0 & t_{2,2} & \cdots & t_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{n,n} \end{bmatrix} \begin{bmatrix} \overline{t_{1,1}} & 0 & \cdots & 0 \\ 0 & \overline{t_{2,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \overline{t_{2,n}} & \cdots & \overline{t_{n,n}} \end{bmatrix}\right]_{2,2}$$
$$= |t_{2,2}|^2 + |t_{2,3}|^2 + \cdots + |t_{2,n}|^2,$$

Spectral Decomposition (Real)



Visualization of Spectral Decomposition

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

$$S$$

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!!!

Normal Matrices have Orthogonal Eigenspaces

Theorem

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

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 errors.

Definition

A factorization of a square matrix A as

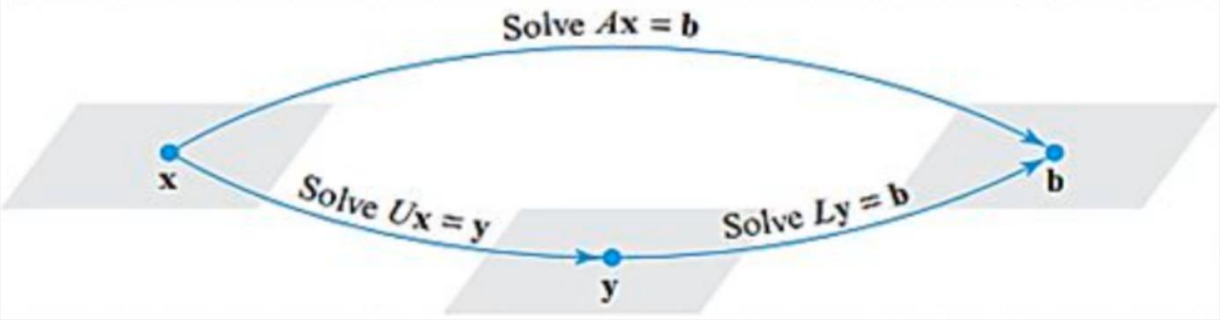
$$A = LU$$

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

Method of LU Factorization

Important

1) Rewrite the system $\boldsymbol{Ax} = \boldsymbol{b}$ as $\boldsymbol{LUx} = \boldsymbol{b}$
2) Define a new $\boldsymbol{n} \times \boldsymbol{1}$ 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 $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$

Example

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

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

❑ denotes an unknown entry of L.

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

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

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

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

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

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

We show by induction that every $n \times n$ matrix A with nonsingular leading principal minors has a factorization $A = LU$ where L is strictly lower triangular, U is upper triangular, and L and U are both nonsingular. (This statement, as you show, is an if-and-only-if.)

The 1×1 base case is just factoring $a = 1 \cdot a$. To induct, write your $n \times n$ matrix A as a leading principal $(n-1) \times (n-1)$ matrix A' and some leftover entries:

$$A = \left[\begin{array}{c|c} A' & \vec{b} \\ \hline \vec{c}^{\mathsf{T}} & d \end{array}\right].$$

By the inductive hypothesis (since all leading principal minors of A' are also leading principal minors of A), A' has an LU factorization as $A' = L'U'$ with nonsingular L', U'. We want to use this to make the factorization

$$\left[\begin{array}{c|c} A' & \vec{b} \\ \hline \vec{c}^{\mathsf{T}} & d \end{array}\right] = \left[\begin{array}{c|c} L' & \vec{0} \\ \hline \vec{x}^{\mathsf{T}} & 1 \end{array}\right] \left[\begin{array}{c|c} U' & \vec{y} \\ \hline \vec{0}^{\mathsf{T}} & z \end{array}\right]$$

work, by picking appropriate $\vec{x}$, $\vec{y}$, and z.

By doing the block multiplication, we get four equations.

- We have $A' = L'U' + \vec{0}\vec{0}^{\mathsf{T}}$, which we know is true, so that's done.
- We have $\vec{b} = L'\vec{y} + \vec{0}z$, so we want to set $\vec{y} = L'^{-1}\vec{b}$. Fortunately that's possible since L' is invertible.
- We have $\vec{c}^{\mathsf{T}} = \vec{x}^{\mathsf{T}}U' + \vec{0}^{\mathsf{T}}$, so we want to set $\vec{x}^{\mathsf{T}} = \vec{c}^{\mathsf{T}}U'^{-1}$. This is possible since U' is also invertible.
- We have $d = \vec{x}^{\mathsf{T}}\vec{y} + z$, so we want to set $z = d - \vec{x}^{\mathsf{T}}\vec{y}$.

For future inductive steps, we also want to know that the resulting matrices L and U are nonsingular. This is immediate for L since its diagonal is 1; for U, it's not obvious how to check that the value of z we get is nonzero. But once we have $A = LU$ where A and L are nonsingular, we know that $U = L^{-1}A$ is nonsingular.

There are also LU factorizations out there for which U is singular (some of the diagonal entries of U are zero). For these, there is not an if-and-only-if condition this nice.

You can see from the above proof, for instance, that if A is possibly singular but all of its proper leading principal minors are still nonsingular, then we get a factorization $A = LU$ in which the bottom right entry is possibly 0. (This is because arguing $z \neq 0$ is the only place where we needed A to be nonsingular.)

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

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:

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

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

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

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

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

Example

- Let $B = \{b_1, \dots, b_r\} \subset \mathbb{R}^m$ with $r = \operatorname{rank}(A)$ 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 →

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

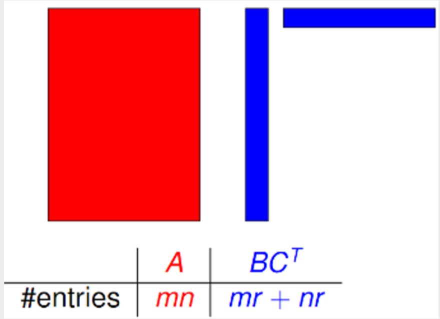
Lemma

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.