Matrix Decomposition

steve canves

yqykrhf@163.com

1 LU decomposition

1.1 Introduction

Take 3×3 matrix for example. Assume **there's no operation that switch one row with another row**, matrix will be transformed into upper triangular matrix U after **gauss elimination**.

$$E_{32}E_{31}E_{21}A = U$$

E_{ij} is a elimination matrix to eliminate the element located in row i, column thus,

$$A = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}U = LU$$

Note: elimination matrix is lower triangular matrix. and the result after inverse or product between two lower triangular matrix is still lower triangular matrix.

why use $A = LU$ instead of $E_{32}E_{31}E_{21}A = U$? here's a example.

$$E_{32} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 1 \end{pmatrix}, E_{31} = I_3, E_{21} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$$E_{32}E_{31}E_{21} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ \boxed{10} & -5 & 1 \end{pmatrix} = E$$

$$E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 5 & 1 \end{pmatrix} = L$$

L only contains elimination information, however $e_{32}e_{31}e_{21}$ contains other information.

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \xrightarrow{r_2 \boxed{-} r_1 \times \boxed{(-\frac{1}{2})}} \begin{pmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & -1 & 2 \end{pmatrix} \xrightarrow{r_3 \boxed{-} r_2 \times \boxed{(-\frac{2}{3})}} \begin{pmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & 0 & \frac{4}{3} \end{pmatrix} = U$$

$$\Longrightarrow A = \begin{pmatrix} 1 & 0 & 0 \\ \boxed{-\frac{1}{2}} & 1 & 0 \\ 0 & \boxed{-\frac{2}{3}} & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & 0 & \frac{4}{3} \end{pmatrix} = LU.$$

sometimes, u is written as $\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \begin{pmatrix} 1 & * & * \\ & \ddots & * \\ & & 1 \end{pmatrix}$ in the above case,

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & \frac{4}{3} \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{pmatrix} = LDU$$

1.2 Solve Equations

If $A = LU$, solving the equation $ax = b$ is equivalent to solve the two equations as follows:

$$Lc = b$$

$$Ub = c$$

triangular matrix is easy to solve. and it is easier to solve two equations than to solve the equation $ax = b$ directly, excluding the time of lu decomposition. actually lu decomposition and gauss elimination take the same time complexity to solve the single equation.

However, if there are many equation like $Ax = b_1, Ax = b_2, Ax = b_3, ...Ax = b_n$ to be solved, gauss elimination take n times to solve independently. take use of LU decomposition, take one time lu decomposition, the left n-1 equations need no do lu decomposition repeatly. and from physical aspect, A is concerned about the system itself, b is from sensor observation. with LU decomposition of A, it's fast to solve the x to different observations.

Time Complexity

Three steps:

1. LU decomposition: $\frac{1}{3}n^3$ addition + $\frac{1}{3}n^3$ multiplication operation
2. solve $Lc = b$: n^2
3. solve $Ux = c$: n^2

1.3 Existence and Uniqueness

Not every square matrix admit lu decomposition, even it's invertible.

eg: $A = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ l_{21} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix} = LU$

$u_{11} = 0, u_{12} = 1, 2 = l_{21} \cdot 0?$

Existence Theorem:

If square matrix A is invertible, then it admits LU (or LDU) factorization if and only if all its leading principal minors are nonzero.

proof:

Using mathematical induction for $a_{n\times n}$

for n=1, l=1, u=$a_{11} \neq 0$

Let us assume $A_{n\times n} = LU$ is ture for n=k.
when n=k+1,

$$A = \begin{pmatrix} A_k & \beta \\ \alpha^T & a_{nn} \end{pmatrix}$$

$$\begin{pmatrix} I_k & 0 \\ -\alpha^T a_k^{-1} & 1 \end{pmatrix} a = \begin{pmatrix} A_k & \beta \\ 0 & a_{nn} - \alpha^T a_k^{-1}\beta \end{pmatrix}$$

$$A = \begin{pmatrix} I_k & 0 \\ \alpha^T a_k^{-1} & 1 \end{pmatrix} \begin{pmatrix} A_k & \beta \\ 0 & a_{nn} - \alpha^T a_k^{-1}\beta \end{pmatrix}$$

Uniqueness Theorem:
If square matrix A is invertible, A=LU, and $l_{ii} = 1, u_{ii} \neq 0$, then the factorization is unique
proof:
Assume, a has two lu decompositions L_1U_1, L_2U_2

$$L_1U_1 = L_2U_2$$

$$L_2^{-1}L_1 = U_2U_1^{-1}$$

The product of two lower(upper) triangular matrix is still lower(upper) triangular. and the diagonal elements of L_1, L_2 are one, so the diagonnal elements of $L_2^{-1}L_1$ are one. at the same time, $L_2^{-1}L_1 = U_2U_1^{-1}$ is upper triangular matrix. the conclusion is $L_2^{-1}L_1 = I$

1.4 PA=LU

Theorem: if square matrix a is invertible, **LU factorization with partial pivoting (LUP)** refers often to lu factorization with row permutations only:

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

Tips: permutation matrix P admits $P^{-1} = P^T$. the inverse of a permutation matrix is a permutation matrix, and the product of a permutation matrix is a permutation matrix.

1.5 Stability

- LU factorization: unstable
- LUP factorization: stable

1.6 Cholesky Decomposition

When A refers to symmetirc positive-definitely matrix(SPD), it admits:

- $A = A^T$
- for any $v \neq 0$, $v^TAv > 0$

$$A = LL^T$$

Note: Diagonal elements in L is not 1

e.g.

$$\begin{pmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 6 & 1 & 0 \\ -8 & 5 & 3 \end{pmatrix} \begin{pmatrix} 2 & 6 & -8 \\ 0 & 1 & 5 \\ 0 & 0 & 3 \end{pmatrix}$$

And here is its LDL^T decomposition:

$$\begin{pmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -4 & 5 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 9 \end{pmatrix} \begin{pmatrix} 1 & 3 & -4 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{pmatrix}$$

No need to partial pivoting like normal matrix, it's just numerically stable.

write out the equation:

$$\begin{aligned} \mathbf{A} = \mathbf{L}\mathbf{L}^T &= \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix} \\ &= \begin{pmatrix} L_{11}^2 & L_{21}L_{11} & L_{31}L_{11} \\ L_{21}L_{11} & L_{21}^2 + L_{22}^2 & L_{31}L_{21} + L_{32}L_{22} \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix} \end{aligned}$$

and therefore the following formulas for the entries of $\mathbf{L}$:

$$L_{j,j} = (\pm)\sqrt{A_{j,j} - \sum_{k=1}^{j1} L_{j,k}^2},$$

$$L_{i,j} = \frac{1}{L_{j,j}} \left(A_{i,j} - \sum_{k=1}^{j-1} L_{i,k} L_{j,k} \right) \qquad \text{for } i > j$$

$L = A$
for $j = 1 : n$ **do**
  $\ell_{jj} = \sqrt{\ell_{jj}}$
  for $i = j + 1 : n$ **do**
    $\ell_{ij} = \ell_{ij} / \ell_{jj}$
  end for
  for $k = j + 1 : n$ **do**
    for $i = k : n$ **do**
      $\ell_{ik} = \ell_{ik} - \ell_{ij}\ell_{kj}$
    end for
  end for
end for

1.7 Symmetric Matrix

If A is a symmetric matrix, then $A = A^T$.

$$A = LDU \quad A^T = U^T D L^T.$$

$A = A^T$. U is upper triangular matrix, so the U^T is lower triangular matrix. And because of the decomposition is unique, $U^T = L$.

$$A = LDL^T$$

2 QR decomposition

3 SVD Decomposition

SVD can be seen as a generalization of eigenvalues and eigenvectors to non-square matrices. The computation of SVD is numerically **well-conditione**.

3.1 Introduce

$m \times n$ matrix **M** is a factorization of the form $\boldsymbol{U\Sigma V^T}$, where $\boldsymbol{U}_{m\times m}$ and $\boldsymbol{V}_{n\times n}$ are orthogonal matrixes, $\Sigma_{m\times n}$ is an rectangular diagonal matrix with non-negative numbers on the diagonla.

r = rank(A)

$$\begin{aligned}
\mathbf{A} &= \mathbf{U\Sigma V}^T \\
&= \mathbf{U} \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0} \end{pmatrix}_{m\times n} \mathbf{V}^T \\
&= \sigma_1 \mathbf{u_1} \mathbf{v_1}^T + \cdots + \sigma_r \mathbf{u_r} \mathbf{v_r}^T
\end{aligned}$$

Customarily, $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r \geqslant 0$, which are call the **singular values**. The front R columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are called **singular vector**

Set $\mathbf{U} = \begin{pmatrix} \mathbf{u_1} & \cdots & \mathbf{u_r} & \mathbf{u_{r+1}} & \cdots & \mathbf{u_m} \end{pmatrix}$, $\mathbf{V} = \begin{pmatrix} \mathbf{v_1} & \cdots & \mathbf{v_r} & \mathbf{v_{r+1}} & \cdots & \mathbf{v_n} \end{pmatrix}$.

$$\begin{aligned}
&\mathbf{AV} = \mathbf{U\Sigma} \\
&\implies \mathbf{Av_i} = \sigma_i \mathbf{u_i}, \quad \mathbf{Av_j} = \mathbf{0} \\
&\mathbf{A^T} = \mathbf{V\Sigma U^T} \\
&\implies \mathbf{A^T u_i} = \sigma_i \mathbf{v_i}, \quad \mathbf{A^T u_k} = \mathbf{0} \\
&(i = 1, \cdots, r) \quad (j = r+1, \cdots n) \quad (k = r+1, \cdots, m) \\
&\implies \mathbf{A^T A v_i} = \sigma_i^2 \mathbf{v_i}, \quad \mathbf{AA^T u_i} = \sigma_i^2 \mathbf{u_i}
\end{aligned}$$

Note that this generalizes the eigenvalue decomposition. While the latter decomposes a symmetric square matrix A with an orthogonal transformation V as:

$$A = V \wedge V^\top, \quad \text{with } V \in O(n), \Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_n\} \tag{2}$$

3.2 Properties

Assume $\mathbf{A}$ is $m \times n$ real matrix with rank r, then $\mathbf{AA^T}$ is $m \times m$ real symmetric matirx, then $\mathbf{A^TA}$ is $n \times n$ real symmetric matirx.

(1) The eigenvalues of $\mathbf{A^TA}$ and $\mathbf{AA^T}$ are non-negative.

Proof:

$$\mathbf{A^TAx} = \lambda\mathbf{x}(x \neq \mathbf{0})$$

Multiply both sides by $\mathbf{x^T}$

$$\begin{aligned} \mathbf{x^TA^TAx} &= \lambda\mathbf{x^Tx} \\ \implies ||\mathbf{Ax}||^2 &= \lambda||x||^2 \\ x \neq \mathbf{0} \implies ||x||^2 \neq 0, &\quad ||\mathbf{Ax}||^2 \geqslant 0 \end{aligned}$$

thus, $\lambda \geqslant 0$

(2) The sets of non-zero eigenvalues of $\mathbf{A^TA}$ and $\mathbf{AA^T}$ are same.

Proof:

$$r(\mathbf{A^TA}) = r(\mathbf{AA^T}) = r(A)$$

$\mathbf{A^TA}$ is symmetirc matrix, which must be similar to diagonalization.

$$\mathbf{A^TA} \sim \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \mathbf{0} \end{pmatrix} = \wedge$$

thus, the number of non-zero eigenvalues of $\mathbf{A^TA}$ is equal to which of $\mathbf{AA^T} = r$

Assume λ is non-zero eigenvalues of $\mathbf{A^TA}$, then

$$\begin{aligned} \mathbf{A^TAx} &= \lambda\mathbf{x} \\ \implies \mathbf{AA^TAx} &= \lambda\mathbf{Ax} \end{aligned}$$

$$\lambda \neq 0 \implies \mathbf{A^TAx} \neq 0 \implies \mathbf{Ax} \neq 0$$

(3) $\sigma_1 \geqq |\lambda|_{max}, \quad \sigma_1 \geqq |a_{ij}| \quad \forall i, j$

Proof:

$$\|A\mathbf{x}\| = \left\|U\Sigma V^T\mathbf{x}\right\| = \left\|\Sigma V^T\mathbf{x}\right\| \leq \sigma_1 \left\|V^T\mathbf{x}\right\| = \sigma_1\|\mathbf{x}\|.$$

If $Ax = \lambda x$, $\|Ax\| = |\lambda| \, \|x\|$,then $\sigma_1 \geq |\lambda|_{max}$

Assume $x = (1, 0, \cdots, 0)$, Ax represent the first column vector, and $\|Ax\| \geq \sigma_1\|x\| = \sigma_1$ then

$$|a_{i1}| \leq \sqrt{a_{11}^2 + \cdots + a_{n1}^2} \leq \sigma_1$$

3.3 Compare With Eigenvalues

the number of non-zero singular values $\Longleftrightarrow$ the rank of matrix.

However, the number of non-zero eigenvalues is less than the rank.(nilpotent matrix has no non-zero eigenvalues)

3.4 Pseudo Inverse

For an arbitrary matrix $A \in \mathbb{R}^{m\times n}$, if its SVD is $A = U\Sigma V^T$, the pseudeo inverse is defined as

$$A^\dagger = V\Sigma^\dagger U^\top, \text{ where } \Sigma^\dagger = \begin{pmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{pmatrix}_{n\times m} \tag{3}$$

which admits

$AA^\dagger A = A, \quad A^\dagger AA^\dagger = A^\dagger, \quad (AA^\dagger)^T = AA^\dagger, \quad (A^\dagger A)^T = A^\dagger A$

On the contrary, the generalized inverse that satisfies these four equations is actually unique.

In addition, the linear system $Ax = b$ with $A \in \mathbb{R}^{m\times n}$of rank $r \leq min(m, n)$ can have multiple or no solutions.

$x_{min} = A^\dagger b$ **is among all minimizers of** $|Ax - b|^2$ **the one with the smalleset norm** $|x|$**.**