LU decomposition

$A = LU =$  ← use Gaussian-Jordan Elimination

$\therefore Ax = LUx = b$ Let $Ux = y$, (i) Solve for $Ly = b$ → y

(ii) Solve for $Ux = y$ → x

- PLU decomposition

if $A = \begin{bmatrix} 0 & \\ & \end{bmatrix}$ i.e. first element $= 0$, cannot find L using Gaussian-Jordan Elimination

$\therefore PA =$ Multiply permutation P to make first element $\neq 0$

e.g. $P = \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix}$ swap row 1 and 2

Since P is orthogonal matrix, $P = P^T = P^{-1}$ $\therefore PA = LU \rightarrow A = PLU$

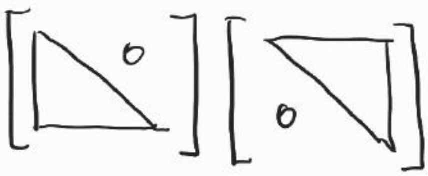
- LDU decomposition

Make diagonal entries of L and U into 1 using diagonal matrix D

$A = LU = L'DU'$

Cholesky decomposition

if A (from $Ax = b$) is symmetric and positive (semi-) definite,

$A = LL^T =$  * Numerically stable decomposition

- LDL^T decomposition

Make diagonal entries of L to 1 using diagonal matrix D

$A = LL^T = L'DL'^T$

QR decomposition

Decompose matrix A into orthogonal matrix Q and upper-triangle matrix R

$A = QR \quad (QQ^T = I)$

* slower than LU decomposition, but efficient to solve least squares problem

- Detailed step

$$A = \begin{bmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{bmatrix} \xrightarrow{\text{Gram-schmidt}} Q = \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix}$$

$$q_1 = a_1 / |a_1|$$

$$q_2 = a_2 - \frac{a_2 \cdot q_1}{q_1 \cdot q_1} q_1$$

$$\vdots$$

$$q_n = a_n - \sum_{i=1}^{n} \frac{a_n \cdot q_i}{q_i \cdot q_i} q_i$$

then,
$$[a_1 \ a_2 \ a_3] = [q_1 \ q_2 \ q_3] \begin{bmatrix} a_1^T q_1 & a_2^T q_1 & a_3^T q_1 \\ & a_2^T q_2 & a_3^T q_2 \\ & & a_3^T q_3 \end{bmatrix} \quad (A = QR)$$

* If A is non-square matrix, $A \in \mathbb{R}^{4\times3} = [q_1 \ q_2 \ q_3 \ q_4] \begin{bmatrix} a_1^T q_1 & a_2^T q_1 & a_3^T q_1 & 0 \\ & a_2^T q_2 & a_3^T q_2 & 0 \\ & & a_3^T q_3 & 0 \\ & & & 0 \end{bmatrix}$

- QR decomposition on least squares problem

$Ax = b \Rightarrow \min_x \|Ax - b\|_2^2$, then $x = (A^TA)^{-1}A^Tb$

if $A = QR = [Q_1 \ Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$, then $\|Ax - b\| = \|QRx - b\|$

$$= \|Q(Rx - Q^Tb)\|$$

$$= \|Rx - Q^Tb\|$$

$$= \left\| \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x - \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} b \right\|$$

$$= \|R_1 x - Q_1^T b\| - \|Q_2^T b\|$$

$\therefore \min_x = R^{-1}Q^Tb$

Eigendecomposition

If square matrix A is diagonalizable,

$A = VDV^{-1}$ $(A, V, D \in \mathbb{R}^{n \times n})$ $\quad \therefore$ A is diagonalizable $\Leftrightarrow$ A is take eigendecomposition

Conditions
- V is invertible (n independent column vectors)
- Each column of V is A's eigenvectors

Singular Value Decomposition

For $A \in \mathbb{R}^{m \times n}$, $A = U \Sigma V^T$
- $U \in \mathbb{R}^{m \times m}$: Each column is ColA's orthonormal basis
- $\Sigma \in \mathbb{R}^{m \times n}$: Diagonal matrix w/ $\sigma_1 \geq \sigma_2 \geq \cdots \geq$ singular value
- $V \in \mathbb{R}^{n \times n}$: Each column is RowA's orthonormal basis

- Sum of outer products

$$A = U \Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T \quad \text{where} \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \quad (\text{if } m \geq n)$$

$\downarrow$ then,

Reduced form of SVD

$$A = U'D'V^T \quad \text{where} \quad U' \in \mathbb{R}^{m \times n},\ D' \in \mathbb{R}^{n \times n}$$

- Another perspective of SVD

$A \in \mathbb{R}^{n \times n}$ $\xrightarrow[\text{Orthogonalization}]{\text{Gram-Schmidt}}$ ColA's orthonormal basis $u_1, u_2, \cdots, u_n$; RowA's orthonormal basis $v_1, v_2, \cdots, v_n$] they are not unique!

$$AV = [Av_1 \ Av_2 \ \cdots \ Av_n]$$

$$U\Sigma = [u_1 \ u_2 \ \cdots \ u_n] \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ & & \ddots & \\ 0 & & & \sigma_n \end{bmatrix} = [\sigma_1 u_1 \ \cdots \ \sigma_n u_n]$$

$$AV = U\Sigma \quad \Leftrightarrow \quad A = U \Sigma V^T$$

- Computing SVD

$$AA^T = U\Sigma V^T V \Sigma^T U^T = U \Sigma \Sigma^T U^T = U \Sigma^2 U^T$$
$$A^TA = V\Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T = V \Sigma^2 V^T$$

$>$ they share equal Σ

$\Rightarrow$ Orthonormal Eigenvectors of A^TA = column vectors of V

square root of corresponding eigenvalues = singular value σ

- Range and nullspace of SVD

Advantage of SVD : Decomposition possible for both singular and non-singular case

Non-singular : $A^{-1} = V \cdot diag(1/\sigma_j) \cdot U^T$

Singular : Set $\sigma_j = 0$ case $\rightarrow$ $1/\sigma_j \Rightarrow 0$

if $\sigma_j \neq 0$, corresponding U's column is A's orthogonal set of basis vector of Range

if $\sigma_j = 0$, corresponding V's column is A's ______________ of Null space

* rank of A = # of singular value s.t. $\sigma_j \neq 0$

- SVD in under-determined system

if A is singular and b is in Range ($Ax=b$), linear system has infinite solutions

$$\min_x \|x\|^2 \rightarrow x = V \cdot diag(1/\sigma_j) \cdot U^T \cdot b$$

- SVD in over-determined system

if A is singular and b is not in Range ($Ax=b$) , linear system has no solution

$$\min_x \|Ax - b\| \rightarrow x = V \cdot diag(1/\sigma_j) \cdot U^T \cdot b$$

Pseudo inverse

if matrix A is non-square matrix

$\rightarrow$ able to find pseudo-inverse if A has full row/column rank

- under-determined system

A : full row rank $\rightarrow$ optimization using lagrange multiplier λ : $\min_x \|x\|^2 + \lambda^T(b - Ax) = f$

$$\frac{df}{dx} = 0$$

$$2x - A^T\lambda = 0$$

$2Ax - AA^T\lambda = 0$ $\quad \because$ A is not square matrix

$2b - AA^T\lambda = 0$ $\quad \because Ax = b$

$\therefore \lambda = 2(AA^T)^{-1}b \longrightarrow x = A^T(AA^T)^{-1}b$

Right pseudo-inverse : $A^+ = A^T(AA^T)^{-1}$

$$AA^+x = A^+b$$
$$x = A^+b$$
$$\therefore x = A^T(AA^T)^{-1}b$$

- Over-determined system

Left pseudo-inverse : $A^+ = (A^TA)^{-1}A^T$

$$A^+Ax = A^+b$$
$$x = A^+b$$
$$\therefore x = (A^TA)^{-1}A^Tb$$

- SVD of pseudo inverse

$A = U\Sigma V^T \rightarrow A^+ = V\Sigma^+U^T$ where $A \in \mathbb{R}^{m\times n}$, $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots) \in \mathbb{R}^{m\times n}$

$\Sigma^+ = \mathrm{diag}(1/\sigma_1, 1/\sigma_2, \cdots) \in \mathbb{R}^{n\times m}$

- Full column rank case

$$A^+A = V\Sigma^+U^TU\Sigma V^T = I_n$$

$$\begin{aligned} A^+ &= (A^TA)^{-1}A^T \\ &= (V\Sigma U^TU\Sigma V^T)^{-1}V\Sigma U^T \\ &= V\Sigma^{-2}\Sigma U^T \\ &= V\Sigma^+U^T \end{aligned}$$

- Full row rank case

$$AA^+ = U\Sigma V^TV\Sigma^+U^T = I_m$$

$$\begin{aligned} A^+ &= A^T(AA^T)^{-1} \\ &= V\Sigma^TU^T(U\Sigma V^TV\Sigma^TU^T)^{-1} \\ &= V\Sigma\Sigma^{-2}U^T \\ &= V\Sigma^+U^T \end{aligned}$$

- Rank deficient case

if $A = U\Sigma V^T = U\begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}V^T$ (i.e. $\mathrm{rank}(A) = 2$)

then, $A^+ = V\Sigma^+U^T = V\begin{bmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}U^T$

$\therefore AA^+ = U\Sigma V^TV\Sigma^+U^T = U\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}U^T = u_1u_1^T + u_2u_2^T$

* if full rank, $u_1u_1^T + u_2u_2^T + u_3u_3^T = I_3$

$$A^+A = v_1v_1^T + v_2v_2^T$$

$\Rightarrow$ for $A \in \mathbb{R}^{m\times n}$, if $m > n$ A^+A is closer to I

if $m < n$ AA^+ is closer to I

- QR decomposition of pseudo inverse when singular case

if A is singular, $Ax=b$ $\rightarrow$ $x = A^{+}b$

$$= (A^TA)^{-1}A^Tb$$

$$= (R^TQ^TQR)^{-1}R^TQ^Tb \qquad \because A = QR$$

$$= (R^TR)^{-1}R^TQ^Tb \qquad \because Q^TQ = I$$

$$= R^{-1}Q^Tb$$

$\therefore x = R^{-1}Q^Tb$ is $\min_x \|Ax-b\|$

Woodbury's identity

⇒ tool that simplifies inverse of the matrix by adding rank 1 matrix

$$(A+uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u} \qquad 1+v^TA^{-1}u \neq 0 \iff A+uv^T \text{ is invertible}$$

$$uv^T = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} [v_1 \; v_2]$$

- Recursive least squares

In $Ax=b$ problem, update A^{-1} efficiently as new data is added

$$\begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \rightarrow x = (A^TA)^{-1}A^Tb$$

if new data a_{m+1}^T is added,

$$x = \left([A \;\; a_{m+1}] \begin{bmatrix} A^T \\ a_{m+1}^T \end{bmatrix} \right)^{-1} [A^T \;\; a_{m+1}] \begin{bmatrix} b \\ b_{m+1} \end{bmatrix} = \underbrace{(A^TA + a_{m+1}a_{m+1}^T)^{-1}}_{\text{update this term efficiently}} (A^Tb + a_{m+1}b_{m+1})$$

$$(A^TA + aa^T)^{-1}$$

$$= P - \frac{Paa^TP}{1+a^TPa} \qquad \text{where } P = (A^TA)^{-1}$$

Above equation becomes $x + Pa(b - a^Tx)$

$\therefore$ Recursive least squares (RLS) : $x \leftarrow x + Pa(b - a^Tx)$

- Matrix inversion lemma

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$$ where $A \in \mathbb{R}^{n\times n}, U \in \mathbb{R}^{n\times k}, C \in \mathbb{R}^{k\times k}, V \in \mathbb{R}^{k\times n}$; $A, C, C^{-1}+VA^{-1}U$ is invertible

* Extension of woodbury's identity

if C is scalar & $B \in \mathbb{R}^{n \times 1}$, $D \in \mathbb{R}^{1 \times n}$, above lemma equals woodbury's identity