→ Angle bet. two vectors $= \cos^{-1}\left[\dfrac{(x, y)}{\|x\|_2 \|y\|_2}\right]$

Norms

→ Properties of a valid norm:
→ Should be real & non-negative → $\|x\| = 0 \Leftrightarrow x = 0$
→ $\|\alpha x\| = \alpha \|x\|$ → $\|x + y\| \le \|x\| + \|y\|$

→ $|x^T y| \le \|x\|_2 \|y\|_2$ (Cauchy-Shwarz Inequality)

→ $\|A\|_1 = \max_{1 \le i \le m} (\|a_i\|_1)$ → $\|A\|_\infty$ = max. row absolute sum

→ If $A = uv^T$, $\|A\|_2 = \|u\|_2 \|v\|_2$ (when $x = \dfrac{v}{\|v\|}$)

→ $\|AB\|_p \le \|A\|_p \|B\|_p$

→ $\|A\|_F = \sqrt{\sum_{i=0}^{m} \sum_{j=0}^{n} |a_{ij}|^2} = \sqrt{\mathrm{tr}(AA^T)} = \sqrt{\mathrm{tr}(A^TA)} = \sqrt{\sum_{i=0}^{m} \|a_i\|^2}$

Conditioning & Stability

→ $\kappa(A) = \dfrac{\sigma_1}{\sigma_m} = \sqrt{\dfrac{\lambda_{max}(A^TA)}{\lambda_{min}(A^TA)}}$

→ $\kappa^R(A) = \|A\| \|A^+\|$, where $\|A^+\| = (A^TA)^{-1}A^T$

→ Backwards stable $\Rightarrow$ Stable $\Rightarrow$ Forward stable
$\not\Leftarrow$ $\not\Leftarrow$

Backward stability: $\tilde{f}(x) = f(x + \delta x)$
Forward stability: $\|\tilde{f}(x) - f(x)\| = O(\epsilon_m)$

→ For a backward stable algo, $f(\tilde{x}) = \tilde{f}(x)$

SVD

$$A = U \Sigma V^T, \quad U \in \mathbb{R}^{m \times m}, \text{ orthogonal}; \quad \Sigma \in \mathbb{R}^{m \times n}, \text{ diagonal}; \quad V^T \in \mathbb{R}^{n \times n}, \text{ orthogonal}$$

→ $\|A\|_2 = \sigma_1$, $\|A\|_F = \sqrt{\sigma_1^2 + \cdots \sigma_n^2}$

Principal Component Analysis

→ Consider a data matrix $A \in \mathbb{R}^{m \times n}$

1) Make the matrix 0-centered by subtracting the mean of the coln. from each element in the coln.
2) Variance of the coln. $= \dfrac{E(X^2) - \cancel{(E(X))^2}}{m} = \dfrac{\|a_j\|_2^2}{m}$ (0)
3) Hence, total variance, $T = \|a_1\|_2^2 + \cdots + \|a_n\|^2$
$= \|A\|_F^2$
$= \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$

Biggest contributor to the variance

→ To get the direction with highest variance, we need to find a vector $\hat{w}_1$ s.t. $t_1 = A\hat{w}_1$, $\|t_1\|_2$ is maximised. The solution is $\hat{w}_1 = v_1$.

$t_1 = A v_1 = \sum_{i=1}^{n} u_i \sigma_i v_i^T v_1 = \sigma_1 u_1$

→ Similarly, $t_2 = \sigma_2 u_2$, where t_2 is the direction along second highest variance.

Projection

→ Projector matrix P, must be:
→ $P^2 = P$ (idempotent) → $Pv - v \in Null(P)$
→ Square matrix → Rank-deficient
→ $Px = x \ \forall x \in Range(P)$ → symmetrical

→ If P is an orthogonal projector, so is $(I - P)$

→ If A is orthogonal, $P = A(\cancel{A^TA})^{-1}A^T = AA^T$ (I)

→ Projection & SVD: If $A = U\Sigma V^T$, U forms a orthonormal basis of A.
Hence, the projection matrix corresponding to A, $P = Q(\cancel{Q^TQ})^{-1}Q^T = QQ^T$ (I)

QR Factorisation

→ Q: matrix w/ orthonormal colns. → R: UTM

→ Gram-Schmidt orthogonalisation:
$v_j = a_j - (q_1^T a_j) q_1 - \cdots - (q_{j-1}^T a_j) q_{j-1}$
$q_j = \dfrac{v_j}{\|v_j\|_2}$

Hence $q_n = \dfrac{a_n - \sum_{i=0}^{n-1} r_{in} q_i}{r_{nn}}$, where $r_{ij} = \begin{cases} q_i^T a_j, & i \ne j \\ \|a_j - \sum_{i=1}^{j-1} r_{ij} q_i\|_2, & i = j \end{cases}$

→ Gram Schmidt with projector: $q_n = \dfrac{P_n a_n}{\|P_n a_n\|}$, where
$P_n = I - \sum_{i=0}^{n-1} q_i q_i^T = I - \hat{Q}_{j-1} \hat{Q}_{j-1}^T$

→ Modified Gram-Schmidt:
$P_n a_n = (I - \sum_{i=0}^{n-1} q_i q_i^T) a_n = \left[\prod_{i=0}^{n-1} (I - q_i q_i^T)\right] a_n$

for $j = 1 \to n$:
$v_j^{(1)} = a_j$
$v_j^{(2)} = P_{\perp q_1} a_j = (I - q_1 q_1^T) a_j = v_j^{(1)} - q_1 q_1^T v_j^{(1)}$
$v_j^{(3)} = P_{\perp q_2} a_j = v_j^{(2)} - q_2 q_2^T v_j^{(2)}$
⋮
$v_j^{(j)} = v_j^{(j-1)} - q_{(j-1)} q_{(j-1)}^T v_j^{(j-1)}$

$q_j = \dfrac{v_j^{(j)}}{\|v_j^{(j)}\|_2}$

→ Householder Triangularisation
→ Perform transformations to convert A to R:

$Q_n \cdots Q_1 A = R$

$Q_k = \begin{bmatrix} I_{(k-1)\times(k-1)} & 0 \\ 0 & \hat{F}_{(m-k+1)(m-k+1)} \end{bmatrix}_{m \times m}$

ensures that prev colns. are not changed.
must be an orthogonal matrix

→ Let $\hat{x} = \begin{bmatrix} - \\ - \\ \vdots \end{bmatrix}$, F should be s.t. $Fx = \begin{bmatrix} \|x\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \|x\| \cdot e_1$

Hence, $y = Fx = (I - 2uu^T)x$, where $u = \dfrac{v}{\|v\|}$,
where $v = \|x\|_2 e_1 - x$

Algo: for $k = 1 \rightarrow n$:

$\tilde{x} = \tilde{A}(k:m, k)$ // row k to row m in the k^{th} coln.

$\tilde{v}_k = \text{sgn}(x_1) \cdot \|\tilde{x}\|_2 \tilde{e}_1 + \tilde{x}$

$\tilde{v}_k = \frac{\tilde{v}_k}{\|\tilde{v}_k\|}$

$\tilde{A}(k:m, k:n) -= 2\tilde{v}_k \tilde{v}_k^T \tilde{A}(k:m, k:n)$

Algo	FLOPs	Stability	Error
CGS	$2mn^2$	Unstable	$O(\kappa(A)^2 \cdot \epsilon_m)$
MGS	"	Backward stable	$O(\kappa(A) \cdot \epsilon_m)$
Householder	$2mn^2 - 2/3 n^3$	"	$O(\epsilon_m)$

Cholesky Decomposition

→ If $\tilde{A}$ is S.P.D, $\tilde{A}$ can be decomposed s.t. $\tilde{A} = \tilde{R}^T \tilde{R}$ UTM

→ $n^3/3$ FLOPs

Symmetric +ve Definite Matrix (SPD)

→ Properties

→ $\tilde{A} = \tilde{A}^T$

→ $(\tilde{x}, \tilde{A}\tilde{y}) = (\tilde{y}, \tilde{A}\tilde{x}) \quad \forall x, y \in \mathbb{R}^m$

→ $\tilde{x}^T \tilde{A} \tilde{x} > 0, \quad \forall \tilde{x} \in \mathbb{R}^m$

→ If $\tilde{A}$ is S.P.D. & $\tilde{X} \in \mathbb{R}^{m \times n}$ is full rank, then $\tilde{X}^T \tilde{A} \tilde{X}$ is also S.P.D

→ All eigenvalues are +ve for any S.P.D.

Linear Least Squares

→ $\tilde{A}\hat{x} = \tilde{P}\tilde{b}$ // get b to coln. space of A

$\tilde{A}\hat{x} = \tilde{A}(\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T\tilde{b} \Rightarrow \hat{x} = (\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T\tilde{b}$

→ Soln. by Cholesky decomposition:

$\tilde{A}^T\tilde{A}\hat{x} = \tilde{A}^T\tilde{b}$ normal equation

$\tilde{R}^T\tilde{R}\hat{x} = \tilde{A}^T\tilde{b}$

$\tilde{R}^T\tilde{w} = \tilde{A}^T\tilde{b}$ // Let $\tilde{w} = \tilde{R}\hat{x}$

↓ Solve for $\tilde{w}$

$\tilde{R}\hat{x} = \tilde{w}$

↓ solve for $\hat{x}$

$\hat{x}$

→ Soln. by $\tilde{Q}\tilde{R}$ factorisation

$\tilde{A}\tilde{x} = \tilde{P}\tilde{b}$

$\tilde{Q}\tilde{R}\tilde{x} = \tilde{Q}\tilde{Q}^T\tilde{b}$

$\tilde{R}\tilde{x} = \tilde{Q}^T\tilde{b}$

↓ solve for $\tilde{x}$

$\tilde{x}$

→ Soln with SVD

$\tilde{A}\tilde{x} = \tilde{P}\tilde{b}$

$\tilde{U}\tilde{\Sigma}\tilde{V}^T\tilde{x} = \tilde{U}\tilde{U}^T\tilde{b}$

$\tilde{\Sigma}\tilde{V}^T\tilde{x} = \tilde{U}^T\tilde{b}$

$\tilde{\Sigma}\tilde{y} = \tilde{U}^T\tilde{b}$ // Let $\tilde{V}^T\tilde{x} = \tilde{y}$

↓ Solve for $\tilde{y}$

$\tilde{V}^T\tilde{x} = \tilde{y}$

↓ Solve for $\tilde{x}$

$\tilde{x}$

Algo	Work
Cholesky	$mn^2 + n^3/3$
QR (Householder)	$2mn^2 - \frac{2n^3}{3}$
SVD	$2mn^2 + 11n^3$

→ If $\tilde{A}$ is close to rank-deficient,

$\boxed{\text{Either } m \geq n \text{ \& } \text{rank}(A) < n \text{, or } m < n}$

$\tilde{A}\tilde{x} = \tilde{b}$

$\tilde{U}\tilde{\Sigma}\tilde{V}^T\tilde{x} = \tilde{b}$

$\boxed{\tilde{U}^T\tilde{U} = \tilde{I} \text{ , whereas } UU^T \text{ will be a projector matrix}}$

$\cancel{\tilde{U}^T\tilde{U}}^{\,I}\tilde{\Sigma}\tilde{V}^T\tilde{x} = \tilde{U}^T\tilde{b}$

$\tilde{V}^T\tilde{x} = \tilde{\Sigma}^{-1}\tilde{U}^T\tilde{b}$

Let $\tilde{x} = \tilde{V}_1\tilde{y} + \tilde{V}_2\tilde{z}$, where $\tilde{V} = [\tilde{V}_1 \quad \tilde{V}_2]$

range (A) null (A)

Hence $\tilde{y} = \tilde{\Sigma}^{-1}\tilde{U}^T\tilde{b}$

Therefore $\tilde{x} = \tilde{V}_1\tilde{\Sigma}^{-1}\tilde{U}^T\tilde{b} + \tilde{V}_2\tilde{z}$

$\|\tilde{x}\|_2 = \|\tilde{V}_1\tilde{\Sigma}^{-1}\tilde{U}^T\tilde{b}\|_2 + \|\tilde{V}_2\tilde{z}\|_2 - \cancel{\|(\tilde{V}_1\tilde{\Sigma}^{-1}\tilde{U}^T\tilde{b}, \tilde{V}_2\tilde{z})\|_2}$ inner product 0

To minimise $\|\tilde{x}\|_2$, we set $\tilde{z} = \tilde{0}$

→ If $b \notin$ Range $(\tilde{A})$, then we solve $\tilde{A}\tilde{x} = \tilde{P}\tilde{b}$.

In this case, $\tilde{x} = \tilde{V}_1\tilde{\Sigma}^{-1}\tilde{U}^T\tilde{P}\tilde{b} + \tilde{V}_2\tilde{z}$ $\boxed{P = UU^T}$

$= \tilde{V}_1\tilde{\Sigma}^{-1}\tilde{U}^T\tilde{b} + \tilde{V}_2\tilde{z}$

Eigen Decomposition

→ $\tilde{A} = \tilde{X}\tilde{\Lambda}\tilde{X}^{-1}$, $\tilde{\Lambda}$ is a diagonal matrix containing eigenvalues and $\tilde{X}$ is a matrix comprising of the respective eigenvectors

→ Characteristic polynomial (P_A) : $P_A(z) = \det(\tilde{A} - z\tilde{I})$

→ Eigenvectors are essentially Nullspace of $(\tilde{A} - \lambda\tilde{I})$. Hence, $(\tilde{A} - \lambda\tilde{I})$ must be a rank-deficient (singular) matrix, with $\det(\tilde{A} - \lambda\tilde{I}) = 0$

→ Geometric multiplicity of λ: No. of L.I. eigenvectors associated with an eigenvalue λ.

→ If λ corresponds to two eigenvectors $\tilde{v}_1$ & $\tilde{v}_2$, any vector $\tilde{x} = \alpha_1\tilde{v}_1 + \alpha_2\tilde{v}_2$ will be an eigenvector of A:

Pf: $\tilde{A}(\alpha_1\tilde{v}_1 + \alpha_2\tilde{v}_2) = \alpha_1\tilde{A}\tilde{v}_1 + \alpha_2\tilde{A}\tilde{v}_2 = \alpha_1\lambda\tilde{v}_1 + \alpha_2\lambda\tilde{v}_2$

$= \lambda(\alpha_1\tilde{v}_1 + \alpha_2\tilde{v}_2)$

$= \lambda\tilde{x}$ (shown)

Similarity Transformation

→ If $\tilde{X} \in \mathbb{R}^{m \times m}$ is non-singular, $\tilde{X}\tilde{A}\tilde{X}^{-1}$ is known as similarity transformation of $\tilde{A}$.

→ Two matrices $\tilde{A}$ & $\tilde{B}$ are said to be similar if there exists a similarity transformation bet. them, i.e $\tilde{B} = \tilde{X}\tilde{A}\tilde{X}^{-1}$.

→ $\tilde{A}$ & $\tilde{B}$ will have same eigenvalues & respective multiplicities

$P_B(z) = \det(z\tilde{I} - \tilde{X}\tilde{A}\tilde{X}^{-1})$

$= \det(z\tilde{X}\tilde{X}^{-1} - \tilde{X}\tilde{A}\tilde{X}^{-1})$

$= \det(\tilde{X}(z\tilde{I} - \tilde{A})\tilde{X}^{-1})$

$= \cancel{\det(\tilde{X})}\det(z\tilde{I} - \tilde{A})\cancel{\det(\tilde{X}^{-1})}$ (shown)

$\boxed{\text{* Eigenvectors may not be the same for } \tilde{A} \text{ \& } \tilde{B}}$