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

→ 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)

→ $\|A\|_1 = \max_{1\le i\le m} (\|a_j\|_1)$

→ $\|A\|_\infty$ = max. row absolute sum

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

→ $\|AB\|_p \le \|A\|_p\|B\|_p$

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

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

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

→ Orthogonal transformation of the feature space.

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

no. of experiments

particular measurement

1) Move the matrix 0-centered by subtracting the mean of the coln. from each element in the coln.

2) Variance of the coln. = $\frac{E(X^2) - \cancel{(E(X))^2}}{m}$ (0) $= \frac{\|a_j\|_2^2}{m}$

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 $\hat{w}_1 = v_1$.

$t_1 = Av_1 = \sum_{i=1}^{n} u_i\sigma_i v_i^T v_1$ $\left[v_i^T v_1 = \begin{cases} 0, & i\ne 1 \\ 1, & i = 1\end{cases}\right]$

$= u_1\sigma_1 \cancel{v_1^T v_1}$ (1)

$= \sigma_1 u_1$

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

Eigen Decomposition

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

→ Characteristic polynomial (P_A) : $P_A(z) = \det(A - zI)$

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

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

→ If λ corresponds to two eigenvectors v_1 & v_2, any vector $x = \alpha_1 v_1 + \alpha_2 v_2$ will be an eigenvector of A:

Pf: $A(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 Av_1 + \alpha_2 Av_2 = \alpha_1\lambda v_1 + \alpha_2\lambda$
$= \lambda(\alpha_1 v_1 + \alpha_2$
$= \lambda x$ (shown)

Similarity Transformation

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

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

→ A & B will have same eigenvalues & respective multiplicities

$P_B(z) = \det(zI - XAX^{-1})$
$= \det(zXX^{-1} - XAX^{-1})$
$= \det(X(zI - A)X^{-1})$
$= \cancel{\det(X)}\det(zI - A)\cancel{\det(X^{-1})}$ (shown)

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

→ Absolute cond. no, $\hat{\kappa} = \frac{\|\delta f\|}{\|\delta x\|} = \frac{\|J(x)\,\delta x\|}{\|\delta x\|} = \|J(x)\|$

→ Relative cond. no. $= \frac{\|\delta f\|}{\|f\|} \div \frac{\|\delta x\|}{\|x\|} = \frac{\|J(x)\|\|x\|}{\|f\|}$

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

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

$\boxed{\begin{array}{l}\text{Backward stability:} \\ \tilde{f}(x) = f(x+\delta x) \\ \text{Forward stability:} \\ \|\tilde{f}(x) - f(x)\| = O(\epsilon_m)\end{array}}$

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

→ 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 a projector, so is $(I-P)$. This the complementary projector of P.

→ Let us have $Ax = b$ has no soln., we can solve $A\hat{x} = Pb$ instead. P projects b onto the coln. space of A, which makes this eqn. solvable

Hence, $P = A(A^TA)^{-1}A^T$

→ 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)

Symmetric +ve Definite Matrix (SPD)

→ Properties

→ $A = A^T$

→ $(x, Ay) = (y, Ax) \;\forall x, y \in \mathbb{R}^m$

→ $x^TAx > 0, \;\forall x \in \mathbb{R}^m$

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

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

Cholesky Decomposition

→ If A is S.P.D., A can be decomposed s.t. $A = R^TR$

U.T.M

→ $n^3/3$ FLOPs

Linear Least Squares

Normal eqn

→ If $Ax = b$ is over determined, we can solve for $\boxed{A^TA\hat{x} = A^Tb}$ instead. This shall minimise the residual, i.e $\|r\|_2$, where $r = A\hat{x} - b$

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

$A\hat{x} = A(A^TA)^{-1}A^Tb \Rightarrow \hat{x} = (A^TA)^{-1}A^Tb$

Soln. by Cholesky decomposition:

$A^TA\hat{x} = A^Tb$
$R^TR\hat{x} = A^Tb$
$R^Tw = A^Tb$ // Let $w = R\hat{x}$
↓ Solve for w
$R\hat{x} = w$
↓ Solve for x
$\hat{x}$

→ Soln. by QR factorisation

$Ax = Pb$
$QRx = QQ^Tb$
$Rx = Q^Tb$
↓ Solve for x
x

Soln with SVD

$Ax = Pb$
$U\Sigma V^Tx = UU^Tb$
$\Sigma V^Tx = U^Tb$
$\Sigma y = U^Tb$ // Let $V^Tx = y$
↓ Solve for y
$V^Tx = y$
↓ Solve for x
x

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

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

→ Gram-Schmidt orthogonalisation:

$$v_j = a_j - (q_1^T a_j)\, q_1 - \cdots - (q_{j-1}^T a_j)\, q_{j-1}$$

$$q_j = \frac{v_j}{\|v_j\|_2}$$

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

→ Modified Gram Schmidt

→ 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:

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

for $j = 1 \to n$:

$$v_j^{(1)} = a_j$$

$$v_j^{(2)} = P_{\perp q_1} a_j = (I - q_1 q_1^T)\, a_j = v_j^{(1)} - q_1 q_1^T v_j^{(1)}$$

$$v_j^{(3)} = P_{\perp q_2} a_j = v_j^{(2)} - q_2 q_2^T v_j^{(2)}$$

$$\vdots$$

$$v_j^{(j)} = v_j^{(j-1)} - q_{(j-1)} q_{(j-1)}^T v_j^{(j-1)}$$

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

→ Householder Triangularisation

→ Perform transformations to convert A to R:

$$Q_n \cdots Q_1 A = R$$

ensures that prev colns. are not changed.

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

must be an orthogonal matrix

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

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

where $v = \|x\|_2 e_1 - x$

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

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

$v_k = \operatorname{sgn}(x_1) \cdot \|x\|_2\, e_1 + x$

$v_k = \dfrac{v_k}{\|v_k\|}$

$A(k:m, k:n) \mathrel{-}= 2 v_k v_k^T A(k:m, k:n)$

Algo	FLOPs	Stability	Error
GS	$2mn^2$	Unstable	$O(\kappa(A)^2 \cdot \varepsilon_m)$
MGS	"	Backward stable	$O(\kappa(A) \cdot \varepsilon_m)$
Householder	$2mn^2 - \frac{2}{3}n^3$	"	$O(\varepsilon_m)$

Linear Least Squares

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

→ If A is close to rank-deficient,

$$Ax = b$$

$$U \Sigma V^T x = b$$

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

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

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

Let $x = V_1 y + V_2 z$, where $V = [\underbrace{V_1}_{\text{range}(A)} \ \underbrace{V_2}_{\text{null}(A)}]$

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

Therefore $x = V_1 \Sigma^{-1} U^T b + V_2 z$

inner product

$$\|x\|_2 = \|V_1 \Sigma^{-1} U^T b\|_2 + \|V_2 z\|_2 - \cancel{\|(V_1 \Sigma^{-1} U^T b, V_2 z)\|_2} \quad 0$$

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

→ If $b \notin \text{Range}(A)$, then we solve $Ax = Pb$

In this case, $x = V_1 \Sigma^{-1} U^T P b + V_2 z$ $\quad \boxed{P = UU^T}$

$$= V_1 \Sigma^{-1} U^T b + V_2 z$$