Linear systems

- Linear Equation

$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$

$a_1, a_2, \cdots, a_n$: unkrown $\mathbb{R}/\mathbb{C}$

$x_1, x_2, \cdots, x_n$: variable

b : coefficient

$\parallel$

$a^T x = b \quad (a = [a_1 \cdots a_n]^T,\ x = [x_1 \cdots x_n]^T)$

- Linear System

= set of linear equations

$$\left.\begin{array}{l} a_1 x = b_1 \\ a_2 x = b_2 \\ \vdots \\ a_n x = b_n \end{array}\right] \rightarrow Ax = b \qquad A \in \mathbb{R}^{n\times m},\ x \in \mathbb{R}^{m\times 1},\ b \in \mathbb{R}^{m\times 1}$$

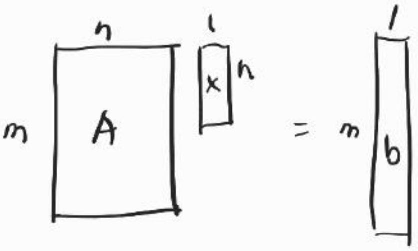
- Homogenous Equation

$$\left[\begin{array}{l} \text{Homogenous} : Ax = 0 \\ \text{Non-homogenous} : Ax = b \end{array}\right.$$

if $Ax=0$ only has trivial solution, $Ax=b$ has max 1 solution

infinite solutions, $Ax=b$ has infinite solutions

- over-determined system



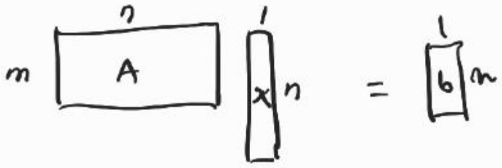
$(m>n)$

$b \notin col(A)$ in most cases

$\Leftrightarrow$ no solution

$\therefore \min_x \|Ax - b\|$

- Under-determined system



$(m<n)$

infinite # of solutions

$\therefore$ select $\min_x \|x\|^2$

- Solving Linear System

$$Ax = b \left[\begin{array}{l} \text{if } A^{-1} \text{ exists (invertible, non-singular)} \quad x = A^{-1}b \\ \text{if } A^{-1} \text{ not exists (non-invertible, singular)} \quad \text{No/Infinite many solution} \\ \quad (\det A = 0) \end{array}\right.$$

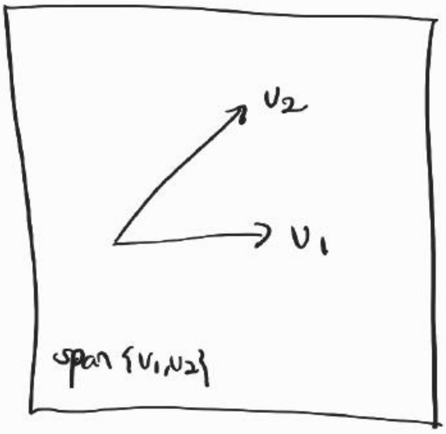
- Linear Combination

for vector $v_1, v_2, \cdots, v_n \in \mathbb{R}^n$, scalar $c_1, c_2, \cdots, c_n \in \mathbb{R}$

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

- Span

span$\{v_1, v_2, \cdots, v_n\}$: set of all possible linear combinations of $v_1, v_2, \cdots, v_n \in \mathbb{R}^n$
subset of $\mathbb{R}^n$



- From Matrix Equation to Vector Equation

Linear system $Ax = b$ $\rightarrow$ $\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = b$ $\rightarrow$ $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$

Linear Combination of Column vectors

solution exists if $b \in \text{span}\{a_1 \cdots a_n\}$

- Several perspectives about Matrix Multiplication

$(Ax)^T = (b)^T$

$x^T A^T = b^T$

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} = a_1^T x_1 + \cdots + a_n^T x_n = b^T$$

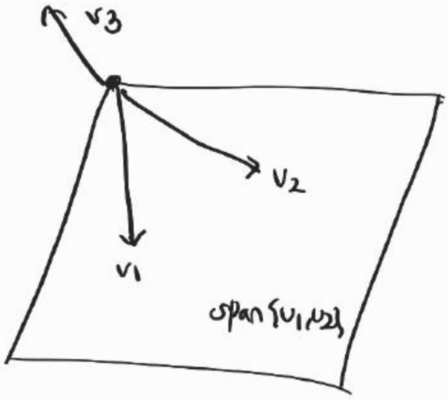
Linear combination of row vectors

---

rank-1 matrix $= ab^T = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}$ multiplication of column and row vector

- Linear independence

for set of vectors $\{v_1, v_2, \cdots, v_n\}$, if there exists $v_j \in \text{span}\{v_1, \cdots v_{j-1}, v_{j+1}, \cdots, v_n\}$ for some $j = 1, \cdots, n$, the set is linearly dependent else, the set is linearly independent



if linearly independent,
$x = [0\ 0 \cdots 0]^T$ is the only solution for
$x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = 0$

$v_3 \notin \text{span}\{v_1, v_2\}$ $\equiv$ $\{v_1, v_2, v_3\}$ are linearly independent
$\equiv$ v_3 increases the dim-space of span

- Linear Dependence

if $v_3 = \text{span}\{v_1, v_2\} = x_1 v_1 + x_2 v_2$, $\text{span}\{v_1, v_2\} = \text{span}\{v_1, v_2, v_3\}$
& $\{v_1, v_2, v_3\}$ are linearly dependent

- Span and subspace

Subspace $H \subset \mathbb{R}^n$: closed under linear combination
if $u_1, u_2 \in H$, $cu_1 + du_2 \in H$ (c, d is scalar)

$\therefore$ $\text{span}\{v_1, v_2, \cdots, v_n\}$ is always a subspace

- Basis of subspace

Conditions
- should span the whole subspace
- basis should be linearly independent

* standard basis vector (e_i : 1 in ith position, 0 otherwise)
e.g. $e_1 = [1\ 0]^T$, $e_2 = [0\ 1]^T$

- Dimension of subspace

Subspace can have different set of basis vectors, but
dim of subspace = # of basis vectors

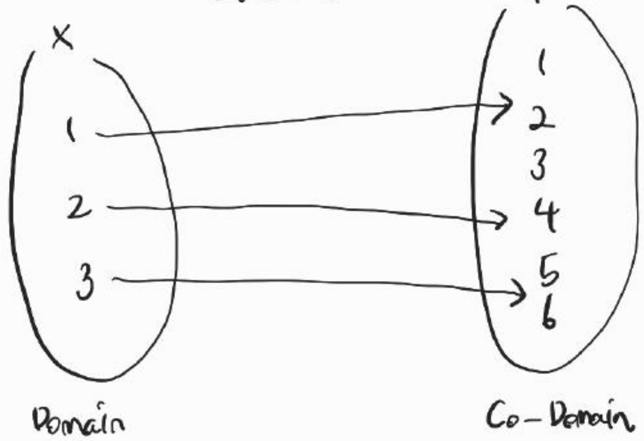
• Column space of Matrix

$$A = \begin{bmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{bmatrix} \rightarrow \quad \text{Col}A = \text{Span}\{a_1, a_2, a_3\}$$

$*$ Rank of A = dim Col$A$ = dim Row A

• Transformation

$T : X \to Y$



Domain : $\{1,2,3\}$

Co-Domain : $\{1,2,3,4,5,6\}$

Range : $\{2,4,6\}$

• Linear transformation

Conditions $\left[\begin{array}{l} T(u+v) = T(u) + T(v) \\ T(cu) = cT(u) \quad (c \text{ is scalar}) \end{array}\right.$

$*$ $T(x) = ax + b \quad (a, b \in \mathbb{R})$ is affine, not linear

↳ can use homogenous systems instead

• Matrix of Linear Transformation

if $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(x) = Ax$ for all $x \in \mathbb{R}^n$

standard Matrix of linear transformation T : $A = [T(e_1) \cdots T(e_n)]$

• onto and one-to-one

onto : Co-Domain = Range , also called surjective

one-to-one : Co-Domain = Range & one-to-one match, both surjective and injective

Least Squares

Method used in over-determined system $Ax=b$

↳ No solution, thus find $\underset{x}{\operatorname{argmin}} \|Ax-b\|^2$

- Inner product

  for $u, v \in \mathbb{R}^n$, $T(u,v) = u \cdot v = u^T v$

* Dot (Inner) product is linear transformation $\because (c_1u_1 + \cdots + c_nu_n) \cdot w$

$$= c_1(u_1 \cdot w) + \cdots + c_n(u_n \cdot w)$$

- Vector Norm

  $\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + \cdots + v_n^2}$

  $\|cv\| = c\|v\|$

- Unit vector

  $u = \dfrac{v}{\|v\|} \rightarrow \|u\| = 1$

- Distance between vectors

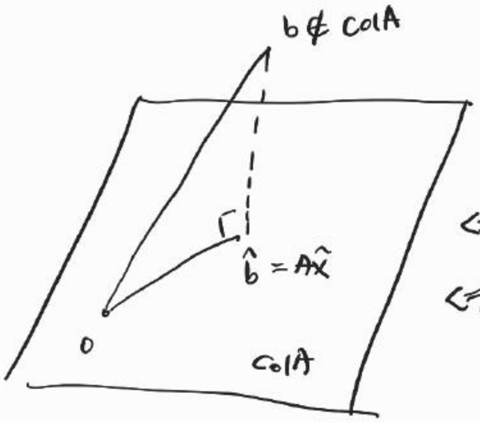
  $dist(u,v) = \|u-v\|$

- Orthogonal vectors

  $u \cdot v = \|u\|\|v\| \cos\theta$ $\therefore u \perp v \Rightarrow u \cdot v = 0$

- Least Square problem

  $\hat{x} = \underset{x}{\operatorname{argmin}} \|b-Ax\|$



$\therefore b - A\hat{x} \perp (x_1a_1 + x_2a_2 + \cdots + x_na_n)$

$\Leftrightarrow (b-A\hat{x}) \perp a_1, \ldots, (b-A\hat{x}) \perp a_n$

$\Leftrightarrow A^T(b-A\hat{x}) = 0$

- Normal Equation

  Solution of least squares : $A^TAx = A^Tb$

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

- Another Derivation of Normal Equation

$$\hat{x} = \underset{x}{\arg\min} \|b - Ax\| = \underset{x}{\arg\min} \|b - Ax\|^2$$

$$\|b-Ax\|^2 = (b-Ax)^T(b-Ax) = b^Tb - x^TA^Tb - b^TAx + x^TA^TAx$$

$$\therefore \frac{d}{dx}\|b-Ax\|^2 = -A^Tb - A^Tb + 2A^TA = 0$$

$$\underline{\quad A^TAx = A^Tb \quad}$$

- A^TA : Non-invertible case

= when ColA is linearly dependent = A^TA is not full rank

This is very rare!

- Orthogonal projection perspective

Projection of b onto colA space : $\hat{b} = f(b) = A\hat{x} = A(A^TA)^{-1}A^Tb$

- orthogonal and orthonormal sets

	Linearly independent	$u_1 \cdot u_2 = 0$	$\|u_1\| = \|u_2\| = 1$
orthogonal	✓	✓	×
orthonormal	✓	✓	✓

Basis vectors set

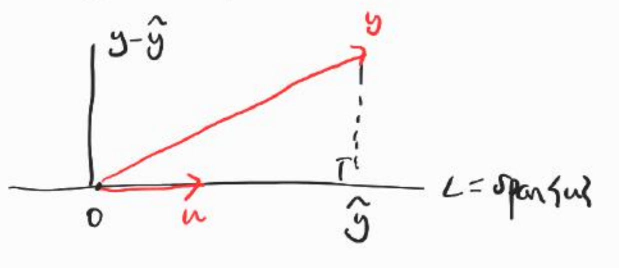
$u_1, u_2, \cdots, u_n$ $\xrightarrow{\text{QR decomposition}}$ orthogonal basis vectors $q_1, q_2, \cdots, q_n$

$\mathrm{span}\{u_1 \cdots u_n\} = W \in \mathbb{R}^n$

$$\begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix} y = \text{Projection of } y \text{ onto } W$$

- Further orthogonal projections

(i) Line



$$\hat{y} = \mathrm{proj}_L y = \underbrace{\frac{y \cdot u}{u \cdot u}}_{\text{scale}} \underbrace{u}_{\text{direction}}$$

(ii) plane

for $W = \text{span}\{u_1, u_2\}$, $\hat{y} = \text{Proj}_W y = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2$

- Transformation : orthogonal projection

for orthonormal basis vectors u_1, u_2, $\text{span}\{u_1, u_2\} = W$

$$\hat{b} = f(b) = (b \cdot u_1) u_1 + (b \cdot u_2) u_2$$

$$= u_1^T u_1 b + u_2^T u_2 b$$

$$= (u_1^T u_1 + u_2^T u_2) b$$

$$= [u_1 \ u_2] \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} b \quad = UU^T b \quad \Rightarrow \text{Linear transformation}$$

* Orthogonal projection of vector b onto $\text{Col}A = \text{Col}U$ $(A = U = [u_1 \ u_2])$

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

Since $A^T A = \begin{bmatrix} u_1^T \\ u_2^T \end{bmatrix} [u_1 \ u_2] = I$,

$$\hat{b} = A\hat{x} = AA^T b = UU^T b$$

- Gram-Schmidt Orthogonalization

Basis vectors (linearly independent) $\rightarrow$ orthogonal basis vectors

$W = \text{span}\{x_1, x_2\}$ $\quad \begin{bmatrix} v_1 = x_1 \\ v_2 = x_2 - \frac{x_2 \cdot x_1}{x_2 \cdot x_2} x_1 \end{bmatrix}$ $\quad W = \text{span}\{v_1, v_2\}$, $v_1 \perp v_2$

Eigenvectors and Eigenvalues

$Ax = \lambda x$: Linear transformation A only changes magnitude, not direction

$(A - \lambda I)x = 0$

$\Rightarrow$ Non-trivial solution exists iff $A - \lambda I$ is non-invertible $\therefore \det(A - \lambda I) = 0$

characteristic Equation

- Null space

Nul A : Solution set of $Ax = 0$ $\Leftrightarrow$ $A = \begin{bmatrix} -a_1^T- \\ \vdots \\ -a_n^T- \end{bmatrix}$, $a_1^T x = \cdots = a_n^T x = 0$

- Orthogonal Complement

set of vector z that is orthogonal to subspace W = $W^\perp$

$\text{Nul}\,A = (\text{Row}\,A)^\perp$

$\text{Nul}\,A^T = (\text{col}\,A)^\perp$

- Eigenspace

Null space of $A - \lambda I$ = Eigenspace of λ

if dim of eigenspace ≥ 1, for all vectors $x \in$ Eigenspace, $T(x) = Ax = \lambda x$

- Diagonalization

$D = V^{-1}AV$ $(A, V, D \in \mathbb{R}^{n \times n})$

Required conditions
- A, V are square matrix
- V is invertible matrix (linearly independent column vectors)

$\Rightarrow$ then, V's column vectors are A's eigenvectors & D has diagonal eigenvalues

$D = V^{-1}AV \Rightarrow VD = AV$

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

$$VD = [v_1 \ v_2 \ \cdots \ v_n] \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} = [\lambda_1 v_1 \ \lambda_2 v_2 \ \cdots \ \lambda_n v_n]$$

$\therefore Av_1 = \lambda_1 v_1, \cdots, Av_n = \lambda_n v_n$ $\Rightarrow$ n different A's eigenvector/eigenvalue pairs

- Eigendecomposition

A is diagonalizable $\iff$ Eigendecomposition of A

$D = V^{-1}AV$ $\qquad$ $A = VDV^{-1}$

- Linear Transformation via Eigendecomposition

$$\underbrace{T(x)}_{\text{Linear Transformation}} = Ax = VDV^{-1}x = V(D(V^{-1}x))$$

- change of basis

$$x = \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad \text{basis} = \left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$$

if $y = Vx = \begin{bmatrix} | & | \\ v_1 & v_2 \\ | & | \end{bmatrix} x$, new basis $= \{v_1, v_2\}$

- Element-wise Scaling

$A = VDV^{-1}$

V^{-1} → change of basis (basis $b_1 \to b_2$)

D → Scaling w/ diagonal element

V → Reverse change of basis (basis $b_2 \to b_1$)

- Linear Transformation via A^K

if A is diagonalizable, $A^K = (VDV^{-1}) \cdots (VDV^{-1}) = VD^KV^{-1}$

where $D^K = \mathrm{diag}([\lambda_1^K, \lambda_2^K, \cdots, \lambda_n^K])$

- Geometric multiplicity and Algebric Multiplicity

For $A \in \mathbb{R}^{n \times n}$,

A is diagonalizable if $\det(A - \lambda I) = 0$ has n real solutions (= n independent eigenvectors)

else if for repeated solutions e.g. $(\lambda - 2)^2 = 0$, requires to have

Algebric Multiplicity (# of repeatition) = Geometric Multiplicity (# of eigenspace dim)

Given λ, calculate # of linearly independent 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

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

↓ Then,

Reduced form of SVD

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

• 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!

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

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

$$AV = U\Sigma \iff A = U\Sigma V^T$$

• Computing SVD

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

> they share equal Σ

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

square root of corresponding eigenvalues = singular value σ

• Diagonalization of Symmetric Matrices

Symmetric matrix $S \in \mathbb{R}^{n \times n}$ $(S^T = S)$ is always diagonalizable

* Spectral Theorem

for $S^T = S \in \mathbb{R}^{n \times n}$, S has n eigenvalues (including repeated roots)

Also, eigenspace's dimension $= AM = GM$

Finally, eigenspace of different λ is orthogonal to each other

$\therefore$ S is orthogonally diagonalizable

- Spectral Decomposition

$$S = UDU^{-1} = UDU^T = [u_1\ u_2\ \cdots\ u_n] \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix}$$

$$= [\lambda_1 u_1\ \lambda_2 u_2\ \cdots\ \lambda_n u_n] \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix}$$

Scaling λ_1

$$= \lambda_1 \underline{u_1 u_1^T} + \cdots + \lambda_n u_n u_n^T$$

Projection to span$\{u_1\}$ direction

- Symmetric Positive Definite Matrices

if S is both symmetric and positive definite, spectral decomposition's eigenvalues are always positive

$S = \lambda_1 u_1 u_1^T + \cdots + \lambda_n u_n u_n^T$ where $\forall \lambda_j > 0$ for $j = 1, 2, \cdots, n$

- Back to Computing SVD

A^TA, AA^T are always symmetric. They are also positive semi-definite,

$x^TAA^Tx = (A^Tx)^T(A^Tx) = \|A^Tx\| \geq 0$

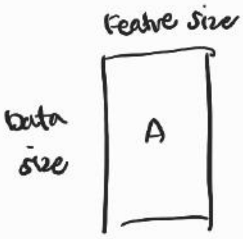
$x^TA^TAx = (Ax)^T(Ax) = \|Ax\| \geq 0$

$\therefore$ As $AA^T = U\Sigma^2U^T$, $A^TA = V\Sigma^2V^T$, singular values are always positive

	Rectangular Matrix	Square Matrix	symmetric Positive-definite matrix
Eigendecomposition	Not possible	Possible	Always possible ‖ same result
SVD decomposition	Always possible	Always possible	Always possible

- Eigendecomposition in ML

Usually use symmetric positive-definite matrix



$\rightarrow$ A^TA : Data relationships

AA^T : Feature relationships $\rightarrow$ Covariance Matrix in PCA

- Low rank approximation of a Matrix

A (rank r) $\xrightarrow{LRA}$ $\hat{A}_r$ (rank $r' < r$)

$$\hat{A}_r = \underset{\hat{A}_r}{\mathrm{argmin}} \|A - A_r\|_F \quad (\text{rank } \hat{A}_r \le r)$$

$$\Rightarrow \hat{A}_r = \sum_{i=1}^{r'} \sigma_i u_i v_i^T \quad (\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_{r'})$$

- Dimension Reducing Transformation

$A \in \mathbb{R}^{m \times n}$, $G \in \mathbb{R}^{m \times r}$ $\quad G^T : x \in \mathbb{R}^m \mapsto y \in \mathbb{R}^n$ $\quad \therefore y_i = G^T a_i$

↳ G's column vectors are orthonormal vectors

Then, Dimension Reducing Transformation is done by reserving $S = A^TA$ similarity.

$$S = A^TA, \quad Y = G^TA \Rightarrow Y^TY = (G^TA)^T(G^TA) = A^TGG^TA$$

$$\therefore \hat{G} = \underset{G}{\mathrm{argmin}} \|S - A^TGG^TA\|_F$$

For $A = U\Sigma V^T = \sum \sigma_i u_i v_i^T$, $\quad \hat{G} = U_r = [u_1\ u_2 \cdots u_r]$

Matrix Algebra

- Identity Matrix

$$I = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots & \\ & & & 1 \end{bmatrix} \in \mathbb{R}^{n \times n} \;/\; \text{for } \forall x \in \mathbb{R}^n ,\; Ix = x$$

- Transpose Matrix

if $[A]_{ij} = a_{ij}$, $[A^T]_{ij} = a_{ji}$

- Determinant

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij}$$

where $C_{ij} = (-1)^{i+j} M_{ij}$ & M_{ij} = determinent of submatrix w/o row i, col j (minor of a_{ij})

- Inverse Matrix

$A^{-1}A = AA^{-1} = I$ $\quad$ A^{-1} only exists iff non-singular and $\det A \neq 0$

$$A^{-1} = \frac{C^T}{\det(A)}$$ where $C \in \mathbb{R}^{n \times n}$ cofactor matrix ($C_{ij} = (-1)^{ij} M_{ij}$)

- Trace of Matrix

$$tr(A) = \sum_i [A]_{ii}$$

- Diagonal Matrix

$$A = \begin{bmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{bmatrix} \rightarrow A^{-1} = \begin{bmatrix} a_{11}^{-1} & & & 0 \\ & a_{22}^{-1} & & \\ & & \ddots & \\ 0 & & & a_{nn}^{-1} \end{bmatrix} \;\&\; \det(A) = \prod_{i=1}^{n} \det(A_{ii})$$

- Idempotent matrix

$A \in \mathbb{R}^{n \times n}$ s.t. $A^2 = A$ $\quad$ ($\therefore A^l = A$ for $l \geq 0$)

e.g. least squares $A = H(H^TH)^{-1}H^T$, $A^2 = A$

- Skew-Symmetric matrix

for vector $v = [a, b, c]^T$, skew-symmetric matrix $[v]_\times = \begin{bmatrix} 0 & -c & b \\ c & 0 & a \\ -b & c & 0 \end{bmatrix}$

$[v]_\times w = v \times w$ / for $R \in SO(3)$, $[Rv]_\times = R[v]_\times R^T$

- Positive definite matrix

for $\forall x \neq 0$ and $A \in \mathbb{R}^{n\times n}$, $x^TAx > 0$ (if $x^TAx \geq 0$, A is positive semi-definite)

A is positive definite $\Leftrightarrow$ for full rank $A = CC^T$, A's eigenvalues are all positive

& A's leading principal minors are all positive

- Toeplitz Matrix

$A_{ij} = a_{i-j}$

$$A = \begin{bmatrix} a_0 & a_{-1} & \cdots & a_{-(n-1)} \\ & a_0 & & \\ \vdots & & \ddots & \vdots \\ a_{n-1} & & \cdots & a_0 \end{bmatrix}$$

Add	$O(n)$
Mul	$O(n^2)$
Solve $Ax=b$	$O(n^2)$
$\det(A)$	$O(n^2)$