Linear Algebra Review

- Basic Concept:

  (1) represent / operate the linear system.

  (2) Notations:

$$\begin{cases} a_{11}x_1 + \cdots \quad a_{1n}x_n = b_1 \\ a_{21}x_1 + \cdots \quad a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + \cdots \quad a_{mn}x_n = b_m \end{cases} \Leftrightarrow \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$Ax = b$$

- Matrix Multiplication:

  (1) $A \in M_{m \times n}$ $\quad B \in M_{n \times p}$

$$C = AB \in M_{m \times p}$$

  (2) Vector-Vector Product:

$$C_{ij} = a_i^T b_j = \sum_{k=1}^{n} A_{ik} B_{kj}$$

  (3) Matrix-Vector Product: (left-row ; right-column)

$$C_{:,j} = A b_j$$

$$C = [A b_1, A b_2, \cdots A b_n]$$

*: linear combination of column vector.

$$C_{i:,} = a_i^T B$$

$$C = \begin{bmatrix} a_1^T B \\ \vdots \\ a_m^T B \end{bmatrix}$$

(4) Matrix - Matrix Products:

$A = [a_1, a_2 \cdots a_n]$

$$B = \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_n^T \end{bmatrix}$$

$C = \sum_{i=1}^{n} a_i b_i^T$ ($a_i b_i^T \Rightarrow$ matrix)

(5) Property:

$(AB)C = A(BC)$

$A(B+C) = AB + AC$

$AB \neq BA$ (not for all)

- Operations and Properties:

(1) Identity matrix:

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Diagonal matrix:

$D = \mathrm{diag}(d_1, d_2, \cdots d_n)$

(2) Transponse:

$(A+B)^T = A^T + B^T$

$(AB)^T = B^T A^T$

(3) Symmetric:

$A = A^T$ $A = -A^T$ (anti-symmetric)

$\forall A \in M_{n \times n}$. $A + A^T$ is symmetric.

(4) Trace:

$A \in M_{n\times n}$. $tr(A) = \sum_{i=1}^{n} A_{ii}$

① $tr(A) = tr(A^T)$

② $tr(A+B) = tr(A) + tr(B)$

③ $tr(AB) = tr(BA)$

$*$: $\nexists A, B$, $AB - BA = I$

(5) Norms:

$\|X\|_p = \left(\sum_{k=1}^{n} |X_k|^p \right)^{\frac{1}{p}}$

Frobenius norm:

$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} = \sqrt{tr(A^T A)}$

◦ Linear Independence and Rank:

(1) linearly independent:

$a_1 x_1 + a_2 x_2 + \cdots a_n x_n = 0$

$\Leftrightarrow a_1 = a_2 = \cdots a_n = 0$.

(2) column rank = row rank = rank(A)

$*$: elimination $\Rightarrow$ don't change rank.

(3) Properties:

① $rank(A) \le \min(m, n)$.

② $rank(A) = rank(A^T)$

③ $rank(AB) \le \min(rank(A), rank(B))$

④ $rank(A+B) \le rank(A) + rank(B)$

- Inverse:

(1) $AA^{-1} = I = A^{-1}A$

A: invertible $\Leftrightarrow$ non-singular.

(2) $\exists A^{-1} \Leftrightarrow \operatorname{rank}(A) = n$

(3) $(AB)^{-1} = B^{-1}A^{-1}$

(4) $(A^{-1})^T = (A^T)^{-1}$

- Orthogonal Matrix:

(1) orthogonal: $x^T y = 0$

(2) orthogonal matrix:

$A \in M_{n\times n}$. $A = [\alpha_1 . \alpha_2 \cdots \alpha_n]$

$$\alpha_i^T \alpha_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

$A^T A = I = A A^T$

$\Rightarrow A^T = A^{-1}$

(3) property:

$\|Ux\| = \|x\|$

Proof: $x^T U^T U x = x^T x = \|x\|^2$ $\square$

- Range and Nullspace:

(1) span: $\langle x_1 . x_2 \cdots x_n \rangle$

$$v = \sum_{i=1}^{n} \alpha_i x_i \quad \forall \alpha_i \in R$$

(2) $y \in R^m \quad x \in R^n$

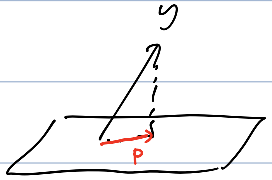
projection of y

$p = \text{argmin}_{v \in span} \|y - v\|_2$

(3) range: span of column vectors.

$R(A) = \{ v \in R^m : v = Ax, x \in R^n \}$

(4) $y \xrightarrow{\text{project}} R(A)$



$p \in \text{range}(A) \Rightarrow p = Ax$

$\forall v \in \text{range}(A) \quad v^T(y-p) = 0.$

$\therefore A^T(y - Ax) = 0$

$\therefore A^T A x = A^T y \iff x = (A^T A)^{-1} A^T y$

$\therefore p = A(A^T A)^{-1} A^T y.$

Let $Q = A(A^T A)^{-1} A^T$

We got: ① $Q^T = Q$ ② $Q^2 = Q$

*: one-dimensional: $Q = \frac{aa^T}{a^T a}$

(5) Nullspace: (or kernel)

$N(A) = \{ x \in R^n \mid Ax = 0 \}.$

*: Fundamental Theorem

$R(A)^\perp = N(A^T)$

$R(A^T)^\perp = N(A)$

*: orthogonal complements.

- Determinant:

(1) intuitively: $f: \mathbb{R}^{n\times n} \to \mathbb{R}$.

propeties (instead of big formula)

① $f(I) = 1$

② $A = \begin{bmatrix} \alpha_1^T \\ \vdots \\ \alpha_n^T \end{bmatrix}$

$$\begin{bmatrix} t\alpha_1^T \\ \vdots \\ \alpha_n^T \end{bmatrix} = t f(A)$$

③ $$\begin{bmatrix} \alpha_2^T \\ \alpha_1^T \\ \vdots \\ \alpha_n^T \end{bmatrix} = -f(A)$$

(2) Properties:

① $|A| = |A^T|$

② $|AB| = |A| \cdot |B|$

③ $|A| = 0 \iff A$ is non-invertible

④ $|A^{-1}| = \frac{1}{|A|}$ $(|A| \neq 0)$

(3) Cofactor:

$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|$

A_{ij}: delete i-row, j-column

recursively: $|A|$ has $n!$ terms.

(4) Adjoint:

$\mathrm{adj}(A) \in \mathbb{R}^{n\times n}$ $A \in \mathbb{R}^{n\times n}$

$[\mathrm{adj}(A)]_{ij} = (-1)^{i+j} |A_{ji}|$

$A \cdot \mathrm{adj}(A) = |A|$

$\Leftrightarrow A^{-1} = \frac{1}{|A|} \cdot \mathrm{adj}(A)$

◦ Quadratic Forms and Positive Semidefinite

(1) quadratic form: $A \in \mathbb{R}^{n \times n}$

$x^T A x = \sum_{i=1}^{n} A_{ij} x_i x_j \in \mathbb{R}$

$x^T A x = (x^T A x)^T = x^T A^T x = x^T (\frac{1}{2}A + \frac{1}{2}A^T) x$

*: symmetric part.

(2) definition:

① positive definite:

for non-zero vectors x. $x^T A x > 0$

② positive semi-definite:

for all vectors x: $x^T A x \geq 0$

PD $\Rightarrow$ PSD

(3) Positive definite $\Rightarrow$ invertible.

Proof: contradiction.

A: not invertible. $A = [\alpha_1 . \alpha_2 \cdots \alpha_n]$.

$\exists x_i$: $\sum_{i=1}^{n} x_i \alpha_i = 0$.

Let $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ $x \neq 0$. $x^T A x = 0$. $\square$.

(4) Gram Matrix:

$G = A^T A$. $\forall x \in \mathbb{R}^n$. $x^T A^T A x \geq 0$.

$\therefore G$ is always positive semi-definite.

- Eigenvalues & Eigenvectors:

(1) Definition:

$Ax = \lambda x$.

λ: eigenvalue $\quad x$: corresponding eigenvector

*: $\lambda, x \in \mathbb{C}$

$\Leftrightarrow (A - \lambda I)x = 0 \quad (x \neq 0)$

non-zero solution $\Leftrightarrow$ non-empty nullspace.

$\Leftrightarrow (\lambda I - A)$ is singular

$\Leftrightarrow \det(\lambda I - A) = 0$

*: find the roots of polynomials.

$\Leftrightarrow \lambda_1, \lambda_2, \lambda_3 \cdots \lambda_n$

(2) Properties:

① $tr(A) = \sum_{i=1}^{n} \lambda_i$

② $\det(A) = \prod_{i=1}^{n} \lambda_i$

③ rank(A) $\Leftrightarrow$ non-zero eigenvalues.

④ $Ax = \lambda x \Leftrightarrow x = \lambda A^{-1} x \Leftrightarrow \frac{1}{\lambda} x = A^{-1} x \quad (\lambda \neq 0)$

$\therefore \frac{1}{\lambda}$ is the eigenvalue of A^{-1}

⑤ eigenvalue of $D = diag(d_1, d_2 \cdot d_n)$

(3) $A x_1 = \lambda_1 x_1$

$A x_2 = \lambda_2 x_2 \quad \Rightarrow \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$

$\cdots$

$Ax_n = \lambda_n x_n$

$\Leftrightarrow AX = X\Lambda$

If the eigenvectors are linearly independent

$\Leftrightarrow A = X\Lambda X^{-1}$ (diagonalizable)

- Eigenstaff of Symmetric Matrices:

(1) 2 Amazing Properties: $A \in S^n$

① All eigenvalues are real

$Ax = \lambda x \Rightarrow A\bar{x} = \bar{\lambda}\bar{x}$.

$\bar{x}^T A^T = \bar{x}^T \bar{\lambda}^T = \bar{x}^T \bar{\lambda}$

$\Leftrightarrow \bar{x}^T A^T x = \bar{x}^T \bar{\lambda} x$

$\Leftrightarrow \bar{x}^T A x = \bar{x}^T \boxed{\bar{\lambda}} x \quad (A = A^T)$

$\bar{x}^T A x = \bar{x}^T \boxed{\lambda} x$

$\bar{\lambda} = \lambda \Rightarrow$ real eigenvalues.

② Eigenvectors are orthonormal.

$Ax_1 = \lambda_1 x_1$. $Ax_2 = \lambda_2 x_2$.

if $\lambda_1 \neq \lambda_2$. then $x_1^T x_2 = 0$

$x_2^T A x_1 = \lambda_1 x_2^T x_1 \quad \cdots\cdots ①$

$x_1^T A x_2 = \lambda_2 x_1^T x_2 \quad \cdots\cdots ②$

① take transpose:

$x_1^T A x_2 = \lambda_1 x_1^T x_2 = \lambda_2 x_1^T x_2$

$\Leftrightarrow (\lambda_1 - \lambda_2) x_1^T x_2 = 0$

$\because \lambda_1 \neq \lambda_2 \quad \therefore x_1^T x_2 = 0$

(2) Definiteness → sign of eigenvalues.

$x^T A x = x^T U \Lambda U^T x = y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2$

① $\forall \lambda_i > 0 \Rightarrow$ PD

② $\forall \lambda_i \geq 0 \Rightarrow$ PSD

(3) Applications: $(A \in S)$

maximizing function of a matrix.

e.g. $\max_{x \in R^n} x^T A x.$

subject to: $\|x\|_2^2 = 1$

◦ Matrix Calculus:

(1) Gradient:

$f: R^{m \times n} \to R.$

$$\nabla_A f(A) \in R^{m \times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & & & \\ \vdots & & & \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

*: gradient is well-defined

when the output is scalar value.

(2) Hessian:

$f: R^n \to R$

$$\nabla_x^2 f(x) \in R^{n \times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & & & \\ \vdots & & & \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

*: partial derivatives

$[\nabla_x^2 f(x)]_{ij} = \frac{\partial f(x)}{\partial x_i \partial x_j}$

Hessian → symmetric.

(3) Analogue:

gradient → first derivative

Hessian → second derivative.

- Hessian of Quadratic and Linear Function.

(1) $x \in \mathbb{R}^n$. $f(x) = b^T x \in \mathbb{R}$.

$$f(x) = \sum_{i=1}^{n} b_i x_i$$

$$\frac{\partial f}{\partial x_i} = b_i$$

$$\nabla_x b^T x = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = b \quad \Rightarrow \quad \nabla_x b^T x = b$$

(2) $x \in \mathbb{R}^n$. $A \in S^n$.

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

$$\frac{\partial f}{\partial x_k} = \sum_{j=1}^{n} a_{ik} x_j + \sum_{i=1}^{n} a_{ki} x_i = 2 \sum_{i=1}^{n} a_{ki} x_i$$

*: inner product of k-th row of A

$$\nabla_x x^T A x = 2Ax.$$

(3) Hessian of the quadratic function.

$$\frac{\partial f}{\partial x_k} = 2 \sum_{i=1}^{n} a_{ki} x_i$$

$$\frac{\partial f}{\partial x_k \partial x_l} = 2 a_{kl}$$

Therefore: $\nabla_x^2 (x^T A x) = 2A$

(analogue: $(ax^2)'' = 2a$)

(4) Summary:

$$\boxed{\nabla_x b^T x = b}$$

$$\boxed{\begin{aligned}&\nabla_x x^T A x = 2Ax\\&\nabla_x^2 (x^T A x) = 2A\end{aligned}}$$

- Least Square:

(1) $A \in \mathbb{R}^{m \times n}$ (assuming full rank)

$\exists\, b \in \mathbb{R}^m$. $b \notin \mathrm{Ran}(A)$.

$Ax = b \Rightarrow$ no solution.

(2) find a vector x

$x = \arg\min \|Ax - b\|_2^2$

$\|Ax - b\|_2^2 = (Ax - b)^T (Ax - b)$

$= (x^T A^T - b^T)(Ax - b)$ $\quad (x^T A^T b = b^T A x)$

$= x^T A^T A x + b^T b - x^T A^T b - b^T A x$

$= x^T A^T A x - 2 b^T A x + b^T b$

$\therefore \nabla_x (\|Ax - b\|_2^2) = 2A^T A x - 2A^T b = 0$

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

- Gradients of the Determinant (TBC)

(1) find grdient with respect to a matrix.

$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{\setminus i, \setminus j}|$

$\frac{\partial}{\partial A_{kl}} |A| = (-1)^{k+l} a_{kl} |A_{\setminus k, \setminus l}| = (\mathrm{adj}(A))_{lk}$

$\nabla_A |A| = (\mathrm{adj}(A))^T = |A| A^{-T}$

(2) function: $f(A) = \log |A|$.

domain: PD matrices.

$$\frac{\partial \log |A|}{\partial A} = \frac{\partial \log |A|}{\partial |A|} \cdot \frac{|A|}{\partial A} = \frac{1}{|A|} \cdot |A| \cdot A^{-T}$$

$A = A^T \Rightarrow A^{-T} = A^{-1}$

Therefore: $\frac{\partial \log |A|}{\partial A} = A^{-1}$

- Eigenvalues as Optimization:

(1) $\max_{x \in \mathbb{R}^n} x^T A x$

subject to $\|x\|_2^2 = 1$, $A \in S^n$

(2) Lagrangian Form:

$L(x, \lambda) = x^T A x - \lambda x^T x$ (Lagrangian multiplier)

$\nabla_x L(x, \lambda) = 2Ax - 2\lambda x = 0$

$\Leftrightarrow Ax = \lambda x.$

$\therefore x \rightarrow$ eigenvector.