Linear Algebra

1 Determinant

1.1 Expansion formula

For any $A \subseteq \mathcal{Y}$,

$$\sum_{A \subseteq Y \subseteq \mathcal{Y}} \det(\boldsymbol{L}_Y) = \det(\boldsymbol{L} + \boldsymbol{I}_{\bar{A}}), \tag{1.1}$$

1.2 Rearrangement

$$\sum_{(I',J') \in \mathcal{S}(I,J)} \det(\boldsymbol{Z}_{Y,I'}) \det(\boldsymbol{Z}_{Y,J'}) \leq \sum_{(I',*) \in \mathcal{S}(I,j)} \det(\boldsymbol{Z}_{Y,I'})^2 \tag{1.2}$$

where $\boldsymbol{I}_{\bar{A}}$ is the diagonal matrix with ones in the diagonal positions with indices in $\bar{A}$ and zeros elsewhere.

1.3 Weinstein-Aronszajin identity

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices of size $m \times n$ and $n \times m$ respectively, given that $\boldsymbol{AB}$ is of trass class, then

$$\det(\boldsymbol{I} + \boldsymbol{AB}) = \det(\boldsymbol{I} + \boldsymbol{BA}) \tag{1.3}$$

1.4 DPP related

1.4.1 Propsosal matrix for NDPP

Given $\boldsymbol{V}, \boldsymbol{B}, \boldsymbol{D}$ such that $\boldsymbol{L} = \boldsymbol{V}\boldsymbol{V}^\top + \boldsymbol{B}(\boldsymbol{D} - \boldsymbol{D}^\top)\boldsymbol{B}^\top$, let $\{\rho_i, \boldsymbol{v}_i\}_{i=1}^K$ be the eigendecomposition of $\boldsymbol{V}\boldsymbol{V}^\top$ and $\{(\sigma_j, \boldsymbol{y}_{2j-1}, \boldsymbol{y}_{2j})\}$ be the Youla decomposition of $\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{D}^\top)\boldsymbol{B}^\top$. Denote $\boldsymbol{Z} := [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_K] \in \mathbb{R}^{M \times 2K}$ and

$$\boldsymbol{X} := \operatorname{diag}\left(\rho, \ldots, \rho_K, \begin{bmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{bmatrix}, \ldots \begin{bmatrix} 0 & \sigma_{K/2} \\ -\sigma_{K/2} & 0 \end{bmatrix}\right),$$
$$\hat{\boldsymbol{X}} := \operatorname{diag}\left(\rho, \ldots, \rho_K, \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_1 \end{bmatrix}, \ldots \begin{bmatrix} \sigma_{K/2} & 0 \\ 0 & \sigma_{K/2} \end{bmatrix}\right),$$

so that $\boldsymbol{L} = \boldsymbol{Z}\boldsymbol{X}\boldsymbol{Z}^\top$ and $\hat{\boldsymbol{L}} = \boldsymbol{Z}\hat{\boldsymbol{X}}\boldsymbol{Z}^\top$. Then, for every subset $\boldsymbol{Y} \subseteq [M]$, it holds that

$$\det(\boldsymbol{L}_Y) \leq \det(\hat{\boldsymbol{L}}_Y) \tag{1.4}$$

and the equality holds when the size of $\boldsymbol{Y}$ is equal to the rank of $\boldsymbol{L}$.

1.4.2 Propsosal matrix for NDPP II

Given $\boldsymbol{X} \in \mathbb{R}^{n \times d}$ and $\boldsymbol{W}^A \in \mathbb{R}^{d \times d}$ Then,

$$\det([\boldsymbol{X}\boldsymbol{W}^A\boldsymbol{X}^\top]_S) \leq \det([\boldsymbol{X}\hat{\boldsymbol{W}}^A\boldsymbol{X}^\top]_S) \tag{1.5}$$

for every $S \subseteq [n]$. In addition, equality holds when $|S| \geq d$.

1.4.3 DPP probability expansion

$$\mathbb{P}_{\hat{\boldsymbol{L}}}(Y) = \frac{\det(\hat{\boldsymbol{L}}_Y)}{\det(\hat{\boldsymbol{L}} + \boldsymbol{I})} = \sum_{E \subseteq [2K], |E|=|Y|} \det(\underbrace{\boldsymbol{Z}_{Y,E}\boldsymbol{Z}_{Y,E}^\top}_{\text{elementary DPP}}) \prod_{i \in E} \frac{\lambda_i}{\lambda_i + 1} \prod_{i \notin E} \frac{1}{\lambda_i + 1} \tag{1.6}$$

1. Choose an elementary DPP according to its mixture weight
2. Sample a subset from the selected elementary DPP

1.4.4 DPP probability expansion II

The probability of sampling $S \in \binom{n}{k}$ from the k-DPP with $\hat{\boldsymbol{L}}$ can be decomposed into the following

$$\frac{\det(\hat{\boldsymbol{L}}_S)}{e_k(\{\lambda_i\}_{i=1}^d)} = \sum_{E \in \binom{[d]}{k}} \frac{\prod_{i \in E} \lambda_i}{e_k(\{\lambda_i\}_{i=1}^d)} \cdot \det(\boldsymbol{K}_S^E) \tag{1.7}$$

where $\boldsymbol{K}^E$ is a rank-k projection matrix consisting of eigenvalues of $\hat{\boldsymbol{L}}$.

1.5 Ratio

Given that $\det(\boldsymbol{QSQ}^\top) \neq 0$

$$\frac{\det(\boldsymbol{Q}(\boldsymbol{S} + \boldsymbol{R})\boldsymbol{Q}^\top)}{\det(\boldsymbol{QSQ}^\top)} \leq \det(\boldsymbol{I}_2 + (\boldsymbol{QSQ}^\top)^{-1/2}\boldsymbol{QRQ}^\top(\boldsymbol{QSQ}^\top)^{-1/2}) \tag{1.8}$$

1.6 Inverse of trace

For an invertible matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$,

$$\operatorname{tr}(\boldsymbol{A}^{-1}) = \sum_{i=1}^n \det(\boldsymbol{A}_{-i}) / \det(\boldsymbol{A}), \tag{1.9}$$

where $\boldsymbol{A}_{-i} \in \mathbb{R}^{n-1 \times n-1}$ is the submatrix of $\boldsymbol{A}$ where the ith row and column of $\boldsymbol{A}$ are removed.

2 Vectorziation

$$\operatorname{vec}(\boldsymbol{AXB}) = (\boldsymbol{B}^\top \otimes \boldsymbol{A})\operatorname{vec}(\boldsymbol{X}) \tag{2.1}$$

Lyapunov Equation.

$$\begin{aligned} \boldsymbol{AX} + \boldsymbol{XB} &= \boldsymbol{C} && (2.2)\\ \boldsymbol{AXI} + \boldsymbol{IXB} &= \boldsymbol{C} && (2.3)\\ (\boldsymbol{I} \otimes \boldsymbol{A})\operatorname{vec}(\boldsymbol{X}) + (\boldsymbol{B}^\top \otimes \boldsymbol{I})\operatorname{vec}(\boldsymbol{X}) &= \operatorname{vec}(\boldsymbol{C}) && (2.4)\\ \operatorname{vec}(\boldsymbol{X}) &= (\boldsymbol{I} \otimes \boldsymbol{A} + \boldsymbol{B}^\top \otimes \boldsymbol{I})^{-1}\operatorname{vec}(\boldsymbol{C}) && (2.5) \end{aligned}$$

3 Trace

$$\operatorname{tr}(\boldsymbol{ABC}) = \operatorname{vec}(\boldsymbol{A}^\top)^\top(\boldsymbol{I} \otimes \boldsymbol{B})\operatorname{vec}(\boldsymbol{C}) \tag{3.1}$$

$$\operatorname{tr}(\boldsymbol{A}^\top\boldsymbol{BCD}^\top) = \operatorname{vec}(\boldsymbol{A})^\top(\boldsymbol{D} \otimes \boldsymbol{B})\operatorname{vec}(\boldsymbol{C}) \tag{3.2}$$

3.1 Von Neumann's trace inequqality

Theorem. If $\boldsymbol{A}, \boldsymbol{B}$ are complex $n \times n$ matrices with singular values

$$\alpha_1 \geq \cdots \geq \alpha_n, \quad \beta_1 \geq \ldots \beta_n, \tag{3.3}$$

respectively, then

$$|\mathrm{tr}\,(\boldsymbol{AB})| \leq \sum_{i=1}^{n} \alpha_i \beta_i \tag{3.4}$$

4 Inversion

4.1 Woodbury identity

$$[\boldsymbol{A} + \boldsymbol{BCD}]^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}[\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B}]^{-1}\boldsymbol{DA}^{-1} \tag{4.1}$$

given that $\boldsymbol{A}^{-1}$ and $\boldsymbol{C}^{-1}$ exist. If $\boldsymbol{B} = \boldsymbol{x}, \boldsymbol{C} = \boldsymbol{I}, \boldsymbol{D} = \boldsymbol{y}^\top$

$$(\boldsymbol{A} + \boldsymbol{xy}^\top)^{-1} = \boldsymbol{A}^{-1} - \frac{(\boldsymbol{A}^{-1}\boldsymbol{x})(\boldsymbol{yA}^{-1})}{1 + \boldsymbol{y}^\top \boldsymbol{A}^{-1}\boldsymbol{x}} \tag{4.2}$$

4.2 Schur Complement

Schur Complement essentially is a block Cholesky factorization of a matrix

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{BD}^{-1} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} (\boldsymbol{A} - \boldsymbol{BD}^{-1}\boldsymbol{C}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{D}^{-1}\boldsymbol{C} & \boldsymbol{I} \end{bmatrix} \tag{4.3}$$

$\boldsymbol{A} - \boldsymbol{BD}^{-1}\boldsymbol{A}$ is called the *Schur complement* of $\boldsymbol{D}$.

5 Hadamard Product

5.1 Quadratic Relation

$$\boldsymbol{x}^\top(\boldsymbol{A} \odot \boldsymbol{B})\boldsymbol{y} = \mathrm{tr}\,(\mathrm{Diag}(\boldsymbol{x})\boldsymbol{A}\mathrm{Diag}(\boldsymbol{y})\boldsymbol{B}^\top) \tag{5.1}$$

By setting $\boldsymbol{x} = \boldsymbol{y}$, it shows that the Hadamard product of two PSD matrices is PSD.

5.2 Rank Relation

$$rank(\boldsymbol{A} \odot \boldsymbol{B}) \leq rank(\boldsymbol{A})rank(\boldsymbol{B}) \tag{5.2}$$

5.3 Spectrum Relation

$$\prod_{i=k}^{n} \lambda_i(\boldsymbol{A} \odot \boldsymbol{B}) \geq \prod_{i=k}^{n} \lambda_i(\boldsymbol{AB}),\ \forall k = 1, \ldots, n \tag{5.3}$$

with $\lambda_i(\cdot)$ denotes PD matrix.

5.4 Determinant

$$|\boldsymbol{A} \odot \boldsymbol{B}| \geq |\boldsymbol{A}|\,|\boldsymbol{B}| \tag{5.4}$$

6 Matrix Calculus

6.1 Matrix Chain rule

$$[\nabla_{\boldsymbol{X}} f(g(\boldsymbol{X}))]_{ij} = \sum_{k=1}^{p} \sum_{\ell=1}^{q} \frac{\partial f(G)}{\partial g_{k\ell}} \frac{\partial g_{k\ell}}{x_{ij}} \tag{6.1}$$

6.2 Differentials

$$\mathrm{d}(\mathrm{tr}\, \boldsymbol{X}) = \mathrm{tr}\ \mathrm{d}\boldsymbol{X} \tag{6.2}$$
$$\mathrm{d}(\boldsymbol{X} \otimes \boldsymbol{Y}) = (\,\mathrm{d}\boldsymbol{X}) \otimes \boldsymbol{Y} + \boldsymbol{X} \otimes (\,\mathrm{d}\boldsymbol{Y}) \tag{6.3}$$
$$\mathrm{d}\boldsymbol{X}^{-1} = -\boldsymbol{X}^{-1} \cdot \mathrm{d}\boldsymbol{X} \cdot \boldsymbol{X}^{-1} \tag{6.4}$$
$$\mathrm{d}(\det(\boldsymbol{X})) = \mathrm{tr}\,(\mathrm{adj}(\boldsymbol{X})\,\mathrm{d}\boldsymbol{X}) \tag{6.5}$$
$$\mathrm{d}\det(\boldsymbol{X}) = \det(\boldsymbol{X})\mathrm{tr}\,(\boldsymbol{X}^{-1}\,\mathrm{d}\boldsymbol{X}) \tag{6.6}$$
$$\mathrm{d}\log(\det(\boldsymbol{X})) = \mathrm{tr}\,(\boldsymbol{X}^{-1}\,\mathrm{d}\boldsymbol{X}) \tag{6.7}$$
$$\mathrm{d}\sigma(a) = (\mathrm{Diag}(\sigma) - \mathrm{Diag}(\sigma)^2)\,\mathrm{d}a \tag{6.8}$$
$$\mathrm{d}(\mathrm{softmax}(\theta)) = (\mathrm{Diag}(\boldsymbol{y}) - \boldsymbol{y}\boldsymbol{y}^\top)\,\mathrm{d}\theta \tag{6.9}$$

Note: Elementwise function은 일단 Diagonal 형태로 바꿔서 생각해 보삼 ㅋ

6.3 Useful first derivatives

$$\frac{\partial \mathrm{tr}\, \boldsymbol{X}}{\partial \boldsymbol{X}} = \boldsymbol{I} \tag{6.10}$$
$$\frac{\partial \mathrm{tr}\, \boldsymbol{X}^{-1}}{\partial \boldsymbol{X}} = -\boldsymbol{X}^{-2} \tag{6.11}$$
$$\frac{\partial \mathrm{tr}\,(\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{A}^\top \tag{6.12}$$
$$\frac{\partial \mathrm{tr}\,(\boldsymbol{X}^k)}{\partial \boldsymbol{X}} = k \cdot (\boldsymbol{X}^\top)^{k-1} \tag{6.13}$$
$$\frac{\partial \mathrm{tr}\,(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})}{\partial \boldsymbol{X}} = \boldsymbol{B}^\top \boldsymbol{X}^\top \boldsymbol{A}^\top + \boldsymbol{A}^\top \boldsymbol{X}^\top \boldsymbol{B}^\top \tag{6.14}$$
$$\frac{\partial \boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}}{\partial \boldsymbol{X}} = -\boldsymbol{X}^\top \boldsymbol{A}^\top \boldsymbol{B}^\top \boldsymbol{X}^{-\top} \tag{6.15}$$
$$\frac{\partial \log\det(\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{X}^{-\top} \tag{6.16}$$
$$\frac{\partial \det(\boldsymbol{X}^{-1})}{\partial \boldsymbol{X}} = \frac{\boldsymbol{X}^{-\top}}{\det \boldsymbol{X}} \tag{6.17}$$
$$\frac{\partial \det(\boldsymbol{X}^k)}{\partial \boldsymbol{X}} = k\det(\boldsymbol{X}^k)\boldsymbol{X}^{-\top} \tag{6.18}$$
$$\frac{\partial \log\det(\boldsymbol{X}\boldsymbol{X}^\top)}{\partial \boldsymbol{X}} = 2\boldsymbol{X}[\boldsymbol{X}^\top \boldsymbol{X}]^{-1} \cdot \det(\boldsymbol{X}\boldsymbol{X}^\top) \tag{6.19}$$
$$\frac{\det(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})}{\partial \boldsymbol{X}} = \det(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})^{-\top}\boldsymbol{B}^\top \tag{6.20}$$

6.4 Quadratic form

$$\frac{\partial(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s})^\top \boldsymbol{W}(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s})}{\partial \boldsymbol{s}} = -2\boldsymbol{A}^\top \boldsymbol{W}(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s}) \tag{6.21}$$

$$\frac{\partial(\boldsymbol{x}-\boldsymbol{s})^\top \boldsymbol{W}(\boldsymbol{x}-\boldsymbol{s})}{\partial \boldsymbol{x}} = 2\boldsymbol{W}(\boldsymbol{x}-\boldsymbol{s}) \tag{6.22}$$

$$\frac{\partial(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s})^\top \boldsymbol{W}(\boldsymbol{s}-\boldsymbol{A}\boldsymbol{s})}{\partial \boldsymbol{x}} = 2\boldsymbol{W}(\boldsymbol{s}-\boldsymbol{A}\boldsymbol{s}) \tag{6.23}$$

$$\frac{\partial(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s})^\top \boldsymbol{W}(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s})}{\partial \boldsymbol{A}} = -2\boldsymbol{W}(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s})\boldsymbol{s}^\top \tag{6.24}$$

6.5 Hessian product rule

Given two functions $f, g : \mathbb{R}^n \to \mathbb{R}$,

$$H_c(fg) = (H_c f)g(c) + \nabla_c f^\top \nabla_c g + \nabla_c g^\top \nabla_c f + f(c) H_c g \tag{6.25}$$

6.6 Integration by parts

Given vector valued function φ and scalar function f with vanishing condition,

$$\int_{\mathbb{R}^d} \varphi(\boldsymbol{x}) \cdot \nabla f(\boldsymbol{x})\, \mathrm{d}x = -\int_{\mathbb{R}^d} (\nabla \cdot \varphi(\boldsymbol{x})) f(\boldsymbol{x})\, \mathrm{d}x \tag{6.26}$$

7 Eigenvalues and Eigenvectors

7.1 General Properties

Assume that $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$,

$$\mathrm{eig}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{eig}(\boldsymbol{B}\boldsymbol{A}) \tag{7.1}$$

$$\mathrm{rank}(\boldsymbol{A}) = r \Rightarrow \text{ At most } r \text{ non-zero } \lambda_i \tag{7.2}$$

7.2 Symmetric

Assume $\boldsymbol{A}$ is symmetric, then

$$\boldsymbol{V}\boldsymbol{V}^\top = \boldsymbol{I} \tag{7.3}$$

$$\lambda_i \in \mathbb{R} \tag{7.4}$$

$$\mathrm{tr}\,(\boldsymbol{A}^p) = \sum_i \lambda_i^p \tag{7.5}$$

$$\mathrm{eig}(\boldsymbol{I} + c\boldsymbol{A}) = 1 + c\lambda_i \tag{7.6}$$

$$\mathrm{eig}(\boldsymbol{A} - c\boldsymbol{I}) = \lambda_i - c \tag{7.7}$$

$$\mathrm{eig}(\boldsymbol{A}^{-1}) = \lambda_i^{-1} \tag{7.8}$$

For a symmetric, positive matrix $\boldsymbol{A}$

$$\mathrm{eig}(\boldsymbol{A}^\top\boldsymbol{A}) = \mathrm{eig}(\boldsymbol{A}\boldsymbol{A}^\top) = \mathrm{eig}(\boldsymbol{A}) \circ \mathrm{eig}(\boldsymbol{A}) \tag{7.9}$$

7.3 Singular Value Decomposition

Any $n \times m$ matrix $\boldsymbol{A}$ can be written as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top \tag{7.10}$$

where

$$\begin{aligned}
\boldsymbol{U} &= \text{ eigenvectors of } \boldsymbol{A}\boldsymbol{A}^\top \ n \times n \\
\boldsymbol{D} &= \sqrt{\text{diag}(\text{eig}(\boldsymbol{A}\boldsymbol{A}^\top))} \quad n \times m \\
\boldsymbol{V} &= \text{ eigenvectors of } \boldsymbol{A}^\top\boldsymbol{A} \ m \times m
\end{aligned}$$

7.3.1 Square decomposed into rectangular

Assume $\boldsymbol{V}_*\boldsymbol{D}_*\boldsymbol{U}_*^\top = 0$ then we can expand the SVD of $\boldsymbol{A}$ into

$$\boldsymbol{A} = \left[\begin{array}{c|c} \boldsymbol{V} & \boldsymbol{V}_* \end{array}\right] \left[\begin{array}{c|c} \boldsymbol{D} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{D}_* \end{array}\right] \left[\begin{array}{c} \boldsymbol{U}^\top \\ \hline \boldsymbol{U}_*^\top \end{array}\right] \tag{7.11}$$

where the SVD of $\boldsymbol{A}$ is $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^\top$

7.4 LU decomposition

Assume $\boldsymbol{A}$ is a square matrix with non-zero leading prcinipal minors, then

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U} \tag{7.12}$$

where $\boldsymbol{L}$ is a unique unit lower triangular matrix and $\boldsymbol{U}$ is a unique upper triangular matrix.

7.5 Cholesky decomposition

Assume $\boldsymbol{A}$ is a symmetric positive definite square matrix, then

$$\boldsymbol{A} = \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{L}\boldsymbol{L}^\top \tag{7.13}$$

where $\boldsymbol{U}$ is an unique upper triangular matrix and $\boldsymbol{L}$ is a lower triangular matrix.

7.6 Eigenvalues of its reverse

Proposition 7.1. Given $M \times K$ matrix $\boldsymbol{A}, \boldsymbol{B}$, the nonzero eigenvalues of $\boldsymbol{A}\boldsymbol{B}^\top \in \mathbb{C}^{M\times M}$ and $\boldsymbol{B}^\top\boldsymbol{A} \in \mathbb{C}^{K\times K}$ are identical. If addition, if $(\lambda, \boldsymbol{v})$ is an eigenpair of $\boldsymbol{B}^\top\boldsymbol{A}$ with $\lambda \neq 0$, then $(\lambda, \boldsymbol{A}\boldsymbol{v}/\left\|\boldsymbol{A}\boldsymbol{v}\right\|_2)$ is an eigenpair of $\boldsymbol{A}\boldsymbol{B}^\top$.

7.7 Row stochastic matrix

Fact. The operator norm of a row-stochastic matrix is 1.

8 Inverses

8.1 Rank-1 update of the inverse of inner product

Denote $\boldsymbol{A} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ and that $\boldsymbol{X}$ is extended to include a new column vector in the end $\tilde{\boldsymbol{X}} = [\boldsymbol{X}, \boldsymbol{v}]$, let $N = \boldsymbol{v}^\top(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top)\boldsymbol{v}$ then

$$(\tilde{\boldsymbol{X}^\top}\tilde{\boldsymbol{X}})^{-1} = N^{-1} \begin{bmatrix} N\boldsymbol{A} + \boldsymbol{A}\boldsymbol{X}^\top\boldsymbol{v}(\boldsymbol{A}\boldsymbol{X}^\top\boldsymbol{v})^\top & -\boldsymbol{A}\boldsymbol{X}^\top\boldsymbol{v} \\ -\boldsymbol{v}^\top\boldsymbol{X}\boldsymbol{A}^\top & 1 \end{bmatrix} \tag{8.1}$$

8.2 Approximations

The following identity is known as the *Neuman series* of a matrix, which holds when $|\lambda_i| < 1$ for all eigenvalues λ_i

$$(\boldsymbol{I} - \boldsymbol{A})^{-1} = \sum_{n=0}^{\infty} \boldsymbol{A}^n \tag{8.2}$$

$$(\boldsymbol{I} + \boldsymbol{A})^{-1} = \sum_{n=0}^{\infty} (-1)^n \boldsymbol{A}^n \tag{8.3}$$

$$\begin{aligned}
\boldsymbol{A} - \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A})^{-1}\boldsymbol{A} &= \boldsymbol{A} - \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A}^{-1})^{-1} && (8.4)\\
&= \boldsymbol{A}(\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{A}^{-1})^{-1}) && (8.5)\\
&\approx \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{I} + \boldsymbol{A}^{-1} - \boldsymbol{A}^{-2}) && (8.6)\\
&= \boldsymbol{I} - \boldsymbol{A}^{-1} && (8.7)
\end{aligned}$$

8.3 Block matrix

Using Schur complements

$$\boldsymbol{C}_1 = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} \tag{8.8}$$

$$\boldsymbol{C}_2 = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \tag{8.9}$$

as

$$\left[\begin{array}{c|c} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \hline \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} \boldsymbol{C}_1^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{C}_2^{-1} \\ \hline -\boldsymbol{C}_2^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{C}_2^{-1} \end{array}\right] \tag{8.10}$$

9 PSD matrix

9.1 Decomposition

1. The matrix is PSD with rank $r \iff$ there exists a matrix $\boldsymbol{B}$ of rank r such that $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}^\top$
2. The matrix is PD $\iff$ there exists an invertible matrix $\boldsymbol{B}$ such that $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}^\top$
3. Given $\boldsymbol{A}$ is an $n \times n$ PSD matrix, there exists an $n \times r$ matrix $\boldsymbol{B}$ of rank r such that $\boldsymbol{B}^\top\boldsymbol{A}\boldsymbol{B} = \boldsymbol{I}$.

9.2 Sylvester's characterization

$$\boldsymbol{A} \succeq 0 \iff \text{All } 2^n - 1 \text{ principal minors are nonnegative.} \tag{9.1}$$

$$\boldsymbol{A} \succ 0 \iff \text{All } n \text{ leading principal minors are positive.} \tag{9.2}$$

9.3 Eqution with zeros

Assume $\boldsymbol{A}$ is PSD, then $\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X} = \boldsymbol{0} \Rightarrow \boldsymbol{A}\boldsymbol{X} = 0$

9.4 Rank of product

Assume $\boldsymbol{A}$ is positive definite, then $\operatorname{rank}(\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^\top) = \operatorname{rank}(\boldsymbol{B})$

9.5 Outer product

If $\boldsymbol{X} \in n \times r$, where $n \leq r$ and $\operatorname{rank}(\boldsymbol{X}) = n$, then $\boldsymbol{X}\boldsymbol{X}^\top$ is positive definite.

9.6 Small pertubations

If $\boldsymbol{A}$ is positive definite, and $\boldsymbol{B}$ is symmetric, then $\boldsymbol{A} - t\boldsymbol{B}$ is positive definite for sufficiently small t.

9.7 Hadamard inequality

If $\boldsymbol{A}$ is a positive definite or semi-definite matrix, then

$$\det(\boldsymbol{A}) \leq \prod_i A_{ii} \tag{9.3}$$

9.8 Loewner order

Fact. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be hermitian positive definite. Then

$$\boldsymbol{A} \succeq \boldsymbol{B} \iff \boldsymbol{I} \succeq \boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2} \tag{9.4}$$

9.9 Inverse of PSD

Fact. Suppose that $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{A} - \boldsymbol{B}$ are all positive definite, then $\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1}$ is also positive definite.

10 Symmetric and skew-symmetric matrix

10.1 Properties of symmetric matrix

1. Every real symmetric matrix can be orthogonally diagonalizable. [1]
2. The rank of a symmetric matrix $\boldsymbol{A}$ is equal to the number of non-zero eigenvalues of $\boldsymbol{A}$.
3. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are $n \times n$ real symmetric matrices that commute, then they can be simultaneously diagonalized by an orthogonal matrix.

10.2 Youla decomposition

Given $\boldsymbol{B} \in \mathbb{R}^{M\times K}$ and $D \in \mathbb{R}^{K\times K}$, consider a rank-$K$ skew-symmetric matrix $\boldsymbol{B}^\top(\boldsymbol{D} - \boldsymbol{D}^\top)\boldsymbol{B}^\top$. Then, we can write

$$\begin{aligned}
\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{D}^\top)\boldsymbol{B} &= \sum_{j=1}^{K/2} i\sigma_j(\boldsymbol{a}_j + i\boldsymbol{b}_j)(\boldsymbol{a}_j + i\boldsymbol{b}_j)^H - i\sigma_j(\boldsymbol{a} - J - i\boldsymbol{b}_j)(\boldsymbol{a}_j - i\boldsymbol{b}_j)^H && (10.1)\\
&= \sum_{j=1}^{K/2} 2\sigma_j(\boldsymbol{a}_j\boldsymbol{b}_j^\top - \boldsymbol{b}_j\boldsymbol{a}_j^\top) && (10.2)\\
&= \sum_{j=1}^{K/2} \begin{bmatrix} \boldsymbol{a}_j - \boldsymbol{b}_j & \boldsymbol{a}_j + \boldsymbol{b}_j \end{bmatrix} \begin{bmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_j^\top - \boldsymbol{b}_j^\top \\ \boldsymbol{a}_j^\top + \boldsymbol{b}_j^\top \end{bmatrix} && (10.3)
\end{aligned}$$

[1] Think of it this way: every symmetric matrix can be triangulated and normality is preserved under a similar transform. When is the triangular matrix normal? Of course, it is the diagonal matrix.

Note that $\boldsymbol{a}_1 \pm \boldsymbol{b}_1, \ldots \boldsymbol{a}_{K/2} \pm \boldsymbol{b}_{K/2}$ are real-valued orthonormal vectors. The pair $\{(\sigma_j, \boldsymbol{a}_j - \boldsymbol{b}_j, \boldsymbol{a}_j + \boldsymbol{b}_j)\}_{j=1}^{K/2}$ is often called the Youla decomposition of $\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{D}^\top)\boldsymbol{B}^\top$.

11 Some techniques

11.1 Binary analysis

어떤 matrix의 operator norm을 분석하기 위해 matrix를 binary matrix로 decomposition하는 것은 유용할 수 있다.

Example

Let $\boldsymbol{v}$ be the unit-normed vector that realizes the operator norm of $\boldsymbol{D}^{-1}\boldsymbol{A}$. We define the sequence of binary matrices $\boldsymbol{B}^0, \boldsymbol{B}^1, \boldsymbol{B}^2$ as follows:

$$B_{i,j}^t := 1_{\left\{2^{-t-1}\sqrt{\alpha/n} < [\boldsymbol{D}^{-1}\boldsymbol{A}]_{i,j} \le 2^{-t}\sqrt{\alpha/n}\right\}} \quad \text{for every integers } t \ge 0, \tag{11.1}$$

where $\sqrt{\alpha/n}$ is the upper bound for entries of $\boldsymbol{D}^{-1}\boldsymbol{A}$. Then we have the following inequalities for the entries and the l_2-norm:

$$[\boldsymbol{D}^{-1}\boldsymbol{A}]_{i,j} \le \sqrt{\alpha/n} \sum_{t=0}^{\infty} 2^{-t} \cdot [\boldsymbol{B}^t]_{i,j} \tag{11.2}$$

$$\left\|\boldsymbol{D}^{-1}\boldsymbol{A} \cdot \boldsymbol{v}\right\|_2 \le \sqrt{\alpha/n} \sum_{t=0}^{\infty} 2^{-t} \cdot \left\|\boldsymbol{B}^t \boldsymbol{v}\right\|_2 \tag{11.3}$$

만약 $\boldsymbol{B}^t$ matrix의 row, column들의 non-zero elements를 estimate 하면 $\|\boldsymbol{B}^t\boldsymbol{v}\|_2^2$도 estimate 할 수 있고 $\boldsymbol{D}^{-1}\boldsymbol{A}$의 operator norm의 bound도 estimate 할 수 있다.