Adaptive Media Processing

2. Mathematical review

Keisuke Kameyama

Department of Computer Science

Keisuke.Kameyama@cs.tsukuba.ac.jp

Topics and schedule

Weeks 1-2

Introduction and reviews on math used in this course.

Weeks 3-7

Theories and techniques for adaptation, recognition and retrieval

- –Basic Pattern Recognition and the Bayes Rule
- –Linear Discrimination and Adaptive Filters
- –Neural Networks and Support Vector Machines
- –Clustering
- –Nearest Neighbor and Subspace Methods

Weeks 8-10

Applications

- –Content-Based Image Retrieval (CBIR)
- –Biometric Authentication
- –Classification of general object images

Math

- Linear maps and approximate solutions of matrix-vector equations
- Projection for dimensionality reduction
- Gradient-based optimization methods
- Lagrange theorems for optimization
- Statistics for pattern recognition
- Fourier transform and signal space
- Convolution and linear filters
- Autocorrelation and other signal features

Linear space (Vector space) (線形空間)

A nonempty set $\boldsymbol{V}$ in which two operations **addition** and **scalar multiplication** satisfying the following eight conditions are defined for its elements and a set of scalars, is called a **Linear Space** or **Vector Space**

Addition:
Operation $\boldsymbol{x}+\boldsymbol{y} \in \boldsymbol{V}$ defined for any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{V}$, satisfies the following four conditions.

1. For any $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \boldsymbol{V}$, $\quad (\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}=\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})$.
2. For any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{V}$, $\quad \boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$.
3. The **zero element** $\boldsymbol{o} \in \boldsymbol{V}$ exists such that for any $\boldsymbol{x} \in \boldsymbol{V}$, $\quad \boldsymbol{x}+\boldsymbol{o}=\boldsymbol{x}$.
4. For any $\boldsymbol{x} \in \boldsymbol{V}$, there exists a $\boldsymbol{x}' \in \boldsymbol{V}$ which is the unique **negative** of $\boldsymbol{x}$ that satisfies $\boldsymbol{x}+\boldsymbol{x}'=\boldsymbol{o}$.

Linear space (Vector space) (線形空間)

Scalar multiplication:
Operation $\lambda \boldsymbol{x} \in \boldsymbol{V}$ defined for any $\boldsymbol{x} \in \boldsymbol{V}$ and scalar λ satisfies the following four conditions.

1. For any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{V}$, and scalar λ,
   $\lambda(\boldsymbol{x} + \boldsymbol{y}) = \lambda \boldsymbol{x} + \lambda \boldsymbol{y}$.
2. For any $\boldsymbol{x} \in \boldsymbol{V}$, and scalars λ and μ,
   $(\lambda + \mu)\boldsymbol{x} = \lambda \boldsymbol{x} + \mu \boldsymbol{x}$.
3. For any $\boldsymbol{x} \in \boldsymbol{V}$, and scalars λ and μ,
   $(\lambda\mu)\boldsymbol{x} = \lambda(\mu \boldsymbol{x})$.
4. For any $\boldsymbol{x} \in \boldsymbol{V}$, scalar 1 is the **unit element** satisfying
   $\boldsymbol{x} = 1\boldsymbol{x}$.

Euclidean space (ユークリッド空間)

Normed vector space:

Normed vector space adds a property called **norm** to vectors of linear space $\boldsymbol{V}$, introducing the notion of **length**.

The norm of a vector is a real-valued function $\boldsymbol{x} \in \boldsymbol{V} \to ||\boldsymbol{x}|| \in \boldsymbol{R}$. Norm $||\boldsymbol{x}||$ satisfies the following :

1. $||\boldsymbol{x}|| \geq 0$. $||\boldsymbol{x}|| = 0$ iff $\boldsymbol{x} = \boldsymbol{o}$.
2. $||\boldsymbol{x} + \boldsymbol{y}|| \leq ||\boldsymbol{x}|| + ||\boldsymbol{y}||$ (triangular inequality)
3. $||\alpha \boldsymbol{x}|| = |\alpha| \cdot ||\boldsymbol{x}||$

Euclidean space $\boldsymbol{E}^n$

The n-dimensional Euclidean space $\boldsymbol{E}^n$ introduces the norm

$$||\boldsymbol{x}|| = ||[x_1, x_2, \ldots, x_n]|| = (\textstyle\sum_{i=1}^{n} x_i^2)^{1/2}$$

to the n-dimensional numerical space $\boldsymbol{R}^n$.

Subspace (部分空間)

When a subset $\boldsymbol{S} \neq \phi$ satisfies the following two conditions, $\boldsymbol{S}$ is a **subspace** of $\boldsymbol{V}$.

1. $\boldsymbol{x} + \boldsymbol{y} \in \boldsymbol{S}$ for any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}$
2. $\lambda \boldsymbol{x} \in \boldsymbol{S}$ for any $\lambda \in \boldsymbol{R}$ and $\boldsymbol{x} \in \boldsymbol{S}$

- **Subspace** of $\boldsymbol{V}$ is a subset of $\boldsymbol{V}$ which is a linear space itself.
- Space $\boldsymbol{V}$ and subspace $\boldsymbol{S}$ share the same zero element (origin).

Bases (基底)

Basis

When $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \in \boldsymbol{V}$ meet the following conditions, the set $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ is the **basis** of $\boldsymbol{V}$.

1. $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ are linearly independent.
2. Any $\boldsymbol{x} \in \boldsymbol{V}$ can be expressed as a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$.

Coordinate

If $\boldsymbol{x} \in \boldsymbol{V}$ is expressed as a linear sum of basis $\{\boldsymbol{a}_i\}_{i=1}^n$ as

$$\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n,$$

$[x_1, x_2, \ldots, x_n]$ is the **coordinate** of $\boldsymbol{x}$ with respect to basis $\{\boldsymbol{a}_i\}_{i=1}^n$.

Bases span space

To span
When any $\boldsymbol{x} \in \boldsymbol{V}$ can be expressed as a linear sum of a certain set of vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n \in \boldsymbol{V}$ as,

$$\boldsymbol{x} = a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2 + \ldots + a_n\boldsymbol{u}_n,$$

by appropriately choosing $a_1, \ldots, a_n$, vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ **span** the linear space $\boldsymbol{V}$.

- Basis $\{\boldsymbol{u}_i\}_{i=1}^n$ of a linear space $\boldsymbol{V}$ always spans $\boldsymbol{V}$.
- Notation "$\boldsymbol{V} < \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n >$" means that vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ span $\boldsymbol{V}$.

Maps (写像)

Map

For two sets $\boldsymbol{S}$ and $\boldsymbol{T}$, if for any $\boldsymbol{x} \in \boldsymbol{S}$, there exists a corresponding $\boldsymbol{y} \in \boldsymbol{T}$, this relation is called a **map from $\boldsymbol{S}$ to $\boldsymbol{T}$**. By signifying this relation (**map**) by f, this correspondence is expressed as,

$$f : \boldsymbol{S} \to \boldsymbol{T}.$$

Here, $\boldsymbol{y} = f(\boldsymbol{x})$ is the **image** of $\boldsymbol{x}$ under f.

$\Rightarrow$ injection (one-to-one), surjection (onto), bijection, inverse map, composition

Linear Map (線形写像)

Linear map

When a map f between two linear spaces $\boldsymbol{V}$ and $\boldsymbol{W}$ meets the following two conditions, map f is called a **linear map**.

1. For $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{V}$, $f(\boldsymbol{x}+\boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$.
2. For $\boldsymbol{x} \in \boldsymbol{V}$ and scalarλ, $f(\lambda \boldsymbol{x}) = \lambda f(\boldsymbol{x})$.

- If f and g are linear maps, $f(\boldsymbol{x})+g(\boldsymbol{x})$, $\lambda f(\boldsymbol{x})$, and $g(f(\boldsymbol{x}))$ are all linear maps.

Linear map – matrix association

Matrix associated with a linear map

Under a linear map $f : \boldsymbol{R}^n \to \boldsymbol{R}^m$, when the image of any $\boldsymbol{x} \in \boldsymbol{R}^n$, namely $\boldsymbol{y} = f(\boldsymbol{x}) \in \boldsymbol{R}^m$ can be expressed as

$$\boldsymbol{y} = f(\boldsymbol{x}) = A\boldsymbol{x} \quad (A \in \boldsymbol{M}_{mn}(\boldsymbol{R})),$$

then, matrix A is **associated with** the linear map f.

Isomorphism of the set of linear maps and matrices

There is an isomorphism (one-to-one and onto mapping) between the set of linear maps $f : \boldsymbol{R}^n \to \boldsymbol{R}^m$, and the set of matrices $\boldsymbol{M}_{mn}(\boldsymbol{R})$.

Inverse Matrix (逆行列)

Inverse matrix and regularity

For a square matrix $A \in \boldsymbol{M}_{nn}(\boldsymbol{R})$, if $X \in \boldsymbol{M}_{nn}(\boldsymbol{R})$ satisfying both $AX = E$ and $XA = E$ exists, matrix X is the **inverse matix** of A, and is written as $X = A^{-1}$. When A^{-1} exists, matrix A is **regular** or **nonsingular**.

Properties of inverse matrices

(1) $(A^{-1})^{-1} = A$
(2) $(AB)^{-1} = B^{-1}A^{-1}$
(3) $(A^T)^{-1} = (A^{-1})^T$
(4) $(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}$

Inverse map and matrix association

When matrix A is associated with linear map f, A^{-1} is associated with the inverse map f^{-1}.

Linear equations (連立1次方程式)

Linear equations in matrix-vector style

A m simultaneous linear equation with n unknowns can be expessed as

$$A\boldsymbol{x} = \boldsymbol{b}, \quad (A \in \boldsymbol{M}_{mn}(\boldsymbol{R}), \boldsymbol{x} \in \boldsymbol{R}^n, \boldsymbol{b} \in \boldsymbol{R}^m)$$

Properties

1. $rank\ A = m = n \Leftrightarrow$ Unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$
2. $rank\ A = rank\ [A|\boldsymbol{b}] \Leftrightarrow$ Solution exists
3. $rank\ A < n \Leftrightarrow$ Nonunique solution.
   $n - rankA = dim$ (solution space)

Linear Maps and Inverse problems

Input ***f*** → Known system *k*(***f***) → Observation ***g*** = *k*(***x***)

Discretized linear system

$$\boxed{\boldsymbol{g} = K\boldsymbol{f} \qquad \boldsymbol{g} \in \boldsymbol{R}^M, \boldsymbol{f} \in \boldsymbol{R}^N, K \in M_{MN}(\boldsymbol{R})}$$

Solve for ***f***

Invertible, nonsingular K**, bijective** $k(f)$

$$\boldsymbol{f} = K^{-1}\boldsymbol{g}$$

Noninvertible, singular K

$$\hat{\boldsymbol{f}} = \underset{\boldsymbol{f}}{argmin}\, ||K\boldsymbol{f} - \boldsymbol{g}||^2$$

Pseudoinverse Matrix

Moore-Penrose' Generalized Inverse matrix (Pseudoinverse matrix)

A generalization of inverse matrices to rectangular and singular matrices.

For matrix $A \in \boldsymbol{M}mn(\boldsymbol{R})$, if $X \in \boldsymbol{M}nm(\boldsymbol{R})$ satisfying

1. $AXA = A$
2. $XAX = X$
3. $(AX)^T = AX$
4. $(XA)^T = XA$

exists, matrix X is the **Moore-Penrose' generalized inverse** or **Pseudoinverse** matrix of A, and is written as $A^\dagger$.

Properties of pseudoinverse matrices

(1) $O^\dagger = O^T$ (O:zero matrix)

(2) $(A^\dagger)^\dagger = A$

(3) $(A^T)^\dagger = (A^\dagger)^T$

(4) $(\alpha A)^\dagger = \alpha^{-1} A^\dagger$ $(\alpha \neq 0)$

(5) $A^\dagger = (A^T A)^\dagger A^T = A^T (A A^T)^\dagger$

(6) $A^\dagger = A^{-1}$ (A:nonsingular)

Columns

(7) $A^\dagger = (A^T A)^{-1} A^T$ (~~Rows~~ of A are independent)

(8) $A^\dagger = A^T (A A^T)^{-1}$ (~~Columns~~ of A are independent)

(9) $rank(A^\dagger) = rank(A) = rank(A^T)$

Rows

Calculation of pseudoinverse (1)

Scalar (pseudo)inverse

$$\alpha^\dagger = \begin{cases} \alpha^{-1} & if(\alpha \neq 0) \\ 0 & if(\alpha = 0) \end{cases}$$

Vector pseudoinverse

$$\boldsymbol{a}^\dagger = \begin{cases} \boldsymbol{a}^T/(\boldsymbol{a}^T\boldsymbol{a}) & if(\boldsymbol{a} \neq \boldsymbol{o}) \\ \boldsymbol{o}^T & if(\boldsymbol{a} = \boldsymbol{o}) \end{cases}$$

Greville's theorem

Put for A with k columns as $A = A_k$, and $A_k = [A_{k-1}|\boldsymbol{a}_k]$ dividing to the rightmost column $\boldsymbol{a}_k$ and the rest A_{k-1}.

Then,

$$A_k^\dagger = \left[\begin{array}{c} A_{k-1}^\dagger(I - \boldsymbol{a}_k\boldsymbol{p}_k^T) \\ \hdashline \boldsymbol{p}_k^T \end{array} \right]$$

where

$$\boldsymbol{p}_k = \begin{cases} \dfrac{(I - A_{k-1}A_{k-1}^\dagger)\boldsymbol{a}_k}{||(I - A_{k-1}A_{k-1}^\dagger)\boldsymbol{a}_k||^2} & (if\ numerator = 0) \\ \dfrac{(A_{k-1}^\dagger)^T A_{k-1}^\dagger \boldsymbol{a}_k}{1 + ||A_{k-1}^\dagger \boldsymbol{a}_k||^2} & (otherwise) \end{cases}$$

Calculation of pseudoinverse (2)

Calculation using SVD $A = USV^T$

(1) The Pseudoinverse of matrix S

$$S^\dagger = \left[\begin{array}{ccc|c} 1/\sigma_1 & & & \\ & \ddots & & 0 \\ & & 1/\sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$$

$\rightarrow$ Check

$SS^\dagger S = S$, $S^\dagger S S^\dagger = S^\dagger$, $SS^\dagger = (SS^\dagger)^T$ and $S^\dagger S = (S^\dagger S)^T$

(2) Pseudoinverse of a general matrix $A = USV^T$ is,

$$A^\dagger = VS^\dagger U^T.$$

$\rightarrow$ Check

$AA^\dagger A = A$, $A^\dagger A A^\dagger = A^\dagger$, $AA^\dagger = (AA^\dagger)^T$ and $A^\dagger A = (A^\dagger A)^T$

Applications of pseudoinverses

(1) Overdetermined equations

When a system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ is **overdetermined**, namely $\boldsymbol{x} \in \boldsymbol{R}^n,\ A \in \boldsymbol{M}_{mn}(\boldsymbol{R}), \boldsymbol{b} \in \boldsymbol{R}^m,\ m > n$, no strict solution exists, generally. In such situations, a minimum- error approximated solution can be useful.

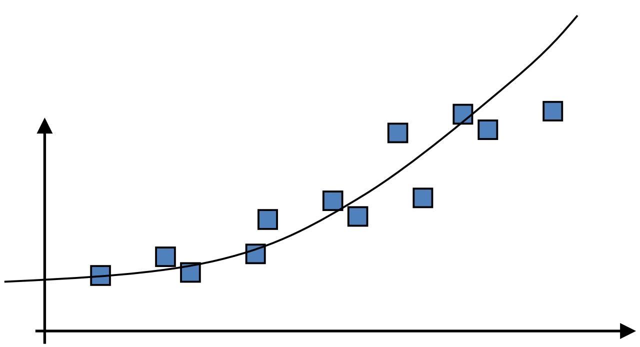
Approximation error

$$E \equiv \frac{1}{2}||A\boldsymbol{x} - \boldsymbol{b}||_2^2 = \frac{1}{2}(A\boldsymbol{x} - \boldsymbol{b})^T(A\boldsymbol{x} - \boldsymbol{b})$$

$$\frac{\partial E}{\partial \boldsymbol{x}} = A^T A\boldsymbol{x} - A^T\boldsymbol{b} = \boldsymbol{o}$$

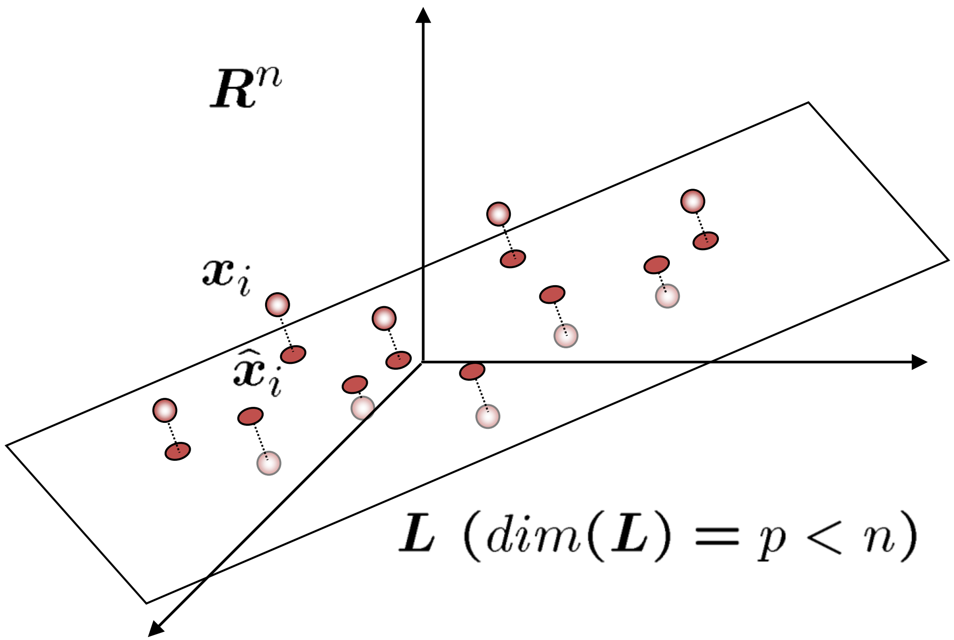
$$\hat{\boldsymbol{x}} = (A^T A)^{-1}A^T\boldsymbol{b} = A^\dagger\boldsymbol{b}$$

(2) Least squared error regression

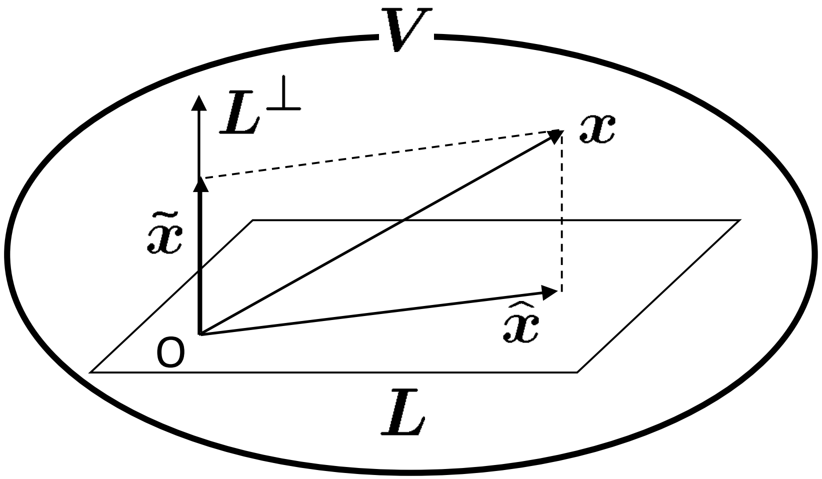


Approximation by projection (PCA)

- Reduction of space dimension for compression, noise reduction, etc.
- Select space $\boldsymbol{L}$ according to the data distribution
- Map each data to $\boldsymbol{L}$ by **orthogonal projection**



Projection



$$\boldsymbol{x} = \boldsymbol{x}' + \boldsymbol{x}'', \ \boldsymbol{x}' \in \boldsymbol{L}, \ \boldsymbol{x}'' \in \boldsymbol{L}^{\perp}$$

$\boldsymbol{x}'$: Approximation of $\boldsymbol{x}$ on $\boldsymbol{L}$

$\boldsymbol{x}''$: Residue

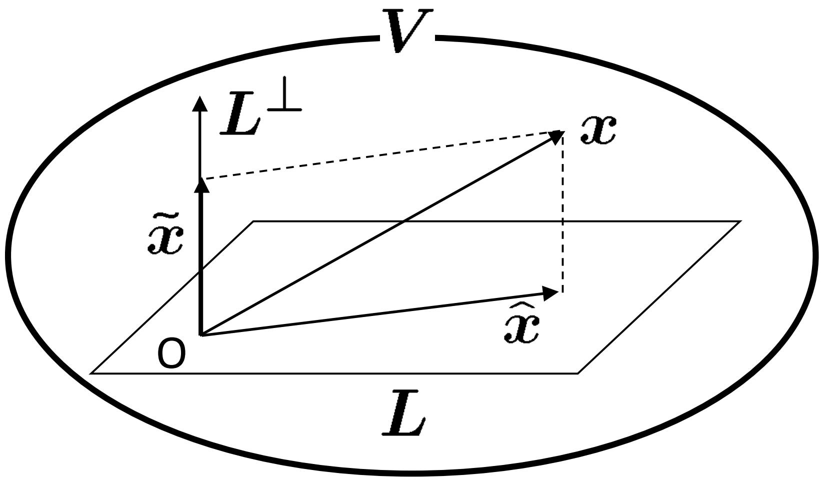
Projection theorem :

Minimize residue ⇔ Orthogonal projection

$$\boldsymbol{x} = \widehat{\boldsymbol{x}} + \tilde{\boldsymbol{x}}, \ \widehat{\boldsymbol{x}} \in \boldsymbol{L}, \ \tilde{\boldsymbol{x}} \in \boldsymbol{L}^{\perp}$$

$$(\widehat{\boldsymbol{x}}, \tilde{\boldsymbol{x}}) = 0$$

Orthogonal projection



When an orthogonal normal basis $\{\boldsymbol{u}_i\}_{i=1}^n$ of $\boldsymbol{L}$ is known,

$$\widehat{\boldsymbol{x}} = \sum_{i=1}^{n} (\boldsymbol{x}, \boldsymbol{u}_i)\boldsymbol{u}_i = \sum_{i=1}^{n} (\boldsymbol{x}^T \boldsymbol{u}_i)\boldsymbol{u}_i = (\sum_{i=1}^{n} \boldsymbol{u}_i \boldsymbol{u}_i^T)\boldsymbol{x}$$

Projection to a space spanned by a set of vectors

- Subspace $\boldsymbol{L} < \boldsymbol{b}_1, \ldots, \boldsymbol{b}_k > \subset \boldsymbol{V}$.

  Assume matrix $B = [\boldsymbol{b}_1 \ \ldots \ \boldsymbol{b}_k] \in M_{nk}(\boldsymbol{R}) \ (k < n)$.
- Let $\boldsymbol{x} \in \boldsymbol{V}$, $\boldsymbol{x} = \hat{\boldsymbol{x}} + \tilde{\boldsymbol{x}}$, $\hat{\boldsymbol{x}} \in \boldsymbol{L}$, $\tilde{\boldsymbol{x}} \in \boldsymbol{L}^{\perp}$. Then,

  $$\tilde{\boldsymbol{x}}^T B = \tilde{\boldsymbol{x}}^T [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k] = 0 \quad \cdots \quad (1)$$
- Penrose solution of Eq. (1) is,

  $$\tilde{\boldsymbol{x}}^T = \boldsymbol{y}^T - \boldsymbol{y}^T B B^{\dagger} = \boldsymbol{y}^T (I_n - B B^{\dagger}) \quad (\boldsymbol{y} \in \boldsymbol{R}^n)$$
- Transpose and have

  $$\tilde{\boldsymbol{x}} = (I_n - B B^{\dagger})^T \boldsymbol{y} = (I_n - B B^{\dagger}) \boldsymbol{y} \quad \text{(by symmetry)}$$
- By choosing $\boldsymbol{y} = \boldsymbol{x}$,

  $$\tilde{\boldsymbol{x}} = (I_n - B B^{\dagger})^T \boldsymbol{x} = \boldsymbol{x} - B B^{\dagger} \boldsymbol{x} \quad \text{gives} \quad \hat{\boldsymbol{x}} = B B^{\dagger} \boldsymbol{x}$$
- Therefore, the projection operator to $\boldsymbol{L}$ is

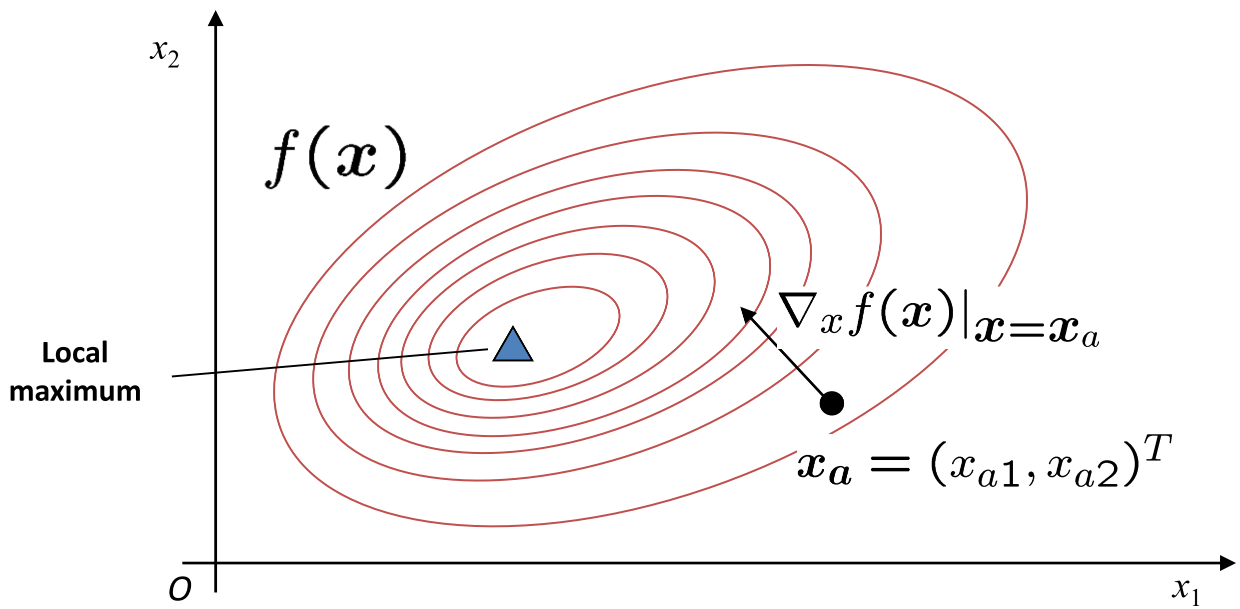
  $$P = B B^{\dagger} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k][\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k]^{\dagger}$$

Gradient of a scalar-function

gradient

$$\boldsymbol{x} = [x_1 \ x_2 \ \cdots \ x_n]^T \in \boldsymbol{R}^n, \ f(\boldsymbol{x}) \in \boldsymbol{R},$$

$$\nabla_x f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left[\frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \ \cdots \frac{\partial f}{\partial x_n}\right]^T \in \boldsymbol{R}^n$$



Gradients of scalar-valued functions

1. $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T \boldsymbol{y}) = \boldsymbol{y}$

2. $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{y}^T \boldsymbol{x}) = \boldsymbol{y}$

3. $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T \boldsymbol{x}) = 2\boldsymbol{x}$

4. $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T P \boldsymbol{y}) = P\boldsymbol{y}$

5. $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{y}^T P \boldsymbol{x}) = (\boldsymbol{y}^T P)^T = P^T \boldsymbol{y}$

6. $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T P \boldsymbol{x}) = P\boldsymbol{x} + P^T \boldsymbol{x}$

7. $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T P \boldsymbol{x}) = 2P\boldsymbol{x}$ for $P^T = P$

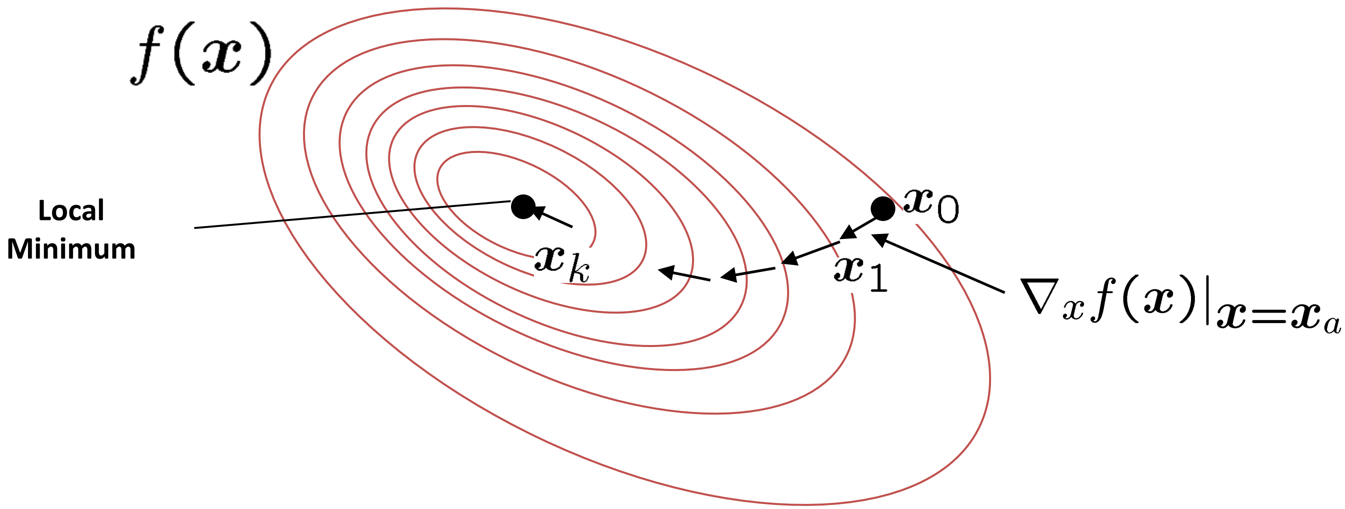
Steepest Descent Method of Optimization

Target function $f(\boldsymbol{x})$

$$\boldsymbol{x} = [x_1 \ x_2 \ \cdots \ x_n]^T \in \boldsymbol{R}^n, \ \ f(\boldsymbol{x}) \in \boldsymbol{R},$$

Search for (local) minimum of $f(\mathbf{x})$

$$\boldsymbol{x}(t+1) = \boldsymbol{x}(t) - \eta \nabla_x f(\boldsymbol{x}) \qquad (\eta > 0)$$



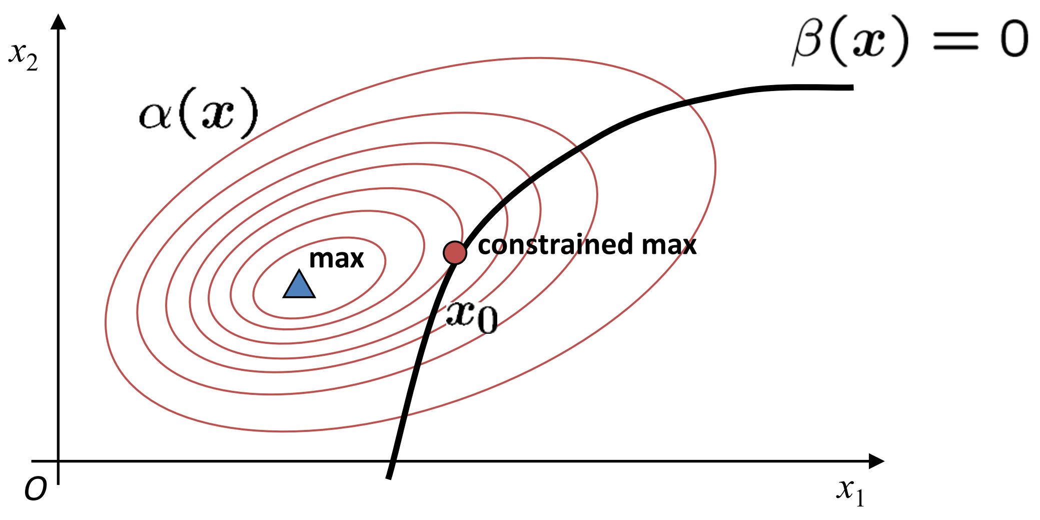
Optimization under equality constraints

Theorem (Lagrange)

When function $\alpha(\boldsymbol{x})$ $(\boldsymbol{x} \in \boldsymbol{R}^n, \alpha(\boldsymbol{x}) \in \boldsymbol{R})$ takes a local maximum (minimum) at $\boldsymbol{x} = \boldsymbol{x}_0$ under condition $\underline{\beta(\boldsymbol{x}) = 0 \ (\in \boldsymbol{R})}$, the following holds.

equality constraint

$$\left.\frac{\partial \alpha(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{x=x_0} - \lambda \left.\frac{\partial \beta(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{x=x_0} = \boldsymbol{o} \quad \text{where } \lambda \in \boldsymbol{R}.$$



Optimization under inequality constraints

Theorem (Kuhn-Tucker)

Let

$$\boldsymbol{x} \in \boldsymbol{R}^n, \quad \alpha(\boldsymbol{x}), \beta_i(\boldsymbol{x}) \in \boldsymbol{R} \quad (i = 1, \ldots, r)$$
$$\boldsymbol{z} = [\lambda_1 \ldots \lambda_r]^T \in \boldsymbol{R}^n, \text{ and } \boldsymbol{b} = [\beta_1(\boldsymbol{x}) \ldots \beta_r(\boldsymbol{x})]^T \in \boldsymbol{R}^r.$$

If $\alpha(\boldsymbol{x})$ is locally minimized under conditions $\beta_i(\boldsymbol{x}) \leq 0 \ (i = 1, \ldots, r)$ at $\boldsymbol{x} = \boldsymbol{x}_0$, then the **Lagrangian**

$$L(\boldsymbol{x}, \boldsymbol{z}) = \alpha(\boldsymbol{x}) + \boldsymbol{z}^T \boldsymbol{b}(\boldsymbol{x})$$

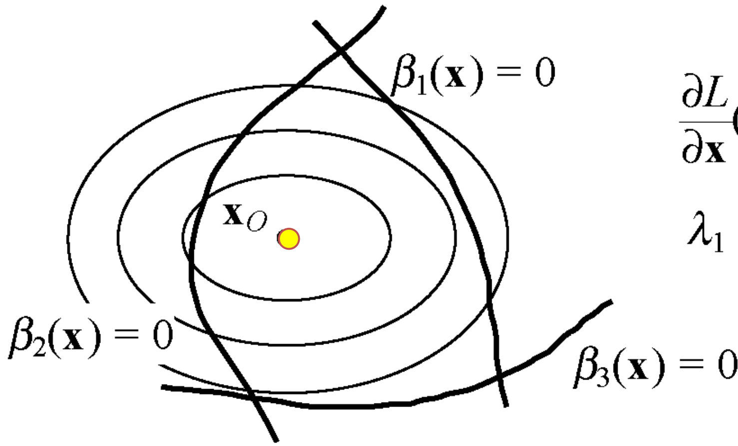
takes a **saddle point** at $\boldsymbol{x} = \boldsymbol{x}_0$, which is a local minimum for $\boldsymbol{x}$ and a local maximum for $\boldsymbol{z}$, respectively.

Saddle point conditions

$$\frac{\partial}{\partial \boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{z}) = \frac{\partial \alpha}{\partial \boldsymbol{x}}(\boldsymbol{x}_0) + \sum_{i=1}^{r} \lambda_i \frac{\partial \beta_i}{\partial \boldsymbol{x}}(\boldsymbol{x}_0) = \boldsymbol{o}$$
$$\frac{\partial}{\partial \boldsymbol{z}} L(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{o}$$

Relations of the optimum and inequality conditions.

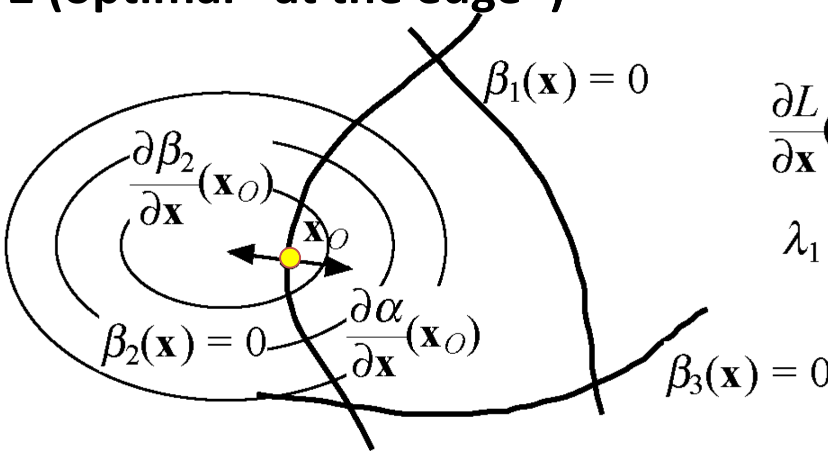
Case 1 (optimum "inside")



$$\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}_O,\ \mathbf{z}) = \frac{\partial \alpha}{\partial \mathbf{x}}(\mathbf{x}_O) = \mathbf{O}$$

$$\lambda_1 = \lambda_2 = \lambda_3 = 0$$

Case 2 (optimal "at the edge")

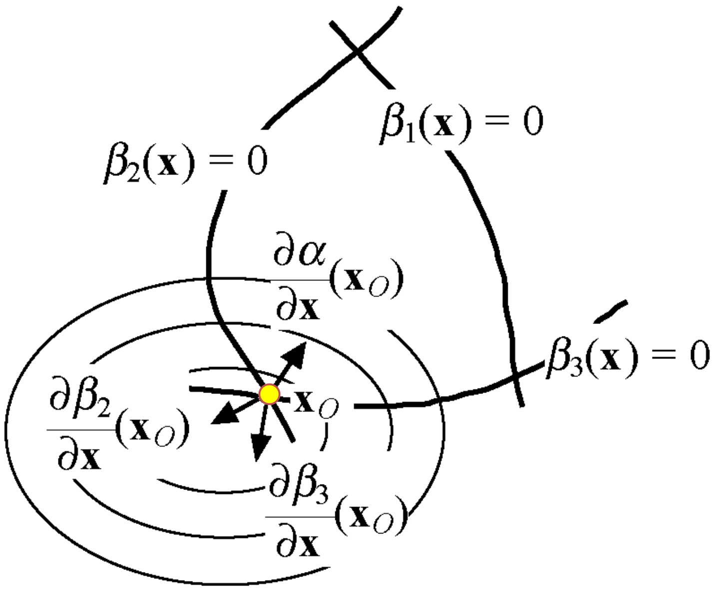


$$\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}_O,\ \mathbf{z}) = \frac{\partial \alpha}{\partial \mathbf{x}}(\mathbf{x}_O) + \boldsymbol{\lambda}_2 \frac{\partial \beta_2}{\partial \mathbf{x}}(\mathbf{x}_O) = \mathbf{O}$$

$$\lambda_1 = \lambda_3 = 0$$

Relations of the optimum and inequality conditions.

Case 3 (optimal "at corner")



$$\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}_O, \mathbf{z}) = \frac{\partial \alpha}{\partial \mathbf{x}}(\mathbf{x}_O) + \boldsymbol{\lambda}_2 \frac{\partial \beta_2}{\partial \mathbf{x}}(\mathbf{x}_O) + \boldsymbol{\lambda}_3 \frac{\partial \beta_3}{\partial \mathbf{x}}(\mathbf{x}_O) = \mathbf{O}$$

$$\lambda_1 = 0$$

Property

Lemma (Kuhn-Tucker)

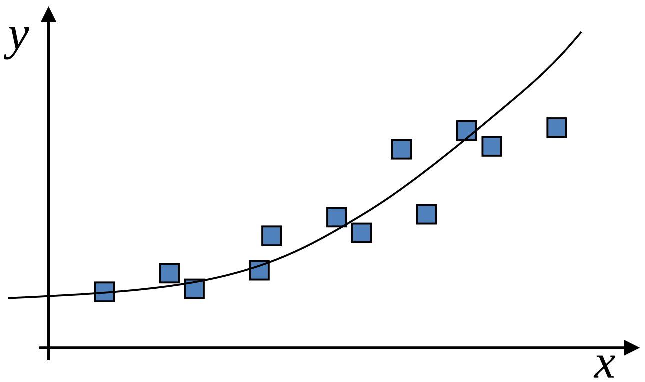
$$\lambda_i \geq 0 \quad \text{and} \quad \lambda_i \beta_i(\boldsymbol{x}_0) = 0 \quad (i = 1, \ldots, r)$$

Proof

1. Vector $\frac{\partial \alpha}{\partial \boldsymbol{x}}(\boldsymbol{x}_0)$ points inward (minimum of $\alpha(\boldsymbol{x})$ sought).
2. Vector $\frac{\partial \beta}{\partial \boldsymbol{x}}(\boldsymbol{x}_0)$ points outward ($\beta_i(\boldsymbol{x}) \leq 0$).
3. 1 & 2 $\Rightarrow \lambda_i \geq 0$
4. $\lambda_i \beta_i(\boldsymbol{x}_0) = 0$ holds as $\lambda_i > 0$ for all conditions effective as **equality** conditions ($\beta_i(\boldsymbol{x}_0) = 0$), and $\lambda_i = 0$ for conditions effective as **inequality** conditions ($\beta_i(\boldsymbol{x}) \leq 0$).

Homework

Least squared error regression



Given a set of samples $\{(x_i, y_i)\}_{i=1}^{N}$, derive a minimum squared error regression of the data using a n-th order polynomial

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

Hint : Find

$$\underset{a_0,\ldots,a_n}{argmin} \left\| \begin{bmatrix} y_1 - (a_0 + a_1 x_1 + a_2 {x_1}^2 + \cdots + a_n {x_1}^n) \\ \vdots \\ y_N - (a_0 + a_1 x_N + a_2 {x_N}^2 + \cdots + a_n {x_N}^n) \end{bmatrix} \right\|^2$$

Singular Value Decomposition : SVD (特異値分解)

Any $A \in \boldsymbol{M}_{mn}(\boldsymbol{R})$ can be decomposed as,

$$A = USV^T,$$

using orthogonal matrices U and V consisting of columns of orthogonal normal bases $\{\boldsymbol{v}_i\}_{i=1}^{n}$ and $\{\boldsymbol{u}_i\}_{i=1}^{m}$, so that matrix S will become

$$S = U^T AV = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]. \quad (\sigma_1 > \sigma_2 > \cdots > \sigma_r)$$

· Scalars σ_i $(i = 1, \ldots, r)$ are called **singular values** of A.

· Vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are called left and right **singular vectors**, respectively.

Calculation of SVD (1)

- Derive eigenvalues and eigenvectors of a symmetric matrix $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{A}^T$. Sort eigenvalues in descending order as,

$$\sigma_1^2 \geq \sigma_2^2 \geq, \ldots, \geq \sigma_r^2 \geq \sigma_{r+1}^2 = \ldots = \sigma_m^2 = 0.$$

- Obtain an orthonormal basis from the eigenvectors as,

$$\boldsymbol{U} = [\boldsymbol{u}_1 \ \boldsymbol{u}_2 \cdots \boldsymbol{u}_m].$$

- As square roots of the eigenvalues are the singular values, we have

$$\boldsymbol{S} = \left[\begin{array}{ccc|ccc} \sigma_1 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_r & & & \\ \hline & 0 & & & 0 & \end{array}\right].$$

Calculation of SVD (2)

- Derive eigenvalues and eigenvectors of another symmetric matrix $\boldsymbol{D} = \boldsymbol{A}^T \boldsymbol{A}$.
  (Eigenvalues are the same, but eigenvectors differ.)
- Obtain an orthonormal basis from the eigenvectors as,

$$\boldsymbol{V} = [\boldsymbol{v}_1 \ \boldsymbol{v}_2 \cdots \boldsymbol{v}_n].$$

- end

Properties and uses of SVD

(1) $A\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j, A^T \boldsymbol{u}_k = \sigma_k \boldsymbol{v}_k$
(2) $A^T A \boldsymbol{v}_j = \sigma_j^2 \boldsymbol{v}_j$ and $AA^T \boldsymbol{u}_k = \sigma_k^2 \boldsymbol{u}_k$
(3) Relations with eigenvalues of square matrices
(4) $A = USV^T$ as a matrix associated with a linear map
(5) Fixed rank approximation of a matrix
(6) Calculation of Pseudoinverse matrices
(7) Calculation of approximated inverse maps (regularization

Orthogonal matrix (直交行列)

Definition

Matrix $A \in \boldsymbol{M}_{nn}(\boldsymbol{R})$ is **orthogonal** if $A^T A = I$.

Properties of orthogonal matrices

1. $|A| = \pm 1$
2. $A^T = A^{-1}$
3. column vectors are orthogonal

Example : 2×2 **orthogonal matrices**

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{or} \quad B = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

General matrix equation *AXB*=*C* and its solution

Theorem

A solution exists for $AXB = C \iff AA^{\dagger}CB^{\dagger}B = C$ holds

Proof

($\Rightarrow$) X is one solution. Thus,

$$AXB = AA^{\dagger}AXBB^{\dagger}B = AA^{\dagger}CB^{\dagger}B = C$$

($\Leftarrow$) $AA^{\dagger}CB^{\dagger}B = A(A^{\dagger}CB^{\dagger})B = C$ holds. So, at least one solution $A^{\dagger}CB^{\dagger}$ does exist.

Penrose solution of general matrix equation *AXB*=*C*

(General solution of *AXB*=*C*) =
(Particular solution of *AXB*=*C*)
+ (Homogeneous sol. of *AXB*=*O*)

Particular solution of *AXB*=*C*

$$AXB = AA^{\dagger}AXBB^{\dagger}B = AA^{\dagger}CB^{\dagger}B$$

$X = A^{\dagger}CB^{\dagger}$ is found to be a particular solution.

Penrose solution of general matrix equation *AXB*=*C*

Homogeneous solution of *AXB*=*O*

- $X = Y - A^{\dagger}AYBB^{\dagger}$ is a sol. of $AXB = O$, for $MM^{\dagger}M = M$ holds for any M ,with Y having same dimensions as X.
- If X is a solution, any sol. of $AXB = O$ can be written as $Y - A^{\dagger}AYBB^{\dagger}$, by choosing a Y because $X = X - A^{\dagger}AXBB^{\dagger}$
- Therefore, any solution to $AXB = O$ is written as $Y - A^{\dagger}AYBB^{\dagger}$ by conveniently choosing Y.

General solution of *AXB*=*C*（Penrose sol.）

$$X = A^{\dagger}CB^{\dagger} + Y - A^{\dagger}AYBB^{\dagger}$$