Appendix A Mathematical background

Last update on 2024-02-26 18:28

Table of contents

Notations

Linear algebra

Calculus

Notations

- numbers

$$\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$$
$$\mathbb{R}_{++} = \{x \in \mathbb{R} \mid x > 0\}$$

- all vectors/matrices are delimited by square brackets, parentheses are used for construct column vectors from comma separated lists
- $\mathbf{1}$ in bold font denotes a vector all of whose components are 1
- matrices

$$\mathbb{S}^k = \{X \in \mathbb{R}^{k \times k} \mid X^T = X\}$$
$$\mathbb{S}^k_+ = \{X \in \mathbb{R}^{k \times k} \mid X^T = X \text{ and } X \succeq 0\} \quad \text{(positive semidefinite symmetric)}$$
$$\mathbb{S}^k_{++} = \{X \in \mathbb{R}^{k \times k} \mid X^T = X \text{ and } X \succ 0\} \quad \text{(positive definite symmetric)}$$

- $f : A \to B$ indicates $\mathbf{dom}\, f \subseteq A$

Symmetric eigenvalue decomposition

- Suppose $A \in \mathbb{S}^n$, then
$$A = Q\Lambda Q^T,$$
where Q is orthogonal and $\Lambda = \mathbf{diag}(\lambda_1, \dots, \lambda_n)$ with $\lambda_1 \geq \cdots \geq \lambda_n$.

- Suppose $A \in \mathbb{S}^n_+$, then $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. We define
$$A^{\frac{1}{2}} = Q\,\mathbf{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_n^{\frac{1}{2}})Q^T.$$

Singular value decomposition

Suppose $A \in \mathbb{R}^{m \times n}$ with $\mathbf{rank}\, A = r$, then

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

where

- $U \in \mathbb{R}^{m \times r}$ with $U^T U = I_r$;
- $\Sigma = \mathbf{diag}(\sigma_1, \ldots, \sigma_r)$ with $\sigma_1 \geq \cdots \geq \sigma_r > 0$;
- $V \in \mathbb{R}^{n \times r}$ with $V^T V = I_r$.

The matrix

$$A^\dagger = V \Sigma^{-1} U^T \in \mathbb{R}^{n \times m}$$

is called the **pseudo-inverse** of A.

Schur complement

Let

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \in \mathbb{S}^n$$

and assume $\det A \neq 0$. Then the matrix

$$S = C - B^T A^{-1} B$$

is called the **Schur complement** of A in X.

- $\det X = \det A \det S$
- $X \succ 0 \quad \Longleftrightarrow \quad A \succ 0$ and $S \succ 0$
- $X \succeq 0 \quad \Longleftrightarrow \quad A \succ 0$ and $S \succeq 0$

Notations

Linear algebra

Calculus

Standard inner products

- For any $x, y \in \mathbb{R}^n$
$$\langle x, y\rangle = x^T y, \qquad \|x\|_2 = \left(x^T x\right)^{\frac{1}{2}}.$$
- For any $X, Y \in \mathbb{R}^{m\times n}$
$$\langle X, Y\rangle = \mathbf{tr}(X^T Y), \qquad \|X\|_F = \left(\mathbf{tr}\left(X^T X\right)\right)^{\frac{1}{2}}.$$

Norms

- ℓ_p norm: $x \in \mathbb{R}^n$, $p \geq 1$ or $p = \infty$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, \qquad \|x\|_\infty = \mathbf{max}\left\{|x_1|, \ldots, |x_n|\right\}$$

- quadratic norm: $x \in \mathbb{R}^n$, $P \in \mathbb{S}^n_{++}$

$$\|x\|_P = \left(x^T P x\right)^{\frac{1}{2}} = \|P^{\frac{1}{2}} x\|_2$$

Facts about norms

- $C \subseteq \mathbb{R}^n$ symmetric, convex, closed, compact with non-empty interior
$\iff C$ is the unit ball of the norm

$$\|x\| = (\mathbf{sup}\{t \geq 0 \mid tx \in C\})^{-1} \qquad \text{for any} \qquad x \in \mathbb{R}^n$$

- let $\|\cdot\|_a$ and $\|\cdot\|_b$ be any norms on $\mathbb{R}^n$, then there exists $\alpha, \beta > 0$ such that

$$\alpha\|x\|_a \leq \|x\|_b \leq \beta\|x\|_a \qquad \text{for any} \qquad x \in \mathbb{R}^n$$

- let $\|\cdot\|$ be any norm on $\mathbb{R}^n$, then there exists a quadratic norm $\|\cdot\|_P$ such that

$$\|x\|_P \leq \|x\| \leq \sqrt{n}\|x\|_P \qquad \text{for any} \qquad x \in \mathbb{R}^n$$

Operator norms

Given $(\mathbb{R}^m, \|\cdot\|_a)$ and $(\mathbb{R}^n, \|\cdot\|_b)$, for any $X \in \mathbb{R}^{m\times n}$, we define

$$\|X\|_{a,b} = \sup_{u\in\mathbb{R}^n} \left\{\|Xu\|_a \mid \|u\|_b \leq 1\right\}.$$

- if $a = b = 2$, it is called ℓ_2-norm or spectral norm

$$\|X\|_2 = \sigma_{\mathbf{max}}(X) = \left(\lambda_{\mathbf{max}}\left(X^T X\right)\right)^{\frac{1}{2}}$$

- if $a = b = \infty$, it is called ℓ_∞-norm or max-row-sum

$$\|X\|_\infty = \max_{1\leq i\leq m}\left(\sum_{j=1}^{n}|x_{ij}|\right)$$

- if $a = b = 1$, it is called ℓ_1-norm or max-column-sum

$$\|X\|_\infty = \max_{1\leq j\leq n}\left(\sum_{i=1}^{m}|x_{ij}|\right)$$

Dual norm

Given $(\mathbb{R}^n, \|\cdot\|)$, the dual norm is given as

$$\|z\|_* = \sup_{x\in\mathbb{R}^n} \left\{ z^T x \mid \|x\| \leq 1 \right\}$$

with properties

$$z^T x \leq \|x\| \|z\|_* \qquad \text{and} \qquad \|\cdot\|_{**} = \|\cdot\|$$

- on $\mathbb{R}^n$, the dual of the p-norm is the q-norm, where $\frac{1}{p} + \frac{1}{q} = 1$
- on $\mathbb{R}^{m\times n}$, the dual of the spectral norm

$$\|X\|_2 = \sigma_{\mathbf{max}}(X)$$

is the nuclear norm

$$\|Z\|_{2*} = \sigma_1(Z) + \cdots + \sigma_r(Z) = \mathbf{tr}\left(\left(Z^T Z\right)^{\frac{1}{2}} \right)$$

Derivatives

- The derivative (Jacobian) of a function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ at x is

$$Df(x) \in \mathbb{R}^{m\times n} \quad \text{with} \quad Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \quad i = 1,\dots,m, \quad j = 1,\dots,n.$$

- The gradient of a function $f\colon \mathbb{R}^n \to \mathbb{R}$ at x is

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

- The Hessian matrix of a function $f\colon \mathbb{R}^n \to \mathbb{R}$ at x is

$$\nabla^2 f(x) = D\nabla f(x) \in \mathbb{S}^n.$$

First order approximation

General form

Suppose $f\colon \mathbb{R}^n \to \mathbb{R}$, and $x, y \in \mathbb{R}^n$, where y is close to x, then

$$\begin{aligned} f(y) &\approx f(x) + \nabla f(x)^T (y - x) \\ &= f(x) + \langle \nabla f(x), y - x \rangle \end{aligned}$$

where both $\nabla f(x)$ and $y - x$ are vectors in $\mathbb{R}^n$.

Special case

Suppose $f\colon \mathbb{R}^{m\times n} \to \mathbb{R}$, and $X, Z \in \mathbb{R}^{m\times n}$, where Z is close to X, then

$$f(Z) \approx f(X) + \mathbf{tr}\left(\nabla f(x)^T (Z - X)\right)$$

where both $\nabla f(x)$ and $Z - X$ are $m \times n$ matrices.

Second order approximation

Suppose $f\colon \mathbb{R}^n \to \mathbb{R}$, and $x, y \in \mathbb{R}^n$, where y is close to x, then

$$\begin{aligned} f(y) &\approx f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}(y-x)^T\nabla^2 f(x)(y-x) \\ &= f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}\langle \nabla^2 f(x)(y-x), y-x\rangle \end{aligned}$$

where both $\nabla^2 f(x)(y-x)$ and $y-x$ are vectors in $\mathbb{R}^n$.

Remark

The term $\nabla^2 f(x)(y-x)$ appears also in the first order approximation of $\nabla f(x)$:

$$\begin{aligned} \nabla f(y) &\approx \nabla f(x) + D\nabla f(x)(y-x) \\ &= \nabla f(x) + \nabla^2 f(x)(y-x), \end{aligned}$$

which is sometimes helpful for the computation.

Example

Consider the function $f\colon \mathbb{S}^n \to \mathbb{R}$ defined as

$$f(X) = \log\det X, \qquad \mathbf{dom}\, f = \mathbb{S}^n_{++}.$$

Compute the second order approximation of f.