

# Optimization Theory

## Lecture 02

Fudan University

luoluo@fudan.edu.cn

# Outline

- 1 Matrix Calculus
- 2 Topology
- 3 Convergence Rates

# Outline

1 Matrix Calculus

2 Topology

3 Convergence Rates

# Matrix Calculus

Given differentiable  $f : \mathbb{R}^{p \times q} \rightarrow \mathbb{R}$  and  $\mathbf{X} \in \mathbb{R}^{p \times q}$ , we define

$$\nabla f(\mathbf{X}) = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1q}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{p1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{pq}} \end{bmatrix} \in \mathbb{R}^{p \times q} \quad \text{and} \quad d\mathbf{X} = \begin{bmatrix} dx_{11} & \cdots & dx_{1q} \\ \vdots & \ddots & \vdots \\ dx_{p1} & \cdots & dx_{pq} \end{bmatrix} \in \mathbb{R}^{p \times q}.$$

We have

$$df(\mathbf{X}) = \sum_{i=1}^p \sum_{j=1}^q \frac{\partial f(\mathbf{X})}{\partial x_{ij}} \cdot dx_{ij} = \langle \nabla f(\mathbf{X}), d\mathbf{X} \rangle = \text{tr}(\nabla f(\mathbf{X})^\top d\mathbf{X})$$

and

$$d(\mathbf{XY}) = (d\mathbf{X})\mathbf{Y} + \mathbf{X}d\mathbf{Y}, \quad d(\mathbf{AXB}) = \mathbf{A} \cdot d\mathbf{X} \cdot \mathbf{B}$$

# The Hessian

Suppose that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a smooth function that takes as input a matrix  $\mathbf{x} \in \mathbb{R}^n$  and returns a real value. Then the Hessian with respect to  $\mathbf{x}$ , written as  $\nabla^2 f(\mathbf{x})$ , which is defined as

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

We write Taylor's expansion for multivariate function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  as

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a})^\top (\mathbf{x} - \mathbf{a}) + \frac{1}{2} (\mathbf{x} - \mathbf{a})^\top \nabla^2 f(\mathbf{a}) (\mathbf{x} - \mathbf{a}).$$

# Outline

1 Matrix Calculus

2 Topology

3 Convergence Rates

# Topology in Euclidean Space

Open set, closed set, bounded set and compact set:

- 1 A subset  $\mathcal{C}$  of  $\mathbb{R}^d$  is called open, if for every  $\mathbf{x} \in \mathcal{C}$  there exists  $\delta > 0$  such that the ball  $\mathcal{B}_\delta(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y} - \mathbf{x}\|_2 \leq \delta\}$  is included in  $\mathcal{C}$ .
- 2 A subset  $\mathcal{C}$  of  $\mathbb{R}^d$  is called closed, if its complement  $\mathcal{C}^c = \mathbb{R}^d \setminus \mathcal{C}$  is open.
- 3 A subset  $\mathcal{C}$  of  $\mathbb{R}^d$  is called bounded, if there exists  $r > 0$  such that  $\|\mathbf{x}\|_2 < r$  for all  $\mathbf{x} \in \mathcal{C}$ .
- 4 A subset  $\mathcal{C}$  of  $\mathbb{R}^d$  is called compact, if it is both bounded and closed.

Is there any subset of  $\mathbb{R}^d$  that is both open and closed?

# Topology in Euclidean Space

Interior, closure and boundary:

- ① The interior of  $\mathcal{C} \in \mathbb{R}^n$  is defined as

$$\mathcal{C}^\circ = \{\mathbf{y} \in \mathbb{R}^n : \text{there exist } \delta > 0 \text{ such that } \mathcal{B}_\delta(\mathbf{y}) \subseteq \mathcal{C}\}$$

- ② The closure of  $\mathcal{C} \in \mathbb{R}^n$  is defined as

$$\overline{\mathcal{C}} = \mathbb{R}^n \setminus (\mathbb{R}^n \setminus \mathcal{C})^\circ.$$

- ③ The boundary of  $\mathcal{C} \in \mathbb{R}^n$  is defined as  $\overline{\mathcal{C}} \setminus \mathcal{C}^\circ$ .



# Topology in General Case

In a metric space, an open set is a set that, along with every point  $\mathbf{x}$ , contains all points that are sufficiently near to  $\mathbf{x}$ .

The other concept also can be generalized in the similar way.

For example, the positive-definite matrices on  $\mathbb{R}^{d \times d}$  with distance under spectral norm is open.

# Outline

- 1 Matrix Calculus
- 2 Topology
- 3 Convergence Rates

# Convergence Rates

Assume the sequence  $\{\mathbf{x}_k\}$  converges to  $\mathbf{x}^*$ . We define the errors

$$z_k = \|\mathbf{x}_k - \mathbf{x}^*\|$$

and suppose

$$\lim_{k \rightarrow +\infty} \frac{z_{k+1}}{z_k^r} = C \quad \text{for some } C \in \mathbb{R}.$$

Q-convergence rates:

- ① linear:  $r = 1, 0 < C < 1$ ;
- ② sublinear:  $r = 1, C = 1$ ;
- ③ superlinear:  $r = 1, C = 0$ ;
- ④ quadratic:  $r = 2$ .

# Convergence Rates

Consider the example

$$x_k = 2^{-\lceil k/2 \rceil} + 1,$$

It should converge to  $x^* = 0$  linearly, however,

$$\lim_{k \rightarrow +\infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|}$$

does not exist.

# Convergence Rates

Suppose that  $\{\mathbf{x}_k\}$  converges to  $\mathbf{x}^*$ . The sequence is said to converge R-linearly to  $\mathbf{x}^*$  if there exists a sequence  $\{\epsilon_k\}$  such that

$$\|\mathbf{x}_k - \mathbf{x}^*\| \leq \epsilon_k$$

for all  $k$  and  $\{\epsilon_k\}$  converges Q-linearly to zero.