Optimization Theory

Lecture 02

Fudan University

luoluo@fudan.edu.cn

Outline

Outline

Matrix Calculus

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

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

We have

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

and

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

The Hessian

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ 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 $f : \mathbb{R}^n \to \mathbb{R}$ 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

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 $\mathcal{C}^{\mathrm{c}} = \mathbb{R}^d \backslash \mathcal{C}$ 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?

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

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

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

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

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

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

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,

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

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.