Optimization Theory

Lecture 05

Fudan University

luoluo@fudan.edu.cn

Outline

Outline

1. Second-Order Characterization

2. Examples and Applications

Second-Order Characterization

Theorem (Smoothness and Convexity)

Let $f(\cdot)$ be a twice differentiable function defined on $\mathbb{R}^d$

1. *It is L-smooth if and only if $-L\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq L\mathbf{I}$ for all $\mathbf{x} \in \mathbb{R}^d$.*
2. *It is convex if and only if $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}^d$.*
3. *It is μ-strongly-convex if and only if $\nabla^2 f(\mathbf{x}) \succeq \mu\mathbf{I}$ for all $\mathbf{x} \in \mathbb{R}^d$.*

Sometimes, we say $f(\cdot)$ is ℓ-weakly convex if the function

$$g(\mathbf{x}) = f(\mathbf{x}) + \frac{\ell}{2}\|\mathbf{x}\|_2^2$$

is convex for some $\ell > 0$.

Second-Order Characterization

Theorem

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable function. Suppose that $\nabla^2 f(\cdot)$ is continuous in an open neighborhood of $\mathbf{x}^* \in \mathbb{R}^d$.

1. *If $\mathbf{x}^*$ is a local minimizer of $f(\cdot)$, then it holds that*

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \qquad \text{and} \qquad \nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}.$$

2. *If it holds that*

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \qquad \text{and} \qquad \nabla^2 f(\mathbf{x}^*) \succ \mathbf{0},$$

   then the point $\mathbf{x}^*$ is a strict local minimizer of $f(\cdot)$.

Outline

Examples

1. For unconstrained quadratic problem

$$\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \triangleq \frac{1}{2}\mathbf{x}^\top\mathbf{A}\mathbf{x} - \mathbf{b}^\top\mathbf{x},$$

   where $\mathbf{A}\in\mathbb{R}^{d\times d}$. We have $\nabla^2 f(\mathbf{x}) = \mathbf{A}$.

2. For regularized generalized linear model

$$\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \triangleq \frac{1}{n}\sum_{i=1}^{n}\phi_i(\mathbf{a}^\top\mathbf{x}) + \frac{\lambda}{2}\|\mathbf{x}\|_2^2.$$

   where $\phi_i:\mathbb{R}^d\to\mathbb{R}$ is twice differentiable. We have

$$\nabla f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\phi_i'(\mathbf{a}_i^\top\mathbf{x})\mathbf{a} + \lambda\mathbf{x}$$

   and

$$\nabla^2 f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\phi_i''(\mathbf{a}_i^\top\mathbf{x})\mathbf{a}_i\mathbf{a}_i^\top + \lambda\mathbf{I}.$$

Applications in Matrix Approximation

Given a symmetric positive-definite matrix $\mathbf{K} \in \mathbb{R}^{d\times d}$ and we sample a subset of columns $\mathbf{C} \in \mathbb{R}^{d\times m}$, where $m < d$.

We want to establish the estimator of $\mathbf{K}$ by the formulation

$$\min_{\mathbf{U}\in\mathbb{R}^{m\times m},\, \delta\in\mathbb{R}} f(\mathbf{U}, \delta) \triangleq \left\|\mathbf{K} - (\mathbf{C}\mathbf{U}\mathbf{C}^\top + \delta\mathbf{I}_d)\right\|_F^2.$$

It has global solution

$$\mathbf{U}^{\mathrm{ss}} = \mathbf{C}^\dagger\mathbf{K}(\mathbf{C}^\dagger)^\top - \delta^{\mathrm{ss}}(\mathbf{C}^\top\mathbf{C})^\dagger$$

and

$$\delta^{\mathrm{ss}} = \frac{1}{d-m}\left(\mathrm{tr}(\mathbf{K}) - \mathrm{tr}(\mathbf{C}^\dagger\mathbf{K}\mathbf{C})\right).$$

Applications in Matrix Approximation

We can show that

$$\mathbf{C}\mathbf{U}^{\mathrm{ss}}\mathbf{C}^\top + \delta^{\mathrm{ss}}\mathbf{I}_d \succ \mathbf{0}$$

and

$$\begin{aligned}&(\mathbf{Q}\mathbf{U}^{\mathrm{ss}}\mathbf{Q}^\top + \delta^{\mathrm{ss}}\mathbf{I}_d)^{-1} \\ =&(\delta^{\mathrm{ss}})^{-1}\mathbf{I}_n - (\delta^{\mathrm{ss}})^{-2}\mathbf{Q}(\mathbf{I}_m + (\delta^{\mathrm{ss}})^{-1}\mathbf{U}^{\mathrm{ss}})^{-1}\mathbf{U}^{\mathrm{ss}}\mathbf{Q}^\top.\end{aligned}$$

is well-defined, where $\mathbf{Q} \in \mathbb{R}^{d\times m}$ is the orthogonal bias of $\mathbf{C} \in \mathbb{R}^{d\times m}$.