Optimization Theory

Lecture 10

Fudan University

luoluo@fudan.edu.cn

Outline

Outline

Damped Newton Method

The damped Newton method is based on

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \frac{1}{1 + M_f \lambda_f(\mathbf{x}_t)} \left(\nabla^2 f(\mathbf{x}_t)\right)^{-1} \nabla f(\mathbf{x}_t),$$

where $M_f > 0$ and

$$\lambda_f(\mathbf{x}_t) = \sqrt{\left\langle \nabla f(\mathbf{x}_t), \left(\nabla^2 f(\mathbf{x}_t)\right)^{-1} \nabla f(\mathbf{x}_t) \right\rangle}.$$

This method has global convergence guarantee under mild assumptions.

Self-Concordant Functions

Definition

We say $f : \mathbb{R}^d \to \mathbb{R}$ is M-strongly self-concordant, if it is twice differentiable and holds

$$\nabla^2 f(\mathbf{x}) - \nabla^2 f(\mathbf{y}) \preceq M \|\mathbf{x} - \mathbf{y}\|_{\nabla^2 f(\mathbf{z})} \nabla^2 f(\mathbf{w}),$$

for any $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w} \in \mathbb{R}^d$ and some $M > 0$.

1. The strong self-concordant property is affine invariant.
2. If $f : \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex and has L_2-Lipschitz continuous Hessian, then it is M-strongly self-concordant with

$$M = \frac{L_2}{\mu^{3/2}}.$$

3. The M-strong self-concordance leads to $(M/2)$-self-concordance.

Self-Concordant Functions

Definition

A function $f : \mathbb{R}^d \to \mathbb{R}$ is called self-concordant if there exists a constant $M_f \geq 0$ such that the inequality

$$|D^3 f(\mathbf{x})[\mathbf{h}, \mathbf{h}, \mathbf{h}]| \leq 2M_f \, \|\mathbf{h}\|^3_{\nabla^2 f(\mathbf{x})}$$

holds for any $\mathbf{x}, \mathbf{h} \in \mathbb{R}^d$.

Lemma

A function $f : \mathbb{R}^d \to \mathbb{R}$ is self-concordant if and only if for any $\mathbf{x} \in \mathbb{R}^d$ and any triple of directions $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3 \in \mathbb{R}^d$, we have

$$|D^3 f(\mathbf{x})[\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3]| \leq 2M_f \prod_{i=1}^{3} \|\mathbf{h}_i\|_{\nabla^2 f(\mathbf{x})}$$

Outline

Global Convergence Analysis

To the ease of presentation, we take $M = 2$ ($M_f = 1$). Then iteration

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \frac{1}{1 + \lambda_f(\mathbf{x}_t)} \left(\nabla^2 f(\mathbf{x}_t)\right)^{-1} \nabla f(\mathbf{x}_t)$$

leads to global convergence of $\lambda_f(\mathbf{x}_t)$.

1. For $\lambda_f(\mathbf{x}_t) \geq 1/4$, we have

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}_t) \leq -\frac{1}{38}.$$

2. For $\lambda_f(\mathbf{x}_t) \leq 1/4$, we have

$$\lambda_f(\mathbf{x}_{t+1}) \leq 2(\lambda_f(\mathbf{x}_t))^2.$$

Global Convergence Analysis

Let $\rho(z) = -\ln(1-z) - z$ and

$$\delta = \sqrt{(\mathbf{y}-\mathbf{x})^\top \nabla^2 f(\mathbf{x})(\mathbf{y}-\mathbf{x})} < 1,$$

then we have

$$\rho(-\delta) \leq f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x} \rangle \leq \rho(\delta),$$
$$(1-\delta)^2 \nabla^2 f(\mathbf{x}) \preceq \nabla^2 f(\mathbf{y}) \preceq \frac{1}{(1-\delta)^2} \nabla^2 f(\mathbf{x})$$

and

$$\left\| \nabla f(\mathbf{x})^{-1/2} \left(\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})(\mathbf{y}-\mathbf{x}) \right) \right\|_2 \leq \frac{\delta^2}{1-\delta}.$$