Columbia University in the City of New York
Optimization Methods for Machine Learning
Instructors: Satyen Kale
Authors: Gurpreet Singh
Email: gurpreet.s@columbia.edu

SCRIBE

5

Properties of Convex Functions

1. Convexity, Strong Convexity and Smoothness

We start with redefining strong convexity and smoothness that were discussed in the previous class.

Definition 5.1 (Strong Convexity) A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be α-strongly convex for some $\alpha > 0$ if and only if $\forall x \in \mathbb{R}^d$, there exists a subgradient g of f at x such that for any $y \in \mathbb{R}^d$, we have

$$f(y) \;\geq\; f(x) + \nabla g^{\mathrm{T}}(y - x) + \frac{1}{2}\alpha \| x - y \|_2^2$$

For example, consider the function $f : \mathbb{R}^d \to \mathbb{R}$ where $f(x) = \| x \|_2^2 = \langle x,\ x \rangle$. The gradient is given by $\nabla f(x) = 2x$. Then, we have

$$\begin{aligned} f(y) - f(x) - \nabla f(x)^{\mathrm{T}}(y - x) &= \| y \|_2^2 - \| x \|_2^2 - 2x^{\mathrm{T}}(y - x) \\ &= \| x \|_2^2 + \| y \|_2^2 - 2 \langle x,\ y \rangle \\ &= \| y - x \|_2^2 \end{aligned}$$

This suggests that the strong convexity constant $\alpha = 2$. In fact f is α-strongly convex for any $\alpha \in (0, 2]$.

Strong convexity ensures that the function remains above a paraboloid defined at any point x with the two curves touching at x.

Definition 5.2 (Smoothness) A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be β-smooth for some $\beta > 0$ if and only if $\forall x, y \in \mathbb{R}^d$, we have

$$f(y) \;\leq\; f(x) + \nabla f(x)^{\mathrm{T}}(y - x) + \frac{1}{2}\beta \| x - y \|_2^2$$

Essentially β-smoothness ensures that the function remains below some paraboloid defined at any point x with the two curves touching at x.

Note A function need not be convex to be β-smooth *i.e.* even non-convex functions can be smooth. For non-convex functions, the definition of β-smoothness needs to be amended to the following:

$\forall x, y \in \mathbb{R}^d$, we have

$$|f(y) - (f(x) + \nabla f(x)^{\mathrm{T}}(y - x))| \leq \frac{1}{2}\beta \,\|\, x - y \,\|_2^2$$

If a function is smooth, then it is necessarily differentiable everywhere.

For example, $f(x) = \|\, x \,\|_2^2$ is β-smooth where $\beta \geq 2$.

Exercise 5.1 If a function $f : \mathbb{R}^d \to \mathbb{R}$ is both α-strongly convex and β-strongly smooth with $\alpha = \beta$ then prove that the form of f is given as

$$f(x) \;=\; \frac{1}{2}\alpha \,\|\, x - \mathbf{a} \,\|^2 + \langle \mathbf{b},\; x \rangle + c$$

It is often difficult to analytically tell if a function is strongly convex or smooth. The following theorem, however, offers a simple condition that implies strong convexity or smoothness for twice differentiable functions.

Theorem 5.1 Suppose a function $f : \mathbb{R}^d \to \mathbb{R}$ is twice differentiable at all points $x \in \mathbb{R}^d$, then

(i) if $\forall x \in \mathbb{R}^d$, $\nabla^2 f(x) \succcurlyeq \alpha\mathbb{I}$, then f is α-strongly convex

(ii) if $\forall x \in \mathbb{R}^d$, $-\beta\mathbb{I} \preccurlyeq \nabla^2 f(x) \preccurlyeq \beta\mathbb{I}$, then f is β-smooth.

Proof. Using Taylor's theorem, we have

$$f(y) \;=\; f(x) + \nabla f(x)^{\mathrm{T}}(y - x) + \frac{1}{2}(y - x)^{\mathrm{T}}\nabla^2 f(\boldsymbol{\gamma})(y - x)$$

where $\boldsymbol{\gamma} \in [\, x, y \,]$

(i) If $\nabla^2 f(\boldsymbol{\gamma}) \succcurlyeq \alpha\mathbb{I}$, then we have

$$\begin{aligned} (y - x)^{\mathrm{T}}\nabla^2 f(\boldsymbol{\gamma})(y - x) &\geq (y - x)^{\mathrm{T}}\alpha\mathbb{I}(y - x) \\ &= \alpha \,\|\, y - x \,\|^2 \end{aligned}$$

Plugging this in the taylor's expansion we wrote above, we have

$$f(y) \;\geq\; f(x) + \nabla f(x)^{\mathrm{T}}(y - x) + \frac{1}{2}\alpha \,\|\, x - y \,\|_2^2$$

(ii) If $-\beta\mathbb{I} \preccurlyeq \nabla^2 f(\boldsymbol{\gamma}) \preccurlyeq \beta\mathbb{I}$, then we have

$$\begin{aligned} (y - x)^{\mathrm{T}}\nabla^2 f(\boldsymbol{\gamma})(y - x) &\leq (y - x)^{\mathrm{T}}\beta\mathbb{I}(y - x) \\ &= \beta \,\|\, y - x \,\|^2 \end{aligned}$$

Plugging this in the Taylor's expansion we wrote above, we have

$$f(y) \;\leq\; f(x) + \nabla f(x)^{\mathrm{T}}(y - x) + \frac{1}{2}\beta \,\|\, x - y \,\|_2^2$$

Similarly one can prove that

$$f(y) \;\geq\; f(x) + \nabla f(x)^{\mathrm{T}}(y - x) - \frac{1}{2}\beta \,\|\, x - y \,\|_2^2 .$$

These two inequalities imply that the function is β smooth.

□

Theorem 5.2 Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function then if x is a local minimum of f, then x is a global minimum.

Proof. Suppose if there is some $y \in \mathbb{R}^d$ such that $f(y) < f(x)$. From convexity of f, we have

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\leq \lambda f(x) + (1-\lambda) f(y) \\ f(x) - (\lambda f(x) + (1-\lambda) f(y)) &\leq f(x) - f(\lambda x + (1-\lambda) y) \\ \implies (1-\lambda)(f(x) - f(y)) &\leq f(x) - f(\lambda x + (1-\lambda) y) \end{aligned}$$

The LHS is greater than 0 for $\lambda < 1$ since $f(y) < f(x)$, whereas the RHS, for λ close to 1, is non-negative since x is a local minimum. Therefore we obtain a contradiction. □

In the next lecture we discuss gradient descent as a solver for optimization problems and we use the theorems discussed above when analysing convergence of descent methods for convex functions.