MATH: Operations Research **2014-15 First Term**

Handout 11: Properties of Lipscitz and convex functions

Instructor: Junfeng Yang *December 2, 2014*

Contents

11.1 Lipschitz continuous functions

Notation.

- $C^k(R^n)$: functions k times continuously differentiable on R^n.
- $C_L^{k,p}(R^n)$: a subset of $C^k(R^n)$, and the pth order derivative of any f in this class is Lipschitz continuous on R^n with the constant $L > 0$, i.e.,

$$\|f^{(p)}(x) - f^{(p)}(y)\| \le L\|x - y\|, \quad \forall x, y \in R^n.$$

- The most important class in $C_L^{k,p}(R^n)$ is $C_L^{1,1}(R^n)$, a function f in which is continuously differentiable and satisfies

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall x, y \in R^n.$$

(thus, ∇f is uniformly continuous on R^n.)

Theorem 11.1 (Hessian uniformly bounded) *Let $f \in C^2(R^n)$. Then $f \in C_L^{2,1}(R^n)$ if and only if $\|\nabla^2 f(x)\| \le L$ for all $x \in R^n$.*

Proof: Necessity. For any $s \in R^n$ and $\alpha > 0$, it holds

$$\left\| \int_0^\alpha \nabla^2 f(x + \tau s) \mathrm{d}\tau \cdot s \right\| = \|\nabla f(x + \alpha s) - \nabla f(x)\| \le \alpha L \|s\|.$$

Dividing both sides by α and letting $\alpha \to 0+$.

Sufficiency. For any $x, s \in R^n$, it holds that

$$\begin{aligned} \|\nabla f(x + s) - \nabla f(x)\| &= \left\| \int_0^1 \nabla^2 f(x + \tau s) \mathrm{d}\tau \cdot s \right\| \\ &\le \int_0^1 \|\nabla^2 f(x + \tau s)\| \mathrm{d}\tau \cdot \|s\| \le L\|s\|. \end{aligned}$$

■

Example 11.1.1

- *Linear function: $f(x) = c^T x + d \in C_0^{1,1}(R^n)$ because*

$$\nabla^2 f(x) \equiv 0.$$

- *Quadratic function: Suppose $A^T = A$. Then*

$$f(x) = \frac{1}{2} x^T A x + b^T x + c \in C_{\|A\|}^{1,1}(R^n)$$

because $\nabla^2 f(x) \equiv A$.

- $f(x) = \sqrt{1 + x^2} \in C_1^{1,1}(R)$ *because* $f''(x) = \frac{1}{(1+x^2)^{3/2}} \le 1$.

Theorem 11.2 *If $f \in C_L^{1,1}(R^n)$, then $\frac{L}{2}\|x\|^2 - f(x)$ is convex.*

Proof: It follows from $f \in C_L^{1,1}(R^n)$ and Cauchy-Schwarz inequality that

$$\langle x - y, \nabla f(x) - \nabla f(y) \rangle \leq L\|x - y\|^2,$$

which is equivalent to

$$\langle x - y, (Lx - \nabla f(x)) - (Ly - \nabla f(y)) \rangle \geq 0,$$

i.e., $Lx - \nabla f(x)$ is monotone. Thus, $\frac{L}{2}\|x\|^2 - f(x)$ is convex. ∎

Theorem 11.3 *Let $f \in C_L^{1,1}(R^n)$. Then for any $x, y \in R^n$ we have*

$$|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq \frac{L}{2}\|y - x\|^2.$$

Proof:

$$f(y) - f(x) - \nabla f(x)^T(y - x) = \int_0^1 \langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x \rangle \, \mathrm{d}\tau.$$

Take absolute value on both sides and amplify the right side properly. ∎

Let $f \in C_L^{1,1}(R^n)$. Define two quadratic functions at the current point $x_k \in R^n$:

$$q(x) = f(x_k) + \nabla f(x_k)^T(x - x_k) - \frac{L}{2}\|x - x_k\|^2,$$
$$Q(x) = f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{L}{2}\|x - x_k\|^2.$$

Then, it holds that

$$q(x) \leq f(x) \leq Q(x), \quad \forall x \in R^n.$$

Theorem 11.4 *Let $f \in C_M^{2,2}(R^n)$. Then for any $x, y \in R^n$ we have*

$$\|\nabla f(y) - \nabla f(x) - \nabla^2 f(x)(y - x)\| \leq \frac{M}{2}\|y - x\|^2,$$
$$|f(y) - q(y; x)| \leq \frac{M}{6}\|y - x\|^3,$$

where

$$q(y; x) := f(x) + \nabla f(x)^T(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x).$$

Proof: Notice that

$$\nabla f(y) - \nabla f(x) - \nabla^2 f(x)(y - x) = \int_0^1 \left(\nabla^2 f(x + \tau(y - x)) - \nabla^2 f(x)\right)(y - x)\mathrm{d}\tau$$

and

$$\begin{aligned}
& f(y) - f(x) - \nabla f(x)^T(y - x) - \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x) \\
= & \int_0^1 \langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x \rangle \, \mathrm{d}\tau - \frac{1}{2}\cdots \\
= & \int_0^1 \left\langle \int_0^1 \nabla^2 f(x + \theta\tau(y - x))(y - x)\mathrm{d}\theta, y - x \right\rangle \tau \mathrm{d}\tau - \frac{1}{2}\cdots \\
= & \int_0^1 \left\langle \int_0^1 \left(\nabla^2 f(x + \theta\tau(y - x)) - \nabla^2 f(x)\right)(y - x)\mathrm{d}\theta, y - x \right\rangle \tau \mathrm{d}\tau.
\end{aligned}$$

Take absolute value on both sides and amplify the right hand side. ∎

Theorem 11.5 *Let $f \in C_M^{2,2}(R^n)$. For any $x, y \in R^n$ with $\|y - x\| = r$, it holds that*

$$\nabla^2 f(x) - M\, r\, I_n \preccurlyeq \nabla^2 f(y) \preccurlyeq \nabla^2 f(x) + M\, r\, I_n,$$

where I_n is the identity matrix. Here $A \succeq B$ means $A - B \succeq 0$.

11.2 Convex Lipschitz continuous functions

Notation.

- $\mathcal{F}(R^n)$: convex functions on R^n.
- $\mathcal{F}^k(R^n)$: the intersection of $C^k(R^n)$ and $\mathcal{F}(R^n)$, i.e., k times continuously differentiable convex functions on R^n.
- $\mathcal{F}_L^{k,p}(R^n)$: a subset of $\mathcal{F}^k(R^n)$, and the pth order derivative of any f in this class is Lipschitz continuous on R^n with the constant L.
- The most important class in $\mathcal{F}_L^{k,p}(R^n)$ is $\mathcal{F}_L^{1,1}(R^n)$, a function f in which is continuously differentiable, convex and satisfies
$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall x, y \in R^n.$$

> *In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity.*
>
> — Rockafellar, 1993.

Consider $\min_{x \in R^n} f(x)$, where $f \in C_L^{1,1}$. In general, gradient type methods converge only to stationary points under this setting. We want to work with a function class $\mathcal{F} \subset C_L^{1,1}$ which satisfies the following assumptions:

Assumption 11.2.1 *1. $\nabla f(x) = 0$ implies that x is a global minimizer of f, $\forall f \in \mathcal{F}$.*

2. If $f_1, f_2 \in \mathcal{F}$ and $\alpha, \beta \ge 0$, then $\alpha f_1 + \beta f_2 \in \mathcal{F}$.

3. Any linear function belongs to $\mathcal{F}$.

Theorem 11.6 *If $\mathcal{F}$ satisfies the above three assumptions, then any $f \in \mathcal{F}$ must be convex.*

Proof: Let $f \in \mathcal{F}$ and x is an arbitrary fixed point. Consider

$$\phi(y) = f(y) - \langle \nabla f(x), y \rangle, \quad y \in R^n.$$

From Assumptions 2 and 3, $\phi \in \mathcal{F}$. Clearly $\nabla \phi(x) = 0$. From Assumption 1, it holds that

$$f(y) - \langle \nabla f(x), y \rangle = \phi(y) \ge \phi(x) = f(x) - \langle \nabla f(x), x \rangle,$$

i.e., f is convex. ■

Theorem 11.7 *Let $f \in C^1(R^n)$. Then $f \in \mathcal{F}_L^{1,1}(R^n)$ if and only if one of the following conditions holds for all $x, y \in R^n$ and $\alpha \in [0,1]$:*

$$0 \le f(y) - f(x) - \nabla f(x)^T(y-x) \le \frac{L}{2}\|y-x\|^2,$$
$$\frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2 \le f(y) - f(x) - \nabla f(x)^T(y-x),$$

$$0 \le \langle \nabla f(x) - \nabla f(y), x-y\rangle \le L\|x-y\|^2,$$
$$\frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \le \langle \nabla f(x) - \nabla f(y), x-y\rangle,$$

$$0 \le \alpha f(x) + (1-\alpha)f(y) - f(\alpha x + (1-\alpha)y) \le \frac{\alpha(1-\alpha)}{2}\cdot L\|x-y\|^2,$$
$$\frac{\alpha(1-\alpha)}{2}\cdot\frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \le \alpha f(x) + (1-\alpha)f(y) - f(\alpha x + (1-\alpha)y).$$

Theorem 11.8 *Let $f \in C^2(R^n)$. Then $f \in \mathcal{F}_L^{2,1}(R^n)$ if and only if*

$$0 \preccurlyeq \nabla^2 f(x) \preccurlyeq LI_n, \quad \forall x \in R^n.$$

Proof: f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in R^n$; ∇f is Lipschitz if and only if $\nabla^2 f(x) \preccurlyeq LI_n$ for all $x \in R^n$. ■

11.3 Strongly convex Lipschitz continuous functions

Notation.

- $\mathcal{S}_\mu^1(R^n)$: continuously differentiable functions that are also strongly convex with constant $\mu > 0$, i.e.,

$$f(y) \ge f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}\|y-x\|^2, \quad \forall x, y \in R^n.$$

- $\mathcal{S}_{\mu,L}^{k,p}(R^n)$: k times continuously differentiable, strongly convex with parameter μ, and the pth order derivative in this class is Lipschitz continuous on R^n with the constant L.
- The most interesting class in $\mathcal{S}_{\mu,L}^{k,p}(R^n)$ is $\mathcal{S}_{\mu,L}^{1,1}(R^n)$, a function f in which satisfies

$$\langle \nabla f(x) - \nabla f(y), x-y\rangle \ge \mu\|x-y\|^2,$$
$$\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|.$$

Definition 11.9 (strongly convex function) *Let $f \in C^1(R^n)$. f is called strongly convex on R^n (denoted by $f \in \mathcal{S}_\mu^1(R^n)$) if there exists a constant $\mu > 0$ such that for any $x, y \in R^n$ we have*

$$f(y) \ge f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}\|y-x\|^2.$$

Theorem 11.10 (strongly convex functions) *Let $f \in C^1$. $f \in \mathcal{S}^1_\mu(R^n)$ if and only if one of the following conditions hold for all $x, y \in R^n$ and $\alpha \in (0,1)$:*

$$\begin{aligned}\langle \nabla f(x) - \nabla f(y), x - y\rangle &\geq \mu \|x-y\|^2,\\ \alpha f(x) + (1-\alpha) f(y) &\geq f(\alpha x + (1-\alpha) y) + \frac{\alpha(1-\alpha)}{2} \cdot \mu \|x-y\|^2.\end{aligned}$$

Theorem 11.11 *If $f \in \mathcal{S}^1_\mu(R^n)$, then for any $x, y \in R^n$ we have*

$$\begin{aligned}f(y) &\leq f(x) + \nabla f(x)^T (y-x) + \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|^2,\\ \langle \nabla f(x) - \nabla f(y), x-y\rangle &\leq \frac{1}{\mu}\|\nabla f(x) - \nabla f(y)\|^2.\end{aligned}$$

Proof: Fix x and consider $\phi(y) := f(y) - \langle \nabla f(x), y\rangle$. First, verify $\phi \in \mathcal{S}^1_\mu(R^n)$. Thus, $\nabla\phi(x) = 0$ implies

$$\phi(x) = \min_z \phi(z) \geq \min_z \{\phi(y) + \langle \nabla \phi(y), z - y\rangle + \frac{\mu}{2}\|z-y\|^2\} = \phi(y) - \frac{1}{2\mu}\|\nabla\phi(y)\|^2,$$

which implies the first. The second follows by adding two copies of the first with x and y interchanged. ■

Theorem 11.12 *If $f \in \mathcal{S}^1_\mu(R^n)$ and $\nabla f(x^*) = 0$, then*

$$\frac{\mu}{2}\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2\mu}\|\nabla f(x)\|^2.$$

Proof: The left inequality follows the definition of strongly convex function, while the right one follows from the last theorem. ■

Theorem 11.13 *If $f \in C^2(R^n)$, then $f \in \mathcal{S}^2_\mu(R^n)$ if and only if for any $x \in R^n$ we have*

$$\nabla^2 f(x) \succeq \mu I_n.$$

Theorem 11.14 *If $f \in C^2(R^n)$, then $f \in \mathcal{S}^{2,1}_{\mu,L}(R^n) \subset \mathcal{S}^{1,1}_{\mu,L}(R^n)$ if and only if*

$$\mu I_n \preccurlyeq \nabla^2 f(x) \preccurlyeq L I_n, \quad \forall x \in R^n.$$

The value $Q_f = L/\mu \geq 1$ is called the condition number of f.

Theorem 11.15 *If $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$ and $\nabla f(x^*) = 0$, then*

$$\begin{aligned}\frac{1}{2L}\|\nabla f(x)\|^2 &\leq f(x) - f(x^*) \leq \frac{1}{2\mu}\|\nabla f(x)\|^2.\\ \frac{\mu}{2}\|x-x^*\|^2 &\leq f(x) - f(x^*) \leq \frac{L}{2}\|x - x^*\|^2.\end{aligned}$$

References

[Nesterov] Yurii Nesterov, Introductory Lectures on Convex Optimization, A Basic Course.