6FMAI19 Nonlinear Optimization	Spring, 2022
Lecture #12 — 13/4, 2022	
Lecturer: Yura Malitsky	*Scribe: Aban Husain*

1 L-smooth functions and strong convexity

Unless otherwise specified, X is a finite dimensional $\mathbb{R}$-vector space equipped with p-norm $\|\cdot\|$.

Definition 1. *The dual space X^* of X is the space of linear forms on X with norm $\|\cdot\|_*$ defined by*

$$\|f\|_* = \max_{\|x\|=1} f(x).$$

As X is assumed to be finite dimensional, there is a natural equivalence between X and X^*, i.e. X^* is an $\mathbb{R}$-vector space of the same dimension as X.

Remark 2. *For X with norm $\|\cdot\|_p$, the dual norm is $\|\cdot\|_q$, where $\frac{1}{p}+\frac{1}{q}=1$ for $p>1$, and $q=\infty$ for $p=1$. In particular, if X has Euclidean norm, then so does X^*.*

Recall that, for an arbitrary function $f : X \to \mathbb{R}$, the **Legendre-Fenchel transform** (or complex conjugate) $f^* : X^* \to \mathbb{R}$ can be constructed as:

$$f^*(y) = \sup_{x\in X}\{\langle y, x\rangle - f(x)\}.$$

Consider the negative entropy function on the n-simplex

$$h(x) = \begin{cases} \sum_{i=1}^n x_i \log x_i & \text{if } x \in \Delta_n = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1\} \\ \infty & \text{else.} \end{cases}$$

Then by straightforward calculation

$$\begin{aligned} h^*(y) &= \sup_{x\in X}\{\langle y, x\rangle - h(x)\} \\ &= \sup_{x\in\Delta_n}\{\langle y, x\rangle - \sum_{i=1}^n x_i \log x_i\} \\ &= \log(\sum_{i=1}^n \exp y_i). \end{aligned}$$

1.1 L-smooth functions

The definition of L-smooth functions can be generalized to X with an unspecified p-norm.

Definition 3. *A differentiable function $f : X \to \mathbb{R}$ is L-smooth with respect to a norm $\|\cdot\|$ if*

$$\|\nabla f(y) - \nabla f(x)\|_* \le L\|y - x\|, \ \forall x, y \in X.$$

Theorem 4. *Let $f : X \to \mathbb{R}$ be convex, and $L > 0$. Then the following are equivalent for all $x, y \in X$ and $\lambda \in [0, 1]$:*

1. *f is L-smooth with respect to $\|\cdot\|$;*

2. $f(y) \le f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y-x\|^2$;
3. $f(y) - f(x) - \langle \nabla f(x), y - x\rangle \ge \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_*^2$;
4. $\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_*^2$;
5. $f(\lambda x + (1-\lambda)y) \ge \lambda f(x) + (1-\lambda) f(y) - \frac{L}{2}\lambda(1-\lambda)\|x-y\|^2$.

Proof. (1) $\Rightarrow$ (2) Let $x_\lambda = x + \lambda(y-x)$ for $\lambda \in [0,1]$. Using the fundamental theorem of calculus and Hölder's inequality:

$$\begin{aligned} f(y) - f(x) - \langle \nabla f(x), y-x\rangle &= \int_0^1 \langle \nabla f(x_\lambda) - \nabla f(x), y - x\rangle \, d\lambda \\ &\le \int_0^1 \|\nabla f(x_\lambda) - \nabla f(x)\|_* \|y-x\| \, d\lambda \\ &\le \int_0^1 L\lambda \|y-x\|^2 \, d\lambda \\ &= \frac{L}{2}\|y-x\|^2. \end{aligned}$$

(2) $\Rightarrow$ (3) For fixed $x \in X$ let

$$\varphi(y) = f(y) - f(x) - \langle \nabla f(x), y - x\rangle.$$

From definition $\nabla\varphi(y) = \nabla f(y) - \nabla f(x)$, and by convexity, $\varphi(x) = 0$ is a minimum value. For $y \in X$, set $z = y - \frac{\|\nabla\varphi(y)\|_*}{L} v$ where v is chosen so that $\langle \nabla\varphi(y), v\rangle = \|\nabla\varphi(y)\|_*$ and $\|v\| = 1$. Then

$$\begin{aligned} 0 &\le \varphi(z) \\ &= \varphi(y) - \langle \nabla\varphi(y), \frac{\|\nabla\varphi(y)\|_*}{L} v\rangle + \frac{L}{2}\|\frac{\|\nabla\varphi(y)\|_*}{L} v\|^2 \\ &= f(y) - f(x) - \langle \nabla f(x), y - x\rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_*^2. \end{aligned}$$

(3) $\Rightarrow$ (4) For each $x, y \in X$,

$$\begin{aligned} f(y) - f(x) - \langle \nabla f(x), y - x\rangle &\ge \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_*^2 \\ f(x) - f(y) - \langle \nabla f(y), x - y\rangle &\ge \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_*^2 \end{aligned}$$

Summation yields (4).

(4) $\Rightarrow$ (1) Using Hölder's inequality,

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

(2) $\Rightarrow$ (5) This follows from the definition of convexity and the inequality in (2).

(5) $\Rightarrow$ (2) Rewrite (5) as

$$f(y) \le f(x) + \frac{f(x + \lambda(y-x)) - f(x)}{\lambda} + \frac{L(1-\lambda)}{2}\|y-x\|^2.$$

The limit as $\lambda \to 0$ results in (2).

□

Claim 5. *The function* $f(x) = \log(\sum_{i=1}^n \exp x_i)$ *is 1-smooth with respect to* $\|\cdot\|_2$ *and* $\|\cdot\|_\infty$.

The first and second order partial derivatives of f are

$$\frac{\partial f}{\partial x_i}(x) = e^{x_i}/(\sum_{k=1}^n e^{x_k}), \quad \frac{\partial f^2}{\partial x_i \partial x_j}(x) = \begin{cases} -e^{x_i}e^{x_j}/(\sum_{i=k}^n e^{x_k})^2, & \text{if } i \neq j \\ -e^{x_i}e^{x_i}/(\sum_{i=k}^n e^{x_k})^2 + e^{x_i}/(\sum_{k=1}^n e^{x_k}) & \text{if } i = j. \end{cases}$$

Fix the notation $\sigma = \nabla f(x)$ and $\nabla^2 f(x) = \text{diag}(\sigma) - \sigma\sigma^T$.

1. In the case of Euclidean norm, L is bounded by the largest eigenvalue of the Hessian. By Weyl's inequality $\nabla^2 f(x) \preccurlyeq \text{diag}(\sigma) \preccurlyeq I$, so f is 1-smooth with respect to $\|\cdot\|_2$.

2. Given $\|\cdot\|_\infty$, for any $d \in \mathbb{R}$ the inequality $\langle \nabla^2 f(x), d\rangle \leq \langle \text{diag}(\sigma), d\rangle \leq \|d\|_\infty$ holds. Since f is twice continuously differentiable, for $x, y \in \mathbb{R}$ there exists some $z \in [x, y]$ such that

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

   By 4, f is 1-smooth with respect to $\|\cdot\|_\infty$.

1.2 μ-strongly convex functions

The definition of strongly convex functions can also be generalized.

Definition 6. *A function* $f : X \to \mathbb{R}$ *is* μ*-strongly convex wrt.* $\|\cdot\|$ *if for all* $x, y \in X$ *and* $\lambda \in [0, 1]$*:*

$$\lambda f(x) + (1-\lambda)f(y) \geq f(\lambda x + (1-\lambda)y) + \frac{\mu}{2}\lambda(1-\lambda)\|y - x\|^2.$$

It is important to note that the equivalence

$$f \text{ is } \mu\text{-strongly convex} \Leftrightarrow f(x) - \frac{\mu}{2}\|x\|^2 \text{ is convex}$$

holds only in the Euclidean case.

Theorem 7. *Let* $f : X \to \mathbb{R} \cup \{\infty\}$*. The following are equivalent for all* $x, y \in X$*:*

1. f *is* μ*-strongly convex with respect to* $\|\cdot\|$*;*
2. $f(y) \geq f(x) + \langle g_x, y - x\rangle + \frac{\mu}{2}\|y - x\|^2, \quad \forall g_x \in \partial f(x)$*;*
3. $\langle g_x - g_y, x - y\rangle \geq \mu\|x - y\|^2, \quad \forall g_x \in \partial f(x), \forall g_y \in \partial f(y)$.

Proof. (1) $\Rightarrow$ (2) Let $x_\lambda = x + \lambda(y - x), \lambda \in [0, 1]$. The definition of μ-strong convexity can be rewritten as

$$f(y) \geq f(x) + \frac{\mu}{2}(1-\lambda)\|y - x\|^2 + \frac{f(x_\lambda) - f(x)}{\lambda}.$$

Allowing $\lambda \to 0$,

$$\begin{aligned} f(y) &\geq f(x) + \frac{\mu}{2}\|y - x\|^2 + \langle \nabla f_{y-x}(x), y - x\rangle \\ &\geq f(x) + \frac{\mu}{2}\|y - x\|^2 + \langle g_x, y - x\rangle \ \forall g_x \in \partial f(x). \end{aligned}$$

(2) $\Rightarrow$ (1) For $x, y \in X$ and $x_\lambda = x + \lambda(y - x),\ \lambda \in [0, 1]$:

$$\lambda f(y) \geq \lambda(f(x_\lambda) + \langle g_{x_\lambda}, y - x_\lambda \rangle + \frac{\mu}{2}\|y - x_\lambda\|^2)$$

$$(1 - \lambda)f(x) \geq (1 - \lambda)(f(x_\lambda) + \langle g_{x_\lambda}, x - x_\lambda \rangle + \frac{\mu}{2}\|x - x_\lambda\|^2).$$

Summation yields (1).

(2) $\Rightarrow$ (3) Monotonicity follows immediately from (2).

(3) $\Rightarrow$ (2) For $\lambda \in [0, 1]$, let $x_\lambda = x + \lambda(y - x)$. Given that f is convex, for $g_{x_\lambda} \in \partial f(x_\lambda)$,

$$f(y) - f(x) = \int_0^1 \langle g_{x_\lambda}, y - x \rangle \, d\lambda.$$

Since $\langle g_{x_\lambda}, y - x \rangle \geq \langle g_x, y - x \rangle + \mu\lambda\|x - y\|^2$, (2) follows. □

2 Fenchel duality of L-smooth and strongly convex functions

In the last lecture, the following relations between subgradients of a function and its convex conjugate were established.

Lemma 8. *[Fenchel Young's equality] For a proper, lower semicontinuous convex function $f : X \to \mathbb{R}$, the following conditions are equivalent:*

1. $f(x) + f^*(y) = \langle y, x \rangle$;
2. $x \in \partial f^*(y)$;
3. $y \in \partial f(x)$.

Theorem 9. *Let $f : X \to \mathbb{R}$. The following statements hold:*

1. *If f is closed and μ-strongly convex with respect to $\|\cdot\|$, then f^* is is $\frac{1}{\mu}$-smooth with respect to $\|\cdot\|_*$;*
2. *If f is convex and L-smooth with respect to $\|\cdot\|$, then f^* is is $\frac{1}{L}$-strongly convex with respect to $\|\cdot\|_*$.*

Proof. Both statements are direct consequences of Fenchel Young, 4 and 7. □

Claim 10. *The negative entropy function $h(x)$ on the n-simplex is 1-stronlgy convex with respect to both $\|\cdot\|_1$ and $\|\cdot\|_2$.*

Since the complex conjuagate of $h(x)$ is 1-smooth with respect to $\|\cdot\|_2$ and $\|\cdot\|_\infty$, 9 ensures 1-strong convexity with respect to the dual norms.