IE 521: Convex Optimization **Spring 2017, UIUC**

Lecture 6: Convex Functions, Part III – February 8

Instructor: Niao He *Scribe: Shuanglong Wang*

Courtesy warning: These notes do not necessarily cover everything discussed in the class. Please email TA (swang157@illinois.edu) if you find any typos or mistakes.

In this lecture, we cover the following topics

- subgradient and subdifferential
  - definition and examples, existence, subdifferential properties, calculus

References: Bertsekas, Nedich & Ozdaglar, 2003, Chapter 4

6.1 Motivation

Convex functions are "essentially" convex sets:

- $f(x)$ is convex $\Longleftrightarrow$ epi(f) is convex
- $\sup_{\alpha \in A} f_\alpha(x)$ is convex $\Longleftrightarrow$ $\cap_{\alpha \in A}$epi(f_α) is convex
- "tight affine minorant" of f $\Longleftrightarrow$ supporting hyperplane of epi(f)

Question: Can you find any affine function that underestimates $f(x)$ and is tight at $x = 0$?

- $f(x) = \frac{1}{2}x^2$
- $f(x) = |x|$
- $f(x) = \begin{cases} -\sqrt{x}, & x \geq 0 \\ +\infty, & o.w. \end{cases}$
- $f(x) = \begin{cases} 1, & x = 0 \\ 0, & x > 0 \\ +\infty, & o.w. \end{cases}$

6.2 Subgradient

Definition 6.1 *Let $f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ be a convex function. A vector $g \in \mathbf{R}^n$ is a subgradient of f at a point $x \in dom(f)$ if*

$$f(y) \geq f(x) + g^T(y - x), \forall y$$

The set of all subgradient at x is called the subdifferential of f at x denoted as ∂f.

Remark (Subgradient and epigraph) Note that for any fixed $x \in dom(f)$

$$\begin{aligned} g \ \in \partial f(x) &\Leftrightarrow f(y) - g^T y \geq f(x) - g^T x, \forall y \\ &\Leftrightarrow t - g^T y \geq f(x) - g^T y, \forall (y,t) \in \text{epi}(f) \\ &\Leftrightarrow \begin{bmatrix} -g \\ 1 \end{bmatrix}^T \begin{bmatrix} y \\ t \end{bmatrix} \geq \begin{bmatrix} -g \\ 1 \end{bmatrix}^T \begin{bmatrix} y \\ f(x) \end{bmatrix}, \forall (y,t) \in \text{epi}(f) \\ &\Leftrightarrow \text{The hyperplane } H := \left\{(y,t) : (-g,1)^T(y,t) = (-g,1)^T(x, f(x))\right\} \\ &\quad\ \text{is a supporint plane of epi}(f) \end{aligned}$$

Remark (Differentiable Case) If f is differentiable at $x \in dom(f)$, then $\partial f(x) = \{\nabla f(x)\}$ is a singleton.

Proof: By definition, let $y = x + \epsilon d$, $g \in \partial f(x)$, $f(x + \epsilon d) \geq f(x) + \epsilon g^T d$.

$$\frac{f(x + \epsilon d) - f(x)}{\epsilon} \geq g^T d, \forall d, \forall \epsilon \overset{\epsilon \to 0}{\Longrightarrow} \nabla f(x)^T d \geq g^T d, \forall d$$

This only holds when $g = \nabla f(x)$. ∎

Examples

1. $f(x) = \frac{1}{2}x^2$, $\partial f(x) = x$
2. $f(x) = |x|$, $\partial f(x) = \begin{cases} \text{sgn}(x), x \neq 0 \\ [-1,1], x = 0 \end{cases}$. Note at $x = 0$, $\forall g \in [-1,1], |y| \geq 0 + g \cdot y, \forall y$
3. $f(x) = \begin{cases} -\sqrt{x}, x \geq 0 \\ +\infty, o.w. \end{cases}$, $\partial f(0) = \emptyset$. Note at $x = 0$, $\nexists g \in \mathbf{R}$, s.t. $\sqrt{y} \geq 0 + g \cdot y, \forall y \geq 0$
4. $f(x) = \begin{cases} 1, x = 0 \\ 0, x > 0 \\ +\infty, o.w. \end{cases}$, $\partial f(0) = \emptyset$. Note at $x = 0$, $\nexists g \in \mathbf{R}$, s.t. $0 \geq 1 + g \cdot y, \forall y \geq 0$

Theorem 6.2 *Let f be a convex function and $x \in dom(f)$. Then*

1. *$\partial f(x)$ is convex and closed*

2. $\partial f(x)$ *is nonempty and bounded if* $x \in rint(dom(f))$

Proof:

1. Convexity and closedness are evident due to

$$\begin{aligned}\partial f(x) &= \left\{g \in \mathbf{R}^n : f(y) \geq f(x) + g^T(y-x), \forall y\right\} \\ &= \cap_y \left\{g \in \mathbf{R}^n f(y) \geq f(x) + g^T(y-x)\right\}\end{aligned}$$

is the solution to an infinite system of linear inequalities. The intersection of arbitrary number of closed and convex sets is still closed and convex.

2. (Non-empty) W.l.o.g., let's assume $dom(f)$ is full-dimensional and $x \in \text{int}(dom(f))$. Since $\text{epi}(f)$ is convex and $(x, f(x))$ belongs to its boundary, by separation theorem, $\exists \alpha = (s, \beta) \neq 0$, s.t.

$$s^T y + \beta t \geq s^T x + \beta f(x), \forall (y,t) \in \text{epi}(f)$$

Clearly, we must have $\beta \geq 0$. Since $x \in int(dom(f))$, we cannot have $\beta = 0$, Otherwise, to ensure $s^T y \geq s^T x, \forall y \in B(x,y), s = 0$, which is impossible.
Hence, $\beta > 0$, setting $g = -\beta^{-1} s$

$$f(y) \geq f(x) + g^T(y-x), \forall y$$

(Bounded) Suppose $\partial f(x)$ is unbounded, i.e. $\exists g_k \in \partial f(x)$, s.t. $\| g_k \|_2 \to \infty$, as $k \to \infty$. Since $x \in int(dom(f))$, $\exists \delta > 0$, s.t. $B(x,\delta) \subseteq dom(f)$. Hence, $y_k = x + \delta \frac{g_k}{\|g_k\|_2} \in dom(f)$. By convexity,

$$f(y_k) \geq f(x) + g_k^T(y_k - x) = f(x) + \delta \| g_k \|_2 \to \infty.$$

However, this contradicts with the continuity of f over $\text{int}(dom(f))$.

■

Remark: On the contrary, if $\forall x \in dom(f), \partial f(x)$ is non-empty, $dom(f)$ is convex, then f is convex. This is because $\forall x, y \in dom(f)$ and $\lambda \in [0,1]$. $z = \lambda x + (1-\lambda)y \in dom(f)$. So $\partial f(z) \neq \emptyset$. Let $g \in \partial f(z)$. we have

$$\left.\begin{array}{l} f(x) \geq f(z) + g^T(x-z) \\ f(y) \geq f(z) + g^T(y-z) \end{array}\right\} \Rightarrow \lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda)y)$$

Remark: **(Continuity of Subdifferential)** Let $f : \mathbf{R}^n \to \mathbf{R}$ be convex and continuous. Suppose $x_k \to x$ and $g_k \in \partial f(x_k)$, $g_k \to g$, then $g \in \partial f(x)$.

6.3 Subdifferential and Directional Derivative

Recall that the directional derivative of a function f at x along direction d is

$$f'(x;d) = \lim_{\delta \to 0^+} \frac{f(x+\delta d) - f(x)}{\delta}$$

If f is differentiable, then $f'(x;d) = \nabla f(x)^T d$

Lemma 6.3 *When f is convex, the ratio $\phi(\delta) := \frac{f(x+\delta d)-f(x)}{\delta}$ is non-decreasing of $\delta > 0$.*

The proof is left as a self exercise.

Theorem 6.4 *Let f be convex and $x \in int(dom(f))$, then*

$$f'(x;d) = \max_{g \in \partial f(x)} g^T d$$

Proof: By definition, We have $f(z+\delta d) - f(x) \geq \delta g^T d$ for all δ and $g \in \partial f(x)$. Hence, $f'(x;d) \geq g^T d, \forall g \in \partial f(x)$. Moreover, this implies that

$$f'(x;d) \geq \max_{g \in \partial f(x)} g^T d.$$

It suffices to show that $\exists \tilde{g} \in \partial f(x)$, s.t. $f'(x;d) \leq \tilde{g}^T d$. Consider the two sets

$$C_1 = \{(y,t) : f(y) < t\}$$
$$C_2 = \left\{(y,t) : y = x + \alpha d, t = f(x) + \alpha f'(x;d), \alpha \geq 0\right\}$$

Claim: $C_1 \cap C_2 = \emptyset$ and C_1, C_2 are closed and nonempty. This is because $f(x+\alpha d) \geq f(x) + \alpha f'(x;d), \forall \alpha \geq 0$. **(Due to Lemma 6.3)**.
By separation theorem, $\exists (g_0, \beta) \neq 0$, s.t.

$$g_0^T(x+\alpha d) + \beta(f(x) + \alpha f'(x;d)) \leq g_0^T y + \beta t, \forall \alpha \geq 0, \forall t > f(y)$$

One can easily show that $\beta > 0$. Let $\tilde{g} = \beta^{-1} g_0$,

$$\tilde{g}^T(x+\alpha d) + f(x) + \alpha f'(x;d) \leq \tilde{g}^T y + f(y), \forall \alpha \geq 0$$

Let $\alpha = 1$ and $y = x$, we have

$$\tilde{g} + f(x) \leq \tilde{g}^T y + f(y) \Leftrightarrow -\tilde{g} \in \partial f(x)$$

Let $\alpha = 1$ and $y = x$, we have

$$\tilde{g}^T d + f'(x;d) \leq 0 \Leftrightarrow f'(x;d) \leq -\tilde{g}^T d$$

Therefore, we have shown that $f'(x;d) = \max_{g \in \partial f(x)} g^T d$ ■

6.4 Calculus of Subgradient

1. Nonnegative summation: If $h(x) = \beta_1 f_1(x) + \beta_2 f_2(x)$, with $\beta_1, \beta_2 \geq 0$, then
$$\partial h(x) = \beta \alpha f_1(x) + \beta_2 \partial f_2(x)$$
2. Affine transformation: If $h(x) = f(Ax+b)$, then
$$\partial h(x) = A^T \partial h(Ax+b)$$
3. Pointwise maximum: If $h(x) = \max_{i \in I} f_i(x), I = \{1, 2, ..., m\}$
$$\partial h(x) = \text{Conv}\left\{\partial f_i(x) | i \in I(x)\right\}, \text{ where } I(x) = \{i \in I : f_i(x) = h(x)\}$$