2.5 Basic Calculus Rules

Proposition: Let $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ be a convex function. Let F be defined by

$$F(x) = f(Ax)$$

where $A \in \mathbb{R}^{m\times n}$. Then

$$A^T \partial f(Ax) \subseteq \partial F(x)$$

Proof. Suppose $A^T g \in A^T \partial f(Ax)$, where $g \in \partial f(Ax)$. Then

$$F(y) - F(x) - \langle A^T g, y - x\rangle = f(Ay) - f(Ax) - \langle g, Ay - Ax\rangle \geq 0$$

□

Theorem:(Moreau-Rockafellar) Let $f, g : \mathbb{R}^n \to (-\infty, \infty]$ be proper convex functions. Then for every $x_0 \in \mathbb{R}^n$

$$\partial f(x_0) + \partial g(x_0) \subset \partial(f + g)(x_0)$$

Moreover, suppose int dom$(f) \cap$ dom$(g) \neq \emptyset$. Then for every $x_0 \in \mathbb{R}^n$,

$$\partial f(x_0) + \partial g(x_0) = \partial(f + g)(x_0)$$

Proof. Let $u_1 \in \partial f(x_0),\ u_2 \in \partial g(x_0)$. Then for every $x \in \mathbb{R}^n$,

$$f(x) \geq f(x_0) + \langle u_1, x - x_0\rangle,\ g(x) \geq g(x_0) + \langle u_2, x - x_0\rangle$$

Hence, adding the two inequalities shows that $u + v \in \partial(f + g)(x_0)$.
Now, let $v \in \partial(f + g)(x_0)$. Note that $f(x_0) \neq \infty$, otherwise this implies that $f + g \equiv \infty$. Similarly, $g(x_0) \neq \infty$. Next, consider the following two sets

$$\Lambda_f := \{(x - x_0, y) :\ y > f(x) - f(x_0) - \langle v, x - x_0\rangle\}$$
$$\Lambda_g := \{(x - x_0, y) :\ -y \geq g(x) - g(x_0)\}.$$

Λ_f, Λ_g are both nonempty and convex (consider epi(f), epi(g)). Also, since $v \in \partial(f + g)(x_0),\ \Lambda_f \cap \Lambda_g = \emptyset$ (otherwise, adding the above two inequalities contradict the fact that $v \in \partial(f + g)$)
Then Λ_f, Λ_g can be separated by a hyperplane. So there exists $(a, b) \neq 0, c$ such that

$$\langle a, x - x_0\rangle + by \leq c,\ \forall(x, y) \text{ such that } y > f(x) - f(x_0) - \langle v, x - x_0\rangle$$

$$\langle a, x - x_0\rangle + by \geq c,\ \forall(x, y) \text{ such that } -y \geq g(x) - g(x_0)$$

Since $(0, 0) \in \Lambda_g$, $c \leq 0$. Since $(0, 1) \in \Lambda_f, b \leq 0$.
For all $\epsilon > 0, (0, \epsilon) \in \Lambda_f$, since $b \leq 0$, letting $\epsilon \to 0$, we get $c \geq 0$. Hence $c = 0$.
Suppose $b = 0$, we have

$$\langle a, x - x_0\rangle \leq 0,\ \forall(x, y) \text{ such that } y > f(x) - f(x_0) - \langle v, x - x_0\rangle$$

$$\langle a, x - x_0\rangle \geq 0,\ \forall (x,y) \text{ such that } -y \geq g(x) - g(x_0)$$

which are equivalent to

$$\langle a, x - x_0\rangle \leq 0,\ \forall x \in \text{dom}(f)$$

$$\langle a, x - x_0\rangle \geq 0,\ \forall x \in \text{dom}(g)$$

Let $\overline{x} \in \text{int dom}(f) \cap \text{dom}(g)$. Then $\langle a, \overline{x} - x_0\rangle = 0$. Since $\overline{x} \in \text{int dom}(f)$, there exists $\delta > 0$ such that $B(\overline{x}, \delta) \subset \text{dom}(f)$. Then

$$\langle a, \frac{\delta a}{2}\rangle = \langle a, \overline{x} + \frac{\delta a}{2} - x_0\rangle \leq 0$$

So $a = 0$. This contradicts the fact that $(a, b) \neq 0$. Hence $b < 0$.
Let $-u_2 = \frac{a}{-b}$, we have

$$\langle -u_2, x - x_0\rangle \leq y,\ \forall (x,y) \text{ such that } y > f(x) - f(x_0) - \langle v, x - x_0\rangle.$$

$$\langle -u_2, x - x_0\rangle \geq y, \forall (x,y) \text{ such that } -y \geq g(x) - g(x_0)$$

Consider $y = g(x_0) - g(x)$, then $u_2 \in \partial g(x_0)$.
By considering $(x, f(x) - f(x_0) - \langle v, x - x_0\rangle + \epsilon$ and letting $\epsilon \to 0$, we have $u_1 = v - u_2 \in \partial f(x_0)$.
Hence $v = u_1 + u_2 \in \partial f(x_0) + \partial g(x_0)$.
Therefore $\partial(f+g)(x_0) \subset \partial f(x_0) + \partial g(x_0)$. □

2.5.1 Directional Derivative

Definition:(Directional Derivative) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function with $x \in \text{dom} f$. The *directional derivative* of f at x with direction d is given by

$$f'(x; d) = \lim_{t \to 0^+} \frac{f(x + td) - f(x)}{t}$$

Lemma: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function with $x \in \text{dom} f$. Then for all direction $d \in \mathbb{R}^n$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ with $\lambda_2 > \lambda_1 > 0$, we have

$$\frac{f(x + \lambda_1 d) - f(x)}{\lambda_1} \leq \frac{f(x + \lambda_2 d) - f(x)}{\lambda_2}$$

Proof. Note that $x + \lambda_1 d = \frac{\lambda_1}{\lambda_2}(x + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2})x$. Then

$$f(x + \lambda_1 d) \leq \frac{\lambda_1}{\lambda_2} f(x + \lambda_2 d) + (1 - \frac{\lambda_1}{\lambda_2}) f(x)$$

The result follows from the above inequality. □

Lemma: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function with $x \in \text{int}(\text{dom} f)$. Then $f'(x; d)$ is finite for every direction $d \in \mathbb{R}^n$.

Proof. Recall that f is locally Lipschitz at x. Then for t small,

$$|\frac{f(x+td)-f(x)}{t}| \le \frac{Lt||d||}{t} \le L||d|| < \infty$$

□

Theorem: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function with $x \in \text{int}(\text{dom} f)$. Then

$$f'(x;d) = \sup_{g \in \partial f(x)} \langle g, d \rangle$$

Proof. By the above proposition, we have $f'(x;d) = \inf_{t>0} \frac{f(x+td)-f(x)}{t}$.
Define $\psi(d) := f'(x;d)$. Then ψ is convex and finite for every d.
Therefore, ψ is continuous and hence closed.
Hence, $\psi = \psi^{**} = \sup_g \{\langle g, d\rangle - \psi^*(g)\}$.
We will show that

$$\psi^*(g) = \begin{cases} 0 & g \in \partial f(x) \\ \infty & \text{otherwise} \end{cases}$$

Note that $\psi(0) = 0$. Then for all g,

$$\psi^*(g) \ge \langle g, 0\rangle - \psi(0) = 0$$

Suppose $g \in \partial f(x)$. Then $\langle g, d\rangle - \psi(d) \le \frac{f(x+td)-f(x)}{t} - \psi(d)$ for all $t > 0$. So

$$\langle g, d\rangle - \psi(d) \le f(x;d) - \psi(d) = 0 \text{ for all } d$$

Therefore, $\psi^*(g) = \sup_d \{\langle g, d\rangle - \psi(d)\} \le 0$.
Suppose $g \notin \partial f(x)$. Then there exists y such that

$$\langle g, y - x\rangle \ge f(y) - f(x)$$

Write $y = x + d_0$, then we have $\langle g, d_0\rangle \ge f(x+d_0) - f(x) \ge f'(x;d_0)$.
Note that $t\psi(d) = \psi(td)$, then

$$\psi^*(g) = \sup_d \{\langle g, d\rangle - \psi(d)\} \ge \sup_{t>0} \{\langle g, td\rangle - \psi(td)\} = \sup_{t>0} \{t(\langle g, d\rangle - \psi(d))\} \ge \infty$$

Consider $\psi^{**}(g) = \sup_d \{\langle g, d\rangle - \psi^*(g)\}$.
It follows that $\psi^{**}(g) = \sup_{g \in \partial f(x)} \langle g, d\rangle$.
Hence, $f'(x;d) = \psi(d) = \psi^{**}(d) = \sup_{g \in \partial f(x)} \langle g, d\rangle$. □

Theorem:(Dubovitskii-Milyutin) Let $f_1, ..., f_m : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex functions and let $\overline{x} \in \cap_m \text{int}(\text{dom} f_i)$. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be given by

$$f(x) := \max_m f_i(x)$$

and let $I(\overline{x}) = \{i |\ f_i(\overline{x}) = f(\overline{x})\}$. Then

$$\partial f(\overline{x}) = \text{conv}\big(\bigcup_{i \in I(\overline{x})} \partial f_i(\overline{x})\big).$$

Proof. Note that if $g \in \partial f_i(\overline{x})$, then $g \in \partial f(\overline{x})$ for all $i \in I(\overline{x})$.
Also, since $\partial f(\overline{x})$ is convex, then $\operatorname{conv}\big(\bigcup_{i\in I(\overline{x})} \partial f_i(\overline{x})\big) \subseteq \partial f(\overline{x})$.
So suppose $g_0 \in \partial f(\overline{x})$ but $g_0 \notin \operatorname{conv}\big(\bigcup_{i\in I(\overline{x})} \partial f_i(\overline{x})\big)$.
Note that $\operatorname{conv}\big(\bigcup_{i\in I(\overline{x})} \partial f_i(\overline{x})\big)$ is compact (Each $\partial f_i(\overline{x})$ is compact).
Then there exists d such that

$$\langle g_0, d\rangle > \max_{i\in I(\overline{x})} \sup_{g\in\partial f_i(\overline{x})} \langle g, d\rangle = \max_{i\in I(\overline{x})} f_i'(\overline{x}; d)$$

We claim that $f'(\overline{x}; d) = \max_{i\in I(\overline{x})} f_i'(\overline{x}; d)$. Then $\langle g_0, d\rangle > f'(\overline{x}; d)$.
But since $g_0 \in \partial f(\overline{x})$, then $f(\overline{x} + td) - f(\overline{x}) \geq \langle g_0, d\rangle$ for all $t > 0$.
Then $f'(\overline{x}; d) \geq \langle g_0, d\rangle$. This is a contradiction.
Therefore $g_0 \in \operatorname{conv}\big(\bigcup_{i\in I(\overline{x})} \partial f_i(\overline{x})\big)$.
It remains to prove that $f'(\overline{x}; d) = \max_{i\in I(\overline{x})} f_i'(\overline{x}; d)$. First for all $t > 0$,

$$\frac{f(\overline{x} + td) - f(\overline{x})}{t} \geq \frac{f_i(\overline{x} + td) - f_i(\overline{x})}{t} \text{ for all } i \in I(\overline{x})$$

Then $f'(\overline{x}; d) \geq f_i'(\overline{x}; d)$. Consider $\{t_k\}$ with $t_k \downarrow 0$ and $x_k = \overline{x} + t_k d$.
Then there exists $\overline{i}$ such that $\overline{i} \in I(x_k)$ for infinitely many k.
Without loss of generality, assume $\overline{i} \in I(x_k)$ for all k.
Then $f_{\overline{i}}(x_k) \geq f_i(x_k)$ for all i, k.
Taking limit and since f_i are continuous at $\overline{x}$, we have

$$f_{\overline{i}}(x) \geq f_i(x) \text{ for all } i$$

Hence

$$f'(\overline{x}; d) = \lim_{k\to\infty} \frac{f(\overline{x} + t_k d) - f(\overline{x})}{t_k} = \lim_{k\to\infty} \frac{f_{\overline{i}}(\overline{x} + t_k d) - f_{\overline{i}}(\overline{x})}{t_k} = f_{\overline{i}}'(\overline{x}; d)$$

Therefore, $f'(\overline{x}; d) = \max_{i\in I(\overline{x})} f_i'(\overline{x}; d)$. $\square$