2.4 Subgradient of Convex Function

In this section, we introduce the crucial concept of subgradient for convex functions. It acts as generalized derivative for nonsmooth functions and has many applications in optimization theory.

Definition:(Subgradient) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let $\overline{x} \in \text{dom} f$. An element $g \in \mathbb{R}^n$ is called a *subgradient* of f at $\overline{x}$ if

$$f(x) - f(\overline{x}) \geq \langle g, x - \overline{x} \rangle \text{ for all } x \in \mathbb{R}^n$$

The collection of all subgradients of f is denoted by $\partial f(\overline{x})$.

Proposition: Let f be a convex function and let $\overline{x} \in \text{int}(\text{dom} f)$, then $\partial f(\overline{x})$ is nonempty and compact.

Proof. Since f is convex, epi$f$ is a convex set.
By the supporting hyperplane theorem to epi$f$ and the point $(\overline{x}, f(\overline{x}))$, there exists $(a, b) \neq 0$ such that

$$\left\langle \begin{bmatrix} a \\ b \end{bmatrix}, \left(\begin{bmatrix} x \\ t \end{bmatrix} - \begin{bmatrix} \overline{x} \\ f(\overline{x}) \end{bmatrix} \right) \right\rangle \leq 0, \text{ for all } (x, t) \in \text{epi} f$$

By considering $(\overline{x}, t) \in \text{epi} f$, we must have $b \leq 0$. Also

$$\langle a, x - \overline{x} \rangle + b(f(x) - f(\overline{x})) \leq 0 \text{ for all } x$$

Suppose $b = 0$, this implies $\langle a, x - \overline{x} \rangle \leq 0$.
This is impossible since $\overline{x} \in \text{int}(\text{dom} f)$. Hence, $b < 0$. Then

$$\left\langle -\frac{a}{b}, x - \overline{x} \right\rangle \leq f(x) - f(\overline{x})$$

Therefore, $-\frac{a}{b} \in \partial f(\overline{x}) \neq \emptyset$.
Recall that a function is locally Lipschitz continuous on the int(dom$f$).
So there exists $\epsilon > 0$ such that

$$f(x) - f(y) \leq L||x - y||, \text{ for all } x, y \in B(\overline{x}; \epsilon)$$

Let $g \in \partial f(\overline{x})$. Consider $x = \overline{x} + \frac{\epsilon g}{||g||}$, then

$$\epsilon||g|| = \langle g, x - \overline{x} \rangle \leq f(x) - f(\overline{x}) \leq L||x - \overline{x}|| = L\epsilon$$

Then we have $||g|| \leq L$. Therefore $\partial f(\overline{x})$ is bounded.
It follows from the definition that $\partial f(\overline{x})$ is closed and hence compact. □

For a differentiable convex function, the subdifferential is just the usual gradient.
Proposition: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and differentiable at $\overline{x} \in \text{int}(\text{dom} f)$. Then $\partial f(\overline{x}) = \{\nabla f(\overline{x})\}$.

Proof. Since f is convex, we have

$$\langle \nabla f(\overline{x}), x - \overline{x} \rangle \leq f(x) - f(\overline{x}) \text{ for all } x \in \mathbb{R}^n$$

So $\nabla f(\overline{x}) \in \partial f(\overline{x})$.
Conversely, suppose $g \in \partial f(\overline{x})$. Since f is differentiable at $\overline{x}$, then for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$f(x) - f(\overline{x}) - \langle \nabla f(\overline{x}), x - \overline{x} \rangle \leq \epsilon ||x - \overline{x}|| \text{ for all } x \text{ with } ||x - \overline{x}|| < \delta$$

Then

$$\langle g - \nabla f(\overline{x}), x - \overline{x} \rangle \leq \epsilon ||x - \overline{x}|| \text{ for all } x \text{ with } ||x - \overline{x}|| < \delta$$

Hence $||g - \nabla f(\overline{x})|| \leq \epsilon$. Since ϵ is arbitrary, this shows that $g = \nabla f(\overline{x})$.
Therefore, $\partial f(\overline{x}) = \{\nabla f(\overline{x})\}$. □

Example: Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be defined by

$$f(x) := \begin{cases} 0 & x \in [-1, 1] \\ |x| - 1 & x \in [-2, 1) \cup (1, 2] \\ \infty & x \in (-\infty, -2) \cup (2, \infty) \end{cases}$$

For $x \in (-2, 1),\ (-1, 1)$ and $(1, 2)$, f is differentiable, hence $\partial f(x) = \{\nabla f(x)\}$.
For $x \in (-\infty, -2) \cup (2, \infty)$, $f(x) = \infty$, hence $\partial f(x) = \emptyset$.
For $x = 1$, we show that $\partial f(x) = [0, 1]$. Let $g \in \partial f(1)$. Then

$$f(y) \geq g(x - 1) \text{ for all } y$$

If $y \in [1, 2]$, then $x - 1 \geq g(x - 1)$, that is $1 \geq g$.
If $y \in [-1, 1]$, then $0 \geq g(x - 1)$, so $g(1 - x) \geq 0$ and $g \geq 0$.
It is easy to check that for $g \in [0, 1],\ g$ satisfies

$$f(y) \geq f(1) + g(x - 1) \text{ for all } y$$

Hence, $\partial f(1) = [0, 1]$.
The subdifferential of other points can be found similarly.
We have

$$\partial f(x) = \begin{cases} \emptyset & x \in (-\infty, -2) \cup (2, \infty) \\ (-\infty, -1] & x = -2 \\ \{-1\} & x \in (-2, -1) \\ [-1, 0] & x = -1 \\ \{0\} & x \in (-1, 1) \\ [0, 1] & x = 1 \\ \{1\} & x \in (1, 2) \\ [1, \infty) & x = 2 \end{cases}$$

The following results show the relationship between subgradients and conjugate of convex functions.

Proposition: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function with $\text{dom} f \neq \emptyset$. Then

$$\langle x, y\rangle \leq f(x) + f^*(y) \text{ for all } x, y$$

Proof. By the definition of conjugate function, $f^*(y) \geq \langle x, y\rangle - f(x)$. □

Theorem: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex with $x \in \text{dom} f$. Then $g \in \partial f(x)$ if and only if

$$f(x) + f^*(g) = \langle g, x\rangle$$

Proof. Suppose $g \in \partial f(x)$, then

$$f(x) + \langle g, y\rangle - f(y) \leq \langle g, x\rangle, \text{ for all } y$$

Then $f(x) + f^*(g) \leq \langle g, x\rangle$. Hence by the above proposition, we have

$$f(x) + f^*(g) = \langle g, x\rangle$$

Suppose $f(x) + f^*(g) = \langle g, x\rangle$, then by the definition of conjugate function,

$$f^*(g) \geq \langle g, y\rangle - f(y) \text{ for all } y$$

Since $f^*(g) = \langle g, x\rangle - f(x)$, we have

$$\langle g, x\rangle - f(x) \geq \langle g, y\rangle - f(y) \text{ for all } y$$

Therefore, $g \in \partial f(x)$. □

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 $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$. 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

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

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

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 $B(\overline{x}, \delta) \subset \mathrm{dom}(f)$. 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,

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

□

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,

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

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

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

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 $\text{conv}\big(\bigcup_{i\in I(\overline{x})} \partial f_i(\overline{x})\big) \subseteq \partial f(\overline{x})$.
So suppose $g_0 \in \partial f(\overline{x})$ but $g_0 \notin \text{conv}\big(\bigcup_{i\in I(\overline{x})} \partial f_i(\overline{x})\big)$.
Note that $\text{conv}\big(\bigcup_{i\in I(\overline{x})} \partial f_i(\overline{x})\big)$ 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 $f(\overline{x} + td) - f(\overline{x}) \ge \langle g_0, d\rangle$ for all $t > 0$.
Then $f'(\overline{x}; d) \ge \langle g_0, d\rangle$. This is a contradiction.
Therefore $g_0 \in \text{conv}\big(\bigcup_{i\in I(\overline{x})} \partial f_i(\overline{x})\big)$.
It remains to prove that $f'(\overline{x}; d) = \max_{i\in I(\overline{x})} f_i'(\overline{x}; d)$. First for all $t > 0$,

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

Then $f'(\overline{x}; d) \ge f_i'(\overline{x}; d)$. 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 $f_{\overline{i}}(x_k) \ge f_i(x_k)$ for all i, k.
Taking limit and since f_i are continuous at $\overline{x}$, we have

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

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, $f'(\overline{x};d) = \max_{i\in I(\overline{x})} f'_i(\overline{x};d)$. □

3 Duality and Optimal Conditions

3.1 Standard forms of optimization problems

In this section, we introduce some of the most basic forms of convex optimization problems.

3.1.1 Linear Programs

A *linear program (LP)* is a problem of the form

$$\begin{aligned} \min_x \; & c^T x \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

3.1.2 Quadratic Programs

A *Quadratic program* is a problem of the form

$$\begin{aligned} \min_x \; & \frac{1}{2} x^T Q x + c^T x \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

3.1.3 Semi-definite Programs (SDP)

A *semi-definite program (SDP)* is a problem of the form

$$\begin{aligned} \min_X \; & C \bullet X \\ & A_i \bullet X = b_i \;\; i = 1, 2..., p \\ & X \succeq \mathbf{0} \end{aligned}$$

3.1.4 Conic Programs

A *conic program* is a problem of the form

$$\begin{aligned} \min_x \; & c^T x \\ & Ax = b \\ & x \in K \end{aligned}$$

where K is a closed convex cone.

3.2 Basics of Convex Optimization

Let's consider the problem

$$\min_{x \in C} f(x)$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function and C is a convex subset of $\mathbb{R}^n$.

Definition: A point $x \in C \cap \mathrm{dom} f$ is called a feasible point.
If there is at least one feasible point, then the problem is called feasible.
A point x^* is called a *minimum* of f over C if

$$x^* \in C \cap \mathrm{dom} f, \quad f(x^*) = \inf_{x \in C} f(x)$$

We may write $x^* \in \arg\min_{x \in C} f(x)$ or even $x^* = \arg\min_{x \in C} f(x)$ if x^* is the unique minimizer.

Other than global minimum, we also have a weaker definition of local minimum, one that is only minimum compared to the points nearby.

Definition:(Local minimizer) We call x^* a local minimum of f over C if $x^* \in C \cap \mathrm{dom} f$ and there exists $\epsilon > 0$ such that

$$f(x^*) \leq f(x), \ \forall x \in C \text{ with } ||x - x^*|| < \epsilon$$

In the convex setting, we have the following nice result.

Proposition: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let C be a convex set. Then a local mimimum of f over C is also a global minimum of f over C. If f is strictly convex, then there exists at most one global minimum of f over C.

Proof. Suppose x^* is a local minimum that is not global.
Then there exists x such that $f(x) < f(x^*)$. Then for $\lambda \in (0, 1)$,

$$f(\lambda x^* + (1 - \lambda)x) \leq \lambda f(x^*) + (1 - \lambda) f(x) < f(x^*)$$

Since f has smaller value on the line connecting x and x^*, this contradicts the local minimality of x^*.
Suppose f is strictly convex, let x^* be a global minimum of f over C. Let $x \in C$ such that $x \neq x^*$. Consider $y = (x + x^*)/2$. Then $y \in C$ and

$$f(y) < \frac{1}{2}(f(x) + f(x^*)) \leq f(x)$$

Since x^* is a global minimum, $f(x^*) \leq f(y)$.
Then $f(x^*) < f(x)$. Hence x^* is the unique global minimum of f over C. □

3.2.1 Existence of solution

Let's consider a general optimization problem

$$\min_{x \in C} f(x)$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $C \subseteq \mathbb{R}^n$.
A basic question is whether a solution to the above problem exists.
Recall the famous Weierstrass theorem. **Proposition:** If f is continuous and C is compact, then there exists a global minimum.

In order to consider cases where C is not bounded (e.g. $\mathbb{R}^n$), we need a new notation.

Definition: (Coercivity) A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called *coercive* if for all sequence $\{x_k\}$ with $\|x_k\| \to \infty$, we have $\lim_{k\to\infty} f(x_k) = \infty$.

Lemma: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a continuous function. Then the following are equivalent.

1. All level sets of f are compact, i.e. $\{x| \ f(x) \leq a\}$ is compact for all a.
2. f is coercive.

Proof. Suppose all level sets of f are compact. Suppose $\{x_k\}$ is a sequence with $||x_k|| \to \infty$. Suppose $f(x_k) \not\to \infty$. Then there exists subsequence x_{k_j} such that $f(x_{k_j})$ is bounded by α for some α. Then $\{x_{k_j}\} \subset V_\alpha$. This contradicts the compactness of V_α. Hence, f is coercive.
Conversely, suppose f is coercive. Suppose V_α is not compact for some α. Since f is continuous, V_α must be closed, this means V_α is not bounded.
Hence, there exists a sequence $\{x_k\} \subset V_\alpha$ such that $||x_k|| \to \infty$. This contradicts the coercivity of f since $f(x_k) \leq \alpha$. $\square$

Proposition: Suppose f is lower-semicontinuous and coercive. Suppose C is non-empty and closed. Then f has a global minimum over C.

Proof. We may assume that $f(x) < \infty$ for some $x \in C$. Then $f^* = \inf_{x \in C} f(x) < \infty$.
Let $\{x_k\} \subset C$ be a sequence such that $\lim f(x_k) = f^* < \infty$. Then since f is coercive, $\{x_k\}$ is bounded. Then there exists a subsequence x_{k_j} converging to a point x^*.
Since C is closed, $x^* \in C$. Then

$$f^* = \lim_{k\to\infty} f(x_k) = \lim_{j\to\infty} f(x_{k_j}) \geq f(x^*)$$

Therefore, x^* is a global minimmum of f over C. $\square$

3.2.2 Optimal condition

For a unconstrained problem, one has a simple optimality test, which is the 'derivative' test in calculus.

Let f be a differentiable convex function on $\mathbb{R}^n$. Then x^* solves

$$\min_{x \in \mathbb{R}^n} f(x)$$

if and only if $\nabla f(x^*) = 0$. How about a constrained problem?
Let's consider the general constrained problem

$$\min_{x \in C} f(x)$$

where C is a convex set, and f is convex.
We have the following result.

Proposition: Let C be a nonempty convex set and let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex differentiable function over an open set that contains C. Then $x^* \in C$ minimizes f over C if and only if

$$\langle \nabla f(x^*), (z - x^*) \rangle \geq 0, \ \forall z \in C.$$

Proof. Suppose $\langle \nabla f(x^*), (z - x^*) \rangle \geq 0, \ \forall z \in C$, then we have,

$$f(z) - f(x^*) \geq \langle \nabla f(x^*), (z - x^*) \rangle \geq 0, \ \forall z \in C.$$

Hence x^* indeed minimizes f over C.
Conversely, suppose x^* minimizes f over C. Suppose on the contrary that $\langle \nabla f(x^*), (z - x^*) \rangle < 0$ for some $z \in C$, then

$$\lim_{\alpha \downarrow 0} \frac{f(x^* + \alpha(z - x^*)) - f(x^*)}{\alpha} = \langle \nabla f(x^*), (z - x^*) \rangle < 0.$$

Then for sufficiently small α, we have $f(x^* + \alpha(z - x^*)) - f(x^*) < 0$, contradicting the optimality of x^*. □

Examples (a) Let's consider the following linear constrained problem.

$$\min_{x \in \mathbb{R}^n} f(x) \text{ subject to } Ax = b$$

where A is a $m \times n$ matrix and $b \in \mathbb{R}^m$.
Suppose we have a solution x^*, then

$$\langle \nabla f(x^*), y - x^* \rangle \geq 0, \ \forall y \text{ such that } Ay = b$$

This is the same as

$$\langle \nabla f(x^*), h \rangle \geq 0, \ \forall h \in N(A).$$

Since $-h \in N(A)$ if $h \in N(A)$, we have

$$\langle \nabla f(x^*), h \rangle = 0, \ \forall h \in N(A).$$

Hence $\nabla f(x^*) \in N(A)^\perp = R(A^T)$.
So there exists $\mu \in \mathbb{R}^m$ such

$$\nabla f(x^*) + A^T \mu = 0.$$

To conclude, x^* is a solution to the minimization problem if and only if

1. $Ax^* = b$
2. There exists $\mu^* \in \mathbb{R}^m$ such that $\nabla f(x^*) + A^T \mu = 0$.

(b) Let's consider the minimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \text{ subject to } x \geq 0.$$

Suppose we have a solution x^*, then

$$\langle \nabla f(x^*), y - x^* \rangle \geq 0, \ \forall y \in \mathbb{R}^n_+.$$

In particular, $0, 2x^* \in \mathbb{R}^n_+$, so

$$\langle \nabla f(x^*), x^* \rangle = 0, \ \langle \nabla f(x^*), y \rangle \geq 0, \ \forall y \in \mathbb{R}^n_+.$$

Hence, $\nabla f(x^*) \geq 0$. This is the same as saying there exists $\lambda^* \geq 0$ such that

$$\nabla f(x^*) - \lambda^* = 0$$

To conclude, x^* is a solution if and only if

1. $x^* \geq 0$
2. There exists $\lambda^* \geq 0$ such that $\nabla f(x^*) - \lambda^* = 0$
3. $\lambda_i^* x_i^* = 0$