Proof:

Since $\hat{\boldsymbol{x}} \notin \text{int}(C)$, there is a sequence $\{\boldsymbol{x}_k\}$ which does not belong to the closure of C, $\bar{C}$, and converges to $\hat{\boldsymbol{x}}$. Now, denote by $p(\boldsymbol{x}_k)$ the orthogonal projection of $\boldsymbol{x}_k$ into $\bar{C}$ by a standard norm. One can see that by the convexity of $\bar{C}$ [Bertsekas]

$$(p(\boldsymbol{x}_k) - \boldsymbol{x}_k)^T(\boldsymbol{x} - p(\boldsymbol{x}_k)) \geq 0, \quad \forall \boldsymbol{x} \in \bar{C}.$$

Hence,

$$(p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T\boldsymbol{x} \geq (p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T p(\boldsymbol{x}_k) = (p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T(p(\boldsymbol{x}_k)-\boldsymbol{x}_k)+(p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T\boldsymbol{x}_k \geq (p(\boldsymbol{x}_k)-\boldsymbol{x}_k)^T\boldsymbol{x}_k.$$

Now, since $\boldsymbol{x}_k \notin \bar{C}$, calling $\boldsymbol{d}_k = \frac{p(\boldsymbol{x}_k)-\boldsymbol{x}_k}{\|p(\boldsymbol{x}_k)-\boldsymbol{x}_k\|}$,

$$\boldsymbol{d}_k^T\boldsymbol{x} \geq \boldsymbol{d}_k^T\boldsymbol{x}_k, \quad \forall \boldsymbol{x} \in \bar{C}.$$

Since $\|\boldsymbol{d}_k\| = 1$, it has a converging subsequence which will converge to let us say $\boldsymbol{d}$. Taking the same indices for this subsequence for $\boldsymbol{x}_k$, we have the desired result. ∎

Theorem 2.2 (Separation Theorem for Convex Sets) Let C_1 and C_2 nonempty non-intersecting convex subsets of $\mathbb{R}^n$. Then, $\exists \boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d} \neq 0$ such that

$$\sup_{\boldsymbol{x}_1 \in C_1} \boldsymbol{d}^T\boldsymbol{x}_1 \leq \inf_{\boldsymbol{x}_2 \in C_2} \boldsymbol{d}^T\boldsymbol{x}_2.$$

Proof:

Consider the set

$$C := \{\boldsymbol{x}_2 - \boldsymbol{x}_1 \in \mathbb{R}^n \mid \boldsymbol{x}_2 \in C_2, \quad \boldsymbol{x}_1 \in C_1\}$$

which is convex by Propositions 1.10 and 1.11.

Since C_1 and C_2 are disjoint, the origin $\boldsymbol{0}$ does not belong to the interior of C. From Proposition 2.1, there is $\boldsymbol{d} \neq \boldsymbol{0}$ such that $\boldsymbol{d}^T\boldsymbol{x} \geq \boldsymbol{0}$, $\forall \boldsymbol{x} \in C$. Therefore

$$\boldsymbol{d}^T\boldsymbol{x}_1 \leq \boldsymbol{d}^T\boldsymbol{x}_2, \quad \forall \boldsymbol{x}_1 \in C_1 \text{ and } \boldsymbol{x}_2 \in C_2.$$

Finally, since both C_1 and C_2 are nonempty, it follows the result. ∎

Remark 2.3 The Separation Theorem for Convex Sets is an essential result to show the strong duality theorem in convex optimization problems (see for example [Bertsekas]).

3 Lipschitz Continuous Differentiable Functions

Definition 3.1 Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{s} \in \mathbb{R}^n$ be a direction (vector) in $\mathbb{R}^n$. Then the one-sided *directional derivative of a function* $f : \mathbb{R}^n \to \mathbb{R}$ *in the direction* $\boldsymbol{s}$ is defined as

$$f'(\boldsymbol{x}; \boldsymbol{s}) := \lim_{\alpha \downarrow 0} \frac{f(\boldsymbol{x} + \alpha\boldsymbol{s}) - f(x)}{\alpha}.$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function on $\mathbb{R}^n$. Then for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + o(\|\boldsymbol{y} - \boldsymbol{x}\|_2),$$

where $o(r)$ is some function of $r > 0$ such that

$$\lim_{r \to 0} \frac{1}{r} o(r) = 0, \ o(0) = 0.$$

We say that the function is continuously differentiable if the function $\boldsymbol{\nabla f} : \mathbb{R}^n \to \mathbb{R}^n$ is continuous.

Hereafter, we define for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, the standard inner product $\langle \boldsymbol{a}, \boldsymbol{b} \rangle := \sum_{i=1}^n a_i b_i$, and the associated norm $\|\boldsymbol{a}\|_2 := \sqrt{\langle \boldsymbol{a}, \boldsymbol{a} \rangle}$ to it.

Definition 3.2 Let Q be a subset of $\mathbb{R}^n$. We denote by $\mathcal{C}_L^{k,p}(Q)$ the class of functions with the following properties:

- Any $f \in \mathcal{C}_L^{k,p}(Q)$ is k times continuously differentiable on Q;
- Its pth derivative is Lipschitz continuous on Q with the constant $L \geq 0$:

$$\|\boldsymbol{f}^{(p)}(\boldsymbol{x}) - \boldsymbol{f}^{(p)}(\boldsymbol{y})\|_2 \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in Q.$$

In particular, $\boldsymbol{f}^{(1)}(\boldsymbol{x}) = \boldsymbol{\nabla f}(\boldsymbol{x})$ and $\boldsymbol{f}^{(2)}(\boldsymbol{x}) = \boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x})$. Observe that if $f_1 \in \mathcal{C}_{L_1}^{k,p}(Q)$, $f_2 \in \mathcal{C}_{L_2}^{k,p}(Q)$, and $\alpha, \beta \in \mathbb{R}$, then for $L_3 = |\alpha|L_1 + |\beta|L_2$ we have $\alpha f_1 + \beta f_2 \in \mathcal{C}_{L_3}^{k,p}(Q)$.

Lemma 3.3 Let $f \in \mathcal{C}^2(\mathbb{R}^n)$. Then $f \in \mathcal{C}_L^{2,1}(\mathbb{R}^n)$ if and only if $\|\boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x})\|_2 \leq L, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$.

Proof:
For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$\begin{aligned} \boldsymbol{\nabla f}(\boldsymbol{y}) &= \boldsymbol{\nabla f}(\boldsymbol{x}) + \int_0^1 \boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x})d\tau \\ &= \boldsymbol{\nabla f}(\boldsymbol{x}) + \left(\int_0^1 \boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x}))d\tau\right)(\boldsymbol{y} - \boldsymbol{x}). \end{aligned}$$

Since $\|\boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x})\|_2 \leq L$,

$$\begin{aligned} \|\boldsymbol{\nabla f}(\boldsymbol{y}) - \boldsymbol{\nabla f}(\boldsymbol{x})\|_2 &\leq \left\|\int_0^1 \boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x}))d\tau\right\|_2 \|\boldsymbol{y} - \boldsymbol{x}\|_2 \\ &\leq \int_0^1 \|\boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x}))\|_2 d\tau \|\boldsymbol{y} - \boldsymbol{x}\|_2 \\ &\leq L\|\boldsymbol{y} - \boldsymbol{x}\|_2. \end{aligned}$$

On the other hand, for $\boldsymbol{s} \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$, $\alpha \neq 0$,

$$\|\boldsymbol{\nabla f}(\boldsymbol{x} + \alpha\boldsymbol{s}) - \boldsymbol{\nabla f}(\boldsymbol{x})\|_2 \leq |\alpha| L \|\boldsymbol{s}\|_2.$$

Dividing both sides by $|\alpha|$ and taking the limit to zero,

$$\|\boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x})\boldsymbol{s}\|_2 \leq L\|\boldsymbol{s}\|_2, \quad \boldsymbol{s} \in \mathbb{R}^n.$$

Therefore, $\|\boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x})\|_2 \leq L$. ∎

Example 3.4

1. The linear function $f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a}, \boldsymbol{x} \rangle \in \mathcal{C}_0^{2,1}(\mathbb{R}^n)$ since

$$\boldsymbol{\nabla f}(\boldsymbol{x}) = \boldsymbol{a}, \quad \boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{O}.$$

2. The quadratic function $f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a}, \boldsymbol{x} \rangle + 1/2\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \rangle$ with $\boldsymbol{A} = \boldsymbol{A}^T$ belongs to $\mathcal{C}_L^{2,1}(\mathbb{R}^n)$ where

$$\boldsymbol{\nabla f}(\boldsymbol{x}) = \boldsymbol{a} + \boldsymbol{A}\boldsymbol{x}, \quad \boldsymbol{\nabla}^2\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}, \quad L = \|\boldsymbol{A}\|_2.$$

3. The function $f(x) = \sqrt{1 + x^2} \in \mathcal{C}_1^{2,1}(\mathbb{R})$ since

$$\nabla f(x) = \frac{x}{\sqrt{1 + x^2}}, \quad \nabla^2 f(x) = \frac{1}{(1 + x^2)^{3/2}} \leq 1.$$

Lemma 3.5 Let $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$. Then for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle| \leq \frac{L}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2.$$

Proof:

For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\begin{aligned} f(\boldsymbol{y}) &= f(\boldsymbol{x}) + \int_0^1 \langle \boldsymbol{\nabla f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})), \boldsymbol{y} - \boldsymbol{x} \rangle d\tau \\ &= f(\boldsymbol{x}) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \int_0^1 \langle \boldsymbol{\nabla f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle d\tau. \end{aligned}$$

Therefore,

$$\begin{aligned} |f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle| &= \left| \int_0^1 \langle \boldsymbol{\nabla f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle d\tau \right| \\ &\leq \int_0^1 |\langle \boldsymbol{\nabla f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle| d\tau \\ &\leq \int_0^1 \|\boldsymbol{\nabla f}(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - \boldsymbol{\nabla f}(\boldsymbol{x})\|_2 \|\boldsymbol{y} - \boldsymbol{x}\|_2 d\tau \\ &\leq \int_0^1 \tau L \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 d\tau = \frac{L}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2. \end{aligned}$$

∎

Consider a function $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$. Let us fix $\boldsymbol{x}_0 \in \mathbb{R}^n$, and define two quadratic functions:

$$\begin{aligned} \phi_1(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0 \rangle - \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2, \\ \phi_2(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0 \rangle + \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2. \end{aligned}$$

Then the graph of the function f is located between the graphs of ϕ_1 and ϕ_2:

$$\phi_1(\boldsymbol{x}) \leq f(\boldsymbol{x}) \leq \phi_2(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^n.$$

Lemma 3.6 Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$. Then for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\|\boldsymbol{\nabla f}(\boldsymbol{y}) - \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})\|_2 \leq \frac{M}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2,$$

$$|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle - \frac{1}{2} \langle \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle| \leq \frac{M}{6} \|\boldsymbol{y} - \boldsymbol{x}\|_2^3.$$

Lemma 3.7 Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, with $\|\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{y})\|_2 \leq M \|\boldsymbol{x} - \boldsymbol{y}\|_2$. Then

$$\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}) - M \|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I} \preceq \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{y}) \preceq \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}) + M \|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I}.$$

Proof:

Since $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, $\|\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x})\|_2 \leq M \|\boldsymbol{y} - \boldsymbol{x}\|_2$. This means that the eigenvalues of the symmetric matrix $\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x})$ satisfy:

$$|\lambda_i(\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}))| \leq M \|\boldsymbol{y} - \boldsymbol{x}\|_2, \quad i = 1, 2, \ldots, n.$$

Therefore,

$$-M \|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I} \preceq \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}) \preceq M \|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I}.$$

∎