Only in this part, we adopt the following rule:

$$0 \cdot \infty = \infty \cdot 0 = 0 \cdot (-\infty) = (-\infty) \cdot 0 = 0. \tag{10}$$

Definition 5.27 A function $f : \mathbb{R}^n \to [-\infty, +\infty]$ is called *lower semicontinuous at* $\boldsymbol{x} \in \mathbb{R}$ if

$$f(\boldsymbol{x}) \leq \lim_{n\to\infty} \inf f(\boldsymbol{x}_n)$$

for any sequence $\{\boldsymbol{x}_n\}_{n=1}^{\infty}$ for which $\boldsymbol{x}_n \to \boldsymbol{x}$. Therefore, a function $f : \mathbb{R}^n \to [-\infty, +\infty]$ is called *lower semicontinous* if it is lower semicontinuous at each point of $\mathbb{R}^n$.

Theorem 5.28 Let $f : \mathbb{R}^n \to [-\infty, +\infty]$. Then the following conditions are equivalent:

1. f is lower semicontinuous.
2. f is closed.
3. For any $\lambda \in \mathbb{R}$, the λ-level sets L_λ of f are closed.

Proof:
$\boxed{1 \Rightarrow 2}$ We need to show that the epigraph E of f is closed. Let $\{\boldsymbol{x}_n, y_n\}_{n=1}^{\infty}$ a sequence of $\mathbb{R}^n \times \mathbb{R}$ such that $(\boldsymbol{x}_n, y_n) \in E$ for any $n \geq 1$ and $(\boldsymbol{x}_n, y_n) \to (\bar{\boldsymbol{x}}, \bar{y})$. Then $f(\boldsymbol{x}_n) \leq y_n$ and taking the lim infimum on both sides of the inequality:

$$f(\bar{\boldsymbol{x}}) \leq \lim_{n\to\infty} \inf f(\boldsymbol{x}_n) \leq \lim_{n\to\infty} \inf y_n = \bar{y},$$

which shows that $(\bar{\boldsymbol{x}}, \bar{y}) \in E$.
$\boxed{2 \Rightarrow 3}$ If $L_\lambda = \emptyset$, there is nothing to do. Therefore, suppose that theres is sequence $\{\boldsymbol{x}_n\}_{n=1}^{\infty} \subseteq L_\lambda$ that converges to $\bar{\boldsymbol{x}}$. That is $f(\boldsymbol{x}_n) \leq \lambda$ and since the epigraph of f is closed, $(\bar{\boldsymbol{x}}, \lambda) \in E$, and therefore, $f(\bar{\boldsymbol{x}}) \leq \lambda$ which implies that $\bar{\boldsymbol{x}} \in L_\lambda$.
$\boxed{3 \Rightarrow 1}$ Suppose to the contrary that f is not lower semicontinuous. That is, there exists $\bar{\boldsymbol{x}}$, a sequence $\{\boldsymbol{x}_n\}_{n=1}^{\infty}$ such that $\boldsymbol{x}_n \to \bar{\boldsymbol{x}}$, and $\lim_{n\to\infty} \inf f(\boldsymbol{x}_n) < f(\bar{\boldsymbol{x}})$. Consider $\lambda \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \inf f(\boldsymbol{x}_n) < \lambda < f(\bar{\boldsymbol{x}}). \tag{11}$$

Therefore, we can consider a subsequence $\{\boldsymbol{x}_{n_k}\}_{k=1}^{\infty}$ such that $f(\boldsymbol{x}_{n_k}) < \lambda$ and then $\boldsymbol{x}_{n_k} \in L_\lambda$. Since the λ-level sets are closed, $\boldsymbol{x}_{n_k} \to \bar{\boldsymbol{x}} \in L_\lambda$ and $f(\bar{\boldsymbol{x}}) \leq \lambda$ contradicting (11). ∎

The definition of convex function for usual functions (Definition 5.1) is valid for extended real-valued functions using the rule (10).

Theorem 5.29

1. Let $f_i : \mathbb{R}^n \to [-\infty, +\infty]$ $(i \in I)$ be a family of (finite or infinite) extended real-valued functions which are closed and convex. Then the function $f(\boldsymbol{x}) := \sup_{i\in I} f_i(\boldsymbol{x})$ is also closed and convex.
2. Let $f_i : \mathbb{R}^n \to [-\infty, +\infty]$ $(1 \leq i \leq m)$ be a family of finite extended real-valued functions which are closed and convex, and $\alpha_1, \alpha_2, \ldots, \alpha_m \geq 0$. Then the function $f(\boldsymbol{x}) = \sum_{i=1}^{m} \alpha_i f_i(\boldsymbol{x})$ is also closed and convex.
3. If $f : \mathbb{R}^m \to [-\infty, +\infty]$ is an extended real-valued function which is closed and convex, $\boldsymbol{b} \in \mathbb{R}^m$, and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, then $\phi(\boldsymbol{x}) := f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ is also a closed and convex function.

5.6 Exercises

1. Given a convex set $S \subseteq \mathbb{R}^n$ and an arbitrarily norm $\|\cdot\|$ in $\mathbb{R}^n$, define the distance of a point $\boldsymbol{x} \in \mathbb{R}^n$ to the set S as
$$\operatorname{dist}(\boldsymbol{x}, S) := \inf_{\boldsymbol{y} \in S} \|\boldsymbol{x} - \boldsymbol{y}\|.$$
Show that the distance function $\operatorname{dist}(\boldsymbol{x}, S)$ is convex on $\boldsymbol{x}$.

2. Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ and a nonempty set $C \subseteq \mathbb{R}$ illustrating each of the following facts:

    (a) f is non convex on $\mathbb{R}$, C is convex, and f is convex on C.

    (b) f is non convex on $\mathbb{R}$, C is non convex, and f is convex on C.

3. Prove Theorem 5.5.
4. Prove Theorem 5.7.
5. Prove Theorem 5.8.
6. Prove Lemma 5.9.
7. Prove Corollary 5.12.
8. Prove Corollary 5.17.
9. Prove Theorem 5.18.
10. Prove Theorem 5.21.
11. Prove Corollary 5.22.
12. Prove Theorem 5.29.

6 Worse Case Analysis for Gradient Based Methods

6.1 Lower Complexity Bound for the class $\mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$

Gradient Based Method: Iterative method $\mathcal{M}$ generated by a sequence such that
$$\boldsymbol{x}_k \in \boldsymbol{x}_0 + \operatorname{span}\{\boldsymbol{\nabla f}(\boldsymbol{x}_0), \boldsymbol{\nabla f}(\boldsymbol{x}_1), \dots, \boldsymbol{\nabla f}(\boldsymbol{x}_{k-1})\}, \quad k \geq 1.$$

Consider the problem class as follows

Model:	$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$ $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$
Oracle:	Only function and gradient values are available
Approximate solution:	Find $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ such that $f(\bar{\boldsymbol{x}}) - f(\boldsymbol{x}^*) < \varepsilon$

Theorem 6.1 For any $1 \leq k \leq \frac{n-1}{2}$, and any $\boldsymbol{x}_0 \in \mathbb{R}^n$, there exists a function $f \in \mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$ such that for any gradient based method of type $\mathcal{M}$, we have
$$\begin{aligned} f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) &\geq \frac{3L\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2}{32(k+1)^2}, \\ \|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 &\geq \frac{1}{8}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2, \end{aligned}$$

where $\boldsymbol{x}^*$ is the minimum of $f(\boldsymbol{x})$.

Proof:

This type of methods are invariant with respect to a simultaneous shift of all objects in the space of variables. Therefore, we can assume that $\boldsymbol{x}_0 = \boldsymbol{0}$.

Consider the family of quadratic functions

$$f_k(\boldsymbol{x}) = \frac{L}{4}\left\{\frac{1}{2}\left[[\boldsymbol{x}]_1^2 + \sum_{i=1}^{k-1}([\boldsymbol{x}]_i - [\boldsymbol{x}]_{i+1})^2 + [\boldsymbol{x}]_k^2\right] - [\boldsymbol{x}]_1\right\}, \quad k = 1, 2, \ldots, n.$$

We can see that

for $k = 1$, $\quad f_1(\boldsymbol{x}) = \frac{L}{4}([\boldsymbol{x}]_1^2 - [\boldsymbol{x}]_1)$,

for $k = 2$, $\quad f_2(\boldsymbol{x}) = \frac{L}{4}([\boldsymbol{x}]_1^2 + [\boldsymbol{x}]_2^2 - [\boldsymbol{x}]_1[\boldsymbol{x}]_2 - [\boldsymbol{x}]_1)$,

for $k = 3$, $\quad f_3(\boldsymbol{x}) = \frac{L}{4}([\boldsymbol{x}]_1^2 + [\boldsymbol{x}]_2^2 + [\boldsymbol{x}]_3^2 - [\boldsymbol{x}]_1[\boldsymbol{x}]_2 - [\boldsymbol{x}]_2[\boldsymbol{x}]_3 - [\boldsymbol{x}]_1)$.

Therefore, $f_k(\boldsymbol{x}) = \frac{L}{4}\left[\frac{1}{2}\langle \boldsymbol{A}_k\boldsymbol{x}, \boldsymbol{x}\rangle - \langle \boldsymbol{e}_1, \boldsymbol{x}\rangle\right]$, where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)^T$, and

$$\boldsymbol{A}_k = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & \\ -1 & 2 & -1 & \cdots & 0 & \\ 0 & -1 & 2 & \ddots & 0 & \boldsymbol{0}_{k,n-k} \\ \vdots & \ddots & \ddots & \ddots & -1 & \\ 0 & \cdots & 0 & -1 & 2 & \\ & & \boldsymbol{0}_{n-k,k} & & & \boldsymbol{0}_{n-k,n-k} \end{pmatrix}.$$

Also, $\boldsymbol{\nabla f}_k(\boldsymbol{x}) = \frac{L}{4}(\boldsymbol{A}_k\boldsymbol{x} - \boldsymbol{e}_1)$ and $\boldsymbol{\nabla}^2\boldsymbol{f}_k(\boldsymbol{x}) = \frac{L}{4}\boldsymbol{A}_k$. After some calculations, we can show that $L\boldsymbol{I} \succeq \boldsymbol{\nabla}^2\boldsymbol{f}_k(\boldsymbol{x}) \succeq \boldsymbol{O}$ for $k = 1, 2, \ldots, n$, and therefore, $f_k(\boldsymbol{x}) \in \mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$, for $k = 1, 2, \ldots, n$, due to Corollary 5.12.

Then

$$\begin{aligned} f_k(\overline{\boldsymbol{x}_k}) &= \frac{L}{8}\left(-1 + \frac{1}{k+1}\right), \\ [\overline{\boldsymbol{x}_k}]_i &= \begin{cases} 1 - \frac{i}{k+1}, & i = 1, 2, \ldots, k \\ 0, & i = k+1, k+2, \ldots, n, \end{cases} \end{aligned}$$

are the minimum value and the minimal solution for $f_k(\cdot)$, respectively.

Now, for $1 \le k \le \frac{n-1}{2}$, let us define $f(\boldsymbol{x}) := f_{2k+1}(\boldsymbol{x})$, and therefore $\boldsymbol{x}^* := \overline{\boldsymbol{x}_{2k+1}}$.

Note that $\boldsymbol{x}_k \in \boldsymbol{x}_0 + \operatorname{span}\{\boldsymbol{\nabla f}(\boldsymbol{x}_0), \boldsymbol{\nabla f}(\boldsymbol{x}_1), \ldots, \boldsymbol{\nabla f}(\boldsymbol{x}_{k-1})\}$ for $\boldsymbol{x}_0 = \boldsymbol{0}$. Moreover, since $\boldsymbol{\nabla f}_k(\boldsymbol{x}) = \frac{L}{4}(\boldsymbol{A}_k\boldsymbol{x} - \boldsymbol{e}_1)$, $[\boldsymbol{x}_k]_p = 0$ for $p > k$. Therefore, $f_p(\boldsymbol{x}_k) = f_k(\boldsymbol{x}_k)$ for $p \ge k$.

Then for $k = 1, 2, \ldots, \lfloor\frac{n-1}{2}\rfloor$,

$$\begin{aligned} f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) &= f_{2k+1}(\boldsymbol{x}_k) - f_{2k+1}(\overline{\boldsymbol{x}_{2k+1}}) = f_k(\boldsymbol{x}_k) - \frac{L}{8}\left(-1 + \frac{1}{2k+2}\right) \\ &\ge f_k(\overline{\boldsymbol{x}_k}) - \frac{L}{8}\left(-1 + \frac{1}{2k+2}\right) = \frac{L}{8}\left(-1 + \frac{1}{k+1}\right) - \frac{L}{8}\left(-1 + \frac{1}{2k+2}\right) \\ &= \frac{L}{16(k+1)}. \end{aligned}$$

We can obtain after some calculations,

$$\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2 \;=\; \|\boldsymbol{x}_0 - \overline{\boldsymbol{x}_{2k+1}}\|_2^2 = \sum_{i=1}^{2k+1}\left(1 - \frac{i}{2k+2}\right)^2$$

$$
\begin{aligned}
&= 2k+1-\frac{2}{2k+2}\sum_{i=1}^{2k+1}\frac{i}{2k+2}+\frac{1}{(2k+2)^2}\sum_{i=1}^{2k+1} i^2 \\
&\leq 2k+1-\frac{2(2k+2)(2k+1)}{(2k+2)2}+\frac{(2k+1+1)^3}{3(2k+2)^2} \\
&\leq \frac{2(k+1)}{3}.
\end{aligned}
$$

Then

$$
\frac{f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)}{\|\boldsymbol{x}_0-\boldsymbol{x}^*\|^2} \geq \frac{L}{16(k+1)}\frac{3}{2(k+1)}.
$$

Also

$$
\begin{aligned}
\|\boldsymbol{x}_k-\boldsymbol{x}^*\|_2^2 &= \|\boldsymbol{x}_k-\overline{\boldsymbol{x}_{2k+1}}\|_2^2 \geq \sum_{i=k+1}^{2k+1}([\overline{\boldsymbol{x}_{2k+1}}]_i)^2 = \sum_{i=k+1}^{2k+1}\left(1-\frac{i}{2k+2}\right)^2 \\
&= k+1-\frac{2}{2k+2}\left[\frac{(2k+2)(2k+1)}{2}-\frac{(k+1)k}{2}\right]-\frac{1}{(2k+2)^2}\sum_{i=k+1}^{2k+1} i^2 \\
&\geq \frac{1}{8}\|\boldsymbol{x}_0-\overline{\boldsymbol{x}_{2k+1}}\|_2^2 = \frac{1}{8}\|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2^2.
\end{aligned}
$$

∎

If we consider very large problems where we can not afford n number of iterations, the above theorem says that:

- The function value can be expected to decrease fast.
- The convergence to the optimal solution $\boldsymbol{x}^*$ can be arbitrarily slow.

6.2 Lower Complexity Bound for the class $\mathcal{S}_{\mu,L}^{\infty,1}(\ell^2)$

Gradient Based Method: Iterative method $\mathcal{M}$ generated by a sequence such that

$$
\boldsymbol{x}_k \in \boldsymbol{x}_0 + \text{span}\{\boldsymbol{\nabla f}(\boldsymbol{x}_0), \boldsymbol{\nabla f}(\boldsymbol{x}_1), \ldots, \boldsymbol{\nabla f}(\boldsymbol{x}_{k-1})\}, \quad k \geq 1.
$$

Let us define

$$
\ell^2 := \left\{ \{x_i\}_{i=1}^{\infty} \;\middle|\; \sum_{i=1}^{\infty} x_i^2 < \infty \right\}.
$$

Consider the problem class as follows

Model:	$\min_{\boldsymbol{x}\in\ell^2} f(\boldsymbol{x})$ $f \in \mathcal{S}_{\mu,L}^{\infty,1}(\ell^2)$
Oracle:	Only function and gradient values are available
Approximate solution:	Find $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ such that $\begin{cases} f(\bar{\boldsymbol{x}})-f(\boldsymbol{x}^*) < \varepsilon \\ \|\bar{\boldsymbol{x}}-\boldsymbol{x}^*\|_2^2 < \varepsilon \end{cases}$