Remark 5.16 Strongly convex functions are different from strictly convex functions. For instance, $f(x) = x^4$ is strictly convex at $x = 0$ but it is not strongly convex at the same point.

Corollary 5.17 If $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$ and $\boldsymbol{\nabla f}(\boldsymbol{x}^*) = 0$, then

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \frac{1}{2}\mu\|\boldsymbol{x} - \boldsymbol{x}^*\|_2^2, \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Proof:
Left for exercise. ▮

Theorem 5.18 Let f be a continuously differentiable function. The following conditions are equivalent:

1. $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$.
2. $\mu\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \leq \langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$
3. $f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) + \alpha(1-\alpha)\frac{\mu}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \leq \alpha f(\boldsymbol{x}) + (1-\alpha)f(\boldsymbol{y}),\ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,\ \forall \alpha \in [0,1].$

Proof:
Left for exercise. ▮

Theorem 5.19 If $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$, we have

1. $f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle + \frac{1}{2\mu}\|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y})\|_2^2,\ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,$
2. $\langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle \leq \frac{1}{\mu}\|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y})\|_2^2,\ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$

Proof:
Let us fix $\boldsymbol{x} \in \mathbb{R}^n$, and define the function $\phi(\boldsymbol{y}) = f(\boldsymbol{y}) - \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y}\rangle$. Clearly, $\phi \in \mathcal{S}^1_\mu(\mathbb{R}^n)$. Also, one minimal solution is $\boldsymbol{x}$. Therefore,

$$\begin{aligned}
\phi(\boldsymbol{x}) &= \min_{\boldsymbol{v} \in \mathbb{R}^n} \phi(\boldsymbol{v}) \geq \min_{\boldsymbol{v} \in \mathbb{R}^n} \left[\phi(\boldsymbol{y}) + \langle \boldsymbol{\nabla \phi}(\boldsymbol{y}), \boldsymbol{v} - \boldsymbol{y}\rangle + \frac{\mu}{2}\|\boldsymbol{v} - \boldsymbol{y}\|_2^2\right] \\
&= \phi(\boldsymbol{y}) - \frac{1}{2\mu}\|\boldsymbol{\nabla \phi}(\boldsymbol{y})\|_2^2
\end{aligned}$$

as wished. Adding two copies of the 1 with $\boldsymbol{x}$ and $\boldsymbol{y}$ interchanged, we get 2. ▮

Remark 5.20 The converse of Theorem 5.19 is not valid. For instance, consider $f(x_1, x_2) = x_1^2 - x_2^2$, $\mu = 1$. Then the inequalities 1. and 2. are satisfied but $f \notin \mathcal{S}^1_\mu(\mathbb{R}^2)$ for any $\mu > 0$.

Theorem 5.21 Let f be a twice continuously differentiable function. Then $f \in \mathcal{S}^2_\mu(\mathbb{R}^n)$ if and only if

$$\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}) \succeq \mu \boldsymbol{I}, \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Proof:
Left for exercise. ▮

Corollary 5.22 Let f be a twice continuously differentiable function. Then $f \in \mathcal{S}^{2,1}_{\mu,L}(\mathbb{R}^n)$ if and only if

$$L\boldsymbol{I} \succeq \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}) \succeq \mu \boldsymbol{I}, \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Proof:
Left for exercise. ▮

Theorem 5.23 If $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, then

$$\frac{\mu L}{\mu + L}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \frac{1}{\mu+L}\|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y})\|_2^2 \leq \langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle, \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

Proof:

If $\mu = L$, from Theorem 5.18 and the definition of $\mathcal{C}^1_\mu(\mathbb{R}^n)$,

$$\begin{aligned} \langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle &\geq \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \\ &\geq \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \frac{1}{2\mu}\|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y})\|_2^2, \end{aligned}$$

and the result follows.

If $\mu < L$, let us define $\phi(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{\mu}{2}\|\boldsymbol{x}\|_2^2$. Then $\boldsymbol{\nabla \phi}(\boldsymbol{x}) = \boldsymbol{\nabla f}(\boldsymbol{x}) - \mu \boldsymbol{x}$ and $\langle \boldsymbol{\nabla \phi}(\boldsymbol{x}) - \boldsymbol{\nabla \phi}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle = \langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle - \mu\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \leq (L-\mu)\|\boldsymbol{x}-\boldsymbol{y}\|_2^2$ since $f \in \mathcal{C}^{1,1}_L(\mathbb{R}^n)$. Also $\langle \boldsymbol{\nabla \phi}(\boldsymbol{x}) - \boldsymbol{\nabla \phi}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle \geq \mu\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 - \mu\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 = 0$ due to Theorem 5.18. Therefore, from Theorem 5.13, $\phi \in \mathcal{F}^{1,1}_{L-\mu}(\mathbb{R}^n)$.

We have now $\langle \boldsymbol{\nabla \phi}(\boldsymbol{x}) - \boldsymbol{\nabla \phi}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle \geq \frac{1}{L-\mu}\|\boldsymbol{\nabla \phi}(\boldsymbol{x}) - \boldsymbol{\nabla \phi}(\boldsymbol{y})\|_2^2$ from Theorem 5.13. Therefore

$$\begin{aligned} \langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle &\geq \mu\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \frac{1}{L-\mu}\|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}) - \mu(\boldsymbol{x}-\boldsymbol{y})\|_2^2 \\ &= \mu\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \frac{1}{L-\mu}\|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y})\|_2^2 - \frac{2\mu}{L-\mu}\langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle \\ &\quad + \frac{\mu^2}{L-\mu}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2, \end{aligned}$$

and the result follows after some simplifications. ∎

5.5 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
$$\mathrm{dist}(\boldsymbol{x}, S) := \inf_{\boldsymbol{y}\in S} \|\boldsymbol{x}-\boldsymbol{y}\|.$$
Show that the distance function $\mathrm{dist}(\boldsymbol{x}, S)$ 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.

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

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

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

$$\boldsymbol{A}_k = \left(\begin{array}{ccccc|c} 2 & -1 & 0 & \cdots & 0 & \\ -1 & 2 & -1 & \cdots & 0 & \\ 0 & -1 & 2 & \ddots & 0 & \boldsymbol{0}_{k,n-k} \\ \vdots & \ddots & \ddots & \ddots & -1 & \\ 0 & \cdots & 0 & -1 & 2 & \\ & & \boldsymbol{0}_{n-k,k} & & & \boldsymbol{0}_{n-k,n-k} \end{array}\right).$$

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

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

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

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

We can obtain after some calculations,

$$
\begin{aligned}
\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2 &= \|\boldsymbol{x}_0 - \overline{\boldsymbol{x}_{2k+1}}\|_2^2 = \sum_{i=1}^{2k+1}\left(1 - \frac{i}{2k+2}\right)^2 \\
&= 2k + 1 - \frac{2}{2k+2}\sum_{i=1}^{2k+1}\frac{i}{2k+2} + \frac{1}{(2k+2)^2}\sum_{i=1}^{2k+1} i^2 \\
&\leq 2k + 1 - \frac{2(2k+2)(2k+1)}{(2k+2)2} + \frac{(2k+1+1)^3}{3(2k+2)^2} \\
&\leq \frac{2(k+1)}{3}.
\end{aligned}
$$

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

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