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

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

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

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

The converse of Theorem 6.18 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 $f \notin \mathcal{S}_\mu^1(\mathbb{R}^2)$ for any $\mu > 0$.

Theorem 6.19 Let f be a twice continuously differentiable function. Then $f \in \mathcal{S}_\mu^2(\mathbb{R}^n)$ if and only if

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

Proof:
Left for exercise. ∎

Corollary 6.20 Let f be a twice continuously differentiable function. Then $f \in \mathcal{S}_{\mu,L}^{2,1}(\mathbb{R}^n)$ if and only if

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

Theorem 6.21 If $f \in \mathcal{S}_{\mu,L}^{1,1}(\mathbb{R}^n)$, then

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

Proof:
If $\mu = L$, from Theorem 6.17 and the definition of $\mathcal{F}_\mu^1(\mathbb{R}^n)$,

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

and the result follows.

If $\mu < L$, let us define $\phi(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{\mu}{2}\|\boldsymbol{x}\|_2^2$. Then $\phi'(\boldsymbol{x}) = f'(\boldsymbol{x}) - \mu\boldsymbol{x}$ and $\langle \phi'(\boldsymbol{x}) - \phi'(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle = \langle f'(\boldsymbol{x}) - f'(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle - \mu\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \leq (L - \mu)\|\boldsymbol{x} - \boldsymbol{y}\|_2^2$ since $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$. Also $\langle \phi'(\boldsymbol{x}) - \phi'(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle \geq \mu\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - \mu\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 = 0$ due to Theorem 6.17. Therefore, from Theorem 6.13, $\phi \in \mathcal{F}_{L-\mu}^{1,1}(\mathbb{R}^n)$.

We have now $\langle \phi'(\boldsymbol{x}) - \phi'(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle \geq \frac{1}{L-\mu}\|\phi'(\boldsymbol{x}) - \phi'(\boldsymbol{y})\|_2^2$ from Theorem 6.13. Therefore

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

and the result follows after some simplifications. ∎

6.5 Exercises

1. Prove Theorem 6.3.
2. Prove Theorem 6.7.
3. Prove Lemma 6.8.
4. Prove Theorem 6.10.
5. Prove Corollary 6.16.
6. Prove Theorem 6.17.
7. Prove Theorem 6.19.

7 Worse Case Analysis for Gradient Based Methods

7.1 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

$$\boldsymbol{x}_k \in \boldsymbol{x}_0 + \mathrm{Lin}\{f'(\boldsymbol{x}_0), f'(\boldsymbol{x}_1), \ldots, f'(\boldsymbol{x}_{k-1})\}, \quad k \geq 1.$$

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 7.1 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

$$\begin{aligned} f(\boldsymbol{x}_k) - f^* &\geq \frac{3L\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2}{32(k+1)^2}, \\ \|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 &\geq \frac{1}{8}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2, \end{aligned}$$

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

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

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

We can see that
for $k=1$, $f_1(\boldsymbol{x}) = \frac{L}{4}([\boldsymbol{x}]_1^2 - [\boldsymbol{x}]_1)$,
for $k=2$, $f_2(\boldsymbol{x}) = \frac{L}{4}([\boldsymbol{x}]_1^2 + [\boldsymbol{x}]_2^2 - [\boldsymbol{x}]_1[\boldsymbol{x}]_2 - [\boldsymbol{x}]_1)$,
for $k=3$, $f_3(\boldsymbol{x}) = \frac{L}{4}([\boldsymbol{x}]_1^2 + [\boldsymbol{x}]_2^2 + [\boldsymbol{x}]_3^2 - [\boldsymbol{x}]_1[\boldsymbol{x}]_2 - [\boldsymbol{x}]_2[\boldsymbol{x}]_3 - [\boldsymbol{x}]_1)$.

Also, $f_k'(\boldsymbol{x}) = \frac{L}{4}(\boldsymbol{A}_k\boldsymbol{x} - \boldsymbol{e}_1)$, where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)^T$, and

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

After some calculations, we can show that $L\boldsymbol{I} \succeq f_k''(\boldsymbol{x}) \succeq \boldsymbol{O}, \quad k = 1, 2, \ldots, \lfloor \frac{n-1}{2} \rfloor$, and therefore, $f_k(\boldsymbol{x}) \in \mathcal{F}_L^{\infty,1}(\mathbb{R}^n)$, for $k = 1, 2, \ldots, \lfloor \frac{n-1}{2} \rfloor$.

Also

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

Let us take $f(\boldsymbol{x}) := f_{2k+1}(\boldsymbol{x})$, and $\boldsymbol{x}^* := \overline{\boldsymbol{x}_{2k+1}}$.

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

Then

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

After some calculations [Nesterov03], we obtain

$$\|\boldsymbol{x}_0 - \overline{\boldsymbol{x}_{2k+1}}\|_2 \leq \frac{2(k+1)}{3}.$$

Also $\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 = \|\boldsymbol{x}_k - \overline{\boldsymbol{x}_{2k+1}}\|_2^2 \geq \sum_{i=k+1}^{2k+1} ([\overline{\boldsymbol{x}_{2k+1}}]_i)^2$.

And then, with more calculations [Nesterov03], we have the results. ∎

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

- The optimal value can be expected to decrease fast.
- The convergence to the optimal solution can be arbitrarily slow.

7.2 Lower Complexity Bound for the class $\mathcal{S}_{\mu,L}^{\infty,1}(\mathbb{R}^\infty)$

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

$$\boldsymbol{x}_k \in \boldsymbol{x}_0 + \mathrm{Lin}\{f'(\boldsymbol{x}_0), f'(\boldsymbol{x}_1), \ldots, f'(\boldsymbol{x}_{k-1})\}, \quad k \geq 1.$$

Consider the problem class as follows