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 $\mathcal{C}_\mu^1(\mathbb{R}^n)$,

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

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 $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$. 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. 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 in $\boldsymbol{x}$.

2. Prove Theorem 6.4.

3. Prove Theorem 6.8.

4. Prove Lemma 6.9.

5. Prove Theorem 6.7.

6. Prove Corollary 6.16.

7. Prove Theorem 6.17.

8. 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
$$f_k(\boldsymbol{x}) = \frac{L}{4}\left\{\frac{1}{2}\left[[\boldsymbol{x}]_1^2 + \sum_{i=1}^{k-1}([\boldsymbol{x}]_i - [\boldsymbol{x}]_{i+1})^2 + [\boldsymbol{x}]_k^2\right] - [\boldsymbol{x}]_1\right\}, \quad k = 1, 2, \ldots, \left\lfloor \frac{n-1}{2} \right\rfloor.$$