5. $0 \le \langle f'(\boldsymbol{x}) - f'(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \le L\|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.

6. $f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) + \frac{\alpha(1-\alpha)}{2L}\|f'(\boldsymbol{x}) - f'(\boldsymbol{y})\|_2^2 \le \alpha f(\boldsymbol{x}) + (1-\alpha)f(\boldsymbol{y})$.

7. $0 \le \alpha f(\boldsymbol{x}) + (1-\alpha)f(\boldsymbol{y}) - f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le \alpha(1-\alpha)\frac{L}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.

Proof:

$\boxed{1 \Rightarrow 2}$ It follows from the definition of convex function and Lemma 3.4.

$\boxed{2 \Rightarrow 3}$ Fix $\boldsymbol{x} \in \mathbb{R}^n$, and consider the function $\phi(\boldsymbol{y}) = f(\boldsymbol{y}) - \langle f'(\boldsymbol{x}), \boldsymbol{y} \rangle$. Clearly $\phi(\boldsymbol{y})$ satisfies 2. Also, $\boldsymbol{y}^* = \boldsymbol{x}$ is a minimal solution. Therefore from 2,

$$
\begin{aligned}
\phi(\boldsymbol{x}) &= \phi(\boldsymbol{y}^*) \le \phi\left(\boldsymbol{y} - \frac{1}{L}\phi'(\boldsymbol{y})\right) \le \phi(\boldsymbol{y}) + \frac{L}{2}\left\|\frac{1}{L}\phi'(\boldsymbol{y})\right\|_2^2 + \langle \phi'(\boldsymbol{y}), -\frac{1}{L}\phi'(\boldsymbol{y}) \rangle \\
&= \phi(\boldsymbol{y}) + \frac{1}{2L}\|\phi'(\boldsymbol{y})\|_2^2 - \frac{1}{L}\|\phi'(\boldsymbol{y})\|_2^2 = \phi(\boldsymbol{y}) - \frac{1}{2L}\|\phi'(\boldsymbol{y})\|_2^2.
\end{aligned}
$$

Since $\phi'(\boldsymbol{y}) = f'(\boldsymbol{y}) - f'(\boldsymbol{x})$, finally we have

$$
f(\boldsymbol{x}) - \langle f'(\boldsymbol{x}), \boldsymbol{x} \rangle \le f(\boldsymbol{y}) - \langle f'(\boldsymbol{x}), \boldsymbol{y} \rangle - \frac{1}{2L}\|f'(\boldsymbol{y}) - f'(\boldsymbol{x})\|_2^2.
$$

$\boxed{3 \Rightarrow 4}$ Adding two copies of 3 with $\boldsymbol{x}$ and $\boldsymbol{y}$ interchanged, we obtain 4.

$\boxed{4 \Rightarrow 1}$ Applying the Cauchy-Schwarz inequality to 4, we obtain $\|f'(\boldsymbol{x}) - f'(\boldsymbol{y})\|_2 \le L\|\boldsymbol{x} - \boldsymbol{y}\|_2$. Also from Theorem 6.7, $f(\boldsymbol{x})$ is convex.

$\boxed{2 \Rightarrow 5}$ Adding two copies of 2 with $\boldsymbol{x}$ and $\boldsymbol{y}$ interchanged, we obtain 5.

$\boxed{5 \Rightarrow 2}$

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

The non-negativity follows from Theorem 6.7.

$\boxed{3 \Rightarrow 6}$ Denote $\boldsymbol{x}_\alpha = \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}$. From 3,

$$
\begin{aligned}
f(\boldsymbol{x}) &\ge f(\boldsymbol{x}_\alpha) + \langle f'(\boldsymbol{x}_\alpha), (1-\alpha)(\boldsymbol{x} - \boldsymbol{y}) \rangle + \frac{1}{2L}\|f'(\boldsymbol{x}) - f'(\boldsymbol{x}_\alpha)\|_2^2 \\
f(\boldsymbol{y}) &\ge f(\boldsymbol{x}_\alpha) + \langle f'(\boldsymbol{x}_\alpha), \alpha(\boldsymbol{y} - \boldsymbol{x}) \rangle + \frac{1}{2L}\|f'(\boldsymbol{y}) - f'(\boldsymbol{x}_\alpha)\|_2^2.
\end{aligned}
$$

Multiplying the first inequality by α, the second by $1-\alpha$, and summing up, we have

$$
\alpha f(\boldsymbol{x}) + (1-\alpha)f(\boldsymbol{y}) \ge f(\boldsymbol{x}_\alpha) + \frac{\alpha}{2L}\|f'(\boldsymbol{x}) - f'(\boldsymbol{x}_\alpha)\|_2^2 + \frac{1-\alpha}{2L}\|f'(\boldsymbol{y}) - f'(\boldsymbol{x}_\alpha)\|_2^2.
$$

Finally, using the inequality

$$
\alpha\|\boldsymbol{b} - \boldsymbol{d}\|_2^2 + (1-\alpha)\|\boldsymbol{c} - \boldsymbol{d}\|_2^2 \ge \alpha(1-\alpha)\|\boldsymbol{b} - \boldsymbol{c}\|_2^2
$$

we have the result.

$$
\left(
\begin{array}{l}
-\alpha(1-\alpha)\|\boldsymbol{b} - \boldsymbol{c}\|_2^2 \ge -\alpha(1-\alpha)(\|\boldsymbol{b} - \boldsymbol{d}\|_2 + \|\boldsymbol{c} - \boldsymbol{d}\|)_2^2 \\
\text{Therefore} \\
\alpha\|\boldsymbol{b} - \boldsymbol{d}\|_2^2 + (1-\alpha)\|\boldsymbol{c} - \boldsymbol{d}\|_2^2 - \alpha(1-\alpha)(\|\boldsymbol{b} - \boldsymbol{d}\|_2 + \|\boldsymbol{c} - \boldsymbol{d}\|_2)^2 \\
= (\alpha\|\boldsymbol{b} - \boldsymbol{d}\|_2 - (1-\alpha)\|\boldsymbol{c} - \boldsymbol{d}\|_2)^2 \ge 0
\end{array}
\right)
$$

$\boxed{6\Rightarrow 3}$ Dividing both sides by $1-\alpha$ and tending α to 1, we obtain 3.

$\boxed{2\Rightarrow 7}$ From 2,

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

Multiplying the first inequality by α, the second by $1-\alpha$, and summing up, we have

$$\alpha f(\boldsymbol{x}) + (1-\alpha)f(\boldsymbol{y}) \leq f(\boldsymbol{x}_\alpha) + \frac{L}{2}\left(\alpha(1-\alpha)^2 + (1-\alpha)\alpha^2\right)\|\boldsymbol{x}-\boldsymbol{y}\|_2^2.$$

The non-negativity follows from Theorem 6.7.

$\boxed{7\Rightarrow 2}$ Dividing both sides by $1-\alpha$ and tending α to 1, we obtain 2. The non-negativity follows from Theorem 6.7. ∎

6.4 Differentiable Strongly Convex Functions

Definition 6.14 A continuously differentiable function $f(\boldsymbol{x})$ is called *strongly convex* on $\mathbb{R}^n$ (notation $f \in \mathcal{S}_\mu^1(\mathbb{R}^n)$) if there exists a constant $\mu > 0$ such that

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

The constant μ is called the *convexity parameter* of the function f.

Example 6.15 The following functions are strongly convex functions:

1. $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|_2^2$.
2. $f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a}, \boldsymbol{x}\rangle + \frac{1}{2}\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}\rangle$, for $\boldsymbol{A} \succeq \mu \boldsymbol{I}, \ \mu > 0$.
3. A sum of a convex and a strongly convex functions.

Corollary 6.16 If $f \in \mathcal{S}_\mu^1(\mathbb{R}^n)$ and $f'(\boldsymbol{x}^*) = 0$, 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 6.17 Let f be a continuously differentiable function. The following conditions are equivalent:

1. $f \in \mathcal{S}_\mu^1(\mathbb{R}^n)$.
2. $\mu\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \leq \langle f'(\boldsymbol{x}) - f'(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.
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 6.18 If $f \in \mathcal{S}_\mu^1(\mathbb{R}^n)$, we have

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