Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $p > 1$, and q such that $\frac{1}{p} + \frac{1}{q} = 1$. Consider

$$a = \frac{|[\boldsymbol{x}]_i|^p}{\sum_{j=1}^n |[\boldsymbol{x}]_j|^p}, \; b = \frac{|[\boldsymbol{y}]_i|^q}{\sum_{j=1}^n |[\boldsymbol{y}]_j|^q}, \; \alpha = \frac{1}{p}, \text{ and } (1-\alpha) = \frac{1}{q}.$$

Then we have

$$\left(\frac{|[\boldsymbol{x}]_i|^p}{\sum_{j=1}^n |[\boldsymbol{x}]_j|^p} \right)^{\frac{1}{p}} \left(\frac{|[\boldsymbol{y}]_i|^q}{\sum_{j=1}^n |[\boldsymbol{y}]_j|^q} \right)^{\frac{1}{q}} \leq \frac{|[\boldsymbol{x}]_i|^p}{p \sum_{j=1}^n [\boldsymbol{x}]_j^p} + \frac{|[\boldsymbol{y}]_i|^q}{q \sum_{j=1}^n |[\boldsymbol{y}]_j|^q}.$$

and summing over i, we obtain the Hölder inequality:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q$$

where $\|\boldsymbol{x}\|_p := \left(\sum_{i=1}^n |[\boldsymbol{x}]_i|^p \right)^{\frac{1}{p}}$.

Theorem 6.13 Let $\{f_i\}_{i \in I}$ be a family of (finite or infinite) functions which are bounded from above and $f_i \in \mathcal{F}(\mathbb{R}^n)$. Then, $f(\boldsymbol{x}) := \sup_{i \in I} f_i(\boldsymbol{x})$ is convex in $\mathbb{R}^n$.

Proof:

For each $i \in I$, since $f_i \in \mathcal{F}(\mathbb{R}^n)$, its epigraph $E_i = \{(\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid f_i(\boldsymbol{x}) \leq y\}$ is convex in $\mathbb{R}^{n+1}$ by Theorem 6.9. Also their intersection

$$\bigcap_{i \in I} E_i = \bigcap_{i \in I} \left\{ (\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid f_i(\boldsymbol{x}) \leq y \right\} = \left\{ (\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \;\middle|\; \sup_{i \in I} f_i(\boldsymbol{x}) \leq y \right\}$$

is convex by Exercise 1 of Section 1, which is exactly the epigraph of $f(\boldsymbol{x})$. ∎

6.2 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 $\boldsymbol{A} \succeq \mu \boldsymbol{I}$.
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. $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$.
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 $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$, we have

1. $f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \langle f'(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle + \frac{1}{2\mu}\|f'(\boldsymbol{x}) - f'(\boldsymbol{y})\|_2^2,\ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,
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, $\phi \in \mathcal{S}^1_\mu(\mathbb{R}^n)$. Also, one minimal solution is $\boldsymbol{x}$. Therefore,

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

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

Theorem 6.19 Let f be a twice continuously differentiable function. Then $f \in \mathcal{S}^2_\mu(\mathbb{R}^n)$ 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 $f \in \mathcal{S}^{2,1}_{\mu,L}(\mathbb{R}^n)$ 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 $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, then

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