

Chapter 5

Smoothness and Strong Convexity

5.1 L -Smooth Functions

We begin with the definition of L -smoothness.

Definition 5.1 (L -smoothness). Let $L \geq 0$. A function $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is said to be L -smooth over a set $D \subseteq \mathbb{E}$ if it is differentiable over D and satisfies

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \leq L\|\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{x}, \mathbf{y} \in D.$$

The constant L is called the **smoothness parameter**.

Obviously, by the definition of differentiability, if f is L -smooth over a set $D \subseteq \mathbb{E}$, this means in particular that $D \subseteq \text{int}(\text{dom} f)$. If a function is L -smooth over \mathbb{E} , then we will just refer to it as L -smooth (without mentioning the entire space). Another frequent terminology in the literature refers to an L -smooth function over D as “a function with Lipschitz gradient with constant L .” The class of L -smooth functions is denoted by $C_L^{1,1}(D)$. When $D = \mathbb{E}$, the class is often denoted by $C_L^{1,1}$ instead of $C_L^{1,1}(\mathbb{E})$. The class of functions which are L -smooth for some $L \geq 0$ is denoted by $C^{1,1}$.

By the definition of L -smoothness, it is clear that if a function is L_1 -smooth, then it is also L_2 -smooth for any $L_2 \geq L_1$. It is therefore sometimes interesting to discuss the value of the *smallest* possible smoothness parameter of a given function.

Example 5.2 (smoothness of quadratic functions). Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. We assume that \mathbb{R}^n is endowed with the l_p -norm ($1 \leq p \leq \infty$). Then, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q = \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{y}\|_q \leq \|\mathbf{A}\|_{p,q} \|\mathbf{x} - \mathbf{y}\|_p,$$

where $\|\cdot\|_{p,q}$ is the induced norm given by (see also Section 1.8.2)

$$\|\mathbf{A}\|_{p,q} = \max\{\|\mathbf{A}\mathbf{x}\|_q : \|\mathbf{x}\|_p \leq 1\},$$

with $q \in [1, \infty]$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$. We can thus conclude that f is $\|\mathbf{A}\|_{p,q}$ -smooth. We will show that $\|\mathbf{A}\|_{p,q}$ is the smallest smoothness parameter. For that, assume that f is L -smooth. Take a vector $\tilde{\mathbf{x}}$ satisfying $\|\tilde{\mathbf{x}}\|_p = 1$ and $\|\mathbf{A}\tilde{\mathbf{x}}\|_q = \|\mathbf{A}\|_{p,q}$. The existence of such a vector is guaranteed by the definition of the induced matrix norm. Then

$$\|\mathbf{A}\|_{p,q} = \|\mathbf{A}\tilde{\mathbf{x}}\|_q = \|\nabla f(\tilde{\mathbf{x}}) - \nabla f(\mathbf{0})\|_q \leq L\|\tilde{\mathbf{x}} - \mathbf{0}\|_p = L.$$

We thus showed that if f is L -smooth, then $L \geq \|\mathbf{A}\|_{p,q}$, proving that $\|\mathbf{A}\|_{p,q}$ is indeed the smallest possible smoothness parameter. ■

Example 5.3 (0-smoothness of affine functions). Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \langle \mathbf{b}, \mathbf{x} \rangle + c$, where $\mathbf{b} \in \mathbb{E}^*$ and $c \in \mathbb{R}$. For any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* = \|\mathbf{b} - \mathbf{b}\|_* = 0 \leq 0\|\mathbf{x} - \mathbf{y}\|,$$

showing that affine functions are 0-smooth. ■

The next example will utilize a well-known result on the orthogonal projection operator, which was introduced in Example 3.31. A more general result will be shown later on in Theorem 6.42.

Theorem 5.4 (see [10, Theorem 9.9]). Let \mathbb{E} be a Euclidean space, and let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Then

(a) **(firm nonexpansiveness)** For any $\mathbf{v}, \mathbf{w} \in \mathbb{E}$,

$$\langle P_C(\mathbf{v}) - P_C(\mathbf{w}), \mathbf{v} - \mathbf{w} \rangle \geq \|P_C(\mathbf{v}) - P_C(\mathbf{w})\|^2. \quad (5.1)$$

(b) **(nonexpansiveness)** For any $\mathbf{v}, \mathbf{w} \in \mathbb{E}$,

$$\|P_C(\mathbf{v}) - P_C(\mathbf{w})\| \leq \|\mathbf{v} - \mathbf{w}\|. \quad (5.2)$$

Example 5.5 (1-smoothness of $\frac{1}{2}d_C^2$). Suppose that \mathbb{E} is a Euclidean space, and let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Consider the function $\varphi_C(\mathbf{x}) = \frac{1}{2}d_C^2(\mathbf{x})$. By Example 3.31, φ_C is differentiable over \mathbb{E} and $\nabla \varphi_C(\mathbf{x}) = \mathbf{x} - P_C(\mathbf{x})$. We will show that φ_C is 1-smooth. Indeed, for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,

$$\begin{aligned} \|\nabla \varphi_C(\mathbf{x}) - \nabla \varphi_C(\mathbf{y})\|^2 &= \|\mathbf{x} - \mathbf{y} - P_C(\mathbf{x}) + P_C(\mathbf{y})\|^2 \\ &= \|\mathbf{x} - \mathbf{y}\|^2 - 2\langle P_C(\mathbf{x}) - P_C(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \|P_C(\mathbf{x}) - P_C(\mathbf{y})\|^2 \\ &\stackrel{(*)}{\leq} \|\mathbf{x} - \mathbf{y}\|^2 - 2\|P_C(\mathbf{x}) - P_C(\mathbf{y})\|^2 + \|P_C(\mathbf{x}) - P_C(\mathbf{y})\|^2 \\ &= \|\mathbf{x} - \mathbf{y}\|^2 - \|P_C(\mathbf{x}) - P_C(\mathbf{y})\|^2 \\ &\leq \|\mathbf{x} - \mathbf{y}\|^2, \end{aligned}$$

where the inequality (*) follows by the firm nonexpansivity of the orthogonal projection operator (Theorem 5.4(a)). ■

Example 5.6 (1-smoothness of $\frac{1}{2}\|\cdot\|^2 - \frac{1}{2}d_C^2$). Suppose that \mathbb{E} is a Euclidean space, and let $C \subseteq \mathbb{E}$ be a nonempty closed convex set. Consider the function

$\psi_C(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 - \frac{1}{2}d_C^2(\mathbf{x})$. By Example 2.17, ψ_C is convex.²³ We will now show that it is 1-smooth. By Example 3.31, $\frac{1}{2}d_C^2(\mathbf{x})$ is differentiable over \mathbb{E} , and its gradient is given by $\mathbf{x} - P_C(\mathbf{x})$. Therefore,

$$\nabla\psi_C(\mathbf{x}) = \mathbf{x} - (\mathbf{x} - P_C(\mathbf{x})) = P_C(\mathbf{x}).$$

The 1-smoothness of ψ_C now follows by the nonexpansivity of the projection operator (Theorem 5.4(b))—for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,

$$\|\nabla\psi_C(\mathbf{x}) - \nabla\psi_C(\mathbf{y})\| = \|P_C(\mathbf{x}) - P_C(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|. \quad \blacksquare$$

5.1.1 The Descent Lemma

An extremely useful result on L -smooth functions is the descent lemma, which states that they can be upper bounded by a certain quadratic function.

Lemma 5.7 (descent lemma). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an L -smooth function ($L \geq 0$) over a given convex set D . Then for any $\mathbf{x}, \mathbf{y} \in D$,*

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.3)$$

Proof. By the fundamental theorem of calculus,

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt.$$

Therefore,

$$f(\mathbf{y}) - f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle dt.$$

Thus,

$$\begin{aligned} |f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| &= \left| \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle dt \right| \\ &\leq \int_0^1 |\langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| dt \\ &\stackrel{(*)}{\leq} \int_0^1 \|\nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x})\|_* \cdot \|\mathbf{y} - \mathbf{x}\| dt \\ &\leq \int_0^1 tL\|\mathbf{y} - \mathbf{x}\|^2 dt \\ &= \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2, \end{aligned}$$

where in $(*)$ we used the generalized Cauchy–Schwarz inequality (Lemma 1.4). \square

²³The convexity of ψ_C actually does not require the convexity of C ; see Example 2.17.

5.1.2 Characterizations of L -Smooth Functions

When f is convex, the next result gives several different and equivalent characterizations of the L -smoothness property of f over the entire space. Note that property (5.3) from the descent lemma is one of the mentioned equivalent properties.

Theorem 5.8 (characterizations of L -smoothness). *Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be a convex function, differentiable over \mathbb{E} , and let $L > 0$. Then the following claims are equivalent:*

- (i) f is L -smooth.
- (ii) $f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2} \|\mathbf{x} - \mathbf{y}\|^2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.
- (iii) $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.
- (iv) $\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \frac{1}{L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.
- (v) $f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{L}{2} \lambda(1 - \lambda) \|\mathbf{x} - \mathbf{y}\|^2$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ and $\lambda \in [0, 1]$.

Proof. (i) \Rightarrow (ii). The fact that (i) implies (ii) is just the descent lemma (Lemma 5.7).

(ii) \Rightarrow (iii). Suppose that (ii) is satisfied. We can assume that $\nabla f(\mathbf{x}) \neq \nabla f(\mathbf{y})$ since otherwise the inequality (iii) is trivial by the convexity of f . For a fixed $\mathbf{x} \in \mathbb{E}$ consider the function

$$g_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \quad \mathbf{y} \in \mathbb{E}.$$

The function $g_{\mathbf{x}}$ also satisfies property (ii). Indeed, for any $\mathbf{y}, \mathbf{z} \in \mathbb{E}$,

$$\begin{aligned} g_{\mathbf{x}}(\mathbf{z}) &= f(\mathbf{z}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{z} - \mathbf{x} \rangle \\ &\leq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle + \frac{L}{2} \|\mathbf{z} - \mathbf{y}\|^2 - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{z} - \mathbf{x} \rangle \\ &= f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{z} - \mathbf{y} \rangle + \frac{L}{2} \|\mathbf{z} - \mathbf{y}\|^2 \\ &= g_{\mathbf{x}}(\mathbf{y}) + \langle \nabla g_{\mathbf{x}}(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle + \frac{L}{2} \|\mathbf{z} - \mathbf{y}\|^2, \end{aligned} \tag{5.4}$$

where we used in the last equality the fact that $\nabla g_{\mathbf{x}}(\mathbf{y}) = \nabla f(\mathbf{y}) - \nabla f(\mathbf{x})$ for any $\mathbf{y} \in \mathbb{E}$. In particular, $\nabla g_{\mathbf{x}}(\mathbf{x}) = \mathbf{0}$, which by the convexity of $g_{\mathbf{x}}$ implies that \mathbf{x} is a global minimizer of g , meaning that

$$g_{\mathbf{x}}(\mathbf{x}) \leq g_{\mathbf{x}}(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{E}. \tag{5.5}$$

Let $\mathbf{y} \in \mathbb{E}$, and let $\mathbf{v} \in \mathbb{E}$ be a vector satisfying $\|\mathbf{v}\| = 1$ and $\langle \nabla g_{\mathbf{x}}(\mathbf{y}), \mathbf{v} \rangle = \|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*$. Substituting

$$\mathbf{z} = \mathbf{y} - \frac{\|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*}{L} \mathbf{v} \tag{5.6}$$

into (5.5) yields

$$0 = g_{\mathbf{x}}(\mathbf{x}) \leq g_{\mathbf{x}}\left(\mathbf{y} - \frac{\|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*}{L} \mathbf{v}\right).$$

Combining the last inequality with (5.4) (using the specific choice of \mathbf{z} given in (5.6)), we obtain

$$\begin{aligned}
 0 &= g_{\mathbf{x}}(\mathbf{x}) \\
 &\leq g_{\mathbf{x}}(\mathbf{y}) - \frac{\|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*}{L} \langle \nabla g_{\mathbf{x}}(\mathbf{y}), \mathbf{v} \rangle + \frac{1}{2L} \|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*^2 \cdot \|\mathbf{v}\|^2 \\
 &= g_{\mathbf{x}}(\mathbf{y}) - \frac{1}{2L} \|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*^2 \\
 &= f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle - \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2,
 \end{aligned}$$

which is claim (iii).

(iii) \Rightarrow (iv). Writing the inequality (iii) for the two pairs (\mathbf{x}, \mathbf{y}) , (\mathbf{y}, \mathbf{x}) yields

$$\begin{aligned}
 f(\mathbf{y}) &\geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2, \\
 f(\mathbf{x}) &\geq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2.
 \end{aligned}$$

Adding the two inequalities and rearranging terms results in (iv).

(iv) \Rightarrow (i). The Lipschitz condition

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \leq L\|\mathbf{x} - \mathbf{y}\|$$

is trivial when $\nabla f(\mathbf{x}) = \nabla f(\mathbf{y})$. We will therefore assume that $\nabla f(\mathbf{x}) \neq \nabla f(\mathbf{y})$. By (iv) and the generalized Cauchy–Schwarz inequality (Lemma 1.4) we have for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \cdot \|\mathbf{x} - \mathbf{y}\| \geq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \frac{1}{L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2.$$

Dividing by $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*$ and multiplying by L , (i) is obtained.

We have just shown the equivalence between (i), (ii), (iii), and (iv). To prove that (v) is also equivalent to each of these four claims, we will establish the equivalence (ii) \Leftrightarrow (v).

(ii) \Rightarrow (v). Let $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ and $\lambda \in [0, 1]$. Denote $\mathbf{x}_\lambda = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$. Then by (ii),

$$\begin{aligned}
 f(\mathbf{x}) &\leq f(\mathbf{x}_\lambda) + \langle \nabla f(\mathbf{x}_\lambda), \mathbf{x} - \mathbf{x}_\lambda \rangle + \frac{L}{2} \|\mathbf{x} - \mathbf{x}_\lambda\|^2, \\
 f(\mathbf{y}) &\leq f(\mathbf{x}_\lambda) + \langle \nabla f(\mathbf{x}_\lambda), \mathbf{y} - \mathbf{x}_\lambda \rangle + \frac{L}{2} \|\mathbf{y} - \mathbf{x}_\lambda\|^2,
 \end{aligned}$$

which is the same as

$$\begin{aligned}
 f(\mathbf{x}) &\leq f(\mathbf{x}_\lambda) + (1 - \lambda) \langle \nabla f(\mathbf{x}_\lambda), \mathbf{x} - \mathbf{y} \rangle + \frac{L(1 - \lambda)^2}{2} \|\mathbf{x} - \mathbf{y}\|^2 \\
 f(\mathbf{y}) &\leq f(\mathbf{x}_\lambda) + \lambda \langle \nabla f(\mathbf{x}_\lambda), \mathbf{y} - \mathbf{x} \rangle + \frac{L\lambda^2}{2} \|\mathbf{x} - \mathbf{y}\|^2.
 \end{aligned}$$

Multiplying the first inequality by λ and the second by $1 - \lambda$ and adding them yields the inequality (iv).

(v) \Rightarrow (ii). Rearranging terms in the inequality (v), we obtain that it is equivalent to

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \frac{f(\mathbf{x} + (1 - \lambda)(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{1 - \lambda} + \frac{L}{2}\lambda\|\mathbf{x} - \mathbf{y}\|^2.$$

Taking $\lambda \rightarrow 1^-$, the last inequality becomes

$$f(\mathbf{y}) \leq f(\mathbf{x}) + f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2,$$

which, by the fact that $f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$ (see Theorem 3.29), implies (ii). \square

Remark 5.9 (necessity of convexity in Theorem 5.8). *The convexity assumption in Theorem 5.8 is essential. Consider, for example, the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = -\frac{1}{2}\|\mathbf{x}\|_2^2$, which is 1-smooth w.r.t. the l_2 -norm but is not L -smooth for $L < 1$ (see Example 5.2). However, f is concave, and hence*

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle,$$

which implies that property (ii) of Theorem 5.8 is satisfied with $L = 0$, although the function is obviously not 0-smooth.

The next example will require the linear approximation theorem, which we now recall.

Theorem 5.10 (linear approximation theorem, [10, Theorem 1.24], [101, Fact 3.3.10]). *Let $f : U \rightarrow \mathbb{R}$ be a twice continuously differentiable function²⁴ over an open set $U \subseteq \mathbb{R}^n$, and let $\mathbf{x} \in U, r > 0$ satisfy $B(\mathbf{x}, r) \subseteq U$. Then for any $\mathbf{y} \in B(\mathbf{x}, r)$ there exists $\boldsymbol{\xi} \in [\mathbf{x}, \mathbf{y}]$ such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\boldsymbol{\xi})(\mathbf{y} - \mathbf{x}).$$

Example 5.11 ($(p - 1)$ -smoothness of the half-squared l_p -norm function).²⁵ Consider the convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_p^2 = \frac{1}{2} \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{2}{p}},$$

where $p \in [2, \infty)$. We assume that \mathbb{R}^n is endowed with the l_p -norm and show that f is $(p - 1)$ -smooth w.r.t. the l_p -norm. The result was already established for the case $p = 2$ in Example 5.2, and we will henceforth assume that $p > 2$. We begin by computing the partial derivatives:

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \begin{cases} \operatorname{sgn}(x_i) \frac{|x_i|^{p-1}}{\|\mathbf{x}\|_p^{p-2}}, & \mathbf{x} \neq \mathbf{0}, \\ 0, & \mathbf{x} = \mathbf{0}, \end{cases}$$

²⁴By “twice continuously differentiable over U ,” we mean that the function has second-order partial derivatives, which are continuous over U .

²⁵The analysis in Example 5.11 follows the derivation of Ben-Tal, Margalit, and Nemirovski [24, Appendix 1].

The partial derivatives are continuous over \mathbb{R}^n , and hence f is differentiable over \mathbb{R}^n (in the sense of Definition 3.28).²⁶ The second-order partial derivatives exist for any $\mathbf{x} \neq \mathbf{0}$ and are given by

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \begin{cases} (2-p) \operatorname{sgn}(x_i) \operatorname{sgn}(x_j) \frac{|x_i|^{p-1} |x_j|^{p-1}}{\|\mathbf{x}\|_p^{2p-2}}, & i \neq j, \\ (p-1) \frac{|x_i|^{p-2}}{\|\mathbf{x}\|_p^{p-2}} + (2-p) \frac{|x_i|^{2p-2}}{\|\mathbf{x}\|_p^{2p-2}}, & i = j. \end{cases}$$

It is easy to see that the second-order partial derivatives are continuous for any $\mathbf{x} \neq \mathbf{0}$. We will show that property (ii) of Theorem 5.8 is satisfied with $L = p - 1$. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be such that $\mathbf{0} \notin [\mathbf{x}, \mathbf{y}]$. Then by the linear approximation theorem (Theorem 5.10)—taking U to be some open set containing $[\mathbf{x}, \mathbf{y}]$ but not containing $\mathbf{0}$ —there exists $\boldsymbol{\xi} \in [\mathbf{x}, \mathbf{y}]$ for which

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{1}{2} (\mathbf{y} - \mathbf{x})^T \nabla^2 f(\boldsymbol{\xi}) (\mathbf{y} - \mathbf{x}). \quad (5.7)$$

We will show that $\mathbf{d}^T \nabla^2 f(\boldsymbol{\xi}) \mathbf{d} \leq (p-1) \|\mathbf{d}\|_p^2$ for any $\mathbf{d} \in \mathbb{R}^n$. Since $\nabla^2 f(t\boldsymbol{\xi}) = \nabla^2 f(\boldsymbol{\xi})$ for any $t \in \mathbb{R}$, we can assume without loss of generality that $\|\boldsymbol{\xi}\|_p = 1$. Now, for any $\mathbf{d} \in \mathbb{R}^n$,

$$\begin{aligned} \mathbf{d}^T \nabla^2 f(\boldsymbol{\xi}) \mathbf{d} &= (2-p) \|\boldsymbol{\xi}\|_p^{2-2p} \left(\sum_{i=1}^n |\xi_i|^{p-1} \operatorname{sgn}(\xi_i) d_i \right)^2 + (p-1) \|\boldsymbol{\xi}\|_p^{2-p} \sum_{i=1}^n |\xi_i|^{p-2} d_i^2 \\ &\leq (p-1) \|\boldsymbol{\xi}\|_p^{2-p} \sum_{i=1}^n |\xi_i|^{p-2} d_i^2, \end{aligned} \quad (5.8)$$

where the last inequality follows by the fact that $p > 2$. Using the generalized Cauchy–Schwarz inequality (Lemma 1.4) with $\|\cdot\| = \|\cdot\|_{\frac{p}{p-2}}$, we have

$$\begin{aligned} \sum_{i=1}^n |\xi_i|^{p-2} d_i^2 &\leq \left(\sum_{i=1}^n (|\xi_i|^{p-2})^{\frac{p}{p-2}} \right)^{\frac{p-2}{p}} \left(\sum_{i=1}^n (d_i^2)^{\frac{p}{2}} \right)^{\frac{2}{p}} \\ &= \left(\sum_{i=1}^n |\xi_i|^p \right)^{\frac{p-2}{p}} \left(\sum_{i=1}^n |d_i|^p \right)^{\frac{2}{p}} \\ &= \|\mathbf{d}\|_p^2. \end{aligned} \quad (5.9)$$

Combining (5.8) and (5.9), we obtain that for any $\mathbf{d} \in \mathbb{R}^n$,

$$\mathbf{d}^T \nabla^2 f(\boldsymbol{\xi}) \mathbf{d} \leq (p-1) \|\mathbf{d}\|_p^2,$$

and specifically, for $\mathbf{d} = \mathbf{x} - \mathbf{y}$,

$$(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\boldsymbol{\xi}) (\mathbf{y} - \mathbf{x}) \leq (p-1) \|\mathbf{x} - \mathbf{y}\|_p^2.$$

Plugging the above inequality into (5.7) implies the inequality

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{p-1}{2} \|\mathbf{x} - \mathbf{y}\|_p^2. \quad (5.10)$$

²⁶See, for example, [112, Theorem 9.21] for a precise statement of this result.

The inequality (5.10) was proven for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ that satisfy $\mathbf{0} \notin [\mathbf{x}, \mathbf{y}]$. We can show that it holds for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ using a continuity argument. Indeed, assume that $\mathbf{0} \in [\mathbf{x}, \mathbf{y}]$. Then we can find a sequence $\{\mathbf{y}_k\}_{k \geq 0}$ converging to \mathbf{y} for which $\mathbf{0} \notin [\mathbf{x}, \mathbf{y}_k]$. Thus, by what was already proven, for any $k \geq 0$,

$$f(\mathbf{y}_k) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y}_k - \mathbf{x}) + \frac{p-1}{2} \|\mathbf{x} - \mathbf{y}_k\|_p^2.$$

Taking $k \rightarrow \infty$ in the last inequality and using the continuity of f , we obtain that (5.10) holds. To conclude, we established that (5.10) holds for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and thus by Theorem 5.8 (equivalence between properties (i) and (ii)) and the convexity of f , it follows that f is $(p-1)$ -smooth w.r.t. the l_p -norm. ■

5.1.3 Second-Order Characterization

We will now consider the space $\mathbb{E} = \mathbb{R}^n$ endowed with the l_p -norm ($p \geq 1$). For twice continuously differentiable functions, it is possible to fully characterize the property of L -smoothness via the norm of the Hessian matrix.

Theorem 5.12 (L -smoothness and boundedness of the Hessian). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice continuously differentiable function over \mathbb{R}^n . Then for a given $L \geq 0$, the following two claims are equivalent:*

- (i) f is L -smooth w.r.t. the l_p -norm ($p \in [1, \infty]$).
- (ii) $\|\nabla^2 f(\mathbf{x})\|_{p,q} \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$, where $q \in [1, \infty]$ satisfies $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. (ii) \Rightarrow (i). Suppose that $\|\nabla^2 f(\mathbf{x})\|_{p,q} \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$. Then by the fundamental theorem of calculus, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\begin{aligned} \nabla f(\mathbf{y}) &= \nabla f(\mathbf{x}) + \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) (\mathbf{y} - \mathbf{x}) dt \\ &= \nabla f(\mathbf{x}) + \left(\int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) dt \right) \cdot (\mathbf{y} - \mathbf{x}). \end{aligned}$$

Then

$$\begin{aligned} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_q &= \left\| \left(\int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) dt \right) \cdot (\mathbf{y} - \mathbf{x}) \right\|_q \\ &\leq \left\| \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) dt \right\|_{p,q} \|\mathbf{y} - \mathbf{x}\|_p \\ &\leq \left(\int_0^1 \|\nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))\|_{p,q} dt \right) \|\mathbf{y} - \mathbf{x}\|_p \\ &\leq L \|\mathbf{y} - \mathbf{x}\|_p, \end{aligned}$$

establishing (i).

(i) \Rightarrow (ii). Suppose now that f is L -smooth w.r.t. the l_p -norm. Then by the fundamental theorem of calculus, for any $\mathbf{d} \in \mathbb{R}^n$ and $\alpha > 0$,

$$\nabla f(\mathbf{x} + \alpha \mathbf{d}) - \nabla f(\mathbf{x}) = \int_0^\alpha \nabla^2 f(\mathbf{x} + t\mathbf{d}) \mathbf{d} dt.$$

Thus,

$$\left\| \left(\int_0^\alpha \nabla^2 f(\mathbf{x} + t\mathbf{d}) dt \right) \mathbf{d} \right\|_q = \|\nabla f(\mathbf{x} + \alpha\mathbf{d}) - \nabla f(\mathbf{x})\|_q \leq \alpha L \|\mathbf{d}\|_p.$$

Dividing by α and taking the limit $\alpha \rightarrow 0^+$, we obtain

$$\|\nabla^2 f(\mathbf{x})\mathbf{d}\|_q \leq L \|\mathbf{d}\|_p \text{ for any } \mathbf{d} \in \mathbb{R}^n,$$

implying that $\|\nabla^2 f(\mathbf{x})\|_{p,q} \leq L$. \square

A direct consequence is that for twice continuously differentiable convex functions, L -smoothness w.r.t. the l_2 -norm is equivalent to the property that the maximum eigenvalue of the Hessian matrix is smaller than or equal to L .

Corollary 5.13. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice continuously differentiable convex function over \mathbb{R}^n . Then f is L -smooth w.r.t. the l_2 -norm if and only if $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$.*

Proof. Since f is convex, it follows that $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in \mathbb{R}^n$. Therefore, in this case,

$$\|\nabla^2 f(\mathbf{x})\|_{2,2} = \sqrt{\lambda_{\max}((\nabla^2 f(\mathbf{x}))^2)} = \lambda_{\max}(\nabla^2 f(\mathbf{x})),$$

which, combined with Theorem 5.12, establishes the desired result. \square

Example 5.14 (1-smoothness of $\sqrt{1 + \|\cdot\|_2^2}$ w.r.t. the l_2 -norm). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be the convex function given by

$$f(\mathbf{x}) = \sqrt{1 + \|\mathbf{x}\|_2^2}.$$

We will show that f is 1-smooth w.r.t. the l_2 -norm. For any $\mathbf{x} \in \mathbb{R}^n$,

$$\nabla f(\mathbf{x}) = \frac{\mathbf{x}}{\sqrt{\|\mathbf{x}\|_2^2 + 1}}$$

and

$$\nabla^2 f(\mathbf{x}) = \frac{1}{\sqrt{\|\mathbf{x}\|_2^2 + 1}} \mathbf{I} - \frac{\mathbf{x}\mathbf{x}^T}{(\|\mathbf{x}\|_2^2 + 1)^{3/2}} \preceq \frac{1}{\sqrt{\|\mathbf{x}\|_2^2 + 1}} \mathbf{I} \preceq \mathbf{I}.$$

Therefore, $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq 1$ for all $\mathbf{x} \in \mathbb{R}^n$, and hence by Corollary 5.13 it follows that f is 1-smooth w.r.t. the l_2 -norm. \blacksquare

Example 5.15 (1-smoothness of the log-sum-exp function w.r.t. the l_2, l_∞ norms). Consider the log-sum-exp function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \log(e^{x_1} + e^{x_2} + \cdots + e^{x_n}).$$

We will first show that it is 1-smooth w.r.t. the l_2 -norm. The partial derivatives of f are

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}, \quad i = 1, 2, \dots, n,$$

and the second-order partial derivatives are

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \begin{cases} -\frac{e^{x_i} e^{x_j}}{(\sum_{k=1}^n e^{x_k})^2}, & i \neq j, \\ -\frac{e^{x_i} e^{x_i}}{(\sum_{k=1}^n e^{x_k})^2} + \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}, & i = j. \end{cases}$$

We can thus write the Hessian matrix as

$$\nabla^2 f(\mathbf{x}) = \text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T,$$

where $w_i = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}$. To show that f is 1-smooth w.r.t. the l_2 -norm, note that for any $\mathbf{x} \in \mathbb{R}^n$,

$$\nabla^2 f(\mathbf{x}) = \text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T \preceq \text{diag}(\mathbf{w}) \preceq \mathbf{I},$$

and hence $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq 1$ for any $\mathbf{x} \in \mathbb{R}^n$. Noting that the log-sum-exp function is convex, we can invoke Corollary 5.13 and conclude that f is 1-smooth w.r.t. the l_2 -norm.

We will show that f is 1-smooth also w.r.t. the l_∞ -norm. For that, we begin by proving that for any $\mathbf{d} \in \mathbb{R}^n$,

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}) \mathbf{d} \leq \|\mathbf{d}\|_\infty^2. \quad (5.11)$$

Indeed,

$$\begin{aligned} \mathbf{d}^T \nabla^2 f(\mathbf{x}) \mathbf{d} &= \mathbf{d}^T (\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T) \mathbf{d} = \mathbf{d}^T \text{diag}(\mathbf{w}) \mathbf{d} - (\mathbf{w}^T \mathbf{d})^2 \\ &\leq \mathbf{d}^T \text{diag}(\mathbf{w}) \mathbf{d} \\ &= \sum_{i=1}^n w_i d_i^2 \\ &\leq \|\mathbf{d}\|_\infty^2 \sum_{i=1}^n w_i \\ &= \|\mathbf{d}\|_\infty^2. \end{aligned}$$

Now, since f is twice continuously differentiable over \mathbb{R}^n , it follows by the linear approximation theorem (Theorem 5.10) that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ there exists $\boldsymbol{\xi} \in [\mathbf{x}, \mathbf{y}]$ for which

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{1}{2} (\mathbf{y} - \mathbf{x})^T \nabla^2 f(\boldsymbol{\xi}) (\mathbf{y} - \mathbf{x}). \quad (5.12)$$

Combining (5.12) (taking $\mathbf{d} = \mathbf{y} - \mathbf{x}$) and (5.11), we obtain the inequality

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_\infty^2,$$

which by Theorem 5.8 (equivalence between properties (i) and (ii)) implies the 1-smoothness of f w.r.t. the l_∞ -norm. ■

5.1.4 Summary of Smoothness Parameter Computations

The table below summarizes the smoothness parameters of the functions discussed in this section. The last function will only be discussed later on in Example 6.62.

$f(\mathbf{x})$	$\text{dom}(f)$	Parameter	Norm	Reference
$\frac{1}{2}\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$ ($\mathbf{A} \in \mathbb{S}^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$)	\mathbb{R}^n	$\ \mathbf{A}\ _{p,q}$	l_p	Example 5.2
$\langle \mathbf{b}, \mathbf{x} \rangle + c$ ($\mathbf{b} \in \mathbb{E}^*, c \in \mathbb{R}$)	\mathbb{E}	0	any norm	Example 5.3
$\frac{1}{2}\ \mathbf{x}\ _p^2, \quad p \in [2, \infty)$	\mathbb{R}^n	$p - 1$	l_p	Example 5.11
$\sqrt{1 + \ \mathbf{x}\ _2^2}$	\mathbb{R}^n	1	l_2	Example 5.14
$\log(\sum_{i=1}^n e^{x_i})$	\mathbb{R}^n	1	l_2, l_∞	Example 5.15
$\frac{1}{2}d_C^2(\mathbf{x})$ ($\emptyset \neq C \subseteq \mathbb{E}$ closed convex)	\mathbb{E}	1	Euclidean	Example 5.5
$\frac{1}{2}\ \mathbf{x}\ ^2 - \frac{1}{2}d_C^2(\mathbf{x})$ ($\emptyset \neq C \subseteq \mathbb{E}$ closed convex)	\mathbb{E}	1	Euclidean	Example 5.6
$H_\mu(\mathbf{x}) \quad (\mu > 0)$	\mathbb{E}	$\frac{1}{\mu}$	Euclidean	Example 6.62

5.2 Strong Convexity

Definition 5.16 (strong convexity). A function $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is called **σ -strongly convex** for a given $\sigma > 0$ if $\text{dom}(f)$ is convex and the following inequality holds for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in [0, 1]$:

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.13)$$

We will sometimes use the terminology “strongly convex with parameter σ ” instead of “ σ -strongly convex.” It is important to note that the strong convexity parameter σ depends on the underlying norm, and we will therefore sometimes refer to it as the “strong convexity parameter w.r.t. $\|\cdot\|$.” Obviously, strongly convex functions are necessarily also convex since their domain is assumed to be convex and inequality (5.13) implies that for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in [0, 1]$, Jensen’s inequality is satisfied:

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}).$$

When the underlying set \mathbb{E} is Euclidean, meaning that $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ for any $\mathbf{x} \in \mathbb{E}$, we can write a different and simple property that is equivalent to strong convexity.

Theorem 5.17. Let \mathbb{E} be a Euclidean space. Then $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is a σ -strongly convex function ($\sigma > 0$) if and only if the function $f(\cdot) - \frac{\sigma}{2}\|\cdot\|^2$ is convex.

Proof. The function $g(\mathbf{x}) \equiv f(\mathbf{x}) - \frac{\sigma}{2}\|\mathbf{x}\|^2$ is convex if and only if its domain $\text{dom}(g) = \text{dom}(f)$ is convex and for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in [0, 1]$,

$$g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}).$$

The latter inequality is the same as

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) + \frac{\sigma}{2} [\|\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\|^2 - \lambda\|\mathbf{x}\|^2 - (1-\lambda)\|\mathbf{y}\|^2]. \quad (5.14)$$

Now, using the identity (which holds since the norm is assumed to be Euclidean)

$$\|\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\|^2 - \lambda\|\mathbf{x}\|^2 - (1-\lambda)\|\mathbf{y}\|^2 = -\lambda(1-\lambda)\|\mathbf{x} - \mathbf{y}\|^2,$$

combined with (5.14), we can conclude that the convexity of g is equivalent to the convexity of $\text{dom}(f)$ and the validity of the inequality

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1-\lambda)\|\mathbf{x} - \mathbf{y}\|^2$$

for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in [0, 1]$, namely, to the σ -strong convexity of f . \square

Remark 5.18. *The assumption that the underlying space is Euclidean is essential in Theorem 5.17. As an example, consider the negative entropy function over the unit simplex*

$$f(\mathbf{x}) \equiv \begin{cases} \sum_{i=1}^n x_i \log x_i, & \mathbf{x} \in \Delta_n, \\ \infty & \text{else.} \end{cases}$$

We will later show (in Example 5.27) that f is a 1-strongly convex function with respect to the l_1 -norm. Regardless of this fact, note that the function

$$g(\mathbf{x}) = f(\mathbf{x}) - \alpha\|\mathbf{x}\|_1^2$$

is convex for any $\alpha > 0$ since over the domain of f , we have that $\|\mathbf{x}\|_1 = 1$. Obviously, it is impossible that a function will be α -strongly convex for any $\alpha > 0$. Therefore, the characterization of strong convexity in Theorem 5.17 is not correct for any norm.

Note that if a function f is σ_1 -strongly convex ($\sigma_1 > 0$), then it is necessarily also σ_2 -strongly convex for any $\sigma_2 \in (0, \sigma_1)$. An interesting problem is to find the largest possible strong convexity parameter of a given function.

Example 5.19 (strong convexity of quadratic functions). Suppose that $\mathbb{E} = \mathbb{R}^n$ is endowed with the l_2 -norm, and consider the quadratic function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c,$$

where $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then by Theorem 5.17, f is strongly convex with parameter $\sigma > 0$ if and only if the function $\frac{1}{2}\mathbf{x}^T (\mathbf{A} - \sigma \mathbf{I}) \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$ is convex, which is equivalent to the matrix inequality $\mathbf{A} - \sigma \mathbf{I} \succeq \mathbf{0}$, namely, to the inequality $\lambda_{\min}(\mathbf{A}) \geq \sigma$. Thus, f is strongly convex if and only if \mathbf{A} is positive definite, and in that case, $\lambda_{\min}(\mathbf{A})$ is its largest possible strong convexity parameter. \blacksquare

A simple result is that the sum of a strongly convex function and a convex function is always a strongly convex function.

Lemma 5.20. *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a σ -strongly convex function ($\sigma > 0$), and let $g : \mathbb{E} \rightarrow (-\infty, \infty]$ be convex. Then $f + g$ is σ -strongly convex.*

Proof. Follows directly from the definitions of strong convexity and convexity. Since f and g are convex, both $\text{dom}(f)$ and $\text{dom}(g)$ are convex sets, and hence also $\text{dom}(f + g) = \text{dom}(f) \cap \text{dom}(g)$ is a convex set. Let $\mathbf{x}, \mathbf{y} \in \text{dom}(f) \cap \text{dom}(g)$ and $\lambda \in [0, 1]$. Then by the σ -strong convexity of f ,

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2.$$

Since g is convex,

$$g(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda g(\mathbf{x}) + (1 - \lambda)g(\mathbf{y}).$$

Adding the two inequalities, we obtain

$$(f + g)(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda(f + g)(\mathbf{x}) + (1 - \lambda)(f + g)(\mathbf{y}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2,$$

showing that $f + g$ is σ -strongly convex. \square

Example 5.21 (strong convexity of $\frac{1}{2}\|\cdot\|^2 + \delta_C$). Suppose that \mathbb{E} is a Euclidean space, and let $C \subseteq \mathbb{E}$ be a nonempty convex set. The function $\frac{1}{2}\|\mathbf{x}\|^2$ is 1-strongly convex (Example 5.19), and by the convexity of C , δ_C is convex. Therefore, by Lemma 5.20, the function $\frac{1}{2}\|\mathbf{x}\|^2 + \delta_C(\mathbf{x})$ is 1-strongly convex. \blacksquare

Theorem 5.24 below describes two properties that are equivalent to σ -strong convexity. The two properties are of a first-order nature in the sense that they are written in terms of the function and its subgradients. The proof uses the following version of the mean-value theorem for one-dimensional functions.

Lemma 5.22 (see [67, p. 26]). *Let $f : \mathbb{R} \rightarrow (-\infty, \infty]$ be a closed convex function, and let $[a, b] \subseteq \text{dom}(f)$ ($a < b$). Then*

$$f(b) - f(a) = \int_a^b h(t)dt,$$

where $h : (a, b) \rightarrow \mathbb{R}$ satisfies $h(t) \in \partial f(t)$ for any $t \in (a, b)$.

Another technical lemma that is being used in the proof is the so-called line segment principle.

Lemma 5.23 (line segment principle [108, Theorem 6.1]). *Let C be a convex set. Suppose that $\mathbf{x} \in \text{ri}(C)$, $\mathbf{y} \in \text{cl}(C)$, and let $\lambda \in (0, 1]$. Then $\lambda \mathbf{x} + (1 - \lambda)\mathbf{y} \in \text{ri}(C)$.*

Theorem 5.24 (first-order characterizations of strong convexity). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function. Then for a given $\sigma > 0$, the following three claims are equivalent:*

(i) f is σ -strongly convex.

(ii)

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle + \frac{\sigma}{2} \|\mathbf{y} - \mathbf{x}\|^2$$

for any $\mathbf{x} \in \text{dom}(\partial f)$, $\mathbf{y} \in \text{dom}(f)$ and $\mathbf{g} \in \partial f(\mathbf{x})$.

(iii)

$$\langle \mathbf{g}_{\mathbf{x}} - \mathbf{g}_{\mathbf{y}}, \mathbf{x} - \mathbf{y} \rangle \geq \sigma \|\mathbf{x} - \mathbf{y}\|^2 \quad (5.15)$$

for any $\mathbf{x}, \mathbf{y} \in \text{dom}(\partial f)$, and $\mathbf{g}_{\mathbf{x}} \in \partial f(\mathbf{x})$, $\mathbf{g}_{\mathbf{y}} \in \partial f(\mathbf{y})$.

Proof. (ii) \Rightarrow (i). Assume that (ii) is satisfied. To show (i), take $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in (0, 1)$. Take some $\mathbf{z} \in \text{ri}(\text{dom}(f))$. Then for any $\alpha \in (0, 1]$, by the line segment principle (Lemma 5.23), the vector $\tilde{\mathbf{x}} = (1 - \alpha)\mathbf{x} + \alpha\mathbf{z}$ is in $\text{ri}(\text{dom}(f))$. At this point we fix α . Using the notation $\mathbf{x}_\lambda = \lambda\tilde{\mathbf{x}} + (1 - \lambda)\mathbf{y}$, we obtain that $\mathbf{x}_\lambda \in \text{ri}(\text{dom}(f))$ for any $\lambda \in (0, 1)$, and hence, by Theorem 3.18, $\partial f(\mathbf{x}_\lambda) \neq \emptyset$, meaning that $\mathbf{x}_\lambda \in \text{dom}(\partial f)$. Take $\mathbf{g} \in \partial f(\mathbf{x}_\lambda)$. Then by (ii),

$$f(\tilde{\mathbf{x}}) \geq f(\mathbf{x}_\lambda) + \langle \mathbf{g}, \tilde{\mathbf{x}} - \mathbf{x}_\lambda \rangle + \frac{\sigma}{2} \|\tilde{\mathbf{x}} - \mathbf{x}_\lambda\|^2,$$

which is the same as

$$f(\tilde{\mathbf{x}}) \geq f(\mathbf{x}_\lambda) + (1 - \lambda)\langle \mathbf{g}, \tilde{\mathbf{x}} - \mathbf{y} \rangle + \frac{\sigma(1 - \lambda)^2}{2} \|\mathbf{y} - \tilde{\mathbf{x}}\|^2. \quad (5.16)$$

Similarly,

$$f(\mathbf{y}) \geq f(\mathbf{x}_\lambda) + \lambda\langle \mathbf{g}, \mathbf{y} - \tilde{\mathbf{x}} \rangle + \frac{\sigma\lambda^2}{2} \|\mathbf{y} - \tilde{\mathbf{x}}\|^2. \quad (5.17)$$

Multiplying (5.16) by λ and (5.17) by $1 - \lambda$ and adding the two resulting inequalities, we obtain that

$$f(\lambda\tilde{\mathbf{x}} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\tilde{\mathbf{x}}) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma\lambda(1 - \lambda)}{2} \|\tilde{\mathbf{x}} - \mathbf{y}\|^2.$$

Plugging the expression for $\tilde{\mathbf{x}}$ in the above inequality, we obtain that

$$g_1(\alpha) \leq \lambda g_2(\alpha) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma\lambda(1 - \lambda)}{2} \|(1 - \alpha)\mathbf{x} + \alpha\mathbf{z} - \mathbf{y}\|^2, \quad (5.18)$$

where $g_1(\alpha) \equiv f(\lambda(1 - \alpha)\mathbf{x} + (1 - \lambda)\mathbf{y} + \lambda\alpha\mathbf{z})$ and $g_2(\alpha) \equiv f((1 - \alpha)\mathbf{x} + \alpha\mathbf{z})$. The functions g_1 and g_2 are one-dimensional proper closed and convex functions, and consequently, by Theorem 2.22, they are also continuous over their domain. Thus, taking $\alpha \rightarrow 0^+$ in (5.18), it follows that

$$g_1(0) \leq \lambda g_2(0) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma\lambda(1 - \lambda)}{2} \|\mathbf{x} - \mathbf{y}\|^2.$$

Finally, since $g_1(0) = f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y})$ and $g_2(0) = f(\mathbf{x})$, we obtain the inequality

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma\lambda(1 - \lambda)}{2} \|\mathbf{x} - \mathbf{y}\|^2,$$

establishing the σ -strong convexity of f .

(i) \Rightarrow (iii). Assume that (i) is satisfied. Let $\mathbf{x}, \mathbf{y} \in \text{dom}(\partial f)$ and $\mathbf{g}_{\mathbf{x}} \in \partial f(\mathbf{x}), \mathbf{g}_{\mathbf{y}} \in \partial f(\mathbf{y})$. We will show that inequality (5.15) is satisfied. Let $\lambda \in [0, 1]$ and denote $\mathbf{x}_{\lambda} = \lambda \mathbf{x} + (1 - \lambda)\mathbf{y}$. By condition (i),

$$f(\mathbf{x}_{\lambda}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2,$$

which is the same as

$$\frac{f(\mathbf{x}_{\lambda}) - f(\mathbf{x})}{1 - \lambda} \leq f(\mathbf{y}) - f(\mathbf{x}) - \frac{\sigma}{2}\lambda\|\mathbf{x} - \mathbf{y}\|^2, \quad (5.19)$$

Since $\mathbf{g}_{\mathbf{x}} \in \partial f(\mathbf{x})$,

$$\frac{f(\mathbf{x}_{\lambda}) - f(\mathbf{x})}{1 - \lambda} \geq \frac{\langle \mathbf{g}_{\mathbf{x}}, \mathbf{x}_{\lambda} - \mathbf{x} \rangle}{1 - \lambda} = \langle \mathbf{g}_{\mathbf{x}}, \mathbf{y} - \mathbf{x} \rangle,$$

which, combined with (5.19), yields the inequality

$$\langle \mathbf{g}_{\mathbf{x}}, \mathbf{y} - \mathbf{x} \rangle \leq f(\mathbf{y}) - f(\mathbf{x}) - \frac{\sigma\lambda}{2}\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.20)$$

Inequality (5.20) holds for any $\lambda \in [0, 1]$. Taking the limit $\lambda \rightarrow 1^-$, we conclude that

$$\langle \mathbf{g}_{\mathbf{x}}, \mathbf{y} - \mathbf{x} \rangle \leq f(\mathbf{y}) - f(\mathbf{x}) - \frac{\sigma}{2}\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.21)$$

Changing the roles of \mathbf{x} and \mathbf{y} yields the inequality

$$\langle \mathbf{g}_{\mathbf{y}}, \mathbf{x} - \mathbf{y} \rangle \leq f(\mathbf{x}) - f(\mathbf{y}) - \frac{\sigma}{2}\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.22)$$

Adding inequalities (5.21) and (5.22), we can finally conclude that

$$\langle \mathbf{g}_{\mathbf{x}} - \mathbf{g}_{\mathbf{y}}, \mathbf{x} - \mathbf{y} \rangle \geq \sigma\|\mathbf{x} - \mathbf{y}\|^2,$$

which is the desired inequality.

(iii) \Rightarrow (ii) Suppose that (iii) is satisfied, and let $\mathbf{x} \in \text{dom}(\partial f), \mathbf{y} \in \text{dom}(f)$, and $\mathbf{g} \in \partial f(\mathbf{x})$. Let \mathbf{z} be any vector in $\text{ri}(\text{dom}(f))$, and define $\tilde{\mathbf{y}} = (1 - \alpha)\mathbf{y} + \alpha\mathbf{z}$ for some $\alpha \in (0, 1)$, which at this point we fix. By the line segment principle (Lemma 5.23), $\tilde{\mathbf{y}} \in \text{ri}(\text{dom}(f))$. Consider now the one-dimensional function

$$\varphi(\lambda) = f(\mathbf{x}_{\lambda}), \quad \lambda \in [0, 1],$$

where $\mathbf{x}_{\lambda} = (1 - \lambda)\mathbf{x} + \lambda\tilde{\mathbf{y}}$. For any $\lambda \in (0, 1)$, let $\mathbf{g}_{\lambda} \in \partial f(\mathbf{x}_{\lambda})$ (whose existence is guaranteed since $\mathbf{x}_{\lambda} \in \text{ri}(\text{dom}(f))$ by the line segment principle). Then $\langle \mathbf{g}_{\lambda}, \tilde{\mathbf{y}} - \mathbf{x} \rangle \in \partial\varphi(\lambda)$, and hence by the mean-value theorem (Lemma 5.22),

$$f(\tilde{\mathbf{y}}) - f(\mathbf{x}) = \varphi(1) - \varphi(0) = \int_0^1 \langle \mathbf{g}_{\lambda}, \tilde{\mathbf{y}} - \mathbf{x} \rangle d\lambda. \quad (5.23)$$

Since $\mathbf{g} \in \partial f(\mathbf{x})$ and $\mathbf{g}_{\lambda} \in \partial f(\mathbf{x}_{\lambda})$, by property (iii),

$$\langle \mathbf{g}_{\lambda} - \mathbf{g}, \mathbf{x}_{\lambda} - \mathbf{x} \rangle \geq \sigma\|\mathbf{x}_{\lambda} - \mathbf{x}\|^2,$$

which is equivalent to

$$\langle \mathbf{g}_\lambda, \tilde{\mathbf{y}} - \mathbf{x} \rangle \geq \langle \mathbf{g}, \tilde{\mathbf{y}} - \mathbf{x} \rangle + \sigma \lambda \|\tilde{\mathbf{y}} - \mathbf{x}\|^2.$$

Plugging the last inequality into (5.23), we obtain that

$$\begin{aligned} f(\tilde{\mathbf{y}}) - f(\mathbf{x}) &\geq \int_0^1 [\langle \mathbf{g}, \tilde{\mathbf{y}} - \mathbf{x} \rangle + \sigma \lambda \|\tilde{\mathbf{y}} - \mathbf{x}\|^2] d\lambda \\ &= \langle \mathbf{g}, \tilde{\mathbf{y}} - \mathbf{x} \rangle + \frac{\sigma}{2} \|\tilde{\mathbf{y}} - \mathbf{x}\|^2. \end{aligned}$$

Recalling the definition of $\tilde{\mathbf{y}}$, we obtain that for any $\alpha \in (0, 1)$,

$$f((1 - \alpha)\mathbf{y} + \alpha\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{g}, (1 - \alpha)\mathbf{y} + \alpha\mathbf{z} - \mathbf{x} \rangle + \frac{\sigma}{2} \|(1 - \alpha)\mathbf{y} + \alpha\mathbf{z} - \mathbf{x}\|^2.$$

Taking $\alpha \rightarrow 0^+$ and using the continuity of the one-dimensional function $\alpha \mapsto f((1 - \alpha)\mathbf{y} + \alpha\mathbf{z})$ over $[0, 1]$ (follows by invoking Theorem 2.22 and recalling that the one-dimensional function is closed and convex), we obtain the desired result

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle + \frac{\sigma}{2} \|\mathbf{y} - \mathbf{x}\|^2. \quad \square$$

The next theorem states that a proper closed and strongly convex function has a unique minimizer and that it satisfies a certain growth property around the minimizer.

Theorem 5.25 (existence and uniqueness of a minimizer of closed strongly convex functions). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and σ -strongly convex function ($\sigma > 0$). Then*

- (a) *f has a unique minimizer;*
- (b) *$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \frac{\sigma}{2} \|\mathbf{x} - \mathbf{x}^*\|^2$ for all $\mathbf{x} \in \text{dom}(f)$, where \mathbf{x}^* is the unique minimizer of f .*

Proof. (a) Since $\text{dom}(f)$ is nonempty and convex, it follows that there exists $\mathbf{x}_0 \in \text{ri}(\text{dom}(f))$ (Theorem 3.17), and consequently, by Theorem 3.18, $\partial f(\mathbf{x}_0) \neq \emptyset$. Let $\mathbf{g} \in \partial f(\mathbf{x}_0)$. Then by the equivalence between σ -strong convexity and property (ii) of Theorem 5.24, it follows that

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \langle \mathbf{g}, \mathbf{x} - \mathbf{x}_0 \rangle + \frac{\sigma}{2} \|\mathbf{x} - \mathbf{x}_0\|^2 \text{ for all } \mathbf{x} \in \mathbb{E}.$$

Since all norms in finite dimensional spaces are equivalent, there exists a constant $C > 0$ such that

$$\|\mathbf{y}\| \geq \sqrt{C} \|\mathbf{y}\|_a,$$

where $\|\cdot\|_a \equiv \sqrt{\langle \cdot, \cdot \rangle}$ denotes the Euclidean norm associated with the inner product of the space \mathbb{E} (which might be different than the endowed norm $\|\cdot\|$). Therefore,

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \langle \mathbf{g}, \mathbf{x} - \mathbf{x}_0 \rangle + \frac{C\sigma}{2} \|\mathbf{x} - \mathbf{x}_0\|_a^2 \text{ for any } \mathbf{x} \in \mathbb{E},$$

which is the same as

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) - \frac{1}{2C\sigma} \|\mathbf{g}\|_a^2 + \frac{C\sigma}{2} \left\| \mathbf{x} - \left(\mathbf{x}_0 - \frac{1}{C\sigma} \mathbf{g} \right) \right\|_a^2 \text{ for any } \mathbf{x} \in \mathbb{E}.$$

In particular, it follows that

$$\text{Lev}(f, f(\mathbf{x}_0)) \subseteq B_{\|\cdot\|_a} \left[\mathbf{x}_0 - \frac{1}{C\sigma} \mathbf{g}, \frac{1}{C\sigma} \|\mathbf{g}\|_a \right].$$

Since f is closed, the above level set is closed (Theorem 2.6), and since it is contained in a ball, it is also bounded. Therefore, $\text{Lev}(f, f(\mathbf{x}_0))$ is compact. We can thus deduce that the optimal set of the problem of minimizing f over $\text{dom}(f)$ is the same as the optimal set of the problem of minimizing f over the nonempty compact set $\text{Lev}(f, f(\mathbf{x}_0))$. Invoking Weierstrass theorem for closed functions (Theorem 2.12), it follows that a minimizer exists. To show the uniqueness, assume that $\tilde{\mathbf{x}}$ and $\hat{\mathbf{x}}$ are minimizers of f . Then $f(\tilde{\mathbf{x}}) = f(\hat{\mathbf{x}}) = f_{\text{opt}}$, where f_{opt} is the minimal value of f . Then by the definition of σ -strong convexity of f ,

$$f_{\text{opt}} \leq f\left(\frac{1}{2}\tilde{\mathbf{x}} + \frac{1}{2}\hat{\mathbf{x}}\right) \leq \frac{1}{2}f(\tilde{\mathbf{x}}) + \frac{1}{2}f(\hat{\mathbf{x}}) - \frac{\sigma}{8}\|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|^2 = f_{\text{opt}} - \frac{\sigma}{8}\|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|^2,$$

implying that $\tilde{\mathbf{x}} = \hat{\mathbf{x}}$ and hence establishing the uniqueness of the minimizer of f .

(b) Let \mathbf{x}^* be the unique minimizer of f . Then by Fermat's optimality condition (Theorem 3.63), $\mathbf{0} \in \partial f(\mathbf{x}^*)$ and hence by using the equivalence between σ -strong convexity and property (ii) of Theorem 5.24, it follows that

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \langle \mathbf{0}, \mathbf{x} - \mathbf{x}^* \rangle + \frac{\sigma}{2} \|\mathbf{x} - \mathbf{x}^*\|^2 = \frac{\sigma}{2} \|\mathbf{x} - \mathbf{x}^*\|^2 \quad (5.24)$$

for any $\mathbf{x} \in \mathbb{E}$, establishing claim (b). \square

5.3 Smoothness and Strong Convexity Correspondence

5.3.1 The Conjugate Correspondence Theorem

An extremely useful connection between smoothness and strong convexity is given in the conjugate correspondence theorem that, loosely speaking, states that f is σ -strongly convex if and only if f^* is $\frac{1}{\sigma}$ -smooth.

Theorem 5.26 (conjugate correspondence theorem). *Let $\sigma > 0$. Then*

- (a) *If $f : \mathbb{E} \rightarrow \mathbb{R}$ is a $\frac{1}{\sigma}$ -smooth convex function, then f^* is σ -strongly convex w.r.t. the dual norm $\|\cdot\|_*$.*
- (b) *If $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is a proper closed σ -strongly convex function, then $f^* : \mathbb{E}^* \rightarrow \mathbb{R}$ is $\frac{1}{\sigma}$ -smooth.*

Proof. (a) Suppose that $f : \mathbb{E} \rightarrow \mathbb{R}$ is a $\frac{1}{\sigma}$ -smooth convex function. To prove that f^* is σ -strongly convex, take $\mathbf{y}_1, \mathbf{y}_2 \in \text{dom}(\partial f^*)$ and $\mathbf{v}_1 \in \partial f^*(\mathbf{y}_1), \mathbf{v}_2 \in \partial f^*(\mathbf{y}_2)$. Then by the conjugate subgradient theorem (Theorem 4.20), using also the properness closedness and convexity of f , it follows that $\mathbf{y}_1 \in \partial f(\mathbf{v}_1)$ and $\mathbf{y}_2 \in \partial f(\mathbf{v}_2)$, which, by the differentiability of f , implies that $\mathbf{y}_1 = \nabla f(\mathbf{v}_1)$ and $\mathbf{y}_2 = \nabla f(\mathbf{v}_2)$ (see Theorem 3.33). By the equivalence between properties (i) and (iv) in Theorem 5.8, we can write

$$\langle \mathbf{y}_1 - \mathbf{y}_2, \mathbf{v}_1 - \mathbf{v}_2 \rangle \geq \sigma \|\mathbf{y}_1 - \mathbf{y}_2\|_*^2.$$

Since the last inequality holds for any $\mathbf{y}_1, \mathbf{y}_2 \in \text{dom}(\partial f^*)$ and $\mathbf{v}_1 \in \partial f^*(\mathbf{y}_1), \mathbf{v}_2 \in \partial f^*(\mathbf{y}_2)$, it follows by the equivalence between σ -strong convexity and property (iii) of Theorem 5.24 that f^* is a σ -strongly convex function.

(b) Suppose that f is a proper closed σ -strongly convex function. By the conjugate subgradient theorem (Corollary 4.21),

$$\partial f^*(\mathbf{y}) = \operatorname{argmax}_{\mathbf{x} \in \mathbb{E}} \{ \langle \mathbf{x}, \mathbf{y} \rangle - f(\mathbf{x}) \} \text{ for any } \mathbf{y} \in \mathbb{E}^*.$$

Thus, by the strong convexity and closedness of f , along with Theorem 5.25(a), it follows that $\partial f^*(\mathbf{y})$ is a singleton for any $\mathbf{y} \in \mathbb{E}^*$. Therefore, by Theorem 3.33, f^* is differentiable over the entire dual space \mathbb{E}^* . To show the $\frac{1}{\sigma}$ -smoothness of f^* , take $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{E}^*$ and denote $\mathbf{v}_1 = \nabla f^*(\mathbf{y}_1), \mathbf{v}_2 = \nabla f^*(\mathbf{y}_2)$. These relations, by the conjugate subgradient theorem (Theorem 4.20), are equivalent to $\mathbf{y}_1 \in \partial f(\mathbf{v}_1), \mathbf{y}_2 \in \partial f(\mathbf{v}_2)$. Therefore, by Theorem 5.24 (equivalence between properties (i) and (iii)),

$$\langle \mathbf{y}_1 - \mathbf{y}_2, \mathbf{v}_1 - \mathbf{v}_2 \rangle \geq \sigma \|\mathbf{v}_1 - \mathbf{v}_2\|^2,$$

that is,

$$\langle \mathbf{y}_1 - \mathbf{y}_2, \nabla f^*(\mathbf{y}_1) - \nabla f^*(\mathbf{y}_2) \rangle \geq \sigma \|\nabla f^*(\mathbf{y}_1) - \nabla f^*(\mathbf{y}_2)\|^2,$$

which, combined with the generalized Cauchy–Schwarz inequality (Lemma 1.4), implies the inequality

$$\|\nabla f^*(\mathbf{y}_1) - \nabla f^*(\mathbf{y}_2)\| \leq \frac{1}{\sigma} \|\mathbf{y}_1 - \mathbf{y}_2\|_*,$$

proving the $\frac{1}{\sigma}$ -smoothness of f^* . \square

5.3.2 Examples of Strongly Convex Functions

We can use the conjugate correspondence theorem (Theorem 5.26) to conclude several results on the strong convexity of several important functions.

Example 5.27 (negative entropy over the unit simplex). Consider the function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ given by

$$f(\mathbf{x}) = \begin{cases} \sum_{i=1}^n x_i \log x_i, & \mathbf{x} \in \Delta_n, \\ \infty & \text{else.} \end{cases}$$

Then, by Section 4.4.10, the conjugate of this function is the log-sum-exp function $f^*(\mathbf{y}) = \log(\sum_{i=1}^n e^{y_i})$, which, by Example 5.15, is a 1-smooth function w.r.t. both the l_∞ - and l_2 -norms. Consequently, by the conjugate correspondence theorem, f is 1-strongly convex w.r.t. both the l_1 - and l_2 -norms. ■

Example 5.28 (squared p -norm for $p \in (1, 2]$). Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_p^2$ ($p \in (1, 2]$). Then, by Section 4.4.15, $f^*(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|_q^2$, where $q \geq 2$ is determined by the relation $\frac{1}{p} + \frac{1}{q} = 1$. By Example 5.11, f^* is a $(q-1)$ -smooth function w.r.t. the l_q -norm, which, by the conjugate correspondence theorem, implies that the function f is $\frac{1}{q-1}$ -strongly convex w.r.t. the l_p -norm. Since $\frac{1}{q-1} = p-1$, we conclude that f is $(p-1)$ -strongly convex w.r.t. the l_p -norm. ■

Example 5.29 (l_2 ball-pen function). Consider the ball-pen function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ given by

$$f(\mathbf{x}) = \begin{cases} -\sqrt{1 - \|\mathbf{x}\|_2^2}, & \|\mathbf{x}\|_2 \leq 1, \\ \infty & \text{else.} \end{cases}$$

By Section 4.4.13, the conjugate of f is

$$f^*(\mathbf{y}) = \sqrt{\|\mathbf{y}\|_2^2 + 1},$$

which, by Example 5.14, is known to be 1-smooth w.r.t. the l_2 -norm, and hence, by the conjugate correspondence theorem, f is 1-strongly convex w.r.t. the l_2 -norm. ■

The table below contains all the strongly convex functions described in this chapter.

$f(\mathbf{x})$	$\text{dom}(f)$	Strong convexity parameter	Norm	Reference
$\frac{1}{2}\mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$ ($\mathbf{A} \in \mathbb{S}_{++}^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$)	\mathbb{R}^n	$\lambda_{\min}(\mathbf{A})$	l_2	Example 5.19
$\frac{1}{2}\ \mathbf{x}\ ^2 + \delta_C(\mathbf{x})$ ($\emptyset \neq C \subseteq \mathbb{E}$ convex)	C	1	Euclidean	Example 5.21
$-\sqrt{1 - \ \mathbf{x}\ _2^2}$	$B_{\ \cdot\ _2}[\mathbf{0}, 1]$	1	l_2	Example 5.29
$\frac{1}{2}\ \mathbf{x}\ _p^2$ ($p \in (1, 2]$)	\mathbb{R}^n	$p-1$	l_p	Example 5.28
$\sum_{i=1}^n x_i \log x_i$	Δ_n	1	l_2 or l_1	Example 5.27

5.3.3 Smoothness and Differentiability of the Infimal Convolution

We will now show that under appropriate conditions, the infimal convolution of a convex function and an L -smooth convex function is also L -smooth; in addition, we will derive an expression for the gradient. The proof of the result is based on the conjugate correspondence theorem.

Theorem 5.30 (smoothness of the infimal convolution). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function, and let $\omega : \mathbb{E} \rightarrow \mathbb{R}$ be an L -smooth convex function. Assume that $f \square \omega$ is real-valued. Then the following hold:*

- (a) $f \square \omega$ is L -smooth.
- (b) Let $\mathbf{x} \in \mathbb{E}$, and assume that $\mathbf{u}(\mathbf{x})$ is a minimizer of

$$\min_{\mathbf{u}} \{f(\mathbf{u}) + \omega(\mathbf{x} - \mathbf{u})\}. \quad (5.25)$$

Then $\nabla(f \square \omega)(\mathbf{x}) = \nabla\omega(\mathbf{x} - \mathbf{u}(\mathbf{x}))$.

Proof. (a) By Theorem 4.19,

$$f \square \omega = (f^* + \omega^*)^*.$$

Since f and ω are proper closed and convex, then so are f^*, ω^* (Theorems 4.3, 4.5). In addition, by the conjugate correspondence theorem (Theorem 5.26), ω^* is $\frac{1}{L}$ -strongly convex. Therefore, by Lemma 5.20, $f^* + \omega^*$ is $\frac{1}{L}$ -strongly convex, and it is also closed as a sum of closed functions; we will prove that it is also proper. Indeed, by Theorem 4.16,

$$(f \square \omega)^* = f^* + \omega^*.$$

Since $f \square \omega$ is convex (by Theorem 2.19) and proper, it follows that $f^* + \omega^*$ is proper as a conjugate of a proper and convex function (Theorem 4.5). Thus, since $f^* + \omega^*$ is proper closed and $\frac{1}{L}$ -strongly convex function, by the conjugate correspondence theorem, it follows that $f \square \omega = (f^* + \omega^*)^*$ is L -smooth.

- (b) Let $\mathbf{x} \in \mathbb{E}$ be such that $\mathbf{u}(\mathbf{x})$ is a minimizer of (5.25), namely,

$$(f \square \omega)(\mathbf{x}) = f(\mathbf{u}(\mathbf{x})) + \omega(\mathbf{x} - \mathbf{u}(\mathbf{x})). \quad (5.26)$$

For convenience, define $\mathbf{z} \equiv \nabla\omega(\mathbf{x} - \mathbf{u}(\mathbf{x}))$. Our objective is to show that $\nabla(f \square \omega)(\mathbf{x}) = \mathbf{z}$. This means that we have to show that for any $\boldsymbol{\xi} \in \mathbb{E}$, $\lim_{\|\boldsymbol{\xi}\| \rightarrow 0} |\phi(\boldsymbol{\xi})|/\|\boldsymbol{\xi}\| = 0$, where $\phi(\boldsymbol{\xi}) \equiv (f \square \omega)(\mathbf{x} + \boldsymbol{\xi}) - (f \square \omega)(\mathbf{x}) - \langle \boldsymbol{\xi}, \mathbf{z} \rangle$. By the definition of the infimal convolution,

$$(f \square \omega)(\mathbf{x} + \boldsymbol{\xi}) \leq f(\mathbf{u}(\mathbf{x})) + \omega(\mathbf{x} + \boldsymbol{\xi} - \mathbf{u}(\mathbf{x})), \quad (5.27)$$

which, combined with (5.26), yields

$$\begin{aligned}
\phi(\boldsymbol{\xi}) &= (f \square \omega)(\mathbf{x} + \boldsymbol{\xi}) - (f \square \omega)(\mathbf{x}) - \langle \boldsymbol{\xi}, \mathbf{z} \rangle \\
&\leq \omega(\mathbf{x} + \boldsymbol{\xi} - \mathbf{u}(\mathbf{x})) - \omega(\mathbf{x} - \mathbf{u}(\mathbf{x})) - \langle \boldsymbol{\xi}, \mathbf{z} \rangle && [(5.26), (5.27)] \\
&\leq \langle \boldsymbol{\xi}, \nabla \omega(\mathbf{x} + \boldsymbol{\xi} - \mathbf{u}(\mathbf{x})) \rangle - \langle \boldsymbol{\xi}, \mathbf{z} \rangle, && [\text{gradient inequality for } \omega] \\
&= \langle \boldsymbol{\xi}, \nabla \omega(\mathbf{x} + \boldsymbol{\xi} - \mathbf{u}(\mathbf{x})) - \nabla \omega(\mathbf{x} - \mathbf{u}(\mathbf{x})) \rangle && [\text{substitution of } \mathbf{z}] \\
&\leq \|\boldsymbol{\xi}\| \cdot \|\nabla \omega(\mathbf{x} + \boldsymbol{\xi} - \mathbf{u}(\mathbf{x})) - \nabla \omega(\mathbf{x} - \mathbf{u}(\mathbf{x}))\|_* && [\text{generalized Cauchy-Schwarz}] \\
&\leq L\|\boldsymbol{\xi}\|^2. && [L\text{-smoothness of } \omega]
\end{aligned}$$

To complete the proof, it is enough to show that we also have $\phi(\boldsymbol{\xi}) \geq -L\|\boldsymbol{\xi}\|^2$. Since $f \square \omega$ is convex, so is ϕ , which, along the fact that $\phi(\mathbf{0}) = 0$, implies that $\phi(\boldsymbol{\xi}) \geq -\phi(-\boldsymbol{\xi})$, and hence the desired result follows. \square

Example 5.31 (revisiting the 1-smoothness of $\frac{1}{2}d_C^2$). Suppose that \mathbb{E} is a Euclidean space and let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Consider the function $\varphi_C(\mathbf{x}) = \frac{1}{2}d_C^2(\mathbf{x})$. We have already shown in Example 5.5 that it is 1-smooth. We will provide here a second proof for this result, which is based on Theorem 5.30. Note that $\varphi_C = \delta_C \square h$, where $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$. Since h is a real-valued 1-smooth convex function, and since δ_C is proper closed and convex, it follows by Theorem 5.30 that φ_C is 1-smooth. \blacksquare