1 Outline

In this lecture, we study

- Smooth approximation of nonsmooth functions,
- Smoothing-based optimization of nonsmooth functions,
- Fenchel conjugate and Moreau decomposition.

2 Smooth approximation

We say that a convex function f is (p, q)-smoothable if for any $\mu > 0$, there exists a convex function f_μ such that

1. f_μ is (p/μ)-smooth, and
2. $f_\mu(x) \leq f(x) \leq f_\mu(x) + q\mu$ for all x.

Example 14.1 (ℓ_1 norm)**.** Consider the Huber function given by

$$L_\mu(z) = \begin{cases} \frac{z^2}{2\mu}, & \text{if } |z| \leq \mu \\ |z| - \frac{\mu}{2}, & \text{otherwise} \end{cases}. \tag{14.1}$$

Note that

$$L_\mu(z) \leq |z| \leq L_\mu(z) + \frac{\mu}{2}.$$

Moreover, L_μ is $(1/\mu)$-smooth. This implies that $|\cdot|$ is $(1, 1/2)$-smoothable. Based on this, note

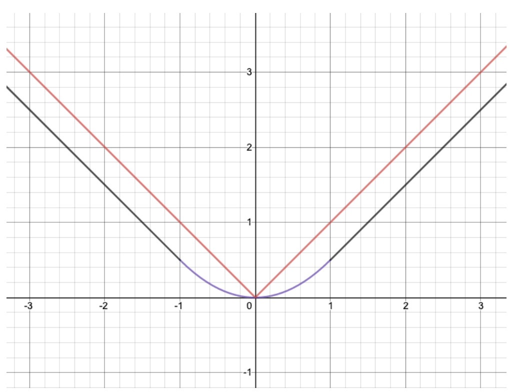

Figure 14.1: Huber function

that for $x \in \mathbb{R}^d$, $f_\mu(x) = \sum_{i=1}^d L_\mu(x_i)$ is $(1/\mu)$-smooth and satisfies

$$f_\mu(x) \leq \|x\|_1 \leq f_\mu(x) + \frac{d\mu}{2}.$$

Hence, $\|\cdot\|_1$ is $(1, d/2)$-smoothable.

Example 14.2 (ℓ_2 norm)**.** For $x \in \mathbb{R}^d$, take

$$f_\mu(x) = \sqrt{\|x\|_2^2 + \mu^2} - \mu.$$

Then for any $\mu \geq 0$, we have

$$f_\mu(x) \leq \|x\|_2 \leq f_\mu(x) + \mu.$$

Moreover, we have

$$\nabla f_\mu(x) = \frac{x}{\sqrt{\|x\|_2^2 + \mu^2}}, \quad \nabla^2 f_\mu(x) = \frac{1}{\sqrt{\|x\|_2^2 + \mu^2}} I - \frac{1}{(\|x\|_2^2 + \mu^2)^{3/2}} x x^\top.$$

Then

$$\nabla^2 f_\mu(x) \preceq \frac{1}{\sqrt{\|x\|_2^2 + \mu^2}} I \preceq \frac{1}{\mu} I.$$

Therefore, f_μ is $(1/\mu)$-smooth. This means that $\|\cdot\|_2$ is $(1,1)$-smoothable.

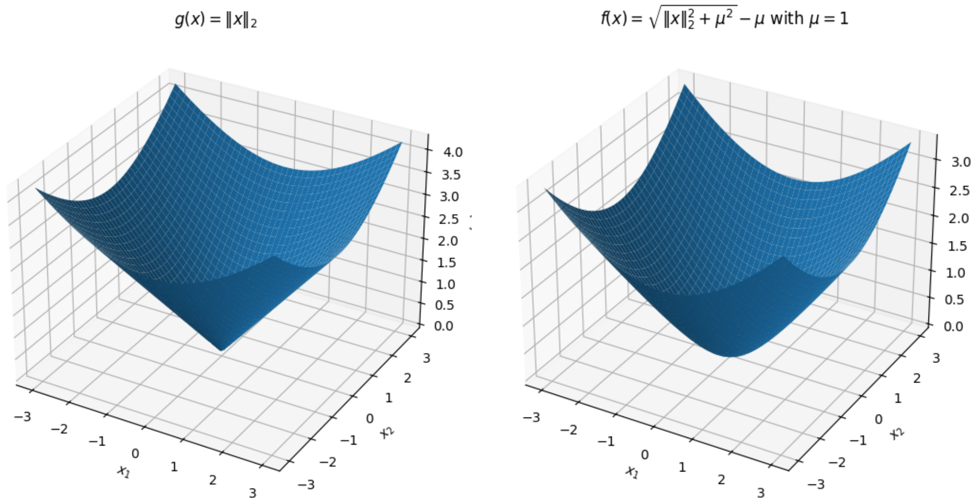

Figure 14.2: Smoothing of the ℓ_2 norm

2.1 Smoothing Lipschitz functions

Suppose that f is a L-Lipschitz convex function for some $L > 0$. In this section, we argue that f is $(1, L^2/2)$-smoothable.

Given a function $f : \mathbb{R}^d \to \mathbb{R}$, the Moreau-Yosida smoothing of f is defined as

$$f_\eta(x) := \inf_u \left\{ f(u) + \frac{1}{2\eta} \|u - x\|_2^2 \right\}$$

for some $\eta > 0$. We will prove that f_η is $(1/\eta)$-smooth. To see that f_η is a right approximation, note that

$$f_\eta(x) \leq f(x) + \frac{1}{2\eta} \|x - x\|_2^2 = f(x).$$

Moreover,

$$\begin{aligned}
f_\eta(x) - f(x) &= \inf_u \left\{ f(u) - f(x) + \frac{1}{2\eta}\|u-x\|_2^2 \right\} \\
&\geq \inf_u \left\{ g_x^\top (u-x) + \frac{1}{2\eta}\|u-x\|_2^2 \right\} \\
&= -\frac{\eta}{2}\|g_x\|_2^2.
\end{aligned}$$

2.2 Smooth optimization of nonsmooth functions

Consider

$$\text{minimize} \quad g(x) + h(x)$$

where g convex and (p,q)-smoothable while h is convex. To solve this, we take the following procedure:

1. We first take the following approximate problem:

$$\text{minimize} \quad g_\mu(x) + h(x)$$

   where g_μ is (p/μ)-smooth and $g_\mu \leq g \leq g_\mu + q\mu$.

2. We apply Nesterov's accelerated gradient descent to minimize $g_\mu + h$.

Theorem 14.3. *Let x_T be the output of the algorithm after $T-1$ iterations. Then*

$$g(x_T) + h(x_T) - \min_x\{g(x) + h(x)\} \leq q\mu + \frac{p}{\mu T^2}\|x_1 - x^*\|_2^2.$$

Proof. Let x^* be a minimizer of $g+h$. Applying Nesterov's algorithm, we have

$$g_\mu(x_T) + h(x_T) - g_\mu(x^*) - h(x^*) \leq \frac{p}{\mu T^2}\|x_1 - x^*\|_2^2.$$

By the choice of g_μ, we have

$$g_\mu(x_T) \geq g(x_T) - q\mu, \quad g_\mu(x^*) \leq g(x^*).$$

This implies that

$$g_\mu(x_T) + h(x_T) - g_\mu(x^*) - h(x^*) \geq g(x_T) - q\mu + h(x_T) - g(x^*) - h(x^*).$$

Hence, it follows that

$$g(x_T) + h(x_T) - g(x^*) - h(x^*) \leq q\mu + \frac{p}{\mu T^2}\|x_1 - x^*\|_2^2,$$

as required. □

Taking $\mu = 1/T$, we have

$$g(x_T) + h(x_T) - \min_x\{g(x) + h(x)\} = O(1/T),$$

which improves the convergnece rate $O(1/\sqrt{T})$ of the subgradient method.

3 Fenchel conjugate

The Fenchel conjugate of a function $f : \mathbb{R}^d \to \mathbb{R}$ is given by

$$f^*(y) = \sup_{x \in \text{dom}(f)} \left\{ y^\top x - f(x) \right\}.$$

As $y^\top x - f(x)$ is linear in y, the conjugate function is always convex, regardless of f.

Lemma 14.4 (Fenchel-Young inequality)**.** *For $x \in dom(f)$ and $y \in dom(f^*)$,*

$$f(x) + f^*(y) \geq y^\top x.$$

Proof. Note that $f^*(y) = \sup_{x \in \text{dom}(f)}(y^\top x - f(x)) \geq y^\top x - f(x)$. □

3.1 Fenchel conjugate examples

Example 14.5. When $f(x) = c^\top x + d$ over $x \in \mathbb{R}^d$,

$$f^*(y) = \sup_{x \in \mathbb{R}^d} (y^\top x - c^\top x - d) = \begin{cases} -d, & \text{if } y = c, \\ +\infty, & \text{otherwise.} \end{cases}$$

Example 14.6. When $f(x) = \log(1 + e^x)$ over $x \in \mathbb{R}$,

$$f^*(y) = \sup_{x \in \mathbb{R}} (yx - \log(1 + e^x)) = \begin{cases} y \log y + (1-y)\log(1-y), & \text{if } 0 < y < 1, \\ 0, & \text{if } y \in \{0, 1\}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Example 14.7. When $f(x) = (1/2)x^\top Qx + p^\top x$ over $x \in \mathbb{R}^d$ for some positive definite Q,

$$f^*(y) = \sup_{x \in \mathbb{R}} \left(y^\top x - \frac{1}{2} x^\top Qx - p^\top x \right).$$

Note that the maximum is attained at $x = Q^{-1}(y - p)$. Therefore,

$$f^*(y) = \frac{1}{2}(y - p)^\top Q^{-1}(y - p).$$

Here,

$$\nabla f^*(y) = Q^{-1}(y - p),$$

which implies that $\nabla f(\nabla f^*(y)) = y$ and

$$\nabla f^*(y) = (\nabla f)^{-1}(y).$$

Example 14.8. When $f(x) = \sum_{i=1}^d x_i \log x_i$ over $x \in \mathbb{R}^d_{++}$,

$$f^*(y) = \sup_{x \in \mathbb{R}^d_{++}} \left(y^\top x - \sum_{i=1}^d x_i \log x_i \right) = \sup_{x \in \mathbb{R}^d_{++}} \left(\sum_{i=1}^d x_i(y_i - \log x_i) \right) = \sum_{i=1}^d e^{y_i - 1}.$$

Example 14.9. When $f(X) = -\log \det X$ over $X \in \mathbb{S}^d_{++}$,

$$f^*(Y) = \sup_{X \in \mathbb{S}^d_{++}} \left(\text{tr}(Y^\top X) + \log \det X \right).$$

It is known that $\nabla \log \det X = X^{-1}$. Then the supremum is attained at $X = -Y^{-1}$, and therefore,

$$f^*(Y) = -d - \log \det(-Y).$$

3.2 Some properties

The following statements hold.

- Let $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$. Then $f^*(y_1, y_2) = f_1^*(y_1) + f_2^*(y_2)$.
- Let $g(x) = f(x) + c^\top x + d$. Then $g^*(y) = f^*(y - c) - d$.
- Let $g(x) = f(x - b)$. Then $g^*(y) = b^\top y + f^*(y)$.
- Let $f(x) = \inf_{u+v=x} \{g(u) + h(v)\}$. Then $f^*(y) = g^*(y) + h^*(y)$.

Lemma 14.10. *For any closed function $f : \mathbb{R}^d \to \mathbb{R}$, its Fenchel conjugate f^* is closed and convex.*

Proof. We have already observed that f^* is convex. Let $h_x : \mathbb{R}^d \to \mathbb{R}$ for any $x \in \mathrm{dom}(f)$ be defined as $h_x(y) = y^\top x - f(x)$. Note that

$$\mathrm{epi}(h_x) = \{(y, t) \in \mathbb{R}^d \times \mathbb{R} : \ t \geq y^\top x - f(x)\}$$

is closed. By definition, we have $f^*(y) = \sup_{x \in \mathrm{dom}(f)} \{h_x(y)\}$, implying in turn that

$$\mathrm{epi}(f^*) = \bigcap_{x \in \mathrm{dom}(f)} \mathrm{epi}(h_x).$$

As the intersection of arbitrarily many closed sets is closed, $\mathrm{epi}(f^*)$ is closed, and therefore, f^* is closed. $\square$

Lemma 14.11. *For any function $f : \mathbb{R}^d \to \mathbb{R}$, we have $f^{**} \leq f$.*

Proof. Let $x \in \mathrm{dom}(f)$. Note that if $x - z \neq 0$, then $\sup_{y \in \mathbb{R}^d} \left\{y^\top (x - z) + f(z)\right\} = +\infty$. If $z = x$, we have $\sup_{y \in \mathbb{R}^d} \left\{y^\top (x - z) + f(z)\right\} = f(x)$. Therefore,

$$f(x) = \inf_{z \in \mathrm{dom}(f)} \sup_{y \in \mathbb{R}^d} \left\{y^\top (x - z) + f(z)\right\}.$$

Note that

$$\begin{aligned}
\inf_{z \in \mathrm{dom}(f)} \sup_{y \in \mathbb{R}^d} \left\{y^\top (x - z) + f(z)\right\} &\geq \sup_{y \in \mathbb{R}^d} \inf_{z \in \mathrm{dom}(f)} \left\{y^\top (x - z) + f(z)\right\} \\
&= \sup_{y \in \mathbb{R}^d} \left\{y^\top x + \inf_{z \in \mathrm{dom}(f)} \left\{-y^\top z + f(z)\right\}\right\} \\
&= \sup_{y \in \mathbb{R}^d} \left\{y^\top x - \sup_{z \in \mathrm{dom}(f)} \left\{y^\top z - f(z)\right\}\right\} \\
&= \sup_{y \in \mathbb{R}^d} \left\{y^\top x - f^*(y)\right\} \\
&\geq \sup_{y \in \mathrm{dom}(f^*)} \left\{y^\top x - f^*(y)\right\} \\
&= f^{**}(x).
\end{aligned}$$

Therefore, $f(x) \geq f^{**}(x)$ for any $x \in \mathrm{dom}(f)$, and thus $f \geq f^{**}$. $\square$

When f is closed and convex, the equality holds, i.e., $f^{**} = f$. To show this, we need the following theorem.

Theorem 14.12 (Strict point-to-convex set separation)**.** *Let $C \subseteq \mathbb{R}^d$ be a closed convex set and $y \notin C$. Then $\inf_{x \in C} \|x - y\| > 0$. Furthermore, there exists $\alpha \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$ such that*

$$\begin{aligned} \alpha^\top x &> \beta \quad \forall x \in C, \\ \alpha^\top y &< \beta. \end{aligned}$$

Lemma 14.13. *For a closed convex function $f : \mathbb{R}^d \to \mathbb{R}$, we have $f^{**} = f$.*

Proof. Next, assume that f is closed and convex. We will show that $\mathrm{epi}(f) = \mathrm{epi}(f^{**})$. As $f \geq f^{**}$, we already know that $\mathrm{epi}(f) \subseteq \mathrm{epi}(f^{**})$. Suppose for a contradiction that there exists $\bar{x}$ such that $(\bar{x}, f^{**}(\bar{x})) \notin \mathrm{epi}(f)$. Then, by Theorem 14.12, there exists $\alpha \in \mathbb{R}^d$, $\gamma \in \mathbb{R}$, and $\beta \in \mathbb{R}$ such that

$$\begin{aligned} \alpha^\top x + \gamma t &> \beta \quad \forall (x,t) \in \mathrm{epi}(f), \\ \alpha^\top \bar{x} + \gamma f^{**}(\bar{x}) &< \beta. \end{aligned}$$

Let $\delta = \beta - \big(\alpha^\top \bar{x} + \gamma f^{**}(\bar{x})\big) > 0$. Then for any $(x,t) \in \mathrm{epi}(f)$,

$$\Big(\alpha^\top x + \gamma t\Big) - \Big(\alpha^\top \bar{x} + \gamma f^{**}(\bar{x})\Big) > \beta - \Big(\alpha^\top \bar{x} + \gamma f^{**}(\bar{x})\Big) = \delta > 0.$$

Here, t can be arbitrarily large with $(x,t) \in \mathrm{epi}(f)$, so $\gamma \geq 0$. Suppose that $\gamma = 0$. Let ϵ be a sufficiently small number and $\bar{y} \in \mathrm{dom}(f^*)$. Now consider

$$\Big((\alpha - \epsilon\bar{y})^\top x + \epsilon t\Big) - \Big((\alpha - \epsilon\bar{y})^\top \bar{x} + \epsilon f^{**}(\bar{x})\Big) > \delta - \epsilon(\bar{y}^\top x - t + \bar{y}^\top \bar{x} + f^{**}(\bar{x})).$$

$$\begin{aligned} &\inf_{(x,t)\in \mathrm{epi}(f)} \Big\{\Big((\alpha - \epsilon\bar{y})^\top x + \epsilon t\Big) - \Big((\alpha - \epsilon\bar{y})^\top \bar{x} + \epsilon f^{**}(\bar{x})\Big)\Big\} \\ &\geq \inf_{(x,t)\in \mathrm{epi}(f)} \Big\{\delta - \epsilon(\bar{y}^\top x - t + \bar{y}^\top \bar{x} + f^{**}(\bar{x}))\Big\} \\ &\geq \inf_{x\in \mathrm{dom}(f)} \Big\{\delta - \epsilon(\bar{y}^\top x - f(x) + \bar{y}^\top \bar{x} + f^{**}(\bar{x}))\Big\} \\ &= \delta - \epsilon(f^*(\bar{y}) - \bar{y}^\top \bar{x} + f^{**}(\bar{x})). \end{aligned}$$

Making ϵ sufficiently small, we have

$$\inf_{(x,t)\in \mathrm{epi}(f)} \Big\{\Big((\alpha - \epsilon\bar{y})^\top x + \epsilon t\Big) - \Big((\alpha - \epsilon\bar{y})^\top \bar{x} + \epsilon f^{**}(\bar{x})\Big)\Big\} > 0.$$

Therefore, we have just argued that there exists $\alpha \in \mathbb{R}^d$, $\gamma \in \mathbb{R}$, and $\delta > 0$ such that $\gamma > 0$ and

$$\inf_{(x,t)\in \mathrm{epi}(f)} \Big\{\Big(\alpha^\top x + \gamma t\Big) - \Big(\alpha^\top \bar{x} + \gamma f^{**}(\bar{x})\Big)\Big\} \geq \delta > 0.$$

Then

$$\inf_{(x,t)\in \mathrm{epi}(f)} \Big\{(\alpha/\gamma)^\top (x - \bar{x}) + t - f^{**}(\bar{x})\Big\} \geq \delta/\gamma > 0.$$

Note that

$$\begin{aligned}
\inf_{(x,t)\in \mathrm{epi}(f)} \left\{ (\alpha/\gamma)^\top (x-\bar{x}) + t - f^{**}(\bar{x}) \right\} &= \inf_{x\in \mathrm{dom}(f)} \left\{ (\alpha/\gamma)^\top (x-\bar{x}) + f(x) - f^{**}(\bar{x}) \right\} \\
&= (-\alpha/\gamma)^\top \bar{x} - f^{**}(\bar{x}) - \sup_{x\in \mathrm{dom}(f)} \left\{ (-\alpha/\gamma)^\top x - f(x) \right\} \\
&= (-\alpha/\gamma)^\top \bar{x} - f^{**}(\bar{x}) - f^*(-\alpha/\gamma) \\
&\le (-\alpha/\gamma)^\top \bar{x} - (-\alpha/\gamma)^\top \bar{x} \\
&= 0
\end{aligned}$$

where the inequality follows from the Fenchel-Young inequality. □

3.3 Moreau decomposition

Remember that for a quadratic function with a positive definite matrix given by

$$f(x) = \frac{1}{2} x^\top Q x + p^\top x,$$

we have $\nabla f^*(y) = (\nabla f)^{-1}(y)$. This is implies that if $y = \nabla f(x)$, then $x = \nabla f^*(y)$. In general, the subdifferential of the conjugate is the inverse of the subdifferential.

Theorem 14.14. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a closed and convex function. Then the following statements are equivalent.*

(i) $y \in \partial f(x)$,

(ii) $x \in \partial f^*(y)$,

(iii) $y^\top x = f(x) + f^*(y)$.

Proof. Assume that $\bar{y} \in \partial f(\bar{x})$. Then $\bar{x} \in \mathrm{dom}(f)$ and $0 \in -\bar{y} + \partial f(\bar{x})$. Consider

$$f^*(\bar{y}) = \sup_{x\in \mathrm{dom}(f)} (\bar{y}^\top x - f(x)) = -\inf_{x\in \mathrm{dom}(f)} (-\bar{y}^\top x + f(x)).$$

Since $0 \in -\bar{y} + \partial f(\bar{x})$, $\bar{x}$ is the minimizer, and therefore,

$$f^*(\bar{y}) = -(-\bar{y}^\top \bar{x} + f(\bar{x})) = \bar{y}^\top \bar{x} - f(\bar{x}).$$

Hence, $\bar{y} \in \mathrm{dom}(f^*)$. Again, the definition of $f^*(y)$ implies that for any $y \in \mathrm{dom}(f^*)$,

$$f^*(y) \ge y^\top \bar{x} - f(\bar{x}) = (y - \bar{y})^\top \bar{x} + f^*(\bar{y}).$$

Therefore, $\bar{x}$ is a subgradient of f^* at $\bar{y}$, and thus $\bar{x} \in \partial f^*(\bar{y})$. Hence, we have just proved the direction $(i) \to (iii) \to (ii)$. Since f is closed and convex, f^* is closed and convex and $f = f^{**}$. Then, by symmetry, we can also argue that $(ii) \to (iii) \to (i)$. Therefore, (i), (ii), and (iii) are all equivalent. □

Using the theorem, we can show the following result.

Theorem 14.15 (Moreau decomposition)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a closed convex function. Then*

$$x = \mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x).$$

Proof. Let $u = \mathrm{prox}_f(x)$, then $x - u \in \partial f(u)$. This implies that $u \in \partial f^*(x-u)$. Let $v = x - u$. Then we have $x - v \in \partial f^*(v)$, implying in turn that $v = \mathrm{prox}_{f^*}(x)$. Therefore,

$$\mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x) = u + v = u + x - u = x,$$

as required. $\square$

Example 14.16. Let $V \subseteq \mathbb{R}^d$ be a linear subspace, and let $f = I_V : \mathbb{R}^d \to \mathbb{R}$ be the indicator function of U. Note that

$$f^*(y) = \sup_{x \in V} \left\{ y^\top x \right\} = I_{V^\perp}(y).$$

Then

$$\mathrm{prox}_f(x) = \operatorname*{argmin}_{u \in \mathbb{R}^d} \left\{ I_V(u) + \frac{1}{2}\|u - x\|_2^2 \right\} = \mathrm{proj}_V(x),$$

and

$$\mathrm{prox}_{f^*}(x) = \operatorname*{argmin}_{u \in \mathbb{R}^d} \left\{ I_{V^\perp}(u) + \frac{1}{2}\|u - x\|_2^2 \right\} = \mathrm{proj}_{V^\perp}(x).$$

Therefore, the Moreau decomposition theorem states that

$$x = \mathrm{proj}_V(x) + \mathrm{proj}_{V^\perp}(x).$$