12 Duality

12.1 Conjugate function

Definition 12.1 (Conjugate function)**.** The *Fenchel conjugate* of a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined as

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{\langle y, x\rangle - f(x)\}.$$

Observe that f^* is always convex and l.s.c. (i.e., $\mathbf{epi}(f^*)$ is closed) since it is a supremum of linear functions.

Note that for any y, we have a lower bound on f, namely $\langle y, x\rangle - f^*(y) \leq f(x)$. Taking the supremum over y tells us that $f^{**}(x) \leq f(x)$. The next theorem tells us that we actually have equality when f is convex and lower semi-continuous.

Theorem 12.1 (Biduality)**.** *If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex and l.s.c. then $f^{**} = f$.*

Sketch of proof. We will show that $\mathbf{epi}(f) = \mathbf{epi}(f^{**})$. The inclusion $\subseteq$ already follows from $f^{**} \leq f$. To prove the reverse inclusion assume $(\bar{x}, \bar{t}) \notin \mathbf{epi}(f)$. We will show that $(\bar{x}, \bar{t}) \notin \mathbf{epi}(f^{**})$. Since $\mathbf{epi}(f)$ is convex and closed (since f is l.s.c.), the separating hyperplane theorem tells us there is $(a, b) \in \mathbb{R}^n \times \mathbb{R} \setminus \{0\}$ such that

$$\begin{cases} \langle a, \bar{x}\rangle - b\bar{t} \geq c + \delta > c \\ \langle a, x\rangle - bt \leq c - \delta < c \quad \forall (x, t) \in \mathbf{epi}(f). \end{cases} \tag{1}$$

Letting $t \to +\infty$ in the second line above tells us that $b \geq 0$.

We can further assume that $b \neq 0$ by perturbing the hyperplane slightly: indeed, assuming $b = 0$, let x_0 be any point where f has a subgradient $g \in \partial f(x_0)$. Then for any $\epsilon > 0$, we can write

$$\begin{cases} \langle a + \epsilon g, \bar{x}\rangle - \epsilon\bar{t} \geq c + \delta + \epsilon(\langle g, \bar{x}\rangle - \bar{t}) = c_1(\epsilon) \\ \forall (x, t) \in \mathbf{epi}(f) \ : \ \langle a + \epsilon g, x\rangle - \epsilon t \leq c - \delta + \epsilon(\langle g, x\rangle - t) \leq c - \delta + \epsilon(\langle g, x_0\rangle - f(x_0)) = c_2(\epsilon). \end{cases}$$

We want $\epsilon > 0$ such that $c_1(\epsilon) > c_2(\epsilon)$, i.e., $2\delta + \epsilon(f(x_0) + \langle g, \bar{x} - x_0\rangle - \bar{t}) > 0$ which can be achieved for small enough ϵ.

Assume thus that $b > 0$: We can assume wlog that $b = 1$. Putting $t = f(x)$ in the second line of (1) tells us that $\langle a, x\rangle - f(x) < c$ for all $x \in \mathbf{dom}(f)$ which implies, $f^*(a) \leq c$. In turn this means that $f^{**}(\bar{x}) \geq \langle a, \bar{x}\rangle - f^*(a) \geq \langle a, \bar{x}\rangle - c > \bar{t}$ where in the last inequality we used (1). This shows that $(\bar{x}, \bar{t}) \notin \mathbf{epi}(f^{**})$ as desired. □

Lemma 1 (Subgradients)**.** *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and lower semicontinuous. For any $x \in \mathbf{dom}(f)$ and $y \in \mathbf{dom}(f^*)$ we have*

$$f^*(y) = \langle y, x\rangle - f(x) \iff y \in \partial f(x) \iff x \in \partial f^*(y). \tag{2}$$

Proof. Fix y. The vector $x \in \mathbb{R}^n$ minimizes the convex function $\xi \mapsto f(\xi) - \langle y, \xi\rangle$ iff the zero element is in the subdifferential at $\xi = x$. This tells us that $f^*(y) = \langle y, x\rangle - f(x)$ iff $y \in \partial f(x)$, which is the first equivalence.

We now show $y \in \partial f(x) \implies x \in \partial f^*(y)$. This is immediate since if $y \in \partial f(x)$ then for any z we have $f^*(z) \geq \langle z, x\rangle - f(x) = f^*(y) + \langle z - y, x\rangle$ which means that $x \in \partial f^*(y)$. The reverse inclusion $x \in \partial f^*(y) \implies y \in \partial f(x)$ follows from $f^{**} = f$. □

Theorem 12.2 (Smoothness of f^*)**.** *Assume $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex and lower semicontinuous. If f is m-strongly convex with respect to $\|\cdot\|$, then $\mathbf{dom}(f^*) = \mathbb{R}^n$ and f^* is differentiable with*

$$\nabla(f^*)(y) = \operatorname*{argmax}_{x\in\mathbb{R}^n} \{\langle y, x\rangle - f(x)\}.$$

Moreover, $\nabla(f^*)$ is $(1/m)$-Lipschitz with respect to $\|\cdot\|$, i.e., $\|\nabla f^*(y_1) - \nabla f^*(y_2)\| \leq \|y_1 - y_2\|_*$, where $\|\cdot\|_*$ is the dual norm

Proof. Since f is strongly convex, the function $x \mapsto f(x) - \langle y, x\rangle$ is bounded from below for any $y \in \mathbb{R}^n$, and thus its infimum is $> -\infty$. This shows that $f^*(y)$ is defined for all y. Since this function is lower semicontinuous and strongly convex, its infimum is attained at a unique point. By (2) we get that $\partial(f^*)(y)$ consists of this unique infimizer, i.e., f^* is smooth and

$$\nabla(f^*)(y) = \operatorname*{argmax}_{x\in\mathbb{R}^n} \{\langle y, x\rangle - f(x)\} = x^*(y).$$

The function $x \mapsto f(x) - \langle y_1, x\rangle$ is m-strongly convex, and so we have, for any x

$$f(x) - \langle y_1, x\rangle \geq f(x^*(y_1)) - \langle y_1, x^*(y_1)\rangle + (m/2)\|x - x^*(y_1)\|^2.$$

$$f(x) - \langle y_2, x\rangle \geq f(x^*(y_2)) - \langle y_2, x^*(y_2)\rangle + (m/2)\|x - x^*(y_2)\|^2.$$

The two combined tell us that

$$m\|x^*(y_1) - x^*(y_2)\|^2 \leq \langle y_1 - y_2, x^*(y_1) - x^*(y_2)\rangle \leq \|y_1 - y_2\|_*\|x^*(y_1) - x^*(y_2)\|$$

i.e.,

$$\|x^*(y_1) - x^*(y_2)\| \leq (1/m)\|y_1 - y_2\|_*.$$

□

12.2 Lagrangian duality

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, and consider the following minimization problem with an affine constraint:

$$\min_{x\in\mathbb{R}^n} \quad f(x) \quad \text{subject to} \quad Ax = b, \tag{3}$$

where $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, and $b \in \mathbb{R}^m$. The *Lagrangian* of this problem is defined as

$$L(x, z) = f(x) + \langle z, b - Ax\rangle$$

where z is the dual variable for the constraint $Ax = b$. The *dual function* is

$$g(z) = \min_{x\in\mathbb{R}^n} L(x, z).$$

For any $z \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ such that $Ax = b$, we clearly have $g(z) \leq f(x)$ since for such x we have $L(x, z) = f(x)$. Thus $g(z)$ gives a *lower bound* on the optimal value of (3). The best such lower bound can be obtained by maximizing $g(z)$. This is the *dual problem*:

$$\max_{z\in\mathbb{R}^n} g(z). \tag{4}$$

Note that

$$\begin{aligned} g(z) = \min_{x\in\mathbb{R}^n} L(x,z) &= \langle b, z\rangle + \min_{x\in\mathbb{R}^n} -\left\langle A^T z, x\right\rangle + f(x) \\ &= \langle b, z\rangle - \max_{x\in\mathbb{R}^n} \left\langle A^T z, x\right\rangle - f(x) \\ &= \langle b, z\rangle - f^*(A^T z). \end{aligned}$$

We note that $g(z)$ is concave in z, and thus the dual problem (4) is a concave maximization problem.

Let $p^* = \min\{f(x) : Ax = b\}$ and let $d^* = \max_{z\in\mathbb{R}^n} g(z)$. Note that $g(z) \leq p^*$ for all z so that $d^* \leq p^*$.

Theorem 12.3 (Strong duality)**.** *Assume f is convex and that $\{x : Ax = b\} \cap \mathbf{int}(\mathbf{dom}\, f)$ is not empty (Slater's condition). Then $p^* = d^*$.*

Proof. We prove the theorem in the special case where the solution of (3) is attained. Let $\bar{x}$ be the optimal solution of (3). Letting $C = \{x : Ax = b\}$, then necessarily we have $0 \in \partial(f + I_C)(\bar{x})$. Since C is polyhedral, and $C \cap \mathbf{int}\,\mathbf{dom}\, f \neq \emptyset$, we have $\partial(f + I_C)(\bar{x}) = \partial f(\bar{x}) + N_C(\bar{x}) = \partial f(\bar{x}) + \mathbf{im} A^T$. So we can write $0 \in \partial f(\bar{x}) - A^T \bar{z}$ for some $\bar{z}$. Then note that for any $x \in \mathbb{R}^n$ we have

$$L(x, \bar{z}) = f(x) + \langle \bar{z}, b - Ax\rangle \geq f(\bar{x}) + \left\langle A^T \bar{z}, x - \bar{x}\right\rangle + \langle \bar{z}, b - Ax\rangle = f(\bar{x}).$$

Hence $g(\bar{z}) = \min_{x\in\mathbb{R}^n} L(x, \bar{z}) \geq f(\bar{x})$. Thus we get

$$d^* \geq g(\bar{z}) \geq f(\bar{x}) = p^*$$

as desired. □