8 Conjugate functions

Theorem 8.1 (Separating hyperplane theorem)**.** *Let $C \subset \mathbb{R}^n$ convex and assume $z \notin C$. Then there exists $y \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$ such that*

$$\begin{cases} y^T z \geq b \\ y^T x \leq b \qquad \forall x \in C. \end{cases} \tag{1}$$

If C is closed, then y and b can be chosen so that inequalities in (1) are strict.

Definition 8.1 (Conjugate function)**.** Given a function $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}^n$, the conjugate of f is defined as

$$f^*(y) = \sup_{x \in D} \; y^T x - f(x).$$

Note that for any y, we have a lower bound on f, namely $y^T x - f^*(y) \leq f(x)$, $\forall x \in D$. Maximizing 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 closed (we say that f is closed when $\text{epi}(f)$ is closed).

Theorem 8.2 (Biduality)**.** *If $f : D \to \mathbb{R}$ is convex and $\text{epi}(f) := \{(x,t) \in D \times \mathbb{R} : t \geq f(x)\}$ is closed, then $f^{**} = f$.*

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

$$\begin{cases} a^T \bar{x} - b\bar{t} > c \\ a^T x - bt < c \quad \forall (x,t) \in \text{epi}(f). \end{cases} \tag{2}$$

Letting $t \to +\infty$ in the second line above tells us that $b > 0$. We assume wlog that $b = 1$. Putting $t = f(x)$ in the second line of (2) tells us that $a^T x - f(x) < c$ for all $x \in D$ which implies, $f^*(a) \leq c$. In turn this means that $f^{**}(\bar{x}) \geq a^T \bar{x} - f^*(a) \geq a^T \bar{x} - c > \bar{t}$ where in the last inequality we used (2). This shows that $(\bar{x}, \bar{t}) \notin \text{epi}(f^{**})$ as desired. □

Lemma 1 (Subgradients)**.** *Let $f : D \to \mathbb{R}$ be convex and closed (i.e., $\text{epi}(f)$ is closed). For any $x \in D$ and y we have*

$$f^*(y) = y^T x - f(x) \;\Leftrightarrow\; y \in \partial f(x) \;\Leftrightarrow\; x \in \partial f^*(y). \tag{3}$$

Proof. Fix y. The vector $x \in D$ maximizes the function $\xi \mapsto y^T \xi - f(\xi)$ iff the zero element is in the subdifferential at $\xi = x$. This tells us that $f^*(y) = y^T x - f(x)$ iff $y \in \partial f(x)$, which is the first equivalence.

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

Theorem 8.3 (Smoothness of f^*)**.** *Assume $f : D \to \mathbb{R}$ is closed and m-strongly convex function. Then f^* is defined everywhere on $\mathbb{R}^n$, smooth, and for any $y \in \mathbb{R}^n$ we have*

$$\nabla f^*(y) = \operatorname*{argmax}_{x \in D} y^T x - f(x).$$

(The argmax has a unique solution.) Furthermore ∇f^* is $(1/m)$-Lipschitz wrt $\| \cdot \|_2$.

Proof. If f is closed and strongly convex then for any fixed y the function $x \mapsto y^T x - f(x)$ has a unique maximizer, $x^*(y) = \operatorname{argmax}_{x \in D} y^T x - f(x)$. Since the maximizer is unique, (3) tells us that $\partial f^*(y) = \{x^*(y)\}$. In other words this means that f^* is smooth at y and $\nabla f^*(y) = x^*(y)$.

For the last statement: we use the fact that for any strongly convex function $\phi(u)$ we have $\phi(u) \geq \phi(u^*) + (m/2)\|u - u^*\|_2^2$ where $u^* = \operatorname{argmin} \phi(u)$. Using this inequality for the strongly convex function $x \mapsto f(x) - y^T x$ gives us $f(x^*(z)) - y^T x^*(z) \geq f(x^*(y)) - y^T x^*(y) + (m/2)\|x^*(y) - x^*(z)\|_2^2$. Using the similar inequality with $x \mapsto f(x) - z^T x$ and adding up gives us

$$m\|x^*(y) - x^*(z)\|_2^2 \leq (x^*(z) - x^*(y))^T (z - y) \leq \|x^*(z) - x^*(y)\|_2 \|z - y\|_2$$

which is what we wanted. ☐

Examples

- If $f(x) = \frac{1}{2} x^T A x + b^T x$ with A positive definite, then $f^*(y) = \frac{1}{2}(y - b)^T A^{-1} (y - b) - c$
- If $f(x) = \|x\|$ for some norm $\| \cdot \|$, then $f^*(y)$ is the indicator function of the unit ball for the dual norm, i.e.,

  $$f^*(y) = \begin{cases} 0 & \text{if } \|y\|_* \leq 1 \\ +\infty & \text{else} \end{cases}$$

  where

  $$\|y\|_* = \sup_{\|x\| = 1} y^T x.$$

  On $\mathbb{R}^n$, the dual norm of $\|x\|_p = (\sum_i x_i^p)^{1/p}$ (for $p \geq 1$) is $\| \cdot \|_{p'}$ where $1/p + 1/p' = 1$ (dual of ℓ_1 norm is ℓ_∞ norm).