1 Outline

In this lecture, we study

- Fenchel duality.
- Fenchel conjugate.

2 Fenchel conjugate

2.1 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 18.1. *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. □

Lemma 18.2. *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 $f\geq f^{**}$. □

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

Theorem 18.3 (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 18.4. *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 18.3, 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 $\delta=\beta-\left(\alpha^\top\bar{x}+\gamma f^{**}(\bar{x})\right)>0$. Then for any $(x,t)\in\mathrm{epi}(f)$,

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

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

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

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

Making ϵ sufficiently small, we have

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

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

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

Then

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

Note that

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

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

2.2 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 18.5. *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 $y \in \mathrm{dom}(f^*)$,

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

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 18.6 (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. □

Example 18.7. 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

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

2.3 Fenchel dual

Consider the following composite optimization problem.

$$\text{minimize} \quad f(x) + g(Ax) \tag{18.1}$$

for some matrix $A \in \mathbb{R}^{m \times d}$. This problem is equivalent to

$$\begin{aligned} \text{minimize} &\quad f(x) + g(y) \\ \text{subject to} &\quad y = Ax. \end{aligned} \tag{18.2}$$

Then the Lagrangian dual function is given by

$$\begin{aligned} \inf_{x,y} \quad f(x) + g(y) + \mu^\top (Ax - y) &= -\sup_{x,y} \left\{ -f(x) - g(y) + \mu^\top(-Ax + y) \right\} \\ &= -\sup_{x,y} \left\{ (-A^\top \mu)^\top x - f(x) + \mu^\top y - g(y) \right\} \\ &= -\sup_{x} \left\{ (-A^\top \mu)^\top x - f(x) \right\} - \sup_{x} \left\{ \mu^\top y - g(y) \right\} \\ &= -f^*(-A^\top \mu) - g^*(\mu). \end{aligned}$$

Therefore, the Lagrangian dual problem is given by

$$\text{maximize} \quad -f^*(-A^\top \mu) - g^*(\mu).$$

Moreover, note that (18.2) is linearly constrained. If f and g are convex, then Slater's condition holds (assuming $\text{dom}(f) = \mathbb{R}^d$ and $\text{dom}(g) = \mathbb{R}^m$), in which case, strong duality holds. Therefore,

$$\begin{aligned} \text{minimize} \quad f(x) + g(Ax) &= \min_{x,y} \max_{\mu} f(x) + g(y) + \mu^\top (Ax - y) \\ &= \max_{\mu} \min_{x,y} f(x) + g(y) + \mu^\top (Ax - y) \\ &= \text{maximize} \quad -f^*(-A^\top \mu) - g^*(\mu). \end{aligned}$$

Example 18.8. Given a convex set C, consider

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax - b \in C. \end{aligned}$$

Using the indicator function, it is equivalent to

$$\text{minimize} \quad f(x) + I_C(Ax - b).$$

We can set $g(y) = I_C(y - b)$. Then

$$g^*(\mu) = \sup_{u - b \in C} \left\{ \mu^\top u \right\} = \sup_{u \in C} \left\{ \mu^\top (u + b) \right\} = b^\top \mu + I_C^*(\mu).$$

Hence, the Fenchel dual is given by

$$\text{maximize} \quad -b^\top \mu - f^*(-A^\top \mu) - I_C^*(\mu).$$

Example 18.9. Consider

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b. \end{aligned}$$

The constraint is equivalent to $Ax - b \in \{0\}$. Since $\{0\}$ is a trivial vector space and $(\{0\})^\perp = \mathbb{R}^d$, we have that $I_{\{0\}}^*(y) = 0$ for any $y \in \mathbb{R}^d$. Then the corresponding dual is

$$\text{maximize} \quad -b^\top \mu - f^*(-A^\top \mu).$$

Example 18.10. Consider

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & \|Ax - b\| \leq 1 \end{aligned}$$

The constraint is equivalent to $Ax - b \in C = \{y : \|y\| \leq 1\}$. Note that

$$I_C^*(\mu) = \sup_{\|y\| \leq 1} \mu^\top y = \|\mu\|_*.$$

In this case, the Fenchel dual is given by

$$\text{maximize} \quad -b^\top \mu - f^*(-A^\top \mu) - \|\mu\|_*.$$

Example 18.11. Consider

$$\text{minimize} \quad f(x) + \|x\|$$

for some $\lambda > 0$. Here, define $g(y) = \|y\|$. Note that

$$g^*(\mu) = \sup_u \left\{ \mu^\top u - \|u\| \right\} = I_C(\mu)$$

where $C = \{u : \|u\|_* \leq 1\}$. Then the corresponding dual is

$$\begin{aligned} \text{maximize} \quad & -f^*(-\mu) \\ \text{subject to} \quad & \|\mu\|_* \leq 1. \end{aligned}$$

3 Dual gradient method

We consider

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b. \end{aligned}$$

We observed that its dual is given by

$$\text{maximize} \quad -f^*(-A^\top \mu) - b^\top \mu.$$

Then the problem is equivalent to

$$(-1) \quad \times \quad \text{minimize} \quad f^*(-A^\top \mu) + b^\top \mu.$$

As f^* is convex, this dual forumulation is a convex minimization problem. Let us apply the subgradient method to the dual.

Given μ_t, let $g_t \in \partial \left(f^*(-A^\top \mu_t) + b^\top \mu_t \right)$. Then the subgradient method applies the following update rule.

$$\mu_{t+1} = \mu_t - \eta_t g_t.$$

Here, what is a subgradient g_t? Note that

$$\underbrace{\partial \left(f^*(-A^\top \mu_t) + b^\top \mu_t \right)}_{\text{subdifferential of } f^*(-A^\top \mu) + b^\top \mu \text{ at } \mu = \mu_t} = -A \underbrace{\partial f^*(-A^\top \mu_t)}_{\text{subdifferential of } f^*(\mu) \text{ at } \mu = -A^\top \mu_t} + b.$$

Hence, $g_t \in \partial \left(f^*(-A^\top \mu_t) + b^\top \mu_t \right)$ if and only if

$$g_t \in -A \partial f^*(-A^\top \mu_t) + b.$$

Therefore,

$$g_t = -Ax_t + b \quad \text{for some } x_t \in \partial f^*(-A^\top \mu_t).$$

Moreover, we have also observed that $x_t \in \partial f^*(-A^\top \mu_t)$ if and only if $-A^\top \mu_t \in \partial f(x_t)$. Here, $-A^\top \mu_t \in \partial f(x_t)$ holds if and only if $0 \in \partial f(x_t) + A^\top \mu_t$ which is equivalent to

$$x_t \in \operatorname*{argmin}_x f(x) + \mu_t^\top Ax.$$

Note that $\mu_t^\top b$ remains constant as x changes, so $x_t \in \operatorname{argmin}_x f(x) + \mu_t^\top Ax$ is equivalent to

$$x_t \in \operatorname*{argmin}_x f(x) + \mu_t^\top (Ax - b).$$

Therefore, the subgradient method applied to the dual problem proceeds with

$$\begin{aligned} &x_t \in \operatorname*{argmin}_x f(x) + \mu_t^\top (Ax - b), \\ &\mu_{t+1} = \mu_t + \eta_t (Ax_t - b). \end{aligned}$$

Here, $f(x) + \mu_t^\top (Ax - b)$ is the Lagrangian function $\mathcal{L}(x, \mu)$ at $\mu = \mu_t$. In words, the subgradient method applied to the dual problem works as follows. At each iteration t with a given dual multiplier μ_t, we find a minimizer of the Lagrangian function $\mathcal{L}(x, \mu_t)$. Then we use the corresponding dual subgradient $Ax_t - b$ to obtain a new multiplier μ_{t+1}.

Algorithm 1 Subgradient method for the dual problem

```
Initialize μ_1.
for t = 1, ..., T − 1 do
    Obtain x_t ∈ argmin_x f(x) + μ_t^⊤(Ax − b),
    Update μ_{t+1} = μ_t + η_t(Ax_t − b) for a step size η_t > 0.
end for
```

At each iteration, we find a minimizer of the Lagrangian function $\mathcal{L}(x, \mu_t)$, which gives rise to an unconstrained optimization problem. Hence, the dual approach is useful when there is a complex system of constraints.