1 Outline

In this lecture, we study

- Lagrangian duality,
- Fenchel duality,
- the dual subgradient method.

2 Lagrangian Duality

In this lecture, we consider problems of the following structure:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0 \quad \text{for } i = 1, \ldots, m \\ & h_j(x) = 0 \quad \text{for } j = 1, \ldots, \ell. \end{aligned} \tag{12.1}$$

We consider the most general setting for which we do not impose the condition that the objective and constraint functions are convex. We may define vector-valued functions $g : \mathbb{R}^d \to \mathbb{R}^m$ and $h : \mathbb{R}^d \to \mathbb{R}^\ell$ such that

- $g(x) = (g_1(x), \ldots, g_m(x))^\top$,
- $h(x) = (h_1(x), \ldots, h_\ell(x))^\top$.

Then (12.1) can be written as

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) \leq 0 \\ & h(x) = 0. \end{aligned} \tag{12.2}$$

The *Lagrangian function* of (12.1) is given by

$$\begin{aligned} \mathcal{L}(x, \lambda, \mu) &= f(x) + \lambda^\top g(x) + \mu^\top h(x) \\ &= f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^\ell \mu_j h_j(x). \end{aligned}$$

When the objective function f is convex, constraint functions $g_1, \ldots, g_m$ are convex, constraint functions $h_1, \ldots, h_\ell$ are affine, and the multiplier $\lambda \geq 0$, the Lagrangian function is convex in x for any fixed λ and μ. Moreover, the Lagrangian function is affine in λ and μ for any fixed x.

The *Lagrangian dual function* of (12.1) is

$$q(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu) = \inf_x \left\{ f(x) + \lambda^\top g(x) + \mu^\top h(x) \right\}.$$

Notice that the Lagrangian dual function is concave in (λ, μ), regardless of f, $g_1, \ldots, g_m$, and $h_1, \ldots, h_\ell$. This is because $\mathcal{L}(x, \lambda, \mu)$ is affine in λ and μ for any fixed x, and $q(\lambda, \mu)$ is a point-wise minimum of affine functions.

Proposition 12.1. *Let x be a feasible solution to (12.1), and $\lambda \geq 0$. Then*

$$f(x) \geq q(\lambda, \mu).$$

Proof. Since x is feasible, $g_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_j(x) = 0$ for $j = 1, \ldots, \ell$. Then for any $\lambda \geq 0$, we have

$$\sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x) \leq 0.$$

This implies that

$$f(x) \geq \mathcal{L}(x, \lambda, \mu).$$

Note that

$$q(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu) \leq \mathcal{L}(x, \lambda, \mu).$$

Therefore, $f(x) \geq q(\lambda, \mu)$. □

By Proposition 12.1, if (12.1) is unbounded below, the Lagrangian dual function $q(\lambda, \mu) = -\infty$ for any $\lambda \geq 0$.

With the Lagrangian dual function, we can provide a lower bound on the problem (12.1). The *Lagrangian dual problem* is defined as

$$\begin{aligned} \text{maximize} \quad & q(\lambda, \mu) \\ \text{subject to} \quad & \lambda \geq 0. \end{aligned} \tag{12.3}$$

We often call (12.1) as *primal* and (12.3) as the *associated (Lagrangian) dual*. The following result states that the optimal value of the primal is lower bounded by the optimal value of the dual.

Theorem 12.2 (Weak duality)**.** *Consider the problem (12.1) and the associated Lagrangian dual problem (12.3). Then the following statement holds.*

$$\min_{x \in C} f(x) \geq \max_{\lambda \geq 0} q(\lambda, \mu)$$

where $C = \{x : g_i(x) \leq 0 \text{ for } i = 1, \ldots, m,\ h_j(x) = 0 \text{ for } j = 1, \ldots, \ell\}$.

Proof. By proposition 12.1, we know that $f(x) \geq q(\lambda, \mu)$ for any $x \in C$ and $\lambda \geq 0$. Then taking the minimum of $f(x)$ over $x \in C$, it follows that $\min_{x \in C} f(x) \geq q(\lambda, \mu)$. Then taking the maximum of $q(\lambda, \mu)$ over $\lambda \geq 0$, we obtain the desired inequality. □

Theorem 12.2 holds regardless of whether the objective and constraint functions are convex or not. Then our next question is whether the equality holds. To answer this, we define the notion of *Slater's condition*.

Definition 12.3 (Slater's condition)**.** Suppose that $g_1, \ldots, g_k$ are affine and $g_{k+1}, \ldots, g_m$ are convex functions that are not affine. Then we say that the problem (12.1) satisfies Slater's condition if there exists a solution $\bar{x}$ such that

$$g_i(\bar{x}) \leq 0 \text{ for } i = 1, \ldots, k, \quad g_i(\bar{x}) < 0 \text{ for } i = k+1, \ldots, m, \quad h_j(\bar{x}) = 0 \text{ for } j = 1, \ldots, \ell.$$

If we assume that the objective f is convex and the constraint functions satisfy Slater's condition, then the inequality given in Theorem 12.2 holds with equality.

Theorem 12.4 (Strong duality). *Consider the primal problem (12.1) and the associated Lagrangian dual problem (12.3). Assume that the objective function f and the constraint functions $g_1, \ldots, g_m$ are convex, and $h_1, \ldots, h_\ell$ are affine. If the primal problem (12.1) has a finite optimal value and Slater's condition, given in Definition 12.3, is satisfied, then there exist $\lambda^* \geq 0$ and μ^* such that*

$$\min_{x \in C} f(x) = q(\lambda^*, \mu^*) = \max_{\lambda \geq 0} q(\lambda, \mu)$$

where $C = \{x : g_i(x) \leq 0 \text{ for } i = 1, \ldots, m,\ h_j(x) = 0 \text{ for } j = 1, \ldots, \ell\}$.

2.1 Example

Consider the following linear program in standard form.

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0. \end{aligned} \tag{12.4}$$

Then the Lagrangian dual function is given by

$$\begin{aligned} q(\lambda, \mu) &= \inf_x \mathcal{L}(x, \lambda, \mu) \\ &= \inf_x \left\{ c^\top x - \lambda^\top x + \mu^\top (Ax - b) \right\} \\ &= -b^\top \mu + \inf_x \left\{ (c - \lambda + A^\top \mu)^\top x \right\}. \end{aligned}$$

Note that $\inf_x \left\{ (c - \lambda + A^\top \mu)^\top x \right\} = -\infty$ unless $c - \lambda + A^\top \mu = 0$. Hence, to maximize the Lagrangian dual function $q(\lambda, \mu)$, it is sufficient to consider (λ, μ) satisfying $c - \lambda + A^\top \mu = 0$. Therefore, the associated Lagrangian dual problem is equivalent to

$$\begin{aligned} \text{maximize} \quad & -b^\top \mu \\ \text{subject to} \quad & c - \lambda + A^\top \mu = 0, \\ & \lambda \geq 0. \end{aligned} \tag{12.5}$$

In fact, we can eliminate the variable from the constraints $c + A^\top \mu \geq \lambda$ and $\lambda \geq 0$, and they can be equivalently written as $c + A^\top \mu \geq 0$. Moreover, the variables μ are unrestricted, so we can replace μ by $-\mu$. Then (12.5) is equivalent to

$$\begin{aligned} \text{maximize} \quad & b^\top \mu \\ \text{subject to} \quad & A^\top \mu \leq c, \end{aligned} \tag{12.6}$$

which is the dual linear program for (12.4).

2.2 Lagrangian Dual for Conic Programming

Consider the following conic programming problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \preceq_{K_i} 0 \quad \text{for } i = 1, \ldots, m \\ & h_j(x) = 0 \quad \text{for } j = 1, \ldots, \ell \end{aligned} \tag{12.7}$$

where $K_1, \ldots, K_m$ are proper cones. Remember that $g_i(x) \leq_{K_i} 0$ means $-g_i(x) \in K_i$. Moreover, recall that the dual cone of a cone K is given by

$$K^* = \{y : \ y^\top x \geq 0 \ \forall x \in K\}.$$

As we picked a nonnegative multiplier $\lambda \geq 0$ to define the Lagrangian function, we pick a multiplier λ from the dual cone K^*. The Lagrangian function of (12.7) is given by

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i^\top g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x)$$

where $\lambda_i \in K_i^*$ is now a vector from the dual cone of K_i for each i. Then the Lagrangian dual function is similarly defined as $q(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu)$. The Lagrangian dual problem is given by

$$\begin{aligned} \text{maximize} \quad & q(\lambda, \mu) \\ \text{subject to} \quad & \lambda_i \geq_{K_i^*} 0 \quad \text{for } i = 1, \ldots, m. \end{aligned} \tag{12.8}$$

As an example, we consider the following semidefinite program.

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & \sum_{i=1}^{d} x_i A_i \geq_{S_+^m} B \end{aligned} \tag{12.9}$$

where S_+^m denotes the PSD cone containing all $m \times m$ PSD matrices. We learned that the PSD cone is self-dual, so the dual of S_+^m is itself. Let $Y \in S_+^m$, and consider the associated Lagrangian dual function.

$$q(Y) = \inf_x \mathcal{L}(x, Y) = \inf_x \left\{ c^\top x - \sum_{i=1}^{d} x_i \mathrm{tr}(Y^\top A_i) + \mathrm{tr}(Y^\top B) \right\}.$$

Note that the Lagrangian dual function $q(Y)$ has a finite value if and only if $c_i = \mathrm{tr}(Y^\top A_i)$ for every $i \in [d]$. Then the Lagrangian dual problem is given by

$$\begin{aligned} \text{maximize} \quad & \mathrm{tr}(Y^\top B) \\ \text{subject to} \quad & \mathrm{tr}(Y^\top A_i) = c_i \quad \text{for } i = 1, \ldots, d. \\ & Y \in S_+^m \end{aligned} \tag{12.10}$$

3 Fenchel Duality

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

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

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

3.1 Optimization with Linear Constraints

We discussed Lagrangian duality, and in fact, we can derive the Lagrangian dual function based on the conjugate function. Consider

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \\ & Cx \leq d. \end{aligned} \tag{12.11}$$

Then the associated Lagrangian dual function is given by

$$\begin{aligned} q(\lambda, \mu) &= \min_x \left\{ f(x) + \lambda^\top (Cx - d) + \mu^\top (Ax - b) \right\} \\ &= -d^\top \lambda - b^\top \mu + \min_x \left\{ f(x) + (C^\top \lambda + A^\top \mu)^\top x \right\} \\ &= -d^\top \lambda - b^\top \mu - \sup_x \left\{ -f(x) - (C^\top \lambda + A^\top \mu)^\top x \right\} \\ &= -d^\top \lambda - b^\top \mu - f^*(-C^\top \lambda - A^\top \mu). \end{aligned}$$

Note that the domain of $q(\lambda, \mu)$ is

$$\mathrm{dom}(q) = \left\{ (\lambda, \mu) : \ -C^\top \lambda - A^\top \mu \in \mathrm{dom}(f^*) \right\}.$$

Then the Lagrangian dual problem is given by

$$\begin{aligned} \text{maximize} \quad & -d^\top \lambda - b^\top \mu - f^*(-C^\top \lambda - A^\top \mu) \\ \text{subject to} \quad & \lambda \geq 0 \\ & -C^\top \lambda - A^\top \mu \in \mathrm{dom}(f^*). \end{aligned} \tag{12.12}$$

In particular, when there is no inequality constraint, the associated Lagrangian dual function is given by

$$q(\mu) = -b^\top \mu - f^*(-A^\top \mu),$$

and the Lagrangian dual problem is given by

$$\begin{aligned} \text{maximize} \quad & -b^\top \mu - f^*(-A^\top \mu) \\ \text{subject to} \quad & -A^\top \mu \in \mathrm{dom}(f^*). \end{aligned} \tag{12.13}$$

3.2 Composition Optimization

Consider the following composite optimization problem.

$$\text{minimize} \quad f(x) + g(Ax) \tag{12.14}$$

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

$$\begin{aligned} \text{minimize} \quad & f(x) + g(y) \\ \text{subject to} \quad & y = Ax. \end{aligned} \tag{12.15}$$

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

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

4 Dual Subgradient Method

We have

Lemma 12.5. *For a closed convex function $f:\mathbb{R}^d\to\mathbb{R}$, we have $f^{**}=f$.*

Based on this lemma, we may prove the following.

Theorem 12.6. *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. ∎

4.1 Linearly Constrained Problem

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

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

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}.

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.

Algorithm 1 Dual Subgradient Method for the Linearly Constrained problem

Initialize μ_1.
for $t = 1, \ldots, T-1$ **do**
  Obtain $x_t \in \operatorname{argmin}_x f(x) + \mu_t^\top (Ax - b)$,
  Update $\mu_{t+1} = \mu_t + \eta_t (Ax_t - b)$ for a step size $\eta_t > 0$.
end for

4.2 Composition Problem

We consider

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

which is equivalent to

$$\begin{array}{rl} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Ax = y. \end{array}$$

Moreover, it can be rewritten as

$$\begin{array}{rl} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Ax - y = 0. \end{array}$$

Then we may apply the dual gradient method developed for separable objective functions.

Algorithm 2 Dual Subgradient Method for the Composite Problem

Initialize μ_1.
for $t = 1, \ldots, T-1$ **do**
  Obtain $x_t \in \operatorname{argmin}_x f(x) + \mu_t^\top Ax$ and $y_t \in \operatorname{argmin}_y g(y) - \mu_t^\top y$.
  Update $\mu_{t+1} = \mu_t + \eta_t (Ax_t - y_t)$ for a step size $\eta_t > 0$.
end for

Basically, at each iteration, we minimize the Lagrangian function at $\mu = \mu_t$:

$$f(x) + g(y) + \mu_t^\top (Ax - y).$$