1 Outline

In this lecture, we study

- Lagrangian duality,
- dual algorithms.

2 Lagrangian Duality

We consider problems of the following structure.

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0 \quad \text{for } i = 1, \ldots, m \\ & h_j(x) = 0 \quad \text{for } j = 1, \ldots, \ell. \end{aligned} \tag{14.1}$$

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 (14.1) can be written as

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) \le 0 \\ & h(x) = 0. \end{aligned} \tag{14.2}$$

2.1 Lagrangian Dual Problem

The *Lagrangian function* of (14.1) 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 $\lambda \ge 0$, 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 (14.1) 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 14.1. *Let x be a feasible solution to (14.1), 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 14.1, if (14.1) 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 (14.1). The *Lagrangian dual problem* is defined as

$$\begin{aligned} \text{maximize} \quad & q(\lambda, \mu) \\ \text{subject to} \quad & \lambda \geq 0. \end{aligned} \tag{14.3}$$

We often call (14.1) as *primal* and (14.3) 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 14.2 (Weak duality)**.** *Consider the problem (14.1) and the associated Lagrangian dual problem (14.3). 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 14.1, 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 14.2 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 14.3 (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 (14.1) 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 14.2 holds with equality.

Theorem 14.4 (Strong duality)**.** *Consider the primal problem (14.1) and the associated Lagrangian dual problem (14.3). 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 (14.1) has a finite optimal value and Slater's condition, given in Definition 14.3, 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 $C = \{x : g_i(x) \leq 0 \text{ for } i = 1, \ldots, m,\ h_j(x) = 0 \text{ for } j = 1, \ldots, \ell\}$.

2.2 Karush-Kuhn-Tucker (KKT) Conditions

Remember that x^* is an optimal solution to

$$\min_{x \in C} \quad f(x)$$

where C is a convex set and f is differentiable if and only if

$$\nabla f(x^*)^\top (x - x^*) \geq 0 \quad \forall x \in C.$$

However, the structure of C may be arbitrary, which makes the condition difficult to verify. In this section, we present another way of verifying optimality. Namely, Karush-Kuhn-Tucker conditions, often referred to as KKT conditions.

We consider problems of the following structure.

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0 \quad \text{for } i = 1, \ldots, m \\ & h_j(x) = 0 \quad \text{for } j = 1, \ldots, \ell \end{aligned} \tag{14.4}$$

where

- f is convex,
- $g_1, \ldots, g_m$ are convex,
- $h_1, \ldots, h_\ell$ are affine.

Theorem 14.5 (KKT conditions for convex constrained problems)**.** *The convex programming problem as in (14.4) satisfies the following.*

1. *(Necessity) Assume that Slater's condition is satisfied. If x^* is a feasible optimal solution to (14.4), then there exist $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^\ell$ such that*

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^\ell \mu_j^* \nabla h_j(x^*) = 0 \quad \& \quad \lambda_i^* g_i(x^*) = 0 \text{ for all } i = 1, \ldots, m. \tag{$\star\star$}$$

2. *(Sufficiency) If x^* is a feasible solution to (14.4) and there exist $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^\ell$ satisfying ($\star\star$), then x^* is an optimal solution to (14.4).*

2.3 KKT Conditions for Linear Constraints

We consider problems of the following structure.

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \leq b \\ & Cx = d \end{array} \tag{14.5}$$

where

- $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$,
- $C \in \mathbb{R}^{\ell \times d}$ and $d \in \mathbb{R}^\ell$.

Theorem 14.6 (KKT conditions for linearly constrained problems)**.** *The linearly constrained problem as in (14.5) satisfies the following.*

1. *(Necessity) If x^* is a feasible solution to (14.5) and $f(x^*)$ is a local minimum, then there exist $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^\ell$ such that*

$$\nabla f(x^*)^\top + \lambda^{*\top} A + \mu^{*\top} C = 0 \quad \& \quad \lambda^{*\top}(Ax - b) = 0. \tag{$\star$}$$

2. *(Sufficiency) If f is convex, x^* is a feasible solution to (14.5), and there exist $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^\ell$ satisfying ($\star$), then x^* is an optimal solution to (14.5).*

3 Dual Methods

We consider

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

For a dual multiplier μ, the Lagrangian is given by

$$\mathcal{L}(x, \mu) = f(x) + \mu^\top (Ax - b).$$

Here, we may interprete the Lagrangian as a penalized objective function.

3.1 Dual Subgradient Method

The first algorithm for the constrained minimization problem is what we call the *dual subgradient method*. The idea behind the dual subgradient method is to adapt the dual multiplier μ which controls the level of penalization. Namely, we start with an initial μ_1 and update μ_t for $t \geq 1$. Given μ_t, we apply

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

Here, what is g_t? The dual subgradient method proceeds with

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

Algorithm 1 Subgradient method for the dual 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

Here, $f(x) + \mu_t^\top (Ax - b)$ is the Lagrangian function $\mathcal{L}(x, \mu)$ at $\mu = \mu_t$. 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.

3.2 Augmented Lagrangian Method

The next algorithm for the constrained minimization problem is as follows.

$$\begin{aligned}
&x_t \in \operatorname*{argmin}_x \left\{ f(x) + \mu_t^\top (Ax - b) + \frac{\eta}{2} \|Ax - b\|_2^2 \right\} \\
&\mu_{t+1} = \mu_t + \eta (Ax_t - b).
\end{aligned}$$

This is precisely, the *augmented Lagrangian method* (ALM).

Algorithm 2 Augmented Lagrangian method

Initialize μ_1.
for $t = 1, \ldots, T$ **do**
  Find $x_t \in \operatorname{argmin}_x \left\{ f(x) + \mu_t^\top (Ax - b) + \frac{\eta}{2} \|Ax - b\|_2^2 \right\}$.
  Update $\mu_{t+1} = \mu_t + \eta (Ax_t - b)$.
end for

Notice that the augmented Lagrangian method is the dual gradient method applied to the following equivalent formulation of the primal problem.

$$\begin{aligned}
\text{minimize} \quad & f(x) + \frac{\eta}{2} \|Ax - b\|_2^2 \\
\text{subject to} \quad & Ax = b.
\end{aligned}$$

4 Composite Minimization

We consider

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

which is equivalent to

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

Moreover, it can be rewritten as

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

Here, the Lagrangian function is given by

$$\mathcal{L}(x, y, \mu) = f(x) + g(y) + \mu^\top (Ax - y).$$

Then we may apply the dual subgradient method developed for separable objective functions. Basically, at each iteration, we minimize the Lagrangian function at $\mu = \mu_t$. The dual subgradient method works with the update rule

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

for some step size $\eta_t > 0$.

Instead, the augmented Lagrangian method considers the augmented Lagrangian function given by

$$f(x) + g(y) + \mu_t^\top (Ax - y) + \frac{\eta}{2}\|Ax - y\|_2^2.$$

Here, μ_t changes over iterations while η remains constant. ALM works with the update rule

$$\begin{aligned} & (x_t, y_t) \in \operatorname*{argmin}_{(x,y)} f(x) + g(y) + \mu_t^\top (Ax - y) + \frac{\eta}{2}\|Ax - y\|_2^2, \\ & \mu_{t+1} = \mu_t + \eta(Ax_t - y_t). \end{aligned}$$

Lastly, we discuss the alternating direction method of multipliers (ADMM). The algorithm works with the following update rule.

$$\begin{aligned} & x_t \in \operatorname*{argmin}_x \left\{ f(x) + g(y_{t-1}) + \mu_t^\top (Ax - y_{t-1}) + \frac{\eta}{2}\|Ax - y_{t-1}\|_2^2 \right\}, \\ & y_t \in \operatorname*{argmin}_y \left\{ f(x_t) + g(y) + \mu_t^\top (Ax_t - y) + \frac{\eta}{2}\|Ax_t - y\|_2^2 \right\}, \\ & \mu_{t+1} = \mu_t + \eta(Ax_t - y_t). \end{aligned}$$

Algorithm 3 Alternating direction method of multipliers

Initialize μ_1 and y_0.
for $t = 1, \ldots, T-1$ **do**
  Obtain $x_t \in \operatorname{argmin}_x \left\{ f(x) + g(y_{t-1}) + \mu_t^\top (Ax - y_{t-1}) + \frac{\eta}{2}\|Ax - y_{t-1}\|_2^2 \right\}$,
  Obtain $y_t \in \operatorname{argmin}_y \left\{ f(x_t) + g(y) + \mu_t^\top (Ax_t - y) + \frac{\eta}{2}\|Ax_t - y\|_2^2 \right\}$,
  Update $\mu_{t+1} = \mu_t + \eta(Ax_t - y_t)$.
end for