3.2 Duality

3.2.1 Lagrangian and Dual Function

We consider a standard optimization problem (P):

$$\begin{aligned} \min\ & f(x) \\ \text{subject to } & g_i(x) \leq 0,\ \ i = 1, ..., h \\ & h_j(x) = 0,\ \ j = 1, ..., k \end{aligned}$$

The optimal value p^* of (P) is called the primal optimal value.
Definition: (Lagrangian) The *Lagrangian* associated with the above problem is defined as

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{h} \lambda_i g_i(x) + \sum_{j=1}^{k} \mu_j h_j(x)$$

The vectors λ, μ are called the *dual variables* or *Lagrange multipliers.*

Definition: (Dual function) The *dual function* is defined as

$$q(\lambda, \mu) = \inf_x L(x, \lambda, \mu)$$

Note that q is always concave, being the pointwise infimum of affine functions.

Let p^* be the optimal value of (P). The dual function gives a lower bound on p^*.

Proposition: For all $\lambda \geq 0$ and μ, we have

$$q(\lambda, \mu) \leq p^*$$

Proof. Let x be a feasible point. Then $g_i(x) \leq 0$ and $h_j(x) = 0$. Then

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

Hence for all $\lambda \geq 0$ and μ,

$$q(\lambda, \mu) \leq L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{h} \lambda_i g_i(x) + \sum_{j=1}^{k} \mu_j h_j(x) \leq f(x)$$

Since this holds for all feasible points, we have $q(\lambda, \mu) \leq p^*$. □

We next consider the dual problem.
Definition:(Dual Problem) The following optimization problem (D) is called the *dual problem* associated to (P):

$$\begin{aligned} \max\ & q(\lambda, \mu) \\ \text{subject to } & \lambda \geq 0 \end{aligned}$$

A pair (λ, μ) such that $\lambda \geq 0$ and $q(\lambda, \mu) > -\infty$ is called *dual feasible*.
A optimal solution (λ^*, μ^*) is called *dual optimal*.

Example: (Linear Program)
Consider a standard linear program (LP):

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \langle c, x \rangle \\ \text{subject to } & Ax = b \\ & x \geq 0 \end{aligned}$$

The Lagrangian is given by

$$L(x, \lambda, \mu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \mu^T (Ax - b) = (A^T \mu + c - \lambda)^T x - b^T \mu$$

If $c + A^T \mu - \lambda \neq 0$, then $L(x, \lambda, \mu)$ is unbounded below. Hence the dual function is given by

$$q(\lambda, \mu) = \begin{cases} -b^T \mu & c + A^T \mu - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Therefore, the dual problem is given by

$$\begin{aligned} \max \ & -b^T \mu \\ \text{subject to } & A^T \mu + c - \lambda = 0 \\ & \lambda \geq 0 \end{aligned}$$

It can also be written in this form:

$$\begin{aligned} \max \ & -b^T \mu \\ & A^T + c\mu \geq 0 \end{aligned}$$

Example: Duality and Conjugate function
Consider the following optimization problem

$$\begin{aligned} \min \ & f(x) \\ \text{subject to } & Ax \leq b \\ & Cx = d \end{aligned}$$

The dual function is

$$\begin{aligned} q(\lambda, \mu) &= \inf_x (f(x) + \lambda^T (Ax - b) + \mu^T (Cx - d)) \\ &= -b^T \lambda - d^T \mu + \inf_x (f(x) + (A^T \lambda + C^T \mu)^T x) \end{aligned}$$

Note that

$$\inf_x (f(x) + (A^T \lambda + C^T \mu)^T x) = -\sup_x (-(A^T \lambda + C^T \mu)^T x - f(x)) = -f^*(-(A^T \lambda + C^T \mu))$$

Hence, we have

$$q(\lambda, \mu) = -b^T \lambda - d^T \mu - f^*(-(A^T \lambda + C^T \mu))$$

3.2.2 Strong and Weak Duality

Let d^* be the optimal value of the dual problem. We have the following inequality.

Proposition:(Weak Duality) Let p^* be the primal optimal value and d^* be the dual optimal value. Then

$$d^* \leq p^*$$

The difference $p^* - d^*$ is called the *duality gap*.
If $p^* = d^*$, then we say that *strong duality* holds.

This leads us to ask the question when do strong duality holds.
Such conditions are called constraint qualification. We will study one simple qualification: Slater's condition.

Consider a convex problem of the form:

$$\begin{aligned} \min\ & f(x) \\ \text{subject to } & g_i(x) \leq 0,\ i = 1, ..., h \\ & Ax = b \end{aligned}$$

where f, g_i are convex.

***Slater's Condition*:** There exists $x \in \mathrm{ri}(D)$ such that

$$g_i(x) < 0,\ i = 1, ..., h, \quad Ax = b$$

where $D = \mathrm{dom} f \cap (\cap_i \mathrm{dom} g_i)$.

Theorem:(Slater's Theorem) If the problem is convex and Slater's condition is satisfied, then strong duality holds.

3.2.3 Geometric Interpretation

Consider the following set

$$A := \{(u, v, t) |\ \exists x\ g_i(x) \leq u_i,\ i = 1, ..., h,\ h_j(x) = v_j,\ j = 1, ..., k,\ f(x) \leq t\}$$

We can show that A is convex if the problem is convex.
Note that

$$p^* = \inf\{t |\ (0, 0, t) \in A\}$$

that is the lowest point where A intersects the 'vertical'-axis.
We can also interpret the dual function through this geometric setting:

$$q(\lambda, \mu) = \inf\{\langle(\lambda, \mu, 1), (u, v, t)\rangle |\ (u, v, t) \in A\}$$

For fixed (λ, μ), we can define a hyperplane

$$\langle (\lambda, \mu, 1), (u, v, t) \rangle = q$$

Then $q(\lambda, \mu)$ is where a supporting hyperplane to A with 'slope' (λ, μ) intersects the 'vertical' axis.
Therefore, strong duality holds if and only if there is a nonvertical supporting hyperplane to A at $(0, 0, p^*)$.

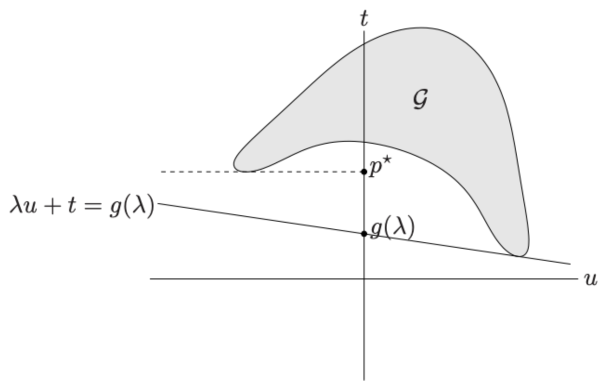

Figure 1: Geometric picture of the set G and dual function

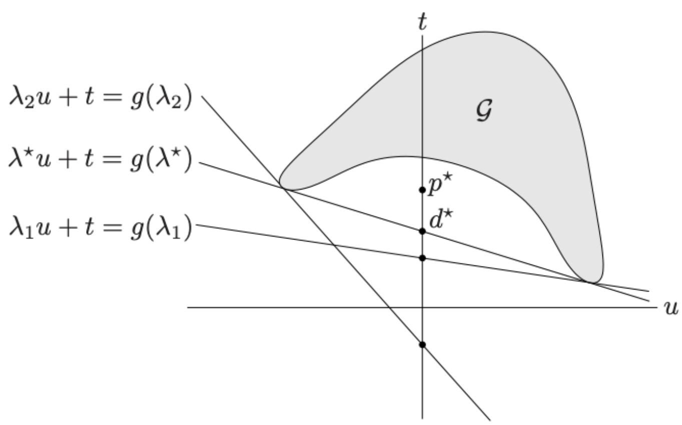

Figure 2: Primal and dual optimal value

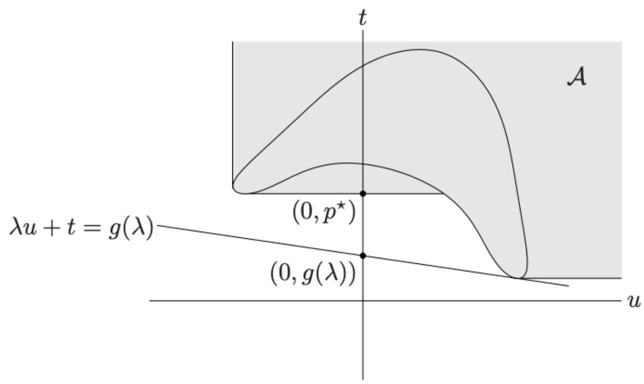

Figure 3: Geometric picture of the set A

Example: Consider the problem

$$\min_{x,y \geq 0} \; e^{-\sqrt{xy}}$$
$$\text{subject to } x = 0$$

The optimal value p^* is 1.
The dual function is given by

$$q(\lambda) = \inf_{x,y \geq 0} \{e^{-\sqrt{xy}} + \lambda x\} = \begin{cases} 0 & \lambda \geq 0 \\ -\infty & \lambda < 0 \end{cases}$$

Hence, the dual optimal value d^* is 0.
Therefore, the strong duality does not hold.
Note that Slater's Condition is not satisfied for this example.

3.3 KKT conditions

Let's consider the general convex problem again

$$\begin{aligned} \min \; & f(x) \\ \text{subject to } & g_i(x) \leq 0, \;\; i = 1, ..., h \\ & h_j(x) = 0, \;\; j = 1, ..., k \end{aligned}$$

where are the functions are convex. We also assume that h_j are affine.
Note that

$$\sup_{\lambda \geq 0, \mu} L(x, \lambda, \mu) = \begin{cases} f(x) & g_i(x) \leq 0, \;\; h_j(x) = 0 \\ \infty & \text{otherwise} \end{cases}$$

Then $p^* = \inf_x \sup_{\lambda \geq 0, \mu} L(x, \lambda, \mu)$
On the other hand, $d^* = \sup_{\lambda \geq 0, \mu} \inf_x L(x, \lambda, \mu)$.
Therefore, strong duality is equivalent to

$$\sup_{\lambda \geq 0, \mu} \inf_x L(x, \lambda, \mu) = \inf_x \sup_{\lambda \geq 0, \mu} L(x, \lambda, \mu)$$

Suppose strong duality holds. Let x^* be primal optimal and (λ^*, μ^*) be dual optimal. Then

$$\begin{aligned} f(x^*) &= q(\lambda^*, \mu^*) \\ &= \inf_x (f(x) + \sum_{i=1}^{h} \lambda_i^* g_i(x) + \sum_{j=1}^{k} \mu_j^* h_j(x)) \\ &\leq f(x^*) + \sum_{i=1}^{h} \lambda_i^* g_i(x^*) + \sum_{j=1}^{k} \mu_j^* h_j(x^*) \\ &\leq f(x^*) \end{aligned}$$

Therefore, we have equality for each line. In particular, we have

$$\sum_{i=1}^{h} \lambda_i^* g_i(x^*) = 0$$

Since each term is nonpositive, we have $\lambda_i^* g_i(x^*) = 0$ for all i.
This is called *complementary slackness*.
Suppose all the functions are also differentiable. Then since x^* minimize $L(x, \lambda^*, \mu^*)$, we have

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0$$

That is

$$\nabla f(x^*) + \sum_{i=1}^{h} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{k} \mu_j^* h_j(x^*) = 0$$

Combining with the complementary slackness condition, we have the following *Karush-Kuhn-Tucker*(KKT) condition:

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{h} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{k} \mu_j^* \nabla h_j(x^*) = 0 \\ g_i(x^*) \leq 0, \ i = 1, ..., h \\ h_j(x^*) = 0, \ j = 1, ..., k \\ \lambda_i^* \geq 0 \\ \lambda_i^* g_i(x^*) = 0, \ i = 1, ..., h \end{aligned}$$

Conversely, suppose $x^*, (\lambda^*, \mu^*)$ satisfy the KKT conditions.
Since $L(x, \lambda^*, \mu^*)$ is convex in x and $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$, then x^* minimizes

$L(x, \lambda^*, \mu^*)$. Then

$$q(\lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*) = f(x^*) + \sum_{i=1}^{h} \lambda_i^* g_i(x^*) + \sum_{j=1}^{k} \mu_j^* h_j(x^*) = f(x^*)$$

Therefore, there is no duality gap and x^*, (λ^*, μ^*) are primal optimal and dual optimal respectively.
To conclude, we have the following optimal condition:

Theorem: Consider the convex problem (P). Suppose strong duality holds. Then $x^*, (\lambda^*, \mu^*)$ are primal and dual optimal if and only if $x^*, (\lambda^*, \mu^*)$ satisfy the KKT conditions.

Remark: If the functions are not differentiable, we can replace the first KKT condition by $0 \in \partial f(x^*) + \sum_{i=1}^{h} \lambda_i^* \partial g_i(x^*) + \sum_{j=1}^{k} \mu_j^* \partial h_j(x^*)$.

Example Consider the problem

$$\begin{aligned} \min\ & x^2 + y^2 \\ \text{subject to } & x + y = 1 \\ & x, y \geq 0 \end{aligned}$$

The KKT condition can be written as

$$\begin{aligned} 2x - \lambda_1 + \mu &= 0 \\ 2y - \lambda_2 + \mu &= 0 \\ x + y &= 1 \\ x, y &\geq 0 \\ \lambda_1, \lambda_2 &\geq 0 \\ \lambda_1 x = \lambda_2 y &= 0 \end{aligned}$$

By the first two conditions, we have,

$$\lambda_1 = 2x + \mu,\ \ \lambda_2 = 2y + \mu.$$

By the complementary slackness conditions, we have

$$2x^2 + \mu x = 0,\ \ 2y^2 + \mu y = 0.$$

Then $x = 0$ or $-\mu/2$, and $y = 0$ or $-\mu/2$.
We cannot have x, y both equal to 0 since otherwise $x + y = 0 \neq 1$.
Suppose exactly one of x, y is zero, say x.
Then $y = -\mu/2$. Since $x + y = 1$, then $\mu = -2$. But since $x = 0$, this implies that $\lambda_1 = \mu = -2 < 0$. This violates the dual feasibility.
Therefore, we must have x, y both nonzero. That is $x = y = -\mu/2$.
Since $x + y = 1$, we have $\mu = -1$. So $x = y = 1/2$ and $\lambda_1 = \lambda_2 = 0$ satisfy the KKT conditions. Therefore the global minimum is obtained at $(1/2, 1/2)$.