Convex Optimization

Chapter 5: Duality

Notes by Renxiong Liu

In this notes, we mainly discuss Lagrangian duality that are widely used for constrained optimization problem.

5.1 Lagrangian Duality

Firstly, we introduce some notions that will be frequently used in this note. For optimization problem (note: not necessary convex),

$$\begin{aligned}(\Pi_P)\ p^\star = \quad & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \le 0,\ i = 1, 2 \ldots, m \leftarrow \lambda_i \\ &&& h_j(x) = 0,\ j = 1, 2 \ldots, n \leftarrow v_i\end{aligned}$$

We call f_0 the primal function, $x \in \mathbb{R}^p$ the primal variables. Note the domain for this problem is

$$D = (\bigcap_{i=0}^{m} \mathbf{dom}(f_i)) \cap (\bigcap_{j=1}^{n} \mathbf{dom}(h_j))$$

For problems above, one very useful method is to consider the Lagrangian function,

$$L : \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$$

with domain $\mathbf{dim}(L) = D \times \mathbb{R}^m \times \mathbb{R}^n$:

$$L(x, \lambda_1, \ldots, \lambda_m, v_1, \ldots, v_n) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{n} v_j h_j(x)$$

where we call $\lambda = (\lambda_1, \ldots, \lambda_m)$ and $v = (v_1 \ldots, v_n)$ the dual variables.

To proceed, we define the dual function as

$$g(\lambda_1, \lambda_m, v_1, \ldots, v_n) = \inf_{x \in D} L(x, \lambda_1, \ldots, \lambda_m, v_1, \ldots, v_n)$$

Note for any fixed $x \in D$, $L(x, \lambda_1, \ldots, \lambda_m, v_1, \ldots, v_n)$ is an affine function with respect to $(\lambda_1, \ldots, \lambda_m, v_1, \ldots, v_n)$, we have that $g(\lambda_1, \lambda_m, v_1, \ldots, v_n)$ must be convace with respect to $(\lambda_1, \lambda_m, v_1, \ldots, v_n)$.

5.2 Weak Duality

Given the dual function $g(\lambda_1, \lambda_m, v_1, \ldots, v_n)$ and corresponding dual variables $(\lambda_1, \ldots, \lambda_m, v_1, \ldots, v_n)$, we define the following problem as the dual problem

$$\begin{aligned}(\Pi_D)\ d^\star = \quad & \text{maximize} && g(\lambda_1, \lambda_m, v_1, \ldots, v_n) \\ & \text{subject to} && \lambda_i \ge 0,\ i = 1, 2 \ldots, m\end{aligned}$$

Now, we may wonder what's the relationship between primal problem and dual problem. Indeed, we can see that for any $x \in D$, $\lambda_i \geq 0$ and $v_j \in \mathbb{R}$, we have that

$$\begin{aligned} g(\lambda_1, \ldots, \lambda_m, v_1, \ldots, v_n) &\leq L(x, \lambda_1, \ldots, \lambda_m, v_1, \ldots, v_n) \\ &= f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{n} v_j h_j(x) \\ &\leq f_0(x) \end{aligned}$$

which means

$$g(\lambda_1, \ldots, \lambda_m, v_1, \ldots, v_n) \leq d^\star \leq p^\star \leq f_0(x)$$

In general, $d^\star < p^\star$, we define $p^\star - d^\star$ as the duality gap. Moreover, if we find $\lambda^\star, v^\star$ and $x^\star$ such that $g(\lambda^\star, v^\star) = f_0(x^\star)$, then we will have that

$$\begin{aligned} x^\star &\in \arg\min_{x \in \mathcal{F}} f_0(x) \\ (\lambda^\star, v^\star) &\in \arg\max_{\lambda \geq 0} g(\lambda, v) \end{aligned}$$

To understand the duality, let's first look at one example. To proceed, for any l_p norm $\|\cdot\|_p$, we define the associated dual norm as

$$\|y\|_\star = \text{maximize}\, \langle y, x \rangle, \text{ subject to } \|x\|_p \leq 1 \tag{5.1}$$

Recall the Holder inequality:

$$\langle x, y \rangle \leq \|x\|_p \cdot \|y\|_q$$

where $p, q > 0$ and $\frac{1}{p} + \frac{1}{q} = 1$, we have that the dual norm of $\|\cdot\|_p$ is $\|\cdot\|_q$. Now, given any (convex) norm $\|\cdot\|$, we claim the conjugate function of $f(x) = \|x\|$ is

$$f^\star(y) = \begin{cases} 0, & \text{if } \|y\|_\star \leq 1 \\ \infty, & \text{otherwise} \end{cases}$$

This is because

$$f^\star(y) = \sup_x \langle y, x \rangle - \|x\| = \sup_x \|x\|(\langle y, \frac{x}{\|x\|} \rangle - 1) = \sup_x \|x\|(\|y\|_\star - 1) = \begin{cases} 0, & \text{if } \|y\|_\star \leq 1 \\ \infty, & \text{otherwise} \end{cases}$$

Now we are ready to discuss the following example.

Example 5.1. Find out the dual problem of the following optimization problem.

$$\begin{aligned} (\Pi_P) \qquad &\text{minimize} \quad \|x\| \\ &\text{subject to} \quad Ax = b \end{aligned}$$

Note the Lagrangian function is

$$L(x, v) = \|x\| + v^\top (Ax - b) = (\langle A^\top v, x \rangle + \|x\|) - v^\top b$$

The dual fucntion can be represented as

$$g(v) = \inf_x (\langle A^\top v, x \rangle + \|x\|) - v^\top b = -v^\top b - \sup_x (\langle -A^\top v, x \rangle - \|x\|) = -v^\top b - f^\star(A^\top v)$$

Hence the dual problem is

$$\begin{array}{lll}(\Pi_D) & \text{minimize} & v^\top b \\ & \text{subject to} & \|A^\top v\|_\star \leq 1\end{array}$$

Particularly, if we choose $\|\cdot\|$ as $\|\cdot\|_1$, then $\|\cdot\|_\star$ will be $\|\cdot\|_\infty$. In this case, the primal problem is called the basis pursuit problem. The dual problem will provide an efficient method to calculate the solution of basis pursuit problem.

5.3 Strong duality

If the duality gap $p^\star - d^\star = 0$, then it is said that strong duality holds. One sufficient condition for strong duality to hold is the Slater's condition. Consider the following **convex** program:

$$\begin{array}{lll}(\Pi_P) & \text{minimize} & f_0(x) \\ & \text{subject to} & f_i(x) \leq 0, \ i = 1, 2 \ldots, m \\ & & Ax = b\end{array}$$

where f_i, $i = 0, 1, \ldots, m$ are convex. if there exists $x \in \text{relint}(D)$ ($D = \cap_{i=0}^m \text{dom}(f_i)$) such that $f_i(x) < 0$, $i = 1, \ldots, m$ and $Ax = b$, then we say the Slater's condition holds. When Slater's condition holds for a convex program, then the strong duality will hold (important property!).

5.4 Karush-Kuhn-Tuker (KKT) conditions

Now we have seen the general relationship between primal problem and corresponding dual problem. Next, we establish the important *KKT conditions* that will be widely used in constrained optimization problems. Denote

$$L(x, \lambda_1, \ldots, \lambda_m, v_1, \ldots, v_n) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^n v_j h_j(x)$$

Formally speaking, KKT conditions tell

$$\begin{cases} \nabla_x L(x^\star, \lambda^\star, v^\star) = 0 \\ f_i(x^\star) \leq 0, i = 1, \ldots, m \text{ and } h_j(x^\star) = 0, j = 1, \ldots, n \\ \lambda_i \geq 0, i = 1, \ldots, m \\ \lambda_i f_i(x^\star) = 0, i = 1, \ldots, m \end{cases}$$

Implications of strong duality Firstly, we assume the strong duality holds, which means

$$f_0(x^\star) = p^\star = d^\star = g(\lambda^\star, v^\star)$$

Note

$$\begin{aligned} f_0(x^\star) &= g(\lambda^\star, v^\star) = \inf_x L(x, \lambda^\star, v^\star) \\ &\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i f_i(x^\star) + \sum_{j=1}^n v_j h_j(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

We have that $\inf_x L(x, \lambda^\star, v^\star) = L(x^\star, \lambda^\star, v^\star)$ and $\lambda_i f_i(x^\star) = 0$ for all $i = 1, 2 \ldots, m$. When f_i and h_j are all differentiable, this implies $\nabla_x L(x^\star, \lambda^\star, v^\star) = 0$. Note the feasibility condition holds naturally, hence strong duality implies KKT conditions, even if the problem is nonconvex. Moreover, we can see that $x^\star = \arg\min_x L(x, \lambda^\star, v^\star)$.

Implications of KKT for convex program. Note L is convex with respect to x, we have that $\nabla_x L(x^\star, \lambda^\star, v^\star) = 0$ implies $L(x^\star, \lambda^\star, v^\star) = \inf_x L(x, \lambda^\star, v^\star)$. This together with condition $\lambda_i f_i(x^\star) = 0, i = 1, \ldots, m$ implies that

$$L(x^\star, \lambda^\star, v^\star) = g(\lambda^\star, v^\star) = f_0(x^\star)$$

hence the strong duality holds.

In summary, when strong duality holds (e.g. Slater's condition holds in convex program), then KKT conditions is necessary; for convex program, any solution $(x^\star, \lambda^\star, v^\star)$ satisfying KKT conditions must be optimal solution to primal and dual problems.