1. Convex Optimization, Saddle Point Theory, and Lagrangian Duality

In this section we extend the duality theory for linear programming to general problmes of convex optimization. This is accomplished using the saddle point properties of the Lagrangian in convex optimization. Again, consider the problem

$$\begin{array}{lll} \mathcal{P} & \text{minimize} & f_0(x) \\ & \text{subject to} & f_i \leq 0, \ i = 1, 2, \ldots, s \\ & & f_i(x) = 0, \ i = s+1, \ldots, m, \end{array}$$

where it is assumed that the functions $f_0, f_1, \ldots, f_s$ are convex functions mapping $\mathbb{R}^n$ to $\bar{\mathbb{R}}$, and $f_{s+1}, \ldots, f_m$ are affine mappings from $\mathbb{R}^n$ to $\mathbb{R}$. We denote the constraint region for $\mathcal{P}$ by Ω.

The Lagrangian for $\mathcal{P}$ is the function

$$L(x, y) = f_0(x) + y_1 f_1(x) + y_2 f_2(x) + \cdots + y_m f_m(x),$$

where it is always assumed that $0 \leq y_i, \ i = 1, 2, \ldots, s$. Set $K = \mathbb{R}^s_+ \times \mathbb{R}^{m-s} \subset \mathbb{R}^m$. A pair $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times K$ is said to be a saddle point for L if

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, y) \qquad \forall \ (x, y) \in \mathbb{R}^n \times K.$$

We have the following basic saddle point theorem for L.

Theorem 1.1 (Saddle Point Theorem). *Let $\bar{x} \in \mathbb{R}^n$. If there exists $\bar{y} \in K$ such that $(\bar{x}, \bar{y})$ is a saddle point for the Lagrangian L, then $\bar{x}$ solves $\mathcal{P}$. Conversely, if $\bar{x}$ is a solution to $\mathcal{P}$ at which the Slater C.Q. is satisfied, then there is a $\bar{y} \in K$ such that $(\bar{x}, \bar{y})$ is a saddle point for L.*

Proof. If $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times K$ is a saddle point for $\mathcal{P}$ then

$$\sup_{y \in K} L(\bar{x}, y) = \sup_{y \in K} f_0(\bar{x}) + y_1 f_1(\bar{x}) + y_2 f_2(\bar{x}) + \cdots + y_m f_m(\bar{x}) \leq L(\bar{x}, \bar{y}).$$

If for some $i \in \{1, \ldots, s\}$ such that $f_i(\bar{x}) > 0$, then we could send $y_i \uparrow +\infty$ to find that the supremum on the left is $+\infty$ which is a contradiction, so we must have $f_i(\bar{x}) \leq 0, \ i = 1, \ldots, s$. Moreover, if $f_i(\bar{x}) \neq 0$ for some $i \in \{s+1, \ldots, m\}$, then we could send $y_i \uparrow -\text{sign}(f_i(\bar{x}))\infty$ to again find that the supremum on the left is $+\infty$ again a contradiction, so we must have $f_i(\bar{x}) = 0, \ i = s+1, \ldots, m$. That is, we must have $\bar{x} \in \Omega$. Since $L(\bar{x}, \bar{y}) = \sup_{y \in K} L(\bar{x}, y)$, we must have $\sum_{i=1}^m \bar{y}_i f_i(\bar{x}) = 0$. Therefore the right half of the saddle point condition implies that

$$f_0(\bar{x}) = L(\bar{x}, \bar{y}) \leq \inf_x L(x, \bar{y}) \leq \inf_{x \in \Omega} L(x, \bar{y}) \leq \inf_{x \in \Omega} f_0(x) \leq f_0(\bar{x}),$$

and so $\bar{x}$ solves $\mathcal{P}$.

Conversely, if $\bar{x}$ is a solution to $\mathcal{P}$ at which the Slater C.Q. is satisfied, then there is a vector $\bar{y}$ such that $(\bar{x}, \bar{y})$ is a KKT pair for $\mathcal{P}$. Primal feasibility ($\bar{x} \in \Omega$), dual feaasibility ($\bar{y} \in K$), and complementarity ($\bar{y}_i f_i(\bar{x}), \ i = 1, \ldots, s$) imply that

$$L(\bar{x}, y) \leq f_0(\bar{x}) = L(\bar{x}, \bar{y}) \quad \forall \ y \in K.$$

On the other hand, dual feasibility and convexity imply the convexity of the function $L(x, \bar{y})$ in x. Hence the condition $0 = \nabla_x L(\bar{x}, \bar{y})$ implies that $\bar{x}$ is a global minimizer for the function $x \to L(x, \bar{y})$, that is

$$L(\bar{x}, \bar{y}) \le L(x, \bar{y}) \quad \forall\ x \in \mathbb{R}^n .$$

Therefore, $(\bar{x}, \bar{y})$ is a saddle point for L. $\square$

Note that it is always the case that

$$\sup_{y \in K} \inf_{x \in \mathbb{R}^n} L(x, y) \le \inf_{x \in \mathbb{R}^n} \sup_{y \in K} L(x, y)$$

since the largest minimum is always smaller that the smallest maximum. On the other hand, if $(\bar{x}, \bar{y})$ is a saddle point for L, then

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in K} L(x, y) \le \sup_{y \in K} L(\bar{x}, y) \le L(\bar{x}, \bar{y}) \le \inf_{x \in \mathbb{R}^n} L(x, \bar{y}) \le \sup_{y \in K} \inf_{x \in \mathbb{R}^n} L(x, y).$$

Hence, if a saddle point for L exists on $\mathbb{R}^n \times K$, then

$$\sup_{y \in K} \inf_{x \in \mathbb{R}^n} L(x, y) = \inf_{x \in \mathbb{R}^n} \sup_{y \in K} L(x, y).$$

Such a result is called a mini-max theorem and provides conditions under which one can exchange and inf-sup for a sup-inf. This mini-max result can be used as a basis for convex duality theory.

Observe that we have already shown that

$$\sup_{y \in K} L(x, y) = \begin{cases} +\infty & \text{if } x \notin \Omega, \\ f_0(x) & \text{if } x \in \Omega. \end{cases}$$

Therefore,

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in K} L(x, y) = \inf_{x \in \Omega} f_0(x) \ .$$

We will call this the *primal* problem. This is the inf-sup side of the saddle point problem. The other side, the sup-inf problem, we will call the *dual* problem with dual objective function

$$g(y) = \inf_{x \in \mathbb{R}^n} L(x, y) \ .$$

The Saddle Point Theorem says that if $(\bar{x}, \bar{y})$ is a saddle point for L, then $\bar{x}$ solves the primal problem, $\bar{y}$ solves the dual problem, and the optimal values in the primal and dual problems coincide. This is a *Weak Duality Theorem*. The Strong Duality Theorem follows from the second half of the Saddle Point Theorem and requires the use of the Slater Constraint Qualification.

1.1. **Linear Programming Duality.** We now show how the Lagrangian Duality Theory described above gives linear programming duality as a special case. Consider the following LP:

$$\begin{array}{lll} \mathcal{P} & \text{minimize} & b^T x \\ & \text{subject to} & A^T x \ge c, \ 0 \le x \ . \end{array}$$

The Lagrangian is

$$L(x, y, v) = b^T x + y^T (c - A^T x) - v^T x, \quad \text{where } 0 \le y, \ 0 \le v \ .$$

The dual objective function is

$$g(y,u) = \min_{x\in\mathbb{R}^n} L(x,y,v) = \min_{x\in\mathbb{R}^n} b^T x + y^T(c - A^T x) - v^T x \ .$$

Our first goal is to obtain a closed form expression for $g(y,u)$. This is accomplished by using the optimality conditions for minimizing $L(x,y,u)$ to eliminate x from the definition of L. Since $L(x,y,v)$ is a convex function in x, the global solution to $\min_{x\in\mathbb{R}^n} L(x,y,v)$ is obtained by solving the equation $0 = \nabla_x L(x,y,u) = b - Ay - v$ with $0 \le y,\ 0 \le v$. Using this condition in the definition of L we get

$$L(x,y,u) = b^T x + y^T(c - A^T x) - v^T x = (b - Ay - v)^T x + c^T y = c^T y,$$

subject to $b - A^T y = v$ and $0 \le y,\ 0 \le v$. Hence the Lagrangian dual problem

$$\begin{array}{ll} \text{maximize} & g(y,v) \\ \text{subject to} & 0 \le y,\ 0 \le v \end{array}$$

can be written as

$$\begin{array}{lll} \mathcal{D} & \text{maximize} & c^T y \\ & \text{subject to} & b - Ay = v,\ 0 \le y,\ 0 \le v \ . \end{array}$$

Note that we can treat the variable v as a slack variable in this LP and write

$$\begin{array}{lll} \mathcal{D} & \text{maximize} & c^T y \\ & \text{subject to} & Ay \le b,\ 0 \le y \ . \end{array}$$

The linear program $\mathcal{D}$ is the dual to the linear program $\mathcal{P}$.

1.2. **Convex Quadratic Programming Duality.** One can also apply the Lagrangian Duality Theory in the context of Convex Quadratic Programming. To see how this is done let $Q \in \mathbb{R}^{n\times n}$ be symmetric and positive definite, and let $c \in \mathbb{R}^n$. Consider the convex quadratic program

$$\begin{array}{lll} \mathcal{D} & \text{minimize} & \frac{1}{2}x^T Q x + c^T x \\ & \text{subject to} & Ax \le b,\ 0 \le x \ . \end{array}$$

The Lagrangian is given by

$$L(x,y,v) = \frac{1}{2}x^T Q x + c^T x + y^T(A^T x - b) - v^T x \quad \text{where } 0 \le y,\ 0 \le v.$$

The dual objective function is

$$g(y,v) = \min_{x\in\mathbb{R}^n} L(x,y,v) \ .$$

The goal is to obtain a closed form expression for g with the variable x removed by using the first-order optimality condition $0 = \nabla_x L(x,y,v)$. This optimality condition completely identifies the solution since L is convex in x. We have

$$0 = \nabla_x L(x,y,v) = Qx + c + A^T y - v.$$

Since Q is invertible, we have

$$x = Q^{-1}(v - A^T y - c).$$

Plugging this expression for x into $L(x, y, v)$ gives

$$\begin{aligned}
g(y,v) &= L(Q^{-1}(v - A^T y - c), y, v) \\
&= \frac{1}{2}(v - A^T y - c)^T Q^{-1}(v - A^T y - c) \\
&\qquad + c^T Q^{-1}(v - A^T y - c) + y^T(AQ^{-1}(v - A^T y - c) - b) - v^T Q^{-1}(v - A^T y - c) \\
&= \frac{1}{2}(v - A^T y - c)^T Q^{-1}(v - A^T y - c) - (v - A^T y - c)^T Q^{-1}(v - A^T y - c) - b^T y \\
&= -\frac{1}{2}(v - A^T y - c)^T Q^{-1}(v - A^T y - c) - b^T y \ .
\end{aligned}$$

Hence the dual problem is

$$\begin{array}{ll}
\text{maximize} & -\frac{1}{2}(v - A^T y - c)^T Q^{-1}(v - A^T y - c) - b^T y \\
\text{subject to} & 0 \leq y,\ 0 \leq v \ .
\end{array}$$

Moreover, $(\bar{y}, \bar{v})$ solve the dual problem if an only if $\bar{x} = Q^{-1}(\bar{v} - A^T\bar{y} - c)$ solves the primal problem with the primal and dual optimal values coinciding.