# 2 Convex and Linear Optimization

## 2.1 Convexity and Strong Duality

Let $S \subseteq \mathbb{R}^n$. $S$ is called a *convex set* if for all $\delta \in [0,1]$, $x, y \in S$ implies that $\delta x + (1-\delta)y \in S$. A function $f : S \to \mathbb{R}$ is called *convex function* if for all $x, y \in S$ and $\delta \in [0,1]$, $\delta f(x) + (1-\delta)f(y) \geqslant f(\delta x + (1-\delta)y)$. A point $x \in S$ is called an *extreme point* of $S$ if for all $y, z \in S$ and $\delta \in (0,1)$, $x = \delta y + (1-\delta)z$ implies that $x = y = z$. A point $x \in S$ is called an *interior point* of $S$ if there exists $\epsilon > 0$ such that $\{y : \|y - x\|_2 \leqslant \epsilon\} \subseteq S$. The set of all interior points of $S$ is called the *interior* of $S$.

We saw in the previous lecture that strong duality is equivalent to the existence of a supporting hyperplane. The following result establishes a sufficient condition for the latter.

THEOREM 2.1 (Supporting Hyperplane Theorem). *Suppose that $\phi$ is convex and $b \in \mathbb{R}$ lies in the interior of the set of points where $\phi$ is finite. Then there exists a (non-vertical) supporting hyperplane to $\phi$ at $b$.*

The following result identifies a condition that guarantees convexity of $\phi$.

THEOREM 2.2. *Consider the optimization problem to*

$$\begin{aligned} \textit{minimize} \quad & f(x) \\ \textit{subject to} \quad & h(x) \leqslant b \\ & x \in X, \end{aligned}$$

*and let $\phi$ be given by $\phi(b) = \inf_{x \in X(b)} f(x)$. Then, $\phi$ is convex when $X$, $f$, and $h$ are convex.*

*Proof.* Consider $b_1, b_2 \in \mathbb{R}^m$ such that $\phi(b_1)$ and $\phi(b_2)$ are defined, and let $\delta \in [0,1]$ and $b = \delta b_1 + (1-\delta)b_2$. Further consider $x_1 \in X(b_1)$, $x_2 \in X(b_2)$, and let $x = \delta x_1 + (1-\delta)x_2$. Then convexity of $X$ implies that $x \in X$, and convexity of $h$ that

$$\begin{aligned} h(x) &= h(\delta x_1 + (1-\delta)x_2) \\ &\leqslant \delta h(x_1) + (1-\delta)h(x_2) \\ &= \delta b_1 + (1-\delta)b_2 \\ &= b. \end{aligned}$$

Thus $x \in X(b)$, and by convexity of $f$,

$$\phi(b) \leqslant f(x) = f(\delta x_1 + (1-\delta)x_2) \leqslant \delta f(x_1) + (1-\delta)f(x_2).$$

This holds for all $x_1 \in X(b_1)$ and $x_2 \in X(b_2)$, so taking infima on the right hand side yields

$$\phi(b) \leqslant \delta\phi(b_1) + (1-\delta)\phi(b_2).$$ □

Observe that an equality constraint $h(x) = b$ is equivalent to constraints $h(x) \leqslant b$ and $-h(x) \leqslant -b$. In this case, the above result requires that $X$, $f$, $h$, and $-h$ are all convex, which in particular requires that $h$ is linear.

## 2.2 Linear Programs

A *linear program* is an optimization problem in which the objective and all constraints are linear. It has the form

$$\begin{array}{lll} \text{minimize} & c^T x & \\ \text{subject to} & a_i^T x \geqslant b_i, & i \in M_1 \\ & a_i^T x \leqslant b_i, & i \in M_2 \\ & a_i^T x = b_i, & i \in M_3 \\ & x_j \geqslant 0, & j \in N_1 \\ & x_j \leqslant 0, & j \in N_2 \end{array}$$

where $c \in \mathbb{R}^n$ is a cost vector, $x \in \mathbb{R}^n$ is a vector of decision variables, and constraints are given by $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for $i \in \{1, \dots, m\}$. Index sets $M_1, M_2, M_3 \subseteq \{1, \dots, m\}$ and $N_1, N_2 \subseteq \{1, \dots, n\}$ are used to distinguish between different types of contraints.

An equality constraint $a_i^T x = b_i$ is equivalent to the pair of constraints $a_i^T \leqslant b_i$ and $a_i^T x \geqslant b_i$, and a constraint of the form $a_i^T x \leqslant b_i$ can be rewritten as $(-a_i)^T x \geqslant -b_i$. Each occurrence of an unconstrained variable $x_j$ can be replaced by $x_j^+ + x_j^-$, where $x_j^+$ and $x_j^-$ are two new variables with $x_j^+ \geqslant 0$ and $x_j^- \leqslant 0$. We can thus write every linear program in the *general form*

$$\min\{\, c^T x : Ax \geqslant b, x \geqslant 0 \,\} \tag{2.1}$$

where $x, c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$. Observe that constraints of the form $x_j \geqslant 0$ and $x_j \leqslant 0$ are just special cases of constraints of the form $a_i^T x \geqslant b_i$, but we often choose to make them explicit.

A linear program of the form

$$\min\{\, c^T x : Ax = b, x \geqslant 0 \,\} \tag{2.2}$$

is said to be in *standard form*. The standard form is of course a special case of the general form. On the other hand, we can also bring every general form problem into the standard form by replacing each inequality constraint of the form $a_i^T x \leqslant b_i$ or $a_i^T x \geqslant b_i$ by a constraint $a_i^T x + s_i = b_i$ or $a_i^T x - s_i = b_i$, where $s_i$ is a new so-called *slack variable*, and an additional constraint $s_i \geqslant 0$.

The general form is typically used to discuss the theory of linear programming, while the standard form is often more convenient when designing algorithms for linear programming.

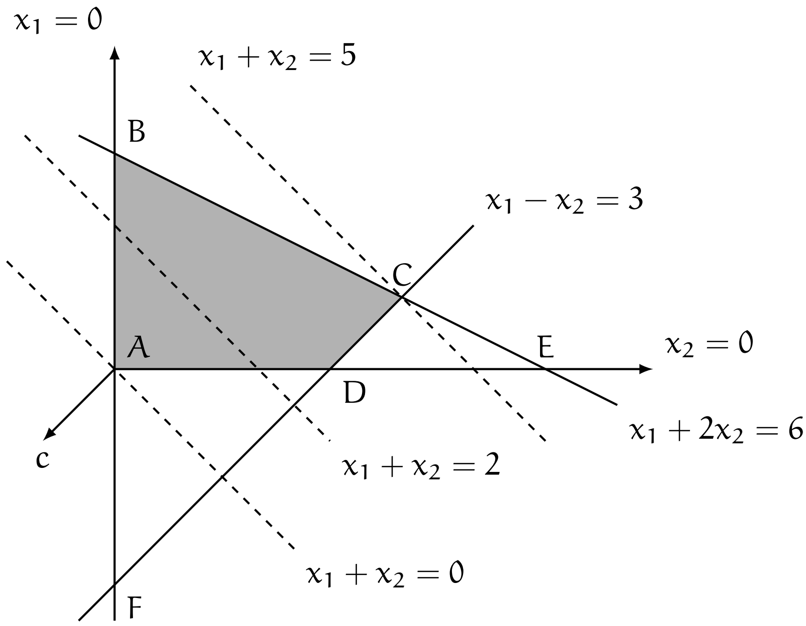


Figure 2.1: Geometric interpretation of the linear program of Example 2.3

Example 2.3. Consider the following linear program, which is illustrated in Figure 2.1:

$$\begin{array}{ll} \text{minimize} & -(x_1 + x_2) \\ \text{subject to} & x_1 + 2x_2 \leqslant 6 \\ & x_1 - x_2 \quad \leqslant 3 \\ & x_1, x_2 \qquad \geqslant 0 \end{array}$$

Solid lines indicate sets of points for which one of the constraints is satisfied with equality. The feasible set is shaded. Dashed lines, orthogonal to the cost vector $c$, indicate sets of points for which the value of the objective function is constant. The optimal value over the feasible set is attained at point $C$.

## 2.3 Linear Program Duality

Consider problem (2.1) and introduce slack variables $z$ to turn it into

$$\min\{\, c^T x : Ax - z = b, x, z \geqslant 0 \,\}.$$

We have $X = \{(x, z) : x \geqslant 0, z \geqslant 0\} \subseteq \mathbb{R}^{m+n}$. The Lagrangian is given by

$$L((x, z), \lambda) = c^T x - \lambda^T(Ax - z - b) = (c^T - \lambda^T A)x + \lambda^T z + \lambda^T b$$

and has a finite minimum over $X$ if and only if

$$\lambda \in Y = \{\, \mu \in \mathbb{R}^m : c^T - \mu^T A \geqslant 0, \mu \geqslant 0 \,\}.$$

For $\lambda \in Y$, the minimum of $L((x,z),\lambda)$ is attained when both $(c^T - \lambda^T A)x = 0$ and $\lambda^T z = 0$, and thus

$$g(\lambda) = \inf_{(x,z)\in X} L((x,z),\lambda) = \lambda^T b.$$

We obtain the dual

$$\max\{\, b^T\lambda : A^T\lambda \leqslant c, \lambda \geqslant 0 \,\}. \tag{2.3}$$

The dual of (2.2) can be determined analogously as

$$\max\{\, b^T\lambda : A^T\lambda \leqslant c \,\}.$$

## 2.4 Complementary Slackness

An important relationship between primal and dual solutions is provided by conditions known as *complementary slackness.* Complementary slackness requires that slack does not occur simultaneously in a variable, of the primal or dual, and the corresponding constraint, of the dual or primal. Here, a variable is said to have slack if its value is non-zero, and an inequality constraint is said to have slack if it does not hold with equality. It is not hard to see that complementary slackness is a necessary condition for optimality. Indeed, if complementary slackness was violated by some variable and the corresponding contraint, reducing the value of the variable would reduce the value of the Lagrangian, contradicting optimality of the current solution. The following result formalizes this intuition.

THEOREM 2.4. *Let $x$ and $\lambda$ be feasible solutions for the primal* (2.1) *and the dual* (2.3)*, respectively. Then $x$ and $\lambda$ are optimal if and only if they satisfy complementary slackness, i.e., if*

$$(c^T - \lambda^T A)x = 0 \quad \text{and} \quad \lambda^T(Ax - b) = 0.$$

*Proof.* If $x$ and $\lambda$ are optimal, then

$$\begin{aligned} c^T x &= \lambda^T b \\ &= \inf_{x'\in X}\big(c^T x' - \lambda^T(Ax' - b)\big) \\ &\leqslant c^T x - \lambda^T(Ax - b) \\ &\leqslant c^T x. \end{aligned}$$

Since the first and last term are the same, the two inequalities must hold with equality. Therefore, $\lambda^T b = c^T x - \lambda^T(Ax - b) = (c^T - \lambda^T A)x + \lambda^T b$, and thus $(c^T - \lambda^T A)x = 0$. Furthermore, $c^T x - \lambda^T(Ax - b) = c^T x$, and thus $\lambda^T(Ax - b) = 0$.

If on the other hand $(c^T - \lambda^T A)x = 0$ and $\lambda^T(Ax - b) = 0$, then

$$c^T x = c^T x - \lambda^T(Ax - b) = (c^T - \lambda^T A)x + \lambda^T b = \lambda^T b,$$

and by weak duality $x$ and $\lambda$ must be optimal. □

## 2.5 Shadow Prices

A more intuitive understanding of Lagrange multipliers can be obtained by again viewing (1.1) as a family of problems parameterized by $b \in \mathbb{R}^m$. As before, let $\phi(b) = \inf\{f(x) : h(x) = b, x \in \mathbb{R}^n\}$. It turns out that at the optimum, the Lagrange multipliers equal the partial derivatives of $\phi$.

THEOREM 2.5. *Suppose that* $f$ *and* $h$ *are continuously differentiable on* $\mathbb{R}^n$*, and that there exist unique functions* $x^* : \mathbb{R}^m \to \mathbb{R}^n$ *and* $\lambda^* : \mathbb{R}^m \to \mathbb{R}^m$ *such that for each* $b \in \mathbb{R}^m$, $h(x^*(b)) = b$, $\lambda^*(b) \leqslant 0$ *and* $f(x^*(b)) = \phi(b) = \inf\{f(x) - \lambda^*(b)^T(h(x) - b) : x \in \mathbb{R}^n\}$*. If* $x^*$ *and* $\lambda^*$ *are continuously differentiable, then*

$$\frac{\partial \phi}{\partial b_i}(b) = \lambda_i^*(b).$$

*Proof.* We have that

$$\begin{aligned}\phi(b) &= f(x^*(b)) - \lambda^*(b)^T(h(x^*(b)) - b) \\ &= f(x^*(b)) - \lambda^*(b)^T h(x^*(b)) + \lambda^*(b)^T b.\end{aligned}$$

Taking partial derivatives of each term,

$$\frac{\partial f(x^*(b))}{\partial b_i} = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x^*(b)) \frac{\partial x_j^*}{\partial b_i}(b),$$

$$\begin{aligned}\frac{\partial \lambda^*(b)^T h(x^*(b))}{\partial b_i} &= \lambda^*(b)^T \frac{\partial h(x^*(b))}{\partial b_i} + h(x^*(b)) \frac{\partial \lambda^*(b)^T}{\partial b_i} \\ &= \left( \sum_{j=1}^{n} \left( \lambda^*(b)^T \frac{\partial h}{\partial x_j}(x^*(b)) \right) \frac{\partial x_j^*}{\partial b_i}(b) \right) + h(x^*(b)) \frac{\partial \lambda^*(b)^T}{\partial b_i},\end{aligned}$$

$$\frac{\partial \lambda^*(b)^T b}{\partial b_i} = \lambda^*(b)^T \frac{\partial b}{\partial b_i} + b \frac{\lambda^*(b)^T}{\partial b_i}.$$

By summing and re-arranging,

$$\begin{aligned}\frac{\partial \phi(b)}{\partial b_i} = \sum_{j=1}^{n} &\left( \frac{\partial f}{\partial x_j}(x^*(b)) - \lambda^*(b)^T \frac{\partial h}{\partial x_j}(x^*(b)) \right) \frac{\partial x_j^*}{\partial b_i}(b) \\ &- (h(x^*(b)) - b) \frac{\partial \lambda^*(b)^T}{\partial b_i} + \lambda^*(b)^T \frac{\partial b}{\partial b_i}.\end{aligned}$$

The first term on the right-hand side is zero, because $x^*(b)$ minimizes $L(x, \lambda^*(b))$ and thus

$$\frac{\partial L(x^*(b), \lambda^*(b))}{\partial x_j} = \frac{\partial f}{\partial x_j}(x^*(b)) - \left( \lambda^*(b)^T \frac{\partial h}{\partial x_j}(x^*(b)) \right) = 0$$

for $j = 1, \dots, n$. The second term is zero as well, because $x^*(b)$ is feasible and thus $(h(x^*(b)) - b)_k = 0$ for $k = 1, \dots, m$, and the claim follows. □

This result continues to hold when the functional constraints are inequalities: if the $i$th constraint is not satisfied with equality, then $\lambda_i^* = 0$ by complementary slackness, and therefore also $\partial\lambda_i^*/\partial b_i = 0$.

In light of Theorem 2.5, Lagrange multipliers are also known as *shadow prices*, due to an economic interpretation of the problem to

$$\begin{array}{ll} \text{maximize} & f(x) \\ \text{subject to} & h(x) \leqslant b \\ & x \in X. \end{array}$$

Consider a firm that produces $n$ different goods from $m$ different raw materials. Vector $b \in \mathbb{R}^m$ describes the amount of each raw material available to the firm, vector $x \in \mathbb{R}^n$ the quantity produced of each good. Functions $h : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^n \to \mathbb{R}$ finally describe the amounts of raw material required to produce, and the profit derived from producing, particular quantities of the goods. The goal in the above problem thus is to maximize the profit of the firm for given amounts of raw materials available to it. The *shadow price* of raw material $i$ then is the price the firm would be willing to pay per additional unit of this raw material, which of course should be equal to the additional profit derived from it, i.e., to $\partial\phi(b)/\partial b_i$. In this context, complementary slackness corresponds to the basic economic principle that a particular raw material has a non-zero price if and only if it is scarce, in the sense that increasing its availability would increase profit.