Part 2

Optimality Conditions and Duality

Notes and References

In this chapter we develop the Kuhn–Tucker conditions for problems with inequality constraints and problems with both equality and inequality constraints. This is done directly by imposing a suitable constraint qualification, as opposed to first developing the Fritz John conditions and then the Kuhn–Tucker conditions.

The Kuhn–Tucker optimality conditions were originally developed by imposing the constraint qualification that for every direction vector **d** in the cone G', there is a feasible arc that points along **d**. Since then, many authors have developed the Kuhn–Tucker conditions under different constraint qualifications. For a thorough study of this subject refer to the works of Abadie [1967b], Arrow, Hurwicz, and Uzawa [1961], Canon, Cullum, and Polak [1966], Cottle [1963a], Evans [1970], Evans and Gould [1970], Guignard [1969], Mangasarian [1969a], Mangasarian and Fromovitz [1967], and Zangwill [1969]. For a comparison and further study of these constraint qualifications, see the survey articles of Bazaraa, Goode, and Shetty [1972], Gould and Tolle [1972], and Peterson [1973].

Gould and Tolle [1971] showed that the constraint qualification of Guignard [1969] is the weakest possible in the sense that it is both necessary and sufficient for the validation of the Kuhn–Tucker conditions.

Chapter 6

Lagrangian Duality and Saddle Point Optimality Conditions

Given a nonlinear programming problem, there is another nonlinear programming problem closely associated with it. The former is called the *primal problem*, and the latter is called the *Lagrangian dual problem*. Under certain convexity assumptions, the primal and dual problems have equal optimal objective values, and hence it is possible to solve the primal problem indirectly by solving the dual problem.

Several properties of the dual problem are developed in this chapter. They are used to provide general solution strategies for solving the primal and dual problems. As a by-product of one of the duality theorems, we obtain saddle point necessary optimality conditions without any differentiability assumptions.

The following is an outline of the chapter.

SECTION 6.1: **The Lagrangian Dual Problem** We introduce the Lagrangian dual problem, give its geometric interpretation, and illustrate it by several numerical examples.

SECTION 6.2: **Duality Theorems and Saddle-Point Optimality** We prove the weak and strong duality theorems. The latter shows that the primal and dual objectives are equal under suitable convexity assumptions.

SECTION 6.3: **Properties of the Dual Function** We study several important properties of the dual function, such as, concavity, differentiability, and subdifferentiability. We then give necessary and sufficient characterizations of ascent and steepest ascent directions.

SECTION 6.4: **Solving the Dual Problem** Several procedures for solving the dual problem are discussed. In particular, we discuss the gradient method, the ascent procedure, and the cutting plane algorithm.

SECTION 6.5: **Getting the Primal Solution** We show that the points generated during the course of solving the dual problem yield optimal solutions to perturbations of the primal problem. For convex programs, we show how to obtain primal feasible solutions that are near-optimal.

SECTION 6.6: **Linear and Quadratic Programs** We give the Lagrangian dual formulation for linear and quadratic programming.

6.1 The Lagrangian Dual Problem

Consider the following nonlinear programming problem P, which is called the *primal problem.*

Primal Problem P

Minimize	$f(\mathbf{x})$	
subject to	$g_i(\mathbf{x}) \leq 0$	for $i = 1, \ldots, m$
	$h_i(\mathbf{x}) = 0$	for $i = 1, \ldots, l$
	$\mathbf{x} \in X$	

Several problems, closely related to the above primal problem, have been proposed in the literature and are called *dual problems.* Among the various duality formulations, the Lagrangian duality formulation has perhaps attracted the most attention. It has led to several algorithms for solving large-scale linear problems, as well as convex and nonconvex nonlinear problems. More recently, it has proved useful in discrete optimization where all or some of the variables are further restricted to be integers. The *Lagrangian dual problem D* is presented below.

Lagrangian Dual Problem D

Maximize $\theta(\mathbf{u}, \mathbf{v})$
subject to $\mathbf{u} \geq \mathbf{0}$
where $\theta(\mathbf{u}, \mathbf{v}) = \inf\{f(\mathbf{x}) + \sum_{i=1}^m u_i g_i(\mathbf{x}) + \sum_{i=1}^l v_i h_i(\mathbf{x}) : \mathbf{x} \in X\}$

Note that the *Lagrangian dual function* θ may assume the value of $-\infty$ for some vector $(\mathbf{u}, \mathbf{v})$. In the expression for $\theta(\mathbf{u}, \mathbf{v})$, the constraints $g_i(\mathbf{x}) \leq 0$ and $h_i(\mathbf{x}) = 0$ have been incorporated in the objective function using the *Lagrangian multipliers* u_i and v_i. Also note that the multiplier u_i associated with the inequality constraint $g_i(\mathbf{x}) \leq 0$ is nonnegative, whereas the multiplier v_i associated with the equality constraint $h_i(\mathbf{x}) = 0$ is unrestricted in sign.

Since the dual problem consists of maximizing the infimum (greatest lower bound) of the function $f(\mathbf{x}) + \sum_{i=1}^m u_i g_i(\mathbf{x}) + \sum_{i=1}^l v_i h_i(\mathbf{x})$, it is sometimes referred to as the *max-min dual problem.*

The primal and Lagrangian dual problems could be written in the following form using vector notation, where $f: E_n \to E_1$, $\mathbf{g}: E_n \to E_m$ is a vector function whose ith component is g_i, and $\mathbf{h}: E_n \to E_l$ is a vector function whose ith component is h_i. For the sake of convenience, we will use this form throughout the remainder of this chapter.

Primal Problem P

Minimize	$f(\mathbf{x})$
subject to	$\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$
	$\mathbf{h}(\mathbf{x}) = \mathbf{0}$
	$\mathbf{x} \in X$

Lagrangian Dual Problem D

Maximize $\theta(\mathbf{u}, \mathbf{v})$
subject to $\mathbf{u} \geq \mathbf{0}$
where $\theta(\mathbf{u}, \mathbf{v}) = \inf\{f(\mathbf{x}) + \mathbf{u}'\mathbf{g}(\mathbf{x}) + \mathbf{v}'\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\}$.

Given a nonlinear programming problem, several Lagrangian dual problems can be devised, depending on which constraints are handled as $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ and which constraints are treated by the set X. The choice would affect the effort expended in evaluating and updating the dual function θ during the course of solving the dual problem. Hence, an appropiate selection of the set X would depend on the structure of the problem.

Geometric Interpretation of the Dual Problem

We now briefly discuss the geometric interpretation of the dual problem. For the sake of simplicity, we will consider only one inequality constraint and assume that no equality constraints exist. Then, the primal problem is to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in X$ and $g(\mathbf{x}) \leq 0$.

In the (z_1, z_2) plane, the set $\{(z_1, z_2) : z_1 = g(\mathbf{x}),\ z_2 = f(\mathbf{x})$ for some $\mathbf{x} \in X\}$ is denoted by G in Figure 6.1. Then, G is the image of X under the (g, f) map. The primal problem asks us to find a point in G to the left of the z_2 axis with minimum ordinate. Obviously, this point is $(\bar{z}_1, \bar{z}_2)$ in Figure 6.1.

Now suppose that $u \geq 0$ is given. To determine $\theta(u)$, we need to minimize $f(\mathbf{x}) + ug(\mathbf{x})$ over all $\mathbf{x} \in X$. Letting $z_1 = g(\mathbf{x})$ and $z_2 = f(\mathbf{x})$ for $\mathbf{x} \in X$, we want to

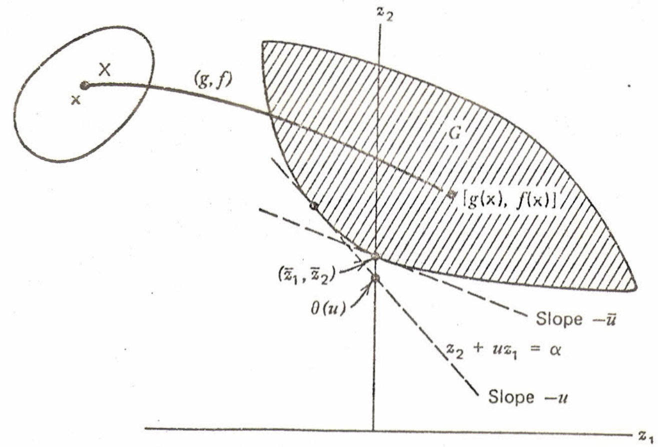

Figure 6.1 Geometric interpretation of Lagrangian duality.

minimize $z_2 + uz_1$ over points in G. Note that $z_2 + uz_1 = \alpha$ is an equation of a straight line with slope $-u$ and intercept α on the z_2 axis. In order to minimize $z_2 + uz_1$ over G, we need to move the line $z_2 + uz_1 = \alpha$ parallel to itself as far down as possible so that it supports G. In other words, the set G is above the line and touches it. Then the intercept on the z_2 axis gives $\theta(u)$, as seen in Figure 6.1. The dual problem is therefore equivalent to finding the slope of the supporting hyperplane such that its intercept on the z_2 axis is maximal. In Figure 6.1, such a hyperplane has slope $-\bar{u}$ and supports the set G at the point $(\bar{z}_1, \bar{z}_2)$. Thus the optimal dual solution is $\bar{u}$, and the optimal dual objective value is $\bar{z}_2$. Furthermore, the optimal primal and dual objectives are equal.

6.1.1 Example

Consider the following primal problem:

$$\begin{aligned} \text{Minimize} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & -x_1 - x_2 + 4 \le 0 \\ & x_1, x_2 \ge 0 \end{aligned}$$

Note that the optimal solution occurs at the point $(x_1, x_2) = (2, 2)$, whose objective is equal to 8.

Letting $g(x) = -x_1 - x_2 + 4$ and $X = \{(x_1, x_2) : x_1, x_2 \ge 0\}$, the dual function is given by

$$\begin{aligned} \theta(u) &= \inf\{x_1^2 + x_2^2 + u(-x_1 - x_2 + 4) : x_1, x_2 \ge 0\} \\ &= \inf\{x_1^2 - ux_1 : x_1 \ge 0\} + \inf\{x_2^2 - ux_2 : x_2 \ge 0\} + 4u \end{aligned}$$

Note that the above infima are achieved at $x_1 = x_2 = u/2$ if $u \ge 0$ and at

$x_1 = x_2 = 0$ if $u < 0$. Hence,

$$\theta(u) = \begin{cases} -\frac{1}{2}u^2 + 4u & \text{for } u \ge 0 \\ 4u & \text{for } u < 0 \end{cases}$$

Note that θ is a concave function, and its maximum over $u \ge 0$ occurs at $\bar{u} = 4$. Note also that the optimal primal and dual objectives are both equal to 8.

Now let us consider the problem in the (z_1, z_2) plane, where $z_1 = g(x)$ and $z_2 = f(x)$. We are interested in finding G, the image of $X = \{(x_1, x_2) : x_1 \ge 0, x_2 \ge 0\}$, under the (g, f) map. We do this by deriving explicit expressions for the lower and upper envelopes of G, denoted, respectively, by α and β.

Given z_1, note that $\alpha(z_1)$ and $\beta(z_1)$ are the optimal objective values of the following problems P_1 and P_2, respectively.

Problem P_1		*Problem* P_2	
Minimize	$x_1^2 + x_2^2$	Maximize	$x_1^2 + x_2^2$
subject to	$-x_1 - x_2 + 4 = z_1$	subject to	$-x_1 - x_2 + 4 = z_1$
	$x_1, x_2 \ge 0$		$x_1, x_2 \ge 0$

The reader can verify that $\alpha(z_1) = (4 - z_1)^2/2$ and $\beta(z_1) = (4 - z_1)^2$ for $z_1 \le 4$. The set G is illustrated in Figure 6.2. Note that $\mathbf{x} \in X$ implies that $x_1, x_2 \ge 0$, so that $-x_1 - x_2 + 4 \le 4$. Thus, every point $\mathbf{x} \in X$ corresponds to $z_1 \le 4$.

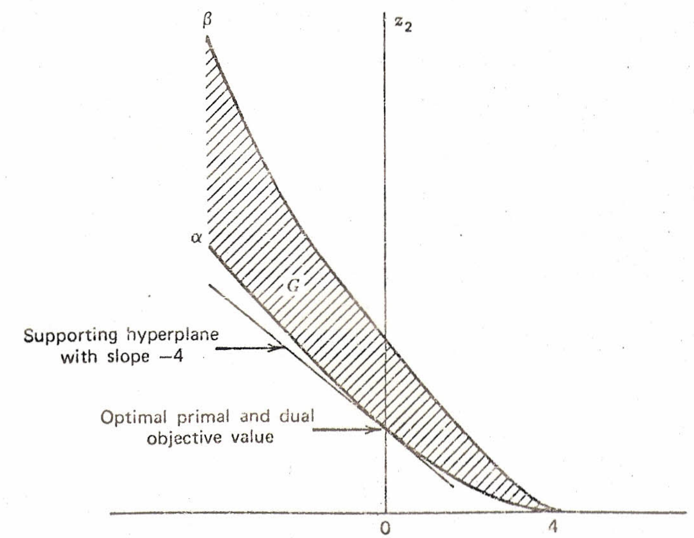

Figure 6.2 Geometric illustration of Example 6.1.1.

Note that the optimal dual solution is $\bar{u}=4$, which is the negative of the slope of the supporting hyperplane shown in Figure 6.2. The optimal dual objective is $\alpha(0)=8$ and is equal to the optimal primal objective.

6.2 Duality Theorems and Saddle Point Optimality

In this section we investigate the relationships between the primal and dual problems and develop saddle point optimality conditions for the primal problem.

Theorem 6.2.1 below, referred to as the *weak duality theorem*, shows that the objective value of any feasible solution to the dual problem yields a lower bound on the objective value of any feasible solution to the primal problem. Several important results follow as corollaries.

6.2.1 Theorem (Weak Duality Theorem)

Let $\mathbf{x}$ be a feasible solution to *Problem P*, that is $\mathbf{x}\in X$, $\mathbf{g}(\mathbf{x})\leq \mathbf{0}$, and $\mathbf{h}(\mathbf{x})=\mathbf{0}$. Also let $(\mathbf{u},\mathbf{v})$ be a feasible solution to *Problem D*, that is, $\mathbf{u}\geq \mathbf{0}$. Then $f(\mathbf{x})\geq \theta(\mathbf{u},\mathbf{v})$.

Proof

By definition of θ, and since $\mathbf{x}\in X$, we have

$$\begin{aligned}\theta(\mathbf{u},\mathbf{v}) &= \inf\{f(\mathbf{y})+\mathbf{u}^t\mathbf{g}(\mathbf{y})+\mathbf{v}^t\mathbf{h}(\mathbf{y}):\mathbf{y}\in X\}\\ &\leq f(\mathbf{x})+\mathbf{u}^t\mathbf{g}(\mathbf{x})+\mathbf{v}^t\mathbf{h}(\mathbf{x})\leq f(\mathbf{x})\end{aligned}$$

since $\mathbf{u}\geq\mathbf{0}$, $\mathbf{g}(\mathbf{x})\leq\mathbf{0}$, and $\mathbf{h}(\mathbf{x})=\mathbf{0}$. This completes the proof.

Corollary 1

$$\inf\{f(\mathbf{x}):\mathbf{x}\in X, \mathbf{g}(\mathbf{x})\leq\mathbf{0}, \mathbf{h}(\mathbf{x})=\mathbf{0}\}\geq\sup\{\theta(\mathbf{u},\mathbf{v}):\mathbf{u}\geq\mathbf{0}\}$$

Corollary 2

If $f(\bar{\mathbf{x}})\leq\theta(\bar{\mathbf{u}},\bar{\mathbf{v}})$, where $\bar{\mathbf{u}}\geq\mathbf{0}$ and $\bar{\mathbf{x}}\in\{\mathbf{x}\in X:\mathbf{g}(\mathbf{x})\leq\mathbf{0}, \mathbf{h}(\mathbf{x})=\mathbf{0}\}$, then $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}},\bar{\mathbf{v}})$ solve the primal and dual problems, respectively.

Corollary 3

If $\inf\{f(\mathbf{x}):\mathbf{x}\in X, \mathbf{g}(\mathbf{x})\leq\mathbf{0}, \mathbf{h}(\mathbf{x})=\mathbf{0}\}=-\infty$, then $\theta(\mathbf{u},\mathbf{v})=-\infty$ for each $\mathbf{u}\geq\mathbf{0}$.

Corollary 4

If $\sup\{\theta(\mathbf{u},\mathbf{v}):\mathbf{u}\geq\mathbf{0}\}=\infty$, then the primal problem has no feasible solution.

Duality Gap

From Corollary 1 to Theorem 6.2.1 above, the optimal objective value of the primal problem is greater than or equal to the optimal objective value of the dual problem. If strict inequality holds true then a *duality gap* is said to exist. Figure 6.3 illustrates the case of a duality gap for a problem with a single inequality constraint and no equality constraints.

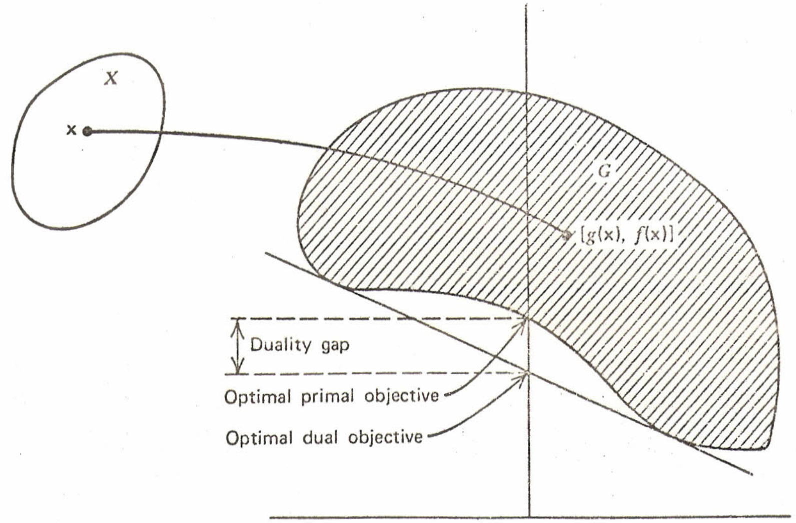

Figure 6.3 Illustration of a duality gap.

6.2.2 Example

Consider the following problem:

Minimize $-2x_1+x_2$

subject to $x_1+x_2-3=0$

$(x_1,x_2)\in X$

where $X=\{(0,0),(0,4),(4,4),(4,0),(1,2),(2,1)\}$.

It is easy to verify that $(2,1)$ is the optimal solution to the primal problem with objective value equal to -3. The dual objective function θ is given by

$$\theta(v)=\text{minimum}\{(-2x_1+x_2)+v(x_1+x_2-3):(x_1,x_2)\in X\}$$

The reader may verify that the explicit expression for θ is given by

$$\theta(v)=\begin{cases}-4+5v & \text{for } v\leq -1\\ -8+v & \text{for } -1\leq v\leq 2\\ -3v & \text{for } v\geq 2\end{cases}$$

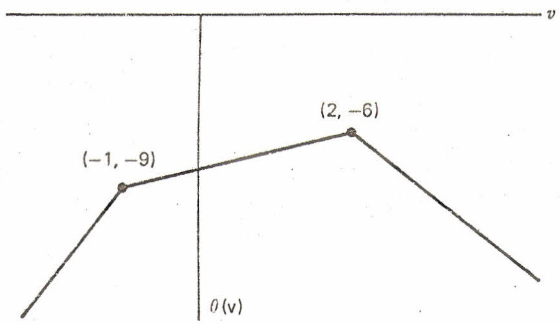

Figure 6.4 Dual function for Example 6.2.2.

The dual function is shown in Figure 6.4, and the optimal solution is $\bar{v}=2$ with objective value -6. Note, in this example, that there exists a duality gap.

In this case the set G consists of finite number of points, each corresponding to a point in X. This is shown in Figure 6.5. The supporting hyperplane, whose intercept on the vertical axis is maximal, is shown in the figure. Note that the intercept is equal to -6 and that the slope is equal to -2. Thus the optimal dual solution is $\bar{v}=2$ with objective value -6. Furthermore, note that the points in the set G on the vertical axis correspond to the primal feasible points, and hence the minimal primal objective value is equal to -3.

Conditions that guarantee the absence of a duality gap are given in Theorem 6.2.4. First, however, the following lemma is needed.

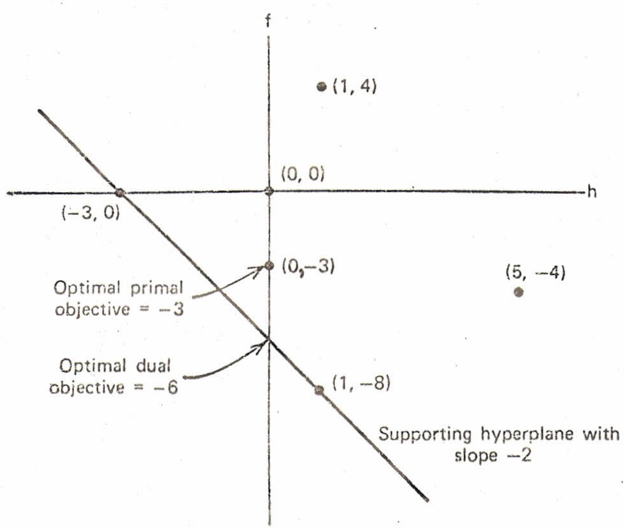

Figure 6.5 Geometric interpretation of Example 6.2.2.

6.2.3 Lemma

Let X be a nonempty convex set in E_n. Let $\alpha: E_n \to E_1$ and $\mathbf{g}: E_n \to E_m$ be convex, and let $\mathbf{h}: E_n \to E_l$ be affine; that is, h is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{Ax} - \mathbf{b}$. If *System 1* below has no solution $\mathbf{x}$, then *System 2* has a solution $(u_0, \mathbf{u}, \mathbf{v})$. The converse holds if $u_0 > 0$.

System 1: $\alpha(\mathbf{x}) < 0, \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}$ for some $\mathbf{x} \in X$

System 2: $u_0\alpha(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \ge 0$ for all $\mathbf{x} \in X$

$(u_0, \mathbf{u}) \ge \mathbf{0}, \quad (u_0, \mathbf{u}, \mathbf{v}) \ne \mathbf{0}$

Proof

Suppose that *System 1* has no solution, and consider the following set:

$$\Lambda = \{(p, \mathbf{q}, \mathbf{r}) : p > \alpha(\mathbf{x}), \quad \mathbf{q} \ge \mathbf{g}(\mathbf{x}), \quad \mathbf{r} = \mathbf{h}(\mathbf{x}) \quad \text{for some } \mathbf{x} \in X\}$$

Noting that X, α, and $\mathbf{g}$ are convex and that $\mathbf{h}$ is affine, it can easily be shown that Λ is convex. Since *System 1* has no solution, then $(0, \mathbf{0}, \mathbf{0}) \notin \Lambda$. By the corollary to Theorem 2.3.7, there exists a nonzero $(u_0, \mathbf{u}, \mathbf{v})$ such that

$$u_0 p + \mathbf{u}^t\mathbf{q} + \mathbf{v}^t\mathbf{r} \ge 0 \quad \text{for each } (p, \mathbf{q}, \mathbf{r}) \in \text{cl}\,\Lambda \tag{6.1}$$

Now, fix an $\mathbf{x} \in X$. Since p and $\mathbf{q}$ can be made arbitrarily large, (6.1) holds true only if $u_0 \ge 0$ and $\mathbf{u} \ge \mathbf{0}$. Furthermore, $(p, \mathbf{q}, \mathbf{r}) = [\alpha(\mathbf{x}), \mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x})]$ belongs to cl Λ. Therefore, from (6.1), we get

$$u_0\alpha(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \ge 0$$

Since the above inequality is true for each $\mathbf{x} \in X$, *System 2* has a solution.

To prove the converse, assume that *System 2* has a solution $(u_0, \mathbf{u}, \mathbf{v})$ such that $u_0 > 0$ and $\mathbf{u} \ge \mathbf{0}$, satisfying

$$u_0\alpha(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \ge 0 \quad \text{for each } \mathbf{x} \in X$$

Now let $\mathbf{x} \in X$ be such that $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. From the above inequality, since $\mathbf{u} \ge \mathbf{0}$, we conclude that $u_0\alpha(\mathbf{x}) \ge 0$. Since $u_0 > 0$, $\alpha(\mathbf{x}) \ge 0$, and hence *System 1* has no solution. This completes the proof.

Theorem 6.2.4 below, referred to as the *strong duality theorem*, shows that under suitable convexity assumptions and under a constraint qualification, the optimal objective function values of the primal and dual problems are equal.

6.2.4 Theorem (Strong Duality Theorem)

Let X be a nonempty convex set in E_n, let $f: E_n \to E_1$ and $\mathbf{g}: E_n \to E_m$ be convex, and let $\mathbf{h}: E_n \to E_l$ be affine; that is, $\mathbf{h}$ is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{Ax} - \mathbf{b}$.

Suppose that the following constraint qualification holds true. There exists an $\hat{\mathbf{x}} \in X$ such that $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$, and $\mathbf{0} \in \operatorname{int} \mathbf{h}(X)$, where $\mathbf{h}(X) = \{\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\}$. Then

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\} = \sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \ge \mathbf{0}\} \tag{6.2}$$

Furthermore, if the inf is finite, then $\sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \ge \mathbf{0}\}$ is achieved at $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{u}} \ge \mathbf{0}$. If the inf is achieved at $\bar{\mathbf{x}}$, then $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) = 0$.

Proof

Let $\gamma = \inf\{f(\mathbf{x}) : \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$. If $\gamma = -\infty$, then by Corollary 3 to Theorem 6.2.1, $\sup\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \ge \mathbf{0}\} = -\infty$, and hence (6.2) holds true. Now suppose that γ is finite, and consider the following system:

$$f(\mathbf{x}) - \gamma < 0, \qquad \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \qquad \mathbf{h}(\mathbf{x}) = \mathbf{0}, \qquad \mathbf{x} \in X$$

By definition of γ, this system has no solution. Hence, from Lemma 6.2.3, there exists a nonzero vector $(u_0, \mathbf{u}, \mathbf{v})$ with $(u_0, \mathbf{u}) \ge \mathbf{0}$ such that

$$u_0[f(\mathbf{x}) - \gamma] + \mathbf{u}'\mathbf{g}(\mathbf{x}) + \mathbf{v}'\mathbf{h}(\mathbf{x}) \ge 0 \qquad \text{for all } \mathbf{x} \in X \tag{6.3}$$

We first show that $u_0 > 0$. By contradiction, suppose that $u_0 = 0$. By assumption, there exists an $\hat{\mathbf{x}} \in X$ such that $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$. Substituting in (6.3), it follows that $\mathbf{u}'\mathbf{g}(\hat{\mathbf{x}}) \ge 0$. Since $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{u} \ge \mathbf{0}$, $\mathbf{u}'\mathbf{g}(\hat{\mathbf{x}}) \ge 0$ is only possible if $\mathbf{u} = \mathbf{0}$. But from (6.3), $u_0 = 0$ and $\mathbf{u} = \mathbf{0}$, which implies that $\mathbf{v}'\mathbf{h}(\mathbf{x}) \ge 0$ for all $\mathbf{x} \in X$. But since $\mathbf{0} \in \operatorname{int} \mathbf{h}(X)$, we can pick an $\mathbf{x} \in X$ such that $\mathbf{h}(\mathbf{x}) = -\lambda \mathbf{v}$, where $\lambda > 0$. Therefore, $0 \le \mathbf{v}'\mathbf{h}(\mathbf{x}) = -\lambda \|\mathbf{v}\|^2$, which implies that $\mathbf{v} = \mathbf{0}$. Thus, we have shown that $u_0 = 0$ implies that $(u_0, \mathbf{u}, \mathbf{v}) = \mathbf{0}$, which is impossible. Hence, $u_0 > 0$. Dividing (6.3) by u_0 and denoting $\mathbf{u}/u_0$ and $\mathbf{v}/u_0$ by $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$, respectively, we get

$$f(\mathbf{x}) + \bar{\mathbf{u}}'\mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}'\mathbf{h}(\mathbf{x}) \ge \gamma \qquad \text{for all } \mathbf{x} \in X \tag{6.4}$$

This shows that $\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \inf\{f(\mathbf{x}) + \bar{\mathbf{u}}'\mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}'\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\} \ge \gamma$. In view of Theorem 6.2.1, it is then clear that $\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = \gamma$, and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ solves the dual problem.

To complete the proof, suppose that $\bar{\mathbf{x}}$ is an optimal solution to the primal problem, that is, $\bar{\mathbf{x}} \in X$, $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and $f(\bar{\mathbf{x}}) = \gamma$. From (6.4), letting $\mathbf{x} = \bar{\mathbf{x}}$, we get $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) \ge 0$. Since $\bar{\mathbf{u}} \ge \mathbf{0}$ and $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) = 0$, and the proof is complete.

In the above theorem, the assumption $\mathbf{0} \in \operatorname{int} \mathbf{h}(X)$ and that there exists an $\hat{\mathbf{x}} \in X$ such that $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$ can be viewed as a generalization of the Slater constraint qualification of Chapter 5. In particular, if $X = E_n$, then $\mathbf{0} \in \operatorname{int} \mathbf{h}(X)$ automatically holds true, so that the constraint qualification asserts the existance of a point $\hat{\mathbf{x}}$ such that $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$. To see this, suppose that $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. Without loss of generality, assume that rank $\mathbf{A} = m$, because otherwise any redundant constraints could be deleted. Now, any $\mathbf{y} \in E_m$ could be represented as $\mathbf{y} = \mathbf{A}\mathbf{x} - \mathbf{b}$, where $\mathbf{x} = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}(\mathbf{y} + \mathbf{b})$. Thus, $\mathbf{h}(X) = E_m$, and in particular, $\mathbf{0} \in \operatorname{int} \mathbf{h}(X)$.

Saddle Point Criteria

As a consequence of Theorem 6.2.4, we develop the well-known saddle point optimality criteria. Note that the necessary part of the criteria requires convexity plus a constraint qualification, whereas the sufficiency part of the theorem needs no such assumptions.

6.2.5 Theorem (Saddle Point Theorem)

Let X be a nonempty set in E_n, and let $f : E_n \to E_1$, $\mathbf{g} : E_n \to E_m$, and $\mathbf{h} : E_n \to E_l$. Suppose that there exist $\bar{\mathbf{x}} \in X$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{u}} \ge \mathbf{0}$, such that

$$\phi(\bar{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \le \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \le \phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \tag{6.5}$$

for all $\mathbf{x} \in X$ and all $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \ge \mathbf{0}$, where $\phi(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}'\mathbf{g}(\mathbf{x}) + \mathbf{v}'\mathbf{h}(\mathbf{x})$. Then $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ solve the primal *Problem P* and the dual *Problem D*, respectively. Conversely, suppose that X, f, and $\mathbf{g}$ are convex and that $\mathbf{h}$ is affine; that is, $\mathbf{h}$ is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. Further, suppose that $\mathbf{0} \in \operatorname{int} \mathbf{h}(X)$ and that there exists an $\hat{\mathbf{x}} \in X$ with $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$. If $\bar{\mathbf{x}}$ is an optimal solution to the primal *Problem P*, then there exists $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{u}} \ge \mathbf{0}$, so that (6.5) holds true.

Proof

Suppose that there exist $\bar{\mathbf{x}} \in X$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{u}} \ge \mathbf{0}$ such that (6.5) holds true. Since

$$f(\bar{\mathbf{x}}) + \mathbf{u}'\mathbf{g}(\bar{\mathbf{x}}) + \mathbf{v}'\mathbf{h}(\bar{\mathbf{x}}) = \phi(\bar{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \le \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$$

for all $\mathbf{u} \ge \mathbf{0}$ and all $\mathbf{v} \in E_l$, it follows that $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$ and $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$. Therefore, $\bar{\mathbf{x}}$ is a feasible solution to *Problem P*. Also by letting $\mathbf{u} = \mathbf{0}$ in the above inequality, it follows that $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) \ge 0$. Since $\bar{\mathbf{u}} \ge \mathbf{0}$ and $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, then $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) = 0$. Noting (6.5), then for each $\mathbf{x} \in X$, we get

$$\begin{aligned} f(\bar{\mathbf{x}}) &= f(\bar{\mathbf{x}}) + \bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) + \bar{\mathbf{v}}'\mathbf{h}(\bar{\mathbf{x}}) \\ &= \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \\ &\le \phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \\ &= f(\mathbf{x}) + \bar{\mathbf{u}}'\mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}'\mathbf{h}(\mathbf{x}) \end{aligned} \tag{6.6}$$

Since (6.6) holds true for each $\mathbf{x} \in X$, it then follows that $f(\bar{\mathbf{x}}) \le \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$. Noting that $\bar{\mathbf{x}}$ is feasible to the primal *Problem P* and that $\bar{\mathbf{u}} \ge \mathbf{0}$, from Corollary 2 to

Theorem 6.2.1, it then follows that $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ are optimal to the primal and dual problems, respectively.

Conversely, suppose that $\bar{\mathbf{x}}$ is an optimal solution to the primal *Problem P*. By Theorem 6.2.4, there exists $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{u}} \geq \mathbf{0}$ such that $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ and $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$. By definition of θ, we must have

$$f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) \leq f(\mathbf{x}) + \bar{\mathbf{u}}^t\mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}^t\mathbf{h}(\mathbf{x}) = \phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \qquad \text{for each } \mathbf{x} \in X$$

But since $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = \bar{\mathbf{v}}^t\mathbf{h}(\bar{\mathbf{x}}) = 0$,

$$\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\bar{\mathbf{x}}) + \bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) + \bar{\mathbf{v}}^t\mathbf{h}(\bar{\mathbf{x}}) \leq \phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \qquad \text{for all } \mathbf{x} \in X$$

which is the second inequality in (6.5). The first inequality in (6.5) holds true trivially by noting that $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, $\mathbf{g}(\bar{\mathbf{x}}) \leq \mathbf{0}$, and $\mathbf{u} \geq \mathbf{0}$. This completes the proof.

Relationship Between the Saddlepoint Criteria and the Kuhn–Tucker Conditions

In Chapters 4 and 5 we discussed the Kuhn–Tucker optimality conditions for *Problem P* defined below.

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ & \mathbf{x} \in X \end{aligned}$$

Furthermore, in Theorem 6.2.5 above, we developed the saddle-point optimality conditions for the same problem. Theorem 6.2.6 below gives the relationship between these two types of optimality conditions.

6.2.6 Theorem

Let $S = \{\mathbf{x} \in X : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, and consider *Problem P* to minimize $f(\mathbf{x})$ subject to $\mathbf{x} \in S$. Suppose that $\bar{\mathbf{x}} \in S$ satisfies the Kuhn–Tucker conditions, that is, there exist $\bar{\mathbf{u}} \geq \mathbf{0}$ and $\bar{\mathbf{v}}$ such that

$$\begin{aligned} \nabla f(\bar{\mathbf{x}}) + \nabla\mathbf{g}(\bar{\mathbf{x}})\bar{\mathbf{u}} + \nabla\mathbf{h}(\bar{\mathbf{x}})\bar{\mathbf{v}} &= \mathbf{0} \\ \bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) &= 0 \end{aligned} \tag{6.7}$$

Suppose that f, g_i for $i \in I$ are convex at $\bar{\mathbf{x}}$, where $I = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. Further suppose that if $\bar{v}_i \neq 0$, then h_i is affine. Then, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ satisfy the saddle point conditions

$$\phi(\bar{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \leq \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \leq \phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \tag{6.8}$$

for all $\mathbf{x} \in X$ and for all $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \geq \mathbf{0}$, where $\phi(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x})$.

Conversely, suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{x}} \in \text{int } X$ and $\bar{\mathbf{u}} \geq \mathbf{0}$ satisfy the saddlepoint conditions (6.8). Then $\bar{\mathbf{x}}$ is feasible to *Problem P* and furthermore, $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ satisfy the Kuhn–Tucker conditions specified by (6.7).

Proof

Suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{x}} \in S$ and $\bar{\mathbf{u}} \geq \mathbf{0}$ satisfy the Kuhn–Tucker conditions specified by (6.7). By convexity at $\bar{\mathbf{x}}$ of f and g_i for $i \in I$, and since h_i is affine for $\bar{v}_i \neq 0$, we get

$$f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \tag{6.9}$$

$$g_i(\mathbf{x}) \geq g_i(\bar{\mathbf{x}}) + \nabla g_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \qquad \text{for } i \in I \tag{6.10}$$

$$h_i(\mathbf{x}) = h_i(\bar{\mathbf{x}}) + \nabla h_i(\bar{\mathbf{x}})^t(\mathbf{x} - \bar{\mathbf{x}}) \qquad \text{for } i = 1, \ldots, l,\ \bar{v}_i \neq 0 \tag{6.11}$$

for all $\mathbf{x} \in X$. Multiplying (6.10) by $\bar{u}_i \geq 0$, (6.11) by $\bar{v}_i$, adding (6.9) and noting (6.7), it follows from the definition of ϕ that $\phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \geq \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for all $\mathbf{x} \in X$. Also, since $\mathbf{g}(\bar{\mathbf{x}}) \leq \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$ and $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$, it follows that $\phi(\bar{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \leq \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for all $\mathbf{u} \geq \mathbf{0}$. Hence $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ satisfy the saddlepoint conditions given by (6.8).

To prove the converse, suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ with $\bar{\mathbf{x}} \in \text{int } X$ and $\bar{\mathbf{u}} \geq \mathbf{0}$ satisfy (6.8). Since $\phi(\bar{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \leq \phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for all $\mathbf{u} \geq \mathbf{0}$ and all $\mathbf{v}$, the reader can easily verify that $\mathbf{g}(\bar{\mathbf{x}}) \leq \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$ and $\bar{\mathbf{u}}^t\mathbf{g}(\bar{\mathbf{x}}) = 0$. This shows that $\bar{\mathbf{x}}$ is feasible to *Problem P*. Since $\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) \leq \phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ for all $\mathbf{x} \in X$, then $\bar{\mathbf{x}}$ solves the problem to minimize $\phi(\mathbf{x}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$ subject to $\mathbf{x} \in X$. Since $\bar{\mathbf{x}} \in \text{int } X$, then $\nabla_x\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}}) = \mathbf{0}$, that is $\nabla f(\bar{\mathbf{x}}) + \nabla\mathbf{g}(\bar{\mathbf{x}})\bar{\mathbf{u}} + \nabla\mathbf{h}(\bar{\mathbf{x}})\bar{\mathbf{v}} = \mathbf{0}$, and hence (6.7) holds. This completes the proof.

Theorem 6.2.6 above shows that if $\bar{\mathbf{x}}$ is a Kuhn–Tucker point, under certain convexity assumptions, the Lagrangian multipliers in the Kuhn–Tucker conditions also serve as the multipliers in the saddlepoint criteria. Conversely, the multipliers in the saddlepoint conditions are the Lagrangian multipliers of the Kuhn–Tucker conditions. Moreover, in view of Theorems 6.2.4, 6.2.5 and 6.2.6 the optimal dual variables for the Lagrangian dual problem are precisely the Lagrangian multipliers for the Kuhn–Tucker conditions and also the multipliers for the saddlepoint conditions.

6.3 Properties of the Dual Function

In Section 6.2 we studied the relationships between the primal and dual problems. Under certain conditions, Theorem 6.2.4 showed that the optimal objectives of the primal and dual problems are equal and, hence, it would be possible to solve the primal problem indirectly by solving the dual problem. In order to facilitate the solution of the dual problem, we need to examine the properties of the dual function. In particular, we show that θ is concave, discuss its differentiability and subdifferentiability properties, and characterize its ascent and steepest ascent directions.

Throughout the rest of this chapter, we will assume that the set X is compact. This will simplify the proofs of several of the theorems. Note that this assumption is not unduly restrictive, since if X were not bounded, one could add suitable lower and upper bounds on the variables such that the feasible region would not be affected. For convenience, we will also combine the vectors $\mathbf{u}$ and $\mathbf{v}$ as $\mathbf{w}$ and the functions $\mathbf{g}$ and $\mathbf{h}$ as $\boldsymbol{\beta}$. Theorem 6.3.1 below shows that θ is concave.

6.3.1 Theorem

Let X be a nonempty compact set in E_n, and let $f: E_n \to E_1$, and $\boldsymbol{\beta}: E_n \to E_{m+l}$ be continuous. Then θ, defined by

$$\theta(\mathbf{w}) = \inf \{f(\mathbf{x}) + \mathbf{w}'\boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\}$$

is concave over E_{m+l}.

Proof

Since f and $\boldsymbol{\beta}$ are continuous and X is compact, θ is finite everywhere on E_{m+l}. Let $\mathbf{w}_1, \mathbf{w}_2 \in E_{m+l}$, and let $\lambda \in (0, 1)$. We then have

$$\begin{aligned}
\theta[\lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2] &= \inf \{f(\mathbf{x}) + [\lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2]'\boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\} \\
&= \inf \{\lambda[f(\mathbf{x}) + \mathbf{w}_1'\boldsymbol{\beta}(\mathbf{x})] + (1-\lambda)[f(\mathbf{x}) + \mathbf{w}_2'\boldsymbol{\beta}(\mathbf{x})] : \mathbf{x} \in X\} \\
&\geq \lambda \inf \{f(\mathbf{x}) + \mathbf{w}_1'\boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\} \\
&\quad + (1-\lambda) \inf \{f(\mathbf{x}) + \mathbf{w}_2'\boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\} \\
&= \lambda\theta(\mathbf{w}_1) + (1-\lambda)\theta(\mathbf{w}_2)
\end{aligned}$$

Thus θ is concave, and the proof is complete.

Since θ is concave, by Theorem 3.4.2, a local optimal of θ is also a global optimal. This makes the maximization of θ an attractive proposition. However, the main difficulty in solving the dual problem is that the dual function is not explicitly available, since θ could be evaluated at a point only after a minimization subproblem is solved. In the remainder of this section, we study differentiability and subdifferentiability properties of the dual function. These properties will aid us in maximizing the dual function.

Differentiability of θ

We now address the question of differentiability of θ defined by $\theta(\mathbf{w}) = \inf \{f(\mathbf{x}) + \mathbf{w}'\boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\}$. It will be convenient to introduce the following set:

$$X(\mathbf{w}) = \{\mathbf{y} : \mathbf{y} \text{ minimizes } f(\mathbf{x}) + \mathbf{w}'\boldsymbol{\beta}(\mathbf{x}) \text{ over } \mathbf{x} \in X\}$$

The differentiability of θ at any given point $\bar{\mathbf{w}}$ depends on the elements of $X(\bar{\mathbf{w}})$. In particular, if the set $X(\bar{\mathbf{w}})$ is a singleton, then Theorem 6.3.3 below shows that θ is differentiable at $\bar{\mathbf{w}}$. First, however, the following lemma is needed.

6.3.2 Lemma

Let X be a nonempty compact set in E_n and let $f: E_n \to E_1$ and $\boldsymbol{\beta}: E_n \to E_{m+l}$ be continuous. Let $\bar{\mathbf{w}} \in E_{m+l}$, and suppose that $X(\bar{\mathbf{w}})$ is the singleton $\{\bar{\mathbf{x}}\}$. Suppose that $\mathbf{w}_k \to \bar{\mathbf{w}}$, and let $\mathbf{x}_k \in X(\mathbf{w}_k)$ for each k. Then $\mathbf{x}_k \to \bar{\mathbf{x}}$.

Proof

By contradiction, suppose that $\mathbf{w}_k \to \bar{\mathbf{w}}, \mathbf{x}_k \in X(\mathbf{w}_k)$, and $\|\mathbf{x}_k - \bar{\mathbf{x}}\| > \varepsilon > 0$ for $k \in \mathscr{K}$ where $\mathscr{K}$ is some index set. Since X is compact, then the sequence $\{\mathbf{x}_k\}_{\mathscr{K}}$ has a convergent subsequence $\{\mathbf{x}_k\}_{\mathscr{K}'}$, with limit $\mathbf{y}$ in X. Note that $\|\mathbf{y} - \bar{\mathbf{x}}\| \geq \varepsilon > 0$, and hence $\mathbf{y}$ and $\bar{\mathbf{x}}$ are distinct. Furthermore, for each $\mathbf{w}_k$ with $k \in \mathscr{K}'$ we have

$$f(\mathbf{x}_k) + \mathbf{w}_k'\boldsymbol{\beta}(\mathbf{x}_k) \leq f(\bar{\mathbf{x}}) + \mathbf{w}_k'\boldsymbol{\beta}(\bar{\mathbf{x}})$$

Taking the limit as k in $\mathscr{K}'$ approaches ∞, and noting that $\mathbf{x}_k \to \mathbf{y}$, $\mathbf{w}_k \to \bar{\mathbf{w}}$, and that f and $\boldsymbol{\beta}$ are continuous, it follows that

$$f(\mathbf{y}) + \bar{\mathbf{w}}'\boldsymbol{\beta}(\mathbf{y}) \leq f(\bar{\mathbf{x}}) + \bar{\mathbf{w}}'\boldsymbol{\beta}(\bar{\mathbf{x}})$$

Therefore $\mathbf{y} \in X(\bar{\mathbf{w}})$, contradicting the assumption that $X(\bar{\mathbf{w}})$ is a singleton. This completes the proof.

6.3.3 Theorem

Let X be a nonempty compact set in E_n, and let $f: E_n \to E_1$, and $\boldsymbol{\beta}: E_n \to E_{m+l}$ be continuous. Let $\bar{\mathbf{w}} \in E_{m+l}$, and suppose that $X(\bar{\mathbf{w}})$ is the singleton $\{\bar{\mathbf{x}}\}$. Then, θ is differentiable at $\bar{\mathbf{w}}$ with gradient $\nabla\theta(\bar{\mathbf{w}}) = \boldsymbol{\beta}(\bar{\mathbf{x}})$.

Proof

Since f and $\boldsymbol{\beta}$ are continuous and X is compact, then for any given $\mathbf{w}$, there exists an $\mathbf{x}_{\mathbf{w}} \in X(\mathbf{w})$. From the definition of θ, the following two inequalities hold true:

$$\theta(\mathbf{w}) - \theta(\bar{\mathbf{w}}) \leq f(\bar{\mathbf{x}}) + \mathbf{w}'\boldsymbol{\beta}(\bar{\mathbf{x}}) - f(\bar{\mathbf{x}}) - \bar{\mathbf{w}}'\boldsymbol{\beta}(\bar{\mathbf{x}}) = (\mathbf{w} - \bar{\mathbf{w}})'\boldsymbol{\beta}(\bar{\mathbf{x}}) \tag{6.12}$$

$$\theta(\bar{\mathbf{w}}) - \theta(\mathbf{w}) \leq f(\mathbf{x}_{\mathbf{w}}) + \bar{\mathbf{w}}'\boldsymbol{\beta}(\mathbf{x}_{\mathbf{w}}) - f(\mathbf{x}_{\mathbf{w}}) - \mathbf{w}'\boldsymbol{\beta}(\mathbf{x}_{\mathbf{w}}) = (\bar{\mathbf{w}} - \mathbf{w})'\boldsymbol{\beta}(\mathbf{x}_{\mathbf{w}}) \tag{6.13}$$

From (6.12) and (6.13) and the Schwartz inequality, it follows that

$$\begin{aligned}
0 \geq \theta(\mathbf{w}) - \theta(\bar{\mathbf{w}}) - (\mathbf{w} - \bar{\mathbf{w}})'\boldsymbol{\beta}(\bar{\mathbf{x}}) &\geq (\mathbf{w} - \bar{\mathbf{w}})'[\boldsymbol{\beta}(\mathbf{x}_{\mathbf{w}}) - \boldsymbol{\beta}(\bar{\mathbf{x}})] \\
&\geq -\|\mathbf{w} - \bar{\mathbf{w}}\| \|\boldsymbol{\beta}(\mathbf{x}_{\mathbf{w}}) - \boldsymbol{\beta}(\bar{\mathbf{x}})\|
\end{aligned}$$

This further implies that

$$0 \geq \frac{\theta(\mathbf{w}) - \theta(\bar{\mathbf{w}}) - (\mathbf{w} - \bar{\mathbf{w}})'\boldsymbol{\beta}(\bar{\mathbf{x}})}{\|\mathbf{w} - \bar{\mathbf{w}}\|} \geq -\|\boldsymbol{\beta}(\mathbf{x}_w) - \boldsymbol{\beta}(\bar{\mathbf{x}})\| \tag{6.14}$$

As $\mathbf{w} \to \bar{\mathbf{w}}$, then by Lemma 6.3.2, $\mathbf{x}_w \to \bar{\mathbf{x}}$, and by continuity of $\boldsymbol{\beta}$, $\boldsymbol{\beta}(\mathbf{x}_w) \to \boldsymbol{\beta}(\bar{\mathbf{x}})$. Therefore, from (6.14), we get

$$\lim_{\mathbf{w} \to \bar{\mathbf{w}}} \frac{\theta(\mathbf{w}) - \theta(\bar{\mathbf{w}}) - (\mathbf{w} - \bar{\mathbf{w}})'\boldsymbol{\beta}(\bar{\mathbf{x}})}{\|\mathbf{w} - \bar{\mathbf{w}}\|} = 0$$

Hence θ is differentiable at $\bar{\mathbf{w}}$ with gradient $\boldsymbol{\beta}(\bar{\mathbf{x}})$. This completes the proof.

Subgradients of θ

We have shown in Theorem 6.3.1 that θ is concave, and hence, by Theorem 3.2.5, θ is subdifferentiable; that is, it has subgradients. As will be seen later, subgradients play an important role in the maximization of the dual function, since they lead naturally to the characterization of the directions of ascent. Theorem 6.3.4 below shows that each $\bar{\mathbf{x}} \in X(\bar{\mathbf{w}})$ yields a subgradient of θ at $\bar{\mathbf{w}}$.

6.3.4 Theorem

Let X be a nonempty compact set in E_n, and let $f: E_n \to E_1$, and $\boldsymbol{\beta}: E_n \to E_{m+l}$ be continuous so that for any $\bar{\mathbf{w}} \in E_{m+l}$, $X(\bar{\mathbf{w}})$ is not empty. If $\bar{\mathbf{x}} \in X(\bar{\mathbf{w}})$, then $\boldsymbol{\beta}(\bar{\mathbf{x}})$ is a subgradient of θ at $\bar{\mathbf{w}}$.

Proof

Since f and $\boldsymbol{\beta}$ are continuous and X is compact, $X(\mathbf{w}) \neq \emptyset$ for any $\mathbf{w} \in E_{m+l}$. Now, let $\mathbf{w} \in E_{m+l}$, and let $\bar{\mathbf{x}} \in X(\bar{\mathbf{w}})$. Then

$$\begin{aligned}\theta(\mathbf{w}) &= \inf\{f(\mathbf{x}) + \mathbf{w}'\boldsymbol{\beta}(\mathbf{x}) : \mathbf{x} \in X\}\\ &\leq f(\bar{\mathbf{x}}) + \mathbf{w}'\boldsymbol{\beta}(\bar{\mathbf{x}})\\ &= f(\bar{\mathbf{x}}) + (\mathbf{w} - \bar{\mathbf{w}})'\boldsymbol{\beta}(\bar{\mathbf{x}}) + \bar{\mathbf{w}}'\boldsymbol{\beta}(\bar{\mathbf{x}})\\ &= \theta(\bar{\mathbf{w}}) + (\mathbf{w} - \bar{\mathbf{w}})'\boldsymbol{\beta}(\bar{\mathbf{x}})\end{aligned}$$

Therefore $\boldsymbol{\beta}(\bar{\mathbf{x}})$ is a subgradient of θ at $\bar{\mathbf{w}}$ and the proof is complete.

6.3.5 Example

Consider the following primal problem:

$$\begin{aligned}&\text{Minimize} && -x_1 - x_2\\ &\text{subject to} && x_1 + 2x_2 - 3 \leq 0\\ &&& x_1, x_2 = 0, 1, 2, \text{ or } 3\end{aligned}$$

Letting $g(x_1, x_2) = x_1 + 2x_2 - 3$ and $X = \{(x_1, x_2): x_1, x_2 = 0, 1, 2, \text{ or } 3\}$ the dual function is given by

$$\theta(u) = \inf\{-x_1 - x_2 + u(x_1 + 2x_2 - 3): x_1, x_2 = 0, 1, 2, \text{ or } 3\}$$

$$= \begin{cases} -6 + 6u & \text{if } 0 \leq u \leq \frac{1}{2} \\ -3 & \text{if } \frac{1}{2} \leq u \leq 1 \\ -3u & \text{if } u \geq 1 \end{cases}$$

Let $\bar{u} = \frac{1}{2}$. In order to find a subgradient of θ at $\bar{u}$, consider the following subproblem:

$$\begin{aligned}&\text{Minimize} && -x_1 - x_2 + \tfrac{1}{2}(x_1 + 2x_2 - 3)\\ &\text{subject to} && x_1, x_2 = 0, 1, 2, \text{ or } 3\end{aligned}$$

Note that the set $X(\bar{u})$ of optimal solutions of the above problem is $\{(3, 0), (3, 1), (3, 2), (3, 3)\}$. Thus, from Theorem 6.3.4, $g(3, 0) = 0$, $g(3, 1) = 2$, $g(3, 2) = 4$, and $g(3, 3) = 6$ are subgradients of θ at $\bar{u}$. Note, however, that $\frac{3}{2}$ is also a subgradient of θ at $\bar{u}$, but $\frac{3}{2}$ cannot be represented as $g(\bar{\mathbf{x}})$ for any $\bar{\mathbf{x}} \in X(\bar{u})$.

From the above example, it is clear that Theorem 6.3.4 gives only a sufficient characterization of subgradients. A necessary and sufficient characterization of subgradients is given in Theorem 6.3.7 below. First, however, the following important result is needed.

6.3.6 Theorem

Let X be a nonempty compact set in E_n, and let $f: E_n \to E_1$ and $\boldsymbol{\beta}: E_n \to E_{m+l}$ be continuous. Let $\bar{\mathbf{w}}$, $\mathbf{d} \in E_{m+l}$. Then

$$\theta'(\bar{\mathbf{w}}; \mathbf{d}) \geq \mathbf{d}'\boldsymbol{\beta}(\bar{\mathbf{x}}) \qquad \text{for some } \bar{\mathbf{x}} \in X(\bar{\mathbf{w}})$$

Proof

Consider $\bar{\mathbf{w}} + \lambda_k \mathbf{d}$, where $\lambda_k \to 0^+$. For each k, there exists an $\mathbf{x}_k \in X(\bar{\mathbf{w}} + \lambda_k \mathbf{d})$, and since X is compact, there is a convergent subsequence $\{\mathbf{x}_k\}_{\mathscr{K}}$ with limit $\bar{\mathbf{x}}$ in X. Given an $\mathbf{x} \in X$, note that

$$f(\mathbf{x}) + (\bar{\mathbf{w}} + \lambda_k \mathbf{d})'\boldsymbol{\beta}(\mathbf{x}) \geq f(\mathbf{x}_k) + (\bar{\mathbf{w}} + \lambda_k \mathbf{d})'\boldsymbol{\beta}(\mathbf{x}_k)$$

for each $k \in \mathscr{K}$. Taking the limit as $k \to \infty$, it follows that

$$f(\mathbf{x}) + \bar{\mathbf{w}}'\boldsymbol{\beta}(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \bar{\mathbf{w}}'\boldsymbol{\beta}(\bar{\mathbf{x}})$$

that is, $\bar{\mathbf{x}} \in X(\bar{\mathbf{w}})$. Furthermore, by definition of $\theta(\bar{\mathbf{w}} + \lambda_k \mathbf{d})$ and $\theta(\bar{\mathbf{w}})$, we get

$$\begin{aligned}\theta(\bar{\mathbf{w}} + \lambda_k \mathbf{d}) - \theta(\bar{\mathbf{w}}) &= f(\mathbf{x}_k) + (\bar{\mathbf{w}} + \lambda_k \mathbf{d})'\boldsymbol{\beta}(\mathbf{x}_k) - \theta(\bar{\mathbf{w}})\\ &\geq \lambda_k \mathbf{d}'\boldsymbol{\beta}(\mathbf{x}_k)\end{aligned}$$

The above inequality holds true for each $k \in \mathscr{K}$. Noting that $\mathbf{x}_k \to \bar{\mathbf{x}}$ as $k \in \mathscr{K}$ approaches ∞, we get

$$\lim_{\substack{k \in \mathscr{K} \\ k \to \infty}} \frac{\theta(\bar{\mathbf{w}} + \lambda_k \mathbf{d}) - \theta(\bar{\mathbf{w}})}{\lambda_k} \geq \mathbf{d}'\boldsymbol{\beta}(\bar{\mathbf{x}})$$

By Lemma 3.1.5, $\theta'(\bar{\mathbf{w}}; \mathbf{d}) = \lim_{\lambda \to 0^+} \dfrac{\theta(\bar{\mathbf{w}} + \lambda \mathbf{d}) - \theta(\bar{\mathbf{w}})}{\lambda}$ exists. In view of the above inequality, the proof is complete.

Corollary

Let $\partial\theta(\bar{\mathbf{w}})$ be the collection of subgradients of θ at $\bar{\mathbf{w}}$, and suppose that the assumptions of the theorem hold true. Then,

$$\theta'(\bar{\mathbf{w}}; \mathbf{d}) = \inf \{\mathbf{d}'\boldsymbol{\xi} : \boldsymbol{\xi} \in \partial\theta(\bar{\mathbf{w}})\}$$

Proof

Let $\bar{\mathbf{x}}$ be as specified in the theorem. By Theorem 6.3.4 $\boldsymbol{\beta}(\bar{\mathbf{x}}) \in \partial\theta(\bar{\mathbf{w}})$, and hence Theorem 6.3.6 implies that $\theta'(\bar{\mathbf{w}}; \mathbf{d}) \geq \inf \{\mathbf{d}'\boldsymbol{\xi} : \boldsymbol{\xi} \in \partial\theta(\bar{\mathbf{w}})\}$. Now let $\boldsymbol{\xi} \in \partial\theta(\bar{\mathbf{w}})$. Since θ is concave, $\theta(\bar{\mathbf{w}} + \lambda \mathbf{d}) - \theta(\bar{\mathbf{w}}) \leq \lambda \mathbf{d}'\boldsymbol{\xi}$. Dividing by $\lambda > 0$ and taking the limit as $\lambda \to 0^+$, it follows that $\theta'(\bar{\mathbf{w}}; \mathbf{d}) \leq \mathbf{d}'\boldsymbol{\xi}$. Since this is true for each $\boldsymbol{\xi} \in \partial\theta(\bar{\mathbf{w}})$, $\theta'(\bar{\mathbf{w}}; \mathbf{d}) \leq \inf \{\mathbf{d}'\boldsymbol{\xi} : \boldsymbol{\xi} \in \partial\theta(\bar{\mathbf{w}})\}$, and the proof is complete.

6.3.7 Theorem

Let X be a nonempty compact set in E_n, and let $f : E_n \to E_1$, and $\boldsymbol{\beta} : E_n \to E_{m+l}$ be continuous. Then $\boldsymbol{\xi}$ is a subgradient of θ at $\bar{\mathbf{w}} \in E_{m+l}$ if and only if $\boldsymbol{\xi}$ belongs to the convex hull of $\{\boldsymbol{\beta}(\mathbf{y}) : \mathbf{y} \in X(\bar{\mathbf{w}})\}$.

Proof

Denote the set $\{\boldsymbol{\beta}(\mathbf{y}) : \mathbf{y} \in X(\bar{\mathbf{w}})\}$ by Λ and its convex hull by $H(\Lambda)$. By Theorem 6.3.4, $\Lambda \subset \partial\theta(\bar{\mathbf{w}})$, and since $\partial\theta(\bar{\mathbf{w}})$ is convex, $H(\Lambda) \subset \partial\theta(\bar{\mathbf{w}})$. Using the facts that X is compact and $\boldsymbol{\beta}$ is continuous, it can be verified that Λ is compact. Furthermore, the convex hull of a compact set is closed. Therefore $H(\Lambda)$ is a closed convex set.

We shall now show that $H(\Lambda) \supset \partial\theta(\bar{\mathbf{w}})$. By contradiction, suppose that there is a $\boldsymbol{\xi}' \in \partial\theta(\bar{\mathbf{w}})$ but not in $H(\Lambda)$. By Theorem 2.3.4, there exist a scalar α and a nonzero vector $\mathbf{d}$ such that

$$\mathbf{d}'\boldsymbol{\beta}(\mathbf{y}) \geq \alpha \qquad \text{for each } \mathbf{y} \in X(\bar{\mathbf{w}}) \tag{6.15}$$

$$\mathbf{d}'\boldsymbol{\xi}' < \alpha \tag{6.16}$$

By Theorem 6.3.6, there exists a $\mathbf{y} \in X(\bar{\mathbf{w}})$ such that $\theta'(\bar{\mathbf{w}}; \mathbf{d}) \geq \mathbf{d}'\boldsymbol{\beta}(\mathbf{y})$, and by (6.15) above, we must have $\theta'(\bar{\mathbf{w}}; \mathbf{d}) \geq \alpha$. But by the corollary to Theorem 6.3.6 and (6.16), we get

$$\theta'(\bar{\mathbf{w}}; \mathbf{d}) = \inf \{\mathbf{d}'\boldsymbol{\xi} : \boldsymbol{\xi} \in \partial\theta(\mathbf{w})\} \leq \mathbf{d}'\boldsymbol{\xi}' < \alpha$$

which is a contradiction. Therefore $\boldsymbol{\xi}' \in H(\Lambda)$, and $\partial\theta(\bar{\mathbf{w}}) = H(\Lambda)$. This completes the proof.

6.3.8 Example

Consider the following primal problem:

$$\begin{aligned} \text{Minimize} \quad & -(x_1 - 4)^2 - (x_2 - 4)^2 \\ \text{subject to} \quad & x_1 - 3 \leq 0 \\ & -x_1 + x_2 - 2 \leq 0 \\ & x_1 + x_2 - 4 \leq 0 \\ & x_1, x_2 \geq 0 \end{aligned}$$

In this example, we let $g_1(x_1, x_2) = x_1 - 3$, $g_2(x_1, x_2) = -x_1 + x_2 - 2$, and $X = \{(x_1, x_2) : x_1 + x_2 - 4 \leq 0;\ x_1, x_2 \geq 0\}$. Thus the dual function is given by

$$\theta(u_1, u_2) = \inf \{-(x_1 - 4)^2 - (x_2 - 4)^2 + u_1(x_1 - 3) + u_2(-x_1 + x_2 - 2) : \mathbf{x} \in X\}$$

We utilize Theorem 6.3.7 above to determine the set of subgradients of θ at $\bar{\mathbf{u}} = (1, 5)'$. In order to find the set $X(\bar{\mathbf{u}})$, we need to solve the following problem:

$$\begin{aligned} \text{Minimize} \quad & -(x_1 - 4)^2 - (x_2 - 4)^2 - 4x_1 + 5x_2 - 13 \\ \text{subject to} \quad & x_1 + x_2 - 4 \leq 0 \\ & x_1, x_2 \geq 0 \end{aligned}$$

The above objective function is concave, and by Theorem 3.4.6, it assumes its minimum over a compact polyhedral set at one of the extreme points. The polyhedral set X has three extreme points, namely $(0, 0)$, $(4, 0)$, and $(0, 4)$. Noting that $f(0, 0) = f(4, 0) = -45$ and $f(0, 4) = -9$, it is evident that the optimal solutions of the above subproblem are $(0, 0)$ and $(4, 0)$, that is, $X(\bar{\mathbf{u}}) = \{(0, 0), (4, 0)\}$. By Theorem 6.3.7, the subgradients of θ at $\bar{\mathbf{u}}$ are thus given by the convex combinations of $\mathbf{g}(0, 0)$ and $\mathbf{g}(4, 0)$, that is, by convex combinations of the two vectors $(-3, -2)'$ and $(1, -6)'$. Figure 6.6 illustrates the set of subgradients.

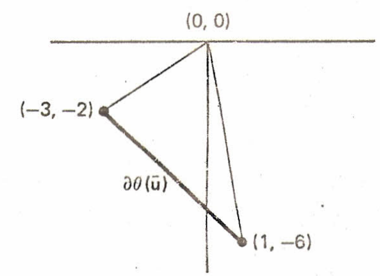

Figure 6.6 Illustration of subgradients.

Ascent and Steepest Ascent Directions

The dual problem is concerned with the maximization of θ subject to the constraint $\mathbf{u} \geq \mathbf{0}$. Given a point $\mathbf{w}^t = (\mathbf{u}^t, \mathbf{v}^t)$, we would like to investigate the directions along which θ increases. For the sake of clarity, first consider the following definition of an ascent direction.

6.3.9 Definition

A vector $\mathbf{d}$ is called an *ascent direction* of θ at $\mathbf{w}$ if there exists a $\delta > 0$ such that

$$\theta(\mathbf{w} + \lambda \mathbf{d}) > \theta(\mathbf{w}) \qquad \text{for each } \lambda \in (0, \delta)$$

Note that if θ is concave, a vector $\mathbf{d}$ is an ascent direction of θ at $\mathbf{w}$ if and only if $\theta'(\mathbf{w}; \mathbf{d}) > 0$. Furthermore, θ assumes its maximum at $\mathbf{w}$ if and only if it has no ascent directions at $\mathbf{w}$, that is, if and only if $\theta'(\mathbf{w}; \mathbf{d}) \leq 0$ for each $\mathbf{d}$.

Using the corollary to Theorem 6.3.6, it follows that a vector $\mathbf{d}$ is an ascent direction of θ at $\mathbf{w}$ if and only if $\inf \{\mathbf{d}^t \boldsymbol{\xi} \in \partial\theta(\mathbf{w})\} > 0$, that is, if and only if the following inequality holds for some $\varepsilon > 0$.

$$\mathbf{d}^t \boldsymbol{\xi} \geq \varepsilon > 0 \qquad \text{for each } \boldsymbol{\xi} \in \partial\theta(\mathbf{w})$$

To illustrate, consider Example 6.3.8. The collection of subgradients of θ at the point $(1, 5)$ is illustrated in Figure 6.6. A vector $\mathbf{d}$ is an ascent direction of θ if and only if $\mathbf{d}^t \boldsymbol{\xi} \geq \varepsilon$ for each subgradient $\boldsymbol{\xi}$, where $\varepsilon > 0$. In other words, $\mathbf{d}$ is an ascent direction if it makes an angle strictly less than 90° with each subgradient. The cone of ascent directions for this example is given in Figure 6.7. In this case, note that each subgradient is an ascent direction. However, this is not necessarily the case in general.

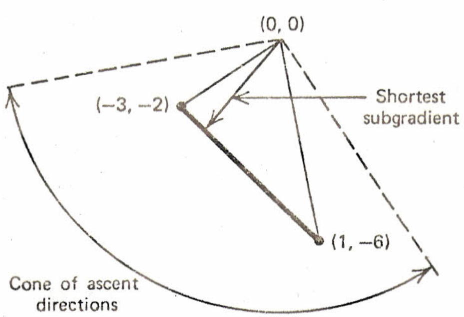

Figure 6.7 The cone of ascent directions in Example 6.3.8.

Since θ is to be maximized, we are interested not only in an ascent direction but also in the direction along which θ increases the most.

6.3.10 Definition

A vector $\bar{\mathbf{d}}$ is called a *direction of steepest ascent* of θ at $\mathbf{w}$ if

$$\theta'(\mathbf{w}; \bar{\mathbf{d}}) = \underset{\|\mathbf{d}\| \leq 1}{\text{maximum}}\ \theta'(\mathbf{w}; \mathbf{d})$$

Theorem 6.3.11 below shows that the direction of steepest ascent of the Lagrangian dual function is given by the subgradient with the smallest Euclidean norm.

6.3.11 Theorem

Let X be a nonempty compact set in E_n, and let $f: E_n \to E_1$, and $\boldsymbol{\beta}: E_n \to E_{m+l}$ be continuous. The direction of steepest ascent $\bar{\mathbf{d}}$ of θ at $\mathbf{w}$ is given below, where $\bar{\boldsymbol{\xi}}$ is the subgradient in $\partial\theta(\mathbf{w})$ with the smallest Euclidean norm.

$$\bar{\mathbf{d}} = \begin{cases} \mathbf{0} & \text{if } \bar{\boldsymbol{\xi}} = \mathbf{0} \\ \dfrac{\bar{\boldsymbol{\xi}}}{\|\bar{\boldsymbol{\xi}}\|} & \text{if } \bar{\boldsymbol{\xi}} \neq \mathbf{0} \end{cases}$$

Proof

By Definition 6.3.10 and by the corollary to Theorem 6.3.6, the steepest ascent direction can be obtained from the following expression:

$$\underset{\|\mathbf{d}\| \leq 1}{\text{maximum}}\ \theta'(\mathbf{w}; \mathbf{d}) = \underset{\|\mathbf{d}\| \leq 1}{\text{maximum}}\ \underset{\boldsymbol{\xi} \in \partial\theta(\mathbf{w})}{\text{infimum}}\ \mathbf{d}^t \boldsymbol{\xi}$$

The reader can easily verify that

$$\begin{aligned} \underset{\|\mathbf{d}\| \leq 1}{\text{maximum}}\ \theta'(\mathbf{w}; \mathbf{d}) &= \underset{\|\mathbf{d}\| \leq 1}{\text{maximum}}\ \underset{\boldsymbol{\xi} \in \partial\theta(\mathbf{w})}{\text{infimum}}\ \mathbf{d}^t \boldsymbol{\xi} \\ &\leq \underset{\boldsymbol{\xi} \in \partial\theta(\mathbf{w})}{\text{infimum}}\ \underset{\|\mathbf{d}\| \leq 1}{\text{maximum}}\ \mathbf{d}^t \boldsymbol{\xi} \\ &= \underset{\boldsymbol{\xi} \in \partial\theta(\mathbf{w})}{\text{infimum}}\ \|\boldsymbol{\xi}\| \\ &= \|\bar{\boldsymbol{\xi}}\| \end{aligned} \qquad (6.17)$$

If we construct a direction $\bar{\mathbf{d}}$ such that $\theta'(\mathbf{w}; \bar{\mathbf{d}}) = \|\bar{\boldsymbol{\xi}}\|$, then by (6.17) $\bar{\mathbf{d}}$ is the steepest ascent direction. If $\bar{\xi}=\mathbf{0}$, then for $\bar{\mathbf{d}}=\mathbf{0}$, we obviously have $\theta'(\mathbf{w};\bar{\mathbf{d}})=\|\bar{\xi}\|$. Now suppose that $\bar{\xi}\neq\mathbf{0}$, and let $\bar{\mathbf{d}}=\bar{\xi}/\|\bar{\xi}\|$. Note that

$$\begin{aligned}\theta'(\mathbf{w};\bar{\mathbf{d}}) &= \inf\{\bar{\mathbf{d}}'\xi:\xi\in\partial\theta(\mathbf{w})\}\\ &= \inf\left\{\frac{\bar{\xi}'\xi}{\|\bar{\xi}\|}:\xi\in\partial\theta(\mathbf{w})\right\}\\ &= \frac{1}{\|\bar{\xi}\|}\inf\{\|\bar{\xi}\|^2+\bar{\xi}'(\xi-\bar{\xi}):\xi\in\partial\theta(\mathbf{w})\}\\ &= \|\bar{\xi}\|+\frac{1}{\|\bar{\xi}\|}\inf\{\bar{\xi}'(\xi-\bar{\xi}):\xi\in\partial\theta(\mathbf{w})\} \qquad (6.18)\end{aligned}$$

Since $\bar{\xi}$ is the shortest vector in $\partial\theta(\mathbf{w})$, then by Theorem 2.3.1, $\bar{\xi}'(\xi-\bar{\xi})\geq 0$ for each $\xi\in\partial\theta(\mathbf{w})$. Hence, $\inf\{\bar{\xi}'(\xi-\bar{\xi}):\xi\in\partial\theta(\mathbf{w})\}=0$ is achieved at $\bar{\xi}$. From (6.18), it then follows that $\theta'(\mathbf{w};\bar{\mathbf{d}})=\|\bar{\xi}\|$. Thus, we have shown that the vector $\bar{\mathbf{d}}$ specified in the theorem is the direction of steepest ascent both when $\bar{\xi}=\mathbf{0}$ and when $\bar{\xi}\neq\mathbf{0}$.

6.4 Solving the Dual Problem

We have described several properties of the dual function in the previous section. In this section, we utilize these properties to develop various schemes for maximizing the dual function θ over the region $\{(\mathbf{u},\mathbf{v}):\mathbf{u}\geq\mathbf{0}\}$. In particular we discuss some ascent procedures, as well as the cutting plane method for solving the dual problem.

Gradient Method

Given $(\bar{\mathbf{u}},\bar{\mathbf{v}})$, the dual function can be evaluated by solving the following subproblem:

$$\begin{aligned}&\text{Minimize} \quad f(\mathbf{x})+\bar{\mathbf{u}}'\mathbf{g}(\mathbf{x})+\bar{\mathbf{v}}'\mathbf{h}(\mathbf{x})\\ &\text{subject to} \quad \mathbf{x}\in X\end{aligned}$$

Suppose that $\bar{\mathbf{x}}$ is the optimal solution. Then by Theorem 6.3.3, $\nabla\theta(\bar{\mathbf{u}},\bar{\mathbf{v}})'=[\mathbf{g}(\bar{\mathbf{x}})',\mathbf{h}(\bar{\mathbf{x}})']$. If $\nabla\theta(\bar{\mathbf{u}},\bar{\mathbf{v}})\neq\mathbf{0}$, then by Theorem 4.1.2, it is an ascent direction and θ will increase by moving from $(\bar{\mathbf{u}},\bar{\mathbf{v}})$ along $\nabla\theta(\bar{\mathbf{u}},\bar{\mathbf{v}})$. However, if some components of $\bar{\mathbf{u}}$ are equal to zero, and any of the corresponding components of $\mathbf{g}(\bar{\mathbf{x}})$ is negative, then $\bar{\mathbf{u}}+\lambda\mathbf{g}(\bar{\mathbf{x}})\ngeq\mathbf{0}$ for $\lambda\geq 0$, thus violating the nonnegativity restriction. In order to handle this difficulty, we use the modified direction $[\hat{\mathbf{g}}(\bar{\mathbf{x}}),\mathbf{h}(\bar{\mathbf{x}})]$, where $\hat{\mathbf{g}}(\bar{\mathbf{x}})$ is defined below as

$$\hat{g}_i(\bar{\mathbf{x}})=\begin{cases}g_i(\bar{\mathbf{x}}) & \text{if } \bar{u}_i>0\\ \text{maximum }[0,\ g_i(\bar{\mathbf{x}})] & \text{if } \bar{u}_i=0\end{cases} \qquad (6.19)$$

The following theorem shows that $[\hat{\mathbf{g}}(\bar{\mathbf{x}}),\mathbf{h}(\bar{\mathbf{x}})]$ is a feasible ascent direction of θ at $(\bar{\mathbf{x}},\bar{\mathbf{v}})$. Furthermore, $[\hat{\mathbf{g}}(\bar{\mathbf{x}}),\mathbf{h}(\bar{\mathbf{x}})]$ is zero only when the dual maximum is reached.

6.4.1 Theorem

Let $(\bar{\mathbf{u}},\bar{\mathbf{v}})\in E_{m+l}$, where $\bar{\mathbf{u}}\geq\mathbf{0}$. Suppose that θ is differentiable at $(\bar{\mathbf{u}},\bar{\mathbf{v}})$ with gradient $[\mathbf{g}(\bar{\mathbf{x}}),\mathbf{h}(\bar{\mathbf{x}})]$. If $[\hat{\mathbf{g}}(\bar{\mathbf{x}}),\mathbf{h}(\bar{\mathbf{x}})]\neq(\mathbf{0},\mathbf{0})$ then $[\hat{\mathbf{g}}(\bar{\mathbf{x}}),\mathbf{h}(\bar{\mathbf{x}})]$ is a feasible ascent direction of θ at $(\bar{\mathbf{u}},\bar{\mathbf{v}})$. If $[\hat{\mathbf{g}}(\bar{\mathbf{x}}),\mathbf{h}(\bar{\mathbf{x}})]=(\mathbf{0},\mathbf{0})$, then θ achieves its maximum over the region $\{(\mathbf{u},\mathbf{v}):\mathbf{u}\geq\mathbf{0}\}$ at $(\bar{\mathbf{u}},\bar{\mathbf{v}})$.

Proof

Let $\mathbf{d}'=[\hat{\mathbf{g}}(\bar{\mathbf{x}})',\mathbf{h}(\bar{\mathbf{x}})']$. By construction of $\hat{\mathbf{g}}$, $\mathbf{d}$ is a feasible direction. Furthermore, if $\mathbf{d}\neq\mathbf{0}$, then $\nabla\theta(\bar{\mathbf{u}},\bar{\mathbf{v}})'\mathbf{d}>0$ and by Theorem 4.1.2, $\mathbf{d}$ is an ascent direction. Now, suppose that $[\hat{\mathbf{g}}(\bar{\mathbf{x}}),\mathbf{h}(\bar{\mathbf{x}})]=(\mathbf{0},\mathbf{0})$. Since $\hat{g}_i(\bar{\mathbf{x}})=0$ for each i, it follows that $g_i(\bar{\mathbf{x}})\leq 0$ and $\bar{u}_i g_i(\bar{\mathbf{x}})=0$ for each i. In other words

$$\mathbf{g}(\bar{\mathbf{x}})\leq\mathbf{0} \qquad \text{and} \qquad \bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}})=0 \qquad (6.20)$$

Now consider the Lagrangian dual problem to maximize $\theta(\mathbf{u},\mathbf{v})$ subject to $\mathbf{u}\geq\mathbf{0}$. The Kuhn–Tucker conditions hold true at $(\bar{\mathbf{u}},\bar{\mathbf{v}})$ if there exists a vector $\mathbf{b}\leq\mathbf{0}$ such that $\nabla\theta(\bar{\mathbf{u}},\bar{\mathbf{v}})=(\mathbf{b},\mathbf{0})$ and $\bar{\mathbf{u}}'\mathbf{b}=0$. Noting (6.20), these conditions clearly hold true by letting $\mathbf{b}=\mathbf{g}(\bar{\mathbf{x}})$. Since θ is concave, by Theorem 4.2.11, the Kuhn–Tucker conditions are sufficient for optimality, and $(\bar{\mathbf{u}},\bar{\mathbf{v}})$ is an optimal solution. This completes the proof.

Summary of the Gradient Method

If the assumptions of Theorems 6.3.3 hold true, then θ is differentiable, and the following scheme could be used to maximize θ over the region $\{(\mathbf{u},\mathbf{v}):\mathbf{u}\geq\mathbf{0}\}$.

In step 2 of the algorithm, a one-dimensional problem in the variable λ is to be solved. For simplicity of presentation, we assumed that a finite optimal solution λ_k exists. If this were not the case, either the optimal objective value is unbounded, or else the optimal objective value is bounded but not achieved at any particular λ. In the first case, we stop with the conclusion that the dual problem is unbounded and the primal is infeasible. In the latter case, λ_k could be taken as a sufficiently large number.

Initialization Step Choose a vector $(\mathbf{u}_1,\mathbf{v}_1)$ with $\mathbf{u}_1\geq\mathbf{0}$, let $k=1$, and go to the main step.

Main Step 1. Given $(\mathbf{u}_k, \mathbf{v}_k)$, solve the following subproblem:

Minimize $f(\mathbf{x})+\mathbf{u}_k^t\mathbf{g}(\mathbf{x})+\mathbf{v}_k^t\mathbf{h}(\mathbf{x})$

subject to $\mathbf{x}\in X$

Let $\mathbf{x}_k$ be the unique optimal solution and form the vector $[\hat{\mathbf{g}}(\mathbf{x}_k), \mathbf{h}(\mathbf{x}_k)]$ using (6.19). If this vector is zero, then stop; $(\mathbf{u}_k, \mathbf{v}_k)$ is an optimal solution. Otherwise, go to step 2.

2. Consider the following problem:

Maximize $\theta[(\mathbf{u}_k, \mathbf{v}_k)+\lambda(\hat{\mathbf{g}}(\mathbf{x}_k), \mathbf{h}(\mathbf{x}_k))]$

subject to $\mathbf{u}_k+\lambda\hat{\mathbf{g}}(\mathbf{x}_k)\geq \mathbf{0}$

$\lambda\geq 0$

Let λ_k be an optimal solution, and let $(\mathbf{u}_{k+1}, \mathbf{v}_{k+1})=(\mathbf{u}_k, \mathbf{v}_k)+\lambda_k[\hat{\mathbf{g}}(\mathbf{x}_k), \mathbf{h}(\mathbf{x}_k)]$, replace k by $k+1$, and repeat step 1.

We shall now illustrate the gradient method for maximizing the dual function by the following example:

6.4.2 Example

Consider the following problem

Minimize $x_1^2+x_2^2$

subject to $-x_1-x_2+4\leq 0$

$x_1+2x_2-8\leq 0$

Note that the optimal solution occurs at the point (2, 2) where the objective function value is equal to 8. The Langrangian dual problem is to maximize $\theta(u_1, u_2)$ subject to $u_1, u_2\geq 0$, where

$$\theta(u_1, u_2)=\underset{x_1, x_2}{\text{minimum}}\,\{x_1^2+x_2^2+u_1(-x_1-x_2+4)+u_2(x_1+2x_2-8)\}$$

We shall solve the dual problem by the gradient method described above starting from $\mathbf{u}_1=(0, 0)^t$. Note that the function θ is differentiable by Theorem 6.3.3.

For $\mathbf{u}_1=(0, 0)^t$, $\theta(\mathbf{u}_1)=\text{minimum}_{x_1, x_2}\{x_1^2+x_2^2\}=0$, and is achieved at the unique optimal point $\mathbf{x}_1=(0, 0)^t$. From Theorem 6.3.3, $\nabla\theta(\mathbf{0})=\mathbf{g}(\mathbf{x}_1)=(4, -8)^t$. In this case $\hat{\mathbf{g}}(\mathbf{x}_1)=(4, 0)^t$. Note that

$$\begin{aligned}\theta(4\lambda, 0)&=\underset{x_1}{\text{minimum}}\,\{x_1^2-4\lambda x_1\}+\underset{x_2}{\text{minimum}}\,\{x_2^2-4\lambda x_2\}+16\lambda\\&=-4\lambda^2-4\lambda^2+16\lambda\\&=-8\lambda^2+16\lambda\end{aligned}$$

Hence, the optimal solution to the problem to maximize $\theta(4\lambda, 0)$ subject to $\lambda\geq 0$† is achieved at $\lambda_1=1$, so that

$$\mathbf{u}_2=\mathbf{u}_1+\lambda_1\hat{\mathbf{g}}(\mathbf{x}_1)=(0, 0)^t+1(4, 0)^t=(4, 0)^t$$

For $\mathbf{u}_2=(4, 0)^t$, $\theta(\mathbf{u}_2)=\text{minimum}_{x_1, x_2}\{x_1^2+x_2^2+4(-x_1-x_2+4)\}=8$, and is achieved at the unique optimal point $\mathbf{x}_2=(2, 2)^t$. From Theorem 6.3.3, $\nabla\theta(\mathbf{u}_2)=\mathbf{g}(\mathbf{x}_2)=(0, -2)^t$. In this case $\hat{\mathbf{g}}(\mathbf{x}_2)=(0, 0)^t$, and hence $\mathbf{u}_2=(4, 0)^t$ is an optimal solution to the Lagrangian dual problem.

Ascent Method for a Nondifferentiable Dual Function

In Section 6.3 we showed that $\mathbf{d}$ is an ascent direction of θ at $(\mathbf{u}, \mathbf{v})$ if $\mathbf{d}^t\boldsymbol{\xi}\geq\varepsilon>0$ for each $\boldsymbol{\xi}\in\partial\theta(\mathbf{u}, \mathbf{v})$. The following problem could be used for finding such a direction.

Maximize ε

subject to $\mathbf{d}^t\boldsymbol{\xi}\geq\varepsilon$ for $\boldsymbol{\xi}\in\partial\theta(\mathbf{u}, \mathbf{v})$

$d_i\geq 0$ if $u_i=0$

$-1\leq d_i\leq 1$ for $i=1, \ldots, m+l$

Note that the constraints $d_i\geq 0$ if $u_i=0$ ensure that the vector $\mathbf{d}$ is a feasible direction, and the normalization constraints $-1\leq d_i\leq 1$ will guarantee a finite solution to the problem.

The reader may note the following difficulties associated with the above direction finding problem.

1. The set $\partial\theta(\mathbf{u}, \mathbf{v})$ and, hence, the constraints of the problem are not explicitly known in advance. However, Theorem 6.3.7, which fully characterizes the subgradient set, could be of use.
2. The set $\partial\theta(\mathbf{u}, \mathbf{v})$ usually admits an infinite number of subgradients, so that we have a linear program with an infinite number of constraints. However, if $\partial\theta(\mathbf{u}, \mathbf{v})$ is a compact polyhedral set, then the constraints $\mathbf{d}^t\boldsymbol{\xi}\geq\varepsilon$ for $\boldsymbol{\xi}\in\partial\theta(\mathbf{u}, \mathbf{v})$ could be replaced by the constraints

$$\mathbf{d}^t\boldsymbol{\xi}_j\geq\varepsilon \qquad \text{for } j=1, \ldots, \gamma$$

where $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_\gamma$ are the extreme points of $\partial\theta(\mathbf{u}, \mathbf{v})$. Thus, in this case the problem reduces to a finite linear program.

† In general, an explicit expression for $\theta(\mathbf{u}_1+\lambda\hat{\mathbf{g}}(\mathbf{x}_1))=\theta(4\lambda, 0)$ is not available. However, for any given λ, θ can be evaluated by solving an unconstrained optimization problem. To find the optimal λ_1, a suitable line search procedure could be used. In Chapter 8, both line search methods and unconstrained optimization methods will be discussed in detail.

Summary of the Ascent Procedure

We present below an ascent procedure for maximizing θ over the region $\{(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}\}$. The method applies whether $\partial\theta(\mathbf{u}, \mathbf{v})$ is polyhedral or not. Step 1 attempts to generate an ascent direction by solving a linear program with finite number of constraints. Step 2 verifies whether the direction obtained from step 1 is indeed an ascent direction.

In view of Theorem 6.3.7, the implementation of step 2 requires the knowledge of all optimal solutions to the problem to minimize $f(\mathbf{x}) + \mathbf{u}_k^t \mathbf{g}(\mathbf{x}) + \mathbf{v}_k^t \mathbf{h}(\mathbf{x})$ subject to $\mathbf{x} \in X$, which may not be readily available. Step 3 maximizes θ along the ascent direction found in steps 1 and 2. At this step a one-dimensional problem in the variable λ is to be solved. For simplicity of presentation, we assumed that a finite optimal solution λ_k exists. If this were not the case, either the optimal objective value is unbounded, or else the optimal objective value is bounded but not achieved at any particular λ. In the first case, we stop with the conclusion that the dual problem is unbounded and the primal is infeasible. In the latter case, λ_k could be taken as a sufficiently large number.

The procedure is summarized below. It is assumed that f, $\mathbf{g}$, and $\mathbf{h}$ are continuous and that X is compact, so that the set $X(\mathbf{u}, \mathbf{v})$ is not empty for each $(\mathbf{u}, \mathbf{v})$.

Initialization Step Choose a vector $\mathbf{w}_1^t = (\mathbf{u}_1^t, \mathbf{v}_1^t)$ with $\mathbf{u}_1 \geq \mathbf{0}$. Solve the problem to minimize $f(\mathbf{x}) + \mathbf{u}_1^t \mathbf{g}(\mathbf{x}) + \mathbf{v}_1^t \mathbf{h}(\mathbf{x})$ subject to $\mathbf{x} \in X$. Let $\mathbf{x}_1$ be an optimal solution, and let $\xi_1^t = [\mathbf{g}(\mathbf{x}_1)^t, \mathbf{h}(\mathbf{x}_1)^t]$. Let $k = \gamma = 1$, and go to the main step.

Main Step 1. Given $\xi_1, \ldots, \xi_\gamma$, solve the following problem:

$$\begin{array}{lll} \text{Maximize} & \varepsilon & \\ \text{subject to} & \mathbf{d}^t \xi_j \geq \varepsilon & \text{for } j = 1, \ldots, \gamma \\ & d_i \geq 0 & \text{if } i\text{th component of } \mathbf{u}_k \text{ is } 0 \\ & -1 \leq d_i \leq 1 & \text{for } i = 1, \ldots, m + l \end{array}$$

Let $(\mathbf{d}_\gamma, \varepsilon_\gamma)$ be an optimal solution. If $\varepsilon_\gamma = 0$, stop; there exists no ascent direction and $\mathbf{w}_k^t = (\mathbf{u}_k^t, \mathbf{v}_k^t)$ is an optimal solution. Otherwise, $\varepsilon_\gamma > 0$; go to step 2.

2. Solve the following subproblem:

$$\begin{array}{ll} \text{Minimize} & \mathbf{d}_\gamma^t \xi \\ \text{subject to} & \xi \in \partial\theta(\mathbf{u}_k, \mathbf{v}_k) \end{array}$$

Let $\xi_{\gamma+1}$ be an optimal solution. If $\mathbf{d}_\gamma^t \xi_{\gamma+1} > 0$, then $\mathbf{d}_\gamma$ is an ascent direction, and proceed to step 3. If $\mathbf{d}_\gamma^t \xi_{\gamma+1} \leq 0$, then replace γ by $\gamma + 1$ and go to step 1.

3. Let $\mathbf{d}_k = \mathbf{d}_\gamma$, and solve the following problem:

$$\begin{array}{lll} \text{Maximize} & \theta(\mathbf{w}_k + \lambda \mathbf{d}_k) & \\ \text{subject to} & (\mathbf{w}_k + \lambda \mathbf{d}_k)_i \geq 0 & \text{for } i = 1, \ldots, m \\ & \lambda \geq 0 & \end{array}$$

where $(\mathbf{w}_k + \lambda \mathbf{d}_k)_i$ is the ith component of $\mathbf{w}_k + \lambda \mathbf{d}_k$. Let λ_k be an optimal solution, and let $\mathbf{w}_{k+1} = \mathbf{w}_k + \lambda_k \mathbf{d}_k$ and go to step 4.

4. Let $\mathbf{x}_{k+1}$ be an optimal solution to the problem to minimize $f(\mathbf{x}) + \mathbf{u}_{k+1}^t \mathbf{g}(\mathbf{x}) + \mathbf{v}_{k+1}^t \mathbf{h}(\mathbf{x})$ subject to $\mathbf{x} \in X$, where $(\mathbf{u}_{k+1}^t, \mathbf{v}_{k+1}^t) = \mathbf{w}_{k+1}^t$. Let $\xi_1^t = [\mathbf{g}(\mathbf{x}_{k+1})^t, \mathbf{h}(\mathbf{x}_{k+1})^t]$. Replace k by $k+1$, let $\gamma = 1$, and go to step 1.

We shall now illustrate the ascent procedure discussed above by the following example.

6.4.3 Example

Consider the following problem.

$$\begin{array}{ll} \text{Minimize} & x_1 - 4x_2 \\ \text{subject to} & -x_1 - x_2 + 2 \leq 0 \\ & x_2 - 1 \leq 0 \\ & 0 \leq x_1, x_2 \leq 3 \end{array}$$

Here the Lagrangian dual problem is to maximize $\theta(u_1, u_2)$ subject to u_1, $u_2 \geq 0$, where

$$\theta(u_1, u_2) = \text{minimum}\,\{(x_1 - 4x_2) + u_1(-x_1 - x_2 + 2) + u_2(x_2 - 1) : 0 \leq x_1, x_2 \leq 3\}$$

$$= \underset{0 \leq x_1 \leq 3}{\text{minimum}}\, x_1(1 - u_1) + \underset{0 \leq x_2 \leq 3}{\text{minimum}}\, x_2(-4 - u_1 + u_2) + 2u_1 - u_2$$

We shall solve the dual problem by the ascent procedure starting with $\mathbf{u}_1 = (0, 4)^t$.

For $\mathbf{u}_1 = (0, 4)^t$, $\theta(\mathbf{u}_1) = -4$ and is achieved at points of the form $(0, \alpha)$ for $0 \leq \alpha \leq 3$. Choosing $\alpha = 0$, we get an optimal solution $\mathbf{x} = (0, 0)^t$ and the associated subgradient $\xi_1 = \mathbf{g}(\mathbf{x}) = (2, -1)^t$. At step 1 of the procedure, we consider the problem

$$\begin{array}{ll} \text{Maximize} & \varepsilon \\ \text{subject to} & 2d_1 - d_2 \geq \varepsilon \\ & 0 \leq d_1 \leq 1 \\ & -1 \leq d_2 \leq 1 \end{array}$$

The optimal solution to this problem is $\mathbf{d}_1 = (1, -1)'$ and $\varepsilon = 3$. Since $\varepsilon > 0$, we solve the problem in step 2. By Theorem 6.3.7, $\partial\theta(\mathbf{u}_1) = \{\xi = (-\alpha+2, \alpha-1)' : 0 \leq \alpha \leq 3\}$. Thus the objective function of the problem in step 2 is $\xi'\mathbf{d}_1 = (-\alpha+2)1 + (\alpha-1)(-1) = -2\alpha+3$. Hence, the problem is to minimize $-2\alpha+3$ subject to $0 \leq \alpha \leq 3$. The optimal solution is $\alpha = 3$ with optimal objective value of $-3 < 0$. The associated optimum subgradient $\xi_2 = (-1, 2)'$.

Returning to step 1, we solve the problem

$$\begin{array}{ll} \text{Maximize} & \varepsilon \\ \text{subject to} & 2d_1 - d_2 \geq \varepsilon \\ & -d_1 + 2d_2 \geq \varepsilon \\ & 0 \leq d_1 \leq 1 \\ & -1 \leq d_2 \leq 1 \end{array}$$

The optimum objective value is 1 and is achieved at $\mathbf{d}_2 = (1, 1)'$. Since $\varepsilon > 0$, we solve the problem in step 2. The objective function is $\xi'\mathbf{d}_2 = (-\alpha+2)1 + (\alpha-1)1 = 1$. Since the objective function is greater than 0 for each ξ, then $\mathbf{d}_2 = (1, 1)'$ is an ascent direction.

We now maximize the dual function along the ascent direction (1, 1); that is, solve the problem to maximize $\theta((0, 4) + \lambda(1, 1))$ subject to $\lambda \geq 0$. The reader can verify that

$$\theta((0, 4) + \lambda(1, 1)) = \begin{cases} \lambda - 4 & \text{for } 0 \leq \lambda \leq 1 \\ -2\lambda - 1 & \text{for } \lambda \geq 1 \end{cases}$$

The optimum solution is $\lambda_1 = 1$, and hence $\mathbf{u}_2 = (1, 5)'$.

We now repeat the process to find an ascent direction at $\mathbf{u}_2$. For $\mathbf{u}_2 = (1, 5)'$, $\theta(\mathbf{u}_2) = -3$ and is achieved at any point of the form (α_1, α_2), where $0 \leq \alpha_1, \alpha_2 \leq 3$. That is an optimal solution is a convex combination of the extreme points (0, 0), (0, 3), (3, 0), and (3, 3). Choosing the optimal solution $\mathbf{x} = (0, 0)'$, we get the associated subgradient $\xi_1 = \mathbf{g}(\mathbf{x}) = (2, -1)'$. At step 1 of the procedure we solve the problem

$$\begin{array}{ll} \text{Maximize} & \varepsilon \\ \text{subject to} & 2d_1 - d_2 \geq \varepsilon \\ & -1 \leq d_1, d_2 \leq 1 \end{array}$$

The optimal solution is $\mathbf{d}_1 = (1, -1)'$ and $\varepsilon = 3 > 0$. By Theorem 6.3.7, $\partial\theta(\mathbf{u}_2)$ is the convex combination of the points $\boldsymbol{\beta}_1 = \mathbf{g}(0, 0) = (2, -1)'$, $\boldsymbol{\beta}_2 = \mathbf{g}(0, 3) = (-1, 2)'$, $\boldsymbol{\beta}_3 = \mathbf{g}(3, 0) = (-1, -1)'$, and $\boldsymbol{\beta}_4 = \mathbf{g}(3, 3) = (-4, 2)'$. Noting that the optimal of the problem in step 2 is achieved at one of the above extreme points, the optimum is equal to minimum $\{\mathbf{d}_1'\boldsymbol{\beta}_1, \mathbf{d}_1'\boldsymbol{\beta}_2, \mathbf{d}_1'\boldsymbol{\beta}_3, \mathbf{d}_1'\boldsymbol{\beta}_4\}$ = minimum $\{3, -3, 0, -6\} = -6$. Therefore, the optimal solution ξ_2 to the problem at step 2 is given by $\boldsymbol{\beta}_4 = (-4, 2)'$.

Repeating step 1 with a new constraint corresponding to ξ_2, we get the following problem:

$$\begin{array}{ll} \text{Maximize} & \varepsilon \\ \text{subject to} & 2d_1 - d_2 \geq \varepsilon \\ & -4d_1 + 2d_2 \geq \varepsilon \\ & -1 \leq d_1, d_2 \leq 1 \end{array}$$

The optimal objective value of this problem is equal to zero and hence there exists no ascent direction. Thus the optimal solution to the dual problem is $\mathbf{u}_2 = (1, 5)'$.

The Cutting Plane Method

The methods discussed above for solving the dual problem generate at each iteration a feasible direction along which the Lagrangian dual function increases. We now discuss another strategy for solving the dual problem, in which at each iteration, a function that approximates the dual function is optimized.

Recall that the dual function θ is defined by

$$\theta(\mathbf{u}, \mathbf{v}) = \inf\{f(\mathbf{x}) + \mathbf{u}'\mathbf{g}(\mathbf{x}) + \mathbf{v}'\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\}$$

Letting $z = \theta(\mathbf{u}, \mathbf{v})$, the inequality $z \leq f(\mathbf{x}) + \mathbf{u}'\mathbf{g}(\mathbf{x}) + \mathbf{v}'\mathbf{h}(\mathbf{x})$ must hold true for each $\mathbf{x} \in X$. Hence the dual problem of maximizing $\theta(\mathbf{u}, \mathbf{v})$ over $\mathbf{u} \geq \mathbf{0}$ is equivalent to the following problem:

$$\begin{array}{lll} \text{Maximize} & z & \\ \text{subject to} & z \leq f(\mathbf{x}) + \mathbf{u}'\mathbf{g}(\mathbf{x}) + \mathbf{v}'\mathbf{h}(\mathbf{x}) & \text{for } \mathbf{x} \in X \\ & \mathbf{u} \geq \mathbf{0} & \end{array} \qquad (6.21)$$

Note that the above problem is a linear program in the variables z, $\mathbf{u}$, and $\mathbf{v}$. Unfortunately, however, the constraints are infinite and are not known explicitly. Suppose that we have the points $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$ in X, and consider the following problem:

$$\begin{array}{lll} \text{Maximize} & z & \\ \text{subject to} & z \leq f(\mathbf{x}_j) + \mathbf{u}'\mathbf{g}(\mathbf{x}_j) + \mathbf{v}'\mathbf{h}(\mathbf{x}_j) & \text{for } j = 1, \ldots, k-1 \\ & \mathbf{u} \geq \mathbf{0} & \end{array} \qquad (6.22)$$

The above problem is a linear program with a finite number of constraints and can be solved by the simplex method. Let $(z_k, \mathbf{u}_k, \mathbf{v}_k)$ be an optimal solution. If this solution satisfies (6.21), then it is an optimal solution to the Lagrangian dual problem. To check whether (6.21) is satisfied, consider the following subproblem:

$$\begin{array}{ll}\text{Minimize} & f(\mathbf{x})+\mathbf{u}_k^t\mathbf{g}(\mathbf{x})+\mathbf{v}_k^t\mathbf{h}(\mathbf{x})\\ \text{subject to} & \mathbf{x}\in X\end{array}$$

Let $\mathbf{x}_k$ be an optimal solution of the above problem, so that $\theta(\mathbf{u}_k, \mathbf{v}_k) = f(\mathbf{x}_k)+\mathbf{u}_k^t\mathbf{g}(\mathbf{x}_k)+\mathbf{v}_k^t\mathbf{h}(\mathbf{x}_k)$. If $z_k \le \theta(\mathbf{u}_k, \mathbf{v}_k)$, then $(\mathbf{u}_k, \mathbf{v}_k)$ is an optimal solution to the Lagrangian dual problem. Otherwise, for $(\mathbf{u}, \mathbf{v}) = (\mathbf{u}_k, \mathbf{v}_k)$ the inequality (6.21) is not satisfied for $\mathbf{x}=\mathbf{x}_k$. Thus, we add the constraint

$$z \le f(\mathbf{x}_k)+\mathbf{u}^t\mathbf{g}(\mathbf{x}_k)+\mathbf{v}^t\mathbf{h}(\mathbf{x}_k)$$

to the constraints in (6.22), and resolve the linear program. Obviously the current optimal point $(z_k, \mathbf{u}_k, \mathbf{v}_k)$ contradicts this added constraint. Thus, this point is cut away and hence the name, cutting plane algorithm.

Summary of the Cutting Plane Method

The cutting plane method is summarized below. It is assumed that f, $\mathbf{g}$, and $\mathbf{h}$ are continuous, and that X is compact, so that the set $X(\mathbf{u}, \mathbf{v})$ is not empty for each $(\mathbf{u}, \mathbf{v})$.

Initialization Step Find a point $\mathbf{x}_0 \in X$ such that $\mathbf{g}(\mathbf{x}_0) \le \mathbf{0}$ and $\mathbf{h}(\mathbf{x}_0)=\mathbf{0}$. Let $k=1$, and go to the main step.

Main Step 1. Solve the following problem, which is usually referred to as the *master problem*.

$$\begin{array}{lll}\text{Maximize} & z & \\ \text{subject to} & z \le f(\mathbf{x}_j)+\mathbf{u}^t\mathbf{g}(\mathbf{x}_j)+\mathbf{v}^t\mathbf{h}(\mathbf{x}_j) & \text{for } j=0,\ldots,k-1\\ & \mathbf{u}\ge\mathbf{0} & \end{array}$$

Let $(z_k, \mathbf{u}_k, \mathbf{v}_k)$ be an optimal solution and go to step 2.

2. Solve the following *subproblem*.

$$\begin{array}{ll}\text{Minimize} & f(\mathbf{x})+\mathbf{u}_k^t\mathbf{g}(\mathbf{x})+\mathbf{v}_k^t\mathbf{h}(\mathbf{x})\\ \text{subject to} & \mathbf{x}\in X\end{array}$$

Let $\mathbf{x}_k$ be an optimal point, and let $\theta(\mathbf{u}_k, \mathbf{v}_k) = f(\mathbf{x}_k)+\mathbf{u}_k^t\mathbf{g}(\mathbf{x}_k)+\mathbf{v}_k^t\mathbf{h}(\mathbf{x}_k)$. If $z_k = \theta(\mathbf{u}_k, \mathbf{v}_k)$, then stop; $(\mathbf{u}_k, \mathbf{v}_k)$ is an optimal dual solution. Otherwise, if $z_k > \theta(\mathbf{u}_k, \mathbf{v}_k)$, then replace k by $k+1$, and repeat step 1.

At each iteration, a cut (constraint) is added to the master problem, and hence the size of the master problem increases monotonically. If the size of the master problem becomes excessively large, all constraints that are not binding may be thrown away. Also note that the optimal solutions of the master problem form a nonincreasing sequence $\{z_k\}$. Since each z_k is an upper bound on the optimal value of the dual problem, we may stop if $z_k - \text{maximum}_{1\le j\le k}\, \theta(\mathbf{u}_j, \mathbf{v}_j) < \varepsilon$, where ε is a small positive number.

Interpretation as a Tangential Approximation Technique

The cutting plane algorithm for maximizing the dual function can be interpreted as a tangential approximation technique. By definition of θ, we must have

$$\theta(\mathbf{u}, \mathbf{v}) \le f(\mathbf{x})+\mathbf{u}^t\mathbf{g}(\mathbf{x})+\mathbf{v}^t\mathbf{h}(\mathbf{x}) \qquad \text{for } \mathbf{x}\in X$$

Thus, for any fixed $\mathbf{x}\in X$, the hyperplane

$$\{(\mathbf{u}, \mathbf{v}, z) : \mathbf{u}\in E_m,\ \mathbf{v}\in E_l,\ z = f(\mathbf{x})+\mathbf{u}^t\mathbf{g}(\mathbf{x})+\mathbf{v}^t\mathbf{h}(\mathbf{x})\}$$

bounds the function θ from above.

The master problem at iteration k is equivalent to solving the following problem.

$$\begin{array}{ll}\text{Maximize} & \hat{\theta}(\mathbf{u}, \mathbf{v})\\ \text{subject to} & \mathbf{u}\ge\mathbf{0}\end{array}$$

where $\hat{\theta}(\mathbf{u}, \mathbf{v}) = \text{minimum}\,\{f(\mathbf{x}_j)+\mathbf{u}^t\mathbf{g}(\mathbf{x}_j)+\mathbf{v}^t\mathbf{h}(\mathbf{x}_j):\ j=1,\ldots,k-1\}$. Note that $\hat{\theta}$ is a piecewise linear function that approximates θ by considering only $k-1$ of the bounding hyperplanes.

Let the optimal solution to the master problem be $(z_k, \mathbf{u}_k, \mathbf{v}_k)$. Now, the subproblem is solved yielding $\theta(\mathbf{u}_k, \mathbf{v}_k)$ and $\mathbf{x}_k$. If $z_k > \theta(\mathbf{u}_k, \mathbf{v}_k)$, then the new constraint $z \le f(\mathbf{x}_k)+\mathbf{u}^t\mathbf{g}(\mathbf{x}_k)+\mathbf{v}^t\mathbf{h}(\mathbf{x}_k)$ is added to the master problem, giving a new and tighter piecewise linear approximation to θ. Since $\theta(\mathbf{u}_k, \mathbf{v}_k) = f(\mathbf{x}_k)+\mathbf{u}_k^t\mathbf{g}(\mathbf{x}_k)+\mathbf{v}_k^t\mathbf{h}(\mathbf{x}_k)$, the hyperplane $\{(z, \mathbf{u}, \mathbf{v}):\ z = f(\mathbf{x}_k)+\mathbf{u}^t\mathbf{g}(\mathbf{x}_k)+\mathbf{v}^t\mathbf{h}(\mathbf{x}_k)\}$ is tangential to the graph of θ at $(z_k, \mathbf{u}_k, \mathbf{v}_k)$.

We now present an example of the cutting plane method and the interpretation given above.

6.4.4 Example

$$\begin{array}{ll}\text{Minimize} & (x_1-2)^2+\frac{1}{4}x_2^2\\ \text{subject to} & x_1-\frac{7}{2}x_2-1\le 0\\ & 2x_1+3x_2 \quad = 4\end{array}$$

TABLE 6.1 **Summary of Computations for Example 6.4.4**

Iteration k	Constraint Added	Step 1 Solution (z_k, u_k)	Step 2 Solution x_k^t	$\theta(u_k)$
1	$z \le \frac{5}{8} - \frac{3}{2}u$	$(\frac{5}{8}, 0)$	$(2, 0)$	0
2	$z \le 0 + u$	$(\frac{1}{4}, \frac{1}{4})$	$(\frac{13}{8}, \frac{1}{4})$	$\frac{3}{32}$
3	$z \le \frac{5}{32} - \frac{1}{4}u$	$(\frac{1}{8}, \frac{1}{8})$	$(\frac{29}{16}, \frac{1}{8})$	$\frac{11}{128}$
4	$z \le \frac{5}{128} + \frac{3}{8}u$	$(\frac{7}{64}, \frac{3}{16})$	$(\frac{55}{32}, \frac{3}{16})$	$\frac{51}{512}$

We let $X = \{(x_1, x_2): 2x_1 + 3x_2 = 4\}$, so that the Lagrangian dual function is given by

$$\theta(u) = \text{minimum } \{(x_1 - 2)^2 + \tfrac{1}{4}x_2^2 + u(x_1 - \tfrac{7}{2}x_2 - 1): 2x_1 + 3x_2 = 4\} \quad (6.23)$$

The cutting plane method is initiated with a feasible solution $\mathbf{x}_0 = (\frac{5}{4}, \frac{1}{2})^t$. At step 1 of the first iteration, we solve the following problem:

$$\begin{aligned} &\text{Maximize} && z \\ &\text{subject to} && z \le \tfrac{5}{8} - \tfrac{3}{2}u \\ & && u \ge 0 \end{aligned}$$

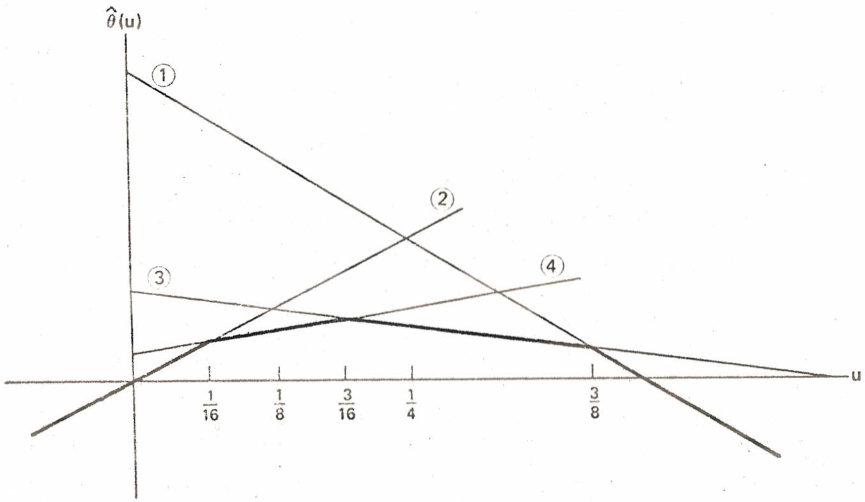

Figure 6.8 Tangential approximation of θ.

The optimal solution is $(z_1, u_1) = (\frac{5}{8}, 0)$. At step 2, we solve (6.23) for $u = u_1 = 0$, yielding an optimal solution $\mathbf{x}_1 = (2, 0)^t$ with $\theta(u_1) = 0 < z_1$. Hence, more iterations are needed. The summary of the first four iterations are given in Table 6.1.

The approximating function $\hat{\theta}$ at the end of the fourth iteration is shown in darkened lines in Figure 6.8. The reader can easily verify that the Langrangian dual function for this problem is given by $\theta(u) = -\frac{5}{2}u^2 + u$ and that the hyperplanes added at iteration 2 onward are indeed tangential to the graph of θ at the point (z_k, u_k). Incidentally, the dual objective function is maximized at $\bar{u} = \frac{1}{5}$ with $\theta(\bar{u}) = \frac{1}{10}$. Note that the sequence $\{u_k\}$ converges to the optimal point $\bar{u} = \frac{1}{5}$.

6.5 Getting the Primal Solution

So far, we have studied several properties of the dual function and described some procedures for solving the dual problem. However, our main concern is finding an optimal solution to the primal problem.

In this section we develop some theorems that will aid us in finding a solution to the primal problem, as well as solutions to perturbations of the primal problem. However, for nonconvex programs, as a result of the possible presence of a duality gap, additional work is usually needed to find an optimal primal solution.

Solutions to Perturbed Primal Problems

During the course of solving the dual problem, the following problem, which is used to evaluate the function θ at $(\mathbf{u}, \mathbf{v})$, is solved frequently.

$$\begin{aligned} &\text{Minimize} && f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x}) \\ &\text{subject to} && \mathbf{x} \in X \end{aligned}$$

Theorem 6.5.1 below shows that an optimal solution $\bar{\mathbf{x}}$ to the above problem is also an optimal solution to a problem that is similar to the primal problem, in which some of the constraints are perturbed.

6.5.1 Theorem

Let $(\mathbf{u}, \mathbf{v})$ be a given vector with $\mathbf{u} \ge \mathbf{0}$. Consider the problem to minimize $f(\mathbf{x}) + \mathbf{u}^t\mathbf{g}(\mathbf{x}) + \mathbf{v}^t\mathbf{h}(\mathbf{x})$ subject to $\mathbf{x} \in X$. Let $\bar{\mathbf{x}}$ be an optimal solution. Then $\bar{\mathbf{x}}$ is an optimal solution to the following problem, where $I = \{i : u_i > 0\}$.

$$\begin{aligned} &\text{Minimize} && f(\mathbf{x}) \\ &\text{subject to} && g_i(\mathbf{x}) \le g_i(\bar{\mathbf{x}}) && \text{for } i \in I \\ & && h_i(\mathbf{x}) = h_i(\bar{\mathbf{x}}) && \text{for } i = 1, \ldots, l \\ & && \mathbf{x} \in X \end{aligned}$$

Proof

Let $\mathbf{x} \in X$ be such that $h_i(\mathbf{x}) = h_i(\bar{\mathbf{x}})$ for $i = 1, \ldots, l$, and $g_i(\mathbf{x}) \le g_i(\bar{\mathbf{x}})$ for $i \in I$. Note that

$$f(\mathbf{x}) + \mathbf{u}'\mathbf{g}(\mathbf{x}) + \mathbf{v}'\mathbf{h}(\mathbf{x}) \ge f(\bar{\mathbf{x}}) + \mathbf{u}'\mathbf{g}(\bar{\mathbf{x}}) + \mathbf{v}'\mathbf{h}(\bar{\mathbf{x}}) \tag{6.24}$$

But since $\mathbf{h}(\mathbf{x}) = \mathbf{h}(\bar{\mathbf{x}})$, and $\mathbf{u}'\mathbf{g}(\mathbf{x}) = \sum_{i \in I} u_i g_i(\mathbf{x}) \le \sum_{i \in I} u_i g_i(\bar{\mathbf{x}}) = \mathbf{u}'\mathbf{g}(\bar{\mathbf{x}})$, from (6.24), we get

$$f(\mathbf{x}) + \mathbf{u}'\mathbf{g}(\bar{\mathbf{x}}) \ge f(\mathbf{x}) + \mathbf{u}'\mathbf{g}(\mathbf{x}) \ge f(\bar{\mathbf{x}}) + \mathbf{u}'\mathbf{g}(\bar{\mathbf{x}})$$

which shows that $f(\mathbf{x}) \ge f(\bar{\mathbf{x}})$, and the proof is complete.

Corollary

Under the assumptions of the theorem, suppose that $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, and $\mathbf{u}'\mathbf{g}(\bar{\mathbf{x}}) = 0$. Then, $\bar{\mathbf{x}}$ is an optimal solution to the following problem:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0 \quad \text{for } i \in I \\ & h_i(\mathbf{x}) = 0 \quad \text{for } i = 1, \ldots, l \\ & \mathbf{x} \in X \end{aligned}$$

In particular, $\bar{\mathbf{x}}$ is an optimal solution to the original primal problem, and $(\mathbf{u}, \mathbf{v})$ is an optimal solution to the dual problem.

Proof

Note that $\mathbf{u}'\mathbf{g}(\bar{\mathbf{x}}) = 0$ implies that $g_i(\bar{\mathbf{x}}) = 0$ for $i \in I$, and from the theorem, it follows that $\bar{\mathbf{x}}$ solves the stated problem. Also, since the feasible region of the primal problem is contained in that of the above problem, and since $\bar{\mathbf{x}}$ is a feasible solution to the primal problem, then $\bar{\mathbf{x}}$ is an optimal solution to the primal problem. Furthermore, $f(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}) + \mathbf{u}'\mathbf{g}(\bar{\mathbf{x}}) + \mathbf{v}'\mathbf{h}(\bar{\mathbf{x}}) = \theta(\mathbf{u}, \mathbf{v})$ so that $(\mathbf{u}, \mathbf{v})$ solves the dual problem.

From the above theorem, as the dual function θ is evaluated at a given point $(\mathbf{u}, \mathbf{v})$, we obtain a point $\bar{\mathbf{x}}$ that is an optimal solution to a problem closely related to the original problem, in which the constraints are perturbed from $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ and $g_i(\mathbf{x}) \le 0$ for $i = 1, \ldots, m$ to $\mathbf{h}(\mathbf{x}) = \mathbf{h}(\bar{\mathbf{x}})$ and $g_i(\mathbf{x}) \le g_i(\bar{\mathbf{x}})$ for $i \in I$.

During the course of solving the dual problem, suppose that for a given $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \ge \mathbf{0}$, we have $\hat{\mathbf{x}} \in X(\mathbf{u}, \mathbf{v})$. Furthermore, for some $\varepsilon > 0$, suppose that $|g_i(\hat{\mathbf{x}})| \le \varepsilon$ for $i \in I$, $g_i(\hat{\mathbf{x}}) \le \varepsilon$ for $i \notin I$, and $|h_i(\hat{x})| \le \varepsilon$ for $i = 1, \ldots, l$. Note that if ε is sufficiently small, then $\hat{\mathbf{x}}$ is *near-feasible*. Now suppose $\bar{\mathbf{x}}$ is an optimal solution to the primal *Problem P*. Then, by definition of $\theta(\mathbf{u}, \mathbf{v})$,

$$f(\hat{\mathbf{x}}) + \sum_{i \in I} u_i g_i(\hat{\mathbf{x}}) + \sum_{i=1}^{l} v_i h_i(\hat{\mathbf{x}}) \le f(\bar{\mathbf{x}}) + \sum_{i \in I} u_i g_i(\bar{\mathbf{x}}) + \sum_{i=1}^{l} v_i h_i(\bar{\mathbf{x}}) \le f(\bar{\mathbf{x}})$$

since $h_i(\bar{\mathbf{x}}) = 0$, $g_i(\hat{\mathbf{x}}) \le 0$, and $u_i \ge 0$. The above inequality thus implies that

$$f(\bar{\mathbf{x}}) \le f(\hat{\mathbf{x}}) + \varepsilon\left[\sum_{i \in I} u_i + \sum_{i=1}^{l} |v_i|\right]$$

Thus, if ε is sufficiently small so that $\varepsilon[\sum_{i \in I} u_i + \sum_{i=1}^{l} |v_i|]$ is small enough, then we have a *near-optimal* solution. In many practical problems, such a solution is acceptable.

In the absence of a duality gap, Theorem 6.5.2 below shows that the complementary slackness condition is necessary for optimality.

6.5.2 Theorem

Suppose that $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ are optimal solutions of the primal and dual problems, respectively, and suppose that $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$. Then $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) = 0$ and $\bar{\mathbf{x}} \in X(\bar{\mathbf{u}}, \bar{\mathbf{v}})$; that is, $\bar{\mathbf{x}}$ solves the problem to minimize $f(\mathbf{x}) + \bar{\mathbf{u}}'\mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}'\mathbf{h}(\mathbf{x})$ subject to $\mathbf{x} \in X$.

Proof

Note, by definition of $\theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, that

$$\begin{aligned} f(\bar{\mathbf{x}}) + \bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) + \bar{\mathbf{v}}'\mathbf{h}(\bar{\mathbf{x}}) &\ge \inf\{f(\mathbf{x}) + \bar{\mathbf{u}}'\mathbf{g}(\mathbf{x}) + \bar{\mathbf{v}}'\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\} \\ &= \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}}) = f(\bar{\mathbf{x}}) \end{aligned} \tag{6.25}$$

Thus, $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) + \bar{\mathbf{v}}'\mathbf{h}(\bar{\mathbf{x}}) \ge 0$, and since $\mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}$, then $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) \ge 0$. Since $\bar{\mathbf{u}} \ge \mathbf{0}$ and $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) = 0$. Thus, from (6.25), it follows that $\bar{\mathbf{x}} \in X(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, and the proof is complete.

It may be noted that in the absence of a duality gap, the above theorem also shows that there exists an optimal solution to the primal problem among points in the set $X(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, where $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ is an optimal solution to the dual problem.

Generating Primal Feasible Solutions in the Convex Case

Under suitable convexity assumptions, we could easily obtain primal feasible solutions at each iteration of the dual problem by solving a linear program. In particular, suppose that we are given a point $\mathbf{x}_0$ which is feasible to the original problem, and let the points $\mathbf{x}_j \in X(\mathbf{u}_j, \mathbf{v}_j)$ for $j = 1, \ldots, k$ be generated. This is done during the process of maximizing the dual function by using any of the algorithms discussed in Section 6.4. Theorem 6.5.3 below shows that a feasible solution to the primal problem could be obtained by solving the following linear programming *problem P′*.

$$\textit{Problem } P': \quad \text{Minimize} \quad \sum_{j=0}^{k} \lambda_j f(\mathbf{x}_j)$$

$$\text{subject to} \quad \sum_{j=0}^{k} \lambda_j \mathbf{g}(\mathbf{x}_j) \le \mathbf{0}$$

$$\sum_{j=0}^{k} \lambda_j \mathbf{h}(\mathbf{x}_j) = \mathbf{0}$$

$$\sum_{j=0}^{k} \lambda_j = 1$$

$$\lambda_j \ge 0 \qquad \text{for } j = 0, \ldots, k$$

6.5.3 Theorem

Let X be a nonempty convex set in E_n, let $f: E_n \to E_1$ and $\mathbf{g}: E_n \to E_m$ be convex, and let $\mathbf{h}: E_n \to E_l$ be affine; that is, $\mathbf{h}$ is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. Let $\mathbf{x}_0$ be an initial feasible solution to *Problem P*, and suppose that $\mathbf{x}_j \in X(\mathbf{u}_j, \mathbf{v}_j)$ for $j = 1, \ldots, k$ are generated by any of the algorithms for solving the dual problem. Furthermore, let $\bar{\lambda}_j$ for $j = 0, \ldots, k$ be an optimal solution to *Problem P′*, and let $\bar{\mathbf{x}}_k = \sum_{j=0}^{k} \bar{\lambda}_j \mathbf{x}_j$. Then $\bar{\mathbf{x}}_k$ is a feasible solution to the primal *Problem P*. Furthermore, if $z_k - \theta(\mathbf{u}, \mathbf{v}) \le \varepsilon$ for some $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \ge \mathbf{0}$, then $f(\bar{\mathbf{x}}_k) \le \gamma + \varepsilon$, where $z_k = \sum_{j=0}^{k} \bar{\lambda}_j f(\mathbf{x}_j)$, and $\gamma = \inf\{f(\mathbf{x}) : \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$.

Proof

Since X is convex and $\mathbf{x}_j \in X$ for each j, then $\bar{\mathbf{x}}_k \in X$. Since $\mathbf{g}$ is convex and $\mathbf{h}$ is affine, and noting the constraints of *Problem P′*, $\mathbf{g}(\bar{\mathbf{x}}_k) \le \mathbf{0}$ and $\mathbf{h}(\bar{\mathbf{x}}_k) = \mathbf{0}$. Thus, $\bar{\mathbf{x}}_k$ is a feasible solution to the primal problem. Now suppose that $z_k - \theta(\mathbf{u}, \mathbf{v}) \le \varepsilon$ for some $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \ge \mathbf{0}$. Noting the convexity of f and Theorem 6.2.1, we get

$$f(\bar{\mathbf{x}}_k) \le \sum_{j=0}^{k} \bar{\lambda}_j f(\mathbf{x}_j) = z_k \le \theta(\mathbf{u}, \mathbf{v}) + \varepsilon \le \gamma + \varepsilon$$

and the proof is complete.

At each iteration of the dual maximization problem, we thus can obtain a primal feasible solution by solving the linear programming *problem P′*. Even though the primal objective values $\{f(\bar{\mathbf{x}}_k)\}$ of the generated primal feasible points are not necessarily decreasing, they form a sequence that is bounded from above by the nonincreasing sequence $\{z_k\}$.

Note that if z_k is close enough to the dual objective value evaluated at any dual feasible point $(\mathbf{u}, \mathbf{v})$, where $\mathbf{u} \ge \mathbf{0}$, then $\bar{\mathbf{x}}_k$ is a near-optimal primal feasible

solution. Also note that we need not solve *problem P′* in the case of the cutting plane algorithm, since it is precisely the linear programming dual of the master problem stated in step 1 of the cutting plane algorithm. Thus the optimal variables $\bar{\lambda}_0, \ldots, \bar{\lambda}_k$ could be retrieved easily from the solution to the master problem and $\bar{\mathbf{x}}_k$ computed as $\sum_{j=0}^{k} \bar{\lambda}_j \mathbf{x}_j$. It is also worth mentioning that the termination criterion $z_k = \theta(\mathbf{u}_k, \mathbf{v}_k)$ in the cutting plane algorithm could be interpreted as letting $(\mathbf{u}, \mathbf{v}) = (\mathbf{u}_k, \mathbf{v}_k)$ and $\varepsilon = 0$ in the above theorem.

To illustrate the above procedure, consider Example 6.4.4. At the end of iteration $k = 1$, we have the points $\mathbf{x}_0 = (5/4, 1/2)'$ and $\mathbf{x}_1 = (2, 0)'$. The associated primal point $\bar{\mathbf{x}}_1$ could be obtained by solving the following linear programming problem

$$\text{Minimize} \quad \tfrac{5}{8}\lambda_0$$

$$\text{subject to} \quad -\tfrac{3}{2}\lambda_0 + \lambda_1 \le 0$$

$$\lambda_0 + \lambda_1 = 1$$

$$\lambda_0, \ \lambda_1 \ge 0$$

The optimal solution to this problem is given by $\bar{\lambda}_0 = 2/5$ and $\bar{\lambda}_1 = 3/5$. This yields a primal feasible solution

$$\bar{\mathbf{x}} = \tfrac{2}{5}(\tfrac{5}{4}, \tfrac{1}{2})' + \tfrac{3}{5}(2, 0)' = (\tfrac{17}{10}, \tfrac{2}{10})'$$

As pointed out earlier, the above linear program need not be solved separately to find the values of $\bar{\lambda}_0$ and $\bar{\lambda}_1$ since its dual has already been solved during the course of the cutting plane algorithm.

6.6 Linear and Quadratic Programs

In this section we discuss some special cases of Lagrangian duality. In particular, we briefly discuss duality in linear and quadratic programming.

Linear Programming

Consider the following primal linear program:

$$\text{Minimize} \quad \mathbf{c}'\mathbf{x}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\mathbf{x} \ge \mathbf{0}$$

Letting $X = \{\mathbf{x} : \mathbf{x} \ge \mathbf{0}\}$, the Lagrangian dual of this problem is to maximize $\theta(\mathbf{v})$, where

$$\theta(\mathbf{v}) = \inf\{\mathbf{c}'\mathbf{x} + \mathbf{v}'(\mathbf{b} - \mathbf{A}\mathbf{x}) : \mathbf{x} \ge \mathbf{0}\} = \mathbf{v}'\mathbf{b} + \inf\{(\mathbf{c}' - \mathbf{v}'\mathbf{A})\mathbf{x} : \mathbf{x} \ge \mathbf{0}\}$$

Clearly,

$$\theta(\mathbf{v}) = \begin{cases} \mathbf{v}'\mathbf{b} & \text{if } (\mathbf{c}' - \mathbf{v}'\mathbf{A}) \geq \mathbf{0} \\ -\infty & \text{otherwise} \end{cases}$$

Hence, the dual problem can be stated as follows:

Maximize $\mathbf{v}'\mathbf{b}$

subject to $\mathbf{A}'\mathbf{v} \leq \mathbf{c}$

Thus, in the case of linear programs, the dual problem does not involve the primal variables. Furthermore, the dual problem itself is a linear program, and the reader could verify that the dual of the dual problem is the original primal program. Theorem 6.6.1 below summarizes the relationships between the primal and dual problems.

6.6.1 Theorem

Consider the primal and dual linear problems stated above. One of the following mutually exclusive cases will occur.

1. The primal problem admits a feasible solution and has an unbounded optimum objective value, in which case the dual problem is infeasible.
2. The dual problem admits a feasible solution and has an unbounded optimum objective value, in which case the primal problem is infeasible.
3. Both problems admit feasible solutions, in which case both problems have optimal solutions $\bar{\mathbf{x}}$ and $\bar{\mathbf{v}}$ such that $\mathbf{c}'\bar{\mathbf{x}} = \bar{\mathbf{v}}'\mathbf{b}$ and $(\mathbf{c}' - \bar{\mathbf{v}}'\mathbf{A})\bar{\mathbf{x}} = 0$.
4. Both problems are infeasible.

Proof

Let $\mathbf{x}$ and $\mathbf{v}$ be such that $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, and $\mathbf{A}'\mathbf{v} \leq \mathbf{c}$. Then, $\mathbf{v}'\mathbf{b} = \mathbf{v}'\mathbf{Ax} \leq \mathbf{c}'\mathbf{x}$. Therefore

$$\inf\{\mathbf{c}'\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\} \geq \sup\{\mathbf{v}'\mathbf{b} : \mathbf{A}'\mathbf{v} \leq \mathbf{c}\} \tag{6.26}$$

If the primal problem has an unbounded objective value, then from (6.26) the dual problem is infeasible. Similarly, from (6.26), if the dual problem has an unbounded objective value, then the primal is infeasible. Now suppose that both problems admit feasible solutions. Again from (6.26), $\inf\{\mathbf{c}'\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ is finite, and hence the primal problem must have an optimal solution, say $\bar{\mathbf{x}}$. From the Kuhn–Tucker conditions, there must exist a vector $\bar{\mathbf{v}} \in E_m$ such that

$$\mathbf{c} - \mathbf{A}'\bar{\mathbf{v}} \geq \mathbf{0} \tag{6.27}$$

$$(\mathbf{c}' - \bar{\mathbf{v}}'\mathbf{A})\bar{\mathbf{x}} = 0 \tag{6.28}$$

From (6.27), we see that $\bar{\mathbf{v}}$ is a feasible solution to the dual problem. Furthermore, from (6.28), we have

$$\mathbf{c}'\bar{\mathbf{x}} = \bar{\mathbf{v}}'\mathbf{A}\bar{\mathbf{x}} = \bar{\mathbf{v}}'\mathbf{b}$$

In view of (6.26), then, $\bar{\mathbf{v}}$ is an optimal solution to the dual problem. The last possible case is for both problems to be infeasible, and the proof is complete.

We now show how the optimal dual variables can be obtained if the simplex method is used for solving the primal problem. Let the optimal basic solution be $\bar{\mathbf{x}}' = (\bar{\mathbf{x}}'_B, \bar{\mathbf{x}}'_N)$, where $\bar{\mathbf{x}}_N = 0$ and $\bar{\mathbf{x}}_B = \mathbf{B}^{-1}\mathbf{b}$. The matrix $\mathbf{A}$ and vector $\mathbf{c}$ are also partitioned to give $\mathbf{A} = [\mathbf{B}, \mathbf{N}]$ and $\mathbf{c}' = (\mathbf{c}'_B, \mathbf{c}'_N)$.

The Kuhn–Tucker condition (6.28) can be written as $(\mathbf{c}'_B - \bar{\mathbf{v}}'\mathbf{B})\bar{\mathbf{x}}_B + (\mathbf{c}'_N - \bar{\mathbf{v}}'\mathbf{N})\bar{\mathbf{x}}_N = 0$. Since $\bar{\mathbf{x}}_N = \mathbf{0}$, this condition is satisfied by letting $\mathbf{c}'_B - \bar{\mathbf{v}}'\mathbf{B} = \mathbf{0}$ or

$$\bar{\mathbf{v}}' = \mathbf{c}'_B\mathbf{B}^{-1} \tag{6.29}$$

The Kuhn–Tucker condition (6.27) can be written as $\mathbf{c}'_B - \bar{\mathbf{v}}'\mathbf{B} \geq \mathbf{0}$ and $\mathbf{c}'_N - \bar{\mathbf{v}}'\mathbf{N} \geq \mathbf{0}$. Letting $\bar{\mathbf{v}}' = \mathbf{c}'_B\mathbf{B}^{-1}$, the first inequality is satisfied, and the latter becomes

$$\mathbf{c}'_B\mathbf{B}^{-1}\mathbf{N} - \mathbf{c}'_N \leq \mathbf{0}$$

which is precisely the optimality condition for the simplex method.

Recall that at each iteration of the simplex method, row 0 displays the vector $\mathbf{c}'_B\mathbf{B}^{-1}\mathbf{A} - \mathbf{c}'$. Suppose that the matrix $\mathbf{A}$ contains an identity submatrix, and suppose that the cost coefficients of the corresponding variables are given by the vector $\mathbf{c}_I$. Then from (6.29), $\hat{\mathbf{c}}'_I = \mathbf{c}'_B\mathbf{B}^{-1} - \mathbf{c}'_I = \bar{\mathbf{v}}' - \mathbf{c}'_I$ is given in row 0 in the updated tableau under the original identity matrix. Adding $\mathbf{c}'_I$ to $\hat{\mathbf{c}}'_I$ in the final tableau yields the optimal values of the dual variables.

Quadratic Programming

Consider the following quadratic programming problem:

Minimize $\frac{1}{2}\mathbf{x}'\mathbf{Hx} + \mathbf{d}'\mathbf{x}$

subject to $\mathbf{Ax} \leq \mathbf{b}$

where $\mathbf{H}$ is symmetric and positive definite, so that the objective function is strictly convex. The Lagrangian dual problem is to maximize $\theta(\mathbf{u})$ over $\mathbf{u} \geq \mathbf{0}$, where

$$\theta(\mathbf{u}) = \inf\{\tfrac{1}{2}\mathbf{x}'\mathbf{Hx} + \mathbf{d}'\mathbf{x} + \mathbf{u}'(\mathbf{Ax} - \mathbf{b}) : \mathbf{x} \in E_n\} \tag{6.30}$$

Note that for a given $\mathbf{u}$, the function $\frac{1}{2}\mathbf{x}'\mathbf{Hx} + \mathbf{d}'\mathbf{x} + \mathbf{u}'(\mathbf{Ax} - \mathbf{b})$ is strictly convex and achieves its minimum at a point satisfying

$$\mathbf{Hx} + \mathbf{A}'\mathbf{u} + \mathbf{d} = \mathbf{0} \tag{6.31}$$

Thus the dual problem could be written as follows:

$$\begin{aligned} \text{Maximize} \quad & \tfrac{1}{2}\mathbf{x}'\mathbf{H}\mathbf{x}+\mathbf{d}'\mathbf{x}+\mathbf{u}'(\mathbf{A}\mathbf{x}-\mathbf{b}) \\ \text{subject to} \quad & \mathbf{H}\mathbf{x}+\mathbf{A}'\mathbf{u}=-\mathbf{d} \\ & \mathbf{u}\geq\mathbf{0} \end{aligned}$$

We now develop an alternative form of the Lagrangian dual problem. Since $\mathbf{H}$ is positive definite, then $\mathbf{H}^{-1}$ exists, and the unique solution to (6.31) is given by

$$\mathbf{x}=-\mathbf{H}^{-1}(\mathbf{d}+\mathbf{A}'\mathbf{u})$$

Substituting in (6.30), it follows that

$$\theta(\mathbf{u})=\tfrac{1}{2}\mathbf{u}'\mathbf{D}\mathbf{u}+\mathbf{u}'\mathbf{c}-\tfrac{1}{2}\mathbf{d}'\mathbf{H}^{-1}\mathbf{d}$$

where $\mathbf{D}=-\mathbf{A}\mathbf{H}^{-1}\mathbf{A}'$, and $\mathbf{c}=-\mathbf{b}-\mathbf{A}\mathbf{H}^{-1}\mathbf{d}$. The dual problem is thus given by:

$$\begin{aligned} \text{Maximize} \quad & \tfrac{1}{2}\mathbf{u}'\mathbf{D}\mathbf{u}+\mathbf{u}'\mathbf{c}-\tfrac{1}{2}\mathbf{d}'\mathbf{H}^{-1}\mathbf{d} \\ \text{subject to} \quad & \mathbf{u}\geq\mathbf{0} \end{aligned}$$

The dual problem can be solved relatively easily by the following scheme. Given $\mathbf{u}$, let $\nabla\theta(\mathbf{u})=\mathbf{D}\mathbf{u}+\mathbf{c}=\mathbf{g}$. Consider $\hat{\mathbf{g}}$ as defined below:

$$\hat{g}_i=\begin{cases} g_i & \text{if } u_i>0 \text{ or } g_i\geq 0 \\ 0 & \text{if } u_i=0 \text{ and } g_i<0 \end{cases}$$

By Theorem 6.4.1, if $\hat{\mathbf{g}}=\mathbf{0}$, then $\mathbf{u}$ is an optimal solution. Otherwise $\hat{\mathbf{g}}$ is an improving feasible direction. Optimizing θ starting from $\mathbf{u}$ along the direction $\hat{\mathbf{g}}$ without violating the nonnegativity restriction leads to a new point. The process is then repeated.

Exercises

6.1 Consider the problem to minimize x_1 subject to $x_1^2+x_2^2=1$. Derive explicitly the dual function, and verify its concavity. Find the optimal solutions to both the primal and dual problems, and compare their objective values.

6.2 Consider the following problem:

$$\begin{aligned} \text{Maximize} \quad & 2x_1+3x_2+x_3 \\ \text{subject to} \quad & x_1+x_2-x_3\leq 1 \\ & x_1+x_2 \leq 4 \\ & x_3\leq 2 \\ & x_1,\ x_2,\ x_3\geq 0 \end{aligned}$$

a. Find explicitly the dual function, where $X=\{(x_1, x_2, x_3): x_1+x_2-x_3\leq 1;\ x_1, x_2, x_3\geq 0\}$.

b. Repeat part a for $X=\{(x_1, x_2, x_3): x_1+x_2\leq 4;\ x_1, x_2, x_3\geq 0\}$

c. In parts a and b, note that the difficulty in evaluating the dual function at a given point depends on which constraints are handled via the set X. Propose some general guidelines that could be used in selecting the set X to make the solution easier.

6.3 Consider the problem to minimize e^{-x} subject to $-x\leq 0$.

a. Solve the above primal problem.

b. Letting $X=E_1$, find the explicit form of the Lagrangian dual function, and solve the dual problem.

6.4 Consider the primal *problem P* discussed in Section 6.1. Introducing the slack vector $\mathbf{s}$, the problem can be formulated as follows:

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x})+\mathbf{s}=\mathbf{0} \\ & \mathbf{h}(\mathbf{x})=\mathbf{0} \\ & (\mathbf{x},\mathbf{s})\in X' \end{aligned}$$

where $X'=\{(\mathbf{x},\mathbf{s}): \mathbf{x}\in X, \mathbf{s}\geq\mathbf{0}\}$. Formulate the dual of the above problem and show that it is equivalent to the dual problem discussed in Section 6.1.

6.5 In the proof of Lemma 6.2.3, show that the set Λ is convex.

6.6 Under the assumptions of Theorem 6.2.5, suppose that $\bar{\mathbf{x}}$ is an optimal solution to the primal problem, and that f and $\mathbf{g}$ are differentiable at $\bar{\mathbf{x}}$. Show that there exists a vector $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ such that

$$\left[\nabla f(\bar{\mathbf{x}})+\sum_{i=1}^{m}\bar{u}_i\nabla g_i(\bar{\mathbf{x}})+\sum_{i=1}^{l}\bar{v}_i\nabla h_i(\bar{\mathbf{x}})\right]'(\mathbf{x}-\bar{\mathbf{x}})\geq 0 \qquad \text{for each } \mathbf{x}\in X$$

$$u_i g_i(\bar{\mathbf{x}})=0 \qquad \text{for } i=1,\ldots,m$$

$$\bar{\mathbf{u}}\geq\mathbf{0}$$

Show that these conditions reduce to the Kuhn–Tucker conditions if X is open.

6.7 Prove the following saddle point optimality condition. Let X be a nonempty convex set in E_n, and let $f: E_n\to E_1$, $\mathbf{g}: E_n\to E_m$ be convex, and $\mathbf{h}: E_n\to E_l$ be affine. If $\bar{\mathbf{x}}$ is an optimal solution to the problem to minimize $f(\mathbf{x})$ subject to

$g(x) \le 0$, $h(x) = 0$, $x \in X$, then there exist $(\bar{u}_0, \bar{u}, \bar{v}) \neq 0$, $(\bar{u}_0, \bar{u}) \ge 0$ such that:

$$\phi(\bar{u}_0, u, v, \bar{x}) \le \phi(\bar{u}_0, \bar{u}, \bar{v}, \bar{x}) \le \phi(\bar{u}_0, \bar{u}, \bar{v}, x)$$

for all $u \ge 0$, $v \in E_l$, and $x \in X$, where $\phi(u_0, u, v, x) = u_0 f(x) + u'g(x) + v'h(x)$.

6.8 Consider the problem to minimize $f(x)$ subject to $g(x) \le 0$, $x \in X$. Theorem 6.2.4 shows that the primal and dual objective values are equal at optimality under the assumptions that f, g, and X are convex and the constraint qualification that $g(\hat{x}) < 0$ for some $\hat{x} \in X$. Suppose that the convexity assumptions on f and g are replaced by continuity of f and g and that X is assumed to be convex and compact. Does the result of the theorem hold? Prove or give a counterexample.

6.9 Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & -2x_1 + 2x_2 + x_3 - 3x_4 \\ \text{subject to} & x_1 + x_2 + x_3 + x_4 \le 8 \\ & x_1 - 2x_3 + 4x_4 \le 2 \\ & x_1 + x_2 \le 8 \\ & x_3 + 2x_4 \le 6 \\ & x_1, \; x_2, \; x_3, \; x_4 \ge 0 \end{array}$$

Let $X = \{(x_1, x_2, x_3, x_4) : x_1 + x_2 \le 8, x_3 + 2x_4 \le 6; x_1, x_2, x_3, x_4 \ge 0\}$.

a. Find the function θ explicitly.
b. Verify that θ is differentiable at $(4, 0)$, and find $\nabla\theta(4, 0)$.
c. Verify that $\nabla\theta(4, 0)$ is an infeasible direction, and find an improving feasible direction.
d. Starting from $(4, 0)$ maximize θ in the direction obtained in part c.

6.10 Consider the problem to minimize $x_1^2 + x_2^2$ subject to $x_1 + x_2 - 4 \le 0$, and $x_1, x_2 \ge 0$.

a. Verify that the optimal solution is $\bar{x} = (2, 2)'$ with $f(\bar{x}) = 8$.
b. Letting $X = \{(x_1, x_2) : x_1 \ge 0, x_2 \ge 0\}$, write the Lagrangian dual problem. Show that the dual function is $\theta(u) = -u^2/2 - 4u$. Verify that there is no duality gap for this problem.
c. Solve the dual problem by the cutting plane algorithm of Section 6.4. Start with $x = (1, 1)'$.
d. Show that θ is differentiable everywhere, and solve the problem using the gradient method of Section 6.4.

6.11 Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & x_1 + x_2 \\ \text{subject to} & 2x_1 + x_2 \le 8 \\ & 3x_1 + 2x_2 \le 10 \\ & x_1, \; x_2 \ge 0 \\ & x_1, \; x_2 \text{ integers} \end{array}$$

Let $X = \{(x_1, x_2) : 3x_1 + 2x_2 \le 10, x_1, x_2 \ge 0 \text{ and integer}\}$. At $u = 2$, is θ differentiable? If not characterize its ascent directions.

6.12 Consider the following problem.

$$\begin{array}{ll} \text{Minimize} & (x_1 - 3)^2 + (x_2 - 5)^2 \\ \text{subject to} & x_1^2 - x_2 \le 0 \\ & -x_1 \le 1 \\ & x_1 + 2x_2 \le 10 \\ & x_1, \; x_2 \ge 0 \end{array}$$

a. Find the optimal solution geometrically, and verify it by using the Kuhn–Tucker conditions.
b. Formulate the dual problem in which $X = \{(x_1, x_2) : x_1 + 2x_2 \le 10;\; x_1, x_2 \ge 0\}$.
c. Perform three iterations of the gradient maximization technique described in Section 6.4, starting with $(u_1, u_2) = (0, 0)$. Describe the perturbed optimization problems corresponding to the generated primal infeasible points.

6.13 In reference to Exercise 6.12 above, perform three iterations of the cutting plane algorithm and compare the results with those obtained by gradient maximization. Also identify the primal feasible solutions generated by the algorithm.

6.14 Consider the following problem.

$$\begin{array}{ll} \text{Maximize} & 3x_1 + 6x_2 + 2x_3 + 4x_4 \\ \text{subject to} & x_1 + x_2 + x_3 + x_4 \le 12 \\ & -x_1 + x_2 + 2x_4 \le 4 \\ & x_1 + x_2 \le 12 \\ & x_2 \le 4 \\ & x_3 + x_4 \le 6 \\ & x_1, \; x_2, \; x_3, \; x_4 \ge 0 \end{array}$$

a. Formulate the dual problem in which $X = \{(x_1, x_2, x_3, x_4) : x_1 + x_2 \le 12, x_2 \le 4, x_3 + x_4 \le 6;\; x_1, x_2, x_3, x_4 \ge 0\}$.
b. Starting from the point $(0, 0)$, solve the Lagrangian dual problem by optimizing along the direction of steepest ascent discussed in Section 6.4.
c. At optimality of the dual, find the optimal primal solution.

6.15 Consider the problem to minimize x subject to $g(x) \le 0$ and $x \in X = \{x : x \ge 0\}$. Derive the explicit forms of the Lagrangian dual function, and determine the collection of subgradients at $u = 0$ for each of the following cases.

a. $g(x) = \begin{cases} -1/x & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$

b. $g(x) = \begin{cases} -1/x & \text{for } x \neq 0 \\ -1 & \text{for } x = 0 \end{cases}$

c. $g(x) = \begin{cases} 1/x & \text{for } x \neq 0 \\ 1 & \text{for } x = 0 \end{cases}$

6.16 Suppose that $\theta : E_m \to E_1$ is concave.

a. Show that θ achieves its maximum at $\bar{\mathbf{u}}$ if and only if

$$\text{maximum } \{\theta'(\bar{\mathbf{u}}; \mathbf{d}) : \|\mathbf{d}\| \le 1\} = 0$$

b. Show that θ achieves its maximum over the region $U = \{\mathbf{u} : \mathbf{u} \ge \mathbf{0}\}$ at $\bar{\mathbf{u}}$ if and only if

$$\text{maximum } \{\theta'(\bar{\mathbf{u}}; \mathbf{d}) : \mathbf{d} \in D, \|\mathbf{d}\| \le 1\} = 0$$

where D is the cone of feasible directions of U at $\bar{\mathbf{u}}$.

(Note that the above results could be used as stopping criteria for maximizing the Lagrangian dual function.)

6.17 Consider the following problem, in which X is a compact polyhedral set and f is a concave function.

$$\begin{aligned} &\text{Minimize} && f(\mathbf{x}) \\ &\text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b} \\ &&& \mathbf{x} \in X \end{aligned}$$

a. Formulate the Lagrangian dual problem.
b. Show that the dual function is concave and piecewise linear.
c. Characterize the subgradients, the ascent directions, and the steepest ascent direction for the dual function.
d. Generalize the result in part b to the case where X is not compact.

6.18 Construct a numerical problem in which a subgradient of the dual function is not an ascent direction. Is it possible that the collection of subgradients and the cone of ascent directions are disjoint?

(*Hint:* Consider the shortest subgradient.)

6.19 In Section 6.3 we showed that the shortest subgradient $\boldsymbol{\xi}$ of θ at $\bar{\mathbf{u}}$ is the steepest ascent direction. The following modification of $\boldsymbol{\xi}$ is proposed to maintain feasibility:

$$\bar{\xi}_i = \begin{cases} \text{maximum } (0, \xi_i) & \text{if } \bar{u}_i = 0 \\ \xi_i & \text{if } \bar{u}_i \ge 0 \end{cases}$$

Is $\bar{\boldsymbol{\xi}}$ an ascent direction? Is it the direction of steepest ascent with the added nonnegativity restriction? Prove or give a counterexample.

6.20 Consider the following problem, in which X is a compact polyhedral set.

$$\begin{aligned} &\text{Minimize} && \mathbf{c}'\mathbf{x} \\ &\text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b} \\ &&& \mathbf{x} \in X \end{aligned}$$

For a given vector $\mathbf{v}$, suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are the extreme points in X that belong to $X(\mathbf{v})$. Show that the extreme points of $\partial\theta(\mathbf{v})$ are contained in the set $\Lambda = \{\mathbf{A}\mathbf{x}_j - \mathbf{b} : j = 1, \ldots, k\}$. Give an example where the extreme points of $\partial\theta(\mathbf{v})$ form a proper subset of Λ.

6.21 Suppose that the shortest subgradient $\bar{\boldsymbol{\xi}}$ of θ at $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ is not equal to zero. Show that there exists an $\varepsilon > 0$ such that $\|\boldsymbol{\xi} - \bar{\boldsymbol{\xi}}\| < \varepsilon$ implies that $\boldsymbol{\xi}$ is an ascent direction of θ at $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$.

(From the above exercise, if an iterative procedure is used to find $\bar{\boldsymbol{\xi}}$, then it would find an ascent direction after a sufficient number of iterations.)

6.22 Consider the primal and Lagrangian dual problems discussed in Section 6.1. Let $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ be an optimal solution to the dual problem. Given $(\mathbf{u}, \mathbf{v})$, suppose that $\bar{\mathbf{x}} \in X(\mathbf{u}, \mathbf{v})$. Show that there exists a $\delta > 0$ such that $\|(\bar{\mathbf{u}}, \bar{\mathbf{v}}) - (\mathbf{u}, \mathbf{v}) - \lambda[\mathbf{g}(\bar{\mathbf{x}}), \mathbf{h}(\bar{\mathbf{x}})]\|$ is a nonincreasing function of λ over the interval $[0, \delta]$. Interpret the result geometrically, and illustrate by the following problem, in which $(u_1, u_2) = (3, 1)$ are the dual variables corresponding to the first two constraints.

$$\begin{aligned} &\text{Minimize} && -2x_1 - 2x_2 - 5x_3 \\ &\text{subject to} && x_1 + x_2 + x_3 \le 10 \\ &&& x_1 \qquad + 2x_3 \ge 6 \\ &&& x_1, \quad x_2, \quad x_3 \le 3 \\ &&& x_1, \quad x_2, \quad x_3 \ge 0 \end{aligned}$$

6.23 From Exercise 6.22 above, it is clear that moving a small step in the direction of any subgradient leads us closer to an optimal dual solution. Consider the following algorithm for maximizing the dual of the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, $\mathbf{x} \in X$.

Main Step

Given $\mathbf{v}_k$, let $\mathbf{x}_k \in X(\mathbf{v}_k)$. Let $\mathbf{v}_{k+1} = \mathbf{v}_k + \lambda \mathbf{h}(\mathbf{x}_k)$, where $\lambda > 0$ is a small scalar. Replace k by $k+1$ and repeat the main step.

a. Discuss some possible ways of choosing a suitable step size λ. Do you see any advantages in reducing the step size during later iterations? If so, propose a scheme for doing that.
b. Does the dual function necessarily increase from one iteration to another? Discuss.
c. Devise a suitable termination criterion.
d. Apply the above algorithm, starting from $\mathbf{v} = (1, 2)'$ to solve the following problem:

$$\begin{aligned} &\text{Minimize} && x_1^2 + x_2^2 + 2x_3 \\ &\text{subject to} && x_1 + x_2 + x_3 = 6 \\ &&& -x_1 + x_2 + x_3 = 4 \end{aligned}$$

(This procedure is referred to as a *subgradient optimization* technique.)

6.24 Consider the problem to minimize $f(\mathbf{x})$ subject to $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$, $\mathbf{x} \in X$.

a. In Exercise 6.23 above, a subgradient optimization technique was discussed for the equality case. Modify the procedure for the above inequality constrained problem.

Hint: Given $\mathbf{u}$, let $\mathbf{x} \in X(\mathbf{u})$. Replace $g_i(\mathbf{x})$ by maximum $[0, g_i(\mathbf{x})]$ for each i with $u_i = 0$.

b. Illustrate the procedure given in part a by solving the problem in Exercise 6.14 starting from $\mathbf{u} = (0, 0)'$.
c. Extend the subgradient optimization technique to handle both equality and inequality constraints.

6.25 Consider the following warehouse location problem. We are given destinations $1, \ldots, k$, where the known demand for a certain product at destination j is d_j. We are also given m possible sites for building warehouses. If we decide to build a warehouse at site i, its capacity has to be k_i and incurs a fixed cost f_i. The unit shipping cost from warehouse i to destination j is c_{ij}. The problem is to determine how many warehouses to build, where to locate them, and what shipping patterns to use so that the demand is satisfied and the total cost is minimized.
The problem can be stated mathematically as follows:

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^{m}\sum_{j=1}^{k} c_{ij}x_{ij} + \sum_{i=1}^{m} f_i y_i \\ \text{subject to} \quad & \sum_{j=1}^{k} x_{ij} \le k_i y_i && \text{for } i = 1, \ldots, m \\ & \sum_{i=1}^{m} x_{ij} \ge d_j && \text{for } j = 1, \ldots, k \\ & x_{ij} \ge 0 && \text{for } i = 1, \ldots, m; j = 1, \ldots, k \\ & y_i = 0 \text{ or } 1 && \text{for } i = 1, \ldots, m \end{aligned}$$

a. Formulate a suitable Lagrangian dual problem.
b. Make use of the results of this chapter to devise a special scheme for maximizing the dual of the warehouse location problem.
c. Illustrate by a small numerical example.

6.25 A company wants to plan its production rate of a certain item over the planning period $[0, T]$ such that the sum of its production and inventory costs is minimized. In addition, the known demand must be met, the production rate must fall in the acceptable interval $[l, u]$, the inventory must not exceed d, and it must be at least equal to b at the end of the planning period. The problem can be formulated as follows:

$$\begin{aligned} \text{Minimize} \quad & \int_0^T [c_1 x(t) + c_2 y^2(t)]\, dt \\ \text{subject to} \quad & x(t) = x_0 + \int_0^t [y(\tau) - z(\tau)]\, d\tau && \text{for } t \in [0, T] \\ & x(T) \ge b \\ & 0 \le x(t) \le d && \text{for } t \in (0, T) \\ & l \le y(t) \le u && \text{for } t \in (0, T) \end{aligned}$$

where $x(t)$ = inventory at time t
$y(t)$ = production rate at time t
$z(t)$ = known demand rate at time t
x_0 = known initial inventory
c_1, c_2 = known coefficients

a. Make the above control problem discrete, as done in Section 1.2, and formulate a suitable Lagrangian dual problem.
b. Make use of the results of this chapter to develop a scheme of solving the primal and dual problems.
c. Apply your algorithm to the following data:
$T = 6$, $x_0 = 0$, $b = 4$, $c_1 = 1$, $c_2 = 2$, $l = 2$, $u = 5$, $d = 6$, and $z(t) = 4$ over $[0, 4]$ and $z(t) = 3$ over $(4, 6]$.

6.26 Consider the primal and dual linear programming problems discussed in Section 6.6. Show directly that:
a. If the primal is inconsistent and the dual admits a feasible solution, then the dual has an unbounded optimal objective value.
b. If the dual is inconsistent and the primal admits a feasible solution, then the primal has an unbounded optimal objective value.
(*Hint:* Use Farkas' theorem.)

6.27 Consider the linear program to minimize $\mathbf{c}'\mathbf{x}$ subject to $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \ge \mathbf{0}$. Write the dual problem. Show that the dual of the dual problem is equivalent to the primal problem.

6.28 Consider the following problem:

$$\begin{aligned} \text{Minimize} \quad & -x_1 - 2x_2 - x_3 \\ \text{subject to} \quad & x_1 + x_2 + x_3 \le 16 \\ & x_1 - x_2 + 3x_3 \le 12 \\ & x_1 + x_2 \le 4 \\ & x_1, \quad x_2, \quad x_3 \ge 0 \end{aligned}$$

Solve the primal problem by the simplex method. At each iteration identify the dual variables from the simplex tableau. Show that the dual variables satisfy the complementary slackness conditions but violate the dual constraints. Verify that dual feasibility is reached at termination.

6.29 Consider the following quadratic programming problem:

$$\begin{aligned} \text{Minimize} \quad & 2x_1^2 + x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \\ \text{subject to} \quad & x_1 + x_2 \le 8 \\ & -x_1 + 2x_2 \le 10 \\ & x_1, \quad x_2 \ge 0 \end{aligned}$$

Solve the Lagrangian dual problem by the method of Section 6.6. At each iteration, identify the corresponding primal infeasible point as well as the primal feasible point. Develop a suitable measure of infeasibility and check its progress. Can you draw any general conclusions?

6.30 Consider the problem to find

$$\underset{\mathbf{x}\in X}{\text{minimum}}\ \underset{\mathbf{y}\in Y}{\text{maximum}}\ \phi(\mathbf{x}, \mathbf{y}) \quad \text{and} \quad \underset{\mathbf{y}\in Y}{\text{maximum}}\ \underset{\mathbf{x}\in X}{\text{minimum}}\ \phi(\mathbf{x}, \mathbf{y})$$

where X and Y are nonempty compact convex sets in E_n and E_m, respectively, and ϕ is convex in $\mathbf{x}$ for any given $\mathbf{y}$, and concave in $\mathbf{y}$ for any given $\mathbf{x}$.

a. Show that $\underset{x \in X}{\text{minimum}}\ \underset{y \in Y}{\text{maximum}}\ \phi(\mathbf{x}, \mathbf{y}) \geq \underset{y \in Y}{\text{maximum}}\ \underset{x \in X}{\text{minimum}}\ \phi(\mathbf{x}, \mathbf{y})$ without any convexity assumptions.

b. Show that $\underset{y \in Y}{\text{maximum}}\ \phi(., \mathbf{y})$ is a convex function in $\mathbf{x}$ and that $\underset{x \in X}{\text{minimum}}\ \phi(\mathbf{x}, .)$ is a concave function in $\mathbf{y}$.

c. Show that

$$\underset{x \in X}{\text{minimum}}\ \underset{y \in Y}{\text{maximum}}\ \phi(\mathbf{x}, \mathbf{y}) = \underset{y \in Y}{\text{maximum}}\ \underset{x \in X}{\text{minimum}}\ \phi(\mathbf{x}, \mathbf{y})$$

(*Hint:* Use part b and the necessary optimality conditions of Section 3.4.)

6.31 Let X and Y be nonempty sets in E_n, and let $f, g: E_n \to E_1$. Consider the *conjugate functions* f^* and g^* defined as follows:

$$f^*(\mathbf{u}) = \inf\{f(\mathbf{x}) - \mathbf{u}'\mathbf{x} : \mathbf{x} \in X\}$$

$$g^*(\mathbf{u}) = \sup\{g(\mathbf{x}) - \mathbf{u}'\mathbf{x} : \mathbf{x} \in Y\}$$

a. Interpret f^* and g^* geometrically.

b. Show that f^* is concave over X^* and g^* is convex over Y^*, where $X^* = \{\mathbf{u} : f^*(\mathbf{u}) > -\infty\}$ and $Y^* = \{\mathbf{u} : g^*(\mathbf{u}) < \infty\}$.

c. Prove the following *conjugate weak duality theorem:*

$$\inf\{f(\mathbf{x}) - g(\mathbf{x}) : \mathbf{x} \in X \cap Y\} \geq \sup\{f^*(\mathbf{u}) - g^*(\mathbf{u}) : \mathbf{u} \in X^* \cap Y^*\}.$$

d. Now suppose that f is convex, g is concave, $\text{int}\, X \cap \text{int}\, Y \neq \varnothing$, and that $\inf\{f(\mathbf{x}) - g(\mathbf{x}) : \mathbf{x} \in X \cap Y\}$ is finite. Show that equality in part c above holds true and that $\sup\{f^*(\mathbf{u}) - g^*(\mathbf{u}) : \mathbf{u} \in X^* \cap Y^*\}$ is achieved.

e. By suitable choices of f, g, X, and Y, formulate a nonlinear programming problem as follows:

Minimize $\quad f(\mathbf{x}) - g(\mathbf{x})$

subject to $\quad \mathbf{x} \in X \cap Y$

What is the form of the conjugate dual problem? Devise some strategies for solving the dual problem.

6.32 Consider a single constrained problem to minimize $f(\mathbf{x})$ subject to $g(\mathbf{x}) \leq 0$ and $\mathbf{x} \in X$, where X is a compact set. The Lagrangian dual problem is to maximize $\theta(u)$ subject to $u \geq 0$, where $\theta(u) = \inf\{f(\mathbf{x}) + ug(\mathbf{x}) : \mathbf{x} \in X\}$

a. Let $\hat{u} \geq 0$, and let $\hat{\mathbf{x}} \in X(\hat{u})$. Show that if $g(\hat{\mathbf{x}}) > 0$ then $\bar{u} > \hat{u}$, and if $g(\hat{\mathbf{x}}) < 0$ then $\bar{u} < \hat{u}$, where $\bar{u}$ is an optimal solution to the Lagrangian dual.

b. Use the result of part a to find an interval $[a, b]$ that contains all the optimal solutions to the dual problem or else concludes that the dual problem is unbounded.

c. Now consider the problem to maximize $\theta(u)$ subject to $a \leq u \leq b$. The following scheme is used to solve the problem.

Let $\bar{u} = (a+b)/2$, and let $\bar{\mathbf{x}} \in X(\bar{u})$. If $g(\bar{\mathbf{x}}) > 0$ then replace a by $\bar{u}$, and repeat the process. If $g(\bar{\mathbf{x}}) < 0$, replace b by $\bar{u}$, and repeat the process. If $g(\bar{\mathbf{x}}) = 0$, stop; $\bar{u}$ is an optimal dual solution.

Show that the procedure converges to an optimal solution, and illustrate by solving the dual of the following problem.

Minimize $\quad x_1^2 + x_2^2$

subject to $\quad -x_1 - x_2 + 1 \leq 0$

d. An alternative approach to solving the problem to maximize $\theta(u)$ subject to $a \leq u \leq b$ is to specialize the tangential approximation method discussed in Section 6.4. Show that at each iteration only two supporting hyperplanes need be considered, and that the method could be stated as follows.

Let $\mathbf{x}_a \in X(a)$ and $\mathbf{x}_b \in X(b)$. Let $\bar{u} = (f(\mathbf{x}_a) - f(\mathbf{x}_b))/(g(\mathbf{x}_b) - g(\mathbf{x}_a))$. If $\bar{u} = a$ or $\bar{u} = b$, stop; $\bar{u}$ is an optimal solution to the dual problem. Otherwise, let $\bar{\mathbf{x}} \in X(\bar{u})$. If $g(\bar{\mathbf{x}}) > 0$, replace a by $\bar{u}$, and repeat the process. If $g(\bar{\mathbf{x}}) < 0$, replace b by $\bar{u}$, and repeat the process. If $g(\bar{\mathbf{x}}) = 0$, stop; $\bar{u}$ is an optimal dual solution.

Show that the procedure converges to an optimal solution, and illustrate by solving the problem in part (c).

Notes and References

The powerful results of duality in linear programming and the saddle point optimality criteria for convex programming sparked a great deal of interest in duality in nonlinear programming. Early results in this area include the work of Dorn [1960], Hanson [1961], Mangasarian [1962], Stoër [1963], and Wolfe [1961].

More recently, several duality formulations that enjoy many of the properties of linear dual programs have evolved. These include the Lagrangian dual problem, the conjugate dual problem, and the surrogate dual problem. In this chapter we concentrated on the Lagrangian dual formulation because, in our judgment, it is the most promising formulation from a computational standpoint and also because the results of this chapter give the general flavor of the results that one would obtain using other duality formulations. Those interested in studying the subject of conjugate duality may refer to Fenchel [1949], Rockafellar [1964, 1966, 1968, 1969, 1970], and Whinston [1967]. For the subject of surrogate duality, where the constraints are grouped into a single constraint by the use of Lagrangian multipliers, refer to Greenberg and Pierskalla [1970b]. Several authors have developed duality formulations that retain the symmetry between the primal and dual problems. The works of Dantzig, Eisenberg, and Cottle [1965], Mangasarian and Ponstein [1965], and Stoër [1963] are in this class.

The reader will find the work of Geoffrion [1971b] and Karamardian [1967] as excellent references on various duality formulations and their interrelationships. See Everett [1963], Falk [1967, 1969], and Lasdon [1968] for further study on duality. The relationship between the Lagrangian duality formulation and other duality formulations is examined in Bazaraa, Goode, and Shetty [1971], Magnanti [1974], and Whinston [1967]. The economic interpretation of duality is covered by Balinski and Baumol [1968], Beckman and Kapur [1972], Peterson [1970], and Williams [1970].

In Sections 6.1 and 6.2, the dual problem is presented, and some of its properties are developed. As a by-product of the main duality theorem, we develop the saddle point optimality criteria for convex programs. These criteria were first developed by Kuhn and Tucker [1951]. For the related concept of min–max duality, see Mangasarian and Ponstein [1965], Ponstein [1965], Rockafellar [1968], and Stoër [1963].

In Section 6.3, we examine several properties of the dual function. We characterize the collection of subgradients at any given point, and use that to determine both ascent directions and the steepest ascent direction. We show that the steepest ascent direction is the shortest subgradient. This result is essentially given by Demyanov [1968]. In Section 6.4, we use these properties to develop several schemes for maximizing the dual function. In particular, we describe the gradient method and a decomposition method for generating ascent directions. For further study of this subject see Demyanov [1968, 1971], Fisher, Northrup and Shapiro [1975], and Lasdon [1970]. There are other procedures for solving the dual problem. One cutting plane method discussed in Section 6.4 is a row generation procedure. In its dual form, it is precisely the column generation generalized programming method of Wolfe, see Dantzig [1963]. Another procedure is the subgradient optimization which is briefly introduced in Exercises 6.22, 6.23, and 6.24. See Held, Wolfe, and Crowder [1974] and Polyak [1967] for validation of subgradient optimization. For other related work, see Bazaraa and Goode [1977], Fisher, Northrup, and Shapiro [1975], and Held and Karp [1970].

One of the pioneering works for using the Lagrangian formulation to develop computational schemes is credited to Everett [1963]. Under certain conditions, he showed how the primal solution could be retrieved. The result and its extensions are given in Section 6.5.