NONLINEAR PROGRAMMING

THEORY AND ALGORITHMS

•

SECOND EDITION

Mokhtar S. Bazaraa
Hanif D. Sherali
C. M. Shetty

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 and suitable constraint qualifications, 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. We also develop the saddle point optimality conditions along with necessary and sufficient conditions for the absence of a duality gap, and interpret this in terms of a suitable perturbation function.

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: Formulating and Solving the Dual Problem Several procedures for solving the dual problem are discussed. In particular, we briefly describe gradient and subgradient-based methods, and present a tangential approximation 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 Lagrangian dual formulations for linear and quadratic programming, relating them to other standard duality formulations.

6.1 The Lagrangian Dual Problem

Consider the following nonlinear programming Problem P, which we call the *primal problem*.

Primal Problem P

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

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. It has also 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

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

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 $(\mathbf{u}, \mathbf{v})$. The optimization problem that evaluates $\theta(\mathbf{u}, \mathbf{v})$ is sometimes referred to as the *Lagrangian dual subproblem*. In this problem, 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*. We remark here that, strictly speaking, we should write D as sup $\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}\}$, rather than max $\{\theta(\mathbf{u}, \mathbf{v}) : \mathbf{u} \geq \mathbf{0}\}$, since the maximum may not exist (see Example 6.2.8). However, we shall specifically identify such cases wherever necessary.

The primal and Lagrangian dual problems can 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 shall use this form throughout the remainder of this chapter.

Primal Problem P

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

Lagrangian Dual Problem D

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

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

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. This choice can affect both the optimal value of D (as in nonconvex situations) and the effort expended in evaluating and updating the dual function θ during the course of solving the dual problem. Hence, an appropriate selection of the set X must be made, depending on the structure of the problem and the purpose for solving D (see the Notes and References section).

Geometric Interpretation of the Dual Problem

We now briefly discuss the geometric interpretation of the dual problem. For the sake of simplicity, we shall 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 (y, z) plane, the set $\{(y, z) : y = g(\mathbf{x}), z = f(\mathbf{x}) \text{ for some } \mathbf{x} \in X\}$ 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 with $y \leq 0$ that has a minimum ordinate. Obviously, this point is $(\bar{y}, \bar{z})$ 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 $y = g(\mathbf{x})$ and $z = f(\mathbf{x})$ for $\mathbf{x} \in X$, we want to minimize $z + uy$ over points in G. Note that $z + uy = \alpha$ is an equation of a straight line with slope $-u$ and intercept α on the z axis. To minimize $z + uy$ over G, we need to move the line $z + uy = \alpha$ parallel to itself as far down (along its negative gradient) as possible while it remains in contact with G. In other words, we move this line parallel to itself until it supports G from below, that is, the set G is above the line and touches it. Then, the intercept on the z 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 axis is maximal. In Figure 6.1, such a hyperplane has slope $-\bar{u}$ and supports the set

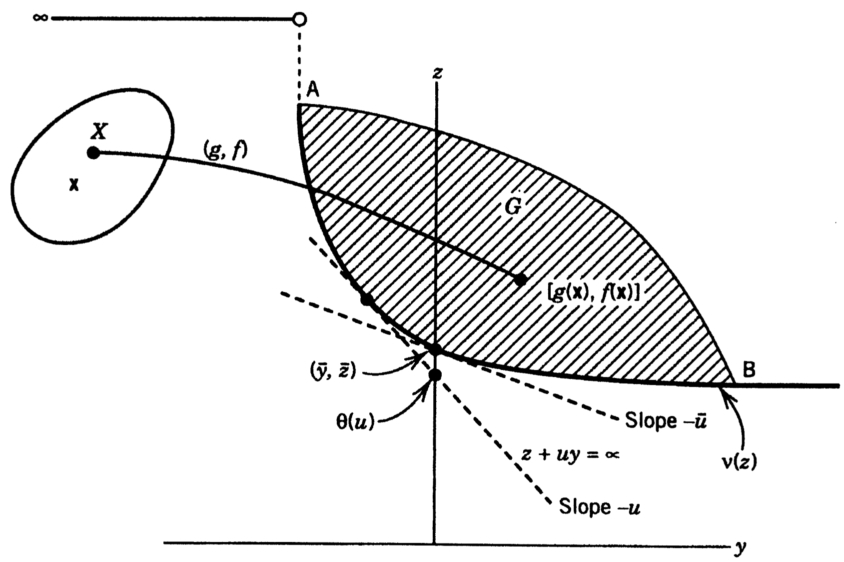

Figure 6.1 Geometric interpretation of Lagrangian duality.

G at the point $(\bar{y}, \bar{z})$. Thus, the optimal dual solution is $\bar{u}$, and the optimal dual objective value is $\bar{z}$. Furthermore, the optimal primal and dual objectives are equal in this case.

There is another related interesting interpretation that provides an important conceptual tool in this context. For the problem under consideration, define the function

$$\nu(y) = \text{minimum } \{f(\mathbf{x}) : g(\mathbf{x}) \leq y, \mathbf{x} \in X\}$$

The function ν is called a *perturbation function* since it is the optimal value function of a problem obtained from the original problem by perturbing the right-hand side of the inequality constraint $g(\mathbf{x}) \leq 0$ to y from the value of zero. Note that $\nu(y)$ is a nonincreasing function of y since, as y increases, the feasible region of the perturbed problem enlarges (or stays the same). For the present case, this function is illustrated in Figure 6.1. Observe that ν corresponds here to the lower envelope of G between points A and B because this envelope is itself monotone-decreasing. Moreover, ν remains constant at the value at point B for values of y higher than that at B, and becomes ∞ for points to the left of A because of infeasibility. In particular, if ν is differentiable at the origin, we observe that $\nu'(0) = -\bar{u}$. Hence, the marginal rate of change in objective function value with an increase in the right-hand side of the constraint from its present value of zero is given by $-\bar{u}$, the negative of the Lagrangian multiplier value at optimality. If ν is convex but is not differentiable at the origin, then $-\bar{u}$ is evidently a subgradient of ν at $y = 0$. In either case, we know that $\nu(y) \geq \nu(0) - \bar{u}y$ for all $y \in E_1$. As we shall see later, ν can be nondifferentiable and/or nonconvex, but the condition $\nu(y) \geq \nu(0) - \bar{u}y$ holds true for all $y \in E_1$, if and only if $\bar{u}$ is a KKT Lagrangian multiplier corresponding to an optimal solution $\bar{\mathbf{x}}$ such that it solves the dual problem with equal primal and dual objective values. As seen above, this happens to be the case in Figure 6.1.

6.1.1 Example

Consider the following primal problem:

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

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

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

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

Note that the above infima are achieved at $x_1 = x_2 = u/2$ if $u \geq 0$ and at $x_1 = x_2 = 0$ if $u < 0$. Hence,

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

Note that θ is a concave function, and its maximum over $u \geq 0$ occurs at $\bar{u} = 4$. Figure 6.2 illustrates the situation. Note also that the optimal primal and dual objective values are both equal to 8.

Now let us consider the problem in the (y, z) plane, where $y = g(x)$ and $z = f(x)$. We are interested in finding G, the image of $X = \{(x_1, x_2): x_1 \geq 0, x_2 \geq 0\}$, 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 y, note that $\alpha(y)$ and $\beta(y)$ 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	$-x_1 - x_2 + 4 = y$	subject to	$-x_1 - x_2 + 4 = y$
	$x_1, x_2 \geq 0$		$x_1, x_2 \geq 0$

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

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 value is $\alpha(0) = 8$ and is equal to the optimal primal objective value.

Again, in Figure 6.2, the perturbation function $\nu(y)$ for $y \in E_1$ corresponds to the lower envelope $\alpha(y)$ for $y \leq 4$, and $\nu(y)$ remains constant at the value 0 for $y \geq 4$. The slope $\nu'(0)$ equals -4, the negative of the optimal Lagrange multiplier value. Moreover, we have $\nu(y) \geq \nu(0) - 4y$ for all $y \in E_1$. As we shall see in the next section, this is a necessary and sufficient condition for the primal and dual objective values to match at optimality.

6.2 Duality Theorems and Saddle Point Optimality Conditions

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.

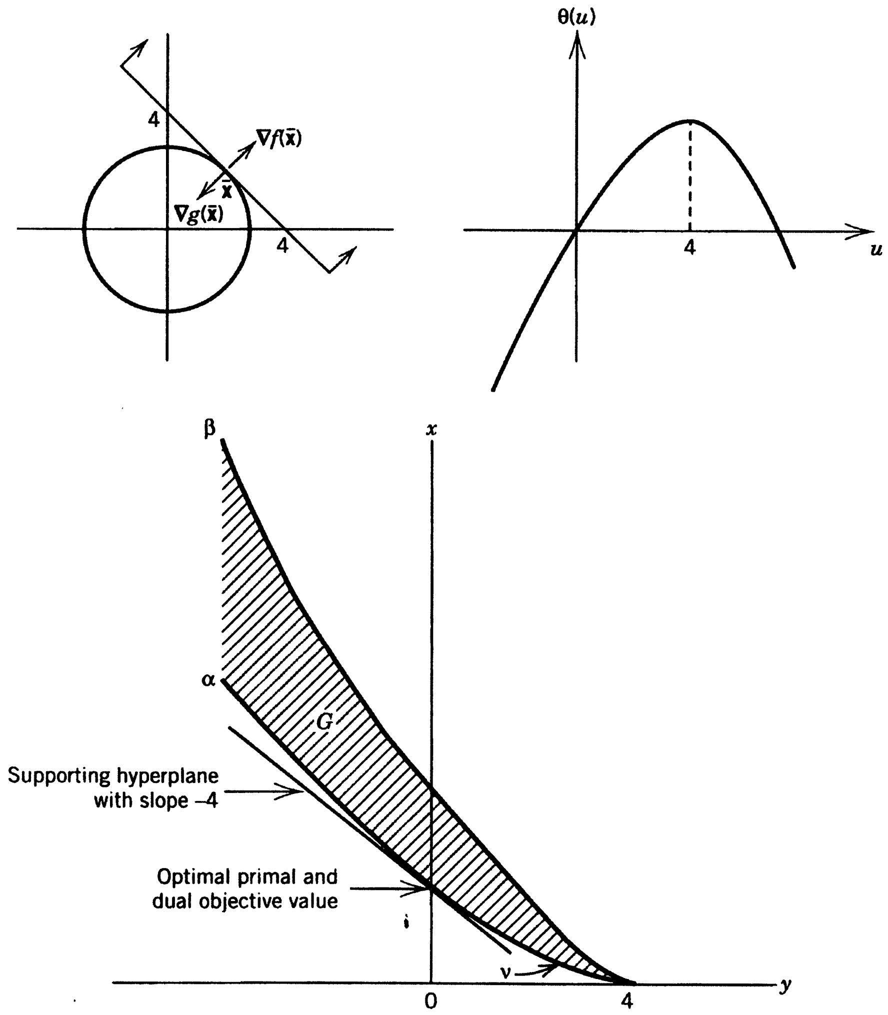

Figure 6.2 Geometric illustration of Example 6.1.1.

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 the 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 $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, 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. The perturbation function $\nu(y)$ for $y \in E_1$ is as shown in the figure. Note that, by definition, *this is the greatest monotone nonincreasing function that envelopes* G *from below* (see Exercise 6.5). The optimal primal value is $\nu(0)$. The greatest intercept on the ordinate z axis achieved by a hyperplane that supports G from below gives the optimal dual objective value as shown. In particular, observe that there does not exist a $\bar{u}$ such that $\nu(y) \geq \nu(0) - \bar{u}y$ for all $y \in E_1$, as we had in Figures 6.1 and 6.2. Exercise 6.6 asks the reader to construct G and ν for the instance illustrated in Figure 4.13 that results in a situation similar to that of Figure 6.3.

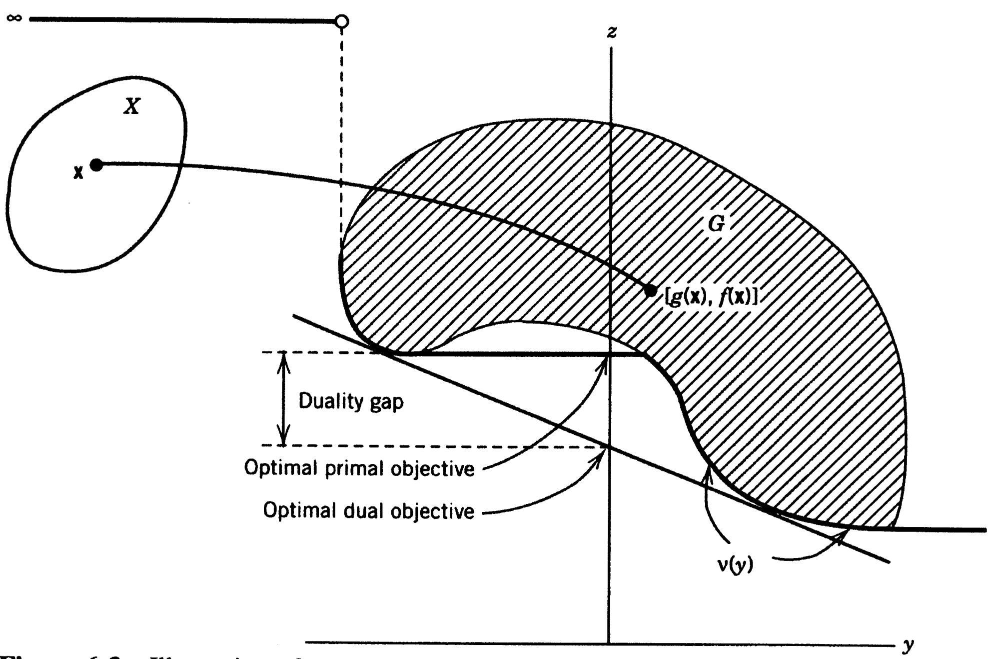

Figure 6.3 Illustration of a duality gap.

6.2.2 Example

Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & f(x) = -2x_1 + x_2 \\ \text{subject to} & h(x) = x_1 + x_2 - 3 = 0 \\ & (x_1, x_2) \in X \end{array}$$

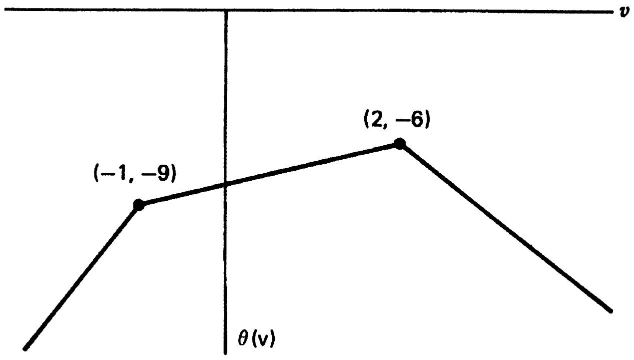

Figure 6.4 Dual function for Example 6.2.2.

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

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

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

$$\theta(v) = \begin{bmatrix} -4 + 5v & \text{for} & v \le -1 \\ -8 + v & \text{for} & -1 \le v \le 2 \\ -3v & \text{for} & v \ge 2 \end{bmatrix}$$

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 a 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

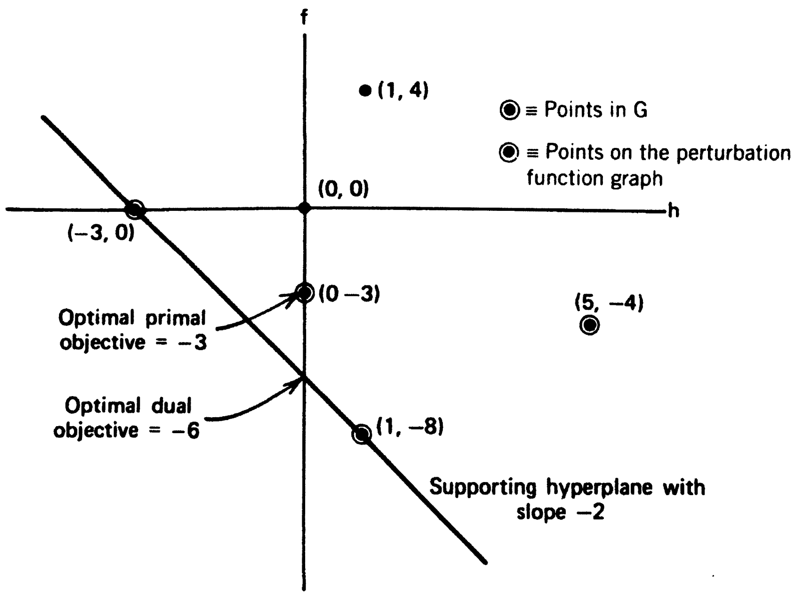

Figure 6.5 Geometric interpretation of Example 6.2.2.

correspond to the primal feasible points and, hence, the minimal primal objective value is equal to -3.

Similar to the inequality constrained case, the perturbation function here is defined as $\nu(y) = \text{minimum}\,\{f(\mathbf{x}) : h(\mathbf{x}) = y, \mathbf{x} \in X\}$. Because of the discrete nature of X, $h(\mathbf{x})$ can take on only a finite possible number of values. Hence, noting G in Figure 6.5, we obtain $\nu(-3) = 0$, $\nu(0) = -3$, $\nu(1) = -8$, and $\nu(5) = -4$, with $\nu(y) = \infty$ for all $y \in E_1$ otherwise. Again, the optimal primal value is $\nu(0) = -3$, and there does not exist a $\bar{\nu}$ such that $\nu(y) \geq \nu(0) - \bar{\nu}y$. Hence, a duality gap exists.

Conditions that guarantee the absence of a duality gap are given in Theorem 6.2.4. Then, Theorem 6.2.7 relates this to the perturbation function. First, however, the following lemma is needed.

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, $\mathbf{h}$ is of the form $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$. 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: $\alpha(\mathbf{x}) < 0 \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}$ for some $\mathbf{x} \in X$

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

$(u_0, \mathbf{u}) \geq \mathbf{0} \quad (u_0, \mathbf{u}, \mathbf{v}) \neq \mathbf{0}$

Proof

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

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

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, ((0, $\mathbf{0}$, $\mathbf{0}$) $\notin \Lambda$. By Corollary 1 to Theorem 2.4.7, there exists a nonzero $(u_0, \mathbf{u}, \mathbf{v})$ such that

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

Now fix an $\mathbf{x} \in X$. Since p and $\mathbf{q}$ can be made arbitrarily large, (6.1) holds true only if $u_0 \geq 0$ and $\mathbf{u} \geq \mathbf{0}$. 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

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

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 $\mathbf{u} \geq \mathbf{0}$, satisfying

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

Now let $\mathbf{x} \in X$ be such that $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. From the above inequality, since $\mathbf{u} \geq \mathbf{0}$, we conclude that $u_0\alpha(\mathbf{x}) \geq 0$. Since $u_0 > 0$, $\alpha(\mathbf{x}) \geq 0$; 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 $\gamma = \inf\{f(\mathbf{x}):\mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$. By assumption, $\gamma < \infty$. 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, therefore (6.2) holds true. Hence, 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 $\mathbf{g}(\bar{\mathbf{x}}) \le \mathbf{0}$, we get $\bar{\mathbf{u}}'\mathbf{g}(\bar{\mathbf{x}}) = 0$, 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 Slater's constraint qualification of Chapter 5. In particular, if $X = E_n$, then $\mathbf{0} \in \operatorname{int} \mathbf{h}(X)$ automatically holds true (if redundant equations are deleted), so that the constraint qualification asserts the existence 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 $\mathbf{h}(\mathbf{x}) = \mathbf{Ax} - \mathbf{b}$. 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 $\mathbf{y} = \mathbf{Ax} - \mathbf{b}$, where $\mathbf{x} = \mathbf{A}'(\mathbf{AA}')^{-1}(\mathbf{y} + \mathbf{b})$. Thus, $\mathbf{h}(X) = E_m$ and, in particular, $\mathbf{0} \in \operatorname{int} \mathbf{h}(X)$.

Saddle Point Criteria

The foregoing theorem shows that under convexity assumptions and under a suitable constraint qualification, the primal and dual objective function values match at optimality.