Nonlinear and Mixed-Integer Optimization

Fundamentals and Applications

CHRISTODOULOS A. FLOUDAS

New York Oxford
OXFORD UNIVERSITY PRESS
1995

Chapter 4 Duality Theory

Nonlinear optimization problems have two different representations, the primal problem and the dual problem. The relation between the primal and the dual problem is provided by an elegant duality theory. This chapter presents the basics of duality theory. Section 4.1 discusses the primal problem and the perturbation function. Section 4.2 presents the dual problem. Section 4.3 discusses the weak and strong duality theorems, while section 4.4 discusses the duality gap.

4.1 Primal Problem

This section presents the formulation of the primal problem, the definition and properties of the perturbation function, the definition of stable primal problem, and the existence conditions of optimal multiplier vectors.

4.1.1 Formulation

The primal problem (P) takes the form:

$$(P)\left\{\begin{array}{lll}\min & f(x) & \\ \text{s.t.} & h(x) & = \mathbf{0} \\ & g(x) & \leq \mathbf{0} \\ & x & \in X\end{array}\right.$$

where x is a vector of n variables;

$h(x)$ is a vector of m real valued functions;
$g(x)$ is a vector of p real valued functions;
$f(x)$ is a real valued function; and
X is a nonempty convex set.

4.1.2 Perturbation Function and Its Properties

The perturbation function, $v(y)$, that is associated with the primal problem (P) is defined as:

$$v(y) = \begin{cases} \inf & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) \leq y \\ & x \in X \end{cases}$$

and y is an p-dimensional perturbation vector.

Remark 1 For $y = 0$, $v(0)$ corresponds to the optimal value of the primal problem (P). Values of the perturbation function $v(y)$ at other points different than the origin $y = 0$ are useful on the grounds of providing information on sensitivity analysis or parametric effects of the perturbation vector y.

Property 4.1.1 (Convexity of $v(y)$)
Let Y be the set of all vectors y for which the perturbation function has a feasible solution, that is,

$$Y = \{y \in R^p \mid h(x) = 0,\ g(x) \leq y \text{ for some } x \in X\}.$$

If Y is a convex set, then $v(y)$ is a convex function on Y.

Remark 2 $v(y) = \infty$ if and only if $y \notin Y$.

Remark 3 Y is a convex set if $h(x)$ are linear and $g(x)$ are convex on the convex set X.

Remark 4 The convexity property of $v(y)$ is the fundamental element for the relationship between the primal and dual problems. A number of additional properties of the perturbation function $v(y)$ that follow easily from its convexity are

- **(i)** If Y is a finite set then $v(y)$ is continuous on Y.
- **(ii)** the directional derivative of $v(y)$ exists in every direction at every point at which $v(y)$ is finite.
- **(iii)** $v(y)$ has a subgradient at every interior point of Y at which $v(y)$ is finite.
- **(iv)** $v(y)$ is differentiable at a point $\bar{y}$ in Y if and only if it has a unique subgradient at $\bar{y}$.

Note also that these additional properties hold for any convex function defined on a convex set.

4.1.3 Stability of Primal Problem

Definition 4.1.1 The primal problem (P) is stable if $v(0)$ is finite and there exists a scalar $L > 0$ such that

$$\frac{v(0) - v(y)}{\|y\|} \leq L \text{ for all } y \neq 0.$$

Remark 1 The above definition of stability does not depend on the particular norm $||y||$ that is used. In fact, it is necessary and sufficient to consider a *one- dimensional* choice of y.

Remark 2 The property of stability can be interpreted as a Lipschitz continuity condition on the perturbation function $v(y)$.

Remark 3 If the stability condition does not hold, then $[v(0) - v(y)]$ can be made as large as desired even with small perturbation of the vector y.

4.1.4 Existence of Optimal Multipliers

Theorem 4.1.1
Let the primal problem (P) have an optimum solution x^*. Then, an optimal multiplier vector $(\bar{\lambda}, \bar{\mu})$ exists if and only if (P) is stable. Furthermore, $(\bar{\lambda}, \bar{\mu})$ is an optimal multiplier vector for (P) if and only if $(-\bar{\lambda}, -\bar{\mu})$ is a subgradient of $v(y)$ at $y = 0$.

Remark 1 This theorem points out that stability is not only a necessary but also sufficient constraint qualification, and hence it is implied by any constraint qualification used to prove the necessary optimality conditions.

Remark 2 If the objective function $f(x)$ is sufficiently well behaved, then the primal problem (P) will be stable regardless of how poorly behaved the constraints are. For instance, if $f(x) = c$, then (P) is stable as long as the constraints are feasible.

Remark 3 If the constraints are linear, it is still possible for (P) to be unstable. Consider the problem:

Illustration 4.1.1 Consider the problem:

$$\begin{cases} \min f(x) &= -x^{\frac{1}{3}} \\ \text{s.t. } g(x) &= x \leq 0 \\ x \in X &= \{x | x \geq 0\} \end{cases}$$

Note that as x approaches zero, the objective function has infinite steepness.

By perturbing the right-hand side in the positive direction, we obtain

$$\frac{0 - (-y^{\frac{1}{3}})}{|y|} = \frac{1}{y^{\frac{2}{3}}},$$

which can be made as large as desired by making the perturbation y sufficiently small.

4.2 Dual Problem

This section presents the formulation of the dual problem, the definition and key properties of the dual function, and a geometrical interpretation of the dual problem.

4.2.1 Formulation

The dual problem of (P), denoted as (D), takes the following form:

$$(D)\begin{cases} \max\limits_{\substack{\lambda \\ \mu \geq 0}} \quad \inf\limits_{x \in X} \quad L(x,\ \lambda,\ \mu) \\ \text{s.t. } L(x,\ \lambda,\ \mu) = f(x) + \lambda^T \mathbf{h}(x) + \mu^T \mathbf{g}(x) \end{cases}$$

where $L(x,\ \lambda,\ \mu)$ is the Lagrange function, λ is the m-vector of Lagrange multipliers associated with the equality constraints, and μ is the p-vector of Lagrange multipliers associated with the inequality constraints.

Remark 1 Note that the inner problem of the dual

$$\inf_{x \in X} \quad L(x,\ \lambda,\ \mu)$$

is a function of λ and μ (i.e., it is parametric in λ and μ). Hence it may take the value of $(-\infty)$ for some λ and μ. If the infimum is attained and is finite for all (λ, μ) then the inner problem of the dual can be written as

$$\min_{x \in X} \quad L(x,\ \lambda,\ \mu)$$

Remark 2 The dual problem consists of (i) an inner minimization problem of the Lagrange function with respect to $x \in X$ and (ii) on outer maximization problem with respect to the vectors of the Lagrange multipliers (unrestricted λ, and $\mu \geq 0$). The inner problem is parametric in λ and μ. For fixed x at the infimum value, the outer problem becomes linear in λ and μ.

4.2.2 Dual Function and Its Properties

Definition 4.2.1 (Dual function) The dual function, denoted as $\phi(\lambda,\ \mu)$, is defined as the inner problem of the dual:

$$\phi(\lambda,\ \mu) = \inf_{x \in X} \quad L(x,\ \lambda,\ \mu)$$

Property 4.2.1 (Concavity of $\phi(\lambda,\ \mu)$)
Let $f(x)$, $h(x)$, $g(x)$ be continuous and X be a nonempty compact set in R^n. Then the dual function $\phi(\lambda,\ \mu)$ is concave.

Remark 1 Note that the dual function $\phi(\lambda, \mu)$ is concave without assuming any type of convexity for the objective function $f(x)$ and constraints $h(x), g(x)$.

Remark 2 Since $\phi(\lambda,\ \mu)$ is concave, and the outer problem is a maximization problem over λ and $\mu \geq 0$, then a local optimum of $\phi(\lambda,\ \mu)$ is also a global one. The difficulty that arises though is that we only have $\phi(\lambda,\ \mu)$ as a parametric function of λ and μ and not as an explicit functionality of λ and μ.

Remark 3 The dual function $\phi(\lambda,\ \mu)$ is concave since it is the pointwise infimum of a collection of functions that are linear in λ and μ.

Remark 4 Since the dual function of $\phi(\lambda,\ \mu)$ is concave, it has a subgradient at $(\bar{\lambda},\ \bar{\mu})$ that is defined as the vector d_1, d_2:

$$\phi(\lambda,\ \mu) \leq \phi(\bar{\lambda},\ \bar{\mu}) + d_1^T(\lambda - \bar{\lambda}) + d_2^T(\mu - \bar{\mu})$$

for all λ, and $\mu \geq \boldsymbol{0}$.

Property 4.2.2 (Subgradient of dual function)
Let $f(x)$, $h(x)$, $g(x)$ be continuous, and X be a nonempty, compact set in R^n.
Let

$$Y(\lambda,\ \mu) = \{x^* : x^* \text{ minimizes } L(x,\ \lambda,\ \mu) \text{ over } x \in X\}$$

If for any $(\bar{\lambda},\ \bar{\mu})$, $Y(\bar{\lambda},\ \bar{\mu})$ is nonempty, and $x^* \in Y(\bar{\lambda},\ \bar{\mu})$, then

$$(h(x^*), g(x^*)) \ \text{ is a subgradient of } \phi(\lambda, \mu) \text{ at } (\bar{\lambda}, \bar{\mu}).$$

Remark 5 This property provides a sufficient condition for a subgradient.

Property 4.2.3 (Differentiability of dual function)
Let $f(x)$, $h(x)$, $g(x)$ be continuous, and X be a nonempty compact set. If $Y(\bar{\lambda},\ \bar{\mu})$ reduces to a single element at the point $(\bar{\lambda},\ \bar{\mu})$, then the dual function $\phi(\lambda, \mu)$ is differentiable at $(\bar{\lambda},\ \bar{\mu})$ and its gradient is

$$\nabla\phi(\bar{\lambda}, \bar{\mu}) = (h(x^*), g(x^*)) \,.$$

4.2.3 Illustration of Primal-Dual Problems

Consider the following constrained problem:

$$\left\{\begin{array}{lrl} \min & (x_1-1)^2 + (x_2-1)^2 + (x_3-1)^2 & \\ \text{s.t.} & 2x_1 + 4x_2 & = 10 \\ & 2x_1 + 2x_2 - 4x_3 & \leq 0 \\ & x_1, x_2, x_3 & \geq 0 \end{array}\right.$$

where

$$\begin{aligned} f(x) &= (x_1-1)^2 + (x_2-1)^2 + (x_3-1)^2, \\ h(x) &= 2x_1 + 4x_2 - 10, \\ g(x) &= 2x_1 + 2x_2 - 4x_3, \text{ and} \\ X &= \{x_1, x_2, x_3 | x_1 \geq 0, x_2 \geq 0, x_3 \geq 0\}\,. \end{aligned}$$

The Lagrange function takes the form:

$$\begin{aligned} L(x_1, x_2, x_3, \lambda, \mu) &= (x_1-1)^2 + (x_2-1)^2 + (x_3-1)^2 \\ &\quad +\lambda\,(2x_1 + 4x_2 - 10) \\ &\quad +\mu\,(2x_1 + 2x_2 - 4x_3)\,. \end{aligned}$$

The minimum of the $L(x_1, x_2, x_3, \lambda, \mu)$ is given by

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= 2(x_1 - 1) + 2\lambda + 2\mu &&= 0 \Rightarrow x_1 = 1 - \lambda - \mu, \\
\frac{\partial L}{\partial x_2} &= 2(x_2 - 1) + 4\lambda + 2\mu &&= 0 \Rightarrow x_2 = 1 - 2\lambda - \mu, \\
\frac{\partial L}{\partial x_3} &= 2(x_3 - 1) - 4\mu &&= 0 \Rightarrow x_3 = 1 + 2\mu.
\end{aligned}$$

The dual function $\phi(\lambda, \mu)$ then becomes

$$\begin{aligned}
\phi(\lambda, \mu) = & \; (-\lambda - \mu)^2 + (-2\lambda - \mu)^2 + (2\mu)^2 \\
& + \lambda[2(1 - \lambda - \mu) + 4(1 - 2\lambda - \mu) - 10] \\
& + \mu[2(1 - \lambda - \mu) + 2(1 - 2\lambda - \mu) - 4(1 + 2\mu)] \\
= & \; -5\lambda^2 - 6\mu^2 - 6\lambda\mu - 4\lambda.
\end{aligned}$$

This dual function is concave in λ and μ. (The reader can verify it by calculating the eigenvalues of the Hessian of this quadratic function in λ and μ, which are both negative.)

The maximum of the dual function $\phi(\lambda, \mu)$ can be obtained when

$$\begin{aligned}
\frac{\partial \phi(\lambda,\mu)}{\partial \lambda} &= -10\lambda - 6\mu - 4 = 0, \\
\frac{\partial \phi(\lambda,\mu)}{\partial \mu} &= -12\mu - 6\lambda = 0,
\end{aligned}$$

which results in

$$\begin{aligned}
\lambda^* &= -\tfrac{4}{7}, \\
\mu^* &= \tfrac{2}{7}, \\
\text{and } \phi(\lambda^*, \mu^*) &= (8/7).
\end{aligned}$$

$$\text{Also, } \left\{ \begin{array}{l} x_1^* = (9/7) \\ x_2^* = (13/7) \\ x_3^* = (11/7) \end{array} \right\} \text{ and } f(x_1^*, x_2^*, x_3^*) = (8/7).$$

4.2.4 Geometrical Interpretation of Dual Problem

The geometrical interpretation of the dual problem provides important insight with respect to the dual function, perturbation function, and their properties. For illustration purposes, we will consider the primal problem (P) consisting of an objective function $f(x)$ subject to constraints $g_1(x) \le 0$ and $g_2(x) \le 0$ in a single variable x.

The geometrical portrayal is based upon the image of X under $f(x), g_1(x)$, and $g_2(x)$ that is represented by the image set I:

$$\begin{aligned}
I = \{(z_1, z_2, z_3) \in R^3 : z_1 \ge g_1(x), z_2 \ge g_2(x) \text{ and} \\
z_3 = f(x) \text{ for some } x \in X\},
\end{aligned}$$

which is shown in Figure 4.1.

Geometrical interpretation of primal problem (P):

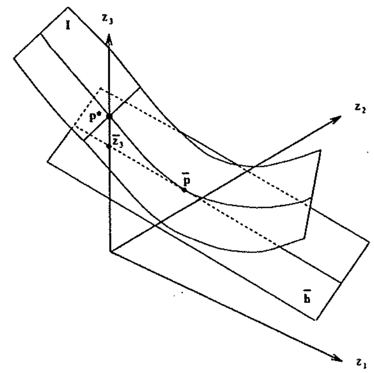

Figure 4.1: Image set I

Note that the intersection point of I and z_3, denoted as (P^*) in the figure, is the image of the optimal solution of the primal problem (P),

$$P^* = [g_1(x^*), g_2(x^*), f(x^*)],$$

where x^* is the minimizer of the primal problem (P). Hence the primal problem (P) can be explained as follows: Determine the point in the image set I which minimizes z_3 subject to $z_1 \leq 0$ and $z_2 \leq 0$.

For the geometrical interpretation of the dual problem, we will consider particular values for the Lagrange multipliers μ_1, μ_2 associated with the two inequality constraints $(\mu_1 \geq 0, \mu_2 \geq 0)$, denoted as $\bar{\mu_1}, \bar{\mu_2}$.

To evaluate the dual function at $\bar{\mu_1}, \bar{\mu_2}$ (i.e., the maximum of (D) at $\bar{\mu_1}, \bar{\mu_2}$), we have

$$\begin{aligned} &\min \{f(x) + \bar{\mu_1} g_1(x) + \bar{\mu_2} g_2(x)\} \\ &s.t.\ x \in X. \end{aligned}$$

This is equivalent to

$$\begin{aligned} \min \quad & z_3 + \bar{\mu_1} z_1 + \bar{\mu_2} z_2 \\ s.t. \quad & (z_1, z_2, z_3) \in I \end{aligned}$$

Note that the objective function

$$z_3 + \bar{\mu_1} z_1 + \bar{\mu_2} z_2$$

is a plane in (z_1, z_2, z_3) with slope $(-\bar{\mu_1}, -\bar{\mu_2})$ as illustrated in Figure 4.1.

Geometrical Interpretation of Dual Function at $(\bar{\mu_1}, \bar{\mu_2})$:

The dual function at $(\bar{\mu}_1, \bar{\mu}_2)$ corresponds to determining the lowest plane with slope $(-\bar{\mu}_1, -\bar{\mu}_2)$ which intersects the image set I. This corresponds to the supporting hyperplane $\bar{h}$ which is tangent to the image set I at the point $\bar{P}$, as shown in Figure 4.1.

Note that the point $\bar{P}$ is the image of

$$\begin{array}{ll} \min & f(x) + \mu_1 g_1(x) + \mu_2 g_2(x) \\ \text{s.t.} & x \in X \end{array}$$

The minimum value of this problem is the value of z_3 where the supporting hyperplane $\bar{h}$ intersects the ordinate, denoted as $\bar{z}_3$ in Figure 4.1.

Geometrical Interpretation of Dual Problem:

Determine the value of $(\bar{\mu_1}, \bar{\mu_2})$, which defines the slope of a supporting hyperplane to the image set I, such that it intersects the ordinate at the highest possible value. In other words, identify $(\bar{\mu_1}, \bar{\mu_2})$ so as to maximize $\bar{z}_3$.

Remark 1 The value of $(\bar{\mu_1}, \bar{\mu_2})$ that intersects the ordinate at the maximum possible value in Figure 4.1 is the supporting hyperplane of I that goes through the point P^*, which is the optimal solution to the primal problem (P).

Remark 2 It is not always possible to obtain the optimal value of the dual problem being equal to the optimal value of the primal problem. This is due to the form that the image set I can take for different classes of mathematical problems (i.e., form of objective function and constraints). This serves as a motivation for the weak and strong duality theorems to be presented in the following section.

4.3 Weak and Strong Duality

In the previous section we have discussed geometrically the nature of the primal and dual problems. In this section, we will present the weak and strong duality theorems that provide the relationship between the primal and dual problem.

Theorem 4.3.1 (Weak duality)
Let $\bar{x}$ be a feasible solution to the primal problem (P), and $(\bar{\lambda}, \bar{\mu})$ be a feasible solution to the dual problem (D). Then, the objective function of (P) evaluated at $\bar{x}$ is greater or equal to the objective function of (D) evaluated at $(\bar{\lambda}, \bar{\mu})$; that is,

$$f(\bar{x}) \geq \phi(\bar{\lambda}, \bar{\mu}).$$

Remark 1 Any feasible solution of the dual problem (D) represents a lower bound on the optimal value of the primal problem (P).

Remark 2 Any feasible solution of the primal problem (P) represents an upper bound on the optimal value of the dual problem (D).

Remark 3 This lower-upper bound feature between the dual and primal problems is very important in establishing termination criteria in computational algorithms. In particular applications , if at some iteration feasible solutions exist for both the primal and the dual problems and are close to each other in value, then they can be considered as being *practically* optimal for the problem under consideration.

Remark 4 This important lower-upper bound result for the dual-primal problems that is provided by the weak duality theorem, is *not* based on any convexity assumption. Hence, it is of great use for nonconvex optimization problems as long as the dual problem can be solved efficiently.

Remark 5 If the optimal value of the primal problem (P) is $-\infty$, then the dual problem must be infeasible (i.e., essentially infeasible).

Remark 6 If the optimal value of the dual problem (D) is $+\infty$, then the primal problem (P) must be infeasible.

The weak duality theorem provides the lower-upper bound relationship between the dual and the primal problem. The conditions needed so as to attain equality between the dual and primal solutions are provided by the following strong duality theorem.

Theorem 4.3.2 (Strong duality)
Let $f(x), g(x)$ be convex, $h(x)$ be affine, and X be a nonempty convex set in R^n. If the primal problem (P) is stable, then

(i) *The dual problem (D) has an optimal solution.*

(ii) *The optimal values of the primal problem (P) and dual problem (D) are equal.*

(iii) *$(\bar{\lambda}, \bar{\mu})$ are an optimal solution of the dual problem if and only if $(-\bar{\lambda}, -\bar{\mu})$ is a subgradient of the perturbation function $v(y)$ at $y = 0$.*

(iv) *Every optimal solution $(\bar{\lambda}, \bar{\mu})$ of the dual problem (D) characterizes the set of all optimal solutions (if any) of the primal problem (P) as the minimizers of the Lagrange function:*

$$L(x, \bar{\lambda}, \bar{\mu}) = f(x) + \bar{\lambda}^T h(x) + \bar{\mu}^T g(x)$$

over $x \in X$ which also satisfy the feasibility conditions

$$h(x) = 0$$
$$g(x) \leq 0$$

and the complementary slackness condition

$$\bar{\mu}^T g(x) = 0.$$

Remark 1 Result (ii) precludes the existence of a gap between the primal problem and dual problem values which is denoted as duality gap. It is important to note that nonexistence of duality gap is guaranteed under the assumptions of convexity of $f(x), g(x)$, affinity of $h(x)$, and stability of the primal problem (P).

Remark 2 Result (iii) provides the relationship between the perturbation function $v(y)$ and the set of optimal solutions $(\bar{\lambda}, \bar{\mu})$ of the dual problem (D).

Remark 3 If the primal problem (P) has an optimal solution and it is stable, then using the theorem of existence of optimal multipliers (see section 4.1.4), we have an alternative interpretation of the optimal solution $(\bar{\lambda}, \bar{\mu})$ of the dual problem (D): that $(\bar{\lambda}, \bar{\mu})$ are the optimal Lagrange multipliers of the primal problem (P).

Remark 4 Result (iii) holds also under a weaker assumption than stability; that is, if $v(0)$ is finite and the optimal values of (P) and (D) are equal.

Remark 5 Result (iv) can also be stated as
If $(\bar{\lambda}, \bar{\mu})$ are optimal in the dual problem (D), then x is optimal in the primal problem (P) if and only if $(x, \bar{\lambda}, \bar{\mu})$ satisfies the optimality conditions of (P).

Remark 6 The geometrical interpretation of the primal and dual problems clarifies the weak and strong duality theorems. More specifically, in the vicinity of $y = 0$, the perturbation function $v(y)$ becomes the z_3-ordinate of the image set I when z_1 and z_2 equal y. In Figure 4.1, this ordinate does not decrease infinitely steeply as y deviates from zero. The slope of the supporting hyperplane to the image set I at the point P^*, $(-\bar{\mu}_1, -\bar{\mu}_2)$, corresponds to the subgradient of the perturbation function $v(y)$ at $y = 0$.

Remark 7 An instance of unstable problem (P) is shown in Figure 4.2 The image set I is tangent to the ordinate z_3 at the point P^*. In this case, the supporting hyperplane is perpendicular, and the value of the perturbation function $v(y)$ decreases infinitely steeply as y begins to increase above zero. Hence, there does not exist a subgradient at $y = 0$. In this case, the strong duality theorem does not hold, while the weak duality theorem holds.

4.3.1 Illustration of Strong Duality

Consider the problem

$$\left\{\begin{aligned} \min f(x) = & \quad (x_1 - 1)^2 + (x_2 - 1)^2 + (x_3 - 1)^2 \\ \text{s.t. } h(x) = & \quad 2x_1 + 4x_2 - 10 = 0 \\ g(x) = & \quad 2x_1 + 2x_2 - 4x_3 \leq 0 \\ & \quad x_1, x_2, x_3 \geq 0 \end{aligned}\right.$$

that was used to illustrate the primal-dual problems in section 4.2.3.

The objective function $f(x)$ and the inequality constraint $g(x)$ are convex since $f(x)$ is separable quadratic (sum of quadratic terms, each of which is a linear function of x_1, x_2, x_3, respectively) and $g(x)$ is linear. The equality constraint $h(x)$ is linear. The primal problem is also stable since $v(0)$ is finite and the additional stability condition (Lipschitz continuity-like) is satisfied since $f(x)$ is well behaved and the constraints are linear. Hence, the conditions of the strong duality theorem are satisfied. This is why

$$\phi(\lambda^*, \mu^*) = f(x_1^*, x_2^*, x_3^*) = (8/7),$$

as was calculated in section 4.2.3.

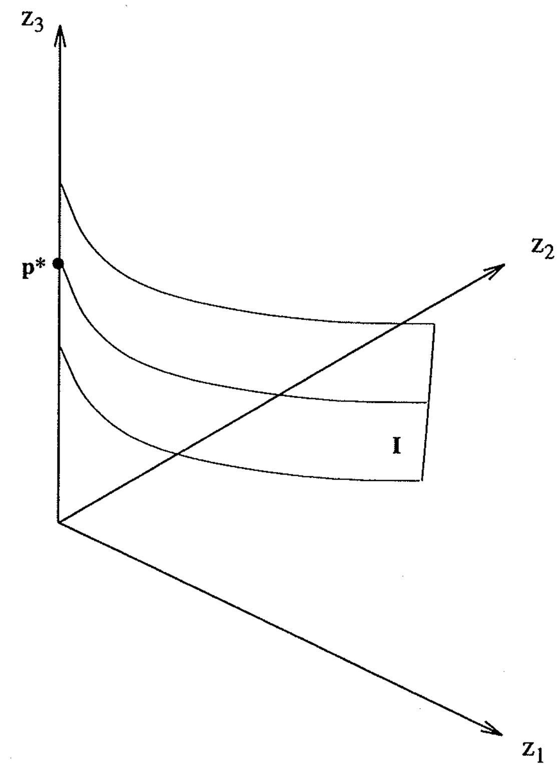

Figure 4.2: Unstable problem (P)

Also notice that since $\phi(\lambda, \mu)$ is differentiable at $(\lambda^*, \mu^*) = \left(-\frac{4}{7}, \frac{2}{7}\right)$ and its gradient is

$$\nabla\phi(\lambda^*, \mu^*) = [h(x^*), g(x^*)].$$

4.3.2 Illustration of Weak and Strong Duality

Consider the following bilinearly (quadratically) constrained problem

$$\begin{cases} \min & -x_1 - x_2 \\ \text{s.t.} & x_1 x_2 \leq 4 \\ & 0 \leq x_1 \leq 4 \\ & 0 \leq x_2 \leq 8 \end{cases}$$

where

$$\begin{aligned} f(x) &= -x_1 - x_2, \\ g(x) &= x_1 x_2 - 4, \text{ and} \\ X &= \{(x_1, x_2) : 0 \leq x_1 \leq 4, 0 \leq x_2 \leq 8\}. \end{aligned}$$

This problem is nonconvex due to the bilinear constraint and its global solution is $(x_1^*, x_2^*) = (0.5, 8)$, $f(x_1^*, x_2^*) = -8.5$.

The Lagrange function takes the form (by dualizing only the bilinear constraint):

$$L(x_1, x_2, \mu) = -x_1 - x_2 + \mu(x_1 x_2 - 4),$$

The minimum of the Lagrange function with respect to x_1, x_2 is given by

$$\frac{\partial L}{\partial x_1} = -1 + \mu x_2 = 0 \Rightarrow x_2 = \frac{1}{\mu},$$
$$\frac{\partial L}{\partial x_2} = -1 + \mu x_1 = 0 \Rightarrow x_1 = \frac{1}{\mu}.$$

The dual function then becomes

$$\begin{aligned}\phi(\mu) &= -\tfrac{1}{\mu} - \tfrac{1}{\mu} + \mu\left(\tfrac{1}{\mu} \cdot \tfrac{1}{\mu} - 4\right) \\ &= -\tfrac{1}{\mu} - 4\mu.\end{aligned}$$

The maximum of the dual function can be obtained when

$$\frac{\partial \phi}{\partial \mu} = \frac{1}{\mu^2} - 4 = 0 \Rightarrow \bar{\mu} = 0.5.$$

Then, $\phi(\bar{\mu}) = -4$

$$\begin{aligned}\bar{x}_1 &= \tfrac{1}{\mu} = 2, \\ \bar{x}_2 &= \tfrac{1}{\mu} = 2, \\ &\Downarrow \\ f(\bar{x}_1, \bar{x}_2) = -4 &\geq \phi(\bar{\mu}) = -4.\end{aligned}$$

4.3.3 Illustration of Weak Duality

Consider the following constrained problem:

$$\begin{cases} \min & x_1^{0.5} + x_2^{0.5} - 6x_1 \\ s.t. & x_2 - 3x_1 = 0 \\ & x_2 \leq 4 \\ & x_1, x_2 \geq 0 \end{cases}$$

where

$$\begin{aligned} f(x) &= x_1^{0.5} + x_2^{0.5} - 6x_1, \\ h(x) &= x_2 - 3x_1, \\ g(x) &= x_2 - 4, \\ X &= \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0\}. \end{aligned}$$

The global optimum solution is obtained at $(x_1, x_2) = (4/3, 4)$ and the objective function value is $(\frac{2}{\sqrt{3}} - 6)$.

The Lagrange function takes the form

$$L(x_1, x_2, \lambda, \mu) = x_1^{0.5} + x_2^{0.5} - 6x_1 + \lambda(x_2 - 3x_1) + \mu(x_2 - 4).$$

The minimum of the $L(x_1, x_2, \lambda, \mu)$ over $x \in X$ is given by

$$\frac{\partial L}{\partial x_1} = 0.5x_1^{-0.5} - 6 - 3\lambda = 0 \Rightarrow x_1^{0.5} = \frac{1}{6(\lambda+2)},$$
$$\frac{\partial L}{\partial x_2} = 0.5x_2^{-0.5} + \lambda + \mu = 0 \Rightarrow x_2^{0.5} = -\frac{1}{2(\lambda+\mu)}.$$

Since $x_1, x_2 \geq 0$, we must have $\lambda + 2 \geq 0$ and $\lambda + \mu \leq 0$.

Since $L(x_1, x_2, \lambda, \mu)$ is strictly concave in x_1, x_2, it achieves its minimum at boundary. Now,

$$L(x_1, x_2, \lambda, \mu) = \sqrt{x_1} - (3\lambda + 6)x_1 + \sqrt{x_2} + (\lambda + \mu)\, x_2 - 4\mu.$$

Clearly if $3\lambda + 6 > 0$ or $\lambda + \mu < 0$, $\phi(\lambda, \mu) = -\infty$ and when $3\lambda + 6 \leq 0$, and $\lambda + \mu \geq 0$, the minimum of L is achieved at $x_1 = x_2 = 0$, with $\phi(\lambda, \mu) = -4\mu$. Notice that $\lambda = -2 \Rightarrow \mu \geq -\lambda \geq 2$:

$$\phi(\lambda, \mu) \leq -8,$$
$$\max_{\lambda, \mu \geq 0} \phi(\lambda, \mu) = -8 < \frac{2}{\sqrt{3}} - 6.$$

4.4 Duality Gap and Continuity of Perturbation Function

Definition 4.4.1 (Duality gap) If the weak duality inequality is strict:

$$f(\bar{x}) > \phi(\bar{\lambda}, \bar{\mu}),$$

then, the difference $f(\bar{x}) - \phi(\bar{\lambda}, \bar{\mu})$ is called a duality gap.

Remark 1 The difference in the optimal values of the primal and dual problems can be due to a lack of continuity of the perturbation function $v(y)$ at $y = 0$. This lack of continuity does not allow the existence of supporting hyperplanes described in the geometrical interpretation section.

Remark 2 The perturbation function $v(y)$ is a convex function if Y is a convex set (see section 4.1.2). A convex function can be discontinuous at points on the boundary of its domain. For $v(y)$, the boundary corresponds to $y = 0$. The conditions that provide the relationship between gap and continuity of $v(y)$ are presented in the following theorem.

Theorem 4.4.1 (Continuity of perturbation function)
Let the perturbation function $v(y)$ be finite at $y = 0$, that is, $v(0)$ is finite. The optimal values of the primal and dual problems are equal (i.e., there is no duality gap) if and only if $v(y)$ is lower semicontinuous at $y = 0$.

Remark 3 Conditions for $v(y)$ to be lower semicontinuous at $y = 0$ are

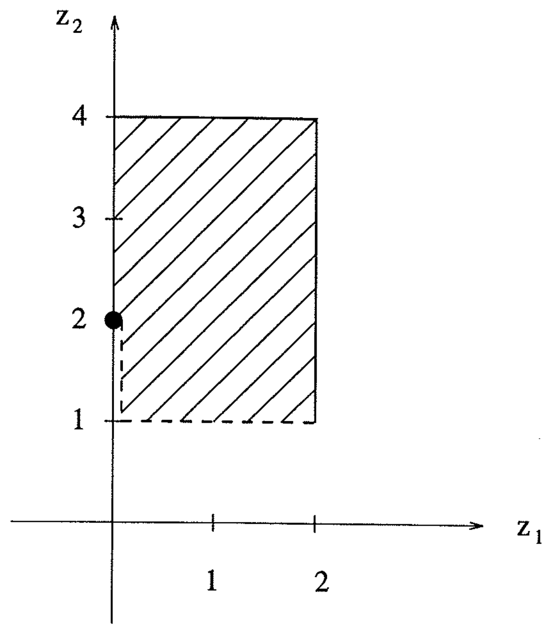

Figure 4.3: Illustration of duality gap

1. X is closed, and
2. $f(x), g(x)$ are continuous on X, and
3. $f(x^*)$ is finite, and
4. $\{x \in X : h(x) = 0, g(x) \leq 0, \text{ and } f(x) \leq \alpha\}$ is a bounded and nonempty convex set for some scalar $\alpha \geq v(0)$.

4.4.1 Illustration of Duality Gap

Consider the following problem

$$\left\{\begin{array}{c} \min\limits_{(x_1, x_2) \in X} \quad x_2 \\ s.t. \quad x_1 \leq 0 \\ X = \{(x_1, x_2) : 0 \leq x_1 \leq 2, 1 < x_2 \leq 4, \text{ and } \\ x_2 \geq 2 \text{ if } x_1 = 0\} \end{array}\right.$$

where $z_1 = x_1$ and $z_2 = x_2$, and the plot of z_1 vs z_2 is shown in Figure 4.3.

The optimal solution of the primal problem is attained at $x_2 = 2, x_1 = 0$.

Notice though that the optimal value of the dual problem cannot equal that of the primal due to the loss of lower semicontinuity of the perturbation function $v(y)$ at $y = 0$.

Summary and Further Reading

This chapter introduces the fundamentals of duality theory. Section 4.1 presents the formulation of the primal problem, defines the perturbation function associated with the primal problem and discuss its properties, and establishes the relationship between the existence of optimal multipliers and the stability of the primal problem. Section 4.2 presents the formulation of the dual problem and introduces the dual function and its associated properties along with its geometrical interpretation. Section 4.3 presents the weak and strong duality theorems, while section 4.4 defines the duality gap and establishes the connection between the continuity of the perturbation function and the existence of the duality gap. Further reading in these subjects can be found in Avriel (1976), Bazaraa *et al.* (1993), Geoffrion (1971), Geoffrion (1972b), Minoux (1986), and Walk (1989).