

Duality Theory of Constrained Optimization

Robert M. Freund

March, 2004

2004 Massachusetts Institute of Technology.

1 Overview

- The Practical Importance of Duality
- Definition of the Dual Problem
- Steps in the Construction of the Dual Problem
- Examples of Dual Constructions
- The Column Geometry of the Primal and Dual Problems
- The Dual is a Concave Maximization Problem
- Weak Duality
- Saddlepoint Optimality Criteria
- Strong Duality for Convex Problems
- Duality Strategies
- Illustration of Lagrange Duality in Discrete Optimization
- Conic Duality

2 The Practical Importance of Duality

Duality arises in nonlinear (and linear) optimization models in a wide variety of settings. Some immediate examples of duality are in:

- **Models of electrical networks.** The current flows are “primal variables” and the voltage differences are the “dual variables” that arise in consideration of optimization (and equilibrium) in electrical networks.
- **Models of economic markets.** In these models, the “primal” variables are production levels and consumption levels, and the “dual” variables are prices of goods and services.

- **Structural design.** In these models, the tensions on the beams are “primal” variables, and the nodal displacements are the “dual” variables.

Nonlinear (and linear) duality is very useful. For example, dual problems and their solutions are used in connection with:

- **Identifying near-optimal solutions.** A good dual solution can be used to bound the values of primal solutions, and so can be used to actually identify when a primal solution is near-optimal.
- **Proving optimality.** Using a strong duality theorem, one can prove optimality of a primal solution by constructing a dual solution with the same objective function value.
- **Sensitivity analysis of the primal problem.** The dual variable on a constraint represents the incremental change in the optimal solution value per unit increase in the RHS of the constraint.
- **Karush-Kuhn-Tucker (KKT) conditions.** The optimal solution to the dual problem is a vector of KKT multipliers.
- **Convergence of improvement algorithms.** The dual problem is often used in the convergence analysis of algorithms.
- **Good Structure.** Quite often, the dual problem has some good mathematical, geometric, or computational structure that can be exploited in computing solutions to both the primal and the dual problem.
- **Other uses, too**

3 The Dual Problem

3.1 Warning: Conventions involving $\pm\infty$

Suppose that we have a problem:

$$\begin{aligned} \text{(P)} : \quad z^* &= \text{maximum}_x \quad f(x) \\ &\text{s.t.} \quad x \in S . \end{aligned}$$

We could define the function:

$$h(x) := \begin{cases} f(x) & \text{if } x \in S \\ -\infty & \text{if } x \notin S . \end{cases}$$

Then we can rewrite our problem as:

$$\begin{aligned} \text{(P)} : \quad z^* &= \text{maximum}_x \quad h(x) \\ &\text{s.t.} \quad x \in \mathbb{R}^n . \end{aligned}$$

Conversely, suppose that we have a function $k(\cdot)$ that takes on the value $-\infty$ outside of a certain region S , but that $k(\cdot)$ is finite for all $x \in S$. Then the problem:

$$\begin{aligned} \text{(P)} : \quad z^* &= \text{maximum}_x \quad k(x) \\ &\text{s.t.} \quad x \in \mathbb{R}^n \end{aligned}$$

is equivalent to:

$$\begin{aligned} \text{(P)} : \quad z^* &= \text{maximum}_x \quad k(x) \\ &\text{s.t.} \quad x \in S . \end{aligned}$$

Similar logic applies to minimization problems over domains where the function values might take on the value $+\infty$.

3.2 Defining the Dual Problem

Recall the basic constrained optimization model:

$$\begin{aligned}
 \text{OP : } & \text{minimum}_x \quad f(x) \\
 & \text{s.t.} \\
 & \quad g_1(x) \leq 0, \\
 & \quad \cdot = \\
 & \quad \cdot \geq \\
 & \quad \vdots \\
 & \quad g_m(x) \leq 0, \\
 & \quad x \in X,
 \end{aligned}$$

In this model, we have $f(x) : \mathbb{R}^n \mapsto \mathbb{R}$ and $g_i(x) : \mathbb{R}^n \mapsto \mathbb{R}, i = 1, \dots, m$.

Of course, we can always convert the constraints to be “ \leq ”, and so for now we will presume that our problem is of the form:

$$\begin{aligned}
 \text{OP : } & z^* = \text{minimum}_x \quad f(x) \\
 & \text{s.t.} \\
 & \quad g_1(x) \leq 0, \\
 & \quad \vdots \\
 & \quad g_m(x) \leq 0, \\
 & \quad x \in X,
 \end{aligned}$$

Here X could be any set, such as:

- $X = \mathbb{R}^n$

- $X = \{x \mid x \geq 0\}$
- $X = \{x \mid g_i(x) \leq 0, i = m + 1, \dots, m + k\}$
- $X = \{x \mid x \in Z_+^n\}$
- $X = \{x \mid Ax \leq b\}$

Let z^* denote the optimal objective value of OP.

For a nonnegative vector u , form the *Lagrangian* function:

$$L(x, u) := f(x) + u^T g(x) = f(x) + \sum_{i=1}^m u_i g_i(x)$$

by taking the constraints out of the body of the model, and placing them in the objective function with costs/prices/multipliers u_i for $i = 1, \dots, m$.

We then solve the presumably easier problem:

$$\begin{array}{ll} L^*(u) := \text{minimum}_x & L(x, u) = \text{minimum}_x \quad f(x) + u^T g(x) \\ \text{s.t.} & x \in X \qquad \qquad \text{s.t.} \quad x \in X \end{array}$$

The function $L^*(u)$ is called the *dual function*.

We presume that computing $L^*(u)$ is an easy task.

The dual problem is then defined to be:

$$\begin{array}{ll} \text{D: } v^* = \text{maximum}_u & L^*(u) \\ \text{s.t.} & u \geq 0 \end{array}$$

4 Steps in the Construction of the Dual Problem

We start with the primal problem:

$$\begin{aligned} \text{OP : } z^* = \text{minimum}_x \quad & f(x) \\ \text{s.t.} \quad & \\ & g_i(x) \leq 0, \quad i = 1, \dots, m, \\ & x \in X, \end{aligned}$$

Constructing the dual involves a three-step process:

- **Step 1.** Create the Lagrangian

$$L(x, u) := f(x) + u^T g(x) .$$

- **Step 2.** Create the dual function:

$$\begin{aligned} L^*(u) := \text{minimum}_x \quad & f(x) + u^T g(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

- **Step 3.** Create the dual problem:

$$\begin{aligned} \text{D : } v^* = \text{maximum}_u \quad & L^*(u) \\ \text{s.t.} \quad & u \geq 0 \end{aligned}$$

5 Examples of Dual Constructions of Optimization Problems

5.1 The Dual of a Linear Problem

Consider the linear optimization problem:

$$\begin{aligned} \text{LP : } \quad & \text{minimum}_x \quad c^T x \\ & \text{s.t.} \quad Ax \geq b \end{aligned}$$

What is the dual of this problem?

$$L(x, u) = c^T x + u^T (b - Ax) = u^T b + (c - A^T u)^T x.$$

$$L^*(u) = \inf_{x \in \mathbb{R}^n} L(x, u) = \begin{cases} -\infty, & \text{if } A^T u \neq c \\ u^T b, & \text{if } A^T u = c \end{cases}$$

The dual problem (D) is:

$$(D) \max_{u \geq 0} L^*(u) = \max u^T b \text{ s.t. } A^T u = c, u \geq 0.$$

5.2 The Dual of a Binary Integer Problem

Consider the binary integer problem:

$$\begin{aligned} \text{IP : } \quad & \text{minimum}_x \quad c^T x \\ & \text{s.t.} \quad Ax \geq b \\ & x_j \in \{0, 1\}, \quad j = 1, \dots, n. \end{aligned}$$

What is the dual of this problem?

5.3 The Dual of a Log-Barrier Problem

Consider the following logarithmic barrier problem:

$$\begin{aligned} \text{BP : } \quad & \text{minimum}_{x_1, x_2, x_3} \quad 5x_1 + 7x_2 - 4x_3 - \sum_{j=1}^3 \ln(x_j) \\ & \text{s.t.} \quad \quad \quad x_1 + 3x_2 + 12x_3 = 37 \\ & \quad \quad \quad x_1 > 0, \quad x_2 > 0, \quad x_3 > 0 . \end{aligned}$$

What is the dual of this problem?

5.4 The Dual of a Quadratic Problem

Consider the quadratic optimization problem:

$$\begin{aligned} \text{QP : } \quad & \text{minimum}_x \quad \frac{1}{2}x^T Qx + c^T x \\ & \text{s.t.} \quad \quad \quad Ax \geq b \end{aligned}$$

where Q is SPSD (symmetric and positive semi-definite).

What is the dual of this problem?

$$L(x, u) = \frac{1}{2}x^T Qx + c^T x + u^T (b - Ax) = u^T b + (c - A^T u)^T x + \frac{1}{2}x^T Qx.$$

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

Assuming that Q is positive definite, we solve: $Q\tilde{x} = -(c - A^T u)$, i.e., $\tilde{x} = -Q^{-1}(c - A^T u)$ and

$$L^*(u) = L(\tilde{x}, u) = -\frac{1}{2}(c - A^T u)^T Q^{-1}(c - A^T u) + u^T b.$$

The dual problem (D) is:

$$(D) \max_{u \geq 0} L^*(u) = \max_{u \geq 0} -\frac{1}{2}(c - A^T u)^T Q^{-1}(c - A^T u) + u^T b.$$

5.5 Remarks on Problems with Different Formats of Constraints

Suppose that our problem has some inequality and equality constraints. Then just as in the case of linear optimization, we assign nonnegative dual variables u_i to constraints of the form $g_i(x) \leq 0$, unrestricted dual variables u_i to equality constraints $g_i(x) = 0$, and non-positive dual variables u_i to constraints of the form $g_i(x) \geq 0$. For example, suppose our problem is:

$$\begin{aligned} \text{OP : } & \text{minimum}_x && f(x) \\ & \text{s.t.} && \\ & && g_i(x) \leq 0, \quad i \in L \\ & && g_i(x) \geq 0, \quad i \in G \\ & && g_i(x) = 0, \quad i \in E \\ & && x \in X, \end{aligned}$$

Then we form the Lagrangian:

$$L(x, u) := f(x) + u^T g(x) = f(x) + \sum_{i \in L} u_i g_i(x) + \sum_{i \in G} u_i g_i(x) + \sum_{i \in E} u_i g_i(x)$$

and construct the dual function $L^*(u)$:

$$\begin{aligned} L^*(u) := & \text{minimum}_x \quad f(x) + u^T g(x) \\ & \text{s.t.} \quad x \in X \end{aligned}$$

The dual problem is then defined to be:

$$\begin{aligned} \text{D : } & \text{maximum}_u \quad L^*(u) \\ & \text{s.t.} \\ & u_i \geq 0, \quad i \in L \\ & u_i \leq 0, \quad i \in G \\ & u_i \begin{matrix} > \\ < \end{matrix}, \quad i \in E \end{aligned}$$

For simplicity, we presume that the constraints of OP are of the form $g_i(x) \leq 0$. This is only for ease of notation, as the results we develop pertain to the general case just described.

6 The Column Geometry of the Primal and Dual Problems

Let us consider the primal problem from a “resource-cost” point of view. For each $x \in X$, we have an array of resources and costs associated with x , namely:

$$\begin{pmatrix} s \\ z \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \\ z \end{pmatrix} = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \\ f(x) \end{pmatrix} = \begin{pmatrix} g(x) \\ f(x) \end{pmatrix} .$$

We can think of each of these arrays as an array of resources and cost in \mathbb{R}^{m+1} . Define the set I :

$$I := \left\{ (s, z) \in \mathbb{R}^{m+1} \mid \text{there exists } x \in X \text{ for which } s \geq g(x) \text{ and } z \geq f(x) \right\} .$$

This region is illustrated in Figure 1.

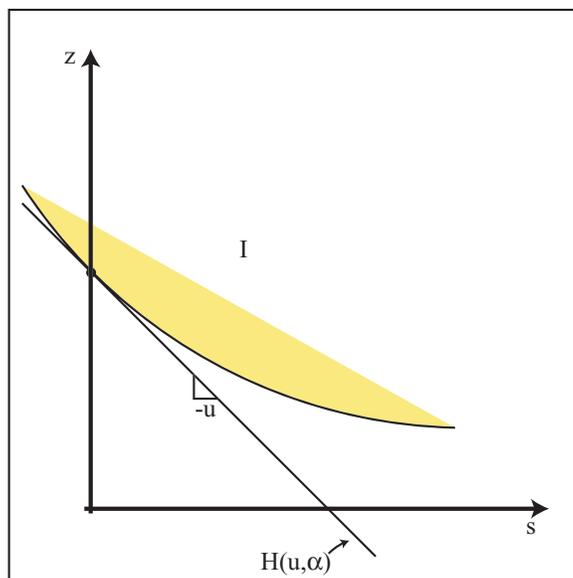


Figure 1: The set I .

Proposition 6.1 *If X is a convex set, and $f(\cdot), g_1(\cdot), \dots, g_m(\cdot)$ are convex functions on X , then I is a convex set. ■*

Let $H_{u,\alpha} = \{(s, z) \in \mathbb{R}^{m+1} : u^T s + z = \alpha\}$.

We call $H_{u,\alpha}$ is a *lower support* of I if $u^T s + z \geq \alpha$ for all $(s, z) \in I$.

Let $L = \{(s, z) \in \mathbb{R}^{m+1} : s = 0\}$. Note that $H_{u,\alpha} \cap L = \{(0, \alpha)\}$.

The problem of determining a hyperplane $H_{u,\alpha}$ that is a lower support of I , and whose intersection with L is the highest is:

$$\begin{aligned}
& \text{maximum}_{u,\alpha} \quad \alpha \\
& \quad \text{s.t.} \quad H_{u,\alpha} \text{ is a lower support of } I \\
= & \text{maximum}_{u,\alpha} \quad \alpha \\
& \quad \text{s.t.} \quad u^T s + z \geq \alpha \quad \text{for all } (s, z) \in I \\
= & \text{maximum}_{u \geq 0, \alpha} \quad \alpha \\
& \quad \text{s.t.} \quad u^T s + z \geq \alpha \quad \text{for all } (s, z) \in I \\
= & \text{maximum}_{u \geq 0, \alpha} \quad \alpha \\
& \quad \text{s.t.} \quad u^T g(x) + f(x) \geq \alpha \quad \text{for all } x \in X \\
= & \text{maximum}_{u \geq 0, \alpha} \quad \alpha \\
& \quad \text{s.t.} \quad L(x, u) \geq \alpha \quad \text{for all } x \in X \\
= & \text{maximum}_{u \geq 0, \alpha} \quad \alpha \\
& \quad \text{s.t.} \quad \inf_{x \in X} L(x, u) \geq \alpha \\
= & \text{maximum}_{u \geq 0, \alpha} \quad \alpha \\
& \quad \text{s.t.} \quad L^*(u) \geq \alpha \\
= & \text{maximum}_u \quad L^*(u) \\
& \quad \text{s.t.} \quad u \geq 0 .
\end{aligned}$$

This last expression is exactly the dual problem. Therefore:

The dual problem corresponds to finding a hyperplane $H_{u,\alpha}$ that is a lower support of I , and whose intersection with L is the highest. This highest value is exactly the value of the dual problem, namely v^ .*

7 The Dual is a Concave Maximization Problem

We start with the primal problem:

$$\begin{aligned} \text{OP : } & \text{minimum}_x \quad f(x) \\ & \text{s.t.} \\ & \quad g_i(x) \leq 0, \quad i = 1, \dots, m \\ & \quad x \in X, \end{aligned}$$

We create the Lagrangian:

$$L(x, u) := f(x) + u^T g(x)$$

and the dual function:

$$\begin{aligned} L^*(u) := & \text{minimum}_x \quad f(x) + u^T g(x) \\ & \text{s.t.} \quad x \in X \end{aligned}$$

The dual problem then is:

$$\begin{aligned} \text{D : } & \text{maximum}_u \quad L^*(u) \\ & \text{s.t.} \quad u \geq 0 \end{aligned}$$

Theorem 7.1 *The dual function $L^*(u)$ is a concave function.*

Proof: Let $u_1 \geq 0$ and $u_2 \geq 0$ be two values of the dual variables, and let $u = \lambda u_1 + (1 - \lambda)u_2$, where $\lambda \in [0, 1]$. Then

$$\begin{aligned} L^*(u) &= \min_{x \in X} f(x) + u^T g(x) \\ &= \min_{x \in X} \lambda [f(x) + u_1^T g(x)] + (1 - \lambda) [f(x) + u_2^T g(x)] \\ &\geq \lambda [\min_{x \in X} f(x) + u_1^T g(x)] + (1 - \lambda) [\min_{x \in X} (f(x) + u_2^T g(x))] \\ &= \lambda L^*(u_1) + (1 - \lambda) L^*(u_2) \quad . \end{aligned}$$

Therefore we see that $L^*(u)$ is a concave function. ■

8 Weak Duality

Let z^* and v^* be the optimal values of the primal and the dual problems:

$$\begin{aligned}
 \text{OP : } \quad z^* &= \text{minimum}_x \quad f(x) \\
 \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m \\
 & x \in X,
 \end{aligned}$$

$$\begin{aligned}
 \text{D : } \quad v^* &= \text{maximum}_u \quad L^*(u) \\
 \text{s.t.} \quad & u \geq 0
 \end{aligned}$$

Theorem 8.1 Weak Duality Theorem: *If \bar{x} is feasible for OP and \bar{u} is feasible for D, then*

$$f(\bar{x}) \geq L^*(\bar{u})$$

In particular,

$$z^* \geq v^* .$$

Proof: If \bar{x} is feasible for OP and \bar{u} is feasible for D, then

$$f(\bar{x}) \geq f(\bar{x}) + \bar{u}^T g(\bar{x}) \geq \min_{x \in X} f(x) + \bar{u}^T g(x) = L^*(\bar{u}) .$$

Therefore $z^* \geq v^*$. ■

Corollary 8.1 *If \bar{x} is feasible for OP and $\bar{u} \geq 0$ is feasible for D, and $f(\bar{x}) = L^*(\bar{u})$, then \bar{x} and \bar{u} are optimal solutions of OP and D, respectively.*

Corollary 8.2 *If $z^* = -\infty$, then D has no feasible solution.*

Corollary 8.3 *If $v^* = +\infty$, then OP has no feasible solution.*

Unlike in linear optimization, strong duality theorem cannot necessarily be established for general nonlinear optimization problems.

9 Saddlepoint Optimality Criteria

The pair (\bar{x}, \bar{u}) is called a *saddlepoint* of the Lagrangian $L(x, u)$ if $\bar{x} \in X$, $\bar{u} \geq 0$, and

$$L(\bar{x}, u) \leq L(\bar{x}, \bar{u}) \leq L(x, \bar{u}) \text{ for all } x \in X \text{ and } u \geq 0 .$$

Theorem 9.1 *A pair (\bar{x}, \bar{u}) is a saddlepoint of the Lagrangian if and only if*

1. $L(\bar{x}, \bar{u}) = L^*(\bar{u})$
2. $g(\bar{x}) \leq 0$, and
3. $\bar{u}^T g(\bar{x}) = 0$.

Moreover, (\bar{x}, \bar{u}) is a saddlepoint if and only if \bar{x} and \bar{u} are, respectively, optimal solutions of OP and D with no duality gap, that is, $z^* = f(\bar{x}) = L^*(\bar{u}) = v^*$.

Proof: Suppose that (\bar{x}, \bar{u}) is a saddlepoint. By definition, condition (1.) must be true. Also, for any $u \geq 0$,

$$f(\bar{x}) + \bar{u}^T g(\bar{x}) \geq f(\bar{x}) + u^T g(\bar{x}).$$

This implies that $g(\bar{x}) \leq 0$, since otherwise the above inequality can be violated by picking $u \geq 0$ appropriately. This establishes (2.). Taking $u = 0$, we get $\bar{u}^T g(\bar{x}) \geq 0$; hence, $\bar{u}^T g(\bar{x}) = 0$, establishing (3.).

Also, note that from the above observations that \bar{x} and \bar{u} are feasible for OP and D, respectively, and $f(\bar{x}) = L(\bar{x}, \bar{u}) = L^*(\bar{u})$, which implies that they are optimal solutions of their respective problems, and there is no duality gap.

Suppose now that \bar{x} and $\bar{u} \geq 0$ satisfy conditions (1.)–(3.). Then $L(\bar{x}, \bar{u}) \leq L(x, \bar{u})$ for all $x \in X$ by (1.). Furthermore,

$$L(\bar{x}, \bar{u}) = f(\bar{x}) \geq L(\bar{x}, u) \quad \forall u \geq 0.$$

Thus (\bar{x}, \bar{u}) is a saddlepoint.

Finally, suppose \bar{x} and \bar{u} are optimal solutions of OP and D with no duality gap. Primal and dual feasibility implies that

$$L^*(\bar{u}) \leq f(\bar{x}) + \bar{u}^T g(\bar{x}) \leq f(\bar{x}).$$

Since there is no duality gap, equality holds throughout, implying conditions (1.)–(3.), and hence (\bar{x}, \bar{u}) is a saddlepoint. ■

Remark 1 *Note that up until this point, we have made no assumptions about the functions $f(x), g_1(x), \dots, g_m(x)$ or the structure of the set X .*

Suppose that our constrained optimization problem OP is a convex program, namely $f(x), g_1(x), \dots, g_m(x)$ are convex functions, and that $X = \mathbb{R}^n$:

$$\begin{aligned} \text{OP : } \quad z^* = \text{minimum}_x \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0, \\ & x \in \mathbb{R}^n . \end{aligned}$$

Suppose that \bar{x} together with \bar{u} satisfy the KKT conditions:

$$(i) \quad \nabla f(\bar{x}) + \bar{u}^T \nabla g(\bar{x}) = 0$$

$$(ii) \quad g(\bar{x}) \leq 0$$

$$(iii) \quad \bar{u} \geq 0$$

$$(iv) \quad \bar{u}^T g(\bar{x}) = 0 .$$

Then because $f(x), g_1(x), \dots, g_m(x)$ are convex functions, condition (i) is equivalent to “ $L(\bar{x}, \bar{u}) = \min_x L(x, \bar{u}) = L^*(\bar{u})$ ”. Therefore the KKT conditions imply conditions (1.)–(3.) of Theorem 9.1, and so we have:

Theorem 9.2 *Suppose $X = \mathbb{R}^n$, and $f(x), g_1(x), \dots, g_m(x)$ are convex differentiable functions, and \bar{x} together with \bar{u} satisfy the KKT conditions. Then \bar{x} and \bar{u} are optimal solutions of OP and D, with no duality gap.*

10 Strong Duality for Convex Optimization Problems

We now assume throughout this section that X is an open convex set and $f(x), g_1(x), \dots, g_m(x)$ are convex functions. In this section we will explore and state sufficient conditions that will guarantee that the dual problem will have an optimal solution with no duality gap.

10.1 The Perturbation Function and the Strong Duality Theorem

We begin with the concept of perturbing the RHS vector of the optimization problem OP. Our standard optimization problem OP is:

$$\begin{aligned} \text{OP : } z^* &= \text{minimum}_x && f(x) \\ & \text{s.t.} && g_i(x) \leq 0, \quad i = 1, \dots, m \\ & && x \in X . \end{aligned}$$

For a given vector $y = (y_1, \dots, y_m)$, we perturb the RHS and obtain the new problem OP_y :

$$\begin{aligned} \text{OP}_y : z^*(y) &= \text{minimum}_x && f(x) \\ & \text{s.t.} && g_i(x) \leq y_i, \quad i = 1, \dots, m \\ & && x \in X . \end{aligned}$$

We call OP_y the *perturbed primal problem* and $z^*(y)$ the *perturbation function*. Note that our original problem OP is just OP_0 and that $z^* = z^*(0)$.

Next define

$$Y := \{y \in \mathbb{R}^m \mid \text{there exists } x \in X \text{ for which } g(x) \leq y\} .$$

Lemma 10.1 *Y is a convex set, and $z(y)$ is a convex function whose domain is Y .*

Proof: Let $y^1, y^2 \in Y$ and let $y^3 = \lambda y^1 + (1 - \lambda)y^2$ where $\lambda \in [0, 1]$. Let $x^1, x^2 \in X$ satisfy $y^1 \geq g(x^1)$, $y^2 \geq g(x^2)$ and let $x^3 = \lambda x^1 + (1 - \lambda)x^2$. Then $g(x^3) \leq y^3$, whereby $y^3 \in Y$, proving that Y is convex.

To show that $z(y)$ is a convex function, let y^1, y^2, y^3 be as given above. The aim is to show that $z(y^3) \leq \lambda z(y^1) + (1 - \lambda)z(y^2)$. First assume that $z(y^1)$ and $z(y^2)$ are finite. For any $\epsilon > 0$, there exist x^1 and x^2 for which $x^1, x^2 \in X$, $g(x^1) \leq y^1$, $g(x^2) \leq y^2$, and $|z(y^1) - f(x^1)| \leq \epsilon$, $|z(y^2) - f(x^2)| \leq \epsilon$. Now let $x^3 = \lambda x^1 + (1 - \lambda)x^2$. Then $g(x^3) \leq \lambda g(x^1) + (1 - \lambda)g(x^2) \leq \lambda y^1 + (1 - \lambda)y^2 = y^3$, and

$$\begin{aligned} z(y^3) &\leq f(x^3) \leq \lambda f(x^1) + (1 - \lambda)f(x^2) \leq \lambda(z(y^1) + \epsilon) + (1 - \lambda)(z(y^2) + \epsilon) \\ &= \lambda z(y^1) + (1 - \lambda)z(y^2) + \epsilon . \end{aligned}$$

As this is true for any $\epsilon > 0$, $z(y^3) \leq \lambda z(y^1) + (1 - \lambda)z(y^2)$, and so $z(y)$ is convex. If $z(y^1)$ and/or $z(y^2)$ is not finite, the proof goes through with appropriate modifications. ■

The perturbation function $z(y)$ is illustrated in Figure 2.

Corollary 10.1 1. $z(y)$ is continuous on $\text{int}Y$

2. $z'(y; d)$ exists wherever $z(y)$ is finite

3. for every $\bar{y} \in \text{int}Y$ there exists a subgradient of $z(\cdot)$ at $y = \bar{y}$

4. the subgradient of $z(y)$ at $y = \bar{y}$ is unique if and only if $z(\cdot)$ is differentiable at $y = \bar{y}$.

■

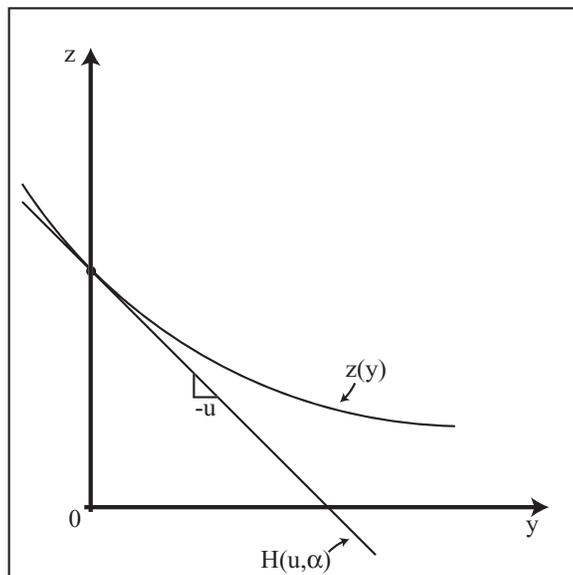


Figure 2: The perturbation function $z(y)$.

Theorem 10.1 (Strong Duality Theorem) *Suppose that X is an open convex set and $f(x), g_1(x), \dots, g_m(x)$ are convex functions. Under the hypothesis that $\partial z(0) \neq \emptyset$, then:*

1. D attains its optimum at some $\bar{u} \geq 0$.
2. \bar{u} is an optimal solution of D if and only if $(-\bar{u}) \in \partial z(0)$.
3. $z^* = v^*$
4. Let \bar{u} be any dual optimal solution. The set of primal optimal solutions (if any) is characterized as the intersection $S^1 \cap S^2$, where:

$$\begin{aligned} S^1 &= \{x \in X \mid g(x) \leq 0, \bar{u}^T g(x) = 0\} \\ S^2 &= \{x \in X \mid L(x, \bar{u}) = L^*(\bar{u})\} . \end{aligned}$$

Proof: Since $\partial z(0) \neq \emptyset$, let $-\bar{u} \in \partial z(0)$. Then

$$z(y) \geq z(0) - \bar{u}^T (y - 0).$$

Noting that $z(y)$ is non-increasing in y , let e^i denote the i^{th} unit vector, and observe:

$$z(0) \geq z(e^i) \geq z(0) - \bar{u}^{-T}(e^i - 0) = z(0) - \bar{u}_i,$$

which shows that $\bar{u} \geq 0$. Thus \bar{u} is feasible for D .

Next, note that for any $x \in X$,

$$z(g(x)) = \inf_{\substack{\tilde{x} \\ \text{s.t. } g(\tilde{x}) \leq g(x) \\ \tilde{x} \in X}} f(\tilde{x}) \leq f(x)$$

which implies that

$$f(x) \geq z(g(x)) \geq z(0) - \bar{u}^T(g(x) - 0),$$

that is,

$$z(0) \leq f(x) + \bar{u}^T g(x).$$

Thus

$$z^* = z(0) \leq \inf_{x \in X} \{f(x) + \bar{u}^T g(x)\} = L^*(\bar{u}).$$

By weak duality, however, $z^* \geq L^*(\bar{u})$, so $z^* = L^*(\bar{u})$ and so \bar{u} is optimal for D . This shows (i) and (iii) and (\Leftarrow) of (ii).

We now show (\Rightarrow) of (ii). Suppose \bar{u} is optimal for D . Then from (iii) we have

$$\inf_{x \in X} \{f(x) + \bar{u}^T g(x)\} = L^*(\bar{u}) = z^* = z(0).$$

Thus, $f(x) + \bar{u}^T g(x) \geq z(0)$ for every $x \in X$.

For any given fixed y , if $x \in X$ and $g(x) \leq y$, then

$$f(x) + \bar{u}^T y \geq f(x) + \bar{u}^T g(x) \geq z(0).$$

Thus for fixed y ,

$$\bar{u}^T y + z(y) = \inf_{\substack{x \\ \text{s.t. } g(x) \leq y \\ x \in X}} f(x) + \bar{u}^T y \geq z(0),$$

that is,

$$z(y) \geq z(0) - \bar{u}^T(y - 0),$$

and so $\bar{u} \in \partial z(0)$. This completes (ii).

We now show (iv). Suppose $\bar{x} \in S^1 \cap S^2$. Then $\bar{u} \geq 0$, \bar{x} satisfies $L^*(\bar{u}) = L(\bar{x}, \bar{u})$, and $\bar{u}^T g(\bar{x}) = 0$, whereby (\bar{x}, \bar{u}) satisfy the conditions of Theorem 9.1 and hence \bar{x} is optimal for the primal problem OP. Conversely, suppose \bar{x} is optimal for the primal problem. Then $z^* = v^*$, and so again invoking Theorem 9.1 we have $g(\bar{x}) \leq 0$, $L^*(\bar{u}) = L(\bar{x}, \bar{u})$, and $\bar{u}^T g(\bar{x}) = 0$. Therefore $\bar{x} \in S_1 \cap S_2$. ■

10.2 Stability

The only hypothesis of the Strong Duality Theorem is that $z(y)$ has a subgradient at $y = 0$, or equivalently, $\partial z(0) \neq \emptyset$. We call an optimization problem OP *stable* if $\partial z(0) \neq \emptyset$. An equivalent and generally more useful characterization of stability is given in the following:

Lemma 10.2 *Let $h(y) : Y \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$ be a convex function. If $h(y)$ is finite at $y = \bar{y}$, then $\partial h(\bar{y}) \neq \emptyset$ if and only if there exists a positive scalar M for which:*

$$\frac{h(\bar{y}) - h(y)}{\|y - \bar{y}\|} \leq M \quad \text{for all } y \in Y, y \neq \bar{y}.$$

Proof: First suppose that $\partial h(\bar{y}) \neq \emptyset$. Let $\bar{\gamma} \in \partial h(\bar{y})$. Then

$$h(y) \geq h(\bar{y}) + \bar{\gamma}^T(y - \bar{y}) \quad \text{for all } y \in Y,$$

and so $h(y) > -\infty$ for all $y \in Y$. Let $M = \|\bar{\gamma}\|$. Then

$$\frac{h(\bar{y}) - h(y)}{\|y - \bar{y}\|} \leq \frac{\bar{\gamma}^T(\bar{y} - y)}{\|y - \bar{y}\|} \leq \frac{\|\bar{\gamma}\| \|\bar{y} - y\|}{\|y - \bar{y}\|} = \|\bar{\gamma}\| = M.$$

To prove the reverse implication, we suppose that there exists a scalar $M > 0$ satisfying the stated inequality. Then $h(y) > -\infty$ for all $y \in Y$. Define the sets

$$S = \{(y, z) \in \mathbb{R}^{m+1} \mid y \in Y \text{ and } h(y) + z \leq h(\bar{y})\}$$

and

$$T = \{(y, z) \in \mathbb{R}^{m+1} \mid M\|y - \bar{y}\| < z\} .$$

It is easy to see that S and T are convex sets, and that if $(y, z) \in S$, then $M\|y - \bar{y}\| \geq h(\bar{y}) - h(y) \geq z$, so that $(y, z) \notin T$. As $S \cap T = \emptyset$, there exists a hyperplane H separating S and T . That is, there exists $(\bar{\gamma}, \rho) \neq (0, 0)$ and α such that

$$(y, z) \in S \Rightarrow \bar{\gamma}^T y + \rho z \leq \alpha$$

and

$$(y, z) \in T \Rightarrow \bar{\gamma}^T y + \rho z \geq \alpha .$$

In particular $(\bar{y}, 0) \in S$ which implies that

$$\bar{\gamma}^T \bar{y} \leq \alpha .$$

Also $(\bar{y}, \theta) \in T$ for any $\theta > 0$ which implies that $\bar{\gamma}^T \bar{y} + \rho\theta > \alpha$ which implies that $\bar{\gamma}^T \bar{y} \geq \alpha$, and so $\bar{\gamma}^T \bar{y} = \alpha$. This also implies that $\rho \geq 0$.

If $\rho > 0$ we can assume without loss of generality that $\rho = 1$. Now for any $y \in Y$, let $z = h(\bar{y}) - h(y)$. Then $(y, z) \in S$ and so

$$\bar{\gamma}^T y + h(\bar{y}) - h(y) \leq \bar{\gamma}^T \bar{y} ,$$

and so

$$h(y) \geq h(\bar{y}) + \bar{\gamma}^T (y - \bar{y}), \text{ so } \bar{\gamma} \in \partial h(\bar{y}) ,$$

which proves the result in this case.

It remains to prove that $\rho = 0$ is not possible. To see this, assume that $\rho = 0$. For any $y \in Y$, let $z = M\|y - \bar{y}\| + 1$. Then $(y, z) \in T$ and so $\bar{\gamma}^T y + \rho z = \bar{\gamma}^T y \geq \alpha = \bar{\gamma}^T \bar{y}$. As this is true for any $y \in Y$, it follows that $\bar{\gamma} = 0$ and so $(\bar{\gamma}, \rho) = (0, 0)$, which is a contradiction. ■

Therefore, we can equivalently say that an optimization problem OP is stable if $z^* = z(0)$ is finite and there is a scalar $M > 0$ for which the perturbation function $z(y)$ satisfies:

$$\frac{z(0) - z(y)}{\|y\|} \leq M \text{ for all } y \in Y, y \neq 0 .$$

This alternative characterization of stability says that OP is stable if the function $z(y)$ does not decrease infinitely steeply in any direction away from $y = 0$.

10.3 Slater Points and Stability

Now consider the problem:

$$\begin{aligned} \text{OPE : } z^* = \text{minimum}_x \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \\ & Ax - b = 0 \\ & x \in \mathbb{R}^n , \end{aligned}$$

where we assume here that A has full row rank and $f(x), g_1(x), \dots, g_m(x)$ are convex functions as before. A point x^0 is called a *Slater point* of OPE if x^0 satisfies:

- $g(x^0) < 0$ and
- $Ax^0 = b$.

We now state and prove the following result which shows that the existence of a Slater point in a convex problem implies stability, which in turn implies strong duality.

Lemma 10.3 *Suppose that $f(x), g_1(x), \dots, g_m(x)$ are all convex functions, that A has full row rank, and problem OPE has a Slater point. If z^* is finite, then problem OPE is stable.*

Proof: We can rewrite OPE as OPE_0 where OPE_y is:

$$\begin{aligned} \text{OPE}_y : z^*(y) = \text{minimum}_x \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq y \\ & x \in X , \end{aligned}$$

where $X = \{x \in \mathbb{R}^n \mid Ax = b\}$.

We are concerned with finding an M such that

$$\frac{z(0) - z(y)}{\|y\|} \leq M .$$

Let $h(\epsilon) = z(\epsilon, \dots, \epsilon) = z(y)$ where $y = (\epsilon, \dots, \epsilon)$. As $z(\cdot)$ is convex, so is $h(\epsilon)$. There exists $\bar{\epsilon} > 0$ such that $|\epsilon| \leq \bar{\epsilon}$ implies $g(x^0) < (\epsilon, \dots, \epsilon)$, and so x^0 is feasible for $\text{OPE}_{y=\text{OPE}_{(\epsilon, \dots, \epsilon)}}$, whereby

$$h(\epsilon) = z(\epsilon, \dots, \epsilon) \leq f(x^0).$$

Also $z^* = h(0) \leq \frac{1}{2}h(\bar{\epsilon}) + \frac{1}{2}h(-\bar{\epsilon}) \leq \frac{1}{2}h(\bar{\epsilon}) + \frac{1}{2}f(x^0)$. Thus, $h(\bar{\epsilon}) \geq 2z^* - f(x^0) > -\infty$, and therefore $|\epsilon| \leq \bar{\epsilon}$ implies $h(\epsilon) \geq h(\bar{\epsilon}) > -\infty$, so $h(\epsilon)$ is convex and finite for all ϵ satisfying $|\epsilon| \leq \bar{\epsilon}$. Therefore, $h(\epsilon)$ has a subgradient at $\epsilon = 0$, i.e., there is a K such that

$$h(\epsilon) \geq h(0) + K(\epsilon - 0).$$

Let $M = |K|$, and for any y define $\tilde{y} = \max_i |y_i|$. We have

$$\frac{z(0) - z(y)}{\|y\|} \leq \frac{h(0) - h(\tilde{y})}{\|y\|} \leq \frac{-K\tilde{y}}{\|y\|} \leq \frac{M\tilde{y}}{\|y\|} \leq M .$$

Thus OPE is stable. ■

11 Duality Strategies

11.1 Dualizing “Bad” Constraints

Suppose we wish to solve:

$$\begin{aligned} \text{OP : } & \text{minimum}_x \quad c^T x \\ & \text{s.t.} \quad Ax \leq b \\ & \quad \quad Nx \leq g . \end{aligned}$$

Suppose that optimization over the constraints “ $Nx \leq g$ ” is easy, but that the addition of the constraints “ $Ax \leq b$ ” makes the problem much more difficult. This can happen, for example, when the constraints “ $Nx \leq g$ ” are network constraints, and when “ $Ax \leq b$ ” are non-network constraints in the model.

Let

$$X = \{x \mid Nx \leq g\}$$

and re-write OP as:

$$\begin{aligned} \text{OP : } & \text{minimum}_x \quad c^T x \\ & \text{s.t.} \quad Ax \leq b \\ & \quad \quad x \in X . \end{aligned}$$

The Lagrangian is:

$$L(x, u) = c^T x + u^T (Ax - b) = -u^T b + (c^T + u^T A)x ,$$

and the dual function is:

$$\begin{aligned} L^*(u) := & \text{minimum}_x \quad -u^T b + (c^T + u^T A)x \\ & \text{s.t.} \quad x \in X , \end{aligned}$$

which is

$$\begin{aligned} L^*(u) := & \text{minimum}_x \quad -u^T b + (c^T + u^T A)x \\ & \text{s.t.} \quad Nx \leq g . \end{aligned}$$

Notice that $L^*(u)$ is easy to evaluate for any value of u , and so we can attempt to solve OP by designing an algorithm to solve the dual problem:

$$\begin{aligned} \text{D : } & \text{maximum}_u \quad L^*(u) \\ & \text{s.t.} \quad u \geq 0 . \end{aligned}$$

11.2 Dualizing A Large Problem into Many Small Problems

Suppose we wish to solve:

$$\begin{aligned} \text{OP : } & \text{minimum}_{x^1, x^2} \quad (c^1)^T x^1 + (c^2)^T x^2 \\ & \text{s.t.} \quad B^1 x^1 + B^2 x^2 \leq d \\ & \quad \quad A^1 x^1 \leq b^1 \\ & \quad \quad \quad A^2 x^2 \leq b^2 \end{aligned}$$

Notice here that if it were not for the constraints “ $B^1 x^1 + B^2 x^2 \leq d$ ”, that we would be able to separate the problem into two separate problems. Let us dualize on these constraints. Let:

$$X = \{(x^1, x^2) \mid A^1 x^1 \leq b^1, A^2 x^2 \leq b^2\}$$

and re-write OP as:

$$\begin{aligned} \text{OP : } & \text{minimum}_{x^1, x^2} \quad (c^1)^T x^1 + (c^2)^T x^2 \\ & \text{s.t.} \quad B^1 x^1 + B^2 x^2 \leq d \\ & \quad \quad (x^1, x^2) \in X \end{aligned}$$

The Lagrangian is:

$$\begin{aligned} L(x, u) &= (c^1)^T x^1 + (c^2)^T x^2 + u^T (B^1 x^1 + B^2 x^2 - d) \\ &= -u^T d + ((c^1)^T + u^T B^1) x^1 + ((c^2)^T + u^T B^2) x^2, \end{aligned}$$

and the dual function is:

$$\begin{aligned} L^*(u) &= \text{minimum}_{x^1, x^2} -u^T d + ((c^1)^T + u^T B^1) x^1 + ((c^2)^T + u^T B^2) x^2 \\ &\text{s.t.} \quad (x^1, x^2) \in X \end{aligned}$$

which can be re-written as:

$$\begin{aligned} L^*(u) &= -u^T d \\ &+ \text{minimum}_{A^1 x^1 \leq b^1} ((c^1)^T + u^T B^1) x^1 \\ &+ \text{minimum}_{A^2 x^2 \leq b^2} ((c^2)^T + u^T B^2) x^2 \end{aligned}$$

Notice once again that $L^*(u)$ is easy to evaluate for any value of u , and so we can attempt to solve OP by designing an algorithm to solve the dual problem:

$$\begin{aligned} \text{D : } &\text{maximum}_u L^*(u) \\ &\text{s.t.} \quad u \geq 0 \end{aligned}$$

12 Illustration of Lagrange Duality in Discrete Optimization

In order to suggest the computational power of duality theory and to illustrate duality constructs in discrete optimization, let us consider the simple constrained shortest path problem portrayed in Figure 3.

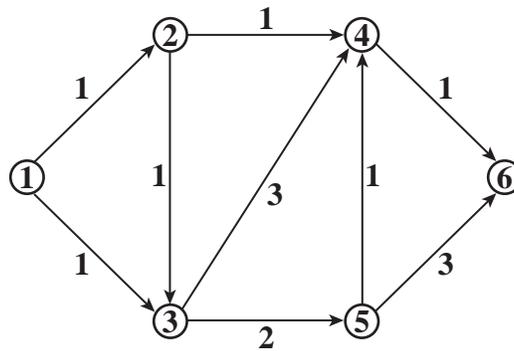


Figure 3: A constrained shortest path problem.

The objective in this problem is to find the shortest (least cost) path from node 1 to node 6 subject to the constraint that the chosen path uses exactly four arcs. In an applied context the network might contain millions of nodes and arcs and billions of paths. For example, in one application setting that can be solved using the ideas considered in this discussion, each arc would have both a travel cost and travel time (which for our example is 1 for each arc). The objective is to find the least cost path between a given pair of nodes from all those paths whose travel time does not exceed a prescribed limit (which in our example is 4 and must be met exactly).

Table 1 shows that only three paths, namely 1-2-3-4-6, 1-2-3-5-6, and 1-3-5-4-6 between nodes 1 and 6 contain exactly four arcs. Since path 1-3-5-4-6 has the least cost of these three paths (at a cost of 5), this path is the optimal solution of the constrained shortest path problem.

Path	Number of Arcs	Cost
1-2-4-6	3	3
1-3-4-6	3	5
1-3-5-6	3	6
1-2-3-4-6	4	6
1-2-3-5-6	4	7
1-3-5-4-6	4	5
1-2-3-5-4-6	5	6

Table 1: Paths from node 1 to node 6.

Now suppose that we have available an efficient algorithm for solving (unconstrained) shortest path problems (indeed, such algorithms are readily available in practice), and that we wish to use this algorithm to solve the given problem.

Let us first formulate the constrained shortest path problem as an integer program:

CSPP :

$$\begin{aligned}
 z^* = \min_{x_{ij}} \quad & x_{12} + x_{13} + x_{23} + x_{24} + 3x_{34} + 2x_{35} + x_{54} + x_{46} + 3x_{56} \\
 \text{s.t.} \quad & x_{12} + x_{13} + x_{23} + x_{24} + x_{34} + x_{35} + x_{54} + x_{46} + x_{56} = 4 \\
 & x_{12} + x_{13} = 1 \\
 & x_{12} - x_{23} - x_{24} = 0 \\
 & x_{13} + x_{23} - x_{34} - x_{35} = 0 \\
 & x_{24} + x_{34} + x_{54} - x_{46} = 0 \\
 & x_{35} - x_{54} - x_{56} = 0 \\
 & x_{46} + x_{56} = 1 \\
 & x_{ij} \in \{0, 1\} \text{ for all } x_{ij} .
 \end{aligned}$$

In this formulation, x_{ij} is a binary variable indicating whether or not arc $i - j$ is chosen as part of the optimal path. The first constraint in this formulation specifies that the optimal path must contain exactly four arcs. The remaining constraints are the familiar constraints describing a shortest path problem. They indicate that one unit of flow is available at node 1 and must be sent to node 6; nodes 2, 3, 4, and 5 act as intermediate nodes that neither generate nor consume flow. Since the decision variables x_{ij} must be integer, any feasible solution to these constraints traces out a path from node 1 to node 6. The objective function determines the cost of any such path.

Notice that if we eliminate the first constraint from this problem, then the problem becomes an unconstrained shortest path problem that can be solved by our available algorithm. Suppose, for example, that we dualize

the first (the complicating) constraint by moving it to the objective function multiplied by a Lagrange multiplier u . The objective function then becomes:

$$\text{minimum}_{x_{ij}} \sum_{i,j} c_{ij}x_{ij} + u \left(4 - \sum_{i,j} x_{ij} \right)$$

and collecting the terms, the modified problem becomes:

$$L^*(u) = \min_{x_{ij}} \begin{aligned} &(1-u)x_{12} + (1-u)x_{13} + (1-u)x_{23} \\ &+ (1-u)x_{24} + (3-u)x_{34} + (2-u)x_{35} \\ &+ (1-u)x_{54} + (1-u)x_{46} + (3-u)x_{56} + 4u \end{aligned}$$

s.t.

$$\begin{array}{rcccccccc} x_{12} & +x_{13} & & & & & & & = & 1 \\ x_{12} & & -x_{23} & -x_{24} & & & & & = & 0 \\ & x_{13} & +x_{23} & & -x_{34} & -x_{35} & & & = & 0 \\ & & & x_{24} & +x_{34} & & +x_{54} & -x_{46} & = & 0 \\ & & & & & x_{35} & -x_{54} & & -x_{56} & = & 0 \\ & & & & & & & x_{46} & +x_{56} & = & 1 \end{array}$$

$$x_{ij} \in \{0, 1\} \text{ for all } x_{ij} \quad .$$

Let $L^*(u)$ denote the optimal objective function value of this problem, for any given multiplier u .

Notice that for a fixed value of u , $4u$ is a constant and the problem becomes a shortest path problem with modified arc costs given by

$$c'_{ij} = c_{ij} - u \quad .$$

Note that for a fixed value of u , this problem can be solved by our available algorithm for the shortest path problem. Also notice that for any fixed value of u , $L^*(u)$ is a lower bound on the optimal objective value z^* to the constrained shortest path problem defined by the original formulation. To obtain the best lower bounds, we need to solve the Lagrangian dual problem:

$$\begin{aligned} \text{D : } \quad v^* &= \text{maximum}_u \quad L^*(u) \\ &\text{s.t.} \quad u \leq 0 \end{aligned}$$

To relate this problem to the earlier material on duality, let the decision set X be given by the set of feasible paths, namely the solutions of:

$$\begin{array}{rcccccccc} x_{12} & +x_{13} & & & & & & & = & 1 \\ x_{12} & & -x_{23} & -x_{24} & & & & & = & 0 \\ & x_{13} & +x_{23} & & -x_{34} & -x_{35} & & & = & 0 \\ & & & x_{24} & +x_{34} & & +x_{54} & -x_{46} & = & 0 \\ & & & & & x_{35} & -x_{54} & & -x_{56} & = & 0 \\ & & & & & & & x_{46} & +x_{56} & = & 1 \end{array}$$

$$x_{ij} \in \{0, 1\} \text{ for all } x_{ij} \quad .$$

Of course, this set will be a finite (discrete) set. That is, P is the set of paths connecting node 1 to node 6. Suppose that we plot the cost z and the number of arcs r used in each of these paths, see Figure 4.

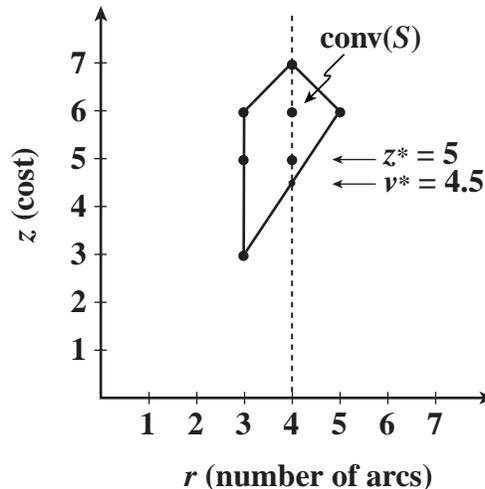


Figure 4: Resources and costs.

We obtain the seven dots in this figure corresponding to the seven paths listed in Table 1. Now, because our original problem was formulated as a minimization problem, the geometric dual is obtained by finding the line that lies on or below these dots and that has the largest intercept with the feasibility line. Note from Figure 4 that the optimal value of the dual is $v^* = 4.5$, whereas the optimal value of the original primal problem is $z^* = 5$. In this example, because the cost-resource set is not convex (it is simply the seven dots), we encounter a duality gap. In general, integer optimization models will usually have such duality gaps, but they are typically not very large in practice. Nevertheless, the dual bounding information provided by $v^* \leq z^*$ can be used in branch and bound procedures to devise efficient algorithms for solving integer optimization models.

It might be useful to re-cap. The original constrained shortest path problem might be very difficult to solve directly (at least when there are hundreds of thousands of nodes and arcs). Duality theory permits us to relax (eliminate) the complicating constraint and to solve a much easier Lagrangian shortest path problem with modified arc costs $c'_{ij} = c_{ij} - u$. Solving for any u generates a lower bound $L^*(u)$ on the value z^* to the

original problem. Finding the best lower bound $\max_u L^*(u)$ requires that we develop an efficient procedure for solving for the optimal u . Duality theory permits us to replace a single and difficult problem (in this instance, a constrained shortest-path problem) with a sequence of much easier problems (in this instance, a number of unconstrained shortest path problems), one for each value of u encountered in our search procedure. Practice on a wide variety of problem applications has shown that the dual problem and subsequent branch and bound procedures can be extremely effective in many applications.

To conclude this discussion, note from Figure 4 that the optimal value for u in the dual problem is $u = 1.5$. Also note that if we subtract $u = 1.5$ from the cost of every arc in the network, then the shortest path (with respect to the modified costs) has cost -1.5 . Both paths 1-2-4-6 and 1-2-3-5-4-6 are shortest paths for this modified problem. Adding $4u = 4 \times 1.5 = 6$ to the modified shortest path cost of -1.5 gives a value of $6 - 1.5 = 4.5$ which equals v^* . Note that if we permitted the flow in the original problem to be split among multiple paths (i.e., formulate the problem as a linear optimization problem instead of as an integer optimization problem), then the optimal solution would send one half of the flow on both paths 1-2-4-6 and 1-2-3-5-4-6 and incur a cost of:

$$\frac{1}{2} \times 3 + \frac{1}{2} \times 6 = 4.5 .$$

In general, the value of the dual problem is always identical to the value of the original (primal) problem if we permit convex combinations of primal solutions. This equivalence between convexification and dualization is easy to see geometrically.

13 Conic Duality

13.1 Cones and Conic Convex Optimization

We say that $K \subset \mathbb{R}^n$ is a convex *cone* if:

$$x, y \in K \text{ and } \alpha, \beta \geq 0 \Rightarrow \alpha x + \beta y \in K .$$

Some examples of convex cones that are useful to us are:

- $\mathbb{R}_+^n := \{x \in \mathbb{R}^n \mid x_j \geq 0, j = 1, \dots, n\}$
- $Q^n = \{x \in \mathbb{R}^n \mid x_1 \geq \sqrt{\sum_{j=2}^n x_j^2}\}$
- \mathbb{R}^n
- $\{0\}^n \in \mathbb{R}^n$
- $S_+^{k \times k} = \{X \in S^{k \times k} \mid v^T X v \geq 0 \text{ for all } v \in \mathbb{R}^k\}$
- $K = K_1 \times K_2 \times \dots \times K_l$ where K_j is a closed convex cone, $j = 1, \dots, l$

We consider the following convex optimization problem in conic form:

$$\begin{aligned} \text{CP : } \quad z^* &= \text{minimum}_x \quad c^T x \\ &\text{s.t.} \quad Ax = b \\ &\quad \quad \quad x \in K , \end{aligned}$$

where K is a closed convex cone.

13.2 Dual Cones and the Conic Dual Problem

Let K^* denote the *dual cone* of the closed convex cone $K \subset \mathbb{R}^n$, defined by:

$$K^* := \{y \in \mathbb{R}^n \mid y^T x \geq 0 \text{ for all } x \in K\} .$$

Proposition 13.1 *If K is a nonempty closed convex cone, then K^* is a nonempty closed convex cone.*

Proof: Notice that $0 \in K^*$, which shows that $K^* \neq \emptyset$. If $y^1, y^2 \in K^*$, then for every $x \in K$ and every $\alpha, \beta \geq 0$ we have $(\alpha y^1 + \beta y^2)^T x \geq 0$, which shows that K^* is a convex cone. Suppose that $y^1, y^2, \dots \in K^*$ and $\lim_{j \rightarrow \infty} y^j = \bar{y}$. Then for every $x \in K$ we have $(y^j)^T x \geq 0$ and so $\bar{y}^T x \geq 0$, whereby $\bar{y} \in K^*$, which shows that K^* is closed. ■

Proposition 13.2 *If K is a nonempty closed convex cone, then $(K^*)^* = K$.*

Proof: We have

$$K^* := \{y \in \mathbb{R}^n \mid y^T x \geq 0 \text{ for all } x \in K\}$$

and

$$(K^*)^* := \{z \in \mathbb{R}^n \mid z^T y \geq 0 \text{ for all } y \in K^*\}.$$

For every $x \in K$ we have $x^T y \geq 0$ for all $y \in K^*$, which shows that $K \subset (K^*)^*$. Suppose that $K \neq (K^*)^*$. Then there exists $\bar{z} \in (K^*)^*$ for which $\bar{z} \notin K$. Since $\bar{z} \notin K$ and K is a closed convex set, there exists a hyperplane that separates \bar{z} from K . Thus there exists $y \neq 0$ and α for which $y^T \bar{z} < \alpha$ and $y^T x > \alpha$ for all $x \in K$. But since K is a cone this means that $\alpha = 0$, and so $y \in K^*$. However, $y^T \bar{z} < \alpha = 0$, which implies that $\bar{z} \notin (K^*)^*$, which is a contradiction. Therefore $(K^*)^* = K$. ■

Here are the dual cones associated with the above examples:

- $(\mathbb{R}_+^n)^* = \mathbb{R}_+^n$
- $(Q^n)^* = Q^n$
- $(\mathbb{R}^n)^* = \{0\}^n$
- $(\{0\}^n)^* = \mathbb{R}^n$
- $(S_+^{k \times k})^* = S_+^{n \times n}$
- $(K_1 \times K_2 \times \dots \times K_l)^* = K_1^* \times K_2^* \times \dots \times K_l^*$

We form the Lagrangian:

$$L(x, u) := c^T x + u^T (b - Ax) .$$

We define the dual function by:

$$L^*(u) := \min_{x \in K} L(x, u) = \min_{x \in K} c^T x + u^T (b - Ax) .$$

Then notice that

$$\begin{aligned} L^*(u) &= \min_{x \in K} c^T x + u^T (b - Ax) \\ &= u^T b + \min_{x \in K} (c^T - u^T A)x \\ &= \begin{cases} b^T u & \text{if } c - A^T u \in K^* \\ -\infty & \text{if } c - A^T u \notin K^* . \end{cases} \end{aligned}$$

From this we write down the dual problem:

$$\begin{aligned} \text{CD : } v^* &= \text{maximum}_{u,s} b^T u \\ &\text{s.t.} \quad A^T u + s = c \\ &\quad s \in K^* . \end{aligned}$$

Proposition 13.3 (Weak Duality) *If x is feasible for CP and (u, s) is feasible for CD, then $c^T x \geq b^T u$. Consequently, $z^* \leq v^*$.*

Proof: If x is feasible for CP and (u, s) is feasible for CD, then $c^T x - b^T u = u^T Ax + s^T x - b^T u = s^T x \geq 0$ because $x \in K$ and $s \in K^*$. ■

Remark 1 *The conic dual of CD is CP.*

13.3 Slater Points and Strong Duality

A *Slater point* for CP is a point x^0 that satisfies:

$$Ax^0 = b \quad \text{and} \quad x^0 \in \text{int}K .$$

Theorem 2 (Strong Duality Theorem) *If CP has a Slater point, then $v^* = z^*$ and the dual attains its optimum.*

Proof: Define the following set:

$$S := \left\{ (w, \lambda, \alpha) \mid \text{there exists } x \text{ for which } Ax = b - w, x + \lambda \in K, c^T x < z^* + \alpha \right\} ,$$

and notice that S is a nonempty convex set. Also notice that $(w, \lambda, \alpha) = (0, 0, 0) \notin S$, and so there is a hyperplane separating $(0, 0, 0)$ from S . This means that there exists $(u, s, \theta) \neq 0$ for which

$$u^T w + s^T \lambda + \theta \alpha \geq 0 \quad \text{for all } (w, \lambda, \alpha) \in S .$$

This is the same as

$$u^T (b - Ax) + s^T (v - x) + \theta (c^T x - z^* + \delta) \geq 0 \quad \text{for all } x, v \in K, \delta > 0 .$$

Rearranging terms we have

$$u^T b + (\theta c - A^T u - s)^T x + s^T v + \theta (\delta - z^*) \geq 0 \quad \text{for all } x, v \in K, \delta > 0 .$$

This then implies that $\theta \geq 0$, $A^T u + s = \theta c$ and $s \in K^*$. If $\theta > 0$ then we can assume (by rescaling if necessary) that $\theta = 1$, whereby we see that (u, s) is feasible for the dual problem and in also from the above that $b^T u - z^* \geq -\delta$ for all $\delta > 0$. This in turn implies that $v^* \geq b^T u \geq z^*$, which by weak duality, shows that $z^* = v^*$ and that (u, s) is an optimal solution of the dual problem.

If instead $\theta = 0$ then we have $A^T u + s = 0$. Therefore $0 \leq s^T x^0 = -u^T Ax^0 = -u^T b \leq 0$, and so $s^T x^0 = 0$. Then since $x^0 \in \text{int}K$ and there exists $\epsilon > 0$ for which $x^0 - \epsilon s \in K$, we have $0 \leq s^T (x^0 - \epsilon s) = -\epsilon s^T s$ which implies that $s = 0$. But then $u = 0$ since we can assume without

loss of generality that A has full row rank, and then $(u, s, \theta) = 0$, which is a contradiction. Thus $\theta > 0$ and the proof is complete. ■

A *Slater point* for CD is a point (u^0, s^0) that satisfies:

$$A^T u^0 + s^0 = c \quad \text{and} \quad s^0 \in \text{int}K^* .$$

Corollary 3 *If CD has a Slater point, then $v^* = z^*$ and the primal attains its optimum.*

Corollary 4 *If CP and CD each have a Slater point, then $v^* = z^*$ and the primal and dual problems attain their optima.*

14 Duality Theory Exercises

1. Consider the problem

$$\begin{aligned} \text{(P)} \quad & \min_x f(x) = -\sqrt{x} \\ \text{s.t.} \quad & g(x) = x \leq 0 \\ & x \in X = [0, \infty) . \end{aligned}$$

Formulate the dual of this problem.

2. Consider the problem

$$\begin{aligned} \text{(P)} \quad & \min_x c^T x \\ \text{s.t.} \quad & b - Ax \leq 0 \\ & x \geq 0 . \end{aligned}$$

- a. Formulate a dual based on $g(x) = b - Ax$, $X = \{x \in \mathbb{R}^n : x \geq 0\}$.

- b.** Formulate a dual based on $g(x) = (\bar{g}(x), \tilde{g}(x)) = (b - Ax, -x)$ and $X = \mathbb{R}^n$.

3. Consider the two equivalent problems

$$\begin{array}{ll} (P^1) & \min_x \|x\| \\ & b - Ax \leq 0 \\ & x \in \mathbb{R}^n \end{array} \quad \begin{array}{ll} (P^2) & \min_x \frac{1}{2}x^T x \\ & b - Ax \leq 0 \\ & x \in \mathbb{R}^n . \end{array}$$

Derive the dual problems (D_1) and (D_2) of (P^1) and (P^2) . What is the relation between (D_1) and (D_2) ?

4. Let P be given as a function of the parameter y :

$$\begin{array}{ll} P_y & z^*(y) = \min \quad -x_1 + 5x_2 - 7x_3 \\ & \text{s.t.} \quad x_1 - 3x_2 + x_3 - y \leq 0, \\ & \quad \quad x \in X = \{x \in \mathbb{R}^3 : x_j = 0 \text{ or } 1, j = 1, 2, 3\}. \end{array}$$

Construct the dual D_y for all $y \in [0, 8]$. Graph the function $z^*(y)$ in (y, z) space. For what values of y is there a duality gap?

5. Consider the problem

$$\begin{array}{ll} (P) & \min_x c^T x + \frac{1}{2}x^T Qx \\ & \text{s.t.} \quad b - Ax \leq 0 . \end{array}$$

where Q is symmetric and positive semidefinite.

- a.** What is the dual?
b. Show that the dual of the dual is the primal.

6. Consider the program

$$\begin{array}{ll} (P) & \min_x f(x) \\ & \text{s.t.} \quad g_i(x) \leq 0, i = 1, \dots, m \\ & \quad \quad x \in X, \end{array}$$

where $f(\cdot)$ and $g_i(\cdot)$ are convex functions and the set X is a convex set. Consider the matrix below:

Property of P	Property of D					
	v^* finite attained $v^* = z^*$	v^* finite not attained $v^* = z^*$	v^* finite attained $v^* < z^*$	v^* finite not attained $v^* < z^*$	$v^* = +\infty$	$v^* = -\infty$
z^* finite P stable	1	5	9	13	17	21
z^* finite P not stable	2	6	10	14	18	22
$z^* = -\infty$	3	7	11	15	19	23
$z^* = +\infty$ (infeasible)	4	8	12	16	20	24

Prove by citing relevant theorems from the notes, etc., that the following cases cannot occur: 2, 3, 4, 5, 7, 8, 9, 11, 13, 15, 17, 18, 19, and 21.

7. Let $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$, $f(x, y) : (X \times Y) \rightarrow \mathbb{R}$. A *saddlepoint* (\bar{x}, \bar{y}) of $f(x, y)$ is a point $(\bar{x}, \bar{y}) \in X \times Y$ such that

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq f(x, \bar{y}),$$

for any $x \in X$ and $y \in Y$. Prove the equivalence of the following three results:

- (a) There exists a saddlepoint (\bar{x}, \bar{y}) of $f(x, y)$.
- (b) There exists $\bar{x} \in X, \bar{y} \in Y$ for which $\max_{y \in Y} f(\bar{x}, y) = \min_{x \in X} f(x, \bar{y})$.
- (c) $\min_{x \in X} (\max_{y \in Y} f(x, y)) = \max_{y \in Y} (\min_{x \in X} f(x, y))$.

8. Consider the logarithmic barrier problem:

$$P(\theta) : \text{minimize}_x \quad c^T x - \theta \sum_{j=1}^n \ln(x_j)$$

$$\text{s.t.} \quad Ax = b \\ x > 0.$$

Compute a dual of $P(\theta)$ by dualizing on the constraints “ $Ax = b$.” What is the relationship between this dual and the dual of the original linear program (without the logarithmic barrier function)?

9. Consider the program

$$P(\gamma) : \text{ minimize}_x \quad - \sum_{j=1}^n \ln(x_j)$$

$$\text{ s.t.} \quad \begin{aligned} Ax &= b \\ c^T x &= \gamma \\ x &> 0 . \end{aligned}$$

Compute a dual of $P(\gamma)$ by dualizing on the constraints $Ax = b$ and $c^T x = \gamma$. What is the relationship between the optimality conditions of $P(\theta)$ in the previous exercise and $P(\gamma)$?

Consider the program

$$P(\delta) : \text{ minimize}_x \quad - \sum_{j=1}^n \ln(x_j) - \ln(\delta - c^T x)$$

$$\text{ s.t.} \quad \begin{aligned} Ax &= b \\ c^T x &< \delta \\ x &> 0 . \end{aligned}$$

Compute a dual of $P(\delta)$ by dualizing on the constraints $Ax = b$. What is the relationship between the optimality conditions of $P(\delta)$ and $P(\theta)$ and $P(\gamma)$ of the previous two exercises?

10. Consider the conic form problem CP of convex optimization

$$\text{CP} : \quad z^* = \text{minimum}_x \quad c^T x$$

$$\text{ s.t.} \quad \begin{aligned} Ax &= b \\ x &\in K , \end{aligned}$$

and the associated conic dual problem:

$$\text{CD} : \quad v^* = \text{maximum}_{u,s} \quad b^T u$$

$$\text{ s.t.} \quad \begin{aligned} A^T y + s &= c \\ s &\in K^* , \end{aligned}$$

where K is a closed convex cone. Show that “the dual of the dual is the primal,” that is, that the conic dual of CD is CP.

11. Prove Corollary 3, which asserts that the existence of Slater point for the conic dual problem guarantees strong duality and that the primal attains its optimum.
12. Consider the following “minimax” problems:

$$\min_{x \in X} \max_{y \in Y} \phi(x, y) \quad \text{and} \quad \max_{y \in Y} \min_{x \in X} \phi(x, y)$$

where X and Y are nonempty compact convex sets in \mathbb{R}^n and \mathbb{R}^m , respectively, and $\phi(x, y)$ is convex in x for fixed y , and is concave in y for fixed x .

- (a) Show that $\min_{x \in X} \max_{y \in Y} \phi(x, y) \geq \max_{y \in Y} \min_{x \in X} \phi(x, y)$ in the absence of any convexity/concavity assumptions on X , Y , and/or $\phi(\cdot, \cdot)$.
- (b) Show that $f(x) := \max_{y \in Y} \phi(x, y)$ is a convex function in x and that $g(y) := \min_{x \in X} \phi(x, y)$ is a concave function in y .
- (c) Use a separating hyperplane theorem to prove:

$$\min_{x \in X} \max_{y \in Y} \phi(x, y) = \max_{y \in Y} \min_{x \in X} \phi(x, y) .$$

13. Let X and Y be nonempty sets in \mathbb{R}^n , and let $f(\cdot), g(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$. Consider the following *conjugate functions* $f^*(\cdot)$ and $g^*(\cdot)$ defined as follows:

$$f^*(u) := \inf_{x \in X} \{f(x) - u^t x\} ,$$

and

$$g^*(u) := \sup_{x \in X} \{f(x) - u^t x\} .$$

- (a) Construct a geometric interpretation of $f^*(\cdot)$ and $g^*(\cdot)$.
- (b) Show that $f^*(\cdot)$ is a concave function on $X^* := \{u \mid f^*(u) > -\infty\}$, and $g^*(\cdot)$ is a convex function on $Y^* := \{u \mid g^*(u) < +\infty\}$.
- (c) Prove the following *weak duality theorem* between the *conjugate primal problem* $\inf\{f(x) - g(x) \mid x \in X \cap Y\}$ and the *conjugate dual problem* $\sup\{f^*(u) - g^*(u) \mid u \in X^* \cap Y^*\}$:

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

- (d) Now suppose that $f(\cdot)$ is a convex function, $g(\cdot)$ is a concave function, $\text{int}X \cap \text{int}Y \neq \emptyset$, and $\inf\{f(x) - g(x) \mid x \in X \cap Y\}$ is finite. Show that equality in part (13c) holds true and that $\sup\{f^*(u) - g^*(u) \mid u \in X^* \cap Y^*\}$ is attained for some $u = u^*$.

- (e) Consider a standard inequality constrained nonlinear optimization problem using the following notation:

$$\begin{aligned} \text{OP : } \quad & \text{minimum}_x \quad \bar{f}(x) \\ & \text{s.t.} \quad \bar{g}_1(x) \leq 0, \\ & \quad \quad \quad \vdots \\ & \quad \quad \quad \bar{g}_m(x) \leq 0, \\ & \quad \quad \quad x \in \bar{X} . \end{aligned}$$

By suitable choices of $f(\cdot), g(\cdot), X$, and Y , formulate this problem as an instance of the conjugate primal problem $\inf\{f(x) - g(x) \mid x \in X \cap Y\}$. What is the form of the resulting conjugate dual problem $\sup\{f^*(u) - g^*(u) \mid u \in X^* \cap Y^*\}$?

14. Consider the following problem:

$$\begin{aligned} z^* = \min \quad & x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 4, \\ & -2x_1 - x_2 \leq 4 . \end{aligned}$$

- (a) Formulate the Lagrange dual of this problem by incorporating both constraints into the objective function via multipliers u_1, u_2 .
- (b) Compute the gradient of $L^*(u)$ at the point $\bar{u} = (1, 2)$.
- (c) Starting from $\bar{u} = (1, 2)$, perform one iteration of the steepest ascent method for the dual problem. In particular, solve the following problem where $\bar{d} = \nabla L^*(\bar{u})$:

$$\begin{aligned} \max_{\alpha} \quad & L^*(\bar{u} + \alpha \bar{d}) \\ \text{s.t.} \quad & \bar{u} + \alpha \bar{d} \geq 0 , \\ & \alpha \geq 0 . \end{aligned}$$