# Applied Lagrange Duality for Constrained Optimization

Robert M. Freund

February 10, 2004

# 1 Overview

- The Practical Importance of Duality

- Review of Convexity

- A Separating Hyperplane Theorem

- Definition of the Dual Problem

- Steps in the Construction of the Dual Problem

- Examples of Dual Constructions

- The Dual is a Concave Maximization Problem

- Weak Duality

- The Column Geometry of the Primal and Dual Problems

- Strong Duality

- Duality Strategies

- Illustration of Lagrange Duality in Discrete Optimization

# 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  Review of Convexity

## 3.1  Local and Global Optima of a Function

The *ball* centered at $\bar{x}$ with radius $\epsilon$ is the set:

$$B(\bar{x}, \epsilon) := \{x | \|x - \bar{x}\| \leq \epsilon\}.$$

Consider the following optimization problem over the set $\mathcal{F}$:

$$P: \quad \min_x \text{ or } \max_x \quad f(x)$$

$$\text{s.t.} \qquad x \quad \in \quad \mathcal{F}$$

We have the following definitions of local/global, strict/non-strict minima/maxima.

**Definition 3.1** $x \in \mathcal{F}$ *is a* local minimum *of P if there exists $\epsilon > 0$ such that $f(x) \leq f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$.*

**Definition 3.2** $x \in \mathcal{F}$ *is a* global minimum *of P if $f(x) \leq f(y)$ for all $y \in \mathcal{F}$.*

**Definition 3.3** $x \in \mathcal{F}$ *is a* strict local minimum *of P if there exists $\epsilon > 0$ such that $f(x) < f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$, $y \neq x$.*

**Definition 3.4** $x \in \mathcal{F}$ *is a* strict global minimum *of P if $f(x) < f(y)$ for all $y \in \mathcal{F}$, $y \neq x$.*

**Definition 3.5** $x \in \mathcal{F}$ *is a* local maximum *of P if there exists $\epsilon > 0$ such that $f(x) \geq f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$.*

**Definition 3.6** $x \in \mathcal{F}$ *is a* global maximum *of P if $f(x) \geq f(y)$ for all $y \in \mathcal{F}$.*

**Definition 3.7** $x \in \mathcal{F}$ *is a* strict local maximum *of P if there exists $\epsilon > 0$ such that $f(x) > f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$, $y \neq x$.*

**Definition 3.8** $x \in \mathcal{F}$ *is a* strict global maximum *of P if $f(x) > f(y)$ for all $y \in \mathcal{F}$, $y \neq x$.*

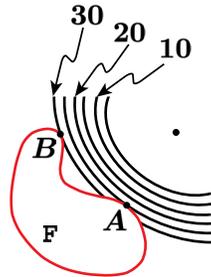The phenomenon of local versus global optima is illustrated in Figure 1.

## 3.2 Convex Sets and Functions

Convex sets and convex functions play an extremely important role in the study of optimization models. We start with the definition of a convex set:

**Definition 3.9** *A subset $S \subset \Re^n$ is a* convex set *if*

$$x, y \in S \Rightarrow \lambda x + (1 - \lambda)y \in S$$

*for any $\lambda \in [0, 1]$.*

**A** is a global minimum
**B** is a local minimum

Figure 1: Illustration of local versus global optima.
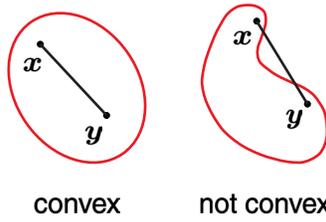


convex  not convex

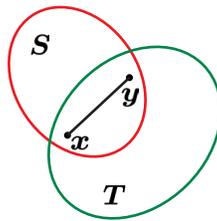Figure 2: Illustration of convex and non-convex sets.



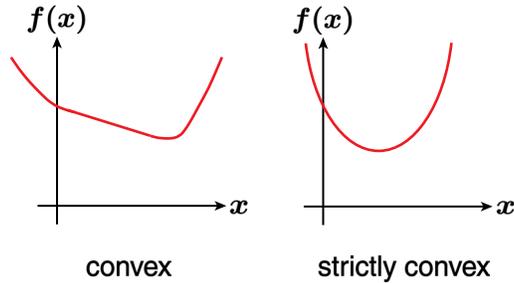Figure 3: Illustration of the intersection of convex sets.

Figure 4: Illustration of convex and strictly convex functions.

Figure 2 shows a convex set and a non-convex set.

**Proposition 3.1** *If $S, T$ are convex sets, then $S \cap T$ is a convex set.*

This proposition is illustrated in Figure 3.

**Proposition 3.2** *The intersection of any collection of convex sets is a convex set.*

We now turn our attention to convex functions, defined below.

**Definition 3.10** *A function $f(x)$ is a* convex function *if*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

*for all $x$ and $y$ and for all $\lambda \in [0, 1]$.*

**Definition 3.11** *A function $f(x)$ is a* strictly convex function *if*

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$$

*for all $x$ and $y$, $x \neq y$, and for all $\lambda \in (0, 1)$.*

Figure 4 illustrates convex and strictly convex functions.

Now consider the following optimization problem, where the feasible region is simply described as the set $\mathcal{F}$:

6

$$P: \quad \text{minimize}_x \quad f(x)$$

$$\text{s.t.}$$
$$x \quad \in \quad \mathcal{F}$$

**Theorem 3.1** *Suppose that $\mathcal{F}$ is a convex set, $f : \mathcal{F} \to \Re$ is a convex function, and $\bar{x}$ is a local minimum of $P$. Then $\bar{x}$ is a global minimum of $f$ over $\mathcal{F}$.*

**Proof:** Suppose $\bar{x}$ is not a global minimum, i.e., there exists $y \in \mathcal{F}$ for which $f(y) < f(\bar{x})$. Let $y(\lambda) = \lambda \bar{x} + (1-\lambda)y$, which is a convex combination of $\bar{x}$ and $y$ for $\lambda \in [0,1]$ (and therefore, $y(\lambda) \in \mathcal{F}$ for $\lambda \in [0,1]$). Note that $y(\lambda) \to \bar{x}$ as $\lambda \to 1$.

From the convexity of $f(\cdot)$,

$$f(y(\lambda)) = f(\lambda \bar{x} + (1-\lambda)y) \le \lambda f(\bar{x}) + (1-\lambda)f(y) < \lambda f(\bar{x}) + (1-\lambda)f(\bar{x}) = f(\bar{x})$$

for all $\lambda \in (0,1)$. Therefore, $f(y(\lambda)) < f(\bar{x})$ for all $\lambda \in (0,1)$, and so $\bar{x}$ is not a local minimum, resulting in a contradiction.
**q.e.d.**

Some examples of convex functions of one variable are:

- $f(x) = ax + b$

- $f(x) = x^2 + bx + c$

- $f(x) = |x|$

- $f(x) = -\ln(x)$ for $x > 0$

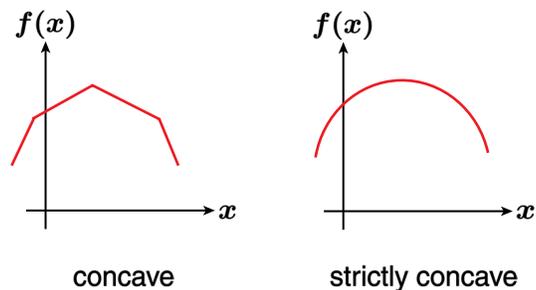- $f(x) = \frac{1}{x}$ for $x > 0$

- $f(x) = e^x$

Figure 5: Illustration of concave and strictly concave functions.

## 3.3  Concave Functions and Maximization

The "opposite" of a convex function is a concave function, defined below:

**Definition 3.12** *A function $f(x)$ is a* concave function *if*

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$$

*for all $x$ and $y$ and for all $\lambda \in [0, 1]$.*

**Definition 3.13** *A function $f(x)$ is a* strictly concave function *if*

$$f(\lambda x + (1 - \lambda)y) > \lambda f(x) + (1 - \lambda)f(y)$$

*for all $x$ and $y$, $x \neq y$, and for all $\lambda \in (0, 1)$.*

Figure 5 illustrates concave and strictly concave functions.

Now consider the maximization problem $P$:

$$P: \quad \text{maximize}_x \quad f(x)$$

$$\text{s.t.}$$
$$x \quad \in \quad \mathcal{F}$$

**Theorem 3.2** *Suppose that $\mathcal{F}$ is a convex set, $f : \mathcal{F} \to \Re$ is a concave function, and $\bar{x}$ is a local maximum of $P$. Then $\bar{x}$ is a global maximum of $f$ over $\mathcal{F}$.*

8

## 3.4  Convex Programs

Now consider the optimization problem $P$ which might be either a minimization problem or a maximization problem:

$$P: \quad \min_x \text{ or } \max_x \quad f(x)$$

$$\text{s.t.}$$
$$x \quad \in \quad \mathcal{F}$$

**Definition 3.14** *We call $P$ a* convex program *if*

- *$\mathcal{F}$ is a convex set, and*

- *we are minimizing and $f(x)$ is a convex function, or we are maximizing and $f(x)$ is a concave function.*

Then from the above theorems, every local optima of the objective function of a convex program is a global optima of the objective function.

The class of convex programs is the class of "well-behaved" optimization problems.

## 3.5  Linear Functions, Convexity, and Concavity

**Proposition 3.3** *A linear function $f(x) = a^T x + b$ is both convex and concave.*

**Proposition 3.4** *If $f(x)$ is both convex and concave, then $f(x)$ is a linear function.*

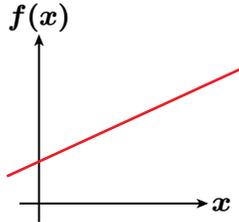These properties are illustrated in Figure 6.

Figure 6: A linear function is convex and concave.

# 4    A Separating Hyperplane Theorem

The main tool that is used in developing duality, analyzing dual problems, etc., is the "separating hyperplane" theorem for convex sets. This theorem states that a point outside of a convex set can be separated from the set by hyperplane. The theorem is intuitively obvious, and is very easy to prove.

**Theorem 4.1  Strong Separating Hyperplane Theorem:** *Suppose that $S$ is a convex set in $\Re^n$, and that we are given a point $\bar{x} \notin S$. Then there exists a vector $u \neq 0$ and a scalar $\alpha$ for which the following hold:*

- $u^T x > \alpha$ *for all $x \in S$*

- $u^T \bar{x} < \alpha$

This is illustrated in Figure 7.

A weaker version of this theorem is as follows:

**Theorem 4.2  Weak Separating Hyperplane Theorem:** *Suppose that $S$ is a convex set in $\Re^n$, and that we are given a point $\bar{x} \notin S$ or $\bar{x} \in \partial S$. Then there exists a vector $u \neq 0$ and a scalar $\alpha$ for which the following hold:*

- $u^T x \geq \alpha$ *for all $x \in S$*

- $u^T \bar{x} \leq \alpha$
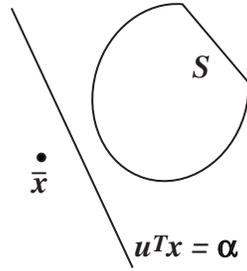
Figure 7: Illustration of strong separation of a point from a convex set.
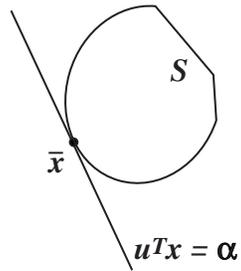


Figure 8: Illustration of weak separation of a point from a convex set.

This is illustrated in Figure 8.

Now consider the following simple primal and dual problems, that will be prototypical of the geometry of duality that we will develop in this lecture.

$$\text{Primal}: \quad \text{minimum}_x \quad \text{height of } x$$

$$\text{s.t.}$$

$$
\begin{aligned}
x &\in S, \\
x &\in L
\end{aligned}
$$

$$\text{Dual}: \quad \text{maximum}_\pi \quad \text{intercept of line } \pi \text{ with } L$$

$$\text{s.t.}$$
$$\text{line } \pi \text{ lies below } S$$

This is illustrated in Figure 9.



Figure 9: Illustration intersecting line dual.

# 5 Definition of the Dual Problem

Recall the basic constrained optimization model:

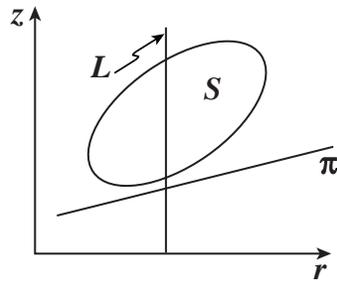$$\text{OP}: \quad \text{minimum}_x \quad f(x)$$

$$\text{s.t.}$$

$$
\begin{array}{ccc}
g_1(x) & \leq & 0, \\
\cdot & = & \\
\cdot & \geq & \\
\vdots & & \\
g_m(x) & \leq & 0, \\
\end{array}
$$

$$x \in P,$$

In this model, we have $f(x) : \Re^n \mapsto \Re$ and $g_i(x) : \Re^n \mapsto \Re, i = 1, \ldots, m$.

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:

$$\text{OP}: \quad z^* = \text{minimum}_x \quad f(x)$$

$$\text{s.t.}$$

$$
\begin{array}{ccc}
g_1(x) & \leq & 0, \\
& \vdots & \\
g_m(x) & \leq & 0, \\
\end{array}
$$

$$x \in P,$$

Here $P$ could be any set, such as:

- $P = \Re^n$

- $P = \{x \mid x \geq 0\}$

13

- $P = \{x \mid g_i(x) \leq 0, i = m+1, \ldots, m+k\}$

- $P = \{x \mid x \in Z_+^n\}$

- $P = \{x \mid Ax \leq b\}$

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 $i = 1, \ldots, m$.

We then solve the presumably easier problem:

$$L^*(u) := \quad \text{minimum}_x \quad f(x) + u^T g(x)$$

$$\text{s.t.} \qquad x \in P$$

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:

$$\text{D}: \quad v^* = \text{maximum}_u \quad L^*(u)$$

$$\text{s.t.} \qquad u \geq 0$$

# 6 Steps in the Construction of the Dual Problem

We start with the primal problem:

$$\text{OP}: \quad z^* = \text{minimum}_x \quad f(x)$$

$$\text{s.t.}$$

$$g_i(x) \quad \leq \quad 0, \quad i = 1, \ldots, m,$$

$$x \in P,$$

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:

$$L^*(u) := \quad \text{minimum}_x \quad f(x) + u^T g(x)$$

$$\text{s.t.} \qquad x \in P$$

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

$$\text{D}: \quad v^* = \text{maximum}_u \quad L^*(u)$$

$$\text{s.t.} \qquad u \geq 0$$

# 7 Examples of Dual Constructions of Optimization Problems

## 7.1 The Dual of a Linear Problem

Consider the linear optimization problem:

$$\text{LP}: \quad \text{minimum}_x \quad c^T x$$

$$\text{s.t.} \quad Ax \geq b$$

What is the dual of this problem?

## 7.2 The Dual of a Binary Integer Problem

Consider the binary integer problem:

$$\text{IP}: \quad \text{minimum}_x \quad c^T x$$

$$\text{s.t.} \quad Ax \geq b$$

$$x_j \in \{0, 1\}, \; j = 1, \ldots, n \; .$$

What is the dual of this problem?

## 7.3 The Dual of a Log-Barrier Problem

Consider the following logarithmic barrier problem:

$$\text{BP}: \quad \text{minimum}_{x_1, x_2, x_3} \quad 5x_1 + 7x_2 - 4x_3 - \sum_{j=1}^{3} \ln(x_j)$$

$$\text{s.t.} \quad x_1 + 3x_2 + 12x_3 = 37$$

$$x_1 > 0, \; x_2 > 0, \; x_3 > 0 \; .$$

What is the dual of this problem?

## 7.4 The Dual of a Quadratic Problem

Consider the quadratic optimization problem:

$$\text{QP}: \quad \text{minimum}_x \quad \tfrac{1}{2}x^T Q x + c^T x$$

$$\text{s.t.} \quad \quad Ax \geq b$$

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

What is the dual of this problem?

## 7.5 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:

$$\text{OP}: \quad \text{minimum}_x \quad f(x)$$

$$\text{s.t.}$$
$$g_i(x) \quad \leq \quad 0, \quad i \in L$$

$$g_i(x) \quad \geq \quad 0, \quad i \in G$$

$$g_i(x) \quad = \quad 0, \quad i \in E$$

$$x \in P,$$

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 $L^*(u)$:

$$L^*(u) := \quad \text{minimum}_x \quad f(x) + u^T g(x)$$

$$\text{s.t.} \qquad x \in P$$

The dual problem is then defined to be:

$$\text{D}: \quad \text{maximum}_u \qquad L^*(u)$$

$$\text{s.t.}$$

$$u_i \geq 0, \quad i \in L$$

$$u_i \leq 0, \quad i \in G$$

$$u_i \overset{>}{<}, \quad i \in E$$

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.

# 8    The Dual is a Concave Maximization Problem

We start with the primal problem:

$$\text{OP}: \quad \text{minimum}_x \quad f(x)$$

$$\text{s.t.}$$

$$g_i(x) \quad \leq \quad 0, \quad i = 1, \ldots, m$$

$$x \in P,$$

We create the Lagrangian:

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

and the dual function:

$$L^*(u) := \quad \text{minimum}_x \quad f(x) + u^T g(x)$$

$$\text{s.t.} \qquad x \in P$$

The dual problem then is:

$$\text{D}: \quad \text{maximum}_u \quad L^*(u)$$

$$\text{s.t.} \qquad u \geq 0$$

**Theorem 8.1** *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

$$
\begin{aligned}
L^*(u) &= \min_{x \in P} f(x) + u^T g(x) \\[2mm]
&= \min_{x \in P} \lambda \left[ f(x) + u_1^T g(x) \right] + (1 - \lambda) \left[ f(x) + u_2^T g(x) \right] \\[2mm]
&\geq \lambda \left[ \min_{x \in P} f(x) + u_1^T g(x) \right] + (1 - \lambda) \left[ \min_{x \in P} (f(x) + u_2^T g(x)) \right] \\[2mm]
&= \lambda L^*(u_1) + (1 - \lambda) L^*(u_2) \quad .
\end{aligned}
$$

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

## 9  Weak Duality

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

$$
\begin{aligned}
\text{OP}: \quad z^* = \text{minimum}_x \quad & f(x) \\[2mm]
\text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\[2mm]
& x \in P,
\end{aligned}
$$

$$
\begin{aligned}
\text{D}: \quad v^* = \text{maximum}_u \quad & L^*(u) \\[2mm]
\text{s.t.} \quad & u \geq 0
\end{aligned}
$$

**Theorem 9.1 Weak Duality Theorem:** *If $\bar{x}$ is feasible for OP and $\bar{u}$ is feasible for D, then*

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

20

*In particular,*

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

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

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

Therefore $z^* \geq v^*$.
**q.e.d.**

# 10 The Column Geometry of the Primal and Dual Problems

For this section only, we will assume that our constraints are all equality constraints, and so our problems are:

$$\text{OP}: \quad z^* = \text{minimum}_x \quad f(x)$$

$$\text{s.t.} \qquad g_i(x) \quad = \quad 0, \quad i = 1, \ldots, m$$

$$x \in P,$$

$$\text{D}: \quad v^* = \text{maximum}_u \quad L^*(u)$$

$$\text{s.t.} \qquad u \gtrless 0$$

where

$$L^*(u) := \quad \text{minimum}_x \quad f(x) + u^T g(x)$$

$$\text{s.t.} \qquad x \in P$$

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

$$\begin{pmatrix} r \\ z \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \\ z \end{pmatrix} = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \\ f(x) \end{pmatrix} = \begin{pmatrix} g(x) \\ f(x) \end{pmatrix} \ .$$

We can think of each of these arrays as an array of resources and cost in $\Re^{m+1}$. If we plot all of these arrays over all values of $x \in P$, we will obtain a region in $\Re^{m+1}$, which we will call the region $S$:

$$S := \left\{ (r,z) \in \Re^{m+1} \mid (r,z) = (g(x), f(x)) \text{ for some } x \in P \right\} \ .$$

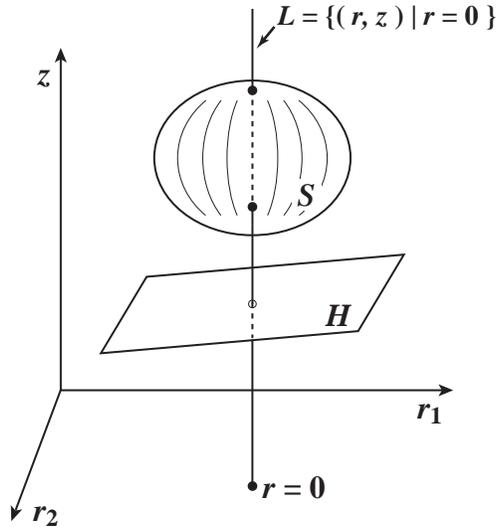This region is illustrated in Figure 10.



Figure 10: The column geometry of the primal and dual problem.

Where is the feasible region of the primal in this problem? Well, we are in the wrong space to observe this. But notice that the figure contains the origin and a vertical line $L$ going up from the origin. The points on this vertical line have resource utilization of 0, and their costs are their heights on the vertical line. Therefore, every point in the intersection of $S$ and $L$ corresponds to a feasible solution, and the feasible solution with the lowest cost corresponds to the lowest point on the intersection of $S$ and $L$. We re-state this for emphasis:

*Every point in the intersection of $S$ and $L$ corresponds to a feasible solution of OP, and the feasible solution with the lowest cost corresponds to the lowest point in the intersection of $S$ and $L$. This lowest value is exactly the value of the primal problem, namely $z^*$.*

In Figure 10, notice that there is also drawn a hyperplane $H$, and that this hyperplane $H$ lies below the region $S$. Suppose that we try to find that hyperplane $H$ lying below $S$ whose intersection with $L$ is as large as possible. Let us formulate this problem mathematically. We will let

$$H = H_{u,\alpha} = \left\{ (r, z) \in \Re^{m+1} \mid z + u^T r = \alpha \right\} .$$

The line $L$ is:

$$L = \left\{ (r, z) \in \Re^{m+1} \mid r = 0 \right\} .$$

The intersection of $H_{u,\alpha}$ with $L$ is:

$$H \cap L$$

and it occurs at the point

$$(r, z) = (0, \alpha) .$$

Then the problem of determining the hyperplane $H_{u,\alpha}$ lying below $S$ whose intersection with $L$ is the highest is:

$$\text{maximum}_{u,\alpha} \qquad \alpha$$

$$\text{s.t.} \qquad H_{u,\alpha} \text{ lies below } S$$

$$= \text{maximum}_{u,\alpha} \qquad \alpha$$

$$\text{s.t.} \qquad z + u^T r \qquad \geq \quad \alpha, \quad \text{for all } (r,z) \in S$$

$$= \text{maximum}_{u,\alpha} \qquad \alpha$$

$$\text{s.t.} \qquad f(x) + u^T g(x) \qquad \geq \quad \alpha, \quad \text{for all } x \in P$$

$$= \text{maximum}_{u,\alpha} \qquad \alpha$$

$$\text{s.t.} \qquad L^*(u) \qquad \geq \quad \alpha,$$

$$= \text{maximum}_{u} \qquad L^*(u)$$

and this last expression is exactly the dual problem (for the special case of this section). Therefore:

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

# 11   Strong Duality

When the set $S$ is a convex set, then generally speaking, we would expect $z^* = v^*$, that is, the primal and the dual have identical optimal objective function values. When the set $S$ is not convex, we would not necessarily expect this to happen, see the illustrations in Figures 11 and 12.
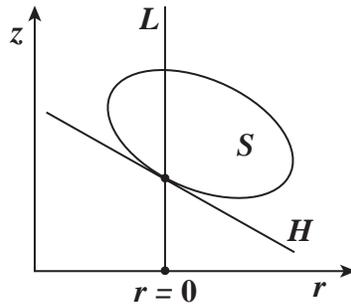


Figure 11: The column geometry of the dual problem when $S$ is convex.
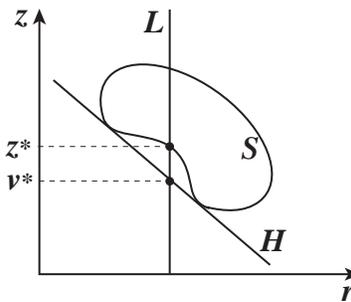


Figure 12: The column geometry of the dual problem when $S$ is not convex.

As it turns out, convexity is the key to strong duality. When the problem OP is a convex program, then strong duality will hold in general, as the following theorem states:

25

**Theorem 11.1 Strong Duality Theorem:** *Suppose that OP is a convex optimization problem, that is, all constraints of OP are of the form $g_i(x) \leq 0, i = 1, \ldots, m$, $f(x)$ as well as $g_i(x)$ are convex functions, $i = 1, \ldots, m$, and P is a convex set. Then under very mild additional conditions, $z^* = v^*$.*

# 12   Duality Strategies

## 12.1   Dualizing "Bad" Constraints

Suppose we wish to solve:

$$\text{OP}: \quad \text{minimum}_x \quad c^T x$$

$$\text{s.t.} \quad Ax \leq b$$

$$Nx \leq g.$$

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

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

and re-write OP as:

$$\text{OP}: \quad \text{minimum}_x \quad c^T x$$

$$\text{s.t.} \quad Ax \leq b$$

$$x \in P.$$

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:

$$L^*(u) := \quad \text{minimum}_x \quad -u^T b + (c^T + u^T A)x$$

$$\text{s.t.} \qquad\qquad x \qquad\qquad \in \quad P \ .$$

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:

$$\text{D}: \quad \text{maximum}_u \quad L^*(u)$$

$$\text{s.t.} \qquad u \geq 0 \ .$$

## 12.2   Dualizing A Large Problem into Many Small Problems

Suppose we wish to solve:

$$\text{OP}: \quad \text{minimum}_{x^1, x^2} \quad (c^1)^T x^1 \quad +(c^2)^T x^2$$

$$\text{s.t.} \qquad B^1 x^1 \quad +B^2 x^2 \quad \leq \quad d$$

$$A^1 x^1 \qquad\qquad \leq \quad b^1$$

$$A^2 x^2 \quad \leq \quad b^2$$

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:

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

and re-write OP as:

$$\text{OP}: \quad \text{minimum}_{x^1, x^2} \quad (c^1)^T x^1 \quad +(c^2)^T x^2$$

$$\text{s.t.} \quad B^1 x^1 \quad +B^2 x^2 \quad \leq \quad d$$

$$(x^1, x^2) \quad \in \quad P$$

The Lagrangian is:

$$L(x, u) \;=\; (c^1)^T x^1 + (c^2)^T x^2 + u^T (B^1 x^1 + B^2 x^2 - d)$$

$$=\; -u^T d + ((c^1)^T + u^T B^1) x^1 + ((c^2)^T + u^T B^2) x^2 \; ,$$

and the dual function is:

$$L^*(u) = \quad \text{minimum}_{x^1, x^2} \quad -u^T d + ((c^1)^T + u^T B^1) x^1 + ((c^2)^T + u^T B^2) x^2$$

$$\text{s.t.} \quad (x^1, x^2) \quad \in \quad P$$

which can be re-written as:

$$L^*(u) = \quad -u^T d$$

$$+ \; \text{minimum}_{A^1 x^1 \leq b^1} \quad ((c^1)^T + u^T B^1) x^1$$

$$+ \; \text{minimum}_{A^2 x^2 \leq b^2} \quad ((c^2)^T + u^T B^2) x^2$$

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:

$$\text{D}: \quad \text{maximum}_u \quad L^*(u)$$

$$\text{s.t.} \quad u \geq 0$$

28

# 13 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 13.
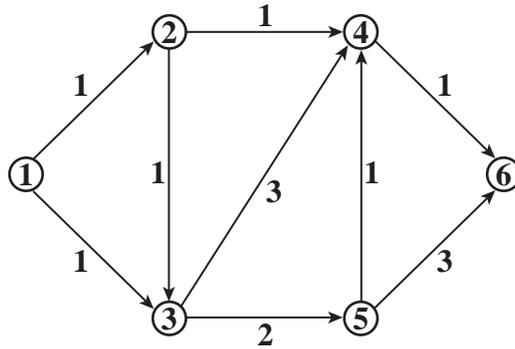


Figure 13: A constrained shortest path problem.

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.

Now suppose that we have available an efficient algorithm for solving

| 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.

(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:

$CSPP$ :

$$z^* = \min_{x_{ij}} \quad x_{12} + x_{13} + x_{23} + x_{24} + 3x_{34} + 2x_{35} + x_{54} + x_{46} + 3x_{56}$$

$$\text{s.t.} \quad x_{12} + x_{13} + x_{23} + x_{24} + x_{34} + x_{35} + x_{54} + x_{46} + x_{56} = 4$$

$$x_{12} + x_{13} = 1$$

$$x_{12} - x_{23} - x_{24} = 0$$

$$x_{13} + x_{23} - x_{34} - x_{35} = 0$$

$$x_{24} + x_{34} + x_{54} - x_{46} = 0$$

$$x_{35} - x_{54} - x_{56} = 0$$

$$x_{46} + x_{56} = 1$$

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

30

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:

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

s.t.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_{12}$ | $+x_{13}$ | | | | | | | | | $=$ | $1$ |
| $x_{12}$ | | $-x_{23}$ | $-x_{24}$ | | | | | | | $=$ | $0$ |
| | $x_{13}$ | $+x_{23}$ | | $-x_{34}$ | $-x_{35}$ | | | | | $=$ | $0$ |
| | | | $x_{24}$ | $+x_{34}$ | | $+x_{54}$ | $-x_{46}$ | | | $=$ | $0$ |
| | | | | | $x_{35}$ | $-x_{54}$ | | $-x_{56}$ | | $=$ | $0$ |
| | | | | | | | $x_{46}$ | $+x_{56}$ | | $=$ | $1$ |

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

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

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

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:

$$\text{D}: \quad v^* = \text{maximum}_u \quad L^*(u)$$

$$\text{s.t.} \qquad u \overset{<}{>} 0$$

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

$$
\begin{array}{rcl}
x_{12} \quad +x_{13} & = & 1 \\[2mm]
x_{12} \qquad -x_{23} \quad -x_{24} & = & 0 \\[2mm]
x_{13} \quad +x_{23} \qquad -x_{34} \quad -x_{35} & = & 0 \\[2mm]
x_{24} \quad +x_{34} \qquad +x_{54} \quad -x_{46} & = & 0 \\[2mm]
x_{35} \quad -x_{54} \qquad -x_{56} & = & 0 \\[2mm]
x_{46} \quad +x_{56} & = & 1
\end{array}
$$

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

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 14.

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 14 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
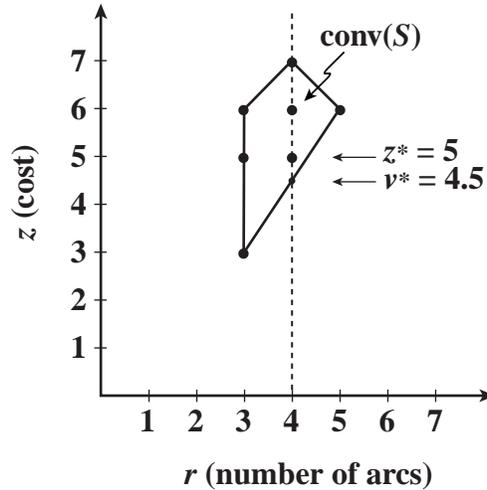
Figure 14: Resources and costs.

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

34

applications.

To conclude this discussion, note from Figure 14 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.