BUDT 758P: Decision Analytics
A guide to duality

Every time we formulate and solve a linear program, it is possible to create a second linear program, known as the *dual problem*, that is somehow related to the original LP. The dual problem often plays an auxiliary role. This second LP does not directly address the problem we set out to solve, but it often addresses closely related issues, or allows us to look at the same problem from a different angle.

Example 1. Consider the following simple problem:

"A California grower has a 50-acre farm on which to plant strawberries and tomatoes. The grower has available 300 hours of labor per week and 800 tons of fertilizer, and he has contracted for shipping space for a maximum of 26 acres' worth of strawberries and 37 acres' worth of tomatoes. An acre of strawberries requires 10 hours of labor and 8 tons of fertilizer, whereas an acre of tomatoes requires 3 hours of labor and 20 tons of fertilizer. The profit from an acre of strawberries is \$400, and the profit from an acre of tomatoes is \$300. The farmer wants to know the number of acres of strawberries and tomatoes to plan to maximize profit."

Letting x_1 be the number of acres of strawberries to plan, and x_2 be the number of acres of tomatoes, we formulate the following LP (let us refer to it by the name **original LP** or **grower's LP**):

$$\begin{array}{rrcrcl}
\max & 400x_1 & + & 300x_2 & & \\
s.t. & x_1 & + & x_2 & \leq & 50 \\
 & 10x_1 & + & 3x_2 & \leq & 300 \\
 & 8x_1 & + & 20x_2 & \leq & 800 \\
 & x_1 & & & \leq & 26 \\
 & & & x_2 & \leq & 37 \\
 & & & x_1, x_2 & \geq & 0
\end{array}$$

Using Excel, we can find the optimal solution $(x_1^*, x_2^*) = (21.429, 28.571)$ and the optimal value $Z^* = 17,142.86$.

We will now consider a different, but related problem. Instead of the grower's point of view, we will consider the problem faced by a potential buyer who wishes to make an offer to the grower to buy the entire farm. We assume that the value of the farm depends only on the five assets mentioned in the original problem: land, labor, fertilizer, and two types of shipping space.

The buyer makes a separate price offer for each individual asset. We can now define five quantities:

y_1 : price offered per acre of land
y_2 : price offered per hour of labor
y_3 : price offered per ton of fertilizer
y_4 : price offered per acre of shipping space for strawberries
y_5 : price offered per acre of shipping space for tomatoes

If these are the price offers for each asset, the total amount of the buyer's offer is

$$50y_1 + 300y_2 + 800y_3 + 26y_4 + 37y_5.$$

Since the grower owns 50 acres of land, and we offer a price per acre, our total cost of buying all the land on the farm is $50y_1$. We add this to the cost of buying the other four assets.

As the buyer, we would like to pay as little as possible for the five assets. Thus, the buyer has to solve a problem with the objective

$$\min 50y_1 + 300y_2 + 800y_3 + 26y_4 + 37y_5.$$

However, there are certain constraints that have to be satisfied by our price offer. We would like to pay as little possible, but our offer should be large enough that the grower would be willing to accept it. For example, the grower would probably not be willing to give away the farm for free.

More specifically, the constraints in the buyer's purchase problem should reflect the economics of the grower's production problem. When we buy the grower's assets, we should offer at least as much money as the grower would be able to make with those assets on his/her own. This is ensured by two constraints

$$\begin{array}{ccccccccccc} y_1 & + & 10y_2 & + & 8y_3 & + & y_4 & & & \geq & 400, \\ y_1 & + & 3y_2 & + & 20y_3 & + & & + & y_5 & \geq & 300. \end{array}$$

To get an idea of where this is coming from, let us consider a particular combination (or "package") of resources: 1 acre of land, 10 hours of labor, 8 tons of fertilizer, and 1 acre of strawberry shipping space. The buyer offers a total of

$$y_1 + 10y_2 + 8y_3 + y_4$$

for this combination of resources. However, the grower could also use this exact package of resources to produce 1 acre of strawberries: the original problem tells us that one acre of strawberries requires exactly 10 hours of labor and 8 tons of fertilizer (and also 1 acre of land and shipping space). One acre of strawberries generates \$400 in profit. Thus, if

$$y_1 + 10y_2 + 8y_3 + y_4 < 400,$$

the buyer's offer will be too low: the grower will prefer to keep these resources, use them for production, and bring in \$400 in profit. In order for the grower to accept the buyer's offer, we must have

$$y_1 + 10y_2 + 8y_3 + y_4 \geq 400.$$

The second constraint is based on the same argument. We consider a package containing 1 acre of land, 3 hours of labor, 20 tons of fertilizer, and 1 acre of tomato shipping space. The grower has a choice between using this package to produce 1 acre of tomatoes for a profit of \$300, or selling it to the buyer for a profit of $y_1 + 3y_2 + 20y_3 + y_5$. The grower will prefer to sell only if

$$y_1 + 3y_2 + 20y_3 + y_5 \geq 300.$$

We also require

$$y_1, y_2, y_3, y_4, y_5 \geq 0,$$

since our price offer obviously has to be positive.

Thus, we have formulated the buyer's problem as another linear program (let us call it the **dual LP** or the **buyer's LP**):

$$\begin{array}{lrcrcrcrcrcl}
\min & 50y_1 & + & 300y_2 & + & 800y_3 & + & 26y_4 & + & 37y_5 & & \\
s.t. & y_1 & + & 10y_2 & + & 8y_3 & + & y_4 & & & \geq & 400 \\
 & y_1 & + & 3y_2 & + & 20y_3 & + & & + & y_5 & \geq & 300 \\
 & & & & & & & & & y_i & \geq & 0, \qquad i = 1,2,3,4,5.
\end{array}$$

This is exactly the *dual problem* of the profit-maximization LP solved by the grower. In this example, the dual provides us with *fair prices* for each of the grower's resources.

If we now solve the dual problem in Excel, we will obtain the answer $(y_1^*, y_2^*, y_3^*, y_4^*, y_5^*) = (257.14, 14.29, 0, 0, 0)$. These are the fair prices for the five assets that comprise the farm. Moreover, these five numbers are also the *shadow prices* of the five constraints in the grower's LP.

Recall that the optimal solution to the grower's LP is given by $(x_1^*, x_2^*) = (21.429, 28.571)$. If we plug these numbers into the constraints of the grower's LP, we will see that

$$\begin{aligned}
x_1^* + x_2^* &= 50 \\
10x_1^* + 3x_2^* &= 300,
\end{aligned}$$

however,

$$\begin{aligned}
8x_1^* + 20x_2^* &< 800 \\
x_1^* &< 26 \\
x_2^* &< 37.
\end{aligned}$$

In other words, the first two constraints are **binding**, and the other three are **non-binding**. We have already discussed that the shadow price of a non-binding constraint must be zero: if we are not using all the shipping space that we already have, we would not be willing to pay any money to obtain even more. The only two assets that are valuable to the grower are land and labor. The buyer's offer puts strictly positive prices on the valuable assets, and zero prices on the non-valuable ones. Essentially, the grower is saying, "If you buy all my land and labor for y_1^* and y_2^*, I'll throw in the fertilizer and shipping space for free."

Furthermore, the optimal value of the dual LP is $50y_1^* + 300y_2^* + 800y_3^* + 26y_4^* + 37y_5^* = 17,142.86$, exactly the same as the optimal value of the grower's problem. This makes economic sense: the smallest total offer we can make to the grower should be equal to the total profit the grower can make from the farm on his/her own.

Fact. The optimal solution of the dual LP gives the shadow prices for the original LP. Also, both LPs have the same optimal value.

We can also observe a kind of "mirror" relationship between the grower's LP and the buyer's LP. The grower maximizes; the buyer minimizes. The grower's objective coefficients become the right-hand side quantities for the buyer. Similarly, the grower's resource constraints become the buyer's objective coefficients. The grower's LP has 2 variables and 5 constraints; the buyer's LP has 5 variables and 2 constraints. The numbers are "flipped" from one problem to the other. For example, if we look at the constraints of the grower's LP, and we read the numbers in "column" corresponding to the variable x_1, these numbers are 1, 10, 8, 1, and 0. The same numbers can be found in the first constraint of the buyer's LP.

This general relationship always holds in duality. Starting with any LP (we call it the **primal**), we can follow some simple rules to create a dual LP, the mirror image of the original problem. Duality is symmetric: if you follow the rules and create the dual LP, and then you attempt to follow the rules again to create the dual of the dual, you will just end up with the primal LP that you originally started with. For that reason, it does not really matter which LP we call the "primal," and which we call the "dual." One of them will always be a maximization problem; the other will be a minimization problem.

The following table explains the relationship between these two problems. Note that the table does not specify which problem is the primal or dual.

Maximization	**Minimization**
Objective function	Right-hand side
Right-hand side	Objective function
Numbers in column	Numbers in row
Numbers in row	Numbers in column
$\leq$ constraint	Variable ≥ 0
$\geq$ constraint	Variable ≤ 0
$=$ constraint	"Free" or unrestricted variable
Variable ≥ 0	$\geq$ constraint
Variable ≤ 0	$\leq$ constraint
"Free" or unrestricted variable	$=$ constraint

The most important thing to remember about these rules is that **a variable in one problem always matches a constraint in the other**. The grower's problem has 2 variables and 5 constraints, so the dual must have 5 variables and 2 constraints. The sign of the inequality on a constraint is related to the sign of a variable in the other problem. A "free" variable is simply a variable that is not required to be non-negative or non-positive.

It seems like there are a lot of rules, but what we are really doing is taking all the numbers related to a variable in one problem, and putting them into a constraint in the other problem. With some practice, writing down a dual LP becomes very easy. Consider the following example:

Example 2. Consider a generic LP

$$\begin{array}{rrcrcrcl}
\min & 5x_1 & + & 3x_2 & + & -x_3 & & \\
s.t. & 4x_1 & + & 2x_2 & + & 7x_3 & \geq & 40 \\
 & 2x_1 & + & 6x_2 & + & 3x_3 & \leq & 30 \\
 & & & 8x_2 & + & 10x_3 & \leq & 120 \\
 & x_1 & + & 3x_2 & + & x_3 & = & 55 \\
 & & & & & x_1, x_2 & \geq & 0 \\
 & & & & & x_3 & \leq & 0.
\end{array}$$

We have not said anything about what this LP could possibly mean, but we already have everything we need to write down the dual. Our primal has 3 variables and 4 constraints (not counting the signs of individual variables). Our dual must therefore have 4 variables and 3 constraints. It is formulated as follows:

$$\begin{array}{rrcrcrcrcl}
\max & 40y_1 & + & 30y_2 & + & 120y_3 & + & 55y_4 & & \\
s.t. & 4y_1 & + & 2y_2 & & & + & y_4 & \leq & 5 \\
 & 2y_1 & + & 6y_2 & + & 8y_3 & + & 3y_4 & \leq & 3 \\
 & 7y_1 & + & 3y_2 & + & 10y_3 & + & y_4 & \geq & -1 \\
 & & & & & & & y_1 & \geq & 0 \\
 & & & & & & & y_2, y_3 & \leq & 0 \\
 & & & & & & & y_4 & free &
\end{array}$$

The signs of y_1, y_2, y_3, y_4 are determined **only** by the signs of the **constraints** in the primal problem. The signs of the constraint inequalities in the dual are determined only by the signs of the primal **variables**. Observe that, in this example, the dual was maximizing and the primal was minimizing.

In general, we can always write down a dual LP by following these rules, but the interpretation of the dual may vary greatly from problem to problem. In the example of the California grower, we interpreted the dual LP as finding fair prices for five assets. In the next example, the dual LP has a very different meaning.

Example 3. Consider the following problem:

"A pharmaceutical company uses two ingredients in its production process. Each ingredient contributes to the same three antibiotics, in different proportions. One gram of ingredient 1 costs \$80 and contributes 3 units of antibiotic A, 1 unit of antibiotic B, and 2 units of antibiotic C. One gram of ingredient 2 costs \$50 and contributes 1 unit of A, 1 unit of B, and 6 units of C. The company has contracted to supply 6, 4, and 12 units of antibiotics A, B, and C, respectively. We wish to find the minimum-cost production plan that will meet the above requirements."

Letting x_1 and x_2 be the purchase quantities of ingredients 1 and 2, it is easy to see that the problem is solved by the following LP:

$$\begin{array}{rrcrcl}
\min & 80x_1 & + & 50x_2 & & \\
s.t. & 3x_1 & + & x_2 & \geq & 6 \\
 & x_1 & + & x_2 & \geq & 4 \\
 & 2x_1 & + & 6x_2 & \geq & 12 \\
 & & & x_1, x_2 & \geq & 0
\end{array}$$

Without even needing to think about the problem, we can follow the rules to obtain the dual LP. Since the primal problem has 2 variables and 3 constraints, it must be the case that the dual has 3 variables and 2 constraints. Since the primal constraints represent the three antibiotics (for example, the first constraint ensures that the requirement for antibiotic A is met), we will denote the corresponding dual variables by y_A, y_B, and y_C. (Remember: one primal constraint is matched to one dual variable.)

The formulation is:

$$
\begin{array}{rrcrcrcl}
\max & 6y_A & + & 4y_B & + & 12y_C & & \\
s.t. & 3y_A & + & y_B & + & 2y_C & \leq & 80 \\
 & y_A & + & y_B & + & 6y_C & \leq & 50 \\
 & & & & & y_A, y_B, y_C & \geq & 0
\end{array}
$$

But what is the meaning of this problem? It is no longer possible to interpret this dual in terms of a buyer who wants to purchase the company. For one thing, the dual is a maximization problem; one would think that any prospective buyer would prefer to minimize purchase costs rather than maximize them. The problem must have a different interpretation.

To understand this problem, suppose that the pharmaceutical company solving the primal problem has the ability to outsource production to an outside contractor. That is, instead of purchasing the two ingredients and using its in-house production process to make the antibiotics, the company will instead simply buy the antibiotics from another manufacturer and use them to meet its contractual obligations. The dual LP is precisely the contractor's **optimal pricing problem**. The dual variables y_A, y_B, y_C represent per-unit prices that the contractor charges to the pharmaceutical company. The deal covers 6, 4, and 12 units of the antibiotics (this is how much the pharmaceutical company needs and wishes to purchase), so the objective function is the total revenue earned by the contractor.

The constraints ensure that the deal goes through. In order for the pharmaceutical company to choose to outsource, the cost of outsourcing must be lower than the cost of manufacturing the antibiotics in-house. Suppose that the first constraint is violated, that is, the dual variables take values such that

$$3y_A + y_B + 2y_C > 80.$$

This means that the total asking price for 3 units of A, 1 unit of B, and 2 units of C is above \$80. If this is the case, however, the pharmaceutical company can spend \$80 to buy one gram of ingredient 1. Then, the company can make 3, 1, and 2 units of the antibiotics in-house. In other words, the same amount of antibiotics can be obtained more cheaply through production than through outsourcing. Likewise, the second dual constraint ensures that the asking price for 1 unit of A, 1 unit of B, and 6 units of C is lower than the company's cost of making that many antibiotics themselves. To summarize, the dual variables in this problem are still valuating the constraints, but now the constraints represent requirements instead of resources, and so the precise interpretation of the dual LP is also different.

Example 4. Consider the following problem:

"Universal Claims Processors processes insurance claims for large national insurance companies. Most claim processing is done by a large pool of computer operators, some of whom are permanent and some of whom are temporary. A permanent operator can process 16 claims per day, whereas a temporary operator can process 12 per day, and on average the company processes at least 450 claims each day. The company has 40 computer workstations. A permanent operator generates about 0.5 claim with errors each day, whereas a temporary operator averages about 1.4 defective claims per day. The company wants to limit claims with errors to 25 per day. A permanent operator is paid \$64 per day, and a temporary operator is paid \$42 per day. The company wants to determine the number of permanent and temporary operators to hire in order to minimize costs."

Letting x_1 be the number of full-time operators, and x_2 be the number of part-time or temporary operators, we formulate the LP

$$\begin{array}{rrcrcl} \min & 64x_1 & + & 42x_2 & & \\ s.t. & 16x_1 & + & 12x_2 & \geq & 450 \\ & x_1 & + & x_2 & \leq & 40 \\ & 0.5x_1 & + & 1.4x_2 & \leq & 25 \\ & & & x_1, x_2 & \geq & 0. \end{array}$$

Again, the primal has 2 variables and 3 constraints, so the dual must have 3 variables and 2 constraints. The formulation is:

$$\begin{array}{rrcrcrcl} \max & 450y_1 & + & 40y_2 & + & 25y_3 & & \\ s.t. & 16y_1 & + & y_2 & + & 0.5y_3 & \leq & 64 \\ & 12y_1 & + & y_2 & + & 1.4y_3 & \leq & 42 \\ & & & & & y_1 & \geq & 0 \\ & & & & & y_2, y_3 & \leq & 0. \end{array}$$

This problem does not fall into either of the preceding interpretations. Observe that the dual variables y_2 and y_3 are required to be *non-positive*. The numbers $40y_2$ and $25y_3$ in the objective function are actually **negative numbers**. The objective $450y_1 + 40y_2 + 25y_3$ means that we are taking a positive number $450y_1$, and **subtracting** something from it.

In fact, the dual LP in this example is a way to **valuate worker performance**. A worker generates value for the company by processing claims (this being the company's only product). However, in doing so, the worker occupies a workstation, an asset that could potentially be used for another purpose. The worker also generates a certain number of defective claims, which have negative value for the company. The dual variables should be interpreted as follows:

y_1 : value generated by processing a single claim
y_2 : penalty for occupying a workstation
y_3 : penalty for generating a single defective claim

The total daily value of all workers for the company is exactly

$$450y_1 + 40y_2 + 25y_3,$$

because the company processes 450 claims, owns 40 workstations, and generates 25 defective claims. The value of a full-time worker is

$$16y_1 + y_2 + 0.5y_3,$$

because every full-time worker generates $16y_1$ in value by processing 16 claims, occupies 1 workstation, and causes $0.5y_3$ in negative value by generating 0.5 defective claim. The constraint requires that

$$16y_1 + y_2 + 0.5y_3 \leq 64$$

because, by an economic argument, a full-time worker only has incentive to generate at most \$64 in value. Similarly, a part-time worker can only be expected to generate \$42 in value. The quantity y_1 can thus be viewed as the **fair value** of a processed claim.

By solving the dual LP and obtaining the optimal solution (y_1^*, y_2^*, y_3^*), we have a simple formula for **performance review**. The LP was calculated by using average figures, but we may have some individual workers who process more than 16 claims per day, or fewer than 12. Some workers may also make fewer errors than others. We can use the insights we obtained from the dual LP to calculate the value of any individual worker:

$$\text{Value} = (\text{no. of claims processed}) \times y_1^* + y_2^* + (\text{no. of defective claims}) \times y_3^*.$$

We can use this formula to identify workers who perform particularly well, and perhaps single them out for a raise or bonus.

Summary. For any linear program, we can write down a dual LP by following certain rules. The rules are always the same regardless of the meaning or context of the original problem. However, the meaning of the dual heavily depends on the context. In one example, the dual finds fair prices for assets to be purchased by a buyer; in a second example, the dual calculates a fair way to evaluate worker performance. There is a universal way to write down a dual problem, but there is no universal way to interpret it. In order to explain the meaning of the dual, it is necessary to carefully examine the formulation and think back to the meaning of each number in the original problem.