Linear Programming III: Duality Theory and Zero-Sum Games

Linear Programming Duality

The Dual of a Linear Program

Every linear program has a *dual* linear program. We call the original linear program the *primal*. A maximization problem's dual is a minimization problem. There are a bunch of amazing properties that come from LP duality.

We have the following optimization problem: You're selling nutrients to the BU population and deciding what to price each macro at. The decision variables x_i will indicate the price per nutrient. The constraints indicate that these prices together cannot exceed the prices for the grains that you're extracting the nutrients from, since that's already the market price. The goal is to maximize your profits from a population that is buying exactly the nutrient diet of 8kg starch, 15kg proteins, and 3kg vitamins.

Primal:

$$\begin{aligned} \max \quad & 8x_1 + 15x_2 + 3x_3 & & \\ \text{subject to} \quad & 5x_1 + 4x_2 + 2x_3 \leq 0.6 & \text{(grain 1)} \quad & (y_1) \\ & 7x_1 + 2x_2 + 1x_3 \leq 0.35 & \text{(grain 2)} \quad & (y_2) \\ & x_1, x_2, x_3 \geq 0 & \text{(non-negativity)} & \end{aligned}$$

Dual:

$$\begin{aligned} \min \quad & 0.6y_1 + 0.35y_2 & & \\ \text{subject to} \quad & 5y_1 + 7y_2 \geq 8 & \text{(starch)} \quad & (x_1) \\ & 4y_1 + 2y_2 \geq 15 & \text{(proteins)} \quad & (x_2) \\ & 2y_1 + 1y_2 \geq 3 & \text{(vitamins)} \quad & (x_3) \\ & y_1, y_2 \geq 0 & \text{(non-negativity)} & \end{aligned}$$

To take the dual: Label each primal constraint with a new dual variable. In our new linear program, each dual constraint will correspond to a primal variable. For the left-hand side, count up the appearances of this constraint's primal variable (e.g., x_1) in each of the primal constraints and multiply them by the dual variable for those constraints. That is, if x_1 appears 5 times ($5x_1$) in constraint for y_1, then add $5y_1$ to x_1's constraint. Don't forget to include its appearance in the primal's objective function, but this will be the right-hand side of the constraint. Finally, the dual objective function is given by the right-hand side coefficients and their correspondence to the dual variables via the constraints in the primal. (See above).

Sometimes, the dual can even be interpreted as a related problem. In fact, this dual can be interpreted as exactly our nutrition example from Lecture #18: BU has hired you to optimize nutrition for campus dining. There are two possible grains they can offer, grain 1 and grain 2, and each contains the macronutrients given in the table in Lecture #18, plus cost per kg for each of the grains. The nutrition requirement per day of starch, proteins, and vitamins is 8, 15, and 3 respectively. Determine how much of each grain to buy such that BU spends as little but meets its nutrition requirements.

The following is the normal form for a maximization problem primal and its dual:

$$\begin{aligned} \max \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \end{aligned} \qquad\qquad \begin{aligned} \min \quad & \mathbf{b}^T\mathbf{y} \\ \text{subject to} \quad & \mathbf{A}^T\mathbf{y} \ge \mathbf{c} \end{aligned}$$

For the above example:

$$\mathbf{A} = \begin{bmatrix} 5 & 4 & 2 \\ 7 & 2 & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 0.6 \\ 0.35 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 8 \\ 15 \\ 3 \end{bmatrix}$$

Example: Maximum Matching

Given a graph $G = (V, E)$ choose a maximum size matching—a set of edges S such that no vertex is covered by more than one edge.

Decision variables: x_e indicating whether edge e is in the matching.

Primal Linear Program:

$$\begin{aligned} \max \quad & \sum_{e \in E} x_e \\ \text{subject to} \quad & \sum_{e: v \in e} x_e \le 1 && \forall v \quad \text{(vertex matched at most once)} \quad (y_v) \\ & x_e \ge 0 && \forall e \quad \text{(non-negativity)} \end{aligned}$$

Taking the dual of the above primal, we get the following linear program:

$$\begin{aligned} \min \quad & \sum_{v \in V} y_v \\ \text{subject to} \quad & \sum_{v \in e} y_v \ge 1 && \forall e \quad \text{(edge covered)} \quad (x_e) \\ & y_v \ge 0 && \forall v \quad \text{(non-negativity)} \end{aligned}$$

What problem is this? (Fractional) Vertex Cover!

Conditions for Optimality

Weak Duality

Theorem 1 (Weak Duality)**.** *If* $\mathbf{x}$ *is feasible in (P) and* $\mathbf{y}$ *is feasible in (D) then* $\mathbf{c}^T\mathbf{x} \le \mathbf{b}^T\mathbf{y}$.

This theorem says that *any* feasible solution to the primal is a *lower bound* to *any* feasible solution to the dual, and likewise, any feasible solution to the dual is an *upper bound* to the primal.

That is, fractional vertex cover gives an upper bound on how large the (fractional) maximum matching can be, and likewise, fractional maximum matching gives a lower bound on how small the minimum (fractional) vertex cover can be.

Strong Duality

Strong duality states that everything in fact needs to hold with equality to be optimal.

Theorem 2 (Strong Duality)**.** *A pair of solutions* $(\mathbf{x}^*, \mathbf{y}^*)$ *are optimal for the primal and dual respectively if and only if* $\mathbf{c}^T\mathbf{x}^* = \mathbf{b}^T\mathbf{y}^*$.

Proof. ($\Leftarrow$) The *if* direction is easy to see: we know that the dual gives an upper bound on the primal, so if these objectives are equal, then the primal objective that we are trying to maximize could not possible get any larger, as it's always *at most* the dual's objective. This is *as tight as possible.*

($\Rightarrow$) The *only if* direction is harder to prove, and we'll skip it for now. □

Complementary Slackness

We rewrite the primal and dual with each constraint separated, and then formalize another condition for optimality called *complementary slackness*, which states that for each corresponding constraint and variable, at most one can be slack in an optimal solution.

Primal (P):

$$\begin{aligned} \max \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \sum_i a_{ji}x_i \le b_j \quad \forall j \quad (y_j) \\ & x_i \ge 0 \quad \forall i \end{aligned}$$

Dual (D):

$$\begin{aligned} \min \quad & \mathbf{b}^T\mathbf{y} \\ \text{subject to} \quad & \sum_i a_{ij}y_i \ge c_i \quad \forall i \quad (x_i) \\ & y_j \ge 0 \quad \forall j \end{aligned}$$

Theorem 3 (Complementary Slackness)**.** *A pair of solutions* $(\mathbf{x}^*, \mathbf{y}^*)$ *are optimal for the primal and dual respectively if and only if the following complementary slackness conditions (1) and (2) hold:*

$$\sum_i a_{ji}x_i = b_j \quad \text{or} \quad y_j = 0 \qquad (1) \qquad\qquad \sum_i a_{ij}y_i = c_i \quad \text{or} \quad x_i = 0. \qquad (2)$$

Proof. ($\Rightarrow$) According to complementary slackness, by rearranging our constraint, either $\sum_i a_{ji}x_i - b_j = 0$ or $y_j = 0$. This ensures that the multiplied quantity $(\sum_i a_{ji}x_i - b_j)\, y_j = 0$, as *one* of the two terms on the left-hand side must be 0. Then multiplying out and rearranging gives that $y_j \sum_i a_{ji}x_i = y_j b_j$. This process with all rows gives the equality from complementary slackness that $\mathbf{y}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T\mathbf{b}$.

Similarly, using the condition that $\sum_i a_{ij}y_i = c_i$ or $x_i = 0$ gives that $\mathbf{c}^T\mathbf{x} = (\mathbf{A}^T\mathbf{y})\mathbf{x}$.

Then following our inequalities in the proof of weak duality, they now all hold with equality, so by Strong Duality, $(\mathbf{x}, \mathbf{y})$ are optimal solutions to the primal and dual.

$$\mathbf{c}^T\mathbf{x} \quad = \quad (\mathbf{A}^T\mathbf{y})\mathbf{x} \quad = \quad \mathbf{y}^T\mathbf{A}\mathbf{x} \quad = \quad \mathbf{y}^T\mathbf{b} \quad = \quad \mathbf{b}^T\mathbf{y}.$$

($\Leftarrow$) Similarly, if Strong Duality holds, the above inequalities hold with equality, in which case it must be that $y_j \sum_i a_{ji}x_i = y_j b_j$ for all j and $\sum_i a_{ij}y_i x_i = c_i x_i$ for all i, and hence that either $\sum_i a_{ji}x_i - b_j = 0$ or $y_j = 0$ for all j and that either $\sum_i a_{ij}y_i = c_i$ or $x_i = 0$ for all i. □

Zero-Sum Games and the Minimax Theorem

Consider the game *Rock-Paper-Scissors*, where as usual, paper covers rock, scissors cuts paper, and rock breaks scissors (that is: the former beats the latter in the comparison). In a face-off, the winner earns +1 and the loser earns -1. If two of the same type face each other, then there is a tie, and both earn 0.

The matrix below shows the game of Rock-Paper Scissors depicted as a *zero-sum-game*. Suppose that brothers Ron and Charlie Weasley are facing off. Each brother must choose a strategy. In the language of the *payoff matrix* below, Ron is the *row player*, and he must choose a row to play as his strategy. Similarly, Charlie is the *column player* and he just choose which column to play. If Ron chooses row i and Charlie chooses column j, then the payoff to Ron will be a_{ij}, and the payoff to Charlie will be $-a_{ij}$, hence the term “zero-sum.” Thus, the row and column players prefer bigger and smaller numbers, respectively.

	Rock	Paper	Scissors
Rock	0	-1	1
Paper	1	0	-1
Scissors	-1	1	0

Order of Turns

- Typically, RPS is played by both players simultaneously choosing their strategies.
- But what if I made you go first? That’s obviously unfair—whatever you do, I can respond with the winning move.
- Now what if I only forced you to commit to a *probability distribution* over rock, paper, and scissors? (Then I respond choosing a strategy, and *then* nature flips coins on your behalf.)

  You can protect yourself by randomizing uniformly among the three options—then, no matter what I do, I’m equally likely to win, lose, or tie.

The *minimax theorem* states that, in general games of “pure competition,” a player moving first can always protect herself by randomizing appropriately.

The Minimax Theorem

Notation:

- $m \times n$ payoff matrix $\mathbf{A}$—a_{ij} is the row player’s payoff for outcome (i, j) when row player plays strategy i and column player plays strategy j

- mixed row strategy $\mathbf{x}$ (a distribution over rows)
- mixed column strategy $\mathbf{y}$ (a distribution over columns)

Expected payoff of the row player:

$$\begin{aligned}\sum_{i=1}^{m}\sum_{j=1}^{n}\Pr[\text{outcome } (i,j)]\, a_{ij} &= \sum_{i=1}^{m}\sum_{j=1}^{n}\underbrace{\Pr[\text{row } i \text{ chosen}]}_{=x_i}\,\underbrace{\Pr[\text{column } j \text{ chosen}]}_{=y_j}\, a_{ij} \\ &= \mathbf{x}^T\mathbf{A}\mathbf{y}\end{aligned}$$

The minimax theorem is the amazing statement that turn order *doesn't matter*.

Theorem 4 (Minimax Theorem)**.** *For every two-player zero-sum game* $\mathbf{A}$*,*

$$\max_{\mathbf{x}}\left(\min_{\mathbf{y}}\mathbf{x}^T\mathbf{A}\mathbf{y}\right) = \min_{\mathbf{y}}\left(\max_{\mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{y}\right). \tag{1}$$

On the left, the row player goes first, choosing a strategy to maximize their payoff and protect against the fact that the column player goes second and adapts to their strategy. The right is the opposite situation. The *value of the game* (value that both sides will equal) is 0 in this case: the first player will play randomly and the second will respond arbitrarily.

From LP Duality to Minimax

This is not the original or only argument, but we will now derive Theorem 4 from LP duality arguments. The first step is to formalize the problem of computing the best strategy for the player forced to go first.

Two issues: (1) the nested min/max, and (2) the quadratic (nonlinear) character of $\mathbf{x}^T\mathbf{A}\mathbf{y}$ in the decision variables $\mathbf{x}$, $\mathbf{y}$.

Observation 5. *The second player never needs to randomize. If the row player goes first and chooses any distribution* $\mathbf{x}$*, the column player can then simply compute the expected payoff (with respect to* $\mathbf{x}$*) of each column and choose the best.*

In math, we have argued that

$$\max_{\mathbf{x}}\left(\min_{\mathbf{y}}\mathbf{x}^T\mathbf{A}\mathbf{y}\right) = \max_{\mathbf{x}}\left(\min_{j=1}^{n}\mathbf{x}^T\mathbf{A}\mathbf{e}_j\right) \tag{2}$$

$$= \max_{\mathbf{x}}\left(\min_{j=1}^{n}\sum_{i=1}^{m}a_{ij}x_i\right) \tag{3}$$

where $\mathbf{e}_j$ is the jth standard basis vector, corresponding to the column player deterministically choosing column j.

We've solved one of our problems by getting rid of $\mathbf{y}$. But there is still the nested max/min.

Specifically, we introduce a decision variable v, intended to be equal to (2), and

$$\max v$$

subject to

$$v - \sum_{i=1}^{m} a_{ij} x_i \leq 0 \quad \text{for all } j = 1, \ldots, n$$

$$\sum_{i=1}^{m} x_i = 1$$

$$x_1, \ldots, x_m \geq 0 \quad \text{and} \quad v \in \mathbb{R}.$$

Note that this is a linear program. Rewriting the constraints in the form

$$v \leq \sum_{i=1}^{m} a_{ij} x_i \quad \text{for all } j = 1, \ldots, n$$

makes it clear that they force v to be at most $\min_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i$.

If $(v^*, \mathbf{x}^*)$ is an optimal solution, then $v^* = \min_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i$. By feasibility, v^* cannot be larger than $\min_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i^*$. If it were strictly less, then we can increase v^* slightly without destroying feasibility, yielding a better feasible solution (contradicting optimality).

Since the linear program explicitly maximizes v over all distributions $\mathbf{x}$, its optimal objective function value is

$$v^* = \max_{\mathbf{x}} \left(\min_{j=1}^{n} \mathbf{x}^T \mathbf{A} \mathbf{e}_j \right) = \max_{\mathbf{x}} \left(\min_{\mathbf{y}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right) \tag{4}$$

Now, we do the same thing for the column player, where the column player moves first:

$$\min w$$

subject to

$$w - \sum_{j=1}^{n} a_{ij} y_j \geq 0 \quad \text{for all } i = 1, \ldots, m$$

$$\sum_{j=1}^{n} y_j = 1$$

$$y_1, \ldots, y_n \geq 0 \quad \text{and} \quad w \in \mathbb{R}.$$

At an optimal solution $(w^*, \mathbf{y}^*)$, $\mathbf{y}^*$ is the optimal strategy for the column player (when going first, assuming optimal play by the row player) and

$$w^* = \min_{\mathbf{y}} \left(\max_{i=1}^{m} \mathbf{e}_i^T \mathbf{A} \mathbf{y} \right) = \min_{\mathbf{y}} \left(\max_{\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right) \tag{5}$$

These two linear programs are duals! For example, the one unrestricted variable (v or w) corresponds to the one equality constraint in the other linear program ($\sum_{j=1}^{n} y_j = 1$ or $\sum_{i=1}^{m} x_i = 1$, respectively). The n x variables correspond to the remaining dual constraints, and the m y variables correspond to the remaining primal constraints. Then strong duality implies that $v^* = w^*$; in light of (4) and (5), the minimax theorem follows directly.