Game Theory

Lecture 09

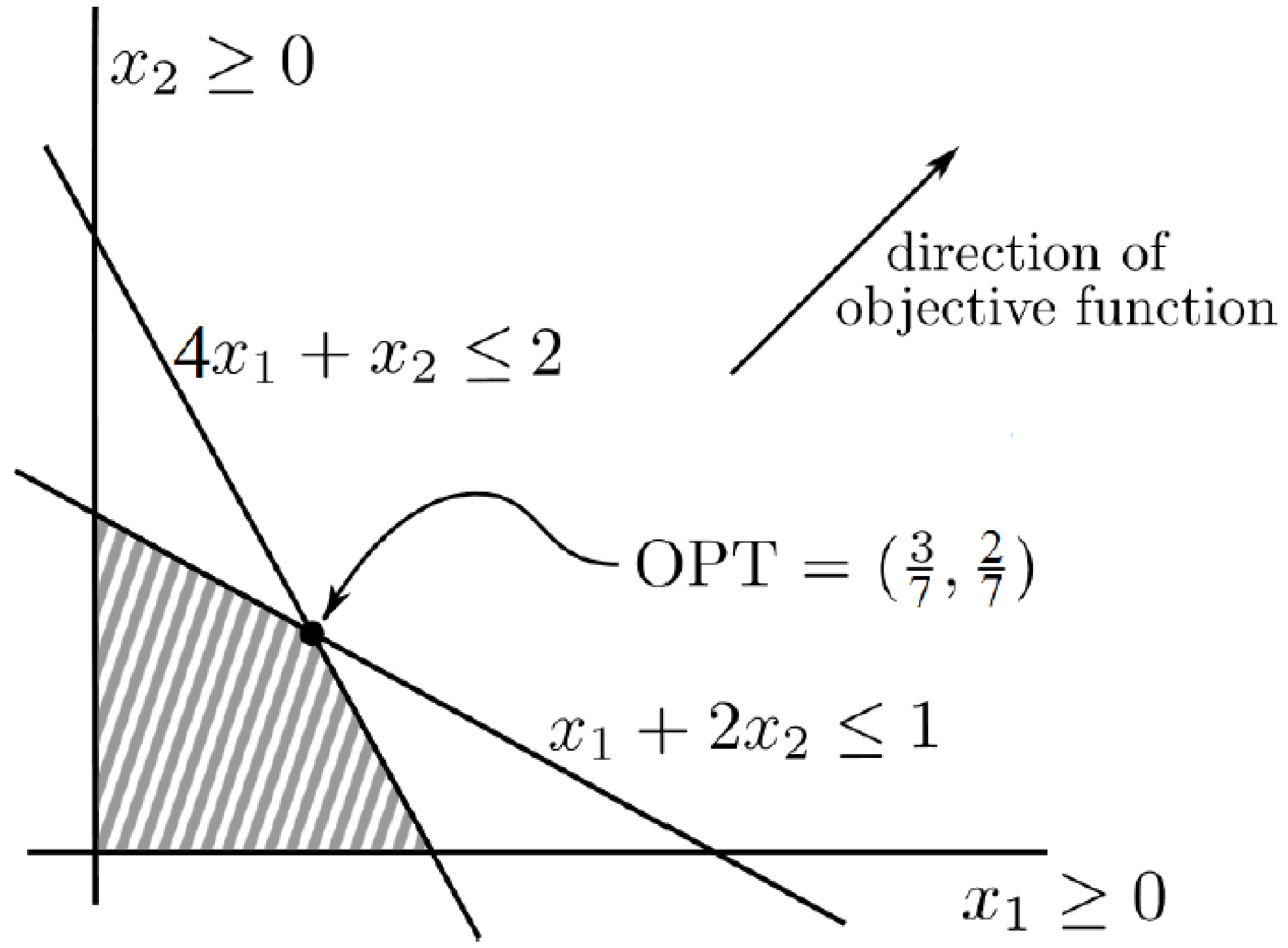
LP Duality and the Minimax Theorem

A toy example to illustrate duality

$$\max x_1 + x_2$$

subject to

$$4x_1 + x_2 \leq 2$$
$$x_1 + 2x_2 \leq 1$$
$$x_1 \geq 0$$
$$x_2 \geq 0.$$



- What's an easy and convincing proof that the optimal objective function value of the linear program can't be too large?

$$\underbrace{x_1 + x_2}_{\text{objective}} \leq 4x_1 + x_2 \leq \underbrace{2}_{\text{upper bound}},$$

- But actually, it's obvious that we can do better by using the second constraint instead: $x_1 + x_2 \leq x_1 + 2x_2 \leq 1,$
- Can we do better?
  - ➢ There's no reason we need to stop at using just one constraint at a time, and are free to blend two or more constraints.
  - ➢ The best blending takes 1/7 of the first and 3/7 of the second to give:

$$x_1 + x_2 \leq \tfrac{1}{7}\underbrace{(4x_1 + x_2)}_{\leq 2} + \tfrac{3}{7}\underbrace{(x_1 + 2x_2)}_{\leq 1} \leq \frac{1}{7} \cdot 2 + \frac{3}{7} \cdot 1 = \frac{5}{7}.$$

➢ This is a convincing proof that the optimal objective function value is at most 5/7. Given the feasible point (3/7;2/27) that actually does realize this upper bound, we can conclude that 5/7 really is the optimal value for the LP.

The Dual Linear Program

- We now generalize the ideas of the previous slide. Consider an arbitrary linear program (call it (P)) of the form:

$$
\begin{aligned}
\max \sum_{j=1}^{n} c_j x_j & \\
\text{subject to} \quad \sum_{j=1}^{n} a_{1j} x_j &\le b_1 \\
\sum_{j=1}^{n} a_{2j} x_j &\le b_2 \\
\vdots &\le \vdots \\
\sum_{j=1}^{n} a_{mj} x_j &\le b_m \\
x_1, \ldots, x_n &\ge 0.
\end{aligned}
$$

Matrix-Vector Notation

$$
\begin{aligned}
\max \mathbf{c}^T \mathbf{x} \\
\text{subject to} \quad & \\
\mathbf{A}\mathbf{x} \le \mathbf{b} \\
\mathbf{x} \ge 0,
\end{aligned}
$$

where c and x are n-vectors, b is an m-vector, A is an mxn matrix (of the a_{ij}'s), and the inequalities are component-wise.

- Remember our strategy for deriving upper bounds on the optimal objective function value of our toy example:
  - ➢ take a nonnegative linear combination of the constraints that (component-wise) dominates the objective function.
  - ➢ In general, for the above linear program with m constraints, we denote by $y_1, \ldots, y_m \ge 0$ the corresponding multipliers that we use.

The Dual Linear Program (Cont'd)

- The goal of dominating the objective function translates to the conditions

$$\sum_{i=1}^{m} y_i a_{ij} \geq c_j \qquad (*)$$

for each objective function coefficient (i.e. for $j = 1, 2, \ldots, n$).

In matrix notation, we are interested in nonnegative m-vectors $\mathbf{y} \geq 0$ such that

$$\mathbf{A}^T \mathbf{y} \geq \mathbf{c};$$

note the sum in $(*)$ is over the rows i of A, which corresponds to an inner product with the jth column of $\mathbf{A}$, or equivalently with the jth row of $\mathbf{A}^T$.

By design, every such choice of multipliers $y_1, \ldots, y_m$ implies an upper bound on the optimal objective function value of the linear program : for every feasible solution $(x_1, \ldots, x_n)$,

$$\underbrace{\sum_{j=1}^{n} c_j x_j}_{\mathbf{x}\text{'s obj fn}} \leq \sum_{j=1}^{n} \left(\sum_{i=1}^{m} y_i a_{ij} \right) x_j$$

$$= \sum_{i=1}^{m} y_i \cdot \left(\sum_{j=1}^{n} a_{ij} x_j \right)$$

$$\leq \underbrace{\sum_{i=1}^{m} y_i b_i}_{\text{upper bound}}. \ .$$

The Dual Linear Program (Cont'd)

- Alternatively, the derivation may be more transparent in matrix-vector notation: $\mathbf{c}^T\mathbf{x} \leq (\mathbf{A}^T\mathbf{y})^T\mathbf{x} = \mathbf{y}^T(\mathbf{A}\mathbf{x}) \leq \mathbf{y}^t\mathbf{b}.$

The upshot is that, whenever $\mathbf{y} \geq 0$ and $(\ast)$ holds,

$$\text{OPT of (P)} \leq \sum_{i=1}^{m} b_i y_i.$$

- Obviously, the most interesting of these upper bounds is the tightest (i.e., smallest) one. So we really want to range over all possible y's and consider the minimum such upper bound.
- Here's the key point: *the tightest upper bound on OPT is itself the optimal solution to a linear program*. Namely:

$$\min \sum_{i=1}^{m} b_i y_i \quad \text{subject to} \quad \begin{aligned} \sum_{i=1}^{m} a_{i1} y_i &\geq c_1 \\ \sum_{i=1}^{m} a_{i2} y_i &\geq c_2 \\ \vdots &\geq \vdots \\ \sum_{i=1}^{m} a_{in} y_i &\geq c_n \\ y_1, \ldots, y_m &\geq 0. \end{aligned}$$

The Dual Linear Program (Cont'd)

in matrix-vector form:

$$\min \mathbf{b}^T\mathbf{y}$$
subject to
$$\mathbf{A}^T\mathbf{y} \geq \mathbf{c}$$
$$\mathbf{y} \geq 0.$$

This linear program is called the *dual* to (P), and we sometimes denote it by (D).

For example, to derive the dual to our toy linear program, we just swap the objective and the right-hand side and take the transpose of the constraint matrix:

$$\max x_1 + x_2$$
subject to
$$4x_1 + x_2 \leq 2$$
$$x_1 + 2x_2 \leq 1$$
$$x_1 \geq 0$$
$$x_2 \geq 0.$$

DUAL

$$\min 2y_1 + y_2$$
subject to
$$4y_1 + y_2 \geq 1$$
$$y_1 + 2y_2 \geq 1$$
$$y_1, y_2 \geq 0.$$

The objective function values of the feasible solutions $(1, 0)$, $(0, 1)$, and $(\frac{1}{7}, \frac{3}{7})$ (of 2, 1, and $\frac{5}{7}$) correspond to our three upper bounds.

Weak Duality

The following important result follows from the definition of the dual and **our** derivation

(Weak Duality) *For every linear program of the form (P) and corresponding dual linear program (D),*

$$\textit{OPT value for (P)} \leq \textit{OPT value for (D)}.$$

$$c^T x^* \leq b^T y^*$$

The LP Duality Theorem

Theorem("Strong Duality", von Neumann'47) One of the following four situations holds:

1. Both the primal and dual LPs are feasible, and for any optimal solutions x^* of the primal and y^* of the dual:
$$c^T x^* = b^T y^*$$
2. The primal is unbounded and the dual is infeasible.
3. The primal is infeasible and the dual is unbounded.
4. Both LPs are infeasible.

Complementary Slackness

- Complementary slackness conditions are a corollary of LP duality, and are another <u>sufficient condition for optimality</u>.
- Let (P),(D) be a primal-dual pair of linear programs.
  - If **x**, **y** are feasible solutions to (P),(D), and the following two conditions hold then both **x** and **y** are optimal:

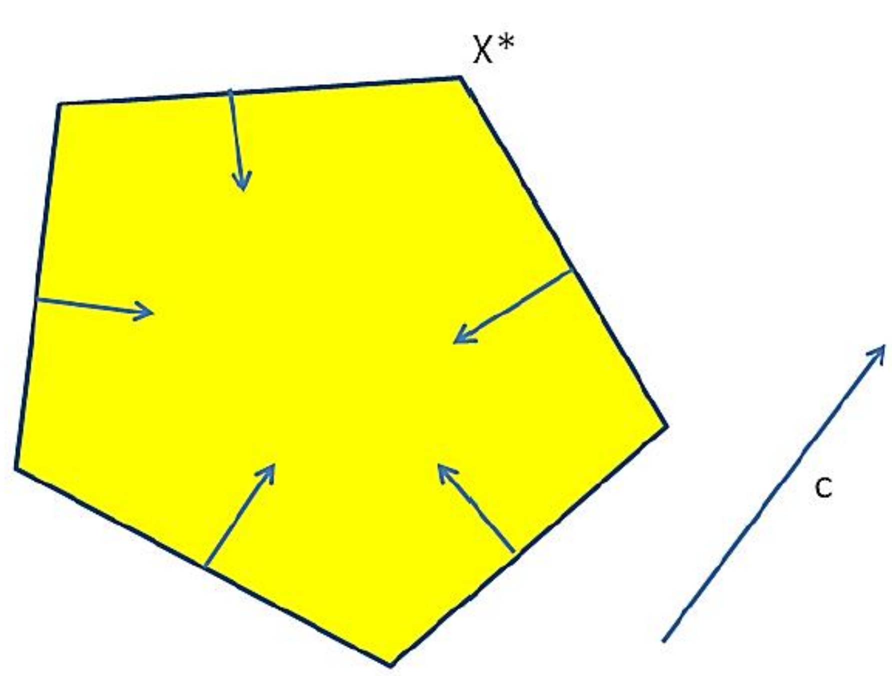
(1) Whenever $x_j \neq 0$, $\mathbf{y}$ *satisfies the jth constraint of (D) with equality.*

(2) Whenever $y_i \neq 0$, $\mathbf{x}$ *satisfies the ith constraint of (P) with equality.*

- The conditions assert that no decision variable and corresponding constraint are simultaneously “slack" (i.e., it forbids that the decision variable is not 0 and also the constraint is not tight).

<u>Physical interpretation of complementary slackness</u>

- The objective function pushes a particle in the direction c until it rests at x*. Walls also exert a force on the particle, and complementary slackness asserts that only walls touching the particle exert a force, and sum of forces is equal to 0.



Proof: Complementary slackness ➔ optimality

- Recall for every pair **x**, **y** of feasible solutions to (P),(D), we have

$$\underbrace{\sum_{j=1}^{n} c_j x_j}_{\mathbf{x}\text{'s obj fn}} \leq \sum_{j=1}^{n} \left(\sum_{i=1}^{m} y_i a_{ij} \right) x_j \qquad (1)$$

$$= \sum_{i=1}^{m} y_i \left(\sum_{j=1}^{n} a_{ij} x_j \right) \qquad (2)$$

$$\leq \underbrace{\sum_{i=1}^{m} y_i b_i}_{\mathbf{y}\text{'s obj fn}} . \qquad (3)$$

- From the first complementary slackness condition, we have:

$$c_j x_j = \left(\sum_{i=1}^{m} y_i a_{ij} \right) x_j$$

for each $j = 1, \ldots, n$ (either $x_j = 0$ or $c_j = \sum_{i=1}^{m} y_i a_{ij}$).

Hence, inequality (1) holds with equality.

- Similarly, the second condition implies that:

$$y_i \left(\sum_{j=1}^{n} a_{ij} x_i \right) = y_i b_i \text{ for each } i = 1, \ldots, m.$$

Hence inequality (3) also holds with equality.

Thus $\mathbf{c}^T\mathbf{x} = \mathbf{y}^T\mathbf{b}$, and implies that both $\mathbf{x}$ and $\mathbf{y}$ are optimal.

general recipe for LP duals

- As we've mentioned, different types of linear programs are easily converted to each other.
- So one perfectly legitimate way to take the dual of an arbitrary linear program is to first convert it into the canonical form, and then apply the duality definition.
- But it's more convenient to be able to take the dual of any linear program directly, using a general recipe. The high-level points of the recipe are familiar:
  - dual variables correspond to primal constraints,
  - dual constraints correspond to primal variables,
  - maximization and minimization get exchanged,
  - the objective function and right-hand side get exchanged,
  - and the constraint matrix gets transposed.
  - The details concern the different type of constraints (inequality vs. equality) and whether or not decision variables are nonnegative.

general recipe for LP duals

- Here is the general recipe for maximization linear programs:

Primal	Dual
variables $x_1, \dots, x_n$	n constraints
m constraints	variables $y_1, \dots, y_m$
objective function $\mathbf{c}$	right-hand side $\mathbf{c}$
right-hand side $\mathbf{b}$	objective function $\mathbf{b}$
$\max \mathbf{c}^T \mathbf{x}$	$\min \mathbf{b}^T \mathbf{y}$
constraint matrix $\mathbf{A}$	constraint matrix $\mathbf{A}^T$
ith constraint is "$\leq$"	$y_i \geq 0$
ith constraint is "$\geq$"	$y_i \leq 0$
ith constraint is "$=$"	$y_i \in \mathbb{R}$
$x_j \geq 0$	jth constraint is "$\geq$"
$x_j \leq 0$	jth constraint is "$\leq$"
$x_j \in \mathbb{R}$	jth constraint is "$=$"

- For minimization linear programs, we define the dual as the reverse operation (from the right column to the left).

- Thus, by definition, the dual of the dual is the original primal.

From LP Duality to Minimax

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

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

- We now proceed to derive the Minimax Theorem from LP duality.
  - ➢ The first step is to formalize the problem of computing the best strategy for the max-player:

$$\max_{\mathbf{x}} \left(\min_{\mathbf{y}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right) = \max_{\mathbf{x}} \left(\min_{j=1}^{n} \mathbf{x}^T \mathbf{A} \mathbf{e}_j \right) = \max_{\mathbf{x}} \left(\min_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i \right),$$

Why? Because the min-player never needs to randomize.

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

From LP Duality to Minimax

$$\max_{\mathbf{x}} \left(\min_{\mathbf{y}} \mathbf{x}^T \mathbf{A} \mathbf{y} \right) = \max_{\mathbf{x}} \left(\min_{j=1}^{n} \mathbf{x}^T \mathbf{A} \mathbf{e}_j \right) = \max_{\mathbf{x}} \left(\min_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i \right),$$

- To get rid of the nested max/min, we recall a trick from before, that a minimum or maximum can often be simulated by additional variables and constraints. The same trick works here, in exactly the same way:

- Specifically, we introduce a decision variable v, intended to be equal to $\min_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i$ and we will have:

$$\begin{aligned}
&\max v \\
&\text{subject to} \\
&\quad v - \sum_{i=1}^{m} a_{ij} x_i \leq 0 \qquad \text{for all } j = 1, \ldots, n \\
&\quad \sum_{i=1}^{m} x_i = 1 \\
&\quad x_1, \ldots, x_m \geq 0 \quad \text{and} \quad v \in \mathbb{R}.
\end{aligned}$$

- Note that this is a linear program with optimal (v^*, x^*).

From LP Duality to Minimax

- Repeating the exercise for the column player gives the linear program:

$$\min w$$

subject to

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

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

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

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

$$w^* = \min_{\mathbf{y}} \left(\max_{i=1}^{m} e_i^\top \mathbf{A} \mathbf{y} \right) = \min_{\mathbf{y}} \left(\max_{\mathbf{x}} \mathbf{x}^\top \mathbf{A} \mathbf{y} \right).$$

Here's the punch line: *these two linear programs are duals.*

This can be seen by looking up our recipe for taking duals and verifying that these two linear programs conform to the recipe

- You will do this verification as an Exercise!
- Accepting this,

Strong duality implies that $v^* = w^*$;