Lecture 5, Wed., April 6, 2022.

Duality of the (discrete) OT problem

- Duality in linear programming (LP): standard form
- Duality of the OT problem. Some properties

Standard form of the LP problem.

$$\text{Minimize } \quad c^T x \qquad\qquad (P)$$
$$\text{Subject to } \quad Ax = b \text{ and } x \geq 0$$

Here: $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$.

<u>Theorem</u> (Optimality condition) $x^* \in \mathbb{R}^n$ is a minimizer of (P) if and only if the following hold true:

(1) $Ax^* = b$ and $x^* \geq 0$ (i.e., x^* is a feasible soln);

(2) $\exists \lambda^* \in \mathbb{R}^m$ s.t. $A^T\lambda^* \leq c$ and $(A^T\lambda^* - c)_i x_i^* = 0$ $(1 \leq i \leq n)$.

<u>QED</u>

<u>Primal and Dual</u>

Primal prob. <u>Minimize</u> $c^T x$ (P)
<u>Subject to</u> $Ax = b$ and $x \geq 0$

Dual prob. <u>Maximize</u> $b^T \lambda$ (D)
subject to $A^T \lambda \leq c$

Other LP problem:

Primal: <u>Minimize</u> $c^T x$ (P')
subject to $Ax \geq b$

Dual: <u>Maximize</u> $b^T \lambda$ (D')
subject to $A^T\lambda = c$ and $\lambda \geq 0$

Theorem The dual of the dual is the primal.

Proof Follow the definition: b in (D) $\leftrightarrow$ $-c$ in (P'), and A in (D) $\leftrightarrow$ $-A$ in (P'). QED

Theorem (Weak duality, or the duality gap). If $Ax=b$ and $x\geq 0$, and $A^T\lambda \leq c$, then $b^T\lambda \leq c^T x$.

Proof $b^T\lambda = (Ax)^T\lambda = x^T A^T \lambda \leq x^T c = c^T x$. QED

Theorem (strong duality) (P) has a minimizer $\Leftrightarrow$ (D) has a maximizer. In this case, the costs are the same. Moreover, if x^* and λ^* are feasible solutions to (P) and (D), then they are respective optimizers $\Leftrightarrow$ $(A^T\lambda^* - c)_i = 0$ if $x^*_i > 0$ $(1\leq i\leq n)$. QED

The discrete OT (in K-form): Given $C \in \mathbb{R}^{m\times n}_{\geq 0}$, $a \in P_m$, $b \in P_n$.

$$\min_{P \in \mathcal{A}(a,b)} \langle P, C \rangle .$$

$$\mathcal{A}(a,b) = \Big\{ P \in \mathbb{R}^{m\times n}_{\geq 0} : \overbrace{\sum_{j=1}^{n} P_{ij} = a_i \ (1\leq i\leq m), \ \sum_{i=1}^{m} P_{ij} = b_j \ (1\leq j\leq n)}^{(*)} \Big\}$$

Convert it into a LP problem (in standard form)

vector $c \in \mathbb{R}^{mn} \longleftrightarrow C \in \mathbb{R}^{m\times n}$

$$\begin{bmatrix} C_{11} & \cdots & C_{1n} \\ & \cdots & \\ C_{m1} & \cdots & C_{mn} \end{bmatrix} \leftrightarrow \begin{bmatrix} c_1 & \cdots & c_n \\ & \cdots & \\ c_{(m-1)n+1} & \cdots & c_{mn} \end{bmatrix}$$

$$c_{j+n(i-1)} = C_{ij} \quad (i=1,\dots,m,\ j=1,\dots,n)$$

variable $x \in \mathbb{R}^{mn} \longleftrightarrow P = [P_{ij}] \in \mathbb{R}^{m\times n}$

$$x_{j+n(i-1)} = P_{ij} \quad (i=1,\dots,m,\ j=1,\dots,n)$$

Find A and b in Problem (P) from equations (*)

Try $m=3$, $n=2$. $a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$.

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \\ p_{31} & p_{32} \end{bmatrix} \longleftrightarrow \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \\ x_5 & x_6 \end{bmatrix} \qquad \left.\begin{array}{l} x_1 + x_2 = a_1 \\ x_3 + x_4 = a_2 \\ x_5 + x_6 = a_3 \\ x_1 + x_3 + x_5 = b_1 \\ x_2 + x_4 + x_6 = b_2 \end{array}\right\}$$

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{(m+n)\times mn} \qquad \begin{bmatrix} a \\ b \end{bmatrix} = \hat{b} \in \mathbb{R}^{m+n}$$

$\lambda \in \mathbb{R}^{3+2}$. $\lambda = \begin{bmatrix} f \\ g \end{bmatrix}$ $f \in \mathbb{R}^3$, $g \in \mathbb{R}^2$.

Maximize $\lambda^T \hat{b} = f^T a + g^T b = \langle f, a \rangle + \langle g, b \rangle$.

subject to $A^T \begin{bmatrix} f \\ g \end{bmatrix} \le C$

$$A^T \begin{bmatrix} f \\ g \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} f_1 + g_1 \\ f_1 + g_2 \\ f_2 + g_1 \\ f_2 + g_2 \\ f_3 + g_1 \\ f_3 + g_2 \end{bmatrix} \le \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \\ c_{31} \\ c_{32} \end{bmatrix}$$

$\Leftrightarrow f_i + g_j \le C_{ij} \quad \forall i, j.$

General case: $A \in \mathbb{R}^{m\times n}$. $\quad 1_n = [1\ 1 \cdots 1]^T \in \mathbb{R}^n$

$$A = \left[\begin{array}{cccc} \overbrace{1 \cdots 1}^{n} & \overbrace{0 \cdots 0}^{n} & \cdots & \overbrace{0 \cdots 0}^{n} \\ 0 \cdots 0 & 1 \cdots 1 & \cdots & 0 \cdots 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 \cdots 0 & 0 \cdots 0 & \cdots & 1 \cdots 1 \\ 1\, 0 \cdots 0 & 1\, 0 \cdots 0 & \cdots & 1\, 0 \cdots 0 \\ 0\, 1 \cdots 0 & 0\, 1 \cdots 0 & \cdots & 0\, 1 \cdots 0 \\ \vdots & \vdots & & \vdots \\ 0\, 0 \cdots 1 & 0\, 0 \cdots 1 & \cdots & 0\, 0 \cdots 1 \end{array}\right] \begin{array}{l} \}\, m \\ \\ \\ \}\, n \end{array} = \left[\begin{array}{cccc} \overbrace{1_n^T}^{n} & \overbrace{0_n^T}^{n} & \cdots & \overbrace{0_n^T}^{n} \\ 0_n^T & 1_n^T & \cdots & 0_n^T \\ \vdots & \vdots & & \vdots \\ 0_n^T & 0_n^T & \cdots & 1_n^T \\ I_n & I_n & \cdots & I_n \end{array}\right] \begin{array}{l} \}\, m \\ \\ \\ \}\, n \end{array}$$

$$A^T \begin{bmatrix} f \\ g \end{bmatrix} = mn\left\{ \left[\begin{array}{cccc} 1_n & 0_n & \cdots & 0_n & I_n \\ 0_n & 1_n & \cdots & 0_n & I_n \\ \vdots & \vdots & & \vdots & \vdots \\ 0_n & 0_n & \cdots & 1_n & I_n \end{array}\right] \right. \begin{array}{l} \}\, n \\ \}\, n \\ \vdots \\ \}\, n \end{array} \begin{bmatrix} f \\ g \end{bmatrix} \begin{array}{l} \}\, m \\ \}\, n \end{array} = \begin{bmatrix} f_1 + g_1 \\ f_1 + g_2 \\ f_1 + g_n \\ \vdots \\ f_m + g_1 \\ \vdots \\ f_m + g_n \end{bmatrix} \le \begin{bmatrix} c_{11} \\ c_{12} \\ c_{1n} \\ \vdots \\ c_{m1} \\ \vdots \\ c_{mn} \end{bmatrix}$$

(the first m block columns are marked "m", the last column "n"; $\begin{bmatrix} f \\ g \end{bmatrix}$ is marked "1"; the result vector is $mn \times 1$)

The dual OT prob.

Maximize $\langle f, a \rangle + \langle g, b \rangle \quad (f \in \mathbb{R}^m,\ g \in \mathbb{R}^n)$

subject to: $f_i + g_j \le C_{ij} \quad (1 \le i \le m,\ 1 \le j \le n)$.

Denote $\mathcal{R}(C) = \{(f,g) \in \mathbb{R}^m \times \mathbb{R}^n : f \oplus g \le C\}$.
$f \oplus g = [f_i + g_j] \in \mathbb{R}^{m \times n}$

So, the dual prob. is $\max_{(f,g) \in \mathcal{R}(C)} [\langle f, a \rangle + \langle g, b \rangle]$.

<u>Proposition</u> (Weak Duality, or Duality Gap). We have

$$\langle f, a \rangle + \langle g, b \rangle \le \langle C, P \rangle \quad \forall (f,g) \in \mathcal{R}(C), \forall P \in \mathcal{A}(a,b)$$

<u>Proof</u>. $\langle C, P \rangle = \sum_{i,j} P_{ij} C_{ij} \ge \sum_{i,j} P_{ij}(f_i + g_j)$

$= \sum_{i,j} f_i P_{ij} + \sum_{i,j} g_j P_{ij} = \sum_i f_i (\sum_j P_{ij}) + \sum_j g_j (\sum_i P_{ij})$

$= \sum_i f_i a_i + \sum_j g_j b_j = \langle f, a \rangle + \langle g, b \rangle$. <u>QED</u>

<u>Proposition</u> The primal and dual problems admit their respective optimizers. Moreover, the optimal costs are the same.

<u>Proof</u> The existence of a minimizer for the primal problem was shown before. The rest follows from the Strong Duality Theorem. <u>QED</u>

<u>Proposition</u> Let $P^* \in \mathcal{A}(a,b)$ and $(f^*, g^*) \in \mathcal{R}(C)$. Then they are optimizers of the primal and dual OT problems, respectively, if and only if

$$(C_{ij} - f_i^* - g_j^*) P_{ij}^* = 0, \quad 1 \le i \le m, \ 1 \le j \le n. \qquad (*)$$

This is equivalent to: $f_i^* + g_j^* = C_{ij}$ if $P_{ij}^* > 0$, $1 \le i \le m$, $1 \le j \le n$.

Call $(*)$ the complementarity condition.

This follows from the strong duality thm for general linear programming problems. Here, we give a direct proof.

<u>Proof</u> Suppose $p^* = \arg\min_{p \in \mathcal{A}(a,b)} \langle C, p \rangle$,

$$(f^*, g^*) = \arg\max_{(f,g) \in \mathcal{R}(C)} [\langle f, a\rangle + \langle g, b \rangle].$$

Then, by the Strong Duality Thm, $\min(\mathcal{P}) = \max(D)$.

$$\langle C, p^* \rangle = \langle f^*, a \rangle + \langle g^*, b \rangle.$$

But $\langle f^*, a\rangle + \langle g^*, b \rangle = \sum_i f_i^* a_i + \sum_j g_j^* b_j$

$$= \sum_i f_i^* \left(\sum_{j=1}^m p_{ij}^*\right) + \sum_j g_j^* \left(\sum_{i=1}^n p_{ij}^*\right)$$

$$= \sum_{i,j} p_{ij}^* (f_i^* + g_j^*) = \langle f^* \oplus g, p^* \rangle$$

Hence, $\langle C - f^* \oplus g, p^* \rangle = 0$, which is (*).

Suppose (*) holds true. Then

$$\min_{p \in \mathcal{A}(a,b)} \langle C, p \rangle \le \langle C, p^* \rangle \overset{(*)}{=} \langle f^* + g^*, p^* \rangle$$

$$\overset{\text{[same as above]}}{=} \langle f^*, a \rangle + \langle g^*, b \rangle \le \min_{p \in \mathcal{A}(a,b)} \langle C, p \rangle$$

Hence, all are equal. Thus

$$p^* = \arg\min_{p \in \mathcal{A}(a,b)} \langle C, p \rangle,$$

and by the weak duality,

$$(f^*, g^*) = \arg\max_{(f,g) \in \mathcal{R}(C)} [\langle f, a \rangle + \langle g, b \rangle]$$

QED