Lecture 2, Wed. 3/30/2022

Review

Given $a = (a_i) \in \mathbb{R}^m$, all $a_i \geq 0$, $\sum_{i=1}^m a_i = 1$

$b = (b_j) \in \mathbb{R}^n$, all $b_j \geq 0$, $\sum_{j=1}^n b_j = 1$

$X = \{x_1, \dots, x_m\}$, $Y = \{y_1, \dots, y_n\}$

Cost function $c : X \times Y \to [0, \infty)$.

feasible set of transp. maps

$$\mathcal{T} = \mathcal{T}(a,b) = \left\{ T : X \to Y : \ b_j = \sum_{i : T(x_i) = y_j} a_i, \quad j = 1, \dots, n \right\}.$$

Monge's formulation: $\min_{T \in \mathcal{T}} \sum_{i=1}^m a_i c(x_i, T(x_i))$.

The optimal assignment problem

$$\min_{\sigma \in S_n} \frac{1}{n} \sum_{i=1}^n C_{i, \sigma(i)}.$$

The feasible set of transport plans

$$\mathcal{A}(a,b) = \{ P = [P_{ij}] \in \mathbb{R}^{m \times n} : \text{ all } P_{ij} \geq 0$$
$$\sum_{i=1} P_{ij} = b_j \ \forall j. \ \sum_{j=1} P_{ij} = a_i, \forall i \}$$

$C = [C_{ij}]$ – the cost function (or matrix).

Kantorovich's formulation:

$$\min_{P \in \mathcal{A}(a,b)} \sum_{i=1}^m \sum_{j=1}^n P_{ij} C_{ij}$$

Today: ① K-OT as linear programming

① Monge vs. Kantorovich

K-OT is a linear programming problem.

A canonical form of linear programming (LP) prob.

Given $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$ (Here, m, n are different from those above.)

$$\boxed{\begin{array}{ll} \text{Minimize} & f(x) \triangleq c^T x \\ \text{subj. to} & Ax \geq b \text{ and } x \geq 0. \end{array}}$$

component wise.

Define the feasible set

$$S = \{ x \in \mathbb{R}^n : Ax \geq b \text{ and } x \geq 0 \}$$

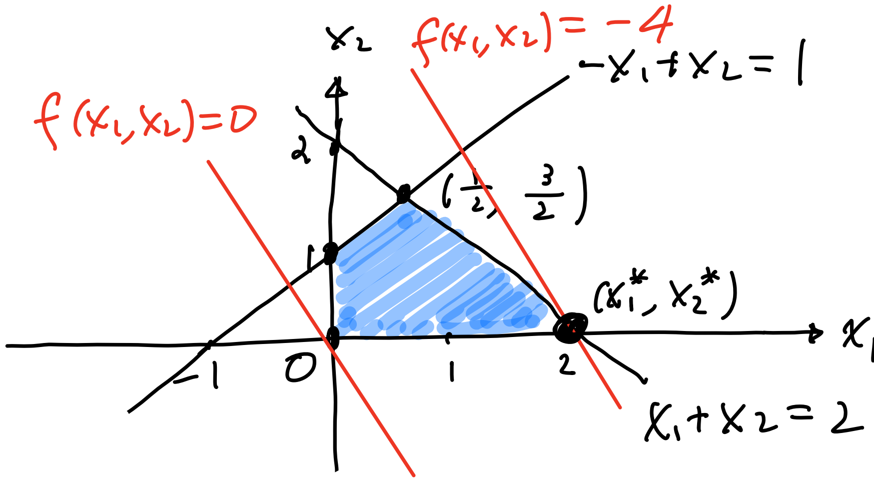
S is a convex polyhedron (boundary composed of hyperplanes).

A convex polyhedron is the convex hull of its extreme points (vertices). (Straszewicz Thm)

Example Minimize $f(x_1, x_2) = -2x_1 - x_2$

subject to $x_1 - x_2 \geq -1$, $-x_1 - x_2 \geq -2$

$x_1 \geq 0$, $x_2 \geq 0$

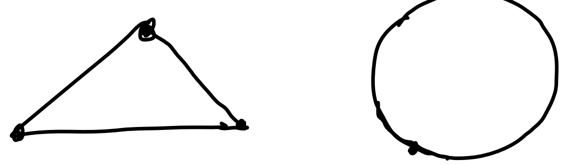


Solution: $(x_1^*, x_2^*) = (2, 0)$

Thm If the feasible set $S \neq \phi$ is bounded, then the minimum is attained at an extreme point of S.

Definition ⊙ A subset A of a vector space X is convex if $u, v \in A \Longrightarrow$ $[u, v] := \{ t u + (1-t) v : 0 \le t \le 1 \} \subseteq A$.

⊙ Let A be a convex subset of a vector space X and $e \in A$. e is an extreme point of A, if $\nexists$ no $u, v \in A$ and $t \in (0,1)$ s.t. $e = t u + (1-t) V$.



Exercise Prove the above Thm.

Methods for solving linear programming problems: The simplex method

Interior-point method

In general, not efficient enough for higher dim. problems.

An equivalent formulation for the linear programming problem, using equalities instead of inequalities as constraints.

$Ax \geq b$. Let $A = [a_{ij}] \in \mathbb{R}^{m\times n}$, $b = (b_j) \in \mathbb{R}^m$.
$x = (x_i) \in \mathbb{R}^n$.

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geq b_i, \quad i = 1, \cdots, m$$

Introduce $t_i \geq 0$ $(i = 1, \cdots, m)$ such that

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - t_i = b_i, \quad i = 1, \cdots, m.$$

Let $B = \begin{bmatrix} A & \begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix} \end{bmatrix}_{m\times(m+n)} \in \mathbb{R}^{m\times(n+m)}$.

$$y = \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^{n+m} \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad t = \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}.$$

Define $g = [c_1, \cdots, c_n, 0, \cdots, 0]^T \in \mathbb{R}^{m+n}$

Equivalen form, often called <u>the standard form</u>:

Minimize $g^T y$

subject to $By = b$, $y \geq 0$.

For the K-OT problem: each $P = [P_{ij}] \in \mathcal{A}(a, b)$ corresponds to a unique $x \in \mathbb{R}^{mn}$:

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ & \cdots & \\ P_{m1} & \cdots & P_{mn} \end{bmatrix} \to \begin{bmatrix} x_1 & \cdots & x_n \\ & \cdots & \\ x_{(m-1)n+1} & \cdots & x_{mn} \end{bmatrix} \to x = \begin{bmatrix} x_1 \\ \vdots \\ x_{mn} \end{bmatrix} \in \mathbb{R}^{mn}.$$

Similarly, $C \leftrightarrow c \in \mathbb{R}^{mn}$.

The $m+n$ equations for P_{ij} in the order:

$$\begin{aligned} \cdots\cdots &= a_1 \\ &\ \vdots \\ \cdots\cdots &= a_m \\ \cdots\cdots &= b_1 \\ &\ \vdots \\ \cdots\cdots &= b_n \end{aligned}$$

leads to $Ax = \begin{bmatrix} a \\ b \end{bmatrix}$ for the coefficient matrix $A \in \mathbb{R}^{(m+n)\times mn}$ ($A_{ij} = 0$ or 1).

The K-OT is now a linear programming prob:

minimize: $c^T x$

subject to $x \geq 0$ and $Ax = \begin{bmatrix} a \\ b \end{bmatrix}$.

The relation between Monge and Kantorovich formulations

First, focus on the optimal assignment problem with $n = m$, and $a = b = (\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n})$. (See the discussion in Lecture 1.)

Permutation matrix: square matrix that has exactly one entry of 1 in each row and each column and 0s elsewhere.

Any permutation $\sigma = (i_1 \cdots i_n) \in S_n$ defines a unique permutation matrix $P_\sigma = (P_\sigma)_{ij}$

$$(P_\sigma)_{ij} = \begin{cases} 1 & \text{if } j = \sigma_i \\ 0 & \text{otherwise} \end{cases} = \delta_{j,\sigma_i}.$$

Note. There are $n!$ elements in S_n. So, there are $n!$ permutation matrices of order $n \times n$.

$n=3$: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

The optimal assignment problem (Monge's formulation)

Find $\hat{\sigma} \in S_n$ s.t.

$$\hat{\sigma} = \arg\min_{\sigma \in S_n} \frac{1}{n} \sum_{i=1}^{n} C_{i,\sigma(i)}.$$

Recall $C = [C_{ij}]$ is the cost function

Now, define

$\mathcal{A}_0(a,b) = \{ \frac{1}{n} P \in \mathbb{R}^{n\times n} : P$ is an $n\times n$ permutation matrix $\}$

Clearly, $\mathcal{A}_0(a,b) \subseteq \mathcal{A}(a,b)$.

row sum $= a$, column sum $= b$. $\left[a = b = (\frac{1}{n}, \cdots, \frac{1}{n})\right]$

Recall: the total cost in M- and K-formulations

$E_M[T] = \sum_i a_i C(x_i, T(x_i)) = \frac{1}{n} \sum_i C_{i,\sigma_i}$, where $T = \sigma \in S_n$

$E_K[Q] = \sum_{i,j} Q_{ij} C_{ij}$, where $Q \in \mathcal{A}(a,b)$.

If $\frac{1}{n} P \in \mathcal{A}_0(a,b)$ then

$$E_K[\tfrac{1}{n}P] = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{n} P_{ij} C_{ij} = \frac{1}{n} \sum_{i,j} \delta_{j,\sigma_i} C_{ij}$$

$$= \frac{1}{n} \sum_{i=1}^{n} C_{i,\sigma_i} = E_M[T].$$

where $T = \sigma \in S_n$ that defines P.

$$\min_{Q \in \mathcal{A}(a,b)} \sum_{i,j} Q_{ij} C_{ij} \leq \min_{\frac{1}{n}P \in \mathcal{A}_0(a,b)} \frac{1}{n} \sum_{i,j} P_{ij} C_{ij}$$

$$= \min_{\sigma \in S_n} \frac{1}{n} \sum_i C_{i,\sigma(i)}.$$

<u>Thm</u>. All are equal!

Pf It follows from

(1) For a linear programming problem with a compact convex feasible set, min. value is reached by an extreme point of the set.

(2) Birkhoff-von Neumann Thm: The set of $n \times n$ bi-stochastic matrices $=$ the convex hull of the set of $n \times n$ permutation matrices. QED

Now, the general case. For the simplicity of notation, let $X = \{1, \cdots, m\}$, $Y = \{1, \cdots, n\}$. The common cost function is

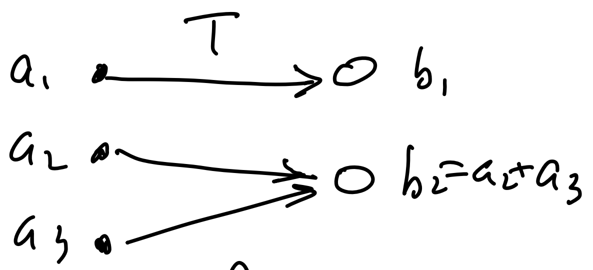
$$c(i, j) = C_{ij} \geq 0 \quad \forall i, j.$$

$\forall T \in \mathcal{T}(a, b)$: the Monge total cost is

$$E_M[T] = \sum_{i=1}^{m} a_i C_{i, T(i)}.$$

$\forall P \in \mathcal{A}(a, b)$: the Kantorovich total cost is

$$E_K[P] = \sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} C_{ij}.$$



Given $T \in \mathcal{J}(a,b)$:

$$b_j = \sum_{i:\, T(i)=j} a_i \quad (*)$$

Construct $P = P(T)$:

$$P_{ij} = a_i \delta_{j, T(i)} \quad \forall i, j$$

check: $P_{ij} \geq 0 \quad \forall i, j$

$$\forall i: \sum_{j=1}^{n} P_{ij} = a_i \sum_{j=1}^{n} \delta_{j, T(i)} = a_i \quad \text{row sum}(P) = a.$$

$$\forall j: \sum_{i=1}^{m} P_{ij} = \sum_{i:\, T(i)=j} a_i = b_j \quad \text{col. sum}(P) = b.$$

Thus, $P \in \mathcal{A}(a, b)$.

Now, the K-cost of $P = P(T)$ is

$$E_K[P(T)] = \sum_i \sum_j P_{ij} C_{ij} = \sum_{i,j} a_i \delta_{j, T(i)} C_{ij}$$

$$= \sum_i a_i C_{i, T(i)} = E_M[T].$$

<u>Thm</u>. $\min_{\mathcal{A}(a,b)} E_K[P] \leq \min_{\mathcal{J}(a,b)} E_K[P(T)]$

$$= \min_{\mathcal{J}(a,b)} E_M[T].$$ <u>QED</u>

From now on, we only consider the Kantorovich's formulation for the discrete OT problem.