1 Lecture 1: Ôn tập đại số tuyến tính, giới thiệu bài toán Optimal Transport

1.1 Phép nhân ma trận

Thông thường, ta sẽ sử dụng các chữ cái in hoa $(A, B, C, ...)$ cho các ma trận 2 chiều và các chữ cái in thường cho vector $(\mathbf{v}, \mathbf{u}, \mathbf{w}, ...)$.

Với $\mathbf{u} = (u_1, \cdots, u_n)^\top, \mathbf{v} = (v_1, \cdots, v_n)^\top$ là các vector thuộc $\mathbb{R}^n$, là trường hợp đặc biệt của ma trận (ma trận cột cột) $(\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n\times 1}$, **tích vô hướng** (inner product/dot product) của chúng là:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \mathbf{v}^\top \mathbf{u} = \sum_{i=1}^{n} u_i v_i.$$

Trong đó, 2 biểu thức đầu là kí hiệu thường gặp của tích vô hướng, 2 biểu thức sau là cách biểu diễn theo phép nhân ma trận. Khái niệm này thường xuất hiện trong vật lý, cụ thể là công thức tính công của lực tác dụng lên vật, hoặc momentum xoay tác dụng lên bản lề khi mở cửa.
<u>Ví dụ:</u> Với $\mathbf{u} = (2, 3, -1)^\top, \mathbf{v} = (0, 2, 6)^\top, \langle \mathbf{u}, \mathbf{v} \rangle = 2 \times 0 + 3 \times 2 + (-1) \times 6 = 0$. (Trường hợp này tích vô hướng của $\mathbf{u}, \mathbf{v}$ bằng 0, ta nói hai vector $\mathbf{u}, \mathbf{v}$ vuông góc nhau)

Phép nhân ma trận $A = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_m \end{pmatrix} = (\mathbf{c}_1 | \mathbf{c}_2 | ... | \mathbf{c}_n) \in \mathbb{R}^{m \times n}$ (ma trận có m hàng, n cột) với vector $\mathbf{v}$ là:

$$A\mathbf{v} = \begin{pmatrix} \langle \mathbf{r}_1, \mathbf{v} \rangle \\ \langle \mathbf{r}_2, \mathbf{v} \rangle \\ \vdots \\ \langle \mathbf{r}_m, \mathbf{v} \rangle \end{pmatrix} = \sum_{i=1}^{n} v_i \mathbf{c}_i.$$

Ta có thể hiểu một cách "cơ học" rằng $A\mathbf{v}$ là lặp lại n lần việc lấy tích vô hướng từng hàng của A với $\mathbf{v}$ với dấu bằng đầu tiên. Với dấu bằng thứ 2, ta có $A\mathbf{v}$ là một **tổ hợp tuyến tính** của các cột trong A.

<u>Ví dụ:</u> Với $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 3 & 0 \\ 4 & 5 & 7 \end{pmatrix}$ (một ma trận tam giác dưới) và vector $\mathbf{v} = (1, 2, 3)^\top$,

$A\mathbf{v} = \begin{pmatrix} \langle (1,0,0), (1,2,3) \rangle \\ \langle (1,3,0), (1,2,3) \rangle \\ \langle (4,5,7), (1,2,3) \rangle \end{pmatrix} = \begin{pmatrix} 1 \\ 7 \\ 35 \end{pmatrix}$

hoặc $A\mathbf{v} = 1 \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 3 \\ 5 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 0 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 \\ 7 \\ 35 \end{pmatrix}$

Với thêm một ma trận $B = \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_n \end{pmatrix} = (\mathbf{d}_1 | \mathbf{d}_2 | ... | \mathbf{d}_p) \in \mathbb{R}^{n \times p}$,

$$AB = (\langle \mathbf{r}_i, \mathbf{d}_j \rangle)_{ij} = (A\mathbf{d}_1 | A\mathbf{d}_2 | ... | A\mathbf{d}_p) = \begin{pmatrix} \mathbf{r}_1 B \\ \mathbf{r}_2 B \\ \vdots \\ \mathbf{r}_n B \end{pmatrix}$$

Ví dụ: Với $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 3 & 0 \\ 4 & 5 & 7 \end{pmatrix}, B = \begin{pmatrix} 2 & 0 & 0 \\ 3 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$

$$AB = \left(A \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \middle| A \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \middle| A \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right).$$

$$A \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 3 \\ 5 \end{pmatrix} + 1 \begin{pmatrix} 0 \\ 0 \\ 7 \end{pmatrix} = \begin{pmatrix} 2 \\ 11 \\ 30 \end{pmatrix}$$

$$\text{Tương tự, } A \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix}, A \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 7 \end{pmatrix}.$$

$$\text{Vậy, } AB = \begin{pmatrix} 2 & 0 & 0 \\ 11 & 3 & 0 \\ 30 & 12 & 7 \end{pmatrix}$$

Nhận xét: Tích của 2 ma trận tam giác dưới cũng là một ma trận tam giác dưới.

1.2 Norm

Norm của một vector $\mathbf{v} \in \mathbb{R}_n$ là một đại lượng mô tả độ lớn của vector, cũng là mô tả khoảng cách (ví dụ như khoảng cách giữa các tỉnh thành, "khoảng cách"xe chạy giữa các vị trí trong thành phố...) , kí hiệu là $\|\mathbf{v}\|$. Một norm mà chúng ta thường thấy nhất được phát biểu dựa trên định lý Pythagore: $\|\mathbf{v}\|_2 = \sqrt{\sum_{i=1}^n \mathbf{v}_i^2}$. Một norm bất kỳ phải thỏa các tính chất sau:

- $\|\mathbf{v}\| \geq 0 \quad \forall \mathbf{v} \in \mathbb{R}^n, \ \|\mathbf{v}\| = 0 \iff \mathbf{v} = 0.$
- $\|k\mathbf{v}\| = |k| \, \|\mathbf{v}\| \, \forall k \in \mathbb{R}.$
- $\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$ (Bất đẳng thức tam giác)

.
Với k là một số thực dương, ta gọi norm k của $\mathbf{v}$ là $\|\mathbf{v}\|_k = (\sum_{i=1}^n |\mathbf{v}_i|^k)^{\frac{1}{k}}$. Ở trên, norm được đề cập đến là norm 2 (khoảng cách Euclid), đây cũng là norm mặc định sử dụng nếu không có ghi chú k cụ thể. Tuy nhiên, trong chủ đề Optimal Transport được giới thiệu ở phần sau, norm thường được sử dụng là norm 1 (khoảng cách Manhattan).
Ví dụ: Cho $\mathbf{v} = (2, -3)^\top, \|\mathbf{v}\|_1 = |2| + |-3| = 5, \|\mathbf{v}\|_2 = \sqrt{2^2 + (-3)^2} = \sqrt{13}.$

Để ý rằng bất đẳng thức Cauchy-Schwarz cho ta $\|\mathbf{v}\|_1^2 \leq \|\mathbf{v}\|_2^2 \leq n \|\mathbf{v}\|_1^2$. Áp dụng vào ví dụ trên ta có $13 \leq 5^2 = 25 \leq 26$ là một bất đẳng thức đúng.

1.3 Optimal Transport

1.3.1 Ý tưởng

Một trường hợp thường gặp trong đời sống là với một mặt hàng (ví dụ như chuối), ta có nhiều nguồn vườn chuối và có nhiều nơi có muốn ăn chuối nhưng trái mùa nên không có. Với mỗi lựa chọn chở chuối từ nơi có đến nơi cần, ta đều phải trả giá tương ứng (dựa vào tiền xăng do khoảng cách, tiền công, khối lượng chuối). Tất nhiên, ta luôn muốn hạ tổng lượng hao phí xuống thấp nhất có thể. Vậy làm sao ta có thể mô hình hóa vấn đề trên?

1.3.2 Phát biểu

Dạng tổng quát (Kantorovich's formulation):

$$\inf \left\{ \int_{X \times Y} C(x, y) d\gamma(x, y) | \gamma \in \Pi(\mathbf{r}, \mathbf{c}) \right\} \tag{1}$$

với $\Pi(r, c)$ là tập tất cả các hàm xác suất f trên $X \times Y$ thỏa mãn $f(x, .) = \int_Y f(x, y) dy = \mathbf{r}, f(., y) = \mathbf{c}$.
Máy tính thường gặp khó khăn trong việc tính chính xác tính phân (kèm theo tốn thời gian chạy), do đó người ta tìm cách để tạo ra một phiên bản rời rạc của bài toán này. Một số quan sát (chủ yếu đến từ cách tính tích phân):

- $f(x_0, .)$ giống như tổng của tất cả các $f(x_0, y)$ với y chạy, tương tự với $f(., y_0)$.
- $d\gamma(x, y)$ giống như một mảnh vuông nhỏ liên kết giữa x trong X và y trong Y nếu ta chia không gian $X \times Y$ thành lưới vuông.

Do đó, nếu ta chia không gian $X \times Y$ thành một lưới vuông $n \times n$, ta có một phiên bản khác như sau:

Định nghĩa: $\mathbb{R}_+^{m\times n}$ là tập hợp các ma trận m hàng, n cột và có các giá trị tại mỗi vị trí đều là số thực dương. Cho $\mathbf{C} \in \mathbb{R}_+^{n\times n}$ là ma trận biểu diễn cost và $\mathbf{r}, \mathbf{c} \in \mathbb{R}_+^n$ là các vector biểu diễn khối lượng, ta cần tìm

$$\min_{\mathbf{X}\geq 0} \langle \mathbf{C}, \mathbf{X} \rangle$$

sao cho: $\mathbf{X}\mathbf{1}_n = \mathbf{r}$ và $\mathbf{X}^\top \mathbf{1}_n = \mathbf{c}$.
Ví dụ: Với $\mathbf{C} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 1 \end{pmatrix}, \mathbf{r} = (0, 0.8, 0.2)^\top, \mathbf{c} = (0.5, 0.3, 0.2)^\top,$
$\mathbf{X} = \begin{pmatrix} 0 & 0 & 0 \\ 0.4 & 0.3 & 0.1 \\ 0.1 & 0 & 0.1 \end{pmatrix}$ là một ma trận thỏa mãn các yêu cầu.
Lúc này $\langle \mathbf{C}, \mathbf{X} \rangle = 1 \times 0 + 2 \times 0 + 1 \times 0 + 2 \times 0.4 + 3 \times 0.3 + 2 \times 0.1 + 1 \times 0.1 + 2 \times 0 + 1 \times 0.1 = 2.1$
Nhận xét: Tiếp tục minh họa bằng ví dụ về chuối thì X_{ij} chính là lượng chuối được chở từ vườn i ra chợ j.

- X là ma trận không âm do không thể vận chuyển khối lượng âm chuối từ vườn nào sang chợ nào được.
- Tổng lượng chuối chở ra từ các vườn ứng với $\mathbf{c}$.
- Tổng lượng chuối chở vào các chợ ứng với $\mathbf{r}$.

Để mô tả rõ hơn sự vận chuyển chuối, ta có các bảng sau để mô tả (đơn vị: tấn):

	Chợ 1	Chợ 2	Chợ 3
Vườn 1	0	0	0
Vườn 2	0.4	0.3	0.1
Vườn 3	0.1	0	0.1

Toán tử $\langle \mathbf{C}, \mathbf{X} \rangle$ ở đây được hiểu gần giống với định nghĩa của tích vô hướng, là tổng của các tích của các cặp phần tử của C và X tại cùng vị trí.
Exercise: Đầu vào $\mathbf{r}, \mathbf{c}$ có cần điều kiện gì không?

Exercise: Viết lại yêu cầu và điều kiện của bài toán OT dưới dạng bài toán Linear Programming.

Exercise: *Trong Python, làm thế nào để tìm* $\langle \mathbf{C}, \mathbf{X} \rangle$*?*

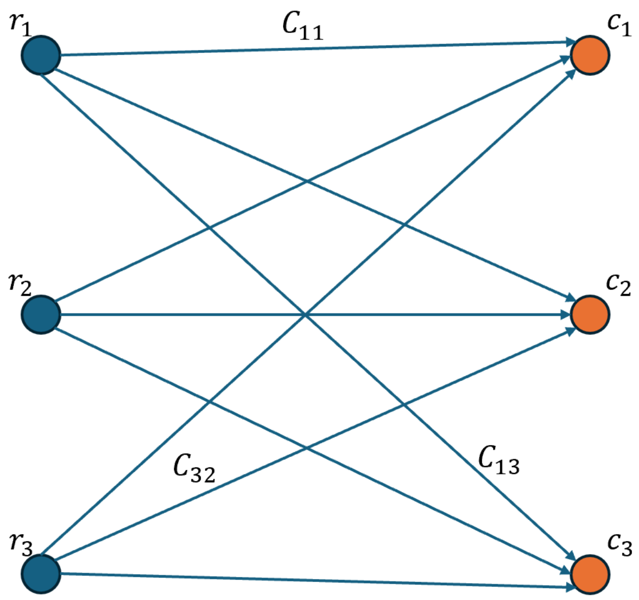
1.4 Độ phức tạp của một số thuật toán giải Linear Programming

Để ý rằng Optimal Transport là một bài toán Linear Programming, ta cần cân nhắc chọn thuật toán để giải nhanh nhất có thể. Một số thuật toán kèm với độ phức tạp được thể hiện trong bảng sau:

Algorithm	Worst-case Computation complexity
Interior-point method	$n^3 \log n$
Karmarkar	$n^{3.5}$
Simplex	Exponential

Bảng 1: Độ phức tạp của một số thuật toán giải Linear Programming.

Câu hỏi: Liệu có cách nào đẩy nhanh tốc độ của các thuật toán thuần túy giải Linear Programming như trên không?



Hình 1: Hình minh họa cho trường hợp có 3 vườn chuối và 3 chợ, với C_{11}, C_{32} biểu thị lượng hàng hóa từ vườn 1 đến chợ 1 và từ vườn 3 đến chợ 2.

1.5 Bài tập

1. Tìm AB với A, B là các ma trận sau:
   - $A = \begin{pmatrix} 1 & 7 & 4 \\ 2 & 9 & 6 \\ 3 & 5 & 8 \end{pmatrix}, B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 3 \end{pmatrix}$
   - $A = \begin{pmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix}, B = A^\top$
2. Tìm A^n với $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ và n là một số nguyên dương.
3. Lập trình ngắn bằng Python để tính $\langle A, B \rangle$ với
   $A = \begin{pmatrix} 1 & 7 & 4 \\ 2 & 9 & 6 \\ 3 & 5 & 8 \end{pmatrix}, B = \begin{pmatrix} 2 & 1 & 9 \\ 4 & 7 & 5 \\ 6 & 8 & 3 \end{pmatrix}$ (Ưu tiên sử dụng các hàm từ thư viện NumPy).
   Nếu có thể, hãy thử lập trình một hàm nhận vào hai ma trận A, B cùng kích cỡ và trả về $\langle A, B \rangle$.
4. Tìm hiểu về bài toán luồng cực đại (max flow) trong lý thuyết đồ thị cũng như mối quan hệ của nó với bài toán OT.