Matrix scaling 1

Tuesday, September 28, 2021 7:27 PM

Matrix scaling

Problem: Given $A \in \mathbb{R}^{n \times n}_{\geq 0}$, scale the rows & columns to make all the row & column sums 1 (doubly stochastic).

$$\begin{array}{c|c} & \beta_1 \; \cdots \; \beta_n \\ \hline \alpha_1 & \\ \vdots & A \\ \alpha_n & \end{array}$$

$$\sum_{j=1}^{n} A_{ij} \alpha_i \beta_j = 1 \quad \forall i \in [n]$$

$$\sum_{i=1}^{n} A_{ij} \alpha_i \beta_j = 1 \quad \forall j \in [n]$$

$(\alpha_i, \beta_i \geq 0)$

Examples:

①
$$\begin{array}{c|cc} & \sqrt{2}-1 & 2-\sqrt{2} \\ \hline 1/\sqrt{2} & 2 & 1 \\ 1 & 1 & 1 \end{array} = \begin{bmatrix} 2-\sqrt{2} & \sqrt{2}-1 \\ \sqrt{2}-1 & 2-\sqrt{2} \end{bmatrix}$$

② $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \rightarrow$ cannot scale

③ If a row or column is all-zero, cannot scale.

④ $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \rightarrow$ this also cannot be scaled

⑤ $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \longrightarrow$ cannot be scaled (exactly)

$\begin{bmatrix} \cdot & \cdot \\ 1 & 0 \end{bmatrix} \longrightarrow$ cannot be scaled (exactly)

$$\begin{array}{c|cc} & 1 & 1/\varepsilon \\ \hline \varepsilon & 1 & 1 \\ 1 & 1 & 0 \end{array} \quad = \begin{bmatrix} \varepsilon & 1 \\ 1 & 0 \end{bmatrix} \xrightarrow[\varepsilon \to 0]{\lim} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Call such matrices "scalable in limit".

Orbits vs orbit-closures → similar to that

Question: When is $A \in \mathbb{R}_{\geq 0}^{n \times n}$ scalable in limit?

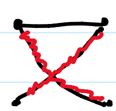
Theorem: TFAE

① $A \in \mathbb{R}_{\geq 0}^{n \times n}$ is scalable in limit.

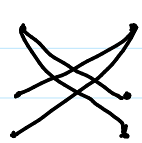
② $\mathrm{supp}(A)$ has a perfect matching.

($\mathrm{supp}(A)$ can viewed as an $n \times n$ bipartite graph)

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



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



→ no perfect matching

hence not scalable in limit.

Proof of theorem: ① ⇒ ②

Suppose A is scalable in limit. Let the limit be B.

$supp(B) \subseteq supp(A)$ & B is doubly stochastic.

Birkhoff - von Neumann theorem : A doubly stochastic matrix is a convex combination of permutations.

In particular, $supp(B)$ contains a perfect matching.

Connection to invariant theory

Consider the action of $ST_n(\mathbb{C}) \times ST_n(\mathbb{C})$ on $M_n(\mathbb{C})$.

$$ST_n(\mathbb{C}) = \left\{ diag(t_1, \dots, t_n) : \prod_i t_i = 1 \right\}$$

$$(A, B) \cdot M = AMB$$

Question : What is the null cone of this group action?

$$\text{Null cone} = \left\{ M \in M_n(\mathbb{C}) \text{ s.t. } \exists \text{ sequence } (A_1, B_1), \dots, (A_k, B_k), \dots \text{ s.t. } \lim_{k \to \infty} (A_k, B_k) \cdot M = 0 \right\}$$

Theorem : Let $M \in M_n(\mathbb{C})$. Let $A \in \mathbb{R}_{\geq 0}^{n \times n}$ given by $A_{ij} = |M_{ij}|^2$. TFAE.

1. A is scalable in limit.
2. $supp(A)$ contains a perfect matching.
3. M is not in the null cone of the above group action.

Proof: ② $\Rightarrow$ ③

supp(M) contains a perfect matching.

$\Rightarrow \exists \ \sigma \in S_n$ s.t. $\prod_{i=1}^{n} M_{i,\sigma(i)} \neq 0$

Let $(A,B) \in ST_n \times ST_n$ & let $M' = (A,B).M$

$$\prod_{i=1}^{n} M'_{i,\sigma(i)} = \prod_{i=1}^{n} M_{i,\sigma(i)} A_{ii} B_{\sigma(i),\sigma(i)}$$

$$= \left(\prod_{i=1}^{n} M_{i,\sigma(i)}\right)\left(\prod_i A_{i,i}\right)\left(\prod_i B_{\sigma(i),\sigma(i)}\right)$$

$$= \prod_{i=1}^{n} M_{i,\sigma(i)}$$

($\prod_{i=1}^{n} M_{i,\sigma(i)}$ is an invariant polynomial)

Consider $(A_1,B_1), \dots, (A_k,B_k), \dots$

Let $M^{(k)} := (A_k,B_k).M$

Then $\prod_i M^{(k)}_{i,\sigma(i)} = \prod_i M_{i,\sigma(i)}$

Let $M^* = \lim_{k\to\infty} M^{(k)}$. By continuity $\prod_i M^*_{i,\sigma(i)} = \prod_i M_{i,\sigma(i)} \neq 0$

$\Rightarrow M^* \neq 0$.

③ $\Rightarrow$ ②

Let us prove the contrapositive

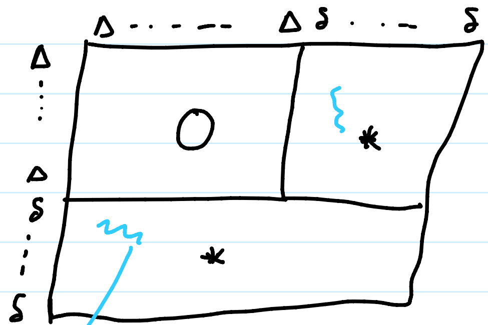
if supp(M) does not contain a perfect matching

$\Rightarrow$ M is in the null cone.

We will use Hall's theorem.

$\exists$ subsets $S, T \subseteq [n]$ s.t. $|S| + |T| \geq n+1$

& $M_{S,T} = 0$.



Let us ensure the product of all the scalars is 1.

$$\Delta^{|S|} \Delta^{|T|} \cdot \delta^{n-|S|} \delta^{n-|T|} = 1$$

$\delta\Delta = \frac{\delta}{\delta^c} = \delta^{1-c} \to 0$ as $\delta \to 0$

$$\Leftrightarrow \quad \Delta^{|S|+|T|} \delta^{2n-(|S|+|T|)} = 1$$

Because $|S| + |T| \geq n+1$, $|S| + |T| > 2n - (|S|+|T|)$

$$\Rightarrow \quad \Delta = \frac{1}{\delta^{\frac{2n-(|S|+|T|)}{|S|+|T|}}} = \frac{1}{\delta^c} \text{ for some } c < 1.$$

We have proven: $\exists$ sequence $(A_1, B_1), \dots (A_k, B_k), \dots$

s.t. ① $(A_k, B_k) \in T_n \times T_n$

② $\det(A_k) \det(B_k) = 1$

③ $\lim_{k\to\infty} A_k M B_k = 0$

$$\tilde{A}_k = \frac{A_k}{\det(A_k)^{1/n}}, \quad \tilde{B}_k = \frac{B_k}{\det(B_k)^{1/n}}$$

$$\tilde{A}_k M \tilde{B}_k = \frac{A_k M B_k}{(\det(A_k)\det(B_k))^{1/n}} = A_k M B_k$$

$$A_k M \tilde{B}_k = \frac{A_k M B_k}{(\det(A_k)\det(B_k))^{1/n}} = A_k M B_k$$

$\square$

Algorithms

Sinkhorn's algorithm: $A \in \mathbb{R}_{\geq 0}^{n \times n}$.

Row-normalize: Let $r_1, \ldots r_n$ be the row sums of A

$$\begin{array}{c|c} r_1^{-1} & \\ \vdots & A \\ r_n^{-1} & \end{array}$$

Column normalize: Let $c_1, \ldots c_n$ be the column sums of A

$$\begin{array}{ccc} c_1^{-1} & \cdots & c_n^{-1} \\ \hline & A & \end{array}$$

Sinkhorn's algorithm [64] Repeat for N steps:

(1) row-normalize

(2) column-normalize

output $\hat{A}$.

Definition: $$ds(A) := \sum_{i=1}^{n} (r_i - 1)^2 + \sum_{j=1}^{n} (c_j - 1)^2$$

$r_1, \ldots r_n$ are the row sums of A

$c_1, \ldots c_n$ are the column sums of A

Input model: $A \in \mathbb{Q}_{\geq 0}^{n \times n}$. Each entry $A_{ij} = \frac{p_{ij}}{q_{ij}}$

$p_{ij}, q_{ij} \in \mathbb{Z}_{\geq 0}$ & can be described using b bits.

$$\boxed{2^b \geq A_{ij} \geq 2^{-b}}$$

Theorem: Suppose $A \in \mathbb{Q}_{\geq 0}^{n \times n}$ s.t. $\mathrm{supp}(A)$ has a perfect matching. Then in $N = O\left(n(\log(n)+b)/\varepsilon\right)$ steps, the algorithm will output $\hat{A}$ s.t. $ds(\hat{A}) \leq \varepsilon$.

[Linial, Samorodnitsky, Wigderson 00]

Proof: $\exists\ \sigma \in S_n$ s.t. $\prod_i A_{i,\sigma(i)} \neq 0$

Wlog $\sigma = id$, so that $\prod_i A_{i,i} \neq 0$.

$A' = $ row-normalize(A). each $A'_{ii} \neq 0$

$$\Rightarrow \quad A'_{i,i} \geq \frac{1}{n\, 2^{2b}}$$

$$\Rightarrow \quad \prod_{i=1}^{n} A'_{i,i} \geq \left(\frac{1}{n\, 2^{2b}}\right)^n.$$

Lemma: Suppose A' is row (or column) normalized. & $\prod_i A'_{i,i} \neq 0$. Also $ds(A') \geq \varepsilon$.

Then column (or row) normalization increases

$\prod_i A'_{i,i}$ by a factor of $e^{\varepsilon/6}$.

Pf: Assume A' is row normalized.

$A'' :=$ column-normalize (A')

$c_1, \ldots c_n$ be the column sums of A'.

$$\prod_{i=1}^{n} A''_{i,i} = \prod_{i=1}^{n} A'_{i,i} \Big/ \prod_{i=1}^{n} c_i$$

Since A' is row-normalized $\Rightarrow \sum_{i=1}^{n} c_i = \sum_{j=1}^{n} r_j = n$

$$\Rightarrow \prod_{i=1}^{n} c_i \le \left(\frac{\sum_{i=1}^{n} c_i}{n}\right)^n = 1$$ (AM-GM inequality)

$$\sum_{i=1}^{n} (c_i - 1)^2 = ds(A') \geq \varepsilon$$

Exercise: Let $\lambda_1, \ldots \lambda_n \geq 0$ s.t. $\sum_{i=1}^{n} \lambda_i = n$

& $\sum_{i=1}^{n} (\lambda_i - 1)^2 \geq \varepsilon$.

$$\Rightarrow \prod_{i=1}^{n} \lambda_i \le e^{-\varepsilon/6}.$$

Start with $A' =$ row-normalize (A).

① $\prod_{i=1}^{n} A'_{i,i} \geq \left(\frac{1}{n\, 2^{2b}}\right)^n$.

② As long as $ds(A') \geq \varepsilon$, each operation will

(2) As long as $ds(A') \geq \varepsilon$, each operation will increase $\prod_{i=1}^{n} A'_{i,i}$ by a factor of $e^{\varepsilon/6}$.

(3) $\prod_{i=1}^{n} A'_{i,i} \leq 1$ if A' is row or column normalised.

$\Rightarrow$ in $O\left(\frac{n(\log(n)+b)}{\varepsilon}\right)$ steps, the algorithm will output $\hat{A}$ s.t. $ds(\hat{A}) \leq \varepsilon$.