Lecture 9, Friday, 4/15/2022

⊙ Sinkhorn's algorithm: Proof of Convergence.

Let $a \in \mathcal{P}_m$, $b \in \mathcal{P}_n$, $C \in \mathbb{R}^{m\times n}$, $C > 0$.

$\mathcal{A}(a,b) = \{ P \in \mathbb{R}^{m\times n} : P \geq 0,\ \sum_j P_{ij} = a_i\ \forall i,\ \sum_i P_{ij} = b_j\ \forall j \}$.

Let $\varepsilon > 0$. Define

$$E_\varepsilon[P] = \langle P, C \rangle + \varepsilon \langle P(\log P - 1), \mathbb{1}_{m\times n} \rangle .$$

Entropically regularized OT prob: $\min_{P \in \mathcal{A}(a,b)} E_\varepsilon[P]$.

$$\mathcal{L}(P, \lambda) \overset{=(f,g)}{=} E_\varepsilon[P] - \sum_i f_i \Big(\sum_j P_{ij} - a_i\Big) - \sum_j g_j \Big(\sum_i P_{ij} - b_j\Big).$$

$\partial_{f_i} \mathcal{L} = 0 \Longrightarrow$ row sum of $P = a$,

$\partial_{g_j} \mathcal{L} = 0 \Longrightarrow$ col sum of $P = b$.

$$\frac{\partial \mathcal{L}}{\partial P_{kl}} = 0 \Longrightarrow P_{kl} = e^{-(C_{kl} - f_k - g_l)/\varepsilon} = \underbrace{e^{f_k/\varepsilon}}_{u_k} \underbrace{e^{-C_{kl}/\varepsilon}}_{K_{kl}} \underbrace{e^{g_l/\varepsilon}}_{v_l}$$

$P = \mathrm{diag}(u)\, K\, \mathrm{diag}(v)$. i.e. $P_{ij} = u_i K_{ij} v_j$.

Assume $a > 0$, $b > 0$.

<u>Sinkhorn's algorithm</u> Choose $v^{(0)} \in \mathbb{R}^n$, $v^{(0)} > 0$.

For $k = 1, 2, \cdots$: $u^{(k)} = \dfrac{a}{K v^{(k-1)}} \in \mathbb{R}^m$, $v^{(k)} = \dfrac{b}{K^T u^{(k)}} \in \mathbb{R}^n$. (*)

Set $A^{(k)} = \mathrm{diag}(u^{(k)})\, K\, \mathrm{diag}(v^{(k-1)})$, $k = 1, 2, \cdots$.

$B^{(k)} = \mathrm{diag}(u^{(k)})\, K\, \mathrm{diag}(v^{(k)})$,

Then $A^{(k)} \to P_\varepsilon = \operatorname{argmin}_{\mathcal{A}(a,b)} E_\varepsilon[\cdot]$, and $B^{(k)} \to P_\varepsilon$. (This will be proved.). So, $A^{(k)}$ or $B^{(k)}$ for $k \gg 1$ approximates $P_\varepsilon$.

<u>Remark</u> We can choose $u^{(0)} \in \mathbb{R}^m$, $u^{(0)} > 0$ and define

$v^{(k)} = \dfrac{b}{K^T u^{(k-1)}}$, $u^{(k)} = \dfrac{a}{K v^{(k)}}$, $k = 1, 2, \cdots$. (**)

<u>Theorem</u> (Convergence of Sinkhorn's algorithm) Let $K \in \mathbb{R}^{m\times n}$, $a \in \mathcal{P}_m$, and $b \in \mathcal{P}_n$ with $K > 0$, $a > 0$, and $b > 0$. Define

$u^{(k)}, v^{(k)}$ $(k=1,2,\dots)$ by $(*)$ (or $**$). Then $\exists\, u \in \mathbb{R}^m_+$ and $v \in \mathbb{R}^n_+$ such that $u_i^{(k)} v_j^{(k)} \to u_i v_j$ $\forall i,j$. Moreover
$A^{(k)} := \mathrm{diag}(u^{(k)}) K \,\mathrm{diag}(v^{(k-1)}) \to \mathrm{diag}(u) K \,\mathrm{diag}(v) =: P \in \mathcal{A}(a,b)$.
and $B^{(k)} := \mathrm{diag}(u^{(k)}) K \,\mathrm{diag}(v^{(k)}) \to P \in \mathcal{A}(a,b)$.

<u>Question</u> Do we have $u^{(k)} \to u$, $v^{(k)} \to v$ (Euclid dist) for some $u \in \mathbb{R}^m_+$, $v \in \mathbb{R}^n_+$ ?

<u>Let's first understand the iteration process</u>: normalizing process!

$v^{(0)}$
$\downarrow$
$u^{(1)} = \frac{a}{Kv^{(0)}}$, $A_{ij}^{(1)} = u_i^{(1)} K_{ij} v_j^{(0)}$, row-$i$ sum $A^{(1)} = a_i$
$\downarrow$
$v^{(1)} = \frac{b}{K^T u^{(1)}}$, $B_{ij}^{(1)} = u_i^{(1)} K_{ij} v_j^{(1)}$, col-$j$ sum $B^{(1)} = b_j$, $B_{ij}^{(1)} = A_{ij}^{(1)} / \mu_j^{(1)}$, $\mu_j^{(1)} = \frac{v_j^{(0)}}{v_j^{(1)}}$
$\downarrow$
$u^{(2)} = \frac{a}{Kv^{(1)}}$, $A_{ij}^{(2)} = u_i^{(2)} K_{ij} v_j^{(1)}$, row-$i$ sum $A^{(2)} = a_i$, $A_{ij}^{(2)} = B_{ij}^{(1)} / \lambda_i^{(1)}$, $\lambda_i^{(1)} = \frac{u_i^{(1)}}{u_i^{(2)}}$
$\downarrow$
$v^{(2)} = \frac{b}{K^T u^{(2)}}$, $B^{(2)} = u_i^{(2)} K_{ij} v_j^{(2)}$, col-$j$ sum $B^{(2)} = b_j$, $B_{ij}^{(2)} = A_{ij}^{(2)} / \mu_j^{(2)}$, $\mu_j^{(2)} = \frac{v_j^{(1)}}{v_j^{(2)}}$.
$\downarrow$

What are $\lambda_i^{(k)}$ and $\mu_j^{(k)}$ $\forall i,j,k$ ?

$$\mu_j^{(1)} = \frac{v_j^{(0)}}{v_j^{(1)}} = v_j^{(0)} \frac{(K^T u^{(1)})_j}{b_j} = \frac{1}{b_j} \sum_i u_i^{(1)} K_{ij} v_j^{(0)}$$
$$= \frac{1}{b_j} \sum_i A_{ij}^{(1)} = \frac{1}{b_j} \text{ col-}j \text{ sum } A^{(1)}.$$
$$\lambda_i^{(1)} = \frac{u_i^{(1)}}{u_i^{(2)}} = u_i^{(1)} \frac{1}{a_i} \sum_j K_{ij} v_j^{(1)} = \frac{1}{a_i} \sum_j u_i^{(1)} K_{ij} v_j^{(1)}$$
$$= \frac{1}{a_i} \sum B_{ij}^{(1)} = \frac{1}{a_i} \text{ row-}i \text{ sum } B^{(1)}$$

$\odot$ row sum $A^{(k+1)} = a$, col. sum $B^{(k)} = b$. $\left(\Rightarrow \sum_{i,j} A_{ij} = \sum_{i,j} B_{ij} = 1\right)$

$\odot$ $B_{ij}^{(k)} = u_i^{(k)} K_{ij} v_j^{(k)} = A_{ij}^{(k)} / \mu_j^{(k)}$, $A_{ij}^{(k+1)} = u_i^{(k+1)} K_{ij} v_j^{(k)} = B_{ij}^{(k)} / \lambda_i^{(k)}$,

$\mu_j^{(k)} = \frac{1}{b_j}$ col-$j$ sum $A^{(k)}$, $\lambda_i^{(k)} = \frac{1}{a_i}$ row-$i$ sum $B^{(k)}$.

$\Rightarrow \sum_j b_j \mu_j^{(k)} = \sum_{i,j} A_{ij}^{(k)} = 1$, $\sum_i a_i \lambda_i^{(k)} = \sum_{i,j} B_{ij}^{(k)} = 1$.

$$
A^{(k)} \qquad \begin{array}{c|cccc|l} & & & j & & \\ i & \cdot & \cdot & \cdot & \cdot & \text{row-}i\text{ sum} = a_i \\ & \cdot & \cdot & \cdot & \cdot & \\ & \cdot & \cdot & \cdot & \cdot & \\ \hline & \multicolumn{4}{c|}{\frac{1}{b_j}\,\text{col-}j\text{ sum} = \mu_j^{(k)}} & \end{array}
\qquad\qquad
B^{(k)} \qquad \begin{array}{c|cccc|l} & & & j & & \\ i & \cdot & \cdot & \cdot & \cdot & \frac{1}{a_i}\,\text{row-}i\text{ sum} = \lambda_i^{(k)} \\ & \cdot & \cdot & \cdot & \cdot & \\ & \cdot & \cdot & \cdot & \cdot & \\ \hline & \multicolumn{4}{c|}{\text{col-}j\text{ sum} = b_j} & \end{array}
$$

$$\lambda_i^{(k)} = \frac{u_i^{(k)}}{u_i^{(k+1)}}. \quad \lambda_i^{(1)} \cdots \lambda_i^{(k)} = u_i^{(1)} / u_i^{(k+1)} \quad (k=1,2,\cdots)$$

$$\mu_j^{(0)} = \frac{v_j^{(0)}}{v_j^{(1)}}, \quad \mu_j^{(k)} = \frac{v_j^{(k-1)}}{v_j^{(k)}}, \quad \mu_j^{(1)} \cdots \mu_j^{(k)} = \frac{v_j^{(0)}}{v_j^{(k)}} \quad (k=1,2,\cdots)$$

Denote $x_i^{(k)} = \left(\prod_{p=1}^{k} \lambda_i^{(p)}\right)^{-1}$ and $y_j^{(k)} = \left(\prod_{p=1}^{k} \mu_j^{(p)}\right)^{-1}$

Then $u_i^{(k+1)} = u_i^{(1)} x_i^{(k)}$ and $v_j^{(k)} = v_j^{(0)} y_j^{(k)}$ $(k=1,2,\cdots)$

$$\odot \quad A_{ij}^{(k)} = \frac{1}{\lambda_i^{(k-1)}} B_{ij}^{(k-1)} = \frac{1}{\lambda_i^{(k-1)}} A_{ij}^{(k-1)} \frac{1}{\mu_j^{(k-1)}} = \cdots = x_i^{(k-1)} A_{ij}^{(1)} y_j^{(k-1)}.$$

$$B_{ij}^{(k)} = \frac{1}{\mu_j^{(k)}} A_{ij}^{(k)} = \frac{1}{\mu_j^{(k)}} B_{ij}^{(k-1)} \frac{1}{\lambda_i^{(k-1)}} = \cdots = x_i^{(k-1)} A_{ij}^{(1)} y_j^{(k)}.$$

<u>Proof of Thm</u> Define $A_{ij}^{(k)}, B_{ij}^{(k)}, \lambda_i^{(k)}, \mu_j^{(k)}, x_i^{(k)}, y_j^{(k)}$ as above.

<u>Step 1</u> Show all $\lambda^{(k)}(m)$, $\lambda^{(k)}(M)$, $\mu^{(k)}(m)$, $\mu^{(k)}(M)$ all converge. Note that

$$\sum_j b_j = 1 \Rightarrow \begin{array}{l} \min_j \mu_j^{(k)} =: \mu^{(k)}(m) \le 1, \\ \max_j \mu_j^{(k)} =: \mu^{(k)}(M) \ge 1, \end{array} \qquad \sum_i a_i = 1 \Rightarrow \begin{array}{l} \min_i \lambda_i^{(k)} = \lambda^{(k)}(m) \le 1. \\ \max_i \lambda_i^{(k)} = \lambda^{(k)}(M) \ge 1. \end{array}$$

Since col-$j$ sum $B^{(k-1)} = b_j$,

$$\mu_j^{(k)} = \frac{1}{b_j}\,\text{col-}j\text{ sum } A^{(k)} = \frac{1}{b_j} \sum_i B_{ij}^{(k-1)} / \lambda_i^{(k-1)}$$

is a convex combination of $1/\lambda_i^{(k-1)}$. Similarly, $\lambda_i^{(k)}$ is a convex combination of $1/\mu_j^{(k)}$. So,

$$\min_i \lambda_i^{(k)} = \lambda_{i_m}^{(k)} \le 1 \le \lambda_{i_M}^{(k)} = \max_i \lambda_i^{(k)}.$$

$$\Rightarrow \lambda^{(k)}(m) \le \lambda^{(k+1)}(m) \le 1 \le \lambda^{(k+1)}(M) \le \lambda^{(k)}(M)$$

Similarly, $\mu^{(k)}(m) \le \mu^{(k+1)}(m) \le 1 \le \mu^{(k+1)}(M) \le \mu^{(k)}(M)$

So, all $\lambda^{(k)}(m)$, $\lambda^{(k)}(M)$, $\mu^{(k)}(m)$, and $\mu^{(k)}(M)$ converge.

Show they all equal 1 in step 3.

<u>Step 2</u> Show $\inf_{i,j,k}(A_{ij}^{k}, B_{ij}^{(k)}) =: \sigma > 0$.

Col. sum $B^{(k)} = b \Rightarrow \sum_i B_{ij}^{(k)} = \sum_i x_i^{(k-1)} A_{ij}^{(1)} y_j^{(k)} = b_j$

$$\Rightarrow y_j^{(k)} = b_j / \sum_i x_i^{(k-1)} A_{ij}^{(1)} \le b_j / A_{ij}^{(1)} x_i^{(k-1)}$$
$$\le b_j / \sigma^{(1)} x_i^{(k-1)} \quad \forall i, j,$$

where $\sigma^{(1)} = \min_{i,j}(A_{ij}^{(1)}, B_{ij}^{(1)}) > 0$. So, $y_j^{(k)} \le \dfrac{b_j}{\sigma^{(1)} x^{(k-1)}(M)}$, $\forall j$.

Now, $\frac{1}{a_i}\lambda_i^{(k)} =$ row-$i$ sum of $B^{(k)}$

$$= \sum_j x_i^{(k-1)} A_{ij}^{(1)} y_j^{(k)} = x_i^{(k-1)} \sum_j A_{ij}^{(1)} y_j^{(k)}.$$

Since $\lambda_i^{(k)} \ge \lambda^{(k)}(m) \ge \lambda^{(1)}(m)$,

$$x_i^{(k-1)} \ge \frac{\lambda_i^{(k)}}{a_i} \Big/ \sum_j A_{ij} y_j \ge \frac{\lambda^{(1)}(m)}{a_i} \Big/ \left( n \|A^{(1)}\|_\infty b_j / \sigma^{(1)} x^{(k-1)}(M) \right)$$

$$\ge \frac{\lambda^{(1)}(m)\, \sigma^{(1)}}{n \|a\|_\infty \|b\|_\infty \|A^{(1)}\|_\infty} x^{(k-1)}(M), \quad \forall i.$$

Again, col-$j$ sum $B^{(k)} = b_j \Rightarrow y_j^{(k)} = b_j \Big/ \sum_i x_i^{(k-1)} A_{ij}^{(1)}$.

$\Rightarrow y_j^{(k)} \geq b_j / m \|A^{(1)}\|_\infty x^{(k-1)}(M)$. Hence,

$$B_{ij}^{(k)} = x_i^{(k-1)} A_{ij}^{(1)} y_j^{(k)} \geqslant \frac{\lambda^{(1)}(m)(\sigma^{(1)})^2 b(m)}{m\, n \|a\|_\infty \|b\|_\infty \|A^{(1)}\|_\infty^2} > 0,$$

where $b(m) = \min_i b_i > 0$.

Thus $\inf_{i,j,k} B_{ij}^{(k)} > 0$. Similarly, $\inf_{i,j,k} A_{ij}^{(k)} > 0$.

<u>Step 3</u> Show the limits all equal 1.

Clearly, $\lambda^{(k)}(M) = \max_i \lambda^{(k)} \searrow 1 + c$ for some $c \geq 0$.

Let $\lambda^{(k)}(M) = 1 + c_k$ $(k = 1, 2, \cdots)$. So, $c_k \downarrow c$ by Step 1. Since $\lambda_i^{(k)} \leq \lambda^{(k)}(M) = 1 + c_k$, and col-$j$ sum $B^{(k)} = b$, we have

$$\mu_j^{(k+1)} = \frac{1}{b_j} \text{ col-}j \text{ sum } A^{(k)} = \frac{1}{b_j} \sum_i B_{ij}^{(k)} / \lambda_i^{(k)}$$

$$\geq \frac{1}{b_j} \sum_{i:\lambda_i^{(k)} \leq 1} B_{ij}^{(k)} + \frac{1}{b_j} \sum_{i:\lambda_i^{(k)} \geq 1} B_{ij}^{(k)} \frac{1}{1 + c_k}$$

$$= \frac{1}{b_j(1 + c_k)} \left( \sum_{\text{all } i} B_{ij}^{(k)} + c_k \sum_{i:\lambda_i^{(k)} \leq 1} B_{ij}^{(k)} \right)$$

$$\overset{\text{Step 2}}{\geqslant} \frac{b_j + c_k b_j(\sigma / b_j)}{b_j(1 + c_k)} = \frac{1 + c_k(\sigma / \|b\|_\infty)}{1 + c_k}.$$

Let $i_0$ be such that $\lambda^{(k+1)}(M) = \lambda_{i_0}^{(k+1)}$. Then,

$$1 + c \leq \lambda^{(k+1)}(M) = \lambda_{i_0}^{(k+1)} = \frac{1}{a_{i_0}} \text{ row-}i_0 \text{ sum of } B^{(k+1)}$$

$$= \frac{1}{a_{i_0}} \sum_j B_{i_0 j}^{(k+1)} = \frac{1}{a_{i_0}} \sum_j A_{i_0 j}^{(k+1)} / \mu_j^{(k+1)}$$

$$= \frac{1}{a_{i_0}} \sum_j B_{i_0 j}^{(k)} / \lambda_{i_0 j}^{(k)} \mu_j^{(k+1)}$$
$$\le \frac{1}{a_{i_0}} \left( \sum_j B_{i_0 j}^{(k)} / \lambda_{i_0 j}^{(k)} \right) \cdot \frac{1+c_k}{1+c_k(\sigma/\|b\|_\infty)}$$
$$= \frac{1+c_k}{1+c_k(\sigma/\|b\|_\infty)} \longrightarrow \frac{1+c}{1+c(\sigma/\|b\|_\infty)}$$

Hence $c=0$, and $\lambda^{(k)}(M) \to 1$.

Since $a_i \lambda_i^{(k)} =$ row-$i$ sum of $B^{(k)} = \sum_j B_{ij}^{(k)}$ and the col. sum of $B^{(k)}$ is $b$, we get

$$\sum_i a_i \lambda_i^{(k)} = \sum_i \sum_j B_{ij}^{(k)} = \sum_j \sum_i B_{ij}^{(k)} = \sum_j b_j = 1.$$

Thus, if $\lambda^{(k)}(m)$ $(\le 1)$ does not converge to 1, then since $\lambda^{(k)}(M) \to 1$, $\liminf_{k\to\infty} \sum_i a_i \lambda_i^{(k)} < \sum_i a_i = 1$, a contradiction. Hence, $\lambda^{(k)}(m) \to 1$.

Similarly, $\mu^{(k)}(M) \to 1$, and $\mu^{(k)}(m) \to 1$.

Step 4 Show $\exists u \in \mathbb{R}_+^m$ and $v \in \mathbb{R}_+^n$ s.t. $u_i^{(k)} v_j^{(k)} \to u_i v_j$ $\forall i,j$. These will imply that $u_i^{(k)} K_{ij} v_j^{(k)} \to u_i K_{ij} v_j$, $\forall i,j$, i.e.,

$B^{(k)} = \mathrm{diag}(u^{(k)}) K \,\mathrm{diag}(v^{(k)}) \to P = \mathrm{diag}(u) K \,\mathrm{diag}(v)$

Since col. sum of $B^{(k)} = b$, col. sum of $P = b$. But $B_{ij}^{(k)} = A_{ij}^{(k)} / \mu_j^{(k)}$ $\forall i,j,k$ and $\mu_j^{(k)} \to 1$ $\forall j$ by Step 3, so, $A^{(k)} \to P$. Since row sum of $A^{(k)} = a$, row sum of $P = a$, and $P \in \mathcal{A}(a,b)$. This will complete the proof.

Note that all $B_{ij}^{(k)} = u_i^{(k)} K_{ij} v_j^{(k)} \ge \sigma$ $\forall i,j,k$ (cf. Step 2). Moreover $\sum_{ij} B_{ij}^{(k)} = 1$ $\forall k$. Suppose $\{B_1^{(k)}\}$ and $\{B_2^{(k)}\}$ are two subsequences of $\{B^{(k)}\}$ such that $B_1^{(k)} \to B_1$ and $B_2^{(k)} \to B_2$. Since row-$i$ sum of $B^{(k)} = a_i \lambda_i^{(k)} \to a_i$ as $k \to \infty$, and col. sum of $B^{(k)} = b$, $B_1, B_2 \in \mathcal{A}(a,b)$. Now,

$B_1^{(k)} = (\hat{u}_i^{(k)} K_{ij} \hat{v}_j^{(k)})$ and $B_2^{(k)} = (\hat{\hat{u}}_i^{(k)} K_{ij} \hat{\hat{v}}_j)$, where $\{\hat{u}^{(k)}\}, \{\hat{\hat{u}}^{(k)}\}, \{\hat{v}^{(k)}\}, \{\hat{\hat{v}}^{(k)}\}$ are the corresponding subseq. of $\{u^{(k)}\}, \{v^{(k)}\}$, respectively. Since $B_1^{(k)} \to B_1$ and $B_2^{(k)} \to B_2$, $\{\hat{u}_i^{(k)} \hat{v}_j^{(k)}\}$ and $\{\hat{\hat{u}}_i^{(k)} \hat{\hat{v}}_j^{(k)}\}$ converge to some positive numbes (as all $B_{ij}^{(k)} \ge \sigma$). Now, by the lemma below, $\exists\, \hat{w}^{(k)}, \hat{\hat{w}}^{(k)} \in \mathbb{R}_+^m$ and $\exists\, \hat{z}^{(k)}, \hat{\hat{z}}^{(k)} \in \mathbb{R}_+^n$ such that $\hat{w}^{(k)} \to \hat{w}$, $\hat{z}^{(k)} \to \hat{z}$, $\hat{\hat{w}}^{(k)} \to \hat{\hat{w}}$, $\hat{\hat{z}} \to \hat{\hat{z}}$ for some $\hat{w}, \hat{\hat{w}} \in \mathbb{R}_+^m$ and $\hat{z}, \hat{\hat{z}} \in \mathbb{R}_+^n$. Thus,

$$B_{1,ij} = \hat{w}_i K_{ij} \hat{z}_j, \quad B_{2ij} = \hat{\hat{w}}_i K_{ij} \hat{\hat{z}}_j, \quad \forall i,j.$$

Since $B_1, B_2 \in \mathcal{A}(a,b)$, the uniqueness of matrix equivalence (cf. Lecture 8) implies ~~that~~ $B_1 = B_2$. Thus, $B^{(k)}$ converges to some $P$. $P > 0$ as $B_{ij}^{(k)} \ge \sigma > 0$ $\forall i,j,k$. Moreover, since $B_{ij}^{(k)} = u_i^{(k)} K_{ij} v_j^{(k)}$, $u_i^{(k)} v_j^{(k)}$ converges for any $i,j$. By the lemma again, $\exists\, u \in \mathbb{R}_+^m$, $v \in \mathbb{R}_+^n$ s.t. $u_i^{(k)} v_j^{(k)} \to u_i v_j$ $\forall i,j$. <u>QED</u>

<u>Remark</u> Steps 1–3 are from Sinkhorn's 1964 paper. Step 4, and the lemma below, are from Sinkhorn-Knopp 1967 paper.

<u>Lemma</u> Suppose $u^{(k)} \in \mathbb{R}_+^m$ and $v^{(k)} \in \mathbb{R}_+^n$ such that $u_i^{(k)} v_j^{(k)} \to l_{ij} > 0$ for some $l_{ij}$ for all $i,j$. Then, $\exists\, \tilde{u}^{(k)} \in \mathbb{R}_+^m$, $\tilde{v}^{(k)} \in \mathbb{R}_+^n$ s.t. $\tilde{u}_i^{(k)} \tilde{v}_j^{(k)} = u_i^{(k)} v_j^{(k)}$ $\forall i,j$, $\tilde{u}^{(k)} \to u$, $\tilde{v}^{(k)} \to v$ for some $u \in \mathbb{R}_+^m$ and $v \in \mathbb{R}_+^n$, and $u_i v_j = l_{ij}$, $\forall i,j$.

**Proof** Set $\tilde{u}_i^{(k)} = \dfrac{u_i^{(k)}}{u_1^{(k)}}$, $\tilde{v}_j^{(k)} = u_1^{(k)} v_j^{(k)}$ $\forall i, j, k$

Then $\tilde{u}_i^{(k)} \tilde{v}_j^{(k)} = u_i^{(k)} v_j^{(k)}$ $\forall i, j, k$.

$\tilde{u}_1^{(k)} = 1 =: u_1$. $\tilde{v}_j^{(k)} = u_1^{(k)} v_j^{(k)} \to l_{1j} =: v_j$ $\forall j$.

For $i \geq 2$: $\tilde{u}_i^{(k)} = \dfrac{u_i^{(k)}}{u_1^{(k)}} \cdot \dfrac{v_1^{(k)}}{v_1^{(k)}} \longrightarrow \dfrac{l_{i1}}{l_{11}} =: u_i$.

So, $u = (u_i) \in \mathbb{R}_+^m$, $v = (v_j) \in \mathbb{R}_+^n$. $\tilde{u}^{(k)} \to u$, $\tilde{v}^{(k)} \to v$,

and $u_i v_j = \lim_{k \to \infty} \tilde{u}_i^{(k)} \tilde{v}_j^{(k)} = \lim_{k \to \infty} u_i^{(k)} v_j^{(k)} = l_{ij}$ $\forall i, j$.

QED