Supplementary Material for the Submission "A Ranking Model Motivated by Nonnegative Matrix Factorization with Applications to Tennis Tournaments"

Rui Xia[1], Vincent Y. F. Tan[1], Louis Filstroff[2], and Cédric Févotte[2]

[1]Department of Mathematics, National University of Singapore
rui.xia@u.nus.edu, vtan@nus.edu.sg
[2]IRIT, Université de Toulouse, CNRS, France
{louis.filstroff, cedric.fevotte}@irit.fr

S-1 Proof that likelihood is non-decreasing after truncation to zero

We first simplify the term $f_\epsilon(\mathbf{W}, \tilde{\mathbf{H}}^{(l)}) - u_2(\tilde{\mathbf{H}}^{(l+1)}, \tilde{\mathbf{H}}^{(l)}|\mathbf{W})$ by writing out these terms as follows. Firstly, for $f_\epsilon(\mathbf{W}, \tilde{\mathbf{H}}^{(l)})$, we have

$$\begin{aligned}
&f_\epsilon(\mathbf{W}, \tilde{\mathbf{H}}^{(l)}) \\
&= \sum_{m=1}^{M} \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \Big[-\log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi}) + \log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mj}) \Big] \\
&= \sum_{m} \sum_{\substack{(i',j'):\\ i',j'\neq i}} b_{i'j'}^{(m)} \Big[-\log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi'}) + \log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi'} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mj'}) \Big] \\
&\quad + \sum_{m} \sum_{j\neq i:(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \Big[-\log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi}) \Big] \\
&\quad + \sum_{m} \sum_{j\neq i:(i,j)\in\mathcal{P}_m} (b_{ij}^{(m)} + b_{ji}^{(m)}) \Big[\log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mj}) \Big]
\end{aligned} \tag{S-1}$$

Next for $u_2(\tilde{\mathbf{H}}^{(l+1)}, \tilde{\mathbf{H}}^{(l)}|\mathbf{W})$, we have

$$\begin{aligned}
&u_2(\tilde{\mathbf{H}}^{(l+1)}, \tilde{\mathbf{H}}^{(l)}|\mathbf{W}) \\
&= \sum_{m=1}^{M} \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \Big[-\sum_k \frac{w_{mk}(\tilde{h}_{ki}^{(l)} + \epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi}} \log\left(\frac{\tilde{h}_{ki}^{(l+1)} + \epsilon}{\tilde{h}_{ki}^{(l)} + \epsilon} [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi} \right) \\
&\quad + \log\left([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mj} \right) + \frac{\sum_k w_{mk}(\tilde{h}_{ki}^{(l+1)} + \tilde{h}_{kj}^{(l+1)} + 2\epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mj}} - 1 \Big] \\
&= \sum_{m=1}^{M} \sum_{\substack{(i',j'):\\ i',j'\neq i}} b_{i'j'}^{(m)} \Big[-\sum_k \frac{w_{mk}(\tilde{h}_{ki'}^{(l)} + \epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi'}} \log\left(\frac{\tilde{h}_{ki'}^{(l)} + \epsilon}{\tilde{h}_{ki'}^{(l)} + \epsilon} [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi'} \right) \\
&\quad + \log\left([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi'} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mj'} \right) + \frac{\sum_k w_{mk}(h_{ki'}^{(l)} + h_{kj'}^{(l)} + 2\epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi'} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mj'}} - 1 \Big] \\
&\quad + \sum_{m=1}^{M} \sum_{j\neq i:(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \Big[-\sum_k \frac{w_{mk}(\tilde{h}_{ki}^{(l)} + \epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi}} \log\left(\frac{\tilde{h}_{ki}^{(l+1)} + \epsilon}{\tilde{h}_{ki}^{(l)} + \epsilon} [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi} \right) \Big] \\
&\quad + \sum_{m=1}^{M} \sum_{j\neq i:(i,j)\in\mathcal{P}_m} (b_{ij}^{(m)} + b_{ji}^{(m)}) \Big[\log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon\mathbb{1})]_{mj}) + \frac{\sum_k w_{mk}(\tilde{h}_{ki}^{(l+1)} + \tilde{h}_{kj}^{(l+1)} + 2\epsilon)}{\sum_k w_{mk}(\tilde{h}_{ki}^{(l)} + \tilde{h}_{kj}^{(l)} + 2\epsilon)} - 1 \Big]
\end{aligned}$$

$$= \sum_{m=1}^{M} \sum_{\substack{(i',j'):\\ i',j' \neq i}} b_{i'j'}^{(m)} \Big[-\log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi'}) + \log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi'} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mj'}) \Big] \tag{S-2}$$

$$+ \sum_{m=1}^{M} \sum_{\substack{j \neq i:\\ (i,j) \in \mathcal{P}_m}} b_{ij}^{(m)} \Big[-\sum_{k' \neq k} \frac{w_{mk'}(\tilde{h}_{k'i}^{(l)} + \epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}} \log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi})$$

$$- \frac{w_{mk}(\tilde{h}_{ki}^{(l)} + \epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}} \log(\frac{\epsilon}{\tilde{h}_{ki}^{(l)} + \epsilon} [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}) \Big]$$

$$+ \sum_{m=1}^{M} \sum_{\substack{j \neq i:\\ (i,j) \in \mathcal{P}_m}} (b_{ij}^{(m)} + b_{ji}^{(m)}) \log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mj})$$

$$+ \sum_{m=1}^{M} \sum_{\substack{j \neq i:\\ (i,j) \in \mathcal{P}_m}} (b_{ij}^{(m)} + b_{ji}^{(m)}) \Big[\frac{\sum_{k' \neq k} w_{mk'}(\tilde{h}_{ki}^{(l)} + \epsilon) + w_{mk}(\epsilon) + \sum_{k} w_{mk}(\tilde{h}_{kj}^{(l)} + \epsilon)}{\sum_k w_{mk}(\tilde{h}_{ki}^{(l)} + \tilde{h}_{kj}^{(l)} + 2\epsilon)} - 1 \Big]$$

When we calculate $f_\epsilon(\mathbf{W}, \tilde{\mathbf{H}}^{(l)}) - u_2(\tilde{\mathbf{H}}^{(l+1)}, \tilde{\mathbf{H}}^{(l)} | \mathbf{W})$, the expressions in (S-1) and (S-2) cancel and we have:

$$\begin{aligned}
& f_\epsilon(\mathbf{W}, \tilde{\mathbf{H}}^{(l)}) - u_2(\tilde{\mathbf{H}}^{(l+1)}, \tilde{\mathbf{H}}^{(l)} | \mathbf{W}) \\
&= \sum_m \sum_{j \neq i : (i,j) \in \mathcal{P}_m} b_{ij}^{(m)} \Big[-\log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}) \Big] \\
&+ \sum_m \sum_{j \neq i : (i,j) \in \mathcal{P}_m} (b_{ij}^{(m)} + b_{ji}^{(m)}) \Big[\log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mj}) \Big] \\
&- \sum_{m=1}^{M} \sum_{\substack{j \neq i:\\ (i,j) \in \mathcal{P}_m}} b_{ij}^{(m)} \Big[-\sum_{k' \neq k} \frac{w_{mk'}(\tilde{h}_{k'i}^{(l)} + \epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}} \log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}) \\
&\qquad - \frac{w_{mk}(\tilde{h}_{ki}^{(l)} + \epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}} \log(\frac{\epsilon}{\tilde{h}_{ki}^{(l)} + \epsilon} [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}) \Big] \\
&- \sum_{m=1}^{M} \sum_{\substack{j \neq i:\\ (i,j) \in \mathcal{P}_m}} (b_{ij}^{(m)} + b_{ji}^{(m)}) \log([\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mj}) \\
&- \sum_{m=1}^{M} \sum_{\substack{j \neq i:\\ (i,j) \in \mathcal{P}_m}} (b_{ij}^{(m)} + b_{ji}^{(m)}) \Big[\frac{\sum_{k' \neq k} w_{mk'}(\tilde{h}_{ki}^{(l)} + \epsilon) + w_{mk}(\epsilon) + \sum_{k} w_{mk}(\tilde{h}_{kj}^{(l)} + \epsilon)}{\sum_k w_{mk}(\tilde{h}_{ki}^{(l)} + \tilde{h}_{kj}^{(l)} + 2\epsilon)} - 1 \Big] \\
&= \sum_{m=1}^{M} \sum_{\substack{j \neq i:\\ (i,j) \in \mathcal{P}_m}} b_{ij}^{(m)} \Big[\frac{w_{mk}(\tilde{h}_{ki}^{(l)} + \epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}} \log(\frac{\epsilon}{\tilde{h}_{ki}^{(l)} + \epsilon}) \Big] \\
&+ \sum_{m=1}^{M} \sum_{\substack{j \neq i:\\ (i,j) \in \mathcal{P}_m}} (b_{ij}^{(m)} + b_{ji}^{(m)}) \Big[\frac{w_{mk} \tilde{h}_{ki}^{(l)}}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mj}} \Big]
\end{aligned}$$

Using the inequality when truncation to zero is invoked, i.e.,

$$\frac{\sum_m \sum_{j \neq i} b_{ij}^{(m)} \frac{w_{mk}(\tilde{h}_{ki}^{(l)} + \epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi}}}{\sum_m \sum_{j \neq i} (b_{ij}^{(m)} + b_{ji}^{(m)}) \frac{w_{mk}}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mi} + [\mathbf{W}(\tilde{\mathbf{H}}^{(l)} + \epsilon \mathbb{1})]_{mj}}} \leq \epsilon$$

we obtain

$$
\begin{aligned}
& f_\epsilon(\mathbf{W}, \tilde{\mathbf{H}}^{(l)}) - u_2(\tilde{\mathbf{H}}^{(l+1)}, \tilde{\mathbf{H}}^{(l)}|\mathbf{W}) \\
&= -\log\Big(\frac{\tilde{h}_{ki}^{(l)}+\epsilon}{\epsilon}\Big)\Bigg[\sum_{m=1}^{M}\sum_{\substack{j\neq i:\\(i,j)\in\mathcal{P}_m}} b_{ij}^{(m)}\frac{w_{mk}(\tilde{h}_{ki}^{(l)}+\epsilon)}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mi}}\Bigg] \\
&\quad + \sum_{m=1}^{M}\sum_{\substack{j\neq i:\\(i,j)\in\mathcal{P}_m}} (b_{ij}^{(m)}+b_{ji}^{(m)})\Bigg[\frac{w_{mk}\tilde{h}_{ki}^{(l)}}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mj}}\Bigg] \\
&\geq -\log\Big(\frac{h_{ki}^{(l)}+\epsilon}{\epsilon}\Big)\cdot\epsilon\cdot\sum_{m=1}^{M}\sum_{\substack{j\neq i:\\(i,j)\in\mathcal{P}_m}} (b_{ij}^{(m)}+b_{ji}^{(m)})\frac{w_{mk}}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mj}} \\
&\quad + \sum_{m=1}^{M}\sum_{\substack{j\neq i:\\(i,j)\in\mathcal{P}_m}} (b_{ij}^{(m)}+b_{ji}^{(m)})\Bigg[\frac{w_{mk}\tilde{h}_{ki}^{(l)}}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mj}}\Bigg] \\
&= \sum_{m=1}^{M}\sum_{\substack{j\neq i:\\(i,j)\in\mathcal{P}_m}} (b_{ij}^{(m)}+b_{ji}^{(m)})\Bigg[\frac{w_{mk}}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mj}}\Bigg(-\epsilon\log(\frac{\tilde{h}_{ki}^{(l)}+\epsilon}{\epsilon})+\tilde{h}_{ki}^{(l)}\Bigg)\Bigg] \geq 0
\end{aligned}
$$

The proof is done and the last inequality is satisfied since $x \geq \log(x+1)$ for all $x \geq 0$ with equality at $x = 0$.

S-2 Convergence Analysis: Proof of Theorem 1

S-2.1 Conditions for Convergence

The paper [1] shows that given a function $f(x)$ to be minimized on domain $\mathcal{X} = \prod_{i=1}^{n}\mathcal{X}_i$, if the surrogate function (auxiliary function/majorizer) $F_i(\cdot|\cdot)$ satisfies the following five properties:

(P1) $F_i(\tilde{x}_i|\tilde{x}) = f(\tilde{x})$, for any $\tilde{x} \in \mathcal{X}$

(P2) $F_i(x_i|\tilde{x}) \leq f(\tilde{x}_1, ..., x_i, ..., \tilde{x}_n)$, for any $(x_i, \tilde{x}) \in \mathcal{X}_i \times \mathcal{X}$

(P3) $F_i(\cdot|\cdot)$ is differentiable on $\textbf{int}\,\mathcal{X}_i\times$ $\textbf{int}\,\mathcal{X}$, there exists a function $g(\cdot|\tilde{x}) : \nabla F_i(\cdot|\tilde{x}) = g(\cdot/\tilde{x}_i|\tilde{x})$

(P4) Define $f_i(\cdot|\tilde{x}) : x_i \mapsto f(\tilde{x}_1, ..., x_i, ..., \tilde{x}_n)$, for any $x_i \in \mathcal{X}_i$, and $\tilde{x} \in \mathcal{X}$. Then for any $\hat{x}_i \in \mathcal{X}_i$, $F_i'(x_i; \hat{x}_i - x_i|\tilde{x})|_{x_i=\tilde{x}_i} = f_i'(x_i; \hat{x}_i - x_i|\tilde{x})|_{x_i=\tilde{x}_i}$ where the directional derivative $f'(x; d)$ is defined as $f'(x; d) := \lim_{\delta\to 0}[f(x + \delta d) - f(x)]/\delta$

(P5) $F_i(\cdot|\tilde{x})$ is strictly convex on $\mathcal{X}_i$, for any $\tilde{x} \in \mathcal{X}$

where,

$$
\begin{aligned}
\mathcal{X} &= \mathbb{R}_{++}^{M\times K} \times \mathbb{R}_{++}^{K\times N} \\
f_\epsilon(\mathbf{W}, \mathbf{H}) &= \sum_{m=1}^{M}\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\Bigg[-\log\left([\mathbf{W}(\mathbf{H}+\epsilon\mathbb{1})]_{mi}\right) + \log\left([\mathbf{W}(\mathbf{H}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\mathbf{H}+\epsilon\mathbb{1})]_{mj}\right)\Bigg] \\
F_1(x_1|\tilde{x}) &= u_1(\mathbf{W}, \mathbf{W}^{(l)}|\mathbf{H}^{(l)}) \\
F_2(x_2|\tilde{x}) &= u_2(\mathbf{H}, \mathbf{H}^{(l)}|\mathbf{W}^{(l+1)})
\end{aligned}
$$

the sequence of iterates $\{(\mathbf{W}^{(l)}, \mathbf{H}^{(l)})\}_{l=0}^{\infty}$ generated by Algorithm 1 converges to the set of stationary points that minimizes the negative log-likelihood $f_\epsilon(\mathbf{W}, \mathbf{H})$.

Although there are truncations to zero during the update of $\mathbf{H}$, the update:

$$
\mathbf{H}^{(l+1)} = \underset{\mathbf{H}\in\mathbb{R}_{++}^{K\times N}}{\text{argmin}}\; u_2(\mathbf{H}, \mathbf{H}^{(l)}|\mathbf{W}^{(l+1)}) \tag{S-3}
$$

is still satisfied. However, those h_{ki} that do not involve truncations are updated such that $\left[u_2(\mathbf{H}, \mathbf{H}^{(l)}|\mathbf{W}^{(l+1)})\right]_{ki}$ is minimized. For those h_{ki} that involve truncations,

$$\left[u_2(\mathbf{H}, \mathbf{H}^{(l)}|\mathbf{W}^{(l+1)})\right]_{ki}\Big|_{h_{ki}=0} \leq \left[u_2(\mathbf{H}, \mathbf{H}^{(l)}|\mathbf{W}^{(l+1)})\right]_{ki}\Big|_{h_{ki}}, \qquad \forall h_{ki} \geq 0$$

is always satisfied by the fact that $x \geq \log(x+1), \forall x \geq 0$. Hence, the update rule in equation (S-3) is maintained. We only need to check the five properties.

S-2.2 Surrogate Functions

We have

$$\begin{aligned}
u_1(\mathbf{W}, \mathbf{W}^{(l)}|\mathbf{H}^{(l)}) = &\sum_m \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \Bigg[-\sum_k \frac{w_{mk}^{(l)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}} \log\left(\frac{w_{mk}}{w_{mk}^{(l)}}[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1}]_{mi}\right) \\
&+ \log\left([\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\right) \\
&+ \frac{[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}} - 1\Bigg]
\end{aligned}$$

and

$$\begin{aligned}
u_2(\mathbf{H}, \mathbf{H}^{(l)}|\mathbf{W}^{(l+1)}) = &\sum_m \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \Bigg[-\sum_k \frac{w_{mk}^{(l+1)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}} \log\left(\frac{h_{ki}+\epsilon}{h_{ki}^{(l)}+\epsilon}[\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}\right) \\
&+ \log\left([\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\right) \\
&+ \frac{[\mathbf{W}^{(l+1)}(\mathbf{H}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}^{(l+1)}(\mathbf{H}+\epsilon\mathbb{1})]_{mj}}{[\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}} - 1\Bigg]
\end{aligned}$$

P1 and P2 are satisfied by the fact that u_1 and u_2 are surrogate functions for $\mathbf{W}$ and $\mathbf{H}$ respectively.

S-2.3 Checking P3 & P4

Firstly, we check slope with respect to $\mathbf{W}$

$$\begin{aligned}
&\left[\nabla_{\mathbf{W}} f_\epsilon(\mathbf{W}, \mathbf{H}^{(l)})\right]_{mk} \\
&= \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \frac{(h_{ki}^{(l)} + h_{kj}^{(l)} + 2\epsilon)[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} - (h_{ki}^{(l)}+\epsilon)\left[[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\right]}{([\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj})[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}} \\
&= \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \frac{h_{kj}^{(l)}[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} - h_{ki}^{(l)}[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj} + \epsilon\left[[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1}]_{mi} - [\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\right]}{\left[[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\right][\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}
\end{aligned}$$

as well as the slope with respect to u_1

$$\begin{aligned}
&\left[\nabla_{\mathbf{W}} u_1(\mathbf{W}, \mathbf{W}^{(l)}|\mathbf{H}^{(l)})\right]_{mk} \\
&= \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \left[-\frac{w_{mk}^{(l)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}} \frac{1}{w_{mk}} + \frac{h_{ki}^{(l)} + h_{kj}^{(l)} + 2\epsilon}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}\right]
\end{aligned}$$

It is easy to check that:

$$\begin{aligned}
&\left[\nabla_{\mathbf{W}} u_1(\mathbf{W}, \mathbf{W}^{(l)}|\mathbf{H}^{(l)})\Big|_{\mathbf{W}=\mathbf{W}^{(l)}}\right]_{mk} \\
=& \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \left[-\frac{h_{ki}^{(l)}+\epsilon}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}} + \frac{h_{ki}^{(l)}+h_{kj}^{(l)}+2\epsilon}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}} \right] \\
=& \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \frac{h_{kj}^{(l)}[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} - h_{ki}^{(l)}[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj} + \epsilon\Big[[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1}]_{mi} - [\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\Big]}{\Big[[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\Big][\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}} \\
=& \left[\nabla_{\mathbf{W}} f(\mathbf{W}, \mathbf{H}^{(l)})\Big|_{\mathbf{W}=\mathbf{W}^{(l)}}\right]_{mk}
\end{aligned}$$

This implies that

$$\nabla_{\mathbf{W}} u_1(\mathbf{W}, \mathbf{W}^{(l)}|\mathbf{H}^{(l)})\Big|_{\mathbf{W}=\mathbf{W}^{(l)}} = \nabla_{\mathbf{W}} f(\mathbf{W}, \mathbf{H}^{(l)})\Big|_{\mathbf{W}=\mathbf{W}^{(l)}} \tag{S-4}$$

Next define $f_{1,\epsilon}(\cdot|\mathbf{H}^{(l)}) : \mathbf{W} \mapsto f(\mathbf{W}, \mathbf{H}^{(l)})$. This is evaluated as follows

$$\begin{aligned}
&f'_{1,\epsilon}(\mathbf{W}; \widehat{\mathbf{W}} - \mathbf{W}|\mathbf{H}^{(l)}) \\
&= \lim_{\delta\to 0} \frac{f(\mathbf{W}+\delta(\widehat{\mathbf{W}}-\mathbf{W})|\mathbf{H}^{(l)}) - f(\mathbf{W}|\mathbf{H}^{(l)})}{\delta} \\
&= \lim_{\delta\to 0} \frac{1}{\delta}\left[\sum_m \sum_{(i,j)} b_{ij}^{(m)} \log\left(\frac{[(\mathbf{W}+\delta(\widehat{\mathbf{W}}-\mathbf{W}))(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi} + [(\mathbf{W}+\delta(\widehat{\mathbf{W}}-\mathbf{W}))(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}{[(\mathbf{W}+\delta(\widehat{\mathbf{W}}-\mathbf{W}))(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}\right.\right. \\
&\qquad \left.\left. \times \frac{[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}{[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}\right)\right]
\end{aligned}$$

For simplification, we denote $a = [\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}$, $b = [\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}$, $c = [\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}$, $d = [\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}$. Then the above simplifies to

$$\begin{aligned}
f'_{1,\epsilon}(\mathbf{W}; \widehat{\mathbf{W}} - \mathbf{W}|\mathbf{H}^{(l)}) &= \lim_{\delta\to 0} \sum_m \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \frac{(1-\delta)a+\delta c+(1-\delta)b+\delta d}{(1-\delta)a+\delta c} \frac{a}{a+b} \\
&= \lim_{\delta\to 0} \sum_m \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \frac{1}{\delta} \log \frac{(1-\delta)a^2+(1-\delta)ab+\delta ac+\delta ad}{(1-\delta)a^2+(1-\delta)ab+\delta ac+\delta bc} \\
&= \lim_{\delta\to 0} \sum_m \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \frac{1}{\delta} \log\left(1+\frac{\delta ad-\delta bc}{(1-\delta)a^2+(1-\delta)ab+\delta ac+\delta bc}\right)
\end{aligned}$$

For further simplification, we denote $s = ad - bc$, $r = a(a+b)$, $t = c(a+b)$. The above simplifies to

$$\begin{aligned}
f'_{1,\epsilon}(\mathbf{W}; \widehat{\mathbf{W}} - \mathbf{W}|\mathbf{H}^{(l)}) &= \lim_{\delta\to 0} \sum_m \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \frac{1}{\delta} \log\left(1+\frac{\delta s}{(1-\delta)r+\delta t}\right) \\
&= \lim_{\delta\to 0} \sum_m \sum_{(i,j)\in\mathcal{P}_m} \log\left(1+\frac{\frac{s}{r}}{\frac{1}{\delta}+\frac{t-r}{r}}\right)^{\frac{1}{\delta}} \\
&= \lim_{\delta\to 0} \sum_m \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \log\left(1+\frac{\frac{s}{r}}{\frac{1}{\delta}+\frac{t-r}{r}}\right)^{\frac{1}{\delta}+\frac{t-r}{r}} - \log\left(1+\frac{\frac{s}{r}}{\frac{1}{\delta}+\frac{t-r}{r}}\right)^{\frac{t-r}{r}}
\end{aligned}$$

since $\left(1+\frac{t}{f(x)}\right)^{f(x)} \to e^t$ as $f(x)\to\infty$ and $\frac{1}{\delta}+\frac{t-r}{r}\to\infty$ as $\delta\to 0$. Thus,

$$
\begin{aligned}
&f'_{1,\epsilon}(\mathbf{W};\widehat{\mathbf{W}}-\mathbf{W}|\mathbf{H}^{(l)})\\
&=\sum_m\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\frac{s}{r}\\
&=\sum_m\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\frac{ad-bc}{a(a+b)}\\
&=\sum_m\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\frac{[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}[\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}-[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}[\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}{[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}\Big[[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\Big]}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&u'_1(\mathbf{W};\widehat{\mathbf{W}}-\mathbf{W},\mathbf{W}^{(l)}|\mathbf{H}^{(l)})\\
&=\lim_{\delta\to0}\frac{\tilde{u}_1(\mathbf{W}+\delta(\widehat{\mathbf{W}}-\mathbf{W})|\mathbf{W}^{(l)},\mathbf{H}^{(l)})-\tilde{u}_1(\mathbf{W}|\mathbf{W}^{(l)},\mathbf{H}^{(l)})}{\delta}\\
&=\lim_{\delta\to0}\sum_m\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\frac{1}{\delta}\Bigg[-\sum_k\frac{w_{mk}^{(l)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}\log\left(\frac{(1-\delta)w_{mk}+\delta\hat{w}_{mk}}{w_{mk}}\right)\\
&\quad+\frac{\delta\Big([\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})-\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})-\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\Big)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}\Bigg]\\
&=\lim_{\delta\to0}\sum_m\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\Bigg[-\sum_k\frac{w_{mk}^{(l)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}\log\left(1+\frac{\frac{\hat{w}_{mk}-w_{mk}}{w_{mk}}}{\frac{1}{\delta}}\right)^{\frac{1}{\delta}}\\
&\quad+\frac{\Big([\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})-\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})-\mathbf{W}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\Big)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}\Bigg]
\end{aligned}
$$

where the last equality again follows from the fact that $\left(1+\frac{t}{f(x)}\right)^{f(x)}\to e^t$ as $f(x)\to\infty$. Thus,

$$
\begin{aligned}
u'_1(\mathbf{W};\widehat{\mathbf{W}}-\mathbf{W},\mathbf{W}^{(l)}|\mathbf{H}^{(l)})=\sum_m\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\Bigg[&-\sum_k\frac{w_{mk}^{(l)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}\frac{\hat{w}_{mk}-w_{mk}}{w_{mk}}\\
&+\frac{[(\widehat{\mathbf{W}}-\mathbf{W})(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[(\widehat{\mathbf{W}}-\mathbf{W})(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}\Bigg]
\end{aligned}
$$

Now we evaluate the expressions above at the point $\mathbf{W}=\mathbf{W}^{(l)}$ as follows

$$
\begin{aligned}
&u'_1(\mathbf{W};\widehat{\mathbf{W}}-\mathbf{W},\mathbf{W}^{(l)}|\mathbf{H}^{(l)})\Big|_{\mathbf{W}=\mathbf{W}^{(l)}}\\
&=\sum_m\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\Bigg[-\sum_k\frac{w_{mk}^{(l)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}\frac{\hat{w}_{mk}-w_{mk}^{(l)}}{w_{mk}^{(l)}}+\frac{[(\widehat{\mathbf{W}}-\mathbf{W})(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[(\widehat{\mathbf{W}}-\mathbf{W})(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}\Bigg]\\
&=\sum_m\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\Bigg[-\frac{[(\widehat{\mathbf{W}}-\mathbf{W})(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}+\frac{[(\widehat{\mathbf{W}}-\mathbf{W})(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[(\widehat{\mathbf{W}}-\mathbf{W})(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}\Bigg]\\
&=\sum_m\sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\frac{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}[\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}-[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}[\widehat{\mathbf{W}}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}\Big[[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}\Big]}\\
&=f'_{1,\epsilon}(\mathbf{W};\widehat{\mathbf{W}}-\mathbf{W}|\mathbf{H}^{(l)})\Big|_{\mathbf{W}=\mathbf{W}^{(l)}}
\end{aligned}
$$

Together with equation (S-4), we have proved P4.

The same idea can be applied to the surrogate function with respect to $\mathbf{H}$, hence we conclude that:

$$
\begin{aligned}
u_1'(\mathbf{W};\widehat{\mathbf{W}}-\mathbf{W},\mathbf{W}^{(l)}|\mathbf{H}^{(l)})\Big|_{\mathbf{W}=\mathbf{W}^{(l)}} &= f_{1,\epsilon}'(\mathbf{W};\widehat{\mathbf{W}}-\mathbf{W}|\mathbf{H}^{(l)})\Big|_{\mathbf{W}=\mathbf{W}^{(l)}} \\
u_2'(\mathbf{H};\widehat{\mathbf{H}}-\mathbf{H},\mathbf{H}^{(l)}|\mathbf{W}^{(l+1)})\Big|_{\mathbf{H}=\mathbf{H}^{(l)}} &= f_{2,\epsilon}'(\mathbf{H};\widehat{\mathbf{H}}-\mathbf{H}|\mathbf{W}^{(l+1)})\Big|_{\mathbf{H}=\mathbf{H}^{(l)}}
\end{aligned}
$$

S-2.4 Checking P5

We have,

$$
\begin{aligned}
\frac{\partial^2}{\partial w_{mk}^2}u_1(\mathbf{W},\mathbf{W}^{(l)}|\mathbf{H}^{(l)}) &= \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\left(\frac{w_{mk}^{(l)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}\frac{1}{w_{mk}^2}\right) \\
\frac{\partial^2}{\partial h_{ki}^2}u_2(\mathbf{H},\mathbf{W}^{(l+1)}|\mathbf{H}^{(l)}) &= \sum_m \sum_{j\neq i:(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\left(\frac{w_{mk}^{(l+1)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}\frac{1}{(h_{ki}+\epsilon)^2}\right)
\end{aligned}
$$

Hence, u_1, u_2 are strictly convex if $w_{mk} > 0$ and $h_{ki} \geq 0$. This is satisfied since $(\mathbf{W},\mathbf{H})$ are both initialized to be positive, i.e., $\mathbf{W}^{(0)} \in \mathbb{R}_{++}^{M\times K}$, $\mathbf{H}^{(0)} \in \mathbb{R}_{++}^{K\times N}$, and during the update of $\mathbf{W}$ and $\mathbf{H}$, entries are kept positive for $\mathbf{W}$ and nonnegative for $\mathbf{H}$.

The five properties are satisfied and hence by [1] the sequence converges to the set of stationary points.

Remark 1: While we have verified that all 5 properties hold, the proof of Theorem 1 differs slightly from the original proof in [1]. Firstly, in the orginal proof of [1], the regularization terms in the objective function guarantees that the set $S_0 \triangleq \left\{(\mathbf{W},\mathbf{H}) \in \mathbb{R}_{++}^{M\times K} \times \mathbb{R}_{++}^{K\times N} \mid \ell(\mathbf{W},\mathbf{H}) \leq \ell(\mathbf{W}^0,\mathbf{H}^0)\right\}$ is bounded. However, in our theorem, boundedness of the corresponding set is ensured by normalization where $\sum_k w_{mk} = 1$ or $\sum_k w_{mk} = 1$, and $\sum_{k,i} h_{ki} = 1$. Secondly, in the last step of Algorithm 1, we perform normalization for both $\mathbf{W}$ and $\mathbf{H}$. Denote the result after the **Update** step in Algorithm 1 as $(\tilde{\mathbf{W}},\tilde{\mathbf{H}})$ and the result after the **Normalization** step as $(\mathbf{W},\mathbf{H})$. Although $f_\epsilon(\tilde{\mathbf{W}},\tilde{\mathbf{H}}) = f_\epsilon(\mathbf{W},\mathbf{H})$, the derivatives differ. This does not affect the last part of the proof in [1], because if we denote the normalization operation as a function, where $\mathbf{W} = N(\tilde{\mathbf{W}})$, application of the chain rule suggests that $f_{1,\epsilon}'(\mathbf{W};\hat{\mathbf{W}}-\mathbf{W}|H) = f_{1,\epsilon}'(\tilde{\mathbf{W}};\hat{\mathbf{W}}-\tilde{\mathbf{W}}|H)\cdot N'(\tilde{\mathbf{W}})$, where $N'(\tilde{\mathbf{W}})$ are positive as one can easily observe from the **Normalization** step in Algorithm 1. Hence, the directional derivatives of $f_{1,\epsilon}'$ and $f_{2,\epsilon}'$ are still non-negative at the limit points.

S-3 Non-decreasing likelihood

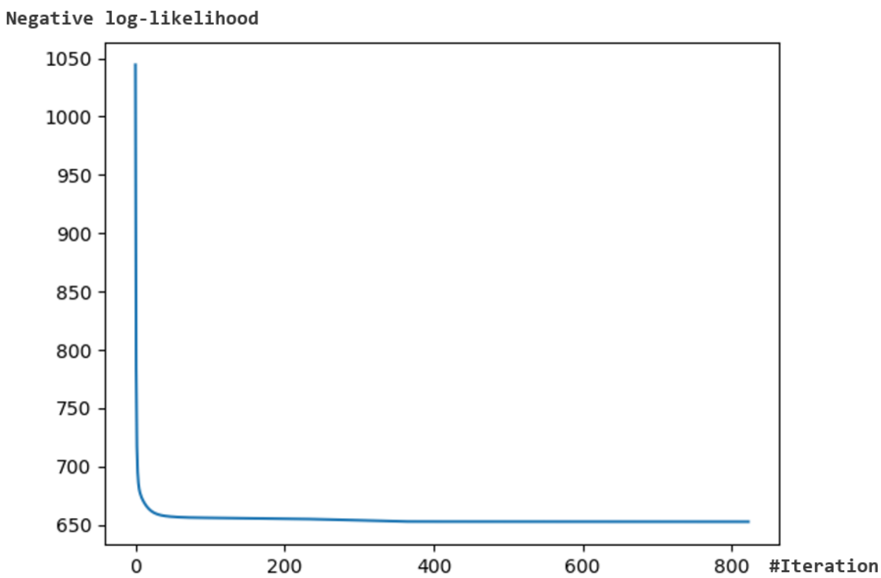

Figure S-1: Plot of the evolution of negative log-likelihoods

Fig. S-1 shows that the likelihood is non-decreasing (negative log-likelihood is non-increasing).

S-4 Numerical Results for $K = 3$ for men

Tables S-1 and S-2 show the $\mathbf{W}$ and $\mathbf{H}$ matrices for the men when $K = 3$. The matrix $\mathbf{\Lambda} = \mathbf{WH}$ when $K = 3$ is displayed in Tables S-3 and S-4.

Tournaments	Row Normalization			Column Normalization		
Australian Open	2.25E-01	3.61E-01	4.14E-01	1.87E-02	4.00E-02	3.65E-02
Indian Wells Masters	2.74E-01	2.79E-01	4.47E-01	4.21E-02	8.02E-02	5.24E-02
Miami Open	6.28E-02	8.45E-01	9.25E-02	1.56E-02	2.68E-02	2.56E-01
Monte-Carlo Masters	8.32E-01	6.06E-101	1.68E-01	1.96E-01	4.63E-02	2.02E-72
Madrid Open	0.00E+00	1.00E-00	1.89E-33	0.00E+00	8.73E-14	2.03E-01
Italian Open	7.65E-01	2.35E-01	0.00E+00	1.64E-01	0.00E+00	6.14E-02
French Open	6.17E-01	1.02E-01	2.81E-01	9.10E-02	4.83E-02	1.83E-02
Wimbledon	3.94E-01	6.11E-10	6.06E-01	5.51E-02	9.92E-02	1.40E-20
Canadian Open	0.00E+00	0.00E+00	1.00E-00	0.00E+00	2.55E-01	0.00E+00
Cincinnati Masters	4.95E-01	3.31E-16	5.05E-01	1.51E-01	1.80E-01	2.30E-38
US Open	3.36E-01	2.82E-01	3.82E-01	5.60E-02	7.42E-02	5.73E-02
Shanghai Masters	9.85E-167	6.45E-01	3.55E-01	1.04E-141	7.92E-02	1.51E-01
Paris Masters	7.48E-01	1.47E-01	1.06E-01	1.88E-01	3.09E-02	4.49E-02
The ATP Finals	1.46E-01	6.32E-01	2.21E-01	2.26E-02	3.99E-02	1.19E-01

Table S-1: Learned dictionary matrix $\mathbf{W}$ for the men's dataset

Players	matrix $\mathbf{H}^T$		
Novak Djokovic	3.66E-02	3.37E-02	1.81E-01
Rafael Nadal	1.38E-01	1.30E-02	4.03E-02
Roger Federer	6.44E-03	6.25E-02	6.03E-02
Andy Murray	4.40E-32	4.45E-02	3.31E-02
Tomas Berdych	1.62E-02	0.00E+00	1.03E-02
David Ferrer	2.29E-02	0.00E+00	1.04E-02
Stan Wawrinka	2.87E-02	0.00E+00	1.02E-02
Jo-Wilfried Tsonga	1.77E-03	3.05E-02	6.69E-04
Richard Gasquet	1.05E-02	4.02E-03	3.09E-03
Juan Martin del Potro	3.62E-03	1.56E-02	1.84E-02
Marin Cilic	0.00E+00	1.86E-02	7.38E-10
Fernando Verdasco	5.53E-03	1.26E-02	1.81E-03
Kei Nishikori	0.00E+00	1.56E-03	3.09E-02
Gilles Simon	3.16E-03	1.10E-02	2.07E-03
Milos Raonic	7.13E-03	1.10E-02	9.22E-04
Philipp Kohlschreiber	4.58E-03	4.09E-07	1.12E-06
John Isner	1.63E-02	2.13E-03	0.00E+00
Feliciano Lopez	0.00E+00	6.14E-03	1.02E-02
Gael Monfils	9.27E-03	0.00E+00	2.63E-03
Nicolas Almagro	2.25E-06	5.91E-03	0.00E+00

Table S-2: Learned transpose of coefficient matrix $\mathbf{H}^T$ with column normalization of $\mathbf{W}$ for the men's dataset

S-5 Numerical Results for $\mathbf{\Lambda}$ for $K = 2$ for women

The matrix $\mathbf{\Lambda} = \mathbf{WH}$ when $K = 2$ for the women players is displayed in Tables S-5 and S-6.

Tournament	Novak Djokovic	Rafael Nadal	Roger Federer	Andy Murray	Tomas Berdych	David Ferrer	Stan Wawrinka	Jo-Wilfried Tsonga	Richard Gasquet	Juan Martin del Potro
Australian Open	8.64E-03	4.58E-03	4.82E-03	2.99E-03	6.78E-04	8.09E-04	9.11E-04	1.28E-03	4.70E-04	1.36E-03
Indian Wells Masters	1.37E-02	8.97E-03	8.45E-03	5.31E-03	1.22E-03	1.51E-03	1.74E-03	2.56E-03	9.26E-04	2.36E-03
Miami Open	4.78E-02	1.28E-02	1.72E-02	9.68E-03	2.88E-03	3.03E-03	3.07E-03	1.02E-03	1.06E-03	5.18E-03
Monte-Carlo Masters	8.75E-03	2.77E-02	4.16E-03	2.06E-03	3.18E-03	4.49E-03	5.64E-03	1.76E-03	2.24E-03	1.43E-03
Madrid Open	3.68E-02	8.18E-03	1.23E-02	6.73E-03	2.09E-03	2.12E-03	2.08E-03	1.36E-04	6.29E-04	3.73E-03
Italian Open	1.71E-02	2.51E-02	4.76E-03	2.03E-03	3.28E-03	4.39E-03	5.33E-03	3.31E-04	1.91E-03	1.72E-03
French Open	8.28E-03	1.39E-02	4.71E-03	2.76E-03	1.66E-03	2.27E-03	2.80E-03	1.65E-03	1.20E-03	1.42E-03
Wimbledon	5.36E-03	8.90E-03	6.56E-03	4.41E-03	8.94E-04	1.26E-03	1.58E-03	3.13E-03	9.76E-04	1.74E-03
Canadian Open	8.59E-03	3.31E-03	1.60E-02	1.14E-02	0.00E+00	0.00E+00	0.00E+00	7.80E-03	1.03E-03	3.97E-03
Cincinnati Masters	1.16E-02	2.32E-02	1.22E-02	8.00E-03	2.45E-03	3.46E-03	4.34E-03	5.76E-03	2.31E-03	3.34E-03
US Open	1.49E-02	1.10E-02	8.46E-03	5.20E-03	1.50E-03	1.88E-03	2.19E-03	2.40E-03	1.06E-03	2.41E-03
Shanghai Masters	3.00E-02	7.11E-03	1.41E-02	8.53E-03	1.55E-03	1.58E-03	1.54E-03	2.52E-03	7.86E-04	4.01E-03
Paris Masters	1.60E-02	2.81E-02	5.85E-03	2.86E-03	3.50E-03	4.76E-03	5.84E-03	1.31E-03	2.23E-03	1.99E-03
ATP World Tour Finals	2.38E-02	8.44E-03	9.83E-03	5.72E-03	1.59E-03	1.76E-03	1.87E-03	1.34E-03	7.66E-04	2.89E-03

Table S-3: Learned $\mathbf{\Lambda} = \mathbf{WH}$ matrix for first 10 men players

Tournament	Marin Cilic	Fernando Verdasco	Gilles Simon	Milos Raonic	John Isner	Philipp Kohlschreiber	John Isner	Feliciano Lopez	Gael Monfils	Nicolas Almagro
Australian Open	7.45E-04	6.75E-04	1.19E-03	5.73E-04	6.09E-04	8.59E-05	3.90E-04	6.16E-04	2.70E-04	2.37E-04
Indian Wells Masters	1.49E-03	1.34E-03	1.74E-03	1.12E-03	1.23E-03	1.93E-04	8.55E-04	1.03E-03	5.28E-04	4.75E-04
Miami Open	4.99E-04	8.89E-04	7.94E-03	8.73E-04	6.43E-04	7.18E-05	3.11E-04	2.77E-03	8.18E-04	1.59E-04
Monte-Carlo Masters	8.61E-04	1.67E-03	7.20E-05	1.13E-03	1.91E-03	9.00E-04	3.29E-03	2.84E-04	1.82E-03	2.74E-04
Madrid Open	1.50E-10	3.68E-04	6.27E-03	4.20E-04	1.87E-04	2.28E-07	1.86E-16	2.07E-03	5.34E-04	5.16E-16
Italian Open	4.53E-11	1.02E-03	1.90E-03	6.45E-04	1.22E-03	7.50E-04	2.66E-03	6.24E-04	1.68E-03	3.68E-07
French Open	9.00E-04	1.15E-03	6.41E-04	8.55E-04	1.20E-03	4.17E-04	1.58E-03	4.83E-04	8.91E-04	2.86E-04
Wimbledon	1.85E-03	1.56E-03	1.54E-04	1.26E-03	1.49E-03	2.53E-04	1.11E-03	6.09E-04	5.11E-04	5.87E-04
Canadian Open	4.75E-03	3.22E-03	3.97E-04	2.80E-03	2.82E-03	1.04E-07	5.44E-04	1.57E-03	0.00E+00	1.51E-03
Cincinnati Masters	3.35E-03	3.11E-03	2.80E-04	2.45E-03	3.06E-03	6.92E-04	2.84E-03	1.10E-03	1.40E-03	1.06E-03
US Open	1.38E-03	1.35E-03	1.88E-03	1.11E-03	1.27E-03	2.57E-04	1.07E-03	1.04E-03	6.70E-04	4.39E-04
Shanghai Masters	1.47E-03	1.27E-03	4.78E-03	1.18E-03	1.01E-03	2.02E-07	1.69E-04	2.02E-03	3.97E-04	4.69E-04
Paris Masters	5.76E-04	1.51E-03	1.43E-03	1.02E-03	1.72E-03	8.59E-04	3.11E-03	6.46E-04	1.86E-03	1.83E-04
ATP World Tour Finals	7.43E-04	8.45E-04	3.74E-03	7.55E-04	7.12E-04	1.04E-04	4.53E-04	1.46E-03	5.23E-04	2.36E-04

Table S-4: Learned $\mathbf{\Lambda} = \mathbf{WH}$ matrix for last 10 men players

Tournament	Serena Williams	Agnieszka Radwanska	Victoria Azarenka	Caroline Wozniacki	Maria Sharapova	Simona Halep	Petra Kvitova	Angelique Kerber	Samantha Stosur	Ana Ivanovic
Australia Open	6.13E-03	2.48E-03	7.27E-03	3.13E-03	8.67E-04	1.55E-03	2.47E-03	7.05E-04	4.29E-05	9.88E-04
Qatar Open	1.30E-02	3.23E-03	6.80E-03	3.92E-03	5.18E-03	3.01E-03	3.94E-03	2.26E-03	2.13E-03	2.26E-03
Dubai Tennis Championships	4.53E-03	1.83E-03	5.38E-03	2.31E-03	6.41E-04	1.15E-03	1.83E-03	5.21E-04	3.17E-05	7.30E-04
Indian Wells Open	1.12E-02	2.66E-03	5.36E-03	3.22E-03	4.63E-03	2.57E-03	3.31E-03	1.98E-03	1.94E-03	1.95E-03
Miami Open	9.69E-03	2.36E-03	4.87E-03	2.86E-03	3.93E-03	2.24E-03	2.90E-03	1.70E-03	1.63E-03	1.69E-03
Madrid Open	9.80E-03	2.20E-03	4.12E-03	2.64E-03	4.29E-03	2.23E-03	2.81E-03	1.79E-03	1.83E-03	1.72E-03
Italian Open	1.32E-02	2.61E-03	4.06E-03	3.08E-03	6.35E-03	2.96E-03	3.56E-03	2.54E-03	2.82E-03	2.34E-03
French Open	1.59E-02	2.74E-03	3.11E-03	3.16E-03	8.37E-03	3.52E-03	4.03E-03	3.23E-03	3.83E-03	2.87E-03
Wimbledon	2.18E-02	3.76E-03	4.28E-03	4.34E-03	1.15E-02	4.82E-03	5.52E-03	4.42E-03	5.25E-03	3.93E-03
Canadian Open	1.07E-02	2.31E-03	4.11E-03	2.76E-03	4.84E-03	2.43E-03	3.02E-03	1.99E-03	2.10E-03	1.89E-03
Cincinnati Open	2.20E-02	3.30E-03	2.25E-03	3.71E-03	1.23E-02	4.77E-03	5.24E-03	4.62E-03	5.75E-03	3.99E-03
US Open	9.75E-03	2.44E-03	5.17E-03	2.96E-03	3.85E-03	2.26E-03	2.96E-03	1.68E-03	1.57E-03	1.69E-03
Pan Pacific Open	2.54E-02	3.81E-03	2.60E-03	4.29E-03	1.42E-02	5.52E-03	6.06E-03	5.35E-03	6.66E-03	4.61E-03
Wuhan Open	6.77E-03	2.73E-03	8.03E-03	3.45E-03	9.57E-04	1.71E-03	2.73E-03	7.77E-04	4.73E-05	1.09E-03
China Open	2.08E-02	3.68E-03	4.55E-03	4.27E-03	1.07E-02	4.60E-03	5.33E-03	4.16E-03	4.87E-03	3.72E-03
WTA Finals	8.56E-03	1.40E-03	1.36E-03	1.60E-03	4.61E-03	1.88E-03	2.12E-03	1.76E-03	2.13E-03	1.54E-03

Table S-5: Learned $\mathbf{\Lambda} = \mathbf{WH}$ matrix for first 10 women players

Tournament	Jelena Jankovic	Anastasia Pavlyuchenkova	Carla Suarez Navarro	Dominika Cibulkova	Lucie Safarova	Elina Svitolina	Sara Errani	Karolina Pliskova	Roberta Vinci	Marion Bartoli
Australia Open	1.21E-04	7.15E-04	3.63E-03	3.07E-03	1.27E-24	5.20E-04	8.27E-05	1.03E-03	4.28E-03	1.50E-03
Qatar Open	1.29E-03	1.33E-03	2.99E-03	3.10E-03	1.75E-03	1.53E-03	1.56E-03	2.15E-03	3.52E-03	2.16E-03
Dubai Tennis Championships	8.92E-05	5.28E-04	2.68E-03	2.27E-03	1.40E-40	3.85E-04	6.11E-05	7.58E-04	3.17E-03	1.11E-03
Indian Wells Open	1.16E-03	1.13E-03	2.30E-03	2.48E-03	1.60E-03	1.34E-03	1.42E-03	1.85E-03	2.72E-03	1.80E-03
Miami Open	9.84E-04	9.83E-04	2.12E-03	2.23E-03	1.34E-03	1.15E-03	1.19E-03	1.60E-03	2.50E-03	1.59E-03
Madrid Open	1.09E-03	9.77E-04	1.71E-03	1.94E-03	1.52E-03	1.20E-03	1.34E-03	1.62E-03	2.01E-03	1.51E-03
Italian Open	1.65E-03	1.28E-03	1.48E-03	2.03E-03	2.35E-03	1.69E-03	2.04E-03	2.17E-03	1.75E-03	1.86E-03
French Open	2.21E-03	1.51E-03	8.07E-04	1.74E-03	3.21E-03	2.14E-03	2.75E-03	2.62E-03	9.52E-04	2.04E-03
Wimbledon	3.02E-03	2.07E-03	1.12E-03	2.39E-03	4.39E-03	2.92E-03	3.77E-03	3.59E-03	1.32E-03	2.79E-03
Canadian Open	1.24E-03	1.06E-03	1.65E-03	1.97E-03	1.75E-03	1.34E-03	1.53E-03	1.77E-03	1.94E-03	1.61E-03
Cincinnati Open	3.28E-03	2.03E-03	8.71E-07	1.59E-03	4.83E-03	3.04E-03	4.11E-03	3.60E-03	1.74E-124	2.56E-03
US Open	9.57E-04	9.95E-04	2.28E-03	2.36E-03	1.30E-03	1.14E-03	1.16E-03	1.61E-03	2.69E-03	1.63E-03
Pan Pacific Open	3.79E-03	2.35E-03	1.01E-06	1.83E-03	5.58E-03	3.52E-03	4.76E-03	4.17E-03	2.62E-34	2.97E-03
Wuhan Open	1.33E-04	7.89E-04	4.00E-03	3.39E-03	5.67E-63	5.74E-04	9.12E-05	1.13E-03	4.72E-03	1.65E-03
China Open	2.81E-03	1.98E-03	1.33E-03	2.46E-03	4.07E-03	2.76E-03	3.50E-03	3.42E-03	1.56E-03	2.71E-03
WTA Finals	1.22E-03	8.03E-04	2.64E-04	8.09E-04	1.78E-03	1.16E-03	1.53E-03	1.41E-03	3.11E-04	1.06E-03

Table S-6: Learned $\mathbf{\Lambda} = \mathbf{WH}$ matrix for last 10 women players

Players	$K = 2$ Trial 1		$K = 2$ Trial 2	
Novak Djokovic	1.20E-01	2.91E-02	3.40E-05	9.42E-05
Rafael Nadal	1.07E-01	2.25E-02	1.47E-05	1.15E-04
Roger Federer	1.53E-01	1.11E-02	9.29E-03	1.83E-05
Andy Murray	1.43E-01	4.39E-03	2.46E-05	1.52E-05
Tomas Berdych	2.37E-12	6.59E-03	6.51E-19	1.60E-05
David Ferrer	4.74E-02	2.19E-03	1.56E-05	5.89E-06
Stan Wawrinka	6.26E-07	7.21E-03	2.11E-05	6.29E-06
Jo-Wilfried Tsonga	2.03E-01	5.88E-04	9.90E-01	1.04E-06
Richard Gasquet	4.98E-04	1.62E-03	5.30E-08	4.81E-06
Juan Martin del Potro	4.26E-06	8.01E-03	1.90E-05	7.19E-06
Marin Cilic	1.56E-09	2.12E-03	3.49E-16	4.11E-06
Fernando Verdasco	2.75E-17	7.12E-03	6.54E-05	9.72E-07
Kei Nishikori	1.83E-12	8.58E-03	4.18E-23	1.77E-05
Gilles Simon	5.14E-06	1.31E-03	2.47E-10	4.13E-06
Milos Raonic	2.07E-07	2.84E-03	3.99E-08	6.00E-06
Philipp Kohlschreiber	0.00E+00	1.13E-03	7.99E-06	5.0E-324
John Isner	6.93E-02	3.21E-04	1.73E-22	9.47E-06
Feliciano Lopez	3.67E-02	4.93E-04	8.57E-06	1.38E-06
Gael Monfils	6.05E-14	2.85E-03	1.06E-12	4.00E-06
Nicolas Almagro	4.18E-07	2.14E-04	1.04E-14	8.10E-07
Assignments	4.72E-01	5.28E-01	3.32E-01	6.68E-01
Log-likelihoods	-657.56		-656.47	

Table S-7: Learned $\boldsymbol{\lambda}$ skill vectors for the mixture-BTL model with $K = 2$

S-6 Numerical Results for BTL and Mixture-BTL models

The learned $\boldsymbol{\lambda}$ vectors for the mixture-BTL models with $K = 2$ across two almost-optimal trials are displayed in Table S-7.

References

[1] R. Zhao and V. Y. F. Tan. A unified convergence analysis of the multiplicative update algorithm for nonnegative matrix factorization. *IEEE Transactions on Signal Processing*, 66(1):129–138, 2018.