A Ranking Model Motivated by NMF with Applications to Tennis Tournaments

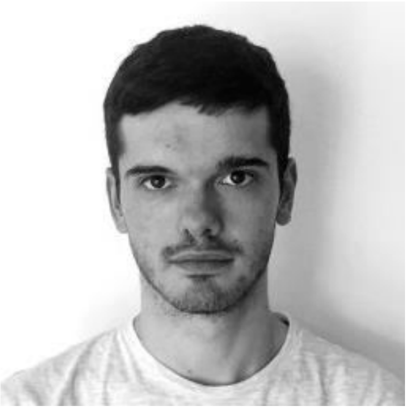
Vincent Tan

National University of Singapore

LRMA 2019

Mons, Belgium

13 Sept 2019



Rui Xia
NUS, Maths

Louis Filstroff
IRIT, Toulouse

Cédric Févotte
IRIT, Toulouse

Motivation

Rod Laver
3 Grand Slam

Roger Federer
3 Grand Slam

Novak Djokovic
5 Grand Slam

1960s

2005 2006

2010 2011

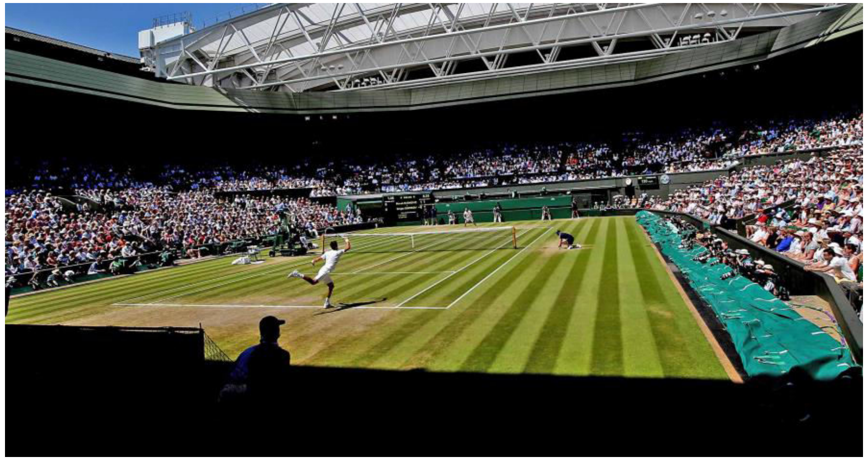
2015 2016

Rafael Nadal
3 Grand Slam

Who is the greatest players
of all time?

52 weeks
10 yeas

Latent Variables: Surface Type?



Wimbledon
Grass Outdoors

Australian Open
Hard Outdoors

French Open
Clay Outdoors

US Open
Hard Outdoors

1. **BTL-NMF Algorithm**
    - Majorization-Minimization Algorithm
    - Resolution of Numerical Problems
    - Normalization
    - Convergence Analysis

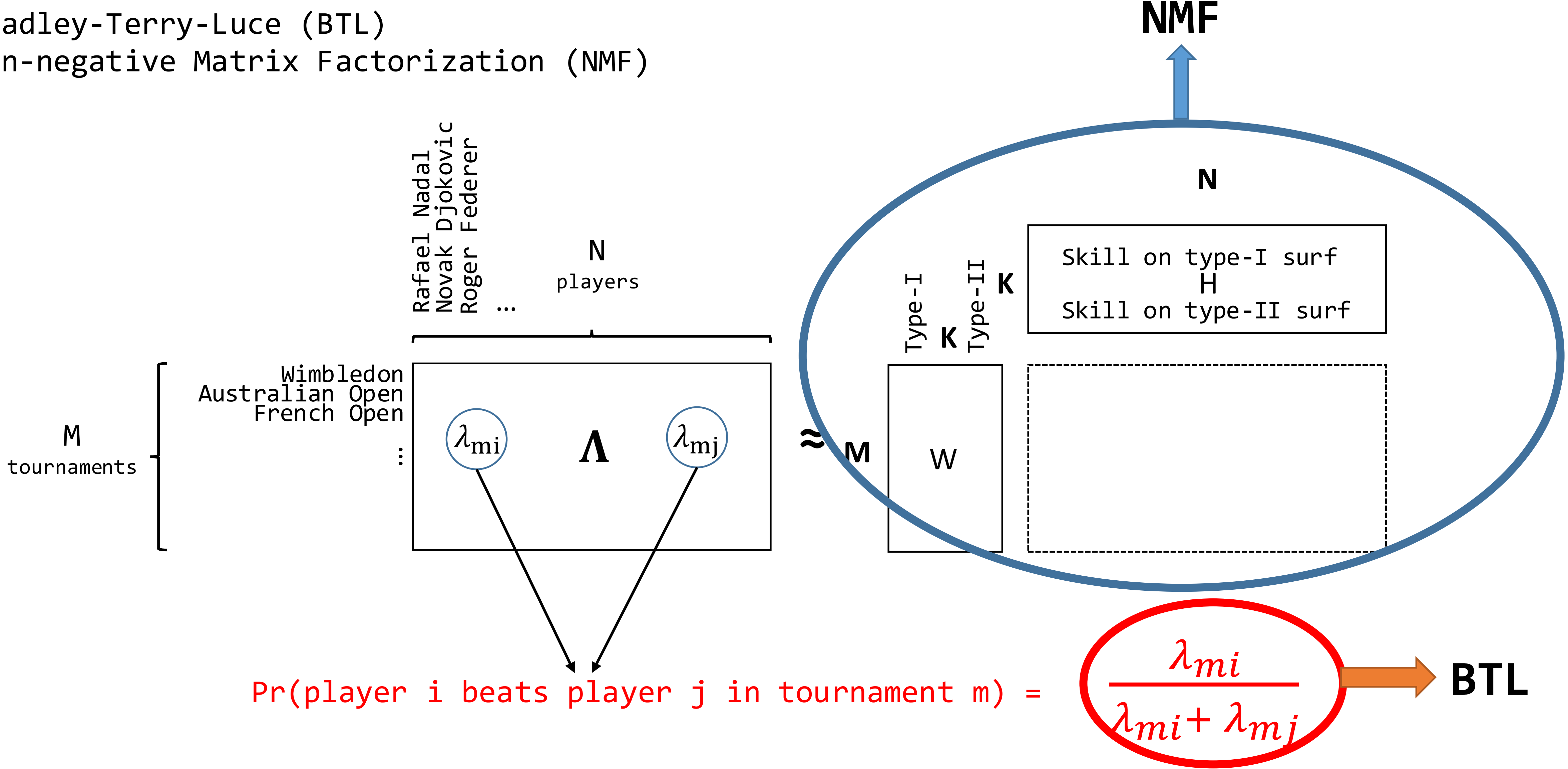
2. Experiments Using Real Life Dataset
    - Dataset Information
    - Running of Algorithm
    - Men Players
    - Women Players

3. Comparison with Mixture BTL
    - Qualitative Comparison
    - Prediction Task & Result Comparison

Hybrid of BTL and NMF

Bradley-Terry-Luce (BTL)
Non-negative Matrix Factorization (NMF)

λ_{mi}: skill of player i in tournament m

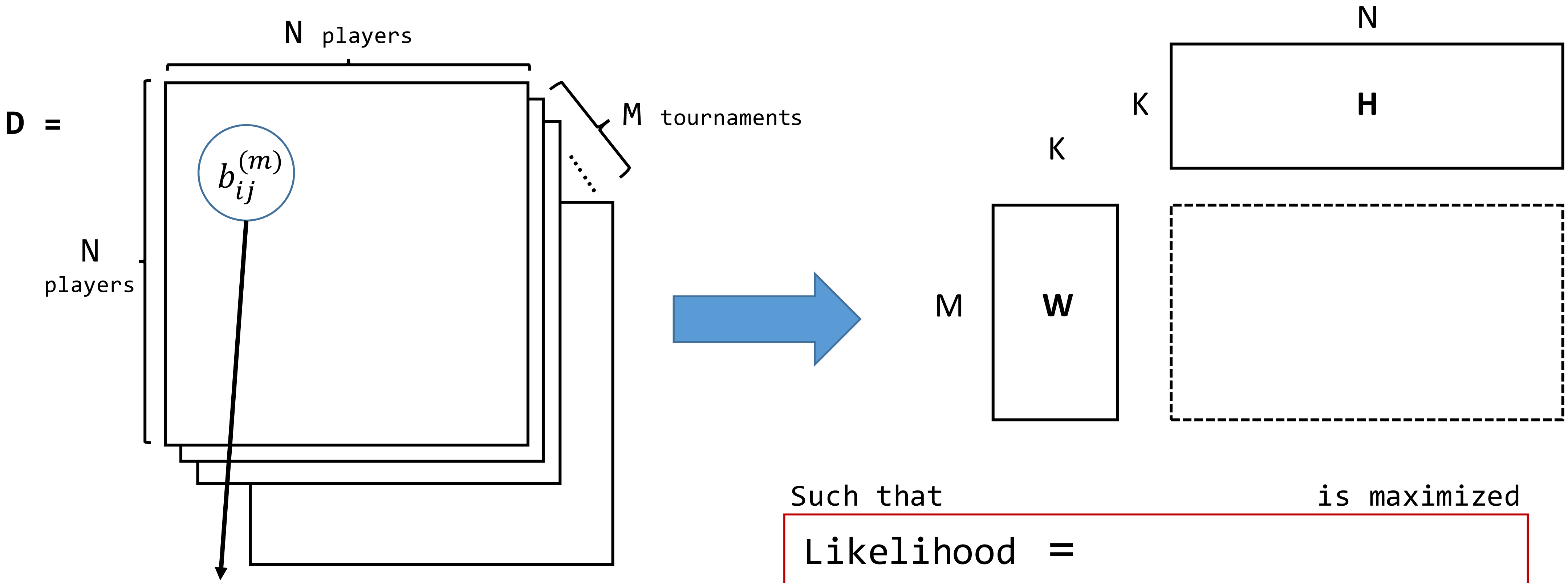


Pr(player i beats player j in tournament m) = $\frac{\lambda_{mi}}{\lambda_{mi}+\lambda_{mj}}$

R. Bradley and M. Terry. Rank analysis of incomplete block designs I: The method of paired comparisons. Biometrika, 35:324--345, 1952.
R. Luce. Individual choice behavior: A theoretical analysis. Wiley, 1959.
D. D. Lee and H. S. Seung. Learning the parts of objects with nonnegative matrix factorization. Nature, 401:788—791, 1999.

GIVEN: D ~ M (N x N) matrices

Output: W ~ (M x K) matrix
H ~ (K x N) matrix



Number of times
player i beat player j
in tournament m

Such that is maximized

Likelihood =

$$\prod_{m=1}^{M}\prod_{(i,j)}\left(\frac{[\mathbf{WH}]_{mi}}{[\mathbf{WH}]_{mi}+[\mathbf{WH}]_{mj}}\right)^{b_{ij}^{(m)}}$$

Task: Minimize the negative log-likelihood

Find $$\operatorname*{argmin}_{\mathbf{W},\mathbf{H} \geq \mathbf{0}} \left(-\log P(\mathbf{W}, \mathbf{H} | \mathcal{D}) \right)$$

$$= \operatorname*{argmin}_{\mathbf{W},\mathbf{H} \geq \mathbf{0}} \left(\sum_{m=1}^{M} \sum_{(i,j) \in \mathcal{P}_m} b_{ij}^{(m)} \left[-\log \left([\mathbf{WH}]_{mi} \right) + \log \left([\mathbf{WH}]_{mi} + [\mathbf{WH}]_{mj} \right) \right] \right)$$

Convex? No!

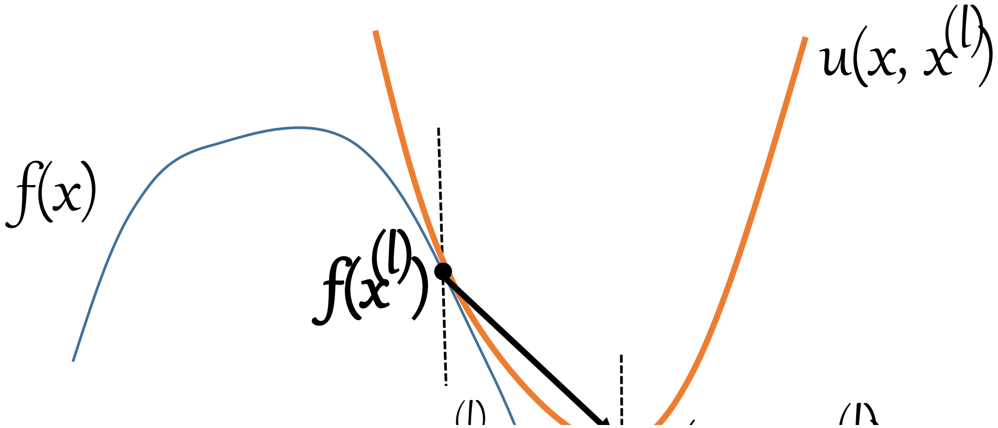
Objective function not guaranteed to decrease using standard gradient-based algorithms

1. BTL-NMF Algorithm
    - **Majorization-Minimization Algorithm**
    - Resolution of Numerical Problems
    - Normalization
    - Convergence Analysis

2. Experiments Using Real Life Dataset
    - Dataset Information
    - Running of Algorithm
    - Men Players
    - Women Players

3. Comparison with Mixture BTL
    - Qualitative Comparison
    - Prediction Task & Result Comparison

Majorization-Minimization algorithm



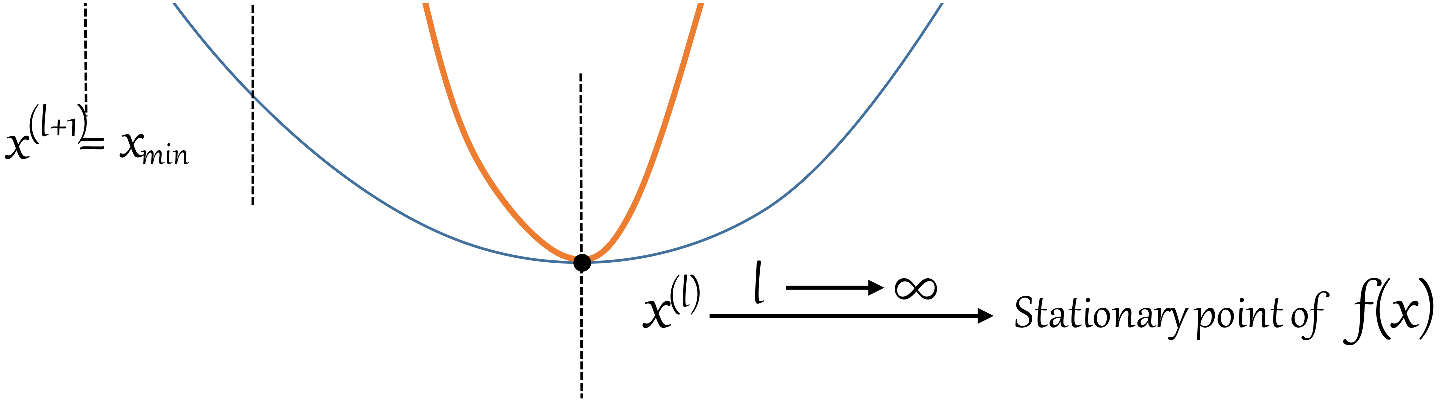
1. $u(x, x) = f(x)$ for all $x \in \mathcal{X}$;

2. $u(x, x^{(l)}) \geq f(x)$ for all $(x, x^{(l)}) \in \mathcal{X}^2$.

3. $u'(x, x^{(l)}; d)|_{x=x^{(l)}} = f'(x^{(l)}; d)$ for all d such that $x^{(l)} + d \in \mathcal{X}$, where

$$u'(x, x^{(l)}; d) = \frac{\partial}{\partial x} u(x, x^{(l)}; d), \quad \text{and} \quad f'(x; d) = \frac{\partial}{\partial x} f(x; d)$$

4. $u(x, x^{(l)})$ is (jointly) continuous in $(x, x^{(l)})$.



$$\min \ f(\mathbf{W}, \mathbf{H} | \mathcal{D})$$

$$\mathbf{W}^{(l+1)} \longleftarrow u_1(\mathbf{W}, \mathbf{W}^{(l)} | \mathbf{H}^{(l)})$$

$$\mathbf{H}^{(l+1)} \longleftarrow u_2(\mathbf{H}, \mathbf{H}^{(l)} \, | \mathbf{W}^{(l+1)})$$

$$\sum_{m=1}^{M} \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \left[-\log\left([\mathbf{WH}]_{\boxed{mi}}\right) + \log\left([\mathbf{WH}]_{\boxed{mi}} + [\mathbf{WH}]_{\boxed{mj}}\right)\right]$$

$$\Big\Downarrow$$

$$-\sum_{(i,j)} b_{ij} [\log \boxed{\theta_i} - \log(\underbrace{\boxed{\theta_i} + \boxed{\theta_j}}_{y})]$$

By Taylor's Theorem $\longrightarrow$ $\log y \leq \log x + \frac{1}{x}(y - x).$

$$u(\boldsymbol{\theta}, \boldsymbol{\theta}^{(l)}) = -\sum_{i,j} b_{ij} \left[\log \theta_i - \log(\underbrace{\theta_i^{(l)} + \theta_j^{(l)}}_{x}) - \frac{\overbrace{\theta_i + \theta_j}^{y}}{\underbrace{\theta_i^{(l)} + \theta_j^{(l)}}_{x}} + 1 \right] \longrightarrow \theta_i^{(l+1)} \leftarrow \frac{\sum_{j \neq i} b_{ij}}{\sum_{j \neq i} (b_{ij} + b_{ji}) / (\theta_i^{(l)} + \theta_j^{(l)})}$$

$$\sum_{m=1}^{M} \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \left[-\log\left([\mathbf{WH}]_{mi}\right) + \log\left([\mathbf{WH}]_{mi} + [\mathbf{WH}]_{mj}\right)\right]$$

⬇

$$\tilde{u}_1(\mathbf{W}, \mathbf{W}^{(l)} | \mathbf{H}^{(l)}) = \sum_{m=1}^{M} \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \Bigg[\boxed{-\log\left([\mathbf{WH}]_{mi}\right)} + \log\left([\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mi} + [\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mj}\right) + \frac{[\mathbf{WH}^{(l)}]_{mi} + [\mathbf{WH}^{(l)}]_{mj}}{[\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mi} + [\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mj}} - 1 \Bigg]$$

By Jensen's Inequality
- log(x) is a convex function

$$\text{With } \lambda_{mki} = \frac{w_{mk}^{(l)} h_{ki}}{[\mathbf{W}^{(l)}\mathbf{H}]_{mi}}$$

⬇

$$-\log[\sum_k w_{mk} h_{ki}] = -\log\left[\sum_k \lambda_{mki} \frac{w_{mk} h_{ki}}{\lambda_{mki}}\right] \leq -\sum_k \lambda_{mki} \log\left[\frac{w_{mk} h_{ki}}{\lambda_{mki}}\right] = -\sum_k \frac{w_{mk}^{(l)} h_{ki}}{[\mathbf{W}^{(l)}\mathbf{H}]_{mi}} \log\left[\frac{w_{mk} h_{ki}}{w_{mk}^{(l)} h_{ki}} [\mathbf{W}^{(l)}\mathbf{H}]_{mi}\right]$$

$$f(\mathbb{E}(X)) \leqslant \mathbb{E}(f(X))$$

Auxiliary Functions

$$u_1(\mathbf{W},\mathbf{W}^{(l)}|\mathbf{H}^{(l)})$$

$$= \sum_m \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \left[-\sum_k \frac{w_{mk}^{(l)} h_{ki}^{(l)}}{[\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mi}} \log\left(\frac{w_{mk}}{w_{mk}^{(l)}}[\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mi}\right) \right.$$

$$\left. + \log\left([\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mi} + [\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mj}\right) + \frac{[\mathbf{W}\mathbf{H}^{(l)}]_{mi}+[\mathbf{W}\mathbf{H}^{(l)}]_{mj}}{[\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mi}+[\mathbf{W}^{(l)}\mathbf{H}^{(l)}]_{mj}} - 1\right]$$

$$u_2(\mathbf{H},\mathbf{H}^{(l)}|\mathbf{W}^{(l+1)})$$

$$= \sum_m \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \left[-\sum_k \frac{w_{mk}^{(l+1)} h_{ki}^{(l)}}{[\mathbf{W}^{(l+1)}\mathbf{H}^{(l)}]_{mi}} \log\left(\frac{h_{ki}}{h_{ki}^{(l)}}[\mathbf{W}^{(l+1)}\mathbf{H}^{(l)}]_{mi}\right) \right.$$

$$+ \log\left([\mathbf{W}^{(l+1)}\mathbf{H}^{(l)}]_{mi} + [\mathbf{W}^{(l+1)}\mathbf{H}^{(l)}]_{mj}\right)$$

$$\left. + \frac{[\mathbf{W}^{(l+1)}\mathbf{H}^{(l)}]_{mi}+[\mathbf{W}^{(l+1)}\mathbf{H}^{(l)}]_{mj}}{[\mathbf{W}^{(l+1)}\mathbf{H}^{(l)}]_{mi}+[\mathbf{W}^{(l+1)}\mathbf{H}^{(l)}]_{mj}} - 1\right]$$

Updates for MM algorithm

$$\tilde{w}_{mk}^{(l+1)} \leftarrow \frac{\sum_{(i,j)\in P_m : j\neq i} b_{ij}^{(m)} \frac{w_{mk}^{(l)} h_{ki}}{[\mathbf{W}^{(l)}\mathbf{H}]_{mi}}}{\sum_{(i,j)\in P_m : j\neq i} b_{ij}^{(m)} \frac{h_{ki}+h_{kj}}{[\mathbf{W}^{(l)}\mathbf{H}]_{mi}+[\mathbf{W}^{(l)}\mathbf{H}]_{mj}}}$$

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

1. BTL-NMF Algorithm
    - Majorization-Minimization Algorithm
    - **Resolution of Numerical Problems**
    - Normalization
    - Convergence Analysis

2. Experiments Using Real Life Dataset
    - Dataset Information
    - Running of Algorithm
    - Men Players
    - Women Players

3. Comparison with Mixture BTL
    - Qualitative Comparison
    - Prediction Task & Result Comparison

May Divide by 0 or small numbers

$$\tilde{w}_{mk}^{(l+1)} \leftarrow \frac{\displaystyle\sum_{(i,j)\in P_m : j\neq i} b_{ij}^{(m)} \frac{w_{mk}^{(l)} h_{ki}}{\boxed{[\mathbf{W}^{(l)}\mathbf{H}]_{mi}}}}{\boxed{\displaystyle\sum_{(i,j)\in P_m : j\neq i} b_{ij}^{(m)} \frac{h_{ki}+h_{kj}}{\boxed{[\mathbf{W}^{(l)}\mathbf{H}]_{mi}} + \boxed{[\mathbf{W}^{(l)}\mathbf{H}]_{mj}}}}}$$

$\longrightarrow 0$

$\longrightarrow \infty$

$\longrightarrow 0$

$$\tilde{h}_{ki}^{(l+1)} \leftarrow \frac{\displaystyle\sum_{m}\sum_{j\neq i} b_{ij}^{(m)} \frac{w_{mk} h_{ki}^{(l)}}{[\mathbf{W}\mathbf{H}^{(l)}]_{mi}}}{\displaystyle\sum_{m}\sum_{j\neq i} \left(b_{ij}^{(m)} + b_{ji}^{(m)}\right) \frac{w_{mk}}{[\mathbf{W}\mathbf{H}^{(l)}]_{mi} + [\mathbf{W}\mathbf{H}^{(l)}]_{mj}}}$$

Desired Properties

Property 1: Likelihood is always non-decreasing
(Objective Function should be non-increasing)

Property 2: W,H are non-negative

Property 3: No division by zero/Numerical problems

Proposed Solution

$+\,\epsilon$

Likelihood $(\mathbf{W},\mathbf{H}|\mathcal{D})$ = $\displaystyle\prod_{m=1}^{M}\prod_{(i,j)\in P_m}(\frac{[\mathbf{WH}]_{mi}}{[\mathbf{WH}]_{mi}+[\mathbf{WH}]_{mj}})^{b_{ij}^{(m)}}$

↓

$$\prod_{m=1}^{M}\prod_{(i,j)\in P_m}(\frac{[\mathbf{W}(\mathbf{H}+\epsilon\mathbb{1})]_{mi}}{[\mathbf{W}(\mathbf{H}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}(\mathbf{H}+\epsilon\mathbb{1}))_{mj}})^{b_{ij}^{(m)}}$$

$$\sum_{m=1}^{M}\sum_{(i,j)\in P_m} b_{ij}^{(m)}\Big[-\log(\sum_k w_{mk}h_{ki}) + \log(\sum_k w_{mk}h_{ki} + \sum_k w_{mk}h_{kj})\Big]$$

$$f_\epsilon(\mathbf{W},\mathbf{H}) := \sum_{m=1}^{M}\sum_{(i,j)\in P_m} b_{ij}^{(m)}\Big[-\log(\sum_k w_{mk}(h_{ki}+\epsilon)) + \log(\sum_k w_{mk}(h_{ki}+\epsilon) + \sum_k w_{mk}(h_{kj}+\epsilon))\Big]$$

$$\tilde{w}_{mk}^{(l+1)} \leftarrow \frac{\sum_{(i,j)\in P_m:j\neq i} b_{ij}^{(m)}\frac{w_{mk}^{(l)}h_{ki}}{[\mathbf{W}^{(l)}\mathbf{H}]_{mi}}}{\sum_{(i,j)\in P_m:j\neq i} b_{ij}^{(m)}\frac{h_{ki}+h_{kj}}{[\mathbf{W}^{(l)}\mathbf{H}]_{mi}+[\mathbf{W}^{(l)}\mathbf{H}]_{mj}}}$$

$$\tilde{w}_{mk}^{(l+1)} = \frac{\sum_{(i,j)} b_{ij}^{(m)}\frac{w_{mk}^{(l)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}}{\sum_{(i,j)} b_{ij}^{(m)}\frac{h_{ki}^{(l)}+h_{kj}^{(l)}+2\epsilon}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}}$$

$$\tilde{h}_{ki}^{(l+1)} \leftarrow \frac{\sum_{m} \sum_{j\neq i} b_{ij}^{(m)} \frac{w_{mk} h_{ki}^{(l)}}{[\mathbf{WH}^{(l)}]_{mi}}}{\sum_{m} \sum_{j\neq i} \left(b_{ij}^{(m)} + b_{ji}^{(m)}\right) \frac{w_{mk}}{[\mathbf{WH}^{(l)}]_{mi} + [\mathbf{WH}^{(l)}]_{mj}}}$$

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

$$\tilde{h}_{ki}^{(l+1)} \leftarrow \max \{\tilde{h}_{ki}^{(l+1)}, 0\}$$

Desired Properties

Property 1: Likelihood is always non-decreasing
(Objective Function should be non-increasing)

Property 2: W,H are non-negative √

Property 3: No division by zero √

How to make sure the likelihood is non-decreasing
or the negative log-likelihood is non-increasing?

Suppose $\tilde{h}_{ki}^{(l+1)} = 0$ and $\tilde{h}_{k',i'}^{(l+1)} = \tilde{h}_{k',i'}^{(l)}$ for all $(k',i') \neq (k,i)$.

Want to show:

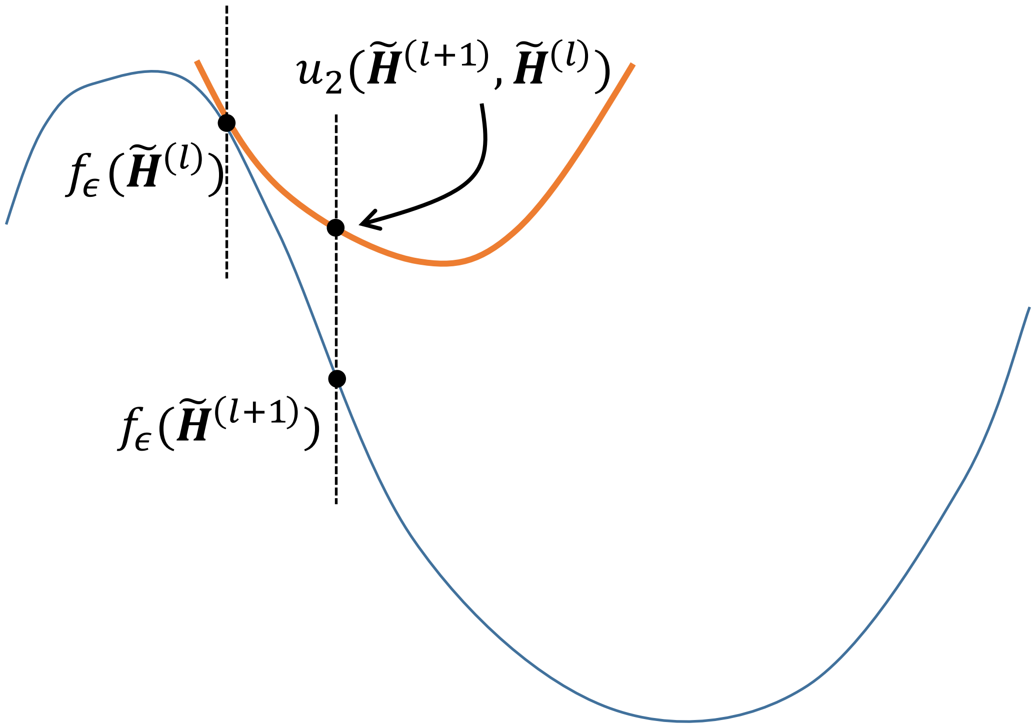
$$f_\epsilon\left(\boldsymbol{W},\widetilde{\boldsymbol{H}}^{(l+1)}\right) \leq f_\epsilon\left(\boldsymbol{W},\widetilde{\boldsymbol{H}}^{(l)}\right)$$

Always satisfied
by the property of auxiliary function:

$$f_\epsilon\left(\boldsymbol{W},\widetilde{\boldsymbol{H}}^{(l+1)}\right) \leq u_2\left(\widetilde{\boldsymbol{H}}^{(l+1)},\widetilde{\boldsymbol{H}}^{(l)}|\boldsymbol{W}\right)$$

Suffices to show:

$$u_2\left(\widetilde{\boldsymbol{H}}^{(l+1)},\widetilde{\boldsymbol{H}}^{(l)}|\boldsymbol{W}\right) \leq f_\epsilon\left(\boldsymbol{W},\widetilde{\boldsymbol{H}}^{(l)}\right)$$



$$f_\epsilon(\mathbf{W},\tilde{\mathbf{H}}^{(l)}) - u_2(\tilde{\mathbf{H}}^{(l+1)},\tilde{\mathbf{H}}^{(l)}|\mathbf{W}) \quad \geq 0$$

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

same

same

Notice that the truncation is invoked only when
$$\frac{\boxed{\sum_{m} \sum_{j\neq i:(i,j)\in\mathcal{P}_m} b_{ij}^{(m)} \frac{w_{mk}^{(l+1)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}}}{\boxed{\sum_{m} \sum_{j\neq i:(i,j)\in\mathcal{P}_m} \frac{(b_{ij}^{(m)}+b_{ji}^{(m)})w_{mk}^{(l+1)}}{[\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}^{(l+1)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mj}}}} \leq \epsilon.$$

$$f_\epsilon(\mathbf{W},\tilde{\mathbf{H}}^{(l)}) - u_2(\tilde{\mathbf{H}}^{(l+1)},\tilde{\mathbf{H}}^{(l)}|\mathbf{W})$$

$$\geq -\log\left(\frac{h_{ki}^{(l)}+\epsilon}{\epsilon}\right)\cdot\epsilon\cdot\sum_{m=1}^{M}\sum_{\substack{j\neq i:\\(i,j)\in\mathcal{P}_m}}(b_{ij}^{(m)}+b_{ji}^{(m)})\frac{w_{mk}}{[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mi}+[\mathbf{W}(\tilde{\mathbf{H}}^{(l)}+\epsilon\mathbb{1})]_{mj}}$$

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

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

≥ 0

$$-\epsilon\log(\frac{\tilde{h}_{ki}^{(l)}+\epsilon}{\epsilon})+\tilde{h}_{ki}^{(l)}\geq 0$$

$$\Rightarrow\frac{\tilde{h}_{ki}^{(l)}}{\epsilon}\geq\log(\frac{\tilde{h}_{ki}^{(l)}}{\epsilon}+1) \quad \Longleftarrow \quad e^x\geq x+1,\forall x\geq 0$$

$$\Rightarrow\exp(\frac{\tilde{h}_{ki}^{(l)}}{\epsilon})\geq\frac{\tilde{h}_{ki}^{(l)}}{\epsilon}+1$$

`Property 1: likelihood is always non-decreasing` √

`Property 2: W,H are non-negative` √

`Property 3: No division by zero` √

1. BTL-NMF Algorithm
    - Majorization-Minimization Algorithm
    - Resolution of Numerical Problems
    - **Normalization**
    - Convergence Analysis

2. Experiments Using Real Life Dataset
    - Dataset Information
    - Running of Algorithm
    - Men Players
    - Women Players

3. Comparison with Mixture BTL
    - Qualitative Comparison
    - Prediction Task & Result Comparison

Pr(i beats j in tournament m) =

$$\frac{\sum_k \tilde{w}_{mk}(\tilde{h}_{ki}+\epsilon)}{\sum_k \tilde{w}_{mk}(\tilde{h}_{ki}+\epsilon)+\sum_k \tilde{w}_{mk}(\tilde{h}_{kj}+\epsilon)}$$

Row Normalization of **W** and Global Normalization of **H**:

Keep likelihood unchanged $\sum_k w_{mk} = 1, \sum_{k,i} h_{ki} = 1$

$$\frac{\sum_k \tilde{w}_{mk}(\tilde{h}_{ki}+\epsilon)}{\sum_k \tilde{w}_{mk}(\tilde{h}_{ki}+\epsilon) + \sum_k \tilde{w}_{mk}(\tilde{h}_{kj}+\epsilon)}$$

$$\boxed{c_m := \sum_k \tilde{w}_{mk}} \quad \Downarrow$$

$$\frac{\sum_k \frac{\tilde{w}_{mk}}{c_m}(\tilde{h}_{ki}+\epsilon)}{\sum_k \frac{\tilde{w}_{mk}}{c_m}(\tilde{h}_{ki}+\epsilon) + \sum_k \frac{\tilde{w}_{mk}}{c_m}(\tilde{h}_{kj}+\epsilon)}$$

$$\boxed{\alpha = ?} \quad \Downarrow$$

$$\frac{\sum_k w_{mk}\frac{(\tilde{h}_{ki}+\epsilon)}{\alpha}}{\sum_k w_{mk}\frac{(\tilde{h}_{ki}+\epsilon)}{\alpha} + \sum_k w_{mk}\frac{(\tilde{h}_{ki}+\epsilon)}{\alpha}}$$

$$w_{mk} \leftarrow \frac{\tilde{w}_{mk}}{c_m}$$

$$h_{ki} \leftarrow \frac{\tilde{h}_{ki} + (1-\alpha)\epsilon}{\alpha}$$

$$\frac{\tilde{h}_{ki}+\epsilon}{\alpha} = \frac{\tilde{h}_{ki} + (1-\alpha)\epsilon}{\alpha} + \epsilon \Rightarrow \sum_{k,i} \frac{\tilde{h}_{ki} + (1-\alpha)\epsilon}{\alpha} = 1 \Rightarrow \alpha := \frac{\sum_{k,i} \tilde{h}_{ki} + KN\epsilon}{1 + KN\epsilon}$$

Column Normalization of **W** and Global Normalization of **H**:

Keep likelihood unchanged $\sum_m w_{mk} = 1, \sum_{k,i} h_{ki} = 1$

$$\frac{\sum_k \tilde{w}_{mk}(\tilde{h}_{ki}+\epsilon)}{\sum_k \tilde{w}_{mk}(\tilde{h}_{ki}+\epsilon)+\sum_k \tilde{w}_{mk}(\tilde{h}_{kj}+\epsilon)}$$

$$\boxed{d_k := \sum_m \tilde{w}_{mk}} \quad \Big\downarrow$$

$$\frac{\sum_k \frac{\tilde{w}_{mk}}{d_k}(\tilde{h}_{ki}+\epsilon)d_k}{\sum_k \frac{\tilde{w}_{mk}}{d_k}(\tilde{h}_{ki}+\epsilon)d_k+\sum_k \frac{\tilde{w}_{mk}}{d_k}(\tilde{h}_{kj}+\epsilon)d_k}$$

$$\boxed{\alpha = ?} \quad \Big\downarrow$$

$$\frac{\sum_k w_{mk}\frac{(\hat{h}_{ki}+\epsilon)}{\beta}}{\sum_k w_{mk}\frac{(\hat{h}_{ki}+\epsilon)}{\beta}+\sum_k w_{mk}\frac{(\hat{h}_{ki}+\epsilon)}{\beta}}$$

$$w_{mk} \leftarrow \frac{\tilde{w}_{mk}}{d_k} \qquad \hat{h}_{ki} \leftarrow \tilde{h}_{ki}d_k+\epsilon(d_k-1)$$

$$h_{ki} \leftarrow \frac{\hat{h}_{ki}+(1-\beta)\epsilon}{\beta}$$

$$\beta := \frac{\sum_{k,i}\hat{h}_{ki}+KN\epsilon}{1+KN\epsilon}$$

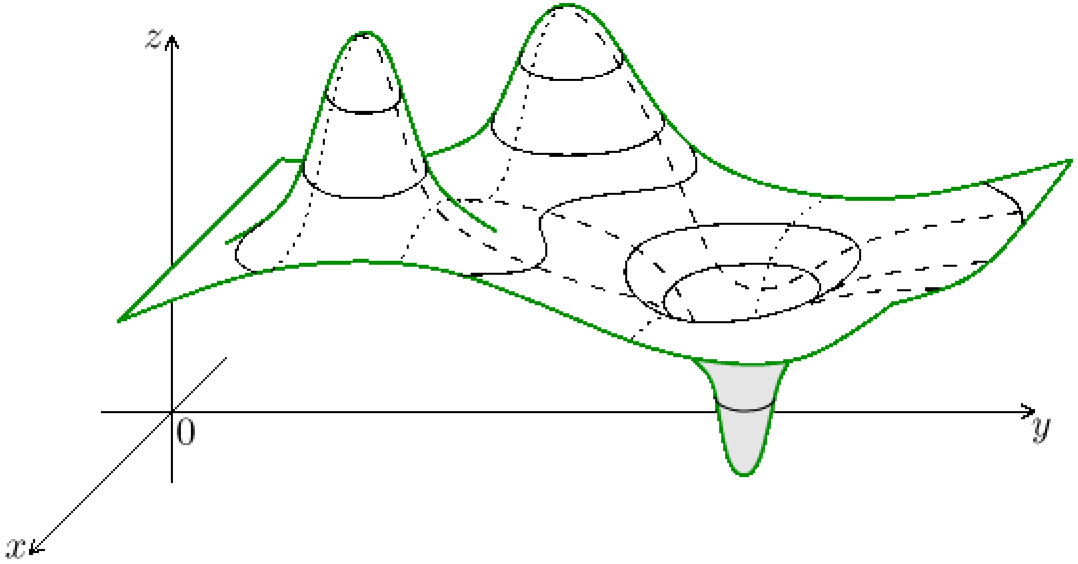
$$\sum_m \sum_i [\mathbf{\Lambda}]_{mi} = \sum_m \sum_i \sum_k w_{mk}h_{ki} = \sum_i \sum_k h_{ki} \sum_m w_{mk} = \sum_{k,i} h_{ki} = 1$$

1. BTL-NMF Algorithm
    - Majorization-Minimization Algorithm
    - Resolution of Numerical Problems
    - Normalization
    - **Convergence Analysis**
2. Experiments Using Real Life Dataset
    - Dataset Information
    - Running of Algorithm
    - Men Players
    - Women Players
3. Comparison with Mixture BTL
    - Qualitative Comparison
    - Prediction Task & Result Comparison

$$x^{(l)} \xrightarrow{l \longrightarrow \infty} \text{Stationary point of } f(x)$$

Directional Derivative:

$$f'(x;d) = \lim_{\lambda \to 0} \frac{f(x+\lambda d) - f(x)}{\lambda}$$



Stationary Point $(\overline{\boldsymbol{W}}, \overline{\boldsymbol{H}})$ of:

$$\min_{\boldsymbol{W} \in \boldsymbol{R}_{++}^{M\times K},\ \boldsymbol{H} \in \boldsymbol{R}_{++}^{K\times N}} f_\epsilon\ (\boldsymbol{W}, \boldsymbol{H})$$

If for $f_1(\boldsymbol{W} |\ \overline{\boldsymbol{H}}) = f_\epsilon(\boldsymbol{W}, \overline{\boldsymbol{H}})$, $f_2(\boldsymbol{H} |\ \overline{\boldsymbol{W}}) = f_\epsilon(\overline{\boldsymbol{W}}, \boldsymbol{H})$:

$$\begin{array}{ll} f_1'(\overline{\boldsymbol{W}}; \boldsymbol{W} - \overline{\boldsymbol{W}} |\ \overline{\boldsymbol{H}}) \geq 0, & \forall\, \boldsymbol{W} \in \boldsymbol{R}_{++}^{M\times K} \\ f_2'(\overline{\boldsymbol{H}}; \boldsymbol{H} - \overline{\boldsymbol{H}} |\ \overline{\boldsymbol{W}}) \geq 0, & \forall\, \boldsymbol{H} \in \boldsymbol{R}_{++}^{K\times N} \end{array}$$

Convergence analysis of block successive minimization methods

Given $f(x)$ to be minimized on domain $\chi = \prod_{i=1}^{n} \chi_i$

$\boldsymbol{n = 2}$, $\boldsymbol{\chi = R_{++}^{M\times K} \times R_{++}^{K\times N}}$

$\boldsymbol{f(x) = f_{\epsilon}(W, H)}$

√ **(P1)** $F_i(\tilde{x}_i|\tilde{x}) = f(\tilde{x})$, for any $\tilde{x} \in \chi$

$\boldsymbol{f_1(W|H^{(l)}) = f_{\epsilon}(W, H^{(l)})}$

$\boldsymbol{f_2(H|W^{(l+1)}) = f_{\epsilon}(W^{(l+1)}, H)}$

√ **(P2)** $F_i(x_i|\tilde{x}) \le f(\tilde{x}_1, \dots, x_i, \dots, \tilde{x}_n)$, for any $(x_i, \tilde{x}) \in \chi_i \times \chi$

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

$\boldsymbol{F_1(x_1|\tilde{x}) = u_1(W, W^{(l)}|H^{(l)})}$

$\boldsymbol{F_2(x_2|\tilde{x}) = u_2(H, H^{(l)}|W^{(l+1)})}$

(P4) Define $f_i(\cdot\,|\tilde{x}) : x_i \mapsto f(\tilde{x}_1, \dots, x_i, \dots, \tilde{x}_n)$, for any $(x_i, \tilde{x}) \in \chi_i \times \chi$
Then for any $\hat{x} \in \chi_i$, $F_i'(x_i; \hat{x}_i - x_i|\tilde{x}_i)|_{x_i = \tilde{x}_i} = f_i'(x_i; \hat{x}_i - x_i|\tilde{x}_i)\,|_{x_i = \tilde{x}_i}$

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

M. Razaviyayn, M. Hong, and Z.-Q. Luo. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. SIAM Journal on Optimization , 23(2):1126--1153, 2013.

✓ **(P3)** u_1 and u_2 are both differentiable

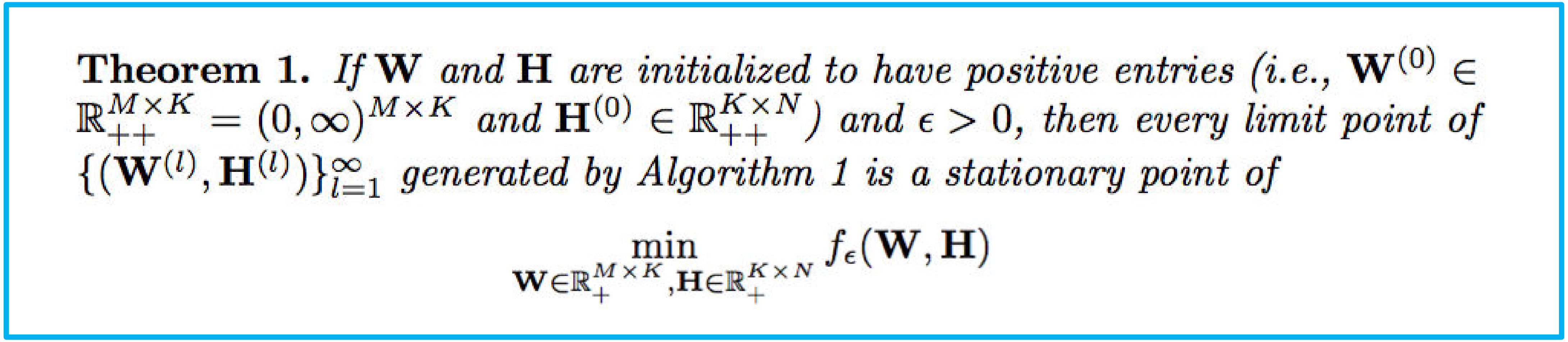
✓ **(P4)** $$u_1'(\boldsymbol{W};\widehat{\boldsymbol{W}}-\boldsymbol{W}|\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})|_{\boldsymbol{W}=\widetilde{\boldsymbol{W}}} = f_1'(\boldsymbol{W};\widehat{\boldsymbol{W}}-\boldsymbol{W}|\widetilde{\boldsymbol{H}})$$

$$u_2'(\boldsymbol{H};\widehat{\boldsymbol{H}}-\boldsymbol{H}|\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})|_{\boldsymbol{H}=\widetilde{\boldsymbol{H}}} = f_2'(\boldsymbol{H};\widehat{\boldsymbol{H}}-\boldsymbol{H}|\widetilde{\boldsymbol{W}})$$

✓ **(P5)** $$\frac{\partial^2}{\partial w_{mk}^2}u_1(\mathbf{W},\mathbf{W}^{(l)}|\mathbf{H}^{(l)}) = \sum_{(i,j)\in\mathcal{P}_m} b_{ij}^{(m)}\left(\frac{w_{mk}^{(l)}(h_{ki}^{(l)}+\epsilon)}{[\mathbf{W}^{(l)}(\mathbf{H}^{(l)}+\epsilon\mathbb{1})]_{mi}}\frac{1}{w_{mk}^2}\right)$$

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

$w_{mk} > 0, h_{ki} \geq 0$



Theorem 1. *If* $\mathbf{W}$ *and* $\mathbf{H}$ *are initialized to have positive entries (i.e.,* $\mathbf{W}^{(0)} \in \mathbb{R}_{++}^{M\times K} = (0,\infty)^{M\times K}$ *and* $\mathbf{H}^{(0)} \in \mathbb{R}_{++}^{K\times N}$*) and* $\epsilon > 0$*, then every limit point of* $\{(\mathbf{W}^{(l)},\mathbf{H}^{(l)})\}_{l=1}^{\infty}$ *generated by Algorithm 1 is a stationary point of*

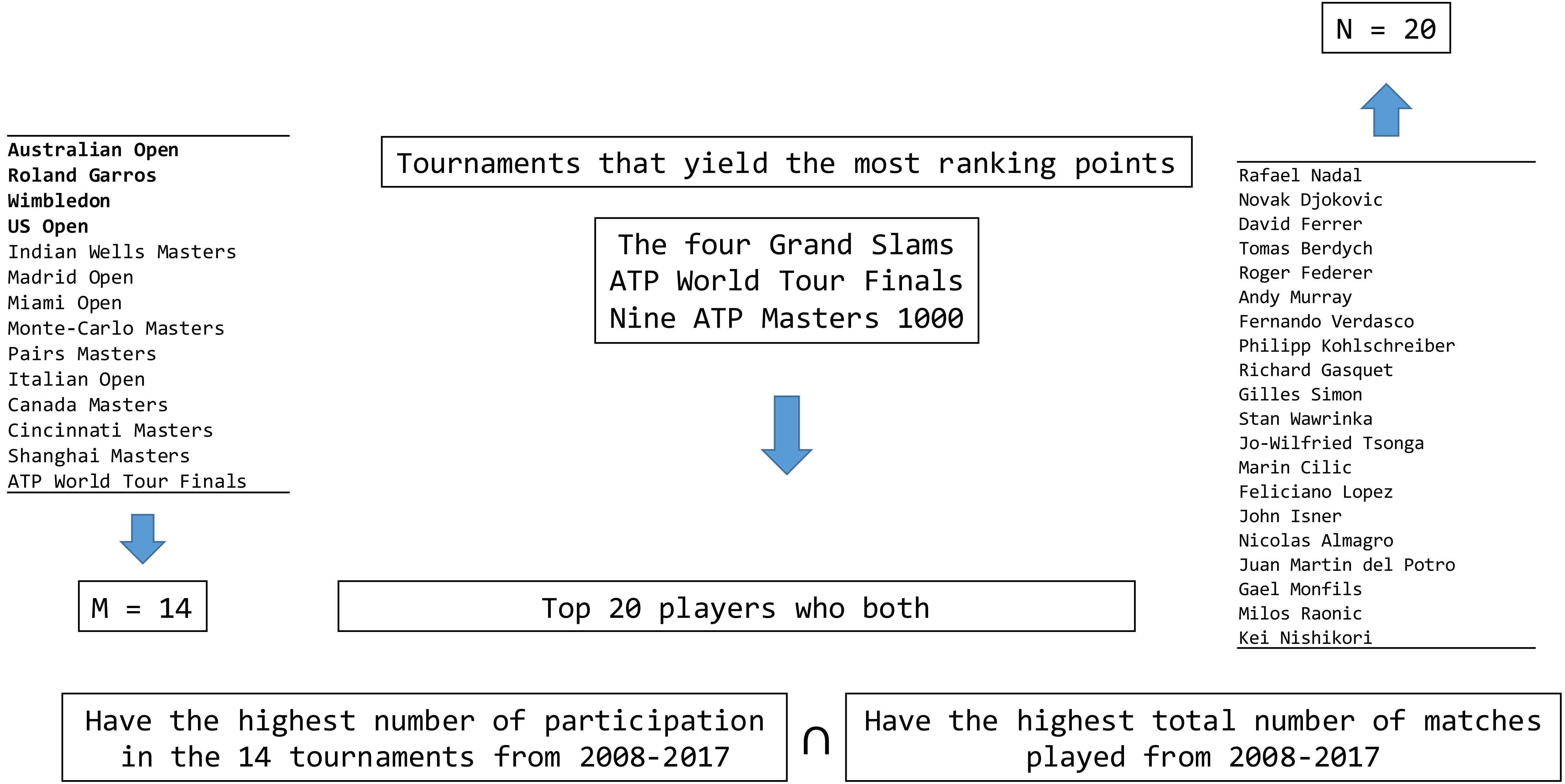
$$\min_{\mathbf{W}\in\mathbb{R}_+^{M\times K},\mathbf{H}\in\mathbb{R}_+^{K\times N}} f_\epsilon(\mathbf{W},\mathbf{H})$$

R. Zhao and V. Y. F. Tan, “A Unified Convergence Analysis of the Multiplicative Update Algorithm for Regularized NMF”, IEEE Transactions on Signal Processing, Vol. 66, No. 1, Pages 129 – 138, Jan 2018

1. BTL-NMF Algorithm
    - Majorization-Minimization Algorithm
    - Resolution of Numerical Problems
    - Normalization
    - Convergence Analysis

2. Experiments Using Real Life Dataset
    - **Dataset Information**
    - Running of Algorithm
    - Men Players
    - Women Players

3. Comparison with Mixture BTL
    - Qualitative Comparison
    - Prediction Task & Result Comparison



Tournaments that yield the most ranking points

The four Grand Slams
ATP World Tour Finals
Nine ATP Masters 1000

Australian Open
Roland Garros
Wimbledon
US Open
Indian Wells Masters
Madrid Open
Miami Open
Monte-Carlo Masters
Pairs Masters
Italian Open
Canada Masters
Cincinnati Masters
Shanghai Masters
ATP World Tour Finals

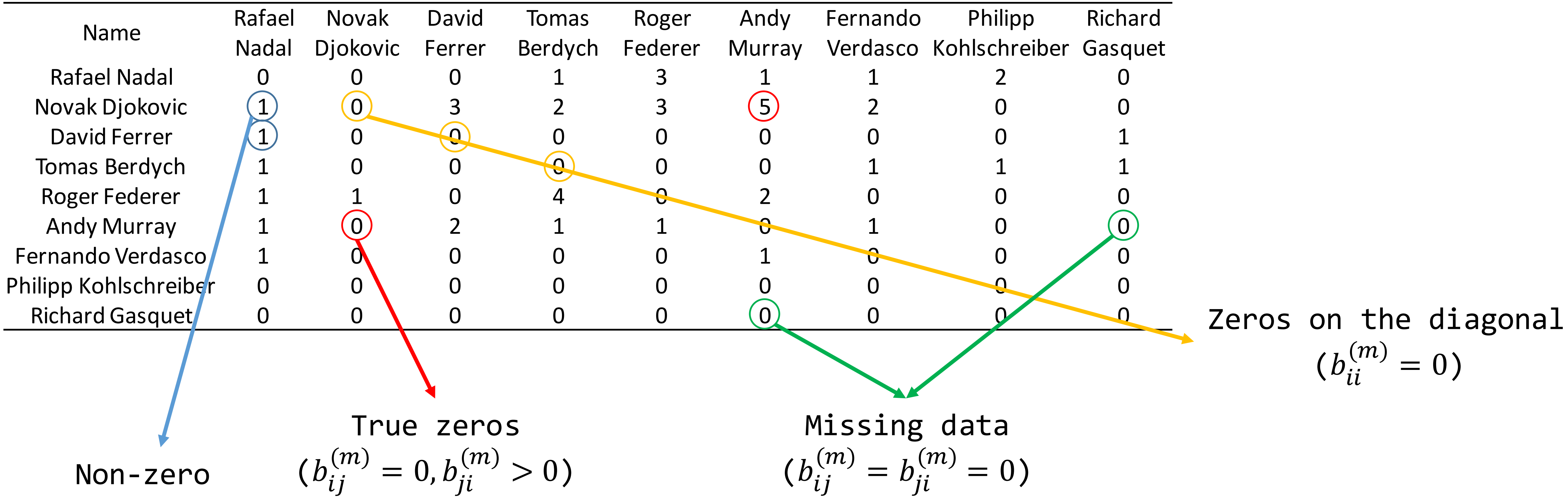
M = 14

Top 20 players who both

Have the highest number of participation in the 14 tournaments from 2008-2017 ∩ Have the highest total number of matches played from 2008-2017

Rafael Nadal
Novak Djokovic
David Ferrer
Tomas Berdych
Roger Federer
Andy Murray
Fernando Verdasco
Philipp Kohlschreiber
Richard Gasquet
Gilles Simon
Stan Wawrinka
Jo-Wilfried Tsonga
Marin Cilic
Feliciano Lopez
John Isner
Nicolas Almagro
Juan Martin del Potro
Gael Monfils
Milos Raonic
Kei Nishikori

N = 20



Name	Rafael Nadal	Novak Djokovic	David Ferrer	Tomas Berdych	Roger Federer	Andy Murray	Fernando Verdasco	Philipp Kohlschreiber	Richard Gasquet
Rafael Nadal	0	0	0	1	3	1	1	2	0
Novak Djokovic	1	0	3	2	3	5	2	0	0
David Ferrer	1	0	0	0	0	0	0	0	1
Tomas Berdych	1	0	0	0	0	0	1	1	1
Roger Federer	1	1	0	4	0	2	0	0	0
Andy Murray	1	0	2	1	1	0	1	0	0
Fernando Verdasco	1	0	0	0	0	1	0	0	0
Philipp Kohlschreiber	0	0	0	0	0	0	0	0	0
Richard Gasquet	0	0	0	0	0	0	0	0	0

Zeros on the diagonal

$(b_{ii}^{(m)} = 0)$

Non-zero

True zeros

$(b_{ij}^{(m)} = 0, b_{ji}^{(m)} > 0)$

Missing data

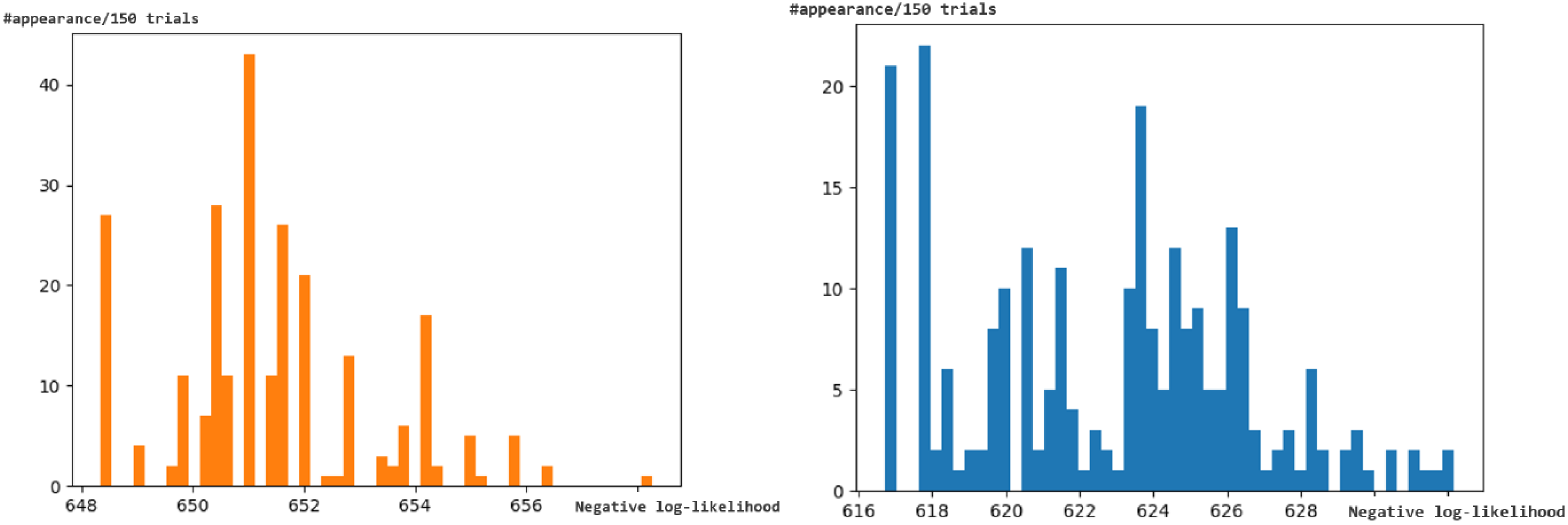
$(b_{ij}^{(m)} = b_{ji}^{(m)} = 0)$

	Male		**Female**	
Total Entries	$14 \times 20 \times 20 = 5600$		$16 \times 20 \times 20 = 6400$	
	Number	Percentage	Number	Percentage
Non-zero	1024	18.30%	788	12.31%
Zeros on the diagonal	280	5.00%	320	5.00%
Missing data	3478	62.10%	4598	71.84%
True zeros	818	14.60%	694	10.85%

1. BTL-NMF Algorithm
   - Majorization-Minimization Algorithm
   - Resolution of Numerical Problems
   - Normalization
   - Convergence Analysis
2. Experiments Using Real Life Dataset
   - Dataset Information
   - **Running of Algorithm**
   - Men Players
   - Women Players
3. Comparison with Mixture BTL
   - Qualitative Comparison
   - Prediction Task & Result Comparison

Stopping condition:

$$\text{diff} \leftarrow \max\left\{ \max_{m,k} \left| w_{mk}^{(l+1)} - w_{mk}^{(l)} \right|, \max_{k,i} \left| h_{ki}^{(l+1)} - h_{ki}^{(l)} \right| \right\} < \mathbf{10^{-6}}$$



(A) $K = 2$

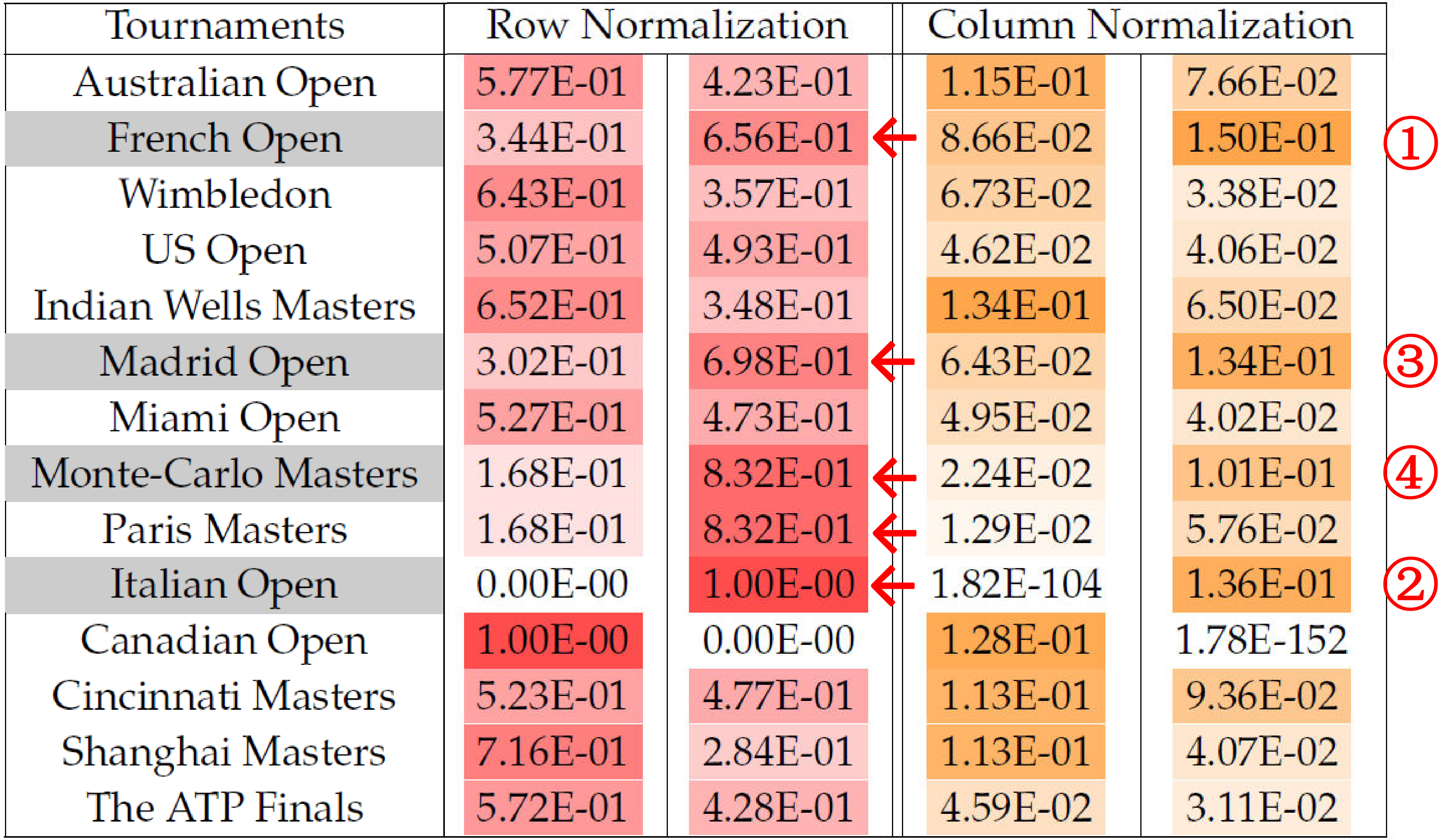
(B) $K = 3$

1. BTL-NMF Algorithm
    - Majorization-Minimization Algorithm
    - Resolution of Numerical Problems
    - Normalization
    - Convergence Analysis

2. Experiments Using Real Life Dataset
    - Dataset Information
    - Running of Algorithm
    - **Men Players**
    - Women Players

3. Comparison with Mixture BTL
    - Qualitative Comparison
    - Prediction Task & Result Comparison

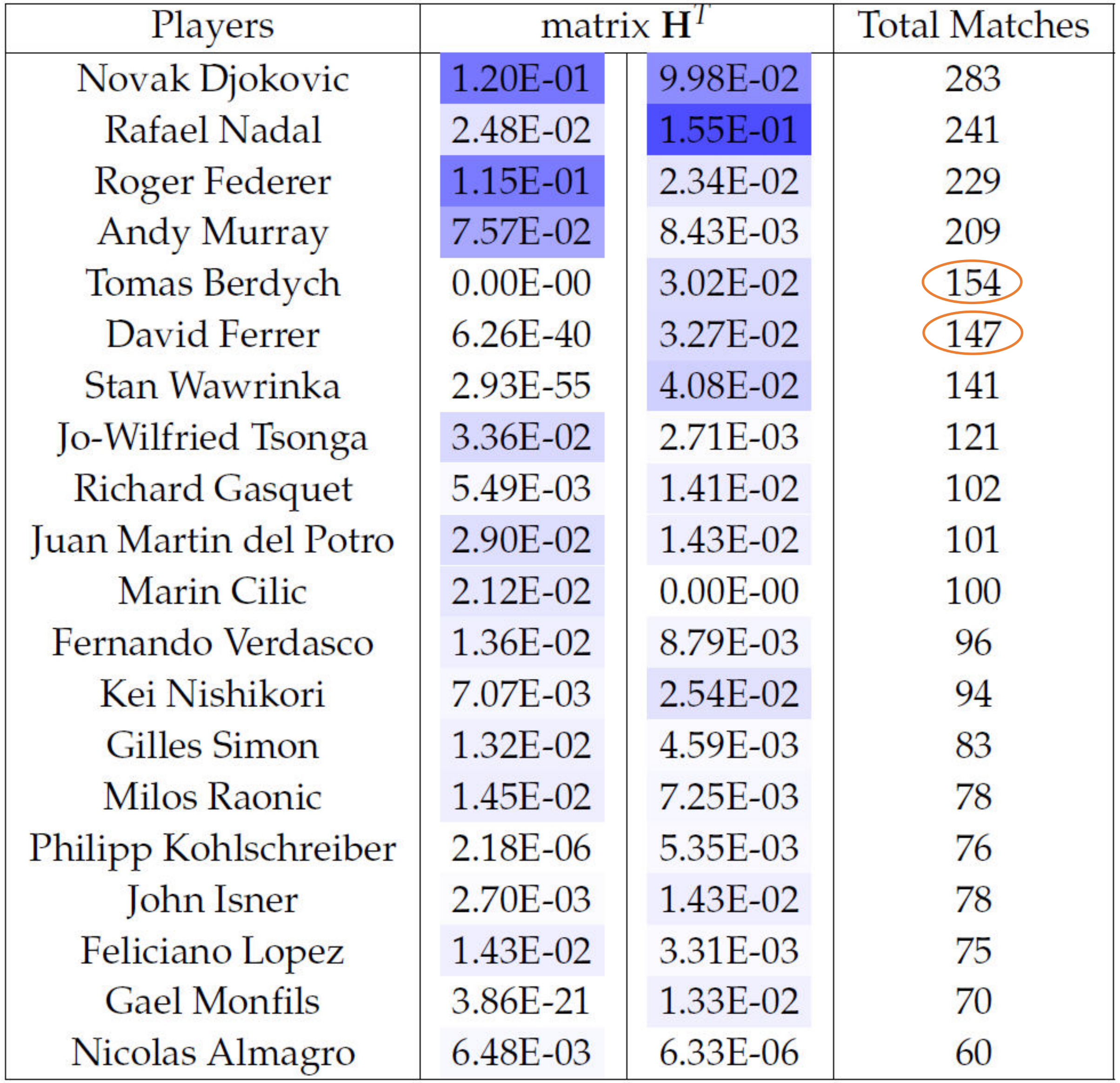
Learned W Dictionary Matrix for men



Tournaments	Row Normalization			Column Normalization		
Australian Open	5.77E-01	4.23E-01		1.15E-01	7.66E-02	
French Open	3.44E-01	6.56E-01	←	8.66E-02	1.50E-01	①
Wimbledon	6.43E-01	3.57E-01		6.73E-02	3.38E-02	
US Open	5.07E-01	4.93E-01		4.62E-02	4.06E-02	
Indian Wells Masters	6.52E-01	3.48E-01		1.34E-01	6.50E-02	
Madrid Open	3.02E-01	6.98E-01	←	6.43E-02	1.34E-01	③
Miami Open	5.27E-01	4.73E-01		4.95E-02	4.02E-02	
Monte-Carlo Masters	1.68E-01	8.32E-01	←	2.24E-02	1.01E-01	④
Paris Masters	1.68E-01	8.32E-01	←	1.29E-02	5.76E-02	
Italian Open	0.00E-00	1.00E-00	←	1.82E-104	1.36E-01	②
Canadian Open	1.00E-00	0.00E-00		1.28E-01	1.78E-152	
Cincinnati Masters	5.23E-01	4.77E-01		1.13E-01	9.36E-02	
Shanghai Masters	7.16E-01	2.84E-01		1.13E-01	4.07E-02	
The ATP Finals	5.72E-01	4.28E-01		4.59E-02	3.11E-02	

Hard Court player →
Clay player →
Grass player →
Non-clay player →

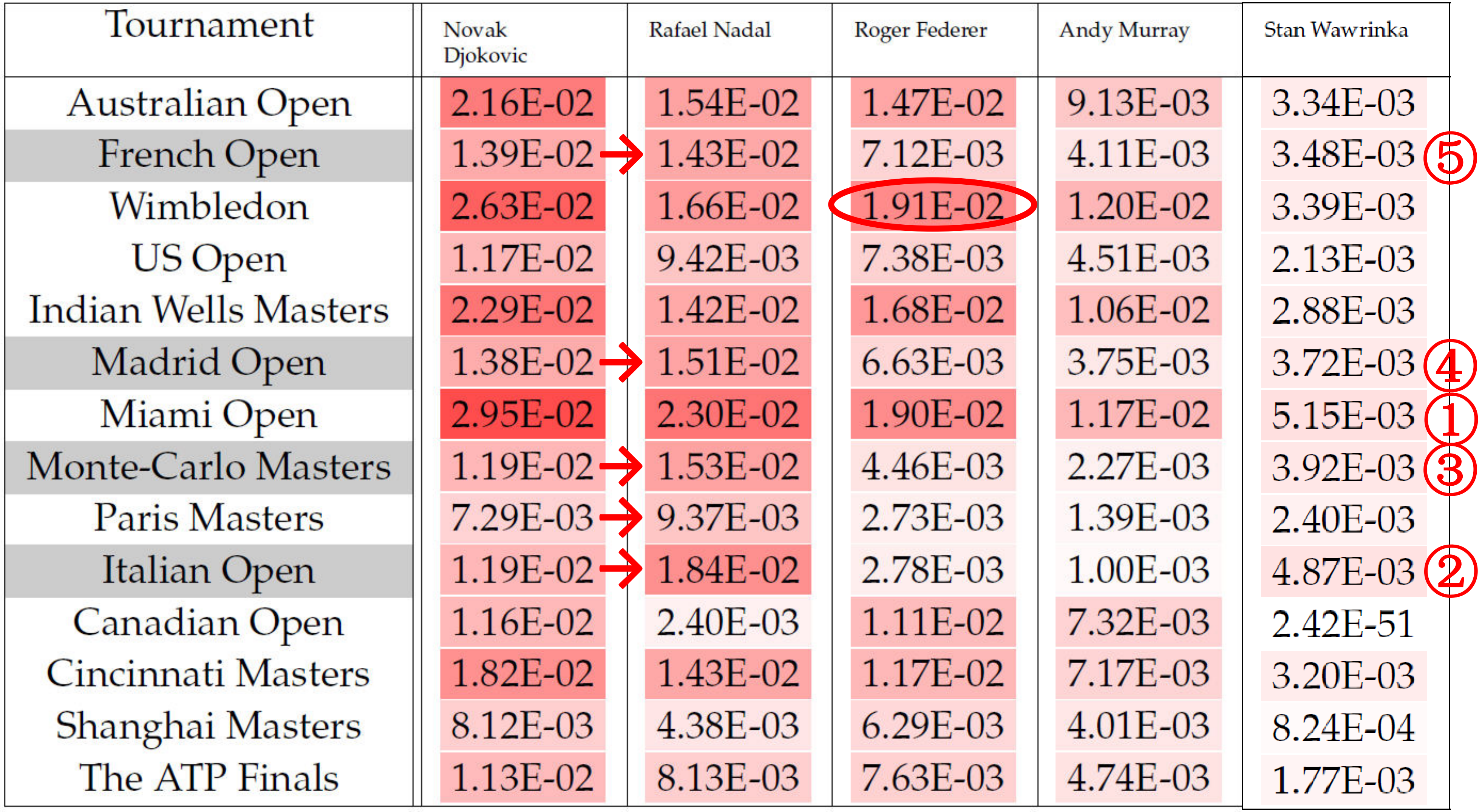
Clay player →



Players	matrix $\mathbf{H}^T$		Total Matches
	non-clay	clay	
Novak Djokovic	1.20E-01	9.98E-02	283
Rafael Nadal	2.48E-02	1.55E-01	241
Roger Federer	1.15E-01	2.34E-02	229
Andy Murray	7.57E-02	8.43E-03	209
Tomas Berdych	0.00E-00	3.02E-02	154
David Ferrer	6.26E-40	3.27E-02	147
Stan Wawrinka	2.93E-55	4.08E-02	141
Jo-Wilfried Tsonga	3.36E-02	2.71E-03	121
Richard Gasquet	5.49E-03	1.41E-02	102
Juan Martin del Potro	2.90E-02	1.43E-02	101
Marin Cilic	2.12E-02	0.00E-00	100
Fernando Verdasco	1.36E-02	8.79E-03	96
Kei Nishikori	7.07E-03	2.54E-02	94
Gilles Simon	1.32E-02	4.59E-03	83
Milos Raonic	1.45E-02	7.25E-03	78
Philipp Kohlschreiber	2.18E-06	5.35E-03	76
John Isner	2.70E-03	1.43E-02	78
Feliciano Lopez	1.43E-02	3.31E-03	75
Gael Monfils	3.86E-21	1.33E-02	70
Nicolas Almagro	6.48E-03	6.33E-06	60

Learned Transpose H Coefficient Matrix for men

$\boldsymbol{\Lambda} = \mathbf{WH}$



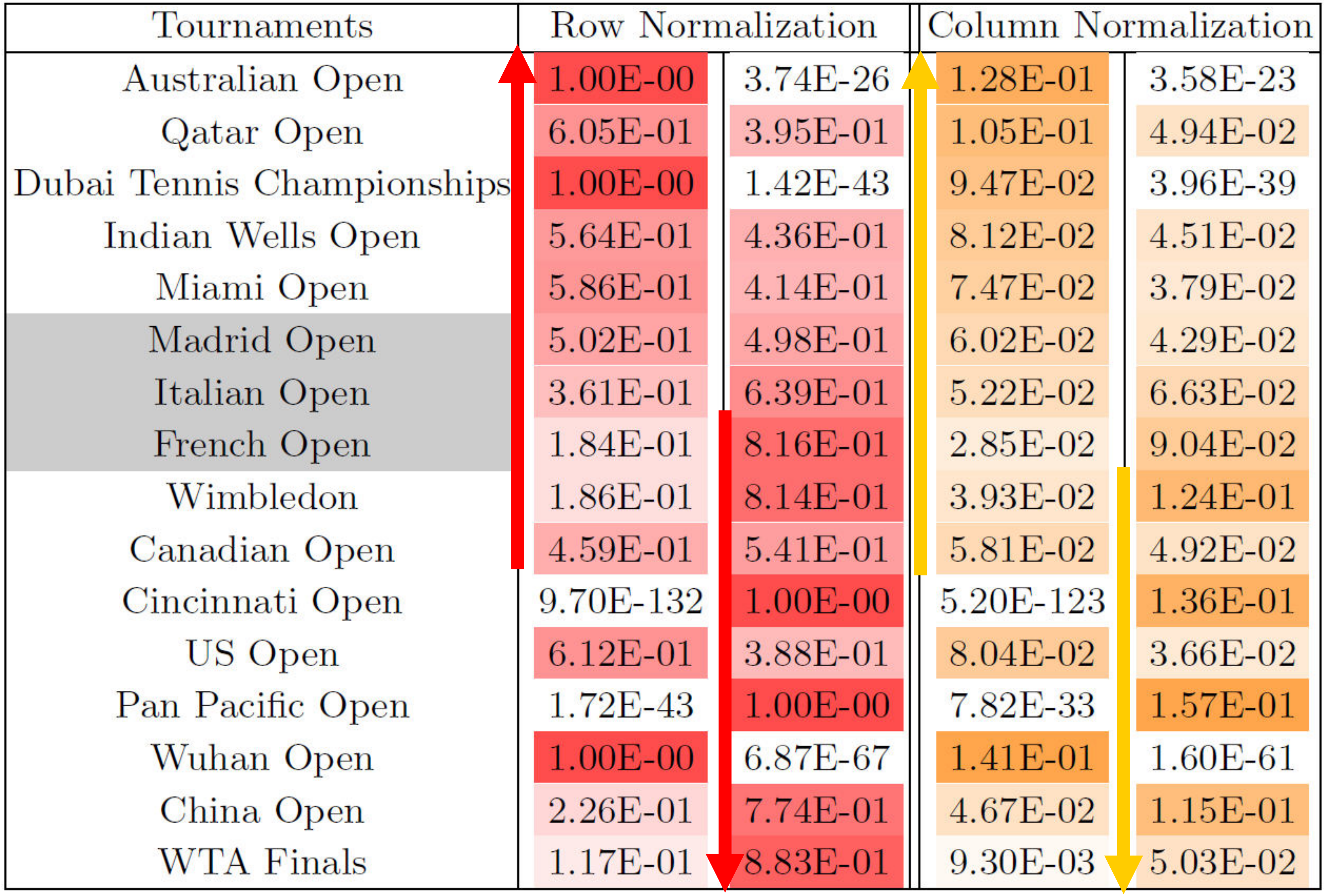
Tournament	Novak Djokovic	Rafael Nadal	Roger Federer	Andy Murray	Stan Wawrinka
Australian Open	2.16E-02	1.54E-02	1.47E-02	9.13E-03	3.34E-03
French Open	1.39E-02 →	1.43E-02	7.12E-03	4.11E-03	3.48E-03 ⑤
Wimbledon	2.63E-02	1.66E-02	1.91E-02	1.20E-02	3.39E-03
US Open	1.17E-02	9.42E-03	7.38E-03	4.51E-03	2.13E-03
Indian Wells Masters	2.29E-02	1.42E-02	1.68E-02	1.06E-02	2.88E-03
Madrid Open	1.38E-02 →	1.51E-02	6.63E-03	3.75E-03	3.72E-03 ④
Miami Open	2.95E-02	2.30E-02	1.90E-02	1.17E-02	5.15E-03 ①
Monte-Carlo Masters	1.19E-02 →	1.53E-02	4.46E-03	2.27E-03	3.92E-03 ③
Paris Masters	7.29E-03 →	9.37E-03	2.73E-03	1.39E-03	2.40E-03
Italian Open	1.19E-02 →	1.84E-02	2.78E-03	1.00E-03	4.87E-03 ②
Canadian Open	1.16E-02	2.40E-03	1.11E-02	7.32E-03	2.42E-51
Cincinnati Masters	1.82E-02	1.43E-02	1.17E-02	7.17E-03	3.20E-03
Shanghai Masters	8.12E-03	4.38E-03	6.29E-03	4.01E-03	8.24E-04
The ATP Finals	1.13E-02	8.13E-03	7.63E-03	4.74E-03	1.77E-03

1. BTL-NMF Algorithm
    - Majorization-Minimization Algorithm
    - Resolution of Numerical Problems
    - Normalization
    - Convergence Analysis

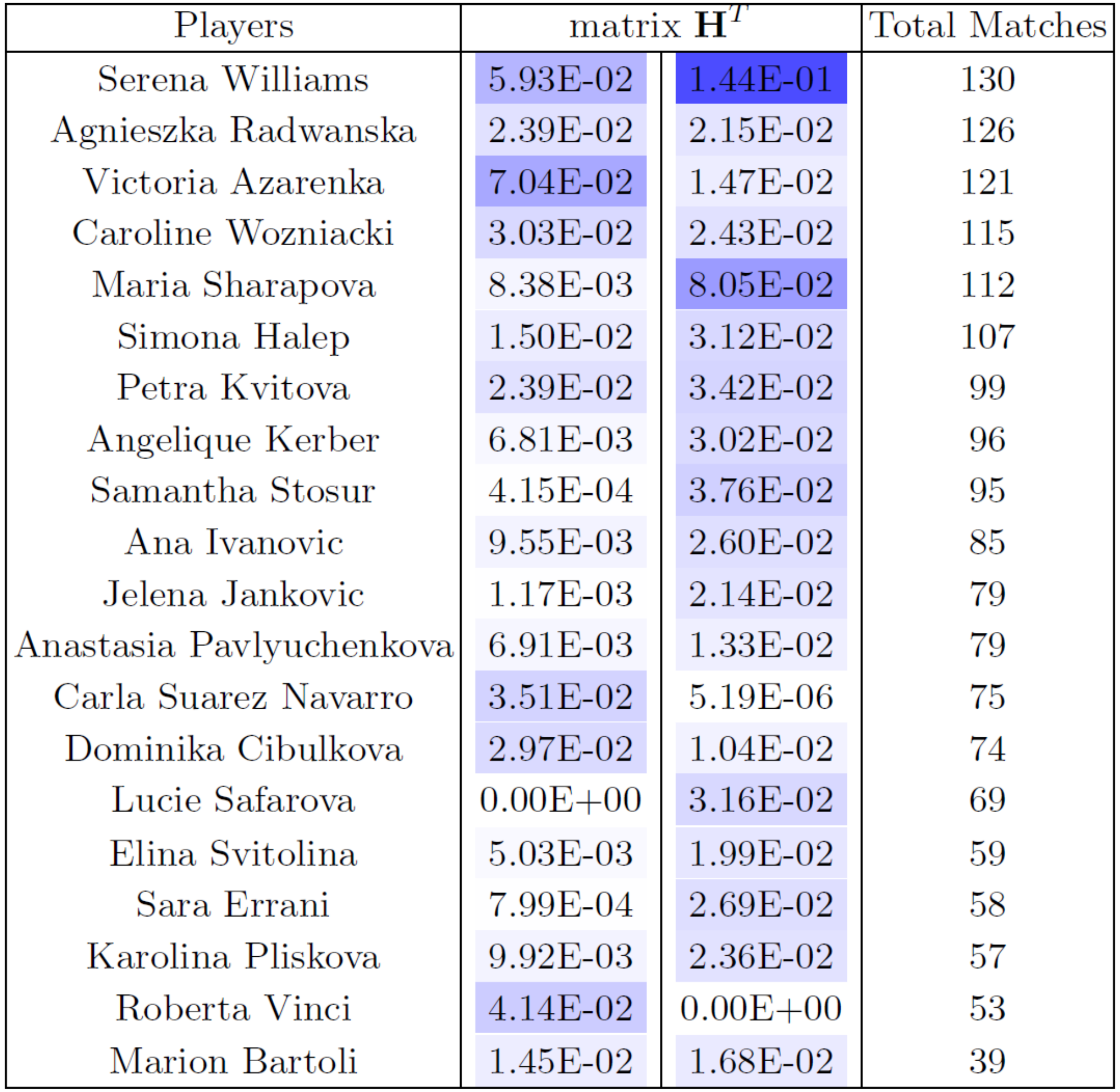
2. Experiments Using Real Life Dataset
    - Dataset Information
    - Running of Algorithm
    - Men Players
    - **Women Players**

3. Comparison with Mixture BTL
    - Qualitative Comparison
    - Prediction Task & Result Comparison

Chronological Order



Tournaments	Row Normalization		Column Normalization	
Australian Open	1.00E-00	3.74E-26	1.28E-01	3.58E-23
Qatar Open	6.05E-01	3.95E-01	1.05E-01	4.94E-02
Dubai Tennis Championships	1.00E-00	1.42E-43	9.47E-02	3.96E-39
Indian Wells Open	5.64E-01	4.36E-01	8.12E-02	4.51E-02
Miami Open	5.86E-01	4.14E-01	7.47E-02	3.79E-02
Madrid Open	5.02E-01	4.98E-01	6.02E-02	4.29E-02
Italian Open	3.61E-01	6.39E-01	5.22E-02	6.63E-02
French Open	1.84E-01	8.16E-01	2.85E-02	9.04E-02
Wimbledon	1.86E-01	8.14E-01	3.93E-02	1.24E-01
Canadian Open	4.59E-01	5.41E-01	5.81E-02	4.92E-02
Cincinnati Open	9.70E-132	1.00E-00	5.20E-123	1.36E-01
US Open	6.12E-01	3.88E-01	8.04E-02	3.66E-02
Pan Pacific Open	1.72E-43	1.00E-00	7.82E-33	1.57E-01
Wuhan Open	1.00E-00	6.87E-67	1.41E-01	1.60E-61
China Open	2.26E-01	7.74E-01	4.67E-02	1.15E-01
WTA Finals	1.17E-01	8.83E-01	9.30E-03	5.03E-02



Players	matrix $\mathbf{H}^T$		Total Matches
Serena Williams	5.93E-02	1.44E-01	130
Agnieszka Radwanska	2.39E-02	2.15E-02	126
Victoria Azarenka	7.04E-02	1.47E-02	121
Caroline Wozniacki	3.03E-02	2.43E-02	115
Maria Sharapova	8.38E-03	8.05E-02	112
Simona Halep	1.50E-02	3.12E-02	107
Petra Kvitova	2.39E-02	3.42E-02	99
Angelique Kerber	6.81E-03	3.02E-02	96
Samantha Stosur	4.15E-04	3.76E-02	95
Ana Ivanovic	9.55E-03	2.60E-02	85
Jelena Jankovic	1.17E-03	2.14E-02	79
Anastasia Pavlyuchenkova	6.91E-03	1.33E-02	79
Carla Suarez Navarro	3.51E-02	5.19E-06	75
Dominika Cibulkova	2.97E-02	1.04E-02	74
Lucie Safarova	0.00E+00	3.16E-02	69
Elina Svitolina	5.03E-03	1.99E-02	59
Sara Errani	7.99E-04	2.69E-02	58
Karolina Pliskova	9.92E-03	2.36E-02	57
Roberta Vinci	4.14E-02	0.00E+00	53
Marion Bartoli	1.45E-02	1.68E-02	39

Learned Transpose **H** Coefficient Matrix for women

1. BTL-NMF Algorithm
    - Majorization-Minimization Algorithm
    - Resolution of Numerical Problems
    - Normalization
    - Convergence Analysis

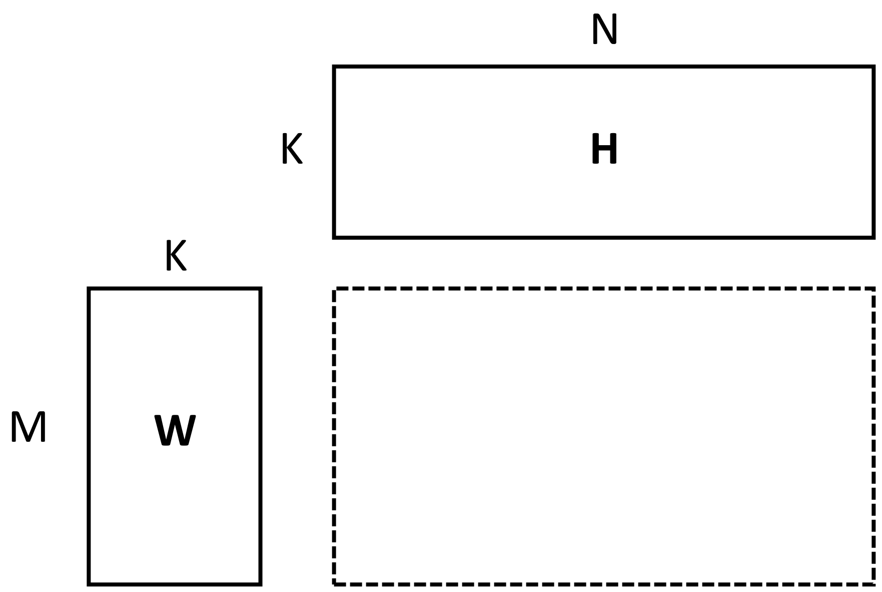
2. Experiments Using Real Life Dataset
    - Dataset Information
    - Running of Algorithm
    - Men Players
    - Women Players

3. Comparison with Mixture BTL
    - **Qualitative Comparison & Algorithm**
    - Prediction Task & Result Comparison

Different number of parameters

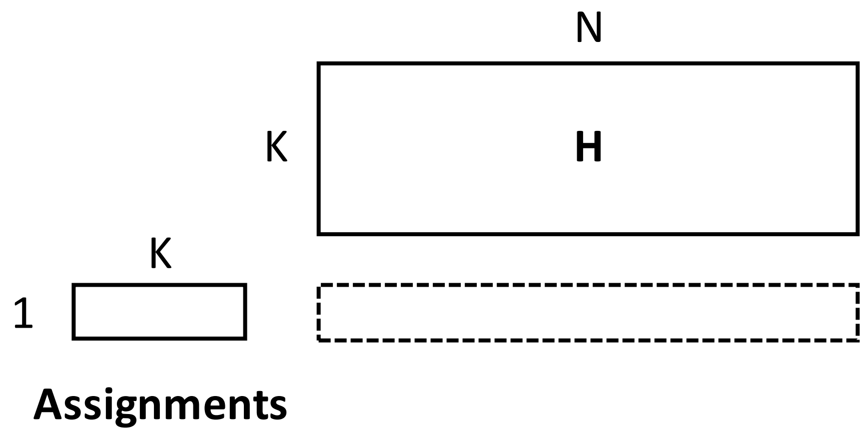
BTL-NMF

$$\Pr(i \text{ beats } j \text{ in tournament } m) = \frac{\sum_k w_{mk} h_{ki}}{\sum_k w_{mk} h_{ki} + \sum_k w_{mk} h_{kj}}$$



Mixture BTL

$$\Pr(i \text{ beats } j \text{ in tournament } m) = \sum_k P(k) \frac{\lambda_{ki}}{\lambda_{ki} + \lambda_{kj}}$$



EM algorithm $\theta = \{\{P(k)\},\{\lambda_{ki}\}\}$

Initialization: $P^{(0)}(K) = 1/K$, randomize $\lambda_{ki}^{(0)}$ nonnegative

$$\ell(\mathcal{D};\theta) = \sum_{(i,j)\in E} b_{ij} \log \left[\sum_k P(k) \left(\frac{\lambda_{ki}}{\lambda_{ki}+\lambda_{kj}} \right) \right]$$

Hard Assignments

$$\ell(\mathcal{D},\mathcal{K};\theta) = \sum_k \sum_{(i,j)\in E} b_{ij}\delta(k|ij) \log \left[P(k) \left(\frac{\lambda_{ki}}{\lambda_{ki}+\lambda_{kj}} \right) \right]$$

E-STEP

Soft posterior Assignments $\boxed{p^{(l)}(k|ij) = P(k|ij,\theta^{(l)})}$

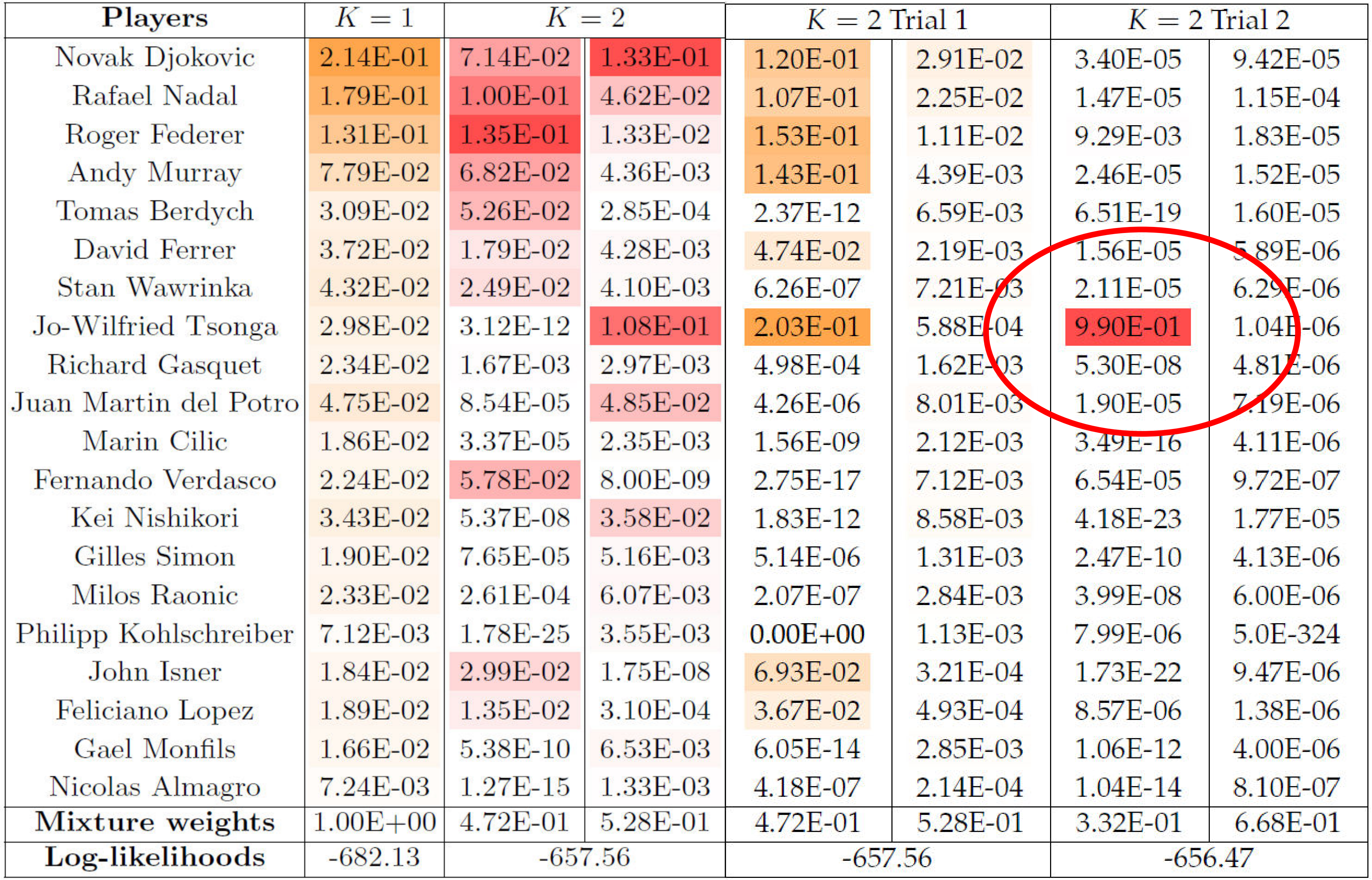
$$\mathbb{E}\{\ell(\mathcal{D},\mathcal{K};\theta)|\mathcal{D},\theta^{(l)}\} = \sum_k \sum_{(i,j)\in E} b_{ij} p^{(l)}(k|ij) \log \left[P(k) \left(\frac{\lambda_{ki}}{\lambda_{ki}+\lambda_{kj}} \right) \right]$$

M-STEP

s.t. $\sum_k P(k) = 1$

$$P^{(l+1)}(k) = \frac{\sum_{(i,j)\in E} b_{ij} p^{(l)}(k|ij)}{\sum_{(i,j)\in E} b_{ij}} \qquad \lambda_{ki}^{(l+1)} = \frac{\sum_{j\neq i} \sum_k p^{(l)}(k|ij) b_{ij}}{\sum_{j\neq i} \left(\sum_k p^{(l)}(k|ij) b_{ij} + \sum_k p^{(l)}(k|ji) b_{ji} \right) \frac{1}{\lambda_{ki}^{(l)}+\lambda_{kj}^{(l)}}}$$

Mixture BTL solution → Unstable



Players	$K=1$	$K=2$		$K=2$ Trial 1		$K=2$ Trial 2	
Novak Djokovic	2.14E-01	7.14E-02	1.33E-01	1.20E-01	2.91E-02	3.40E-05	9.42E-05
Rafael Nadal	1.79E-01	1.00E-01	4.62E-02	1.07E-01	2.25E-02	1.47E-05	1.15E-04
Roger Federer	1.31E-01	1.35E-01	1.33E-02	1.53E-01	1.11E-02	9.29E-03	1.83E-05
Andy Murray	7.79E-02	6.82E-02	4.36E-03	1.43E-01	4.39E-03	2.46E-05	1.52E-05
Tomas Berdych	3.09E-02	5.26E-02	2.85E-04	2.37E-12	6.59E-03	6.51E-19	1.60E-05
David Ferrer	3.72E-02	1.79E-02	4.28E-03	4.74E-02	2.19E-03	1.56E-05	5.89E-06
Stan Wawrinka	4.32E-02	2.49E-02	4.10E-03	6.26E-07	7.21E-03	2.11E-05	6.29E-06
Jo-Wilfried Tsonga	2.98E-02	3.12E-12	1.08E-01	2.03E-01	5.88E-04	9.90E-01	1.04E-06
Richard Gasquet	2.34E-02	1.67E-03	2.97E-03	4.98E-04	1.62E-03	5.30E-08	4.81E-06
Juan Martin del Potro	4.75E-02	8.54E-05	4.85E-02	4.26E-06	8.01E-03	1.90E-05	7.19E-06
Marin Cilic	1.86E-02	3.37E-05	2.35E-03	1.56E-09	2.12E-03	3.49E-16	4.11E-06
Fernando Verdasco	2.24E-02	5.78E-02	8.00E-09	2.75E-17	7.12E-03	6.54E-05	9.72E-07
Kei Nishikori	3.43E-02	5.37E-08	3.58E-02	1.83E-12	8.58E-03	4.18E-23	1.77E-05
Gilles Simon	1.90E-02	7.65E-05	5.16E-03	5.14E-06	1.31E-03	2.47E-10	4.13E-06
Milos Raonic	2.33E-02	2.61E-04	6.07E-03	2.07E-07	2.84E-03	3.99E-08	6.00E-06
Philipp Kohlschreiber	7.12E-03	1.78E-25	3.55E-03	0.00E+00	1.13E-03	7.99E-06	5.0E-324
John Isner	1.84E-02	2.99E-02	1.75E-08	6.93E-02	3.21E-04	1.73E-22	9.47E-06
Feliciano Lopez	1.89E-02	1.35E-02	3.10E-04	3.67E-02	4.93E-04	8.57E-06	1.38E-06
Gael Monfils	1.66E-02	5.38E-10	6.53E-03	6.05E-14	2.85E-03	1.06E-12	4.00E-06
Nicolas Almagro	7.24E-03	1.27E-15	1.33E-03	4.18E-07	2.14E-04	1.04E-14	8.10E-07
Mixture weights	1.00E+00	4.72E-01	5.28E-01	4.72E-01	5.28E-01	3.32E-01	6.68E-01
Log-likelihoods	-682.13	-657.56		-657.56		-656.47	

1. BTL-NMF Algorithm
    - Majorization-Minimization Algorithm
    - Resolution of Numerical Problems
    - Normalization
    - Convergence Analysis

2. Experiments Using Real Life Dataset
    - Dataset Information
    - Running of Algorithm
    - Men Players
    - Women Players

3. Comparison with Mixture BTL
    - Qualitative Comparison & Algorithm
    - **Prediction Task & Result Comparison**

S = subset of the collected dataset D

$|S| \approx 10\% * |\text{non-zero } D|$

Training on $D' = D \setminus S$ → Learned (WH) matrix

↓

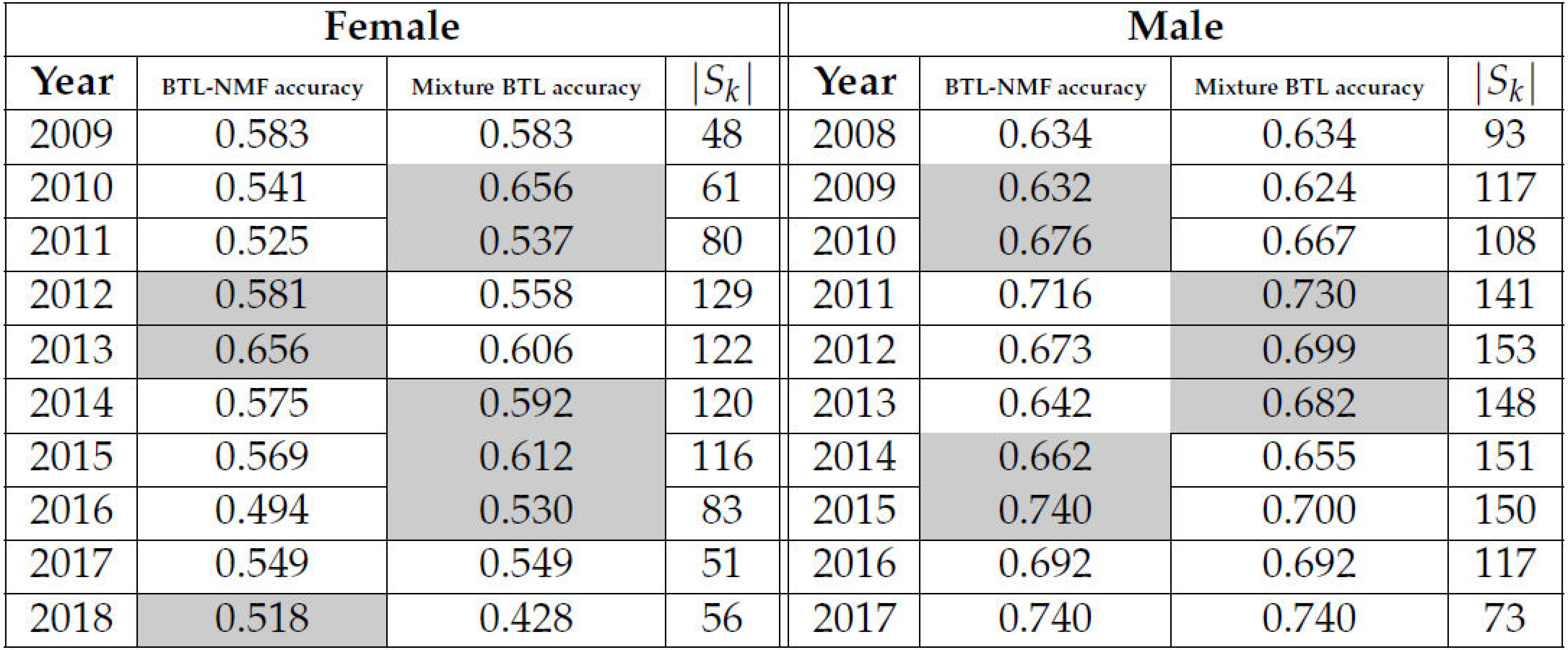
In S, If Pr(i beats j in tournament m) > 0.5

↓

i is predicted to beat j in that match

$$Accuracy = \frac{\#\, correct\ predictions}{|S|}$$

$S_k =$ all the matches played in year k



Female				Male			
Year	BTL-NMF accuracy	Mixture BTL accuracy	$\|S_k\|$	**Year**	BTL-NMF accuracy	Mixture BTL accuracy	$\|S_k\|$
2009	0.583	0.583	48	2008	0.634	0.634	93
2010	0.541	0.656	61	2009	0.632	0.624	117
2011	0.525	0.537	80	2010	0.676	0.667	108
2012	0.581	0.558	129	2011	0.716	0.730	141
2013	0.656	0.606	122	2012	0.673	0.699	153
2014	0.575	0.592	120	2013	0.642	0.682	148
2015	0.569	0.612	116	2014	0.662	0.655	151
2016	0.494	0.530	83	2015	0.740	0.700	150
2017	0.549	0.549	51	2016	0.692	0.692	117
2018	0.518	0.428	56	2017	0.740	0.740	73

Future Plans

1. Larger and longer dataset: Greatest Players of all time
2. Epsilon : Bayesian Interpretation?

Thank you!

For future details, please see the full paper at
https://arxiv.org/abs/1903.06500

Datasets and Code:
https://github.com/XiaRui1996/btl-nmf

Will be presented at ECML/PKDD next week