A. Supplement to “A Listwise Approach to Collaborative Ranking”

A.1. Proofs in Theory Section

Proof of Theorem 1. Notice that \( \Pi_{i1} \) is the argument, \( k \), that minimizes \( Y_{ik} \), and \( \mathbb{P}\{\Pi_{i1} = k\} = \mathbb{P}\{Y_{ik} \leq \min\{Y_{ij}\}_{j \neq k}\} \). Furthermore, \( \min\{Y_{ij}\}_{j \neq k} \) is exponential with rate parameter \( \sum_{j \neq k} \phi(X_{ij}) \) and is independent of \( Y_{ik} \). Hence,

\[
\mathbb{P}\{Y_{ik} \leq \min\{Y_{ij}\}_{j \neq k}\} = \int_0^\infty \phi(X_{ik}) e^{-u\phi(X_{ik})} e^{-\sum_{j \neq k} u\phi(X_{ij})} du = \frac{\phi(X_{ik})}{\sum_j \phi(X_{ij})}.
\]

Furthermore,

\[
\mathbb{P}\{\Pi_{i1} = \ldots = \Pi_{im}\} = \mathbb{P}\{Y_{i1} \geq \ldots \geq Y_{im}\} = \mathbb{P}\{Y_{i1} < Y_{i2} < \ldots < Y_{im}\}.
\]

By the memorylessness property, we have that the joint distribution of \( Y_{i1}, Y_{i2}, \ldots, Y_{im} \) is equivalent to the distribution of \( Y_{i1}, \ldots, Y_{i1}, Z \), where \( Z \) is independent of \( Y_{i1} \). Hence, we can apply induction with the previous argument, and the tower property of conditional probability.

Proof of Lemma 1. By optimality,

\[
\frac{1}{n} \sum_{i=1}^n - \log P_{X_i^*}(\Pi_i) \leq \frac{1}{n} \sum_{i=1}^n - \log P_{X_{i1}^*}(\Pi_i) = \sum_{i=1}^n - \log \frac{P_{X_i^*}(\Pi_i)}{P_{X_{i1}^*}(\Pi_i)} \geq 0.
\]

Thus, we can subtract the expectation,

\[
\frac{1}{n} \sum_{i=1}^n \mathbb{E} \log \frac{P_{X_i^*}(\Pi_i)}{P_{X_{i1}^*}(\Pi_i)} \leq - \frac{1}{n} \sum_{i=1}^n \left( \log \frac{P_{X_i^*}(\Pi_i)}{P_{X_{i1}^*}(\Pi_i)} - \mathbb{E} \log \frac{P_{X_i^*}(\Pi_i)}{P_{X_{i1}^*}(\Pi_i)} \right)
\]

where the expectation \( \mathbb{E} \) is with respect to the draw of \( \Pi_i \) conditional on \( X \).

Lemma 2. Let \( \pi \) be a permutation vector and \( x \) be a score vector each of length \( m \). Suppose that \( |\log \phi(x_j)| \leq C \) for all \( j = 1, \ldots, m \). Define the relative loss function,

\[
L_{x,x'}(\pi) := \log \frac{P_x(\pi)}{P_{x'}(\pi)}.
\]

Consider translating an item in position \( \ell \) to position \( \ell' \) in the permutation \( \pi \), thus forming \( \pi' \) where \( \pi'_{\ell} = \pi_{\ell'} \). Specifically, \( \pi'_{k} = \pi_k \) if \( k < \min\{\ell, \ell'\} \) or \( k > \max\{\ell, \ell'\} \); if \( \ell < \ell' \) then \( \pi'_{k} = \pi_{k+1} \) for \( k = \ell, \ldots, \ell' - 1 \); if \( \ell' < \ell \) then \( \pi'_{k} = \pi_{k-1} \) for \( k = \ell' + 1, \ldots, \ell \). The relative loss function has bounded differences in the sense that

\[
|L_{x,x'}(\pi) - L_{x,x'}(\pi')| \leq C_0 \| \log (\phi(x)) - \log (\phi(x')) \|_{\infty},
\]

where \( \phi(x) \) is applied elementwise and \( C_0 = 2 + e^{2C} \ln(m+1) \).

Proof. Suppose that \( \ell < \ell' \), and define the following shorthand,

\[
\lambda_j = \phi(x_{\pi_j}), \quad \Lambda_j = \sum_{k=j}^m \lambda_k,
\]

and let \( \lambda'_j, \Lambda'_j \) be similarly defined with \( x' \). Then by replacing the permutation \( \pi \) with \( \pi' \) causes the \( \Lambda_j \) to be replaced with \( \Lambda_j - \lambda_j + \lambda_{\ell} \) for \( j = \ell + 1, \ldots, \ell' \). Hence,

\[
\log Q_x(\pi) - \log Q_{x'}(\pi') = \sum_{j=\ell+1}^{\ell'} \log (\lambda_j - \lambda_{j-1} + \lambda_{\ell}) - \log (\lambda_j) = \sum_{j=\ell+1}^{\ell'} \log \left( 1 + \frac{\lambda_j}{\Lambda_j} \right) + \log \Lambda_{\ell} - \log \Lambda_{\ell'}.
\]

So we can bound the difference,

\[
|L_{x,x'}(\pi) - L_{x,x'}(\pi')| \leq \left| \log \frac{\Lambda_{\ell}}{\Lambda_{\ell'}} - \log \frac{\lambda_j}{\lambda_{j'}} \right| + \sum_{j=\ell}^{\ell'-1} \left| \log \left( 1 + \frac{\lambda_j}{\Lambda_j} \right) - \log \left( 1 + \frac{\lambda_j}{\Lambda_j} \right) \right|.
\]

Suppose that for each \( j \), \( |\log \lambda_j - \log \lambda'_{j'}| \leq \delta \) and that \( |\log \Lambda_j| \leq C \). Then we have that

\[
\left| \log \frac{\Lambda_{\ell}}{\Lambda_{\ell'}} \right| \leq \max_{j \geq x} \left| \log \frac{\lambda_j}{\lambda_{j'}} \right| \leq \delta.
\]

The same equation can be made for this term with \( \ell' \). Let

\[
\alpha_j = \max \left\{ \frac{\lambda_j}{\Lambda_j}, \frac{\lambda_{j'}}{\Lambda_j} \right\}.
\]

Then we have

\[
\left| \log \left( 1 + \frac{\lambda_j}{\Lambda_j} \right) - \log \left( 1 + \frac{\lambda_{j'}}{\Lambda_j} \right) \right| \leq |\log(1+\alpha_j) - \log(1+e^{-\delta}\alpha_j)| \leq |1-e^{-\delta}| |\alpha_j|.
\]
Furthermore, because \( \Lambda_j \geq (m - j + 1)e^{-c} \) then \( |\alpha_j| \leq (m - j + 1)^{-1}e^{2c} \) and
\[
|L_{x,x'}(\pi) - L_{x,x'}(\pi')| \leq 2\delta \\
+ \sum_{j=\ell}^{\ell-1} |1 - e^{-\delta}| \frac{e^{2c}}{m-j+1} \\
\leq 2\delta + |1 - e^{-\delta}|e^{2c}H_m \\
\leq \delta(2 + e^{2c}\ln(m+1)).
\]
In the above equation, \( H_m \) is the \( m \)th harmonic number.
A similar calculation can be made when \( \ell > \ell \). Setting \( \delta = ||\log \phi(x) - \log \phi(x')||_\infty \) concludes the proof. \( \square \)

**Proof of Theorem 2.** Define the empirical process function to be
\[
\rho_n(x) := \frac{1}{n} \sum_{i=1}^{n} \left( \log \frac{P_{X_i}(1)}{P_{X_i}(1)} - \log \frac{P_{X_i}(1)}{P_{X_i}(1)} \right).
\]
By the listwise representation theorem, \( \rho_n(x) \) is a function of \( n \times m \) independent exponential random variables. Moreover, if we were to change the value of a single element \( y_{ik} \) then this would result in a change of permutation of the type described in Lemma 2. Notice that the bound on the change in the relative loss is \( C_0||\log \phi(X_i) - \log \phi(X_i')||_\infty \), where \( C_0 = 2 + e^{2c}\ln(m+1) \), and notice that the sum of squares of these bounds are
\[
\sum_{i,k} C_0^2||\log \phi(X_i) - \log \phi(X_i')||_\infty^2 \\
= mC_0^2 \sum_{i=1}^{n} ||\log \phi(X_i) - \log \phi(X_i')||_\infty^2 \\
= mC_0^2 ||Z - Z'||_{\infty,2}^2,
\]
where \( Z, Z' \) are \( \log \phi \) applied elementwise to \( X, X' \) respectively. By Lemma 2 and McDiarmid’s inequality,
\[
\mathbb{P}(\rho_n(x) - \rho_n(x') > \epsilon) \leq \exp \left( -\frac{2\epsilon^2}{mC_0^2 ||Z - Z'||_{\infty,2}^2} \right).
\]
Hence, the stochastic process \( \{\rho_n(X) : X \in \mathcal{X}\} \) is a subGaussian field with canonical distance,
\[
d(X, X') := \sqrt{mC_0} ||Z - Z'||_{\infty,2}.
\]
The result follows by Dudley’s chaining (Talagrand, 2006). \( \square \)

**Lemma 3.** If \( \log \phi \) is \( 1 \)-Lipschitz then we have that \( g(Z) \leq g(\mathcal{X}) \).

**Proof.** Let \( X, X' \in \mathcal{X} \), and \( Z = \log \phi(X), Z' = \log \phi(X') \). Then
\[
|z_{ij} - z'_{ij}| \leq |x_{ij} - x'_{ij}|
\]
by the Lipschitz property. Hence, \( ||Z - Z'||_{\infty,2} \leq ||X - X'||_{\infty,2} \), and so
\[
\mathcal{N}(u, Z, ||.,||_{\infty,2}) \leq \mathcal{N}(u, X, ||.,||_{\infty,2}).
\]
\( \square \)

**Proof of Corollary 1.** Consider two matrices \( X, X' \) in the model (7) then
\[
||x_i - x'_i||_{\infty} = \max_j |\beta z_{ij} - \beta' z_{ij}| \leq ||\beta - \beta'||_2 ||Z_i||_{2,\infty}.
\]
Let \( \zeta = \max_i ||Z_i||_{2,\infty} \) then
\[
||X - X'||_{\infty,2} \leq \zeta ||\beta - \beta'||_2.
\]
The covering number of \( \mathcal{X} \) is therefore bounded by
\[
\mathcal{N}(\mathcal{X}, u, ||.,||_{\infty,2}) \leq \mathcal{N}(B_{c_0, u/\zeta}, ||.,||_2) \leq \left( \frac{CC_0c_k}{u} \right)^s,
\]
for an absolute constant \( C_0 \), where \( B_{c_0} \) is the \( \ell_2 \) ball of radius \( c_0 \). The result follows by Lemma 3 and Theorem 2. \( \square \)

**Proof of Corollary 2.** Let \( X = UV^T \) and \( X' = U'V'^T \) such that \( U, V, U', V' \) bounded by 1 in Frobenius norm (modifying the Frobenius norm bound does not change the substance of the proof). Consider
\[
|u_i v_j - u'_i v'_j| \leq |u_i v_j - u'_i v_j| + |u'_i v_j - u'_i v'_j| \\
\leq ||u_i - u'_i||_2 ||v_j|| + ||u'_i||_2 ||v_j - v'_j||.
\]
Maximizing this over the selection of \( j \),
\[
\max_j |u_i v_j - u'_i v'_j| \\
\leq ||u_i - u'_i||_2 ||V||_{2,\infty} + ||u'_i||_2 ||V' - V'||_{2,\infty}.
\]
Hence,
\[
||X - X'||_{\infty,2} \leq ||U - U'||_F ||V||_{2,\infty} + ||U'||_F ||V' - V'||_{2,\infty} \\
\leq ||U - U'||_F + ||V - V'||_{2,\infty}.
\]
Consider the vectorization mapping from the \( m \times r \) matrix to the \( mr \) dimensional vectors. The Frobenius norm is mapped to the \( \ell_2 \) norm, and we can consider the \( 2, \infty \) norm to be, the norm \( ||x||_{\rho} = \max_j \| (x_{j+1}, \ldots, x_{j+r+1}) \|_2 \). The \( \rho \)-norm unit ball \( B_{p, \rho} \) is just the Cartesian product of the \( \ell_2 \) norm ball in \( K \) dimensions. The volume of a \( d \)-dimensional ball, \( V_d \), is bounded by
\[
C_d \leq \frac{V_d}{(\pi d/2)^{d/2}} \leq C_u,
\]
where $C_l < C_u$ are universal constants. So the volume ratio between the $\ell_2$ norm ball and the $\rho$ norm ball is bounded by
\[
\frac{V(B_2)}{V(B_\rho)} \leq C \left( \frac{\rho^{r/2}}{(e\pi)^{r/2}} \right)^m / \left( \frac{(rm)^{rm/2}}{(e\pi)^{rm/2}} \right)
\leq Cm^{-rm/2},
\]
where $C = C_u/C_l$.

\[
\mathcal{N}(\epsilon, B_2, \|\cdot\|_\rho) \leq C \left( \frac{2}{\epsilon} + 1 \right)^{rm} m^{-rm/2}
\leq C \left( \frac{3}{\epsilon \sqrt{m}} \right)^{rm},
\]
for $\epsilon \leq 1$. This is also the covering number of the Frobenius norm ball in the $2, \infty$ norm. Moreover, we know that the covering number of the unit Frobenius norm ball in $n \times K$ matrices ($B_F$) in the Frobenius norm is
\[
\mathcal{N}(\epsilon, B_F, \|\cdot\|_F) \leq C \left( \frac{\epsilon}{\epsilon} \right)^{nr},
\]
for some constant $c$. Consider covering the space $X$, by selecting centers $U, V$ from the $\epsilon/2$-coverings of $B_F$ in the $F$-norm and $2, \infty$ norm respectively. By the above norm bound, this produces an $\epsilon$-covering in the $\infty, 2$ norm. Dudley’s entropy bound is thus
\[
\int_0^\infty \sqrt{\log \mathcal{N}(\epsilon, B_F, \|\cdot\|_F) + \log \mathcal{N}(\epsilon, B_F, \|\cdot\|_{2, \infty})} \, d\epsilon
\leq \int_0^c \sqrt{nr \log(c/\epsilon)} \, d\epsilon + \int_c^{3/\sqrt{m}} \sqrt{-mr \log(\epsilon/\sqrt{m})} \, d\epsilon.
\]
So that
\[
\int_0^c \sqrt{nr \log(c/\epsilon)} \, d\epsilon
\leq c' \sqrt{m}
\]
for some absolute constant $c'$ and
\[
\int_0^{3/\sqrt{m}} \sqrt{-mr \log(\epsilon/\sqrt{m})} \, d\epsilon \leq \int_0^3 \sqrt{r \log(u/3)} \, du
\leq c' \sqrt{r}.
\]
Hence, for $g(X) \leq c' \sqrt{m}$ and we have the result.

A.2. Algorithms

**Algorithm 3** Compute gradient for $V$ when $U$ fixed

**Input:** $\Pi, U, V, \lambda, \rho$

**Output:** $g \{ g \in \mathbb{R}^{r \times m} is the gradient for f(V) \}$

$g = \lambda \cdot V$

for $i = 1$ to $n$

Precompute $h_t = u^T v_{||i}$ for $1 \leq t \leq \tilde{m}$ (For implicit feedback, it should be $(1 + \rho) \cdot \tilde{m}$ instead of $\tilde{m}$, since $\rho \cdot \tilde{m}$ 0’s are appended to the back)

Initialize $total = 0, tt = 0$

for $t = \tilde{m}$ to 1 do

$\text{total} += \exp(h_t)$

$tt += 1/\text{total}$

end for

Initialize $c[t] = 0$ for $1 \leq t \leq \tilde{m}$

for $t = \tilde{m}$ to 1 do

$c[t] += h_t \cdot (1 - h_t)$

$c[t] += \exp(h_t) \cdot h_t \cdot (1 - h_t) \cdot tt$

$\text{total} += \exp(h_t)$

$tt -= 1/\text{total}$

end for

for $t = 1$ to $\tilde{m}$ do

$g[:; \Pi_u] += c[t] \cdot u_i$

end for

end for

Return $g$

**Algorithm 4** Gradient update for $V$ (Same procedure for updating $U$)

**Input:** $V, ss, rate \{ rate \text{ refers to the decaying rate of the step size } ss \}$

**Output:** $V$

Compute gradient $g$ for $V$ (see alg 3)

$V -= ss \cdot g$

$ss *= rate$

Return $V$
**Figure 4.** Effectiveness of Stochastic Queuing Process.

**Figure 5.** Effectiveness of using full lists.