A. Proof

A.1. Proof of Lemma 2

It is relatively hard to represent all $\boldsymbol{x}_j - \boldsymbol{x}_k$, so we define $\tilde{X}$ to be an $d \times c$ matrix, with $c = C_2^n$, and we use q to denote each possible (j,k) pairs (the q-th column of $\tilde{X}$ is $\boldsymbol{x}_j - \boldsymbol{x}_k$). For each observed comparison $\alpha = 1, \ldots, m$, we use $(i_\alpha, j_\alpha, k_\alpha)$ to denote the (task, item j, item k) tuple, $q_\alpha := (j_\alpha, k_\alpha)$ to denote the encoding of item pairs, and y_α is the observed +1/-1 outcome. The problem can then be rewritten as

$$\min_{W \in \mathbb{R}^{T\times d}} \sum_{\alpha=1}^{m} \ell((W\tilde{X})_{i_\alpha, q_\alpha}, y_\alpha), \text{ such that } \|W\|_* \leq \mathcal{W} \tag{11}$$

First we rewrite $\mathfrak{R}(F_W)$ as follows:

$$\begin{aligned}\mathfrak{R}(F_\Theta) &= \mathbb{E}_\sigma[\sup_{f\in F_W} \frac{1}{m}\sum_{\alpha=1}^{m} \sigma_\alpha \ell(f(i_\alpha, j_\alpha, k_\alpha), y_\alpha)] \\ &= \mathbb{E}_\sigma[\sup_{W:\|W\|_*\leq \mathcal{W}} \frac{1}{m}\sum_{(i,q)} \Gamma_{iq}\ell((W\tilde{X})_{i,q}, y_{i,q})]\end{aligned}$$

Where $\Gamma \in \mathbb{R}^{T\times c}$ with each entry $\Gamma_{iq} = \sum_{\alpha: i_\alpha = i, q_\alpha = q} \sigma_\alpha$. Now, using the same trick in (Shamir & Shalev-Shwartz, 2014), we can divide Γ based on the "hit-time" on entry (i,q) of Ω, with some threshold $p > 0$ (we will discuss the optimal choice of p later). Let $h_{iq} = |\alpha : i_\alpha = i, q_\alpha = q|$, and let $A, B \in \mathbb{R}^{T\times c}$ be defined as:

$$A_{iq} = \begin{cases} \Gamma_{iq}, & \text{if } h_{iq} > p \\ 0, & \text{otherwise.}\end{cases} \qquad B_{iq} = \begin{cases} 0, & \text{if } h_{iq} > p \\ \Gamma_{iq}, & \text{ohterwise.}\end{cases}$$

Since $\Gamma = A + B$, we can rewrite $\mathfrak{R}(F_W)$ as:

$$\begin{aligned}\mathfrak{R}(F_W) &= \mathbb{E}_\sigma[\sup_{W:\|W\|_*\leq \mathcal{W}} \frac{1}{m}\sum_{(i,q)} A_{iq}\ell((W\tilde{X})_{i,q}, y_{i,q})] \\ &+ \mathbb{E}_\sigma[\sup_{W:\|W\|_*\leq \mathcal{W}} \frac{1}{m}\sum_{(i,q)} B_{iq}\ell((W\tilde{X})_{i,q}, y_{i,q})]\end{aligned} \tag{12}$$

By Lemma 10 in (Shamir & Shalev-Shwartz, 2014), the first term of (12) can be upper bounded by:

$$\frac{\mathcal{B}}{m}\mathbb{E}_\sigma[\sum_{(i,j)} |A_{ij}|] \leq \frac{\mathcal{B}}{\sqrt{p}}$$

Also, we can bound the second term in (12) by:

$$\begin{aligned}&\mathbb{E}_\sigma[\sup_{W:\|W\|_*\leq \mathcal{W}} \frac{1}{m}\sum_{(i,q)} B_{iq}\ell((W\tilde{X})_{iq}, y_{i,q})] \\ &\leq \frac{L_\ell}{m}\mathbb{E}_\sigma[\sup_{W:||W||_*\leq \mathcal{W}} \sum_{(i,j)} B_{ij}(W\tilde{X})_{iq}] \\ &\leq \frac{L_\ell}{m}\mathbb{E}_\sigma[\sup_{W:||W||_*\leq \mathcal{W}} \|B\|_2\|W\tilde{X}\|_*] \\ &= \frac{L_\ell}{m}||W\tilde{X}||_*\mathbb{E}_\sigma[||B||_2] \\ &\leq \frac{2L_\ell}{m}\mathcal{W}\mathbb{E}_\sigma[\|B\|_2] \\ &\leq 4.4\frac{L_\ell}{m}C\mathcal{W}\sqrt{p}(\sqrt{T} + \sqrt{c}).\end{aligned} \tag{13}$$

Note that the last inequality is using Lemma 11 in (Shamir & Shalev-Shwartz, 2014), where C is a universal constant used in their paper. Therefore, with p chosen to be $m\mathcal{B}/(4.4L_\ell C\mathcal{W}(\sqrt{T}+\sqrt{c}))$, we can get $\mathfrak{R}(F_W)$ bounded by:

$$\mathbb{E}_\Omega[\mathfrak{R}(F_W)] \leq \sqrt{\frac{9L_\ell\mathcal{B}C\mathcal{W}(\sqrt{T}+\sqrt{c})}{m}}. \tag{14}$$

Also, we can bound $\mathbb{E}_\Omega[\mathfrak{R}(F_W)]$ using another way:

$$\begin{aligned}\mathfrak{R}(F_W) &\leq \frac{L_\ell}{m}\mathbb{E}_\sigma[\sup_{W:||W||_*\leq \mathcal{W}} \sum_{\alpha=1}^{m} \sigma_\alpha (W\tilde{X})_{i_\alpha, q_\alpha}] \\ &= L_\ell \mathbb{E}_\sigma[\sup_{W:||W||_*\leq \mathcal{W}} \frac{1}{m}\sum_{\alpha=1}^{m} \sigma_\alpha \text{trace}(W\tilde{X}_{q_\alpha} I_{i_\alpha}^T)] \\ &\leq L_\ell \mathcal{W} \max_{q,i} \|\tilde{X}_q I_i^T\|_2 \sqrt{\frac{\log 2d}{m}} \\ &\leq 2L_\ell\mathcal{W}\sqrt{\frac{\log 2d}{m}}\end{aligned} \tag{15}$$

where we use the assumption that $\|\boldsymbol{x}_j\| \leq 1$ for all j so $\|\boldsymbol{x}_j - \boldsymbol{x}_k\| \leq 2$. Therefore, Rademacher complexity can be upper bounded by:

$$\mathbb{E}_\Omega[\mathfrak{R}_{F_\Theta}] \leq \min\{2L_\ell\mathcal{W}\sqrt{\frac{\log 2d}{m}}, \sqrt{\frac{9L_\ell\mathcal{B}\mathcal{W}(\sqrt{T}+\sqrt{c})}{m}}\}, \tag{16}$$

which then implies our theorem statement since $c \leq n^2$.

A.2. Proof of Theorem 1

Combining Lemma 2 and Lemma 1, we get

$$\begin{aligned}R_\ell(f) \leq& \hat{R}_\ell(f) + \mathcal{B}\sqrt{\frac{\log\frac{1}{\delta}}{2m}} \\ &+ \min\left\{4L_\ell\mathcal{W}\sqrt{\frac{\log 2d}{m}}, \sqrt{\frac{36L_\ell\mathcal{W}C\mathcal{B}(\sqrt{T}+\sqrt{c})}{m}}\right\}.\end{aligned}$$

Therefore

$$R_\ell(f) - R_\ell^* \le (\hat{R}_\ell(f) - R_\ell^*) + \mathcal{B}\sqrt{\frac{\log\frac{1}{\delta}}{2m}}$$
$$+ \min\left\{4L_\ell\mathcal{W}\sqrt{\frac{\log 2d}{m}}, \sqrt{\frac{36L_\ell\mathcal{W}C\mathcal{B}(\sqrt{T}+\sqrt{c})}{m}}\right\}.$$

Next we use the following lemma to bound the expected ranking loss.

Lemma 5 (Consistency of Excess Risk (Bartlett et al., 2006))**.** *Let ℓ be a convex surrogate loss function. Then there exists a strictly increasing function Ψ, $\Psi(0) = 0$, such that for all measurable f:*

$$R(f) - R^* \le \Psi(R_\ell(f) - R_\ell^*),$$

where $R^* = \inf_f R(f)$ and $R_\ell^* = \inf_f R_\ell(f)$.

Using Lemma 5, and assume $\Psi(\cdot)$ is L_Ψ is the Lipchitz constant of Ψ in the domain $\{f : R_\ell(f^*) - R_\ell^*\}$ (which is always bounded in this case since we only consider f_W with $\|W\|_* \le \mathcal{W}$), we get

$$R(f^*) - R^* \le L_\Psi(R_\ell(f^*) - R_\ell^*)$$
$$\le L_\Psi\left((\hat{R}_\ell(f) - R_\ell^*) + \mathcal{B}\sqrt{\frac{\log\frac{1}{\delta}}{2m}}\right.$$
$$\left.+ \min\left\{4L_\ell\mathcal{W}\mathcal{X}\sqrt{\frac{\log 2d}{m}}, \sqrt{\frac{36L_\ell\mathcal{W}C\mathcal{B}(\sqrt{T}+\sqrt{c})}{m}}\right\}\right)$$

Which can be simplified to

$$R(f^*) - R^* = O(\hat{R}_\ell(f^*) - R_\ell^*) + O(\mathcal{B}\sqrt{\frac{\log(1/\delta)}{m}})$$
$$+ O(\frac{\mathcal{W}\log d + \sqrt{\mathcal{W}\mathcal{B}(\sqrt{T}+\sqrt{c})}}{\sqrt{m}}).$$

A.3. Proof of Lemma 3

Under the conditions listed in the lemma, taking $\hat{W} = W^*$ will result in zero empirical error, so

$$\hat{R}_\ell(f_{\hat{W}}) \le \hat{R}_\ell(f_{W^*}) = 0.$$

Furthermore, under these conditions $R_\ell^* = 0$. Combining with Theorem 1 proves this lemma.

A.4. Proof of Lemma 4

Since each task $i = 1, \ldots, T$ is just a standard rankSVM problem with L2 regularization, we can follow the standard derivation for Rademacher complexity. From (Kakade et al., 2009)), the complexity for using L2 regularization is

$$E_\Omega[\Re(F_W)] \le wL_\ell\sqrt{\frac{1}{m/T}} = \sqrt{T}wL_\ell\sqrt{\frac{1}{m}}. \quad (17)$$

Assume there are m/T pairs, and then use Lemma 1 we get the following error bound:

$$R_\ell(f) \le \hat{R}_\ell(f) + 2\mathbb{E}_\Omega[\Re(F_W)] + \mathcal{B}\sqrt{\frac{T\log\frac{1}{\delta}}{2m}} \quad (18)$$

Combine with the Rademacker complexity proved in (17), we get

$$R_\ell(f) \le \hat{R}_\ell(f) + 4L_\ell\sqrt{T}w\mathcal{X}\sqrt{\frac{\log 2d}{m}} + \mathcal{B}\sqrt{\frac{T\log\frac{1}{\delta}}{2m}}.$$

Therefore,

$$\mathrm{R}_\ell(f) - R_\ell^* \le (\hat{R}_\ell(f) - R_\ell^*) + 4L_\ell\sqrt{T}w\mathcal{X}\sqrt{\frac{\log 2d}{m}} + \mathcal{B}\sqrt{\frac{T\log\frac{1}{\delta}}{2m}}.$$

Use Lemma 5 we get

$$\mathrm{R}_\ell(f^*) - R_\ell^* \le L_\Psi\left((\hat{R}_\ell(f) - R_\ell^*) + 4L_\ell\sqrt{T}w\mathcal{X}\sqrt{\frac{\log 2d}{m}} + \mathcal{B}\sqrt{\frac{T\log\frac{1}{\delta}}{2m}}\right).$$

Under the condition provided in the theorem, the first term in the right hand side becomes 0, and then using the big-O notation we can prove this theorem.