Supplementary material for "Horde of Bandits using Gaussian Markov Random Fields"

A Learning the Graph

In the main paper, we assumed that the graph is known, but in practice such a user-user graph may not be available. In such a case, we explore a heuristic to learn the graph on the fly. The computational gains described in the main paper make it possible to simultaneously learn the user-preferences and infer the graph between users in an efficient manner. Our approach for learning the graph is related to methods proposed for multitask and multilabel learning in the batch setting [8, 7] and multitask learning in the online setting [14]. However, prior works that learn the graph in related settings only tackle problem with tens or hundreds of tasks/labels while we learn the graph and preferences across thousands of users.

Let $V_t \in \mathbb{R}^{n \times n}$ be the inverse covariance matrix corresponding to the graph inferred between users at round t. Since zeroes in the inverse covariance matrix correspond to conditional independences between the corresponding nodes (users) [13], we use L1 regularization on V_t for encouraging sparsity in the inferred graph. We use an additional regularization term $\Delta(V_t||V_{t-1})$ to encourage the graph to change smoothly across rounds. This encourages V_t to be close to V_{t-1} according to a distance metric Δ. Following [14], we choose Δ to be the log-determinant Bregman divergence given by $\Delta(X||Y) = \text{Tr}(XY^{-1}) - \log|XY^{-1}| - dn$. If $W_t \in R^{d \times n} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_n]$ corresponds to the matrix of user preference estimates, the combined objective can be written as:

$$[\mathbf{w}_t, V_t] = \underset{\mathbf{w}, V}{\text{argmin}} \, ||\mathbf{r}_t - \Phi_t \mathbf{w}||_2^2 + \text{Tr}\left(V(\lambda W^T W + V_{t-1}^{-1})\right) + \lambda_2 ||V||_1 - (dn+1)\ln|V| \tag{1}$$

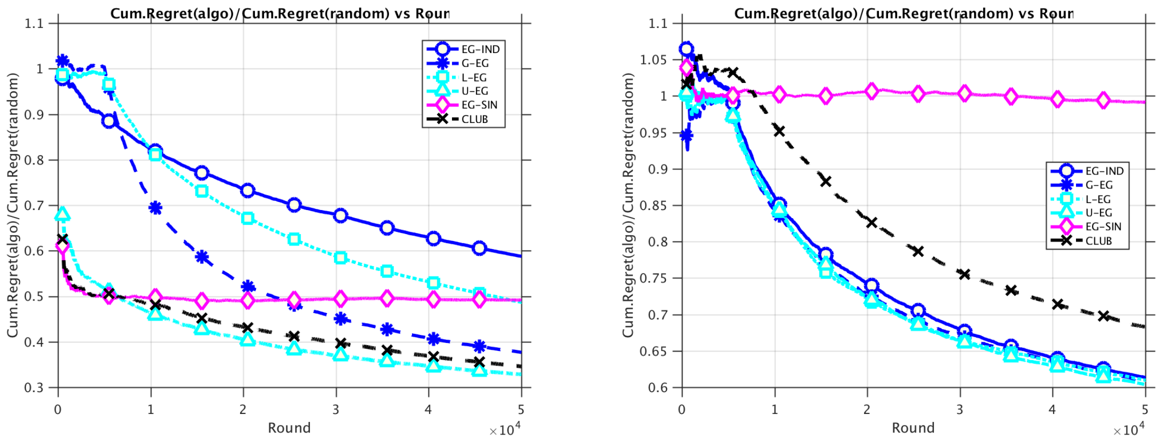
The first term in (1) is the data fitting term. The second term imposes the smoothness constraint across the graph and ensures that the changes in V_t are smooth. The third term ensures that the learnt precision matrix is sparse, whereas the last term penalizes the complexity of the precision matrix. This function is independently convex in both $\mathbf{w}$ and V (but not jointly convex), and we alternate between solving for $\mathbf{w}_t$ and V_t in each round. With a fixed V_t, the $\mathbf{w}$ sub-problem is the same as the MAP estimation in the main paper and can be done efficiently. For a fixed $\mathbf{w}_t$, the V sub-problem is given by

$$V_t = \underset{V}{\text{argmin}} \, \text{Tr}\left((V[\lambda \overline{W}_t^T \overline{W}_t + V_{t-1}^{-1})\right) + \lambda_2 ||V||_1 - (dn+1)\ln|V| \tag{2}$$

Here $\overline{W}_t$ refers to the mean subtracted (for each dimension) matrix of user preferences. This problem can be written as a graphical lasso problem [6], $\min_X \text{Tr}(SX) + \lambda_2 ||X||_1 - \log|X|$, where the empirical covariance matrix S is equal to $\lambda \overline{W}_t^T \overline{W}_t + V_{t-1}^{-1}$. We use the highly-scalable second order methods described in [9, 10] to solve (2). Thus, both sub-problems in the alternating minimization framework at each round can be solved efficiently.

For our preliminary experiments in this direction, we use the most scalable epoch-greedy algorithm for learning the graph on the fly and denote this version as L-EG. We also consider another variant, U-EG in which we start from the Laplacian matrix L corresponding to the given graph and allow it to change by re-estimating the graph according to (2). Since U-EG has the flexibility to infer a better graph than the one given, such a variant is important for cases where the prior is meaningful but somewhat misspecified (the given graph accurately reflects some but not all of the user similarities). Similar to [14], we start off with an empty graph and start learning the graph only after the preference vectors have become stable, which happens in this case after each user has received 10 recommendations. We update the graph every 1K rounds. For both datasets, we allow the learnt graph to contain at most 100K edges and tune λ_2 to achieve a sparsity level equal to 0.05 in both cases.

To avoid clutter, we plot all the variants of the EG algorithm, L-EG and U-EG, and use EG-IND, G-EG, EG-SIN as baselines. We also plot CLUB as a baseline. For the Last.fm dataset (Figure 1(b)(a)), U-EG performs slightly better than G-EG, which already performed well. The regret for L-EG is lower compared to LINUCB-IND



(a) Last.fm

(b) Delicious

Figure 1: Regret Minimization while learning the graph

indicating that learning the graph helps, but is worse as compared to both CLUB and LINUCB-SIN. On the other hand, for Delicious (Figure 1(b)(b)), L-EG and U-EG are the best performing methods. L-EG slightly outperforms EG-IND, underscoring the importance of learning the user-user graph and transferring information between users. It also outperforms G-EG, which implies that it is able to learn a graph which reflects user similarities better than the existing social network between users. For both datasets, U-EG is among the top performing methods, which implies that allowing modifications to a good (in that it reflects user similarities reasonably well) initial graph to model the obtained data might be a good method to overcome prior misspecification. From a scalability point of view, for Delicious the running time for L-EG is 0.1083 seconds/iteration (averaged across T) as compared to 0.04 seconds/iteration for G-EG. This shows that even in the absence of an explicit user-user graph, it is possible to achieve a low regret in an efficient manner.

B Regret bound for Epoch-Greedy

Theorem 1. *Under the additional assumption that $||w_t||_2 \leq 1$ for all rounds t, the expected regret obtained by epoch-greedy in the GOB framework is given as:*

$$R(T) = \tilde{O}\left(n^{1/3} \left(\frac{\text{Tr}(L^{-1})}{\lambda n} \right)^{\frac{1}{3}} T^{\frac{2}{3}} \right) \tag{3}$$

Proof. Let $\mathcal{H}$ be the class of hypotheses of linear functions (one for each user) coupled with Laplacian regularization. Let $\mu(\mathcal{H}, q, s)$ represent the regret or cost of performing s exploitation steps in epoch q. Let the number of exploitation steps in epoch q be s_q.

Lemma 1 (Corollary 3.1 from [11])**.** *If $s_q = \lfloor \frac{1}{\mu(\mathcal{H},q,1)} \rfloor$ and Q_T is the minimum Q such that $Q + \sum_{q=1}^{Q} s_q \geq T$, then the regret obtained by Epoch Greedy is bounded by $R(T) \leq 2Q_T$.*

We now bound the quantity $\mu(\mathcal{H}, q, 1)$. Let $Err(q, \mathcal{H})$ be the generalization error for $\mathcal{H}$ after obtaining q unbiased samples in the exploration rounds. Clearly,

$$\mu(\mathcal{H}, q, s) = s \cdot Err(q, \mathcal{H}). \tag{4}$$

Let ℓ_{LS} be the least squares loss. Let the number of unbiased samples per user be equal to p. The empirical Rademacher complexity for our hypotheses class $\mathcal{H}$ under ℓ_{LS} can be given as $\hat{\mathcal{R}}_p^n(\ell_{LS} \circ \mathcal{H})$. The generalization error for $\mathcal{H}$ can be bounded as follows:

Lemma 2 (Theorem 1 from [12])**.** *With probability* $1-\delta$,

$$Err(q,\mathcal{H}) \leq \hat{\mathcal{R}}_p^n(\ell_{LS}\circ\mathcal{H}) + \sqrt{\frac{9\ln(2/\delta)}{2pn}} \tag{5}$$

Assume that the target user is chosen uniformly at random. This implies that the expected number of samples per user is at least $p = \lfloor\frac{q}{n}\rfloor$. For simplicity, assume q is exactly divisible by n so that $p = \frac{q}{n}$ (this only affects the bound by a constant factor). Substituting p in (5), we obtain

$$Err(q,\mathcal{H}) \leq \hat{\mathcal{R}}_p^n(\ell_{LS}\circ\mathcal{H}) + \sqrt{\frac{9\ln(2/\delta)}{2q}}. \tag{6}$$

The Rademacher complexity can be bounded using Lemma 3 (see below) as follows:

$$\hat{\mathcal{R}}_p^n(\ell_{LS}\circ\mathcal{H}) \leq \frac{1}{\sqrt{p}}\sqrt{\frac{48\,\mathrm{Tr}(L^{-1})}{\lambda n}} = \frac{1}{\sqrt{q}}\sqrt{\frac{48\,\mathrm{Tr}(L^{-1})}{\lambda}} \tag{7}$$

Substituting this into (6) we obtain

$$Err(q,\mathcal{H}) \leq \frac{1}{\sqrt{q}}\Bigg[\sqrt{\frac{48\,\mathrm{Tr}(L^{-1})}{\lambda}} + \sqrt{\frac{9\ln(2/\delta)}{2}}\Bigg]. \tag{8}$$

We set $s_q = \frac{1}{Err(q,\mathcal{H})}$. Denoting $\left[\sqrt{\frac{48\,\mathrm{Tr}(L^{-1})}{\lambda}} + \sqrt{\frac{9\ln(2/\delta)}{2}}\right]$ as C, $s_q = \frac{\sqrt{q}}{C}$.

Recall that from Lemma 1, we need to determine Q_T such that

$$Q_T + \sum_{q=1}^{Q_T} s_q \geq T \implies \sum_{q=1}^{Q_T}(1+s_q) \geq T$$

Since $s_q \geq 1$, this implies that $\sum_{q=1}^{Q_T} 2s_q \geq T$. Substituting the value of s_q and observing that for all q, $s_{q+1} \geq s_q$, we obtain the following:

$$2Q_T s_{Q_T} \geq T \implies 2\frac{Q_T^{3/2}}{C} \geq T \implies Q_T \geq \left(\frac{CT}{2}\right)^{\frac{2}{3}}$$

$$Q_T = \Bigg[\sqrt{\frac{12\,\mathrm{Tr}(L^{-1})}{\lambda}} + \sqrt{\frac{9\ln(2/\delta)}{8}}\Bigg]^{\frac{2}{3}} T^{\frac{2}{3}} \tag{9}$$

Using the above equation with Lemma 1, we can bound the regret as

$$R(T) \leq 2\Bigg[\sqrt{\frac{12\,\mathrm{Tr}(L^{-1})}{\lambda}} + \sqrt{\frac{9\ln(2/\delta)}{8}}\Bigg]^{\frac{2}{3}} T^{\frac{2}{3}} \tag{10}$$

To simplify this expression, we suppress the term $\sqrt{\frac{9\ln(2/\delta)}{8}}$ in the $\tilde{O}$ notation, implying that

$$R(T) = \tilde{O}\left(2\left[\frac{12\,\mathrm{Tr}(L^{-1})}{\lambda}\right]^{\frac{1}{3}} T^{\frac{2}{3}}\right) \tag{11}$$

To present and interpret the result, we keep only the factors which are dependent on n, λ, L and T. We then obtain

$$R(T) = \tilde{O}\left(n^{1/3}\left(\frac{\mathrm{Tr}(L^{-1})}{\lambda n}\right)^{\frac{1}{3}} T^{\frac{2}{3}}\right) \tag{12}$$

$\square$

This proves Theorem 1. We now prove Lemma 3, which was used to bound the Rademacher complexity.

Lemma 3. *The empirical Rademacher complexity for $\mathcal{H}$ under ℓ_{LS} on observing p unbiased samples for each of the n users can be given as:*

$$\hat{\mathcal{R}}_p^n(\ell_{LS} \circ \mathcal{H}) \leq \frac{1}{\sqrt{p}} \sqrt{\frac{48\,\mathrm{Tr}(L^{-1})}{\lambda n}} \tag{13}$$

Proof. The Rademacher complexity for a class of linear predictors with graph regularization for a 0/1 loss function $\ell_{0,1}$ can be bounded using Theorem 2 of [12]. Specifically,

$$\hat{\mathcal{R}}_p^n(\ell_{0,1} \circ \mathcal{H}) \leq \frac{2M}{\sqrt{p}} \sqrt{\frac{\mathrm{Tr}((\lambda L)^{-1})}{n}} \tag{14}$$

where M is the upper bound on the value of $\frac{||L^{\frac{1}{2}}W^*||_2}{\sqrt{n}}$ and W^* is the $d \times n$ matrix corresponding to the true user preferences.

$$\tag{15}$$

We now upper bound $\frac{||L^{\frac{1}{2}}W^*||_2}{\sqrt{n}}$.

$$||L^{\frac{1}{2}}W^*||_2 \leq ||L^{\frac{1}{2}}||_2||W^*||_2$$

$$||W^*||_2 \leq ||W^*||_F = \sqrt{\sum_{i=1}^{n} ||w_i^*||_2^2}$$

$$||W^*||_2 \leq \sqrt{n} \qquad \text{(Using assumption 1: For all } i\text{, } ||w_i^*||_2 \leq 1\text{)}$$

$$||L^{\frac{1}{2}}|| \leq \nu_{max}(L^{\frac{1}{2}}) = \sqrt{\nu_{max}(L)} \leq \sqrt{3}$$

(The maximum eigenvalue of any normalized Laplacian L_G is 2 [4] and recall that $L = L_G + I_n$)

$$\implies \frac{||L^{\frac{1}{2}}W^*||_2}{\sqrt{n}} \leq \sqrt{3} \implies M = \sqrt{3} \tag{16}$$

Since we perform regression using a least squares loss function instead of classification, the Rademacher complexity in our case can be bounded using Theorem 12 from [3]. Specifically, if ρ is the Lipschitz constant of the least squares problem,

$$\hat{\mathcal{R}}_p^n(\ell_{LS} \circ \mathcal{H}) \leq 2\rho \cdot \mathcal{R}_p^n(\ell_{0,1} \circ \mathcal{H}) \tag{17}$$

Since the estimates $w_{i,t}$ are bounded from above by 1 (additional assumption in the theorem), $\rho = 1$. From Equations 15, 17 and the bound on M, we obtain that

$$\hat{\mathcal{R}}_p^n(\ell_{LS} \circ \mathcal{H}) \leq \frac{4}{\sqrt{p}} \sqrt{\frac{3\,\mathrm{Tr}(L^{-1})}{\lambda n}} \tag{18}$$

which proves the lemma. □

C Regret bound for Thompson Sampling

Theorem 2. *Under the following additional technical assumptions: (a) $\log(K) < (dn-1)\ln(2)$ (b) $\lambda < dn$ (c) $\log\left(\frac{3+T/\lambda dn}{\delta}\right) \leq \log(KT)\log(T/\delta)$, with probability $1-\delta$, the regret obtained by Thompson Sampling in the GOB framework is given as:*

$$R(T) = \tilde{O}\left(\frac{dn}{\sqrt{\lambda}}\sqrt{T}\sqrt{\log\left(\frac{\mathrm{Tr}(L^{-1})}{n}\right) + \log\left(3 + \frac{T}{\lambda dn\sigma^2}\right)}\right) \tag{19}$$

Proof. We can interpret graph-based TS as being equivalent to solving a single dn-dimensional contextual bandit problem, but with a modified prior covariance ($(L \otimes I_d)^{-1}$ instead of I_{dn}). Our argument closely follows the proof structure in [2], but is modified to include the prior covariance. For ease of exposition, assume that the target user at each round is implicit. We use j to index the available items. Let the index of the optimal item at round t be j_t^*, whereas the index of the item chosen by our algorithm is denoted j_t.

Let $\hat{r}_t(j)$ be the estimated rating of item j at round t. Then, for all j,

$$\hat{r}_t(j) \sim \mathcal{N}(\langle \mathbf{w}_t, \phi_j \rangle, s_t(j)) \tag{20}$$

Here, $s_t(j)$ is the standard deviation in the estimated rating for item j at round t. Recall that Σ_{t-1} is the covariance matrix at round t. $s_t(j)$ is given as:

$$s_t(j) = \sqrt{\phi_j^T \Sigma_{t-1}^{-1} \phi_j} \tag{21}$$

We drop the argument in $s_t(j_t)$ to denote the standard deviation and estimated rating for the selected item j_t i.e. $s_t = s_t(j_t)$ and $\hat{r}_t = \hat{r}_t(j_t)$.

Let Δ_t measure the immediate regret at round t incurred by selecting item j_t instead of the optimal item j_t^*. The immediate regret is given by:

$$\Delta_t = \langle \mathbf{w}^*, \phi_{j_t^*} \rangle - \langle \mathbf{w}^*, \phi_{j_t} \rangle \tag{22}$$

Define $\mathcal{E}^\mu(t)$ as the event such that for all j,

$$\mathcal{E}^\mu(t) : \quad |\langle \mathbf{w}_t, \phi_j \rangle - \langle \mathbf{w}^*, \phi_j \rangle| \le l_t s_t(j) \tag{23}$$

Here $l_t = \sqrt{dn \log\left(\frac{3+t/\lambda dn}{\delta}\right)} + \sqrt{3\lambda}$. If the event $\mathcal{E}^\mu(t)$ holds, it implies that the expected rating at round t is close to the true rating with high probability.

Recall that $|\mathcal{C}_t| = K$ and that $\tilde{\mathbf{w}}_t$ is a sample drawn from the posterior distribution at round t. Define $\rho_t = \sqrt{9dn \log\left(\frac{t}{\delta}\right)}$ and $g_t = \min\{\sqrt{4dn \ln(t)}, \sqrt{4 \log(tK)}\}\rho_t + l_t$. Define $\mathcal{E}^\theta(t)$ as the event such that for all j,

$$\mathcal{E}^\theta(t) : \quad |\langle \tilde{\mathbf{w}}_t, \phi_j \rangle - \langle \mathbf{w}_t, \phi_j \rangle| \le \min\{\sqrt{4dn \ln(t)}, \sqrt{4 \log(tK)}\}\rho_t s_t(j) \tag{24}$$

If the event $\mathcal{E}^\theta(t)$ holds, it implies that the estimated rating using the sample $\tilde{\mathbf{w}}_t$ is close to the expected rating at round t.

$$\tag{25}$$

In lemma 6, we prove that the event $\mathcal{E}^\mu(t)$ holds with high probability. Formally, for $\delta \in (0, 1)$,

$$\Pr(\mathcal{E}^\mu(t)) \ge 1 - \delta \tag{26}$$

To show that the event $\mathcal{E}^\theta(t)$ holds with high probability, we use the following lemma from [2].

Lemma 4 (Lemma 2 of [2])**.**

$$\Pr(\mathcal{E}^\theta(t)) | \mathcal{F}_{t-1}) \ge 1 - \frac{1}{t^2} \tag{27}$$

Next, we use the following lemma to bound the immediate regret at round t.

Lemma 5 (Lemma 4 in [2])**.** *Let $\gamma = \frac{1}{4e\sqrt{\pi}}$. If the events $\mathcal{E}^{\mu}(t)$ and $\mathcal{E}^{\theta}(t)$ are true, then for any filtration $\mathcal{F}_{t-1}$, the following inequality holds:*

$$\mathbb{E}[\Delta_t|\mathcal{F}_{t-1}] \leq \frac{3g_t}{\gamma}\mathbb{E}[s_t|\mathcal{F}_{t-1}] + \frac{2g_t}{\gamma t^2} \tag{28}$$

Define $\mathcal{I}(\mathcal{E})$ to be the indicator function for an event $\mathcal{E}$. Let $regret(t) = \Delta_t \cdot \mathcal{I}(\mathcal{E}^{\mu}(t))$. We use Lemma 7 (proof is given later) which states that with probability at least $1 - \frac{\delta}{2}$,

$$\sum_{t=1}^{T} regret(t) \leq \sum_{t=1}^{T} \frac{3g_t}{\gamma} s_t + \sum_{t=1}^{T} \frac{2g_t}{\gamma t^2} + \sqrt{2\sum_{t=1}^{T} \frac{36g_t^2}{\gamma^2} \ln(2/\delta)} \tag{29}$$

From Lemma 6, we know that event $\mathcal{E}^{\mu}(t)$ holds for all t with probability at least $1 - \frac{\delta}{2}$. This implies that, with probability $1 - \frac{\delta}{2}$, for all t

$$regret(t) = \Delta_t \tag{30}$$

From Equations 29 and 30, we have that with probability $1 - \delta$,

$$R(T) = \sum_{t=1}^{T} \Delta_t \leq \sum_{t=1}^{T} \frac{3g_t}{\gamma} s_t + \sum_{t=1}^{T} \frac{2g_t}{\gamma t^2} + \sqrt{2\sum_{t=1}^{T} \frac{36g_t^2}{\gamma^2} \ln(2/\delta)}$$

Note that g_t increases with t i.e. for all t, $g_t \leq g_T$

$$R(T) \leq \frac{3g_T}{\gamma} \sum_{t=1}^{T} s_t + \frac{2g_T}{\gamma} \sum_{t=1}^{T} \frac{1}{t^2} + \frac{6g_T}{\gamma} \sqrt{2T \ln(2/\delta)} \tag{31}$$

Using Lemma 8 (proof given later), we have the following bound on $\sum_{t=1}^{T} s_t$, the variance of the selected items:

$$\sum_{t=1}^{T} s_t \leq \sqrt{dnT} \sqrt{C \log\left(\frac{\mathrm{Tr}(L^{-1})}{n}\right) + \log\left(3 + \frac{T}{\lambda dn\sigma^2}\right)} \tag{32}$$

where $C = \frac{1}{\lambda \log\left(1 + \frac{1}{\lambda\sigma^2}\right)}$.

$$\tag{33}$$

Substituting this into Equation 31, we get

$$R(T) \leq \frac{3g_T}{\gamma} \sqrt{dnT} \sqrt{C \log\left(\frac{\mathrm{Tr}(L^{-1})}{n}\right) + \log\left(3 + \frac{T}{\lambda dn\sigma^2}\right)} + \frac{2g_T}{\gamma} \sum_{t=1}^{T} \frac{1}{t^2} + \frac{6g_T}{\gamma} \sqrt{2T \ln(2/\delta)}$$

Using the fact that $\sum_{t=1}^{T} \frac{1}{t^2} < \frac{\pi^2}{6}$

$$R(T) \leq \frac{3g_T}{\gamma} \sqrt{dnT} \sqrt{C \log\left(\frac{\mathrm{Tr}(L^{-1})}{n}\right) + \log\left(3 + \frac{T}{\lambda dn\sigma^2}\right)} + \frac{\pi^2 g_T}{3\gamma} + \frac{6g_T}{\gamma} \sqrt{2T \ln(2/\delta)} \tag{34}$$

We now upper bound g_T. By our assumption on K, $\log(K) < (dn-1)\ln(2)$. Hence for all $t \geq 2$, $\min\{\sqrt{4dn\ln(t)}, \sqrt{4\log(tK)}\} = \sqrt{4\log(tK)}$. Hence,

$$
\begin{aligned}
g_T &= 6\sqrt{dn\log(KT)\log(T/\delta)} + l_T \\
&= 6\sqrt{dn\log(KT)\log(T/\delta)} + \sqrt{dn\log\left(\frac{3+T/\lambda dn}{\delta}\right)} + \sqrt{3\lambda}
\end{aligned}
$$

By our assumption on λ, $\lambda < dn$. Hence,

$$
g_T \leq 8\sqrt{dn\log(KT)\log(T/\delta)} + \sqrt{dn\log\left(\frac{3+T/\lambda dn}{\delta}\right)}
$$

Using our assumption that $\log\left(\frac{3+T/\lambda dn}{\delta}\right) \leq \log(KT)\log(T/\delta)$,

$$
g_T \leq 9\sqrt{dn\log(KT)\log(T/\delta)} \tag{35}
$$

Substituting the value of g_T into Equation 34, we obtain the following:

$$
\begin{aligned}
R(T) \leq{}& \frac{27dn}{\gamma}\sqrt{T}\sqrt{C\log\left(\frac{\mathrm{Tr}(L^{-1})}{n}\right) + \log\left(3 + \frac{T}{\lambda dn\sigma^2}\right)} \\
&+ \frac{3\pi^2\sqrt{dn\ln(T/\delta)\ln(KT)}}{\gamma} + \frac{54\sqrt{dn\ln(T/\delta)\ln(KT)}\sqrt{2T\ln(2/\delta)}}{\gamma}
\end{aligned}
$$

For ease of exposition, we keep the just leading terms on d, n and T. This gives the following bound on $R(T)$.

$$
R(T) = \tilde{O}\left(\frac{27dn}{\gamma}\sqrt{T}\sqrt{C\log\left(\frac{\mathrm{Tr}(L^{-1})}{n}\right) + \log\left(3 + \frac{T}{\lambda dn\sigma^2}\right)}\right)
$$

Rewriting the bound to keep only the terms dependent on d, n, λ, T and L. We thus obtain the following equation.

$$
R(T) = \tilde{O}\left(\frac{dn}{\sqrt{\lambda}}\sqrt{T}\sqrt{\log\left(\frac{\mathrm{Tr}(L^{-1})}{n}\right) + \log\left(3 + \frac{T}{\lambda dn\sigma^2}\right)}\right) \tag{36}
$$

This proves the theorem. □

We now prove the the auxiliary lemmas used in the above proof.

In the following lemma, we prove that $\mathcal{E}^\mu(t)$ holds with high probability, i.e., the expected rating at round t is close to the true rating with high probability.

Lemma 6.

The following statement is true for all $\delta \in (0,1)$:

$$
\Pr(\mathbb{E}^\mu(t)) \geq 1 - \delta \tag{37}
$$

Proof.

Recall that $r_t = \langle \mathbf{w}^*, \phi_{j_t} \rangle + \eta_t$ (Assumption 2) and that $\Sigma_t \mathbf{w}_t = \frac{b_t}{\sigma^2}$. Define $\mathbf{S}_{t-1} = \sum_{l=1}^{t-1} \eta_l \phi_{j_l}$.

$$\mathbf{S}_{t-1} = \sum_{l=1}^{t-1} \left(r_l - \langle \mathbf{w}^*, \phi_{j_l} \rangle\right) \phi_{j_l} = \sum_{l=1}^{t-1} \left(r_l \phi_{j_l} - \phi_{j_l} \phi_{j_l}^T \mathbf{w}^*\right)$$

$$\mathbf{S}_{t-1} = b_{t-1} - \sum_{l=1}^{t-1} \left(\phi_{j_l} \phi_{j_l}^T\right) \mathbf{w}^* = b_{t-1} - \sigma^2 (\Sigma_{t-1} - \Sigma_0) \mathbf{w}^* = \sigma^2 (\Sigma_{t-1} \mathbf{w}_t - \Sigma_{t-1} \mathbf{w}^* + \Sigma_0 \mathbf{w}^*)$$

$$\hat{\mathbf{w}}_t - \mathbf{w}^* = \Sigma_{t-1}^{-1} \left(\frac{\mathbf{S}_{t-1}}{\sigma^2} - \Sigma_0 \mathbf{w}^*\right)$$

The following holds for all j:

$$\begin{aligned}
|\langle \mathbf{w}_t, \phi_j \rangle - \langle \mathbf{w}^*, \phi_j \rangle| &= |\langle \phi_j, \mathbf{w}_t - \mathbf{w}^* \rangle| \\
&\leq \left| \phi_j^T \Sigma_{t-1}^{-1} \left(\frac{\mathbf{S}_{t-1}}{\sigma^2} - \Sigma_0 \mathbf{w}^*\right) \right| \\
&\leq ||\phi_j||_{\Sigma_{t-1}^{-1}} \left(\left|\left| \frac{\mathbf{S}_{t-1}}{\sigma^2} - \Sigma_0 \mathbf{w}^* \right|\right|_{\Sigma_{t-1}^{-1}} \right) && \text{(Since } \Sigma_{t-1}^{-1} \text{ is positive definite)}
\end{aligned}$$

By triangle inequality,

$$|\langle \mathbf{w}_t, \phi_j \rangle - \langle \mathbf{w}^*, \phi_j \rangle| \leq ||\phi_j||_{\Sigma_{t-1}^{-1}} \left(\left|\left| \frac{\mathbf{S}_{t-1}}{\sigma^2} \right|\right|_{\Sigma_{t-1}^{-1}} + ||\Sigma_0 \mathbf{w}^*||_{\Sigma_{t-1}^{-1}} \right) \tag{38}$$

We now bound the term $||\Sigma_0 \mathbf{w}^*||_{\Sigma_{t-1}^{-1}}$

$$\begin{aligned}
||\Sigma_0 \mathbf{w}^*||_{\Sigma_{t-1}^{-1}} &\leq ||\Sigma_0 \mathbf{w}^*||_{\Sigma_0^{-1}} = \sqrt{\mathbf{w}^{*T} \Sigma_0^T \Sigma_0^{-1} \Sigma_0 \mathbf{w}^*} && \text{(Since } \phi_{j_t} \phi_{j_t}^T \text{ is positive definite for all } t\text{)} \\
&= \sqrt{\mathbf{w}^{*T} \Sigma_0 \mathbf{w}^*} && \text{(Since } \Sigma_0 \text{ is symmetric)} \\
&\leq \sqrt{\nu_{max}(\Sigma_0)} ||\mathbf{w}^*||_2 \\
&\leq \sqrt{\nu_{max}(\lambda L \otimes I_d)} && (||\mathbf{w}^*||_2 \leq 1) \\
&= \sqrt{\nu_{max}(\lambda L)} && (\nu_{max}(A \otimes B) = \nu_{max}(A) \cdot \nu_{max}(B)) \\
&\leq \sqrt{\lambda \cdot \nu_{max}(L)} \\
||\Sigma_0 \mathbf{w}^*||_{\Sigma_{t-1}^{-1}} &\leq \sqrt{3\lambda}
\end{aligned}$$

(The maximum eigenvalue of any normalized Laplacian is 2 [4] and recall that $L = L_G + I_n$)

For bounding $||\phi_j||_{\Sigma_{t-1}^{-1}}$, note that

$$||\phi_j||_{\Sigma_{t-1}^{-1}} = \sqrt{\phi_j^T \Sigma_{t-1}^{-1} \phi_j} = s_t(j)$$

Using the above relations, Equation 38 can thus be rewritten as:

$$|\langle \mathbf{w}_t, \phi_j \rangle - \langle \mathbf{w}^*, \phi_j \rangle| \leq s_t(j) \left(\frac{1}{\sigma} ||\mathbf{S}_{t-1}||_{\Sigma_{t-1}^{-1}} + \sqrt{3\lambda} \right) \tag{39}$$

To bound $||\mathbf{S}_{t-1}||_{\Sigma_{t-1}^{-1}}$, we use Theorem 1 from [1] which we restate in our context. Note that using this theorem with the prior covariance equal to I_{dn} gives Lemma 8 of [2].

Theorem 3 (Theorem 1 of [1]). *For any $\delta > 0$, $t \geq 1$, with probability at least $1-\delta$,*

$$||\mathbf{S}_{t-1}||^2_{\Sigma^{-1}_{t-1}} \leq 2\sigma^2 \log\left(\frac{\det(\Sigma_t)^{1/2}\det(\Sigma_0)^{-1/2}}{\delta}\right)$$
$$||\mathbf{S}_{t-1}||^2_{\Sigma^{-1}_{t-1}} \leq 2\sigma^2 \Bigg(\log\left(\det(\Sigma_t)^{1/2}\right) + \log\left(\det(\Sigma_0^{-1})^{1/2}\right) - \log(\delta)\Bigg)$$

Rewriting the above equation,

$$||\mathbf{S}_{t-1}||^2_{\Sigma^{-1}_{t-1}} \leq \sigma^2 \Bigg(\log\left(\det(\Sigma_t)\right) + \log\left(\det(\Sigma_0^{-1})\right) - 2\log(\delta)\Bigg)$$

We now use the trace-determinant inequality. For any $n \times n$ matrix A, $\det(A) \leq \left(\frac{Tr(A)}{n}\right)^n$ which implies that $\log(\det(A)) \leq n \log\left(\frac{Tr(A)}{n}\right)$. Using this for both Σ_t and Σ_0^{-1}, we obtain:

$$||\mathbf{S}_{t-1}||_{\Sigma^{-1}_{t-1}} \leq dn\sigma^2 \Bigg(\log\left(\left(\frac{\mathrm{Tr}(\Sigma_t)}{dn}\right)\right) + \log\left(\left(\frac{\mathrm{Tr}(\Sigma_0^{-1})}{dn}\right)\right) - \frac{2}{dn}\log(\delta)\Bigg) \tag{40}$$

Next, we use the fact that

$$\Sigma_t = \Sigma_0 + \sum_{l=1}^{t} \phi_{j_l}\phi_{j_l}^T \implies \mathrm{Tr}(\Sigma_t) \leq \mathrm{Tr}(\Sigma_0) + t \qquad \text{(Since } ||\phi_{j_l}||_2 \leq 1)$$

Note that $\mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A) \cdot \mathrm{Tr}(B)$. Since $\Sigma_0 = \lambda L \otimes I_d$, it implies that $\mathrm{Tr}(\Sigma_0) = \lambda d \cdot \mathrm{Tr}(L)$. Also note that $\mathrm{Tr}(\Sigma_0^{-1}) = \mathrm{Tr}((\lambda L)^{-1} \otimes I_d) = \frac{d}{\lambda}\mathrm{Tr}(L^{-1})$. Using these relations in Equation 40,

$$\begin{aligned}
||\mathbf{S}_{t-1}||^2_{\Sigma^{-1}_{t-1}} &\leq dn\sigma^2 \Bigg(\log\left(\frac{\lambda d\,\mathrm{Tr}(L) + t}{dn}\right) + \log\left(\frac{\mathrm{Tr}(L^{-1})}{\lambda n}\right) - \frac{2}{dn}\log(\delta)\Bigg) \\
&\leq dn\sigma^2 \Bigg(\log\left(\frac{\mathrm{Tr}(L)\,\mathrm{Tr}(L^{-1})}{n^2} + \frac{t\,\mathrm{Tr}(L^{-1})}{\lambda dn^2}\right) - \log(\delta^{\frac{2}{dn}})\Bigg) && (\log(a) + \log(b) = \log(ab)) \\
&= dn\sigma^2 \log\left(\frac{\mathrm{Tr}(L)\,\mathrm{Tr}(L^{-1})}{n^2\delta} + \frac{t\,\mathrm{Tr}(L^{-1})}{\lambda dn^2\delta}\right) && (\text{Redefining } \delta \text{ as } \delta^{\frac{2}{dn}})
\end{aligned}$$

If $L = I_n$, $\mathrm{Tr}(L) = \mathrm{Tr}(L^{-1}) = n$, we recover the bound in [2] i.e.

$$||\mathbf{S}_{t-1}||^2_{\Sigma^{-1}_{t-1}} \leq dn\sigma^2 \log\left(\frac{1 + t/\lambda dn}{\delta}\right) \tag{41}$$

The upper bound for $\mathrm{Tr}(L)$ is $3n$, whereas the upper bound on $\mathrm{Tr}(L^{-1})$ is n. We thus obtain the following relation.

$$||\mathbf{S}_{t-1}||^2_{\Sigma^{-1}_{t-1}} \leq dn\sigma^2 \log\left(\frac{3}{\delta} + \frac{t}{\lambda dn\delta}\right)$$
$$||\mathbf{S}_{t-1}||_{\Sigma^{-1}_{t-1}} \leq \sigma\sqrt{dn\log\left(\frac{3 + t/\lambda dn}{\delta}\right)} \tag{42}$$

Combining Equations 39 and 42, we have with probability $1-\delta$,

$$|\langle \mathbf{w}_t, \phi_j \rangle - \langle \mathbf{w}^*, \phi_j \rangle| \leq s_t(k) \left(\sqrt{dn \log\left(\frac{3 + t/\lambda dn}{\delta}\right)} + \sqrt{3\lambda} \right)$$
$$|\langle \mathbf{w}_t, \phi_j \rangle - \langle \mathbf{w}^*, \phi_j \rangle| \leq s_t(k) l_t$$

where $l_t = \sqrt{dn \log\left(\frac{3+t/\lambda dn}{\delta}\right)} + \sqrt{3\lambda}$. This completes the proof.

(43)

□

Lemma 7. *With probability* $1-\delta$,

$$\sum_{t=1}^{T} regret(t) \leq \sum_{t=1}^{T} \frac{3g_t}{\gamma} s_t + \sum_{t=1}^{T} \frac{2g_t}{\gamma t^2} + \sqrt{2 \sum_{t=1}^{T} \frac{36 g_t^2}{\gamma^2} \ln \frac{2}{\delta}} \tag{44}$$

Proof.

Let Z_l and Y_t be defined as follows:

$$Z_l = regret(l) - \frac{3g_l}{\gamma} s_l - \frac{2g_l}{\gamma l^2}$$
$$Y_t = \sum_{l=1}^{t} Z_l \tag{45}$$
$$\mathbb{E}[Y_t - Y_{t-1} | \mathcal{F}_{t-1}] = \mathbb{E}[X_t] = \mathbb{E}[regret(t) | \mathcal{F}_{t-1}] - \frac{3g_t}{\gamma} s_t - \frac{2g_t}{\gamma t^2}$$
$$\mathbb{E}[regret(t) | \mathcal{F}_{t-1}] \leq \mathbb{E}[\Delta_t | \mathcal{F}_{t-1}] \leq \frac{3g_t}{\gamma} s_t - \frac{2g_t}{\gamma t^2} \qquad \text{(Definition of } regret(t) \text{ and using lemma 5)}$$
$$\mathbb{E}[Y_t - Y_{t-1} | \mathcal{F}_{t-1}] \leq 0$$

Hence, Y_t is a super-martingale process. We now state and use the Azuma-Hoeffding inequality for Y_t

(46)

Inequality 1 (Azuma-Hoeffding)**.** *If a super-martingale* Y_t *(with* $t \geq 0$*) and its the corresponding filtration* $\mathcal{F}_{t-1}$*, satisfies* $|Y_t - Y_{t-1}| \leq c_t$ *for some constant* c_t*, for all* $t = 1, \dots T$*, then for any* $a \geq 0$*,*

$$Pr(Y_T - Y_0 \geq a) \leq exp\left(\frac{-a^2}{2 \sum_{t=1}^{T} c_t^2} \right) \tag{47}$$

We define $Y_0 = 0$. Note that $|Y_t - Y_{t-1}| = |Z_l|$ is bounded by $1 + \frac{3g_l}{\gamma} - \frac{2g_l}{\gamma l^2}$. Hence, $c_t = \frac{6g_t}{\gamma}$. Setting $a = \sqrt{2 \ln(2/\delta) \sum_{t=1}^{T} c_t^2}$ in the above inequality, we obtain that with probability $1 - \frac{\delta}{2}$,

$$Y_T \leq \sqrt{2 \sum_{t=1}^{T} \frac{36 g_t^2}{\gamma^2} \ln(2/\delta)}$$
$$\sum_{t=1}^{T} \left(regret(t) - \frac{3g_t}{\gamma} s_t - \frac{2g_t}{\gamma t^2} \right) \leq \sqrt{2 \sum_{t=1}^{T} \frac{36 g_t^2}{\gamma^2} \ln(2/\delta)} \tag{48}$$
$$\sum_{t=1}^{T} regret(t) \leq \sum_{t=1}^{T} \frac{3g_t}{\gamma} s_t + \sum_{t=1}^{T} \frac{2g_t}{\gamma t^2} + \sqrt{2 \sum_{t=1}^{T} \frac{36 g_t^2}{\gamma^2} \ln(2/\delta)} \tag{49}$$

$\square$

Lemma 8.

$$\sum_{t=1}^{T} s_t \leq \sqrt{dnT}\sqrt{C\log\left(\frac{\mathrm{Tr}(L^{-1})}{n}\right) + \log\left(3 + \frac{T}{\lambda dn\sigma^2}\right)} \tag{50}$$

Proof.

Following the proof in [5, 15],

$$\begin{aligned}
\det\left[\Sigma_t\right] &\geq \det\left[\Sigma_{t-1} + \frac{1}{\sigma^2}\phi_{j_t}\phi_{j_t}^T\right] \\
&= \det\left[\Sigma_{t-1}^{\frac{1}{2}}\left(I + \frac{1}{\sigma^2}\Sigma_{t-1}^{-\frac{1}{2}}\phi_{j_t}\phi_{j_t}^T\Sigma_{t-1}^{-\frac{1}{2}}\right)\Sigma_{t-1}^{\frac{1}{2}}\right] \\
&= \det\left[\Sigma_{t-1}\right]\det\left[I + \frac{1}{\sigma^2}\Sigma_{t-1}^{-\frac{1}{2}}\phi_{j_t}\phi_{j_t}^T\Sigma_{t-1}^{-\frac{1}{2}}\right] \\
\det\left[\Sigma_t\right] &= \det\left[\Sigma_{t-1}\right]\left(1 + \frac{1}{\sigma^2}\phi_{j_t}^T\Sigma_{t-1}^{-1}\phi_{j_t}\right) = \det\left[\Sigma_{t-1}\right]\left(1 + \frac{s_t^2}{\sigma^2}\right) \\
\log\left(\det\left[\Sigma_t\right]\right) &\geq \log\left(\det\left[\Sigma_{t-1}\right]\right) + \log\left(1 + \frac{s_t^2}{\sigma^2}\right) \\
\log\left(\det\left[\Sigma_T\right]\right) &\geq \log\left(\det\left[\Sigma_0\right]\right) + \sum_{t=1}^{T}\log\left(1 + \frac{s_t^2}{\sigma^2}\right) \qquad (51)
\end{aligned}$$

If A is an $n \times n$ matrix, and B is an $d \times d$ matrix, then $\det[A \otimes B] = \det[A]^d \det[B]^n$. Hence,

$$\begin{aligned}
&\det[\Sigma_0] = \det[\lambda L \otimes I_d] = \det[\lambda L]^d \\
&\det[\Sigma_0] = [\lambda^n \det(L)]^d = \lambda^{dn}[\det(L)]^d \\
&\log\left(\det[\Sigma_0]\right) = dn\log\left(\lambda\right) + d\log\left(\det[L]\right) \qquad (52)
\end{aligned}$$

From Equations 51 and 52,

$$\log\left(\det\left[\Sigma_T\right]\right) \geq \left(dn\log\left(\lambda\right) + d\log\left(\det[L]\right)\right) + \sum_{t=1}^{T}\log\left(1 + \frac{s_t^2}{\sigma^2}\right) \tag{53}$$

We now bound the trace of $\mathrm{Tr}(\Sigma_{T+1})$.

$$\begin{aligned}
&\mathrm{Tr}(\Sigma_{t+1}) = \mathrm{Tr}(\Sigma_t) + \frac{\mathbf{1}}{\sigma^2}\phi_{j_t}\phi_{j_t}^T \implies \mathrm{Tr}(\Sigma_{t+1}) \leq \mathrm{Tr}(\Sigma_t) + \frac{\mathbf{1}}{\sigma^2} \qquad (\text{Since } ||\phi_{j_t}|| \leq 1) \\
&\mathrm{Tr}(\Sigma_T) \leq \mathrm{Tr}(\Sigma_0) + \frac{\mathbf{T}}{\sigma^2}
\end{aligned}$$

Since $\mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A) \cdot \mathrm{Tr}(B)$

$$\mathrm{Tr}(\Sigma_T) \leq \mathrm{Tr}\left(\lambda(L \otimes I_d)\right) + \frac{T}{\sigma^2} \implies \mathrm{Tr}(\Sigma_T) \leq \lambda d\,\mathrm{Tr}(L) + \frac{T}{\sigma^2} \tag{54}$$

Using the determinant-trace inequality, we have the following relation:

$$\begin{aligned}
&\left(\frac{1}{dn}\mathrm{Tr}(\Sigma_T)\right)^{dn} \geq \left(\det[\Sigma_T]\right) \\
&dn\log\left(\frac{1}{dn}\mathrm{Tr}(\Sigma_T)\right) \geq \log\left(\det[\Sigma_T]\right) \qquad (55)
\end{aligned}$$

Using Equations 53, 54 and 55, we obtain the following relation.

$$dn \log\left(\frac{\lambda d \operatorname{Tr}(L) + \frac{T}{\sigma^2}}{dn}\right) \geq \left(dn \log(\lambda) + d \log(\det[L])\right) + \sum_{t=1}^{T} \log\left(1 + \frac{s_t^2}{\sigma^2}\right)$$

$$\begin{aligned}
\sum_{t=1}^{T} \log\left(1 + \frac{s_t^2}{\sigma^2}\right) &\leq dn \log\left(\frac{\lambda d \operatorname{Tr}(L) + \frac{T}{\sigma^2}}{dn}\right) - dn \log(\lambda) - d \log(\det[L]) \\
&\leq dn \log\left(\frac{\lambda d \operatorname{Tr}(L) + \frac{T}{\sigma^2}}{dn}\right) - dn \log(\lambda) + d \log\left(\det[L^{-1}]\right) && (det[L^{-1}] = 1/det[L]) \\
&\leq dn \log\left(\frac{\lambda d \operatorname{Tr}(L) + \frac{T}{\sigma^2}}{dn}\right) - dn \log(\lambda) + dn \log\left(\frac{1}{n} \operatorname{Tr}(L^{-1})\right) \\
& && \text{(Using the determinant-trace inequality for } \log(\det[L^{-1}])) \\
&\leq dn \log\left(\frac{\lambda d \operatorname{Tr}(L) \operatorname{Tr}(L^{-1}) + \frac{\operatorname{Tr}(L^{-1})T}{\sigma^2}}{\lambda d n^2}\right) && (\log(a) + \log(b) = \log(ab)) \\
&\leq dn \log\left(\frac{\operatorname{Tr}(L) \operatorname{Tr}(L^{-1})}{n^2} + \frac{\operatorname{Tr}(L^{-1})T}{\lambda d n^2 \sigma^2}\right)
\end{aligned}$$

The maximum eigenvalue of any Laplacian is 2. Hence $\operatorname{Tr}(L)$ is upper-bounded by $3n$.

$$\sum_{t=1}^{T} \log\left(1 + \frac{s_t^2}{\sigma^2}\right) \leq dn \log\left(\frac{3 \operatorname{Tr}(L^{-1})}{n} + \frac{\operatorname{Tr}(L^{-1})T}{\lambda d n^2 \sigma^2}\right) \tag{56}$$

$$\tag{57}$$

$$\begin{aligned}
s_t^2 &= \phi_j^T \Sigma_t^{-1} \phi_j \leq \phi_j^T \Sigma_0^{-1} \phi_j && \text{(Since we are making positive definite updates at each round } t) \\
&\leq \|\phi_j\|^2 \nu_{max}(\Sigma_0^{-1}) \\
&= \|\phi_j\|^2 \frac{1}{\nu_{min}(\lambda L \otimes I_d)} \\
&= \|\phi_j\|^2 \frac{1}{\nu_{min}(\lambda L)} && (\nu_{min}(A \otimes B) = \nu_{min}(A)\nu_{min}(B)) \\
&\leq \frac{1}{\lambda} \cdot \frac{1}{\nu_{min}(L)} && (||\phi_j||_2 \leq 1) \\
s_t^2 &\leq \frac{1}{\lambda} && \text{(Minimum eigenvalue of a normalized Laplacian } L_G \text{ is 0. } L = L_G + I_n)
\end{aligned}$$

Moreover, for all $y \in [0, 1/\lambda]$, we have $\log\left(1 + \frac{y}{\sigma^2}\right) \geq \lambda \log\left(1 + \frac{1}{\lambda\sigma^2}\right) y$ based on the concavity of $\log(\cdot)$. To see this, consider the following function:

$$h(y) = \frac{\log\left(1 + \frac{y}{\sigma^2}\right)}{\lambda \log\left(1 + \frac{1}{\lambda\sigma^2}\right)} - y \tag{58}$$

Clearly, $h(y)$ is concave. Also note that, $h(0) = h(1/\lambda) = 0$. Hence for all $y \in [0, 1/\lambda]$, the function $h(y) \geq 0$. This implies that $\log\left(1 + \frac{y}{\sigma^2}\right) \geq \lambda \log\left(1 + \frac{1}{\lambda\sigma^2}\right) y$. We use this result by setting $y = s_t^2$.

$$\log\left(1 + \frac{s_t^2}{\sigma^2}\right) \geq \lambda \log\left(1 + \frac{1}{\lambda\sigma^2}\right) s_t^2$$

$$s_t^2 \leq \frac{1}{\lambda \log\left(1 + \frac{1}{\lambda\sigma^2}\right)} \log\left(1 + \frac{s_t^2}{\sigma^2}\right) \tag{59}$$

Hence,

$$\sum_{t=1}^{T} s_t^2 \leq \frac{1}{\lambda \log\left(1+\frac{1}{\lambda\sigma^2}\right)} \sum_{t=1}^{T} \log\left(1+\frac{s_t^2}{\sigma^2}\right) \tag{60}$$

By Cauchy Schwartz,

$$\sum_{t=1}^{T} s_t \leq \sqrt{T}\sqrt{\sum_{t=1}^{T} s_t^2} \tag{61}$$

From Equations 60 and 61,

$$\begin{aligned}
\sum_{t=1}^{T} s_t &\leq \sqrt{T}\sqrt{\frac{1}{\lambda \log\left(1+\frac{1}{\lambda\sigma^2}\right)} \sum_{t=1}^{T} \log\left(1+\frac{s_t^2}{\sigma^2}\right)} \\
\sum_{t=1}^{T} s_t &\leq \sqrt{T}\sqrt{C \sum_{t=1}^{T} \log\left(1+\frac{s_t^2}{\sigma^2}\right)}
\end{aligned} \tag{62}$$

where $C = \frac{1}{\lambda \log\left(1+\frac{1}{\lambda\sigma^2}\right)}$. Using Equations 56 and 62,

$$\begin{aligned}
\sum_{t=1}^{T} s_t &\leq \sqrt{dnT}\sqrt{C \log\left(\frac{3\,\mathrm{Tr}(L^{-1})}{n} + \frac{\mathrm{Tr}(L^{-1})T}{\lambda d n^2 \sigma^2}\right)} \\
\sum_{t=1}^{T} s_t &\leq \sqrt{dnT}\sqrt{C \log\left(\frac{\mathrm{Tr}(L^{-1})}{n}\right) + \log\left(3 + \frac{T}{\lambda d n \sigma^2}\right)}
\end{aligned} \tag{63}$$

□

References

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

[2] Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. *arXiv preprint arXiv:1209.3352*, 2012.

[3] Peter L Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *The Journal of Machine Learning Research*, 3:463–482, 2003.

[4] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc.

[5] Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic linear optimization under bandit feedback. In *21st Annual Conference on Learning Theory - COLT 2008, Helsinki, Finland, July 9-12, 2008*, pages 355–366, 2008.

[6] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[7] Andre R Goncalves, Puja Das, Soumyadeep Chatterjee, Vidyashankar Sivakumar, Fernando J Von Zuben, and Arindam Banerjee. Multi-task sparse structure learning. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 451–460. ACM, 2014.

[8] André R Gonçalves, Fernando J Von Zuben, and Arindam Banerjee. Multi-label structure learning with ising model selection. In *Proceedings of the 24th International Conference on Artificial Intelligence*, pages 3525–3531. AAAI Press, 2015.

[9] Cho-Jui Hsieh, Inderjit S Dhillon, Pradeep K Ravikumar, and Mátyás A Sustik. Sparse inverse covariance matrix estimation using quadratic approximation. In *Advances in Neural Information Processing Systems*, pages 2330–2338, 2011.

[10] Cho-Jui Hsieh, Mátyás A Sustik, Inderjit S Dhillon, Pradeep K Ravikumar, and Russell Poldrack. Big & quic: Sparse inverse covariance estimation for a million variables. In *Advances in Neural Information Processing Systems*, pages 3165–3173, 2013.

[11] John Langford and Tong Zhang. The epoch-greedy algorithm for multi-armed bandits with side information. In *Advances in neural information processing systems*, pages 817–824, 2008.

[12] Andreas Maurer. The rademacher complexity of linear transformation classes. In *Learning Theory*, pages 65–78. Springer, 2006.

[13] Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC Press, 2005.

[14] Avishek Saha, Piyush Rai, Suresh Venkatasubramanian, and Hal Daume. Online learning of multiple tasks and their relationships. In *International Conference on Artificial Intelligence and Statistics*, pages 643–651, 2011.

[15] Zheng Wen, Branislav Kveton, and Azin Ashkan. Efficient learning in large-scale combinatorial semi-bandits. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 1113–1122, 2015.