Efficient Algorithms for Logistic Contextual Slate Bandits with Bandit Feedback

Tanmay Goyal[1] **Gaurav Sinha**[1]

[1]Microsoft Research India

Abstract

We study the Logistic Contextual Slate Bandit problem, where, at each round, an agent selects a slate of N items from an exponentially large set (of size $2^{\Omega(N)}$) of candidate slates provided by the environment. A single binary reward, determined by a logistic model, is observed for the chosen slate. Our objective is to develop algorithms that maximize cumulative reward over T rounds while maintaining low per-round computational costs. We propose two algorithms, `Slate-GLM-OFU` and `Slate-GLM-TS`, that accomplish this goal. These algorithms achieve $N^{O(1)}$ per-round time complexity via "local planning" (independent slot selections), and low regret through "global learning" (joint parameter estimation). We provide theoretical and empirical evidence supporting these claims. Under a well-studied diversity assumption, we prove that `Slate-GLM-OFU` incurs only $\tilde{O}(\sqrt{T})$ regret. Extensive experiments across a wide range of synthetic settings demonstrate that our algorithms consistently outperform state-of-the-art baselines, achieving both the lowest regret and the fastest runtime. Furthermore, we apply our algorithm to select in-context examples in prompts of Language Models for solving binary classification tasks such as sentiment analysis. Our approach achieves competitive test accuracy, making it a viable alternative in practical scenarios.

1 INTRODUCTION

Online slate bandit problems provide a popular framework for modeling decision-making scenarios where multiple items must be selected in each round. A slate consists of multiple slots, each with its own pool of candidate items, which may change over time. In each round, the learner selects one item per slot, thereby forming a slate. A single reward drawn from a logistic model with unknown parameters is then received for the entire slate. The learner's objective is to adaptively optimize their slate selection policy to maximize the cumulative reward (or equivalently, minimize the cumulative regret) over time. Online slate bandits naturally model various real-world applications. A prominent example is landing page optimization [Hill et al., 2017], where the goal is to optimize the selection of components for each part of a landing page to maximize conversions. Another important application is the automatic optimization of advertising creatives [Chen et al., 2021], which requires advertisers to automatically compose ads from multiple elements, such as product images, text descriptions, and titles. Beyond these practical applications, slate bandits have been extensively studied in the academic literature, leading to the development of many interesting algorithms in diverse settings [Kale et al., 2010, Dimakopoulou et al., 2019, Rhuggenaath et al., 2020].

Although good progress has been made on a variety of online slate bandit settings, some significant challenges still remain that limit the applicability of these algorithms. In applications such as those mentioned above, at each round, the learner has access to some contextual information (such as user query, user history, or demographics) which influences the set of available items per slot. To the best of our knowledge, the current literature focuses heavily on the non-contextual (fixed arms[1]) setting, i.e., they do not assume access to such contexts and therefore keep the set of items unchanged over time. Another limitation is that most of the prior work assumes that the reward of a slate is a function (known or unknown) of rewards of the items in the slate which are themselves either adversarially chosen or are stochastic but disjoint from each other (i.e., each item's reward comes from a different distribution). This assumption neglects the inherent similarities between items. A more realistic approach is to assume a unified parametric reward model shared across all slates. This model allows the learner

[1]We use the terms arms and actions interchangeably.

to leverage shared information, significantly simplifying the learning process. Specifically, for binary rewards, models based on the logistic or probit function can effectively capture the reward structure.

A third, and equally important, limitation is the prevalent focus on the semi-bandit feedback setting. This setting provides separate reward feedback for each item within a selected slate. However, many practical applications (e.g., the ad creatives problem [Chen et al., 2021]) offer only a single, slate-level reward (i.e., bandit feedback). Although there are some methods for converting bandit feedback to semi-bandit feedback [Dimakopoulou et al., 2019], these are often heuristic and lack theoretical guarantees. The item-level feedback in the semi-bandit setting facilitates per-slot exploration and exploitation, enabling the development of algorithms with $N^{O(1)}$ per-round complexity (e.g., [Kale et al., 2010, Rhuggenaath et al., 2020]) by avoiding explicit iteration over the entire slate space. It remains unclear how to achieve similar efficiency in the more challenging bandit feedback setting. For example, directly applying state-of-the-art bandit algorithms [Lattimore and Szepesvári, 2020] to the slate bandit problem (treating slates as arms) and selecting a slate by iterating through the $2^{\Omega(N)}$ sized set of all possible slates, results in exponential per-round time complexity.

Motivated by these challenges, our work introduces efficient and optimal algorithms for the logistic contextual slate bandit problem under bandit feedback, assuming time-varying item features and rewards generated from a global logistic model. We make the following contributions.

1.1 OUR CONTRIBUTIONS

1. We propose two new algorithms `Slate-GLM-OFU` and `Slate-GLM-TS` that solve the logistic contextual slate bandit problem under bandit feedback. While `Slate-GLM-OFU` is based on the OFU (Optimization in the Face of Uncertainty) paradigm, `Slate-GLM-TS` follows the Thompson Sampling (TS) paradigm. Under a diversity assumption (Assumption 2.1), we prove that `Slate-GLM-OFU` incurs a regret of $\tilde{O}(dN\sqrt{T})$ with high probability. Here, d is the dimensionality of the items in the slate, N is the number of slots and T is the total number of rounds the algorithm is run for. Both algorithms explore and exploit at the slot level and thus have a per round time complexity that grows polynomially in N and $\log T$, making them feasible in practice.
2. We also propose a fixed arm version `Slate-GLM-TS-Fixed` of the `Slate-GLM-TS` algorithm for the non-contextual (fixed arm) setting. Using an assumption similar to Assumption 2.1, we prove an $O(d^{3/2}N^{3/2}\sqrt{T})$ regret guarantee for `Slate-GLM-TS-Fixed`. Similar to `Slate-GLM-TS`, `Slate-GLM-TS-Fixed` also explores and exploits at the slot level and has per round complexity polynomial in N and $\log T$.
3. We perform extensive experiments to validate the performance of our algorithms for both the contextual and the non-contextual settings. Under a wide range of randomly selected instances, we see that `Slate-GLM-OFU` incurs the least regret compared to all baselines and `Slate-GLM-TS`, `Slate-GLM-TS-Fixed` are competitive with other state-of-the-art algorithms. We also evaluate the maximum and average per round time complexity of our algorithms and compare it to the time complexities of the baselines. Our algorithms are exponentially (most of the time) faster than all baselines.
4. Finally, we use our algorithm `Slate-GLM-OFU` to select in-context examples for tuning prompts of language models, applied to binary classification tasks. We perform experiments on two datasets *SST2* and *Yelp Review* and achieve a competitive test accuracy of $\sim 80\%$ making it a possible alternative in practical prompt tuning scenarios.

1.2 RELATED WORK

Online slate bandits have received significant attention due to their wide applicability in applications such as recommendations and advertising [Hill et al., 2017, Chen et al., 2021, Dimakopoulou et al., 2019], however, there are only a few theoretical studies that provide regret guarantees [Kale et al., 2010, Rhuggenaath et al., 2020]. While these papers make progress on the slate bandit problem, neither do they address the contextual setting, nor do they accommodate bandit feedback which are the main motivations of our work. Theoretical analysis might be feasible for the Thompson Sampling approach in Dimakopoulou et al. [2019], but proving optimal guarantees might still be hard since their algorithm assigns equal rewards to all slots in order to maintain slot level policies for efficiency purposes. However, we would like to acknowledge that in our experiments (Section 5), for the fixed arms setting, we found their algorithm to be quite competitive to ours on the instances we considered.

One way of achieving optimal regret guarantees for the slate bandit problem is to reduce it to the canonical logistic bandit problem by considering each candidate slate as a separate arm and then using state of the art algorithms such as those in [Faury et al., 2020, Abeille et al., 2021, Faury et al., 2022]. While these algorithms do achieve optimal (κ free) regret, they are infeasible in practice. During the arm selection step they require an iteration through all the arms which is a $2^{\Omega(N)}$ sized set, thereby incurring exponential time per round. Even though these algorithms are inefficient for the slate bandit problem, we combine some of their key ideas with an efficient planning approach to design our algorithms. In Section 5, we demonstrate that our algorithms perform

better than these state of the art logistic bandit algorithms both in regret and time complexity, when applied to a wide variety of slate bandit instances.

Recently a large number of works [Swaminathan et al., 2017, Kiyohara et al., 2024, Vlassis et al., 2024] have studied the slate bandit problem in the off-policy setting, wherein they utilize a dataset collected using some base policy to find optimal slate bandit policies. While these works have made significant progress both from the theoretical and practical sides, they are not relevant to our work since we focus on the online setting only.

2 PRELIMINARIES

In this section, we define the notations used in the paper. Following this, we formulate the Slate Bandits problem and present the assumptions that enable us to prove the regret guarantee provided in Theorem 3.1 and Theorem C.1.

Notations The set $\{1, 2 \ldots, N\}$ is denoted as $[N]$. Unless otherwise specified, we use bold upper case letters for matrices, bold lower case letters for vectors, and upper case calligraphic symbols or greek letters for sets. For any matrix $\mathbf{A}$, we denote its minimum and maximum eigenvalues as $\lambda_{min}(\mathbf{A})$ and $\lambda_{max}(\mathbf{A})$ respectively. We write $\mathbf{A} \succcurlyeq 0$, if matrix $\mathbf{A}$ is positive semi-definite and $\mathbf{A} \succcurlyeq \mathbf{B}$, if $\mathbf{A} - \mathbf{B} \succcurlyeq 0$. For a positive semi-definite matrix $\mathbf{A}$, we define the norm of a vector $\mathbf{x}$ with respect to $\mathbf{A}$ as $\|\mathbf{x}\|_{\mathbf{A}} = \sqrt{\mathbf{x}^\top \mathbf{A} \mathbf{x}}$ and the spectral norm of $\mathbf{A}$ as $\|\mathbf{A}\|_2 = \sqrt{\lambda_{max}(\mathbf{A}^\top \mathbf{A})}$. We use $\mathbf{I}_m$ and $\mathbf{0}_m$ to denote the $m \times m$ identity and zero matrices respectively. When the dimension m is clear from the context, we use $\mathbf{I}$ and $\mathbf{0}$ instead. The symbols $\mathbb{P}$ and $\mathbb{E}$ denote probability and expectation respectively. For sets $\mathcal{A}, \mathcal{X}$ that are subsets of some ambient space $\mathbb{R}^m$, we define the diameter of $\mathcal{X}$ as $diam(\mathcal{X}) = \max_{\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$ and the diameter with respect to $\mathcal{A}$ as $diam_{\mathcal{A}}(\mathcal{X}) = \max_{\boldsymbol{a} \in \mathcal{A}} \max_{\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}} |\boldsymbol{a}^\top (\boldsymbol{x}_1 - \boldsymbol{x}_2)|$.

2.1 SLATE BANDITS

In the Slate Bandits problem, a learner interacts with the environment over T rounds. At each round $t \in [T]$, the learner is presented with N finite sets $\mathcal{X}_t^i$ $(\subset \mathbb{R}^d), i \in [N]$, of *items* and is expected to select one item (say $\mathbf{x}_t^i$) from each $\mathcal{X}_t^i$. Based on the selected N-tuple $\mathbf{x}_t = (\mathbf{x}_t^1, \ldots \mathbf{x}_t^N)$ (called a "slate") the learner receives a stochastic binary reward $y_t(\mathbf{x}_t)$. The learner's goal is to select slates $\mathbf{x}_t, t \in [T]$ such that her expected regret,

$$Regret(T) = \sum_{t=1}^{T} \left\{ \max_{\mathbf{x} \in \mathcal{X}_t} \mathbb{E}[y_t(\mathbf{x})] - \mathbb{E}[y_t(\mathbf{x}_t)] \right\}$$

is minimized[2]. Here, $\mathcal{X}_t$ denotes the set $\mathcal{X}_t^1 \times \ldots \times \mathcal{X}_t^N$ of all possible slates at round t. When the chosen slate $\mathbf{x}_t$ is clear from the context, for simplicity, we will denote $y_t(\mathbf{x}_t)$ as y_t. For convenience, we say that the slate $\mathbf{x}_t$ comprises of N "slots", and the item $\mathbf{x}_t^i$ is placed in slot i in the slate.

In this work, we consider two well known settings; (a) **Stochastic Contextual** and (b) **Non-Contextual** (also known as Fixed-Arm setting). In the first setting, we assume that at every round $t \in [T]$, the set $\mathcal{X}_t^i$ is constructed by sampling from a distribution (unknown to the learner) $\mathbb{D}_i$, in an i.i.d fashion. Moreover, $\mathcal{X}_t^i$ and $\mathcal{X}_s^j$ are sampled independently of one another, for all $s, t \in [T]$ and $i, j \in [N]$. In the second setting, we assume $\mathcal{X}_t^i$ remains fixed over time. Thus, in this setting, for simplicity, we denote $\mathcal{X}_t^i$ by $\mathcal{X}^i$.

Logistic rewards In this paper, we assume that the binary reward variable y_t comes from a Logistic Model. Therefore, $\mathbb{P}[y_t = 1 \mid \mathbf{x}_t] = \mu(\mathbf{x}_t^\top \boldsymbol{\theta}^\star)$, where $\mu : \mathbb{R} \to \mathbb{R}$ is the logistic function, i.e., $\mu(a) = 1/(1 + \exp(-a))$, and $\theta^\star \in \mathbb{R}^{dN}$ is an unknown $d \times N$ dimensional parameter vector. Similar to prior works on Logistic bandits [Faury et al., 2020, Abeille et al., 2021, Faury et al., 2022], we assume that $\|\theta^\star\|_2 \leq S$, where S is known to the learner, and $\left\|\mathbf{x}^i\right\|_2 \leq 1/\sqrt{N}$, for all $\mathbf{x}^i \in \mathcal{X}_t^i, i \in [N], t \in [T]$[3]. Recent logistic bandit literature [Filippi et al., 2010, Faury et al., 2020, Abeille et al., 2021, Faury et al., 2022] also identifies a critical parameter κ, that captures the non-linearity of the reward for the given problem instance, defined as follows.

$$\kappa = \max_{t \in [T]} \max_{\mathbf{x} \in \mathcal{X}_t, \boldsymbol{\theta} \in \Theta} \frac{1}{\dot{\mu}(\mathbf{x}^\top \boldsymbol{\theta})} \tag{1}$$

where $\Theta = \{\|\theta\|_2 \leq S\} \subset \mathbb{R}^{dN}$. The parameter κ can be intuitively seen as the mismatch between the true reward function and a linear approximation of the same. Developing algorithms with regret independent of κ has gained significant attention recently [Faury et al., 2020, Abeille et al., 2021, Faury et al., 2022, Sawarni et al., 2024] and is an active area of research. We refer the reader to Section 2 of Faury et al. [2020] for a thorough discussion on κ and its implications on regret analysis.

Assumption 2.1. ***(Diversity Assumption)*** *We describe a key assumption that enables us to design algorithms with low per-round computational complexity and strong regret guarantees (Theorem 3.1 in Section 3 and Theorem C.1 in Appendix C). Let $\mathcal{F}_t$ be the sigma algebra generated by $\{\mathbf{x}_1, y_1, \ldots, \mathbf{x}_{t-1}, y_{t-1}\}$ and $\phi = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \mathcal{F}_T$, be the associated filtration. For each $i \in [N]$, $t \in [T]$, we assume that,*

$$\mathbb{E}[\mathbf{x}_t^i \mid \mathcal{F}_t] = \mathbf{0} \quad \textit{and} \quad \mathbb{E}[\mathbf{x}_t^i {\mathbf{x}_t^i}^\top \mid \mathcal{F}_t] \succcurlyeq \rho\kappa\mathbf{I}$$

[2]We also use $R(T)$ for shorthand.

[3]This implies the usual assumption $\|\mathbf{x}\|_2 \leq 1$ for all $\mathbf{x} \in \mathcal{X}_t$.

where $\rho > 0$ is a fixed constant and κ is the non-linearity parameter defined earlier in Section 2.

Remarks on Assumption 2.1: The assumption intuitively means that for each slot $i \in [N]$ and round $t \in [T]$, the item features $\mathbf{x}_t^i$ that can be selected by the algorithm are sufficiently "diverse", i.e., the expected matrix $\mathbb{E}[\mathbf{x}_t^i \mathbf{x}_t^{i\top} \mid \mathcal{F}_t]$ is full rank and has sufficiently large eigenvalues. In our proofs, this assumption is used to first prove that with high probability the minimum eigenvalue of certain design matrices $\mathbf{W}_t^i = \gamma \mathbf{I} + \sum_{s \in [t]} \dot{\mu}(\mathbf{x}_s^\top \theta_{s+1}) \mathbf{x}_s^i \mathbf{x}_s^{i\top}$ used by our algorithms (Algorithms 1, 3, 4) grows (sufficiently) linearly with t. In particular, we show that (Lemma D.1, Appendix D) $\lambda_{min}(\mathbf{W}_t^i) \geq \gamma + c\rho\kappa t$, for a fixed constant $c > 0$. We critically utilize this linear growth of the minimum eigenvalue (Lemma B.9, Appendix B.1 and Lemma C.2, Appendix C.2) to prove multiplicative equivalence between the block diagonal matrix $\mathbf{U}_t = diag(\mathbf{W}_t^1, \ldots, \mathbf{W}_t^N)$ and a similarly defined slate-level design matrix $\mathbf{W}_t = \gamma \mathbf{I} + \sum_{s \in [t]} \dot{\mu}(\mathbf{x}_s^\top \theta_{s+1}) \mathbf{x}_s \mathbf{x}_s^\top$. As a result of this multiplicative equivalence, we are able to use slot level exploration bonuses[4] (leading to low per round time complexity in Algorithms 1, 3 and 4), and still continue to have optimal regret. Details of the algorithm and the regret proof can be found in Sections 3, 4 and Appendix C. We would like to highlight that many similar diversity assumptions have been used in the literature and connections between them have also been studied (Section 3 Papini et al. [2021]). Depending on the strength of the assumption, novel and stronger regret guarantees for well-known algorithms have been established, (e.g., Lemma 2, Papini et al. [2021] and Corollary 4, Das and Sinha [2024]). Interestingly, their regret proofs also proceed by first showing a linear lower bound on the minimum eigenvalue of the design matrix. Since the assumption is instance/algorithm dependent, there could be instances where the linear lower bound might not hold. To study this, we empirically examine the growth of the minimum eigenvalues ($\lambda_{min}(\mathbf{W}_t^i)$) for a large number of randomly chosen instances and see a clear linear trend validating the assumption, at least for these randomly picked instances. More details can be found in Appendix G.

3 SLATE-GLM-OFU

In this section, we present our first algorithm `Slate-GLM-OFU` (Algorithm 1) based on the OFU (Optimization in the Face of Uncertainty) paradigm [Abbasi-yadkori et al., 2011] used in bandit algorithms. At a high level, `Slate-GLM-OFU` (along with sub-routine Algorithm 2) builds upon the `ada-OFU-ECOLog` algorithm (Algorithm 2 in Faury et al. [2022]) which achieves an optimal (κ-free) $O(\sqrt{T})$ regret guarantee for logistic reward models and incurs $O(K \log T)$ per round computational cost, where K is the total number of actions to choose from. In the slate bandit setting, K is exponential in N, the number of slots in the slate, making a direct application of `ada-OFU-ECOLog` infeasible when N is large. To address this, `Slate-GLM-OFU` selects an item for each slot independently, reducing the per round computational cost to $N^{O(1)}$. Interestingly, despite the independent selection of items to build the slate, `Slate-GLM-OFU` (via sub-routine Algorithm 2) estimates only a single reward model using the slate level reward feedback. This is a critical difference with respect to prior works on slate bandits with bandit feedback [Dimakopoulou et al., 2019] which attribute the single slate level reward feedback to individual items in the slate and estimates N separate reward models.

Input to `Slate-GLM-OFU` are T, δ and S, where T is the time horizon i.e., the total number of rounds, δ is the error probability and S is a known upper bound for $\|\boldsymbol{\theta}^\star\|_2$. Similar to `ada-OFU-ECOLog` [Faury et al., 2022],

[4]Instead of slate level exploration.

Algorithm 1 `Slate-GLM-OFU`

Inputs: T, δ, S
Initialize $\mathbf{W}_1^1 = \ldots = \mathbf{W}_1^N = \mathbf{I}_d$, $\mathbf{W}_1 = I_{dN}$, $\Theta_1 = \{\|\theta\|_2 \leq S\}$, $\theta_1 \in \Theta_1$, $\eta_t(\delta) = O(S^2 N d \log(t/\delta))$, and $\mathcal{H}_1 = \emptyset$
for each round $t \in [T]$ **do**
  Obtain the set of items $\mathcal{X}_t^i, \forall i \in [N]$, and find $\mathbf{x}_t^i = \arg\max_{\mathbf{x} \in \mathcal{X}_t^i} \langle \mathbf{x}^\top \theta_t^i \rangle + \sqrt{\eta_t(\delta)} \|\mathbf{x}\|_{(\mathbf{W}_t^i)^{-1}}$
  Select slate $\mathbf{x}_t = (\mathbf{x}_t^1, \ldots, \mathbf{x}_t^N)$ and get reward y_t.
  Obtain $\boldsymbol{\theta}_{t+1}$, $\{\mathbf{W}_{t+1}^i\}_{i=1}^N$, Θ_{t+1}, $\mathcal{H}_{t+1}$ by calling Algorithm 2 with inputs $\mathbf{x}_t$, y_t, $\boldsymbol{\theta}_t$, $\mathbf{W}_t$, $\{\mathbf{W}_t^i\}_{i=1}^N$, Θ_t, $\mathcal{H}_t$
end for

Algorithm 2 `ada-OFU-ECOLog-Updates`

Inputs: $\mathbf{x}_t, y_t, \boldsymbol{\theta}_t, \mathbf{W}_t, \{\mathbf{W}_t^i\}_{i=1}^N, \Theta_t, \mathcal{H}_t$
Initialize $\gamma_t(\delta) = O(S^2 N d \log(t/\delta))$ and $\beta_t(\delta) = O(S^6 N d \log(t/\delta))$.
Compute $\bar{\theta}_t$, θ_t^0, and θ_t^1 using 3 and 4
if $\dot{\mu}(\mathbf{x}_t^\top \bar{\theta}_t) \leq 2\dot{\mu}(\mathbf{x}_t^\top \theta_t^u)$ for $u \in \{0, 1\}$ **then**
  Let θ_{t+1} be solution of 5 up to precision $1/t$.
  $\mathbf{W}_{t+1}^i = \mathbf{W}_t^i + \dot{\mu}(\mathbf{x}_t^\top \theta_{t+1}) \mathbf{x}_t^i \mathbf{x}_t^{i\top}, \forall i \in [N]$
  $\mathbf{W}_{t+1} = \mathbf{W}_t + \dot{\mu}(\mathbf{x}_t^\top \theta_{t+1}) \mathbf{x}_t \mathbf{x}_t^\top$
  $\mathcal{H}_{t+1} = \mathcal{H}_t$ and $\Theta_{t+1} = \Theta_t$
else
  $\mathcal{H}_{t+1} = \mathcal{H}_t \cup \{(\mathbf{x}_t, y_t)\}$.
  Let $\theta_{t+1}^{\mathcal{H}}$ be solution of 6 up to precision $1/t$.
  $\mathbf{V}_t^{\mathcal{H}} = \sum_{\mathbf{x} \in \mathcal{H}_t} \mathbf{x}\mathbf{x}^\top / \kappa + \gamma_t(\delta) \mathbf{I}_{Nd}$
  $\Theta_{t+1} = \left\{ \left\|\theta - \theta_{t+1}^{\mathcal{H}}\right\|_{\mathbf{V}_t^{\mathcal{H}}}^2 \leq \beta_t(\delta) \right\} \cap \Theta_1$
  $\theta_{t+1} = \theta_t$, $\mathbf{W}_{t+1} = \mathbf{W}_t$, $\mathbf{W}_{t+1}^i = \mathbf{W}_t^i, \forall i \in [N]$
end if
return $\boldsymbol{\theta}_{t+1}, \mathbf{W}_{t+1}, \{\mathbf{W}_{t+1}^i\}_{i=1}^N, \Theta_{t+1}, \mathcal{H}_{t+1}$

`Slate-GLM-OFU` maintains vectors $\boldsymbol{\theta}_t$, and sets Θ_t and $\mathcal{H}_t$. The vector $\boldsymbol{\theta}_t$ provides an estimate of $\boldsymbol{\theta}^\star$ during the tth round. Set $\Theta_t \subseteq \Theta_1 = \{\|\boldsymbol{\theta}\|_2 \leq S\}$ is an admissible set for the values of $\boldsymbol{\theta}_{t+1}$ and contains the true reward parameter $\boldsymbol{\theta}^\star$ with high probability (See Proposition 7 in Faury et al. [2022] for more details). In order to facilitate adaptivity, `ada-OFU-ECOLog` introduced the set $\mathcal{H}_t$ comprising pairs $(\mathbf{x}_s, y_s(\mathbf{x}_s))$ $(s \leq t)$ at which an inequality criterion (described in *Step 3* of Algorithm 2) fails. In addition to these, `ada-OFU-ECOLog` also introduces a matrix $\mathbf{W}_t = \lambda \mathbf{I} + \sum_{s=1}^{t-1} \dot{\mu}(\mathbf{x}_s^\top \theta_{s+1}) \mathbf{x}_s \mathbf{x}_s^\top$ as on-policy proxy for the *concentration matrix* $\mathbf{H}_t = \lambda \mathbf{I} + \sum_{s=1}^{t-1} \dot{\mu}(\mathbf{x}_s^\top \theta^\star) \mathbf{x}_s \mathbf{x}_s^\top$, to enable efficient per round computation of parameter estimates. In `Slate-GLM-OFU`, along with $\mathbf{W}_t$, we also maintain N other such matrices (one for each slot $i \in [N]$), $\mathbf{W}_t^i = \lambda I + \sum_{s=1}^{t-1} \dot{\mu}(\mathbf{x}_s^\top \theta_{s+1}) \mathbf{x}_s^i \mathbf{x}_s^{i\top}$. These matrices help us in the explore-exploit trade-off while selecting the item for the ith slot.

Next we go through the steps of `Slate-GLM-OFU` (Algorithm 1) and its sub-routine (Algorithm 2) to provide a more detailed explanation. *Steps 3-7* (Algorithm 1) is where `Slate-GLM-OFU` differs significantly from `ada-OFU-ECOLog`. Instead of getting the set of arm features $\mathcal{X}_t$ (slates in our case) directly from the environment (as in `ada-OFU-ECOLog`), `Slate-GLM-OFU` receives N different sets of items $\mathcal{X}_t^i$, for each slot $i \in [N]$. Then, it picks one item $\mathbf{x}_t^i \in \mathcal{X}_t^i$, using the optimistic rule mentioned in *Step 4* (Algorithm 1). Note that, the underlying optimization problem for slot i, only requires the candidate items in $\mathcal{X}_t^i$ and the components θ_t^i of $\boldsymbol{\theta}_t$ that correspond to the ith slot, and thus, can be solved independently and in parallel for all slots. Why the selection of items independently at the slot level leads to optimal selection at the slate level is quite interesting and constitutes the core technical part of our regret guarantee (Theorem 3.1). Essentially, we can show that, under our diversity assumption (Assumption 2.1), the positive definite matrices $\mathbf{W}_t$ and $diag(\mathbf{W}_t^1, \ldots, \mathbf{W}_t^N)$ are multiplicatively equivalent, further implying that, for all slates $\mathbf{x}_t = (\mathbf{x}_t^1, \ldots, \mathbf{x}_t^N)$, the quantities $\|\mathbf{x}_t\|_{\mathbf{W}_t}$ and $\sum_{i \in [N]} \|\mathbf{x}_t^i\|_{\mathbf{W}_t^i}$ are multiplicatively equivalent. This observation is exploited in our algorithm to convert an optimistic selection rule at the slate level into an equivalent optimistic selection rule for each slot. In *Step 5* (Algorithm 1), we select the slate $\mathbf{x}_t = (\mathbf{x}_t^1, \ldots, \mathbf{x}_t^N)$, yielding a reward y_t. At this point, `Slate-GLM-OFU` calls a sub-routine described in Algorithm 2 which updates the parameters $\boldsymbol{\theta}_t$, $\mathbf{W}_t$, $(\mathbf{W}_t^1, \ldots, \mathbf{W}_t^N)$, Θ_t, and $\mathcal{H}_t$. The update rules in Algorithm 2 largely follow the one in `ada-OFU-ECOLog`, which is based on the following inequality criterion.

$$\dot{\mu}(\mathbf{x}_t^\top \bar{\boldsymbol{\theta}}_t) \leq 2 \min\{\dot{\mu}(\mathbf{x}_t^\top \theta_t^0), \dot{\mu}(\mathbf{x}_t^\top \theta_t^1)\} \tag{2}$$

Here $\bar{\boldsymbol{\theta}}_t, \boldsymbol{\theta}_t^0, \boldsymbol{\theta}_t^1 \in \mathbb{R}^{dN}$, are $\mathcal{F}_t$-adapted parameters that enable adaptivity. They are obtained as follows.

$$\bar{\theta}_t = \underset{\boldsymbol{\theta} \in \Theta_t}{\arg\min} \left[\eta \|\theta - \theta_t\|_{\mathbf{W}_t}^2 + \sum_{u \in \{0,1\}} \ell(\mathbf{x}_t^\intercal \theta, u) \right] \tag{3}$$

$$\theta_t^u = \underset{\boldsymbol{\theta} \in \Theta_t}{\arg\min} \left[\eta \|\theta - \theta_t\|_{\mathbf{W}_t}^2 + \ell(\mathbf{x}_t^\intercal \theta, u) \right] \tag{4}$$

where $\ell(\mathbf{x}, y) = -y \log \mu(\mathbf{x}) - (1-y) \log(1 - \mu(\mathbf{x}))$ is the cross entropy loss and $\eta = (2 + diam(\Theta_t))^{-1}$. When the inequality in 2 holds, θ_t, $\mathbf{W}_t$ and $\mathbf{W}_t^i$ $(i \in [N])$ are updated as described in *Steps 4-6* (Algorithm 2). First, in *Step 4*, $\boldsymbol{\theta}_{t+1}$ is computed by solving the following optimization problem up to a precision of $1/t$.

$$\theta_{t+1} = \arg\min_{\Theta_t} \left[\eta \|\theta - \theta_t\|_{\mathbf{W}_t}^2 + \ell(\mathbf{x}_t^\intercal \theta, y_t) \right] \tag{5}$$

Following this, $\mathbf{W}_t^i$ $(i \in [N])$ and $\mathbf{W}_t$ are updated in *Step 5* and *Step 6* as per their definitions provided earlier. When the inequality in 2 does not hold, $\mathcal{H}_t$ and Θ_t are updated as described in *Steps 9-12* (Algorithm 2). In *Step 9*, since the inequality criterion failed, $\mathcal{H}_t$ is updated to $\mathcal{H}_{t+1}$ by appending the pair $(\mathbf{x}_t, y_t)$ to it. Using $\mathcal{H}_{t+1}$, in *Step 10*, another estimate $\boldsymbol{\theta}_{t+1}^{\mathcal{H}}$ of $\theta^\star$ is computed by minimizing the regularized cross-entropy loss (up to a precision $1/t$).

$$\theta_{t+1}^{\mathcal{H}} = \arg\min \sum_{(\mathbf{x}, y) \in \mathcal{H}_{t+1}} \ell(\mathbf{x}^\intercal \theta, r) + \gamma_t(\delta) \|\theta\|_2^2 \tag{6}$$

Using this estimate, and a design matrix $\mathbf{V}_t^{\mathcal{H}}$ computed in *Step 11*, in *Step 12* the set Θ_t is updated to Θ_{t+1} by taking an intersection between a confidence set of radius $\beta_t(\delta) = O(dN \log(t/\delta))$ around the new estimate $\boldsymbol{\theta}_{t+1}^{\mathcal{H}}$ (that contains $\theta^\star$ with probability $1 - \delta$) and the initial set $\Theta_1 = \{\|\boldsymbol{\theta}\|_2 \leq S\}$. In Lemma 8, Faury et al. [2022] show that $|\mathcal{H}_T| = \tilde{O}(\kappa dNS^6)$. The rounds corresponding to $\mathcal{H}_T$, therefore, incur at most $\tilde{O}(\kappa dNS^6)$ regret.

In Theorem 3.1, we provide a regret guarantee for `Slate-GLM-OFU` and present its proof in Appendix B.1.

Theorem 3.1 (Regret of `Slate-GLM-OFU`)**.** *Let $\mathcal{T}$ denote the set of rounds until round T where the inequality condition in 2 fails, i.e., $\mathcal{T} = \{s \in [T] : (\mathbf{x}_s, y_s) \in \mathcal{H}_T\}$. Let $\mathbf{x}_{\star,t} = \arg\max_{\mathbf{x} \in \mathcal{X}_t} \mu(\mathbf{x}^\top \boldsymbol{\theta}^\star)$, be the optimal slate at round $t \in [T]$. Under the diversity assumption (Assumption 2.1), at the end of T rounds, with probability at least $1 - 6\delta$, the regret $R(T)$ of `Slate-GLM-OFU` satisfies,*

$$R(T) = \tilde{O}\left(SdN \sqrt{\sum_{t \notin \mathcal{T}} \dot{\mu}(\mathbf{x}_{\star,t}^\top \boldsymbol{\theta}^\star)} + S^6 d^2 N^2 \kappa \right)$$

Remark: Let $\mathcal{T}$ be as defined in Theorem 3.1. The per-round time complexity of `Slate-GLM-OFU` is $O((dN \log t)^2)$ for rounds $t \in [T] \setminus \mathcal{T}$ and it is $O(Ndt)$ for rounds $t \in \mathcal{T}$. Lemma 8 in Faury et al. [2022] implies that $|\mathcal{T}| = O(\kappa dNS^6)$. Thus, the $O(Ndt)$ per-round complexity is incurred for only these many rounds.

4 SLATE-GLM-TS

In this section, we present our second algorithm, Slate-GLM-TS (Algorithm 3) based on the Thompson Sampling paradigm [Thompson, 1933, Russo et al., 2018] used in bandit algorithms. Slate-GLM-TS builds upon the TS-ECOLog algorithm (Algorithm 3 in Appendix $D.2$, Faury et al. [2022]) while adapting to the changing action sets using the update strategy in Algorithm 2. TS-ECOLog adapts the Linear Thompson Sampling algorithm from Abeille and Lazaric [2017] (Figure 1 in Abeille and Lazaric [2017]) that perturbs the estimated parameter vector by adding an appropriately transformed noise vector sampled from a suitable multivariate distribution $\mathcal{D}^{TS}$ satisfying some nice properties (See Definition 1 of Abeille and Lazaric [2017]). Following this, the optimal action (slate in our case) with respect to the new perturbed parameter vector is chosen. While TS-ECOLog also achieves an optimal $O(\sqrt{T})$ regret guarantee for logistic reward models (for fixed action sets), similar to ada-OFU-ECOLog it also incurs per round computational cost proportional to the number of actions K (recall $K = 2^{\Omega(N)}$ in our setting) due to its selection at the slate level. To circumvent this, Slate-GLM-TS operates at the slot level and for each slot $i \in [N]$, it perturbs the components of the estimated parameter vector (corresponding to the ith slot) using a noise vector sampled independently of all other slots. This is followed by selecting the optimal items for each slot independently, thereby incurring an $N^{O(1)}$ per round time complexity in choosing the slate. While the items for each slot are independently determined, similar to Slate-GLM-OFU, Slate-GLM-TS also estimates a single reward model and updates the parameter vector for this model jointly using the slate level reward y_t, by employing the update strategy in Algorithm 2.

Input to Slate-GLM-TS are T, δ, S and $\mathcal{D}^{TS}$, where T is the time horizon i.e., the total number of rounds, δ is the error probability, S is a known upper bound for $\|\boldsymbol{\theta}^\star\|_2$ and $\mathcal{D}^{TS}$ is a multivariate distribution satisfying properties in Definition 1 in Abeille and Lazaric [2017]. During the course of the algorithm, Slate-GLM-TS maintains vectors $\boldsymbol{\theta}_t$, matrices $\mathbf{W}_t$, $\mathbf{W}_t^i$ ($i \in [N]$) and sets $\Theta_t, \mathcal{H}_t$ with exactly the same definition as in Slate-GLM-OFU.

Next, we go through the steps of Slate-GLM-TS (Algorithm 3). *Steps 3-10* is where Slate-GLM-TS differs significantly from TS-ECOLog. Instead of getting the set of arm features $\mathcal{X}_t$ (slates in our case) directly from the environment (as in TS-ECOLog), Slate-GLM-TS receives N different sets of items $\mathcal{X}_t^i, i \in [N]$ in *Step 4*. While TS-ECOLog samples one noise vector $\eta \in \mathbb{R}^{dN}$ from $\mathcal{D}^{TS}$ and perturbs the estimated parameter vector $\boldsymbol{\theta}_t$ by adding (a scalar multiple of) $(\boldsymbol{W}_t)^{-1/2}\boldsymbol{\eta}$, Slate-GLM-TS independently samples N such vectors $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_N$ and perturbs the components $\boldsymbol{\theta}_t^i$ of $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_t^1, \ldots, \boldsymbol{\theta}_t^N)$ (corresponding to the item features in the ith slot) to $\tilde{\boldsymbol{\theta}}_t^i \in \mathbb{R}^d$ by adding to it (a scalar multiple of) $(\boldsymbol{W}_t^i)^{-1/2}\boldsymbol{\eta}_i$ (*Step 7* and *8*). The algorithm continues to sample these noise vectors until the perturbed vector $\tilde{\boldsymbol{\theta}}_t = (\tilde{\boldsymbol{\theta}}_t^1, \ldots, \tilde{\boldsymbol{\theta}}_t^N)$ belongs to the admissible set Θ_t. Once this happens, in *Step 11*, it picks the item $\mathbf{x}_t^i \in \mathcal{X}_t^i$, which is optimal with respect to the perturbed parameter vector $\tilde{\boldsymbol{\theta}}_t^i$. Note that, the underlying optimization problem for slot i, only requires the candidate items in $\mathcal{X}_t^i$ and the perturbed vectors $\tilde{\theta}_t^i$, and thus, can be solved independently and in parallel for all slots.

In *Step 12*, we select the slate $\mathbf{x}_t = (\mathbf{x}_t^1, \ldots, \mathbf{x}_t^N)$, yielding a reward y_t. At this point, Slate-GLM-OFU calls a sub-routine described in Algorithm 2 which performs updates to $\boldsymbol{\theta}_t$, $\mathbf{W}_t$, $(\mathbf{W}_t^1, \ldots, \mathbf{W}_t^N)$, Θ_t, $\mathcal{H}_t$. We make a few additional remarks about Slate-GLM-TS below.

Algorithm 3 Slate-GLM-TS

1: **Inputs:** $T, \delta, S, \mathcal{D}^{TS}$
2: Initialize $\mathbf{W}_1^1 = \ldots = \mathbf{W}_1^N = \mathbf{I}_d$, $\mathbf{W}_1 = I_{dN}$, $\Theta_1 = \{\|\theta\|_2 \leq S\}$, $\theta_1 \in \Theta_1$, $\eta_t(\delta) = O(S^2 N d \log(t/\delta))$, and $\mathcal{H}_1 = \emptyset$
3: **for** each round $t \in [T]$ **do**
4: Obtain the set of items $\mathcal{X}_t^i, \forall i \in [N]$
5: Set reject = True
6: **while** reject **do**
7: Sample $\eta^1, \ldots, \eta^N \overset{\text{iid}}{\sim} \mathcal{D}^{TS}$
8: Define $\tilde{\theta}_t^i = \theta_t^i + \eta_t(\delta)(\mathbf{W}_t^i)^{-1/2}\eta^i, \forall i \in [N]$
9: If $\tilde{\theta}_t = (\tilde{\theta}_t^1, \ldots, \tilde{\theta}_t^N) \in \Theta_t$, reject = False
10: **end while**
11: For each $i \in [N]$, find item $\mathbf{x}_t^i = \arg\max_{\mathbf{x} \in \mathcal{X}_t^i} \mathbf{x}^\intercal \tilde{\theta}_t^i$
12: Select slate $\mathbf{x}_t = (\mathbf{x}_t^1, \ldots, \mathbf{x}_t^N)$ and get reward y_t
13: Obtain $\boldsymbol{\theta}_{t+1}$, $\mathbf{W}_{t+1}$, $(\mathbf{W}_{t+1}^1, \ldots, \mathbf{W}_{t+1}^N)$, Θ_{t+1}, $\mathcal{H}_{t+1}$ by calling Algorithm 2 with inputs $\mathbf{x}_t$, y_t, $\boldsymbol{\theta}_t$, $\mathbf{W}_t$, $(\mathbf{W}_t^1, \ldots, \mathbf{W}_t^N)$, Θ_t, $\mathcal{H}_t$
14: **end for**

Remark: It's easy to see that the per round time complexity of Slate-GLM-TS is $N(d \log T)^{O(1)}$. This is significantly lower than that of TS-ECOLog which runs in time exponential in N. The improvement comes as a result of the slot-level selection in Slate-GLM-TS. This along with the efficient estimation of $\boldsymbol{\theta}_t$ in Algorithm 2, ensures that the algorithm has low per-round time complexity making it useful in practical scenarios. This is validated by our Synthetic and Real-World experiments in Section 5. We also observe that in almost all experiments we performed, the regret of Slate-GLM-TS was quite competitive and better than most baselines. Even though we do not provide a theoretical guarantee for the regret of Slate-GLM-TS, in Appendix C.1, we provide a fixed-arms version of Slate-GLM-TS called Slate-GLM-TS-Fixed which operates in the non-contextual setting, like TS-ECOLog i.e., the action (slate) features do not change over time. It

uses the short warm-up procedure from TS-ECOLog and the slot-level selection technique from Slate-GLM-TS resulting in a per round time complexity linear in N. By utilizing the multiplicative equivalence of $\boldsymbol{W}_t$ and $diag(\boldsymbol{W}_t^1, \ldots, \boldsymbol{W}_t^N)$ that we showed in the proof of Theorem 3.1 (using the diversity assumption (Assumption 2.1)), and adapting the proof of TS-ECOLog (Theorem 5, Faury et al. [2022]), we prove an optimal ($O(\sqrt{T})$) dependence on the number of rounds T. For brevity, we discuss details of Slate-GLM-TS-Fixed (Algorithm 4) and its regret guarantee (Theorem C.1) in Appendix C.1.

5 EXPERIMENTS

In this section, we perform a wide range of synthetic (**Experiments 1,2,3**) and real-world experiments (**Experiment 4**) to demonstrate the empirical performance of our algorithms Slate-GLM-OFU, Slate-GLM-TS and Slate-GLM-TS-Fixed. Details of each experiment are in the respective paragraphs.[5]

Experiment 1 ($R(T)$ vs. T, Contextual Setting): In this experiment, we compare our algorithms Slate-GLM-OFU and Slate-GLM-TS to their counterparts ada-OFU-ECOLog (Algorithm 2, Faury et al. [2022]) and TS-ECOLog (Section D.2, Faury et al. [2022]). These are the only logistic bandit algorithms that achieve optimal ($\kappa-$free) regret and are also computationally efficient ($O((\log T)^2)$ per round time complexity). We perform experiments for the following two settings.

Finite Contexts: We assume the contexts come from the set $\mathcal{C} = \{1, \ldots, C\}$. For each $c \in \mathcal{C}$ and $i \in [N]$, a set of items $\mathcal{X}^{i,c}$ is constructed before hand by randomly sampling K vectors from the $d-$dimensional ball with radius $1/\sqrt{N}$. At each round t, a context c is sampled uniformly at random from $\mathcal{C}$ and the sets $\mathcal{X}^{1,c}, \ldots, \mathcal{X}^{N,c}$ are presented to the learner.

Infinite Contexts: At each round $t \in [T]$, and for each slot $i \in [N]$, set $\mathcal{X}_t^i$ is constructed by sampling K vectors randomly from the $d-$dimensional ball with radius $1/\sqrt{N}$. The learner is then presented with $\mathcal{X}_t^i$.

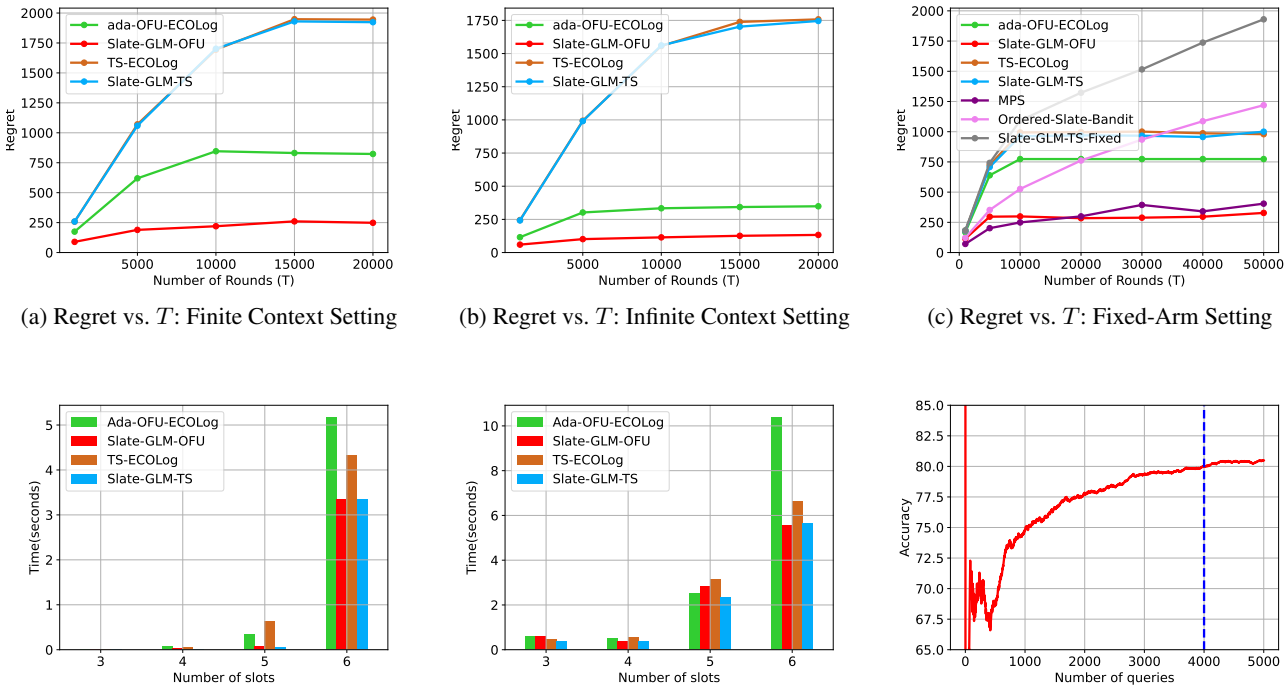
For the finite context setting, we fix $C = 5$. For both settings, we fix the number of slots $N = 3$, the number of items per slot $K = 5$, and the dimension of item features to $d = 5$. To simulate the reward, we select $\boldsymbol{\theta}^\star$ by randomly sampling from $[-1, 1]^{15}$. We run our algorithms by varying the time horizon T in $\{1000, 5000, 10000, 15000, 20000\}$. For each T, we average the regret obtained at the end of T rounds over 20 diferent seeds used to sample the rewards. The results for the Finite and Infinite context settings are shown in Figures 1a and 1b respectively. We can see that in both instances, Slate-GLM-OFU performs the best, while Slate-GLM-TS performs on par with TS-ECOLog. Further, in Section F of the appendix, we report the average results along with two standard deviations.

Experiment 2 (Per-Round Time vs. N): In this experiment, we compare the average and maximum time taken (per round) by our algorithms Slate-GLM-OFU and Slate-GLM-TS, with respect to their counterparts ada-OFU-ECOLog and TS-ECOLog [Faury et al., 2022] respectively[6]. While doing this comparison, we vary the number of slots N in the set $\{3, \ldots, 6\}$. The number of items ($K = |\mathcal{X}_t^i|$) per slot is fixed to 7 and the dimension d of each item is fixed to 5. The item features are selected by randomly sampling from $[-1, 1]^5$ and normalized to have norm $1/\sqrt{N}$. For each $N \in \{3, 4, 5, 6\}$, we select a different reward parameter vector $\boldsymbol{\theta}^\star$ by randomly sampling from $[-1, 1]^{5N}$. Note that the number of possible slates is K^N and thus, varying N in $\{3, 4, 5, 6\}$ results in 343, 2401, 16807, and 117649 slates respectively. We perform this experiment in the infinite context setting (See **Experiment 1** for details). We run all the algorithms for $T = 1000$ rounds and average the results over 10 different seeds for sampling rewards. We the average per round running time in Figure 1d and maximum per round running time in Figure 1e. As expected, we observe much lower running times for Slate-GLM-OFU and Slate-GLM-TS compared to their counterparts. Moreover, the plots also indicate exponential growth in the per-round running time for both ada-OFU-ECOLog and TS-ECOLog. Further, there is a significant gap between the maximum and average per-round time of Slate-GLM-OFU and Slate-GLM-TS, implying that the actual per-round time for these algorithms is generally much lower than their maximum values. In Section F of the appendix, we report the results with two standard deviations, along with each algorithm's average time for choosing an arm to pull and updating its parameters speerately.

Experiment 3 ($R(T)$ vs. T, Non-Contextual Setting): In this experiment, we compare our algorithms Slate-GLM-OFU, Slate-GLM-TS, and Slate-GLM-TS-Fixed (Algorithm 4, Appendix C) to a number of state-of-the-art baseline algorithms, in the non-contextual setting, i.e., the set of candidate slates remains fixed throughout the course of the algorithm. Like previous experiments, our baselines include ada-OFU-ECOLog and TS-ECOLog from Faury et al. [2022]. However, for the non-contextual setting, we also include other state-of-the-art baselines such as the MPS algorithm (Algorithm 3, Dimakopoulou et al. [2019]) and the Ordered Slate Bandit algorithm (Figure 3, Kale et al. [2010]). The

[5]The codes for the experiments can be found at https://github.com/tanmaygoyal258/Logistic_Slate_Bandits.git and https://github.com/tanmaygoyal258/Prompt_Optimization_Slate_Bandits.git

[6]The per-round time is calculated as the sum of the per-round pull and per-round update times.



(a) Regret vs. T: Finite Context Setting (b) Regret vs. T: Infinite Context Setting (c) Regret vs. T: Fixed-Arm Setting

(d) Average running time (per-round) (e) Maximum running time (per-round) (f) Accuracy vs. T: Prompt Optimization

Figure 1

latter is designed for semi-bandit feedback, and hence, we adapt it to the bandit feedback setting as explained in Appendix F. We fix the number of slots N to 3 and the number of items in each slot $K = |\mathcal{X}_t^i|$ to 5. The dimension d of items for each slot is fixed to 5. The items for each slot are randomly sampled from $[-1, 1]^5$ and normalized to have norm $1/\sqrt{3}$, while $\boldsymbol{\theta}_\star$ is randomly sampled from $[-1, 1]^{15}$ and normalized. We run all the algorithms for $T \in \{1000, 5000, 10000, 20000, 30000, 40000, 50000\}$ rounds and average the results over 50 different seeds for sampling rewards. The rewards are shown in Figure 1c. We see that `Slate-GLM-OFU` has the best performance, with the only algorithm having comparable performance being `MPS`. Also, `Slate-GLM-TS` performs worse than `ada-OFU-ECOLog` and `MPS` while being on par with `TS-ECOLog`. In Section F, we showcase the average results with two standard deviations, which also demonstrates that `MPS` showcases a high variance in results, hence, being less reliable in practice.

Experiment 4 (Prompt Tuning): In this experiment, we apply our contextual slate bandit algorithm `Slate-GLM-OFU` to select in-context examples for tuning prompts of Language Models, applied to binary classification tasks. Typically, for such applications, a labeled training set of (input query, output label) pairs is used to learn policies of editing different parts of the prompt (instruction, in-context examples, verbalizers) [Zhang et al., 2022] depending on a provided test input query. To simplify our task, we fix the instruction and the verbalizer and only select N in-context examples from an available pool of K examples. There are N available positions (slots) in the prompt. Given a test input query (context), we create context-dependent features for the K pool examples and independently select one (with repetition) per slot. This matches the contextual slate bandit problem setting (See Section 2) and therefore `Slate-GLM-OFU` can be applied. We experiment on a sampled subset of size 5000 from two popular sentiment analysis datasets, *SST2* and *Yelp Review*. We randomly order the set and use about $\sim 80\%$ (4128 for *SST2*, 4000 for *Yelp Review*) of them for "warm-up" training and the remaining 20% for testing. Like most prompt tuning experiments [Zhang et al., 2022], we report our results only on the test set, however, our algorithm continues to learn throughout the 5000 rounds. The warm-up rounds help us to start with a good estimate of the hidden reward parameter vector. We fix $N = 4$ and vary K in the set $\{8, 16, 32\}$. All the slots choose an example from the same K-sized example pool. At each round, given an input query $\boldsymbol{q}$ that needs to be solved for, item features for each in-context example $\boldsymbol{e} = (\boldsymbol{x}, y)$, is constructed by embedding each of $\boldsymbol{q}$, $\boldsymbol{x}$, and y into 64 dimensions [Nussbaum et al., 2024] and concatenating them, thereby resulting in a 192-dimensional item feature vector. After selecting the 4 items (slate), the resulting prompt (also containing the input query $\boldsymbol{q}$) is passed through the RoBERTa [Zhuang et al., 2021] model and a possible answer for $\boldsymbol{q}$ is generated. Hence, we are learning

to choose best the in-context examples for RoBERTa. At each round, we use GPT-3.5-Turbo to provide feedback (binary, 0 or 1) for the generated answer. This is treated as the reward for the chosen slate and utilized by the rest of the `Slate-GLM-OFU` algorithm. Figure 1f shows the increase in cumulative accuracy as we sequentially proceed through the 5000 data points in the *Yelp Review* dataset. The data points to the left of the dotted blue line are the warm-up points and those to the right are the test points. We can see that the cumulative accuracy increases consistently as we sequentially proceed through the points. Also, on the test set, the accuracy stays well above 80%. We vary K in the set $\{8, 16, 32\}$ and report test accuracy for both datasets in Table 1. It can be seen that the cumulative test accuracies for `Slate-GLM-OFU` are much higher compared to the Random Allocation baseline where each in-context example is chosen randomly and no learning is performed. Also, we see that the accuracy generally increases when the pool size increases since better examples can be available. We do see a small dip for the *Yelp Review* dataset when K increases from 16 to 32 and hypothesize that this may be happening due to more exploration.

Pool	**SST2**		**Yelp Review**	
Size	Random	`Slate-GLM-OFU`	Random	`Slate-GLM-OFU`
8	54.22	69.15	62.90	74.00
16	54.46	80.96	63.30	82.50
32	53.82	81.42	62.00	79.50

Table 1: Prompt Tuning Test Accuracy

6 CONCLUSIONS

We proposed three algorithms `Slate-GLM-OFU`, `Slate-GLM-TS`, `Slate-GLM-TS-Fixed` for the slate bandit problem with logistic rewards. While the first two work in both the contextual and non-contextual settings, the third is designed for the non-contextual setting. All our algorithms perform explore-exploit at the slot level, making their average per round time complexity logarithmic in the number of candidate slates. By building on algorithms from Faury et al. [2022], the average time per round is also logarithmic in the number of rounds T. As a result, our algorithms run much faster than state of the art logistic bandit algorithms (having $2^{\Omega(N)}$ per round time complexity). We also show that under a popular diversity assumption (Assumption 2.1), which we also empirically validate, `Slate-GLM-OFU` and `Slate-GLM-TS-Fixed` achieve κ independent $\tilde{O}(\sqrt{T})$ regret, making them both optimal and computationally efficient.

References

Yasin Abbasi-yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper_files/paper/2011/file/e1d5be1c7f2f456670de3d53c7b54f4a-Paper.pdf.

Marc Abeille and Alessandro Lazaric. Linear Thompson Sampling Revisited. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 176–184. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/abeille17a.html.

Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-wise minimax-optimal algorithms for logistic bandits, 2021. URL https://arxiv.org/abs/2010.12642.

Niladri S. Chatterji, Vidya Muthukumar, and Peter L. Bartlett. Osom: A simultaneously optimal algorithm for multi-armed and linear contextual bandits, 2020. URL https://arxiv.org/abs/1905.10040.

Jin Chen, Ju Xu, Gangwei Jiang, Tiezheng Ge, Zhiqiang Zhang, Defu Lian, and Kai Zheng. Automated creative optimization for e-commerce advertising. *Proceedings of the Web Conference 2021*, 2021. URL https://api.semanticscholar.org/CorpusID:232076065.

Nirjhar Das and Gaurav Sinha. *Linear Contextual Bandits with Hybrid Payoff: Revisited*, page 441–455. Springer Nature Switzerland, 2024. ISBN 9783031703652. doi: 10.1007/978-3-031-70365-2_26. URL http://dx.doi.org/10.1007/978-3-031-70365-2_26.

Maria Dimakopoulou, Nikos Vlassis, and Tony Jebara. Marginal posterior sampling for slate bandits. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 2223–2229. International Joint Conferences on Artificial Intelligence Organization, 7 2019. doi: 10.24963/ijcai.2019/308. URL https://doi.org/10.24963/ijcai.2019/308.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits, 2020. URL https://arxiv.org/abs/2002.07530.

Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clément Calauzènes. Jointly efficient and optimal algorithms for logistic bandits, 2022. URL https://arxiv.org/abs/2201.01985.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010. URL https://proceedings.neurips.cc/paper_files/paper/2010/file/c2626d850c80ea07e7511bbae4c76f4b-Paper.pdf.

Daniel N. Hill, Houssam Nassif, Yi Liu, Anand Iyer, and S.V.N. Vishwanathan. An efficient bandit algorithm for realtime multivariate optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17. ACM, August 2017. doi: 10.1145/3097983.3098184. URL http://dx.doi.org/10.1145/3097983.3098184.

Satyen Kale, Lev Reyzin, and Robert E Schapire. Non-stochastic bandit slate problems. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010. URL https://proceedings.neurips.cc/paper_files/paper/2010/file/390e982518a50e280d8e2b535462ec1f-Paper.pdf.

Haruka Kiyohara, Masahiro Nomura, and Yuta Saito. Off-policy evaluation of slate bandit policies via optimizing abstraction. In *Proceedings of the ACM Web Conference 2024*, WWW '24, page 3150–3161, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400701719. doi: 10.1145/3589334.3645343. URL https://doi.org/10.1145/3589334.3645343.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

Zach Nussbaum, John X. Morris, Brandon Duderstadt, and Andriy Mulyar. Nomic embed: Training a reproducible long context text embedder, 2024.

Matteo Papini, Andrea Tirinzoni, Marcello Restelli, Alessandro Lazaric, and Matteo Pirotta. Leveraging good representations in linear contextual bandits, 2021. URL https://arxiv.org/abs/2104.03781.

Jason Rhuggenaath, Alp Akcay, Yingqian Zhang, and Uzay Kaymak. Algorithms for slate bandits with non-separable reward functions, 04 2020.

Daniel J. Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, and Zheng Wen. A tutorial on thompson sampling. *Found. Trends Mach. Learn.*, 11(1):1–96, July 2018. ISSN 1935-8237. doi: 10.1561/2200000070. URL https://doi.org/10.1561/2200000070.

Ayush Sawarni, Nirjhar Das, Siddharth Barman, and Gaurav Sinha. Generalized linear bandits with limited adaptivity, 2024. URL https://arxiv.org/abs/2404.06831.

Adith Swaminathan, Akshay Krishnamurthy, Alekh Agarwal, Miro Dudik, John Langford, Damien Jose, and Imed Zitouni. Off-policy evaluation for slate recommendation. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/5352696a9ca3397beb79f116f3a33991-Paper.pdf.

William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933. ISSN 00063444. URL http://www.jstor.org/stable/2332286.

Joel Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16, 01 2011a. doi: 10.1214/ECP.v16-1624.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, August 2011b. ISSN 1615-3383. doi: 10.1007/s10208-011-9099-z. URL http://dx.doi.org/10.1007/s10208-011-9099-z.

Nikos Vlassis, Ashok Chandrashekar, Fernando Amat Gil, and Nathan Kallus. Control variates for slate off-policy evaluation. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9781713845393.

Tianjun Zhang, Xuezhi Wang, Denny Zhou, Dale Schuurmans, and Joseph E. Gonzalez. Tempera: Test-time prompting via reinforcement learning, 2022. URL https://arxiv.org/abs/2211.11890.

Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. A robustly optimized BERT pre-training approach with post-training. In Sheng Li, Maosong Sun, Yang Liu, Hua Wu, Kang Liu, Wanxiang Che, Shizhu He, and Gaoqi Rao, editors, *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–1227, Huhhot, China, August 2021. Chinese Information Processing Society of China. URL https://aclanthology.org/2021.ccl-1.108/.

Efficient Algorithms for Logistic Slate Contextual Bandits with Bandit Feedback (Supplementary Material)

Tanmay Goyal[1] **Gaurav Sinha**[1]

[1]Microsoft Research India

A GENERAL NOTATIONS AND RESULTS

This section presents some general notations and results for the logistic function that would be used throughout the Appendix. For a matrix $\boldsymbol{A}$, let $\lambda_{max}(\boldsymbol{A})$ and $\lambda_{min}(\boldsymbol{A})$ denote the maximum and minimum eigenvalue of $\boldsymbol{A}$ respectively. Similarly, we define $\sigma_{max}(\boldsymbol{A})$ and $\sigma_{min}(\boldsymbol{A})$ to be the maximum and minimum singular values respectively. We also define the following functions, borrowed from Faury et al. [2022]:

1. $\gamma_t(\delta) = O(S^2 N d \log(t/\delta))$
2. $\beta_t(\delta) = O(S^6 N d \log(t/\delta))$
3. $\eta_t(\delta) = O(S^2 N d \log(t/\delta))$

Claim A.1. *Let $\mu : \mathbb{R} \to \mathbb{R}$ be the logistic function, i.e., $\mu(x) = 1/(1+\exp(-x))$ and $\dot{\mu}, \ddot{\mu}$ be the first and second derivative of μ. The following are true.*

1. $|\ddot{\mu}(x)| \leq \dot{\mu}(x)$, $\forall x \in \mathbb{R}$
2. $\dot{\mu}(x) \leq \dot{\mu}(y)\exp(|x-y|)$, $\forall x, y \in \mathbb{R}$

Definition A.1. Let $\dot{\mu}$ be the derivative of the logistic function. Define functions $\alpha : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $\tilde{\alpha} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ as follows.

1. $\alpha(x, y) = \int_0^1 \dot{\mu}(x + v(y-x))\,\mathrm{d}v$
2. $\tilde{\alpha}(x, y) = \int_0^1 (1-v)\dot{\mu}(x + v(y-x))\,\mathrm{d}v$

Definition A.2. (Exact Taylor Expansion for the Logistic Function) The logistic function $\mu(x)$ can be expanded using an Exact Taylor Expansion as follows:

$$\mu(x) = \mu(y) + \dot{\mu}(y)(x-y) + \int_0^1 (1-v)\ddot{\mu}(x + v(y-x))\,\mathrm{d}v(x-y)^2$$

Definition A.3. (Mean Value Theorem for the Logistic Function) The logistic function μ can be expanded using the Mean Value theorem as follows:

$$\mu(x) = \mu(y) + \alpha(x, y)(x-y)$$

Recall the following notations from Section 3:

1. $\boldsymbol{W}_t = \boldsymbol{I} + \sum_{s=1}^{t-1} \dot{\mu}(\boldsymbol{x}_s^\top \boldsymbol{\theta}_{s+1}) \boldsymbol{x}_s \boldsymbol{x}_s^\top$
2. $\boldsymbol{W}_t^i = \boldsymbol{I} + \sum_{s=1}^{t-1} \dot{\mu}(\boldsymbol{x}_s^\top \boldsymbol{\theta}_{s+1}) \boldsymbol{x}_s^i \boldsymbol{x}_s^{i\top}$
3. $\boldsymbol{V}_t^{\mathcal{H}} = \gamma_t(\delta)\boldsymbol{I} + \sum_{\boldsymbol{x} \in \mathcal{H}_t} \boldsymbol{x}\boldsymbol{x}^\top / \kappa$

We define the following additional matrices.

1. $\boldsymbol{U}_t = \text{diag}(\boldsymbol{W}_t^1, \ldots, \boldsymbol{W}_t^N)$
2. $\boldsymbol{W}_t^{i,j} = \sum_{s=1}^{t-1} \dot{\mu}(\boldsymbol{x}_s^\top \boldsymbol{\theta}_{s+1}) \boldsymbol{x}_s^i \boldsymbol{x}_s^{j\top}$
3. $\boldsymbol{V}_t^{\mathcal{H},i} = \gamma_t(\delta)\boldsymbol{I} + \sum_{\boldsymbol{x} \in \mathcal{H}_t} \boldsymbol{x}^i \boldsymbol{x}^{i\top} / \kappa$
4. $\boldsymbol{V}_t^{\mathcal{H},i,j} = \gamma_t(\delta)\boldsymbol{I} + \sum_{\boldsymbol{x} \in \mathcal{H}_t} \boldsymbol{x}^i \boldsymbol{x}^{j\top} / \kappa$
5. $\boldsymbol{U}_t^{\mathcal{H}} = \text{diag}(\boldsymbol{V}_t^{\mathcal{H},1}, \ldots, \boldsymbol{V}_t^{\mathcal{H},N})$

B SLATE-GLM-OFU

Let $\boldsymbol{x}^i \in \mathbb{R}^d$, we define the "*lift*" $\tilde{\boldsymbol{x}}^i \in \mathbb{R}^{dN}$, of $\boldsymbol{x}^i$ as follows,

$$\tilde{\boldsymbol{x}}^i(j) = \begin{cases} 0 & \text{if } j \notin [(i-1)d, id-1] \\ \boldsymbol{x}(j-(i-1)d) & \text{otherwise} \end{cases}$$

In other words, consider $\tilde{\boldsymbol{x}}^i$ to be a vector with N slots of dimension d, such that the ith slot is $\boldsymbol{x}^i$ while the rest of the slots are assigned the zero vector. Then, for any vector $\boldsymbol{z} = (\boldsymbol{z}^1, \ldots, \boldsymbol{z}^N) \in \mathbb{R}^{dN}$, with $\boldsymbol{z}^i \in \mathbb{R}^d$, $\forall i \in [N]$, we get that $\boldsymbol{z} = \tilde{\boldsymbol{z}}^1 + \ldots + \tilde{\boldsymbol{z}}^N$.

Let $T_0 \in \mathbb{N}$ be a constant (depending on N and ρ) such that $\forall t \geq T_0, t \geq \frac{3+2\rho N}{3\rho^2}(N-1)^2 \log\left(\frac{2dNT}{\delta}\right)$. We assume that the total rounds T satisfies $T \geq T_0$.

We now prove that the regret for `Slate-GLM-OFU` can be bounded above by the quantity mentioned in Theorem 3.1 (restated and expanded below). Define the following events:

$$\mathcal{E}_1 = \left\{ \forall i, j \in [N], i \neq j, \forall t \in [T] : \left\| \boldsymbol{W}_t^{i,j} \right\| \leq \sqrt{\frac{t}{2N^2} \log\left(\frac{dN(N-1)}{\delta}\right)} \text{and} \left\| \boldsymbol{V}_t^{\mathcal{H},i,j} \right\| \leq \sqrt{\frac{8t}{\kappa^2 N^2} \log\left(\frac{dN(N-1)}{\delta}\right)} \right\}$$

$$\mathcal{E}_2 = \left\{ \forall i \in [N], \forall t \in [T_0, T] : \lambda_{min}\left(\boldsymbol{V}_t^i\right) \geq 1 + \frac{\rho t}{2} \text{ and } \lambda_{min}\left(\boldsymbol{V}_t^{\mathcal{H},i}\right) \geq \gamma_t(\delta) + \frac{\rho t}{2} \right\}$$

$$\mathcal{E}_3 = \left\{ \forall t \in [T], \|\boldsymbol{\theta}^\star - \boldsymbol{\theta}_{t+1}\|_{\boldsymbol{W}_{t+1}}^2 \leq CS^2 d \log(t/\delta) \text{ and } \boldsymbol{\theta}^\star \in \Theta \right\}$$

$$\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$$

Theorem B.1 (Regret of `Slate-GLM-OFU`)**.** *At the end of $T(\geq T_0)$ rounds and assuming event $\mathcal{E}$ holds, the regret of `Slate-GLM-OFU` is bounded by*

$$Regret(T) \leq T_0 + CSNd^{1/2} \sqrt{\left(d\log(T/4N) + \frac{1}{2\rho}\log T\right)\log(T/\delta)} \sqrt{\sum_{t \notin \mathcal{T}} \dot{\mu}\left(\boldsymbol{x}_t^{\star\top}\boldsymbol{\theta}^\star\right)}$$
$$+ (1+\kappa)CS^2N^2 d\log(T/\delta)\kappa\left(d\log(T/4N) + \frac{1}{\rho}\log(T)\right) + CS^6N^2d^2\kappa\log(T/\delta)\log(T/\kappa N)$$

Proof. Recall from Section 3 that $\mathcal{T}$ is the set of all rounds in $[T]$, where the inequality condition in *Step 2* of Algorithm 2 does not hold. Using the bound on $|\mathcal{T}|$ provided in Lemma B.15, we get that,

$$Regret(T) \leq |\mathcal{T}| + \sum_{t \notin \mathcal{T}} \mu(\boldsymbol{x}_t^{\star\top}\boldsymbol{\theta}^\star) - \mu(\boldsymbol{x}_t^\top \boldsymbol{\theta}^\star) \leq CS^6 d^2 \kappa \log(T/\delta)\log(T/\kappa N) + R(T)$$

where $\boldsymbol{x}_t^\star = \arg\max_{\boldsymbol{x}\in\mathcal{X}_t} \mu(\boldsymbol{x}^\top\boldsymbol{\theta}^\star)$ and $R(T) = \sum_{t\notin\mathcal{T}} \mu(\boldsymbol{x}_t^{\star\top}\boldsymbol{\theta}^\star) - \mu(\boldsymbol{x}_t^\top\boldsymbol{\theta}^\star)$.

Now, recall from event $\mathcal{E}$ that all our *good events* are defined for $t \in [T_0, T]$ (where T_0 is some constant in N and ρ) . Hence, for rounds $t \leq T_0$, we can trivially bound the regret as T_0.

Now, we shift our attention to $t \in [T_0, T]$. **From here on, we assume that $\mathbf{t} \in [\mathbf{T_0}, \mathbf{T}]$.**

Now, expanding $R(T)$ using an exact Taylor expansion (Definition A.2) along with the fact that $|\ddot{\mu}(.)| \leq \dot{\mu}(.)$ gives us,

$$R(T) \leq \sum_{t\notin\mathcal{T}} \dot{\mu}(\boldsymbol{x}_t^\top\boldsymbol{\theta}^\star)(\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\top\boldsymbol{\theta}^\star + \sum_{t\notin\mathcal{T}} \tilde{\alpha}(\boldsymbol{x}_t^{\star\top}\boldsymbol{\theta}^\star, \boldsymbol{x}_t^\top\boldsymbol{\theta}^\star)((\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\top\boldsymbol{\theta}^\star)^2$$

So we bound $R(T)$ by bounding the two quantities $R_1(T) = \sum_{t\notin\mathcal{T}} \dot{\mu}(\boldsymbol{x}_t^\top\boldsymbol{\theta}^\star)(\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\top\boldsymbol{\theta}^\star$ and $R_2(T) = \sum_{t\notin\mathcal{T}} \tilde{\alpha}(\boldsymbol{x}_t^{\star\top}\boldsymbol{\theta}^\star, \boldsymbol{x}_t^\top\boldsymbol{\theta}^\star)((\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\top\boldsymbol{\theta}^\star)^2$ separately.

<u>Bounding $R_1(T)$:</u> To bound $R_1(T)$, we define $\mathcal{T}_1 = \{t \in [T_0, T] : t \notin \mathcal{T} \text{ and } \dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star\right) \geq \dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right)\}$ and $\mathcal{T}_2 = \{t \in [T_0, T] : t \notin \mathcal{T} \text{ and } \dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star\right) \leq \dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right)\}$. Note that, $\mathcal{T}_1 \cap \mathcal{T}_2 = \phi$, and $[T_0, T] \setminus \mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$. Bu summing over rounds in $\mathcal{T}_1$ we obtain,

$$\sum_{t\in\mathcal{T}_1} \dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star\right)\left(\boldsymbol{x}_t^\star - \boldsymbol{x}_t\right)^\intercal\boldsymbol{\theta}^\star \overset{(i)}{=} \sum_{t\in\mathcal{T}_1} \left[\dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right) + \ddot{\mu}\left(z_t\right)\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star - \boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right)\right]\left(\boldsymbol{x}_t^\star - \boldsymbol{x}_t\right)^\intercal\boldsymbol{\theta}^\star$$

for some z_t between $\boldsymbol{x}_t^\top\boldsymbol{\theta}^\star$ and $\boldsymbol{x}_t^\top\boldsymbol{\theta}_{t+1}$. Here, (i) follows from the mean value theorem. Let $R_1(T)_1 = \sum_{t\in\mathcal{T}_1} \dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right)\left[\left(\boldsymbol{x}_t^\star - \boldsymbol{x}_t\right)^\intercal\boldsymbol{\theta}^\star\right]$ and $R_1(T)_2 = \sum_{t\in\mathcal{T}_1} \ddot{\mu}(z)\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star - \boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right)\left(\boldsymbol{x}_t^\star - \boldsymbol{x}_t\right)^\intercal\boldsymbol{\theta}^\star$. We bound these separately.

$$
\begin{aligned}
R_1(T)_1 &= \sum_{t\in\mathcal{T}_1} \dot{\mu}(\boldsymbol{x}_t^\top \boldsymbol{\theta}_{t+1}) \left[(\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\top \boldsymbol{\theta}^\star \right] \leq \sum_{t\in\mathcal{T}_1} \dot{\mu}(\boldsymbol{x}_t^\top \boldsymbol{\theta}_{t+1}) \dot{\mu}(z_t) (\boldsymbol{x}_t^{\star\top} \boldsymbol{\theta}^\star - \boldsymbol{x}_t^\top \boldsymbol{\theta}^\star) \\
&\overset{(i)}{\leq} \sum_{t\in\mathcal{T}_1} \dot{\mu}(\boldsymbol{x}_t^\top \boldsymbol{\theta}_{t+1}) \left\{ |\boldsymbol{x}_t^{\star\top} \boldsymbol{\theta}^\star - \boldsymbol{x}_t^{\star\top} \boldsymbol{\theta}_t| + |\boldsymbol{x}_t^\top \boldsymbol{\theta}^\star - \boldsymbol{x}_t^\top \boldsymbol{\theta}_t| + \boldsymbol{x}_t^{\star\top} \boldsymbol{\theta}_t - \boldsymbol{x}_t^\top \boldsymbol{\theta}_t \right\} \\
&\overset{(ii)}{\leq} \sum_{t\in\mathcal{T}_1} \dot{\mu}(\boldsymbol{x}_t^\top \boldsymbol{\theta}_{t+1}) \left\{ \|\boldsymbol{x}_t^\star\|_{\boldsymbol{W}_t^{-1}} \sqrt{\eta_t(\delta)} + \|\boldsymbol{x}_t\|_{\boldsymbol{W}_t^{-1}} \sqrt{\eta_t(\delta)} + \sum_{i=1}^N \left(\tilde{\boldsymbol{x}}_t^{\star,i} - \tilde{\boldsymbol{x}}_t^i \right)^\top \boldsymbol{\theta}_t \right\} \\
&\overset{(iii)}{\leq} \sum_{t\in\mathcal{T}_1} \dot{\mu}(\boldsymbol{x}_t^\top \boldsymbol{\theta}_{t+1}) \left\{ \sum_{i=1}^N \sqrt{\eta_t(\delta)} \left(\|\boldsymbol{x}_t^{\star,i}\|_{(\boldsymbol{W}_t^i)^{-1}} + \|\boldsymbol{x}_t^i\|_{(\boldsymbol{W}_t^i)^{-1}} \right) + \sum_{i=1}^N \left(\tilde{\boldsymbol{x}}_t^{\star,i} - \tilde{\boldsymbol{x}}_t^i \right)^\top \boldsymbol{\theta}_t^i \right\} \\
&\overset{(iv)}{\leq} \sum_{t\in\mathcal{T}_1} \dot{\mu}(\boldsymbol{x}_t^\top \boldsymbol{\theta}_{t+1}) \left\{ \sum_{i=1}^N \sqrt{\eta_t(\delta)} \left(\|\boldsymbol{x}_t^{\star,i}\|_{(\boldsymbol{W}_t^i)^{-1}} + \|\boldsymbol{x}_t^i\|_{(\boldsymbol{W}_t^i)^{-1}} \right) + \sum_{i=1}^N \left(\sqrt{\eta_t(\delta)} \|\boldsymbol{x}_t^i\|_{(\boldsymbol{W}_t^i)^{-1}} - \sqrt{\eta_t(\delta)} \|\boldsymbol{x}_t^{\star,i}\|_{(\boldsymbol{W}_t^i)^{-1}} \right) \right\} \\
&\leq C\sqrt{\eta_T(\delta)} \sum_{t\in\mathcal{T}_1} \dot{\mu}(\boldsymbol{x}_t^\top \boldsymbol{\theta}_{t+1}) \sum_{i=1}^N 2\|\boldsymbol{x}_t^i\|_{(\boldsymbol{W}_t^i)^{-1}} \leq C\sqrt{\eta_T(\delta)} \sqrt{\sum_{t\in\mathcal{T}_1} \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \sqrt{\sum_{t\in\mathcal{T}_1} \left(\sum_{i=1}^N \sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}} \right)^2} \\
&\overset{(v)}{\leq} C\sqrt{\eta_T(\delta)} \sqrt{\sum_{t\in\mathcal{T}_1} \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \sqrt{Nd\log(T/4N) + M(T)} \overset{(vi)}{\leq} C\sqrt{\eta_T(\delta)} \sqrt{Nd\log(T/4N) + M(T)} \sqrt{\sum_{t\in\mathcal{T}_1} \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star\right)} \\
&\overset{(vii)}{\leq} C\sqrt{\eta_T(\delta)} \sqrt{Nd\log(T/4N) + M(T)} \left(\sqrt{R(T)} + \sqrt{\sum_{t\notin\mathcal{T}} \dot{\mu}\left(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star\right)} \right) \\
&\overset{(viii)}{\leq} CSN^{1/2} d^{1/2} \sqrt{Nd\log(T/4N) + M(T)} \sqrt{\log(T/\delta)} \left(\sqrt{R(T)} + \sqrt{\sum_{t\notin\mathcal{T}} \dot{\mu}\left(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star\right)} \right)
\end{aligned}
$$

where $M(T) = \sum_{t\in\mathcal{T}_1} \sum_{i=1}^N \sum_{j=1;j\neq i}^N \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}} \left\|\boldsymbol{x}_t^j\right\|_{(\boldsymbol{W}_t^j)^{-1}}$.

Here, (i) follows from the fact that $\dot{\mu}(.) \leq 1$, (ii) follows from an application of the Cauchy-Schwarz inequality and the fact that $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}^\star \in \mathcal{C}_t(\delta)$, (iii) follows from a direct application of Lemma B.10 and the definition of $\tilde{\boldsymbol{x}}^i$, (iv) follows from the UCB rule, i.e since in slot i, $\boldsymbol{x}_t^i$ was chosen, we have $\boldsymbol{x}_t^{i\top} \boldsymbol{\theta}_t^i + \sqrt{\eta_t(\delta)} \left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}} \geq \boldsymbol{x}_t^{\star,i\top} \boldsymbol{\theta}_t^i + \sqrt{\eta_t(\delta)} \left\|\boldsymbol{x}_t^{\star,i}\right\|_{(\boldsymbol{W}_t^i)^{-1}}$, (v) is a direct application of Lemma E.4 on $\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \boldsymbol{x}_t^i$ and the fact that $\left\|\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \boldsymbol{x}_t^i\right\|_2 \leq \frac{1}{2\sqrt{N}}$, (vi) holds due to the definition of $\mathcal{T}_1$, (vii) follows from Lemma B.14, and $(viii)$ follows from $\eta_t(\delta) \leq CS^2 Nd\log(T/\delta)$.

Turning to $M(T)$, we can bound the term using Rayleigh's quotient and Lemma B.6 (since event $\mathcal{E}_0$ holds) as follows:

$$
\begin{aligned}
M(T) &= \sum_{t\in\mathcal{T}_1} \sum_{i=1}^N \sum_{j=1;j\neq i}^N \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}} \left\|\boldsymbol{x}_t^j\right\|_{(\boldsymbol{W}_t^j)^{-1}} \\
&\overset{(i)}{\leq} \sum_{t\in\mathcal{T}_1} \sum_{i=1}^N \sum_{j=1;j\neq i}^N \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \left\|\boldsymbol{x}_t^i\right\|_2 \left\|\boldsymbol{x}_t^j\right\|_2 \sqrt{\lambda_{max}\left(\boldsymbol{W}_t^i\right)^{-1} \lambda_{max}\left(\boldsymbol{W}_t^j\right)^{-1}} \\
&\overset{(ii)}{\leq} \sum_{t\in\mathcal{T}_1} \sum_{i=1}^N \sum_{j=1;j\neq i}^N \frac{1}{4N} \frac{1}{\sqrt{\lambda_{min}\left(\boldsymbol{W}_t^i\right) \lambda_{min}\left(\boldsymbol{W}_t^j\right)}} \overset{(iii)}{\leq} \frac{N^2}{4N} \sum_{t\in\mathcal{T}_1} \frac{1}{1+\frac{\rho t}{2}} \overset{(iv)}{\leq} \frac{N}{2\rho} \log(T)
\end{aligned}
$$

Here, (i) follows from Rayleigh's Quotient, (ii) follows from $\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \leq \frac{1}{4}$ and $\left\|\boldsymbol{x}_t^i\right\|_2 \leq \frac{1}{\sqrt{N}}$, (iii) follows from a direct application of Lemma B.6, and (iv) follows from the sum of Harmonic Series.

Thus, we get

$$R_1(T)_1 \leq CSNd^{1/2}\sqrt{\left(d\log(T/4N) + \frac{1}{2\rho}\log T\right)\log(T/\delta)}\left(\sqrt{R(T)} + \sqrt{\sum_{t\notin\mathcal{T}} \dot{\mu}\left(\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star\right)}\right)$$

The bound on $R_1(T)_2$ is as follows:

$$\begin{aligned}
R_1(T)_2 &= \sum_{t\in\mathcal{T}_1} \ddot{\mu}(z_t)\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star - \boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right)\left(\boldsymbol{x}_t^\star - \boldsymbol{x}_t\right)^\intercal\boldsymbol{\theta}^\star \overset{(i)}{\leq} C\sqrt{\eta_t(\delta)}\sum_{t\in\mathcal{T}_1}\left|\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star - \boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right)\right|\sum_{i=1}^N 2\left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}} \\
&\overset{(ii)}{\leq} C\sqrt{\eta_t(\delta)}\sum_{t\in\mathcal{T}_1}\left(\sum_{i=1}^N\left\|\tilde{\boldsymbol{x}}_t^i\right\|_{\boldsymbol{W}_t^{-1}}\left\|\boldsymbol{\theta}^\star - \boldsymbol{\theta}_{t+1}\right\|_{\boldsymbol{W}_t}\right)\sum_{i=1}^N 2\left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}} \\
&\overset{(iii)}{\leq} C\sqrt{\eta_t(\delta)}\sum_{t\in\mathcal{T}_1}\left(\sum_{i=1}^N\left\|\tilde{\boldsymbol{x}}_t^i\right\|_{\boldsymbol{W}_t^{-1}}\left\|\boldsymbol{\theta}^\star - \boldsymbol{\theta}_{t+1}\right\|_{\boldsymbol{W}_{t+1}}\right)\sum_{i=1}^N\left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}} \overset{(iv)}{\leq} C\eta_t(\delta)\sum_{t\in\mathcal{T}_1}\left(\sum_{i=1}^N\left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}}\right)^2 \\
&\overset{(v)}{\leq} C\eta_t(\delta)\kappa\sum_{t\in\mathcal{T}_1}\left(\sum_{i=1}^N\left\|\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right)}\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}}\right)^2 \overset{(vi)}{\leq} C\eta_t(\delta)\kappa\left(Nd\log(T/4N) + \frac{N}{2\rho}\log T\right) \\
&\overset{(vii)}{\leq} CS^2Nd\kappa\log(T/\delta)\left(Nd\log(T/4N) + \frac{N}{2\rho}\log T\right)
\end{aligned}$$

Here, (i) follows in the same manner as the regret bound for rounds $t \leq T_0$, and uses the fact that $|\ddot{\mu}(.)| \leq 1$, (ii) is obtained by an application of Cauchy-Schwarz followed by triangle inequality, (iii) follows using the fact that $\tilde{\boldsymbol{\theta}}_t^i, \boldsymbol{\theta}^\star \in \mathcal{C}_t(\delta)$ and $\boldsymbol{W}_t \preccurlyeq \boldsymbol{W}_{t+1}$, (iv) follows since $\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}^\star \in \mathcal{C}_{t+1}(\delta)$ and from Lemma B.10, (v) follows from the definition of κ, i.e $\kappa \geq \frac{1}{\dot{\mu}(\boldsymbol{x}^\intercal\boldsymbol{\theta})}$, (vi) follows similar to the technique used in bounding $R_1(T)_1$, and (vii) follows from $\eta_t(\delta) \leq CS^2Nd\log(T/\delta)$.

Similarly, summing over all indices in $\mathcal{T}_2$, we get:

$$\begin{aligned}
&\sum_{t\in\mathcal{T}_2}\dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star\right)\left(\boldsymbol{x}_t^\star - \boldsymbol{x}_t\right)^\intercal\boldsymbol{\theta}^\star \overset{(i)}{\leq} \sum_{t\in\mathcal{T}_2}\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star\right)}\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}_{t+1}\right)}\left(\boldsymbol{x}_t^\star - \boldsymbol{x}_t\right)^\intercal\boldsymbol{\theta}^\star \\
&\overset{(ii)}{\leq} CSNd^{1/2}\sqrt{\left(d\log(T/4N) + \frac{1}{2\rho}\log T\right)\log(T/\delta)}\left(\sqrt{R(T)} + \sqrt{\sum_{t\notin\mathcal{T}}\dot{\mu}\left(\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star\right)}\right)
\end{aligned}$$

Here, (i) follows from the definition of $\mathcal{T}_2$, (ii) follows using the same steps as followed for $R_1(T)_1$.

Combining all the bounds on $R_1(T)$, we get,

$$\begin{aligned}
R_1(T) \leq\ & CSNd^{1/2}\sqrt{\left(d\log(T/4N) + \frac{1}{2\rho}\log T\right)\log(T/\delta)}\left(\sqrt{R(T)} + \sqrt{\sum_{t\notin\mathcal{T}}\dot{\mu}\left(\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star\right)}\right) \\
&+ CS^2N^2d\kappa\log(T/\delta)\left(d\log(T/4N) + \frac{1}{\rho}\log(T)\right)
\end{aligned}$$

We now bound $R_2(T)$.

$$
\begin{aligned}
R_2(T) &= \sum_{t\notin\mathcal{T}} \int_0^1 (1-v)\dot{\mu}\left(v\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star + (1-v)\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star\right)\mathrm{d}v\left((\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\intercal\boldsymbol{\theta}^\star\right)^2 \\
&\overset{(i)}{\leq} \eta_t(\delta)\sum_{t\notin\mathcal{T}} \int_0^1 (1-v)\left|\dot{\mu}\left(v\boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star + (1-v)\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star\right)\right|\mathrm{d}v\left(\sum_{i=1}^N 2\left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}}\right)^2 \\
&\overset{(ii)}{\leq} \eta_t(\delta)\sum_{t\notin\mathcal{T}} \int_0^1 (1-v)\,\mathrm{d}v\left(\sum_{i=1}^N 2\left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}}\right)^2 \leq 2\eta_t(\delta)\sum_{t\notin\mathcal{T}}\left(\sum_{i=1}^N \left\|\boldsymbol{x}_t^i\right\|_{(\boldsymbol{W}_t^i)^{-1}}\right)^2 \\
&\overset{(iii)}{\leq} CS^2N^2d\kappa\log(T/\delta)\left(d\log(T/4N) + \frac{1}{\rho}\log T\right)
\end{aligned}
$$

Here, (i) follows in a manner similar to the one used in bounding the regret for rounds $t \leq T_0$, $|ab| \leq |a|\,|b|$ and $\left|\int f(x)\,\mathrm{d}x\right| \leq \int |f(x)|\,\mathrm{d}x$, (ii) follows from the fact that $|\dot{\mu}(.)| \leq 1$, and (iii) follows in the same manner as steps $(i), (ii),$ and (iii) follows in a similar manner as the bound for $R_1(T)_2$.

Combining all the bounds, we get

$$
\begin{aligned}
R(T) \leq\ & CSNd^{1/2}\sqrt{\left(d\log(T/4N) + \frac{1}{2\rho}\log T\right)\log(T/\delta)}\left(\sqrt{R(T)} + \sqrt{\sum_{t\notin\mathcal{T}}\dot{\mu}\left(\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star\right)}\right) \\
&+ CS^2N^2d\kappa\log(T/\delta)\left(d\log(T/4N) + \frac{1}{\rho}\log T\right)
\end{aligned}
$$

Applying Lemma E.5 for $R(T)$, we get that

$$
\begin{aligned}
R(T) \leq\ & CSNd^{1/2}\sqrt{\left(d\log(T/4N) + \frac{1}{2\rho}\log T\right)\log(T/\delta)}\sqrt{\sum_{t\notin\mathcal{T}}\dot{\mu}\left(\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star\right)} \\
&+ (1+\kappa)CS^2N^2d\log(T/\delta)\kappa\left(d\log(T/4N) + \frac{1}{\rho}\log(T)\right)
\end{aligned}
$$

Thus, our overall Regret is

$$
\begin{aligned}
Regret(T) \leq\ & T_0 + CSNd^{1/2}\sqrt{\left(d\log(T/4N) + \frac{1}{2\rho}\log T\right)\log(T/\delta)}\sqrt{\sum_{t\notin\mathcal{T}}\dot{\mu}\left(\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star\right)} \\
&+ (1+\kappa)CS^2N^2d\log(T/\delta)\kappa\left(d\log(T/4N) + \frac{1}{\rho}\log(T)\right) + CS^6N^2d^2\kappa\log(T/\delta)\log(T/\kappa N)
\end{aligned}
$$

□

B.1 SUPPORTING LEMMAS FOR THEOREM B.1

Lemma B.1. $\boldsymbol{U}_t^{-\frac{1}{2}}\boldsymbol{W}_t\boldsymbol{U}_t^{-\frac{1}{2}} = \boldsymbol{I}_d + \boldsymbol{A}_t$

$$
\textit{where } \boldsymbol{A}_t = \begin{bmatrix}
\boldsymbol{0}_d & (\boldsymbol{W}_t^1)^{-\frac{1}{2}}\boldsymbol{W}_t^{1,2}(\boldsymbol{W}_t^2)^{-\frac{1}{2}} & \dots & (\boldsymbol{W}_t^1)^{-\frac{1}{2}}\boldsymbol{W}_t^{1,N}(\boldsymbol{W}_t^N)^{-\frac{1}{2}} \\
(\boldsymbol{W}_t^2)^{-\frac{1}{2}}\boldsymbol{W}_t^{2,1}(\boldsymbol{W}_t^1)^{-\frac{1}{2}} & \boldsymbol{0}_d & \dots & (\boldsymbol{W}_t^2)^{-\frac{1}{2}}\boldsymbol{W}_t^{2,N}(\boldsymbol{W}_t^N)^{-\frac{1}{2}} \\
\vdots & \vdots & \dots & \vdots \\
(\boldsymbol{W}_t^N)^{-\frac{1}{2}}\boldsymbol{W}_t^{N,1}(\boldsymbol{W}_t^1)^{-\frac{1}{2}} & (\boldsymbol{W}_t^N)^{-\frac{1}{2}}\boldsymbol{W}_t^{N,2}(\boldsymbol{W}_t^2)^{-\frac{1}{2}} & \dots & \boldsymbol{0}_d
\end{bmatrix}
$$

Proof. It is enough to show $\boldsymbol{W}_t = \boldsymbol{U}_t + \boldsymbol{U}_t^{\frac{1}{2}} \boldsymbol{A}_t \boldsymbol{U}_t^{\frac{1}{2}}$ to prove the claim. We can decompose $\boldsymbol{W}_t$ as follows:

$$
\begin{aligned}
\boldsymbol{W}_t &= \boldsymbol{I}_{Nd} + \sum_{s=1}^{t-1} \dot{\mu}\left(\boldsymbol{x}_s^\intercal \boldsymbol{\theta}_{s+1}\right) \boldsymbol{x}_s \boldsymbol{x}_s^\top \\
&\overset{(i)}{=} \boldsymbol{I}_{Nd} + \sum_{s=1}^{t-1} \dot{\mu}\left(\boldsymbol{x}_s^\intercal \boldsymbol{\theta}_{s+1}\right) \left(\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_s^i\right) \left(\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_s^{i^\top}\right) \\
&\overset{(ii)}{=} \boldsymbol{I}_{Nd} + \sum_{s=1}^{t-1} \dot{\mu}\left(\boldsymbol{x}_s^\intercal \boldsymbol{\theta}_{s+1}\right) \begin{bmatrix} \boldsymbol{x}_s^1 \boldsymbol{x}_s^{1^\top} & \boldsymbol{x}_s^1 \boldsymbol{x}_s^{2^\top} & \dots & \boldsymbol{x}_s^1 \boldsymbol{x}_s^{N^\top} \\ \boldsymbol{x}_s^2 \boldsymbol{x}_s^{1^\top} & \boldsymbol{x}_s^2 \boldsymbol{x}_s^{2^\top} & \dots & \boldsymbol{x}_s^2 \boldsymbol{x}_s^{N^\top} \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{x}_s^N \boldsymbol{x}_s^{1^\top} & \boldsymbol{x}_s^N \boldsymbol{x}_s^{2^\top} & \dots & \boldsymbol{x}_s^N \boldsymbol{x}_s^{N^\top} \end{bmatrix} \\
&\overset{(iii)}{=} \begin{bmatrix} \boldsymbol{W}_t^1 & \boldsymbol{W}_t^{1,2} & \dots & \boldsymbol{W}_t^{1,N} \\ \boldsymbol{W}_t^{2,1} & \boldsymbol{W}_t^2 & \dots & \boldsymbol{W}_t^{2,N} \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{W}_t^{N,1} & \boldsymbol{W}_t^{N,2} & \dots & \boldsymbol{W}_t^N \end{bmatrix} \overset{(iv)}{=} \boldsymbol{U}_t + \boldsymbol{B}_t
\end{aligned}
$$

Here, (i) follows using the fact $\boldsymbol{x}_s = \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_s^i$, (ii) follows from the definition of $\tilde{\boldsymbol{x}}_s^i$, (iii) follows from the definitions of $\boldsymbol{W}_t^i$ and $\boldsymbol{W}_t^{i,j}$ and the fact that $\boldsymbol{I}_{Nd} = \text{diag}(\boldsymbol{I}_d, \dots \boldsymbol{I}_d)$, and (iv) follows from the definition of $\boldsymbol{U}_t$.

We finish the claim by showing $\boldsymbol{B}_t = \begin{bmatrix} \boldsymbol{0}_d & \boldsymbol{W}_t^{1,2} & \dots & \boldsymbol{W}_t^{1,N} \\ \boldsymbol{W}_t^{2,1} & \boldsymbol{0}_d & \dots & \boldsymbol{W}_t^{2,N} \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{W}_t^{N,1} & \boldsymbol{W}_t^{N,2} & \dots & \boldsymbol{0}_d \end{bmatrix} = \boldsymbol{U}_t^{\frac{1}{2}} \boldsymbol{A}_t \boldsymbol{U}_t^{\frac{1}{2}}$, i.e , $\boldsymbol{A}_t = \boldsymbol{U}_t^{-\frac{1}{2}} \boldsymbol{B}_t \boldsymbol{U}_t^{-\frac{1}{2}}$.

Note that since $\boldsymbol{U}_t$ is a diagonal block matrix, $\boldsymbol{U}_t^{-\frac{1}{2}} = \text{diag}\left((\boldsymbol{W}_t^1)^{-\frac{1}{2}}, \dots, (\boldsymbol{W}_t^N)^{-\frac{1}{2}}\right)$. We can write the $(i,j)^{\text{th}}$ element (in this case, $d \times d$ block) of $\boldsymbol{U}_t^{-\frac{1}{2}} \boldsymbol{B}_t \boldsymbol{U}_t^{-\frac{1}{2}}$ as:

$$
\begin{aligned}
\left[\boldsymbol{U}_t^{-\frac{1}{2}} \boldsymbol{B}_t \boldsymbol{U}_t^{-\frac{1}{2}}\right]_{i,j} &= \sum_{k=1}^{N} \sum_{l=1}^{N} \left[\boldsymbol{U}_t^{-\frac{1}{2}}\right]_{i,k} [\boldsymbol{B}_t]_{k,l} \left[\boldsymbol{U}_t^{-\frac{1}{2}}\right]_{l,j} \\
&= \delta_{i,k} \bar{\delta}_{k,l} \delta_{l,j} \left[\boldsymbol{U}_t^{-\frac{1}{2}}\right]_{i,k} [\boldsymbol{B}_t]_{k,l} \left[\boldsymbol{U}_t^{-\frac{1}{2}}\right]_{l,j} \\
&= \begin{cases} (\boldsymbol{W}_t^j)^{-\frac{1}{2}} \boldsymbol{W}_t^{i,j} (\boldsymbol{W}_t^j)^{-\frac{1}{2}} & i \neq j \\ \boldsymbol{0}_{d \times d} & i = j \end{cases} \\
&= [\boldsymbol{A}_t]_{i,j}
\end{aligned}
$$

where $\delta_{i,j}$ denotes the Kronecker Delta, which takes a value of 1 if $i = j$ and 0 otherwise. Likewise, $\bar{\delta}(i,j)$ denotes the complement of the Kronecker Delta. The second equality follows from the fact that the off-diag entries in $\boldsymbol{U}_t^{\frac{-1}{2}}$ are zero matrices and likewise, the diagonal entries in $\boldsymbol{B}_t$ are zero matrices. This completes the proof.

□

Proposition B.1. *Let $\Lambda(\boldsymbol{A})$ denote the set of eigenvalues of $\boldsymbol{A}$. Then,*

$$\Lambda(\boldsymbol{A}) = \Lambda\left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} \end{bmatrix}\right)$$

Proposition B.2. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two symmetric matrices. Then,*

$$\lambda_{max}(\boldsymbol{A}+\boldsymbol{B}) \leq \lambda_{max}(\boldsymbol{A}) + \lambda_{max}(\boldsymbol{B}) \text{ and } \lambda_{min}(\boldsymbol{A}+\boldsymbol{B}) \geq \lambda_{min}(\boldsymbol{A}) + \lambda_{min}(\boldsymbol{B})$$

Lemma B.2. *Define the matrix recurrence relation as follows:*

$$\boldsymbol{A}^{(k)} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{Z}_k \\ \boldsymbol{Z}_k{}^\top & \boldsymbol{A}^{(k-1)} \end{bmatrix} \text{ and } \boldsymbol{A}^{(1)} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{Z}_1 \\ \boldsymbol{Z}_1{}^\top & \boldsymbol{0} \end{bmatrix}$$

Then, $\lambda_{max}\left(\boldsymbol{A}^{(k)}\right) \leq \sum_{i=1}^{k} \sigma_{max}(\boldsymbol{Z}_i)$ and $\lambda_{min}\left(\boldsymbol{A}^{(k)}\right) \geq -\sum_{i=1}^{k} \sigma_{max}(\boldsymbol{Z}_i)$.

Proof. The proof follows by induction. For $k=1$, we see that the statement indeed holds from Lemma E.1.

Assume that the statement holds for $k=n$, i.e $\lambda_{max}\left(A^{(n)}\right) \leq \sum_{i=1}^{n} \sigma_{max}(\boldsymbol{Z}_i)$ and $\lambda_{min}\left(A^{(n)}\right) \geq -\sum_{i=1}^{n} \sigma_{min}(\boldsymbol{Z}_i)$

Consider $\boldsymbol{A}^{(n+1)} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{Z}_{n+1} \\ \boldsymbol{Z}_{n+1}{}^\top & \boldsymbol{A}^{(n)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{Z}_{n+1} \\ \boldsymbol{Z}_{n+1}{}^\top & \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}^{(n)} \end{bmatrix}$

We have that,

$$\begin{aligned}
\lambda_{max}\left(\boldsymbol{A}^{(n+1)}\right) &\overset{(i)}{\leq} \lambda_{max}\left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{Z}_{n+1} \\ \boldsymbol{Z}_{n+1}{}^\top & \boldsymbol{0} \end{bmatrix}\right) + \lambda_{max}\left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}^{(n)} \end{bmatrix}\right) \overset{(ii)}{=} \sigma_{max}(\boldsymbol{Z}_{n+1}) + \lambda_{max}\left(\boldsymbol{A}^{(n)}\right) \\
&\overset{(iii)}{\leq} \sigma_{max}(\boldsymbol{Z}_{n+1}) + \sum_{i=1}^{n} \sigma_{max}(\boldsymbol{Z}_{\boldsymbol{i}}) = \sum_{i=1}^{n+1} \sigma_{max}(\boldsymbol{Z}_{\boldsymbol{i}})
\end{aligned}$$

where (i) follows from Proposition B.2, (ii) follows from Lemma E.1 and Proposition B.1, and (iii) follows from the induction hypothesis.

Similarly,

$$\begin{aligned}
\lambda_{min}\left(\boldsymbol{A}^{(n+1)}\right) &\overset{(i)}{\geq} \lambda_{min}\left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{Z}_{n+1} \\ \boldsymbol{Z}_{n+1}{}^\top & \boldsymbol{0} \end{bmatrix}\right) + \lambda_{min}\left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}^{(n)} \end{bmatrix}\right) \overset{(ii)}{=} -\sigma_{max}(\boldsymbol{Z}_{n+1}) + \lambda_{min}\left(\boldsymbol{A}^{(n)}\right) \\
&\overset{(iii)}{\geq} -\sigma_{max}(\boldsymbol{Z}_{n+1}) - \sum_{i=1}^{n} \sigma_{max}(\boldsymbol{Z}_{\boldsymbol{i}}) = -\sum_{i=1}^{n+1} \sigma_{max}(\boldsymbol{Z}_{\boldsymbol{i}})
\end{aligned}$$

where (i) follows from Proposition B.2, (ii) follows from Lemma E.1 and Proposition B.1, and (iii) follows from the induction hypothesis.

□

Lemma B.3. *The items chosen at round t in two different slots, say i and j, where $i, j \in [N]$ and $i \neq j$ are independent of one another, conditioned on $\mathcal{F}_t$. In other words,*

$$\mathbb{E}\left[\boldsymbol{x}_t^i \boldsymbol{x}_t^{j\top} | \mathcal{F}_t\right] = \boldsymbol{0}_d$$

Proof. It is easy to see that the item chosen in slot i during round t only depends on $\{\boldsymbol{x}_s\}_{s=1}^{t-1}$, $\{\boldsymbol{\theta}_{s+1}\}_{s=1}^{t-1}$, and $\left\{\boldsymbol{x}_s^i\right\}_{s=1}^{t}$. Since, $\mathcal{F}_t$ accounts for all of these terms, conditioned on $\mathcal{F}_t$, the items being chosen in two different slots are independent of one another.

Because of the independence, we can say that

$$\mathbb{E}\left[\boldsymbol{x}_t^i \boldsymbol{x}_t^{j\top} | \mathcal{F}_t\right] = \mathbb{E}\left[\boldsymbol{x}_t^i | \mathcal{F}_t\right] \mathbb{E}\left[\boldsymbol{x}_t^j | \mathcal{F}_t\right] = \mathbf{0}_d$$

where the last equality follows from Assumption 2.1.

□

Lemma B.4. *The diversity assumptions in Assumption 2.1 can be extended to the set of vectors* $\left\{\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \boldsymbol{x}_t^i\right\}_{i=1}^{N}$*, i.e, we can show the following:*

1. $\mathbb{E}\left[\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \boldsymbol{x}_t^i | \mathcal{F}_t\right] = \mathbf{0}_d$
2. $\mathbb{E}\left[\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \boldsymbol{x}_t^i \boldsymbol{x}_t^{j\top} | \mathcal{F}_t\right] = \mathbf{0}_d$ *where* $i \neq j$
3. $\mathbb{E}\left[\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \boldsymbol{x}_t^i \boldsymbol{x}_t^{i\top} | \mathcal{F}_t\right] \succcurlyeq \rho\kappa \boldsymbol{I}_d$

Proof. We attempt to bound $\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)$.

Using the Cauchy-Schwarz inequality, it is easy to see that $-S \leq \boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1} \leq S$. Since $\dot{\mu}\left(.\right)$ is an increasing function on $(-\infty, 0]$ and a decreasing function on $[0, \infty)$, we have that

$$\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \in \begin{cases} \left[\dot{\mu}\left(S\right), \frac{1}{4}\right] & \text{if } \boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1} \in [0, S] \\ \left[\dot{\mu}\left(-S\right), \frac{1}{4}\right] & \text{if } \boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1} \in [-S, 0] \end{cases}$$

Since $\dot{\mu}\left(-S\right) = \dot{\mu}\left(S\right)$, we have that $\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \in \left[\dot{\mu}\left(S\right), \frac{1}{4}\right]$.

Now, we have that

$$\sqrt{\dot{\mu}\left(S\right)} \mathbb{E}\left[\boldsymbol{x}_t^i | \mathcal{F}_t\right] \leq \mathbb{E}\left[\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \boldsymbol{x}_t^i | \mathcal{F}_t\right] \leq \sqrt{\frac{1}{4}} \mathbb{E}\left[\boldsymbol{x}_t^i | \mathcal{F}_t\right] \implies \mathbf{0}_d \leq \mathbb{E}\left[\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \boldsymbol{x}_t^i | \mathcal{F}_t\right] \leq \mathbf{0}_d$$
$$\implies \mathbb{E}\left[\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \boldsymbol{x}_t^i | \mathcal{F}_t\right] = \mathbf{0}_d$$

Similarly, from Lemma B.3,

$$\dot{\mu}\left(S\right) \mathbb{E}\left[\boldsymbol{x}_t^i \boldsymbol{x}_t^{j\top} | \mathcal{F}_t\right] \leq \mathbb{E}\left[\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \boldsymbol{x}_t^i \boldsymbol{x}_t^{j\top} | \mathcal{F}_t\right] \leq \frac{1}{4} \mathbb{E}\left[\boldsymbol{x}_t^i \boldsymbol{x}_t^{j\top} | \mathcal{F}_t\right] \implies \mathbf{0}_d \leq \mathbb{E}\left[\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \boldsymbol{x}_t^i \boldsymbol{x}_t^{j\top} | \mathcal{F}_t\right] \leq \mathbf{0}_d$$
$$\implies \mathbb{E}\left[\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \boldsymbol{x}_t^i \boldsymbol{x}_t^{j\top}\right] = \mathbf{0}_d$$

Finally, since $\kappa = \max_{\boldsymbol{x}} \max_{\boldsymbol{\theta}} \frac{1}{\dot{\mu}(\boldsymbol{x}^\intercal \boldsymbol{\theta})}$, we have that $\kappa \geq \frac{1}{\dot{\mu}(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1})}$. Hence,

$$\mathbb{E}\left[\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \boldsymbol{x}_t^i \boldsymbol{x}_t^{i\top} | \mathcal{F}_t\right] \preccurlyeq \frac{1}{\kappa} \mathbb{E}\left[\boldsymbol{x}_t^i \boldsymbol{x}_t^{i\top} | \mathcal{F}_t\right] \preccurlyeq \rho \boldsymbol{I}_d$$

where the last inequality follows from Assumption 2.1.

□

Lemma B.5. *For all* $i \in [N]$, $j \in [i+1, N]$*, and* $t \geq 0$*,* $\left\|\boldsymbol{W}_t^{i,j}\right\| \leq \sqrt{\frac{t}{2N^2} \log\left(\frac{dN(N-1)}{\delta}\right)}$ *with probability at least* $1 - \delta$*.*

Proof. To prove this lemma, we would invoke Lemma D.2. We have already shown in Lemma B.4 that $\mathbb{E}\left[\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \boldsymbol{x}_t^i {\boldsymbol{x}_t^i}^\top | \mathcal{F}_t\right] = \mathbf{0}_d$. Thus, invoking Lemma D.2 with $\boldsymbol{x}_s = \sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)}\boldsymbol{x}_t^i$, $\boldsymbol{z}_s = \sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)}\boldsymbol{x}_t^j$, $m_1 = m_2 = \sqrt{\frac{\dot{\mu}(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1})}{N}} \leq \frac{1}{2\sqrt{N}}$, $d_1 = d_2 = d$, and $\delta = \frac{2\delta}{N(N-1)}$, we get that

$$\mathbb{P}\left\{\exists t \geq 1 : \left\|\boldsymbol{W}_t^{i,j}\right\| \geq \sqrt{\frac{t}{2N^2}\log\left(\frac{2dN(N-1)}{2\delta}\right)}\right\} \leq \frac{2\delta}{N(N-1)}$$

Performing a union bound over all $i \in [N]$ and $j \in [i+1, N]$ results in the following:

$$\mathbb{P}\left\{\forall t : \left\|\boldsymbol{W}_t^{i,j}\right\| \leq \sqrt{\frac{t}{2N^2}\log\left(\frac{dN(N-1)}{\delta}\right)}\right\} \geq 1 - \delta$$

This finishes the proof.

□

Lemma B.6. *For all $i \in [N]$, $\mathbb{P}\left\{\forall t \geq T_0 : \lambda_{min}\left(\boldsymbol{W}_t^i\right) \leq 1 + \frac{\rho t}{2}\right\} \leq \delta$.*

Proof. To prove this claim, we invoke Lemma D.1. We have already shown in Lemma B.4 that $\mathbb{E}\left[\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)}\boldsymbol{x}_t^i | \mathcal{F}_t\right] = \mathbf{0}_d$ and $\mathbb{E}\left[\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \boldsymbol{x}_t^i {\boldsymbol{x}_t^i}^\top | \mathcal{F}_t\right] \succcurlyeq \rho \boldsymbol{I}_d$. Thus, invoking Lemma D.1 with $\boldsymbol{x}_t = \sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)}\boldsymbol{x}_t^i$, $m = \frac{1}{2\sqrt{N}}$, $d = d$, $\gamma = 1$, $c = \frac{1}{2}$, and $\delta = \frac{\delta}{N}$, we get that with probability atleast $1 - \frac{\delta}{N}$,

$$\lambda_{min}\left(\boldsymbol{W}_t^i\right) \geq 1 + \frac{\rho t}{2}, \forall t \geq \frac{3 + 2N\rho}{3\rho^2 N^2}\log\left(\frac{2dNT}{\delta}\right)$$

Performing a union bound over all $i \in [N]$ and using the fact that $(N-1)^2 \geq 1/N^2$ gives us:

$$\mathbb{P}\left\{\forall t \geq \frac{3 + 2\rho N}{3\rho^2}(N-1)^2\log\left(\frac{2dNT}{\delta}\right), \forall i \in [N] : \lambda_{min}\left(\boldsymbol{W}_t^i\right) \geq 1 + \frac{\rho t}{2}\right\} \geq 1 - \delta$$

Since $T_0 \geq \frac{3+2\rho N}{3\rho^2}(N-1)^2\log\left(\frac{2dNT}{\delta}\right)$, we can say the same for $t \geq T_0$. This finishes the claim.

□

Let us define the following events: $\mathcal{E}_1 = \left\{\forall i \in [N], \forall j \in [i+1, N], \forall t \geq 0 : \left\|\boldsymbol{W}_t^{i,j}\right\| \leq \sqrt{\frac{t}{2N^2}\log\left(\frac{dN(N-1)}{\delta}\right)}\right\}$, $\mathcal{E}_2 = \left\{\forall i \in [N], \forall t \geq T_0 : \lambda_{min}\left(\boldsymbol{W}_t^i\right) \geq 1 + \frac{\rho t}{2}\right\}$, and $\mathcal{E}_0 = \mathcal{E}_1 \cap \mathcal{E}_2$

Lemma B.7. $\mathbb{P}\left\{\mathcal{E}_0\right\} \geq 1 - 2\delta$

Proof. $\mathbb{P}\left\{\overline{\mathcal{E}_0}\right\} = \mathbb{P}\left\{\overline{\mathcal{E}}_1 \cup \overline{\mathcal{E}}_2\right\} \leq \mathbb{P}\left\{\overline{\mathcal{E}}_1\right\} + \mathbb{P}\left\{\overline{\mathcal{E}}_2\right\} \leq 2\delta$ using a union bound.

□

Lemma B.8. *Define the matrix $\boldsymbol{Z}_t^{(i)} = \left[(\boldsymbol{W}_t^i)^{-\frac{1}{2}}\boldsymbol{W}_t^{i,i+1}(\boldsymbol{W}_t^{i+1})^{-\frac{1}{2}}, \ldots (\boldsymbol{W}_t^i)^{-\frac{1}{2}}\boldsymbol{W}_t^{i,N}(\boldsymbol{W}_t^N)^{-\frac{1}{2}}\right]$ Then, under event $\mathcal{E}_0$, for $t \geq T_0$ and $\rho \geq \frac{12}{N}$, we have that*

$$\left\|\boldsymbol{Z}_t^{(i)}\right\| \leq \frac{N-i}{2N(N-1)}$$

Proof. The idea of the proof is borrowed from Das and Sinha [2024]. We know that $\|\boldsymbol{Z}\| = \sup_{\|\boldsymbol{b}\|_2 \le 1} \|\boldsymbol{Z}\boldsymbol{b}\|_2$. Thus,

$$\left\|\boldsymbol{Z}_t^{(i)}\right\| = \sup_{\|\boldsymbol{b}\|_2 \le 1} \left\|\boldsymbol{Z}_t^{(i)}\boldsymbol{b}\right\|_2 = \sup_{\sum_{j=1}^{N-i} \|b_j\|_2 \le 1} \left\|\sum_{j=1}^{N-i} (\boldsymbol{W}_t^i)^{-\frac{1}{2}} \boldsymbol{W}_t^{i,i+j} (\boldsymbol{W}_t^{i+j})^{-\frac{1}{2}} b_j\right\|_2$$

$$\overset{(i)}{\le} \sup_{\sum_{j=1}^{N-i} \|b_j\|_2 \le 1} \sum_{j=1}^{N-i} \left\|(\boldsymbol{W}_t^i)^{-\frac{1}{2}} \boldsymbol{W}_t^{i,i+j} (\boldsymbol{W}_t^{i+j})^{-\frac{1}{2}} b_j\right\|_2 \le \sum_{j=1}^{N-i} \sup_{\|b_j\|_2 \le 1} \left\|(\boldsymbol{W}_t^i)^{-\frac{1}{2}} \boldsymbol{W}_t^{i,i+j} (\boldsymbol{W}_t^{i+j})^{-\frac{1}{2}} b_j\right\|_2$$

$$\overset{(ii)}{\le} \sum_{j=1}^{N-i} \left\|(\boldsymbol{W}_t^i)^{-\frac{1}{2}}\right\| \left\|\boldsymbol{W}_t^{i,i+j}\right\| \left\|(\boldsymbol{W}_t^{i+j})^{-\frac{1}{2}}\right\| \overset{(iii)}{\le} \sum_{j=1}^{N-i} \frac{\left\|\boldsymbol{W}_t^{i,i+j}\right\|}{\sqrt{\lambda_{min}\left(\boldsymbol{W}_t^i\right) \lambda_{min}\left(\boldsymbol{W}_t^{i+j}\right)}} \overset{(iv)}{\le} \sum_{j=1}^{N-i} \frac{\sqrt{\frac{t}{2N^2} \log\left(\frac{dN(N-1)}{\delta}\right)}}{1 + \frac{\rho t}{2}}$$

$$\overset{(v)}{\le} \sum_{j=1}^{N-i} \sqrt{\frac{\frac{1}{2N^2} \log\left(\frac{dN(N-1)}{\delta}\right)}{\frac{3+2\rho N}{12}(N-1)^2 \log\left(\frac{2dNT}{\delta}\right)}} = \frac{1}{N(N-1)} \sum_{j=1}^{N-i} \sqrt{\frac{6}{3+2\rho N}} \times \sqrt{\frac{\log\left(\frac{dN(N-1)}{\delta}\right)}{\log\left(\frac{2dNT}{\delta}\right)}}$$

$$\le \frac{N-i}{N(N-1)} \sqrt{\frac{6}{3+2\rho N}} \le \frac{N-i}{N(N-1)} \sqrt{\frac{3}{\rho N}} \overset{(vi)}{\le} \frac{N-i}{2N(N-1)}$$

where (i) follows from triangle inequality, (ii) follows from the sub-multiplicativity of the norm, (iii) follows from the fact that $\|\boldsymbol{A}\| = \lambda_{max}(\boldsymbol{A})$ and $\lambda_{max}\left(\boldsymbol{A}^{-1}\right) = \frac{1}{\lambda_{min}(\boldsymbol{A})}$, (iv) follows from Lemma D.2, (v) follows from $\frac{1}{1+\frac{\rho t}{2}} \le \frac{1}{\frac{\rho t}{2}}$ and $t \ge T_0$, and (vi) follows from the fact that $\rho N \ge 12$.

□

Lemma B.9. *Under event $\mathcal{E}_0$, for all $t \ge T_0$, we have*

$$\frac{3}{4}\boldsymbol{U}_t \preccurlyeq \boldsymbol{W}_t \preccurlyeq \frac{5}{4}\boldsymbol{U}_t$$

Proof. Define the matrix recurrence relation:

$$\boldsymbol{A}_t^{(i)} = \begin{bmatrix} \boldsymbol{0}_{d\times d} & \boldsymbol{Z}_t^{(i)} \\ \boldsymbol{Z}_t^{(i)\top} & \boldsymbol{A}_t^{(i-1)} \end{bmatrix}$$

where $\boldsymbol{Z}_t^{(i)} = \left[(\boldsymbol{W}_t^i)^{-\frac{1}{2}} \boldsymbol{W}_t^{i,i+1} (\boldsymbol{W}_t^{i+1})^{-\frac{1}{2}}, \ldots (\boldsymbol{W}_t^i)^{-\frac{1}{2}} \boldsymbol{W}_t^{i,N} (\boldsymbol{W}_t^N)^{-\frac{1}{2}}\right]$. Then, it is easy to see that $\boldsymbol{A}_t$ from Lemma B.1 is the same as $\boldsymbol{A}_t^{(1)}$. From Lemma B.2, we have that

$$\lambda_{max}(\boldsymbol{A}_t) \le \sum_{i=1}^{N} \sigma_{max}\left(\boldsymbol{Z}_t^{(i)}\right) = \sum_{i=1}^{N} \left\|\boldsymbol{Z}_t^{(i)}\right\| \le \sum_{i=1}^{N} \frac{N-i}{2N(N-1)} = \frac{1}{4}$$

Similarly,

$$\lambda_{min}(\boldsymbol{A}_t) \ge -\sum_{i=1}^{N} \sigma_{max}\left(\boldsymbol{Z}_t^{(i)}\right) = -\sum_{i=1}^{N} \left\|\boldsymbol{Z}_t^{(i)}\right\| \ge -\sum_{i=1}^{N} \frac{N-i}{2N(N-1)} = -\frac{1}{4}$$

Thus, we can write

$$-\frac{1}{4}\boldsymbol{I}_d \preccurlyeq \boldsymbol{A}_t \preccurlyeq \frac{1}{4}\boldsymbol{I}_d \implies -\frac{1}{4}\boldsymbol{I}_d \preccurlyeq \boldsymbol{U}_t^{-\frac{1}{2}} \boldsymbol{W}_t \boldsymbol{U}_t^{-\frac{1}{2}} - \boldsymbol{I}_d \preccurlyeq \frac{1}{4}\boldsymbol{I}_d \implies \frac{3}{4}\boldsymbol{U}_t \preccurlyeq \boldsymbol{W}_t \preccurlyeq \frac{5}{4}\boldsymbol{U}_t$$

□

Lemma B.10. *Let $\tilde{\boldsymbol{x}}^i$ be the lift of $\boldsymbol{x}^i$. Then,*

$$\left\|\tilde{\boldsymbol{x}}^i\right\|_{\boldsymbol{W}^{-1}} \leq \frac{4}{3}\left\|\boldsymbol{x}^i\right\|_{(\boldsymbol{W}^i)^{-1}}$$

Proof. From Lemma B.9, we have

$$\left\|\tilde{\boldsymbol{x}}^i\right\|_{\boldsymbol{W}^{-1}} \leq \frac{4}{3}\left\|\tilde{\boldsymbol{x}}^i\right\|_{\boldsymbol{U}^{-1}} = \frac{4}{3}\left\|\boldsymbol{x}^i\right\|_{(\boldsymbol{W}^i)^{-1}}$$

where the last inequality follows from the definition of the lift of $\boldsymbol{x}$ and the structure of $\boldsymbol{U}$.

□

Lemma B.11. *With probability at least $1-2\delta$, for all $t \geq T_0$ and $\rho \geq \frac{12}{N}$, we have*

$$\frac{3}{4}\boldsymbol{U}_t^{\mathcal{H}} \preccurlyeq \boldsymbol{V}_t^{\mathcal{H}} \preccurlyeq \frac{5}{4}\boldsymbol{U}_t^{\mathcal{H}}$$

Proof. First, notice the similarity in structures between $\boldsymbol{V}_t^{\mathcal{H}}$ and $\boldsymbol{W}_t$, as well as between $\boldsymbol{U}_t^{\mathcal{H}}$ and $\boldsymbol{U}_t$. Thus, we can perform a decomposition similar to the one in Lemma B.1. We first show that the diversity conditions hold. It is enough to obtain a bound on the norm of the matrices $\boldsymbol{V}_t^{\mathcal{H},i,j}$ and $\boldsymbol{V}_t^{\mathcal{H},i}$ to prove the claim.

We first show that the diversity assumptions also hold for the set of vectors $\left\{\frac{1}{\sqrt{\kappa}}\boldsymbol{x}_t^i\right\}_{i=1}^N$. For this, we show that $\frac{1}{\sqrt{\kappa}}$ is bounded.

From the proof of Lemma B.4, we have shown that $\dot{\mu}\left(\boldsymbol{x}^\intercal\boldsymbol{\theta}\right) \in \left[\dot{\mu}\left(S\right), \frac{1}{4}\right]$. Since, $\kappa = \max_{\boldsymbol{x}}\max_{\boldsymbol{\theta}} \frac{1}{\dot{\mu}(\boldsymbol{x}^\intercal\boldsymbol{\theta})}, \kappa \in \left[4, \frac{1}{\dot{\mu}(S)}\right]$. Hence, $\frac{1}{\kappa} \in \left[\dot{\mu}\left(S\right), \frac{1}{4}\right]$ and we can show:

$$\sqrt{\dot{\mu}\left(S\right)}\mathbb{E}\left[\boldsymbol{x}_s^i|\mathcal{F}_s\right] \leq \mathbb{E}\left[\frac{1}{\sqrt{\kappa}}\boldsymbol{x}_t^i|\mathcal{F}_s\right] \leq \frac{1}{2}\mathbb{E}\left[\boldsymbol{x}_s^i|\mathcal{F}_s\right] \implies \boldsymbol{0}_d \leq \mathbb{E}\left[\frac{1}{\sqrt{\kappa}}\boldsymbol{x}_t^i|\mathcal{F}_s\right] \leq \boldsymbol{0}_d \implies \mathbb{E}\left[\frac{1}{\sqrt{\kappa}}\boldsymbol{x}_t^i|\mathcal{F}_s\right] = \boldsymbol{0}_d$$

Similarly, from Lemma B.3,

$$\dot{\mu}\left(S\right)\mathbb{E}\left[\boldsymbol{x}_s^i\boldsymbol{x}_s^{j\top}|\mathcal{F}_s\right] \leq \mathbb{E}\left[\frac{1}{\kappa}\boldsymbol{x}_s^i\boldsymbol{x}_s^{j\top}|\mathcal{F}_s\right] \leq \frac{1}{4}\mathbb{E}\left[\boldsymbol{x}_s^i\boldsymbol{x}_s^{j\top}|\mathcal{F}_s\right] \implies \boldsymbol{0}_d \leq \mathbb{E}\left[\frac{1}{\kappa}\boldsymbol{x}_s^i\boldsymbol{x}_s^{j\top}|\mathcal{F}_s\right] \leq \boldsymbol{0}_d \implies \mathbb{E}\left[\frac{1}{\kappa}\boldsymbol{x}_s^i\boldsymbol{x}_s^{j\top}|\mathcal{F}_s\right] = \boldsymbol{0}_d$$

Finally,

$$\mathbb{E}\left[\boldsymbol{x}_s^i\boldsymbol{x}_s^{i\top}|\mathcal{F}_s\right] \succcurlyeq \rho\kappa\boldsymbol{I}_d \implies \mathbb{E}\left[\frac{1}{\kappa}\boldsymbol{x}_s^i\boldsymbol{x}_s^{i\top}|\mathcal{F}_s\right] \succcurlyeq \rho\boldsymbol{I}_d$$

Using an idea similar to Lemma B.5, we can define the event

$$\mathcal{E}_1' = \left\{\forall i \in [N], \forall j \in [i+1, N], \forall t \geq T_0 : \left\|\boldsymbol{V}_t^{\mathcal{H},i,j}\right\| \leq \sqrt{\frac{8t}{\kappa^2 N^2}\log\left(\frac{dN(N-1)}{\delta}\right)}\right\}$$

Similarly, using an idea similar to Lemma B.6, we can define the event

$$\mathcal{E}_2' = \left\{\forall i \geq 0, \forall t \geq \frac{48+8\kappa N\rho}{3\rho^2\kappa^2}(N-1)^2\log\left(\frac{2dNT}{\delta}\right) : \lambda_{min}\left(\boldsymbol{V}_t^{\mathcal{H},i}\right) \geq \gamma_t(\delta) + \frac{\rho t}{2}\right\}$$

Since, $\kappa \geq 4$, we have that $T_0 \geq \frac{3+2N\rho}{3\rho^2}(N-1)^2\log\left(\frac{2dNT}{\delta}\right) \geq \frac{48+8\kappa N\rho}{3\rho^2\kappa^2}(N-1)^2\log\left(\frac{2dNT}{\delta}\right)$, and hence, we have

$$\mathcal{E}_2' = \left\{\forall i \geq 0, \forall t \geq T_0 : \lambda_{min}\left(\boldsymbol{V}_t^{\mathcal{H},i}\right) \geq \gamma_t(\delta) + \frac{\rho t}{2}\right\}$$

Define $\mathcal{E}_0' = \mathcal{E}_1' \cap \mathcal{E}_2'$. Then, it is easy to see $\mathbb{P}\{\mathcal{E}_0'\} \geq 1 - 2\delta$.

Finally, following the same line of thought as Lemma B.8 and Lemma B.9, and using the fact that $\frac{1}{\kappa} \leq \frac{1}{4}$, we obtain

$$\frac{3}{4}\boldsymbol{U}_t^{\mathcal{H}} \preccurlyeq \boldsymbol{V}_t^{\mathcal{H}} \preccurlyeq \frac{5}{4}\boldsymbol{U}_t^{\mathcal{H}}$$

□

Lemma B.12. *(Faury et al. [2022], Proposition 7) Let $\delta \in (0,1)$ and $\{(\boldsymbol{\theta}_t, \boldsymbol{W}_t, \boldsymbol{\theta}_t)\}_r$ be maintained by the ada-OFU-ECOLog algorithm. Then,*

$$\mathbb{P}\left\{\forall t \geq 1 : \boldsymbol{\theta}^\star \in \boldsymbol{\theta}_t \text{ and } \|\boldsymbol{\theta}^\star - \boldsymbol{\theta}_{t+1}\|_{\boldsymbol{W}_{t+1}} \leq CS^2 d\log(t/\delta)\right\} \geq 1 - 2\delta$$

Lemma B.13. *Define the following events:*

$$\mathcal{E}' = \left\{\forall t \geq 1, \|\boldsymbol{\theta}^\star - \boldsymbol{\theta}_{t+1}\|^2_{\boldsymbol{W}_{t+1}} \leq CS^2 d\log(t/\delta) \text{ and } \boldsymbol{\theta}^\star \in \Theta\right\}$$

$$\mathcal{E} = \mathcal{E}_0 \cap \mathcal{E}_0' \cap \mathcal{E}'$$

Then, we have that $\mathbb{P}\{\mathcal{E}\} \leq 6\delta$.

Proof.

$$\mathbb{P}\{\overline{\mathcal{E}}\} = \mathbb{P}\left\{\overline{\mathcal{E}_0 \cap \mathcal{E}_0' \cap \mathcal{E}'}\right\} = \mathbb{P}\left\{\overline{\mathcal{E}_0} \cup \overline{\mathcal{E}_0'} \cup \overline{\mathcal{E}'}\right\} \leq \mathbb{P}\{\overline{\mathcal{E}_0}\} + \mathbb{P}\left\{\overline{\mathcal{E}_0'}\right\} + \mathbb{P}\{\overline{\mathcal{E}'}\} \leq 2\delta + 2\delta + 2\delta = 6\delta$$

where the last inequality follows from Lemma B.7 , B.11, and B.12 respectively.

□

Lemma B.14. *(Abeille et al. [2021], Theorem 1)* $\sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star) \leq R(T) + \sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star)$ *where* $R_T = \sum_{t=1}^{T} \mu(\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star) - \mu(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star)$

Proof. We provide a brief proof for the sake of completeness

$$\begin{aligned}
\sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star) &= \sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star) + \sum_{t=1}^{T} \int_0^1 \ddot{\mu}(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star + v(\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\intercal \boldsymbol{\theta}^\star)\ \mathrm{d}v\, (\boldsymbol{x}_t - \boldsymbol{x}_t^\star)^\intercal \boldsymbol{\theta}^\star \\
&\leq \sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star) + \sum_{t=1}^{T} \left| \int_0^1 \ddot{\mu}(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star + v(\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\intercal \boldsymbol{\theta}^\star)\ \mathrm{d}v\, (\boldsymbol{x}_t - \boldsymbol{x}_t^\star)^\intercal \boldsymbol{\theta}^\star \right| \\
&\overset{(i)}{\leq} \sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star) + \sum_{t=1}^{T} \int_0^1 |\ddot{\mu}(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star + v(\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\intercal \boldsymbol{\theta}^\star)|\ \mathrm{d}v\, |(\boldsymbol{x}_t - \boldsymbol{x}_t^\star)^\intercal \boldsymbol{\theta}^\star| \\
&\overset{(ii)}{\leq} \sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star) + \sum_{t=1}^{T} \int_0^1 |\ddot{\mu}(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star + v(\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\intercal \boldsymbol{\theta}^\star)|\ \mathrm{d}v\, (\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\intercal \boldsymbol{\theta}^\star \\
&\overset{(iii)}{\leq} \sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star) + \sum_{t=1}^{T} \int_0^1 \dot{\mu}(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star + v(\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\intercal \boldsymbol{\theta}^\star)\ \mathrm{d}v\, (\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\intercal \boldsymbol{\theta}^\star \\
&\overset{(iv)}{=} \sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star) + \sum_{t=1}^{T} \mu(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star) - \mu(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}^\star) \\
&= \sum_{t=1}^{T} \dot{\mu}(\boldsymbol{x}_t^{\star\intercal} \boldsymbol{\theta}^\star) + R(T)
\end{aligned}$$

Here, (i) follows from $\left|\int f(x)\ \mathrm{d}x\right| \leq \int |f(x)|\ \mathrm{d}x$, (ii) follows from $\boldsymbol{x}_t^{\star\intercal}\boldsymbol{\theta}^\star \geq \boldsymbol{x}_t^\intercal\boldsymbol{\theta}^\star$, (iii) follows since $|\ddot{\mu}(.)| \leq \dot{\mu}(.)$, and (iv) follows from applying the Mean-Value Theorem on the expression for $R(T)$.

□

Lemma B.15. *Let $\mathcal{T}$ represent the set of all time instances where the data-dependent condition fails, i.e $\forall t \in \mathcal{T}$, $\dot{\mu}\left(\boldsymbol{x}_t^\intercal\bar{\boldsymbol{\theta}}_t\right) \geq 2\dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}_t^u\right)$ for all $u \in \{0,1\}$. Then,*

$$|\mathcal{T}| \leq CS^6N^2d^2\kappa\log(T/\delta)\log(T/\kappa N)$$

Proof. The proof follows along the lines of Faury et al. [2022].

By the self-concordance property of the logistic function, we know that

$$\dot{\mu}\left(\boldsymbol{x}_t^\intercal\bar{\boldsymbol{\theta}}_t\right) \leq \dot{\mu}\left(\boldsymbol{x}_t^\intercal\boldsymbol{\theta}_t^u\right)\exp\left(\left|\boldsymbol{x}_t^\intercal(\bar{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^u)\right|\right)$$

Thus, if $t \in \mathcal{T}$, we have that $\left|\boldsymbol{x}_t^\intercal(\bar{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^u)\right| \geq \log 2$.

Summing this over all indices in $\mathcal{T}$, we get that

$$\begin{aligned}
\sum_{t\in\mathcal{T}}\log^2 2 = |\mathcal{T}|\log^2 2 &\leq \sum_{t\in\mathcal{T}}\left|\boldsymbol{x}_t^\intercal(\bar{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^u)\right|^2 \overset{(i)}{\leq} \sum_{t\in\mathcal{T}}\|\boldsymbol{x}_t\|^2_{(\boldsymbol{V}_t^{\mathcal{H}})^{-1}}\left\|\bar{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^u\right\|^2_{\boldsymbol{V}_t^{\mathcal{H}}} \overset{(ii)}{\leq} 4\beta_T(\delta)\sum_{t\in\mathcal{T}}\|\boldsymbol{x}_t\|^2_{(\boldsymbol{V}_t^{\mathcal{H}})^{-1}} \\
&\overset{(iii)}{\leq} C\beta_T(\delta)\sum_{t\in\mathcal{T}}\|\boldsymbol{x}_t\|^2_{(\boldsymbol{U}_t^{\mathcal{H}})^{-1}} \overset{(iv)}{\leq} C\beta_T(\delta)\sum_{t\in\mathcal{T}}\left\|\sum_{i=1}^N\tilde{\boldsymbol{x}}_t^i\right\|^2_{(\boldsymbol{U}_{t-1}^{\mathcal{H}})^{-1}} \overset{(v)}{\leq} C\beta_T(\delta)\sum_{i=1}^N\sum_{t\in\mathcal{T}}\left\|\tilde{\boldsymbol{x}}_t^i\right\|^2_{(\boldsymbol{U}_t^{\mathcal{H}})^{-1}} \\
&\leq C\beta_T(\delta)\sum_{i=1}^N\sum_{t\in\mathcal{T}}\left\|\boldsymbol{x}_t^i\right\|^2_{(\boldsymbol{V}_t^{\mathcal{H},i})^{-1}} \overset{(vi)}{\leq} CNd\beta_T(\delta)\kappa\log(t/\kappa N) \overset{(vii)}{\leq} CS^6N^2d^2\kappa\log(T/\delta)\log(T/\kappa N)
\end{aligned}$$

Here (i) follows from the Cauchy-Schwarz Inequality, (ii) follows from the fact that $\boldsymbol{\theta}_t^u, \bar{\boldsymbol{\theta}}_t \in \Theta_t$, $(a+b)^2 \leq 2a^2 + 2b^2$, (iii) follows due to event $\mathcal{E}_0'$, (iv) follows from the definition of the lift of $\boldsymbol{x}_s^i$, i.e $\boldsymbol{x}_s = \sum_{i=1}^N \tilde{\boldsymbol{x}}_s^i$, (v) follows from the triangle inequality, (vi) follows from a direct application of Lemma E.4 on $\frac{1}{\sqrt{\kappa}}\boldsymbol{x}_t^i$ and the fact that $\left\|\frac{1}{\sqrt{\kappa}}\boldsymbol{x}_t^i\right\|_2 \leq \frac{1}{\sqrt{N\kappa}}$, and (vii) follows from the definition $\beta_T(\delta) \leq CS^6Nd\log(T/\delta)$.

□

C SLATE-GLM-TS AND SLATE-GLM-TS-FIXED

C.1 ALGORITHM IN A FIXED-ARM SETTING

We present a Thompson Sampling based algorithm `Slate-GLM-TS-Fixed` in the non-contextual (fixed-arm) setting in Algorithm 4. Following this, we analyze the regret of this algorithm in Theorem C.1. Since we are in the non-contextual setting, we directly use the minimum eigenvalue bound in Assumption C.1. (See Remarks on Assumption 2.1 in Section 2).

Algorithm 4 `Slate-GLM-TS-Fixed`

1: **Inputs:** Number of rounds T, Failure probability δ , Distribution $\mathcal{D}^{TS}$, warm-up length τ
2: Initialize $\boldsymbol{V}_0^{\mathcal{H},i} = \lambda \boldsymbol{I}_d \ \forall i \in [N]$ and $\boldsymbol{V}_0^{\mathcal{H}} = \lambda \boldsymbol{I}_{Nd}$
3: Obtain the set of items $\mathcal{X}^i, \forall i \in [N]$
4: **for** each round t in $[1, \tau]$ **do**
5: For each slot $i \in [N]$, choose $\boldsymbol{x}_t^i = \arg\max_{\boldsymbol{x} \in \mathcal{X}^i} \|\boldsymbol{x}\|_{(\boldsymbol{V}_t^{\mathcal{H},i})^{-1}}$, select slate $\boldsymbol{x}_t = (\boldsymbol{x}_t^1, \ldots, \boldsymbol{x}_t^N)$, and get reward y_t.
6: Update $\boldsymbol{V}_t^{\mathcal{H}} \leftarrow \boldsymbol{V}_{t-1}^{\mathcal{H}} + \frac{1}{\kappa} \boldsymbol{x}_t \boldsymbol{x}_t^\top$ and $\boldsymbol{V}_t^{\mathcal{H},i} \leftarrow \boldsymbol{V}_{t-1}^{\mathcal{H},i} + \frac{1}{\kappa} \boldsymbol{x}_t^i \boldsymbol{x}_t^{i\top}, \forall i \in [N]$
7: **end for**
8: Compute $\widehat{\boldsymbol{\theta}}_\tau = \arg\min \sum_{s=1}^{\tau} l_{s+1}(\boldsymbol{\theta}) + \frac{\lambda}{2} \|\boldsymbol{\theta}\|_2^2$ and set $\Theta = \left\{ \left\| \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_\tau \right\|_{\boldsymbol{V}_\tau^{\mathcal{H}}} \leq \beta_\tau(\delta) \right\}$
9: Initialize $\boldsymbol{W}_\tau = \boldsymbol{I}_{dN}, \boldsymbol{W}_\tau^i = \boldsymbol{I}_d, \forall i \in [N]$ and $\boldsymbol{\theta}_{\tau+1} \in \Theta$
10: **for** each round $t \in [\tau + 1, T]$ **do**
11: Set reject = True
12: **while** reject **do**
13: For each slot $i \in [N]$, sample $\boldsymbol{\eta}^i \overset{\text{iid}}{\sim} \mathcal{D}^{TS}$, and set $\tilde{\boldsymbol{\theta}}_t^i = \boldsymbol{\theta}_t^i + \eta_t(\delta)(\boldsymbol{W}_t^i)^{-1/2} \boldsymbol{\eta}^i$
14: If $\tilde{\boldsymbol{\theta}}_t = (\tilde{\boldsymbol{\theta}}_t^1, \ldots, \tilde{\boldsymbol{\theta}}_t^N) \in \Theta_t$, set reject = False
15: **end while**
16: For each slot $i \in [N]$, choose $\boldsymbol{x}_t^i = \arg\max_{\boldsymbol{x} \in \mathcal{X}^i} \boldsymbol{x}^\intercal \tilde{\boldsymbol{\theta}}_t^i$, select slate $\boldsymbol{x}_t = (\boldsymbol{x}_t^1, \ldots, \boldsymbol{x}_t^N)$, and get reward y_t
17: Let $\boldsymbol{\theta}_{t+1}$ be solution of 5 up to precision $1/t$.
18: Update $\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \dot{\mu}(\boldsymbol{x}_t^T \boldsymbol{\theta}_{t+1}) \boldsymbol{x}_t \boldsymbol{x}_t^T$, and $\boldsymbol{W}_{t+1}^i = \boldsymbol{W}_t^i + \dot{\mu}(\boldsymbol{x}_t^\top \boldsymbol{\theta}_{t+1}) \boldsymbol{x}_t^i \boldsymbol{x}_t^{i\top}, \forall i \in [N]$
19: **end for**

Assumption C.1. *The minimum eigenvalue of the design matrices grows linearly, i.e*

$$\lambda_{min}\left(\boldsymbol{V}_t^{(i)}\right) = \lambda_{min}\left(\boldsymbol{W}_t^{(i)}\right) \geq \rho t \text{ and } \lambda_{min}\left(\boldsymbol{V}_t^{\mathcal{W}\,(i)}\right) \geq \rho t$$

Define $T_0 = max\left\{\frac{(N-1)^2}{2\rho^2} \log \frac{dN(N-1)}{\delta}, \frac{8(N-1)^2}{\kappa^2 \rho^2} \log \frac{dN(N-1)}{\delta}\right\} = \frac{(N-1)^2}{2\rho^2} \log \frac{dN(N-1)}{\delta}$ since $\kappa > 4$.

Theorem C.1. *(Regret of `Slate-GLM-TS-Fixed`) At the end of $T \geq T_0$ rounds, the regret of `Slate-GLM-TS-Fixed` is bounded by*

$$Regret(T) \leq \max\{CS^6 N^2 d^2 \kappa \log(T/\delta)^2, T_0\} + CSN^{3/2} d^{3/2} \sqrt{\log(T/\delta)\log(T/2)} \sqrt{T \dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)} + CN^3 d^3 S^2 \log(T/\delta) \log(T/2)$$

Proof. We have that the *good events* are defined for $t \in [T_0, T]$. Since the first $|\mathcal{T}| = \tau$ rounds constitute a warm-up (*Steps 4-7* in Algorithm 4), we can trivially bound the regret of these rounds (warm-up as well as first T_0) by $1 \cdot \max\{\tau, T_0\}$. Going forward, let $\max\{\tau, T_0\} = T'$ Hence, we have

$$\begin{aligned}
Regret(T) &\leq \max\{\tau, T_0\} + \sum_{t=T'+1}^{T} \mu\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right) - \mu\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_\star\right) \\
&\leq \max\{CS^6 N^2 d^2 \kappa \log(T/\delta)^2, T_0\} + \sum_{t=T'+1}^{T} \left\{\mu\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right) - \mu\left(\boldsymbol{x}_t^\intercal \tilde{\boldsymbol{\theta}}_t\right)\right\} + \sum_{t=T'+1}^{T} \left\{\mu\left(\boldsymbol{x}_t^\intercal \tilde{\boldsymbol{\theta}}_t\right) - \mu\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_\star\right)\right\} \\
&= \max\{CS^6 N^2 d^2 \kappa \log(T/\delta)^2, T_0\} + R^{TS}(T) + R^{PRED}(T) = \max\{CS^6 N^2 d^2 \kappa \log(T/\delta)^2, T_0\} + R(T)
\end{aligned}$$

where $R(T) = R^{TS}(T) + R^{PRED}(T)$, $R^{TS}(T) = \sum_{t=T'+1}^{T} \mu\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right) - \mu\left(\boldsymbol{x}_t^\intercal \tilde{\boldsymbol{\theta}}_t\right)$, and $R^{PRED}(T) = \sum_{t=T'+1}^{T} \mu\left(\boldsymbol{x}_t^\intercal \tilde{\boldsymbol{\theta}}_t\right) - \mu\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_\star\right)$. The first inequality follows from Lemma C.1.

We first bound $R^{PRED}(T)$ as follows:

$$
\begin{aligned}
R^{PRED}(T) &= \sum_{t=T'+1}^{T} \mu\left(\boldsymbol{x}_t^\intercal \tilde{\boldsymbol{\theta}}_t\right) - \mu\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_\star\right) \leq \sum_{t=T'+1}^{T} \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_\star\right) \left|\boldsymbol{x}_t^\intercal \left(\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_\star\right)\right| \\
&\overset{(i)}{\leq} \sum_{t=T'+1}^{T} \sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_\star\right)} \sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \exp\left(\left|\boldsymbol{x}_t^\intercal \left(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_{t+1}\right)\right|\right)} \left\|\boldsymbol{x}_t\right\|_{\boldsymbol{W}_t^{-1}} \left\|\boldsymbol{\theta}_\star - \tilde{\boldsymbol{\theta}}_t\right\|_{\boldsymbol{W}_t} \\
&\overset{(ii)}{\leq} C\sqrt{e}\sqrt{\sigma_t(\delta)}\sqrt{Nd} \sum_{t=T'+1}^{T} \sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_\star\right)} \sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)} \left\|\boldsymbol{x}_t\right\|_{\boldsymbol{W}_t^{-1}} \\
&\overset{(iii)}{\leq} C\sqrt{\sigma_t(\delta)}\sqrt{Nd} \sqrt{\sum_{t=T'+1}^{T} \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_\star\right)} \sqrt{\sum_{t=T'+1}^{T} \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right) \left\|\boldsymbol{x}_t\right\|_{\boldsymbol{W}_t^{-1}}^2} \\
&\overset{(iv)}{\leq} C\sqrt{\sigma_t(\delta)} Nd \sqrt{\log(T/2)} \left(\sqrt{R(T)} + \sqrt{T\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)}\right) \\
&\overset{(v)}{\leq} CSN^{3/2} d^{3/2} \sqrt{\log(T/\delta)\log(T/2)} \left(\sqrt{R(T)} + \sqrt{T\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)}\right)
\end{aligned}
$$

where (i) follows from the Self-Concordance result and uses Cauchy-Schwarz, (ii) follows from the fact that $\left|\boldsymbol{x}_t^\top \left(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_{t+1}\right)\right| \leq \text{diam}_{\mathcal{X}}\left(\Theta\right) \leq 1$ (Lemma C.1) and Lemma C.5, (iii) follows from Cauchy-Schwarz, (iv) follows from Lemma E.4 on $\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)}\boldsymbol{x}_t$ and Lemma B.14, and (v) follows from the fact that $\sigma_t(\delta) \leq CS^2 Nd\log(T/\delta)$.

We now turn to bounding $R^{TS}(T)$. Define $J(\boldsymbol{\theta}) = \max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{x}^\intercal \boldsymbol{\theta}$. Then, it is easy to see that $J(\boldsymbol{\theta}_\star) = \boldsymbol{x}_\star^\top \boldsymbol{\theta}_\star$. Also, note that

$$
J(\tilde{\boldsymbol{\theta}}_t) = \max_{\boldsymbol{x}\in\mathcal{X}} \sum_{i=1}^{N} {\boldsymbol{x}^i}^\top \tilde{\boldsymbol{\theta}}_t^i = \sum_{i=1}^{N} \max_{\boldsymbol{x}\in\mathcal{X}^i} \boldsymbol{x}^\top \tilde{\boldsymbol{\theta}}_t^i = \sum_{i=1}^{N} \boldsymbol{x}_t^\top \tilde{\boldsymbol{\theta}}_t^i = \boldsymbol{x}_t^\top \tilde{\boldsymbol{\theta}}_t
$$

which uses the fact that the selection of the item in each slot is independent of the rest of the slots.

Hence, we have

$$
R^{TS}(T) = \sum_{t=T'+1}^{T} \mu\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right) - \mu\left(\boldsymbol{x}_t^\intercal \tilde{\boldsymbol{\theta}}_t\right) = \alpha\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star, \boldsymbol{x}_t^\intercal \tilde{\boldsymbol{\theta}}_t\right)\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star - \boldsymbol{x}_t^\intercal \tilde{\boldsymbol{\theta}}_t\right) = \alpha(J(\boldsymbol{\theta}_\star), J(\tilde{\boldsymbol{\theta}}_t))\left(J(\boldsymbol{\theta}_\star) - J(\tilde{\boldsymbol{\theta}}_t)\right)
$$

Similar to Section D.2 of the Appendix in Faury et al. [2022] and Section C of Abeille and Lazaric [2017], using the convexity of J gives us:

$$
\begin{aligned}
\left|J(\boldsymbol{\theta}_\star) - J(\tilde{\boldsymbol{\theta}}_{t+1})\right| &\leq \max\left\{\left|\nabla J(\boldsymbol{\theta}_\star)^\intercal \left(\boldsymbol{\theta}_\star - \tilde{\boldsymbol{\theta}}_t\right)\right|, \left|\nabla J(\tilde{\boldsymbol{\theta}}_{t+1})^\intercal \left(\boldsymbol{\theta}_\star - \tilde{\boldsymbol{\theta}}_t\right)\right|\right\} \overset{(i)}{\leq} \max\left\{\left|\boldsymbol{x}_\star^\intercal \left(\boldsymbol{\theta}_\star - \tilde{\boldsymbol{\theta}}_t\right)\right|, \left|\boldsymbol{x}_t^\intercal \left(\boldsymbol{\theta}_\star - \tilde{\boldsymbol{\theta}}_t\right)\right|\right\} \\
&\leq \text{diam}_{\mathcal{X}}\left(\Theta\right) \overset{(ii)}{\leq} 1
\end{aligned}
$$

where (i) follows from the fact that $\nabla J(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{x}^\top \boldsymbol{\theta}$ (Abeille and Lazaric [2017]), and (ii) follows from Lemma C.1. Thus, we have that

$$
\begin{aligned}
\alpha(J(\boldsymbol{\theta}_\star), J(\tilde{\boldsymbol{\theta}}_t)) &= \int_0^1 \dot{\mu}\left(J(\boldsymbol{\theta}_\star) + v\left(J(\boldsymbol{\theta}_\star) - J(\tilde{\boldsymbol{\theta}}_t)\right)\right) \mathrm{d}v \leq \dot{\mu}\left(J(\boldsymbol{\theta}_\star)\right) \int_0^1 \exp\left(v\left|J(\boldsymbol{\theta}_\star) - J(\tilde{\boldsymbol{\theta}}_t)\right|\right) \mathrm{d}v \\
&\leq \dot{\mu}\left(J(\boldsymbol{\theta}_\star)\right) \int_0^1 \exp\left(v\right) \mathrm{d}v \leq 2\dot{\mu}\left(J(\boldsymbol{\theta}_\star)\right) = 2\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)
\end{aligned}
$$

where the first inequality follows from self-concordance. Substituting this into the original bound, we get

$$R^{TS}(T) \leq 2\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right) \sum_{t=T'+1}^{T} J(\boldsymbol{\theta}_\star) - J(\tilde{\boldsymbol{\theta}}_t)$$

Following the same steps as the proof in Abeille and Lazaric [2017] and referring to Section D.2 in Faury et al. [2022], we get that

$$\sum_{t=T'+1}^{T} J(\boldsymbol{\theta}_\star) - J(\tilde{\boldsymbol{\theta}}_t) \lesssim C\sqrt{Nd}\sqrt{\sigma_t(\delta)} \sum_{t=T'+1}^{T} \|\boldsymbol{x}_t\|_{\boldsymbol{W}_t^{-1}} + \sqrt{T}$$

Substituting this into the original equation, we get that:

$$\begin{aligned}
R^{TS}(T) &\leq 2\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)\left(C\sqrt{Nd}\sqrt{\sigma_t(\delta)} \sum_{t=T'+1}^{T} \|\boldsymbol{x}_t\|_{\boldsymbol{W}_t^{-1}} + \sqrt{T}\right) \\
&\overset{(i)}{\leq} C\sqrt{Nd}\sqrt{\sigma_t(\delta)} \sum_{t=T'+1}^{T} \sqrt{\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)}\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)\exp\left(|\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1} - \boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star|\right)}\,\|\boldsymbol{x}_t\|_{\boldsymbol{W}_t^{-1}} + 2\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)\sqrt{T} \\
&\overset{(ii)}{\leq} C\sqrt{Nd}\sqrt{\sigma_t(\delta)} \sum_{t=T'+1}^{T} \sqrt{\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)}\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)}\,\|\boldsymbol{x}_t\|_{\boldsymbol{W}_t^{-1}} + 2\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)\sqrt{T} \\
&\overset{(iii)}{\leq} C\sqrt{Nd}\sqrt{\sigma_t(\delta)}\sqrt{\sum_{t=T'+1}^{T} \dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)}\sqrt{\sum_{t=T'+1}^{T} \dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)\|\boldsymbol{x}_t\|^2_{\boldsymbol{W}_t^{-1}}} + 2\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)\sqrt{T} \\
&\overset{(iv)}{\leq} CNd\sqrt{\sigma_t(\delta)}\sqrt{T\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)}\sqrt{\log(T/2)} + 2\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)\sqrt{T} \\
&\overset{(v)}{\leq} CN^{3/2}d^{3/2}S\sqrt{\log(T/2)\log(T/\delta)}\sqrt{T\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)} + 2\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)\sqrt{T}
\end{aligned}$$

where (i) follows from self-concordance, (ii) follows from the fact that $|\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1} - \boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star| \leq 2\text{diam}_{\mathcal{X}}(\Theta)$, (iii) follows from Cauchy-Schwarz, (iv) follows from Lemma E.4 on $\sqrt{\dot{\mu}\left(\boldsymbol{x}_t^\intercal \boldsymbol{\theta}_{t+1}\right)}\boldsymbol{x}_t^i$, and (v) follows from the fact that $\sigma_t(\delta) \leq CS^2Nd\log(T/\delta)$

Combining the bounds on $R(T)$, we get

$$R(T) \leq CSN^{3/2}d^{3/2}\sqrt{\log(T/\delta)\log(T/2)}\left(\sqrt{R(T)} + \sqrt{T\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)}\right)$$

Using Lemma E.5, we get

$$R(T) \leq CSN^{3/2}d^{3/2}\sqrt{\log(T/\delta)\log(T/2)}\sqrt{T\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)} + CN^3d^3S^2\log(T/\delta)\log(T/2)$$

Finally, combining the bound for $Regret(T)$ gives us:

$$Regret(T) \leq \max\{CS^6N^2d^2\kappa\log(T/\delta)^2, T_0\} + CSN^{3/2}d^{3/2}\sqrt{\log(T/\delta)\log(T/2)}\sqrt{T\dot{\mu}\left(\boldsymbol{x}_\star^\intercal \boldsymbol{\theta}_\star\right)} + CN^3d^3S^2\log(T/\delta)\log(T/2)$$

□

C.2 SUPPORTING LEMMAS FOR THEOREM C.1

Lemma C.1. *Let $\delta \in (0,1)$, then, setting $\tau = CS^6N^2d^2\kappa\log(T/\delta)^2$ ensures that Θ returned after the warm-up phase satisfies the following:*

1. $\mathbb{P}\left\{\boldsymbol{\theta}_\star \in \Theta\right\} \geq 1-\delta$
2. $diam_{\mathcal{X}}\left(\Theta\right) \leq 1$

Proof. The proof for the first part is the same as the proof for the first part in Proposition 5 in Faury et al. [2022] since the proof does not depend on the manner in which the arm is selected.

For the second part notice that:

$$
\begin{aligned}
\operatorname{diam}_{\mathcal{X}}\left(\Theta\right) &= \max_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta} \left|\boldsymbol{x}^\intercal\left(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\right)\right| \overset{(i)}{\leq} \max_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{x}\|_{(\boldsymbol{V}_\tau^{\mathcal{H}})^{-1}} \max_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta} \|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_{\boldsymbol{V}_\tau^{\mathcal{H}}} \overset{(ii)}{\leq} \sqrt{\beta_t(\delta)} \max_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{x}\|_{(\boldsymbol{V}_\tau^{\mathcal{H}})^{-1}} \\
&\leq \sqrt{\beta_t(\delta)}\sqrt{\max_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{x}\|^2_{(\boldsymbol{V}_\tau^{\mathcal{H}})^{-1}}} \leq \sqrt{\beta_t(\delta)}\frac{1}{\sqrt{\tau}}\sqrt{\sum_{t=1}^{\tau}\max_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{x}\|^2_{(\boldsymbol{V}_\tau^{\mathcal{H}})^{-1}}} \overset{(iii)}{\leq} \sqrt{\beta_t(\delta)}\frac{1}{\sqrt{\tau}}\sqrt{2\sum_{t=1}^{\tau}\max_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{x}\|^2_{(\boldsymbol{U}_\tau^{\mathcal{H}})^{-1}}} \\
&\leq \sqrt{\beta_t(\delta)}\frac{1}{\sqrt{\tau}}\sqrt{2\sum_{t=1}^{\tau}\sum_{i=1}^{N}\max_{\boldsymbol{x}\in\mathcal{X}} \|\tilde{\boldsymbol{x}}^i\|^2_{(\boldsymbol{U}_\tau^{\mathcal{H}})^{-1}}} \leq \sqrt{\beta_t(\delta)}\frac{1}{\sqrt{\tau}}\sqrt{2\sum_{t=1}^{\tau}\sum_{i=1}^{N}\max_{\boldsymbol{x}\in\mathcal{X}^i} \|\boldsymbol{x}\|^2_{(\boldsymbol{V}_\tau^{\mathcal{H},i})^{-1}}} \\
&\overset{(iv)}{\leq} \sqrt{\beta_t(\delta)}\frac{1}{\sqrt{\tau}}\sqrt{2\sum_{t=1}^{\tau}\sum_{i=1}^{N}\left\|\boldsymbol{x}_t^i\right\|^2_{(\boldsymbol{V}_\tau^{\mathcal{H},i})^{-1}}} \leq \sqrt{\beta_t(\delta)}\frac{1}{\sqrt{\tau}}\sqrt{\kappa}\sqrt{2\sum_{t=1}^{\tau}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{\kappa}}\boldsymbol{x}_t^i\right\|^2_{(\boldsymbol{V}_\tau^{\mathcal{H},i})^{-1}}} \\
&\overset{(v)}{\leq} \sqrt{\beta_t(\delta)}\frac{1}{\sqrt{\tau}}\sqrt{\kappa}\sqrt{2\sum_{i=1}^{N} d\log(T/\kappa N)} \leq C\sqrt{\frac{Nd\beta_t(\delta)\kappa\log(T/\kappa N)}{\tau}}
\end{aligned}
$$

where (i) follows from an application of Cauchy-Schwarz, (ii) follows from the definition of Θ, (iii) follows from Lemma C.3, (iv) follows from how items in each slot are selected, (v) follows from Lemma E.4 on $\frac{1}{\sqrt{\kappa}}\boldsymbol{x}_t^i$.

Thus, setting $\tau \leq Nd\beta_t(\delta)\kappa\log(T/\kappa N) \leq CS^6N^2d^2\kappa\log(T/\kappa N)\log(T/\delta)$ ensures $\operatorname{diam}_{\mathcal{X}}\left(\Theta\right) \leq 1$.

□

Lemma C.2. *For $t \geq \frac{(N-1)^2}{2\rho^2}\log\frac{dN(N-1)}{\delta}$, we have*

$$
\frac{1}{2}\boldsymbol{U}_t \preccurlyeq \boldsymbol{W}_t \preccurlyeq \frac{3}{2}\boldsymbol{U}_t
$$

Proof. Following the same line of thought as Lemma B.3, Lemma B.4, and Lemma B.5, we have that

$$
\left\|\boldsymbol{W}_t^{i,j}\right\| \leq \sqrt{\frac{t}{2N^2}\log\frac{dN(N-1)}{\delta}}
$$

Following the same line of thought as Lemma B.8 and making use of Assumption C.1, we can derive

$$
\left\|\boldsymbol{Z}_t^{(i)}\right\| \leq \sum_{j=1}^{N-i}\frac{\left\|\boldsymbol{W}_t^{i,j}\right\|}{\sqrt{\lambda_{min}\left(\boldsymbol{W}_t^i\right)\lambda_{min}\left(\boldsymbol{W}_t^j\right)}} \leq \sum_{j=1}^{N-i}\frac{\sqrt{\frac{t}{2N^2}\log\frac{dN(N-1)}{\delta}}}{\rho t} \leq \frac{(N-i)}{N(N-1)}
$$

where the last inequality follows from the fact that $t \geq \frac{(N-1)^2}{2\rho^2}\log\frac{dN(N-1)}{2\delta}$

Finally, using the same line of thought as Lemma B.9, we get

$$
\frac{1}{2}\boldsymbol{U}_t \preccurlyeq \boldsymbol{W}_t \preccurlyeq \frac{3}{2}\boldsymbol{U}_t
$$

□

Lemma C.3. *For* $t \geq \frac{8(N-1)^2}{\kappa^2\rho^2}\log\frac{dN(N-1)}{\delta}$.

$$\frac{1}{2}\boldsymbol{U}_t^{\mathcal{H}} \preccurlyeq \boldsymbol{V}_t^{\mathcal{H}} \preccurlyeq \frac{3}{2}\boldsymbol{U}_t^{\mathcal{H}}$$

Proof. Following the same line of thought as Lemma B.11 and making use of Assumption C.1, we get

$$\left\|\boldsymbol{Z}_t^{(i)}\right\| \leq \sum_{j=1}^{N-i} \frac{\sqrt{\frac{8t}{\kappa^2 N^2}\log\left(\frac{dN(N-1)}{\delta}\right)}}{\rho t} \leq \frac{(N-i)}{N(N-1)}$$

□

where the last inequality follows from the fact that $t \geq \frac{8(N-1)^2}{\kappa^2\rho^2}\log\frac{dN(N-1)}{\delta}$

Finally, we can show that

$$\frac{1}{2}\boldsymbol{U}_t^{\mathcal{H}} \preccurlyeq \boldsymbol{V}_t^{\mathcal{H}} \preccurlyeq \frac{3}{2}\boldsymbol{U}_t^{\mathcal{H}}$$

Lemma C.4. *Define the distribution* $\mathcal{D} = \bigtimes_{i=1}^{N} \mathcal{D}^{TS}$ *where* $\mathcal{D}^{TS}$ *is a multivariate distribution that satisfies the properties given in Definition E.2. Then,* $\mathcal{D}$ *also satisfies the properties given in Definition E.2, making it a suitable distribution for Thompson Sampling.*

Proof. Define $\boldsymbol{\eta} = \left(\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^N\right) \in \mathbb{R}^{Nd}$ where $\boldsymbol{\eta}^i \sim \mathcal{D}^{TS}$. Then, it is easy to see that sampling $\boldsymbol{\eta}^i, i \in [N]$ in an iid fashion from $\mathcal{D}^{TS}$ is the same as sampling $\boldsymbol{\eta}$ from $\mathcal{D}$.

We begin by showing the Concentration property, i.e $\exists C, C'$ such that

$$\mathbb{P}_{\boldsymbol{\eta}\sim\mathcal{D}}\left\{\|\boldsymbol{\eta}\|_2 \leq \sqrt{C(Nd)\log\frac{C'(Nd)}{\delta'}}\right\} \geq 1-\delta'$$

Since $\mathcal{D}^{TS}$ satisfies the concentration property, we know that $\left\|\boldsymbol{\eta}^i\right\|_2 \geq \sqrt{cd\log\frac{c'd}{\delta}}$ with probability at most δ. Hence, it is easy to see that

$$\|\boldsymbol{\eta}\|_2 = \sqrt{\sum_{i=1}^{N}\|\boldsymbol{\eta}_i\|_2^2} \geq \sqrt{cNd\log\frac{c'd}{\delta}}$$

with probability at most δ^N. Setting $C = \frac{c}{N}, C' = \frac{(c')^N d^{N-1}}{N}$ and $\delta' = \delta^N$, we get that

$$\|\boldsymbol{\eta}\|_2 \leq \sqrt{CN^2 d\log\left(\frac{C'Nd}{\delta'}\right)^{1/N}} = \sqrt{C(Nd)\log\frac{C'(Nd)}{\delta'}}$$

with probability at least $1-\delta'$. This proves that $\mathcal{D}$ satisfies the concentration property.

We now show that $\mathcal{D}$ satisfies the Anti-Concentration property, i.e $\exists P \in (0,1)$ such that $\forall \boldsymbol{u} \in \mathbb{R}^{Nd}$:

$$\mathbb{P}_{\boldsymbol{\eta}\in\mathcal{D}}\left\{\boldsymbol{u}^\top\boldsymbol{\eta} \geq \|\boldsymbol{u}\|_2\right\} \geq P$$

Assume $\boldsymbol{u} = \left(\boldsymbol{u}^1, \ldots, \boldsymbol{u}^N\right)$ such that $\|\boldsymbol{u}\|_2 = 1$. This implies that $\sum_{i=1}^{N}\left\|\boldsymbol{u}^i\right\|_2^2 = 1$ which in turn implies that $\left\|\boldsymbol{u}^i\right\|_2 \leq 1$.

Since, $\left\|\boldsymbol{u}^i\right\|_2 \leq 1$, we have that $\left\|\boldsymbol{u}^i\right\|_2^2 \leq \left\|\boldsymbol{u}^i\right\|_2$, and since $\boldsymbol{\eta}^i \sim \mathcal{D}^{TS}$, we have that

$$\mathbb{P}\left\{{\boldsymbol{u}^i}^\top\boldsymbol{\eta}^i \leq \left\|\boldsymbol{u}^i\right\|_2^2\right\} \leq \mathbb{P}\left\{{\boldsymbol{u}^i}^\top\boldsymbol{\eta}^i \leq \left\|\boldsymbol{u}^i\right\|_2\right\} \leq 1-p$$

Hence, we have that

$$\mathbb{P}\left\{\boldsymbol{u}^\top\boldsymbol{\eta} \le \|\boldsymbol{u}\|_2\right\} = \mathbb{P}\left\{\boldsymbol{u}^\top\boldsymbol{\eta} \le \|\boldsymbol{u}\|_2^2\right\} = \mathbb{P}\left\{\sum_{i=1}^N {\boldsymbol{u}^i}^\top\boldsymbol{\eta}^i \le \sum_{i=1}^N \left\|\boldsymbol{u}^i\right\|_2^2\right\} = \mathbb{P}\left\{\bigcap_{i=1}^N \left\{{\boldsymbol{u}^i}^\top\boldsymbol{\eta}^i \le \left\|\boldsymbol{u}^i\right\|_2^2\right\}\right\}$$
$$= \prod_{i=1}^N \mathbb{P}\left\{{\boldsymbol{u}^i}^\top\boldsymbol{\eta}^i \le \left\|\boldsymbol{u}^i\right\|_2^2\right\} \le (1-p)^N$$

Thus, we have that $\mathbb{P}\left\{\boldsymbol{u}^\top\boldsymbol{\eta} \ge \|\boldsymbol{u}\|_2\right\} \ge 1-(1-p)^N$, and setting $P = 1-(1-p)^N$ finishes the claim.

□

Lemma C.5. *At round $t \ge T_0$, let $\tilde{\boldsymbol{\theta}}^i = \boldsymbol{\theta}_t^i + \sqrt{\sigma_t(\delta)}(\boldsymbol{W}_t^i)^{-\frac{1}{2}}\boldsymbol{\eta}^i$ for all $i \in [N]$, where $\boldsymbol{\eta}^i \sim \mathcal{D}^{TS}$, as given in* Steps 7-8 *of Algorithm 4. Define $\tilde{\boldsymbol{\theta}}_t = \left(\tilde{\boldsymbol{\theta}}_t^1, \ldots, \tilde{\boldsymbol{\theta}}_t^N\right)$. Assuming event $\mathcal{E}$ holds, we have that,*

$$\left\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_t\right\|_{\boldsymbol{W}_t} \le C\sqrt{\sigma_t(\delta)}\sqrt{Nd}$$

Proof. We can write $\tilde{\boldsymbol{\theta}}_t = \left(\tilde{\boldsymbol{\theta}}_t^1, \ldots, \tilde{\boldsymbol{\theta}}_t^N\right)$ as the following:

$$\tilde{\boldsymbol{\theta}}_t = \boldsymbol{\theta}_t + \sqrt{\sigma_t(\delta)}\begin{bmatrix}(\boldsymbol{W}_t^1)^{-\frac{1}{2}}\boldsymbol{\eta}^1\\ \vdots \\ (\boldsymbol{W}_t^N)^{-\frac{1}{2}}\boldsymbol{\eta}^N\end{bmatrix} = \boldsymbol{\theta}_t + \sqrt{\sigma_t(\delta)}\text{diag}((\boldsymbol{W}_t^1)^{-\frac{1}{2}}, \ldots, (\boldsymbol{W}_t^N)^{-\frac{1}{2}})\boldsymbol{\eta} = \boldsymbol{\theta}_t + \sqrt{\sigma_t(\delta)}\boldsymbol{U}_t^{-\frac{1}{2}}\boldsymbol{\eta}$$

where $\boldsymbol{\eta} = \left(\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^N\right)$.

Thus, we get

$$\left\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_t\right\|_{\boldsymbol{W}_t} = \sqrt{\sigma_t(\delta)}\left\|\boldsymbol{U}_t^{-\frac{1}{2}}\boldsymbol{\eta}\right\|_{\boldsymbol{W}_t} \overset{(i)}{\le} \frac{3}{2}\sqrt{\sigma_t(\delta)}\left\|\boldsymbol{U}_t^{-\frac{1}{2}}\boldsymbol{\eta}\right\|_{\boldsymbol{U}_t} = \frac{3}{2}\sqrt{\sigma_t(\delta)}\left\|\boldsymbol{\eta}\right\|_2 \overset{(ii)}{\le} C\sqrt{\sigma_t(\delta)}\sqrt{Nd}$$

where (i) follows from Lemma C.2 and (ii) follows from the concentration property shown in Lemma C.4.

□

D CONCENTRATION RESULTS FOR RANDOM MATRICES AND VECTORS

Lemma D.1. *(Chatterji et al. [2020], Generalization of Lemma 7) Let $\{\boldsymbol{x}_s\}_{s=1}^\top$ be a stochastic process in $\mathbb{R}^d$ such that for filtration $\mathcal{F}_t$, we have that $\mathbb{E}\left[\boldsymbol{x}_s|\mathcal{F}_{s-1}\right] = \boldsymbol{0}_d$ and $\mathbb{E}\left[\boldsymbol{x}_s\boldsymbol{x}_s^\top|\mathcal{F}_{s-1}\right] \succcurlyeq \rho\boldsymbol{I}_d$. Further, let $\|\boldsymbol{x}_s\|_2 \le m$ for all $s \ge 1$. Also, define the matrix*

$$\boldsymbol{Q}_t = \gamma\boldsymbol{I}_d + \sum_{s=1}^t \boldsymbol{x}_s\boldsymbol{x}_s^\top$$

Then, with probability atleast $1-\delta$, we have that

$$\lambda_{min}\left(\boldsymbol{Q}_t\right) \ge \gamma + c\rho t$$

for $0 \le c \le 1$ and for all t such that $\frac{12m^4+4m^2\rho(1-c)}{3(1-c)^2\rho^2}\log\left(\frac{2dT}{\delta}\right) \le t \le T$

Proof. The proof follows on the same lines as that of Chatterji et al. [2020].

Assume $\mathbb{E}\left[\boldsymbol{x}_s\boldsymbol{x}_s^\top|\mathcal{F}_{s-1}\right] = \boldsymbol{\Sigma}_c \succcurlyeq \rho\boldsymbol{I}_d$. Define the matrix martingale $\boldsymbol{Z}_s = \sum_{s=1}^t \left[\boldsymbol{x}_s\boldsymbol{x}_s^\top - \boldsymbol{\Sigma}_c\right]$ with $\boldsymbol{Z}_0 = 0$ and the corresponding martingale difference sequence $\boldsymbol{X}_s = \boldsymbol{Z}_s - \boldsymbol{Z}_{s-1}$ for all $s \ge 1$.

We have that $\|\boldsymbol{x}_s\|_2 \leq m$. Also, $\|\boldsymbol{\Sigma}_c\| = \left\|\mathbb{E}\left[\boldsymbol{x}_s\boldsymbol{x}_s^\top|\mathcal{F}_{t-1}\right]\right\| \leq \|\boldsymbol{x}_s\|_2^2 \leq m^2$

Therefore, using triangle inequality, $\|\boldsymbol{X}_s\| = \left\|\boldsymbol{x}_s\boldsymbol{x}_s^\top - \boldsymbol{\Sigma}_c\right\| \leq \left\|\boldsymbol{x}_s\boldsymbol{x}_s^\top\right\| + \|\boldsymbol{\Sigma}_c\| \leq 2m^2$

Finally, we have that

$$\begin{aligned}
\sum_{s=1}^{t}\left\|\mathbb{E}\left[\boldsymbol{X}_s\boldsymbol{X}_s^\top|\mathcal{F}_{s-1}\right]\right\| &= \sum_{s=1}^{t}\left\|\mathbb{E}\left[\boldsymbol{X}_s^\top\boldsymbol{X}_s|\mathcal{F}_{s-1}\right]\right\| \\
&= \sum_{s=1}^{t}\left\|\mathbb{E}\left[\boldsymbol{x}_s\boldsymbol{x}_s^\top\boldsymbol{x}_s\boldsymbol{x}_s^\top - \boldsymbol{x}_s\boldsymbol{x}_s^\top\boldsymbol{\Sigma}_c^\top - \boldsymbol{\Sigma}_c\boldsymbol{x}_s\boldsymbol{x}_s^\top + \boldsymbol{\Sigma}_c\boldsymbol{\Sigma}_c^\top|\mathcal{F}_{s-1}\right]\right\| \\
&\leq \sum_{s=1}^{t}\left\|\mathbb{E}\left[\left(\boldsymbol{x}_s^\top\boldsymbol{x}_s\right)\boldsymbol{x}_s\boldsymbol{x}_s^\top + \boldsymbol{\Sigma}_c\boldsymbol{\Sigma}_c^\top|\mathcal{F}_{s-1}\right]\right\| \\
&\leq 2m^4t
\end{aligned}$$

Thus, applying the Matrix Freedman Inequality (Lemma E.2) with $R = 2m^2, \omega^2 = 2m^4t, {}_1^d = d_2 = d$ and $u = (1-c)\rho t$, we get

$$\mathbb{P}\left\{\left\|\sum_{s=1}^{t}\left[\boldsymbol{x}_s\boldsymbol{x}_s^\top - \boldsymbol{\Sigma}_c\right]\right\| \geq (1-c)\rho t\right\} \leq 2d\exp\left(-\frac{(1-c)^2\rho^2t^2/2}{2m^4t + 2m^2(1-c)\rho t/3}\right)$$

Choosing $t \geq \frac{12m^4+4m^2\rho(1-c)}{3(1-c)^2\rho^2}\log\left(\frac{2dT}{\delta}\right)$, we get that with probability at least $1-\frac{\delta}{T}$,

$$(1-c)\rho t \geq \left\|\sum_{s=1}^{t}\left[\boldsymbol{x}_s\boldsymbol{x}_s^\top - \boldsymbol{\Sigma}_c\right]\right\|$$

Now, recall the definition of the norm: $\|\boldsymbol{A}\| = \sup\limits_{\|\boldsymbol{y}\|_2\leq 1}\boldsymbol{A}\boldsymbol{y}$. Substituting this definition results in:

$$(1-c)\rho t \geq \sup_{\|y\|_2\leq 1}\left[\left(\sum_{s=1}^{t}\boldsymbol{x}_s\boldsymbol{x}_s^\top\right)\boldsymbol{y} - t\boldsymbol{\Sigma}_c\boldsymbol{y}\right] \geq \left|\inf_{\|y\|_2\leq 1}\left(\sum_{s=1}^{t}\boldsymbol{x}_s\boldsymbol{x}_s^\top\right)\boldsymbol{y} - t\cdot\inf_{\|y\|_2\leq 1}\boldsymbol{\Sigma}_c\boldsymbol{y}\right|$$

which uses the inequality $\sup\limits_A|f-g| \geq \left|\inf\limits_A f - \inf\limits_A g\right|$. Now, using Rayleigh's quotient, we also know that $\inf\limits_{\|\boldsymbol{y}\|_2\leq 1}\boldsymbol{A}\boldsymbol{y} = \lambda_{min}\left(\boldsymbol{A}\right)$. Thus,

$$(1-c)\rho t \geq \left|\lambda_{min}\left(\sum_{s=1}^{t}\boldsymbol{x}_s\boldsymbol{x}_s^\top\right) - t\lambda_{min}\left(\boldsymbol{\Sigma}_c\right)\right| \implies \lambda_{min}\left(\sum_{s=1}^{t}\boldsymbol{x}_s\boldsymbol{x}_s^\top\right) \geq c\rho t$$

using the fact that $\boldsymbol{\Sigma}_c \succcurlyeq \rho\boldsymbol{I}$. This holds with probability $1-\frac{\delta}{T}$. Performing a union bound over all time indices finishes the claim.

□

Lemma D.2. *(Das and Sinha [2024], Lemma 17) Let $\delta \in (0,1)$, $\boldsymbol{x}_s \in \mathbb{R}^{d_1}$ and $\boldsymbol{z}_s \in \mathbb{R}^{d_2}$ such that $\mathbb{E}\left[\boldsymbol{x}_s\boldsymbol{z}_s^\top|\mathcal{F}_{s-1}\right] = \boldsymbol{0}_{d_1\times d_2}$. Define $\boldsymbol{M}_t = \sum\limits_{s=1}^{t}\boldsymbol{x}_s\boldsymbol{z}_s^\top$. Further, assume that $\|\boldsymbol{x}_s\|_2 \leq m_1$ and $\|\boldsymbol{z}_s\|_2 \leq m_2$. Then, with probability at least $1-\delta$*

$$\|M_t\| \leq 2(m_1\wedge m_2)^2\sqrt{2t\log\left(\frac{d_1+d_2}{\delta}\right)}$$

Proof. Denote $\boldsymbol{X}_s = \boldsymbol{x}_s \boldsymbol{z}_s^\top$. Since $\mathbb{E}\left[\boldsymbol{X}_s | \mathcal{F}_{s-1}\right] = \boldsymbol{0}_{d_1 \times d_2}$, $\boldsymbol{X}_s$ is a Martingale Difference sequence. Further, $\boldsymbol{M}_t = \sum_{s=1}^{t} \boldsymbol{X}_s$ is the sum of Martingale Difference Sequences.

Consider the square of the Hermitian Dilation (see Definition E.1) of $\boldsymbol{X}_s$

$$\begin{aligned}
\mathcal{H}(\boldsymbol{X}_s)^2 = \begin{bmatrix} \boldsymbol{0}_{d_1 \times d_1} & \boldsymbol{X}_s \\ \boldsymbol{X}_s^\top & \boldsymbol{0}_{d_2 \times d_2} \end{bmatrix}^2 &= \begin{bmatrix} \boldsymbol{X}_s \boldsymbol{X}_s^\top & \boldsymbol{0}_{d_1 \times d_2} \\ \boldsymbol{0}_{d_2 \times d_1} & \boldsymbol{X}_s^\top \boldsymbol{X}_s \end{bmatrix} \\
&= \begin{bmatrix} \|\boldsymbol{z}_s\|_2^2 \, \boldsymbol{x}_s \boldsymbol{x}_s^\top & \boldsymbol{0}_{d_1 \times d_2} \\ \boldsymbol{0}_{d_2 \times d_1} & \|\boldsymbol{x}_s\|_2^2 \, \boldsymbol{z}_s \boldsymbol{z}_s^\top \end{bmatrix} \\
&\preccurlyeq (m_1 \wedge m_2)^2 \begin{bmatrix} \boldsymbol{x}_s \boldsymbol{x}_s^\top & \boldsymbol{0}_{d_1 \times d_2} \\ \boldsymbol{0}_{d_2 \times d_1} & \boldsymbol{z}_s \boldsymbol{z}_s^\top \end{bmatrix} \\
&\preccurlyeq (m_1 \wedge m_2)^4 \boldsymbol{I}_{d_1 + d_2}
\end{aligned}$$

Applying the Matrix Azuma inequality (Lemma E.3) with $\boldsymbol{A}_s = (m_1 \wedge m_2)^2 \boldsymbol{I}_{d_1+d_2}$, we have that $\sigma_t^2 = (m_1 \wedge m_2)^4 t$ and thus,

$$\mathbb{P}\left\{\exists t \geq 1 : \sigma_{max}\left(\boldsymbol{M}_t\right) \geq \epsilon\right\} \leq (d_1 + d_2) \exp\left(-\frac{\epsilon^2}{8(m_1 \wedge m_2)^4 t}\right)$$

Choosing $\epsilon = \sqrt{8(m_1 \wedge m_2)^4 t \log\left(\frac{d_1+d_2}{\delta}\right)}$ finishes the proof.

□

E OTHER USEFUL RESULTS AND DEFINITIONS

Definition E.1. (Hermitian Dilation) The Hermitian matrix for a matrix $\boldsymbol{A}$ is defined as

$$\mathcal{H}(\boldsymbol{A}) = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{A} \\ \boldsymbol{A}^\top & \boldsymbol{0} \end{bmatrix}$$

Lemma E.1. *(Das and Sinha [2024], Lemma 16) Let* $\boldsymbol{\mathcal{H}(Z)} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{Z} \\ \boldsymbol{Z}^\top & \boldsymbol{0} \end{bmatrix}$ *where* $\boldsymbol{Z}$ *has positive singular values. Then, it holds almost surely,* $\lambda_{max}\left(\mathcal{H}(\boldsymbol{Z})\right) = -\lambda_{min}\left(\mathcal{H}(\boldsymbol{Z})\right) = \sigma_{max}\left(\boldsymbol{Z}\right)$

Lemma E.2. *(Matrix Freedman Inequality Tropp [2011a] Corollary 1.3) Define a matrix martingale* $\boldsymbol{Z}_s \in \mathbb{R}^{d_1 \times d_2}$ *with respect to the filtration* $\mathcal{F}_s$ *and a martingale difference sequence* $\boldsymbol{X}_s = \boldsymbol{Z}_s - \boldsymbol{Z}_{s-1}$*. Assume that the difference sequence is almost surely uniformly bounded, i.e* $\|\boldsymbol{X}_s\| \leq R$*. Define the quantities*

$$\boldsymbol{W}_{row,t} = \sum_{s=1}^{t} \mathbb{E}\left[\boldsymbol{X}_s \boldsymbol{X}_s^\top | \mathcal{F}_{s-1}\right]$$

$$\boldsymbol{W}_{col,t} = \sum_{s=1}^{t} \mathbb{E}\left[\boldsymbol{X}_s^\top \boldsymbol{X}_s | \mathcal{F}_{s-1}\right]$$

Then, for all $u \geq 0$ *and* $\omega^2 > 0$*, we have*

$$\mathbb{P}\left\{\exists t \geq 0 : \|\boldsymbol{Z}_t\| \geq u \text{ and } \max\left\{\|\boldsymbol{W}_{row,t}\|, \|\boldsymbol{W}_{row,t}\|\right\} \leq \omega^2\right\} \leq (d_1 + d_2) \exp\left(-\frac{u^2/2}{\omega^2 + Ru/3}\right)$$

Lemma E.3. *(Matrix Azuma Inequality, Tropp [2011b], Theorem 7.1) Let* $\{\boldsymbol{X}_s\}_{s=1}^{\infty}$ *be a matrix martingale difference sequence in* $\mathbb{R}^{d_1 \times d_2}$ *and let* $\mathcal{H}(\boldsymbol{X}_s)$ *represent the Hermitian Dilation (see def. E.1) of* $\boldsymbol{X}_s$*. Let* $\{\boldsymbol{A}_s\}_{s=1}^{\infty}$ *be a sequence of*

matrices in $\mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$ *such that* $\mathbb{E}\left[\boldsymbol{X}_s|\mathcal{F}_{s-1}\right] = \mathbf{0}$ *and* $\mathcal{H}(\boldsymbol{X}_s)^2 \preccurlyeq \boldsymbol{A}_s^2$. *Let* $\sigma_t^2 = \lambda_{max}\sum_{s=1}^{t}\boldsymbol{A}_k^2$ *for* $t \geq 1$. *Then, for all* $\epsilon \geq 0$*:*

$$\mathbb{P}\left\{\exists t \geq 1 : \sigma_{max}\left(\sum_{s=1}^{t}\boldsymbol{X}_s\right) \geq \epsilon\right\} \leq (d_1+d_2)\exp\left(-\frac{\epsilon^2}{8\sigma_t^2}\right)$$

Lemma E.4. *(Elliptical Potential Lemma, Abbasi-yadkori et al. [2011], Lemma 11) Let* $\{\boldsymbol{x}_s\}_{s=1}^{\top}$ *represent a set of vectors in* $\mathbb{R}^d$ *and let* $\|\boldsymbol{x}_s\|_2 \leq L$. *Let* $\boldsymbol{V}_s = \lambda\boldsymbol{I}_d + \sum_{m=1}^{s-1}\boldsymbol{x}_m\boldsymbol{x}_m^{\top}$. *Then, for* $\lambda \geq 1$

$$\sum_{s=1}^{t}\|\boldsymbol{x}_s\|^2_{\boldsymbol{V}_s^{-1}} \leq 2d\log\left(1+\frac{tL^2}{\lambda d}\right) \leq 4d\log(tL^2)$$

Lemma E.5. *(Abeille et al. [2021], Proposition 7) Let* $b, c \geq 0$ *and* $x^2 - bx - c \leq 0$. *Then,* $x^2 \leq 2b^2 + 2c$.

Proof. Since the coefficient of the quadratic term is 1, the quadratic expression can attain non-positive values only if it has two distinct or equal real roots. We denote the roots by α_1 and α_2. Without loss of generality, assume $\alpha_1 = \frac{b-\sqrt{b^2+4c}}{2}$ and $\alpha_2 = \frac{b+\sqrt{b^2+4c}}{2}$. Then, the set of x for which $x^2 - bx - c \leq 0$ is true is $x \in [\alpha_1, \alpha_2]$. Thus, we can say

$$x \leq \alpha_2 = \frac{b+\sqrt{b^2+4c}}{2} \leq b + \sqrt{c}$$

using the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \geq 0$. Finally,

$$x^2 \leq b^2 + c + 2b\sqrt{c} \leq 2b^2 + 2c$$

using the fact that $(b-\sqrt{c})^2 \geq 0 \implies 2b\sqrt{c} \leq b^2 + c$

□

Definition E.2. (Multivariate distribution for Thompson Sampling, Abeille and Lazaric [2017], Definition 1) $\mathcal{D}^{TS}$ is a suitable multivariate distribution on $\mathbb{R}^d$ for Thompson Sampling if it is absolutely continuous with respect to the Lebesgue measure and satisfies the following properties:

1. Concentration: There exist constants c and $c^{'}$ such that $\forall\delta \in (0,1)$

$$\mathbb{P}_{\boldsymbol{\eta}\sim\mathcal{D}^{TS}}\left\{\|\boldsymbol{\eta}\|_2 \leq \sqrt{cd\log\frac{c^{'}d}{\delta}}\right\} \geq 1-\delta$$

2. Anti-Concentration: There exists a strictly positive probability p such that for any $\boldsymbol{u} \in \mathbb{R}^d$

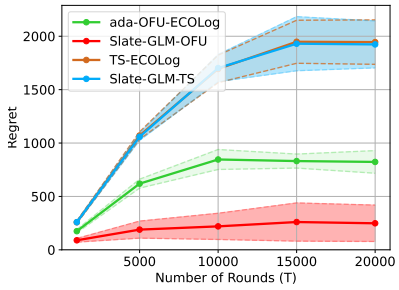
$$\mathbb{P}_{\boldsymbol{\eta}\sim\mathcal{D}^{TS}}\left\{\boldsymbol{u}^{\top}\boldsymbol{\eta} \geq \|\boldsymbol{u}\|_2\right\} \geq p$$

F ADDITIONAL EXPERIMENTS AND EXPERIMENTAL DETAILS

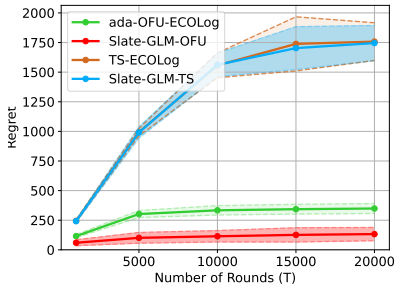
In this section, we provide additional plots to back the experiments shown in Section 5. Also, we provide additional details about our experimental setup.

In all of the figures, the shaded regions represent two standard deviations. Figures 2a and 2b depict the graphs from **Experiment 1** (Section 5) wherein we compare our algorithms `Slate-GLM-OFU` and `Slate-GLM-TS` to their counterparts `ada-OFU-ECOLog` and `TS-ECOLog` in the finite and infinite context settings.

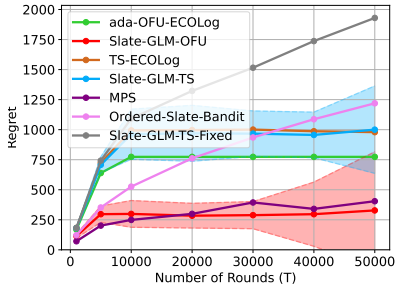
Figures 2c and 2d depict the graphs from **Experiment 3**(Section 5), wherein we compare our algorithms `Slate-GLM-OFU`, `Slate-GLM-TS`, and `Slate-GLM-TS-Fixed` to several state-of-the-art non-contextual logistic bandit algorithms. In Figure 2c, we only show the uncertainity involved in `Slate-GLM-OFU` and `Slate-GLM-TS`. We see that



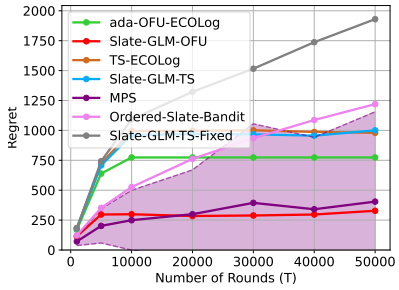
(a) Regret vs. T: Finite Context Setting



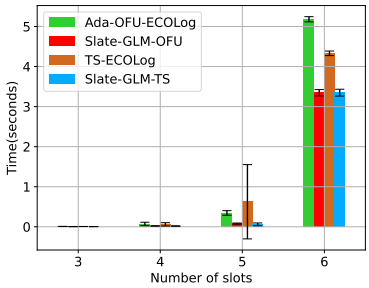
(b) Regret vs. T: Infinite Context Setting



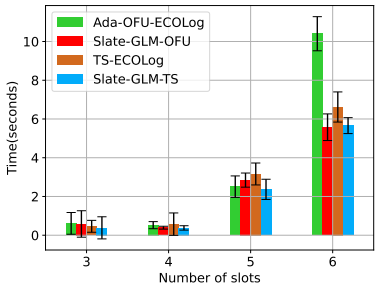
(c) Regret vs. T: Fixed-Arm Setting



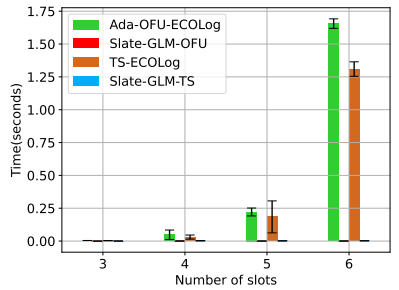
(d) Regret vs. T: Fixed-Arm Setting



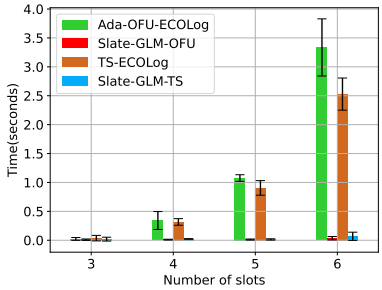
(e) Average running time (per-round)



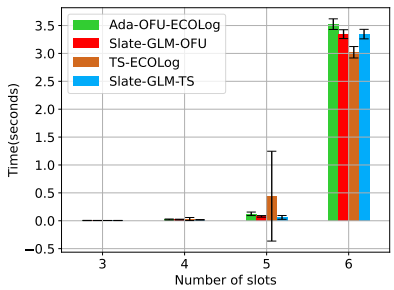
(f) Maximum running time (per-round)



(g) Average time taken to pull an arm



(h) Maximum time taken to pull an arm



(i) Average time taken to update parameters

Figure 2

`Slate-GLM-OFU` has the best performance, with the only algorithm having comparable performance being `MPS`. On the other hand, `Slate-GLM-TS` performs worse than `ada-OFU-ECOLog` and `MPS`, while being on par with `TS-ECOLog`. However, in Figure 2d, we showcase that the variance of `MPS` is very high, hence, making the algorithm less reliable in practice.

Figures 2e and 2f showcase two standard deviations in the average and maximum (per-round) running time of the algorithms. We see that both `ada-OFU-ECOLog` and `TS-ECOLog` show an exponential increase in their running times. Further, the significant gap between the average and maximum (per-round) running times of `Slate-GLM-OFU` and `Slate-GLM-TS` (as highlighted in the table below) indicates that the true per-round time is much lower than the maximum. As we have mentioned in the main paper, we calculate the per-round running time for an algorithm as the sum of the per-round pull and update times. Figures 2g and 2h show the average and maximum pull times (per round), while Figure 2i display the average per-round update times. We see that the pull time for `ada-OFU-ECOLog` and `TS-ECOLog` increases exponentially with the number of slots, whereas the update times remain similar for all algorithms. Hence, the differences in per-round running times can be majorly attributed to the pulling times for each algorithm, which is in line with our theoretical claims. We also tabulate the average and maximum per-round pulling times for each algorithm in Table 2 for more clarity.

Slots	`ada-OFU-ECOLog`		`Slate-GLM-OFU`		`TS-ECOLog`		`Slate-GLM-TS`	
	Average (ms)	**Maximum (ms)**	**Average (ms)**	**Maximum (ms)**	**Average (ms)**	**Maximum (ms)**	**Average (ms)**	**Maximum (ms)**
3	4.3 ± 0.2	23.0 ± 24.5	0.3 ± 0.0	9.5 ± 12.5	3.1 ± 0.1	36.6 ± 47.7	0.6 ± 0.1	19.2 ± 33.4
4	47.5 ± 36.4	341.8 ± 154.8	0.8 ± 0.9	10.5 ± 7.3	30.3 ± 15.6	316.7 ± 57.1	2.2 ± 1.1	22.8 ± 5.4
5	221.4 ± 30.1	1075.7 ± 57.8	0.6 ± 0.2	12.0 ± 9.7	184.1 ± 121.8	905.3 ± 126.5	1.2 ± 0.5	13.8 ± 11.5
6	1655.6 ± 36.3	3335.5 ± 494.6	0.9 ± 0.2	35.8 ± 28.9	1309.4 ± 55.3	2528.3 ± 278.2	1.9 ± 0.2	68.3 ± 71.1

Table 2: Average and Maximum per-round running times (in milliseconds), averaged over 10 different seeds for sampling rewards, with 2 standard deviations

Now, we provide additional details about our experimental setup. In **Experiment 3**, we implement `Ordered Slate Bandit` and `ETC-Slate` from Kale et al. [2010] and Rhuggenaath et al. [2020] respectively. Since these algorithms are designed for semi-bandit feedback, we make modifications to implement these algorithms in our setting. These modifications are detailed below:

`Ordered Slate Bandit`: The original algorithm in Kale et al. [2010] assumes that there exists a base set $\mathcal{X}$ such that $|\mathcal{X}| = K$ and the learner picks a slate of N items from $\mathcal{X}$. Hence, their algorithm assumes that each base item is equally likely to be placed in any slot. Thus, they start with the initial distribution P such that $P_{i,j} = 1 \; \forall i \in [N] \; \forall j \in [K]$. On the other hand, we cannot make the same assumption since we get a different set of items $\mathcal{X}_t^i$ for each slot $i \in [N]$. Thus, we change the initial distribution to P such that $P_{i,j} = 1$ if and only if $j \in [K(i-1)+1, K(i)]$. This modification restricts the items that can be selected for a particular slot. A similar modification is made for the exploratory distribution in each round. There is a significant difference in the manner in which the loss matrix is constructed. Since the algorithm is designed for semi-bandit feedback, the algorithm propagates the loss for the item chosen in each slot at each round. We make use of the fact that the loss is the additive inverse of the reward, and hence, we have two choices for the loss we wish to propagate. Since we operate in the logistic setting, the obvious choice is to propagate the non-linear losses to the algorithm. However, since the total loss for a slate is assumed to be the sum of the loss obtained for each slot, the linear loss seems more suitable. We experiment with both these choices, and find that the algorithm with non-linear losses incurs very high regret. Hence, we only compare our algorithms to the Ordered Slate Bandit algorithm with linear losses, referred to as `Ordered Slate Bandit`.

`ETC-Slate`: The original algorithm in Rhuggenaath et al. [2020] is also designed for semi-bandit feedback, wherein, it is assumed that the reward for each slot is sampled from a distribution such as the uniform distribution (see Example 1 in Rhuggenaath et al. [2020]). However, in our case, we do not have a notion of a reward distribution at the slot level. Hence, to create a reward distribution at the slot level, we assume that the reward for slot i is sampled from $\mathcal{N}({\boldsymbol{x}_s^i}^\top \boldsymbol{\theta}_\star^i, 0.0001)$. This ensures that, in expectation, the reward attributed to a particular slot is the linear reward for the item played. We set the slate-level reward function f to simply be the sigmoid function applied to the sum of the rewards obtained at the slot levels and then proceed with the algorithm. We find that `ETC-Slate` incurs very high regret, and hence, do not include the algorithm in our comparisions.

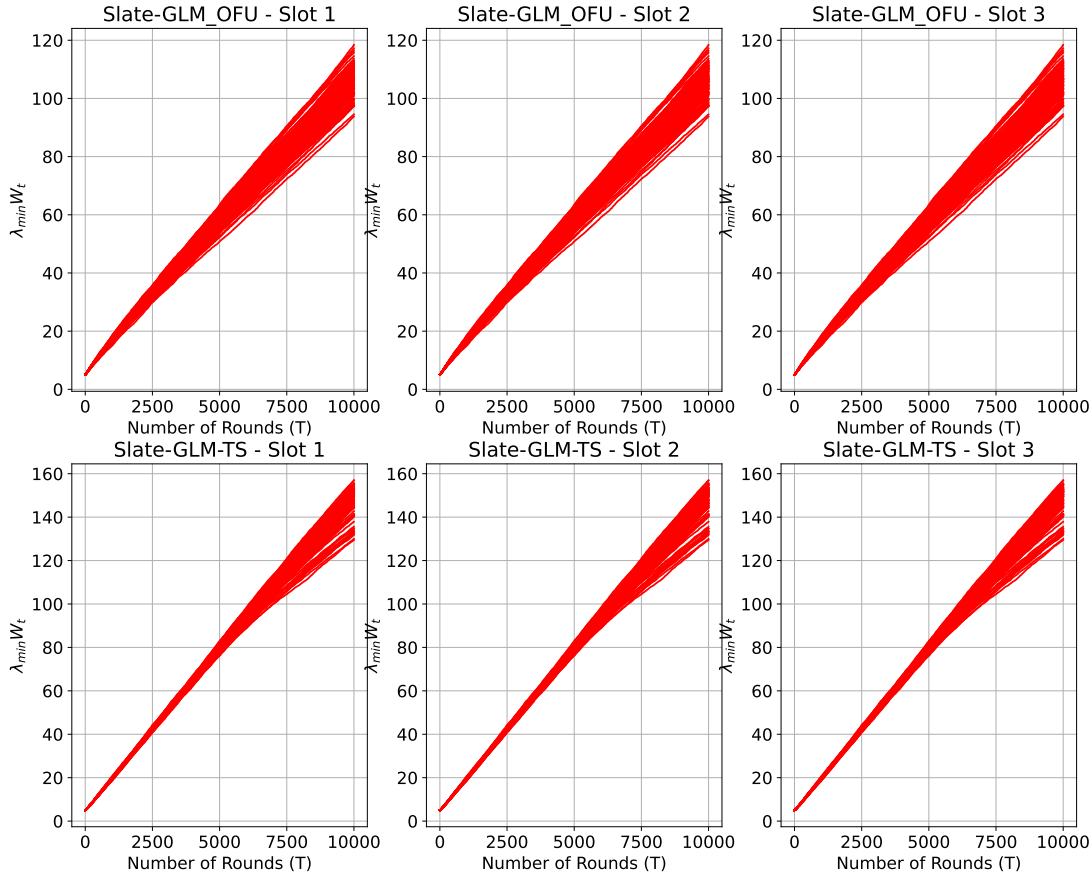


Figure 3: Demonstration of the algorithm-dependent assumption for `Slate-GLM-OFU` and `Slate-GLM-TS` wherein we plot the minimum eigenvalues of $\boldsymbol{W}_t^i$ as a function of the time round for 100 independent runs

G EMPERICAL VALIDATION OF THE DIVERSITY ASSUMPTION (ASSUMPTION 2.1)

In this section, we show that our (instance and algorithm dependent) diversity assumption we make indeed holds for a lot of instances. We choose the number of slots N to be 3 and the number of items in each slot $\left|\mathcal{X}_t^i\right|$ is fixed to 5. The dimension of items for each slot is fixed to 5, resulting in the slate having a dimension $d = 15$. The items for each slot are randomly sampled from $[-1, 1]^5$ and normalized to have norm $1/\sqrt{3}$, while $\boldsymbol{\theta}_\star$ is randomly sampled from $[-1, 1]^{15}$. We operate in the Infinite context setting, wherein the items in each slot change every time round (check **Experiment 1** in Section 5 for more details). We run both `Slate-GLM-OFU` and `Slate-GLM-TS` 100 times with different seeds for a horizon of $T = 10000$ rounds. For each run of the algorithm, we plot the minimum eigenvalue of $\boldsymbol{W}_t^i$ for $i \in [3]$ as a function of the time round t and show our results in Figure 3. The figures clearly depict a (near) linear growth in the eigenvalues of the matrices $\boldsymbol{W}_t^i$ for all the slots $i \in [3]$ and all rounds $t \in [T]$.