Thompson Sampling for High-Dimensional Sparse Linear Contextual Bandits

Sunrit Chakraborty [*1] **Saptarshi Roy** [*1] **Ambuj Tewari** [1]

Abstract

We consider the stochastic linear contextual bandit problem with high-dimensional features. We analyze the Thompson sampling algorithm using special classes of sparsity-inducing priors (e.g., spike-and-slab) to model the unknown parameter and provide a nearly optimal upper bound on the expected cumulative regret. To the best of our knowledge, this is the first work that provides theoretical guarantees of Thompson sampling in high-dimensional and sparse contextual bandits. For faster computation, we use variational inference instead of Markov Chain Monte Carlo (MCMC) to approximate the posterior distribution. Extensive simulations demonstrate the improved performance of our proposed algorithm over existing ones.

1. Introduction

Sequential decision-making, including bandits problems and reinforcement learning, has been one of the most active areas of research in machine learning. It formalizes the idea of selecting actions based on current knowledge to optimize some long term reward over sequentially collected data. On the other hand, the abundance of personalized information allows the learner to make decisions while incorporating this contextual information, a setup that is mathematically formalized as contextual bandits. Moreover, in the big data era, the personal information used as contexts often has a much larger size, which can be modeled by viewing the contexts as high-dimensional vectors. Examples of such models cover internet marketing and treatment assignment in personalized medicine, among many others.

A particularly interesting special case of the contextual bandit problem is the linear contextual bandit problem, where the expected reward is a linear function of the features (Abe et al., 2003; Auer, 2002). Under this setting, (Dani et al., 2008), (Chu et al., 2011) and (Abbasi-Yadkori et al., 2011) showed polynomial dependence of the cumulative regret on ambient dimension d and time horizon T in low dimensional case. Specifically, (Dani et al., 2008) and (Abbasi-Yadkori et al., 2011) proved a regret upper bound scaling as $O(d\sqrt{T})$, while (Chu et al., 2011) showed a regret upper bound of the order $O(\sqrt{dT})$. It is worthwhile to mention that all of the aforementioned algorithms fall under a certain class of algorithms known as upper confidence bound (UCB) type algorithms that rely on the construction of a specific confidence set for the unknown parameter. In contrast, Thompson Sampling (TS) maintains uncertainty about the unknown parameter in the form of a posterior distribution. The first TS algorithm under this setting was proposed by (Agrawal & Goyal, 2013) where they established a regret bound of the order $O(d^2\delta^{-1}\sqrt{T^{1+\delta}})$ for any $\delta \in (0, 1)$.

There is also a large body of work present in high-dimensional sparse linear contextual bandit setup, where the reward only depends on a small subset of features of the observed contexts. This area has recently attracted considerable attention due to its abundance in modern reinforcement learning applications (e.g, clinical trials, personalized recommendation systems, etc.) and has quite naturally spawned theoretical research in this direction. Some of the important references include (Bastani & Bayati, 2020; Wang et al., 2018; Hao et al., 2020; Chen et al., 2022; Ariu et al., 2022; Kim & Paik, 2019; Li et al., 2022; Oh et al., 2021; Li et al., 2021) among others. A more detailed discussion on existing literature in the high dimensional bandit field is provided in Section 3.3. However, there has been very limited work dedicated to analyzing TS algorithms in high-dimensional sparse bandit setups. (Hao et al., 2021) proposed a sparse information-directed sampling (IDS) algorithm which under a special case reduces to a TS algorithm based on a spike-and-slab Gaussian-Laplace prior. However, the regret bound of IDS scales polynomially in d, which is sub-optimal in the high-dimensional regime. In related work, (Gilton & Willett, 2017) proposed a linear TS algorithm based on a relevance vector machine (RVM) which again suffers from sub-optimal dependence on d.

In this paper, we specifically focus on the high-dimensional sparse linear contextual bandit (SLCB) setup and propose a TS algorithm based on a sparsity-inducing prior that en-

[*]Equal contribution [1]Department of Statistics, University of Michigan, Ann Arbor, USA. Correspondence to: Saptarshi Roy <roysapta@umich.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

joys almost dimension-independent regret bound. While TS algorithms have been known to empirically perform better than optimism-based algorithms (Chapelle & Li, 2011; Kaufmann et al., 2012), theoretical understanding of these is challenging due to the complex dependence structure of the bandit environment. Moreover, posterior sampling in high-dimensional regression (which is the crucial step for TS in high-dimensional SLCB), using MCMC, generally suffers from computational bottleneck. Our work overcomes all these challenges and makes the following contributions:

1. We use the sparsity inducing prior proposed in (Castillo et al., 2015) for posterior sampling and establish posterior contraction result for *non-i.i.d. observations* coming from bandit environment and for a wide class of noise distributions.
2. Using the posterior contraction result, we establish an almost *dimension free* regret bound for our proposed TS algorithm under different arm-separation regimes parameterized by ω. The algorithm enjoys minimax optimal performance for $\omega \in [0, 1)$. To the best of our knowledge, this is the first work that proposes a novel TS algorithm with desirable regret guarantees in high-dimensional and sparse SLCB setup.
3. Our algorithm does *not* need the knowledge of model sparsity level, unlike other algorithms such as LASSO-bandit, MCP-bandit, ESTC, etc.
4. Finally, the prior allows us to design a computationally efficient TS algorithm based on Variational Bayes.

The rest of the paper is organized as follows. In Section 2, we introduce the problem formally and discuss the assumptions and prior distribution. In Section 3, we present the crucial posterior contraction result and the main regret bound for our proposed algorithm. In Section 4, we discuss the challenges of drawing samples from the posterior in such problems and present a faster alternative relying on variational inference. In Section 5, we present simulation studies under different setups comparing our proposed method with existing algorithms. Detailed proofs of results and technical lemmas are deferred to the appendix.

Notation: Let $\mathbb{R}$ and $\mathbb{R}^+$ denote the set of real numbers and the set of non-negative real numbers respectively. Denote by $\mathbb{R}^p$ the p-dimensional Euclidean space and $\mathbb{S}^{p-1} := \{x \in \mathbb{R}^p \mid \|x\|_2 = 1\}$ the $(p-1)$-dimensional unit sphere. For a positive integer K, denote by $[K]$ the set $\{1, 2, \ldots, K\}$. Regarding vectors and matrices, for a vector $v \in \mathbb{R}^p$, we denote by $\|v\|_0, \|v\|_1, \|v\|_2, \|v\|_\infty$ the $\ell_0, \ell_1, \ell_2, \ell_\infty$ norms of v respectively. We use $\mathbb{I}_p$ to denote the p-dimensional identity matrix. For a $p \times q$ matrix M we define the $\|M\| := \max_{j\in[q]} \sqrt{(M^\top M)_{j,j}}$.

Regarding distributions, $\mathsf{N}(\mu, \sigma)$ denotes the Gaussian distribution with mean μ and standard deviation σ, $\mathsf{Lap}(\lambda)$ denotes the Laplace distribution with density $f(x) = (\lambda/2)\exp(-\lambda|x|)$, and $\mathsf{Beta}(a, b)$ denotes the beta distribution with parameters a, b. Regarding random variables, $\|Z\|_{\psi_2}$ denotes the Orlicz norm of the reandom variable Z, i.e., $\|Z\|_{\psi_2} = \inf\{\lambda > 0 : \mathbb{E}\exp(Z^2/\lambda^2) \le 2\}$. For a random vector $X \in \mathbb{R}^p$, the corresponding Orlicz norm is defined as $\|X\|_{\psi_2} := \sup_{u\in\mathbb{S}^{p-1}} \left\|u^\top X\right\|_{\psi_2}$.

Throughout the paper, let $O(\cdot)$ denote the standard big-O notation, i.e., we say $a_n = O(b_n)$ if there exists a universal constant $C > 0$, such that $a_n \le Cb_n$ for all $n \in \mathbb{N}$. Sometimes for notational convenience, we write $a_n \lesssim b_n$ in place of $a_n = O(b_n)$. We write $a_n \asymp b_n$ or $a_n = \Theta(b_n)$ if $a_n = O(b_n)$ and $b_n = O(a_n)$.

2. Problem Formulation

We consider a linear stochastic contextual bandit with K arms. At time $t \in [T]$, context vectors $\{x_i(t)\}_{i\in[K]}$ are revealed for every arm i. We assume $x_i(t) \in \mathbb{R}^d$ for all $i \in [K], t \in [T]$ and for every i, $\{x_i(t)\}_{t\in[T]}$ are i.i.d. from some distribution $\mathcal{P}_i$. At every time step t, an action $a_t \in [K]$ is chosen by the learner and a reward $r(t)$ is generated according to the following linear model:

$$r(t) = x_{a_t}(t)^\top \beta^* + \epsilon(t) \tag{1}$$

where $\beta^* \in \mathbb{R}^d$ is the unknown true signal and $\{\epsilon(t)\}_{t\in[T]}$ are independent sub-Gaussian random noise, also independent of all the other processes. We assume that the true parameter β^* is s^*–sparse, i.e., $\|\beta^*\|_0 = s^*$. We denote by S^* the true support of β^*, i.e., $S^* = \{j : \beta^*_j \neq 0\}$.

The goal is to design a sequential decision-making policy π that maximizes the expected cumulative reward over the time horizon. To formalize the notion, we define the history $\mathcal{H}_t$ up to time t as follows:

$$\mathcal{H}_t := \left\{(a_\tau, r(\tau), \{x_i(\tau)\}_{i\in[K]}) : \tau \in [t]\right\},$$

and an admissible policy π generates a sequence of random variables $a_1, a_2, \ldots$ taking values in $[K]$ such that a_t is measurable with respect to the σ-algebra generated by the previous feature vectors from each arm, observed rewards of the chosen arms till the previous round and the current feature vectors, i.e., measurable with respect to the filtration $\mathcal{F}_t := \sigma\left(x_{a_\tau}(\tau), r(\tau), x_i(t); \tau \in [t-1], i \in [K]\right)$.

Thus, an algorithm for contextual bandits is a policy π, which at every round t, chooses an action (arm) a_t based on history $\mathcal{H}_{t-1}$ and current contexts. We note that although contexts of the previous round corresponding to arms that were not chosen are in $\mathcal{F}_t$, however, they do not provide useful information on the parameter, since we do not observe rewards corresponding to them under the bandit feedback, and hence are not included in

the history $\mathcal{H}_t$. To measure the quality of performance, we compare it with the oracle policy π^* which uses the knowledge of the true β^* to choose the optimal action $a_t^* := \arg\max_{i\in[K]} x_i(t)^\top \beta^*$. Define $\Delta_i(t)$ to be the difference between the mean rewards of the optimal arm and ith arm at time t, i.e., $\Delta_i(t) = x_{a_t^*}(t)^\top \beta^* - x_i(t)^\top \beta^*$. Note that under the random-design assumption, a_t^* is also random. Then the regret at time t is defined as $\text{regret}(t) = \Delta_{a_t}(t)$ and the objective of the learner is to minimize the total regret till time T, defined as $R(T) = \sum_{t\in[T]} \text{regret}(t)$. We also define the matrix $X_t := (x_{a_1}(1), \ldots, x_{a_t}(t))^\top$. The time horizon T is finite but possibly unknown, but much smaller compared to the ambient dimension of the parameter, i.e. $d \gg T$. (Hao et al., 2021) refers to this regime as "data-poor" regime; such a regime adds an extra layer of hardness on top of the difficulty incurred by the sparse structure of β. We also assume that K is fixed and much smaller compared to both d and T.

2.1. Assumptions

In this section, we discuss the assumptions of our model.

Definition 2.1 (Sparse Riesz Condition (SRC))**.** Let M be a $d \times d$ positive semi-definite matrix. The maximum and minimum sparse eigenvalues of M with parameter $s \in [d]$ are defined as follows:

$$\phi_{\min}(s; M) := \inf_{\delta:\delta\neq 0, \|\delta\|_0 \leq s} \frac{\delta^\top M \delta}{\|\delta\|_2^2},$$
$$\phi_{\max}(s; M) := \sup_{\delta:\delta\neq 0, \|\delta\|_0 \leq s} \frac{\delta^\top M \delta}{\|\delta\|_2^2}.$$

We say M satisfies the SRC if $0 < \phi_{\min}(s, M) \leq \phi_{\max}(s; M) < \infty$.

Now we are ready to state the assumptions on the context distributions, which are as follows:

Assumption 2.2 (Assumptions on Context Distributions)**.** We assume that

(a) For some constant $x_{\max} \in \mathbb{R}^+$, we have that for all $i \in [K]$, $\mathcal{P}_i(\|x\|_\infty \leq x_{\max}) = 1$.

(b) For all arms $i \in [K]$, the distribution $\mathcal{P}_i$ is sub-Gaussian, i.e., there exists a constant $\vartheta > 0$ such that $\max_{i\in[K]} \|x_i(t)\|_{\psi_2} \leq \vartheta$ for all $t \in [T]$.

(c) There exists a constant $\xi \in \mathbb{R}^+$ such that for each $u \in \mathbb{S}^{d-1} \cap \{v \in \mathbb{R}^d : \|v\|_0 \leq Cs^*\}$ and $h \in \mathbb{R}^+$ $\mathcal{P}_i(\langle x, u\rangle^2 \leq h) \leq \xi h$, for all $i \in [K]$, ,where $C \in (2, \infty)$.

(d) The matrix $\Sigma_i := \mathbb{E}_{x\sim\mathcal{P}_i}[xx^\top]$ has bounded maximum sparse eigenvalue, i.e., $\phi_{\max}(Cs^*, \Sigma_i) \leq \phi_u < \infty$, for all $i \in [K]$, where C is the same constant as in part (c).

Assumption 2.2(a) basically tells that the contexts are bounded; such assumptions are standard in the bandit literature to obtain results on regret bound that are independent of the scaling of the contexts (or parameter). Assumption 2.2(b) says that all the arm-contexts are generated from sub-Gaussian distributions with a common bound of the order $O(\vartheta^2)$ on the proxy-variance, for all time point t. This is indeed a very mild assumption on the context distribution and a broad class of distributions enjoys such property. For example, truncated multivariate normal distribution with covariance matrix $\mathbb{I}_d$, where the truncation is over the set $\{u \in \mathbb{R}^d : \|u\|_\infty \leq 1\}$ is a valid distribution for the contexts. In comparison, most of the previous literature such as (Kim & Paik, 2019; Oh et al., 2021; Li et al., 2022) assume that $\|x_i(t)\|_2 \leq L$, for some constant $L > 0$. This condition automatically implies that Assumption 2.2(b) holds with $\vartheta = (L/\log 2)^{1/2}$. As a result, the theory in (Kim & Paik, 2019; Oh et al., 2021; Li et al., 2022) can not accommodate the aforementioned truncated multivariate normal distribution as in this case $\|x_i(t)\|_2 = \Theta(\sqrt{d})$ and their analysis yields $O(\sqrt{d})$ dependence in the regret bound. Assumption 2.2(c) talks about anti-concentration condition that plays a critical role in controlling the estimation accuracy of β^*. This condition is also assumed by (Li et al., 2021) and a variant (diverse covariates) of this condition is assumed in (Ren & Zhou, 2020). Intuitively, this condition prohibits the context features to fall along a singular direction. When u is not constrained to be sparse, this condition implies that the distribution of the contexts is not supported on a lower dimensional sub-space, allowing diversity and in the contexts and leading to inherent exploration. Assumption 2.2(c) captures this notion using the weaker condition where u is only sparse. Existing works like (Oh et al., 2021; Kim & Paik, 2019) try to capture this notion via compatibility/RE condition which is a somewhat stronger assumption and cannot be easily checked in practice. Assumption 2.2(c) is more interpretable - under the mere existence of bounded density for $u^\top x_i(t)$ (for all sparse u), the condition holds. Moreover, The entire Assumption 2.2 does not need any existence of pdf, whereas (Oh et al., 2021; Ariu et al., 2022) need the relaxed symmetry assumption which requires the existence of pdf. Lastly, from the discussion in Section 2.3 of (Ren & Zhou, 2020), it follows that minimum sparse eigenvalue assumption and relaxed symmetry assumption (both used in (Ariu et al., 2022)) implies diverse covariate property when $K = 2$. This suggests that Assumption 2.2(b) is very mild. Lastly, Assumption 2.2(d) imposes an upper bound on the maximum sparse eigenvalue of Σ_i which is a common assumption in high-dimensional literature (Zhang & Huang, 2008; Zhang, 2010).

Next, we come to the assumptions on the true parameter β^*

Assumption 2.3 (Assumptions on the true parameter)**.** We assume the followings:

(a) Sparsity and Soft-sparsity: There exist positive constants $s^* \in \mathbb{N}$ and $b_{\mathsf{max}} \in \mathbb{R}^+$ such that $\|\beta^*\|_0 = s^*$ and $\|\beta^*\|_1 \leq b_{\mathsf{max}}$.

(b) Margin condition: There exists positive constants Δ_*, A and $\omega \in [0, \infty]$, such that for $h \in \left[A\sqrt{\log(d)/T}, \Delta_*\right]$ and for all $t \in [T]$,

$$\mathbb{P}\left(x_{a_t^*}(t)^\top \beta^* \leq \max_{i \neq a_t^*} x_i(t)^\top \beta^* + h\right) \leq \left(\frac{h}{\Delta^*}\right)^\omega.$$

The first part of the assumption requires boundedness of the true parameter β^* to make the final regret bound scale free. Such an assumption is also standard in bandit literature (Bastani & Bayati, 2020; Abbasi-Yadkori et al., 2011).

The second part of the assumption imposes a margin condition on the arm distributions. Essentially, this assumption controls the probability of the optimal arm falling into h-neighborhood of the sub-optimal arms. As ω increases, the margin condition becomes stronger as the sub-optimal arms are less likely to fall close to the optimal arms. As a result, it becomes easier for any bandit policy to distinguish the optimal arm. As an illustration, consider the two extreme cases $\omega = \infty$ and $\omega = 0$. The $\omega = \infty$ case tells that there is a deterministic gap between rewards corresponding to the optimal arm and sub-optimal arms. This is the same as the "gap assumption" in (Abbasi-Yadkori et al., 2011). Thus, quite evidently it is easy for any bandit policy to recognize the optimal arm. This phenomenon is reflected in the regret bound of Theorem 5 in (Abbasi-Yadkori et al., 2011), where the regret depends on the time horizon T only though poly-logarithmic terms. In contrast, $\omega = 0$ corresponds to the case when there is no apriori information about the separation between the arms, and as a consequence, we pay the price in regret bound by a $\sqrt{T}$ term (Hao et al., 2021; Agrawal & Goyal, 2013; Chu et al., 2011).

The margin condition with $\omega = 1$ has been assumed in (Goldenshluger & Zeevi, 2013; Bastani & Bayati, 2020; Wang et al., 2018) and will be satisfied when the density of $x_i(t)^\top \beta^*$ is uniformly bounded for all $i \in [K]$. (Li et al., 2021) also discusses an example where the margin condition holds for different values of ω.

The final assumption is on the noise variables:

Assumption 2.4 (Assumption on Noise)**.** We assume that the random variables $\{\epsilon(t)\}_{t\in[T]}$ are independent and also independent of the other processes and each one is $\sigma-$Sub-Gaussian, i.e., $\mathbb{E}[e^{a\epsilon(t)}] \leq e^{\sigma^2 a^2/2}$ for all $t \in [T]$ and $a \in \mathbb{R}$.

Various families of distribution satisfy such a requirement, including normal distribution and bounded distributions, which are commonly chosen noise distributions. Note that such a requirement automatically implies that for every $t \in [T]$, $\mathbb{E}[\epsilon(t)] = 0$ and $\mathrm{Var}[\epsilon(t)] \leq \sigma^2$.

2.2. Thompson Sampling and Prior

We discuss the basics of Thompson sampling and introduce the specific structure of the prior that we use and analyze. Typically, we place a prior Π on the unknown parameter (β in our case) along with a *specified likelihood model* on the data, and do the following: while taking action, we draw a sample from the posterior distribution of the parameter given the data and use that as the proxy for the unknown parameter value, hence in our case at time t, we draw a sample $\hat{\beta}_t \sim \Pi(\beta|\mathcal{H}_{t-1})$ and choose $a_t = \arg\max_i x_i(t)^\top \hat{\beta}_t$ as the action. While simple enough to describe, Thompson sampling has been difficult to analyze theoretically, particularly because of the complex dependence between the observations due to the bandit structure. The choice of prior plays a crucial role, as we shall see, in providing the correct exploration-exploitation trade-off. In the high-dimensional sparse case that we are dealing with, this choice is specifically important since we do not wish to have a linear dependence on the dimension d in our regret bound - which would be incurred if we use the normal prior-likelihood setup of (Agrawal & Goyal, 2013), which analyzes Thompson sampling in contextual bandits.

While there is a rich literature on Bayesian priors for high dimensional regression, including horseshoe priors and slab-and-spike priors among others, we shall be using the complexity prior introduced in (Castillo et al., 2015). Specifically, we consider a prior Π on β that first selects a dimension s from a prior π_d on the set $[d]$, next a random subset $S \subset [d]$ of size $|S| = s$ and finally, given S, a set of nonzero values $\beta_S := \{\beta_i : i \in S\}$ from a prior density g_S on $\mathbb{R}^S$. Formally, the prior on (S, β) can be written as

$$(S, \beta) \mapsto \pi_d(|S|) \frac{1}{\binom{d}{|S|}} g_S(\beta_S)\delta_0(\beta_{S^c}), \tag{2}$$

where the term $\delta_0(\beta_{S^c})$ refers to coordinates β_{S^c} being set to 0. Moreover, we choose g_S as a product of Laplace densities on $\mathbb{R}$ with parameter λ/σ, i.e., $\beta_i \mapsto (2\sigma)^{-1}\lambda \exp(-\lambda|\beta_i|/\sigma)$ for all $i \in S$. Note that, here we assume that the noise level σ is known. In practice, one can add another level of hierarchy by setting a prior on σ but in this paper we do not pursue that direction.

The prior π_d plays the role of expressing the sparsity of the parameter. This is in contrast to other priors like product of independent Laplace densities over the coordinates (typically known as Bayesian LASSO), where the Laplace parameter plays the role of shrinking the coefficients towards 0. However, in our case, the scale parameter λ of the Laplace does not have this role and we assume that during the tth round we use $\lambda \in [(5/3)\overline{\lambda}_t, 2\overline{\lambda}_t]$, where $\overline{\lambda}_t \asymp \sqrt{t \log d}$, which is the usual order of the regularization parameter used in the LASSO.

The choice of the prior π_d is very critical; it should down

weight big models but at the same time give enough mass to the true model. Following (Castillo et al., 2015), we assume that there are constants $A_1, A_2, A_3, A_4 > 0$ such that $\forall\, s \in [d]$

$$A_1 d^{-A_3}\pi_d(s-1) \le \pi_d(s) \le A_2 d^{-A_4}\pi_d(s-1). \quad (3)$$

Complexity priors of the form $\pi_d(s) \propto c^{-s}d^{-as}$ for constants a, c satisfy the above requirement. Moreover, slab and spike priors of the form $(1-r)\delta_0 + r\mathsf{Lap}(\lambda/\sigma)$ independently over the coordinates satisfy the requirement with hyperprior on r being $\mathsf{Beta}(1, d^u)$.

Finally, we specify the data likelihood that is crucial for the TS algorithm. At each time point $t \in [T]$, given the observations coming from model (1), we model the $\{\epsilon(\tau)\}_{\tau \le t}$ as i.i.d. $\mathsf{N}(0, \sigma^2)$. We emphasize that this Gaussian assumption is *only required for likelihood modeling* and our main results hold under any true error distribution satisfying Assumption 2.4. The same strategy is also used in (Agrawal & Goyal, 2013) for the LinTS algorithm in low-dimensional setting.

3. Main Results

3.1. Posterior contraction

Now, we present an informal version of the main posterior contraction result for the estimation of β. A more detailed version of the result with exact rates, along with the measure theoretic details, is in Appendix C.

Theorem 3.1 (Informal)**.** *Write* $\mathbf{r}_t = (r(1), \ldots, r(t))^\top$, *and let the Assumption 2.2–2.4 hold with* $C = \Theta(\phi_u \vartheta^2 \xi K \log K)$, *and* $K \ge 2, d \ge T$. *With* $\lambda \asymp x_{\mathsf{max}}(t \log d)^{1/2}$ *and* $\varepsilon_{t,d,s} = s^*\{(\log d + \log t)/t\}\}^{1/2}$, *the following holds as* $t \to \infty$*:*

$$\mathbb{E}_{\mathbf{r}_t}\Pi\left(\|\beta - \beta^*\|_1 \gtrsim \sigma\varepsilon_{t,d,s} \,\middle|\, \mathbf{r}_t, X_t\right) \overset{a.s.}{\to} 0.$$

The above result is similar to Theorem 3 in (Castillo et al., 2015) under classical linear regression setup with i.i.d. observations and Gaussian noise. However, we generalize their result under bandit setup and sub-Gaussian noise by carefully controlling the correlation between noise and observed contexts, which is crucial for our regret analysis.

3.2. Algorithm and regret bound

In this section we introduce the Thomson sampling algorithm for high-dimensional contextual bandit, a pseudo-code for which is provided below in Algorithm 1. Similar to the Thompson sampling algorithm in (Agrawal & Goyal, 2013), in the tth round Algorithm 1 sets the a specific prior on β and updates it sequentially based on the observed rewards and contexts. In particular, it chooses the prior described in (2) with an appropriate choice of round-specific prior scaling λ_t and updates the posterior using the observed rewards and contexts until $(t-1)$th round. Then a sample is generated from the posterior and an arm a_t is chosen greedily based on the generated sample.

Now, we show that the Thompson sampling algorithm achieves desirable regret upper bound.

Theorem 3.2. *Let the Assumption 2.2–2.4 hold with* $C = \Theta(\phi_u \vartheta^2 \xi K \log K)$, *and* $K \ge 2, d \ge T$. *Define the quantity* $\kappa(\xi, \vartheta, K) := \min\{(4c_3 K \xi \vartheta^2)^{-1}, 1/2\}$ *where* c_3 *is a universal positive constant. Also, set the prior scaling* λ_t *as follows:*

$$(5/3)\overline{\lambda}_t \le \lambda_t \le 2\overline{\lambda}_t, \quad \overline{\lambda}_t = x_{\mathsf{max}}\sqrt{2t(\log d + \log t)}.$$

Then there exists a universal constant $C_0 > 0$ *such that we have the following regret bound for Algorithm 1:*

$$\mathbb{E}\{R(T)\} \lesssim I_b + I_\omega,$$

where,

$$I_b = \left\{\frac{b_{\mathsf{max}} x_{\mathsf{max}} \phi_u \vartheta^2 \xi (K \log K)}{\min\{\kappa^2(\xi, \vartheta, K), \log K\}}\right\} s^* \log(Kd),$$

$$I_\omega = \begin{cases} \Phi^{1+\omega}\left(\frac{s^{*1+\omega}(\log d)^{\frac{1+\omega}{2}} T^{\frac{1-\omega}{2}}}{\Delta_*^\omega}\right), & \text{for } \omega \in [0,1), \\ \Phi^2\left(\frac{s^{*2}[\log d + \log T]\log T}{\Delta_*}\right), & \text{for } \omega = 1, \\ \frac{\Phi^2}{(\omega-1)}\left(\frac{s^{*2}[\log d + \log T]}{\Delta_*}\right), & \text{for } \omega \in (1, \infty) \\ \Phi^2\left(\frac{s^{*2}[\log d + \log T]}{\Delta_*}\right), & \text{for } \omega = \infty, \end{cases}$$

and $\Phi = \sigma x_{\mathsf{max}}^2 \xi K\left(2 + 40A_4^{-1} + C_0 K \xi x_{\mathsf{max}}^2 A_4^{-1}\right)$.

Discussion on the above result: The regret bound provided by Theorem 3.2 shows that the regret of the algorithm grows poly-logarithmically in d, i.e., $\mathbb{E}\{R(T)\} = O((\log d)^{\frac{1+\omega}{2}})$, when $\omega \in [0, 1)$; logarithmically in d, i.e., $O(\log d)$ when $\omega \in [1, \infty]$. Meanwhile, the expected cumulative regret depends polynomially in T, i.e., $\mathbb{E}\{R(T)\} = O(T^{\frac{1-\omega}{2}})$ when $\omega \in [0, 1)$; ploy-logarithmically in T; i.e., $\mathbb{E}\{R(T)\} = O((\log T)^2)$, when $\omega = 1$. In $\omega \in (1, \infty]$ regime, the expected cumulative regret depends poly-logarithmically in both the time horizon T and ambient dimension d. As $T \ll d$, the expected regret ultimately scales as $O(\log d)$. Comparing our upper bound result with minimax regret lower bound established in Theorem 1 of (Li et al., 2021), it follows that our algorithm enjoys optimal dependence on both ambient dimension d and time-horizon T when $\omega \in [0, 1)$. In $\omega = 1$ region, the regret upper bound in the above theorem is optimal up to a $O(\log T)$ term. To the best of our knowledge, there does not exist any result on minimax lower bound in the regime $\omega > 1$ in the high-dimensional linear contextual bandit literature. It is worth

mentioning that this is an upper bound on the expected (frequentist) regret, as compared to Bayesian regret which is often considered for Thompson sampling based algorithms.

Algorithm 1 Thompson Sampling Algorithm

$\mathcal{H}_0 = \emptyset$.
for $t = 1, \cdots, T$ **do**
  if $t \leq 1$ **then**
    Choose action a_t uniformly over $[K]$.
  end if
  if $t > 1$ **then**
    Set $\overline{\lambda}_t = x_{\max}\sqrt{2t(\log d + \log t)}$ and choose $\lambda_t \in (5\overline{\lambda}_t/3, 2\overline{\lambda}_t)$.
    Generate sample $\tilde{\beta}_t \sim \Pi(\cdot \mid \mathcal{H}_{t-1})$ with prior Π in (2)-(3), $\lambda = \lambda_t$ and Gaussian likelihood.
    Play arm: $a_t = \arg\max_{i\in[K]} x_i(t)^\top \tilde{\beta}_t$.
  end if
  Observe reward $r(t)$.
  Update $\mathcal{H}_t \leftarrow \mathcal{H}_{t-1} \cup \{(a_t, r_{a_t}(t), x_{a_t}(t))\}$.
end for

Intuitively, the initial term I_b in regret upper bound in Theorem 3.2 describes the regret caused by the "burn-in" period of exploring the space of contexts and it does not contribute to the asymptotic regret growth. Note that we consider running Thompson sampling from the very beginning, without an explicit random exploration phase, in contrast to most of the existing algorithms; the distinction between the burn-in phase and the subsequent phase is only a construct of our theoretical analysis. Furthermore, the constant Δ_* plays the role of gap parameter which commonly appears in a problem-dependent regret bound (Abbasi-Yadkori et al., 2011). Note that, for $\omega = 0$, we get a problem-independent regret bound of the order $O(s^*\sqrt{T}\log d)$. The appearance of the $\sqrt{T}$, term is not surprising, as the condition $\omega = 0$ poses no prior knowledge on the arm-separability, Thus, in the worst case, the context vectors may fall into each other, making the bandit environment harder to learn. In contrast, as ω increases the optimal arm becomes more distinguishable than the sub-optimal arms and the bandit environment becomes easier to learn. As a result, the effect of the time horizon becomes less and less severe as ω increases. In particular, when $\omega \in [1, \infty]$, the time horizon does not affect the asymptotic growth of the regret bound. Finally, as we mainly focus on the case when the number of arms is very small, the quantity Φ roughly has an inflating effect of $O(1)$ on the regret bound.

Sketch of the proof of Theorem 3.2: While a self-contained and detailed proof of the above result is given in the Appendix, here we go through the main steps and ideas of the proof. The proof is broadly divided into 3 parts for clarity:

(i) In Section B.1 we will first show that the estimated covariance matrix $\widehat{\Sigma}_t := X_t^\top X_t/t$ enjoys SRC condition with high probability for sufficiently large t. In our analysis, we carefully decouple this complex dependent structure and exploit the special temporal dependence structure of the bandit environment to establish SRC property of $\widehat{\Sigma}_t$.

(ii) Next, in Section B.2 we will establish a compatibility condition for the matrix $\widehat{\Sigma}_t$. We use a Transfer Lemma (Lemma B.9) which essentially translates the uniform lower bound on $\phi_{\min}(Cs^*, \widehat{\Sigma}_t)$ to a certain compatibility number.

(iii) Finally, in Section B.3, under the compatibility condition we use the posterior contraction result in Theorem 3.1 to give bound on the per round regret $\Delta_{a_t}(t)$.

3.3. Comparison with existing literature

Over the past few years, the problem of high dimensional stochastic linear contextual bandit has attracted significant attention and has quite evidently generated a large body of work in this field under different problem settings. However, there are mainly two types of settings that have been considered in high-dimensional linear bandit literature: **(S1)** Each arm has different parameters β_i^* for $i \in [K]$ and only one context vector $x(t)$ is generated at every time point t, **(S2)** K different contexts $x_i(t)$ are generated for each arm $i \in [K]$ at every time point t and all of the arms have one common parameter β^*, which is also the setting of this paper.

There has been an ample amount of work in both of these settings. To mention a few, (Wang et al., 2018; Bastani & Bayati, 2015; Wang & Cheng, 2020) consider the setting in **(S1)**, whereas (Kim & Paik, 2019; Li et al., 2021; Oh et al., 2021) consider the setting in **(S2)**. It is worth mentioning that most of these works assume very strong compatibility or restricted eigenvalue (RE) conditions on the feature distribution, which is in general hard to check in real-world applications. Instead, in this paper we show that TS algorithm enjoys desirable regret bound under much weaker and easily interpretable assumptions on the feature distribution. There is also a parallel line of work that considers the set of features or contexts to be infinite but fixed (Hao et al., 2020; 2021; Jang et al., 2022), which is in sharp contrast to the setting considered in this paper. Moreover, in the setup of these works, the optimal arm remains the same for every round. On the other hand, in our setting, due to the randomness of the observed contexts, the optimal arm does not necessarily remain the same in every round. Lastly, (Hao et al., 2021) and (Hao et al., 2022) study the properties of information-directed sampling and provide a guarantee for Bayesian regret, which is much weaker than the result in Theorem 3.2.

Now we compare the results and assumptions of this paper with existing literature in SLCB setting. Table 1 shows the

Table 1. This table compares the regret bounds and working assumptions of this paper with existing works under different SLCB settings. We focus four most important assumptions: (1) 'Margin' - similar to Assumption 2.2(b) with $\omega \in \{0, 1\}$, (2)'Comp/RE'- Compatibility or RE condition, (3) 'ℓ_2-bound'- boundedness of contexts in ℓ_2-norm, (4) 'Pdf exst'- existence of pdf. ✓ symbol indicates that the corresponding condition is assumed in the paper. ✓$(\star)$ symbol indicates that (Chen et al., 2022) assumes that the coordinates of the contexts are i.i.d and the second moments are lower bounded, which is typically much stronger than compatibility or RE condition.

Setting	Paper	Regret Bound	Margin	Comp/RE	ℓ_2-bound	Pdf exst
(S1)	Bastani & Bayati (2015)	$O(s^{*2}[\log d + \log T]^2)$	✓	✓		
	Wang et al. (2018)	$O(s^{*2}[\log d + \log T]\log T)$	✓	✓		
	Wang & Cheng (2020)	$O(s^{*2}[\log d + \log T]^2)$	✓	✓		
(S2)	Kim & Paik (2019)	$O(s^*\sqrt{T}\log(dT))$		✓	✓	
	Oh et al. (2021)	$O(\sqrt{s^*T\log(dT)})$		✓	✓	✓
	Li et al. (2021)	$O(s^*\sqrt{T\log d})$				
		$O(s^{*2}[\log d + \log T]\log T)$	✓			
	Li et al. (2022)	$O(s^{*1/3}T^{2/3}\sqrt{\log(dT)})$		✓	✓	
	Ariu et al. (2022)	$O(s^{*2}\log d + \sqrt{s^*T})$		✓		✓
		$O(s^{*2}\log d + s^*\log T)$	✓	✓		✓
	Chen et al. (2022)	$O(s^*\sqrt{T}\log^2(Td))$		✓$(\star)$		
	This paper	$O(s^*\sqrt{T\log d})$				
		$O(s^{*2}[\log d + \log T]\log T)$	✓			

comparison of regret bound and working assumptions of different papers under both **(S1)** and **(S2)** settings. Under the setup of **(S1)**, (Bastani & Bayati, 2015) and (Wang et al., 2018) proposed the LASSO-bandit algorithm and MCP bandit algorithm respectively, and under the margin condition $\omega = 1$, they established a regret bound of the order $\widetilde{O}((s^*\log T)^2)$[1]. However, Theorem 3.2 accommodates a broader range of ω and our method does not need the knowledge of ω, but enjoys the same regret upper bound for $\omega = 1$. Moreover, unlike LASSO-bandit or MCP-bandit, our method does not require forced sampling which could be expensive in certain marketing applications.

Under the setting in **(S2)**, (Kim & Paik, 2019) and (Ariu et al., 2022) proposed Doubly-robust LASSO and Threshold LASSO bandit algorithms respectively. Under strong compatibility or RE condition they established $\widetilde{O}(\sqrt{T})$ regret bound. With margin condition $\omega = 1$, (Ariu et al., 2022) improved their bound to $\widetilde{O}(\log T)$. (Li et al., 2022) proposed "Explore Structure then Commit" framework and established regret bound of the order $\widetilde{O}(T^{2/3})$. However, all of these algorithms require the knowledge of true sparsity s^*, and as mentioned before, also need some type of compatibility/RE conditions or some other strong conditions on the covariance structure and density functions of the contexts. In comparison, our theory does not assume any strong compatibility/RE condition on the context distribution and the TS algorithm also does not require the knowledge of true sparsity but still enjoys better or comparable regret bound. In some recent works, (Oh et al., 2021; Li et al., 2021) also proposed LASSO-based algorithms which do not require the knowledge of true sparsity s^*. It is worth mentioning that under a similar set of assumptions as in this paper, (Li et al., 2021) showed that the LASSO-L1 confidence ball algorithm enjoys similar regret bounds as in Theorem 3.2 for different values of ω. However, the Sparsity Agnostic LASSO algorithm proposed in (Oh et al., 2021) needs a strong RE and balanced covariance assumptions. Lastly, (Chen et al., 2022) recently proposed Sparse LinUCB and SupLinUCB algorithm which relies on best subset selection and showed that it enjoys $\widetilde{O}(\sqrt{T})$ regret bound. However, they assume that the contexts are sub-Gaussian with independent coordinates, which is far stronger than compatibility condition and even unrealistic in most real-world applications.

[1] $\widetilde{O}(\cdot)$ hides the logarithmic dependence on d or T.

4. Computation

In this section, we discuss the computational challenges and how these are overcome by using Variational Bayes (VB). While priors as (2) have been shown to perform well, both empirically and in theory, the discrete model selection component of the prior makes it challenging to allow computation and inference on the posterior. For $\beta \in \mathbb{R}^d$, inference using the slab and spike prior requires a combinatorial search over 2^d possible models, which in the case of high dimension is computationally infeasible. Fast algorithms are known only in the special diagonal design case and traditional Markov Chain Monte Carlo methods have very slow mixing in such high dimensional cases. Thus, following (Ray & Szabó, 2021) we use Variational Bayes to make computations faster. Specifically, in the sampling

step of Algorithm 1, we consider the VB approximation of the posterior $\Pi(\cdot \mid \mathcal{H}_{t-1})$ arising from slab and spike prior with $\mathsf{Lap}(\lambda/\sigma)$ slab in the mean-field family

$$\left\{ \bigotimes_{j=1}^{d} [\gamma_j \mathsf{N}(\mu_j, \sigma_j^2) + (1-\gamma_j)\delta_0] : (\mu_j, \sigma_j, \gamma_j) \in \mathcal{R} \right\},$$

where $\mathcal{R} = \mathbb{R} \times \mathbb{R}^+ \times [0,1]$. We use the `sparsevb` package (Clara et al., 2021) in R to use the Coordinate Ascent Variational Inference (CAVI) algorithm proposed in (Ray & Szabó, 2021) to obtain the VB posterior. This makes the Thompson sampling algorithm much faster as one can efficiently obtain samples from the VB posterior due to its structure. The details of the algorithm for the Variational Bayes Thompson Sampling (VBTS)[2] are in the appendix (see Section E).

5. Numerical Experiments

In both simulations and real data experiments, we present results corresponding to $\lambda_t = 1$ for all $t \in [T]$. Recall that Theorem 3.2 suggests that in tth round $\lambda_t \asymp \sqrt{t \log d}$ is a reasonable choice for the exact Thompson sampling algorithm. However, in practice, we noticed that such choices of λ_t lead to numerical instability. Some recent findings in (Ray & Szabó, 2021) suggest that λ_t in the order of $O(\sqrt{t \log d}/s^*)$ should be an appropriate choice, which is smaller than the predicted order of λ_t in our main theorem. Motivated by this, we also present the simulation results for synthetic data experiments with $\lambda_t = \lambda_* \sqrt{t}$ for $\lambda_* \in \{0.2, 0.3, 0.4, 0.5\}$ in the Appendix A.2. We found the performance of VBTS to be robust with respect to the choice of the tuning parameter λ_t.

5.1. Synthetic data

In this section, we illustrate the performance of the VBTS algorithm on a simulated data set. As a benchmark, we consider, DR-LASSO (Kim & Paik, 2019), LASSO-L1 confidence ball algorithm (Li et al., 2021), ESTC (Li et al., 2022), sparsity agnostic (SA) LASSO (Oh et al., 2021), thresholded (TH) LASSO (Ariu et al., 2022), and TS algorithm based on Bayesian LASSO (Park & Casella, 2008) (BLASSO TS) to compare the performance of VBTS (Algorithm 2). In this section, we only include the methods that are designed for the high-dimensional linear contextual bandit. Simulation results for LinUCB (Abbasi-Yadkori et al., 2011) and LinTS (Agrawal & Goyal, 2013) can be found in Appendix A.4.

[2]Codes are available online: `Github link`.

EQUICORRELATED (EC) STRUCTURE

We set the number of arms $K = 10$ and we generate the context vectors $\{x_i(t)\}_{i=1}^K$ from multivariate d-dimensional Gaussian distribution $\mathsf{N}_d(\mathbf{0}, \Sigma)$, where $\Sigma_{ij} = \rho^{|i-j| \wedge 1}$ and $\rho = 0.3$. We consider $d = 1000$ and the sparsity $s^* = 5$. We choose the set of active indices S^* uniformly over all the subsets of $[d]$ of size s^*. Next, for each choice of d, we consider two types generating scheme for β:

- **Setup 1:** $\{U_i\}_{i \in S^*} \overset{i.i.d.}{\sim} \text{Uniform}(0.3, 1)$ and set $\beta_j = U_j (\sum_{\ell \in S^*} U_\ell^2)^{-1/2} \mathbb{1}(j \in S^*)$.
- **Setup 2:** $\{Z_i\}_{i \in S^*} \overset{i.i.d.}{\sim} \mathsf{N}(0,1)$ and set $\beta_j = Z_j (\sum_{\ell \in S^*} Z_\ell^2)^{-1/2} \mathbb{1}(j \in S^*)$.

We run 40 independent simulations and plot the mean cumulative regret with 95% confidence band in Figure 1a-1b. In all the setups, we see that VBTS outperforms its competitors by a wide margin.

AUTOREGRESSIVE (AR) STRUCTURE

We consider the same setups as in the EC structure above, with the exception of the context distribution. Here we generate the context vectors $\{x_i(t)\}_{i=1}^K$ from $\mathsf{N}_d(\mathbf{0}, \Sigma)$ where $\Sigma_{ij} = \phi^{|i-j|}$ and $\phi = 0.3$. VBTS also enjoys superior empirical performance under this setup (see Figure 1c-1d). Table 2 shows the mean execution time (across Setup 1 and 2) of all TS algorithms for both EC and AR structure simulations. Among the class of TS algorithms, VBTS outperforms its other competing algorithms.

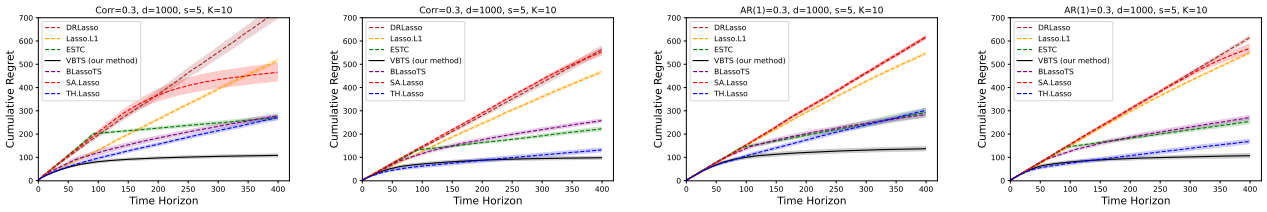
Table 2. Time comparison among the competing algorithms.

Type	Algorithm	Mean time of execution (seconds)	
		Equicorrelated	Auto-regressive
TS	LinTS	1344.39	1346.46
	BLASSO TS	1511.68	1455.53
	VBTS	**29.33**	**27.65**

5.2. Real data - `gravier` Breast Carcinoma Data

We consider breast cancer data `gravier` (`microarray` package in `R`) for 168 patients to predict metastasis of breast carcinoma based on 2905 gene expressions (bacterial artificial chromosome or BAC array). The goal of the learner is to identify the positive cases.

Similar to (Kuzborskij et al., 2019; Chen et al., 2021), in our experimental setup, we convert the breast cancer classification problem into 2-armed contextual bandit problem. More details about the data and reward generation process are provided in Appendix A.3. We perform 10 independent Monte Carlo simulations and plot the expected regret of



(a) EC (Setup 1) (b) EC (Setup 2) (c) AR (Setup 1) (d) AR (Setup 2)

Figure 1. Cumulative regret of competing algorithms.

VBTS in Figure 2 along with its competitors. In this experiment, we omit LinUCB and LinTS algorithms as they were performing far worse compared to the existing ones in Figure 2. The figure shows that VBTS and LASSO-L1 confidence ball algorithms are by far the clear winners in terms of cumulative regret. However, upon closer look, we see that VBTS is slightly better than LASSO-L1 confidence in terms of cumulative regret. In terms of accuracy, LASSO-L1 and VBTS are in the same ball park as seen in Table 3.

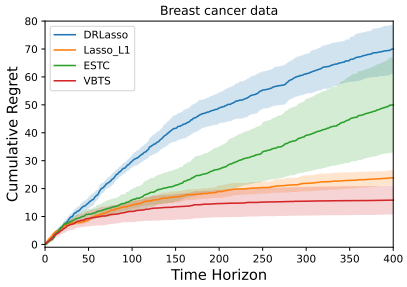


Figure 2. Cumulative regret plot for breast cancer data set.

Table 3. Classification accuracy of competing algorithms.

Algorithm	DR-LASSO	LASSO-L1	ESTC	**VBTS**
Accuracy(%)	65.63	81.20	73.32	**81.88**

6. Conclusion

In this paper, we consider the stochastic linear contextual bandit problem with high-dimensional sparse features and a fixed number of arms. We propose a Thompson sampling algorithm for this problem by placing a suitable *sparsity-inducing* prior on the unknown parameter to induce sparsity. We also develop a crucial posterior contraction result for *non-i.i.d.* data that allows us to obtain an almost *dimension independent* regret bound for our proposed algorithm. We explicitly point out the dependences on d and T for different arm-separation regimes parameterized by ω, which is also minimax optimal for $\omega \in [0, 1)$. Moreover, the choice of prior allows us to devise a Variational Bayes algorithm that enjoys computational expediency over traditional MCMC. We demonstrate the superior performance of our algorithm through extensive simulation studies. We finally perform an experiment on the `gravier` dataset, for which our method performs better compared to other existing algorithms.

Now we point the readers toward some of the natural research directions that we plan to cover in our future works. The regret analysis, similar to most of the recent works in high dimensional contextual bandits, relies on upper bounding the regret through estimation of the parameter, i.e., we rely on the estimation of β^* to be able to provide meaningful regret bound. However, this should not be required - as an example, consider the case where the first coordinate of $x_i(t)$ is 0 for all $i \in [K], t \in [T]$. Then the first coordinate of β^* is not estimable, however, this does not pose any problem to designing a sensible policy since this coordinate does not appear in the regret. Unfortunately, Assumption 2.2(c) is not satisfied for such degeneracy in the contexts and as a result, it would require a modified analysis of the regret bound. Secondly, we underscore the fact that in our setup we adopt the Variational Bayes framework only to sidestep the computational hurdles of MCMC arising from a myriad of challenges such as slow mixing times of the chains, lack of easy implementation, etc. However, in high-dimensional regression setup (Yang et al., 2016) has proposed Metropolis-Hastings algorithms based on truncated sparsity priors that do not meet the above roadblocks. It could be very well possible that some other prior structure will allow us to design more efficient MCMC algorithms with faster mixing times in the high-dimensional SLCB setup along with theoretical guarantees. Finally, we also plan to analyze Thompson sampling for high-dimensional generalized contextual bandit problems.

Acknowledgements

A.T. acknowledges the support of NSF via grant IIS-2007055. We also thank the four anonymous reviewers whose comments helped to significantly improve the paper.

Author Contribution: All authors conceived and carried out the research project jointly. S.C. and S.R. jointly wrote the paper and code for numerical experiments. A.T. helped edit the paper.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 24, 2011.

Abe, N., Biermann, A. W., and Long, P. M. Reinforcement learning with immediate rewards and linear hypotheses. *Algorithmica*, 37(4):263–293, 2003.

Agrawal, S. and Goyal, N. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pp. 127–135. PMLR, 2013.

Ariu, K., Abe, K., and Proutière, A. Thresholded lasso bandit. In *International Conference on Machine Learning*, pp. 878–928. PMLR, 2022.

Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Bastani, H. and Bayati, M. Online decision-making with high-dimensional covariates, 2015. *Available at SSRN 2661896*, 2015.

Bastani, H. and Bayati, M. Online decision making with high-dimensional covariates. *Operations Research*, 68 (1):276–294, 2020.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. Simultaneous analysis of lasso and dantzig selector. *The Annals of statistics*, 37(4):1705–1732, 2009.

Castillo, I., Schmidt-Hieber, J., and Van der Vaart, A. Bayesian linear regression with sparse priors. *The Annals of Statistics*, 43(5):1986–2018, 2015.

Chapelle, O. and Li, L. An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, 24, 2011.

Chen, C., Luo, L., Zhang, W., Yu, Y., and Lian, Y. Efficient and robust high-dimensional linear contextual bandits. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 4259–4265, 2021.

Chen, Y., Wang, Y., Fang, E. X., Wang, Z., and Li, R. Nearly dimension-independent sparse linear bandit over small action spaces via best subset selection. *Journal of the American Statistical Association*, (just-accepted):1–31, 2022.

Chu, W., Li, L., Reyzin, L., and Schapire, R. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 208–214. JMLR Workshop and Conference Proceedings, 2011.

Clara, G., Szabo, B., and Ray, K. sparsevb: Spike-and-slab variational bayes for linear and logistic regression: R package version 0.1. 0. 2021.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. 2008.

Gilton, D. and Willett, R. Sparse linear contextual bandits via relevance vector machines. In *2017 International Conference on Sampling Theory and Applications (SampTA)*, pp. 518–522. IEEE, 2017.

Goldenshluger, A. and Zeevi, A. A linear response bandit problem. *Stochastic Systems*, 3(1):230–261, 2013.

Gravier, E., Pierron, G., Vincent-Salomon, A., Gruel, N., Raynal, V., Savignoni, A., De Rycke, Y., Pierga, J.-Y., Lucchesi, C., Reyal, F., et al. A prognostic dna signature for t1t2 node-negative breast cancer patients. *Genes, chromosomes and cancer*, 49(12):1125–1134, 2010.

Hao, B., Lattimore, T., and Wang, M. High-dimensional sparse linear bandits. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 10753–10763. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/7a006957be65e608e863301eb98e1808-Paper.pdf.

Hao, B., Lattimore, T., and Deng, W. Information directed sampling for sparse linear bandits. *Advances in Neural Information Processing Systems*, 34:16738–16750, 2021.

Hao, B., Lattimore, T., and Qin, C. Contextual information-directed sampling. In *International Conference on Machine Learning*, pp. 8446–8464. PMLR, 2022.

Jang, K., Zhang, C., and Jun, K.-S. Popart: Efficient sparse regression and experimental design for optimal sparse linear bandits. In *Advances in Neural Information Processing Systems*, 2022.

Kaufmann, E., Korda, N., and Munos, R. Thompson sampling: An asymptotically optimal finite-time analysis. In *International conference on algorithmic learning theory*, pp. 199–213. Springer, 2012.

Kim, G.-S. and Paik, M. C. Doubly-robust lasso bandit. *Advances in Neural Information Processing Systems*, 32: 5877–5887, 2019.

Kuzborskij, I., Cella, L., and Cesa-Bianchi, N. Efficient linear bandits through matrix sketching. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 177–185. PMLR, 2019.

Li, K., Yang, Y., and Narisetty, N. N. Regret lower bound and optimal algorithm for high-dimensional contextual linear bandit. *Electronic Journal of Statistics*, 15(2):5652–5695, 2021.

Li, W., Barik, A., and Honorio, J. A simple unified framework for high dimensional bandit problems. In *International Conference on Machine Learning*, pp. 12619–12655. PMLR, 2022.

Oh, M.-h., Iyengar, G., and Zeevi, A. Sparsity-agnostic lasso bandit. In *International Conference on Machine Learning*, pp. 8271–8280. PMLR, 2021.

Oliveira, R. I. The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties. *arXiv preprint arXiv:1312.2903*, 2013.

Park, T. and Casella, G. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

Raskutti, G., Wainwright, M. J., and Yu, B. Restricted eigenvalue properties for correlated gaussian designs. *The Journal of Machine Learning Research*, 11:2241–2259, 2010.

Ray, K. and Szabó, B. Variational bayes for high-dimensional linear regression with sparse priors. *Journal of the American Statistical Association*, pp. 1–12, 2021.

Ren, Z. and Zhou, Z. Dynamic batch learning in high-dimensional sparse linear contextual bandits. *arXiv preprint arXiv:2008.11918*, 2020.

Vershynin, R. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Wang, C.-H. and Cheng, G. Online batch decision-making with high-dimensional covariates. In *International Conference on Artificial Intelligence and Statistics*, pp. 3848–3857. PMLR, 2020.

Wang, X., Wei, M., and Yao, T. Minimax concave penalized multi-armed bandit model with high-dimensional covariates. In *International Conference on Machine Learning*, pp. 5200–5208. PMLR, 2018.

Yang, Y., Wainwright, M. J., and Jordan, M. I. On the computational complexity of high-dimensional bayesian variable selection. *The Annals of Statistics*, 44(6):2497–2532, 2016.

Zhang, C.-H. Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics*, 38(2): 894–942, 2010.

Zhang, C.-H. and Huang, J. The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics*, 36(4):1567–1594, 2008.

A. Details of Simulations

A.1. Simulation details for AR(1) structure

We set the number of arms $K = 10$ and we generate the context vectors $\{x_i(t)\}_{i=1}^K$ from multivariate d-dimensional Gaussian distribution $\mathsf{N}_d(\mathbf{0}, \Sigma)$, where $\Sigma_{ij} = \phi^{|i-j|}$ and $\phi = 0.3$. We consider $d = 1000$ and the sparsity $s^* = 5$. We choose the set of active indices $\mathcal{S}^*$ uniformly over all the subsets of $[d]$ of size s^*. Next, for each choice of d, we consider two types generating scheme for β:

- Setup 1: $\{U_i\}_{i\in\mathcal{S}^*} \overset{i.i.d.}{\sim} \text{Uniform}(0.3, 1)$ and set β as the following:

$$\beta_j = \begin{cases} \frac{U_j}{\sqrt{\sum_{\ell\in\mathcal{S}^*} U_\ell^2}}, & \text{if } j \in \mathcal{S}^*, \\ 0, & \text{otherwise.} \end{cases}.$$

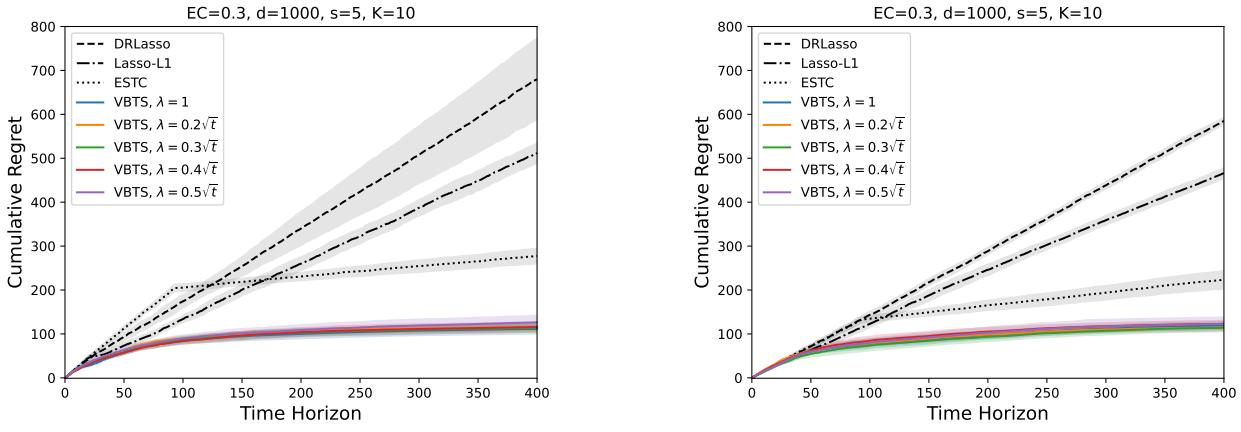
- Setup 2: $\{Z_i\}_{i\in\mathcal{S}^*} \overset{i.i.d.}{\sim} \text{Normal}(0, 1)$ and set β as the following:

$$\beta_j = \begin{cases} \frac{Z_j}{\sqrt{\sum_{\ell\in\mathcal{S}^*} Z_\ell^2}}, & \text{if } j \in \mathcal{S}^*, \\ 0, & \text{otherwise.} \end{cases}.$$

We run 40 independent simulations and plot the mean cumulative regret with 95% confidence band in Figure 1. In all the setups, we see that VBTS outperforms its competitors by a wide margin. Similar to the previous simulation example, in this case also Table 2 shows that VBTS is far better in terms of mean execution time than its competitors in the class of TS algorithms.

A.2. Siumlation for different choices of λ

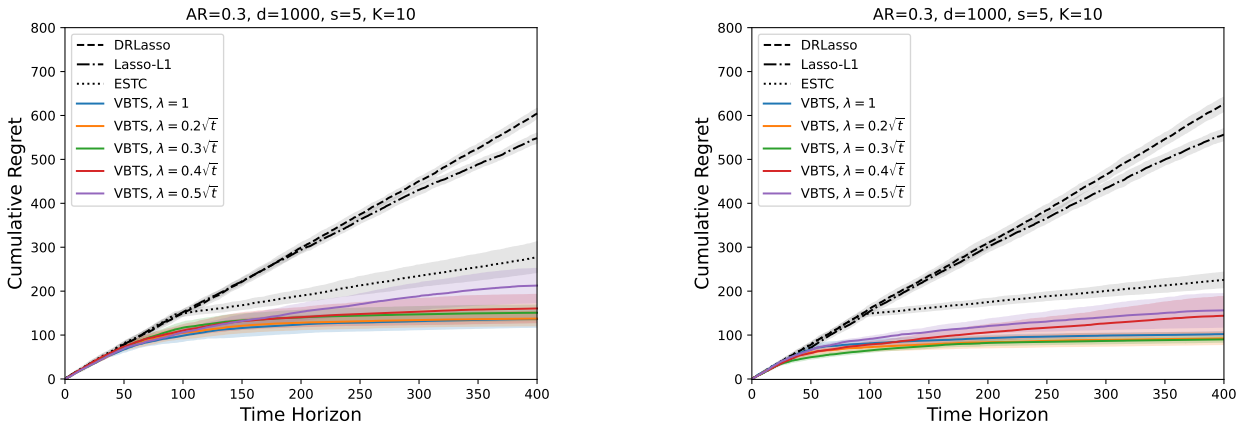
As discussed in the first paragraph of Section 5, for each of these simulation settings, we tried a few choices for the tuning parameter λ_t. In addition to the default choice of $\lambda_t = 1$ (for all time points t), we also explored the performance of the algorithm under growing λ, as required by our theoretical results. In particular, we tried $\lambda_t = \lambda_*\sqrt{t}$ for $\lambda_* \in \{0.2, 0.3, 0.4, 0.5\}$. For comparison, we only kept the faster optimism based methods DRLasso, Lasso-L1 and ESTC. We found the results to be roughly robust to the choice of this tuning parameter. The results are summarized in Figure 3 and Figure 4 below. However, we found that larger values of λ_* lead to numerical issues, we conjecture that this is an artifact of the variational Bayes approximation, rather than the prior itself. For our simulation settings, the choice $\sqrt{t \log d}/s^* \approx 0.5\sqrt{t}$ and hence by the findings in (Ray & Szabó, 2021), values of λ_* higher than this may yield inaccurate Variational Bayes estimation.



(a) Setup 1

(b) Setup 2

Figure 3. Regret bound for equi-correlated design for different tuning parameter choices



(a) Setup 1

(b) Setup 2

Figure 4. Regret bound for auto-regressive design for different tuning parameter choices

A.3. Details of real data experiment

We consider breast cancer data `gravier` (`microarray` package in `R`) for 168 patients to predict metastasis of breast carcinoma based on 2905 gene expressions (bacterial artificial chromosome or BAC array). (Gravier et al., 2010) considered small, invasive ductal carcinomas without axillary lymph node involvement (T1T2N0) to predict metastasis of small node-negative breast carcinoma. Using comparative genomic hybridization arrays, they examined 168 patients over a five-year period. The 111 patients with no event after diagnosis were labeled good (class 0), and the 57 patients with early metastasis were labeled poor (class 1). The 2905 gene expression levels were normalized with a $\log_2$ transformation.

Similar to (Kuzborskij et al., 2019; Chen et al., 2021), in our experimental setup we convert the breast cancer classification problem into 2-armed contextual bandit problem as follows: Given the `gravier` data set with 2 classes, we first set Class 1 as the target class. In each round, the environment randomly draws one sample from each class and composes a set of contexts of 2 samples. The learner chooses one sample and observes the reward following a logit model. In particular, we model the reward as

$$r(t) := \log\left\{\frac{\mathbb{P}(\text{Selected class} = 1)}{\mathbb{P}(\text{Selected class} = 0)}\right\} = x_{a_t}^\top \beta^* + \epsilon(t),$$

where $a_t \in \{1, 2\}$ is the selected arm at round t. Thus, small cumulative regret insinuates that the learner is able to differentiate the positive patients eventually. Such concepts can be used for constructing online classifiers to differentiate carcinoma metastasis from healthy patients based on gene expression data. However, in practice, we can not measure the regret defined in (2), unless we have the knowledge of β^*. To resolve this issue, we first fit a logit model on the whole `gravier` data set and consider the estimated $\hat{\beta}$ as the ground truth and report the expected regret with respect to the estimated β. As reported in (Gravier et al., 2010), 24 (out of 2905) BACs showed statistically significant difference (comparing Cy3/Cy5 values) between the two groups, motivating the use of a sparse logit model in our case. The estimated β^* in our sparse logistic model on the dataset had a sparsity of 18 using the dataset. In addition to this, we also treat the estimated noise variance from the fitted logit model as the true noise variance of the error induced by the environment in each round.

A.4. More simulations

We set the number of arms $K = 10$ and we generate the context vectors $\{x_i(t)\}_{i=1}^K$ from multivariate d-dimensional Gaussian distribution $\mathsf{N}_d(\mathbf{0}, \Sigma)$, where $\Sigma_{ij} = \rho^{|i-j|\wedge 1}$ and $\rho = 0.3$. We consider $d = 1000$ and the sparsity $s^* = 5$. We choose the set of active indices S^* uniformly over all the subsets of $[d]$ of size s^*. Next, for each choice of d, we consider two types generating scheme for β:

- **Setup 1:** $\{U_i\}_{i\in S^*} \overset{i.i.d.}{\sim} \text{Uniform}(0.3, 1)$ and set $\beta_j = U_j(\sum_{\ell\in S^*} U_\ell^2)^{-1/2}\mathbb{1}(j \in S^*)$.
- **Setup 2:** $\{Z_i\}_{i\in S^*} \overset{i.i.d.}{\sim} \mathsf{N}(0, 1)$ and set $\beta_j = Z_j(\sum_{\ell\in S^*} Z_\ell^2)^{-1/2}\mathbb{1}(j \in S^*)$.

We run 40 independent simulations and plot the mean cumulative regret with 95% confidence band in Figure 5. In all the setups, we see that VBTS outperforms its competitors by a wide margin. VBTS also enjoys superior empirical performance under the autoregressive (AR) model (see Figure 6) with auto-correlation coefficient 0.3.

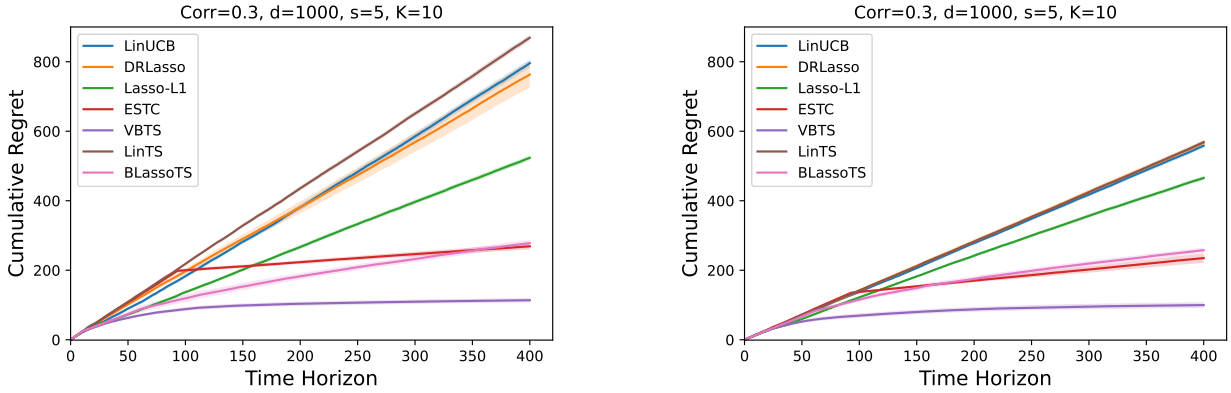


Figure 5. Regret bound for equi-correlated design: (Left) Setup 1, (Right) Setup 2

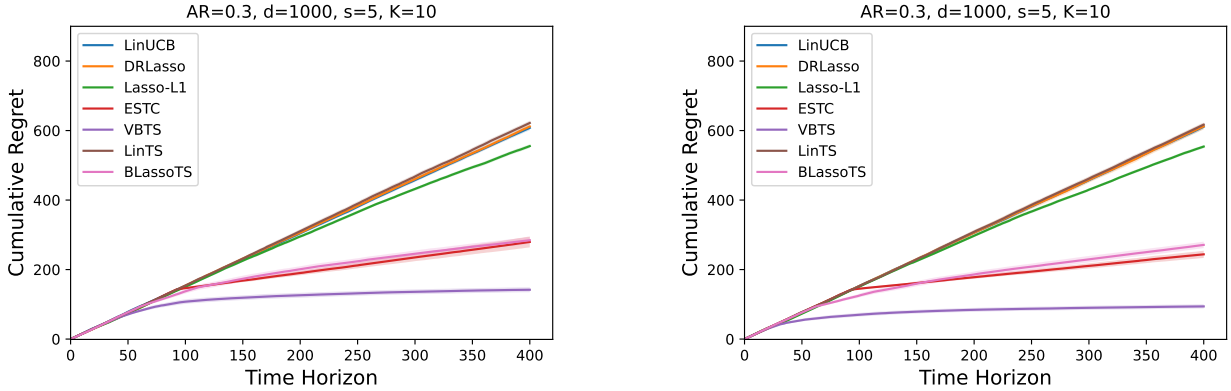


Figure 6. Regret bound for AR(1) design: (Left) Setup 1, (Right) Setup 2

B. Proof of Theorem 3.2

In this section, we present the detailed proof of Theorem 3.2. First, for clarity of presentation, we introduce some notations. We use X_t to denote the matrix $(x_{a_1}(1), \ldots, x_{a_t}(t))^\top \in \mathbb{R}^{t\times d}$. Given this, we denote the covariance matrix $\widehat{\Sigma}_t = X_t^\top X_t/t$. Next, we define the set

$$\mathbb{S}_0^{d-1}(s) \triangleq \mathbb{S}^{d-1} \cap \{v \ : \ \|v\|_0 \leq s\}.$$

We also define the following:

Definition B.1. For a index set $I \subseteq [d]$ and $\alpha \in \mathbb{R}^+$, we define the restricted cone as

$$\mathbb{C}_\alpha(I) := \{v \in \mathbb{R}^d \ : \ \|v_{I^c}\|_1 \leq \alpha \|v_I\|_1, v_I \neq 0\}$$

.

In high-dimensional literature one typically assumes compatibility condition on the design matrix X, i.e.,

$$\phi_{\mathsf{comp}}(S^*; X) := \inf_{\delta \in \mathbb{C}_7(S^*)} \frac{\|X\delta\|_2 |S^*|^{1/2}}{t^{1/2}\|\delta\|_1} > 0, \tag{4}$$

where $S^* = \{j : \beta_j^* \neq 0\}$. This is mainly to guarantee the estimation accuracy of high-dimensional estimators like LASSO (Bickel et al., 2009) or to show the posterior consistency in Bayesian high dimensional literature (Castillo et al., 2015).

As discussed in the main paper, we prove the theorem in three parts, the subsequent sections deal with each part separately.

*Proof outline In this section we first give a brief outline of the proof and discuss technical difficulties. We have three major components in this proof:

(i) In Section B.1 we will first show that that the gram matrix $\widehat{\Sigma}_t$ enjoys SRC condition with high probability for sufficiently large t. Based on this threshold on t, the burning time T_0 is chosen. The main difficulty here lies in that fact that the context sequence $(x_{a_1}(1), \ldots, x_{a_t}(t))$ are in general highly dependent on each other. Thus, existing high-dimensional concentration results are not directly applicable due to the inherent bandit environment. In our analysis we carefully decouple this complex dependent structure and exploit the special temporal independence structure of the bandit environment to establish SRC property of $\widehat{\Sigma}_t$.

(ii) Next, in Section B.2 we will establish the compatibility condition (4) for the matrix $\widehat{\Sigma}_t$. Again, due to the temporal dependence structure, we can not use the results from (Raskutti et al., 2010). Moreover, those results are only applicable for *independent multivariate Gaussian* design matrices, which may not be true in our case. Hence, we resort to different tools to show the compatibility condition. Essentially the main ingredient of the proof is the Transfer Lemma B.9 which essentially translates the uniform lower bound on $\phi_{\min}(Cs^*, \widehat{\Sigma}_t)$ to $\hat{\phi}_{\mathsf{comp}}(S^*)$, for sufficiently large C.

(iii) Finally, in Section B.3, under the compatibility condition we use the posterior contraction result from (Castillo et al., 2015) to give bound on the per round regret $\Delta_{a_t}(t)$. However, it is not readily trivial that the posterior contraction result is applicable due to the same temporal dependence of the bandit environment. It turns out that the contraction result essentially hinges on controlling the correlation between X_t and the vector $(\epsilon(1), \ldots, \epsilon(t))^\top$ at each time point t. The only challenge lies in controlling this correlation as the contexts X_t are not independent anymore. We resolve this issue by considering a suitable martingale difference sequence with respect to proper filtration.

In the subsequent sections we will prove the above three facts separately.

B.1. Proof of part (i)

In this section we will show that the matrix $\widehat{\Sigma}_t$ enjoys the SRC condition with high probability. As a warm up, we recall the definition of Orlicz norms:

Definition B.2 (Orlicz norms)**.** For random variable Z we have the followings:

(a) The *sub-Gaussian* norm of a random variable Z, denoted $\|Z\|_{\psi_2}$, is defined as

$$\|Z\|_{\psi_2} := \inf\{\lambda > 0 : \mathbb{E}\{\exp(Z^2/\lambda^2)\} \leq 2\}.$$

(b) The *sub-Exponential* norm of a random variable Z, denoted $\|Z\|_{\psi_1}$, is defined as

$$\|Z\|_{\psi_1} := \inf\{\lambda > 0 : \mathbb{E}\{\exp(|Z|/\lambda)\} \leq 2\}.$$

The details and related properties can be found in Section 2.5.2 and Section 2.7 in (Vershynin, 2018). The following is a relationship between sub-gaussian and sub-exponential random variables.

Lemma B.3 (Sub-Exponential and sub-Gaussian squared)**.** *A random variable Z is sub-Gaussian iff Z^2 is sub-Exponential. Moreover,*

$$\left\|Z^2\right\|_{\psi_1} = \|Z\|_{\psi_2}^2 .$$

Proof. The proof can be found in Lemma 2.7.4 of (Vershynin, 2018) □

Lemma B.4 (Bernstein's inequality)**.** *Let $Z_1, \ldots, Z_N$ be independent mean-zero sub-exponential random variable. Then for every $\delta \geq 0$ we have*

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} Z_i\right| \geq \delta\right) \leq 2\exp\left\{-c_2 \min\left(\frac{\delta^2}{K_0^2}, \frac{\delta}{K_0}\right) N\right\},$$

where $K_0 = \max_{i\in[N]} \|Z_i\|_{\psi_1}$.

Proof. The proof can be found in Corollary 2.8.3 of (Vershynin, 2018). □

Proposition B.5 (Empirical SRC). *Let $\varepsilon = \min\{1/4, 1/(\tilde{c}\phi_u\vartheta^2\xi K\log K + 3)\}$ for some universal constant $\tilde{c} > 0$ and also define the quantity $\kappa(\xi, \vartheta, K) \triangleq \min\{(4c_3K\xi\vartheta^2)^{-1}, 1/2\}$ for some universal positive constants c_3. Then the followings are true for any constant $C > 0$:*

$$\mathbb{P}\left(\phi_{\max}(Cs^*; \widehat{\Sigma}_t) \geq \frac{c_9\vartheta^2\phi_u \log K}{1-3\varepsilon}\right) \leq \exp\left\{-c_8 t\log K + Cs^*\log K + Cs^*\log(3d/\varepsilon)\right\},$$

$$\begin{aligned}\mathbb{P}\left(\phi_{\min}(Cs^*; \widehat{\Sigma}_t) \leq \frac{1}{8K\xi}\right) &\leq 2\exp\{-c_2\kappa^2(\xi,\vartheta,K)t + Cs^*\log K + Cs^*\log(3d/\varepsilon)\} \\ &\quad + \exp\left\{-c_8\log(K)t + Cs^*\log K + Cs^*\log(3d/\varepsilon)\right\},\end{aligned}$$

where all c_j's in the above display are universal positive constants.

Proof. We will first show the SRC condition for a fixed vector $v \in \mathbb{S}_0^{d-1}(Cs^*)$. Then, the whole argument will be extended via a ε-net argument.

Analysis for a fixed vector v: Let $v \in \mathbb{S}_0^{d-1}(Cs^*)$ be a fixed vector. Now note that following fact:

$$v^\top\widehat{\Sigma}_t v = \frac{1}{t}\sum_{\tau=1}^{t}\{v^\top x_{a_\tau}(\tau)\}^2 \geq \frac{1}{t}\sum_{\tau=1}^{t}\min_{i\in[K]}\{v^\top x_i(\tau)\}^2.$$

We define $Z_{\tau,v} \triangleq \min_{i\in[K]}\{v^\top x_i(\tau)\}^2$ and note that for a fixed $v \in \mathbb{S}_0^{d-1}(Cs^*)$, the random variables $\{Z_{\tau,v}\}_{\tau=1}^t$ are i.i.d. across the time points. Moreover, due to Assumption 2.2(b) and Lemma B.3, we have

$$\|Z_{\tau,v}\|_{\psi_1} = \left\|\min_{i\in[K]}\left|v^\top x_i(\tau)\right|\right\|_{\psi_2}^2 \leq c_3\vartheta^2.$$

Thus, $\{Z_{\tau,v}\}_{\tau=1}^t$ are i.i.d sub-exponential random variables. First we will show that $\mathbb{E}(Z_{\tau,v})$ is uniformly lower bounded. Due to Assumption 2.2(c) we have

$$\mathbb{P}(Z_{\tau,v} \leq h) \leq \sum_{i=1}^{K}\mathbb{P}\{(v^\top x_i(\tau))^2 \leq h\} \leq K\xi h.$$

Thus we have the following:

$$\begin{aligned}\mathbb{E}(Z_{\tau,v}) &= \int_0^\infty \mathbb{P}(Z_{\tau,v} \geq u)\, du \\ &\geq \int_0^h \mathbb{P}(Z_{\tau,v} \geq u)\, du \\ &\geq \int_0^h (1 - K\xi u)\, du \\ &= h(1 - K\xi h/2).\end{aligned} \tag{5}$$

Setting $h = 1/(K\xi)$ in Equation (5) yields $\mathbb{E}(Z_{\tau,v}) \geq \frac{1}{2K\xi}$. Now, using Lemma B.4, we have the following for a $\mu \in (0, c_1/\|Z_{1,v}\|_{\psi_1})$ and $\delta > 0$:

$$\begin{aligned}\mathbb{P}\left(\frac{1}{t}\sum_{\tau=1}^{t}\{Z_{\tau,v} - \mathbb{E}(Z_{\tau,v})\} \geq \delta\right) &\leq 2\exp\left\{-c_2\min\left(\frac{\delta^2}{\|Z_{1,v}\|_{\psi_1}^2}, \frac{\delta}{\|Z_{1,v}\|_{\psi_1}}\right)t\right\} \\ &\leq 2\exp\left\{-c_2\min\left(\frac{\delta^2}{c_3^2\vartheta^4}, \frac{\delta}{c_3\vartheta^2}\right)t\right\}.\end{aligned}$$

Now choose $\delta = \min\{(4K\xi)^{-1}, c_3\vartheta^2/2\}$ to finally get

$$\mathbb{P}\left(\left|\frac{1}{t}\sum_{\tau=1}^{t}\{Z_{\tau,v} - \mathbb{E}(Z_{\tau,v})\}\right| \geq \frac{1}{4K\xi}\right) \leq 2\exp\left\{-c_2\kappa^2(\xi,\vartheta,K)t\right\}, \tag{6}$$

where $\kappa(\xi,\vartheta,K) = \min\{(4c_3K\xi\vartheta^2)^{-1}, 1/2\}$. Now recall that $\mathbb{E}(Z_{\tau,v}) \geq 1/(2K\xi)$, which shows that

$$\mathbb{P}\left\{\frac{1}{t}\sum_{\tau=1}^{t}Z_{\tau,v} \geq \frac{1}{4K\xi}\right\} \leq 2\exp\left\{-c_2\kappa^2(\xi,\vartheta,K)t\right\}, \quad \forall\, v \in \mathbb{S}_0^{d-1}(Cs^*). \tag{7}$$

ε-net argument: We consider a ε-net of the space $\mathbb{S}_0^{d-1}(Cs^*)$ constructed in a specific way which will be described shortly. We denote it by $\mathcal{N}_\varepsilon$. Let $J \subseteq [d]$ such that $|J| = Cs^*$ and consider the set $E_J = \mathbb{S}^{d-1} \cap \mathsf{span}\{e_j \ : \ j \in J\}$. Here e_j denotes the jth canonical basis of $\mathbb{R}^d$. Thus we have

$$\mathbb{S}_0^{d-1}(Cs^*) = \bigcup_{J:|J|=Cs^*} E_J.$$

Now we describe the procedure of constructing a net for $\mathbb{S}_0^{d-1}(Cs^*)$ which is essential for controlling the parse eigenvalues.
Greedy construction of net:

- Construct a ε-net of E_J for each J of size Cs^*. We denote this net by $\mathcal{N}_{\varepsilon,J}$. Note that $|\mathcal{N}_{\varepsilon,J}| \leq (3/\varepsilon)^{Cs^*}$ (Vershynin, 2018, Corollary 4.2.13) for $\varepsilon \in (0,1)$ as E_J can be viewed as an unit ball embedded in $\mathbb{R}^{Cs^*}$.
- Then the net $\mathcal{N}_\varepsilon$ is constructed by taking union over all the $\mathcal{N}_{\varepsilon,J}$, i.e.,

$$\mathcal{N}_\varepsilon = \bigcup_{J:|J|=Cs^*} \mathcal{N}_{\varepsilon,J}.$$

Thus, from from the construction we have

$$|\mathcal{N}_\varepsilon| \leq \binom{d}{Cs^*}\left(\frac{3}{\varepsilon}\right)^{Cs^*} \leq \exp\{Cs^*\log(3d/\varepsilon)\}, \tag{8}$$

whenever $\varepsilon \in (0,1)$. Now, we state an useful lemma on evaluating minimum eigenvalue on ε-net.

Lemma B.6. *Let A be a $m \times m$ symmetric positive-definite matrix and $\varepsilon \in (0,1)$. Then, for ε-net $\mathcal{N}_\varepsilon$ of $\mathbb{S}_0^{d-1}(s)$ constructed in greedy way, we have*

$$\phi_{\min}(s;A) \geq \min_{u\in\mathcal{N}_\varepsilon} u^\top Au - 3\varepsilon\phi_{\max}(s;A).$$

The proof of the lemma is deferred to Appendix D.1. Note that from Equation (7) and an union bound argument we get

$$\mathbb{P}\left(\min_{u\in\mathcal{N}_\varepsilon} v^\top\widehat{\Sigma}_t v \geq \frac{1}{4K\xi}\right) \geq 1 - 2\exp\{-c_2\kappa^2(\xi,\vartheta,K)t + Cs^*\log K + Cs^*\log(3d/\varepsilon)\}. \tag{9}$$

If $\phi_{\max}(Cs^*, \widehat{\Sigma}_t)$ is bounded with high probability, then for small ε, then along with Lemma B.6 we will readily have an uniform lower bound on $\phi_{\min}(Cs^*, \widehat{\Sigma}_t)$.
Bounding $\phi_{\max}(Cs^*, \widehat{\Sigma}_t)$: Here we gain start with $v \in \mathcal{N}_\varepsilon$. Similar, to previous discussion we have

$$v^\top\widehat{\Sigma}_t v = \frac{1}{t}\sum_{\tau=1}^{t}\{v^\top x_{a_\tau}(\tau)\}^2 \leq \frac{1}{t}\sum_{\tau=1}^{t}\max_{i\in[K]}\{v^\top x_i(\tau)\}^2.$$

We define $W_{\tau,v} \triangleq \max_{i\in[K]}\{v^\top x_i(\tau)\}^2$ and note that for a fixed $v \in \mathbb{S}_0^{d-1}(Cs^*)$, the random variables $\{W_{\tau,v}\}_{\tau=1}^t$ are i.i.d. across the time points. Moreover, due to Assumption 2.2(b) and Lemma B.3, we have

$$\|W_{\tau,v}\|_{\psi_1} = \left\|\max_{i\in[K]}\{v^\top x_i(\tau)\}^2\right\|_{\psi_1} \le c_4 K\vartheta^2.$$

Thus, $\{W_{\tau,v}\}_{\tau=1}^t$ are i.i.d sub-exponential random variables. Recall, that $\phi_{\max}(Cs^*,\Sigma_i) \le \phi_u$ for all $i \in [K]$. The next lemma provides an upper bound on the moment generating function (MGF) of sub-Exponential random variables.

Lemma B.7. *(Vershynin, 2018, Lemma 2.8.1) Let X be a mean-zero, sub-Exponential random variable. Then there exists positive constants c_5, c_6, such that for any λ with $|\lambda| \le c_5/\|X\|_{\psi_1}$, the following is true:*

$$\mathbb{E}\{\exp(\lambda X)\} \le \exp(c_6\lambda^2\|X\|_{\psi_1}^2).$$

Equipped with the above lemma we have the following;

$$\begin{aligned}\mathbb{P}\left(\frac{1}{t}\sum_{\tau=1}^t W_{\tau,v} - \phi_u \ge \delta\right) &= \mathbb{P}\left(\sum_{\tau=1}^t\{W_{\tau,v}-\phi_u\} \ge \delta t\right)\\ &= \mathbb{P}\left(\exp\left\{\mu\sum_{\tau=1}^t(W_{\tau,v}-\phi_u)\right\} \ge e^{\mu\delta t}\right)\\ &\le e^{-\mu\delta t}\prod_{\tau=1}^t \mathbb{E}\{e^{\mu(W_{\tau,v}-\phi_u)}\}.\end{aligned} \tag{10}$$

For a fixed $\tau \in [t]$ we note the following;

$$\begin{aligned}\mathbb{E}\{e^{\mu(W_{\tau,v}-\phi_u)}\} &\le \sum_{i=1}^K \mathbb{E}\{e^{\mu[(v^\top x_i(\tau))^2-\phi_u]}\}\\ &\le \sum_{i=1}^K \mathbb{E}\{e^{\mu[(v^\top x_i(\tau))^2-v^\top\Sigma_i v]}\}.\end{aligned}$$

For brevity let $\kappa_i \triangleq \left\|\{v^\top x_i(\tau)\}^2\right\|_{\psi_1}$. If we choose $\mu \le c_5/\max_{i\in[K]}\kappa_i$, then by Lemma B.7 we have

$$\mathbb{E}\{e^{\mu(W_{\tau,v}-\phi_u)}\} \le \exp\left(c_6\mu^2\max_{i\in[K]}\kappa_i^2 + \log K\right).$$

Using the above inequality in Equation (10), it follows that

$$\mathbb{P}\left(\frac{1}{t}\sum_{\tau=1}^t W_{\tau,v} - \phi_u \ge \delta\right) \le \exp\left(-\mu\delta t + c_6 t\mu^2\max_{i\in[K]}\kappa_i^2 + t\log K\right). \tag{11}$$

The right hand side of Equation (11) is minimized at $\mu = \delta/(2c_2\max_{i\in[K]}\kappa_i^2)$ with the minimum value of

$$\exp\left(-\frac{\delta^2 t}{4c_6\max_{i\in[K]}\kappa_i^2} + t\log K\right).$$

If $\delta/(2c_6\max_{i\in[K]}\kappa_i^2) > c_5/\max_{i\in[K]}\kappa_i^2$, then the right hand side of Equation (11) is minimized at $\mu = c_5/\max_{i\in[K]}\kappa_i^2$ and we get

$$\mathbb{P}\left(\frac{1}{t}\sum_{\tau=1}^t W_{\tau,v} - \phi_u \ge \delta\right) \le \exp\left(-\frac{c_5\delta t}{\max_i\kappa_i} + c_6c_5^2 t + t\log K\right).$$

Using the fact that $\delta/(2c_6 \max_{i\in[K]} \kappa_i^2) > c_5/\max_{i\in[K]} \kappa_i^2$, the right hand side of the above display can be upper bounded by

$$\exp\left(-\frac{c_5\delta t}{2\max_i \kappa_i} + t\log K\right).$$

Thus we have for all $\delta > 0$

$$\mathbb{P}\left(\frac{1}{t}\sum_{\tau=1}^{t} W_{\tau,v} - \phi_u \geq \delta\right) \leq \exp\left(-\min\left\{\frac{\delta^2 t}{4c_6 \max_{i\in[K]} \kappa_i^2}, \frac{c_5\delta t}{2\max_i \kappa_i}\right\} + t\log K\right). \tag{12}$$

Next we set $\delta = c_7\vartheta^2\phi_u \log K$ for sufficiently large $c_7 > 0$. Then Equation (12) yields

$$\mathbb{P}\left(\frac{1}{t}\sum_{\tau=1}^{t} W_{\tau,v} - \phi_u \geq \delta\right) \leq \exp\left(-c_8 t\log K\right).$$

Finally taking union bond over all vectors in $\mathcal{N}_\varepsilon$ we get

$$\mathbb{P}\left(\forall v \in \mathcal{N}_\varepsilon : v^\top \widehat{\Sigma}_t v \geq c_9\vartheta^2\phi_u \log K\right) \leq \exp\left\{-c_8 t\log K + Cs^*\log K + Cs^*\log(3d/\varepsilon)\right\}. \tag{13}$$

Next, to prove the same for all $v \in \mathbb{S}_0^{d-1}(Cs^*)$ we need the following lemma.

Lemma B.8 (maximum sparse eigenvalue on net)**.** *Let A be a $m \times m$ symmetric positive-definite matrix and $\varepsilon \in (0, 1/3)$. Then, for ε-net $\mathcal{N}_\varepsilon$ of $\mathbb{S}_0^{d-1}(s)$ constructed in greedy way, we have*

$$\max_{v\in\mathcal{N}_\varepsilon} v^\top A v \leq \max_{v\in\mathbb{S}_0^{d-1}(Cs^*)} v^\top A v \leq \frac{1}{1-3\varepsilon}\max_{v\in\mathcal{N}_\varepsilon} v^\top A v.$$

The proof of the above lemma is deferred to Appendix D.2. Now we set some $\varepsilon \in (0, 1/3)$. In light of the above lemma we immediately have that

$$\mathbb{P}\left(\phi_{\max}(Cs^*; \widehat{\Sigma}_t) \geq \frac{c_9\vartheta^2\phi_u\log K}{1-3\varepsilon}\right) \leq \exp\left\{-c_8 t\log K + Cs^*\log K + Cs^*\log(3d/\varepsilon)\right\}. \tag{14}$$

Finally using Equation (9), (14) and Lemma B.6 we have

$$\begin{aligned}
&\mathbb{P}\left(\phi_{\min}(Cs^*; \widehat{\Sigma}_t) \geq \frac{1}{4K\xi} - \frac{3\varepsilon c_9\vartheta^2\log K}{1-3\varepsilon}\phi_u\right) \\
&\geq 1 - 2\exp\{-c_2\kappa^2(\xi,\vartheta,K)t + Cs^*\log K + Cs^*\log(3d/\varepsilon)\} \\
&\quad - \exp\left\{-c_8 t\log K + Cs^*\log K + Cs^*\log(3d/\varepsilon)\right\}.
\end{aligned} \tag{15}$$

Now the result follows from taking $\varepsilon = \min\{1/4, 1/(24\phi_u\vartheta^2\xi K\log K + 3)\}$. □

B.2. Proof of part (ii)

In this section we will show that the matrix $\widehat{\Sigma}_t$ enjoys the compatibility condition (4) with high probability. This is equivalent to showing that the quantity

$$\Psi(S^*; \widehat{\Sigma}_t) \triangleq \inf_{\delta\in\mathbb{C}_7(S^*)}\left(\frac{\delta^\top\widehat{\Sigma}_t\delta}{\|\delta\|_1^2}\right)s^* = \phi_{\mathsf{comp}}(S^*; X_t)^2.$$

is bounded away from 0 with high probability. First we present the Transfer lemma (Oliveira, 2013, Lemma 5) below.

Lemma B.9 (Transfer lemma)**.** *Suppose $\widehat{\Sigma}_t$ and Σ are matrix with non-negative diagonal entries, and assume $\eta \in (0, 1)$, $m \in [d]$ are such that*

$$\forall v \in \mathbb{R}^d \text{ with } \|v\|_0 \leq m, v^\top\widehat{\Sigma}_t v \geq (1-\eta)v^\top\Sigma v. \tag{16}$$

Assume D is a diagonal matrix whose diagonal entries $D_{j,j}$ are non-negative and satisfy $D_{j,j} \geq (\widehat{\Sigma}_t)_{j,j} - (1-\eta)\Sigma_{j,j}$. Then

$$\forall x \in \mathbb{R}^d, x^\top\widehat{\Sigma}_t x \geq (1-\eta)x^\top\Sigma x - \frac{\|D^{1/2}x\|_1^2}{m-1}. \tag{17}$$

Condition (16) basically demands that $\widehat{\Sigma}_t$ enjoys SRC condition with the sparsity parameter m. Then under the proper choice of diagonal matrix D with sufficiently large diagonal elements $\{D_{j,j}\}_{j=1}^d$, Equation (17) will yield the desired compatibility condition for $\widehat{\Sigma}_t$. We formally state the result in the following lemma:

Proposition B.10 (Empirical compatibility condition)**.** *Assume the conditions of Proposition B.5 hold. Also assume that Assumption 2.2(d) holds with $C = C_0\phi_u\vartheta^2\xi K\log K$ for some sufficiently large universal constant $C_0 > 0$. Then there exists a positive constant C_1 such that the following is true:*

$$\begin{aligned}
&\mathbb{P}\left(\Psi(S^*;\widehat{\Sigma}_t) \geq \frac{1}{C_1\xi K}\right)\\
&= 1 - 2\exp\{-c_2\kappa^2(\xi,\vartheta,K)t + Cs^*\log K + Cs^*\log(3d/\varepsilon)\}\\
&\quad - 2\exp\left\{-c_8 t\log K + Cs^*\log K + Cs^*\log(3d/\varepsilon)\right\}
\end{aligned}$$

with $\varepsilon = \min\{1/4, 1/(\tilde{c}\phi_u\vartheta^2\xi K\log K + 3)\}$ for the same universal constant $\tilde{c} > 0$ in Proposition B.5 and $\kappa(\xi,\vartheta,K) = \min\{(4c_3K\xi\vartheta^2)^{-1}, 1/2\}$.

Proof. As suggested before we will make use of Lemma B.9. Towards this, we set $\Sigma = \frac{1}{4K\xi}\mathbb{I}_d$ and $D = \mathsf{diag}(\widehat{\Sigma}_t)$. Next, we define the following two events:

$$\mathcal{G}_{t,1} := \left\{\phi_{\max}(Cs^*;\widehat{\Sigma}_t) \leq \frac{c_9\vartheta^2\phi_u\log K}{1-3\varepsilon}\right\},$$

$$\mathcal{G}_{t,2} := \left\{\phi_{\min}(Cs^*;\widehat{\Sigma}_t) \geq \frac{1}{8K\xi}\right\},$$

where the constants ε and c_9 are same as in Proposition B.5. Under $\mathcal{G}_{t,1}$ and $\mathcal{G}_{t,2}$, the inequality in Equation (16) holds with $\eta = 1/2$ and $m = Cs^*$. Also, due construction of D, we trivially have

$$D_{j,j} \geq (\widehat{\Sigma}_t)_{j,j} - (1-\eta)\Sigma_{j,j}.$$

Lastly, note that

$$\max_{j\in[d]} D_{j,j} = \max_{j\in[d]}(\widehat{\Sigma}_t)_{j,j} = \max_{j\in[d]} e_j^\top\widehat{\Sigma}_t e_j \leq \frac{c_9\vartheta^2\phi_u\log K}{1-3\varepsilon} \tag{18}$$

under $\mathcal{G}_{t,1}$. Equipped with Lemma B.9, under $\mathcal{G}_{t,1}$ and $\mathcal{G}_{t,2}$, for all $x \in \mathbb{C}_7(S^*)\cap\mathbb{S}^{d-1}$ we have the following:

$$\begin{aligned}
x^\top\widehat{\Sigma}_t x &\geq \frac{1}{8K\xi} - \frac{\left\|D^{1/2}x\right\|_1^2}{Cs^*-1}\\
&\geq \frac{1}{8K\xi} - \frac{\left(\frac{c_9\vartheta^2\phi_u\log K}{1-3\varepsilon}\right)\|x\|_1^2}{Cs^*-1}\\
&\geq \frac{1}{8K\xi} - \frac{\left(\frac{c_9\vartheta^2\phi_u\log K}{1-3\varepsilon}\right)64s^*}{Cs^*-1}.
\end{aligned}$$

The last inequality follows form the fact that

$$\|x\|_1 = \|x_{S^*}\|_1 + \left\|x_{(S^*)^c}\right\|_1 \leq 8\|x_{S^*}\|_1 \leq 8\sqrt{s^*}\|x_{S^*}\|_2 \leq 8\sqrt{s^*}. \tag{19}$$

Thus, if $C \gtrsim \frac{\phi_u\vartheta^2\xi K\log K}{1-3\varepsilon} + \frac{1}{s^*}$ then $x^\top\widehat{\Sigma}_t x \geq 1/(16K\xi)$. Also, note that from the choice of ε in Proposition B.5, we have $\varepsilon < 1/4$. This further tells that if $C = C_0\phi_u\vartheta^2\xi K\log K$ for large enough $C_0 > 0$, then

$$\inf_{\delta\in\mathbb{C}_7(S^*)}\frac{\delta^\top\widehat{\Sigma}_t\delta}{\|\delta\|_2^2} \geq \frac{1}{16K\xi}.$$

Then, the result follows from Proposition B.5. Using this and Equation (19) we also have

$$\Psi(S^*;\widehat{\Sigma}_t) \geq \frac{1}{64}\inf_{\delta\in\mathbb{C}_7(S^*)}\frac{\delta^\top\widehat{\Sigma}_t\delta}{\|\delta\|_2^2} \geq \frac{1}{C_1\xi K}$$

where $C_1 = 1024$. Finally, the result follows from conditioning over the events $\mathcal{G}_{t,1}$ and $\mathcal{G}_{t,2}$ and using Proposition B.5.

□

B.3. Proof of part (iii)

In this section we will establish the desired regret bound in Theorem 3.2. The main tools that has been used to prove the regret bound is the Bayesian contraction in high-dimensional linear regression problem. In particular, we will use Theorem C.1 to control the ℓ_1-distance between $\tilde{\beta}_t$ and β^* at each time point $t \in [T]$.

We recall that $X_t = (x_{a_1}(1), \ldots, x_{a_t}(t))^\top$. Also, note that the sequence $\{x_{a_\tau}(\tau)\}_{\tau=1}^t$ forms an adapted sequence of observations, i.e., $x_{a_\tau}(\tau)$ may depend on the history $\{x_{a_u}(u), r(u)\}_{u=1}^{\tau-1}$. Also, recall that $\{\epsilon(\tau)\}_{\tau=1}^t$ are mean-zero σ-sub-Gaussian errors.

Lemma B.11 (Bernstein Concentration)**.** *Let $\{D_k, \mathcal{F}_k\}_{k=1}^\infty$ be a martingale difference sequence, and suppose that D_k is a σ-sub-Gaussian in adapted sense, i.e., for all $\alpha \in \mathbb{R}, \mathbb{E}[e^{\alpha D_k} \mid \mathcal{F}_{k-1}] \leq e^{\alpha^2\sigma^2/2}$ almost surely. Then, for all $t \geq 0$, $\mathbb{P}\left(\left|\sum_{k=1}^t D_k\right| \geq \delta\right) \leq 2\exp\{-\delta^2/(2t\sigma^2)\}$.*

Proof. Proof of Lemma B.11 follows from Theorem 2.19 of (Wainwright, 2019) by setting $\alpha_k = 0$ and $\nu_k = \sigma$ for all k. □

Lemma B.11 is the main tool that is used to control the correlation between $\epsilon_t := (\epsilon(1), \ldots, \epsilon(t))^\top$ and the chosen contexts X_t which is important to control the Bayesian contraction of the posterior distribution in each round. To elaborate, let $X_t^{(j)}$ be the jth column for $j \in [d]$ and define $D_{t,j} := \epsilon(t)x_{a_t,j}(t)$. Note that for a fixed $j \in [d]$, $\{D_{\tau,j}\}_{\tau=1}^t$ forms a martingale difference sequence with respect to the filtration $\{\mathcal{F}_\tau\}_{\tau=1}^{t-1}$ with $\mathcal{F}_\tau := \sigma(\mathcal{H}_\tau)$ is the σ-algebra generated by $\mathcal{H}_\tau$ and $\mathcal{F}_1 = \emptyset$. Also note that

$$\mathbb{E}(e^{\alpha D_{t,j}} \mid \mathcal{F}_{t-1}) \leq \mathbb{E}\{e^{\alpha^2\sigma^2 x_{a_t,j}^2(t)/2}\} \leq \mathbb{E}\{e^{\alpha^2\sigma^2 x_{\max}^2/2}\}.$$

Thus, using Lemma B.11, we have the following proposition:

Proposition B.12 (Lemma EC.2, (Bastani & Bayati, 2020))**.** *Define the event*

$$\mathcal{T}_t(\lambda_0(\gamma)) := \left\{\max_{j\in[d]}\frac{\left|\epsilon_t^\top X_t^{(j)}\right|}{t} \leq \lambda_0(\gamma)\right\},$$

where $\lambda_0(\gamma) = x_{\max}\sigma\sqrt{(\gamma^2 + 2\log d)/t}$. Then we have $\mathbb{P}\{\mathcal{T}_t(\lambda_0(\gamma))\} \geq 1 - 2\exp(-\gamma^2/2)$.

The proof of the above proposition mainly relies on the martingale difference structure and Lemma B.4. It is important to mention that the proof does not depend on any particular algorithm.

For notational brevity we define $\|X_t\| := \max_{j\in[d]}\sqrt{(X_t^\top X_t)_{j,j}}$. Next, we will set $\gamma = \gamma_t := \sqrt{2\log t}$. Hence by Proposition B.12 we have $\mathbb{P}\{\mathcal{T}_t(\lambda_0(\gamma_t))\} \geq 1 - 2t^{-1}$. Also recall that, under $\mathcal{G}_{t,1}$ and $\mathcal{G}_{t,2}$, we have

$$\frac{1}{\sqrt{8K\xi}} \leq \|X_t\|/\sqrt{t} \leq \sqrt{4c_9\phi_u\vartheta^2\log K}. \tag{20}$$

Also, under $\mathcal{G}_{t,2} \cap \mathcal{T}_t(\lambda_0(\gamma_t))$ it follows that

$$\max_{j\in[d]}\frac{\left|\epsilon_t^\top X_t^{(j)}\right|}{\sigma} \leq x_{\max}\sqrt{2t(\log d + \log t)} = \overline{\lambda}_t \tag{21}$$

Now, we are ready to present the proof of the main regret bound.

MAIN REGRET BOUND

Recall the definition of regret is $R(T) = \sum_{t=1}^{T} \Delta_{a_t}(t)$, where $\Delta_{a_t}(t) = x_{a_t^*}(t)^\top \beta^* - x_{a_t}(t)^\top \beta^*$. Next, we partition the whole time horizon $[T]$ in to two parts, namely $\{t : 1 \le t \le T_0\}$ and $\{t : T_0 \le t \le T\}$, where T_0 will be chosen later. Thus, the regret can be written as

$$R(T) = \underbrace{\sum_{t=1}^{T_0} \Delta_{a_t}(t)}_{R(T_0)} + \underbrace{\sum_{t=T_0+1}^{T} \Delta_{a_t}(t)}_{\tilde{R}(T)}.$$

All notations for expectation operators and probability measures are given in Appendix C.

Now by Assumption 2.2(a) and 2.3(a) we have the following inequality:

$$\mathbb{E}\{R(T_0)\} \le 2x_{\max} b_{\max} T_0. \tag{22}$$

Next, we focus on the term $\tilde{R}(T)$. First, we define a few quantities below:

$$\begin{aligned} \overline{\phi}_t(s) &:= \inf_{\delta} \left\{ \frac{\|X_t\delta\|_2 \, |S_\delta|^{1/2}}{t^{1/2} \, \|\delta\|_1} : 0 \ne |S_\delta| \le s \right\}, \\ \widetilde{\phi}_t(s) &:= \inf_{\delta} \left\{ \frac{\|X_t\delta\|_2}{t^{1/2} \, \|\delta\|_2} : 0 \ne |S_\delta| \le s \right\}. \end{aligned} \tag{23}$$

Now set

$$\overline{\psi}_t(S) = \bar{\phi}_t \left(\left(2 + \frac{40}{A_4} + \frac{128 A_4^{-1} x_{\max}^2}{\Psi(S, \widehat{\Sigma}_t)} \right) |S| \right),$$

$$\widetilde{\psi}_t(S) = \widetilde{\phi}_t \left(\left(2 + \frac{40}{A_4} + \frac{128 A_4^{-1} x_{\max}^2}{\Psi(S, \widehat{\Sigma}_t)} \right) |S| \right).$$

Note that $\bar{\phi}_t(s) \ge \widetilde{\phi}_t(s)$, hence $\bar{\psi}_t(S) \ge \widetilde{\psi}_t(S)$.

Recall that

$$5\overline{\lambda}_t/3 \le \lambda_t \le 2\overline{\lambda}_t, \tag{24}$$

under $\mathcal{G}_{t,1}$ and $\mathcal{G}_{t,2}$. Also define the following events:

$$\mathcal{E}_t := \left\{ \left\| \tilde{\beta}_{t+1} - \beta^* \right\|_1 \le Q_4 \sigma x_{\max} K \xi (D_* + s^*) \sqrt{\frac{\log d + \log t}{t}} \right\}.$$

where $D_* = \{1 + (40/A_4) + 128 A_4^{-1} x_{\max}^2 / \Psi(S^*, \widehat{\Sigma}_t)\} s^*$ and Q_4 is large enough universal constant as specified in Theorem C.1. Also we have

$$\bar{\psi}_t(S) \le \bar{\phi}_t \left(\left(2 + \frac{40}{A_4} + \frac{64 A_4^{-1} x_{\max}^2}{\Psi(S, \widehat{\Sigma}_t)} \frac{\lambda}{\overline{\lambda}} \right) |S| \right), \text{ and}$$

$$\widetilde{\psi}_t(S) \le \widetilde{\phi}_t \left(\left(2 + \frac{40}{A_4} + \frac{64 A_4^{-1} x_{\max}^2}{\Psi(S, \widehat{\Sigma}_t)} \frac{\lambda}{\overline{\lambda}} \right) |S| \right).$$

Next, by Proposition B.10, the event

$$\mathcal{G}_{t,3} := \left\{ \Psi(S^*; \widehat{\Sigma}_t) \ge \frac{1}{C_1 \xi K} \right\}$$

holds with probability of at least $1 - 2\exp\{-c_2 \kappa^2(\xi, \vartheta, K) t + C s^* \log K + C s^* \log(3d/\varepsilon)\} - 2\exp\{-c_8 t \log K + C s^* \log K + C s^* \log(3d/\varepsilon)\}$. For, notational brevity, we define

$$\tilde{C} := C_1 \xi K.$$

Also, define the event

$$\mathcal{G}_{t,4} := \left\{ \widetilde{\psi}_t^2(S^*) \geq \frac{1}{8K\xi} \right\}.$$

Noting that $\widetilde{\psi}_t^2(S^*) = \phi_{\min}(\tilde{C}_1 s^*; \widehat{\Sigma}_t)$ with

$$\tilde{C}_1 = 2 + \frac{40}{A_4} + 128 A_4^{-1} x_{\mathsf{max}}^2 \tilde{C},$$

an argument similar to the proof of Proposition B.5 yields

$$\begin{aligned} \mathbb{P}\left(\mathcal{G}_{t,4}\right) \geq 1 - 2\exp\{-c_2\kappa^2(\xi,\vartheta,K)t + \tilde{C}_1 s^* \log K + \tilde{C}_1 s^* \log(3d/\varepsilon)\} \\ - \exp\left\{-c_8 \log(K) t + \tilde{C}_1 s^* \log K + \tilde{C}_1 s^* \log(3d/\varepsilon)\right\}. \end{aligned} \tag{25}$$

Finally, Using Proposition B.12 with $\gamma = \gamma_d$ and the result of Theorem C.1, we have the following:

$$\begin{aligned} \mathbb{P}(\mathcal{E}_t^c) &= \mathbb{E}_{X_t} \mathbb{E}_t^X \left(\mathbb{1}_{\mathcal{E}_t^c} \right) \\ &= \mathbb{E}_{X_t} \mathbb{E}_{t,\mathbf{r}_t}^X \left\{ \Pi_t^X \left(\mathcal{E}_t^c \mid \mathbf{r}_t \right) \right\} \\ &= \mathbb{E}_{X_t} \mathbb{E}_{t,\mathbf{r}_t}^X \left\{ \Pi_t^X \left(\mathcal{E}_t^c \mid \mathbf{r}_t \right) \mathbb{1}_{\mathcal{T}_t(\lambda_0(\gamma_t)) \cap \mathcal{G}_{t,2}} \right\} + \mathbb{E}_{X_t} \mathbb{E}_{t,\mathbf{r}_t}^X \left\{ \Pi_t^X \left(\mathcal{E}_t^c \mid \mathbf{r}_t \right) \mathbb{1}_{\mathcal{T}_t^c(\lambda_0(\gamma_t)) \cup \mathcal{G}_{t,2}^c} \right\} \\ &\leq \frac{M_1}{d^{s^*}} + \frac{2}{t} + 2\exp\{-c_2\kappa^2(\xi,\vartheta,K)t + Cs^* \log K + Cs^* \log(3d/\varepsilon)\} \\ &\quad + \exp\left\{-c_8 t \log K + Cs^* \log K + Cs^* \log(3d/\varepsilon)\right\} \end{aligned} \tag{26}$$

for some large universal constant $M_1 > 0$. Next, let $\mathcal{G}_t = \cap_{i=1}^4 \mathcal{G}_{t,i}$. Under the event $\mathcal{E}_t \cap \mathcal{G}_t$, we have

$$D_* + s^* \leq \underbrace{\left(2 + \frac{40}{A_4} + \frac{128 C_1 K \xi x_{\mathsf{max}}^2}{A_4} \right)}_{:=\rho} s^*,$$

and,

$$\left\| \tilde{\beta}_{t+1} - \beta^* \right\|_1 \leq M_2 \rho \sigma x_{\mathsf{max}} \xi K \left\{ \frac{s^{*2}(\log d + \log t)}{t} \right\}^{1/2},$$

where M_2 is an universal constant depending on A_4. Now we set

$$\delta_t = M_2 \rho \sigma x_{\mathsf{max}}^2 \xi K \left\{ \frac{s^{*2}(\log d + \log t)}{t} \right\}^{1/2}.$$

It follows that under $\mathcal{E}_{t-1} \cap \mathcal{G}_{t-1}$, we have the following almost sure inequality:

$$\begin{aligned} \Delta_{a_t}(t) &= x_{a_t^*}^\top(t)\beta^* - x_{a_t}^\top(t)\beta^* \\ &= x_{a_t^*}^\top(t)\beta^* - x_{a_t^*}^\top(t)\tilde{\beta}_t + \underbrace{\left(x_{a_t^*}^\top(t)\tilde{\beta}_t - x_{a_t}^\top(t)\tilde{\beta}_t\right)}_{\leq 0} + x_{a_t^*}^\top(t)\tilde{\beta}_t - x_{a_t}^\top(t)\beta^* \\ &\leq \left\| x_{a_t^*}(t) \right\|_\infty \|\tilde{\beta}_t - \beta^*\|_1 + \|x_{a_t}(t)\|_\infty \|\tilde{\beta}_t - \beta^*\|_1 \\ &\leq 2\delta_{t-1}. \end{aligned}$$

Finally define the event

$$\mathcal{M}_t := \left\{ x_{a_t^*}^\top \beta^* > \max_{i \neq a_t^*} x_{a_t}^\top \beta^* + h_{t-1} \right\}.$$

Under $\mathcal{M}_t \cap \mathcal{E}_{t-1} \cap \mathcal{G}_{t-1}$, we have the following for any $i \neq a_t^*$:

$$\begin{aligned} x_{a_t^*}^\top(t)\tilde{\beta}_t - x_i^\top(t)\tilde{\beta}_t &= \left\langle x_{a_t^*}(t), \tilde{\beta}_t - \beta^* \right\rangle + \langle x_{a_t}(t) - x_i(t), \beta^* \rangle + \left\langle x_i(t), \beta^* - \tilde{\beta}_t \right\rangle \\ &\geq -\delta_{t-1} + h_{t-1} - \delta_{t-1}. \end{aligned}$$

Thus, if we set $h_{t-1} = 3\delta_{t-1}$ then $x_{a_t^*}^\top(t)\tilde{\beta}_t - \max_{i \neq a_t^*} x_i^\top(t)\tilde{\beta}_t \geq \delta_{t-1}$. As a result, in tth round the regret is 0 almost surely as the optimal arm will be chosen with probability 1. Thus, finally using Assumption 2.3(b), we have

$$\begin{aligned} \mathbb{E}(\Delta_{a_t}(t)) &= \mathbb{E}\{\Delta_{a_t}(t)\mathbb{1}_{\mathcal{M}_t^c}\} \\ &= \mathbb{E}\{\Delta_{a_t}(t)\mathbb{1}_{\mathcal{M}_t^c \cap \mathcal{E}_{t-1} \cap \mathcal{G}_{t-1}}\} + \mathbb{E}\{\Delta_{a_t}(t)\mathbb{1}_{\mathcal{M}_t^c \cap (\mathcal{E}_{t-1} \cap \mathcal{G}_{t-1})^c}\} \\ &\leq 2\delta_{t-1}\mathbb{P}(\mathcal{M}_t^c) + 2x_{\mathsf{max}}b_{\mathsf{max}}\mathbb{P}(\mathcal{E}_t^c \cup \mathcal{G}_t^c) \\ &\leq 2\delta_{t-1}\mathbb{P}(\mathcal{M}_t^c) + \frac{2M_1 x_{\mathsf{max}}b_{\mathsf{max}}}{d^{s^*}} + \frac{2x_{\mathsf{max}}b_{\mathsf{max}}}{d} \\ &\quad + M_3 x_{\mathsf{max}}b_{\mathsf{max}} \exp\{-c_2\kappa^2(\xi, \vartheta, K)t + Ds^* \log K + Ds^* \log(3d/\varepsilon)\} \\ &\quad + M_4 x_{\mathsf{max}}b_{\mathsf{max}} \exp\left\{-c_8 \log(K)t + Ds^* \log K + Ds^* \log(3d/\varepsilon)\right\}, \end{aligned} \tag{27}$$

where M_3, M_4 are large enough universal positive constants and $D = \max\{C, \tilde{C}_1\} = \Theta(\phi_u \vartheta^2 \xi K \log K)$. Thus, if we set

$$T_0 = M_5 \max\left\{\frac{1}{\kappa^2(\xi, \vartheta, K)}, \frac{1}{\log K}\right\}(Ds^* \log K + Ds^* \log(3d/\varepsilon)), \tag{28}$$

for some large universal constant $M_5 > 0$. Thus, we have

$$\mathbb{E}\{\tilde{R}(T)\} \leq \underbrace{2\sum_{t=T_0+1}^{T} \delta_{t-1}\mathbb{P}(\mathcal{M}_t^c)}_{I_\omega} + M_6 x_{\mathsf{max}}b_{\mathsf{max}} \exp\left\{-M_7(Ds^* \log K + Ds^* \log(3d/\varepsilon))\right\} + O(\log T).$$

Recall that

$$\delta_t = M_2 \rho\sigma x_{\mathsf{max}}^2 \xi K \left\{\frac{s^{*2}(\log d + \log t)}{t}\right\}^{1/2}.$$

For $\omega \in [0, 1]$ we have

$$\begin{aligned} I_\omega &\leq 2\sum_{t=T_0+1}^{T} \delta_{t-1}\left(\frac{3\delta_{t-1}}{\Delta_*}\right)^\omega \\ &\asymp \frac{\left\{3M_2\rho\sigma x_{\mathsf{max}}^2 \xi K\right\}^{1+\omega} s^{*1+\omega}}{\Delta_*^\omega} \int_{T_0}^{T} (\log d + \log u)^{\frac{1+\omega}{2}} u^{-\frac{1+\omega}{2}}\, du \\ &\lesssim \begin{cases} \frac{\left[3M_2\rho\sigma x_{\mathsf{max}}^2 \xi K\right]^{1+\omega} s^{*1+\omega} (\log d)^{\frac{1+\omega}{2}} T^{\frac{1-\omega}{2}}}{\Delta_*^\omega}, & \text{for } \omega \in [0, 1), \\ \frac{\left[3M_2\rho\sigma x_{\mathsf{max}}^2 \xi K\right]^2 s^{*2}(\log d + \log T)\log T}{\Delta_*}, & \text{for } \omega = 1. \end{cases} \end{aligned} \tag{29}$$

For $\omega \in (1, \infty)$ we have

$$I_\omega \leq 2\sum_{t=T_0+1}^{T} \delta_{t-1} \min\left\{1, \left(\frac{3\delta_{t-1}}{\Delta_*}\right)^\omega\right\}. \tag{30}$$

Note that

$$\frac{3\delta_{t-1}}{\Delta_*} \leq 1 \Rightarrow t \geq T_1 := \left[3M_2\rho\sigma x_{\mathsf{max}}^2 \xi K\right]^2 \frac{s^{*2} \log d}{\Delta_*^2} + 1.$$

Thus. from Equation (30) we have

$$\begin{aligned} I_\omega &\le 2\sum_{t=T_0+1}^{T_1}\delta_{t-1} + 2\sum_{t=T_1+1}^{T}\delta_{t-1}\left(\frac{3\delta_{t-1}}{\Delta_*}\right)^\omega \\ &\le 2\int_{T_0}^{T_1} M_2\rho\sigma x_{\mathsf{max}}^2\xi K\left\{\frac{s^{*2}(\log d+\log u)}{u}\right\}^{1/2} du + \quad 2\sum_{t=T_1+1}^{T}\delta_{t-1}\left(\frac{3\delta_{t-1}}{\Delta_*}\right)^\omega \\ &\lesssim 4\left[M_2\rho\sigma x_{\mathsf{max}}^2\xi K\right]^2\left\{\frac{s^{*2}(\log d+\log T)}{\Delta_*}\right\} + 2J_\omega, \end{aligned}$$

where $J_\omega := \sum_{t=T_1+1}^{T}\delta_{t-1}\left(\frac{3\delta_{t-1}}{\Delta_*}\right)^\omega$.

Finally, we give bound on the term J_ω:

$$\begin{aligned} J_\omega &= \sum_{t=T_1+1}^{T}\delta_{t-1}\left(\frac{3\delta_{t-1}}{\Delta_*}\right)^\omega \\ &= \sum_{t=2}^{T}\delta_{t-1}\left(\frac{3\delta_{t-1}}{\Delta_*}\right)^\omega \mathbb{1}\{3\delta_{t-1}/\Delta_* \le 1\} \\ &\le \left(\frac{3^{\frac{\omega}{1+\omega}}M_2\rho\sigma x_{\mathsf{max}}^2\xi K}{\Delta_*^{\frac{\omega}{1+\omega}}}\right)^{1+\omega}\int_1^T\{s^{*2}(\log d+\log u)\}^{\frac{1+\omega}{2}}u^{-\frac{1+\omega}{2}}\mathbb{1}\{u\ge T_1\}\,du \\ &\le \left(\frac{3^{\frac{\omega}{1+\omega}}M_2\rho\sigma x_{\mathsf{max}}^2\xi K}{\Delta_*^{\frac{\omega}{1+\omega}}}\right)^{1+\omega}\{s^{*2}(\log d+\log T)\}^{\frac{1+\omega}{2}}\int_{T_1}^{\infty}u^{-\frac{1+\omega}{2}}\,du \\ &= 2\left(\frac{3^{\frac{\omega}{1+\omega}}M_2\rho\sigma x_{\mathsf{max}}^2\xi K}{\Delta_*^{\frac{\omega}{1+\omega}}}\right)^{1+\omega}\{s^{*2}(\log d+\log T)\}^{\frac{1+\omega}{2}}\frac{T_1^{-\frac{\omega-1}{2}}}{\omega-1} \\ &\asymp \left\{\frac{6\left[M_2\rho\sigma x_{\mathsf{max}}^2\xi K\right]^2}{(\omega-1)}\right\}\left(\frac{s^{*2}\log d}{\Delta_*}\right). \end{aligned} \tag{31}$$

Finally, for $\omega=\infty$ it is easy to see that $J_\omega = 0$. Hence, the result follows from combing Equation (22), (29), (30) and (31).

C. Posterior contraction result

We briefly describe the probability space under which we are working. Given β, the bandit environment (along with the specific policy π) gives rise to the chosen contexts X_t and rewards $\mathbf{r}_t$. Here we note that the chosen contexts depend not only on the arm-specific distributions, but also on the sequence of actions taken under π till time t. Let $\mathcal{Q}_t$ denote the joint distribution of $(\beta, X_t, \mathbf{r}_t)$ under $\beta\sim\Pi$ (prior) and $(X_t,\mathbf{r}_t)\mid\beta\sim\text{SLCB}_t(\beta,\pi,\mathcal{P}_\epsilon)$ where the latter indicates the joint distribution of the observed contexts and rewards (till time t) under the SLCB environment with policy π, true parameter β and $\mathcal{P}_\epsilon$ denotes the noise distribution. We work under a likelihood misspecified regime, which we now discuss.

We assume that the true parameter is β^* and the observations $(X_t,\mathbf{r}_t)$ is generated from $\mathcal{Q}_t^* := \text{SLCB}_t(\beta^*,\pi,\mathcal{P}_\epsilon^*)$, where π is the policy given by the TS and $\mathcal{P}_\epsilon^*$ is an arbitrary sub-Gaussian distribution. We denote by $\mathcal{Q}_{t,\mathbf{r}_t}^{*X}$ the conditional distribution of $\mathbf{r}_t$ given X_t arising from the joint $\mathcal{Q}_t^*$ and $\mathbb{E}_{t,\mathbf{r}_t}^X$ to be the expectation under $\mathcal{Q}_{t,\mathbf{r}_t}^{*X}$. Furthermore, we denote by $\mathcal{Q}_{X_t}^*$ the marginal distribution of X_t under the joint $\mathcal{Q}_t^*$ and $\mathbb{E}_{X_t}$ to be the corresponding expectation.

For modelling purpose, we place prior Π on β and model the likelihood as $(X_t,\mathbf{r}_t)\mid\beta\sim\text{SLCB}_t(\beta,\pi,\mathcal{P}_\epsilon)$, where $\mathcal{P}_\epsilon$ is taken to be $\mathsf{N}(0,\sigma^2)$. This gives rise to a joint distribution $\mathcal{Q}_t$, as discussed above. Now, let $\Pi_t^X(\cdot\mid\mathbf{r}_t)$ denote the posterior distribution of β given all others, i.e. it is the conditional measure of β given $X_t,\mathbf{r}_t$ arising from the joint $\mathcal{Q}_t$.

Thus, given a measurable set B, $\Pi_t^X(B|\mathbf{r}_t)$ is a random measure, whose randomness is due to (X_t,r_t). In the following result, we consider $\mathbb{E}_{t,\mathbf{r}_t}^X\Pi_t^X(B|\mathbf{r}_t)$, which is the expectation of the above under $\mathcal{Q}_{t,\mathbf{r}_t}^{*X}$. Thus, this quantity itself is a random

variable, whose randomness is due to X_t. The following result shows that, for B taken as the complement of an appropriate ball around the true β^*, this random variable is small, almost surely $\mathcal{Q}^*_{X_t}$.

Theorem C.1. *Consider the bandit problem in (1) and let Assumption 2.2-2.4 hold. Also, assume that the prior on parameter β is modeled as (2) with*

$$(5/3)\overline{\lambda}_t \le \lambda \le 2\overline{\lambda}_t, \quad \overline{\lambda}_t = x_{\max}\sqrt{2t(\log d + \log t)}$$

Then the following is true:

$$\mathbb{E}^X_{t,\mathbf{r}_t}\left\{\Pi^X_t\left(\|\beta-\beta^*\|_1 \ge Q_4\sigma x_{\max}K\xi(D_*+s^*)\sqrt{\frac{\log d+\log t}{t}}\;\middle|\;\mathbf{r}_t\right)\mathbb{1}_{\mathcal{G}_t\cap\mathcal{T}_t(\lambda_0(\gamma_t))}\right\} \lesssim d^{-s^*},$$

almost sure X_t, where Q_4 is a universal constant and $D_* = D_1 s^* + \frac{D_2 x^2_{\max} s^*}{\phi^2_{\mathsf{comp}}(S^*;X_t)}$ with $D_1 = 1+(40/A_4)$ and $D_2 = 128A_4^{-1}$.

Proof. Without loss of generality we assume that $\sigma = 1$ as the bandit reward model can be viewed as

$$(\mathbf{r}_t/\sigma) = X_t(\beta^*/\sigma) + (\epsilon_t/\sigma).$$

In this case $\overline{\lambda}_t = x_{\max}\sqrt{2t(\log d+\log t)} =: \overline{\lambda}$.

Next, define the event

$$\mathcal{T}_0 := \left\{\max_{j\in[d]}\left|\epsilon_t^\top X_t^{(j)}\right| \le \bar{\lambda}\right\}.$$

By Lemma B.12 and condition (21), it follows that fro any measurable set $\mathcal{B}\subset\mathbb{R}^d$,

$$\mathbb{E}_{t,\mathbf{r}_t}\Pi^X_t(\mathcal{B}\mid\mathbf{r}_t) \le \left[\mathbb{E}_{t,\mathbf{r}_t}\left\{\Pi^X_t(\mathcal{B}\mid\mathbf{r}_t)\mathbb{1}_{\mathcal{T}_0}\right\}\right]^{1/2} + \frac{2}{t}.$$

Recall that the errors ϵ_t is modeled as isotropic standard Gaussian independent of the features. Thus, conditioned on the matrix X_t, model likelihood ration takes the following form:

$$\mathcal{L}_{t,\beta,\beta^*}(\mathbf{r}_t) := \exp\left\{-\frac{\|X_t\beta - X_t\beta^*\|_2^2}{2} + (\mathbf{r}_t - X_t\beta^*)^\top(X_t\beta - X_t\beta^*)\right\}.$$

Then by Lemma 2 of (Castillo et al., 2015) it follows that

$$\int \mathcal{L}_{t,\beta,\beta^*}(\mathbf{r}_t)\; d\Pi(\beta) \ge \frac{\pi_d(s^*)}{p^{2s^*}}e^{-\lambda\|\beta^*\|_1}e^{-1},$$

where Π is given by (2). The only change that is needed in their proof to run the argument in our case is the following upper bound:

$$\|X\beta\|_2 \le \|\beta\|_1\|X\| \le c_9\vartheta^2\phi_u\log K/(1-3\varepsilon).$$

The last inequality follows from the fact that we are on the event $\mathcal{G}_{t,1}$ by assumption. The rest of the proof follows from the fact that $\lambda\in(5\overline{\lambda}/3, 2\overline{\lambda})$.

Thus by Bayes's formula it follows that

$$\begin{aligned}\Pi^X_t(\mathcal{B}\mid\mathbf{r}_t) &= \frac{\int_{\mathcal{B}}\mathcal{L}_{t,\beta,\beta^*}(\mathbf{r}_t)\;d\Pi(\beta)}{\int\mathcal{L}_{t,\beta,\beta^*}(\mathbf{r}_t)\;d\Pi(\beta)}\\ &\le \frac{ed^{2s^*}}{\pi_d(s^*)}e^{\lambda\|\beta^*\|_1}\int_{\mathcal{B}}\exp\left\{-\frac{\|X_t\beta - X_t\beta^*\|_2^2}{2} + (\mathbf{r}_t - X_t\beta^*)^\top(X_t\beta - X_t\beta^*)\right\}\;d\Pi(\beta).\end{aligned} \tag{32}$$

Using Holder's inequality, we see that on $\mathcal{T}_0$,

$$\begin{aligned}(\mathbf{r}_t - X_t\beta^*)^\top X_t(\beta-\beta^*) &= \epsilon_t^\top X_t(\beta-\beta^*)\\ &\le \left\|\epsilon_t^\top X_t\right\|_\infty\|\beta-\beta^*\|_1\\ &\le \bar{\lambda}\|\beta-\beta^*\|_1.\end{aligned} \tag{33}$$

Therefore, on the event $\mathcal{T}_0$, the expected value under $\mathbb{E}_{\beta^*}$ of the integrand on the right hand side of (32) is bounded above by

$$e^{-(1/2)\|X_t(\beta-\beta^*)\|_2^2+\overline{\lambda}\|\beta-\beta^*\|_1}.$$

Thus, we have

$$\Pi_t^X(\mathcal{B} \mid \mathbf{r}_t)\mathbb{1}_{\mathcal{T}_0} \leq \frac{ep^{2s^*}}{\pi_d(s^*)}\int_{\mathcal{B}} e^{\lambda\|\beta\|_1-(1/2)\|X_t(\beta-\beta^*)\|_2^2+\overline{\lambda}\|\beta-\beta^*\|_1}\, d\Pi(\beta). \tag{34}$$

Now, by triangle inequality,

$$\begin{aligned}
\lambda\|\beta^*\|_1+\overline{\lambda}\|\beta-\beta^*\|_1 &= \lambda\|\beta^*_{S^*}\|_1+\overline{\lambda}\|\beta-\beta^*\|_1 \\
&\leq \lambda\|\beta^*_{S^*}-\beta_{S^*}\|_1+\lambda\|\beta_{S^*}\|_1+\overline{\lambda}\|\beta_{S^*}-\beta^*_{S^*}\|_1+\overline{\lambda}\|\beta_{S^{*c}}\| \\
&= \lambda\|\beta_{S^*}\|_1+\overline{\lambda}\|\beta_{S^{*c}}\|_1+(\lambda+\overline{\lambda})\|\beta_{S^*}-\beta^*_{S^*}\|_1 \\
&= \lambda\|\beta\|_1+(\overline{\lambda}-\lambda)\|\beta_{S^{*c}}\|_1+(\lambda+\overline{\lambda})\|\beta_{S^*}-\beta^*_{S^*}\|_1.
\end{aligned} \tag{35}$$

Case 1: Suppose $7\|\beta_{S^*}-\beta^*_{S^*}\|_1 \leq \|\beta_{S^{*c}}\|_1$. Then the following holds:

$$\begin{aligned}
(\lambda+\overline{\lambda})\|\beta_{S^*}-\beta^*_{S^*}\|_1 &= (\overline{\lambda}-3\lambda/4)\|\beta_{S^*}-\beta^*_{S^*}\|_1+(7\lambda/4)\|\beta_{S^*}-\beta^*_{S^*}\|_1 \\
&\leq (\overline{\lambda}-3\lambda/4)\|\beta_{S^*}-\beta^*_{S^*}\|_1+(\lambda/4)\|\beta_{S^{*c}}\|_1.
\end{aligned}$$

Using the above inequality in (35) we get

$$\begin{aligned}
\lambda\|\beta^*\|_1+\overline{\lambda}\|\beta-\beta^*\|_1 &\leq \lambda\|\beta\|_1+(\overline{\lambda}-3\lambda/4)\|\beta_{S^{*c}}\|_1+(\overline{\lambda}-3\lambda/4)\|\beta_{S^*}-\beta^*_{S^*}\|_1 \\
&= \lambda\|\beta\|_1+(\overline{\lambda}-3\lambda/4)\|\beta-\beta^*\|_1
\end{aligned} \tag{36}$$

Case 2: Now assume $7\|\beta_{S^*}-\beta^*_{S^*}\|_1 > \|\beta_{S^{*c}}\|_1$. We again focus on the inequality (35), i.e.,

$$\begin{aligned}
\lambda\|\beta^*\|_1+\overline{\lambda}\|\beta-\beta^*\|_1 &\leq \lambda\|\beta\|_1+(\overline{\lambda}-\lambda)\|\beta_{S^{*c}}\|_1+(\lambda+\overline{\lambda})\|\beta_{S^*}-\beta^*_{S^*}\|_1 \\
&= \lambda\|\beta\|_1+(\overline{\lambda}-\lambda)\|\beta_{S^{*c}}\|_1+(\overline{\lambda}-\lambda)\|\beta_{S^*}-\beta^*_{S^*}\|_1 \\
&\quad +2\lambda\|\beta_{S^*}-\beta^*_{S^*}\|_1 \\
&= \lambda\|\beta\|_1+(\overline{\lambda}-\lambda)\|\beta-\beta^*\|_1+2\lambda\|\beta_{S^*}-\beta^*_{S^*}\|_1 \\
&\leq \lambda\|\beta\|_1+(\overline{\lambda}-3\lambda/4)\|\beta-\beta^*\|_1+2\lambda\|\beta_{S^*}-\beta^*_{S^*}\|_1.
\end{aligned} \tag{37}$$

Finally, by compatibility condition and Young's inequality we get

$$2\lambda\|\beta_{S^*}-\beta^*_{S^*}\| \leq 2\lambda\frac{\|X_t(\beta-\beta^*)\|_2 s^{*1/2}}{t^{1/2}\phi_{\mathsf{comp}}(S^*;X_t)} \leq \frac{\|X_t(\beta-\beta^*)\|_2^2}{2}+\frac{2s^*\lambda^2}{t\phi^2_{\mathsf{comp}}(S^*;X_t)}.$$

Using the above inequality in (37) we get

$$\lambda\|\beta^*\|_1+\overline{\lambda}\|\beta-\beta^*\|_1 \leq \lambda\|\beta\|_1+(\overline{\lambda}-3\lambda/4)\|\beta-\beta^*\|_1+\frac{\|X_t(\beta-\beta^*)\|_2^2}{2}+\frac{2s^*\lambda^2}{t\phi^2_{\mathsf{comp}}(S^*;X_t)}. \tag{38}$$

Thus combining (36) and (38) we can conclude

$$\lambda\|\beta^*\|_1+\overline{\lambda}\|\beta-\beta^*\|_1 \leq \lambda\|\beta\|_1+(\overline{\lambda}-3\lambda/4)\|\beta-\beta^*\|_1+\frac{\|X_t(\beta-\beta^*)\|_2^2}{2}+\frac{2s^*\lambda^2}{t\phi^2_{\mathsf{comp}}(S^*;X_t)}. \tag{39}$$

Using the above result and recalling that $5\overline{\lambda}/3 \leq \lambda \leq 2\overline{\lambda}$, we see that the right hand side of (34) is bounded by

$$\Pi_t^X(\mathcal{B} \mid \mathbf{r}_t)\mathbb{1}_{\mathcal{T}_0} \leq \frac{ed^{2s^*}}{\pi_d(s^*)}e^{\frac{2s^*\lambda^2}{t\phi^2_{\mathsf{comp}}(S^*;X_t)}}\int_{\mathcal{B}} e^{\lambda\|\beta\|_1-(\lambda/4)\|\beta-\beta^*\|_1}\, d\Pi(\beta)$$

Controlling sparsity: For the set $\mathcal{B} = \{\beta : |S_\beta| > L\}$ and $L \geq s^*$, the above integral can be bounded by

$$\begin{aligned}
&\sum_{S:s=|S|>L} \frac{\pi_d(s)}{\binom{d}{s}} \left(\frac{\lambda}{2}\right)^s \int e^{-(\lambda/4)\|\beta_S - \beta_S^*\|} \, d\beta_S \\
&\leq \sum_{s=L+1}^{\infty} \pi_d(s) 4^s \\
&\leq \pi_d(s^*) 4^{s^*} \left(\frac{4A_2}{d^{A_4}}\right)^{L+1-s^*} \sum_{j=0}^{\infty} \left(\frac{4A_2}{d^{A_4}}\right)^j
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\mathbb{E}^X_{t,\mathbf{r}_t}\left\{\Pi^X_t(\mathcal{B} \mid \mathbf{r}_t)\mathbb{1}_{\mathcal{T}_0}\right\} \\
&\lesssim \exp\left\{4s^* \log d + \frac{2s^*\lambda^2}{t\phi^2_{\mathsf{comp}}(S^*; X_t)} + s^* \log 4 - (A_4/4)(L+1-s^*)\log d\right\} \\
&\leq \exp\left\{5s^* \log d + \frac{4s^*\lambda\overline{\lambda}}{t\phi^2_{\mathsf{comp}}(S^*; X_t)} - (A_4/4)(L+1-s^*)\log d\right\}
\end{aligned}$$

Now recall that $\overline{\lambda}^2 = 2tx^2_{\mathsf{max}}(\log d + \log t) \leq 4tx^2_{\mathsf{max}} \log d$. Using this inequality in the above display we have

$$\mathbb{E}^X_{t,\mathbf{r}_t}\left\{\Pi^X_t(\mathcal{B} \mid \mathbf{r}_t)\mathbb{1}_{\mathcal{T}_0}\right\} \lesssim \exp\left\{5s^* \log d + \frac{16s^*(\lambda/\overline{\lambda})x^2_{\mathsf{max}} \log d}{\phi^2_{\mathsf{comp}}(S^*; X_t)} - (A_4/4)(L+1-s^*)\log d\right\}.$$

Thus, setting $L \geq 40s^*/A_4 + s^* + \frac{64A_4^{-1}s^*x^2_{\mathsf{max}}}{\phi^2_{\mathsf{comp}}(S^*;X_t)}(\lambda/\overline{\lambda})$ then there exists a universal constant $Q_1 > 0$ such that

$$\mathbb{E}^X_{t,\mathbf{r}_t}\left\{\Pi^X_t(\mathcal{B} \mid \mathbf{r}_t)\mathbb{1}_{\mathcal{T}_0}\right\} \leq Q_1 d^{-s^*}.$$

Control on prediction: Recall that $\lambda/\overline{\lambda} \leq 2$. Using this and the result in the previous part, we can conclude that the posterior distribution is asymptotically supported on the even $\mathcal{B}_1 = \{\beta : |S_\beta| \leq D_*\}$, where $D_* = D_1 s^* + \frac{D_2 x^2_{\mathsf{max}} s^*}{\phi^2_{\mathsf{comp}}(S^*;X_t)}$ where $D_1 = 1 + (40/A_4)$ and $D_2 = 128A_4^{-1}$. By combining (32), (33) and the inequality $\lambda\|\beta^*\|_1 \leq 2\overline{\lambda}\|\beta - \beta^*\|_1 + \lambda\|\beta\|_1$ we can conclude that any Borel set $\mathcal{B}$,

$$\Pi^X_t(\mathcal{B} \mid \mathbf{r}_t)\mathbb{1}_{\mathcal{T}_0} \leq \frac{ed^{2s^*}}{\pi_d(s^*)} \int_{\mathcal{B}} \exp\left\{-\frac{\|X_t\beta - X_t\beta^*\|_2^2}{2} + 3\overline{\lambda}\|\beta - \beta^*\|_1 + \lambda\|\beta\|_1\right\} d\Pi(\beta).$$

By the definition in (23) we have,

$$\begin{aligned}
(4-1)\overline{\lambda}\|\beta - \beta^*\|_1 &\leq \frac{4\overline{\lambda}\|X_t(\beta - \beta^*)\|_2 |S_{\beta-\beta^*}|^{1/2}}{t^{1/2}\overline{\phi}_t(|S_{\beta-\beta^*}|)} - \overline{\lambda}\|\beta - \beta^*\|_1 \\
&\leq \frac{1}{4}\|X_t(\beta - \beta^*)\|_2^2 + \frac{16\overline{\lambda}^2 |S_{\beta-\beta^*}|}{t\overline{\phi}_t(|S_{\beta-\beta^*}|)^2} - \overline{\lambda}\|\beta - \beta^*\|_1.
\end{aligned} \tag{40}$$

Since $|S_{\beta-\beta^*}| \leq |S_\beta| + s^* \leq D_* + s^*$ on the event $\mathcal{B}_1$, it follows that

$$\begin{aligned}
\Pi^X_t(\mathcal{B} \mid \mathbf{r}_t)\mathbb{1}_{\mathcal{T}_0} \leq \frac{ed^{2s^*}}{\pi_d(s^*)} &e^{16\overline{\lambda}^2(D_*+s^*)/(t\overline{\psi}_t(S^*)^2)} \\
&\times \int_{\mathcal{B}} e^{-(1/4)\|X_t(\beta-\beta^*)\|_2^2 - \overline{\lambda}\|\beta-\beta^*\|_1 + \lambda\|\beta\|_1} \, d\Pi(\beta).
\end{aligned} \tag{41}$$

Now we set $\mathcal{B} = \mathcal{B}_2 := \{\beta \in \mathcal{B}_1 : \|X_t(\beta - \beta^*)\|_2 > L\}$, where L will be chosen shortly. Recall that $\pi_d(s^*) \geq (A_1 p^{-A_3})^{s^*} \pi_p(0)$. It follows that for set $\mathcal{B}$, the right hand side of (41) is upper bounded by

$$\begin{aligned}
&\frac{ed^{2s^*}}{\pi_d(s^*)} e^{16\overline{\lambda}^2(D_*+s^*)/(t\overline{\psi}_t(S^*)^2)} e^{-(1/4)L^2} \int e^{-\overline{\lambda}\|\beta-\beta^*\|_1+\lambda\|\beta\|_1}\, d\Pi(\beta) \\
&\lesssim d^{(2+A_3)s^*} A_1^{-s^*} e^{16\overline{\lambda}^2(D_*+s^*)/(t\overline{\psi}_t(S^*)^2)} e^{-(1/4)L^2} \underbrace{\sum_{s=0}^{d} \pi_d(s^*) 2^s}_{O(1)}.
\end{aligned}$$

Hence by a calculation similar to previous discussion yields that for

$$\frac{1}{4}L^2 = (3+A_3)s^* \log d + \frac{16\overline{\lambda}^2(D_*+s^*)}{t\overline{\psi}_t(S^*)^2} \leq Q_2 x_{\mathsf{max}}^2 (D_*+s^*) \frac{\log d + \log t}{\overline{\psi}_t(S^*)^2} =: L_*^2,$$

where $Q_2 > 0$ is sufficiently large universal constant, then we have

$$\mathbb{E}_{t,\mathbf{r}_t}^X \left\{ \Pi_t^X(\mathcal{B}_2 \mid \mathbf{r}_t) \mathbb{1}_{\mathcal{T}_0} \right\} \leq \frac{1}{d^{s^*}}.$$

Control on estimation: Similar to (40) we have

$$\begin{aligned}
\overline{\lambda}\,\|\beta - \beta^*\|_1 &\leq \frac{\overline{\lambda}\,\|X_t(\beta-\beta^*)\|_2\,|S_{\beta-\beta^*}|^{1/2}}{t^{1/2}\overline{\phi}_t(|S_{\beta-\beta^*}|)} \\
&\leq \frac{1}{4}\,\|X_t(\beta-\beta^*)\|_2^2 + \frac{\overline{\lambda}^2\,|S_{\beta-\beta^*}|}{t\overline{\phi}_t(|S_{\beta-\beta^*}|)^2}.
\end{aligned}$$

On the event $\mathcal{B}_2$, we thus have

$$\overline{\lambda}\,\|\beta - \beta^*\|_1 \leq Q_3 x_{\mathsf{max}}^2 (D_*+s^*) \frac{\log d + \log t}{\overline{\psi}_t(S^*)^2}.$$

Finally on the event $\mathcal{B}_2 \cap \mathcal{G}_t$ we have $\overline{\lambda} = x_{\mathsf{max}}\sqrt{2t(\log d + \log t)}$ and $\overline{\psi}_t(S^*)^2 \gtrsim (K\xi)^{-1}$ and it follows that

$$\|\beta - \beta^*\|_1 \leq Q_4 K\xi x_{\mathsf{max}} (D_*+s^*) \sqrt{\frac{\log d + \log t}{t}}.$$

□

D. Technical lemmas

D.1. Proof of Lemma B.6

As A is symmetric positive definite matrix, by Cholesky decomposition there exists as lower triangular matrix L such that $A = LL^\top$. Let $v \in \mathbb{S}_0^{d-1}(s)$. Then there exists a index set J of size s, such that $\mathsf{supp}(v) \subseteq J$. Hence we have $v \in E_J$. Now consider the net $\mathcal{N}_{\varepsilon,J}$ and let u be the nearest point to v in $\mathcal{N}_{\varepsilon,J}$. Thus we have $\|v-u\|_2 \leq \varepsilon < 1$ and $\|v-u\|_0 \leq s$. Then we have the following:

$$\begin{aligned}
v^\top A v &= (v-u)^\top A(v-u) + 2(v-u)^\top Au + u^\top Au \\
&\geq u^\top Au - 3\varepsilon\phi_{\max}(s;A).
\end{aligned} \tag{42}$$

The second inequality follows from the following facts:

$$\begin{aligned}
\left|(v-u)^\top A(v-u)\right| &\leq \|v-u\|_2^2\,\phi_{\max}(s;A) \leq \varepsilon\phi_{\max}(s;A), \\
\left|(v-u)^\top Au\right| &= \left|(v-u)^\top LL^\top u\right| \\
&\leq \sqrt{(v-u)^\top LL^\top (v-u)}\sqrt{u^\top LL^\top u} \\
&= \sqrt{(v-u)^\top A(v-u)}\sqrt{u^\top Au} \\
&\leq \varepsilon\phi_{\max}(s;A)
\end{aligned}$$

Then the result follows from taking infimum over u and v in both sides.

D.2. Proof of Lemma B.8

The lower bound result is trivial. Hence, we focus on the upper bound part. As A is symmetric positive definite matrix, by Cholesky decomposition there exists as lower triangular matrix L such that $A = LL^\top$. Let $v \in \mathbb{S}_0^{d-1}(s)$. Then there exists a index set J of size s, such that $\mathsf{supp}(v) \subseteq J$. Hence we have $v \in E_J$. Now consider the net $\mathcal{N}_{\varepsilon,J}$ and let u be the nearest point to v in $\mathcal{N}_{\varepsilon,J}$. Thus we have $\|v-u\|_2 \leq \varepsilon < 1/3$ and $\|v-u\|_0 \leq s$. By a similar argument as in the proof of Lemma B.6, we can conclude that

$$v^\top A v \leq 3\varepsilon \phi_{\max}(s; A) + \max_{u \in \mathcal{N}_\varepsilon} u^\top A u.$$

Thus by taking supremum over v on the left-hand side of the above display we get

$$(1-3\varepsilon)\phi_{\max}(s; A) \leq \max_{u \in \mathcal{N}_\varepsilon} u^\top A u \iff \phi_{\max}(s; A) \leq \frac{1}{1-3\varepsilon} \max_{u \in \mathcal{N}_\varepsilon} u^\top A u.$$

E. Pseudo code of VBTS and other tables

Algorithm 2 Variational Bayes Thompson Sampling

Set $\mathcal{H}_0 = \emptyset$.
for $t = 1, \cdots, T$ **do**
  if $t \leq 1$ **then**
    Choose action a_t uniformly over $[K]$.
  end if
  if $t > 1$ **then**
    Compute the VB posterior $\widetilde{\Pi}_{t-1}$ from $\Pi(\cdot \mid \mathcal{H}_{t-1})$ using CAVI.
    Generate sample $\tilde{\beta}_t \sim \widetilde{\Pi}_{t-1}$.
    Play arm: $a_t = \arg\max_{i \in [K]} x_i(t)^\top \tilde{\beta}_t$.
  end if
end for
Observe reward $r(t)$.
Update $\mathcal{H}_t \leftarrow \mathcal{H}_{t-1} \cup \{(a_t, r_{a_t}(t), x_{a_t}(t))\}$.

Table 4. Time comparison among the competing algorithms.

Type	Algorithm	Mean time of execution (seconds)	
		Equicorrelated	Auto-regressive
Non-TS	LinUCB	15.41	16.30
	DR Lasso	3.12	3.13
	Lasso-L1	3.57	3.59
	ESTC	1.01	1.05
TS	LinTS	1344.39	1346.46
	BLasso TS	1511.68	1455.53
	VBTS	**29.33**	**27.65**