ONLINE THOMPSON SAMPLING FOR GENERALIZED LINEAR BANDITS

Anonymous authors
Paper under double-blind review

ABSTRACT

Applying Thompson sampling (TS) to generalized linear bandits (GLBs) is computationally challenging because exact posterior updates are generally intractable. Existing GLB algorithms with strong regret guarantees typically rely on batch estimators built from accumulated data, while computationally efficient one-pass methods often require problem-dependent quantities, such as an upper bound on the context norm and a radius of a ball containing the true parameter, as explicit inputs to their implementation. We propose OnlineTS, an online TS algorithm for GLBs based on Bayesian one-pass online learning. Our method uses a Gaussian approximation induced by a local quadratic expansion of the log-likelihood, leading to a recursive posterior update that depends only on the current estimator and the current observation. We show that this Bayesian online one-pass estimator admits a high probability concentration bound in the bandit setting, which yields a frequentist regret bound of $\widetilde{O}(\sqrt{T})$ for OnlineTS over horizon T. Thus, OnlineTS combines strong statistical performance with computational efficiency without requiring such problem-dependent quantities as inputs to its implementation. Through numerical experiments, we demonstrate that OnlineTS achieves regret competitive with the existing GLB methods at substantially lower computational cost.

1 INTRODUCTION

Contextual bandits provide a fundamental framework for sequential decision-making under partial feedback and arise in a wide range of applications, including recommendation, advertising, and clinical trials (Woodroofe, 1979; Li et al., 2010; Tewari & Murphy, 2017). At each round t, the learner observes contextual information $\mathcal{X}_t$ for the available actions, selects one action a_t, and receives only the reward of the chosen action r_t. While linear contextual bandits have been studied extensively (Dani et al., 2008; Rusmevichientong & Tsitsiklis, 2010; Chu et al., 2011; Abbasi-Yadkori et al., 2011; Agrawal & Goyal, 2013), many practical problems involve nonlinear reward structures. Generalized linear bandits (GLBs) provide the natural extension of the linear model to such settings and accommodate a broad class of reward models, including Bernoulli and Poisson observations (Filippi et al., 2010; Abeille & Lazaric, 2017; Li et al., 2017; Lee et al., 2024). Beyond their direct practical relevance, GLBs also serve as fundamental building blocks for other sequential decision-making problems, including dynamic pricing and reinforcement learning with function approximation (Wang et al., 2021; Chen et al., 2022; Kim et al., 2025).

In many sequential decision-making problems, data arrive continuously and models are updated online (Hoi et al., 2021; Block et al., 2023). In such settings, storing the full history may be impractical, and recomputing a batch estimator after each observation can become increasingly expensive as t increases (Jun et al., 2017; Choi et al., 2025; Zhang et al., 2025; Lee et al., 2026a;b). This motivates the one-pass setting, in which each update uses only the current estimator and the current observation without revisiting past data.

Existing GLB algorithms with strong regret guarantees are typically based on batch estimators constructed from the accumulated data (Filippi et al., 2010; Faury et al., 2020; Jun et al., 2021; Lee et al., 2024). As data accumulate, these methods generally require increasing time and memory for model updates. By contrast, one-pass update methods are computationally attractive, since each update uses only the current estimator and the current observation (Jun et al., 2017; Zhang et al., 2025). However, existing one-pass methods often depend on problem-dependent theoretical quantities

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

that are difficult to specify in practice. Typical examples include upper bounds on the context norm and the true parameter norm, such as constants L and R^* in conditions of the form $\|x_t\|_2 \le L$ and $\|\theta^*\|_2 \le R^*$.

A natural exploration strategy in bandit problems is Thompson sampling (TS), which samples a parameter from the current posterior distribution and then selects an action that is optimal under the sampled parameter. For GLBs, however, exact posterior updates are generally intractable. Existing TS-based and other randomized exploration methods therefore typically rely on Laplace approximations, randomized perturbations, or other estimators built from the accumulated data (Abeille & Lazaric, 2017; Kveton et al., 2020; Kim et al., 2023; Janz et al., 2024). As a result, they do not naturally yield a one-pass procedure in which the posterior approximation itself is updated online using only the current estimator and the current observation.

In this paper, we propose *OnlineTS*, a novel online Thompson sampling algorithm for GLBs based on Bayesian one-pass online learning. Our starting point is Bayesian one-pass online learning framework, in which an intractable posterior update is replaced by a Gaussian approximation induced by a local quadratic expansion of the log-likelihood (Lee et al., 2026b). This yields a simple recursive Gaussian update that uses only the current posterior approximation and the current observation. We adapt this idea to the bandit setting and use the resulting approximated posterior for Thompson sampling. Since the local quadratic approximation is reliable only in a suitable neighborhood of θ^*, we introduce a warm-start phase based on a batch Laplace approximation and switch to the one-pass recursion only after a phase transition time t_0. The resulting procedure is a hybrid method consisting of an initial stabilization phase followed by a fully online Bayesian update that never revisits past data.

Our main theoretical contribution is to establish a high probability concentration result for the Bayesian online one-pass estimator in the bandit setting, where the selected contexts depend on past observations. Building on this concentration result, we prove that, under the frequentist regret criterion with a fixed unknown parameter θ^*, OnlineTS achieves a regret upper bound of $\widetilde{O}(d\sqrt{T})$ over the horizon T, where d is the dimension of the context vectors. From a practical perspective, OnlineTS is computationally efficient and can be implemented without explicitly specifying problem-dependent theoretical quantities that are typically unknown in practice. Its unique tuning parameter, the phase transition time t_0, is straightforward to set: while larger values are theoretically preferable, guarantees hold as long as t_0 is sufficiently large. Furthermore, our empirical results demonstrate that the algorithm already achieves strong performance even with relatively small values of t_0.

The remainder of the paper is organized as follows. In the following subsections, we introduce our notation and briefly summarize related work. In Section 2, we introduce the GLB model and the Bayesian online learning framework underlying our approach. Section 3 presents the Bayesian online one-pass estimator and the OnlineTS algorithm. Section 4 develops the concentration and regret analyses. Section 5 reports numerical experiments, including comparisons with existing GLB methods and a sensitivity study for the phase transition time t_0.

1.1 Notations

Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and for $n \in \mathbb{N}$, let $[n] = \{1, \dots, n\}$. For $z \in \mathbb{R}$, let $\lceil z \rceil$ denote the smallest integer greater than or equal to z. For two positive sequences (a_n) and (b_n), $a_n \lesssim b_n$ (or $a_n = O(b_n)$) means that $a_n \le C b_n$ for some constant $C \in (0, \infty)$, and we write $a_n = \Omega(b_n)$ if $a_n \gtrsim b_n$. The notation $\widetilde{O}(\cdot)$ denotes the corresponding bound that ignores logarithmic factors. Also, $a_n \asymp b_n$ indicates that $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For $1 \le p \le \infty$, $\|\cdot\|_p$ denotes the ℓ_p-norm on $\mathbb{R}^d$. Let $\mathcal{S}_+^d$ and $\mathcal{S}_{++}^d$ denote the sets of all $d \times d$ symmetric positive semidefinite and symmetric positive definite matrices, respectively, and let $\mathbf{I}_d \in \mathbb{R}^{d \times d}$ denote the identity matrix. For $\mathbf{A} \in \mathcal{S}_+^d$ and $v \in \mathbb{R}^d$, define $\|v\|_{\mathbf{A}}^2 = v^\top \mathbf{A} v$. For a symmetric matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$, let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote its smallest and largest eigenvalues, respectively. We write $C = C(\cdot)$ to denote a positive constant that may depend only on the displayed arguments.

1.2 Related works

Generalized linear bandit algorithms Generalized linear bandits were introduced by Filippi et al. (2010), who proposed a UCB-based algorithm and established finite-time regret guarantees. Subsequent work obtained sharper confidence bounds and improved regret guarantees, especially

Table 1: Comparison of representative GLB algorithms. Here, $O(1)$ means constant in the horizon T, possibly depending on the dimension d. The last column indicates whether the algorithm can be implemented without explicitly specifying problem-dependent theoretical quantities. † indicates the average per-round cost over t rounds.

Method	Regret in T	Time per round	Memory	Directly implementable
OFUGLB (Lee et al., 2024)	$\widetilde{O}(\sqrt{T})$	$O(t)$	$O(t)$	$\times$
RS-GLinCB (Sawarni et al., 2024)	$\widetilde{O}(\sqrt{T})$	$O((\log t)^2)^\dagger$	$O(t)$	$\times$
GLOC (Jun et al., 2017)	$\widetilde{O}(\sqrt{T})$	$O(1)$	$O(1)$	$\times$
GLB-OMD (Zhang et al., 2025)	$\widetilde{O}(\sqrt{T})$	$O(1)$	$O(1)$	$\times$
OnlineTS (ours)	$\widetilde{O}(\sqrt{T})$	$O(1)$	$O(1)$	$\bigcirc$

for logistic bandits and self-concordant GLBs Li et al. (2017); Faury et al. (2020); Jun et al. (2021); Abeille et al. (2021); Faury et al. (2022); Lee et al. (2024). These results show that GLB algorithms can achieve strong regret guarantees under suitable assumptions. However, the corresponding algorithms are largely built on batch estimators formed from the accumulated data, and thus are not naturally suited to a strict one-pass setting.

Thompson sampling and randomized exploration for GLBs Thompson sampling is a widely used randomized exploration strategy in bandits and has been studied extensively Thompson (1933); Chapelle & Li (2011); Agrawal & Goyal (2012; 2017); Russo et al. (2018). For GLBs, however, exact posterior sampling is generally intractable, and existing TS-based methods therefore rely on various approximations or surrogate estimators. Abeille & Lazaric (2017) analyzed TS for generalized linear models, while Kveton et al. (2020) and Janz et al. (2024) studied randomized exploration methods based on Laplace approximations and perturbed-history estimators. Kim et al. (2023) proposed a Thompson sampling algorithm based on a double doubly robust estimator that uses contexts from all arms. These methods provide useful randomized exploration schemes for GLBs, but they still rely on batch estimators and cannot be implemented in a one-pass manner, where one-pass refers to both computational and memory efficiency.

Computationally efficient GLB algorithms Several works have aimed to reduce the computational burden of GLB algorithms by developing one-pass or memory-efficient update procedures. The works most closely related to ours are Jun et al. (2017) and Zhang et al. (2025), both of which propose one-pass algorithms with constant time and memory costs per round. In particular, GLOC Jun et al. (2017) achieves efficient online computation with a $\widetilde{O}(\sqrt{T})$ regret guarantee in the horizon T, while GLB-OMD Zhang et al. (2025) attains a nearly optimal regret bound. However, both methods require a known upper bound on the true parameter norm for implementation, and GLOC additionally requires a curvature lower bound. These problem-dependent theoretical quantities are typically unknown in practice and may substantially affect empirical performance when misspecified. A different direction was taken by Sawarni et al. (2024), who proposed a rare-update strategy that updates the policy only $O(\log^2 T)$ times while achieving $\widetilde{O}(\sqrt{T})$ regret. However, since the method remains MLE-based, its implementation still requires $O(t)$ memory at round t. Our work is motivated by the goal of combining one-pass online updates with strong regret guarantees without requiring problem-dependent theoretical quantities as explicit inputs to the implementation. A summary of these comparisons is provided in Table 1.

2 Preliminaries

2.1 Generalized linear bandits

We consider a K-armed generalized linear bandit problem. At each round $t \in [T]$, the learner observes a collection of context vectors $\mathcal{X}_t = \{x_{t,a} \in \mathbb{R}^d : a \in [K]\}$. After observing $\mathcal{X}_t$ and the past history, the learner selects an action $a_t \in [K]$ and then receives a reward $r_t \in \mathbb{R}$. We assume that the reward distribution follows a generalized linear model (GLM; McCullagh & Nelder (1989)) with an unknown true parameter $\theta^* \in \Theta \subseteq \mathbb{R}^d$. For each $\theta \in \Theta$ and $t \in \mathbb{N}$, let $\mathbf{D}_{1:t} = (\mathcal{X}_1, a_1, r_1, \ldots, \mathcal{X}_t, a_t, r_t)$

denote the observed data up to round t. We write $\mathbb{P}_\theta^{(t)}$ and $\mathbb{E}_\theta^{(t)}$ for the probability measure and expectation induced by the reward model and the bandit policy under parameter θ, and omit the superscript when t is clear. In particular, we assume that the observed data $\mathbf{D}_{1:t}$ are generated under the true probability measure $\mathbb{P}_{\theta^*}^{(t)}$. Under parameter $\theta \in \Theta$, conditional on $\mathbf{D}_{1:t-1}$, $\mathcal{X}_t$, and a_t, the reward r_t has density with respect to ν

$$p_{t,\theta}(r_t \mid \mathbf{D}_{1:t-1}, \mathcal{X}_t, a_t) = \exp\left(\frac{r_t x_{t,a_t}^\top \theta - m(x_{t,a_t}^\top \theta)}{g(\tau)} + h(r_t, \tau)\right), \tag{2.1}$$

where ν is a known base measure, $\tau > 0$ is a known dispersion parameter, and $m(\cdot)$, $g(\cdot)$, and $h(\cdot, \cdot)$ are known functions. In particular, the conditional distribution of r_t depends on the past only through the currently selected context x_{t,a_t}. Under (2.1), the conditional mean reward is

$$\mathbb{E}_\theta[r_t \mid \mathbf{D}_{1:t-1}, \mathcal{X}_t, a_t] = m'(x_{t,a_t}^\top \theta) =: \mu(x_{t,a_t}^\top \theta),$$

where $\mu(\cdot)$ denotes the inverse link function. For the regular GLM considered in this paper, we have $\mu'(z) > 0$ for all $z \in \mathbb{R}$. Let $a_t^* \in \operatorname{argmax}_{a \in [K]} \mu(x_{t,a}^\top \theta^*)$ be an optimal action at round t. Our goal is to minimize the cumulative regret over T rounds, defined as

$$\mathrm{Reg}(T) = \sum_{t=1}^{T} \left\{\mu(x_{t,a_t^*}^\top \theta^*) - \mu(x_{t,a_t}^\top \theta^*)\right\}.$$

For each $\theta \in \Theta$, we define the one-step negative log-likelihood function under (2.1) by

$$\ell_t(\theta) = g(\tau)^{-1}\left\{m(x_{t,a_t}^\top \theta) - r_t \cdot x_{t,a_t}^\top \theta\right\}, \tag{2.2}$$

where we omit the additive term independent of θ. Whenever $\ell_t(\theta)$ is twice differentiable, we write

$$\nabla \ell_t(\theta) = g(\tau)^{-1}\big(\mu(x_{t,a_t}^\top \theta) - r_t\big) x_{t,a_t}, \qquad \nabla^2 \ell_t(\theta) = g(\tau)^{-1} \mu'(x_{t,a_t}^\top \theta) x_{t,a_t} x_{t,a_t}^\top. \tag{2.3}$$

Fix $R > 0$, and define the local neighborhood of θ^* by $\Theta_R = \{\theta \in \Theta : \|\theta - \theta^*\|_2 \le R\}$. The radius R is regarded as a fixed constant so that the regularity condition (A3) stated in Section 4 hold on Θ_R. Accordingly, every subsequent statement involving Θ_R is understood with respect to this fixed radius R.

2.2 Bayesian online learning

In Bayesian online learning framework, each newly observed data updates the current posterior distribution, and the resulting posterior is then used as the prior for the next step. However, unless the model-prior pair is conjugate, exact posterior updates are generally computationally intractable. In practice, Bayesian online learning methods introduce an additional approximation step in which the updated posterior is replaced by a simpler distribution, most commonly through variational approximation (Broderick et al., 2013; Lin, 2013; Lambert et al., 2022; Lee et al., 2026a).

Let $\mathbf{D}_t = (\mathcal{X}_t, a_t, r_t)$ denote the data collected at round t. Let $\mathcal{Q}$ denote the variational family, and let the initial prior $\Pi_0 \in \mathcal{Q}$. With $\Pi_0^{\mathrm{VB}} = \Pi_0$, for each $t \ge 1$, let $\widetilde{\Pi}_t(\cdot \mid \mathbf{D}_t)$ denote the posterior obtained by updating the prior Π_{t-1}^{VB} with the new data $\mathbf{D}_t$, defined as

$$\widetilde{\Pi}_t(\mathcal{A} \mid \mathbf{D}_t) = \frac{\int_{\mathcal{A}} \exp(-\ell_t(\theta))\, \mathrm{d}\Pi_{t-1}^{\mathrm{VB}}(\theta)}{\int_{\Theta} \exp(-\ell_t(\theta))\, \mathrm{d}\Pi_{t-1}^{\mathrm{VB}}(\theta)}, \tag{2.4}$$

for any measurable $\mathcal{A} \subset \Theta$. Since the posterior distribution in (2.4) is typically intractable, we approximate it by projecting onto $\mathcal{Q}$ via

$$\Pi_t^{\mathrm{VB}} = \operatorname*{argmin}_{Q \in \mathcal{Q}} K\left(Q;\ \widetilde{\Pi}_t(\cdot \mid \mathbf{D}_t)\right), \tag{2.5}$$

where $K(P;\ Q)$ denotes the Kullback-Leibler (KL) divergence. Here, Π_t^{VB} denotes the approximated posterior, which is subsequently used as the prior for the next step. This general online variational approximation procedure has been applied to many popular Bayesian models (Broderick et al., 2013; Lin, 2013; Nguyen et al., 2018; Choi et al., 2025). We note that minimizing the KL divergence in (2.5) is equivalent to maximizing the evidence lower bound (ELBO), defined as

$$\Pi_t^{\mathrm{VB}} = \operatorname*{argmax}_{Q \in \mathcal{Q}} \left\{\mathbb{E}_Q\left[-\ell_t(\theta)\right] - K(Q;\ \Pi_{t-1}^{\mathrm{VB}})\right\}, \tag{2.6}$$

where $\mathbb{E}_Q$ denotes the expectation with respect to Q. Here, we take the variational family $\mathcal{Q}$ to be Gaussian distributions. Formally, we define $\mathcal{Q} = \left\{\mathcal{N}(\mu, \mathbf{\Omega}^{-1}) : \mu \in \Theta,\ \mathbf{\Omega} \in \mathcal{S}_+^d\right\}$, where $\mathcal{N}(\mu, \mathbf{\Omega}^{-1})$ denotes a normal distribution with mean μ and precision matrix $\mathbf{\Omega}$.

3 PROPOSED ALGORITHM

3.1 BAYESIAN ONE-PASS ONLINE LEARNING

To obtain a one-pass online update, we replace the negative log-likelihood ℓ_t in (2.6) by its quadratic approximation $\widetilde{\ell}_t$. Specifically, let the prior at step $t-1$ be $\Pi_{t-1} = \mathcal{N}(\theta_{t-1}, \mathbf{\Omega}_{t-1}^{-1})$. We define the quadratic approximation of ℓ_t around θ_{t-1} by

$$\widetilde{\ell}_t(\theta) = \ell_t(\theta_{t-1}) + \nabla \ell_t(\theta_{t-1})^\top (\theta - \theta_{t-1}) + \frac{1}{2}(\theta - \theta_{t-1})^\top \nabla^2 \ell_t(\theta_{t-1})(\theta - \theta_{t-1}).$$

Replacing ℓ_t by $\widetilde{\ell}_t$ in (2.6) leads to

$$\Pi_t = \operatorname*{argmax}_{Q \in \mathcal{Q}} \left\{ \mathbb{E}_Q \left[-\widetilde{\ell}_t(\theta) \right] - K(Q;\, \Pi_{t-1}) \right\}. \tag{3.1}$$

The solution to (3.1) can be obtained in the following closed-form updates:

$$\mathbf{\Omega}_t = \mathbf{\Omega}_{t-1} + \nabla^2 \ell_t(\theta_{t-1}), \qquad \theta_t = \theta_{t-1} - \mathbf{\Omega}_t^{-1} \nabla \ell_t(\theta_{t-1}). \tag{3.2}$$

Hence, at each round t, we obtain the Gaussian posterior approximation $\Pi_t = \mathcal{N}(\theta_t, \mathbf{\Omega}_t^{-1})$. See Appendix B for the detailed derivation of (3.2).

Although (3.2) yields a computationally efficient one-pass update, we do not apply it from the outset, since the local quadratic approximation may be inaccurate when the estimator is still far from θ^*. We therefore use a warm-start phase and switch to the online update phase after the estimator has entered a more reliable local region; this is also empirically important in certain settings; see Section 5.2.

Specifically, for each $t \le t_0$, we use a batch Gaussian approximation to the posterior based on a Laplace approximation around the maximum a posteriori (MAP) estimator. Let $\Pi_0 = \mathcal{N}(\theta_0, \mathbf{\Omega}_0^{-1})$ be the initial prior. Throughout the paper, we take $\mathbf{\Omega}_0 = \lambda \mathbf{I}_d$ for an initial prior precision parameter $\lambda > 0$. We then define

$$\widehat{\theta}_t^{\mathrm{MAP}} := \operatorname*{argmin}_{\theta \in \Theta} \sum_{s=1}^{t} \ell_s(\theta) + \frac{1}{2}\|\theta - \theta_0\|_{\mathbf{\Omega}_0}^2. \tag{3.3}$$

We then set

$$\mathbf{\Omega}_t = \mathbf{\Omega}_0 + \sum_{s=1}^{t} \nabla^2 \ell_s\left(\widehat{\theta}_t^{\mathrm{MAP}}\right), \qquad \theta_t = \widehat{\theta}_t^{\mathrm{MAP}}, \qquad t \le t_0. \tag{3.4}$$

For $t > t_0$, we initialize (3.2) at $(\theta_{t_0}, \mathbf{\Omega}_{t_0})$ obtained from the batch approximation and then apply the one-pass online update sequentially thereafter. Here, t_0 is the phase transition time that determines the length of the warm-start phase. Its choice will be specified in Section 4.

3.2 ONLINE THOMPSON SAMPLING

We now combine the Gaussian posterior approximation introduced in the previous subsection with Thompson sampling. Specifically, at round t, the Bayesian online one-pass estimator provides an approximate posterior $\Pi_{t-1} = \mathcal{N}(\theta_{t-1}, \mathbf{\Omega}_{t-1}^{-1})$. We then sample $\widetilde{\theta}_{t-1} \sim \Pi_{t-1}$ and choose the action

$$a_t \in \operatorname*{argmax}_{a \in [K]} \mu(x_{t,a}^\top \widetilde{\theta}_{t-1}).$$

After observing the reward r_t, we update the approximate posterior using the batch update (3.4) when $t \le t_0$, and the online update (3.2) when $t > t_0$. We refer to the resulting procedure as OnlineTS, which is summarized in Algorithm 1.

Warm-start phase Several prior GLB algorithms use an explicit warm-up or forced-exploration phase to obtain an initial estimator before the main online procedure begins (Li et al., 2017; Kveton et al., 2020; Jun et al., 2021; Faury et al., 2022). In contrast, our warm-start phase is not a separate exploration-only preprocessing step. For the first t_0 rounds, we use a batch Gaussian posterior approximation based on a Laplace approximation, and this approximate posterior is already used for action selection. Thus, the data collected during the warm-start phase are used directly to construct the posterior approximation employed by the algorithm, rather than being used only to certify an initialization condition for a later online update.

Algorithm 1 OnlineTS

Input: $\theta_0, \mathbf{\Omega}_0$: mean and precision matrix of the initial prior; t_0: phase transition time
for $t = 1, 2, \ldots$ **do**
  Observe $\mathcal{X}_t = \{x_{t,a} : a \in [K]\}$.
  Sample $\widetilde{\theta}_{t-1}$ from $\mathcal{N}(\theta_{t-1}, \mathbf{\Omega}_{t-1}^{-1})$ and select action $a_t = \mathrm{argmax}_{a\in[K]}\{\mu(x_{t,a}^\top \widetilde{\theta}_{t-1})\}$.
  Observe reward r_t.
  if $t \le t_0$ **then**
    Update $(\theta_t, \mathbf{\Omega}_t)$ by the batch update (3.4).
  else
    Update $(\theta_t, \mathbf{\Omega}_t)$ by the online update (3.2).
  end if
end for

Computational cost Under model (2.1), the precision update in (3.2) takes the form

$$\mathbf{\Omega}_t = \mathbf{\Omega}_{t-1} + g(\tau)^{-1}\mu'(x_{t,a_t}^\top \theta_{t-1})\, x_{t,a_t} x_{t,a_t}^\top.$$

Since the increment $\nabla^2 \ell_t(\theta_{t-1})$ is rank one under (2.1), we can update $\mathbf{\Omega}_t^{-1}$ using the Sherman–Morrison formula once $\mathbf{\Omega}_{t-1}^{-1}$ is available. Thus, the posterior update, namely the update of θ_t and $\mathbf{\Omega}_t^{-1}$, requires $O(d^2)$ computation per round in the online phase. Moreover, after time t_0, the one-pass procedure stores only the current estimator and precision matrix, requiring $O(d^2)$ memory. The warm-start phase incurs additional cost because it repeatedly computes batch updates up to time t_0 and temporarily stores the first t_0 observations. However, as will be shown in Section 4, t_0 does not grow with the horizon T. Hence, the computational cost in the warm-start phase is independent of T, and the overall procedure remains one-pass after initialization.

4 Theoretical results

We impose the following assumptions:

- (A0) $\|\theta^*\|_2 \le R^*$ for some fixed unknown $R^* > 0$.
- (A1) For all $a \in [K]$ and $t \in \mathbb{N}$, $x_{t,a} \in \mathcal{X} := \{x \in \mathbb{R}^d : \|x\|_2 \le L\}$ for some $L > 0$.
- (A2) μ is twice differentiable and there exists a non-negative constant S such that $|\mu''(z)| \le S\mu'(z)$ for all $z \in \mathbb{R}$.
- (A3) There exist constants $C_1, C_2 > 0$ such that for all $t \in \mathbb{N}$ with $t \ge C_2 d \log d$,

$$\inf_{\{\bar{\theta}_s\}_{s=1}^t \subset \Theta_R} \lambda_{\min}\left(\sum_{s=1}^{t} \nabla^2 \ell_s(\bar{\theta}_s)\right) \ge C_1\, t.$$

(A0) and (A1) are standard boundedness conditions in contextual bandits. (A0) only assumes the existence of a fixed bound R^*; it does not require R^* to be known in advance, unlike some existing works on GLBs (Jun et al., 2017; Janz et al., 2024; Lee et al., 2024; Zhang et al., 2025). (A2) is a self-concordance condition on the inverse link function. It controls how the local curvature of the loss varies with the linear predictor, and is standard in second-order analysis of GLBs (Faury et al., 2020; Russac et al., 2021; Lee et al., 2024). This condition is easily verified in several standard models. For instance, the logistic and Poisson models satisfy (A2) with $S = 1$, while the Gaussian model satisfies it with $S = 0$.

(A3) plays a different role. Since our one-pass update is based on a local quadratic approximation of the log-likelihood, our analysis is facilitated by a condition ensuring that the cumulative curvature remains uniformly stable throughout the local region Θ_R. (A3) provides such a condition by imposing a linear lower bound on the accumulated Hessian over Θ_R. In this sense, it should be viewed as a local regularity condition. In Appendix E.1, we show that this condition holds with high probability under mild assumptions on the distribution of the context set $\mathcal{X}_t$, for example when $x_{t,a}$ follows a multivariate Gaussian distribution truncated on the ℓ_2-ball of radius L. While this distributional assumption has been adopted in several prior works (Bastani et al., 2021; Kim et al., 2021; 2023), the

technique we use to derive (A3) from this assumption is novel. In our work, we establish a lower bound on the minimum eigenvalue of the Hessian constructed from contexts selected by TS, whereas prior works establish lower bounds on the minimum eigenvalue of the Gram matrix constructed either from contexts chosen by greedy policies Bastani et al. (2021) or from contexts of all arms (Kim et al., 2021; 2023).

Intuition of t_0 Our analysis further relies on a warm-start phase before the one-pass update begins. The reason is that the online update (3.2) is driven by a local quadratic approximation, and such an approximation becomes reliable only after the estimator has entered a suitable local region such as Θ_R. Motivated by this, we define the phase transition time by

$$t_0 = \left\lceil M\big(d\log(d \vee 3) + \log^2(1/\delta)\big) \right\rceil, \tag{4.1}$$

where $M > 0$ is a constant used to control the size of t_0 and $\delta \in (0, 1]$ is a confidence level. Theoretically, M does not need to be fixed at a specific value; it is enough that M be sufficiently large, as formalized in Theorems 4.1 and 4.2. Thus, t_0 should be understood as the sample size required for the initial batch phase to place the estimator in a local region where the one-pass update can be analyzed reliably.

4.1 Concentration result

We begin by establishing a high probability concentration result for the Bayesian online one-pass estimator. The proof is based on a recursive analysis of $V_t := \|\theta_t - \theta^*\|^2_{\mathbf{\Omega}_t}$. Using the online update (3.2) together with a Taylor expansion of $\ell_t(\theta^*)$ around θ_{t-1}, we obtain the one-step bound

$$V_t \le V_{t-1} + a_t V_{t-1}^{3/2} + \|\nabla \ell_t(\theta_{t-1})\|^2_{\mathbf{\Omega}_t^{-1}} + 2\big(\ell_t(\theta^*) - \ell_t(\theta_{t-1})\big),$$

where $a_t \asymp t^{-3/2}$; see Lemma C.1. The term $a_t V_{t-1}^{3/2}$ is a superlinear remainder arising from the local quadratic approximation, and its contribution becomes negligible only after the estimator has entered a sufficiently small local region. This is precisely the role of the warm-start phase, and the definition of t_0 in (4.1) is chosen accordingly so that the recursion starts from a point where this superlinear term can be controlled. The remaining two terms are handled through cumulative bounds on the self-normalized gradients and the excess loss under our assumptions. Combining these ingredients with an inductive argument shows that, once the estimator enters the local region after time t_0, it remains there and satisfies the desired uniform concentration bound for all subsequent times; see Appendix C.1 for the induction device and Appendix C for the full proof.

Theorem 4.1. *Let $\delta \in (0, 1]$ and $\lambda > 0$. Let $R > 0$ denote the radius of the local neighborhood Θ_R. Under the GLM in* (2.1)*, suppose that* (A0)*,* (A1)*,* (A2) *and* (A3) *hold. Then, there exist constants $M_0 = M_0(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0$ and $C = C(\lambda, \tau, S, L, R, R^*) > 0$ such that, if t_0 is defined by* (4.1) *with $M \ge M_0$, then with probability at least $1 - \delta$, for all $t \in \mathbb{N}$ satisfying $t \ge t_0$,*

$$\theta_t \in \Theta_R, \qquad \left\|\theta_t - \theta^*\right\|^2_{\mathbf{\Omega}_t} \le C\left(d\log t + \log^2 t + \log^2(1/\delta)\right).$$

Theorem 4.1 shows that, once the warm-start phase has placed the estimator in the local region, the one-pass recursion remains uniformly stable and the estimation error grows only logarithmically in t. This result is obtained in the bandit setting, where the selected contexts are collected adaptively, and therefore differs in nature from the non-adaptive setting studied in Bayesian one-pass online learning (Lee et al., 2026b). This uniform control of the estimator is the key input for the regret analysis in the next subsection.

4.2 Regret analysis

We now turn to the regret analysis of OnlineTS. The concentration result in Theorem 4.1 provides the key uniform control of the estimator needed for the regret analysis after the phase transition time t_0. Combining this with a standard regret decomposition yields the following high probability regret bound; see Appendix D for the full proof.

Theorem 4.2. *Let $\delta \in (0, 1]$ and $\lambda > 0$. Let $R > 0$ denote the radius of the local neighborhood Θ_R, Under the GLM in* (2.1)*, suppose that* (A0)*,* (A1)*,* (A2) *and* (A3) *hold. Then, there exist constants*

$M_0 = M_0(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0$ *and* $C = C(\lambda, \tau, S, L, R, R^*) > 0$ *such that, if* t_0 *is defined by* (4.1) *with* $M \geq M_0$*, then with probability at least* $1 - 2\delta$*, for* $T \in \mathbb{N}$ *satisfying* $T \geq t_0$*,*

$$\textit{Reg}(T) \leq C\left(\left((\kappa^{*-1/2} + C_1^{-1/2})\sqrt{d} + \sqrt{\log K}\right)\sqrt{T\log T}\log(T/\delta) + C_1^{-1} d \log T \log^2(T/\delta)\right),$$

where $\kappa^* = 1/(T^{-1}\sum_{t\in[T]} \mu'(x_{t,a_t^*}^\top \theta^*))$.

Theorem 4.2 implies that OnlineTS achieves a $\widetilde{O}(d\sqrt{T})$ regret bound whenever the curvature constant in (A3) satisfies $C_1^{-1} = O(d)$. This condition holds under certain random design settings, such as the truncated Gaussian distribution; see Proposition E.1. We discuss the curvature dependence of this bound in Appendix G. For K-armed GLBs, randomized exploration algorithms such as GLM-TSL and GLM-FPL of Kveton et al. (2020) are also known to achieve regret of order $\widetilde{O}(d\sqrt{T})$. However, those methods rely on Laplace approximations or perturbed-history estimators built from the accumulated data, whereas OnlineTS maintains an approximate posterior using the one-pass online update after the warm-start phase. Thus, OnlineTS attains the same regret order while remaining computationally efficient of its one-pass update in the online phase.

5 Experiments

In this section, we evaluate the empirical performance of OnlineTS. Our experiments address two questions. First, we compare OnlineTS with existing GLB algorithms and examine the practical effect of problem-dependent quantities required for their implementation. Second, we study the sensitivity of OnlineTS to the phase transition time t_0. We consider both logistic and Poisson models. In the experiments presented in this section, we fix θ^* with $\|\theta^*\|_2 = 1$ and use $K = 10$ arms, $\lambda = 5$, and $\delta = 0.01$. Contexts are drawn either from the Gaussian distribution $x_{t,a} \sim \mathcal{N}(0, \mathbf{I}_d)$ or from the Uniform distribution $x_{t,a} \sim \text{Unif}(\|x\|_2 \leq 3)$. All results are averaged over 10 independent runs.

5.1 Comparison with existing methods

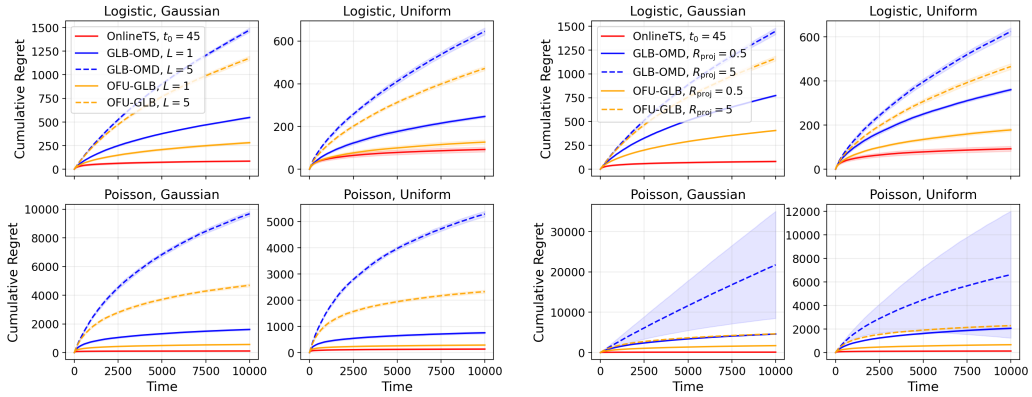
We compare OnlineTS with GLB-OMD (Zhang et al., 2025) and OFUGLB (Lee et al., 2024), which are strong representative baselines for one-pass and batch GLB algorithms, respectively. Here, GLB-OMD is a one-pass algorithm, whereas OFUGLB uses a batch estimator based on the MLE. Both methods also showed competitive empirical performance in (Zhang et al., 2025). In addition, both methods require problem-dependent quantities in their implementation. In our implementations, we supply a context norm bound L and a parameter radius R_{proj}, with the parameter set $\{\theta \in \mathbb{R}^d : \|\theta\|_2 \leq R_{\text{proj}}\}$ assumed to contain the true parameter θ^*. Details of how L and R_{proj} enter the implementations are provided in Appendix H.

In practice, however, neither the exact context bound nor a suitable projection radius is known in advance. To examine the effect of such misspecification, we vary $L \in \{1, 5\}$ and $R_{\text{proj}} \in \{0.5, 5\}$. The choice $R_{\text{proj}} = 0.5$ corresponds to a misspecified setting since $\|\theta^*\|_2 = 1$, while $R_{\text{proj}} = 5$ contains θ^* but can be overly conservative. Similarly, in the Uniform setting, $L = 1$ is misspecified since the support is not contained in $\{\|x\|_2 \leq 1\}$, whereas $L = 5$ is sufficiently large to cover the support. For OnlineTS, we fix $M = 1$ throughout, which corresponds to $t_0 = 45$ through (4.1) for the present choice of d and δ.

Figure 1 shows that GLB-OMD and OFUGLB are highly sensitive to the choice of L and R_{proj}. When these quantities are misspecified, both methods can incur substantially larger regret than OnlineTS. Moreover, even when L and R_{proj} are chosen to cover the true instance, conservative choices can still lead to significantly worse performance. In these settings, OnlineTS consistently achieves substantially lower regret than GLB-OMD and OFUGLB. Supplementary comparisons with GLB-OMD and OFUGLB under well-specified norm bounds and alternative context distributions are reported in Appendices I.2 and I.1, respectively. We extend the comparison to additional GLB algorithms in Appendix I.3, evaluating both cumulative regret and runtime.

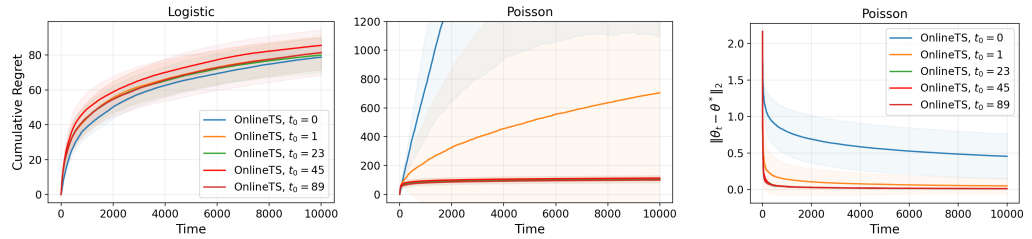
5.2 Sensitivity to the phase transition time

We next study the sensitivity of OnlineTS to the phase transition time t_0, which is determined by M through (4.1). To assess the practical role of the warm-start phase, we vary $M \in \{0,\ 0.01,\ 0.5,\ 1,\ 2\}$,



(a) Sensitivity to L (b) Sensitivity to R_{proj}

Figure 1: Comparison with existing methods under different choices of L and R_{proj}.



(a) Cumulative regret for varying M (b) ℓ_2 estimation error with Poisson

Figure 2: Sensitivity of OnlineTS to the phase transition parameter M.

which correspond to $t_0 \in \{0,\ 1,\ 23,\ 45,\ 89\}$, respectively, and compare the resulting cumulative regret. Here, $t_0 = 0$ means that the algorithm starts the one-pass update immediately without any warm-start phase, while $t_0 = 1$ corresponds to a minimal warm-start using only one sample. In this experiment, we use the Gaussian context setting.

As shown in Figure 2, OnlineTS performs well across a broad range of choices of the phase transition time once $M \geq 0.5$. In particular, for logistic bandits, the cumulative regret curves are similar across all choices of M, indicating little sensitivity to the phase transition time in these experiments. For Poisson bandits, the practical role of the warm-start phase becomes more pronounced. When $M = 0$, the cumulative regret is substantially worse, and the ℓ_2 estimation error is also noticeably larger. However, a minimal warm-start with $M = 0.01$ reduces average regret relative to $M = 0$, whereas both regret and estimation error become stable for $M \geq 0.5$. Overall, these results suggest that a small positive choice of M, such as $M = 1$, is sufficient for stable performance in the settings we considered.

6 Conclusion

We proposed OnlineTS, an online Thompson sampling algorithm for generalized linear bandits based on Bayesian one-pass online learning. The method combines a warm-start phase based on a batch Laplace approximation with a fully online Gaussian posterior update that uses only the current estimator and the current observation. As a result, OnlineTS provides a computationally efficient approach to Thompson sampling for GLBs that can be implemented without explicitly specifying problem-dependent theoretical quantities. We also showed that OnlineTS achieves a frequentist regret upper bound of $\widetilde{O}(d\sqrt{T})$. Our numerical experiments further demonstrated that OnlineTS combines competitive regret with the computational efficiency of one-pass updates and is reasonably robust to the choice of the phase transition time t_0. An interesting direction for future work is to extend the present framework beyond the fixed finite-action setting, including infinitely many actions, and to investigate whether it can be extended to broader classes of bandit models beyond GLBs.

AI USE STATEMENT

We used generative AI tools to aid and polish the writing and to assist with literature retrieval and discovery. We reviewed the AI-assisted text and checked the retrieved references against the original sources. We take responsibility for the final content of this paper, including its text, claims, and citations.

REPRODUCIBILITY STATEMENT

The assumptions underlying our theoretical results are stated in Section 4, and the proofs are provided in the appendix. Experimental settings and implementation details are described in Section 5 and the appendix. We provide the source code for reproducing our experiments as supplementary material, together with a README containing instructions for running the experiments and generating the figures.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Proc. Neural Information Processing Systems*, 2011.

Marc Abeille and Alessandro Lazaric. Linear Thompson sampling revisited. In *Proc. International Conference on Artificial Intelligence and Statistics*, 2017.

Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-wise minimax-optimal algorithms for logistic bandits. In *Proc. International Conference on Artificial Intelligence and Statistics*, 2021.

Shipra Agrawal and Navin Goyal. Analysis of Thompson sampling for the multi-armed bandit problem. In *Proc. Conference on Learning Theory*, 2012.

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *Proc. International Conference on Machine Learning*, 2013.

Shipra Agrawal and Navin Goyal. Near-optimal regret bounds for Thompson sampling. *Journal of the ACM*, 64(5):1–24, 2017.

Hamsa Bastani, Mohsen Bayati, and Khashayar Khosravi. Mostly exploration-free algorithms for contextual bandits. *Management Science*, 67(3):1329–1349, 2021.

Adam Block, Max Simchowitz, and Alexander Rakhlin. Oracle-efficient smoothed online learning for piecewise continuous decision making. In *Proc. Conference on Learning Theory*, 2023.

Tamara Broderick, Nicholas Boyd, Andre Wibisono, Ashia C Wilson, and Michael I Jordan. Streaming variational Bayes. In *Proc. Neural Information Processing Systems*, 2013.

Olivier Chapelle and Lihong Li. An empirical evaluation of Thompson sampling. In *Proc. Neural Information Processing Systems*, 2011.

Xi Chen, David Simchi-Levi, and Yining Wang. Privacy-preserving dynamic personalized pricing with demand learning. *Management Science*, 68(7):4878–4898, 2022.

Junhyeok Choi, Jeyong Lee, Yongdai Kim, and Minwoo Chae. Online Bayesian inference for Cox proportional hazards model. *Journal of Computational and Graphical Statistics*, pp. 1–9, 2025.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proc. International Conference on Artificial Intelligence and Statistics*, 2011.

Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *Proc. Conference on Learning Theory*, 2008.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *Proc. International Conference on Machine Learning*, 2020.

Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clément Calauzènes. Jointly efficient and optimal algorithms for logistic bandits. In *Proc. International Conference on Artificial Intelligence and Statistics*, 2022.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In *Proc. Neural Information Processing Systems*, 2010.

Steven CH Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. Online learning: A comprehensive survey. *Neurocomputing*, 459:249–289, 2021.

David Janz, Shuai Liu, Alex Ayoub, and Csaba Szepesvári. Exploration via linearly perturbed loss minimisation. In *Proc. International Conference on Artificial Intelligence and Statistics*, 2024.

Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, and Rebecca Willett. Scalable generalized linear bandits: Online computation and hashing. In *Proc. Neural Information Processing Systems*, 2017.

Kwang-Sung Jun, Lalit Jain, Blake Mason, and Houssam Nassif. Improved confidence bounds for the linear logistic model and applications to bandits. In *Proc. International Conference on Machine Learning*, 2021.

Dongguen Kim, Young-Geun Choi, and Minwoo Chae. A Bayesian approach to contextual dynamic pricing using the proportional hazards model with discrete price data. In *Proc. Neural Information Processing Systems*, 2025.

Wonyoung Kim, Gi-Soo Kim, and Myunghee Cho Paik. Doubly robust thompson sampling with linear payoffs. *Advances in neural information processing systems*, 34:15830–15840, 2021.

Wonyoung Kim, Kyungbok Lee, and Myunghee Cho Paik. Double doubly robust Thompson sampling for generalized linear contextual bandits. In *Proc. AAAI Conference on Artificial Intelligence*, 2023.

Branislav Kveton, Manzil Zaheer, Csaba Szepesvari, Lihong Li, Mohammad Ghavamzadeh, and Craig Boutilier. Randomized exploration in generalized linear bandits. In *Proc. International Conference on Artificial Intelligence and Statistics*, 2020.

Marc Lambert, Silvere Bonnabel, and Francis Bach. The recursive variational Gaussian approximation (R-VGA). *Statistics and Computing*, 32(1):10, 2022.

Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, 28(5):1302–1338, 2000.

Jeyong Lee, Junhyeok Choi, and Minwoo Chae. Online Bernstein-von Mises theorem. *Journal of Machine Learning Research*, 27(49):1–124, 2026a.

Jeyong Lee, Junhyeok Choi, Dongguen Kim, and Minwoo Chae. Bayesian online learning in the one-pass regime: Frequentist validity and uncertainty quantification. *arXiv preprint arXiv:2604.27442*, 2026b.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. A unified confidence sequence for generalized linear models, with applications to bandits. In *Proc. Neural Information Processing Systems*, 2024.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proc. International Conference on World Wide Web*, 2010.

Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *Proc. International Conference on Machine Learning*, 2017.

Dahua Lin. Online learning of nonparametric mixture models via sequential variational approximation. In *Proc. Neural Information Processing Systems*, 2013.

Liam Madden, Emiliano Dall'Anese, and Stephen Becker. High probability convergence bounds for non-convex stochastic gradient descent with sub-Weibull noise. *Journal of Machine Learning Research*, 25(241):1–36, 2024.

Peter McCullagh and John Ashworth Nelder. *Generalized Linear Models*. Chapman & Hall, 1989.

Cuong V Nguyen, Yingzhen Li, Thang D Bui, and Richard E Turner. Variational continual learning. In *Proc. International Conference on Learning Representations*, 2018.

David Pollard. Empirical processes: Theory and applications. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 2:1–86, 1990.

Paat Rusmevichientong and John N Tsitsiklis. Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411, 2010.

Yoan Russac, Louis Faury, Olivier Cappé, and Aurélien Garivier. Self-concordant analysis of generalized linear bandits with forgetting. In *Proc. International Conference on Artificial Intelligence and Statistics*, 2021.

Daniel J. Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, and Zheng Wen. A tutorial on Thompson sampling. *Foundations and Trends in Machine Learning*, 11(1):1–96, 2018.

Ayush Sawarni, Nirjhar Das, Siddharth Barman, and Gaurav Sinha. Generalized linear bandits with limited adaptivity. In *Proc. Neural Information Processing Systems*, 2024.

Ambuj Tewari and Susan A Murphy. From ads to interventions: Contextual bandits in mobile health. In *Mobile Health: Sensors, Analytic Methods, and Applications*, pp. 495–517. Springer, 2017.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2 edition, 2026.

Yining Wang, Ruosong Wang, Simon Shaolei Du, and Akshay Krishnamurthy. Optimism in reinforcement learning with generalized linear function approximation. In *Proc. International Conference on Learning Representations*, 2021.

Michael Woodroofe. A one-armed bandit problem with a concomitant variable. *Journal of the American Statistical Association*, 74(368):799–806, 1979.

Yu-Jie Zhang, Sheng-An Xu, Peng Zhao, and Masashi Sugiyama. Generalized linear bandits: Almost optimal regret with one-pass update. In *Proc. Neural Information Processing Systems*, 2025.

CONTENTS

A Notations **13**

B Online update formula **13**

C Proof of Theorem 4.1 **14**

C.1 Induction devices 19

D Proof of Theorem 4.2 **23**

E Justifications for conditions **28**

E.1 Justification of (A3) 29

E.2 Justification of (B1) 35

E.3 Justification of (B2) 36

E.4 Justification of (B3) 43

F Technical lemmas **48**
F.1 Lemmas for main results 49
F.2 Lemmas for Section E.1 49
F.3 Lemmas for Section E.2 51
F.4 Lemmas for Section E.3 51
F.5 Lemmas for Section E.4 55

G Curvature dependence of the regret bound **57**

H Implementation details of the comparison methods **57**

I Additional experiments **58**
I.1 Alternative context distributions 58
I.2 Additional comparisons in well-specified settings 58
I.3 Comparison with additional GLB algorithms 59

J Computational resources used **59**

A NOTATIONS

We collect additional notation used in the appendix. Let $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ denote the unit sphere in $\mathbb{R}^d$. For symmetric matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{d\times d}$, we write $\mathbf{A} \succeq \mathbf{B}$ if $\mathbf{A} - \mathbf{B}$ is positive semidefinite. For $\mathbf{A} \in \mathcal{S}_+^d$, let $\det(\mathbf{A})$ denote its determinant. For a matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$, let $\|\mathbf{A}\|_{\text{op}} = \sup_{\|u\|_2=1} \|\mathbf{A}u\|_2$ denote its operator norm. For $x \in \mathbb{R}$, we write $x_+ = \max\{x, 0\}$. For an event $\mathcal{E}$, let $\mathbb{1}[\mathcal{E}]$ denote its indicator function.

B ONLINE UPDATE FORMULA

In this section, we derive the closed-form update (3.2) from (3.1). For $Q = \mathcal{N}(\mu, \mathbf{\Omega}^{-1}) \in \mathcal{Q}$, define the negative ELBO objective

$$F(Q) = F(\mu, \mathbf{\Omega}^{-1}) = \mathbb{E}_Q\left[\widetilde{\ell}_t(\theta)\right] + K(Q;\, \Pi_{t-1}), \tag{B.1}$$

where θ denotes the integration variable under Q. Since both Q and Π_{t-1} are Gaussian, the KL divergence is expressed by the closed-form

$$K(Q;\, \Pi_{t-1}) = \frac{1}{2}\Big(\log\det(\mathbf{\Omega}_{t-1}^{-1}) - \log\det(\mathbf{\Omega}^{-1}) + (\mu - \theta_{t-1})^\top \mathbf{\Omega}_{t-1}(\mu - \theta_{t-1}) + \operatorname{tr}(\mathbf{\Omega}_{t-1}\mathbf{\Omega}^{-1}) - d\Big).$$

Moreover, the expectation term in (B.1) can be written as

$$\mathbb{E}_Q\left[\widetilde{\ell}_t(\theta)\right] = \left(\nabla\ell_t(\theta_{t-1}) - \nabla^2\ell_t(\theta_{t-1})\theta_{t-1}\right)^\top \mu + \frac{1}{2}\operatorname{tr}\left(\nabla^2\ell_t(\theta_{t-1})(\mu\mu^\top + \mathbf{\Omega}^{-1})\right) + C,$$

where C denotes a term independet of μ and $\mathbf{\Omega}$. Differentiating $F(\mu, \mathbf{\Omega}^{-1})$ with respect to μ and $\mathbf{\Omega}^{-1}$, we obtain

$$\nabla_\mu F(\mu, \mathbf{\Omega}^{-1}) = \nabla\ell_t(\theta_{t-1}) + \nabla^2\ell_t(\theta_{t-1})(\mu - \theta_{t-1}) + \mathbf{\Omega}_{t-1}(\mu - \theta_{t-1}),$$
$$\nabla_{\mathbf{\Omega}^{-1}} F(\mu, \mathbf{\Omega}^{-1}) = \frac{1}{2}\left(\nabla^2\ell_t(\theta_{t-1}) - \mathbf{\Omega} + \mathbf{\Omega}_{t-1}\right).$$

Setting these gradients equal to zero yields

$$\mathbf{\Omega}_t = \mathbf{\Omega}_{t-1} + \nabla^2\ell_t(\theta_{t-1}), \quad \theta_t = \theta_{t-1} - \mathbf{\Omega}_t^{-1}\nabla\ell_t(\theta_{t-1}),$$

which is the update formula in (3.2).

C Proof of Theorem 4.1

Our concentration result is stated under the structural assumptions (A0), (A1), (A2) and (A3). To prove it, we first introduce three intermediate conditions, (B1), (B2) and (B3), which will later be verified from the assumptions (A0), (A1), (A2) and (A3) in Section E. We then show that they are sufficient to derive the desired concentration bound.

Define

$$L_\mu = L_\mu(L, R, R^*) := \sup_{x \in \mathcal{X}, \theta \in \Theta_R} \mu'(x^\top \theta). \tag{C.1}$$

We also define the stopping time

$$t_{\texttt{exit}} = \inf\{t \in \mathbb{N} : t > t_0, \theta_t \notin \Theta_R\}. \tag{C.2}$$

The stopping time $t_{\texttt{exit}}$ is introduced to localize the analysis to the region Θ_R. We will first derive the desired concentration bound up to $t_{\texttt{exit}}$, and then show that, for a sufficiently large choice of t_0, $\theta_t \in \Theta_R$ for all $t \geq t_0$.

We next introduce three intermediate conditions used in the proof. These are not additional assumptions of the model; rather, they will be verified later from (A0), (A1), (A2) and (A3).

(B1) Let $\delta \in (0, 1]$, and fix a phase transition time $t_0 \in \mathbb{N}$. There exists a constant $K_1 > 0$ such that, on an event $\mathcal{E}_1(\delta)$ satisfying $\mathbb{P}(\mathcal{E}_1(\delta)) \geq 1 - \delta$, for all $t \in \mathbb{N}$ satisfying $t > t_0$,

$$\sum_{s=t_0+1}^{t} (\ell_s(\theta^*) - \ell_s(\theta_{s-1})) \leq K_1 \log \frac{1}{\delta}.$$

(B2) Let $\delta \in (0, 1]$, and fix a phase transition time $t_0 \in \mathbb{N}$. There exists a constant $K_2 > 0$ such that, on an event $\mathcal{E}_2(\delta)$ satisfying $\mathbb{P}(\mathcal{E}_2(\delta)) \geq 1 - \delta$, for all $t \in \mathbb{N}$ satisfying $t_0 < t \leq t_{\texttt{exit}}$,

$$\sum_{s=t_0+1}^{t} \|\nabla \ell_s(\theta_{s-1})\|^2_{\mathbf{\Omega}_s^{-1}} \leq K_2 \left(d \log \left(1 + \frac{t - t_0}{d} \right) + \log^2 \left(\frac{t - t_0}{\delta} \right) \right).$$

(B3) Let $\delta \in (0, 1]$. There exists a large enough constant $M = M(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0$ such that, on an event $\mathcal{E}_3(\delta)$ with $\mathbb{P}(\mathcal{E}_3(\delta)) \geq 1 - \delta$,

$$\begin{aligned} &t_0 = t_0(M, d, \delta) = \lceil M(d \log(d \vee 3) + \log^2(1/\delta)) \rceil, \\ &\left\| \widehat{\theta}^{\texttt{MAP}}_{t_0} - \theta^* \right\|_{\mathbf{\Omega}_{t_0}} \leq K_3 \sqrt{d + \log(1/\delta)}, \\ &\widehat{\theta}^{\texttt{MAP}}_t \in \Theta_R, \qquad \forall t \in \mathbb{N} \text{ with } t \geq t_0, \end{aligned}$$

where $K_3 > 0$.

(B1) controls the cumulative excess loss relative to the truth, (B2) bounds the accumulated self-normalized gradient fluctuations, and (B3) provides the required initialization and localization from the phase transition time t_0. For $\delta \in (0, 1]$, define the event

$$\mathcal{E}(\delta) = \mathcal{E}_1(\delta) \cap \mathcal{E}_2(\delta) \cap \mathcal{E}_3(\delta).$$

On the event $\mathcal{E}(\delta)$, all three conditions hold simultaneously.

The next lemma provides the key one-step recursive inequality for $\|\theta_t - \theta^*\|^2_{\mathbf{\Omega}_t}$.

Lemma C.1. *Let $\delta \in (0, 1]$. Under the GLM in* (2.1)*, suppose that* (A0)*,* (A1)*,* (A2)*,* (A3) *and* (B3) *hold. Then, on the event $\mathcal{E}(\delta)$, it holds that*

$$\left\|\theta_t - \theta^*\right\|^2_{\mathbf{\Omega}_t} \leq \left\|\theta_{t-1} - \theta^*\right\|^2_{\mathbf{\Omega}_{t-1}} + a_{t-1} \left\|\theta_{t-1} - \theta^*\right\|^3_{\mathbf{\Omega}_{t-1}} + b_t, \quad \forall t \in \mathbb{N} \text{ with } t_0 < t \leq t_{\texttt{exit}},$$

where

$$\begin{aligned} a_t &= C\, t^{-3/2}, \\ b_t &= \left\|\nabla \ell_t(\theta_{t-1})\right\|^2_{\mathbf{\Omega}_t^{-1}} + 2\left(\ell_t(\theta^*) - \ell_t(\theta_{t-1})\right) \end{aligned}$$

and $C = C(\tau, S, L, R, R^*, C_1) > 0$ is a constant.

Proof. Fix t with $t_0 < t \le t_{\texttt{exit}}$. For simplicity of notation, define

$$V_t = \big\|\theta_t - \theta^*\big\|_{\mathbf{\Omega}_t}^2, \qquad \Delta_t = \theta_t - \theta^*, \qquad g_t = \nabla \ell_t(\theta_{t-1}), \qquad \mathbf{H}_t = \nabla^2 \ell_t(\theta_{t-1}).$$

By the online update formula (3.2), we have

$$\begin{aligned} V_t &= \big\|\theta_t - \theta^*\big\|_{\mathbf{\Omega}_t}^2 \\ &= \big\|\theta_{t-1} - \mathbf{\Omega}_t^{-1} g_t - \theta^*\big\|_{\mathbf{\Omega}_t}^2 \\ &= \big\|\Delta_{t-1}\big\|_{\mathbf{\Omega}_t}^2 + \big\|g_t\big\|_{\mathbf{\Omega}_t^{-1}}^2 - 2\langle g_t, \Delta_{t-1}\rangle \\ &= V_{t-1} + \big\|\Delta_{t-1}\big\|_{\mathbf{H}_t}^2 + \big\|g_t\big\|_{\mathbf{\Omega}_t^{-1}}^2 - 2\langle g_t, \Delta_{t-1}\rangle. \end{aligned} \tag{C.3}$$

By Taylor's theorem with the integral form of the remainder, we have

$$\ell_t(\theta^*) = \ell_t(\theta_{t-1}) - \langle g_t, \Delta_{t-1}\rangle + \big\|\Delta_{t-1}\big\|_{\widetilde{\mathbf{H}}_t}^2, \tag{C.4}$$

where

$$\widetilde{\mathbf{H}}_t = \int_0^1 (1-u)\nabla^2 \ell_t(\theta_{t-1} + u(\theta^* - \theta_{t-1}))\mathrm{d}u.$$

By the definition of the Hessian in (2.3), we can further express $\widetilde{\mathbf{H}}_t$ as

$$\widetilde{\mathbf{H}}_t = g(\tau)^{-1}\widetilde{\alpha}(\theta_{t-1}, \theta^*, x_t) x_t x_t^\top, \tag{C.5}$$

where

$$\widetilde{\alpha}(\theta_{t-1}, \theta^*, x_t) = \int_0^1 (1-u)\mu'\big(x_t^\top(\theta_{t-1} + u(\theta^* - \theta_{t-1}))\big)\mathrm{d}u.$$

By Lemma F.1, we have

$$\widetilde{\alpha}(\theta_{t-1}, \theta^*, x_t) \ge \frac{\mu'(x_t^\top \theta_{t-1})}{2 + S|x_t^\top \Delta_{t-1}|},$$

which implies that

$$\begin{aligned} \mathbf{H}_t &\preceq \big(2 + S|x_t^\top \Delta_{t-1}|\big)\widetilde{\mathbf{H}}_t \\ &= (2+\phi_t)\,\widetilde{\mathbf{H}}_t, \end{aligned} \tag{C.6}$$

where we denote $\phi_t := S|x_t^\top \Delta_{t-1}|$ for simplicity. Substituting (C.6) into (C.3), we have

$$\begin{aligned} V_t &\le V_{t-1} + (2+\phi_t)\big\|\Delta_{t-1}\big\|_{\widetilde{\mathbf{H}}_t}^2 + \big\|g_t\big\|_{\mathbf{\Omega}_t^{-1}}^2 - 2\langle g_t, \Delta_{t-1}\rangle \\ &= V_{t-1} + \phi_t\big\|\Delta_{t-1}\big\|_{\widetilde{\mathbf{H}}_t}^2 + \big\|g_t\big\|_{\mathbf{\Omega}_t^{-1}}^2 + 2\big(\ell_t(\theta^*) - \ell_t(\theta_{t-1})\big), \end{aligned} \tag{C.7}$$

where the last equality follows from (C.4). Recall that Θ_R is a neighborhood of θ^*, and by definition of $t_{\texttt{exit}}$ and (B3), $\theta_s \in \Theta_R$ for all $s \in \mathbb{N}$ with $t_0 \le s < t_{\texttt{exit}}$. Then, by (A1), we have

$$\mu'\big(x_t^\top(\theta_{t-1} + u(\theta^* - \theta_{t-1}))\big) \le L_\mu, \quad \forall u \in [0,1],$$

where $L_\mu = L_\mu(L, R, R^*) > 0$ is defined as (C.1). Substituting it into (C.5), we have

$$\begin{aligned} \big\|\widetilde{\mathbf{H}}_t\big\|_{\text{op}} &\le g(\tau)^{-1}\int_0^1 (1-u)L_\mu \mathrm{d}u\,\big\|x_t\big\|_2^2 \\ &\le (2g(\tau))^{-1} L_\mu L^2, \end{aligned} \tag{C.8}$$

where the last inequality holds by (A1). Since $\widehat{\theta}_{t_0}^{\texttt{MAP}} \in \Theta_R$ and $\{\theta_{s-1}\}_{s=t_0+1}^{t_{\texttt{exit}}} \subset \Theta_R$, for any $t \in \mathbb{N}$ with $t_0 \le t \le t_{\texttt{exit}}$, we have

$$\begin{aligned} \lambda_{\min}(\mathbf{\Omega}_t) &\ge \lambda_{\min}(\mathbf{\Omega}_0) + \lambda_{\min}\left(\sum_{s=1}^{t_0} \nabla^2 \ell_s(\widehat{\theta}_{t_0}^{\texttt{MAP}}) + \sum_{s=t_0+1}^{t} \nabla^2 \ell_s(\theta_{s-1})\right) \\ &\ge C_1 t, \end{aligned} \tag{C.9}$$

where the last inequality holds by (A3). Hence, for $t_0 < t \le t_{\texttt{exit}}$, we have

$$\begin{aligned}\big\|\Delta_{t-1}\big\|_2 &\le \big\|\mathbf{\Omega}_{t-1}^{-1/2}\big\|_{\text{op}}\,\big\|\mathbf{\Omega}_{t-1}^{1/2}\Delta_{t-1}\big\|_2 \\ &= \lambda_{\min}(\mathbf{\Omega}_{t-1})^{-1/2}\,V_{t-1}^{1/2} \\ &\le C_1^{-1/2}(t-1)^{-1/2}\,V_{t-1}^{1/2},\end{aligned} \tag{C.10}$$

where the last inequality holds by (C.9). Recall the definition of ϕ_t:

$$\begin{aligned}\phi_t &= S|x_t^\top \Delta_{t-1}| \\ &\le SL\|\Delta_{t-1}\|_2 \\ &\le SLC_1^{-1/2}(t-1)^{-1/2}V_{t-1}^{1/2},\end{aligned} \tag{C.11}$$

where the last inequality holds by (C.10). Combining (C.7), (C.8), (C.10) and (C.11), we have

$$\begin{aligned}V_t &\le V_{t-1} + \phi_t\big\|\Delta_{t-1}\big\|_{\widetilde{\mathbf{H}}_t}^2 + \big\|g_t\big\|_{\mathbf{\Omega}_t^{-1}}^2 + 2\big(\ell_t(\theta^*) - \ell_t(\theta_{t-1})\big) \\ &\le V_{t-1} + \phi_t\big\|\widetilde{\mathbf{H}}_t\big\|_{\text{op}}\big\|\Delta_{t-1}\big\|_2^2 + \big\|g_t\big\|_{\mathbf{\Omega}_t^{-1}}^2 + 2\big(\ell_t(\theta^*) - \ell_t(\theta_{t-1})\big) \\ &\le V_{t-1} + \frac{SL_\mu L^3}{2g(\tau)}C_1^{-3/2}(t-1)^{-3/2}V_{t-1}^{3/2} + \big\|g_t\big\|_{\mathbf{\Omega}_t^{-1}}^2 + 2\big(\ell_t(\theta^*) - \ell_t(\theta_{t-1})\big) \\ &\le V_{t-1} + a_{t-1}V_{t-1}^{3/2} + b_t,\end{aligned}$$

where

$$\begin{aligned}a_t &= \frac{SL_\mu L^3}{2g(\tau)}C_1^{-3/2}\,t^{-3/2} = C\,t^{-3/2}, \\ b_t &= \big\|g_t\big\|_{\mathbf{\Omega}_t^{-1}}^2 + 2\big(\ell_t(\theta^*) - \ell_t(\theta_{t-1})\big),\end{aligned}$$

and $C = C(\tau, S, L, L_\mu, C_1) = C(\tau, S, L, R, R^*, C_1) > 0$ is a constant. It holds for arbitrary t with $t_0 < t \le t_{\texttt{exit}}$, which completes the proof.

□

Lemma C.1 reduces the proof of the concentration result to the analysis of a nonlinear recursive inequality. The next theorem shows that this recursive inequality, together with the three conditions (B1), (B2) and (B3), is sufficient to yield the desired concentration bound.

Theorem C.2. *Let $\delta \in (0, 1]$. Under the GLM in* (2.1)*, suppose that* (A0)*,* (A1)*,* (A2)*,* (A3)*,* (B1)*,* (B2) *and* (B3) *hold. If t_0 is defined by* (4.1) *with $M \ge M_0$, then on the event $\mathcal{E}(\delta)$, for all $t \in \mathbb{N}$ satisfying $t \ge t_0$,*

$$\big\|\theta_t - \theta^*\big\|_{\mathbf{\Omega}_t}^2 \le C\left(d\log t + \log^2 t + \log^2\frac{1}{\delta}\right), \qquad \theta_t \in \Theta_R,$$

where $C = C(K_1, K_2, K_3) > 0$ and $M_0 = M_0(\lambda, \tau, S, L, R, R^*, C_1, C_2, K_1, K_2, K_3) > 0$.

Proof. Throughout the proof, we work on the event $\mathcal{E}(\delta)$. First, we verify that the conditions of Lemma C.4 hold. For simplicity of notation, define

$$V_t = \big\|\theta_t - \theta^*\big\|_{\mathbf{\Omega}_t}^2, \qquad \Delta_t = \theta_t - \theta^*, \qquad g_t = \nabla\ell_t(\theta_{t-1}), \qquad \mathbf{H}_t = \nabla^2\ell_t(\theta_{t-1}).$$

Since (A0), (A1), (A2), (A3) and (B3) hold, by Lemma C.1, we have

$$V_t \le V_{t-1} + a_{t-1}V_{t-1}^{3/2} + b_t, \quad \forall t \in \mathbb{N} \text{ with } t_0 < t \le t_{\texttt{exit}},$$

where

$$\begin{aligned}a_t &= c_1 t^{-3/2}, \\ b_t &= \big\|g_t\big\|_{\mathbf{\Omega}_t^{-1}}^2 + 2\big(\ell_t(\theta^*) - \ell_t(\theta_{t-1})\big),\end{aligned} \tag{C.12}$$

and $c_1 = c_1(\tau, S, L, R, R^*, C_1) > 0$. Let $t_1, t_2 \in \mathbb{N}$ with $t_0 \le t_1 < t_2 \le t_{\texttt{exit}}$. Note that

$$\sum_{s=t_1}^{t_2} a_s = c_1 \sum_{s=t_1}^{t_2} s^{-3/2} \le c_1 \sum_{s=t_1}^{\infty} s^{-3/2} \le c_2 t_1^{-1/2},$$

where we set $c_2 = 2c_1$ using $\sum_{s=t_1}^{\infty} s^{-3/2} \le 2t_1^{-1/2}$. By (B1) and (B2), we have

$$\begin{aligned}
\sum_{s=t_0+1}^{t_2} b_s &= \sum_{s=t_0+1}^{t_2} \left\|g_s\right\|_{\mathbf{\Omega}_s^{-1}}^2 + 2 \sum_{s=t_0+1}^{t_2} \left(\ell_s(\theta^*) - \ell_s(\theta_{s-1})\right) \\
&\le K_2 \left(d \log \left(1 + \frac{t_2 - t_0}{d} \right) + \log^2 \left(\frac{t_2 - t_0}{\delta} \right) \right) + 2K_1 \log \frac{1}{\delta} \\
&\le \widetilde{K}_1 \left(d \log \left(1 + \frac{t_2 - t_0}{d} \right) + \log^2 \left(\frac{t_2 - t_0}{\delta} \right) + \log \frac{1}{\delta} \right),
\end{aligned}$$

where $\widetilde{K}_1 = \widetilde{K}_1(K_1, K_2) > 0$. In addition, by (B3), there exists a constant $M_0' = M_0'(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0$ such that if $M \ge M_0'$, then

$$V_{t_0} \le K_3^2 \left(d + \log \frac{1}{\delta} \right).$$

Combining the last five displays, we verify the conditions of Lemma C.4. Therefore, by Lemma C.4, there exists a constant $M_1 = M_1 = M_1(c_2, K_3, \widetilde{K}_1) = M_1(\tau, S, L, R, R^*, C_1, K_1, K_2, K_3) > 0$ such that if $M > 0$ is large enough such that $M \ge M_0' \vee M_1$, for all $t \in \mathbb{N}$ satisfying $t_0 \le t \le t_{\texttt{exit}}$,

$$\begin{aligned}
V_t &\le \tilde{c}_2 \left(d \log \left(1 + \frac{2t - t_0}{d} \right) + \log^2 (2t - t_0) + \log^2 \left(\frac{1}{\delta} \right) \right) \\
&\le C \left(d \log t + \log^2 t + \log^2 \frac{1}{\delta} \right),
\end{aligned} \tag{C.13}$$

where $\tilde{c}_2 = \tilde{c}_2(K_3, \widetilde{K}_1) = \tilde{c}_2(K_1, K_2, K_3) > 0$ and $C = C(\tilde{c}_2) = C(K_1, K_2, K_3) > 0$.

We have thus obtained the weighted-norm bound (C.13) up to the exit time $t_{\texttt{exit}}$. Now, we will show that $t_{\texttt{exit}} = \infty$ by contradiction. Suppose that $t_{\texttt{exit}} < \infty$. As shown in the proof of Lemma C.1, (A3) and (B3) imply that

$$\lambda_{\min}(\mathbf{\Omega}_t) \ge C_1 t, \quad \forall t \in \mathbb{N} \text{ with } t_0 \le t \le t_{\texttt{exit}}. \tag{C.14}$$

Hence, for $t \in \mathbb{N}$ with $t_0 \le t \le t_{\texttt{exit}}$, we have

$$\begin{aligned}
\left\|\theta_t - \theta^*\right\|_2^2 &\le \lambda_{\min}(\mathbf{\Omega}_t)^{-1} \left\|\theta_t - \theta^*\right\|_{\mathbf{\Omega}_t}^2 \\
&\le \tilde{c}_2 C_1^{-1} t^{-1} \left(d \log \left(1 + \frac{2t - t_0}{d} \right) + \log^2 (2t - t_0) + \log^2 \left(\frac{1}{\delta} \right) \right),
\end{aligned}$$

where the last inequality holds by (C.13) and (C.14). Note that the map

$$t \mapsto t^{-1} \left(d \log \left(1 + \frac{2t - t_0}{d} \right) + \log^2 (2t - t_0) + \log^2 \left(\frac{1}{\delta} \right) \right)$$

is a decreasing function for $t \ge t_0$. Then, for $t \in \mathbb{N}$ with $t_0 \le t \le t_{\texttt{exit}}$, we have

$$\begin{aligned}
\left\|\theta_t - \theta^*\right\|_2^2 &\le \tilde{c}_2 C_1^{-1} t^{-1} \left(d \log \left(1 + \frac{2t - t_0}{d} \right) + \log^2 (2t - t_0) + \log^2 \left(\frac{1}{\delta} \right) \right) \\
&\le \tilde{c}_2 C_1^{-1} t_0^{-1} \left(d \log \left(1 + \frac{t_0}{d} \right) + \log^2 (t_0) + \log^2 \left(\frac{1}{\delta} \right) \right) \\
&\le \tilde{c}_2 \tilde{c}_3 C_1^{-1} M^{-1/2} \\
&\le \tilde{c}_2 \tilde{c}_3 C_1^{-1} M_2^{-1/2} = R^2,
\end{aligned}$$

where the third inequality holds by the universal constant $\widetilde{c}_3 > 0$, the fourth inequality holds for $M \geq M_2$, and the last equality holds by taking $M_2 = \tilde{c}_2^2 \tilde{c}_3^2 C_1^{-2} R^{-4} = M_2(R, C_1, K_1, K_2, K_3) > 0$. We take

$$M_0 := M_0' \vee M_1 \vee M_2 = M_0(\lambda, \tau, S, L, R, R^*, C_1, C_2, K_1, K_2, K_3) > 0.$$

Then, if $M > 0$ is large enough with $M \geq M_0$, $\theta_{t_{\texttt{exit}}} \in \Theta_R$, which contradicts the definition of $t_{\texttt{exit}}$. Therefore, $t_{\texttt{exit}} = \infty$ and hence $\theta_t \in \Theta_R$ for all $t \geq t_0$, which completes the proof.

□

Theorem C.2 is an abstract concentration statement. It shows that once the conditions (B1), (B2), and (B3) have been established, the desired bound follows. It remains to verify these conditions under the assumptions (A0), (A1), (A2) and (A3), which we now do using the results of Section E.

Proposition C.1. *Let $\delta \in (0, 1]$ and $\lambda > 0$. Under the GLM in* (2.1)*, suppose that* (A0)*,* (A1)*,* (A2)*, and* (A3) *hold. Then,* (B1)*,* (B2)*, and* (B3) *hold with constants*

$$K_1 = 1, \qquad K_2 = K_2(\lambda, \tau, S, L, R, R^*), \qquad K_3 = K_3(\lambda, \tau, S, L, R, R^*),$$

and with

$$t_0 = \lceil M\big(d \log(d \vee 3) + \log^2(1/\delta)\big) \rceil$$

for any sufficiently large $M > 0$ satisfying $M \geq M_0$, where

$$M_0 = M_0(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0.$$

Proof. Proposition E.2 verifies (B1) with $K_1 = 1$. Proposition E.3 verifies (B2) with $K_2 = K_2(\lambda, \tau, S, L, R, R^*)$. Proposition E.4 verifies (B3) with $K_3 = K_3(\lambda, \tau, S, L, R, R^*)$ and

$$t_0 = \lceil M\big(d \log(d \vee 3) + \log^2(1/\delta)\big) \rceil$$

for sufficiently large $M > 0$ satisfying $M \geq M_0$, where

$$M_0 = M_0(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0.$$

Therefore, (B1), (B2), and (B3) hold with the stated constants. □

Combining Theorem C.2 with Proposition C.1, we immediately obtain the desired concentration result under the structural assumptions of the model.

Proof of Theorem 4.1. Fix $\delta \in (0, 1]$ and $\lambda > 0$. Under (A0), (A1), (A2) and (A3), by Proposition C.1, (B1), (B2) and (B3) hold with

$$K_1 = 1, \quad K_2 = K_2(\lambda, \tau, S, L, R, R^*), \quad K_3 = K_3(\lambda, \tau, S, L, R, R^*),$$

and

$$t_0 = \lceil M(d \log(d \vee 3) + \log^2(1/\delta)) \rceil$$

for sufficiently large $M > 0$ satisfying $M \geq M_0'$, where

$$M_0' = M_0'(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0.$$

Therefore, Theorem C.2 yields that if t_0 is defined by (4.1) with $M \geq M_0$, then on the event $\mathcal{E}(\delta)$, for all $t \in \mathbb{N}$ satisfying $t \geq t_0$,

$$\big\|\theta_t - \theta^*\big\|_{\mathbf{\Omega}_t}^2 \leq \widetilde{C}\left(d \log t + \log^2 t + \log^2 \frac{1}{\delta}\right), \qquad \theta_t \in \Theta_R,$$

where $\widetilde{C} = \widetilde{C}(\lambda, \tau, S, L, R, R^*) > 0$ and $M_0 = M_0(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0$. By construction of $\mathcal{E}(\delta)$, we have

$$\mathbb{P}(\mathcal{E}(\delta)) \geq 1 - 3\delta.$$

Replacing δ by $\delta/3$, we conclude that there exists a constant $C = C(\lambda, \tau, S, L, R, R^*) > 0$ such that with probability at least $1 - \delta$, for all $t \in \mathbb{N}$ satisfying $t \geq t_0$,

$$\big\|\theta_t - \theta^*\big\|_{\mathbf{\Omega}_t}^2 \leq C\left(d \log t + \log^2 t + \log^2 \frac{1}{\delta}\right), \qquad \theta_t \in \Theta_R.$$

This completes the proof. □

We now combine Proposition E.1 with Theorem 4.1 to obtain a concrete concentration result under the truncated Gaussian design. In particular, under the random design (T1) and (T2), the eigenvalue condition (A3) is no longer an additional assumption, but a consequence of the random design itself.

Corollary C.1. *Let $\delta \in (0,1]$ and $\lambda > 0$. Let $\sigma_{\min} > 0$ and $\sigma_{\max} > 0$ are constants defined in* (E.2)*. Under the GLM in* (2.1)*, suppose that* (A0)*,* (A2)*,* (T1)*, and* (T2) *hold. Then, there exist constants $M_0 = M_0(\lambda, \tau, S, L, K, R, R^*, \sigma_{\min}, \sigma_{\max}) > 0$ and $C = C(\lambda, \tau, S, L, R, R^*) > 0$ such that, if t_0 is defined by* (4.1) *with $M \geq M_0$, then with probability at least $1 - \delta$, for all $t \in \mathbb{N}$ satisfying $t \geq t_0$,*

$$\left\|\theta_t - \theta^*\right\|_{\mathbf{\Omega}_t}^2 \leq C\left(d \log t + \log^2 t + \log^2 \frac{1}{\delta}\right), \qquad \theta_t \in \Theta_R.$$

Proof. Under (T1), we have $\|X_{t,a}\|_2 \leq L$ almost surely for all $t \in \mathbb{N}$ and $a \in [K]$. Hence (A1) holds almost surely. Next, by Proposition E.1, under (T1) and (T2), with probability at least $1 - \delta/2$, (A3) holds with constants

$$C_1 = C_1(\tau, L, K, R, R^*, \sigma_{\min}, \sigma_{\max}) > 0, \qquad C_2 = C_2(L, K, \sigma_{\min}, \sigma_{\max}) > 0$$

for all

$$t \geq C_2\big(d \log d + \log(2/\delta)\big).$$

Therefore, on this event that (A3) holds, the assumptions of Theorem 4.1 are satisfied under (A0), (A2), (T1) and (T2). Applying Theorem 4.1 with confidence level $\delta/2$ conditionally on this event, we obtain that there exist constants

$$M_0 = M_0(\lambda, \tau, S, L, K, R, R^*, \sigma_{\min}, \sigma_{\max}) > 0, \quad C = C(\lambda, \tau, S, L, R, R^*) > 0$$

such that if t_0 is defined by (4.1) with $M \geq M_0$, then with probability at least $1 - \delta/2$, for all $t \geq t_0$,

$$\|\theta_t - \theta^*\|_{\mathbf{\Omega}_t}^2 \leq C\left(d \log t + \log^2 t + \log^2 \frac{2}{\delta}\right), \qquad \theta_t \in \Theta_R.$$

Finally, a union bound shows that the two events above hold simultaneously with probability at least $1 - \delta$. Absorbing the numerical factor 2 inside the logarithms into the constants C and M_0, we obtain the claim. □

C.1 Induction devices

We conclude this section with two lemmas that formalize the induction argument used in the proof of Theorem C.2. The underlying induction scheme originates from Lee et al. (2026b), and we record here the precise form needed for the present analysis.

The first lemma is quoted directly from Lee et al. (2026b).

Lemma C.3. *Let $(\tau_k)_{k \in \mathbb{N}_0}$ be a sequence in $\mathbb{N}$ such that $\tau_k < \tau_{k+1}$ for all $k \in \mathbb{N}_0$. Let $\{V_t\}_{t \in \mathbb{N}_0}$ be a sequence in $\mathbb{R}_+$ satisfying*

$$V_t \leq V_{t-1} + a_{t-1} V_{t-1}^{3/2} + b_t, \quad \forall t \in \mathbb{N},$$

where $a_t \in \mathbb{R}_+$ and $b_t \in \mathbb{R}$ for all $t \in \mathbb{N}_0$. For $t \in \mathbb{N}$ with $t \geq \tau_0$, let

$$B_t = \left(\max_{\tau_0 \leq s \leq t} \sum_{r=\tau_0+1}^{s} b_r\right) \vee 0, \quad H_t = V_{\tau_0} + B_t.$$

For $k \in \mathbb{N}_0$, let

$$A_k = \sum_{t=\tau_k}^{\tau_{k+1}} a_t, \quad S_k = H_{\tau_{k+1}} + \sum_{t=\tau_0+1}^{\tau_k} a_{t-1} V_{t-1}^{3/2}, \quad p_k > 1.$$

Fix $k \in \mathbb{N}_0$ and assume that

$$A_k \leq \frac{p_k - 1}{p_k^{3/2} \sqrt{S_k \vee 1}}.$$

Then,

$$V_t \leq p_k S_k, \quad \forall t \in \{\tau_k, \tau_k + 1, \ldots, \tau_{k+1}\}.$$

The next lemma adapts the above induction device to our setting. For completeness, we provide a self-contained proof tailored to the present notation and scaling.

Lemma C.4. *Let $\delta \in (0,1]$. Let*

$$\tau_0 = \lceil M(d\log(d \vee 3) + \log^2(1/\delta))\rceil,$$

for some $M > 0$. Let $\tau_* \in \mathbb{N} \cup \{\infty\}$ satisfying $\tau_* > \tau_0$. Let $(V_t)_{t\in\mathbb{N}_0}$ be a non-negative sequence satisfying

$$V_t \le V_{t-1} + a_{t-1}V_{t-1}^{3/2} + b_t, \quad \forall t \in \mathbb{N} \text{ with } \tau_0 < t \le \tau_*,$$

where $a_t \in \mathbb{R}_+$ and $b_t \in \mathbb{R}$ for $\tau_0 \le t \le \tau_*$. Suppose that there exist constants $C_0, C_1, C_2 > 0$ such that for all $t_1, t_2 \in \mathbb{N}$ with $\tau_0 \le t_1 < t_2 \le \tau_*$,

$$\begin{aligned} V_{\tau_0} &\le C_0\left(d + \log\left(\frac{1}{\delta}\right)\right), \\ \sum_{s=t_1}^{t_2} a_s &\le C_1 t_1^{-1/2}, \\ \sum_{s=\tau_0+1}^{t_2} b_s &\le C_2\left(d\log\left(1+\frac{t_2-\tau_0}{d}\right) + \log^2\left(\frac{t_2-\tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right). \end{aligned}$$

Then, there exist constants $M_0 = M_0(C_0, C_1, C_2) > 0$ and $C = C(C_0, C_2) > 0$ such that if $M \ge M_0$, then

$$V_t \le C\left(d\log\left(1+\frac{2t-\tau_0}{d}\right) + \log^2(2t-\tau_0) + \log^2\left(\frac{1}{\delta}\right)\right), \quad \forall t \in \mathbb{N} \text{ with } \tau_0 \le t \le \tau_*.$$

Proof. Let $(\tau_k)_{k\in\mathbb{N}_0}$ be a sequence in $\mathbb{N}$ such that $\tau_k = (2^k\tau_0) \wedge \tau_*$ for all $k \in \mathbb{N}_0$. For $t \in \mathbb{N}$ with $t \ge \tau_0$, let

$$\begin{aligned} B_t &= \left(\max_{\tau_0 \le s \le t} \sum_{r=\tau_0+1}^{s} b_r\right) \vee 0, \quad H_t = V_{\tau_0} + B_t, \\ \mathrm{B}(t) &= C_2\left(d\log\left(1+\frac{t-\tau_0}{d}\right) + \log^2\left(\frac{t-\tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right). \end{aligned}$$

Since $(\mathrm{B}(t))_{t\ge\tau_0}$ are non-decreasing w.r.t. t, it is easy to show that $B_t \le \mathrm{B}(t)$ for any $t \ge \tau_0$. For $k \in \mathbb{N}_0$, let

$$A_k = \sum_{t=\tau_k}^{\tau_{k+1}} a_t, \quad S_k = H_{\tau_{k+1}} + \sum_{t=\tau_0+1}^{\tau_k} a_{t-1}V_{t-1}^{3/2}, \quad p_k = 1 + \frac{2^{-k/4}}{7}, \quad P_k = \prod_{j=0}^{k-1} p_j,$$

where $P_0 = 1$ and $S_0 = H_{\tau_1}$. By the assumption, for any $k \in \mathbb{N}_0$, we have

$$\begin{aligned} A_k &\le C_1\tau_k^{-1/2}, \\ P_k &\le P_\infty := \prod_{j=0}^{\infty} p_j \le 2.5. \end{aligned} \tag{C.15}$$

First, we will show that if

$$A_k \le \frac{p_k - 1}{p_k^{3/2}\sqrt{S_k \vee 1}}, \quad S_k \le P_k H_{\tau_{k+1}} \tag{C.16}$$

for a given $k \in \mathbb{N}_0$, then

$$\begin{aligned} V_t \le p_k S_k \le C\left(d + d\log\left(1+\frac{2t-\tau_0}{d}\right) + \log^2\left(\frac{2t-\tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right), \\ \forall t \in \{\tau_k, \dots, \tau_{k+1}\}, \end{aligned} \tag{C.17}$$

where $C = C(C_0, C_2) > 0$. Suppose that (C.16) holds for a given $k \in \mathbb{N}_0$. Then, by Lemma C.3, we have

$$V_t \le p_k S_k, \quad \forall t \in \{\tau_k, \dots, \tau_{k+1}\}.$$

Also, we have

$$\begin{aligned}
p_k S_k &\le p_k P_k H_{\tau_{k+1}} = P_{k+1} H_{\tau_{k+1}} \\
&\le 2.5(V_{\tau_0} + B_{\tau_{k+1}}) \\
&\le 2.5(V_{\tau_0} + \mathrm{B}(\tau_{k+1})) \\
&\le 2.5(V_{\tau_0} + \mathrm{B}(2t)) \\
&\le 2.5\Bigg(C_0 \left(d + \log\left(\frac{1}{\delta}\right)\right) \\
&\qquad + C_2 \left(d \log\left(1 + \frac{2t - \tau_0}{d}\right) + \log^2\left(\frac{2t-\tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right)\Bigg) \\
&\le C_3 \left(d + d\log\left(1 + \frac{2t-\tau_0}{d}\right) + \log^2\left(\frac{2t-\tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right),
\end{aligned}$$

where the first inequality holds by (C.16), the second inequality holds by (C.15), the fourth inequality holds by the monotonicity of $t \mapsto \mathrm{B}(t)$ and $\tau_{k+1} \le 2t$, the fifth inequality holds by the assumptions, and the last inequality holds by the constant $C_3 = 2.5(C_0 + C_2)$. This completes the proof of (C.17).

Let $k_* = \min\{k \in \mathbb{N} : \tau_* \le 2^k \tau_0\}$ so that $\tau_* = \tau_{k_*}$. Now, we will show that for all $k \in \mathbb{N}_0$ with $k \le k_* - 1$,

$$A_k \le \frac{p_k - 1}{p_k^{3/2}\sqrt{S_k \vee 1}}, \quad S_k \le P_k H_{\tau_{k+1}}. \tag{C.18}$$

Let $k = 0$. Note that

$$S_0 = H_{\tau_1} = P_0 H_{\tau_1}, \tag{C.19}$$

and

$$\begin{aligned}
S_0 \vee 1 &\le H_{\tau_1} + 1 = V_{\tau_0} + B_{\tau_1} + 1 \\
&\le V_{\tau_0} + \mathrm{B}(\tau_1) + 1 \\
&\le C_0 \left(d + \log\left(\frac{1}{\delta}\right)\right) + C_2 \left(d \log\left(1 + \frac{\tau_0}{d}\right) + \log^2\left(\frac{\tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right) + 1 \\
&\le c_1^2 \left(d + d\log\left(1 + \frac{\tau_0}{d}\right) + \log^2\left(\frac{\tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right),
\end{aligned}$$

where the third inequality holds by the assumptions, and the last inequality holds by the constant $c_1 = c_1(C_0, C_2) > 0$. Combining it with (C.16), we have

$$\begin{aligned}
A_0\sqrt{S_0 \vee 1} &\le c_1 C_1 \tau_0^{-1/2} \left(d + d\log\left(1 + \frac{\tau_0}{d}\right) + \log^2\left(\frac{\tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right)^{1/2} \\
&\le c_1 C_1 \left(\sqrt{\frac{d}{\tau_0}} + \sqrt{\frac{d\log\left(1 + \tau_0/d\right)}{\tau_0}} + \sqrt{\frac{\log^2\left(\tau_0/\delta\right)}{\tau_0}} + \sqrt{\frac{\log\left(1/\delta\right)}{\tau_0}}\right) \\
&\le c_1 C_1 \left(M^{-1/2} + M^{-1/4} + c_0 M^{-1/4} + M^{-1/2}\right) \\
&\le c_2 M^{-1/4} \\
&\le c_2 M_0^{-1/4},
\end{aligned}$$

where the third inequality holds by the universal constant $c_0 > 0$, the fourth inequality holds by the constant $c_2 = c_2(c_0, c_1, C_1) = c_2(C_0, C_1, C_2) > 0$, and the last inequality holds for $M \ge M_0$. Note that

$$\frac{p_0 - 1}{p_0^{3/2}} = 7^{-1}(1 + 7^{-1})^{-3/2} \ge 0.1,$$

and then we have

$$A_0\sqrt{S_0 \vee 1} \le c_2 M_0^{-1/4} \le 0.1 \le \frac{p_0 - 1}{p_0^{3/2}}, \tag{C.20}$$

where the second inequality holds by large enough $M_0 = M_0(c_2) = M_0(C_0, C_1, C_2)$. Combining (C.19) and (C.20), (C.18) holds for $k = 0$.

Suppose that (C.18) holds for given $k \in \mathbb{N}_0$ with $k \le k_* - 2$. We will show that

$$A_{k+1} \le \frac{p_{k+1} - 1}{p_{k+1}^{3/2}\sqrt{S_{k+1} \vee 1}}, \quad S_{k+1} \le P_{k+1} H_{\tau_{k+1}}. \tag{C.21}$$

We have

$$\begin{aligned} S_k - H_{\tau_{k+1}} &= \sum_{t=\tau_0+1}^{\tau_k} a_{t-1} V_{t-1}^{3/2}, \\ \sum_{t=\tau_k+1}^{\tau_{k+1}} a_{t-1} V_{t-1}^{3/2} &\le \sum_{t=\tau_k+1}^{\tau_{k+1}} a_{t-1} (p_k S_k)^{3/2} \\ &\le A_k (p_k S_k)^{3/2} \\ &\le (p_k - 1) S_k, \end{aligned} \tag{C.22}$$

where the first inequality holds by (C.17) and the last inequality holds by (C.18) for given k. Then, we have

$$\begin{aligned} S_{k+1} &= H_{\tau_{k+2}} + \sum_{t=\tau_0+1}^{\tau_{k+1}} a_{t-1} V_{t-1}^{3/2} \\ &= H_{\tau_{k+2}} + \sum_{t=\tau_0+1}^{\tau_k} a_{t-1} V_{t-1}^{3/2} + \sum_{t=\tau_k+1}^{\tau_{k+1}} a_{t-1} V_{t-1}^{3/2} \\ &\le H_{\tau_{k+2}} + \left(S_k - H_{\tau_{k+1}}\right) + (p_k - 1) S_k \\ &= p_k S_k + H_{\tau_{k+2}} - H_{\tau_{k+1}}, \end{aligned}$$

where the inequality holds by (C.22). Substituting (C.18) into the last display,

$$\begin{aligned} S_{k+1} &\le p_k \left(P_k H_{\tau_{k+1}}\right) + H_{\tau_{k+2}} - H_{\tau_{k+1}} \\ &\le \left(P_{k+1} - 1\right) H_{\tau_{k+1}} + H_{\tau_{k+2}} \\ &\le \left(P_{k+1} - 1\right) H_{\tau_{k+2}} + H_{\tau_{k+2}} \\ &= P_{k+1} H_{\tau_{k+2}}, \end{aligned} \tag{C.23}$$

where the last inequality holds since $P_j > 1$ for $j \in \mathbb{N}$ and $(H_t)_{t \ge \tau_0}$ is non-decreasing sequence. Thus, it completes the proof of the second assertion in (C.21). Also, note that

$$\begin{aligned} S_{k+1} \vee 1 &\le P_{k+1} H_{\tau_{k+2}} + 1 \\ &\le 2.5 \left(V_{\tau_0} + B_{\tau_{k+2}}\right) + 1 \\ &\le 2.5 \left(V_{\tau_0} + \mathrm{B}(\tau_{k+2})\right) + 1 \\ &\le 2.5 \Bigg(C_0 \left(d + \log\left(\frac{1}{\delta}\right)\right) \\ &\qquad + C_2 \left(d \log\left(1 + \frac{\tau_{k+2} - \tau_0}{d}\right) + \log^2\left(\frac{\tau_{k+2} - \tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right)\Bigg) + 1 \\ &\le c_3^2 \left(d + d \log\left(1 + \frac{\tau_{k+2} - \tau_0}{d}\right) + \log^2\left(\frac{\tau_{k+2} - \tau_0}{\delta}\right) + \log\left(\frac{1}{\delta}\right)\right), \end{aligned}$$

where the first inequality holds by (C.23), the second inequality holds by (C.15), the fourth inequality holds by the assumptions, and the last inequality holds by the constant $c_3 = c_3(C_0, C_2) > 0$.

Combining it with (C.15), we have

$$\begin{aligned}
&A_{k+1}\sqrt{S_{k+1} \vee 1} \\
&\leq c_3 C_1 \tau_{k+1}^{-1/2} \left(d + d \log \left(1 + \frac{\tau_{k+2} - \tau_0}{d} \right) + \log^2 \left(\frac{\tau_{k+2} - \tau_0}{\delta} \right) + \log \left(\frac{1}{\delta} \right) \right)^{1/2} \\
&\leq c_3 C_1 2^{-(k+1)/4} \left(\sqrt{\frac{d}{2^{(k+1)/2}\tau_0}} + \sqrt{\frac{d \log\left(1 + 2^{k+2}\tau_0/d\right)}{2^{(k+1)/2}\tau_0}} \right. \\
&\qquad \left. + \sqrt{\frac{\log^2\left(2^{k+2}\tau_0/\delta\right)}{2^{(k+1)/2}\tau_0}} + \sqrt{\frac{\log(1/\delta)}{2^{(k+1)/2}\tau_0}} \right) \\
&\leq c_3 C_1 2^{-(k+1)/4} \left(M^{-1/2} + \tilde{c}_1 M^{-1/4} + \tilde{c}_2 M^{-1/4} + M^{-1/2} \right) \\
&\leq c_4 2^{-(k+1)/4} M^{-1/4} \\
&\leq c_4 2^{-(k+1)/4} M_0^{-1/4},
\end{aligned}$$

where the second inequality holds by $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, the third inequality holds by the universal constants $\tilde{c}_1, \tilde{c}_2 > 0$, the fourth inequality holds by the constant $c_4 = c_4(c_3, C_1, \tilde{c}_1, \tilde{c}_2) = c_4(C_0, C_1, C_2) > 0$. In addition, note that

$$\frac{p_{k+1} - 1}{p_{k+1}^{3/2}} = \frac{7^{-1} 2^{-(k+1)/4}}{\left(1 + 7^{-1} 2^{-(k+1)/4}\right)^{3/2}} \geq c_5 2^{-(k+1)/4},$$

where $c_5 > 0$ is a universal constant. Combining the last two displays,

$$\begin{aligned}
A_{k+1}\sqrt{S_{k+1} \vee 1} &\leq c_4 2^{-(k+1)/4} M_0^{-1/4} \\
&\leq c_5 2^{-(k+1)/4} \\
&\leq \frac{p_{k+1} - 1}{p_{k+1}^{3/2}},
\end{aligned}$$

where the second inequality holds by large enough $M_0 = M_0(c_4, c_5) = M_0(C_0, C_1, C_2)$. Thus, it completes the proof of the first assertion in (C.21). By induction, it completes the proof of (C.18) for all $k \in \mathbb{N}_0$ with $k \leq k_* - 1$. Combining (C.18) and (C.17), for all $t \in \{\tau_0, \ldots, \tau_*\}$,

$$\begin{aligned}
V_t &\leq C_3 \left(d + d \log \left(1 + \frac{2t - \tau_0}{d} \right) + \log^2 \left(\frac{2t - \tau_0}{\delta} \right) + \log \left(\frac{1}{\delta} \right) \right) \\
&\leq C \left(d \log \left(1 + \frac{2t - \tau_0}{d} \right) + \log^2 (2t - \tau_0) + \log^2 \left(\frac{1}{\delta} \right) \right),
\end{aligned}$$

where $C = C(C_3) = C(C_0, C_2) > 0$. This completes the proof. □

D Proof of Theorem 4.2

Fix $\delta \in (0, 1]$, and let $(\mathcal{F}'_t)_{t \in \mathbb{N}}$ be the filtration defined by $\mathcal{F}'_t = \sigma(\mathcal{X}_1, a_1, r_1, \ldots, \mathcal{X}_t, a_t, r_t, \mathcal{X}_{t+1})$. This filtration is chosen so that sample $\widetilde{\theta}_t$ in the TS algorithm and the action a_{t+1} are $\mathcal{F}'_t$-measurable. Set $\delta_t = 6\delta/\pi^2 t^2$ for all $t \in \mathbb{N}$. Define the event

$$\mathcal{E}_0(\delta) := \left\{ \forall t \geq t_0,\ \|\theta_t - \theta^*\|_{\mathbf{\Omega}_t} \leq \beta_t(\delta),\ \theta_t \in \Theta_R \right\},$$

where

$$\beta_t(\delta) = c_0 \left(d \log t + \log^2 t + \log^2 \frac{1}{\delta} \right)^{1/2},$$

and $c_0 = c_0(\lambda, \tau, S, L, R, R^*) > 0$ is the constant from Theorem 4.1. On $\mathcal{E}_0(\delta)$, the online estimator remains in the local region Θ_R and gives the concentration bound provided by Theorem 4.1. Theorem 4.1 yields

$$\mathbb{P}(\mathcal{E}_0(\delta)) \geq 1 - \delta.$$

For each $t \geq 1$, define

$$\mathcal{E}_{t,G}(\delta_t) := \left\{ \forall a \in [K], \left| x_{t,a}^\top (\widetilde{\theta}_{t-1} - \theta_{t-1}) \right| \leq \sqrt{2 \log \frac{2K}{\delta_t}} \|x_{t,a}\|_{\mathbf{\Omega}_{t-1}^{-1}} \right\}.$$

The event $\mathcal{E}_{t,G}(\delta_t)$ controls the Gaussian Thompson perturbation uniformly over all actions at round t. Thus, we have

$$\mathbb{P}\big(\mathcal{E}_{t,G}(\delta_t) \mid \mathcal{F}'_{t-1}\big) \geq 1 - \delta_t.$$

The proof below is carried out on the intersection of these events.

Proof of Theorem 4.2. We begin with the decomposition

$$\begin{aligned}
\mathrm{Reg}(T) &= \underbrace{\sum_{t=t_0+1}^{T} \big(\mu(x_t^{*\top}\theta^*) - \mu(x_t^\top \theta^*)\big)}_{=:\mathrm{Reg}_{t_0}(T)} + \sum_{t=1}^{t_0} \big(\mu(x_t^{*\top}\theta^*) - \mu(x_t^\top \theta^*)\big) \\
&= \underbrace{\sum_{t=t_0+1}^{T} \big(\mu(x_t^{*\top}\theta^*) - \mu(x_t^\top \theta_{t-1})\big)}_{=:A_1} + \underbrace{\sum_{t=t_0+1}^{T} \big(\mu(x_t^\top \theta_{t-1}) - \mu(x_t^\top \theta^*)\big)}_{=:A_2} \qquad \text{(D.1)} \\
&\quad + \underbrace{\sum_{t=1}^{t_0} \big(\mu(x_t^{*\top}\theta^*) - \mu(x_t^\top \theta^*)\big)}_{=:A_3} \qquad \text{(D.2)}
\end{aligned}$$

We first bound the term A_3. Recall that

$$t_0 = \lceil M(d \log(d \vee 3) + \log^2(1/\delta)) \rceil,$$

where $M > 0$ is sufficiently large depending only on $\lambda, \tau, S, L, R, R^*, C_1, C_2$ by Proposition C.1. Hence

$$\begin{aligned}
A_3 &\leq \sum_{t=1}^{t_0} L_\mu \big| (x_t^* - x_t)^\top \theta^* \big| \\
&\leq \sum_{t=1}^{t_0} 2 L_\mu L R^* \\
&\leq 2 L_\mu L R^* M (d \log(d \vee 3) + \log^2(1/\delta)). \qquad \text{(D.3)}
\end{aligned}$$

We next turn to A_1. We write

$$A_1 = \sum_{t=t_0+1}^{T} \underbrace{\big(\mu(x_t^{*\top}\theta^*) - \mu(x_t^\top \theta_{t-1})\big)}_{=:\mathrm{reg}_{t,1}}. \qquad \text{(D.4)}$$

Fix $t \geq t_0 + 1$, and work on the event $\mathcal{E}_0(\delta) \cap \mathcal{E}_{t,G}(\delta_t)$. Since $\mu(\cdot)$ is increasing, we have

$$\mu(z_1) - \mu(z_2) \leq \sup_{z \in [-B,B]} \mu'(z) \cdot (z_1 - z_2)_+,$$

for any $z_1, z_2 \in [-B, B]$ and some $B > 0$. Recalling $L_\mu = \sup_{x \in \mathcal{X}, \theta \in \Theta_R} \mu'(x^\top \theta)$, it follows that

$$\begin{aligned}
\mathrm{reg}_{t,1} &= \mu(x_t^{*\top}\theta^*) - \mu(x_t^\top \theta_{t-1}) \\
&\leq L_\mu \big(x_t^{*\top}\theta^* - x_t^\top \theta_{t-1}\big)_+ \\
&= L_\mu \big(x_t^{*\top}\theta^* - x_t^{*\top}\widetilde{\theta}_{t-1} + x_t^{*\top}\widetilde{\theta}_{t-1} - x_t^\top \widetilde{\theta}_{t-1} + x_t^\top \widetilde{\theta}_{t-1} - x_t^\top \theta_{t-1}\big)_+ \\
&\leq L_\mu \big(x_t^{*\top}\theta^* - x_t^{*\top}\widetilde{\theta}_{t-1} + x_t^\top \widetilde{\theta}_{t-1} - x_t^\top \theta_{t-1}\big)_+ \\
&\leq L_\mu \big(x_t^{*\top}(\theta^* - \widetilde{\theta}_{t-1})\big)_+ + L_\mu \big(x_t^\top (\widetilde{\theta}_{t-1} - \theta_{t-1})\big)_+, \qquad \text{(D.5)}
\end{aligned}$$

where the first inequality holds since $\|x_t\|_2 \le L$, $\|x_t^*\|_2 \le L$ by (A1) and $\theta_{t-1} \in \Theta_R$ on $\mathcal{E}_0(\delta)$, the second inequality uses

$$x_t^{*\top}\widetilde{\theta}_{t-1} - x_t^\top \widetilde{\theta}_{t-1} \le 0$$

since $x_t^\top \widetilde{\theta}_{t-1} = \max_{a\in[K]} x_{t,a}^\top \widetilde{\theta}_{t-1}$, and the last inequality holds since $(x+y)_+ \le x_+ + y_+$ for any $x, y \in \mathbb{R}$. On the event $\mathcal{E}_0(\delta) \cap \mathcal{E}_{t,G}(\delta_t)$, the first term on the right-hand side of (D.5) is bounded by

$$\begin{aligned}
\left(x_t^{*\top}(\theta^* - \widetilde{\theta}_{t-1})\right)_+ &\le \left|x_t^{*\top}(\theta^* - \widetilde{\theta}_{t-1})\right| \\
&\le \left|x_t^{*\top}(\theta^* - \theta_{t-1})\right| + \left|x_t^{*\top}(\theta_{t-1} - \widetilde{\theta}_{t-1})\right| \\
&\le \left\|x_t^*\right\|_{\mathbf{\Omega}_{t-1}^{-1}} \left\|\theta^* - \theta_{t-1}\right\|_{\mathbf{\Omega}_{t-1}} + \sqrt{2\log\frac{2K}{\delta_t}} \left\|x_t^*\right\|_{\mathbf{\Omega}_{t-1}^{-1}} \\
&\le \left(\beta_{t-1}(\delta) + \sqrt{2\log\frac{2K}{\delta_t}}\right) \left\|x_t^*\right\|_{\mathbf{\Omega}_{t-1}^{-1}},
\end{aligned}$$

Similarly, on $\mathcal{E}_{t,G}(\delta_t)$, the second term on the right-hand side of (D.5) is bounded by

$$\left(x_t^\top(\widetilde{\theta}_{t-1} - \theta_{t-1})\right)_+ \le \left|x_t^\top(\widetilde{\theta}_{t-1} - \theta_{t-1})\right| \le \sqrt{2\log\frac{2K}{\delta_t}} \left\|x_t\right\|_{\mathbf{\Omega}_{t-1}^{-1}}.$$

Substituting these into (D.5), on the event $\mathcal{E}_0(\delta) \cap \mathcal{E}_{t,G}(\delta_t)$, we obtain

$$\mathrm{reg}_{t,1} \le L_\mu \left(\beta_{t-1}(\delta) + \sqrt{2\log\frac{2K}{\delta_t}}\right) \left\|x_t^*\right\|_{\mathbf{\Omega}_{t-1}^{-1}} + L_\mu \sqrt{2\log\frac{2K}{\delta_t}} \left\|x_t\right\|_{\mathbf{\Omega}_{t-1}^{-1}}.$$

Combining this with (D.4), on the event $\mathcal{E}_0(\delta) \cap \left(\bigcap_{t=t_0+1}^T \mathcal{E}_{t,G}(\delta_t)\right)$, we have

$$\begin{aligned}
A_1 &\le \sum_{t=t_0+1}^T L_\mu \left(\beta_{t-1}(\delta) + \sqrt{2\log\frac{2K}{\delta_t}}\right) \left\|x_t^*\right\|_{\mathbf{\Omega}_{t-1}^{-1}} + \sum_{t=t_0+1}^T L_\mu \sqrt{2\log\frac{2K}{\delta_t}} \left\|x_t\right\|_{\mathbf{\Omega}_{t-1}^{-1}} \\
&\le L_\mu \left(\beta_{T-1}(\delta) + \sqrt{2\log(2K/\delta_T)}\right) \sum_{t=t_0+1}^T \left\|x_t^*\right\|_{\mathbf{\Omega}_{t-1}^{-1}} + L_\mu \sqrt{2\log(2K/\delta_T)} \sum_{t=t_0+1}^T \left\|x_t\right\|_{\mathbf{\Omega}_{t-1}^{-1}}
\end{aligned}$$

It remains to control the cumulative normalized design terms $\sum \|x_t^*\|_{\mathbf{\Omega}_{t-1}^{-1}}$ and $\sum \|x_t\|_{\mathbf{\Omega}_{t-1}^{-1}}$, which is done using the linear eigenvalue growth implied by (A3). Using Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\sum_{t=t_0+1}^T \left\|x_t^*\right\|_{\mathbf{\Omega}_{t-1}^{-1}} &\le \sqrt{T}\sqrt{\sum_{t=t_0+1}^T \left\|x_t^*\right\|_{\mathbf{\Omega}_{t-1}^{-1}}^2} \\
&\le \sqrt{T}\sqrt{\sum_{t=t_0+1}^T \frac{\|x_t^*\|_2^2}{\lambda_{\min}(\mathbf{\Omega}_{t-1})}} \\
&\le \sqrt{T}\sqrt{C_1^{-1}L^2 \sum_{t=t_0+1}^T (t-1)^{-1}} \\
&\le C_1^{-1/2} L\sqrt{T(1+\log(T/t_0))},
\end{aligned}$$

where the third inequality holds by (A1) and (A3). Likewise,

$$\sum_{t=t_0+1}^T \left\|x_t\right\|_{\mathbf{\Omega}_{t-1}^{-1}}^2 \le C_1^{-1}L^2(1+\log(T/t_0)), \tag{D.6}$$

and hence

$$\begin{aligned}
\sum_{t=t_0+1}^T \left\|x_t\right\|_{\mathbf{\Omega}_{t-1}^{-1}} &\le \sqrt{T}\sqrt{\sum_{t=t_0+1}^T \left\|x_t\right\|_{\mathbf{\Omega}_{t-1}^{-1}}^2} \\
&\le C_1^{-1/2} L\sqrt{T(1+\log(T/t_0))}.
\end{aligned}$$

Therefore, on the event $\mathcal{E}_0(\delta) \cap \big(\bigcap_{t=t_0+1}^{T} \mathcal{E}_{t,G}(\delta_t)\big)$,

$$A_1 \le C_1^{-1/2} L L_\mu \left(\beta_{T-1}(\delta) + 2\sqrt{2\log(2K/\delta_T)}\right) \sqrt{T(1+\log(T/t_0))}. \tag{D.7}$$

We next consider A_2. Recall that

$$A_2 = \sum_{t=t_0+1}^{T} \underbrace{\big(\mu(x_t^\top \theta_{t-1}) - \mu(x_t^\top \theta^*)\big)}_{=:\mathrm{reg}_{t,2}}.$$

By Taylor's theorem with the integral form of the remainder,

$$\begin{aligned}
\mathrm{reg}_{t,2} &= \mu(x_t^\top \theta_{t-1}) - \mu(x_t^\top \theta^*) \\
&\le \mu'(x_t^\top \theta_{t-1}) \big| x_t^\top (\theta_{t-1} - \theta^*) \big| + \bar{\alpha}(\theta_{t-1}, \theta^*, x_t) \big| x_t^\top (\theta_{t-1} - \theta^*) \big|^2,
\end{aligned}$$

where

$$\bar{\alpha}(\theta_{t-1}, \theta^*, x_t) = \int_0^1 (1-u) \mu''\big(x_t^\top (\theta_{t-1} + u(\theta^* - \theta_{t-1}))\big) \mathrm{d}u.$$

Since $\theta_{t-1} \in \Theta_R$ on $\mathcal{E}_0(\delta)$, the convexity of Θ_R implies

$$\theta_{t-1} + u(\theta^* - \theta_{t-1}) \in \Theta_R, \qquad \forall u \in [0,1].$$

Thus, by (A2),

$$\begin{aligned}
\bar{\alpha}(\theta_{t-1}, \theta^*, x_t) &\le S \int_0^1 (1-u) \mu'\big(x_t^\top (\theta_{t-1} + u(\theta^* - \theta_{t-1}))\big) \mathrm{d}u \\
&\le S L_\mu \int_0^1 (1-u) \mathrm{d}u = \frac{S L_\mu}{2},
\end{aligned}$$

Combining the last three displays, on $\mathcal{E}_0(\delta)$, we have

$$\begin{aligned}
\mathrm{reg}_{t,2} &\le \mu'(x_t^\top \theta_{t-1}) \big| x_t^\top (\theta_{t-1} - \theta^*) \big| + \frac{S L_\mu}{2} \big| x_t^\top (\theta_{t-1} - \theta^*) \big|^2 \\
&\le \mu'(x_t^\top \theta_{t-1}) \big\| x_t \big\|_{\mathbf{\Omega}_{t-1}^{-1}} \big\| \theta_{t-1} - \theta^* \big\|_{\mathbf{\Omega}_{t-1}} + \frac{S L_\mu}{2} \big\| x_t \big\|_{\mathbf{\Omega}_{t-1}^{-1}}^2 \big\| \theta_{t-1} - \theta^* \big\|_{\mathbf{\Omega}_{t-1}}^2 \\
&\le \beta_{t-1}(\delta) \mu'(x_t^\top \theta_{t-1}) \big\| x_t \big\|_{\mathbf{\Omega}_{t-1}^{-1}} + \frac{S L_\mu \beta_{t-1}(\delta)^2}{2} \big\| x_t \big\|_{\mathbf{\Omega}_{t-1}^{-1}}^2,
\end{aligned}$$

Summing over t gives

$$A_2 \le \underbrace{\beta_{T-1}(\delta) \sum_{t=t_0+1}^{T} \mu'(x_t^\top \theta_{t-1}) \big\| x_t \big\|_{\mathbf{\Omega}_{t-1}^{-1}}}_{=:B_1} + \underbrace{\frac{S L_\mu \beta_{T-1}(\delta)^2}{2} \sum_{t=t_0+1}^{T} \big\| x_t \big\|_{\mathbf{\Omega}_{t-1}^{-1}}^2}_{=:B_2}. \tag{D.8}$$

By (D.6), B_2 is directly bounded by

$$B_2 \le \frac{C_1^{-1} S L_\mu L^2 \beta_{T-1}(\delta)^2}{2} (1 + \log(T/t_0)). \tag{D.9}$$

To control B_1, we first compare the local curvature at θ_{t-1} with that at θ^* using the self-concordance property. Using the self-concordance property (A2), we have

$$\mu'(x_t^\top \theta_{t-1}) \le e^{SLR} \mu'(x_t^\top \theta^*),$$

for all $t \ge t_0 + 1$ on $\mathcal{E}_0(\delta)$, and therefore

$$\mu'(x_t^\top \theta_{t-1}) \le e^{SLR/2} \sqrt{\mu'(x_t^\top \theta^*)} \sqrt{\mu'(x_t^\top \theta_{t-1})}.$$

Using Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
&\sum_{t=t_0+1}^{T} \mu'(x_t^\top \theta_{t-1}) \|x_t\|_{\mathbf{\Omega}_{t-1}^{-1}} \\
&\le e^{SLR/2} \sum_{t=t_0+1}^{T} \sqrt{\mu'(x_t^\top \theta^*)} \sqrt{\mu'(x_t^\top \theta_{t-1}) \|x_t\|^2_{\mathbf{\Omega}_{t-1}^{-1}}} \\
&\le g(\tau)^{1/2} e^{SLR/2} \sqrt{\sum_{t=t_0+1}^{T} \mu'(x_t^\top \theta^*)} \sqrt{\sum_{t=t_0+1}^{T} g(\tau)^{-1} \mu'(x_t^\top \theta_{t-1}) \|x_t\|^2_{\mathbf{\Omega}_{t-1}^{-1}}}.
\end{aligned}$$

By Lemma F.4, we have

$$\begin{aligned}
\sum_{t=t_0+1}^{T} \mu'(x_t^\top \theta^*) &\le T\kappa^{*-1} + S \sum_{t=t_0+1}^{T} \big(\mu(x_t^{*\top} \theta^*) - \mu(x_t^\top \theta^*)\big) \\
&= T\kappa^{*-1} + S\,\mathrm{Reg}_{t_0}(T),
\end{aligned}$$

where the last equality holds by the definition of $\mathrm{Reg}_{t_0}(T)$. Moreover, since $\mu'(x_t^\top \theta_{t-1}) \le L_\mu$, (D.6) implies

$$\begin{aligned}
\sum_{t=t_0+1}^{T} g(\tau)^{-1} \mu'(x_t^\top \theta_{t-1}) \|x_t\|^2_{\mathbf{\Omega}_{t-1}^{-1}} &\le g(\tau)^{-1} L_\mu \sum_{t=t_0+1}^{T} \|x_t\|^2_{\mathbf{\Omega}_{t-1}^{-1}} \\
&\le g(\tau)^{-1} C_1^{-1} L_\mu L^2 (1 + \log(T/t_0)).
\end{aligned}$$

Combining the last three displays, we have

$$\begin{aligned}
B_1 &\le g(\tau)^{1/2} e^{SLR/2} \beta_{T-1}(\delta) \sqrt{\sum_{t=t_0+1}^{T} \mu'(x_t^\top \theta^*)} \sqrt{\sum_{t=t_0+1}^{T} g(\tau)^{-1} \mu'(x_t^\top \theta_{t-1}) \|x_t\|^2_{\mathbf{\Omega}_{t-1}^{-1}}} \\
&\le c_1 \beta_{T-1}(\delta) \sqrt{T\kappa^{*-1} + S\,\mathrm{Reg}_{t_0}(T)} \sqrt{1 + \log(T/t_0)} \\
&\le c_1 \beta_{T-1}(\delta) \sqrt{1 + \log(T/t_0)} \left(\sqrt{T\kappa^{*-1}} + \sqrt{S\,\mathrm{Reg}_{t_0}(T)} \right), \qquad \text{(D.10)}
\end{aligned}$$

where $c_1 = \left(e^{SLR} C_1^{-1} L_\mu L^2\right)^{1/2}$. Combining (D.7), (D.8), (D.9) and (D.10), on the event $\mathcal{E}_0(\delta) \cap \left(\bigcap_{t=t_0+1}^{T} \mathcal{E}_{t,G}(\delta_t)\right)$, we have

$$\begin{aligned}
\mathrm{Reg}_{t_0}(T) &= A_1 + A_2 \\
&\le C_1^{-1/2} L L_\mu \left(\beta_{T-1}(\delta) + 2\sqrt{2 \log \frac{2K}{\delta_T}} \right) \sqrt{T \left(1 + \log \frac{T}{t_0}\right)} \\
&\quad \underbrace{+ \frac{C_1^{-1} S L_\mu L^2}{2} \beta_{T-1}(\delta)^2 \left(1 + \log \frac{T}{t_0}\right)}_{=:D_1} \\
&\quad + \underbrace{c_1 \beta_{T-1}(\delta) \sqrt{1 + \log \frac{T}{t_0}}}_{=:D_2} \left(\sqrt{T\kappa^{*-1}} + \sqrt{S\,\mathrm{Reg}_{t_0}(T)} \right) \\
&= D_1 + D_2 \sqrt{T\kappa^{*-1}} + D_2 \sqrt{S\,\mathrm{Reg}_{t_0}(T)}.
\end{aligned}$$

This is of the form $x^2 \le bx + c$ with

$$x^2 = \mathrm{Reg}_{t_0}(T), \quad b = D_2\sqrt{S}, \quad c = D_1 + D_2\sqrt{T\kappa^{*-1}}.$$

Hence, by Lemma F.5,

$$\begin{aligned}\mathrm{Reg}_{t_0}(T) &\le (b+\sqrt{c})^2 \\ &\le 2b^2 + 2c \\ &\le 2SD_2^2 + 2\big(D_1 + D_2\sqrt{T\kappa^{*-1}}\big).\end{aligned}$$

Combining this with (D.2) and (D.3), we conclude that on $\mathcal{E}_0(\delta) \cap \big(\bigcap_{t=t_0+1}^{T} \mathcal{E}_{t,G}(\delta_t)\big)$,

$$\mathrm{Reg}(T) \le 2SD_2^2 + 2\big(D_1 + D_2\sqrt{T\kappa^{*-1}}\big) + 2L_\mu L R^* M(d\log(d\vee 3) + (\log(1/\delta))^2). \quad \text{(D.11)}$$

To simplify this expression, define

$$\Gamma_T(\delta) := \left(d\log T + \log^2 T + \log^2\frac{1}{\delta}\right)^{1/2}.$$

By the definition of $\beta_t(\delta)$, we have

$$\beta_{T-1}(\delta) \le c_0\Gamma_T(\delta).$$

Moreover, since $\delta_T = 6\delta/(\pi^2 T^2)$,

$$\log\frac{2K}{\delta_T} = \log\left(\frac{\pi^2 KT^2}{3\delta}\right) \lesssim \log\left(\frac{KT}{\delta}\right),$$

and we also have

$$1 + \log(T/t_0) \lesssim \log T.$$

Therefore,

$$D_1 \lesssim C_1^{-1/2}\Big(\Gamma_T(\delta) + \sqrt{\log(KT/\delta)}\Big)\sqrt{T\log T} + \Gamma_T(\delta)^2\log T, \quad D_2 \lesssim C_1^{-1/2}\Gamma_T(\delta)\sqrt{\log T}.$$

Since

$$\Gamma_T(\delta) \lesssim \sqrt{d}\log(T/\delta),$$

substituting these bounds into (D.11), on $\mathcal{E}_0(\delta) \cap \big(\bigcap_{t=t_0+1}^{T} \mathcal{E}_{t,G}(\delta_t)\big)$, we have

$$\begin{aligned}\mathrm{Reg}(T) &\lesssim \sqrt{T\log T}\left(C_1^{-1/2}\Gamma_T(\delta) + \sqrt{\log(KT/\delta)} + \kappa^{*-1/2}\Gamma_T(\delta)\right) + C_1^{-1}\Gamma_T(\delta)^2\log T \\ &\lesssim \left((\kappa^{*-1/2} + C_1^{-1/2})\sqrt{d} + \sqrt{\log K}\right)\sqrt{T\log T}\log(T/\delta) + C_1^{-1}d\log T\log^2(T/\delta).\end{aligned} \quad \text{(D.12)}$$

Here the notation $\lesssim$ hides a constant depending only on $\lambda, \tau, S, L, R, R^*$.

Finally, since $\mathbb{P}(\mathcal{E}_0(\delta)) \ge 1-\delta$ and $\sum_{t=1}^{\infty}\delta_t = \delta$, we have

$$\mathbb{P}\left(\mathcal{E}_0(\delta) \cap \bigcap_{t=t_0+1}^{T} \mathcal{E}_{t,G}(\delta_t)\right) \ge 1 - 2\delta.$$

Therefore, (D.12) holds with probability at least $1-2\delta$. This completes the proof. □

E JUSTIFICATIONS FOR CONDITIONS

In this section, we verify the abstract conditions (A3), (B1), (B2) and (B3) used in the concentration argument. Each subsection isolates one condition and shows how it follows from the structural assumptions (A0), (A1) and (A2). We first introduce the quantity

$$\kappa = \kappa(L, R, R^*) = \frac{1}{\inf_{x\in\mathcal{X},\theta\in\Theta_R}\mu'(x^\top\theta)}, \quad \text{(E.1)}$$

which will be used repeatedly to lower bound the local curvature on Θ_R.

E.1 JUSTIFICATION OF (A3)

In this subsection, we verify the eigenvalue condition (A3) under a random design with random contexts $\{X_{t,a}\}_{t\in\mathbb{N},a\in[K]}$. Throughout this section, capital letters (e.g., $X_{t,a}$) denote random variables and lowercase letters (e.g., $x_{t,a}$) denote their realized values.

The argument proceeds in three steps. We first obtain a uniform lower bound on the conditional second moment of each arm under the truncated Gaussian design (Lemma E.1). We then transfer this lower bound to the played context under Thompson sampling (Lemma E.2), and finally combine it with a directional concentration argument to prove the desired eigenvalue growth (Lemma E.4 and Proposition E.1).

E.1.1 TRUNCATED GAUSSIAN DESIGN

We consider a finite-action truncated Gaussian design as follows:

(T1) Let $L > 0$. For each $a \in [K]$, $\{X_{t,a}\}_{t\geq 1}$ are drawn i.i.d. from the truncated Gaussian $\mathcal{N}(0, \mathbf{\Sigma}_a)$ conditioned on $\{\|X_{t,a}\|_2 \leq L\}$, and for each $t \geq 1$, $\{X_{t,a}\}_{a\in[K]}$ are independent.

(T2) $\mathbf{\Sigma}_a \succ 0$ for all $a \in [K]$.

For simplicity of notation, define

$$\sigma^2_{\min} = \min_{a\in[K]} \lambda_{\min}(\mathbf{\Sigma}_a), \qquad \sigma^2_{\max} = \max_{a\in[K]} \lambda_{\max}(\mathbf{\Sigma}_a), \tag{E.2}$$

and denote the played context by $X_t := X_{t,a_t}$. By (T2), $\sigma^2_{\min}$ and $\sigma^2_{\max}$ are finite and positive. Let $(\mathcal{F}''_t)_{t\in\mathbb{N}}$ be a filtration defined by $\mathcal{F}''_t = \sigma(\mathcal{X}_1, a_1, r_1, \ldots, \mathcal{X}_t, a_t, r_t)$.

Remark. Under (T1), we have $\|X_{t,a}\|_2 \leq L$ almost surely for all $t \in \mathbb{N}, a \in [K]$. Hence, for any realization $x_{t,a}$ of $X_{t,a}$, $x_{t,a} \in \{x \in \mathbb{R}^d : \|x\|_2 \leq L\}$, and therefore (A1) holds.

Lemma E.1. *Suppose that* (T1) *and* (T2) *hold. Then, for every $t \geq 1$ and every $a \in [K]$,*

$$\mathbb{E}\left[X_{t,a}X_{t,a}^\top \mid \mathcal{F}''_{t-1}\right] \succeq C\,\mathbf{I}_d \qquad \text{a.s.}$$

where $C = C_1 d^{-1}$ and $C_1 = C_1(L, \sigma_{\min}) > 0$.

Proof. Fix $t \geq 1$ and $a \in [K]$. Under (T1), the random vector $x_{t,a}$ has the law of X_a conditioned on $\{\|X_a\|_2 \leq L\}$, where $X_a \sim \mathcal{N}(0, \mathbf{\Sigma}_a)$, and this conditional law is independent of $\mathcal{F}''_{t-1}$. Hence

$$\mathbb{E}\left[X_{t,a}X_{t,a}^\top \mid \mathcal{F}''_{t-1}\right] = \mathbb{E}\left[X_a X_a^\top \mid \|X_a\|_2 \leq L\right]. \tag{E.3}$$

By Lemma F.8, we have

$$\begin{aligned}
\mathbb{E}\left[X_a X_a^\top \mid \|X_a\|_2 \leq L\right] &\succeq \exp\left(-\frac{L^2}{2\lambda_{\min}(\mathbf{\Sigma}_a)}\right)\frac{L^2}{d+2}\mathbf{I}_d \\
&\succeq \exp\left(-\frac{L^2}{2\sigma^2_{\min}}\right)\frac{L^2}{d+2}\mathbf{I}_d \\
&= C\,\mathbf{I}_d,
\end{aligned}$$

where $C = C_1 d^{-1} > 0$ and $C_1 = C_1(L, \sigma_{\min}) > 0$. This completes the proof. □

Lemma E.2. *Suppose that* (T1) *and* (T2) *hold. At time t, the algorithm samples $\widetilde{\theta}_{t-1}$ conditionally on $\mathcal{F}''_{t-1}$, and chooses the TS action $a_t \in \operatorname{argmax}_{a\in[K]} \mu(X_{t,a}^\top \widetilde{\theta}_{t-1})$. Then, for every $t \geq 1$,*

$$\mathbb{E}\left[X_t X_t^\top \mid \mathcal{F}''_{t-1}\right] \succeq C\,\mathbf{I}_d, \qquad \text{a.s.}$$

where $C = C_1 d^{-1}$ and $C_1 = C_1(K, L, \sigma_{\min}) > 0$.

Proof. Fix $t \geq 1$. Define the enlarged filtration

$$\widetilde{\mathcal{F}}''_{t-1} = \sigma\left(\mathcal{F}''_{t-1}, \widetilde{\theta}_{t-1}\right).$$

By (T1), the context $\{X_{t,a}\}_{a\in[K]}$ are independent of $\mathcal{F}''_{t-1}$ and also independent of $\widetilde{\theta}_{t-1}$ conditional on $\mathcal{F}''_{t-1}$; equivalently, $\{X_{t,a}\}_{a\in[K]}$ is independent of $\widetilde{\mathcal{F}}''_{t-1}$. Using the tower rule, we have

$$\mathbb{E}\left[X_t X_t^\top \mid \mathcal{F}''_{t-1}\right] = \mathbb{E}_{\widetilde{\theta}_{t-1}}\left[\mathbb{E}\left[X_t X_t^\top \mid \widetilde{\mathcal{F}}''_{t-1}\right] \mid \mathcal{F}''_{t-1}\right], \tag{E.4}$$

where $\mathbb{E}_{\widetilde{\theta}_{t-1}}[\cdot \mid \mathcal{F}''_{t-1}]$ denotes the conditional expectation with respect to $\widetilde{\theta}_{t-1}$ given $\mathcal{F}''_{t-1}$. Set the unit vector

$$u_{t-1} = \frac{\widetilde{\theta}_{t-1}}{\|\widetilde{\theta}_{t-1}\|_2}.$$

Since $\mu'(z) > 0$ for all $z \in \mathbb{R}$, the function μ is strictly increasing. Hence, the TS decision is equivalent to

$$a_t \in \operatorname*{argmax}_{a\in[K]} u_{t-1}^\top X_{t,a}.$$

For each $a \in [K]$, define the event

$$\mathcal{E}_{t,a} = \left\{u_{t-1}^\top X_{t,a} > 0\right\} \cap \bigcap_{b\in[K]\setminus\{a\}} \left\{u_{t-1}^\top X_{t,b} < 0\right\}. \tag{E.5}$$

On the event $\mathcal{E}_{t,a}$, we have $u_{t-1}^\top X_{t,a} > 0 > u_{t-1}^\top X_{t,b}$ for all $b \neq a$, hence a is the unique maximizer, and thus

$$\mathcal{E}_{t,a} \subseteq \{a_t = a\}.$$

Therefore, we have

$$X_t X_t^\top = \sum_{a\in[K]} X_{t,a} X_{t,a}^\top \mathbb{1}[a_t = a] \succeq \sum_{a\in[K]} X_{t,a} X_{t,a}^\top \mathbb{1}[\mathcal{E}_{t,a}].$$

Taking conditional expectation given $\widetilde{\mathcal{F}}''_{t-1}$ gives

$$\mathbb{E}\left[X_t X_t^\top \mid \widetilde{\mathcal{F}}''_{t-1}\right] \succeq \sum_{a\in[K]} \mathbb{E}\left[X_{t,a} X_{t,a}^\top \mathbb{1}[\mathcal{E}_{t,a}] \mid \widetilde{\mathcal{F}}''_{t-1}\right]. \tag{E.6}$$

Fix $a \in [K]$. By the definition of $\mathcal{E}_{t,a}$ in (E.5) and the independence across arms in (T1),

$$\begin{aligned}
&\mathbb{E}\left[X_{t,a} X_{t,a}^\top \mathbb{1}[\mathcal{E}_{t,a}] \mid \widetilde{\mathcal{F}}''_{t-1}\right] \\
&\quad = \mathbb{E}\left[X_{t,a} X_{t,a}^\top \mathbb{1}[u_{t-1}^\top X_{t,a} > 0] \mid \widetilde{\mathcal{F}}''_{t-1}\right] \cdot \prod_{b\neq a} \mathbb{P}\left(u_{t-1}^\top X_{t,b} < 0 \mid \widetilde{\mathcal{F}}''_{t-1}\right).
\end{aligned} \tag{E.7}$$

For any arm $b \in [K]$, conditional on $\widetilde{\mathcal{F}}''_{t-1}$, $X_{t,b}$ has a density proportional to $\exp(-x^\top \mathbf{\Sigma}_b^{-1} x/2)$ on the ball $\{\|x\|_2 \le L\}$. This density and support are invariant under $x \mapsto -x$, hence

$$X_{t,b} \overset{d}{=} -X_{t,b} \quad \text{given } \widetilde{\mathcal{F}}''_{t-1}.$$

Then, we have

$$\mathbb{P}\left(u_{t-1}^\top X_{t,b} < 0 \mid \widetilde{\mathcal{F}}''_{t-1}\right) = \mathbb{P}\left(u_{t-1}^\top X_{t,b} > 0 \mid \widetilde{\mathcal{F}}''_{t-1}\right) = \frac{1}{2},$$

and

$$\prod_{b\neq a} \mathbb{P}\left(u_{t-1}^\top X_{t,b} < 0 \mid \widetilde{\mathcal{F}}''_{t-1}\right) = 2^{-(K-1)}. \tag{E.8}$$

Using the symmetry $X_{t,a} \overset{d}{=} -X_{t,a}$, we have

$$\begin{aligned}
\mathbb{E}\left[X_{t,a} X_{t,a}^\top \mathbb{1}[u_{t-1}^\top X_{t,a} > 0] \mid \widetilde{\mathcal{F}}''_{t-1}\right] &= \mathbb{E}\left[(-X_{t,a})(-X_{t,a})^\top \mathbb{1}[u_{t-1}^\top (-X_{t,a}) > 0] \mid \widetilde{\mathcal{F}}''_{t-1}\right] \\
&= \mathbb{E}\left[X_{t,a} X_{t,a}^\top \mathbb{1}[u_{t-1}^\top X_{t,a} < 0] \mid \widetilde{\mathcal{F}}''_{t-1}\right].
\end{aligned}$$

Adding the two equal quantities gives

$$\mathbb{E}\left[X_{t,a}X_{t,a}^\top \mathbb{1}[u_{t-1}^\top X_{t,a} > 0] \mid \widetilde{\mathcal{F}}''_{t-1}\right] = \frac{1}{2}\mathbb{E}\left[X_{t,a}X_{t,a}^\top \mid \widetilde{\mathcal{F}}''_{t-1}\right]. \tag{E.9}$$

Combining (E.7), (E.8) and (E.9), we have

$$\mathbb{E}\left[X_{t,a}X_{t,a}^\top \mathbb{1}[\mathcal{E}_{t,a}] \mid \widetilde{\mathcal{F}}''_{t-1}\right] = 2^{-K}\mathbb{E}\left[X_{t,a}X_{t,a}^\top \mid \widetilde{\mathcal{F}}''_{t-1}\right]. \tag{E.10}$$

Since (E.10) holds for arbitrary $a \in [K]$, combining (E.4), (E.6) and (E.10),

$$\begin{aligned}
\mathbb{E}\left[X_t X_t^\top \mid \mathcal{F}''_{t-1}\right] &= \mathbb{E}_{\widetilde{\theta}_{t-1}}\left[\mathbb{E}\left[X_t X_t^\top \mid \widetilde{\mathcal{F}}''_{t-1}\right] \mid \mathcal{F}''_{t-1}\right] \\
&= \sum_{a\in[K]} \mathbb{E}_{\widetilde{\theta}_{t-1}}\left[\mathbb{E}\left[X_{t,a}X_{t,a}^\top \mathbb{1}[\mathcal{E}_{t,a}] \mid \widetilde{\mathcal{F}}''_{t-1}\right] \mid \mathcal{F}''_{t-1}\right] \\
&= 2^{-K}\sum_{a\in[K]} \mathbb{E}_{\widetilde{\theta}_{t-1}}\left[\mathbb{E}\left[X_{t,a}X_{t,a}^\top \mid \widetilde{\mathcal{F}}''_{t-1}\right] \mid \mathcal{F}''_{t-1}\right] \\
&= 2^{-K}\sum_{a\in[K]} \mathbb{E}\left[X_{t,a}X_{t,a}^\top \mid \mathcal{F}''_{t-1}\right] \\
&\succeq 2^{-K} K c\, \mathbf{I}_d \\
&\succeq C\, \mathbf{I}_d,
\end{aligned}$$

where the first inequality holds by Lemma E.1 with $c = c_1 d^{-1}$ and $c_1 = c_1(L, \sigma_{\min}) > 0$, and the last equality holds by $C = C_1 d^{-1}$ and $C_1 = C_1(K, L, \sigma_{\min}) > 0$. This completes the proof. □

Lemma E.3. *Suppose that* (T1) *and* (T2) *hold. Then, there exist universal constants $C_1, C_2 > 0$ such that for any $u \in \mathbb{S}^{d-1}$ and $t \in \mathbb{N}$,*

$$\mathbb{E}\left[\exp\left(\lambda\Big((u^\top X_t)^2 - \mathbb{E}[(u^\top X_t)^2 \mid \mathcal{F}''_{t-1}]\Big)\right) \mid \mathcal{F}''_{t-1}\right] \le \exp\left(\frac{\nu^2\lambda^2}{2}\right), \qquad \forall|\lambda| \le \frac{1}{b},$$

where

$$\nu = C_1\sigma_{\max}^2\big(1 + \log K + \log(1/p_0)\big), \qquad b = C_2\sigma_{\max}^2\big(1 + \log K + \log(1/p_0)\big),$$

and $p_0 = \min_{a\in[K]} \mathbb{P}(\|Z_a\|_2 \le L)$ with $Z_a \sim \mathcal{N}(0, \mathbf{\Sigma}_a)$.

Proof. Fix $t \in \mathbb{N}$ and a unit vector $u \in \mathbb{S}^{d-1}$. Let

$$Y_t := u^\top X_t, \qquad Y_{t,a} := u^\top X_{t,a}, \qquad \forall a \in [K].$$

Then, for every $r \ge 0$,

$$\{|Y_t| \ge r\} = \{|u^\top X_{t,a_t}| \ge r\} \subseteq \bigcup_{a=1}^K \{|u^\top X_{t,a}| \ge r\}.$$

Using the conditional independence of $\{X_{t,a}\}_{a\in[K]}$ given $\mathcal{F}''_{t-1}$,

$$\begin{aligned}
\mathbb{P}\left(|Y_t| \ge r \mid \mathcal{F}''_{t-1}\right) &\le \sum_{a=1}^K \mathbb{P}\left(|Y_{t,a}| \ge r \mid \mathcal{F}''_{t-1}\right) \\
&= \sum_{a=1}^K \mathbb{P}\left(|Y_{t,a}| \ge r\right).
\end{aligned} \tag{E.11}$$

For each $a \in [K]$, let

$$Z_a \sim \mathcal{N}(0, \mathbf{\Sigma}_a), \quad p_0 := \min_{a\in[K]} \mathbb{P}(\|Z_a\|_2 \le L).$$

Fix $a \in [K]$. Since $X_{t,a}$ is distributed as Z_a conditioned on $\{\|Z_a\|_2 \le L\}$, we have

$$Y_{t,a} \overset{d}{=} u^\top Z_a\{\|Z_a\|_2 \le L\}.$$

Let $s_a^2 := u^\top \mathbf{\Sigma}_a u$, then

$$0 < s_a^2 \le \lambda_{\max}(\mathbf{\Sigma}_a) \le \sigma_{\max}^2.$$

Therefore, for every $r \ge 0$,

$$\begin{aligned}
\mathbb{P}(|Y_{t,a}| \ge r) &= \mathbb{P}\big(|u^\top Z_a| \ge r \mid \|Z_a\|_2 \le L\big) \\
&= \frac{\mathbb{P}\big(|u^\top Z_a| \ge r,\ \|Z_a\|_2 \le L\big)}{\mathbb{P}(\|Z_a\|_2 \le L)} \\
&\le \frac{\mathbb{P}\big(|u^\top Z_a| \ge r\big)}{p_0} \\
&\le \frac{2\exp\big(-r^2/(2s_a^2)\big)}{p_0} \\
&\le 2\exp\left(\log(1/p_0) - \frac{r^2}{2\sigma_{\max}^2}\right).
\end{aligned}$$

Summing over $a \in [K]$ and combining it with (E.11), we obtain

$$\begin{aligned}
\mathbb{P}\big(|Y_t| \ge r \mid \mathcal{F}''_{t-1}\big) &\le \min\left\{1, 2K\exp\left(\log(1/p_0) - \frac{r^2}{2\sigma_{\max}^2}\right)\right\} \\
&= \min\left\{1, \exp\left(\log K + \log(1/p_0) - \frac{r^2}{2\sigma_{\max}^2}\right)\right\}.
\end{aligned}$$

Now set

$$\rho = c_0\sigma_{\max}\sqrt{1 + \log K + \log(1/p_0)}, \tag{E.12}$$

where $c_0 > 0$ is a sufficiently large universal constant. Then the preceding bound implies

$$\mathbb{P}\big(|Y_t| \ge r \mid \mathcal{F}''_{t-1}\big) \le 2\exp\left(-\frac{r^2}{\rho^2}\right), \qquad \forall r \ge 0,$$

and hence, conditionally on $\mathcal{F}''_{t-1}$, the random variable Y_t is $c_1\rho$-sub-Gaussian with a universal constant $c_1 > 0$. Therefore, Lemma F.7 yields

$$\mathbb{E}\left[\exp\left(\lambda\Big((u^\top X_t)^2 - \mathbb{E}[(u^\top X_t)^2 \mid \mathcal{F}''_{t-1}]\Big)\right) \mid \mathcal{F}''_{t-1}\right] \le \exp\left(c_2\lambda^2\rho^4\right), \quad \forall|\lambda| \le c_3/\rho^2,$$

where $c_2, c_3 > 0$ are universal constants. Substituting ρ in (E.12) into this inequality, we have

$$\mathbb{E}\left[\exp\left(\lambda\Big((u^\top X_t)^2 - \mathbb{E}[(u^\top X_t)^2 \mid \mathcal{F}''_{t-1}]\Big)\right) \mid \mathcal{F}''_{t-1}\right] \le \exp\left(\frac{\nu^2\lambda^2}{2}\right), \quad \forall|\lambda| \le \frac{1}{b},$$

with

$$\nu = c_4\sigma_{\max}^2\big(1 + \log K + \log(1/p_0)\big), \qquad b = c_5\sigma_{\max}^2\big(1 + \log K + \log(1/p_0)\big), \tag{E.13}$$

where $c_4, c_5 > 0$ are universal constants. Taking $C_1 = c_4$ and $C_2 = c_5$, this completes the proof. □

Lemma E.4. *Let $\delta \in (0, 1]$. Suppose that* (T1) *and* (T2) *hold. At time t, the algorithm samples $\widetilde{\theta}_{t-1}$ conditionally on $\mathcal{F}''_{t-1}$, and chooses the TS action $a_t \in \operatorname{argmax}_{a\in[K]} \mu(x_{t,a}^\top \widetilde{\theta}_{t-1})$. Then, there exist constants $C_0 = C_0(K, L, \sigma_{\min}) > 0$ and $C_2 = C_2(K, L, \sigma_{\min}, \sigma_{\max}) > 0$ such that with probability at least $1 - \delta$, for all $t \in \mathbb{N}$ with $t \ge C_2(d\log d + \log(1/\delta))$,*

$$\lambda_{\min}\left(\sum_{s=1}^{t} X_s X_s^\top\right) \ge C_1 t, \tag{E.14}$$

where $C_1 = C_0 d^{-1}$.

Proof. For $t \in \mathbb{N}$, let

$$\mathbf{X}_t = \sum_{s=1}^{t} X_s X_s^\top.$$

By Lemma E.2, there exists a constant $c_1 = c_1(K, L, \sigma_{\min}) > 0$ such that for any unit vector $u \in \mathbb{S}^{d-1}$ and $t \in \mathbb{N}$,

$$\mathbb{E}\left[(u^\top X_t)^2 \mid \mathcal{F}''_{t-1}\right] \geq c \qquad \text{a.s.} \tag{E.15}$$

where $c = c_1 d^{-1}$. Moreover, by Lemma E.3, there exist universal constants $c_2, c_3 > 0$ such that for any unit vector $u \in \mathbb{S}^{d-1}$ and $t \in \mathbb{N}$,

$$\mathbb{E}\left[\exp\left(\lambda\Big((u^\top X_t)^2 - \mathbb{E}[(u^\top X_t)^2 \mid \mathcal{F}''_{t-1}]\Big)\right) \mid \mathcal{F}''_{t-1}\right] \leq \exp\left(\frac{\nu^2\lambda^2}{2}\right), \qquad \forall|\lambda| \leq \frac{1}{b}, \tag{E.16}$$

where

$$\nu = c_2\sigma^2_{\max}(1 + \log K + \log(1/p_0)), \quad b = c_3\sigma^2_{\max}(1 + \log K + \log(1/p_0)),$$

and $p_0 = \min_{a\in[K]} \mathbb{P}(\|Z_a\|_2 \leq L)$ with $Z_a \sim \mathcal{N}(0, \mathbf{\Sigma}_a)$.

Fix $\delta \in (0, 1]$ and an integer $t_1 \geq 1$. For each $u \in \mathbb{S}^{d-1}$, define

$$S_t(u) = \sum_{s=1}^{t}(u^\top X_s)^2, \quad M_t(u) = \sum_{s=1}^{t}\left\{(u^\top X_s)^2 - \mathbb{E}[(u^\top X_s)^2 \mid \mathcal{F}_{s-1}]\right\}.$$

Then

$$S_t(u) = \sum_{s=1}^{t}\mathbb{E}[(u^\top X_s)^2 \mid \mathcal{F}_{s-1}] + M_t(u).$$

By (E.15), for $t \in \mathbb{N}$

$$\sum_{s=1}^{t}\mathbb{E}[(u^\top X_s)^2 \mid \mathcal{F}_{s-1}] \geq ct$$

almost surely. Hence

$$\left\{S_t(u) \leq \frac{c}{2}t\right\} \subseteq \left\{M_t(u) \leq -\frac{c}{2}t\right\}. \tag{E.17}$$

We next derive a lower-tail bound for $M_t(u)$. By (E.16), for any $\lambda \in [0, 1/b]$,

$$\mathbb{E}\Big[e^{-\lambda M_t(u)}\Big] = \mathbb{E}\left[\prod_{s=1}^{t}\exp\Big(-\lambda\Big((u^\top X_s)^2 - \mathbb{E}[(u^\top X_s)^2 \mid \mathcal{F}_{s-1}]\Big)\Big)\right] \leq \exp\Big(\frac{t\nu^2\lambda^2}{2}\Big),$$

Therefore, by Markov's inequality,

$$\mathbb{P}\Big(M_t(u) \leq -\frac{c}{2}t\Big) = \mathbb{P}\Big(e^{-\lambda M_t(u)} \geq e^{\lambda ct/2}\Big) \leq \exp\Big(-\frac{\lambda ct}{2} + \frac{t\nu^2\lambda^2}{2}\Big)$$

for every $\lambda \in [0, 1/b]$. Choosing $\lambda^* = \min\{\frac{c}{2\nu^2}, \frac{1}{b}\}$ we obtain

$$\mathbb{P}\Big(M_t(u) \leq -\frac{c}{2}t\Big) \leq e^{-c_4 t},$$

where

$$c_4 = \min\left\{\frac{c^2}{8\nu^2}, \frac{c}{4b}\right\}.$$

Combining this with (E.17), for every $u \in \mathbb{S}^{d-1}$ and $t \in \mathbb{N}$,

$$\mathbb{P}\Big(S_t(u) \leq \frac{c}{2}t\Big) \leq e^{-c_4 t}. \tag{E.18}$$

We now pass from a fixed direction to all directions. Let $\epsilon = \frac{c}{8L^2}$ and let $\mathcal{N}$ be an ϵ-net of $\mathbb{S}^{d-1}$ in Euclidean norm. Since $\|X_t\|_2 \leq L$ almost surely by (T1), we have

$$\|\mathbf{X}_t\|_{\text{op}} \leq \sum_{s=1}^{t}\|X_s X_s^\top\|_{\text{op}} = \sum_{s=1}^{t}\|X_s\|_2^2 \leq tL^2$$

almost surely. Also, by the standard covering-number bound,

$$|\mathcal{N}| \le \left(1+\frac{2}{\epsilon}\right)^d = \left(1+\frac{16L^2}{c}\right)^d. \tag{E.19}$$

For each $t \in \mathbb{N}$, define the event

$$\mathcal{E}_t = \bigcap_{v\in\mathcal{N}} \left\{ v^\top \mathbf{X}_t v \ge \frac{c}{2}t \right\}. \tag{E.20}$$

Let $u \in \mathbb{S}^{d-1}$ be arbitrary, and choose $v \in \mathcal{N}$ such that $\|u-v\|_2 \le \epsilon$. Since $\mathbf{X}_t$ is symmetric,

$$|u^\top \mathbf{X}_t u - v^\top \mathbf{X}_t v| = |(u-v)^\top \mathbf{X}_t u + v^\top \mathbf{X}_t (u-v)| \le 2\|\mathbf{X}_t\|_{\mathrm{op}}\|u-v\|_2 \le 2tL^2\epsilon = \frac{c}{4}t.$$

Therefore, on $\mathcal{E}_t$,

$$u^\top \mathbf{X}_t u \ge v^\top \mathbf{X}_t v - \frac{c}{4}t \ge \frac{c}{4}t.$$

Taking the infimum over all $u \in \mathbb{S}^{d-1}$, on the event $\mathcal{E}_t$,

$$\lambda_{\min}(\mathbf{X}_t) = \inf_{\|u\|_2=1} u^\top \mathbf{X}_t u \ge \frac{c}{4}t. \tag{E.21}$$

It remains to control the failure probability of $\bigcap_{t\ge t_1} \mathcal{E}_t$. By (E.18), (E.19) and (E.20),

$$\mathbb{P}(\mathcal{E}_t^c) \le |\mathcal{N}| e^{-c_4 t} \le \exp\left(d\log\left(1+\frac{16L^2}{c}\right) - c_4 t \right).$$

Hence

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{t=t_1}^{\infty} \mathcal{E}_t^c\right) &\le \sum_{t=t_1}^{\infty} \mathbb{P}(\mathcal{E}_t^c) \\
&\le \exp\left(d\log\left(1+\frac{16L^2}{c}\right)\right) \sum_{t=t_1}^{\infty} e^{-c_4 t} \\
&\le c_5 \exp\left(d\log\left(1+\frac{16L^2}{c}\right) - c_4 t_1 \right),
\end{aligned}$$

where the last inequality holds by $\sum_{t=t_1}^{\infty} e^{-c_4 t} \le c_5 e^{-c_4 t_1}$ with a constant $c_5 = c_5(c_4) > 0$. Therefore, there exists a constant $c_6 = c_6(c_4, c_5) > 0$ such that whenever $t_1 \ge c_6(d\log(1 + 16L^2/c) + \log(1/\delta))$,

$$\mathbb{P}\left(\bigcap_{t=t_1}^{\infty} \mathcal{E}_t\right) \ge 1-\delta.$$

Note that since $c = c_1 d^{-1}$, there exists $c_7 = c_7(c_1, c_6, L) > 0$ such that $c_7(d\log d + \log(1/\delta)) \ge c_6(d\log(1+16L^2/c) + \log(1/\delta))$. Hence, on this event, by (E.21), for all $t \in \mathbb{N}$ with $t \ge c_7(d\log d + \log(1/\delta))$,

$$\lambda_{\min}(\mathbf{X}_t) \ge \frac{c}{4}t = C_1 t,$$

where $C_1 = C_0 d^{-1}$ and $C_0 = c_1/4 = C_0(K, L, \sigma_{\min}) > 0$. Taking $C_2 = c_7 = C_2(K, L, \sigma_{\min}, \sigma_{\max}) > 0$, this completes the proof. □

Proposition E.1. *Let $\delta \in (0,1]$. Suppose that* (A0)*,* (T1) *and* (T2) *hold. Then, there exist constants $C_0 = C_0(\tau, K, L, R, R^*, \sigma_{\min}) > 0$ and $C_2 = C_2(K, L, \sigma_{\min}, \sigma_{\max}) > 0$ such that with probability at least $1-\delta$, for all $t \in \mathbb{N}$ with $t \ge C_2\big(d\log d + \log(1/\delta)\big)$,*

$$\inf_{\{\bar\theta_s\}_{s=1}^t \subset \Theta_R} \lambda_{\min}\left(\sum_{s=1}^{t} \nabla^2 \ell_s(\bar\theta_s)\right) \ge C_1 t,$$

where $C_1 = C_0 d^{-1}$.

Proof. By Lemma E.4, there exist constants $c_0 = c_0(K, L, \sigma_{\min}) > 0$ and $c_2 = c_2(K, L, \sigma_{\min}, \sigma_{\max}) > 0$ such that with probability at least $1 - \delta$, for all integers $t \geq c_2\big(d \log d + \log(1/\delta)\big)$,

$$\lambda_{\min}\left(\sum_{s=1}^{t} X_s X_s^\top\right) \geq c_1 t,$$

where $c_1 = c_0 d^{-1}$. Fix any sequence $\{\bar{\theta}_s\}_{s=1}^t \subset \Theta_R$. Under (T1), we have $\|X_s\|_2 \leq L$ a.s. for all $s \in \mathbb{N}$. Since $\bar{\theta}_s \in \Theta_R$, it follows that

$$\mu'(X_s^\top \bar{\theta}_s) \geq \kappa^{-1} \qquad \text{for all } s \in \mathbb{N}.$$

Using (2.3), we obtain

$$\begin{aligned}
\lambda_{\min}\left(\sum_{s=1}^{t} \nabla^2 \ell_s(\bar{\theta}_s)\right) &= g(\tau)^{-1} \lambda_{\min}\left(\sum_{s=1}^{t} \mu'(X_s^\top \bar{\theta}_s)\, X_s X_s^\top\right) \\
&\geq g(\tau)^{-1} \kappa^{-1} \lambda_{\min}\left(\sum_{s=1}^{t} X_s X_s^\top\right) \\
&\geq g(\tau)^{-1} \kappa^{-1} c_1 t.
\end{aligned}$$

Since the above bound holds uniformly over all sequences $\{\bar{\theta}_s\}_{s=1}^t \subset \Theta_R$, we conclude that

$$\inf_{\{\bar{\theta}_s\}_{s=1}^t \subset \Theta_R} \lambda_{\min}\left(\sum_{s=1}^{t} \nabla^2 \ell_s(\bar{\theta}_s)\right) \geq g(\tau)^{-1} \kappa^{-1} c_1 t.$$

This proves the claim with $C_1 = C_0 d^{-1}$, $C_0 = g(\tau)^{-1}\kappa^{-1} c_0$ and $C_2 = c_2$.

For the final claim, suppose that $L = C_3 \sigma_{\max} \sqrt{d}$ for some $C_3 > 1$. Then, by the second part of Lemma E.4, there exist constants $c_1' = c_1'(C_3, K, \sigma_{\min}, \sigma_{\max}) > 0$ and $c_2' = c_2'(C_3, K, \sigma_{\min}, \sigma_{\max}) > 0$ such that with probability at least $1 - \delta$, for all integers $t \geq c_2'\big(d \log d + \log(1/\delta)\big)$,

$$\lambda_{\min}\left(\sum_{s=1}^{t} X_s X_s^\top\right) \geq c_1' t.$$

Repeating the same argument as above, we obtain

$$\inf_{\{\bar{\theta}_s\}_{s=1}^t \subset \Theta_R} \lambda_{\min}\left(\sum_{s=1}^{t} \nabla^2 \ell_s(\bar{\theta}_s)\right) \geq g(\tau)^{-1} \kappa^{-1} c_1' t.$$

Hence the conclusion holds with $C_1' = g(\tau)^{-1}\kappa^{-1} c_1'$ and $C_2' = c_2'$. This completes the proof. □

E.2 Justification of (B1)

Proposition E.2. *For any $\delta \in (0, 1]$ and any fixed $t_0 \in \mathbb{N}$, there exists an event $\mathcal{E}_1(\delta)$ satisfying $\mathbb{P}(\mathcal{E}_1(\delta)) \geq 1 - \delta$ such that, on $\mathcal{E}_1(\delta)$, for all $t \in \mathbb{N}$ with $t > t_0$,*

$$\sum_{s=t_0+1}^{t} \big(\ell_s(\theta^*) - \ell_s(\theta_{s-1})\big) \leq \log \frac{1}{\delta}.$$

Proof. Let $(\mathcal{F}_t'')_{t \geq 0}$ be a filtration defined by $\mathcal{F}_t'' = \sigma\left(\{\mathcal{X}_s, a_s, r_s\}_{s=1}^t\right)$. Fix $\delta \in (0, 1]$ and $t_0 \in \mathbb{N}$. For $s \geq 1$, define the shifted filtration and shifted losses by

$$\overline{\mathcal{F}}_s'' = \mathcal{F}_{t_0+s}'', \qquad \bar{\ell}_s(\theta) = \ell_{t_0+s}(\theta).$$

Also define, for each $s \geq 0$, a random probability measure $\overline{P}_s$ on Θ by

$$\overline{P}_s = \delta_{\theta_{t_0+s}},$$

the Dirac measure at θ_{t_0+s}. Since θ_{t_0+s} is $\mathcal{F}''_{t_0+s}$-measurable by assumption, $\overline{P}_s$ is $\overline{\mathcal{F}}''_s$-measurable in the sense of Lemma F.9. Applying Lemma F.9 to the sequence $(\overline{\mathcal{F}}''_s, \overline{P}_s, \bar{\ell}_s)_{s\geq 1}$, we obtain an event $\mathcal{E}_1(\delta)$ with $\mathbb{P}\big(\mathcal{E}_1(\delta)\big) \geq 1-\delta$ such that, on $\mathcal{E}_1(\delta)$, for all $m \geq 1$,

$$\sum_{s=1}^{m} \bar{\ell}_s(\theta^*) \leq -\sum_{s=1}^{m} \log\left(\int_{\Theta} e^{-\bar{\ell}_s(\theta)}\, \overline{P}_{s-1}\mathrm{d}\theta\right) + \log\frac{1}{\delta}.$$

Now, since $\overline{P}_{s-1} = \delta_{\theta_{t_0+s-1}}$,

$$\int_{\Theta} e^{-\bar{\ell}_s(\theta)}\, \overline{P}_{s-1}\mathrm{d}\theta = e^{-\bar{\ell}_s(\theta_{t_0+s-1})},$$

and hence

$$-\log\left(\int_{\Theta} e^{-\bar{\ell}_s(\theta)}\, \overline{P}_{s-1}\mathrm{d}\theta\right) = \bar{\ell}_s(\theta_{t_0+s-1}).$$

Therefore, on $\mathcal{E}_1(\delta)$, for all $m \geq 1$,

$$\sum_{s=1}^{m}\Big(\bar{\ell}_s(\theta^*) - \bar{\ell}_s(\theta_{t_0+s-1})\Big) \leq \log\frac{1}{\delta}.$$

Finally, taking $m = t - t_0$ for any $t > t_0$, then we have

$$\sum_{s=t_0+1}^{t} \big(\ell_s(\theta^*) - \ell_s(\theta_{s-1})\big) \leq \log\frac{1}{\delta}.$$

Thus, (B1) holds with $K_1 = 1$. □

E.3 Justification of (B2)

In this subsection, we verify (B2) by separating the stochastic fluctuation of the reward noise from the deterministic estimation error. The main probabilistic step is a self-normalized bound for the weighted squared noise term, which is then combined with the GLM structure of the gradient.

We first state an abstract bound for the weighted squared noise process. The following lemma is formulated for a general adapted noise sequence and will later be specialized to the GLM setting.

Lemma E.5. *Let $\delta \in (0,1]$ and $\lambda > 0$. Suppose that* (A0) *and* (A1) *hold. Let $(\nu_t)_{t\in\mathbb{N}_0}$ be a positive random sequence. Let $(\mathcal{F}_t)_{t\in\mathbb{N}_0}$ be a filtration such that, for each $t \geq 1$, x_t is $\mathcal{F}_{t-1}$-measurable, and r_t, ν_t are $\mathcal{F}_t$-measurable. Let θ_t and $\mathbf{\Omega}_t$ be given according to the online update rule in* (3.2)*, and $\mathbf{\Omega}_0 = \lambda \mathbf{I}_d$. Define*

$$\epsilon_t := r_t - \mathbb{E}[r_t \mid \mathcal{F}_{t-1}].$$

Assume that there exists a constant $\alpha > 0$ such that, for every $t \geq 1$,

$$\mathbb{E}\big[\exp(\gamma\epsilon_t) \mid \mathcal{F}_{t-1}\big] \leq \exp\left(\frac{\gamma^2\nu_{t-1}^2}{2}\right), \qquad \forall\, |\gamma| \leq \frac{1}{\alpha},$$

almost surely. Assume that there exists a deterministic non-decreasing sequence $(\bar{\nu}_t)_{t\in\mathbb{N}_0}$ such that

$$\nu_t \leq \bar{\nu}_t \qquad \textit{almost surely for all } t \geq 0.$$

Then, with probability at least $1-\delta$, for all $t \in \mathbb{N}$ satisfying $t_0 < t \leq t_{\mathtt{exit}}$,

$$\sum_{s=t_0+1}^{t} \|\epsilon_s x_s\|^2_{\mathbf{\Omega}_s^{-1}} \leq C(\bar{\nu}_{t-1}+\alpha)^2 \left(d\log\left(1+\frac{t-t_0}{d}\right) + \log^2\left(\frac{t-t_0}{\delta}\right)\right),$$

where $C = C(\lambda, \tau, L, R, R^*) > 0$.

Proof. For simplicity of notation, define

$$w_s := \|x_s\|^2_{\mathbf{\Omega}_s^{-1}}, \qquad S_t := \sum_{s=t_0+1}^{t} \epsilon_s^2 w_s.$$

Recall the definition of $t_{\texttt{exit}}$ in (C.2):

$$t_{\texttt{exit}} = \inf\{t \in \mathbb{N} : t > t_0, \theta_t \notin \Theta_R\}.$$

Since $(\theta_t)_{t\in\mathbb{N}_0}$ is adapted to $(\mathcal{F}_t)_{t\in\mathbb{N}_0}$, $t_{\texttt{exit}}$ is a stopping time with respect to $(\mathcal{F}_t)_{t\in\mathbb{N}_0}$.

We first derive a deterministic upper bound for $\sum_{s=t_0+1}^{t} w_s$. Fix an integer $t \in \mathbb{N}$ such that $t_0 < t \le t_{\texttt{exit}}$. Since $\{\theta_{s-1}\}_{s=t_0+1}^{t} \subset \Theta_R$, by (E.1) and (A1), $\mu'(x_s^\top \theta_{s-1}) \ge \kappa^{-1} > 0$ for all $s = t_0+1, \dots, t$. Hence, by (3.2) and (2.3), for each $s = t_0+1, \dots, t$,

$$\begin{aligned}
\mathbf{\Omega}_s &= \mathbf{\Omega}_{s-1} + g(\tau)^{-1}\mu'(x_s^\top\theta_{s-1})x_s x_s^\top \\
&\succeq \mathbf{\Omega}_{s-1} + (\kappa g(\tau))^{-1} x_s x_s^\top \\
&\succeq \mathbf{\Omega}_{t_0} + (\kappa g(\tau))^{-1} \sum_{r=t_0+1}^{s} x_r x_r^\top \\
&\succeq (\kappa g(\tau))^{-1}\mathbf{V}_s,
\end{aligned} \tag{E.22}$$

where

$$\mathbf{V}_s := \mathbf{V}_{t_0} + \sum_{r=t_0+1}^{s} x_r x_r^\top, \qquad \mathbf{V}_{t_0} := \kappa g(\tau)\mathbf{\Omega}_{t_0}.$$

Then, we have

$$\begin{aligned}
\sum_{s=t_0+1}^{t} w_s &= \sum_{s=t_0+1}^{t} \|x_s\|^2_{\mathbf{\Omega}_s^{-1}} \\
&\le \kappa g(\tau) \sum_{s=t_0+1}^{t} \|x_s\|^2_{\mathbf{V}_s^{-1}}.
\end{aligned}$$

Note that for $s = t_0, \dots, t$, $\mathbf{V}_s \succeq \kappa g(\tau)\mathbf{\Omega}_0 = \lambda\kappa g(\tau)\mathbf{I}_d$ and $\|x_s\|_2 \le L$ by (A1). By Lemma F.2, we have

$$\begin{aligned}
\sum_{s=t_0+1}^{t} \|x_s\|^2_{\mathbf{V}_s^{-1}} &\le \sum_{s=t_0+1}^{t} \|x_s\|^2_{\mathbf{V}_{s-1}^{-1}} \\
&\le 2\max\left\{1, (\lambda\kappa g(\tau))^{-1}L^2\right\} \log\frac{\det\mathbf{V}_t}{\det\mathbf{V}_{t_0}} \\
&\le c_1 d\,\log\left(1 + \frac{(t-t_0)L^2}{d\,\lambda_{\min}(\mathbf{V}_{t_0})}\right) \\
&\le c_1 d\,\log\left(1 + \frac{(t-t_0)L^2}{d\lambda\kappa g(\tau)}\right),
\end{aligned}$$

where $c_1 = 2\max\left\{1, (\lambda\kappa g(\tau))^{-1}L^2\right\} > 0$, and the third inequality holds by Lemma F.3. Combining the last two displays,

$$\sum_{s=t_0+1}^{t} w_s \le c_2 d \log\left(1 + \frac{(t-t_0)L^2}{d\lambda\kappa g(\tau)}\right) =: W_t, \tag{E.23}$$

where $c_2 = \kappa g(\tau)c_1 > 0$. Furthermore, for $s = t_0+1, \dots, t$, we have

$$\begin{aligned}
\|x_s\|^2_{\mathbf{\Omega}_s^{-1}} &\le \kappa g(\tau)\|x_s\|^2_{\mathbf{V}_s^{-1}} \\
&= \kappa g(\tau) x_s^\top \left(\mathbf{V}_{s-1}^{-1} - \frac{\mathbf{V}_{s-1}^{-1}x_s x_s^\top \mathbf{V}_{s-1}^{-1}}{1 + \|x_s\|^2_{\mathbf{V}_{s-1}^{-1}}}\right) x_s \\
&= \kappa g(\tau)\frac{\|x_s\|^2_{\mathbf{V}_{s-1}^{-1}}}{1 + \|x_s\|^2_{\mathbf{V}_{s-1}^{-1}}} \\
&\le \kappa g(\tau),
\end{aligned}$$

where the first inequality holds by (E.22), and the first equality holds by Sherman-Morrison formula. On the other hand, since $\mathbf{\Omega}_s \succeq \lambda \mathbf{I}_d$, we have

$$\|x_s\|^2_{\mathbf{\Omega}_s^{-1}} \leq \lambda^{-1}\|x_s\|_2^2 \leq \lambda^{-1}L^2,$$

where the last inequality holds by (A1). Combining the last two displays, we conclude that for $s = t_0 + 1, \ldots, t$,

$$w_s \leq \min\{\kappa g(\tau), \lambda^{-1}L^2\} =: c_0. \tag{E.24}$$

Combining (E.23) and (E.24), we obtain

$$\sum_{s=t_0+1}^{t} w_s^2 \leq c_0 W_t. \tag{E.25}$$

We now decompose S_t into a predictable part and a martingale part:

$$S_t = A_t + M_t, \tag{E.26}$$

where

$$A_t := \sum_{s=t_0+1}^{t} w_s \mathbb{E}\big[\epsilon_s^2 \mid \mathcal{F}_{s-1}\big], \quad M_t := \sum_{s=t_0+1}^{t} \xi_s, \quad \xi_s := w_s\big(\epsilon_s^2 - \mathbb{E}\big[\epsilon_s^2 \mid \mathcal{F}_{s-1}\big]\big).$$

To bound the predictable part A_t, for each $s \geq 1$, define

$$H_s(\gamma) := \exp\left(\frac{\gamma^2 \nu_{s-1}^2}{2}\right) - \mathbb{E}\big[e^{\gamma \epsilon_s} \mid \mathcal{F}_{s-1}\big].$$

By assumption, $H_s(\gamma) \geq 0$ for all $|\gamma| \leq 1/\alpha$ a.s. Moreover, $H_s(0) = 0$ and $H_s'(0) = 0$. Hence $H_s''(0) \geq 0$, which implies

$$\mathbb{E}\big[\epsilon_s^2 \mid \mathcal{F}_{s-1}\big] \leq \nu_{s-1}^2 \leq \bar{\nu}_{t-1}^2, \qquad \forall s = t_0 + 1, \ldots, t,$$

where the last inequality follows from the monotonicity of $(\bar{\nu}_s)_{s \in \mathbb{N}_0}$. Combining this with (E.23), we have

$$A_t \leq \bar{\nu}_{t-1}^2 \sum_{s=t_0+1}^{t} w_s \leq \bar{\nu}_{t-1}^2 W_t \tag{E.27}$$

almost surely. Since t is arbitrary, it holds for any $t_0 < t \leq t_{\texttt{exit}}$.

We next turn to the martingale part M_t. For each $s \geq t_0 + 1$, Lemma F.10 implies that ξ_s is conditionally centered and conditionally k_s-sub-Weibull(2) given $\mathcal{F}_{s-1}$, where

$$k_s = c_3 w_s (\bar{\nu}_{s-1} + \alpha)^2,$$

and $c_3 > 0$ is the universal constant. We now define the stopped shifted sequence

$$\begin{aligned}
\xi_j^{(t_{\texttt{exit}})} &:= \xi_{t_0+j}\mathbb{1}[j + t_0 \leq t_{\texttt{exit}}], \\
k_j^{(t_{\texttt{exit}})} &:= k_{t_0+j}\mathbb{1}[j + t_0 \leq t_{\texttt{exit}}], \\
\mathcal{F}_j^{(t_{\texttt{exit}})} &:= \mathcal{F}_{t_0+j}, \qquad j \geq 1.
\end{aligned}$$

By construction, $(\xi_j^{(t_{\texttt{exit}})})_{j \in \mathbb{N}}$ is a martingale difference sequence adapted to $(\mathcal{F}_j^{(t_{\texttt{exit}})})_{j \in \mathbb{N}_0}$. By the convention that an almost surely zero random variable is conditionally 0-sub-Weibull(2), $\xi_j^{(t_{\texttt{exit}})}$ is conditionally $k_j^{(t_{\texttt{exit}})}$-sub-Weibull(2) given $\mathcal{F}_j^{(t_{\texttt{exit}})}$ for every $j \geq 1$. Since $(\bar{\nu}_t)_{t \in \mathbb{N}_0}$ is deterministic and non-decreasing, we have

$$\begin{aligned}
k_j^{(t_{\texttt{exit}})} &\leq c_3 w_{t_0+j}(\bar{\nu}_{t_0+j-1} + \alpha)^2 \\
&\leq c_0 c_3 (\bar{\nu}_{t_0+j-1} + \alpha)^2 =: \bar{k}_j, \qquad j \geq 1,
\end{aligned}$$

where the last inequality holds by (E.24). Moreover, for every $m \in \mathbb{N}$,

$$\begin{aligned}\sum_{j=1}^{m}(k_j^{(t_{\texttt{exit}})})^2 &= \sum_{j=1}^{m\wedge(t_{\texttt{exit}}-t_0-1)} k_{t_0+j}^2 \\ &\le c_3^2(\overline{\nu}_{t_0+m-1}+\alpha)^4 \sum_{j=1}^{m\wedge(t_{\texttt{exit}}-t_0-1)} w_{t_0+j}^2 \\ &\le c_0 c_3^2(\overline{\nu}_{t_0+m-1}+\alpha)^4 W_{t_0+m} \\ &=: \overline{K}_m^2,\end{aligned}$$

where the last inequality holds by (E.25). Note that $(\bar{k}_m)_{m\in\mathbb{N}}$ and $(\overline{K}_m)_{m\in\mathbb{N}}$ are deterministic and non-decreasing. We now apply Lemma F.11. It follows that, with probability at least $1-\delta$,

$$\sum_{j=1}^{m}\xi_j^{(t_{\texttt{exit}})} \le \overline{K}_m\sqrt{2c_4\log\left(\frac{2\pi^2m^2}{3\delta}\right)} + 4\bar{k}_m\log^2\left(\frac{2\pi^2m^2}{3\delta}\right) =: \overline{M}_m, \quad \forall m \ge 1, \tag{E.28}$$

where $c_4 > 0$ is the universal constant.

For $t \in \mathbb{N}$ satisfying $t_0 < t \le t_{\texttt{exit}}$, define the event

$$\mathcal{B}_t(\delta) := \left\{M_t > \overline{M}_{t-t_0}\right\}.$$

By the definition of $\xi_j^{(t_{\texttt{exit}})}$, for $t \in \mathbb{N}$ satisfying $t_0 < t \le t_{\texttt{exit}}$, we have

$$M_t = \sum_{s=t_0+1}^{t}\xi_s = \sum_{j=1}^{t-t_0}\xi_{t_0+j} = \sum_{j=1}^{t-t_0}\xi_j^{(t_{\texttt{exit}})}.$$

Hence, by (E.28), we have

$$\bigcup_{t_0<t\le t_{\texttt{exit}}}\mathcal{B}_t(\delta) = \bigcup_{t_0<t\le t_{\texttt{exit}}}\left\{\sum_{j=1}^{t-t_0}\xi_j^{(t_{\texttt{exit}})} > \overline{M}_{t-t_0}\right\} \le \delta.$$

Therefore, with probability at least $1-\delta$, for all $t \in \mathbb{N}$ with $t_0 < t \le t_{\texttt{exit}}$,

$$M_t \le c_3(\overline{\nu}_{t-1}+\alpha)^2\left(\sqrt{2c_0c_4W_t\log\left(\frac{2\pi^2(t-t_0)^2}{3\delta}\right)} + 4c_0\log^2\left(\frac{2\pi^2(t-t_0)^2}{3\delta}\right)\right). \tag{E.29}$$

Combining (E.26), (E.27) and (E.29), with probability at least $1-\delta$, for all $t \in \mathbb{N}$ satisfying $t_0 < t \le t_{\texttt{exit}}$,

$$\begin{aligned}S_t &\le \overline{\nu}_{t-1}^2 W_t \\ &\quad + c_3(\overline{\nu}_{t-1}+\alpha)^2\left(\sqrt{2c_0c_4W_t\log\left(\frac{2\pi^2(t-t_0)^2}{3\delta}\right)} + 4c_0\log^2\left(\frac{2\pi^2(t-t_0)^2}{3\delta}\right)\right) \\ &\le C(\overline{\nu}_{t-1}+\alpha)^2\left(d\log\left(1+\frac{t-t_0}{d}\right) + \log^2\left(\frac{t-t_0}{\delta}\right)\right),\end{aligned}$$

where $C = C(c_0, c_2, c_3, c_4, \lambda, \kappa, \tau, L) = C(\lambda, \tau, L, R, R^*) > 0$. This completes the proof.

□

We now specialize Lemma E.5 to the GLM setting. For each round $t \in \mathbb{N}$, define

$$\mathcal{F}_t = \sigma(\mathcal{X}_1, a_1, r_1, \ldots, \mathcal{X}_t, a_t, r_t, \mathcal{X}_{t+1}, a_{t+1})$$

so that x_t and a_t are $\mathcal{F}_{t-1}$-measurable. We also define the centered noise

$$\epsilon_t := r_t - \mathbb{E}[r_t \mid \mathcal{F}_{t-1}] = r_t - \mu(x_t^\top\theta^*).$$

The next lemma identifies the conditional tail behavior of the GLM noise. It shows that the centered reward is conditionally sub-exponential with parameters determined by the local curvature $\mu'(x_t^\top\theta^*)$ under (A2).

Lemma E.6. *Suppose that* (A2) *holds. Then, for all* $t \in \mathbb{N}$*, conditionally on* $\mathcal{F}_{t-1}$*,* ϵ_t *is sub-exponential with parameters*

$$\left(\sqrt{2g(\tau)\,\mu'(x_t^\top \theta^*)},\ \frac{Sg(\tau)}{\log 2}\right).$$

Proof. Let $z_t := x_t^\top \theta^*$. By the definition of $\mathcal{F}_{t-1}$, the action a_t and the played context x_{t,a_t} are $\mathcal{F}_{t-1}$-measurable, and hence so is z_t. Conditionally on $\mathcal{F}_{t-1}$, the reward r_t has density from (2.1) as follows

$$p_t(r) = \exp\left(\frac{rz_t - m(z_t)}{g(\tau)} + h(r,\tau)\right) \tag{E.30}$$

with respect to the base measure ν. Since (E.30) is a probability density, for every $z \in \mathbb{R}$,

$$\int \exp\left(\frac{rz - m(z)}{g(\tau)} + h(r,\tau)\right) \mathrm{d}\nu(r) = 1. \tag{E.31}$$

We start with the conditional moment generating function of ϵ_t. Let $s \in \mathbb{R}$ be such that $z_t + g(\tau)s$ belongs to the parameter domain of the model. Using (E.30), we have

$$\begin{aligned}\mathbb{E}\left[e^{sr_t} \mid \mathcal{F}_{t-1}\right] &= \int e^{sr} \exp\left(\frac{rz_t - m(z_t)}{g(\tau)} + h(r,\tau)\right) \mathrm{d}\nu(r) \\ &= \int \exp\left(\frac{r(z_t + g(\tau)s) - m(z_t)}{g(\tau)} + h(r,\tau)\right) \mathrm{d}\nu(r).\end{aligned}$$

Applying (E.31) with $z = z_t + g(\tau)s$, we obtain

$$\mathbb{E}\left[e^{sr_t} \mid \mathcal{F}_{t-1}\right] = \exp\left(\frac{m(z_t + g(\tau)s) - m(z_t)}{g(\tau)}\right).$$

Since $\epsilon_t = r_t - \mu(z_t)$ and $\mu(z_t)$ is $\mathcal{F}_{t-1}$-measurable, it follows that

$$\begin{aligned}\mathbb{E}\left[e^{s\epsilon_t} \mid \mathcal{F}_{t-1}\right] &= e^{-s\mu(z_t)}\,\mathbb{E}\left[e^{sr_t} \mid \mathcal{F}_{t-1}\right] \\ &= \exp\left(\frac{m(z_t + g(\tau)s) - m(z_t) - g(\tau)s\,\mu(z_t)}{g(\tau)}\right).\end{aligned} \tag{E.32}$$

We next bound the exponent in (E.32). Set $u := g(\tau)s$. By Taylor's theorem, there exists a point ξ_t lying between z_t and $z_t + u$ such that

$$m(z_t + u) = m(z_t) + um'(z_t) + \frac{u^2}{2}m''(\xi_t).$$

Since $m' = \mu$ and $m'' = \mu'$, we deduce from (E.32) that

$$\log \mathbb{E}\left[e^{s\epsilon_t} \mid \mathcal{F}_{t-1}\right] = \frac{u^2}{2g(\tau)}m''(\xi_t) = \frac{g(\tau)s^2}{2}\mu'(\xi_t). \tag{E.33}$$

It remains to compare $\mu'(\xi_t)$ with $\mu'(z_t)$. By assumption (A2),

$$|\mu''(z)| \le S\mu'(z) \qquad \forall z \in \mathbb{R}.$$

Since $\mu'(z) > 0$ for all $z \in \mathbb{R}$, we may divide by $\mu'(z)$ and obtain

$$\left|\frac{\mathrm{d}}{\mathrm{d}z}\log \mu'(z)\right| = \left|\frac{\mu''(z)}{\mu'(z)}\right| \le S.$$

Integrating this bound between z_t and ξ_t gives

$$|\log \mu'(\xi_t) - \log \mu'(z_t)| \le S|\xi_t - z_t|.$$

Therefore,

$$\mu'(\xi_t) \le e^{S|\xi_t - z_t|}\mu'(z_t) \le e^{S|u|}\mu'(z_t) \le e^{Sg(\tau)|s|}\mu'(z_t).$$

Hence, for all $|s| \le \log 2/(Sg(\tau))$, we have

$$\mu'(\xi_t) \le 2\mu'(z_t). \tag{E.34}$$

Substituting (E.34) into (E.33), we obtain

$$\mathbb{E}[e^{s\epsilon_t} \mid \mathcal{F}_{t-1}] \le \exp\left(\frac{g(\tau)s^2}{2} \cdot 2\mu'(z_t)\right) = \exp\left(\frac{1}{2}\left(2g(\tau)\mu'(z_t)\right)\right) s^2.$$

Thus, conditionally on $\mathcal{F}_{t-1}$, the random variable ϵ_t is sub-exponential with parameters

$$\left(\sqrt{2g(\tau)\,\mu'(x_t^\top \theta^*)},\ \frac{S\,g(\tau)}{\log 2}\right).$$

This completes the proof. □

Combining the previous lemma with Lemma E.5, we obtain the required control of the weighted squared noise term in the GLM setting.

Lemma E.7. *Let $\delta \in (0,1]$ and $\lambda > 0$. Suppose that* (A0)*,* (A1) *and* (A2) *hold. Let θ_t and $\mathbf{\Omega}_t$ be given according to the online update rule in* (3.2)*, and let $\mathbf{\Omega}_0 = \lambda \mathbf{I}_d$. Then, with probability at least $1-\delta$, for all $t \in \mathbb{N}$ satisfying $t_0 < t \le t_{\texttt{exit}}$,*

$$\sum_{s=t_0+1}^{t} \|\epsilon_s x_s\|^2_{\mathbf{\Omega}_s^{-1}} \le C\left(d \log\left(1 + \frac{t-t_0}{d}\right) + \log^2\left(\frac{t-t_0}{\delta}\right)\right),$$

where $C = C(\lambda, \tau, S, L, R, R^*) > 0$.

Proof. For each $s \in \mathbb{N}$, define

$$\nu_{s-1} := \sqrt{2g(\tau)\mu'(x_s^\top \theta^*)}, \quad \alpha := \frac{Sg(\tau)}{\log 2}.$$

By Lemma E.6, conditionally on $\mathcal{F}_{s-1}$, the noise ϵ_s is sub-exponential with parameters (ν_{s-1}, α). Since $\theta^* \in \Theta_R$, by (A1), $\mu'(x_s^\top \theta^*) \le L_\mu$, hence

$$\nu_{s-1} = \sqrt{2g(\tau)\mu'(x_s^\top \theta^*)} \le \sqrt{2L_\mu g(\tau)} \quad \forall s \in \mathbb{N}.$$

For each $s \in \mathbb{N}$, let $\bar{\nu}_{s-1} := \sqrt{2L_\mu g(\tau)}$, then $(\bar{\nu}_{s-1})_{s\in\mathbb{N}}$ is a deterministic non-decreasing sequence. Therefore, all assumptions of Lemma E.5 are satisfied. Applying Lemma E.5, we obtain that with probability at least $1-\delta$, for all $t \in \mathbb{N}$ satisfying $t_0 < t \le t_{\texttt{exit}}$, we have

$$\begin{aligned}\sum_{s=t_0+1}^{t} \|\epsilon_s x_s\|^2_{\mathbf{\Omega}_s^{-1}} &\le c_2(\bar{\nu}_{t-1} + \alpha)^2 \left(d \log\left(1 + \frac{t-t_0}{d}\right) + \log^2\left(\frac{t-t_0}{\delta}\right)\right) \\ &= c_2\left(\sqrt{2c_1 g(\tau)} + \alpha\right)^2 \left(d \log\left(1 + \frac{t-t_0}{d}\right) + \log^2\left(\frac{t-t_0}{\delta}\right)\right),\end{aligned}$$

where $c_2 = c_2(\lambda, \tau, L, R) > 0$. Setting $C = c_2\left(\sqrt{2L_\mu g(\tau)} + \alpha\right)^2$ completes the proof. □

We now return to the gradient process itself. The following proposition combines the noise bound above with a deterministic control of the estimation error term to verify (B2).

Proposition E.3. *Let $\delta \in (0,1]$ and $\lambda > 0$. Suppose that* (A0)*,* (A1) *and* (A2) *hold. Let θ_t and $\mathbf{\Omega}_t$ be given according to the online update rule in* (3.2)*, and let $\mathbf{\Omega}_0 = \lambda \mathbf{I}_d$. Then, with probability at least $1-\delta$, for all $t \in \mathbb{N}$ satisfying $t_0 < t \le t_{\texttt{exit}}$, we have*

$$\sum_{s=t_0+1}^{t} \|\nabla \ell_s(\theta_{s-1})\|^2_{\mathbf{\Omega}_s^{-1}} \le C\left(d \log\left(1 + \frac{t-t_0}{d}\right) + \log^2\left(\frac{t-t_0}{\delta}\right)\right),$$

where $C = C(\lambda, \tau, S, L, R, R^*) > 0$ is a constant.

Proof. For each $s \in \mathbb{N}$ with $t_0 + 1 \le s \le t$, by (2.3), we have

$$\begin{aligned}\|\nabla \ell_s(\theta_{s-1})\|^2_{\mathbf{\Omega}_s^{-1}} &= \|g(\tau)^{-1}(\mu(x_s^\top \theta_{s-1}) - \mu(x_s^\top \theta^*) + \mu(x_s^\top \theta^*) - r_s)x_s\|^2_{\mathbf{\Omega}_s^{-1}} \\ &\le 2g(\tau)^{-2}\|(\mu(x_s^\top \theta_{s-1}) - \mu(x_s^\top \theta^*))x_s\|^2_{\mathbf{\Omega}_s^{-1}} + 2g(\tau)^{-2}\|\epsilon_s x_s\|^2_{\mathbf{\Omega}_s^{-1}}.\end{aligned}$$

Hence, we have

$$\begin{aligned}&\sum_{s=t_0+1}^{t} \|\nabla \ell_s(\theta_{s-1})\|^2_{\mathbf{\Omega}_s^{-1}} \\ &\le 2g(\tau)^{-2}\left(\sum_{s=t_0+1}^{t} \|(\mu(x_s^\top \theta_{s-1}) - \mu(x_s^\top \theta^*))x_s\|^2_{\mathbf{\Omega}_s^{-1}} + \sum_{s=t_0+1}^{t} \|\epsilon_s x_s\|^2_{\mathbf{\Omega}_s^{-1}}\right).\end{aligned} \tag{E.35}$$

By Lemma (E.7), with probability at least $1-\delta$, for all $t \in \mathbb{N}$ satisfying $t_0 < t \le t_{\texttt{exit}}$, we have

$$\sum_{s=t_0+1}^{t} \|\epsilon_s x_s\|^2_{\mathbf{\Omega}_s^{-1}} \le c_1\left(d\log\left(1 + \frac{t-t_0}{d}\right) + \log^2\left(\frac{t-t_0}{\delta}\right)\right), \tag{E.36}$$

where $c_1 = c_1(\lambda, \tau, S, L, R, R^*) > 0$. By the definition of $t_{\texttt{exit}}$, $\theta_{s-1} \in \Theta_R$ for all $t_0 + 1 \le s \le t$. Hence, by (A1), we have

$$\begin{aligned}\sum_{s=t_0+1}^{t} \|(\mu(x_s^\top \theta_{s-1}) - \mu(x_s^\top \theta^*))x_s\|^2_{\mathbf{\Omega}_s^{-1}} &\le \sum_{s=t_0+1}^{t} \|L_\mu |x_s^\top(\theta_{s-1} - \theta^*)| x_s\|^2_{\mathbf{\Omega}_s^{-1}} \\ &\le L_\mu^2 L^2 R^2 \sum_{s=t_0+1}^{t} \|x_s\|^2_{\mathbf{\Omega}_s^{-1}} \\ &\le \kappa g(\tau) L_\mu^2 L^2 R^2 \sum_{s=t_0+1}^{t} \|x_s\|^2_{\mathbf{V}_s},\end{aligned} \tag{E.37}$$

where the last inequality holds by (E.22) in the proof of Lemma E.5, and

$$\mathbf{V}_s = \mathbf{V}_{t_0} + \sum_{r=t_0+1}^{s} x_r x_r^\top, \quad \mathbf{V}_{t_0} = \kappa g(\tau)\mathbf{\Omega}_{t_0}.$$

Since $\mathbf{V}_s \succeq \kappa g(\tau)\mathbf{\Omega}_0 = \lambda\kappa g(\tau)\mathbf{I}_d$ and $\|x_s\|_2 \le L$ for all $t_0 + 1 \le s \le t$, by Lemma F.2 and Lemma F.3, we have

$$\begin{aligned}\sum_{s=t_0+1}^{t} \|x_s\|^2_{\mathbf{V}_s} &\le \sum_{s=t_0+1}^{t} \|x_s\|^2_{\mathbf{V}_{s-1}} \\ &\le 2\max\{1, (\lambda\kappa g(\tau))^{-1}L^2\} d\log\left(1 + \frac{(t-t_0)L^2}{d\lambda\kappa g(\tau)}\right) \\ &\le c_2 d\log\left(1 + \frac{t-t_0}{d}\right),\end{aligned} \tag{E.38}$$

where $c_2 = c_2(\lambda, \kappa, \tau, L) = c_2(\lambda, \tau, L, R, R^*) > 0$. Combining (E.35), (E.36), (E.37) and (E.38), with probability at least $1-\delta$, for all $t \in \mathbb{N}$ satisfying $t_0 < t \le t_{\texttt{exit}}$, we have

$$\begin{aligned}\sum_{s=t_0+1}^{t} \|\nabla \ell_s(\theta_{s-1})\|^2_{\mathbf{\Omega}_s^{-1}} &\le c_3 d\log\left(1 + \frac{t-t_0}{d}\right) + c_4\left(d\log\left(1 + \frac{t-t_0}{d}\right) + \log^2\left(\frac{t-t_0}{\delta}\right)\right) \\ &\le c_5\left(d\log\left(1 + \frac{t-t_0}{d}\right) + \log^2\left(\frac{t-t_0}{\delta}\right)\right),\end{aligned}$$

where $c_3 = 2\kappa g(\tau)^{-1} L_\mu^2 L^2 R^2 c_2$ and $c_4 = 2g(\tau)^{-2} c_1$, and $c_5 = c_5(c_3, c_4) = c_5(\lambda, \tau, S, L, R, R^*) > 0$. This completes the proof. □

E.4 JUSTIFICATION OF (B3)

In this subsection, we verify (B3) for the batch MAP estimator. The proof has two parts. We first obtain a high probability bound for the score-based quantity $f_t(\widehat{\theta}_t^{\mathrm{MAP}}) - f_t(\theta^*)$, and then combine it with a local invertibility argument to show both localization in Θ_R and the required bound at the phase transition time t_0.

We begin by rewriting the batch MAP estimator in (3.3) under the choices $\mathbf{\Omega}_0 = \lambda \mathbf{I}_d$ and $\theta_0 = 0$:

$$\widehat{\theta}_t^{\mathrm{MAP}} = \operatorname*{argmin}_{\theta \in \Theta} \; L_{1:t}(\theta) + \frac{\lambda}{2}\|\theta\|_2^2, \tag{E.39}$$

where $L_{1:t}(\theta) = \sum_{s=1}^t \ell_s(\theta)$ and $\lambda > 0$ is the initial prior precision parameter. Differentiating the objective function with respect to θ, we obtain that $\widehat{\theta}_t^{\mathrm{MAP}}$ satisfies

$$\sum_{s=1}^t g(\tau)^{-1}\big(\mu(x_s^\top \widehat{\theta}_t^{\mathrm{MAP}})x_s - r_s x_s\big) + \lambda \widehat{\theta}_t^{\mathrm{MAP}} = 0.$$

Define the function $f_t : \Theta \to \mathbb{R}^d$ as

$$f_t(\theta) := g(\tau)^{-1} \sum_{s=1}^t \mu(x_s^\top \theta)x_s + \lambda\theta. \tag{E.40}$$

Then, it follows that

$$f_t(\widehat{\theta}_t^{\mathrm{MAP}}) = g(\tau)^{-1} \sum_{s=1}^t r_s x_s. \tag{E.41}$$

Furthermore, for each $\theta \in \Theta$, we define

$$\begin{aligned} \mathbf{F}_{t,\theta} &:= \nabla^2 L_{1:t}(\theta) + \mathbf{\Omega}_0 \\ &= g(\tau)^{-1} \sum_{s=1}^t \mu'(x_s^\top \theta) x_s x_s^\top + \lambda \mathbf{I}_d. \end{aligned} \tag{E.42}$$

The next lemma provides a uniform concentration bound for the transformed score map $f_t(\widehat{\theta}_t^{\mathrm{MAP}}) - f_t(\theta^*)$ in the Fisher-type norm $\mathbf{F}_{t,\theta^*}^{-1}$.

Lemma E.8. *Let $\delta \in (0,1]$ and $\lambda_1, \lambda_2 > 0$. Let $\widehat{\theta}_t^{\mathrm{MAP}}$ and f_t be defined as in* (E.39) *and* (E.40)*, respectively, with λ_1, and let $\mathbf{F}_{t,\theta^*}$ be defined as in* (E.42) *with λ_2. Suppose that* (A0)*,* (A1) *and* (A2) *hold. Then, with probability at least $1-\delta$, for all $t \in \mathbb{N}$,*

$$\begin{aligned} \big\| f_t(\widehat{\theta}_t^{\mathrm{MAP}}) - f_t(\theta^*) \big\|_{\mathbf{F}_{t,\theta^*}^{-1}} \le \sqrt{\lambda_2}\frac{\log 2}{2SL} + \sqrt{\lambda_1}R^* \\ + \frac{2SL}{\sqrt{\lambda_2}\log 2}\left(d + \frac{d}{2}\log\left(1 + \frac{tL_\mu L^2}{\lambda_2 g(\tau) d}\right) + \log\frac{1}{\delta}\right). \end{aligned}$$

Proof. Let $\epsilon_t = r_t - \mu(x_t^\top \theta^*)$ and $S_t = \sum_{s=1}^t \epsilon_s x_s$ for all $t \in \mathbb{N}$. By (E.41) and the definition of f_t, we have

$$\begin{aligned} f_t(\widehat{\theta}_t^{\mathrm{MAP}}) - f_t(\theta^*) &= g(\tau)^{-1}\sum_{s=1}^t r_s x_s - g(\tau)^{-1}\sum_{s=1}^t \mu(x_s^\top \theta^*)x_s - \lambda\theta^* \\ &= g(\tau)^{-1}\sum_{s=1}^t \epsilon_s x_s - \lambda_1 \theta^* \\ &= g(\tau)^{-1} S_t - \lambda_1 \theta^*. \end{aligned}$$

Hence, we have

$$\left\|f_t(\widehat{\theta}_t^{\mathrm{MAP}}) - f_t(\theta^*)\right\|_{\mathbf{F}_{t,\theta^*}^{-1}} \le g(\tau)^{-1}\left\|S_t\right\|_{\mathbf{F}_{t,\theta^*}^{-1}} + \lambda_1\left\|\theta^*\right\|_{\mathbf{F}_{t,\theta^*}^{-1}} \le g(\tau)^{-1}\left\|S_t\right\|_{\mathbf{F}_{t,\theta^*}^{-1}} + \frac{\lambda_1}{\sqrt{\lambda_2}}R^*, \tag{E.43}$$

where the last inequality holds by $\mathbf{F}_{t,\theta^*} \succeq \lambda_2 \mathbf{I}_d$.

We now verify the conditions required to apply Lemma F.12. By Lemma E.6, conditionally on $\mathcal{F}_{t-1}$, ϵ_t is sub-exponential with parameters $\left(\sqrt{2g(\tau)\mu'(x_t^\top\theta^*)}, Sg(\tau)/\log 2\right)$. Equivalently,

$$\mathbb{E}\left[\exp(\gamma\epsilon_t) \mid \mathcal{F}_{t-1}\right] \le \exp\left(\gamma^2 g(\tau)\mu'(x_t^\top\theta^*)\right) \qquad \forall|\gamma| \le \log 2/(Sg(\tau)),\ t \in \mathbb{N}.$$

Therefore, (F.7) is satisfied with $\nu_{t-1} = g(\tau)\mu'(x_t^\top\theta^*)$ and $M = Sg(\tau)/\log 2$.

Let

$$\overline{\mathbf{F}}_{t,\theta^*} = g(\tau)\sum_{s=1}^{t}\mu'(x_s^\top\theta^*)x_s x_s^\top + g(\tau)^2\lambda_2\mathbf{I}_d.$$

Then, we have

$$\begin{aligned}\mathbf{F}_{t,\theta^*} &= g(\tau)^{-1}\sum_{s=1}^{t}\mu'(x_s^\top\theta^*)x_s x_s^\top + \lambda_2\mathbf{I}_d \\ &= g(\tau)^{-2}\left(g(\tau)\sum_{s=1}^{t}\mu'(x_s^\top\theta^*)x_s x_s^\top + g(\tau)^2\lambda_2\mathbf{I}_d\right) \\ &= g(\tau)^{-2}\overline{\mathbf{F}}_{t,\theta^*}.\end{aligned} \tag{E.44}$$

By Lemma F.12, with probability at least $1-\delta$, for all $t \in \mathbb{N}$,

$$\left\|S_t\right\|_{\overline{\mathbf{F}}_{t,\theta^*}^{-1}} \le \frac{\sqrt{\lambda_2}\log 2}{2SL} + \frac{2SL}{\sqrt{\lambda_2}\log 2}\left(d + \frac{1}{2}\log\frac{\det(\overline{\mathbf{F}}_{t,\theta^*})}{(\lambda_2 g(\tau)^2)^d} + \log\frac{1}{\delta}\right). \tag{E.45}$$

Combining (E.43), (E.44) and (E.45), with probability at least $1-\delta$, for all $t \in \mathbb{N}$,

$$\begin{aligned}\left\|f_t(\widehat{\theta}_t^{\mathrm{MAP}}) - f_t(\theta^*)\right\|_{\mathbf{F}_{t,\theta^*}^{-1}} &\le \left\|S_t\right\|_{\overline{\mathbf{F}}_{t,\theta^*}^{-1}} + \frac{\lambda_1}{\sqrt{\lambda_2}}R^* \\ &\le \sqrt{\lambda_2}\frac{\log 2}{2SL} + \frac{\lambda_1}{\sqrt{\lambda_2}}R^* \\ &\quad + \frac{2SL}{\sqrt{\lambda_2}\log 2}\left(d + \frac{1}{2}\log\frac{\det(\overline{\mathbf{F}}_{t,\theta^*})}{(\lambda_2 g(\tau)^2)^d} + \log\frac{1}{\delta}\right).\end{aligned}$$

By (A1), we have the bound $\mu'(x_s^\top\theta^*) \le L_\mu$ and $\|x_s\|_2 \le L$. By Lemma F.3, we have

$$\frac{\det(\overline{\mathbf{F}}_{t,\theta^*})}{(\lambda_2 g(\tau)^2)^d} \le \left(1 + \frac{tL_\mu L^2}{\lambda_2 g(\tau) d}\right)^d.$$

Substituting it into the above bound, this completes the proof.

□

We now combine the concentration bound above with the eigenvalue growth condition (A3) to verify (B3). The key point is that once f_t is sufficiently well concentrated and $\mathbf{F}_{t,\theta^*}$ is sufficiently non-degenerate, the MAP estimator must remain inside Θ_R and satisfy the required bound at t_0.

Proposition E.4. *Let $\delta \in (0,1]$ and $\lambda > 0$. Suppose that* (A0)*,* (A1)*,* (A2) *and* (A3) *hold. Then, there exists a sufficiently large $M = M(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0$ such that with probability at least $1-\delta$,*

$$\begin{aligned}&t_0 = \lceil M(d\log(d \vee 3) + \log^2(1/\delta))\rceil, \\ &\left\|\widehat{\theta}_{t_0}^{\mathrm{MAP}} - \theta^*\right\|_{\mathbf{\Omega}_{t_0}} \le C\sqrt{d + \log(1/\delta)}, \\ &\widehat{\theta}_t^{\mathrm{MAP}} \in \Theta_R, \qquad \forall t \in \mathbb{N} \text{ with } t \ge t_0,\end{aligned}$$

where $C = C(\lambda, \tau, S, L, R, R^*) > 0$.

Proof. Fix $\delta \in (0,1]$ and $\lambda > 0$. For the localization argument below, introduce an auxiliary initiail prior precision parameter $\bar{\lambda} > 0$, which will be chosen later. Fix $\theta \in \Theta_R$. To avoid confusion with the arbitrary λ in the statement, all Fisher-type matrices appearing in the localization argument below are defined with respect to $\bar{\lambda}$, whereas the batch MAP estimator remains defined with respect to λ.

By Taylor's theorem with the integral form of the remainder, we have

$$\begin{aligned} f_t(\theta) - f_t(\theta^*) &= g(\tau)^{-1} \sum_{s=1}^{t} \big(\mu(x_s^\top \theta) - \mu(x_s^\top \theta^*)\big) x_s + \bar{\lambda}(\theta - \theta^*) \\ &= \mathbf{F}_t(\theta, \theta^*, \bar{\lambda})(\theta - \theta^*), \end{aligned} \tag{E.46}$$

where

$$\mathbf{F}_t(\theta, \theta^*, \bar{\lambda}) := g(\tau)^{-1} \sum_{s=1}^{t} \int_{u=0}^{1} \mu'\big(x_s^\top(\theta^* + u(\theta - \theta^*))\big) \mathrm{d}u \; x_s x_s^\top + \bar{\lambda} \mathbf{I}_d.$$

Define

$$\mathbf{F}_{t,\theta^*}(\bar{\lambda}) := g(\tau)^{-1} \sum_{s=1}^{t} \mu'(x_s^\top \theta^*) x_s x_s^\top + \bar{\lambda} \mathbf{I}_d.$$

Note that $\|x_s\|_2 \le L$ and $\|\theta - \theta^*\|_2 \le R$ by (A1) and $\theta \in \Theta_R$. By Lemma F.1, we have

$$\begin{aligned} \mathbf{F}_t(\theta, \theta^*, \bar{\lambda}) &\succeq (1 + SLR)^{-1} g(\tau)^{-1} \sum_{s=1}^{t} \mu'(x_s^\top \theta^*) x_s x_s^\top + \bar{\lambda} \mathbf{I}_d \\ &\succeq (1 + SLR)^{-1} \mathbf{F}_{t,\theta^*}(\bar{\lambda}). \end{aligned} \tag{E.47}$$

Combining (E.46) and (E.47), we have

$$\begin{aligned} \big\| f_t(\theta) - f_t(\theta^*) \big\|^2_{\mathbf{F}_{t,\theta^*}(\bar{\lambda})^{-1}} &= (\theta - \theta^*)^\top \mathbf{F}_t(\theta, \theta^*, \bar{\lambda}) \mathbf{F}_{t,\theta^*}(\bar{\lambda})^{-1} \mathbf{F}_t(\theta, \theta^*, \bar{\lambda})(\theta - \theta^*) \\ &\ge (1 + SLR)^{-2} (\theta - \theta^*)^\top \mathbf{F}_{t,\theta^*}(\bar{\lambda})(\theta - \theta^*) \\ &\ge (1 + SLR)^{-2} \lambda_{\min}(\mathbf{F}_{t,\theta^*}(\bar{\lambda})) \big\| \theta - \theta^* \big\|_2^2. \end{aligned} \tag{E.48}$$

This implies that

$$\left\{ \theta \in \Theta : \big\| f_t(\theta) - f_t(\theta^*) \big\|_{\mathbf{F}_{t,\theta^*}(\bar{\lambda})^{-1}} \le R(1 + SLR)^{-1} \sqrt{\lambda_{\min}(\mathbf{F}_{t,\theta^*}(\bar{\lambda}))} \right\} \subseteq \Theta_R. \tag{E.49}$$

Define the event

$$\mathcal{E}(\delta, \bar{\lambda}) := \left\{ \forall t \in \mathbb{N},\ \big\| f_t(\widehat{\theta}_t^{\mathrm{MAP}}) - f_t(\theta^*) \big\|_{\mathbf{F}_{t,\theta^*}(\bar{\lambda})^{-1}} \le \gamma_t(\delta, \bar{\lambda}) \right\}, \tag{E.50}$$

where

$$\gamma_t(\delta, \bar{\lambda}) := \sqrt{\bar{\lambda}} \frac{\log 2}{2SL} + \frac{\lambda}{\sqrt{\bar{\lambda}}} R^* + \frac{2SL}{\sqrt{\bar{\lambda}} \log 2} \left(d + \frac{d}{2} \log\left(1 + \frac{t L_\mu L^2}{\bar{\lambda} g(\tau) d} \right) + \log \frac{1}{\delta} \right).$$

By Lemma E.8 with $\lambda_1 = \lambda$ and $\lambda_2 = \bar{\lambda}$, the event $\mathcal{E}(\delta, \bar{\lambda})$ holds with probability at least $1 - \delta$. On the event $\mathcal{E}(\delta, \bar{\lambda})$, by (E.49) and (E.50),

$$\gamma_t(\delta, \bar{\lambda}) \le R(1 + SLR)^{-1} \sqrt{\lambda_{\min}(\mathbf{F}_{t,\theta^*}(\bar{\lambda}))} \tag{E.51}$$

implies that $\widehat{\theta}_t^{\mathrm{MAP}} \in \Theta_R$. In addition, by (A3), there exist constants $C_1, C_2 > 0$ such that, for all $t \in \mathbb{N}$ with $t \ge C_2(d \log d + \log(1/\delta))$,

$$\begin{aligned} \lambda_{\min}(\mathbf{F}_{t,\theta^*}(\bar{\lambda})) &\ge \lambda_{\min}\left(\sum_{s=1}^{t} \nabla^2 \ell_s(\theta^*) \right) + \bar{\lambda} \\ &\ge C_1 t + \bar{\lambda}. \end{aligned} \tag{E.52}$$

Hence, on the event $\mathcal{E}(\delta, \overline{\lambda})$,

$$\gamma_t(\delta, \overline{\lambda}) \le R(1+SLR)^{-1}\sqrt{C_1 t + \overline{\lambda}} = c_1\sqrt{C_1 t + \overline{\lambda}} \tag{E.53}$$

implies (E.51), where $c_1 = R(1+SLR)^{-1} > 0$.

Now, we check the sufficient conditions to verify (E.53). For simplicity, we write

$$\gamma_t(\delta, \overline{\lambda}) = c_2\sqrt{\overline{\lambda}} + \frac{c_3}{\sqrt{\overline{\lambda}}}\left(d + \frac{d}{2}\log\left(1 + \frac{c_4 t}{\overline{\lambda} d}\right) + \log\frac{1}{\delta}\right), \tag{E.54}$$

where

$$c_2 = \frac{\log 2}{2SL}, \quad c_3 = \max\left\{\lambda R^*, \frac{2SL}{\log 2}\right\}, \quad c_4 = \frac{L_\mu L^2}{g(\tau)}.$$

Set

$$D_\delta := d + \log(1/\delta), \quad \overline{D}_\delta := d\log d + \log(1/\delta).$$

Choose $\overline{\lambda} = c_0 D_\delta$, where $c_0 > 0$ is a sufficiently large constant to be specified later, and write $t = M'\overline{D}_\delta$ with $M' \ge 1$. Then, since $D_\delta \ge d$ and $\overline{D}_\delta/(D_\delta d) \le 1$, we have

$$\begin{aligned}
\gamma_t(\delta, \overline{\lambda}) &= c_2\sqrt{c_0 D_\delta} + \frac{c_3}{\sqrt{c_0 D_\delta}}\left(D_\delta + \frac{d}{2}\log\left(1 + \frac{c_4 M'\overline{D}_\delta}{c_0 D_\delta d}\right)\right) \\
&\le \sqrt{D_\delta}\left(c_2\sqrt{c_0} + \frac{c_3}{\sqrt{c_0}}\left(1 + \frac{1}{2}\log\left(1 + \frac{c_4 M'}{c_0}\right)\right)\right) \\
&\le \sqrt{D_\delta}\left(c_2\sqrt{c_0} + \frac{c_3}{\sqrt{c_0}} + \frac{c_3\sqrt{c_4 M'}}{c_0}\right),
\end{aligned}$$

where the last inequality follows from $\log(1+x) \le 2\sqrt{x}$ for all $x \ge 0$. On the other hand,

$$c_1\sqrt{C_1 t + \overline{\lambda}} = c_1\sqrt{D_\delta}\sqrt{C_1 M' \frac{\overline{D}_\delta}{D_\delta} + c_0}.$$

Since $\overline{D}_\delta \ge D_\delta$, we further have

$$c_1\sqrt{C_1 t + \overline{\lambda}} \ge c_1\sqrt{D_\delta}\sqrt{C_1 M' + c_0}.$$

Choose $c_0 = c_0(c_1, c_3, c_4, C_1) = c_0(\lambda, \tau, S, L, R, R^*, C_1) > 0$ sufficiently large so that

$$\frac{c_3\sqrt{c_4}}{c_0} \le \frac{c_1\sqrt{C_1}}{2},$$

explicitly,

$$c_0 = \frac{2c_3\sqrt{c_4}}{c_1\sqrt{C_1}} + \lambda e^{-SLR}. \tag{E.55}$$

Then, we have

$$\frac{c_3\sqrt{c_4}}{c_0}\sqrt{M'} \le \frac{c_1\sqrt{C_1}}{2}\sqrt{M'} \le \frac{c_1}{2}\sqrt{C_1 M' + c_0}.$$

Combining the last four displays, if

$$c_2\sqrt{c_0} + \frac{c_3}{\sqrt{c_0}} \le \frac{c_1\sqrt{C_1}}{2}\sqrt{M'}, \tag{E.56}$$

then (E.53) follows. Hence, there exists a constant

$$M_0 = M_0(c_0, c_1, c_2, c_3, C_1, C_2) = M_0(\lambda, \tau, S, L, R, R^*, C_1, C_2) > 0,$$

explicitly,

$$M_0 = \frac{4}{c_1^2 C_1}\left(c_2^2 c_0 + \frac{c_3^2}{c_0} + 2c_2c_3\right) + 1,$$

such that $t = M'\overline{D}_\delta$ with $M' \geq M_0$ implies both (E.56) and $t \geq C_2(d\log d + \log(1/\delta))$. Therefore, for all $t \in \mathbb{N}$ with $t \geq M_0(d\log d + \log(1/\delta))$, on the event $\mathcal{E}(\delta, \bar{\lambda})$, (E.51) holds. Combining (E.49), (E.50) and (E.51), with probability at least $1-\delta$, we have

$$\widehat{\theta}_t^{\texttt{MAP}} \in \Theta_R, \qquad \forall t \geq M_0(d\log d + \log(1/\delta)). \tag{E.57}$$

The previous argument shows that, with high probability, the MAP estimator remains in the local region Θ_R after a sufficiently large time.

It remains to prove the bound on $\|\widehat{\theta}_{t_0}^{\texttt{MAP}} - \theta^*\|_{\mathbf{\Omega}_{t_0}}$. Recall the definition of t_0:

$$t_0 = \lceil MH_\delta \rceil,$$

where $H_\delta = d\log(d \vee 3) + \log^2(1/\delta)$. By (E.57), there exists a universal constant $\tilde{c}_0 \geq 1$ such that, if $M \geq \tilde{c}_0 M_0$, then with probability at least $1 - \delta/3$,

$$\widehat{\theta}_t^{\texttt{MAP}} \in \Theta_R, \qquad \forall t \geq t_0. \tag{E.58}$$

Let $\mathcal{E}_1'(\delta)$ denote the event in (E.58). Then, on the event $\mathcal{E}_1'(\delta)$, $\widehat{\theta}_{t_0}^{\texttt{MAP}} \in \Theta_R$. From the immediate result in (E.48), we have

$$\begin{aligned}\big\|\widehat{\theta}_{t_0}^{\texttt{MAP}} - \theta^*\big\|_{\mathbf{F}_{t_0,\theta^*}(\lambda)} &\leq (1+SLR)\big\|f_{t_0}(\widehat{\theta}_{t_0}^{\texttt{MAP}}) - f_{t_0}(\theta^*)\big\|_{\mathbf{F}_{t_0,\theta^*}(\lambda)^{-1}} \\ &\leq (1+SLR)\left(g(\tau)^{-1}\big\|S_{t_0}\big\|_{\mathbf{F}_{t_0,\theta^*}(\lambda)^{-1}} + \sqrt{\lambda}R^*\right),\end{aligned} \tag{E.59}$$

where the last inequality holds from (E.43) in the proof of Lemma E.8.

Now compare $\widetilde{\mathbf{F}}_{t_0,\theta^*}$ with $\mathbf{\Omega}_{t_0}$. Recall that

$$\mathbf{\Omega}_{t_0} = g(\tau)^{-1}\sum_{s=1}^{t_0}\mu'(x_s^\top \widehat{\theta}_{t_0}^{\texttt{MAP}})x_s x_s^\top + \lambda \mathbf{I}_d.$$

Since $\widehat{\theta}_{t_0}^{\texttt{MAP}} \in \Theta_R$ on $\mathcal{E}_1'(\delta)$, we have $|x_s^\top(\widehat{\theta}_{t_0}^{\texttt{MAP}} - \theta^*)| \leq LR$. Hence, by (A2), we have

$$\mu'(x_s^\top\theta^*) \geq e^{-SLR}\mu'(x_s^\top\widehat{\theta}_{t_0}^{\texttt{MAP}}), \qquad s \in [t_0],$$

which implies

$$\mathbf{F}_{t_0,\theta^*}(\lambda) \succeq e^{-SLR}\mathbf{\Omega}_{t_0}.$$

Combining this with (E.59), we have

$$\big\|\widehat{\theta}_{t_0}^{\texttt{MAP}} - \theta^*\big\|_{\mathbf{\Omega}_{t_0}} \leq e^{SLR/2}(1+SLR)\left(g(\tau)^{-1}\big\|S_{t_0}\big\|_{\mathbf{F}_{t_0,\theta^*}(\lambda)^{-1}} + \sqrt{\lambda}R^*\right). \tag{E.60}$$

We next apply Lemma F.14. By Lemma E.6, conditionally on $\mathcal{F}_{s-1}$, ϵ_s is sub-exponential with parameters $\big(\sqrt{2g(\tau)\mu'(x_s^\top\theta^*)}, Sg(\tau)/\log 2\big)$. Equivalently,

$$\mathbb{E}\big[\exp(\gamma\epsilon_s) \mid \mathcal{F}_{s-1}\big] \leq \exp\left(\gamma^2 g(\tau)\mu'(x_s^\top\theta^*)\right) \qquad \forall|\gamma| \leq \log 2/(Sg(\tau)),\ s \in \mathbb{N}.$$

Thus Lemma F.14 applies with

$$\nu_{s-1} = g(\tau)\mu'(x_s^\top\theta^*), \qquad \overline{\mathbf{\Omega}}_t = \sum_{s=1}^{t}\nu_{s-1}x_s x_s^\top + \lambda'\mathbf{I}_d,$$

where $\lambda' = g(\tau)^2\lambda$. Since

$$\begin{aligned}\overline{\mathbf{\Omega}}_{t_0} &= g(\tau)\sum_{s=1}^{t_0}\mu'(x_s^\top\theta^*)x_s x_s^\top + g(\tau)^2\lambda\mathbf{I}_d \\ &= g(\tau)^2\mathbf{F}_{t_0,\theta^*}(\lambda),\end{aligned}$$

for any vector $v \in \mathbb{R}^d$,

$$\|v\|_{\overline{\mathbf{\Omega}}_{t_0}} = g(\tau)^{-1}\|v\|_{\mathbf{F}_{t_0,\theta^*}(\lambda)}.$$

By Lemma F.14, with probability at least $1-\delta/3$,

$$\|S_{t_0}\|_{\mathbf{F}_{t_0,\theta^*}(\lambda)^{-1}} \le 4g(\tau)\sqrt{x_\delta} + \frac{4Sg(\tau)}{\log 2}\,\xi_{t_0}\,x_\delta, \tag{E.61}$$

where

$$x_\delta = d\log 6 + \log(6/\delta), \qquad \xi_{t_0} = \max_{1\le s\le t_0}\|x_s\|_{\mathbf{F}_{t_0,\theta^*}(\lambda)^{-1}}.$$

Let $\mathcal{E}_2'(\delta)$ denote the event in (E.61). To control the term ξ_{t_0}, use (A3). There exists an event $\mathcal{E}_3'(\delta)$ with probability at least $1-\delta/3$ such that for all $t \ge C_2(d\log d + \log(3/\delta))$,

$$\lambda_{\min}\left(g(\tau)^{-1}\sum_{s=1}^{t}\mu'(x_s^\top\theta^*)x_s x_s^\top\right) \ge C_1 t.$$

Hence, on $\mathcal{E}_3'(\delta)$,

$$\lambda_{\min}(\mathbf{F}_{t_0,\theta^*}(\lambda)) \ge C_1 t_0 + \lambda \ge C_1 M H_\delta + \lambda,$$

provided M is chosen so that $t_0 \ge C_2(d\log d + \log(3/\delta))$. Therefore,

$$\xi_{t_0} \le \frac{L}{\sqrt{\lambda_{\min}(\mathbf{F}_{t_0,\theta^*}(\lambda))}} \le \frac{L}{\sqrt{C_1 M H_\delta + \lambda}}.$$

Since

$$x_\delta \le c_5 D_\delta$$

for a universal constant $c_5 > 0$, and $H_\delta \ge D_\delta$, we get

$$\xi_{t_0}\,x_\delta \le \frac{c_5 L}{\sqrt{C_1 M}}\sqrt{D_\delta}. \tag{E.62}$$

Choose

$$M_1 = \frac{c_5 S^2 L^2}{C_1 \log^2 2}.$$

Combining the last two displays, if $M \ge M_1$, then we have

$$\frac{4Sg(\tau)}{\log 2}\,\widetilde{\xi}_{t_0}\,x_\delta \le 4g(\tau)\sqrt{c_5 D_\delta}.$$

Substituting this into (E.61), we obtain that on $\mathcal{E}_2'(\delta)\cap\mathcal{E}_3'(\delta)$,

$$\|S_{t_0}\|_{\mathbf{F}_{t_0,\theta^*}(\lambda)^{-1}} \le 8g(\tau)\sqrt{c_5 D_\delta}.$$

Combining this with (E.60), we obtain on $\mathcal{E}_1'(\delta)\cap\mathcal{E}_2'(\delta)\cap\mathcal{E}_3'(\delta)$,

$$\begin{aligned}
\|\widehat{\theta}_{t_0}^{\mathrm{MAP}} - \theta^*\|_{\mathbf{\Omega}_{t_0}} &\le e^{SLR/2}(1+SLR)\left(g(\tau)^{-1}\big\|S_{t_0}\big\|_{\mathbf{F}_{t_0,\theta^*}(\lambda)^{-1}} + \sqrt{\lambda}R^*\right)\\
&\le e^{SLR/2}(1+SLR)\left(8\sqrt{c_5} + \sqrt{\lambda}\,R^*\right)\sqrt{D_\delta}\\
&= C\sqrt{d+\log(1/\delta)},
\end{aligned} \tag{E.63}$$

where $C = C(\lambda,\tau,S,L,R,R^*) > 0$. Therefore, if $M \ge \max\{\tilde{c}_0 M_0, M_1\}$, on $\mathcal{E}_1'(\delta)\cap\mathcal{E}_2'(\delta)\cap\mathcal{E}_3'(\delta)$, the above inequality holds. Note that $\mathcal{E}_1'(\delta)\cap\mathcal{E}_2'(\delta)\cap\mathcal{E}_3'(\delta)$ has probability at least $1-\delta$ by the union bound, and on this event both (E.58) and (E.63) hold. This completes the proof.

□

F TECHNICAL LEMMAS

This section collects the auxiliary lemmas used in the verification arguments of Section E. For the reader's convenience, we first state several generic tools that are used repeatedly throughout the proofs, and then group the remaining lemmas according to the verification in which they are used.

F.1 LEMMAS FOR MAIN RESULTS

This subsection collects standard probabilistic and linear-algebraic tools that are used repeatedly in the proofs of the verification results.

Lemma F.1 (Self-concordance control. Lemma 8 of Zhang et al. (2025)). *Suppose that* (A2) *holds. Then, for any $z_1, z_2 \in \mathbb{R}$ and $z \in \{z_1, z_2\}$, we have*

$$\int_0^1 \mu'(z_1 + u(z_2 - z_1))du \geq \frac{\mu'(z)}{1 + S|z_1 - z_2|},$$

and

$$\frac{\mu'(z)}{2 + S|z_1 - z_2|} \leq \int_{u=0}^1 (1-u)\mu'(z_1 + u(z_2 - z_1))du \leq \exp(S^2|z_1 - z_2|^2) \cdot \mu'(z).$$

Lemma F.2 (Elliptical potential lemma. Lemma 15 of Faury et al. (2020)). *Let $L > 0$. Let $\{a_t\}_{t=1}^{\infty}$ be a sequence in $\{a \in \mathbb{R}^d : \|a\|_2 \leq L\}$ and let $\mathbf{V}_0 \succ 0$. Define $\mathbf{V}_t = \mathbf{V}_0 + \sum_{s=1}^t a_s a_s^\top$ for each $t \in \mathbb{N}$. Then, for all $t \in \mathbb{N}$,*

$$\sum_{s=1}^t \|a_s\|^2_{\mathbf{V}_{s-1}^{-1}} \leq 2\max\left\{1, \frac{L^2}{\lambda_{\min}(\mathbf{V}_0)}\right\} \log \frac{\det(\mathbf{V}_t)}{\det(\mathbf{V}_0)}.$$

Lemma F.3 (Determinant-Trace inequality. Lemma 16 of Faury et al. (2020)). *Let $L > 0$. Let $(a_s)_{s\in\mathbb{N}}$ be a sequence in $\{a \in \mathbb{R}^d : \|a\|_2 \leq L\}$ and let $\mathbf{V}_0 \succ 0$. Define $\mathbf{V}_t := \mathbf{V}_0 + \sum_{s=1}^t a_s a_s^\top$ for $t \in \mathbb{N}$. Then, for $t \in \mathbb{N}$,*

$$\frac{\det(\mathbf{V}_t)}{\det(\mathbf{V}_0)} \leq \left(1 + \frac{tL^2}{d\lambda_{\min}(\mathbf{V}_0)}\right)^d.$$

Lemma F.4 (Claim 14 in Janz et al. (2024)). *Let $\tau, T \in \mathbb{N}$ with $\tau < T$. Suppose that* (A2) *holds. Then,*

$$\sum_{t=\tau+1}^T \mu'(x_t^\top \theta^*) \leq T\kappa^{*-1} + S\sum_{t=\tau+1}^T \big(\mu(x_t^{*\top}\theta^*) - \mu(x_t^\top \theta^*)\big).$$

Lemma F.5 (Claim 6 in Janz et al. (2024)). *Let $x \in \mathbb{R}$ and $b, c \geq 0$. If $x^2 \leq bx + c$, then $x \leq b + \sqrt{c}$.*

F.2 LEMMAS FOR SECTION E.1

The lemmas in this subsection provide the main technical ingredients for establishing the eigenvalue growth condition under the truncated Gaussian design.

Lemma F.6 (Exponential inequality for chi-square distributions. Laurent & Massart (2000)). *Let X be a χ^2 statistic with d degrees of freedom. Then, for any $u > 0$,*

$$\mathbb{P}\left(X \geq d + 2\sqrt{du} + 2u\right) \leq e^{-u}.$$

Lemma F.7 (Square of a sub-Gaussian variable). *Let Y be a mean zero R-sub-Gaussian random variable for some $R > 0$. Then there exist universal constants $C_1, C_2 > 0$ such that*

$$\mathbb{E}\big[\exp\big(\lambda\big(Y^2 - \mathbb{E}[Y^2]\big)\big)\big] \leq \exp\big(C_1\lambda^2 R^4\big) \qquad \forall|\lambda| \leq C_2/R^2$$

Proof. Here and below, the notation $\lesssim$ hides only universal constants. Since Y is sub-Gaussian, by Lemma 2.8.5 in Vershynin (2026), Y^2 is sub-exponential and

$$\|Y^2\|_{\psi_1} \lesssim \|Y\|^2_{\psi_2} \lesssim R^2.$$

Hence

$$\|Y^2 - \mathbb{E}[Y^2]\|_{\psi_1} \lesssim R^2.$$

Applying the standard mgf bound for a centered sub-exponential random variable yields

$$\mathbb{E}\left[\exp\big(\lambda(Y^2 - \mathbb{E}[Y^2])\big)\right] \leq \exp(C_1\lambda^2 R^4) \qquad \forall|\lambda| \leq C_2/R^2,$$

for some universal constants $C_1, C_2 > 0$; see Proposition 2.8.1 in Vershynin (2026). This completes the proof.

□

Lemma F.8. *Let $L > 0$ and let $X \sim \mathcal{N}(0, \mathbf{\Sigma})$, where $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ is positive definite. Then*

$$\mathbb{E}\big[XX^\top \mid \|X\|_2 \le L\big] \succeq \exp\left(-\frac{L^2}{2\lambda_{\min}(\mathbf{\Sigma})}\right)\frac{L^2}{d+2}\mathbf{I}_d.$$

Proof. The density of X is

$$\phi_{\mathbf{\Sigma}}(x) = \frac{1}{(2\pi)^{d/2}\det(\mathbf{\Sigma})^{1/2}}\exp\left(-\frac{1}{2}x^\top\mathbf{\Sigma}^{-1}x\right).$$

Hence, the conditional density of X given $\|X\|_2 \le L$ is

$$p_L(x) = \frac{\phi_{\mathbf{\Sigma}}(x)\cdot\mathbb{1}[\|x\|_2 \le L]}{\mathbb{P}(\{\|X\|_2 \le L\})}.$$

Let $\mathcal{B}_L^d := \{x \in \mathbb{R}^d : \|x\|_2 \le L\}$. For any $x \in \mathcal{B}_L^d$, note that

$$x^\top\mathbf{\Sigma}^{-1}x \le \lambda_{\max}(\mathbf{\Sigma}^{-1})\|x\|_2^2 = \frac{\|x\|_2^2}{\lambda_{\min}(\mathbf{\Sigma})} \le \frac{L^2}{\lambda_{\min}(\mathbf{\Sigma})}.$$

Hence, for $x \in \mathcal{B}_L^d$, we have

$$\phi_{\mathbf{\Sigma}}(x) \ge \phi_{\mathbf{\Sigma}}(0)\exp\left(-\frac{L^2}{2\lambda_{\min}(\mathbf{\Sigma})}\right).$$

On the other hand, since $\phi_{\mathbf{\Sigma}}(z) \le \phi_{\mathbf{\Sigma}}(0)$ for all $z \in \mathbb{R}^d$, we have

$$\mathbb{P}(\|X\|_2 \le L) = \int_{\mathcal{B}_L^d}\phi_{\mathbf{\Sigma}}(z)\,\mathrm{d}z \le \phi_{\mathbf{\Sigma}}(0)\mathrm{vol}(\mathcal{B}_L^d),$$

where $\mathrm{vol}(S) := \int_S \mathrm{d}x$ is the volume of a set $S \subset \mathbb{R}^d$. Therefore, for every $x \in \mathcal{B}_L^d$,

$$p_L(x) \ge \exp\left(-\frac{L^2}{2\lambda_{\min}(\mathbf{\Sigma})}\right)\mathrm{vol}(\mathcal{B}_L^d)^{-1}.$$

It follows that

$$\begin{aligned}\mathbb{E}[XX^\top \mid \|X\|_2 \le L] &= \int_{\mathcal{B}_L^d} xx^\top p_L(x)\,\mathrm{d}x \\ &\succeq \exp\left(-\frac{L^2}{2\lambda_{\min}(\mathbf{\Sigma})}\right)\mathrm{vol}(\mathcal{B}_L^d)^{-1}\int_{\mathcal{B}_L^d} xx^\top\,\mathrm{d}x.\end{aligned}$$

By rotational symmetry of the Euclidean ball,

$$\begin{aligned}\int_{\mathcal{B}_L^d} xx^\top\,\mathrm{d}x &= \frac{1}{d}\left(\int_{\mathcal{B}_L^d}\|x\|_2^2\,\mathrm{d}x\right)\mathbf{I}_d \\ &= \frac{1}{d}\cdot\frac{d}{d+2}L^2\mathrm{vol}(\mathcal{B}_L^d)\mathbf{I}_d \\ &= \frac{L^2}{d+2}\mathrm{vol}(\mathcal{B}_L^d)\mathbf{I}_d.\end{aligned}$$

Substituting this into the previous display gives

$$\mathbb{E}[XX^\top \mid \|X\|_2 \le L] \succeq \exp\left(-\frac{L^2}{2\lambda_{\min}(\mathbf{\Sigma})}\right)\frac{L^2}{d+2}\mathbf{I}_d.$$

This completes the proof. □

F.3 Lemmas for Section E.2

Lemma F.9 (Lemma 5 of Zhang et al. (2025)). *Let $(\mathcal{F}_t'')_{t\geq 0}$ be a filtration defined by $\mathcal{F}_t'' = \sigma\left(\{\mathcal{X}_s, a_s, r_s\}_{s=1}^t\right)$. For each $t \geq 0$, let $P_t(\cdot)$ be an $\mathcal{F}_t''$-measurable random probability measure on the parameter space Θ, meaning that for every measurable $A \subseteq \Theta$, the map $P_t(A)$ is $\mathcal{F}_t''$-measurable. Let $\ell_t(\cdot)$ be the loss function defined by (2.2). For each $t \geq 1$, define*

$$L_t(\theta^*) = \sum_{s=1}^{t} \ell_s(\theta^*) \quad \textit{and} \quad F_t = -\sum_{s=1}^{t} \log\left(\int_\Theta e^{-\ell_s(\theta)} P_{s-1}(d\theta)\right).$$

Then, for any $\delta \in (0,1]$, we have

$$\mathbb{P}\left(\forall t \geq 1, L_t(\theta^*) \leq F_t + \log\left(\frac{1}{\delta}\right)\right) \geq 1 - \delta.$$

F.4 Lemmas for Section E.3

The lemmas in this subsection are used to control the self-normalized gradient fluctuation term appearing in the proof of Proposition E.3.

Definition F.1 (Sub-Weibull). A random variable X is K-sub-Weibull(θ) with parameters $K > 0$ and $\theta > 0$ if

$$\mathbb{E}\left[\exp\left((|X|/K)^{1/\theta}\right)\right] \leq 2.$$

We adopt the convention that if $X = 0$ almost surely, then X is said to be 0-sub-Weibull(θ).

Lemma F.10. *Let $(\mathcal{F}_t)_{t\in\mathbb{N}_0}$ be a filtration. Fix $s \in \mathbb{N}$. Suppose that*

(i) w_s is a non-negative $\mathcal{F}_{s-1}$-measurable random variable,

(ii) ϵ_s is $\mathcal{F}_s$-measurable and satisfies

$$\mathbb{E}[e^{\gamma\epsilon_s} \mid \mathcal{F}_{s-1}] \leq \exp\left(\frac{\gamma^2\nu_{s-1}^2}{2}\right), \qquad \forall\, |\gamma| \leq \frac{1}{\alpha},$$

almost surely, for some $\alpha > 0$,

(iii) ν_{s-1} is a non-negative $\mathcal{F}_{s-1}$-measurable and satisfies $\nu_{s-1} \leq \bar{\nu}_{s-1}$ almost surely, where $\bar{\nu}_{s-1}$ is deterministic.

Define

$$\xi_s := w_s\left(\epsilon_s^2 - \mathbb{E}\left[\epsilon_s^2 \mid \mathcal{F}_{s-1}\right]\right).$$

Then, ξ_s is conditionally centered, and ξ_s is conditionally k_s-sub-Weibull(2) given $\mathcal{F}_{s-1}$, where $k_s = Cw_s(\bar{\nu}_{s-1} + \alpha)^2$ and $C > 0$ is a universal constant.

Proof.

$$k_s := 16\rho^2 w_s \beta_s^2, \qquad \rho := \frac{\log 3}{\log 2},$$

We first prove the conditional centering. Since w_s is $\mathcal{F}_{s-1}$-measurable,

$$\begin{aligned}\mathbb{E}[\xi_s \mid \mathcal{F}_{s-1}] &= w_s\mathbb{E}\left[\epsilon_s^2 - \mathbb{E}\left[\epsilon_s^2 \mid \mathcal{F}_{s-1}\right] \mid \mathcal{F}_{s-1}\right] \\ &= 0\end{aligned}$$

almost surely. This proves the conditional centering.

We now prove that ξ_s is conditionally sub-Weibull(2). Define

$$F_s(\gamma) := \mathbb{E}\left[e^{\gamma\epsilon_s} \mid \mathcal{F}_{s-1}\right], \quad G_s(\gamma) := e^{\gamma^2\nu_{s-1}^2/2}, \quad H_s(\gamma) := G_s(\gamma) - F_s(\gamma).$$

By assumption,

$$F_s(\gamma) \leq G_s(\gamma) \quad \forall |\gamma| \leq 1/\alpha. \tag{F.1}$$

Moreover, $F_s(0) = G_s(0) = 1$ and $F_s'(0) = \mathbb{E}[\epsilon_s \mid \mathcal{F}_{s-1}] = 0 = G_s'(0)$ Hence, $H_s(\gamma)$ satisfies $H_s(\gamma) \geq 0$ in a neighborhood of 0, with $H_s(0) = 0$ and $H_s'(0) = 0$. Therefore, $H_s''(0) \geq 0$, which implies

$$\mathbb{E}[\epsilon_s^2 \mid \mathcal{F}_{s-1}] \leq \nu_{s-1}^2 \leq \bar{\nu}_{s-1}^2$$

almost surely. Using the preceding inequality,

$$\begin{aligned}\big|\xi_s\big| &\leq w_s\big(\epsilon_s^2 + \mathbb{E}\big[\epsilon_s^2 \mid \mathcal{F}_{s-1}\big]\big) \\ &\leq w_s\big(\epsilon_s^2 + \bar{\nu}_{s-1}^2\big).\end{aligned}$$

Now define

$$k_s' := 16 w_s \beta_s^2, \quad \beta_s := \bar{\nu}_{s-1} + \alpha.$$

Then, we have

$$\sqrt{\frac{|\xi_s|}{k_s'}} \leq \frac{\sqrt{\epsilon_s^2 + \bar{\nu}_{s-1}^2}}{4\beta_s} \leq \frac{|\epsilon_s| + \bar{\nu}_{s-1}}{4\beta_s} \leq \frac{|\epsilon_s|}{4\beta_s} + \frac{1}{4},$$

where the second inequality holds by the elementary inequality $\sqrt{a^2 + b^2} \leq a + b$ for $a, b \geq 0$, and the last inequality holds since $\beta_s \geq \bar{\nu}_{s-1}$. It follows that

$$\begin{aligned}\mathbb{E}\big[\exp\big(\sqrt{|\xi_s|/k_s'}\big) \mid \mathcal{F}_{s-1}\big] &\leq e^{1/4}\mathbb{E}\big[e^{|\epsilon_s|/(4\beta_s)} \mid \mathcal{F}_{s-1}\big] \\ &\leq e^{1/4}\mathbb{E}\big[e^{\epsilon_s/(4\beta_s)} + e^{-\epsilon_s/(4\beta_s)} \mid \mathcal{F}_{s-1}\big] \\ &\leq 2e^{1/4} e^{\nu_{s-1}^2/(32\beta_s^2)} \\ &\leq 2e^{1/4} e^{\bar{\nu}_{s-1}^2/(32\beta_s^2)} \\ &\leq 2e^{1/4+1/32} \\ &< 3,\end{aligned}$$

where the third inequality holds by (F.1) since $1/(4\beta_s) \leq 1/\alpha$, the fourth inequality holds since $\nu_{s-1} \leq \bar{\nu}_{s-1}$, and the fifth inequality holds since $\beta_s \geq \bar{\nu}_{s-1}$. Set

$$\rho := \frac{\log 3}{\log 2}, \quad k_s := \rho^2 k_s' = 16\rho^2 w_s \beta_s^2.$$

Since $1/\rho \in (0,1)$, the map $x \mapsto x^{1/\rho}$ is concave on $\mathbb{R}_+$, and conditional Jensen's inequality gives

$$\begin{aligned}\mathbb{E}\big[\exp\big(\sqrt{|\xi_s|/k_s}\big) \mid \mathcal{F}_{s-1}\big] &= \mathbb{E}\big[\big(\exp\big(\sqrt{|\xi_s|/k_s'}\big)\big)^{1/\rho} \mid \mathcal{F}_{s-1}\big] \\ &\leq \mathbb{E}\big[\exp\big(\sqrt{|\xi_s|/k_s'}\big) \mid \mathcal{F}_{s-1}\big]^{1/\rho} \\ &\leq 3^{1/\rho} = 2.\end{aligned}$$

Thus, ξ_s is conditionally k_s-sub-Weibull(2) given $\mathcal{F}_{s-1}$. This completes the proof. □

We slightly modify the Theorem 11 in Madden et al. (2024).

Lemma F.11. *Let $(\xi_t)_{t\in\mathbb{N}}$ be a martingale difference sequence adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}_0}$. Let $(k_t)_{t\in\mathbb{N}}$ be a positive sequence such that k_t is $\mathcal{F}_{t-1}$-measurable for each $t \geq 1$. Assume that ξ_t is conditionally k_t-sub-Weibull(2) given $\mathcal{F}_{t-1}$. Define*

$$S_t := \sum_{s=1}^{t} \xi_s, \qquad K_t^2 := \sum_{s=1}^{t} k_s^2.$$

Suppose that there exist deterministic non-decreasing sequences $(\bar{k}_t)_{t\in\mathbb{N}}$ and $(\bar{K}_t)_{t\in\mathbb{N}}$ such that

$$k_t \leq \bar{k}_t, \qquad K_t^2 \leq \bar{K}_t^2$$

almost surely for all $t \geq 1$. Then, for any $\delta \in (0,1]$,

$$\mathbb{P}\left(\forall t \geq 1 : S_t \leq \bar{K}_t \sqrt{2C \log\left(\frac{2\pi^2 t^2}{3\delta}\right)} + 4\bar{k}_t \log^2\left(\frac{2\pi^2 t^2}{3\delta}\right)\right) \geq 1 - \delta,$$

where $C > 0$ is a universal constant.

Proof. Fix $\delta \in (0,1]$. For each $t \geq 1$, define the truncation level

$$c_t := u_t^2, \quad u_t := \log(2\pi^2 t^2/(3\delta)),$$

and let

$$\bar{\xi}_t := \xi_t \mathbb{1}[\xi_t \leq c_t k_t].$$

We first bound the probability that ξ_t exceeds the truncation level. We obtain

$$\begin{aligned}
\mathbb{P}\Big(|\xi_t| \geq c_t k_t\Big) &= \mathbb{P}\Big(\exp\big(\sqrt{|\xi_t|/k_t}\big) \geq \exp\big(\sqrt{c_t}\big)\Big) \\
&\leq \exp\big(-\sqrt{c_t}\big)\mathbb{E}\Big[\exp\big(\sqrt{|\xi_t|/k_t}\big)\Big] \\
&= \exp\big(-\sqrt{c_t}\big)\mathbb{E}\Big[\mathbb{E}\Big[\exp\big(\sqrt{|\xi_t|/k_t}\big) \mid \mathcal{F}_{t-1}\Big]\Big] \\
&\leq 2\exp\big(-\sqrt{c_t}\big),
\end{aligned}$$

where the first inequality holds by Markov's inequality, and the last inequality holds by the definition of sub-Weibull(2). Hence,

$$\mathbb{P}\big(\xi_t \geq c_t k_t\big) \leq \mathbb{P}\big(|\xi_t| \geq c_t k_t\big) \leq 2\exp\big(-\sqrt{c_t}\big) = \frac{3\delta}{\pi^2 t^2}.$$

Using the union bound,

$$\mathbb{P}\big(\exists t \geq 1 : \xi_t \geq c_t k_t\big) \leq \sum_{t=1}^{\infty} \frac{3\delta}{\pi^2 t^2} = \frac{\delta}{2}. \tag{F.2}$$

To control the truncated sequence $(\bar{\xi}_t)_{t\in\mathbb{N}}$, we apply the conditional moment generating function bound for $\bar{\xi}_t$. By Lemma 12 in Madden et al. (2024), we have

$$\mathbb{E}\big[e^{\lambda\bar{\xi}_t} \mid \mathcal{F}_{t-1}\big] \leq \exp\left(\frac{\lambda^2}{2} a_t k_t^2\right), \quad \forall \lambda \in \left[0, \frac{1}{b_t k_t}\right], \tag{F.3}$$

where

$$\begin{aligned}
a_t &= (2^5 + 2)\,\Gamma(5) + \frac{2^6\,\Gamma(7)}{3\sqrt{c_t}}, \\
b_t &= 2\sqrt{c_t}.
\end{aligned}$$

Since $c_t \geq (\log(2\pi^2/3))^2 \geq 1$ for all $t \in \mathbb{N}$, there exists a universal constant $C_0 > 0$ such that $a_t \leq C_0$ for all $t \in \mathbb{N}$.

Fix $t \geq 1$ and define

$$\begin{aligned}
\overline{S}_t &:= \sum_{s=1}^{t} \bar{\xi}_s, \\
\gamma_t &:= \overline{K}_t\sqrt{2C_0 u_t} + 2b_t\bar{k}_t u_t = \overline{K}_t\sqrt{2C_0 u_t} + 4\bar{k}_t u_t^2.
\end{aligned}$$

We claim that

$$\mathbb{P}\big(\overline{S}_t \geq \gamma_t\big) \leq \frac{3\delta}{2\pi^2 t^2}. \tag{F.4}$$

Let

$$\lambda_t^* := \min\left\{\frac{\gamma_t}{C_0\overline{K}_t^2}, \frac{1}{b_t\bar{k}_t}\right\}.$$

Since $b_s \leq b_t$ and $k_s \leq \bar{k}_t$ a.s. for all $s \leq t$, we have

$$\lambda_t^* \leq \frac{1}{b_t\bar{k}_t} \leq \frac{1}{b_t k_t}, \quad s = 1, \ldots, t,$$

almost surely. Hence (F.3) applies to each $s \le t$ with $\lambda = \lambda_t^*$. Using the tower property repeatedly,

$$\begin{aligned}\mathbb{E}\Big[\exp\left(\lambda_t^* \overline{S}_t\right)\Big] &= \mathbb{E}\left[\prod_{s=1}^{t} \exp\left(\lambda_t^* \bar{\xi}_s\right)\right] \\ &\le \exp\left(\frac{(\lambda_t^*)^2}{2}\sum_{s=1}^{t} a_s k_s^2\right) \\ &\le \exp\left(\frac{(\lambda_t^*)^2}{2} C_0 \overline{K}_t^2\right),\end{aligned}$$

where the last inequality holds since $a_s \le C_0$ for all $s \ge 1$ and $\sum_{s=1}^{t} k_s^2 \le \overline{K}_t^2$ a.s. by the assumption. Therefore, by Markov's inequality,

$$\mathbb{P}\big(\overline{S}_t \ge \gamma_t\big) \le \exp\left(-\lambda_t^* \gamma_t + \frac{(\lambda_t^*)^2}{2} C_0 \overline{K}_t^2\right). \tag{F.5}$$

We now bound the exponent in (F.5). In case $\gamma_t/(C_0\overline{K}_t^2) \le 1/(b_t \bar{k}_t)$ holds, $\lambda_t^* = \gamma_t/(C_0\overline{K}_t^2)$. Then, we have

$$-\lambda_t^*\gamma_t + \frac{(\lambda_t^*)^2}{2} C_0 \overline{K}_t^2 = -\frac{\gamma_t^2}{2C_0\overline{K}_t^2} \le -u_t,$$

where the inequality holds since $\gamma_t \ge \overline{K}_t\sqrt{2C_0 u_t}$. In case $\gamma_t/(C_0\overline{K}_t^2) > 1/(b_t\bar{k}_t)$ holds, $\lambda_t^* = 1/(b_t\bar{k}_t)$ and this inequality implies $C_0\overline{K}_t^2 < \gamma_t b_t \bar{k}_t$. Hence, we have

$$\begin{aligned}-\lambda_t^*\gamma_t + \frac{(\lambda_t^*)^2}{2} C_0 \overline{K}_t^2 &= -\frac{\gamma_t}{b_t\bar{k}_t} + \frac{C_0\overline{K}_t^2}{2b_t^2\bar{k}_t^2} \\ &< -\frac{\gamma_t}{2b_t\bar{k}_t} \\ &\le -u_t,\end{aligned}$$

where the last inequality holds since $\gamma_t \ge 2b_t\bar{k}_t u_t$. Thus, in both cases, the exponent in (F.5) less than $-u_t$. Substituting this into (F.5), we have

$$\mathbb{P}\big(\overline{S}_t \ge \gamma_t\big) \le e^{-u_t} = \frac{3\delta}{2\pi^2 t^2},$$

which proves (F.4). Using the union bound,

$$\mathbb{P}\big(\exists t \ge 1 : \overline{S}_t \ge \gamma_t\big) \le \sum_{t=1}^{\infty} \frac{3\delta}{2\pi^2 t^2} = \frac{\delta}{4}. \tag{F.6}$$

If for some $t \ge 1$,

$$S_t = \sum_{s=1}^{t} \ge \gamma_t$$

and simultaneously

$$\xi_s < c_s k_s \quad \forall s \ge 1,$$

then $\bar{\xi}_s = \xi_s$ for all $s \le t$, so $S_t = \overline{S}_t$, and hence

$$\overline{S}_t \ge \gamma_t.$$

Therefore,

$$\big\{\exists t \ge 1 : S_t \ge \gamma_t\big\} \subseteq \big\{\exists t \ge 1 : \overline{S}_t \ge \gamma_t\big\} \cup \big\{\exists s \ge 1 : \xi_s \ge c_s k_s\big\}.$$

Using (F.2) and (F.6), we conclude that

$$\mathbb{P}\big(\exists t \ge 1 : S_t \ge \gamma_t\big) \le \frac{\delta}{2} + \frac{\delta}{4} \le \delta.$$

This completes the proof.

□

F.5 LEMMAS FOR SECTION E.4

Lemma F.12 (Theorem 2 in Janz et al. (2024)). *Let $\delta \in (0,1]$ and $\lambda, L, M > 0$. Let $(x_t)_{t\in\mathbb{N}}$ be a sequence in $\{x \in \mathbb{R}^d : \|x\|_2 \le L\}$, $(r_t)_{t\in\mathbb{N}}$ a real-valued sequence, and $(\nu_t)_{t\in\mathbb{N}_0}$ a non-negative sequence. Let $(\mathcal{F}_t)_{t\in\mathbb{N}_0}$ be a filtration such that x_t is $\mathcal{F}_{t-1}$-measurable and r_t, ν_t are $\mathcal{F}_t$-measurable. Let $\epsilon_t = r_t - \mathbb{E}[r_t \mid \mathcal{F}_{t-1}]$ and assume that the following condition holds:*

$$\mathbb{E}\big[\exp(\gamma\epsilon_t) \mid \mathcal{F}_{t-1}\big] \le \exp(\gamma^2\nu_{t-1}) \qquad \forall|\gamma| \le 1/M,\ t \in \mathbb{N}. \tag{F.7}$$

Let $\overline{\mathbf{\Omega}}_t = \sum_{s=1}^t \nu_{s-1}x_s x_s^\top + \lambda\mathbf{I}_d$ and $S_t = \sum_{s=1}^t \epsilon_s x_s$. Then,

$$\mathbb{P}\left(\exists t \in \mathbb{N} : \|S_t\|_{\overline{\mathbf{\Omega}}_t^{-1}} \ge \frac{\sqrt{\lambda}}{2ML} + \frac{2ML}{\sqrt{\lambda}}\left(d + \frac{1}{2}\log\left(\frac{\det(\overline{\mathbf{\Omega}}_t)}{\lambda^d}\right) + \log\frac{1}{\delta}\right)\right) \le \delta.$$

Lemma F.13. *Let X be a real-valued random variable. Suppose that there exist $\sigma > 0$ and $b > 0$ such that*

$$\mathbb{E}\exp(\eta X) \le \exp(\eta^2\sigma^2), \qquad \textit{for all } |\eta| \le \frac{1}{b}.$$

Then, for every $x > 0$,

$$\mathbb{P}\left(|X| \ge 2\sigma\sqrt{x} + 2bx\right) \le 2e^{-x}.$$

Proof. Fix $x > 0$. By Chernoff's bound, for any $\eta \in [0, 1/b]$,

$$\mathbb{P}(X \ge r) \le \exp(-\eta r)\mathbb{E}e^{\eta X} \le \exp(-\eta r + \eta^2\sigma^2).$$

Choose

$$r = 2\sigma\sqrt{x} + 2bx.$$

We consider two cases. First, suppose $\sqrt{x}/\sigma \le 1/b$. Taking $\eta = \sqrt{x}/\sigma$ gives

$$-\eta r + \eta^2\sigma^2 = -\frac{\sqrt{x}}{\sigma}\left(2\sigma\sqrt{x} + 2bx\right) + x = -x - \frac{2bx^{3/2}}{\sigma} \le -x.$$

Hence $\mathbb{P}(X \ge r) \le e^{-x}$.

Second, suppose $\sqrt{x}/\sigma > 1/b$. Taking $\eta = 1/b$ gives

$$-\eta r + \eta^2\sigma^2 = -\frac{2\sigma\sqrt{x}}{b} - 2x + \frac{\sigma^2}{b^2}.$$

Since $\sqrt{x}/\sigma > 1/b$, we have $\sigma^2/b^2 < x$. Therefore

$$-\frac{2\sigma\sqrt{x}}{b} - 2x + \frac{\sigma^2}{b^2} \le -x.$$

Thus, $\mathbb{P}(X \ge r) \le e^{-x}$. Applying the same argument to $-X$ gives

$$\mathbb{P}(-X \ge r) \le e^{-x}.$$

A union bound yields

$$\mathbb{P}(|X| \ge r) \le 2e^{-x}.$$

This completes the proof. □

Lemma F.14. *Let $\delta \in (0,1]$, $\lambda > 0$, and $M > 0$. Let $(x_s)_{s\in\mathbb{N}}$ be a sequence in $\mathbb{R}^d$, $(r_s)_{s\in\mathbb{N}}$ a real-valued sequence, and $(\nu_s)_{s\in\mathbb{N}_0}$ a non-negative sequence. Let $(\mathcal{F}_s)_{s\in\mathbb{N}_0}$ be a filtration such that x_s is $\mathcal{F}_{s-1}$-measurable and r_s, ν_s are $\mathcal{F}_s$-measurable. Define*

$$\epsilon_s = r_s - \mathbb{E}[r_s \mid \mathcal{F}_{s-1}], \qquad S_t = \sum_{s=1}^t \epsilon_s x_s,$$

and assume that

$$\mathbb{E}\big[\exp(\gamma\epsilon_s) \mid \mathcal{F}_{s-1}\big] \le \exp(\gamma^2\nu_{s-1}) \qquad \forall|\gamma| \le 1/M,\ s \in \mathbb{N}.$$

For a fixed $t \in \mathbb{N}$, define

$$\overline{\mathbf{\Omega}}_t = \sum_{s=1}^t \nu_{s-1}x_s x_s^\top + \lambda\mathbf{I}_d, \qquad \xi_t = \max_{1\le s\le t}\|x_s\|_{\overline{\mathbf{\Omega}}_t^{-1}}.$$

Then, with probability at least $1-\delta$,

$$\|S_t\|_{\overline{\mathbf{\Omega}}_t^{-1}} \le 4\sqrt{d\log 6 + \log(2/\delta)} + 4M\xi_t\{d\log 6 + \log(2/\delta)\}.$$

Proof. Fix $t \in \mathbb{N}$. Let $Z_t := \overline{\mathbf{\Omega}}_t^{-1/2} S_t$. Then, we have

$$\|S_t\|_{\overline{\mathbf{\Omega}}_t^{-1}} = \|Z_t\|_2.$$

Fix $u \in \mathbb{S}^{d-1}$ and define

$$a_s := u^\top \overline{\mathbf{\Omega}}^{-1/2} x_s, \qquad s = 1, \dots, t.$$

Then, we have

$$u^\top Z_t = \sum_{s=1}^{t} a_s \epsilon_s.$$

Moreover,

$$\sum_{s=1}^{t} \nu_{s-1} a_s^2 = u^\top \overline{\mathbf{\Omega}}^{-1/2} \left(\sum_{s=1}^{t} \nu_{s-1} x_s x_s^\top \right) \overline{\mathbf{\Omega}}^{-1/2} u \le \|u\|_2^2 = 1. \tag{F.8}$$

Also,

$$|a_s| \le \|x_s\|_{\overline{\mathbf{\Omega}}^{-1}} \le \xi_t.$$

For any η satisfying $|\eta| \le 1/(M\xi_t)$, we have $|\eta a_s| \le 1/M$ for all $s \le t$. Hence, taking $\gamma = \eta a_s$ gives

$$\mathbb{E}\left[\exp(\eta a_s \epsilon_s) \mid \mathcal{F}_{s-1}\right] \le \exp(\eta^2 a_s^2 \nu_{s-1}), \qquad \forall |\eta| \le 1/(M\xi_t).$$

Iterating this conditional mgf bound gives

$$\mathbb{E} \exp(\eta u^\top Z_t) \le \exp\left(\eta^2 \sum_{s=1}^{t} \nu_{s-1} a_s^2 \right) \le \exp(\eta^2), \qquad |\eta| \le \frac{1}{M\xi_t},$$

where the last inequality holds by (F.8). Therefore $u^\top Z_t$ is sub-exponential with $(1, M\xi_t)$. By Lemma F.13, for every $x > 0$,

$$\mathbb{P}\left(|u^\top Z_t| \ge 2\sqrt{x} + 2M\xi_t x \right) \le 2e^{-x}.$$

Let $\mathcal{N}$ be a $1/2$-net of $\mathbb{S}^{d-1}$. Then, by Lemma 4.1 in Pollard (1990), $|\mathcal{N}| \le 6^d$. For each $u \in \mathcal{N}$, we have already shown that for every $x > 0$,

$$\mathbb{P}\left(|u^\top Z_t| \ge 2\sqrt{x} + 2M\xi_t x \right) \le 2e^{-x}.$$

Set

$$x_\delta = d \log 6 + \log(2/\delta).$$

Since $|\mathcal{N}| \le 6^d$, we have

$$|\mathcal{N}|\, 2e^{-x_\delta} \le 6^d \cdot 2 \exp\{-d \log 6 - \log(2/\delta)\} = \delta.$$

Therefore, by the union bound,

$$\mathbb{P}\left(\exists u \in \mathcal{N} : |u^\top Z_t| > 2\sqrt{x_\delta} + 2M\xi_t x_\delta \right) \le \delta.$$

Equivalently, with probability at least $1 - \delta$,

$$\sup_{u \in \mathcal{N}} |u^\top Z_t| \le 2\sqrt{x_\delta} + 2M\xi_t x_\delta. \tag{F.9}$$

Fix any $v \in \mathbb{S}^{d-1}$. Since $\mathcal{N}$ is a $1/2$-net, there exists $u \in \mathcal{N}$ such that

$$\|v - u\|_2 \le \frac{1}{2}.$$

Hence,

$$|v^\top Z_t| \le |u^\top Z_t| + |(v-u)^\top Z_t| \le |u^\top Z_t| + \|v-u\|_2 \|Z_t\|_2 \le \sup_{w \in \mathcal{N}} |w^\top Z_t| + \frac{1}{2}\|Z_t\|_2.$$

Taking the supremum over $v \in \mathbb{S}^{d-1}$ on the left-hand side yields

$$\|Z_t\|_2 = \sup_{v \in \mathbb{S}^{d-1}} |v^\top Z_t| \le \sup_{w \in \mathcal{N}} |w^\top Z_t| + \frac{1}{2}\|Z_t\|_2.$$

Therefore, we have

$$\|Z_t\|_2 \le 2 \sup_{w \in \mathcal{N}} |w^\top Z_t|.$$

Combining this with (F.9), we conclude that with probability at least $1-\delta$,

$$\|Z_t\|_2 \le 4\sqrt{x_\delta} + 4M\xi_t x_\delta.$$

Since $\|S_t\|_{\overline{\mathbf{\Omega}}_t^{-1}} = \|Z_t\|_2$, this proves the first claim.

□

G Curvature dependence of the regret bound

The curvature dependence of Theorem 4.2 enters through C_1. Recall from (E.1) that

$$\kappa = \frac{1}{\inf_{x \in \mathcal{X},\, \theta \in \Theta_R} \mu'(x^\top \theta)}, \qquad \Theta_R = \{\theta \in \Theta : \|\theta - \theta^*\|_2 \le R\}.$$

Unlike the worst-case curvature constants defined over an origin-centered parameter ball specified by a known parameter norm bound (Faury et al., 2022; Zhang et al., 2025), our κ is defined over a fixed neighborhood of the true parameter.

Under the truncated Gaussian design in Proposition E.1, its proof gives

$$C_1 = \frac{c_0}{g(\tau)\kappa d}, \qquad c_0 = c_0(K, L, \sigma_{\min}) > 0.$$

Thus, the $C_1^{-1/2}\sqrt{dT}$ term in Theorem 4.2, up to logarithmic factors and the common constant C, retains dependence on this local worst-case curvature.

In contrast, κ^* is determined by the average curvature at the optimal arms evaluated at θ^*. Existing analyses obtain leading terms governed by this optimal-arm curvature, with worst-case curvature appearing only in lower-order terms (Faury et al., 2022; Zhang et al., 2025). Obtaining a regret bound for OnlineTS whose leading term depends on curvature only through κ^* remains an interesting direction for future work.

H Implementation details of the comparison methods

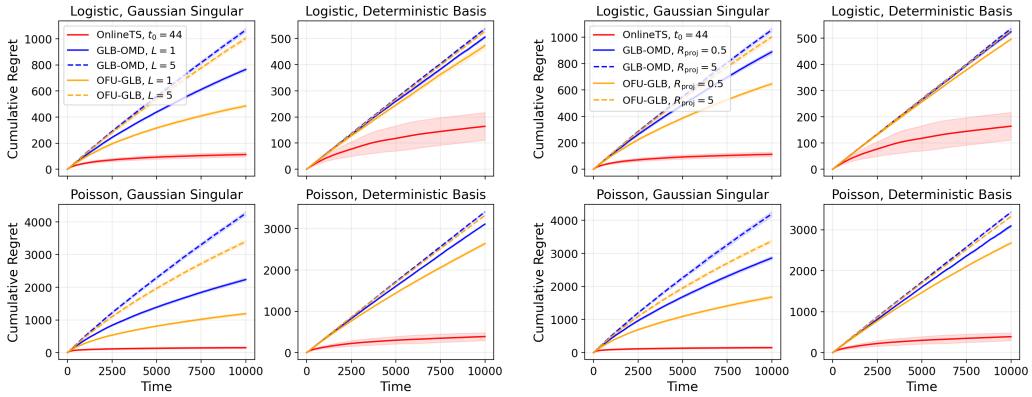
We describe how the supplied context norm bound L and parameter radius R_{proj} enter the implementations of GLB-OMD and OFUGLB. Both methods are originally formulated under $\|x\|_2 \le 1$. For general bounds $\|x\|_2 \le L$ and $\|\theta\|_2 \le R_{\text{proj}}$, the linear predictor satisfies

$$|x^\top \theta| \le L R_{\text{proj}}.$$

The identity $x^\top \theta = (x/L)^\top (L\theta)$ also shows that LR_{proj} is the corresponding parameter norm bound under unit context normalization. This motivates using LR_{proj} in the predictor-dependent constants of the two methods, while the parameter constraint remains $\{\theta \in \mathbb{R}^d : \|\theta\|_2 \le R_{\text{proj}}\}$.

GLB-OMD The OMD update is constrained to this parameter set. Using the self-concordance constant S in (A2), we set the step size to $\eta = 1 + SLR_{\text{proj}}$. The confidence radius is computed using LR_{proj} and an upper bound on μ' over $[-LR_{\text{proj}}, LR_{\text{proj}}]$. Thus, R_{proj} determines the constraint on the estimator, while both supplied bounds enter the step size and confidence radius.

OFUGLB We compute the constrained MLE over the same parameter set and use the ellipsoidal confidence set for action selection. The radius-dependent terms in the confidence bound calculation are evaluated using LR_{proj}. Consequently, R_{proj} determines the MLE constraint, while LR_{proj} enters the confidence bound and the resulting exploration bonus.



(a) Sensitivity to L (b) Sensitivity to R_{proj}

Figure 3: Comparison with existing methods under additional context distributions.

I Additional Experiments

I.1 Alternative context distributions

In this section, we report additional comparisons under alternative context distributions. These experiments are supplementary to the main results in Section 5.1. Let $j(a) := ((a-1) \bmod d) + 1$, so that $j(a) \in [d]$ assigns to each arm $a \in [K]$ a coordinate index cyclically. We consider the following two settings. In the Gaussian singular setting, for each arm $a \in [K]$,

$$x_{t,a} \sim \mathcal{N}(0, e_{j(a)} e_{j(a)}^\top),$$

where $e_1, \ldots, e_d$ denote the standard basis vectors in $\mathbb{R}^d$. In the deterministic basis setting, for each arm $a \in [K]$,

$$x_{t,a} = e_{j(a)}.$$

In both settings, we compare OnlineTS, GLB-OMD, and OFUGLB under the same experimental protocol as in Section5.1.

Figure 3 shows that OnlineTS continues to achieve lower regret than GLB-OMD and OFUGLB in both settings. Thus, the same qualitative conclusion as in the main text continues to hold under these additional context distributions.

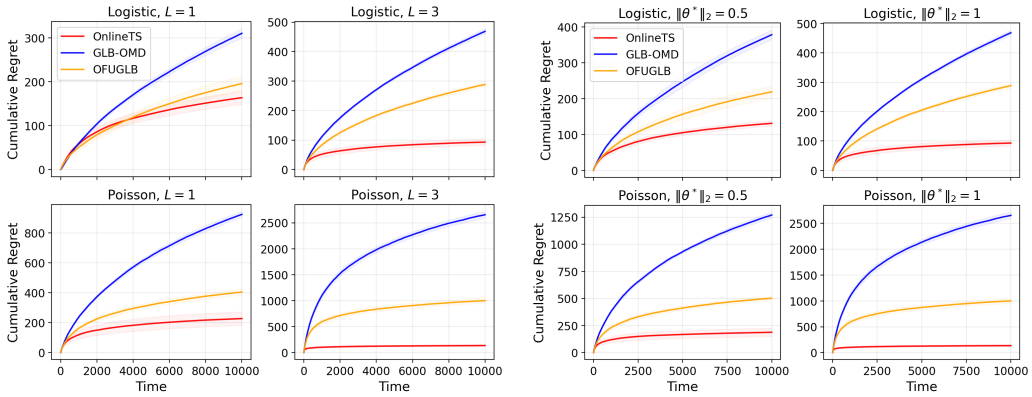
These settings are also of interest because (A3) is less naturally aligned with such context distributions. The fact that OnlineTS still performs well in these experiments suggests that (A3) should be viewed as a sufficient condition for our current theory rather than as a necessary condition for good empirical performance. Providing a refined theoretical explanation for this phenomenon is left for future work.

I.2 Additional comparisons in well-specified settings

In this subsection, we report additional comparisons in well-specified settings for GLB-OMD and OFUGLB. These experiments are intended to complement the misspecified and conservative settings considered in Section 5.1.

Recall that both GLB-OMD and OFUGLB require a context norm bound L and a projection radius R_{proj} satisfying $\|\theta^*\|_2 \le R_{\text{proj}}$. Here, we consider well-specified settings under the Uniform context distribution $x_{t,a} \sim \text{Unif}(\|x\|_2 \le L)$, where these quantities are set to match the true problem instance. Specifically, we vary the true context bound and the norm of the true parameter, and set the corresponding constants in GLB-OMD and OFUGLB accordingly.

The left panel of Figure 4 fixes $\|\theta^*\|_2 = 1$ and varies the context bound $L \in \{1, 3\}$. The right panel fixes $L = 3$ and varies $\|\theta^*\|_2 \in \{0.5, 1\}$, with R_{proj} matched to the true parameter norm in each case. In both logistic and Poisson bandits, OnlineTS consistently achieves lower cumulative regret than GLB-OMD and OFUGLB across all tested settings.



(a) Varying the true context bound L (b) Varying the true parameter norm $\|\theta^*\|_2$

Figure 4: Additional comparisons in well-specified settings.

These results complement the main text in two ways. First, the results show that the empirical advantage of OnlineTS is not limited to settings where the problem-dependent constants of the baselines are misspecified or overly conservative. Second, even when these constants are chosen in a theoretically favorable manner for GLB-OMD and OFUGLB, OnlineTS remains empirically competitive and typically achieves the best regret among the compared methods.

I.3 Comparison with additional GLB algorithms

In this subsection, we extend the comparison to GLM-TSL, GLM-FPL, and GLOC, together with GLB-OMD and OFUGLB. We consider logistic and Poisson bandits with $d = K = 10$ and $T = 10^4$. Contexts are sampled uniformly from the ball of radius 3, and $\|\theta^*\|_2 = 1$. We supply the comparison methods with the corresponding norm bounds, $L = 3$ and $R_{\text{proj}} = 1$, where required. For these additional experiments, we use $\lambda = 10$, $\delta = 0.01$, and $t_0 = 45$ for OnlineTS. All results are averaged over 10 independent runs.

For GLM-TSL and GLM-FPL, we consider informal and theory-based exploration settings based on Kveton et al. (2020), adapted to the reward models in our experiments. These settings are labeled "informal" and "theory" in the figures. GLOC uses its theory-based confidence radius (Jun et al., 2017).

Figures 5 and 6 show the cumulative regret and the corresponding regret–runtime trade-off, respectively. OnlineTS achieves competitive cumulative regret across both reward models, with lower regret than GLB-OMD, OFUGLB, and GLOC. In the logistic setting, its cumulative regret is comparable to the informal settings of GLM-TSL and GLM-FPL. In the Poisson setting, OnlineTS achieves cumulative regret comparable to GLM-TSL, while GLM-FPL with the informal setting achieves lower regret. However, the informal settings of these batch-based methods require approximately 16–20 times as much computation time as OnlineTS across the two reward models. Thus, the lower regret of GLM-FPL in the Poisson setting comes with substantially greater computation time. OnlineTS retains the low computational cost of one-pass methods while achieving substantially lower regret than GLB-OMD and GLOC in these experiments. These results demonstrate that OnlineTS combines competitive regret relative to batch-based randomized methods with the computational efficiency of one-pass updates.

J Computational resources used

All experiments in this paper were conducted using a machine equipped with an Intel(R) Core(TM) i9-10900X CPU. No GPU was used.

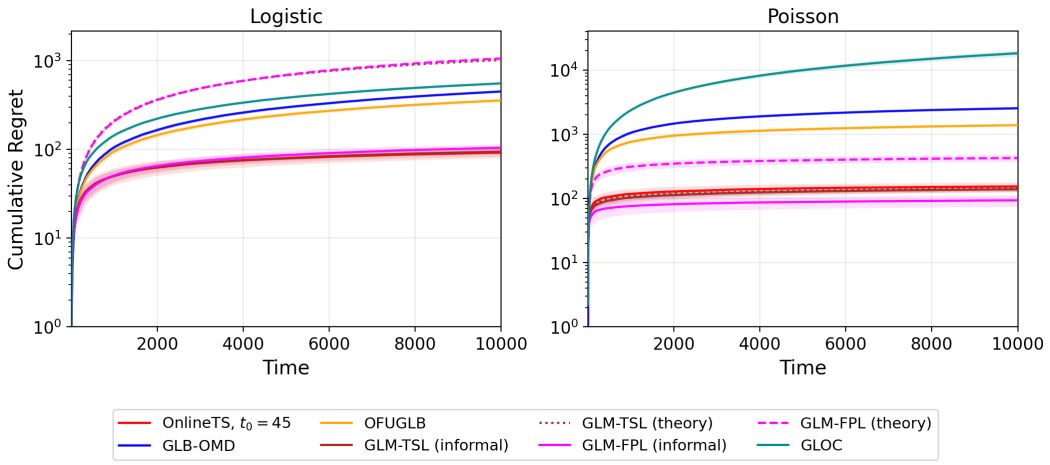


Figure 5: Cumulative regret comparison with additional GLB algorithms.

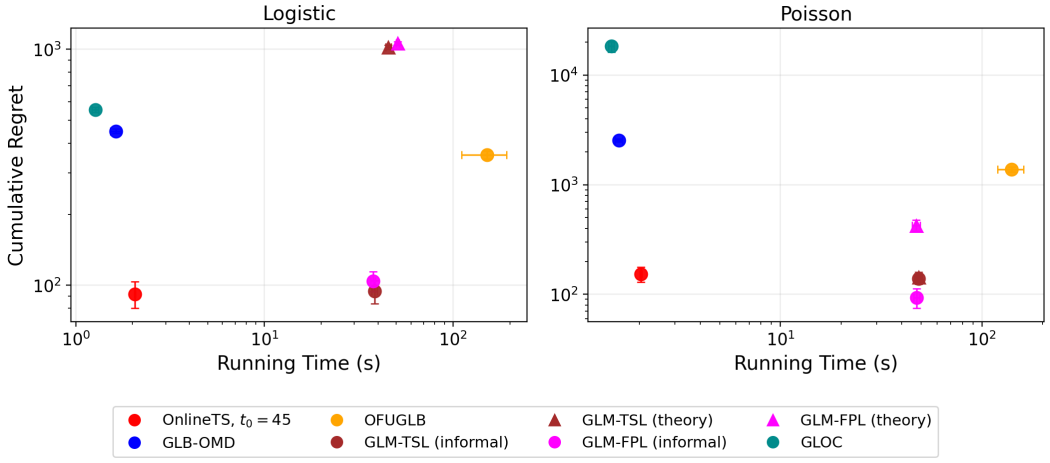


Figure 6: Regret–runtime comparison with additional GLB algorithms.