Frequentist Regret Analysis of Thompson Sampling with Fractional Posteriors for Generalized Linear Bandit

author names withheld

Editor: Under Review for COLT 2024

Abstract

We propose a variant of the Thompson Sampling (TS) algorithm for solving stochastic generalized linear bandit problems. We call the proposed algorithm α-TS, where we use fractional or α-posterior instead of the standard posterior in TS. Our main contribution is to identify general regularity conditions on the prior and reward distributions to generalize the analysis of α-TS without assuming any tractable representation (or approximation) of the posterior distribution, unlike previous works. The exponential (generalized linear model) and sub-Gaussian family of distributions satisfy our general conditions on the reward distributions. Our general regret bound yields a regret bound of $O(d^{7/4}\sqrt{T})$ for the exponential family and of $O(d^2\sqrt{T})$ for the sub-Gaussian family of reward distributions, which is away by a factor of $d^{3/4}$ and d, respectively, from the lower bound. Our proof technique combines the recent advancements in the analysis of linear bandit problems with the first and second-order posterior concentration theory in the Bayesian statistics literature.

Keywords: Thompson sampling, generalized linear models, frequentist regret bounds, linear bandits, finite-sample Bernstein-von Mises, posterior concentration

1. Introduction

The bandit framework (Robbins, 1952) is widely used to model various sequential decision-making problems with many applications in healthcare, advertising, resource allocation, robotics, material discovery, etc. To solve such sequential decision-making problems, the decision-maker must balance exploration and exploitation. In particular, to make a sequence of decisions that maximizes the total outcome, the decision maker has to choose between exploring lesser-explored decisions to gain new knowledge and exploiting the previous knowledge to choose the (statistically) best decision. This work considers a stochastic generalized linear bandit problem with compact decision space (in $\mathbb{R}^d$) and reward distribution with generalized linear mean.

Thompson Sampling (TS) (Thompson, 1933) and *optimism-in-the-face-of-uncertainty* (OFU) (Lai et al., 1985) are two broad classes of algorithms to solve various versions of the bandit problem. The OFU algorithms compute a confidence set for the unknown coefficients in the reward model and then select an optimistic estimate of the coefficient and decision that maximizes the estimated outcome. On the other hand, TS uses a Bayesian heuristic (posterior distribution) to address the exploration-exploitation dilemma. TS computes a randomized decision-making policy, which samples a decision with the posterior probability of that decision being the best. The theoretical performance of such algorithms is typically measured by computing a bound on a term called regret, which is the total sum of the loss incurred by the algorithm over past trials. At each trial, the algorithm incurs loss because of taking a sub-optimal decision instead of the oracle's best decision.

In this work, we propose and theoretically analyze a version of TS to solve the generalized linear bandit (LB) problem.

Agrawal and Goyal (2013), in their pioneering work, analyze TS for LB problems and derive state-of-the-art regret bound of $O(d^{3/2}\sqrt{T})$, where T is the number of trials. Their regret bound is very close to the lower bound of $\Omega(d\sqrt{T})$ for LB derived in Dani et al. (2008). The other important work by Abeille and Lazaric (2017) computes a similar upper bound by using a completely novel proof technique. Both works consider a TS algorithm for LB with sub-Gaussian errors[1]; however, the sampling distribution coincides with the posterior distribution only when the errors are centered Gaussian and the prior distribution on the unknown coefficients is also Gaussian (conjugate setting). The closed-form expression of the Gaussian sampling distribution enables them to compute anti-concentration bounds[2]; the availability of which is a crucial component for the analysis in both works. However, recall that the Bayesian heuristic in TS is more general as it is applicable for any prior and reward combination (Li and Chapelle, 2012; Chapelle and Li, 2011; Urteaga and Wiggins, 2018; Hong et al., 2022). This work contributes to further generalizing the analysis of TS with a more general reward and prior model and, sticks to sampling from the Bayesian posterior instead of the Gaussian approximation used in the previous works on TS. Our regularity conditions on the reward distributions are satisfied by both sub-Gaussian and exponential families. While the sub-Gaussian family has been extensively used in most of the previous works on LB (Dani et al., 2008; Abbasi-Yadkori et al., 2011; Agrawal and Goyal, 2013; Abeille and Lazaric, 2017), the use of the exponential family (generalized linear model (GLM) (Nelder and Wedderburn, 1972)) is novel to the best of our knowledge.

We derive the regret bound for a version of the TS algorithm that uses fractional or α-posterior (Bhattacharya et al., 2019) instead of the standard posterior distribution. To compute the α-posterior distribution, the likelihood in the definition of the standard posterior distribution is tempered with a factor α. We name our new algorithm α-TS. For α-TS, we compute a regret bound for any $\alpha \in (0, 1)$, but it grows exponentially with d. However, by optimally choosing α for the exponential and sub-Gaussian families of distribution, we show that the regret bound is of $O(d^{7/4}\sqrt{T})$ (for $\alpha = d^{-3/2}$) and $O(d^2\sqrt{T})$ (for $\alpha = d^{-2}$), respectively. Intuitively, setting α to a value in $(0, 1)$ inflates the variance of the posterior distribution. A similar variance inflation factor is also considered in the sampling distribution (or posterior distribution) for the analysis in Agrawal and Goyal (2013) and Abeille and Lazaric (2017). This work formalizes the idea of introducing the variance inflation factor by replacing the standard posterior with α-posterior.

Our regret analysis uses the idea of separating the actions into saturated and unsaturated sets as introduced in Agrawal and Goyal (2013). However, to derive posterior concentration and anti-concentration properties, we adapt the state-of-the-art first and second-order analysis of the Bayesian posterior from the Bayesian statistics literature. In particular, we leverage ideas developed in Bhattacharya et al. (2019); Zhang and Gao (2020); Ghosal et al. (2000) to derive first-order concentration properties and in Spokoiny (2012); Panov and Spokoiny (2015) to derive the second-order properties of the α-posterior distribution. We would also like to note that these results hold under fairly

1. Abeille and Lazaric (2017) also extend their analysis to a generalized case with sub-Gaussian errors, which is different than the canonical exponential family generalized linear models(GLM).
2. Abeille and Lazaric (2017) also proposes two other non-Gaussian sampling distributions for which the anti-concentration bounds are available

general conditions on the prior and the reward model, which enabled us to generalize the analysis of α-TS.

2. Literature review

In addition to the works discussed in the introduction, there are various research directions that recent works have pursued for TS to solve LB problems. Luo and Bayati (2023); Hamidi and Bayati (2023) focus on improving the regret bound for TS so that it matches the lower bound of $\Omega(d\sqrt{T})$. In particular, Luo and Bayati (2023) proposes a new algorithm that switches between OFU and TS algorithms based on the history of observations and computes minimax optimal frequentist regret guarantees. Zhang (2021) proposes feel-good TS and generalizes the theoretical properties of TS without using any structural assumptions on the posterior sampling distribution. Zhang (2021) assumes that the negative reward log-likelihood is a weighted squared difference between the true reward and the random reward. The modified negative log-likelihood favors sampling of the optimal model from the posterior distribution and enables the author to derive a regret bound that matches the lower bound for LB. There is also increasing interest in using more complex reward models, such as semi-parametric (Greenewald et al., 2017), nonparametric (Rigollet and Zeevi, 2010; Kim et al., 2021), and high-dimensional (Bastani and Bayati, 2020; Chakraborty et al., 2023) reward models in contextual bandit settings.

3. Problem setup

We consider a stochastic generalized linear bandit problem. We denote the arbitrary action space as $\mathbb{A} \subset \mathbb{R}^d$ and the time horizon as T. For any $n \in \mathbb{N}$, we denote $[n]$ to represent $\{1, 2, 3, \ldots, n\}$. In this problem, at each time step $t \in [T]$, the learner chooses an action $a_t \in \mathbb{A}$ (according to some rule, possibly randomized) and observes a reward r_t as feedback from an unknown environment in response to the learner's action. We model the environment by assuming that the reward at time t is generated from a distribution $P_{\theta_0}(\cdot|a_t)$ with conditional density function denoted as $p_0(\cdot|a_t)$. Here, $\theta_0 \in \Theta \subseteq \mathbb{R}^d$ is an unknown but fixed parameter. While interacting with the environment, the learner collects the history of observations, which we denote as $X_t := \{a_s, r_s\}_{s=1}^t$. We define the filtration generated by X_t as $\mathcal{F}_t := \mathcal{F}_0 \cup \{\sigma(X_t)\}$, that contains all the information till time t including any prior information $\mathcal{F}_0$. At time $t+1$, the learner takes an action according to a randomized action selection policy $q(\cdot|\mathcal{F}_t)$, that is $a_{t+1} \sim q(\cdot|\mathcal{F}_t)$. We also assume that the conditional mean reward under distribution $P_\theta(\cdot|a_t)$ can be represented using a strictly monotonic and Lipschitz link function $g : \mathbb{R} \mapsto \mathbb{R}$ as $g(a_t^\top \theta)$.

We evaluate a learner (or algorithm) by comparing its action at time t to the oracle best action $a_0 := \arg\max_{a \in \mathbb{A}} g(a^\top \theta_0)$. In particular, the learner's objective is to minimize the following cumulative regret on a sample path of observations,

$$\boldsymbol{R}(T) = \sum_{t=1}^{T} (g(a_0^\top \theta_0) - g(a_t^\top \theta_0)), \tag{1}$$

which quantifies the total regret of taking suboptimal action a_t instead of the oracle's best action a_0. We measure the algorithm's performance by analyzing the $\mathbb{E}[\boldsymbol{R}(T)]$, where the expectation is taken with respect to the distribution that generates the sequence of observations X_{T-1}.

3.1. α-Thompson Sampling

We study a variant of the standard Thompson sampling (TS) algorithm, where we use α-posterior instead of the standard posterior distribution ($\alpha = 1$). We call this new algorithm as α-TS. Recall that TS was proposed by Thompson (1933) to solve a multiarmed bandit problem, where the exploration-exploitation dilemma is addressed using a posterior distribution. The pseudo-code of α-TS is provided in Algorithm 1. In α-TS, we posit a prior distribution $\Pi(\cdot)$ on a measurable space $(\Theta, \mathcal{Q})$. For any $t \geq 1$ and $\alpha \in (0, 1)$, we define the α-posterior distribution $\Pi_{t,\alpha}(\cdot)$ for any measurable set $Q \in \mathcal{Q}$ as

$$\Pi_{t,\alpha}(Q|\mathcal{F}_t) = \frac{\int_Q \prod_{s=1}^t [p_\theta(r_s|a_s)q(a_s|\mathcal{F}_{s-1})]^\alpha \Pi(d\theta)}{\int_\Theta \prod_{s=1}^t [p_\theta(r_s|a_s)q(a_s|\mathcal{F}_{s-1})]^\alpha \Pi(d\theta)} = \frac{\int_Q \prod_{s=1}^t [p_\theta(r_s|a_s)]^\alpha \Pi(d\theta)}{\int_\Theta \prod_{s=1}^t [p_\theta(r_s|a_s)]^\alpha \Pi(d\theta)}, \quad (2)$$

where $q(a_s|\mathcal{F}_{s-1})$ is the likelihood (or the randomized action selection rule) of choosing $a_s \in \mathbb{A}$ given $\mathcal{F}_{s-1}$ and $p_\theta(r_s|a_s)$ is the likelihood of observing r_s given a_s. Note that a_s is adapted to $\mathcal{F}_{s-1}$. Essentially, step 3 and 4 of Algorithm 1, combines together to sample from the randomized action selection rule $q(a_s|\mathcal{F}_{s-1})$. In case if the action space $\mathbb{A}$ is a discrete set, then $q(a_s|\mathcal{F}_{s-1})$ can be precisely defined as the α-posterior probability of choosing a_s as the optimal action, that is equals to $\Pi_{t,\alpha}(\{a_s^\top \theta = \max_{a\in\mathbb{A}} a^\top \theta\} | \mathcal{F}_{s-1})$. Observe that the definition of the α-posterior distribution in equation 2 is unaffected by the probability of choosing the best action a_s. In the rest of the paper, we write $E[\cdot|\mathcal{F}_t]$ as the expectation and $P(\cdot|\mathcal{F}_t)$ as probability w.r.t. α-posterior distribution.

Algorithm 1: α-Thompson Sampling (α-TS)

Initialization: Set $X_0 = \{\}$, $g(\cdot)$, and prior distribution Π on Θ

1 **for** *time step* $t = 1, 2, \ldots T$ **do**
2 Update α-posterior $\Pi_{t,\alpha}(\cdot|\mathcal{F}_{t-1})$;
3 Generate a sample $\theta_t \sim \Pi_{t,\alpha}(\cdot|\mathcal{F}_{t-1})$;
4 Compute an optimal action $a_t = \arg\max_{a\in\mathbb{A}} g(a^\top \theta_t)$;
5 Observe r_t from $p_0(\cdot|\mathcal{F}_{t-1}, a_t)$;
6 Update $X_t = X_{t-1} \cup \{a_t, r_t\}$.;

It is evident from Algorithm 1 that the true likelihood of generating data in α-TS is $p_0^{(T)}(X_T) := \prod_{t=1}^T [p_0(r_t|a_t)q_\alpha(a_t|\mathcal{F}_{t-1})]$, and we denote the corresponding data-generating distribution as $P_0^{(T)}$ till time T. The expectation of the regret is taken with respect to $P_0^{(T-1)}$. For any $\theta \in \Theta$, we also denote the reward generating distribution given a sequence of actions $a^{(T)} := \{a_1, a_2, \ldots, a_T\}$ as $P_{\theta,a}^{(T)}$ and the corresponding density as $p_\theta^{(T)}(r^{(T)}|a^{(T)}) := \prod_{t=1}^T p_\theta(r_t|a_t)$. For any $\alpha > 0$, we denote the α-Rényi divergence between two reward densities $p_\theta^{(t)}(r^{(t)}|a^{(t)})$ and $p_\vartheta^{(t)}(r^{(t)}|a^{(t)})$ for any $\theta, \vartheta \in \Theta$ as

$$\begin{aligned} D_\alpha^{(t)}(\theta, \vartheta) &:= \frac{1}{\alpha - 1} \log \int p_\theta^{(t)}(r^{(t)}|a^{(t)})^\alpha p_\vartheta^{(t)}(r^{(t)}|a^{(t)})^{1-\alpha} dr^{(t)} \\ &= \sum_{s=1}^t \frac{1}{\alpha - 1} \log \int p_\theta(r|a_s)^\alpha p_\vartheta(r|a_s)^{1-\alpha} dr := \sum_{s=1}^t D_\alpha^t(\theta, \vartheta), \end{aligned} \quad (3)$$

where $r^{(t)} := \{r_1, r_2, \ldots, r_t\}$ and the second equality is due to the additivity property of α-Rényi divergence (Van Erven and Harremos, 2014, Theorem 28).

4. Assumptions

In this section, we provide the list of regularity conditions that we impose on the action space, prior, and the reward model to derive our general regret bound. We first assume that the action space is compact, which is a standard assumption in the linear bandit literature (Agrawal and Goyal, 2013; Abeille and Lazaric, 2017; Dani et al., 2008).

Assumption 4.1 (Action space) *We assume the $\mathbb{A}$ is a closed and bounded subset of $\mathbb{R}^d$. In particular, without loss of generality, we assume that for any $a \in \mathbb{A}$, $\|a\| \leq 1$, where $\|\cdot\|$ is the Euclidean norm.*

We prove our regret bounds under some mild regularity conditions on the reward and prior distributions. The first assumption is on the link function $g(\cdot)$.

Assumption 4.2 (Link function) *The link function is Lipschitz continuous and strictly monotonic, that is, for any $x, y \in \mathbb{R}$, there exists a constant $C_g > 0$, such that $|g(x) - g(y)| \leq C_g|x - y|$, and there exists an $m > 0$ such that $g'(x) > m$.*

In many works on LB, it is common to assume that the rewards are generated from the following linear model: $r_t = a_t^\top \theta_0 + \eta_t$, where η_t is a zero mean sub-Gaussian error. Note that for such models, the link function is identity. Moreover, for exponential family reward distributions (defined in Definition 6 formally), the link function is a standard terminology that defines the mean of the respective distribution in terms of its parameters.

Next, we impose another technical condition on the reward distribution that can be shown to be easily satisfied by both sub-Gaussian (Lemma 15) and exponential (Lemma 11) family distributions.

Assumption 4.3 (Renyi divergence) *At any time t, given $a^{(t)}$, we assume that for any $\theta, \vartheta \in \Theta$ and for some constant $C > 0$, $|a_t^\top \theta - a_t^\top \vartheta| \leq C\sqrt{\frac{2}{\alpha} D_\alpha^t(\theta, \vartheta)}$.*

Regularity conditions for α-posterior contraction. The following regularity condition is a joint condition on the prior and the reward-generating distribution, which is crucial for obtaining near optimal minimax rate of convergence of the α-posterior distribution.

Assumption 4.4 (Prior thickness) *We assume that for a fix $\alpha \in (0, 1)$ and a positive sequence $\{\epsilon_t\}_{t \geq 0}$, there exist a $t_0 \geq 1$, such that for all $t \geq t_0$, $t\epsilon_t^2 \geq 2.4/\alpha$ and $\Pi(B(\theta_0, \epsilon_t)) \geq e^{-\frac{D\alpha t\epsilon_t^2}{4}}$, where $B(\theta_0, \epsilon_t) := \left\{\theta \in \Theta : D_2^{(t)}(\theta_0, \theta) \leq \frac{D\alpha t\epsilon_t^2}{4}\right\}$ is a neighborhood of θ_0 defined for any given $a^{(t)}$ and a positive constant D.*

The above assumption requires that, for any action sequence, the prior distribution places an exponentially decaying mass to a shrinking neighborhood of the true model parameter θ_0. The neighborhood $B(\theta_0, \epsilon_t)$ of θ_0 is defined using a 2-Rényi divergence measure $D_2^{(t)}(\theta_0, \theta)$ (see Definition in equation 3 for $\alpha = 2$). Such type of assumptions are common in the works studying non-asymptotic convergence rate of the Bayesian posteriors (Ghosal et al., 2000; Zhang and Gao, 2020; Bhattacharya et al., 2019) and are satisfied by a large class of prior-likelihood combination for both

parametric and non-parametric problems. Typically, $\epsilon_t \to 0$ as $t \to \infty$ and determines the rate of convergence of the α-posterior distribution. We will later see in Lemma 4 that this assumption enables us to construct (near) minimax optimal rate of convergence of the α-posterior distribution that is required for computing the bounds on the regret of α-TS. In particular, observe that if ϵ_t satisfies (B1), then any $\epsilon_t' \geq \epsilon_t$ satisfies (B1), and thus ϵ_t is optimal. A common recipe that we follow to show this assumption is to locally bound the 2-Rényi divergence term by $t\|\theta - \theta_0\|^2$ (assuming Θ is an Euclidean space) and then appropriately control the rate at which prior mass of the shrinking 2-norm neighborhood decays. We show that the assumption above is satisfied by any bounded prior density with both exponential and sub-Gaussian families in Lemmas 12 and 17, respectively. We also impose another regularity condition on the prior density.

Assumption 4.5 (sub-Gaussian prior) *We assume that* $\int_\Theta \Pi(d\theta) e^{\frac{\gamma}{2}\|\theta\|^2} < C_\pi$ *for any* $\gamma \leq 1$.

The assumption above can be easily satisfied by any sub-Gaussian prior distribution.

Regularity conditions for the finite Sample Bernstein-von Mises for α-posterior distribution. In this section, we specify the regularity conditions developed in Spokoiny (2012); Panov and Spokoiny (2015) for finite sample Bernstein-von Mises (BvM) theorem to hold for parametric models. Given $a^{(t)}$, we denote the log-likelihood of generating $r^{(t)}$ as $\mathbb{L}(\theta) = \sum_{s=1}^t \log p_\theta(r_t|a_t)$. $\nabla\mathbb{L}(\theta)$ denotes the gradient of $\mathbb{L}(\cdot)$ evaluated at θ and $\nabla^2\mathbb{E}_a\mathbb{L}(\theta)$ stands for the Hessian of the expected log-likelihood, where $\mathbb{E}_a[] := \mathbb{E}[\cdot|a^{(t)}]$ is the expectation with respect to $P_{0,a}^{(t)}$. Define

$$\mathbb{D}_0^2 := -\nabla^2\mathbb{E}_a\mathbb{L}(\theta_0). \tag{4}$$

Note that $\mathbb{D}_0^2$ is defined akin to the Fisher information matrix of $P_{\theta,a}^{(t)}$ at θ_0. Also define $\mathbb{D}_0^2(\theta) := -\nabla^2\mathbb{E}_a\mathbb{L}(\theta)$ and note that $\mathbb{D}_0^2 = \mathbb{D}_0^2(\theta_0)$. Denote the maximum likelihood estimator as $\hat{\theta}_n := \arg\max_{\theta\in\Theta}\mathbb{L}(\theta)$ and $\theta_0 = \arg\max_{\theta\in\Theta}\mathbb{E}_a[\mathbb{L}(\theta)]$. The stochastic part of the log-likelihood is denoted as $\zeta(\theta) := \mathbb{L}(\theta) - \mathbb{E}[\mathbb{L}(\theta)]$.

The regularity conditions required for the finite sample BvM to hold are divided into local and global. The local conditions only describe the properties of $\mathbb{L}(\theta)$ for $\theta \in \Theta(\mathtt{r}_0)$ with some fixed value $\mathtt{r}_0$, where

$$\Theta_0(\mathtt{r}_0) := \big\{\theta \in \Theta\colon \|\mathbb{D}_0(\theta - \theta_0)\| \leq \mathtt{r}_0\big\}. \tag{5}$$

The global conditions have to be fulfilled on the whole Θ. These conditions are constructed to replace the celebrated *local asymptotic normality* (LAN) condition (Van der Vaart, 2000, Chapter 7) on the $\mathbb{L}(\theta)$, which is required for the asymptotic Bernstein-von Mises theorem. Recall that LAN is essentially a local quadratic approximation of $\mathbb{L}(\theta)$ in the vicinity of θ_0.

Assumption 4.6 *We start with some exponential moment conditions.*

(ED_0) *There exists a constant* $\nu_0 > 0$ *, and a constant* $\mathtt{g} > 0$ *such that*

$$\sup_{\gamma\in\mathbb{R}^d} \log \mathbb{E}_a \exp\left\{\mathtt{m}\frac{\langle\nabla\zeta(\theta_0),\gamma\rangle}{\|\mathbb{D}_0\gamma\|}\right\} \leq \frac{\nu_0^2\mathtt{m}^2}{2},\ |\mathtt{m}| \leq \mathtt{g}. \tag{6}$$

(ED_1) *There are constants* $\omega > 0$ *and for each* $\mathtt{r} > 0$ *a constant* $\mathtt{g}(\mathtt{r}) > 0$ *such that for all* $\theta \in \Theta_0(\mathtt{r})$*:*

$$\sup_{\gamma_1,\gamma_2\in\mathbb{R}^d}\sup_{\theta\in\Theta_0(\mathtt{u})} \log \mathbb{E}_a \exp\left\{\frac{\mathtt{m}}{\omega}\frac{\gamma_1^\top \nabla^2\zeta(\theta)\gamma_2}{\|\mathbb{D}_0\gamma_1\|\|\mathbb{D}_0\gamma_2\|}\right\} \le \frac{\nu_0^2\mathtt{m}^2}{2},\ |\mathtt{m}| \le \mathtt{g}(\mathtt{r}). \tag{7}$$

$(\mathcal{L}_0)$ *There exists a constant* $\delta(\mathtt{r})$ *such that it holds on the set* $\Theta_0(\mathtt{r})$ *for all* $\mathtt{r} \le \mathtt{r}_0$

$$\left|\mathbb{D}_0^{-1}\mathbb{D}_0^2(\theta)\mathbb{D}_0^{-1} - I_d\right| \le \delta(\mathtt{r}). \tag{8}$$

The following is the required global condition.

$(\mathcal{L}\mathtt{r})$ *For any* $\mathtt{r} > 0$ *there exists a value* $\mathtt{b}(\mathtt{r}) > 0$*, such that* $\mathtt{r}b(\mathtt{r}) \to \infty$ *as* $\mathtt{r} \to \infty$ *and*

$$-\mathbb{E}\mathbb{L}(\theta,\theta_0) \ge \mathtt{r}^2 b(\mathtt{r}) \quad \textit{for all } \theta \textit{ with } \mathtt{r} = \|\mathbb{D}_0(\theta-\theta_0)\|. \tag{9}$$

At a very high level, the conditions (ED_0) and (ED_1) are the exponential moment conditions that essentially require the tails of the reward distribution to be exponentially decaying. Moreover, $(\mathcal{L}_0)$ imposes a local identifiablity condition, that ensures that $-\mathbb{E}_a[\mathbb{L}(\theta)-\mathbb{L}(\theta_0)]$ are bounded above and below by a quadratic function of $\|\theta-\theta_0\|$. $(\mathcal{L}\mathtt{r})$ is a global identifiability condition that requires the gap $-\mathbb{E}_a[\mathbb{L}(\theta)-\mathbb{L}(\theta_0)]$ grows in a controlled fashion as $\mathtt{r} = \|\mathbb{D}_0(\theta-\theta_0)\|$ increases. More, discussion regarding these assumptions can be found in Section 2 of Spokoiny (2012) and partly in Panov and Spokoiny (2015).

Note that Assumption 4.4, which is needed to derive first-order concentration properties of the α-posterior distribution, effectively requires no regularity condition on the reward distribution. However, to conduct a more refined second-order BvM-type analysis, a more comprehensive set of local and global moment and identifiability conditions is required. In Section 6, we show that the exponential (Definition 6) and sub-Gaussian (Definition 8 with some mild smoothness conditions on the density of the error distribution) families satisfy all the conditions in Assumption 4.6. More examples can be found in Spokoiny (2012); Panov and Spokoiny (2015).

To extend the lower bound computed in (Panov and Spokoiny, 2015, Theorem 4) to general priors, we need to control the prior density on the set $\Theta_0(\mathtt{r}_0)$. In particular, observe that

$$\Pi_{t,\alpha}(\Theta_0(\mathtt{r}_0)|F_t) = \frac{\int_\Theta \exp\{\alpha\mathbb{L}(\theta,\theta_0)\}\Pi(d\theta)\mathbf{1}\{\theta\in\Theta_0(\mathtt{r}_0)\}}{\int_\Theta \exp\{\alpha\mathbb{L}(\theta,\theta_0)\}\}\Pi(d\theta)}. \tag{10}$$

Note, that $\Theta_0(\mathtt{r}_0) = \{\theta\in\Theta\colon \|\mathbb{D}_0(\theta-\theta_0)\| \le \mathtt{r}_0\}$. So for a fixed $\mathtt{r}_0$, the prior density ($\pi(\theta) := \Pi(d\theta)$) can easily be lower bounded on $\Theta_0(\mathtt{r}_0)$ by fixed number that depends on $\mathtt{r}_0$ and θ_0. In fact, if $\mathtt{r}_0 = \frac{c}{\sqrt{t}}$, for some constant $c > 0$, then note that $\min_{\theta\in\Theta_0(\mathtt{r}_0)}\pi(\theta) \ge \min_{\theta\in\Theta_0(c)}\pi(\theta)$, because, $\Theta_0(\mathtt{r}_0) \subseteq \Theta_0(c)$ for all $t \ge 1$. Therefore,

Assumption 4.7 (Prior) *We make following assumption on the prior distribution.*

1. *The prior distribution is continuous on* Θ.

2. *There exists a positive constant* $\mathtt{M}$*, such that the prior density,* $\pi(\theta) \leq \mathtt{M}$ *for all* $\theta \in \Theta$.

3. *For any compact set* $K \subset \Theta$, $\pi(\theta) > 0$ *for all* $\theta \in K$.

As our focus is solely on the lower bound outcome described in Panov and Spokoiny (2015), we observe that, under the stated assumption, the outcome in Theorem 4 of Panov and Spokoiny (2015) readily follows with a factor denoted by $\frac{\min_{\theta \in \Theta_0(c)} \pi(\theta)}{\mathtt{M}} := \mathbb{C}$. For the parametric problems of interest in this study, the two conditions (stated in Assumptions 4.4 and 4.7) imposed on the prior distribution are quite lenient, merely necessitating the prior density to be positive, continuous, and bounded. However, Assumption 4.5 requires the prior distribution to have sub-Gaussian tails.

5. Frequentist Regret Bound

In this section, we present our main result on bounding the regret of α-TS. The general bound on the expected regret of α-TS is provided below.

Theorem 1 (General regret bound) *Under Assumptions 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7, we show for any* $\eta > 0$ *and* $\alpha(1-\alpha)\lambda < 1$ *that*

$$\mathbb{E}[\boldsymbol{R}(T)] \leq \left(2d \log\left(1 + \lambda^{-1} T\right) \sum_{t=1}^{T} t\epsilon_t^2\right)^{1/2} \Big[\left(\bar{p}_0(d)(1 - 2e^{-\eta})\right)^{-1/2} + C_g \left((1-\alpha)^{-1}(D+2)C(\alpha, \lambda, \theta_0)\right)^{1/2} \left(1 + \left(\bar{p}_0(d)(1-2e^{-\eta})\right)^{-1/2}\right) \Big]. \tag{11}$$

where $\bar{p}_0(d) = \mathbb{C} \exp\left(-2\alpha\Delta(d,\eta) - 8e^{-\eta}\right) \left\{ \frac{C\sqrt{\alpha(d+\eta)}}{1+C^2\alpha(d+\eta)} e^{-C^2\alpha(d+\eta)/2} \right\} - e^{-\eta} - C(\alpha, \lambda, \theta_0) 2e^{-\frac{\alpha}{4} t_0 \epsilon_{t_0}^2}$ *for some fixed* $\Delta(d,\eta)$ *and* t_0*, and* $C(\alpha, \lambda, \theta_0)$ *grows with* $\alpha(1-\alpha)$ *and also depends on* C_π.

The upper bound on the regret above depends on the time horizon T and the dimension of the action space d, among other terms defined in the regularity conditions. In many parametric modeling instances, the term ϵ_t^2 is $O\left(\frac{d}{\alpha}\frac{\log t}{t}\right)$. Observe that $d\sum_{t=1}^{T} t\epsilon_t^2 = O(\frac{d^2}{\alpha}\sum_{t=1}^{T}\log t) = O(\frac{d^2}{\alpha} T \log T)$, which could provide the desired upper bound that matches the lower bound in Dani et al. (2008) for linear bandit problems with respect to both d and T. However, note that the term $\bar{p}_0(d)$ also depends on d, and for any $\alpha \in (0,1)$, it will result in a regret bound that grows exponentially with d. To balance this, we must choose a d-dependent α. We will observe in Corollary 7 for exponential family that $\Delta(d,\eta) = O((\eta+d)^{3/2})$. Therefore, we choose $\alpha = d^{-3/2}$, which results in a regret bound for exponential family to be of $O(d^{7/4}\sqrt{T}\log T)$, which is away by a factor of $d^{3/4}$ from the lower bound. For this choice of α, other α-dependent terms such as $C(\alpha,\lambda,\theta_0)$ and $(1-\alpha)^{-1}$ do not grow with d and are bounded away by 0. Note that $\alpha = d^{-3/2}$ is an optimal choice of α because to nullify the exponential effect due to $\Delta(d,\eta)$, we must choose $\alpha = d^{-(3/2+\varepsilon)}$ for any $\varepsilon \geq 0$ which would results into an upper bound of $O(d^{7/4+\varepsilon/2}\sqrt{T}\log T)$.

Similarly, in Corollary 9 we choose $\alpha = d^{-2}$ for the sub-Gaussian family to compute a regret bound of $O(d^2\sqrt{T}\log T)$, which is sub-optimal to the well-known regret upper bounds in Agrawal and Goyal (2013) and Abeille and Lazaric (2017) by a factor of $d^{1/2}$. Moreover, observe that if the posterior would have been a Gaussian, the $\bar{p}_0(d)$ would reduce to the expression in $\{\cdot\}$ and therefore, the choice of $\alpha = d^{-1}$ would suffice; and it will result into an upper bound that matches the existing upper bounds of $O(d^{3/2}\sqrt{T}\log T)$.

As noted in the introduction, our analysis adapts the idea of constructing saturated and unsaturated sets of action introduced by Agrawal and Goyal (2013) and combines it with the first and second-order properties of the α-posterior distribution. To present the proof sketch of the theorem above, we state the following two definitions.

Definition 2 (Precision matrix) *Define* $V_t = \lambda I + \sum_{s=1}^{t-1} a_s a_s^\top$ *for any* $\lambda > 0$, *and* $\|x\|_A := \sqrt{x^\top A x}$ *is the matrix norm of* x *for any square matrix* A.

This is a standard definition of the precision matrix that is derived while solving a Bayesian linear regression problem with Gaussian prior (with covariance λI) and standard Gaussian error. Next, we define the set of saturated actions.

Definition 3 (Saturated actions) *Any action* $a \in \mathbb{A}$ *is called saturated at time* t *or* $a \in C_t$, *where* C_t *is the set of saturated actions, if* $g(a_0^\top \theta_0) - g(a^\top \theta_0) \geq \sqrt{t}\epsilon_t \|a\|_{V_t^{-1}}$ *and otherwise it is called unsaturated.*

This definition is adapted from Agrawal and Goyal (2013). Here, ϵ_t is the same as defined in Assumption 4.4, which characterizes the α-posterior contraction rate. Note that at any time t, the best action a_0 is always unsaturated. Next, we provide a proof sketch of our main result, which also highlights the need for computing first and second-order properties of the α-posterior.

Proof sketch of Theorem 1. Recall the definition $\boldsymbol{R}(T)$ and let $\bar{a}_t := \arg\min_{a \notin C_t} \|a\|_{V_t^{-1}}$. Now observe that

$$g(a_0^\top \theta_0) - g(a_t^\top \theta_0) = \underbrace{g(a_0^\top \theta_0) - g(\bar{a}_t^\top \theta_0)}_{\text{(I)}} + \underbrace{g(\bar{a}_t^\top \theta_0) - g(a_t^\top \theta_t)}_{\text{(II)}} + \underbrace{g(a_t^\top \theta_t) - g(a_t^\top \theta_0)}_{\text{(III)}}. \quad (12)$$

Using Definition 3, observe that the expected cumulative sum of (I) can be bounded above by $\left(\sum_{t=1}^T t\epsilon_t^2\right)^{1/2} \left(\sum_{t=1}^T \mathbb{E}[\|\bar{a}_t\|^2_{V_t^{-1}}]\right)^{1/2}$. The expected cumulative sum of (II) can be bounded by $C_g \left(\mathbb{E}\left[\sum_{t=1}^T \|\bar{a}_t\|^2_{V_t^{-1}}\right]\right)^{1/2} \left(\sum_{t=1}^T \mathbb{E}\left[\|\theta_t - \theta_0\|^2_{V_t}\right]\right)^{1/2}$ using Assumption 4.2 and the fact that $g(a_t^\top \theta_t) > g(\bar{a}_t^\top \theta_t)$. Similarly, the expected cumulative sum of (III) can be bounded above by $C_g \left(\mathbb{E}\left[\sum_{t=1}^T \|a_t\|^2_{V_t^{-1}}\right]\right)^{1/2} \left(\sum_{t=1}^T \mathbb{E}\left[\|\theta_t - \theta_0\|^2_{V_t}\right]\right)^{1/2}$.

To further simplify the bound observe that the term $\left(\mathbb{E}\left[\sum_{t=1}^T \|a_t\|^2_{V_t^{-1}}\right]\right)^{1/2}$ can be bounded by $\sqrt{2d \log\left(1 + \lambda^{-1} T\right)}$ using elliptical potential lemma from (Auer et al., 2002, Lemma 3.1) and (Carpentier et al., 2020, Proposition 1). Observe that quantifying a bound on $\left(\sum_{t=1}^T \mathbb{E}\left[\|\theta_t - \theta_0\|^2_{V_t}\right]\right)^{1/2}$ requires understanding the first order concentration properties of the α-posterior, that we derive in Lemma 4. Now, the only term that remains to be analyzed is $\mathbb{E}\left[\|\bar{a}_t\|^2_{V_t^{-1}}\right]$. Observe that

$$\begin{aligned} \mathbb{E}\left[\|a_t\|^2_{V_t^{-1}} | \mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\|a_t\|^2_{V_t^{-1}} | \mathcal{F}_{t-1}, a_t \notin C_t\right] P(a_t \notin C_t | \mathcal{F}_{t-1}) \\ &\geq \|\bar{a}_t\|^2_{V_t^{-1}} P(a_t \notin C_t | \mathcal{F}_{t-1}), \end{aligned} \quad (13)$$

where the second inequality uses the definition of $\bar{a}_t$ and the fact that it is completely determined by $\mathcal{F}_{t-1}$. The remaining steps establish a lower bound on $P(a_t \notin C_t | \mathcal{F}_{t-1})$ to upper bound $\mathbb{E}\left[\|\bar{a}_t\|^2_{V_t^{-1}}\right]$ by $\mathbb{E}\left[\|a_t\|^2_{V_t^{-1}}\right]$ upto some time-dependent term and then use elliptical potential lemma.

If $g(a_0^\top \theta_t) > g(a_j^\top \theta_t)$ for all saturated actions, i.e.$\forall j \in C_t$, then one of the unsaturated actions must be played because a_0 is always unsaturated. Using this fact, we establish a lower bound on the probability of selecting an unsaturated action as

$$\begin{aligned} P(a_t \notin C_t|\mathcal{F}_{t-1}) &\geq P(g(a_0^\top \theta_t) > g(a_j^\top \theta_t), \forall j \in C_t|\mathcal{F}_{t-1}) \\ &\geq P(a_0^\top \theta_t > a_0^\top \theta_0|\mathcal{F}_{t-1}) - P(\|\theta_t - \theta_0\|_{V_t} \geq \sqrt{t}\epsilon_t|\mathcal{F}_{t-1}). \end{aligned} \tag{14}$$

To establish a lower bound on $P(a_t \notin C_t|\mathcal{F}_{t-1})$, we upper bound $P(\|\theta_t - \theta_0\|_{V_t} \geq \sqrt{t}\epsilon_t|\mathcal{F}_{t-1})$ by using α-posterior contraction result in Lemma 4 and lower bound $P(a_0^\top \theta_t > a_0^\top \theta_0|\mathcal{F}_{t-1})$ using the lower bound of the finite sample BvM stated in Lemma 5. The detailed proof of the theorem and the associated lemmas are provided in the Appendix A.

5.1. First and second-order properties of the α-posterior

In this section, we present two important results that characterize the first and second-order properties of the α-posterior distribution using the regularity conditions on the prior and reward distribution in the previous section.

Lemma 4 (α-posterior contraction) *Fix $\alpha \in (0,1)$. Under Assumptions 4.4 and 4.5 for any given sequence of actions $a^{(t)}$ that is adapted to $\mathcal{F}_t$ and $t > 0$, we have for all $j > 0$ and $\alpha(1-\alpha)\lambda < 1$ that,*

$$P_0\left(\Pi_{t,\alpha}\left(\frac{\lambda\alpha}{2}\|\theta - \theta_0\|^2 + D_\alpha^{(t)}(\theta, \theta_0) \geq \frac{(D+j)\alpha}{(1-\alpha)2} t\epsilon_t^2 |\mathcal{F}_t\right)\right) \leq 2C(\alpha, \lambda, \theta_0) e^{-\frac{j\alpha}{4}t\epsilon_t^2}, \tag{15}$$

and

$$\mathbb{E}\left[E\left[\frac{\lambda\alpha}{2}\|\theta - \theta_0\|^2 + D_\alpha^{(t)}(\theta, \theta_0)|\mathcal{F}_t\right]\right] \leq \frac{(D+2)\alpha}{(1-\alpha)2} t\epsilon_t^2 + \frac{4C(\alpha, \lambda, \theta_0) e^{-\frac{\alpha}{4}t\epsilon_t^2}}{(1-\alpha)}. \tag{16}$$

The first result computes a bound on the posterior probability of θ being away from θ_0. This bound computes the rate at which the α-posterior shrinks. The second result quantifies a concentration bound in expectation. Intuitively, since, $D_\alpha^{(t)}(\theta, \theta_0)$ scales with t, ϵ_t determines the rate at which the α-posterior distribution concentrates to the truth. The proof of both results is adapted from the posterior concentration analysis in the Bayesian statistics literature.

Remark: Typical analysis of the α-posterior concentration bounds measure deviation using only $D_\alpha^{(t)}(\theta, \theta_0)$ term. However, in the analysis of the regret, we needed to quantify a bound on the expectation of $\|\theta - \theta_0\|_{V_t}$, where V_t is the precision matrix defined in Definition 2, therefore, we modify the analysis to incorporate this additional term $\lambda\|\theta - \theta_0\|$ due to V_t-norm. Also, it is only due to this modified definition of the discrepancy measure; we need an additional assumption (Assumption 4.5) on the prior that requires its tails to be sub-Gaussian. Recall that C_π is the bound on the exponential moment on the prior as defined in Assumption 4.5.

Next, we present below the relevant part of the (Panov and Spokoiny, 2015, Theorem 4).

Lemma 5 (Finite BvM) *Under Assumptions $(ED_0), (ED_1), (\mathcal{L}_0), (\mathcal{L}\mathtt{r})$ on the reward distribution, and Assumption 4.7 on the prior density, for any measurable set $A \subseteq \mathbb{R}^d$ and $\alpha \in (0,1)$,*

$$\Pi_{t,\alpha}(\mathbb{D}_0(\theta - \theta_t^*) \in A|\mathcal{F}_t) \geq \mathbb{C}\exp\left(-2\alpha\Delta_t(d,\eta) - 8e^{-\eta}\right)\left\{P(\phi_d/\sqrt{\alpha} \in A)\right\} - e^{-\eta}, \tag{17}$$

with $P_{0,a}^{(t)}$ probability of at least $1-e^{-\eta}$, where ϕ_d is a standard Gaussian random variable in $\mathbb{R}^d$, $\theta_t^* = \theta_0 + \mathbb{D}_0^{-2}\nabla\mathbb{L}(\theta_0)$, $\Delta_t(d,\eta)$ is a known non-random function of η,d, and t and it decreases as t increases. Moreover, $\|\mathbb{D}_0(\theta_0-\theta_t^*)\| \le C\sqrt{d+\eta}$ with $P_{0,a}^{(t)}$-probability of at least $1-e^{-\eta}$ for some positive constant C.

The proof of the result above is a direct consequence of (Panov and Spokoiny, 2015, Theorem 4) and the result in the display (33) of (Panov and Spokoiny, 2015, Theorem 9). Analogous to the classical BvM result (Van der Vaart, 2000, Section 10.2), observe that $\mathbb{D}_0(\theta-\theta_t^*)$ perturbs parameter θ by θ_t^*, which is the first order approximation of the maximum likelihood estimate and scale it by $\mathbb{D}_0$ which accounts for the Fisher information matrix. Intuitively, as $t\to\infty$, the result above recovers (lower limit of) the classical Bernstein-von Mises theorem. The factor α accounts for the effect of tempering the likelihood in the definition of the α-posterior distribution. In particular, it can be observed from the lower bound above that $\alpha\in(0,1)$ inflates the posterior variance by a factor α^{-1}. Recall, $\Delta(d,\eta)$ appears in the expression for $\bar{p}_0(d)$ in Theorem 1. We will show in the examples that $\Delta_t(d,\eta)\le\Delta(d,\eta)$, for any $t\ge 1$, and to be precise, $\Delta_1(d,\eta)=\Delta(d,\eta)$.

6. Examples

We consider two broad classes of reward distributions: the sub-Gaussian and the exponential families. Below, we clearly state the definitions and assumptions for these families.

6.1. Exponential family

First, we specify the conditions for the exponential family reward models.

Definition 6 (Exponential) *We assume for $\Theta\subseteq\mathbb{R}^d$ that*

i. $p_\theta(r|a_t) = e^{ra_t^\top\theta - A(a_t^\top\theta)+C(r)}$, with conditional mean as $A'(a_t^\top\theta)$ and conditional variance as $A''(a_t^\top\theta)$, where $A(\cdot)$ is the log-partition function and $C(\cdot)$ is some known mapping. We also define the link function $g(\cdot):=A'(\cdot)$.

ii. the link function and its derivative are Lipschitz continuous, that is, for any $x,y\in\Omega$, (a) there exists a constant $C_g>0$, such that $|g(x)-g(y)|\le C_g|x-y|$ and (b) there exists a constant $L>0$, such that$|g'(x)-g'(y)|\le L|x-y|$.

iii. the log-partition function is strongly convex with parameter m, that is for any $\alpha\in(0,1)$, $x,y\in\Omega$, there exists an $m>0$, such that

$$\alpha A(x)+(1-\alpha)A(y)-A(\alpha x+(1-\alpha)y)\ge\frac{1}{2}\alpha(1-\alpha)m|x-y|^2, \tag{18}$$

Typically, if $\max_{x\in\mathbb{R}}|g'(x)|<\infty$, then the condition *ii*(a) above is satisfied for $C_g=\max_{x\in\mathbb{R}}|g'(x)|$. Moreover, when $|A''(\cdot)|\ge m>0$, then $A(\cdot)$ is strongly convex with parameter m. Note that conditions (*ii*(a)) and (*iii*) above follow when $|A''(\cdot)|\in[m,C_g]$. This condition is restrictive for exponential family distribution as it requires the variance of many exponential family distributions (such as Poisson or Exponential) to lie in a compact space.

We show in the Appendix B that the above exponential family of distributions satisfies all the conditions required for the first and second-order properties of α-posterior. Consequently, we have the following corollary of Theorem 1.

Corollary 7 *For the exponential family reward distributions (Definition 6), under Assumptions 4.4, 4.5, and 4.7 on the prior distribution and for some prior dependent positive constant C_ϕ, we show for any $\eta > 0$ and $\alpha(1-\alpha)\lambda < 1$ that $\epsilon_t^2 = \frac{4dC_g \log(t)}{D\alpha C_\phi t}$, $\Delta(d,\eta) = O((\eta+d)^{3/2})$, and for $\alpha = d^{-3/2}$, $\mathbb{E}[\boldsymbol{R}(T)] = O(d^{7/4}\sqrt{T}\log(T))$.*

The expression for ϵ_t is derived in Lemma 12 and $\Delta(d,\eta)$ is computed in Lemma 14, where we satisfy all the conditions required for the second-order analysis of α-posterior for exponential family models.

6.2. Sub-Gaussian Family

We begin with the definition of the sub-Gaussian family and state the required conditions.

Definition 8 (Sub-Gaussian) *For any $\Theta \subseteq \mathbb{R}^d$, we assume that*

i. the reward distribution is conditionally sub-Gaussian with bounded sub-Gaussian parameter σ_θ, that is for any $t \geq 0$, and given a_t, $\mathbb{E}[e^{sr_t}|a_t] \leq e^{sa_t^\top\theta + \sigma_\theta s^2/2}$ for all $s \in \mathbb{R}$,

ii. σ_θ is bounded by 1, which implies $\mathbb{E}[e^{sr_t}|a_t] \leq e^{sa_t^\top\theta + s^2/2}$ for all $s \in \mathbb{R}$, and

iii. the true mean reward $a^\top\theta_0$ lies in $[0,1]$ for any $a \in \mathbb{A}$.

We show in the Appendix C that the above sub-Gaussian family of distributions satisfies all the conditions required for the first and second-order properties of α-posterior. Consequently, we have the following corollary of Theorem 1 for the sub-Gaussian family.

Corollary 9 *For the sub-Gaussian family of reward distributions (Definition 8) with some additional regularity conditions on the error density, under Assumptions 4.4, 4.5, and 4.7 on the prior distribution and for some prior dependent positive constant C_ϕ and a positive constant $\tilde{C}$, we show for any $\eta > 0$ and $\alpha(1-\alpha)\lambda < 1$ that $\epsilon_t^2 = \frac{4d\tilde{C}\log(t)}{D\alpha C_\phi t}$, $\Delta(d,\eta) = O(d^2)$, and for $\alpha = d^{-2}$, $\mathbb{E}[\boldsymbol{R}(T)] = O(d^2\sqrt{T}\log(T))$.*

We derive the expression for ϵ_t in Lemma 17 and for $\Delta(d,\eta)$ in Lemma 18, where we show that sub-Gaussian family satisfies all the conditions required for the second-order analysis of α-posterior. A similar regret bound can be computed when the mean of the reward distribution is generalized to $g(a_t^\top\theta)$ by assuming that g satisfies Assumption 4.2.

7. Conclusion

This work connects the rich literature in Bayesian statistics on posterior concentration theory and recent advancements in the regret analysis of Thompson Sampling for generalized linear bandit problems. In addition to computing regret bounds, this research direction encourages using advanced Bayesian statistical inference techniques to model complex sequential decision-making problems and propose novel TS-type algorithms with complex and useful priors. In future work, we aim to extend this analysis to incorporate nonparametric and misspecified reward distributions. This paper assumes that the exact samples are available from the posterior distribution, which may not be possible for many interesting combinations of the prior and reward distributions. Therefore, it would be interesting to extend the analysis in this paper to incorporate approximate samples from the posterior distribution (by using variational Bayesian or Markov Chain Monte Carlo methods).

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Marc Abeille and Alessandro Lazaric. Linear thompson sampling revisited. In *Artificial Intelligence and Statistics*, pages 176–184. PMLR, 2017.

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*, pages 127–135. PMLR, 2013.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

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

Anirban Bhattacharya, Debdeep Pati, and Yun Yang. Bayesian fractional posteriors. *The Annals of Statistics*, 47(1):39–66, 2019.

Alexandra Carpentier, Claire Vernade, and Yasin Abbasi-Yadkori. The Elliptical Potential Lemma Revisited, October 2020. URL http://arxiv.org/abs/2010.10182. arXiv:2010.10182 [cs, stat].

Sunrit Chakraborty, Saptarshi Roy, and Ambuj Tewari. Thompson sampling for high-dimensional sparse linear contextual bandits. In *International Conference on Machine Learning*, pages 3979–4008. PMLR, 2023.

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

Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic linear optimization under bandit feedback. In *COLT*, 2008.

Subhashis Ghosal, Jayanta K. Ghosh, and Aad W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531, 2000. ISSN 00905364. URL http://www.jstor.org/stable/2674039.

Kristjan Greenewald, Ambuj Tewari, Susan Murphy, and Predag Klasnja. Action centered contextual bandits. *Advances in neural information processing systems*, 30, 2017.

Nima Hamidi and Mohsen Bayati. On Frequentist Regret of Linear Thompson Sampling, April 2023. URL http://arxiv.org/abs/2006.06790. arXiv:2006.06790 [cs, stat].

Joey Hong, Branislav Kveton, Manzil Zaheer, Mohammad Ghavamzadeh, and Craig Boutilier. Thompson sampling with a mixture prior. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 7565–7586. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/hong22b.html.

Dong Woo Kim, Tze Leung Lai, and Huanzhong Xu. Multi-armed bandits with covariates. *Statistica Sinica*, 31:2275–2287, 2021.

Tze Leung Lai, Herbert Robbins, et al. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

Lihong Li and Olivier Chapelle. Open problem: Regret bounds for thompson sampling. In *Conference on Learning Theory*, pages 43–1. JMLR Workshop and Conference Proceedings, 2012.

Yuwei Luo and Mohsen Bayati. Geometry-Aware Approaches for Balancing Performance and Theoretical Guarantees in Linear Bandits, December 2023. URL http://arxiv.org/abs/2306.14872. arXiv:2306.14872 [cs, stat].

J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370, 1972. ISSN 0035-9238. doi: 10.2307/2344614. URL http://dx.doi.org/10.2307/2344614.

Maxim Panov and Vladimir Spokoiny. Finite sample bernstein – von mises theorem for semiparametric problems. *Bayesian Analysis*, 10(3), September 2015. doi: 10.1214/14-ba926. URL https://doi.org/10.1214/14-ba926.

Philippe Rigollet and Assaf Zeevi. Nonparametric bandits with covariates. *COLT 2010*, page 54, 2010.

Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527 – 535, 1952. doi: bams/1183517370. URL https://doi.org/.

Vladimir Spokoiny. Parametric estimation. finite sample theory. *The Annals of Statistics*, 40 (6), December 2012. doi: 10.1214/12-aos1054. URL https://doi.org/10.1214/12-aos1054.

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, December 1933. doi: 10.2307/2332286. URL https://doi.org/10.2307/2332286.

Iñigo Urteaga and Chris H Wiggins. Nonparametric gaussian mixture models for the multi-armed contextual bandit. *stat*, 1050:8, 2018.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Tim Van Erven and Peter Harremos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

Fengshuo Zhang and Chao Gao. Convergence rates of variational posterior distributions. *The Annals of Statistics*, 48(4), August 2020. doi: 10.1214/19-aos1883. URL https://doi.org/10.1214/19-aos1883.

Tong Zhang. Feel-good thompson sampling for contextual bandits and reinforcement learning. *arXiv preprint arXiv:2110.00871*, 2021.

Appendix A. Proof of Theorem 1

Proof [Proof of Lemma 4] Let us first define a set for any $D > 0$, $j > 0$, and $a^{(t)}$ adapted to $\mathcal{F}_t$ as

$$F_{t,\epsilon_t} = \left\{ \theta \in \Theta : \frac{\alpha\lambda}{2}\|\theta - \theta_0\|^2 + D_\alpha^{(t)}(\theta, \theta_0) \geq \frac{(D+j)\alpha}{(1-\alpha)2} t\epsilon_t^2 \right\}. \tag{19}$$

Note that the above set is adapted to $\mathcal{F}_t$ because the definition of $D_\alpha^{(t)}(\theta, \theta_0)$ requires knowledge of $a^{(t)}$. The α-posterior measure of the set F_{t,ϵ_t} can be defined as

$$\begin{aligned}\Pi_{t,\alpha}(F_{t,\epsilon_t}|\mathcal{F}_t) &= \frac{\int_{F_{t,\epsilon_t}} \prod_{s=1}^t [p_\theta(r_s|a_s)]^\alpha \Pi(d\theta)}{\int \prod_{s=1}^t [p_\theta(r_s|a_s)]^\alpha \Pi(d\theta)} \\ &= \frac{\int_{F_{t,\epsilon_t}} \Pi(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)}}{\int \Pi(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)}},\end{aligned} \tag{20}$$

where $\ell_t(\theta, \theta_0) = \log \frac{\prod_{s=1}^t p_0(r_s|\mathcal{F}_{s-1}, a_s)}{\prod_{s=1}^t p_\theta(r_s|a_s)}$.

Now define a set

$$A_{t,\epsilon_t} = \left\{ X_t : \int_\Theta \tilde{\Pi}(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)} \leq e^{-\frac{(D+j)\alpha}{4} t\epsilon_t^2} \right\},$$

where for any $B \subseteq \Theta$, $\tilde{\Pi}(B) = \Pi(B \cap \Theta)/\Pi(B(\theta_0, \epsilon_t))$ and $B(\theta_0, \epsilon_t)$ is as defined in Assumption 4.4. (Note: Assumption 4.4 is a stronger statement as it requires condition (B1) to be satisfied for any $a^{(t)}$; however, for this lemma, we need this assumption only for $a^{(t)}$ that are adapted to $\mathcal{F}_t$.)

Observe that

$$\mathbb{E}[\Pi_{t,\alpha}(F_{t,\epsilon_t}|\mathcal{F}_t)] = P_0^{(t)}(A_{t,\epsilon_t}) + \mathbb{E}\left[\frac{\int_{F_{t,\epsilon_t}} \Pi(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)}}{\int \Pi(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)}} \mathbf{1}(A_{t,\epsilon_t}^c) \right]. \tag{21}$$

First, let us analyze the second term in equation 21. Note that on the set A_{t,ϵ_t}^c,

$$\int \Pi(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)} \geq \Pi(B(\theta_0, \epsilon_t)) e^{-\frac{(D+j)\alpha}{4} t\epsilon_t^2}.$$

Therefore, using Assumption 4.4, it follows that

$$\begin{aligned}\mathbb{E}\left[\frac{\int_{F_{t,\epsilon_t}} \Pi(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)}}{\int \Pi(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)}} \mathbf{1}(A_t^c) \right] &\leq e^{\frac{(D+j)\alpha}{4} t\epsilon_t^2} \mathbb{E}\left[\frac{\int_{F_{t,\epsilon_t}} \Pi(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)}}{\Pi(B(\theta_0, \epsilon_t))} \right] \\ &\leq e^{(\frac{j\alpha}{4} + \frac{D\alpha}{2}) t\epsilon_t^2} \mathbb{E}\left[\int_{F_{t,\epsilon_t}} \Pi(d\theta) e^{-\alpha\ell_t(\theta,\theta_0)} \right].\end{aligned} \tag{22}$$

Now it follows from the definition of α-Rényi divergence that

$$\mathbb{E}\left[e^{-\alpha\ell_t(\theta,\theta_0)} | a^{(t)} \right] = e^{-(1-\alpha)D_\alpha^{(t)}(\theta,\theta_0)}.$$

Hence, using Fubini's theorem and the observation above, it follows that

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[\int_{F_{t,\epsilon_t}}\Pi(d\theta)e^{-\alpha\ell_t(\theta,\theta_0)}\right]|a^{(t)}\right] &= \mathbb{E}\left[\int_{F_{t,\epsilon_t}}\Pi(d\theta)e^{-(1-\alpha)D_\alpha^{(t)}(\theta,\theta_0)}\right]\\ &\leq \mathbb{E}\left[\int_{F_{t,\epsilon_t}}\Pi(d\theta)e^{-(1-\alpha)\left(\frac{(D+j)\alpha}{(1-\alpha)2}t\epsilon_t^2-\frac{\alpha\lambda}{2}\|\theta-\theta_0\|^2\right)}\right]\\ &\leq e^{-\frac{(D+j)\alpha}{2}t\epsilon_t^2}\mathbb{E}\left[\int_{F_{t,\epsilon_t}}\Pi(d\theta)e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta-\theta_0\|^2}\right], \qquad (23)\end{aligned}$$

where the penultimate inequality is due to the definition of F_{t,ϵ_t}. Substituting equation 23 into equation 22 yields,

$$\mathbb{E}\left[\frac{\int_{F_{t,\epsilon_t}}\Pi(d\theta)e^{-\alpha\ell_t(\theta,\theta_0)}}{\int\Pi(d\theta)e^{-\alpha\ell_t(\theta,\theta_0)}}\mathbf{1}(A_t^c)\right] \leq e^{-\frac{j\alpha}{4}t\epsilon_t^2}\int\Pi(d\theta)e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta-\theta_0\|^2}. \qquad (24)$$

Next, we analyze the first term in equation 21. It follows from the Markov inequality that

$$\begin{aligned}P_0^{(t)}&\left(\left[\int\tilde{\Pi}(d\theta)e^{-\alpha\ell_t(\theta,\theta_0)}\right]^{-1/\alpha}\geq e^{\frac{D+j}{4}t\epsilon_t^2}\right)\\ &\leq e^{-\frac{D+j}{4}t\epsilon_t^2}\mathbb{E}\left(\left[\int\tilde{\Pi}(d\theta)e^{-\alpha\ell_t(\theta,\theta_0)}\right]^{-1/\alpha}\right)\\ &\leq e^{-\frac{D+j}{4}t\epsilon_t^2}\mathbb{E}\left[\int\tilde{\Pi}(d\theta)\mathbb{E}\left(e^{\ell_t(\theta,\theta_0)}|a^{(t)}\right)\right]\\ &= e^{-\frac{D+j}{4}t\epsilon_t^2}\mathbb{E}\left[\int\tilde{\Pi}(d\theta)e^{D_2^{(t)}(\theta_0,\theta)}\right]\\ &\leq e^{-\frac{(D+j)\alpha}{4}t\epsilon_t^2}e^{\frac{D\alpha}{4}t\epsilon_t^2}=e^{-\frac{j\alpha}{4}t\epsilon_t^2}, \qquad (25)\end{aligned}$$

where the second inequality is due to Jensen's and Fubini's theorems, and the penultimate inequality uses the definition of the set $B(\theta_0,\epsilon_t)$ and the fact that $\alpha\in(0,1)$.

Combining equation 24 and equation 25, it follows from equation 21 that for all $j>0$ and $\alpha\in(0,1)$,

$$\begin{aligned}\mathbb{E}\left[E\left[\mathbf{1}\{\frac{\lambda\alpha}{2}\|\theta-\theta_0\|^2+D_\alpha^{(t)}(\theta,\theta_0)\geq\frac{(D+j)\alpha}{(1-\alpha)2}t\epsilon_t^2\}|\mathcal{F}_t\right]\right] &= \mathbb{E}[\Pi_{t,\alpha}(F_{t,\epsilon_t}|\mathcal{F}_t)]\\ &\leq 2e^{-\frac{j\alpha}{4}t\epsilon_t^2}\int\Pi(d\theta)e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta-\theta_0\|^2}.\end{aligned} \qquad (26)$$

This proves the first assertion of the lemma. Now for the second assertion, note that the RHS above is non-increasing in j; therefore it follows from the inequality above that for all $s\geq 1$,

$$\mathbb{E}\left[E\left[\mathbf{1}\{\frac{\lambda\alpha}{2}\|\theta-\theta_0\|^2+D_\alpha^{(t)}(\theta,\theta_0)\geq\frac{(D+s)\alpha}{(1-\alpha)2}t\epsilon_t^2\}|\mathcal{F}_t\right]\right]\leq 2e^{-\frac{(s-1)\alpha}{4}t\epsilon_t^2}\int\Pi(d\theta)e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta-\theta_0\|^2}. \qquad (27)$$

Since $\frac{\lambda\alpha}{2}\|\theta-\theta_0\|^2 + D_\alpha^{(t)}(\theta,\theta_0) > 0$, it is straightforward to see that

$$E\left[\frac{\lambda\alpha}{2}\|\theta-\theta_0\|^2 + D_\alpha^{(t)}(\theta,\theta_0)|\mathcal{F}_t\right] = \int_0^\infty E\left[\mathbf{1}\{\frac{\lambda\alpha}{2}\|\theta-\theta_0\|^2 + D_\alpha^{(t)}(\theta,\theta_0) \geq u\}|\mathcal{F}_t\right] du$$
$$\leq \frac{(D+2)\alpha}{(1-\alpha)2}t\epsilon_t^2 + \int_{\frac{(D+2)\alpha}{(1-\alpha)2}t\epsilon_t^2}^\infty E\left[\mathbf{1}\{\frac{\lambda\alpha}{2}\|\theta-\theta_0\|^2 + D_\alpha^{(t)}(\theta,\theta_0) \geq u\}|\mathcal{F}_t\right] du.$$

Now using Fubini's theorem, we have

$$\begin{aligned}
&\mathbb{E}\left[E\left[\frac{\lambda\alpha}{2}\|\theta-\theta_0\|^2 + D_\alpha^{(t)}(\theta,\theta_0)|\mathcal{F}_t\right]\right] \\
&\leq \frac{(D+2)\alpha}{(1-\alpha)2}t\epsilon_t^2 + \int_{\frac{(D+2)\alpha}{(1-\alpha)2}t\epsilon_t^2}^\infty \mathbb{E}\left[E\left[\mathbf{1}\{\frac{\lambda\alpha}{2}\|\theta-\theta_0\|^2 + D_\alpha^{(t)}(\theta,\theta_0) \geq u\}|\mathcal{F}_t\right]\right] du \\
&= \frac{(D+2)\alpha}{(1-\alpha)2}t\epsilon_t^2 + \int_2^\infty \mathbb{E}\left[E\left[\mathbf{1}\{\frac{\lambda\alpha}{2}\|\theta-\theta_0\|^2 + D_\alpha^{(t)}(\theta,\theta_0) \geq \frac{(D+s)\alpha}{(1-\alpha)2}t\epsilon_t^2\}|\mathcal{F}_t\right]\right] \frac{\alpha t\epsilon_t^2}{(1-\alpha)2}ds \\
&\leq \frac{(D+2)\alpha}{(1-\alpha)2}t\epsilon_t^2 + \frac{\alpha t\epsilon_t^2}{(1-\alpha)}e^{-\frac{\alpha}{4}t\epsilon_t^2}\int \Pi(d\theta)e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta-\theta_0\|^2}\int_2^\infty e^{-\frac{(s-2)\alpha}{4}t\epsilon_t^2}ds \\
&\leq \frac{(D+2)\alpha}{(1-\alpha)2}t\epsilon_t^2 + \frac{4e^{-\frac{\alpha}{4}t\epsilon_t^2}}{(1-\alpha)}\int \Pi(d\theta)e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta-\theta_0\|^2}, \qquad (28)
\end{aligned}$$

where the penultimate inequality is due to equation 27 and the last inequality holds for all $\alpha t\epsilon_t^2/4 \geq 0.6$. Now the result follows by using the observation that $\int \Pi(d\theta)e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta-\theta_0\|^2} \leq C_\pi f e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta_0\|^2} := C(\alpha,\lambda,\theta_0)$ for some increasing mapping f of $\alpha(1-\alpha)\lambda\|\theta_0\|$.

■

The next result computes a lower bound on the probability of sampling from the posterior along the best action sequence.

Lemma 10 *Under the conditions of Lemma 5 for $\theta_t^* = \theta_0 + \mathbb{D}_0^{-2}\nabla\mathbb{L}(\theta_0)$, we have*

$$\Pi_\alpha(a_0^\top\theta_t > a_0^\top\theta_0)|\mathcal{F}_{t-1}) = \Pi_\alpha(a_0^\top\mathbb{D}_0^{-1}\mathbb{D}_0\,(\theta_t-\theta_t^*) > -a_0^\top\mathbb{D}_0^{-2}\nabla\mathbb{L}(\theta_0)|\mathcal{F}_{t-1})$$
$$\geq \mathbb{C}\exp\left(-2\alpha\Delta(d,\eta) - 8e^{-\eta}\right)\left\{\frac{C\sqrt{\alpha(d+\eta)}}{1+C^2\alpha(d+\eta)}e^{-C^2\alpha(d+\eta)/2}\right\} - e^{-\eta} \qquad (29)$$

with $P_{0,a}^{(t)}$ probability of at least $1-2e^{-\eta}$ for any $\eta > 0$.

Proof [Proof of Lemma 10] Using Lemma 5, we have

$$\begin{aligned}
&\Pi_\alpha(a_0^\top \mathbb{D}_0^{-1}\mathbb{D}_0\,(\theta_t-\theta_t^*) > -a_0^\top \mathbb{D}_0^{-2}\nabla\mathbb{L}(\theta_0)|\mathcal{F}_{t-1})\\
&\quad\geq \Pi_\alpha(a_0^\top \mathbb{D}_0^{-1}\mathbb{D}_0\,(\theta_t-\theta_t^*) > \|a_0^\top \mathbb{D}_0^{-1}\|\|\mathbb{D}_0^{-1}\nabla\mathbb{L}(\theta_0)\||\mathcal{F}_{t-1})\\
&\quad\geq \Pi_\alpha(a_0^\top \mathbb{D}_0^{-1}\mathbb{D}_0\,(\theta_t-\theta_t^*) > C\|a_0^\top \mathbb{D}_0^{-1}\|\sqrt{d+\eta}|\mathcal{F}_{t-1})\\
&\quad\geq \mathbb{C}\exp\left(-2\alpha\Delta_t(d,\eta)-8e^{-\eta}\right)\left\{P(\phi_d/\sqrt{\alpha}\in A)\right\}-e^{-\eta}\\
&\quad= \mathbb{C}\exp\left(-2\alpha\Delta_t(d,\eta)-8e^{-\eta}\right)\left\{P(\phi\geq C\sqrt{\alpha(d+\eta)})\right\}-e^{-\eta} \qquad (30)\\
&\quad\geq \mathbb{C}\exp\left(-2\alpha\Delta_t(d,\eta)-8e^{-\eta}\right)\left\{\frac{C\sqrt{\alpha(d+\eta)}}{1+C^2\alpha(d+\eta)}e^{-C^2\alpha(d+\eta)/2}\right\}-e^{-\eta} \qquad (31)
\end{aligned}$$

with $P_{0,a}^{(t)}$ probability of at least $1-2e^{-\eta}$ for any $\eta>0$, where $A:=\{x\in\mathbb{R}^d : a_0^\top\mathbb{D}_0^{-1}x\geq C\|a_0^\top\mathbb{D}_0^{-1}\|\sqrt{d+\eta}\}$. Since, $\Delta_t(d,\eta)$ is a decreasing function of t, the result follows for $\Delta(d,\eta):=\Delta_1(d,\eta)\geq\Delta_t(d,\eta)$. ■

Proof [Proof of Theorem 1] Let $\bar{a}_t:=\arg\min_{a\notin C_t}\|a\|_{V_t^{-1}}$ and observe that

$$\left(g(a_0^\top\theta_0)-g(a_t^\top\theta_0)\right)=\left(g(a_0^\top\theta_0)-g(\bar{a}_t^\top\theta_0)\right)+\left(g(\bar{a}_t^\top\theta_0)-g(a_t^\top\theta_t)\right)+\left(g(a_t^\top\theta_t)-g(a_t^\top\theta_0)\right). \qquad (32)$$

Using the definition of the unsaturated actions 3 observe that the first term in equation 36 is bounded as

$$\begin{aligned}
\sum_{t=1}^T\mathbb{E}\left[\left(g(a_0^\top\theta_0)-g(\bar{a}_t^\top\theta_0)\right)\right]&\leq\sum_{t=1}^T\sqrt{t}\epsilon_t\mathbb{E}[\|\bar{a}_t\|_{V_t^{-1}}]\leq\sum_{t=1}^T\sqrt{t}\epsilon_t\mathbb{E}[\|\bar{a}_t\|^2_{V_t^{-1}}]^{1/2}\\
&\leq\left(\sum_{t=1}^T t\epsilon_t^2\right)^{1/2}\left(\sum_{t=1}^T\mathbb{E}[\|\bar{a}_t\|^2_{V_t^{-1}}]\right)^{1/2}, \qquad (33)
\end{aligned}$$

where the last inequality uses Cauchy-Schwarz (CS) inequality. Consider the third term in equation 36 and note that

$$\begin{aligned}
\sum_{t=1}^T\mathbb{E}\left[(g(a_t^\top\theta_t)-g(a_t^\top\theta_0))\right]&\leq C_g\sum_{t=1}^T\mathbb{E}\left[\left|(a_t^\top\theta_t-a_t^\top\theta_0)\right|\right]\\
&\leq C_g\sum_{t=1}^T\mathbb{E}\left[\|a_t\|_{V_t^{-1}}\|\theta_t-\theta_0\|_{V_t}\right]\\
&\leq C_g\mathbb{E}\left[\left(\sum_{t=1}^T\|a_t\|^2_{V_t^{-1}}\right)^{1/2}\left(\sum_{t=1}^T\|\theta_t-\theta_0\|^2_{V_t}\right)^{1/2}\right]\\
&\leq C_g\left(\mathbb{E}\left[\sum_{t=1}^T\|a_t\|^2_{V_t^{-1}}\right]\right)^{1/2}\left(\sum_{t=1}^T\mathbb{E}\left[\|\theta_t-\theta_0\|^2_{V_t}\right]\right)^{1/2}\\
&\leq C_g\sqrt{2d\log\left(1+\lambda^{-1}T\right)}\left(\sum_{t=1}^T\mathbb{E}\left[\|\theta_t-\theta_0\|^2_{V_t}\right]\right)^{1/2}, \qquad (34)
\end{aligned}$$

where the first inequality is due to Assumption 4.2, the second inequality uses CS with $V_t = \lambda I + \sum_{s=1}^{t-1} a_s a_s^\top$ for any $\lambda > 0$, and $\|x\|_A = \sqrt{x^T A x}$, the third and fourth inequality are also due to CS. The bound on $\left(\mathbb{E}\left[\sum_{t=1}^T \|a_t\|^2_{V_t^{-1}}\right]\right)^{1/2}$ uses elliptical potential lemma from (Auer et al., 2002, Lemma 3.1) and (Carpentier et al., 2020, Proposition 1).

Combined with the fact that $g(a_t^\top \theta_t) > g(\bar{a}_t^\top \theta_t)$, the second term in equation 36 can be bounded in a similar way to obtain

$$\sum_{t=1}^T \mathbb{E}\left[\left(g(\bar{a}_t^\top \theta_0) - g(a_t^\top \theta_t)\right)\right] \leq C_g \left(\mathbb{E}\left[\sum_{t=1}^T \|\bar{a}_t\|^2_{V_t^{-1}}\right]\right)^{1/2} \left(\sum_{t=1}^T \mathbb{E}\left[\|\theta_t - \theta_0\|^2_{V_t}\right]\right)^{1/2} \tag{35}$$

Now observe that, using the second result in Lemma 4, we have

$$\begin{aligned}
\sum_{t=1}^T \mathbb{E}\left[\|\theta_t - \theta_0\|^2_{V_t}\right] &= \sum_{t=1}^T \mathbb{E}\left[(\theta_t - \theta_0)^\top V_t (\theta_t - \theta_0)\right] \\
&= \sum_{t=1}^T \mathbb{E}\left[\lambda \|\theta_t - \theta_0\|^2 + \sum_{s=1}^{t-1} \|(\theta_t - \theta_0)^\top a_s\|^2\right] \\
&\leq \sum_{t=1}^T \mathbb{E}\left[\lambda \|\theta_t - \theta_0\|^2 + \frac{2}{\alpha} D_\alpha^{t-1}(\theta_t, \theta_0)\right] \\
&\leq \frac{2}{\alpha} \sum_{t=1}^T \left[\frac{(D+2)\alpha}{(1-\alpha)2} t\epsilon_t^2 + \frac{4e^{-\frac{\alpha}{4} t\epsilon_t^2}}{(1-\alpha)} \int \Pi(d\theta) e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta - \theta_0\|^2}\right] \\
&\leq \frac{(D+2)C(\alpha, \lambda, \theta_0)}{(1-\alpha)} \sum_{t=1}^T t\epsilon_t^2,
\end{aligned} \tag{36}$$

where the first inequality uses Assumption 4.3. (Note: We omit factor C while using Assumption 4.3 just for ease of exposition. The arguments can be easily updated to incorporate this factor.)

The only term that remains to be analyzed is $\mathbb{E}\left[\|\bar{a}_t\|^2_{V_t^{-1}}\right]$. Observe that

$$\begin{aligned}
\mathbb{E}\left[\|a_t\|^2_{V_t^{-1}} | \mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\|a_t\|^2_{V_t^{-1}} | \mathcal{F}_{t-1}, a_t \notin C_t\right] P(a_t \notin C_t | \mathcal{F}_{t-1}) \\
&\geq \|\bar{a}_t\|^2_{V_t^{-1}} P(a_t \notin C_t | \mathcal{F}_{t-1}),
\end{aligned} \tag{37}$$

where the second inequality uses the definition of $\bar{a}_t$ and the fact that it is completely determined by $\mathcal{F}_{t-1}$.

Note that if we can establish a lower bound on $P(a_t \notin C_t | \mathcal{F}_{t-1})$ then we can upper bound $\mathbb{E}\left[\|\bar{a}_t\|^2_{V_t^{-1}}\right]$ by $\mathbb{E}\left[\|a_t\|^2_{V_t^{-1}}\right]$ upto some time dependent term.

Since a_0 is always unsaturated, therefore if $g(a_0^\top \theta_t) > g(a_j^\top \theta_t)$ for all saturated actions, i.e.$\forall j \in C_t$, then one of the unsaturated actions must be played. Consequently,

$$
\begin{aligned}
P(a_t \notin C_t|\mathcal{F}_{t-1}) &\geq P(g(a_0^\top \theta_t) > g(a_j^\top \theta_t), \forall j \in C_t|\mathcal{F}_{t-1}) \\
&\geq P(\{\forall j \in C_t : g(a_0^\top \theta_t) > g(a_j^\top \theta_t)\}, \{\forall j : g(a_j^\top \theta_0) \geq g(a_j^\top \theta_t) - C_g\sqrt{t}\epsilon_t \|a_j\|_{V_t^{-1}}\}|\mathcal{F}_{t-1}) \\
&\geq P(\{\forall j \in C_t : g(a_0^\top \theta_t) > g(a_j^\top \theta_0)\}\{\forall j : g(a_j^\top \theta_0) \geq g(a_j^\top \theta_t) - C_g\sqrt{t}\epsilon_t \|a_j\|_{V_t^{-1}}\}|\mathcal{F}_{t-1}) \\
&\geq P(a_0^\top \theta_t > a_0^\top \theta_0, \|\theta_t - \theta_0\|_{V_t} \leq \sqrt{t}\epsilon_t|\mathcal{F}_{t-1}) \\
&\geq P(a_0^\top \theta_t > a_0^\top \theta_0|\mathcal{F}_{t-1}) - P(\|\theta_t - \theta_0\|_{V_t} \geq \sqrt{t}\epsilon_t|\mathcal{F}_{t-1}) \\
&\geq p - P\left(\lambda\frac{\alpha}{2}\|\theta_t - \theta_0\|^2 + D_\alpha^{t-1}(\theta_t, \theta_0) \geq \frac{\alpha}{2}\sqrt{t}\epsilon_t|\mathcal{F}_{t-1}\right) \\
&\geq p - 2e^{-\frac{\alpha}{4}t\epsilon_t^2}\int \Pi(d\theta)e^{\frac{\alpha(1-\alpha)\lambda}{2}\|\theta-\theta_0\|^2} \geq p - C(\alpha, \lambda, \theta_0)2e^{-\frac{\alpha}{4}t\epsilon_t^2} \qquad (38)
\end{aligned}
$$

with P_0- probability of at least $1-2e^{-\eta}$ for $p = \mathbb{C}\exp\left(-2\alpha\Delta(d,\eta) - 8e^{-\eta}\right)\left\{\frac{C\sqrt{\alpha(d+\eta)}}{1+C^2\alpha(d+\eta)}e^{-C^2\alpha(d+\eta)/2}\right\} - e^{-\eta}$. The penultimate inequality uses the definition of the saturated actions and the result in Lemmas 10 and 4. So, combining equation 37 and equation 38 and denoting $\bar{p} = p - C(\alpha, \lambda, \theta_0)2e^{-\frac{\alpha}{4}t\epsilon_t^2}$ for brevity, we have

$$
\begin{aligned}
\mathbb{E}\left[\|a_t\|^2_{V_t^{-1}}|\mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\|a_t\|^2_{V_t^{-1}}|\mathcal{F}_{t-1}, a_t \notin C_t\right]P(a_t \notin C_t|\mathcal{F}_{t-1}) \qquad (39) \\
&\geq \|\bar{a}_t\|^2_{V_t^{-1}}\bar{p} \qquad (40)
\end{aligned}
$$

with P_0- probability of at least $1-2e^{-\eta}$. Since, $t\epsilon_t^2$ increases with t, for sufficiently large t (say t_1), $\bar{p}$ will be positive. Denoting the above high probability event as $\mathbf{E}$ observe that

$$
\mathbb{E}\left[\mathbb{E}\left[\|a_t\|^2_{V_t^{-1}}|\mathcal{F}_{t-1}\right]\right] \geq \mathbb{E}\left[\mathbb{E}\left[\|a_t\|^2_{V_t^{-1}}|\mathcal{F}_{t-1}\right]|\mathbf{E}\right]P_0(\mathbf{E}) \geq \mathbb{E}\left[\|\bar{a}_t\|^2_{V_t^{-1}}\right]\bar{p}P_0(\mathbf{E}).
$$

Therefore, we have

$$
\mathbb{E}\left[\sum_{t=1}^T \|\bar{a}_t\|^2_{V_t^{-1}}\right] \leq \frac{1}{\bar{p}_0(1-2e^{-\eta})}\mathbb{E}\left[\sum_{t=1}^T \|a_t\|^2_{V_t^{-1}}\right] \leq \frac{2d\log\left(1+\lambda^{-1}T\right)}{\bar{p}_0(1-2e^{-\eta})}, \qquad (41)
$$

where $\bar{p}_0 = \mathbb{C}\exp\left(-2\alpha\Delta(d,\eta) - 8e^{-\eta}\right)\left\{\frac{C\sqrt{\alpha(d+\eta)}}{1+C^2\alpha(d+\eta)}e^{-C^2\alpha(d+\eta)/2}\right\} - e^{-\eta} - C(\alpha, \lambda, \theta_0)2e^{-\frac{\alpha}{4}t_0\epsilon_{t_0}^2}$ for a sufficiently small $t_0 (\geq t_1)$ and the second inequality uses *elliptical potential lemma*. We can choose such t_0 because $\bar{p}$ increases as t increases (since $t\epsilon_t^2$ increases as t increases).

Now, combining equations 32, 33, 34, 35, 36, and 37, we have

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{R}(T)] \leq& \left(\sum_{t=1}^T t\epsilon_t^2\right)^{1/2}\left(\frac{2d\log\left(1+\lambda^{-1}T\right)}{\bar{p}_0(1-2e^{-\eta})}\right)^{1/2} \\
&+ C_g\sqrt{2d\log\left(1+\lambda^{-1}T\right)}\left(\frac{(D+2)C(\alpha,\lambda,\theta_0)}{(1-\alpha)}\sum_{t=1}^T t\epsilon_t^2\right)^{1/2} \\
&+ C_g\left(\frac{2d\log\left(1+\lambda^{-1}T\right)}{\bar{p}_0(1-2e^{-\eta})}\right)^{1/2}\left(\frac{(D+2)C(\alpha,\lambda,\theta_0)}{(1-\alpha)}\sum_{t=1}^T t\epsilon_t^2\right)^{1/2}. \qquad (42)
\end{aligned}
$$

■

Appendix B. Verifying assumptions for the Exponential family

We first provide the expression for the α-Rényi divergence for exponential family distributions. Observe that for any $\alpha \in (0,1)$,

$$D_\alpha^t(\theta,\vartheta) = \frac{1}{1-\alpha}\left[\alpha A(a_t^\top\theta) + (1-\alpha)A(a_t^\top\vartheta) - A(\alpha a_t^\top\theta + (1-\alpha)a_t^\top\vartheta)\right]. \tag{43}$$

Next, we show that the exponential family satisfies Assumption 4.2.

Lemma 11 *Fix $\alpha \in (0,1)$. For distribution lying in exponential family distribution satisfying conditions in Definition 6, we have,*

$$|a_t^\top\theta - a_t^\top\vartheta| \le \sqrt{\frac{1}{m}}\sqrt{\frac{2}{\alpha}D_\alpha^t(\theta,\vartheta)}. \tag{44}$$

Proof The proof of the lemma is a direct consequence of the condition (iii) in Definition 6. ■

Our next result shows that the exponential family satisfies Assumption 4.4.

Lemma 12 *Under Assumption 4.1, the exponential family of distribution as defined in Definition 6, satisfies Assumption 4.4 for $\epsilon_t^2 = \frac{4dC_g\log(t)}{D\alpha C_\phi t}$, where C_g is the Lipschitz constant of the link function and $C_\phi = \min_{i\in[d],x\in B(\theta_0^i,1)}\phi_i(x)$, $B(\theta_0^i,u)\subset\mathbb{R}$ is a ball of radius u centered at $\theta_0^i\in\Theta$ and $\phi_i(\cdot)$ is the density of Π_t^i.*

Proof [Proof of Lemma 12] Observe that

$$\begin{aligned}
D_2^{(t)}(\theta,\theta_0) &= \log\int p_\theta^{(t)}(r^{(t)}|a^{(t)})^2 p_0^{(t)}(r^{(t)}|a^{(t)})^{-1}d\mu^{(t)} \\
&= \sum_{s=1}^t \log\int p_\theta(r_s|\mathcal{F}_{t-1},a_s)^2 p_0(r_s|\mathcal{F}_{s-1},a_s)^{-1}d\mu \\
&= \sum_{s=1}^t \left[A(a_s^\top(2\theta-\theta_0)) - 2A(a_s^\top\theta) + A(a_s^\top\theta_0)\right] \\
&\le C_g\sum_{s=1}^T \left|(a_s^\top(\theta-\theta_0)\right|^2 \\
&\le tC_g\|(\theta-\theta_0)\|^2.
\end{aligned} \tag{45}$$

where the third equality uses the definition of 2-Rényi divergence for an exponential family of distributions, the first inequality follows due to condition (ii) in Definition 6 (since $|A''(x)| \le C_g$,

the result follows by second-order mean value theorem), the last inequality follows from Cauchy-Schwartz inequality and the Assumption 4.1 that $\sum_{t=1}^{T} \|a_t\| \leq T$. Now using equation 45 observe that

$$\begin{aligned}\Pi\left(D_2^{(t)}(\theta,\theta_0) \leq \frac{D\alpha}{4} t\epsilon_t^2\right) &\geq \Pi\left(tC_g\|\theta-\theta_0)\|^2 \leq \frac{D\alpha}{4} t\epsilon_t^2\right) \\ &= \Pi\left((\theta-\theta_0)^\top(\theta-\theta_0) \leq \frac{D\alpha}{4C_g}\epsilon_t^2\right) \\ &\geq \prod_{i=1}^{d} \Pi^i\left((\theta^i-\theta_0^i)^2 \leq \frac{D\alpha}{4dC_g}\epsilon_t^2\right). \end{aligned} \tag{46}$$

Fix $\epsilon_t^2 = \frac{\log(t)}{C^2 t C_\phi}$ for $C^2 = \frac{D\alpha}{4dC_g}$ and observe that $C\epsilon_t = \sqrt{\frac{\log(t)}{tC_\phi}} \leq 1$ for all $t \geq 1$. Now it follows that $\min_{i\in[d], x\in B(\theta_0^i, C\epsilon_T)} \phi(x) \geq \min_{i\in[d], x\in B(\theta_0^i,1)} \phi_i(x) = C_\phi$ for all $t \geq t_0$ and any $i \in [d]$, where $B(\theta_0^i, u) \subset \mathbb{R}$ is a ball of radius u centered at $\theta_0^i \in \Theta$ and $\phi_i(\cdot)$ is the density of Π_t^i. Note that $C_\phi > 0$ since the prior density is strictly positive everywhere. Consequently, for all $t \geq t_0$, it follows from our choice of ϵ_t that

$$\Pi^i\left(|\theta^i-\theta_0^i|^2 \leq \frac{D\alpha}{4dC_g}\epsilon_t^2\right) \geq 2 \min_{x\in B(\theta_0^i, C\epsilon_t)} \phi(x) C\epsilon_t \geq C_\phi \sqrt{\frac{4\log(t)}{tC_\phi^2}} \geq \sqrt{\frac{\log(t)}{t}}. \tag{47}$$

Now the assertion of the lemma follows using the fact that $\sqrt{\frac{\log n}{n}} \geq \frac{1}{n} = e^{-C^2 C_\phi n\epsilon_n^2} = e^{-\frac{D\alpha C_\phi}{4dC_g} n\epsilon_n^2}$ for any $n \geq 2$. In particular, we have from equation 46, equation 47 and the arguments above that

$$\Pi\left(D_2^{(t)}(\theta,\theta_0) \leq \frac{D\alpha}{4} t\epsilon_t^2\right) \geq \prod_{i=1}^{d} \Pi^i\left((\theta^i-\theta_0^i)^2 \leq \frac{D\alpha}{4dC_g}\epsilon_T^2\right) \geq e^{-\frac{D\alpha C_\phi}{4C_g} t\epsilon_t^2}, \tag{48}$$

and the result follows for any prior for which $C_\phi \leq C_g$ (otherwise, this term will appear in the form of a constant in the main result).

■

To show that the exponential family satisfies Assumption 4.6, the conditions required for the finite BvM to hold, we first write down various expressions used in their definition for exponential family models. First recall the definition of the conditional log-likelihood of generating $X_t|a^{(t)}$, that is $\mathbb{L}(\theta) = \log \prod_{s=1}^{t} [p_\theta(r_s|a_s)]$. Using this definition, the gradient of the log-likelihood can be derived as

$$\nabla\mathbb{L}(\theta) = \sum_{s=1}^{t} \nabla \log p_\theta(r_s|a_s) = \sum_{s=1}^{t} \left(r_s a_s^\top - A'(a_s^\top\theta)a_t^\top\right). \tag{49}$$

Similarly, $\nabla^2\mathbb{L}(\theta) = -\sum_{s=1}^{t} A''(a_t^\top\theta)a_t a_t^\top$. Recall that $\mathbb{E}_a[] = \mathbb{E}[|a^{(t)}]$. Now observe that

$$\mathbb{E}_a[\nabla\mathbb{L}(\theta)] = \sum_{s=1}^{t} \mathbb{E}_a[\nabla\mathbb{L}(\theta)] = \sum_{s=1}^{t} \left(A'(a_s^\top\theta_0)a_s^\top - A'(a_s^\top\theta)a_s^\top\right).$$

and $\mathbb{D}_0^2$ and $\mathbb{D}_0^2(\theta) = -\nabla^2\mathbb{E}_a[\mathbb{L}(\theta)] = \sum_{s=1}^t A''(a_s^\top\theta)a_s a_s^\top$. The stochastic part of the conditional log-likelihood is denoted as $\zeta(\theta) := \mathbb{L}(\theta) - \mathbb{E}_a[\mathbb{L}(\theta)]$ and $\nabla\zeta(\theta) = \nabla\mathbb{L}(\theta) - \mathbb{E}_a[\nabla\mathbb{L}(\theta)] = \sum_{s=1}^t \left(r_s a_s^\top - A'(a_s^\top\theta)a_s^\top\right) - \mathbb{E}_a[\sum_{s=1}^t \left(r_s a_s^\top - A'(a_s^\top\theta)a_s^\top\right)] = \sum_{s=1}^t \left(r_s a_s^\top - \mathbb{E}_a[r_s a_s^\top]\right) = \sum_{s=1}^t \left(r_s a_s^\top - A'(a_s^\top\theta_0)a_s^\top\right)$ and therefore $\mathbb{E}_a[\nabla\zeta(\theta_0)] = 0$. Note that $\nabla\zeta(\theta)$ is independent of θ, therefore $\nabla^2\zeta(\theta) = 0$. First, we present a technical lemma that is satisfied by exponential families.

Lemma 13 *There exists some constant $v_0 > 0$ and $\mathtt{g}_1 > 0$, for every s a constant $\sigma_s^2 = A''(a_s^\top\theta_0)$ (given $\mathcal{F}_{s-1}$ and a_s) such that $\mathbb{E}[(r_s - A'(a_s^\top\theta_0))^2/\sigma_i^2|\mathcal{F}_{s-1}, a_s] \leq 1$ and*

$$\log\mathbb{E}[\exp(\lambda(r_s - A'(a_s^\top\theta_0))/\sigma_s)|\mathcal{F}_{s-1}, a_s] \leq v_0^2\lambda^2/2, \quad |\lambda| < \mathtt{g}_1. \tag{50}$$

Proof The proof is a direct consequence of Lemma 2.14 Spokoiny (2012)(in their supplement). ■

Next, we have the main result that verifies all the conditions in Assumption 4.6.

Lemma 14 *The exponential family of distribution (Definition 6) satisfies Assumption 4.6. The condition (ED_0) is satisfied with $\mathtt{g} := \mathtt{g}_1\mathbb{T}^{1/2}$, where $\mathbb{T}^{-1/2} := \max_{s\in[t]}\sup_{\gamma\in\mathbb{R}^d}\frac{A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|}$ and condition (ED_1) can be satisfied for any $\omega > 0$ and $\mathtt{g} > 0$. Condition $\mathcal{L}_0$ follows for $\delta(\mathtt{r}) = L\mathbb{T}_2^{-1/2}\mathtt{r}$, where $\mathbb{T}^{-1/2} := \max_{s\in[t]}\sup_{\gamma\in\mathbb{R}^d}\frac{A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|}$.*

Proof We derive all the conditions in seriatim. The proofs are adapted from Panov and Spokoiny (2015), and Spokoiny (2012).

1. Proof for (ED_0): Observe using the definition of $\nabla\zeta(\theta_0)$ that

$$\begin{aligned}\mathbb{E}_a\exp\left\{\mathtt{m}\frac{\langle\nabla\zeta(\theta_0),\gamma\rangle}{\|\mathbb{D}_0\gamma\|}\right\} &= \mathbb{E}_a\exp\left\{\mathtt{m}\frac{\sum_{s=1}^t\left(r_s a_s^\top\gamma - A'(a_s^\top\theta_0)a_s^\top\gamma\right)}{\|\mathbb{D}_0\gamma\|}\right\}\\ &= \mathbb{E}_a\prod_{s=1}^t\exp\left\{\mathtt{m}\frac{a_s^\top\gamma\left(r_s - A'(a_s^\top\theta_0)\right)}{\|\mathbb{D}_0\gamma\|}\right\}\\ = \mathbb{E}_a\left[\prod_{s=1}^{t-1}\exp\left\{\mathtt{m}\frac{a_s^\top\gamma\left(r_s - A'(a_s^\top\theta_0)\right)}{\|\mathbb{D}_0\gamma\|}\right\}\mathbb{E}_a\left[\exp\left\{\mathtt{m}\frac{a_s^\top\gamma\left(r_s - A'(a_s^\top\theta_0)\right)}{\|\mathbb{D}_0\gamma\|}\right\}\right]\right].\end{aligned} \tag{51}$$

Define $\mathbb{T}^{-1/2} := \max_{s\in[t]}\sup_{\gamma\in\mathbb{R}^d}\frac{A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|}$. By this definition, $\mathbb{T}^{-1/2} \geq \frac{A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|}$ and therefore $\mathtt{m}\frac{A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|} \leq \mathtt{m}\mathbb{T}^{-1/2} \leq \mathtt{g}_1$, for $\mathtt{g} := \mathtt{g}_1\mathbb{T}^{1/2}$. Therefore, using Lemma 13 it follows that

$$\mathbb{E}_a\left[\exp\left\{\mathtt{m}\frac{a_s^\top\gamma\left(r_s - A'(a_s^\top\theta_0)\right)}{\|\mathbb{D}_0\gamma\|}\right\}\right] \leq \exp\left(v_0^2\mathtt{m}^2\frac{A''(a_s^\top\theta_0)^2|a_s^\top\gamma|^2}{2\|\mathbb{D}_0\gamma\|^2}\right). \tag{52}$$

Now substituting equation 52 into equation 51, we have for $|\mathtt{m}| \leq \mathtt{g} := \mathtt{g}_1 \mathbb{T}^{1/2}$,

$$\begin{aligned}
&\mathbb{E}_a \exp\left\{\mathtt{m}\frac{\langle \nabla\zeta(\theta_0), \gamma\rangle}{\|\mathbb{D}_0\gamma\|}\right\} \\
&\quad\leq \mathbb{E}_a\left[\prod_{s=1}^{t-1}\exp\left\{\mathtt{m}\frac{a_s^\top\gamma\left(r_s - A'(a_s^\top\theta_0)\right)}{\|\mathbb{D}_0\gamma\|}\right\}\exp\left(v_0^2\mathtt{m}^2\frac{A''(a_t^\top\theta_0)^2|a_t^\top\gamma|^2}{2\|\mathbb{D}_0\gamma\|^2}\right)\right] \\
&\quad\leq \exp\left(v_0^2\mathtt{m}^2\frac{\sum_{s=1}^{t}A''(a_s^\top\theta_0)^2|a_s^\top\gamma|^2}{2\|\mathbb{D}_0\gamma\|^2}\right) = \exp\left(v_0^2\mathtt{m}^2/2\right). \qquad (53)
\end{aligned}$$

The result follows for any γ, so it must follow for the supremum over $\gamma \in \mathbb{R}^d$.

2. Proof for (ED_1): This assumption follows immediately for any $\omega > 0$ because $\nabla^2\zeta(\theta) = 0$ by definition.

3. Proof for $(\mathcal{L}_0)$: For $I_d = \mathbb{D}_0^{-1}\mathbb{D}_0^2\mathbb{D}_0^{-1}$, we have by the definition of operator norm

$$\begin{aligned}
\|\mathbb{D}_0^{-1}(\mathbb{D}_0^2(\theta) - \mathbb{D}_0^2)\mathbb{D}_0^{-1}\| &= \sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1}|\gamma^\top\mathbb{D}_0^{-1}(\mathbb{D}_0^2(\theta) - \mathbb{D}_0^2)\mathbb{D}_0^{-1}\gamma| \\
&= \sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1}\left|\sum_{s=1}^{t}(A''(a_s^\top\theta) - A''(a_s^\top\theta_0))\gamma^\top\mathbb{D}_0^{-1}a_sa_s^\top\mathbb{D}_0^{-1}\gamma\right| \\
&\leq \sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1}\sum_{s=1}^{t}\left|A''(a_s^\top\theta) - A''(a_s^\top\theta_0))\right|\gamma^\top\mathbb{D}_0^{-1}a_sa_s^\top\mathbb{D}_0^{-1}\gamma \\
&\leq \sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1}\sum_{s=1}^{t}L\left|a_s^\top(\theta-\theta_0)\right|\gamma^\top\mathbb{D}_0^{-1}a_sa_s^\top\mathbb{D}_0^{-1}\gamma \qquad (54)
\end{aligned}$$

Define $\mathbb{T}_2^{-1/2} := \max_s \sup_{\gamma\in\mathbb{R}^d}\frac{|a_s^\top\gamma|}{A''(a_s^\top\theta_0)\|\mathbb{D}_0\gamma\|} \geq \frac{|a_s^\top(\theta-\theta_0)|}{A''(a_s^\top\theta_0)\|\mathbb{D}_0(\theta-\theta_0)\|}$ and observe that

$$\begin{aligned}
\|\mathbb{D}_0^{-1}(\mathbb{D}_0^2(\theta) - \mathbb{D}_0^2)\mathbb{D}_0^{-1}\| &\leq \sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1}\sum_{s=1}^{t}L\left|a_s^\top(\theta-\theta_0)\right|\gamma^\top\mathbb{D}_0^{-1}a_sa_s^\top\mathbb{D}_0^{-1}\gamma \\
&\leq L\mathbb{T}_2^{-1/2}\|\mathbb{D}_0(\theta-\theta_0)\|\sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1}\gamma^\top\mathbb{D}_0^{-1}\sum_{s=1}^{t}A''(a_s^\top\theta_0)a_sa_s^\top\mathbb{D}_0^{-1}\gamma \\
&= L\mathbb{T}_2^{-1/2}\mathtt{r}. \qquad (55)
\end{aligned}$$

Consequently, $\delta(\mathtt{r}) = L\mathbb{T}_2^{-1/2}\mathtt{r}$.

4. Proof for $(\mathcal{L}\mathtt{r})$: For any $\mathtt{r} > 0$ there exists a value $\mathtt{b}(\mathtt{r}) > 0$, such that $\mathtt{r}\mathtt{b}(\mathtt{r}) \to \infty$ as $\mathtt{r} \to \infty$ and

$$-\mathbb{E}_a\mathbb{L}(\theta,\theta_0) \geq \mathtt{r}^2 b(\mathtt{r}) \quad \text{for all } \theta \text{ with } \mathtt{r} = \|\mathbb{D}_0(\theta-\theta_0)\|. \qquad (56)$$

Note that

$$\begin{aligned}\mathbb{E}_a[\mathbb{L}(\theta,\theta_0)] = \mathbb{E}_a[\mathbb{L}(\theta)-\mathbb{L}(\theta_0)] &= \sum_{s=1}^{t}\mathbb{E}_a\left[\log\left(\frac{p_\theta(r_s|a_s)}{p_{\theta_0}(r_s|\mathcal{F}_{s-1},a_s)}\right)\right]\\ &= \sum_{s=1}^{t}\mathbb{E}_a\left[r_s a_s^\top\theta - A(a_s^\top\theta) - r_s a_s^\top\theta_0 + A(a_s^\top\theta_0)\right]\\ &= \sum_{s=1}^{t}\left[A(a_s^\top\theta_0) - A(a_s^\top\theta) + A'(a_s^\top\theta_0)a_s^\top(\theta-\theta_0)\right]\\ &= -\sum_{s=1}^{t}A''(a_s^\top\theta^*)(\theta-\theta_0)^\top a_s a_s^\top(\theta-\theta_0)\\ &= -\|\mathbb{D}_0(\theta^*)(\theta-\theta_0)\|^2, \end{aligned} \tag{57}$$

where the penultimate equality uses second-'order Taylor's theorem for a θ^* that lies between θ and θ_0. Now observe that $\frac{-\mathbb{E}_a[\mathbb{L}(\theta,\theta_0)]}{\mathtt{r}^2} = \frac{\|\mathbb{D}_0(\theta^*)(\theta-\theta_0)\|^2}{\|\mathbb{D}_0(\theta-\theta_0)\|^2}$. Note that, by definition θ^* is a mapping from $\mathtt{r}$ and so is $\frac{\|\mathbb{D}_0(\theta^*)(\theta-\theta_0)\|^2}{\|\mathbb{D}_0(\theta-\theta_0)\|^2}$. Therefore, we can always find a mapping $\mathtt{b}(\mathtt{r})$ such that $\mathtt{r}\mathtt{b}(\mathtt{r}) \to \infty$ as $\mathtt{r}\to\infty$ and $\frac{-\mathbb{E}_a[\mathbb{L}(\theta,\theta_0)]}{\mathtt{r}^2}$.

■

Now recall the definition of $\Delta(\mathtt{r}_0,\eta)$ from (Panov and Spokoiny, 2015, Theorem 9), that is $\Delta(\mathtt{r}_0,\eta) = \{\delta(\mathtt{r}_0) + 6v_0 z_{\mathbb{H}}(\eta)\omega\}\,\mathtt{r}_0^2$, where $z_{\mathbb{H}}(\eta) := 2\sqrt{d} + \sqrt{2\eta} + \mathtt{g}^{-1}(\mathtt{g}^{-2}\eta+1)4d$. Note that, we denote $\Delta(\mathtt{r}_0,\eta)$ as $\Delta_t(d,\eta)$ to explicitly show the dependence of d, t, and η. Since ω (from $(ED1)$) can be any positive number, we choose $\omega = \frac{\delta(\mathtt{r}_0)}{6v_0 z_{\mathbb{H}}(\eta)}$. Therefore, $\Delta(\mathtt{r}_0,\eta) = 2\delta(\mathtt{r}_0)\mathtt{r}_0^2 \le L\frac{\mathtt{r}_0^3}{\mathbb{T}_2^{1/2}}$. Moreover, $\mathtt{r}_0^2 = C(\eta+d)$ for some known constant C (Spokoiny, 2012, Section 5.2). Note that $\|a_s\| = 1$ because g is monotonic and the optimizer of $a^\top\theta$ on $a\in\mathbb{R}^d : \|a\|\le 1$ is nothing but $\theta/\|\theta\|$, which lies on a d-dimensional sphere ($\mathcal{S}^{d-1}$). Observe that for some $\gamma^*\in\mathbb{R}^d\backslash\{0\}$, $\mathbb{T}^{-1/2} := \max_{s\in[t]}\sup_{\gamma\in\mathbb{R}^d}\frac{A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|} = \max_{s\in[t]}\frac{A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma^*|}{\|\mathbb{D}_0\gamma^*\|} \le \frac{1}{\sqrt{t}}\frac{\max_{s\in[t]}A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma^*|}{\min_{s\in[t]}A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma^*|} \le \frac{1}{\sqrt{t}}\sqrt{\frac{C_g}{m}}\frac{\max_{s\in[t]}|a_s^\top\gamma^*|}{\min_{s\in[t]:|a_s^\top\gamma^*|\neq 0}|a_s^\top\gamma^*|} \le \frac{1}{\sqrt{t}}\sqrt{\frac{C_g}{m}}\frac{\|\gamma^*\|}{\min_{s\in[t]:|a_s^\top\gamma^*|\neq 0}|a_s^\top\gamma^*|}$. Since, $\gamma^*\neq 0$ and $a_s\in\mathcal{S}^{d-1}$, we can compute a t-independent lower bound on $\min_{s\in[t]:|a_s^\top\gamma^*|\neq 0}|a_s^\top\gamma^*| \ge \min_{a\in\mathcal{S}^{d-1}:|a^\top\gamma^*|\neq 0}|a_s^\top\gamma^*|$. Therefore, $\mathbb{T}^{-1/2} = O(t^{-1/2})$. By definition of $\mathbb{D}_0^2$, observe that $\|\mathbb{D}_0\gamma^*\|^2 = \gamma^\top\mathbb{D}_0^2\gamma$ $= \sum_{s=1}^t A''(a_s^\top\theta_0)|\gamma^\top a_s||a_s^\top\gamma| \ge A''(a_s^\top\theta_0)|a_s^\top\gamma|^2$. This also implies that $T^{-1/2}\le 1$ and thus $\Delta_t(d,\eta)\le L\mathtt{r}_0^3 := \Delta(d,\eta)$.

Appendix C. Verifying assumptions for the sub-Gaussian family

First, we show a general result for the sub-Gaussian family of distribution that implies Assumption 4.2.

Lemma 15 *Fix $\alpha \in (0,1)$. For any two sub-Gaussian measures μ and ν with sub-Gaussian parameter σ_μ and σ_ν respectively, such that ν is absolutely continuous wrt μ, the $\alpha-$Rényi divergence can be bounded below by the absolute difference between the respective means. In particular, for any random variable X having measure μ and ν, we have*

$$|\mathbb{E}_\nu[X] - \mathbb{E}_\mu[X]| \leq \sqrt{(\sigma_\mu^2\alpha + \sigma_\nu^2(1-\alpha))}\sqrt{\frac{2}{\alpha}D_\alpha(\nu\|\mu)}. \tag{58}$$

Proof [Proof of Lemma 15] For any $\alpha \in (0,1)$, recall the definition of α-Rényi divergence $D_\alpha(\nu\|\mu) = \log \int g^\alpha d\mu = \frac{1}{\alpha-1}\log \int \left(\frac{d\nu}{d\mu}\right)^\alpha d\mu$, where $g \equiv \frac{d\nu}{d\mu}$. Now observe that

$$\begin{aligned}
(\alpha-1)D_\alpha(\nu\|\mu) &= \log \mathbb{E}_\nu[g^{\alpha-1}e^{-(\alpha-1)X}e^{(\alpha-1)X}] = \log \int (ge^{-X})^{\alpha-1}e^{(\alpha-1)X}d\nu \\
&\leq \log \left(\int e^{(\alpha-1)X}d\nu\right)^\alpha \left(\int (ge^{-X})^{-1}e^{(\alpha-1)X}d\nu\right)^{1-\alpha} \\
&= \log \left(\mathbb{E}_\nu[e^{(\alpha-1)X}]^\alpha \mathbb{E}_\nu[g^{-1}e^{\alpha X}]^{1-\alpha}\right) \\
&= \log \left(\mathbb{E}_\nu[e^{(\alpha-1)X}]^\alpha \mathbb{E}_\mu[e^{\alpha X}]^{1-\alpha}\right) \\
&= \alpha \log \mathbb{E}_\nu[e^{(\alpha-1)X}] + (1-\alpha)\log \mathbb{E}_\mu[e^{\alpha X}],
\end{aligned} \tag{59}$$

where the first inequality follows from the Hölder's inequality wrt measure $e^{(\alpha-1)X}d\nu$.

Now fix $X = sX$ for any $s \in \mathbb{R}$ (without loss of generality). Since, $\alpha \in (0,1)$, it follows from the inequality above that

$$\begin{aligned}
D_\alpha(\nu\|\mu) &\geq \left[\frac{\alpha}{\alpha-1}\log \mathbb{E}_\nu[e^{(\alpha-1)sX}] - \log \mathbb{E}_\mu[e^{\alpha sX}]\right] \\
&\geq \left[\frac{\alpha}{\alpha-1}[s(\alpha-1)\mathbb{E}_\nu[X] + \sigma_\nu^2 s^2(\alpha-1)^2/2] - [s\alpha\mathbb{E}_\mu[X] + \sigma_\mu^2 s^2\alpha^2/2]\right] \\
&= \left[s\alpha[\mathbb{E}_\nu[X] - \mathbb{E}_\mu[X]] - \frac{s^2\alpha}{2}[\sigma_\nu^2(1-\alpha) + \sigma_\mu^2\alpha]\right],
\end{aligned} \tag{60}$$

where the second inequality uses the fact that μ and ν are sub-Gaussian measures. Recall if $X \sim \mu$, is a sub-Gaussian random variable, then $\mathbb{E}_\mu[e^{sX}] \leq e^{s\mathbb{E}_\mu[X] + \sigma_\mu^2 s^2/2}$ for all $s \in \mathbb{R}$. Now it follows from above that

$$\begin{aligned}
|\mathbb{E}_\nu[X] - \mathbb{E}_\mu[X]| &\leq \inf_{|s|} \frac{D_\alpha(\nu\|\mu)}{|s|\alpha} + \frac{|s|}{2}(\sigma_\mu^2\alpha + \sigma_\nu^2(1-\alpha)) \\
&= \sqrt{(\sigma_\mu^2\alpha + \sigma_\nu^2(1-\alpha))}\sqrt{\frac{2}{\alpha}D_\alpha(\nu\|\mu)}.
\end{aligned} \tag{61}$$

■

Next, we provide a technical result that bounds 2-Rényi divergence between two random variables that are modeled with additive noise.

Lemma 16 *For a given $a \in \mathbb{A}$ and any $\theta \in \Theta$, let $X = g(a^\top\theta) + \eta$, where $g(\cdot)$ is a known twice differentiable mapping and η is a random variable with density $p(\cdot)$. Then for any $\epsilon > 0$, there exist a $\theta^*(\epsilon)$ in $B(\theta_0, \epsilon)$, a convex ball centred at θ_0 with radius ϵ, such that for any $\theta \in B(\theta_0, \epsilon)$,*

$$D_2(\theta, \theta_0) = (\theta - \theta_0)^\top a^\top \mathcal{I}(\theta^*(\epsilon)) a(\theta - \theta_0), \tag{62}$$

where $\mathcal{I}(\theta) = \mathbb{E}_\theta \left[(\nabla_u \log p(x - g(u)))(\nabla_u \log p(x - g(u)))^\top \Big|_{u=a^\top\theta} \right]$ and $\mathcal{I}(\theta_0) aa^\top$ is the Fisher information about θ_0 that X contains.

Proof Recall from the definition of the $2-$Rényi divergence that, for a given $a \in \mathbb{A}$,

$$D_2(\theta, \theta_0) = \log \int p_\theta(x|a)^2 p_{\theta_0}(x|a)^{-1} dx = \log \mathbb{E}_{\theta_0} \left[\frac{p_\theta(x|a)^2}{p_{\theta_0}(x|a)^2} \right] = \log \mathbb{E}_{\theta_0} \left[\frac{p(x - g(a^\top\theta))^2}{p(x - g(a^\top\theta_0))^2} \right]. \tag{63}$$

For brevity, we denote $\dot{f}$ and $\ddot{f}$ as the first and second derivative of f for $f = \{g, p\}$. It is straightforward to compute the gradient of D_2 with respect to θ as

$$\nabla_\theta D_2 = -\left(\mathbb{E}_{\theta_0} \left[\frac{p(x - g(a^\top\theta))^2}{p(x - g(a^\top\theta_0))^2} \right] \right)^{-1} \mathbb{E}_{\theta_0} \left[\frac{2p(x - g(a^\top\theta)) \dot{p}(x - g(a^\top\theta)) \dot{g}(a^\top\theta) a^\top}{p(x - g(a^\top\theta_0))^2} \right]. \tag{64}$$

Similarly, the Hessian of D_2 is computed as,

$$\begin{aligned} \nabla^2_\theta D_2 = &-\left(\mathbb{E}_{\theta_0} \left[\frac{p(x - g(a^\top\theta))^2}{p(x - g(a^\top\theta_0))^2} \right] \right)^{-2} \left(\mathbb{E}_{\theta_0} \left[\frac{2p(x - g(a^\top\theta)) \dot{p}(x - g(a^\top\theta)) \dot{g}(a^\top\theta)}{p(x - g(a^\top\theta_0))^2} \right] \right)^2 aa^\top \\ &+ \left(\mathbb{E}_{\theta_0} \left[\frac{p(x - g(a^\top\theta))^2}{p(x - g(a^\top\theta_0))^2} \right] \right)^{-1} \\ &\mathbb{E}_{\theta_0} \left[\frac{\begin{array}{c} 2p(x - g(a^\top\theta)) \ddot{p}(x - g(a^\top\theta)) \dot{g}(a^\top\theta)^2 + 2\left(\dot{p}(x - g(a^\top\theta)) \dot{g}(a^\top\theta) \right)^2 \\ -2p(x - g(a^\top\theta)) \dot{p}(x - g(a^\top\theta)) \ddot{g}(a^\top\theta) \end{array}}{p(x - g(a^\top\theta_0))^2} \right] aa^\top. \end{aligned} \tag{65}$$

When we evaluate the above Hessian at $\theta = \theta_0$ using the fact that $\nabla_\theta \mathbb{E}_{\theta_0}[\log p(x - g(a^\top\theta))]\Big|_{\theta=\theta_0} = 0$ and $\mathbb{E}_{\theta_0} \left[\frac{2\nabla^2_\theta p(x - g(a^\top\theta))}{p(x - g(a^\top\theta_0))} \right] = \nabla^2_\theta \int p_\theta(x|a) dx = 0$, then

$$\nabla^2_\theta D_2 \Big|_{\theta=\theta_0} = \mathbb{E}_{\theta_0} \left[2(\nabla_u \log p(x - g(u)))(\nabla_u \log p(x - g(u)))^\top \Big|_{u=a^\top\theta_0} \right] aa^\top = 2\mathcal{I}(\theta_0) aa^\top. \tag{66}$$

Now using second-order Taylor's theorem, it follows that for any $\epsilon > 0$, there exists a $\theta^*(\epsilon)$ in $B(\theta_0, \epsilon)$ such that for any $\theta \in B(\theta_0, \epsilon)$,

$$D_2(\theta, \theta_0) = 2(\theta - \theta_0)^\top a^\top \mathcal{I}(\theta^*(\epsilon)) a(\theta - \theta_0). \tag{67}$$

■

Our next result shows that the sub-Gaussian family satisfies Assumption 4.4.

Lemma 17 *Under Assumption 4.1, the sub-Gaussian family of distributions as defined in Definition 8, satisfies Assumption 4.4 for* $\epsilon_t^2 = \frac{4d\tilde{C}\log(t)}{D\alpha C_\phi t}$ *and sufficiently large* t *(*$t \geq \log(t)/C_\phi$*), where* $\tilde{C} \geq \mathcal{I}(\theta)$ *for any* $\theta \in B(\theta_0, 1)$ *and* $C_\phi = \min_{i\in[d], x\in B(\theta_0^i,1)} \phi_i(x)$*,* $B(\theta_0^i, u) \subset \mathbb{R}$ *is a ball of radius* u *centered at* $\theta_0^i \in \Theta$ *and* $\phi_i(\cdot)$ *is the density of* Π_t^i*.*

Proof [Proof of Lemma 17] Using Lemma 16, there exist a $\theta^*(1) \in B(\theta_0, 1)$ (defined later), a convex ball centred at θ_0 with radius 1, such that for any $\theta \in B(\theta_0, 1)$,

$$\begin{aligned} D_2^{(t)}(\theta, \theta_0) &= \log \int p_\theta^{(t)}(r^{(t)}|a^{(t)})^2 p_0^{(t)}(r^{(t)}|a^{(t)})^{-1} d\mu^{(t)} \\ &= \sum_{s=1}^t \log \int p_\theta(r_s|\mathcal{F}_{t-1}, a_s)^2 p_0(r_s|\mathcal{F}_{s-1}, a_s)^{-1} d\mu \\ &\leq \tilde{C}(\theta - \theta_0)^\top \left(\sum_{s=1}^t a_s a_s^\top\right)(\theta - \theta_0) \\ &\leq t\tilde{C}\|\theta - \theta_0\|^2. \end{aligned} \tag{68}$$

where in the first inequality $\tilde{C}$ is the bound on $\mathcal{I}(\theta)$ for any $\theta \in B(\theta_0, 1)$ and the last inequality follows due to CS and Assumption 4.1. Now observe for $B(\theta_0, 1) = \{\theta \in \Theta : \|\theta - \theta_0)\| \leq 1\}$ that

$$\begin{aligned} \Pi\left(D_2^{(t)}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq \frac{D\alpha}{4} t\epsilon_t^2\right) &\geq \Pi\left(B(\theta_0, 1), D_2^{(t)}(\theta, \theta_0) \leq \frac{D\alpha}{4} t\epsilon_t^2\right) \\ &\geq \Pi\left(B(\theta_0, 1), \|\theta - \theta_0\|^2 \leq \frac{D\alpha}{4\tilde{C}}\epsilon_t^2\right) \\ &= \Pi\left((\theta - \theta_0)^\top(\theta - \theta_0) \leq \frac{D\alpha}{4\tilde{C}}\epsilon_t^2\right) \\ &\geq \prod_{i=1}^d \Pi^i\left((\theta^i - \theta_0^i)^2 \leq \frac{D\alpha}{4d\tilde{C}}\epsilon_T^2\right), \end{aligned} \tag{69}$$

where the second inequality follows from equation 68 and the first equality is due to the assumption that $\frac{D\alpha}{4\tilde{C}d}\epsilon_t^2 < 1$ (for sufficiently large t). Now the result follows for $\epsilon_t^2 = \frac{4d\tilde{C}\log(t)}{D\alpha t C_\phi}$.

■

To show that the sub-Gaussian family satisfies Assumption 4.6, we first write down various expressions used in their definition for this family. The conditional log-likelihood of generating $X_t|a^{(t)}$ as $\mathbb{L}(\theta) = \log\prod_{s=1}^t \left[p_\eta(r_s - a_s^\top\theta)\right]$, where p_η is the density of the sub-Gaussian error with sub-Gaussian parameter 1. Also, denote $\log p_\eta := h_\eta$. The stochastic part of the conditional log-likelihood is denoted as $\zeta(\theta) := \mathbb{L}(\theta) - \mathbb{E}[\mathbb{L}(\theta)|a^{(t)}]$ and $\nabla\zeta(\theta) = \nabla\mathbb{L}(\theta) - \mathbb{E}[\nabla\mathbb{L}(\theta)|a^{(t)}] = \sum_{s=1}^t -h_\eta'(r_s - a_s^\top\theta)a_s^\top + \mathbb{E}[h_\eta'(r_s - a_s^\top\theta)a_s^\top|a^{(t)}]$ and therefore $\mathbb{E}[\nabla\zeta(\theta_0)|a^{(t)}] = -\sum_{s=1}^t h_\eta'(r_s - a_s^\top\theta)a_s^\top$ (since $\mathbb{E}[\nabla\mathbb{L}(\theta_0)|a^{(t)}] = 0$). We assume that h_η is twice continuously differentiable and let $\mathtt{h}^2 := -\int h_\eta''(z)p_\eta(z)dz < \infty$. Also,

$$\mathbb{D}_0^2 = \mathbb{D}_0^2(\theta_0) = -\nabla^2\mathbb{E}[\mathbb{L}(\theta_0)|a^{(t)}] = -\sum_{s=1}^t \mathbb{E}[h_\eta''(r_s - a_s^\top\theta_0)|a^{(t)}]a_s a_s^\top = \mathtt{h}^2\sum_{s=1}^t a_s a_s^\top. \tag{70}$$

Similarly, $\mathbb{D}_0^2(\theta) = -\sum_{s=1}^t \mathbb{E}[h''_\eta(r_s - a_s^\top\theta)|a^{(t)}]a_s a_s^\top = -\sum_{s=1}^t \mathbb{E}[h''_\eta(\eta + a_s^\top(\theta_0 - \theta))|a^{(t)}]a_s a_s^\top$. Henceforth, use $\mathbb{E}_a$ in place of the conditional expectation $\mathbb{E}[\cdot|a^{(t)}]$.

Next, we assume two other conditions on the error density p_η.

Assumption C.1

1. *There exists some constant v_0 and $\mathtt{g}_1 > 0$, such that a random variable $\eta \sim p_\eta$, it holds that*

$$\log \mathbb{E}\exp(\mu h'_\eta(\eta)/\mathtt{h}) \le v_0^2\mu^2/2, \quad |\mu| < \mathtt{g}_1. \tag{71}$$

2. *There exists some constant v_0 and for every $\mathtt{r} > 0$, there exists $\mathtt{g}_1(\mathtt{r}) > 0$, such that for all δ with $|\delta| < \mathbb{T}_2^{-1/2}\mathtt{r}/\vartheta_s$ it holds that*

$$\log \mathbb{E}\left[\exp\left(\frac{\mu}{\vartheta_s^2}\left\{h''_\eta(\eta_s+\delta) - \mathbb{E}[h''_\eta(\eta_s+\delta)]\right\}\right)\right] \le v_0^2\mu^2/2, \quad |\mu| < \mathtt{g}_1(\mathtt{r}), \tag{72}$$

where ϑ_s is known value (given $a^{(t)}$) and $\mathbb{T}_2^{-1/2} := \max_s \sup_{\gamma\in\mathbb{R}^d} \frac{\vartheta_s|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|}$.

The above assumption essentially requires that the error distribution has an exponentially decaying tail. Since we assume that the error distribution is sub-Gaussian due to Assumption 8, the above condition is automatically satisfied.

Lemma 18 *The sub-Gaussian family of distributions (Definition 8) satisfies Assumption 4.6. The condition (ED_0) is satisfied with $\mathtt{g} := \mathtt{g}_1\mathbb{T}^{1/2}$, where $\mathbb{T}^{-1/2} := \max_{s\in[t]}\sup_{\gamma\in\mathbb{R}^d}\frac{A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|}$ and condition (ED_1) can be satisfied for any $\omega > 0$ and $\mathtt{g} > 0$. Condition $\mathcal{L}_0$ follows for $\delta(\mathtt{r}) = L\mathbb{T}_2^{-1/2}\mathtt{r}$, where $\mathbb{T}^{-1/2} := \max_{s\in[t]}\sup_{\gamma\in\mathbb{R}^d}\frac{A''(a_s^\top\theta_0)^{1/2}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|}$.*

Proof We derive all the conditions in seriatim. The proofs are adapted from Panov and Spokoiny (2015).

1. Proof for (ED_0):

   Using the definition of $\nabla\zeta(\theta_0)$ and $\mathbb{D}_0$, observe that

$$\mathbb{E}_a\exp\left\{\mathtt{m}\frac{\langle\nabla\zeta(\theta_0),\gamma\rangle}{\|\mathbb{D}_0\gamma\|}\right\} = \mathbb{E}_a\exp\left\{-\mathtt{m}\frac{\sum_{s=1}^t h'_\eta(r_s - a_s^\top\theta)a_s^\top\gamma}{\|\mathbb{D}_0\gamma\|}\right\} \tag{73}$$

$$= \mathbb{E}_a\exp\left\{\frac{-\mathtt{mh}\sum_{s=1}^t a_s^\top\gamma}{\|\mathbb{D}_0\gamma\|}\frac{h'_\eta(\eta)}{\mathtt{h}}.\right\} \tag{74}$$

   Define $\mathbb{T}^{-1/2} := \max_s\sup_{\gamma\in\mathbb{R}^d}\frac{\mathtt{h}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|} \ge \frac{\mathtt{h}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|}$ and $\mathtt{g} = \mathtt{g}_1\mathbb{T}^{1/2}$. Consequently, for $\mu = \frac{-\mathtt{mh}a_s^\top\gamma}{\|\mathbb{D}_0\gamma\|}$, $|\mu| \le |\mathtt{m}|\frac{\mathtt{h}|a_s^\top\gamma|}{\|\mathbb{D}_0\gamma\|} < \mathtt{g}\mathbb{T}^{1/2} = \mathtt{g}_1$. Therefore, using Assumption C.1, the result follows as

$$\mathbb{E}_a\exp\left\{\frac{-\mathtt{mh}\sum_{s=1}^t a_s^\top\gamma}{\|\mathbb{D}_0\gamma\|}\frac{h'_\eta(\eta)}{\mathtt{h}}.\right\} \le \exp\left(\frac{v_0^2\mathtt{m}^2}{2}\frac{\mathtt{h}^2\sum_{s=1}^t|a_s^\top\gamma|^2}{\|\mathbb{D}_0\gamma\|^2}\right) = \exp\left(\frac{v_0^2\mathtt{m}^2}{2}\right). \tag{75}$$

2. Proof for (ED_1):

Using the definition of $\zeta(\theta)$, observe that

$$\begin{aligned}
&\mathbb{E}_a \exp\left\{\frac{\mathtt{m}}{\omega}\frac{\gamma_1^\top \nabla^2\zeta(\theta)\gamma_2}{\|\mathbb{D}_0\gamma_1\|\|\mathbb{D}_0\gamma_2\|}\right\}\\
&\quad= \mathbb{E}_a \exp\left\{\frac{\mathtt{m}}{\omega}\frac{\sum_{s=1}^t \left(h_\eta''(r_s - a_s^\top\theta) - \mathbb{E}_a[h_\eta''(r_s - a_s^\top\theta)]\right)\gamma_1^\top a_s a_s^\top\gamma_2}{\|\mathbb{D}_0\gamma_1\|\|\mathbb{D}_0\gamma_2\|}\right\}\\
&\quad= \mathbb{E}_a \exp\left\{\frac{\mathtt{m}}{\omega}\frac{\sum_{s=1}^t \left(h_\eta''(\eta_s - a_s^\top(\theta-\theta_0)) - \mathbb{E}_a[h_\eta''(\eta_s - a_s^\top(\theta-\theta_0))]\right)\gamma_1^\top a_s a_s^\top\gamma_2}{\|\mathbb{D}_0\gamma_1\|\|\mathbb{D}_0\gamma_2\|}\right\}.
\end{aligned} \tag{76}$$

Using the definition of $\mathbb{T}_2$ for $\theta \in \Theta_0(\mathtt{r})$, observe that $\mathbb{T}_2^{-1/2} \geq \frac{\vartheta_s|a_s^\top(\theta-\theta_0)|}{\|\mathbb{D}_0(\theta-\theta_0)\|}$. Therefore, $|a_s^\top(\theta-\theta_0)| \leq \mathbb{T}_2^{-1/2}\frac{\|\mathbb{D}_0(\theta-\theta_0)\|}{\vartheta_s} \leq \mathbb{T}_2^{-1/2}\frac{\mathtt{r}}{\vartheta_s}$. For $\mu = \frac{\mathtt{m}\vartheta_s^2\gamma_1^\top a_s a_s^\top\gamma_2}{\omega\|\mathbb{D}_0\gamma_1\|\|\mathbb{D}_0\gamma_2\|}$, observe that $|\mu| \leq \frac{|m|}{\omega}\frac{\vartheta_s|a_s^\top\gamma_1|}{\|\mathbb{D}_0\gamma_1\|}\frac{\vartheta_s|a_s^\top\gamma_2|}{\|\mathbb{D}_0\gamma_2\|} \leq \frac{\mathtt{g}(r)}{\omega}\mathbb{T}^{-1} := \mathtt{g}_1(\mathtt{r})$. It follows from Assumption C.1(2), that

$$\mathbb{E}_a \exp\left\{\frac{\mathtt{m}}{\omega}\frac{\gamma_1^\top \nabla^2\zeta(\theta)\gamma_2}{\|\mathbb{D}_0\gamma_1\|\|\mathbb{D}_0\gamma_2\|}\right\} \leq \exp\left(\sum_{s=1}^t v_0^2\frac{\mathtt{m}^2\vartheta_s^4|\gamma_1^\top a_s|^2|a_s^\top\gamma_2|^2}{2\omega^2\|\mathbb{D}_0\gamma_1\|^2\|\mathbb{D}_0\gamma_2\|^2}\right) \leq \exp\left(\frac{v_0^2\mathtt{m}^2}{2\omega^2}\frac{t}{\mathbb{T}_2^2}\right). \tag{77}$$

By fixing $\omega = \frac{\sqrt{t}}{\mathbb{T}_2}$, the result follows.

3. Proof for $(\mathcal{L}_0)$:

For $I_d = \mathbb{D}_0^{-1}\mathbb{D}_0^2\mathbb{D}_0^{-1}$, we have by the definition of the operator norm

$$\begin{aligned}
&\|\mathbb{D}_0^{-1}(\mathbb{D}_0^2(\theta) - \mathbb{D}_0^2)\mathbb{D}_0^{-1}\|\\
&\qquad= \sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1} |\gamma^\top\mathbb{D}_0^{-1}(\mathbb{D}_0^2(\theta) - \mathbb{D}_0^2)\mathbb{D}_0^{-1}\gamma|\\
&\qquad= \sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1} \left|\sum_{s=1}^t(-\mathbb{E}[h_\eta''(\eta + a_s^\top(\theta_0-\theta))|a^{(t)}] + \mathbb{E}[h_\eta''(\eta)|a^{(t)}])\gamma^\top\mathbb{D}_0^{-1}a_s a_s^\top\mathbb{D}_0^{-1}\gamma\right|\\
&\qquad\leq \sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1} \sum_{s=1}^t \mathbb{E}\left[\left|h_\eta''(\eta) - h_\eta''(\eta + a_s^\top(\theta_0-\theta))\right| |a^{(t)}\right]\gamma^\top\mathbb{D}_0^{-1}a_s a_s^\top\mathbb{D}_0^{-1}\gamma\\
&\qquad\leq \sup_{\gamma\in\mathbb{R}^d:\|\gamma\|=1} \sum_{s=1}^t L\left|a_s^\top(\theta-\theta_0)\right|\gamma^\top\mathbb{D}_0^{-1}a_s a_s^\top\mathbb{D}_0^{-1}\gamma.
\end{aligned} \tag{78}$$

Define $\mathbb{T}_1^{-1/2} := \max_s \sup_{\gamma \in \mathbb{R}^d} \frac{\mathtt{h}|a_s^\top \gamma|}{\|\mathbb{D}_0 \gamma\|} \geq \frac{\mathtt{h}|a_s^\top(\theta-\theta_0)|}{\|\mathbb{D}_0(\theta-\theta_0)\|}$ and observe that

$$\begin{aligned}
\|\mathbb{D}_0^{-1}(\mathbb{D}_0^2(\theta) - \mathbb{D}_0^2)\mathbb{D}_0^{-1}\| &\leq \sup_{\gamma \in \mathbb{R}^d : \|\gamma\|=1} \sum_{s=1}^t L \left| a_s^\top(\theta - \theta_0) \right| \gamma^\top \mathbb{D}_0^{-1} a_s a_s^\top \mathbb{D}_0^{-1} \gamma \\
&\leq \frac{L\mathbb{T}_1^{-1/2}}{\mathtt{h}} \|\mathbb{D}_0(\theta - \theta_0)\| \sup_{\gamma \in \mathbb{R}^d : \|\gamma\|=1} \gamma^\top \mathbb{D}_0^{-1} \sum_{s=1}^t \mathtt{h}^2 a_s a_s^\top \mathbb{D}_0^{-1} \gamma \\
&= \frac{L}{\mathbb{T}_1^{1/2}\mathtt{h}} \mathtt{r}. \qquad (79)
\end{aligned}$$

Consequently, $\delta(\mathtt{r}) = L\mathtt{h}^{-1}\mathbb{T}_2^{-1/2}\mathtt{r}$.

4. Proof for $(\mathcal{L}\mathtt{r})$: Note that

$$\begin{aligned}
&\mathbb{E}_a[\mathbb{L}(\theta, \theta_0)] = \mathbb{E}_a[\mathbb{L}(\theta) - \mathbb{L}(\theta_0)] \\
&= \sum_{s=1}^t \mathbb{E}_a \left[\log \left(\frac{p_\theta(r_s|a_s)}{p_{\theta_0}(r_s|\mathcal{F}_{s-1}, a_s)} \right) \right] \\
&= \sum_{s=1}^t \mathbb{E}_a \left[h_\eta(r_s - a_s^\top\theta) - h_\eta(r_s - a_s^\top\theta_0) \right] \\
&= \sum_{s=1}^t \mathbb{E}_a \left[-h'_\eta(r_s - a_s^\top\theta_0) a_s^\top(\theta - \theta_0) + h''_\eta(r_s - a_s^\top\theta^*)(\theta - \theta_0)^\top a_s a_s^\top(\theta - \theta_0) \right] \\
&= \sum_{s=1}^t \mathbb{E}_a[h''_\eta(r_s - a_s^\top\theta^*)](\theta - \theta_0) a_s a_s^\top(\theta - \theta_0) = -\|\mathbb{D}_0(\theta^*)(\theta - \theta_0)\|^2, \qquad (80)
\end{aligned}$$

where the third equality uses second-order Taylor's theorem for a θ^* that lies between θ and θ_0 and the penultimate inequality uses the fact that $\sum_{s=1}^t \mathbb{E}_a \left[-h'_\eta(r_s - a_s^\top\theta_0) a_s^\top \right] = \nabla \mathbb{E}_a[\mathbb{L}(\theta_0)] = 0$.

Now observe that $\frac{-\mathbb{E}_a[\mathbb{L}(\theta,\theta_0)]}{\mathtt{r}^2} = \frac{\|\mathbb{D}_0(\theta^*)(\theta-\theta_0)\|^2}{\|\mathbb{D}_0(\theta-\theta_0)\|^2}$. Note that, by definition θ^* is a mapping from $\mathtt{r}$ and so is $\frac{\|\mathbb{D}_0(\theta^*)(\theta-\theta_0)\|^2}{\|\mathbb{D}_0(\theta-\theta_0)\|^2}$. Therefore, we can always find a mapping $\mathtt{b}(\mathtt{r})$ such that $\mathtt{r}\mathtt{b}(\mathtt{r}) \to \infty$ as $\mathtt{r} \to \infty$ and $\frac{-\mathbb{E}_a[\mathbb{L}(\theta,\theta_0)]}{\mathtt{r}^2} > \mathtt{r}^2\mathtt{b}(\mathtt{r})$.

■

The bound on $\Delta_t(d, \eta)$ can be computed using similar steps as used for exponential family models for $\mathtt{r}_0^2 = C(d + \eta)$ (Panov and Spokoiny, 2015, Theorem 6). However, since, ω cannot be chosen arbitrarily as we did in the exponential case, we will have $\Delta_t(d, \eta) \leq \{L\mathtt{h}^{-1}\mathbb{T}_2^{-1/2}\mathtt{r}_0 + 6v_0(2\sqrt{d} + \sqrt{2\eta} + \mathtt{g}^{-1}(\mathtt{g}^{-2}\eta + 1)4d))(\eta)\omega\}\mathtt{r}_0^2 \leq \{L\mathtt{h}^{-1}\mathtt{r}_0 + 6v_0(2\sqrt{d} + \sqrt{2\eta} + \mathtt{g}_1^{-1}(\mathtt{g}_1^{-2}\eta + 1)4d))(\eta)\}\mathtt{r}_0^2 := \Delta(\eta, d)$, where we used the definition of ω and $\mathtt{g}$ from the last lemma. Consequently, using the definition of $\mathtt{r}_0^2$, obsever that $\Delta(\eta, d) = O(d^2)$.