Identifying near-optimal decisions in linear-in-parameter bandit models with continuous decision sets: Supplementary material

Sanjay P. Bhat[1] **Chaitanya Amballa**[1]

[1]TCS Research, Hyderabad, India

Please note that the numbering of equations, figures and tables in the supplementary material is in continuation with that in the main paper.

A PROOF OF THEOREM 3.2

We begin by recalling properties of the measures $\mathbb{E}^{\mathcal{A},\mu}$ and $\mathbb{E}_t^{\mathcal{A},\mu}$ introduced in Section 3. By Proposition 7.28 in Bertsekas and Shreve [1996], the measures $\mathbb{P}_t^{\mathcal{A},\mu}$, $t \in \mathbb{Z}_+$, and $\mathbb{P}^{\mathcal{A},\mu}$ satisfy the following properties (see also Proposition V.1.1 of Neveu [1965]):

1. For every real-valued function q that is integrable on $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1^{\mathcal{A},\mu})$, we have

$$\int_{\Omega_1} q(h_1)\mathbb{P}_1^{\mathcal{A},\mu}(\mathrm{d}h_1) = \int_{[0,1]^n}\int_{\mathcal{D}}\int_{\mathbb{R}} q(u,s,y) \times Q^{\mu}(\mathrm{d}y|s)\pi_1(\mathrm{d}s|u)\lambda(\mathrm{d}u) \tag{12}$$

2. For every $t > 1$ and every integrable real-valued function q on $(\Omega_t, \mathcal{F}_t, \mathbb{P}_t^{\mathcal{A},\mu})$, we have

$$\int_{\Omega_t} q(h_t)\mathbb{P}_t^{\mathcal{A},\mu}(\mathrm{d}h_t) = \int_{\Omega_{t-1}}\int_{[0,1]^n}\int_{\mathcal{D}}\int_{\mathbb{R}} q(h_{t-1},u,s,y) \times Q^{\mu}(\mathrm{d}y|s)\pi_1(\mathrm{d}s|u)\lambda(\mathrm{d}u)\mathbb{P}_{t-1}^{\mathcal{A},\mu}(\mathrm{d}h_{t-1}). \tag{13}$$

3. For every $t \in \mathbb{Z}_+$ and every Borel subset A of Ω_t, we have $\mathbb{P}_t^{\mathcal{A},\mu}(A) = \mathbb{P}^{\mathcal{A},\mu}(A \times \mathcal{S} \times \mathcal{S} \times \cdots)$.

Next, suppose Assumption 2 holds, and let $\rho_{z,\nu}(\cdot)$ denote the Gaussian density on $\mathbb{R}$ with mean $z \in \mathbb{R}$ and standard deviation $\nu > 0$. Then, for every $\zeta \in \mathbb{R}^f$ and $s \in \mathcal{D}$, the measure $Q^{\zeta}(\cdot|s)$ is a Gaussian measure on $\mathbb{R}$ having density $\rho_{g_\zeta(s),\sigma}(\cdot)$ w.r.t. the Lebesgue measure on $\mathbb{R}$. Consequently, for every $\zeta \in \mathbb{R}^f$ and $s \in \mathcal{D}$, the measures $Q^{\zeta}(\cdot|s)$ and $Q^{\mu}(\cdot|s)$ are mutually absolutely continuous, and

$$\left.\frac{\mathrm{d}Q^{\mu}(\cdot|s)}{\mathrm{d}Q^{\zeta}(\cdot|s)}\right|_y = \frac{\rho_{g_\mu(s),\sigma}(y)}{\rho_{g_\zeta(s),\sigma}(y)}. \tag{14}$$

We are now ready to begin the proof of Theorem 3.2.

Proof of Theorem 3.2 Consider an alternative reward model given by $\zeta \in \mathrm{Alt}_\varepsilon(\mu)$. For each $t \in \mathbb{Z}_+$, define the log-likelihood ratio $\mathcal{L}_t^{\mu,\zeta} \stackrel{\text{def}}{=} \ln \frac{\mathrm{d}\mathbb{P}_t^{\mathcal{A},\mu}}{\mathrm{d}\mathbb{P}_t^{\mathcal{A},\zeta}}$, and note that $\mathcal{L}_t^{\mu,\zeta}$ is a random variable on $(\Omega_t, \mathcal{F}_t, \mathbb{P}_t^{\mathcal{A},\mu})$. It is now easy to see from (12), (13) and (14) that, for each $t \in \mathbb{Z}_+$ and each $h_t = \{(s_i, y_i, u_i)\}_{i=1}^t \in \Omega_t$, we have

$$\mathcal{L}_t^{\mu,\zeta}(h_t) = \sum_{i=1}^{t}[\ln \rho_{g_\mu(s_i),\sigma}(y_i) - \ln \rho_{g_\zeta(s_i),\sigma}(y_i)]. \tag{15}$$

Next, define the event $\mathcal{E}^\mu \stackrel{\text{def}}{=} \{\arg\max_{s\in\mathcal{D}} g_\mu(s) \subseteq \mathbb{F}(h_\tau) \subseteq \mathcal{O}_\varepsilon(\mu)\}$, and note that $\mathcal{E}^\mu$ is contained in the σ-algebra $\mathcal{F}_\tau$ generated by the stopping time τ. It follows from Lemma 19 of Kaufmann et al. [2016] that

$$\mathbb{E}^{\mathcal{A},\mu}(\mathcal{L}_\tau^{\mu,\zeta}) \geq \mathrm{kl}(\mathbb{P}^{\mathcal{A},\mu}(\mathcal{E}^\mu), \mathbb{P}^{\mathcal{A},\zeta}(\mathcal{E}^\mu)), \tag{16}$$

where $\mathrm{kl}(\nu_1, \nu_2)$ is the KL-divergence between two Bernoulli distributions having parameters $\nu_1, \nu_2 \in [0, 1]$.

Next, define the event $\mathcal{E}^\zeta$ by replacing μ in the definition of $\mathcal{E}^\mu$ with ζ. Since $\mathcal{A}$ is a (ε, δ)-PAC algorithm, we have $\mathbb{P}^{\mathcal{A},\mu}(\mathcal{E}^\mu) \geq 1 - \delta$ and $\mathbb{P}^{\mathcal{A},\zeta}(\mathcal{E}^\zeta) \geq 1 - \delta$. By our choice of ζ, we have $\mathcal{E}^\mu \cap \mathcal{E}^\zeta \subseteq \{\mathbb{F}(h_\tau) \subseteq \mathcal{O}_\varepsilon(\mu) \cap \mathcal{O}_\varepsilon(\zeta)\} = \varnothing$. As a result, we infer that $\mathbb{P}^{\mathcal{A},\zeta}(\mathcal{E}^\mu) < \delta$. Monotonicity properties of the KL divergence now imply that $\mathrm{kl}(\mathbb{P}^{\mathcal{A},\mu}(\mathcal{E}^\mu), \mathbb{P}^{\mathcal{A},\zeta}(\mathcal{E}^\mu)) \geq \mathrm{kl}(\delta, 1-\delta)$ By inequality (3) in Kaufmann et al. [2016], we further have $\mathrm{kl}(\delta, 1-\delta) \geq \ln(1/2.4\delta)$. Using this in (16), we get

$$\mathbb{E}^{\mathcal{A},\mu}(\mathcal{L}_\tau^{\mu,\zeta}) \geq \ln\left(\frac{1}{2.4\delta}\right). \tag{17}$$

Accepted for the 38[th] Conference on Uncertainty in Artificial Intelligence (UAI 2022).

Combining (17) with (18) from Lemma A.1 below yields

$$\frac{1}{2\sigma^2}\mathbb{E}^{\mathcal{A},\mu}(\tau)\|g_\mu - g_\zeta\|_\infty^2 \geq \ln\left(\frac{1}{2.4\delta}\right).$$

Inequality (2) now follows by taking an infimum over $\zeta \in \mathrm{Alt}_\varepsilon(\mu)$ on the left hand side in the inequality above and rearranging the resulting inequality. □

Lemma A.1. *Let algorithm $\mathcal{A}$ and $\mu, \zeta \in \mathbb{R}^f$ be as in the proof of Theorem (4.2). Suppose $\{\mathcal{L}_t^{\mu,\zeta}\}_{t=1}^\infty$ is defined as in (15), and let τ be a stopping time with respect to the filtration $\{\mathcal{F}_t\}_{t=0}^\infty$. Then we have*

$$\mathbb{E}^{\mathcal{A},\mu}(\mathcal{L}_\tau^{\mu,\zeta}) \leq \frac{1}{2\sigma^2}\mathbb{E}^{\mathcal{A},\mu}(\tau)\|g_\mu - g_\zeta\|_\infty^2. \tag{18}$$

Proof. For each $t \in \mathbb{Z}_+$, denote $\ell_t = \ln \rho_{g_\mu(s_t),\sigma}(y_t) - \ln \rho_{g_\zeta(s_t),\sigma}(y_t)$, and let $\mathcal{G}_t$ denote the σ-algebra on Ω generated by (h_{t-1}, u_t, s_t). Note that, for each $t \in \mathbb{Z}_+$, ℓ_t is a $\mathcal{F}_t$-measurable random variable, while $\mathcal{F}_{t-1} \subseteq \mathcal{G}_t \subseteq \mathcal{F}_t$.

Next, define the process $\{M_t\}_{t=0}^\infty$ by $M_0 = 0$ and $M_t = \sum_{i=1}^t [\ell_i - \mathbb{E}^{\mathcal{A},\mu}(\ell_i|\mathcal{G}_i)]$ for each $t \in \mathbb{Z}_+$. The inclusions $\mathcal{F}_{t-1} \subseteq \mathcal{G}_t \subseteq \mathcal{F}_t$ along with the tower property of conditional expectations show that the process $\{M_t\}_{t=0}^\infty$ is adapted to the filtration $\{\mathcal{F}_t\}_{t=0}^\infty$ and is a martingale under the measure $\mathbb{P}^{\mathcal{A},\mu}$. The optional stopping theorem now implies that $\mathbb{E}^{\mathcal{A},\mu}(M_\tau) = \mathbb{E}^{\mathcal{A},\mu}(M_0) = 0$. This immediately yields

$$\mathbb{E}^{\mathcal{A},\mu}\left[\sum_{i=1}^{\tau}\ell_i\right] = \mathbb{E}^{\mathcal{A},\mu}\left[\sum_{i=1}^{\tau}\mathbb{E}^{\mathcal{A},\mu}(\ell_i|\mathcal{G}_i)\right]. \tag{19}$$

Substituting the expression for a Gaussian density in the expression for ℓ_i yields $2\sigma^2\ell_i = 2y_i[g_\mu(s_i) - g_\zeta(s_i)] + [g_\zeta(s_i)]^2 - [g_\mu(s_i)]^2$ for each $i \in \mathbb{Z}_+$. Using the fact that $\mathbb{E}^{\mathcal{A},\mu}(y_i) = g_\mu(s_i)$ gives $2\sigma^2\mathbb{E}^{\mathcal{A},\mu}(\ell_i|\mathcal{G}_i) = [g_\zeta(s_i) - g_\mu(s_i)]^2 \leq \max_{s\in\mathcal{D}}[g_\mu(s) - g_\zeta(s)]^2$ for each $i \in \mathbb{Z}_+$. Using the last inequality in (19) and recognizing the left hand side of (19) to be $\mathbb{E}^{\mathcal{A},\mu}(\mathcal{L}_\tau^{\mu,\zeta})$ yields (18). □

B PROOFS FOR SUBSECTIONS 4.1 AND 4.3

Proof of Proposition 4.1. Choose $s^* \in \arg\max_{s\in\mathcal{D}} q(s)$, and consider $s \in \mathcal{D}'$. We have $q(s^*) \leq \hat{q}(s^*) + \frac{\varepsilon}{4} \leq \hat{q}(\hat{s}) + \frac{\varepsilon}{4} \leq \hat{q}(s) + \frac{3\varepsilon}{4} \leq q(s) + \varepsilon$, where the first and last inequalities follow from $\|\hat{q} - q\|_\infty \leq \frac{\varepsilon}{4}$, the second inequality follows from the definition of $\hat{s}$, and the third follows from the definition of $\mathcal{D}'$ and our choice $s \in \mathcal{D}'$. We have thus shown that every $s \in \mathcal{D}'$ is ε-optimal for q.

Next, we have $\hat{q}(s^*) \geq q(s^*) - \frac{\varepsilon}{4} \geq q(\hat{s}) - \frac{\varepsilon}{4} \geq \hat{q}(\hat{s}) - \frac{\varepsilon}{2}$, where the first and the last inequalities follow from $\|\hat{q} - q\|_\infty \leq \frac{\varepsilon}{4}$ while the second inequality follows from the fact that s^* is a maximizer of q. We have thus shown that $s^* \in \mathcal{D}'$. Since $s^* \in \arg\max_{s\in\mathcal{D}} q(s)$ was chosen arbitrarily, the last assertion of the result follows. □

Proof of Lemma 4.2. First, we recall a definition from the theory of optimal designs. A *design* is a probability measure on the Borel σ-algebra of $\mathcal{D}$. Given a design ξ on $\mathcal{D}$, we denote $V_\xi = \int_{\mathcal{D}} \phi(s)\phi^{\mathrm{T}}(s)\xi(\mathrm{d}s)$. Note that the integral is defined since $\mathcal{D}$ is compact and ϕ is continuous.

Next, suppose $C = \{\phi(p_1), \ldots, \phi(p_m)\}$ is a (L, m)-volumetric spanner for some $L > 0$ and $m \geq f$. Consider the design ξ which places mass $1/m$ at each of the points of C, and let $X = [\phi(p_1), \ldots, \phi(p_m)] \in \mathbb{R}^{f\times m}$. Note that $XX^{\mathrm{T}} = mV_\xi$.

By the definition of a (L, m)-volumetric spanner, we have $\max_{z\in\phi(\mathcal{D})}\|X^{\mathrm{T}}(XX^{\mathrm{T}})^{-1}z\|_2^2 \leq L^2$. A simple calculation shows that, for each $z \in \phi(\mathcal{D})$, we have $\|X^{\mathrm{T}}(XX^{\mathrm{T}})^{-1}z\|_2^2 = z^{\mathrm{T}}(XX^{\mathrm{T}})^{-1}z = m^{-1}z^{\mathrm{T}}V_\xi z$. The Keifer-Wolfowitz theorem [Kiefer and Wolfowitz, 1960], [Lattimore and Szepesvári, 2020, Thm. 21.1] implies that $\max_{z\in\phi(\mathcal{D})} x^{\mathrm{T}}V_\xi x \geq f$. Putting everything together, we have $L^2 \geq m^{-1}\max_{z\in\phi(\mathcal{D})} z^{\mathrm{T}}V_\xi z \geq f/m$. This completes the proof. □

C PROOF OF PROPOSITION 4.3

By way of preparation for the proof of Proposition 4.3, we will find it convenient to rewrite (3) and (4) by grouping together observations made during each round. To this end, let $B_{L,m} = [\phi(p_1), \ldots, \phi(p_m)] \in \mathbb{R}^{f\times m}$ and, for each $j \in \mathbb{Z}_+$, let $\bar{y}^j = [y_{(j-1)m+1}, \ldots, y_{jm}]^{\mathrm{T}} \in \mathbb{R}^m$ and $\bar{\eta}^j = [\eta_{(j-1)m+1}, \ldots, \eta_{jm}]^{\mathrm{T}} \in \mathbb{R}^m$ denote the vectors of rewards and noise samples, respectively, encountered in the jth round. The decision epoch at the end of $k > 0$ rounds is $t = km$. In the notation of (3), we have

$$X_t = [\underbrace{B_{L,m}|\cdots|B_{L,m}}_{k \text{ times}}].$$

Equations (3)-(4) now become

$$\hat{\mu}_{km} = (B_{L,m}B_{L,m}^{\mathrm{T}})^{-1}B_{L,m}\left[\frac{1}{k}\sum_{j=1}^{k}\bar{y}^j\right], \tag{20}$$

$$\hat{\mu}_{km} - \mu = (B_{L,m}B_{L,m}^{\mathrm{T}})^{-1}B_{L,m}\left[\frac{1}{k}\sum_{j=1}^{k}\bar{\eta}^j\right]. \tag{21}$$

The proof of the sub-Gaussian part of Proposition 4.3 essentially applies to the right hand side of (21) the tail concentration inequality below for the norm of the average of k random vectors having independent σ-sub-Gaussian components. The proof is given later in this appendix.

Proposition C.1. *Suppose $\xi^1, \ldots, \xi^k$ are f-dimensional random vectors such that the random variables $\{\xi_i^j : i = 1, \ldots, f, j = 1, \ldots, k\}$ are independent and σ-sub Gaussian. Let $S_k = \left(\xi^1 + \cdots + \xi^k\right)$. Then the following statements hold.*

1. *$\exp\left(\lambda \|S_k\|_2^2\right)$ is integrable for each $\lambda \in \left(0, 1/2\sigma^2 k\right)$.*
2. *For every $\varepsilon > 0$, we have $\mathbb{P}\left(\frac{1}{k}\|S_k\|_2 > \varepsilon\right) \leq \beta(k, \varepsilon)$, where β is given by (7).*

We are now ready to prove Proposition 4.3.

Proof of Proposition 4.3. First, suppose Assumption 1 holds. We have

$$\begin{aligned}\|g_{\hat{\mu}_{km}} - g_\mu\|_\infty &= \max_{s \in \mathcal{D}} |\phi^{\mathrm{T}}(s)(\hat{\mu}_{km} - \mu)| \\ &= \max_{s \in \mathcal{D}} \left| \phi^{\mathrm{T}}(s)(B_{L,m} B_{L,m}^{\mathrm{T}})^{-1} B_{L,m} \left[\frac{1}{k} S_k\right] \right|, \end{aligned} \quad (22)$$

where $S_k = \sum_{j=1}^k \bar{\eta}^j$, and the last equality uses (21). Since the columns of $B_{L.m}$ form a (L, m)-volumetric spanner for $\phi(\mathcal{D})$, it follows that $\|B_{L,m}^{\mathrm{T}}(B_{L,m}B_{L,m}^{\mathrm{T}})^{-1}\phi(s)\|_2 \leq L$ for all $s \in \mathcal{D}$. Using this fact along with the Cauchy-Schwarz inequality in (22) gives $\|g_{\hat{\mu}_{km}} - g_\mu\|_\infty \leq \frac{L}{k}\|S_k\|_2$. The assertion of the proposition now follows immediately from Proposition C.1. □

For proving Proposition C.1, we first recollect a few preliminary results. Though these results are known, we state them to make the constants explicit, and provide proofs for easy reference.

Lemma C.2. *Suppose X is σ-sub Gaussian for some $\sigma > 0$. If $\lambda \in (0, 1/2\sigma^2)$, then $\exp\left(\lambda X^2\right)$ is integrable, and $\mathbb{E}\left[\exp\left(\lambda X^2\right)\right] \leq 2^{2\sigma^2\lambda}(1 - 2\sigma^2\lambda)^{-1}$.*

Proof. Let $\lambda \in (0, 1/2\sigma^2)$. Since X is σ-sub Gaussian, we have $\mathbb{P}(|X| > t) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$ for all $t > 0$ Next, note that $\exp\left(\lambda X^2\right) \geq 1$. Let $s \geq 1$. Then

$$\begin{aligned}\mathbb{P}\left(\exp\left(\lambda X^2\right) > s\right) &= \mathbb{P}\left(|X| > \sqrt{\frac{\ln s}{\lambda}}\right) \\ &\leq 2\exp\left(-\frac{1}{2\sigma^2}\frac{\ln s}{\lambda}\right) = 2s^{-\frac{1}{2\sigma^2\lambda}}.\end{aligned}$$

Thus, we conclude that

$$\mathbb{P}\left(\exp\left(\lambda X^2\right) > s\right) \leq \begin{cases} 1, & \text{if } s \leq 2^{2\sigma^2\lambda}, \\ 2s^{-\frac{1}{2\sigma^2\lambda}}, & \text{if } s > 2^{2\sigma^2\lambda}. \end{cases} \quad (23)$$

Since $2\sigma^2\lambda < 1$, the integral $\int_0^\infty \mathbb{P}\left(\exp\left(\lambda x^2\right) > s\right) \mathrm{d}s$ exists. Indeed, (23) implies that

$$\begin{aligned}&\int_0^\infty \mathbb{P}\left(\exp\left(\lambda x^2\right) > s\right) \mathrm{d}s \\ &\leq \int_0^{2^{2\sigma^2\lambda}} 1 \mathrm{d}s + \int_{2^{2\sigma^2\lambda}}^\infty 2s^{-\frac{1}{2\sigma^2\lambda}} \mathrm{d}s \\ &= 2^{2\sigma^2\lambda}(1 - 2\sigma^2\lambda)^{-1}.\end{aligned}$$

Since $\exp\left(\lambda x^2\right)$ is a non-negative random variable, it follows that $\mathbb{E}\left[\exp\left(\lambda x^2\right)\right] = \int_0^\infty \mathbb{P}\left(\exp\left(\lambda x^2\right) > s\right) \mathrm{d}s$, and the result follows. □

Lemma C.3. *Suppose ζ is a random vector of dimension f such that $\zeta_1, \ldots, \zeta_f$ are independent σ-sub Gaussian random variables. Then $\mathbb{E}\left[\exp\left(\lambda x^{\mathrm{T}}\zeta\right)\right] \leq \exp\left(\frac{\lambda^2\|x\|_2^2\sigma^2}{2}\right)$ for all $x \in \mathbb{R}^f$ and $\lambda \in \mathbb{R}$. Furthermore, if $\lambda \in (0, 1/2\sigma^2)$, then $\exp\left(\lambda\|\zeta\|_2^2\right)$ is integrable, and $\mathbb{E}\left[\exp\left(\lambda\|\zeta\|_2^2\right)\right] \leq \left(\frac{2^{2\sigma^2\lambda}}{1-2\sigma^2\lambda}\right)^f$.*

Proof. By independence and σ-sub Gaussianity, we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda x^{\mathrm{T}}\zeta\right)\right] &= \prod_{i=1}^f \mathbb{E}\left[\exp\left(\lambda x_i \zeta_i\right)\right] \\ &\leq \prod_{i=1}^f \exp\left(\frac{\lambda^2 x_i^2 \sigma^2}{2}\right) = \exp\left(\frac{\lambda^2\|x\|_2^2\sigma^2}{2}\right).\end{aligned}$$

This proves the first assertion. To prove the second assertion, let $\lambda \in \left(0, 1/2\sigma^2\right)$. By Lemma C.2, $\exp\left(\lambda\zeta_i^2\right)$ is integrable for each i. Hence it follows by independence that $\exp\left(\lambda\|\zeta\|_2^2\right)$ is also integrable, and $\mathbb{E}\left[\exp\left(\lambda\|\zeta\|_2^2\right)\right] = \prod_{i=1}^f \mathbb{E}\left[\exp\left(\lambda\zeta_i^2\right)\right] \leq \left(\frac{2^{2\sigma^2\lambda}}{1-2\sigma^2\lambda}\right)^f$. □

The next lemma, which we state without proof, is a conditional version of the first part of Lemma C.3.

Lemma C.4. *Suppose ζ is a random vector of dimension f such that $\zeta_1, \ldots, \zeta_f$ are independent, σ-sub Gaussian random variables. Let Y be a $\mathcal{G}$-measurable f-dimensional random vector, where $\mathcal{G}$ is a σ-algebra such that ζ is independent of $\mathcal{G}$. Then*

$$\mathbb{E}\left[\exp\left(\lambda Y^{\mathrm{T}}\zeta\right) |\mathcal{G}\right] \leq \exp\left(\frac{\lambda^2\|Y\|_2^2\sigma^2}{2}\right) \text{ a.s.}$$

The proof of Proposition C.1 follows next.

Proof of Proposition C.1. The ith component of S_k is a sum of k independent σ-sub Gaussian random variables. Applying the first part of Lemma C.3 with $\zeta = \left[\xi_i^1, \ldots, \xi_i^k\right]$ and $x = [1, \ldots, 1]$ lets us conclude that the i^{th} element of S_k is $\sigma\sqrt{k}$-sub Gaussian. Applying the second part of

Lemma C.3 with $\zeta = S_k$ shows that $\exp\left(\lambda \|S_k\|_2^2\right)$ is integrable for $\lambda \in \left(0, 1/2\sigma^2 k\right)$. This proves the first assertion.

To prove the second statement, choose $x \in \mathbb{R}^f$, and define the process $\{M_j(x)\}_{j=0}^k$ by $M_0(x) = 1$ and

$$M_j(x) = \exp\left(x^{\mathrm{T}} S_j - j\frac{\sigma^2}{2}\|x\|_2^2\right), j = 1, \ldots, k,$$

where $S_j = \xi^1 + \cdots + \xi^j$ for each j. It follows from the first part of Lemma C.3 that $M_j(x)$ is integrable for each j. Next, let $\mathcal{G}_j$ denote the σ-algebra generated by $\xi^1, \ldots, \xi^j$, with $\mathcal{G}_0$ denoting the trivial σ-algebra, and note that $M_j(x)$ is $\mathcal{F}_j$-measurable. For each $j = 1, \ldots, k$, we have

$$\begin{aligned}
&\mathbb{E}\left[M_j(x)|\mathcal{G}_{j-1}\right] \\
&= \mathbb{E}\left[M_{j-1}(x)\exp\left(x^{\mathrm{T}}\xi^j - \frac{\sigma^2}{2}\|x\|_2^2\right)\middle|\mathcal{G}_{j-1}\right] \\
&= M_{j-1}(x)\mathbb{E}\left[\exp\left(x^{\mathrm{T}}\xi^j - \frac{\sigma^2}{2}\|x\|_2^2\right)\right] \\
&\leq M_{j-1}(x),
\end{aligned}$$

where the second equality follows from the $\mathcal{G}_{j-1}$-measurability of $M_{j-1}(x)$ and the $\mathcal{G}_{j-1}$-independence of ξ^j (see Lemma C.4), while the last inequality follows by applying the first part of Lemma C.3 with $\zeta = \xi^j$. Thus, $\{M_j(x)\}_{j=0}^k$ is a supermartingale with respect to the filtration $\{\mathcal{G}_j\}_{j=0}^k$.

Next, define $\{\bar{M}_j\}_{j=0}^k$ by

$$\bar{M}_j = \left(\frac{k\sigma^2}{2\pi}\right)^{f/2} \int_{\mathbb{R}^f} M_j(x)\exp\left(\frac{-k\sigma^2}{2}x^{\mathrm{T}}x\right)\mathrm{d}x, \tag{24}$$

and note that $\bar{M}_0 = 1$. Substituting for $M_j(x)$ in (24), completing the square in the exponent and rearranging terms yields

$$\bar{M}_j = \left[\left(\frac{k}{(j+k)}\right)^{\frac{f}{2}} \exp\left(\frac{\|S_j\|_2^2}{2(j+k)\sigma^2}\right)\right] \times J, \tag{25}$$

where J is the integral over x of the F-dimensional Gaussian density over x with mean $[(j+k)\sigma^2]^{-1}S_j$ and covariance matrix $[(j+k)\sigma^2]^{-1}I$, with I denoting the $f \times f$ identity matrix. Thus, J evaluates to 1. Next, S_j is a random vector with independent $\sigma\sqrt{j}$-sub Gaussian components. Also, $\frac{1}{2(j+k)\sigma^2} < \frac{1}{2j\sigma^2}$. Hence, by Lemma C.3, $\bar{M}_j$ is integrable. In addition, it follows from Lemma 20.3 in Lattimore and Szepesvári [2020] that $\{\bar{M}_j\}_{j=1}^k$ is a submartingale.

Letting $j = k$ in (25) gives $\bar{M}_k = 2^{-\frac{f}{2}}\exp(\frac{\|S_k\|_2^2}{4k\sigma^2})$. By Ville's maximal inequality (see Theorem 3.9 in Lattimore and Szepesvári [2020]), we have

$$\begin{aligned}
\mathbb{P}\left(\|S_k\|_2 > \epsilon\right) &= \mathbb{P}\left(\bar{M}_k > \frac{1}{2^{f/2}}\exp\left(\frac{\epsilon^2}{4k\sigma^2}\right)\right) \\
&\leq \mathbb{P}\left(\max_j \bar{M}_j > \frac{1}{2^{f/2}}\exp\left(\frac{\epsilon^2}{4k\sigma^2}\right)\right) \\
&\leq \frac{\mathbb{E}\left[\bar{M}_0\right]}{\frac{1}{2^{f/2}}\exp\left(\frac{\epsilon^2}{4k\sigma^2}\right)} \\
&= 2^{f/2}\exp\left(\frac{-\epsilon^2}{4k\sigma^2}\right).
\end{aligned}$$

Replacing ε by $k\varepsilon$ in the last inequality completes the proof of the second assertion. □

D PROOF OF PROPOSITION 5.1

The proof of Proposition 5.1 uses the following lemma.

Lemma D.1. *Let $s \in [p_{\min}, p_{\max}]$ and suppose $p_1, \ldots, p_f \in [p_{\min}, p_{\max}]$ are such that $p_i \neq p_j$ for all $i \neq j$. Then $c_1, \ldots, c_{n+1} \in \mathbb{R}$ satisfy*

$$c_1\phi(p_1) + \cdots + c_{n+1}\phi(p_f) = \phi(s) \tag{26}$$

if and only if $c_i = l_i(s, \mathbf{p})$ for each $i = 1, \ldots, f$ where $\mathbf{p} = [p_1, \ldots, p_f]^{\mathrm{T}}$, and $l_i(\cdot, \mathbf{p})$ is the ith Lagrange basis polynomial for the points $\{p_1, p_2, \ldots, p_f\}$ given by

$$l_i(s, \mathbf{p}) \stackrel{\mathrm{def}}{=} \frac{\prod_{j\neq i}(s - p_j)}{\prod_{j\neq i}(p_i - p_j)}. \tag{27}$$

Proof. Equation (26) may be rewritten as

$$V(\mathbf{p})c(s) = \phi(s), \tag{28}$$

where $V(\mathbf{p}) \stackrel{\mathrm{def}}{=} [\phi(p_1), \ldots, \phi(p_f)] \in \mathbb{R}^{f\times f}$. Note that $V(\mathbf{p})$ is a Vandermonde matrix, and its determinant is given by (see Fact 7.18.5 from Bernstein [2018])

$$\det(V(\mathbf{p})) = \prod_{1\leq i<j\leq f}(p_j - p_i), \tag{29}$$

The determinant of $V(\mathbf{p})$ in (29) is nonzero since $p_i \neq p_j$ for $j \neq i$. Equation (28) thus has a unique solution. Applying Cramer's rule (see Fact 3.16.12 from Bernstein [2018]) gives this solution to be

$$c_i = \frac{\det(V(\mathbf{p}_i^s))}{\det(V(\mathbf{p}))} \tag{30}$$

where $\mathbf{p}_i^s$ is the vector obtained by replacing the ith element of $\mathbf{p}$ by s. Using (29) to expand the determinants of the two Vandermonde matrices in (30) and canceling common terms gives $c_i = l_i(s, \mathbf{p})$. □

The proof of Proposition 5.1 follows.

Proof of Proposition 5.1. To show 1) implies 2), suppose $p_1, \ldots, p_f \in \mathcal{D}$ are $(1, f)$-volumetric points for the pair $(\phi, \mathcal{D})$. Choose $s \in \mathcal{D} = [p_{\min}, p_{\max}]$ arbitrarily. Applying the definition of $(1, f)$-volumetric points, it follows that there exist $c_1, \ldots, c_f \in \mathbb{R}$ such that $c_1\phi(p_1) + \cdots + c_f\phi(p_f) = \phi(s)$ and $c_1^2 + \ldots + c_f^2 \leq 1$. Clearly, $|c_i| \leq 1$ for all $i = 1, \ldots, f$. Since $s \in \mathcal{D}$ was chosen arbitrarily, it follows that $\{\phi(p_1), \ldots, \phi(p_f)\}$ is a barycentric spanner for $\phi(\mathcal{D})$ (see Amballa et al. [2021] for a definition). Theorem 1 of Amballa et al. [2021] now implies that 2) holds.

To prove that 2) implies 1), suppose $p_{\min} = p_1 \leq p_2 \leq \cdots \leq p_f = p_{\max}$ satisfy (11). Define $\mathbf{p}$ as in Lemma D.1. The Lagrange polynomials defined in Lemma D.1 satisfy

$$
\begin{aligned}
l_i(p_i, \mathbf{p}) &= 1,\ i = 1, \ldots, f, && (31)\\
l_i(p_j, \mathbf{p}) &= 0,\ i, j = i = 1, \ldots, f,\ i \neq j, && (32)\\
\frac{\mathrm{d}l_i}{\mathrm{d}s}(p_i, \mathbf{p}) &= 0,\ i = 2, \ldots, f-1, && (33)
\end{aligned}
$$

$$\frac{\mathrm{d}l_1}{\mathrm{d}s}(p_1, \mathbf{p}) < 0 < \frac{\mathrm{d}l_f}{\mathrm{d}s}(p_f, \mathbf{p}). \tag{34}$$

Equations (31), (32) and the inequalities in (34) follow by substituting appropriate values for s in (27), while (33) follows by differentiating (27) with respect to s, substituting appropriately for s, and then using (11).

Next, define the function $G : \mathcal{D} \to \mathbb{R}$ by $G(s) \stackrel{\text{def}}{=} l_1^2(s, \mathbf{p}) + \cdots + l_f^2(s, \mathbf{p}) - 1$. We claim that $G(s) \leq 0$ for all $s \in \mathcal{D}$. In light of Lemma D.1 and the definition of $(1, f)$-volumetric points, our claim implies that 1) holds. Hence, to complete the proof, it is sufficient to prove our claim.

To prove our claim, note that G is a polynomial of degree $2(f-1)$. Also, we observe from (31), (33) and (34) that p_1 and p_f are roots of G of multiplicity 1, while each p_i is a root of G of multiplicity at least 2 for $i \neq 1, f$. Thus, the polynomial $H(s) \stackrel{\text{def}}{=} (s-p_1)(s-p_2)^2 \cdots (s-p_{f-1})^2(s-p_f)$ divides G. Since H also clearly has degree $2(f-1)$, it follows that there exists $K \in \mathbb{R}$ such that $G(s) = KH(s)$ for all $s \in \mathcal{D}$. The value of K may be computed as $K = \frac{G'(p_1)}{H'(p_1)}$, where $'$ indicates the derivative. It is easy to use (31) and (32) to verify that $G'(p_1) = 2\frac{\mathrm{d}l_1}{\mathrm{d}s}(p_1, \mathbf{p})$, which is negative by (34). An easy calculation also yields $H'(p_1) = (p_1-p_2)^2 \cdots (p_1-p_{f-1})^2(p_1-p_f)$ which is negative since $p_1 < p_f$. These arguments show that $K > 0$. Our claim now follows by noting that H takes only non-positive values on $\mathcal{D}$. This completes the proof. □

E ALGORITHM 2

Algorithm 1 VSBAI-Poly: Best Arm Identification for Polynomial Rewards

Input: $\varepsilon > 0$, $\delta \in (0, 1)$, sub-Gaussianity parameter σ, $(1, f)$-volumetric points $p_1, \ldots, p_f$ for $(\phi, \mathcal{D})$
Set $B_{1,f} = [\phi(p_1), \ldots, \phi(p_f)]$
Initialize $k \leftarrow 1$, $r \leftarrow 0$
Set STOP = False
while STOP==False **do**
  $\bar{y}^k$ = []
  for $t = 1, \ldots, f$ **do**
    $y_{(k-1)m+t} = g_\mu(p_t) + \eta_t$
    $\bar{y}^k \leftarrow [(\bar{y}^k)^{\mathrm{T}}; y_{(k-1)f+t}]^{\mathrm{T}}$
  end for
  $r = r + \bar{y}^k$
  if $\beta(k, \frac{\varepsilon}{4L}) < \delta$ **then**
    STOP = True
  else
    $k = k + 1$
  end if
end while
$\tau^* = kd$
$\hat{\mu}_{\tau^*} = \frac{1}{k} {B_{1,f}}^{-\mathrm{T}} r$
$\hat{s}$ = global_optimizer($\hat{\mu}_{\tau^*}, p_{\min}, p_{\max}$)
$\mathcal{D}_{\tau^*}$ = get_dtau($\hat{\mu}_{\tau^*}, \hat{s}, p_{\min}, p_{\max}, \varepsilon$)
Output: $\mathcal{D}_{\tau^*}$

Function global_optimizer($\hat{\mu}_{\tau^*}, p_{\min}, p_{\max}$)
$\hat{\mu}'_{\tau^*}$ = differentiate($\hat{\mu}_{\tau^*}$)
roots = find_roots($\hat{\mu}'_{\tau^*}$)
roots.add($p_{\min}, p_{\max}$)
values = $g_{\hat{\mu}_{\tau^*}}$(roots)
opt_value = argmax(values)
return opt_value

Function get_dtau($\hat{\mu}_{\tau^*}, \hat{s}, p_{\min}, p_{\max}, \varepsilon$)
d_tau = []
find_roots($g_{\hat{\mu}_{\tau^*}}(s) - g_{\hat{\mu}_{\tau^*}}(\hat{s}) + \varepsilon/2$)
roots.add($p_{\min}, p_{\max}$)
roots = sort(roots)
root_left = get_closest_left_root_to_$\hat{s}$(roots, $\hat{s}$)
root_right = get_closest_right_root_to_$\hat{s}$(roots, $\hat{s}$)
d_tau.add(root_left, root_right)
d_tau.add(every_pair_to_the_left_of_root_left)
d_tau.add(every_pair_to_the_right_of_root_right)
return d_tau

F MULTI-ARM SETTING CONFIGURATIONS

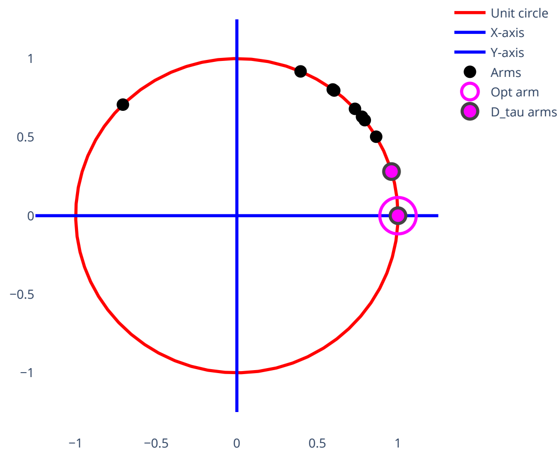

Figure 3: 10 arm setting when the angles ϕ of the arms (3 to 10) are sampled from $\mathcal{N}(0, .09)$. $(\varepsilon, \delta) = (0.1, 0.05)$ for the VSBAI algorithm

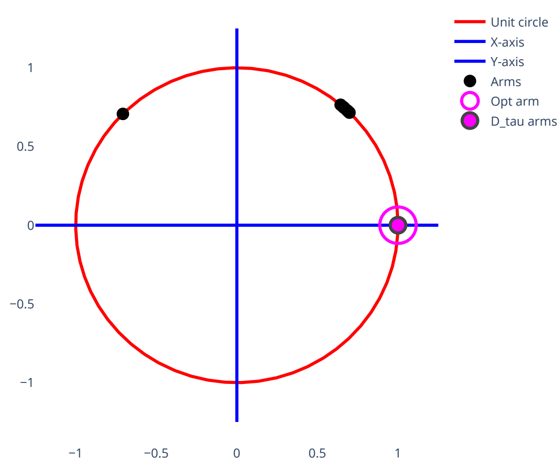

Figure 4: 10 arm setting when the angles ϕ of the arms (3 to 10þ) are sampled uniformly from $[0, 0.1]$. $(\varepsilon, \delta) = (0.1, 0.05)$ for the VSBAI algorithm

G OTHER EXPERIMENTS

We consider the setting outlined in subsection 6.1 and present results for a different configuration of the problem instances. We first note that the implementation of the baseline algorithms presented in Fiez et al. [2019], Jedra and Proutiere [2020a], and Soare et al. [2014] for the setting in subsection 6.1 is true when the angles ϕ_i for $i = 3, \ldots, n$ are sampled from a uniform distribution $[0, 0.1]$ rather than a Gaussian distribution as in subsection 6.1. We therefore present experimental results for this uniform setting and provide a comparison of sample complexity and run time as in subsection 6.1. Note that we are able reproduce the results reported in Jedra and Proutiere [2020b] (see Table 2 and Table 3 of Jedra and Proutiere [2020b]).

We observe from tables 3 that the sample complexity of VSBAI is greater than the other baselines. However, we argue that the instances generated in this setting are simple and in situations where it is difficult to separate out the best-arm from the next best (like when the angles ϕ_i of the arms are sampled from Gaussian Gaussian setting in 6.1), all these baselines suffer from huge sample complexities and run-times. In other words VSBAI is independent of the way the instances are generated but on the other hand all the other baselines are not robust, hence can potentially perform badly in adversarial environments. Table 4 gives a comparison of the run-times for this setting. The results shown are obtained after averaging over 20 seeds.

Algorithm	LazyTS		Rage		Oracle		VSBAI	
Arms	Mean	Std	Mean	Std	Mean	Std	Mean	Std
10	335.1	22.71	524.1	33.84	347.1	32.22	47693.4	105.32
20	423.05	28.92	683.05	92.07	356.15	32.31	47424.1	41.32
100	421.75	32.11	1038.75	148.01	426.4	39.76	47271.4	13.49
1000	419.65	28.43	1152.7	50.23	476.45	40.74	47222.7	1.27
2500	446.15	29.06	1447.3	150.46	481.8	41.33	47219.8	0.37
5000	431.65	32.23	1546.9	160.17	510.05	48.87	47219.9	0.41

Table 3: Expected sample complexity for the setting described in Appendix G

Algorithm	LazyTS		Rage		Oracle		VSBAI	
Arms	Mean	Std	Mean	Std	Mean	Std	Mean	Std
10	0.27	0.01	0.05	0.001	0.001	0.0	1.30	0.02
20	0.33	0.02	0.06	0.00	0.001	0.00	1.39	0.04
100	0.39	0.02	1.09	0.07	0.02	0.00	1.41	0.04
1000	34.78	1.06	27.61	0.33	0.69	0.02	1.44	0.04
2500	211.24	8.22	335.82	2.46	0.65	0.02	1.56	0.04
5000	422.35	22.45	884.32	3.85	0.89	0.03	2.17	0.03

Table 4: Run-time in seconds for the setting described in Appendix G

References

Chaitanya Amballa, Manu K. Gupta, and Sanjay P. Bhat. Computing an efficient exploration basis for learning with univariate polynomial features. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8):6636–6643, May 2021.

Dennis S. Bernstein. *Scalar, Vector, and Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, revised edition, 2018. ISBN 9780691151205.

Dimitri P Bertsekas and Steven E Shreve. *Stochastic optimal control: the discrete-time case*, volume 5. Athena Scientific, 1996. ISBN 9781886529038.

Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems*, 32, 2019.

Yassir Jedra and Alexandre Proutiere. Optimal best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 33:10007–10017, 2020a.

Yassir Jedra and Alexandre Proutiere. Optimal best-arm identification in linear bandits. *arXiv preprint arXiv:2006.16073*, 2020b.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *J. Mach. Learn. Res.*, 17(1):1–42, jan 2016. ISSN 1532-4435.

Jack Kiefer and Jacob Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020. doi: 10.1017/9781108571401.

Jacques Neveu. *Mathematical foundations of the calculus of probability*. Holden-day, 1965.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 27:828–836, 2014.