Electronic Companion to Online Decision Making with Local Differential Privacy

Appendix A: The LDP-OLS Algorithm for Linear Contextual Bandit

In this section, we provide the detailed algorithm for single-parameter setting with the LDP OLS update in Algorithm 9 and Algorithm 10, whose privacy guarantee is analyzed in the Appendix C and regret is analyzed in Appendix D.

Algorithm 9: LDP Epoch-Greedy Algorithm with OLS

Input: Privacy Level ε, δ; Exploration Schedule T_S; Confidence Level η and regularization c_n as in Proposition 5.

Initialization: Setting $\hat{\theta}_0 = \mathbf{0}$.

for $t \leftarrow 1$ *to* T **do**

The server sets $B_t = \mathbf{1}\{t \in T_S\}$

The server sends $(\hat{\theta}_{t-1}, B_t, \varepsilon, \delta)$ to the user-side terminal

The user-side terminal executes the subroutine in Algorithm 10 with input $(\hat{\theta}_{t-1}, B_t, \varepsilon, \delta)$ and generates $(\widetilde{V}_t, \widetilde{b}_t)$.

The server receives the user feedback $(\widetilde{V}_t, \widetilde{b}_t)$ and updates $\hat{\theta}_t$ via

$$\hat{\theta}_t \leftarrow \left(\sum_{i \in T_S \cap [t]} \widetilde{V}_i + c_{n_t} I \right)^{-1} \sum_{i \in T_S \cap [t]} \widetilde{b}_i.$$

end

In Algorithm 9, only observations from exploration rounds are used for estimation. For the purely greedy algorithm, we set $B_t = 0$ and update the estimator at every round. This gives $T_S \cap [t] = [t]$ and $n_t = t$ in the update above.

Algorithm 10: Local Decision Subroutine for OLS

Input: Current Estimator $\hat{\theta}_{t-1}$; Exploration Criterion B_t; Privacy Level ε, δ;

Take action

$$a_t = \begin{cases} \text{argmax}_{a\in[K]} \phi(X_t, a)^\top \hat{\theta}_{t-1} & \text{if } B_t = 0 \\ a \sim \text{UniformDist}[K] & \text{if } B_t = 1 \end{cases}$$

Receive the reward r_t

Generate the sufficient statistics

$$\widetilde{V}_t = \phi(X_t, a_t)\phi(X_t, a_t)^\top + W_t, \qquad \widetilde{b}_t = \phi(X_t, a_t) r_t + w_t. \tag{19}$$

return $(\widetilde{V}_t, \widetilde{b}_t)$

The noises $W_t \sim \text{Wigner}(0, \sigma_W^2)$ and $w_t \sim \mathcal{N}(0, \sigma_w^2 I_d)$ are drawn independently at each round, with scales from Proposition 5, where $0 < \varepsilon \le 1$ and $0 < \delta < 1$.

Appendix B: The LDP-UCB Algorithm for Linear Contextual Bandit

In this section, we give Algorithms 11 and 12 for the single-parameter setting, with c_t from Proposition 5, $\eta = \alpha$, and R_t from (25).

Algorithm 11: LDP LinUCB Algorithm

Input: Privacy Level ε, δ; Confidence Level $\alpha \in (0, 1)$.

Initialization: $M_0 = 0$, $\hat{\theta}_0 = 0$, $C_0 = C_B^4 I$, and $\rho_0 = C_B^2$.

for $t \leftarrow 1$ *to* T **do**

The server sends $(\hat{\theta}_{t-1}, C_{t-1}, \varepsilon, \delta, \rho_{t-1})$ to the user-side terminal.

The user runs Algorithm 12 and returns $(\widetilde{V}_t, \widetilde{b}_t)$.

The server receives the feedback $(\widetilde{V}_t, \widetilde{b}_t)$ and updates

$$M_t = \sum_{i=1}^{t} \widetilde{V}_i, \quad \hat{\theta}_t = (M_t + c_t I)^{-1} \sum_{i=1}^{t} \widetilde{b}_i, \quad C_t = \frac{(M_t + c_t I)^2}{t+1}, \quad \rho_t = \frac{R_t}{\sqrt{t+1}}.$$

end

Algorithm 12: Local Decision Subroutine for UCB

Input: Current Estimator $\hat{\theta}_{t-1}$; Confidence Parameters C_{t-1}, ρ_{t-1}; Privacy Level ε, δ.

1 Take UCB action

$$a_t \in \arg\max_{a\in[K]} \left\{ \phi(X_t, a)^\top \hat{\theta}_{t-1} + \rho_{t-1} \|\phi(X_t, a)\|_{C_{t-1}^{-1}} \right\}.$$

Receive the reward r_t.

2 Generate $(\tilde{V}_t, \tilde{b}_t)$ according to (19).

3 **return** $(\widetilde{V}_t, \widetilde{b}_t)$

Appendix C: Proof of Privacy Guarantee

In this section, we prove the ε-LDP guarantee of Algorithm 1, Algorithm 3 and the (ε, δ)-LDP guarantee of Algorithm 9, Algorithm 11.

We would first introduce several background results in local differential privacy, then prove the ε-LDP guarantee and the (ε, δ)-LDP guarantee in the next two sections.

C.1. Background Results for LDP

We first recall two well-adopted LDP mechanisms in previous literature, which are also used in our algorithm design.

LEMMA 5 **(Gaussian Mechanism (Dwork et al. 2006a, Dwork and Roth 2014))**. *Let* $0 < \varepsilon \leq 1$ *and* $0 < \delta < 1$. *For any* $f : \mathcal{X} \to \mathbb{R}^n$, *let* $\sigma_{\varepsilon,\delta} = \frac{1}{\varepsilon} \sup_{x,x'\in\mathcal{X}} \|f(x) - f(x')\|_2 \sqrt{2\ln(1.25/\delta)}$. *The Gaussian mechanism, which adds random noise independently drawn from distribution* $\mathcal{N}(0, \sigma_{\varepsilon,\delta}^2 I_n)$ *to each output of* f, *ensures* (ε, δ)*-LDP.*

LEMMA 6 **(Private Releaser for l_2-ball (Duchi et al. 2018))**. *For any* $R > 0$ *and* $\varepsilon > 0$, *let* $r_{\varepsilon,d} = R\frac{\sqrt{\pi}}{2}\frac{e^\varepsilon + 1}{e^\varepsilon - 1}\frac{d\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2}+1)}$ *where* Γ *is the Gamma function. For any* $x \in B_R^d$, *consider the mechanism* $\Psi_{\varepsilon,R} : B_R^d \to S_{r_{\varepsilon,d}}^{d-1}$ *of generating* Z_x *as follows. For* $x \neq 0$, *first generate a random vector* $\tilde{X} = (2b-1)x$ *where* b *is a Bernoulli random variable with success probability* $\frac{1}{2} + \frac{\|x\|_2}{2R}$. *Next, generate a random vector* Z_x *via*

$$Z_x \sim \begin{cases} \textit{Unif}\{z \in \mathbb{R}^d : z^T\tilde{X} > 0, \|z\|_2 = r_{\varepsilon,d}\} \textit{ with probability } e^\varepsilon/(1+e^\varepsilon), \\ \textit{Unif}\{z \in \mathbb{R}^d : z^T\tilde{X} \leq 0, \|z\|_2 = r_{\varepsilon,d}\} \textit{ with probability } 1/(1+e^\varepsilon). \end{cases}$$

For $x = 0$, *draw* $Z_x \sim \mathrm{Unif}(S_{r_{\varepsilon,d}}^{d-1})$. *Then* $\Psi_{\varepsilon,R}$ *is* ε*-LDP and* $\mathbb{E}[\Psi_{\varepsilon,R}(x)] = x$. *In addition, by Stirling's approximation we have* $r_{\varepsilon,d} \lesssim \frac{R\sqrt{d}}{\varepsilon}$ *for* $\varepsilon \leq 1$.

C.2. Proof of ε-LDP Guarantees

Proof of Proposition 1. Firstly, since the server only accesses X_t through g_t, it is sufficient to show that g_t is ε-LDP by the post-processing property. Now if conditions (i) to (iii) of Assumption 1 are satisfied, it is guaranteed that $\|\nabla f_t\|_2 \le R$. Thus, applying Lemma 6 leads to the desired claim.

□

Proof of Proposition 3. As in the proof of Proposition 1, since the server only accesses X_t through $\{g_{t,a}\}_{a\in[K]}$, it is sufficient to show that $\{g_{t,a}\}_{a\in[K]}$ is ε-LDP by the post-processing property. Now if conditions (i) to (iii) of Assumption 2 are satisfied, it is guaranteed that $\|\nabla f_t\|_2 \le R$. When $B_t = 1$, condition on the uniformly drawn arm, which is independent of D. It is therefore ε-LDP by Lemma 6.

When $B_t = 0$, fix the historical observation D before round t. Let $p_D(g_1, \ldots, g_K)$ be the output density under D. All unselected arms return independent vectors uniformly distributed on $S^{d-1}_{r_{\varepsilon,d}}$. By the construction in Lemma 6, we have then for any D, D',

$$\frac{p_D(g_1, \ldots, g_K)}{p_{D'}(g_1, \ldots, g_K)} \le \frac{2e^{\varepsilon}/(1+e^{\varepsilon})}{2/(1+e^{\varepsilon})} = e^{\varepsilon}.$$

This gives the desired ε-LDP guarantee. □

C.3. Proof of (ε, δ)-LDP Guarantees

PROPOSITION 7. *Algorithm 11, with the noise scales in Proposition 5, is (ε, δ)-LDP in each user's data.*

Proof of Propositions 5 and 7. Since in both Algorithms 9 and 11 the server accesses X_t only through $(\widetilde{V}_t, \widetilde{b}_t)$ in (19), it suffices by post-processing to show that this pair is (ε, δ)-LDP. This result follows directly from Theorem 2 of Zheng et al. (2020). □

Appendix D: Proof of Regret Guarantees

D.1. Regret Analysis of Algorithm 1

We first provide the following general error bound result of the projected stochastic gradient descent for later reference.

LEMMA 7. *Let $\{\mathcal{F}_n\}_{n\ge 0}$ be a filtration, and let $\Theta \subseteq B_1^d$ be a nonempty closed convex, $\mathcal{F}_0$-measurable set containing $\theta^\star$ almost surely. Let $\hat{\theta}_0 \in B_1^d$ be $\mathcal{F}_0$-measurable and, for $n \ge 1$, set*

$$\hat{\theta}_n = \Pi_\Theta\left(\hat{\theta}_{n-1} - \frac{G_n}{m(n+1)}\right), \qquad e_n = \hat{\theta}_n - \theta^\star,$$

where G_n is $\mathcal{F}_n$-measurable. Suppose $m > 0$, $r \geq 2m$, and

$$\|G_n\|_2 \leq r, \qquad \langle e_{n-1}, \mathbb{E}[G_n \mid \mathcal{F}_{n-1}]\rangle \geq m\|e_{n-1}\|_2^2 \quad \textit{almost surely}.$$

For any integer $N \geq 1$ and $0 < \alpha < 1$, with probability at least $1-\alpha$, simultaneously for $0 \leq n \leq N$,

$$\|e_n\|_2^2 \leq \min\left\{4, \frac{64r^2}{m^2(n+1)} \log\frac{2N}{\alpha}\right\}.$$

Proof. Put $h = r^2/m^2$, $A = 64h\log\frac{2N}{\alpha}$, $V_n = \|e_n\|_2^2$, and $Z_n = \langle \mathbb{E}[G_n \mid \mathcal{F}_{n-1}] - G_n, e_{n-1}\rangle$. Nonexpansiveness of projection and the conditional gradient bound give

$$V_n \leq \left(1 - \frac{2}{n+1}\right) V_{n-1} + \frac{2Z_n}{m(n+1)} + \frac{h}{(n+1)^2}.$$

Multiplying by $n(n+1)$ and summing yields

$$n(n+1)V_n \leq \frac{2}{m}\sum_{i=1}^{n} iZ_i + hn. \tag{20}$$

Let $\rho = \inf\{n \geq 1 : V_n > A/(n+1)\}$ and $S_n = \sum_{i=1}^{n} iZ_i\mathbf{1}\{\rho \geq i\}$, with $S_0 = 0$. The indicators are $\mathcal{F}_{i-1}$-measurable, and $V_{i-1} \leq A/i$ on $\{\rho \geq i\}$, including $i = 1$ since $V_0 \leq 4 \leq A$. Conditional on $\mathcal{F}_{i-1}$, $S_i - S_{i-1}$ has mean zero and lies in an interval of length at most $2r\sqrt{Ai}$. If $\rho = n$, (20) implies $S_n > mn(A-h)/2$. Azuma–Hoeffding's inequality therefore gives, for $1 \leq n \leq N$,

$$\mathbb{P}\{\rho = n\} \leq \exp\left\{-\frac{n(A-h)^2}{4hA(n+1)}\right\} \leq e^{-A/(32h)} \leq \frac{\alpha}{N},$$

where we used $A \geq 2h$ and $n/(n+1) \geq 1/2$. Finally, noticing that $V_n \leq 4$ for all $n \geq 0$ by the projection, taking union bounds over n then gives the desired result. □

D.1.1. Proof of the SGD Estimation Error Result in Epoch-Greedy Algorithm

Proof of Proposition 2. For simplicity of the notation, we first focus on the case $n_t = t$, and let $\mathcal{F}_{t-1}$ contain all information before observing X_t. Then, by unbiasedness of the mechanism as shown in Lemma 6, we have

$$\begin{aligned}\mathbb{E}[g_t^T(\hat{\theta}_{t-1} - \theta^\star)|\mathcal{F}_{t-1}] =& \mathbb{E}[(\mu(\phi(X_t,a_t)^T\hat{\theta}_{t-1}) - \mu(\phi(X_t,a_t)^T\theta^\star))\phi(X_t,a_t)^T(\hat{\theta}_{t-1} - \theta^\star)|\mathcal{F}_{t-1}] \\ \geq& \zeta\mathbb{E}[[\phi(X_t,a_t)^T(\hat{\theta}_{t-1} - \theta^\star)]^2|\mathcal{F}_{t-1}] \geq \zeta\kappa_\Phi\|\hat{\theta}_{t-1} - \theta^\star\|^2,\end{aligned}$$

where the last inequality is from the condition (iv) of Assumption 1. Moreover, notice that $\|g_t\| = r_{\varepsilon,d}$. The projected update in (5) has step size $\eta_t = [\zeta\kappa_\Phi(t+1)]^{-1}$ for these exploration updates. Since

$r_{\varepsilon,d} \geq R \geq 2LC_B^2 \geq 2\zeta\kappa_\Phi$, we can apply Lemma 7 with $\Theta = B_1^d$, $m = \zeta\kappa_\Phi$, $r = r_{\varepsilon,d}$ and $N = T$ to obtain

$$\|\hat{\theta}_t - \theta^\star\|_2^2 \leq \min\left\{4, \frac{64 r_{\varepsilon,d}^2}{\zeta^2\kappa_\Phi^2(t+1)} \log\frac{2T}{\alpha}\right\}, \qquad 0 \leq t \leq T.$$

Substituting $r_{\varepsilon,d} \lesssim R\sqrt{d}/\varepsilon$ gives the stated rate.

Finally, when $n_t \neq t$, the result holds directly with t replaced by n_t in RHS since the estimator remains unchanged between exploration steps. This finishes the proof of Proposition 2. □

D.1.2. Proof of Theorem 1

Proof of Theorem 1. Fix $T \geq 1$ and $0 < \alpha < 1$. Note that for every T, we have

$$\mathrm{Reg}(T) \leq \sum_{t\in[T]\setminus T_S} \left[\max_{a\in[K]} \mu(\phi(X_t,a)^\top\theta^\star) - \mu(\phi(X_t,a_t)^\top\theta^\star)\right] + 2LC_B|T_S \cap [T]|.$$

By our selection $T_S = \{t_i : t_i = t_{i-1} + \lceil\sqrt{i}\rceil, t_0 = 1, i \in \mathbb{N}\}$, $n_t = |T_S \cap [t]| \leq 3t^{2/3}$. At an exploitation round $t_j \leq t < t_{j+1}$, we have $n_{t-1} \geq j$ and $t \leq 1 + \sum_{i=1}^{j+1}\lceil\sqrt{i}\rceil \leq 3(n_{t-1}+1)^{3/2}$. On the other hand, since a_t is the greedy action for $t \notin T_S$, we have

$$\max_{a\in[K]} \mu(\phi(X_t,a)^\top\theta^\star) - \mu(\phi(X_t,a_t)^\top\theta^\star) \leq 2L \max_{a\in[K]} |\phi(X_t,a)^\top(\theta^\star - \hat{\theta}_{t-1})| \leq 2LC_B\|\hat{\theta}_{t-1} - \theta^\star\|_2. \tag{21}$$

This follows by adding and subtracting the two estimated rewards, using the greedy choice and Lipschitz continuity. Proposition 2 and its proof imply, with probability at least $1-\alpha$, simultaneously at all exploitation rounds up to T,

$$\begin{aligned}\max_{a\in[K]} \mu(\phi(X_t,a)^\top\theta^\star) - \mu(\phi(X_t,a_t)^\top\theta^\star) &\leq \frac{16LC_B r_{\varepsilon,d}}{\zeta\kappa_\Phi}\sqrt{\frac{1}{n_{t-1}+1}\log\frac{2T}{\alpha}} \\ &\leq \frac{16\cdot 3^{1/3}LC_B r_{\varepsilon,d}}{\zeta\kappa_\Phi}\sqrt{\log\frac{2T}{\alpha}}\, t^{-1/3}.\end{aligned}$$

Since $\sum_{t=1}^s t^{-1/3} \leq (3/2)s^{2/3}$, on this same event,

$$\mathrm{Reg}(s) \leq 6LC_B\left(1 + \frac{8r_{\varepsilon,d}}{\zeta\kappa_\Phi}\sqrt{\log\frac{2T}{\alpha}}\right)s^{2/3}, \qquad 1 \leq s \leq T.$$

Taking $\alpha = T^{-2}$ and using $r_{\varepsilon,d} \lesssim R\sqrt{d}/\varepsilon$ gives the desired result. □

D.1.3. Proof of the SGD Estimation Error Result under the Diversity Condition

Proof of Lemma 3. As in the proof of Proposition 2, we have

$$\begin{aligned}
&\mathbb{E}[g_t^\top(\hat{\theta}_{t-1}-\theta^\star)\mid\mathcal{F}_{t-1}]\\
&=\mathbb{E}[(\mu(\phi(X_t,a_t)^\top\hat{\theta}_{t-1})-\mu(\phi(X_t,a_t)^\top\theta^\star))\phi(X_t,a_t)^\top(\hat{\theta}_{t-1}-\theta^\star)\mid\mathcal{F}_{t-1}]\\
&\geq\zeta\mathbb{E}[(\phi(X_t,a_t)^\top(\hat{\theta}_{t-1}-\theta^\star))^2\mid\mathcal{F}_{t-1}]\geq\frac{\zeta\kappa_l p_*}{d}\|\hat{\theta}_{t-1}-\theta^\star\|_2^2,
\end{aligned}$$

where the last inequality is from the diversity condition in Assumption 3 and Markov's inequality. Thus then replacing κ_Φ by $\kappa_l p_*/d$ in proof of Proposition 2 finishes the proof of Lemma 3. In particular, Lemma 7 applies with $m=\zeta\kappa_l p_*/d$ and $r=r_{\varepsilon,d}\geq 2LC_B^2\geq 2\zeta\kappa_l p_*/d$, including the initial estimate. □

D.1.4. Proof of Theorem 2 and Theorem 4

Proof of Theorem 2. Since $\phi(X_t,a_t)$ is selected using a greedy strategy, we have $\mu(\phi(X_t,a_t)^\top\hat{\theta}_{t-1})\geq\mu(\phi(X_t,a)^\top\hat{\theta}_{t-1})$ for all $t\in[T]$ and $a\in[K]$. Consequently we have the following upper bound for the instantaneous regret at every time t,

$$\max_{a\in[K]}\mu(\phi(X_t,a)^\top\theta^\star)-\mu(\phi(X_t,a_t)^\top\theta^\star)\leq 2LC_B\|\hat{\theta}_{t-1}-\theta^\star\|_2\leq\frac{16LC_B d r_{\varepsilon,d}}{\zeta\kappa_l p_*}\sqrt{\frac{1}{t}\log\frac{2T}{\alpha}},$$

by (21) and Lemma 3. Summing over t gives

$$\mathrm{Reg}(T)\leq\frac{32LC_B d r_{\varepsilon,d}}{\zeta\kappa_l p_*}\sqrt{T\log\frac{2T}{\alpha}}$$

with probability at least $1-\alpha$. Taking $\alpha=T^{-2}$ proves the desired result. □

Proof of Theorem 4. Since $\phi(X_t,a_t)$ is selected using a greedy strategy, we have $\mu(\phi(X_t,a_t)^\top\hat{\theta}_{t-1})\geq\mu(\phi(X_t,a)^\top\hat{\theta}_{t-1})$ for all $t\in[T]$ and $a\in[K]$. By the bounded-feature condition,

$$\max_a|\phi(X_t,a)^\top(\hat{\theta}_{t-1}-\theta^\star)|\leq C_B\|\hat{\theta}_{t-1}-\theta^\star\|_2.$$

Combine this with Lemma 3, using failure probability $\alpha/3$, and (21). With

$$U:=\frac{16LC_B d r_{\varepsilon,d}}{\zeta\kappa_l p_*}\sqrt{\log\frac{6T}{\alpha}},\quad h_t:=\min\{2LC_B,U/\sqrt{t}\},$$

on an event of probability at least $1-\alpha/3$, we have $\max_{a\in[K]}\mu(\phi(X_t,a)^\top\theta^\star)-\mu(\phi(X_t,a_t)^\top\theta^\star)\leq h_t$. On this event, a positive regret implies $\triangle_t\leq h_t$. Therefore $\mathrm{Reg}(T)\leq\sum_{t=1}^T A_t$, with $A_t=h_t\mathbf{1}\{\triangle_t\leq h_t\}$. The A_t are independent, with $0\leq A_t\leq 2LC_B$ and $\mathrm{Var}(A_t)\leq 2LC_B\mathbb{E}[A_t]$. The local margin condition

gives $\mathbb{E}[A_t] \le \bar{\gamma} h_t^{1+\beta}$ for every t: if $h_t > b$, then $\bar{\gamma} h_t^{\beta} \ge 1$. Bernstein's inequality therefore gives, with probability at least $1 - \alpha/3$,

$$\sum_{t=1}^{T} A_t \le 2 \sum_{t=1}^{T} \mathbb{E}[A_t] + 2LC_B \log \frac{3}{\alpha} \le 2\bar{\gamma} U^{1+\beta} \sum_{t=1}^{T} t^{-(1+\beta)/2} + 2LC_B \log \frac{3}{\alpha}.$$

Intersecting the two events gives probability at least $1 - 2\alpha/3 \ge 1 - \alpha$. Finally,

$$\sum_{t=1}^{T} t^{-(1+\beta)/2} \le \begin{cases} \log(eT), & \beta = 1, \\ 2T^{(1-\beta)/2}/(1-\beta), & 0 \le \beta < 1, \end{cases}$$

and substituting the expression for U proves the result. □

D.2. Regret Analysis of Algorithm 3

D.2.1. Proof of Theorem 5

Proof of Theorem 5. For each arm $a \in [K]$, we denote by $n_{t,a}$ the number of times arm a has been pulled during exploration up to time t. Fix $T \ge 2$ and $0 < \alpha < 1$, and let $\mathcal{F}_{t-1}$ contain the history before observing X_t. At an exploration update for arm a, its current estimate is determined before observing the context, and the independent uniform arm choice leaves the context distributed as P_X. The conditional expectation calculation in the proof of Proposition 2 therefore gives

$$\mathbb{E}[g_{t,a}^{\top}(\hat{\theta}_{t-1,a} - \theta_a^{\star}) \mid \mathcal{F}_{t-1}, a_t = a] \ge \zeta \kappa_C \|\hat{\theta}_{t-1,a} - \theta_a^{\star}\|_2^2, \qquad t \in T_S.$$

The projected updates are counted separately for each arm, and all other rounds leave that arm's estimate unchanged. Since $r_{\varepsilon,d} \ge R \ge 2LC_B^2 \ge 2\zeta\kappa_C$, Lemma 7, with failure probability $\alpha/(2K)$ for each arm, and a union bound give, with probability at least $1 - \alpha/2$,

$$\|\hat{\theta}_{t,a} - \theta_a^{\star}\|_2^2 \le \min\left\{4, \frac{64 r_{\varepsilon,d}^2}{\zeta^2 \kappa_C^2 (n_{t,a}+1)} \log \frac{4KT}{\alpha}\right\}, \qquad 0 \le t \le T,\ a \in [K]. \tag{22}$$

By t_i there are at least i exploration rounds, with independent uniform arm choices. Thus $\mathbb{P}\{n_{t_i,a} < i/(2K)\} \le e^{-i/(8K)}$. A union bound over all arms and all $i \ge 8K \log(4KT/\alpha)$ with $t_i \le T$ shows that, with probability at least $1 - \alpha/2$,

$$n_{t_i,a} + 1 \ge \frac{i}{8K \log(4KT/\alpha)}, \qquad i \ge 1,\ t_i \le T,\ a \in [K].$$

For smaller i this inequality follows from $n_{t_i,a} + 1 \ge 1$. Intersect this event with (22).

On the other hand, since a_t is the greedy action for $t \notin T_S$, if we denote $a_t^\star = \operatorname{argmax}_{a\in[K]}\mu(X_t^\top\theta_a^\star)$, then

$$\mu(X_t^\top\theta_{a_t^\star}^\star) - \mu(X_t^\top\theta_{a_t}^\star) \le L\left(|X_t^\top(\theta_{a_t^\star}^\star - \hat\theta_{t-1,a_t^\star})| + |X_t^\top(\hat\theta_{t-1,a_t} - \theta_{a_t}^\star)|\right)$$
$$\le 2LC_B \max_a \|\hat\theta_{t-1,a} - \theta_a^\star\|_2 \lesssim \frac{LC_B r_{\varepsilon,d}}{\zeta\kappa_C}\sqrt{\frac{K}{i}}\log\frac{4KT}{\alpha}, \qquad t_i < t < t_{i+1}.$$

There are at most $\lceil\sqrt{i+1}\rceil - 1 \le \sqrt{2i}$ such greedy rounds after t_i. Each exploration round and the initial round has regret at most $2LC_B$. Moreover, $t_i = 1 + \sum_{j=1}^{i}\lceil\sqrt{j}\rceil \ge 1 + 2i^{3/2}/3$, so $|T_S \cap [T]| \lesssim T^{2/3}$. Summing gives, with probability at least $1-\alpha$,

$$\mathrm{Reg}(T) \lesssim LC_B\left(1 + \frac{r_{\varepsilon,d}\sqrt{K}}{\zeta\kappa_C}\log\frac{4KT}{\alpha}\right)T^{2/3}.$$

Take $\alpha = T^{-2}$ and use $r_{\varepsilon,d} \lesssim R\sqrt{d}/\varepsilon$ for $0 < \varepsilon \le 1$ and $r_{\varepsilon,d}/(\zeta\kappa_C) \ge 2$ to obtain the result. □

D.2.2. Proof of Theorem 6

Proof of Theorem 6. Similar to the single-parameter case, we can first show an improved estimation error result for $\hat\theta_{t,a}$ under the self-exploration condition in Assumption 5: Since the gradient estimator in (10) is unbiased, Assumptions 2 and 5 give

$$\mathbb{E}[g_{t,a}^\top(\hat\theta_{t-1,a} - \theta_a^\star) \mid \mathcal{F}_{t-1}] \ge \zeta\mathbb{E}[(X_t^\top(\hat\theta_{t-1,a} - \theta_a^\star))^2\mathbf{1}\{a_t = a\} \mid \mathcal{F}_{t-1}] \ge \zeta\kappa p_*\|\hat\theta_{t-1,a} - \theta_a^\star\|_2^2.$$

Since $\|g_{t,a}\|_2 = r_{\varepsilon,d} \ge 2\zeta\kappa p_*$, apply Lemma 7 with $\Theta = B_1^d$, $m = \zeta\kappa p_*$ and failure probability $\alpha/(2K)$ to each arm, and take a union bound. With probability at least $1-\alpha/2$,

$$\max_a \|\hat\theta_{t-1,a} - \theta_a^\star\|_2 \le \min\{2, U/\sqrt{t}\}, \quad \text{with } U := \frac{8r_{\varepsilon,d}}{\zeta\kappa p_*}\sqrt{\log\frac{4KT}{\alpha}}, \quad \forall 1 \le t \le T. \tag{23}$$

The greedy comparison used above now bounds each round's regret by $2LC_BU/\sqrt{t}$, so $\mathrm{Reg}(T) \le 4LC_BU\sqrt{T}$.

Under the margin condition, a decision with positive regret has regret at least $\triangle_t$. On (23), therefore,

$$\mathrm{Reg}(T) \le \sum_{t=1}^{T} A_t, \quad \text{with } A_t = \min\{2LC_B, 2LC_BU/\sqrt{t}\}\mathbf{1}\{\triangle_t \le \min\{2LC_B, 2LC_BU/\sqrt{t}\}\}.$$

These thresholds are deterministic, so the A_t are independent and belong to $[0, 2LC_B]$, without conditioning on the estimation event. The local margin condition extends to all positive thresholds with the existing constant $\bar\gamma = \max\{\gamma, b^{-\beta}\}$, since $\bar\gamma u^\beta \ge 1$ for $u > b$. Consequently,

1. $\mathbb{E}[\sum_t A_t] \le \bar{\gamma}(2LC_BU)^2 \log(eT)$ for $\beta = 1$.
2. $\mathbb{E}[\sum_t A_t] \le \dfrac{2\bar{\gamma}}{1-\beta}(2LC_BU)^{1+\beta}T^{(1-\beta)/2}$ for $0 \le \beta < 1$.

Applying Bernstein's bound as in the proof of Theorem 4, we then obtain with probability at least $1 - \alpha/2$,

$$\sum_{t=1}^{T} A_t \le (e-1)\sum_{t=1}^{T}\mathbb{E}[A_t] + 2LC_B\log(2/\alpha).$$

Intersecting the event on which this inequality holds with the estimation event gives the desired regret bound. □

D.2.3. Proof of Lemma 4 and Theorem 8

Proof of Lemma 4. Fix $T \ge 2$, $0 < \varepsilon \le 1$, and put $\alpha = T^{-2}$. Use U from (23), and choose

$$n_w = \left\lceil \frac{64C_B^2(\kappa p_*)^2U^2}{\kappa_C^2 h^2} \right\rceil, \qquad s_0 = \left(\frac{\kappa_C}{\kappa p_*}\right)^2 (n_w + 1).$$

Here s_0 is the offset in the step sizes below. This choice gives $n_w = O(\log(2KT))$ with the other parameters fixed, $s_0 \ge 1$, and $2U/\sqrt{s_0} \le h/(4C_B)$. During warm-up, pull each arm n_w times in a fixed order with $B_t = 1$ and the exploration step sizes in Theorem 5. For $T > Kn_w$, set $B_t = 0$ after warm-up and use (11) with

$$\eta_t = [\zeta\kappa p_*(s_0 + t - Kn_w)]^{-1}, \qquad t > Kn_w.$$

Since $n_w < T$, Lemma 7, applied to each arm's warm-up updates with failure probability $\alpha/(2K)$, gives with probability at least $1 - \alpha/2$,

$$s_0 \max_a \|\hat{\theta}_{Kn_w,a} - \theta_a^\star\|_2^2 \le U^2.$$

Condition on a warm-up history satisfying this bound, and let

$$\tau = \inf\left\{ t > Kn_w : \max_a \|\hat{\theta}_{t,a} - \theta_a^\star\|_2 > \frac{2U}{\sqrt{s_0 + t - Kn_w}} \right\}.$$

For $Kn_w < t \le \tau$, each estimation error at time $t-1$ is at most $h/(4C_B)$. On U_a, the estimated gap over every other arm is therefore greater than $h - 2C_B \cdot h/(4C_B) = h/2$, so greedy selects a. Assumption 7 and the preceding conditional expectation calculation then give

$$\mathbb{E}[g_{t,a}^\top(\hat{\theta}_{t-1,a} - \theta_a^\star) \mid \mathcal{F}_{t-1}] \ge \zeta\kappa p_*\|\hat{\theta}_{t-1,a} - \theta_a^\star\|_2^2 \quad \text{on } \{\tau \ge t\}.$$

Now, using the same telescoping and stopped Azuma–Hoeffding argument as in the proof of Lemma 7, for each $Kn_w < t \le T$, a union bound over arms gives

$$\mathbb{P}\{\tau = t \mid \mathcal{F}_{Kn_w}\} \le K \exp\left\{-\frac{\zeta^2 \kappa^2 p_*^2 U^2}{16 r_{\varepsilon,d}^2}\right\} \le \frac{\alpha}{4T}.$$

A union bound over these rounds, combined with the warm-up bound, gives with probability at least $1 - 3\alpha/4$,

$$\max_a \|\hat{\theta}_{t,a} - \theta_a^\star\|_2 \le \frac{2U}{\sqrt{s_0 + t - Kn_w}}, \qquad Kn_w \le t \le T.$$

Since $s_0 \ge 1$ and $\log(4KT/\alpha) \le 3\log(2KT)$, this proves the desired result. □

Finally, for Theorem 8, the estimation bound in the proof of Lemma 4 allows us to repeat the regret argument in the proof of Theorem 6 using failure probability $\alpha/4$ for the margin bound. Adding the warm-up regret, bounded by $2LC_BKn_w$, then gives the desired result.

D.3. Regret Analysis of Algorithm 9

D.3.1. Proof of Theorem 7 It suffices to establish the OLS estimation bound and apply the regret arguments above. Use the update index n and features x_n from (15), with $0 < \varepsilon \le 1$, the parameters in Proposition 5, and $\eta = \alpha \in (0, 1)$. Write

$$A_n = M_n + c_n I_d, \tag{24}$$

and define the Gram matrix, accumulated matrix noise, and its bound by

$$G_n = \sum_{j=1}^n x_j x_j^\top, \qquad E_n = \sum_{j=1}^n W_j, \qquad b_W(n) = 4\sigma_W \sqrt{n\left(d\log 9 + \log\frac{16T}{\alpha}\right)}.$$

For the confidence region in Appendix B, set

$$R_n = C_B^2 + \left[12\sigma_W + 2\sqrt{2}(\sigma_w + 2C_B)\right]\sqrt{n\left(d\log 9 + \log\frac{16T}{\alpha}\right)}. \tag{25}$$

In particular, $R_n \le R_1\sqrt{\max\{1, n\}}$. Let $\mathcal{F}_{j-1}$ contain the history before the context at the j-th update is observed. We have the following concentration results.

LEMMA 8. *Suppose* $\mathbb{E}[x_j x_j^\top \mid \mathcal{F}_{j-1}] \succeq \lambda I_d$ *for some* $\lambda > 0$. *With probability at least* $1 - \alpha/4$,

$$\lambda_{\min}(G_n) \ge \frac{n\lambda}{2} \qquad \textit{for all} \quad \frac{8C_B^2}{\lambda}\log\frac{4dT}{\alpha} \le n \le T.$$

Proof. Since $0 \preceq x_j x_j^\top \preceq C_B^2 I_d$, the matrix Chernoff bound for adapted matrices (see, e.g., (Tropp 2011)) gives $\mathbb{P}\{\lambda_{\min}(G_n) \le n\lambda/2\} \le d\exp[-n\lambda/(8C_B^2)]$. A union bound over $n \le T$ gives the claim. □

LEMMA 9. *With probability at least* $1 - 3\alpha/8$*, simultaneously for* $0 \le n \le T$*,*

$$\begin{aligned} \|E_n\|_{\mathrm{op}} &\le b_W(n), \\ \left\|\sum_{j=1}^{n} w_j\right\|_2 &\le 2\sqrt{2}\sigma_w\sqrt{n\left(d\log 9 + \log\frac{16T}{\alpha}\right)}, \\ \left\|\sum_{j=1}^{n} x_j \epsilon_{\tau_j}\right\|_2 &\le 4\sqrt{2}C_B\sqrt{n\left(d\log 9 + \log\frac{16T}{\alpha}\right)}. \end{aligned} \tag{26}$$

In particular, under the above inequalities,

$$c_n = C_B^2 + 2b_W(n), \qquad A_n \succeq G_n + [C_B^2 + b_W(n)]I_d, \qquad \|A_n(\widetilde{\theta}_n - \theta^\star)\|_2 \le R_n.$$

Proof. The operator norm concentration result for symmetric Gaussian matrices (see, e.g., Corollary 4.4.8 in Vershynin (2018)) gives the first bound for each n with failure probability $\alpha/(8T)$. For the second, apply the Gaussian tail bound to unit projections of $\sum_j w_j$, which have variance $n\sigma_w^2$. For the third, apply the Azuma–Hoeffding inequality to $v^\top x_j \epsilon_{\tau_j}$, which has conditional mean zero and lies in $[-2C_B, 2C_B]$ for every unit v, and then apply a standard $1/4$-net argument. A union bound over n proves (26).

Finally, to show the last three inequalities, noting that $M_n = G_n + E_n \succeq -b_W(n)I_d$, we obtain the first two inequalities directly. And the last inequality then follows from the decomposition

$$A_n(\widetilde{\theta}_n - \theta^\star) = \sum_{j=1}^{n} x_j \epsilon_{\tau_j} + \sum_{j=1}^{n} w_j - (E_n + c_n I_d)\theta^\star.$$

□

In particular, this gives the following estimation error guarantee.

PROPOSITION 8. *Under the condition of Lemma 8, with probability at least* $1 - \alpha$*, simultaneously for* $1 \le n \le T$*,*

$$\|\widetilde{\theta}_n - \theta^\star\|_2 \le \frac{2R_n}{n\lambda} \le \frac{2R_1}{\lambda\sqrt{n}}.$$

Proof. Intersect the events in Lemmas 8 and 9, whose total failure probability is at most $\alpha/4 + 3\alpha/8 < \alpha$. Above the threshold in Lemma 8, $A_n \succeq (C_B^2 + n\lambda/2)I_d$. Below it, the same bound follows from Lemma 9 and

$$\frac{n\lambda}{2} \le \sqrt{2C_B^2 \lambda n \log \frac{4dT}{\alpha}} \le b_W(n),$$

because $\lambda \le C_B^2$ and Proposition 5 gives $\sigma_W \ge C_B^2$ for $0 < \varepsilon \le 1$. Dividing the residual bound by $\lambda_{\min}(A_n)$ proves the claim. □

To finish the proof of Theorem 7, we consider the epoch-greedy case: under uniform exploration, $\mathbb{E}[x_j x_j^\top \mid \mathcal{F}_{j-1}] = \Phi_X \succeq \kappa_\Phi I_d$. Proposition 8 with $\lambda = \kappa_\Phi$ therefore gives $\|\widetilde{\theta}_n - \theta^\star\|_2 \le \frac{2R_1}{\kappa_\Phi \sqrt{n}}$. Bringing this bound to the proof of Theorem 1 and taking $\alpha = T^{-2}$ for $T \ge 2$ gives a regret bound similar to that in Theorem 1. The purely greedy case and the margin bound case follow from the similar estimation results by replacing κ_Φ with $\kappa_l p_*/d$ under Assumption 3 and applying the proofs of Theorems 2 and 4.

D.4. Regret Analysis of Algorithm 11

D.4.1. Proof of Theorem 9

Proof. Using the same notation and parameters as in Appendix D.3, define $\mathcal{G}_n = \{\|E_n\|_{\mathrm{op}} \le b_W(n)\}$,

$$H_n = G_n + [C_B^2 + 2b_W(n)]I_d, \quad L_n = \mathrm{tr}(H_n^{-1}), \quad q_n = x_{n+1}^\top H_n^{-2} x_{n+1},$$
$$n_0 = \lceil 72 C_B^2 d^2 \log(4T/\alpha)/(\kappa_l p_*) \rceil.$$

We first show that $L_n = O(1/n)$ uniformly for $n_0 \le n \le T$ with high probability, then combine this fact and the OLS estimation error bound to control regret. Under $\mathcal{G}_n$, $M_n \succeq -b_W(n)I_d$ and (24) give $A_n = H_n + E_n$. Since $\|H_n^{-1}E_n\|_{\mathrm{op}} \le 1/2$, we have

$$\frac{2}{3}\|H_n^{-1}x\|_2 \le \|A_n^{-1}x\|_2 \le 2\|H_n^{-1}x\|_2. \tag{27}$$

For the feature $\tilde{x}_{n+1}$ corresponding to the greedy action at the $n+1$-th round, the UCB rule gives $\|A_n^{-1}x_{n+1}\|_2 \ge \|A_n^{-1}\tilde{x}_{n+1}\|_2$. Thus (27) and Assumption 3 imply

$$\mathbb{E}[q_n \mid \mathcal{F}_n] \ge \frac{\kappa_l p_*}{9d}\mathrm{tr}(H_n^{-2}) \ge \frac{\kappa_l p_*}{9d^2} L_n^2$$

over $\mathcal{G}_n$. As $b_W(n)$ is nondecreasing and $H_n \succeq C_B^2 I_d$, we have

$$L_n - L_{n+1} \ge \frac{q_n}{2}, \qquad \frac{1}{L_{n+1}} - \frac{1}{L_n} \ge \frac{q_n}{2L_n^2}. \tag{28}$$

Define $Z_n = \frac{q_n}{L_n^2}\mathbf{1}_{\mathcal{G}_n} + \frac{\kappa_l p_*}{9d^2}\mathbf{1}_{\mathcal{G}_n^c}$. Since $q_n \le C_B^2 L_n^2$ and $\mathcal{G}_n$ is $\mathcal{F}_n$-measurable, $0 \le Z_n \le C_B^2$ and $\mathbb{E}[Z_n \mid \mathcal{F}_n] \ge \kappa_l p_*/(9d^2)$. The conditional Chernoff bound (Tropp 2011) gives, for $n \ge n_0$,

$$\mathbb{P}\left\{\sum_{j=0}^{n-1} Z_j \le \frac{n\kappa_l p_*}{18d^2}\right\} \le \exp\left(-\frac{n\kappa_l p_*}{72d^2 C_B^2}\right) \le \frac{\alpha}{4T}.$$

Taking a union bound over $n \le T$, and taking intersection with the events in Lemma 9, we have with probability at least $1-\alpha$, $Z_j = q_j/L_j^2$. Summing (28) then yields

$$\frac{1}{L_n} \ge \frac{1}{2}\sum_{j=0}^{n-1}\frac{q_j}{L_j^2} \ge \frac{n\kappa_l p_*}{36d^2}, \qquad n_0 \le n \le T.$$

This gives the desired bound for L_n. It remains to control the regret. Under the same event, combining Lemma 9 and (27) gives

$$\max_{a\in[K]} \phi(X_{n+1}, a)^\top \theta^\star - x_{n+1}^\top \theta^\star \le 2R_n \|A_n^{-1} x_{n+1}\|_2 \le 4R_n\sqrt{q_n}.$$

For $T > n_0$, (28) and the bound on L_n imply

$$\sum_{n=n_0}^{T-1} n q_n \le 2n_0 L_{n_0} + 2\sum_{n=n_0+1}^{T-1} L_n \lesssim \frac{d^2\log(eT)}{\kappa_l p_*}.$$

Using $R_n \le R_1\sqrt{n}$ for $n \ge 1$, Cauchy–Schwarz and the $2C_B$ regret bound for the first $\min\{T, n_0\}$ rounds give

$$\mathrm{Reg}(T) \lesssim C_B \min\{T, n_0\} + R_1 d\sqrt{\frac{T\log(eT)}{\kappa_l p_*}} \lesssim R_1 d\sqrt{\frac{T\log(eT)}{\kappa_l p_*}}.$$

For $T \le n_0$, the first term alone bounds the regret. The last inequality uses $\min\{T, n_0\} \le \sqrt{Tn_0}$, $\kappa_l p_* \le C_B^2$, and $R_1 \ge C_B^2\sqrt{\log(4T/\alpha)}$. Taking $\alpha = T^{-2}$ and substituting (25) gives the claimed $\widetilde{O}(d^{3/2}\sqrt{T}/(\varepsilon\sqrt{\kappa_l p_*}))$ rate. Finally, $A_n \succeq H_n/2 \succeq (2L_n)^{-1} I_d$ on the same event, so for $n \ge n_0$,

$$\|\widetilde{\theta}_n - \theta^\star\|_2^2 \le 4L_n^2 R_n^2 = \widetilde{O}(1/n), \qquad \|\widetilde{\theta}_n - \theta^\star\|_{A_n}^2 \le 2L_n R_n^2 = \widetilde{O}(1).$$

Together with $\rho_n = R_n/\sqrt{n+1} \le R_1$, this gives the desired result. □

D.5. Regret Analysis of Algorithm 7

To prove Theorem 10, first define

$$\Phi_m = \mathbb{E}[\phi(X_t, a_t)\phi(X_t, a_t)^\top \mid \mathcal{F}_{(m-1)W}], \qquad (m-1)W < t \le mW.$$

Conditioning on the history at each epoch's start, Hoeffding's inequality and the Gaussian operator norm bound used in the proof of Lemma 9, together with a $1/4$-net argument and a union bound over epochs, give, with probability at least $1 - 1/(4T^2)$,

$$\|\bar{V}_m - \Phi_m\|_{\mathrm{op}} \le C_B^2\sqrt{\frac{2v_T}{W}} + 4\sigma_W\sqrt{\frac{v_T}{W}} \le \frac{\kappa_0}{4d}$$

simultaneously for every tested epoch, by the choice of W and v_T. On this event, a passed test implies $\Phi_m \succeq \kappa_0 I_d/(4d)$. Conversely, under Assumption 3 with $\kappa_l p_* \ge \kappa_0$, we have $\Phi_m \succeq (\kappa_l p_*/d) I_d$, so all tests pass under this condition.

Proof of Theorem 10. Based on the above covariance concentration result, we first bound the parameter estimation error, then control the regret before and after switching.

Control the Parameter Estimation Error. Using the error notation e_m from Lemma 7, for each complete epoch before switching set $\xi_m = \bar{g}_m - \mathbb{E}[\bar{g}_m \mid \mathcal{F}_{(m-1)W}]$. The same calculation as in the proof of Proposition 2 gives, whenever $\Phi_m \succeq \kappa_0 I_d/(4d)$,

$$e_{m-1}^\top(\bar{g}_m - \xi_m) \ge \frac{\zeta\kappa_0}{4d}\|e_{m-1}\|_2^2, \quad \|\bar{g}_m - \xi_m\|_2^2 \le LC_B^2 e_{m-1}^\top(\bar{g}_m - \xi_m).$$

Since $\|g_t\|_2 = r_{\varepsilon/2,d}$, conditional Hoeffding bounds give

$$\mathbb{E}[e^{su^\top\xi_m} \mid \mathcal{F}_{(m-1)W}] \le \exp\left(\frac{s^2 r_{\varepsilon/2,d}^2\|u\|_2^2}{2W}\right)$$

for every $\mathcal{F}_{(m-1)W}$-measurable u and $s \in \mathbb{R}$. The same net argument gives $\|\xi_m\|_2^2 \le 8r_{\varepsilon/2,d}^2 v_T/W$ uniformly over these epochs, with probability $1 - 1/(8T^2)$.

For updates satisfying the covariance bound and $m \ge 8LC_B^2 d/(\zeta\kappa_0)$, we have $\eta_m LC_B^2 \le 1/2$, and therefore

$$\|e_m\|_2^2 \le \frac{m}{m+1}\|e_{m-1}\|_2^2 - 2\eta_m e_{m-1}^\top\xi_m + 2\eta_m^2\|\xi_m\|_2^2.$$

Apply the stopped Azuma–Hoeffding argument of Lemma 7 to this recursion using the refined $1/W$ bound of ξ. With projection controlling the earlier epochs and $\sum_{i=1}^m i^{-1} \le \log(eT)$, a union bound over epochs gives, with probability at least $1 - 1/(4T^2)$,

$$\|e_m\|_2^2 \lesssim \frac{1}{m+1}\left[1 + \frac{LC_B^2 d}{\zeta\kappa_0} + \frac{r_{\varepsilon/2,d}^2 d^2 v_T\log(eT)}{\zeta^2\kappa_0^2 W}\right]$$

simultaneously for all updates m before switching for which the covariance bound has held throughout.

Control the Regret before Switching. On the covariance event established above, this controls the estimate used in every played epoch, including the failed and final incomplete epochs, since all preceding updates passed their tests. By (21), the first epoch costs at most $2LC_B \min\{W,T\}$; summing the remaining bounds uses $W\sum_{m=2}^{M} m^{-1/2} \le 2\sqrt{WT}$, with an empty sum if $W \ge T$. Substituting W and $r_{\varepsilon/2,d} \lesssim R\sqrt{d}/\varepsilon$, and using $\min\{W,T\} \le \sqrt{WT}$, gives regret before switching $\widetilde{O}(d^2\sqrt{T}/\varepsilon)$ with the other model parameters fixed.

Control the Regret after Switching. After a switch, condition on its history and apply the uniform bound in the proof of Theorem 1 with horizon bound T and failure probability $1/(4T^2)$. The restarted algorithm has the same model conditions, an initializer in B_1^d, and privacy budget ε, so its additional regret is

$$\widetilde{O}\left(\frac{LC_B R\sqrt{d}}{\zeta\kappa_\Phi\varepsilon}T^{2/3}\right).$$

The total failure probability is at most $3/(4T^2)$, proving the general claim. Under diversity there is no switch, leaving only the first bound, which implies the stated bound since $\varepsilon \le 1$. □

Appendix E: Proof of Lower Bounds

E.1. Proof of Theorem 3

In this section, we give a proof of Theorem 3 by combining the argument in Han et al. (2020) and the divergence contraction inequality in Duchi et al. (2018).

LEMMA 10 **(Lemma 8 in Han et al. (2020))**. *Let $x_{t,1} - x_{t,2} = R_t u_t$ with $R_t > 0, \|u_t\|_2 = 1$. For $0 < \Delta \le 1$, let Q_0 be uniform on ΔS_1^{d-1}, and define*

$$\frac{dQ_1}{dQ_0}(\theta) = \frac{((x_{t,1}-x_{t,2})^\top\theta)_+}{Z_0}, \qquad \frac{dQ_2}{dQ_0}(\theta) = \frac{((x_{t,2}-x_{t,1})^\top\theta)_+}{Z_0},$$

where $Z_0 = \mathbb{E}_{Q_0}[((x_{t,1}-x_{t,2})^\top\theta)_+] = \mathbb{E}_{Q_0}[((x_{t,2}-x_{t,1})^\top\theta)_+]$ and $(x)_+ = \max\{x,0\}$. Then $\theta \sim Q_1$ if and only if $\theta - 2\left(u_t^\top\theta\right)u_t \sim Q_2$.

Moreover, we have

$$Z_0 = R_t\Delta \cdot \begin{cases} \frac{2^d}{\pi d}\binom{d}{d/2}^{-1}, & \text{if } d \text{ is even} \\ \frac{1}{2^d}\binom{d-1}{(d-1)/2}, & \text{if } d \text{ is odd} \end{cases} \ge \frac{R_t\Delta}{5\sqrt{d}},$$

and

$$\mathbb{E}_{Q_1}\left(u_t^\top\theta\right)^2 = \mathbb{E}_{Q_2}\left(u_t^\top\theta\right)^2 = \frac{2\Delta^2}{d+1}.$$

Proof of Theorem 3. Consider the two-arm stochastic contextual bandit environment: for each d-dimensional context $i = 1$ or 2, $x_{t,i} \sim \mathcal{N}(0, \frac{1}{d} I_d)$ independently. If action a_t is chosen at time t, the reward r_t is generated via $r_t = x_{t,a_t}^T \theta + \epsilon_t$ with $\epsilon_t \sim_{i.i.d.} \mathcal{N}(0,1)$. Given any fixed ε-LDP bandit algorithm π with $\varepsilon \le 1$, we denote its decision at t-th step by a_t. Besides the current X_t, this decision can use only the private observations

$$S_t = (M_1(X_1, a_1, r_1), \ldots, M_{t-1}(X_{t-1}, a_{t-1}, r_{t-1})),$$

where each ε-LDP mechanism and each action rule may depend on earlier private observations. We denote the distribution of S_t by Q_θ^t, and we have

$$\begin{aligned} &\mathbb{E}_{\theta\sim Q_0}[\mathbb{E}_{Q_\theta^t}[(x_{t,a_t^*} - x_{t,a_t})^T \theta | x_{t,1}, x_{t,2}]] \\ &= \mathbb{E}_{\theta\sim Q_0}[((x_{t,1} - x_{t,2})^T \theta)_+ Q_\theta^t(a_t(S_t, X_t) = 2) + ((x_{t,2} - x_{t,1})^T \theta)_+ Q_\theta^t(a_t(S_t, X_t) = 1)]. \end{aligned} \tag{29}$$

Here Q_0, Q_1, Q_2, Z_0 are as in Lemma 10, with $\Delta > 0$ to be determined later. Denote $R_t = \|x_{t,1} - x_{t,2}\|$, $u_t = R_t^{-1}(x_{t,1} - x_{t,2})$. Then the right hand side of (29) is lower bounded by

$$\begin{aligned} &Z_0(Q_1 \circ Q_\theta^t(a_t(S_t, X_t) = 2) + Q_2 \circ Q_\theta^t(a_t(S_t, X_t) = 1)) \\ &\ge_{(a)} Z_0(1 - \mathrm{TV}(Q_1 \circ Q_\theta^t, Q_2 \circ Q_\theta^t)) \\ &\ge_{(b)} \frac{Z_0}{2} \exp\{-D_{KL}(Q_1 \circ Q_\theta^t \| Q_2 \circ Q_\theta^t)\} \\ &=_{(c)} \frac{Z_0}{2} \exp\{-D_{KL}(Q_1 \circ Q_\theta^t \| Q_1 \circ Q_{\theta - 2(u_t^T\theta)u_t}^t)\} \\ &\ge_{(d)} \frac{Z_0}{2} \exp\{-\mathbb{E}_{Q_1}[D_{KL}(Q_\theta^t \| Q_{\theta - 2(u_t^T\theta)u_t}^t)]\}, \end{aligned} \tag{30}$$

where $D_{KL}(\cdot\|\cdot)$ denotes the KL-divergence, $TV(\cdot,\cdot)$ denotes the total variation distance and $Q_i \circ Q_\theta^t$ means $\mathbb{E}_{\theta\sim Q_i}[Q_\theta^t]$. The (a) inequality comes from the fundamental limit of two-point testing (see, e.g., Section 15.2 in Wainwright (2019)), and the (b) inequality comes from Lemma 2.6 of Tsybakov (2008), the (c) equality comes from Lemma 10 and the (d) inequality comes from the joint convexity of KL-divergence.

Given the same private history $\mathcal{H}_s$ before round s, let $P_\theta^s(\cdot \mid h)$ denote the distribution of (X_s, a_s, r_s). The conditional action rule is the same for every parameter. We have

$$D_{KL}(P_\theta^s(\cdot \mid h) \| P_{\theta'}^s(\cdot \mid h)) = \frac{1}{2}\mathbb{E}[(x_{s,a_s}^\top(\theta - \theta'))^2 | \mathcal{H}_s] \le \frac{1}{2}\sum_{a=1}^{2} \mathbb{E}[(x_{s,a}^\top(\theta - \theta'))^2] = \frac{\|\theta - \theta'\|_2^2}{d}.$$

Applying the chain rule and the divergence contraction inequality in Theorem 1 of Duchi et al. (2018) then gives

$$D_{KL}(Q_\theta^t \| Q_{\theta'}^t) \le \frac{2(e^\varepsilon - 1)^2 (t-1)}{d} \|\theta - \theta'\|_2^2.$$

Applying the formula for $\mathbb{E}_{Q_1}[(u_t^\top \theta)^2]$ in Lemma 10 gives

$$\mathbb{E}_{Q_1} D_{KL}(Q_\theta^t \| Q_{\theta - 2(u_t^T\theta)u_t}^t) \le \frac{16(e^\varepsilon - 1)^2 (t-1)\Delta^2}{d(d+1)} \le \frac{1}{4},$$

with the selection $\Delta = \min\left\{1, \frac{d}{8(e^\varepsilon - 1)\sqrt{T}}\right\}$.

Thus (30) is at least $e^{-1/4} Z_0/2$. For every $\theta \in \Delta S_1^{d-1}$, the Gaussian contexts give $(x_{t,1} - x_{t,2})^\top \theta \sim \mathcal{N}(0, 2\Delta^2/d)$, hence $\mathbb{E}_X Z_0 = \Delta/\sqrt{\pi d}$. Averaging over the current contexts and summing under the same prior Q_0 yields

$$\sup_{\|\theta\|_2 \le 1} \mathbb{E}_\theta \operatorname{Reg}_\pi(T;\theta) \ge \frac{e^{-1/4} T\Delta}{2\sqrt{\pi d}} \gtrsim \min\left\{\frac{T}{\sqrt{d}}, \frac{\sqrt{dT}}{e^\varepsilon - 1}\right\}.$$

Since $e^\varepsilon - 1 \le (e-1)\varepsilon$ for $0 < \varepsilon \le 1$ and $T \ge d^2/\varepsilon^2$, this proves the regret lower bound. □

E.2. Proof of Proposition 6

Proof of Proposition 6. Since our focus is the rate with respect to t here, we use the noise variance and c_t as in Proposition 5, with $C_B = 1$. Consider the case that $\theta^\star = e_2 := \begin{pmatrix} 0 & 1 \end{pmatrix}^\top$ and $\phi(X_t, a_t) \equiv e_1 := \begin{pmatrix} 1 & 0 \end{pmatrix}^\top$, and for simplicity we can assume no observation noise is added, i.e., $r_t = \phi(X_t, a_t)^\top \theta^\star \equiv 0$, then with

$$M_t = t e_1 e_1^\top + \sum_{i=1}^t W_i, \quad A_t = M_t + c_t I, \quad \tilde{\theta}_t = A_t^{-1} \sum_{i=1}^t w_i,$$

it holds that

$$\begin{aligned} \|\tilde{\theta}_t - \theta^\star\|_{A_t}^2 &= \big(A_t^{-1} \sum_{i=1}^t w_i - e_2\big)^\top A_t \big(A_t^{-1} \sum_{i=1}^t w_i - e_2\big) \\ &= \underbrace{(\sum_{i=1}^t w_i)^\top A_t^{-1} (\sum_{i=1}^t w_i)}_{\ge 0} + e_2^\top A_t e_2 - 2 e_2^\top (\sum_{i=1}^t w_i). \end{aligned}$$

Since $c_t \ge 1 - \lambda_{\min}(M_t)$, we have $A_t \succeq I$, so the first term is non-negative. For the last two terms, we have

$$e_2^\top A_t e_2 = \sum_{i=1}^t (W_i)_{22} + c_t, \qquad \sum_{i=1}^t (W_i)_{22} \sim \mathcal{N}(0, \sigma_W^2 t),$$

and $2e_2^\top(\sum_{i=1}^t w_i) = 2\sum_{i=1}^t w_{i,2} \sim \mathcal{N}(0, 4\sigma_w^2 t)$, $e_2^\top A_t e_2 - c_t - 2e_2^\top\Big(\sum_{i=1}^t w_i\Big) \sim \mathcal{N}\big(0, (\sigma_W^2 + 4\sigma_w^2)t\big)$. That implies

$$\mathbb{P}\left\{e_2^\top A_t e_2 - 2e_2^\top\Big(\sum_{i=1}^t w_i\Big) \geq \sqrt{(\sigma_W^2 + 4\sigma_w^2)t}\right\} \geq 0.15,$$

because $c_t > 0$ and for a standard normal variable $Z \sim \mathcal{N}(0,1)$, we can verify $\mathbb{P}(Z > 1) \geq 0.15$ for the standard Gaussian distribution CDF. On this event, $\|\tilde{\theta}_t - \theta^\star\|_{A_t} \geq (\sigma_W^2 + 4\sigma_w^2)^{1/4} t^{1/4}$. Thus, at $t = T$, any $\rho_T \asymp T^\nu$ with $\nu < 1/4$ is smaller than this error for all sufficiently large T. This gives the desired claim. □

References

Duchi JC, Jordan MI, Wainwright MJ (2018) Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association* 113(521):182–201.

Dwork C, Kenthapadi K, McSherry F, Mironov I, Naor M (2006a) Our data, ourselves: Privacy via distributed noise generation. *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, 486–503 (Springer).

Dwork C, McSherry F, Nissim K, Smith A (2006b) Calibrating noise to sensitivity in private data analysis. *Theory of cryptography conference*, 265–284 (Springer).

Dwork C, Roth A (2014) The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* 9(3-4):211–407, URL `http://dx.doi.org/10.1561/0400000042`.

Han Y, Zhou Z, Zhou Z, Blanchet J, Glynn PW, Ye Y (2020) Sequential batch learning in finite-action linear contextual bandits. *arXiv:2004.06321* .

Tropp JA (2011) User-friendly tail bounds for matrix martingales. Technical report, CALIFORNIA INST OF TECH PASADENA.

Tsybakov AB (2008) *Introduction to nonparametric estimation* (Springer Science & Business Media).

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

Wainwright MJ (2019) *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48 (Cambridge University Press).

Zheng K, Cai T, Huang W, Li Z, Wang L (2020) Locally differentially private (contextual) bandits learning. *Advances in Neural Information Processing Systems* 33:12300–12310.