Supplementary Material: Stochastic Linear Contextual Bandits with Diverse Contexts

Weiqiang Wu
London Stock Exchange

Jing Yang
The Pennsylvania State University

Cong Shen
University of Virginia

A Proof of Proposition 1

For the constrained convex optimization problem in (6), the corresponding Lagrangian can be formulated as

$$\mathcal{L}(\boldsymbol{\beta}, \lambda) = \boldsymbol{\beta}^{\intercal}\mathbf{N}_t^{-1}(a)\boldsymbol{\beta} + (\mathbf{x}(a, c_t) - \mathbf{X}_t(a)\boldsymbol{\beta})^{\intercal}\boldsymbol{\lambda}, \tag{10}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^d$ is the Lagrangian multiplier vector.

Taking derivative with respect to $\boldsymbol{\beta}$, we have

$$\mathbf{N}_t^{-1}(a)\boldsymbol{\beta} - \mathbf{X}_t^{\intercal}(a)\boldsymbol{\lambda} = 0, \tag{11}$$

i.e.,

$$\boldsymbol{\beta} = \mathbf{N}_t(a)\mathbf{X}_t^{\intercal}(a)\boldsymbol{\lambda}. \tag{12}$$

Then, to satisfy the first constraint in (6), we have

$$\mathbf{x}(a, c_t) = \mathbf{X}_t(a)\boldsymbol{\beta} = \mathbf{X}_t(a)\mathbf{N}_t(a)\mathbf{X}_t^{\intercal}(a)\boldsymbol{\lambda}, \tag{13}$$

which implies that

$$\boldsymbol{\lambda} = \left[\mathbf{X}_t(a)\mathbf{N}_t(a)\mathbf{X}_t^{\intercal}(a)\right]^{-1}\mathbf{x}(a, c_t) := \mathbf{V}_t^{-1}(a)\mathbf{x}(a, c_t). \tag{14}$$

Note that the definitions of $\mathbf{X}_t(a)$ and $\mathbf{N}_t(a)$ ensure that $\mathbf{V}_t(a)$ is positive definite and invertible for every t.

Plugging (14) into (12), we have

$$\boldsymbol{\beta} = \mathbf{N}_t(a)\mathbf{X}_t^{\intercal}(a)\mathbf{V}_t^{-1}(a)\mathbf{x}(a, c_t),$$

which is the unique optimal solution to the optimization problem in (6).

B One the Relationship between LinUCB-d and LinUCB

Major difference. One major difference between LinUCB-d and LinUCB (Li et al., 2010) is as follows: Under LinUCB, at each time t, it will first estimate the true parameter of arm a (i.e., $\boldsymbol{\theta}_a$) by solving a ridge regression and then use it to derive the UCB for the expected reward. The criterion to select the estimate is to minimize the penalized mean squared error in fitting the past observations; On the other hand, under LinUCB-d, the learner will directly estimate the expected reward through a linear combination of the rewards obtained when arm a was pulled under all contexts. The criterion of selecting the estimate is to minimize the uncertainty (or "variance") of the estimation. It avoids the intermediate step of trying to estimate $\boldsymbol{\theta}_a$ first in LinUCB.

Essential equivalence. Although linUCB-d and linUCB view the problem from different angles, they actually produce the same estimate on the expected reward and confidence bound at every time t under the same realizations of context arrivals and rewards, as shown below.

Based on (12), (13) and (14), the estimated mean reward $\hat{r}_t(a)$ in Algorithm 1 can be alternatively expressed as

$$\hat{r}_t(a) = \mathbf{s}_t^{\intercal}(a)\mathbf{X}_t^{\intercal}(a)\boldsymbol{\lambda} = \mathbf{s}_t^{\intercal}(a)\mathbf{X}_t^{\intercal}(a)\mathbf{V}_t^{-1}(a)\mathbf{x}(a, c_t) := \hat{\boldsymbol{\theta}}_t^{\intercal}(a)\mathbf{x}(a, c_t),$$

where $\hat{\boldsymbol{\theta}}_t(a) := \mathbf{V}_t^{-1}(a)\mathbf{X}_t(a)\mathbf{s}_t(a)$. We can verify that this is exactly the estimate of $\boldsymbol{\theta}(a)$ obtained by applying the ridge regression with penalty factor l^2 to the historical data $\{(\mathbf{x}(a_\tau, c_\tau), y_\tau)\}_{\tau=1}^{t-1}$.

Besides, for the $\hat{\sigma}_t(a)$ in Algorithm 1, we have

$$\hat{\sigma}_t(a) = \sqrt{\boldsymbol{\lambda}^\intercal \mathbf{X}_t(a)\mathbf{N}_t(a)\mathbf{X}_t^\intercal(a)\boldsymbol{\lambda}} = \|\mathbf{x}(a, c_t)\|_{\mathbf{V}_t^{-1}(a)}, \tag{15}$$

where we follow the convention to denote $\mathbf{x}\mathbf{V}\mathbf{x}^\intercal$ as $\|\mathbf{x}\|_{\mathbf{V}}^2$.

Thus, if we let $l = 1$, both $\hat{r}_t(a)$ and $\hat{\sigma}_t(a)$ share the same form as the corresponding quantities in LinUCB. As a *reformulation* of LinUCB, LinUCB-d automatically inherits all properties of LinUCB.

Computation and analytical issues. Computationally LinUCB-d is the same as LinUCB if we first compute the Lagrangian multiplier in (14) through $\mathbf{V}_t(a)$, which can be equivalently computed by summing $\mathbf{x}(a, c_t)\mathbf{x}^\intercal(a, c_t)$ over the time slots when a is pulled. The advantage of LinUCB-d as an alternative form of LinUCB is on the analytical side. The prediction uncertainty minimization nature shown in Proposition 1 gives us a unique angle to elucidate the impact of context diversity on the corresponding learning regret, as elaborated in Lemma 3, Lemma 4, Lemma 9 and Lemma 10.

C Proof of Theorem 1

In the following, we will derive regret bounds for those error events individually, and then assemble them together to obtain the regret bound in Theorem 1.

C.1 Bound the Regret over $\mathcal{A}_T$

First, based on Hoeffding's inequality, and the independent and uniform arrival of contexts assumption, we have

$$\mathbb{P}\left[N_{F_k}(c) \le \frac{1}{2n} \cdot 2^{k-1}\right] \le \exp\left(-\frac{2^{k-1}}{2n^2}\right). \tag{16}$$

Denote $R(\mathcal{A}_T)$ as the regret incurred over $\mathcal{A}_T$, and M as the maximum per-step regret. Then,

$$\begin{aligned}
\mathbb{E}[R(\mathcal{A}_T)] &\le M \sum_{k=2}^{\lceil \log_2 T \rceil} \sum_{t \in F_k} \mathbb{E}[\mathbf{1}\{t \in \mathcal{A}_T\}] \le M \sum_{k=1}^{\lfloor \log_2 T \rfloor} \sum_{t \in F_{k+1}} \sum_{a, c \in \bar{\mathcal{C}}_a} \mathbb{P}\left[N_{F_k}(c) \le \frac{2^{k-1}}{2n}\right] \\
&\le MKd \sum_{k=1}^{\lfloor \log_2 T \rfloor} \sum_{t \in F_{k+1}} \exp\left(-\frac{2^{k-1}}{2n^2}\right) \le Mn \sum_{t=2}^{\infty} \exp\left(-\frac{t}{8n^2}\right) &(17)\\
&\le MKd \int_0^\infty \exp\left(-\frac{t}{8n^2}\right) \mathrm{d}t = 8MKdn^2, &(18)
\end{aligned}$$

where (17) follows from (16).

C.2 Bound the Regret over $\mathcal{B}_T$

First, we define $\tilde{\mathbf{s}}_t(a)$ and $\bar{\mathbf{s}}_t(a)$ as follows:

$$\tilde{\mathbf{s}}_t(a) = [S_t(a, 1) - N_t(a, 1)r(a, 1), \ldots, S_t(a, n_t) - N_t(a, n_t)r(a, n_t), \mathbf{0}_d]^\intercal \tag{19}$$

$$\bar{\mathbf{s}}_t(a) = [\mathbf{0}_{n_t}, -l\mathbf{e}_1^\intercal \boldsymbol{\theta}(a), \ldots, -l\mathbf{e}_d^\intercal \boldsymbol{\theta}(a)]^\intercal. \tag{20}$$

Intuitively, $\tilde{\mathbf{s}}_t(a)$ corresponds to the accumulated noise in the observations when arm a is pulled under different contexts up to time t, and $\bar{\mathbf{s}}_t(a)$ corresponds to the bias contributed by the feature vectors associated with the dummy contexts, which were added to ensure the existence of the unique solution in (6) for every t.

Then, the reward estimation error can be expressed as

$$\hat{r}_t(a) - r(a, c_t) = \mathbf{s}_t^\intercal(a)\mathbf{N}_t^{-1}(a)\boldsymbol{\beta}_t(a) - \boldsymbol{\theta}^\intercal(a)\mathbf{x}(a, c_t)$$

$$= \mathbf{s}_t^\intercal(a)\mathbf{N}_t^{-1}(a)\boldsymbol{\beta}_t(a) - \boldsymbol{\theta}^\intercal(a)\mathbf{X}_t(a)\boldsymbol{\beta}_t(a) \tag{21}$$
$$= \left(\mathbf{s}_t(a) - \mathbf{N}_t(a)\mathbf{X}^\intercal(a)\boldsymbol{\theta}(a)\right)^\intercal \mathbf{N}_t^{-1}(a)\boldsymbol{\beta}_t(a)$$
$$:= (\tilde{\mathbf{s}}_t(a) + \bar{\mathbf{s}}_t(a))^\intercal \mathbf{X}_t^\intercal(a)\boldsymbol{\lambda}, \tag{22}$$

where (21) is due to the fact that $\mathbf{x}(a, c_t) = \mathbf{X}_t(a)\boldsymbol{\beta}_t(a)$ according to Proposition 1, and the $\boldsymbol{\lambda}$ in (22) is the Lagrangian multiplier involved in the proof of Proposition 1 in Appendix A and satisfies (11). In the following, we will bound the contribution from $\bar{\mathbf{s}}_t(a)$ and $\tilde{\mathbf{s}}_t(a)$ in the estimation error, respectively.

We note that at any time t, $N_t(a, c) = 1$ for $c = n_t + 1, \ldots, n_t + d$. Besides, according to (15) in Appendix B,

$$\hat{\sigma}_t(a) = \sqrt{\boldsymbol{\lambda}^\intercal \mathbf{X}_t(a)\mathbf{N}_t(a)\mathbf{X}_t^\intercal(a)\boldsymbol{\lambda}} = \|\mathbf{N}_t^{1/2}(a)\mathbf{X}_t^\intercal(a)\boldsymbol{\lambda}\|_2.$$

Thus,

$$|\bar{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\boldsymbol{\lambda}| = |\bar{\mathbf{s}}_t^\intercal(a)\mathbf{N}_t^{1/2}(a)\mathbf{X}_t^\intercal(a)\boldsymbol{\lambda}| \tag{23}$$
$$\leq \|\bar{\mathbf{s}}_t^\intercal(a)\|_2 \cdot \|\mathbf{N}_t^{1/2}(a)\mathbf{X}_t^\intercal(a)\boldsymbol{\lambda}\|_2 \tag{24}$$
$$= \|\boldsymbol{\theta}(a)\|_2\hat{\sigma}_t(a) \leq ls\hat{\sigma}_t(a), \tag{25}$$

where (24) follows from the Cauchy-Schwarz inequality.

Before we proceed to bound $\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\boldsymbol{\lambda}$, we first introduce the following notations. Recall that $\mathbf{V}_t(a) := \mathbf{X}_t(a)\mathbf{N}_t(a)\mathbf{X}_t^\intercal(a)$. Let $\mathbf{V}_t^{1/2}(a)$ be its square root, i.e., $\mathbf{V}_t^{1/2}(a)\mathbf{V}_t^{1/2}(a) = \mathbf{V}_t(a)$. Let $\tilde{\mathbf{V}}_t(a) := \sum_{c=1}^n N_t(a, c)\mathbf{x}(a, c)\mathbf{x}^\intercal(a, c)$, $\mathbf{V}_0 := l^2\mathbf{I}$. Then, $\mathbf{V}_t(a) = \tilde{\mathbf{V}}_t(a) + \mathbf{V}_0$. We have

$$|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\boldsymbol{\lambda}| = |\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\mathbf{V}_t^{-1/2}(a)\mathbf{V}_t^{1/2}(a)\boldsymbol{\lambda}|$$
$$\leq \|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\|_{\mathbf{V}_t^{-1}(a)}\|\mathbf{V}_t^{1/2}(a)\boldsymbol{\lambda}\|_2 \tag{26}$$
$$= \|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\|_{\mathbf{V}_t^{-1}(a)}\hat{\sigma}_t(a). \tag{27}$$

We then adopt the Laplace method (Lattimore and Szepesvári, 2019) to bound $\|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\|_{\mathbf{V}_t^{-1}(a)}$ as follows.

Lemma 1 *Denote* $M_t(\mathbf{u}) := \exp\left(\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\mathbf{u} - \frac{1}{2}\mathbf{u}^\intercal\tilde{\mathbf{V}}_t(a)\mathbf{u}\right)$ *for any* $\mathbf{u}$. *Let* $h(\mathbf{u})$ *be a probability measure over* $\mathbb{R}^d$. *Then,* $\bar{M}_t := \mathbb{E}_h[M_t(\mathbf{u})]$ *is a super martingale with* $\bar{M}_0 = 1$.

Proof. First, we note that $M_t(\mathbf{u}) - M_{t-1}(\mathbf{u})$ equals zero if arm a is not pulled at time t. Then, for any fixed $\mathbf{u}$, we have

$$\mathbb{E}[M_t(\mathbf{u}) \mid \mathcal{F}_{t-1}]$$
$$= \mathbb{E}\left[\exp\left(\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\mathbf{u} - \frac{1}{2}\mathbf{u}^\intercal\tilde{\mathbf{V}}_t(a)\mathbf{u}\right)\middle|\mathcal{F}_{t-1}\right] \tag{28}$$
$$= \mathbb{E}\left[\exp\left(\left(\tilde{\mathbf{s}}_t^\intercal(a) - \tilde{\mathbf{s}}_{t-1}^\intercal(a)\right)\mathbf{X}_t^\intercal(a)\mathbf{u} - \frac{1}{2}\mathbf{u}^\intercal\left(\tilde{\mathbf{V}}_t(a) - \tilde{\mathbf{V}}_{t-1}(a)\right)\mathbf{u}\right)\right] M_{t-1}(\mathbf{u}) \tag{29}$$

Based on the definition of $\tilde{\mathbf{s}}_t$ in (27), $\tilde{\mathbf{s}}_t^\intercal(a) - \tilde{\mathbf{s}}_{t-1}^\intercal(a)$ equals $\eta_t\mathbf{e}_c^\intercal$ if $a_t = a, c_t = c$, and zero otherwise; Similarly, for $\tilde{\mathbf{V}}_t(a) - \tilde{\mathbf{V}}_{t-1}(a)$, it equals $\mathbf{x}(a, c)\mathbf{x}^\intercal(a, c)$ if $a_t = t, c_t = c$, and zero otherwise. Therefore, if $a_t \neq a$,

$$\mathbb{E}\left[\exp\left(\left(\tilde{\mathbf{s}}_t^\intercal(a) - \tilde{\mathbf{s}}_{t-1}^\intercal(a)\right)\mathbf{X}^\intercal(a)\mathbf{u} - \frac{1}{2}\mathbf{u}^\intercal\left(\tilde{\mathbf{V}}_t(a) - \tilde{\mathbf{V}}_{t-1}(a)\right)\mathbf{u}\right)\middle|a_t \neq a\right] = 1. \tag{30}$$

If $a_t = a$, $c_t = c$,

$$\mathbb{E}\left[\exp\left(\left(\tilde{\mathbf{s}}_t^\intercal(a) - \tilde{\mathbf{s}}_{t-1}^\intercal(a)\right)\mathbf{X}^\intercal(a)\mathbf{u} - \frac{1}{2}\mathbf{u}^\intercal\left(\tilde{\mathbf{V}}_t(a) - \tilde{\mathbf{V}}_{t-1}(a)\right)\mathbf{u}\right)\middle|a_t = a, c_t = c\right] \tag{31}$$
$$= \mathbb{E}\left[\exp\left(\eta_t\mathbf{e}_c\mathbf{X}^\intercal(a)\mathbf{u} - \frac{1}{2}\mathbf{u}^\intercal\mathbf{x}(a, c)\mathbf{x}^\intercal(a, c)\mathbf{u}\right)\right] \tag{32}$$

$$= \mathbb{E}\left[\exp\left(\eta_t \mathbf{x}^\intercal(a,c)\mathbf{u}\right)\right] \cdot \exp\left(-\frac{1}{2}\mathbf{u}^\intercal \mathbf{x}(a,c)\mathbf{x}^\intercal(a,c)\mathbf{u}\right) \tag{33}$$

$$\leq \exp\left(\frac{1}{2}\left(\mathbf{x}^\intercal(a,c)\mathbf{u}\right)^2\right) \cdot \exp\left(-\frac{1}{2}\mathbf{u}^\intercal \mathbf{x}(a,c)\mathbf{x}^\intercal(a,c)\mathbf{u}\right) = 1 \tag{34}$$

where the last inequality follows from Assumption 1.3 that η_t is conditionally 1-subgaussian.

Combining (30)(34) with (29), for every fixed $\mathbf{u}$, we have $\mathbb{E}[M_t(\mathbf{u}) \mid \mathcal{F}_{t-1}] \leq M_{t-1}(\mathbf{u})$. Thus, $\{M_t(\mathbf{u})\}_t$ is a super-martingale, and

$$\mathbb{E}[M_t(\mathbf{u})] \leq M_0(\mathbf{u})] = 1. \tag{35}$$

Since this holds for every $\mathbf{u}$, after taking expectation with respect to $\mathbf{u}$, $\bar{M}_t$ is a super-martingale as well. Thus, $\mathbb{E}[\bar{M}_t] \leq \bar{M}_0 = 1$. ■

Lemma 2 *Under Algorithm 1,*

$$\mathbb{P}\left[\|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\|_{\mathbf{V}_t^{-1}(a)} \geq \sqrt{2u + \log\frac{\det \mathbf{V}_t(a)}{\det \mathbf{V}_0}}\right] \leq e^{-u}.$$

Proof. Assume h is the probability density function of a Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{V}_0)$, i.e.,

$$h(\mathbf{u}) = \frac{1}{\sqrt{(2\pi)^d \det \mathbf{V}_0}} \exp\left(-\frac{1}{2}\mathbf{u}^\intercal \mathbf{V}_0 \mathbf{u}\right).$$

Then,

$$\begin{aligned}
\bar{M}_t &= \int_{\mathbb{R}^d} M_t(\mathbf{u}) h(\mathbf{u})\,\mathrm{d}\mathbf{u} \\
&= \frac{1}{\sqrt{(2\pi)^d \det \mathbf{V}_0^{-1}}} \int_{\mathbb{R}^d} \exp\left(\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}^\intercal(a)\mathbf{u} - \frac{1}{2}\mathbf{u}^\intercal \tilde{\mathbf{V}}_t(a)\mathbf{u} - \frac{1}{2}\mathbf{u}^\intercal \mathbf{V}_0 \mathbf{u}\right)\mathrm{d}\mathbf{u} \\
&= \frac{1}{\sqrt{(2\pi)^d \det \mathbf{V}_0^{-1}}} \int_{\mathbb{R}^d} \exp\left(\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}^\intercal(a)\mathbf{u} - \frac{1}{2}\mathbf{u}^\intercal \mathbf{V}_t(a)\mathbf{u}\right)\mathrm{d}\mathbf{u} \\
&= \frac{1}{\sqrt{(2\pi)^d \det \mathbf{V}_0^{-1}}} \int_{\mathbb{R}^d} \exp\left(\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}^\intercal(a)\mathbf{V}_t^{-1/2}(a)\mathbf{V}_t^{1/2}(a)\mathbf{u} - \frac{1}{2}\|\mathbf{u}^\intercal \mathbf{V}_t(a)\|_{\mathbf{V}_t^{-1}(a)}\right)\mathrm{d}\mathbf{u} \\
&= \frac{1}{\sqrt{(2\pi)^d \det \mathbf{V}_0^{-1}}} \\
&\quad \times \int_{\mathbb{R}^d} \exp\left(\frac{1}{2}\|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}^\intercal(a)\|^2_{\mathbf{V}_t^{-1}(a)} - \frac{1}{2}\|\mathbf{u}^\intercal \mathbf{V}_t(a) - \tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}^\intercal(a)\|^2_{\mathbf{V}_t^{-1}(a)}\right)\mathrm{d}\mathbf{u} \\
&= \sqrt{\frac{\det \mathbf{V}_t^{-1}(a)}{\det \mathbf{V}_0^{-1}}} \exp\left(\frac{1}{2}\|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}^\intercal(a)\|^2_{\mathbf{V}_t^{-1}(a)}\right) \\
&= \exp\left(\frac{1}{2}\|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}^\intercal(a)\|^2_{\mathbf{V}_t^{-1}(a)} + \frac{1}{2}\log\frac{\det \mathbf{V}_0}{\det \mathbf{V}_t(a)}\right).
\end{aligned}$$

Therefore, according to Lemma 1, we have

$$\begin{aligned}
&\mathbb{P}\left[\|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\|_{\mathbf{V}_t^{-1}(a)} \geq \sqrt{2u + \log\frac{\det \mathbf{V}_t(a)}{\det \mathbf{V}_0}}\right] \\
&= \mathbb{P}\left[\frac{1}{2}\|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}_t^\intercal(a)\|^2_{\mathbf{V}_t^{-1}(a)} + \frac{1}{2}\log\frac{\det \mathbf{V}_0}{\det \mathbf{V}_t(a)} \geq u\right]
\end{aligned}$$

$$\leq e^{-u}\mathbb{E}[\bar{M}_t] \leq e^{-u}. \tag{36}$$

■

Next, we will provide a bound on $\frac{\det \mathbf{V}_0}{\det \mathbf{V}_t(a)}$. The definition of $\mathbf{V}_0$ indicates that $\det \mathbf{V}_0 = l^{2d}$. For $\mathbf{V}_t(a)$, we note that for any $\mathbf{y} \in \mathbb{R}^d$, we have

$$\begin{aligned}\mathbf{y}^\intercal \mathbf{V}_t(a)\mathbf{y} &= \sum_{c=1}^{n} N_t(a,c)\mathbf{y}^\intercal \mathbf{x}(a,c)\mathbf{x}^\intercal(a,c)\mathbf{y} + l^2\mathbf{y}^\intercal\mathbf{y}\\ &\leq tl^2\|\mathbf{y}\|_2^2 + l^2\|\mathbf{y}\|_2^2 = (t+1)l^2\|\mathbf{y}\|_2^2, \end{aligned}\tag{37}$$

where the inequality follows from Assumption 1.1.

Eqn. (37) indicates that the maximum eigenvalue of $\mathbf{V}_t(a)$ is upper bound by $(t+1)l^2$. Therefore, we have $\det \mathbf{V}_t(a) \leq (l^2 + tl^2)^d$, which implies that

$$\log \frac{\det \mathbf{V}_t(a)}{\det \mathbf{V}_0} \leq d\log(1+t). \tag{38}$$

Combining (38) with (22)(25)(27) and Lemma 2, we have

$$\begin{aligned}&\mathbb{P}\left[|\hat{r}_t(a) - r(a,c_t)| \geq \left(ls + \sqrt{2u + d\log(1+t)}\right)\hat{\sigma}_t(a)\right]\\ &\leq \mathbb{P}\left[\|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}^\intercal(a)\|_{\mathbf{V}_t^{-1}(a)}\hat{\sigma}_t(a) \geq \sqrt{2u + d\log(1+t)}\hat{\sigma}_t(a)\right]\\ &\leq \mathbb{P}\left[\|\tilde{\mathbf{s}}_t^\intercal(a)\mathbf{X}^\intercal(a)\|_{\mathbf{V}_t^{-1}(a)} \geq \sqrt{2u + \log\frac{\det \mathbf{V}_t(a)}{\det \mathbf{V}_0}}\right] \leq e^{-u}. \end{aligned}\tag{39}$$

Set $u = \log f(t)$ and $\alpha_t = ls + \sqrt{(2+d)\log f(t)}$. When $t > 2$, $f(t) > 1 + t$, therefore, (39) implies that

$$\mathbb{P}\left[|\hat{r}_t(a) - r(a,c_t)| \geq \alpha_t\hat{\sigma}_t(a)\right] \leq \frac{1}{f(t)}. \tag{40}$$

Thus,

$$\mathbb{E}[|\mathcal{B}_T|] \leq 2 + K\sum_{t=3}^{\infty}\frac{1}{f(t)} \leq 2 + 2.5K, \quad \mathbb{E}[R(\mathcal{B}_T)] \leq M\mathbb{E}[|\mathcal{B}_T|] \leq M(2+2.5K). \tag{41}$$

C.3 Bound the Regret over $\mathcal{C}_T$

Recall $B_k := |\mathcal{B}_T \cap F_k|$, i.e., the number of bad estimates in frame k. Then, according to Markov's inequality, we have

$$\mathbb{P}\left[B_k \geq \frac{2^{k-1}}{4n}\right] \leq \frac{\mathbb{E}[B_k]4n}{2^{k-1}}. \tag{42}$$

The definitions of B_k and $\mathcal{B}_T$ also imply that $\sum_{k=1}^{\lceil \log_2 T\rceil}\mathbb{E}[B_k] = \mathbb{E}[\mathcal{B}_T]$. Therefore,

$$\mathbb{E}[R(\mathcal{C}_T)] \leq M\mathbb{E}[|\mathcal{C}_T|] = M\sum_{k=1}^{\lfloor \log_2 T\rfloor}|F_{k+1}| \cdot \mathbb{P}\left[B_k \geq \frac{2^{k-1}}{4n}\right] \tag{43}$$

$$\leq M\sum_{k=1}^{\lfloor \log_2 T\rfloor} 2^k\frac{\mathbb{E}[B_k]4n}{2^{k-1}} \leq 8nM\mathbb{E}[|\mathcal{B}_T|] \leq 8nM(2+2.5K), \tag{44}$$

where (43) follows from the definition of $\mathcal{C}_T$, and (44) is due to (42) and (41).

C.4 Bound the Regret over $\mathcal{D}_T$

Let $\bar{N}_t(a,c)$ be the total number of time slots before t when arm a is pulled under context c, and all estimates are good, i.e., $\bar{N}_t(a,c) = |\{\tau \mid a_\tau = a, c_\tau = c, \tau \notin \mathcal{B}_t, 1 \leq \tau < t\}|$. We have the following observations.

Lemma 3 *For any a, $c \notin \mathcal{C}_a$, $\bar{N}_t(a,c) \leq \frac{4\alpha_t^2}{\Delta^2}$ for all t.*

Proof. We first consider a time slot $t \notin \mathcal{B}_T$ at which a sub-optimal action a_t is taken under c_t. Then, according to the LinUCB-d Algorithm, we must have

$$\hat{r}_t(a_t) + \alpha_t \hat{\sigma}_t(a_t) \geq \hat{r}_t(a_t^*) + \alpha_t \hat{\sigma}_t(a_t^*). \tag{45}$$

Besides, $t \notin \mathcal{B}_T$ implies that

$$|\hat{r}_t(a_t) - r(a_t, c_t)| \leq \alpha_t \hat{\sigma}_t(a_t), \quad |\hat{r}_t(a_t^*) - r(a_t^*, c_t)| \leq \alpha_t \hat{\sigma}_t(a_t^*). \tag{46}$$

Putting (45)(46) together, we have

$$r(a_t, c_t) + 2\alpha_t \hat{\sigma}_t(a_t) \geq \hat{r}_t(a_t) + \alpha_t \hat{\sigma}_t(a_t) \geq \hat{r}_t(a_t^*) + \alpha_t \hat{\sigma}_t(a_t^*) \geq r(a_t^*, c_t). \tag{47}$$

Therefore,

$$\Delta \leq r(a_t^*, c_t) - r(a_t, c_t) \leq 2\alpha_t \hat{\sigma}_t(a_t) = 2\alpha_t \sqrt{\boldsymbol{\beta}_t^\intercal(a_t)\mathbf{N}_t^{-1}(a_t)\boldsymbol{\beta}_t(a_t)}. \tag{48}$$

Denote $\tilde{\boldsymbol{\beta}} \in \mathbb{R}^{n_t+d}$ as a unit vector whose c_t-th entry takes value 1. Then, when $N_t(a_t, c_t) \neq 0$, $\tilde{\boldsymbol{\beta}}$ satisfies the constraints in (6). According to Proposition 1, we must have

$$\boldsymbol{\beta}_t^\intercal(a_t)\mathbf{N}_t^{-1}(a_t)\boldsymbol{\beta}_t(a_t) \leq \tilde{\boldsymbol{\beta}}^\intercal(a_t)\mathbf{N}_t^{-1}(a_t)\tilde{\boldsymbol{\beta}}(a_t) = \frac{1}{N_t(a_t, c_t)} \leq \frac{1}{\bar{N}_t(a_t, c_t)}. \tag{49}$$

Combining (48) and (49), we have

$$\bar{N}_t(a_t, c_t) \leq \frac{1}{\boldsymbol{\beta}_t^\intercal(a_t)\mathbf{N}_t^{-1}(a_t)\boldsymbol{\beta}_t(a_t)} \leq \frac{4\alpha_t^2}{\Delta^2}. \tag{50}$$

When $N_t(a_t, c_t) = 0$, we must have $\bar{N}_t(a_t, c_t) = 0$, thus (50) is satisfied as well.

Hence, Lemma 3 holds for all time slots $t \notin \mathcal{B}_T$. Since $\bar{N}_t(a,c)$ is a step function for any fixed (a,c) pair and α_t monotonically increases in t, Lemma 3 hold for all t as well. ∎

Lemma 3 indicates that the total number of times that a is pulled as a sub-optimal arm up to t is bounded by $O(\log f(t))$. Based on this result, we will then show that the total number of times that a is pulled as an optimal arm grows linearly in t, as described in Lemma 4. Next, we utilize Lemma 4 to show the diminishing estimation uncertainty in Lemma 6, which eventually leads to the finite regret bound over $\mathcal{D}_T$ in Theorem 4.

Lemma 4 *For any a, $c \in \mathcal{C}_a$ and any time slot $t \in \mathcal{D}_T$, we must have $N_t(a,c) \geq \frac{t}{16n} - \frac{8K\alpha_t^2}{\Delta^2}$.*

Proof. Assume t lies in the $(k+1)$th time frame. Then, based on the definition of $N_t(a,c)$, we must have

$$N_t(a,c) \geq N_{F_k}(a,c) = N_{F_k}(c) - \sum_{b:b\neq a} N_{F_k}(b,c) \geq \frac{2^{k-1}}{2n} - \Big[B_k + \sum_{b:b\neq a} \bar{N}_{2^k}(b,c)\Big] \tag{51}$$

$$\geq \frac{2^{k-1}}{2n} - \frac{2^{k-1}}{4n} - K\frac{4\alpha_t^2}{\Delta^2} \geq \frac{t}{16n} - \frac{4K\alpha_t^2}{\Delta^2}, \tag{52}$$

where (51) follows from the assumption that $t \notin \mathcal{A}_T$, and (52) follows from the assumption that $t \notin \mathcal{C}_T$ and Lemma 3. ∎

Before we proceed, we introduce the following lemma.

Lemma 5 *Let $\{\phi_1, \phi_2, \ldots, \phi_d\}$ be a basis for $\mathbb{R}^d$, and $\Phi := [\phi_1, \phi_2, \ldots, \phi_d]$. Then, for any $\mathbf{x} \in \mathbb{R}^d$, $\|\mathbf{x}\|_2 \leq l$, we can express it as $\mathbf{x} = \Phi\boldsymbol{\beta}$, where $\boldsymbol{\beta} \in \mathbb{R}^d$, $\|\boldsymbol{\beta}\|_1 \leq \frac{l\sqrt{d}}{\sqrt{\lambda_{\min}(\Phi^\intercal\Phi)}}$.*

Proof. Since

$$l^2 \geq \|\mathbf{x}\|_2^2 = \boldsymbol{\beta}^\intercal\Phi^\intercal\Phi\boldsymbol{\beta} \geq \lambda_{\min}(\Phi^\intercal\Phi)\boldsymbol{\beta}^\intercal\boldsymbol{\beta} \geq \frac{\lambda_{\min}(\Phi^\intercal\Phi)\|\boldsymbol{\beta}\|_1^2}{d}, \tag{53}$$

we have $\|\boldsymbol{\beta}\|_1 \leq \frac{l\sqrt{d}}{\sqrt{\lambda_{\min}(\Phi^\intercal\Phi)}}$. ■

Lemma 6 *For any arm $a \in [K]$, any time slot $t \in \mathcal{D}_T$, we must have $\boldsymbol{\beta}_t^\intercal(a)\mathbf{N}_t^{-1}(a)\boldsymbol{\beta}_t(a) \leq \frac{\delta^2}{\frac{t}{16n} - \frac{4K\alpha_t^2}{\Delta^2}}$, where $\delta := l\sqrt{d/\lambda_0}$.*

Proof. For any $a \in [K], c \in \mathcal{C}$, let $\bar{\boldsymbol{\beta}}(a, c)$ be the solution to the following equation

$$\mathbf{x}(a, c) = \mathbf{X}_t(a)\bar{\boldsymbol{\beta}}, \quad \bar{\boldsymbol{\beta}}[c] = 0, \text{ for } c \notin \bar{\mathcal{C}}_a. \tag{54}$$

Note that we use $\bar{\boldsymbol{\beta}}[c]$ to denote the entry associated with context c in $\bar{\boldsymbol{\beta}}$.

Consider a time slot $t \in \mathcal{D}_T$. Based on the definitions of the error events in Section 4, we note that all contexts in $\bar{\mathcal{C}}_a$ must have appeared before time slot t. Thus, $\mathbf{X}_t(a)$ contains all columns in $\bar{\Phi}_a$. Therefore, $\bar{\boldsymbol{\beta}}(a, c)$ is simply the coefficient vector if we express $\mathbf{x}(a, c)$ as a linear combination of the feature vectors in $\bar{\Phi}_a$. The diversity assumption in Assumption 1.5 guarantees that there exists a unique solution $\bar{\boldsymbol{\beta}}(a, c)$ for each (a, c) pair. Besides, Lemma 5 implies that $\|\bar{\boldsymbol{\beta}}(a, c)\|_1 \leq \frac{l\sqrt{d}}{\sqrt{\lambda_{\min}(\bar{\Phi}_a^\intercal\bar{\Phi}_a)}}$.

Then, according to Proposition 1, Lemma 4 and Lemma 5, we must have

$$\boldsymbol{\beta}_t^\intercal(a)\mathbf{N}_t^{-1}(a)\boldsymbol{\beta}_t(a) \leq \bar{\boldsymbol{\beta}}^\intercal(a, c_t)\mathbf{N}_t^{-1}(a)\bar{\boldsymbol{\beta}}(a, c_t) \leq \frac{\|\bar{\boldsymbol{\beta}}(a, c_t)\|_1^2}{\frac{t}{16n} - \frac{4K\alpha_t^2}{\Delta^2}} \leq \frac{\delta^2}{\frac{t}{16n} - \frac{4K\alpha_t^2}{\Delta^2}}, \tag{55}$$

where the first inequality in (55) follows from Proposition 1, the second inequality follows from Lemma 4, and the last inequality follows from Lemma 5. ■

We then have the following bound on $\mathbb{E}[R(\mathcal{D}_T)]$.

Theorem 3 *Let*

$$t_1 = \max\left\{\frac{384(2+d)n(\delta^2+K)}{\Delta^2}, 10\right\}, \quad t_2 = \max\left\{t_1\log t_1, \exp\left(\frac{12l^2s^2}{2+d}\right)\right\}.$$

Then, under Algorithm 1,

$$\mathbb{E}[R(\mathcal{D}_T)] \leq t_2 M = O\left(\frac{dn(\delta^2+K)}{\Delta^2}\log\frac{dn(\delta^2+K)}{\Delta^2}\right).$$

Proof. For any $t \geq t_2$, we have

$$t \geq \exp\left(\frac{12l^2s^2}{2+d}\right) \geq \exp\left(\frac{(\sqrt{2}+\sqrt{3})^2l^2s^2}{2+d}\right),$$

which implies that

$$ls \leq \sqrt{2+d}(\sqrt{3}-\sqrt{2})\sqrt{\log t}. \tag{56}$$

Meanwhile, since $\log f(t) \leq 2\log t$, combining with (56), we have

$$\alpha_t := ls + \sqrt{(2+d)\log f(t)} \leq \sqrt{3(2+d)\log t}. \tag{57}$$

Since $\frac{t_2}{\log t_2} \geq \frac{t_1 \log t_1}{\log(t_1 \log t_1)} \geq \frac{t_1}{2}$, for any $t \geq t_2$, we have

$$\frac{t}{\log t} \geq \frac{t_1}{2} = \frac{192(2+d)n(\delta^2+K)}{\Delta^2}, \tag{58}$$

i.e.,

$$t > \frac{192(2+d)n(\delta^2+K)}{\Delta^2} \log t \geq \frac{64n(\delta^2+K)}{\Delta^2}\alpha_t^2,$$

where the last inequality is due to (57).

Thus,

$$\frac{t}{16n} \geq \frac{4\alpha_t^2\delta^2}{\Delta^2} + \frac{4K\alpha_t^2}{\Delta^2}. \tag{59}$$

Rearranging the terms, we have

$$\Delta^2 > \frac{4\alpha_t^2\delta^2}{\frac{t}{16n} - \frac{4K\alpha_t^2}{\Delta^2}} \geq (2\alpha_t\hat{\sigma}_t(a))^2, \quad \forall a \in [K], \tag{60}$$

where the last inequality follows from Lemma 6.

Since $\Delta \geq 2\alpha_t\hat{\sigma}_t(a)$ for any $t \geq t_2$, arm a will not be pulled as a suboptimal arm at any time $t \notin \mathcal{B}_T$, according to Eqn. (48). Therefore, $\mathcal{D}_T$ can only include time indices before t_2. The expected regret over $\mathcal{D}_T$ can thus be bounded by Mt_2. ■

C.5 Put Everything Together

After obtaining bounds on the expected regret over $\mathcal{A}_T$, $\mathcal{B}_T$, $\mathcal{C}_T$ and $\mathcal{D}_T$, we are ready to prove our main result in Theorem 1. We have

$$\begin{aligned}
\mathbb{E}[R_T] &\leq \mathbb{E}[R(\mathcal{A}_T)] + \mathbb{E}[R(\mathcal{B}_T)] + \mathbb{E}[R(\mathcal{C}_T)] + \mathbb{E}[R(\mathcal{D}_T)] \\
&\leq 8MKdn^2 + (8n+1)M(2+2.5K) + t_2M \qquad (61)\\
&= O\left(Kdn^2 + \frac{dn(K+\delta^2)}{\Delta^2}\log\frac{dn(K+\delta^2)}{\Delta^2}\right). \qquad (62)
\end{aligned}$$

We point out that the $O\left(\exp\left(\frac{12l^2s^2}{2+d}\right)\right)$ term from t_2 is dropped in (62), since it mainly depends on the bounds on $\|\boldsymbol{\theta}(a)\|_2$ and $\|\mathbf{x}(a,c)\|_2$, and does not scale with the system dimensions d or K.

D Proof of Theorem 2

Before we proceed, we will first introduce the following lemma, which will play a critical role in the analysis afterwards.

Lemma 7 *Let $\{\phi_1, \phi_2, \ldots, \phi_d\}$ be a basis for $\mathbb{R}^d$, and $\Phi := [\phi_1, \phi_2, \ldots, \phi_d]$. Let $B(\phi_i, r)$ be an ℓ_2 ball centered at ϕ_i with radius $r < \sqrt{\lambda_{\min}(\Phi^\intercal\Phi)/d}$, i.e., $B(\phi_i, r) := \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \phi_i\|_2 \leq r\}$. Let $\hat{\phi}_i$ be any vector lying in $B(\phi_i, r)$ and $\hat{\Phi} := [\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_d]$. Then, $\lambda_{\min}(\hat{\Phi}^\intercal\hat{\Phi}) \geq (\sqrt{\lambda_{\min}(\Phi^\intercal\Phi)} - \sqrt{d}r)^2$.*

Proof. Denote $\gamma_i := \hat{\phi}_i - \phi_i$. Then, based on the definition of $\hat{\phi}_i$, we have $\|\gamma_i\|_2 \leq r$. Let $\Gamma = [\gamma_1, \gamma_2, \ldots, \gamma_d]$, and $\Gamma(j)$ be its jth row. Then, for any $\boldsymbol{\beta} \in \mathbb{R}^d$,

$$\|\hat{\Phi}\boldsymbol{\beta}\|_2 = \|\Phi\boldsymbol{\beta} + \Gamma\boldsymbol{\beta}\|_2 \geq \|\Phi\boldsymbol{\beta}\|_2 - \|\Gamma\boldsymbol{\beta}\|_2 \geq \sqrt{\lambda_{\min}(\Phi^\intercal\Phi)}\|\boldsymbol{\beta}\|_2 - \sqrt{\sum_j |\Gamma(j)\boldsymbol{\beta}|^2} \tag{63}$$

$$\geq \sqrt{\lambda_{\min}(\Phi^\intercal \Phi)}\|\boldsymbol{\beta}\|_2 - \sqrt{\sum_j \|\Gamma(j)\|_2^2 \|\boldsymbol{\beta}\|_2^2} \tag{64}$$

$$= \sqrt{\lambda_{\min}(\Phi^\intercal \Phi)}\|\boldsymbol{\beta}\|_2 - \|\boldsymbol{\beta}\|_2 \sqrt{\sum_{i=1}^{d} \gamma_i^\intercal \gamma_i} \tag{65}$$

$$\geq \left(\sqrt{\lambda_{\min}(\Phi^\intercal \Phi)} - \sqrt{d}r\right)\|\boldsymbol{\beta}\|_2, \tag{66}$$

where (63) follows from (53), and (64) follows from the Cauchy-Schwartz inequality; Rearranging the terms involved in the summation, we obtain (65), which can be further bounded by (66) due to the definition of $B(\phi_i, r)$. Thus, the eigenvalues of $\hat{\Phi}^\intercal \hat{\Phi}$ are lower bounded by $(\sqrt{\lambda_{\min}(\Phi^\intercal \Phi)} - \sqrt{d}r)^2 > 0$. ■

Remark: Lemma 7 implies that $\{\hat{\phi}_i\}$ are linearly independent, thus forming a valid basis for $\mathbb{R}^d$.

D.1 Bound the Regret over $\mathcal{A}_T$

First, based on Hoeffding's inequality, we have

$$\mathbb{P}\left[N_{F_k}(\bar{\mathcal{C}}_a^{(i)}) \leq \frac{p}{2} \cdot 2^{k-1}\right] \leq \exp\left(-p^2 2^{k-2}\right). \tag{67}$$

Recall that M is the maximum per-step regret. Thus, by extending the proof in Appendix C.1, we have

$$\mathbb{E}[R(\mathcal{A}_T)] \leq M(Kd)\sum_{t=2}^{\infty} \exp\left(-\frac{p^2}{8}t\right) \leq \frac{8MKd}{p^2}. \tag{68}$$

D.2 Bound the Regret over $\mathcal{C}_T$

According to Markov's inequality, we have

$$\mathbb{P}\left[B_k \geq \frac{p \cdot 2^{k-1}}{4}\right] \leq \frac{\mathbb{E}[B_k] \cdot 4}{p \cdot 2^{k-1}}. \tag{69}$$

Therefore, by following similar steps in Appendix C.3, we have

$$\mathbb{E}[R(\mathcal{C}_T)] \leq \frac{8M}{p}\mathbb{E}[\mathcal{B}_T] \leq \frac{8M(2+2.5K)}{p}. \tag{70}$$

D.3 Bound the Regret over $\mathcal{D}_T$

Before we proceed, we first state an adapted version of the celebrated elliptical potential lemma below, which will play a key role to analysis afterwards.

Lemma 8 (Elliptical Potential (Lattimore and Szepesvári, 2019)) *Let $\mathbf{V}_0$ be positive definite and $\mathbf{V}_t = \mathbf{V}_{t-1} + \mathbf{x}_t \mathbf{x}_t^\intercal$, where $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ is a sequence of vectors with $\|\mathbf{x}_t\|_2 \leq l < \infty$ for all t. Then,*

$$\sum_{t=1}^{n}\left(1 \wedge \|\mathbf{x}_t\|_{\mathbf{V}_{t-1}^{-1}}^2\right) \leq 2\log\left(\frac{\det \mathbf{V}_n}{\det \mathbf{V}_0}\right) \leq 2d\log\left(\frac{trace\mathbf{V}_0 + nl^2}{d\det^{1/d}\mathbf{V}_0}\right),$$

where $x \wedge y = \min\{x, y\}$.

Let $\mathcal{T}_t(a, \mathcal{C}_b^{(i)})$ be the time slots before t when arm a is pulled under a context lying in $\mathcal{C}_b^{(i)}$, and at the same time, all estimates are good, i.e.,

$$\mathcal{T}_t(a, \mathcal{C}_b^{(i)}) := \{\tau \mid a_\tau = a, c_\tau \in \mathcal{C}_b^{(i)}, \tau \notin \mathcal{B}_t, 1 \leq \tau < t\}, \tag{71}$$

and denote $\bar{N}_t(a, \mathcal{C}_b^{(i)}) := |\mathcal{T}(a, \mathcal{C}_b^{(i)})|$.

We have the following lemma analogue to Lemma 3.

Lemma 9 *For any a, $b \neq a$, $\bar{N}_t(a, \mathcal{C}_b^{(i)}) \leq \frac{8\alpha_t^2}{\Delta^2} d \log\left(\frac{d+t}{d}\right)$ for all t.*

Proof. First, following steps similar to the proof of Lemma 3, for any $\tau \in \mathcal{T}_t(a, \mathcal{C}_b^{(i)})$, we have

$$\Delta \leq r(a_\tau^*, c_\tau) - r(a, c_\tau) \leq 2\alpha_\tau \hat{\sigma}_\tau(a). \tag{72}$$

Next, we consider the solution to the following optimization problem, denoted as $\tilde{\boldsymbol{\beta}}_t(a)$:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{n_t + d}} \quad \boldsymbol{\beta}^\intercal \mathbf{N}_t^{-1}(a) \boldsymbol{\beta}, \quad \text{s.t.} \quad \mathbf{x}(a, c_t) = \mathbf{X}_t(a)\boldsymbol{\beta}, \quad \boldsymbol{\beta}[c] = 0 \text{ if } c \notin \mathcal{C}_b^{(i)} \cup \mathcal{C}_0. \tag{73}$$

Compared with the optimization problem in (6), we have one additional constraint, i.e., we only restrict to the contexts in $\mathcal{C}_b^{(i)}$ and $\mathcal{C}_0$. The inclusion of the dummy contexts $\mathcal{C}_0$ ensures the existence of at least one feasible solution to (73). Due to the additional constraint, the corresponding minimum value of the objective function must increase, i.e.,

$$\tilde{\sigma}_t(a) := \sqrt{\tilde{\boldsymbol{\beta}}_t^\intercal(a) \mathbf{N}_t^{-1}(a) \tilde{\boldsymbol{\beta}}_t(a)} \geq \hat{\sigma}_t(a), \quad \forall t. \tag{74}$$

Note that

$$r(a_\tau^*, c_\tau) - r(a, c_\tau) \leq 2ls \leq 2\alpha_\tau, \tag{75}$$

where the first inequality in (75) follows from Assumption 1.1 and the second inequality follows from the definition of α_t.

Combining (74)(75) with (72), we have

$$\Delta \leq 2\alpha_\tau \left(1 \wedge \tilde{\sigma}_\tau(a)\right), \quad \tau \in \mathcal{T}_t(a, \mathcal{C}_{b,i}). \tag{76}$$

Summing over all $\tau \in \mathcal{T}_t(a, \mathcal{C}_b^{(i)})$, we have

$$\begin{aligned} \bar{N}_t(a, \mathcal{C}_b^{(i)})\Delta &\leq \sum_{\tau \in \mathcal{T}_t(a, \mathcal{C}_b^{(i)})} 2\alpha_\tau \left(1 \wedge \tilde{\sigma}_\tau(a)\right) \\ &\leq 2\alpha_t \sqrt{\bar{N}_t(a, \mathcal{C}_b^{(i)}) \left(\sum_{\tau \in \mathcal{T}_t(a, \mathcal{C}_b^{(i)})} \left(1 \wedge \tilde{\sigma}_\tau(a)\right)^2 \right)}, \end{aligned} \tag{77}$$

where (77) follows from the monotonicity of α_t and the Cauchy-Schewartz inequality.

Consider the sequence of feature vectors $\{\mathbf{x}(a, c_\tau)\}_{\tau \in \mathcal{T}_t(a, \mathcal{C}_b^{(i)})}$. Label the times indices in $\mathcal{T}_t(a, \mathcal{C}_b^{(i)})$ as τ_1, τ_2, ... Let $\tilde{\mathbf{V}}_0 = l^2\mathbf{I}$, $\tilde{\mathbf{V}}_{\tau_i} = \tilde{\mathbf{V}}_{\tau_{i-1}} + \mathbf{x}(a, c_{\tau_i})\mathbf{x}(a, c_{\tau_i})^\intercal$. Then, similar to (15), we have $\tilde{\sigma}_{\tau_i}(a) = \|\mathbf{x}(a, c_{\tau_i})\|_{\tilde{\mathbf{V}}_{\tau_{i-1}}^{-1}}$. Following Lemma 8, we have

$$\sum_{\tau \in \mathcal{T}_t(a, \mathcal{C}_b^{(i)})} \left(1 \wedge (\tilde{\sigma}_\tau(a))^2\right) \leq 2d \log\left(\frac{dl^2 + \bar{N}_t(a, \mathcal{C}_b^{(i)})l^2}{dl^2}\right) \leq 2d \log\left(\frac{d+t}{d}\right). \tag{78}$$

Plugging (78) into (77) and rearranging the terms, we have $\bar{N}_t(a, \mathcal{C}_b^{(i)}) \leq \frac{8\alpha_t^2}{\Delta^2} d \log\left(\frac{d+t}{d}\right)$ for all t. ■

Lemma 10 *For any $a \in [K]$, any time slot $t \in \mathcal{D}_T$, $N_t(a, \bar{\mathcal{C}}_a^{(i)}) \geq \frac{tp}{16} - \frac{8K\alpha_t^2}{\Delta^2} d \log\left(\frac{d+t}{d}\right)$.*

Proof. Assume t lies in the $(k+1)$th time frame. Then, based on the definition of $N_t(a, \bar{\mathcal{C}}_a^{(i)})$, we have

$$N_t(a, \bar{\mathcal{C}}_a^{(i)}) \geq N_{F_k}(a, \bar{\mathcal{C}}_a^{(i)}) \geq N_{F_k}(\bar{\mathcal{C}}_a^{(i)}) - \sum_{b: a \neq b} N_{F_k}(b, \bar{\mathcal{C}}_a^{(i)}) \tag{79}$$

$$\geq \frac{2^{k-1}}{2}p - \left[B_k + \sum_{b:a\neq b} \bar{N}_{2^k}(b, \mathcal{C}_a^{(i)})\right] \tag{80}$$

$$\geq \frac{2^{k-1}}{2}p - \frac{2^{k-1}}{4}p - K\frac{8\alpha_t^2}{\Delta^2} d\log\left(\frac{d+t}{d}\right) \tag{81}$$

$$= \frac{tp}{16} - \frac{8K\alpha_t^2}{\Delta^2} d\log\left(\frac{d+t}{d}\right), \tag{82}$$

where (79) follows from the assumption that $t \notin \mathcal{A}_T$, (80) follows from the fact that $\bar{\mathcal{C}}_a^{(i)} \subseteq \mathcal{C}_a^{(i)}$ thus $N_{F_k}(b, \bar{\mathcal{C}}_a^{(i)}) \leq N_{F_k}(b, \mathcal{C}_a^{(i)})$, and (81) follows from Lemma 9. ■

Lemma 11 *For any arm $a \in [K]$, and any time slot $t \in \mathcal{D}_T$, we have $\hat{\sigma}_t(a)^2 \leq \frac{4\delta^2}{\frac{tp}{16} - \frac{8K\alpha_t^2}{\Delta^2} d\log\left(\frac{d+t}{d}\right)}$, where $\delta = l\sqrt{d/\lambda_0(\{\Phi_a\})}$.*

Proof. Let

$$\hat{\phi}_t^{(i)}(a) := \frac{\sum_{\tau \in \mathcal{T}_t(a, \bar{\mathcal{C}}_a^{(i)})} \mathbf{x}(a, c_\tau)}{N_t(a, \bar{\mathcal{C}}_a^{(i)})} \tag{83}$$

be the empirical average of the feature vectors over the time slots before t when arm a is pulled under a context in $\bar{\mathcal{C}}_a^{(i)}$. Since $B(\phi_a^{(i)}, r)$ is convex, $\hat{\phi}_t^{(i)}(a) \in B(\phi_a^{(i)}, r)$. Thus, according to Lemma 7, $\{\hat{\phi}_t^{(i)}(a)\}_i$ form a valid basis for $\mathcal{X}_a$. Let $\hat{\Phi}_t(a)$ be the matrix whose columns are $\hat{\phi}_t^{(i)}(a)$. Then, we can always obtain a vector $\bar{\boldsymbol{\beta}}$, such that $\mathbf{x}(a, c_t) = \hat{\Phi}_t(a)\bar{\boldsymbol{\beta}}$. Besides,

$$\|\bar{\boldsymbol{\beta}}\|_1 \leq \frac{l\sqrt{d}}{\sqrt{\lambda_0(\{\Phi_a\})} - \sqrt{d}r} = 2\delta. \tag{84}$$

Expanding $\hat{\phi}_t^{(i)}(a)$, we have

$$\mathbf{x}(a, c_t) = \sum_{i=1}^{d} \frac{\sum_{\tau \in \mathcal{T}_t(a, \bar{\mathcal{C}}_a^{(i)})} \mathbf{x}(a, c_\tau)}{N_t(a, \bar{\mathcal{C}}_a^{(i)})} \bar{\boldsymbol{\beta}}[i], \tag{85}$$

i.e., $\mathbf{x}(a, c_t)$ can be expressed as a linear combination of the feature vectors $\{\mathbf{x}(a, c_\tau)\}$ for $\tau \in \cup_i \mathcal{T}_t(a, \bar{\mathcal{C}}_a^{(i)})$, where the corresponding coefficients are $\bar{\boldsymbol{\beta}}[i]/N_t(a, \bar{\mathcal{C}}_a^{(i)})$.

Thus, according to Proposition 1, we have

$$\hat{\sigma}_t(a)^2 \leq \sum_{i=1}^{d} \frac{\bar{\boldsymbol{\beta}}[i]^2}{N(a, \bar{\mathcal{C}}_a^{(i)})} \leq \frac{\sum_{i=1}^{d} \bar{\boldsymbol{\beta}}[i]^2}{\frac{tp}{16} - \frac{8K\alpha_t^2}{\Delta^2} d\log\left(\frac{d+t}{d}\right)} \leq \frac{\|\bar{\boldsymbol{\beta}}\|_1^2}{\frac{tp}{16} - \frac{8K\alpha_t^2}{\Delta^2} d\log\left(\frac{d+t}{d}\right)} \tag{86}$$

$$\leq \frac{4\delta^2}{\frac{tp}{16} - \frac{8K\alpha_t^2}{\Delta^2} d\log\left(\frac{d+t}{d}\right)}, \tag{87}$$

where (86) follows from Lemma 10, and (87) follows from (84). ■

Theorem 4 *Let*

$$t_3 = \max\left\{\frac{1728(2+d)(\delta^2 + 2Kd)}{\Delta^2 p}, 10\right\}, \quad t_4 = \max\left\{t_3 \log^2 t_3, \exp\left(\frac{12l^2 s^2}{2+d}\right)\right\}.$$

Then, under Algorithm 1,

$$\mathbb{E}[R(\mathcal{D}_T)] \leq t_4 M = O\left(\frac{d(\delta^2 + 2Kd)}{\Delta^2 p} \log^2\left(\frac{d(\delta^2 + 2Kd)}{\Delta^2 p}\right)\right).$$

Proof. First, we note that

$$\frac{t_4}{(\log t_4)^2} \geq \frac{t_3 \log^2 t_3}{(\log t_3 + 2\log\log t_3)^2} > \frac{t_3 \log^2 t_3}{9 \log^2 t_3} = \frac{t_3}{9}. \tag{88}$$

Thus, for any $t \geq t_4$, we have

$$\frac{t}{(\log t)^2} \geq \frac{t_3}{9} = \frac{192(2+d)(\delta^2 + 2Kd)}{\Delta^2 p}, \tag{89}$$

which is equivalent to

$$t \geq \frac{192(2+d)(\delta^2 + 2Kd)}{\Delta^2 p} \log^2 t. \tag{90}$$

According to (57), $3(2+d)\log t \geq \alpha_t^2$ for $t > t_4$. Thus, (90) can be further bounded as

$$t \geq \frac{64\alpha_t^2}{\Delta^2 p}(\delta^2 + 2Kd)\log t \tag{91}$$

$$\geq \frac{64\alpha_t^2}{\Delta^2 p}\left(\delta^2 + 2Kd \log \frac{d+t}{d}\right), \tag{92}$$

where the last inequality follows from the fact that when $d > 1$, $\log t \geq \log \frac{d+t}{d}$. Rearranging the terms, we have

$$\Delta^2 > \frac{4\alpha_t^2 \delta^2}{\frac{tp}{16} - \frac{8K\alpha_t^2}{\Delta^2} d \log\left(\frac{d+t}{d}\right)} \geq 4\alpha_t^2 \hat{\sigma}_t(a)^2, \quad \forall a \in [K], \tag{93}$$

where the last inequality follows from Lemma 11.

Thus, $\mathcal{D}_T$ can only include time slots $t < t_4$. The bound on $\mathbb{E}[R(\mathcal{D}_T)]$ then follows. ■

D.4 Put Everything Together

After obtaining bounds on the expected regret over $\mathcal{A}_T$, $\mathcal{B}_T$, $\mathcal{C}_T$ and $\mathcal{D}_T$, we are ready to obtain the result in Theorem 2. We have

$$\mathbb{E}[R_T] \leq \mathbb{E}[R(\mathcal{A}_T)] + \mathbb{E}[R(\mathcal{B}_T)] + \mathbb{E}[R(\mathcal{C}_T)] + \mathbb{E}[R(\mathcal{D}_T)]$$

$$\leq \frac{8MKd}{p^2} + \left(\frac{8}{p} + 1\right)M(2 + 2.5K) + t_4 M \tag{94}$$

$$= O\left(\frac{Kd}{p^2} + \frac{d(2\delta^2 + Kd)}{\Delta^2 p}\log^2\left(\frac{d(2\delta^2 + Kd)}{\Delta^2 p}\right)\right). \tag{95}$$