Supplementary Material: Stochastic Linear Contextual Bandits with Diverse Contexts

Weiqiang Wu Jing Yang Cong Shen
London Stock Exchange The Pennsylvania State University University of Virginia

A Proof of Proposition 1

For the constrained convex optimization problem in (6), the corresponding Lagrangian can be formulated as
\[
L(\beta, \lambda) = \beta^T N_t^{-1}(a)\beta + (x(a, c_t) - X_t(a)\beta)^T \lambda,
\]
where $\lambda \in \mathbb{R}^d$ is the Lagrangian multiplier vector.

Taking derivative with respect to $\beta$, we have
\[
N_t^{-1}(a)\beta - X_t^T(a)\lambda = 0,
\]
which implies that
\[
\beta = N_t(a)X_t^T(a)\lambda.
\]

Then, to satisfy the first constraint in (6), we have
\[
x(a, c_t) = X_t(a)\beta = X_t(a)N_t(a)X_t^T(a)\lambda,
\]
which implies that
\[
\lambda = [X_t(a)N_t(a)X_t^T(a)]^{-1} x(a, c_t) := V_t^{-1}(a)x(a, c_t).
\]

Note that the definitions of $X_t(a)$ and $N_t(a)$ ensure that $V_t(a)$ is positive definite and invertible for every $t$.

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., $\theta_a$) 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 $\theta_a$ 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
\[
\hat{r}_t(a) = s_t^T(a)X_t^T(a)\lambda = s_t^T(a)X_t^T(a)V_t^{-1}(a)x(a, c_t) := \hat{\theta}_t^T(a)x(a, c_t),
\]
where \( \hat{\theta}_t(a) := V_t^{-1}(a)X_t(a)s_t(a) \). We can verify that this is exactly the estimate of \( \theta(a) \) obtained by applying the ridge regression with penalty factor \( \lambda^2 \) to the historical data \( \{(x_t, c_t), y_t\}_{t=1}^{t-1} \).

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

\[
\hat{\sigma}_t(a) = \sqrt{\lambda^T X_t(a)N_t(a)X_t^T(a)\lambda} = \|x(a, c_t)\|V_t^{-1}(a),
\]

where we follow the convention to denote \( xVx^T \) as \( \|x\|_V^2 \).

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 \( V_t(a) \), which can be equivalently computed by summing \( x(a, c_t)x^T(a, c_t) \) 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

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

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

\[
\mathbb{E}[R(\mathcal{A}_T)] \leq M \sum_{k=2}^{\lfloor \log_2 T \rfloor} \sum_{t \in F_k} \mathbb{E}[1\{t \in \mathcal{A}_T\}] \leq M \sum_{k=1}^{\lfloor \log_2 T \rfloor} \sum_{t \in F_{k+1}} \sum_{a, c \in \mathcal{C}_a} \mathbb{P}\left[ N_{F_k}(c) \leq \frac{2^{k-1}}{2n} \right]
\]

\[
\leq MKd \sum_{k=1}^{\lfloor \log_2 T \rfloor} \sum_{t \in F_{k+1}} \exp \left( -\frac{2^{k-1}}{2n^2} \right) \leq Mdn \sum_{t=2}^{\infty} \exp \left( -\frac{t}{8n^2} \right) \tag{17}
\]

\[
\leq MKd \int_{0}^{\infty} \exp \left( -\frac{t}{8n^2} \right) dt = 8MKdn^2, \tag{18}
\]

where (17) follows from (16).

**C.2 Bound the Regret over \( \mathcal{B}_T \)**

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

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

\[
\bar{s}_t(a) = [0_{n_x}, -I_1^T \theta(a), \ldots, -I_d^T \theta(a)]^T. \tag{20}
\]

Intuitively, \( \bar{s}_t(a) \) corresponds to the accumulated noise in the observations when arm \( a \) is pulled under different contexts up to time \( t \), and \( \bar{s}_t(a) \) 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

\[
\hat{r}_t(a) - r(a, c_t) = \bar{s}_t^T(a)N_t^{-1}(a)\hat{\sigma}_t(a) - \theta^T(a)x(a, c_t)
\]
Proof. We then adopt the Laplace method (Lattimore and Szepesvári, 2019) to bound we have

\[ X = (s_t(a) - N_t(a)X_t(a)\theta(a))\hat{\theta}_t(a) \]

:= \left( \hat{s}_t(a) + \bar{s}_t(a) \right)\hat{\theta}_t(a), \tag{22} \]

where (21) is due to the fact that \( x(a, c_t) = X_t(a)\beta_t(a) \) according to Proposition 1, and the \( \lambda \) 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 \( \bar{s}_t(a) \) and \( \hat{s}_t(a) \) 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,

\[ \hat{s}_t(a) = \sqrt{\lambda^T X_t(a) N_t(a) X_t^T(a)} \lambda = \| N_t^{1/2}(a) X_t(a) \lambda \|_2. \]

Thus,

\[ |\hat{s}_t^T(a) X_t(a)\lambda| = |\hat{s}_t^T(a) N_t^{1/2}(a) X_t^T(a)\lambda| \]

\[ \leq \| \hat{s}_t^T(a) \|_2 \cdot \| N_t^{1/2}(a) X_t^T(a) \lambda \|_2 \]

\[ = \| \theta(a) \|_2 \sigma_t(a) \leq t \sigma_t(a), \tag{25} \]

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

Before we proceed to bound \( \hat{s}_t^T(a) X_t^T(a) \lambda \), we first introduce the following notations. Recall that \( V_t(a) := X_t(a)N_t(a)X_t^T(a) \). Let \( V_t^{1/2}(a) \) be its square root, i.e., \( V_t^{1/2}(a)V_t^{1/2}(a) = V_t(a) \). Let \( \bar{V}_t(a) := \sum_{c=1}^n N_t(a, c)x(a, c)x^T(a, c) \). Then, \( \bar{V}_t(a) = \bar{V}_t(a) + V_0 \). We have

\[ |\hat{s}_t^T(a) X_t^T(a)\lambda| = |\hat{s}_t^T(a) X_t^T(a)V_t^{-1/2}(a)V_t^{1/2}(a)\lambda| \]

\[ \leq |\hat{s}_t^T(a) X_t^T(a)| \| V_t^{-1/2}(a) \| \| V_t^{1/2}(a) \lambda \|_2 \]

\[ = |\hat{s}_t^T(a) X_t^T(a)| \| V_t^{-1/2}(a) \hat{\sigma}_t(a). \tag{26} \]

We then adopt the Laplace method (Lattimore and Szepesvári, 2019) to bound \( |\hat{s}_t^T(a) X_t^T(a)| \| V_t^{-1/2}(a) \) as follows.

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

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

\[ \mathbb{E}[M_t(u) \mid \mathcal{F}_{t-1}] \]

\[ = \mathbb{E} \left[ \exp \left( \hat{s}_t^T(a) X_t^T(a)u - \frac{1}{2} u^T \bar{V}_t(a)u \right) \mid \mathcal{F}_{t-1} \right] \]

\[ = \mathbb{E} \left[ \exp \left( (\hat{s}_t(a) - \hat{s}_{t-1}(a)) X_t(a)u - \frac{1}{2} u^T \left( \bar{V}_t(a) - \bar{V}_{t-1}(a) \right) u \right) \right] M_{t-1}(u) \]

\[ \tag{28} \]

\[ \tag{29} \]

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

\[ \mathbb{E} \left[ \exp \left( (\hat{s}_t(a) - \hat{s}_{t-1}(a)) X_t(a)u - \frac{1}{2} u^T \left( \bar{V}_t(a) - \bar{V}_{t-1}(a) \right) u \right) \right] = 1. \]

\[ a_t \neq a \]

\[ \tag{30} \]

If \( a_t = a, c_t = c \),

\[ \mathbb{E} \left[ \exp \left( (\hat{s}_t(a) - \hat{s}_{t-1}(a)) X_t(a)u - \frac{1}{2} u^T \left( \bar{V}_t(a) - \bar{V}_{t-1}(a) \right) u \right) \right] = \mathbb{E} \left[ \exp \left( \eta_t e^T \bar{x}(a, c)x^T(a, c)u \right) \right] \]

\[ a_t = a, c_t = c \]

\[ \tag{31} \]

\[ \tag{32} \]
= \mathbb{E} \left[ \exp \left( \eta_t \mathbf{x}^\top (a,c) \mathbf{u} \right) \right] \cdot \exp \left( -\frac{1}{2} \mathbf{u}^\top \mathbf{x} (a,c) \mathbf{x}^\top (a,c) \mathbf{u} \right) \\
\leq \exp \left( \frac{1}{2} \left( \mathbf{x}^\top (a,c) \mathbf{u} \right)^2 \right) \cdot \exp \left( -\frac{1}{2} \mathbf{u}^\top \mathbf{x} (a,c) \mathbf{x}^\top (a,c) \mathbf{u} \right) = 1 \tag{34} \\
\end{align}

where the last inequality follows from Assumption 1.3 that \( \eta_t \) 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 \( \mathbf{u} \), \( M_t \) is a super-martingale as well. Thus, \( \mathbb{E}[M_t] \leq M_0 = 1. \) \( \blacksquare \)

**Lemma 2** Under Algorithm 1,

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

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

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

Then,

\[
\begin{align*}
M_t &= \int_{\mathbb{R}^d} M_t(\mathbf{u}) h(\mathbf{u}) \, d\mathbf{u} \\
&= \frac{1}{\sqrt{(2\pi)^d \det \mathbf{V}_0}} \int_{\mathbb{R}^d} \exp \left( \hat{s}_t^\top (a) \mathbf{X}_t^\top (a) \mathbf{u} - \frac{1}{2} \mathbf{u}^\top \mathbf{V}_t(a) \mathbf{u} - \frac{1}{2} \mathbf{u}^\top \mathbf{V}_0 \mathbf{u} \right) \, d\mathbf{u} \\
&= \frac{1}{\sqrt{(2\pi)^d \det \mathbf{V}_0}} \int_{\mathbb{R}^d} \exp \left( \hat{s}_t^\top (a) \mathbf{X}_t^\top (a) \mathbf{u} - \frac{1}{2} \mathbf{u}^\top \mathbf{V}_t(a) \mathbf{u} \right) \, d\mathbf{u} \\
&= \frac{1}{\sqrt{(2\pi)^d \det \mathbf{V}_0}} \int_{\mathbb{R}^d} \exp \left( \frac{1}{2} \| \hat{s}_t^\top (a) \mathbf{X}_t^\top (a) \|^2_{\mathbf{V}_t^{-1} (a)} - \frac{1}{2} ||\mathbf{u}^\top \mathbf{V}_t(a) - \hat{s}_t^\top (a) \mathbf{X}_t^\top (a)\|^2_{\mathbf{V}_t^{-1} (a)} \right) \, d\mathbf{u} \\
&= \frac{\det \mathbf{V}_t^{-1} (a)}{\det \mathbf{V}_0} \exp \left( \frac{1}{2} \| \hat{s}_t^\top (a) \mathbf{X}_t^\top (a) \|^2_{\mathbf{V}_t^{-1} (a)} \right) \\
&= \exp \left( \frac{1}{2} \| \hat{s}_t^\top (a) \mathbf{X}_t^\top (a) \|^2_{\mathbf{V}_t^{-1} (a)} + \frac{1}{2} \log \frac{\det \mathbf{V}_0}{\det \mathbf{V}_t(a)} \right). \\
\end{align*}
\]

Therefore, according to Lemma 1, we have

\[ \mathbb{P} \left[ \| \hat{s}_t^\top (a) \mathbf{X}_t^\top (a) \| \mathbf{V}_t^{-1} (a) \geq \sqrt{2u + \log \frac{\det \mathbf{V}_t(a)}{\det \mathbf{V}_0}} \right] \]

\[ = \mathbb{P} \left[ \frac{1}{2} \| \hat{s}_t^\top (a) \mathbf{X}_t^\top (a) \|^2_{\mathbf{V}_t^{-1} (a)} + \frac{1}{2} \log \frac{\det \mathbf{V}_0}{\det \mathbf{V}_t(a)} \geq u \right] \]
where (43) follows from the definition of \(\hat{M}_t\). The definitions of \(\hat{M}_t\) indicates that \(\det \hat{V}_0 = t^{2d}\). For \(\hat{V}_t(a)\), we note that for any \(y \in \mathbb{R}^d\), we have
\[
y^\top \hat{V}_t(a)y = \sum_{c=1}^{n} N_t(a,c)y^\top x(a,c)x^\top(a,c)y + t^2 y^\top y \\
\leq t^2 \|y\|_2^2 + t^2 \|y\|_2^2 = (t + 1)t^2 \|y\|_2^2,
\]
where the inequality follows from Assumption 1.1.

Eqn. (37) indicates that the maximum eigenvalue of \(\hat{V}_t(a)\) is upper bound by \((t + 1)t^2\). Therefore, we have \(\det \hat{V}_t(a) \leq (t^2 + t^2)^d\), which implies that
\[
\log \frac{\det \hat{V}_t(a)}{\det \hat{V}_0} \leq d \log(1 + t).
\]

Combining (38) with (22)(25)(27) and Lemma 2, we have
\[
P \left[ s_t(a) - r(a,c) \geq \left( l\alpha + \sqrt{2\alpha + d \log(1 + t)} \right) \hat{\alpha}_t(a) \right] \\
\leq P \left[ \|s_t(a)X^\top(a)\|_{V_t^{-1}(a)} \hat{\alpha}_t(a) \geq \sqrt{2\alpha + d \log(1 + t)} \hat{\alpha}_t(a) \right] \\
\leq P \left[ \|s_t(a)X^\top(a)\|_{V_t^{-1}(a)} \geq \sqrt{2\alpha + d \log \frac{\det \hat{V}_t(a)}{\det \hat{V}_0}} \right] \leq e^{-u}.
\]

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
\[
P[|s_t(a) - r(a,c)| \geq \alpha_t \hat{\alpha}_t(a)] \leq \frac{1}{f(t)}.
\]
Thus,
\[
\mathbb{E}[|B_T|] \leq 2 + K \sum_{t=3}^{\infty} \frac{1}{f(t)} \leq 2 + 2.5K, \quad \mathbb{E}[R(B_T)] \leq M\mathbb{E}[|B_T|] \leq M(2 + 2.5K).
\]

C.3 Bound the Regret over \(C_T\)

Recall \(B_k := |B_T \cap F_k|\), i.e., the number of bad estimates in frame \(k\). Then, according to Markov’s inequality, we have
\[
P \left[ B_k \geq \frac{2^{k-1}}{4n} \right] \leq \frac{\mathbb{E}[B_k]4n}{2^{k-1}}.
\]

The definitions of \(B_k\) and \(B_T\) also imply that \(\sum_{k=1}^{\lfloor \log_2 T \rfloor} \mathbb{E}[B_k] = \mathbb{E}[B_T]\). Therefore,
\[
\mathbb{E}[R(C_T)] \leq M\mathbb{E}[|C_T|] = M \sum_{k=1}^{\lfloor \log_2 T \rfloor} |F_{k+1}| \cdot P \left[ B_k \geq \frac{2^{k-1}}{4n} \right] \\
\leq M \sum_{k=1}^{\lfloor \log_2 T \rfloor} 2^k \frac{\mathbb{E}[B_k]4n}{2^{k-1}} \leq 8nM\mathbb{E}[|B_T|] \leq 8nM(2 + 2.5K),
\]
where (43) follows from the definition of \(C_T\), 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 $a$, $c \notin \mathcal{C}_a$, $\bar{N}_t(a,c) \leq \frac{4\alpha_\tau^2}{\Delta^2}$ for all $t$.

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

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

(45)

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

$$|\hat{r}_t(a_i) - r(a_i, c_i)| \leq \alpha_t \hat{\sigma}_t(a_i), \quad |\hat{r}_t(a_i^*) - r(a_i^*, c_i)| \leq \alpha_t \hat{\sigma}_t(a_i^*).$$

(46)

Putting (45)(46) together, we have

$$r(a_i, c_i) + 2\alpha_t \hat{\sigma}_t(a_i) \geq \hat{r}_t(a_i) + \alpha_t \hat{\sigma}_t(a_i) \geq \hat{r}_t(a_i^*) + \alpha_t \hat{\sigma}_t(a_i^*) \geq r(a_i^*, c_i).$$

(47)

Therefore,

$$\Delta \leq r(a_i^*, c_i) - r(a_i, c_i) \leq 2\alpha_t \hat{\sigma}_t(a_i) = 2\alpha_t \sqrt{\beta^T_t(a_i) N^{-1}_t(a_i) \beta_t(a_i)}.$$ 

(48)

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

$$\beta^T_t(a_i) N^{-1}_t(a_i) \beta_t(a_i) \leq \frac{1}{N_t(a_i, c_i)} \leq \frac{1}{\bar{N}_t(a_i, c_i)}.$$ 

(49)

Combining (48) and (49), we have

$$\bar{N}_t(a_i, c_i) \leq \frac{1}{\beta^T_t(a_i) N^{-1}_t(a_i) \beta_t(a_i)} \leq \frac{4\alpha_\tau^2}{\Delta^2}. $$ 

(50)

When $N_t(a_i, c_i) = 0$, we must have $\bar{N}_t(a_i, c_i) = 0$, 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 $\alpha_t$ 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 $N_t(a,c) \geq \frac{t}{16n} - \frac{8K\alpha_\tau^2}{\Delta^2}$.

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

$$N_t(a,c) \geq N_{F_k}(a,c) = N_{F_k}(b,c) - \sum_{b \neq a} N_{F_k}(b,c) \geq \frac{2^{k-1}}{2n} - \left[B_k \sum_{b \neq a} \bar{N}_{\tau^*}(b,c)\right] $$

(51)

$$ \geq \frac{2^{k-1} - 2^{k-1}}{2n} - K \frac{4\alpha_\tau^2}{\Delta^2} \geq \frac{t}{16n} - \frac{4K\alpha_\tau^2}{\Delta^2},$$

(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 \( x \in \mathbb{R}^d \), \( ||x||_2 \leq l \), we can express it as \( x = \Phi \beta \), where \( \beta \in \mathbb{R}^d \), \( ||\beta||_1 \leq \frac{l \sqrt{d}}{\sqrt{\lambda_{\min}(\Phi^T \Phi)}} \).

Proof. Since

\[
l^2 \geq ||x||^2 = \beta^T \Phi^T \Phi \beta \geq \lambda_{\min}(\Phi^T \Phi) ||\beta||_1^2 \geq \frac{\lambda_{\min}(\Phi^T \Phi)||\beta||_1^2}{d},
\]

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

Lemma 6 For any arm \( a \in [K] \), any time slot \( t \in D_T \), we must have \( \beta_t^T (a) N_t^{-1}(a) \beta_t(a) \leq \frac{\delta^2}{16 n \Delta^2} \), where \( \delta := l \sqrt{d/\lambda_0} \).

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

\[
x(a,c) = X_t(a) \bar{\beta}, \quad \bar{\beta}[c] = 0, \text{ for } c \notin \mathcal{C}_a.
\]

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

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

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

\[
\beta_t^T (a) N_t^{-1}(a) \beta_t(a) \leq \bar{\beta}^T (a,c_t) N_t^{-1}(a) \bar{\beta}(a,c_t) \leq \frac{\|\bar{\beta}(a,c_t)\|_1^2}{t_{16n} - \frac{4K\alpha_t^2}{\Delta^2}} \leq \frac{\delta^2}{t_{16n} - \frac{4K\alpha_t^2}{\Delta^2}},
\]

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 \( \mathbb{E}[R(D_T)] \).

Theorem 3 Let

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

Then, under Algorithm 1,

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

Proof. For any \( t \geq t_2 \), we have

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

which implies that

\[
l \leq \sqrt{2 + d} (\sqrt{3} - \sqrt{2}) \sqrt{\log t}.
\]

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

\[
\alpha_t := ls + \sqrt{(2 + d) \log f(t)} \leq \sqrt{3(2 + d) \log t}.
\]
We point out that the

Theorem 1. We have

where the last inequality follows from Lemma 6.

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

C.5 Put Everything Together

After obtaining bounds on the expected regret over \( A_T, B_T, C_T \) and \( D_T \), we are ready to prove our main result in Theorem 1. We have

\[
\mathbb{E}[R_T] \leq \mathbb{E}[R(A_T)] + \mathbb{E}[R(B_T)] + \mathbb{E}[R(C_T)] + \mathbb{E}[R(D_T)] \\
\leq 8MKdn^2 + (8n + 1)M(2 + 2.5K) + t_2M \\
= O \left( Kdn^2 + \frac{dn(K + \delta^2)}{\Delta^2} \log \frac{dn(K + \delta^2)}{\Delta^2} \right).
\]

We point out that the \( O \left( \exp \left( \frac{12a^2 \Delta^2}{8 + \delta^2} \right) \right) \) term from \( t_2 \) is dropped in (62), since it mainly depends on the bounds on \( \|\theta(a)\|_2 \) and \( \|x(a, c)\|_2 \), 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 \( t_2 \) ball centered at \( \phi_i \) with radius \( r < \sqrt{\lambda_{\text{min}}(\Phi^T \Phi)/d} \), i.e., \( B(\phi_i, r) := \{x \in \mathbb{R}^d \mid \|x - \phi_i\|_2 \leq r \} \). 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, \( \lambda_{\text{min}}(\hat{\Phi}^T \hat{\Phi}) \geq (\sqrt{\lambda_{\text{min}}(\Phi^T \Phi)} - \sqrt{dr})^2 \).

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 \( j \)th row. Then, for any \( \beta \in \mathbb{R}^d \),

\[
\|\hat{\Phi} \beta\|_2 = \|\Phi \beta + \Gamma \beta\|_2 \geq \|\Phi \beta\|_2 - \|\Gamma \beta\|_2 \geq \sqrt{\lambda_{\text{min}}(\Phi^T \Phi)} \|\beta\|_2 - \sqrt{\sum_j |\Gamma(j) \beta|^2}.
\]
We have the following lemma analogue to Lemma 3.

\[ \geq \sqrt{\lambda_{\min}(\Phi^T\Phi)} \|\beta\|_2 - \sqrt{\sum_j \|\Gamma(j)\|_2^2 \|\beta\|_2^2} \]

\[ = \sqrt{\lambda_{\min}(\Phi^T\Phi)} \|\beta\|_2 - \|\beta\|_2 \sqrt{\sum_{i=1}^d \gamma_i^2 \gamma_i^2} \]

\[ \geq \left( \sqrt{\lambda_{\min}(\Phi^T\Phi)} - \sqrt{d}r \right) \|\beta\|_2, \]

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 \(B(\phi, r)\).

Thus, the eigenvalues of \(\Phi^T\Phi\) are lower bounded by \(\sqrt{\lambda_{\min}(\Phi^T\Phi)} - \sqrt{d}r > 0\). \(\blacksquare\)

**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

\[ \Pr\left[ N_{\mathcal{A}_T}(\mathcal{C}^{(i)}_a) \leq \frac{p}{2} \cdot 2^{k-1} \right] \leq \exp \left( -\frac{p^2 2^{k-2}}{8} \right). \]

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

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

### D.2 Bound the Regret over \(\mathcal{C}_T\)

According to Markov’s inequality, we have

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

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

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

### 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 in analysis afterwards.

**Lemma 8 (Elliptical Potential (Lattimore and Szepesvári, 2019))** Let \(V_0\) be positive definite and \(V_t = V_{t-1} + x_t x_t^T\), where \(x_1, \ldots, x_n \in \mathbb{R}^d\) is a sequence of vectors with \(\|x_i\|_2 \leq l < \infty\) for all \(t\). Then,

\[ \sum_{t=1}^{n} (1 \wedge \|x_t\|^2_{V_{t-1}}) \leq 2\log \left( \frac{\det V_n}{\det V_0} \right) \leq 2d\log \left( \frac{\text{trace} V_0 + nl^2}{d \det^{1/d} V_0} \right). \]

where \(x \wedge y = \min\{x, y\}\).

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

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

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

We have the following lemma analogue to Lemma 3.
Lemma 9 For any \( a, b \neq a \), \( N_t(a, \tilde{C}_b^{(i)}) \leq \frac{8a^2}{2d} \log \left( \frac{d+t}{d} \right) \) for all \( t \).

Proof. First, following steps similar to the proof of Lemma 3, for any \( \tau \in T_t(a, \mathcal{C}_b^{(i)}) \), we have
\[
\Delta \leq r(a_\tau^*, c_\tau) - r(a, c_\tau) \leq 2\alpha_\tau \tilde{\sigma}_\tau(a). \tag{72}
\]
Next, we consider the solution to the following optimization problem, denoted as \( \hat{\beta}_t(a) \):
\[
\min_{\beta \in \mathbb{R}^{n_t+d}} \beta^T N_t^{-1}(a) \beta, \quad \text{s.t.} \quad x(a, c_i) = X_t(a) \beta, \quad \beta[c] = 0 \text{ if } c \notin C_b^{(i)} \cup C_0. \tag{73}
\]
Compared with the optimization problem in (6), we have one additional constraint, i.e., we only restrict to the contexts in \( C_b^{(i)} \) and \( C_0 \). The inclusion of the dummy contexts \( C_0 \) 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.,
\[
\tilde{\sigma}_t(a) := \sqrt{\beta_t^T(a) N_t^{-1}(a) \beta_t(a)} \geq \tilde{\sigma}_t(a), \quad \forall t. \tag{74}
\]
Note that
\[
\Delta \leq 2\alpha_\tau (1 \land \tilde{\sigma}_\tau(a)), \quad \tau \in T_t(a, \mathcal{C}_b^{(i)}). \tag{75}
\]
Combining (74)(75) with (72), we have
\[
\Delta \leq 2\alpha_\tau (1 \land \tilde{\sigma}_\tau(a)), \quad \tau \in T_t(a, \mathcal{C}_b^{(i)}). \tag{76}
\]
Summing over all \( \tau \in T_t(a, \mathcal{C}_b^{(i)}) \), we have
\[
\tilde{N}_t(a, \mathcal{C}_b^{(i)}) \Delta \leq \sum_{\tau \in T_t(a, \mathcal{C}_b^{(i)})} 2\alpha_\tau (1 \land \tilde{\sigma}_\tau(a)) \leq 2\alpha_t \left( \frac{\sum_{\tau \in T_t(a, \mathcal{C}_b^{(i)})} (1 \land \tilde{\sigma}_\tau(a))^2}{\sum_{\tau \in T_t(a, \mathcal{C}_b^{(i)})} 1 \land \tilde{\sigma}_\tau(a)^2} \right), \tag{77}
\]
where (77) follows from the monotonicity of \( \alpha_t \) and the Cauchy-Schwarz inequality.

Consider the sequence of feature vectors \( \{x(a, c_\tau)\}_{\tau \in T_t(a, \mathcal{C}_b^{(i)})} \). Label the times indices in \( T_t(a, \mathcal{C}_b^{(i)}) \) as \( \tau_1, \tau_2, \ldots \). Let \( \tilde{V}_0 = t^2 I, \tilde{V}_\tau = \tilde{V}_{\tau-1} + x(a, c_\tau) x(a, c_\tau)^T \). Then, similar to (15), we have \( \tilde{\sigma}_\tau(a) = \|x(a, c_\tau)\| \tilde{V}_{\tau-1}^{-1} \). Following Lemma 8, we have
\[
\sum_{\tau \in T_t(a, \mathcal{C}_b^{(i)})} (1 \land \tilde{\sigma}_\tau(a)^2) \leq 2d \log \left( \frac{d^2 + \tilde{N}_t(a, \mathcal{C}_b^{(i)}) t^2}{d^2} \right) \leq 2d \log \left( \frac{d+t}{d} \right). \tag{78}
\]
Plugging (78) into (77) and rearranging the terms, we have \( \tilde{N}_t(a, \mathcal{C}_b^{(i)}) \leq \frac{8a^2}{2d} \log \left( \frac{d+t}{d} \right) \) for all \( t \).

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

Proof. Assume \( t \) lies in the \((k+1)\)th time frame. Then, based on the definition of \( N_t(a, \bar{C}_a^{(i)}) \), we have
\[
N_t(a, \bar{C}_a^{(i)}) \geq N_{F_k}(a, \bar{C}_a^{(i)}) \geq N_{F_k}(\bar{C}_a^{(i)}) - \sum_{b:a \neq b} N_{F_k}(b, \bar{C}_a^{(i)}) \tag{79}
\]
Then, under Algorithm 1,\( C_t \leq j \) follows from Lemma 10, and \( (87) \) follows from (84).

Thus, according to Proposition 1, we have
\[
\text{(79) follows from the assumption that}
\]

\[
\text{Lemma 11 For any arm } a \in [K], \text{ and any time slot } t \in \mathcal{D}_T, \text{ we have}
\]
\[
\hat{\sigma}_t(a)^2 \leq \frac{4 \delta^2}{tp - 8K\alpha_t^2d\log \left( \frac{d + t}{d} \right)},
\]

where \( \delta = \sqrt{d/\lambda_0(\{F_a\})} \).

Proof. Let
\[
\hat{\phi}_t^{(i)}(a) := \frac{\sum_{\tau \in \mathcal{T}_t(a, C_t^{(i)})} x(a, c_{\tau})}{N_t(a, C_t^{(i)})}
\]
be the empirical average of the feature vectors over the time slots before \( t \) when arm \( a \) is pulled under a context in \( C_t^{(i)} \). Since \( B(\phi_t^{(i)}, r) \) is convex, \( \hat{\phi}_t^{(i)}(a) \in B(\phi_t^{(i)}, r) \). Thus, according to Lemma 7, \( \{\hat{\phi}_t^{(i)}(a)\} \), form a valid basis for \( \lambda_a \). Let \( \hat{\Phi}_t(a) \) be the matrix whose columns are \( \hat{\phi}_t^{(i)}(a) \). Then, we can always obtain a vector \( \hat{\beta} \), such that
\[
x(a, c_t) = \sum_{i=1}^d \frac{\sum_{\tau \in \mathcal{T}_t(a, C_t^{(i)})} x(a, c_{\tau})}{N_t(a, C_t^{(i)})} \hat{\beta}[i],
\]
i.e., \( x(a, c_t) \) can be expressed as a linear combination of the feature vectors \( \{x(a, c_{\tau})\} \) for \( \tau \in \cup \mathcal{T}_t(a, C_t^{(i)}) \), where the corresponding coefficients are \( \hat{\beta}[i]/N_t(a, C_t^{(i)}) \).

Thus, according to Proposition 1, we have
\[
\hat{\sigma}_t(a)^2 \leq \sum_{i=1}^d \frac{\hat{\beta}[i]^2}{N_t(a, C_t^{(i)})} \leq \frac{\sum_{i=1}^d \hat{\beta}[i]^2}{tp - 8K\alpha_t^2d\log \left( \frac{d + t}{d} \right)} \leq \frac{\|\hat{\beta}\|^2}{tp - 8K\alpha_t^2d\log \left( \frac{d + t}{d} \right)},
\]

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

Theorem 4 Let
\[
t_3 = \max \left\{ \frac{1728(2 + d)(\delta^2 + 2Kd)}{\Delta^2 p}, 10 \right\}, \quad t_4 = \max \left\{ t_3 \log^2 t_3, \exp \left( \frac{12\delta^2s^2}{2} \right) \right\}.
\]

Then, under Algorithm 1,
\[
\mathbb{E}[R(D_T)] \leq t_4 \mathcal{M} = O \left( \frac{d(\delta^2 + 2Kd)}{\Delta^2 p} \log^2 \left( \frac{d(\delta^2 + 2Kd)}{\Delta^2 p} \right) \right).
\]
Proof. First, we note that
\[
\frac{t_4}{(\log t_4)^2} \geq \frac{t_3 \log^2 t_4}{(\log t_3 + 2 \log \log t_3)^2} > \frac{t_3 \log^2 t_4}{9 \log^2 t_3} = \frac{t_3}{9}
\]
(88)

Thus, for any \( t \geq t_4 \), we have
\[
\frac{t}{(\log t)^2} \geq \frac{t_3}{9} = \frac{192(2 + d)(\delta^2 + 2Kd)}{\Delta^2 p},
\]
(89)

which is equivalent to
\[
t \geq \frac{192(2 + d)(\delta^2 + 2Kd)}{\Delta^2 p} \log^2 t.
\]
(90)

According to (57), \( 3(2 + d) \log t \geq \alpha_i^2 \) for \( t > t_4 \). Thus, (90) can be further bounded as
\[
t \geq \frac{64\alpha_i^2}{\Delta^2 p} (\delta^2 + 2Kd) \log t
\]
(91)
\[
\geq \frac{64\alpha_i^2}{\Delta^2 p} \left( \delta^2 + 2Kd \log \frac{d + t}{d} \right),
\]
(92)

where the last inequality follows from the fact that when \( d > 1 \), \( \log t \geq \log \frac{d+1}{d} \). Rearranging the terms, we have
\[
\Delta^2 > \frac{4\alpha_i^2 \delta^2}{\frac{1}{p} - \frac{8K\alpha_i^2}{\Delta^2 p} d \log \left( \frac{d+1}{d} \right)} \geq 4\alpha_i^2 \hat{\sigma}(a)^2, \quad \forall a \in [K],
\]
(93)

where the last inequality follows from Lemma 11.

Thus, \( D_T \) can only include time slots \( t < t_4 \). The bound on \( \mathbb{E}[R(D_T)] \) then follows. \( \blacksquare \)

D.4 Put Everything Together

After obtaining bounds on the expected regret over \( A_T, B_T, C_T \) and \( D_T \), we are ready to obtain the result in Theorem 2. We have
\[
\mathbb{E}[R_T] \leq \mathbb{E}[R(A_T)] + \mathbb{E}[R(B_T)] + \mathbb{E}[R(C_T)] + \mathbb{E}[R(D_T)]
\]
\[
\leq \frac{8MKd}{p^2} + \left( \frac{8}{p} + 1 \right) M(2 + 2.5K) + t_4 M
\]
(94)
\[
= O \left( \frac{Kd}{p^2} + \frac{d(2\delta^2 + Kd)}{\Delta^2 p} \log^2 \left( \frac{d(2\delta^2 + Kd)}{\Delta^2 p} \right) \right).
\]
(95)