Contextual Multi-armed Bandit Algorithm for Semiparametric Reward Model: Supplementary Material

Gi-Soo Kim [1] **Myunghee Cho Paik** [1]

A. Preliminaries

Lemma A.1. *(Lemma 11 of Abbasi-Yadkori et al., 2011) Let $\{X_t\}_{t=1}^T$ be a sequence in $\mathbb{R}^d$ with $||X_t||_2 \leq 1$, Q a $d \times d$ positive definite matrix with $det(Q) \geq 1$ and $A(t) = \sum_{\tau=1}^{t-1} X_\tau X_\tau^T$. Then, we have*

$$\sum_{t=1}^{T} X_t^T \{Q + A(t)\}^{-1} X_t \leq 2\log\Big(\frac{det(Q + A(T+1))}{det(Q)}\Big).$$

Lemma A.2. *(Lemma 2.1 of Bercu and Touati, 2008) Let x be a square integrable random variable with mean 0 and variance $\sigma^2 > 0$. Then,*

$$\mathbb{E}\Big[\exp\Big(x - \frac{1}{2}x^2 - \frac{1}{2}\sigma^2\Big)\Big] \leq 1.$$

Lemma A.3. *(Lemma 7 of de la Peña et al., 2009) Let $X_\tau \in \mathbb{R}^d$ be $\mathcal{F}_\tau$-measurable for some filtration $\{\mathcal{F}_\tau\}_{\tau=1}^t$, $\mathbb{E}\big[X_\tau|\mathcal{F}_{\tau-1}\big] = 0$, and $||X_\tau||_2 \leq B$ for some constant B, $\tau = 1, \cdots, t$. Let $c_\tau \in \mathbb{R}$ be $\mathcal{F}_\tau$-measurable, $|c_\tau| \leq 1$ and $X_\tau \perp c_\tau|\mathcal{F}_{\tau-1}$. Then for any $\lambda \in \mathbb{R}^d$,*

$$\mathbb{E}\Big[\exp\Big\{\lambda^T \sum_{\tau=1}^{t} X_\tau c_\tau - \frac{1}{2}\lambda^T\Big(\sum_{\tau=1}^{t} X_\tau X_\tau^T + \sum_{\tau=1}^{t}\mathbb{E}\big[X_\tau X_\tau^T|\mathcal{F}_{\tau-1}\big]\Big)\lambda\Big\}\Big] \leq 1.$$

Proof. Taking $x = \lambda^T X_\tau c_\tau$, we have from Lemma A.2,

$$\mathbb{E}\Big[\exp\Big\{\lambda^T X_\tau c_\tau - \frac{1}{2}\lambda^T\Big(c_\tau^2 X_\tau X_\tau^T + \mathbb{E}\big[c_\tau^2 X_\tau X_\tau^T|\mathcal{F}_{\tau-1}\big]\Big)\lambda\Big\}\Big|\mathcal{F}_{\tau-1}\Big] \leq 1.$$

Since $c_\tau^2 \leq 1$ and $X_\tau X_\tau^T$ is positive semi-definite,

$$\mathbb{E}\Big[\exp\Big\{\lambda^T X_\tau c_\tau - \frac{1}{2}\lambda^T\Big(X_\tau X_\tau^T + \mathbb{E}\big[X_\tau X_\tau^T|\mathcal{F}_{\tau-1}\big]\Big)\lambda\Big\}\Big|\mathcal{F}_{\tau-1}\Big] \leq 1.$$

□

Lemma A.4. *(Abramowitz and Stegun, 1964) If $Z \sim \mathcal{N}(m, \sigma^2)$, for any $z \geq 1$,*

$$\frac{1}{2\sqrt{\pi}z}\exp\Big(-\frac{z^2}{2}\Big) \leq \mathbb{P}\big(|Z - m| > z\sigma\big) \leq \frac{1}{\sqrt{\pi}z}\exp\Big(-\frac{z^2}{2}\Big).$$

B. Proof of Theorem 4.2

The proof of Theorem 4.2 follows the proof sketch of Section 4.2.

[1]Department of Statistics, Seoul National University, Seoul, Korea. Correspondence to: Myunghee Cho Paik <myunghee-chopaik@snu.ac.kr>.

B.1. Proof of (14)

Take $c_\tau = \left(\frac{\nu(\tau)+\bar{b}(\tau)^T\mu}{2}\right)$. Since $\mathbb{E}\big[X_\tau|\mathcal{F}_{\tau-1}\big]=0$, $|c_\tau|\le 1$, and $X_\tau \perp c_\tau|\mathcal{F}_{\tau-1}$, we can apply Lemma A.3, i.e., for any $\lambda\in\mathbb{R}^d$,

$$\mathbb{E}\Big[\exp\Big\{\lambda^T\sum_{\tau=1}^{t-1}X_\tau c_\tau - \frac{1}{2}\lambda^T\Big(\sum_{\tau=1}^{t-1}X_\tau X_\tau^T + \sum_{\tau=1}^{t-1}\mathbb{E}[X_\tau X_\tau^T|\mathcal{F}_{\tau-1}]\Big)\lambda\Big\}\Big]\le 1.$$

B.2. Proof of Lemma 4.4

By Lemma A.3, for any $\lambda\in\mathbb{R}^d$,

$$\mathbb{E}\Big[\exp\Big\{\lambda^T\sum_{\tau=1}^{t-1}\frac{1}{\sqrt{2}}Y_\tau - \frac{1}{2}\lambda^T\Big(\frac{1}{2}\sum_{\tau=1}^{t-1}Y_\tau Y_\tau^T + \frac{1}{2}\sum_{\tau=1}^{t-1}\mathbb{E}\big[Y_\tau Y_\tau^T|\mathcal{F}_{\tau-1}\big]\Big)\lambda\Big\}\Big]\le 1.$$

Here,

$$\begin{aligned}\lambda^T Y_\tau Y_\tau^T\lambda &= \lambda^T D(\tau)\mu\mu^T D(\tau)\lambda\\ &= \big\{\big(D(\tau)\lambda\big)^T\mu\big\}^2\\ &\le \mu^T\mu\big(D(\tau)\lambda\big)^T\big(D(\tau)\lambda\big)\ \ (\because \text{Cauchy-Schwarz inequality})\\ &\le \big(D(\tau)\lambda\big)^T\big(D(\tau)\lambda\big) = \lambda^T D(\tau)^2\lambda, \qquad (1)\end{aligned}$$

and

$$\lambda^T\mathbb{E}\big[Y_\tau Y_\tau^T|\mathcal{F}_{\tau-1}\big]\lambda \le \lambda^T\mathbb{E}\big[D(\tau)^2|\mathcal{F}_{\tau-1}\big]\lambda. \qquad (2)$$

Let $L = X_\tau X_\tau^T$ and $K=\mathbb{E}\big[X_\tau X_\tau^T|\mathcal{F}_{\tau-1}\big]$. Then,

$$\begin{aligned}\lambda^T D(\tau)^2\lambda &= \lambda^T\big(L-K\big)^2\lambda\\ &= \lambda^T L^2\lambda + \lambda^T K^2\lambda + 2\lambda^T L(-K)\lambda\\ &\le \lambda^T L^2\lambda + \lambda^T K^2\lambda + 2\sqrt{\lambda^T L^2\lambda\ \lambda^T K^2\lambda}\ \ (\because \text{Cauchy-Schwarz inequality})\\ &\le 2\lambda^T L^2\lambda + 2\lambda^T K^2\lambda. \qquad (3)\end{aligned}$$

Also,

$$\begin{aligned}\mathbb{E}\big[D(\tau)^2|\mathcal{F}_{\tau-1}\big] &= \mathbb{E}\big[(L-K)^2|\mathcal{F}_{\tau-1}\big]\\ &= \mathbb{E}\big[L^2|\mathcal{F}_{\tau-1}\big] - \mathbb{E}\big[L|\mathcal{F}_{\tau-1}\big]K - K\mathbb{E}\big[L|\mathcal{F}_{\tau-1}\big] + K^2\\ &= \mathbb{E}\big[L^2|\mathcal{F}_{\tau-1}\big] - K^2\ \ (\because \mathbb{E}\big[L|\mathcal{F}_{\tau-1}\big]=K)\\ \Rightarrow \lambda^T\mathbb{E}\big[D(\tau)^2|\mathcal{F}_{\tau-1}\big]\lambda &\le 2\lambda^T\mathbb{E}\big[D(\tau)^2|\mathcal{F}_{\tau-1}\big]\lambda\\ &= 2\lambda^T\mathbb{E}\big[L^2|\mathcal{F}_{\tau-1}\big]\lambda - 2\lambda^T K^2\lambda. \qquad (4)\end{aligned}$$

Due to (1), (2), (3) and (4),

$$\begin{aligned}\lambda^T\Big(Y_\tau Y_\tau^T + \mathbb{E}\big[Y_\tau Y_\tau^T|\mathcal{F}_{\tau-1}\big]\Big)\lambda &\le 2\lambda^T\Big(L^2 + \mathbb{E}\big[L^2|\mathcal{F}_{\tau-1}\big]\Big)\lambda\\ &\le 2\lambda^T\Big(X_\tau X_\tau^T + \mathbb{E}\big[X_\tau X_\tau^T|\mathcal{F}_{\tau-1}\big]\Big)\lambda,\end{aligned}$$

where the last inequality is due to $L = X_\tau X_\tau^T$ and $X_\tau^T X_\tau \le 1$. Therefore, for any $\lambda\in\mathbb{R}^d$,

$$\mathbb{E}\Big[\exp\Big\{\lambda^T\sum_{\tau=1}^{t-1}\frac{1}{\sqrt{2}}Y_\tau - \frac{1}{2}\lambda^T\Big(\sum_{\tau=1}^{t-1}X_\tau X_\tau^T + \sum_{\tau=1}^{t-1}\mathbb{E}\big[X_\tau X_\tau^T|\mathcal{F}_{\tau-1}\big]\Big)\lambda\Big\}\Big]$$

$$\leq \mathbb{E}\Big[\exp\Big\{\lambda^T \sum_{\tau=1}^{t-1} \frac{1}{\sqrt{2}} Y_\tau - \frac{1}{2}\lambda^T \Big(\frac{1}{2}\sum_{\tau=1}^{t-1} Y_\tau Y_\tau^T + \frac{1}{2}\sum_{\tau=1}^{t-1} \mathbb{E}\big[Y_\tau Y_\tau^T | \mathcal{F}_{\tau-1}\big]\Big)\lambda\Big\}\Big]$$
$$\leq 1.$$

C. Proof of Theorem 4.1

The proof of Theorem 4.1 follows the lines of Agrawal and Goyal (2013) with some modifications. We present the whole proof.

(a) The first stage is the derivation of a high-probability upper bound of $|(b_i(t) - \bar{b}(t))^T(\hat{\mu}(t) - \mu)|$. This is done in Theorem 4.2, which we restate here for concreteness.

Theorem C.1. *Let the event $E^{\hat{\mu}}(t)$ be defined as follows:*

$$E^{\hat{\mu}}(t) = \big\{\forall i : |\big(b_i(t) - \bar{b}(t)\big)^T(\hat{\mu}(t) - \mu)| \leq l(t)s_{t,i}^c\big\},$$

where $s_{t,i}^c = \sqrt{\big(b_i(t) - \bar{b}(t)\big)^T B(t)^{-1}\big(b_i(t) - \bar{b}(t)\big)}$ and $l(t) = (2R+6)\sqrt{d\log(6t^3/\delta)} + 1$. Then for all $t \geq 1$, for any $0 < \delta < 1$, $\mathbb{P}(E^{\hat{\mu}}(t)) \geq 1 - \frac{\delta}{t^2}$.

(b) We next establish a high-probability upper bound for $|(b_i(t) - \bar{b}(t))^T(\tilde{\mu}(t) - \hat{\mu}(t))|$ in the following Proposition C.2. The proof is a simple extension of Agrawal and Goyal (2013), which uses Lemma A.4 for gaussian random variables.

Proposition C.2. *Let the event $E^{\tilde{\mu}}(t)$ be defined as follows:*

$$E^{\tilde{\mu}}(t) = \big\{\forall i : |\big(b_i(t) - \bar{b}(t)\big)^T(\tilde{\mu}(t) - \hat{\mu}(t))| \leq m(T)s_{t,i}^c\big\},$$

where $m(T) = v\sqrt{4d\log(Td)}$. Then for all $t \geq 0$, $\mathbb{P}(E^{\tilde{\mu}}(t)|\mathcal{F}_{t-1}) \geq 1 - \frac{1}{T^2}$.

Proof. Note that given $\mathcal{F}_{t-1}$, the values of $\big(b_i(t) - \bar{b}(t)\big)$, $B(t)$ and $\hat{\mu}(t)$ are fixed. Then,

$$\begin{aligned}
|b_i^c(t)^T(\tilde{\mu}(t) - \hat{\mu}(t))| &= |b_i^c(t)^T vB(t)^{-1/2}\frac{1}{v}B(t)^{1/2}(\tilde{\mu}(t) - \hat{\mu}(t))| \\
&\leq v\sqrt{b_i^c(t)^T B(t)^{-1} b_i^c(t)}\Big|\Big|\frac{1}{v}B(t)^{1/2}(\tilde{\mu}(t) - \hat{\mu}(t))\Big|\Big|_2 \\
&= vs_{t,i}^c\Big|\Big|\frac{1}{v}B(t)^{1/2}(\tilde{\mu}(t) - \hat{\mu}(t))\Big|\Big|_2 \\
&= vs_{t,i}^c\sqrt{\sum_{j=1}^d ||Z_j(t)||_2^2},
\end{aligned}$$

where $Z_j(t)|\mathcal{F}_{t-1} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and the first inequality is due to Cauchy-Schwarz inequality. Due to Lemma A.4, for fixed j and $z \geq 1$,

$$\mathbb{P}\big(|Z_j(t)| > z \mid \mathcal{F}_{t-1}\big) \leq \frac{1}{\sqrt{\pi}z}\exp\Big(-\frac{\mathrm{z}^2}{2}\Big) \leq \exp\Big(-\frac{\mathrm{z}^2}{2}\Big).$$

Setting $\exp\big(-\mathrm{z}^2/2\big) = \frac{1}{dT^2}$, we have $z = \sqrt{2\log(dT^2)} \leq \sqrt{2\log(d^2T^2)} = \sqrt{4\log(dT)}$. Hence,

$$\mathbb{P}\big(|Z_j(t)| > \sqrt{4\log(dT)} \mid \mathcal{F}_{t-1}\big) \leq \frac{1}{dT^2}$$
$$\Rightarrow \mathbb{P}\big(\forall j : |Z_j(t)| > \sqrt{4\log(dT)} \mid \mathcal{F}_{t-1}\big) \leq \frac{1}{T^2}.$$

Thus, with probability at least $1 - \frac{1}{T^2}$, for all $i = 1, \cdots, N$,

$$|\big(b_i(t) - \bar{b}(t)\big)^T(\tilde{\mu}(t) - \hat{\mu}(t))| \leq vs_{t,i}^c\sqrt{4d\log(dT)} = m(T)s_{t,i}^c.$$

$\square$

(c) Before proceeding, we divide the arms at each time into two groups: saturated and unsaturated arms. Let $g(T) = m(T) + l(T)$. An arm i is saturated at time t if

$$\big(b_i(t) - \bar{b}(t)\big)^T \mu + g(T)s^c_{t,i} < \big(b_{a^*(t)}(t) - \bar{b}(t)\big)^T \mu,$$

and unsaturated otherwise. Note that the optimal arm $a^*(t)$ is unsaturated. Note also that from Stage (a) and Stage (b), $(b_i(t) - \bar{b}(t))^T \mu + g(T)s^c_{t,i}$ is an upper bound of $(b_i(t) - \bar{b}(t))^T \tilde{\mu}(t)$. Hence by definition, the saturated arms are the arms that have quite accurate values of $(b_i(t) - \bar{b}(t))^T \tilde{\mu}(t)$ so that their upper bound is lower than $(b_{a^*(t)}(t) - \bar{b}(t))^T \mu$, enabling the algorithm to distinguish between them and the optimal arm.

(d) Next, we show in Proposition C.3 that the probability of playing saturated arms is bounded by a function of the probability of playing unsaturated arms. The proof is a simple extension of Agrawal and Goyal (2013).

Proposition C.3. *Let $C(t)$ be the set of saturated arms at time t, i.e., $C(t) = \{i : \big(b_i(t) - \bar{b}(t)\big)^T \mu + g(T)s^c_{t,i} < \big(b_{a^*(t)}(t) - \bar{b}(t)\big)^T \mu\}$. Given any filtration $\mathcal{F}_{t-1}$ such that $E^{\hat{\mu}}(t)$ is true,*

$$\mathbb{P}\big(a(t) \in C(t)|\mathcal{F}_{t-1}\big) \leq \frac{1}{p}\mathbb{P}\big(a(t) \notin C(t)|\mathcal{F}_{t-1}\big) + \frac{1}{pT^2},$$

where $p = \frac{1}{4e\sqrt{2}\sqrt{\pi}}$.

Proof. Since the algorithm pulls the arm $\underset{i}{\text{argmax}}\{b_i(t)^T \tilde{\mu}(t)\}$, if $b_{a^*(t)}(t)^T \tilde{\mu}(t) > b_j(t)^T \tilde{\mu}(t)$ for every $j \in C(t)$, then $a(t) \notin C(t)$. Hence,

$$\begin{aligned}\mathbb{P}\big(a(t) \notin C(t)|\mathcal{F}_{t-1}\big) &\geq \mathbb{P}\big(b_{a^*(t)}(t)^T \tilde{\mu}(t) > b_j(t)^T \tilde{\mu}(t),\ \forall j \in C(t)|\mathcal{F}_{t-1}\big)\\ &= \mathbb{P}\big(b^c_{a^*(t)}(t)^T \tilde{\mu}(t) > b^c_j(t)^T \tilde{\mu}(t),\ \forall j \in C(t)|\mathcal{F}_{t-1}\big). \end{aligned} \tag{5}$$

If $E^{\tilde{\mu}}(t)$ is additionally true, for $\forall j \in C(t)$,

$$\begin{aligned} b^c_j(t)^T \tilde{\mu}(t) &\leq b^c_j(t)^T \mu + g(T)s^c_{t,j} \quad (\because E^{\hat{\mu}}(t)\ \&\ E^{\tilde{\mu}}(t))\\ &\leq b^c_{a^*(t)}(t)^T \mu. \quad (\because \text{definition of } C(t)) \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathbb{P}\big(b^c_{a^*(t)}(t)^T \tilde{\mu}(t) > b^c_j(t)^T \tilde{\mu}(t),\ \forall j \in C(t)|\mathcal{F}_{t-1}\big) + \Big(1 - \mathbb{P}\big(E^{\tilde{\mu}}(t)|\mathcal{F}_{t-1}\big)\Big)\\ &\geq \mathbb{P}\big(b^c_{a^*(t)}(t)^T \tilde{\mu}(t) > b^c_{a^*(t)}(t)^T \mu|\mathcal{F}_{t-1}\big). \end{aligned} \tag{6}$$

Given $E^{\hat{\mu}}(t)$, $|b^c_{a^*(t)}(t)^T(\hat{\mu}(t) - \mu)| \leq l(T)s^c_{t,a^*(t)}$. Thus by Lemma A.4,

$$\begin{aligned} (6) &= \mathbb{P}\Big(\frac{b^c_{a^*(t)}(t)^T(\tilde{\mu}(t) - \hat{\mu}(t))}{vs^c_{t,a^*(t)}} > \frac{b^c_{a^*(t)}(t)^T(\mu - \hat{\mu}(t))}{vs^c_{t,a^*(t)}}\Big|\mathcal{F}_{t-1}\Big)\\ &\geq \mathbb{P}\Big(Z(t) > \frac{l(T)}{v}\Big|\mathcal{F}_{t-1}\Big)\\ &\geq \frac{1}{4\sqrt{\pi}z}\exp\Big(-\frac{z^2}{2}\Big) \geq p, \end{aligned} \tag{7}$$

where $Z(t)|\mathcal{F}_{t-1} \sim \mathcal{N}(0, 1)$ and $z = l(T)/v$. Therefore, due to (5), (6), (7) and Proposition C.2,

$$\mathbb{P}\big(a(t) \notin C(t)|\mathcal{F}_{t-1}\big) \geq p - \frac{1}{T^2}.$$

$$\Rightarrow \frac{\mathbb{P}\big(a(t) \in C(t)|\mathcal{F}_{t-1}\big)}{\mathbb{P}\big(a(t) \notin C(t)|\mathcal{F}_{t-1}\big) + \frac{1}{T^2}} \leq \frac{1}{p}.$$

□

(e) Next in Proposition C.4, we use Proposition C.3 and the definition of unsaturated arms to show that the regret can be bounded by a factor of $s^c_{t,a(t)}$ in expectation.

Proposition C.4. *Given any filtration $\mathcal{F}_{t-1}$ such that $E^{\hat{\mu}}(t)$ is true,*

$$\mathbb{E}\big[regret(t)|\mathcal{F}_{t-1}\big] \leq \frac{5g(T)}{p}\mathbb{E}\big[s^c_{t,a(t)}|\mathcal{F}_{t-1}\big] + \frac{3g(T)}{pT^2}.$$

Proof. Let $\tilde{a}(t) = \underset{i \notin C(t)}{\text{argmin}}\, s^c_{t,i}$. This value is determined by $\mathcal{F}_{t-1}$. Under both $E^{\hat{\mu}}(t)$ and $E^{\tilde{\mu}}(t)$,

$$\begin{aligned}
b^c_{a^*(t)}(t)^T\mu &= b^c_{a^*(t)}(t)^T\mu - b^c_{\tilde{a}(t)}(t)^T\mu + b^c_{\tilde{a}(t)}(t)^T\mu \\
&\leq g(T)s^c_{t,\tilde{a}(t)} + b^c_{\tilde{a}(t)}(t)^T\mu \\
&\leq g(T)s^c_{t,\tilde{a}(t)} + b^c_{\tilde{a}(t)}(t)^T\tilde{\mu}(t) + g(T)s^c_{t,\tilde{a}(t)} \\
&\leq 2g(T)s^c_{t,\tilde{a}(t)} + b^c_{a(t)}(t)^T\tilde{\mu}(t) \\
&\leq 2g(T)s^c_{t,\tilde{a}(t)} + b^c_{a(t)}(t)^T\mu + g(T)s^c_{t,a(t)} \\
\Rightarrow regret(t) &\leq 2g(T)s^c_{t,\tilde{a}(t)} + g(T)s^c_{t,a(t)},
\end{aligned}$$

where the first inequality follows from the definition of unsaturated arms, the second and fourth inequalities from $E^{\hat{\mu}}(t)$ and $E^{\tilde{\mu}}(t)$, and the third inequality from the action selection mechanism. Therefore, given $\mathcal{F}_{t-1}$ such that $E^{\hat{\mu}}(t)$ holds,

$$\begin{aligned}
\mathbb{E}\big[regret(t)|\mathcal{F}_{t-1}\big] &\leq 2g(T)s^c_{t,\tilde{a}(t)} + g(T)\mathbb{E}\big[s^c_{t,a(t)}|\mathcal{F}_{t-1}\big] + 1 - \mathbb{P}(E^{\tilde{\mu}}(t)|\mathcal{F}_{t-1}) \\
&\leq 2g(T)s^c_{t,\tilde{a}(t)} + g(T)\mathbb{E}\big[s^c_{t,a(t)}|\mathcal{F}_{t-1}\big] + \frac{1}{T^2}.
\end{aligned} \tag{8}$$

Here,

$$\begin{aligned}
s^c_{t,\tilde{a}(t)} &= s^c_{t,\tilde{a}(t)}\big\{\mathbb{P}(a(t) \in C(t)|\mathcal{F}_{t-1}) + \mathbb{P}(a(t) \notin C(t)|\mathcal{F}_{t-1})\big\} \\
&\leq s^c_{t,\tilde{a}(t)}\Big\{\frac{2}{p}\mathbb{P}(a(t) \notin C(t)|\mathcal{F}_{t-1}) + \frac{1}{pT^2}\Big\} \\
&= \frac{2}{p}\mathbb{E}\big(s^c_{t,\tilde{a}(t)}I\{a(t) \notin C(t)\}\big|\mathcal{F}_{t-1}\big) + \frac{s^c_{t,\tilde{a}(t)}}{pT^2} \\
&\leq \frac{2}{p}\mathbb{E}\big(s^c_{t,a(t)}I\{a(t) \notin C(t)\}\big|\mathcal{F}_{t-1}\big) + \frac{s^c_{t,\tilde{a}(t)}}{pT^2} \\
&\leq \frac{2}{p}\mathbb{E}\big(s^c_{t,a(t)}\big|\mathcal{F}_{t-1}\big) + \frac{1}{pT^2},
\end{aligned}$$

where the first inequality is due to Proposition C.3 and the second inequality is due to the definition of $\tilde{a}(t)$. Combining this result with (8), we have

$$\mathbb{E}\big[regret(t)|\mathcal{F}_{t-1}\big] \leq \frac{5g(T)}{p}\mathbb{E}\big(s^c_{t,a(t)}\big|\mathcal{F}_{t-1}\big) + \frac{3g(T)}{pT^2}.$$

□

(f) Let $M_t = regret(t)I(E^{\hat{\mu}}(t)) - \frac{5g(T)}{p}s^c_{t,a(t)} - \frac{3g(T)}{pT^2}$. Then $|M_t|$ is bounded by $\frac{9g(T)}{p}$. Also, due to Proposition C.4, $\{M_t\}_{t=1}^T$ is a bounded super-martingale difference process with respect to the filtration $\{\mathcal{F}_t\}_{t=1}^T$. Hence by Azuma-Hoeffding's inequality, for any $a \geq 0$,

$$\mathbb{P}\Big(\sum_{t=1}^T M_t \geq a\Big) \leq \exp\Big(-\frac{a^2}{2\sum_{t=1}^T c_t^2}\Big),$$

where $c_t = \frac{9}{p}g(T)$. Setting $\exp\Big(-\frac{a^2}{2\sum_{t=1}^T c_t^2}\Big) = \frac{\delta}{2}$, we have $a = \frac{9}{p}g(T)\sqrt{2T\log\big(\frac{2}{\delta}\big)}$. Thus with probability at least $1-\frac{\delta}{2}$,

$$\sum_{t=1}^{T} regret(t)I(E^{\hat{\mu}}(t)) \leq \frac{5g(T)}{p}\sum_{t=1}^{T} s_{t,a(t)}^c + \frac{3g(T)}{pT} + \frac{9}{p}g(T)\sqrt{2Tlog\big(\frac{2}{\delta}\big)}. \tag{9}$$

In Proposition C.5, we show that $\sum_{t=1}^T s_{t,a(t)}^c \leq \sqrt{2dT\log(1+T/d)}$ using Lemma A.1.

Proposition C.5. $\sum_{t=1}^T s_{t,a(t)}^c \leq \sqrt{2dT\log(1+T/d)}$.

Proof. Take $X_t = b_{a(t)}(t) - \bar{b}(t)$, $Q = I_d$, and $A(t) = \sum_{\tau=1}^{t-1} X_\tau X_\tau^T$. Then by Jensen's inequality and Lemma A.1,

$$\begin{aligned}
\sum_{t=1}^{T} s_{t,a(t)}^c &\leq \sqrt{T\sum_{t=1}^{T}\{s_{t,a(t)}^c\}^2} \ \ (\because \text{ Jensen's inequality}) \\
&= \sqrt{T\sum_{t=1}^{T} X_t^T B(t)^{-1} X_t} \\
&\leq \sqrt{T\sum_{t=1}^{T} X_t^T \{Q + A(t)\}^{-1} X_t} \ \ (\because B(t) \succ Q + A(t)) \\
&\leq \sqrt{2T\log\Big(\frac{det(Q + A(T+1))}{det(Q)}\Big)} \ \ (\because \text{ Lemma A.1}) \\
&\leq \sqrt{2dT\log\Big(1+\frac{T}{d}\Big)}. \ \ (\because \text{ determinant-trace inequality.})
\end{aligned}$$

□

Due to (9), Proposition C.5 and the definitions of p and $g(T)$, we have with probability at least $1-\frac{\delta}{2}$,

$$\sum_{t=1}^{T} regret(t)I(E^{\hat{\mu}}(t)) \leq O\Big(d^{3/2}\sqrt{T}\sqrt{\log(Td)\log(T/\delta)}\big(\sqrt{\log(1+T/d)} + \sqrt{\log(1/\delta)}\big)\Big).$$

Since $E^{\hat{\mu}}(t)$ holds for all t with probability at least $1-\frac{\delta}{2}$ (Theorem C.1), $regret(t)I(E^{\hat{\mu}}(t)) = regret(t)$ for all t with probability at least $1-\frac{\delta}{2}$. Hence, with probability at least $1-\delta$,

$$R(T) \leq O\Big(d^{3/2}\sqrt{T}\sqrt{\log(Td)\log(T/\delta)}\big(\sqrt{\log(1+T/d)} + \sqrt{\log(1/\delta)}\big)\Big).$$

References

Abbasi-Yadkori, Y., Pál, D. and Szepesvári, C. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems,* pp. 2312–2320, 2011.

Abramowitz, M. and Stegun, I.A. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Washington, DC: National Bureau of Standards, 1964.

Agrawal, S. and Goyal, N. Thompson sampling for contextual bandits with linear payoffs. In *Proceedings of the 30th International Conference on Machine Learning*, pp. 127–135, 2013.

Bercu, B. and Touati, A. Exponential inequalities for self-normalized martingales with applications. *The Annals of Applied Probability,* 18(5):1848–1869, 2008.

de la Peña ,V. H., Klass, M. J. and Lai, T. L. Theory and applications of multivariate self-normalized processes. *Stochastic Processes and their Applications,* 119(12):4210–4227, 2009.