Online Appendix for "Non-stationary Experimental Design"

EC.1. Technical Details of Section 3

EC.1.1. Proof to Proposition 1

First with the union bound, we can have

$$\mathbb{P}\left(\left\|\hat{\theta}_T^{\texttt{L-EOFU}(\alpha)}-\theta_\nu\right\|_1 \geq \delta_0\right) = \mathbb{P}\left(\left\|\hat{\theta}_{1,T}^{\texttt{L-EOFU}(\alpha)}-\hat{\theta}_{2,T}^{\texttt{L-EOFU}(\alpha)}-\theta_1+\theta_2\right\|_1 \geq \delta_0\right)$$
$$\leq \mathbb{P}\left(\left\|\hat{\theta}_{1,T}^{\texttt{L-EOFU}(\alpha)}-\theta_1\right\|_1 \geq \frac{\delta_0}{2}\right)+\mathbb{P}\left(\left\|\hat{\theta}_{2,T}^{\texttt{L-EOFU}(\alpha)}-\theta_2\right\|_1 \geq \frac{\delta_0}{2}\right). \quad \text{(EC.1)}$$

Therefore, we are going to bound the two terms respectively. Note that since these two terms are almost symmetric. For simplicity, we assume T^α is an even number and $T^\alpha \geq 4$. We start with the first term of Eq. (EC.1). Denote $V_{a,0}=\sum_{(\phi(i),R_i)\in\mathcal{D}_{a,0}}\phi(i)\phi(i)^\top$ for $a\in\{0,1\}$. By the definition of $\hat{\theta}_{1,T}^{\texttt{L-EOFU}(\alpha)}$, we can have

$$\begin{aligned}\hat{\theta}_{1,T}^{\texttt{L-EOFU}(\alpha)} &= V_{1,0}^{-1}\sum_{(\phi(i),R_i)\in\mathcal{D}_{1,0}} R_i\phi(i)\\ &= V_{1,0}^{-1}\sum_{i=1,3,\cdots,T^\alpha-1}(\theta_1^\top\phi(i)+\varepsilon_{1,i})\phi(i)\\ &= \theta_1+V_{1,0}^{-1}\sum_{i=1,3,\cdots,T^\alpha-1}\varepsilon_{1,i}\phi(i).\end{aligned}$$

Therefore, $\hat{\theta}_{1,T}^{\texttt{L-EOFU}(\alpha)}-\theta_1=V_{1,0}^{-1}\sum_{i=1,3,\cdots,T^\alpha-1}\varepsilon_{1,i}\phi(i)$. We now begin to work on the matrix $V_{1,0}$.

$$V_{1,0}=\sum_{i=1,3,\cdots,T^\alpha-1}\begin{pmatrix}1 & i\\ i & i^2\end{pmatrix}=\begin{pmatrix}T^\alpha/2 & T^{2\alpha}/4\\ T^{2\alpha}/4 & T^\alpha(T^{2\alpha}-1)/6\end{pmatrix}=\frac{T^\alpha}{2}\begin{pmatrix}1 & T^\alpha/2\\ T^\alpha/2 & (T^{2\alpha}-1)/3\end{pmatrix}.$$

The inverse of the matrix $V_{1,0}$ can be calculated explicitly as

$$V_{1,0}^{-1}=\frac{2}{T^\alpha}\begin{pmatrix}\frac{4-4T^{2\alpha}}{4-T^{2\alpha}} & \frac{6T^\alpha}{4-T^{2\alpha}}\\ \frac{6T^\alpha}{4-T^{2\alpha}} & \frac{12}{T^{2\alpha}-4}\end{pmatrix}=\frac{4}{T^\alpha(T^{2\alpha}-4)}\begin{pmatrix}2T^{2\alpha}-2 & -3T^\alpha\\ -3T^\alpha & 6\end{pmatrix}.$$

Thus, $\hat{\theta}_{1,T}^{\texttt{L-EOFU}(\alpha)}-\theta_1$ can be further decomposed as

$$\begin{aligned}\hat{\theta}_{1,T}^{\texttt{L-EOFU}(\alpha)}-\theta_1 &= \frac{4}{T^\alpha(T^{2\alpha}-4)}\begin{pmatrix}2T^{2\alpha}-2 & -3T^\alpha\\ -3T^\alpha & 6\end{pmatrix}\begin{pmatrix}\sum_{i=1,3,\cdots,T^\alpha-1}\varepsilon_{1,i}\\ \sum_{i=1,3,\cdots,T^\alpha-1}i\cdot\varepsilon_{1,i}\end{pmatrix}\\ &= \frac{4}{T^\alpha(T^{2\alpha}-4)}\begin{pmatrix}\sum_{i=1,3,\cdots,T^\alpha-1}(2T^{2\alpha}-2-3T^\alpha i)\varepsilon_{1,i}\\ \sum_{i=1,3,\cdots,T^\alpha-1}(-3T^\alpha+6i)\varepsilon_{1,i}\end{pmatrix}.\end{aligned} \quad \text{(EC.2)}$$

Note that all the data in $\mathcal{D}_{a,0}$ is independent. We now can use some standard techniques to bound the probability tails. We start with the following probability, for any $\delta_1>0$,

$$\mathbb{P}\left(\left|\frac{4\sum_{i=1,3,\cdots,T^\alpha-1}(2T^{2\alpha}-2-3T^\alpha i)\varepsilon_{1,i}}{T^\alpha(T^{2\alpha}-4)}\right|\geq\delta_1\right)\leq 2\mathbb{P}\left(\frac{4\sum_{i=1,3,\cdots,T^\alpha-1}(2T^{2\alpha}-2-3T^\alpha i)\varepsilon_{1,i}}{T^\alpha(T^{2\alpha}-4)}\geq\delta_1\right)$$

$$
\begin{aligned}
&\le \frac{2\mathbb{E}[\exp(4s\sum_{i=1,3,\cdots,T^\alpha-1}(2T^{2\alpha}-2-3T^\alpha i)\varepsilon_{1,i})]}{\exp(s\delta_1(T^\alpha(T^{2\alpha}-4)))}\\
&= \frac{2\Pi_{i=1,3,\cdots,T^\alpha-1}\mathbb{E}[\exp(4s(2T^{2\alpha}-2-3T^\alpha i)\varepsilon_{1,i})]}{\exp(s\delta_1(T^\alpha(T^{2\alpha}-4)))}\\
&\le \frac{2\Pi_{i=1,3,\cdots,T^\alpha-1}[\exp(8s^2(2T^{2\alpha}-2-3T^\alpha i)^2\sigma_1^2)]}{\exp(s\delta_1(T^\alpha(T^{2\alpha}-4)))}\\
&\le 2\exp(16s^2T^{5\alpha}\sigma_1^2 - s\delta_1(T^\alpha(T^{2\alpha}-4)))\\
&= 2\exp\left(-\frac{\delta_1^2(T^{2\alpha}-4)^2}{64T^{3\alpha}\sigma_1^2}\right), \qquad \text{(EC.3)}
\end{aligned}
$$

where the first inequality holds by symmetry, the second inequality adopts Markov inequality, the first equation is due to the independence of all the data, the third inequality follows from the definition of the sub-Gaussian noise, the fourth inequality uses $(2T^{2\alpha}-2-3T^\alpha i)^2 \le 2T^{4\alpha}$ and we select $s = \frac{\delta_1(T^{2\alpha}-4)}{64T^{4\alpha}\sigma_1^2}$ in the last step. Similarly, for any $\delta_2 > 0$,

$$
\begin{aligned}
\mathbb{P}\left(\left|\frac{4\sum_{i=1,3,\cdots,T^\alpha-1}(-3T^\alpha+6i)\varepsilon_{1,i}}{T^\alpha(T^{2\alpha}-4)}\right| \ge \delta_2\right) &\le 2\mathbb{P}\left(\frac{4\sum_{i=1,3,\cdots,T^\alpha-1}(-3T^\alpha+6i)\varepsilon_{1,i}}{T^\alpha(T^{2\alpha}-4)} \ge \delta_2\right)\\
&\le \frac{2\mathbb{E}[\exp(4s\sum_{i=1,3,\cdots,T^\alpha-1}(-3T^\alpha+6i)\varepsilon_{1,i})]}{\exp(s\delta_2(T^\alpha(T^{2\alpha}-4)))}\\
&= \frac{2\Pi_{i=1,3,\cdots,T^\alpha-1}\mathbb{E}[\exp(4s(-3T^\alpha+6i)\varepsilon_{1,i})]}{\exp(s\delta_2(T^\alpha(T^{2\alpha}-4)))}\\
&\le \frac{2\Pi_{i=1,3,\cdots,T^\alpha-1}[\exp(8s^2(-3T^\alpha+6i)^2\sigma_1^2)]}{\exp(s\delta_2(T^\alpha(T^{2\alpha}-4)))}\\
&\le 2\exp(36s^2T^{3\alpha}\sigma_1^2 - s\delta_2(T^\alpha(T^{2\alpha}-4)))\\
&= 2\exp\left(-\frac{\delta_2^2(T^{2\alpha}-4)^2}{144T^{\alpha}\sigma_1^2}\right), \qquad \text{(EC.4)}
\end{aligned}
$$

where we choose s to be $\frac{\delta_1(T^{2\alpha}-4)}{72T^{2\alpha}\sigma_1^2}$ at last. By combining Eqs. (EC.3) and (EC.4) and the union bound and setting $\delta_1 = \frac{1/3T^\alpha\delta_0}{1+2/3T^\alpha}$ and $\delta_2 = \frac{1/2\delta_0}{1+2/3T^\alpha}$, we can derive

$$
\mathbb{P}\left(\left\|\hat{\theta}_{1,T}^{\texttt{L-EOFU}(\alpha)} - \theta_1\right\|_1 \ge \frac{\delta_0}{2}\right) \le 4\exp\left(-\frac{(T^{2\alpha}-4)^2\delta_0^2}{144(1+\frac{2}{3}T^\alpha)^2T^\alpha\sigma_1^2}\right) \le 4\exp\left(-\frac{(T^{2\alpha}-4)^2\delta_0^2}{144T^{3\alpha}\sigma_1^2}\right), \qquad \text{(EC.5)}
$$

where the second inequality is based on our assumption that $T^\alpha \ge 4$.

The bound for the second term of Eq. (EC.1) bears strong similarity to Eq. (EC.5), with slight difference on the index of time. Specifically, $\hat{\theta}_{2,T}^{\texttt{L-EOFU}(\alpha)} - \theta_2 = V_{2,0}^{-1}\sum_{i=2,4,\cdots,T^\alpha}\varepsilon_{2,i}\phi(i)$, and

$$
\begin{aligned}
V_{2,0} = \sum_{i=2,4,\cdots,T^\alpha}\begin{pmatrix} 1 & i \\ i & i^2 \end{pmatrix} &= \begin{pmatrix} T^\alpha/2 & (T^{2\alpha}+2T^\alpha)/4 \\ (T^{2\alpha}+2T^\alpha)/4 & T^\alpha(T^\alpha+1)(T^\alpha+2)/6 \end{pmatrix}\\
&= \frac{T^\alpha}{2}\begin{pmatrix} 1 & (T^\alpha+2)/2 \\ (T^\alpha+2)/2 & (T^\alpha+1)(T^\alpha+2)/3 \end{pmatrix}.
\end{aligned}
$$

The inverse of the matrix $V_{2,0}$ is

$$
V_{2,0}^{-1} = \frac{2}{T^\alpha}\begin{pmatrix} \frac{4+4T^\alpha}{T^\alpha-2} & \frac{-6}{T^\alpha-2} \\ \frac{-6}{T^\alpha-2} & \frac{12}{T^{2\alpha}-4} \end{pmatrix} = \frac{4}{T^\alpha(T^\alpha-2)}\begin{pmatrix} 2+2T^\alpha & -3 \\ -3 & 6/(T^\alpha+2) \end{pmatrix}.
$$

Then, naturally we have

$$\begin{aligned}
\hat{\theta}_{2,T}^{\texttt{L-EOFU}(\alpha)} - \theta_2 &= \frac{4}{T^\alpha(T^\alpha-2)} \begin{pmatrix} 2+2T^\alpha & -3 \\ -3 & 6/(T^\alpha+2) \end{pmatrix} \begin{pmatrix} \sum_{i=2,4,\cdots,T^\alpha} \varepsilon_{2,i} \\ \sum_{i=2,4,\cdots,T^\alpha} i\cdot\varepsilon_{2,i} \end{pmatrix} \\
&= \frac{4}{T^\alpha(T^\alpha-2)} \begin{pmatrix} \sum_{i=2,4,\cdots,T^\alpha} (2T^\alpha+2-3i)\varepsilon_{2,i} \\ \sum_{i=2,4,\cdots,T^\alpha} (-3+6i/(T^\alpha+2))\varepsilon_{2,i} \end{pmatrix}.
\end{aligned}$$

Similarly, we start bounding with the first coordinate as, for any $\delta_3 > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left| \frac{4\sum_{i=2,4,\cdots,T^\alpha}(2T^\alpha+2-3i)\varepsilon_{2,i}}{T^\alpha(T^\alpha-2)} \right| \geq \delta_3 \right) &\leq 2\mathbb{P}\left(\frac{4\sum_{i=2,4,\cdots,T^\alpha}(2T^\alpha+2-3i)\varepsilon_{2,i}}{T^\alpha(T^\alpha-2)} \geq \delta_3 \right) \\
&\leq \frac{2\mathbb{E}[\exp(4s\sum_{i=2,4,\cdots,T^\alpha}(2T^\alpha+2-3i)\varepsilon_{2,i})]}{\exp(s\delta_3(T^\alpha(T^\alpha-2)))} \\
&= \frac{2\Pi_{i=2,4,\cdots,T^\alpha}\mathbb{E}[\exp(4s(2T^\alpha+2-3i)\varepsilon_{2,i})]}{\exp(s\delta_3(T^\alpha(T^\alpha-2)))} \\
&\leq \frac{2\Pi_{i=2,4,\cdots,T^\alpha}[\exp(8s^2(2T^\alpha+2-3i)^2\sigma_2^2)]}{\exp(s\delta_3(T^\alpha(T^\alpha-2)))} \\
&\leq 2\exp(16s^2T^{3\alpha}\sigma_2^2 - s\delta_3(T^\alpha(T^\alpha-2))) \\
&= 2\exp\left(-\frac{\delta_3^2(T^\alpha-2)^2}{64T^\alpha\sigma_2^2}\right), \qquad \text{(EC.6)}
\end{aligned}$$

where we use the fact that $(2T^\alpha+2-3i)^2 \leq 4T^\alpha$ and set $s = \frac{\delta_3(T^\alpha-2)}{32T^{2\alpha}\sigma_2^2}$. Moreover, for any $\delta_4 > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left| \frac{4\sum_{i=2,4,\cdots,T^\alpha}(-3+6i/(T^\alpha+2))\varepsilon_{2,i}}{T^\alpha(T^\alpha-2)} \right| \geq \delta_4 \right) &\leq 2\mathbb{P}\left(\frac{12\sum_{i=2,4,\cdots,T^\alpha}(2i-2-T^\alpha)\varepsilon_{2,i}}{T^\alpha(T^\alpha-2)(T^\alpha+2)} \geq \delta_4 \right) \\
&\leq \frac{2\mathbb{E}[\exp(12s\sum_{i=2,4,\cdots,T^\alpha}(2i-2-T^\alpha)\varepsilon_{2,i})]}{\exp(s\delta_4(T^\alpha(T^{2\alpha}-4)))} \\
&\leq \frac{2\Pi_{i=2,4,\cdots,T^\alpha}[\exp(72s^2(2i-2-T^\alpha)^2\sigma_2^2)]}{\exp(s\delta_4(T^\alpha(T^{2\alpha}-4)))} \\
&\leq 2\exp(36s^2T^{3\alpha}\sigma_2^2 - s\delta_4(T^\alpha(T^{2\alpha}-4))) \\
&= 2\exp\left(-\frac{\delta_4^2(T^{2\alpha}-4)^2}{144T^\alpha\sigma_2^2}\right), \qquad \text{(EC.7)}
\end{aligned}$$

where we take $s = \frac{\delta_4(T^{2\alpha}-4)}{72T^{2\alpha}\sigma_2^2}$. By taking $\delta_3 = \frac{\delta_0(T^\alpha+2)}{2T^\alpha+7}$ and $\delta_4 = \frac{3\delta_0}{4T^\alpha+14}$ in Eqs. (EC.6) and (EC.7), we can obtain that

$$\mathbb{P}\left(\left\| \hat{\theta}_{2,T}^{\texttt{L-EOFU}(\alpha)} - \theta_2 \right\|_1 \geq \frac{\delta_0}{2} \right) \leq 4\exp\left(-\frac{\delta_0^2(T^{2\alpha}-4)^2}{64T^\alpha\sigma_2^2(2T^\alpha+7)^2}\right) \leq 4\exp\left(-\frac{\delta_0^2(T^{2\alpha}-4)^2}{1024T^{3\alpha}\sigma_2^2}\right). \qquad \text{(EC.8)}$$

By plugging Eqs. (EC.5) and (EC.8) into Eq. (EC.1), we can have

$$\mathbb{P}\left(\left\| \hat{\theta}_{T}^{\texttt{L-EOFU}(\alpha)} - \theta_\nu \right\|_1 \geq \delta_0 \right) \leq 8\exp\left(-\frac{\delta_0^2(T^{2\alpha}-4)^2}{1024T^{3\alpha}(\sigma_1^2\vee\sigma_2^2)}\right).$$

By simple calculations, we further arrive at that for any $\epsilon > 0$,

$$\mathbb{P}\left(\left\| \hat{\theta}_{T}^{\texttt{L-EOFU}(\alpha)} - \theta_\nu \right\|_1 \geq 32(\sigma_1\vee\sigma_2)\frac{T^{\frac{3\alpha}{2}}}{T^{2\alpha}-4}\sqrt{\ln\frac{8}{\epsilon}} \right) \leq \epsilon,$$

which, with $T^\alpha \geq 4$, further implies,

$$\mathbb{P}\left(\left\|\hat{\theta}_T^{\texttt{L-EOFU}(\alpha)} - \theta_\nu\right\|_1 \geq \frac{128}{3}(\sigma_1 \vee \sigma_2)T^{-\frac{\alpha}{2}}\sqrt{\ln\frac{8}{\epsilon}}\right) \leq \epsilon.$$

We finish the proof. Q.E.D.

EC.1.2. Proof to Theorem 1

Since our L-EOFU operates in a two-stage manner. We are going to bound the regrets for each stage separately. For simplicity, we assume T^α is an integer. For the first exploration stage, since $\|\theta_a\| \leq S$, we can bound the first-stage regret $\mathcal{R}_\nu^{\texttt{L-EOFU}(\alpha),1}(T)$ as

$$\mathcal{R}_\nu^{\texttt{L-EOFU}(\alpha),1}(T) \leq \sum_{t=1}^{T^\alpha}\left|(\theta_1 - \theta_2)^\top \phi(t)\right| \leq 2ST^\alpha + 2S\sum_{t=1}^{T^\alpha} t \leq 4T^{2\alpha}S. \tag{EC.9}$$

For the second stage, by Theorem 2 in Abbasi-Yadkori et al. (2011) (see, also Lemma EC.2 in Appendix EC.3), with probability at least $1-\delta$, for all $t \in [T]$ and $a \in \{1,2\}$,

$$\tilde{R}_t(a) = \tilde{\theta}_{a,t}^\top \phi(t) + \gamma_{a,t}\|\phi(t)\|_{V_{a,t}^{-1}} \geq \theta_a^\top \phi(t). \tag{EC.10}$$

This is the reason why the algorithm is partially called by *OFU*. We first investigate the single-step regret at time t when arm 1 is optimal for illustration.

$$\begin{aligned}
r_t &= (\theta_1 - \theta_2)^\top \phi(t)\mathbb{I}_{A_t^*=1,A_t=2} \\
&\leq \left(\tilde{\theta}_{1,t}^\top \phi(t) + \gamma_{1,t}\|\phi(t)\|_{V_{1,t}^{-1}} - \theta_2^\top \phi(t)\right)\mathbb{I}_{A_t^*=1,A_t=2} \\
&\leq \left(\tilde{\theta}_{2,t}^\top \phi(t) + \gamma_{2,t}\|\phi(t)\|_{V_{2,t}^{-1}} - \theta_2^\top \phi(t)\right)\mathbb{I}_{A_t^*=1,A_t=2} \\
&\leq 2\gamma_{2,t}\|\phi(t)\|_{V_{2,t}^{-1}}\mathbb{I}_{A_t^*=1,A_t=2},
\end{aligned}$$

where the first and the third inequalities hold due to Eq. (EC.10), and the second inequality holds since we are optimistic. We can know the follows about the regret of the second phase $\mathcal{R}_\nu^{\texttt{L-EOFU}(\alpha),2}(T)$:

$$\begin{aligned}
\mathcal{R}_\nu^{\texttt{L-EOFU}(\alpha),2}(T) &= \sum_{t=T^\alpha+1}^{T} r_t \\
&\leq \sum_{t=T^\alpha+1}^{T} 2\left[\gamma_{2,t}\|\phi(t)\|_{V_{2,t}^{-1}}\mathbb{I}_{A_t^*=1,A_t=2} + \gamma_{1,t}\|\phi(t)\|_{V_{1,t}^{-1}}\mathbb{I}_{A_t^*=2,A_t=1}\right] \\
&\leq \sum_{t=T^\alpha+1}^{T} 2\left[\gamma_{2,t}\|\phi(t)\|_{V_{2,t}^{-1}}\mathbb{I}_{A_t^*=1,A_t=2}\right] + \sum_{t=T^\alpha+1}^{T} 2\left[\gamma_{1,t}\|\phi(t)\|_{V_{1,t}^{-1}}\mathbb{I}_{A_t^*=2,A_t=1}\right] \\
&\leq \sum_{t=T^\alpha+1}^{T} 2\left[\gamma_{2,t}\|\phi(t)\|_{V_{2,t}^{-1}}\mathbb{I}_{A_t=2}\right] + \sum_{t=T^\alpha+1}^{T} 2\left[\gamma_{1,t}\|\phi(t)\|_{V_{1,t}^{-1}}\mathbb{I}_{A_t=1}\right] \\
&\leq \sum_{t=T^\alpha+1}^{T} 2\gamma\left[\|\phi(t)\|_{V_{2,t}^{-1}}\mathbb{I}_{A_t=2}\right] + \sum_{t=T^\alpha+1}^{T} 2\gamma\left[\|\phi(t)\|_{V_{1,t}^{-1}}\mathbb{I}_{A_t=1}\right] \\
&\leq \sum_{t=1}^{T} 2\gamma\left[\|\phi(t)\|_{V_{2,t}^{-1}}\mathbb{I}_{A_t=2}\right] + \sum_{t=1}^{T} 2\gamma\left[\|\phi(t)\|_{V_{1,t}^{-1}}\mathbb{I}_{A_t=1}\right],
\end{aligned} \tag{EC.11}$$

where $\gamma = (\sigma_1 \vee \sigma_2)\sqrt{12\ln(T) - 2\ln(4.5\lambda\delta)} + \lambda^{1/2}S$. The last inequality follows from the following facts: $\gamma_{1,t} \le \gamma_{1,T}$ and $\gamma_{2,t} \le \gamma_{2,T}$ for all t and

$$\det(V_{1,T}) \le \left(\frac{\operatorname{trace}(V_{1,T})}{2}\right)^2 \le \left(\frac{T + \sum_{t=1}^T t^2 + 2\lambda}{2}\right)^2 \le \frac{T^6}{9},$$

where the last inequality holds when $T \ge 2$ and $\lambda \le 1$. Then, it is sufficient to bound $\sum_{t=1}^T \|\phi(t)\|_{V_{1,t}^{-1}} \mathbb{I}_{A_t=1}$. Denote $\Psi_1 = \{t : A_t = 1\}$, and index the elements in Ψ_1 by $0 := t_{1,0} < t_{1,1} < t_{1,2} < \cdots < t_{1,|\Psi_1|} \le T$. Then, we can reorganize

$$\begin{aligned} \sum_{t=1}^T \|\phi(t)\|_{V_{1,t}^{-1}} \mathbb{I}_{A_t=1} &= \sum_{i=1}^{|\Psi_1|} \|\phi(t_{1,i})\|_{V_{1,t_{1,i}}^{-1}} \\ &= \sum_{i=1}^{|\Psi_1|} \sqrt{\|\phi(t_{1,i})\|^2_{V_{1,t_{1,i}}^{-1}}} \\ &= \sum_{i=1}^{|\Psi_1|} \sqrt{\frac{\det(V_{1,t_{1,i}})}{\det(V_{1,t_{1,i-1}})} - 1}, \end{aligned} \tag{EC.12}$$

where the last inequality holds based on the following observation:

$$\begin{aligned} \det(V_{1,t_{1,i}}) &= \det(V_{1,t_{1,i}} + \phi(t_{1,i})\phi(t_{1,i})^\top) = \det(V_{1,t_{1,i-1}})\det(I + V_{1,t_{1,i-1}}^{-1/2}\phi(t_{1,i})(V_{1,t_{1,i-1}}^{-1/2}\phi(t_{1,i}))^\top) \\ &= \det(V_{1,t_{1,i-1}})(1 + \|\phi(t_{1,i})\|^2_{V_{1,t_{1,i}}^{-1}}). \end{aligned}$$

The RHS of Eq. (EC.12) can be upper bounded by the optimal value of the optimization problem:

$$\begin{aligned} \max_{h_1,\cdots,h_{|\Psi_1|}} \quad & \sum_{i=1}^{|\Psi_1|} \sqrt{h_i - 1} \\ \text{s.t.} \quad & \Pi_{i=1}^{|\Psi_1|} h_i \le |\Psi_1|^2 T^4, \\ & h_i \ge 1 \qquad \forall i \in [|\Psi_1|]. \end{aligned} \tag{EC.13}$$

The first constraint adopts similar argument as before that

$$\det(V_{1,t_{1,|\Psi_1|}}) \le \left(\frac{\operatorname{trace}(V_{1,t_{1,|\Psi_1|}})}{2}\right)^2 = \left(\frac{2\lambda + |\Psi_1| + \sum_{t\in\Psi_1} t^2}{2}\right)^2 \le |\Psi_1|^2 T^4.$$

By Lemma 8 in Chu et al. (2011), the optimal solution to Eq. (EC.13) is $h_1 = h_2 = \cdots = h_{|\Psi_1|} = (|\Psi_1|^2 T^4)^{1/|\Psi_1|}$. Therefore, the RHS of Eq. (EC.12) can be bounded by

$$\begin{aligned} \sum_{i=1}^{|\Psi_1|} \sqrt{\frac{\det(V_{1,t_{1,i}})}{\det(V_{1,t_{1,i-1}})} - 1} &\le |\Psi_1| \sqrt{(|\Psi_1| T^2)^{\frac{2}{|\Psi_1|}} - 1} \\ &\le |\Psi_1| \sqrt{(2^{\frac{2}{\alpha}} |\Psi_1|^{1+\frac{2}{\alpha}})^{\frac{2}{|\Psi_1|}} - 1} \\ &\le |\Psi_1| \sqrt{\frac{4^{1+\frac{2}{\alpha}}}{\ln 2} \frac{\ln(2|\Psi_1|)}{|\Psi_1|}} \\ &= \sqrt{\frac{4^{1+\frac{2}{\alpha}}}{\ln 2} \ln(2|\Psi_1|)|\Psi_1|}, \end{aligned} \tag{EC.14}$$

where the second inequality follows from the fact that $|\Psi_1| \geq \frac{T^\alpha}{2}$ since in the first stage we at least gain $\frac{T^\alpha}{2}$ observations. In order to establish the third inequality, it is sufficient to show that

$$(2|\Psi_1|)^{(1+\frac{2}{\alpha})\frac{2}{|\Psi_1|}} \leq \frac{4^{1+\frac{2}{\alpha}}}{\ln 2}\frac{\ln(2|\Psi_1|)}{|\Psi_1|}+1,$$

which is equivalent to

$$\left(2+\frac{4}{\alpha}\right)\frac{\ln(2|\Psi_1|)}{|\Psi_1|} \leq \ln\left(\frac{4^{1+\frac{2}{\alpha}}}{\ln 2}\frac{\ln(2|\Psi_1|)}{|\Psi_1|}+1\right).$$

When $|\Psi_1| \geq 2$, the above inequality can be easily verified.

Following exactly the same procedure for arm 2, we can achieve

$$\sum_{t=1}^{T}\|\phi(t)\|_{V_{2,t}^{-1}}\mathbb{I}_{A_t=2} \leq \sqrt{\frac{4^{1+\frac{2}{\alpha}}}{\ln 2}\ln(2|\Psi_2|)|\Psi_2|}, \tag{EC.15}$$

where $|\Psi_2| := \{t : A_t = 2\}$. Plugging Eqs. (EC.14) and (EC.15) back into Eq. (EC.11) with $|\Psi_1| \leq T$ and $|\Psi_2| \leq T$, we derive

$$\mathcal{R}_\nu^{\texttt{L-EOFU}(\alpha),2}(T) \leq 4\gamma\sqrt{\frac{4^{1+\frac{2}{\alpha}}}{\ln 2}\ln(2T)T}. \tag{EC.16}$$

Put Eqs. (EC.9) and (EC.16) together, and we finish the proof. Q.E.D.

A quick remark about why the techniques of Theorem 3 in Abbasi-Yadkori et al. (2011) cannot apply here is that the single-step regret has to be bounded by a constant independent of T in their case, which cannot hold here.

EC.1.3. Proof to Proposition 2

We first construct two instances ν_1 and ν_2 as follows. For ν_1, we set arm 1 to be with $\theta_1^{(1)} = (0,0)$ and arm 2 to be with $\theta_2^{(1)} = (f(T),0)$, where $f(T)$ is a function of T which will be specified later. For ν_2, arm 1 remains the same to be with $\theta_1^{(2)} = (0,0)$ and arm 2 to be with $\theta_2^{(1)} = (2f(T),0)$. All the noises for ν_1 and ν_2 are assumed to be σ_0^2-Gaussain.

$$\begin{aligned}
\inf_{\hat\theta}\sup_{\nu\in\mathcal{E}_1}\mathbb{P}_\nu\left(\left\|\hat\theta-\theta_\nu\right\|_1 > f(T)\right) &\geq \inf_{\hat\theta}\max_{i\in\{1,2\}}\mathbb{P}_{\nu_i}\left(\left\|\hat\theta-\theta_{\nu_i}\right\|_1 > f(T)\right) \\
&\geq \inf_{\Psi}\max_{i\in\{1,2\}}\mathbb{P}_{\nu_i}(\Psi \neq i) \\
&= \frac{1}{2}\left[1-\mathrm{TV}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})\right] \\
&\geq \frac{1}{2}[1-\sqrt{\frac{\mathrm{KL}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})}{2}}] \\
&\geq \frac{1}{2}\left[1-\sqrt{\frac{Tf(T)^2}{2\sigma_0^2}}\right],
\end{aligned}$$

where Ψ is the test for whether the underlying model is ν_1 or ν_2, the third inequality holds due to Pinsker's inequality (see, e.g., Wainwright 2019), and the last equality follows from the KL-divergence of two normal distributions with the same variance. By setting $\sqrt{\frac{Tf(T)^2}{2\sigma_0^2}} = 2c_2$, we can solve that $f(T) = \frac{2\sqrt{2}c_2\sigma_0}{\sqrt{T}}$.

Therefore, we can have

$$\inf_{\hat{\theta}} \sup_{\nu\in\mathcal{E}_1} \mathbb{E}_\nu \left[\left\|\hat{\theta} - \theta_\nu\right\|_1\right] \geq \left(\frac{1}{2} - c_2\right)\frac{2\sqrt{2}c_2\sigma_0}{\sqrt{T}}.$$

By choosing $c_2 = \frac{1}{4}$,

For the regret lower bound, we again construct two instances ν_1 and ν_2 as follows. For ν_1, we set arm 1 to be with $\theta_1^{(1)} = (0,0)$ and arm 2 to be with $\theta_2^{(1)} = (g(T), -\frac{2g(T)}{T})$, where $g(T)$ is a function of T which will be specified soon. For ν_2, arm 1 remains the same to be with $\theta_1^{(2)} = (0,0)$ and arm 2 to be with $\theta_2^{(1)} = (-g(T), \frac{2g(T)}{T})$. Note that when $t \leq \frac{T}{2}$, arm 2 is optimal for ν_1 and arm 1 is optimal for ν_2. All the noises for ν_1 and ν_2 are assumed to be σ_0^2-Gaussain. Define event $\mathcal{A}$ to be that arm 1 is played $\frac{T}{4}$ in the first $\frac{T}{2}$ periods.

$$\begin{aligned}
\inf_\pi \sup_{\nu\in\mathcal{E}_1} \mathbb{E}\left[\mathcal{R}_\nu^\pi(T)\right] &\geq \inf_\pi \max_{i\in\{1,2\}} \mathbb{E}\left[\mathcal{R}_{\nu_i}^\pi(T)\right] \\
&\geq \inf_\pi \mathbb{P}_{\nu_1}^\pi(\mathcal{A})\frac{g(T)(T+4)}{16} + \mathbb{P}_{\nu_2}^\pi(\mathcal{A}^c)\frac{g(T)(T+4)}{16} \\
&\geq \frac{g(T)(T+4)}{32}\exp(-\mathrm{KL}(\mathbb{P}_{\nu_1}, \mathbb{P}_{\nu_2})),
\end{aligned} \tag{EC.17}$$

where the last inequality holds due to Bretagnolle-Huber inequality (see, Bretagnolle and Huber 1979, Lattimore and Szepesvári 2020 and Lemma EC.3 in Appendix EC.3). What remains now is to upper bound $\mathrm{KL}(\mathbb{P}_{\nu_1}, \mathbb{P}_{\nu_2})$.

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_{\nu_1}, \mathbb{P}_{\nu_2}) &= \sum_{t=1}^{T} \frac{(2g(T) - \frac{4g(T)}{T}t)^2}{2\sigma_0^2}\mathbb{P}_{\nu_1}(A_t = 2) \\
&\leq \frac{2g(T)^2}{\sigma_0^2 T^2}\sum_{t=1}^{T}(T-2t)^2 \\
&\leq \frac{2g(T)^2 T}{\sigma_0^2}.
\end{aligned} \tag{EC.18}$$

Plugging Eq. (EC.18) back into Eq. (EC.17) and choosing $g(T) = \frac{\sigma_0}{\sqrt{2T}}$, we derive

$$\inf_\pi \sup_{\nu\in\mathcal{E}_1} \mathbb{E}\left[\mathcal{R}_\nu^\pi(T)\right] \geq \frac{\sigma_0(T+4)}{32e\sqrt{T}} > \frac{\sigma_0\sqrt{T}}{32e}.$$

We finish the proof. Q.E.D.

EC.1.4. Proof to Theorem 2

Without loss of generality, we specify $S=2$ in this proof. We first construct two instances ν_1 and ν_2 as follows. For ν_1, we set arm 1 to be with $\theta_1^{(1)}=(0,0)$ and arm 2 to be with $\theta_2^{(1)}=(0,-1)$. For ν_2, arm 1 remains the same to be with $\theta_1^{(2)}=(0,0)$ and arm 2 to be with $\theta_2^{(2)}=(-2f(T),-1)$. Thus, $\theta_{\nu_1}=(0,1)$ and $\theta_{\nu_2}=(2f(T),1)$. All the noises for ν_1 and ν_2 are assumed to be σ_0^2-Gaussain. Following the traditional statistical testing, we define the minimum distance test $\psi(\hat{\theta}_T)$ with respect to l_1 norm, which is associated to $\hat{\theta}_T$ by $\psi(\hat{\theta}_T)=\arg\min_{i=1,2}\left\|\hat{\theta}_T-\theta_{\nu_i}\right\|_1$. If $\psi(\hat{\theta}_T)=1$, we know that $\left\|\hat{\theta}_T-\theta_{\nu_1}\right\|_1\le\left\|\hat{\theta}_T-\theta_{\nu_2}\right\|_1$. By the triangle inequality, we can have, if $\psi(\hat{\theta}_T)=1$,

$$\left\|\hat{\theta}_T-\theta_{\nu_2}\right\|_1\ge\|\theta_{\nu_1}-\theta_{\nu_2}\|_1-\left\|\hat{\theta}_T-\theta_{\nu_1}\right\|_1\ge\|\theta_{\nu_1}-\theta_{\nu_2}\|_1-\left\|\hat{\theta}_T-\theta_{\nu_2}\right\|_1,$$

which yields that $\left\|\hat{\theta}_T-\theta_{\nu_2}\right\|_1\ge\frac{1}{2}\|\theta_{\nu_1}-\theta_{\nu_2}\|_1=f(T)$. Symmetrically, if $\psi(\hat{\theta}_T)=2$, we can have $\left\|\hat{\theta}_T-\theta_{\nu_1}\right\|_1\ge\frac{1}{2}\|\theta_{\nu_1}-\theta_{\nu_2}\|_1=f(T)$. Therefore, we can have

$$\begin{aligned}\inf_{\hat{\theta}_T}\max_{\nu\in\mathcal{E}_1}\mathbb{P}_\nu^\pi\left(\left\|\hat{\theta}_T-\theta_\nu\right\|_1\ge f(T)\right)&\ge\inf_{\pi\in\Pi_\beta,\hat{\theta}_T}\max_{i\in\{1,2\}}\mathbb{P}_{\nu_i}^\pi\left(\left\|\hat{\theta}_T-\theta_{\nu_i}\right\|_1\ge f(T)\right)\\&\ge\inf_{\hat{\theta}_T}\max_{i\in\{1,2\}}\mathbb{P}_{\nu_i}^\pi(\psi(\hat{\theta}_T)\ne i)\\&\ge\inf_{\psi}\max_{i\in\{1,2\}}\mathbb{P}_{\nu_i}^\pi(\psi\ne i)\\&\ge\frac{1}{2}\inf_{\psi}\left(\mathbb{P}_{\nu_1}^\pi(\psi\ne1)+\mathbb{P}_{\nu_2}^\pi(\psi\ne2)\right)\\&=\frac{1}{2}\left[1-\mathrm{TV}(\mathbb{P}_{\nu_1}^\pi,\mathbb{P}_{\nu_2}^\pi)\right]\\&\ge\frac{1}{2}\left[1-\sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}_{\nu_1}^\pi,\mathbb{P}_{\nu_2}^\pi)}\right],\end{aligned}\tag{EC.19}$$

where the third line is taken over all tests ψ based on $\mathcal{H}_T$ that take values in $\{1,2\}$, the first equality holds due to Neyman-Pearson Lemma (see, Neyman and Pearson 1933 and Lemma EC.4 in Appendix EC.3), the fifth inequality follows from Pinsker's inequality.

We now need to further upper bound $\sup_{\pi\in\Pi_\beta}\mathrm{KL}(\mathbb{P}_{\nu_1}^\pi,\mathbb{P}_{\nu_2}^\pi)$. With the chain rule of KL-divergence, since ν_1 and ν_2 are only different in arm 2, we can decompose $\mathrm{KL}(\mathbb{P}_{\nu_1}^\pi,\mathbb{P}_{\nu_2}^\pi)$ as follows:

$$\mathrm{KL}(\mathbb{P}_{\nu_1}^\pi,\mathbb{P}_{\nu_2}^\pi)=\sum_{t=1}^T\mathbb{E}_{\nu_1}^\pi[\mathrm{KL}(P_{1,A_t},P_{2,A_t})]=\sum_{t=1}^T\frac{2f(T)^2}{\sigma_0^2}\mathbb{E}_{\nu_1}^\pi[\mathbb{I}_{A_t=2}]=\frac{2f(T)^2}{\sigma_0^2}\sum_{t=1}^T\mathbb{P}_{\nu_1}^\pi(A_t=2).\tag{EC.20}$$

Moreover, because of $\pi\in\Pi_\beta$ and $\mathbb{E}\left[\mathcal{R}_{\nu_1}^\pi(T)\right]=\sum_{t=1}^T t\mathbb{P}_{\nu_1}^\pi(A_t=2)$, we can have the constraint that there exists a constant $c_3>0$,

$$\sum_{t=1}^T t\mathbb{P}_{\nu_1}^\pi(A_t=2)\le c_3T^\beta.\tag{EC.21}$$

Therefore, we are now using Eq. (EC.21) to bound the RHS of Eq. (EC.20). The RHS of Eq. (EC.21) can be bounded by the optimal value of the following optimization problem.

$$\begin{aligned} \max_{x_1,\cdots,x_T} \quad & \sum_{t=1}^{T} x_t \\ \text{s.t.} \quad & \sum_{t=1}^{T} t x_t \le c_3 T^{\beta} \\ & 0 \le x_t \le 1 \quad \forall t \in [T]. \end{aligned}$$

By observation, the optimal solution of the above optimization is $x_1 = \cdots = x_{\lfloor \frac{-1+\sqrt{1+8c_3T^\beta}}{2} \rfloor} = 1$ and $x_{\lfloor \frac{-1+\sqrt{1+8c_3T^\beta}}{2} \rfloor + 1} = (c_3T^\beta - \frac{1}{2}(\lfloor \frac{-1+\sqrt{1+8c_3T^\beta}}{2} \rfloor + 1)\lfloor \frac{-1+\sqrt{1+8c_3T^\beta}}{2} \rfloor)/(\lfloor \frac{-1+\sqrt{1+8c_3T^\beta}}{2} \rfloor + 1)$. Hence,

$$\sum_{t=1}^{T} \mathbb{P}^{\pi}_{\nu_1}(A_t = 2) \le \frac{1+\sqrt{1+8c_3T^\beta}}{2} \le \frac{1+4\sqrt{c_4}T^{\frac{\beta}{2}}}{2} \le 4\sqrt{c_4}T^{\frac{\beta}{2}}, \tag{EC.22}$$

where $c_4 = c_3 \vee \frac{1}{8}$. Plugging Eq. (EC.22) back into Eq. (EC.20), we arrive at the point that

$$\inf_{\hat{\theta}_T} \max_{\nu\in\mathcal{E}_1} \mathbb{P}^{\pi}_{\nu}\left(\left\|\hat{\theta}_T - \theta_\nu\right\|_1 \ge f(T)\right) \ge \frac{1}{2}\left[1 - \sqrt{\frac{4\sqrt{c_4}f(T)^2}{\sigma_0^2}T^{\frac{\beta}{2}}}\right].$$

By choosing $f(T) = \frac{\sigma_0\xi}{c_4^{1/4}T^{\beta/4}}$, we finally derive

$$\inf_{\hat{\theta}_T} \max_{\nu\in\mathcal{E}_1} \mathbb{P}^{\pi}_{\nu}\left(\left\|\hat{\theta}_T - \theta_\nu\right\|_1 \ge c_4^{-\frac{1}{4}}\sigma_0\xi T^{-\frac{\beta}{4}}\right) \ge \frac{1}{2} - \xi.$$

Moreover, by choosing $\xi = \frac{1}{4}$, we have

$$\inf_{\hat{\theta}_T} \max_{\nu\in\mathcal{E}_1} \mathbb{E}^{\pi}_{\nu}\left[\left\|\hat{\theta}_T - \theta_\nu\right\|_1\right] \ge \frac{\sigma_0}{16c_4^{\frac{1}{4}}}T^{-\frac{\beta}{4}}.$$

We finish the proof. Q.E.D.

EC.1.5. Proof to Theorem 3

Recall that in Eq. (EC.2) we have established that

$$\hat{\theta}^{\texttt{L-EOFU}(\alpha)}_{1,T} - \theta_1 = \frac{4}{T^\alpha(T^{2\alpha}-4)}\begin{pmatrix}\sum_{i=1,3,\cdots,T^\alpha-1}(2T^{2\alpha}-2-3T^\alpha i)\varepsilon_{1,i} \\ \sum_{i=1,3,\cdots,T^\alpha-1}(-3T^\alpha+6i)\varepsilon_{1,i}\end{pmatrix}.$$

Thus, we can have

$$\begin{pmatrix}\hat{\theta}^{\texttt{L-EOFU}(\alpha)}_{1,T,0} - \theta_{1,0} \\ T^\alpha(\hat{\theta}^{\texttt{L-EOFU}(\alpha)}_{1,T,1} - \theta_{1,1})\end{pmatrix} = \frac{4}{T^\alpha(T^{2\alpha}-4)}\sum_{i=1,3,\cdots,T^\alpha-1}\begin{pmatrix}2T^{2\alpha}-2-3T^\alpha i \\ -3T^{2\alpha}+6T^\alpha i\end{pmatrix}\varepsilon_{1,i}, \tag{EC.23}$$

where $\hat{\theta}^{\texttt{L-EOFU}(\alpha)}_{1,T,0}$ and $\hat{\theta}^{\texttt{L-EOFU}(\alpha)}_{1,T,1}$ are the first and the second component of $\hat{\theta}^{\texttt{L-EOFU}(\alpha)}_{1,T}$. After such an transformation, we are ready to apply Lindeberg-Feller central limit theorem (see, e.g., Van der

Vaart 2000). Define $z_{T^\alpha} = \frac{2}{T^\alpha}\sum_{i=1,3,\cdots,T^\alpha-1}\frac{1}{T^{2\alpha}}(2T^{2\alpha}-2-3T^\alpha i, -3T^{2\alpha}+6T^\alpha i)^\top\varepsilon_{1,i}$. The covariance matrix becomes,

$$\begin{aligned}
\mathbb{V}_{T^\alpha} &= \frac{2\tilde\sigma_1^2}{T^\alpha}\sum_{i=1,3,\cdots,T^\alpha-1}\frac{1}{T^{4\alpha}}\begin{pmatrix}(2T^{2\alpha}-2-3T^\alpha i)^2 & (2T^{2\alpha}-2-3T^\alpha i)(-3T^{2\alpha}+6T^\alpha i)\\ (2T^{2\alpha}-2-3T^\alpha i)(-3T^{2\alpha}+6T^\alpha i) & (-3T^{2\alpha}+6T^\alpha i)^2\end{pmatrix}\\
&= \frac{2\tilde\sigma_1^2}{T^\alpha}\sum_{i=1,3,\cdots,T^\alpha-1}\frac{1}{T^{4\alpha}}\begin{pmatrix}4T^{4\alpha}-12T^{3\alpha}i+9T^{2\alpha}i^2 & -6T^{4\alpha}+21T^{3\alpha}i-18T^{2\alpha}i^2\\ -6T^{4\alpha}+21T^{3\alpha}i-18T^{2\alpha}i^2 & 9T^{4\alpha}-36T^{3\alpha}i+36T^{2\alpha}i^2\end{pmatrix} \quad (T\to\infty)\\
&= \frac{2\tilde\sigma_1^2}{T^{5\alpha}}\begin{pmatrix}2T^{5\alpha}-3T^{5\alpha}+\frac{3}{2}T^{5\alpha} & -3T^{5\alpha}+\frac{21}{4}T^{5\alpha}-3T^{5\alpha}\\ -3T^{5\alpha}+\frac{21}{4}T^{5\alpha}-3T^{5\alpha} & \frac{9}{2}T^{5\alpha}-9T^{5\alpha}+6T^{5\alpha}\end{pmatrix} \quad (T\to\infty)\\
&= \tilde\sigma_1^2\begin{pmatrix}1 & -\frac{3}{2}\\ -\frac{3}{2} & 3\end{pmatrix}.
\end{aligned}$$

Now, we are going to verify the Lindeberg's condition on z_{T^α}. It is straightforward to see that $\frac{1}{T^{2\alpha}}(2T^{2\alpha}-2-3T^\alpha i, -3T^{2\alpha}+6T^\alpha i)^\top\varepsilon_{1,i}$ forms a triangle array. For simplicity, we define $x_i := \frac{1}{T^{2\alpha}}(2T^{2\alpha}-2-3T^\alpha i, -3T^{2\alpha}+6T^\alpha i)^\top \in \mathbb{R}^2$. Note that $\|x_i\|\le 5$ for all i. Then for any $\epsilon>0$,

$$\begin{aligned}
&\frac{2}{T^\alpha}\sum_{i=1,3,\cdots,T^\alpha-1}\mathbb{E}\left\{\|x_i\varepsilon_{1,i}\|^2\,\mathbb{I}\left(\|x_i\varepsilon_{1,i}\|\ge\epsilon\sqrt{\frac{T^\alpha}{2}}\right)\right\}\\
&= \frac{2}{T^\alpha}\sum_{i=1,3,\cdots,T^\alpha-1}\mathbb{E}\left\{\|x_i\|^2\,\varepsilon_{1,i}^2\mathbb{I}\left(\|x_i\|\,|\varepsilon_{1,i}|\ge\epsilon\sqrt{\frac{T^\alpha}{2}}\right)\right\}\\
&\le \frac{2}{T^\alpha}\sum_{i=1,3,\cdots,T^\alpha-1}\mathbb{E}\left\{25\varepsilon_{1,i}^2\mathbb{I}\left(5|\varepsilon_{1,i}|\ge\epsilon\sqrt{\frac{T^\alpha}{2}}\right)\right\}\\
&= \mathbb{E}\left\{25\varepsilon_{1,i}^2\mathbb{I}\left(5|\varepsilon_{1,i}|\ge\epsilon\sqrt{\frac{T^\alpha}{2}}\right)\right\},
\end{aligned}$$

where the last inequality holds due to the i.i.d assumption of $\varepsilon_{1,i}$. $25\varepsilon_{1,i}^2\mathbb{I}\left(5|\varepsilon_{1,i}|\ge\epsilon\sqrt{\frac{T^\alpha}{2}}\right)\to 0$ in probability, since $\frac{\epsilon}{5}\sqrt{\frac{T^\alpha}{2}}\to\infty$. Due to $\mathbb{E}\{25\varepsilon_{1,i}^2\}=25\tilde\sigma_1^2<\infty$, we can apply dominated convergence theorem and get $\mathbb{E}\left\{25\varepsilon_{1,i}^2\mathbb{I}\left(5|\varepsilon_{1,i}|\ge\epsilon\sqrt{\frac{T^\alpha}{2}}\right)\right\}\to 0$. Thus, the Lindeberg's condition holds. We can safely draw the conclusion that

$$\sqrt{\frac{T^\alpha}{2}}z_{T^\alpha}\xrightarrow{d}\mathcal{N}\left(0,\tilde\sigma_1^2\begin{pmatrix}1 & -\frac{3}{2}\\ -\frac{3}{2} & 3\end{pmatrix}\right).$$

Therefore, by Slutsky's theorem, we can have that

$$\sqrt{T^\alpha}\begin{pmatrix}\hat\theta_{1,T,0}^{\texttt{L-EOFU}(\alpha)}-\theta_{1,0}\\ T^\alpha(\hat\theta_{1,T,1}^{\texttt{L-EOFU}(\alpha)}-\theta_{1,1})\end{pmatrix} = \frac{2\sqrt{2}T^{2\alpha}}{T^{2\alpha}-4}\sqrt{\frac{T^\alpha}{2}}z_{T^\alpha}\xrightarrow{d}\mathcal{N}\left(0,\tilde\sigma_1^2\begin{pmatrix}8 & -12\\ -12 & 24\end{pmatrix}\right).$$

By exactly the same procedures, we can obtain that

$$\sqrt{T^\alpha}\begin{pmatrix}\hat\theta_{2,T,0}^{\texttt{L-EOFU}(\alpha)}-\theta_{2,0}\\ T^\alpha(\hat\theta_{2,T,1}^{\texttt{L-EOFU}(\alpha)}-\theta_{2,1})\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(0,\tilde\sigma_2^2\begin{pmatrix}8 & -12\\ -12 & 24\end{pmatrix}\right).$$

Since $\hat\theta_{1,T}^{\texttt{L-EOFU}(\alpha)}$ and $\hat\theta_{2,T}^{\texttt{L-EOFU}(\alpha)}$ are independent, we can directly put them together and finish the proof. Q.E.D.

EC.1.6. Proof to Theorem 4

The proof follows that to Theorem 2. One can easily replicate everything till Eq. (EC.19), which is

$$\inf_{\hat\theta_T}\max_{\nu\in\mathcal{E}_1}\mathbb{P}^\pi_\nu\left(\left\|\hat\theta_T-\theta_\nu\right\|_1\geq f(T)\right)\geq\frac{1}{2}\left[1-\sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}^\pi_{\nu_1},\mathbb{P}^\pi_{\nu_2})}\right].$$

In the setting with growing noise, the $\mathrm{KL}(\mathbb{P}^\pi_{\nu_1},\mathbb{P}^\pi_{\nu_2})$ can be decompose as

$$\mathrm{KL}(\mathbb{P}^\pi_{\nu_1},\mathbb{P}^\pi_{\nu_2})=\sum_{t=1}^T\frac{2f(T)^2}{\sigma_0^2t^{2\rho}}\mathbb{E}^\pi_{\nu_1}[\mathbb{I}_{A_t=2}]=\frac{2f(T)^2}{\sigma_0^2}\sum_{t=1}^T\frac{\mathbb{P}^\pi_{\nu_1}(A_t=2)}{t^{2\rho}}.$$

Since we are considering $\pi\in\Pi_\beta$, there exists a constant $c_4>0$ such that

$$\sum_{t=1}^T t\mathbb{P}^\pi_{\nu_1}(A_t=2)\leq c_5T^\beta.$$

Thus, the KL divergence can be upper bounded by the optimal value if the following optimization problem.

$$\begin{aligned}\max_{x_1,\cdots,x_T}\ &\sum_{t=1}^T\frac{x_t}{t^{2\rho}}\\ \text{s.t.}\quad &\sum_{t=1}^T tx_t\leq c_5T^\beta\\ &0\leq x_t\leq 1\quad\forall t\in[T].\end{aligned}$$

By observation, the optimal solution of the above optimization is $x_1=\cdots=x_{\lfloor\frac{-1+\sqrt{1+8c_5T^\beta}}{2}\rfloor}=1$ and $x_{\lfloor\frac{-1+\sqrt{1+8c_5T^\beta}}{2}\rfloor+1}=(c_5T^\beta-\frac{1}{2}(\lfloor\frac{-1+\sqrt{1+8c_5T^\beta}}{2}\rfloor+1)\lfloor\frac{-1+\sqrt{1+8c_5T^\beta}}{2}\rfloor)/(\lfloor\frac{-1+\sqrt{1+8c_5T^\beta}}{2}\rfloor+1)$. Hence, if $\rho<1/2$,

$$\sum_{t=1}^T\frac{\mathbb{P}^\pi_{\nu_1}(A_t=2)}{t^{2\rho}}\leq\sum_{t=1}^{\lfloor\frac{-1+\sqrt{1+8c_5T^\beta}}{2}\rfloor+1}\frac{1}{t^{2\rho}}=\mathcal{O}(T^{\frac{\beta}{2}(1-2\rho)}).$$

Therefore, we can have

$$\inf_{\hat\theta_T}\max_{\nu\in\mathcal{E}_1}\mathbb{P}^\pi_\nu\left(\left\|\hat\theta_T-\theta_\nu\right\|_1\geq f(T)\right)\geq\frac{1}{2}\left[1-\sqrt{\frac{c_6f(T)^2}{\sigma_0^2}T^{\frac{\beta}{2}(1-2\rho)}}\right].$$

By choosing $f(T)=T^{-\frac{\beta}{4}(1-2\rho)}$, we finish the proof. Q.E.D.

EC.2. Technical Details to Section 4

EC.2.1. `P-EOFU`(k,α) Design

`P-EOFU`(k,α) Design is presented in Algorithm 2.

Algorithm 2: Polynomial Exploration and OFU ($\texttt{P-EOFU}(k,\alpha)$)

Input: Controlling parameter α, instance parameters: σ_a for all $a \in \{1,2\}$, T, S, δ, k
Initialization: $\lambda = 1$, $\mathcal{D}_{1,0} = \mathcal{D}_{2,0} = \mathcal{D}_{1,T} = \mathcal{D}_{2,T} = \emptyset$, $\phi(x) = (1, x, \cdots, x^k)^\top$
for $t = 1, 2, \cdots, \lceil T^\alpha \rceil$ **do** // Phase 1: Exploration
  Select $A_t = 1$ if t is odd; otherwise select $A_t = 2$;
  Observe reward R_t;
  $\mathcal{D}_{A_t,0} \leftarrow \mathcal{D}_{A_t,0} \cup \{(\phi(t), R_t)\}$ and $\mathcal{D}_{A_t,T} \leftarrow \mathcal{D}_{A_t,T} \cup \{(\phi(t), R_t)\}$;
end for
for $t = \lceil T^\alpha \rceil + 1, \lceil T^\alpha \rceil + 2, \cdots, T$ **do** // Phase 2: OFU
  $V_{a,t} = \sum_{(\phi(i),R_i)\in\mathcal{D}_{a,T}} \phi(i)\phi(i)^\top + \lambda I$ for $a \in \{1,2\}$;
  $\gamma_{a,t} = \sigma_a \sqrt{2\ln(\frac{2\det(V_{a,t})^{1/2}\det(\lambda I)^{-1/2}}{\delta})} + \lambda^{1/2} S$ for $a \in \{1,2\}$;
  $\tilde{\theta}_{a,t} = V_{a,t}^{-1} \sum_{(\phi(i),R_i)\in\mathcal{D}_{a,T}} R_i \phi(i)$ for $a \in \{1,2\}$;
  $\tilde{R}_t(a) = \tilde{\theta}_{a,t}^\top \phi(t) + \gamma_{a,t} \|\phi(t)\|_{V_{a,t}^{-1}}$ for $a \in \{1,2\}$;
  Select $A_t = \arg\max_{a\in\{1,2\}} \tilde{R}_t(a)$;
  Observe reward R_t;
  $\mathcal{D}_{A_t,T} \leftarrow \mathcal{D}_{A_t,T} \cup \{(\phi(t), R_t)\}$;
end for
$\hat{\theta}_{a,T}^{\texttt{P-EOFU}(k,\alpha)} = (\sum_{(\phi(i),R_i)\in\mathcal{D}_{a,0}} \phi(i)\phi(i)^\top)^{-1} \sum_{(\phi(i),R_i)\in\mathcal{D}_{a,0}} R_i \phi(i)$ for $a \in \{1,2\}$;
Output: $\hat{\theta}_T^{\texttt{P-EOFU}(k,\alpha)} = \hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)} - \hat{\theta}_{2,T}^{\texttt{P-EOFU}(k,\alpha)}$;

EC.2.2. Proof to Lemma 1

By Rayleigh-Ritz theorem, $\lambda_{\min}(V_{a,0})$ is equal to the optimal value of the optimization problem, $\min_{\|x\|_2=1} x^\top V_{a,0} x$, where $x := (x_0, x_1, \cdots, x_k)^\top \in \mathbb{R}^{k+1}$. We introduce another notation $[[N]] := \{i \in \mathbb{N}^+ : i = 1 \mod 2\}$ and $n := T^\alpha$ for simplicity. From now on, we will mainly focus on the analysis for $a = 1$, since $a = 2$ follows naturally. Moreover, we observe that for $a = 1$,

$$x^\top V_{1,0} x = \sum_{t\in[[n]]} (x_0 + x_1 t + \cdots + x_k t^k)^2.$$

We denote the optimal solution to $\min_{\|x\|_2=1} x^\top V_{1,0} x$ given n as $x_k^* = (x_k^*(0), \cdots, x_k^*(k))^\top$. We claim that for any $i \in [k]$, $|x_k^*(i) n^i| = \mathcal{O}(1)$, which can proved by contradiction as follows. Suppose that there exists an $i \in [k]$, such that $|x_k^*(i) n^i| = \Theta(n^\gamma)$, where $\gamma > 0$. Then,

$$x_k^{*\top} V_{1,0} x_k^* \geq \sum_{t\in[[n]]\setminus[[n^{(i-\gamma)/i}]]} (x_k^*(0) + x_k^*(1) t + \cdots + x_k^*(k) t^k)^2 = \Omega(\sum_{t\in[[n]]\setminus[[n^{(i-\gamma)/i}]]} (x_k^*(i) t^i)^2) = \Omega(n^{2\gamma+1}). \quad \text{(EC.24)}$$

Another straightforward observation is that

$$\min_{\|x\|_2=1} x^\top V_{1,0} x \le \min_{x_0=1, x_1=\cdots=x_k=0} x^\top V_{1,0} x = n. \tag{EC.25}$$

Combing Eqs. (EC.24) and (EC.25) leads to a contradiction, when n is sufficiently large. Therefore, we can derive that $|x_k^*(i)| = \mathcal{O}(n^{-i})$ and $|x_k^*(i)t^i| = \mathcal{O}(t^i/n^i) = \mathcal{O}(t/n)$. Moreover, since $\|x_k^*\|_2 = 1$, $|x_k^*(0)| = \Theta(1)$. Therefore, there exist two constants $|c_1| = \Theta(1)$ and $|c_2| = \Theta(1)$, such that

$$\begin{aligned}
\sum_{t\in[[n]]} (x_k^*(0) + x_k^*(1)t + \cdots + x_k^*(k)t^k)^2 &= \sum_{t\in[[n]]} \left(c_1 + c_2 \frac{t}{n}\right)^2 \\
&= \frac{c_1^2 n}{2} + \frac{c_1 c_2 n}{4} + \frac{n c_2^2}{6} - \frac{1}{6n} \\
&= \left(\left(\frac{4c_1 + c_2}{4\sqrt{2}}\right)^2 + \frac{13}{96} c_2^2\right) n - \frac{1}{6n} \\
&\ge \frac{13}{96} c_2^2 n - \frac{1}{6n} = \Theta(n).
\end{aligned}$$

We finish the proof. Q.E.D.

EC.2.3. Proof to Theorem 5

Similar as before, with the union bound, we can have

$$\begin{aligned}
\mathbb{P}\left(\left\|\hat{\theta}_T^{\texttt{P-EOFU}(k,\alpha)} - \theta_\nu\right\|_1 \ge \delta_0\right) &= \mathbb{P}\left(\left\|\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)} - \hat{\theta}_{2,T}^{\texttt{P-EOFU}(k,\alpha)} - \theta_1 + \theta_2\right\|_1 \ge \delta_0\right) \\
&\le \mathbb{P}\left(\left\|\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)} - \theta_1\right\|_1 \ge \frac{\delta_0}{2}\right) + \mathbb{P}\left(\left\|\hat{\theta}_{2,T}^{\texttt{P-EOFU}(k,\alpha)} - \theta_2\right\|_1 \ge \frac{\delta_0}{2}\right).
\end{aligned} \tag{EC.26}$$

Note that since these two terms are almost symmetric. We focus on the first term of Eq. (EC.26), and the second term can be bounded by almost the same argument. For brevity, without loss of generality, we assume T^α is an even number and is sufficiently large that Lemma 1 holds. We still use the notation $[[n]]$ to denote the set of the odd numbers that are no larger than n. By the definition of $\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)}$, we can have

$$\begin{aligned}
\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)} &= V_{1,0}^{-1} \sum_{(\phi(t),R_t)\in\mathcal{D}_{1,0}} R_t \phi(t) \\
&= V_{1,0}^{-1} \sum_{t\in[[T^\alpha]]} (\theta_t^\top \phi(t) + \varepsilon_{1,t})\phi(t) \\
&= \theta_1 + V_{1,0}^{-1} \sum_{t\in[[T^\alpha]]} \varepsilon_{1,t}\phi(t).
\end{aligned}$$

Therefore, $\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)} - \theta_1 = V_{1,0}^{-1}\sum_{t\in[[T^\alpha]]} \varepsilon_{1,t}\phi(t)$. Let $a_t = V_{1,0}^{-1}\phi(t) \in \mathbb{R}^{k+1}$, and $a_t(i)$ be the i-th coordinate of a_t. Based on these notations,

$$\mathbb{P}\left(\left\|\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)} - \theta_1\right\|_1 \ge \frac{\delta_0}{2}\right) = \mathbb{P}\left(\sum_{i=0}^{k} \left|\sum_{t\in[[T^\alpha]]} \varepsilon_{1,t} a_t(i)\right| \ge \frac{\delta_0}{2}\right)$$

$$\le \sum_{i=0}^{k} \mathbb{P}\left(\left|\sum_{t\in[[T^\alpha]]} \varepsilon_{1,t} a_t(i)\right| \ge \frac{\delta_0}{2(k+1)}\right). \tag{EC.27}$$

For a specific i in the RHS of Eq. (EC.27), we apply the standard techniques for the sum of sub-Gaussian variables as follows.

$$\begin{aligned}
\mathbb{P}\left(\left|\sum_{t\in[[T^\alpha]]} \varepsilon_{1,t} a_t(i)\right| \ge \frac{\delta_0}{2(k+1)}\right) &\le 2\mathbb{P}\left(\sum_{t\in[[T^\alpha]]} \varepsilon_{1,t} a_t(i) \ge \frac{\delta_0}{2(k+1)}\right) \\
&= 2\mathbb{P}\left(\exp\left(s\sum_{t\in[[T^\alpha]]} \varepsilon_{1,t} a_t(i)\right) \ge \exp\left(\frac{s\delta_0}{2(k+1)}\right)\right) \\
&\le 2\frac{\mathbb{E}\left[\exp\left(s\sum_{t\in[[T^\alpha]]} \varepsilon_{1,t} a_t(i)\right)\right]}{\exp\left(\frac{s\delta_0}{2(k+1)}\right)} \\
&= 2\frac{\Pi_{t\in[[T^\alpha]]}\mathbb{E}\left[\exp\left(s\varepsilon_{1,t} a_t(i)\right)\right]}{\exp\left(\frac{s\delta_0}{2(k+1)}\right)} \\
&\le 2\frac{\Pi_{t\in[[T^\alpha]]}\exp(s^2 a_t^2(i)\sigma_1^2/2)}{\exp\left(\frac{s\delta_0}{2(k+1)}\right)} \\
&= 2\exp\left(s^2\sigma_1^2\frac{\sum_{t\in[[T^\alpha]]} a_t^2(i)}{2} - \frac{s\delta_0}{2(k+1)}\right) \\
&= 2\exp\left(-\frac{\delta_0^2}{8(k+1)^2\sigma_1^2\sum_{t\in[[T^\alpha]]} a_t^2(i)}\right) \\
&\le 2\exp\left(-\frac{\delta_0^2}{8(k+1)^2\sigma_1^2\sum_{t\in[[T^\alpha]]} \|a_t\|_2^2}\right),
\end{aligned}$$

where the first inequality due to the symmetry, the second inequality holds from Markov inequality, the second equation adopts the independence of $\varepsilon_{1,t}$, the second last line sets $s = \frac{\delta^2}{2(k+1)\sigma_1^2(\sum_{t\in[[T^\alpha]]} a_t^2(i))}$ and the last inequality holds due to the fact that $a_t^2(i) \le \|a_t\|_2^2$. Therefore, Eq. (EC.27) can be further bounded by

$$\mathbb{P}\left(\left\|\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)} - \theta_1\right\|_1 \ge \frac{\delta_0}{2}\right) \le 2(k+1)\exp\left(-\frac{\delta_0^2}{8(k+1)^2\sigma_1^2\sum_{t\in[[T^\alpha]]} \|a_t\|_2^2}\right). \tag{EC.28}$$

What remains to be controlled is $\sum_{t\in[[T^\alpha]]} \|a_t\|_2^2$. By using the definition of a_t, we can have

$$\begin{aligned}
\sum_{t\in[[T^\alpha]]} \|a_t\|_2^2 &= \sum_{t\in[[T^\alpha]]} a_t^\top a_t \\
&= \sum_{t\in[[T^\alpha]]} (V_{1,0}^{-1}\phi(t))^\top V_{1,0}^{-1}\phi(t) \\
&= \sum_{t\in[[T^\alpha]]} \mathrm{trace}\left(\phi(t)^\top V_{1,0}^{-1} V_{1,0}^{-1}\phi(t)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t\in[[T^\alpha]]} \operatorname{trace}\left(V_{1,0}^{-1}V_{1,0}^{-1}\phi(t)\phi(t)^\top\right) \\
&= \operatorname{trace}\left(V_{1,0}^{-1}V_{1,0}^{-1}\sum_{t\in[[T^\alpha]]}\phi(t)\phi(t)^\top\right) \\
&= \operatorname{trace}(V_{1,0}^{-1}) \\
&\le \frac{k+1}{\lambda_{\min}(V_{1,0})} \\
&\le \frac{c_k(k+1)}{T^\alpha},
\end{aligned} \tag{EC.29}
$$

where the third equality holds due to $(V_{1,0}^{-1}\phi(t))^\top V_{1,0}^{-1}\phi(t)\in\mathbb{R}$, the fourth inequality due to the cyclic property of the trace of a product of matrices, the last equality follows from the fact that $\operatorname{trace}(V_{1,0}^{-1})\le(k+1)\lambda_{\max}(V_{1,0}^{-1})=(k+1)/\lambda_{\min}(V_{1,0})$, and the last inequality is thanks to Lemma 1. By plugging Eq. (EC.29) into Eq. (EC.28), we arrive at the point that

$$
\mathbb{P}\left(\left\|\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)}-\theta_1\right\|_1\ge\frac{\delta_0}{2}\right)\le 2(k+1)\exp\left(-\frac{\delta_0^2T^\alpha}{8c_k(k+1)^3\sigma_1^2}\right).
$$

Following the same argument, we can derive

$$
\mathbb{P}\left(\left\|\hat{\theta}_{2,T}^{\texttt{P-EOFU}(k,\alpha)}-\theta_2\right\|_1\ge\frac{\delta_0}{2}\right)\le 2(k+1)\exp\left(-\frac{\delta_0^2T^\alpha}{8c_k(k+1)^3\sigma_2^2}\right).
$$

Finally, by putting everything together, we draw the conclusion that

$$
\mathbb{P}\left(\left\|\hat{\theta}_{T}^{\texttt{P-EOFU}(k,\alpha)}-\theta_\nu\right\|_1\ge\delta_0\right)\le 4(k+1)\exp\left(-\frac{\delta_0^2T^\alpha}{8c_k(k+1)^3(\sigma_1^2\vee\sigma_2^2)}\right).
$$

We finish the proof. Q.E.D.

EC.2.4. Proof to Theorem 6

For simplicity, we derive the CLT for $a=1$ here and the other arm can follow naturally. First, we know that

$$
\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)}-\theta_1=V_{1,0}^{-1}\sum_{t\in[[T^\alpha]]}\varepsilon_t\phi(t),
$$

where $V_{1,0}=\sum_{t\in[[T^\alpha]]}\phi(t)\phi(t)^\top$. We first analyze the convergence of $D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1}$. The $(i,j)-$th element of $V_{1,0}$ is $(V_{1,0})_{ij}=\sum_{t\in[[T^\alpha]]}t^{i+j}$. Then, the element of $D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1}$ is

$$
(D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})_{ij}=\frac{\sum_{t\in[[T^\alpha]]}t^{i+j}}{T^{\alpha(i+\frac{1}{2})}T^{\alpha(j+\frac{1}{2})}}=\frac{\sum_{t\in[[T^\alpha]]}t^{i+j}}{T^{\alpha(i+j+1)}}.
$$

Now, let us define $S_m(k)=\sum_{t=1}^m t^k$. Then, $\sum_{t\in[[T^\alpha]]}t^{i+j}=S_{T^\alpha}(i+j)-2^{i+j}S_{\lfloor T^\alpha/2\rfloor}(i+j)$. By Faulhaber's formula, we can have

$$
\lim_{T\to\infty}\frac{S_{T^\alpha}(i+j)}{T^{\alpha(i+j+1)}}=\frac{1}{i+j+1},
$$

$$
\lim_{T\to\infty}\frac{2^kS_{\lfloor T^\alpha/2\rfloor}(k)}{T^{\alpha(i+j+1)}}=\lim_{T\to\infty}\frac{2^{i+j}(\frac{T^\alpha}{2})^{i+j+1}}{(i+j+1)T^{\alpha(i+j+1)}}=\frac{1}{2(i+j+1)}.
$$

Thus, $(D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})_{ij} \to \frac{1}{2(i+j+1)}$ and $D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1} \to Q$ as $T \to \infty$.

Now, we can have

$$\begin{aligned} D_{T^\alpha}(\hat{\theta}_{1,T}^{\texttt{P-EOFU}(k,\alpha)} - \theta_1) &= D_{T^\alpha}V_{1,0}^{-1}\sum_{t\in[[T^\alpha]]}\varepsilon_t\phi(t) \\ &= (D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})^{-1}\sum_{t\in[[T^\alpha]]}\varepsilon_t D_{T^\alpha}^{-1}\phi(t), \end{aligned}$$

Note that $\varepsilon_t D_{T^\alpha}^{-1}\phi(t)$ is independent zero-mean random vector. The variance is

$$\mathbb{V}_{T^\alpha} = \sigma_1^2\sum_{t\in[[T^\alpha]]}D_{T^\alpha}^{-1}\phi(t)\phi(t)^\top D_{T^\alpha}^{-1} = \sigma_1^2 D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1} \to \sigma_1^2 Q.$$

Additionally, we define $u_\alpha = \max_{t\in[[T^\alpha]]}\|(D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})^{-1}D_{T^\alpha}^{-1}\phi(t)\|$. Observe that

$$\|(D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})^{-1}D_{T^\alpha}^{-1}\phi(t)\| \le \lambda_{\max}\left((D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})^{-1}\right)\|D_{T^\alpha}^{-1}\phi(t)\| \le \lambda_{\max}\left((D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})^{-1}\right)\frac{k+1}{\sqrt{T^\alpha}}.$$

Since $D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1} \to Q$, for sufficiently large T, we can always have $u_\alpha \le 2(k+1)\lambda_{\max}(Q^{-1})/\sqrt{T^\alpha}$.

Now, we are going to verify the Linderber's condition to establish the central limit theorem. For any $\delta > 0$,

$$\begin{aligned} &\sum_{t\in[[T^\alpha]]}\mathbb{E}\left(\varepsilon_t^2\left\|(D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})^{-1}D_{T^\alpha}^{-1}\phi(t)\right\|^2\mathbb{I}(|\varepsilon_t|\left\|(D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})^{-1}D_{T^\alpha}^{-1}\phi(t)\right\| > \delta)\right) \\ &\le \sum_{t\in[[T^\alpha]]}\left\|(D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})^{-1}D_{T^\alpha}^{-1}\phi(t)\right\|^2\mathbb{E}\left(\varepsilon_t^2\mathbb{I}(|\varepsilon_t|u_\alpha > \delta)\right) \\ &= \mathbb{E}\left(\varepsilon_1^2\mathbb{I}(|\varepsilon_1|u_\alpha > \delta)\right)\sum_{t\in[[T^\alpha]]}\left\|(D_{T^\alpha}^{-1}V_{1,0}D_{T^\alpha}^{-1})^{-1}D_{T^\alpha}^{-1}\phi(t)\right\|^2 \\ &= \mathbb{E}\left(\varepsilon_1^2\mathbb{I}(|\varepsilon_1|u_\alpha > \delta)\right)\sum_{t\in[[T^\alpha]]}\left\|D_{T^\alpha}V_{1,0}^{-1}\phi(t)\right\|^2 \\ &= \mathbb{E}\left(\varepsilon_1^2\mathbb{I}(|\varepsilon_1|u_\alpha > \delta)\right)\sum_{t\in[[T^\alpha]]}\text{trace}\left(\phi^\top(t)V_{1,0}^{-1}D_{T^\alpha}D_{T^\alpha}V_{1,0}^{-1}\phi(t)\right) \\ &= \mathbb{E}\left(\varepsilon_1^2\mathbb{I}(|\varepsilon_1|u_\alpha > \delta)\right)\sum_{t\in[[T^\alpha]]}\text{trace}\left(D_{T^\alpha}V_{1,0}^{-1}\phi(t)\phi^\top(t)V_{1,0}^{-1}D_{T^\alpha}\right) \\ &= \mathbb{E}\left(\varepsilon_1^2\mathbb{I}(|\varepsilon_1|u_\alpha > \delta)\right)\text{trace}\left(D_{T^\alpha}V_{1,0}^{-1}D_{T^\alpha}\right) \\ &\le \mathbb{E}\left(\varepsilon_1^2\mathbb{I}(|\varepsilon_1|2(k+1)\lambda_{\max}(Q^{-1}) > \delta\sqrt{T^\alpha})\right)\text{trace}\left(D_{T^\alpha}V_{1,0}^{-1}D_{T^\alpha}\right) \qquad \text{(for sufficiently large T)} \\ &\to 0, \end{aligned}$$

where the first equation holds since $\varepsilon_t^2\mathbb{I}(|\varepsilon_t|u_\alpha > \delta)$ is i.i.d., the last inequality holds for sufficiently large T, and the convergence is guaranteed by dominated convergence theorem and $\text{trace}\left(D_{T^\alpha}V_{1,0}^{-1}D_{T^\alpha}\right) \to \text{trace}(Q^{-1})$. Therefore, by Lindeberg-Feller Central Limit Theorem, we finish the proof. Q.E.D.

EC.2.5. Proof to Theorem 7

We again are going to bound the regrets for each stage separately. For simplicity, we assume T^α is an integer. For the first exploration stage, since $\|\theta_a\| \le S$, we can bound the first-stage regret $\mathcal{R}_\nu^{\texttt{P-EOFU}(k,\alpha),1}(T)$ as

$$\begin{aligned}\mathcal{R}_\nu^{\texttt{P-EOFU}(k,\alpha),1}(T) &\le \sum_{t=1}^{T^\alpha}\left|(\theta_1-\theta_2)^\top\phi(t)\right| \le 2ST^\alpha + 2Sk\sum_{t=1}^{T^\alpha} t^k \le 2T^\alpha S + 2Sk\int_0^{T^\alpha+1} t^k dt \\ &= 2T^\alpha S + \frac{2Sk}{k+1}(T^\alpha+1)^{k+1} \le \frac{4Sk}{k+1}(T^\alpha+1)^{k+1}. \end{aligned} \tag{EC.30}$$

For the second stage, We repeat Eq. (EC.10) for a higher-dimension version. By Theorem 2 in Abbasi-Yadkori et al. (2011), with probability at least $1-\delta$, for all $t\in[T]$ and $a\in\{1,2\}$,

$$\tilde{R}_t(a) = \tilde{\theta}_{a,t}^\top\phi(t) + \gamma_{a,t}\|\phi(t)\|_{V_{a,t}^{-1}} \ge \theta_a^\top\phi(t).$$

Adopting exactly the same procedure as Eq. (EC.11), we can directly have

$$\mathcal{R}_\nu^{\texttt{P-EOFU}(k,\alpha),2}(T) = \sum_{t=T^\alpha+1}^{T} r_t \le \sum_{t=1}^{T} 2\gamma\left[\|\phi(t)\|_{V_{2,t}^{-1}}\mathbb{I}_{A_t=2}\right] + \sum_{t=1}^{T} 2\gamma\left[\|\phi(t)\|_{V_{1,t}^{-1}}\mathbb{I}_{A_t=1}\right],$$

where $\gamma = (\sigma_1\vee\sigma_2)\sqrt{2\ln(4(T+1)^{2k+1}) - 2\ln((2k+1)\lambda^{\frac{k+1}{2}}\delta)} + \lambda^{1/2}S$, since

$$\det(V_{1,T}) \le \left(\frac{\mathrm{trace}(V_{1,T})}{k+1}\right)^{k+1} \le \left(\frac{\sum_{t=1}^T\sum_{i=0}^k t^{2i}}{k+1}+\lambda\right)^2 \le \left(2\sum_{t=1}^T t^{2k}\right)^2 \le \frac{4(T+1)^{4k+2}}{(2k+1)^2}.$$

In order to bound $\sum_{t=1}^T \|\phi(t)\|_{V_{1,t}^{-1}}\mathbb{I}_{A_t=1}$, we adopt the same notation as Eq. (EC.12) (i.e., $\Psi_1 = \{t : A_t = 1\}$), and index the elements in Ψ_1 by $0 := t_{1,0} < t_{1,1} < t_{1,2} < \cdots < t_{1,|\Psi_1|} \le T$. Eq. (EC.12) still holds here, and we repeat it as follows,

$$\sum_{t=1}^T \|\phi(t)\|_{V_{1,t}^{-1}}\mathbb{I}_{A_t=1} = \sum_{i=1}^{|\Psi_1|}\sqrt{\frac{\det(V_{1,t_{1,i}})}{\det(V_{1,t_{1,i-1}})}-1}. \tag{EC.31}$$

The corresponding optimization problem used to bound the RHS of Eq. (EC.31) is

$$\begin{aligned} \max_{h_1,\cdots,h_{|\Psi_1|}} \quad & \sum_{i=1}^{|\Psi_1|}\sqrt{h_i-1} \\ \text{s.t.} \quad & \Pi_{i=1}^{|\Psi_1|}h_i \le |\Psi_1|^{k+1}T^{2k(k+1)}, \\ & h_i \ge 1 \qquad \forall i\in[|\Psi_1|]. \end{aligned} \tag{EC.32}$$

The first constraint exists because

$$\det(V_{1,t_{1,|\Psi_1|}}) \le \left(\frac{\mathrm{trace}(V_{1,t_{1,|\Psi_1|}})}{k+1}\right)^{k+1} = \left(\lambda + \frac{\sum_{i=0}^k\sum_{t\in\Psi_1}t^{2i}}{k+1}\right)^{k+1} \le |\Psi_1|^{k+1}T^{2k(k+1)}.$$

Therefore, the RHS of Eq. (EC.31) can be bounded by

$$\sum_{i=1}^{|\Psi_1|}\sqrt{\frac{\det(V_{1,t_{1,i}})}{\det(V_{1,t_{1,i-1}})}-1} \le |\Psi_1|\sqrt{(|\Psi_1|T^{2k})^{\frac{k+1}{|\Psi_1|}}-1}$$
$$\le |\Psi_1|\sqrt{(2^{\frac{2k}{\alpha}}|\Psi_1|^{1+\frac{2k}{\alpha}})^{\frac{k+1}{|\Psi_1|}}-1}$$
$$\le |\Psi_1|\sqrt{(2|\Psi_1|)^{(1+\frac{2k}{\alpha})\frac{k+1}{|\Psi_1|}}-1}. \tag{EC.33}$$

If $T^\alpha \ge (1+\frac{2k}{\alpha})^2(k+1)^2$, it is straightforwards to verify that $\frac{2\ln T^\alpha}{T^\alpha} \le \frac{1}{(1+\frac{2k}{\alpha})(k+1)}$, which leads to $(1+\frac{2k}{\alpha})(k+1)\frac{2\ln T^\alpha}{T^\alpha} \le 1$. We can then guarantee that $(1+\frac{2k}{\alpha})(k+1)\frac{\ln 2|\Psi_1|}{|\Psi_1|} \le 1$ since $|\Psi_1| \ge \frac{T^\alpha}{2}$. By the fact that $x \le \ln(1+2x)$ for $x\in[0,1]$, we have

$$\left(1+\frac{2k}{\alpha}\right)(k+1)\frac{\ln 2|\Psi_1|}{|\Psi_1|} \le \ln\left(1+2\left(1+\frac{2k}{\alpha}\right)(k+1)\frac{\ln 2|\Psi_1|}{|\Psi_1|}\right).$$

Therefore, Eq. (EC.33) can be further bounded by,

$$\sum_{i=1}^{|\Psi_1|}\sqrt{\frac{\det(V_{1,t_{1,i}})}{\det(V_{1,t_{1,i-1}})}-1} \le |\Psi_1|\sqrt{2\left(1+\frac{2k}{\alpha}\right)(k+1)\frac{\ln 2|\Psi_1|}{|\Psi_1|}} = \sqrt{2\left(1+\frac{2k}{\alpha}\right)(k+1)|\Psi_1|\ln(2|\Psi_1|)}.$$

Following exactly the same procedure for arm 2, we can achieve

$$\sum_{t=1}^{T}\|\phi(t)\|_{V_{2,t}^{-1}}\mathbb{I}_{A_t=2} \le \sqrt{2\left(1+\frac{2k}{\alpha}\right)(k+1)|\Psi_2|\ln(2|\Psi_2|)},$$

where $|\Psi_2| := \{t : A_t = 2\}$. Therefore, since $|\Psi_1| \le T$ and $|\Psi_2| \le T$, we derive

$$\mathcal{R}_\nu^{\texttt{P-EOFU}(k,\alpha),2}(T) \le 4\gamma\sqrt{2\left(1+\frac{2k}{\alpha}\right)(k+1)T\ln(2T)}. \tag{EC.34}$$

We complete the proof. Q.E.D.

EC.2.6. Proof to Theorem 8

Without loss of generality, we specify $S=2$ in this proof. We first construct two instances ν_1 and ν_2 as follows. For ν_1, we set arm 1 to be with $\theta_1^{(1)} = (0,\cdots,0)$ and arm 2 to be with $\theta_2^{(1)} = (0,\cdots,0,-1)$. For ν_2, arm 1 remains to be with $\theta_1^{(2)} = (0,\cdots,0)$ and arm 2 to be with $\theta_2^{(1)} = (-2f(T),0,\cdots,0,-1)$. Thus, $\theta_{\nu_1} = (0,\cdots,0,1)$ and $\theta_{\nu_2} = (2f(T),0,\cdots,0,1)$. All the noises for ν_1 and ν_2 are assumed to be σ_0^2-Gaussain. We still adopt the minimum distance test $\psi(\hat\theta_T)$ with respect to l_1 norm, which is associated to $\hat\theta_T$ by $\psi(\hat\theta_T) = \arg\min_{i=1,2}\left\|\hat\theta_T-\theta_{\nu_i}\right\|_1$. Following exactly the same argument as Eq. (EC.19), we can have

$$\inf_{\hat\theta_T}\max_{\nu\in\mathcal{E}_1}\mathbb{P}_\nu^\pi\left(\left\|\hat\theta_T-\theta_\nu\right\|_1 \ge f(T)\right) \ge \frac{1}{2}\left[1-\sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}_{\nu_1}^\pi,\mathbb{P}_{\nu_2}^\pi)}\right]. \tag{EC.35}$$

With the chain rule of KL-divergence, since ν_1 and ν_2 are only different in arm 2, we can decompose $\mathrm{KL}(\mathbb{P}^\pi_{\nu_1}, \mathbb{P}^\pi_{\nu_2})$ as follows:

$$\mathrm{KL}(\mathbb{P}^\pi_{\nu_1}, \mathbb{P}^\pi_{\nu_2}) = \sum_{t=1}^{T} \mathbb{E}^\pi_{\nu_1}[\mathrm{KL}(P_{1,A_t}, P_{2,A_t})] = \sum_{t=1}^{T} \frac{2f(T)^2}{\sigma_0^2} \mathbb{E}^\pi_{\nu_1}[\mathbb{I}_{A_t=2}] = \frac{2f(T)^2}{\sigma_0^2} \sum_{t=1}^{T} \mathbb{P}^\pi_{\nu_1}(A_t=2).$$

Moreover, because of $\pi \in \Pi_\beta$ and $\mathbb{E}\left[\mathcal{R}^\pi_{\nu_1}(T)\right] = \sum_{t=1}^T t^k \mathbb{P}^\pi_{\nu_1}(A_t=2)$, we can have the constraint that there exists a constant $c_4 > 0$,

$$\sum_{t=1}^{T} t^k \mathbb{P}^\pi_{\nu_1}(A_t=2) \le c_4 T^\beta.$$

Then, $\sum_{t=1}^T \mathbb{P}^\pi_{\nu_1}(A_t=2)$ can be bounded by the optimal value of the following optimization problem.

$$\begin{aligned} \max_{x_1,\cdots,x_T} \quad & \sum_{t=1}^{T} x_t \\ \text{s.t.} \quad & \sum_{t=1}^{T} t^k x_t \le c_4 T^\beta \\ & 0 \le x_t \le 1 \quad \forall t \in [T]. \end{aligned}$$

The structure of the optimal solution of the optimization is that there exists an i such that $x_0 = \cdots = x_i = 1$ and $0 \le x_{i+1} = c_4 T^\beta - \sum_{t=1}^i t^k \le 1$. The reason is that the first constraint is a weighted sum of x_t with the increasing weights (with respect to t). Therefore, if we want to maximize the sum of x_t, we want to make x_t to be as large as possible when its corresponding weight is small. In the other words, i satisfies,

$$\sum_{t=1}^{i} t^k \le c_4 T^\beta \le \sum_{t=1}^{i+1} t^k.$$

Since $\sum_{t=1}^i t^k = \Theta(i^{k+1})$, we can know that there exists a constant c_5 such that $i+1 = c_5 T^{\frac{\beta}{k+1}}$. Therefore, $\sum_{t=1}^T \mathbb{P}^\pi_{\nu_1}(A_t=2)$ is upper bounded by $c_5 T^{\frac{\beta}{k+1}}$. Then, Eq. (EC.35) becomes

$$\inf_{\hat\theta_T} \max_{\nu\in\mathcal{E}_1} \mathbb{P}^\pi_\nu\left(\left\|\hat\theta_T - \theta_\nu\right\|_1 \ge f(T)\right) \ge \frac{1}{2}\left[1 - \sqrt{\frac{c_5}{2}\frac{2f(T)^2}{\sigma_0^2} T^{\frac{\beta}{k+1}}}\right].$$

By choosing $f(T) = \frac{2\sigma_0\xi}{c_5^{1/2} T^{\beta/(2k+2)}}$, we finally derive

$$\inf_{\hat\theta_T} \max_{\nu\in\mathcal{E}_1} \mathbb{P}^\pi_\nu\left(\left\|\hat\theta_T - \theta_\nu\right\|_1 \ge c_6^{-\frac{1}{2}} \sigma_0 \xi T^{-\frac{\beta}{(2k+2)}}\right) \ge \frac{1}{2} - \xi.$$

Moreover, by choosing $\xi = \frac{1}{4}$, we have

$$\inf_{\hat\theta_T} \max_{\nu\in\mathcal{E}_1} \mathbb{E}^\pi_\nu\left[\left\|\hat\theta_T - \theta_\nu\right\|_1\right] \ge \frac{\sigma_0}{16 c_6^{\frac{1}{2}}} T^{-\frac{\beta}{(2k+2)}}.$$

We finish the proof. Q.E.D.

EC.3. Useful Technical Tools

Lemma EC.1 **(Theorem 1 in Abbasi-Yadkori et al. 2011)**. *Let $\{F_t\}_{t=0}^{\infty}$ be a filtration and $\{\eta_t\}_{t=1}^{\infty}$ be independent R-subGaussain for some $R \geq 0$. Let $\{X_t\}_{t=1}^{\infty}$ be an $\mathcal{R}^d$-valued stochastic process such that X_t is F_{t-1}-measurable. Assume that V is a $d \times d$ positive definite matrix. For any $t \geq 0$, define $\bar{V}_t = V + \sum_{s=1}^{t} X_s X_s^\top$ and $S_t = \sum_{s=1}^{t} \eta_s X_s$. Then, for any $\delta > 0$, with probability at least $1-\delta$, for all $t \geq 0$,*

$$\|S_t\|_{\bar{V}_t^{-1}}^2 \leq 2R^2 \ln\left(\frac{\det(\bar{V}_t)^{1/2}\det(V)^{-1/2}}{\delta}\right).$$

Lemma EC.2 **(Theorem 2 in Abbasi-Yadkori et al. 2011)**. *Based on the notations and assumptions of Lemma EC.1, let $V = I\lambda$, $\lambda > 0$, define $Y_t = \langle X_t, \theta_* \rangle + \eta_t$ and assume $\|\theta_*\|_2 \leq S$. Let $\hat{\theta}$ be the l^2-regularized least-squares estimate of θ_* with regularization parameter λ. Then, for any $\delta > 0$, with probability at least $1-\delta$, for all $t \geq 0$, θ_* lies in the set*

$$C_t = \left\{\theta \in \mathbb{R}^d : \left\|\hat{\theta}_t - \theta\right\|_{\bar{V}_t} \leq R\sqrt{2\ln\left(\frac{\det(\bar{V}_t)^{1/2}\det(\lambda I)^{-1/2}}{\delta}\right)} + \lambda^{1/2} S\right\}.$$

Lemma EC.3 **(Bretagnolle-Huber inequality in Bretagnolle and Huber 1979)**. *Let P and Q be the probability measures on the same measurable space $(\Omega, \mathcal{F})$, and let $A \in \mathcal{F}$ be an arbitrary event. Then,*

$$P(A) + Q(A^c) \geq \frac{1}{2}\exp(-KL(P,Q)),$$

where A^c is the complement of A.

Lemma EC.4 **(Neyman-Pearson Lemma in Neyman and Pearson 1933)**. *Let $\mathbb{P}_0$ and $\mathbb{P}_1$ be two probability measures. Then for any test ψ, it holds*

$$\mathbb{P}_0(\psi = 1) + \mathbb{P}_1(\psi = 0) \geq \int \min(p_0, p_1).$$

The equality holds for the Likelihood Ratio test $\psi^\star = \mathbb{I}(p_1 \geq p_0)$. Moreover,

$$\inf_{\psi}[\mathbb{P}_0(\psi = 1) + \mathbb{P}_1(\psi = 0)] = 1 - TV(\mathbb{P}_0, \mathbb{P}_1).$$