A Appendix

A.1 Proof of Lemma 1

Proof. The proof techniques basically follows [7]. However, since the EXP3 layer and contextual bandit layer are coupled, the result in [7] cannot be directly applied to show our result. We make modifications of the proofs in [7] below.

We first reload some notations in this proof: at time t we are given all the previous information $\mathcal{F}_{t-1}$ generated from using our auto-tune framework shown in Algorithm 1, and then pull an arm according to some exploration hyper-parameter α. Therefore, for convenience we could safely omit $\mathcal{F}_{t-1}$ here, and denote $a_t(\alpha) := a_t(\alpha|\mathcal{F}_{t-1})$ and $X_t(\alpha) := X_t(\alpha|\mathcal{F}_{t-1})$ as the arm pulled and its corresponding feature vector at round t. Furthermore, if arm $a_t(\alpha_j)$ is pulled at round t, we define the corresponding mean reward as $\mu_t(\alpha_j) = \mu\left(X_t(\alpha_j)^T\theta\right)$. The corresponding observed sample reward is $y_t(\alpha_j) = \mu_t(\alpha_j) + \epsilon_{t,j}$, where $\epsilon_{t,j}$ denotes the hypothetical random noise at round t if arm $a_t(\alpha_j)$ is pulled. Note that $\epsilon_t = \epsilon_{t,i_t}$ since $a_t(\alpha_{i_t})$ is the arm pulled by our algorithm and ϵ_t is the associated random noise. By definition, $Y_t = y_t(\alpha_{i_t})$. From the definition of $\hat{y}_t(j)$ in Algorithm 1, we have $\hat{y}_t(j) = y_t(\alpha_j)/p_j(t)$ if $j = i_t$. Otherwise $\hat{y}_t(j) = 0$. Then $w_j(t+1) = w_j(t)\text{exp}(\frac{\beta}{n}\hat{y}_t(i))$ according to Algorithm 1.

Given all the information in the past $\mathcal{F}_{t-1}$, $\left(\hat{\theta}_t, V_t, p_j(t), w_j(t)\right)$ are fixed. Since $0 \leq y_t(\alpha_j) \leq 1$, we have

$$\mathbb{E}\left[\sum_{i=1}^{n} p_i(t)\hat{y}_t(i)|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[p_{i_t}(t)\frac{y_t(\alpha_{i_t})}{p_{i_t}(t)}|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\mu_t(\alpha_{i_t})|\mathcal{F}_{t-1}\right] \tag{4}$$

$$\mathbb{E}\left[\sum_{i=1}^{n} p_i(t)\hat{y}_t(i)^2|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[p_{i_t}(t)\frac{y_t(\alpha_{i_t})}{p_{i_t}(t)}\hat{y}_t(i_t)|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[y_t(\alpha_{i_t})\hat{y}_t(i_t)|\mathcal{F}_{t-1}\right]$$

$$\leq \mathbb{E}\left[\hat{y}_t(i_t)|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\sum_{i=1}^{n}\hat{y}_t(i)|\mathcal{F}_{t-1}\right] \tag{5}$$

$$= \sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}\left[\hat{y}_t(i)|\sigma(\mathcal{F}_{t-1}, \epsilon_{t,i}, a_t(\alpha_i))\right]|\mathcal{F}_{t-1}\right] \tag{6}$$

$$= \sum_{i=1}^{n}\mathbb{E}\left[y_t(\alpha_i)|\mathcal{F}_{t-1}\right] \tag{7}$$

$$= \sum_{i=1}^{n}\mathbb{E}\left[\mu_t(\alpha_i)|\mathcal{F}_{t-1}\right]. \tag{8}$$

Equation 5 holds since $\hat{y}_t(i) \neq 0$ only when $i = i_t$. In Equation 6, $\sigma(\mathcal{F}_{t-1}, \epsilon_{t,i}, a_t(\alpha_i))$ is the smallest σ-algebra induced by $\mathcal{F}_{t-1}, \epsilon_{t,i}$, and $a_t(\alpha_i)$. Equation 7 holds since $\hat{y}_t(i) = y_t(\alpha_i)/p_i(t)\mathbb{1}(i = i_t)$. Meanwhile, since given the hyper-parameter to be used at round t as α_i, the arm to be pulled $a_t(\alpha_i)$ follows a fixed distribution and does not affect the distribution of i_t, so $i = i_t$ is still with probability $p_i(t)$. Now we are ready to use the above results to prove the lemma. Define $W_t = \sum_{i=1}^{n} w_i(t)$. We find the lower bound and upper bound of $\mathbb{E}[\log\frac{W_{T+1}}{W_1}]$ below.

Lower bound. Since $w_i(1) = 1$ for all i, $\mathbb{E}[\log\frac{W_{T+1}}{W_1}] \geq \mathbb{E}[\log w_i(T+1)] - \log n$ for all $i \in [n]$. We take a look at $\mathbb{E}\left[\log\frac{w_i(t+1)}{w_i(t)}\right]$ below.

$$\mathbb{E}\left[\log\frac{w_i(t+1)}{w_i(t)}|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\log\left[\frac{w_i(t)}{w_i(t)}\text{exp}\left(\frac{\beta}{n}\hat{y}_t(i)\right)\right]|\mathcal{F}_{t-1}\right]$$

$$= \mathbb{E}\left[\frac{\beta}{n}\hat{y}_t(i)|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\frac{\beta}{n}y_t(i)|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\frac{\beta}{n}\mu_t(\alpha_i)|\mathcal{F}_{t-1}\right]$$

The third "=" in the above is due to the same reason as in Equation 7. Take expectation on both sides and sum over t, we get

$$\mathbb{E}\left[\log w_i(T+1)\right] = \frac{\beta}{n}\sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_i)\right]$$

Therefore, for all $i = 1, \ldots, n$,

$$\mathbb{E}\left[\log\frac{W_{T+1}}{W_1}\right] \geq \frac{\beta}{n}\sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_i)\right] - \log n. \tag{9}$$

Upper bound. On the other hand, let's look at $\mathbb{E}[\log\frac{W_{t+1}}{W_t}]$:

$$\begin{aligned}
&\mathbb{E}\left[\log\frac{W_{t+1}}{W_t}|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\log\sum_{i=1}^{n}\frac{w_i(t+1)}{W_t}|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\log\sum_{i=1}^{n}\frac{w_i(t)}{W_t}\exp\left(\frac{\beta}{n}\hat{y}_t(i)\right)|\mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\log\sum_{i=1}^{n}\frac{p_i(t)-\frac{\beta}{n}}{1-\beta}\exp\left(\frac{\beta}{n}\hat{y}_t(i)\right)|\mathcal{F}_{t-1}\right] \quad \text{definition of } p_i(t) \\
&\leq \mathbb{E}\left[\log\sum_{i=1}^{n}\frac{p_i(t)-\frac{\beta}{n}}{1-\beta}\left(1+\frac{\beta}{n}\hat{y}_t(i)+\frac{(e-2)\beta^2}{n^2}\hat{y}_t(i)^2\right)|\mathcal{F}_{t-1}\right] \\
&\leq \mathbb{E}\left[\log\left(1+\sum_{i=1}^{n}\left[\frac{\beta}{n(1-\beta)}p_i(t)\hat{y}_t(i)+\frac{(e-2)\beta^2}{n^2(1-\beta)}p_i(t)\hat{y}_t(i)^2\right]\right)|\mathcal{F}_{t-1}\right] \\
&\leq \mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{\beta}{n(1-\beta)}p_i(t)\hat{y}_t(i)+\frac{(e-2)\beta^2}{n^2(1-\beta)}p_i(t)\hat{y}_t(i)^2|\mathcal{F}_{t-1}\right)\right] \\
&\leq \frac{\beta}{n(1-\beta)}\mathbb{E}\left[\mu_t(\alpha_{i_t})|\mathcal{F}_{t-1}\right] + \frac{(e-2)\beta^2}{n^2(1-\beta)}\sum_{i=1}^{n}\mathbb{E}\left[\mu_t(\alpha_i)|\mathcal{F}_{t-1}\right].
\end{aligned}$$

The first inequality in the above holds since $e^x \leq 1 + x + (e-2)x^2$ for $x \in [0,1]$. Here, we have $0 \leq \frac{\beta}{n}\hat{y}_t(i) \leq 1$ because $p_i(t) \geq \frac{\beta}{n}$ and $0 \leq y_t(\alpha_i) \leq 1$. The third inequality "$\leq$" in the above holds since $\log(1+x) \leq x$ when $x \geq 0$. The last inequality is from Equation 4, 8. Take another expectation on both sides, we get

$$\mathbb{E}\left[\log\frac{W_{t+1}}{W_t}\right] \leq \frac{\beta}{n(1-\beta)}\mathbb{E}\left[\mu_t(\alpha_{i_t})\right] + \frac{(e-2)\beta^2}{n^2(1-\beta)}\sum_{i=1}^{n}\mathbb{E}[\mu_t(\alpha_i)]$$

By summing the above over t, we have

$$\mathbb{E}\left[\log\frac{W_{T+1}}{W_1}\right] \leq \frac{\beta}{n(1-\beta)}\sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_{i_t})\right] + \frac{(e-2)\beta^2}{n^2(1-\beta)}\sum_{t=1}^{T}\sum_{i=1}^{n}\mathbb{E}[\mu_t(\alpha_i)] \tag{10}$$

Combining the lower bound (Equation 9) and upper bound (Equation 10) of $\mathbb{E}\left[\log\frac{W_{T+1}}{W_1}\right]$, we get for every $i = 1, \ldots, n$,

$$\frac{\beta}{n}\sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_i)\right] - \log n \leq \frac{\beta}{n(1-\beta)}\sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_{i_t})\right] + \frac{(e-2)\beta^2}{n^2(1-\beta)}\sum_{t=1}^{T}\sum_{i=1}^{n}\mathbb{E}[\mu_t(\alpha_i)] \tag{11}$$

Let

$$G_{\max} = \max_{i\in[n]}\sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_i)\right]$$

Since Equation 11 holds for any i, we have

$$\frac{\beta}{n}G_{\max} - \log n \leq \frac{\beta}{n(1-\beta)}\sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_{i_t})\right] + \frac{(e-2)\beta^2}{n(1-\beta)}G_{\max} \tag{12}$$

Equation 12 can be further simplified as

$$G_{\max} - \sum_{t=1}^{T} \mathbb{E}\left[\mu_t(\alpha_{i_t})\right] \leq (e-1)\beta G_{\max} + \frac{(1-\beta)n\log n}{\beta}$$

Since we choose $\beta = \min\left\{1, \sqrt{\frac{n\log n}{(e-1)T}}\right\}$ and note that $G_{\max} \leq T$, we get

$$G_{\max} - \sum_{t=1}^{T} \mathbb{E}\left[\mu_t(\alpha_{i_t})\right] \leq 2\sqrt{(e-1)Tn\log n} = \tilde{O}(\sqrt{nT}).$$

□

A.2 Proof of Theorem 1

To bound the cumulative regret, we only need to bound Quantity (A) and then combine the results in Lemma 1. In the following, we first list some useful lemmas for bounding Quantity (A) for completeness.

A.2.1 Useful Lemmas

Lemma 2 (Proposition 1 in [18]). *Define $V_{n+1} = \sum_{t=1}^{n} X_t X_t^T$, where X_t is drawn IID from some distribution in unit ball $\mathbb{B}^d$. Furthermore, let $\Sigma := E[X_t X_t^T]$ be the second moment matrix, let $B, \delta_2 > 0$ be two positive constants. Then there exists positive, universal constants C_1 and C_2 such that $\lambda_{\min}(V_{n+1}) \geq B$ with probability at least $1-\delta_2$, as long as*

$$n \geq \left(\frac{C_1\sqrt{d} + C_2\sqrt{\log(1/\delta_2)}}{\lambda_{\min}(\Sigma)}\right)^2 + \frac{2B}{\lambda_{\min}(\Sigma)}.$$

Lemma 3 (Theorem 2 in [1]). *For any $\delta < 1$, under our problem setting in Section 3, it holds that for all $t > 0$,*

$$\left\|\hat{\theta}_t - \theta^*\right\|_{V_t} \leq \beta_t(\delta),$$
$$\forall x \in \mathbb{R}^d, |x^\top(\hat{\theta}_t - \theta^*)| \leq \|x\|_{V_t^{-1}} \beta_t(\delta),$$

with probability at least $1-\delta$, where

$$\beta_t(\delta) = \sigma\sqrt{\log\left(\frac{(\lambda+t)^d}{\delta^2\lambda^d}\right)} + \sqrt{\lambda}S.$$

In this subsection we denote $\alpha^*(\delta) \coloneqq \beta_T(\delta)$.

Lemma 4 ([15]). *Let $\lambda > 0$, and $\{x_i\}_{i=1}^t$ be a sequence in $\mathbb{R}^d$ with $\|x_i\| \leq 1$, then we have*

$$\sum_{s=1}^{t} \|x_s\|_{V_s^{-1}}^2 \leq 2\log\left(\frac{\det(V_{t+1})}{\det(\lambda I)}\right) \leq 2d\log\left(1+\frac{t}{\lambda}\right),$$
$$\sum_{s=1}^{t} \|x_s\|_{V_s^{-1}} \leq \sqrt{T\left(\sum_{s=1}^{t}\|x_s\|_{V_s^{-1}}^2\right)} \leq \sqrt{2dt\log\left(1+\frac{t}{\lambda}\right)}.$$

Lemma 5 ([5]). *For a Gaussian random variable Z with mean m and variance σ^2, for any $z \geq 1$,*

$$P(|Z-m| \geq z\sigma) \leq \frac{1}{\sqrt{\pi}z}e^{-z^2/2}.$$

A.2.2 Formal Proof

Proof. (1). Here we would use LinUCB and LinTS for the detailed proof, and note that regret bound of all other UCB and TS algorithms could be similarly deduced. Since α^* in our regret decomposition could be arbitrary element in J, here we simply take $\alpha^* = \min_{\alpha \in J} \alpha$. For LinUCB, since the Lemma 3 holds for any sequence $(x_1, \dots, x_t)$, and hence we have that with probability at least $1 - \delta$,

$$\left\|\hat{\theta} - \theta\right\|_{V_t} \leq \beta_t(\delta) \leq \alpha(t, \delta),$$

where $\alpha(T, \delta)$ is the theoretical optimal exploration rate at round t we denoted in Eqn. (3) with probability parameter δ. And we would omit δ for simplicity. Recall that for $t > T_1$, we denote the feature vector pulled at round t as X_t, i.e.

$$X_t = \underset{x \in \mathcal{A}_t}{\operatorname{argmax}}\, x^\top \hat{\theta}_t + \alpha_{i_t} \|x\|_{V_t^{-1}},\ X_t = X_t(\alpha_{i_t} | \mathcal{F}_{t-1}).$$

And we also define $\tilde{X}_t = X_t(\alpha^* | \mathcal{F}_{t-1})$, i.e.

$$\tilde{X}_t = \underset{x \in \mathcal{A}_t}{\operatorname{argmax}}\, x^\top \hat{\theta}_t + \alpha^* \|x\|_{V_t^{-1}}.$$

And it turns out that the Quantity (A) can be represented by

$$\mathbb{E}\left[\sum_{t=T_1+1}^{T} \left(\mu\left({x_{t,*}}^T\theta\right) - \mu\left(X_t(\alpha^*|\mathcal{F}_{t-1})^T\theta\right)\right)\right] = \mathbb{E}\left[\sum_{t=T_1+1}^{T} \left(\mu\left({x_{t,*}}^T\theta\right) - \mu\left(\tilde{X}_t^T\theta\right)\right)\right].$$

According to the proof of LinUCB we could similarly argue that

$$\begin{aligned} x_{t,*}^\top\theta - \tilde{X}_t^\top\theta &\leq \alpha^*\left(\left\|\tilde{X}_t\right\|_{V_t^{-1}} - \|\tilde{x}_{t,*}\|_{V_t^{-1}}\right) + \left\|x_{t,*} - \tilde{X}_t\right\|_{V_t^{-1}} \left\|\hat{\theta}_t - \theta\right\|_{V_t} \\ &\leq (\alpha^* + \alpha(T))\left\|\tilde{X}_t\right\|_{V_t^{-1}} + \alpha(T)\|x_{t,*}\|_{V_t^{-1}}. \end{aligned}$$

In conclusion, we have that

$$\sum_{t=T_1+1}^{T}\left(\mu\left({x_{t,*}}^T\theta\right) - \mu\left(\tilde{X}_t^T\theta\right)\right) = \widetilde{O}\left(\sum_{t=T_1+1}^{T}\left\|\tilde{X}_t\right\|_{V_t^{-1}} + \sum_{t=T_1+1}^{T}\|x_{t,*}\|_{V_t^{-1}}\right).$$

By Lemma 4 and choosing $T_1 = T^{2/3}$, it holds that,

$$\sum_{t=T_1+1}^{T}\|x_{t,*}\|_{V_t^{-1}},\ \sum_{t=T_1+1}^{T}\left\|\tilde{X}_t\right\|_{V_t^{-1}} = O(T \times T^{-1/3}) = O(T^{2/3}).$$

Secondly, According to [5], we know that for LinTS we have that

$$\mathbb{E}\left[\sum_{t=T_1+1}^{T}\left(\mu\left({x_{t,*}}^T\theta\right) - \mu\left(\tilde{X}_t^T\theta\right)\right)\right] = \widetilde{O}\left(\sum_{t=T_1+1}^{T}\left\|\tilde{X}_t\right\|_{V_t^{-1}} + \|x_{t,*}\|_{V_t^{-1}}\right).$$

But for completeness we still offer an alternative proof for this equality:

$$\begin{aligned} \tilde{X}_t^\top\hat{\theta}_t + \alpha^*\left\|\tilde{X}_t\right\|_{V_t^{-1}} Z_t &\geq x_{t,*}^\top\theta + \alpha^*\|x_{t,*}\|_{V_t^{-1}} Z_{t,*} + x_{t,*}^\top(\hat{\theta}_t - \theta) \\ &\geq x_{t,*}^\top\theta + \alpha^*\|x_{t,*}\|_{V_t^{-1}} Z_{t,*} + \|x_{t,*}\|_{V_t^{-1}}\left\|\hat{\theta}_t - \theta\right\|_{V_t} \\ &\geq x_{t,*}^\top\theta + (\alpha^* Z_{t,*} - \alpha(T))\|x_{t,*}\|_{V_t^{-1}}, \end{aligned}$$

where Z_t and $Z_{t,*}$ are IID normal random variables, $\forall t$. Therefore, it holds that,

$$\begin{aligned} \tilde{X}_t^\top\theta &\geq x_{t,*}^\top\theta + (\alpha^* Z_{t,*} - \alpha(T))\|x_{t,*}\|_{V_t^{-1}} - \alpha^*\|X_t\|_{V_t^{-1}} Z_t + X_t^\top(\theta - \hat{\theta}_t), \\ (x_{t,*} - X_t)^\top\theta &\leq (\alpha(T) + \alpha^* Z_t)\|X_t\|_{V_t^{-1}} + (\alpha(T) - \alpha^* Z_{t,*})\|x_{t,*}\|_{V_t^{-1}} = K_t, \end{aligned}$$

where K_t is normal random variable with

$$\mathbb{E}(K_t) \le 2\alpha(T)T^{-1/3},\ \mathrm{SD}(K_t) \le \sqrt{2}\alpha^* T^{-1/3}.$$

Consequently, we have

$$\sum_{t=T_1+1}^{T} \left({x_{t,*}}^T \theta - \tilde{X}_t^T \theta \right) \le \sum_{t=T_1+1}^{T} K_t := K$$
$$\mathbb{E}(K) = 2\alpha(T)T^{2/3} = \widetilde{O}(T^{4/7}),\ \mathrm{SD}(K) \le \sqrt{2}\alpha^* T^{1/6} = O(T^{1/6}).$$

We have

$$P(K > (2\alpha^* + \sqrt{2})T^{2/3}) \le \frac{1}{c\sqrt{\pi}\sqrt{T}} e^{-c^2 T/2}.$$

This probability upper bound is ultra small and hence negligible. Therefore, we not only prove the expected cumulative regret could be controlled, but also provide a probability bound.

Note we could use this procedure to bound the regret for other UCB and TS bandit algorithms, since most of the proofs for generalized linear bandits are closely related to the rate of $\sum_{t=T_1+1}^{T} \left\| \tilde{X}_t \right\|_{V_t^{-1}}$. Finally, the cost of pure exploration is also of scale $\tilde{O}(T^{2/3})$, which concludes the proof.

(2). Here we simply take $\alpha^* = \min_{\alpha \in J} \alpha$. We also use LinUCB as an example here since other UCB-based algorithms with exploration hyper-parameters could be identically bounded. Based on the definition of X_t and $\tilde{X}_t$, we have that,

$$\begin{aligned} X_t^\top \hat{\theta}_t + \alpha_{i_t} \|X_t\|_{V_t^{-1}} &= X_t^\top \hat{\theta}_t + (\alpha_{i_t} - \alpha^*) \|X_t\|_{V_t^{-1}} + \alpha^* \|X_t\|_{V_t^{-1}} \\ &\ge \tilde{X}_t^\top \hat{\theta}_t + (\alpha_{i_t} - \alpha^*) \left\|\tilde{X}_t\right\|_{V_t^{-1}} + \alpha^* \left\|\tilde{X}_t\right\|_{V_t^{-1}} \\ &\ge X_t^\top \hat{\theta}_t + (\alpha_{i_t} - \alpha^*) \left\|\tilde{X}_t\right\|_{V_t^{-1}} + \alpha^* \|X_t\|_{V_t^{-1}}, \end{aligned}$$

which implies that

$$(\alpha_{i_t} - \alpha^*) \|X_t\|_{V_t^{-1}} \ge (\alpha_{i_t} - \alpha^*) \left\|\tilde{X}_t\right\|_{V_t^{-1}}.$$

Since we have that $\alpha_{i_t} \ge \alpha^*$, and when $\alpha_{i_t} > \alpha^*$ it holds that

$$\|X_t\|_{V_t^{-1}} \ge \left\|\tilde{X}_t\right\|_{V_t^{-1}}, \quad \forall\, t > 0. \tag{13}$$

On the other hand, when $\alpha_{i_t} = \alpha^*$ it holds that $X_t = \tilde{X}_t$, which consequently implies that

$$\|X_t\|_{V_t^{-1}} = \left\|\tilde{X}_t\right\|_{V_t^{-1}}, \quad \forall\, t > 0.$$

According to the proof of LinUCB we could similarly argue that

$$\begin{aligned} x_{t,*}^\top \theta - \tilde{X}_t^\top \theta &\le \alpha^* \left(\left\|\tilde{X}_t\right\|_{V_t^{-1}} - \|\tilde{x}_{t,*}\|_{V_t^{-1}} \right) + \left\| x_{t,*} - \tilde{X}_t \right\|_{V_t^{-1}} \left\| \hat{\theta}_t - \theta \right\|_{V_t} \\ &\le 2\alpha^* \left\|\tilde{X}_t\right\|_{V_t^{-1}}, \end{aligned}$$

since $\alpha(T) \le \alpha^*$. Therefore, we have

$$\sum_{t=T_1+1}^{T} \left(\mu\left({x_{t,*}}^T \theta\right) - \mu\left(\tilde{X}_t^T \theta\right) \right) \le 2\alpha^* \sum_{t=T_1+1}^{T} \left\|\tilde{X}_t\right\|_{V_t^{-1}} \le 2\alpha^* \sum_{t=T_1+1}^{T} \|X_t\|_{V_t^{-1}} = \tilde{O}(\sqrt{T}).$$

Remark 3. *(1) Intuitively, we can deduce Eqn.* (13) *by choosing $\alpha^* = \min_{\alpha \in J} \alpha$, i.e. α^* is no larger than any exploration hyper-parameter candidate since the best feature vector solved in UCB algorithms tends to have larger value of $\|\cdot\|_{V_t^{-1}}$ at time t if we enlarge α. In other words, under larger α we would more likely to choose arm with greater uncertainty quantified by the value of $\|\cdot\|_{V_t^{-1}}$. (2) On the other hand, for TS bandit algorithms we would expect the similar result: the* *feature vectors of superior arms tend to have smaller value of $\|\cdot\|_{V_t^{-1}}$ since the value of $\|\cdot\|_{V_t^{-1}}$ depicts the standard deviation of the feature vector. And the direction of the optimal arm should be frequently explored in the long run and hence its standard deviation is expected to be smaller than other inferior arms. By enlarging α, we would have more chance to choose those sub-optimal arm with larger standard deviation and smaller estimated reward, which means results in Eqn.* (13) *could happen with high probability.*

And this concludes the proof.

(3). Here we would use LinUCB and LinTS for the detailed proof, and note that regret bound of all other UCB and TS algorithms could be similarly deduced. W.l.o.g. we take $\alpha^* = \min_{\alpha \in J} \alpha$

For LinUCB, since the Lemma 3 holds for any sequence $(x_1, \ldots, x_t)$, and hence we have that with probability at least $1 - \delta$,

$$\left\|\hat{\theta} - \theta\right\|_{V_t} \leq \beta_t(\delta) \leq \alpha(t, \delta).$$

And we would omit δ for simplicity. Recall that for $t > T_1$, we denote the feature vector pulled at round t as X_t, i.e.

$$X_t = \operatorname*{argmax}_{x \in \mathcal{A}_t} x^\top \hat{\theta}_t + \alpha_{i_t} \|x\|_{V_t^{-1}}, \; X_t = X_t(\alpha_{i_t} | \mathcal{F}_{t-1}).$$

And we also define $\tilde{X}_t = X_t(\alpha^* | \mathcal{F}_{t-1})$, i.e.

$$\tilde{X}_t = \operatorname*{argmax}_{x \in \mathcal{A}_t} x^\top \hat{\theta}_t + \alpha^* \|x\|_{V_t^{-1}}.$$

And it turns out that the Quantity (A) can be represented by

$$\mathbb{E}\left[\sum_{t=T_1+1}^{T} \left(\mu\left({x_{t,*}}^T\theta\right) - \mu\left(X_t(\alpha^* | \mathcal{F}_{t-1})^T\theta\right)\right)\right] = \mathbb{E}\left[\sum_{t=T_1+1}^{T} \left(\mu\left({x_{t,*}}^T\theta\right) - \mu\left(\tilde{X}_t^T\theta\right)\right)\right].$$

Note the selection of a_t in LinUCB implies that

$$x_{t,*}^\top \hat{\theta}_t + \alpha_{i_t} \|x_{t,*}\|_{V_t^{-1}} \leq X_t^\top \hat{\theta}_t + \alpha_{i_t} \|X_t\|_{V_t^{-1}}.$$

Therefore, we have

$$\begin{aligned} X_t^\top \hat{\theta}_t + \alpha_{i_t} \|X_t\|_{V_t^{-1}} &\geq x_{t,*}^\top \theta + \alpha_{i_t} \|x_{t,*}\|_{V_t^{-1}} + x_{t,*}^\top (\hat{\theta}_t - \theta) \\ &\geq x_{t,*}^\top \theta + \alpha_{i_t} \|x_{t,*}\|_{V_t^{-1}} - \|x_{t,*}\|_{V_t^{-1}} \left\|\hat{\theta}_t - \theta\right\|_{V_t} \\ &\geq x_{t,*}^\top \theta + (\alpha_{i_t} - \alpha(T)) \|x_{t,*}\|_{V_t^{-1}}. \end{aligned} \tag{14}$$

Therefore, it holds that,

$$\begin{aligned} X_t^\top \theta &\geq x_{t,*}^\top \theta + (\alpha_{i_t} - \alpha(T)) \|x_{t,*}\|_{V_t^{-1}} - \alpha_{i_t} \|X_t\|_{V_t^{-1}} + X_t^\top(\theta - \hat{\theta}_t), \\ (x_{t,*} - X_t)^\top \theta &\leq (\alpha(T) + \alpha_{i_t}) \|X_t\|_{V_t^{-1}} + (\alpha(T) - \alpha_{i_t}) \|x_{t,*}\|_{V_t^{-1}}, \end{aligned}$$

By Lemma 2, we have as long as $T_1 = O(T^{4/7})$, it holds that

$$(x_{t,*} - X_t)^\top \theta \leq 2\alpha(T) T^{-2/7}, \; t > T_1.$$

Similarly, we could also deduce that

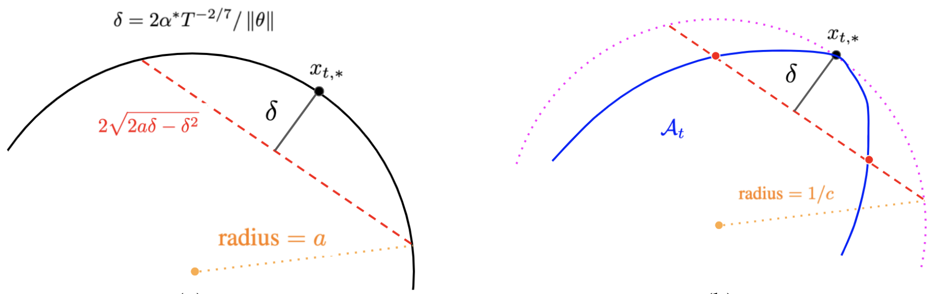
$$(x_{t,*} - \tilde{X}_t)^\top \theta \leq 2\alpha(T) T^{-2/7}, \; t > T_1.$$

Firstly, we take $\mathcal{A}_t = \{x : \|x\| \leq a^2\}, a > 0$ for example, then it holds that $x_{t,*} = \theta / \|\theta\|$, and consequently

$$\|x_{t,*} - X_t\|, \left\|x_{t,*} - \tilde{X}_t\right\| \leq \sqrt{4a\alpha(T) T^{-2/7} / \|\theta\|} = O(\sqrt{\alpha(T)} T^{-1/7}).$$

Please refer to Figure 3 (a) for a 2D visual explanation, and similar argument could be made for higher dimension cases. And this implies that

$$\left\|X_t - \tilde{X}_t\right\| = O(\sqrt{\alpha(T)} T^{-1/7}). \tag{15}$$



(a) (b)

Figure 3: Illustration of our argument in 2D case: (a). explanation of the Eqn. 15, where the red line denotes the maximum distance between X_t and X_t^*; (b). visualization on how to cover the neighborhood of $x_{t,*}$ on $\mathcal{A}_t$, where the blue line denotes the boundary of $\mathcal{A}_t$ and the pink dashed circle is the outer cover with radius $1/c$. In this case, the length of red line gives an upper bound of the maximum distance between X_t and X_t^*.

Generally, if $\mathcal{A}_t$ is some convex set, and we know there exists a small neighborhood of the optimal feature vector $x_{t,*} \in \mathcal{A}_t$ such that the (sectional) principal curvature in this neighborhood can be lowered bounded by some positive constant $c > 0$. Then we can cover this neighborhood by a d-dimensional sphere with radius $1/c$ (Figure 3 (b) for 2D visualization), and hence we could similarly deduce the above result. Note that for the example $\mathcal{A}_t = \{x : \|x\| \leq a\}, a > 0$, all the principal curvatures are equal to $1/a$ anywhere on this sphere, and hence it is a special case. The rest of argument is based on the proof outline of UCB bandits. According to the proof of LinUCB we could similarly argue that

$$\begin{aligned} x_{t,*}^\top \theta - \tilde{X}_t^\top \theta &\leq \alpha^* \left(\left\| \tilde{X}_t \right\|_{V_t^{-1}} - \|\tilde{x}_{t,*}\|_{V_t^{-1}} \right) + \left\| x_{t,*} - \tilde{X}_t \right\|_{V_t^{-1}} \left\| \hat{\theta}_t - \theta \right\|_{V_t} \\ &\leq (\alpha^* + \alpha(T)) \left\| \tilde{X}_t \right\|_{V_t^{-1}} + \alpha(T) \|x_{t,*}\|_{V_t^{-1}} . \end{aligned}$$

In conclusion, we have that

$$\sum_{t=T_1+1}^{T} \left(\mu\left({x_{t,*}}^T \theta \right) - \mu\left(\tilde{X}_t^T \theta \right) \right) = \widetilde{O}\left(\sum_{t=T_1+1}^{T} \left\| \tilde{X}_t \right\|_{V_t^{-1}} + \|x_{t,*}\|_{V_t^{-1}} \right). \tag{16}$$

Note that we have that,

$$\begin{aligned} \sum_{t=T_1+1}^{T} \left\| \tilde{X}_t \right\|_{V_t^{-1}} &\leq \sum_{t=T_1+1}^{T} \|X_t\|_{V_t^{-1}} + \sum_{t=T_1+1}^{T} \left\| X_t - \tilde{X}_t \right\|_{V_t^{-1}} \\ \sum_{t=T_1+1}^{T} \|x_{t,*}\|_{V_t^{-1}} &\leq \sum_{t=T_1+1}^{T} \|X_t\|_{V_t^{-1}} + \sum_{t=T_1+1}^{T} \|X_t - x_{t,*}\|_{V_t^{-1}} , \end{aligned}$$

where the first quantity could be easily bounded by Lemma 4, i.e.

$$\sum_{t=T_1+1}^{T} \|X_t\|_{V_t^{-1}} \leq \sqrt{2dT \log\left(1 + \frac{T}{\lambda}\right)} = \widetilde{O}(\sqrt{T}).$$

And the second quantity could be bounded as

$$\sum_{t=T_1+1}^{T} \left\| X_t - \tilde{X}_t \right\|_{V_t^{-1}} \leq \sum_{t=T_1+1}^{T} \left\| X_t - \tilde{X}_t \right\| \sqrt{\lambda_{\min}(V_t^{-1})} \lesssim \sum_{t=T_1+1}^{T} \sqrt{\alpha^*} T^{-3/7} = \widetilde{O}(T^{4/7}),$$

with high probability. Then by taking $\delta = \delta / T^{3/7}$ we can easily prove that

$$\mathbb{E}\left[\sum_{t=T_1+1}^{T} \left\| X_t - \tilde{X}_t \right\|_{V_t^{-1}} \right] = \widetilde{O}(T^{4/7}).$$

Similarly, it holds that

$$\mathbb{E}\left[\sum_{t=T_1+1}^{T}\|X_t - x_{t,*}\|_{V_t^{-1}}\right] = \widetilde{O}(T^{4/7}).$$

Therefore, we have that

$$\sum_{t=T_1+1}^{T}\left(\mu\left({x_{t,*}}^T\theta\right) - \mu\left(\tilde{X}_t^T\theta\right)\right) = \widetilde{O}(T^{4/7}).$$

For LinTS, the proof could also be similarly deduced. And we modify the definition of $\tilde{X}_t$ as

$$\tilde{X}_t = \operatorname*{argmax}_{x\in\mathcal{A}_t} x^\top\hat{\theta}_t + \alpha^*\|x\|_{V_t^{-1}}\tilde{Z}_t,$$

where Z_t is a standard normal random variable. And we could similarly show that:

$$\begin{aligned}
X_t^\top\hat{\theta}_t + \alpha_{i_t}\|X_t\|_{V_t^{-1}} Z_t &\geq x_{t,*}^\top\theta + \alpha_{i_t}\|x_{t,*}\|_{V_t^{-1}} Z_{t,*} + x_{t,*}^\top(\hat{\theta}_t - \theta) \\
&\geq x_{t,*}^\top\theta + \alpha_{i_t}\|x_{t,*}\|_{V_t^{-1}} Z_{t,*} + \|x_{t,*}\|_{V_t^{-1}}\left\|\hat{\theta}_t - \theta\right\|_{V_t} \\
&\geq x_{t,*}^\top\theta + (\alpha_{i_t}Z_{t,*} - \alpha(T))\|x_{t,*}\|_{V_t^{-1}}.
\end{aligned}$$

Therefore, it holds that,

$$\begin{aligned}
X_t^\top\theta &\geq x_{t,*}^\top\theta + (\alpha_{i_t}Z_{t,*} - \alpha(T))\|x_{t,*}\|_{V_t^{-1}} - \alpha_{i_t}\|X_t\|_{V_t^{-1}} Z_t + X_t^\top(\theta - \hat{\theta}_t), \\
(x_{t,*} - X_t)^\top\theta &\leq (\alpha(T) + \alpha_{i_t}Z_t)\|X_t\|_{V_t^{-1}} + (\alpha(T) - \alpha_{i_t}Z_{t,*})\|x_{t,*}\|_{V_t^{-1}} = K_t,
\end{aligned} \tag{17}$$

where K_t is normal random variable with

$$\mathbb{E}(K_t) \leq 2\alpha(T)T^{-2/7},\ \mathrm{SD}(K_t) \leq \sqrt{2}\alpha_{i_t}T^{-2/7} \leq \sqrt{2}\alpha^*T^{-2/7}.$$

According to Lemma 5, we have that for arbitrary $\xi > 0$

$$\begin{aligned}
&P\left(\max_{t\in T} K_t \geq 2\alpha(T)T^{-2/7} + \left(\sqrt{2\log(T)} + \xi\right)\sqrt{2}\alpha^*T^{-2/7}\right) \\
&\leq P\left(\max_{t\in T}\frac{K_t - \mathbb{E}[K_t]}{SD(K_t)} \geq \sqrt{2\log(T)} + \xi\right) \\
&= T \times P\left(Z \geq \sqrt{2\log(T)} + \xi\right) \quad Z \sim N(0,1) \\
&\leq T\frac{1}{\sqrt{\pi}(\sqrt{2\log(T)} + \xi)}\exp\left(-(\sqrt{2\log(T)} + \xi)^2/2\right) \\
&\leq \frac{1}{\sqrt{2\log(T)} + \xi}\exp\left(-\frac{\xi^2}{2}\right).
\end{aligned}$$

By taking $\xi = 2\sqrt{\log(T)}$, it holds that

$$P\left(\max_{t\in T} K_t \geq 2\alpha(T)T^{-2/7} + \left(\sqrt{2\log(T)} + \xi\right)\sqrt{2}\alpha^*T^{-2/7}\right) \leq \frac{1}{2T^2\sqrt{\log(T)}}.$$

Since this probability upper bound is ultra small and hence negligible, we have

$$(x_{t,*} - X_t)^\top\theta \leq 2\alpha(T)T^{-2/7},\ t > T_1.$$

Similarly, we could also deduce that

$$(x_{t,*} - \tilde{X}_t)^\top\theta \leq 2\alpha(T)T^{-2/7},\ t > T_1.$$

This result similarly implies that

$$\mathbb{E}\left[\sum_{t=T_1+1}^{T}\|X_t - x_{t,*}\|_{V_t^{-1}}\right] = \tilde{O}(T^{4/7}),\ \mathbb{E}\left[\sum_{t=T_1+1}^{T}\left\|X_t - \tilde{X}_t\right\|_{V_t^{-1}}\right] = \tilde{O}(T^{4/7}).$$

According to [5] (or Eqn. (17)), we know that for LinTS we have the similar result as in Eqn. (16):

$$\sum_{t=T_1+1}^{T}\left(\mu\left({x_{t,*}}^T\theta\right)-\mu\left(\tilde{X}_t^T\theta\right)\right)=\tilde{O}\left(\sum_{t=T_1+1}^{T}\left\|\tilde{X}_t\right\|_{V_t^{-1}}\|x_{t,*}\|_{V_t^{-1}}\right).$$

And this directly implies that

$$\sum_{t=T_1+1}^{T}\left(\mu\left({x_{t,*}}^T\theta\right)-\mu\left(\tilde{X}_t^T\theta\right)\right)=\tilde{O}(T^{4/7}).$$

Note we could use this procedure to bound the regret for UCB and TS bandit algorithms under condition in (3), since most of the proofs for generalized linear bandits are closely related to the rate of $\sum_{t=T_1+1}^{T}\left\|\tilde{X}_t\right\|_{V_t^{-1}}$. Finally, the cost of pure exploration is also of scale $\tilde{O}(T^{4/7})$, which concludes the proof.

$\square$

A.3 Analysis of Theorem 2

A.3.1 Useful Conclusions

Proposition 1. *Assume given the past information $\mathcal{F}_{t-1}$ and the hyper-parameters to be used by the contextual bandit algorithm at round t, the arm to be pulled by the contextual bandit algorithm follows a fixed distribution. Denote $R(\alpha^{(1)},\dots,\alpha^{(L)},T,\{\mathcal{F}_{t-1}\})$ as the cumulative regret of the contextual bandit algorithm if it is run with parameters $(\alpha^{(1)},\dots,\alpha^{(L)})$ given the past information $\mathcal{F}_{t-1}$ at round t. Then the auto tuning method in Algorithm 2 has regret that satisfies the following:*

$$\begin{aligned}\mathbb{E}[R(T)]\le&\min_{(\alpha^{(1)},\dots,\alpha^{(L)})\in J_1\times\cdots\times J_L}\mathbb{E}[R(\alpha^{(1)},\dots,\alpha^{(L)},T,\{\mathcal{F}_{t-1}\})]\\&+2\sum_{l=1}^{L}\sqrt{(e-1)n_l(T-T_1)\log n_l}.\end{aligned}$$

Proof. We also reload some notations here for simplicity in the same way as proof of Lemma 1 in Appendix A.1. More specifically, since at iteration t we are given the past information $\mathcal{F}_{t-1}$ to make decision according to different choices of hyper-parameter values, and hence we would omit this notation $\mathcal{F}_{t-1}$ when we refer to the arm or feature vector we pull under different hyper-parameter values: Denote $a_t\left(\alpha_{i_1}^{(1)},\dots,\alpha_{i_L}^{(L)}\right)$ as the pulled arm at round t if the hyper-parameters selected at round t is $\left(\alpha_{i_1}^{(1)},\dots,\alpha_{i_L}^{(L)}\right)$. Denote $X_t\left(\alpha_{i_1}^{(1)},\dots,\alpha_{i_L}^{(L)}\right)$ as the corresponding feature vector and $\mu_t\left(\alpha_{i_1}^{(1)},\dots,\alpha_{i_L}^{(L)}\right)$ as the corresponding expected reward. It suffices to show that for any $l=1,\dots,L$, the following holds.

$$\begin{aligned}&\sum_{t=1}^{T}\mathbb{E}\left[\mu_t\left(\alpha_{i_t(1)}^{(1)},\dots,\alpha_{i_t(l-1)}^{(l-1)},\alpha_*^{(l)},\dots,\alpha_*^{(L)}\right)\right]\\&-\sum_{t=1}^{T}\mathbb{E}\left[\mu_t\left(\alpha_{i_t(1)}^{(1)},\dots,\alpha_{i_t(l-1)}^{(l-1)},\alpha_{i_t(l)}^{(l)},\alpha_*^{(l+1)},\dots,\alpha_*^{(L)}\right)\right]\\&\le 2\sqrt{(e-1)n_lT\log n_l}.\end{aligned}\tag{18}$$

For convenience, we will denote $\left(\alpha_{i_t(1)}^{(1)},\dots,\alpha_{i_t(l-1)}^{(l-1)},\alpha_j^{(l)},\alpha_*^{(l+1)},\dots\alpha_*^{(L)}\right)$ as (α_j) when there is no ambiguity, which means that the first $l-1$ hyper-parameters are chosen as $\alpha_{i_t(s)}^{(s)}$ for $s=1,\dots,l-1$, the l-th hyper-parameter is chosen with index j and the rest of the hyper-parameters are chosen as $\alpha_*^{(s)}$ for $s=l+1,\dots,L$. Then the result we want to show in Equation 18 can be written as

$$\sum_{t=1}^{T}\mathbb{E}\left[\mu_t\left(\alpha^{(1)}_{i_t(1)},\ldots,\alpha^{(l-1)}_{i_t(l-1)},\alpha^{(l)}_{*},\ldots,\alpha^{(L)}_{*}\right)\right]-\sum_{t=1}^{T}\mathbb{E}\left[\mu_t\left(\alpha_{i_t(l)}\right)\right]\leq 2\sqrt{(e-1)n_lT\log n_l}. \tag{19}$$

We will also omit the superscript / subscript (l) for convenience when there is no ambiguity, so $p_j^{(l)}(t), w_j^{(l)}(t), \hat{y}_t^{(l)}(j)$ are abbreviated as $p_j(t), w_j(t), \hat{y}_t(j)$ respectively. Denote $\mathcal{H}_t = \sigma\left(\alpha^{(1)}_{i_t(1)},\ldots,\alpha^{(l-1)}_{i_t(l-1)},\alpha^{(l+1)}_{*},\ldots,\alpha^{(L)}_{*}\right)$ as the σ-algebra induced by the event that at round t, the first $l-1$ hyper-parameters are chosen as $\alpha^{(s)}_{i_t(s)}$ and for $s = l+1,\ldots,L$, the hyper-parameters are chosen as $\alpha^{(s)}_{*}$. Given $\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)$, denote $y_t(\alpha_j)=\mu_t(\alpha_j)+\epsilon'$ as the observed reward at round t if $\alpha^{(l)}$ is chosen as $\alpha^{(l)}_j$ and the rest hyper-parameters given by $\mathcal{H}_t$. Here, ϵ' is a hypothetical random noise if arm $a_t(\alpha_j)$ is pulled at round t.

Given $\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)$, by the above definitions and Algorithm 2, $\hat{y}_t(j) = y_t(\alpha_j)/p_j(t)$ if $j = i_t(l)$. Otherwise, $\hat{y}_t(j) = 0$. Since $p_j(t)\geq \frac{\beta_l}{n_l}$, we have $\hat{y}_t(j)\leq \frac{n_l}{\beta_l}$ for all $j\in[n_l]$ and t. We also have the following two inequalities.

$$\begin{aligned}&\mathbb{E}\left(\sum_{i=1}^{n_l}p_i(t)\hat{y}_t(i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)=\mathbb{E}\left(p_{i_t(l)}(t)\hat{y}_t(i_t(l))|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)\\&=\mathbb{E}\left(y_t(\alpha_{i_t(l)})|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)=\mathbb{E}\left[\mu_t\left(\alpha_{i_t(l)}\right)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]\end{aligned} \tag{20}$$

$$\begin{aligned}&\mathbb{E}\left(\sum_{i=1}^{n_l}p_i(t)\hat{y}_t(i)^2|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)=\mathbb{E}\left(p_{i_t(l)}(t)\hat{y}_t(i_t(l))^2|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)\\&=\mathbb{E}\left(y_t(i_t(l))\hat{y}_t(i_t(l))|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)\leq\mathbb{E}\left(\hat{y}_t(i_t(l))|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)\\&=\mathbb{E}\left(\sum_{i=1}^{n_l}\hat{y}_t(i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)\end{aligned} \tag{21}$$

For a single $i\in[n_l]$, since given $\mathcal{F}_{t-1}$, $p_i^{(l)}(t)$ is already fixed, which means that the choices of other hyper-parameters do not affect the distribution of $i_t(l)$. Moreover, $a_t(\alpha_i)$ follows a fixed distribution due to the conditions in Theorem 2, i.e., the arm to be pulled follows a fixed distribution given the past information and the hyper-parameters to be used at round t. Therefore, given $\sigma(\mathcal{F}_{t-1},\mathcal{H}_t,a_t(\alpha_i),\epsilon')$, $i=i_t(l)$ is still with probability $p_i^{(l)}(t)$ for all $i\in[n_l]$. So

$$\begin{aligned}&\mathbb{E}\left(\hat{y}_t(i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)=\mathbb{E}\left[\mathbb{E}\left(\hat{y}_t(i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t,a_t(\alpha_i),\epsilon')\right)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]\\&=\mathbb{E}\left[y_t(\alpha_i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]\\&=\mathbb{E}\left[\mu_t(\alpha_i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right].\end{aligned} \tag{22}$$

From Equation 21 and 22, we have

$$\mathbb{E}\left(\sum_{i=1}^{n_l}p_i(t)\hat{y}_t(i)^2|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)\leq\mathbb{E}\left(\sum_{i=1}^{n_l}\mu_t(\alpha_i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right) \tag{23}$$

We still look at the lower bound and upper bound of $\mathbb{E}[\log\frac{W_{T+1}}{W_1}]$, but now $W_t=\sum_{i=1}^{n_l}w_i^{(l)}(t)$, and we will use the abbreviation $w_i(t)=w_i^{(l)}(t)$ below for ease of notation.

Lower bound:

$$\begin{aligned}&\mathbb{E}\left[\log\frac{w_i(t+1)}{w_i(t)}|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]=\mathbb{E}\left[\frac{\beta_l}{n_l}\hat{y}_t(i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]\\&=\mathbb{E}\left[\frac{\beta_l}{n_l}\mu_t(\alpha_i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]\quad\text{from Equation 22}\end{aligned}$$

Take an expectation on both sides and sum over t, we have

$$\mathbb{E}\left[\log w_i(T+1)\right] = \frac{\beta_l}{n_l}\sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_i)\right]$$

Therefore, for all $i \in [n_l]$,

$$\mathbb{E}[\log \frac{W_{T+1}}{W_1}] \geq \mathbb{E}[\log w_i(T+1)] - \log n_l = \frac{\beta_l}{n_l}\sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_i)\right] - \log n_l. \tag{24}$$

Upper bound: This part is almost the same as the arguments in Lemma 1, except now that the conditional expectation is taken over $\sigma(\mathcal{F}_{t-1}, \mathcal{H}_t)$. For completeness, we write out the proof of this part below. Again, we will use $p_i(t) = p_i^{(l)}(t)$ and $w_i(t) = w_i^{(l)}(t)$ for convenience.

$$\begin{aligned}
&\mathbb{E}\left[\log\frac{W_{t+1}}{W_t}|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right] = \mathbb{E}\left[\log\sum_{i=1}^{n_l}\frac{w_i(t+1)}{W_t}|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]\\
&= \mathbb{E}\left[\log\sum_{i=1}^{n_l}\frac{w_i(t)}{W_t}\exp\left(\frac{\beta_l}{n_l}\hat{y}_t(i)\right)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]\\
&= \mathbb{E}\left[\log\sum_{i=1}^{n_l}\frac{p_i(t)-\frac{\beta_l}{n_l}}{1-\beta_l}\exp\left(\frac{\beta_l}{n_l}\hat{y}_t(i)\right)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right] \quad \text{definition of } p_i(t)\\
&\leq \mathbb{E}\left[\log\sum_{i=1}^{n_l}\frac{p_i(t)-\frac{\beta_l}{n_l}}{1-\beta_l}\left(1+\frac{\beta_l}{n_l}\hat{y}_t(i)+\frac{(e-2)\beta_l^2}{n_l^2}\hat{y}_t(i)^2\right)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]\\
&\leq \mathbb{E}\left[\log\left(1+\sum_{i=1}^{n_l}\left[\frac{\beta_l}{n_l(1-\beta_l)}p_i(t)\hat{y}_t(i)+\frac{(e-2)\beta_l^2}{n_l^2(1-\beta_l)}p_i(t)\hat{y}_t(i)^2\right]\right)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]\\
&\leq \mathbb{E}\left[\sum_{i=1}^{n_l}\left(\frac{\beta_l}{n_l(1-\beta_l)}p_i(t)\hat{y}_t(i)+\frac{(e-2)\beta_l^2}{n_l^2(1-\beta_l)}p_i(t)\hat{y}_t(i)^2|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right)\right]\\
&\leq \frac{\beta_l}{n_l(1-\beta_l)}\mathbb{E}\left[\mu_t(\alpha_{i_t})|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right]+\frac{(e-2)\beta_l^2}{n_l^2(1-\beta_l)}\sum_{i=1}^{n_l}\mathbb{E}\left[\mu_t(\alpha_i)|\sigma(\mathcal{F}_{t-1},\mathcal{H}_t)\right].
\end{aligned}$$

The first inequality "$\leq$" in the above holds since $e^x \leq 1 + x + (e-2)x^2$ for $x \in [0,1]$. Here, we have $0 \leq \frac{\beta_l}{n_l}\hat{y}_t(i) \leq 1$ because $p_i(t) \geq \frac{\beta_l}{n_l}$, $0 \leq y_t(\alpha_i) \leq 1$ and $\hat{y}_t(i) \leq \frac{y_t(\alpha_i)}{p_i(t)}$. The last inequality is from Equation 20, 23. Take another expectation on both sides, we get

$$\mathbb{E}\left[\log\frac{W_{t+1}}{W_t}\right] \leq \frac{\beta_l}{n_l(1-\beta_l)}\mathbb{E}\left[\mu_t(\alpha_{i_t})\right]+\frac{(e-2)\beta_l^2}{n_l^2(1-\beta_l)}\sum_{i=1}^{n_l}\mathbb{E}[\mu_t(\alpha_i)]$$

By summing the above over t, we have

$$\mathbb{E}[\log\frac{W_{T+1}}{W_1}] \leq \frac{\beta_l}{n_l(1-\beta_l)}\sum_{t=1}^{T}\mathbb{E}[\mu_t(\alpha_{i_t(l)})]+\frac{(e-2)\beta_l^2}{n_l^2(1-\beta_l)}\sum_{i=1}^{n_l}\sum_{t=1}^{T}\mathbb{E}[\mu_t(\alpha_i)] \tag{25}$$

Note that the lower bound in Equation 24 holds for any i, so it also holds for $\alpha_*^{(l)}$. Denote

$$G_{\max} = \sum_{t=1}^{T}\mathbb{E}\left[\mu_t(\alpha_{i_t(1)}^{(1)},\ldots,\alpha_{i_t(l-1)}^{(l-1)},\alpha_*^{(l)},\ldots,\alpha_*^{(L)})\right].$$

Then

$$\frac{\beta_l}{n_l}G_{\max} - \log n_l \leq \frac{\beta_l}{n_l(1-\beta_l)}\sum_{t=1}^{T}\mathbb{E}[\mu_t(\alpha_{i_t(l)})]+\frac{(e-2)\beta_l^2}{n_l^2(1-\beta_l)}\sum_{i=1}^{n_l}\sum_{t=1}^{T}\mathbb{E}[\mu_t(\alpha_i)]$$

We note that $\sum_{t=1}^{T}\mathbb{E}[\mu_t(\alpha_i)] \leq T$ for all i, so

$$\frac{\beta_l}{n_l}G_{\max} - \log n_l \leq \frac{\beta_l}{n_l(1-\beta_l)}\sum_{t=1}^{T}\mathbb{E}[\mu_t(\alpha_{i_t(l)})] + \frac{(e-2)\beta_l^2}{n_l(1-\beta_l)}T$$

Simplify the above inequality and due to the choice of β_l, we have

$$\begin{aligned}
&G_{\max} - \sum_{t=1}^{T}\mathbb{E}[\mu_t(\alpha_{i_t(l)})] \leq \beta_l G_{\max} + (e-2)\beta_l T + \frac{(1-\beta_l)n_l}{\beta_l}\log n_l \\
&\leq 2\sqrt{(e-1)n_l T \log n_l}.
\end{aligned}$$

This concludes the proof of Proposition 1. □

Lemma 6 (Adapted from Lemma 3)**.** *For any $\delta < 1$, under our problem setting in Section 3 with the regularization hyper-parameter $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ ($\lambda_{\min} > 0$), it holds that for all $t > 0$,*

$$\begin{aligned}
\left\|\hat{\theta}_t - \theta^*\right\|_{V_t} &\leq \beta_t(\delta), \\
\forall x \in \mathbb{R}^d, |x^\top(\hat{\theta}_t - \theta^*)| &\leq \|x\|_{V_t^{-1}}\,\beta_t(\delta),
\end{aligned}$$

with probability at least $1-\delta$, where

$$\beta_t(\delta) = \sigma\sqrt{\log\left(\frac{(\lambda_{\min}+t)^d}{\delta^2\lambda_{\min}^d}\right)} + \sqrt{\lambda_{\max}}S.$$

Proof. The proof of this Lemma is trivial given Lemma 3. For any $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, according to Lemma 3 it holds that, for all $t > 0$,

$$\begin{aligned}
\left\|\hat{\theta}_t - \theta^*\right\|_{V_t} &\leq \beta_t(\delta), \\
\forall x \in \mathbb{R}^d, |x^\top(\hat{\theta}_t - \theta^*)| &\leq \|x\|_{V_t^{-1}}\,\beta_t(\delta),
\end{aligned}$$

with probability at least $1-\delta$, where

$$\beta_t(\delta) = \sigma\sqrt{\log\left(\frac{(\lambda+t)^d}{\delta^2\lambda^d}\right)} + \sqrt{\lambda}S \leq \sigma\sqrt{\log\left(\frac{(\lambda_{\min}+t)^d}{\delta^2\lambda_{\min}^d}\right)} + \sqrt{\lambda_{\max}}S.$$

□

A.3.2 Proof of Theorem 2

Proof. We could validate Theorem 2 by extending the proof of Theorem 1 with Proposition 1. Note that most contextual bandit algorithms contain three types of hyper-parameters: one is the exploration rate, which we have throughout discussed in the proof of Theorem 1. The second class is the stepsize of some gradient-based optimization loop (e.g. Laplace-TS [4]), but the output from the loop when the convergent criteria is met is similar. In other words, this kind of hyper-parameter is not critical in the theoretical proof. The last one is the regularization parameter λ, but it can be easily handled by using Lemma 6. Therefore, we only need to consider the case when we tune the exploration rate and the regularization parameter simultaneously. We will take LinUCB with two hyperparameters (i.e. exploration rate and regularization parameter) as an example:

The proof is similar to the one in Appendix A.2. Denote the candidate sets for hyper-parameter α and λ as J_1 and J_2 ($0 < \lambda_{\min} \leq J_2 \leq \lambda_{\max}$). And denote $V_t(\lambda) = \lambda I + \sum_{i=1}^{t-1} X_t X_t^\top$, α_{i_t} and λ_{i_t} as the exploration and regularization rate we tune in our Syndicated framework at round t. Moreover, we define $\alpha^* = \min_{\alpha\in J_1}\alpha, \lambda^* = \min_{\lambda\in J_2}\lambda$. With probability at least $1-\delta$,

$$\left\|\hat{\theta} - \theta\right\|_{V_t(\lambda)} \leq \beta_t(\delta) \coloneqq \alpha(T,\delta), \qquad \forall\lambda \in J_2$$

where the definition of $\beta_t(\delta)$ is reloaded in Lemma 6. And we would omit δ for simplicity. For $t > T_1$, we denote the feature vector pulled at round t as X_t, i.e.

$$X_t = \operatorname*{argmax}_{x\in\mathcal{A}_t} x^\top\hat{\theta}_t + \alpha_{i_t}\|x\|_{V_t^{-1}(\lambda_{i_t})},\ X_t = X_t(\alpha_{i_t}, \lambda_{i_t}|\mathcal{F}_{t-1}).$$

And we also define $\tilde{X}_t = X_t(\alpha^*, \lambda^*|\mathcal{F}_{t-1})$, i.e.

$$\tilde{X}_t = \operatorname*{argmax}_{x\in\mathcal{A}_t} x^\top\hat{\theta}_t + \alpha^*\|x\|_{V_t^{-1}(\lambda^*)}.$$

According to Proposition 1, it holds that

$$\begin{aligned}&\mathbb{E}[R(T)] \le \mathbb{E}[R(\alpha^*, \lambda^*, T, \{\mathcal{F}_{t-1}\})] + O(\sqrt{T-T_1})\\ &\le \mathbb{E}\left[\sum_{t=T_1+1}^{T}\left(\mu\left({x_{t,*}}^T\theta\right) - \mu\left(\tilde{X}_t^T\theta\right)\right)\right] + O(\sqrt{T-T_1})\end{aligned}$$

According to the proof of LinUCB we could similarly argue that

$$\begin{aligned}x_{t,*}^\top\theta - \tilde{X}_t^\top\theta &\le \alpha^*\left(\left\|\tilde{X}_t\right\|_{V_t^{-1}(\lambda^*)} - \|\tilde{x}_{t,*}\|_{V_t^{-1}}(\lambda^*)\right) + \left\|x_{t,*} - \tilde{X}_t\right\|_{V_t^{-1}(\lambda^*)}\left\|\hat{\theta}_t - \theta\right\|_{V_t(\lambda^*)}\\ &\le (\alpha^* + \alpha(T))\left\|\tilde{X}_t\right\|_{V_t^{-1}(\lambda^*)} + \alpha(T)\|x_{t,*}\|_{V_t^{-1}(\lambda^*)}.\end{aligned}$$

In conclusion, we have that

$$\sum_{t=T_1+1}^{T}\left(\mu\left({x_{t,*}}^T\theta\right) - \mu\left(\tilde{X}_t^T\theta\right)\right) = \tilde{O}\left(\sum_{t=T_1+1}^{T}\left\|\tilde{X}_t\right\|_{V_t^{-1}(\lambda^*)} + \sum_{t=T_1+1}^{T}\|x_{t,*}\|_{V_t^{-1}(\lambda^*)}\right).$$

By Lemma 4 and choosing $T_1 = T^{2/3}$, it holds that,

$$\sum_{t=T_1+1}^{T}\|x_{t,*}\|_{V_t^{-1}},\ \sum_{t=T_1+1}^{T}\left\|\tilde{X}_t\right\|_{V_t^{-1}} = O(T\times T^{-1/3}) = O(T^{2/3}).$$

Note we could literally use the identical argument for all UCB and TS bandit algorithms as in the proof of Theorem 1, and the only modification is the value of $\alpha(T)$ we newly defined in Lemma 6.

To prove the Theorem 1 (3) holds, we can also use an exactly identical argument as in the proof of Theorem 1 (3) in Appendix A.2, and the only difference is we replace the value of $\alpha(T)$ in our main paper by the newly defined one in Lemma 6, and hence we would not copy it here again. And this fact concludes our proof. □

A.4 Experimental Settings

Simulations. We use $d = 10$, $K = 100$ and draw $\theta^* \sim$ Uniform$(-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}})$. For linear bandits, we draw the feature vectors $x_{t,a} \sim$ Uniform$(-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}})$ and transform the mean reward of arm a at round t by $\mu_{t,a} \leftarrow \frac{\mu_{t,a}+1}{2}$ to make sure the mean rewards are within $[0,1]$. Each round an arm is pulled, a sample reward $Y_t \sim N(\mu_{t,a_t}, 0.1)$ is revealed to the player. For logistic models, the feature vectors $x_{t,a} \sim$ Uniform$(-1,1)$ and the corresponding mean reward is $\mu_{t,a} = 1/(1+\exp(-x_{t,a}^T\theta^*))$. A sample Bernoulli reward is drawn when an arm is pulled.

Real datasets. We use the benchmark Movielens 100K dataset similarly as in [8]. The Movielens dataset contains 100K ratings on 1,682 movies contributed by 943 users. For data preprocessing, we apply LIBPMF [26, 27] to factorize the ratings matrix to get feature matrices for both users and movies with $d = 20$. We randomly select $K = 1000$ movies (arms) in each round, and the model parameter θ^* is defined as the averaged feature vectors of 100 randomly selected users. For linear models, the mean reward is defined as $\mu_{t,a} = x_{t,a}^T\theta$ and transformed into $[0,1]$. The sample reward is drawn from $N(\mu_{t,a}, 1)$. For logistic models, the mean reward is defined as $\mu_{t,a} = 1/(1+\exp(-x_{t,a}^T\theta^*))$, and the sample reward is drawn from a Bernoulli distribution.

A.5 Additional experiments on tuning SGD-TS

In this section, we show the comparison of different tuning methods in SGD-TS [14], a recently proposed efficient algorithm for generalized linear bandit. We apply SGD-TS with a logistic model to the datasets considered in Section 6. SGD-TS has four tuning parameters, the length of epoch τ, two exploration parameters $\alpha^{(1)}$ and $\alpha^{(2)}$, step size for stochastic gradient descent η_0. In [14], the experiments are conducted by using a grid search of all four parameters, which is not feasible in practice. Since the epoch length has to be pre-determined, it is not applicable to tune it online. We set $\tau = 10 \times \lfloor \max(\log T, d) \rfloor$ as suggested by the grid search set in [14] and fix it for all tuning methods. The tuning set for $\alpha^{(1)}$ and $\alpha^{(2)}$ are the same $\{0, 0.01, 0.1, 1, 10\}$. The tuning set for step size η_0 is set as $\{0.01, 0.1, 1, 10\}$. The theoretical choices of step size η_0 in SGD-TS are intractable, so for the tuning methods in Section 6, we make the following modifications:

1. **OP [9]**: We modify OPLINUCB to tune step size η_0 only.
2. **Corral [3]**: We modify the CORRAL model selection framework to tune step size η_0 only.
3. **Corral-Combined [3]**: We modify the CORRAL model selection framework to tune all three hyper-parameters $\alpha^{(1)}$, $\alpha^{(2)}$ and η_0. And the tuning set contain all possible combinations of these three hyper-parameters.
4. **TL (Our work, Algorithm 1)**: This is our proposed Algorithm 1, where we use the two-layer bandit structure to tune the step size η_0 only.
5. **TL-Combined (Our work, Algorithm 1)**: This method tunes all three hyper-parameters $\alpha^{(1)}$, $\alpha^{(2)}$ and η_0 using Algorithm 1, but with the tuning set containing all the possible combinations of the three hyper-parameters.
6. **Syndicated (Our work, Algorithm 2)**: This method keeps three separate tuning sets for $\alpha^{(1)}$, $\alpha^{(2)}$ and η_0 respectively. It uses the Syndicated Bandits framework in Algorithm 2.

For OP Corral and TL, since they do not tune the two exploration parameters, $\alpha^{(1)}$ and $\alpha^{(2)}$ are set as the theoretical values as in [14]. Results reported in Figure 4 are averaged over 10 repeated experiments. From the plots, we can see that 1) our proposed Syndicated Bandits framework outperforms TL-combined method since now there are in total three hyper-parameters and the regret of TL-combined depends on the number of hyper-parameters exponentially. 2) Tuning all 3 hyper-parameters significantly outperforms tuning only the step size as in OP, Corral and TL. This further indicates that tuning multiple hyper-parameters is better than tuning fewer. On the other hand, it suggests that the theoretical choices of the exploration parameters do not always perform better than the fine-tuned results. 3) Our proposed TL algorithm outperforms OP and Corral when tuning only the step size.

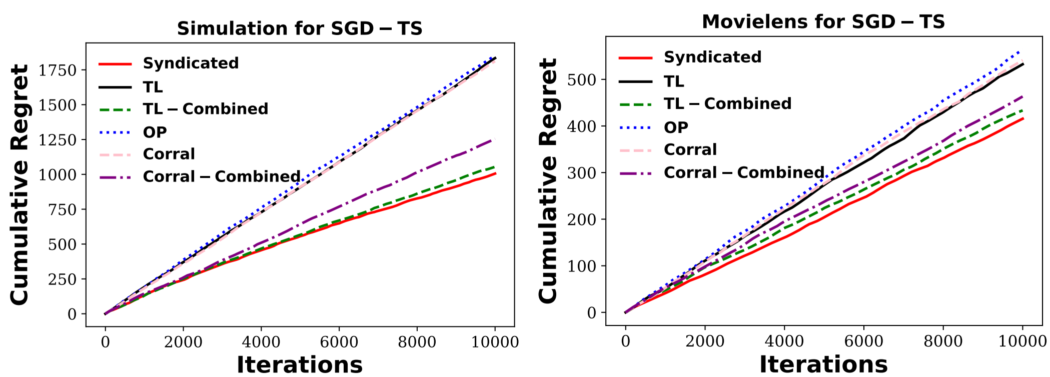


Figure 4: Comparison of hyper-parameters selection methods in SGD-TS.