A Simple Approach for Non-stationary Linear Bandits

A Proof of Lemma 3

Proof of Lemma 3. We first prove the upper bound of $A_t$. The essential proof is actually due to Cheung et al. [2019a] in analyzing sliding window based approach. For self-containedness, we restate here in the notations of our proposed restart strategy.

$$
\left\| V_{t-1}^{-1} \sum_{s=t_0}^{t-1} X_s X_s^T (\theta_s - \theta_t) \right\|_2

= \left\| V_{t-1}^{-1} \left( \sum_{s=t_0}^{t-1} X_s X_s^T \left( \sum_{p=s}^{t-1} (\theta_p - \theta_{p+1}) \right) \right) \right\|_2

= \left\| V_{t-1}^{-1} \sum_{p=t_0}^{t-1} \left( \sum_{s=t_0}^{p} X_s X_s^T (\theta_p - \theta_{p+1}) \right) \right\|_2

\leq \sum_{p=t_0}^{t-1} \left\| V_{t-1}^{-1} \left( \sum_{s=t_0}^{p} X_s X_s^T \right) (\theta_p - \theta_{p+1}) \right\|_2

\leq \sum_{p=t_0}^{t-1} \lambda_{\text{max}} \left( \sum_{s=t_0}^{p} X_s X_s^T \right) \left\| \theta_p - \theta_{p+1} \right\|_2

\leq \left\| \theta_p - \theta_{p+1} \right\|_2,

\tag{21}

\tag{22}

\tag{23}

\tag{24}

where (21) holds by rearranging over the index pair of $(s, p)$, (22) holds due to the triangle inequality, (23) and (24) can be obtained by the same argument in Appendix B of Cheung et al. [2019b]. We thus prove the upper bound of $A_t$.

We proceed to prove the upper bound of $B_t$. From the self-normalized concentration inequality [Abbasi-Yadkori et al., 2011, Theorem 1], restated in Theorem 5 of Appendix C, we know that

$$
\left\| \sum_{s=t_0}^{t-1} \eta_s X_s \right\|_{V_{t-1}^{-1}} \leq \sqrt{2 R^2 \log \left( \frac{\det(V_{t-1})^{1/2} \det(I)'^{-1/2}}{\delta} \right)}

\leq R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{(t-t_0)L^2}{d} \right)},

\tag{32}

$$

where the last inequality is obtained from the analysis of the determinant, as shown in the proof of Lemma 4.

Meanwhile, since $V_{t-1} \succeq I$, we know that

$$
\left\| \lambda \eta V_{t-1}^{-1} \right\|_2 \leq \frac{1}{\lambda \min(V_{t-1})} \left\| \lambda \eta \right\|_2 \leq \frac{1}{\lambda} \left\| \lambda \eta \right\|_2 \leq \lambda S^2.

$$

Therefore, the upper bound of $B_t$ can be immediately obtained by combining the above inequalities. \hfill \Box

B Bandit-over-Bandits Mechanism and Proof of Theorem 4

The RestartUCB algorithm requires prior information of the path-length $P_T$, which is generally unknown. Such a limitation can be avoided by utilizing the Bandits-over-bandits (BOB) mechanism, proposed by Cheung et al. [2019a] in designing parameter-free algorithm for non-stationary linear bandits based on sliding window least square estimator.

In the following, we first describe how to apply the BOB mechanism to eliminate the requirement of the unknown path-length in RestartUCB. Then, we present the proof of Theorem 4.

B.1 RestartUCB with BOB Mechanism

We name the RestartUCB algorithm with Bandit-over-Bandits mechanism as “RestartUCB-BOB”, whose main idea is illustrated in Figure 4.

From a high-level view, although the exact value of the optimal epoch size (or equivalently, the path-length $P_T$) is not clear, we can make some random guesses of its possible value, since the $P_T$ is always bounded. Then, we can use a certain meta-algorithm to adaptively track the best epoch size, based on the returned reward returned. Specifically, The RestartUCB-BOB algorithm first sets an update period $H_0$, and then runs the RestartUCB with a particular epoch size in each period, and the epoch size will be adaptively adjusted by employing EXP3 [Auer et al., 2002] as the meta-algorithm.

We refer the reader to Section 7.3 of Cheung et al. [2019b] for more descriptions of design motivations and algorithmic details.

In the configuration of RestartUCB-BOB, we set $H_0 = [d\sqrt{T}]$ and the pool of epoch sizes $J$ as

$$
J = \{H_i = [(d/(2S))]^{2/3} \cdot 2^{i-1} \mid i = 1, 2, \ldots, N\},

$$

where $N = \left\lfloor \ln(d^{1/3}T^{1/2}(2S)^{2/3}) \right\rfloor + 1$.

Denoted by $H_{\text{min}}$ ($H_{\text{max}}$) the minimal (maximal) epoch size in the pool $J$, we know that

$$
H_{\text{min}} = [(d/(2S))]^{2/3}, H_{\text{max}} = [d\sqrt{T}] \leq H_0.

\tag{25}$$
B.2 Proof of Theorem 4

Proof of Theorem 4. We begin with the following decomposition of the dynamic regret.

\[
\sum_{t=1}^{T} \langle X_t^*, \theta_t \rangle - \langle X_t, \theta_t \rangle = \sum_{i=1}^{T} \langle X_t^*(H^\dagger), \theta_t \rangle - \sum_{i=1}^{T} \langle X_t,(H^\dagger)\rangle, \theta_t \rangle,
\]

where \( H^\dagger \) is the best epoch size to approximate the optimal epoch size \( H^* \) in the pool \( J \), and \( H^* = \lfloor dT/(1 + P_T) \rfloor^{2/3} \). Hence, it suffices to bound terms (i) and (ii). In the following, we consider two cases, either \((1 + P_T) \geq d^{-1/2}T^{1/4} \) or \((1 + P_T) < d^{-1/2}T^{1/4} \).

Case 1. \((1 + P_T) \geq d^{-1/2}T^{1/4} \).

In this case, it is easy to verify that \( H^* \leq H_{\text{max}} \) and we thus conclude that \( H^* \) lies in the the range of \([H_{\text{min}}, H_{\text{max}}]\). Furthermore, from the configuration of the pool \( J \), we confirm that there exists an epoch size \( H^\dagger \in J \) such that \( H^\dagger \leq H^* \leq 2H^\dagger \). So term (ii) can be upper bounded by

\[
\text{term (ii)} \leq \sum_{i=1}^{\lfloor dT/H_0 \rfloor} \tilde{O}(H^\dagger P_T + \frac{dH_0}{\sqrt{H^\dagger}}) \leq \tilde{O}(H^\dagger P_T + \frac{d}{\sqrt{H^\dagger}}) \leq \tilde{O}(d^{1/3}T^{1/3}),
\]

where (26) is due to Theorem 2 and \( P_t \) denotes the path-length in the \( i \)-th update period. (27) follows by summing over all update periods, and the last inequality holds since the optimal epoch size \( H^* \) is provably in the range of \([H_{\text{min}}, H_{\text{max}}]\) and satisfies \( H^\dagger \leq H^* \leq 2H^\dagger \).

Next, we bound the term (i),

\[
\text{term (i)} \leq \tilde{O}(\sqrt{H_0 NT}) \leq \tilde{O}(d^{1/2}T^{3/4}) \leq \tilde{O}(d^{2/3}T^{2/3}(1 + P_T)^{1/3}),
\]

where the first inequality follows by the same argument as in the sliding window based approach [Cheung et al., 2019b, Lemma 13], building upon the of EXP3. In addition, the last inequality holds due to the fact that \((1 + P_T) \geq d^{-1/2}T^{1/4} \) implies,

\[
d^{1/2}T^{3/4} = d^{2/3}T^{2/3}d^{-1/3}T^{1/6} \leq d^{2/3}T^{2/3}(1 + P_T)^{1/3}.
\]

Hence, by combining the upper bounds of term (i) and term (ii), we know that the dynamic regret of RestartUCB-BOB is bounded by \( \tilde{O}(d^{2/3}T^{2/3}(1 + P_T)^{1/3}) \) under the condition of \((1 + P_T) \geq d^{-1/2}T^{1/4} \).

Case 2. \((1 + P_T) < d^{-1/2}T^{1/4} \).

In this case, we cannot guarantee that the optimal epoch size \( H^* \) lies in the range of \([H_{\text{min}}, H_{\text{max}}]\), so we set \( H^\dagger = H_0 \),

\[
\text{term (ii)} \leq \tilde{O}(H^\dagger P_T + \frac{dH_0}{\sqrt{H^\dagger}}) \leq \tilde{O}(H_0 P_T + \frac{d}{\sqrt{H_0}}) = \tilde{O}(d\sqrt{T}P_T + d^{1/2}T^{3/4}) \leq \tilde{O}(d^{1/2}T^{3/4})
\]

where the last inequality holds by exploiting the condition of \((1 + P_T) \leq d^{-1/2}T^{1/4} \). The result in conjunction with the upper bound of term (i) in (28) gives the \( \tilde{O}(d^{1/2}T^{3/4}) \) dynamic regret under this condition.

Finally, note that the dynamic regret of above two cases can be rewritten in the following unified form,

\[
\text{term (i)} + \text{term (ii)} \leq \tilde{O}(d^{2/3}T^{2/3}(\max\{P_T, d^{-1/2}T^{1/4}\})^{\frac{1}{3}}).
\]

Hence, we complete the proof of Theorem 4. \( \square \)

C Technical Lemmas

In this section, we provide several technical lemmas that frequently used in the proofs.

Theorem 5 (Self-Normalized Bound for Vector-Valued Martingales [Abbasi-Yadkori et al., 2011, Theorem 1]). Let \( \{F_t\}_{t=0}^{\infty} \) be a filtration. Let \( \{\eta_t\}_{t=0}^{\infty} \) be a real-valued stochastic process such that \( \eta_t \) is \( F_t \)-measurable and conditionally \( R \)-sub-Gaussian for some \( R > 0 \), namely,

\[
\forall \lambda \in \mathbb{R}, \quad \mathbb{E}[\exp(\lambda \eta_t | F_{t-1})] \leq \exp\left(\frac{\lambda^2 R^2}{2}\right). \tag{29}
\]

Let \( \{X_t\}_{t=1}^{\infty} \) be an \( \mathbb{R}^d \)-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

\[
\bar{V}_t = V + \sum_{\tau=1}^{t} X_{\tau}X_{\tau}^T, \quad S_t = \sum_{\tau=1}^{t} \eta_{\tau}X_{\tau}. \tag{30}
\]
Then, for any $\delta > 0$, with probability at least $1 - \delta$, for all $t \geq 0$,
\[
\|S_t\|_{\tilde{V}^{-1}_t}^2 \leq 2R^2 \log \left( \frac{\det(\tilde{V}_t)^{1/2} \det(V)^{-1/2}}{\delta} \right). \tag{31}
\]

**Lemma 4** (Elliptical Potential Lemma). Suppose $U_0 = \lambda I$, $U_t = U_{t-1} + X_tX_t^T$, and $\|X_t\|_2 \leq L$, then
\[
\sum_{t=1}^T \|U_{t-1}^{-\frac{1}{2}}X_t\|_2 \leq \sqrt{2dT \log \left( 1 + \frac{L^2 T}{\lambda d} \right)} \tag{32}
\]

**Proof.** First, we have the following decomposition,
\[
U_t = U_{t-1} + X_tX_t^T = U_{t-1}^\frac{1}{2}(I + U_{t-1}^{-\frac{1}{2}}X_tX_t^TU_{t-1}^{-\frac{1}{2}})U_{t-1}^\frac{1}{2}.
\]
Taking the determinant on both sides, we get
\[
det(U_t) = det(U_{t-1}) det(I + U_{t-1}^{-\frac{1}{2}}X_tX_t^TU_{t-1}^{-\frac{1}{2}}),
\]
which in conjunction with Lemma 5 yields
\[
det(U_t) = det(U_{t-1})(1 + \|U_{t-1}^{-\frac{1}{2}}X_t\|_2^2)
\]
\[
\geq det(U_{t-1}) \exp(\|U_{t-1}^{-\frac{1}{2}}X_t\|_2^2/2).
\]
Note that in the first inequality, we utilize the fact that $1 + x \geq \exp(x/2)$ holds for any $x \in [0, 1]$. By taking advantage of the telescope structure, we have
\[
\sum_{t=1}^T \|U_{t-1}^{-\frac{1}{2}}X_t\|_2^2 \leq 2 \log \frac{det(U_T)}{det(U_0)} \leq 2d \log \left( 1 + \frac{L^2 T}{\lambda d} \right),
\]
where the last inequality follows from the fact that $\text{Tr}(U_T) \leq \text{Tr}(U_0) + L^2 T = \lambda d + L^2 T$, and thus $\det(U_T) \leq (\lambda + L^2 T/d)^d$.

Therefore, Cauchy-Schwartz inequality gives,
\[
\sum_{t=1}^T \|U_{t-1}^{-\frac{1}{2}}X_t\|_2 \leq \sqrt{T \sum_{t=1}^T \|U_{t-1}^{-\frac{1}{2}}X_t\|_2^2}
\]
\[
\leq \sqrt{2dT \log \left( 1 + \frac{L^2 T}{\lambda d} \right)}.
\]

**Lemma 5.**
\[
det(I + vv^T) = 1 + \|v\|_2^2. \tag{33}
\]

**Proof.** Notice that

(i) $(I + vv^T)v = (1 + \|v\|_2^2)v$, therefore, $v$ is its eigenvector with $(1 + \|v\|_2^2)$ as the eigenvalue;

(ii) $(I + vv^T)v^\perp = v^\perp$, therefore, $v^\perp \perp v$ is its eigenvector with 1 as the eigenvalue.

Consequently, $\det(I + vv^T) = 1 + \|v\|_2^2$. \qed