Revisiting Weighted Strategy for Non-stationary Parametric Bandits

Jing Wang, Peng Zhao, Zhi-Hua Zhou
National Key Laboratory for Novel Software Technology,
Nanjing University, Nanjing 210023, China
{wangjing, zhaop, zhouzh}@lamda.nju.edu.cn

Abstract
Non-stationary parametric bandits have attracted much attention recently. There are three principled ways to deal with non-stationarity, including sliding-window, weighted, and restart strategies. As many non-stationary environments exhibit gradual drifting patterns, the weighted strategy is commonly adopted in real-world applications. However, previous theoretical studies show that its analysis is more involved and the algorithms are either computationally less efficient or statistically suboptimal. This paper revisits the weighted strategy for non-stationary parametric bandits. In linear bandits (LB), we discover that this undesirable feature is due to an inadequate regret analysis, which results in an overly complex algorithm design. We propose a refined analysis framework, which simplifies the derivation and importantly produces a simpler weight-based algorithm that is as efficient as window/restart-based algorithms while retaining the same regret as previous studies. Furthermore, our new framework can be used to improve regret bounds of other parametric bandits, including Generalized Linear Bandits (GLB) and Self-Concordant Bandits (SCB). For example, we develop an optimal algorithm consisting of a non-stationarity detector and a base algorithm that performs well in near-stationary environments. Whenever the detector examines of each arm are structured with an unknown time-varying parameter, which have been extensively studied in recent years [Cheung et al., 2019, Russac et al., 2019, Zhao et al., 2020, Russac et al., 2020, Faury et al., 2021, Russac et al., 2021, Wei and Luo, 2021, Deng et al., 2022, Liu et al., 2022] due to its significance in many real-world non-stationary online applications such as recommendation systems [Tomkins et al., 2021, Huleihel et al., 2021], as well as tight connection with theoretical foundation of reinforcement learning [Jin et al., 2020, Touati and Vincent, 2020].

Linear Bandits (LB) is a fundamental instance of parametric bandits, where the expected reward for pulling a certain arm at time $t$ is the inner product between the arm’s feature vector $X_t$ and an unknown parameter $\theta_t$, namely, $E[r_t | X_t] = X_t^\top \theta_t$. Moreover, Generalized Linear Bandits (GLB) is introduced as a generalization of LB to model a broader range of reward functions such as binary rewards, where the expected reward obeys a generalized linear model as $E[r_t | X_t] = \mu(X_t^\top \theta_t)$ with $\mu(\cdot)$ being an inverse link function. Notably, the non-stationary models allow the parameter $\theta_t$ in the above models to be time-varying. There are two typical non-stationarity measures to quantify the intensity of parameter changes: (i) in gradually drifting cases, path length $P_T = \sum_{t=2}^T \|\theta_{t-1} - \theta_t\|_2$ is used to measure the cumulative variations of the underlying parameters; and (ii) in piecewise-stationary cases, $\Gamma_T$ denotes the number of parameter changes in $T$ rounds.

To deal with non-stationarity, there are three principled ways: sliding-window, weighted, and restart strategies. For the sliding-window strategy, the learner maintains a time window that contains the most recent observed data to discard the outdated data. For the weighted strategy, the learner puts more weight on the most recent data and less weight on the old data to gradually forget the outdated data. For the restart strategy, the learner restarts the algorithm according to a certain period to discard the outdated data. The currently best-known result for non-stationary (generalized) linear bandits is by Wei and Luo [2021], who developed an optimal algorithm consisting of a non-stationarity detector and a base algorithm that performs well in near-stationary environments.

1 INTRODUCTION
Non-stationary parametric bandits model the sequential decision-making problems where the reward distributions...
that the non-stationarity exceeds a certain limit, the algorithm will restart itself to resist the non-stationarity. In this sense, their algorithm can be regarded as an adaptive restart-based algorithm. Building on the RestartUCB algorithm [Zhao et al., 2020] and a carefully designed non-stationarity detector with multi-scale explorations, their algorithm can achieve an $\tilde{O}(\min\{\sqrt{T}T, P_t^{1/3}T^{2/3}\})$ optimal dynamic regret for both LB and GLB.

In real-world scenarios, the distribution change of environments often exhibits gradually drifting patterns [Crammer et al., 2010, Chiang et al., 2013, Gama et al., 2014], in such cases, a soft weighted strategy can be (empirically) more advantageous than a hard restart strategy to deal with the non-stationarity, as can be observed in bands learning [Russac et al., 2019, Zhao et al., 2020, Deng et al., 2022], classification with concept drift [Anagnostopoulos et al., 2012, Zhao et al., 2021a], and adaptive system identification [Guo et al., 1993, Chu and Mak, 2017]. As a result, it will be highly attractive to design an adaptive weight-based algorithm for non-stationary parametric bandits, which imposes weights to discount the importance of past data, and the weights are set adaptively according to environments. Towards this end, we examine existing methods for non-stationary parametric bandits based on the weighted strategy, and (surprisingly) find that current results exhibit unnatural gaps compared to the other strategies, such as restart-based algorithms, as well as unnatural regret analysis transitions from GLB to LB.

Those unnatural phenomena motivate us to revisit the algorithm design and regret analysis of the weighted strategy for non-stationary parametric bandits [Russac et al., 2019, 2021, Faury et al., 2021]. Indeed, the key ingredient is the estimation error analysis for the weight-based estimator, which is usually decomposed into two parts — one is the bias part due to the parameter drift, and the other is the variance part due to the stochastic noise. Generally, the bias part is controlled by non-stationary strategies, and the variance part is handled by carefully designed concentration. Russac et al. [2019] provided the first analysis of the weight-based algorithm for LB, where they introduced a virtual window size in the analysis to control the bias in order to mimic the analysis of sliding-window strategy [Cheung et al., 2019]. Consequently, they have to use different local norms to control bias and variance parts, resulting in unexpected inefficiencies and complications. For LB, this leads to an algorithm D-LinUCB [Russac et al., 2019] requiring to maintain an extra covariance matrix, which is less efficient than the window-based and restart-based algorithms [Cheung et al., 2019, Zhao et al., 2020].

This analysis framework for weighted strategy introduces more severe issues in GLB, due to its more enriched and complicated structure. Specifically, Faury et al. [2021] studied the drifting GLB and designed a highly complex projection operation to control bias and variance parts following the way of Russac et al. [2019] to mimic sliding-window analysis, and finally attained an $\tilde{O}(d^{1/10}P_t^{1/5}T^{3/5})$ dynamic regret. Unfortunately, this cannot recover the $\tilde{O}(d^{2/3}P_t^{1/3}T^{2/3})$ bound enjoyed by the weight-based algorithm for drifting LB (a special case of GLB) [Russac et al., 2019]. Subsequently, Russac et al. [2021] investigated the non-stationary Self-Concordant Bandits (SCB), a subclass of GLB with many attractive structures. They can only conduct analysis under the piecewise-stationary setting, whereas failed in the more challenging drifting setting, due to technical difficulties in bounding bias using conventional analysis. As such, two open questions are proposed in their papers: (i) how to extend weight-based algorithms to drifting SCB; and (ii) how to replicate recent progress in stationary SCB [Abeille et al., 2021] to improve dependence on $c_\mu$ in non-stationary SCB.

**Our Results.** In this paper, we revisit the weighted strategy for the non-stationary parametric bandits. We discover that the earlier analysis framework may be inappropriate due to mimicking the sliding-window analysis, which demands bounding the bias and variance parts using different local norms. We present a refined analysis framework for the weighted strategy, in which a new analysis for the bias part is presented such that it is now allowed to use the same local norm to analyze both bias and variance parts. This refined analysis framework not only simplifies the re-

<table>
<thead>
<tr>
<th>Parametric Bandit Models</th>
<th>Previous Work</th>
<th>Our Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drifting LB</td>
<td>$\tilde{O}(d^{7/8}P_t^{1/4}T^{3/4})$ [Russac et al., 2019]</td>
<td>$\tilde{O}(d^{2/3}P_t^{1/3}T^{2/3})$ [Theorem 1]</td>
</tr>
<tr>
<td>Drifting GLB</td>
<td>$\tilde{O}\left(\frac{k^2}{c_\mu}d^{1/10}P_t^{1/5}T^{4/5}\right)$ [Faury et al., 2021]</td>
<td>$\tilde{O}\left(\frac{k^2}{c_\mu}d^{1/4}P_t^{1/3}T^{3/4}\right)$ [Theorem 2]</td>
</tr>
<tr>
<td>Drifting SCB</td>
<td>$\tilde{O}\left(\frac{k^2}{c_\mu}d^{1/10}P_t^{1/5}T^{4/5}\right)$ [Faury et al., 2021]</td>
<td>$\tilde{O}\left(\frac{k^2}{c_\mu}d^{1/4}P_t^{1/3}T^{3/4}\right)$ [Theorem 3]</td>
</tr>
<tr>
<td>Piecewise Stationary SCB</td>
<td>$\tilde{O}(\frac{1}{c_\mu}d^{2/3}T^{1/3}T^{2/3})$ [Russac et al., 2021]</td>
<td>$\tilde{O}(d^{2/3}T^{1/3}T^{2/3})$ [Theorem 4]</td>
</tr>
</tbody>
</table>
regret analysis but also brings many benefits in algorithm designs, including improving efficiency for LB and resolving the projection problem brought by GLB and SCB. Table 1 summarizes our main results compared to earlier best-known results of weight-based algorithms. Specifically, based on our refined analysis framework, we achieve the following results: (i) for LB, our approach only needs to maintain one covariance instead of two and still enjoys the same regret as [Russac et al., 2019]; (ii) for GLB, our approach enjoys an $\tilde{O}(k_\mu c_\mu^{-1/2}d^{3/4}P_T^{5/4}T^{3/4})$ regret bound, whose order of $d$, $P_T$ and $T$ matches that in LB case; and (iii) for SCB, we achieve an $\tilde{O}(k_\mu c_\mu^{-1/2}d^{3/4}P_T^{5/4}T^{3/4})$ regret bound. In addition, for piecewise stationary SCB, our approach achieves an $\tilde{O}(d^{3/4}P_T^{3/4}T^{5/4})$ regret bound that can get rid of the influence of $c_\mu$, resolving the second open problem asked by Russac et al. [2021].

2 RELATED WORK

Linear Bandits. Non-stationary LB problem was first studied by Cheung et al. [2019]. They established an $\Omega(d^{3/4}P_T^{3/4}T^{5/4})$ minimax lower bound and then proposed SW-UCB algorithm based on the sliding-window strategy. Then Russac et al. [2019] proposed the D-LinUCB algorithm based on weighted strategy and Zhao et al. [2020] proposed the RestartUCB algorithm based on restart strategy. Note that the three works proved an $\tilde{O}(d^{3/4}P_T^{3/4}T^{5/4})$ regret bound, but there exists a subtle technical gap in the regret analysis as identified by Zhao and Zhang [2021]. After fixing the technical gap, all three aforementioned algorithms achieve an $\tilde{O}(d^{3/4}P_T^{3/4}T^{5/4})$ regret bound [Zhao and Zhang, 2021, Zhao et al., 2021b]. However, to achieve this result, all the three algorithms require the knowledge of the path length $P_T$ as an input at the beginning of algorithmic implementation, which is undesired. To address so, Cheung et al. [2019] proposed the bandits-over-bandits (BOB) strategy as a meta-algorithm to learn the unknown parameter $P_T$ which can be combined with the above algorithms to remove the requirement of this prior knowledge. Afterward, Wei and Luo [2021] proposed the MASTER algorithm with theoretically optimal $\tilde{O}(\min\{d\sqrt{T}, dP_T^{3/4}T^{2/3}\})$ regret bound, also without requiring the non-stationarity level of environments (that is, $\Gamma_T$ and $P_T$) in advance. Most recently, there also has been some new progress in the non-stationary (linear) bandits [Liu et al., 2022, Suk and Kpotufe, 2022, Abbasi-Yadkori et al., 2022, Clerici et al., 2023].

Generalized Linear Bandits. GLB problem was first introduced by Filippi et al. [2010]. They proposed GLM-UCB algorithm, achieving an $\tilde{O}(k_\mu c_\mu^{-1}d\sqrt{T})$ regret bound where $k_\mu$, $c_\mu$ are the problem-dependent constants and $k_\mu/c_\mu$ represents the nonlinearity of the generalized linear model. Faury et al. [2021] extended the stationary GLB to the drifting case, and proposed BVD-GLM-UCB algorithm with $\tilde{O}(k_\mu c_\mu^{-1}d^{3/4}P_T^{1/4}T^{4/3})$ regret bound. Faury et al. [2020] studied a specific instance of GLB called Logistic Bandits (LogB), they first pointed out that under GLB setting, the problem-dependent constant $1/c_\mu$ could be very large in some cases like LogB, then they proposed the Logistic-UCB-1 algorithm with an $\tilde{O}(c_\mu^{-1/2}d\sqrt{T})$ regret bound and the Logistic-UCB-2 algorithm with an $\tilde{O}(d\sqrt{T} + c_\mu^{-1})$ regret bound. Subsequently, Abeille et al. [2021] established an $\Omega(d\sqrt{\mu(X^*_t\theta_t)^T})$ regret lower bound for logistic bandits and provided an optimal algorithm OFULog. Russac et al. [2021] generalized the logistic bandits to self-concordant bandits and considered the piecewise-stationary case, their algorithm enjoys an $\tilde{O}(c_\mu^{-1/3}d^{3/4}\Gamma_T^{1/3}T^{2/3})$ regret bound. To deal with $P_T$-unknown cases, Faury et al. [2021] proposed a parameter-free algorithm by combining BVD-GLM-UCB with BOB strategy, but the final result is still suboptimal. Meanwhile, the optimal black-box algorithm [Wei and Luo, 2021] can also adaptively restart stationary algorithm GLM-UCB [Filippi et al., 2010] and achieve an optimal $\tilde{O}(\min\{k_\mu c_\mu^{-1}\sqrt{T}T, k_\mu^{3/4}c_\mu^{-1}(dP_T^{3/4}T^{-3/4})\})$ regret.

3 LINEAR BANDITS

In this section, we first introduce the problem setting of non-stationary LB, and then describe our LB-WeightUCB algorithm and its theoretical guarantee. Our algorithm has the same regret bound as the best-known weight-based algorithm [Russac et al., 2019] but is more efficient.

3.1 Problem Setting

At each round $t$, the learner chooses an arm $X_t$ from a feasible set $\mathcal{X} \subseteq \mathbb{R}^d$ and receives a reward $r_t$ such that

$$r_t = X_t^\top \theta_t + \eta_t,$$

where $\theta_t \in \mathbb{R}^d$ is the unknown time-varying parameter and $\eta_t$ is the $R$-sub-Gaussian noise. The goal of the learner is to minimize the following (pseudo) dynamic regret:

$$R_T = \sum_{t=1}^{T} \max_{x \in \mathcal{X}} X_t^\top \theta_t - \sum_{t=1}^{T} X_t^\top \theta_t,$$

which is the cumulative regret against the optimal strategy that has full information of the unknown parameter. Here we consider the drifting case where we use path length $P_T = \sum_{t=2}^{T} \|\theta_{t-1} - \theta_t\|_2$ as the non-stationarity measure.

We work under the following standard boundedness assumption [Abbasi-Yadkori et al., 2011, Cheung et al., 2019, Russac et al., 2019, Zhao et al., 2020].

**Assumption 1.** The feasible set and unknown parameters are assumed to be bounded: $\forall x \in \mathcal{X}$, $\|x\|_2 \leq L$, and $\theta_t \in \Theta$ holds for all $t \in [T]$ where $\Theta \triangleq \{\theta \mid \|\theta\|_2 \leq S\}$. 
3.2 Algorithm and Regret Guarantee

We propose the LB-WeightUCB algorithm, which attains the same guarantees as earlier methods while having higher efficiency. We first give the employed estimator and then derive its estimation error upper bound by our refined analysis framework, which is the key for algorithm design and regret analysis. Based on the estimation error bound, we propose our selection criterion and finally give the theoretical guarantee on its dynamic regret.

Estimator. We adopt the weighted regularized least square estimator same as D-LinUCB [Russac et al., 2019], the estimator \( \hat{\theta}_t \) is the solution to the following problem,

\[
\min_{\theta} \lambda \|\theta\|_2^2 + \sum_{s=1}^{t-1} \gamma^{t-s-1} (X_s^\top \theta - r_s)^2,
\]

where \( \lambda > 0 \) is the regularization coefficient and \( \gamma \in (0, 1) \) is the discounted factor. Clearly, \( \hat{\theta}_t \) admits a close-form solution \( \hat{\theta}_t = V_{t-1}^{-1}\left(\sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s \right) \), where \( V_t = \lambda I_d + \sum_{s=1}^{t} \gamma^{t-s-1} X_s X_s^\top \), \( V_0 = \lambda I_d \) is the covariance matrix. Note that this close-form solution can be further transformed into a recursive formula [Haykin, 2002, Chapter 10.3]. This allows it to be updated online without the need to store historical data, which is another important computational advantage of the weighted strategy compared to the sliding-window strategy.

Upper Confidence Bounds. For estimator (3), we provide the following estimation error bound. Notably, this is different from the previous result [Russac et al., 2019, Appendix B.3, second and third steps in Proof of Theorem 2], which is the key component being a more efficient algorithm and will be explained later.

Lemma 1. For any \( x \in \mathcal{X} \), \( \gamma \in (0, 1) \) and \( \delta \in (0, 1) \), with probability at least \( 1 - \delta \), the following holds for all \( t \in [T] \)

\[
| X_t^\top (\hat{\theta}_t - \theta_t) | \leq L^2 \sqrt{d} \sum_{p=1}^{t-1} \gamma^{t-p} \sqrt{\frac{\gamma^{p-1} - 1}{1 - \gamma}} \|\theta_p - \theta_{p+1}\|_2 + \beta_{t-1} \| x \|_{V_{t-1}^{-1}},
\]

where \( \beta_t \) is the radius of confidence region set by

\[
\beta_t = \sqrt{\lambda S + R \sqrt{2 \log \frac{1}{\delta}} \sqrt{d \log \left(1 + \frac{L^2 (1 - \gamma^{2t})}{\lambda d (1 - \gamma^2)}\right)}}.
\]

The proof of Lemma 1 is in Appendix A.2. Based on Lemma 1, we can specify the arm selection criterion as

\[
X_t = \arg \max_{x \in \mathcal{X}} \left\{ \left(x, \hat{\theta}_t\right) + \beta_{t-1} \|x\|_{V_{t-1}^{-1}} \right\}.
\]

The overall algorithm is summarized in Algorithm 1. From the update procedure in Line 5 of Algorithm 1, we can observe that our algorithm needs to maintain a single covariance matrix \( V_{t-1} \in \mathbb{R}^{d \times d} \). By contrast, the selection criterion of algorithm proposed in Russac et al. [2019] is like

\[
X_t = \arg \max_{x \in \mathcal{X}} \left\{ \left(x, \hat{\theta}_t\right) + \beta_{t-1} \|x\|_{V_{t-1}^{-1}} \right\},
\]

where \( \beta_{t-1} \) and \( V_{t-1}^{-1} \) are identical with those in our selection criterion (6) and \( \hat{\theta}_t \) is the solution to the following problem, with\( \beta_{t-1} \) and \( V_{t-1}^{-1} \) in advance. To achieve a parameter-

Algorithm 1 LB-WeightUCB

**Input:** time horizon \( T \), discounted factor \( \gamma \), confidence \( \delta \), regularization \( \lambda \), parameters \( S \), \( L \) and \( R \)

1. Set \( V_0 = \lambda I_d \), \( \hat{\theta}_1 = 0 \) and compute \( \beta_0 \) by (5)
2. for \( t = 1, 2, ..., T \) do
3. Select \( X_t = \arg \max_{x \in \mathcal{X}} \left\{ \left(x, \hat{\theta}_t\right) + \beta_{t-1} \|x\|_{V_{t-1}^{-1}} \right\} \)
4. Compute the reward \( r_t \)
5. Update \( V_t = \gamma V_{t-1} + X_t X_t^\top + (1 - \gamma) I_d \)
6. Compute \( \hat{\theta}_{t+1} \) by (3) and \( \beta_t \) by (5)
7. end for
free result for unknown \( P_T \) case, our algorithm can be combined with the BOB strategy \([\text{Cheung et al.}, 2019]\) and achieves an \( \tilde{O}(d^{3/4} P_T^{-1/4} T^{3/4}) \) bound. However, this bound is not optimal, and it is possible to design an adaptive weight-based algorithm based on our result, in the spirit of \( \text{Wei and Luo} [2021] \), to further achieve an optimal dynamic regret without prior knowledge of \( P_T \). This is very challenging since that at each round \( t \in [T] \), we can only receive one data pair \((X_t, r_t)\), which is not adequate for the learner to real-time update the discounted factor \( \gamma_t \). At the same time, MASTER algorithm [\text{Wei and Luo}, 2021] can be considered as a special case of the adaptive weight-based algorithm since it only includes two circumstances: setting \( \gamma_t = 0 \) to restart at time \( t \) and setting \( \gamma_t = 1 \) to keep going. But for the adaptive weight-based algorithm, the choice of the discounted factor \( \gamma_t \) can be continuous in \([0, 1]\), which is more difficult than a binary decision. We leave this as an important open question for future study.

4 GENERALIZED LINEAR BANDITS

In this section, we apply the weighted strategy to drifting GLB. Compared to the best-known weight-based algorithm for drifting GLB [\text{Faury et al.}, 2021], our algorithm is simpler and meanwhile has a better theoretical guarantee.

4.1 Problem Setting

GLB assumes an inverse link function \( \mu : \mathbb{R} \to \mathbb{R} \) such that \( r_t = \mu(X_t^\top \theta_t) + \eta_t \), where \( \theta_t \in \mathbb{R}^d \) is the unknown parameter and can change over time. Similar to LB, we define dynamic regret for GLB as follows:

\[
R_T = \sum_{t=1}^{T} \left( \max_{x \in \mathcal{X}} \mu'(x^\top \theta_t) - \mu'(X_t^\top \theta_t) \right). \tag{7}
\]

Under the GLB setting, we make the same assumptions as those of LB, which include \( R \)-sub-Gaussian noise, boundedness of feasible set and unknown regression parameters (Assumption 1). In addition, we work under the standard boundedness assumption of the inverse link function [Filippi et al., 2010, Li et al., 2017, Faury et al., 2021].

Assumption 2. The inverse link function \( \mu : \mathbb{R} \to \mathbb{R} \) is \( k_\mu \)-Lipschitz, and continuously differentiable with

\[
c_\mu (\tilde{\theta}) \triangleq \inf_{\theta \in \Theta} \mu'((\tilde{\theta}^\top x)) > 0, \quad \Theta = \{ \theta \mid \| \theta \|_2 \leq S \}.
\]

Previous works [\text{Zhao et al.}, 2020, \text{Cheung et al.}, 2022] define a similar parameter \( \tilde{c}_\mu \) and obtain regret upper bound scaling with \( 1/\tilde{c}_\mu \). Clearly, that \( \tilde{c}_\mu \) is smaller than our defined \( c_\mu \) (and can be much smaller) as \( c_\mu \) is defined on \( \Theta \) while \( \tilde{c}_\mu \) is defined on \( \mathbb{R} \). Therefore, \( \tilde{c}_\mu \) is less attractive to appear in the upper bound.

4.2 Algorithm and Regret Guarantee

We propose GLB-WeightUCB, which is a simpler algorithm with better theoretical guarantee compared to previous weight-based algorithm [\text{Faury et al.}, 2021]. The key improvement is owing to our refined analysis framework, which is compatible with a simple projection step.

Estimator. At iteration \( t \), we first adopt the quasi-maximum likelihood estimator (QMLE) without considering the projection onto the feasible domain. Specifically, the estimator \( \hat{\theta}_t \) is the solution of the following weighted regularized estimation equation:

\[
\lambda c_\mu \hat{\theta} + \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu'(X_s^\top \theta) - r_s) X_s = 0. \tag{8}
\]

Given that \( \hat{\theta}_t \) may not belong to the feasible set \( \Theta \) and \( c_\mu \) is defined over the parameter \( \theta \in \Theta \), we need to perform the following projection step

\[
\tilde{\theta}_t = \arg \min_{\theta \in \Theta} \| g_t(\hat{\theta}_t) - g_t(\theta) \|_{V_t^{-1}}, \tag{9}
\]

where \( V_t = \lambda I_d + \sum_{s=1}^{t} \gamma^{t-s} X_s X_s^\top \) and \( g_t(\theta) \) is

\[
g_t(\theta) \triangleq \lambda c_\mu \theta + \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu'(X_s^\top \theta) - r_s) X_s. \tag{10}
\]

However, previous work [\text{Faury et al.}, 2021] cannot conduct the same simple projection in the drifting case as stationary GLB or piecewise-stationary GLB, since they use different local norms to measure the bias and variance parts separately for estimation error analysis. Consequently, they have to design a complicated projection to ensure that the bias and variance parts could be measured by different local norms (see [Faury et al., 2021, Section 4.1], and our restatements in Appendix B.1).
Our refined analysis framework is compatible with this projection operation, thanks to our analysis framework utilizing the same local norm for the bias and variance parts.

Upper Confidence Bounds. For estimator (8) with projection (9), we construct following estimation error bound.

**Lemma 2.** For any \( x \in X, \gamma \in (0, 1) \) and \( \delta \in (0, 1) \), with probability at least \( 1 - \delta \), the following holds for all \( t \in [T] \)

\[
\left| \mu^{\top} \tilde{\theta}_t - \mu(x^{\top} \tilde{\theta}_t) \right| \leq \frac{2k_{\mu}}{c_{\mu}} \left( \sum_{p=1}^{t-1} C(p) \| \theta_p - \theta_{p+1} \|_2 + \beta_{t-1} \| x \|_{V_{t-1}^{-1}} \right),
\]

where \( C(p) = k_{\mu} L^2 \sqrt{\frac{\gamma^{p+1}}{\gamma + 1}} \sqrt{\frac{\gamma^{p+1}}{\gamma + 1}}, \) and \( \beta_t \) is the radius of confidence region set by

\[
\beta_t = \sqrt{X_c \mu} S + R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2 (1 - \gamma^2)}{\lambda d (1 - \gamma^2)} \right)}.
\] *(11)*

The proof of Lemma 2 is in Appendix B.2. Then, based on Lemma 2, we can specify the arm selection criterion as

\[
X_t = \arg \max_{x \in X} \left\{ \mu(x^{\top} \tilde{\theta}_t) + \frac{2k_{\mu}}{c_{\mu}} \beta_{t-1} \| x \|_{V_{t-1}^{-1}} \right\}.
\] *(12)*

The overall algorithm is summarized in Algorithm 2.

Notice that the estimation equation (8) and the confidence radius (11) are the same as those used in Algorithm 1 of Faury et al. [2021]. But importantly, the final (projected) estimators of the two approaches are significantly different. With a simpler projection operation and our refined analysis framework, we can immediately attain an improved regret guarantee for weight-based algorithm.

**Theorem 2.** For all \( \gamma \in (1/T, 1), \lambda = d/c_{\mu}^2 \), the regret of GLB-WeightUCB (Algorithm 2) is bounded with probability at least \( 1 - 1/T \), by

\[
R_T \leq \hat{O} \left( k_{\mu}^2 \frac{1}{(1 - \gamma)^{3/2}} P_T + \frac{k_{\mu}}{c_{\mu}} (d(1 - \gamma)^{1/2} T) \right).
\]

By optimally setting \( \gamma = 1 - \max\{1/T, \sqrt{k_{\mu} c_{\mu} P_T / (dT)} \} \), GLB-WeightUCB achieves the following dynamic regret,

\[
R_T \leq \begin{cases} 
\hat{O} \left( \frac{k_{\mu}^3 d^{3/4}}{c_{\mu}^3} \right) & \text{when } P_T \geq \frac{d}{k_{\mu} c_{\mu} T}, \vspace{1.5mm} \\
\hat{O} \left( \frac{k_{\mu}}{c_{\mu}^3} d \sqrt{T} \right) & \text{when } 0 \leq P_T < \frac{d}{k_{\mu} c_{\mu} T}.
\end{cases}
\]

Compared to BVD-GLM-UCB (the best-known weight-based algorithm for drifting GLB) [Faury et al., 2021], focusing on the dependence on \( d, P_T, \) and \( T \), we can see that our approach improves the regret from \( \hat{O}\left(c_{\mu}^{\gamma/2} (d^{1/2} P_T^{1/2} T^{3/4}) \right) \) to \( \hat{O}\left(d^{3/4} P_T^{1/2} T^{3/4} \right) \). Furthermore, our result also improves their result upon the \( c_{\mu} \) dependence from \( c_{\mu}^{\gamma/4} \) to \( c_{\mu}^{3/4} \).

## 5 SELF-CONCORDANT BANDITS

This section studies Self-Concordant Bandits (SCB), an important subclass of GLB with many attractive structures.

### 5.1 Problem Setting

For SCB, the reward’s distribution belongs to a canonical exponential family: \( \mathbb{P} \{ r | x \} = \exp(\mathbf{x}^\top \theta - b(\mathbf{x}^\top \theta) + c(\mathbf{r})) \) where \( b(\cdot) \) is a twice continuously differentiable function and \( c(\cdot) \) is a real-valued function. Owing to the benign properties of exponential families, we have \( \mathbb{E} \{ r | x \} = b'(\mathbf{x}^\top \theta) \) and \( \text{Var} \{ r | x \} = b''(\mathbf{x}^\top \theta) \) where \( b' \) denotes the first derivative of the function \( b \), and \( b'' \) denotes its second derivative. Then, we can introduce the (inverse) link function \( \mu(\cdot) \equiv b'(\cdot) \) such that at \( t \in [T] \) the following holds

\[
\mathbb{E} \{ r_t | X_t \} = \mu(X_t^\top \theta_t), \quad \text{Var} \{ r_t | X_t \} = \mu'(X_t^\top \theta_t).
\] *(13)*

SCB requires the link function satisfy \( |\mu''| \leq \mu' \), usually referred to general self-concordant property. We further introduce the notation \( \eta_t = \rho_t - \mu(X_t^\top \theta_t) \) to denote the noise. SCB successfully models many important real-world applications and captures the reward structure. For example, choosing \( \mu(x) = (1 + e^{-2x})^{-1} \) yields the Logistic Bandits (LogB), which is often adopted to model the binary-feedback reward in recommendation system [Zhang et al., 2016, Jun et al., 2017, Dong et al., 2019].

We make several standard assumptions same as LB and GLB, including boundedness of feasible set and unknown regression parameters (Assumption 1), and non-linearity measure on link function (Assumption 2). In addition, similar to Russac et al. [2021], we need assumptions on boundedness of reward, and for the convenience of analysis we let \( L = 1 \) which means \( \| x \|_2 \leq 1 \) for all \( x \in X \).

**Assumption 3.** The reward received at each round satisfies

\( 0 \leq r_t \leq m \) for all \( t \in [T] \) and some constant \( m > 0 \).

### 5.2 Algorithm and Regret Guarantee

We propose the SCB-WeightUCB algorithm. Compared to GLB, we use a new local norm for projection and regret analysis which is the key to improving the order of \( c_{\mu}^{-1} \).

**Estimator.** At iteration \( t \), we first adopt the same maximum likelihood estimator as GLB which is defined in (8). Different from GLB, here we use a new local norm to perform the projection onto the feasible set \( \Theta \).

\[
\tilde{\theta}_t = \arg \min_{\theta \in \Theta} \left\{ \| g_t(\tilde{\theta}_t) - g_t(\theta) \|_{H_t^{-1}(\theta)} \right\},
\] *(14)*

where \( g_t(\theta) \) is the same as (10) while \( H_t(\theta) \) is defined as

\[
H_t(\theta) \equiv \lambda c_{\mu} I_d + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu'(X_s^\top \theta) X_s X_s^\top.
\] *(15)*

Notably, compared to \( V_t \), \( H_t(\theta) \) depends on the function
curvature along the dynamics and thus can capture more local information. Combining this projection step with the weighted version self-normalized concentration as restated in Theorem 6 will remove a constant $c_{\mu}^{-1/2}$ in regret bound.

**Upper Confidence Bound.** For estimator (8) with projection (14), we construct following estimation error bound.

**Lemma 3.** For any $\mathbf{x} \in \mathcal{X}$, $\gamma \in (0, 1)$ and $\delta \in (0, 1)$, with probability at least $1 - \delta$, the following holds for all $t \in [T]$

$$\left| \mu(\mathbf{x}^T \bar{\theta}_t) - \mu(\mathbf{x}^T \theta_t) \right| \leq \frac{\sqrt{4 + 8 S} \kappa_\mu}{\sqrt{c_\mu}} \left( \sum_{p=1}^{t-1} C(p) \| \theta_p - \theta_{p+1} \|_2 + \bar{\beta}_{t-1} \| \mathbf{x} \|_{V_{t-1}^{-1}} \right),$$

where $C(p) = L^2 \sqrt{d \kappa_\mu \gamma^{\frac{t-p-1}{1-\gamma}}} \sqrt{\gamma^{-p-1} - \gamma^{-t-1}}$, and $\bar{\beta}_t$ is the radius of confidence region set by

$$\bar{\beta}_t = \frac{\sqrt{\lambda \mu}}{2m} + \frac{2m}{\sqrt{\lambda \mu}} \left( \log \frac{1}{\delta} + d \log 2 \right) + \frac{d \mu}{\sqrt{\lambda \mu}} \log \left( 1 + \frac{\lambda^2 \mu (1 - \gamma^2)}{1 - \gamma^2} \right) \sqrt{\frac{\lambda \mu S}{d}}. \tag{16}$$

The proof of Lemma 3 is in Appendix C.1. Based on Lemma 3, we can specify the arm selection criterion as

$$X_t = \arg \max_{x \in \mathcal{X}} \left\{ \mu(\mathbf{x}^T \bar{\theta}_t) + 2 \sqrt{1 + 2 S} \frac{\kappa_\mu}{\sqrt{c_\mu}} \bar{\beta}_{t-1} \| \mathbf{x} \|_{V_{t-1}^{-1}} \right\}. \tag{17}$$

Our algorithm for SCB (named SCB-WeightUCB) follows the same procedure of Algorithm 2, and the difference is that $\bar{\theta}_t$ is computed by (14), $\bar{\beta}_{t-1}$ is computed by (16) and $\bar{\mathbf{x}}_t$ is computed by (17). Further, we have the following guarantee for SCB-WeightUCB algorithm.

**Theorem 3.** For all $\gamma \in (1/T, 1)$, $\lambda = d \log(T)/c_\mu$, the dynamic regret of SCB-WeightUCB is bounded with probability at least $1 - 1/T$, by

$$R_T \leq \bar{O} \left( \frac{\kappa_\mu^2}{\sqrt{c_\mu}} \frac{1}{1 - \gamma} \frac{\gamma^{1/2}}{\sqrt{P_T}} + \frac{\kappa_\mu}{\sqrt{c_\mu}} d(1 - \gamma)^{1/2} \right). \tag{18}$$

By setting $\gamma = 1 - \max\{1/T, \sqrt{\lambda \mu P_T/(d T)} \}$, we achieve

$$R_T \leq \begin{cases} \bar{O} \left( \frac{\kappa_\mu^2}{c_\mu} \frac{d^{3/4}}{\sqrt{P_T}} \delta^{1/4} \frac{1}{T^{3/4}} \right) & \text{when } P_T \geq \frac{d}{\kappa_\mu}, \\ \bar{O} \left( \frac{\kappa_\mu}{c_\mu} \delta \sqrt{T} \right) & \text{when } 0 \leq P_T < \frac{d}{\kappa_\mu}. \end{cases}$$

Compared to GLB, we improve the order of $c_\mu$ from $c_\mu^{-1}$ to $c_\mu^{-1/2}$ by exploiting the self-concordant properties. At the same time, in near-stationary environments ($P_T$ is small enough), our result can recover to the performance of LogUCB1 algorithm [Faury et al., 2020]. The proof of Theorem 3 is presented in Appendix C.2.

In addition, for the piecewise-stationary SCB, we propose SCB-PW-WeightUCB algorithm that gets rid of influence of $c_\mu$ and thus directly improves upon [Russac et al., 2021].

**Theorem 4.** For all $\gamma \in (1/2, 1)$, $D = \log(T)/\log(1/\gamma)$ and $\lambda = d \log(T)/c_\mu$, the regret of SCB-PW-WeightUCB is bounded with probability at least $1 - 1/T$, by

$$R_T \leq \bar{O} \left( \frac{1}{1 - \gamma} \frac{1}{\sqrt{1 - \gamma}} + d \sqrt{(1 - \gamma) T} \right). \tag{19}$$

By setting $\gamma = 1 - \max\{1/T, (\Gamma_T/(d T))^{1/2} \}$, we achieve

$$R_T \leq \begin{cases} \bar{O} \left( d^{3/4} \Gamma_T^{1/4} T^{3/4} \right) & \text{when } \Gamma_T \geq d / \sqrt{T}, \\ \bar{O} \left( d \sqrt{T} \right) & \text{when } 0 \leq \Gamma_T < d / \sqrt{T}. \end{cases}$$

The overall algorithm and analysis are in Appendix D.

### 6 REFINED ANALYSIS FRAMEWORK

This section presents a proof sketch for Lemma 1 (estimation error analysis for weighted linear bandits), which also serves as a description of our proposed analysis framework.

**Proof Sketch.** From the model assumption (1) and the estimator (3), the estimation error can be split into two parts,

$$\hat{\theta}_t - \theta_t = \sqrt{V_{t-1}^{-1}} \left( \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s \mathbf{x}_s^T (\theta_s - \theta_t) \right) \tag{20}$$

$$\quad + \sqrt{V_{t-1}^{-1}} \left( \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \right),$$

where the bias part is caused by the parameter drifting, and the variance part is due to the stochastic noise. Then, by the Cauchy-Schwarz inequality, for any $\mathbf{x} \in \mathcal{X}$,

$$|\mathbf{x}^T (\hat{\theta}_t - \theta_t)| \leq |\mathbf{x}|_{V_{t-1}^{-1}} (A_t + B_t), \tag{21}$$

where $A_t = \| \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s \mathbf{x}_s^T (\theta_s - \theta_t) \|_{V_{t-1}^{-1}}$ and $B_t = \| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \|_{V_{t-1}^{-1}}$.

Choosing an appropriate local norm for (21) is the key to simplify and improve the estimation error analysis. We note that the previous analysis [Russac et al., 2019] had to use **different** local norms: using $l_2$-norm in the bias part, and $V_{t-1}^{-1} - V_{t-1}^{-1}$-norm in the variance part, namely,

$$|\mathbf{x}^T (\hat{\theta}_t - \theta_t)| \leq |\mathbf{x}|_2 A'_t + |\mathbf{x}|_{V_{t-1}^{-1} - V_{t-1}^{-1}} B'_t, \tag{22}$$

where $A'_t = \| \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s \mathbf{x}_s^T (\theta_s - \theta_t) \|_2$ and $B'_t = \| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \|_{V_{t-1}^{-1} - V_{t-1}^{-1}}$ and $\tilde{V}_t = \lambda I + \sum_{s=1}^{t-1} \gamma^{2(t-s)} X_s X_s^T$. Since the need for using sliding-window analysis to analyze the bias part, they have to use $l_2$-norm to get the format of $A'_t$. For the variance
part, to use weighted version of self-normalized concentration (Theorem 5), they use \( V_{t-1}^{-1} V_{t-1} \)-norm to control x term so that \( B_t \) term can be normed by \( V_{t-1}^{-1} \). As an improvement, we can directly use the same \( V_{t-1}^{-1} \)-norm to control both parts, which benefits from our new analysis for the bias part and modified analysis for the variance part.

**Bias Part Analysis.** The first step is to extract the variations of underlying parameters as follows,

\[
A_t \leq L \sum_{s=1}^{t-1} \gamma^{t-s-1} \|X_s\|_{V_{s-1}^{-1}} \|\theta_p - \theta_{p+1}\|_2.
\]

Term \( \sum_{s=1}^{t-1} \gamma^{t-s-1} \|X_s\|_{V_{s-1}^{-1}} \) should be able to further derive an expression about discounted factor \( \gamma \), which can control the variation item. After some derivation, we get

\[
A_t \leq L \sqrt{d} \sum_{s=1}^{t-1} \frac{\gamma^{t-s-1}}{\sqrt{1 - \gamma}} \|\theta_p - \theta_{p+1}\|_2,
\]

where the variation item is only controlled by the discounted factor \( \gamma \) instead of a virtual window size.

**Variance Part Analysis.** Based on the definition of \( V_t \) and \( \tilde{V}_t \), we can find that \( V_t \geq \tilde{V}_t \), so we have \( B_t \leq B'_t \leq \beta_{t-1} \) where \( \beta_{t-1} \) is the confidence radius (5) and the second inequality is by Theorem 5. So we keep the same confidence bound \( \beta_{t-1} \) while only need to compute \( V_{t-1}^{-1} \) instead of \( V_{t-1}^{-1} V_{t-1} V_{t-1}^{-1} \) when doing the arm selection.

Combining the analysis for bias and variance parts, we can finish the proof of Lemma 1. \( \square \)

**Remark 1.** The key step (18) in our analysis framework also resolves the projection issue in GLB. Specifically, after the projection step, the bias-variance decomposition can only be performed in \( V_{t-1}^{-1} \)-norm. To accommodate previous analysis (19), Faury et al. [2021] have to inject a highly complex projection operation in the algorithm, whereas our framework already satisfies this condition owing to the usage of the same \( V_{t-1}^{-1} \)-norm for the bias and variance parts.

### 7 EXPERIMENTS

In this section, we further empirically examine the performance of our proposed algorithms. We present two synthetic experiments on drifting LB and GLB, respectively. For each experiment, we set the dimension of the feature space to \( d = 2 \), the number of rounds to \( T = 6000 \), and the number of arms to \( n = 50 \). The features of each arm are sampled from the normal distribution \( \mathcal{N}(0, 1) \) and subsequently rescaled to satisfy \( L = 1 \). We initialize the time-varying parameter \( \theta_t \) to \([1, 0]\) and rotate it uniformly counterclockwise around the unit circle, completing one full revolution from 0 to \( 2\pi \) over the course of \( T \) rounds and returning to the starting point \([1, 0]\).

**7.1 Linear Bandits**

**Setting.** We consider the linear model \( r_t = X_t^T \theta_t + \eta_t \) where the random noise \( \eta_t \) is drawn from the normal distribution \( \mathcal{N}(0, 1) \) at each time \( t \) independently. We compare the performance of our proposed LB-WeightUCB algorithm to: (a) the static algorithm OFUL [Abbasi-Yadkori et al., 2011]; (b) the restart-based algorithm RestartUCB [Zhao et al., 2020]; (c) the weight-based algorithm D-LinUCB [Russac et al., 2019]; and (d) the adaptive restart algorithm MASTER+OFUL [Wei and Luo, 2021]. Since \( P_T \) is computable, we set the discounted factor \( \gamma = 1 - \max\{1/T, \sqrt{P_T/(dT)}\} \) for LB-WeightUCB and D-LinUCB, and set the window size \( w \) and restarting period \( H \) as \( w = H = d^{1/4} \sqrt{T/(1 + P_T)} \). For MASTER, there is a parameter \( n \) representing the initial value of a multi-scale exploration parameter (see the input of Procedure 1 in [Wei and Luo, 2021]) and the origin MASTER algorithm lets it start from 0 (i.e., \( n = 0, 1, \ldots \)). However, a small initial value of \( n \) will lead to high-frequent restart and thus achieve poor performance. To address this issue, we experiment with a larger initial value of \( n = 13 \), which leads to greatly improved performance in our case.

**Results.** The experimental results are averaged over 20 independent trials. Figure 1a shows the cumulative dynamic regret performance, where the shaded area denotes the variance of the 20 independent trials of experimental results. Figure 1b reports the average time per run, with each run containing 6000 rounds. Our LB-WeightUCB algorithm performs as well as D-LinUCB but significantly more efficient, with over 1.5 times speedup. Figure 1a also shows that when equipped with a fine-tuned \( n \), MASTER+OFUL \((n = 13)\) performs better than RestartUCB, whereas a vanilla MASTER+OFUL \((n = 0)\) performs worse due to overly active restarts at the beginning. However, a larger initial value of \( n \) results in greater time overhead, since at each restart, MASTER+OFUL needs to do Procedure 1 once, resulting in an \( \mathcal{O}(n2^n) \) time complexity. More importantly, neither adaptive restart (MASTER+OFUL) nor periodical restart (RestartUCB) outperforms our weighted strategy in slowly-evolving environments.
of the weight-based algorithms outperform the static algorithms, and our GLB-WeightUCB and SCB-WeightUCB are better than BVD-GLM-UCB. When \( S \) is large (\( S = 5, c_{\mu}^{-1} \approx 152 \)), SCB-WeightUCB significantly outperforms GLB-WeightUCB, demonstrating the importance of considering self-concordant property (recall that LogB is an instance of SCB). In contrast, the performance of BVD-GLM-UCB drops dramatically, as it does not take the \( c_{\mu}^{-1} \) issue into account. Similar to LB, the experimental results of GLB also demonstrate the empirical advantage of the weighted strategy over (adaptive) restart strategy in slowly-evolving environments. Specifically, we observe that GLB-WeightUCB consistently outperforms MASTER+GLM-UCB, and SCB-WeightUCB consistently outperforms MASTER+LogUCB1.

8 CONCLUSION

This paper revisits the weight-based algorithms for three non-stationary parametric bandit models (LB, GLB, SCB). We identify that the inadequacies of the previous work are due to the inadequate analysis of the estimation error. We thus propose a refined analysis framework that enables the usage of the same local norm for both the bias and variance part in estimation error analysis. Our framework ensures more efficient algorithms for all three bandit models and improves the regret bounds for GLB and SCB settings.

The importance of our work lies in the fact that we have now made the weight-based algorithms for non-stationary LB/GLB/SCB as competitive as the restart-based algorithms, in terms of both computational efficiency and regret guarantee. Given that the weighted strategy is particularly appealing in gradually drifting scenarios that are commonly seen in real-world applications, it is essential to further design adaptive weight-based algorithms for non-stationary parametric bandits with optimal dynamic regret without requiring the knowledge of environmental non-stationarity, in the spirit of the currently best-known result achieved by adaptive restart strategy [Wei and Luo, 2021].

In this work, we employ \( P_T = \sum_{t=2}^{T} \| \theta_{t-1} - \theta_t \|_2 \) as a measure to capture the gradually changing environment. However, this metric may not be precise enough in capturing only the gradual changes in the environment, as it can also include other types of variations, such as abrupt changes. This might be able to serve as an explanation why weight-based algorithms do not exhibit a significant theoretical advantage, yet perform remarkably well in experiments on gradually changing environments compared to restart-based algorithms. To overcome this limitation, future research could explore more refined characterizations of gradual changes, drawing inspiration from the ideas behind Sobolev or Holder classes [Baby and Wang, 2019] or other information-theoretic tools [Liu et al., 2022].
Acknowledgements

This research was supported by NSFC (61921006), JiangsuSF (BK20220776) and Collaborative Innovation Center of Novel Software Technology and Industrialization. The authors greatly appreciate helpful comments and suggestions from anonymous reviewers and meta-reviewer. The authors also thank Yu-Hu Yan and Yu-Jie Zhang for many helpful discussions.

References


A Analysis of LB-WeightUCB

In this section, we provide the analysis for LB-WeightUCB algorithm. In Appendix A.1, we review the D-LinUCB algorithm proposed by Russac et al. [2019] and restate their estimation error analysis. In Appendix A.2, we present our own estimation error analysis for the proposed LB-WeightUCB algorithm, which is captured in Lemma 1. Finally, in Appendix A.3, we provide a proof for our dynamic regret bound, as stated in Theorem 1.

A.1 Review Estimation Error Analysis of D-LinUCB Algorithm

In this part, we review the previous estimation error analysis of the D-LinUCB algorithm [Russac et al., 2019] who has the same estimator as ours (3). The first step is to divide the estimation error into the bias and variance parts, where the bias part represents the error caused by parameter drift and the variance part represents the error caused by stochastic noise. Based on the reward model assumption and the estimator (same as eq (1) and eq (3)), the estimation error of D-LinUCB algorithm can be decomposed as

\[
\hat{\theta}_t - \theta_t = \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s - \theta_t
\]

\[
= V_{t-1}^{-1} \sum_{s=1}^{t-1} \gamma^{t-s-1} (X_s^\top \theta_s + \eta_s) X_s - V_{t-1}^{-1} \left( \lambda I_d + \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s X_s^\top \right) \theta_t
\]

\[
= V_{t-1}^{-1} \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s X_s^\top (\theta_s - \theta_t) + V_{t-1}^{-1} \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t + \lambda \theta_t
\]

Afterward, Russac et al. [2019] use different local norms (we will explain the reason of using different local norms later) for the bias and variance parts as follows,

\[
|x^\top (\hat{\theta}_t - \theta_t)| \leq \|x\|_2 \Lambda'_t + \|x\|_{V_{t-1}^{-1} V_{t-1}^{-1} V_{t-1}^{-1} B'_t} \tag{21}
\]

where \( V_t = \lambda I_d + \sum_{s} \gamma^{2(t-s)} X_s X_s^\top \) and

\[
\Lambda'_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s X_s^\top (\theta_s - \theta_t) \right\|_2, \quad B'_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \right\|_{V_{t-1}^{-1}}.
\]

For the bias part, Russac et al. [2019] divide it into two parts on the timeline by introducing a virtual window size \( D \),

\[
\Lambda'_t \leq \left\| \sum_{s=t-D}^{t-1} V_{t-1}^{-1} \gamma^{t-s-1} X_s X_s^\top (\theta_s - \theta_t) \right\|_2 + \left\| \sum_{s=1}^{t-D-1} V_{t-1}^{-1} \gamma^{t-s-1} X_s X_s^\top (\theta_s - \theta_t) \right\|_2
\]

The first term can be considered as a virtual window containing the most recent data obtained after time \( t - D \), and can be directly analyzed by the analysis of SW-UCB [Cheung et al., 2019] since it corresponds to the bias part of the estimation error of window strategy. The second term reflects the influence formed by the outdated data obtained before time \( t - D \). Since \( \gamma^{t-s-1} \) will be very small when \( s \leq t - D - 1 \), this small term is dominated by the first virtual window term which means the bias part is actually controlled by the virtual window size \( D \).

For the variance part, Russac et al. [2019] extend the previous self-normalized concentration [Abbasi-Yadkori et al., 2011, Theorem 1] to the weighted version which is restated in Theorem 5. This concentration requires to use \( V_t \) as the local norm. To this end, Russac et al. [2019] split the variance part as

\[
\left\| x^\top V_{t-1}^{-1} \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \right\| \leq \|x\|_{V_{t-1}^{-1} V_{t-1}^{-1} V_{t-1}^{-1}} \left\| V_{t-1}^{-1} \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \right\|_{V_{t-1}^{-1} V_{t-1}^{-1} V_{t-1}^{-1}}
\]
Then term $\|\sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s\|_{\tilde{V}_{t-1}}$ can be bounded by Theorem 5. Finally, based on this analysis, D-LinUCB needs to use the following action selection criterion which only depends on the variance part since the bias part doesn’t contain $x$,

$$X_t = \arg \max_{x \in \mathcal{X}} \left\{ (x, \hat{\theta}_t) + \beta_{t-1} \|x\|_{\tilde{V}_{t-1}} \right\},$$

where $\beta_{t-1}$ is the upper bound of $B'_t$ which is the same as (5). From this selection criterion, it can be seen that D-LinUCB needs to maintain two covariance matrices, namely, $V_t$ and $\tilde{V}_t$ at round $t$ during the algorithm running.

In the next section, we present our proof for the estimation error upper bound. The difference between our analysis and D-LinUCB’s analysis mainly starts at step (21), which is the key step of the analysis and our new analysis framework allows us to employ the same local norm for both bias and variance parts.

### A.2 Proof of Lemma 1

**Proof.** Using the same derivation in (20), the estimation error of LB-WeightUCB algorithm can also be decomposed as

$$\hat{\theta}_t - \theta_t = V_{t-1}^{-1} \left( \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s X_s^\top (\theta_s - \theta_t) \right) + V_{t-1}^{-1} \left( \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \right).$$

Therefore, by the Cauchy-Schwarz inequality, we know that for any $x \in \mathcal{X}$,

$$\|x^\top (\hat{\theta}_t - \theta_t)\| \leq \|x\|_{V_{t-1}} (A_t + B_t),$$

where

$$A_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s X_s^\top (\theta_s - \theta_t) \right\|_{V_{t-1}^{-1}}, \quad B_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \right\|_{V_{t-1}^{-1}}.$$

The above two terms can be bounded separately, as summarized in the following two lemmas.

**Lemma 4.** For any $t \in [T]$, we have

$$\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s X_s^\top (\theta_s - \theta_t) \right\|_{V_{t-1}^{-1}} \leq L \sqrt{\frac{d}{\delta}} \sum_{p=1}^{t-1} \frac{\gamma^{t-p} - 1}{1 - \gamma} \|\theta_p - \theta_{p+1}\|_2.$$

**Lemma 5.** For any $\delta \in (0, 1)$, with probability at least $1 - \delta$, the following holds for all $t \in [T]$,

$$\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \right\|_{V_{t-1}^{-1}} \leq \sqrt{\lambda S} + R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2 (1 - \gamma^{2t-2})}{\lambda d (1 - \gamma^2)} \right)}.$$

Based on the inequality (22), Lemma 4, Lemma 5, and the boundedness assumption of the feasible set, for any $x \in \mathcal{X}$, $\gamma \in (0, 1)$ and $\delta \in (0, 1)$, with probability at least $1 - \delta$, the following holds for all $t \in [T]$,

$$|x^\top (\hat{\theta}_t - \theta_t)| \leq L^2 \sqrt{\frac{d}{\lambda}} \sum_{p=1}^{t-1} \frac{\gamma^{t-p} - 1}{1 - \gamma} \|\theta_p - \theta_{p+1}\|_2 + \beta_{t-1} \|x\|_{V_{t-1}^{-1}},$$

where $\beta_t \triangleq \sqrt{\lambda S} + R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2 (1 - \gamma^{2t})}{\lambda d (1 - \gamma^2)} \right)}$ is the confidence radius used in LB-WeightUCB. Hence we complete the proof.
Proof of Lemma 4. The first step is to extract the variations of the parameter \( \theta_t \) as follows,

\[
\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s X_s^\top (\theta_s - \theta_t) \right\|_{V_{t-1}} = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} X_s X_s^\top \sum_{p=s}^{t-1} (\theta_p - \theta_{p+1}) \right\|_{V_{t-1}}
\]

\[
= \left\| \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} X_s X_s^\top (\theta_p - \theta_{p+1}) \right\|_{V_{t-1}}
\]

\[
\leq \sum_{p=1}^{t-1} \left\| \sum_{s=1}^{p} \gamma^{t-s-1} X_s \|X_s\|_2 \|\theta_p - \theta_{p+1}\|_2 \right\|_{V_{t-1}}
\]

\[
\leq L \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} \|X_s\|_{V_{t-1}} \|\theta_p - \theta_{p+1}\|_2,
\]

and term \( \sum_{s=1}^{p} \gamma^{t-s-1} \|X_s\|_{V_{t-1}} \) can be able to further derive an expression about discounted factor \( \gamma \) as follows,

\[
\sum_{s=1}^{p} \gamma^{t-s-1} \|X_s\|_{V_{t-1}} \leq \gamma^{\frac{p-1}{2}} \sqrt{\sum_{s=1}^{p} \gamma^{t-s} \sum_{s=1}^{p} \gamma^{t-s-1} \|X_s\|_2^2} \leq \sqrt{d} \gamma^{\frac{p-1}{2}} \gamma \sqrt{\frac{1}{1-\gamma}}.
\]

(23)

In above, we use the fact that for any \( x \), \( \|x\|_{V_{t-1}} \leq \|x\|_2 / \|V_{t-1}\| \) since \( V_{t-1} \succeq \lambda I_d \). The second last step holds by the Cauchy-Schwarz inequality. Besides, the last step follows the fact,

\[
\forall p \in [t-1], \quad \sum_{s=1}^{p} \gamma^{t-s-1} \|X_s\|_{V_{t-1}}^2 \leq d,
\]

which can be proven by the following argument.

\[
\sum_{s=1}^{p} \gamma^{t-s-1} \|X_s\|_{V_{t-1}}^2 = \sum_{s=1}^{p} \gamma^{t-s-1} \text{Tr}(X_s^\top V_{t-1}^{-1} X_s) = \text{Tr} \left( V_{t-1}^{-1} \sum_{s=1}^{p} \gamma^{t-s-1} X_s X_s^\top \right)
\]

\[
\leq \text{Tr} \left( V_{t-1}^{-1} \sum_{s=1}^{p} \gamma^{t-s-1} X_s X_s^\top \right) + \text{Tr} \left( V_{t-1}^{-1} \sum_{s=p+1}^{t-1} \gamma^{t-s-1} X_s X_s^\top \right)
\]

\[
= \text{Tr}(I_d) = d.
\]

Hence, we complete the proof. 

Proof of Lemma 5. Let \( \bar{V}_t \triangleq \lambda I_d + \sum_{s=1}^{t} \gamma^{2(t-s)} X_s X_s^\top \),

\[
\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda \theta_t \right\|_{V_{t-1}^{-1}} \leq \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{V_{t-1}^{-1}} + \|\lambda \theta_t\|_{V_{t-1}} \leq \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{V_{t-1}^{-1}} + \sqrt{\lambda} S.
\]

Recall that \( V_t = \lambda I_d + \sum_{s=1}^{t} \gamma^{t-s} X_s X_s^\top \), so the last inequality comes from

\[
V_t = \lambda I_d + \sum_{s=1}^{t} \gamma^{t-s} X_s X_s^\top \succeq \lambda I_d + \sum_{s=1}^{t} \gamma^{2(t-s)} X_s X_s^\top = \bar{V}_t.
\]

We emphasize that the \( \bar{V}_t \) is introduced into analysis only, which is actually not required in our algorithmic implementation. From the weighted version maximal deviation inequality [Russac et al., 2019, Theorem 1], restated in Theorem 5, we can get the bound for the first term \( \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{V_{t-1}^{-1}} \) as below by just let \( w_s = \gamma^{t-s-1}, \mu_t = \lambda \),

\[
\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{V_{t-1}^{-1}} \leq R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2 \sum_{s=1}^{t-1} \gamma^{2(t-s-1)}}{\lambda d} \right)} \leq R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2 (1-\gamma^{2t-2})}{\lambda d (1-\gamma^2)} \right)},
\]

which completes the proof. 

A.3 Proof of Theorem 1

**Proof.** Let \( X_t^* = \arg \max_{x \in X} x^T \theta_t \). Due to Lemma 1 and the fact that \( X_t^*, X_t \in \mathcal{X} \), each of the following holds with probability at least \( 1 - \delta \),

\[
\forall t \in [T], X_t^* \theta_t \leq X_t^* \tilde{\theta}_t + L^2 \frac{d}{\lambda} \sum_{p=1}^{t-1} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2 + \beta_{t-1} \| X_t^* \|_{V_{t-1}^{-1}}.
\]

By the union bound, the following holds with probability at least \( 1 - 2\delta \),

\[
\forall t \in [T], X_t^* \theta_t - X_t^* \tilde{\theta}_t \leq X_t^* \tilde{\theta}_t - X_t^* \tilde{\theta}_t + L^2 \frac{d}{\lambda} \sum_{p=1}^{t-1} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2 + \beta_{t-1} \| X_t^* \|_{V_{t-1}^{-1}} + \| X_t \|_{V_{t-1}^{-1}}
\]

\[
\leq L^2 \frac{d}{\lambda} \sum_{p=1}^{t-1} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2 + 2\beta_{t-1} \| X_t \|_{V_{t-1}^{-1}},
\]

where the last step comes from the arm selection criterion (6) such that

\[
X_t^* \tilde{\theta}_t + \beta_{t-1} \| X_t^* \|_{V_{t-1}^{-1}} \leq X_t^* \tilde{\theta}_t + \beta_{t-1} \| X_t \|_{V_{t-1}^{-1}}.
\]

Hence, the following dynamic regret bound holds with probability at least \( 1 - 2\delta \) and can be divided into two parts,

\[
R_T = \sum_{t=1}^{T} (X_t^* \theta_t - X_t^* \tilde{\theta}_t) \leq L^2 \frac{d}{\lambda} \sum_{t=1}^{T} \sum_{p=1}^{t-1} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2 + 2\beta_{T} \sum_{t=1}^{T} \| X_t \|_{V_{t-1}^{-1}},
\]

where \( \beta_T = \sqrt{s} + R \sqrt{2} \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2 (1-\gamma^T)}{\lambda \delta} \right) \) is the confidence radius.

Now we derive the upper bound for the bias and variance parts separately.

**Bias Part.** For the bias part, we need to extract path length \( P_T \) and show the control of the discounted factor \( \gamma \) on \( P_T \).

\[
2L^2 \frac{d}{\lambda} \sum_{t=1}^{T} \sum_{p=1}^{t-1} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2 = 2L^2 \frac{d}{\lambda} \sum_{p=1}^{T-1} \sum_{t=p+1}^{T} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2
\]

\[
= 2L^2 \frac{d}{\lambda} \sum_{p=1}^{T-1} \sum_{t=p+1}^{T} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2
\]

\[
\leq 2L^2 \frac{d}{\lambda} \sum_{p=1}^{T-1} \sum_{t=p+1}^{T} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2
\]

\[
\leq 4L^2 \frac{d}{\lambda} \sum_{p=1}^{T-1} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2
\]

\[
= 4L^2 \frac{d}{\lambda} \sum_{p=1}^{T-1} \frac{1}{(1-\gamma)^{p/2}} P_T.
\]

So for the bias part, we have

\[
2L^2 \frac{d}{\lambda} \sum_{t=1}^{T} \sum_{p=1}^{t-1} \gamma^{p-1} \left( \frac{\gamma^{p-1}}{1-\gamma} \right) \| \theta_p - \theta_{p+1} \|_2 \leq 4L^2 \frac{d}{\lambda} \frac{1}{(1-\gamma)^{T/2}} P_T.
\]  

(25)

**Variance Part.** First, use the Cauchy-Schwarz inequality, we know that

\[
2\beta_T \sum_{t=1}^{T} \| X_t \|_{V_{t-1}^{-1}} \leq 2\beta_T \sqrt{\sum_{t=1}^{T} \| X_t \|_{V_{t-1}^{-1}}}^2.
\]
Then by Lemma 11 (potential lemma), we have the following upper bound:

\[ 2\beta_T \sum_{t=1}^{T} \|X_t\|_{V_{t-1}} \leq 2\beta_T \sqrt{2 \max\{1, L^2 / \lambda\} d T} \sqrt{T \log \frac{1}{\gamma} + \log \left(1 + \frac{L^2}{\lambda d (1 - \gamma)}\right)}. \]  

(26)

Combining the upper bounds of the bias and variance parts and with confidence level \(\delta = 1/(2T)\), by union bound we have the following dynamic regret bound with probability at least \(1 - 1/T\):

\[ R_T \leq 4L^2 \sqrt{\lambda S + R_T^2} \left(2 \log T + 2 \log 2 + d \log \left(1 + \frac{L^2(1 - \gamma^2 T)}{\lambda d (1 - \gamma)}\right)\right). \]

Since that there has a term \(T \sqrt{\log(1/\gamma)}\) in the regret bound, we cannot let \(\gamma\) close to 0, so we set \(\gamma \geq 1/T\) and have \(\log(1/\gamma) \leq C/(1 - \gamma)\), where \(C = \log T/(1 - 1/T)\). Then, ignoring logarithmic factors in time horizon \(T\), and let \(\lambda = d\), we finally obtain

\[ R_T \leq \tilde{O} \left(\frac{1}{(1 - \gamma)^{3/2}} P_T + d(1 - \gamma) ^{1/2} T\right). \]

When \(P_T < d/T\) (which corresponds a small amount of non-stationarity), we simply set \(\gamma = 1 - 1/T\) and achieve an \(\tilde{O}(d\sqrt{T})\) regret bound. Besides, when coming to the non-degenerated case \((P_T \geq d/T)\), We set the discounted factor optimally as \(1 - \gamma = \sqrt{P_T/(dT)}\) and attain an \(\tilde{O}(d^{\gamma/4} P_T^{3/4} T^{3/4})\) dynamic regret bound, which completes the proof. \(\square\)

**B Analysis of GLB-WeightUCB**

In this section, we provide the analysis for GLB-WeightUCB algorithm. In Appendix B.1, we review the projection issue of GLB and restate the BVD-GLM-UCB algorithm of Faury et al. [2021]. In Appendix B.2, we present the proof of the estimation error upper bound of our GLB-WeightUCB algorithm (namely, Lemma 2). Finally, in Appendix B.3, we provide the proof of dynamic regret upper bound as stated in Theorem 2.

**B.1 Review Projection Step of BVD-GLM-UCB Algorithm**

As mentioned in Section 4.2, the main difficulty of GLB is that the result of MLE or QMLE estimator \(\hat{\theta}_t\) may not belong to the feasible set \(\Theta\) and \(c_\mu\) is defined over the parameter \(\theta \in \Theta\). Under stationary environments, Filippi et al. [2010] overcame this difficulty by introducing a projection step as

\[ \tilde{\theta}_t = \arg \min_{\theta \in \Theta} \|g_t(\hat{\theta}_t) - g_t(\theta)\|_{V_{t-1}}, \]  

(27)

where \(V_t = \lambda I_d + \sum_{s=1}^{t-1} X_s X_s^\top\) and \(g_t(\theta) = \lambda c_\mu \theta + \sum_{s=1}^{t-1} \mu(X_s^\top \theta) X_s\) are the static version (by setting \(\gamma = 1\)). Based on the QMLE, we know that

\[ g_t(\hat{\theta}_t) = \lambda c_\mu \hat{\theta}_t + \sum_{s=1}^{t-1} \mu(X_s^\top \hat{\theta}_t) X_s = \sum_{s=1}^{t-1} r_s X_s, \]  

(28)

and then by the mean value theorem, we know that

\[ g_t(\theta_1) - g_t(\theta_2) = G_t(\theta_1, \theta_2)(\theta_1 - \theta_2), \]  

(29)

where \(G_t(\theta_1, \theta_2) \triangleq \int_0^1 \nabla g_t(s \theta_2 + (1 - s) \theta_1) ds \in \mathbb{R}^d \times d\). Notice that for any \(\theta \in \Theta\), the gradient of \(g_t\) satisfies

\[ \nabla g_t(\theta) = \lambda c_\mu I_d + \sum_{s=1}^{t-1} \mu'(X_s^\top \theta) X_s X_s^\top \geq c_\mu V_{t-1}, \]

which clearly implies \(\forall \theta_1, \theta_2 \in \Theta, G_t(\theta_1, \theta_2) \succeq c_\mu V_{t-1}\).
By this projection step, Filippi et al. [2010] can analyze the estimation error like,
\[
|\mu(x^\top \hat{\theta}_t) - \mu(x^\top \theta_t)| \leq k_\mu |x^\top (\hat{\theta}_t - \theta_t)|
\]
\[
= k_\mu |x^\top G_t^{-1}(\hat{\theta}_t, \hat{\theta}_t)(g_t(\hat{\theta}_t) - g_t(\theta_t))|
\]
\[
\leq k_\mu \|x\|_{G_t^{-1}(\hat{\theta}_t, \hat{\theta}_t)} \|g_t(\hat{\theta}_t) - g_t(\theta_t)\|_{V_t^{-1}}
\]
\[
\leq \frac{k_\mu}{c_\mu} \|x\|_{V_t^{-1}} \|g_t(\hat{\theta}_t) - g_t(\theta_t)\|_{V_t^{-1}}
\]
\[
\leq \frac{2k_\mu}{c_\mu} \|x\|_{V_t^{-1}} \|g_t(\hat{\theta}_t) - g_t(\theta_t)\|_{V_t^{-1}},
\]
where the last step comes from the projection step. After doing the projection step, term \(g_t(\hat{\theta}_t) - g_t(\theta_t)\) is the estimation error of the MLE without projection. Notice that in piecewise-stationary case, Russac et al. [2021] can also use this projection step. Faury et al. [2021] believe that these two previous works could use this projection operation mainly due to their stationary or piecewise-stationary setting. They mention that for the drifting case, the estimation error is always divided into the bias (tracking error) and variance (learning error) part, and this simple projection operation ignores the bias part which needs to be generalized to adapt to the two sources of deviation. In the analysis, the problem is that after the projection step estimation error term \(g_t(\hat{\theta}_t) - g_t(\theta_t)\) need to be separate into the bias part and variance parts, and Faury et al. [2021] need to use \(l_2\)-norm for bias part and \(V_t^{-1}\) for variance part. But the whole estimation error is already normed by \(V_t^{-1}\), which means they cannot use the previous analysis of the window strategy for the bias part.

To this end, Faury et al. [2021] propose the BVD-GLM-UCB algorithm for drifting generalized linear bandits, as restated in Algorithm 3, where a new projection step is devised to solve this problem. Specifically, at each round \(t\), the first step is to construct the confidence set \(E_t^\delta(\theta)\) which represents the influence of the stochastic noise.
\[
E_t^\delta(\theta) := \left\{ \theta' \in \mathbb{R}^d \mid \|g_t(\theta') - g_t(\theta)\|_{V_t^{-1}} \leq \beta_t(\delta) \right\}.
\] (30)

The second step is to find a confidence set \(E_t^\delta(\theta_t^p)\) that intersects with the feasible set, and the gap between \(\theta_t^p\) and \(\hat{\theta}_t\) represents the influence of parameter drift.
\[
\theta_t^p \in \arg \min_{\theta_t \in \mathbb{R}^d} \left\{ \|g_t(\theta) - g_t(\hat{\theta}_t)\|_{V_t^{-2}} \right\} \quad \text{s.t. } \Theta \cap E_t^\delta(\theta) \neq \emptyset.
\] (31)

After obtaining the solution \(\theta_t^p\) via computing the optimization problem (31), the third step is to select \(\hat{\theta}_t\) from \(\Theta \cap E_t^\delta(\theta_t^p)\). Based on this projection step, Faury et al. [2021] can separate the bias and variance parts before projection as follows,
\[
|\mu(x^\top \hat{\theta}_t) - \mu(x^\top \theta_t)| \leq k_\mu |x^\top (\hat{\theta}_t - \theta_t)|
\]
\[
= k_\mu |x^\top G_t^{-1}(\hat{\theta}_t, \hat{\theta}_t)(g_t(\hat{\theta}_t) - g_t(\theta_t))|
\] (32)
Their bias-variance decomposition motivates the choice of different local norms for bounding bias and variance parts in their algorithm and analysis. Notably, due to the complications of the projection step (see (30) and (31)), the overall algorithm is fairly complicated and less attractive for practical implementations, and moreover, it needs to maintain two covariance matrices $V_t$ and $V_l$ (due to the constructed confidence region (30)) at each round $t$ during the algorithm running. In the next section, we will show that the simple projection used in the stationary GLB (27) can be sufficient for coping with the drifting GLB via our refined analysis framework.

### B.2 Proof of Lemma 2

**Proof.** Base on the estimator equation (8), we know that

$$g_t(\hat{\theta}_t) = \lambda c_p\hat{\theta}_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \hat{\theta}_t)X_s = \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s,$$

and then by the mean value theorem, we know that

$$g_t(\theta_1) - g_t(\theta_2) = G_t(\theta_1, \theta_2)(\theta_1 - \theta_2),$$

where $G_t(\theta_1, \theta_2) \triangleq \int_0^1 \nabla g_t(s\theta_2 + (1-s)\theta_1)ds \in \mathbb{R}^{d \times d}$. Notice that for any $\theta \in \Theta$, the gradient of $g_t$ is

$$\nabla g_t(\theta) = \lambda c_p I_d + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu'(X_s^\top \theta)X_sX_s^\top \geq c_v V_{t-1},$$

which clearly implies $\forall \theta_1, \theta_2 \in \Theta, G_t(\theta_1, \theta_2) \succeq c_v V_{t-1}$.

By Assumption 2, the mean value theorem (37) on $g_t$ and the projection (9), we have

$$|\mu(X^\top \hat{\theta}_t) - \mu(X^\top \theta_t)| \leq k_\mu||x^\top (\hat{\theta}_t - \theta_t)||$$

then based on the model assumption, the function $g_t(\theta)$ and $g_t(\hat{\theta}_t)$ (36), we have,

$$g_t(\theta_t) - g_t(\hat{\theta}_t) = \lambda c_p \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta_t)X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s$$

and then by the Cauchy-Schwarz inequality, we know that for any $x \in \mathcal{X}$,

$$|\mu(x^\top \hat{\theta}_t) - \mu(x^\top \theta_t)| \leq \frac{2k_\mu}{c_\mu} ||x||_{V_{t-1}^{-1}} (C_t + D_t),$$
where

\[ C_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^T \theta_t) - \mu(X_s^T \theta_s)) X_s \right\|_{V_{t-1}^{-1}}, \quad D_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_{\mu} \theta_t \right\|_{V_{t-1}^{-1}}. \]

This two terms can be bounded separately, as summarized in the following lemmas.

**Lemma 6.** For any \( t \in [T] \), we have

\[ \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^T \theta_t) - \mu(X_s^T \theta_s)) X_s \right\|_{V_{t-1}^{-1}} \leq L k_{\mu} \sqrt{d} \sum_{p=1}^{t-1} \gamma^{t-1} \frac{2 \log 1}{\delta} \left\| \theta_p - \theta_{p+1} \right\|_2, \]  \( (41) \)

**Lemma 7.** For any \( \delta \in (0, 1) \), with probability at least \( 1 - \delta \), the following holds for all \( t \in [T] \),

\[ \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_{\mu} \theta_t \right\|_{V_{t-1}^{-1}} \leq \sqrt{\lambda c_{\mu}} S + R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2 (1 - \gamma^{2^t - 2})}{\lambda d (1 - \gamma^2)} \right)}. \]  \( (42) \)

Based on the inequality (40), Lemma 6, Lemma 7, and the boundedness assumption of the feasible set, we have for any \( x \in \mathcal{X}, \gamma \in (0, 1), \delta \in (0, 1) \), with probability at least \( 1 - \delta \), the following holds for all \( t \in [T] \),

\[ \left\| \mu(x^T \theta_t) - \mu(x^T \theta_t) \right\| \leq 2k_{\mu} c_{\mu} \left\| x \right\|_{V_{t-1}^{-1}} \left( L k_{\mu} \sqrt{d} \sum_{p=1}^{t-1} \gamma^{t-1} \frac{2 \log 1}{\delta} \left\| \theta_p - \theta_{p+1} \right\|_2 + \tilde{\beta}_{t-1} \right) \]

\[ \leq 2k_{\mu} c_{\mu} \left( L^2 k_{\mu} \sqrt{d} \sum_{p=1}^{t-1} \gamma^{t-1} \frac{2 \log 1}{\delta} \left\| \theta_p - \theta_{p+1} \right\|_2 + \tilde{\beta}_{t-1} \left\| x \right\|_{V_{t-1}^{-1}} \right), \]

where \( \tilde{\beta}_{t-1} \equiv \sqrt{\lambda c_{\mu}} S + R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2 (1 - \gamma^{2^t - 2})}{\lambda d (1 - \gamma^2)} \right)} \) is the confidence radius used in GLB-WeightUCB. Hence we complete the proof.  \( \square \)

**Proof of Lemma 6.** Here we need to extract the variations of the time-varying parameter \( \theta_t \)

\[ \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^T \theta_t) - \mu(X_s^T \theta_s)) X_s \right\|_{V_{t-1}^{-1}} \leq \left\| \sum_{s=1}^{t-1} \sum_{p=s}^{t-1} (\mu(X_s^T \theta_p) - \mu(X_s^T \theta_{p+1})) X_s \right\|_{V_{t-1}^{-1}} \]

\[ = \left\| \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} \alpha(X_s, \theta_p, \theta_{p+1}) (X_s^T \theta_p - X_s^T \theta_{p+1}) X_s \right\|_{V_{t-1}^{-1}} \]

\[ = \left\| \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} \alpha(X_s, \theta_p, \theta_{p+1}) X_s X_s^T (\theta_p - \theta_{p+1}) \right\|_{V_{t-1}^{-1}} \]

\[ \leq \left\| \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} \alpha(X_s, \theta_p, \theta_{p+1}) X_s \left\|_{V_{t-1}^{-1}} \right\| \| \theta_p - \theta_{p+1} \|_2 \right\|_{V_{t-1}^{-1}} \]

\[ \leq L \left\| \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} \alpha(X_s, \theta_p, \theta_{p+1}) X_s \left\|_{V_{t-1}^{-1}} \right\| \| \theta_p - \theta_{p+1} \|_2 \right\| \]

\[ \leq L k_{\mu} \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} \| X_s \|_{V_{t-1}^{-1}} \| \theta_p - \theta_{p+1} \|_2. \]

where the forth equation is due to the mean value theorem where \( \alpha(x, \theta, \theta_2) = \int_{\theta_1}^{\theta_2} \mu'(v x^T \theta_2 + (1 - v)x^T \theta_1) dv: \)

\[ \mu(X_s^T \theta_p) - \mu(X_s^T \theta_{p+1}) = \alpha(X_s, \theta_p, \theta_{p+1})(X_s^T \theta_p - X_s^T \theta_{p+1}). \]

Next, the derivation for the bound of term \( \sum_{p=1}^{t-1} \gamma^{t-s-1} \| X_s \|_{V_{t-1}^{-1}} \) is the same as the inequality (23) in A.2, hence we complete the proof.  \( \square \)
Revisiting Weighted Strategy for Non-stationary Parametric Bandits

Proof of Lemma 7. Same as the linear case, we need to use $\hat{V}_t = \lambda I_d + \sum_{s=1}^{t} \gamma^{2(t-s)}X_sX_s^\top$.

$$D_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_\mu \theta_t \right\|_{V_{t-1}^{-1}}$$

$$\leq \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{V_{t-1}^{-1}} + \left\| \lambda c_\mu \theta_t \right\|_{V_{t-1}^{-1}}$$

$$\leq \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{V_{t-1}^{-1}} + \sqrt{\lambda c_\mu} S$$

$$\leq R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2(1 - \gamma^{2t-2})}{\lambda d(1 - \gamma^2)} \right) + \sqrt{\lambda c_\mu} S}.$$  

Again, we emphasize that the $\hat{V}_t$ is introduced into analysis only. The proof here is the same as the proof of Lemma 5 in A.2, the only difference is an extra $c_\mu$ in the second term.

\[\square\]

B.3 Proof of Theorem 2

Proof. Let $X^*_t \triangleq \arg \max_{x \in X} \mu(x^\top \theta_t)$. Due to Lemma 2 and the fact that $X^*_t, X_t \in X$, each of the following holds with probability at least $1 - \delta$,

$$\forall t \in [T], \mu(X^*_t^\top \theta_t) \leq \mu(X_t^*^\top \bar{\theta}_t) + \frac{2k_\mu}{c_\mu} \left( L^2 k_{\mu} \sqrt{\frac{d}{\lambda} \sum_{p=1}^{t-1} \gamma^{t-1-p} \sqrt{\gamma^{-p-1} \frac{1}{1 - \gamma}} \left\| \theta_p - \theta_{p+1} \right\|_2 + \bar{\beta}_{t-1} \left\| X_t^* \right\|_{V_{t-1}^{-1}}} \right).$$

$$\forall t \in [T], \mu(X_t^\top \theta_t) \geq \mu(X_t^*^\top \bar{\theta}_t) - \frac{2k_\mu}{c_\mu} \left( L^2 k_{\mu} \sqrt{\frac{d}{\lambda} \sum_{p=1}^{t-1} \gamma^{t-1-p} \sqrt{\gamma^{-p-1} \frac{1}{1 - \gamma}} \left\| \theta_p - \theta_{p+1} \right\|_2 + \bar{\beta}_{t-1} \left\| X_t \right\|_{V_{t-1}^{-1}}} \right).$$

By the union bound, the following holds with probability at least $1 - 2\delta$: $\forall t \in [T]$

$$\mu(X_t^*^\top \theta_t) - \mu(X_t^\top \theta_t)$$

$$\leq \mu(X_t^*^\top \bar{\theta}_t) - \mu(X_t^\top \theta_t) + 4L^2 k_{\mu}^2 \frac{d}{\lambda} \sum_{p=1}^{t-1} \gamma^{t-1-p} \sqrt{\gamma^{-p-1} \frac{1}{1 - \gamma}} \left\| \theta_p - \theta_{p+1} \right\|_2 + \frac{2k_\mu}{c_\mu} \left( \bar{\beta}_{t-1} \left\| X_t^* \right\|_{V_{t-1}^{-1}} + \bar{\beta}_{t-1} \left\| X_t \right\|_{V_{t-1}^{-1}} \right)$$

$$\leq 4L^2 k_{\mu}^2 \frac{d}{\lambda} \sum_{p=1}^{t-1} \gamma^{t-1-p} \sqrt{\gamma^{-p-1} \frac{1}{1 - \gamma}} \left\| \theta_p - \theta_{p+1} \right\|_2 + \frac{4k_\mu}{c_\mu} \bar{\beta}_{t-1} \left\| X_t \right\|_{V_{t-1}^{-1}},$$

where the last step comes from the arm selection criterion (12) such that

$$\mu(X_t^*^\top \bar{\theta}_t) + \frac{2k_\mu}{c_\mu} \bar{\beta}_{t-1} \left\| X_t^* \right\|_{V_{t-1}^{-1}} \leq \mu(X_t^\top \theta_t) + \frac{2k_\mu}{c_\mu} \bar{\beta}_{t-1} \left\| X_t \right\|_{V_{t-1}^{-1}}.$$ 

Hence the following dynamic regret bound holds with probability at least $1 - 2\delta$ and can be divided into two parts,

$$R_T = \sum_{t=1}^{T} \max_{x \in X} \mu(x^\top \theta_t) - \mu(X_t^\top \theta_t)$$

$$\leq 4L^2 k_{\mu}^2 \frac{d}{\lambda} \sum_{t=1}^{T} \sum_{p=1}^{t-1} \gamma^{t-1-p} \sqrt{\gamma^{-p-1} \frac{1}{1 - \gamma}} \left\| \theta_p - \theta_{p+1} \right\|_2 + \frac{4k_\mu}{c_\mu} \bar{\beta}_{T-1} \sum_{t=1}^{T} \left\| X_t \right\|_{V_{t-1}^{-1}}.$$  \[\text{(43)}\]

where $\bar{\beta}_t = \sqrt{\lambda c_\mu} S + R \sqrt{2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2(1 - \gamma^{2t-2})}{\lambda d(1 - \gamma^2)} \right) \left( \frac{L^2(1 - \gamma^{2t-2})}{\lambda d(1 - \gamma^2)} \right)}$ is the confidence radius.

Now we derive the upper bound for these two parts separately.
Algorithm 4 SCB-WeightUCB

**Input:** time horizon $T$, discounted factor $\gamma$, confidence $\delta$, regularizer $\lambda$, inverse link function $\mu$, parameters $S$, $L$ and $m$

1: Set $V_0 = \lambda I_d$, $\hat{\theta}_1 = 0$ and compute $k_\mu$ and $c_\mu$
2: for $t = 1, 2, ..., T$ do
3: if $\|\hat{\theta}_t\|_2 \leq S$ then
4: let $\theta_t = \hat{\theta}_t$
5: else
6: Do the projection and get $\tilde{\theta}_t$ by (14)
7: end if
8: Compute $\tilde{\beta}_{t-1}$ by (16)
9: Select $X_t$ by (17)
10: Receive the reward $r_t$
11: Update $V_t = \gamma V_{t-1} + X_tX_t^T + (1 - \gamma)\lambda I_d$
12: Compute $\hat{\theta}_{t+1}$ according to (8)
3: end for

**Bias Part.** Similar to the proof of inequality (25), we have

$$\frac{4T^2k_\mu^2}{c_\mu} \sqrt{\frac{d}{\lambda}} \sum_{t=1}^{T} \sum_{p=1}^{t-1} \gamma^p \left( \frac{\gamma - p - 1}{1 - \gamma} \right) \|\theta_p - \theta_{p+1}\|_2 \leq \frac{8L^2k_\mu^2}{c_\mu} \sqrt{\frac{d}{\lambda}} \frac{1}{(1 - \gamma)^{3/2}} P_T.$$  

**Variance Part.** Similar to the proof of inequality (26), we have

$$\frac{4k_\mu}{c_\mu} \tilde{\beta}_T \sqrt{T} \sum_{t=1}^{T} \|X_t\|_2^2 \leq \frac{4k_\mu}{c_\mu} \tilde{\beta}_T \sqrt{2 \max \{1, L^2/\lambda\} dT} \left( T \log \frac{1}{\gamma} + \log \left( 1 + \frac{L^2}{\lambda d(1 - \gamma)} \right) \right).$$

Combine the upper bound for the bias and variance parts, and let $\delta = 1/(2T^2)$, we have the following regret bound with probability at least $1 - 1/T$,

$$R_T \leq \frac{8L^2k_\mu^2}{c_\mu} \sqrt{\frac{d}{\lambda}} \frac{1}{(1 - \gamma)^{3/2}} P_T + \frac{4k_\mu}{c_\mu} \tilde{\beta}_T \sqrt{2 \max \{1, L^2/\lambda\} dT} \left( T \log \frac{1}{\gamma} + \log \left( 1 + \frac{L^2}{\lambda d(1 - \gamma)} \right) \right).$$

where $\tilde{\beta}_t = \sqrt{\lambda c_\mu S + R} \sqrt{4 \log T + 2 \log 2 + 2 \log \left( 1 + \frac{L^2(1 - \gamma^2)}{\lambda d(1 - \gamma)} \right)}$. We set $\gamma \geq 1/T$ and $\lambda = d/c_\mu^2$, and obtain that,

$$R_T \leq \tilde{O} \left( k_\mu^2 \left( \frac{1}{(1 - \gamma)^{3/2}} P_T + \frac{k_\mu}{c_\mu} d (1 - \gamma)^{1/2} T \right) \right).$$

When $P_T < d/(k_\mu c_\mu T)$, we set $\gamma = 1 - 1/T$ and achieve an $\tilde{O}(k_\mu c_\mu^{-1} d \sqrt{T})$ regret bound. When $P_T \geq d/(k_\mu c_\mu T)$, we set $\gamma$ optimally as $1 - \gamma = \sqrt{\frac{E_\mu c_\mu P_T / (dT)}{d}}$ and attain an $\tilde{O}(k_\mu^2 c_\mu^{-3/4} d^{3/4} P_T^{1/4} T^{3/4})$ regret bound. Notice that, if $k_\mu < 1$, we just let $1 - \gamma = \sqrt{c_\mu P_T / (dT)}$ and the regret bound becomes $\tilde{O}(k_\mu^2 c_\mu^{-3/4} d^{3/4} P_T^{1/4} T^{3/4}).$  

**C Analysis of SCB-WeightUCB**

In this section, we first present SCB-WeightUCB algorithm in Algorithm 4. Then, in Appendix C.1 we present the proof of the estimation error upper bound of our SCB-WeightUCB algorithm (Lemma 3). Finally, in Appendix C.2, we provide the proof of dynamic regret upper bound (Theorem 3).

**C.1 Proof of Lemma 3**

**Proof.** Base on the estimator equation (8), we know that

$$g_t(\hat{\theta}_t) = \lambda c_\mu \hat{\theta}_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^T \hat{\theta}_t) X_s = \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s, \quad (44)$$
and then by the mean value theorem, we know that
\[ g_t(\theta_1) - g_t(\theta_2) = G_t(\theta_1, \theta_2)(\theta_1 - \theta_2), \]
where \( G_t(\theta_1, \theta_2) \triangleq \int_0^1 \nabla g_t(s\theta_2 + (1-s)\theta_1)ds \in \mathbb{R}^{d \times d} \). Notice that for any \( \theta \in \Theta \), the gradient of \( g_t \) is
\[ \nabla g_t(\theta) = \lambda c_{\mu} I_d + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu'(X_s^\top \theta)X_s^\top X_s \geq c_{\mu} V_{t-1}, \]
which clearly implies \( \forall \theta_1, \theta_2 \in \Theta, G_t(\theta_1, \theta_2) \succeq c_{\mu} V_{t-1} \) and \( \forall \theta, H_t(\theta) \succeq c_{\mu} V_{t-1} \), where \( H_t(\theta) \) is defined as
\[ H_t(\theta) \triangleq \lambda c_{\mu} I_d + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu'(X_s^\top \theta)X_s^\top X_s. \]

By Assumption 2, the mean value theorem (37) on \( g_t \), the projection (14) and Lemma 14, we have
\[ |\mu(X^\top \tilde{\theta}_t) - \mu(X^\top \theta_t)| \leq k_{\mu} |x^\top (\tilde{\theta}_t - \theta_t)| \]
\[ = k_{\mu}, \|x\|G_{t}^{-1}(\theta_t, \tilde{\theta}_t) (g_t(\tilde{\theta}_t) - g_t(\theta_t)) \]
\[ \leq k_{\mu} \|x\|G_{t}^{-1}(\theta_t, \tilde{\theta}_t) \left(|g_t(\tilde{\theta}_t) - g_t(\theta_t)| + \|g_t(\tilde{\theta}_t) - g_t(\theta_t)\|_{H_{t}^{-1}(\tilde{\theta}_t)}\right) \]
\[ \leq \sqrt{1 + 2SK_{\mu} \|x\|G_{t}^{-1}(\theta_t, \tilde{\theta}_t) \left(|g_t(\tilde{\theta}_t) - g_t(\theta_t)| + \|g_t(\tilde{\theta}_t) - g_t(\theta_t)\|_{H_{t}^{-1}(\tilde{\theta}_t)}\right) \]
\[ \leq 2\sqrt{1 + 2SK_{\mu} \|x\|G_{t}^{-1}(\theta_t, \tilde{\theta}_t) \left(|g_t(\tilde{\theta}_t) - g_t(\theta_t)| + \|g_t(\tilde{\theta}_t) - g_t(\theta_t)\|_{H_{t}^{-1}(\tilde{\theta}_t)}\right) \]
then based on the model assumption (13), the function \( g_t(\theta) \) (10) and the \( g_t(\tilde{\theta}_t) \) (44), we have,
\[ g_t(\theta_t) - g_t(\tilde{\theta}_t) = \lambda c_{\mu} \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta_t)X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s \]
\[ = \lambda c_{\mu} \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta_t)X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) + \eta_s) X_s \]
\[ = \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s + \lambda c_{\mu} \theta_t - \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s, \]
then, by Cauchy-Schwarz inequality, we have
\[ \left|\mu(X^\top \tilde{\theta}_t) - \mu(X^\top \theta_t)\right| \leq 2\sqrt{1 + 2SK_{\mu} \|x\|G_{t}^{-1}(\theta_t, \tilde{\theta}_t) \left(|g_t(\tilde{\theta}_t) - g_t(\theta_t)| + \|g_t(\tilde{\theta}_t) - g_t(\theta_t)\|_{H_{t}^{-1}(\tilde{\theta}_t)}\right) \]

where
\[ E_t = \left\|\sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{H_{t}^{-1}(\tilde{\theta}_t)}, \]
\[ F_t = \left\|\sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_{\mu} \theta_t \right\|_{H_{t}^{-1}(\theta_t)}. \]
This two terms can be bounded separately.

**Lemma 8.** For any \( t \in [T] \), we have
\[ \left\|\sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{H_{t}^{-1}(\theta_t)} \leq L k_{\mu} \sqrt{d} \sum_{p=1}^{t-1} \gamma^{t-1} \sqrt{\gamma^{-p-1} \frac{1}{1 - \gamma}} \|\theta_p - \theta_{p+1}\|. \]
Lemma 9. For any $\delta \in (0, 1)$, with probability at least $1 - \delta$, we have for all $t \in [T]$, 
\[
\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_{\mu} \theta_t \right\|_{H_{t^{-1}}(\theta_t)} \leq \frac{\sqrt{\lambda c_{\mu}}}{2m} + \frac{2m}{\sqrt{\lambda c_{\mu}}} \log \frac{1}{\delta} + \frac{dm}{\sqrt{\lambda c_{\mu}}} \log \left( 1 + \frac{L^2 k_{\mu} (1 - \gamma^{-2(t-1)})}{\lambda c_{\mu} d (1 - \gamma^{-2})} \right) + \frac{2m}{\sqrt{\lambda c_{\mu}}} d \log(2) + \sqrt{\lambda c_{\mu}} S, \tag{49}
\]

Based on the inequality (47), Lemma 6 and Lemma 7, and the boundedness assumption of the feasible set, we have for any $x \in X, \delta \in (0, 1)$, with probability at least $1 - \delta$, we have for all $t \in [T]$, 
\[
\left\| \mu(x^T \tilde{\theta}_t) - \mu(x^T \theta_t) \right\| \leq 2\sqrt{1 + 2S} \frac{k_{\mu}}{\sqrt{\lambda c_{\mu}}} \left( L \frac{k_{\mu}}{\sqrt{\lambda c_{\mu}}} \sqrt{d} \sum_{p=1}^{t-1} \gamma^{t-p} \sqrt{\frac{\gamma^{p-1}}{1 - \gamma} \|\theta_p - \theta_{p+1}\|_2 + \beta_{t-1}} \right),
\]
\[
\leq 2\sqrt{1 + 2S} \frac{k_{\mu}}{\sqrt{\lambda c_{\mu}}} \left( L^2 \frac{k_{\mu}}{\sqrt{\lambda c_{\mu}}} \sqrt{d} \sum_{p=1}^{t-1} \gamma^{t-p} \sqrt{\frac{\gamma^{p-1}}{1 - \gamma} \|\theta_p - \theta_{p+1}\|_2 + \beta_{t-1}} \right),
\]
where $\beta_{t-1} \triangleq \frac{\sqrt{\lambda c_{\mu}}}{2m} + \frac{2m}{\sqrt{\lambda c_{\mu}}} \log \frac{1}{\delta} + \frac{dm}{\sqrt{\lambda c_{\mu}}} \log \left( 1 + \frac{L^2 k_{\mu} (1 - \gamma^{-2})}{\lambda c_{\mu} d (1 - \gamma^{-2})} \right) + \frac{2m}{\sqrt{\lambda c_{\mu}}} d \log(2) + \sqrt{\lambda c_{\mu}} S$ is the confidence radius used in SCB-WeightUCB. Hence we completes the proof. 

Proof of Lemma 8. Since $\forall \theta, H_t(\theta) \geq c_{\mu} V_{t-1}$, we have 
\[
\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^T \theta_t) - \mu(X_s^T \theta_s)) X_s \right\|_{H_{t^{-1}}(\theta_t)} \leq \frac{1}{\sqrt{\lambda c_{\mu}}} \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^T \theta_t) - \mu(X_s^T \theta_s)) X_s \right\|_{V_{t-1}^{-1}}.
\]
Then use Lemma 6 and we complete the proof. 

Proof of Lemma 9. Let $\tilde{H}_t(\theta) \triangleq \lambda c_{\mu} \gamma^{-2(t-1)} I_d + \sum_{s=1}^{t-1} \gamma^{-2s} \mu'(X_s^T \theta) X_s X_s^T$ which is only used in the analysis. 
\[
F_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_{\mu} \theta_t \right\|_{H_{t^{-1}}(\theta_t)} \leq \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{H_{t^{-1}}(\theta_t)} + \left\| \lambda c_{\mu} \theta_t \right\|_{H_{t^{-1}}(\theta_t)} 
\]
\[
\leq \left\| \sum_{s=1}^{t-1} \gamma^{-s} \eta_s X_s \right\|_{H_{t^{-1}}(\theta_t)} + \sqrt{\lambda c_{\mu}} S.
\]
Recall that $H_t(\theta) = \lambda c_{\mu} I_d + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu'(X_s^T \theta) X_s X_s^T$, so the last inequality comes from 
\[
\gamma^{-2(t-1)} H_t(\theta) = \lambda c_{\mu} \gamma^{-2(t-1)} I_d + \sum_{s=1}^{t-1} \gamma^{-t-s} \mu'(X_s^T \theta) X_s X_s^T \succeq \lambda c_{\mu} \gamma^{-2(t-1)} I_d + \sum_{s=1}^{t-1} \gamma^{-2s} \mu'(X_s^T \theta) X_s X_s^T = \tilde{H}_t^{-1}(\theta).
\]
From the weighted version concentration inequality [Russac et al., 2021, Theorem 3], restated in Theorem 6, we can get the bound for the first term $\| \sum_{s=1}^{t-1} \gamma^{-s} \eta_s X_s \|_{H_{t^{-1}}(\theta_t)}$. First by the model assumption (13), we know that $\sigma_t^2 = \mathbb{E}[\eta_t^2 | F_t]$, we have 
\[
Var[\eta_t | F_t] = \mu''(X_t \theta_t), \quad \text{then just let } w_t = \gamma^{-t}, \lambda_t = \lambda c_{\mu} \gamma^{-2t} \text{ and we have,}
\]
\[
\left\| \sum_{s=1}^{t-1} \gamma^{-s} \eta_s X_s \right\|_{H_{t^{-1}}(\theta_t)} \leq \sqrt{\frac{\lambda c_{\mu}}{2m}} + \frac{2m}{\sqrt{\lambda c_{\mu}}} \log \left( \frac{\det(\tilde{H}_t)^{1/2}}{\delta(\lambda c_{\mu} \gamma^{-2(t-1)} d)^{1/2}} \right) + \frac{2m}{\sqrt{\lambda c_{\mu}}} d \log(2).
\]
Then use Lemma 12 and let $w_{t,s} = \gamma^{-2s} \mu'(X_s^T \theta_t)$, $\lambda_t = \lambda c_{\mu} \gamma^{-2(t-1)}$, we get the upper bound for $\det(\tilde{H}_t)$, 
\[
\det(\tilde{H}_t) \leq \left( \lambda c_{\mu} \gamma^{-2(t-1)} + \frac{L^2 k_{\mu} \sum_{s=1}^{t-1} \gamma^{-2s}}{d} \right)^d,
\]
then,
\[
\left\| \sum_{s=1}^{t-1} \gamma^{-s} \eta_s X_s \right\|_{\tilde{H}^{-1}(\theta_t)} \leq \frac{2m \sqrt{\lambda c_\mu}}{\sqrt{\lambda c_\mu}} + \frac{2m}{\sqrt{\lambda c_\mu}} \log \frac{1}{\delta} + \frac{dm}{\sqrt{\lambda c_\mu}} \log \left( \frac{\lambda c_\mu \gamma^{-2(t-1)} + \frac{L^2 k_\mu}{\lambda c_\mu} \sum_{s=1}^{t-1} \gamma^{-s}}{\lambda c_\mu \gamma^{-2(t-1)}} \right) + \frac{2m}{\sqrt{\lambda c_\mu}} d \log(2).
\]

Therefore, we get the upper bound for \( F_t \) term.

\[ \square \]

### C.2 Proof of Theorem 3

**Proof.** Let \( X_t^* \triangleq \arg \max_{X \in \mathcal{X}} \mu(x^T \theta_t) \). Due to Lemma 2 and the fact that \( X_t^*, X_t \in \mathcal{X} \), each of the following holds with probability at least \( 1 - \delta \),
\[
\forall t \in [T], \mu(X_t^T \theta_t) \leq \mu(X_t^T \tilde{\theta}_t) + 2 \sqrt{1 + 2S} \frac{k_\mu}{\sqrt{c_\mu}} \left( L^2 \frac{k_\mu}{\sqrt{\lambda c_\mu}} \sqrt{d} \sum_{t=1}^{t-1} \gamma_{t-1} \sqrt{\frac{\gamma^{-p-1}}{1 - \gamma}} \| \theta_p - \theta_{p+1} \|_2 + \tilde{\beta}_{t-1} \| X_t^* \|_{V^{-1}_{t-1}} \right),
\]
\[
\forall t \in [T], \mu(X_t^T \theta_t) \geq \mu(X_t^T \tilde{\theta}_t) - 2 \sqrt{1 + 2S} \frac{k_\mu}{\sqrt{c_\mu}} \left( L^2 \frac{k_\mu}{\sqrt{\lambda c_\mu}} \sqrt{d} \sum_{t=1}^{t-1} \gamma_{t-1} \sqrt{\frac{\gamma^{-p-1}}{1 - \gamma}} \| \theta_p - \theta_{p+1} \|_2 + \tilde{\beta}_{t-1} \| X_t \|_{V^{-1}_{t-1}} \right).
\]

By the union bound, the following holds with probability at least \( 1 - 2\delta \): \( \forall t \in [T] \)
\[
\mu(X_t^T \theta_t) - \mu(X_t^T \tilde{\theta}_t) \leq 2 \sqrt{1 + 2S} \frac{2L^2 k_\mu^2}{c_\mu} \sqrt{d} \sum_{t=1}^{t-1} \gamma_{t-1} \sqrt{\frac{\gamma^{-p-1}}{1 - \gamma}} \| \theta_p - \theta_{p+1} \|_2 + 2 \sqrt{1 + 2S} \frac{k_\mu}{\sqrt{c_\mu}} \tilde{\beta}_{t-1} \| X_t \|_{V^{-1}_{t-1}},
\]
\[
\leq \frac{4 \sqrt{1 + 2S} L^2 k_\mu^2}{c_\mu} \sqrt{d} \sum_{t=1}^{t-1} \gamma_{t-1} \sqrt{\frac{\gamma^{-p-1}}{1 - \gamma}} \| \theta_p - \theta_{p+1} \|_2 + 4 \sqrt{1 + 2S} \frac{k_\mu}{\sqrt{c_\mu}} \tilde{\beta}_{t-1} \| X_t \|_{V^{-1}_{t-1}},
\]
where the last step comes from the arm selection criterion (17) such that
\[
\mu(X_t^T \tilde{\theta}_t) + 2 \sqrt{1 + 2S} \frac{k_\mu}{\sqrt{c_\mu}} \tilde{\beta}_{t-1} \| X_t^* \|_{V^{-1}_{t-1}} \leq \mu(X_t^* \tilde{\theta}_t) + 2 \sqrt{1 + 2S} \frac{k_\mu}{\sqrt{c_\mu}} \tilde{\beta}_{t-1} \| X_t \|_{V^{-1}_{t-1}}.
\]

Hence, the following dynamic regret bound holds with probability at least \( 1 - 2\delta \) and can be divided into two parts,
\[
R_T = \sum_{t=1}^{T} \mu(X_t^T \theta_t) - \mu(X_t^* \tilde{\theta}_t)
\]
\[
\leq \frac{4 \sqrt{1 + 2S} L^2 k_\mu^2}{c_\mu} \sqrt{d} \sum_{t=1}^{T} \gamma_{t-1} \sqrt{\frac{\gamma^{-p-1}}{1 - \gamma}} \| \theta_p - \theta_{p+1} \|_2 + 4 \sqrt{1 + 2S} \frac{k_\mu}{\sqrt{c_\mu}} \tilde{\beta}_T \sum_{t=1}^{T} \| X_t \|_{V^{-1}_{t-1}}.
\]

where \( \tilde{\beta}_T = \frac{\sqrt{\lambda c_\mu}}{2m} + \frac{2m}{\sqrt{\lambda c_\mu}} \log \frac{1}{\delta} + \frac{dm}{\sqrt{\lambda c_\mu}} \log \left( 1 + \frac{L^2 k_\mu}{\lambda c_\mu} (1 - \gamma^2) \right) + \frac{2m}{\sqrt{\lambda c_\mu}} d \log(2) + \sqrt{\lambda c_\mu} S \) is the confidence radius.

Now we derive the upper bound for these two parts separately.

**Bias Part.** Similar to the proof of inequality (25), we have
\[
\frac{4 \sqrt{1 + 2S} L^2 k_\mu^2}{c_\mu} \sqrt{d} \sum_{t=1}^{T} \gamma_{t-1} \sqrt{\frac{\gamma^{-p-1}}{1 - \gamma}} \| \theta_p - \theta_{p+1} \|_2 \leq \frac{8 \sqrt{1 + 2S} L^2 k_\mu^2}{c_\mu} \sqrt{d} \left( \frac{1}{\lambda (1 - \gamma)^{\gamma/2} P_T} \right).
\]
**Variance Part.** First use the Cauchy-Schwarz inequality, we know that
\[
\frac{4\sqrt{1+2Sk}}{\sqrt{c_\mu}} \beta_T \sum_{t=1}^{T} \|X_t\|_{V_t-1} \leq \frac{4\sqrt{1+2Sk}}{\sqrt{c_\mu}} \beta_T \sqrt{T} \left[ \sum_{t=1}^{T} \|X_t\|_{V_t-1}^2 \right].
\]

Then for the term \(\sum_{t=1}^{T} \|X_t\|_{V_t-1}^2\), we can directly use the Lemma 11 to bound it,
\[
\frac{4\sqrt{1+2Sk}}{\sqrt{c_\mu}} \beta_T \sqrt{T} \left[ \sum_{t=1}^{T} \|X_t\|_{V_t-1}^2 \right] \leq \frac{4\sqrt{1+2Sk}}{\sqrt{c_\mu}} \beta_T \sqrt{2 \max \{1, L^2/\lambda\} dT} \sqrt{T \log \frac{1}{\gamma} + \log \left(1 + \frac{L^2}{\lambda d(1-\gamma)}\right)}.
\]

Combining the upper bound for the bias and variance parts, and letting \(\delta = 1/(2T)\), we have the following regret bound with probability at least \(1 - 1/T\),
\[
R_T \leq \frac{8\sqrt{1+2S}L^2k_d^2}{c_\mu} \sqrt{\frac{d}{\lambda (1-\gamma)^{3/2}}} P_T + \frac{4\sqrt{1+2Sk}}{\sqrt{c_\mu}} \beta_T \sqrt{2 \max \{1, L^2/\lambda\} dT} \sqrt{T \log \frac{1}{\gamma} + \log \left(1 + \frac{L^2}{\lambda d(1-\gamma)}\right)}.
\]

where \(\tilde{\beta}_t = \frac{\sqrt{\bar{X}_n}}{2m} + \frac{dm}{\sqrt{\lambda c_\mu}} \log (2T) + \frac{dm}{\sqrt{\lambda c_\mu}} \log \left(1 + \frac{L^2k_r(1-\gamma^2)}{\lambda c_\mu d(1-\gamma)}\right) + \frac{dm}{\sqrt{\lambda c_\mu}} d \log (2) + \sqrt{\lambda c_\mu} S\). Since that there is a \(T \sqrt{\log(1/\gamma)}\) term in the regret bound, which means that we cannot let \(\gamma\) close to 0, so we set \(\gamma \geq 1/T\), then we have \(\log(1/\gamma) \leq C(1-\gamma)\), where \(C = \log T/(1-1/T)\). Then, ignoring logarithmic factors in time horizon \(T\), and let \(\lambda = d \log(T)/c_\mu\), we finally obtain that,
\[
R_T \leq \tilde{O} \left( \frac{k_d^2}{\sqrt{\lambda c_\mu}} \frac{1}{(1-\gamma)^{3/2}} P_T + \frac{k_r}{\sqrt{\lambda c_\mu}} d(1-\gamma)^{3/4} T\right).
\]

When \(P_T < d/(k_r T)\) (which corresponds a small amount of non-stationarity), we simply set \(\gamma = 1 - 1/T\) and achieve an \(\tilde{O}(k_r c_\mu^{-1/2} d \sqrt{T})\) regret bound. Besides, when coming to the non-degenerated case of \(P_T \geq d/(k_r T)\), We set the discounted factor optimally as \(1 - \gamma = \sqrt{k_r P_T/(dT)}\) and attain an \(\tilde{O}(k_r c_\mu^{-1/2} d^{1/4} P_T^{1/4} T^{3/4})\) regret bound, which completes the proof.

---

**D Piecewise-Stationary SCB**

In this section, we study SCB under piecewise-stationary environment and our work is a direct improvement over [Russac et al., 2021]. Next, we will first propose our SCB-PW-WeightUCB algorithm, and then, present the analysis of the confidence set. Finally, we give the proof of the dynamic regret upper bound.

**D.1 SCB-PW-WeightUCB Algorithm**

Inspired by Abeille et al. [2021], we make a direct improvement over Russac et al. [2021]. Just like Russac et al. [2021], for \(D \geq 1\), define \(T(D) = \{1 \leq t \leq T\}\), such that \(\theta_t = \theta_t^*\) for \(t - D \leq s \leq t - 1\). \(t \in T(D)\) when \(t\) is at least \(D\) steps away from the previous closest changing point. But the difference is that Russac et al. [2021] considers \(D\) as an analysis parameter, and we treat \(D\) as a tunable algorithm parameter. Notice that, the \(D\) here is not a virtual window size, but the algorithm’s estimate of how durable the environment is stationary.

**Estimator.** At iteration \(t\), we adopt the same maximum likelihood estimator as in the drifting case as defined in (8).

**Confidence Set.** We further construct confidence set for the real \(\theta_t^*\). For \(\delta \in (0, 1)\), we define,
\[
C_t(\delta) \triangleq \{\theta \in \Theta \mid \|g_t(\theta) - g_t(\tilde{\theta}_t)\|_{H_{\theta_t^*}^{-1}} \leq \rho_t\},
\]
where \(\rho_t = \frac{2L^2k}{\sqrt{\lambda c_\mu}} \frac{\gamma^2}{1-\gamma} + \frac{L m}{\sqrt{\lambda c_\mu}} \frac{\gamma D}{1-\gamma} + \tilde{\beta}_t\) and \(\tilde{\beta}_t = \frac{dm}{\sqrt{\lambda c_\mu}} \log \left(1 + \frac{L^2k_r(1-\gamma^2)}{\lambda c_\mu d(1-\gamma)}\right) + \frac{\lambda c_\mu}{2m} + \frac{2m}{\sqrt{\lambda c_\mu}} \log \frac{1}{\gamma} + \frac{2m}{\sqrt{\lambda c_\mu}} d \log (2) + \sqrt{\lambda c_\mu} S\).
Lemma 10. For any $\delta \in (0, 1)$, with probability at least $1 - \delta$, we have $\forall t \in T(D), \theta_t \in C_t(\delta)$.

\[
C_t(\delta) = \left\{ \theta \in \Theta \mid \|g_t(\theta) - g_t(\hat{\theta}_t)\|_{H_t^{-1}(\theta)} \leq \frac{2L^2 S k_{\mu}}{\sqrt{\lambda c_{\mu}}} \frac{\gamma^D}{1 - \gamma} + \frac{Lm}{\sqrt{\lambda c_{\mu}}} \frac{\gamma^D}{1 - \gamma} + \beta_t \right\},
\]

where $\beta_t = \frac{\sqrt{\lambda c_{\mu}}}{2m} + \frac{2m}{\sqrt{\lambda c_{\mu}}} \log \frac{1}{\delta} + \frac{dm}{\sqrt{\lambda c_{\mu}}} \log \left(1 + \frac{L^2 k_{\mu} (1 - \gamma^2)}{\lambda c_{\mu} d (1 - \gamma)} \right) + \frac{2m}{\sqrt{\lambda c_{\mu}}} d \log(2) + \sqrt{\lambda c_{\mu} S}$. The proof of Lemma 10 is presented in Appendix D.2.

**Selection Criteria.** Algorithms discussed earlier for drifting cases are using bonus-based selection criteria. But here we use a parameter-based selection criterion as follows,

\[
(X_t, \tilde{\theta}_t) = \arg \max_{x \in X, \theta \in C_t(\delta)} \mu(x^\top \theta).
\]

(51)

The main difference between parameter-based and bonus-based selection criteria is discussed in Section 3.2 of Abeille et al. [2021]. The overall algorithm is summarized in Algorithm 5.

**D.2 Proof of Lemma 10**

*Proof.* Based on the model assumption (13), the function $g_t$ (10) and the $g_t(\hat{\theta}_t)$ (44), we have,

\[
g_t(\theta_t) - g_t(\hat{\theta}_t) = \lambda c_{\mu} \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_{s}^\top \theta_t) X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s
\]

\[
= \lambda c_{\mu} \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_{s}^\top \theta_t) X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_{s}^\top \theta_s) + \eta_s) X_s
\]

\[
= \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_{s}^\top \theta_t) - \mu(X_{s}^\top \theta_s)) X_s + \lambda c_{\mu} \theta_t - \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s.
\]

Then,

\[
\|g_t(\theta_t) - g_t(\hat{\theta}_t)\|_{H_t^{-1}(\theta_t)} = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_{s}^\top \theta_t) - \mu(X_{s}^\top \theta_s)) X_s + \lambda c_{\mu} \theta_t - \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{H_t^{-1}(\theta_t)}
\]

\[
\leq \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_{s}^\top \theta_t) - \mu(X_{s}^\top \theta_s)) X_s \right\|_{H_t^{-1}(\theta_t)} + \left\| \lambda c_{\mu} \theta_t - \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{H_t^{-1}(\theta_t)}
\]

\[
\leq \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_{s}^\top \theta_t) - \mu(X_{s}^\top \theta_s)) X_s \right\|_{H_t^{-1}(\theta_t)} + \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{H_t^{-1}(\theta_t)}
\]

\[
+ \left\| \sum_{s=t-D}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_{\mu} \theta_t \right\|_{H_t^{-1}(\theta_t)}.
\]
Term (a). Since \( t \in \mathcal{T}(D) \), we have
\[
\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{H_t^{-1}(\theta_t)} = \left\| \sum_{s=1}^{t-D-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{H_t^{-1}(\theta_t)} 
\leq \left\| \sum_{s=1}^{t-D-1} \gamma^{t-s-1} k_{\mu} X_s^\top (\theta_t - \theta_s) X_s \right\|_{H_t^{-1}(\theta_t)} 
\leq \sum_{s=1}^{t-D-1} \gamma^{t-s-1} k_{\mu} \|X_s\|_2 \|((\theta_t - \theta_s))_2 \|X_s\|_{H_t^{-1}(\theta_t)} 
\leq \frac{2L^2 S k_{\mu}}{\sqrt{\lambda c_{\mu}}} \frac{\gamma}{1-\gamma}.
\]

Term (b).
\[
\left\| \sum_{s=1}^{t-D-1} \gamma^{t-s-1} \eta_s X_s \right\|_{H_t^{-1}(\theta_t)} \leq \sum_{s=1}^{t-D-1} \gamma^{t-s-1} m \|X_s\|_{H_t^{-1}(\theta_t)} \leq \frac{L m}{\sqrt{\lambda c_{\mu}}} \sum_{s=1}^{t-D-1} \gamma^{t-s-1} \leq \frac{L m}{\sqrt{\lambda c_{\mu}}} \frac{\gamma}{1-\gamma}.
\]

Term (c). Let \( \bar{H}_{t-D,t}(\theta) = \lambda c_{\mu} \gamma^{-2(t-1)} I_d + \sum_{s=t-D}^{t-1} \gamma^{-2s} \mu'(X_s^\top \theta) X_s X_s^\top \)
\[
\left\| \sum_{s=t-D}^{t-1} \gamma^{-s-1} \eta_s X_s - \lambda c_{\mu} \theta \right\|_{H_t^{-1}(\theta_t)} \leq \left\| \sum_{s=t-D}^{t-1} \gamma^{-s-1} \eta_s X_s \right\|_{H_t^{-1}(\theta_t)} + \sqrt{\lambda c_{\mu} S} 
\leq \left\| \sum_{s=t-D}^{t-1} \gamma^{-s} \eta_s X_s \right\|_{\bar{H}_{t-D,t}(\theta_t)} + \sqrt{\lambda c_{\mu} S} 
\leq \left\| \sum_{s=t-D}^{t-1} \gamma^{-s} \eta_s X_s \right\|_{\bar{H}_{t-D,t}(\theta_t)} + \sqrt{\lambda c_{\mu} S}.
\]

We already proof that \( \gamma^{-2(t-1)} H_t(\theta) \geq \bar{H}_{t}(\theta) \) in (50), and obviously \( \bar{H}_{t}(\theta) \geq \bar{H}_{t-D,t}(\theta) \). Next, we need to bound the term \( \left\| \sum_{s=t-D}^{t-1} \gamma^{-s} \eta_s X_s \right\|_{\bar{H}_{t-D,t}(\theta_t)} \) using self-normalization bound [Russac et al., 2021, Theorem 3], restated in Theorem 6 by let \( \eta_t = \gamma^{-t}, \lambda_t = \lambda c_{\mu} \gamma^{-2t} \), then we have
\[
\left\| \sum_{s=t-D}^{t-1} \gamma^{-s} \eta_s X_s \right\|_{\bar{H}_{t-D,t}(\theta_t)} 
\leq \frac{\sqrt{\lambda c_{\mu}}}{2m} + \frac{2m}{\sqrt{\lambda c_{\mu}}} \log \left( \frac{\det (\bar{H}_{t-D,t})^{1/2}}{\delta (\lambda c_{\mu} \gamma^{-2(t-1)} d/2)} \right) + \frac{2m}{\sqrt{\lambda c_{\mu}}} d \log(2) 
\leq \frac{\sqrt{\lambda c_{\mu}}}{2m} + \frac{2m}{\sqrt{\lambda c_{\mu}}} \log \left( \frac{\lambda c_{\mu} \gamma^{-2(t-1)} + \frac{L^2 k_{\mu}}{\lambda c_{\mu}} \sum_{s=t-D}^{t-1} \gamma^{-2s}}{\lambda c_{\mu} \gamma^{-2(t-1)} d} \right) + \frac{2m}{\sqrt{\lambda c_{\mu}}} d \log(2) 
\leq \frac{\sqrt{\lambda c_{\mu}}}{2m} + \frac{2m}{\sqrt{\lambda c_{\mu}}} \log \left( \frac{1}{\delta} + \frac{L^2 k_{\mu} (1 - \gamma^2 D)}{\lambda c_{\mu} d (1 - \gamma)} \right) + \frac{2m}{\sqrt{\lambda c_{\mu}}} d \log(2).
\]

Let \( \tilde{\beta}_t \triangleq \frac{\sqrt{\lambda c_{\mu}}}{2m} + \frac{2m}{\sqrt{\lambda c_{\mu}}} \log \left( \frac{1}{\delta} + \frac{L^2 k_{\mu} (1 - \gamma^2 D)}{\lambda c_{\mu} d (1 - \gamma)} \right) + \frac{2m}{\sqrt{\lambda c_{\mu}}} d \log(2) + \sqrt{\lambda c_{\mu} S} \), finally we have,
\[
\left\| g_t(\theta_t) - g_t(\bar{\theta}_t) \right\|_{H_t^{-1}(\theta_t)} \leq \frac{2L^2 S k_{\mu}}{\sqrt{\lambda c_{\mu}}} \frac{\gamma}{1-\gamma} + \frac{L m}{\sqrt{\lambda c_{\mu}}} \frac{\gamma}{1-\gamma} + \tilde{\beta}_t,
\]
which completes the proof.
D.3 Proof of Theorem 4

Proof. Let \( R_t = \mu(X_t^\top \theta_t) - \mu(X_{t-1}^\top \theta_{t-1}) \)
\[
R_T = \sum_{t=1}^T R_t = \sum_{t \in \mathcal{T}(D)} R_t + \sum_{t \in \mathcal{T}(D)} R_t = \Gamma_T D + \sum_{t \in \mathcal{T}(D)} R_t.
\]

For \( t \in \mathcal{T}(D) \), by selection criterion (51),
\[
R_t = \mu(X_t^\top \theta_t) - \mu(X_{t-1}^\top \theta_{t-1}) \\
\leq \mu(X^\top \tilde{\theta}_t) + \mu(X_t^\top \tilde{\theta}_t) - \mu(X_{t-1}^\top \tilde{\theta}_{t-1}) \\
\leq \alpha(X_t, \tilde{\theta}_t, \tilde{\theta}_t) \bigg| X_t^\top \tilde{\theta}_t - \tilde{\theta}_t \bigg| + \alpha(X_t, \theta_t, \tilde{\theta}_t) \bigg| X_t^\top \theta_t - \tilde{\theta}_t \bigg| \\
\leq \sqrt{1 + 2S} \left( \alpha(X_t, \tilde{\theta}_t, \tilde{\theta}_t) \| X_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)} \| g_t(\tilde{\theta}_t) - g_t(\tilde{\theta}_t) \|_{H_{t-1}^{-1}(\tilde{\theta}_t)} \\
+ \alpha(X_t, \theta_t, \tilde{\theta}_t) \| X_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)} \| g_t(\theta_t) - g_t(\tilde{\theta}_t) \|_{H_{t-1}^{-1}(\tilde{\theta}_t)} \right).
\]

where \( \alpha(x, \theta_1, \theta_2) = \int_0^1 \mu'\left(vx^\top \theta_2 + (1 - v)x^\top \theta_1 \right) dv \), and the last second inequality comes from the mean value theorem \( \mu(x^\top \theta_1) - \mu(x^\top \theta_2) = \alpha(x, \theta_1, \theta_2)(x^\top \theta_1 - x^\top \theta_2) \). Since that \( \tilde{\theta}_t \in C_t(\delta) \) and with probability at least \( 1 - \delta \), \( \forall t \in [T], \theta_t \in C_t(\delta) \), and by union bound, the following dynamic regret bound hold with probability at least \( 1 - \delta \),
\[
\sum_{t \in \mathcal{T}(D)} R_t \leq \sqrt{1 + 2S} \left( \sum_{t \in \mathcal{T}(D)} \alpha(X_t, \tilde{\theta}_t, \tilde{\theta}_t) \| X_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)} \| g_t(\tilde{\theta}_t) - g_t(\tilde{\theta}_t) \|_{H_{t-1}^{-1}(\tilde{\theta}_t)} + \sum_{t \in \mathcal{T}(D)} \alpha(X_t, \theta_t, \tilde{\theta}_t) \| X_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)} \| g_t(\theta_t) - g_t(\tilde{\theta}_t) \|_{H_{t-1}^{-1}(\tilde{\theta}_t)} \right).
\]

Now we try to derive the upper bound for term \( \sum_{t \in \mathcal{T}(D)} \alpha(X_t, \tilde{\theta}_t, \tilde{\theta}_t) \| X_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)} \| g_t(\tilde{\theta}_t) - g_t(\tilde{\theta}_t) \|_{H_{t-1}^{-1}(\tilde{\theta}_t)} \).

Based on the definition of \( g_t \) (10), we have
\[
g_t(\theta_1) - g_t(\theta_2) = \lambda c_\mu (\theta_1 - \theta_2) + \sum_{s=1}^{t-1} \gamma^{t-s-1}(\mu(X_s^\top \theta_1) - \mu(X_s^\top \theta_2)) X_s \\
= \lambda c_\mu (\theta_1 - \theta_2) + \sum_{s=1}^{t-1} \gamma^{t-s-1} \alpha(X_s, \theta_1, \theta_2) X_s^\top X_s (\theta_1 - \theta_2) \\
= \left( \lambda c_\mu + \sum_{s=1}^{t-1} \gamma^{t-s-1} \alpha(X_s, \theta_1, \theta_2) X_s^\top X_s \right) (\theta_1 - \theta_2).
\]

Then based on the definition of \( G_t \) (45), we know \( G_t(\theta_1, \theta_2) = \lambda c_\mu I_d + \sum_{s=1}^{t-1} \gamma^{t-s-1} \alpha(X_s, \theta_1, \theta_2) X_s^\top X_s \), which means \( G_t(\tilde{\theta}_t, \tilde{\theta}_t) = \lambda c_\mu I_d + \sum_{s=1}^{t-1} \gamma^{t-s-1} \alpha(X_s, \tilde{\theta}_t, \tilde{\theta}_t) X_s^\top X_s \), if we let \( \bar{X}_s = \sqrt{\alpha(X_s, \tilde{\theta}_t, \tilde{\theta}_t)} X_s \), then
\[
\sum_{t \in \mathcal{T}(D)} \alpha(X_t, \tilde{\theta}_t, \tilde{\theta}_t) \| X_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)} \| g_t(\tilde{\theta}_t) - g_t(\tilde{\theta}_t) \|_{H_{t-1}^{-1}(\tilde{\theta}_t)} \\
\leq \sqrt{k_\mu T} \sum_{t=1}^T \| \bar{X}_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)}^2.
\]

Then for the term \( \sqrt{\sum_{t=1}^T \| \bar{X}_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)}^2} \), we can directly use the Lemma 11 to bound it,
\[
\sqrt{k_\mu T} \sum_{t=1}^T \| \bar{X}_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)}^2 \leq \sqrt{2k_\mu \max\{1, L^2 k_\mu/(\lambda c_\mu)\}} \int_0^T \sum_{t=1}^T \| \bar{X}_t \|_{G_{t-1}^{-1}(\tilde{\theta}_t, \tilde{\theta}_t)}^2 \left( T \log \frac{1}{\gamma} + \log \left( 1 + \frac{L^2 k_\mu}{\lambda c_\mu d(1 - \gamma)} \right) \right).
\]
We can bound term \( \sum_{t \in T(D)} \alpha(X_t, \theta_t, \hat{\theta}_t) \|X_t\|_{G_t^{-1}(\theta_t, \hat{\theta}_t)} \) in the same way and get,
\[
\sum_{t \in T(D)} \alpha(X_t, \theta_t, \hat{\theta}_t) \|X_t\|_{G_t^{-1}(\theta_t, \hat{\theta}_t)} \leq 2k_\mu \max \{1, L^2 k_\mu / (\lambda c_\mu)\} dT \sqrt{T \log \frac{1}{\gamma} + \log \left( 1 + \frac{L^2 k_\mu}{\lambda c_\mu d(1 - \gamma)} \right)}.
\]
Combine these two bound and let \( \delta = 1/T \), we have the following regret bound with probability at least \( 1 - 1/T \),
\[
R_T \leq \Gamma_T D + 2\sqrt{T} + 25 \rho_T \sqrt{2k_\mu \max \{1, L^2 k_\mu / (\lambda c_\mu)\} dT} \sqrt{T \log \frac{1}{\gamma} + \log \left( 1 + \frac{L^2 k_\mu}{\lambda c_\mu d(1 - \gamma)} \right)},
\]
where \( \rho_T = \frac{2L^2 k_{\mu}}{\sqrt{\lambda c_\mu}} \frac{\gamma^D}{1 - \gamma} + \frac{Lm}{\sqrt{\lambda c_\mu}} \frac{\gamma^D}{1 - \gamma} + \tilde{\beta}_t \) and \( \tilde{\beta}_t = \frac{dm}{\sqrt{\lambda c_\mu}} \log \left( 1 + \frac{L^2 k_\mu (1 - \gamma^D)}{\lambda c_\mu d(1 - \gamma)} \right) + \frac{\sqrt{\lambda c_\mu}}{2m} + \frac{2m}{\sqrt{\lambda c_\mu}} \log (T) + \frac{2m}{\sqrt{\lambda c_\mu}} d \log (2) + \sqrt{\lambda c_\mu} S \). Since that there is a \( T \sqrt{\log(1/\gamma)} \) term in the regret bound, which means that we cannot let \( \gamma \) close to 0, so we set \( \gamma \geq 1/2 \), then we have \( \log(1/\gamma) \leq 2 \log(2)(1 - \gamma) \). Then, we set \( D = \log(T) / \log(1/\gamma) \), noticing that \( 0 < 1/\gamma - 1 < 1 \) and using \( \log(1 + x) \geq x/2 \) for \( 0 < x < 1 \), we have
\[
\log \left( \frac{1}{\gamma} \right) = \log(1 + 1/\gamma - 1) \geq \frac{1 - \gamma}{2\gamma}.
\]
Therefore, we have \( D \leq \frac{2\gamma \log(\Gamma)}{1 - \gamma} \). Then, ignoring logarithmic factors in time horizon \( T \), and let \( \lambda = \log(\log(T))/c_\mu \), we finally obtain that,
\[
R_T \leq \tilde{O} \left( \frac{1}{1 - \gamma} \Gamma_T + \left( \frac{1}{\sqrt{d}} \frac{1}{1 - \gamma} T + \sqrt{d(1 - \gamma) T} \right) \right) \leq \tilde{O} \left( \frac{1}{1 - \gamma} \Gamma_T + \frac{1}{\sqrt{1 - \gamma}} + d \sqrt{(1 - \gamma) T} \right).
\]
When \( \Gamma_T < d / \sqrt{T} \) (which corresponds a small amount of non-stationarity), we simply set \( \gamma = 1 - 1/T \) and achieve an \( \tilde{O}(d / \sqrt{T}) \) regret bound. Besides, when coming to the non-degenerated case of \( \Gamma_T > d / \sqrt{T} \), we set the discounted factor optimally as \( 1 - \gamma = (\Gamma_T / (dT))^{1/3} \) and attain an \( \tilde{O}(d^{2/3} \Gamma_T^{-1/3} T^{1/3}) \) dynamic regret bound, which completes the proof.

E  Thechnical Lemmas

In this section, we provide several useful lemmas, mainly about weighted version self-normalized concentration, weighted version potential lemma and some derivatives of self-concordant property.

E.1 Weighted Version Self-normalized Concentration

**Theorem 5** (Weighted Version Self-normalized Bound for Vector-Valued Martingales [Russac et al., 2019, Theorem 1]). Let \( \{F_t\}_{t=0}^\infty \) be a filtration, \( \{\eta_t\}_{t=0}^\infty \) be a real-valued stochastic process such that \( \eta_t \) is \( F_t \)-measurable and \( \eta_t \) is conditionally \( R \)-sub-Gaussian for some \( R \geq 0 \), such that
\[
\forall \lambda \in \mathbb{R}, \mathbb{E} \left[ \exp(\lambda \eta_t) \mid X_{1:t}, \eta_{1:t-1} \right] \leq \exp \left( \frac{\lambda^2 R^2}{2} \right).
\]
Let \( \{X_t\}_{t=1}^\infty \) be an \( \mathbb{R}^d \)-valued stochastic process such that \( X_t \) is \( F_{t-1} \)-measurable. For any \( t \geq 0 \), define
\[
\bar{V}_t = \mu_t I_d + \sum_{s=1}^t w_s^2 X_s X_s^\top, \quad S_t = \sum_{s=1}^t w_s \eta_s X_s,
\]
where \( s \geq 0, t \geq 0, w_s, \mu_t > 0 \). Then, for any \( \delta > 0 \), with probability at least \( 1 - \delta \), we have
\[
\forall t \geq 0, \|S_t\|_{\bar{V}_{t-1}} \leq R \left( 2 \log \frac{1}{\delta} + d \log \left( 1 + \frac{L^2 \sum_{s=1}^t w_s^2}{\delta} \right) \right).
\]

**Theorem 6** (Theorem 3 of Russac et al. [2021]). Let \( t \) be a fixed time instant. Let \( \{F_t\}_{t=0}^\infty \) be a filtration. Let \( \{X_t\}_{t=0}^\infty \) be a stochastic process on \( \mathbb{R}^d \) such that \( X_t \) is \( F_{t-1} \)-measurable and \( \|X_t\|_2 \leq 1 \). Let \( \{\eta_t\}_{t=0}^\infty \) be a martingale difference sequence such that \( \eta_t \) is \( F_t \)-measurable. Assume that the weights are non-decreasing, strictly positive and the time horizon
is known. Furthermore, assume that conditionally on $\mathcal{F}_t$ we have $\|\eta_t\| \leq m$ a.s. Let $\{\lambda_t\}_{t=0}^{\infty}$ be a deterministic sequence of regularization terms and denote $\sigma_t^2 = \mathbb{E}[\|\eta_t^2| \mathcal{F}_t]$. Let $\tilde{H}_t = \sum_{s=1}^{t-1} w_s^2 \sigma_s^2 X_s X_s^\top + \lambda_{t-1} I_d$ and $S_t = \sum_{s=1}^{t-1} w_s \eta_s X_s$, then for any $\delta \in (0,1]$, with probability at least $1 - \delta$, 
\[
\forall t \geq 0, \|S_t\|_{\tilde{H}_t^{-1}} \leq \frac{\sqrt{\lambda_t - 1}}{2mw_t-1} + \frac{2wmw_t-1}{\sqrt{\lambda_t-1}} \log \left( \frac{\det(\tilde{H}_t)^{1/2}}{\delta \lambda_t^{-d/2}} \right) + \frac{2wmw_t-1}{\sqrt{\lambda_t-1}} d \log(2).
\]

E.2 Weighted Version Potential Lemma

**Lemma 11** (Weighted Version Potential Lemma [Faury et al., 2021, Lemma 8]). Suppose $V_t = \sum_{s=1}^{t} \gamma^{t-s} X_s X_s^\top + \lambda I_d, V_0 = \lambda I_d, \gamma \in (0,1]$ and $\|X_t\|_2 \leq L$ for all $t \geq 1$, then the following inequality holds,
\[
\sum_{i=1}^{T} \|X_t\|_2^2 \leq 2 \max\{1, L^2/\lambda\} d \left( T \log \frac{1}{\gamma} + \log \left( 1 + \frac{L^2}{\lambda d(1 - \gamma)} \right) \right).
\]

**Lemma 12** (Determinant inequality). Let $V_t = \sum_{s=1}^{t} w_{t,s} X_s X_s^\top + \lambda_t I_d, V_0 = \lambda_0 I_d$. Assume $\|X_t\|_2 \leq L$ and we have,
\[
\det(V_t) \leq \left( \lambda_t + \frac{L^2 \sum_{s=1}^{t} w_{t,s}}{d} \right)^d.
\]

**Proof.** Now we have $V_t = \sum_{s=1}^{t} w_{t,s} X_s X_s^\top + \lambda_t I_d$, take the trace on both sides, and get the upper bound of $\text{Tr}(V_t)$
\[
\text{Tr}(V_t) = \text{Tr}(\lambda_t I_d) + \sum_{s=1}^{t} w_{t,s} \text{Tr}(X_s X_s^\top) = \lambda_t d + \sum_{s=1}^{t} w_{t,s} \|X_s\|_2^2 \leq \lambda_t d + L^2 \sum_{s=1}^{t} w_{t,s}.
\]

Base on the definition of determinant and the upper bound of $\text{Tr}(V_t)$ (54), we can get the upper bound for $\det(V_t)$,
\[
\det(V_t) = \prod_{i=1}^{d} \lambda_i \leq \left( \frac{\sum_{i=1}^{d} \lambda_i}{d} \right)^d = \left( \frac{\text{Tr}(V_t)}{d} \right)^d \leq \left( \lambda_t + \frac{L^2 \sum_{s=1}^{t} w_{t,s}}{d} \right)^d.
\]

E.3 Self-Concordant Properties

Based on the generalized self-concordant property of the (inverse) link function $\mu(\cdot)$, we have the following lemma, which will be later used to derive Lemma 14.

**Lemma 13** (Lemma 9 of Faury et al. [2020]). For any $z_1, z_2 \in \mathbb{R}$, we have the following inequality:
\[
\mu'(z_1) \frac{1 - \exp(-|z_1 - z_2|)}{|z_1 - z_2|} \leq \int_0^1 \mu'(z_1 + v(z_2 - z_1)) dv \leq \mu'(z_1) \frac{\exp(|z_1 - z_2|) - 1}{|z_1 - z_2|}.
\]

Furthermore, $\int_0^1 \mu'(z_1 + v(z_2 - z_1)) dv \geq \mu'(z_1)(1 + |z_1 - z_2|)^{-1}$.

The following lemma provides a weighted version of Lemma 10 of Faury et al. [2020] which can be easily proven.

**Lemma 14.** With $G_t$ defined in (45) and $H_t$ defined in (46), the following inequalities hold
\[
\forall \theta_1, \theta_2 \in \Theta, \quad G_t(\theta_1, \theta_2) \geq (1 + 2S)^{-1} H_t(\theta_1), \quad G_t(\theta_1, \theta_2) \geq (1 + 2S)^{-1} H_t(\theta_2).
\]