arXiv:2303.02691v2 [cs.LG] 7 Jun 2023

修订 加权策略：非平稳参数化老虎机 s

Jing Wang、Peng Zhao、Zhi-Hua Zhou新型软件
技术国家重点实验室，南京大学，南京 210023，中国{
wangjing、zhaop、zhouzh}@lamda.nju.edu.cn

摘要

近年来，非平稳参数化老虎机受到了广泛关注。应对非平稳性主要有三种方法，包括滑动窗口策略、加权策略和重启策略。由于许多非平稳环境呈现渐进漂移模式，加权策略在实际应用中得到了广泛采用。然而，已有理论研究表明，其分析更加复杂，并且相关算法要么计算效率较低，要么在统计上并非最优。本文重新审视了非平稳参数化老虎机中的加权策略。在<glossary>线性多臂老虎机</glossary>（LB）中，我们发现，这一不理想特征源于不充分的遗憾分析，进而导致了过于复杂的算法设计。我们提出了一个精细化分析框架，该框架简化了推导过程，更重要的是，产生了一种更简单的基于权重的算法：其效率与基于窗口或重启的算法相当，同时保持了与已有研究相同的遗憾。此外，我们的新框架还可用于改进其他参数化老虎机的遗憾界，包括广义线性老虎机（GLB）和自协调老虎机（SCB）。例如，我们设计了一种简单的加权 GLB 算法，其遗憾为 $\widetilde{\mathcal{O}}\left(k_\mu^{5/4} c_\mu^{-3/4} d^{3/4} P_T^{1/4} T^{3/4}\right)$，改进了已有工作中的 $\widetilde{\mathcal{O}}\left(k_\mu^2 c_\mu^{-1} d^{9/10} P_T^{1/5} T^{4/5}\right)$ 界，其中 k_μ 和 c_μ 刻画奖励模型的非线性，P_T 衡量非平稳性，d 和 T 分别表示维度和时间范围。

1 引言

非平稳参数化老虎机对序贯决策问题进行建模，其中每个臂的奖励分布由未知的时变参数构成结构，这类问题近年来得到了广泛研究 [Cheung 等人，2019，Russac 等人，2019，Zhao 等人，2020，Russac 等人，2020，Faury 等人，2021，Rus- sac 等人，2021，Wei 和 Luo，2021，Deng 等人，2022，Liu 等人，2022]这是由于其在许多现实世界中的非平稳在线应用中具有重要意义，例如推荐系统 [Tomkins 等人，2021，Huleihel 等人，2021], 以及与强化学习理论基础的紧密联系 [Jin 等人，2020，Touati 和 Vincent，2020]。

线性多臂老虎机（LB）是参数化老虎机的一个基本实例。在时刻 t 拉动某一臂所获得的期望奖励，是该臂的特征向量 X_t 与未知参数 θ_t 的内积，即 $\mathbb{E}[r_t \mid X_t] = X_t^\top \theta_t$。此外，广义线性老虎机（GLB）是对 LB 的推广，用于建模更广泛的奖励函数，例如二元奖励；其期望奖励遵循广义线性模型 $\mathbb{E}[r_t \mid X_t] = \mu(X_t^\top \theta_t)$，其中 $\mu(\cdot)$ 是逆链接函数。值得注意的是，非平稳模型允许上述模型中的参数 θ_t 随时间变化。通常使用两种非平稳性度量来量化参数变化的程度：（i）在逐渐漂移的情形中，使用路径长度 $P_T = \sum_{t=2}^{T} \|\theta_{t-1} - \theta_t\|_2$ 衡量底层参数的累积变化；（ii）在分段平稳情形中，Γ_T 表示在 T 轮中参数变化的次数。

应对非平稳性有三种有原则的方法：滑动窗口、加权和重启策略。采用滑动窗口策略时，学习者维护一个包含最新观测数据的时间窗口，以丢弃过时数据。采用加权策略时，学习者为最新数据赋予更高权重，为旧数据赋予更低权重，从而逐渐遗忘过时数据。采用重启策略时，学习者按照一定周期重启算法，以丢弃过时数据。目前，关于非平稳（广义）线性多臂老虎机的最佳已知结果由Wei 和 Luo [2021], 提出。他们设计了一种最优算法，由非平稳性检测器和基础算法组成；后者在近似平稳的环境中表现良好。每当检测器检测到

《 26th国际人工智能与统计学会议（AISTATS）2023》论文集，西班牙瓦伦西亚。PMLR：第206卷。版权所有 © 2023，作者所有。

表1：在不同的非平稳参数化老虎机模型下，我们的动态遗憾界与基于权重的算法此前已知的最优结果的比较。下文中，k_μ/c_μ 表示非线性度，在 LB 情形下为 1；d 表示维度，路径长度 P_T 和变化次数 Γ_T 分别是漂移情形和分段平稳情形的非平稳性度量。

参数化老虎机模型	先前工作	我们的结果
漂移线性老虎机	$\widetilde{\mathcal{O}}(d^{7/8}P_T^{1/4}T^{3/4})$ [Russac 等人，2019]	$\widetilde{\mathcal{O}}(d^{3/4}P_T^{1/4}T^{3/4})$ [定理 1]
漂移广义线性赌博机	$\widetilde{\mathcal{O}}\left(\frac{k_\mu^2}{c_\mu}d^{9/10}P_T^{1/5}T^{4/5}\right)$ [Faury 等人，2021]	$\widetilde{\mathcal{O}}\left(\frac{k_\mu^{5/4}}{c_\mu^{3/4}}d^{3/4}P_T^{1/4}T^{3/4}\right)$ [定理 2]
漂移半参数老虎机	$\widetilde{\mathcal{O}}\left(\frac{k_\mu^2}{c_\mu}d^{9/10}P_T^{1/5}T^{4/5}\right)$ [Faury 等人，2021]	$\widetilde{\mathcal{O}}\left(\frac{k_\mu^{5/4}}{c_\mu^{1/2}}d^{3/4}P_T^{1/4}T^{3/4}\right)$ [定理 3]
分段平稳 SCB	$\widetilde{\mathcal{O}}\left(\frac{1}{c_\mu^{1/3}}d^{2/3}\Gamma_T^{1/3}T^{2/3}\right)$ [Russac 等人，2021]	$\widetilde{\mathcal{O}}(d^{2/3}\Gamma_T^{1/3}T^{2/3})$ [定理 4]

非平稳性超过某个阈值时，算法就会重启自身，以抵御非平稳性。从这个意义上说，他们的算法可视为一种自适应重启算法。在 RestartUCB 算法的基础上 [Zhao 等人，2020] 并精心设计了结合多尺度探索的非平稳性检测器，其算法能够为 LB 和 GLB 实现 $\widetilde{\mathcal{O}}(\min\{\sqrt{\Gamma_T T}, P_T^{1/3}T^{2/3}\})$ 最优动态遗憾。

在现实场景中，环境的分布变化往往呈现逐渐漂移的模式 [Crammer 等人，2010，Chiang 等人，2013，Gama 等人，2014]，在这种情况下，软加权策略在应对非平稳性时（从经验上看）可能比硬重启策略更具优势，这一点可从赌博机学习 [Russac 等人，2019，Zhao 等人，2020，Deng 等人，2022]，概念漂移分类 [Anagnostopoulos 等人，2012，Zhao 等人，2021a]，和自适应系统辨识 [Guo 等人，1993，Chu 和 Mak，2017]中观察到。因此，设计一种用于非平稳参数化赌博机的自适应加权算法将极具吸引力。该算法通过赋予权重来降低过去数据的重要性，并根据环境自适应地设置权重。为此，我们考察了现有基于加权策略的非平稳参数化赌博机方法，并（令人惊讶地）发现，与其他策略（如基于重启的算法）相比，当前结果存在不自然的差距，而且遗憾分析从 GLB 过渡到线性赌博机时也显得不自然。

这些不自然的现象促使我们重新审视非平稳参数化老虎机加权策略的算法设计与遗憾分析 [Russac 等人，2019，2021，Faury 等人，2021]。事实上，关键在于基于权重的估计量的估计误差分析，通常将其分解为两部分——一部分是由参数漂移导致的偏差部分，另一部分是由随机噪声导致的方差部分。通常，偏差部分通过非平稳策略加以控制，而方差部分则通过精心设计的集中不等式加以处理。Russac 等人 [2019] 首次分析了线性赌博机的基于权重的算法，并在分析中引入了虚拟窗口大小，以控制偏差，从而模仿滑动窗口策略的分析 [Cheung 等人，2019]。因此，他们不得不使用不同的局部范数来控制偏差部分和方差部分，导致了意料之外的低效与复杂性。对于线性赌博机，这催生了 D-LinUCB 算法 [Russac 等人，2019] 该算法需要维护一个额外的协方差矩阵，其效率低于基于窗口和基于重启的算法 [Cheung 等人，2019，Zhao 等人，2020]。

这一加权策略的分析框架在 GLB 中引发了更严峻的问题，因为 GLB 的结构更加丰富和复杂。具体而言，Faury 等人 [2021]研究了漂移广义线性赌博机，并设计了一种高度复杂的投影操作，以控制偏差部分和方差部分，沿用了Russac 等人 [2019]的做法来模拟滑动窗口分析，最终获得了一个$(d^{9/10}P_T^{1/5}T^{4/5})$动态遗憾界。然而，这种方法无法恢复$\widetilde{\mathcal{O}}(d^{7/8}P_T^{1/4}T^{3/4})$ 漂移线性赌博机（GLB 的一个特例）中基于权重的算法所享有的界 [Russac 等人，2019]。随后，Russac 等人 [2021] 研究了非平稳自协调老虎机（SCB）。SCB 是 GLB 的一个子类，具有许多吸引人的结构。由于技术上的困难，无法用传统分析方法界定偏差，因此他们只能在分段平稳设定下进行分析，而未能处理更具挑战性的漂移设定。因此，他们在论文中提出了两个开放问题：（i）如何将基于权重的算法扩展到漂移 SCB；以及（ii）如何将平稳 SCB 中的近期进展 [Abeille 等人，2021] 加以复现，以改善c_μ 非平稳 SCB 对。

我们的结果。 本文重新审视了非平稳参数化老虎机的加权策略。我们发现，早先的分析框架可能并不恰当，因为它模仿了滑动窗口分析，要求使用不同的 局部范数来界定偏差部分和方差部分。我们为加权策略提出了一个精细化分析框架，并针对偏差部分给出了一种新的分析方法，因此现在可以使用相同的 局部范数来分析偏差部分和方差部分。这个精细化分析框架不仅简化了遗

憾分析，还为算法设计带来了诸多益处，包括提升线性赌博机的效率，以及解决 GLB 和 SCB 中出现的投影问题。表 1 总结了我们的主要结果，并与基于权重的算法此前已知的最佳结果进行了比较。具体而言，基于我们精细化的分析框架，我们取得了以下结果：(i) 对于线性赌博机，我们的方法只需维护一个协方差矩阵而非两个，同时仍能达到与 [Russac 等人， 2019]；(ii) 对于 GLB，我们的方法获得了一个 $\widetilde{\mathcal{O}}(k_\mu^{5/4}c_\mu^{-3/4}d^{3/4}P_T^{1/4}T^{3/4})$ 遗憾界，其 d， P_T和 T的阶与线性赌博机情形相同；(iii) 对于 SCB，我们获得了一个 $\widetilde{\mathcal{O}}(k_\mu^{5/4}c_\mu^{-1/2}d^{3/4}P_T^{1/4}T^{3/4})$ 遗憾界。此外，对于分段平稳 SCB，我们的方法获得了一个 $\widetilde{\mathcal{O}}(d^{2/3}\Gamma_T^{1/3}T^{2/3})$ 遗憾界，能够消除 c_μ^{-1}的影响，从而解决了 Russac 等人 [2021] 提出的第二个开放问题。

2 相关工作

线性多臂老虎机。非平稳线性赌博机问题最早由 Cheung 等人 [2019] 研究。他们建立了 $\Omega(d^{2/3}P_T^{1/3}T^{2/3})$ 极小极大下界，并提出了基于滑动窗口策略的 SW-UCB 算法。随后，Russac 等人 [2019] 提出了基于加权策略的 D-LinUCB 算法，Zhao 等人 [2020] 提出了基于重启策略的 RestartUCB 算法。请注意，这三项工作证明了 $\widetilde{\mathcal{O}}(d^{2/3}P_T^{1/3}T^{2/3})$ 遗憾界，但正如 Zhao 和 Zhang [2021] 所指出的，遗憾分析中存在一个细微的技术漏洞。在修正这一技术漏洞后，上述三个算法均达到了 $\widetilde{\mathcal{O}}(d^{7/8}P_T^{1/4}T^{3/4})$ 遗憾界 [Zhao 和 Zhang，2021，Zhao 等人，2021b]。然而，为了实现这一结果，这三个算法都要求在算法实现开始时将路径长度 P_T 作为输入，这并不理想。为了解决这一问题，Cheung 等人 [2019] 提出了赌博机之上的赌博机（BOB）策略，将其作为一种元算法来学习未知参数 P_T，该策略可以与上述算法结合，从而消除对这一先验知识的要求。随后，Wei 和 Luo [2021] 提出了 MASTER 算法，该算法具有理论上最优的 $\widetilde{\mathcal{O}}(\min\{d\sqrt{\Gamma_T T}, dP_T^{1/3}T^{2/3}\})$ 遗憾界，同时也无需事先知道环境的非平稳程度（即，Γ_T 和 P_T）。最近，非平稳（残性）赌博机领域也取得了一些新进展 [Liu 等，2022，Suk 和 Kpotufe，2022，Abbasi-Yadkori 等人，2022，Clerici 等人， 2023]。此外，Cheung 等人 [2022, Remark 4] 通过重启（高效的）对抗性线性赌博机算法，绕过了上述技术漏洞。然而，需要注意的是，这一方法仅适用于线性赌博机，并且仍然要求预先知道 P_T。

广义线性老虎机。 GLB 问题最早由 Filippi 等人 [2010] 提出。他们提出了 GLM-UCB 算法，实现了 $\widetilde{\mathcal{O}}(k_\mu c_\mu^{-1}d\sqrt{T})$ 遗憾界，其中 k_μ、 c_μ 是与问题相关的常数， k_μ/c_μ 表示广义线性模型的非线性程度。Faury 等人 [2021] 将平稳 GLB 扩展到漂移情形，并提出了具有 $\widetilde{\mathcal{O}}(k_\mu^2 c_\mu^{-1}d^{9/10}P_T^{1/5}T^{4/5})$ 遗憾界的 BVD-GLM-UCB 算法。Faury 等人 [2020] 研究了 GLB 的一个特定实例，即逻辑斯蒂老虎机（LogB）。他们首先指出，在 GLB 设置下，与问题相关的常数 $1/c_\mu$ 在某些情形下可能非常大，例如 LogB；随后，他们提出了具有$\widetilde{\mathcal{O}}(c_\mu^{-1/2}d\sqrt{T})$ 遗憾界的 Logistic-UCB-1 算法，以及具有 $\widetilde{\mathcal{O}}(d\sqrt{T}+c_\mu^{-1})$ 遗憾界的 Logistic-UCB-2 算法。随后，Abeille 等人 [2021] 为逻辑回归老虎机建立了 $\Omega(d\sqrt{\mu'(X_*^\top\theta_*)T})$ 遗憾下界，并提出了最优算法 OFULog。Russac 等人 [2021] 将逻辑回归老虎机推广到自协调老虎机，并考虑了分段平稳情形；他们的算法具有 $\widetilde{\mathcal{O}}(c_\mu^{-1/3}d^{2/3}\Gamma_T^{1/3}T^{2/3})$ 遗憾界。为处理 P_T-未知的情形，Faury 等人 [2021] 通过将 BVD-GLM-UCB 与 BOB 策略相结合，提出了一种无参数算法，但最终结果仍然不是最优的。与此同时，Wei 和 Luo 提出的最优黑盒算法 [2021]还可以自适应地重启 Filippi 等人 [提出的平稳 GLM-UCB 算法，并达到最优的 $\widetilde{\mathcal{O}}(\min\{k_\mu c_\mu^{-1}\sqrt{\Gamma_T T}, k_\mu^{4/3}c_\mu^{-1}dP_T^{1/3}T^{2/3}\})$ 遗憾。

3 线性多臂老虎机

本节首先介绍非平稳线性赌博机的问题设定，然后描述我们的 LB-WeightUCB 算法及其理论保证。我们的算法具有与目前已知最佳的基于权重的算法相同的遗憾界 [Russac 等人， 2019] 但效率更高。

3.1 问题设定

在每一轮 t，学习者从可行集合 $\mathcal{X}\subseteq\mathbb{R}^d$ 中选择一条臂 X_t，并获得奖励 r_t，满足

$$r_t = X_t^\top\theta_t + \eta_t, \tag{1}$$

其中， $\theta_t\in\mathbb{R}^d$ 是未知的时变参数，η_t 是 R-次高斯噪声。学习者的目标是最小化以下（伪）动态遗憾：

$$R_T = \sum_{t=1}^{T}\max_{\mathbf{x}\in\mathcal{X}}\mathbf{x}^\top\theta_t - \sum_{t=1}^{T}X_t^\top\theta_t, \tag{2}$$

这是相对于最优策略的累积遗憾，该策略掌握未知参数的全部信息。这里我们考虑漂移情形，并使用路径长度 $P_T=\sum_{t=2}^{T}\|\theta_{t-1}-\theta_t\|_2$ 作为非平稳性度量。

我们采用以下标准有界性假设 [Abbasi-Yadkori 等人，2011，Cheung 等人， 2019，Russac 等人，2019， Zhao 等人，2020]。

假设1。 假设可行集和未知参数均有界： $\forall \mathbf{x} \in \mathcal{X}$, $\|\mathbf{x}\|_2 \leq L$, 并且$\theta_t \in \Theta$ 对所有 $t \in [T]$ 均成立，其中 $\Theta \triangleq \{\theta \mid \|\theta\|_2 \leq S\}$。

3.2 算法与遗憾保证

我们提出了 LB-WeightUCB 算法，该算法在实现与先前方法相同保证的同时，效率更高。我们首先给出所采用的估计量，然后通过改进的分析框架推导其估计误差上界，这是算法设计与遗憾分析的关键。基于估计误差上界，我们提出选择准则，并最终给出其动态遗憾的理论保证。

估计量。 我们采用加权正则化最小二乘估计器，与 D-LinUCB [Russac 等人的相同， 2019], 该估计量 θ_t 是以下问题的解:

$$\min_{\theta} \lambda \|\theta\|_2^2 + \sum_{s=1}^{t-1} \gamma^{t-s-1} \left(X_s^\top \theta - r_s\right)^2, \tag{3}$$

其中 $\lambda > 0$ 是正则化系数， $\gamma \in (0,1)$ 是折扣因子。显然， θ_t 存在闭式解 $\widehat{\theta}_t = V_{t-1}^{-1}(\sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s)$, 其中 $V_t = \lambda I_d + \sum_{s=1}^{t} \gamma^{t-s} X_s X_s^\top$, $V_0 = \lambda I_d$ 是协方差矩阵。注意，该闭式解还可以进一步转化为递推公式 [Haykin， 2002，第 10 章。3]。这样便可以在线更新，而无需存储历史数据；与滑动窗口策略相比，这也是加权策略的一项重要计算优势。

上置信界。对于估计量（3），我们给出如下估计误差界。值得注意的是，该结果不同于先前的结果 [Russac 等人， 2019，附录 B.3，定理证明中的第二步和第三步 2],是构成更高效算法的关键要素，稍后将对此进行解释。

引理 1。对于任意$\mathbf{x} \in \mathcal{X}, \gamma \in (0, 1)$ 和 $\delta \in (0, 1)$，以至少 $1-\delta$的概率，以下结论对所有 $t \in [T]$均成立

$$\left|\mathbf{x}^\top(\widehat{\theta}_t - \theta_t)\right| \leq L^2 \sqrt{\frac{d}{\lambda}} \sum_{p=1}^{t-1} \gamma^{\frac{t-1}{2}} \sqrt{\frac{\gamma^{-p}-1}{1-\gamma}} \|\theta_p - \theta_{p+1}\|_2 + \beta_{t-1} \|\mathbf{x}\|_{V_{t-1}^{-1}}, \tag{4}$$

其中 β_t 是所设置信心区域的半径

$$\beta_t = \sqrt{\lambda} S + R\sqrt{2\log\frac{1}{\delta} + d\log\left(1 + \frac{L^2(1-\gamma^{2t})}{\lambda d(1-\gamma^2)}\right)}. \tag{5}$$

引理1的证明见附录A.2。根据引理1，我们可以将选臂准则指定为

$$X_t = \arg\max_{\mathbf{x} \in \mathcal{X}} \left\{\langle \mathbf{x}, \widehat{\theta}_t \rangle + \beta_{t-1} \|\mathbf{x}\|_{V_{t-1}^{-1}}\right\}. \tag{6}$$

整体算法总结于算法1中。从

算法 1LB-WeightUCB

输入： 时间范围 T、折扣因子 γ、置信度 δ, 正则化项 λ, 参数 S、 L 和 R

1: 设置 $V_0 = \lambda I_d$, $\widehat{\theta}_1 = \mathbf{0}$ 并计算 β_0 根据（5）
2: **对于** $t = 1, 2, ..., T$ **do**
3: 选择 $X_t = \arg\max_{\mathbf{x} \in \mathcal{X}} \{\langle \mathbf{x}, \widehat{\theta}_t \rangle + \beta_{t-1} \|\mathbf{x}\|_{V_{t-1}^{-1}}\}$
4: 获得奖励 r_t
5: 更新 $V_t = \gamma V_{t-1} + X_t X_t^\top + (1-\gamma)\lambda I_d$
6: 计算 θ_{t+1}，通过 (3) 以及 β_t，通过 (5)
7: **结束循环**

从算法1第5行的更新过程可以观察到，我们的算法只需维护一个单一的协方差矩阵 $V_{t-1} \in \mathbb{R}^{d \times d}$。相比之下， Russac 等人 [2019]提出的算法，其选择准则类似

$$X_t = \arg\max_{\mathbf{x} \in \mathcal{X}} \left\{\langle \mathbf{x}, \widehat{\theta}_t \rangle + \beta_{t-1} \|\mathbf{x}\|_{V_{t-1}^{-1} \widetilde{V}_{t-1} V_{t-1}^{-1}}\right\},$$

其中， β_{t-1}， V_{t-1}^{-1} 与我们的选择准则（6）中的相同, $\widetilde{V}_{t-1} = \lambda I_d + \sum_{s=1}^{t-1} \gamma^{2(t-s-1)} X_s X_s^\top \in \mathbb{R}^{d \times d}$ 是一个额外的协方差矩阵。因此，我们的算法比他们的算法更高效，因为它只需维护一个协方差矩阵，而非两个。这是因为引理 1 的分析仅使用 V_{t-1}^{-1}作为局部范数来分析偏差部分和方差部分，而 Russac 等人 [2019] 的算法则要求分别使用 l_2- 范数和 $V_{t-1}^{-1}\widetilde{V}_{t-1}V_{t-1}^{-1}$-范数来控制偏差部分和方差部分。在第 6节中，我们简要介绍了引理1 的分析框架，更详细的讨论见附录 A.1。此外，我们证明了我们的算法具有与 d算法相同（甚至略优）的遗憾Russac 等人 [2019]。

定理 1。 对于所有 $\gamma \in (1/T, 1)$, $\lambda = d$, *LB-WeightUCB*（算法 *1*）的动态遗憾以至少 $1 - 1/T$的概率有界为

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

此外，通过将折扣因子最优地设置为$\gamma = 1 - \max\{1/T, \sqrt{P_T/(dT)}\}$, *LB-WeightUCB* 可保证

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

与先前工作相比 [Cheung 等人、2019、 Russac 等人、2019以及Zhao 等人， 2020], 我们的方法将$\widetilde{\mathcal{O}}$（$d^{7/8} P_T^{1/4} T^{3/4}$）改进为$\widetilde{\mathcal{O}}$（$d^{3/4} P_T^{1/4} T^{3/4}$），当 $P_T \geq d/T$。我们指出，这种维度依赖性的改进仅仅源于对折扣因子进行比 [Russac 等人所采用的方法更为精细的调节， 2019], 后者在调节时没有考虑维度因素。他们的算法及遗憾界也同样可以从这种精细调节中获益。定理1 的证明见附录A.3。

此外，请注意，折扣因子的最优选择γ需要预先知道P_T。为了在未知 P_T 的情况下获得与参数无关的结果，我们的算法可以与 BOB 策略 [Cheung 等人, 2019] 相结合，并达到 $\widetilde{\mathcal{O}}\left(d^{3/4}P_T^{1/4}T^{3/4}\right)$ 的界。然而，这一界并非最优；基于我们的结果，沿用 Wei 和 Luo [2021], 的思路，可以设计一种自适应加权算法，进一步在不了解 P_T的情况下实现最优动态遗憾。这极具挑战性，因为在每一轮 $t \in [T]$中，我们只能接收一个数据对（X_t， r_t），这不足以让学习者实时更新折扣因子 γ_t。与此同时，MASTER 算法 [Wei 和 Luo， 2021]可以被视为自适应加权算法的一个特例，因为它只包含两种情形：将 $\gamma_t = 0$ 设置为在时间 t 重新启动，以及将 $\gamma_t = 1$ 设置为继续运行。但对于自适应加权算法而言，折扣因子 γ_t 在$[0, 1]$, 中的选择可以是连续的，这比二元决策更加困难。我们将其留作未来研究中的一个重要开放问题。

4 广义线性老虎机

本节将加权策略应用于漂移广义线性赌博机。与目前已知的漂移广义线性赌博机最佳基于权重的算法相比 [Faury 等人, 2021], 我们的算法更简单，同时具有更好的理论保证。

4.1 问题设定

广义线性老虎机假设存在一个逆链接函数 $\mu: \mathbb{R} \to \mathbb{R}$，使得 $r_t = \mu(X_t^\top \theta_t) + \eta_t$，其中 $\theta_t \in \mathbb{R}^d$ 是未知参数，且可能随时间变化。与线性赌博机类似，我们将广义线性老虎机的动态遗憾定义如下：

$$R_T = \sum_{t=1}^{T} \left(\max_{\mathbf{x} \in \mathcal{X}} \mu(\mathbf{x}^\top \theta_t) - \mu(X_t^\top \theta_t) \right). \tag{7}$$

在广义线性老虎机设定下，我们采用与线性赌博机相同的假设，包括 R-次高斯噪声、可行集和未知回归参数的有界性（假设1）。此外，我们还采用逆链接函数的标准有界性假设 [Fil- ippi 等人，2010年，Li等人，2017，Faury 等人， 2021]。

假设2。 逆链接函数 $\mu: \mathbb{R} \to \mathbb{R}$ 是k_μ-利普希茨连续的，并且连续可微，满足

$$c_\mu \triangleq \inf_{\{\theta \in \Theta, \mathbf{x} \in \mathcal{X}\}} \mu'(\theta^\top \mathbf{x}) > 0, \quad \Theta = \{\theta \mid \|\theta\|_2 \leq S\}.$$

先前工作 [Zhao 等人、2020、Cheung 等人， 2022] 定义了一个类似的参数 $\widetilde{c}_\mu \triangleq \inf_{\{\theta \in \mathbb{R}^d, \mathbf{x} \in \mathcal{X}\}} \mu'(\theta^\top \mathbf{x}) > 0$，并得到了随 $1/\widetilde{c}_\mu$变化的遗憾上界。显然， $\widetilde{c}_\mu$ 比我们定义的 c_μ 更小（而且可能小得多），因为 c_μ 定义在 Θ 上，而 $\widetilde{c}_\mu$ 定义在 $\mathbb{R}$上。因此， $\widetilde{c}_\mu$ 不太适合出现在上界中。

算法 2GLB-WeightUCB

输入： 时间范围 T、折扣因子 γ、置信度 δ、正则化项 λ、逆链接函数 μ、参数 S、 L 以及 R

设置 $V_0 = \lambda I_d$ $\widehat{\theta}_1 = \mathbf{0}$ 并计算 k_μ 和 c_μ
对于 $t = 1, 2, ..., T$ **do**
  如果 $\|\theta_t\|_2 \leq S$ **则**
    令 $\widetilde{\theta}_t = \theta_t$
  else
    $\widetilde{\theta}_t$进行投影并得到 θ_t 由 (9)
  结束如果
  计算 β_{t-1} 依据（11）
  选择 X_t 依据（12）
  接收奖励 r_t
  更新 $V_t = \gamma V_{t-1} + X_t X_t^\top + (1-\gamma)\lambda I_d$
  计算 θ_{t+1} 依据（8）
结束循环

4.2 算法与遗憾保证

我们提出了 GLB-WeightUCB，与先前的基于权重的算法相比，它更简单，且具有更好的理论保证 [Faury 等人， 2021]。关键改进得益于我们精细化的分析框架，该框架与简单的投影步骤兼容。

估计量。 在第 t 次迭代中，我们首先采用准最大似然估计（QMLE），暂不考虑向可行域的投影。具体而言，$\widehat{\theta}_t$估计量 θ_t 是以下加权正则化估计方程的解：

$$\lambda c_\mu \theta + \sum_{s=1}^{t-1} \gamma^{t-s-1} \left(\mu(X_s^\top \theta) - r_s \right) X_s = 0. \tag{8}$$

鉴于 $\widehat{\theta}_t$ 可能不属于可行集 Θ，且 c_μ 是在参数 $\theta \in \Theta$上定义的，我们需要执行以下投影步骤

$$\widetilde{\theta}_t = \arg\min_{\theta \in \Theta} \|g_t(\widehat{\theta}_t) - g_t(\theta)\|_{V_{t-1}^{-1}}, \tag{9}$$

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

$$g_t(\theta) \triangleq \lambda c_\mu \theta + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta) X_s. \tag{10}$$

然而，先前工作 [Faury 等人， 2021]在漂移情形下，无法像平稳 GLB 或分段平稳 GLB 那样进行同样简单的投影，因为它们使用不同的局部范数分别度量偏差部分和方差部分，以进行估计误差分析。因此，它们必须设计一种复杂的投影，以确保偏差部分和方差部分能够由不同的局部范数度量（见 [Faury 等人, 2021，第 4 节。1], 以及我们在附录中的重述B.1）。

我们改进后的分析框架与这种投影操作兼容，这得益于我们的分析框架对偏差部分和方差部分采用了相同的局部范数。

上置信界。 对于估计量（8）及其投影（9），我们构造如下估计误差界。

引理 2。 对于任意 $\mathbf{x} \in \mathcal{X}$、$\gamma \in (0, 1)$ 以及 $\delta \in (0, 1)$，以至少 $1-\delta$ 的概率，对于所有 $t \in [T]$，以下结论成立

$$\left|\mu(\mathbf{x}^\top \widetilde{\theta}_t) - \mu(\mathbf{x}^\top \theta_t)\right| \leq \frac{2k_\mu}{c_\mu}\left(\sum_{p=1}^{t-1} C(p)\left\|\theta_p - \theta_{p+1}\right\|_2 + \bar{\beta}_{t-1}\left\|\mathbf{x}\right\|_{V_{t-1}^{-1}}\right),$$

其中，$C(p) = k_\mu L^2\sqrt{\frac{d}{\lambda}}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}$，且 $\bar{\beta}_t$ 是由以下方式设定的置信区域半径

$$\bar{\beta}_t = \sqrt{\lambda}c_\mu S + R\sqrt{2\log\frac{1}{\delta} + d\log\left(1 + \frac{L^2(1-\gamma^{2t})}{\lambda d(1-\gamma^2)}\right)}. \tag{11}$$

引理 2 的证明见附录 B.2。随后，基于引理 2，我们可以将臂选择准则具体表示为

$$X_t = \arg\max_{\mathbf{x}\in\mathcal{X}}\left\{\mu(\mathbf{x}^\top\widetilde{\theta}_t) + \frac{2k_\mu}{c_\mu}\bar{\beta}_{t-1}\left\|\mathbf{x}\right\|_{V_{t-1}^{-1}}\right\}. \tag{12}$$

整体算法总结于算法 m 2.

请注意，估计方程（8）和置信半径（11）与Faury 等人 [2021]的算法 1 中所使用的相同。但重要的是，这两种方法最终得到的（投影后）估计量存在显著差异。借助更简单的投影操作和我们改进后的分析框架，我们可以立即为基于权重的算法获得更优的遗憾保证。

定理 2。 对于所有 $\gamma \in (1/T, 1)$，$\lambda = d/c_\mu^2$，*GLB-WeightUCB*（算法 2）的遗憾以至少 $1 - 1/T$的概率受到如下界的约束：

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

通过最优地设置 $\gamma = 1-\max\{1/T, \sqrt{k_\mu c_\mu P_T/(dT)}\}$，*GLB-WeightUCB* 可实现如下动态遗憾：

$$R_T \leq \begin{cases}\widetilde{\mathcal{O}}\left(\frac{k_\mu^{5/4}}{c_\mu^{3/4}}d^{3/4}P_T^{1/4}T^{3/4}\right) & \textit{when } P_T \geq \frac{d}{k_\mu c_\mu T},\\ \widetilde{\mathcal{O}}\left(\frac{k_\mu}{c_\mu}d\sqrt{T}\right) & \textit{when } 0 \leq P_T < \frac{d}{k_\mu c_\mu T}.\end{cases}$$

与 BVD-GLM-UCB（针对漂移广义线性赌博机已知的最佳基于权重的算法）相比，[Faury 等人，2021], 关注对 d的依赖关系，P_T、以及 T，我们可以看到，我们的方法将遗憾从 $\widetilde{\mathcal{O}}(d^{9/10}P_T^{1/5}T^{4/5})$ 降低至 $\widetilde{\mathcal{O}}(d^{3/4}P_T^{1/4}T^{3/4})$。此外，我们的结果在依赖关系方面也优于他们的结果，从c_μ 降至 c_μ^{-1}再到 $c_\mu^{-3/4}$。

5 自协调老虎机

本节研究自协调老虎机（SCB），这是 GLB 的一个重要子类，具有许多吸引人的结构。

5.1 问题设定

对于 SCB，奖励的分布属于规范指数族：$\mathbb{P}_\theta[r|\mathbf{x}] = \exp(r\mathbf{x}^\top\theta - b(\mathbf{x}^\top\theta) + c(r))$，其中 $b(\cdot)$ 是二阶连续可微函数，而 $c(\cdot)$ 是一个实值函数。由于指数族的良好性质，我们有 $\mathbb{E}[r \mid \mathbf{x}] = b'(\mathbf{x}^\top\theta)$ 和 $\mathrm{Var}[r \mid \mathbf{x}] = b''(\mathbf{x}^\top\theta)$，其中 b' 表示函数 b 的一阶导数，b'' 表示其二阶导数。于是，我们可以引入（逆）链接函数 $\mu(\cdot) \triangleq b'(\cdot)$，使得在 $t \in [T]$ 时满足以下关系

$$\mathbb{E}[r_t \mid X_t] = \mu(X_t^\top\theta_t), \mathrm{Var}[r_t \mid X_t] = \mu'(X_t^\top\theta_t). \tag{13}$$

SCB 要求链接函数满足 $|\mu''| \leq \mu'$，这通常称为广义自协调性质。我们进一步引入记号 $\eta_t = r_t - \mu(X_t^\top\theta_t)$ 表示噪声。SCB 成功地对许多重要的现实应用进行建模，并捕捉奖励结构。例如，选择 $\mu(x) = (1+e^{-x})^{-1}$ 可得到逻辑赌博机（LogB），它常用于对推荐系统中的二元反馈奖励进行建模 [Zhang 等人，2016, Jun 等人，2017，Dong 等人，2019]。

我们采用与线性赌博机和 GLB 相同的若干标准假设，包括可行集和未知回归参数的有界性（假设 1），以及链接函数上的非线性度量（假设2）。此外，与Russac 等人 [2021], 类似，我们还需要对奖励的有界性作出假设。为便于分析，我们令$L = 1$ 这意味着 $\|\mathbf{x}\|_2 \leq 1$ 对所有 $\mathbf{x} \in \mathcal{X}$。

假设 3。 每轮获得的奖励满足$0 \leq r_t \leq m$，对所有 $t \in [T]$ 以及某个常数 $m > 0$均成立。

5.2 算法与遗憾保证

我们提出了 SCB-WeightUCB 算法。与 GLB 相比，我们在投影和遗憾分析中采用了一种新的局部范数，这是改进 c_μ^{-1} 阶数的关键。

估计量。 在迭代 t时，我们首先采用与 GLB 相同的极大似然估计量，其定义见（8）。与 GLB 不同的是，此处我们使用一种新的局部范数，将点投影到可行集 Θ上，

$$\widetilde{\theta}_t = \arg\min_{\theta\in\Theta}\left\|g_t(\widehat{\theta}_t) - g_t(\theta)\right\|_{H_t^{-1}(\theta)}, \tag{14}$$

其中 $g_t(\theta)$ 与（10）相同，而 $H_t(\theta)$ 定义为

$$H_t(\theta) \triangleq \lambda c_\mu I_d + \sum_{s=1}^{t-1}\gamma^{t-s-1}\mu'(X_s^\top\theta)X_sX_s^\top. \tag{15}$$

值得注意的是，与 V_t 相比，$H_t(\theta)$ 取决于函数

沿着动态过程的曲率，因此能够捕捉更多局部 信息。将此投影步骤与定理 6 中重新表述的加权版本自归一化集中不等式相结合，将消除遗憾界中的一个常数 $c_\mu^{-1/2}$。

上置信界。 对于估计量 (8) 及其投影 (14，我们构造如下估计误差界。

引理3。 对于任意 $\mathbf{x} \in \mathcal{X}$, $\gamma \in (0,\ 1)$ 以及 $\delta \in (0,\ 1)$，以至少 $1-\delta$ 的概率，以下结论对所有 $t \in [T]$ 均成立

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

其中 $C(p) = L^2\sqrt{\frac{d}{\lambda}}\frac{k_\mu}{\sqrt{c_\mu}}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}$，且 $\widetilde{\beta}_t$ 是置信区域的半径，由下式确定

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

引理的证明3 见附录 C.1。基于引理 3，我们可以将选臂准则具体表述为

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

我们的 SCB 算法（命名为 SCB-WeightUCB）遵循算法 2的相同流程，区别在于$\widetilde{\ }$ $\widetilde{\ }$由 θ_t 根据 (14) 计算， β_{t-1} 根据 (16) 计算，且X_t 根据 (17) 计算。此外，我们有如下关于 SCB-WeightUCB 算法的保证。

定理3。 对于所有 $\gamma \in (1/T,\ 1)$、 $\lambda = d\log(T)/c_\mu$, *SCB-WeightUCB* 的动态遗憾以至少 $1-1/T$ 的概率有如下界:

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

令 $\gamma = 1 - \max\{1/T, \sqrt{k_\mu P_T/(dT)}\}$，则可得

$$R_T \leq \begin{cases}\widetilde{\mathcal{O}}\left(\frac{k_\mu^{5/4}}{c_\mu^{1/2}}d^{3/4}P_T^{1/4}T^{3/4}\right) & \textit{when } P_T \geq \frac{d}{k_\mu T},\\ \widetilde{\mathcal{O}}\left(\frac{k_\mu}{c_\mu^{1/2}}d\sqrt{T}\right) & \textit{when } 0 \leq P_T < \frac{d}{k_\mu T}.\end{cases}$$

与 GLB 相比，通过利用自协调性质，我们将 c_μ 的阶从 c_μ^{-1}改进为$c_\mu^{-1/2}$。同时，在近似平稳环境中（P_T 足够小），我们的结果可以恢复到 LogUCB1 算法的性能 [Faury 等人， 2020]。定理 3 的证明见附录C.2。

此外，对于分段平稳 SCB，我们提出了能够摆脱.....影响的 SCB-PW-WeightUCB算法c_μ，从而直接改进了 [Russac 等人的 2021]。

定理 4。 对于所有 $\gamma \in (1/2,\ 1)$、 $D = \log(T)/\log(1/\gamma)$以及 $\lambda = d\log(T)/c_\mu$，*SCB-PW-WeightUCB* 的遗憾以至少 $1-1/T$ 的概率有如下界:

$$R_T \leq \widetilde{\mathcal{O}}\left(\frac{1}{1-\gamma}\Gamma_T + \frac{1}{\sqrt{1-\gamma}} + d\sqrt{(1-\gamma)T}\right).$$

通过设置 $\gamma = 1 -$ 最大值$\{1/T, (\Gamma_T/(dT))^{2/3}\}$，我们得到

$$R_T \leq \begin{cases}\widetilde{\mathcal{O}}\left(d^{2/3}\Gamma_T^{1/3}T^{2/3}\right) & \textit{when } \Gamma_T \geq d/\sqrt{T},\\ \widetilde{\mathcal{O}}\left(d\sqrt{T}\right) & \textit{when } 0 \leq \Gamma_T < d/\sqrt{T}.\end{cases}$$

完整的算法和分析见附录D。

6 精细分析框架

本节给出引理 1 （加权线性老虎机的估计误差分析）的证明概要，同时也介绍我们提出的分析框架。

证明概要。根据模型假设（1）和估计量（3），估计误差可以分为两部分:

$$\widehat{\theta}_t - \theta_t = \underbrace{V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}X_sX_s^\top(\theta_s - \theta_t)\right)}_{\texttt{bias part}} + \underbrace{V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_sX_s - \lambda\theta_t\right)}_{\texttt{variance part}},$$

其中，偏差部分由参数漂移引起，而方差部分则源于随机噪声。接着，根据柯西-施瓦茨不等式，对于任意$\mathbf{x} \in \mathcal{X}$,

$$|\mathbf{x}^\top(\widehat{\theta}_t - \theta_t)| \leq \|\mathbf{x}\|_{V_{t-1}^{-1}}(A_t + B_t), \tag{18}$$

where $A_t = \|\sum_{s=1}^{t-1}\gamma^{t-s-1}X_sX_s^\top(\theta_s - \theta_t)\|_{V_{t-1}^{-1}}$ and $B_t = \|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_sX_s - \lambda\theta_t\|_{V_{t-1}^{-1}}$.

为（18）选择合适的局部范数，是简化并改进估计误差分析的关键。我们注意到，先前的分析 [Russac 等人， 2019] 不得不使用 不同的 局部范数：在偏差部分使用 l_2-范数，在方差部分使用 $V_{t-1}^{-1}\widetilde{V}_{t-1}V_{t-1}^{-1}$-范数，即:

$$|\mathbf{x}^\top(\widehat{\theta}_t - \theta_t)| \leq \|\mathbf{x}\|_2 A_t' + \|\mathbf{x}\|_{V_{t-1}^{-1}\widetilde{V}_{t-1}V_{t-1}^{-1}}B_t', \tag{19}$$

其中 $A_t' = \|V_{t-1}^{-1}\sum_{s=1}^{t-1}\gamma^{t-s-1}X_sX_s^\top(\theta_s - \theta_t)\|_2$和 $B_t' = \|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_sX_s - \lambda\theta_t\|_{\widetilde{V}_{t-1}^{-1}}$ 以及 $\widetilde{V}_t = \lambda I_d + \sum_{s=1}^{t}\gamma^{2(t-s)}X_sX_s^\top$。由于需要使用滑动窗口分析来分析偏差部分，因此他们必须使用 l_2-范数来得到 A_t' 的形式。对于方差

部分，为了使用加权版本的自归一化集中不等式（定理5），他们使用 $V_{t-1}^{-1}\widetilde{V}_{t-1}V_{t-1}^{-1}$-范数来控制 $\mathbf{x}$ 项，从而使 B_t' 项可以通过 $\widetilde{V}_{t-1}^{-1}$进行范数化。作为改进，我们可以直接使用 同一个 V_{t-1}^{-1}-范数来控制这两个部分，这得益于我们对偏差部分的新分析以及对方差部分的改进分析。

偏差部分分析。 第一步是提取底层参数的变化，如下所示，

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

项 $\sum_{s=1}^{p}\gamma^{t-s-1}\left\|X_s\right\|_{V_{t-1}^{-1}}$ 应能够进一步推导出折扣因子 γ 的表达式，从而控制变化项。经过一些推导，我们得到

$$A_t \leq L\sqrt{d}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2,$$

其中，变化项仅受折扣因子 γ 控制，而非虚拟窗口大小。

方差部分分析。 根据 V_t 和 $\sim$ $\sim V_t$的定义，我们可以发现 $V_t \succeq V_t$，因此有 $B_t \leq B_t' \leq \beta_{t-1}$其中 β_{t-1} 是置信半径（5），第二个不等式由定理5得出。因此，我们保留相同的置信界 β_{t-1}，但在进行臂选择时，只需计算 V_{t-1}^{-1}，而非$V_{t-1}^{-1}\widetilde{V}_{t-1}V_{t-1}^{-1}$。

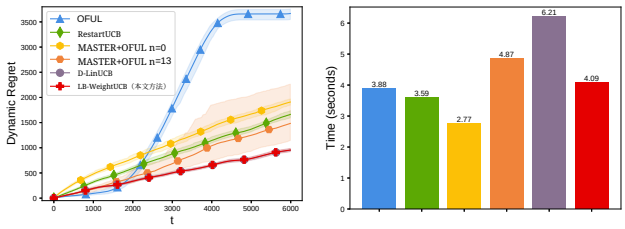
结合偏差部分和方差部分的分析，我们可以完成引理1的证明。 □

注1。我们分析框架中的关键步骤（18）也解决了 GLB 中的投影问题。具体而言，在投影步骤之后，只能在 V_{t-1}^{-1}-范数下进行偏差-方差分解。为了适应先前的分析（19），Faury 等人 [2021] 不得不在算法中加入一个高度复杂的投影操作，而我们的框架由于偏差部分和方差部分使用相同的 V_{t-1}^{-1}-范数，已经满足了这一条件。

7 实验

本节进一步通过实证分析考察我们提出的算法的性能。我们分别针对漂移线性赌博机和 GLB 展示了两个合成实验。对于每个实验，我们将特征空间的维度设为 $d=2$，轮数设为 $T=6000$，臂数设为 $n=50$。每个臂的特征均从正态分布 $\mathcal{N}(0,1)$ 中采样，随后重新缩放以满足 $L=1$。我们将时变参数 θ_t 初始化为 $[1,0]$，并使其沿单位圆均匀逆时针旋转，在 T 轮内从 0 旋转一周至 2π，最终回到起点

$[1,0]$。



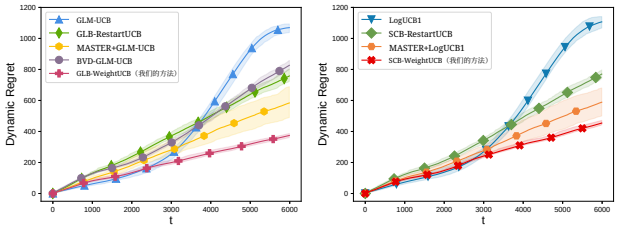
（a）线性赌博机：累积遗憾　　（b）线性赌博机：平均运行时间

图 1：线性多臂老虎机实验。

7.1 线性多臂老虎机

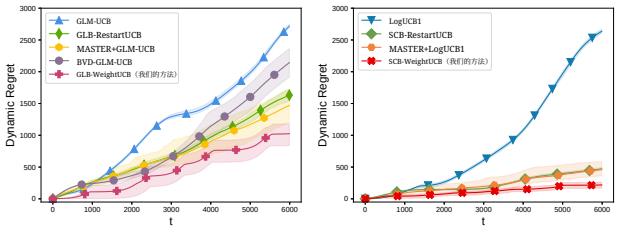
设定。我们考虑线性模型 $r_t = X_t^\top\theta_t+\eta_t$，其中随机噪声 η_t 在每个时刻 t 独立地服从正态分布 $\mathcal{N}(0,1)$。我们将所提出的 LB-WeightUCB 算法与以下算法进行性能比较：(a) 静态算法 OFUL [Abbasi-Yadkori 等人，2011]；(b) 基于重启的算法 RestartUCB [Zhao 等人，2020]；(c) 基于权重的算法 D-LinUCB [Russac 等人，2019]；以及 (d) 自适应重启算法 MASTER+OFUL [Wei 和 Luo，2021]。由于 P_T 是可计算的，我们将 LB-WeightUCB 和 D-LinUCB 的折扣因子设为 $\gamma=1-\max\{1/T,\sqrt{P_T/(dT)}\}$，并将窗口大小 w 和重启周期 H 设为 $w=H=d^{1/4}\sqrt{T/(1+P_T)}$。对于 MASTER，其中有一个参数 n 表示多尺度探索参数的初始值（参见 [Wei 和 Luo，2021]中的过程 1 输入），而原始 MASTER算法令其从 0 开始（即 $n=0$、1、.....）。然而，较小的 n 初始值会导致频繁重启，从而带来较差的性能。为了解决这一问题，我们尝试使用更大的 $n=13$ 初始值，在我们的实验中，这显著提升了性能。

结果。 实验结果取自 20次独立试验的平均值。图 1a 展示了累积动态遗憾表现，阴影区域表示 20次独立试验实验结果的方差。图 1b 报告了每次运行的平均时间，每次运行包含 6000轮。我们的 LB-WeightUCB 算法表现与 D-LinUCB 相当，但效率显著更高，加速比超过 1.5 倍。图 1a还显示，配备经过精细调优的 n时，MASTER+OFUL($n=13$) 的表现优于 RestartUCB；而原始的 MASTER+OFUL ($n=0$) 表现较差，这是由于初期重启过于频繁。然而，较大的初始值 n 会带来更高的时间开销，因为每次重启时，MASTER+OFUL 都需要执行一次过程1，从而导致 $\mathcal{O}(n2^n)$ 时间复杂度。更重要的是，在缓慢演化环境中，自适应重启（MASTER+OFUL）和周期性重启（RestartUCB）均不如我们的加权策略。



(a) GLB算法 ($S = 1$)

(b) SCB 算法 ($S = 1$)



(c) GLB算法 ($S = 5$)

(d) SCB 算法 ($S = 5$)

图 2：广义线性老虎机实验。

7.2 广义线性老虎机

设置。我们在 GLB 实验中采用逻辑模型，即奖励满足 $r_t \sim$ 伯努利分布($\mu(X_t^\top \theta_t)$)，其中逻辑函数为 $\mu(x) = (1 + e^{-x})^{-1}$。我们分别考虑 $S = 1$ 和 $S = 5$两种情况。我们将所提出的 GLB-WeightUCB 和 SCB-WeightUCB 算法与以下算法进行比较：(a) GLM-UCB，适用于 GLB 的静态算法 [Fil- lippi 等人， 2010]；(b) LogUCB1，适用于 LogB 的静态算法 [Faury 等人， 2020]；(c) BVD-GLM-UCB，适用于 GLB 的基于权重的算法 [Faury 等人， 2021]；(d) GLB-RestartUCB，适用于 GLB 的重启算法 [Zhao 等人， 2020]；(e) SCB-RestartUCB，适用于 SCB 的重启算法 [Zhao 等人， 2020]；(f) MASTER+GLM-UCB，适用于 GLB 的自适应重启算法 [Wei 和 Luo， 2021]；以及 (g) MASTER+LogUCB1，适用于 LogB 的自适应重启算法 [Wei 和 Luo， 2021]。我们为 GLB-WeightUCB 设置折扣因子$\gamma = 1 - \max\{1/T, \sqrt{c_\mu P_T/(dT)}\}$，为 BVD-GLM-UCB 设置$\gamma = 1 - (P_T/(\sqrt{dT}))^{2/5}$，并为 SCB-WeightUCB 设置$\gamma = 1 - \max\{1/T, \sqrt{P_T/(dT)}\}$。我们为 GLB-RestartUCB 和 SCB-RestartUCB 设置重启周期 $H = d^{1/4}\sqrt{T/(1 + P_T)}$。我们为 GLM-UCB、BVD-GLM-UCB、GLB-RestartUCB 和 MASTER+GLM-UCB 设置正则化项 $\lambda = d$；为 GLB-WeightUCB 设置 $\lambda = d/c_\mu^2$；为 LogUCB1、SCB-RestartUCB、MASTER+LogUCB1 和 SCB-WeightUCB 设置 $\lambda = d \log T/c_\mu$。注意，对于 LogB，$k_\mu = 1/4 < 1$，因此我们无需控制 k_μ的阶数。对于两个 MASTER 算法，我们设置 $n = 13$。

结果。 我们在 20 次独立试验中进行实验，并在图 2中展示了平均累积动态遗憾结果。当 S 较小时（$S = 1$, $c_\mu^{-1} \approx 5$），所有基于权重的算法都优于静态算法，并且我们的 GLB-WeightUCB 和 SCB-WeightUCB 优于 BVD-GLM-UCB。当 S 较大时（$S = 5$, $c_\mu^{-1} \approx 152$），SCB-WeightUCB 显著优于 GLB-WeightUCB，表明考虑自协调性质的重要性（请注意，LogB 是 SCB 的一个实例）。相比之下，BVD-GLM-UCB 的性能大幅下降，因为它没有考虑 c_μ^{-1}问题。与线性赌博机类似，GLB 的实验结果也表明，在缓慢演化环境中，加权策略相较于（自适应）重启策略具有经验优势。具体而言，我们观察到 GLB-WeightUCB 始终优于 MASTER+GLM-UCB，而 SCB-WeightUCB 始终优于 MASTER+LogUCB1。

8 结论

本文重新审视了三种非平稳参数化老虎机模型（LB、GLB、SCB）的基于权重的算法。我们发现，先前工作的不足源于对估计误差分析不充分。因此，我们提出了一种精细化分析框架，使得在估计误差分析中，偏差部分和方差部分可以使用相同的局部范数。我们的框架确保了针对这三种老虎机模型的算法具有更高的效率，并改进了 GLB 和 SCB 设置下的遗憾界。

我们工作的重要性在于，如今我们已经使非平稳 LB/GLB/SCB 的基于权重的算法在计算效率和遗憾保证两个方面都能够与基于重启的算法相媲美。鉴于加权策略在渐进漂移场景中尤其具有吸引力，而这类场景在现实世界应用中十分常见，因此有必要进一步设计自适应的非平稳参数化老虎机基于权重的算法，在无需了解环境非平稳性的情况下实现最优动态遗憾，这与当前通过自适应重启策略取得的已知最佳结果 [Wei 和 Luo， 2021]一致。

在本研究中，我们采用 $P_T = \sum_{t=2}^{T} \|\theta_{t-1} - \theta_t\|_2$ 作为刻画渐变环境的度量。然而，该度量在仅捕捉环境中的渐变方面可能不够精确，因为它也可能包含其他类型的变化，例如突变。这或许可以解释为什么基于权重的算法在理论上并未表现出显著优势，但与基于重启的算法相比，它们在渐变环境的实验中却表现得非常出色。为克服这一局限，未来的研究可以借鉴 Sobolev 类或 Hölder 类背后的思想，探索对渐变变化更为精细的刻画 [Baby 和 Wang， 2019] 或其他信息论工具 [Liu 等人， 2022]。

致谢

本研究得到国家自然科学基金（61921006）、江苏省自然科学基金（BK20220776）以及新型软件技术与产业化协同创新中心的资助。作者衷心感谢匿名审稿人和元审稿人提出的宝贵意见与建议。作者还感谢 Yu-Hu Yan 和 Yu-Jie Zhang 的多次有益讨论。

参考文献

Yasin Abbasi-Yadkori、D´avid P´al 和 Csaba Szepesv´ari。线性随机老虎机的改进算法。载于神经信息处理系统进展第*24*卷（*NIPS*），第2312–2320页，2011年。

Yasin Abbasi-Yadkori、Andr´as Gy¨orgy 和 Nevena Lazic。重新审视非平稳随机老虎机的动态遗憾。*arXiv* 预印本，arXiv：2201.06532，2022年。

Marc Abeille、Louis Faury 和 Cl´ement Calauz`enes。逻辑回归老虎机的逐实例极小极大最优算法。载于第*24*届国际人工智能与统计学会议（*AISTATS*）论文集，第3691–3699页，2021年。

Christoforos Anagnostopoulos、Dimitris K Tasoulis、Niall M Adams、Nicos G Pavlidis 和 David J Hand。采用自适应遗忘进行流式分类的在线线性与二次判别分析。统计分析与数据挖掘，5(2):139–166，2012。

Dheeraj Baby 和 Yu-Xiang Wang。全变差有界序列的在线预测。载于神经信息处理系统进展第 *32* 卷（*NeurIPS*），第 11069–11079 页，2019。

Wang Chi Cheung、David Simchi-Levi 和 Ruihao Zhu。在非平稳性下学习优化。载于第 *22* 届国际人工智能与统计会议论文集（*AISTATS*），第 1079–1087 页，2019。

Wang Chi Cheung、David Simchi-Levi 和 Ruihao Zhu。对冲漂移：学习在非平稳性下进行优化。管理科学，68(3):1696–1713，2022。

Chao-Kai Chiang、Chia-Jung Lee 和 Chi-Jen Lu。在渐进演化环境中击败多臂老虎机。载于 第 *26* 届学习理论年度会议（*COLT*）论文集，第 210–227 页，2013 年。

Yijing Chu 和 Cheuk Ming Mak。用于分布式估计的可变遗忘因子扩散递归最小二乘算法。信号处理，140：219–225，2017 年。

Giulia Clerici、Pierre Laforgue 和 Nicol`o Cesa-Bianchi。带记忆的线性老虎机：从衰退到增长。*arXiv*预印本，arXiv:2302.08345，2023年。

Koby Crammer、Yishay Mansour、Eyal Even-Dar 和 Jen- nifer Wortman Vaughan。带概念漂移的遗憾最小化。载于第*23*届学习理论会议（*COLT*）论文集，第168–180页，2010年。Yuntian Deng、Xingyu Zhou、Baekjin Kim、Ambuj Tewari、Abhishek Gupta 和 Ness Shroff。面向非平稳环境的加权高斯过程老虎机。载于第*25*届人工智能与统计国际会议（*AISTATS*）论文集，第6909–6932页，2022年。Shi Dong、Tengyu Ma 和 Benjamin Van Roy。汤普森采样在逻辑回归老虎机上的性能。载于第*32*届学习理论会议（*COLT*）论文集，第1158–1160页，2019年。

Louis Faury、Marc Abeille、Cl´ement Calauz`enes 和 Olivier Fercoq。改进的逻辑斯蒂多臂老虎机乐观算法。载于第 *37* 届国际机器学习大会（*ICML*）论文集，第 3052–3060 页，2020 年。

Louis Faury、Yoan Russac、Marc Abeille 和 Cl´ement Calauz`enes。参数漂移下广义线性多臂老虎机的遗憾界。*ArXiv* 预印本，arXiv:2103.05750，2021 年。

Sarah Filippi、Olivier Capp´e、Aur´elien Garivier 和 Csaba Szepesv´ari。参数化老虎机：广义线性情形。载于神经信息处理系统进展 *23*（*NIPS*），第 586–594 页，2010年。

Jo˜ao Gama、Indre Zliobaite、Albert Bifet、Mykola Pech- enizkiy 和 Abdelhamid Bouchachia。概念漂移适应综述。*ACM*计算调查，46(4)：44:1–44:37，2014年。

Lei Guo、Lennart Ljung 和 Pierre Priouret。遗忘因子递归最小二乘算法的性能分析。国际自适应控制与信号处理期刊，7(6)：525–537，1993年。

Simon S Haykin。自适应滤波理论。培生印度教育，2002年。

Wasim Huleihel、Soumyabrata Pal 和 Ofer Shayevitz。学习非平稳环境中的用户偏好。载于 第 *24* 届国际人工智能与统计学会议论文集（*AISTATS*），第 1432–1440 页，2021。

Chi Jin、Zhuoran Yang、Zhaoran Wang 和 Michael I. Jordan。基于线性函数近似的强化学习具有可证明的效率。载于 第 *33* 届学习理论会议论文集（*COLT*），第 2137–2143 页，2020。

Kwang-Sung Jun、Aniruddha Bhargava、Robert D. Nowak 和 RebeccaWillett。可扩展的广义线性老虎机：在线计算与哈希。载于神经信息处理系统进展第*30*卷（*NIPS*），第 99–109 页，2017。

Lihong Li、Yu Lu 和 Dengyong Zhou。广义线性上下文老虎机的可证明最优算法。载于第*34*届国际机器学习大会（*ICML*）论文集，第 2071–2080 页，2017。Yueyang Liu、Benjamin Van Roy 和 Kuang Xu。通过预测采样进行非平稳老虎机学习。*arXiv* 预印本，arXiv:2205.01970，2022。Yoan Russac、Claire Vernade 和 Olivier Capp´e。面向非平稳环境的加权线性老虎机。载于神经信息处理系统进展第*32*卷（*NeurIPS*），第 12040–12049 页，2019。Yoan Russac、Olivier Capp´e 和 Aur´elien Garivier。非平稳广义线性老虎机算法。*arXiv* 预印本，arXiv:2003.10113，2020。Yoan Russac、Louis Faury、Olivier Capp´e 和 Aur´elien Garivier。对采用遗忘机制的广义线性老虎机进行自协调分析。载于第*24*届人工智能与统计国际会议（*AISTATS*）论文集，第 658–666 页，2021。Joe Suk 和 Samory Kpotufe。追踪老虎机中最显著的臂切换。载于第*35*届学习理论会议（*COLT*）论文集，第 2160–2182 页，2022。Sabina Tomkins、Peng Liao、Predrag V. Klasnja 和 Susan A. Murphy。Intelligentpooling：用于移动医疗的实用汤普森采样。机器学习，110(9)：2685–2727，2021。Ahmed Touati 和 Pascal Vincent。非平稳线性马尔可夫决策过程中的高效学习。*arXiv* 预印本，arXiv:2010.12870，2020。Chen-Yu Wei 和 Haipeng Luo。无需先验知识的非平稳强化学习：一种最优黑箱方法。载于第*34*届学习理论会议（*COLT*）论文集，第 4300–4354 页，2021。Lijun Zhang、Tianbao Yang、Rong Jin、Yichi Xiao 和 Zhi-Hua Zhou。单比特反馈下的在线随机线性优化。载于第*33*届国际机器学习大会（*ICML*）论文集，第 392–401 页，2016。Peng Zhao 和 Lijun Zhang。重新审视非平稳线性老虎机。*arXiv* 预印本，arXiv:2103.05324，2021。Peng Zhao、Lijun Zhang、Yuan Jiang 和 Zhi-Hua Zhou。一种简单的非平稳线性老虎机方法。载于第*23*届人工智能与统计国际会议（*AISTATS*）论文集，第 746–755 页，2020。Peng Zhao、Xinqiang Wang、Siyu Xie、Lei Guo 和 Zhi-Hua Zhou。分布无关单遍学习。*IEEE*知识与数据工程汇刊，33：951–963，2021a。Peng Zhao、LijunZhang、Yuan Jiang 和 Zhi-Hua Zhou。非平稳线性老虎机的一种简单方法。*arXiv* 预印本，arXiv:2103.05324，2021b。

A LB-WeightUCB 分析

本节将对 LB-WeightUCB 算法进行分析。在附录 A.1 中，我们回顾 Russac 等人 [2019] 提出的 D-LinUCB 算法，并重述其估计误差分析。在附录 A.2 中，我们针对所提出的 LB-WeightUCB 算法给出自己的估计误差分析，该分析由引理 1 描述。最后，在附录 A.3 中，我们证明定理 1 中所述的动态遗憾界。

A.1 回顾 D-LinUCB 算法的估计误差分析

在本节中，我们回顾 D-LinUCB 算法先前的估计误差分析 [Russac 等人， 2019] 他们使用的估计量与我们相同（3）。第一步是将估计误差分解为偏差部分和方差部分，其中偏差部分表示参数漂移导致的误差，方差部分表示随机噪声导致的误差。基于奖励模型假设和估计量（与公式（1）和公式（3）相同），D-LinUCB 算法的估计误差可分解为

$$\begin{aligned}
\widehat{\theta}_t - \theta_t &= V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1} r_s X_s\right) - \theta_t \\
&= V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}\left(X_s^\top\theta_s + \eta_s\right)X_s\right) - V_{t-1}^{-1}\left(\lambda I_d + \sum_{s=1}^{t-1}\gamma^{t-s-1}X_s X_s^\top\right)\theta_t \\
&= V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}X_s X_s^\top\theta_s + \sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s\right) - V_{t-1}^{-1}\left(\lambda I_d + \sum_{s=1}^{t-1}\gamma^{t-s-1}X_s X_s^\top\right)\theta_t \\
&= \underbrace{V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}X_s X_s^\top\left(\theta_s - \theta_t\right)\right)}_{\texttt{bias part}} + \underbrace{V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s - \lambda\theta_t\right)}_{\texttt{variance part}}.
\end{aligned} \tag{20}$$

随后，Russac 等人 [2019] 对偏差部分和方差部分使用不同的局部范数（稍后我们将解释使用不同局部范数的原因），如下所示，

$$|\mathbf{x}^\top(\widehat{\theta}_t - \theta_t)| \le \|\mathbf{x}\|_2 A'_t + \|\mathbf{x}\|_{V_{t-1}^{-1}\widetilde{V}_{t-1}V_{t-1}^{-1}} B'_t, \tag{21}$$

其中 $\widetilde{V}_t = \lambda I_d + \sum_{s=1}^{t}\gamma^{2(t-s)}X_s X_s^\top$和

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

对于偏差部分，Rus sac 等人 [2019] 通过引入虚拟窗口大小，将其沿时间线划分为两部分 D ，

$$A'_t \le \underbrace{\left\|\sum_{s=t-D}^{t-1}V_{t-1}^{-1}\gamma^{t-s-1}X_s X_s^\top\left(\theta_s - \theta_t\right)\right\|_2}_{\texttt{virtual window}} + \underbrace{\left\|\sum_{s=1}^{t-D-1}V_{t-1}^{-1}\gamma^{t-s-1}X_s X_s^\top\left(\theta_s - \theta_t\right)\right\|_2}_{\texttt{small term}},$$

第一项可以视为一个虚拟窗口，其中包含在时间 $t-D$之后获取的最新数据，并且可以直接通过对 SW-UCB 的分析来分析 [Cheung 等人， 2019]因为它对应于窗口策略估计误差中的偏差部分，这也是他们对偏差部分使用 l_2-范数的原因。第二项反映了在时间 $t-D$之前获取的过时数据所产生的影响。由于 γ^{t-s-1} 在 $s \le t-D-1$时会非常小，因此这一小项会被第一项虚拟窗口项所支配，这意味着偏差部分实际上受虚拟窗口大小控制 D。

对于方差部分，Russac 等人 [2019] 将先前的自归一化集中不等式 [Abbasi-Yadkori 等人，2011，~定理 1] 推广到加权形式，并在定理5中重新陈述。该集中不等式要求使用 V_t 作为局部范数。为此，Russac 等人 [2019]将方差部分拆分为

$$\left|\mathbf{x}^\top V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s - \lambda\theta_t\right)\right| \le \|\mathbf{x}\|_{V_{t-1}^{-1}\widetilde{V}_{t-1}V_{t-1}^{-1}}\left\|V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s - \lambda\theta_t\right)\right\|_{V_{t-1}\widetilde{V}_{t-1}^{-1}V_{t-1}},$$

其中

$$\left\|V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s-\lambda\theta_t\right)\right\|_{V_{t-1}\widetilde{V}_{t-1}^{-1}V_{t-1}}=\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s-\lambda\theta_t\right\|_{\widetilde{V}_{t-1}^{-1}}\leq\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s\right\|_{\widetilde{V}_{t-1}^{-1}}+\sqrt{\lambda}S.$$

那么，项 $\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s\|_{\widetilde{V}_{t-1}^{-1}}$ 可由定理5进行界定。最后，基于上述分析，D-LinUCB 需要采用以下动作选择准则，该准则仅取决于方差部分，因为偏差部分不包含$\mathbf{x}$,

$$X_t=\arg\max_{\mathbf{x}\in\mathcal{X}}\left\{\langle\mathbf{x},\widehat{\theta}_t\rangle+\beta_{t-1}\|\mathbf{x}\|_{V_{t-1}^{-1}\widetilde{V}_{t-1}V_{t-1}^{-1}}\right\},$$

其中 β_{t-1} 是 B_t' 的上界，与（5）相同。从该选择准则可以看出，D-LinUCB$^{\sim}$需要维护两个协方差矩阵，即 V_t 和 V_t 在第 t 轮算法运行期间。

下一节将给出估计误差上界的证明。我们的分析与 D-LinUCB 的分析主要从步骤（21）开始有所不同，这是分析的关键步骤，而我们提出的新分析框架使我们能够采用相同的局部范数来处理偏差部分和方差部分。

A.2 引理1<style>的证明

证明。 使用相同的 d （20）中的推导表明，LB-WeightUCB 算法的估计误差也可以分解为

$$\widehat{\theta}_t-\theta_t=\underbrace{V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}X_s X_s^\top(\theta_s-\theta_t)\right)}_{\texttt{bias part}}+\underbrace{V_{t-1}^{-1}\left(\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s-\lambda\theta_t\right)}_{\texttt{variance part}}.$$

因此，根据柯西-施瓦茨不等式，我们知道，对于任意 $\mathbf{x}\in\mathcal{X}$,

$$\left|\mathbf{x}^\top\left(\widehat{\theta}_t-\theta_t\right)\right|\leq\|\mathbf{x}\|_{V_{t-1}^{-1}}(A_t+B_t),\tag{22}$$

其中

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

上述两项可以分别进行界定，如下两个引理所述，

引理 4. 对于任意 $t\in[T]$，有

$$\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}X_s X_s^\top(\theta_s-\theta_t)\right\|_{V_{t-1}^{-1}}\leq L\sqrt{d}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\|\theta_p-\theta_{p+1}\|_2.$$

引理 5. 对于任意 $\delta\in(0,1)$，以至少 $1-\delta$ 的概率，以下结论对所有 $t\in[T]$ 均成立，

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

根据不等式（22）、引理 4、引理 5以及可行集的有界性假设，对于任意 $\mathbf{x}\in\mathcal{X}$，$\gamma\in(0,1)$ 和 $\delta\in(0,1)$，以至少 $1-\delta$的概率，以下结论对所有 $t\in[T]$均成立，

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

其中 $\beta_t\triangleq\sqrt{\lambda}S+R\sqrt{2\log\frac{1}{\delta}+d\log\left(1+\frac{L^2}{(1-\gamma^{2t})\lambda d(1-\gamma^2)}\right)}$ 是 LB-WeightUCB 中使用的置信半径。因此，我们完成了证明。

□

引理 4 的证明。第一步是提取参数 θ_t 的变化量，如下所示：

$$\begin{aligned}\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}X_sX_s^\top(\theta_s-\theta_t)\right\|_{V_{t-1}^{-1}} &= \left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}X_sX_s^\top\sum_{p=s}^{t-1}(\theta_p-\theta_{p+1})\right\|_{V_{t-1}^{-1}}\\ &= \left\|\sum_{p=1}^{t-1}\sum_{s=1}^{p}\gamma^{t-s-1}X_sX_s^\top(\theta_p-\theta_{p+1})\right\|_{V_{t-1}^{-1}}\\ &\le \sum_{p=1}^{t-1}\left\|\sum_{s=1}^{p}\gamma^{t-s-1}X_s\|X_s\|_2\|\theta_p-\theta_{p+1}\|_2\right\|_{V_{t-1}^{-1}}\\ &\le L\sum_{p=1}^{t-1}\sum_{s=1}^{p}\gamma^{t-s-1}\|X_s\|_{V_{t-1}^{-1}}\|\theta_p-\theta_{p+1}\|_2,\end{aligned}$$

以及项 $\sum_{s=1}^{p}\gamma^{t-s-1}\|X_s\|_{V_{t-1}^{-1}}$ 可以进一步推导出折扣因子 γ 的表达式如下 s,

$$\sum_{s=1}^{p}\gamma^{t-s-1}\|X_s\|_{V_{t-1}^{-1}} \le \gamma^{\frac{t-1}{2}}\sqrt{\sum_{s=1}^{p}\gamma^{-s}}\sqrt{\sum_{s=1}^{p}\gamma^{t-s-1}\|X_s\|_{V_{t-1}^{-1}}^2} \le \sqrt{d}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}. \tag{23}$$

上述推导中，我们利用了以下事实：对于任意 $\mathbf{x}$，$\|\mathbf{x}\|_{V_{t-1}^{-1}} \le \|\mathbf{x}\|_2/\sqrt{\lambda}$，因为 $V_{t-1}\succeq\lambda I_d$。倒数第二步根据柯西-施瓦茨不等式成立。此外，最后一步依据以下事实：

$$\forall p\in[t-1],\quad \sum_{s=1}^{p}\gamma^{t-s-1}\|X_s\|_{V_{t-1}^{-1}}^2 \le d, \tag{24}$$

这一点可通过以下论证证明。

$$\begin{aligned}\sum_{s=1}^{p}\gamma^{t-s-1}\|X_s\|_{V_{t-1}^{-1}}^2 &= \sum_{s=1}^{p}\gamma^{t-s-1}\mathrm{Tr}(X_s^\top V_{t-1}^{-1}X_s) = \mathrm{Tr}\left(V_{t-1}^{-1}\sum_{s=1}^{p}\gamma^{t-s-1}X_sX_s^\top\right)\\ &\le \mathrm{Tr}\left(V_{t-1}^{-1}\sum_{s=1}^{p}\gamma^{t-s-1}X_sX_s^\top\right)+\mathrm{Tr}\left(V_{t-1}^{-1}\sum_{s=p+1}^{t-1}\gamma^{t-s-1}X_sX_s^\top\right)+\mathrm{Tr}\left(V_{t-1}^{-1}\lambda\sum_{i=1}^{d}\mathbf{e}_i\mathbf{e}_i^\top\right)\\ &= \mathrm{Tr}(I_d)=d.\end{aligned}$$

因此，证明完毕。 □

引理的证明 5。 令 $\widetilde{V}_t \triangleq \lambda I_d+\sum_{s=1}^{t}\gamma^{2(t-s)}X_sX_s^\top$，

$$\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_sX_s-\lambda\theta_t\right\|_{V_{t-1}^{-1}} \le \left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_sX_s\right\|_{V_{t-1}^{-1}}+\|\lambda\theta_t\|_{V_{t-1}^{-1}} \le \left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_sX_s\right\|_{\widetilde{V}_{t-1}^{-1}}+\sqrt{\lambda}S.$$

回忆一下，$V_t=\lambda I_d+\sum_{s=1}^{t}\gamma^{t-s}X_sX_s^\top$，因此最后一个不等式由下式得出

$$V_t=\lambda I_d+\sum_{s=1}^{t}\gamma^{t-s}X_sX_s^\top \succeq \lambda I_d+\sum_{s=1}^{t}\gamma^{2(t-s)}X_sX_s^\top=\widetilde{V}_t.$$

~我们强调，V_t 仅在分析中引入，实际上并不是我们的算法实现所必需的。根据 [Russac 等人，2019，定理 1]，在定理 5 中重述的加权版最大偏差不等式，我们只需令 $w_s=\gamma^{t-s-1}$，$\mu_t=\lambda$，

$$\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_sX_s\right\|_{\widetilde{V}_{t-1}^{-1}} \le R\sqrt{2\log\frac{1}{\delta}+d\log\left(1+\frac{L^2\sum_{s=1}^{t-1}\gamma^{2(t-s-1)}}{\lambda d}\right)} \le R\sqrt{2\log\frac{1}{\delta}+d\log\left(1+\frac{L^2(1-\gamma^{2t-2})}{\lambda d(1-\gamma^2)}\right)},$$

从而完成证明。 □

A.3 定理1的证明

证明。 设 $X_t^* \triangleq$ 使其最大化$\mathbf{x}\in\mathcal{X}\mathbf{x}^\top\theta_t$。 根据引理 1 以及 X_t^*， $X_t \in \mathcal{X}$以下每一项成立的概率至少为 $1-\delta$,

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

$$\forall t \in [T], X_t^{\top}\theta_t \ge X_t^{\top}\widehat{\theta}_t - L^2\sqrt{\frac{d}{\lambda}}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2 - \beta_{t-1}\|X_t\|_{V_{t-1}^{-1}}.$$

根据并集界，以下结论以至少 $1-2\delta$ 的概率成立,

$$\begin{aligned}\forall t \in [T], X_t^{*\top}\theta_t - X_t^{\top}\theta_t &\le X_t^{*\top}\widehat{\theta}_t - X_t^{\top}\widehat{\theta}_t + 2L^2\sqrt{\frac{d}{\lambda}}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2 + \beta_{t-1}(\|X_t^*\|_{V_{t-1}^{-1}} + \|X_t\|_{V_{t-1}^{-1}})\\ &\le 2L^2\sqrt{\frac{d}{\lambda}}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2 + 2\beta_{t-1}\|X_t\|_{V_{t-1}^{-1}},\end{aligned}$$

其中，最后一步由臂选择准则（6）得出，即

$$X_t^{*\top}\widehat{\theta}_t + \beta_{t-1}\|X_t^*\|_{V_{t-1}^{-1}} \le X_t^{\top}\widehat{\theta}_t + \beta_{t-1}\|X_t\|_{V_{t-1}^{-1}}.$$

因此，以下动态遗憾界以至少 $1-2\delta$ 的概率成立，并可分为两部分,

$$R_T = \sum_{t=1}^{T}\left(X_t^{*\top}\theta_t - X_t^{\top}\theta_t\right) \le \underbrace{2L^2\sqrt{\frac{d}{\lambda}}\sum_{t=1}^{T}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2}_{\text{bias part}} + \underbrace{2\beta_T\sum_{t=1}^{T}\|X_t\|_{V_{t-1}^{-1}}}_{\text{variance part}},$$

其中 $\beta_T = \sqrt{\lambda}S + R\sqrt{2\log\frac{1}{\delta} + d\log\left(1+\frac{L^2}{(1-\gamma^{2T})\lambda d(1-\gamma^2)}\right)}$是置信半径。

现在我们分别推导偏差部分和方差部分的上界。

偏差部分。 对于偏差部分，我们需要提取路径长度 P_T，并展示折扣因子 γ 对 P_T 的控制作用。

$$\begin{aligned}2L^2\sqrt{\frac{d}{\lambda}}\sum_{t=1}^{T}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2 &= 2L^2\sqrt{\frac{d}{\lambda}}\sum_{p=1}^{T-1}\sum_{t=p+1}^{T}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2\\ &= 2L^2\sqrt{\frac{d}{\lambda}}\sum_{p=1}^{T-1}\frac{\gamma^{\frac{p}{2}}-\gamma^{\frac{T}{2}}}{1-\gamma^{\frac{1}{2}}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2\\ &\le 2L^2\sqrt{\frac{d}{\lambda}}\sum_{p=1}^{T-1}\frac{\gamma^{\frac{p}{2}}-\gamma^{\frac{T}{2}}}{(1-\gamma^{\frac{1}{2}})\frac{1+\gamma^{\frac{1}{2}}}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2\\ &\le 4L^2\sqrt{\frac{d}{\lambda}}\sum_{p=1}^{T-1}\frac{\gamma^{\frac{p}{2}}\gamma^{-\frac{p}{2}}}{(1-\gamma)^{3/2}}\left\|\theta_p-\theta_{p+1}\right\|_2\\ &= 4L^2\sqrt{\frac{d}{\lambda}}\frac{1}{(1-\gamma)^{3/2}}P_T.\end{aligned}$$

因此，对于偏差部分，我们有

$$2L^2\sqrt{\frac{d}{\lambda}}\sum_{t=1}^{T}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2 \le 4L^2\sqrt{\frac{d}{\lambda}}\frac{1}{(1-\gamma)^{3/2}}P_T. \tag{25}$$

方差部分。首先，利用柯西-施瓦茨不等式，我们知道 $2\beta_T\sum_{t=1}^{T}\|X_t\|_{V_{t-1}^{-1}} \le 2\beta_T\sqrt{T\sum_{t=1}^{T}\|X_t\|^2_{V_{t-1}^{-1}}}$.

然后根据引理11 （势函数引理），我们得到以下上界：

$$2\beta_T \sum_{t=1}^{T} \|X_t\|_{V_{t-1}^{-1}} \leq 2\beta_T \sqrt{2\max\{1, L^2/\lambda\} dT} \sqrt{T \log \frac{1}{\gamma} + \log\left(1 + \frac{L^2}{\lambda d(1-\gamma)}\right)}. \tag{26}$$

结合偏差部分和方差部分的上界，并取置信水平 $\delta = 1/(2T)$，根据并集界，我们以至少 $1 - 1/T$ 的概率得到如下动态遗憾界，

$$R_T \leq 4L^2 \sqrt{\frac{d}{\lambda}} \frac{1}{(1-\gamma)^{3/2}} P_T + 2\beta_T \sqrt{2\max\{1, L^2/\lambda\} dT} \sqrt{T \log \frac{1}{\gamma} + \log\left(1 + \frac{L^2}{\lambda d(1-\gamma)}\right)}.$$

其中 $\beta_T = \sqrt{\lambda} S + R\sqrt{2\log T + 2\log 2 + d\log\left(1 + \frac{L^2}{(1-\gamma^{2T})\lambda d(1-\gamma^2)}\right)}$。由于遗憾界中含有一项 $T\sqrt{\log(1/\gamma)}$，因此不能令 γ 接近 0，所以我们设定 $\gamma \geq 1/T$，并有 $\log(1/\gamma) \leq C(1-\gamma)$，其中 $C = \log T/(1-1/T)$。

然后，忽略时间范围 T 中的对数因子，并令 $\lambda = d$，最终得到

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

当 $P_T < d/T$（这对应于较小程度的非平稳性）时，我们直接令 $\gamma = 1 - 1/T$，并获得 $\widetilde{\mathcal{O}}(d\sqrt{T})$ 遗憾界。此外，在非退化情形（$P_T \geq d/T$）下，我们将折扣因子最优地设为 $1 - \gamma = \sqrt{P_T/(dT)}$，并得到 $\widetilde{\mathcal{O}}(d^{3/4} P_T^{1/4} T^{3/4})$ 动态遗憾界，从而完成证明。 □

B GLB-WeightUCB 分析

在本节中，我们对 GLB-WeightUCB 算法进行分析。在附录 B.1 中，我们回顾 GLB 的投影问题，并重述 Faury 等人 [2021] 提出的 BVD-GLM-UCB 算法。在附录 B.2 中，我们给出 GLB-WeightUCB 算法估计误差上界的证明（即引理2）。最后，在附录 B.3 中，我们给出定理2 所述动态遗憾上界的证明。

B.1 回顾 BVD-GLM-UCB 算法的投影步骤

如第 4.2 节所述，GLB 的主要难点在于，MLE 或 QMLE 估计量的结果 $\widehat{\theta}_t$ 可能不属于可行集 Θ，且 c_μ 定义在参数 $\theta \in \Theta$上。在平稳环境下，Filippi 等人 [2010]通过引入投影步骤克服了这一难题，具体如下：

$$\widetilde{\theta}_t = \arg\min_{\theta \in \Theta} \|g_t(\widehat{\theta}_t) - g_t(\theta)\|_{V_{t-1}^{-1}}, \tag{27}$$

其中 $V_t = \lambda I_d + \sum_{s=1}^{t} X_s X_s^\top$和 $g_t(\theta) = \lambda c_\mu \theta + \sum_{s=1}^{t-1} \mu(X_s^\top \theta) X_s$ 是静态版本（通过令 $\gamma = 1$得到）。根据 QMLE，我们知道

$$g_t(\widehat{\theta}_t) = \lambda c_\mu \widehat{\theta}_t + \sum_{s=1}^{t-1} \mu(X_s^\top \widehat{\theta}_t) X_s = \sum_{s=1}^{t-1} r_s X_s, \tag{28}$$

然后，根据中值定理，我们知道

$$g_t(\theta_1) - g_t(\theta_2) = G_t(\theta_1, \theta_2)(\theta_1 - \theta_2), \tag{29}$$

其中 $G_t(\theta_1, \theta_2) \triangleq \int_0^1 \nabla g_t(s\theta_2 + (1-s)\theta_1) \mathrm{d}s \in \mathbb{R}^{d \times d}$。注意，对于任意 $\theta \in \Theta$，g_t 的梯度满足

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

这显然意味着 $\forall \theta_1, \theta_2 \in \Theta$，$G_t(\theta_1, \theta_2) \succeq c_\mu V_{t-1}$。

算法3BVD-GLM-UCB [Faury 等人， 2021]

输入：时间范围 T、折扣因子 γ、置信度 δ、正则化项 λ、逆链接函数 μ、参数 S， L 和 R1: 设置 $V_0 = \lambda I_d$， $\theta_1 = \mathbf{0}$ 并计算 k_μ 和 c_μ2: **对于** $t = 1\ 2\ \ldots\ T$ **执行**

3: Find θ_t^p by solving $\theta_t^p \in \arg\min_{\theta\in\mathbb{R}^d} \left\{ \left\| g_t(\theta) - g_t(\widehat{\theta}_t) \right\|_{V_t^{-2}} \text{ s.t } \Theta \cap \mathcal{E}_t^\delta(\theta) \neq \emptyset \right\}$

4: Select $\widetilde{\theta}_t \in \Theta \cap \mathcal{E}_t^\delta(\theta_t^p)$ where $\mathcal{E}_t^\delta(\theta) := \left\{ \theta' \in \mathbb{R}^d \,\middle|\, \|g_t(\theta') - g_t(\theta)\|_{\widetilde{V}_t^{-1}} \leq \bar{\beta}_t(\delta) \right\}$

5: Compute $\bar{\beta}_{t-1}$ by $\bar{\beta}_t = \sqrt{\lambda} c_\mu S + R\sqrt{2\log\frac{1}{\delta} + d\log\left(1 + \frac{L^2(1-\gamma^{2t})}{\lambda d(1-\gamma^2)}\right)}$

6: Select X_t by $X_t = \arg\max_{\mathbf{x}\in\mathcal{X}} \left\{ \mu(\mathbf{x}^\top \widetilde{\theta}_t) + \frac{2k_\mu}{c_\mu} \bar{\beta}_{t-1} \|\mathbf{x}\|_{V_{t-1}^{-1}} \right\}$

7: Receive the reward r_t

8: Update $V_t = \gamma V_{t-1} + X_t X_t^\top + (1-\gamma)\lambda I_d$, $\widetilde{V}_t = \gamma^2 V_{t-1} + X_t X_t^\top + (1-\gamma^2)\lambda I_d$

9: Compute $\widehat{\theta}_{t+1}$ according to $\lambda c_\mu \theta + \sum_{s=1}^t \gamma^{t-s} \left(\mu(X_s^\top \theta) - r_s\right) X_s = 0$

10: **end 或**

通过这一投影步骤，Filippi 等人 [2010] 可以像下面这样分析估计误差：

$$\begin{aligned}
|\mu(\mathbf{x}^\top \widetilde{\theta}_t) - \mu(\mathbf{x}^\top \theta_t)| &\leq k_\mu |\mathbf{x}^\top (\widetilde{\theta}_t - \theta_t)| \\
&= k_\mu |\mathbf{x}^\top G_t^{-1}(\theta_t, \widetilde{\theta}_t)(g_t(\widetilde{\theta}_t) - g_t(\theta_t))| \\
&\leq k_\mu \|\mathbf{x}\|_{G_t^{-1}(\theta_t,\widetilde{\theta}_t)} \|g_t(\widetilde{\theta}_t) - g_t(\theta_t)\|_{G_t^{-1}(\theta_t,\widetilde{\theta}_t)} \\
&\leq \frac{k_\mu}{c_\mu} \|\mathbf{x}\|_{V_{t-1}^{-1}} \|g_t(\widetilde{\theta}_t) - g_t(\theta_t)\|_{V_{t-1}^{-1}} \\
&\leq \frac{2k_\mu}{c_\mu} \|\mathbf{x}\|_{V_{t-1}^{-1}} \|g_t(\widehat{\theta}_t) - g_t(\theta_t)\|_{V_{t-1}^{-1}},
\end{aligned}$$

其中，最后一步来自投影步骤。执行投影步骤后，项 $g_t(\theta_t) - g_t(\theta_t)$ 是未经投影的最大似然估计的估计误差。请注意，在分段平稳情形下，Russac 等人 [2021] 也可以使用这一投影步骤。Faury 等人 [2021] 认为，这两项先前工作之所以能够使用这一投影操作，主要是因为它们处于平稳或分段平稳设定中。他们指出，对于漂移情形，估计误差总是分为偏差（跟踪误差）和方差（学习误差）两部分，而这一简单的投影操作忽略了偏差部分，因此需要加以推广，以适应这两种偏差来源。在分析中，问题在于，经过投影步骤后，估计误差项 $g_t(\theta_t) - g_t(\theta_t)$ 需要拆分为偏差部分和方差部分，而 Faury 等人 [2021] 需要对偏差部分使用 l_2- 范数，对方差部分使用 V_{t-1}^{-1}。但整个估计误差已经按 V_{t-1}^{-1}进行范数化，这意味着他们无法对偏差部分沿用先前的窗口策略分析。

为此，Faury 等人 [2021] 为漂移广义线性赌博机提出了 BVD-GLM-UCB 算法，如算法 3所述，其中设计了一个新的投影步骤来解决这一问题。具体而言，在每一轮 t中，第一步是构造置信集 $\mathcal{E}_t^\delta(\theta)$，它表示随机噪声的影响。

$$\mathcal{E}_t^\delta(\theta) := \left\{ \theta' \in \mathbb{R}^d \,\middle|\, \|g_t(\theta') - g_t(\theta)\|_{\widetilde{V}_t^{-1}} \leq \bar{\beta}_t(\delta) \right\}. \tag{30}$$

第二步是找到一个与可行集相交的置信集 $\mathcal{E}_t^\delta(\theta_t^p)$，而 θ_t^p 与 $\widehat{\theta}_t$之间的差距表示参数漂移的影响。

$$\theta_t^p \in \arg\min_{\theta\in\mathbb{R}^d} \left\{ \left\| g_t(\theta) - g_t(\widehat{\theta}_t) \right\|_{V_t^{-2}} \text{ s.t } \Theta \cap \mathcal{E}_t^\delta(\theta) \neq \emptyset \right\}. \tag{31}$$

在获得解 θ_t^p（通过求解优化问题（31））后，第三步是从 $\widetilde{\theta}_t$ 中选择 $\Theta \cap \mathcal{E}_t^\delta$（$\theta_t^p$）。基于这一投影步骤，Faury 等人 [2021] 可以在投影前将偏差部分和方差部分分离如下：

$$\begin{aligned}
|\mu(\mathbf{x}^\top \widetilde{\theta}_t) - \mu(\mathbf{x}^\top \theta_t)| &\leq k_\mu |\mathbf{x}^\top (\widetilde{\theta}_t - \theta_t)| \\
&= k_\mu |\mathbf{x}^\top G_t^{-1}(\theta_t, \widetilde{\theta}_t)(g_t(\widetilde{\theta}_t) - g_t(\theta_t))|
\end{aligned} \tag{32}$$

$$\leq k_\mu |\mathbf{x}^\top G_t^{-1}(\theta_t, \widetilde{\theta}_t)(g_t(\widetilde{\theta}_t) - g_t(\theta_t^p) + g_t(\theta_t^p) - g_t(\widehat{\theta}_t) + g_t(\widehat{\theta}_t) - g_t(\bar{\theta}_t) + g_t(\bar{\theta}_t) - g_t(\theta_t))| \tag{33}$$

$$\leq \underbrace{k_\mu |\mathbf{x}^\top G_t^{-1}(\theta_t, \widetilde{\theta}_t)(g_t(\widetilde{\theta}_t) - g_t(\theta_t^p) + g_t(\widehat{\theta}_t) - g_t(\bar{\theta}_t)|}_{\text{bias part}} \tag{34}$$

$$+ \underbrace{k_\mu |\mathbf{x}^\top G_t^{-1}(\theta_t, \widetilde{\theta}_t)(g_t(\theta_t^p) - g_t(\widehat{\theta}_t) + g_t(\bar{\theta}_t) - g_t(\theta_t))|}_{\text{variance part}}. \tag{35}$$

他们的偏差-方差分解促使其在算法和分析中选择 不同的 局部范数，以界定偏差部分和方差部分。值得注意的是，由于投影步骤的复杂性（见（30）和（31）），整体算法相当复杂，不太适合实际实现；此外，在算法运行期间的每一轮 ˜协方差矩阵 V_t 和 V_t （由于所构造的置信区域（30）） t 都需要维护。下一节将说明，平稳 GLB 中使用的简单投影（27）通过我们改进的分析框架，便足以应对漂移 GLB。

B.2 引理 2 的证明

证明。 根据估计量方程（8），我们知道

$$g_t(\widehat{\theta}_t) = \lambda c_\mu \widehat{\theta}_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \widehat{\theta}_t) X_s = \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s, \tag{36}$$

然后根据中值定理，我们知道

$$g_t(\theta_1) - g_t(\theta_2) = G_t(\theta_1, \theta_2)(\theta_1 - \theta_2), \tag{37}$$

其中 $G_t(\theta_1, \theta_2) \triangleq \int_0^1 \nabla g_t(s\theta_2 + (1-s)\theta_1)\mathrm{d}s \in \mathbb{R}^{d\times d}$。注意，对于任意 $\theta \in \Theta$， g_t 的梯度 is

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

这显然意味着 $\forall \theta_1, \theta_2 \in \Theta$， $G_t(\theta_1, \theta_2) \succeq c_\mu V_{t-1}$。

根据假设 2，对37式应用中值定理， g_t 并结合投影（9，我们得到

$$\begin{aligned}
|\mu(\mathbf{x}^\top \widetilde{\theta}_t) - \mu(\mathbf{x}^\top \theta_t)| &\leq k_\mu |\mathbf{x}^\top (\widetilde{\theta}_t - \theta_t)| \\
&= k_\mu |\mathbf{x}^\top G_t^{-1}(\theta_t, \widetilde{\theta}_t)(g_t(\widetilde{\theta}_t) - g_t(\theta_t))| \\
&\leq k_\mu \|\mathbf{x}\|_{G_t^{-1}(\theta_t, \widetilde{\theta}_t)} \|g_t(\widetilde{\theta}_t) - g_t(\theta_t)\|_{G_t^{-1}(\theta_t, \widetilde{\theta}_t)} \\
&\leq \frac{k_\mu}{c_\mu} \|\mathbf{x}\|_{V_{t-1}^{-1}} \|g_t(\widetilde{\theta}_t) - g_t(\theta_t)\|_{V_{t-1}^{-1}} \\
&\leq \frac{2k_\mu}{c_\mu} \|\mathbf{x}\|_{V_{t-1}^{-1}} \|g_t(\widehat{\theta}_t) - g_t(\theta_t)\|_{V_{t-1}^{-1}},
\end{aligned}$$

˜那么，根据模型假设，函数 g_t (10) 和 $g_t(\theta_t)$ (36)，我们有，

$$\begin{aligned}
g_t(\theta_t) - g_t(\widehat{\theta}_t) &= \lambda c_\mu \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta_t) X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s \\
&= \lambda c_\mu \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta_t) X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_s) + \eta_s) X_s \qquad (38) \\
&= \underbrace{\sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s}_{\text{bias part}} + \underbrace{\lambda c_\mu \theta_t - \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s}_{\text{variance part}}. \qquad (39)
\end{aligned}$$

那么，根据柯西-施瓦茨不等式，我们知道，对于任意 $\mathbf{x} \in \mathcal{X}$，

$$\left|\mu(\mathbf{x}^\top \widetilde{\theta}_t) - \mu(\mathbf{x}^\top \theta_t)\right| \leq \frac{2k_\mu}{c_\mu} \|\mathbf{x}\|_{V_{t-1}^{-1}} (C_t + D_t), \tag{40}$$

其中

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

这两项可以分别进行界定，如以下引理所述。

引理 6. 对于任意 $t \in [T]$，我们有

$$\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{V_{t-1}^{-1}} \leq L k_\mu \sqrt{d} \sum_{p=1}^{t-1} \gamma^{\frac{t-1}{2}} \sqrt{\frac{\gamma^{-p}-1}{1-\gamma}} \left\| \theta_p - \theta_{p+1} \right\|_2. \tag{41}$$

引理 7. 对于任意 $\delta \in (0,\ 1)$，以至少 $1-\delta$ 的概率，以下结论对所有 $t \in [T]$ 均成立，

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

根据不等式 (40)、引理6、引理 7以及可行集的有界性假设，对于任意$\mathbf{x} \in \mathcal{X}$，$\gamma \in (0,\ 1)$，$\delta \in (0,\ 1)$，以至少 $1-\delta$ 的概率，以下结论对所有 $t \in [T]$ 均成立，

$$\begin{aligned} \left| \mu(\mathbf{x}^\top \widetilde{\theta}_t) - \mu(\mathbf{x}^\top \theta_t) \right| &\leq \frac{2k_\mu}{c_\mu} \|\mathbf{x}\|_{V_{t-1}^{-1}} \left(L k_\mu \sqrt{d} \sum_{p=1}^{t-1} \gamma^{\frac{t-1}{2}} \sqrt{\frac{\gamma^{-p}-1}{1-\gamma}} \left\| \theta_p - \theta_{p+1} \right\|_2 + \bar{\beta}_{t-1} \right) \\ &\leq \frac{2k_\mu}{c_\mu} \left(L^2 k_\mu \sqrt{\frac{d}{\lambda}} \sum_{p=1}^{t-1} \gamma^{\frac{t-1}{2}} \sqrt{\frac{\gamma^{-p}-1}{1-\gamma}} \left\| \theta_p - \theta_{p+1} \right\|_2 + \bar{\beta}_{t-1} \|\mathbf{x}\|_{V_{t-1}^{-1}} \right), \end{aligned}$$

其中 $\bar{\beta}_t \triangleq \sqrt{\lambda} c_\mu S + R \sqrt{2 \log \frac{1}{\delta} + d \log \left(1 + \frac{L^2}{(1-\gamma^{2t}) \lambda d (1-\gamma^2)} \right)}$ 是 GLB-WeightUCB 中使用的置信半径。因此，我们完成了证明。

□

引理6的证明。 这里我们需要提取时变参数的变化 θ_t

$$\begin{aligned} \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{V_{t-1}^{-1}} &\leq \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \sum_{p=s}^{t-1} (\mu(X_s^\top \theta_p) - \mu(X_s^\top \theta_{p+1})) X_s \right\|_{V_{t-1}^{-1}} \\ &= \left\| \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} \alpha(X_s, \theta_p, \theta_{p+1}) (X_s^\top \theta_p - X_s^\top \theta_{p+1}) X_s \right\|_{V_{t-1}^{-1}} \\ &= \left\| \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} \alpha(X_s, \theta_p, \theta_{p+1}) X_s X_s^\top (\theta_p - \theta_{p+1}) \right\|_{V_{t-1}^{-1}} \\ &\leq \sum_{p=1}^{t-1} \left\| \sum_{s=1}^{p} \gamma^{t-s-1} \alpha(X_s, \theta_p, \theta_{p+1}) X_s \|X_s\|_2 \|\theta_p - \theta_{p+1}\|_2 \right\|_{V_{t-1}^{-1}} \\ &\leq L \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} |\alpha(X_s, \theta_p, \theta_{p+1})| \left\| X_s \right\|_{V_{t-1}^{-1}} \|\theta_p - \theta_{p+1}\|_2 \\ &\leq L k_\mu \sum_{p=1}^{t-1} \sum_{s=1}^{p} \gamma^{t-s-1} \left\| X_s \right\|_{V_{t-1}^{-1}} \|\theta_p - \theta_{p+1}\|_2. \end{aligned}$$

其中第四个等式由中值定理得出，其中 $\alpha(\mathbf{x}\ \theta_1, \theta_2) = \int_0^1 \mu'(v\mathbf{x}^\top \theta_2 + (1-v) x^\top \theta_1) \mathrm{d}v$:

$$\mu(X_s^\top \theta_p) - \mu(X_s^\top \theta_{p+1}) = \alpha(X_s, \theta_p, \theta_{p+1}) (X_s^\top \theta_p - X_s^\top \theta_{p+1}).$$

接下来，项 $\sum_{s=1}^{p} \gamma^{t-s-1} \left\| X_s \right\|_{V_{t-1}^{-1}}$ 的界的推导与 A.2 中的不等式（23）相同，因此证明完成。

□

引理 7 的证明。 与线性情形相同，我们需要使用 $\widetilde{V}_t = \lambda I_d + \sum_{s=1}^{t} \gamma^{2(t-s)} X_s X_s^\top$。

$$\begin{aligned} D_t &= \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_\mu \theta_t \right\|_{V_{t-1}^{-1}} \\ &\le \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{V_{t-1}^{-1}} + \|\lambda c_\mu \theta_t\|_{V_{t-1}^{-1}} \\ &\le \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{\widetilde{V}_{t-1}^{-1}} + \sqrt{\lambda} c_\mu S \\ &\le R\sqrt{2\log\frac{1}{\delta} + d\log\left(1 + \frac{L^2(1-\gamma^{2t-2})}{\lambda d(1-\gamma^2)}\right)} + \sqrt{\lambda} c_\mu S. \end{aligned}$$

~再次强调， V_t 是引入分析中的唯一因素。此处的证明与引理5 在A.2中的证明相同，唯一的区别是第二项中多了一个c_μ 。 □

B.3 定理2的证明

证明。 令 $X_t^* \triangleq \arg\max_{\mathbf{x}\in\mathcal{X}} \mu(\mathbf{x}^\top \theta_t)$。根据引理 2 以及 $X_t^*, X_t \in \mathcal{X}$，以下各项成立的概率至少为 $1-\delta$,

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

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

根据并集界，以下结论成立的概率至少为 $1-2\delta$: $\forall t \in [T]$

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

其中，最后一步来自臂选择准则（12），因此

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

因此，以下动态 遗憾界以至少 $1-2\delta$ 的概率成立，并可分为两部分，

$$\begin{aligned} R_T &= \sum_{t=1}^{T}\max_{\mathbf{x}\in\mathcal{X}}\mu(\mathbf{x}^\top\theta_t) - \mu(X_t^\top\theta_t) \\ &\le \underbrace{\frac{4L^2k_\mu^2}{c_\mu}\sqrt{\frac{d}{\lambda}}\sum_{t=1}^{T}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\|\theta_p-\theta_{p+1}\|_2}_{\text{bias part}} + \underbrace{\frac{4k_\mu}{c_\mu}\bar{\beta}_T\sum_{t=1}^{T}\|X_t\|_{V_{t-1}^{-1}}}_{\text{variance part}}. \end{aligned} \tag{43}$$

其中 $\bar{\beta}_t = \sqrt{\lambda}c_\mu S + R\sqrt{2\log\frac{1}{\delta} + d\log\left(1 + \frac{L^2}{(1-\gamma^{2t})\lambda d(1-\gamma^2)}\right)}$ 是置信半径。

现在我们分别推导这两部分的上界。

算法4SCB -WeightUCB

输入：时间范围 T，折扣因子 γ，置信度 δ，正则化项 λ，逆链接函数 μ，参数 S、L 和 m 1：设置 $V_0 = \lambda I_d$、$\theta_1 = \mathbf{0}$ 并计算 k_μ 和 c_μ

2：对于 $t = 1, 2, \ldots, T$，执行 3：如果 $\|\theta_t\|_2 \le S$，则 4：令 $\theta_t = \theta_t$ 5：否则 6：执行投影，并根据 (14) 得到 θ_t 7：结束 if 8：根据 (16) 计算 β_{t-1} 9：根据 (17) 选择 X_t 10：接收奖励 r_t 11：更新 $V_t = \gamma V_{t-1} + X_t X_t^\top + (1-\gamma)\lambda I_d$ 12：根据 (8) 计算 θ_{t+1} 13：结束 for

偏差部分。 类似于不等式（25）的证明，我们有

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

方差部分。 类似于不等式（26）的证明，我们有

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

结合偏差部分和方差部分的上界，并令 $\delta = 1/(2T^2)$，我们以至少 $1-1/T$ 的概率得到如下遗憾界，

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

其中 $\bar{\beta}_t = \sqrt{\lambda}c_\mu S + R\sqrt{4\,\text{对数}\ T + 2\,\text{对数}\ 2 + d\,\text{对数}\left(1+\frac{L^2}{\lambda d}\frac{(1-\gamma^{2t})}{(1-\gamma^2)}\right)}$。我们令 $\gamma \ge 1/T$ 和 $\lambda = d/c_\mu^2$，从而得到，

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

当 $P_T < d/(k_\mu c_\mu T)$ 时，我们令 $\gamma = 1-1/T$，并得到一个 $\widetilde{\mathcal{O}}(k_\mu c_\mu^{-1}\sqrt{dT})$ 遗憾界。当 $P_T \ge d/(k_\mu c_\mu T)$ 时，我们将 γ 最优地设为 $1-\gamma = \sqrt{k_\mu c_\mu P_T/(dT)}$，并得到一个 $\widetilde{\mathcal{O}}(k_\mu^{5/4}c_\mu^{-3/4}d^{3/4}P_T^{1/4}T^{3/4})$ 遗憾界。注意，如果 $k_\mu < 1$，我们只需令 $1-\gamma = \sqrt{c_\mu P_T/(dT)}$，此时遗憾界变为 $\widetilde{\mathcal{O}}(k_\mu^2 c_\mu^{-3/4}d^{3/4}P_T^{1/4}T^{3/4})$。 □

C SCB-WeightUCB分析

本节首先在算法4中介绍 SCB-WeightUCB 算法。然后，在附录C.1 中给出 SCB-WeightUCB算法估计误差上界的证明（引理3）。最后，在附录C.2中给出动态遗憾上界的证明（定理3）。

C.1 引理3的证明

证明。根据估计量方程（8），我们知道

$$g_t(\widehat{\theta}_t) = \lambda c_\mu\widehat{\theta}_t + \sum_{s=1}^{t-1}\gamma^{t-s-1}\mu(X_s^\top\widehat{\theta}_t)X_s = \sum_{s=1}^{t-1}\gamma^{t-s-1}r_sX_s, \tag{44}$$

然后根据中值定理，我们知道

$$g_t(\theta_1) - g_t(\theta_2) = G_t(\theta_1, \theta_2)(\theta_1 - \theta_2), \tag{45}$$

其中 $G_t(\theta_1, \theta_2) \triangleq \int_0^1 \nabla g_t(s\theta_2 + (1-s)\theta_1)\mathrm{d}s \in \mathbb{R}^{d\times d}$。注意，对于任意 $\theta \in \Theta$， g_t 的梯度为

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

这显然意味着 mplies $\forall \theta_1, \theta_2 \in \Theta, G_t(\theta_1, \theta_2) \succeq c_\mu V_{t-1}$ 且 $\forall \theta, H_t(\theta) \succeq c_\mu V_{t-1}$，其中 $H_t(\theta)$ 定义为

$$H_t(\theta) \triangleq \lambda c_\mu I_d + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu'(X_s^\top \theta) X_s X_s^\top. \tag{46}$$

根据假设 2，对37应用中值定理 g_t，投影（14）以及引理 14，可得

$$\begin{aligned}
|\mu(\mathbf{x}^\top \widetilde{\theta}_t) - \mu(\mathbf{x}^\top \theta_t)| &\le k_\mu |\mathbf{x}^\top (\widetilde{\theta}_t - \theta_t)| \\
&= k_\mu |\mathbf{x}^\top G_t^{-1}(\theta_t, \widetilde{\theta}_t)(g_t(\widetilde{\theta}_t) - g_t(\theta_t))| \\
&\le k_\mu \|\mathbf{x}\|_{G_t^{-1}(\theta_t, \widetilde{\theta}_t)} \|g_t(\widetilde{\theta}_t) - g_t(\theta_t)\|_{G_t^{-1}(\theta_t, \widetilde{\theta}_t)} \\
&\le k_\mu \|\mathbf{x}\|_{G_t^{-1}(\theta_t, \widetilde{\theta}_t)} \left(\|g_t(\widetilde{\theta}_t) - g_t(\widehat{\theta}_t)\|_{G_t^{-1}(\theta_t, \widetilde{\theta}_t)} + \|g_t(\widehat{\theta}_t) - g_t(\theta_t)\|_{G_t^{-1}(\theta_t, \widetilde{\theta}_t)} \right) \\
&\le \sqrt{1+2S} k_\mu \|\mathbf{x}\|_{G_t^{-1}(\theta_t, \widetilde{\theta}_t)} \left(\|g_t(\widetilde{\theta}_t) - g_t(\widehat{\theta}_t)\|_{H_t^{-1}(\widetilde{\theta}_t)} + \|g_t(\widehat{\theta}_t) - g_t(\theta_t)\|_{H_t^{-1}(\theta_t)} \right) \\
&\le 2\sqrt{1+2S} \frac{k_\mu}{\sqrt{c_\mu}} \|\mathbf{x}\|_{V_{t-1}^{-1}} \|g_t(\widehat{\theta}_t) - g_t(\theta_t)\|_{H_t^{-1}(\theta_t)},
\end{aligned}$$

那么，根据模型假设（13）、函数 g_t （10）以及 g_t（θ_t） （44），可得，

$$\begin{aligned}
g_t(\theta_t) - g_t(\widehat{\theta}_t) &= \lambda c_\mu \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta_t) X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s \\
&= \lambda c_\mu \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta_t) X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_s) + \eta_s) X_s \\
&= \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s + \lambda c_\mu \theta_t - \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s,
\end{aligned}$$

则，由柯西-施瓦茨不等式可得

$$\left| \mu(\mathbf{x}^\top \widetilde{\theta}_t) - \mu(\mathbf{x}^\top \theta_t) \right| \le 2\sqrt{1+2S} \frac{k_\mu}{\sqrt{c_\mu}} \|\mathbf{x}\|_{V_{t-1}^{-1}} \left(E_t + F_t \right), \tag{47}$$

其中

$$E_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{H_t^{-1}(\theta_t)}, \quad F_t = \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_\mu \theta_t \right\|_{H_t^{-1}(\theta_t)}.$$

这两项可以分别进行界定。

引理 8。 对于任意 $t \in [T]$，我们有

$$\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{H_t^{-1}(\theta_t)} \le L \frac{k_\mu}{\sqrt{c_\mu}} \sqrt{d} \sum_{p=1}^{t-1} \gamma^{\frac{t-1}{2}} \sqrt{\frac{\gamma^{-p} - 1}{1-\gamma}} \|\theta_p - \theta_{p+1}\|_2. \tag{48}$$

引理 9。 对于任意 $\delta \in (0,1)$，以至少 $1-\delta$ 的概率，对于所有 $t \in [T]$，我们有

$$\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s - \lambda c_\mu \theta_t\right\|_{H_t^{-1}(\theta_t)} \leq \frac{\sqrt{\lambda c_\mu}}{2m} + \frac{2m}{\sqrt{\lambda c_\mu}}\log\frac{1}{\delta} + \frac{dm}{\sqrt{\lambda c_\mu}}\log\left(1+\frac{L^2 k_\mu(1-\gamma^{2t-2})}{\lambda c_\mu d(1-\gamma^2)}\right) + \frac{2m}{\sqrt{\lambda c_\mu}} d\log(2) + \sqrt{\lambda c_\mu} S, \tag{49}$$

根据不等式（47）、引理 6 和引理 7，以及可行集的有界性假设，对于任意$\mathbf{x} \in \mathcal{X}$， $\delta \in (0,1)$，以至少 $1-\delta$ 的概率，对于所有 $t \in [T]$,

$$\begin{aligned}\left|\mu(\mathbf{x}^\top \widetilde{\theta}_t) - \mu(\mathbf{x}^\top \theta_t)\right| &\leq 2\sqrt{1+2S}\frac{k_\mu}{\sqrt{c_\mu}}\|\mathbf{x}\|_{V_{t-1}^{-1}}\left(L\frac{k_\mu}{\sqrt{c_\mu}}\sqrt{d}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\|\theta_p-\theta_{p+1}\|_2 + \widetilde{\beta}_{t-1}\right), \\ &\leq 2\sqrt{1+2S}\frac{k_\mu}{\sqrt{c_\mu}}\left(L^2\frac{k_\mu}{\sqrt{\lambda c_\mu}}\sqrt{d}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\|\theta_p-\theta_{p+1}\|_2 + \widetilde{\beta}_{t-1}\|\mathbf{x}\|_{V_{t-1}^{-1}}\right),\end{aligned}$$

其中 $\widetilde{\beta}_t \triangleq \frac{\sqrt{\lambda c_\mu}}{2m} + \frac{2m}{\sqrt{\lambda c_\mu}}\log\frac{1}{\delta} + \frac{dm}{\sqrt{\lambda c_\mu}}\log\left(1+\frac{L^2 k_\mu}{\lambda c_\mu d(1-\gamma^{2t})(1-\gamma^2)}\right) + \frac{2m}{\sqrt{\lambda c_\mu}} d\log(2) + \sqrt{\lambda c_\mu} S$ 是 SCB-WeightUCB 中使用的置信半径。因此，我们完成了证明。 □

引理 8 的证明。 由于 $\forall \theta$， $H_t(\theta) \succeq c_\mu V_{t-1}$，我们有

$$\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}(\mu(X_s^\top\theta_t)-\mu(X_s^\top\theta_s))X_s\right\|_{H_t^{-1}(\theta_t)} \leq \frac{1}{\sqrt{c_\mu}}\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}(\mu(X_s^\top\theta_t)-\mu(X_s^\top\theta_s))X_s\right\|_{V_{t-1}^{-1}}.$$

然后使用引理6 即可完成证明。 □

引理 9的证明。 令 $\widetilde{H}_t(\theta) \triangleq \lambda c_\mu \gamma^{-2(t-1)} I_d + \sum_{s=1}^{t-1}\gamma^{-2s}\mu'(X_s^\top\theta)X_s X_s^\top$，其仅在分析中使用 s.

$$\begin{aligned}F_t &= \left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s - \lambda c_\mu\theta_t\right\|_{H_t^{-1}(\theta_t)} \leq \left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}\eta_s X_s\right\|_{H_t^{-1}(\theta_t)} + \|\lambda c_\mu\theta_t\|_{H_t^{-1}(\theta_t)} \\ &\leq \left\|\sum_{s=1}^{t-1}\gamma^{-s}\eta_s X_s\right\|_{\widetilde{H}_t^{-1}(\theta_t)} + \sqrt{\lambda c_\mu}S.\end{aligned}$$

回忆一下， $H_t(\theta) = \lambda c_\mu I_d + \sum_{s=1}^{t-1}\gamma^{t-s-1}\mu'(X_s^\top\theta)X_s X_s^\top$，因此最后一个不等式来自

$$\gamma^{-2(t-1)}H_t(\theta) = \lambda c_\mu\gamma^{-2(t-1)}I_d + \sum_{s=1}^{t-1}\gamma^{-t-s+1}\mu'(X_s^\top\theta)X_s X_s^\top \succeq \lambda c_\mu\gamma^{-2(t-1)}I_d + \sum_{s=1}^{t-1}\gamma^{-2s}\mu'(X_s^\top\theta)X_s X_s^\top = \widetilde{H}_t^{-1}(\theta). \tag{50}$$

根据加权版集中不等式 [Russac 等人，2021，定理 3]，（在定理 6中重新陈述），我们可以得到第一项的界 $\|\sum_{s=1}^{t-1}\gamma^{-s}\eta_s X_s\|_{\widetilde{H}_t^{-1}(\theta_t)}$。首先，根据模型假设（13），我们知道 $\sigma_t^2 = \mathbb{E}[\eta_t^2|\mathcal{F}_t] =$方差$[r_t|\mathcal{F}_t] = \mu''(X_t\theta_t)$，然后令 $w_t = \gamma^{-t}$， $\lambda_t = \lambda c_\mu\gamma^{-2t}$ 便有:

$$\left\|\sum_{s=1}^{t-1}\gamma^{-s}\eta_s X_s\right\|_{\widetilde{H}_t^{-1}(\theta_t)} \leq \frac{\sqrt{\lambda c_\mu}}{2m} + \frac{2m}{\sqrt{\lambda c_\mu}}\log\left(\frac{\det(\widetilde{H}_t)^{1/2}}{\delta(\lambda c_\mu\gamma^{-2(t-1)})^{d/2}}\right) + \frac{2m}{\sqrt{\lambda c_\mu}}d\log(2).$$

然后利用引理 12 并令 $w_{t,\ s} = \gamma^{-2s}\mu'(X_s^\top\theta_t)$ $\lambda_t = \lambda c_\mu\gamma^{-2(t-1)}$，可得 $\det(\widetilde{H}_t)$ 的上界，

$$\det(\widetilde{H}_t) \leq \left(\lambda c_\mu\gamma^{-2(t-1)} + \frac{L^2 k_\mu\sum_{s=1}^{t-1}\gamma^{-2s}}{d}\right)^d,$$

于是，

$$\begin{aligned}\left\|\sum_{s=1}^{t-1}\gamma^{-s}\eta_s X_s\right\|_{\widetilde{H}_t^{-1}(\theta_t)} &\le \frac{\sqrt{\lambda c_\mu}}{2m}+\frac{2m}{\sqrt{\lambda c_\mu}}\log\frac{1}{\delta}+\frac{dm}{\sqrt{\lambda c_\mu}}\log\left(\frac{\lambda c_\mu\gamma^{-2(t-1)}+\frac{L^2k_\mu\sum_{s=1}^{t-1}\gamma^{-2s}}{d}}{\lambda c_\mu\gamma^{-2(t-1)}}\right)+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2)\\ &\le \frac{\sqrt{\lambda c_\mu}}{2m}+\frac{2m}{\sqrt{\lambda c_\mu}}\log\frac{1}{\delta}+\frac{dm}{\sqrt{\lambda c_\mu}}\log\left(1+\frac{L^2k_\mu(1-\gamma^{2t-2})}{\lambda c_\mu d(1-\gamma^2)}\right)+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2).\end{aligned}$$

因此，我们得到 F_t 项的上界。 □

C.2 定理3的证明

证明。令 $X_t^* \triangleq \arg\max_{\mathbf{x}\in\mathcal{X}}\mu(\mathbf{x}^\top\theta_t)$。根据引理2 以及 $X_t^*, X_t \in \mathcal{X}$这一事实，以下各项成立的概率至少为 $1-\delta$,

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

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

根据并集界，以下结论成立的概率至少为 $1-2\delta$： $\forall t\in[T]$

$$\begin{aligned}&\mu(X_t^{*\top}\theta_t)-\mu(X_t^{\top}\theta_t)\\ &\le \mu(X_t^{*\top}\widetilde{\theta}_t)-\mu(X_t^{\top}\widetilde{\theta}_t)+2\sqrt{1+2S}\Bigg(\frac{2L^2k_\mu^2}{c_\mu}\sqrt{\frac{d}{\lambda}}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2\\ &\qquad\qquad+\frac{k_\mu}{\sqrt{c_\mu}}\left(\widetilde{\beta}_{t-1}\|X_t^*\|_{V_{t-1}^{-1}}+\widetilde{\beta}_{t-1}\|X_t\|_{V_{t-1}^{-1}}\right)\Bigg)\\ &\le \frac{4\sqrt{1+2S}L^2k_\mu^2}{c_\mu}\sqrt{\frac{d}{\lambda}}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2+\frac{4\sqrt{1+2S}k_\mu}{\sqrt{c_\mu}}\widetilde{\beta}_{t-1}\|X_t\|_{V_{t-1}^{-1}},\end{aligned}$$

其中最后一步来自臂选择准则（17），满足

$$\mu(X_t^{*\top}\widetilde{\theta}_t)+2\sqrt{1+2S}\frac{k_\mu}{\sqrt{c_\mu}}\widetilde{\beta}_{t-1}\|X_t^*\|_{V_{t-1}^{-1}} \le \mu(X_t^{\top}\widetilde{\theta}_t)+2\sqrt{1+2S}\frac{k_\mu}{\sqrt{c_\mu}}\widetilde{\beta}_{t-1}\|X_t\|_{V_{t-1}^{-1}}.$$

因此，下面的 动态遗憾界以至少 $1-2\delta$ 的概率成立，并可分为两部分，

$$\begin{aligned}R_T &= \sum_{t=1}^{T}\mu(X_t^{*\top}\theta_t)-\mu(X_t^{\top}\theta_t)\\ &\le \underbrace{\frac{4\sqrt{1+2S}L^2k_\mu^2}{c_\mu}\sqrt{\frac{d}{\lambda}}\sum_{t=1}^{T}\sum_{p=1}^{t-1}\gamma^{\frac{t-1}{2}}\sqrt{\frac{\gamma^{-p}-1}{1-\gamma}}\left\|\theta_p-\theta_{p+1}\right\|_2}_{\text{bias part}}+\underbrace{\frac{4\sqrt{1+2S}k_\mu}{\sqrt{c_\mu}}\widetilde{\beta}_T\sum_{t=1}^{T}\|X_t\|_{V_{t-1}^{-1}}}_{\text{variance part}}.\end{aligned}$$

其中 $\widetilde{\beta}_t=\frac{\sqrt{\lambda c_\mu}}{2m}+\frac{2m}{\sqrt{\lambda c_\mu}}\log\frac{1}{\delta}+\frac{dm}{\sqrt{\lambda c_\mu}}\log\left(1+\frac{L^2k_\mu}{\lambda c_\mu d}\frac{(1-\gamma^{2t})}{(1-\gamma^2)}\right)+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2)+\sqrt{\lambda c_\mu}S$ 是置信半径。

现在我们分别推导这两部分的上界。

偏差部分。 类似于不等式（25）的证明，我们有

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

方差部分。 首先使用柯西-施瓦茨不等式，我们知道

$$\frac{4\sqrt{1+2S}k_\mu}{\sqrt{c_\mu}}\widetilde{\beta}_T\sum_{t=1}^{T}\|X_t\|_{V_{t-1}^{-1}}\le\frac{4\sqrt{1+2S}k_\mu}{\sqrt{c_\mu}}\widetilde{\beta}_T\sqrt{T}\sqrt{\sum_{t=1}^{T}\|X_t\|^2_{V_{t-1}^{-1}}}.$$

那么，对于项 $\sqrt{\sum_{t=1}^{T}\|X_t\|^2_{V_{t-1}^{-1}}}$，我们可以直接使用引理 11 对其进行界定，

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

结合偏差部分和方差部分的上界，并令 $\delta=1/(2T)$，我们以至少 $1-1/T$ 的概率得到如下遗憾界，

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

其中 $\widetilde{\beta}_t=\frac{\sqrt{\lambda c_\mu}}{2m}+\frac{2m}{\sqrt{\lambda c_\mu}}\log(2T)+\frac{dm}{\sqrt{\lambda c_\mu}}\log\left(1+\frac{L^2k_\mu}{\lambda c_\mu d(1-\gamma^{2t})(1-\gamma^2)}\right)+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2)+\sqrt{\lambda c_\mu}S$。由于遗憾界中存在一个$T\sqrt{\log(1/\gamma)}$项，这意味着我们不能令 γ 接近 0，因此设 $\gamma\ge1/T$，于是有 $\log(1/\gamma)\le C(1-\gamma)$，其中 $C=\log T/(1-1/T)$。然后，忽略时间范围 T 中的对数因子，并令$\lambda=d\log(T)/c_\mu$，最终得到，

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

当 $P_T<d/(k_\mu T)$（对应较小程度的非平稳性）时，我们只需设置 $\gamma=1-1/T$，即可得到 $\widetilde{\mathcal{O}}(k_\mu c_\mu^{-1/2}d\sqrt{T})$ 遗憾界。此外，对于 $P_T\ge d/(k_\mu T)$ 的非退化情形，我们将折扣因子最优地设为$1-\gamma=\sqrt{k_\mu P_T/(dT)}$，从而得到 $\widetilde{\mathcal{O}}(k_\mu^{5/4}c_\mu^{-1/2}d^{3/4}P_T^{1/4}T^{3/4})$ 遗憾界，至此证明完成。

□

D 分段平稳 SCB

本节研究分段平稳环境下的 SCB，我们的工作是对 [Rus- sac 等人的 2021]直接改进。接下来，我们将首先提出 SCB-PW-WeightUCB 算法，然后分析置信集。最后，我们给出动态遗憾上界的证明。

D.1 SCB-PW-WeightUCB算法

受 Abeille 等人 [2021], 的启发，我们对 Russac 等人 [2021]的方法进行了直接改进。与 Russac 等人 [2021],针对 $D\ge1$的方法类似，定义 $\mathcal{T}(D)=\{1\le t\le T,$，使得对于 $t-D\le s\le t-1\}$，有 $\theta_s=\theta_t$。当 t 距离上一个最近的变化点至少 D 步时， $t\in\mathcal{T}(D)$。但不同之处在于，Russac 等人 [2021] 将 D 视为分析参数，而我们将 D 作为可调算法参数。请注意，这里的 D 不是虚拟窗口大小，而是算法对环境保持平稳持久程度的估计。

估计量。在迭代 过程中 t，我们采用与漂移情形中相同的最大似然估计量，其定义见（8）。

置信集。我们进一步为真实的 θ_t构造置信集。对于 $\delta\in(0,1)$，定义如下:

$$\mathcal{C}_t(\delta)\triangleq\left\{\theta\in\Theta\,\middle|\,\|g_t(\theta)-g_t(\widehat{\theta}_t)\|_{H_t^{-1}(\theta)}\le\rho_t\right\},$$

其中 $\rho_t=\frac{2L^2Sk_\mu}{\sqrt{\lambda c_\mu}}\frac{\gamma^D}{1-\gamma}+\frac{Lm}{\sqrt{\lambda c_\mu}}\frac{\gamma^D}{1-\gamma}+\breve{\beta}_t$ 和 $\breve{\beta}_t=\frac{dm}{\sqrt{\lambda c_\mu}}\log\left(1+\frac{L^2k_\mu}{(1-\gamma^{2D})\lambda c_\mu d(1-\gamma)}\right)+\frac{\sqrt{\lambda c_\mu}}{2m}+\frac{2m}{\sqrt{\lambda c_\mu}}\log\frac{1}{\delta}+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2)+\sqrt{\lambda c_\mu}S$。

算法 5SCB-PW-WeightUCB

输入： 时间范围 T，折扣因子 γ，置信度 δ，正则化项 λ，逆链接函数 μ，参数 S、L 和 m，变化置信度 D

设置 $\widehat{\theta}_0 = \mathbf{0}$ 并计算 k_μ 和 c_μ

对于 $t = 1, 2, 3, ..., T$ **do**

  计算 $(X_t, \widetilde{\theta}_t) = \arg\max_{\mathbf{x}\in\mathcal{X}, \theta\in\mathcal{C}_t(\delta)} \mu(\mathbf{x}^\top\theta)$

  选择 X_t 并获得奖励 r_t

  计算 θ_{t+1} 根据 (8)

结束循环

引理 10. 对于任意 $\delta \in (0, 1)$，以至少 $1-\delta$ 的概率，有 $\forall t \in \mathcal{T}(D)$ $\theta_t \in \mathcal{C}_t(\delta)$。

$$\mathcal{C}_t(\delta) = \left\{\theta \in \Theta \mid \|g_t(\theta) - g_t(\widehat{\theta}_t)\|_{H_t^{-1}(\theta)} \le \frac{2L^2 S k_\mu}{\sqrt{\lambda c_\mu}} \frac{\gamma^D}{1-\gamma} + \frac{Lm}{\sqrt{\lambda c_\mu}} \frac{\gamma^D}{1-\gamma} + \breve{\beta}_t \right\},$$

where $\breve{\beta}_t = \frac{\sqrt{\lambda c_\mu}}{2m} + \frac{2m}{\sqrt{\lambda c_\mu}} \log\frac{1}{\delta} + \frac{dm}{\sqrt{\lambda c_\mu}} \log\left(1 + \frac{L^2 k_\mu (1-\gamma^{2D})}{\lambda c_\mu d(1-\gamma)}\right) + \frac{2m}{\sqrt{\lambda c_\mu}} d\log(2) + \sqrt{\lambda c_\mu} S.$

引理10的证明见附录 D.2。

选择准则。 前文针对漂移情形讨论的算法使用基于奖励的选择准则。但在此，我们采用如下基于参数的选择准则，

$$(X_t, \widetilde{\theta}_t) = \underset{\mathbf{x}\in\mathcal{X}, \theta\in\mathcal{C}_t(\delta)}{\arg\max} \mu(\mathbf{x}^\top\theta). \tag{51}$$

基于参数的选择准则与基于奖励的选择准则之间的主要区别，在 Abeille 等人 [2021]的第 3.2 节中进行了讨论。整体算法总结于算法5中。

D.2 引理10的证明

证明。 根据模型假设（13）、函数 g_t (10) 和 $g_t(\theta_t)$ (44)，我们有：

$$\begin{aligned}
g_t(\theta_t) - g_t(\widehat{\theta}_t) &= \lambda c_\mu \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta_t) X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} r_s X_s \\
&= \lambda c_\mu \theta_t + \sum_{s=1}^{t-1} \gamma^{t-s-1} \mu(X_s^\top \theta_t) X_s - \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_s) + \eta_s) X_s \\
&= \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s + \lambda c_\mu \theta_t - \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s.
\end{aligned}$$

那么，

$$\begin{aligned}
\|g_t(\theta_t) - g_t(\widehat{\theta}_t)\|_{H_t^{-1}(\theta_t)} &= \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s + \lambda c_\mu \theta_t - \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{H_t^{-1}(\theta_t)} \\
&\le \left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{H_t^{-1}(\theta_t)} + \left\| \lambda c_\mu \theta_t - \sum_{s=1}^{t-1} \gamma^{t-s-1} \eta_s X_s \right\|_{H_t^{-1}(\theta_t)} \\
&\le \underbrace{\left\| \sum_{s=1}^{t-1} \gamma^{t-s-1} (\mu(X_s^\top \theta_t) - \mu(X_s^\top \theta_s)) X_s \right\|_{H_t^{-1}(\theta_t)}}_{\texttt{term (a)}} + \underbrace{\left\| \sum_{s=1}^{t-D-1} \gamma^{t-s-1} \eta_s X_s \right\|_{H_t^{-1}(\theta_t)}}_{\texttt{term (b)}} \\
&\quad + \underbrace{\left\| \sum_{s=t-D}^{t-1} \gamma^{t-s-1} \eta_s X_s - \lambda c_\mu \theta_t \right\|_{H_t^{-1}(\theta_t)}}_{\texttt{term (c)}}.
\end{aligned}$$

条件（a）。由于 $t \in \mathcal{T}(D)$，我们有

$$\begin{aligned}\left\|\sum_{s=1}^{t-1}\gamma^{t-s-1}(\mu(X_s^\top\theta_t)-\mu(X_s^\top\theta_s))X_s\right\|_{H_t^{-1}(\theta_t)} &= \left\|\sum_{s=1}^{t-D-1}\gamma^{t-s-1}(\mu(X_s^\top\theta_t)-\mu(X_s^\top\theta_s))X_s\right\|_{H_t^{-1}(\theta_t)}\\ &\le \left\|\sum_{s=1}^{t-D-1}\gamma^{t-s-1}k_\mu X_s^\top(\theta_t-\theta_s)X_s\right\|_{H_t^{-1}(\theta_t)}\\ &\le \sum_{s=1}^{t-D-1}\gamma^{t-s-1}k_\mu\|X_s\|_2\|(\theta_t-\theta_s)\|_2\,\|X_s\|_{H_t^{-1}(\theta_t)}\\ &\le \frac{2L^2Sk_\mu}{\sqrt{\lambda c_\mu}}\frac{\gamma^D}{1-\gamma}.\end{aligned}$$

项 (b)。

$$\left\|\sum_{s=1}^{t-D-1}\gamma^{t-s-1}\eta_s X_s\right\|_{H_t^{-1}(\theta_t)} \le \sum_{s=1}^{t-D-1}\gamma^{t-s-1}m\,\|X_s\|_{H_t^{-1}(\theta_t)} \le \frac{Lm}{\sqrt{\lambda c_\mu}}\sum_{s=1}^{t-D-1}\gamma^{t-s-1} \le \frac{Lm}{\sqrt{\lambda c_\mu}}\frac{\gamma^D}{1-\gamma}.$$

项 (c)。令 $\widetilde{H}_{t-D:t}(\theta)=\lambda c_\mu\gamma^{-2(t-1)}I_d+\sum_{s=t-D}^{t-1}\gamma^{-2s}\mu'(X_s^\top\theta)X_sX_s^\top$

$$\begin{aligned}\left\|\sum_{s=t-D}^{t-1}\gamma^{t-s-1}\eta_sX_s-\lambda c_\mu\theta_t\right\|_{H_t^{-1}(\theta_t)} &\le \left\|\sum_{s=t-D}^{t-1}\gamma^{t-s-1}\eta_sX_s\right\|_{H_t^{-1}(\theta_t)}+\sqrt{\lambda c_\mu}S\\ &\le \left\|\sum_{s=t-D}^{t-1}\gamma^{-s}\eta_sX_s\right\|_{\widetilde{H}_t^{-1}(\theta_t)}+\sqrt{\lambda c_\mu}S\\ &\le \left\|\sum_{s=t-D}^{t-1}\gamma^{-s}\eta_sX_s\right\|_{\widetilde{H}_{t-D:t}^{-1}(\theta_t)}+\sqrt{\lambda c_\mu}S.\end{aligned}$$

我们已在 (50) 中证明 $\gamma^{-2(t-1)}H_t(\theta)\succeq\widetilde{H}_t^{-1}(\theta)$，显然 $\widetilde{H}_t(\theta)\succeq\widetilde{H}_{t-1:t}(\theta)$。接下来，我们需要使用自归一化界 [Russac 等人，2021，定理 3]，（在定理6中重述）来界定项 $\|\sum_{s=t-D}^{t-1}\gamma^{-s}\eta_sX_s\|_{\widetilde{H}_{t-D:t}^{-1}(\theta_t)}$，令 $w_t=\gamma^{-t}$，$\lambda_t=\lambda c_\mu\gamma^{-2t}$，于是有

$$\begin{aligned}&\left\|\sum_{s=t-D}^{t-1}\gamma^{-s}\eta_sX_s\right\|_{\widetilde{H}_{t-D:t}^{-1}(\theta_t)}\\ \le&\frac{\sqrt{\lambda c_\mu}}{2m}+\frac{2m}{\sqrt{\lambda c_\mu}}\log\left(\frac{\det\left(\widetilde{H}_{t-D:t}\right)^{1/2}}{\delta(\lambda c_\mu\gamma^{-2(t-1)})^{d/2}}\right)+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2)\\ \le&\frac{\sqrt{\lambda c_\mu}}{2m}+\frac{2m}{\sqrt{\lambda c_\mu}}\log\frac{1}{\delta}+\frac{dm}{\sqrt{\lambda c_\mu}}\log\left(\frac{\lambda c_\mu\gamma^{-2(t-1)}+\frac{L^2k_\mu\sum_{s=t-D}^{t}\gamma^{-2s}}{d}}{\lambda c_\mu\gamma^{-2(t-1)}}\right)+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2)\\ \le&\frac{\sqrt{\lambda c_\mu}}{2m}+\frac{2m}{\sqrt{\lambda c_\mu}}\log\frac{1}{\delta}+\frac{dm}{\sqrt{\lambda c_\mu}}\log\left(1+\frac{L^2k_\mu(1-\gamma^{2D})}{\lambda c_\mu d(1-\gamma)}\right)+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2).\end{aligned}$$

令 $\breve{\beta}_t \triangleq \frac{\sqrt{\lambda c_\mu}}{2m}+\frac{2m}{\sqrt{\lambda c_\mu}}\log\frac{1}{\delta}+\frac{dm}{\sqrt{\lambda c_\mu}}\log\left(1+\frac{L^2k_\mu}{\lambda c_\mu d}\frac{(1-\gamma^{2D})}{(1-\gamma)}\right)+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2)+\sqrt{\lambda c_\mu}S$，最终可得，

$$\|g_t(\theta_t)-g_t(\widehat{\theta}_t)\|_{H_t^{-1}(\theta_t)} \le \frac{2L^2Sk_\mu}{\sqrt{\lambda c_\mu}}\frac{\gamma^D}{1-\gamma}+\frac{Lm}{\sqrt{\lambda c_\mu}}\frac{\gamma^D}{1-\gamma}+\breve{\beta}_t,$$

由此完成证明。 □

D.3 定理4的证明

证明。设 $R_t = \mu(X_t^{*\top}\theta_t) - \mu(X_t^\top\theta_t)$

$$R_T = \sum_{t=1}^{T} R_t = \sum_{t\notin\mathcal{T}(D)} R_t + \sum_{t\in\mathcal{T}(D)} R_t = \Gamma_T D + \sum_{t\in\mathcal{T}(D)} R_t.$$

对于 $t \in \mathcal{T}(D)$，根据选择准则（51），

$$\begin{aligned}
R_t &= \mu(X_t^{*\top}\theta_t) - \mu(X_t^\top\theta_t) \\
&\le \mu(X_t^\top\widetilde{\theta}_t) - \mu(X_t^\top\widehat{\theta}_t) + \mu(X_t^\top\widehat{\theta}_t) - \mu(X_t^\top\theta_t) \\
&\le \alpha(X_t,\widetilde{\theta}_t,\widehat{\theta}_t)\left|X_t^\top\left(\widetilde{\theta}_t-\widehat{\theta}_t\right)\right| + \alpha(X_t,\theta_t,\widehat{\theta}_t)\left|X_t^\top\left(\theta_t-\widehat{\theta}_t\right)\right| \\
&\le \sqrt{1+2S}\Bigg(\alpha(X_t,\widetilde{\theta}_t,\widehat{\theta}_t)\,\|X_t\|_{G_t^{-1}(\widetilde{\theta}_t,\widehat{\theta}_t)}\left\|g_t(\widetilde{\theta}_t)-g_t(\widehat{\theta}_t)\right\|_{H_t^{-1}(\widetilde{\theta}_t)} \\
&\qquad + \alpha(X_t,\theta_t,\widehat{\theta}_t)\,\|X_t\|_{G_t^{-1}(\theta_t,\widehat{\theta}_t)}\left\|g_t(\theta_t)-g_t(\widehat{\theta}_t)\right\|_{H_t^{-1}(\theta_t)}\Bigg).
\end{aligned}$$

其中 $\alpha(\mathbf{x}\ \theta_1,\theta_2) = \int_0^1 \mu'(v\mathbf{x}^\top\theta_2 + (1-v)x^\top\theta_1)\mathrm{d}v$，且倒数第二个不等式由中值定理得出$\mu(\mathbf{x}^\top\theta_1)-\mu(\mathbf{x}^\top\theta_2) = \alpha(\mathbf{x}\ \theta_1,\theta_2)(\mathbf{x}^\top\theta_1-\mathbf{x}^\top\theta_2)$。由于 $\widetilde{\theta}_t \in \mathcal{C}_t(\delta)$，且以至少 $1-\delta$ 的概率，$\forall t\in[T]$，$\theta_t \in \mathcal{C}_t(\delta)$，再根据并集界，以下动态遗憾界以至少 $1-\delta$ 的概率成立，

$$\begin{aligned}
\sum_{t\in\mathcal{T}(D)} R_t &\le \sum_{t\in\mathcal{T}(D)} \sqrt{1+2S}\left(\alpha(X_t,\widetilde{\theta}_t,\widehat{\theta}_t)\,\|X_t\|_{G_t^{-1}(\widetilde{\theta}_t,\widehat{\theta}_t)}\,\rho_t + \alpha(X_t,\theta_t,\widehat{\theta}_t)\,\|X_t\|_{G_t^{-1}(\theta_t,\widehat{\theta}_t)}\,\rho_t\right) \\
&\le \sqrt{1+2S}\rho_T\left(\sum_{t\in\mathcal{T}(D)}\alpha(X_t,\widetilde{\theta}_t,\widehat{\theta}_t)\,\|X_t\|_{G_t^{-1}(\widetilde{\theta}_t,\widehat{\theta}_t)} + \sum_{t\in\mathcal{T}(D)}\alpha(X_t,\theta_t,\widehat{\theta}_t)\,\|X_t\|_{G_t^{-1}(\theta_t,\widehat{\theta}_t)}\right).
\end{aligned}$$

现在我们尝试推导项 $\sum_{t\in\mathcal{T}(D)}\alpha(X_t,\theta_t,\theta_t)\,\|X_t\|_{G_t^{-1}(\widetilde{\theta}_t,\widehat{\theta}_t)}$ 的上界。

根据 g_t (10的定义，我们有

$$\begin{aligned}
g_t(\theta_1) - g_t(\theta_2) &= \lambda c_\mu(\theta_1-\theta_2) + \sum_{s=1}^{t-1}\gamma^{t-s-1}(\mu(X_s^\top\theta_1)-\mu(X_s^\top\theta_2))X_s \\
&= \lambda c_\mu(\theta_1-\theta_2) + \sum_{s=1}^{t-1}\gamma^{t-s-1}\alpha(X_s,\theta_1,\theta_2)X_s^\top X_s(\theta_1-\theta_2) \\
&= \left(\lambda c_\mu + \sum_{s=1}^{t-1}\gamma^{t-s-1}\alpha(X_s,\theta_1,\theta_2)X_s^\top X_s\right)(\theta_1-\theta_2).
\end{aligned} \tag{52}$$

那么，根据 G_t (45) 的定义，我们知道 $G_t(\theta_1,\theta_2) = \lambda c_\mu + \sum_{s=1}^{t-1}\gamma^{t-s-1}\alpha(X_s,\theta_1,\theta_2)X_s^\top X_s$. 这意味着$G_t(\widetilde{\theta}_t,\widehat{\theta}_t) = \lambda c_\mu I_d + \sum_{s=1}^{t-1}\gamma^{t-s-1}\alpha(X_s,\widetilde{\theta}_t,\widehat{\theta}_t)X_sX_s^\top$，若令 $\widetilde{X}_s = \sqrt{\alpha(X_s,\widetilde{\theta}_t,\widehat{\theta}_t)}X_s$，则

$$\begin{aligned}
\sum_{t\in\mathcal{T}(D)}\alpha(X_t,\widetilde{\theta}_t,\widehat{\theta}_t)\,\|X_t\|_{G_t^{-1}(\widetilde{\theta}_t,\widehat{\theta}_t)} &\le \sqrt{\sum_{t=1}^{T}\alpha(X_t,\widetilde{\theta}_t,\widehat{\theta}_t)}\sqrt{\sum_{t=1}^{T}\alpha(X_t,\widetilde{\theta}_t,\widehat{\theta}_t)\,\|X_t\|^2_{G_t^{-1}(\widetilde{\theta}_t,\widehat{\theta}_t)}} \\
&\le \sqrt{k_\mu T}\sqrt{\sum_{t=1}^{T}\left\|\widetilde{X}_t\right\|^2_{G_t^{-1}(\widetilde{\theta}_t,\widehat{\theta}_t)}}.
\end{aligned} \tag{53}$$

然后，对于项 $\sqrt{\sum_{t=1}^{T}\|\widetilde{X}_t\|^2_{G_t^{-1}(\widetilde{\theta}_t,\widehat{\theta}_t)}}$，我们可以直接使用引理 11 对其进行界定，

$$\sqrt{k_\mu T}\sqrt{\sum_{t=1}^{T}\left\|\widetilde{X}_t\right\|^2_{G_t^{-1}(\widetilde{\theta}_t,\widehat{\theta}_t)}} \le \sqrt{2k_\mu\max\{1,L^2k_\mu/(\lambda c_\mu)\}dT}\sqrt{T\log\frac{1}{\gamma}+\log\left(1+\frac{L^2k_\mu}{\lambda c_\mu d(1-\gamma)}\right)}.$$

^我们可以用同样的方法对项 $\sum_{t\in\mathcal{T}(D)}\alpha(X_t,\theta_t,\theta_t)\|X_t\|_{G_t^{-1}(\theta_t,\widehat{\theta}_t)}$ 进行界定，得到：

$$\sum_{t\in\mathcal{T}(D)}\alpha(X_t,\theta_t,\widehat{\theta}_t)\|X_t\|_{G_t^{-1}(\theta_t,\widehat{\theta}_t)}\le\sqrt{2k_\mu\max\{1,L^2k_\mu/(\lambda c_\mu)\}dT}\sqrt{T\log\frac{1}{\gamma}+\log\left(1+\frac{L^2k_\mu}{\lambda c_\mu d(1-\gamma)}\right)}.$$

结合这两 个界并令 $\delta=1/T$，我们以至少 $1-1/T$ 的概率得到以下遗憾界 ，

$$R_T\le\Gamma_T D+2\sqrt{1+2S}\rho_T\sqrt{2k_\mu\max\{1,L^2k_\mu/(\lambda c_\mu)\}dT}\sqrt{T\log\frac{1}{\gamma}+\log\left(1+\frac{L^2k_\mu}{\lambda c_\mu d(1-\gamma)}\right)},$$

其中 $\rho_t=\frac{2L^2Sk_\mu}{\sqrt{\lambda c_\mu}}\frac{\gamma^D}{1-\gamma}+\frac{Lm}{\sqrt{\lambda c_\mu}}\frac{\gamma^D}{1-\gamma}+\breve{\beta}_t$ 和 $\breve{\beta}_t=\frac{dm}{\sqrt{\lambda c_\mu}}\log\left(1+\frac{L^2k_\mu(1-\gamma^{2D})}{\lambda c_\mu d(1-\gamma)}\right)+\sqrt{\lambda c_\mu}+\frac{2m}{\sqrt{\lambda c_\mu}}+\frac{2m}{\sqrt{\lambda c_\mu}}\log(T)+\frac{2m}{\sqrt{\lambda c_\mu}}d\log(2)+\sqrt{\lambda c_\mu}S$。由于遗憾界中存在一个 $T\sqrt{\log(1/\gamma)}$ 项，这意味着我们不能令 γ 接近 0，因此我们设 $\gamma\ge1/2$，于是有 $\log(1/\gamma)\le2\log(2)(1-\gamma)$。接着，我们设 $D=\log(T)/\log(1/\gamma)$，注意到 $0<1/\gamma-1<1$，并利用对于 $0<x<1$，$\log(1+x)\ge x/2$，可得

$$\log\frac{1}{\gamma}=\log(1+1/\gamma-1)\ge\frac{1-\gamma}{2\gamma}.$$

因此，我们有 $D\le\frac{2\gamma}{\log(T)1-\gamma}$。然后，忽略时间范围 T 中的对数因子，并令 $\lambda=d\log(T)/c_\mu$，最终可得：

$$\begin{aligned}R_T&\le\widetilde{\mathcal{O}}\left(\frac{1}{1-\gamma}\Gamma_T+\left(\frac{1}{\sqrt{d}}\frac{1}{1-\gamma}\frac{1}{T}+\sqrt{d}\right)\sqrt{d(1-\gamma)T}\right)\\&\le\widetilde{\mathcal{O}}\left(\frac{1}{1-\gamma}\Gamma_T+\frac{1}{\sqrt{1-\gamma}}+d\sqrt{(1-\gamma)T}\right).\end{aligned}$$

当 $\Gamma_T<d/\sqrt{T}$ （对应于较小程度的非平稳性）时，我们只需设置 $\gamma=1-1/T$，即可得到一个$\widetilde{\mathcal{O}}(d\sqrt{T})$遗憾界。此外，对于 $\Gamma_T>d/\sqrt{T}$ 的非退化情形，我们将折扣因子最优地设置为 $1-\gamma=(\Gamma_T/(dT))^{2/3}$，并得到一个 $\widetilde{\mathcal{O}}(d^{2/3}\Gamma_T^{1/3}T^{2/3})$ 动态遗憾界，从而完成证明。 □

E 技术引理

在本节中，我们给出若干有用的引理，主要涉及加权自归一化集中、加权版本势引理，以及自协调性质的一些推论。

E.1 加权版本的自归一化集中不等式

定理 5 （向量值鞅的加权自归一化界 [Russac 等人，2019，定理 1]）。设 $\{\mathcal{F}_t\}_{t=0}^{\infty}$ 为一个滤过，$\{\eta_t\}_{t=0}^{\infty}$ 为一个实值随机过程，满足 η_t 是 $\mathcal{F}_t$-可测的，且 η_t 在条件上是 R-次高斯的，其中 $R\ge0$，满足

$$\forall\lambda\in\mathbb{R},\mathbb{E}\left[\exp(\lambda\eta_t)\mid X_{1:t},\eta_{1:t-1}\right]\le\exp\left(\frac{\lambda^2R^2}{2}\right).$$

令 $\{X_t\}_{t=1}^{\infty}$ 为一个取值于 $\mathbb{R}^d$ 的随机过程，满足 X_t 是 $\mathcal{F}_{t-1}$-可测的。对于任意 $t\ge0$，定义

$$\widetilde{V}_t=\mu_tI_d+\sum_{s=1}^{t}w_s^2X_sX_s^\top,\qquad S_t=\sum_{s=1}^{t}w_s\eta_sX_s.$$

其中 $\forall s\ge0$、$t\ge0$、w_s、$\mu_t>0$。那么，对于任意 $\delta>0$，以至少 $1-\delta$ 的概率，有

$$\forall t\ge0,\|S_t\|_{\widetilde{V}_t^{-1}}\le R\sqrt{2\log\frac{1}{\delta}+d\log\left(1+\frac{L^2\sum_{s=1}^{t}w_s^2}{d\mu_t}\right)}.$$

定理6 (Russac 等人的定理3 [2021])。 设 t 为一个固定时刻。设 $\{\mathcal{F}_t\}_{t=0}^{\infty}$ 为一个滤过。设 $\{X_t\}_{t=0}^{\infty}$为定义在 $\mathbb{R}^d$ 上的随机过程，满足 X_t 是 $\mathcal{F}_{t-1}$ 可测的，且 $\|X_t\|_2\le1$。设 $\{\eta_t\}_{t=0}^{\infty}$ 为一个鞅差序列，满足 η_t 是 $\mathcal{F}_t$ 可测的。假设权重单调不减且严格为正，并且时间范围

已知。此外，假设在给定 $\mathcal{F}_t$ 的条件下，我们有 $|\eta_t| \leq m$ 几乎必然成立。设 $\{\lambda_t\}_{u=0}^{\infty}$ 为一个确定性的正则化项序列，并记 $\sigma_t^2 = \mathbb{E}\left[\eta_t^2 \mid \mathcal{F}_t\right]$。设 $\widetilde{H}_t = \sum_{s=1}^{t-1} w_s^2 \sigma_s^2 X_s X_s^\top + \lambda_{t-1}\mathbf{I}_d$ 和 $S_t = \sum_{s=1}^{t-1} w_s \eta_s X_s$，则对于任意 $\delta \in (0,1]$,，其成立概率至少为 $1-\delta$,

$$\forall t \geq 0, \|S_t\|_{\widetilde{H}_t^{-1}} \leq \frac{\sqrt{\lambda_{t-1}}}{2mw_{t-1}} + \frac{2mw_{t-1}}{\sqrt{\lambda_{t-1}}} \log\left(\frac{\det\left(\widetilde{H}_t\right)^{1/2}}{\delta \lambda_{t-1}^{d/2}}\right) + \frac{2mw_{t-1}}{\sqrt{\lambda_{t-1}}} d\log(2).$$

E.2 加权版本势引理

引理 11 (加权版本势引理 [Faury 等人，2021，引理 8])。假设 $V_t = \sum_{s=1}^{t} \gamma^{t-s} X_s X_s^\top + \lambda I_d$，$V_0 = \lambda I_d$，$\gamma \in (0,1]$ 且 $\|X_t\|_2 \leq L$ 对所有 $t \geq 1$均成立，则有以下不等式成立：

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

引理 12 (行列式不等式)。令 $V_t = \sum_{s=1}^{t} w_{t,\ s} X_s X_s^\top + \lambda_t I_d$, $V_0 = \lambda_0 I_d$。假设 $\|\mathbf{x}\|_2 \leq L$ 且有：

$$\det(V_t) \leq \left(\lambda_t + \frac{L^2 \sum_{s=1}^{t} w_{t,s}}{d}\right)^d.$$

证明。现在我们有 $V_t = \sum_{s=1}^{t} w_{t,\ s} X_s X_s^\top + \lambda_t I_d$，对两边取迹，得到 $\mathrm{Tr}(V_t)$ 的上界

$$\mathrm{Tr}(V_t) = \mathrm{Tr}(\lambda_t I_d) + \sum_{s=1}^{t} w_{t,s} \mathrm{Tr}\left(X_s X_s^\top\right) = \lambda_t d + \sum_{s=1}^{t} w_{t,s} \|X_s\|_2^2 \leq \lambda_t d + L^2 \sum_{s=1}^{t} w_{t,s}. \tag{54}$$

根据行列式的定义以及 $\mathrm{Tr}(V_t)$ 的上界(54)，我们可以得到 $\det(V_t)$ 的上界，

$$\det(V_t) = \prod_{i=1}^{d} \lambda_i \leq \left(\frac{\sum_{i=1}^{d} \lambda_i}{d}\right)^d = \left(\frac{\mathrm{Tr}(V_t)}{d}\right)^d \leq \left(\lambda_t + \frac{L^2 \sum_{s=1}^{t} w_{t,s}}{d}\right)^d.$$

□

E.3 自协调性质

根据（逆）链接函数 $\mu(\cdot)$的广义自协调性质，我们有以下引理，稍后将用它来推导引理 14。

Lem引理 13 （Faury 等人 [2020]的引理 9）。对于任意 z_1，$z_2 \in \mathbb{R}$，有以下不等式：

$$\mu'(z_1)\frac{1-\exp(-|z_1-z_2|)}{|z_1-z_2|} \leq \int_0^1 \mu'(z_1 + v(z_2-z_1))\mathrm{d}v \leq \mu'(z_1)\frac{\exp(|z_1-z_2|)-1}{|z_1-z_2|}.$$

此外，$\int_0^1 \mu'(z_1 + v(z_2-z_1))\mathrm{d}v \geq \mu'(z_1)(1+|z_1-z_2|)^{-1}$。

以下引理给出了 Faury 等人 [2020] 引理 10 的加权版本，它可以　　很容易证明 n.

引理 14. 其中 G_t 定义见 (45)，且 H_t 定义见 (46)，下列不等式成立

$$\forall \theta_1, \theta_2 \in \Theta, \quad G_t(\theta_1,\theta_2) \geq (1+2S)^{-1} H_t(\theta_1), \quad G_t(\theta_1,\theta_2) \geq (1+2S)^{-1} H_t(\theta_2).$$