arXiv:2603.03843v1 [stat.ML] 4 Mar 2026

基于不变性的动态遗憾最小化

Margherita Lazzaretto[1]*、Jonas Peters[2] 和 Niklas Pfister[3]

[1]哥本哈根大学， [2]苏黎世联邦理工学院， [3]Lakera Zürich

摘要

我们研究随机非平稳线性老虎机问题，其中连接上下文与奖励的线性参数会随时间变化。现有算法通过逐步丢弃或降低过去数据的权重来局部化策略，实际上缩短了可用于学习的时间范围。然而，在许多情境下，历史数据仍可能包含有关奖励模型的部分信息。我们假设奖励模型可分解为平稳分量和非平稳分量，并据此提出一种在适应变化的同时利用此类数据的方法。基于这一假设，我们引入 ISD-linUCB，这是一种利用过去数据学习奖励模型中不变性、随后加以利用以提升在线性能的算法。我们通过理论和实证结果表明，利用不变性可以降低问题维度；当有足够的历史数据可用时，这能在快速变化的环境中显著改善遗憾表现。

1 引言

随机上下文老虎机刻画了一种在线决策过程：智能体在$T \in \mathbb{N}$ 轮次中，根据上下文信息依次选择动作（例如，参见 Lattimore 和 Szepesvári，2020年）。智能体的目标是学习一种决策策略，通过平衡对新动作的探索和对已获取知识的利用，以最大化累积奖励。老虎机算法为不确定条件下的决策问题提供了最简单的框架；在这类问题中，智能体的动作不会影响环境，因此老虎机算法常被用作更复杂模型（如强化学习）的起点。随机上下文老虎机算法的设计依赖于一些假设，这些假设规定了智能体学习所针对的可能奖励分布类别。我们考虑线性时变奖励函数的情形，其中假设期望奖励与某个 p维上下文-动作特征呈线性关系。

遗憾是衡量老虎机算法性能的主要指标。它对应于最优（预言机）动作序列的累积奖励与老虎机算法所选动作的累积奖励之间的差异。线性老虎机算法的分析涉及研究关于时间范围 T 和维度p的遗憾有限样本上下界。标准设定假设智能体行动所处的环境是平稳的，这一设定在文献中得到了广泛研究（例如，参见Lattimore 和 Szepesvári，2020年；老虎机文献的详细综述见附录A）。Dani等人（2008年）给出了平稳随机线性老虎机问题遗憾的下界 $\Omega(p\sqrt{T})$。文献中提出了不同的算法，其遗憾上界与该下界仅相差对数因子，例如采用置信上界（UCB）策略的算法（Auer，2002年；

*电子邮箱：mala@math.ku.dk

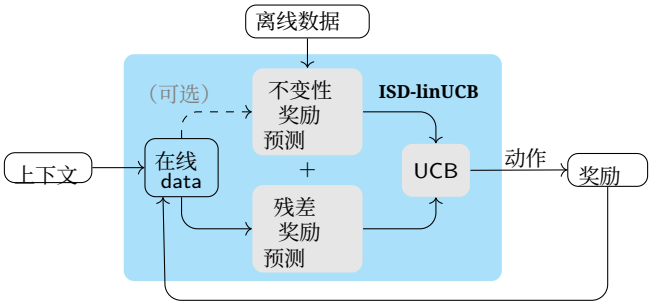

图 1：ISD-linUCB 利用历史数据改进奖励预测，并将其用于置信上界策略。

Dani等人，2008年；Abbasi-Yadkori 等人，2011年）。近期研究放宽了平稳性假设。在这类设定中，对非平稳老虎机算法的评估还会考虑变化预算 B_T，用于衡量环境在 T 轮次中的变化程度。当底层奖励函数随时间变化时，学习并更新可靠的策略可能颇具挑战，因为算法需要持续探索上下文-动作空间，以检测变化并适应变化。Cheung 等人（2019）通过使用滑动窗口正则化最小二乘估计器来处理非平稳性；Russac 等人（2019）则对过去的观测赋予折扣因子权重；Zhao 等人（2020）提出定期重启标准线性老虎机算法，重启间隔取决于 B_T。所有这些算法在 p、T 和 B_T 方面都实现了相同的遗憾上界，表示为 $\tilde{O}(p^{\frac{7}{8}}T^{\frac{3}{4}}B_T^{\frac{1}{4}})$。Zhao 等人（2020）的方法特别表明，与平稳情形相比，额外遗憾源于假设线性参数固定，并因此产生了取决于重启间隔之间参数变化的额外损失。

为更快速地适应环境变化这一总体目标，我们研究非平稳奖励函数的某些部分是否在所有轮次中保持不变，以便利用更早收集的数据，而不是对其降权或将其完全丢弃。例如，不变信息有助于学习选择最坏情形最优动作的策略（参见 Saengkyongam 等人，2023年）。由于纯粹的不变策略在某些特定时刻可能并非最优，我们希望学习一种能够在线更新的不变策略。为此，我们采用Lazzaretto 等人（2025年）提出的 ISD 框架，用于线性非平稳情形下的回归。ISD 将时变参数的学习分解为两个更低维的分量，其中一个分量在时间上保持不变，其维度为 $p^{\text{inv}} < p$：这使我们能够利用所有可用数据进行不变分量估计，从而降低预测误差。所提算法的示意图如图1所示。

我们的贡献如下。

(i) 我们提出一种实用的新型线性上下文多臂老虎机算法（ISD-linUCB），该算法利用 ISD 框架从历史数据中估计不变分量，从而将在线适应限制在维度更低的残差子空间中。(ii) 我们证明，遗憾界随残差维度（$p - p^{\text{inv}}$）而非 p 增长；在有充足历史数据时，这能显著改善算法在快速变化的环境中的表现。

1.1 符号说明

对于所有 $n \in \mathbb{N}$，定义 $[n] := \{1, \ldots, n\}$ 和 $[-n] := \{-n, \ldots, -1\}$。对于所有线性子空间 $\mathcal{S} \subseteq \mathbb{R}^p$，记 $\Pi^{\mathcal{S}} \in \mathbb{R}^{p \times p}$ 为到 $\mathcal{S}$ 的正交投影矩阵，记 $\mathcal{S}^{\perp}$ 为 $\mathcal{S}$ 在 $\mathbb{R}^p$中的正交补。对于所有 $x \in \mathbb{R}^p$ 以及所有半正定矩阵$A \in \mathbb{R}^{p \times p}$，定义范数 $\|x\|_A := \sqrt{x^\top A x}$。进一步地，记 $\lambda_{\min}(A)$ 和 $\lambda_{\max}(A)$ 分别为 A的最小和最大特征值。对于所有 $p \in \mathbb{N}$，记 I_p 为p维单位矩阵。我们用 $O(\cdot)$ 表示忽略多对数因子的渐近界，即若存在常数 $c > 0$ 和 $k \geq 0$，使得对于充分大的 n，有 $b_n = \tilde{O}(a_n)$ $b_n \leq c a_n \log(n)^k$，则记为 $O(\cdot)$。

2 问题设定

在线性随机上下文老虎机中，智能体在轮次 $t \in [T]$ 中依次观测上下文 $X_t \in \mathcal{X} \subseteq \mathbb{R}^p$，其中每个 X_t 都独立于先前的上下文抽取，且服从分布$\mathbb{P}_{X_t}$。给定 X_t 后，智能体选择一个动作 $a_t \in \mathcal{A} = [K]$，并获得含噪奖励 $R_t^{a_t}$。假设所有 $a \in \mathcal{A}$ 的奖励均满足

$$R_t^a = \varphi(X_t, a)^\top \gamma_{0,t} + \epsilon_t \tag{1}$$

其中 φ： $\mathcal{X} \times \mathcal{A} \to \mathbb{R}^p$ 是已知的上下文-动作特征映射， ϵ_t 是相对于 σ-代数 $\mathcal{F}_t = \sigma$（($\varphi(X_\tau$, a_τ) $R_\tau^{a_\tau})_{\tau \in [t-1]}, \varphi(X_t, a_t)$) 条件零均值的 σ-次高斯噪声变量；$\gamma_{0,t} \in \mathbb{R}^p$ 是未知的线性参数。与标准线性老虎机设定（例如，Lattimore 和 Szepesvári，2020年）一致，我们假设上下文-动作特征和线性参数有界，并定义 $L := \max_{t \in \mathbb{N}} \sup_{a \in \mathcal{A}} \|\varphi(X_t, a)\|_2$ 和 $M := \max_{t \in \mathbb{N}} \|\gamma_{0,t}\|_2$。

线性老虎机算法会依次估计线性参数$\gamma_{0,\ t}$ 并利用它在给定观测到的上下文 X_t 时选择最优动作。随后，可以通过考察以下定义的动态遗憾 来评估算法的质量：

$$\mathrm{Reg}_T := \mathbb{E}\left[\sum_{t=1}^{T} (R_t^{a_t^*} - R_t^{a_t})\right] \tag{2}$$

其中 $a_t^* := \arg\max_{a \in \mathcal{A}} \varphi(X_t,\ a)^\top \gamma_{0,\ t}$是未知的最优动作。我们将时间$t$的遗憾 $\mathrm{reg}_t := \mathbb{E}($ $R_t^{a_t^*} - R_t^{a_t})$称为瞬时遗憾。

我们考虑一种非平稳设定，其中参数$\gamma_{0,\ t}$ 以及上下文分布 $\mathbb{P}_{X_t}$ 都可能随轮次变化。为了能够在非平稳设定下有效估计 $\gamma_{0,\ t}$ 我们作出两项假设：（1）假设 $\gamma_{0,\ t}$ 可以分解为变化部分和不变部分；（2）假设变化部分缓慢变化，因此在较短的时间区间内可视为近似恒定。为形式化假设（1），我们在下文提出假设 1，该假设基于 Lazzaretto 等人提出的不变子空间分解框架，(2025年)，并针对线性老虎机设定进行了调整。

假设1（不变子空间分解（ISD））。 存在一个非退化的 不变子空间分解，将 $\mathbb{R}^p$ 用于时变线性赌博机模型 （1），即将$\mathbb{R}^p$ 划分为两个线性子空间 （$\mathcal{S}^{\mathrm{inv}}$， $\mathcal{S}^{\mathrm{res}}$），使得 维数（$\mathcal{S}^{\mathrm{inv}}$）：$= p^{\mathrm{inv}} < p$ 且 维数（$\mathcal{S}^{\mathrm{res}}$）：$= p^{\mathrm{res}} = p - p^{\mathrm{inv}}$，满足

i) $\mathcal{S}^{\mathrm{inv}} = (\mathcal{S}^{\mathrm{res}})^{\perp}$;

ii) $\forall t \in \mathbb{N}$, $\forall x \in \mathcal{X}$, $\forall a \in \mathcal{A}$: $\mathrm{Cov}(\Pi^{\mathcal{S}^{\mathrm{inv}}} \varphi(x, a), \Pi^{\mathcal{S}^{\mathrm{res}}} \varphi(x, a)) = 0$;

iii) $\forall t \in \mathbb{N}$, $\exists \beta^{\text{inv}} \in \mathcal{S}^{\text{inv}}$, $\exists \delta_t^{\text{res}} \in \mathcal{S}^{\text{res}}$: $\gamma_{0,t} = \beta^{\text{inv}} + \delta_t^{\text{res}}$;

iv) $\forall$ 将满足 *(i)—(iii)* 的 $\mathbb{R}^p$ 划分为 $(\tilde{\mathcal{S}}^{\text{inv}}, \tilde{\mathcal{S}}^{\text{res}})$: $\dim(\tilde{\mathcal{S}}^{\text{inv}}) \leq \dim(\mathcal{S}^{\text{inv}})$。

空间 $\mathcal{S}^{\text{inv}}$ 和 $\mathcal{S}^{\text{res}}$ 分别称为 不变性 和残差子空间，而 β^{inv} 和 δ_t^{res}分别称为 不变性 和 残差分量 属于$\gamma_{0,\ t}$。Lazzaretto 等人 (2025年) 表明，不变分量和残差分量都可以表示为相应子空间中回归问题的最小二乘解。

利用假设 1，我们现在可以形式化多臂老虎机设置。我们将问题分解为一个离线阶段，随后是一个较短的在线阶段。

设定 1。 我们可以获取 $T_0 \in \mathbb{N}$ 离线观测 $(\varphi(X_t, a_t)\ a_t, R_t^{a_t})_{t\in[-T_0]}$，这些观测由多臂老虎机智能体与满足模型 (1)的变化环境交互收集。在线时间范围 $T \in \mathbb{N}$ 满足：对于所有 $t \in [T]$，模型γ_0中的参数 t 在模型 (1) 中保持恒定，且假设 *1*成立。

常数$\gamma_{0,\ t}$ 假设简化了分析，其动机源于其他非平稳多臂老虎机算法的设置；在这些设置中，T 可能表示重启算法中的一个单独阶段（Zhao 等人，2020年）或一个滑动窗口（Russac 等人，2019）。在实践中，我们提出的算法会依次将先前观测到的数据移至离线数据中。这种表述也便于与平稳线性多臂老虎机算法进行比较。

对于平稳多臂老虎机，时间范围 T 往往主导界；而对于非平稳多臂老虎机，由于变化分布使时间范围较短， p 项变得更为显著。利用离线数据，我们能够将上界从 $\tilde{O}$（$p\sqrt{T}$）降低至 $\tilde{O}$（$p^{\text{res}}\sqrt{T}$）。其中 p^{res} 是残差子空间的维数（在附录D中，我们证明在假设1 下，线性多臂老虎机算法的遗憾下界为$\Omega(p^{\text{res}}\sqrt{T})$）。

我们先回顾与本文相关的平稳线性多臂老虎机的一些概念（第2.1节），再介绍我们的方法（第3节），并分两步进行分析：首先假设分解（$\mathcal{S}^{\text{inv}}$， $\mathcal{S}^{\text{res}}$）已知（第4.1.2节），随后将分析扩展到通过数据估计该分解的情形（第4.2节）。

2.1 平稳线性老虎机的遗憾分析

在线性上下文多臂老虎机算法中，探索通常基于线性参数估计中的不确定性。更具体地说，令 $\hat{}$ γ_t 表示 $\gamma_{0,\ t}$ 的估计值，该估计值是通过使用观测值 $(X_\tau,\ R_\tau^{a_\tau})_{\tau=1}^{t-1}$，以 $\lambda > 0$ 为正则化参数进行岭回归所得的解。瞬时遗憾通常以高概率上界于一个与线性参数的估计误差 $\|\hat{}\gamma_t - \gamma_{0,\ t}\|_{\hat{\Sigma}_{t-1}}$ 和上下文-动作特征范数 $\|\varphi(X_t,\ a_t)\|_{\hat{\Sigma}_{t-1}^{-1}}$ 的乘积成正比的项，并分别通过 $\hat{}\ \Sigma_{t-1}$ 及其逆矩阵进行归一化，其中，对于所有 $t \in [T]$，我们定义正则化样本协方差矩阵

$$\hat{\Sigma}_t \coloneqq \lambda I_p + \sum_{\tau=1}^{t} \varphi(X_\tau, a_\tau)\varphi(X_\tau, a_\tau)^\top. \tag{3}$$

例如，Abbasi-Yadkori 等人（2011年）提出的基于上置信界的算法（LinUCB），以及 Agrawal 和 Goyal（2013年）提出的基于汤普森采样的算法（LinTS）都是如此。利用线性参数估计器的显式表达式和三角不等式，可以得到

$$\|\hat{\gamma}_t - \gamma_{0,t}\|_{\hat{\Sigma}_{t-1}} \leq \|\sum_{\tau=1}^{t-1} \varphi(X_\tau, a_\tau)^\top \epsilon_\tau\|_{\hat{\Sigma}_{t-1}^{-1}} + \sqrt{\lambda}\|\gamma_{0,t}\|_2. \tag{4}$$

在界定线性参数估计误差方面发挥关键作用的一个结果，是由Abbasi-Yadkori 等人（2011年）提出的以下引理。

引理 *1*（*Abbasi-Yadkori* 等（*2011*），定理 *1*）。设 $\{F_t\}_{t=0}^{\infty}$ 为一个滤子。设 $\{\epsilon_t\}_{t=1}^{\infty}$为一个实值随机过程，使得 ϵ_t 是关于 F_t可测的，且在给定 F_{t-1} 的条件下为子高斯，其参数为 $\sigma > 0$。设 $\{\varphi(X_t, a_t)\}_{t=1}^{\infty}$ 为一个取值于 $\mathbb{R}^p$的随机过程，使得$\varphi(X_t, a_t)$是关于 F_{t-1}可测的。设 $\lambda \in \mathbb{R}$ 为一个严格正的常数。那么，对于所有 $\eta \in (0, 1)$，以至少 $1-\eta$ 的概率，对所有 $t \in \mathbb{N}$，均有

$$\|\sum_{\tau=1}^{t} \varphi(X_\tau, a_\tau)\epsilon_\tau\|^2_{\hat{\Sigma}_t^{-1}} \leq 2\sigma^2 \log\left(\frac{1}{\eta}\sqrt{\frac{\det(\hat{\Sigma}_t)}{\det(\lambda I_p)}}\right).$$

此外，若对于所有 $t \in [T]$，$\|\varphi(X_t, a_t)\|_2 \leq L$，则

$$\log\left(\frac{\det(\hat{\Sigma}_t)}{\det(\lambda I_p)}\right) \leq p \log\left(1 + \frac{tL^2}{\lambda p}\right).$$

˜引理 1 意味着 $\|\hat{}\gamma_t - \gamma_{0,t}\|_{\hat{\Sigma}_{t-1}}$ 为 $O(\sqrt{p})$，其中对 p 的依赖由矩阵 Σ_t的维度决定。上下文-动作特征的平方范数在整个时间范围内的总和 $\sum_{t=1}^{T} \|\varphi(X_t, a_t)\|^2_{\hat{\Sigma}_{t-1}^{-1}}$ 也被证明为 $\tilde{O}(p)$，同样取决于上下文-动作特征的维度。在对累积遗憾˜进行界定时，该项出现在平方根之下，因此最终界中会多出一项 $O(\sqrt{p})$。我们的目标是证明，在设定 1 中，我们可以使遗憾依赖于 p^{res} 而非 p，同时允许存在某个在 T_0足够大时可忽略的项。下文将重点分析 LinUCB，但类似的推理也适用于其他基于最小二乘估计的线性老虎机算法。

3 ISD-linUCB算法

假设设定 1，我们提出一种算法，将其分为用于估计 $(\mathcal{S}^{\text{inv}}, \mathcal{S}^{\text{res}})$ 和 β^{inv}的离线阶段，以及随后适应非平稳性的在线赌博机阶段。

离线阶段依赖于设定 1 中的 T_0 历史观测。为估计 $(\mathcal{S}^{\text{inv}}, \mathcal{S}^{\text{res}})$，Lazzaretto 等人（2025年）估计一个正交归一矩阵 U，该矩阵能够将协方差矩阵 $(\text{Var}(\varphi(X_t, a_t)))^{t\in[-T_0}$联合分块对角化。$U$ 可以划分为两个子矩阵$U\text{inv} \in \mathbb{R}^{p\times p^{\text{inv}}}$和 $U\text{res} \in \mathbb{R}^{p\times p^{\text{res}}}$，其列分别构成 $\mathcal{S}^{\text{inv}}$和 $\mathcal{S}^{\text{res}}$的基。于是，投影到 $\mathcal{S}^{\text{inv}}$和 $\mathcal{S}^{\text{res}}$上的正交投影矩阵分别为 $\Pi^{\mathcal{S}^{\text{inv}}} = U^{\text{inv}}(U^{\text{inv}})^\top$和$\Pi^{\mathcal{S}^{\text{res}}} = U^{\text{res}}(U\text{res})^\top$。在假设1下，对于所有 $t \in [T]$，我们可以将（1）重写为

$$R_t^{a_t} = \varphi(X_t, a_t)^\top \Pi^{\mathcal{S}^{\text{inv}}} \beta^{\text{inv}} + \varphi(X_t, a_t)^\top \Pi^{\mathcal{S}^{\text{res}}} \delta_t^{\text{res}} + \epsilon_t, \tag{5}$$

其中 $\text{Cov}(\Pi^{\mathcal{S}^{\text{inv}}}\varphi(X_t, a_t)\ \Pi^{\mathcal{S}^{\text{res}}}\varphi(X_t, a_t)) = 0$。因此，子空间分解允许分别估计 β^{inv} 和 δ_t^{res}。对于任意索引集 $\mathcal{T} \subseteq [-T_0] \cup [T]$，我们将样本协方差矩阵定义为

$$\hat{\Sigma}_{\mathcal{T}} := \sum_{\tau\in\mathcal{T}} \varphi(X_\tau, a_\tau)\varphi(X_\tau, a_\tau)^\top.$$

ˆ 我们作出以下假设，以确保 $\Sigma_{[-T_0]}$为严格正定矩阵。

假设 2. 用于收集 T_0 观测的策略满足 $\exists\lambda_0 > 0$ 使得$\lambda_{\min}(\frac{1}{T_0}\hat{}\Sigma_{[-T_0]}) \geq \lambda_0$ 几乎处处。

—

这尤其意味着，用于收集离线数据的策略已充分探索上下文-动作特征空间中的所有方向。这也是能够估计不变子空间分解的首要条件（参见Lazzaretto 等人，2025年）。

我们将在假设 2 下记为我们可以利用估计得到的 β^{inv} 以及离线观测，在不变子空间中求普通最小二乘解，从而估计不变分量 $\hat{U}^{\mathrm{inv}}$，即

$$\hat{\beta}^{\mathrm{inv}} := \hat{U}^{\mathrm{inv}}(\tilde{\Sigma}^{\mathrm{inv}}_{[-T_0]})^{-1} \sum_{t\in[-T_0]} (\hat{U}^{\mathrm{inv}})^\top \varphi(X_t, a_t) R_t^{a_t}, \tag{6}$$

其中 $\tilde{\Sigma}^{\mathrm{inv}}_{[-T_0]} := (\hat{U}^{\mathrm{inv}})^\top \hat{\Sigma}_{[-T_0]} \hat{U}^{\mathrm{inv}}$。我们围绕 $\hat{\beta}^{\mathrm{inv}}$ 定义一个置信集：

$$\hat{\mathcal{C}}^{\beta} := \{\beta \in \hat{\mathcal{S}}^{\mathrm{inv}} \mid \|\hat{\beta}^{\mathrm{inv}} - \beta\|^2_{\hat{\Sigma}_{[-T_0]}} \le \hat{\rho}^{\mathrm{inv}}_{T_0}(\eta, L, M)\}, \tag{7}$$

其中 $\hat{\rho}^{\mathrm{inv}}_{T_0}(\eta, L, M)$ 在附录 25C.3 中的式（）中定义，使得该集合以至少 $\hat{\Pi}^{\mathcal{S}^{\mathrm{inv}}} \beta^{\mathrm{inv}}$ 的概率包含 $1-\eta$。随着新观测不断获得，估计 $\hat{\beta}^{\mathrm{inv}}$ 以及式（7）中的置信集可以在线重新计算。

在在线 ISD-linUCB 算法的第 $t \in [T]$ 轮，我们估计残差分量 a s

$$\hat{\delta}^{\mathrm{res}}_t := \hat{U}^{\mathrm{res}}(\tilde{\Sigma}^{\mathrm{res}}_{t-1})^{-1} \sum_{\tau=1}^{t-1} (\hat{U}^{\mathrm{res}})^\top \varphi_\tau (R_\tau^{a_\tau} - \varphi_\tau^\top \hat{\beta}^{\mathrm{inv}})$$

其中，$\varphi_\tau := \varphi(X_\tau, a_\tau)$ 且对所有 $t \in [T]$，$\tilde{\Sigma}^{\mathrm{res}}_t := (\hat{U}^{\mathrm{res}})^\top \hat{\Sigma}_t \hat{U}_{\text{残差}}$ 且 $\hat{\Sigma}_t$ 定义如式（3）所示。我们随后围绕 δ^{res}_t 定义一个置信集：

$$\hat{\mathcal{C}}^{\delta}_t := \{\delta \in \hat{\mathcal{S}}^{\mathrm{res}} \mid \|\hat{\delta}^{\mathrm{res}}_t - \delta\|^2_{\hat{\Sigma}_{t-1}} \le \hat{\rho}^{\mathrm{res}}_t(\eta, L, M)\},$$

其中，$\hat{\rho}^{\mathrm{res}}_t(\eta, L, M)$ 定义见附录 30）中的式（C.3，该集合以至少 $\hat{\Pi}^{\mathcal{S}^{\mathrm{res}}} \delta^{\mathrm{res}}_t$ 的概率包含 $1-\eta$，并基于 $\hat{\mathcal{C}}^{\beta} \oplus \hat{\mathcal{C}}^{\delta}_t$ 选择动作。假设 1 确保在估计 β^{inv} 时，因省略残差子空间中的上下文-动作特征而引入偏差（对于 δ^{res}_t，情况亦然），因此 $\hat{\mathcal{C}}^{\beta}$ 和 $\hat{\mathcal{C}}^{\delta}_t$ 具有所需的覆盖率。

单个在线步骤中完整的 ISD-linUCB 算法（包含和不包含不变分量更新）见算法 1。由于设定 1 假设 $\gamma_{0,\,t}$ 在 $[T]$中固定，因此 ISD-linUCB 在 $\mathcal{S}^{\mathrm{res}}$ 中执行标准 LinUCB 算法。当该假设不成立时，可修改算法1 ，使其在 $\mathcal{S}^{\mathrm{res}}$ 中像现有的非平稳算法（例如，使用滑动窗口）那样运行，从而提升其维度方面的性能。

4 遗憾分析

为说明我们的方法，我们首先在第 4.1 节分析一种简化算法的遗憾，该算法掌握子空间分解 ($\mathcal{S}^{\mathrm{inv}}$, $\mathcal{S}^{\mathrm{res}}$)的预言机知识。随后，我们将在第 4.3 节讨论完整的遗憾分析。

4.1 动机：oracleISD-linUCB

我们研究算法 1，并假设矩阵 $U = [U_{\mathrm{inv}} U^{\mathrm{res}}]$ 已知。为简单起见，我们沿用上一节引入的符号，以表示已知 U时的相同量。我们

算法 1 ISD-linUCB（时间 t 的迭代）

输入： $(\varphi(X_\tau,\ a_\tau)\ a_\tau,\ R_\tau^{a_\tau})_{\tau\in[t-1]\cup[-T_0]},\ X_t$

参数： $\mathcal{A},\ \lambda,\ \eta,\ L,\ M,$ 重新计算

$\mathbf{X}_t \leftarrow [\varphi(X_1, a_1)\ldots, \varphi(X_{t-1}, a_{t-1})]^\top$

$\mathbf{R}_t \leftarrow [R_1^{a_1}, \ldots, R_{t-1}^{a_{t-1}}]^\top$

如果 重新计算或 $t = 1$ **则**

$\mathcal{T} \leftarrow [-T_0] \cup [t-1]$

$\overline{\mathbf{X}} \leftarrow [\varphi(X_{-T_0}, a_{-T_0})\ldots, \varphi(X_{t-1}, a_{t-1})]^\top$

计算 $\hat{U}^{\text{res}}$ 和 $\hat{U}^{\text{inv}}$ 使用联合块对角化

$\mathbf{X}^{\text{inv}} \leftarrow \mathbf{X}\hat{U}^{\text{inv}}$

$\hat{\beta}^{\text{inv}} \leftarrow \hat{U}^{\text{inv}}\big((\mathbf{X}^{\text{inv}})^\top\mathbf{X}^{\text{inv}}\big)^{-1}(\mathbf{X}^{\text{inv}})^\top\overline{\mathbf{R}}$

$\hat{\mathcal{C}}^\beta \leftarrow \{\beta \in \hat{\mathcal{S}}^{\text{inv}} \mid \|\hat{\beta}^{\text{inv}} - \beta\|^2_{\hat{\Sigma}_{\mathcal{T}}} \le \hat{\rho}^{\text{inv}}_{|\mathcal{T}|}(\eta,\ L,\ M)\}$

else

load $\hat{\beta}^{\text{inv}}$, $\hat{U}^{\text{res}}$ 和 $\hat{\mathcal{C}}^\beta$ 来自上一次迭代

$\mathbf{R}_t^{\text{res}} \leftarrow \mathbf{R}_t - \mathbf{X}_t\hat{\beta}^{\text{inv}}$

$\mathbf{X}_t^{\text{res}} \leftarrow \mathbf{X}_t\hat{U}^{\text{res}}$

$\hat{\delta}_t^{\text{res}} \leftarrow \hat{U}^{\text{res}}\big(\lambda I_{p^{\text{res}}} + (\mathbf{X}_t^{\text{res}})^\top\mathbf{X}_t^{\text{res}}\big)^{-1}(\mathbf{X}_t^{\text{res}})^\top\mathbf{R}_t^{\text{res}}$

$\hat{\mathcal{C}}_t^\delta \leftarrow \{\delta \in \hat{\mathcal{S}}^{\text{res}} \mid \|\hat{\delta}_t^{\text{res}} - \delta\|^2_{\Sigma_{t-1}} \le \hat{\rho}_t^{\text{res}}(\eta,\ L,\ M)\}$

$a_t \leftarrow \arg\max_{a\in\mathcal{A}} \max_{\gamma\in\hat{\mathcal{C}}^\beta\oplus\hat{\mathcal{C}}_t^\delta} \varphi(X_t, a)^\top\gamma$

返回 a_t

定义 $(\bar{\beta}_t, \bar{\delta}_t) \coloneqq \arg\max_{\beta\in\hat{\mathcal{C}}^\beta,\delta\in\hat{\mathcal{C}}^\delta} \varphi(X_t, a_t)^\top(\beta+\delta)$ 为在时间 t 所选动作的上置信界达到最大值时对应的参数。

使用（5）中的奖励分解以及标准的 LinUCB 分析（例如参见 Lattimore 和 Szepesvári, 2020年，第 19.3 节），我们可以得到，在时间 $t \in [T]$，以高概率，瞬时遗憾的上界为

$$\text{reg}_t \le \underbrace{\varphi(X_t, a_t)^\top U^{\text{inv}}(U^{\text{inv}})^\top(\bar{\beta}_t - \beta^{\text{inv}})}_{\text{reg}_t^{\text{inv}}} + \underbrace{\varphi(X_t, a_t)^\top U^{\text{res}}(U^{\text{res}})^\top(\bar{\delta}_t - \delta_t^{\text{res}})}_{\text{reg}_t^{\text{res}}}. \tag{8}$$

直观地说，对于更大的 T_0，置信集$\hat{\mathcal{C}}^\beta$会收缩，并且 $\bar{\beta}_t - \beta^{\text{inv}}$ 趋近于零。因此，对于 $T_0 \gg T$，$\text{reg}_t^{\text{inv}}$相对于$\text{reg}_t^{\text{res}}$变得可以忽略。消除不变分量上的不确定性意味着，在遗憾分析中，对维度 p（即老虎机参数的维度）的依赖被降低为维度 p^{res}（即残差参数的维度）。我们首先在还拥有不变分量 β^{inv} 的预言机知识这一假设下研究遗憾（第4.1.1节）。然后，我们将不变分量的估计纳入遗憾分析（第4.1.2节）。

4.1.1 预言机$(\mathcal{S}^{\text{inv}}, \mathcal{S}^{\text{res}})$ β^{inv}的遗憾分析

如果我们知道 β^{inv}，那么$\hat{\mathcal{C}}^\beta = \{\beta^{\text{inv}}\}$ 和$\text{reg}_t^{\text{inv}} = 0$。因此，算法1 仅在残差子空间中进行探索。如定理1所示，其遗憾仅取决于$\hat{}$在 δ_t^{res}上的不确定性，该不确定性位于一个 p^{res}维空间中。

定理 1。 考虑设定 1，假设 $(\mathcal{S}^{\text{inv}}, \mathcal{S}^{\text{res}})$ 和 β^{inv} 已知，并考虑算法 1 的预言机版本，该版本使用 β^{inv}、U^{inv} 和 U_{res}，而非 $\hat{}\,\beta^{\text{inv}}$、$\hat{}U_{\text{inv}}$ 和 $\hat{}U^{\text{res}}$。那么，该预言机算法在时间范围 T 内、满足 $\eta = 1/T$ 时的遗憾为 $\tilde{O}(p^{\text{res}}\sqrt{T})$。

定理 1 的证明利用了这样一个事实：在已知预言机知识的情况下，$\text{reg}_t = \text{reg}_t^{\text{res}}$，其中 $(U_{\text{res}})^\top \varphi(X_t,\ a_t)$ 和 $(U^{\text{res}})^\top(\bar{\delta}_t - \delta_t^{\text{res}})$ 的维度为 p^{res}。随后，证明沿用了第 2.1 节介绍的标准 LinUCB 算法的遗憾分析中的类似步骤。详细证明见附录 C.1。

4.1.2 从离线数据中估计不变分量

同时考虑使用 T_0 个观测估计不变分量所带来的不确定性时，我们可以将算法 1 的遗憾上界限制如下。

定理 2。 考虑设定 1，假设假设条件 2 成立，且已知 $(\mathcal{S}^{\text{inv}}, \mathcal{S}^{\text{res}})$，并考虑算法 1 的预言机版本，该版本使用 U^{inv} 和 U_{res}，而非 $\hat{}U_{\text{inv}}$ 和 $\hat{}U^{\text{res}}$。那么，该预言机算法在时间范围 T 且满足 $\eta = 1/T$ 时的遗憾为

$$\tilde{O}\left(\sqrt{T}\left(p^{\text{res}} + \left(\sqrt{\tfrac{p^{\text{inv}}}{\lambda_0}} + \tfrac{1}{\lambda_0}\right)\sqrt{\tfrac{T}{T_0}}\right)\right).$$

定理 2 表明，只要 T_0 足够大于 T，相比仅依赖在线数据，利用离线数据估计遗憾函数中的不变信息更具优势。事实上，T_0 越大，我们就越接近第 4.1.1 节中所示的神谕界 $\tilde{O}(p^{\text{res}}\sqrt{T})$。

定理 2 的证明概要。 为了界定 $\text{reg}_t^{\text{inv}}$，我们选择 $\hat{}\rho_{T_0}^{\text{inv}}(\eta, L, M)$ 使得 $\hat{}\mathcal{C}^\beta$ 以至少 $\beta^{\text{inv}}\hat{}$ 的概率包含 $1-\eta$。根据假设 2，我们有 $\Sigma_{[-T_0]} \succeq \lambda_0 T_0 I_p$（其中 $\succeq$ 表示两个矩阵之间的洛温纳序），由此可得 $\tilde{\Sigma}^{\text{inv}}_{[-T_0]} \succeq \lambda_0 T_0 I_{p^{\text{inv}}}$。因此，对于所有 $\beta \in \hat{}\mathcal{C}^\beta$，

$$\begin{aligned}
\|\hat{\beta}^{\text{inv}} - \beta\|^2_{\hat{\Sigma}_{[-T_0]}} =& \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{\frac{1}{2}}(U^{\text{inv}})^\top(\hat{\beta}^{\text{inv}} - \beta)\|_2^2 \\
\geq& \|\sqrt{\lambda_0 T_0}(U^{\text{inv}})^\top(\hat{\beta}^{\text{inv}} - \beta)\|_2^2 \\
=& \lambda_0 T_0 \|(U^{\text{inv}})^\top(\hat{\beta}^{\text{inv}} - \beta)\|_2^2. \qquad (9)
\end{aligned}$$

最后，这意味着

$$\|(U^{\text{inv}})^\top(\hat{\beta}^{\text{inv}} - \beta)\|_2^2 \leq \frac{1}{\lambda_0 T_0}\hat{\rho}_{T_0}^{\text{inv}}(\eta, L, M).$$

因此，以至少 $1-\eta$ 的概率，

$$\text{reg}_t^{\text{inv}} \leq 2\|(U^{\text{inv}})^\top \varphi(X_t, a_t)\|_2 \sqrt{\frac{\hat{\rho}_{T_0}^{\text{inv}}(\eta, L, M)}{\lambda_0 T_0}}.$$

根据引理 1 和假设 2，我们可以将 $\sqrt{\hat{}\rho_{T_0}^{\text{inv}}(\eta, L, M)}$ 定义为 $\tilde{O}(\sqrt{p^{\text{inv}}} + \sqrt{\frac{1}{\lambda_0}})$。$\text{reg}_t^{\text{res}}$ 的分析与定理 1 类似，只是由于引入了 $\hat{}\,\beta^{\text{inv}}$，还需增加一项，即 $\tilde{O}(\sqrt{T\hat{}\rho_{T_0}^{\text{inv}}(\eta, L, M)/(\lambda_0 T_0)})$。这意味着，对 $t \in [T]$ 求和后，累积遗憾为 $\tilde{O}\left(\sqrt{T}\left(p^{\text{res}} + \sqrt{\frac{p^{\text{inv}} T}{\lambda_0 T_0}} + \frac{1}{\lambda_0}\sqrt{\frac{T}{T_0}}\right)\right)$。完整证明见附录 C.2。 □

4.2 子空间分解中的误差核算

在估计 $(\mathcal{S}^{\text{inv}},\ \mathcal{S}^{\text{res}})$ 时，T_0 利用现有观测，我们需要考虑两种误差来源。Lazzaretto 等人 (2025年提出通过联合块对角化来估计子空间。在我们的情形中，即对上下文-动作特征随时间变化的样本协方差矩阵进行联合块对角化。该过程首先对特征进行不可约联合分解，将其分解为正交低维子空间，使得特征在这些子空间上的投影两两不相关。随后，将奖励与特征之间的线性关系具有不变性的子空间归为一组，定义 $\mathcal{S}^{\text{inv}}$；并将其余线性关系随时间变化的子空间归为一组，定义 $\mathcal{S}^{\text{res}}$。

第一种误差来源是将 $\mathbb{R}^p$ 中的某些方向错误地识别为不变或时变方向。将不变方向标记为时变方向问题较小，因为这会导致算法虽非最优但并非错误：在这种情况下，p^{inv} 被低估，使分解无法发挥最大效益。相反，高估 p^{inv}（即假设时变方向具有不变性）意味着算法会在这些方向上使用 T_0 历史观测得到的合并估计参数来预测奖励，而真实参数可能已经发生变化。直观而言，ISD-linUCB 会因此遭受额外损失，其大小取决于在被错误标记的子空间中，$[T]$ 里的真实参数与其在 $[-T_0]$ 中的平均值相差多大。

第二种误差来源是有限样本近似导致的估计误差。为估计矩阵 U，我们首先估计 (Var(φ($X_t,\ a_t$)))$_{t\in[-T_0]}$；在实践中，只需将$[-T_0]$分成 m个窗口，并计算 $\hat{}\,\Sigma_{\mathcal{M}_i}$，其定义与式（3）相同，其中 $\mathcal{M}_i \coloneqq \{-\frac{i}{m}T_0,\dots,-\frac{i-1}{m}T_0-1\}$，$\hat{}\,\hat{}i\in\{1,\dots,m\}$。随后，$U$便是 $(\Sigma_{\mathcal{M}_i})_{i\in\overline{\{1,\dots,m\}}}$的一个估计联合块对角化矩阵。矩阵 $\hat{}U$ 随后用于估计 β^{inv} 和 δ_t^{res}。设 $\hat{}\gamma^{\text{ISD}} = \hat{}\beta^{\text{inv}} + \hat{}\delta_t^{\text{res}}$为使用 ISD 框架得到的$\gamma_{0,\ t}$ 的一个估计量。对于所有 $t\in[T]$，我们可以将估计误差表示为

$$\begin{aligned}\hat{\gamma}_t^{\text{ISD}} - \gamma_{0,t} &= \hat{\beta}^{\text{inv}} - \beta^{\text{inv}} + \hat{\delta}_t^{\text{res}} - \delta_t^{\text{res}} \\ &= (\hat{\beta}^{\text{inv}} - \hat{\Pi}^{\mathcal{S}^{\text{inv}}}\beta^{\text{inv}}) + (\hat{\Pi}^{\mathcal{S}^{\text{inv}}} - \Pi^{\mathcal{S}^{\text{inv}}})\beta^{\text{inv}} \\ &\quad + (\hat{\delta}_t^{\text{res}} - \hat{\Pi}^{\mathcal{S}^{\text{res}}}\delta_t^{\text{res}}) + (\hat{\Pi}^{\mathcal{S}^{\text{res}}} - \Pi^{\mathcal{S}^{\text{res}}})\delta_t^{\text{res}}. \end{aligned} \tag{10}$$

第一项和第三项分别对应于估计子空间上不变部分和残差分量的估计误差。第二项和第四项仅取决于子空间估计误差。在界定算法 1 的遗憾时，我们假设子空间分解已被正确估计，并且仅考虑估计 $(\mathcal{S}^{\text{inv}},\ \mathcal{S}^{\text{res}})$ 时的有限样本误差，从分解式（10）开始。

4.3 完整遗憾分析

为分析式 (10) 中的各项，我们需要量化子空间估计误差。常见的方法是考察真实子空间与估计子空间之间的距离，这一距离可通过主角的概念来描述。令 $U^{\mathcal{S}}\in\mathbb{R}^{p\times k}$ 为 U的子矩阵，其列张成 $\mathbb{R}^p$的子空间 $\mathcal{S}$；令 $\hat{}U^{\mathcal{S}}$ 为通过 $\hat{}U$得到的估计，其 $\hat{}\,\hat{}$ 列张成 $\mathcal{S}$。$\mathcal{S}$ 与 $\mathcal{S}$ 之间的主角，定义为矩阵 $(\hat{}U^{\mathcal{S}})^\top U^{\mathcal{S}}$的非零奇异值的反余弦（直观而言，两个向量夹角的余弦等于它们的归一化内积）。令 $\Theta^{\mathcal{S}}$ 为对角矩阵，其对角线上为 $\hat{}\,\mathcal{S}$ 与 $\mathcal{S}$ 之间的主角。于是，我们用 $\|\sin\Theta^{\mathcal{S}}\|_{\text{op}}$ 度量 $\mathcal{S}$与 $\hat{}\mathcal{S}$之间的距离，也就是 $\mathcal{S}$与 $\hat{}\mathcal{S}$之间最大的主角（例如参见 Stewart and Sun 的定理 4.5，1990），该距离也等于 $\|\hat{}\Pi^{\mathcal{S}} - \Pi^{\mathcal{S}}\|_{\text{op}}$（Stewart and Sun，1990，推论 4.6）。我们用 $\Delta\Pi$ 表示不变子空间和残差子空间之间的这一距离（两者必须

对两个子空间都相同，参见Stewart and Sun (1990，预备知识——推论 5.4））， 并针对子空间分解误差提出以下假设。

假设 3。 对于所有 $\eta \in (0,1)$，以至少 $1-\eta$ 的概率，有 $\Delta\Pi$ 为 $O(\sqrt{\log(p/\eta)/T_0})$。

假设 3 可通过 Davis–Kahan 定理（例如，参见 Yu 等人，2015年）以及关于样本协方差矩阵的集中性结果（例如，参见 Vershynin，2018年，第 5.4 和 5.6 节）来论证。具体而言，Davis–Kahan 定理给出了 $\|\sin\Theta^{\mathcal{S}}\|_{\mathrm{op}}$ 的上界，该上界取决于矩阵 $M \in \mathbb{R}^{p\times p}$ 的估计误差，其中 U 的列是 M 的特征向量。在我们的情形中，U 联合分块对角化 $(\mathrm{Var}(\varphi(X_t,a_t)))_{t\in[-T_0]}$，且对于所有 $t\in[-T_0]$，U^{inv} 和 U^{res} 的列张成的空间与 $\mathrm{Var}(\varphi(X_t,a_t))$ 的若干特征子空间的并集相同。我们通过矩阵 $(\hat{\Sigma}_{\mathcal{M}_i})_{i=1,\dots,m}$ 估计 $\hat{U}$，每个矩阵均使用 T_0/m 个观测计算。由于 $\mathrm{Var}(\varphi(X_t,a_t))$ 不随 t 保持不变，我们无法使用精确的集中性结果；在假设 3 下，我们假设 $\Delta\Pi$ 的尺度与从 T_0 个观测中估计一个固定的 p 维协方差矩阵时的有限样本误差相当（我们在附录 B.1 中通过实证验证了这一点）。

式（10）中第二项与第四项之和等于 $\Delta\Pi\gamma_{0,\,t}$。在假设 3 下，该量以至少 $1-\eta$ 的概率为 $O(\sqrt{\log(p/\eta)/T_0})$。此外，在假设 3 下，我们可以给出 $\hat{\beta}^{\mathrm{inv}}$ 在估计的不变子空间中的估计误差的如下高概率界。

引理 2. 考虑设定 *1* 并假定假设 *2* 与 *3* 成立。则对所有 $\eta\in(0,\,1)$ 以至少 $1-\eta$ 的概率成立：$\|\hat{\beta}^{\mathrm{inv}}-\hat{\Pi}^{\mathcal{S}^{\mathrm{inv}}}\beta^{\mathrm{inv}}\|_{\hat{\Sigma}_{[-T_0]}}$ 为

$$O\left(\sqrt{p^{\mathrm{inv}}\log(\tfrac{1}{p^{\mathrm{inv}}\lambda_0})+\log(\tfrac{1}{\eta})}\right)+O\left(\sqrt{p^{\mathrm{inv}}\log(\tfrac{p}{\eta})}\right)+O\left(\sqrt{\tfrac{1}{\lambda_0}\log(\tfrac{p}{\eta})}\right).$$

证明概要。 我们可以将误差分解为如下三项

$$\begin{aligned}\|\hat{\beta}^{\mathrm{inv}}-\hat{\Pi}^{\mathcal{S}^{\mathrm{inv}}}\beta^{\mathrm{inv}}\|_{\hat{\Sigma}_{[-T_0]}} \le{}& \|(\hat{U}^{\mathrm{inv}})^\top\hat{\Sigma}_{[-T_0]}(\Pi^{\mathcal{S}^{\mathrm{inv}}}-\hat{\Pi}^{\mathcal{S}^{\mathrm{inv}}})\beta^{\mathrm{inv}}\|_{(\tilde{\Sigma}^{\mathrm{inv}}_{[-T_0]})^{-1}}\\ &+\|(\hat{U}^{\mathrm{inv}})^\top\sum_{t=-T_0}^{-1}\varphi(X_t,a_t)\varphi(X_t,a_t)^\top\delta_t^{\mathrm{res}}\|_{(\tilde{\Sigma}^{\mathrm{inv}}_{[-T_0]})^{-1}}\\ &+\|(\hat{U}^{\mathrm{inv}})^\top\sum_{t=-T_0}^{-1}\varphi(X_t,a_t)\epsilon_t\|_{(\tilde{\Sigma}^{\mathrm{inv}}_{[-T_0]})^{-1}}.\end{aligned}$$

最后一项可按预言机情形（定理 2）进行分析。前两项源于真实子空间与估计子空间之间的错位。在假设 3 下，对所有 $\eta\in(0,\,1)$，以至少 $1-\eta$ 的概率，第一项为 $O(\sqrt{p^{\mathrm{inv}}\log(\frac{p}{\eta})})$，第二项为 $O(\sqrt{\frac{1}{\lambda_0}\log(\frac{p}{\eta})})$。完整证明见附录 C.3。 □

将引理2中的所有因子相乘 $\frac{1}{\sqrt{\lambda_0 T_0}}$（参见（9）），即可进一步得到 $\|\hat{\beta}^{\mathrm{inv}}-\hat{\Pi}^{\mathcal{S}^{\mathrm{inv}}}\beta^{\mathrm{inv}}\|_2$ 的上界。

对于残差分量，我们同样可以根据假设 3，将 $\|\hat{\delta}_t^{\mathrm{res}}-\hat{\Pi}^{\mathcal{S}^{\mathrm{res}}}\delta_t^{\mathrm{res}}\|_{\hat{\Sigma}_{t-1}}$ 界定如下。

引理 3。 考虑设定 *1* 并假设假设条件 *2* 和 *3* 成立。那么，对于所有 $t \in [T]$ 以及所有 $\eta \in (0,1)$ 以至少 $1-\eta$ 的概率，有 $\|\hat{\delta}_t^{\text{res}} - \hat{\Pi}^{\mathcal{S}^{\text{res}}} \delta_t^{\text{res}}\|_{\hat{\Sigma}_{t-1}}$ 成立

$$O\left(\sqrt{p^{\text{res}} \log(\tfrac{t}{p^{\text{res}}}) + \log(\tfrac{1}{\eta})}\right) + O\left(\sqrt{p^{\text{res}} t}\left(\sqrt{\tfrac{\log(p/\eta)}{T_0}} + \|\hat{\beta}^{\text{inv}} - \beta^{\text{inv}}\|_2\right)\right).$$

第一项与预言机子空间情况下出现的项相同。第二项则源于子空间估计误差。

证明概要。 我们可以将 $\|\hat{\delta}_t^{\text{res}} - \hat{\Pi}^{\mathcal{S}^{\text{res}}} \delta_t^{\text{res}}\|_{\hat{\Sigma}_{t-1}}$ 上界为以下各项之和

$$\begin{aligned}
\|\hat{\delta}_t^{\text{res}} - \hat{\Pi}^{\mathcal{S}^{\text{res}}} \delta_t^{\text{res}}\|_{\hat{\Sigma}_{t-1}} \le & \|\sum_{\tau=1}^{t-1} (\hat{U}^{\text{res}})^\top \varphi_\tau \varphi_\tau^\top (\Pi^{\mathcal{S}^{\text{res}}} - \hat{\Pi}^{\mathcal{S}^{\text{res}}}) \delta_t^{\text{res}}\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} \\
& + \|\sum_{\tau=1}^{t-1} (\hat{U}^{\text{res}})^\top \varphi_\tau \varphi_\tau^\top (\beta^{\text{inv}} - \hat{\beta}^{\text{inv}})\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} \\
& + \|\sum_{\tau=1}^{t-1} (\hat{U}^{\text{res}})^\top \varphi_\tau \epsilon_\tau\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} + \sqrt{\lambda}\|\delta_t^{\text{res}}\|_2
\end{aligned}$$

其中 $\varphi_\tau := \varphi(X_\tau, a_\tau)$。最后两项的分析与定理 1 中相同，因此，对于所有 $\eta \in (0,1)$，以至少 $1-\eta$ 的概率，这些项为 $O\left(\sqrt{\log(\tfrac{1}{\eta}) + p^{\text{res}} \log(\tfrac{t}{p^{\text{res}}})}\right)$。前两项是由子空间估计误差引入的。在假设 3 下，对于所有 $\eta \in (0,1)$，以至少 $1-\eta$ 的概率，第一项为 $O\left(\sqrt{\tfrac{p^{\text{res}} t}{T_0}} \log(\tfrac{p}{\eta})\right)$，第二项为 $O\left(\sqrt{p^{\text{res}} t} \|\hat{\beta}^{\text{inv}} - \beta^{\text{inv}}\|_2\right)$。完整证明见附录 C.3。 □

引理2 与引理3 蕴含 ISD-linUCB 算法的如下遗憾界。

定理 3. 考虑设定 *1* 并假设假设 *2* 与 *3* 成立。若 $\hat{\mathcal{C}}^\beta$ 被定义为包含 $\hat{\Pi}^{\mathcal{S}^{\text{inv}}} \beta^{\text{inv}}$ 的概率至少为 $1 - 4/T_0$，且 $\hat{\mathcal{C}}_t^\delta$ 被定义为包含 $\hat{\Pi}^{\mathcal{S}^{\text{res}}} \delta_t^{\text{res}}$ 的概率至少为 $1 - (\tfrac{2}{T} + \tfrac{4}{T_0})$，则以至少 $1 - (\tfrac{2}{T} + \tfrac{9}{T_0})$ 的概率，*ISD-linUCB* 的遗憾为

$$\tilde{O}\left(\sqrt{T}\left(p^{\text{res}} + p^{\text{res}} \sqrt{\tfrac{T}{\lambda_0 T_0}}\left(\sqrt{p^{\text{inv}}} + \sqrt{\max\{\lambda_0, \tfrac{1}{\lambda_0}\}}\right)\right)\right).$$

定理 3 特别表明，当 T_0 充分大于 T（例如， $T_0 = \Omega(T^{1+\varepsilon})$，对某个 $\epsilon > 0$）时，ISD-linUCB 的遗憾界由 $p^{\text{res}}\sqrt{T}$主导。

5 仿真实验

为支持我们的理论结果，我们开展了若干仿真实验。首先，我们在第 5.1 节应用具有预言机子空间知识的 ISD-linUCB 算法，并展示遗憾确实按 $p^{\text{res}}\sqrt{T}$ 的尺度增长，从而支持以下结果：若 T_0 的增长速度快于 T，遗憾确实按 $p^{\text{res}}\sqrt{T}$ 的尺度增长。随后，我们将子空间分解的估计纳入算法（第 5.2 节），并展示随着 T_0 增大，其性能会逐渐接近具有预言机子空间知识的 ISD-linUCB。在所有仿真中，我们都与标准 LinUCB 算法进行比较，因为在设定 1 中，我们假设 $\gamma_{0,\,t}$ 在 [T] 中保持固定（在此设定下，其他非平稳算法的性能与 LinUCB 相当，见附录 B.2）。

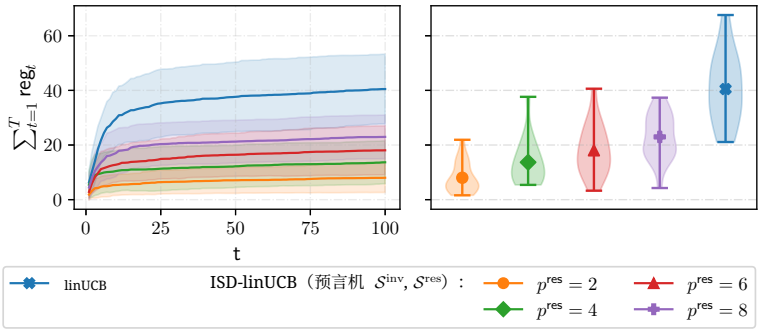

图2：在 $T = 100$ 轮次内，对于 $p^{\text{res}} \in \{2, 4, 6, 8\}$，ISD-linUCB 在拥有预言机（$\mathcal{S}^{\text{inv}}$， $\mathcal{S}^{\text{res}}$）时的遗憾。对于每个 p^{res}，实验重复进行 20 次。左图展示了 20 次重复实验的平均表现和标准差，右图展示了 20 次重复实验中遗憾的分布。

5.1 预言机子空间分解

我们考虑一个设定，其中 $T_0 = 2000$、$T = 100$ 和 $|\mathcal{A}| = 5$ 已知，且已知 $U_{\text{inv}} U^{\text{res}}$。矩阵$U$ 是从随机正交归一矩阵中采样得到的，我们将动作-上下文特征的协方差矩阵生成为 $U\tilde{V}_t U^\top$，其中 $\tilde{V}_t$ 是一个包含两个分块的分块对角矩阵，其维度为p^{inv}， p^{res}。$(U^{\text{inv}})^\top \beta^{\text{inv}}$ 的元素从 $(0.5, 1.5)$ 中均匀采样。对于 $(U^{\text{res}})^\top \delta_t^{\text{res}}$ 的初始值，我们也采用相同的采样方式；对于 $t \in [-T_0]$，我们加上 $-1.5(t/T0)\sin^2(0.25it/T0 + i)$，其中 $i \in \{1, \ldots, p^{\text{res}}\}$ 为元素索引。 $[-T_0]$ 中用于估计 β^{inv} 的数据，是由一个采用策略、以均匀随机方式选择动作的老虎机算法生成的。对于 $t \in [T]$，$(U^{\text{res}})^\top \delta_t^{\text{res}}$ 的元素从 $(0.5, 1.5)$ 中均匀采样。我们设定 $\lambda = 0.1$，并使用 $\eta = 1/T$。随后我们进行两项实验（两项实验中，$\hat{}\ \beta^{\text{inv}}$ 均不在线更新）。

首先，对于上下文-动作特征的固定维度 $p = 10$，我们考虑以下取值：$p^{\text{res}} \in \{2, 4, 6, 8\}$。图 2 显示，ISD-linUCB（拥有预言机子空间信息）的遗憾随 T（左图）呈次线性增长，而随 p^{res}（右图）呈（近似）线性增长，从实验上支持了定理 2的结论。为了进一步支持这些结果，我们进行了第二个实验，考察维度在 p 至 3 和10之间变化的上下文-动作特征，同时保持 $\mathcal{S}^{\text{res}}$ 的维度固定为 $p^{\text{res}} = 2$。结果如图 3所示，该图将累积遗憾与标准 LinUCB 算法的累积遗憾进行了比较。后者随 p线性增长，而 ISD-linUCB 的遗憾则基本保持不变，因为 p^{res} 固定不变。

5.2 估计子空间分解

我们考虑与第 5.1 节相同的数据生成过程，但现在纳入使用 T_0 个观测进行的子空间估计。我们设定 $T = 500$ 并考虑 $T_0 \in \{1000, 3500, 8000\}$。此外，我们将子空间的维度固定为 $p^{\text{inv}} = 7$ 和 $p^{\text{res}} = 3$。图 4 展示了随着 T_0 增大时标准 LinUCB 算法（不受 T_0 影响）以及第4.1.1、4.1.2 和 4.2 节所研究算法的累积遗憾，从而支持所给出的理论结果。事实上，随着 T_0 增大，ISD-linUCB 算法在估计 $(\mathcal{S}^{\text{inv}}$， $\mathcal{S}^{\text{res}})$ 和 β^{inv} 时使用 T_0 个观测，其遗憾更接近具有预言机知识的算法的遗憾。此外，对于所考虑的全部 T_0 取值，使用估计所得 ISD 的算法的遗憾均低于标准 LinUCB 算法。

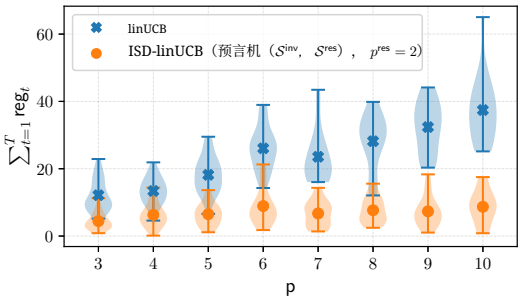

图3：标准 LinUCB 和带预言机（$\mathcal{S}^{\text{inv}}, \mathcal{S}^{\text{res}}$）的 ISD-linUCB 在$T = 100$ 且上下文-动作特征维度 p不断增大时的累积遗憾。对于 ISD-linUCB，使用 $T_0 = 2000$ 次观测估计不变分量 β^{inv} 。 p^{inv} 从 3 变化到 10，而 p^{res} 固定为 2。对于每个 p，实验重复 20 次。

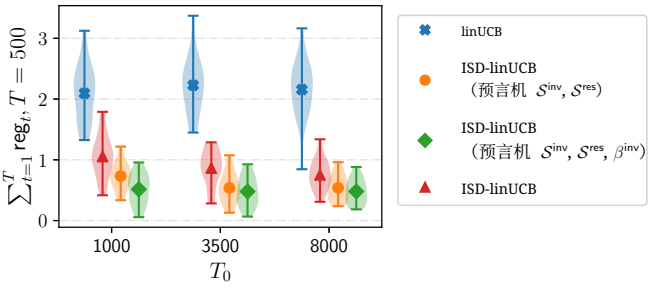

图4：ISD-linUCB 在 $T_0 \in \{1000, 3500, 8000\}$ 下的累积遗憾（重复 20 次），并与具有预言机信息的同一算法以及标准 LinUCB 算法（不受 T_0影响）进行比较。随着 T_0增大，ISD-linUCB 的遗憾逐渐接近预言机版本。

6 结论

在非平稳环境中利用不变性，可使线性上下文老虎机算法更高效地适应奖励模型的变化。我们提出 ISD-linUCB，这是一种新颖算法，它通过将上下文-动作特征空间分解为两个正交子空间，从离线老虎机数据中学习此类不变性：在这两个子空间中维度为p^{inv}的那个子空间内，奖励与上下文-动作特征之间的关系保持平稳。当有足够的离线数据时，ISD-linUCB 的遗憾以 $(p - p^{\text{inv}})$ 而非 p 的量级增长，从而在经历快速变化的环境中带来显著的性能提升。

致谢

NP 和 ML 获得了 Novo Nordisk Fonden 提供的研究资助（0069071）。

参考文献

Y. Abbasi-Yadkori、D. Pál 和 C. Szepesvári。线性随机老虎机的改进算法。神经信息处理系统进展，24，2011年。S. Agrawal 和 N. Goyal。针对具有线性收益的上下文老虎机的Thompson采样。载于国际机器学习会议，第 127–135 页。PMLR，2013年。P. Auer。利用置信界权衡利用与探索。机器学习研究杂志，3(Nov):397–422，2002年。S. Bilaj、S. Dhouib 和 S. Maghsudi。共享仿射子空间中的老虎机元学习。载于第*27*届国际人工智能与统计学会议，第 523–531 页。PMLR，2024年。W. C. Cheung、D. Simchi-Levi 和 R. Zhu。学习在非平稳环境下进行优化。载于第*22*届国际人工智能与统计学会议，第 1079–1087 页。PMLR，2019年。W. Chu、L. Li、L. Reyzin 和 R. Schapire。具有线性收益函数的上下文老虎机。载于第*14*届国际人工智能与统计学会议，第 208–214 页，2011年。V. Dani、T. P. Hayes 和 S. M. Kakade。基于老虎机反馈的随机线性优化。载于第*21*届学习理论年度会议，第 355–366 页，2008年。C. Kausik、K. Tan 和 A. Tewari。在线性潜变量老虎机中利用离线数据。*arXiv*预印本 *arXiv:2405.17324*，2024年。T. Lattimore 和 C. Szepesvári。老虎机算法。剑桥大学出版社，2020年。M. Lazzaretto、J. Peters 和 N. Pfister。不变子空间分解。机器学习研究杂志，26(95):1–56，2025年。L. Li、W. Chu、J. Langford 和 R. E. Schapire。基于上下文老虎机的个性化新闻文章推荐方法。载于第*19*届万维网国际会议论文集，第 661–670 页，2010年。M. Papini、A. Tirinzoni、M. Restelli、A. Lazaric 和 M. Pirotta。在线性上下文老虎机中利用优质表示。载于第*38*届国际机器学习会议，第 8371–8380 页。PMLR，2021年。Y. Qin、T. Menara、S. Oymak、S. Ching 和 F. Pasqualetti。序列线性老虎机中的非平稳表示学习。*IEEE*开放控制系统期刊，1:41–56，2022年。Y. Russac、C. Vernade 和 O. Cappé。适用于非平稳环境的加权线性老虎机。神经信息处理系统进展，32，2019年。S. Saengkyongam、N. Thams、J. Peters 和 N. Pfister。不变策略学习：因果视角。*IEEE*模式分析与机器智能汇刊，45(7):8606–8620，2023年。G. W. Stewart 和 J.-g. Sun。矩阵扰动理论。计算机科学与科学计算。学术出版社，1990年。网址：https://cir.nii.ac.jp/crid/1971149384759715084。

A. L. Trella、W. Dempsey、F. Doshi-Velez 和 S. A. Murphy。非平稳潜在自回归老虎机。*arXiv* 预印本 *arXiv:2402.03110*，2024年。J. A. Tropp。随机矩阵之和的易用尾界。计算数学基础，12(4):389–434，2012年。R. Vershynin。高维概率：数据科学中的应用导论，第47卷。剑桥大学出版社，2018年。Y. Yu、T. Wang 和 R. J. Samworth。一个适用于统计学家的 Davis–Kahan 定理实用变体。*Biometrika*，102(2):315–323，2015年。P. Zhao、L. Zhang、Y. Jiang 和 Z.-H. Zhou。一种解决非平稳线性老虎机问题的简单方法。载于第*23*届国际人工智能与统计学会议，第746–755页。PMLR，2020年。

进一步的相关工作

线性上下文老虎机 随机线性老虎机问题在文献中得到了广泛研究。通常假设（例如，参见 Lattimore 和 Szepesvári，2020年），存在一个维度为 p 、由所有上下文-动作特征共享的未知线性参数：这尤其使得处理任意大的动作空间成为可能，并可获得不依赖于动作空间基数 K 的遗憾界。文献中为解决这一问题提出的两种标准算法是 LinUCB （线性上置信界）和 LinTS （线性汤普森采样）。在基于置信上界的算法中，每一轮智能体都会选择使估计遗憾的上置信界最大的动作。Li 等（2010）考虑有限动作空间，并直接基于每个可能动作的估计遗憾构造此类置信上界。Chu 等（2011）证明了这一版本 LinUCB 的遗憾界为 $\tilde{O}(\sqrt{pKT})$。Abbasi-Yadkori 等（2011）提出了一种更通用的置信上界算法，它通过为估计的线性参数而非奖励构造置信集，从而支持无限大的动作空间，并证明其遗憾界为 $\tilde{O}(p\sqrt{T})$。基于 Thompson 采样的线性老虎机（Agrawal 和 Goyal，2013）在每一轮都会从所有可能臂的奖励后验分布中采样，然后选择使采样奖励最大的动作。研究表明，LinTS 可实现 $\tilde{O}(p^{\frac{3}{2}}\sqrt{T})$ 的遗憾。

在这些算法中，除了上下文-动作特征和老虎机参数有界外，并未对观测到的上下文的性质作出任何假设。 Papini 等人 (2021) 回顾了上下文-动作特征空间中不同的多样性条件，这些条件有助于改进遗憾界，使遗憾在 T 中呈对数级或常数级。其中一个例子是，最优上下文-动作特征的方差为正定矩阵，这意味着对于某些上下文， $\mathbb{R}^p$ 中的所有方向都有可能是最优的。

线性潜变量上下文老虎机 为处理奖励函数的异质性，Kausik 等人 (2024) 考虑了一个潜在线性上下文老虎机框架，其中潜在线性参数$\theta \in \mathbb{R}^{p_l}$（满足 $p_l < p$）在所有异质观测之间共享。他们假设奖励的形式为 $R_t^{a_t} = \varphi(X_t,\ a_t)^\top \gamma_0 + \epsilon_t$，其中 $\gamma_0 = U\theta$，且 $U \in \mathbb{R}^{p\times p_l}$是一个酉矩阵。他们提出利用离线数据学习矩阵 U ，并据此获得遗憾界，该遗憾界取决于潜在参数的维度 p_l ，而非上下文维度 p。

Bilaj 等人 (2024年) 假设智能体依次与不同任务交互，并且对于每个任务，参数 γ_0 都是从同一分布中独立抽取的。他们假设存在一个秩为 p_l的正交投影矩阵 P，满足 $\mathbb{E}_{\gamma_0}[\|(I-P)(\gamma_0-\mathbb{E}[\gamma_0])\|^2] \ll \mathbb{E}_{\gamma_0}[\|P(\gamma_0-\mathbb{E}[\gamma_0])\|^2] \le \mathbb{E}_{\gamma_0}[\|\gamma_0-\mathbb{E}[\gamma_0]\|^2]$，即 γ_0 的分布沿着 $p-p_l$ 正交方向的方差非常小。他们给出的遗憾界取决于 p_l 而非p。

Qin 等人 (2022年) 考虑了一种多任务序贯线性老虎机模型，其中智能体依次执行一系列 S 老虎机任务（从 m 个不同环境中抽取），每个任务持续 N 轮。他们将奖励建模为 $R_t^{a_t} = \varphi(X_t, a_t)^\top \gamma_{s(t)} + \epsilon_t$ ，并假设存在一个共享的潜在参数 $\theta \in \mathbb{R}^{p_l}$ ，使得对于所有 $s \in [S]$ ，都存在一个矩阵 $U_s \in \mathbb{R}^{p\times p_l}$ ，满足 $\gamma_s = U_s\theta$。他们提出了一种算法，可通过序贯方式学习潜在表示，并实现 $\tilde{O}(pp_l\sqrt{mSN} + p_lS\sqrt{N})$遗憾。

Trella 等人(2024年) 考虑了一种由潜在变量驱动的非平稳奖励模型，该变量˜按照阶数为 ℓ 的自回归过程演化。他们证明了一个遗憾界： $O(\ell\sqrt{pT})$。

非平稳线性老虎机 Cheung 等人 (2019年) 和Russac 等人 (2019年研究了环境非平稳情况下的线性老虎机问题，尤其允许线性老虎机参数随时间变化，因此奖励模型的形式为

form $R_t^a = \varphi(X_t, a)^\top \gamma_{0,t} + \epsilon_t$。他们将变化预算 B_T 定义为一个常数，使得$\sum_{t=1}^{T-1} \|\gamma_{0,t+1} - \gamma_{0,t}\|_2 \leq B_T$。Cheung 等人（2019）提出了一种基于滑动窗口正则化最小二乘估计器的算法，而 Russac 等人（2019）则提出对过去的观测采用折扣因子进行加权。两种方法都实现了 $\tilde{O}(p^{\frac{7}{8}} T^{\frac{3}{4}} B_T^{\frac{1}{4}})$ 的遗憾界。正如 Zhao 等人（2020年）所示，通过定期重启标准 LinUCB 算法也能达到相同的遗憾；每次重启前应执行的最优轮次数取决于 B_T。这意味着，与平稳情形相比，额外的遗憾源于采用假设线性参数固定的标准 LinUCB 算法，以及在两次重启之间的时间段内因参数变化而产生的额外损失。Cheung 等人（2019）还证明了这一设定下的下界 $\Omega(p^{\frac{2}{3}} T^{\frac{2}{3}} B_T^{\frac{1}{3}})$。

B 其他实验

B.1 投影误差

假设 3 的关键作用是确保 $\mathcal{S}^{\text{inv}}$ 和 $\mathcal{S}^{\text{res}}$ 上的投影误差会随着用于估计矩阵的样本量 T_0 的增加而减小 U（因此，投影到子空间上的投影矩阵也会随之变化）。在第 5.2 节所述的同一实验中，我们评估了此类误差，即 $\|\hat{\Pi}^{\mathcal{S}^{\text{inv}}} - \Pi^{\mathcal{S}^{\text{inv}}}\|_{\text{op}}$，并展示（图 5）该误差确实近似以 $\sqrt{T_0}$的速度减小，正如假设 3中所设。

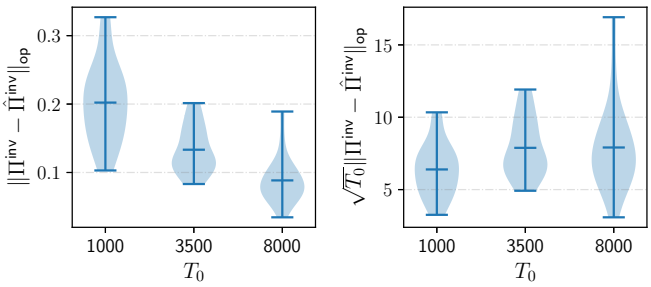

图 5： $T_0 \in \{1000, 3500, 8000\}$ 的投影误差。左图显示乘以 $\sqrt{T_0}$ 后的相同数值，证实了我们的假设。

B.2 其他非平稳算法

在第5.1节所述第二个实验的相同设置下，我们额外加入了两种非平稳算法进行比较：一种是 Cheung 等人Cheung et al. (2019)提出的使用滑动窗口的算法（SW-linUCB），另一种是 Russac 等人 Russac et al. (2019)提出的使用折扣因子的算法（D-linUCB）。在此设置中，线性参数 $\gamma_{0,t}$ 在时间范围 $[T]$内保持固定，因此将 SW-linUCB 的窗口维度设为 T，并将 D-linUCB 的折扣因子设为 0.999。图6 显示，这两种算法的性能确实与 LinUCB 相当。具体而言，我们保持 $\gamma_{0,t}$ 不变，因为我们关注的是降低维度（针对问题中的非平稳部分）。当这一假设不成立时，我们仍可将 ISD 框架整合到现有的非平稳算法中，以提升其维度方面的性能，正如我们在所展示的 LinUCB 实验中所做的那样。

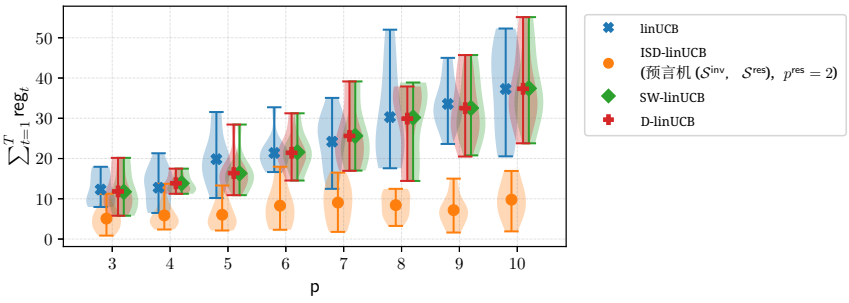

图 6：标准 LinUCB、SW-linUCB、D-linUCB 以及带有预言机（$\mathcal{S}^{\text{inv}}, \mathcal{S}^{\text{res}}$）的 ISD-linUCB 在 $T = 100$ 下、且上下文-动作特征维度 p取值递增时的累积遗憾。对于 ISD-linUCB，不变分量 β^{inv} 使用 $T_0 = 2000$ 个观测进行估计。 p^{inv} 从 3 变化到 10，而 p^{res} 固定为 2。

第 5.2 节所示的仿真实验也有相同情况，如图 7 所示。

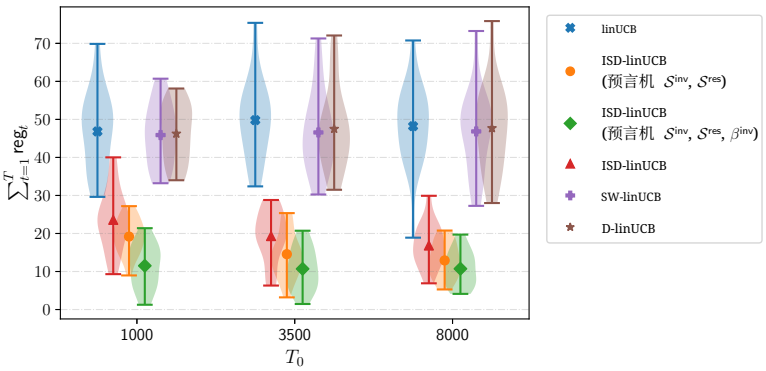

图7：ISD-linUCB 在 $T_0 \in \{1000, 3500, 8000\}$ 下的累积遗憾，并与具有预言机信息的同一算法、标准 linUCB 算法、SW-linUCB 以及 D-linUCB（不受 T_0影响）进行比较。对于每个 T_0，实验重复 20 次。

C 遗憾分析：证明

C.1 预言机 $\mathcal{S}^{\text{inv}}, \mathcal{S}^{\text{res}}, \beta^{\text{inv}}$

定理*1*的证明。我们首先证明，估计得到的置信集$\hat{\mathcal{C}}_t^{\delta}$以$\hat{\delta}_t^{\text{res}}$为中心，$\delta_t^{\text{res}}$以高概率成立。

Let $\tilde{\Sigma}_t^{\text{res}} = (\lambda I_{p^{\text{res}}} + \hat{\Sigma}_t(X^{\text{res}})) \times \tilde{\Sigma}_t^{\text{res}}$ 为 $\hat{\Sigma}_t(X^{\text{res}})$ 的定义取值，并令 $t \in [T]$，我们有

$$\|\hat{\delta}_t^{\text{res}} - \delta_t^{\text{res}}\|_{\hat{\Sigma}_{t-1}} = \|\tilde{\delta}_t^{\text{res}} - (U^{\text{res}})^{\top} \delta_t^{\text{res}}\|_{\tilde{\Sigma}_{t-1}^{\text{res}}}$$

$$
\begin{aligned}
&= \|(\tilde{\Sigma}^{\text{res}}_{t-1})^{-1}\sum_{\tau=1}^{t-1}(U^{\text{res}})^\top\varphi(X_\tau,a_\tau)(\varphi(X_\tau,a_\tau)^\top\delta^{\text{res}}_\tau+\epsilon_\tau)-(U^{\text{res}})^\top\delta^{\text{res}}_t\|_{\tilde{\Sigma}^{\text{res}}_{t-1}}\\
&= \|(\tilde{\Sigma}^{\text{res}}_{t-1})^{-1}\sum_{\tau=1}^{t-1}(U^{\text{res}})^\top\varphi(X_\tau,a_\tau)\epsilon_\tau-\lambda(\tilde{\Sigma}^{\text{res}}_{t-1})^{-1}(U^{\text{res}})^\top\delta^{\text{res}}_t\|_{\tilde{\Sigma}^{\text{res}}_{t-1}}\\
&= \|\sum_{\tau=1}^{t-1}(U^{\text{res}})^\top\varphi(X_\tau,a_\tau)\epsilon_\tau-\lambda(U^{\text{res}})^\top\delta^{\text{res}}_t\|_{(\tilde{\Sigma}^{\text{res}}_{t-1})^{-1}}\\
&\le \|\sum_{\tau=1}^{t-1}(U^{\text{res}})^\top\varphi(X_\tau,a_\tau)\epsilon_\tau\|_{(\tilde{\Sigma}^{\text{res}}_{t-1})^{-1}}+\|\lambda(U^{\text{res}})^\top\delta^{\text{res}}_t\|_{(\tilde{\Sigma}^{\text{res}}_{t-1})^{-1}}\\
&\le \|\sum_{\tau=1}^{t-1}(U^{\text{res}})^\top\varphi(X_\tau,a_\tau)\epsilon_\tau\|_{(\tilde{\Sigma}^{\text{res}}_{t-1})^{-1}}+\sqrt{\lambda}\|\delta^{\text{res}}_t\|_2 \qquad (11)
\end{aligned}
$$

由引理 1可得，对于所有 $\eta\in(0,\ 1)$，以至少 $1-\eta$ 的概率，

$$
\|\hat{\delta}^{\text{res}}_t-\delta^{\text{res}}_t\|_{\hat{\Sigma}_{t-1}}\le\sigma\sqrt{2\log\left(\frac{1}{\eta}\right)+\log\left(\frac{\det(\tilde{\Sigma}^{\text{res}}_{t-1})}{\det(\lambda I_{p^{\text{res}}})}\right)}+\sqrt{\lambda}\|\delta^{\text{res}}_t\|_2
$$

假设 L 和 M 在算法中选取得足够大1 ，使其确实分别界定上下文-动作特征和线性参数，我们得到，以至少 $1-\eta$ 的概率，

$$
\|\hat{\delta}^{\text{res}}_t-\delta^{\text{res}}_t\|_{\hat{\Sigma}_{t-1}}\le\sigma\sqrt{2\log\left(\frac{1}{\eta}\right)+p^{\text{res}}\log\left(1+\frac{tL^2}{\lambda p^{\text{res}}}\right)}+\sqrt{\lambda}M=:\sqrt{\hat{\rho}^{\text{res}}_t(\eta,L,M)}.
$$

假设 1 使我们能够分别在不变性子空间和残差子空间上估计 β^{inv} 和 δ^{res}_t，从而估计 $\gamma_{0,\ t}$。这意味着总遗憾由不变性子空间和残差子空间上的遗憾之和构成，如下所示

$$
\begin{aligned}
\text{reg}_t &= (\varphi(X_t,a^*_t)-\varphi(X_t,a_t))^\top\gamma_{0,t}\\
&\le \varphi(X_t,a_t)^\top(\bar{\gamma}_t-\gamma_{0,t})\\
&= \underbrace{\varphi(X_t,a_t)^\top(\bar{\beta}_t-\beta^{\text{inv}})}_{\text{reg}^{\text{inv}}_t}+\underbrace{\varphi(X_t,a_t)^\top(\bar{\delta}_t-\delta^{\text{res}}_t)}_{\text{reg}^{\text{res}}_t}\\
&= \text{reg}^{\text{res}}_t, \qquad (12)
\end{aligned}
$$

其中 $\bar{\gamma}_t=\arg\max_{\gamma\in\hat{\ }\mathcal{C}^\beta\oplus\hat{\ }\mathcal{C}^\delta_t}\varphi(X_t,\ a_t)^\top\gamma$，$(\bar{\beta}_t,\ \bar{\delta}_t):=\arg\max_{\beta\in\hat{\ }\mathcal{C}^\beta,\ \delta\in\hat{\ }\mathcal{C}^\delta_t}\varphi(X_t,\ a_t)^\top(\beta+\delta)$，并且根据构造我们有$\bar{\ }\quad\bar{\ }$。第一个不等式成立，是因为根据定义，$\bar{\gamma}_t$在时间 t最大化了上置信界。最后一个等式来自这样的假设：我们拥有关于 β^{inv}的预言机知识，因此$\hat{\ }\mathcal{C}^\beta=\{\beta^{\text{inv}}\}$ 且 $\text{reg}^{\text{inv}}_t=0$，我们只需考虑残差子空间上的遗憾。于是，对于所有 $t\in[T]$，残差瞬时遗憾满足

$$
\begin{aligned}
\text{reg}^{\text{res}}_t &= \varphi(X_t,a_t)^\top(\bar{\delta}_t-\delta^{\text{res}}_t)\\
&\le \|(U^{\text{res}})^\top\varphi(X_t,a_t)\|_{(\tilde{\Sigma}^{\text{res}}_{t-1})^{-1}}\|(U^{\text{res}})^\top(\bar{\delta}_t-\delta^{\text{res}}_t)\|_{\tilde{\Sigma}^{\text{res}}_{t-1}}\\
&\le 2\|(U^{\text{res}})^\top\varphi(X_t,a_t)\|_{(\tilde{\Sigma}^{\text{res}}_{t-1})^{-1}}\sqrt{\hat{\rho}^{\text{res}}_t(\eta,L,M)}, \qquad (13)
\end{aligned}
$$

第一个不等式由柯西-施瓦茨不等式得出。最后一个不等式是通过在第二个范数中加上再减去$\hat{}\ \delta_t^{\text{res}}$，并利用$\bar{\delta}_t \in \hat{}\mathcal{C}_t^{\delta}$根据定义成立，且以至少$1-\eta$，$\delta_t^{\text{res}} \in \hat{}\mathcal{C}_t^{\delta}$的概率成立。此外，我们有，对于所有$t \in [T]$，遗憾$_t \le 2LM$，且对于$\eta = 1/T$，当$T > 2$时，有$\hat{}\rho_t^{\text{res}}(\eta,\ L,\ M) \ge 1$。结合式（13），可得

$$\begin{aligned}
\text{reg}_t^{\text{res}} &\le \min\{2LM, 2\sqrt{\hat{\rho}_t^{\text{res}}(\eta,L,M)}\|(U^{\text{res}})^\top\varphi(X_t,a_t)\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}\} \\
&\le 2\max\{LM,1\}\sqrt{\hat{\rho}_t^{\text{res}}(\eta,L,M)}\min\{1,\|(U^{\text{res}})^\top\varphi(X_t,a_t)\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}\} \\
&\le C^{\text{res}}\sqrt{\hat{\rho}_t^{\text{res}}(\eta,L,M)}\min\{1,\|(U^{\text{res}})^\top\varphi(X_t,a_t)\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}\}
\end{aligned}$$

其中$C^{\text{res}} := 2\max\{LM,1\}$。利用这一点，对于所有$x \ge 0$，$\min\{1,x\} \le 2\log(1+x)$，我们可以将累积完全遗憾上界为

$$\begin{aligned}
\text{Reg}_T = \sum_{t\in[T]} \text{reg}_T^{\text{res}} &\le \sqrt{T\sum_{t\in[T]}(\text{reg}_t^{\text{res}})^2} \\
&\le C^{\text{res}}\sqrt{T\rho_T^{\text{res}}(\eta,L,M)\sum_{t\in[T]}\log(1+\|(U^{\text{res}})^\top\varphi(X_t,a_t)\|^2_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}})}. \qquad (14)
\end{aligned}$$

最后一步，我们使用以下由 Lattimore 和 Szepesvári (2020年) 提出的引理，并将符号调整为适用于我们的问题。

引理 4 (Lattimore 和 Szepesvári (2020年)，引理 19.4)。考虑与定理 *1*相同的假设。令$\lambda > 0$和$L < \infty$。那么，

$$\sum_{t\in[T]}\log(1+\|(U^{\text{res}})^\top\varphi(X_t,a_t)\|^2_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}) \le p^{\text{res}}\log\left(1+\frac{TL^2}{\lambda p^{\text{res}}}\right).$$

引理 4 意味着

$$\text{Reg}_T \le C^{\text{res}}\sqrt{Tp^{\text{res}}\log\left(1+\frac{TL^2}{\lambda p^{\text{res}}}\right)}\left(\sigma\sqrt{2\log\left(\frac{1}{\eta}\right)+p^{\text{res}}\log\left(1+\frac{TL^2}{\lambda p^{\text{res}}}\right)}+\sqrt{\lambda}M\right).$$

最后，选择$\eta = 1/T$意味着 Reg$_T$ 为 $\tilde{O}(p^{\text{res}}\sqrt{T})$。 □

C.2 预言机子空间

定理 *2* 的证明。我们首先证明，在已知子空间的预言机知识时，置信集 $\sqrt{\hat{}\rho_{T_0}^{\text{inv}}(\eta,L,M)}$ 的半径 $\hat{}\mathcal{C}^{\beta}$（在（7）中引入）可定义为 $\tilde{O}(\sqrt{p^{\text{inv}}}+\sqrt{\frac{1}{\lambda_0}})$。为此，我们考虑（$\hat{}\ \beta^{\text{inv}}$ 在（6）中定义的估计误差，其中 $\hat{}U^{\text{inv}}$ 被 U^{inv}替换。

$$\begin{aligned}
&\|\hat{\beta}^{\text{inv}}-\beta\|_{\hat{\Sigma}_{[-T_0]}} \qquad (15)\\
&= \|U^{\text{inv}}(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}\sum_{t=-T_0}^{-1}(U^{\text{inv}})^\top\varphi(X_t,a_t)R_t^{a_t} - U^{\text{inv}}(U^{\text{inv}})^\top\beta^{\text{inv}}\|_{\hat{\Sigma}_{[-T_0]}} \\
&= \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}\sum_{t=-T_0}^{-1}(U^{\text{inv}})^\top\varphi(X_t,a_t)R_t^{a_t} - (U^{\text{inv}})^\top\beta^{\text{inv}}\|_{\tilde{\Sigma}^{\text{inv}}_{[-T_0]}}
\end{aligned}$$

$$
\begin{aligned}
&= \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1} \sum_{t=-T_0}^{-1} (U^{\text{inv}})^\top \varphi(X_t, a_t)(\varphi(X_t, a_t)^\top(\beta^{\text{inv}} + \delta_t^{\text{res}}) + \epsilon_t) - (U^{\text{inv}})^\top \beta^{\text{inv}}\|_{\tilde{\Sigma}^{\text{inv}}_{[-T_0]}} \\
&= \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1} \sum_{t=-T_0}^{-1} (U^{\text{inv}})^\top \varphi(X_t, a_t)(\varphi(X_t, a_t)^\top U^{\text{res}}(U^{\text{res}})^\top \delta_t^{\text{res}} + \epsilon_t)\|_{\tilde{\Sigma}^{\text{inv}}_{[-T_0]}} \\
&\le \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1} \sum_{t=-T_0}^{-1} (U^{\text{inv}})^\top \varphi(X_t, a_t)\varphi(X_t, a_t)^\top U^{\text{res}}(U^{\text{res}})^\top \delta_t^{\text{res}}\|_{\tilde{\Sigma}^{\text{inv}}_{[-T_0]}} \\
&\quad + \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1} \sum_{t=-T_0}^{-1} (U^{\text{inv}})^\top \varphi(X_t, a_t)\epsilon_t\|_{\tilde{\Sigma}^{\text{inv}}_{[-T_0]}} \\
&= \| \sum_{t=-T_0}^{-1} (U^{\text{inv}})^\top \varphi(X_t, a_t)\varphi(X_t, a_t)^\top U^{\text{res}}(U^{\text{res}})^\top \delta_t^{\text{res}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \qquad (16) \\
&\quad + \|(\tilde{\mathbf{X}}^{\text{inv}}_{T_0})^\top \boldsymbol{\epsilon}_{T_0}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \qquad (17)
\end{aligned}
$$

其中 $\tilde{\Sigma}^{\text{inv}}_{[-T_0]} := (U^{\text{inv}})^\top \hat{\Sigma}_{[-T_0]} U^{\text{inv}}$，$\tilde{\mathbf{X}}^{\text{inv}}_{T_0} := [(U^{\text{inv}})^\top \varphi(X_{-T_0}, a_{-T_0}) \dots, (U^{\text{inv}})^\top \varphi(X_{-1}, a_{-1})]^\top \in \mathbb{R}^{T_0 \times p^{\text{inv}}}$，且 $\boldsymbol{\epsilon}_{T_0} := [\epsilon_{-T_0}, \dots, \epsilon_{-1}]^\top \in \mathbb{R}^{T_0}$。对于（16）中的项，有

$$
\begin{aligned}
&\| \sum_{t=-T_0}^{-1} (U^{\text{inv}})^\top \varphi(X_t, a_t)\varphi(X_t, a_t)^\top U^{\text{res}}(U^{\text{res}})^\top \delta_t^{\text{res}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \\
\le &\|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-\frac{1}{2}}\|_{\text{op}} \| \sum_{t=-T_0}^{-1} (U^{\text{inv}})^\top \varphi(X_t, a_t)\varphi(X_t, a_t)^\top U^{\text{res}}(U^{\text{res}})^\top \delta_t^{\text{res}}\|_2 \\
\le &\sqrt{\frac{1}{\lambda_0 T_0}} \| \sum_{t=-T_0}^{-1} \alpha_t\|_2 \qquad (18)
\end{aligned}
$$

其中 $\alpha_t := (U_{\text{inv}})^\top \varphi(X_t, a_t)\, \varphi(X_t, a_t)^\top U_{\text{res}} (U_{\text{res}})^\top \delta_t^{\text{res}} \in \mathbb{R}^{p^{\text{inv}}}$。我们给出 $\| \sum_{t=-T_0}^{-1} \alpha_t\|_2$的上界，所用工具为矩阵霍夫丁不等式（例如参见Tropp，2012年，定理 1.3，我们将在下文中陈述该定理）。

引理 5（矩阵霍夫丁不等式）。设 $\{\mathbf{M}_k\}_{k\in[K]}$ 为维数为 p 的独立随机自伴矩阵的有限序列，并设 $\{\mathbf{A}_k\}_{k\in[K]}$ 为维数为 p 的固定自伴矩阵序列。假设对所有 $k \in [K]$，$\mathbb{E}[\mathbf{M}_k] = 0$ 且 $\mathbf{M}_k^2 \preceq \mathbf{A}_k$ 几乎处处成立。则对所有 $\xi > 0$，有

$$
\mathbb{P}\left(\lambda_{\max}\left(\sum_{k\in[K]} \mathbf{M}_k\right) \ge \xi\right) \le p \exp\left(-\frac{\xi^2}{8\sigma^2}\right) \quad \textit{where} \quad \sigma^2 := \left\|\sum_{k\in[K]} \mathbf{A}_k^2\right\|.
$$

根据 ISD 的定义，对于所有 $t \in [-T_0]$，$\mathbb{E}[\alpha_t] = 0$。利用柯西-施瓦茨不等式，我们还可得 $\|\alpha_t\|_2 \le L^2 M$。此外，我们可以为 Tropp，2012年，第 2.6 节中所述的自伴扩张构造为 α_t 如下

$$
\mathscr{S}(\alpha_t) := \begin{bmatrix} 0 & \alpha_t \\ \alpha_t^\top & 0 \end{bmatrix} \in \mathbb{R}^{(p^{\text{inv}}+1)\times(p^{\text{inv}}+1)},
$$

满足 $\|\alpha_t\|_2 = \|\mathscr{S}(\alpha_t)\|_{\mathrm{op}} = \lambda_{\max}(\mathscr{S}(\alpha_t))$，并且

$$\mathscr{S}(\alpha_t)^2 := \begin{bmatrix} \alpha_t\alpha_t^\top & 0 \\ 0 & \alpha_t^\top\alpha_t \end{bmatrix}.$$

现在可以将霍夫丁不等式应用于自伴矩阵序列 $(\mathscr{S}(\alpha_t))_{t\in[-T_0]}$。令 $\sigma_\alpha^2 := \|\sum_{t\in[-T_0]}\mathscr{S}(\alpha_t)^2\|_{\mathrm{op}} \le \sum_{t\in[-T_0]}\|\mathscr{S}(\alpha_t)\|_{\mathrm{op}}^2 \le T_0(L^2M)^2$。那么，对于所有 $\xi \ge 0$，根据霍夫丁不等式可得

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{t\in[-T_0]}\mathscr{S}(\alpha_t)\right) \ge \xi\right) \le (p^{\mathrm{inv}}+1)e^{-\frac{\xi^2}{8\sigma_\alpha^2}}.$$

利用 $\mathscr{S}(\alpha_t)$ 的定义，并令 $\eta := (p^{\mathrm{inv}}+1)e^{-\frac{\xi^2}{8\sigma_\alpha^2}}$，上式等价于—

$$\mathbb{P}\left(\|\sum_{t\in[-T_0]}\alpha_t\|_2 \le 2L^2M\sqrt{2T_0\log\left(\frac{p^{\mathrm{inv}}+1}{\eta}\right)}\right) \ge 1-\eta.$$

因此，以至少 $1-\eta$ 的概率，$\sqrt{\frac{1}{\lambda_0 T_0}}\|\sum_{t=-T_0}^{-1}\alpha_t\|_2$ 为 $O\left(\sqrt{\frac{1}{\lambda_0}\log\left(\frac{p^{\mathrm{inv}}}{\eta}\right)}\right)$。

现在需要对式 $\|(\tilde{\mathbf{X}}_{T_0}^{\mathrm{inv}})^\top\boldsymbol{\epsilon}_{T_0}\|_{(\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}})^{-1}}$（17）中的项进行界定。根据假设2，有 $\hat{\Sigma}_{[-T_0]} \succeq \lambda_0 T_0 I_p$，这意味着 $\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}} \succeq \lambda_0 T_0 I_{p^{\mathrm{inv}}}$，因此 $\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}} \succeq 1/2(\lambda_0 T_0 I_{p^{\mathrm{inv}}} + \tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}})$。这意味着，

$$\|(\tilde{\mathbf{X}}_{T_0}^{\mathrm{inv}})^\top\boldsymbol{\epsilon}_{T_0}\|_{(\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}})^{-1}} \le \sqrt{2}\|(\tilde{\mathbf{X}}_{T_0}^{\mathrm{inv}})^\top\boldsymbol{\epsilon}_{T_0}\|_{(\lambda_0 T_0 I_{p^{\mathrm{inv}}}+\tilde{\Sigma}_{T_0}^{\mathrm{inv}})^{-1}}$$

引理1 表明，对于所有 $\eta\in(0,1)$，其成立的概率至少为 $1-\eta$，

$$\|(\tilde{\mathbf{X}}_{T_0}^{\mathrm{inv}})^\top\boldsymbol{\epsilon}_{T_0}\|_{(\lambda_0 T_0 I_{p^{\mathrm{inv}}}+\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}})^{-1}} \le \sigma\sqrt{\log\left(\frac{1}{\eta}\right)+\frac{1}{2}\log\left(\frac{\det(\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}})}{\det(\lambda_0 T_0 I_{p^{\mathrm{inv}}})}\right)}$$

根据算术-几何平均不等式，有

$$\det(\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}}) \le \left(\frac{1}{p^{\mathrm{inv}}}\operatorname{trace}(\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}})\right)^{p^{\mathrm{inv}}} \le \left(\frac{T_0L^2}{p^{\mathrm{inv}}}\right)^{p^{\mathrm{inv}}}$$

其中，最后一个不等式由迹$(\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}})$ = 迹$((\tilde{\mathbf{X}}_{T_0}^{\mathrm{inv}})^\top\tilde{\mathbf{X}}_{T_0}^{\mathrm{inv}}) \le T_0L^2$ 得出。此外，$\det(\lambda_0 T_0 I_{p^{\mathrm{inv}}}) = (\lambda_0 T_0)^{p^{\mathrm{inv}}}$。因此，其成立的概率至少为 $1-\eta$，

$$\|(\tilde{\mathbf{X}}_{T_0}^{\mathrm{inv}})^\top\boldsymbol{\epsilon}_{T_0}\|_{(\lambda_0 T_0 I_{p^{\mathrm{inv}}}+\tilde{\Sigma}_{[-T_0]}^{\mathrm{inv}})^{-1}} \le \sigma\sqrt{\log\left(\frac{1}{\eta}\right)+\frac{p^{\mathrm{inv}}}{2}\log\left(\frac{T_0L^2}{p^{\mathrm{inv}}\lambda_0 T_0}\right)} \tag{19}$$

并且，以至少 $1-2\eta$ 的概率，

$$\begin{aligned}&\|\hat{\beta}^{\mathrm{inv}}-\beta\|_{\hat{\Sigma}_{[-T_0]}}\\ \le&\sigma\sqrt{2\log\left(\frac{1}{\eta}\right)+p^{\mathrm{inv}}\log\left(\frac{L^2}{p^{\mathrm{inv}}\lambda_0}\right)}+2L^2M\sqrt{\frac{2}{\lambda_0}\log\left(\frac{p^{\mathrm{inv}}+1}{\eta}\right)} =: \hat{\rho}_{T_0}^{\mathrm{inv}}(\eta,L,M)\end{aligned} \tag{20}$$

根据假设 2，我们进一步得到，以至少 $1-2\eta$,

$$\|\hat{\beta}^{\text{inv}} - \beta\|_2 \le \sqrt{\frac{\hat{\rho}^{\text{inv}}_{T_0}(\eta, L, M)}{\lambda_0 T_0}} \tag{21}$$

现在考虑算法 1。对于所有 $a \in \mathcal{A}$，令 $\text{UCB}_t^{\beta\,\delta}(a) := \max_{\beta \in \hat{\mathcal{C}}^{\beta}, \delta \in \hat{\mathcal{C}}_t^{\delta}} \varphi(X_t, a)^\top(\beta+\delta)$，并令 $(\bar{\beta}_t, \bar{\delta}_t) := \arg\max_{\beta \in \hat{\mathcal{C}}^{\beta}, \delta \in \hat{\mathcal{C}}_t^{\delta}} \varphi(X_t, a_t)^\top(\beta+\delta)$。那么，与（12）中一样，对于所有 $t \in [T]$，瞬时遗憾满足 $\text{reg}_t \le \text{reg}_t^{\text{inv}} + \text{reg}_t^{\text{res}}$. 。我们对 $\text{reg}_t^{\text{inv}}$作如下上界估计

$$\begin{aligned}
\text{reg}_t^{\text{inv}} &:= ((U^{\text{inv}})^\top \varphi(X_t, a_t))^\top (U^{\text{inv}})^\top (\bar{\beta}_t - \beta^{\text{inv}}) \\
&\le \|(U^{\text{inv}})^\top \varphi(X_t, a_t)\|_2 \|(U^{\text{inv}})^\top (\bar{\beta}_t - \beta^{\text{inv}})\|_2 \\
&\le \frac{2L}{\sqrt{\lambda_0 T_0}} \left(\sigma \sqrt{2\log\left(\frac{1}{\eta}\right) + p^{\text{inv}} \log\left(\frac{L^2}{p^{\text{inv}}\lambda_0}\right)} + 2L^2 M \sqrt{\frac{2}{\lambda_0}\log\left(\frac{p^{\text{inv}}+1}{\eta}\right)} \right)
\end{aligned}$$

其中，第一个不等式以概率 $1-2\eta$ 成立，且由 $\hat{\mathcal{C}}^{\beta}$ 的定义以及柯西-施瓦茨不等式可得。

对于$\text{reg}_t^{\text{res}}$，定理 1 中的分析仍然成立，但由于使用估计的不变分量而非预言机给出的不变分量，还需考虑一个额外项。具体而言，按照（11）中的相同步骤，我们有，对于所有 $t \in [T]$,

$$\begin{aligned}
\|\hat{\delta}_t^{\text{res}} - \delta_t^{\text{res}}\|_{\hat{\Sigma}_{t-1}} \le{}& \|\sum_{\tau=1}^{t-1} (U^{\text{res}})^\top \varphi(X_\tau, a_\tau)\epsilon_\tau\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} + \sqrt{\lambda}\|\delta_t^{\text{res}}\|_2 \\
&+ \|\sum_{\tau=1}^{t-1} (U^{\text{res}})^\top \varphi(X_\tau, a_\tau)\varphi(X_\tau, a_\tau)^\top (\beta^{\text{inv}} - \hat{\beta}^{\text{inv}})\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}.
\end{aligned}$$

我们分析了定理 1 中的前两项。对于最后一项，我们有

$$\begin{aligned}
&\|\sum_{\tau=1}^{t-1} (U^{\text{res}})^\top \varphi(X_\tau, a_\tau)\varphi(X_\tau, a_\tau)^\top (\beta^{\text{inv}} - \hat{\beta}^{\text{inv}})\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} \\
\le{}& \|(\tilde{\Sigma}_{t-1}^{\text{res}})^{-\frac{1}{2}}\|_{\text{op}} \|\sum_{\tau=1}^{t-1} (U^{\text{res}})^\top \varphi(X_\tau, a_\tau)\varphi(X_\tau, a_\tau)^\top U^{\text{inv}}\|_{\text{op}} \|(U^{\text{inv}})^\top (\beta^{\text{inv}} - \hat{\beta}^{\text{inv}})\|_2.
\end{aligned}$$

使用矩阵霍夫丁不等式（参见引理 5），并按照用于界定（16）的相同步骤可得，对于所有 $\eta \in (0,1)$，以至少 $1-\eta$ 的概率,

$$\|\sum_{\tau=1}^{t-1} (U^{\text{res}})^\top \varphi(X_\tau, a_\tau)\varphi(X_\tau, a_\tau)^\top U^{\text{inv}}\|_{\text{op}} \le 2L^2 \sqrt{2(t-1)\log\left(\frac{p+1}{\eta}\right)}.$$

因此，以至少 $1-3\eta$ 的概率,

$$\begin{aligned}
&\|\sum_{\tau=1}^{t-1} (U^{\text{res}})^\top \varphi(X_\tau, a_\tau)\varphi(X_\tau, a_\tau)^\top (\beta^{\text{inv}} - \hat{\beta}^{\text{inv}})\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} \\
\le{}& \frac{1}{\sqrt{\lambda}} 2L^2 \sqrt{2(t-1)\log\left(\frac{p+1}{\eta}\right)} \sqrt{\frac{\hat{\rho}^{\text{inv}}_{T_0}(\eta, L, M)}{\lambda_0 T_0}}.
\end{aligned}$$

不等式右侧可以加到 $\sqrt{\hat{}}\rho_t^{\text{res}}(\eta, L, M)$ 定理1证明中得到的结果上，从而以至少 $1-4\eta$

$$\sum_{t\in[T]} \text{reg}_t^{\text{res}} \le C^{\text{res}} \sqrt{Tp^{\text{res}} \log\left(1+\frac{TL^2}{\lambda p^{\text{res}}}\right)} \times \left(\sigma\sqrt{2\log\left(\frac{1}{\eta}\right)+p^{\text{res}}\log\left(1+\frac{TL^2}{\lambda p^{\text{res}}}\right)}+\sqrt{\lambda}M2L^2\sqrt{\frac{2T\hat{\rho}_{T_0}^{\text{inv}}(\eta,L,M)}{\lambda\lambda_0 T_0}\log\left(\frac{p+1}{\eta}\right)}\right).$$

最后，这意味着对于所有 $\eta \in (0, 1/5)$，累积遗憾以至少 $1-5\eta$ 的概率可被上界为 –

$$\begin{aligned}\text{Reg}_T &= \sum_{t\in[T]} \text{reg}_t = \sum_{t\in[T]} \text{reg}^{\text{inv}} + \sum_{t\in[T]} \text{reg}^{\text{res}} \\ &\le \frac{2LT}{\sqrt{\lambda_0 T_0}}\left(\sigma\sqrt{2\log\left(\frac{1}{\eta}\right)+p^{\text{inv}}\log\left(\frac{L^2}{p^{\text{inv}}\lambda_0}\right)}+2L^2M\sqrt{\frac{2}{\lambda_0}\log\left(\frac{p^{\text{inv}}+1}{\eta}\right)}\right) \\ &+ C^{\text{res}}\sqrt{Tp^{\text{res}}\log\left(1+\frac{TL^2}{\lambda p^{\text{res}}}\right)} \\ &\times\left(\sigma\sqrt{2\log\left(\frac{1}{\eta}\right)+p^{\text{res}}\log\left(1+\frac{TL^2}{\lambda p^{\text{res}}}\right)}+\sqrt{\lambda}M+2L^2\sqrt{\frac{2T\hat{\rho}_{T_0}^{\text{inv}}(\eta,L,M)}{\lambda\lambda_0 T_0}\log\left(\frac{p+1}{\eta}\right)}\right)\end{aligned}$$

选择 $\eta' = 1/T$，其中 $\eta' := 5\eta$，可得 Reg_T 以至少 $1-\eta'$
$\tilde{O}\left(\sqrt{T}\left(p^{\text{res}}+\sqrt{\frac{p^{\text{inv}}T}{\lambda_0 T_0}}+\frac{1}{\lambda_0}\sqrt{\frac{T}{T_0}}\right)\right)$ 的概率成立。 □

C.3 子空间分解误差的处理

$\hat{}$ 引理 2 的证明。我们希望对估计的不变子空间 $\hat{}\mathcal{S}^{\text{inv}}$ 中不变分量的估计误差的 $\Sigma_{[-T_0]}$-范数进行界定，即 $\|\hat{}\beta^{\text{inv}} - \hat{}\Pi^{\mathcal{S}^{\text{inv}}}\beta^{\text{inv}}\|_{\hat{\Sigma}_{[-T_0]}}$。令 $\tilde{\Sigma}^{\text{inv}}_{[-T_0]} := (\hat{}U^{\text{inv}})^\top \hat{}\Sigma_{[-T_0]} \hat{}U^{\text{inv}}$、$\mathbf{X}_{T_0} := [\varphi(X_{-T_0}, a_{-T_0}) \dots, \varphi(X_{-1}, a_{-1})]^\top \in \mathbb{R}^{T_0\times p}$、$\mathbf{R}_{T_0} := [R_{-T_0}^{a_{-T_0}}, \dots, R_{-1}^{a_{-1}}]^\top \in \mathbb{R}^{T_0}$ 和 $\boldsymbol{\epsilon}_{T_0} := [\epsilon_{-T_0}, \dots, \epsilon_{-1}]^\top \in \mathbb{R}^{T_0}$。于是，我们有

$$\begin{aligned}\|\hat{\beta}^{\text{inv}} - \hat{\Pi}^{\mathcal{S}^{\text{inv}}}\beta^{\text{inv}}\|_{\hat{\Sigma}_{[-T_0]}} &= \|\hat{U}^{\text{inv}}((\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}\sum_{t=-T_0}^{-1}(\hat{U}^{\text{inv}})^\top\varphi(X_t,a_t)R_t^{a_t} - (\hat{U}^{\text{inv}})^\top\beta^{\text{inv}})\|_{\hat{\Sigma}_{[-T_0]}} \\ &\le \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}(\hat{U}^{\text{inv}})^\top\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0}\beta^{\text{inv}} - (\hat{U}^{\text{inv}})^\top\beta^{\text{inv}}\|_{\tilde{\Sigma}^{\text{inv}}_{[-T_0]}} \\ &+ \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}\sum_{t=-T_0}^{-1}(\hat{U}^{\text{inv}})^\top\varphi(X_t,a_t)\varphi(X_t,a_t)^\top\delta_t^{\text{res}}\|_{\tilde{\Sigma}^{\text{inv}}_{[-T_0]}} \\ &+ \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}(\hat{U}^{\text{inv}})^\top\mathbf{X}_{T_0}^\top\boldsymbol{\epsilon}_{T_0}\|_{\tilde{\Sigma}^{\text{inv}}_{[-T_0]}} \\ &= \|(\hat{U}^{\text{inv}})^\top\mathbf{X}_{T_0}^\top\mathbf{X}_{T_0}(U^{\text{inv}}(U^{\text{inv}})^\top - \hat{U}^{\text{inv}}(\hat{U}^{\text{inv}})^\top)\beta^{\text{inv}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \quad (22) \\ &+ \|\sum_{t=-T_0}^{-1}(\hat{U}^{\text{inv}})^\top\varphi(X_t,a_t)\varphi(X_t,a_t)^\top\delta_t^{\text{res}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \quad (23)\end{aligned}$$

$$+\|(\hat{U}^{\text{inv}})^\top \mathbf{X}_{T_0}^\top \boldsymbol{\epsilon}_{T_0}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \tag{24}$$

（24）中的项可以完全按照第 C.2 节（见公式（19））中的方法进行上界估计，因为用 $\hat{U}^{\text{inv}}$ 代替 U^{inv} 不会影响其分析。（22）和（23）中的项源于真实不变子空间与估计的不变子空间之间的错位：前者表示由于 $\hat{\mathcal{S}}^{\text{inv}}$ 与 $\mathcal{S}^{\text{inv}}$之间的差异，我们无法估计出多少不变参数；后者则量化了由于 $\hat{\mathcal{S}}^{\text{inv}}$ 与 $\mathcal{S}^{\text{res}}$相交，有多少真实残差参数进入了不变分量的估计。对于（22），我们有

$$\begin{aligned}
&\|(\hat{U}^{\text{inv}})^\top \mathbf{X}_{T_0}^\top \mathbf{X}_{T_0}(\Pi^{\mathcal{S}^{\text{inv}}} - \hat{\Pi}^{\mathcal{S}^{\text{inv}}})\beta^{\text{inv}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \\
=&\|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-\frac{1}{2}}(\hat{U}^{\text{inv}})^\top \mathbf{X}_{T_0}^\top \mathbf{X}_{T_0}(\Pi^{\mathcal{S}^{\text{inv}}} - \hat{\Pi}^{\mathcal{S}^{\text{inv}}})\beta^{\text{inv}}\|_2 \\
\leq&\sqrt{\sum_{t\in[-T_0]} \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-\frac{1}{2}}(\hat{U}^{\text{inv}})^\top \varphi(X_t,a_t)\|_2^2}\sqrt{\sum_{t\in[-T_0]} (\varphi(X_t,a_t)^\top(\Pi^{\mathcal{S}^{\text{inv}}} - \hat{\Pi}^{\mathcal{S}^{\text{inv}}})\beta^{\text{inv}})^2} \\
\leq&\sqrt{p^{\text{inv}}}\sqrt{T_0 L^2 M^2 (\Delta\Pi)^2} =: \hat{\rho}_1^{\text{inv}}
\end{aligned}$$

第一个不等式用到了柯西-施瓦茨不等式。第二个不等式由以下事实得出：

$$\begin{aligned}
&\sum_{t\in[-T_0]} \|(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-\frac{1}{2}}(\hat{U}^{\text{inv}})^\top \varphi(X_t,a_t)\|_2^2 \\
=&\sum_{t\in[-T_0]} \varphi(X_t,a_t)^\top \hat{U}^{\text{inv}}(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}(\hat{U}^{\text{inv}})^\top \varphi(X_t,a_t) \\
=&\operatorname{trace}\left(\sum_{t\in[-T_0]} (\hat{U}^{\text{inv}})^\top \varphi(X_t,a_t)\varphi(X_t,a_t)^\top \hat{U}^{\text{inv}}(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}\right) \\
=&\operatorname{trace}\left(I_{p^{\text{inv}}}\right) \\
=&p^{\text{inv}},
\end{aligned}$$

并将第二个平方根下各因子单独的范数提取出来。最后，根据假设3，对于所有 $\eta \in (0,1)$，以至少 $1-\eta$ 的概率有$\|(\hat{U}^{\text{inv}})^\top \mathbf{X}_{T_0}^\top \mathbf{X}_{T_0}(\Pi^{\mathcal{S}^{\text{inv}}} - \hat{\Pi}^{\mathcal{S}^{\text{inv}}})\beta^{\text{inv}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}}$为 $O(\sqrt{p^{\text{inv}}\log(\frac{p}{\eta})})$。

现在我们需要对式（23）中的项给出上界，即

$$\begin{aligned}
&\|\sum_{t=-T_0}^{-1} (\hat{U}^{\text{inv}})^\top \varphi(X_t,a_t)\varphi(X_t,a_t)^\top \delta_t^{\text{res}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \\
=&\|\sum_{t=-T_0}^{-1} (\hat{U}^{\text{inv}})^\top (U^{\text{inv}}(U^{\text{inv}})^\top + U^{\text{res}}(U^{\text{res}})^\top)\varphi(X_t,a_t)\varphi(X_t,a_t)^\top U^{\text{res}}(U^{\text{res}})^\top \delta_t^{\text{res}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \\
\leq&\|\sum_{t=-T_0}^{-1} (\hat{U}^{\text{inv}})^\top U^{\text{inv}}(U^{\text{inv}})^\top \varphi(X_t,a_t)\varphi(X_t,a_t)^\top U^{\text{res}}(U^{\text{res}})^\top \delta_t^{\text{res}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}} \\
+&\|\sum_{t=-T_0}^{-1} (\hat{U}^{\text{inv}})^\top U^{\text{res}}(U^{\text{res}})^\top \varphi(X_t,a_t)\varphi(X_t,a_t)^\top U^{\text{res}}(U^{\text{res}})^\top \delta_t^{\text{res}}\|_{(\tilde{\Sigma}^{\text{inv}}_{[-T_0]})^{-1}}
\end{aligned}$$

对于第二项，有

$$
\begin{aligned}
&\| \sum_{t=-T_0}^{-1} (\hat{U}^{\mathrm{inv}})^\top U^{\mathrm{res}} (U^{\mathrm{res}})^\top \varphi(X_t, a_t) \varphi(X_t, a_t)^\top U^{\mathrm{res}} (U^{\mathrm{res}})^\top \delta_t^{\mathrm{res}} \|_{(\tilde{\Sigma}^{\mathrm{inv}}_{[-T_0]})^{-1}} \\
\leq & \|(\tilde{\Sigma}^{\mathrm{inv}}_{[-T_0]})^{-\frac{1}{2}}\|_{\mathrm{op}} \|(\hat{U}^{\mathrm{inv}})^\top U^{\mathrm{res}}\|_{\mathrm{op}} T_0 L^2 M \\
\leq & \sqrt{\frac{1}{\lambda_0 T_0}} \Delta\Pi T_0 L^2 M \\
= & \sqrt{\frac{T_0}{\lambda_0}} \Delta\Pi L^2 M =: \hat{\rho}_2^{\mathrm{inv}}.
\end{aligned}
$$

在假设 3下，对于所有 $\eta \in (0, 1)$，该量以至少 $1-\eta$ 的概率满足 $O\left(\sqrt{\frac{\log(p/\eta}{\lambda_0}}\right)$。对于第一项，我们得到

$$
\begin{aligned}
&\| \sum_{t=-T_0}^{-1} (\hat{U}^{\mathrm{inv}})^\top U^{\mathrm{inv}} (U^{\mathrm{inv}})^\top \varphi(X_t, a_t) \varphi(X_t, a_t)^\top U^{\mathrm{res}} (U^{\mathrm{res}})^\top \delta_t^{\mathrm{res}} \|_{(\tilde{\Sigma}^{\mathrm{inv}}_{[-T_0]})^{-1}} \\
\leq & \|(\tilde{\Sigma}^{\mathrm{inv}}_{[-T_0]})^{-\frac{1}{2}}\|_{\mathrm{op}} \|(\hat{U}^{\mathrm{inv}})^\top U^{\mathrm{inv}}\|_{\mathrm{op}} \| \sum_{t=-T_0}^{-1} (U^{\mathrm{inv}})^\top \varphi(X_t, a_t) \varphi(X_t, a_t)^\top U^{\mathrm{res}} (U^{\mathrm{res}})^\top \delta_t^{\mathrm{res}} \|_2 \\
\leq & \sqrt{\frac{1}{\lambda_0 T_0}} \| \sum_{t=-T_0}^{-1} \alpha_t \|_2
\end{aligned}
$$

其中 $\alpha_t := (U^{\mathrm{inv}})^\top \varphi(X_t, a_t) \varphi(X_t, a_t)^\top U^{\mathrm{res}} (U\mathrm{res})^\top \delta_t^{\mathrm{res}} \in \mathbb{R}^{p^{\mathrm{inv}}}$；在最后一个不等式中，我们利用了 $\|(\hat{U}^{\mathrm{inv}})^\top U\mathrm{inv}\|_{\mathrm{op}} \leq 1$，因为这两个矩阵都有标准正交列。这与预言机子空间情形下式 (18) 中出现的量相同（连同式 (24)，构成式 (20) 给出的完整神谕界，我们在此记为 $\hat{\rho}_3^{\mathrm{inv}}$）。我们已证明，对于所有 $\eta \in (0, 1)$，式 (18) 以至少 $1-\eta$ 的概率满足 $O\left(\sqrt{\frac{1}{\lambda_0} \log \binom{p^{\mathrm{inv}}}{\eta}}\right)$。对于相同的 η 取值，该项不超过 $O\left(\sqrt{\frac{1}{\lambda_0} \log \binom{p}{\eta}}\right)$。因此，式 (23) 中的表达式以至少 $1-2\eta$ 的概率满足 $O\left(\sqrt{\frac{1}{\lambda_0} \log \binom{p^{\mathrm{inv}}}{\eta}}\right)$。

Let

$$
\begin{aligned}
\hat{\rho}_{T_0}^{\mathrm{inv}}(\eta, L, M) := \sum_{i=1}^{3} \hat{\rho}_i^{\mathrm{inv}} = \sigma \sqrt{2 \log\left(\frac{1}{\eta}\right) + p^{\mathrm{inv}} \log\left(\frac{L^2}{p^{\mathrm{inv}} \lambda_0}\right)} + 2L^2 M \sqrt{\frac{2}{\lambda_0} \log\left(\frac{p^{\mathrm{inv}} + 1}{\eta}\right)} \\
+ \sqrt{p^{\mathrm{inv}} T_0} \Delta\Pi L M + \sqrt{\frac{T_0}{\lambda_0}} \Delta\Pi L^2 M.
\end{aligned} \tag{25}
$$

根据并集界，以至少 $1-4\eta$ 的概率，

$$
\|\hat{\beta}^{\mathrm{inv}} - \hat{\Pi}^{S^{\mathrm{inv}}} \beta^{\mathrm{inv}}\|_{\hat{\Sigma}_{[-T_0]}} \leq \sum_{i=1}^{3} \hat{\rho}_i^{\mathrm{inv}} =: \hat{\rho}_{T_0}^{\mathrm{inv}}(\eta, L, M)
$$

is

$$O\left(\sqrt{\left(\log\left(\frac{1}{\eta}\right)+p^{\text{inv}}\log\left(\frac{1}{p^{\text{inv}}\lambda_0}\right)\right)}\right)+O\left(\sqrt{p^{\text{inv}}\log\left(\frac{p}{\eta}\right)}\right)+O\left(\sqrt{\frac{1}{\lambda_0}\log\left(\frac{p}{\eta}\right)}\right).$$

引理的证明至此完成。 □

引理的证明 3。我们希望界定残差子空间中残差分量的估计误差，即 $\|\hat{}\delta_t^{\text{res}}-\hat{}\Pi^{\mathcal{S}^{\text{res}}}\delta_t^{\text{res}}\|_{\hat{\Sigma}_{t-1}}$。回顾我们已定义 $\tilde{\Sigma}_t^{\text{res}}:=(\hat{}U^{\text{res}})^\top\hat{}\Sigma_t\hat{}U^{\text{res}}$。我们首先明确写出 δ_t^{res}的表达式，得到

$$\begin{aligned}
&\|\hat{\delta}_t^{\text{res}}-\hat{\Pi}^{\mathcal{S}^{\text{res}}}\delta_t^{\text{res}}\|_{\hat{\Sigma}_{t-1}}\\
=&\|\hat{U}^{\text{res}}(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)(R_\tau^{a_\tau}-\varphi(X_\tau,a_\tau)^\top\hat{\beta}^{\text{inv}})-\hat{U}^{\text{res}}(\hat{U}^{\text{res}})^\top\delta_t^{\text{res}}\|_{\hat{\Sigma}_{t-1}}\\
=&\|(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)(R_\tau^{a_\tau}-\varphi(X_\tau,a_\tau)^\top\hat{\beta}^{\text{inv}})-(\hat{U}^{\text{res}})^\top\delta_t^{\text{res}}\|_{\tilde{\Sigma}_{t-1}^{\text{res}}}\\
\le&\|\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\varphi(X_\tau,a_\tau)^\top(U^{\text{res}}(U^{\text{res}})^\top-\hat{U}^{\text{res}}(\hat{U}^{\text{res}})^\top)\delta_t^{\text{res}}-\lambda(\hat{U}^{\text{res}})^\top\delta_t^{\text{res}}\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}\\
+&\|\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\varphi(X_\tau,a_\tau)^\top(\beta^{\text{inv}}-\hat{\beta}^{\text{inv}})\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}\\
+&\|\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\epsilon_\tau\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}\\
\le&\|\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\varphi(X_\tau,a_\tau)^\top(U^{\text{res}}(U^{\text{res}})^\top-\hat{U}^{\text{res}}(\hat{U}^{\text{res}})^\top)\delta_t^{\text{res}}\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} &(26)\\
+&\|\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\varphi(X_\tau,a_\tau)^\top(\beta^{\text{inv}}-\hat{\beta}^{\text{inv}})\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} &(27)\\
+&\|\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\epsilon_\tau\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}+\sqrt{\lambda}\|\delta_t^{\text{res}}\|_2 &(28)
\end{aligned}$$

对式（28）中各项的分析遵循与预言机情形下式（11）相同的步骤。特别地，根据定理 1，对于所有 $\eta\in(0,1)$，以至少 $1-\eta$ 的概率，

$$\begin{aligned}
&\|\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\epsilon_\tau\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}+\sqrt{\lambda}\|\delta_t^{\text{res}}\|_2\\
\le&\sigma\sqrt{2\log\left(\frac{1}{\eta}\right)+p^{\text{res}}\log\left(1+\frac{tL^2}{\lambda p^{\text{res}}}\right)}+\sqrt{\lambda}M=:\hat{\rho}_1^{\text{res}}.
\end{aligned}$$

对于（26）中的项，我们有

$$\begin{aligned}
&\|\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\varphi(X_\tau,a_\tau)^\top(\Pi^{\mathcal{S}^{\text{res}}}-\hat{\Pi}^{\mathcal{S}^{\text{res}}})\delta_t^{\text{res}}\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}}\\
\le&\sqrt{\sum_{\tau\in[t-1]}\|(\tilde{\Sigma}_{t-1}^{\text{res}})^{-\frac{1}{2}}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\|_2^2}\sqrt{\sum_{\tau\in[t-1]}(\varphi(X_\tau,a_\tau)^\top(\Pi^{\mathcal{S}^{\text{res}}}-\hat{\Pi}^{\mathcal{S}^{\text{res}}})\delta_t^{\text{res}})^2}\\
\le&\sqrt{p^{\text{res}}}\sqrt{t\max_{\tau\in[t-1]}\|\varphi(X_\tau,a_\tau)\|_2^2\|\delta_t^{\text{res}}\|_2^2\|\Pi^{\mathcal{S}^{\text{res}}}-\hat{\Pi}^{\mathcal{S}^{\text{res}}}\|_{\text{op}}^2}
\end{aligned}$$

$$\leq LM\Delta\Pi\sqrt{p^{\text{res}}t} =: \hat{\rho}_2^{\text{res}}. \tag{29}$$

第一个不等式由柯西-施瓦茨不等式得出。第二个不等式利用了这一事实:

$$\begin{aligned}
&\sum_{\tau\in[t-1]} \|(\tilde{\Sigma}_{t-1}^{\text{res}})^{-\frac{1}{2}}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\|_2^2 \\
=&\sum_{\tau\in[t-1]} \varphi(X_\tau,a_\tau)^\top\hat{U}^{\text{res}}(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau) \\
=&\,\text{trace}\left(\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\varphi(X_\tau,a_\tau)^\top\hat{U}^{\text{res}}(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}\right) \\
\leq&\,\text{trace}\left(\left(\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\varphi(X_\tau,a_\tau)^\top\hat{U}^{\text{res}}+\lambda I_{p^{\text{res}}}\right)(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}\right) \\
=&\,\text{trace}(I_{p^{\text{res}}}) \\
=&p^{\text{res}}.
\end{aligned}$$

在假设 3下，对于所有 $\eta\in(0,\ 1)$，以至少 $1-\eta$ 的概率，(29) 为 $O\left(\sqrt{p^{\text{res}}t\log(\frac{p/\eta}{T_0})}\right)$。为了界定 (27) 中的项，我们使用引理 2 中关于不变分量估计误差的结果，因为

$$\begin{aligned}
&\|\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\varphi(X_\tau,a_\tau)^\top(\beta^{\text{inv}}-\hat{\beta}^{\text{inv}})\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} \\
=&\|(\tilde{\Sigma}_{t-1}^{\text{res}})^{-\frac{1}{2}}\sum_{\tau\in[t-1]}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\varphi(X_\tau,a_\tau)^\top(\beta^{\text{inv}}-\hat{\beta}^{\text{inv}})\|_2 \\
\leq&\sqrt{\sum_{\tau\in[t-1]}\|(\tilde{\Sigma}_{t-1}^{\text{res}})^{-\frac{1}{2}}(\hat{U}^{\text{res}})^\top\varphi(X_\tau,a_\tau)\|_2^2}\sqrt{\sum_{\tau\in[t-1]}(\varphi(X_\tau,a_\tau)^\top(\beta^{\text{inv}}-\hat{\beta}^{\text{inv}}))^2} \\
\leq&\sqrt{p^{\text{res}}}\sqrt{t\max_{\tau\in[t-1]}\|\varphi(X_\tau,a_\tau)\|_2^2\|\beta^{\text{inv}}-\hat{\beta}^{\text{inv}}\|_2^2} \\
\leq&L\|\beta^{\text{inv}}-\hat{\beta}^{\text{inv}}\|_2\sqrt{p^{\text{res}}t} =: \hat{\rho}_3^{\text{res}}.
\end{aligned}$$

应用并集界可得，以至少 $1-2\eta$ 的概率，

$$\|\hat{\delta}_t^{\text{res}}-\hat{\Pi}^{\mathcal{S}^{\text{res}}}\delta_t^{\text{res}}\|_{\hat{\Sigma}_{t-1}}\leq\sum_{i=1}^{3}\hat{\rho}_i^{\text{res}} =: \hat{\rho}_t^{\text{res}}(\eta,L,M). \tag{30}$$

要完成该引理的证明，只需在上述所有项中将 η 替换为 $\eta':=\eta/2$

□

定理 3 的证明。 我们首先考虑一个与式（8）中的分解类似的瞬时遗憾分解。现在，真实子空间分解未知，因此 $\hat{}\mathcal{C}^\beta$ 和 $\hat{}\mathcal{C}_t^\delta$ 是置信集，以至少 $1-\eta$ 的概率分别包含 $\hat{}\Pi^{\mathcal{S}^{\text{inv}}}\beta^{\text{inv}}$ 和 $\hat{}\Pi^{\mathcal{S}^{\text{res}}}\delta_t^{\text{res}}$。与之前一样，我们用 $(\bar{\beta}_t,\bar{\delta}_t)=\arg\max_{\beta\in\hat{}\mathcal{C}^\beta,\delta\in\hat{}\mathcal{C}_t^\delta}\varphi(X_t,a_t)^\top(\beta+\delta)$ 表示，并用 $\bar{\gamma}_t=\bar{\beta}_t+\bar{\delta}_t$ 表示在时间 t 使预测奖励的上置信界达到最大值的参数。此外，

回忆一下，$\tilde{\Sigma}_t^{\text{res}} = (\hat{U}_{\text{残差}})^\top \hat{\Sigma}_t \hat{U}_{\text{残差}}$ 且 $\tilde{\Sigma}_{[-T_0]}^{\text{inv}} = (\hat{U}^{\text{inv}})^\top \hat{\Sigma}_{[-T_0]} \hat{U}^{\text{inv}}$。那么，在时间 $t \in [T]$，reg_t 满足

$$
\begin{aligned}
\text{reg}_t &:= (\varphi(X_t, a_t^*) - \varphi(X_t, a_t))^\top \gamma_{0,t} \\
&\le \varphi(X_t, a_t)^\top (\bar{\gamma}_t - \gamma_{0,t}) \\
&= \varphi(X_t, a_t)^\top (\bar{\beta}_t - \beta^{\text{inv}}) + \varphi(X_t, a_t)^\top (\bar{\delta}_t - \delta_t^{\text{res}}) \\
&= \varphi(X_t, a_t)^\top (\bar{\beta}_t - \hat{\Pi}^{\mathcal{S}^{\text{inv}}} \beta^{\text{inv}} + \hat{\Pi}^{\mathcal{S}^{\text{inv}}} \beta^{\text{inv}} - \beta^{\text{inv}}) \\
&+ \varphi(X_t, a_t)^\top (\bar{\delta}_t - \hat{\Pi}^{\mathcal{S}^{\text{res}}} \delta_t^{\text{res}} + \hat{\Pi}^{\mathcal{S}^{\text{res}}} \delta_t^{\text{res}} - \delta_t^{\text{res}}) \\
&\le \|(\hat{U}^{\text{inv}})^\top \varphi(X_t, a_t)\|_2 \|(\hat{U}^{\text{inv}})^\top \bar{\beta}_t - (\hat{U}^{\text{inv}})^\top \beta^{\text{inv}}\|_2 \\
&+ \|(\hat{U}^{\text{res}})^\top \varphi(X_t, a_t)\|_{((\hat{U}^{\text{res}})^\top \hat{\Sigma}_{t-1} \hat{U}^{\text{res}})^{-1}} \|(\hat{U}^{\text{res}})^\top \bar{\delta}_t - (\hat{U}^{\text{res}})^\top \delta_t^{\text{res}}\|_{(\hat{U}^{\text{res}})^\top \hat{\Sigma}_{t-1} \hat{U}^{\text{res}}} \\
&+ \|\varphi(X_t, a_t)\|_2 \|\gamma_{0,t}\|_2 \Delta\Pi \\
&\le \|(\hat{U}^{\text{inv}})^\top \varphi(X_t, a_t)\|_2 \frac{1}{\sqrt{\lambda_0 T_0}} \|(\hat{U}^{\text{inv}})^\top \bar{\beta}_t - (\hat{U}^{\text{inv}})^\top \beta^{\text{inv}}\|_{\tilde{\Sigma}_{[-T_0]}^{\text{inv}}} \\
&+ \|(\hat{U}^{\text{res}})^\top \varphi(X_t, a_t)\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} \|(\hat{U}^{\text{res}})^\top \bar{\delta}_t - (\hat{U}^{\text{res}})^\top \delta_t^{\text{res}}\|_{\tilde{\Sigma}_{t-1}^{\text{res}}} \\
&+ \|\varphi(X_t, a_t)\|_2 \|\gamma_{0,t}\|_2 \Delta\Pi \\
&\le 2\|(\hat{U}^{\text{inv}})^\top \varphi(X_t, a_t)\|_2 \sqrt{\frac{\hat{\rho}_{T_0}^{\text{inv}}(\eta, L, M)}{\lambda_0 T_0}} + 2\|(\hat{U}^{\text{res}})^\top \varphi(X_t, a_t)\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} \sqrt{\hat{\rho}_t^{\text{res}}(\eta, L, M)} \\
&+ \|\varphi(X_t, a_t)\|_2 \|\gamma_{0,t}\|_2 \Delta\Pi.
\end{aligned}
$$

引理 2 意味着我们可以定义 $\hat{\mathcal{C}}^\beta$ 使其包含 $\hat{\Pi}^{\mathcal{S}^{\text{inv}}} \beta^{\text{inv}}$，其概率至少为 $1-\eta$，从而 $\sqrt{\hat{\rho}_{T_0}^{\text{inv}}(\eta, L, M)}$ 是

$$
O\left(\sqrt{p^{\text{inv}} \log\left(\frac{1}{p^{\text{inv}} \lambda_0}\right) + \log\left(\frac{1}{\eta}\right)}\right) + O\left(\sqrt{p^{\text{inv}} \log\left(\frac{p}{\eta}\right)}\right) + O\left(\sqrt{\frac{1}{\lambda_0} \log\left(\frac{p}{\eta}\right)}\right).
$$

引理 3 则意味着我们可以定义 $\hat{\mathcal{C}}_t^\delta$，使其包含 $\hat{\Pi}^{\mathcal{S}^{\text{res}}} \delta_t^{\text{res}}$，从而 $\sqrt{\hat{\rho}_t^{\text{res}}(\eta, L, M)}$ 对所有 $\eta \in (0, 1/2)$ 均成立，且概率至少为 $1-2\eta$

$$
O\left(\sqrt{p^{\text{res}} \log\left(\frac{t}{p^{\text{res}}}\right) + \log\left(\frac{1}{\eta}\right)}\right) + O\left(\sqrt{\frac{p^{\text{res}} t}{T_0}} \left(\sqrt{\log\left(\frac{p}{\eta}\right)} + \sqrt{\frac{\hat{\rho}_{T_0}^{\text{inv}}(\eta, L, M)}{\lambda_0}}\right)\right). \tag{31}
$$

此外，根据假设 3，我们有，对于所有 $\eta \in (0, 1)$，其概率至少为 $1-\eta$，

$$
\|\varphi(X_t, a_t)\|_2 \Delta\Pi \|\gamma_{0,t}\|_2 \le LM\Delta\Pi
$$

为 $O(\sqrt{\log(p/\eta)/T_0})$。现在对时间范围 T 求和，可得

$$
\begin{aligned}
\text{Reg}_T &= \sum_{t=1}^T \text{reg}_t \\
&\le \sum_{t=1}^T \left(2\|(\hat{U}^{\text{inv}})^\top \varphi(X_t, a_t)\|_2 \sqrt{\frac{\hat{\rho}_{T_0}^{\text{inv}}(\eta, L, M)}{\lambda_0 T_0}} + \|\varphi(X_t, a_t)\|_2 \|\gamma_{0,t}\|_2 \Delta\Pi\right) \\
&+ \sum_{t=1}^T 2\|(\hat{U}^{\text{res}})^\top \varphi(X_t, a_t)\|_{(\tilde{\Sigma}_{t-1}^{\text{res}})^{-1}} \sqrt{\rho_t^{\text{res}}}
\end{aligned}
$$

我们分别考察这两项求和。对于第一项，可得

$$\begin{aligned}&\sum_{t=1}^{T}\left(2\|(\hat{U}^{\mathrm{inv}})^{\top}\varphi(X_t,a_t)\|_2\sqrt{\frac{\hat{\rho}^{\mathrm{inv}}_{T_0}(\eta,L,M)}{\lambda_0 T_0}}+\|\varphi(X_t,a_t)\|_2\|\gamma_{0,t}\|_2\Delta\Pi\right)\\ \leq& T\left(2L\sqrt{\frac{\hat{\rho}^{\mathrm{inv}}_{T_0}(\eta,L,M)}{\lambda_0 T_0}}+LM\Delta\Pi\right).\end{aligned}$$

根据上述结果，对于所有 $\eta\in(0,1\,5)$，该项以至少$1-5\eta$的概率满足，

$$O\left(\frac{T}{\sqrt{T_0\lambda_0}}\sqrt{p^{\mathrm{inv}}+\log(\frac{1}{\eta})}\right)+O\left(\frac{T}{\sqrt{T_0\lambda_0}}\sqrt{p^{\mathrm{inv}}\log(\frac{p}{\eta})}\right)+O\left(\frac{T}{\sqrt{T_0\lambda_0}}\sqrt{\max\{\lambda_0,\frac{1}{\lambda_0}\}\log(\frac{p}{\eta})}\right).$$

对于累积遗憾的其余部分，我们有

$$\sum_{t=1}^{T}2\|(\hat{U}^{\mathrm{res}})^{\top}\varphi(X_t,a_t)\|_{(\tilde{\Sigma}^{\mathrm{res}}_{t-1})^{-1}}\sqrt{\hat{\rho}^{\mathrm{res}}_t(\eta,L,M)}\leq 2\sqrt{T\hat{\rho}^{\mathrm{res}}_T(\eta,L,M)\sum_{t=1}^{T}\|(\hat{U}^{\mathrm{res}})^{\top}\varphi(X_t,a_t)\|^2_{(\tilde{\Sigma}^{\mathrm{res}}_{t-1})^{-1}}}$$

与预言机情形相同（参见引理 3的证明），我们可以利用（14）和引理 4 得到

$$\sum_{t=1}^{T}2\|(\hat{U}^{\mathrm{res}})^{\top}\varphi(X_t,a_t)\|_{(\tilde{\Sigma}^{\mathrm{res}}_{t-1})^{-1}}\sqrt{\hat{\rho}^{\mathrm{res}}_t(\eta,L,M)}\leq C^{\mathrm{res}}\sqrt{\hat{\rho}^{\mathrm{res}}_T(\eta,L,M)Tp^{\mathrm{res}}\log\left(1+\frac{TL^2}{\lambda p^{\mathrm{res}}}\right)}.$$

利用（31），可知对于所有 $\eta\in(0,1\,2)$，上述项以至少 $1-2$ 的概率满足 η，

$$\begin{aligned}&O\left(p^{\mathrm{res}}\sqrt{T\log\left(1+\frac{TL^2}{\lambda p^{\mathrm{res}}}\right)\left(\log(\frac{T}{p^{\mathrm{res}}})+\log(\frac{1}{\eta})\right)}\right)\\ +&O\left(\frac{p^{\mathrm{res}}T}{\sqrt{T_0}}\sqrt{\log\left(1+\frac{TL^2}{\lambda p^{\mathrm{res}}}\right)}\left(\sqrt{\log(\frac{p}{\eta})}+\sqrt{\frac{\hat{\rho}^{\mathrm{inv}}_{T_0}(\eta,L,M)}{\lambda_0}}\right)\right).\end{aligned}$$

因此，忽略对数项后，对于所有 $\eta\in(0,1\,11)$，ISD-linUCB 算法的累积遗憾以至少 $1-11\eta$ 的概率满足

$$\tilde{O}\left(\sqrt{T}\left(p^{\mathrm{res}}+p^{\mathrm{res}}\sqrt{\frac{T}{\lambda_0 T_0}}\left(\sqrt{p^{\mathrm{inv}}}+\sqrt{\max\{\lambda_0,\frac{1}{\lambda_0}\}}\right)\right)\right).$$

□

D 下界

D.1 LinUCB 的下界

Lattimore 和 Szepesvári (2020年) 给出了 LinUCB 算法的下界 $\Omega(p\sqrt{T})$ 。在线性模型中，他们考虑的是单位方差的高斯噪声。于是，对于上下文-动作空间 $\mathcal{X}\times\mathcal{A}$ 和线性参数 $\gamma\in\mathbb{R}^p$，策略的遗憾为

$$\mathrm{Reg}_T(\mathcal{X}\times\mathcal{A},\gamma)=T\max_{\varphi\in\mathcal{X}\times\mathcal{A}}\varphi^{\top}\gamma-\mathbb{E}_{\gamma}\left[\sum_{t=1}^{T}R_t^{a_t}\right]$$

其中，期望是相对于由策略与参数为 γ的老虎机交互所诱导的奖励测度取的。其思路是寻找老虎机问题的一个最坏情况实例，该实例由上下文-动作空间和参数 γ刻画，且累积遗憾至少为 $p\sqrt{T}$ 的量级。Lattimore 和 Szepesvári（2020年）针对上下文-动作空间的两种不同选择证明了这一结果。第一种是超立方体，其中 $\mathcal{X}\times\mathcal{A}=[-1,1]^p$。第二种选择是单位球面，即 $\mathcal{X}\times\mathcal{A}=\{x\in\mathbb{R}^p \mid \|x\|_2\le 1\}$。在第一种情况下，证明较为简单，因为一个方向上的上下文-动作特征值不受其余方向上的值约束：我们暂且着重讨论这一证明。具体而言，针对 $\mathcal{X}\times\mathcal{A}=[-1,1]^p$ 的证明表明，存在一个参数 $\gamma\in\Gamma\coloneqq\left\{\pm\frac{1}{\sqrt{T}}\right\}^p$，使得 $\mathrm{Reg}_T(\mathcal{X}\times\mathcal{A},\gamma)$ 为 $\Omega(p\sqrt{T})$。证明首先考虑概率测度 $\mathbb{P}_\gamma,\mathbb{P}_{\gamma'}$ 之间的相对熵（或 KL散度）。这些测度是由固定策略（即固定的上下文-动作特征序列 $\{\varphi(X_t,a_t)\}_{t=1}^T$）分别与参数为 $\gamma,\gamma'\in\Gamma$ 的两个老虎机交互所诱导的奖励分布（噪声为 $\epsilon_t\sim\mathcal{N}(0,1)$），也就是说，

$$D(\mathbb{P}_\gamma,\mathbb{P}_{\gamma'})=\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_\gamma[(\varphi(X_t,a_t)^\top(\gamma-\gamma'))^2].$$

用下标 i 表示向量的第 i 个分量，随后他们为 $i\in[p]$ 和$\gamma\in\Gamma$ 定义

$$q_{\gamma_i}\coloneqq\mathbb{P}_\gamma\left(\sum_{t=1}^{T}\mathbb{1}\{\mathrm{sign}(\varphi_i(X_t,a_t)\neq\mathrm{sign}(\gamma_i)\}\ge\frac{T}{2}\right),$$

也就是说，所选上下文-动作特征的第 i 个分量的符号非最优的次数超过总学习时域一半的概率。对于固定的 $i\in[p]$ 和 $\gamma\in\Gamma$，他们进一步将 γ' 固定为一个参数：除第 i 个分量设为 $\gamma_i'=-\gamma_i$ 外，其余分量均为 γ。根据相对熵的一个性质（Bretangolle-Huber 不等式），可得

$$q_{\gamma_i}+q_{\gamma_i'}\ge\frac{1}{2}\exp\left(-D(\mathbb{P}_\gamma,\mathbb{P}_{\gamma'})\right)=\frac{1}{2}\exp\left(-\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_\gamma[(\varphi(X_t,a_t)^\top(\gamma-\gamma'))^2]\right)\ge\frac{1}{2}\exp(-2),$$

其中，最后一个不等式由空间 $\mathcal{A}\times\mathcal{X}$和 Γ 的定义得出。对所有可能的参数 $\gamma\in\Gamma$ 取平均，可得

$$\frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\sum_{i\in[p]}q_{\gamma_i}=\frac{1}{|\Gamma|}\sum_{i\in[p]}\sum_{\gamma\in\Gamma}q_{\gamma_i}\ge\frac{1}{|\Gamma|}\sum_{i\in[p]}\frac{|\Gamma|}{2}\frac{1}{2}\exp\left(-2\right)=\frac{p}{4}\exp(-2).$$

最后，这意味着至少存在一个参数 $\gamma\in\Gamma$，使得 $\sum_{i\in[p]}q_{\gamma_i}\ge\frac{p}{4}\exp(-2)$。对于这个 γ的取值，遗憾为

$$\begin{aligned}\mathrm{Reg}_T(\mathcal{X}\times\mathcal{A},\gamma)&=\mathbb{E}_\gamma\left[\sum_{t=1}^{T}\sum_{i=1}^{p}(\mathrm{sign}(\gamma_i)-\varphi_i(X_t,a_t))\gamma_i\right]\\&\ge\frac{1}{\sqrt{T}}\sum_{i=1}^{p}\mathbb{E}_\gamma\left[\sum_{t=1}^{T}\mathbb{1}\{\mathrm{sign}(\varphi_i(X_t,a_t))\neq\mathrm{sign}(\gamma_i)\}\right]\\&\ge\frac{\sqrt{T}}{2}\sum_{i=1}^{p}\mathbb{P}_\gamma\left(\sum_{t=1}^{T}\mathbb{1}\{\mathrm{sign}(\varphi_i(X_t,a_t))\neq\mathrm{sign}(\gamma_i)\}\ge\frac{T}{2}\right)\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sqrt{T}}{2} \sum_{i=1}^{p} q_{\gamma_i} \\
&\geq \frac{\exp -2}{8} p\sqrt{T}.
\end{aligned}
$$

第一个等式由 $\max_{\varphi \in [-1,1]^p} \varphi^\top \theta = \text{sign}(\theta)^\top \theta$ 这一事实得出。第一个不等式根据空间 Γ 的定义，并结合对求和项取值的分类讨论得出。第二个不等式使用了马尔可夫不等式。最后一个等式由 γ 的取值方式得出。至此，下界的证明完成。

D.2 具有预言机子空间和预言机不变分量的 ISD 下界

现在考虑我们可以预言机访问不变子空间分解（$\mathcal{S}^{\text{inv}}$，$\mathcal{S}^{\text{res}}$）和不变参数 β^{inv} 的情形。我们可以修改上述标准线性情形的证明，考虑一个参数空间，其中的参数仅在其残差分量上有所不同，也就是说，它们在 $\mathcal{S}^{\text{res}}$上的投影不同，而不变分量保持不变。在这一附加约束下，Γ 的有效维数从 p 降至 p^{res}，从而得到下界 $\Omega(p^{\text{res}}\sqrt{T})$。

更具体地说，我们再次考虑上下文-动作空间 $\mathcal{X} \times \mathcal{A} = [-1,1]^p$。现在将参数空间定义为 $\Gamma := \left\{ \gamma \in \mathbb{R}^p \mid \gamma = \beta^{\text{inv}} + \delta^{\text{res}},\ (U^{\text{res}})^\top \delta^{\text{res}} \in \left\{ \pm \frac{1}{\sqrt{T}} \right\}^{p^{\text{res}}} \right\}$。于是，对于两个参数 γ，$\gamma' \in \Gamma$，

$$
D(\mathbb{P}_\gamma, \mathbb{P}_{\gamma'}) = \frac{1}{2} \sum_{t=1}^{T} \mathbb{E}_\gamma [(\varphi(X_t, a_t)^\top (\delta^{\text{res}} - \delta'^{\text{res}}))^2].
$$

对于所有 $t \in [T]$，令 $\tilde{\varphi}^{\text{res}}(X_t,\ a_t) := (U^{\text{res}})^\top \varphi(X_t,\ a_t)$ 和 $\tilde{\delta}^{\text{res}} := (U^{\text{res}})^\top \delta^{\text{res}}$。与上文一样，我们可以对 $i \in [p^{\text{res}}]$ 和 $\gamma \in \Gamma$ 定义

$$
q_{\gamma_i} := \mathbb{P}_\gamma \left(\sum_{t=1}^{T} \mathbb{1}\{\text{sign}(\tilde{\varphi}_i^{\text{res}}(X_t, a_t)) \neq \text{sign}(\tilde{\delta}_i^{\text{res}})\} \geq \frac{T}{2} \right).
$$

对于固定的 $i \in [p^{\text{res}}]$ 和 $\gamma \in \Gamma$，将 $\gamma' \in \Gamma$ 定义为通过设置 $\delta_i'^{\text{res}} = -\delta_i^{\text{res}}$ 从 γ 得到的参数（其他部分则与 γ 相同）。于是，

$$
q_{\gamma_i} + q_{\gamma_i'} \geq \frac{1}{2} \exp\left(-\frac{1}{2} \sum_{t=1}^{T} \mathbb{E}_\gamma [(\varphi(X_t, a_t)^\top (\delta^{\text{res}} - \delta'^{\text{res}}))^2] \right) \geq \frac{1}{2} \exp(-2)
$$

这同样可由空间 $\mathcal{X} \times \mathcal{A}$ 和 Γ 的定义推出。对所有可能的参数 $\gamma \in \Gamma$ 取平均后，我们得到

$$
\frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \sum_{i \in [p^{\text{res}}]} q_{\gamma_i} \geq \frac{1}{|\Gamma|} \sum_{i \in [p^{\text{res}}]} \frac{|\Gamma|}{2} \frac{1}{2} \exp(-2) = p^{\text{res}} \frac{\exp(-2)}{4},
$$

这意味着存在 $\gamma \in \Gamma$ 使得 $\sum_{i \in [p^{\text{res}}]} q_{\gamma_i} \geq p^{\text{res}} \exp(-2)/4$。证明的最后一部分与标准情形完全相同，由此得到的遗憾为 $\Omega(p^{\text{res}}\sqrt{T})$。