arXiv:2601.09042v1 [cs.LG] 14 Jan 2026

SCaLE：感知切换成本的学习与探索 n

Neelkamal Bhuyan*,1、Debankur Mukherjee[1],和 Adam Wierman[2]

[1]佐治亚理工学院[2]加州理工学院

关键词 — 在线算法、移动成本、在线学习、矩阵估计、矩阵扰动理论

摘要

本文针对老虎机在线凸优化中的无界度量移动成本这一基本问题展开研究，考察高维动态二次命中成本以及带噪老虎机反馈模型中的 ℓ_2-范数切换成本。对于一类一般的随机环境，我们提出首个算法 SCaLE；即使不了解到达成本结构，该算法也能保证实现分布无关的次线性动态遗憾。此外，我们提出一种新颖的谱遗憾分析方法，分别量化由特征值误差驱动的遗憾和由特征基扰动驱动的遗憾。与在线学习基线方法进行的大量数值实验验证了我们的结论，并凸显了该算法的统计一致性。

1 引言

大规模基础设施的高效管理，越来越需要能够随时学习系统动力学、同时维持最优控制的算法。在许多实际场景中，预先计算的物理模型会因条件随时间变化或复杂交互而失效，因此需要持续进行在线估计。例如，在数据中心热管理和协作式风电场控制中，运营人员必须估计未知的耦合矩阵，这些矩阵代表气流湍流或尾流空气动力学，同时不能破坏当前系统的稳定性。这一挑战也延伸至智能电网拓扑识别，其中阻抗参数会随负载变化 [1, 2] 机器人学领域的运动学模型也以难以通过解析方法定义而闻名 [3], 而在过程控制中，则必须实时、安全地推断系统动力学 [4]。第6节将详细介绍其中几个领域，在这些领域中，学习智能体面临一项基本权衡：一方面需要探测系统以降低模型不确定性，另一方面则要尽量减少快速执行器波动所带来的切换成本。

为应对这一权衡，系统辨识与控制之间的相互作用已得到广泛研究，历史上主要是在自适应控制框架下展开的。然而，现有文献的一项重要局限在于依赖严格的分布假设。大多数现有研究，尤其是基于线性二次高斯（LQG）控制的研究，都假设外部驱动力（或噪声）服从次高斯分布 [5, 6]。尽管这一假设便于通过黎卡提方程求得闭式解，但在实际部署中，扰动可能呈重尾特征、具有结构性或带有对抗性，因此该假设往往不成立。因此，在高斯先验下推导出的算法，在面对现实中变幻莫测、

*电子邮件：nbhuyan3@gatech.edu

非平稳的需求或环境负载统计特性时，往往缺乏稳健性。这种脱节凸显了现实世界对分布无关方法的迫切需求：此类方法无需依赖特定的噪声生成模型，也能确保性能稳定性。

为弥合这一差距，近期的理论进展已将自适应线性二次调节控制与平滑在线凸优化（SOCO）这两个领域联系起来 [7, 8]。具体而言， [8] 研究表明，对于二次代价，只要系统参数已知，即可推导出一种分布无关的在线最优策略，其结构与经典线性二次调节相似，但对任意鞅噪声序列均有效。在此基础上，我们旨在将随机系统辨识与控制扩展到次高斯假设之外。我们研究的问题是：控制系统所依据的成本矩阵未知，必须通过老虎机反馈来学习，从而将鲁棒平滑在线凸优化框架推广到自适应设定。具体而言，我们希望回答以下问题：

> 如何同时做到：*(i)* 从带噪老虎机反馈中估计未知成本函数；*(ii)* 利用该估计进行在线决策；*(iii)* 在改变动作时考虑切换成本惩罚；以及 *(iv)* 确保动态遗憾最小？

为回答这一问题，我们将其与老虎机凸优化（BCO）领域的研究联系起来；该领域旨在以最少的反馈应对动态成本 [9, 10, 11], ，并与逆最优控制（IOC）的最新进展相联系；后者利用数据驱动技术推断潜在的成本结构 [12, 13]。第 7 节进一步阐述了本文与在线算法、学习理论和最优控制领域研究之间的联系。

关注切换成本，是因为大规模序贯决策需要移动高效算法，包括但不限于智能电网成本管理[15, 16], 自适应控制 [5, 6, 17, 18, 19], 数据中心管理 [20, 21], 电动汽车充电 [22], 视频传输 [23] 和电力系统 [15, 24]。许多此类领域的一个核心特征是，改变决策会产生显著成本；这通常源于决策参数的精细化特性，使决策空间连续且无界 [20, 23, 25, 26]。

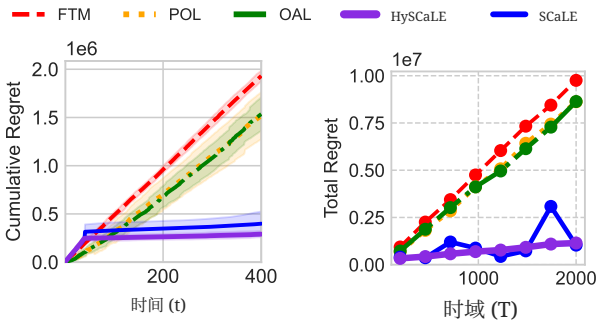


图 1： SCaLE &HySCaLE 性能（本研究）； FTM: 平滑在线优化基准 [14]； POL/OAL: 含噪/理想在线学习基准

然而，学习领域中的控制算法通常是在一些假设下开发的
这些假设使得对移动成本的显式管理沦为次要问题 d。 Th其理论h保证 l
通常针对满足以下任一条件的环境提出： (i) 使用静态基准 [27, 28](ii) 具有有界的动作空间直径 [9, 29]; (iii) 由低漂移的次高斯随机过程驱动 [5, 6]; 或 (iv) 满足某些可控性/稳定性条件 [5, 6, 18]。违反其中任何一项假设都可能导致移动成本失控；图1展示了基于学习的技术这一缺陷，第4.2节的案例研究也对此进行了记录。

挑战。本文聚焦于二次命中成本函数这一类函数，其研究动机源自大规模应用所采用的成本模型 [7, 8, 15, 20, 30]。在每一轮 t，玩家选择一个动作 $x_t \in \mathbb{R}^d$并产生二次命中成本 $f_t(x) = \frac{1}{2}(x - v_t)^T A(x - v_t)$，以及一个 ℓ_2-范数切换成本 $c(x_t, x_{t-1}) = \frac{1}{2}\|x_t - x_{t-1}\|_2^2$。未知矩阵 $A \in \mathbb{R}^{d\times d}$是半正定的，而命中成本的极小化点 $\{v_t\}_{t=1}^T$，我们称之为位置

参数以在线方式揭示，并在有限时间范围内演化为一个随机过程 $\mathbb{R}^d$ 。该曲率矩阵的估计值 A 必须通过每轮仅观察一次已产生命中成本的带噪声观测值 $f_t(x_t) + \eta_t$，以在线方式逐轮改进。面对这种带噪声的零阶反馈，智能体面临两项技术挑战：

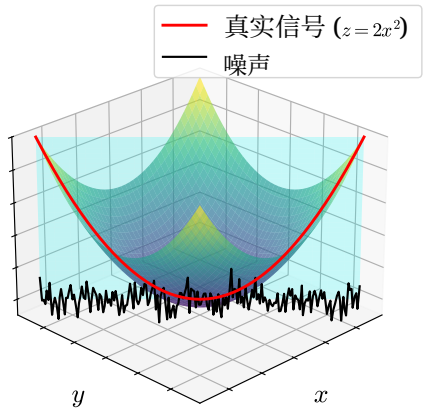
（i）基本权衡。为了有效地学习 $d \times d$ 矩阵 A，学习智能体必须生成一系列具有足够多样性的查询点 $\{x_t\}_t$[31, 32, 33, 34, 35]。然而，这种探索需求与最小化切换成本 $\|x_t - x_{t-1}\|_2^2$ 的目标直接冲突，因为切换成本会对大幅移动施加重罚。这形成了一种在标准多臂老虎机问题或在线系统辨识中从未见过的新型张力：收集信息所需的动作，恰恰受到目标函数核心组成部分的明确抑制。有趣的是，这一现象体现为一种基本的性能极限（定理2），并通过附录D中的详细示例加以说明。

(ii) 高维退化。当 $f_t(x)$存在零曲率区域时（见图 2），这种矛盾在高维情形下会加剧。当 A 秩亏或奇异时，会出现两个关键问题：(i)老虎机反馈无法为学习提供有用信号，实际上退化为纯噪声；(ii)切换成本超过命中成本。因此，智能体可能付出高昂的切换成本，却只得到无用的信号。

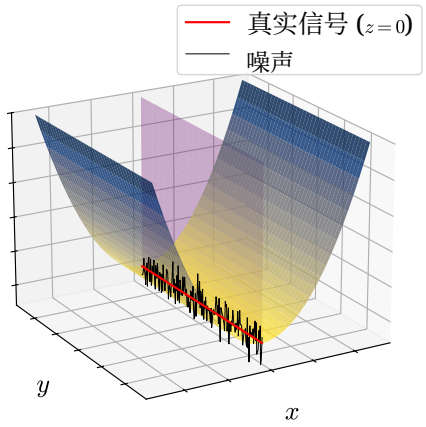
上述问题是度量移动成本文献中常见的障碍[26, 36, 37, 38, 39, 40]。以往研究已在对抗性设定下确立了一个强有力的 $\Omega(\sqrt{d})$ 竞争比下界，尽管这是通过病态的非光滑构造得出的 [39]。然而，实际场景并不会出现如此极端的情况。在调度等相关领域中，研究者已使用随机建模来捕捉已知的数据模式，并据此设计出能在这些已知模型下表现良好的算法 [41, 42, 43, 44, 45, 46]。

这促使我们将重点放在随机设定上，近期的进展已展现出一定的前景。值得注意的是， [26] 利用最小化器轨迹的统计特性，绕过了这一秩亏障碍，尽管这需要完整掌握矩阵 A. 基于这一见解，我们采用随机框架，以应对这些退化高维环境中估计与移动所面临的综合挑战。

贡献。本文做出了四项贡献：(a)算法方面：我们设计了一种新颖的算法，*SCaLE* (算法 *1*)，它在无界动作空间中实现了与分布无关的次线性动态遗憾(定理 1)，且无需事先了解命中成本结构。



（a）在噪声下可以检测曲率信号



(b) 奇异方向上的纯噪声

图 2：秩亏阻碍曲率估计

(b)分析方面：SCaLE 的卓越性能得到了严格性能保证的支持：

1. 一种 $\mathcal{O}(T^{2/3})$ 适用于秩亏二次命中成本的动态遗憾，即当秩为$(A) < d$。这是有限反馈在线服务成本优化中首个克服曲率奇异性的高维保证（定理4）。

2. 改进且紧致的 $\Theta(T^{1/2})$ 适用于满秩（秩为$(A) = d$）情形的动态遗憾（定理5），与下界（定理 2）相匹配。

(c) 新颖的遗憾分解：这些分析结果得益于关键的方法论进展，这些进展本身也可能具有独立的研究价值。我们提出：

1. 一种新颖的在线遗憾分解（定理 3），能够将总遗憾精确拆分为以下几个部分：(i) 统计一致性，(ii) 环境探索，以及 (iii) 利用。

2. 对扰动进行细致分析（命题1），揭示了对退化在线二次优化的一种新颖且紧致的刻画（定理 4 和附录 5.2 ）。这一新技术分别量化了曲率矩阵特征基 A 中的估计误差与其特征值中的误差所导致的遗憾。

(d) 实证见解：第 4 节中的实验巩固了动态估计中统计一致性的必要性：第 4.2 节展示了基于梯度的技术在重尾随机环境中的失效，而我们提出的算法 SCaLE 仍能实现次线性动态遗憾。对于轻尾环境，我们提出一种混合形式 HySCaLE （算法 2），在 SCaLE 的利用阶段引入学习，从而（在实证上）提升性能

2 模型与预备知识

考虑一个动作空间为 $\mathbb{R}^d, d \geq 1$ 的在线决策问题，其有限时间范围为 T 轮。在每一轮中，玩家选择一个动作 x_t，并承受命中成本 $f_t(x_t) = \frac{1}{2}(x_t - v_t)^T A(x_t - v_t)$，其中 A 是一个半正定 $d \times d$ 矩阵，而 v_t 是一个以在线方式揭示的位置参数 。此外，每一轮中，玩家在动作之间切换时还需承担 $\frac{1}{2}\|x_t - x_{t-1}\|_2^2$ 的切换成本。我们分别用 $\|\cdot\|_*$ 和 $\|\cdot\|_F$ 表示核范数和弗罗贝尼乌斯范数，并用 $tr(\cdot)$表示矩阵迹算子。

定义 1。随机环境由极小化序列 $\{v_t\}_{t=1}^T$ 定义为任意鞅

$$\mathbb{E}[v_t|\mathcal{F}_{t-1}] = v_{t-1} \ and \ v_t - v_{t-1} \sim \mathrm{F}_t$$

其中 $\{\mathcal{F}_s\}_{s=0}^T$ 是由极小化序列生成的自然滤过，中心化分布 $\{\mathrm{F}_s\}_{s=1}^T$ 可以是动态的，也可以具有轻尾或重尾。

我们所考虑的随机设定推广了随机控制文献中的漂移模型，该模型考虑由加性白噪声驱动的动力学 [5, 6, 17, 18, 33]。

定义 2。对于任意 t，我们将老虎机反馈或秩一测量定义为：玩家在选择 x_t 之前只知道位置 v_t，之后会收到所产生的命中成本的带噪声数值，即 $f_t(x_t) + \eta_t$；其中 $\eta_t \in [-\bar{\eta}, \bar{\eta}]$ 可以是随机的，也可以是对抗性的。

在此信息模型下，玩家在每一轮次 t 计算一个在线动作 $x_t^{\texttt{ALG}}$，以求解以下优化问题：

$$\min_{(x_t)_{t=1}^T} \sum_{t=1}^{T} \frac{1}{2}(x_t - v_t)^T A(x_t - v_t) + \frac{1}{2}\|x_t - x_{t-1}\|_2^2 \tag{2.1}$$

我们选择的性能指标如下：

定义 3。 我们将任意在线算法 *ALG* 的动态遗憾 定义为

$$Regret_T^{\texttt{ALG}} := \mathbb{E}[Cost_{[1,T]}^{\texttt{ALG}}] - \mathbb{E}[Cost_{[1,T]}^*],$$

其中， $\mathbb{E}[$成本$^*[1,T]]$ 是在线最优动作序列 $\{x_t^*\}_{t=1}^T$. 的总成本

请注意，在线最优 $\{x_t^*\}_{t=1}^T$ 由 [8], 刻画，其通过随机动态规划求解了（2.1），但需要对 A. 由于篇幅限制，相关文献的详细概述将推迟至第 7 节。

3 算法与遗憾保证

本节提出我们的主要算法 SCaLE，并在满秩和秩亏设置下证明其具有次线性动态遗憾。

3.1 SCaLE 算法

SCaLE（算法 1）采用了一种经过精心平衡的先探索后利用策略：

1. 探索（第 1–5 行）：对于轮次 $\{1,\dots,m\}$，SCaLE 通过在揭示的极小点 v_t 附近注入经过精细缩放的独立同分布高斯噪声 z_t，主动探测成本函数。此处的目标是实现统计一致性（参见备注 1），其中 $m = c_1 \cdot rd$。尽管 c_1 是一个普适常数，我们提出了一种线搜索技术（4.1），以便于实际应用。

2. 估计与利用（第 6–14 行）：收集到足够的数据后，算法进入主要的估计与控制阶段。

(i) 统计估计（第 6 行）：通过将矩阵恢复问题表述为凸迹范数最小化，求解估计值 $\hat{}A^{\texttt{SCaLE}}$

$$\hat{A}^{\texttt{SCaLE}} = \operatorname*{argmin}_{\substack{M \succeq 0 \\ \|Y - \mathcal{A}(M)\|_1 \le \bar{\eta} m}} Tr(M) \tag{3.1}$$

(ii) 规划（第 7 行）：利用 $\hat{}A^{\texttt{SCaLE}}$，SCaLE 通过 RecurGen（子程序 1，借助动态规划可证明其为最优）计算整个（近似）最优矩阵序列 $\{\hat{}C_k\}_{k=1}^T$

（iii）**利用**（第 8—14 行）：在剩余的 $T-m$ 轮次中，SCaLE 使用（近似）最优的系数 $\{\hat{}C_k\}_{k=m+1}^T$，模拟在线最优 LAI。

SCaLE 结合了（3.1）——一种最初用于静态矩阵估计的稳健统计工具 [47, 48],与 LAI [8], 的分布无关的在线最优结构，并采用先探索后利用策略。由此得到如下动态遗憾，即使在重尾极小化序列下也成立$\{v_t\}_t$。

算法 1 SCaLE

输入： 噪声上限 $\bar{\eta}$、秩 r、下限 $\underline{\sigma_r^A}$、时域 T

Initialize: $m = c_1 rd$, $\hat{C}_{T+1} = I_{d\times d}$, $\gamma^2 = \begin{cases} \sqrt{\bar{\eta}}\max\left\{T^{2/3}, \frac{1}{\underline{\sigma_r^A}}\right\} & r < d \\ \sqrt{\bar{\eta}}\max\left\{T^{1/2}, \frac{1}{\underline{\sigma_r^A}}\right\} & r = d \end{cases}$

1: 对于 $t = 1, 2, \ldots, m$ **执行 2:** $z_t \sim \mathcal{N}$ **(0,** $I_{d\times d}$**) 3:** $x_t \leftarrow v_t + \gamma z_t$ **4:** $y_t \leftarrow f_t$**(**x_t**)** $+ \eta_t$ **5: 结束循环 6:** $\hat{A}^{\texttt{SCaLE}} \leftarrow$ **argmin** $Tr(M)$ $_{\|Y - \mathcal{A}\overset{M\succeq 0}{M}(\,)\|_1 \le \bar{\eta}m}$ **7:** $\{\hat{C}_k\}_{k=1}^T \leftarrow$ **RecurGen(** $\hat{A}^{\texttt{SCaLE}}$**, 1) 8: 对于** $\tau = 1, \ldots,$ m **执行 9:** $x^{\texttt{LAI}(\hat{A}^{\texttt{SCaLE}})}_{\tau} \leftarrow \hat{C}_\tau x^{\texttt{LAI}(\hat{A}^{\texttt{SCaLE}})}_{\tau-1} + (I - \hat{C}_\tau)v_\tau$ **10: 结束循环 11:** $x_{m+1} \leftarrow \hat{C}_{m+1} x^{\texttt{LAI}(\hat{A}^{\texttt{SCaLE}})}_{m} + (I - \hat{C}_m)v_m$ **12: 对于** $T = M + 2, \ldots, T$ **执行** $\hat{C}_t \leftarrow$ **13:** $x_t \leftarrow C_t x_{t-1} + (I - C_t)v_t$ **14: 结束循环**

子程序 1 RecurGen

输入：矩阵 A，轮次 t

初始化： $C_{T+1} = I$ **1: 对** $\tau = T, \ldots, t$ **执行 2:** $(\hat{C}_\tau)^{-1} = 2I + \hat{A} - \hat{C}_{\tau+1}$ **3: 结束循环 4: 返回** $\{\hat{C}_k\}_{k=t}^T$

定理 1. *SCaLE*$(\bar{\eta}, r, \sigma_r^A)$ 的动态遗憾，对于秩为 r 的 A，满足以下上界：

$$Regret_T^{\texttt{SCaLE}} \le \begin{cases} \alpha_1(\sqrt{\bar{\eta}}, \underline{\sigma_r^A}, \Sigma, d)(\boldsymbol{T - c_1 rd})^{\boldsymbol{2/3}} + \alpha_2(\sqrt{\bar{\eta}}, \underline{\sigma_r^A}, \Sigma, d) & r < d \\ \beta_1(\sqrt{\bar{\eta}}, \underline{\sigma_d^A}, \Sigma, d)(\boldsymbol{T - c_1 d^2})^{\boldsymbol{1/2}} + \beta_2(\sqrt{\bar{\eta}}, \underline{\sigma_d^A}, \Sigma, d) & r = d. \end{cases}$$

以高概率成立（$1 - \exp(-C_0 m)$），其中 C_0, c_1 是矩阵估计理论中的普适常数 *[47, 48]*, 且 Σ 是扰动过程协方差的上界 $\{v_t - v_{t-1}\}_t$，即 $\Sigma_{v_t - v_{t-1}} \preceq \Sigma$。各项 $\alpha_1, \alpha_2, \beta_1, \beta_2 > 0$，具体依赖关系见附录 *C.4*。

备注1. 下界 *[48,* 定理 *2.4]*关于“m”表明， $\Theta(r \cdot d)$ 秩-1 测量对于确保估计器 $\hat{}$的“受限一致有界性”（*RUB*）性质是必要的，而该估计器 A是$\hat{}$在噪声下实现稳定恢复所必需的。若缺少这一性质，任何估计器 A 都在统计上不一致，也就是说，

$$\inf_{\hat{A}} \sup_{\substack{A\in\mathbb{R}^{d\times d} \\ rank(A)=r}} \mathbb{E}\|\hat{A} - A\|_F^2 = \infty.$$

第 4.2 节阐述了这种不一致性对基于在线梯度下降的算法的动态遗憾所产生的下游影响。

特别是，除了协方差上界外，该结果不依赖于分布的其他特性$tr(\Sigma)$。这种尾部无关保证是上述结果区别于以往线性二次高斯控制尝试的关键特征 [5, 6]。

不同的遗憾保证源于谱误差的两个主要来源——特征值的扰动，以及估计曲率矩阵特征基的扰动。我们在第5.1 节说明，这两个误差分量随估计误差（ϵ）和时域（T）变化时呈现不同的缩放规律，正是这种差异导致了一种情形下的 $\mathcal{O}(T^{2/3})$ 速率，而另一种情形下的 $\mathcal{O}(T^{1/2})$ 速率。特别是，定理 4 建立了一种新颖的误差速率，其数量级比ϵ^2低一个数量级（ ϵ，而非 [8] ），这是标准遗憾分析所得的结果。

此外，满秩情形下的 $\mathcal{O}(T^{1/2})$ 速率证实了我们方法的最优性，因为它与该设定的信息论下界相匹配。

3.2 遗憾下界

为了确立该问题的基本极限，我们现在通过为一大类在线算法补充遗憾下界，来完善算法的上界分析——这类算法通过矩阵 $C_t \preceq I$ 在其先前动作与当前最小化解之间进行插值，从而做出决策：

$$x_t = C_t x_{t-1} + (I - C_t)v_t$$

这种结构具有一般性，涵盖了所有最先进的二次代价 SOCO 方法。以下定理形式化地刻画了任何此类算法在老虎机反馈约束下运行时必须付出的固有探索成本。

定理 2. 如果上述类别中的某个“插值”算法 *ALG* 执行秩为1 的测量，以将最优矩阵 $\{C_t^*\}_t$ 学习到误差不超过 ϵ 的程度（即 $\frac{\epsilon}{2} \leq \sigma^{C_t - C_t^*}_{\text{最小}} \leq \sigma^{C_t - C_t^*}_{\text{最大}} \leq \epsilon$），那么存在一个“$A$”，使得 *ALG* 遭受的遗憾至少为

$$Regret_T^{ALG} = \frac{\alpha_1}{\epsilon} + \alpha_2 \epsilon T + \alpha_3 = \Omega(\sqrt{T}),$$

其中 α_1、$\alpha_2 \in \mathbb{R}^+$ 和 $\alpha_3 \in \mathbb{R}$ 均与 T. 无关

该 α_1/ϵ 术语表示探索成本——精度越高（ ϵ越小），所需的前期成本就越高。相反， $\alpha_2 \epsilon T$ 术语表示利用期间的收益。附录C.5证明，这种估计—控制权衡是问题本身不可避免的一部分，并由此确立了在满秩情形下， $\mathcal{O}(T^{1/2})$ SCaLE 所达到的遗憾（定理 1）具有阶最优性。

3.3 HySCaLE：启发式自适应利用

SCaLE 严格遵循先探索后利用结构，在利用阶段有意舍弃收到的所有反馈，以保留 ϵ 关于 A 的估计保证。这引出了一个自然的问题：我们能否通过在利用阶段进一步利用反馈来提升实际性能？

为探究这一问题，我们提出了 HySCaLE（混合式 SCaLE），这是一种启发式算法，通过持续学习组件增强 SCaLE 的利用阶段。在初始探索和估计之后，HySCaLE 利用每个后续轮次的反馈来优化对 A 的估计。

算法 2 HySCaLE

输入： 噪声上限 $\bar{\eta}$，秩 r，下限 $\underline{\sigma_r^A}$，时域 T

Receive $(x_{m+1}^{\texttt{SCaLE}}, \hat{A}^{\texttt{SCaLE}})$ from $\texttt{SCaLE}(\bar{\eta}, r, \underline{\sigma_r^A}, T)$
$\hat{A}^{(m+1)} \leftarrow \hat{A}^{\texttt{SCaLE}}$
$x_{m+1} \leftarrow x_{m+1}^{\texttt{SCaLE}}$
$y_{m+1} \leftarrow f_{m+1}(x_{m+1}) + \eta_{m+1}$
for $t = m+2, \ldots, T$ **do**
  $M^{(t)} \leftarrow \hat{A}^{(t-1)} - \xi \cdot \nabla \ell_2\text{-loss}(y_{t-1}, \hat{A}^{(t-1)})$
  $\hat{A}^{(t)} \leftarrow \Pi_{M \succeq 0}\left(M^{(t)}\right)$
  $\{\hat{C}_k\}_{k=t}^{T} \leftarrow \text{RecurGen}(\hat{A}^{(t)}, t)$
  $x_t \leftarrow \hat{C}_t x_{t-1} + (I - \hat{C}_t) v_t$
  $y_t \leftarrow f_t(x_t) + \eta_t$
end for

在第 $t > m$ 轮，它使用上一步的含噪观测代价 y_{t-1}，对先前的估计 $\hat{A}^{(t-1)}$进行投影梯度下降 更新:

$$A^{(t)} = \Pi_{M \succeq 0}\Big\{ \hat{A}^{(t-1)} - \xi (x_{t-1} - v_{t-1})(x_{t-1} - v_{t-1})^T \times$$
$$\ldots \times \Big(\langle \hat{A}^{(t-1)}, (x_{t-1} - v_{t-1})(x_{t-1} - v_{t-1})^T \rangle - y_{t-1} \Big) \Big\}$$

然后使用 $A^{(t)}$ 重新计算 $\hat{C}_t(\hat{A}^{(t)})$，作为当前轮次的估计，具体如算法 2所示。正如我们接下来将要展示的，这种启发式方法在轻尾过程中能够提升性能，因为利用阶段的反馈是可靠的信号。

注释 2. 基于含噪梯度的更新可能使 *SCaLE* 对估计值 $\hat{A}^{SCaLE}$ *(*来自 (3.1)*)* 的调整可能导致 *HySCaLE* 出现统计不一致性。我们在图 *4*的重尾过程中观察到了这一现象。

4 数值实验

为验证我们的理论发现并评估所提出的算法，我们针对若干基线方法开展了一系列全面的数值模拟。

4.1 实验设置

我们的模拟 环境以及所考察的算法配置如下 s:

环境生成。 在线测试环境由一般的高维二次命中成本 $f_t($ $x_t - v_t$ c x_t x_{t-1} $=$ $\frac{1}{2}\|x_t - x_{t-1}\|^2$ $-$ $\{v_t\}_{t=1}^{T}$ A $) = \frac{1}{2}($ $)$ $\bar{T}$ $A()$、切换成本 $(,)$ 以及其极小化点的在线序列构成。(a) **成本矩阵：** 为实现精确控制，我们通过谱分解构造 A， $A = USU^T$。 U 是随机生成的正交归一矩阵， S 是由 $r = rank(A)$ 个奇异值构成的对角矩阵。

(b) **极小化序列** $\{v_t\}_t$**：** 在 $v_t = v_{t-1} + \delta_t$ 下，我们考虑底层$\{\delta_t\}_t$-过程的四种情况：(i) 非相关轻尾（$\delta_t \sim$独立同分布 $\mathcal{N}(\mathbf{0}, \sigma_v^2 I)$，(ii) 相关轻尾（$\delta_t =$$(1-\alpha)\delta_{t-1} + \alpha\mathcal{N}(\mathbf{0}, \sigma_v^2 I)$），(iii) 重尾（拉普拉斯分布），以及 (iv) 肥尾（柯西分布）。

(c) **馈送 back Noise：** 反馈 y_t 包含加性噪声 η_t，从 $[-\bar{\eta}, \bar{\eta}$ 中均匀抽取 $]$.

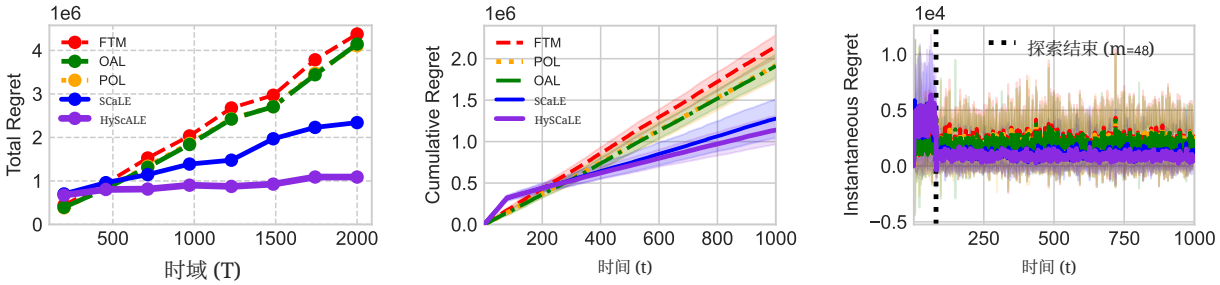


图 3：当 $r = d = 4$、$\sigma_r^A = 10^{-2}$、$\sigma_v = 50$、$\bar{\eta} = 10$ 和 c_1均设为 3 时的 Regret$_T$、Regret$_{1000}(t)$ 以及 Δ Regret$_{1000}$ (t) Δt

算法与基线。 我们将提出的两种算法的性能与理论最优值和三个具有代表性的基线方法进行比较。

- 动态在线最优（LAI）：理论基准 [8]，其完全知晓矩阵 A，以此作为衡量 Regret$_T$ 的基准。
- 跟随最小化器（FTM）：老虎机反馈下的自然基线，其中 $x_{\text{FTM}t} = v_t$.
- 被动在线学习器（POL）：旨在模拟在线梯度下降。它从一个 $\hat{}$ 朴素估计值 $A_0 = I$ 开始，并在每个轮次 t执行以下操作：(i) 通过对观测到的噪声反馈 y_t执行一步投影梯度下降，将其估计值从 $\hat{}A$ $(t-1)$ 更新为 $\hat{}A$ (t) 。

$$\hat{A}^{(t)} = \Pi_{M \succeq 0} \left\{ \hat{A}^{(t-1)} - \xi \cdot \nabla \ell_2\text{-loss}\left(y_t, \hat{A}^{(t-1)}\right) \right\}$$

  (ii) 重新计算 $\{\hat{}C_k\}_{k=t}^{T}$（使用 RecurGen($\hat{}A^{(t)}$, t)），并执行 $x_t = \hat{}C_t x_{t-1} + (I - \hat{}C_t)v_t$,
- 预言机辅助学习器（OAL）：POL 的强大变体，可访问**无噪声**反馈 y_t^*
- SCaLE 和HySCaLE：我们的算法 1 和 2

超参数与求解器设置。学习率 ξ 对于HySCaLE、POL 和OAL 均固定为 $\simeq 10^{-6}$。探索参数 c_1定义了探索长度 $m = c_1 rd$，其设为固定值（≈ 3），本节稍后将对此作进一步阐述。所有凸优化问题均使用 MOSEK 求解器求解 [49]（或者使用开源锥分裂求解器（SCS） [50]，通过 CVXPY 接口 [51]。

4.2 案例研究

我们在此通过案例研究展示主要的实证发现，所有结果均见附录A。

主要评估。 我们针对所有算法，生成多次运行后取平均的累积遗憾曲线和瞬时遗憾曲线。为了验证遗憾随时域变化的理论缩放关系（定理1），我们分析遗憾$_T$ 关于 $T \in [200, 2000]$ 的函数。

图 3展示了我们考虑 相关 $(\alpha = 0.7)$ 轻尾过程时的结果，表明 SCaLE 和HySCaLE 显著优于所有基线方法，尤其是在时域较长时。累积遗憾和

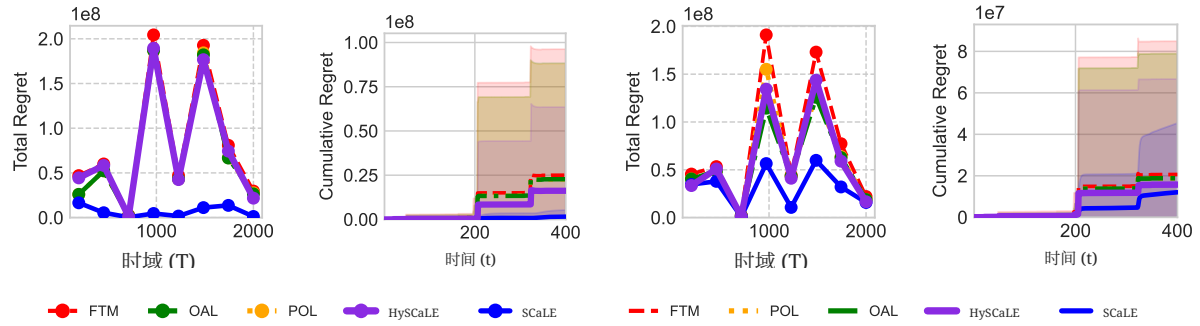


图4： $v_t - v_{t-1} \sim$ 柯西分布，带有 $\sigma_v = 1$。左图为 $\sigma_r^A = 10^{-2}$，右图为$\sigma_r^A = 1$。测试设置： $r = 1$、$d = 4$、 $\bar{\eta} = 50$、 $c = 10$。

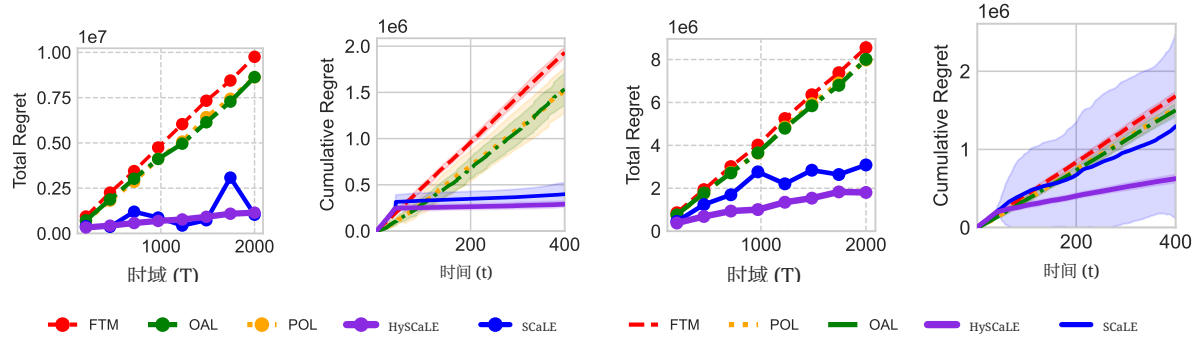


图5：SCaLE 和 HySCaLE，针对 $r = 1$， $d = 4$. 左图为 $\sigma_r^A = 10^{-2}$，右图为 $\sigma_r^A = 1$。测试设置：$\bar{\eta} = 50$、$v_t - v_{t-1} \sim \mathcal{N}(0,\ 50 \cdot I)$、 $c = 10$

瞬时遗憾揭示了 SCaLE 和 HySCaLE 初期付出的高昂代价（最高达 $m = 80$），这显著降低了利用阶段的遗憾。将 HySCaLE 与 POL 和 OAL 进行比较可见，只有在确保基线统计估计 $\hat{}A^{\texttt{SCaLE}}$ 后，投影梯度下降的优势才会显现。

重尾鲁棒性。我们的目标是在在线设置下验证备注1 ：

> $\hat{}$在老虎机在线二次优化中， A 的受限一致有界性（*RUB*）性质是否是确保稳定性的必要条件？

具体而言，POL 和 OAL 采取的动作形式如下：

$$x_t = \hat{C}_t x_{t-1} + (I - \hat{C}_t) v_t$$

并非中心化次高斯。因此，这些基线方法无法保证受限一致有界性（RUB）性质（以及统计一致性）。由于含噪梯度更新，HySCaLE 也有同样的问题（见备注 2）。

图4 显示，当HySCaLE 等基于梯度的方法，以及基线方法POL 和OAL 在 $\{v_t\}_t$ 服从柯西分布时表现不佳。SCaLE的表现则显著优于其他方法，尤其是在 σ_r^A较小时。这体现了算法1中探索阶段（第 1–5 行）和迹范数最小化（第 6 行）的重要性。

秩亏设置。 为测试曲率信号 （A） 恢复，我们考虑如下设置：$r \ll d$ （回顾图 2）。图 5 描绘了所考虑算法在 $r = 1$

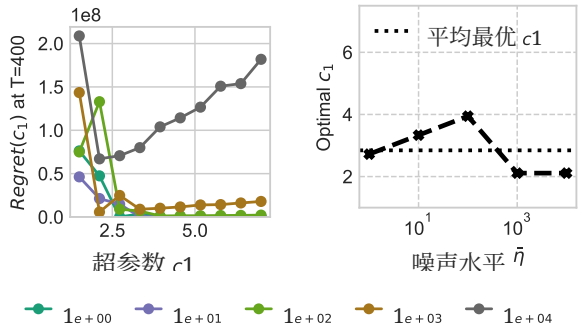


图 6： c_1 v/s $\bar{\eta}$ 敏感性分析，环境设置为 $T = 400$. $\sigma_r^A = 1.0$、 $r = 4$、 $d = 4$、$\sigma_v = 50$

以及 $d = 4$. 在这一退化设置中，定理 1 所声称的阶差异在SCaLE 和HySCaLE的动态遗憾中清晰可见。

超参数调优。 尽管声称 c_1 具有普适性，但在实际环境中它是一个超参数，需要进行调优。

因此，我们通过线搜索为上述每项实验设定 c_1 ，以最小化遗憾$_T$:

$$c_1 = \underset{c \in \mathbb{R}^+}{\operatorname{argmin}} \operatorname{Regret}_T(c). \tag{4.1}$$

我们的敏感性分析揭示了两个关键见解，如图 6所示。首先，遗憾$_T(c)$ 呈 U 型趋势，表明围绕“m”存在探索—利用权衡。其次，观察发现，最优的c_1 在广泛的噪声水平范围内几乎保持不变 (≈ 3) $\bar{\eta} \in [1, 10^4]$，这证实了 c_1 确实是一个通用参数。

5 分阶段遗憾分析与谱分解

5.1 证明概要

本节给出主要遗憾保证的证明。我们的分析将 SCaLE 的总遗憾分解为探索阶段和利用阶段的贡献。随后，我们运用谱扰动分析，揭示谱误差随估计质量（$\|A - A\| = \epsilon$）和时域（T）变化时呈现的不同尺度，正是造成遗憾行为差异的原因

定理 3. *SCaLE* 的遗憾满足以下界

$$Regret_T^{SCaLE} \leq \underbrace{m\left(\frac{(1/2)tr(\Sigma)}{1+\sigma_{\min}^A}\right)}_{Irreducible\ Cost\ of\ NOT\ knowing\ A} + \underbrace{m\gamma^2 \cdot ((1/2)tr(A)+d) + \frac{\gamma^2}{2}d}_{Cost\ of\ Data\ Collection\ \propto \frac{1}{\epsilon}} + \underbrace{Regret_{m+1:T}^{SCaLE}}_{\substack{Potential\ Benefit:\\ o(T)\ Regret}}$$

这一分解明确揭示了信噪比参数 γ^2 的作用。要想最优地选择 γ^2，就需要权衡第二项和第三项；这一选择在实践中的影响可从图 3(iii) 中看出。我们将上述分解的证明留到附录C.3 ，以便聚焦于主要的分析难点：建立一个关于估计误差 $_{m+1:T}^{\texttt{SCaLE}}$ 的紧致且非平凡的遗憾界 ϵ。

我们提出了一种新的谱遗憾分析，将利用遗憾进一步拆分为两个正交分量：一个由 A 的估计特征值误差导致，另一个则由其估计特征基中的扰动导致。定义算法类 LAI(A)：$\{M \succeq \mathbf{0}_{d\times d}\} \to \mathbb{R}^{d\times T}$，利用在线最优 LAI 基准的结构，并估计 $\hat{A}$：

$$\{\hat{C}_t\}_{t=1}^{T} \leftarrow \texttt{RecurGen}(\hat{A}, 1)$$
$$x_t^{\texttt{LAI}(\hat{A})} = \hat{C}_t x_{t-1}^{\texttt{LAI}(\hat{A})} + (I - \hat{C}_t)v_t \;\forall\, t$$

设特征值分解为 $A = QDQ^T$和 $\hat{A} = P\hat{D}P^T$。

命题 1. 对于任意鞅序列 $\{v_t\}_{t=1}^{T}$，*LAI*($\hat{A}$) 的动态遗憾恰好为

$$Regret_T^{\texttt{LAI}(\hat{A})} = \underbrace{\sum_{s=1}^{T} \frac{1}{2}\mathbb{E}\|v_s - v_{s-1}\|^2_{C_s - \hat{C}_s}}_{R_T^{val}:\ dominated\ by\ \Delta_{val}=\hat{D}-D} + \underbrace{\sum_{s=1}^{T} \frac{1}{2}\mathbb{E}\|x_{s-1} - v_s\|^2_{\hat{C}_s(A-\hat{A})\hat{C}_s}}_{R_T^{basis}:\ dominated\ by\ \Delta_{basis}=P-Q}$$

该分解在附录 C.1中得到证明，是对[8]中分析的一项非平凡扩展；在整个时域 $\{1, \ldots, T\}$ 上解耦谱分析，以生成各个遗憾项R_T^{val} 和 R_T^{basis}.

接下来的挑战是将核范数误差 $\|A - \hat{A}\|_*$ 精确地转化为两个遗憾分量：

$$\|\hat{A} - A\|_* \le \epsilon \begin{cases} \Delta_{val}(\epsilon) \longrightarrow R_T^{val}(\epsilon, T) \\ \Delta_{basis}(\epsilon) \longrightarrow R_T^{basis}(\epsilon, T) \end{cases}$$

矩阵扰动理论 [52] 是这里的关键工具。我们的工作有一个核心发现：对于秩亏（$r < d$）和满秩 A. 这两种情形，$R_T^{val}(\epsilon, T)$ 与 $R_T^{basis}(\epsilon, T)$ 对总遗憾的贡献相同

我们先讨论更具挑战性的秩亏情形。A 中存在奇异区域，这使标准控制/遗憾分析 [5, 6] 所依赖的"误差衰减传播"无法实现。朴素分析会导致估计误差在每一轮 t线性累积（$\simeq \epsilon t$），从而造成无法控制的总遗憾 ϵT^2。我们通过精确刻画特征基扰动所导致的误差累积来克服这一问题：

定理 4. 设 $rank(A) = r < d$。对于任意满足 $\|A - \hat{A}\|_* \le \epsilon$ 的估计 $\hat{A}$，*LAI*($\hat{A}$) 的动态遗憾有界于

$$Regret_T^{\texttt{LAI}(\hat{A})} \le \underbrace{\left(\frac{2d\sigma^A_{\max}\|\Sigma\|_F}{(\sigma^A_{\min} - \epsilon)^2}\right)\epsilon^2 T^2}_{R_T^{val}} + \underbrace{tr(\Sigma)\sqrt{\epsilon}T}_{R_T^{basis}} + \epsilon T \cdot \frac{(3\sqrt{2d} + 2\epsilon d + 2\sigma^A_{\max}\sqrt{2d})\|\Sigma\|_F}{(\sigma^A_{\min} - \epsilon)^2}$$

关键在于，由特征基扰动导致的主导误差项，受控于的并非 ϵ，而是小得多的 ϵ^2 因子。在时域内，该误差演变为 $\mathcal{O}(\epsilon^2 T^2)$ 遗憾，而非（不可控的） $\mathcal{O}(\epsilon T^2)$。揭示这一结构的细致谱分析将在接下来的第 5.2 节中展开。

当 A 为满秩时，由于误差在整个时域内几何衰减传播，每一步的误差都是常数。应用针对秩亏情形（第5.2节）所建立的分析，可得到以下改进的遗憾保证，具体内容见附录C.2。

ˆ ˆ**定理 5**。 设 $rank(A) = d$。对于任意满足 $\|A - A\|_* \le \epsilon$ 的估计 A， ***LAI***ˆ 的利用遗憾(A)有界于

$$Regret_T^{\mathit{LAI}(\hat{A})} \le \underbrace{\epsilon T \cdot \frac{(1 + (\sqrt{2} + \epsilon\sqrt{d})\sigma^A_{\max})\sqrt{d}\|\Sigma\|_F}{(\sigma^A_{\min} - \epsilon)^3}}_{R_T^{val}} + \underbrace{\epsilon T \cdot \frac{tr(\Sigma)}{2\sigma^A_{\min}}}_{R_T^{basis}}.$$

ˆ最后，要通过秩一测量实现 $\|A - A\|_* \le \epsilon$ 的估计误差，探索阶段必须付出 $\Theta(1/\epsilon)$的代价。以下推论表明，将这一代价与定理4 和5 中的利用遗憾相平衡，会使两个谱误差项的贡献相等：

推论 1. 探索遗憾 $\mathit{Regret}_{1:m} = \Theta(1/\epsilon)$与利用遗憾 $\mathit{Regret}^{(\mathit{ALG}\ \hat{}A)}_{m+1:T}$之间的最优平衡在以下情况下实现：

1. 当满足以下条件时，秩亏$(r < d)$ A

$$R_T^{val} = \mathcal{O}(T^{2/3}) \quad and \quad R_T^{basis} = \mathcal{O}(T^{2/3}),$$

2. 当满足以下条件时，满秩$(r = d)$ A

$$R_T^{val} = \mathcal{O}(T^{1/2}) \quad and \quad R_T^{basis} = \mathcal{O}(T^{1/2}).$$

附录C.4 中的详细证明给出了精确的遗憾表达式，以及 γ^2的最优取值，该取值用于SCaLE (算法1)。它将本节的详细分析与定理1中给出的主要保证联系起来。谱分解揭示的关键见解是：当特征值误差（R_T^{val}）和特征基误差（R_T^{basis}）的遗憾贡献处于同一量级时，探索与利用达到平衡。

5.2 主要分析贡献：定理证明4

本节重点介绍本文的核心方法论贡献：在误差增大的情况下实现次线性遗憾，而这通常源于目标函数的曲率退化。为此，我们直接运用标准的矩阵扰动结果（参见附录B）和命题1 （证明见附录C.1）。现在，让我们再次回到精确的广义遗憾形式：

$$\mathrm{Regret}[1,T] = \sum_{s=1}^{T} \frac{1}{2}\mathbb{E}\|x_{s-1} - v_s\|^2_{\hat{C}_s(A-\hat{A})\hat{C}_s} + \sum_{s=1}^{T} \frac{1}{2}\mathbb{E}\|v_{s-1} - v_s\|^2_{C_s - \hat{C}_s}$$

其中

$$\mathbb{E}\|x_{s-1} - v_s\|^2_{\hat{C}_s(A-\hat{A})\hat{C}_s} = \sum_{k=0}^{s-1} \mathbb{E}\|v_{s-k-1} - v_{s-k}\|^2_{\prod_{l=k}^{0} \hat{C}_{s-l}(A-\hat{A}) \prod_{l=0}^{k} \hat{C}_{s-l}}.$$

我们来看看

$$\prod_{l=k}^{0} \hat{C}_{s-l}(A - \hat{A}) \prod_{l=0}^{k} \hat{C}_{s-l}$$

$$
\begin{aligned}
=&P \times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1},\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right)\times\underbrace{(P^TQDQ^TP-\hat{D})}_{P^TQDQ^TP-D+D-\hat{D}}\times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1},\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right)\times P^T.\\
=&\underbrace{P \times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1},\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right)\times\underbrace{(P^TQDQ^TP-D)}_{\Delta_2}\times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1},\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right)P^T}_{X_1}\\
&+\underbrace{P \times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1},\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right)\times(D-\hat{D})\times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1},\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right)P^T}_{X_2}.
\end{aligned}
$$

我们先看看 X_2：

$$
\begin{aligned}
D-\hat{D}&=diag(\lambda^{A,1}-\hat{\lambda}^1,\ldots,\lambda^{A,d}-\hat{\lambda}^d)\\
&\preceq diag(\underbrace{\epsilon,\ldots,\epsilon}_{r},\underbrace{0,\ldots,0}_{d-r})
\end{aligned}
$$

此外，以下结论成立

$$
\hat{\lambda}_t^i\leq\begin{cases}\frac{1}{1+\lambda^{A,i}-\epsilon} & i\in\{1,\ldots,r\}\\ 1 & i\in\{r+1,\ldots,d\}\end{cases}
$$

这将得到

$$
\begin{aligned}
&diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1},\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right)\times(D-\hat{D})\times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1},\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right)\\
&\preceq\underbrace{diag\left(\frac{\epsilon}{(1+\lambda^{A,1}-\epsilon)^{2k+2}},\ldots,\frac{\epsilon}{(1+\lambda^{A,r}-\epsilon)^{2k+2}},0,\ldots,0\right)}_{\Delta_1}
\end{aligned}
$$

从而得到

$$
\begin{aligned}
\mathbb{E}\|v_{s-k-1}-v_{s-k}\|_{X_2}^2&=\langle X_2,\Sigma\rangle\\
&=\langle\Delta_1,P^T\Sigma P\rangle\\
&=\sum_{i=1}^{r}\frac{\epsilon}{(1+\lambda^{A,i}-\epsilon)^{2k+2}}p_i^T\Sigma p_i\\
&\leq\frac{\epsilon}{(1+\sigma_{\min}^A-\epsilon)^{2k+2}}tr(P^T\Sigma P)\\
&=\epsilon\cdot\frac{tr(\Sigma)}{(1+\sigma_{\min}^A-\epsilon)^{2k+2}}
\end{aligned}
$$

现在我们来看看 X_1，更具体地说，是 Δ_2:

$$
\begin{aligned}
\Delta_2&=P^TQDQ^TP-D\\
&=(P^TQ-I+I)D(Q^TP-I+I)-D\\
&=(P^TQ-I)D(Q^TP-I)+(P^TQ-I)D+D(Q^TP-I)+D-D
\end{aligned}
$$

$$= \underbrace{(P^TQ - I)D(Q^TP - I)}_{\Delta_{2a}} + \underbrace{(P^TQ - I)D}_{\Delta_{2b}} + \underbrace{D(Q^TP - I)}_{\Delta_{2c}}$$

现在请注意,

$$\begin{aligned}\Delta_{3b} &= diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right) \times \Delta_{2b} \times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right) \\ &= diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right) \times (P^TQ - I) \times diag\left(\underbrace{\lambda^{A,1}\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \lambda^{A,r}\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{r}}_{r}, \underbrace{0, \ldots, 0}_{d-r}\right)\end{aligned}$$

and

$$\begin{aligned}\Delta_{3c} &= diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right) \times \Delta_{2c} \times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right) \\ &= diag\left(\underbrace{\lambda^{A,1}\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \lambda^{A,r}\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{r}}_{r}, \underbrace{0, \ldots, 0}_{d-r}\right) \times (P^TQ - I)diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right).\end{aligned}$$

最后，相应地定义 Δ_{3a} ，

$$\Delta_{3a} = diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right) \times \Delta_{2a} \times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^{1}, \ldots, \prod_{l=k}^{0}\hat{\lambda}_{s-l}^{d}\right).$$

现在将这三者结合起来

$$\begin{aligned}\mathbb{E}\|v_{s-k-1} - v_{s-k}\|_{X_1}^2 &= \langle\Delta_{3a}, \mathbb{P}^T\Sigma P\rangle + \langle\Delta_{3b}, \mathbb{P}^T\Sigma P\rangle + \langle\Delta_{3c}, P^T\Sigma P\rangle \\ &\leq \|\Sigma\|_F \cdot (\|\Delta_{3a}\|_F + \|\Delta_{3b}\|_F + \|\Delta_{3c}\|_F)\end{aligned}$$

现在， $\|P^TQ - I\|_F = \|Q^TP - I\|_F = \|P - Q\|_F \leq \sqrt[\top]{d} \cdot \|P - Q\|_{op} \leq \frac{\sqrt[\top]{2d}\cdot\epsilon}{\sigma_{\min}^A - \epsilon}$。 因此，我们有

$$\|\Delta_{3b}\|_F = \|\Delta_{3c}\|_F \leq \frac{\sigma_{\max}^A\sqrt{2d} \cdot \epsilon}{(\sigma_{\min}^A - \epsilon)(1 + \sigma_{\min}^A - \epsilon)^{k+1}}$$

and

$$\begin{aligned}\|\Delta_{3a}\|_F &\leq \|\Delta_{2a}\|_F \\ &\leq \|P^TQ - I\|_F \cdot \|D(Q^TP - I)\|_F \\ &\leq \frac{2d\sigma_{\max}^A \cdot \epsilon^2}{(\sigma_{\min}^A - \epsilon)^2}.\end{aligned}$$

现在将 X_1 和 X_2 放在一起,

$$\mathbb{E}\|x_{s-1} - v_s\|_{\hat{C}_s(A-\hat{A})\hat{C}_s}^2$$

$$\leq \sum_{k=0}^{s-1} \frac{2d\sigma_{\max}^A \|\Sigma\|_F \cdot \epsilon^2}{(\sigma_{\min}^A - \epsilon)^2} + 2\sum_{k=0}^{s-1} \frac{\|\Sigma\|_F \sigma_{\max}^A \sqrt{2d} \cdot \epsilon}{(\sigma_{\min}^A - \epsilon)(1 + \sigma_{\min}^A - \epsilon)^{k+1}} + \sum_{k=0}^{s-1} \epsilon \cdot \frac{tr(\Sigma)}{(1 + \sigma_{\min}^A - \epsilon)^{2k+2}}$$
$$\leq \left(\frac{2d\sigma_{\max}^A \|\Sigma\|_F}{(\sigma_{\min}^A - \epsilon)^2} \right) \epsilon^2 \cdot s + \frac{2\|\Sigma\|_F \sigma_{\max}^A \sqrt{2d} \cdot \epsilon}{(\sigma_{\min}^A - \epsilon)(1 + \sigma_{\min}^A - \epsilon)} + \frac{tr(\Sigma)\epsilon}{(1 + \sigma_{\min}^A - \epsilon)^2 - 1}$$

这导致了一个遗憾:

$$\text{Regret}[1,T] \leq \left(\frac{d\sigma_{\max}^A \|\Sigma\|_F}{(\sigma_{\min}^A - \epsilon)^2} \right) \epsilon^2 \cdot T(T+1) + \frac{2\|\Sigma\|_F \sigma_{\max}^A \sqrt{2d} \cdot \epsilon T}{(\sigma_{\min}^A - \epsilon)(1 + \sigma_{\min}^A - \epsilon)} + \frac{tr(\Sigma)\epsilon T}{(1 + \sigma_{\min}^A - \epsilon)^2 - 1}$$
$$+ \sum_{s=1}^{T} \frac{1}{2} \mathbb{E} \|v_{s-1} - v_s\|_{C_s - \hat{C}_s}^2 .$$

现在我们来考察这一项 $\sum_{s=1}^{T} \frac{1}{2} \mathbb{E} \|v_{s-1} - v_s\|_{C_s - \hat{C}_s}^2$:

$$\begin{aligned} C_s - \hat{C}_s &= Q D_s Q^T - P \hat{D}_s P^T \\ &= P \left(P^T D_s Q^T P - \hat{D}_s \right) P^T \end{aligned}$$

其中

$$P^T D_s Q^T P - \hat{D}_s = \underbrace{(P^T Q - I) D_s (Q^T P - I)}_{\Delta_{4a}} + \underbrace{(P^T Q - I) D_s}_{\Delta_{4b}} + \underbrace{D_s (Q^T P - I)}_{\Delta_{4c}} + \underbrace{D_s - \hat{D}_s}_{\Delta_{4d}}$$

根据我们此前的计算:

$$\begin{aligned} \|\Delta_{4a}\|_F &\leq \|P^T Q - I\|_F \cdot \|D_s (P^T Q - I)\|_F \\ &\leq \frac{\sqrt{2d} \cdot \epsilon}{\sigma_{\min}^A - \epsilon} \cdot 1 \cdot \frac{\sqrt{2d} \cdot \epsilon}{\sigma_{\min}^A - \epsilon} \\ &\leq \frac{2d\epsilon^2}{(\sigma_{\min}^A - \epsilon)^2} \\ \|\Delta_{4b}\|_F = \|\Delta_{4c}\|_F &\leq \frac{\sqrt{2d} \cdot \epsilon}{\sigma_{\min}^A - \epsilon}. \end{aligned}$$

其中我们用到这 样一个事实: D_s 的每个对角元素都不大于 1。现在我们来关注最后一项 :

$$D_s - \hat{D}_s = diag(\lambda_t^1 - \hat{\lambda}_t^1, \ldots, \lambda_r^1 - \hat{\lambda}_t^r, \lambda_t^{r+1} - \hat{\lambda}_t^{r+1}, \ldots, \lambda_t^d - \hat{\lambda}_t^d).$$

对于 $i \in \{1, \ldots, r\}$,

$$\begin{aligned} \frac{1}{\lambda_t^i} &= 2 + \lambda^{A,i} - \lambda_{t+1}^i \\ \frac{1}{\hat{\lambda}_t^i} &= 2 + \hat{\lambda}^{A,i} - \hat{\lambda}_{t+1}^i \\ \implies |\lambda_t^i - \hat{\lambda}_t^i| &\leq \lambda_t^i \hat{\lambda}_t^i \cdot |\lambda_{t+1}^i - \hat{\lambda}_{t+1}^i| + \lambda_t^i \hat{\lambda}_t^i \cdot |\lambda^{A,i} - \hat{\lambda}^{A,i}| \\ &\leq \frac{|\lambda_{t+1}^i - \hat{\lambda}_{t+1}^i|}{1 + \lambda^{A,i}} + \frac{|\lambda^{A,i} - \hat{\lambda}^{A,i}|}{1 + \lambda^{A,i}}. \end{aligned}$$

递归执行此操作会得到

$$|\lambda_t^i - \hat{\lambda}_t^i| \leq \frac{|\lambda^{A,i} - \hat{\lambda}^{A,i}|}{\lambda^{A,i}} \leq \frac{\epsilon}{\min_{i \leq r} \lambda^{A,i}} = \frac{\epsilon}{\sigma_{\min}^A}$$

对于 $i \in \{1, \ldots\}$. 现在，这对这些索引成立的唯一原因是 $\lambda^{A,\ i} = 0$ 对于 $i > r$. 这些索引， $\lambda_t^i = 1$ 对于所有 t. 因此，我们只考察关于 $\hat{}\ \lambda_t^i$ 的方程

$$\frac{1}{\hat{\lambda}_t^i} = 2 + \hat{\lambda}^{A,i} - \hat{\lambda}_{t+1}^i$$

其中我们知道 $0 \leq \hat{}_{\lambda^{A,\ i}} < \epsilon$ 对于所有 $i > r$. 对 的随机动态规划分析 [8]表明

$$\frac{\hat{\lambda}^{A,i} + 2 - \sqrt{(\hat{\lambda}^{A,i})^2 + 4\hat{\lambda}^{A,i}}}{2} = \hat{\lambda}_L^i \leq \hat{\lambda}_t^i < \frac{1}{1 + \hat{\lambda}^{A,i}}$$

and

$$\frac{\hat{\lambda}^{A,i} + 2 - \sqrt{(\hat{\lambda}^{A,i})^2 + 4\hat{\lambda}^{A,i}}}{2} > \frac{\epsilon + 2 - \sqrt{\epsilon^2 + 4\epsilon}}{2} = 1 - \sqrt{\epsilon} + \frac{\epsilon}{2} > 1 - \sqrt{\epsilon}$$

由此得到

$$\lambda_t^i - \hat{\lambda}_t^i \leq \sqrt{\epsilon} \ \forall \ i > r$$

因此我们有

$$(D_s - \hat{D}_s)_{i,i} \leq \max\left\{\sqrt{\epsilon}, \frac{\epsilon}{\sigma_{\min}^A}\right\}$$

我们将其代入 $\mathbb{E}\|v_{s-1} - v_s\|^2_{C_s - \hat{} C_s}$， 得到

$$\mathbb{E}\|v_{s-1} - v_s\|^2_{C_s - \hat{C}_s} \leq \left(\frac{2d\epsilon^2}{(\sigma_{\min}^A - \epsilon)^2} + \frac{\sqrt{2d} \cdot \epsilon}{\sigma_{\min}^A - \epsilon}\right) \|\Sigma\|_F + \max\left\{\sqrt{\epsilon}, \frac{\epsilon}{\sigma_{\min}^A}\right\} \cdot tr(\Sigma).$$

$\hat{}\ \hat{}$ 估计 A 的最终遗憾上界（误差率为 $\|A - A\|_* \leq \epsilon$ ）是

$$\text{Regret}_T \leq \underbrace{\left(\frac{d\sigma_{\max}^A \|\Sigma\|_F}{(\sigma_{\min}^A - \epsilon)^2}\right) \epsilon^2 \cdot T(T+1) + \frac{2\|\Sigma\|_F \sigma_{\max}^A \sqrt{2d} \cdot \epsilon T}{(\sigma_{\min}^A - \epsilon)(1 + \sigma_{\min}^A - \epsilon)} + \frac{tr(\Sigma)\epsilon T}{(1 + \sigma_{\min}^A - \epsilon)^2 - 1}}_{\text{Regret}_T^A}$$

$$+ \underbrace{T\left(\left[\frac{2d\epsilon^2}{(\sigma_{\min}^A - \epsilon)^2} + \frac{\sqrt{2d} \cdot \epsilon}{\sigma_{\min}^A - \epsilon}\right] \|\Sigma\|_F + \max\left\{\sqrt{\epsilon}, \frac{\epsilon}{\sigma_{\min}^A}\right\} \cdot tr(\Sigma)\right)}_{\text{Regret}_T^B}.$$

将上述表达式与探索阶段的遗憾相结合，可得以下动力学方程，总结了总遗憾以及我们对 $1/\gamma^2 = \Theta(\epsilon) \ \propto T^{-2/3}$ 在 $r < d$ 算法 1中的选择：

$$\text{Regret}_T^{\texttt{SCaLE}} = \mathcal{O}(\gamma^2) + \mathcal{O}(\epsilon^2 T^2) + \mathcal{O}(\epsilon^{1/2} \cdot T).$$

定理5 的证明以及 γ^2 在满秩（$r = d$）情形下的相应选取采用了类似的技巧，详见附录C.2 和 C.4。

6 激励性应用

本节详细介绍两个直接促成本研究问题建模的现实控制问题——数据中心热管理和协作式风电场控制。在这两种场景中，控制器都必须在显著切换成本的约束下，跟踪一个随机极小值点，同时优化一个由未知物理耦合矩阵决定的二次目标函数。

6.1 数据中心热管理（分区气流控制)

现代数据中心消耗大量能耗，其中相当一部分用于冷却基础设施。控制目标是调节计算机房空调（CRAC）设备的气流速率，记为 $x_t \in \mathbb{R}^d$，在维持服务器安全温度的同时，最大限度地降低能耗 [53]。

- **最小化器（v_t）：** 最优气流配置由分布在服务器机架上的 IT负载（热量产生）决定。由于计算任务遵循随机到达过程——通常建模为随机游走或突发性布朗运动——“目标”冷却曲线 v_t 表现为一个鞅过程，该过程通过控制执行前的 CPU利用率指标进行观测 [54]。

- **未知矩阵（A）：** 矩阵 $A \succeq 0$ 表示计算机房空调设备与服务器机架之间的对流热耦合（或热交叉干扰矩阵），其中 A_{ij} 量化了风扇 i 对机架 j [55, 56]的冷却影响。尽管该矩阵归设施所有，但实际上它是未知的，并且通常会随时间变化，这是由复杂的湍流气流、回流和物理退化（例如堵塞的过滤器、电缆移位）造成的。因此，基于计算流体力学（CFD）的静态离线模型往往无法捕捉实时动态，因而需要进行在线辨识 [57, 58]。

- **成本结构：** 命中成本 $f_t(x_t)$ 描述了功耗与热风险之间的权衡。偏离最佳气流 v_t 会导致功率受到三次方惩罚（过度冷却），或硬件故障风险呈指数级上升（冷却不足）；通常使用二次惩罚函数 [25]对其进行近似。至关重要的是，系统会施加与$\|x_t - x_{t-1}\|^2$成正比的切换成本。风扇转速快速波动会显著加剧变频驱动器（VFD）的损耗并引发声疲劳，因此需要平滑的控制轨迹，以最大限度延长设备寿命。

6.2 协作式风电场控制（尾流偏转）

在风力涡轮机密集的风电场中，上游风力涡轮机会产生“尾流”——湍流、低风速的气流区域——显著降低下游风力涡轮机的功率输出。“尾流偏转”是一种协作控制策略，上游风力涡轮机会有意调整其偏航角（记为 $x_t \in \mathbb{R}^d$），使尾流偏离下游相邻风力涡轮机，从而最大限度地提高风电场总功率 [59, 60]。

- **最小化器（v_t）：** 最优偏航配置严格取决于瞬时自由来流风向。风向是一种随机环境变量，会呈现非平稳漂移，实际上充当了由机舱式风速计观测到的鞅驱动力 [61, 62]。

- **未知矩阵（A）：** 矩阵 A 描述了风力涡轮机之间的气动相互作用（或灵敏度雅可比矩阵）。它决定了风力涡轮机的偏航角发生变化时

i 对下游风力涡轮机功率捕获的影响 j。虽然存在 Jensen 或高斯尾流模型等解析替代模型 [63, 64]，但实际的相互作用取决于湍流强度、风切变和热稳定性等无法直接观测的大气参数 [65]。随着大气条件变化 A，有效耦合也会改变，因此控制器需要持续估计局部气动灵敏度 [66]。

- **成本结构：**目标函数是最大化总功率，而总功率在工作点附近局部呈凸性。命中成本（$x_t - v_t)^\top A(x_t - v_t)$）会对偏离最优尾流偏转角的情况进行惩罚。在这一领域，切换成本至关重要。偏转庞大的机舱结构会产生巨大的陀螺载荷，导致塔架和叶片出现结构疲劳（等效损伤载荷） [67]。为防止机械故障并延长设备寿命，控制器必须通过 $\|x_t - x_{t-1}\|_2^2$ [68] 对偏航速率进行惩罚。

7 相关研究

本节详细概述了与我们工作交叉的研究领域，即在线算法、控制理论、在线学习和逆向最优控制。

7.1 带切换成本的在线算法

这一领域的文献最直接地探讨了在线决策中管理移动成本这一挑战，主要包括三个相关子领域：度量任务系统（MTS）、凸体追逐（CBC）和平滑在线凸优化（SOCO）。

度量任务系统（MTS）作为一种用于研究带切换成本的在线问题的基础模型，MTS 将问题描述为：一名在线玩家在度量空间中移动服务器，以处理一系列请求（任务） [37, 69]。目标是最小化移动成本与任务处理成本之和。该领域历史悠久，致力于开发竞争算法并理解这一问题的根本限制 [70, 71, 72], ，其成果在分页 [73, 74], 、k服务器问题 [75, 76] 和缓存 [77, 78]，等领域有着重要应用。近期文献的研究重点已转向利用机器学习建议，设计既稳健又能把握机会的算法 [72, 79]。

凸体追逐（CBC）CBC 可视为 MTS 在连续空间中的几何推广：玩家必须在在线揭示的一系列凸集（凸体）中始终维持一个点的位置，同时尽量减少移动的总距离 [39]。这一框架推动了重要的理论进展，并与优化的多个领域建立了联系 [80, 81]。然而，否定性结果表明，在高维空间中，这一问题已被证明可能很难求解 [39]。

平滑在线凸优化（SOCO）SOCO 从函数优化的角度审视 MTS 和 CBC，将问题定义为：最小化一系列命中成本函数之和，以及连续动作之间的切换成本 [8, 26, 38, 40, 82]。这一框架与我们的工作尤为相关，并已成功应用于涉及（可能无界）决策空间的领域，例如数据中心管理 [20], 、自适应控制 [7, 83]和分布式控制 [26, 84]。

传统上，大多数研究都考虑对抗性设定，并针对动态事后最优解提供最坏情况保证。早期研究 [38, 39] 为一维设定下的凸命中代价给出了精确刻画。高维分析 [7, 39] 证明了其困难性

在没有对命中成本结构作出额外假设的情况下，对抗性设定具有困难性。人们对高维设定的不同方面进行了探索，并在以下方面取得了显著进展：(i)未来预测 [84, 85], (ii)不可信建议 [72]以及 (iii)多智能体场景 [26, 84]。

关键区别 这三个领域有一个关键且几乎普遍采用的假设：在线算法在每一步都完全了解当前的命中成本函数（或凸体）[7, 8, 40, 72]。虽然近期少数研究考虑了梯度 [86]或延迟反馈 [86, 87], 等较弱的反馈模型，但它们仍假设存在无噪声预言机。我们的工作则大不相同：我们在更弱且带噪声的零阶（赌博机）反馈模型下研究 SOCO问题，此时必须在线学习成本函数的结构。

7.2 控制中的学习：系统辨识与线性二次高斯

该领域关注在控制系统的同时学习其未知的动力学[5, 6, 17]。核心目标是从输入输出数据中，以在线方式学习系统状态空间模型的参数，同时对所采取的动作 $\{u_t\}_t$ 和由此产生的状态 $\{x_{t+1}\}_t$承担已知的二次成本 $x_t^T Q x_t + u_t^T R u_t$。近期研究 [18, 88] 越来越多地采用在线学习技术，为线性动力系统 $x_{t+1} = Ax_t + Bu_t + w_t$中未知的状态转移矩阵 A以及 B 提供性能保证，其中 w_t 是独立同分布高斯漂移。这对自适应控制 [18, 19]应用至关重要。这类系统与前述二次代价相结合，构成了“线性二次高斯”（LQG）控制范式。在 LQG 设置中，目标是在高斯漂移下控制具有二次成本的线性系统。采用一阶反馈 [5, 6]的遗憾最小化保证已得到确立。

关键区别尽管这些领域将学习与控制相结合，但其主要关注点是学习**系统动力学**（例如，状态转移矩阵），而我们的工作则侧重于学习**成本函数**本身。此外，它们的理论保证通常依赖于一些假设，使得不受限移动成为次要考量，例如系统可控性/稳定性 [5, 6, 18]或低漂移的次高斯过程 [5, 6], ，而我们并不作出这些假设。

7.3 在线学习与在线凸优化（OCO）

老虎机凸优化（BCO）代表了信息受限的极端情形：算法必须选择一个动作，且只能收到由此产生的成本值（零阶反馈）[89, 90]。这与我们的反馈模型一致。在为这一设定设计具有最优遗憾的算法方面，已有显著进展 [9, 11]。这种方法已直接应用于非随机控制，其中对抗性扰动会得到处理 [10, 91]„ 但需要满足可控性假设。大量在线凸优化和老虎机凸优化研究在两个基本假设下提供次线性遗憾保证：(i) 动作集是紧致的有界域 [9, 29, 91] ；(ii) 遗憾采用静态基准 [91]。直到最近，研究才开始探讨自适应基准 [29] ，但仍然依赖有界空间和梯度。

关键区别 主要区别在于如何处理切换成本。绝大多数 OCO 和 BCO 文献都没有将切换成本纳入目标函数；目标只是最小化累积命中成本。这类文献普遍假设存在**有界动作空间** ，这实际上限制了任何可能的移动，因此无需显式管理切换成本。我们的研究则探讨了更具挑战性的问题：在无界域中管理显式且无界的切换成本。这一问题在 BCO 文献中鲜有研究。

7.4 逆最优控制（IOC）与逆强化学习

该研究领域关注如何根据智能体行为的示范推断其目标（即其成本函数或奖励函数）。逆最优控制（IOC），尤其是在线性二次调节器（LQR）背景下，在学习潜在二次代价矩阵方面取得了显著进展（Q和 R），这些矩阵是根据观测到的状态-动作轨迹 [12]推断得出的。这与机器人学[92, 93]和人类行为建模 [94, 95]等应用密切相关。IOC 及更广泛的逆强化学习（IRL）领域的主流范式是离线且数据驱动的。最先进的算法通常假设可以访问预先收集的专家或事后最优轨迹数据集，并据此学习成本函数[12, 13], ，其表述属于“离策略”或数据驱动范畴。直到最近，同策略 IOC/IRL 才开始取得进展；在这类方法中，对专家轨迹的依赖有所降低，但并未完全消除 [96, 97]。

关键区别 根本区别在于问题是**在线还是离线**。IOC/IRL 从过去行为的静态数据集中学习。我们的算法以**在线**方式运行，必须在没有现成专家示范的情况下，同时学习成本函数并做出决策。算法收到的反馈由其自身动作产生，从而形成探索、估计与控制之间的紧密闭环，而离线 IOC 场景中不存在这种闭环。

8 结论

在这项工作中，我们提出了 SCaLE，这是首个兼顾动态二次命中代价的老虎机探索与移动高效利用的算法，可降低无界度量切换代价，并在秩亏和满秩两种设定下实现次线性动态遗憾。我们的关键发现是，秩亏成本对应的 $\mathcal{O}(T^{2/3})$ 遗憾是该领域的首例，这得益于从谱视角审视传统遗憾分析。未来的研究方向包括在轻尾环境中为混合利用方案 HySCaLE 提供遗憾保证，以及将先探索后利用方法扩展到对抗环境。

参考文献

[1] Deepjyoti Deka、Scott Backhaus 和 Michael Chertkov。配电网络中的结构学习。*IEEE*网络系统控制汇刊，5(3):1061–1074，2018年。doi: 10.1109/TCNS.2017.2673546。[2] Guido Cavraro、Andrey Bernstein、Vassilis Kekatos 和 Yingchen Zhang。在有限监测下实时识别配电网络拓扑。*IEEE*控制系统快报，4(2):325–330，2019年。[3] Duy Nguyen-Tuong 和 Jan Peters。机器人控制的模型学习：综述。认知加工，12(4):319–340，2011年。[4] Lukas Hewing、Kim P Wabersich、Marcel Menner 和 Melanie N Zeilinger。基于学习的模型预测控制：迈向控制领域的安全学习。控制、机器人与自主系统年度评论，3(1):269–296，2020年。[5] Sahin Lale、Kamyar Azizzadenesheli、Babak Hassibi 和 Anima Anandkumar。线性二次高斯（LQG）设定下的自适应控制与遗憾最小化。载于*2021*年美国控制会议（*ACC*），第2517–2522页，2021年。doi: 10.23919/ACC50511.2021.9483309。

[6] Archith Athrey、Othmane Mazhar、Meichen Guo、Bart De Schutter 和 Shengling Shi。具有加性探索的基于学习的线性二次高斯控制的遗憾分析。载于2024年欧洲控制会议（ECC），第1795–1801页，2024年。doi: 10.23919/ECC64448.2024. 10590739。[7] Gautam Goel 和 Adam Wierman。用于平滑回归和线性二次调节控制的在线算法。第89卷，第2504–2513页。PMLR，2019年7月。[8] Neelkamal Bhuyan、Debankur Mukherjee 和 Adam Wierman。平滑在线二次优化的两全其美保证。载于第41届国际机器学习会议论文集，《机器学习研究论文集》第235卷，第3850–3888页。PMLR，2024年7月21–27日。网址：https://proceedings.mlr.press/v235/bhuyan24a.html。[9] Elad Hazan 和 Yuanzhi Li。老虎机凸优化的最优算法。arXiv预印本 arXiv:1603.04350，2016年。[10] Y Jennifer Sun 和 Zhou Lu。超越二次情形的bandit控制紧确速率。神经信息处理系统进展，37:43894–43924，2024年。[11] Xiaoqi Liu、Dorian Baudry、Julian Zimmert、Patrick Rebeschini 和 Arya Akhavan。非平稳老虎机凸优化：全面研究。arXiv预印本 arXiv:2506.02980，2025年。[12] Bosen Lian、Wenqian Xue、Yijing Xie、Frank L Lewis 和 Ali Davoudi。针对离散时间对抗性未知系统的离策略逆Q学习。Automatica，155:111171，2023年。[13] Hamed Jabbari Asl 和 Eiji Uchibe。连续时间非线性系统的数据驱动逆最优控制。arXiv预印本 arXiv:2503.09090，2025年。[14] Lijun Zhang、Wei Jiang、Shiyin Lu 和 Tianbao Yang。重新审视平滑在线学习。arXiv:2102.06933，2021年。[15] Seung-Jun Kim 和 Geogios B Giannakis。采用在线凸优化方法实现需求响应的实时能源定价。IEEE智能电网汇刊，8:2784–2793，2017年。doi: 10.1109/TSG.2016.2539948。[16] Hao Wang、Jianwei Huang、Xiaojun Lin 和 Hamed Mohsenian-Rad。探索智能电网与数据中心的交互，以实现电力负荷平衡。SIGMETRICS Perform. Eval. Rev.，41:89–94，2014年1月。ISSN 0163-5999。doi: 10.1145/2567529.2567556。[17] Yasin Abbasi-Yadkori 和 Csaba Szepesvári。线性二次系统自适应控制的遗憾界。第19卷，第1–26页。PMLR，2011年1月。网址：https://proceedings. mlr. press/v19/abbasi-yadkori11a.html。[18] Sahin Lale、Kamyar Azizzadenesheli、Babak Hassibi 和 Anima Anandkumar。通过更多探索降低线性二次调节器的遗憾。ArXiv，abs/2007.12291，2020年。网址：https: //api.semanticscholar.org/CorpusID:220768718。[19] Alon Cohen、Tomer Koren 和 Yishay Mansour。仅需t遗憾即可高效学习线性二次调节器。第97卷，第1300–1309页。PMLR，2019年1月。网址：https://proceedings. mlr.press/v97/cohen19b.html。

[20] Minghong Lin、Zhenhua Liu、Adam Wierman 和 Lachlan L H Andrew。地理负载均衡在线算法。第1–10页，2012年。doi: 10.1109/IGCC.2012.6322266。[21] Susanne Albers 和 Jens Quedenfeld。异构数据中心合理规模配置算法。载于第33届ACM算法与体系结构并行性研讨会论文集，第48–58页，2021年。[22] Lingwen Gan、Ufuk Topcu 和 Steven H Low。电动汽车充电的最优去中心化协议。《IEEE电力系统汇刊》，28:940–951，2013年。doi: 10.1109/TPWRS.2012.2210288。[23] Tianyu Chen、Yiheng Lin、Nicolas Christianson、Zahaib Akhtar、Sharath Dharmaji、Mohammad Hajiesmaili、Adam Wierman 和 Ramesh K. Sitaraman。Soda：一种用于实现稳定高品质视频流媒体的自适应比特率控制器。载于 ACM SIGCOMM 2024，2024年。[24] Lian Lu、Jinlong Tu、Chi-Kin Chau、Minghua Chen 和 Xiaojun Lin。面向采用间歇性能源和热电联产的微电网的在线能源发电调度。第53–66页。计算机协会，2013年。ISBN 9781450319003。doi: 10.1145/2465529. 2465551。[25] Francesco Zanini、David Atienza、Colin N Jones、Luca Benini 和 Giovanni De Micheli。多处理器系统的在线热控制方法。《ACM电子系统设计自动化汇刊》（TODAES），18(1):1–26，2013年。[26] Neelkamal Bhuyan、Debankur Mukherjee 和 Adam Wierman。最优去中心化平滑在线凸优化。arXiv预印本 arXiv:2411.08355，2024年。[27] Aditya Bhaskara、Ashok Cutkosky、Ravi Kumar 和 Manish Purohit。提示对具有移动成本的在线学习的作用。载于人工智能与统计国际会议，第2818–2826页。PMLR，2021年。[28] Idan Amir、Guy Azov、Tomer Koren 和 Roi Livni。具有切换成本的多臂老虎机的更优双全兼顾界。神经信息处理系统进展，35:15800–15810，2022年。[29] Peng Zhao、Yu-Hu Yan、Yu-Xiang Wang 和 Zhi-Hua Zhou。具有记忆的非平稳在线学习与非随机控制。《机器学习研究期刊》，24(206):1–70，2023年。URL http://jmlr.org/papers/v24/22-0218.html。[30] Aren Karapetyan、Diego Bolliger、Anastasios Tsiamis、Efe C. Balta 和 John Lygeros。具有遗憾保证的在线线性二次跟踪。《IEEE控制系统快报》，2023年。doi: 10.1109/LCSYS.2023.3345809。URL https://arxiv.org/abs/2303.10260。arXiv:2303.10260。[31] J Moore。扩展最小二乘法中的持续激励。《IEEE自动控制汇刊》，28(1):60–68，2003年。[32] Max Simchowitz、Karan Singh 和 Elad Hazan。非随机控制的不适当学习。载于学习理论会议，第 3320–3436 页。PMLR，2020年。[33] Sahin Lale、Kamyar Azizzadenesheli、Babak Hassibi 和 Anima Anandkumar。非线性动力系统中的模型学习预测控制。载于2021年IEEE第60届决策与控制会议（CDC），第757–762页。IEEE，2021年。

[34] Deepan Muthirayan、Jianjun Yuan、Dileep Kalathil 和 Pramod P Khargonekar。具有可证明遗憾保证的预测控制在线学习。载于2022年IEEE第61届决策与控制会议（CDC），第6666–6671页。IEEE，2022年。[35] Mohammad Akbari、Bahman Gharesifard 和 Tamas Linder。通过提示在含噪线性二次调节系统的在线学习中实现对数遗憾。载于2022年IEEE第61届决策与控制会议（CDC），第4700–4705页。IEEE，2022年。[36] Allan Borodin、Nathan Linial 和 Michael E Saks。度量任务系统的最优在线算法。《ACM期刊》，39:745–763，1992年10月。ISSN 0004-5411。doi: 10.1145/146585.146588。URL https://doi.org/10.1145/146585.146588。[37] Avrim Blum 和 Carl Burch。在线学习与度量任务系统问题。《机器学习》，39(1):35–58，2000年。[38] Nikhil Bansal、Anupam Gupta、Ravishankar Krishnaswamy、Kirk Pruhs、Kevin Schewior 和 Cliff Stein。用于具有切换成本的在线凸优化的2-竞争算法。第40卷，第96–109页。Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik，2015年。ISBN 978-3-939897-89-7。doi: 10.4230/LIPIcs.APPROX-RANDOM.2015.96。关键词：随机，调度。[39] Kevin Antoniadis Antonios 和 Schewior。具有切换成本的在线凸优化的紧下界。第164–175页。Springer International Publishing，2018年。ISBN 978-3-319- 89441-6。[40] Gautam Goel、Yiheng Lin、Haoyuan Sun 和 Adam Wierman。超越在线平衡下降：一种用于平滑在线优化的最优算法。神经信息处理系统进展，32，2019年。[41] Iordanis Koutsopoulos 和 Leandros Tassiulas。控制与优化相遇于智能电网：面向最优能源管理的电力需求调度。第41–50页。计算机协会，2011年。ISBN 9781450313131。doi: 10.1145/2318716.2318723。URL https://doi.org/10.1145/2318716.2318723。[42] Rahul Urgaonkar、Bhuvan Urgaonkar、Michael J Neely 和 Anand Sivasubramaniam。利用数据中心中的储存能源实现最优电力成本管理。第221–232页。计算机协会，2011年。ISBN 9781450308144。doi: 10.1145/1993744.1993766。URL https://doi.org/10.1145/1993744.1993766。[43] Otis B Jennings、Avishai Mandelbaum、William A Massey 和 Ward Whitt。满足时变需求的服务器人员配置。《管理科学》，42:1383–1394，1996年。ISSN 00251909, 15265501。URL http://www.jstor.org/stable/2634372。[44] Anshul Gandhi、Varun Gupta、Mor Harchol-Balter 和 Michael A Kozuch。服务器场管理中能源与性能权衡的最优性分析。《性能评估》，67(11):1155–1171，2010年。[45] Vinay Joseph 和 Gustavo de Veciana。联合优化无线系统视频传输的多用户速率自适应：均值、公平性与波动性权衡。第567–575页，2012年。doi: 10.1109/INFCOM.2012.6195799。

[46] Neelkamal Bhuyan、Randeep Bhatia、Murali Kodialam 和 TV Lakshman。通过优先级调度，在云中优化异构需求下竞价实例的节省。在2025 IEEE第26届国际高性能交换与路由会议（HPSR）论文集，第1–6页，2025年。doi: 10.1109/HPSR64165.2025.11038902。[47] Yuxin Chen、Yuejie Chi 和 Andrea J. Goldsmith。通过凸规划从二次采样中进行精确且稳定的协方差估计。《IEEE 信息论汇刊》，61(7):4034–4059，2015年。doi: 10.1109/TIT.2015.2429594。[48] T. Tony Cai 和 Anru Zhang。ROP：通过秩一投影进行矩阵恢复。《统计年鉴》，43(1):102–138，2015年。doi: 10.1214/14-AOS1267。URL https://doi.org/10.1214/14-AOS1267。[49] MOSEK。Mosek 求解器。https://www.mosek.com/。[50] Brendan O’Donoghue。针对线性互补问题齐次嵌入的算子分裂。《SIAM 优化期刊》，31:1999–2023，2021年8月。[51] CVXPY。用于凸优化的开源嵌入 Python 的建模语言 CVXPY。https://www.cvxpy.org/。[52] T Tony Cai 和 Anru Zhang。奇异子空间的速率最优扰动界及其在高维统计中的应用。2018年。[53] Nevena Lazic、Craig Boutilier、Tyler Lu、Eehern Wong、Binz Roy、MK Ryu 和 Greg Imwalle。利用模型预测控制实现数据中心冷却。《神经信息处理系统进展》，31，2018年。[54] Steven Pelley、David Meisner、Thomas F Wenisch、James W VanGilder 等。理解并抽象化数据中心总功耗。在能效设计研讨会上，第11卷，第1–6页，2009年。[55] Qinghui Tang、Sandeep KS Gupta 和 Georgios Varsamopoulos。通过最小化热量回流实现数据中心的热感知任务调度。在2007年IEEE国际集群计算会议论文集，第129–138页。IEEE，2007年。[56] Justin Moore、Jeffrey S Chase 和 Parthasarathy Ranganathan。Weatherman：面向数据中心的自动化在线预测热力映射与管理。在2006年IEEE自主计算国际会议论文集，第155–164页。IEEE，2006年。[57]Richard Evans 和 Jim Gao。DeepMind AI 将 Google 数据中心的冷却费用降低了40%。DeepMind 博客，2016年。https://deepmind.google/discover/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40/。[58]Chris Gamble 和 Jim Gao。面向自主数据中心冷却和工业控制的安全优先人工智能。DeepMind 博客，2018年。https://deepmind.google/blog/safety-first-ai-for-autonomous-data-centre-cooling-and-industrial-control/。[59] Paul Fleming、Jennifer Annoni、J Shah、L Wang、O Anaya-Lara 等。在一座海上风电场进行尾流偏转现场测试。风能科学，2(1)，2017年。

[60] Pieter MO Gebraad、Floris W Teeuwisse、JW Van Wingerden、Paul A Fleming、Shalom D Ruben、Jason R Marden 和 Lucy Y Pao。利用针对尾流效应的参数化模型，通过偏航控制优化风电场功率——一项计算流体力学模拟研究。《风能》，19(1):95–114，2016年。[61] Julija Tastu、Pierre Pinson、Pierre-Julien Trombe 和 Henrik Madsen。考虑地理分散信息的风力发电概率预测。《IEEE智能电网汇刊》，5(1):480–489，2013年。[62] Yan Jiang 和 Guoqing Huang。短期风速预测：集合经验模态分解、特征选择与误差校正的混合方法。《能源转换与管理》，144:340–350，2017年。[63] Niels Otto Jensen。关于风力发电机相互作用的说明。Risø国家实验室，1983年。[64] Majid Bastankhah 和 Fernando Porté-Agel。一种新的风力涡轮机尾流解析模型。《可再生能源》，70:116–123，2014年。[65] Michael F Howland、Sanjiva K Lele 和 John O Dabiri。通过尾流偏转优化风电场功率。《美国国家科学院院刊》（PNAS），116(29)，2019年。[66] Judith Ebegbulem 和 Martin Guay。用于风电场功率最大化的分布式极值寻优控制。IFAC-PapersOnLine，50(1):147–152，2017年。[67] Ervin Bossanyi。用于载荷降低的风力涡轮机控制。《风能》，6(3)，2003年。[68] SK Kanev、FJ Savenije 和 WP Engels。主动尾流控制：一种优化风电场全寿命运行的方法。《风能》，21(7):488–501，2018年。[69] Yair Bartal、Avrim Blum、Carl Burch 和 Andrew Tomkins。一种针对度量任务系统的多对数（n）竞争算法。在第二十九届ACM计算理论研讨会论文集，第711–719页，1997年。[70] Allan Borodin、Nathan Linial 和 Michael E Saks。度量任务系统的最优在线算法。《ACM期刊》，39(4):745–763，1992年10月。ISSN 0004-5411。doi: 10.1145/146585.146588。[71] Susanne Albers。在线算法：综述。《数学规划》，97(1):3–26，2003年。[72] Nicolas Christianson、Junxuan Shen 和 Adam Wierman。学习增强型度量任务系统中鲁棒性—一致性的最优权衡。第206卷，第9377–9399页。PMLR，2023年7月。[73] Zhihao Jiang、Debmalya Panigrahi 和 Kevin Sun。带预测的加权分页在线算法。载于 Artur Czumaj、Anuj Dawar 和 Emanuela Merelli（编），第47届自动机、语言与编程国际研讨会（ICALP 2020），莱布尼茨国际信息学论文集第168卷，第69:1–69:18页，德国达格施图尔，2020年。Schloss Dagstuhl–Leibniz-Zentrum für Informatik。ISBN 978-3-95977-138-2。doi: 10.4230/LIPIcs.ICALP.2020.69。URL https://drops.dagstuhl.de/opus/volltexte/2020/12476。[74] Nikhil Bansal、Christian Coester、Ravi Kumar、Manish Purohit 和 Erik Vee。尺度无关分配、摊销凸性与短视加权分页。CoRR，abs/2011.0，2020年。URL https://arxiv.org/abs/2011.09076。

[75] Amos Fiat、Yuval Rabani 和 Yiftach Ravid。Competitive k 服务器算法。《计算机与系统科学杂志》，48(3):410–428，1994年。[76] Elias Koutsoupias 和 Christos H Papadimitriou。关于 k 服务器猜想。《ACM期刊》，42:971–983，1995年9月。ISSN 0004-5411。doi：10.1145/210118.210128。URL https://doi.org/10.1145/210118.210128。[77] Thodoris Lykouris 和 Sergei Vassilvtiskii。基于机器学习建议的竞争性缓存。载于 ICML '18 论文集，第3296–3305页。PMLR，2018年。[78] Thodoris Lykouris 和 Sergei Vassilvitskii。基于机器学习建议的竞争性缓存。《ACM期刊》，68(4)，2021年。ISSN 0004-5411。doi：10.1145/3447579。URL https://doi.org/10.1145/3447579。[79] Antonios Antoniadis、Themis Gouleakis、Pieter Kleer 和 Pavel Kolev。带有机器学习建议的秘书问题与在线匹配问题。载于 H Larochelle、M Ranzato、R Hadsell、M F Balcan 和 H Lin（编），《神经信息处理系统进展》，第33卷，第7933–7944页。Curran Associates, Inc.，2020年。URL
h
t
t
p
s://proceedings.neurips.cc/paper/2020/file/5a378f8490c8d6af8647a753812f6e31-Paper.pdf。[80] Mark Sellke。最优追逐凸体。载于 SODA '20 论文集，{{SODA}} '20，第1509–1518页，美国，2020年1月。工业与应用数学学会。[81] C J Argue、Anupam Gupta、Ziye Tang 和 Guru Guruganesh。具有线性竞争比的追逐凸体。《ACM期刊》，68(5):1–10，2021年。doi：10.1145/3450349。[82] Neelkamal Bhuyan、Debankur Mukherjee 和 Adam Wierman。估计后决策：矩阵补全驱动的平滑在线二次优化。载于神经信息处理系统大会 2025年研讨会 MLxOR：面向不确定性感知决策的机器学习数学基础与运营集成，2025年。URL https://openreview.net/forum?id=CHglLRbEGO。[83]Yiheng Lin、Yang Hu、Guanya Shi、Haoyuan Sun、Guannan Qu 和 Adam Wierman。线性时变系统中预测控制的基于扰动的遗憾分析。《神经信息处理系统进展》，34:5174–5185，2021年。[84] Yiheng Lin、Judy Gan、Guannan Qu、Yash Kanoria 和 Adam Wierman。网络化系统中的分布式在线凸优化。载于国际机器学习会议，第13356–13393页。PMLR，2022年。[85] Niangjun Chen、Joshua Comden、Zhenhua Liu、Anshul Gandhi 和 Adam Wierman。利用预测进行在线优化：以前瞻之眼回望过去。第193–206页。计算机协会，2016年。ISBN 9781450342667。doi：10.1145/2896377.2901464。[86] Spandan Senapati 和 Rahul Vaze。带有切换成本和延迟梯度的在线凸优化。《性能评估》，162:102371，2023年。[87] Harsh Shah、Purna Chandrasekhar 和 Rahul Vaze。仅进行一次梯度评估的带切换成本在线凸优化。arXiv 预印本 arXiv:2507.04133，2025年。[88] Sahin Lale、Kamyar Azizzadenesheli、Babak Hassibi 和 Anima Anandkumar。自回归外生系统中的有限时间系统辨识与自适应控制。载于《

第三届动力学与控制学习会议论文集》，《机器学习研究论文集》第144卷，第967–979页。PMLR，2021年6月7日至8日。URL https://proceedings.mlr.press/v144/lale21b.html。[89] Nicolò Cesa-Bianchi、Gábor Lugosi 和 Gergely Neu。在线优化中的多臂老虎机信息代价。载于《神经信息处理系统进展》，第1–9页，2006年。URL
h
t
t
p://
p
apers.neurips.cc/paper/3371-the-price-of-bandit-infor
mation-for-online-optimization.pdf。[90] Tor Lattimore 和 Csaba Szepesvári。多臂老虎机算法。剑桥大学出版社，2020年。[91] Asaf Cassel 和 Tomer Koren。多臂老虎机线性控制。《神经信息处理系统进展》，33:8872–8882，2020年。[92] M Cody Priess、Richard Conway、Jongeun Choi、John M Popovich 和 Clark Radcliffe。逆线性二次调节器问题的解及其在生物系统分析中的应用。《IEEE控制系统技术汇刊》，23(2):770–777，2014年。[93] Haitham El-Hussieny 和 Jee-Hwan Ryu。用于学习人类运动行为的基于折扣的逆线性二次调节器算法。《应用智能》，49(4):1489–1501，2019年。[94] Huai-Ning Wu 和 Mi Wang。通过积分并发自适应逆强化学习进行人在回路行为建模。《IEEE神经网络与学习系统汇刊》，35(8):11359–11370，2023年。[95] Wen-Hua Li 和 Huai-Ning Wu。通过移动时域估计和状态依赖黎卡提方程，对一类含噪离散时间非线性人在回路系统进行人类行为学习。SSRN 5245076，2025年。[96] Zhongshi Sun 和 Guangyan Jia。通过专家模仿，为随机线性二次最优控制问题进行逆强化学习。《神经计算》，633:129758，2025年。[97] Shicheng Liu 和 Minghui Zhu。轨迹内逆强化学习：在进行中的轨迹终止前逐步学习。载于《神经信息处理系统进展》（NeurIPS），2025年。URL https://arxiv.org/abs/2410.15612。

A 附加数值实验

我们广泛测试了所提出的算法 SCaLE 和 HySCaLE 的性能，并将其与三种基线方法进行比较：(i) 跟随最小化器（FTM）、(ii) 被动在线学习器（POL）以及 (iii) 预言机辅助学习器（OAL）。本节将介绍我们的全部研究结果。

A.1 非相关轻尾过程：最小化器差分 $\delta_t = v_{t-1} - v_t$ 服从独立同分布高斯分布

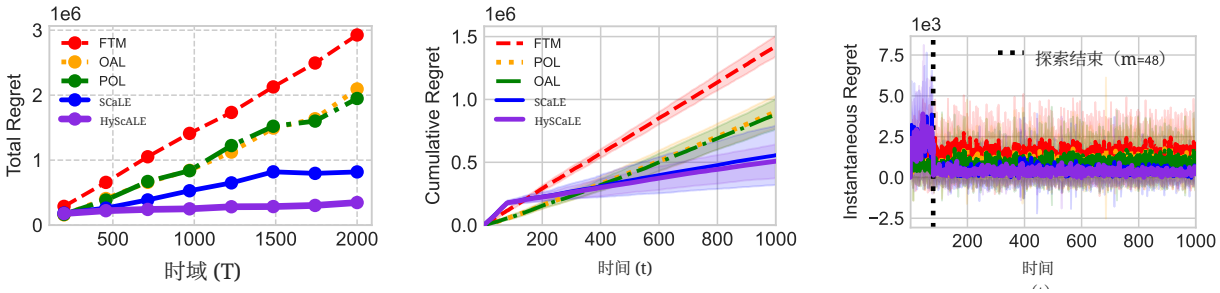


图 7：Regret$_T$、Regret$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，对于 $r = d = 4$，$\sigma_r^A = 1$，$\sigma_v = 50$，$\bar{\eta} = 10$ 以及将 c_1 设为 3

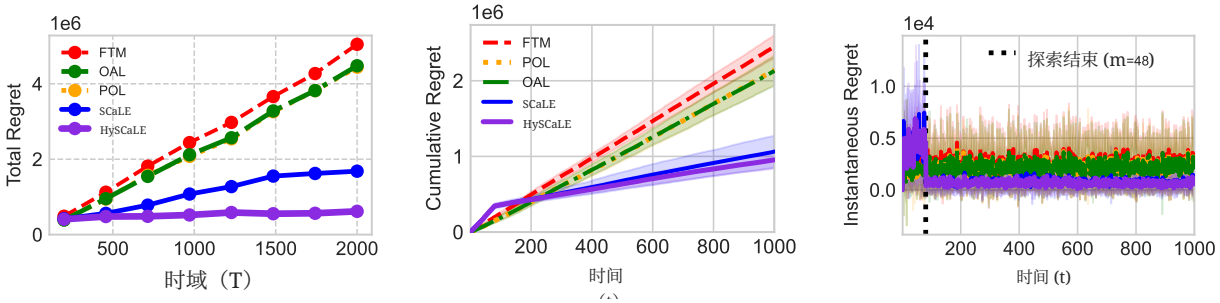


图 8：Regret$_T$、Regret$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，对于 $r = d = 4$，$\sigma_r^A = 10^{-2}$，$\sigma_v = 50$，$\bar{\eta} = 10$ 以及将 c_1 设为 3

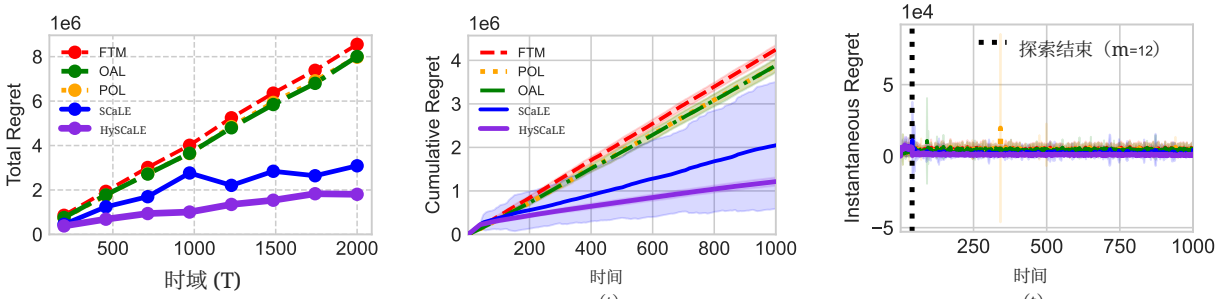


图 9：Regret$_T$、Regret$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，对于 $r = 1$，$d = 4$，$\sigma_r^A = 1$，$\sigma_v = 50$，$\bar{\eta} = 10$ 以及将 c_1 设为 3

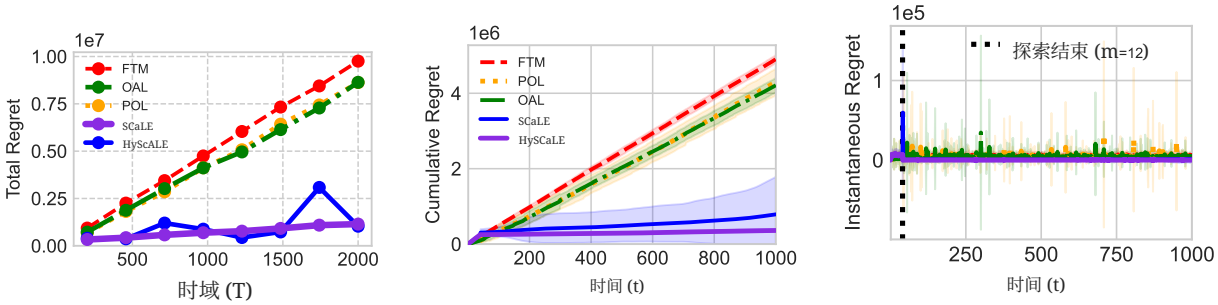


图10：遗憾$_T$、遗憾$_{1000}(t)$和Δ遗憾$_{1000}(t)$ Δt，当 $r=1, d=4, \sigma_r^A=10^{-2}, \sigma_v=50, \bar{\eta}=10$和 c_1 设为3时

A.2 相关 d 轻尾过程：非独立同分布增量 $\delta_t = 0.3\delta_{t-1} + \mathcal{N}(\mathbf{0}, \sigma_v^2 I)$

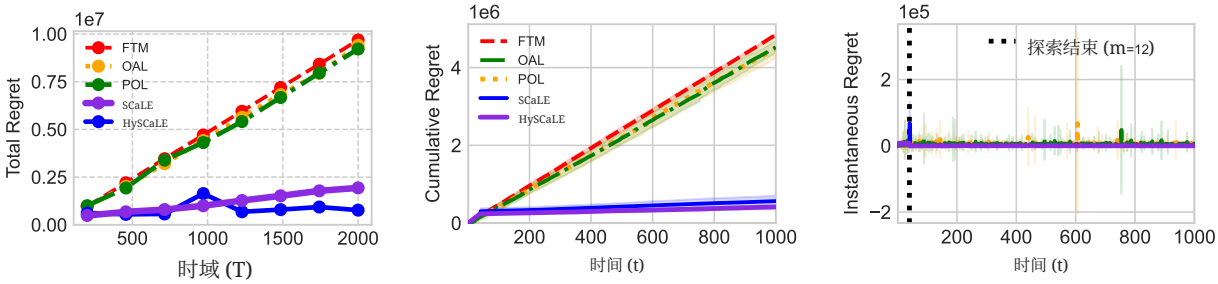


图11：遗憾$_T$、遗憾$_{1000}(t)$和Δ遗憾$_{1000}(t)$ Δt，当 $r=1, d=4, \sigma_r^A=10^{-2}, \sigma_v=50, \bar{\eta}=10$和 c_1 设为3时

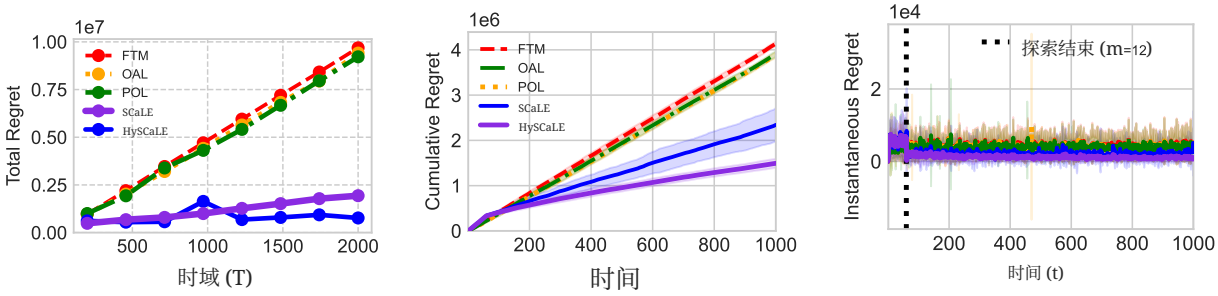


图12：遗憾$_T$、遗憾$_{1000}(t)$和Δ遗憾$_{1000}(t)$ Δt，当 $r=1, d=4, \sigma_r^A=1, \sigma_v=50, \bar{\eta}=10$和 c_1设为3时

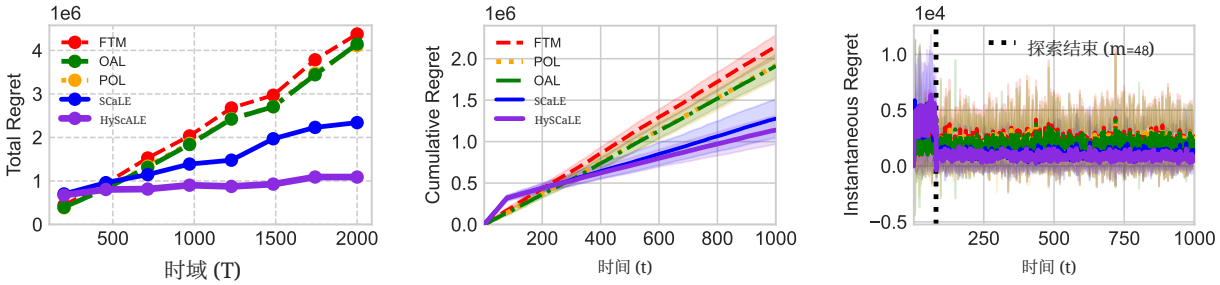


图13：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，当 $r = d = 4, \sigma_r^A = 10^{-2}, \sigma_v = 50, \bar{\eta} = 10$ 和c_1 设为 3 时

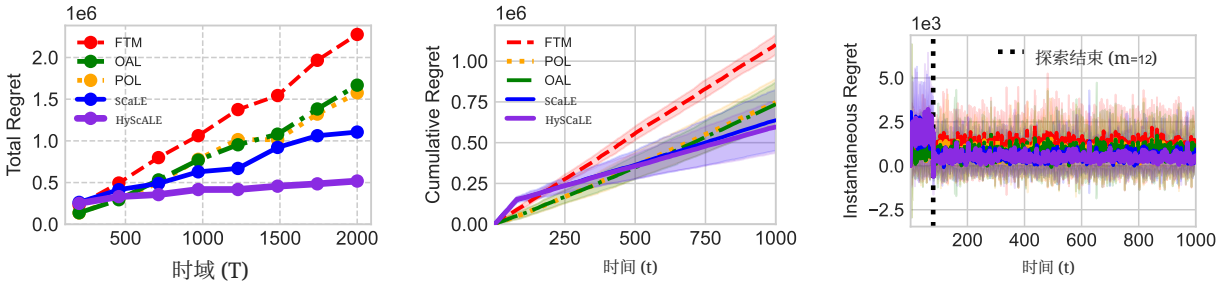


图14：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，当 $r = d = 4, \sigma_r^A = 1, \sigma_v = 50, \bar{\eta} = 10$ 和 c_1设为 3 时

A.3 厚尾：柯西分布

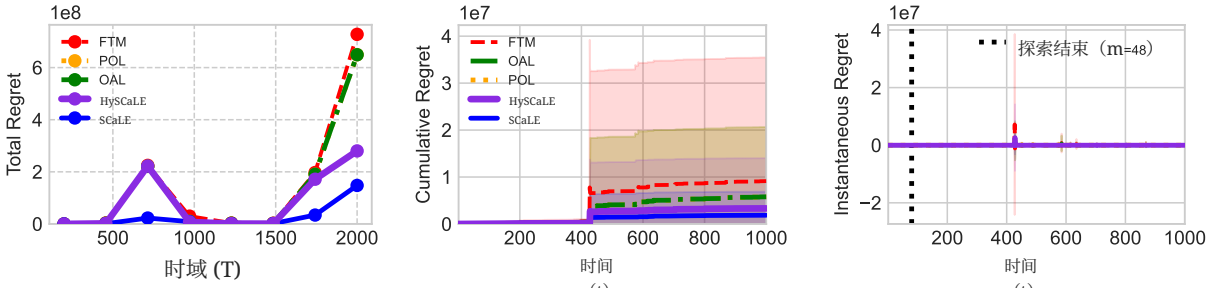


图15：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，当 $r = 4, d = 4, \sigma_r^A = 1, \sigma_v = 1, \bar{\eta} = 1$ 和 c_1设为 5 时

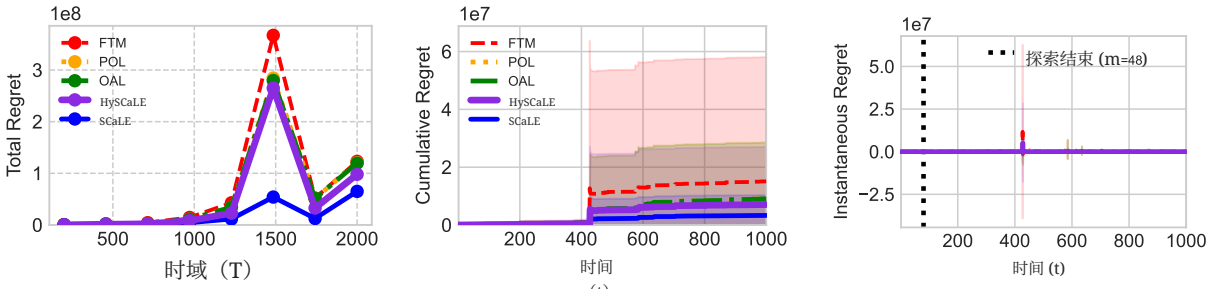


图16：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，其中 $r = 4,\ d = 4,\ \sigma_r^A = 10^{-2},\ \sigma_v = 1,\ \bar{\eta} = 1$ 和 c_1 均设为 5

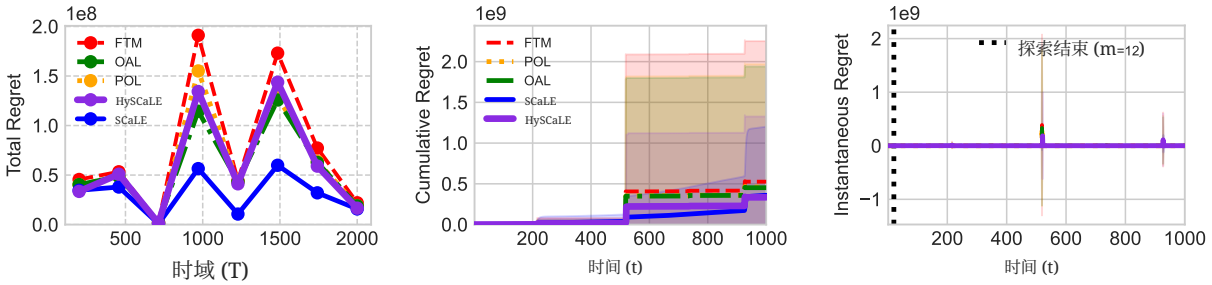


图 17：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，其中 $r = 1, d = 4, \sigma_r^A = 1, \sigma_v = 1, \bar{\eta} = 1$ 和 c_1设为 5

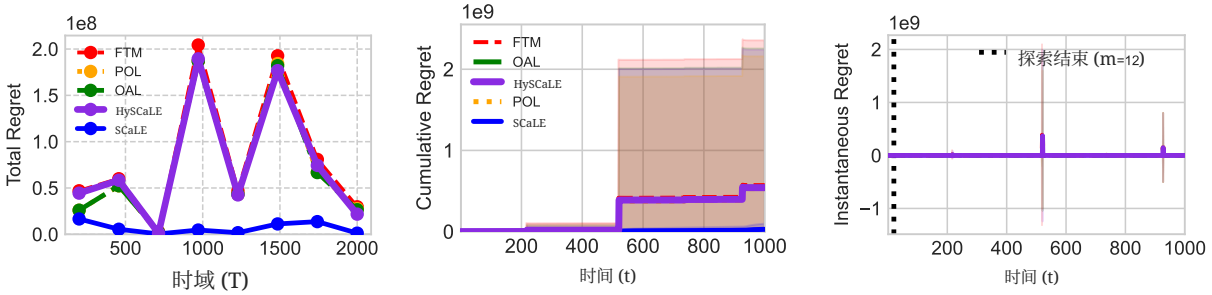


图 18：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，其中 $r = 1, d = 4, \sigma_r^A = 10^{-2}, \sigma_v = 1, \bar{\eta} = 1$ 和c_1 设为 5

A.4 重尾：拉普拉斯分布

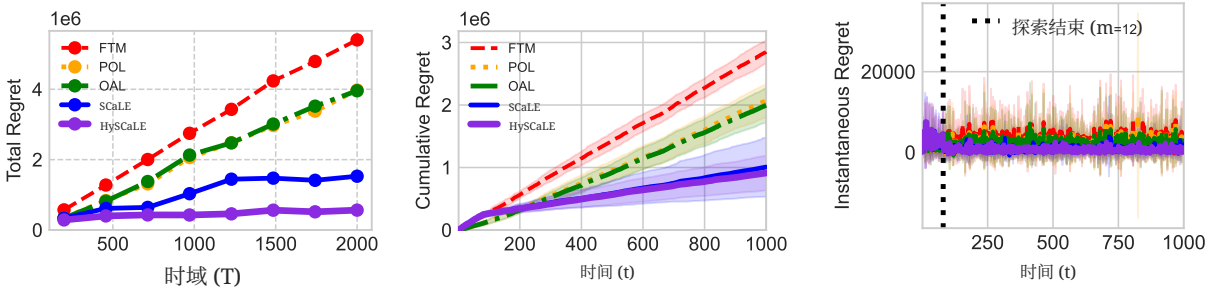


图19：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$ Δt，其中 $r = 4, d = 4, \sigma_r^A = 1, \sigma_v = 50, \bar{\eta} = 1$ 和 c_1设为 5

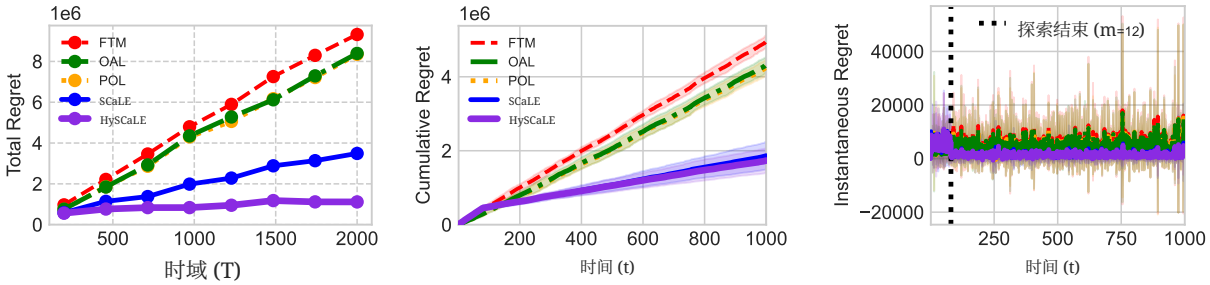


图 20：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$$\Delta t$，当 $r = 4, d = 4, \sigma_r^A = 10^{-2}, \sigma_v = 50, \bar{\eta} = 1$ 和c_1 设为 5 时

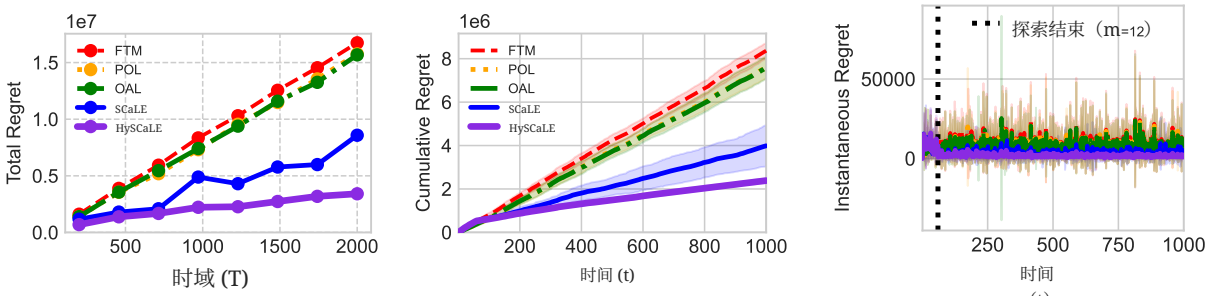


图 21：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$$\Delta t$，当 $r = 1, d = 4, \sigma_r^A = 1, \sigma_v = 50, \bar{\eta} = 1$ 和 c_1设为 5 时

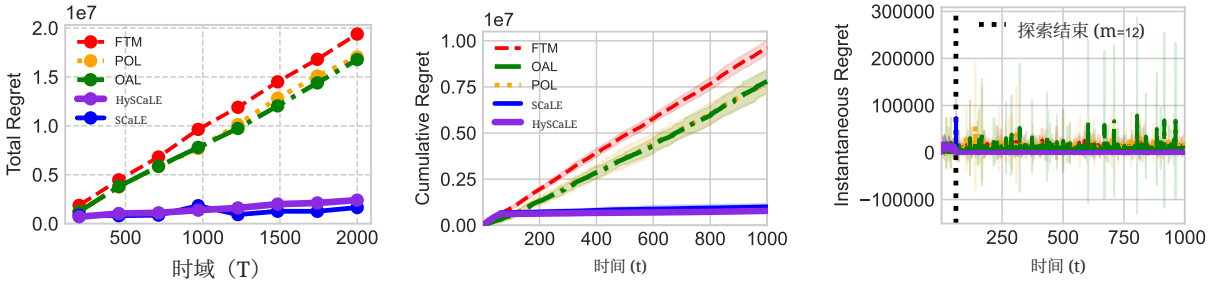


图 22：遗憾$_T$、遗憾$_{1000}(t)$ 和 Δ遗憾$_{1000}(t)$$\Delta t$，当 $r = 1, d = 4, \sigma_r^A = 10^{-2}, \sigma_v = 50, \bar{\eta} = 1$ 和c_1 设为 5 时

B 矩阵扰动基础知识

考虑任意一个 $A = QDQ^T$及其估计值 $\hat{A} = P\hat{D}P^T$.。 A和 $\hat{A}$均为半正定矩阵。不失一般性， $\lambda^1, A \geq \ldots \lambda^d, A$且 $\hat{\lambda}^1 \geq \ldots \hat{\lambda}^d$.。我们假设 $\|A - \hat{A}\|_* \leq \epsilon$.。奇异值的 Weyl 不等式表明

$$|\hat{\lambda}^i - \lambda^{i,A}| \leq \|\hat{A} - A\|_{op} \leq \|\hat{A} - A\|_* \leq \epsilon$$

由此可得

$$\hat{\lambda}^i \in [\lambda^{i,A} - \epsilon, \lambda^{i,A} + \epsilon]$$

对于 $i \in \{1, \ldots, r\}$ 和

$$\hat{\lambda}^j \in [0, \epsilon]$$

对于 $j > r$，当 $\lambda^{j,A} = 0$ 且 $\hat{A} \succeq 0$ 接下来，我们利用以下事实 $\|\hat{A} - A\|_F \leq \|\hat{A} - A\|_* / \sqrt{d}$ 来证明

$$|\hat{A}_{i,j} - A_{i,j}| \leq \epsilon.$$

对于 $A = A + Z$（其中 $Z = A - A$），根据 [52], 我们有

$$\frac{1}{\sqrt{2}}\|P - Q\|_{op} \leq \|\sin\Theta(P, Q)\|_{op} \leq \frac{\alpha z_{12} + \beta z_{21}}{\alpha^2 - \beta^2 - \min(z_{12}^2, z_{21}^2)}$$

其中 $z_{ij} \leq \|Z\|_F$、 $\alpha = \sigma_{\min}^A - \epsilon$ 和 $\beta \leq \epsilon$. 这意味着

$$\|P - Q\|_{op} \lessapprox \frac{\sqrt{2} \cdot \epsilon}{\sigma_{\min}^A - \epsilon}$$

C 矩阵扰动驱动的遗憾分析框架

C.1 命题 1 的证明

在第 T 轮，近似策略的成本为

$$\begin{aligned}
\text{Cost}[T, T] &= \frac{1}{2}\|x_T - v_T\|_A^2 + \frac{1}{2}\|x_T - x_{T-1}\|_2^2 \\
&= \frac{1}{2}\|x_{T-1} - v_T\|_{\hat{C}_T A \hat{C}_T}^2 + \frac{1}{2}\|v_T - x_{T-1}\|_{(I-\hat{C}_T)^2}^2 \\
&= \frac{1}{2}\|x_{T-1} - v_T\|_{\hat{C}_T \hat{A} \hat{C}_T}^2 + \frac{1}{2}\|x_T - x_{T-1}\|_{(I-\hat{C}_T)^2}^2 + \frac{1}{2}\|x_{T-1} - v_T\|_{\hat{C}_T(A-\hat{A})\hat{C}_T}^2 \\
&= \frac{1}{2}\|x_{T-1} - v_T\|_{I-\hat{C}_T}^2 + \underbrace{\frac{1}{2}\|x_{T-1} - v_T\|_{\hat{C}_T(A-\hat{A})\hat{C}_T}^2}_{\Delta_T}
\end{aligned}$$

对 $\mathcal{F}_{T-1}$ 取期望，得到

$$\mathbb{E}[\text{Cost}[T, T]|\mathcal{F}_{T-1}] \leq \frac{1}{2}\|x_{T-1} - v_{T-1}\|_{I-\hat{C}_T}^2 + \frac{1}{2}\mathbb{E}\|v_{T-1} - v_T\|_{I-\hat{C}_T}^2 + \mathbb{E}[\Delta_T|\mathcal{F}_{T-1}].$$

我们断言

$$\mathbb{E}[\text{Cost}[T - t + 1, T]|\mathcal{F}_{T-t}] \leq \frac{1}{2}\|x_{T-t} - v_{T-t}\|_{I-\hat{C}_{T-t+1}}^2 + \sum_{s=0}^{t-1}\frac{1}{2}\mathbb{E}\|v_{T-s-1} - v_{T-s}\|_{I-\hat{C}_{T-s}}^2 + \sum_{s=0}^{t-1}\mathbb{E}[\Delta_{T-s}|\mathcal{F}_{T-t}].$$

上述结论对 $t = 1$. 成立。假设它对某个 $t \geq 1$成立，则

$$\frac{1}{2}\|x_{T-t} - v_{T-t}\|_A^2 + \frac{1}{2}\|x_{T-t} - x_{T-t-1}\|_2^2 + \frac{1}{2}\|x_{T-t} - v_{T-t}\|_{I-\hat{C}_{T-t+1}}^2$$

$$= \frac{1}{2}\|x_{T-t-1} - v_{T-t}\|^2_{\hat{C}_{T-t}(I+A-\hat{C}_{T-t})\hat{C}_{T-t}} + \frac{1}{2}\|v_{T-t} - x_{T-t-1}\|^2_{(I-\hat{C}_{T-t})^2}$$
$$= \frac{1}{2}\|x_{T-t-1} - v_{T-t}\|^2_{\hat{C}_{T-t}(I+\hat{A}-\hat{C}_{T-t})\hat{C}_{T-t}} + \frac{1}{2}\|v_{T-t} - x_{T-t-1}\|^2_{(I-\hat{C}_{T-t})^2} + \frac{1}{2}\|x_{T-t-1} - v_{T-t}\|^2_{\hat{C}_{T-t}(A-\hat{A})\hat{C}_{T-t}}$$

$$= \frac{1}{2}\|x_{T-t-1} - v_{T-t}\|^2_{I-\hat{C}_{T-t}} + \underbrace{\frac{1}{2}\|x_{T-t-1} - v_{T-t}\|^2_{\hat{C}_{T-t}(A-\hat{A})\hat{C}_{T-t}}}_{\Delta_{T-t}}$$

由此可得

$$\begin{aligned}
&\mathbb{E}[\mathrm{Cost}[T-t,T]|\mathcal{F}_{T-t-1}] \\
&= \mathbb{E}\left[\frac{1}{2}\|x_{T-t} - v_{T-t}\|^2_A + \frac{1}{2}\|x_{T-t} - x_{T-t-1}\|^2_2 \middle| \mathcal{F}_{T-t-1}\right] + \mathbb{E}[\mathrm{Cost}[T-t+1,T]|\mathcal{F}_{T-t-1}] \\
&= \frac{1}{2}\|x_{T-t-1} - v_{T-t-1}\|^2_{I-\hat{C}_{T-t}} + \sum_{s=0}^{t}\frac{1}{2}\mathbb{E}\|v_{T-s-1} - v_{T-s}\|^2_{I-\hat{C}_{T-s}} + \sum_{s=0}^{t}\mathbb{E}[\Delta_{T-s}|\mathcal{F}_{T-t-1}]
\end{aligned}$$

证明　通过归纳法证明我们的结论。因此，我们得到总成本的上界为

$$\mathbb{E}[\mathrm{Cost}[1,T]|\mathcal{F}_0] = \sum_{s=1}^{T}\frac{1}{2}\mathbb{E}\|v_s - v_{s-1}\|^2_{I-\hat{C}_s} + \sum_{k=1}^{T}\mathbb{E}[\Delta_k|\mathcal{F}_0]$$

这给出的遗憾等于

$$\mathrm{Regret}_T = \sum_{s=1}^{T}\frac{1}{2}\mathbb{E}\|v_s - v_{s-1}\|^2_{C_s-\hat{C}_s} + \sum_{k=1}^{T}\frac{1}{2}\mathbb{E}\|x_{k-1} - v_k\|^2_{\hat{C}_k(A-\hat{A})\hat{C}_k}$$

C.2 定理5的证明

该算法为

$$x_t = \hat{C}_t x_{t-1} + (I - \hat{C}_t)v_t$$

其中，在第 t 轮，玩家进行估计

$$\hat{C}_\tau^{-1} = 2I + \hat{A} - \hat{C}_{\tau+1}$$

对于 $\tau \geq t$. 与DP序列之间的差距:

$$C_t^{-1} = 2I + A - C_{t+1}$$

可以量化为

$$\hat{C}_t - C_t = \sum_{\tau=t}^{T}\prod_{m=t}^{\tau} C_m(A-\hat{A})\prod_{m=\tau}^{t}\hat{C}_m$$

对于 $\sigma^A_{\min} > 0$ ，可以量化为

$$\|\hat{C}_t - C_t\|_{op} \leq \frac{\|A-\hat{A}\|_{op}}{\sigma^A_{\min}} \leq \frac{\epsilon}{\sigma^A_{\min}}$$

回想一下，遗憾恰好等于

$$\text{Regret}_T = \underbrace{\sum_{s=1}^{T} \frac{1}{2}\mathbb{E}\|v_s - v_{s-1}\|^2_{C_s - \hat{C}_s}}_{R_1} + \underbrace{\sum_{s=1}^{T} \frac{1}{2}\mathbb{E}\|x_{s-1} - v_s\|^2_{\hat{C}_s(A-\hat{A})\hat{C}_s}}_{R_2}$$

对于 R_1，我们有

$$\begin{aligned} R_1 &\leq \sum_{s=1}^{T} \frac{\epsilon}{2\sigma^A_{\min}} \mathbb{E}\|v_s - v_{s-1}\|_2^2 \\ &\leq \frac{tr(\Sigma)}{2\sigma^A_{\min}} \cdot \epsilon T \end{aligned}$$

对于 R_2，我们考察每一项

$$\mathbb{E}\|x_{s-1} - v_s\|^2_{\hat{C}_s(A-\hat{A})\hat{C}_s} = \sum_{k=0}^{s-1} \mathbb{E}\|v_{s-k-1} - v_{s-k}\|^2_{\prod_{l=k}^{0} \hat{C}_{s-l}(A-\hat{A}) \prod_{l=0}^{k} \hat{C}_{s-l}}.$$

我们考察

$$\begin{aligned} &\prod_{l=k}^{0} \hat{C}_{s-l}(A-\hat{A}) \prod_{l=0}^{k} \hat{C}_{s-l} \\ =&P \times diag\left(\prod_{l=k}^{0} \hat{\lambda}^1_{s-l}, \ldots, \prod_{l=k}^{0} \hat{\lambda}^d_{s-l}\right) \times \underbrace{(P^T Q D Q^T P - \hat{D})}_{P^T Q D Q^T P - D + D - \hat{D}} \times diag\left(\prod_{l=k}^{0} \hat{\lambda}^1_{s-l}, \ldots, \prod_{l=k}^{0} \hat{\lambda}^d_{s-l}\right) \times P^T. \\ =&\underbrace{P \times diag\left(\prod_{l=k}^{0} \hat{\lambda}^1_{s-l}, \ldots, \prod_{l=k}^{0} \hat{\lambda}^d_{s-l}\right) \times \underbrace{(P^T Q D Q^T P - D)}_{\Delta_2} \times diag\left(\prod_{l=k}^{0} \hat{\lambda}^1_{s-l}, \ldots, \prod_{l=k}^{0} \hat{\lambda}^d_{s-l}\right) P^T}_{X_1} \\ &\underbrace{+ P \times diag\left(\prod_{l=k}^{0} \hat{\lambda}^1_{s-l}, \ldots, \prod_{l=k}^{0} \hat{\lambda}^d_{s-l}\right) \times (D - \hat{D}) \times diag\left(\prod_{l=k}^{0} \hat{\lambda}^1_{s-l}, \ldots, \prod_{l=k}^{0} \hat{\lambda}^d_{s-l}\right) P^T}_{X_2}. \end{aligned}$$

我们先看 X_2：

$$\begin{aligned} D - \hat{D} &= diag(\lambda^{A,1} - \hat{\lambda}^1, \ldots, \lambda^{A,d} - \hat{\lambda}^d) \\ &\preceq \epsilon I_{d \times d} \end{aligned}$$

此外，以下结论对每个 i 都成立

$$\hat{\lambda}^i_t \leq \frac{1}{1 + \lambda^{A,i} - \epsilon}$$

这将为我们带来

$$diag\left(\prod_{l=k}^{0} \hat{\lambda}^1_{s-l}, \ldots, \prod_{l=k}^{0} \hat{\lambda}^d_{s-l}\right) \times (D - \hat{D}) \times diag\left(\prod_{l=k}^{0} \hat{\lambda}^1_{s-l}, \ldots, \prod_{l=k}^{0} \hat{\lambda}^d_{s-l}\right)$$

$$\preceq \underbrace{diag\left(\frac{\epsilon}{(1+\lambda^{A,1}-\epsilon)^{2k+2}},\ldots,\frac{\epsilon}{(1+\lambda^{A,r}-\epsilon)^{2k+2}},0,\ldots,0\right)}_{\Delta_1}$$

从而导致

$$\begin{aligned}
\mathbb{E}\|v_{s-k-1}-v_{s-k}\|_{X_2}^2 &= \langle X_2,\Sigma\rangle \\
&= \langle \Delta_1, P^T\Sigma P\rangle \\
&= \sum_{i=1}^{r}\frac{\epsilon}{(1+\lambda^{A,i}-\epsilon)^{2k+2}}p_i^T\Sigma p_i \\
&\le \frac{\epsilon}{(1+\sigma_{\min}^A-\epsilon)^{2k+2}}tr(P^T\Sigma P) \\
&= \epsilon\cdot\frac{tr(\Sigma)}{(1+\sigma_{\min}^A-\epsilon)^{2k+2}}
\end{aligned}$$

现在我们来看 X_1，更具体地说，是 Δ_2:

$$\begin{aligned}
\Delta_2 &= P^TQDQ^TP-D \\
&= (P^TQ-I+I)D(Q^TP-I+I)-D \\
&= (P^TQ-I)D(Q^TP-I)+(P^TQ-I)D+D(Q^TP-I)+D-D \\
&= \underbrace{(P^TQ-I)D(Q^TP-I)}_{\Delta_{2a}}+\underbrace{(P^TQ-I)D}_{\Delta_{2b}}+\underbrace{D(Q^TP-I)}_{\Delta_{2c}}
\end{aligned}$$

现在，注意到

$$\begin{aligned}
\Delta_{3b} &= diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right)\times\Delta_{2b}\times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right) \\
&= diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right)\times(P^TQ-I)\times diag\left(\lambda^{A,1}\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\lambda^{A,d}\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right)
\end{aligned}$$

and

$$\begin{aligned}
\Delta_{3c} &= diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right)\times\Delta_{2c}\times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right) \\
&= diag\left(\lambda^{A,1}\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\lambda^{A,d}\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right)\times(P^TQ-I)diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right).
\end{aligned}$$

最后， Δ_{3a} 相应地定义为,

$$\Delta_{3a} = diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right)\times\Delta_{2a}\times diag\left(\prod_{l=k}^{0}\hat{\lambda}_{s-l}^1,\ldots,\prod_{l=k}^{0}\hat{\lambda}_{s-l}^d\right).$$

现在把这三者放在一起

$$\begin{aligned}
\mathbb{E}\|v_{s-k-1}-v_{s-k}\|_{X_1}^2 &= \langle\Delta_{3a},\mathbb{P}^T\Sigma P\rangle+\langle\Delta_{3b},\mathbb{P}^T\Sigma P\rangle+\langle\Delta_{3c},P^T\Sigma P\rangle \\
&\le \|\Sigma\|_F\cdot(\|\Delta_{3a}\|_F+\|\Delta_{3b}\|_F+\|\Delta_{3c}\|_F)
\end{aligned}$$

现在，$\|P^TQ - I\|_F = \|Q^TP - I\|_F = \|P - Q\|_F \le \sqrt{d} \cdot \|P - Q\|_{op} \le \frac{\sqrt{2d}\cdot\epsilon}{\sigma^A_{\min} - \epsilon}$。因此，我们有

$$\|\Delta_{3b}\|_F = \|\Delta_{3c}\|_F \le \frac{\sigma^A_{\max}\sqrt{2d}\cdot\epsilon}{(\sigma^A_{\min} - \epsilon)(1 + \sigma^A_{\min} - \epsilon)^{2k+2}}$$

and

$$\begin{aligned}
\|\Delta_{3a}\|_F &\le \frac{1}{(1 + \sigma^A_{\min} - \epsilon)^{2k+2}} \|\Delta_{2a}\|_F \\
&\le \frac{1}{(1 + \sigma^A_{\min} - \epsilon)^{2k+2}} \|P^TQ - I\|_F \cdot \|D(Q^TP - I)\|_F \\
&\le \frac{1}{(1 + \sigma^A_{\min} - \epsilon)^{2k+2}} \frac{2d\sigma^A_{\max}\cdot\epsilon^2}{(\sigma^A_{\min} - \epsilon)^2}.
\end{aligned}$$

现在将 X_1 和 X_2 结合起来,

$$\begin{aligned}
&\mathbb{E}\|x_{s-1} - v_s\|^2_{\hat{C}_s(A-\hat{A})\hat{C}_s} \\
&\le \sum_{k=0}^{s-1} \frac{2d\sigma^A_{\max}\|\Sigma\|_F\cdot\epsilon^2}{(\sigma^A_{\min} - \epsilon)^2(1 + \sigma^A_{\min} - \epsilon)^{2k+2}} + 2\sum_{k=0}^{s-1} \frac{\|\Sigma\|_F\sigma^A_{\max}\sqrt{2d}\cdot\epsilon}{(\sigma^A_{\min} - \epsilon)(1 + \sigma^A_{\min} - \epsilon)^{2k+2}} + \sum_{k=0}^{s-1} \epsilon \cdot \frac{tr(\Sigma)}{(1 + \sigma^A_{\min} - \epsilon)^{2k+2}} \\
&\le \frac{1}{(1 + \sigma^A_{\min} - \epsilon)^2 - 1} \left(\frac{2d\sigma^A_{\max}\|\Sigma\|_F}{(\sigma^A_{\min} - \epsilon)^2} \cdot \epsilon^2 + \frac{2\|\Sigma\|_F\sigma^A_{\max}\sqrt{2d}\cdot\epsilon}{\sigma^A_{\min} - \epsilon} + \frac{tr(\Sigma)\epsilon}{(1 + \sigma^A_{\min} - \epsilon)^2 - 1} \right)
\end{aligned}$$

这会导致遗憾为

$$\begin{aligned}
\mathrm{Regret}[1, T] \le &\frac{\epsilon T}{(1 + \sigma^A_{\min} - \epsilon)^2 - 1} \left(\frac{2d\sigma^A_{\max}\|\Sigma\|_F}{(\sigma^A_{\min} - \epsilon)^2} \cdot \epsilon + \frac{2\|\Sigma\|_F\sigma^A_{\max}\sqrt{2d}}{\sigma^A_{\min} - \epsilon} + \frac{tr(\Sigma)}{(1 + \sigma^A_{\min} - \epsilon)^2 - 1} \right) \\
&+ \frac{tr(\Sigma)}{2\sigma^A_{\min}} \cdot \epsilon T.
\end{aligned}$$

C.3 定理3的证明

现在，若前 m 轮次通过动作序列进行数据收集 ce:

$$x_t = v_t + \gamma \cdot z_t$$

使用 $z_t \sim \mathcal{N}(\mathbf{0}, I)$，，该流程所涉及的成本为

$$\begin{aligned}
&\sum_{t=1}^{m} \frac{1}{2}\mathbb{E}\|x_t - v_t\|^2_A + \frac{1}{2}\mathbb{E}\|x_t - x_{t-1}\|^2_2 \\
&\quad= \sum_{t=1}^{m} \frac{\gamma^2}{2}\mathbb{E}[z_t^T A z_t] + \frac{1}{2}\mathbb{E}\|v_t + \gamma z_t - v_{t-1} - \gamma z_{t-1}\|^2_2 \\
&\quad= \sum_{t=1}^{m} \frac{\gamma^2}{2}\mathbb{E}\langle A, z_t z_t^T\rangle + \frac{1}{2}\mathbb{E}\|v_t - v_{t-1}\|^2_2 + \frac{\gamma^2}{2}\mathbb{E}\|z_t\|^2_2 + \frac{\gamma^2}{2}\mathbb{E}\|z_{t-1}\|^2_2 \\
&\quad= \sum_{t=1}^{m} \frac{\gamma^2}{2}\langle A, I\rangle + (1/2)tr(\Sigma) + \gamma^2 d
\end{aligned}$$

$$= m \cdot (1/2)tr(\Sigma) + m\gamma^2 \cdot ((1/2)tr(A) + d)$$

玩家以以下形式使用这些测量：

$$\begin{aligned} y_t &= \frac{\gamma^2}{2} z_t^T A z_t + \eta_t \\ \frac{y_t}{\gamma^2} &= \frac{1}{2} z_t^T A z_t + \frac{\eta_t}{\gamma^2} \\ y_t' &= \frac{1}{2}\langle A, z_t z_t^T \rangle + \eta_t' \end{aligned}$$

其中 总测量误差之和为 $\epsilon_m = \sum_{t=1}^m |\eta_t'| = \frac{\|\boldsymbol{\eta}_{1:m}\|_1}{\gamma^2}$。解出 ———

$$\hat{A} = \underset{\substack{M \succeq 0 \\ \|Y' - \mathcal{A}(M)\|_1 \le \epsilon_m}}{\operatorname{argmin}} Tr(M)$$

我们可以保证，以概率 $1 - \exp(-C_0 m)$ 有 $\|\hat{A} - A\|_* \le \frac{\epsilon_m}{m} = \frac{\|\boldsymbol{\eta}_{1:m}\|_1}{m\gamma^2}$，其中 C_0 是一个普适常数。该结论仅对 $m > c_1 \cdot d^2$ 成立，其中 c_1. 是一个已知的普适常数。由此得到的最终遗憾保证为：

$$\text{Regret}[1,T] = \text{Regret}[1,m] + \frac{1}{2}\|x_m - v_m\|^2_{I - \hat{C}_{m+1}} + \sum_{s=m+1}^{T} \frac{1}{2}\mathbb{E}\|v_s - v_{s-1}\|^2_{I-\hat{C}_s} + \frac{1}{2}\mathbb{E}\|x_{s-1} - v_s\|^2_{\hat{C}_s(A-\hat{A})\hat{C}_s}$$

$$\begin{aligned} &= \text{Regret}[1,m] + \frac{1}{2}\|x_m - v_m\|^2_{I-\hat{C}_{m+1}} + \text{Regret}[m+1,T] \\ &\le m\gamma^2 \cdot ((1/2)tr(A) + d) + \sum_{t=1}^{m} \left\{ (1/2) \underbrace{tr(\Sigma)}_{\langle \Sigma, I \rangle} - \frac{1}{2} \underbrace{\mathbb{E}\|v_{s-1} - v_s\|^2_{I-C_s}}_{\langle \Sigma, I - C_s \rangle} \right\} + \frac{\gamma^2}{2} d \\ &\quad + \text{Regret}[m+1,T] \\ &= m\gamma^2 \cdot ((1/2)tr(A) + d) + \sum_{t=1}^{m} (1/2)\langle \Sigma, C_s \rangle + \frac{\gamma^2}{2} d + \text{Regret}[m+1,T] \\ &\le \underbrace{m\left(\frac{(1/2)tr(\Sigma)}{1 + \sigma^A_{\min}}\right)}_{\text{Irreducible Cost of NOT knowing } A} + \underbrace{m\gamma^2 \cdot ((1/2)tr(A) + d) + \frac{\gamma^2}{2} d}_{\text{Additional cost of data collection}} + \underbrace{\text{Regret}[m+1,T]}_{o(T) \text{ Regret}} \end{aligned}$$

C.4 推论1的证明

对于满秩情形，即 $rank(A) = r = d$，我们有

$$\begin{aligned} \text{Regret}[m+1,T] \le & \frac{\epsilon(T-m)}{(1 + \sigma^A_{\min} - \epsilon)^2 - 1} \left(\frac{2d\sigma^A_{\max}\|\Sigma\|_F}{(\sigma^A_{\min} - \epsilon)^2} \cdot \epsilon + \frac{2\|\Sigma\|_F \sigma^A_{\max}\sqrt{2d}}{\sigma^A_{\min} - \epsilon} + \frac{tr(\Sigma)}{(1 + \sigma^A_{\min} - \epsilon)^2 - 1} \right) \\ & + \frac{tr(\Sigma)}{2\sigma^A_{\min}} \cdot \epsilon(T-m). \end{aligned}$$

其中 $\epsilon = \|\hat{A} - A\|_* \leq \frac{\|\boldsymbol{\eta}_{1:m}\|_1}{m\gamma^2}$。考虑最坏噪声模型，此时总噪声随测量数量线性增长：$\|\boldsymbol{\eta}_{1:m}\|_1 = \bar{\eta}m$，我们选择 $\gamma^2 = \sqrt{\bar{\eta}}\max\left\{(T-m)^{1/2}, \frac{2}{\sigma^A_{\min}}\right\} \implies \epsilon \leq \min\left\{\frac{1}{\sqrt{T-m}}, \frac{\sigma^A_{\min}}{2}\right\}$ 和 m作为使估计值 $\hat{A}$保持一致性的最小轮次数，即 $m = c_1 \cdot d^2$。由此得到遗憾保证

$$\begin{aligned}\text{Regret}_T \leq & \sqrt{\bar{\eta}(T - c_1 d^2)}\left(\frac{4\sigma^A_{\max}\|\Sigma\|_F(d+\sqrt{2d}) + tr(\Sigma)}{(\sigma^A_{\min})^2} + \frac{tr(\Sigma)}{2\sigma^A_{\min}}\right) \\ & + \max\left\{\sqrt{\bar{\eta}(T-c_1d^2)}, \frac{2\sqrt{\bar{\eta}}}{\sigma^A_{\min}}\right\} \cdot \left((c_1/2)d^2(tr(A)+d) + d/2\right) + c_1 d^2\left(\frac{(1/2)tr(\Sigma)}{1+\sigma^A_{\min}}\right).\end{aligned}$$

对于秩亏情形，即 $rank(A) = r < d$,

$$\begin{aligned}\text{Regret}[m+1, T] \leq & \left(\frac{2d\sigma^A_{\max}\|\Sigma\|_F}{(\sigma^A_{\min}-\epsilon)^2}\right)\epsilon^2(T-m)^2 + \sqrt{\epsilon}(T-m)\left(\frac{2\sqrt{\epsilon}\|\Sigma\|_F\sigma^A_{\max}\sqrt{2d}}{(\sigma^A_{\min}-\epsilon)(1+\sigma^A_{\min}-\epsilon)} + \frac{\sqrt{\epsilon}tr(\Sigma)}{(1+\sigma^A_{\min}-\epsilon)^2-1}\right) \\ & + \sqrt{\epsilon}(T-m)\left(\left[\frac{2d\epsilon^{3/2}}{(\sigma^A_{\min}-\epsilon)^2} + \frac{\sqrt{2d}\cdot\sqrt{\epsilon}}{\sigma^A_{\min}-\epsilon}\right]\|\Sigma\|_F + \max\left\{1, \frac{\sqrt{\epsilon}}{\sigma^A_{\min}}\right\}\cdot tr(\Sigma)\right).\end{aligned}$$

现在，回想一下，在我们的估计在线优化算法中 $\epsilon = \frac{\|\boldsymbol{\eta}_{1:m}\|}{m\gamma^2} \leq \frac{\bar{\eta}}{\gamma^2} +$。测量轮次的代价为 $m\gamma^2$.。我们选择 $\gamma^2 = \sqrt{\bar{\eta}}\max\left\{(T-m)^{2/3}, \frac{2}{\sigma^A_{\min}}\right\} \implies \epsilon \leq \min\left\{\frac{1}{T-m()^{2/3}}, \frac{\sigma^A_{\min}}{2}\right\}$并将 $\hat{m}$ 作为使估计 A保持一致性所需的最小轮次数，即 $m = c_1 \cdot rd$，其中$r < d$ 是 A的秩。这给出了遗憾上界

$$\begin{aligned}\text{Regret}_T \leq & \sqrt{\bar{\eta}}(T - c_1 rd)^{2/3}\left(\frac{8d\sigma^A_{\max}\|\Sigma\|_F}{(\sigma^A_{\min})^2} + \frac{((4\sigma^A_{\max}+2)\sqrt{d} + 4\sqrt{2}d)\|\Sigma\|_F + (\sigma^A_{\min}+\sqrt{2})tr(\Sigma)}{(\sigma^A_{\min})^{1/2}}\right) \\ & + \sqrt{\bar{\eta}}\max\left\{(T-c_1rd)^{2/3}, \frac{2}{\sigma^A_{\min}}\right\}\left((c_1/2)rd(tr(A)+d) + d/2\right) + c_1 rd\left(\frac{(1/2)tr(\Sigma)}{1+\sigma^A_{\min}}\right)\end{aligned}$$

C.5 定理2的证明

考虑任意一个如下形式的算法：

$$x_t = \hat{C}_t x_{t-1} + (I - \hat{C}_t)v_t$$

其中 $\frac{\epsilon}{2} \leq \sigma^{C_t - \hat{C}_t\min} \leq \sigma^{C_t - \hat{C}_t\max} \leq \epsilon$. 对 [8] 中的插值算法（定理3.7）的分析可以扩展为以下朴素遗憾界（在第一步中）。使用 $A = \mathbf{0}_{d\times d}$ 的具体实例（在第二步中）：

$$\begin{aligned}\text{Regret}_T &\geq \sum_{s=1}^{T}\frac{1}{2}\mathbb{E}\|x_{s-1} - v_s\|^2_{(\hat{C}_s - C_s)A(\hat{C}_s - C_s) + (\hat{C}_s - C_s)^2} \\ &\geq \frac{\epsilon^2}{4}\sum_{s=1}^{T}\frac{1}{2}\mathbb{E}\|x_{s-1} - v_s\|_2^2\end{aligned}$$

$$\geq \frac{\epsilon^2}{8} \sum_{s=1}^{T} \frac{1-(1-\epsilon)^{2s}}{\epsilon(2-\epsilon)}$$
$$\geq \frac{\epsilon T}{8} - \text{ constant}$$

由于探索阶段的遗憾为 $\propto \frac{1}{\epsilon}$，总遗憾为

$$\text{Regret}_T \geq \frac{1}{\epsilon} + \frac{\epsilon T}{8} - \text{ constant} = \Omega(\sqrt{T})$$

D 基本权衡

考虑动作空间 $\mathbb{R}^d$，玩家从原点出发。时间范围固定为 $T = 2d-1$，底层的 A 矩阵是对角矩阵，具有 d 个互不相同的正值。现在，环境提供一系列极小化点 $\{v_t\}_{t=1}^{2d-1}$，使得在前 d 轮次（v_t）$_2 =$（v_t）$_3 = \ldots =$（v_t）$_d = 0$，而从第（$d+1$）轮开始，在第 $t =$（$d+i$）轮，v_t 满足 $v_t - x_{t-1}$ 与 e_{i+1} 平行，其中 $i \in \{1, \ldots, d-1\}$.。这样一系列极小化点可以通过对抗式地或随机地生成(，例如，$v_{d+i} \sim \mathcal{N}$（x_{t-1}，$e_{i+1}e_{i+1}^T$）)。

选项 1：不进行估计的跟踪：SOCO 文献中的大多数在线算法[7, 8, 14, 26, 40, 84] 具有如下通用形式：

$$x_t^{\texttt{ALG}} = \operatorname*{argmin}_{x \in \mathbb{R}^d} f_t(x) + c(x, x_{t-1}) + g(x, v_t, x_{t-1}),$$

对于二次成本函数，这会使 x_t 位于 x_{t-1} 和 v_t之间的直线上。这意味着，任何鲁棒在线算法 ALG 都会指示玩家在最初的 d 轮次中采取一系列与 e_1 平行的动作（$x_1^{\texttt{ALG}}$，.....，$x_d^{\texttt{ALG}}$）。在这些轮次中，老虎机反馈模型会收集信息：

$$\left\{(1/2) \cdot (x_k^{\texttt{ALG}} - v_k)^T A (x_k^{\texttt{ALG}} - v_k) + \eta_k\right\}_{k=1}^{d} = \{c_k A_{1,1} + \eta_k\}_{k=1}^{d}.$$

在第（$d+1$）轮，$x_{d+1}^{\texttt{ALG}}$ 应位于 x_d 和 v_{d+1}之间的直线上，该直线与 e_2平行。因此，x_{d+1} 与未知矩阵 A 的第二个对角线元素 $A_{2,\ 2}$存在**直接依赖**。玩家现在只能求助于迄今收集到的上述秩一数据，因为这是它掌握的关于矩阵 A. 的唯一信息

然而，事实证明，最初的 d 轮次只产生了关于 $A_{1,\ 1,}$ 的信息，迫使玩家对 x_{d+1}. 进行盲目猜测 。事实上，情况还会更糟：这种情况会反复发生，第 $d+i$ 轮需要知道 $A_{i+1,\ i+1}$ 的值才能计算出鲁棒动作，但玩家却只知道 $\{A_{1,1,}, \ldots, A_{i,i}\}$。

选项 2：全力进行估计：或者，玩家可以花费若干轮次$\{2, 3 \ldots, d\}$ 探测其余方向上的命中成本，从而收集整个矩阵 A的信息。然而，由于老虎机反馈中存在噪声，这种机制的成本非常高

。考虑玩家使用动作 $x_k = v_k + \gamma \cdot u_k$ 探测命中成本 $f_t(x) = \frac{1}{2}(x - v_k)^T A(x - v_k)$ 的轮次 k（及其前一轮 $k-1$）。这次测量产生的成本为：

$$f_k(x_k) + c(x_k, x_{k-1}) > \gamma^2 \left(\frac{\lambda_{\min}^A}{2} + \frac{\|x_k - x_{k-1}\|_2^2}{2} \right)$$

其中，估计要求 γ 和 u_k 满足以下条件，以避免数据冗余

$$\gamma^2 \gg \frac{2\eta_k}{\lambda_{\min}^A} \to \text{high signal-to-noise ratio (SNR)}$$
$$u_k \not\parallel u_{k-1} \to \text{avoid directional redundancy.}$$

数据收集机制需要满足受限一致有界性（RUB）性质[48],才能恢复原始矩阵 A:

$$C_1 \leq \frac{\frac{1}{m}\sum_{k=1}^{m} \frac{\gamma^2}{2} u_k^T A u_k}{\|A\|_F} \leq C_2 \ \forall \ A \in \mathbb{R}^{d \times d} \text{ having } rank(A) = r$$

为了使这一结论以高概率对任意秩为 r的对称矩阵 A 成立， $\{u_k\}_k$ 必须从方向均匀分布中独立同分布（IID）地采样，通常采用正态分布。这会导出：

$$\|x_k - x_{k-1}\|_2^2 > \frac{\gamma^2}{2}\|u_k - u_{k-1}\|_2^2 - \|v_k - v_{k-1}\|_2^2$$

其期望为

$$\mathbb{E}\|x_k - x_{k-1}\|_2^2 > \gamma^2 - tr(\Sigma)$$

因此可以确定，每轮测量的成本约为 $\Theta(\gamma^2)$，而进行 m 轮测量会导致线性遗憾，其大小为

$$\text{Regret}_{\text{Estimation}}[1, m] \simeq \Theta(\gamma^2) \cdot m$$

由于忽视了较高的命中成本和切换成本，实际轨迹与最优轨迹 [8, 40]之间存在显著偏差。因此，我们得到以下权衡：

> “收集高保真度的秩-1数据用于矩阵估计，会产生高昂的在线成本，但从长远来看可能会有所回报。仅仅采用鲁棒在线算法或许在初期有益，但如果算法需要了解矩阵 A.，玩家就会因此处于不利境地”

这表明，SOQO 与数据驱动学习是彼此正交的任务，极难将二者结合起来。