arXiv:2601.00461v1 [cs.LG] 1 Jan 2026

拉普拉斯核化老虎机

Shuang Wu
系 f S
加州大学洛杉矶分校，洛杉矶，加利福尼亚州 90025
shuangwu222@ucla.edu

Arash A. Amini统计系，
加州大学洛杉矶分校，洛杉矶，加利福尼亚州 90025
aaamini@stat.ucla.edu

摘要

我们研究多用户上下文多臂老虎机问题，其中用户由一张图关联，其奖励函数同时表现出非线性行为和图同质性。我们为用户奖励函数集合 $\{f_u\}$引入一种有理论依据的联合惩罚项，将基于再生核希尔伯特空间距离的图平滑项与个体粗糙度惩罚相结合。我们的核心贡献是证明，该惩罚等价于单一统一的多用户再生核希尔伯特空间中的平方范数。我们明确推导出其再生核，该核巧妙地融合了图拉普拉斯算子与基础臂核。这种统一化使我们能够将问题重新表述为学习一个单一的“提升”函数，并据此设计有理论依据的算法 LK-GP-UCB 和 LK-GP-TS，利用基于这一新核的高斯过程后验进行探索。我们给出了高概率遗憾界，其增长取决于多用户核的有效维数，而不再依赖用户数量或环境维数。实验表明，在非线性情形下，我们的方法优于表现强劲的线性基线方法和未考虑图结构的基线方法；即使真实奖励为线性函数，我们的方法也仍具竞争力。我们的工作提供了一个统一、具有坚实理论基础且实用的框架，将拉普拉斯正则化与核化老虎机结合起来，以实现结构化探索。

1 引言

图在现代序贯决策中无处不在，能够编码用户、物品或传感器等实体之间的相似性或交互关系。在多用户上下文老虎机设定中，图结构具有重要信息，因为它提供了共享信息的途径，使算法能够比将每位用户孤立处理时更高效地学习。我们研究已知用户图能够促进同质性的问题，即相连用户往往具有相似的奖励函数。在每一轮 t，学习者观察到一位用户 u_t 和一组可用臂（上下文）$\mathcal{D}_t \subset \mathbb{R}^d$，选择一个臂 $\boldsymbol{x}_t \in \mathcal{D}_t$，并获得带噪奖励 y_t。为每位用户单独学习模型的朴素做法效率低下，导致遗憾随用户数量增长。然而，利用图结构能够显著提升样本效率和性能 [1, 2, 3, 4]。

这一问题最初被形式化为 Gang of Bandits (GOB) [5], , 用于对用户奖励函数集合 $\{f_u(\cdot)\}_{u=1}^n$ 进行建模，将其视为图上的平滑信号。GoB.Lin 等开创性工作 [5]假设奖励函数为线性函数，$f_u(\boldsymbol{x}) = \boldsymbol{\theta}_u^\top \boldsymbol{x}$，并通过图拉普拉斯算子惩罚不平滑性，从而得到有效的线性赌博机解法。后续研究通过改进计算扩展性 [6, 7], 拓展了这一方法，但总体上仍局限于线性范式。然而，在从推荐系统到个性化医疗的许多应用中，奖励函数都呈现出复杂的非线性行为。尽管已有大量文献研究如何利用核化老虎机处理单个智能体的非线性奖励 [8, 9, 10, 11, 12], , 但针对多用户图设置的系统性方法仍不够成熟。现有方法通过启发式方式，将用户核与臂核的乘积构造成多用户核 [13], , 使直观的建模目标与最终算法之间仍存在差距。相关工作的进一步讨论请参见附录 A。

预备加州大学洛杉矶分校统计与数据科学系的初步成果。请勿传播。（版权@2025年 ）

我们的工作从这一问题的自然第一性原理出发，弥合了这一差距。理想的奖励函数集合 $\{f_u\}_{u=1}^n$ 中，每个函数 f_u 都属于再生核希尔伯特空间（RKHS） $\mathcal{H}_x$，并应受到联合正则化：它们在图上应保持平滑（同质性），同时各自具有良好的性质（低复杂度）。我们通过一种直观的加性惩罚项对此进行形式化，该惩罚项结合了图平滑项与标准岭惩罚。在不含臂特征的标量情形下，已知这种基于拉普拉斯的正则化会诱导出一个核，其矩阵是图的（正则化）格林函数 [14]。基于这一联系，我们证明，同一原理可以推广到多用户上下文设置：联合惩罚项定义了统一的多用户再生核希尔伯特空间中单个提升函数 $f(\boldsymbol{x},\ u) := f_u(\boldsymbol{x})$ 的平方范数。我们明确推导了该空间的再生核，它优雅地融合了图拉普拉斯算子 L 与基础臂核 K_x：

$$K((\boldsymbol{x},u),(\boldsymbol{x}',u')) = [\boldsymbol{L}_\rho^{-1}]_{u,u'}\, K_x(\boldsymbol{x},\boldsymbol{x}'),$$

其中 $\boldsymbol{L}_\rho = \boldsymbol{L} + \rho\boldsymbol{I}$ 是正则化拉普拉斯。

这一统一视角将学习 n 相关函数的问题，转化为在定义明确的核空间中学习单个函数这一优雅的问题。它使我们能够直接应用高斯过程（GP）多臂老虎机的强大工具 [15, 16, 17]。我们提出了 LK-GP-UCB 和 LK-GP-TS 这两种算法，其原理性的不确定性估计来自这一统一核的高斯过程后验，使它们能够自然而然地联合利用非线性臂结构和拉普拉斯同质性。我们为这些算法提供了遗憾保证，其界限由有效维数 表示，该维数刻画了图与核之间的谱相互作用。实验表明，我们的方法在线性机制下具有竞争力；当奖励呈非线性但具有图平滑性时，其表现显著优于线性基线方法和未考虑图结构的基线方法。

我们的主要贡献如下：

- 我们对广义赌博机群问题进行了形式化，并为奖励函数集合构建了一个原理性的联合惩罚项，将图平滑性与再生核希尔伯特空间正则性相结合$\{f_u\}_{u=1}^n$。
- 我们证明，该惩罚项等价于单个多用户再生核希尔伯特空间中的平方范数，并明确推导出其再生核，从而统一图结构与臂结构。
- 我们提出 LK-GP-UCB 和 LK-GP-TS，这两种基于高斯过程的老虎机算法利用这一统一核进行有原则且有效的探索。
- 我们根据有效维数给出了新颖的遗憾界，该有效维数取决于核和图拉普拉斯算子的谱性质。
- 我们通过实验证实了所提出方法的有效性：在具有非线性、图平滑奖励结构的场景中，相较于强基线方法，其性能获得了显著提升。

记号。令 $[n]$ 为集合 $\{1,2,...,n\}$。对于集合或事件 $\mathcal{E}$，我们将其补集记为 $\mathcal{E}$。向量均假设为列向量。 $\boldsymbol{e}_i$ 是 $\mathbb{R}^n$中的第 i 个标准基向量。 $\boldsymbol{I}$ 是单位矩阵。$\lambda_{\min}(\boldsymbol{A})$表示矩阵 $\boldsymbol{A}$的最小特征值。 $\otimes$ 是克罗内克积。将截至（但不包括）轮次 t 的随机性历史记录记为 $\mathcal{F}_t$，并将 $\mathbb{P}_t(\cdot) := \mathbb{P}(\cdot\,|\mathcal{F}_t)$ 和 $\mathbb{E}_t(\cdot) := \mathbb{E}[\cdot\,|\mathcal{F}_t]$ 分别表示给定 $\mathcal{F}_t$时的条件概率与期望。我们使用 $\mathcal{O}$ 表示忽略对数因子的 big-O记号，并使用 $\asymp$ 表示任意两个函数在增长速率上的渐近等价。

2 问题形式化

2.1 具有非线性奖励的 Gang of Bandits

我们考虑一个多用户上下文老虎机问题，通常称为 Gang of Bandits（GOB） [5]，，其中有 n 个用户和一个可能无限的臂集合。我们将用户集合记为 $\mathcal{U} = \{1,\dots,n\}$，臂集合记为 $\mathcal{D} \subseteq \mathbb{R}^d$，其中每个臂都由一个特征向量 $\boldsymbol{x} \in \mathcal{D}$ 表示。用户通过一个已知的无向图 $G = (\mathcal{U},\ \mathcal{E})$ 相连，其中 $\mathcal{E}$ 是边集合。令 $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ 为非负边权重 w_{ij} 构成的矩阵， $\boldsymbol{D}$ 为对角度矩阵，其元素为 $d_i :\ = \sum_j w_{ij}$。相应的图拉普拉斯算子为 $\boldsymbol{L} :\ = \boldsymbol{D} - \boldsymbol{W}$。

学习过程持续 T轮次。在每一轮 $t \in \{1,\dots,T\}$，环境会呈现一位用户 $u_t \in \mathcal{U}$（例如，随机/均匀分布地选择一位用户）以及一个可用臂的有限子集

$\mathcal{D}_t \subseteq \mathcal{D}$。学习者按照某种决策策略 π 选择一个臂 $\boldsymbol{x}_t \in \mathcal{D}_t$，并观测到一个带噪奖励：$y_t = f_{u_t}(\boldsymbol{x}_t) + \varepsilon_t$ 其中 $\{f_u : \mathcal{D} \to \mathbb{R}\}_{u=1}^n$ 是一组未知的奖励函数，每位用户对应一个。假设噪声项 ε_t 在给定交互历史 $\mathcal{F}_t$的条件下具有条件零均值，且为方差代理为 σ^2 的次高斯噪声。为便于说明，我们使用$f_{1:n} := \{f_u\}_{u=1}^n$ 作为用户级奖励函数的集合。

学习者的目标是最小化累积遗憾。在时间 t 产生的瞬时遗憾为$\Delta_t = f_{u_t}(\boldsymbol{x}_t^*) - f_{u_t}(\boldsymbol{x}_t)$，其中 $\boldsymbol{x}_t^* = \mathrm{argmax}_{\boldsymbol{x}\in\mathcal{D}_t} f_{u_t}(\boldsymbol{x})$，而在 T 轮次内的累积遗憾定义为 $\mathcal{R}_T = \sum_{t=1}^T \Delta_t$。成功的算法必须实现次线性遗憾，$\mathcal{R}_T/T \to 0 \,\mathrm{as} T \to \infty$，从而确保每轮平均遗憾趋于零。

2.2 图同质性的原则性正则性模型

为了使学习过程易于处理，我们需要对未知函数 $f_{1:n}$施加正则性。我们提出两个核心假设。首先，我们假设每个函数 f_u 都具有良好的性质，属于同一个再生核希尔伯特空间（RKHS），记作 $\mathcal{H}_x$，其对应的半正定核为 $K_x : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$。关联的特征映射记作 φ，使得$K_x(\boldsymbol{x}, \boldsymbol{x}') = \langle \varphi(\boldsymbol{x}), \varphi(\boldsymbol{x}') \rangle_{\mathbb{R}^d}$。这刻画了奖励相对于臂特征的非线性结构。

其次，我们通过假设在G 中由一条边相连的用户具有相似的奖励函数，来形式化图同质性的概念。此用户相似性由再生核希尔伯特空间中函数之间的平方距离 $\|f_i - f_j\|_{\mathcal{H}_x}^2$ 来度量。结合这些原则，我们将真实奖励函数建模为具有较小联合惩罚项，该惩罚项平衡了图平滑性与单个函数的复杂度：

$$\mathrm{PEN}(f_{1:n}; \rho) := \underbrace{\frac{1}{2}\sum_{i,j=1}^n w_{ij}\|f_i - f_j\|_{\mathcal{H}_x}^2}_{\mathrm{PEN}_{graph}(f_{1:n})} + \rho \underbrace{\sum_{i=1}^n \|f_i\|_{\mathcal{H}_x}^2}_{\mathrm{PEN}_{ridge}(f_{1:n})} = \sum_{i,j=1}^n [\boldsymbol{L}_\rho]_{ij} \langle f_i, f_j \rangle_{\mathcal{H}_x}, \tag{1}$$

其中，$\rho > 0$是一个正则化超参数，$\boldsymbol{L}_\rho := \boldsymbol{L} + \rho \boldsymbol{I}$是正则化图拉普拉斯矩阵。此惩罚项是我们框架的核心，因为它为建模相关的非线性函数提供了清晰、可解释的目标。

2.3 从联合惩罚项到统一的多用户核

我们的关键理论洞见是，式 (1) 中直观的加性惩罚项并非仅仅是一个临时拼凑的正则项。事实上，它是用户-臂乘积域 $\mathcal{U} \times \mathcal{D}$ 上一个统一希尔伯特空间中的平方范数。这使我们能够将问题重新表述为：不再学习 n 个相关函数，而是在这个新空间中学习一个“提升后”的函数 $f(\boldsymbol{x}, u) := f_u(\boldsymbol{x})$。我们将在以下定理中证明，它是乘积空间 $\mathcal{H} = \mathcal{H}_G \otimes \mathcal{H}_x$ 上的再生核希尔伯特空间范数平方，其中 $\mathcal{H}_G$ 是核为 $K_G(u, u') = [\boldsymbol{L}_\rho^{-1}]_{u,\,u'}$ 的再生核希尔伯特空间。

定理 2.1（多用户核）。设 $\mathcal{H}_x$ 是定义在 $\mathcal{D}$ 上、核为 K_x 的再生核希尔伯特空间。函数集合 $\mathcal{H} := \{(f_1, \ldots, f_n) : f_u \in \mathcal{H}_x, \forall u \in \mathcal{U}\}$ 所构成的向量空间配备内积

$$\langle f, g \rangle_{\mathcal{H}} := \sum_{i,j=1}^n [\boldsymbol{L}_\rho]_{ij} \langle f_i, g_j \rangle_{\mathcal{H}_x}$$

是定义在 $\mathcal{U} \times \mathcal{D}$ 上的函数再生核希尔伯特空间。其对应的 *RKHS* 范数平方恰好是 (1) 中的惩罚项，其再生核 $K : (\mathcal{D} \times \mathcal{U})^2 \to \mathbb{R}$ 由下式给出：

$$K((\boldsymbol{x}, u), (\boldsymbol{x}', u')) = [\boldsymbol{L}_\rho^{-1}]_{u,u'} K_x(\boldsymbol{x}, \boldsymbol{x}'). \tag{2}$$

这一结果非常强大：它为融合图和特征信息的多用户核提供了一种直接且规范的构造方式。核 K_x 捕捉臂之间的相似性，而矩阵$\boldsymbol{L}_\rho^{-1}$（图的格林函数）捕捉用户之间的相似性，其中 $[\boldsymbol{L}_\rho^{-1}]_{u,\,u'}$ 度量用户 u 和 u' 之间通过图中所有路径建立的连接强度。更多背景信息请参见附录 A。

这种统一性使我们能够通过特征映射 $\phi(\boldsymbol{x}, u)$ 表示提升奖励函数 $f(\boldsymbol{x}, u)$，使得 $f(\boldsymbol{x}, u) = \langle \boldsymbol{\theta}, \phi(\boldsymbol{x}, u) \rangle$ 对于某个（可能是无限维的）参数 $\boldsymbol{\theta}$，并且

$K((\boldsymbol{x},\ u),\ (\boldsymbol{x}',\ u'))=\langle\phi(\boldsymbol{x},\ u)\ \phi(\boldsymbol{x}',\ u')\rangle$。形式上，对于上下文-用户对 $(\boldsymbol{x},\ u)\in\mathcal{D}\times\mathcal{U}$，特征映射 ϕ 定义为 $\phi(\boldsymbol{x},\ u):=\boldsymbol{L}_\rho^{-1/2}\boldsymbol{e}_u\otimes\varphi(\boldsymbol{x})$。现在，问题被转化为在多用户再生核希尔伯特空间 $\mathcal{H}$中学习单个函数。这一见解为基于高斯过程的原理性算法方法铺平了道路，下面将对此进行详细说明。

3 拉普拉斯核化老虎机算法

对具有显式核 K 的多用户再生核希尔伯特空间的识别，为 GOB 问题提供了一个强大且统一的框架。它使我们能够使用单个高斯过程（GP）对整个系统——涵盖所有用户和臂——进行建模，从而避免管理 n 个彼此独立但相关的模型所带来的复杂性。

3.1 高斯过程视角

我们提出了基于高斯过程（GP）的算法，其动机来自核化老虎机文献 [8]。我们的贝叶斯建模仅用于推导估计量，并不一定代表真实模型。我们在未知的提升奖励函数f 上置入高斯过程先验： $\mathcal{D}\times\mathcal{U}\to\mathbb{R}$，记作

$$[f_1(\cdot),\dots,f_n(\cdot)]\sim\mathcal{GP}(0,K(\cdot,\cdot)).$$

其中， K 是式（2）中定义的多用户核。对于任意有限的用户-臂对集合 $\{(\boldsymbol{x}_i,\ u_i)\}_{i=1}^t$，该先验意味着 $\boldsymbol{f}_t:=[f_{u_1}(\boldsymbol{x}_1)\ \cdots,\ f_{u_t}(\boldsymbol{x}_t)]^\top\sim\mathcal{N}(\boldsymbol{0},\ \boldsymbol{K}_t)$，其中 $\boldsymbol{K}_t\in\mathbb{R}^{t\times t}$的元素为$[\boldsymbol{K}_t]_{ij}=K((\boldsymbol{x}_i,\ u_i),\ (\boldsymbol{x}_j,\ u_j))$，即核矩阵。

在第 t 轮，给定用户 u_t 和选定的臂 $\boldsymbol{x}_t$，贝叶斯模型假设奖励模型为 $y_t=f(\boldsymbol{x}_t,\ u_t)+\varepsilon_t$其中 $\varepsilon_t\sim\mathcal{N}(0,\ \lambda)$ 为噪声。因此，给定历史记录 $\mathcal{F}_t$， $f_u(\boldsymbol{x})$ 的后验分布为 $\mathcal{N}(\mu_{u,\ t-1}(\boldsymbol{x})\ \sigma^2_{u,\ t-1}(\boldsymbol{x}))$，其后验均值和方差为:

$$\begin{aligned}\mu_{u,t}(\boldsymbol{x})&=\boldsymbol{k}_t(\boldsymbol{x},u)^\top(\boldsymbol{K}_t+\lambda\boldsymbol{I}_t)^{-1}\boldsymbol{y}_t\\ \sigma^2_{u,t}(\boldsymbol{x})&=K((\boldsymbol{x},u),(\boldsymbol{x},u))-\boldsymbol{k}_t(\boldsymbol{x},u)^\top(\boldsymbol{K}_t+\lambda\boldsymbol{I}_t)^{-1}\boldsymbol{k}_t(\boldsymbol{x},u).\end{aligned}\tag{3}$$

这里， $\boldsymbol{k}_t(\boldsymbol{x},\ u):=[K((\boldsymbol{x}_1,\ u_1),\ (\boldsymbol{x},\ u))\dots,\ K((\boldsymbol{x}_t,\ u_t),\ (\boldsymbol{x},\ u))]^\top\in\mathbb{R}^t$是过去选定的用户-动作对 $\{(\boldsymbol{x}_s,\ u_s)\}_{s=1}^t$与新用户-动作对 $(\boldsymbol{x},\ u)$ 之间的核向量，而 $\boldsymbol{y}_t=[y_1,\dots,y_t]^\top\in\mathbb{R}^t$ 是观测奖励。

注 1. 当 $\{(u_t,\ \boldsymbol{x}_t)\}_{t=1}^T$ 是固定的（确定性）序列时，在该模型下有$\boldsymbol{y}_t\,|\,\boldsymbol{f}_t\sim N(\boldsymbol{f}_t,\ \lambda\boldsymbol{I}_t)$和 $\boldsymbol{f}_t\sim N(\boldsymbol{0},\ \boldsymbol{K}_t)$。于是， $\boldsymbol{y}_t$ 与 $\boldsymbol{f}_t$ 之间的互信息为： $I(\boldsymbol{y}_t;\ \boldsymbol{f}_t)=\frac{1}{2}$ 对数行列式 $\big(\boldsymbol{I}_t+\lambda^{-1}\boldsymbol{K}_t\big)$，通常称为贝叶斯模型 *[15,* 第 *2* 节 *1]*的信息增益。为方便起见，我们记作

$$\gamma_t:=\log\det\big(\boldsymbol{I}_t+\lambda^{-1}\boldsymbol{K}_t\big),\tag{4}$$

并将其称为第 t轮的信息增益，尽管它是文献中通常所称信息增益的两倍。此外，在我们的记号中， γ_t 取决于序列，而在文献中，该符号通常用于表示所有长度为 T 的序列$\{(u_t,\boldsymbol{x}_t)\}_{t=1}^T$ 上的最大信息增益。

与正则化回归的联系。值得注意的是，式（3）中的高斯过程后验均值估计器等价于离线核拉普拉斯正则化回归（KLRR）问题的解。具体而言，使正则化最小二乘目标函数达到最小值的函数 $f\in\mathcal{H}$

$$\min_{f\in\mathcal{H}}\sum_{s=1}^t(f(\boldsymbol{x}_s,u_s)-y_s)^2+\lambda\|f\|^2_{\mathcal{H}}\tag{5}$$

正是后验均值函数 $\mu_{t-1}(\boldsymbol{x},u)$。这一等价关系证实，我们基于高斯过程的在线算法与批量学习原则密切相关：该原则旨在最小化预测误差，并通过我们在 (1) 中提出的多用户再生核希尔伯特空间范数进行正则化。

3.2 决策策略：UCB 和 Thompson采样

基于这些后验估计，我们可以设计能够有效平衡探索与利用的老虎机算法。我们提出了两种基于常见且强大的启发式方法的算法：上置信界（UCB）和 Thompson采样（TS）。完整流程见附录 E.1

拉普拉斯核化 GP-UCB（LK-GP-UCB）。遵循"不确定性下的乐观主义"原则，我们的 UCB 算法会选择奖励乐观估计值最高的臂。在第 t 轮，观察到用户 u_t 和臂集合 $\mathcal{D}_t$ 后，它会选择：

$$\boldsymbol{x}_t = \underset{\boldsymbol{x}\in\mathcal{D}_t}{\operatorname{argmax}} \Big(\mu_{u_t,t-1}(\boldsymbol{x}) + \beta_t \sigma_{u_t,t-1}(\boldsymbol{x})\Big), \tag{6}$$

其中，β_t 是超参数，用于通过置信宽度 $\sigma_{u_t,\,t-1}(\boldsymbol{x})$ 确保适当的探索尺度。我们的理论分析在定理 4.2 中给出了 β_t 的显式形式，以保证较低的遗憾；不过在实践中，它通常被视为可调超参数。

拉普拉斯核化 GP-TS（LK-GP-TS）。Thompson采样 [18, 19] 遵循"概率匹配"原则。每一轮，它都会从后验分布中抽取一个随机函数，并根据该样本采取贪心行动。一种实用的实现方式是，选择使奖励的后验预测分布样本最大化的臂：

$$\boldsymbol{x}_t = \underset{\boldsymbol{x}\in\mathcal{D}_t}{\operatorname{argmax}} \Big(\mu_{u_t,t-1}(\boldsymbol{x}) + \nu_t z_t(\boldsymbol{x}) \sigma_{u_t,t-1}(\boldsymbol{x})\Big), \tag{7}$$

其中，ν_t 是用于探索的尺度超参数，$z_t(\boldsymbol{x}) \sim \mathcal{N}(0,\,1)$为高斯扰动。与常见的 Thompson采样文献一致，我们在式 (7) 中的决策策略可分为两步：对所有 $\boldsymbol{x} \in \mathcal{D}_t$，从 $\mathcal{N}(\mu_{u_t,\,t-1}(\boldsymbol{x}),\, \nu_t^2 \sigma^2_{u_t,\,t-1}(\boldsymbol{x}))$ 中采样 $\widetilde{\mu}_t(\boldsymbol{x})$，然后通过 $\boldsymbol{x}_t = \operatorname{argmax}_{\boldsymbol{x}\in\mathcal{D}_t} \widetilde{\mu}_t(\boldsymbol{x})$ 选择一个臂。与置信上界算法类似，我们也在定理 4.3 中对 ν_t 采用明确的理论取值，而在实际应用中，它是一个调优超参数。

3.3 实际实现

对式 (3) 中的后验更新进行朴素实现，计算开销很大，因为每一步都需要进行一次 $\mathcal{O}(t^3)$ 矩阵求逆。为确保实际应用中的可扩展性，我们可以使用递归公式，以 $\mathcal{O}(t^2)$ 的方式更新后验均值和方差；对于固定网格上的点，还可以进一步提高效率。具体而言，我们可以维护并更新逆矩阵 $(\boldsymbol{K}_t + \lambda \boldsymbol{I}_t)^{-1}$，或者使用以下递归公式更新后验估计量 [8]:

$$\begin{aligned}
\mu_{u,t}(\boldsymbol{x}) &= \mu_{u,t-1}(\boldsymbol{x}) + \frac{q_{t-1}((\boldsymbol{x},u),(\boldsymbol{x}_t,u_t))}{\lambda + \sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}(y_t - \mu_{u_t,t-1}(\boldsymbol{x}_t)) \\
q_t((\boldsymbol{x},u),(\boldsymbol{x}',u')) &= q_{t-1}((\boldsymbol{x},u),(\boldsymbol{x}',u')) - \frac{q_{t-1}((\boldsymbol{x},u),(\boldsymbol{x}_t,u_t)) q_{t-1}((\boldsymbol{x}_t,u_t),(\boldsymbol{x}',u'))}{\lambda + \sigma^2_{u_t,t-1}(\boldsymbol{x}_t)} \\
\sigma^2_{u,t}(\boldsymbol{x}) &= \sigma^2_{u,t-1}(\boldsymbol{x}) - \frac{q^2_{t-1}((\boldsymbol{x},u),(\boldsymbol{x}_t,u_t))}{\lambda + \sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}.
\end{aligned} \tag{8}$$

其中，$q_t((\boldsymbol{x},\ u),\ (\boldsymbol{x}',\ u'))$ 是第 t 轮的估计后验协方差。我们在附录E.2中说明如何获得这些更新。对于较小的 t，采用精确求逆；对于较大的 t，则切换为递归更新，这种混合方法可以兼顾数值稳定性和计算效率。有关实现的更多细节见附录 F.4。

4 遗憾分析

现在，我们为所提出的算法提供理论保证。我们的分析基于针对高斯过程后验估计的高概率置信界，并由此为 LK-GP-UCB 和 LK-GP-TS 推导出次线性遗憾界。

4.1 假设

我们的结果依赖于以下标准假设。

假设 1 (次高斯噪声)。 噪声过程 $\{\varepsilon_t\}_{t=1}^{T}$ 是一个 $\mathcal{F}_t$-可测随机过程，并且在条件下是次高斯的，其常数为 σ^2。

假设 2 （有界基核）。 基础臂核 $K_x(\cdot,\cdot)$ 是半正定的，且其对角线一致有界：$\sup_{\boldsymbol{x}\in\mathcal{D}} K_x(\boldsymbol{x},\boldsymbol{x}) \le \alpha^2$，其中 $\alpha > 0$ 为某个常数。

假设 3 （有界多用户再生核希尔伯特空间范数）。 真实的提升奖励函数 f 在多用户再生核希尔伯特空间 $\mathcal{H}$中的范数有界： $\|f\|_{\mathcal{H}}^2 = PEN(f_{1:n};\ \rho) \le B_\rho^2$，其中 $B_\rho > 0$ 为某个常数。

假设 1 是老虎机文献中的常见假设。假设 2 和 3 与核化老虎机和图平滑性文献中的正则性假设间接一致 [20, 21]。这些假设意味着奖励和多用户核均有界。具体而言，我们有

$$\sup_{(\boldsymbol{x},u)\in\mathcal{D}\times\mathcal{U}} K((\boldsymbol{x},u),(\boldsymbol{x},u)) \le K_{\max} := \alpha^2 \cdot \max_{u\in\mathcal{U}} [\boldsymbol{L}_\rho^{-1}]_{u,u}.$$

4.2 高概率置信界

遗憾分析的核心是将真实函数 f 与后验均值估计量 μ_t联系起来的置信界。该结果量化了模型的不确定性，并为 UCB 算法的探索策略提供了依据。

定理 4.1 （置信界）。 假设假设 *1*、*2* 和 *3* 成立。令 $\{(\boldsymbol{x}_t, u_t)\}_{t=1}^{\infty}$ 为$\mathcal{F}_{t-1}$-可测的离散时间随机过程。那么，在 (3) 中使用后验估计量 $\mu_{u,t}(\boldsymbol{x})$ 和$\sigma_{u,t}(\boldsymbol{x})$ 可得一个高概率上界：对于任意 $\delta \in (0,1)$，以至少$1-\delta$的概率，对所有 $t \ge 1$ 以及所有 $(\boldsymbol{x},u) \in \mathcal{D}\times\mathcal{U}$，均有：

$$|\mu_{u,t}(\boldsymbol{x}) - f(\boldsymbol{x},u)| \le \beta_t \cdot \sigma_{u,t}(\boldsymbol{x}) \tag{9}$$

其中，置信参数 β_t 定义为

$$\beta_t := B_\rho + \sqrt{\frac{\sigma^2}{\lambda}\Big(2\log\frac{1}{\delta} + \log\det(\boldsymbol{I}_t + \lambda^{-1}\boldsymbol{K}_t)\Big)}. \tag{10}$$

该置信界的结构与核化老虎机文献 [8, 22, 13],中的结果相似，但我们的分析有两点关键区别。首先，我们的证明不需要某些先前工作中的约束 $\lambda \ge 1$。更重要的是，我们将 log det $\big(\boldsymbol{I}_t + \lambda^{-1}\boldsymbol{K}_t\big)$ 项直接保留在置信宽度 β_t中。这与经典方法形成对比：经典方法通常会进一步用信息论量来界定该项，因此可能得到较松的界。通过保留精确项，我们为基于有效维数进行更紧致、依赖于数据的分析奠定了基础。

4.3 通过有效维数界定遗憾

To o为获得具体的遗憾率，我们使用

维数。

有效

˜**定义 4.1** （有效维数）。 *</style}*学习问题在给定截至时间 T 的动作序列时，其有效维数 d 定义如下：

$$\tilde{d} := \frac{\log\det(\boldsymbol{I}_T + \boldsymbol{K}_T/\lambda)}{\log(1 + TK_{\max}/\lambda)}. \tag{11}$$

这一量受近期核方法和过参数化模型研究的启发 [23, 24, 25],，用于衡量学习问题的内在复杂度。它可以解释为矩阵 $\boldsymbol{I}_T + \boldsymbol{K}_T/\lambda$ 的对数特征值之和与最大可能对数特征值上界之比（$TK_{\max}$ 是 $\boldsymbol{K}_T$ 最大特征值的上界）。因此，它是一种稳健且依赖于图的矩阵秩度量，能够捕捉算法实际探索的函数空间的“维数”。

˜利用定理 4.1 中的置信界以及定义 4.1 中的 d，我们给出 LK-GP-UCB 和 LK-GP-TS 的遗憾上界如下。

定理 4.2 （LK-GP-UCB的遗憾界）。 假设假设 *1*、*2* 和 *3* 成立，且不对臂的数量作任何假设。根据定理 *4.1*，将 *LK-GP-UCB* 中的探索参数 β_t 设为 β_t，则累积遗憾以高概率满足以下界：

$$\mathcal{R}_T = \mathcal{O}(\tilde{d}\log(T)\sqrt{T}) = \tilde{\mathcal{O}}(\tilde{d}\sqrt{T})$$

定理 4.3 (LK-GP-TS 的遗憾界 LK-GP-TS)。 假设假设 *1*、*2* 和 *3* 成立，且决策集 $\mathcal{D}_t$ 均为有限集。将 *LK-GP-TS* 中的探索参数 ν_t 设为定理 *4.1* 中的 β_t 后，以高概率可得累积遗憾的上界如下：

$$\mathcal{R}_T = \mathcal{O}(\tilde{d}\log(T)^{3/2}\sqrt{T}) = \tilde{\mathcal{O}}(\tilde{d}\sqrt{T})$$

这些界展现了我们方法的效率。遗憾的增长并不取决于˜用户数量 n 或环境特征维数，而取决于有效维数 d。对于˜图和核结构使谱快速衰减的问题， d 可能会显著更小，从而大幅提升样本效率。

˜ ˜按照注 1 中的记号，有效维数 d的增长率为： $d = \gamma_T/\log(1 + TK_{\max}/\lambda) \asymp \gamma_T\log T$，其中该近似假设 $\lambda = \Theta(1)$。在线性情形下（$n = 1$ 在 $\mathbb{R}^d$ 上使用线性核），将 d 理解为维数显然是合理的，此时 $\gamma_T = \mathcal{O}(d\log T)$ [15, 定理 5],给出 $\tilde{d} = \mathcal{O}(d)$。这一例子表明，对于无限动作空间，我们的界 $\tilde{\mathcal{O}}(\tilde{d}\sqrt{T})$ 除对数因子外是紧的，与线性老虎机的极小极大最优速率 $\tilde{\mathcal{O}}(d\sqrt{T})$ 相匹配 [26]。

对于一致有限动作空间（$|\mathcal{D}_t| \le M$ 对所有 t），使用 SupKernelUCB [22] 等算法，可以实现更紧的遗憾界 $\tilde{\mathcal{O}}(\sqrt{\tilde{d}T})$。这一改进依赖于将探索参数缩放为 $\beta_t \propto 1/\sqrt{\lambda}$，而非使用 (10)，从而有效消除一个 $\sqrt{\gamma_T}$ 因子。由于我们的主要贡献是构造统一多用户核，核老虎机文献中的此类算法改进可以直接应用于我们的框架。

4.4 多用户核的谱分析

˜为了解释有效维数 d，我们分析多用户核 K 的谱。根据定理 2.1， $K = K_G \otimes K_x$ 是与矩阵 $\boldsymbol{K}_G = \boldsymbol{L}_\rho^{-1}$ 相关联的用户核 K_G 和臂核 K_x 的张量积。因此，与 K 相关联的积分算子的特征值是边缘特征值两两相乘所得。设 $\{\lambda_i^G\}_{i=1}^n$ 为 $\boldsymbol{L}_\rho^{-1}$ 的特征值， $\{\nu_j^x\}_{j=1}^\infty$ 为 K_x 的特征值。那么， K 的算子特征值为 $\{\mu_{ij} = \lambda_i^G\nu_j^x\}_{i,\ j}$。归一化矩阵 $\boldsymbol{K}_T/T$ 的特征值近似于这些算子特征值[1]。

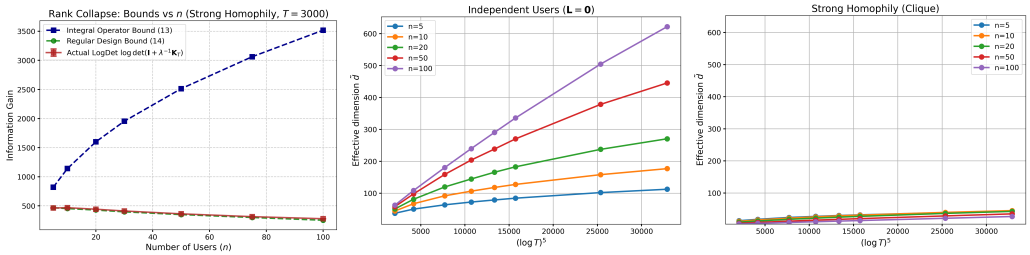
特别地，我们得到如下关于信息增益 γ_T 的近似上界：

$$\gamma_T = \log\det\left(\boldsymbol{I} + \lambda^{-1}\boldsymbol{K}_T\right) \lessapprox \sum_{i=1}^{n}\sum_{j=1}^{\infty}\log\left(1 + \frac{T}{\lambda}\lambda_i^G\nu_j^x\right) = \sum_{i=1}^{n}\Psi\left(\frac{T\lambda_i^G}{\lambda}\right), \tag{12}$$

其中， $\Psi(s) := \sum_{j=1}^\infty \log(1 + s\nu_j^x)$ 表示有效信号强度为 s 时单用户问题的信息增益。我们知道， $\Psi(s)$ 是凹且次线性的；例如，对于 $\mathbb{R}^d$ 上的平方指数核， $\Psi(s) \lesssim (\log s)^{d+1}$ [15, 定理 5])，因此作为 T 的函数， γ_T 随 T 缓慢增长。那么，值得关注的是它对 n 的依赖关系。

虽然（12）中的界具有参考价值，但对于有限的 T，它可能较为保守（见图 1）。通过考虑正则设计，可以得到一个类似但更紧的界：假设我们恰好对每个用户观测 $m := T/n$ 次，并为所有用户选择相同的动作集 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$。通过置换轮次索引，使用户 1 的所有观测排在最前面，然后依次排列用户 2 等用户的观测， $\boldsymbol{K}_T$的特征值保持不变。在此设置下， $\boldsymbol{K}_T = \boldsymbol{K}_G \otimes \boldsymbol{K}_x^{\text{base}}$，其中 $\otimes$ 是矩阵克罗内克积， $\boldsymbol{K}_x^{\text{base}}$是根据共同动作集计算得到的 $m \times m$ 核矩阵。令$\{\hat{}\nu_j^x\}_{j=1}^m$ 为 $\boldsymbol{K}_x^{\text{base}}/m$的特征值。除以 m 进行归一化，可确保当 m较大时，$\hat{}\nu_j^x$趋于总体特征值 ν_j^x。

[1]当 $T \to \infty$ 时，在独立同分布抽样 [27] 下，这一结论渐近成立； [15, 定理 5] 的结果表明，对于最坏情况序列，也有类似的近似结论。



(a) 两个界与实际值　　(b) $\tilde{d}$ 与 $(\log T)^d$ (空图) (c) $\tilde{d}$ 与 $(\log T)^d$ (完全图)

图 1：秩坍塌：(a) 比较独立同分布设计中实际信息增益 γ_T 与 n 的增长（红色），以及完全图中的两个界 (12)（蓝色；粗略）和 (13)（绿色；近乎精确）。核为 $\exp(-\|x-y\|^2/2)$，$u_i \sim$ 均匀分布$([n])$，且 $x_i \sim$ 均匀分布$[0,1]^d$，其中 $d=5$。面板 (b) 和 (c) 分别展示空图和完全图下 $\tilde{d}$ 与 $(\log T)^d$ 的增长。注意~，在完全图下，d 会随着 n 增大而略微减小。

因此，$\boldsymbol{K}_T/T$ 的特征值为 $\lambda_i^G \hat{\nu}_j^x/n$，从而得到精确表达式：

$$\gamma_T = \sum_{i=1}^{n}\sum_{j=1}^{m} \log\left(1 + \frac{T}{n\lambda}\lambda_i^G \hat{\nu}_j^x\right) = \sum_{i=1}^{n} \hat{\Psi}\left(\frac{T}{n\lambda}\lambda_i^G\right), \tag{13}$$

其中 $\hat{\Psi}(s) := \sum_{j=1}^{m} \log\left(1 + s\hat{\nu}_j^x\right)$ 表示具有共同动作的单用户问题的“经验”信息增益。对于足够大的 $m = T/n$，有 $\hat{\nu}_j^x \approx \nu_j^x$ 和 $\hat{\Psi}(s) \approx \Psi(s)$。表达式 (13) 对正则设计是精确的，并且如图 1 所示，它为独立同分布抽样情形提供了精确近似~。我们利用 (13) 分析 d 在不同图结构下的表现。

情形 1：独立用户（最坏情况）。如果 $\boldsymbol{L} = \boldsymbol{0}$，那么 $\mathbf{K}_G = \rho^{-1}\boldsymbol{I}$，且对于所有 $i \in [n]$，$\lambda_i^G = \rho^{-1}$。增益呈线性累加：$\gamma_T^{\text{indep}} = \sum_{i=1}^{n} \hat{\Psi}\left(\frac{T}{n\rho\lambda}\right) = n \cdot \hat{\Psi}\left(\frac{T}{n\rho\lambda}\right)$. 因此，有效维数为单用户有效维数的 n 倍。例如，对于平方指数核，$\tilde{d} = \mathcal{O}(n(\log(T/n))^d)$，其相对于 T 仍呈次线性增长。

情形 2：强同质性（完全图）。为了在同质性先验下单独考察极其稠密的用户图所产生的影响，考虑边权重为 $w_{ij} = 1$ 的完全图。拉普拉斯特征值为 0（重数为 1）和 n（重数为 $n-1$）。核特征值则呈现相反的结构，其中 $\lambda_1^G = 1/\rho$，且对于 $i \geq 2$，$\lambda_i^G = 1/(n+\rho)$。当 n 较大时，这会得到一个近似秩为 1 的矩阵。代入式 (13)，可得“头部 + 尾部”分解：

$$\gamma_T^{\text{clique}} = \hat{\Psi}\left(\frac{T}{n\rho\lambda}\right) + (n-1)\hat{\Psi}\left(\frac{T}{n(n+\rho)\lambda}\right). \tag{14}$$

由此得到以下推论：

命题 4.1. 考虑满足 $T \leq Cn$ （其中 C 为某个常数）的情形。那么，在正则设计下：$\gamma_T^{\text{clique}} \lesssim \frac{C}{\lambda}\left(\frac{1}{\rho} + 1\right) = \mathcal{O}(1)$。

证明. 利用对于 $x \geq 0$ 有 $\log(1+x) \leq x$，可得 $\hat{\Psi}(s) \leq s(\sum_{j=1}^{m} \hat{\nu}_j^x)$。于是，对于 $T \leq Cn$，

$$(n-1)\hat{\Psi}\left(\frac{T}{n(n+\rho)\lambda}\right) \leq n\hat{\Psi}\left(\frac{C}{(n+\rho)\lambda}\right) \leq n \cdot \frac{C}{(n+\rho)\lambda}\sum_{j=1}^{m}\hat{\nu}_j^x \lesssim \frac{C}{\lambda},$$

因为 $\sum_{j=1}^{m} \hat{\nu}_j^x = \mathcal{O}(\sum_{j=1}^{\infty} \nu_j^x) = \mathcal{O}(1)$[2]。类似地，对于第一项，$\hat{\Psi}\left(\frac{T}{n\rho\lambda}\right) \lesssim \frac{C}{\rho\lambda}$。 □

[2]该界适用于积分算子为迹类的任意核。对于如假设 2 中所述的一致有界核，我们有更直接的界 $\sum_{j=1}^{m} \hat{\nu}_j^x = \operatorname{tr}\left(\boldsymbol{K}_x^{\text{base}}\right)/m \leq \alpha^2$。

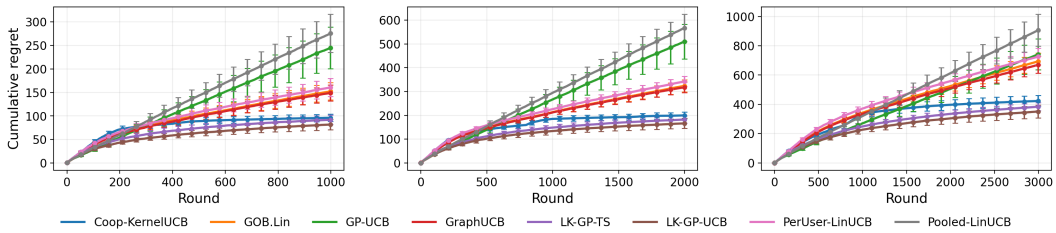


图 2：*Linear-GOB* 情形下的累积遗憾。从左到右依次为难度较低、中等和较高的任务。

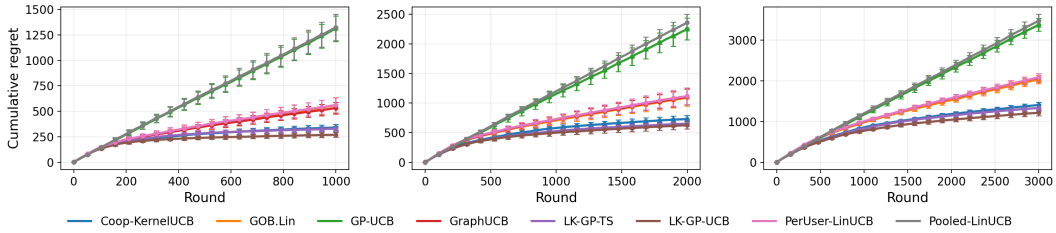


图 3：使用高斯过程抽样时，拉普拉斯核 情形下的累积遗憾。从左到右依次为难度较低、中等和较高的任务。

这一结果意义重大：在 $T \leq Cn$ 的情形下，信息增益既不随 n 也不随 T 增长，˜这意味着 $d = \mathcal{O}(1/\log\ T)$ （缓慢下降）。这种现象在单用户设置中并无对应情况，并证实了在强同质性下，遗憾与 n 无关。我们的图 1 实证图表也印证了这些理论发现。

推广到簇。 如果图包含 k 个内部连通性较高的互不相交簇，$\mathbf{K}_G$ 将有 k 个大小为 $\mathcal{O}(1)$ 的特征值，以及 $n-k$ 个大小为 $\mathcal{O}(1/n)$ 的特征值。类似的论证 ˜ ˜表明，当 $T \leq Cn$时， $d = \mathcal{O}(k/\log\ T)$。因此， d 实质上统计了归一化核 $\mathbf{K}_G$中显著特征值的数量，可作为不同用户簇数量的软性近似。

与独立老虎机的比较。 将此方法与不共享任何信息的独立学习器进行比较很有启发性。由于每个用户平均产生 T/n 个观测，学习每个函数的遗憾至少为 $\sqrt{T/n}$，因此总体遗憾为 $\sum_{u=1}^{n}\sqrt{T/n} = \sqrt{nT}$。在最坏情况下（案例 1），我们的遗憾界 $\tilde{d}\sqrt{T}$ 按 $n\sqrt{T}$ 增长（忽略对数因子），比独立基线宽松 $\sqrt{n}$倍。不过，如果假设动作空间有限且一致有界，我们就能达到 $\sqrt{\tilde{d}T} \asymp \sqrt{nT}$的遗憾界，¯与独立情形下的最优速率相匹配。

在强同质性下，我们的方法优势显而易见。对于处于 $T \asymp n$ 这一情形的独立学习器，遗憾的增长率为 $\sqrt{nT} \asymp T$，这意味着无法进行学习。相比之下，我们证明，在这一情形下，我们的拉普拉斯核化老虎机可实现 $\mathcal{O}(\sqrt{T})$ 的遗憾（忽略对数因子）。当存在 $k = \mathcal{O}(1)$ 个强簇时，也能获得类似的改进。

5 实验

我们在多个合成数据环境中评估拉普拉斯核化老虎机算法 LK-GP-UCB 和 LK-GP-TS。这些环境在已知图上体现用户间的同质性，同时变化奖励结构（线性与非线性）和问题难度。基线算法包括 GraphUCB[7], GoB.Lin[5],COOP-KernelUCB[13], GP-UCB[8], Pooled LinUCB 和逐用户 LinUCB。完整的实现细节见附录F。

环境。 我们首先从 $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$ 中抽样，构建上下文池 $\mathcal{D}$，然后对上下文向量进行归一化。在第 t 轮，我们从 $\mathcal{D}$ 中无放回地抽取 M_t 个不同的物品，并将其作为 $\mathcal{D}_t$ 呈现。我们使用 Erd˝os–Rényi（ER）随机图模型或径向

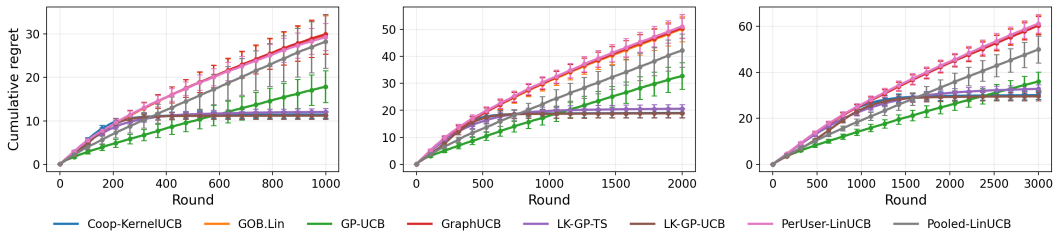


图 4：使用表示子抽样时，拉普拉斯核情形下的累积遗憾。从左到右依次为简单、中等和困难级别的任务。

基函数（RBF）随机图模型生成用户图。生成图后，我们考虑一种线性机制和两种核化（非线性）机制来进行合成数据模拟。第一个合成数据环境称为 *Linear–GOB*。我们考虑模拟真实的图平滑用户参数 $\boldsymbol{\Theta} = (\boldsymbol{I} + \eta \boldsymbol{L})^{-1}\boldsymbol{\Theta}_0$，它们对随机初始参数 $\boldsymbol{\Theta}_0 \in \mathbb{R}^{n\times d}$ [7]施加图同质性。在 *Linear–GOB* 机制中，同质性强度由 η 控制。我们还通过模拟多用户核来生成真实奖励函数，这种机制称为拉普拉斯核机制。我们首先使用平方指数核作为基核 K_x，作用于臂 $\mathcal{U}$，并根据式（2）构建多用户核。接下来，我们设计了两种生成 f 的方式，包括高斯过程抽样和表示子抽样。数据模拟的所有细节见附录 F.1。

任务设计。我们的实验对多臂老虎机任务进行了如下设计，以便进行一般性比较。在这些任务中，奖励噪声设为 $\sigma = 0.1$，用户数量为 $n = 20$。最简单级别的任务是一个 10 臂多臂老虎机问题（$m = 10$），所有用户在每个轮次的可见率为 50%（$M_t = 5$），共进行 $T = 1000$ 个交互轮次。中等难度的任务是一个 20 臂多臂老虎机问题（$m = 20$），所有用户在每个轮次的可见率为 25%（$M_t = 5$），共进行 $T = 3000$ 个交互轮次。困难任务是一个 50 臂多臂老虎机问题（$m = 50$），所有用户在每个轮次的可见率为 10%（$M_t = 5$），共进行 $T = 5000$ 个交互轮次。在图 2、3 和 4 中，从左到右依次为简单、中等和困难级别的任务。

算法配置。我们提出的 LK-GP-UCB 和 LK-GP-TS 分别见附录 E.1 中的算法 1 和算法 2。我们使用式（8）中的实用递归更新和式（3）中的精确更新（采用乔列斯基分解）实现混合更新。详情见附录 F.4。我们对超参数 ν 和 β 进行了调优。对于 COOP-KernelUCB，我们最初选取了五种相似性核 K_z，并开展实验（见附录中的图）以验证逆拉普拉斯矩阵 $\boldsymbol{L}_\rho^{-1}$ 是最佳选择，而经验最大均值差异法也接近最佳选择。在实验中，除非另有说明，K_z 设为经验 MMD 方法，以学习相似性核 K_z。用于 GOB 问题的经典基线算法 GoB.Lin、GraphUCB，以及其余所有基线方法 Pooled LinUCB、逐用户 LinUCB 和 GP-UCB，均为基于 UCB 的算法。我们还对它们的置信界超参数进行了调优。正则化参数 λ 设计为一种调度 $\lambda_t = \lambda_{\text{base}} \cdot S_{\text{spec}} \cdot \frac{T}{T+t}$，其中 S_{spec} 是最小非零特征值与最大特征值之比，而 λ_{base} 经过调优。附录 F.5 讨论了超参数调优。所有方法均在集中式、无延迟设置下运行。

主要发现。我们的算法 LK-GP-UCB 和 LK-GP-TS 在全部 9 个数据环境中均表现稳健。在线性 *GOB*情形下（这是线性赌博机算法的首选设置），我们的算法明显优于大多数基线方法。在拉普拉斯核情形下，我们的算法始终是最佳选择。在 GP 抽样设置中，我们的算法在实验中始终名列前茅。在采用表示子抽样的设置中，LK-GP-UCB 和 LK-GP-TS 能实现次线性遗憾，而大多数基线方法难以实现次线性遗憾。我们认为，由于实现了明显的次线性遗憾，我们提出的算法从长远来看显然能够胜过其他算法。最后，尽管我们通过实证研究为 Coop-KernelUCB 选择了一个最佳方案，并在比较中采用该方案，使 Coop-KernelUCB 的表现名列前茅（接近我们的算法），但 LK-GP-UCB 的表现始终优于 Coop-KernelUCB。

参考文献

[1] Balazs Szorenyi、Róbert Busa-Fekete、István Hegedus、Róbert Ormándi、Márk Jelasity 和 Balázs Kégl。基于八卦协议的分布式随机赌博机算法。载于国际机器学习会议，第 19–27 页。PMLR，2013 年。[2] Peter Landgren、Vaibhav Srivastava 和 Naomi Ehrich Leonard。关于多臂赌博机中的分布式协作决策。载于*2016*年欧洲控制会议（*ECC*），第 243–248 页。IEEE，2016 年。[3] Xueping Gong 和 Jiheng Zhang。利用侧观测和切换约束实现高效的图赌博机学习。载于*AAAI*人工智能会议论文集，第 39 卷，第 16871–16879 页，2025 年。[4] Yao Wang、Jiannan Li、Yue Kang、Shanxing Gao 和 Zhenxin Xiao。结合图信息的广义低秩矩阵上下文赌博机。*arXiv* 预印本 *arXiv:2507.17528*，2025 年。[5] Nicolo Cesa-Bianchi、Claudio Gentile 和 Giovanni Zappella。一群赌博机。《神经信息处理系统进展》，第 26 卷，2013 年。[6] Sharan Vaswani、Mark Schmidt 和 Laks Lakshmanan。基于高斯马尔可夫随机场的一群赌博机。载于人工智能与统计，第 690–699 页。PMLR，2017 年。[7] Kaige Yang、Laura Toni 和 Xiaowen Dong。拉普拉斯正则化图赌博机：算法与理论分析。载于人工智能与统计国际会议，第 3133–3143 页。PMLR，2020 年。[8] Sayak Ray Chowdhury 和 Aditya Gopalan。关于核化多臂赌博机。载于国际机器学习会议，第 844–853 页。PMLR，2017 年。[9] Yihan Du、Wei Chen、Yuko Kuroki 和 Longbo Huang。核赌博机中的协作纯探索。*arXiv* 预印本 *arXiv:2110.15771*，2021 年。[10] Sébastien Bubeck、Ronen Eldan 和 Yin Tat Lee。用于赌博机凸优化的基于核的方法。美国计算机学会会刊（*JACM*），68(4):1–35，2021 年。[11] Chuanhao Li、Huazheng Wang、Mengdi Wang 和 Hongning Wang。面向核化上下文赌博机的通信高效分布式学习。《神经信息处理系统进展》，第 35 卷：19773–19785，2022 年。[12] Xingyu Zhou 和 Bo Ji。带约束的核化多臂赌博机。《神经信息处理系统进展》，第 35 卷：14–26，2022 年。[13] Abhimanyu Dubey 等人。用于协作式多智能体上下文赌博机的核方法。载于国际机器学习会议，第 2740–2750 页。PMLR，2020 年。[14] Alexander J. Smola 和 Risi Kondor。图上的核与正则化。载于计算学习理论与核机器，计算机科学讲义第 2777 卷，计算机科学讲义，第 144–158 页。施普林格，2003 年。[15] Niranjan Srinivas、Andreas Krause、Sham M Kakade 和 Matthias Seeger。赌博机设定下的高斯过程优化：无遗憾与实验设计。*arXiv* 预印本 *arXiv:0912.3995*，2009 年。[16] Andreas Krause 和 Cheng Ong。上下文高斯过程赌博机优化。《神经信息处理系统进展》，第 24 卷，2011 年。[17] Sattar Vakili、Kia Khezeli 和 Victor Picheny。关于高斯过程赌博机中的信息增益与遗憾界。载于人工智能与统计国际会议，第 82–90 页。PMLR，2021 年。[18] William R Thompson。基于两组样本的证据，一个未知概率超过另一个的可能性。*Biometrika*，25(3/4):285–294，1933 年。

[19] Daniel J Russo、Benjamin Van Roy、Abbas Kazerouni、Ian Osband、Zheng Wen 等。Thompson采样教程。《机器学习基础与趋势》，11(1):1–96，2018。[20] Mikhail Belkin、Partha Niyogi 和 Vikas Sindhwani。流形正则化：从有标签和无标签示例中学习的几何框架。《机器学习研究杂志》，7(11)，2006。[21] Tomáš Kocák、Rémi Munos、Branislav Kveton、Shipra Agrawal 和 Michal Valko。谱老虎机。《机器学习研究杂志》，21(218):1–44，2020。[22] Michal Valko、Nathaniel Korda、Rémi Munos、Ilias Flaounas 和 Nelo Cristianini。核化上下文老虎机的有限时间分析。*arXiv*预印本 *arXiv:1309.6869*，2013。[23]Shuang Wu 和 Arash A Amini。图神经汤普森采样。*arXiv*预印本 *arXiv:2406.10686*，2024。[24] Alberto Bietti 和 Julien Mairal。关于神经切线核的归纳偏差。《神经信息处理系统进展》，32，2019。[25] Lin Yang 和 Mengdi Wang。特征空间中的强化学习：矩阵老虎机、核与遗憾界。载于 国际机器学习会议，第10746–10756页。PMLR，2020。[26] Varsha Dani、Thomas P. Hayes 和 Sham M. Kakade。老虎机反馈下的随机线性优化。载于 第*21*届学习理论年度会议（*COLT*）论文集，第355–366页。Omnipress，2008。[27] Vladimir Koltchinskii 和 Evarist Giné。积分算子谱的随机矩阵近似。*Bernoulli*，6(1):113–167，2000。[28] Yasin Abbasi-Yadkori、Dávid Pál 和 Csaba Szepesvári。改进的线性随机老虎机算法。《神经信息处理系统进展》，24，2011。[29] Mauricio A Alvarez、Lorenzo Rosasco、Neil D Lawrence 等。向量值函数的核：综述。《机器学习基础与趋势》，4(3):195–266，2012。[30] Yael Yankelevsky 和 Michael Elad。双图正则化字典学习。*IEEE*《网络信号与信息处理汇刊》，2(4):611–624，2016。[31] Shuang Wu、Chi-Hua Wang、Yuantong Li 和 Guang Cheng。随机线性老虎机的残差自助法探索。载于 人工智能不确定性会议，第2117–2127页。PMLR，2022。[32] Branislav Kveton、Csaba Szepesvari、Mohammad Ghavamzadeh 和 Craig Boutilier。随机线性老虎机中的扰动历史探索。*arXiv*预印本 *arXiv:1903.09132*，2019。[33] Arun Venkitaraman、Saikat Chatterjee 和 Peter Handel。图上的高斯过程。载于*ICASSP 2020-2020 IEEE*国际声学、语音与信号处理会议（*ICASSP*），第5640–5644页。IEEE，2020。[34] Yin-Cong Zhi、Yin Cheng Ng 和 Xiaowen Dong。通过谱核学习实现图上的高斯过程。*IEEE*《网络信号与信息处理汇刊》，9:304–314，2023。[35] David Blanco-Mulero、Markus Heinonen 和 Ville Kyrki。演化图高斯过程。*arXiv*预印本 *arXiv:2106.15127*，2021。[36] Noga Alon、Nicolo Cesa-Bianchi、Claudio Gentile、Shie Mannor、Yishay Mansour 和 Ohad Shamir。具有图结构反馈的非随机多臂老虎机。《*SIAM*计算杂志》，46(6):1785–1826，2017。[37] Qingyun Wu、Huazheng Wang、Quanquan Gu 和 Hongning Wang。协作环境中的上下文多臂老虎机。载于 第*39*届国际*ACM SIGIR*信息检索研究与发展会议论文集，第529–538页，2016。

[38] Ravi Kumar Kolla、Krishna Jagannathan 和 Aditya Gopalan。社交网络上的随机多臂老虎机协作学习。*IEEE/ACM*网络汇刊，26(4):1782–1795，2018。[39] Noga Alon、Nicolo Cesa-Bianchi、Claudio Gentile 和 Yishay Mansour。从多臂老虎机到专家：关于支配与独立性的故事。《神经信息处理系统进展》，26，2013。[40] Fang Liu、Swapna Buccapatnam 和 Ness Shroff。具有图反馈的随机多臂老虎机的信息导向采样。载于 *AAAI*人工智能会议论文集，第32卷，2018。[41] Fang Liu、Zizhan Zheng 和 Ness Shroff。无图条件下图多臂老虎机的Thompson采样分析。*arXiv*预印本 *arXiv:1805.08930*，2018。[42] Raman Arora、Teodor Vanislavov Marinov 和 Mehryar Mohri。具有反馈图和切换成本的多臂老虎机。《神经信息处理系统进展》，32，2019。[43] Shiyin Lu、Yao Hu 和 Lijun Zhang。非平稳环境下具有图反馈的随机多臂老虎机。载于 *AAAI*人工智能会议论文集，第35卷，第8758–8766页，2021。[44] Mengxiao Zhang、Yuheng Zhang、Olga Vrousgou、Haipeng Luo 和 Paul Mineiro。具有反馈图的实用上下文多臂老虎机。《神经信息处理系统进展》，36:30592–30617，2023。[45] Huazheng Wang、Qingyun Wu 和 Hongning Wang。用于交互式推荐的因子分解多臂老虎机。载于 *AAAI*人工智能会议论文集，第31卷，2017。[46] Shuai Li、Alexandros Karatzoglou 和 Claudio Gentile。协同过滤多臂老虎机。载于第*39*届国际 *ACM SIGIR*信息检索研究与发展会议论文集，第539–548页，2016。[47] Ronshee Chawla、Daniel Vial、Sanjay Shakkottai 和 R Srikant。协同多智能体异质多臂老虎机。载于 国际机器学习会议，第4189–4217页。PMLR，2023。[48] Konstantina Christakopoulou 和 Arindam Banerjee。学习与用户交互：一种协同多臂老虎机方法。载于*2018*年*SIAM*国际数据挖掘会议论文集，第612–620页。SIAM，2018。[49] Guanghong Xu。从农场到冰箱：肯尼亚乳制品价值链的数字化可追溯性与质量升级。*SSRN 5722923*，2024。[50] Liang Wang 和 Luis Carvalho。偏差矩阵分解。电子统计学期刊，17(2):3762–3810，2023。[51] Shuang Wu、Mingxuan Zhang、Yuantong Li、Carl Yang 和 Pan Li。通过隐藏表示共享实现图联邦学习。*arXiv*预印本 *arXiv:2212.12158*，2022。[52] Liang Wang 和 Luis Carvalho。指数族因子分析的计算方法。*arXiv*预印本 *arXiv:2403.14925*，2024。[53] Parker Glenn、Parag Dakle、Liang Wang 和 Preethi Raghavan。BlendSQL：一种在关系代数中统一混合式问答的可扩展方言。载于 计算语言学协会论文集：*ACL 2024*，第453–466页，2024。[54] Liang Wang、Ivano Lauriola 和 Alessandro Moschitti。利用排序用户的反馈准确训练基于网络的问答系统。载于 计算语言学协会第*61*届年会论文集（第*5*卷：工业赛道），第660–667页，2023。

相关工作

图上的高斯过程 我们的核构造建立在图正则化的奠基性工作之上。 [14]最初证明，对离散图范数 $\|f\|_L^2 = f^\top \mathbf{L} f$施加惩罚会诱导出一个再生核希尔伯特空间（RKHS），其核为拉普拉斯算子的伪逆。我们的定理 2.1 通过张量积再生核希尔伯特空间，为向量值情形下的这种对偶关系提供了形式化表述。我们指出，这一结构性结果实际上可以从 Alvarez 等人 [29]对向量值函数的全面综述中推知。
根据 [14],，顶点上的任何拉普拉斯算子的函数形式的半正定核，都可以在 $\boldsymbol{L}$ 的特征基下写为 $\boldsymbol{K}_G = \sum_{i=1}^{n} r(\lambda_i)\, \boldsymbol{q}_i \boldsymbol{q}_i^\top$，其中 $\{(\lambda_i, \boldsymbol{q}_i)\}_{i=1}^{n}$是 $\boldsymbol{L}$ 的特征值对， $r(\cdot) \geq 0$ 是谱传递函数。我们的选择 $\boldsymbol{K}_G = (\boldsymbol{L} + \rho \boldsymbol{I})^{-1}$ 对应于$r(\lambda) = 1/(\lambda + \rho)$，它是单调递减的，因此会更强烈地收缩高频分量，从而施加平滑的／同质性先验。原则上，非单调或带通传递函数 r 可以编码更复杂、可能不平滑或呈异配关系的用户间关系；在多臂老虎机设定下分析此类先验，是一个值得探索的未来研究方向。

图信号处理领域较新的研究采用了相关的基于拉普拉斯算子的构造，但并未将诱导再生核希尔伯特空间范数作为分析的主要工具。 [33] 基于拉普拉斯先验构造图上的高斯过程， [34] 则进一步将其推广为学习一个直接作用于拉普拉斯算子的谱滤波器g（$\boldsymbol{L}$）。这两种方法都侧重于批量回归和信号重构；其底层正则项可以根据与 g相关的传递函数，从谱的角度加以刻画，具体而言是 [14]„，但这并非其研究的主要对象。相比之下，我们采用特定的格林函数核 $\boldsymbol{K}_G = （\boldsymbol{L} + \rho \boldsymbol{I}）^{-1}$，它对应于经典的狄利克雷能量正则项，并施加同质性先验。这一选择带来了简单而明确的再生核希尔伯特空间范数，使我们能够在整个分析过程中对其进行追踪，并将其与多用户多臂老虎机设定中的有效维数和遗憾直接联系起来。

图结构多臂老虎机 基于图的多臂老虎机模型也与此相关，但在概念上有所不同。在具有图结构反馈的非随机多臂老虎机中，学习者选择一个臂（节点），并观察反馈图中该臂及其邻居的损失，在全信息与标准多臂老虎机之间形成一种折中 [36]。这类研究中的遗憾界通常随独立数 $\alpha(G)$ 等图论量或相关的可观测性参数 [39] 而变化。后续关于反馈图多臂老虎机和图多臂老虎机的研究进一步改进了这些保证，并将其扩展到随机环境、切换成本、对抗性污染、非平稳环境和上下文多臂老虎机；遗憾由支配数和弱支配数、团覆盖数和独立数，或最大无环子图类的量等参数控制 [40, 41, 42, 43, 44]。在我们的问题设定中，用户图通过拉普拉斯核编码用户价值函数之间的先验相关性；反馈仍严格属于多臂老虎机反馈（我们只能观察所选用户–臂对的奖励）。因此，图仅通过用户核的谱及由此得到的有效维数进入我们的分析，而非通过图多臂老虎机遗憾界中使用的此类侧信息参数。

协作式老虎机 我们的方法与图上的协作式上下文老虎机相关，这类方法利用用户之间的关系加速学习。协作式上下文老虎机 [37]使用用户邻接图在线共享上下文和奖励信息，从而实质上为线性上下文老虎机模型添加了拉普拉斯型正则化项。其他研究则考虑低秩或基于分解的协作式老虎机，例如用于交互式推荐的矩阵分解老虎机 [45] ，以及在老虎机框架中对用户和物品进行协同聚类的协同过滤多臂老虎机 [46]。另一类相关研究探讨社交网络上的多智能体老虎机，其中多个参与者观测或共享彼此的动作和奖励，以降低遗憾[38, 47, 48]。这类方法通常要么（i）采用线性模型并手动选择正则化项，要么（ii）通过潜在因子、聚类或消息传递来建模协作，却没有明确的多输出再生核希尔伯特空间／高斯过程诠释。相比之下，我们的拉普拉斯核化构造为协作提供了有原则的核视角：已知用户图定义了一个正定用户核，并与灵活的上下文核相结合，由此得到的算法，其不确定性量化和遗憾明确取决于图拉普拉斯算子与基核的联合谱，而非用户数、簇数或潜在维度。

协作式多智能体核化老虎机 [13] 研究的是一种在固定图 $G = (V, E)$上、具有延迟通信的协作式多智能体核化上下文老虎机。在其模型中，每个智能体 $v \in V$在每一轮 t都会采取行动，选择一个动作 $x_{v,t}$ 并获得奖励 $y_{v,t}$，因此经过T 轮后共有 $|V|T$ 条观测；图 G 仅用于限制消息传递，并通过图论量（例如图幂的团数）影响遗憾，但并未参与构造智能体之间的相似性核，也不用于奖励函数本身的建模。相反，Dubey 等人针对每个智能体假设一个潜在的"网络上下文" z_v，并假设存在一个全局函数 $F(x, z)$，它属于乘积核 $K((x, z), (x', z')) = K_x(x, x')K_z(z, z')$的再生核希尔伯特空间。当网络上下文（或核 K_z）不可用时，他们提出从上下文 $x_{v,t}$ 中估计这些信息：将每个智能体的上下文分布 P_v 嵌入到 K_x 的再生核希尔伯特空间中，并将 K_z 定义为这些均值嵌入上的径向基函数核。因此，智能体核归根结底是基于（估计得到的）上下文分布学习出的相似性，而底层通信图并未直接参与定义任务相似性，也未用于对 $(f_v)_{v\in V}$施加平滑性惩罚。

相比之下，我们的设定遵循 Gang-of-Bandits 模型：在每个时间步随机抽取一名用户，我们为该用户选择一个动作，并且只观测一个奖励，因此经过 T轮次后，我们得到 T个观测，而不是 $|V|T$。我们还充当集中式学习器，而不是去中心化老虎机网络。最重要的是，我们不引入或估计任何潜在网络上下文；相反，我们假设给定一个用户图，并将智能体核 固定为正则化拉普拉斯矩阵的逆，

$$K_z(u, v) = [\boldsymbol{L}_\rho^{-1}]_{u,v}.$$

该核与奖励函数向量上的全局平滑性惩罚紧密耦合，并产生一个具有明确平滑性解释的再生核希尔伯特空间范数。这种有理论依据的基于图的构造使我们能够对所得的多用户核进行谱分析，将遗憾与 $\boldsymbol{L}_\rho$ 的谱联系起来，并突出有效维数如何适应用户图的簇结构，而不是将网络信息简化为从上下文分布中通过临时方法推断出的潜在特征。

更广泛的应用。虽然我们的重点在于方法论，但用于序贯决策的图结构化共享也与更广泛的场景相关，这些场景涉及关系数据和噪声反馈，包括供应链可追溯性、针对异质结果的 [49], 指数族潜在因子建模[50, 52], , 以及采用结构化表示共享的联邦学习 [51]。此外，自然语言处理系统根据排序用户反馈进行训练 [54] ，以及基于关系数据的混合式问答 [53]，也是相关但相互正交的研究方向。

B 定理 2.1 的证明

证明。证明分为三个主要步骤：(1) 将多用户问题的希尔伯特空间构造为用户空间与上下文空间的张量积；(2) 定义映射到该空间的特征映射，并证明其内积给出核 K。由此可知，我们构造的空间确实是再生核希尔伯特空间 $\mathcal{H}$；(3) 刻画 $\mathcal{H}$ 中的元素，并推导其范数表达式。

步骤 1：通过张量积构造希尔伯特空间。设 $\mathcal{H}_G = \mathbb{R}^n$ 为用户对应的有限维希尔伯特空间，并配备标准欧几里得内积$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathcal{H}_G} = \boldsymbol{u}^\top \boldsymbol{v}$。 $\{\boldsymbol{e}_i\}_{i=1}^n$ 构成 $\mathcal{H}_G$的标准正交基。我们的多用户再生核希尔伯特空间 $\mathcal{H}$是 $\mathcal{H}_G$ 与 $\mathcal{H}_x$的张量积：

$$\mathcal{H} := \mathcal{H}_G \otimes \mathcal{H}_x = \mathbb{R}^n \otimes \mathcal{H}_x.$$

$\mathcal{H}$ 的元素是形如 $\boldsymbol{u} \otimes h$ 的初等张量的有限线性组合（的极限），其中 $\boldsymbol{u} \in \mathcal{H}_U$ 和 $h \in \mathcal{H}_x$。$\mathcal{H}$ 中的内积定义在这些初等张量上，并通过线性性扩展：

$$\langle \boldsymbol{u}_1 \otimes h_1, \boldsymbol{u}_2 \otimes h_2 \rangle_{\mathcal{H}} := \langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle_{\mathcal{H}_G} \langle h_1, h_2 \rangle_{\mathcal{H}_x}.$$

步骤 2：定义特征映射并验证核。 令 $\boldsymbol{L}_\rho^{1/2}$为 $\boldsymbol{L}_\rho$唯一的对称正定平方根。我们将特征映射 ϕ 定义为：$(\mathcal{U} \times \mathcal{D}) \to \mathcal{H}$

$$\phi(\boldsymbol{x}, u) := \left(L_\rho^{-1/2} \boldsymbol{e}_i\right) \otimes \varphi(\boldsymbol{x}).$$

这是 $\mathcal{H}$ 的一个有效元素，因为 $\boldsymbol{L}_\rho^{-1/2}\boldsymbol{e}_i \in \mathbb{R}^n = \mathcal{H}_G$ 且 $\varphi(\boldsymbol{x}) \in \mathcal{H}_x$。让我们计算 $\mathcal{H}$中两个此类特征映射的内积:

$$\begin{aligned}\langle \phi(\boldsymbol{x}, i), \phi(\boldsymbol{y}, j)\rangle_{\mathcal{H}} &= \langle (\boldsymbol{L}_\rho^{-1/2}\boldsymbol{e}_i) \otimes \varphi(\boldsymbol{x}), (\boldsymbol{L}_\rho^{-1/2}\boldsymbol{e}_j) \otimes \varphi(\boldsymbol{y})\rangle_{\mathcal{H}} \\ &= \langle \boldsymbol{L}_\rho^{-1/2}\boldsymbol{e}_i, \boldsymbol{L}_\rho^{-1/2}\boldsymbol{e}_j\rangle_{\mathcal{H}_G} \cdot \langle \varphi(\boldsymbol{x}), \varphi(\boldsymbol{y})\rangle_{\mathcal{H}_x} \\ &= \boldsymbol{e}_i^\top \boldsymbol{L}_\rho^{-1}\boldsymbol{e}_j \cdot K_x(\boldsymbol{x}, \boldsymbol{y}) \\ &= [\boldsymbol{L}_\rho^{-1}]_{ij} \cdot K_x(\boldsymbol{x}, \boldsymbol{y}) = K((\boldsymbol{x}, i), (\boldsymbol{y}, j)).\end{aligned}$$

根据RKHS的基本性质，由于核 K 是由希尔伯特空间 $\mathcal{H}$中、特征映射 ϕ 的内积生成的，因此 $\mathcal{H}$ 是与 K相关联的唯一再生核希尔伯特空间。

步骤 3：刻画 $\mathcal{H}$ 中的函数及其范数。 $\mathcal{H}$ 中的一个元素是函数$f: (\mathcal{U} \times \mathcal{D}) \to \mathbb{R}$。根据里斯表示定理，对于每个 $f \in \mathcal{H}$，都存在唯一的元素 $\boldsymbol{\theta} \in \mathcal{H}$，使得 $f(\cdot, \cdot) = \langle \boldsymbol{\theta}, \phi(\cdot, \cdot)\rangle_{\mathcal{H}}$且 $\|f\|_{\mathcal{H}} = \|\boldsymbol{\theta}\|_{\mathcal{H}}$。对于某些分量函数 $\{g_k\}_{k=1}^n \subset \mathcal{H}_x$，我们可以将 $\boldsymbol{\theta}$唯一地表示为

$$\boldsymbol{\theta} = \sum_{k=1}^n \boldsymbol{e}_k \otimes g_k$$

$\boldsymbol{\theta}$ 在 $\mathcal{H}$ 中的平方范数为:

$$\|\boldsymbol{\theta}\|_{\mathcal{H}}^2 = \Big\langle \sum_k \boldsymbol{e}_k \otimes g_k, \sum_l \boldsymbol{e}_l \otimes g_l \Big\rangle_{\mathcal{H}} = \sum_{k,l} \langle \boldsymbol{e}_k, \boldsymbol{e}_l\rangle_{\mathcal{H}_G} \langle g_k, g_l\rangle_{\mathcal{H}_x} = \sum_{k=1}^n \|g_k\|_{\mathcal{H}_x}^2.$$

接下来，我们可以将奖励函数 $f_{1:n}$ 与分量函数 $\{g_k\}_{k=1}^n$联系起来:

$$\begin{aligned}f_i(\boldsymbol{x}) &= \langle \boldsymbol{\theta}, \phi((\boldsymbol{x}, i))\rangle_{\mathcal{H}} \\ &= \langle \sum_{k=1}^n \boldsymbol{e}_k \otimes g_k, (\boldsymbol{L}_\rho^{-1/2}\boldsymbol{e}_i) \otimes \varphi(\boldsymbol{x})\rangle_{\mathcal{H}} \\ &= \sum_{k=1}^n \langle \boldsymbol{e}_k, \boldsymbol{L}_\rho^{-1/2}\boldsymbol{e}_i\rangle_{\mathcal{H}_G} \cdot \langle g_k, \varphi(\boldsymbol{x})\rangle_{\mathcal{H}_x} \\ &= \sum_{k=1}^n [\boldsymbol{L}_\rho^{-1/2}]_{ki} \cdot \langle g_k, \varphi(\boldsymbol{x})\rangle_{\mathcal{H}_x} \\ &= \sum_{k=1}^n [\boldsymbol{L}_\rho^{-1/2}]_{ki} g_k(\boldsymbol{x}) \quad (\text{since } \langle g, \varphi(\boldsymbol{x})\rangle_{\mathcal{H}_x} = g(\boldsymbol{x}))\end{aligned}$$

由此可得

$$g_k(\boldsymbol{x}) = \sum_{j=1}^n [\boldsymbol{L}_\rho^{1/2}]_{kj} f_j(\boldsymbol{x}).$$

对于作为 $\mathcal{H}_\kappa$中元素的函数，此等式成立： $g_k = \sum_{j=1}^n [L_\rho^{1/2}]_{kj} f_j$。

最后，我们计算 f 在 $\mathcal{H}$中的范数：

$$\begin{aligned}
\|f\|_{\mathcal{H}}^2 = \|\boldsymbol{\theta}\|_{\mathcal{H}}^2 &= \sum_{k=1}^{n} \|g_k\|_{\mathcal{H}_x}^2 \\
&= \sum_{k=1}^{n} \left\| \sum_{j=1}^{n} [\boldsymbol{L}_\rho^{1/2}]_{kj} f_j \right\|_{\mathcal{H}_x}^2 \\
&= \sum_{k=1}^{n} \langle \sum_{j=1}^{n} [\boldsymbol{L}_\rho^{1/2}]_{kj} f_j, \sum_{l=1}^{n} [\boldsymbol{L}_\rho^{1/2}]_{kl} f_l \rangle_{\mathcal{H}_x} \\
&= \sum_{k=1}^{n} \sum_{j,l=1}^{n} [L_\rho^{1/2}]_{kj} [\boldsymbol{L}_\rho^{1/2}]_{kl} \langle f_j, f_l \rangle_{\mathcal{H}_x} \\
&= \sum_{j,l=1}^{n} \left(\sum_{k=1}^{n} [\boldsymbol{L}_\rho^{1/2}]_{kj} [L_\rho^{1/2}]_{kl} \right) \langle f_j, f_l \rangle_{\mathcal{H}_x} \\
&= \sum_{j,l=1}^{n} [\boldsymbol{L}_\rho]_{jl} \langle f_j, f_l \rangle_{\mathcal{H}_x}
\end{aligned}$$

其中，最后一步是因为括号中的项是矩阵乘积 $(\boldsymbol{L}_\rho^{1/2})^\top \boldsymbol{L}_\rho^{1/2} = \boldsymbol{L}_\rho^{1/2} \boldsymbol{L}_\rho^{1/2} = \boldsymbol{L}_\rho$ 的第 (j, l) 个元素。根据极化恒等式， $\mathcal{H}$ 中对应的内积为：

$$\langle f, g \rangle_{\mathcal{H}} := \sum_{i,j=1}^{n} [\boldsymbol{L}_\rho]_{ij} \langle f_i, g_j \rangle_{\mathcal{H}_x}.$$

为说明 $\|f\|_{\mathcal{H}}^2$ 恰好是式 (1) 中的惩罚项，我们展开 $\boldsymbol{L}_\rho = \rho I_n + \boldsymbol{L}$：

$$\begin{aligned}
\|f\|_{\mathcal{H}}^2 &= \sum_{j,l=1}^{n} (\rho \mathbb{I}\{j = l\} + [\boldsymbol{L}]_{jl}) \langle f_j, f_l \rangle_{\mathcal{H}_x} \\
&= \rho \sum_{j=1}^{n} \|f_j\|_{\mathcal{H}_x}^2 + \sum_{j,l=1}^{n} [\boldsymbol{L}]_{jl} \langle f_j, f_l \rangle_{\mathcal{H}_x}.
\end{aligned}$$

利用拉普拉斯二次型的标准恒等式，上式中的第二项恰好是 $\frac{1}{2} \sum_{i,j} w_{ij} \|f_i - f_j\|_{\mathcal{H}_x}^2$，因此得到：

$$\|f\|_{\mathcal{H}}^2 = \rho \sum_{j=1}^{n} \|f_j\|_{\mathcal{H}_x}^2 + \frac{1}{2} \sum_{i,j} w_{ij} \|f_i - f_j\|_{\mathcal{H}_x}^2.$$

至此，证明完成。 □

C 分析中的证明

我们首先定义以下附加记号

$$\boldsymbol{\Phi}_t := [\phi(\boldsymbol{x}_1, u_1), \cdots, \phi(\boldsymbol{x}_t, u_t)]^\top \tag{15}$$

$$\boldsymbol{J}_t := \boldsymbol{\Phi}_t^\top \boldsymbol{\Phi}_t \tag{16}$$

$$\boldsymbol{\Gamma}_t := \boldsymbol{J}_t + \lambda \boldsymbol{I}_\infty \tag{17}$$

$$\boldsymbol{\Sigma}_t := \boldsymbol{K}_t + \lambda \boldsymbol{I}_t \tag{18}$$

这里有 $\boldsymbol{\Phi}_t \in \mathbb{R}^{t \times \infty}$ 和 $\boldsymbol{J}_t$， $\boldsymbol{\Gamma}_t$ 来自 $\mathbb{R}^{\infty \times \infty}$。

接下来，我们定义一些有助于集中性的事件：

$$\mathcal{E}_t^{\text{ts}} = \{|z_t(\boldsymbol{x})| \le \sqrt{2\log(t^2|\mathcal{D}_t|)}, \text{ for all } \boldsymbol{x} \in \mathcal{D}_t\}$$
$$\mathcal{E}_t^{a} = \{\mu_{u_t,t-1}(\boldsymbol{x}_t^*) + \beta_t z_t(\boldsymbol{x}_t^*)\sigma_{u_t,t-1}(\boldsymbol{x}_t^*) > f(\boldsymbol{x}_t^*, u_t)\}$$

其中，$z_t(\boldsymbol{x}) \sim \mathcal{N}(0,1)$ 表示 Thompson采样中的重采样随机性。我们还定义轮次 t 时的置信集：

$$\mathcal{C}_t := \{|\mu_{u_t,t-1}(\boldsymbol{x}_t) - f(\boldsymbol{x}_t, u_t)| \le \beta_t \cdot \sigma_{u_t,t-1}(\boldsymbol{x}_t)\} \tag{19}$$

其中

$$\beta_t := \left(B_\rho + \sqrt{\frac{\sigma^2}{\lambda} \cdot \log\det(\boldsymbol{I}_{t-1} + \lambda^{-1}\boldsymbol{K}_{t-1}) + \frac{2\sigma^2}{\lambda}\log\frac{1}{\delta}}\right).$$

此外，回顾以下有效维数

$$\tilde{d} := \frac{\log\det(\boldsymbol{I}_T + \boldsymbol{K}_T/\lambda)}{\log(1 + TK_{\max}/\lambda)}$$

以及最优性差距的上界：

$$|\Delta_t| \le B_\Delta := 2B_\rho K_{\max}^{1/2}.$$

最后，我们给出以下引理，它们在遗憾分析中经常用到。

引理 C.1 (TS 的集中性). 对所有 $t \in [T]$，都有 $\mathbb{P}_t(\bar{\mathcal{E}}_t^{ts}) \le t^{-2}$ 且 $\mathbb{P}_t(\mathcal{E}_t^a|\mathcal{C}_t) \ge (4e\sqrt{\pi})^{-1}$。

引理 C.2 （TS 的单步遗憾界）。 假设 $\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^{ts}) > 0$。那么对于任意 t，几乎必然有，

$$\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{C}_t}] \le \mathbb{I}_{\mathcal{C}_t} \cdot \left\{\left(\frac{2}{\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^{ts})} + 1\right) \cdot \mathbb{E}_t[\gamma_t \sigma_{u_t,t-1}(\boldsymbol{x}_t)] + B_\Delta \cdot \mathbb{P}_t(\bar{\mathcal{E}}_t^{ts})\right\}$$

其中 $\gamma_t := \beta_t + \beta_t\sqrt{2\log(t^2|\mathcal{D}_t|)}$ 以及 $B_\Delta := 2B_\rho K_{\max}^{1/2}$

引理 C.3 （累积不确定性界）。 我们得到累积估计不确定性的上界：

$$\sum_{t=1}^{T} \sigma_{u_t,t-1}(\boldsymbol{x}_t) \le \sqrt{2T\max\{1, K_{\max}\} \cdot \log\det(\boldsymbol{I}_T + \lambda^{-1}\boldsymbol{K}_T)}$$

引理 C.4 （对偶恒等式）。 使用 （15） 中定义的记号，我们有两个关键恒等式 ：

$$\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Phi}_t = \boldsymbol{\Phi}_t\boldsymbol{\Gamma}_t^{-1}, \text{ and } \sigma_{u,t}^2(\boldsymbol{x}) = \lambda\|\phi(\boldsymbol{x}, u)\|_{\boldsymbol{\Gamma}_t^{-1}}^2.$$

C.1 置信集的证明

定理 *4.1* 的证明。我们首先进行分解

$$\begin{aligned}
\mu_{u,t}(\boldsymbol{x}) - f(\boldsymbol{x}, u) &= \boldsymbol{k}_t(\boldsymbol{x}, u)^\top\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\Phi}_t\boldsymbol{\theta} + \boldsymbol{\epsilon}_t) - \boldsymbol{\theta}^\top\phi(\boldsymbol{x}, u) \\
&= (\boldsymbol{\Phi}_t^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{k}_t(\boldsymbol{x}, u))^\top\boldsymbol{\theta} + \boldsymbol{k}_t(\boldsymbol{x}, u)^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\epsilon}_t - \boldsymbol{\theta}^\top\phi(\boldsymbol{x}, u) \\
&= \underbrace{\langle\boldsymbol{\theta}, \delta_t(\boldsymbol{x}, u)\rangle}_{\text{bias}_t(\boldsymbol{x},u)} + \underbrace{\boldsymbol{k}_t(\boldsymbol{x}, u)^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\epsilon}_t}_{\text{noise}_t(\boldsymbol{x},u)}
\end{aligned}$$

其中 $\delta_t(\boldsymbol{x}, u) = \boldsymbol{\Phi}_t^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{k}_t(\boldsymbol{x}, u) - \phi(\boldsymbol{x}, u) \in \ell^2$。我们的目标是界定偏差$_t(\boldsymbol{x}, u)$ 和噪声$_t(\boldsymbol{x}, u)$。我们陈述以下引理：

引理 C.5 (偏差恒等式). 偏差平方就是噪声的退化方差：

$$\|\delta_t(\boldsymbol{x}, u)\|_{\ell^2}^2 = \sigma_{u,t}^2(\boldsymbol{x}) - \lambda\boldsymbol{k}_t(\boldsymbol{x}, u)^\top\boldsymbol{\Sigma}_t^{-2}\boldsymbol{k}_t(\boldsymbol{x}, u) \tag{20}$$

特别地，我们有 $\|\delta_t(\boldsymbol{x}, u)\|_{\ell^2} \le \sigma_{u,\,t}(\boldsymbol{x})$ 以及 $\lambda\boldsymbol{k}_t(\boldsymbol{x}, u)^\top\boldsymbol{\Sigma}_t^{-2}\boldsymbol{k}_t(\boldsymbol{x}, u) < \sigma_{u,\,t}^2(\boldsymbol{x})$。

引理 C.6 (噪声界)。 以高概率，我们有以下噪声向量 $\boldsymbol{\epsilon}_t$ 的范数上界：

$$\|\boldsymbol{\Phi}_t\boldsymbol{\epsilon}_t\|_{\boldsymbol{\Gamma}_t^{-1}} \leq \sqrt{\sigma^2 \log\det(\boldsymbol{I}_t + \lambda^{-1}\boldsymbol{K}_t) + 2\sigma^2 \log\frac{1}{\delta}}$$

根据引理 C.5，我们可以对偏差进行如下界定：

$$\text{bias}_t(\boldsymbol{x}, u) \leq \|\boldsymbol{\theta}\|_{\ell^2}\|\delta_t(\boldsymbol{x}, u)\|_{\ell^2} \leq B_\rho \sigma_{u,t}(\boldsymbol{x}). \tag{21}$$

利用上述引理 C.4 中的恒等式，我们注意到：

$$\begin{aligned}
\text{noise}_t(\boldsymbol{x}, u) &= \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\epsilon}_t \\
&= \phi(\boldsymbol{x}, u)^\top \boldsymbol{\Gamma}_t^{-1}\boldsymbol{\Phi}_t\boldsymbol{\epsilon}_t \\
&= \langle \phi(\boldsymbol{x}, u), \boldsymbol{\Phi}_t\boldsymbol{\epsilon}_t\rangle_{\boldsymbol{\Gamma}_t^{-1}} \\
&\leq \|\phi(\boldsymbol{x}, u)\|_{\boldsymbol{\Gamma}_t^{-1}} \cdot \|\boldsymbol{\Phi}_t\boldsymbol{\epsilon}_t\|_{\boldsymbol{\Gamma}_t^{-1}} \\
&= \frac{\sigma_{u,t}(\boldsymbol{x})}{\sqrt{\lambda}} \cdot \|\boldsymbol{\Phi}_t\boldsymbol{\epsilon}_t\|_{\boldsymbol{\Gamma}_t^{-1}}
\end{aligned}$$

其中，不等式由内积的柯西-施瓦茨不等式得出 $\langle\cdot,\ \cdot\rangle_{\boldsymbol{\Gamma}_t^{-1}}$。

我们的引理 C.6 给出了范数 $\|\boldsymbol{\Phi}_t\boldsymbol{\epsilon}_t\|_{\boldsymbol{\Gamma}_t^{-1}}$ 的高概率上界，因此可得：

$$\text{noise}_t(\boldsymbol{x}, u) \leq \frac{\sigma_{u,t}(\boldsymbol{x})}{\sqrt{\lambda}} \cdot \sqrt{\sigma^2 \log\det(\boldsymbol{I}_t + \lambda^{-1}\boldsymbol{K}_t) + 2\sigma^2 \log\frac{1}{\delta}} \tag{22}$$

现在将 (21) 和 (22) 结合起来，可得：

$$\begin{aligned}
|\mu_{u,t}(\boldsymbol{x}) - f(\boldsymbol{x}, u)| &\leq |\text{bias}_t(\boldsymbol{x}, u)| + |\text{noise}_t(\boldsymbol{x}, u)| \\
&\leq \sigma_{u,t}(\boldsymbol{x})\left(B_\rho + \sqrt{\frac{\sigma^2}{\lambda} \cdot \log\det(\boldsymbol{I}_t + \lambda^{-1}\boldsymbol{K}_t) + \frac{2\sigma^2}{\lambda}\log\frac{1}{\delta}}\right).
\end{aligned}$$

□

C.2 LK-GP-UCB 遗憾界的证明

定理 *4.2* 的证明。 回顾时刻 t 的瞬时遗憾为 $\Delta_t = f(\boldsymbol{x}_t^*,\ u_t) - f(\boldsymbol{x}_t,\ u_t)$，时间范围 T 内的累积遗憾为 $\mathcal{R}_T = \sum_{t=1}^T \Delta_t$。
根据定理，事件 $\mathcal{C}_t := \{|\mu_{u_t,t-1}(\boldsymbol{x}_t)$
$f\ \boldsymbol{x}_t\ u_t\ | \leq \beta_t \cdot \sigma_{u_t\ t-1}\ \boldsymbol{x}_t\ \}$ $1-\delta$ 4 1, ¯ (,). ( ) 以高概率发生 ( )。

$$\beta_t := \left(B_\rho + \sqrt{\frac{\sigma^2}{\lambda} \cdot \log\det(\boldsymbol{I}_{t-1} + \lambda^{-1}\boldsymbol{K}_{t-1}) + \frac{2\sigma^2}{\lambda}\log\frac{1}{\delta}}\right) \tag{23}$$

根据定理 4.1，对于所有 $t \geq 2$，其概率至少为 $1-\delta$，

$$\begin{aligned}
\Delta_t = f(\boldsymbol{x}_t^*, u_t) - f(\boldsymbol{x}_t, u_t) &\leq \mu_{u_t,t-1}(\boldsymbol{x}_t^*) + \beta_t\sigma_{u_t,t-1}(\boldsymbol{x}_t^*) - f(\boldsymbol{x}_t, u_t) \\
&\leq \mu_{u_t,t-1}(\boldsymbol{x}_t) + \beta_t\sigma_{u_t,t-1}(\boldsymbol{x}_t) - f(\boldsymbol{x}_t, u_t) \\
&\leq 2\beta_t\sigma_{u_t,t-1}(\boldsymbol{x}_t).
\end{aligned}$$

因此，我们得到了累积遗憾的高概率界

$$\mathcal{R}_T \leq 2\mathbb{E}\Big[\beta_t \sum_{t=2}^T \sigma_{u_t,t-1}(\boldsymbol{x}_t)\Big] + B_\Delta.$$

接着，我们应用引理 C.3 和式（11）中的有效维数定义。

$$\begin{aligned}
\sum_{t=1}^T \sigma_{u_t,t-1}(\boldsymbol{x}_t) &\leq \sqrt{2T\max\{1, K_{\max}\} \cdot \log\det(\boldsymbol{I}_T + \lambda^{-1}\boldsymbol{K}_T)} \\
&= \sqrt{2T\max\{1, K_{\max}\} \cdot \tilde{d}\log(1 + T\lambda^{-1}K_{\max})}.
\end{aligned}$$

因此，我们得到了遗憾的最终高概率上界：

$$\mathcal{R}_T \leq 2\mathbb{E}[\beta_T]\sqrt{2T\max\{1, K_{\max}\}\cdot \tilde{d}\log(1+T\lambda^{-1}K_{\max})} + B_\Delta.$$

˜下一步是分析上界的阶。再次利用有效维数 d 并忽略常数项，我们得到

$$\beta_t \leq B_\rho + \sqrt{\frac{\sigma^2}{\lambda}\cdot \tilde{d}\log(1+T\lambda^{-1}K_{\max}) + \frac{2\sigma^2}{\lambda}\log\frac{1}{\delta}} = \mathcal{O}(\sqrt{\tilde{d}\log(T)})$$
$$\Rightarrow \mathcal{R}_T = \mathcal{O}(\tilde{d}\log(T)\sqrt{T}) = \tilde{\mathcal{O}}(\tilde{d}\sqrt{T}).$$

□

C.3 LK-GP-TS遗憾界的证明

定理 *4.3* 的证明。我们从累积遗憾的分解开始

$$\mathcal{R}_T = \sum_{t=1}^{T}\mathbb{E}[\Delta_t] = \sum_{t=1}^{T}\mathbb{E}[\Delta_t \mathbb{I}_{\mathcal{C}_t}] + \sum_{t=1}^{T}\mathbb{E}[\Delta_t \mathbb{I}_{\bar{\mathcal{C}}_t}].$$

By 根据定理 4.1 和最优性差距的上界，我们知道第二项有界 :

$$\sum_{t=1}^{T}\mathbb{E}[\Delta_t \mathbb{I}_{\bar{\mathcal{C}}_t}] \leq \delta B_\Delta$$

在定理 4.1 中，对所有 t 令 $\mathbb{P}(\mathcal{C}_t) \leq \delta/T$。

对于事件 $\mathcal{C}_t$ 上的遗憾，根据引理 C.2，几乎必然有

$$\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{C}_t}] \leq \mathbb{I}_{\mathcal{C}_t}\cdot\left\{\left(\frac{2}{\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^{\text{ts}})} + 1\right)\cdot \mathbb{E}_t[\gamma_t \sigma_{u_t,t-1}(\boldsymbol{x}_t)] + B_\Delta \cdot \mathbb{P}_t(\bar{\mathcal{E}}_t^{\text{ts}})\right\}$$

其中 $\gamma_t := \beta_t + \beta_t\sqrt{2\log(t^2|\mathcal{D}_t|)}$。注意，根据引理 C.1 以及对所有 $t \geq 5$ 均有 $t^2 \geq 5e\sqrt{\pi}$ 这一事实，$\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^{\text{ts}}) \geq \frac{1}{4e\sqrt{\pi}} - \frac{1}{t^2} \geq \frac{1}{20e\sqrt{\pi}}$。因此我们有

$$\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{C}_t}] \leq \mathbb{I}_{\mathcal{C}_t}\cdot\left\{194\mathbb{E}_t[\gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}_t)] + B_\Delta t^{-2}\right\}$$

by 利用 $40e\sqrt{\pi}+1 \leq 194$。对目标累积遗憾两边求和，我们得到 t

$$\begin{aligned}
\sum_{t=1}^{T}\mathbb{E}[\Delta_t \mathbb{I}_{\mathcal{C}_t}] &= \mathbb{E}[\sum_{t=1}^{T}\mathbb{E}_t[\Delta_t \mathbb{I}_{\mathcal{C}_t}]] \\
&\leq \mathbb{E}[\sum_{t=5}^{T}\left(194\mathbb{E}_t[\gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}_t)] + B_\Delta t^{-2}\right) + 4B_\Delta] \\
&\leq \mathbb{E}[194\sum_{t=5}^{T}\mathbb{E}_t[\gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}_t)] + (4+\frac{\pi^2}{6})B_\Delta] \\
&\leq \mathbb{E}[194\gamma_T\mathbb{E}_t[\sum_{t=1}^{T}\sigma_{u_t,t-1}(\boldsymbol{x}_t)] + (4+\frac{\pi^2}{6})B_\Delta]
\end{aligned}$$

其中，第二个等式利用了 $\sum_{t=1}^{\infty} t^{-2} = \pi^2/6$，最后一步则源于 γ_t 的单调性以及 σ_u 的非负性. $_t(\boldsymbol{x})$。接下来我们将重点对不确定性的求和进行界定。与定理 4.2 的证明采用相同的方法，我们应用引理 C.3 以及式 (11) 中有效维数的定义

$$\begin{aligned}
\sum_{t=1}^{T}\sigma_{u_t,t-1}(\boldsymbol{x}_t) &\leq \sqrt{2T\max\{1,K_{\max}\}\cdot\log\det(\boldsymbol{I}_T + \lambda^{-1}\boldsymbol{K}_T)} \\
&= \sqrt{2T\max\{1,K_{\max}\}\cdot\tilde{d}\log(1+T\lambda^{-1}K_{\max})}.
\end{aligned}$$

因此，我们有

$$\sum_{t=1}^{T}\mathbb{E}[\Delta_t\mathbb{I}_{\mathcal{C}_t}] \le 194\mathbb{E}[\gamma_T]\sqrt{2T\max\{1,K_{\max}\}\cdot\tilde{d}\log(1+T\lambda^{-1}K_{\max})} + (4+\frac{\pi^2}{6})B_\Delta$$

从而得到高概率（$1-\delta$）遗憾上界:

$$\mathcal{R}_T \le 194\mathbb{E}[\gamma_T]\sqrt{2T\max\{1,K_{\max}\}\cdot\tilde{d}\log(1+T\lambda^{-1}K_{\max})} + (4+\frac{\pi^2}{6})B_\Delta + \delta B_\Delta.$$

关于上界的阶，我们首先利用有效维数的定义 $\mathbb{E}[\gamma_T]$，再次分析d，并忽略常数项

$$\begin{aligned}\gamma_T &\le \left(1+\sqrt{2\log(T^2M)}\right)\cdot\left(B_\rho+\sqrt{\frac{\sigma^2}{\lambda}\cdot\tilde{d}\log(1+T\lambda^{-1}K_{\max})+\frac{2\sigma^2}{\lambda}\log\frac{1}{\delta}}\right)\\ &= \mathcal{O}(\log(T)\sqrt{\tilde{d}}).\end{aligned}$$

其中， M是时刻 t动作集大小的上界，即对于所有 $t\le T$，均有 $|\mathcal{D}_t|\le M$。因此，

$$\mathcal{R}_T = \mathcal{O}(\tilde{d}\log(T)^{3/2}\sqrt{T}) = \tilde{\mathcal{O}}(\tilde{d}\sqrt{T}).$$

□

D 引理的证明

D.1 引理 C.1 的证明

证明。利用标准的高斯尾界和经典的并集界，我们有

$$\mathbb{P}_t(|z_t(\boldsymbol{x})|>u)\le|\mathcal{D}_t|e^{-u^2/2}.$$

令 $u=\sqrt{2\log(t^2|\mathcal{D}_t|)}$，我们得到 $\mathbb{P}_t(\bar{\mathcal{E}}_t^{ts})\le t^{-2}$。

对于事件 $\mathcal{E}_t^a$的结果，我们有

$$\begin{aligned}\mathbb{P}_t\left(\mu_{u_t,t-1}(\boldsymbol{x}_t^*)+\beta_t z_t(\boldsymbol{x}_t^*)\sigma_{u_t,t-1}(\boldsymbol{x}_t^*)>f(\boldsymbol{x}_t^*,u_t)|\mathcal{C}_t\right) &= \mathbb{P}_t\left(z_t(\boldsymbol{x}_t^*)>\frac{f(\boldsymbol{x}_t^*,u_t)-\mu_{u_t,t-1}(\boldsymbol{x}_t^*)}{\beta_t\sigma_{u_t,t-1}(\boldsymbol{x}_t^*)}|\mathcal{C}_t\right)\\ &\ge \mathbb{P}_t(z_t(\boldsymbol{x}_t^*)>1)\\ &\ge (4e\sqrt{\pi})^{-1}\end{aligned}$$

其中，第一个不等式源于 $\mathcal{C}_t$ 成立这一事实，最后一步直接由以下事实得到： $\mathbb{P}(Z\ge1)\ge(4e\sqrt{\pi})^{-1}$ 对于 $Z\sim\mathcal{N}(0,1)$。

□

D.2 引理 C.2 的证明

证明。 该证明遵循 Thompson采样算法 [32, 31, 23]的经典分析。我们首先回顾 $\mathbb{E}_t[\cdot]=\mathbb{E}[\cdot|\mathcal{F}_t]$。给定来自历史记录 $\mathcal{F}_t$的随机性，事件 $\mathcal{C}_t$ 变为确定性的，随机性仅来自重新采样步骤。因此我们有

$$\begin{aligned}\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{C}_t}] &= \mathbb{I}_{\mathcal{C}_t}\cdot\mathbb{E}_t[\Delta_t]\\ &= \mathbb{I}_{\mathcal{C}_t}\cdot\left(\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{E}_t^{\text{ts}}}]+\mathbb{E}_t[\Delta_t\mathbb{I}_{\bar{\mathcal{E}}_t^{\text{ts}}}]\right)\\ &\le \mathbb{I}_{\mathcal{C}_t}\cdot\left(\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{E}_t^{\text{ts}}}]+B_\Delta\cdot\mathbb{P}_t(\bar{\mathcal{E}}_t^{\text{ts}})\right)\end{aligned}$$

其中，最后一步源于最优性差距 $\Delta_t\le B_\Delta$的有界性。接下来我们将重点界定 $\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{E}_t^{\text{ts}}}]$，这表明在证明的其余部分 $\mathcal{C}_t$ 成立。

We 然后定义“最不确定的采样不足”动作这一概念，称为未饱和动作 d acti，定义为

$$\mathcal{U}_t := \{\boldsymbol{x}\in\mathcal{D}_t : f(\boldsymbol{x}_t^*,u_t) < f(\boldsymbol{x},u_t)+\gamma_t\sigma_{u_t,t-1}(\boldsymbol{x})\}$$

其中

$$\gamma_t := \beta_t + \beta_t\sqrt{2\log(t^2|\mathcal{D}_t|)}$$

并令 $\bar{\boldsymbol{x}}_t$ 为时刻 t不确定性最小的未饱和动作：

$$\bar{\boldsymbol{x}}_t = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{U}_t} \gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}).$$

回顾重采样索引的记号：$\tilde{\mu}_t(\boldsymbol{x}) = \mu_{u_t,\,t-1}(\boldsymbol{x}) + \beta_t z_t(\boldsymbol{x})\sigma_{u_t,\,t-1}(\boldsymbol{x})$。在良好情形 $\mathcal{C}_t \cap \mathcal{E}_t^{\text{ts}}$下，我们有

$$|\tilde{\mu}_t(\boldsymbol{x}) - f(\boldsymbol{x}, u_t)| \le |\tilde{\mu}_t(\boldsymbol{x}) - \mu_{u_t,t-1}(\boldsymbol{x})| + |\mu_{u_t,t-1}(\boldsymbol{x}) - f(\boldsymbol{x}, u_t)| \le \gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}).$$

因此，我们可以给出遗憾的一个初始上界

$$\begin{aligned}
\Delta_t &= f(\boldsymbol{x}_t^*, u_t) - f(\boldsymbol{x}_t, u_t) \\
&= f(\boldsymbol{x}_t^*, u_t) - f(\bar{\boldsymbol{x}}_t, u_t) + f(\bar{\boldsymbol{x}}_t, u_t) - f(\boldsymbol{x}_t, u_t) \\
&\le \gamma_t\sigma_{u_t,t-1}(\bar{\boldsymbol{x}}_t) + f(\bar{\boldsymbol{x}}_t, u_t) - f(\boldsymbol{x}_t, u_t) + \tilde{\mu}_t(\boldsymbol{x}_t) - \tilde{\mu}_t(\boldsymbol{x}_t) \quad (\text{by } \bar{\boldsymbol{x}}_t \in \mathcal{U}_t) \\
&\le 2\gamma_t\sigma_{u_t,t-1}(\bar{\boldsymbol{x}}_t) + \gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}_t) + \tilde{\mu}_t(\bar{\boldsymbol{x}}_t) - \tilde{\mu}_t(\boldsymbol{x}_t) \quad (\text{ since } \mathcal{C}_t \cap \mathcal{E}_t^{\text{ts}}) \\
&\le 2\gamma_t\sigma_{u_t,t-1}(\bar{\boldsymbol{x}}_t) + \gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}_t) \quad (\text{ by } \tilde{\mu}_t(\bar{\boldsymbol{x}}_t) < \tilde{\mu}_t(\boldsymbol{x}_t)).
\end{aligned} \tag{24}$$

注意，

$$\gamma_t\sigma_{u_t,t-1}(\bar{\boldsymbol{x}}_t)\mathbb{I}\{\boldsymbol{x}_t \in \mathcal{U}_t\} \le \gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}_t)$$

并在两边乘以 $\mathbb{I}_{\mathcal{E}_t^{\text{ts}}}$后取 $\mathbb{E}_t[\cdot]$，可得

$$\sigma_{u_t,t-1}(\bar{\boldsymbol{x}}_t)\mathbb{P}_t(\{\boldsymbol{x}_t \in \mathcal{U}_t\} \cap \mathcal{E}_t^{\text{ts}}) \le \mathbb{E}_t[\sigma_{u_t,t-1}(\boldsymbol{x}_t)\mathbb{I}_{\mathcal{E}_t^{\text{ts}}}].$$

因此，接下来只需从下方界定概率 $\mathbb{P}_t(\{\boldsymbol{x}_t \in \mathcal{U}_t\} \cap \mathcal{E}_t^{\text{ts}})$。

我们注意到以下两个事实。首先，如果对所有 $\boldsymbol{x} \in \mathcal{U}_t$ 都有 $\tilde{\mu}_t(\boldsymbol{x}_t^*) > \tilde{\mu}_t(\boldsymbol{x})$，那么 $\boldsymbol{x}_t$ 必属于 $\mathcal{U}_t$，这意味着 $\{\tilde{\mu}_t(\boldsymbol{x}_t^*) > \max_{\boldsymbol{x}\in\bar{\mathcal{U}}_t} \tilde{\mu}_t(\boldsymbol{x})\} \subseteq \{\boldsymbol{x}_t \in \mathcal{U}_t\}$。其次，对于任意 $\boldsymbol{x} \in \mathcal{U}_t$，在良好情形 $\mathcal{C}_t \cap \mathcal{E}_t^{\text{ts}} \cap \mathcal{E}_t^a$ 下，我们有

$$\tilde{\mu}_t(\boldsymbol{x}) \le f(\boldsymbol{x}, u_t) + \gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}) \le f(\boldsymbol{x}_t^*, u_t) < \tilde{\mu}_t(\boldsymbol{x}^*)$$

这意味着 $\mathcal{E}_t^a \subseteq \{\tilde{\mu}_t(\boldsymbol{x}_t^*) > \max_{\boldsymbol{x}\in\bar{\mathcal{U}}_t} \tilde{\mu}_t(\boldsymbol{x})\}$

因此，在事件 $\mathcal{C}_t$ 上，我们有

$$\begin{aligned}
\mathbb{P}_t(\{\boldsymbol{x}_t \in \mathcal{U}_t\} \cap \mathcal{E}_t^{\text{ts}}) &\ge \mathbb{P}_t(\{\tilde{\mu}_t(\boldsymbol{x}_t^*) > \max_{\boldsymbol{x}\in\bar{\mathcal{U}}_t} \tilde{\mu}_t(\boldsymbol{x})\} \cap \mathcal{E}_t^{\text{ts}}) \\
&\ge \mathbb{P}_t(\mathcal{E}_t^a \cap \mathcal{E}_t^{\text{ts}}) \\
&\ge \mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^{\text{ts}})
\end{aligned}$$

现在我们得到了 $\sigma_{u_t,t-1}(\bar{\boldsymbol{x}}_t)$ 的一个上界：

$$\sigma_{u_t,t-1}(\bar{\boldsymbol{x}}_t) \le \frac{\mathbb{E}_t[\sigma_{u_t,t-1}(\boldsymbol{x}_t)\mathbb{I}_{\mathcal{E}_t^{\text{ts}}}]}{\mathbb{P}_t(\{\boldsymbol{x}_t \in \mathcal{U}_t\} \cap \mathcal{E}_t^{\text{ts}})} \le \frac{\mathbb{E}_t[\sigma_{u_t,t-1}(\boldsymbol{x}_t)]}{\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^{\text{ts}})}$$

将上述结果代入 (24)，即可得到瞬时遗憾的上界：

$$\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{E}_t^{\text{ts}}}] \le \left(\frac{2}{\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^{\text{ts}})} + 1\right) \cdot \mathbb{E}_t[\gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}_t)].$$

因此，

$$\mathbb{E}_t[\Delta_t\mathbb{I}_{\mathcal{C}_t}] \le \mathbb{I}_{\mathcal{C}_t} \cdot \left\{\left(\frac{2}{\mathbb{P}_t(\mathcal{E}_t^a) - \mathbb{P}_t(\bar{\mathcal{E}}_t^{\text{ts}})} + 1\right) \cdot \mathbb{E}_t[\gamma_t\sigma_{u_t,t-1}(\boldsymbol{x}_t)] + B_\Delta \cdot \mathbb{P}_t(\bar{\mathcal{E}}_t^{\text{ts}})\right\}$$

□

D.3 引理 C.3 的证明

证明。 我们首先应用柯西-施瓦茨不等式，得到

$$\sum_{t=1}^{T}\sigma_{u_t,t-1}(\boldsymbol{x}_t)\le\sqrt{T\sum_{t=1}^{T}\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}=\sqrt{\lambda T\sum_{t=1}^{T}\frac{\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}{\lambda}}.$$

若 $\lambda\ge K_{\max}$，利用 $\sigma^2_{u_t,\ t-1}(\boldsymbol{x}_t)\ \le|K((\boldsymbol{x}_t,\ \ u_t)(\boldsymbol{x}_t,\ \ u_t))|\le K_{\max}$，可知 $\frac{\sigma^2_{u_t,\ t-1}(\boldsymbol{x}_t)}{\lambda}\le 1$，从而得到

$$\sum_{t=1}^{T}\frac{\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}{\lambda}\le 2\sum_{t=1}^{T}\log\Big(1+\frac{1}{\lambda}\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)\Big)\le\frac{2K_{\max}}{\lambda}\sum_{t=1}^{T}\log\Big(1+\frac{1}{\lambda}\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)\Big)$$

利用以下事实：若 $x\le 1$，则 $x\le 2\log(1+x)$。

若 $\lambda\le K_{\max}$，仍利用 $\sigma^2_{u_t,\ t-1}(\boldsymbol{x}_t)\ \le|K((\boldsymbol{x}_t,\ \ u_t)(\boldsymbol{x}_t,\ \ u_t))|\le K_{\max}$，可知

$$\frac{\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}{\lambda}\le\min\{\frac{K_{\max}}{\lambda},\frac{\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}{\lambda}\}\le\frac{K_{\max}}{\lambda}\min\{1,\frac{\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}{\lambda}\}$$

从而得到

$$\sum_{t=1}^{T}\frac{\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}{\lambda}\le\frac{K_{\max}}{\lambda}\sum_{t=1}^{T}\min\{1,\frac{1}{\lambda}\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)\}\le\frac{2K_{\max}}{\lambda}\sum_{t=1}^{T}\log\Big(1+\frac{1}{\lambda}\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)\Big).$$

利用以下事实：对于 $x\ge 0$，$\min\{1,x\}\le 2\log(1+x)$。

我们可以将上述关于 λ 的两个条件归纳起来，得到

$$\sum_{t=1}^{T}\sigma_{u_t,t-1}(\boldsymbol{x}_t)\le\sqrt{2T\max\{1,K_{\max}\}\sum_{t=1}^{T}\log\Big(1+\frac{1}{\lambda}\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)\Big)}.\tag{25}$$

现在我们可以利用 K_t 的舒尔补性质：

$$\begin{aligned}\det\Big(\boldsymbol{I}_t+\frac{1}{\lambda}\boldsymbol{K}_t\Big)=&\det\Big(\boldsymbol{I}_{t-1}+\frac{1}{\lambda}\boldsymbol{K}_{t-1}\Big)\\&\times\Big[1+\frac{1}{\lambda}\big(\underbrace{K((\boldsymbol{x}_t,u_t),(\boldsymbol{x}_t,u_t))-\boldsymbol{k}_{t-1}(\boldsymbol{x}_t,u_t)^\top(\boldsymbol{K}_{t-1}+\lambda\boldsymbol{I})^{-1}\boldsymbol{k}_{t-1}(\boldsymbol{x}_t,u_t)}_{\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)}\big)\Big]\end{aligned}$$

由此可得

$$\sum_{t=1}^{T}\log\Big(1+\frac{1}{\lambda}\sigma^2_{u_t,t-1}(\boldsymbol{x}_t)\Big)=\sum_{t=1}^{T}\log\frac{\det\big(\boldsymbol{I}_t+\frac{1}{\lambda}\boldsymbol{K}_t\big)}{\det\big(\boldsymbol{I}_{t-1}+\frac{1}{\lambda}\boldsymbol{K}_{t-1}\big)}=\log\det\big(\boldsymbol{I}_T+\lambda^{-1}\boldsymbol{K}_T\big).$$

因此，将上述结果与式 (25) 结合，可得

$$\sum_{t=1}^{T}\sigma_{u_t,t-1}(\boldsymbol{x}_t)\le\sqrt{2T\max\{1,K_{\max}\}\cdot\log\det(\boldsymbol{I}_T+\lambda^{-1}\boldsymbol{K}_T)}$$

□

D.4 引理 C.5 的证明

证明。 我们注意到

$$\|\delta_t(\boldsymbol{x},u)\|^2_{\ell^2}=\|\phi((\boldsymbol{x},u))\|^2_{\ell^2}+\Big\|\boldsymbol{\Phi}_t^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{k}_t(\boldsymbol{x},u)\Big\|^2_{\ell^2}-2\langle\phi((\boldsymbol{x},u)),\boldsymbol{\Phi}_t^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{k}_t(\boldsymbol{x},u)\rangle_{\ell^2}$$

并且我们有

$$\begin{aligned}\left\|\boldsymbol{\Phi}_t^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u)\right\|_{\ell^2}^2 &= \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Phi}_t \boldsymbol{\Phi}_t^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u) \\ &= \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{K}_t \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u) \\ &= \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u) - \lambda \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-2} \boldsymbol{k}_t(\boldsymbol{x}, u) \\ &= \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u) - \lambda \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-2} \boldsymbol{k}_t(\boldsymbol{x}, u)\end{aligned}$$

and

$$\langle \phi((\boldsymbol{x}, u)), \boldsymbol{\Phi}_t^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u) \rangle_{\ell^2} = \phi((\boldsymbol{x}, u))^\top \boldsymbol{\Phi}_t^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u) = \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u).$$

将上述等式合并可得

$$\begin{aligned}\|\delta_t(\boldsymbol{x}, u)\|_{\ell^2}^2 &= \|\phi((\boldsymbol{x}, u))\|_{\ell^2}^2 - \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u) - \lambda \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-2} \boldsymbol{k}_t(\boldsymbol{x}, u) \\ &= K((\boldsymbol{x}, u), (\boldsymbol{x}, u)) - \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k}_t(\boldsymbol{x}, u) - \lambda \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-2} \boldsymbol{k}_t(\boldsymbol{x}, u) \\ &= \sigma_{u,t}^2(\boldsymbol{x}) - \lambda \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-2} \boldsymbol{k}_t(\boldsymbol{x}, u) \\ &\leq \sigma_{u,t}^2(\boldsymbol{x})\end{aligned}$$

因为 $\boldsymbol{\Sigma}_t^{-1}$ 是半负定的。

□

D.5 引理 C.6 的证明

证明。 我们首先定义

$$\boldsymbol{s}_t = \boldsymbol{\Phi}_t \boldsymbol{\epsilon}_t = \sum_{s=1}^{t} \phi(\boldsymbol{x}_s, u_s) \epsilon_s.$$

注意， $\boldsymbol{s}_t$ 是关于 $\mathcal{F}_t$的鞅。

我们还定义一个超鞅

$$M_t(\boldsymbol{g}) = exp\big(\sum_{s=1}^{t} \frac{1}{\sigma} \langle \boldsymbol{g}, \boldsymbol{s}_t \rangle - \frac{1}{2} \|\boldsymbol{g}\|^2 \big)$$

它有一种等价形式

$$M_t(\boldsymbol{g}) = exp\big(\sum_{s=1}^{t} \frac{1}{\sigma} \langle \boldsymbol{g}, \phi(\boldsymbol{x}_s, u_s) \rangle \epsilon_s - \frac{1}{2} \|\boldsymbol{g}\|^2 \big)$$

其中 $\boldsymbol{g}$ 是函数向量，其元素为

我们沿用经典线性老虎机 [28],的方法，即对 $\boldsymbol{g}$上的高斯分布求 $M_t(\boldsymbol{g})$的平均值。关键的技术难点在于函数向量 $\boldsymbol{g}$的维数是无限的。我们首先进行截断处理，这样便能与经典结果精确对应。令 d为特征映射的维数。我们的目标是在 $d \to \infty$时获得极限结果。现在假设$\boldsymbol{g}^d \sim \mathcal{N}(\mathbf{0}, \frac{1}{\lambda} \boldsymbol{I}_d)$，且与其他所有变量独立，并定义

$$M_t^{(d)} = \mathbb{E}_{\boldsymbol{g}^d}[M_t(\boldsymbol{g}^d)] = \int M_t^{(d)}(\boldsymbol{g}) d\rho^d(\boldsymbol{g})$$

由迭代期望（即富比尼定理），我们有

$$\mathbb{E}[M_t^{(d)} | \mathcal{F}_t] \leq M_{t-1}$$

这表明 M_t 是一个超鞅。

接着，我们将 $\Psi \quad \ell^2 \to \mathbb{R}^d \qquad d \qquad \Psi_d \boldsymbol{\theta} =$ 定义为：投影到前几个 $[\boldsymbol{\Theta}_1, \cdots, \boldsymbol{\Theta}_d]^\top \qquad \boldsymbol{\theta} \in \ell^2$ 坐标上的截断投影：对任意 。我们进一步记

$$\boldsymbol{\Psi}_d \boldsymbol{\Phi}_t^\top = [\Psi_d \phi(\boldsymbol{x}_1, u_1), \cdots, \Psi_d \phi(\boldsymbol{x}_t, u_t)] \in \mathbb{R}^{d \times t}$$

and

$$\boldsymbol{\Psi}_d \boldsymbol{J}_t \boldsymbol{\Psi}_d^\top = \boldsymbol{\Psi}_d \boldsymbol{\Phi}_t^\top \boldsymbol{\Phi}_t \boldsymbol{\Psi}_d.$$

我们注意到

$$\frac{\det(\lambda \boldsymbol{I}_d)}{\det\Big(\lambda \boldsymbol{I}_d + \boldsymbol{\Psi}_d \boldsymbol{J}_t \boldsymbol{\Psi}_d^\top\Big)} = \frac{1}{\det\Big(\boldsymbol{I}_d + \lambda^{-1} \boldsymbol{\Psi}_d \boldsymbol{J}_t \boldsymbol{\Psi}_d^\top\Big)}$$

由此得到

$$\begin{aligned} M_t^{(d)} &= \Big(\frac{\det(\lambda \boldsymbol{I}_d)}{\det\Big(\lambda \boldsymbol{I}_d + \boldsymbol{\Psi}_d \boldsymbol{J}_t \boldsymbol{\Psi}_d^\top\Big)}\Big)^{1/2} \exp\left(\frac{1}{2\sigma^2} \|\boldsymbol{\Psi}_d \boldsymbol{\Phi}_t \boldsymbol{\epsilon}_t\|^2_{(\lambda \boldsymbol{I}_d + \boldsymbol{\Psi}_d \boldsymbol{J}_t \boldsymbol{\Psi}_d^\top)^{-1}}\right) \\ &= \det\Big(\boldsymbol{I}_d + \lambda^{-1} \boldsymbol{\Psi}_d \boldsymbol{J}_t \boldsymbol{\Psi}_d^\top\Big)^{-1/2} \exp\left(\frac{1}{2\sigma^2} \|\boldsymbol{\Psi}_d \boldsymbol{\Phi}_t \boldsymbol{\epsilon}_t\|^2_{(\lambda \boldsymbol{I}_d + \boldsymbol{\Psi}_d \boldsymbol{J}_t \boldsymbol{\Psi}_d^\top)^{-1}}\right). \end{aligned}$$

令 M_t 为 $M_t^{(d)}$ 在 $d \to \infty$时的极限，我们有

$$\begin{aligned} M_t &= \det\big(\boldsymbol{I}_\infty + \lambda^{-1} \boldsymbol{J}_t\big)^{-1/2} \exp\left(\frac{1}{2\sigma^2} \|\boldsymbol{\Phi}_t \boldsymbol{\epsilon}_t\|^2_{(\lambda \boldsymbol{I}_\infty + \boldsymbol{J}_t)^{-1}}\right) \\ &= \det\big(\boldsymbol{I}_t + \lambda^{-1} \boldsymbol{K}_t\big)^{-1/2} \exp\left(\frac{1}{2\sigma^2} \|\boldsymbol{\Phi}_t \boldsymbol{\epsilon}_t\|^2_{\Gamma_t^{-1}}\right) \end{aligned}$$

其中第二步由（Slyvestr）或 Weinstein–Aronszajn恒等式得出。
根据 Ville不等式,

$$\mathbb{P}\big(\sup_{t=0,1,2,\cdots} M_t \geq \frac{1}{\delta}\big) \leq \mathbb{E}[M_0] \cdot \delta$$

且 $M_0 = 1$。因此我们知道，以至少 $1-\delta$ 的概率，对于所有 $t = 0$，1，2，$\cdots$

$$\log(M_t) \leq \log\left(\frac{1}{\delta}\right)$$

由此得到

$$-\frac{1}{2} \log\det\big(\boldsymbol{I}_t + \lambda^{-1} \boldsymbol{K}_t\big) + \frac{1}{2\sigma^2} \|\boldsymbol{\Phi}_t \boldsymbol{\epsilon}_t\|^2_{\Gamma_t^{-1}} \leq \log\left(\frac{1}{\delta}\right).$$

重新排列后，我们得到

$$\|\boldsymbol{\Phi}_t \boldsymbol{\epsilon}_t\|^2_{\Gamma_t^{-1}} \leq 2\sigma^2 \log \frac{\sqrt{\det(\boldsymbol{I}_t + \lambda^{-1} \boldsymbol{K}_t)}}{\delta}.$$

这表明我们的结果成立。

□

D.6 引理 C.4 的证明

证明。 将 $\boldsymbol{\Phi}_t = \boldsymbol{U}_t \boldsymbol{\Lambda}_t \boldsymbol{V}_t^\top$写作 $\boldsymbol{\Phi}_t$的奇异值分解（SVD）。我们有$\boldsymbol{\Lambda}_t = [\boldsymbol{\Lambda}_{1,t}, \boldsymbol{0}]$ 其中 $\boldsymbol{\Lambda}_{1,t}$是一个对角矩阵，其对角元素为 $\boldsymbol{\Phi}_t$的奇异值。我们还注意到 $\boldsymbol{\Sigma}_t \in \mathbb{R}^{t \times t}$和 $\boldsymbol{U}_t \in \mathbb{R}^{t \times t}$。我们还有

$$\boldsymbol{J}_t = \boldsymbol{\Phi}_t^\top \boldsymbol{\Phi}_t = \boldsymbol{V}_t \begin{bmatrix} \boldsymbol{\Lambda}_{1,t}^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{V}_t^\top$$

类似地

$$\boldsymbol{K}_t = \boldsymbol{\Phi}_t \boldsymbol{\Phi}_t^\top = \boldsymbol{U}_t \boldsymbol{\Lambda}_{1,t}^2 \boldsymbol{U}_t^\top.$$

于是，我们有

$$\boldsymbol{\Gamma}_t = \boldsymbol{V}_t \begin{bmatrix} \boldsymbol{\Lambda}_{1,t}^2 + \lambda \boldsymbol{I}_t & \boldsymbol{0} \\ \boldsymbol{0} & \lambda \boldsymbol{I}_\infty \end{bmatrix} \boldsymbol{V}_t^\top, \quad \boldsymbol{\Sigma}_t = \boldsymbol{U}_t (\boldsymbol{\Lambda}_{1,t}^2 + \lambda \boldsymbol{I}_t) \boldsymbol{U}_t^\top.$$

显然，有如下恒等式:

$$\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Phi}_t = \boldsymbol{\Phi}_t\boldsymbol{\Gamma}_t^{-1}$$

由于两边都等于 $\boldsymbol{U}_t[\boldsymbol{D}_t, \boldsymbol{0}]\boldsymbol{V}_t^\top$where $\boldsymbol{D}_t = \boldsymbol{\Lambda}_{1,\ t}(\boldsymbol{\Lambda}_{1,\ t}^2 + \lambda\boldsymbol{I}_t)^{-1}$，这是一个对角矩阵。接下来，我们注意到

$$\begin{aligned}
\sigma_{u,t}^2(\boldsymbol{x}) &= K((\boldsymbol{x},u),(\boldsymbol{x},u)) - \boldsymbol{k}_t(\boldsymbol{x},u)^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{k}_t(\boldsymbol{x},u) \\
&= \phi(\boldsymbol{x},u)^\top(\boldsymbol{I}_\infty - \boldsymbol{\Phi}_t^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Phi}_t)\phi(\boldsymbol{x},u) \\
&= \phi(\boldsymbol{x},u)^\top(\boldsymbol{I}_\infty - \boldsymbol{\Phi}_t^\top\boldsymbol{\Phi}_t\boldsymbol{\Gamma}_t^{-1})\phi(\boldsymbol{x},u)
\end{aligned}$$

这是由矩阵诱导的 $\phi(\boldsymbol{x},\ u)$ 的范数

$$\begin{aligned}
\boldsymbol{I}_\infty - \boldsymbol{\Phi}_t^\top\boldsymbol{\Phi}_t\boldsymbol{\Gamma}_t^{-1} &= \boldsymbol{I}_\infty - \boldsymbol{J}_t\boldsymbol{\Gamma}_t^{-1} \\
&= \boldsymbol{V}_t\begin{bmatrix}\lambda\boldsymbol{I}_t(\boldsymbol{\Lambda}_{1,t}^2 + \lambda\boldsymbol{I}_t)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \lambda\boldsymbol{I}_\infty\end{bmatrix}\boldsymbol{V}_t^\top \\
&= \lambda\boldsymbol{V}_t\begin{bmatrix}(\boldsymbol{\Lambda}_{1,t}^2 + \lambda\boldsymbol{I}_t)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_\infty\end{bmatrix}\boldsymbol{V}_t^\top \\
&= \lambda\boldsymbol{\Gamma}_t^{-1}.
\end{aligned}$$

因此，我们得到另一个恒等式

$$\sigma_{u,t}^2(\boldsymbol{x}) = \lambda\|\phi(\boldsymbol{x},u)\|_{\boldsymbol{\Gamma}_t^{-1}}^2.$$

□

E 其他内容

E.1 算法

算法 1 LK-GP-UCB

1：输入： T, λ, $\{\beta_t\}_{t=1}^T$**2：初始化：** μ_u**,0(**$\boldsymbol{x}$**),**σ_u**,0(**$\boldsymbol{x}$**) 3：对于** $t = 1$**,...,**T**，执行 4：观察用户** u_t **和臂集合** $\mathcal{D}_t$**。 5：选择臂** $\boldsymbol{x}_t =$ **argmax**$_{\boldsymbol{x}\in\mathcal{D}_t}$ μ_{u_t}**,**$t-1$**(**$\boldsymbol{x}$**) +** $\beta_t\sigma_{u_t}$**,**$t-1$**(**$\boldsymbol{x}$**)。 6：接收反馈** $y_t = f(\boldsymbol{x}_t, u_t) + \varepsilon_t$**。 7：更新** μ_{u_t}**,**t**(**$\boldsymbol{x}$**) 和** $\sigma_{u_t}^2$**,**t**(**$\boldsymbol{x}$**)。 8：结束循环**

算法 2 LK-GP-TS

1：输入： T, λ, $\{\nu_t\}_{t=1}^T$**2：初始化：** μ_u**,0(**$\boldsymbol{x}$**),**σ_u**,0(**$\boldsymbol{x}$**) 3：对于** $t = 1$**,...,**T**，执行 4：观察用户** u_t **和臂集合** $\mathcal{D}_t$**。 5：从** $\mathcal{N}$**(**μ_{u_t}**,** $_{t-1}$**(**$\boldsymbol{x}$**)** $\nu_t^2\sigma_{u_t}^2$**,**$t-1$**(**$\boldsymbol{x}$**)) 中为所有** $\boldsymbol{x} \in \mathcal{D}_t$**采样** $\widetilde{\mu}_t$**(**$\boldsymbol{x}$**) 6：选择臂** $\boldsymbol{x}_t =$ **argmax**$_{\boldsymbol{x}\in\mathcal{D}_t}$ $\widetilde{\mu}_t$**(**$\boldsymbol{x}$**)。 7：接收反馈** $y_t = f(\boldsymbol{x}_t, u_t) + \varepsilon_t$**。 8：更新** μ_{u_t}**,**t**(**$\boldsymbol{x}$**) 和** $\sigma_{u_t}^2$**,**t**(**$\boldsymbol{x}$**)。 9：结束循环**

E.2 后验均值和方差的递归更新

本节介绍如何利用舒尔补的性质，推导后验均值和后验方差的增量更新方法 [8], 。回想一下，我们需要处理其求逆,

$\boldsymbol{\Sigma}_t = \boldsymbol{I} + \lambda \boldsymbol{K}_t \in \mathbb{R}^{t\times t}$ 它会随着轮次增加而增大。为了高效地计算 $\boldsymbol{\Sigma}_t$ 的逆，我们利用 $\boldsymbol{\Sigma}_{t-1}$ 中的递归公式，该公式基于分块矩阵求逆公式。

$$\boldsymbol{\Sigma}_t^{-1} = \begin{bmatrix} \boldsymbol{M}_{11,t} & \boldsymbol{M}_{12,t} \\ \boldsymbol{M}_{12,t}^\top & d_t^{-1} \end{bmatrix} \tag{26}$$

其中

$$\begin{aligned} \boldsymbol{M}_{11,t} &= \boldsymbol{\Sigma}_{t-1}^{-1} + d_t^{-1}\boldsymbol{G}_t \\ \boldsymbol{M}_{12,t} &= -d_t^{-1}\boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{k}_{t-1}(\boldsymbol{x}_t, u_t) \end{aligned} \tag{27}$$

and

$$\begin{aligned} d_t &= K((\boldsymbol{x}_t, u_t), (\boldsymbol{x}_t, u_t)) - \boldsymbol{k}_{t-1}(\boldsymbol{x}_t, u_t)^\top \boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{k}_{t-1}(\boldsymbol{x}_t, u_t) + \lambda = \sigma_{u_t,t-1}^2(\boldsymbol{x}_t) + \lambda \\ \boldsymbol{G}_t &= \boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{k}_{t-1}(\boldsymbol{x}_t, u_t)\boldsymbol{k}_{t-1}(\boldsymbol{x}_t, u_t)^\top \boldsymbol{\Sigma}_{t-1}^{-1} \end{aligned}$$

这里， d_t 是舒尔补。

因此，利用式 (26) 可得后验均值

$$\begin{aligned} \mu_{u,t}(\boldsymbol{x}) =& \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1}\boldsymbol{y}_t \\ =& \begin{bmatrix} \boldsymbol{k}_{t-1}(\boldsymbol{x}, u)^\top & K((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t)) \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_{11,t} & \boldsymbol{M}_{12,t} \\ \boldsymbol{M}_{12,t}^\top & d_t^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_{t-1} \\ y_t \end{bmatrix} \\ =& \boldsymbol{k}_{t-1}(\boldsymbol{x}, u)^\top \boldsymbol{M}_{11,t}\boldsymbol{y}_{t-1} + K((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t))\boldsymbol{M}_{12,t}^\top \boldsymbol{y}_{t-1} \\ &+ \boldsymbol{k}_{t-1}(\boldsymbol{x}, u)^\top \boldsymbol{M}_{12,t} y_t + K((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t)) d_t^{-1} y_t \\ =& \underbrace{\boldsymbol{k}_{t-1}(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{y}_{t-1}}_{\mu_{u,t-1}(\boldsymbol{x})} + d_t^{-1}(\beta_1 \boldsymbol{y}_{t-1} - \beta_2 \boldsymbol{y}_{t-1} - \beta_3 y_t + \beta_4 y_t) \end{aligned}$$

其中

$$\begin{aligned} &\beta_1 = \boldsymbol{k}_{t-1}(\boldsymbol{x}, u)^\top \boldsymbol{G}_t \Rightarrow \beta_1 \boldsymbol{y}_{t-1} = \Big(\boldsymbol{k}_{t-1}(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{k}_{t-1}(\boldsymbol{x}_t, u_t)\Big)\mu_{u_t,t-1}(\boldsymbol{x}_t) \\ &\beta_2 = K((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t))\boldsymbol{k}_{t-1}(\boldsymbol{x}_t, u_t)^\top \boldsymbol{\Sigma}_{t-1}^{-1} \Rightarrow \beta_2 \boldsymbol{y}_{t-1} = K((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t))\mu_{u_t,t-1}(\boldsymbol{x}_t) \\ &\beta_3 = \boldsymbol{k}_{t-1}(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{k}_{t-1} \\ &\beta_4 = K((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t)). \end{aligned}$$

因此，我们得到后验均值的递归更新

$$\begin{aligned} \mu_{u,t}(\boldsymbol{x}) &= \mu_{u,t-1}(\boldsymbol{x}) + d_t^{-1} \\ &\quad \times \Big(\boldsymbol{k}_{t-1}(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{k}_{t-1}(\boldsymbol{x}_t, u_t)(\mu_{u_t,t-1}(\boldsymbol{x}_t) - y_t) + K((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t))(y_t - \mu_{u_t,t-1}(\boldsymbol{x}_t))\Big) \\ &\quad = \mu_{u,t-1}(\boldsymbol{x}) + d_t^{-1} q_{t-1}((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t))(y_t - \mu_{u_t,t-1}(\boldsymbol{x}_t)) \end{aligned}$$

其中， $q_{t-1}((\boldsymbol{x}, u),(\boldsymbol{x}_t, u_t))$ 定义为

$$q_t((\boldsymbol{x}, u), (\boldsymbol{x}', u')) = K((\boldsymbol{x}, u), (\boldsymbol{x}', u')) - \boldsymbol{k}_t(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_t^{-1}\boldsymbol{k}_t(\boldsymbol{x}', u')$$

利用（26）可以将其转化为递归形式

$$\begin{aligned} &q_t((\boldsymbol{x}, u), (\boldsymbol{x}', u')) \\ =& K((\boldsymbol{x}, u), (\boldsymbol{x}', u')) - \Big(\boldsymbol{k}_{t-1}(\boldsymbol{x}, u)^\top \boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{k}_{t-1}(\boldsymbol{x}', u') \\ &\quad + d_t^{-1}(\beta_1 \boldsymbol{k}_{t-1}(\boldsymbol{x}', u') - \beta_2 \boldsymbol{k}_{t-1}(\boldsymbol{x}', u') - \beta_3 K((\boldsymbol{x}_t, u_T), (\boldsymbol{x}', u')) + \beta_4 K((\boldsymbol{x}_t, u_T), (\boldsymbol{x}', u')))\Big) \\ =& q_{t-1}((\boldsymbol{x}, u), (\boldsymbol{x}', u')) - d_t^{-1} q_{t-1}((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t)) q_{t-1}((\boldsymbol{x}_t, u_t), (\boldsymbol{x}', u')). \end{aligned}$$

现在，利用后验协方差的增量更新，我们可以轻松得到后验方差的递归更新

$$\sigma_{u,t}^2(\boldsymbol{x}) = \sigma_{u,t-1}^2(\boldsymbol{x}) - d_t^{-1} q_{t-1}^2((\boldsymbol{x}, u), (\boldsymbol{x}_t, u_t)).$$

现在将 d_t 替换为 $\sigma_{u_t,\ t-1}^2(\boldsymbol{x}_t) + \lambda$ ，即可得到式（8）中的递归更新。

F 实验补充材料

本附录详细介绍了我们的合成环境、算法配置、超参数选择、实现方案、消融实验和报告规范。

F.1 合成环境

令 $\mathcal{U} = \{1, \ldots, n\}$ 表示用户， $\mathcal{D} \subset \mathbb{R}^d$ 表示臂（上下文）空间， $M_t := |\mathcal{D}_t|$ 表示第 t 轮展示的候选项数量。我们抽取一个全局归一化上下文池 $\mathcal{D} = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$，其中包含 $\boldsymbol{x}^{(i)} \sim \mathcal{N}$（$\mathbf{0}, \boldsymbol{I}_d$）和 $\boldsymbol{x}^{(i)} \leftarrow \boldsymbol{x}^{(i)}/\|\boldsymbol{x}^{(i)}\|$。在第 t 轮，我们从 $\mathcal{D}$ 中无放回地抽取 M_t 个不同物品，展示给 $\mathcal{D}_t$。每轮服务一位用户 u_t；除非另有说明，否则从 $\mathcal{U}$ 中均匀抽取。观测到的奖励包含加性噪声 $y_t = f$（$\boldsymbol{x}_t, u_t$）$+ \varepsilon_t$。我们在一种线性机制（*Linear–GOB*）和两种核化机制（拉普拉斯核，分别采用高斯过程抽样和表示子抽样）下生成图、上下文和真实奖励。

用户图。我们考虑在 $\mathcal{U}$ 上使用两种随机图生成器。第一种随机图族是 Erd˝os– Rényi（ER）随机图：每条（无向）边以 p 的概率出现，边权为 $w_{ij} = 1$。我们在实验中设定 $p = 0.2$。第二种是径向基函数（RBF）随机图：抽取潜在的 $\boldsymbol{z}_i \sim \mathcal{N}$（$\mathbf{0}, \boldsymbol{I}_q$），设定 $w_{ij} = \exp\big(-\rho_L \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2^2\big)$，并通过保留满足 $w_{ij} \geq s$ 的边进行稀疏化。我们在模拟中选取 $s = 0.1$、$\rho_L = 0.1$ 和 $q = 4$。

任务设计。我们设计了不同难度级别的任务。最简单的情况是 (M, M_t, n, d, T) =(10, 5, 20, 5, 1000)。这是一个 10 臂多臂老虎机问题，对所有用户而言，每轮的可见率为 50%。中等难度的情况是 (M, M_t, n, d, T) =(20, 5, 20, 10, 3000)，对应一个 20 臂多臂老虎机问题，对所有用户而言，每轮的可见率为 25%。难度最高的情况则使用 (M, M_t, n, d, T) =(50, 5, 20, 20, 3000)，对应一个 50 臂多臂老虎机问题，对所有用户而言，每轮的可见率为 10%。除非另有说明，否则 σ 设为 0.1。

实际场景。尽管我们的实证研究使用合成环境，但我们所考虑的多用户图老虎机设定受到多种实际应用的启发。例如，在推荐系统中，用户通过社交图或相似性图相连，并反复与共用的商品目录互动；在区域个性化问题中，商店或地理区域构成一张图，而臂对应商品组合或定价行动；在医疗保健或教育领域，患者或学生通过相似性网络相连，治疗方案或练习则构成臂集合。在这些领域中，所提出的拉普拉斯核化老虎机能够利用用户图共享统计强度，同时捕捉非线性上下文效应。

F.1.1 机制 1：*Linear–GOB*（图平滑线性奖励）

对初始用户参数 $\boldsymbol{\Theta}_0 \in \mathbb{R}^{n \times d}$ 进行采样，其行数为 $\boldsymbol{\theta}_{0,i} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$。通过蒂霍诺夫平滑[30]来施加图同质性约束：

$$\boldsymbol{\Theta} \;=\; \arg\min_{\tilde{\boldsymbol{\Theta}}} \; \|\tilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\|_F^2 + \eta \,\mathrm{tr}(\tilde{\boldsymbol{\Theta}}^\top \boldsymbol{L} \tilde{\boldsymbol{\Theta}}) \;=\; (\boldsymbol{I}_n + \eta \boldsymbol{L})^{-1} \boldsymbol{\Theta}_0.$$

因此， $f(\boldsymbol{x}, u) = \boldsymbol{x}^\top \boldsymbol{\theta}_u$，其中 $\boldsymbol{\theta}_u$ 是 $\boldsymbol{\Theta}$ 的第 u 行。图同质性的强度 η 默认设为 1.0。

F.1.2 机制2：拉普拉斯核

我们选择在臂上使用基核 K_x，即平方指数核，其定义如下：

$$K_{\mathrm{SE}}(\boldsymbol{x}, \boldsymbol{x}') = \exp\Big(-\|\boldsymbol{x} - \boldsymbol{x}'\|^2 / 2\ell^2\Big)$$

其中长度尺度 $\ell > 0$在我们的实验中设为 1.0。然后，我们根据定义构造多用户核：

$$K((\boldsymbol{x}, u), (\boldsymbol{x}', u')) \;=\; [\boldsymbol{L}_\rho^{-1/2}]_{u,u'} \, K_x(\boldsymbol{x}, \boldsymbol{x}')$$

其中，我们在实验中设定 $\rho = 0.01$。

选项 A：使用高斯过程抽样的拉普拉斯核

我们从协方差由 K 导出的零均值高斯过程中抽取联合值 $\{f(\boldsymbol{x},\ u)\}_{u\in\mathcal{U},\ \boldsymbol{x}\in\mathcal{D}}$，并通过在 $\mathcal{D}\times\mathcal{U}$ 上进行插值来固定 f。噪声为 $\varepsilon_t\sim\mathcal{N}(0,\ \sigma^2)$，其范围为 $\sigma=0.01\cdot\mathrm{range}(f)$。

选项 B：拉普拉斯核 **结合表示子抽样**我们考虑再生核希尔伯特空间（RKHS）的表示定理，并通过 $\alpha_{\boldsymbol{x},\ u}\sim\mathcal{N}(0,\ 1)$ 对 $\mathcal{D}\times\mathcal{U}$ 中的独立同分布系数进行采样，并设置

$$f(\boldsymbol{x},u)\ =\ \sum_{u'\in\mathcal{U},\boldsymbol{x}'\in\mathcal{D}}\alpha_{\boldsymbol{x},u}\,K\big((\boldsymbol{x},u),(\boldsymbol{x}',u')\big).$$

F.2 基线方法

在每个合成环境的每次试验中，所有方法都面对相同的序列 $\{u_t,\mathcal{D}_t,\varepsilon_t\}_{t=1}^{T}$，以确保公平比较。我们的实验包括以下基线方法。

逐用户 LinUCB（无图）。：我们实现了逐用户 LinUCB，该方法忽略整个图，并针对每个用户独立执行线性赌博机算法。

Pooled LinUCB（无图）。：我们实现了 Pooled LinUCB，该方法忽略图和个性化，将多用户问题视为单智能体赌博机问题。简而言之，该问题由全局线性 UCB 算法求解。

GP-UCB（无图）。我们实现了 GP-UCB[8],，即先前关于高斯过程和 UCB 的研究 [8] 中的 IGP-UCB。这是一种核化基线方法，仅在臂上使用 K_x，忽略用户之间的相似性（拉普拉斯算子）。

GoB.Lin. 我们实现了 GoB.Lin，这是群组多臂老虎机问题中的经典方法 [5]。它是一种基于图白化特征的拉普拉斯正则化线性 UCB 算法（等价于 GraphUCB，且 $\rho=1$ 即 $\boldsymbol{A}=\boldsymbol{I}+\boldsymbol{L}$）。算法中的置信度尺度根据表格进行调整。

GraphUCB. 我们实现了 GraphUCB[7],，即拉普拉斯正则化 LinUCB。此外，算法中的置信度尺度根据表格进行调整。

COOP-KernelUCB. 我们实现了 COOP-KernelUCB[13],，它利用智能体 $\times$ 臂上的乘积核。这里我们沿用其工作中的记号。我们考虑 K_z 的五种选择（如下所示）；完整的核为 $K=K_z\otimes K_x$，并采用与 LK-GP-UCB 相同的 UCB 规则。

智能体核 K_z 的五种半正定选项：

1. **拉普拉斯算子_inv**：$\boldsymbol{K}_z=(\boldsymbol{L}+\rho\boldsymbol{I})^{-1},\qquad\rho>0$。
2. **热核**：$\boldsymbol{K}_z=\exp(-\tau\boldsymbol{L})$，通过对 L进行谱分解得到。
3. **谱_径向基函数**：使用 k 最低的非平凡拉普拉斯特征向量 $Z\in\mathbb{R}^{n\times}$嵌入节点 ^{k} 并设置
$$K_z[u,u']\ =\ \exp\Big(-\tfrac{\|Z_u-Z_{u'}\|^2}{2\sigma_z^2}\Big).$$
4. **全一**：完全协作，$\boldsymbol{K}_z=\mathbf{1}\mathbf{1}^{\top}$。
5. **学习得到的_最大均值差异 (网络上下文，忠实于 [13])**：定义每个用户对 $\mathcal{H}_x$中观测到的臂-上下文分布的核均值嵌入Ψ_u，并令
$$K_{z,t}(u,u')\ =\ \exp\Big(-\tfrac{\|\hat{\Psi}_u(t)-\hat{\Psi}_{u'}(t)\|^2}{2\sigma_z^2}\Big),$$
其中，$\hat{}\,\Psi_u(t)$ 是截至时间 t 用户 u 所观测上下文的经验均值嵌入。在我们的实现中，我们使用高效的随机傅里叶特征近似来近似 K_x，并按固定计划更新 $K_{z,\ t}$；观测数量少于某个较低阈值的用户只与自身协作（即仅保留对角项）。

对于时变的 K_z（学习得到的 _mmd），高斯过程状态会在 K_z 个刷新点利用算法自身的历史记录重新构建，以确保格拉姆矩阵与当前核保持一致。图展示了 COOP-KernelUCB 中 K_z 的选择对比。

F.3 集中式协议

每个 t：采样 $u_t\sim$ 均匀分布$(\mathcal{U})$，呈现 $\mathcal{D}_t$ (大小为 M_t)，按照算法选择 $\boldsymbol{x}_t\in\mathcal{D}_t$，观察$y_t$，更新我们的决策策略（模型），并记录 $\Delta_t=\max_{\boldsymbol{x}\in\mathcal{D}_t}f(\boldsymbol{x},\ u_t)-f(\boldsymbol{x}_t,\ u_t)$。每种配置重复进行 R 次试验（最终结果使用 $R=20$；初步/试点调优使用$R\in[5,10]$）。

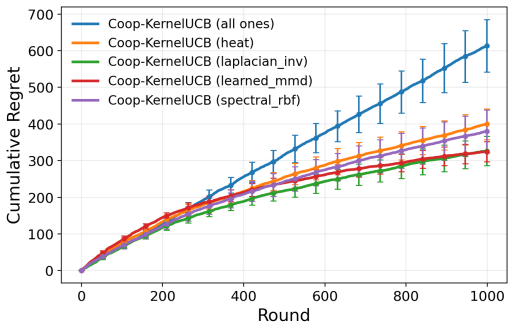


图 5：COOP-KernelUCB 用户相似性核选择的比较。

F.4 后验更新与数值细节

动机。 对于第 t 轮的原始更新 (3)，求逆需要 $\mathcal{O}(t^3|\mathcal{D}_t|)$ 时间。实用更新对每一对 $(\boldsymbol{x},\ u)$ 都很高效，但它需要更新所有配对，因此需要 $\mathcal{O}(|\mathcal{D}_t|n)$ 时间。因此，高层次的思路是在 $t \le n^{1/3}$ 时执行原始更新 (3)，并在 $t \le n^{1/3}$ 时执行实用更新 (8)。因此，对于基于高斯过程的方法，我们采用混合实现，具体如下所述。

精确（乔列斯基）阶段： 维护 $\boldsymbol{\Sigma}_t = K_t + \lambda \boldsymbol{I}$，并通过秩一乔列斯基分解更新 $t < t_*$（每步开销为 $O(t^2)$；初始求逆开销为 $O(t^3)$）。

递归阶段： 切换到式 (8) 中的秩一递推，并将 $q_0 = K$ 限制为 $\mathcal{D} \times \mathcal{U}$。应用于整个网格 $\mathcal{D} \times \mathcal{U}$时，每次更新的开销为 $O(n|\mathcal{D}_t|)$。

默认情况下，我们取 $t_\star = \min\{1500,\ \lfloor n^{1/3} \rfloor |\mathcal{D}|\}$ 作为阶段切换点。我们使用乔列斯基抖动 10^{-8}，将负方差截为零，并缓存 $K_x(\mathcal{D},\ \mathcal{D})$。对于较大的 n，我们还可以选择应用图谱截断 $\boldsymbol{L}_\rho \approx \boldsymbol{U}_r \Lambda_r \boldsymbol{U}_r^\top$（前 r 个特征值对），从而得到 $K \approx (\boldsymbol{U}_r \Lambda_r^{-1} \boldsymbol{U}_r^\top) \otimes K_x$。

F.5 超参数与调优

各算法中固定的内容。 为确保公平，在每个环境中固定基础核超参数：长度尺度 ℓ 使用针对 $\mathcal{D}$ 的中位数启发式，拉普拉斯岭参数 $\rho = 0.1$ 固定。对于 K_z，拉普拉斯_逆使用 $\rho = 0.1$；热核使用 $\tau = 1.0$；谱_径向基函数使用 $k = 8$和中位数带宽；学习型_*MMD*使用 256 维随机特征、中位数带宽启发式、约每 200 轮更新一次，并要求用户至少有 5 条观测后才能参与协作。

设计思路。 为避免将未知噪声尺度用作先验，所有高斯过程类方法都使用一种感知图结构和时间的拉普拉斯岭参数调度

$$\lambda_t \ = \ \lambda_{\text{base}} \cdot S_{\text{spec}} \cdot \frac{T}{T+t}, \qquad S_{\text{spec}} \ = \ \frac{\lambda_2(\boldsymbol{L})}{\lambda_{\max}(\boldsymbol{L})} \in [0,1],$$

其中 $\lambda_{Fiedler}(\boldsymbol{L})$ 是费德勒值（最小非零特征值）。我们将 λ_t 裁剪至 $[\lambda_{\min}, \lambda_{\max}]$，其中 $\lambda_{\min} = 10^{-6}$ 和 $\lambda_{\max} = 10^{-1}$。为限制重新分解的次数，我们按照倍增周期调度更新 λ（大约在 $t \approx$200、400、800、.....），且仅当变化超过 20% 时才重新构建。

调优内容。 仅通过网格搜索在试点时域（中等/困难任务使用 $T_{\text{pilot}} = 1500$；简单任务使用 $T_{\text{pilot}} = 1000$）上调优探索尺度，使用 $R_{\text{pilot}} \in \{5, 10\}$：

算法	网格（试点）
LK-GP-UCB、GP-UCB、Coop-KernelUCB	$\beta \in \{0.5, 1, 2, 4\}$
LK-GP-TS	$\nu \in \{0.5, 1, 2, 4\}$
GoB.Lin、GraphUCB、LinUCB变体	$\alpha \in \{0.5, 1, 2, 4\}$

随后，在完整时域评估中冻结表现最佳的试点设置（以试点累计遗憾的均值为准）。高斯过程更新中的噪声/岭正则化 λ_{base} 使用试点上的 $\lambda_{\text{base}} \in \{0.001, 0.005, 0.01, 0.05, 0.1\}$。

F.6 消融与压力测试

我们报告了两项消融研究。一项是在中等规模、拉普拉斯核与高斯过程抽样环境下进行的消融研究（ER图，固定 ℓ 和 ρ），考察**用户规模可扩展性（n）：** $n \in \{20, 50, 100, 200\}$固定（$M$， M_t， d， T）和图生成器。我们给出了最终累积遗憾与 n 的关系，并在表1中报告了最后一步的累积遗憾。另一项研究考察随机图模型的影响。我们的标准实验采用两种随机图生成器：Erd˝os–Rényi（ER）随机图和 F.1 中提到的径向基函数（RBF）随机图。我们在这项消融研究中加入了随机块模型（SBM）。我们仍采用中等规模、拉普拉斯核与高斯过程抽样环境。结果见图6。

表1：用户数量的消融实验 n（最终累积遗憾；均值±标准误）。

算法	$n = 20$	$n = 50$	$n = 100$	$n = 200$
LK-GP-UCB	627.22 ± 32.98	892.43 ± 21.73	1062.69 ± 18.29	1157.74 ± 23.02
LK-GP-TS	634.46 ± 22.78	943.41 ± 19.56	1176.23 ± 15.77	1260.35 ± 16.23
Coop-KernelUCB	730.06 ± 31.02	1015.35 ± 22.18	1273.28 ± 17.36	1358.48 ± 14.22
GOB.Lin	1092.86 ± 71.70	1203.32 ± 18.57	1370.51 ± 16.78	1432.48 ± 18.72
GraphUCB	1105.20 ± 68.54	1192.30 ± 22.12	1360.02 ± 15.32	1453.21 ± 17.81
GP-UCB	2222.20 ± 90.26	1964.65 ± 61.40	1641.43 ± 37.43	1444.83 ± 36.33
Pooled-LinUCB	2360.95 ± 70.55	1909.81 ± 49.49	1723.27 ± 40.23	1438.74 ± 26.44
PerUser-LinUCB	1117.87 ± 72.04	1221.99 ± 22.03	1432.89 ± 18.81	1527.04 ± 17.61

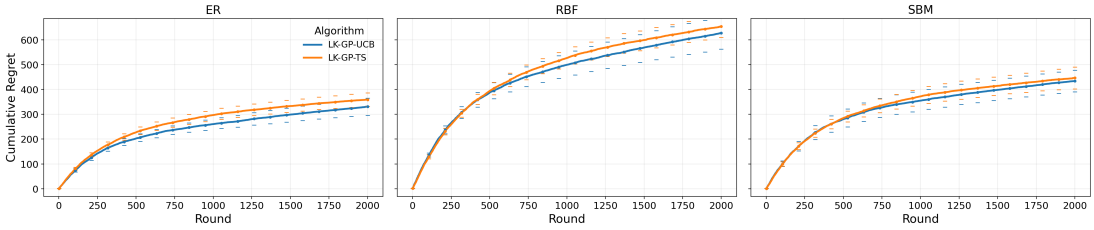


图6：随机图模型选择的比较。