arXiv:2302.00392v1 [stat.ML] 1 Feb 2023

# 核老虎机中的延迟反馈

**Sattar Vakili**
联发科技研究院
sattar.vakili@mtkresearch.com

**Danyal Ahmed**联发科技研究院
danyal.ahmed@mtkresearch.com

**Alberto Bernacchia**
联发科技研究院
alberto.bernacchi@mtkresearch.com

**Ciara Pike-Burke**
伦敦帝国学院
c.pike-burke@imperial.ac.uk

2023年2月2日

**摘要**

通过昂贵且带噪声的评估对未知函数进行黑箱优化，是机器学习、学术研究和工业生产中普遍存在的问题。该问题可以抽象为基于核的老虎机问题（也称为贝叶斯优化）：学习者旨在通过序贯噪声观测来优化核化函数。现有研究大多假设反馈可以立即获得，但在许多现实场景中这一假设并不成立，包括推荐系统、临床试验和超参数调优。我们研究随机延迟反馈下的核老虎机问题，并提出一种算法，其遗憾界为 $\tilde{\mathcal{O}}(\sqrt{\Gamma_k(T)T + \mathbb{E}[\tau]})$，其中 $T$表示时间步数，$\Gamma_k(T)$ 表示具有 $T$ 个观测的核的最大信息增益，而 $\tau$ 表示延迟随机变量。这显著优于 $\tilde{\mathcal{O}}(\Gamma_k(T)\sqrt{T + \mathbb{E}[\tau]\Gamma_k(T)})$所报告的当前最优遗憾界，见Verma 等人 (2022年）。尤其是，对于非常不光滑的核，信息增益会随时间近乎线性增长，使现有结果变得毫无意义。我们还通过模拟验证了理论结果。

## 1 引言

核老虎机问题是一种灵活的框架，涵盖了通过连续输入查询来学习优化未知函数的问题。通常，游戏按轮次进行，每一轮中，学习者选择一个输入点进行查询，然后立即收到该点处函数的噪声观测。学习者可以立即利用这一观测，改进其对下一步查询哪个点的决策。由于核老虎机问题具有普适性，因此在实践中非常流行。具体而言，它使我们能够依次学习优化各种不同的函数，而无需了解函数形式的诸多细节。

然而，在许多可能需要使用核老虎机的场景中，我们还必须应对延迟观测。例如，考虑使用核老虎机按序学习如何为化学实验选择最优条件。化学反应可能不会即时发生，而是在未来某个随机时刻才发生。如果我们开始按顺序开展一系列实验，就可以在收到前一次实验的随机延迟反馈之前启动新的实验。然而，在这种情况下，我们必须在收到前几次实验的所有反馈之前，就为后续实验更新实验条件。推荐系统、临床试验和超参数调优中也会出现类似情况，因此，设计能够处理延迟反馈的核老虎机算法具有实际意义。此外，现有核老虎机算法面临的一大挑战是计算复杂度。一般来说，在每一轮 $t$ 中，核老虎机算法都需要针对 $t$ 个已观测数据点拟合核模型，其复杂度可能达到 $\mathcal{O}(t^3)$。为降低复杂度，近期研究开始关注核老虎机算法的批处理版本。这些算法使用同一个模型选择 $\tau$ 个输入值，然后在收到批次中的所有观测后更新模型。这相当于每个观测的接收延迟至多为 $\tau$，因此可以视为核老虎机中的延迟反馈情形。

本文研究具有随机延迟反馈的核老虎机问题。我们提出带延迟的批量纯探索（BPE-Delay）——将批量纯探索算法（Li 和 Scarlett，2022年）改编至延迟反馈设定——并证明，在对未知延迟分布作出温和假设的条件下，BPE-Delay 能够实现近乎最优的遗憾。具体而言，我们证明 BPE-Delay 的遗憾按 $\tilde{\mathcal{O}}(\sqrt{\Gamma_k(T)T + \mathbb{E}[\tau]})$[1] 的量级增长，其中 $T$ 为时间步数，$\Gamma_k(T)$ 为核 $k$ 在具有 $T$ 个观测时的最大信息增益（见第 2 节），而 $\tau$ 为延迟随机变量。该结果从根本上表明，延迟反馈对该算法遗憾的影响与时域 $T$ 和问题参数 $\Gamma_k(T)$ 无关。这一理想性质意味着，随着 $T$ 或 $\Gamma_k(T)$ 增大，延迟的影响仍保持不变。

我们注意到，对于线性模型，$\Gamma_k(T) = \mathcal{O}(d\log(T))$，其中 $d$ 是输入维度。对于平方指数核（SE）等非常光滑的核，$\Gamma_k(T) = \text{poly}\log(T)$。然而，对于光滑度较低的核，$\Gamma_k(T)$ 可以变得任意接近 $T$ 上的线性。例如，对于光滑度参数为 $\nu$ 的 Matérn 核，我们有 $\Gamma_k(T) = \tilde{\mathcal{O}}(T^{\frac{d}{2\nu+d}})$（Vakili 等人，2021c）。因此，与现有的延迟反馈核老虎机研究相比，我们的结果在理论上有显著改进。现有研究针对一种基于上置信界（UCB）观测点获取的算法，给出了 $\tilde{\mathcal{O}}(\Gamma_k(T)\sqrt{T + \mathbb{E}[\tau]\Gamma_k(T)})$ 遗憾界（Verma 等人，~2022）。具体而言，对于非光滑核，我们的结果将延迟惩罚从可能接近 $\mathcal{O}(\mathbb{E}[\tau]T)$ 降至仅 $\mathcal{O}(\mathbb{E}[\tau])$。我们在第7节中证明，这些理论改进在实验中也有所体现。此外，将我们的结果应用于线性老虎机这一特殊情形（采用线性核的核老虎机问题）时，与 Howson 等人（2022）的当前最佳结果相比，我们将遗憾界中延迟相关项对输入维度的依赖降低了 $d^{3/2}$ 倍。第2节对相关工作进行了详细比较。

## 2 背景与相关工作

本节概述即时反馈核老虎机的相关背景，以及较简单的随机老虎机形式（即 $K$ 臂老虎机和线性老虎机）中的延迟反馈。随后，我们将详细比较与延迟反馈核老虎机最相关的工作。

即时反馈核老虎机。在典型的即时反馈核老虎机设定中，经典算法（例如，基于上置信界（高斯过程置信上界）选择查询点的算法（Srinivas 等人，2010），以及汤普森采样（Chowdhury 和 Gopalan, _2017））可实现 $\tilde{\mathcal{O}}(\Gamma_k(T)\sqrt{T})$ 的遗憾界。该遗憾界并非最优（详情参见 Vakili 等人，2021e），通过较新的工作可以将其改进为 $\tilde{\mathcal{O}}(\sqrt{\Gamma_k(T)T})$。具体而言，Li 和 Scarlett（2022）提出的批量纯探索（BPE），以及 Salgia 等人（2021）提出的阈值域收缩（GP-ThreDS），均可实现这一改进后的遗憾界。其中，$\Gamma_k(T)$ 是核特定的复杂度项，可解释为信息增益或有效维度。

**信息增益与有效维度**。尽管常见核的特征空间表示具有无限维度，但对于有限数据集，只有有限个特征会对基于核的回归产生显著影响。这便引出了有效维度的定义。具体而言，考虑一个有限的集合 $\mathbf{X}_T = \{X_1, \dots, X_T\}$，其中包含观测点，以及一个正定核 $k\colon\ \mathcal{X} \times \mathcal{X} \to \mathbb{R}$。令 $\mathbf{K}_{\mathbf{X}_T} = [k(X_t, X_{t'})]_{t,t'=1}^T$ 为由观测点之间两两计算的核值构成的核矩阵。给定核和观测集时，有效维度通常定义为（Zhang，2005年；Valko 等人，2013年）

$$\tilde{d}_k(T) = \operatorname{tr}\left(\mathbf{K}_{\mathbf{X}_T}(\mathbf{K}_{\mathbf{X}_T} + \lambda^2\mathbf{I}_T)^{-1}\right), \tag{1}$$

其中，$\lambda > 0$ 是一个自由参数，$\mathbf{I}_T$ 表示维度为 $T$的单位矩阵。

定义一个相关的信息增益概念也很有用。为此，假设 $f$ 是定义域 $\mathcal{X}$ 上的一个零均值高斯过程（GP），其核为 $k$。信息增益指的是数据值 $\mathbf{y}_T = [y_t = f(X_t) + \epsilon_t]_{t=1}^T$ 与 $f$之间的互信息 $\mathcal{I}(\mathbf{y}_T; f)$（假设采用替代高斯过程分布，并且零均值高斯噪声 $\epsilon_t$ 的方差为 $\lambda^2$）。根据两个多元高斯分布之间互信息的闭式表达式（Cover，1999年），可得 $\mathcal{I}(\mathbf{y}_T; f) = \frac{1}{2}\log\det\left(\mathrm{I}_T + \frac{1}{\lambda^2}\mathbf{K}_{\mathbf{X}_T}\right)$.。接下来，我们将数据无关且特定于核的最大信息增益定义如下（Srinivas 等人，2010年）：

$$\Gamma_k(T) = \sup_{\mathbf{X}_T \subset \mathcal{X}} \mathcal{I}(\mathbf{y}_T; f). \tag{2}$$

众所周知，信息增益与有效维度至多相差对数因子。具体而言，我们~有 $d_k(T) \le \mathcal{I}(\mathbf{y}_T; f)$，且 $\mathcal{I}(\mathbf{y}_T; f) = \mathcal{O}(d_k(T)\log(T))$（Calandriello 等人，2019年）。对于特定的核，Srinivas 等人（2010年）和 Vakili 等人（2021b,c）推导出了 $\Gamma_k(T)$ 的显式界。我们以信息增益的形式给出遗憾界；至多相差对数因子时，这些遗憾界也可轻松改写为有效维度的形式。

[1]我们使用记号 $\mathcal{O}$ 和 $\tilde{\mathcal{O}}$ 分别表示阶以及忽略对数因子后的阶。

**遗憾下界**。 对于常用的平方指数核和 Matérn 核，Scarlett 等人(2017年) 推导了遗憾下界（在即时反馈设定下）。具体而言，他们证明了 $\Omega(\sqrt{T}(\log(T))^{d/2})$ 和 $\Omega(T^{\frac{\nu+d}{2\nu+d}})$ 分别是平方指数核和 Matérn 核情况下的遗憾下界。这些下界与 GP-ThreDS （ Salgia 等人，2021 年）和 BPE （ Li 和 Scarlett，2022 年）的遗憾界相匹配，差距仅为对数因子。

**随机多臂老虎机中的延迟反馈**。 在存在 $K$ 独立臂的随机多臂老虎机问题中，延迟反馈已得到广泛研究。 Joulani 等人(2013年)提出了一种基于队列的包装算法，将其与任意多臂老虎机算法结合使用时，相较于无延迟设定，遗憾会增加 $\mathbb{E}[\tau]$ 的附加惩罚。他们还表明，仅使用已观测数据的置信上界算法（Auer等人，2002）也能得到类似的遗憾界 $\tilde{\mathcal{O}}(\sqrt{KT}+\mathbb{E}[\tau])$。 Mandel等人 (~~2015~~)也提出了一种具有相同遗憾界的基于队列的算法。 Vernade等人 (2017年)开发了一种置信上界算法，用于处理延迟反馈；当观测的延迟超过阈值时，这些观测会被删失。值得注意的是，我们无法将这些基于队列的方法直接扩展到核老虎机问题，因为该问题中存在大量相关臂。不过，我们证明，即使在我们更具挑战性的环境中，也能保持相同的附加惩罚。此外，尽管可以将基于置信上界的算法扩展到我们的环境，但在第7节中，我们通过实验表明，所提出的算法比使用延迟版本的高斯过程置信上界表现更好。

此外，也有研究探讨了延迟反馈对广义线性老虎机的影响。Zhou 等人(2019年）和Howson 等人(2022年）对乐观（UCB）算法进行了调整，以应对次指数延迟下的延迟反馈。Zhou 等人(2019年) 得到了一个按 $\tilde{\mathcal{O}}(d\sqrt{T}+\sqrt{dT\mathbb{E}[\tau]})$增长的遗憾界，而Howson 等人(2022年）得到了一个改进的遗憾界 $\tilde{\mathcal{O}}(d\sqrt{T}+d^{\frac{3}{2}}\mathbb{E}[\tau])$，但在该研究中，延迟惩罚仍取决于维度；而在具有延迟反馈的随机 $K$臂老虎机问题中，情况并非如此。将其应用于这一设定时，我们证明，所提出的算法消除了维度与延迟之间的这种相互作用。具体而言，在特殊情形——线性老虎机（采用线性核的核老虎机问题）——下，我们的结果将遗憾界中与延迟相关的项改进了 $d^{\frac{3}{2}}$ 倍。我们还指出，Vernade等人(2020年）将其关于带删失延迟反馈的研究扩展到了上下文线性老虎机设定，而Dudik等人(2011）研究了上下文老虎机设定中的恒定延迟，不过这些设定与我们的设定并不能直接比较。

核老虎机中的延迟反馈。与我们工作最相关的是 Verma 等人（2022），他们也研究了具有随机延迟反馈的核老虎机问题。Verma 等人（2022）针对延迟反馈设定，提出了基于高斯过程置信上界（Srinivas 等人，2010）和汤普森采样（Chowdhury 和 Gopalan，2017）的算法。这些算法称为 GP-UCB-SDF 和 GP-TS-SDF（其中 SDF 表示随机延迟反馈），它们用函数最小值（假设该值已知）替代因延迟而不可用的反馈。他们为该算法给出了遗憾界 $\tilde{\mathcal{O}}(\Gamma_k(T)\sqrt{T}+\mathbb{E}[\tau]\Gamma_k(T))$。这优于直接套用现有算法（其遗憾界为 $\mathcal{O}(\Gamma_k(T)\sqrt{T\mathbb{E}[\tau]})$），但遗憾界中与延迟有关的项仍按 $\Gamma_k(T)$ 的量级增长。相比之下，我们的算法不需要额外获知函数最小值（我们只需忽略不可用的观测）。此外，我们的结果显著优于 Verma 等人（2022），完全解耦了延迟和问题参数的影响。我们的遗憾界在已知遗憾下界的情形下也达到了阶最优，无法进一步改进；特别是，对于平方指数核和 Matérn 核，在即时反馈设定下，Scarlett 等人（2017）给出的界即属此类。

具有批量观测的核老虎机可视为我们延迟反馈框架的一种特殊情况，其延迟为常数（非随机）。具体而言，批次中各点的观测会在固定延迟后可用，且延迟至多为批次长度。典型例子包括 Desautels 等人（2014）；Vakili 等人（2021d）；Daxberger 和 Low（2017）；Chowdhury 和 Gopalan（2019），这些工作均基于 Desautels 等人（2014）提出的幻觉技术。在该技术中，将不可用观测幻觉为仅根据可用反馈得到的基于核的预测。这等价于在更新不确定性估计时使用所有选定的观测点（包括具有延迟观测的点），同时使预测不受不可用观测的影响（Chowdhury 和 Gopalan，2019）。相比之下，Verma 等人（2022）将不可用观测设为函数最小值（假定该值已知）。在我们的算法中，预测时会直接忽略不可用观测，同时采用两种不确定性估计变体（一种使用所有观测点更新，另一种仅使用反馈已可用的观测点更新）。详情见第5节。尽管批量设置因延迟固定且已知而简单得多，我们的遗憾界仍显著优于该设置下已有的结果。具体而言，在批量设置中，Chowdhury 和 Gopalan（2019）给出的已知最佳遗憾界为 $\tilde{\mathcal{O}}(\Gamma_k(T)\sqrt{T}+\tau\sqrt{T\Gamma_k(T)})$；为便于与我们的结果比较，我们在此滥用记号，用 $\tau$ 表示批量大小。定理 6.1 改进了该遗憾界中的两项。

## 3 问题定义

考虑一个正定核 $k$： $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$ 定义在一个紧致的 $d$维集合 $\mathcal{X} \subset \mathbb{R}^d$ 上。设函数希尔伯特空间 $\mathcal{H}_k$ 定义在 $\mathcal{X}$ 上，并配备内积 $\langle \cdot, \cdot \rangle_{\mathcal{H}_k}$；若其满足以下条件，则称其为再生核希尔伯特空间（再生核希尔伯特空间），其再生核为 $k$。对于所有 $x \in \mathcal{X}, k(\cdot, x) \in \mathcal{H}_k$，以及所有 $x \in \mathcal{X}$ 和 $f \in \mathcal{H}_k$，均有$\langle f, k(\cdot, x) \rangle_{\mathcal{H}_k} = f(x)$（再生性质）。

我们考虑一个核老虎机问题。假设存在一个未知目标函数 $f$： $\mathcal{X} \to \mathbb{R}$，它是已知核 $k$的再生核希尔伯特空间中的一个函数。这个假设非常一般，因为常见核的再生核希尔伯特空间能够逼近欧几里得空间紧子集上的几乎所有连续函数（Srinivas 等人，2010年）。我们的目标是找到一个输入 $x^*$，使未知函数 $f$达到最大值，即 $x^* \in \arg\max_{x \in \mathcal{X}} f(x)$。为此，我们可以依次在选定的一系列观测点 $f$处查询 $X_t \in \mathcal{X}$、 $t = 1$、2、.....，并获得带噪反馈 $y_t = f(X_t) + \epsilon_t$，其中 $\epsilon_t$ 是一列相互独立、均值为零的次高斯噪声。
我们用 该问题中任何方法的性能都以其遗憾来衡量。遗憾定义为 s

$$\mathcal{R}(T) = \sum_{t=1}^{T} f(x^*) - f(X_t) \tag{3}$$

并表示在 $X_1$、.....、 $X_T$处查询 $f$而不是在未知最大值 $x^* \in \arg\max_{x \in \mathcal{X}} f(x)$ 处查询所造成的累积损失。

为了使问题易于处理，我们采用以下文献中常见的假设。
**假设 3.1**。 目标函数 $f$的再生核希尔伯特空间范数有界，即 $\|f\|_{\mathcal{H}_k} \leq C_k$，其中某个 $C_k > 0$，记号 $\|\cdot\|^2_{\mathcal{H}_k} = \langle \cdot, \cdot \rangle_{\mathcal{H}_k}$表示再生核希尔伯特空间范数。
**假设 3.2**。 观测噪声项 $\epsilon_t$ 是相互独立的 $\sigma$-次高斯随机变量，且均值为零。也就是说，对于所有 $t \in \mathbb{N}$、所有 $\eta \in \mathbb{R}$，以及某个 $\sigma > 0$， $\epsilon_t$ 的矩母函数满足 $\mathbb{E}[\exp(\eta \epsilon_t)] \leq \exp(\frac{\eta^2 \sigma^2}{2})$。

**延迟反馈**。 本文研究具有随机延迟反馈的核老虎机问题。在这种设定下，每个时刻 $t$，环境除了生成观测 $y_t$ 之外，还会独立生成一个随机延迟 $\tau_t \geq 0$。学习者会在观测时间 $t$ 收到在时刻 $t + \tau_t$所做决策的反馈。我们假设 $\tau_1$，.....， $\tau_T$表示一系列独立同分布的次指数随机变量。为简单起见，我们假设 $\tau_t$ 随机变量是离散的，不过，只需考虑 $\tau_t + t \leq s$ 对所有 $s \geq t$成立的事件，即可轻松将我们的结果扩展到连续延迟。我们对延迟分布的假设将在下文正式给出，该假设与文献中通常采用的假设相同（Zhou 等人，2019年；Howson 等人，2022年）。
**假设 3.3**。 延迟 $\tau_t$ 是独立同分布的次指数随机变量。也就是说，对所有 $t \in \mathbb{N}$，存在某个 $\xi$， $b > 0$，对所有 $|\eta| \leq \frac{1}{b}$，$-\tau_t$的矩母函数满足 $\mathbb{E}\left[\exp\left(\eta(\tau_t - \mathbb{E}[\tau_t])\right)\right] \leq \exp(\frac{\eta^2 \xi^2}{2})$。

## 4 置信界

基于核的回归可提供强大的预测器和不确定性估计。这些结果可用于构造未知目标函数 $f$的置信区间，而置信区间是开发基于核的老虎机问题算法的重要基础。考虑一组观测点 $\mathbf{X}_t = \{X_1, \ldots, X_t\}$，以及对应的观测值向量 $\mathbf{y}_t = [y_1, \ldots, y_t]^\top$。回顾 $\mathbf{K}_{\mathbf{X}_t} = [k(X_s, X_{s'})]_{s,}^{t}$ 它是由观测点之间两两核值构成的正定核矩阵。我们有以下针对 $f$的预测器和不确定性估计，分别可解释为 $f$ 的代理高斯过程模型在核 $k$下的均值和方差（例如，参见 Rasmussen and Williams，2006；Kanagawa 等人，2018）

$$\begin{aligned} \mu_{\mathbf{X}_t, \mathbf{y}_t}(x) &= \mathbf{k}_{\mathbf{X}_t}^\top(x)(\mathbf{K}_{\mathbf{X}_t} + \lambda^2 \mathbf{I}_t)^{-1} \mathbf{y}_t \\ \sigma^2_{\mathbf{X}_t}(x) &= k(x, x) - \mathbf{k}_{\mathbf{X}_t}^\top(x)(\mathbf{K}_{\mathbf{X}_t} + \lambda^2 \mathbf{I}_t)^{-1} \mathbf{k}_{\mathbf{X}_t}(x), \end{aligned}$$

其中，$\mathbf{k}_{\mathbf{X}_t}(x) = [k(x_1, x) \ldots, k(x_t, x)]^\top$ 是观测点与关注点之间的核值向量， $\lambda > 0$ 是一个自由参数。

有了上述预测器和不确定性估计，我们便可得到以下置信界。

**引理 4.1.** *[(Vakili 等人，2021a)]*考虑式 (4)中给出的预测器和不确定性估计。假设观测集 $\mathbf{X}_t$ 的选取独立于观测噪声 $\epsilon_s$，$s = 1, \ldots, t$。在假设 *3.1* 和 *3.2*下，对于任意固定的 $1-\delta$，以下不等式均以概率 $x \in \mathcal{X}$成立，

$$f(x) \leq \mu_{\mathbf{X}_t, \mathbf{y}_t}(x) + \beta(\delta)\sigma_{\mathbf{X}_t}(x),$$
$$f(x) \geq \mu_{\mathbf{X}_t, \mathbf{y}_t}(x) - \beta(\delta)\sigma_{\mathbf{X}_t}(x),$$

其中 $\beta(\delta) = C_k + \frac{\sigma}{\lambda}\sqrt{2\log(\frac{1}{\delta})}$，$C_k$和 $\sigma$是假设 *3.1* 和 *3.2*中给出的参数，而 $\lambda$是 (4)中的参数。

如果定义域 $\mathcal{X}$是有限的，那么通过并集界，可以直接由该结果得到一致置信界，并且可以用 $\frac{\delta}{|\mathcal{X}|}$替代 $\delta$。—对于连续定义域，可以使用技术性的离散化论证，将引理 4.1扩展为一致界，前提是满足以下连续性假设。

**假设 4.2**。对于每个 $t \in \mathbb{N}$，存在一个对 $\mathcal{X}$的离散化 $\mathbb{X}$，使得对于任意满足 $\|f\|_{\mathcal{H}_k} \leq C_k$的 $f \in \mathcal{H}_k$，都有 $f(x) - f([x]) \leq \frac{1}{t}$，其中 $[x] = \text{argmin}_{x' \in \mathbb{X}} ||x' - x||_{l^2}$是 $\mathbb{X}$中距离 $x$最近点，且 $|\mathbb{X}| \leq cC_k^d t^d$，其中 $c$是一个与 $t$ 和 $C_k$无关的常数。

假设 4.2 是一个温和的技术性假设，对于典型的核（例如平方指数核和 Matérn 核）以及 $\nu > 1$ (Srinivas 等人，2010年；Chowdhury 和 Gopalan，2017年；Vakili 等人，2021a）均成立。

**引理 4.3**（Vakili 等人中的推论 3.7 的一个特例，2022年）。**在引理** *4.1*的设定和假设下，在假设 *4.2*下，以下不等式以至少 $x \in \mathcal{X}$的概率对 $1-\delta$一致成立，

$$f(x) \leq \mu_{\mathbf{X}_t, \mathbf{y}_t}(x) + \frac{2}{t} + \beta'_\delta(t)(\sigma_{\mathbf{X}_t}(x) + \frac{2}{\sqrt{t}}),$$
$$f(x) \geq \mu_{\mathbf{X}_t, \mathbf{y}_t}(x) - \frac{2}{t} - \beta'_\delta(t)(\sigma_{\mathbf{X}_t}(x) + \frac{2}{\sqrt{t}}),$$

其中 $\beta'_\delta(t) = \beta(\frac{\delta}{2\Gamma_t})$, $\Gamma_t = c(\tilde{C}_k(\frac{\delta}{2}))^d t^d$, $\tilde{C}_k(\delta) = C_k + \frac{\max\{\sigma, k_{\max}\}\sqrt{t}}{\lambda}\sqrt{2\log(\frac{2t}{\delta})}$, $k_{\max} = \max_{x \in \mathcal{X}} k(x, x$

˜引理 4.3使用关于 $\mu_{\mathbf{X}_t, \mathbf{y}_t}$ 的再生核希尔伯特空间范数的高概率界 $C_k(\delta)$（由于随机噪声，$\mu_{\mathbf{X}_t, \mathbf{y}_t}$ 是一个随机变量），并结合对 $\mu_{\mathbf{X}_t, \mathbf{y}_t}$和 $\sigma_{\mathbf{X}_t}$ 应用连续性假设来构造置信界。实际上，置信区间宽度的乘数 $\beta'_\delta(t)$按以下方式缩放：

$$\beta'_\delta(t) = \mathcal{O}\left(C_k + \frac{\sigma}{\lambda}\sqrt{d\log(\frac{tC_k}{\delta})}\right). \tag{4}$$

在具有延迟反馈的核老虎机设定中，由于反馈延迟，部分观测可能尚不可用。˜令 $\mathbf{X}_t = \{X_s \in \mathbf{X}_t : s + \tau_s \leq t\}$ 表示截至时间 $t$ 已收到反馈的观测点集合。由于延迟是随机的，这实际上是一个随机观测集合。为简化记号，我们用 $\mu_t, \sigma_t$ 表示使用 $\mathbf{X}_t$、$\mathbf{y}_t$ 时的预测器和不确定性估计，并用 $\tilde{\mu}_t, \tilde{\sigma}_t$ 表示使用 $\mathbf{X}_t$, ˜$\mathbf{y}_t$ 时的预测器和不确定性估计，其中 ˜$\mathbf{y}_t = [f(X_s) + \epsilon_s]^\top_{X_s \in}$ ˜$\mathbf{X}_t$是可用观测组成的向量。值得注意的是，在延迟反馈设定中，除了 $\tilde{\mu}_t, \tilde{\sigma}_t$ 之外，我们还可以计算 $\sigma_t$，因为它只取决于选定输入，而不取决于观测。另一方面，$\mu_t$ 无法计算，因为它依赖于部分缺失的反馈。设计算法 BPE-Delay 时，我们会用到这三个可用统计量：$\tilde{\mu}_t, \tilde{\sigma}_t, \sigma_t$。

## 5 带延迟的批量纯探索

在本节中，我们介绍所提出的算法：带延迟的批量纯探索（BPE-Delay）。该算法按轮次进行 $r = 1, 2, \ldots, R$，其中 $R$ 是总轮数。每轮包含 $t_r$ 个时间步，因此 $\sum_{r=1}^{R} t_r = T$。每轮中，观测点根据最大不确定性采集函数（定义见（6））进行选择。在每轮结束时，利用置信区间移除不太可能成为 $f$ 的最大化点。

轮次 $r$ 的长度 $t_r$ 随 $r$ 增加，并经过谨慎选择，以兼顾每一轮所需的探索、随机延迟反馈以及轮数。具体而言，我们设定 $t_r = \lceil q_r + u_T(\delta) \rceil$

其中， $u_T(\delta)$ 是延迟随机变量的一个 $1-\delta$ 上界， $q_r = \sqrt{q_{r-1}T}$ 则通过递归确定，适用于 $r \geq 1$，初始值为 $q_0 = 1$。与延迟相关的量设为 $u_T(\delta) = \mathbb{E}[\tau] + \psi_T(\frac{\delta}{2})$，其中

$$\psi_t(\delta) = \min\left\{\sqrt{2\xi^2 \log(\frac{3t}{2\delta})},\ 2b\log(\frac{3t}{2\delta})\right\}, \tag{5}$$

其中 $\xi$ 和 $b$ 是假设 3.3中指定的参数。在第 6 节的分析中，我们将看到， $u_T(\delta)$是一个关于延迟随机变量的 $1-\frac{\delta}{2}$ 上置信界。事实证明，这种取决于 $t_r$ 的 $u_T(\delta)$ 选择对于改进算法对延迟的依赖至关重要。如果我们知道观测不存在延迟，就可以将 $u_T(\delta)$ 设为零。最后一轮的长度也可以轻松调整，以确保 $\sum_{t=1}^{R} t_r = T$。

BPE-Delay维护一个由 $\mathcal{X}_r$ 个 $f$的潜在最大化点组成的集合。利用围绕每个 $x \in \mathcal{X}_{r-1}$ 的置信区间，该集合会从 $\mathcal{X}_{r-1}$ 中递归剪枝，以高概率剔除那些并非最优的点。我们从完整的输入空间 $\mathcal{X}_0 = \mathcal{X}$开始。在每一轮 $r$中，根据最大不确定性采集，从 $\mathcal{X}_r$ 中选取第 $k$个观测点$X_{k,\ r}$：

$$X_{k,r} \leftarrow \arg\max_{x \in \mathcal{X}_r} \sigma_{k-1,r}(x), \tag{6}$$

其中，$\sigma_{k,r}(\cdot)$ 表示使用点 $\mathrm{X}_{k,r} = \{X_{1,r}, X_{2,r}, \ldots, X_{k,r}\}$计算式 (4) 中的不确定性估计。BPE-Delay 仅使用截至轮次 $r$ 所选的输入来计算采集函数，并根据不确定性估计选择轮次 $r$ 剩余的点。虽然使用所有过往观测可能有效，但要建立相应的遗憾界会变得困难。原因在于，在观测点之间建立这种依赖关系，会使引理 4.1 中给出的紧置信区间失效。尽管如此，(Li 和 Scarlett，2022年) 在实验中表明，丢弃前几轮收集的数据虽然看似浪费，但并不会显著损害性能。根据式 (4) 中不确定性估计的定义可知，它不依赖观测值。因此，即使由于延迟而无法获得所有观测值，$\sigma^2_{k,r}(\cdot)$ 仍有良好定义。为便于说明，我们使用双重索引 $r = 1, 2, \ldots, k = 1, 2, \ldots, t_r$。不过，可以通过将所有轮次中的观测点串联起来，使用 $t = 1, 2, \ldots$ 为观测点编索引（见算法1）。

**算法1**带延迟的批量纯探索（BPE-Delay）

输入：动作集 $\mathcal{X}$，步数 $T$；初始化： $\mathcal{X}_1 \leftarrow \mathcal{X}, q_0 = 1, t = 1$； # 算法按轮次进行 $r = 1, 2, \ldots$。对于 $r = 1, 2, \ldots$， 直到 $t \leq T$do$q_r = \lceil \sqrt{T q_{r-1}} \rceil$ # 使用 $q_r$ $t_r$ 来确定 ( ) # 是轮次 $r$ 的长度对于 $k = 1, \ldots, t_r$ do # 每轮中的点由最大不确定性采集函数选出$X_{k,r} \leftarrow \arg\max_{x \in \mathcal{X}_r} \sigma_{k-1,r}(x) X_t \leftarrow X_{k,r} t = t + 1$end for # 使用置信区间移除不太可能是 $f$ 的最大化点的输入点$\mathcal{X}_{r+1} \leftarrow \{x \in \mathcal{X}_r : U_{r,\delta}(x) \geq L_{r,\delta}(z)\ \forall z \in \mathcal{X}_r\}$end for $t_r = \lceil q_r + u_T\ \delta\ \rceil$ $t_r$

为了对比第 $r$ 轮中根据全部观测（忽略延迟）得出的统计量与根据可用观测（考虑延迟）得出的统计量，我们回顾一下记号 $\tilde{\mu}_{k,r}$ 和 $\tilde{\sigma}_{k,r}$，它们表示仅使用可用观测的统计量。具体而言，$\tilde{\mu}_{k,r}$ 和 $\tilde{\sigma}_{k,r}$ 分别表示在第 $r$ 轮使用集合 $\mathrm{X}_{k,r} = \{X_{s,r} \in \mathrm{X}_{k,r}$ 选取 $k^{\tilde{}}$ 个观测点后得到的预测和不确定性估计：$s + \tau_{s,r} \leq k\}$ 是在第 $r$ 轮选取的点，且在第 $r$ 轮有可用观测 $\tilde{\mathrm{y}}_{k,r} = [y = f(X_{s,r}) + \epsilon_{s,r}]_{X_{s,r} \in \tilde{\mathrm{X}}_{k,r}}$。

在每一轮结束时， $r$利用第 $r$轮的可用观测，我们为 $f(\cdot)$、 $U_r(\cdot)$和 $L_r(\cdot)$分别构造上置信界和下置信界。利用这些界可以移除不太可能成为 $f$最大化点的观测点。具体而言，如果存在一个观测点 $z \in \mathcal{X}_r$，其下置信界大于 $x \in \mathcal{X}_r$的上置信界，那么 $x$ 就不太可能是 $f$ 的最大化点，因此我们可以将其移除。置信区间宽度的选取方式可确保算法中使用的所有置信区间以高概率成立（详情请参见定理6.1 和6.2）。

我们强调，虽然我们使用 $\sigma_{k,r}$ 来选择每一轮中的观测点，但 $\tilde{\mu}_{k,r}$ 和 $\tilde{\sigma}_{k,r}$ 用于在每轮结束时构造置信区间。使用 $\sigma_{k,r}$ 可以避免因反馈不可用而重复选择观测点，而有效的置信区间只能使用 $\tilde{\mu}_{k,r}$ 和 $\tilde{\sigma}_{k,r}$构造。

BPE-Delay 算法的完整伪代码见算法 1。

## 6 分析

本节将分析 BPE-Delay 算法。具体而言，我们证明其遗憾具有高概率$\tilde{\mathcal{O}}(\sqrt{T\Gamma_k(T)} + \mathbb{E}[\tau])$ 界。该遗憾界将延迟的影响与主要的、依赖时间的遗憾项完全解耦。值得注意的是，我们结果中由延迟导致的额外遗憾显著小于已有研究，并且与更简单的 $K$臂老虎机问题中观察到的情况一致（Joulani 等人，2013年）。

由于置信区间存在细微差异，进而导致遗憾界略有不同，我们分别考虑有限和连续 $\mathcal{X}$ 两种情况。特别地，当 $\mathcal{X}$ 为有限时，遗憾界的缩放因子为 $\mathcal{O}(\sqrt{\log(|\mathcal{X}|)})$。对于连续且紧致的 $\mathcal{X} \subset \mathbb{R}^d$，在假设 4.2下，我们证明了一个类似的遗憾界，其缩放因子为 $\mathcal{O}(\sqrt{d})$，而非 $\mathcal{O}(\sqrt{\log(|\mathcal{X}|)})$。在核老虎机设定中，$\sqrt{d}$ 通常在 $\mathcal{O}$ 记号中被隐去，视为一个不会随 $T$增长的常数；而在线性老虎机设定中，$d$ 对 $d$ 的依赖关系应明确体现出来，这对评估算法至关重要。

**定理 6.1**。 考虑第 *3* 节所述的带延迟反馈的基于核的老虎机问题。当 $\mathcal{X}$ 为有限值时，将 *BPE-Delay*（算法 *1*）中的置信区间设为

$$U_{r,\delta}(x) = \tilde{\mu}_{t_r,r}(x) + \tilde{\beta}(\delta)\tilde{\sigma}_{t_r,r}(x),$$
$$L_{r,\delta}(x) = \tilde{\mu}_{t_r,r}(x) - \tilde{\beta}(\delta)\tilde{\sigma}_{t_r,r}(x), \tag{7}$$

当 $\tilde{\beta}(\delta) = C_k + \frac{\sigma}{\lambda}\sqrt{2\log(\frac{4R|\mathcal{X}|}{\delta})}$时。在假设 *3.1*、*3.2* 和 *3.3*下，以至少 $1-\delta$的概率，*BPE-Delay* 的遗憾上界为

$$\mathcal{R}(T) = \mathcal{O}\left(\tilde{\beta}(\delta)\log\log(T)\sqrt{T\Gamma_k(T)} + \mathbb{E}[\tau]\right). \tag{8}$$

注意，当 $\mathcal{X}$为有限值时，$\beta(\delta) = \mathcal{O}(1)$ 不依赖于问题参数，例如 $d$和 $T$。在下一个定理中，我们将给出 $\mathcal{X}$ 不为有限值、而是 $\mathbb{R}^d$的紧子集时的遗憾上界。

**定理 6.2**。 考虑第 *3* 节所述的带延迟反馈的基于核的老虎机问题，在假设 *3.1*、*3.2*、*3.3* 和 *4.2*下。将 *BPE-Delay*（算法 *1*）中的置信区间设为

$$U_{r,\delta}(x) = \tilde{\mu}_{t_r,r}(x) + \frac{2}{u_r} + \tilde{\beta}'_r(\delta)(\tilde{\sigma}_{t_r,r}(x) + \frac{2}{\sqrt{u_r}}),$$
$$U_{r,\delta}(x) = \tilde{\mu}_{t_r,r}(x) + \frac{2}{u_r} + \tilde{\beta}'_r(\delta)(\tilde{\sigma}_{t_r,r}(x) + \frac{2}{\sqrt{u_r}}), \tag{9}$$

在 $\tilde{\beta}'_r(\delta) = \beta'_{u_r}(\frac{\delta}{4R})$且 $\beta'_t$由引理 *4.3*给出时，以至少 $1-\delta$ 的概率，*BPE-Delay* 的遗憾有如下界：

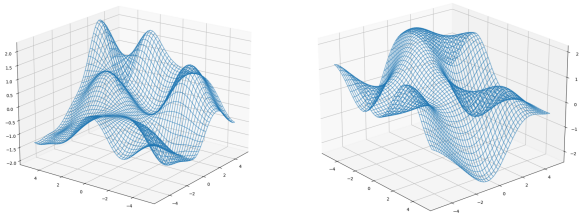
$$\mathcal{R}(T) = \mathcal{O}\left(\tilde{\beta}'_R(\delta)\log\log(T)\sqrt{T\Gamma_k(T)} + \mathbb{E}[\tau]\right). \tag{10}$$

根据引理 4.3，可以看出 $\tilde{\beta}'_R(\delta) = \mathcal{O}(C_k + \frac{\sigma}{\lambda}\sqrt{d\log(\frac{TRC_k}{\delta})})$。因此，遗憾界随 $\mathcal{O}(\sqrt{d})$ 增长；在线性老虎机中应用时，这一点尤为重要。

证明概要。根据每轮的大小 $t_r$，轮次数量 $R$有界于 $R = \mathcal{O}(\log\log(T))$。我们首先证明，利用 $t_r$ 观测得到的不确定性估计 $\tilde{\sigma}_{t_r,\,r}$ 会充分缩小。利用置信区间，我们证明：i) 最大化点 $x^*$ 在任何轮次都不会被移除；即 $x^* \in \mathcal{X}_r$，对所有 $r$均成立；ii) 选择每个点 $X_{k,\,r} \in \mathcal{X}_r$ 所产生的瞬时遗憾有界。 $f(x^*) - f(X_{k,\,r})$的界基于前一轮的探索量，并考虑了延迟。利用瞬时遗憾的界以及轮次 $r$ 的大小 $t_r$，我们可以界定第 $r$轮的遗憾。详细证明见附录。

□

定理 6.1可具体应用于许多常用核，包括 Matérn 核和平方指数核，以及线性核这一特殊情况。



(a) $f_1$ (b) $f_2$

图 1：实验中使用的目标函数。

**Co推论 6.3**。 在定理 *6.2*的设定下，以下结论以至少 $1-$ $\delta$,

- 对于光滑度参数为 $\nu$的 *Matérn* 核，

$$\mathcal{R}(T) = \tilde{\mathcal{O}}\left(T^{\frac{\nu+d}{2\nu+d}} + \mathbb{E}[\tau]\right). \tag{11}$$

- 对于平方指数核，

$$\mathcal{R}(T) = \mathcal{O}\left(\sqrt{T}(\log(T))^{d+1}\log\log(T) + \mathbb{E}[\tau]\right). \tag{12}$$

- 在线性核的情况下，

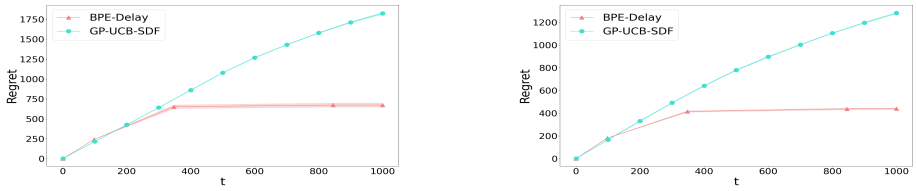
$$\mathcal{R}(T) = \mathcal{O}\left(d\sqrt{T}\log(T)\log\log(T) + \mathbb{E}[\tau]\right). \tag{13}$$

与线性老虎机情形下的下界（Lattimore 和 Szepesvári，2020年）以及 SE 和 Matérn 核情形下核老虎机的下界（Scarlett 等人，2017年）相比，推论 6.3 中给出的遗憾界在阶数上是最优的，至多相差对数因子和加性延迟惩罚；后者与其他所有问题参数无关。因此，我们将遗憾界中与延迟相关的项缩减至仅 $\mathbb{E}[\tau]$，而不影响遗憾界中的主导项。作为比较，在（广义）线性设定中，已知的最优延迟相关遗憾界为 $\mathcal{O}(d^{\frac{3}{2}}\mathbb{E}[\tau])$（Howson 等人，2022年）；我们的结果将其改善了 $d^{\frac{3}{2}}$ 倍。在核老虎机设定中，我们也将延迟相关遗憾 $\mathcal{O}(\Gamma_T\mathbb{E}[\tau])$（Verma 等人，2022年）改善至仅 $\mathcal{O}(\mathbb{E}[\tau])$。这一改进意义重大，因为对于 $\nu$ 较小且 $d$ 较大的 Matérn 核，$\Gamma_T$ 可以任意接近 $T$ 中的线性增长。此外，遗憾的第一项也从 $\tilde{\mathcal{O}}(\Gamma_T\sqrt{T})$ 改善至 $\tilde{\mathcal{O}}(\sqrt{\Gamma_T T})$。

## 7 实验

在理论分析之后，我们通过数值实验评估 BPE-Delay 的性能，并将其与 GP-UCB-SDF（Verma 等人，2022年）进行比较。我们在图 1 所示的目标函数 $f_1$ 和 $f_2$ 上测试这些算法。这些函数是通过对从多元高斯分布中随机生成的点拟合基于核的模型而生成的。这是生成再生核希尔伯特空间元素的一种常用方法（例如，参见 Chowdhury 和 Gopalan，2017年；Vakili 等人，2021a；Li 和 Scarlett，2022年）。我们使用长度尺度参数为 $l = 0.8$ 的平方指数核来生成目标函数 $f_1$，并使用长度尺度参数为 $l = 1.0$ 的平方指数核来生成目标函数 $f_2$。随后，学习者可以从均匀$50 \times 50$ 网格上的 $|\mathcal{X}| = 2500$ 个点中进行选择。采样噪声为均值为零、标准差为 $\sigma = 0.02$的高斯噪声。反馈中的随机延迟服从参数为 $\lambda$的泊松分布。BPE-Delay 的 $\psi_t$ 计算使用假设 4.2 中给出的 $\xi = 9$和 $b = 1$。累积遗憾曲线是 10 次独立实验的平均值。误差棒表示半个标准差。这些实验的代码已发布在一个匿名 GitHub 仓库中（参见补充材料）。

在图 2中，我们将 BPE-Delay 的遗憾性能与最相关的工作（Verma 等人，2022年）中提出的 GP-UCB-SDF 在与我们相同的设置下进行比较。图中显示，在两种情况下，使用 BPE-Delay 都显著改善了遗憾性能。在该实验中，延迟服从泊松分布，参数为 $\lambda = 50$。
在图 3和 4中，我们展示了延迟对这两种算法的影响。具体而言，我们将泊松延迟参数设为 $\lambda = 0$，25 和 50，同时将其余参数保持为与前一次实验相同。BPE-Delay 的遗憾

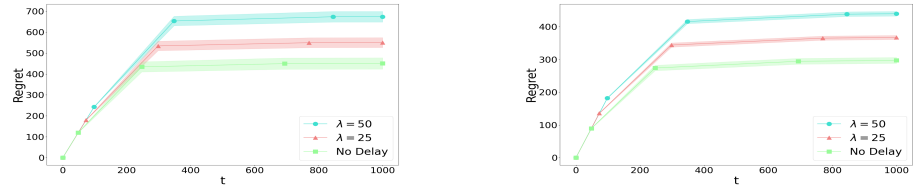



(a) $f_1$ (b) $f_2$

图 2：比较 BPE-Delay 和 GP-UCB-SDF 的遗憾性能。

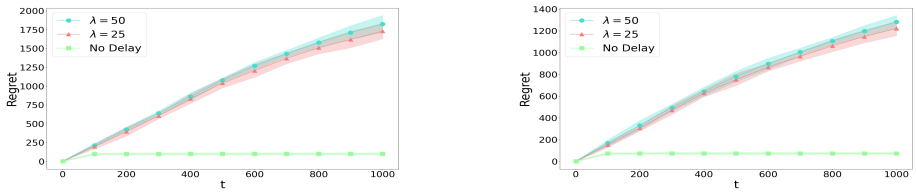
随期望延迟线性增加。另一方面，即使在中等延迟值下，GP-UCB-SDF 的遗憾也会急剧跃升。

我们还比较了 BPE-Delay 和 BPE（Li 和 Scarlett，2022年）的性能。从图5可以看出，在延迟环境中，BPE-Delay 的表现自然优于 BPE。



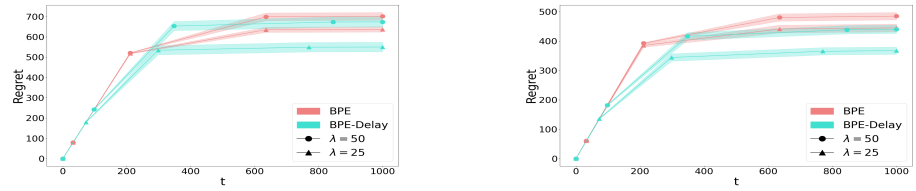

(a) $f_1$ (b) $f_2$

图 3：BPE-Delay 在不同延迟泊松参数下的遗憾性能。




(a) $f_1$ (b) $f_2$

图 4：GP-UCB-SDF 在不同延迟泊松参数下的遗憾性能。




(a) $f_1$ (b) $f_2$

图5：比较延迟反馈设定下 BPE 和 BPE-Delay 的遗憾性能。

## 8 结论

核老虎机因其在机器学习以及学术和工业实验设计中的诸多应用而备受关注。在许多现实场景中，例如推荐系统，决策产生的反馈无法立即获得，这自然促使人们提出具有随机延迟反馈的核老虎机问题。针对这一设定，我们提出了 BPE-Delay，它能够高效处理延迟反馈。我们证明了 BPE-Delay 的遗憾界具有阶数最优性，且延迟反馈带来的额外遗憾非常小，显著优于同一设定下已有的工作。我们还通过多项模拟研究表明，这些理论改进同样适用于实践。此外，在应用于线性核这一特殊情形时，与最先进的方法相比，我们的理论遗憾界将与延迟相关的惩罚降低了 $d^{\frac{3}{2}}$ 倍。在所有情形下，我们的结果首次表明，加性延迟惩罚仅取决于 $\mathbb{E}[\tau]$。这将我们对更简单的 $K$ 臂老虎机设置的认知，推广到了现实性强得多的核老虎机设定。

## 参考文献

P. Auer、N. Cesa-Bianchi 和 P. Fischer。多臂老虎机问题的有限时间分析。机器学习，47（2）：235–256，2002年。D. Calandriello、L. Carratino、A. Lazaric、M. Valko 和 L. Rosasco。采用自适应草图的高斯过程优化：可扩展且无遗憾。载于学习理论会议，2019年。S. R. Chowdhury 和 A. Gopalan。论核化多臂老虎机。载于国际机器学习会议，第844–853页，2017年。S. R. Chowdhury 和 A. Gopalan。论批量贝叶斯优化。*arXiv*预印本 *arXiv:1911.01032*，2019年。T. M. Cover。信息论要素。John Wiley & Sons，1999年。E. A. Daxberger 和 B. K. H. Low。分布式批量高斯过程优化。载于国际机器学习会议，第951–960页。PMLR，2017年。T. Desautels、A. Krause 和 J. W. Burdick。高斯过程老虎机优化中的探索-利用权衡并行化。机器学习研究杂志，15：3873–3923，2014年。M. Dudik、D. Hsu、S. Kale、N. Karampatziakis、J. Langford、L. Reyzin 和 T. Zhang。上下文老虎机的高效最优学习。载于第二十七届人工智能不确定性会议论文集，第169–178页，2011年。B. Howson、C. Pike-Burke 和 S. Filippi。重新审视广义线性老虎机中的延迟反馈。*arXiv*预印本 *arXiv:2207.10786*，2022年。P. Joulani、A. Gyorgy 和 C. Szepesvári。延迟反馈下的在线学习。载于国际机器学习会议，第1453–1461页。PMLR，2013年。M. Kanagawa、P. Hennig、D. Sejdinovic 和 B. K. Sriperumbudur。高斯过程与核方法：联系与等价性的综述。*arXiv*预印本 *arXiv:1807.02582*，2018年。T. Lattimore 和 C. Szepesvári。老虎机算法。剑桥大学出版社，2020年。Z. Li 和 J. Scarlett。少量批次的高斯过程老虎机优化。载于人工智能与统计国际会议，第92–107页。PMLR，2022年。T. Mandel、Y.-E. Liu、E. Brunskill 和 Z. Popovi´c。队列方法：处理老虎机中的延迟、启发式方法、先验数据与评估。载于第二十九届*AAAI*人工智能会议，2015年。C. E. Rasmussen 和 C. K. Williams。机器学习中的高斯过程。MIT出版社，2006年。S. Salgia、S. Vakili 和 Q. Zhao。一种基于域收缩的贝叶斯优化算法，具有阶最优遗憾性能。神经信息处理系统进展，34：28836–28847，2021年。J. Scarlett、I. Bogunovic 和 V. Cevher。含噪高斯过程老虎机优化的遗憾下界。载于学习理论会议，第1723–1742页，2017年。N. Srinivas、A. Krause、S. Kakade 和 M. Seeger。老虎机设置下的高斯过程优化：无遗憾与实验设计。载于国际机器学习会议，第1015–1022页，2010年。S. Vakili、N. Bouziani、S. Jalali、A. Bernacchia 和 D.-s. Shiu。高斯过程老虎机的最优阶简单遗憾。载于神经信息处理系统会议，2021a。S. Vakili、M. Bromberg、J. Garcia、D.-s. Shiu 和 A. Bernacchia。过参数化神经网络的统一泛化界。*arXiv*预印本 *arXiv:2109.06099*，2021b。

S. Vakili、K. Khezeli 和 V. Picheny。高斯过程老虎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，2021c。S. Vakili、H. Moss、A. Artemev、V. Dutordoir 和 V. Picheny。利用稀疏高斯过程模型实现可扩展的汤普森采样。神经信息处理系统进展，34：5631–5643，2021d。S. Vakili、J. Scarlett 和 T. Javidi。开放问题：再生核希尔伯特空间元素的紧致在线置信区间。载于学习理论会议，第4647–4652页。PMLR，2021e。S. Vakili、J. Scarlett、D.-s. Shiu 和 A. Bernacchia。基于核的学习中稀疏近似方法的收敛速率改进。载于国际机器学习会议。PMLR，2022年。M. Valko、N. Korda、R. Munos、I. Flaounas 和 N. Cristianini。核化上下文老虎机的有限时间分析。载于人工智能不确定性会议，第654–663页，2013年。A. Verma、Z. Dai 和 B. K. H. Low。随机延迟反馈下的贝叶斯优化。载于国际机器学习会议，第22145–22167页。PMLR，2022年。C. Vernade、O. Cappé 和 V. Perchet。延迟转化的随机老虎机模型。载于人工智能不确定性会议，2017年。C. Vernade、A. Carpentier、T. Lattimore、G. Zappella、B. Ermis 和 M. Brueckner。具有随机延迟反馈的线性老虎机。载于国际机器学习会议，第9712–9721页。PMLR，2020年。T. Zhang。利用有效数据维数推导核回归的学习界。神经计算，17（9）：2077–2098，2005年。Z. Zhou、R. Xu 和 J. Blanchet。具有随机延迟的广义线性上下文老虎机学习。神经信息处理系统进展，32，2019年。

## A 证明

本节中，我们给出定理 6.1、6.2 和推论6.3的详细证明。定理的证明安排如下。我们首先概述基于核的置信区间，以及关于延迟和轮数的界。在对这些界进行形式化后，我们分三步进行证明。第一步，界定每个轮次结束时任意点 $x \in \mathcal{X}_{r+1}$ 的不确定性上限 $r$。第二步，界定每个轮次选择一个点时的瞬时遗憾。第三步，界定总遗憾。在证明最后，我们讨论推论 6.3。

**置信区间**。回顾 BPE-Delay 算法中用于更新集合 $\mathcal{X}_r$ （第 $r$ 轮的潜在最大化点）的置信区间。根据引理 4.1 和 4.3，并对各轮次应用概率并集界，该算法使用的所有置信区间以至少 $1-\frac{\delta}{2}$ 的概率有效。具体而言，定义事件

$$\mathcal{E} = \{L_{r,\delta}(x) \leq f(x) \leq U_{r,\delta}(x), \forall r = 1, 2, \ldots, R, \forall x \in \mathcal{X}_r\}. \tag{14}$$

那么，我们有 $\Pr[\mathcal{E}] \geq 1 - \frac{\delta}{2}$。

**延迟**。 回顾给定的 $\psi_t(\delta)$ 定义（5）。我们得到次指数随机变量的如下集中界。

**引理 A.1** (Howson 等人 (2022年))。 设 $\tau_t - \mathbb{E}[\tau_t]$, $t = 1, 2, \ldots$, 为参数为 $\xi$ 和 $b$ （如假设 *3.3*中所规定）的独立同分布次指数随机变量。则，

$$\Pr\left[\forall t \geq 1, \tau_t \leq \mathbb{E}[\tau] + \psi_t(\delta)\right] \geq 1 - \delta. \tag{15}$$

令$\tau_{k,r}$ 表示 BPE-Delay 第 $k$轮次中第 $k$ 个观测的随机延迟。定义事件

$$\mathcal{E}' = \{\forall r, \forall k \leq q_r : \tau_{k,r} \leq u_T(\delta)\}. \tag{16}$$

利用引理 A.1 以及 $u_T$的定义，我们有 $\Pr[\mathcal{E}'] \geq 1 - \frac{\delta}{2}$。

因此， $\mathcal{E}$和 $\mathcal{E}'$ 都成立的概率为 $\Pr[\mathcal{E} \cap \mathcal{E}'] \geq 1 - \delta$。

**我们在 $\mathcal{E} \cap \mathcal{E}'$. 的条件下进行证明的其余部分**。

**第 n 轮次数量。** BPE-Delay 中的轮次数量 $R$ 由以下引理限定 a.

**引理 A.2。** 回顾 *BPE-Delay* 中轮次长度 $t_r$ 的选择。我们有 $R = \mathcal{O}(\log\log(T)$ ).

证明遵循 Li 和 Scarlett 1 在 Li 和 Scarlett (2022年) 中命题的证明所采用的类似步骤。
现在我们继续介绍定理证明的主要步骤。

**步骤 1（每轮结束时的最大不确定性）。** 在每轮 $r$中，在 $q_r$ 观测结束时，使用所有观测（包括具有延迟反馈的观测）得到的不确定性之和可以利用 $\Gamma_k(q_r)$进行界定。**引理 A.3。** *[Srinivas 等人 (2010年)]* 回顾 $\Gamma_k(t)$ 在 (2)中给出的定义，对于一组 $t$ 观测 $\mathbf{X}_t$，我们有

$$\sum_{s=1}^{t} \sigma^2_{\mathbf{X}_{s-1}}(X_s) \leq c_1 \Gamma_k(t), \tag{17}$$

其中 $c_1 = \frac{2}{\log(1+\frac{1}{\lambda^2})}$。

因此，考虑第 $r$轮的观测点，我们有

$$\sum_{k=1}^{q_r} \sigma^2_{k-1,r}(X_{k,r}) \leq c_1 \Gamma_k(q_r), \tag{18}$$

其中 $c_1$ 是引理 A.3 中给出的常数。根据基于最大不确定性采集的观测点选择规则（6），对于所有 $x \in \mathcal{X}_r$， $1 \leq k \leq q_r$

$$\sigma_{k-1,r}(x) \leq \sigma_{k-1,r}(X_{k,r}). \tag{19}$$

此外，由于核矩阵的正定性，在更大的观测集上进行条件化会降低不确定性（见（4））。因此，

$$\sigma_{q_r,r}(x) \leq \sigma_{k-1,r}(X_{k,r}). \tag{20}$$

将最后一个不等式与（18）结合，我们得到

$$\sum_{k=1}^{q_r} \sigma^2_{q_r,r}(x) \leq c_1 \Gamma_k(q_r), \tag{21}$$

这意味着

$$\sigma^2_{q_r,r}(x) \leq \frac{c_1 \Gamma_k(q_r)}{q_r}. \tag{22}$$

回顾（16）中给出的事件 $\mathcal{E}'$。在 $\mathcal{E}'$ 下， $\tilde{\sigma}^2_{t_r,\ r}$ 相较于 $\sigma^2_{q_r,\ r}$ 是在更大的观测集上进行条件化。核矩阵的正定性意味着，在更大的观测集上进行条件化会降低不确定性。因此，我们有 $\tilde{\sigma}^2_{t_r,\ r}(x) \leq \sigma^2_{q_r,\ r}(x)\ \forall x \in \mathcal{X}_r$。由此，

$$\tilde{\sigma}^2_{t_r,r}(x) \leq \frac{c_1 \Gamma_k(q_r)}{q_r}, \quad \forall x \in \mathcal{X}_r. \tag{23}$$

**步骤 2（瞬时遗憾）。** 我们可以利用引理 4.1 和 4.3中给出的再生核希尔伯特空间元素的置信区间，按如下方式界定瞬时遗憾。回顾事件 $\mathcal{E}$，其定义见（14）。
我们首先注意到，对所有 $x \in \mathcal{X}_r$,

$$\begin{aligned} U_{r,\delta}(x^*) &\geq f(x^*) \\ &\geq f(x) \\ &\geq L_{r,\delta}(x). \end{aligned}$$

因此，$U_{r,\ \delta}(x^*) \geq \max_{x \in \mathcal{X}_r} L_{r,\ \delta}(x)$ 对所有 $r$ 均成立，并且 $x^*$ 在 BPE-Delay 的任何轮次中都不会被消除。因此 $x^* \in \mathcal{X}_r$，对所有 $r$ 均成立。
现在，我们分别在有限 $\mathcal{X}$ 和连续 $\mathcal{X}$两种不同设定下，着手界定瞬时遗憾。这两种情况分别对应定理6.1 和6.2。

当 $\mathcal{X}$ 为有限时（定理6.1的设定）：

对于所有 $x \in \mathcal{X}_{r+1}$,

$$\begin{aligned} f(x^*) - f(x) &\leq U_{r,\delta}(x^*) - L_{r,\delta}(x) \\ &= \tilde{\mu}_{t_r,r}(x^*) + \tilde{\beta}(\delta)\tilde{\sigma}_{t_r,r}(x^*) - \tilde{\mu}_{t_r,r}(x) - \tilde{\beta}(\delta)\tilde{\sigma}_{t_r,r}(x^*) \\ &= L_{r,\delta}(x^*) - U_{r,\delta}(x) + 2\tilde{\beta}(\delta)\tilde{\sigma}_{t_r,r}(x^*) + 2\tilde{\beta}(\delta)\tilde{\sigma}_{t_r,r}(x) \\ &\leq 2\tilde{\beta}(\delta)\tilde{\sigma}_{t_r,r}(x^*) + 2\tilde{\beta}(\delta)\tilde{\sigma}_{t_r,r}(x). \end{aligned} \tag{24}$$

第一个和第二个等式由定理6.1中置信区间的选取得到。最后一行由 $\mathcal{X}_{r+1}$ 的选取得到，该选取确保对所有 $x$, $z \in \mathcal{X}_{r+1}$, $L_{r,\ \delta}(z) \leq U_{r,\ \delta}(x)$。

结合上一步得到的关于$\tilde{\sigma}_{t_r,\ r}(x)$的界，我们对所有 $x \in \mathcal{X}_{r+1}$的瞬时遗憾进行如下界定。

$$f(x^*) - f(x) \leq 4\tilde{\beta}(\delta)\sqrt{\frac{c_1\Gamma_k(q_r)}{q_r}}. \tag{25}$$

当 $\mathcal{X}$ 为连续时（定理6.2的设定）：

按照与前一种情况类似的步骤，对于所有 $x \in \mathcal{X}_{r+1}$,

$$\begin{aligned} f(x^*) - f(x) &\leq U_{r,\delta}(x^*) - L_{r,\delta}(x) \\ &\leq \tilde{\mu}_{t_r,r}(x^*) + \frac{2}{q_r} + \tilde{\beta}'_r(\delta)(\tilde{\sigma}_{t_r,r}(x^*) + \frac{2}{\sqrt{q_r}}) \\ &\quad - \tilde{\mu}_{t_r,r}(x) + \frac{2}{q_r} + \tilde{\beta}'_r(\delta)(\tilde{\sigma}_{t_r,r}(x) + \frac{2}{\sqrt{q_r}}) \\ &= L_{r,\delta}(x^*) - U_{r,\delta}(x) \\ &\quad + \frac{8}{q_r} + 2\tilde{\beta}'_r(\delta)\left(\tilde{\sigma}_{t_r,r}(x^*) + \tilde{\sigma}_{t_r,r}(x) + \frac{4}{\sqrt{q_r}}\right) \\ &\leq \frac{8}{q_r} + 2\tilde{\beta}'_r(\delta)\left(\tilde{\sigma}_{t_r,r}(x^*) + \tilde{\sigma}_{t_r,r}(x) + \frac{4}{\sqrt{q_r}}\right). \end{aligned}$$

结合上一步得到的 $\tilde{\sigma}_{t_r,\ r}(x)$ 的界，我们可以将所有 $x \in \mathcal{X}_{r+1}$ 的瞬时遗憾界定如下。

$$f(x^*) - f(x) \leq \frac{8}{q_r} + 4\tilde{\beta}'_r(\delta)\left(\sqrt{\frac{c_1\Gamma_k(q_r)}{q_r}} + \frac{2}{\sqrt{q_r}}\right). \tag{26}$$

**步骤 3（界定累积遗憾）：** 令 $\mathcal{R}_r = \sum_{k=1}^{t_r}(f(x^*) - f(X_{k,\ r}))$ 表示第 $r$ 轮的总遗憾。那么，我们有 $\mathcal{R}(T) = \sum_{r=1}^{R}\mathcal{R}_r$。对第 $r$ 轮的遗憾求和，可得如下结果。

当 $\mathcal{X}$ 为有限时（定理6.1的设定）：

将（25）中的瞬时遗憾界替换为 $r \geq 2$,

$$\begin{aligned} \mathcal{R}_r &\leq 4t_r\tilde{\beta}(\delta)\sqrt{\frac{c_1\Gamma_k(q_{r-1})}{q_{r-1}}} \\ &\leq 4\lceil\sqrt{Tq_{r-1}} + u_T(\delta)\rceil\tilde{\beta}(\delta)\sqrt{\frac{c_1\Gamma_k(q_{r-1})}{q_{r-1}}} \\ &\leq 4\tilde{\beta}(\delta)\sqrt{c_1T\Gamma_k(q_{r-1})} + (u_T(\delta)+1)\tilde{\beta}(\delta)\sqrt{\frac{c_1\Gamma_k(q_{r-1})}{q_{r-1}}}. \end{aligned}$$

第二个不等式是通过选择 $t_r$得到的。

当 $\mathcal{X}$ 为连续变量时（定理6.2的设定）：

将（26）中的瞬时遗憾上界替换为 $r \geq 2$,

$$\begin{aligned}
\mathcal{R}_r &\leq t_r \left( \frac{8}{q_{r-1}} + 4\tilde{\beta}'_{r-1}(\delta) \left( \sqrt{\frac{c_1 \Gamma_k(q_{r-1})}{q_{r-1}}} + \frac{2}{\sqrt{q_{r-1}}} \right) \right) \\
&\leq \lceil \sqrt{T q_{r-1}} + u_T(\delta) \rceil \left( \frac{8}{q_{r-1}} + 4\tilde{\beta}'_{r-1}(\delta) \left( \sqrt{\frac{c_1 \Gamma_k(q_{r-1})}{q_{r-1}}} + \frac{2}{\sqrt{q_{r-1}}} \right) \right) \\
&\leq 4\tilde{\beta}'_{r-1}(\delta) \sqrt{c_1 T \Gamma_k(q_{r-1})} + 8\sqrt{\frac{T}{q_{r-1}}} + 2\sqrt{T} \\
&\quad + (u_T(\delta) + 1) \left( \frac{8}{q_{r-1}} + 4\tilde{\beta}'_{r-1}(\delta) \left( \sqrt{\frac{c_1 \Gamma_k(q_{r-1})}{q_{r-1}}} + \frac{2}{\sqrt{q_{r-1}}} \right) \right).
\end{aligned}$$

我们还注意到 $f(x) = \langle f, k(\cdot, x) \rangle$（再生性质）。因此 $|f(x)| \leq \|f\|_{\mathcal{H}_k} \|k(\cdot, x)\|$；也就是说，$|f(x)| \leq C_k k_{\max}$。因此，第一轮的遗憾可以直接用其长度来界定，

$$\begin{aligned}
\mathcal{R}_1 &\leq t_1 C_k k_{\max} \\
&\leq \lceil \sqrt{T} + u_T(\delta) \rceil C_k k_{\max}.
\end{aligned}$$

回忆一下 $u_T(\delta) = \mathcal{O}(\mathbb{E}[\tau] + \log(\frac{T}{\delta}))$，并注意，对于 $r > 1$,

$$\frac{8}{q_{r-1}} + 4\tilde{\beta}'_{r-1}(\delta) \left( \sqrt{\frac{c_1 \Gamma_k(q_{r-1})}{q_{r-1}}} + \frac{2}{\sqrt{q_{r-1}}} \right) = \mathcal{O}(1), \tag{27}$$

因为 $q_{r-1} \geq \sqrt{T}$ for $r > 1$。

将所有轮次的遗憾相加，$r = 1, 2, \ldots\ldots, R$，其中 $R = \log\log(T)$，我们得到

$$\mathcal{R}(T) = \mathcal{O}\left( \tilde{\beta}(\delta) \log\log(T) \sqrt{T \Gamma_k(T)} + C_k \mathbb{E}[\tau] \right), \tag{28}$$

and,

$$\mathcal{R}(T) = \mathcal{O}\left( \tilde{\beta}'_R(\delta) \log\log(T) \sqrt{T \Gamma_k(T)} + C_k \mathbb{E}[\tau] \right), \tag{29}$$

在离散和连续 $\mathcal{X}$ 情形下，分别证明定理 6.1 和 6.2。

推论 6.3 可通过分别在 Matérn 核、平方指数核和线性核的情形下替换 $\Gamma_k$ 的取值直接得出；具体而言，$\Gamma_k(T) = \mathcal{O}(T^{\frac{d}{2\nu+d}} \log(T))$、$\Gamma_k(T) = \mathcal{O}((\log(T))^{\frac{2\nu}{2\nu+d}})$、$\Gamma_k(T) = \mathcal{O}((\log(T))^{d+1})$ 和 $\Gamma_k(T) = \mathcal{O}(d \log(T))$（Srinivas 等人，2010；Vakili 等人，2021c）。在线性核的情形下，我们特别强调 $\tilde{\beta}'_R(\delta)$ 中的 $\mathcal{O}(\sqrt{d})$ 因子。