arXiv:2601.02870v1 [cs.LG] 6 Jan 2026

量子增强神经上下文老虎机算法

Yuqi Huang[1*]，Vincent Y. F Tan[1,2] 和 Sharu Theresa Jose[3]1*新加坡国立大学数学系，新加坡，新加坡 119077。[2]新加坡国立大学电气与计算机工程系，新加坡，新加坡 117583。[3]伯明翰大学计算机科学学院，伯明翰，B15 2TT，英国。

通讯作者。电子邮箱：e0727232@u.nus.edu；贡献作者：vtan@nus.edu.sg；s.t.jose@bham.ac.uk；

摘要

随机上下文多臂老虎机是序贯决策的基础，但现有基于神经网络的算法面临重大挑战；当扩展至量子神经网络（QNN）时，过度参数化、计算不稳定性和贫瘠高原现象等问题尤为突出。本文提出了 *Quantum Neural Tangent Kernel-Upper Confidence Bound* (QNTK-UCB) 算法，这是一种利用量子神经切线核（QNTK）解决上述局限性的新算法。

通过将量子神经网络固定在随机初始化状态，并利用其静态 QNTK 作为岭回归的核，QNTK-UCB 避开了显式参数化量子电路训练中固有的不稳定训练动力学，同时充分利用了独特的量子归纳偏置。对于时域 $\boldsymbol{T}$ 和 $\boldsymbol{K}$ 个动作，我们的理论分析表明，QNTK-UCB 所需的参数规模显著降低至$\boldsymbol{\Omega((TK)^3)}$，相比之下，经典 NeuralUCB 算法要达到类似的遗憾保证则需要$\boldsymbol{\Omega((TK)^8)}$ 。在非线性合成基准和量子原生的变分量子本征求解器任务上的实证评估表明，QNTK-UCB 在低数据量场景中具有更高的样本效率。本研究凸显了 QNTK 的固有属性如何提供隐式正则化并带来更快的谱衰减，为在在线学习中实现“量子优势”铺平了道路。

关键词： 量子神经切线核（QNTK）、量子神经网络（QNNs）、上置信界（UCB）算法、样本效率

1 引言

随机上下文多臂老虎机（SCBs）因其在临床试验等领域的广泛应用，在过去几十年中得到了深入研究。 [1, 2], 电子商务推荐系统 [3, 4], 在线广告和个性化媒体推送 [5]。SCBs 为在线序贯决策提供了一个典型框架，在该框架中，学习者会根据观察到的上下文信息反复选择动作。在每个时间步，每个动作（例如试验药物）都具有取决于环境辅助信息（例如患者的年龄、性别和肿瘤生物标志物）的上下文特征。学习者观察这些依赖于上下文的特征，选择一个动作，并获得随机奖励（例如观察到的治疗结果）。学习者的目标是在有限时间范围内设计一种决策策略，使预期累积奖励最大化， T或者等价地，使相对于基准策略的累积遗憾最小化。实现这一目标需要有效地平衡对其他动作的探索与对当前表现最佳动作的利用。

在经典文献中，SCB通常借助线性奖励模型进行分析，其中未知的平均奖励函数被假定为上下文特征向量的线性函数。这促成了若干强大算法的发展，例如线性上下文UCB [6, 7]和线性上下文汤普森采样 [8, 9]。尽管线性奖励模型在理论上十分便利，但它们往往无法捕捉实际场景中高度非线性的依赖关系。这一局限促使人们探索若干非线性模型，包括广义线性模型 [10,11]、核模型 [12], 以及高斯过程 [13], ，其中奖励函数被假定位于再生核希尔伯特空间（再生核希尔伯特空间）之内。尽管这些方法十分强大，其有效性在很大程度上取决于其归纳偏置与真实底层奖励函数是否契合。

为应对这一挑战，经典神经网络（神经网络）最近被引入老虎机场景 [14–17]，借助其巨大的表征能力来逼近复杂的平均奖励函数。由此得到的神经上下文赌博机算法通常在“神经切线核”（神经切核）机制下运行，其中神经网络被充分过参数化，从而使其训练动力学线性化。尽管在经验上取得了成功，经典神经赌博机仍面临重大挑战：它们需要大规模的过度参数化（参数数量按 $\Omega((TK)^8)$ 缩放，其中 K 表示动作的数量）才能满足收敛性保证，并且由于需要频繁对动态设计矩阵求逆，其在在线场景中的计算成本仍然是关键瓶颈。

近年来，量子神经网络（QNN），又称参数化量子电路（PQC），已成为一种强大的机器学习范式。借助叠加与纠缠原理，量子神经网络能够以可比的参数数量，在表示能力上优于经典神经网络 [18–20]。一方面，量子神经网络可通过复杂的量子特征映射，将经典上下文数据嵌入维度呈指数级增长的特征希尔伯特空间，从而形成独特的“量子归纳偏置”，或能更高效地表示奖励函数 [21]。另一方面，近期研究表明，对于源自本质上属于量子物理的

过程——例如寻找哈密顿量的基态或对量子相进行分类——经典模型可能效率低下 [22, 23]。

受这些优势的启发，本文提出了一类新型量子增强神经上下文赌博机算法，利用基于量子神经网络的奖励模型。然而，从经典神经赌博机转向量子神经赌博机面临着重大的技术挑战。首先，深度量子神经网络饱受“贫瘠高原”现象困扰 [24], ，网络梯度会随着量子比特数量的增加而呈指数级消失 m。因此，要准确估计梯度，就需要进行指数数量的测量，这不仅抵消了潜在的量子优势，还会导致梯度下降训练不稳定。其次，基于量子神经网络的赌博机需要在线训练，这要求计算动态梯度特征映射，而这些映射会随着每次迭代中权重的更新而变化。这迫使系统在每一步都重新计算并求逆设计矩阵，给 NISQ 硬件带来难以承受的计算成本。

克服贫瘠高原的一种可行方法是将架构限制为浅层量子神经网络，其层数至多随量子比特数量呈对数增长 [25]。有趣的是，在这类不存在贫瘠高原的区域，近期研究表明，量子神经网络的训练动力学由一个固定的解析核所支配，该核由初始化时的电路架构决定 [26, 27], ，称为量子神经切线核（QNTK）。这将经典 NTK 理论扩展到了过参数化量子神经网络。重要的是，QNTK 框架适用于深度可随量子比特数量增长的架构，因此能够使用经典模拟困难的电路，从而有望实现量子优势（这类电路的示例见 [27, 第 2 节5] ）。

受这些结果以及在老虎机场景中显式训练深度量子神经网络所面临的挑战启发，我们提出一种核化方法，将 QNTK 用作岭回归的静态核。这一策略使我们能够避开变分电路的非凸优化景观，同时保留量子特征映射固有的独特归纳偏置。至关重要的是，我们证明，与经典特征空间相比，量子特征空间能够实现更高效的线性化。

我们的主要贡献如下：

- 我们提出 QNTK-UCB，这是首个利用经验量子神经切线核进行奖励估计的上下文老虎机算法。该框架使学习者能够利用量子表达能力，同时避免显式参数化量子电路训练带来的不稳定性。
- 我们基于量子有效维数，对 QNTK-UCB 进行了全面的遗憾分析。我们工作的一个关键发现是，QNTK-UCB 只需显著更低的参数规模 $\tilde{\Omega}((TK)^3)$，即可达到经典 NeuralUCB 使用 $\tilde{\Omega}((TK)^8)$ 个参数时才能达到的相同遗憾界 [16]。
- 通过在非线性合成基准和变分量子本征求解器（VQE）任务的量子初始态推荐上开展一系列实验，我们证明，QNTK-UCB 在低数据量场景下展现出更高的样本效率，为在线学习实现“量子优势”指明了清晰的路径。

相关工作：我们的研究处于量子机器学习与在线决策的交汇处，区别于量子老虎机领域中若干成熟的研究方向。

大量研究 [28–31] 提出了用于经典多臂老虎机和上下文老虎机的量子算法。这些研究通常假设存在量子奖励预言机，并借助量子蒙特卡洛或振幅放大等量子算法，实现查询复杂度上的（二次）加速。然而，这些方法的实用性往往受“输入瓶颈”现象制约，即将经典奖励数据编码到量子预言机中的计算成本可能高得令人难以承受，从而抵消任何算法加速。相比之下，我们的研究不假设存在量子预言机，而是利用量子电路作为经典数据的函数近似工具。

另一类研究 [32, 33] 将量子态属性的学习（例如影子层析或哈密顿量估计）建模为随机量子老虎机问题。然而，这类问题通常属于经典线性上下文框架。虽然 QNTK-UCB 能够处理此类任务，但它旨在作为一种通用学习者，适用于经典和量子原生奖励函数。

2 背景设定与量子神经网络

2.1 问题设定：上下文老虎机

我们考虑一个具有有限时域$T \in \mathbb{N}$的随机 K臂上下文老虎机问题。在每一轮 $t \in [T]$：$= \{1, \ldots, T\}$，智能体会观察到一组上下文向量$\mathcal{X}_t = \{\mathbf{x}_{t,a} : a \in [K]\}$，其中$\mathbf{x}_{t,a} \in \mathcal{X}$：$= \cup_{t\in\mathbb{N}} \mathcal{X}_t \subset \mathbb{R}^d$ 表示与臂 a相关联的 d维特征向量。智能体选择一个臂 $a_t \in [K]$，并观察到一个含噪标量奖励r_{t,a_t}。我们假设奖励的生成方式如下

$$r_{t,a_t} = h(\mathbf{x}_{t,a_t}) + \xi_t, \tag{1}$$

其中，h：$\mathbb{R}^d \to [0,1]$ 是一个未知的，$[0,1]$有界的平均奖励函数，而 ξ_t 是ν次高斯噪声，给定历史记录 $\mathcal{H}_{t-1} = \{(\mathbf{x}_{s,a_s}, r_{s,a_s})\}_{s=1}^{t-1}$（截至并包括时间 $t-1$），即它满足 $\mathbb{E}[\xi_t|\mathcal{H}_{t-1}] = 0$ 以及 $\mathbb{E}[\exp(s\xi_t)|\mathcal{H}_{t-1}] \le \exp(\nu^2 s^2/2)$，对所有 $s \in \mathbb{R}$ 均成立。我们注意到，假设平均奖励函数 $h(\cdot)$ 有界是老虎机文献中的标准假设，并且在上下文和模型类满足标准有界性假设时也成立（例如，线性模型中的上下文/参数有界，以及核模型中的再生核希尔伯特空间范数有界）。为进行核分析，我们将整个时域内的所有上下文集合记为一个“向量化”数据集 $\mathcal{X}_{1:TK} := \{\mathbf{x}_{t,a}\}_{t\in[T],a\in[K]}$，其索引可记为 $\{\mathbf{x}^i\}_{i=1}^{TK}$。

令 $a_t^* \in \arg\max_{a\in[K]} h(\mathbf{x}_{t,\,a})$ 表示在第 t 轮次使平均奖励最大化的任意最优臂。智能体的目标是最小化期望累积遗憾，

$$\bar{R}_T := \sum_{t=1}^{T} \mathbb{E}\left[h(\mathbf{x}_{t,a_t^*}) - h(\mathbf{x}_{t,a_t})\right], \tag{2}$$

定义为在时域 T内累积的最优期望奖励与所选臂的期望奖励之间的累积差值。

2.2 量子神经网络

本节将说明所考虑的量子神经网络的结构。具体而言，我们采用 [27]，中考虑的一般量子神经网络框架，该框架保证当量子比特数量趋于无限时，量子神经网络所描述的函数会收敛（依分布）到一个高斯过程。

量子神经网络作用于由 m 个量子比特组成的系统，电路深度为 $L \in \mathbb{N}$。具体而言，我们允许层数 $L(m)$ 随 m 变化。作用于初始态 $|0\rangle^{\otimes m}$ 的总幺正操作 $U(\boldsymbol{\theta}, x)$ 由一系列 L 层组成：

$$U(\boldsymbol{\theta}, \mathbf{x}) = U_L(\boldsymbol{\theta}_L, \mathbf{x}) \dots U_1(\boldsymbol{\theta}_1, \mathbf{x}). \tag{3}$$

每一层 $l \in [L]$ 都由一个参数化模块和一个固定模块构成的幺正算子表示：

$$U_l(\boldsymbol{\theta}_l, \mathbf{x}) = W_l(\boldsymbol{\theta}_l) V_l(\mathbf{x}),$$

其中，$W_l(\boldsymbol{\theta}_l)$ 包含作用于每个量子比特的可训练单量子比特旋转（由 $\boldsymbol{\theta}_l$ 参数化），而 $V_l(\mathbf{x})$由固定的纠缠门（例如受控非门）以及可选的数据编码门组成。对于层数固定为 L（即 L 不随 m 变化）的量子神经网络，容易看出 $\boldsymbol{\theta} \in \mathbb{R}^p$，参数总数为 $p \approx Lm$。

上述量子电路定义了奖励模型 $f(\mathbf{x}; \boldsymbol{\theta})$，具体如下：总幺正操作 $U(\boldsymbol{\theta}, \mathbf{x})$ 作用于初始量子态 $|0^m\rangle$，生成输出量子态 $|\psi(\boldsymbol{\theta}, \mathbf{x})\rangle = U(\boldsymbol{\theta}, \mathbf{x})|0^m\rangle$。随后使用可观测量 $\mathcal{O}$ 对该输出态进行测量。遵循 [27]，中概述的框架，我们将 $\mathcal{O}$ 定义为局域单量子比特可观测量之和，表示为：

$$\mathcal{O} = \sum_{k=1}^{m} \mathcal{O}_k, \quad \text{where } \mathrm{Tr}(\mathcal{O}_k) = 0, \text{ for } k = 1, \dots, m.$$

此外，每个可观测量 $\mathcal{O}_k$ 的特征谱都局限于集合 $\{-1, +1\}$，这意味着 $\mathcal{O}_k$ 的每个特征值只能是 $+1$ 或 -1。随后，将输出模型$f(\mathbf{x}; \boldsymbol{\theta})$定义为全局可观测量 $\mathcal{O}$ 相对于输出态 $|\psi(\boldsymbol{\theta}, \mathbf{x})\rangle$ 的期望值，并除以归一化常数 $N(m)$，即

$$f(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{N(m)} \sum_{k=1}^{m} f_k(\mathbf{x}; \boldsymbol{\theta}), \quad \text{where} \quad f_k(\mathbf{x}; \boldsymbol{\theta}) = \langle 0^m | U^\dagger(\boldsymbol{\theta}, \mathbf{x}) \mathcal{O}_k U(\boldsymbol{\theta}, \mathbf{x}) | 0^m \rangle. \tag{4}$$

这里，$N(m)$ 是一个归一化因子，由量子神经网络模型在初始化时的协方差函数决定，也就是说，在训练开始前，参数 $\boldsymbol{\theta}$ 是随机选取的。模型定义中包含这一归一化，以确保当 $m \to \infty$时，$f(\mathbf{x}; \boldsymbol{\theta})$ 收敛到一个非平凡的高斯过程。具体而言，我们作出以下假设 [27]:

假设1 参数 $\boldsymbol{\theta}$ 的分布、量子电路的架构以及所选择的归一化 $N(m)$满足

$$\mathbb{E}[f_k(\mathbf{x}; \boldsymbol{\theta})] = 0 \quad \forall \mathbf{x} \in \mathcal{X},$$

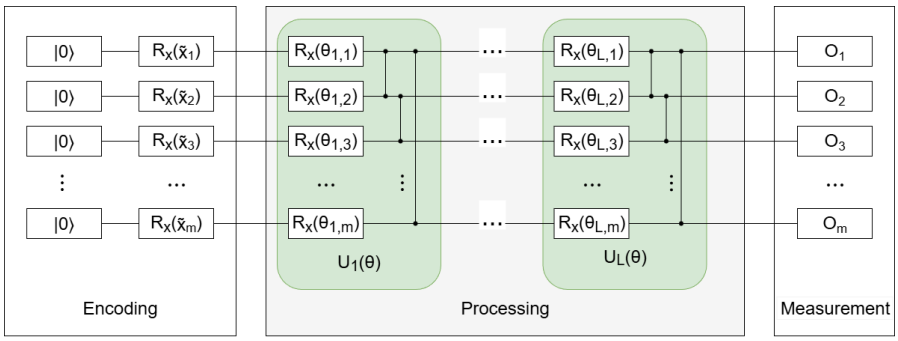


图1：电路结构示例

and

$$\lim_{m\to\infty}\sup_{\mathbf{x},\mathbf{x}'\in\mathcal{X}}\left|\mathbb{E}[f_k(\mathbf{x};\boldsymbol{\theta})f_k(\mathbf{x}';\boldsymbol{\theta})]-\mathcal{K}(\mathbf{x},\mathbf{x}')\right|=0,$$

其中 $\mathcal{K}$： $\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ 是一个从特征空间映射到实数的任意二元函数，其对角元素严格为正，即对于所有$\mathbf{x}$、$\mathbf{x}>0\,\mathbf{x}\in\mathcal{X}$。

根据上述假设， $N(m)$ 的选取使得极限 $m\to\infty$ 产生有限且非平凡的协方差函数。请注意， $N(m)$ 的值取决于具体的量子神经网络架构。例如，图1 中的量子神经网络架构满足 $N(m)=\sqrt{m}$ [26]。

在本文中，我们通过函数 $f(\mathbf{x};\boldsymbol{\theta})$ 对未知奖励函数 $h(\mathbf{x})$ 建模，该函数由满足假设1的量子神经网络架构定义。我们将模型函数关于参数向量的梯度记为$\mathbf{g}(\mathbf{x};\boldsymbol{\theta})=\nabla_{\boldsymbol{\theta}}f(\mathbf{x};\boldsymbol{\theta})$。

2.3 量子神经切核理论

经典神经网络老虎机算法（如 NeuralUCB）的理论分析建立在神经切线核（NTK）[16, 34] 理论之上。该理论认为，在无限宽神经网络的“惰性训练”机制下，网络权重始终接近其初始化值，优化动力学可近似为网络梯度的线性模型。在这种机制下，神经网络实际上作为高维特征映射的线性模型运行： $\boldsymbol{\phi}(\mathbf{x})=\mathbf{g}(\mathbf{x};\boldsymbol{\theta}_0)$，其中该映射由网络在初始化时的梯度 $\boldsymbol{\theta}_0$ 定义。

NTK 理论最近已扩展至量子神经网络模型 [26, 27]。具体而言，Girardi 和 De Palma [27] 表明，在某些假设下，当 $m\to\infty$ 时，量子神经网络函数收敛至高斯过程，其动态由量子神经切线核（QNTK）决定。我们将随机初始化 $\boldsymbol{\theta}_0$ 处的经验量子神经切线核定义为：

$$\hat{\mathbf{K}}_{\boldsymbol{\theta}_0}(\mathbf{x},\mathbf{x}')=\frac{1}{N_K(m)}\mathbf{g}(\mathbf{x};\boldsymbol{\theta}_0)^\top\mathbf{g}(\mathbf{x}';\boldsymbol{\theta}_0)=\frac{1}{N_K(m)}\langle\nabla_{\boldsymbol{\theta}}f(\mathbf{x};\boldsymbol{\theta}_0),\nabla_{\boldsymbol{\theta}}f(\mathbf{x}';\boldsymbol{\theta}_0)\rangle,\quad(5)$$

其中 $\mathbf{g}(\mathbf{x};\ \boldsymbol{\theta}_0) = \nabla_{\boldsymbol{\theta}} f(\mathbf{x};\ \boldsymbol{\theta}_0)$ 是量子神经网络模型（4）在初始化时的梯度，而 $N_K(m)$ 是一个依赖于宽度的归一化因子，其选取旨在确保当 $m \to \infty$。

与之相关的解析量子神经正切核是网络参数随机初始化下经验核的期望，即：

$$\mathbf{K}^{(m)}(\mathbf{x}, \mathbf{x}') = \mathbb{E}_{\boldsymbol{\theta}_0}[\hat{\mathbf{K}}_{\boldsymbol{\theta}_0}(\mathbf{x}, \mathbf{x}')], \tag{6}$$

其中，随机向量 $\boldsymbol{\theta}_0$ 的各元素相互独立，且是在 $[0, \pi]$ 中均匀分布的随机变量。注意，解析 QNTK $\mathbf{K}^{(m)}(\mathbf{x},\ \mathbf{x}')$ 取决于量子神经网络架构。其谱性质以及由此产生的学习归纳偏置，受量子比特数量 m、电路深度 L、纠缠门的连接方式以及数据编码策略 $V(\mathbf{x})$ 的影响。为简洁起见，在下文中我们省略 $\mathbf{K}^{(m)}(\mathbf{x},\ \mathbf{x}')$ 对 m 的依赖，并将其记为 $\mathbf{K}(\mathbf{x},\ \mathbf{x}')$。同样，当上下文已足以说明时，我们省略$\mathbf{K}_{\boldsymbol{\theta}_0}$ 对 $\boldsymbol{\theta}_0$ 的依赖，并将其记为 $\mathbf{K}$。此外，我们作出以下假设 [27]:

假设 2 存在一种 $N_K(m)$ 的取值，使得存在函数 $\bar{\mathbf{K}}(\mathbf{x},\ \mathbf{x}')$，满足

$$\lim_{m\to\infty} \sup_{\mathbf{x},\mathbf{x}'\in\mathcal{X}} |\mathbf{K}(\mathbf{x}, \mathbf{x}') - \bar{\mathbf{K}}(\mathbf{x}, \mathbf{x}')| = 0. \tag{7}$$

换言之，当 $m \to \infty$ 时，解析量子神经正切核会收敛到极限量子神经正切核；许多具有实际应用价值的量子电路架构都满足这一假设 [27]。

此外，我们的量子神经上下文老虎机算法依赖于经验量子神经切线核的以下关键性质：当量子比特数量 m 增大时，经验量子神经切线核 $\mathbf{K}_{\boldsymbol{\theta}_0}$ 会收敛到解析均值 $\mathbf{K}$。如果量子神经网络架构满足某些结构条件，就能保证这种收敛。影响收敛的量子神经网络关键性质包括过去光锥 $\mathcal{N}_k$，其定义为能够影响局部可观测量 f_k 的参数 $\{\boldsymbol{\theta}_i\}$ 集合，其中 $|\mathcal{N}| = \max_k |\mathcal{N}_k|$；以及未来光锥 $\mathcal{M}_i$，即参数 $\boldsymbol{\theta}_i$ 能够影响的可观测量 $\{f_k\}$ 集合，其中 $|\mathcal{M}| = \max_i |\mathcal{M}_i|$。请注意， L、 $|\mathcal{M}|$ 和 $|\mathcal{N}|$ 可能取决于宽度 m。形式化地说，该架构的结构条件以及经验量子神经切线核的收敛性质可表述如下：

假设3 具有 m 个量子比特的量子神经网络架构满足以下条件：

$$\lim_{m\to\infty} \frac{Lm|\mathcal{M}|^4|\mathcal{N}|^2}{N(m)^4} = 0, \tag{8}$$

定理 1 (定理 4.13 [27]) *Under Assumption 3 when the QNN is randomly initialized, i.e., the parameters $\boldsymbol{\theta}$ are independent random variables, the empirical QNTK converges in probability to the analytic QNTK as $m \to \infty$. In particular, there exists a constant $c > 0$ such that, for any $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$, we have*

$$\mathbb{P}\left(\left|\hat{\mathbf{K}}_{\boldsymbol{\theta}_0}(\mathbf{x}, \mathbf{x}') - \mathbf{K}(\mathbf{x}, \mathbf{x}')\right| > \varepsilon\right) \leq \exp\left(-c\varepsilon^2 \frac{N_K(m)^2 N(m)^4}{Lm|\mathcal{M}|^4|\mathcal{N}|^2}\right). \tag{9}$$

请注意，$|\mathcal{N}|$ 和 $|\mathcal{M}|$ 相对于宽度 m 和深度 L的缩放取决于量子试探态的具体连接结构。例如，几何局域电路[1] 使光锥的增长主要受限于电路深度，而全连接架构则允许光锥迅速扩展，覆盖整个系统规模 m。此外，可以验证，对于满足定理 1 和假设 2的电路，$\Omega(1) \leq N_K(m) < O(|\mathcal{N}|)$（参见 [27]中的引理 4.9）。

在全文中，我们考虑满足（8）的量子神经网络电路架构。具体而言，以下量子神经网络架构均满足这一假设：

- 常深度电路（$L = O(1)$）：对于深度固定的几何局域电路 L，光锥大小 $|\mathcal{M}|$ 和 $|\mathcal{N}|$ 均为 $O(1)$。假设归一化$N(m) = \Omega(\sqrt{m})$，则（8）中的收敛条件得到满足。

- 对数深度电路（$L = O(\log m)$）。正如 [27], 所确立的，使电路深度随量子比特数量增长，是实现量子优势的必要条件。在归一化为 $N(m) = \Omega(\sqrt{m})$ 的情况下，可以直接验证，对数深度架构满足我们的理论假设。 [27]的第 2.5 节提供了符合这些标准的电路构造的具体示例。

- 多项式深度电路（$L = O(m^\alpha)$，其中某个较小的 $\alpha > 0$）。在我们的无训练机制中，可以进一步探索更大的（例如多项式）深度缩放，以增强表达能力，并提高实现更显著量子优势的潜力。然而，深度增加也会带来额外的权衡，包括对光锥增长、量子神经切线核集中性和核缩放的担忧。第 3.3 节将详细讨论这些影响。

3 基于量子神经切线核的上置信界算法

本节介绍基于 QNTK 的上下文老虎机新算法——量子神经切线核（QNTK）-上置信界（UCB）。尽管量子神经网络具有通用函数逼近能力 [35, 36]，但由于“贫瘠高原”现象 [24], ，通过梯度下降训练量子神经网络一直极其困难：随着量子比特数量增加，网络梯度会呈指数级消失。因此，直接训练量子神经网络模型$f(\mathbf{x}; \boldsymbol{\theta})$来逼近未知奖励函数 $h(\mathbf{x})$，在计算上代价高昂且不稳定。为解决这一问题，在我们提出的算法中，我们将量子神经网络固定在随机初始化状态，并利用其对应的量子神经切线核（QNTK）进行回归。

3.1 算法描述

我们提出的算法 *QNTK-UCB*是一种核化上置信界策略，其中的核是由随机初始化的量子神经网络诱导的经验量子神经切线核（定义见（5））。其核心优势在于无需对量子神经网络进行基于梯度的训练，从而避开贫瘠高原问题带来的优化难题，同时保留量子神经网络固有的表达能力和归纳偏置。

[1]仅相邻量子比特之间发生相互作用的电路 [26]。

这种核化方法的有效性依托于过参数化网络中观察到的“惰性训练”现象。为此，我们首先证明（见引理 A.4），对于足够宽的量子神经网络，奖励函数 $h(\mathbf{x})$ 可以在切线特征空间中表示为线性函数。具体而言，存在一个参数向量 $\boldsymbol{\theta}^*$，使得对于所有上下文 $\mathbf{x}^i$，奖励都可以通过在初始化 $\boldsymbol{\theta}_0$ 附近进行一阶泰勒展开来近似表示：

$$h(\mathbf{x}^i) = f(\mathbf{x}^i; \boldsymbol{\theta}_0) + \langle \nabla_{\boldsymbol{\theta}} f(\mathbf{x}^i; \boldsymbol{\theta}_0), \boldsymbol{\theta}^* - \boldsymbol{\theta}_0 \rangle. \tag{10}$$

注意，$\boldsymbol{\theta}_0$ 是从随机初始化分布中抽取的参数向量（例如，$\boldsymbol{\theta}_0$ 中的每个元素都服从区间 $[0, \pi]$ 上的均匀分布）。

我们将随机初始化的 QNN 视为静态特征提取器，并将 p 维特征映射 $\boldsymbol{\phi}$: $\mathcal{X} \to \mathbb{R}^p$ 定义为在 $\boldsymbol{\theta}_0$ 处计算的量子神经网络模型的缩放梯度：

$$\boldsymbol{\phi}(\mathbf{x}) := \frac{1}{\sqrt{N_K(m)}} \nabla_{\boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{11}$$

这种构造确保了特征空间中的内积能够恢复经验量子神经切线核，即 $\boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\phi}(\mathbf{x}') = \hat{\mathbf{K}}_{\boldsymbol{\theta}_0}(\mathbf{x}, \mathbf{x}')$。

随后，我们的算法作为显式特征空间上的核化老虎机 [12] 的一个实例运行。在每一轮 t，智能体都会维护一个正则化设计矩阵$\mathbf{Z}_{t-1}$和一个奖励加权特征向量 $\mathbf{b}_{t-1}$，其定义如下：

$$\mathbf{Z}_{t-1} = \lambda \mathbf{I} + \sum_{\tau=1}^{t-1} \boldsymbol{\phi}(\mathbf{x}_{\tau,a_\tau}) \boldsymbol{\phi}(\mathbf{x}_{\tau,a_\tau})^\top \quad \text{and} \quad \mathbf{b}_{t-1} = \sum_{\tau=1}^{t-1} r_{\tau,a_\tau} \boldsymbol{\phi}(\mathbf{x}_{\tau,a_\tau}),$$

其中，$\lambda > 0$ 是正则化参数。未知的线性参数 $\boldsymbol{\theta}^*$ 通过岭回归估计如下：$\hat{\boldsymbol{\theta}}_{t-1} = \mathbf{Z}_{t-1}^{-1} \mathbf{b}_{t-1} + \boldsymbol{\theta}_0$。为平衡探索与利用，智能体选择使某个上置信界达到最大值的臂：

$$a_t = \underset{a \in [K]}{\operatorname{argmax}} \left\{ \boldsymbol{\phi}(\mathbf{x}_{t,a})^\top (\hat{\boldsymbol{\theta}}_{t-1} - \boldsymbol{\theta}_0) + \beta_{t-1} \sqrt{\boldsymbol{\phi}(\mathbf{x}_{t,a})^\top \mathbf{Z}_{t-1}^{-1} \boldsymbol{\phi}(\mathbf{x}_{t,a})} \right\}.$$

其中，β_{t-1} 是控制置信区间宽度的探索半径。其精确值由第3.2节中的遗憾分析推导得出。完整流程总结于算法1中。

与现有算法的比较。

QNTK-UCB 算法可以被视为神经网络老虎机算法 [16], 的量子对应算法，但它在操作和理论特性方面具有明显差异：

- 经典 NeuralUCB 算法 [16] 会定期使用梯度下降训练经典神经网络。尽管这种方法有效，但由于贫瘠高原现象，直接将其应用于量子电路会遇到问题：随着系统规模增大，梯度会呈指数级消失，使训练变得不稳定，甚至无法进行。

算法 1 QNTK-UCB 算法

输入： 正则化参数 $\lambda > 0$，探索参数 ν，置信参数 $\delta \in (0, 1)$，范数参数 S，轮次数量 T。

初始化：

随机初始化 QNN 参数 $\boldsymbol{\theta}_0$。

定义特征映射 $\boldsymbol{\phi}(\mathbf{x}) = \frac{1}{\sqrt{N_K(m)}} \nabla_{\boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta}_0)$。

初始化 $\mathbf{Z}_0 = \lambda \mathbf{I}_p$（其中 $p = \dim(\boldsymbol{\theta}_0)$）和 $\mathbf{b}_0 = \mathbf{0} \in \mathbb{R}^p$。

对于 $t = 1, 2, \ldots, T$ **do**

观察上下文 $\{\mathbf{x}_{t,a}\}_{a=1}^{K}$。

计算岭回归估计：$\hat{\boldsymbol{\theta}}_{t-1} = \mathbf{Z}_{t-1}^{-1} \mathbf{b}_{t-1} + \boldsymbol{\theta}_0$。

计算探索半径 $\beta_{t-1} = \nu \sqrt{\log \frac{\det(\mathbf{Z}_{t-1}) \det(\lambda \mathbf{I}) + 2 \log\left(\frac{1}{\delta}\right) + \sqrt{\lambda} S}{}}$。

对于每个臂 $a \in [K]$ **执行**

计算特征：$\boldsymbol{\phi}_{t,a} = \boldsymbol{\phi}(\mathbf{x}_{t,a})$。

计算预测奖励：$\hat{h}_{t-1}(\mathbf{x}_{t,a}) = \boldsymbol{\phi}_{t,a}^{\top} (\hat{\boldsymbol{\theta}}_{t-1} - \boldsymbol{\theta}_0)$。

计算置信界的宽度：$w_{t,a} = \beta_{t-1} \sqrt{\boldsymbol{\phi}_{t,a}^{\top} \mathbf{Z}_{t-1}^{-1} \boldsymbol{\phi}_{t,a}}$。

计算上置信界：$U_{t,a} = \hat{h}_{t-1}(\mathbf{x}_{t,a}) + w_{t,a}$。

结束循环

选择动作：$a_t = \text{取最大值}_{a \in [K]} U_{t,a}$。

观察奖励 $r_t = r_{t,a_t}$。

更新 $\mathbf{Z}_t = \mathbf{Z}_{t-1} + \boldsymbol{\phi}_{t,a_t} \boldsymbol{\phi}_{t,a_t}^{\top}$。

更新 $\mathbf{b}_t = \mathbf{b}_{t-1} + r_t \boldsymbol{\phi}_{t,a_t}$。

结束循环

QNTK-UCB 通过利用初始化时量子特征空间的固定几何结构彻底规避了这一问题。我们的方法将量子电路视为静态核，而非可训练模型，因此在老虎机交互过程中无需进行梯度更新，既能确保算法稳定性，又能保留量子试探态的表达能力。我们在第3.3节讨论这种固定几何结构对参数效率的影响。

- 尽管我们的算法与 KernelUCB 具有相似的代数结构，[37],但关键区别在于所采用的核。标准经典核（如径向基函数和 Mat´ern 核）往往难以准确建模量子奖励函数。相比之下，QNTK 显式地纳入了量子电路的归纳偏置，并考虑了纠缠结构和数据编码方法等特征。因此，QNTK-UCB 能够在“量子原生”环境中有效学习，例如分析量子态的性质；在这类场景中，经典核无法捕捉重要的关联。实验结果将进一步清晰地展示这一点。

3.2 遗憾分析

本节将为我们的 QNTK-UCB 算法提供理论保证。我们的分析基于经验量子神经切线核向其极限的集中性，从而能够以量子有效维数为依据界定遗憾。

下文中，我们将核函数 $\mathbf{K}(\mathbf{x},\mathbf{x}')$ 与由其诱导的核矩阵 $\mathbf{K} = [\mathbf{K}_{ij}]$，其中第 (i,j) 个元素 $\mathbf{K}_{ij}$ 是核函数在第 i 个和第 j 个数据点处的取值，即 $\mathbf{K}_{ij} = \mathbf{K}(\mathbf{x}^i,\mathbf{x}^j)$。借助这一记号，我们首先定义量子有效维数。

定义 1 数据集 $\mathcal{X}_{1:TK}$ 上量子神经切线核的 *quantum effective dimension* $d_{\mathrm{q}}(\lambda)$ 定义如下：

$$\widetilde{d}_{\mathrm{q}}(\lambda) = \frac{\log\det(\mathbf{I}_{TK} + \bar{\mathbf{K}}/\lambda)}{\log(1 + TK/\lambda)},$$

其中 $\bar{\mathbf{K}}$ 是假设2中定义的极限量子神经正切核。

直观而言，量子有效维数衡量的是特征空间相对于可用数据的“容量”。当上下文明确时，我们省略 $\widetilde{\cdot}$ 对 $d_{\mathrm{q}}(\lambda)$ 的依赖关系 λ，直接记作 d_{q}。这一定义与经典核老虎机中使用的有效维数 [16] 相对应，直观上衡量了由 QNTK 定义的特征空间的复杂度。

除了假设1和2中关于量子神经网络的结构性假设外，我们还作出以下假设；这些假设在核老虎机文献中较为温和且标准 [16, 38]：

假设 4 上下文向量 $\mathbf{x}_{t,\,a}$ 满足 $\|\mathbf{x}_{t,\,a}\|_2 = 1$ 对所有 $t \in [T]$ 和 $a \in [K]$ 均成立。

事实上，适当归一化上下文向量即可作出这一假设，且不失一般性。

假设 5 极限量子神经切线核矩阵 $\bar{\mathbf{K}}$（在数据集 $\mathcal{X}_{1:TK}$ 上计算）为正定矩阵，其最小特征值为 $\lambda_0 > 0$，即 $\bar{\mathbf{K}} \succeq \lambda_0 \mathbf{I}$。

主要结果

定理 2 *Fix any $\delta \in (0,1)$. Let $m = \Omega\left((TK)^3 \log\left((TK)^2/\delta\right)\right)$. Then, with probability atleast $1-\delta$, the cumulative regret of QNTK-UCB (Algorithm 1) satisfies*

$$\begin{aligned} R_T &:= \sum_{t=1}^{T} \left[h(\mathbf{x}_{t,a_t^*}) - h(\mathbf{x}_{t,a_t})\right] \\ &\le 3\sqrt{T}\sqrt{\widetilde{d}_{\mathrm{q}} \log\left(1 + \frac{TK}{\lambda}\right) + 1}\left(\nu\sqrt{\widetilde{d}_{\mathrm{q}} \log\left(1 + \frac{TK}{\lambda}\right) + 1 + 2\log\left(\frac{1}{\delta}\right)} + \sqrt{\lambda}S\right), \end{aligned}$$

where $S \geq \sqrt{2\mathbf{h}^\top \bar{\mathbf{K}}^{-1}\mathbf{h}}$, $\mathbf{h} = [h(\mathbf{x}^1)\ldots, h(\mathbf{x}^{TK})]^\top$, *and* $\lambda \geq \max\{1, S^{-2}\}$. *Ignoring logarithmic terms and constants, this bound simplifies to*

$$R_T = \tilde{\mathcal{O}}\left(\widetilde{d}_{\mathrm{q}}\sqrt{T}\right)$$

证明概要。

证明（详见附录 A）分三步进行:

1. **集中性与量子线性可实现性**。我们基于近期关于量子电路中高斯过程行为的研究 [27], 以ˆ ¯控制经验 QNTK $\mathbf{K}$ 围绕其极限$\mathbf{K}$的集中性。这里的一项技术创新是利用酉群上的测度集中，将谱距离 $\|\hat{\mathbf{K}} - \bar{\mathbf{K}}\|_{\mathrm{F}}$ 界定为仅取决于电路架构、而与优化动力学无关的函数（引理 A.2）。这一收敛性使我们能够有效地将 QNN“线性化”。具体而言，引理 A.4保证存在一个参数向量 $\boldsymbol{\theta}^* \in \mathbb{R}^p$，使得 $h(\mathbf{x}) = \langle \nabla_{\boldsymbol{\theta}} f(\mathbf{x};\, \boldsymbol{\theta}_0)\ \boldsymbol{\theta}^* - \boldsymbol{\theta}_0 \rangle$ 对所有$\mathbf{x} \in \mathcal{X}_{1:TK}$都成立，且满足约束 $\|\boldsymbol{\theta}^*\|_2 \leq S$。

2. **置信椭球与瞬时遗憾**。 随后，我们为未知参数 $\boldsymbol{\theta}^*$ 构造置信椭球，并利用向量值鞅的自归一化鞅不等式（引理 A.5）。经典 NeuralUCB 的标准证明 [16] 必须界定梯度下降过程中 NTK 的漂移，以确保置信集保持有效（通常需要令人望而却步的宽度缩放）；而我们的分析利用了冻结量子拟设的静态几何结构。由于参数在初始化时即固定，核不会发生漂移，因此我们能够以高概率保证，真实参数位于以岭回归估计为中心的有界区域内。

3. **行列式界与总遗憾**。最后，我们将累积遗憾与预测方差之和联系起来，从而建立其上界；而预测方差之和受核矩阵的对数行列式控制（引理 A.7）。这使我们能够˜在引理中引入量子有效维数 d_{q} A.8，并证明˜遗憾主要随 d_{q}增长。该维数有效刻画了量子拟设独特的归纳偏置，从而在形式上将特定量子架构的谱衰减与算法的学习效率联系起来。

推论 3 *Under the same conditions as Theorem 2: the expected cumulative regret satisfies*

$$\bar{R}_T \leq 3\sqrt{T}\sqrt{\widetilde{d}_{\mathrm{q}}\log\left(1+\frac{TK}{\lambda}\right)+1}\left(\nu\sqrt{\widetilde{d}_{\mathrm{q}}\log\left(1+\frac{TK}{\lambda}\right)+1+2\log T}+\sqrt{\lambda}S\right)+1 = \tilde{\mathcal{O}}\left(\widetilde{d}_{\mathrm{q}}\sqrt{T}\right)$$

Proof 由定理 2 令 $\delta = 1/T$可得。 □

3.3 与经典方法的讨论和比较

定理2 中的遗憾保证为分析 QNTK-UCB 的效用奠定了基础。尽管遗憾界的代数形式与标准核化老虎机相似，但 QNTK 的具体性质在参数效率、归纳偏置和隐式正则化方面带来了独特优势。

遗憾与参数效率

定理2 的一个重要推论涉及实现指定保证所需的模型规模。NeuralUCB 等经典方法 [16], 要求神经网络在“惰性训练”或“线性神经切核”机制下运行，以维持遗憾界的理论有效性。在这一机制下，权重相对于其初始化仅做极小调整，从而确保经验核实际上保持静态。

然而，要达到这一机制，对模型规模的限制十分严格。值得注意的是，网络宽度 w 必须极大，才能尽量减小神经网络与其线性化核之间的逼近误差。NeuralUCB 的分析（参见 [16] 中的引理 5.1 和引理 5.4）表明，宽度必须按时域的高阶多项式增长，具体而言为 $w = \tilde{\Omega}\left((TK)^6\right)$。由于全连接网络中的参数数量 p 与宽度的平方成正比，因此参数需求高得令人难以承受：

$$p_{\text{c,train}} = \tilde{\Omega}\left((TK)^{12}\right).$$

即使是静态基线 NeuralUCB0（或经典 NTK）——它依赖于初始化时固定的 NTK——仍然需要相当程度的过度参数化。该方法遵循 [16], 引理 5.1 中概述的条件，该条件要求宽度按 $w = \tilde{\Omega}((TK)^4)$ 进行扩展，从而导致如下参数需求：

$$p_{\text{c,no-train}} = \tilde{\Omega}\left((TK)^8\right).$$

相比之下，我们的 QNTK-UCB 框架以显著更高效的参数规模实现了类似的遗憾保证。如定理 2 所示，我们的界成立的条件是：

$$p_{\text{q,no-train}} = \tilde{\Omega}\left((TK)^3\right).$$

这一显著差异凸显了量子模型在模型紧凑性方面的优势。通过利用相对较小的量子电路（具有较小的 p）所对应的高维希尔伯特空间，QNTK-UCB 提供了一种稳健且具有数学保证的核机制状态，而无需像将经典深度网络线性化那样进行过度参数化。

归纳偏置与表征能力。

除了提升效率外，QNTK 还引入了一种独特的归纳偏置。RBF 和 Mat´ern 等经典核，或标准神经切核，往往偏好经典意义上的平滑函数。相比之下，QNTK 从根本上受到量子

电路架构的影响，尤其是其纠缠结构和数据编码映射 $V(\mathbf{x})$。

量子神经网络将经典输入映射到维度为 2^m的指数级大希尔伯特空间中。这种高维嵌入使 QNTK 能够捕捉反映量子力学过程固有特性的关联，而这些关联对经典模型而言可能难以捕捉。因此，我们预计，在“量子原生”的老虎机任务中，例如优化变分量子本征求解器，或对物质相进行分类（这些任务的底层奖励函数具有与量子电路一致的对称性），QNTK-UCB 将优于经典基准。我们的实验结果支持这一预期。

在上下文老虎机框架中，可以通过观测到的上下文的 QNTK 格拉姆矩阵的谱特征来理解这种量子归纳偏置的优势。当 QNTK 的归纳偏置与真实奖励函数 h有效契合时，与通用的各向同性经典核（例如径向基函数）相比，其特征值衰减更快；后者往往会将概率质量分散到特征空间中。这种更快的谱衰减会使定义 d_{q}的总和$\widetilde{\ }$最小化，从而有效缩窄置信椭球的宽度。因此，算法只需少得多的样本，就能将后验方差降至最优性差距以下，从而降低探索的样本复杂度。

信息增益与有效维数。
为了理解定理 2 中的遗憾界，将对数行列式项改写为核信息增益是很有帮助的。回顾一下，QNTK 特征映射 $\boldsymbol{\phi}(\mathbf{x}) = \frac{1}{\sqrt{N_K(m)}}\nabla_{\boldsymbol{\theta}} f(\mathbf{x};\boldsymbol{\theta}_0)$ 诱导出极限格拉姆矩阵 $\bar{\mathbf{K}} \in \mathbb{R}^{TK\times TK}$。该量

$$\gamma_T^{(\mathrm{q})} := \log\det\left(\mathbf{I} + \frac{1}{\lambda}\bar{\mathbf{K}}\right)$$

是核老虎机分析中常见的信息增益项 [13]。我们的量子有效维数正是其归一化版本：

$$\widetilde{d}_{\mathrm{q}} = \frac{\gamma_T^{(\mathrm{q})}}{\log(1+TK/\lambda)}.$$

因此，定理 2 中的遗憾界由 $\gamma_T^{(\mathrm{q})}$ 控制，等价地，也可由 $\widetilde{d}_{\mathrm{q}}$ 控制。

在经典 NeuralUCB 分析中 [16]，相同的结构也会出现，只不过使用的是经典极限 NTK 格拉姆矩阵 $\mathbf{H}$ [34]，而非 $\mathbf{K}$。经典界取决于：

$$\gamma_T^{(\mathrm{c})} := \log\det\left(\mathbf{I} + \frac{1}{\lambda}\mathbf{H}\right) \quad \text{and} \quad \widetilde{d}_{\mathrm{c}} = \frac{\gamma_T^{(\mathrm{c})}}{\log(1+TK/\lambda)}.$$

我们的定理表明，一旦进入核化（无需训练）机制，“量子”与“经典”的区别主要体现在相应极限核矩阵的谱上，即$\mathbf{K}$ 与$\mathbf{H}$。因此，

$\bar{}$只要$\mathbf{K}$ 的特征值在实际观测到的上下文序列上比$\mathbf{H}$的特征值衰减得更快，我们就能获得更低的信息增益 $\gamma_T^{(\mathrm{q})}$ ，从而得到更紧的遗憾保证。

d_{q}的得为“贫瘠高原”现象中一种量子特有的权衡。在变分量子算法中，更深或结构较弱的电路通常会表现出显著的测度集中效应，导致梯度呈指数级消失，使基于梯度的训练变得困难。然而，在我们的核化设定中，这种集中效应发挥着不同且具有建设性的作用：

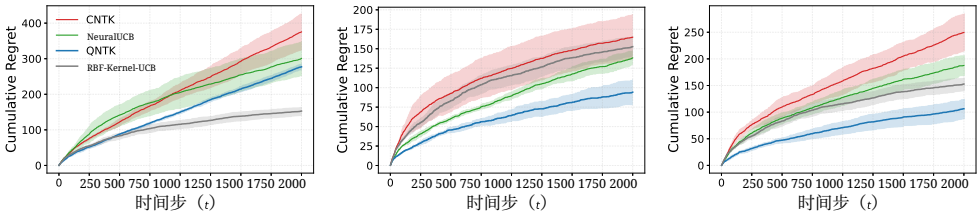
•**通过梯度缩放实现的隐式正则化**。 贫瘠高原现象表明，梯度的幅值 $\|\nabla_{\boldsymbol{\theta}} f(\mathbf{x};\, \boldsymbol{\theta}_0)\|$ 会随着量子比特数量 m 趋于无穷而呈指数级消失。由于经验量子神经切线核基于这些梯度的内积，梯度集中会导致格拉姆矩阵 $\hat{}\mathbf{K}$的矩阵元素和特征值均减小。近期的理论研究 [39] 还表明，量子神经网络具有较强的表达能力时，QNTK 值会呈指数级集中于零。因此，与源自宽而非集中型经典网络的核的信息增益相比，项 $\gamma_{(\mathrm{q})\,T} = \log\det(\mathbf{I} + \bar{}\mathbf{K}/\lambda)$ 会显著减小。这样的谱压缩构成了一种隐式正则化形式，可能会降低遗憾上界。

• 然而，需要注意的是，集中现象引起的谱收缩并不会自动带来收益。考虑定理 2 中引入的可实现性常数 S，它受再生核希尔伯特空间中奖励函数范数的影响，具体如下：$S^2 \approx \mathbf{h}^\top \bar{}\mathbf{K}^{-1}\mathbf{h}$。如果集中现象导致整个核按一个常数因子缩小，那么参数范数 S 会随之增大，抵消有效维数降低所带来的优势。

真正的量子优势源于非均匀集中。有效的量子架构应体现核-目标对齐；也就是说，它应在与奖励函数$\mathbf{h}$对齐的特定方向上保留较大的特征值，从而确保$\mathbf{h}^\top \bar{}\mathbf{K}^{-1}\mathbf{h}$ 始终具有统一上界。与此同时，它应在大多数正交的“无关”方向上实现显著集中。$\tilde{}$在这种情况下，随着谱尾部消失，量子有效维数 d_{q}会迅速下降，而可实现性常数 S 则保持较小，因为信号方向得以保留。这种选择性谱衰减使 QNTK-UCB 得以优于经典基线。

4 实验

为验证我们的理论发现，我们将 QNTK-UCB 的性能与最先进的经典神经网络老虎机算法进行比较。我们的实验旨在检验这一假设：量子核能够为非线性奖励函数提供更优的归纳偏置和更高的参数效率。



(a) $m = 3,\ p = 36$　　(b) $m = 5,\ p = 60$　　(c) $m = 10,\ p = 120$

图 2：基于高斯分位数分类的多臂老虎机任务奖励

4.1 高斯分位数

我们考虑一个包含 K 个臂、具有 $K = 2$ 的上下文老虎机问题，其中奖励函数由 $\mathbb{R}^d$ 中的非线性决策边界定义。基特征向量（记为 $\mathbf{x}_t$）从多维高斯分布中采样。真实类别标签 $y_t \in \{0, 1\}$ 根据高斯分位数赋值。从几何上看，这种设置会在特征空间中形成同心超球壳，类似于区分“圆”或“球体”的非线性二分类任务。

在每一轮次 t，智能体都会接收一个上下文向量$\mathbf{x}_t \in \mathbb{R}^d$。为建模各臂特定的奖励，我们采用标准的不相交上下文编码 [3]：智能体会观察一组臂特征 $\{\mathbf{x}_{t,\ a}\}_{a\in\{0,1\}}$ 其中 $\mathbf{x}_{t,\ 0} = [\mathbf{x}_t,\ \mathbf{0}]$ 以及 $\mathbf{x}_{t,\ 1} = [\mathbf{0}, \mathbf{x}_t]$。智能体选择一个臂 a_t ；当 a_t 与真实类别标签 y_t匹配时，智能体会获得奖励 $r_t = 1$ ，否则获得 $r_t = 0$ 。

我们比较以下四种基准算法：

- QNTK（本文方法）：使用由包含 $L = 4$ 层且量子比特数量可变的强纠缠层拟设推导出的经验量子神经切线核$m \in \{3, 5, 10\}$。该拟设中的可训练参数数量为 $p = 3mL$。
- NeuralUCB：一种经典的神经上下文赌博机，通过梯度下降训练多层感知机（MLP）。该网络包含一个隐藏层，并使用 ReLU 激活函数。优化采用 Adam，学习率为 $\eta = 0.01$。
- CNTK（经典 NTK，或 NeuralUCB0 [16]）：一种核化 UCB 算法，使用随机初始化的经典多层感知机的固定经验神经切线核。
- RBF-Kernel-UCB [12]：使用径向基函数核的核化 UCB 算法。

为公平评估参数效率，我们将经典模型（Neu-ralUCB 和 C-NTK）的可训练参数数量限制为与其量子对应算法相同，即 p_{c}。对于一个具有 p_{q} 个参数的给定量子电路，我们通过解析方式调整经典多层感知机的宽度，使其总参数数量 p_{c} 满足 $p_{\text{c}} \approx p_{\text{q}}$。对于所有模型，我们都进行了网格搜索，以优化正则化参数 $\lambda \in \{0.01, 0.1, 1.0\}$ 和固定探索半径 $\beta_t = \beta \in \{0.05, 0.1, 0.5, 1.0, 3.0\}$。

图 2 展示了在 $T = 2000$ 条独立试验中累积遗憾的平均值。我们分析了模型复杂度三个不同区间内的性能：

• 欠参数化区域（$m = 3,\ \ p = 36$）：如图 2(a) 所示，模型容量受到限制。所有算法都呈现出相对陡峭的遗憾曲线，表明模型过于简单，无法完美捕捉非线性决策边界。然而，QNTK 的遗憾仍低于经典基线。这证实了我们关于参数效率的假设：即使量子比特数量极少，量子特征映射仍能提供比同等规模的经典网络更丰富的表示，从而在严格的资源约束下更好地近似奖励函数。

f

量子比特，量子特征映射仍能提供比同等规模的经典网络更丰富的表示，从而在严格的资源约束下更好地近似奖励函数。

• 最优区域（$m = 5,\ \ p = 60$）：在图 2(b) 中，模型规模增至中等水平。在此，我们观察到所有方法都呈现出明显的次线性遗憾，表明模型具有足够的表达能力来学习该任务。QNTK 始终保持明显优势，展现了量子归纳偏置在此任务上的优越性。

• 过参数化（$m = 10, p = 120$）：图 2(c) 揭示了不同的表现趋势。对于经典方法（NeuralUCB 和 C-NTK），其累积遗憾相比于 $m = 5$ 情况变得更加陡峭。经典学习理论预示了这种性能下降：随着参数数量 p 增加，模型需要更多数据才能收敛，而遗憾界中的方差项（由有效维数决定）也会增大。另一方面，QNTK 依然稳健，其在 $m = 10$下的遗憾与 $m = 5$下的遗憾非常接近，没有出现性能下降或过拟合的迹象。这从实证上验证了第3.3节中讨论的 QNTK 隐式正则化特性。经典有效维数会随着 p 增加，而~量子有效维数 d_{q} 由于核谱的集中而趋于饱和，使量子算法能够抵御较高的模型复杂度。

为进一步探究图中观察到的性能差异。2，我们分析了特征表示的有效维数随量子比特数（因此也就是参数数量）的变化情况。我们绘制了经验特征维数（$\log\det(\mathbf{I}+\hat{\mathbf{K}}/\lambda)/\log(1+TK/\lambda)$），其中 $T = 2000$如图3所示。我们观察到两种截然不同的行为，这为我们的结果提供了更多见解。

以红色显示的经典 NTK，其有效维数呈单调递增，并始终高于 QNTK。尽管这一趋势表明其表达能力很强，但在过参数化区域中观察到的维数过高（$m = 10$）表明模型容量可能存在“过度消耗”。这一现象与图2(c)中所示较差的遗憾性能直接相关；经典模型的方差也更高。

相反，图中以蓝色表示的 QNTK 总体呈下降趋势。当 m 从 3 增至 5 时，曲线最初上升，这反映出有效捕捉非线性决策边界所需的表达能力提升。值得注意的是，超过 $m = 5$ 后，有效维数达到饱和，随后开始下降。这种行为体现了量子特征空间中的测度集中现象，其中内在正则化机制限制了量子模型的复杂度，防止其随参数数量增加而无限增长。

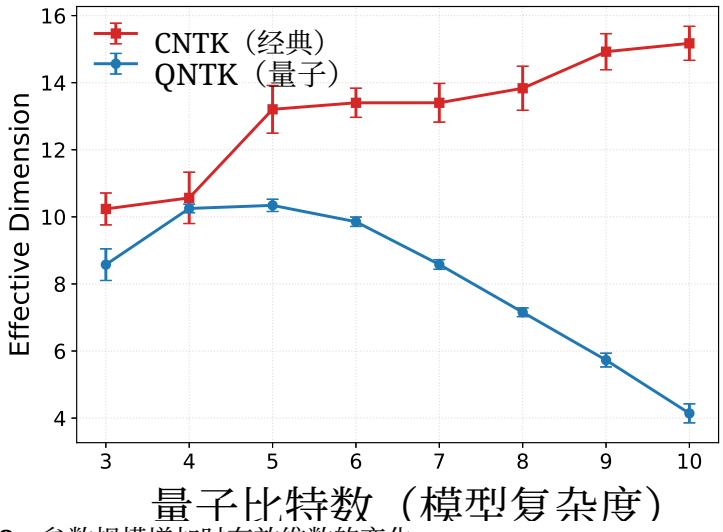


图 3：参数规模增加时有效维数的变化

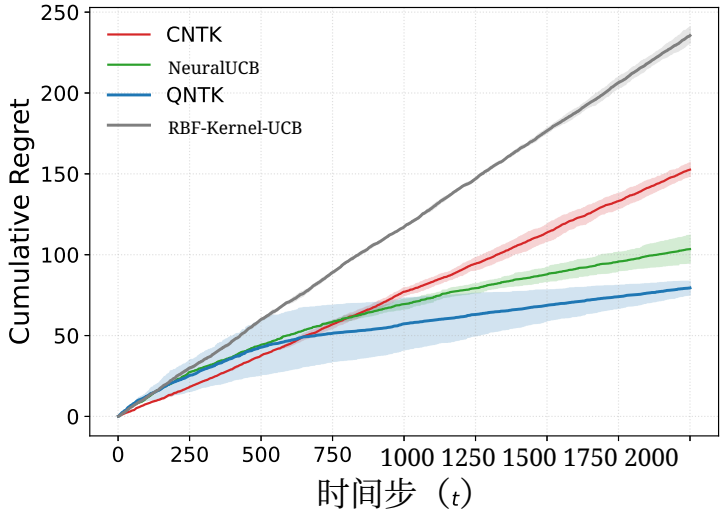


图 4：带有VQE优化起点推荐的多臂老虎机任务

4.2 在线量子初始态推荐

为展示 QNTK 在量子原生任务中的实用性，我们研究如何为变分量子本征求解器（VQE）寻找能够求出哈密顿量 [40] 基态的最优初始态。VQE 优化对初始化高度敏感。我们将 VQE 试探态参数的选择表述为

一个带有侧信息的老虎机问题（或上下文老虎机问题），其中侧信息由问题哈密顿量给出。

我们聚焦于一　族由 4 个量子比特横场伊辛哈密顿　量组成，

$$H(c) = -\sum_{i=1}^{m-1} \sigma_i^Z \sigma_{i+1}^Z - c\sum_{i=1}^{m} \sigma_i^X,$$

其中 σ_i^A 表示作用于第 i 个量子比特的泡利 A 算符，c 为横场强度。在每一轮 t，环境生成一个场强度 c_t，并构造相应的哈密顿量 $H(c_t)$，将其作为共享上下文提供给学习者。

学习者可以选择 $K = 5$ 个“臂”，每个臂对应 VQE 的一个固定且不同的初始态 $|\psi_a\rangle$。在时刻 t，学习者选择一个臂 $a_t \in [K]$，环境以 $|\psi_{a_t}\rangle$ 为初始态，使用 VQE 试探态电路 $U(\boldsymbol{\theta})$ 执行一次短深度 VQE 优化，以近似求得 $H(c_t)$ 的基态。这里，我们在 $m = 4$ 个量子比特上使用浅层双层硬件高效试探态电路 $U(\boldsymbol{\theta})$。该拟设有 2 层，每层都先对所有量子比特施加逐量子比特欧拉旋转 Rot(·)，然后接入线性纠缠链，其中包含受控非门。能量目标是其期望值

$$E(\boldsymbol{\theta}; c_t) = \langle \psi_{a_t} | U(\boldsymbol{\theta})^\dagger H(c_t) U(\boldsymbol{\theta}) | \psi_{a_t} \rangle$$

并且我们执行 $I = 5$ 个梯度步，以获得 $\boldsymbol{\theta}_I$。由此得到的近似基态 $|\psi(\boldsymbol{\theta}_I | a_t, c_t)\rangle := U(\boldsymbol{\theta}_I)|\psi_{a_t}\rangle$ 取决于初始态和哈密顿量侧信息。随后，学习者观察到奖励

$$r_{t,a_t} = -\mathrm{Tr}(H(c_t)|\psi(\boldsymbol{\theta}_I|a_t, c_t)\rangle\langle\psi(\boldsymbol{\theta}_I|a_t, c_t)|) + \xi_t,$$

该奖励是最终能量的相反数，并受到高斯噪声 ξ_t 的影响，用于模拟有限次测量造成的误差。

与前一个实验类似，我们将 QNTK-UCB 与 NeuralUCB、CNTK-UCB（NeuralUCB0）和 RBF-Kernel-UCB 进行比较。为确保比较公平，我们保证量子模型和经典模型的参数数量相同。此外，我们使用与第 4.1 节相同的网格搜索策略，为所有算法优化超参数；从集合中选取正则化参数 λ，$\{0.01, 0.1, 1.0\}$ 并从集合 β 中选取固定的探索半径 $\{0.05, 0.1, 0.5, 1.0, 3.0\}$。图 4 展示了各种算法的累积遗憾。

与经典基线相比，QNTK 智能体显然表现更优。从“哈密顿量参数 c”到“最优初始态”的映射由伊辛模型内在的相变决定。QNTK 源自量子电路，因此能够自然地捕捉希尔伯特空间景观中的关联与对称性。相比之下，经典网络缺乏这种特定的归纳偏置，因此需要更多样本才能学会从哈密顿量参数到最优拟设初始参数的映射。

5 结论

我们提出了一类新型量子增强神经上下文赌博机算法，它们不仅能实现与经典神经UCB方法相当的遗憾性能，而且所需的模型参数数量显著更少。借助QNTK理论的最新进展，我们推导出遗憾界，其缩放形式为 $\mathcal{O}(\tilde{d}_{\mathrm{q}}\sqrt{T})$，其中 $\tilde{d}_{\mathrm{q}}$ 表示 QNTK 的有效维数。该方法有效地运行在 QNTK机制内，并简化为具有静态量子切线核的核化模型。因此，我们成功应对了贫瘠高原带来的挑战，并避免了在上下文老虎机应用中显式训练量子神经网络所需的高昂计算成本。

尽管如此，依赖静态核可能会限制模型在面对复杂奖励函数时的表达能力。未来研究可以探索混合量子-经典模型，在量子表达能力与经典可训练性之间取得平衡。潜在方向包括采用近期研究所提出的架构，将量子特征映射与可训练经典神经网络相结合 [41]。

此外，尽管我们的分析强调参数缩减是量子优势的一个关键来源，但我们的框架也支持电路深度随量子比特数量增加而扩展。这种灵活性使人们能够开发一系列被认为经典模拟困难的量子电路，从而引入一类独特的量子模型，其动力学无法由经典算法高效复现。因此，我们的工作不仅是迈向上下文老虎机量子效率的重要一步，也为探索超越单纯参数效率的量子优势提供了有前景的途径。

致谢。STJ 获得英国皇家学会国际旅行资助（IES\R2\242104）的支持，并部分获得 EPSRC 量子技术职业加速奖学金的资助。

参考文献

[1] Tewari,A., Murphy,S.A.: 从广告到干预：移动健康中的上下文老虎机。在：移动健康：传感器、分析方法与应用，第 495– 517 页。Springer，Cham（2017）[2] Varatharajah, Y., Berry, B.: 一种基于上下文老虎机的方法，用于临床试验中的知情决策。《Life》**12**(8), 1277（2022）[3]Li, L., Chu, W., Langford, J., Schapire, R.E.: 个性化新闻文章推荐的上下文老虎机方法。在：第19届万维网国际会议论文集，第 661–670 页（2010）[4] Tang, L., Jiang, Y., Li, L., Li, T.: 用于个性化推荐的集成上下文老虎机。在：第8届ACM推荐系统会议论文集，第 73–80 页（2014）

[5] Bouneffouf, D., Bouzeghoub, A., Gan.carski, A.L.: 一种用于移动上下文感知推荐系统的上下文老虎机算法。在：神经信息处理国际会议，第 324–331 页（2012）。Springer[6]Abbasi-yadkori, Y., P´al, D., Szepesv´ari, C.: 线性随机老虎机的改进算法。在：《神经信息处理系统进展》，第24卷（2011）[7] Chu, W., Li, L., Reyzin, L., Schapire, R.: 具有线性收益函数的上下文老虎机。在：第十四届人工智能与统计国际会议论文集，第 208–214 页（2011）。JMLR研讨会与会议论文集[8] Agrawal, S., Goyal, N.: 具有线性收益的上下文老虎机的Thompson采样。在：国际机器学习会议，第 127–135 页（2013）。PMLR[9]Oh, M.-h., Iyengar, G.: 多项式Logit上下文老虎机的Thompson采样。《神经信息处理系统进展》**32** （2019）[10] Filippi, S., Cappe, O., Garivier, A., Szepesv´ari, C.: 参数老虎机：广义线性情形。《神经信息处理系统进展》**23**（2010）[11] Li, L., Lu, Y., Zhou, D.: 广义线性上下文老虎机的可证明最优算法。在：国际机器学习会议，第 2071–2080 页（2017）。PMLR[12] Valko,M., Korda,N., Munos, R., Flaounas, I., Cristianini, N.: 核化上下文老虎机的有限时间分析。arXiv预印本 arXiv:1309.6869（2013）[13]Srinivas, N., Krause, A., Kakade, S.M., Seeger, M.W.: 上下文老虎机场景下高斯过程优化的信息论遗憾界。IEEE信息论汇刊 **58**(5), 3250–3265（2012）[14]Riquelme, C., Tucker, G., Snoek, J.: 深度贝叶斯老虎机对决。在：国际学习表征会议，第9卷（2018）[15] Zahavy, T., Mannor, S.: 深度神经线性老虎机：通过似然匹配克服灾难性遗忘。arXiv预印本 arXiv:1901.08612（2019）[16] Zhou, D., Li, L., Gu, Q.: 采用基于UCB的探索的神经上下文老虎机。在：第37届国际机器学习会议论文集。《机器学习研究论文集》，第119卷，第 11492–11502 页。PMLR，???（2020）[17] Zhang, W., Zhou, D., Li, L., Gu, Q.: 神经Thompson采样。arXiv预印本

[18]Schuld, M., Bocharov, A., Svore, K.M., Wiebe, N.：以电路为中心的量子分类器。《物理评论A》**101**(3), 032308 (2020)[19] Du, Y., Hsieh, M.-H., Liu, T., Tao, D.：参数化量子电路的表达能力。《物理评论研究》**2**(3), 033125 (2020)[20] Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., Woerner, S.：量子神经网络的力量。《自然计算科学》**1**(6), 403–409 (2021)[21]Schuld, M., Sweke, R., Meyer, J.J.：数据编码对变分量子机器学习模型表达能力的影响。《物理评论A》**103**(3), 032430 (2021)[22] K¨ubler, J., Buchholz, S., Sch¨olkopf, B.：量子核的归纳偏置。《神经信息处理系统进展》**34**, 12661–12673 (2021)[23] Uvarov, A., Kardashin, A., Biamonte, J.D.：量子处理器中的机器学习相变。《物理评论A》**102**(1), 012415 (2020)[24] McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., Neven, H.：量子神经网络训练景观中的贫瘠高原。《自然通讯》**9**(1), 4812 (2018)[25] Napp, J.：量化无结构变分拟设模型的贫瘠高原现象。arXiv预印本 arXiv:2203.06174 (2022)[26] Abedi, E., Beigi, S., Taghavi, L.：量子惰性训练。Quantum **7**, 989 (2023)[27] Girardi, F., De Palma, G.：经过训练的量子神经网络是高斯过程。《数学物理通讯》**406**(4), 92 (2025)[28] Wan, Z., Zhang, Z., Li, T., Zhang, J., Sun, X.：量子多臂老虎机和随机线性老虎机具有对数遗憾。载于：AAAI人工智能会议论文集，第37卷，第10087–10094页 (2023)[29] Dai, Z., Lau, G.K.R., Verma, A., Shu, Y., Low, B.K.H., Jaillet, P.：量子贝叶斯优化。《神经信息处理系统进展》**36**, 20179–20207 (2023)[30] Hikima, Y., Murao, K., Takemori, S., Umeda, Y.：量子核化老虎机。载于：第40届人工智能不确定性会议 (2024)[31] Siam, Z.S., Guan, C., Liu, C.：量子非线性老虎机优化。arXiv预印本 arXiv:2503.03023 (2025)[32] Lumbreras, J., Haapasalo, E., Tomamichel, M.：多臂量子老虎机：

探索与利用：学习量子态性质时的权衡。Quantum **6**, 749 (2022)[33] Brahmachari, S., Lumbreras, J., Tomamichel, M.：面向量子数据的量子上下文老虎机与推荐系统。Quantum Machine Intelligence**6**(2), 58 (2024)[34] Jacot, A., Gabriel, F., Hongler, C.：神经切线核：神经网络中的收敛与泛化 (2020)[35] Goto, T., Tran, Q.H., Nakajima, K.：量子增强特征空间中量子机器学习模型的通用逼近性质。《物理评论快报》**127**(9), 090506 (2021)[36] Gonon, L., Jacquier, A.：量子神经网络和量子储备池的通用逼近定理与误差界。IEEE神经网络与学习系统汇刊 (2025)[37] Chowdhury, S.R., Gopalan, A.：论核化多臂老虎机。载于：第34届国际机器学习会议论文集——第70卷。ICML'17，第844–853页。JMLR.org，悉尼，新南威尔士州，澳大利亚 (2017)[38] Cao, Y., Gu,Q.：宽而深的神经网络随机梯度下降的泛化界 (2019)[39] Yu, L.-W., Li, W., Ye, Q., Lu, Z., Han, Z., Deng, D.-L.：量子神经切线核的表达能力诱导的集中现象 (2023)[40] Brahmachari, S., Lumbreras, J., Tomamichel, M.：面向量子数据的量子上下文老虎机与推荐系统。Quantum Machine Intelligence**6**(2) (2024)[41] Nakaji, K., Tezuka, H., Yamamoto, N.：神经切线核机制下的量子经典混合神经网络。量子科学与技术**9**(1), 015022 (2023)

附录 A 定理的证明2

本节给出定理 2的证明。首先回顾一些定义：

ˆ 1. 经验 QNTK，$\mathbf{K}$：由一个随机初始化的 QNN 实例计算得到的随机神经切线核（对于给定的固定结构，宽度为 m）。

$$\hat{\mathbf{K}}_{ij} = \frac{1}{N_K(m)} \nabla_{\boldsymbol{\theta}} f(\mathbf{x}^i; \boldsymbol{\theta}_0)^\top \nabla_{\boldsymbol{\theta}} f(\mathbf{x}^j; \boldsymbol{\theta}_0).$$

ˆ 我们省略 $\mathbf{K}$ 对 $\boldsymbol{\theta}_0$ 的依赖。除非另有说明，经验核始终在随机初始化 $\boldsymbol{\theta}_0$ 处进行评估。

2. 解析 QNTK，$\mathbf{K}$：经验核的期望值

$$\mathbf{K}_{ij} = \mathbb{E}_{\boldsymbol{\theta}_0}\left[\hat{\mathbf{K}}_{ij}\right].$$

¯3. 极限量子神经正切核，$\mathbf{K}$：解析核在无限量子比特极限下收敛到的极限核。

$$\bar{\mathbf{K}}_{ij} = \lim_{m\to\infty} \mathbf{K}_{ij}.$$

为方便记号，我们再作一个假设：

假设 6 对所有 $\mathbf{x} \in \mathcal{X}$，上下文均已归一化，使得 $\|\mathbf{x}\|_2 = 1$，且量子神经网络模型在初始化时以如下意义为中心：

$$f(\mathbf{x};\boldsymbol{\theta}_0) = 0 \qquad \forall \mathbf{x} \in \mathcal{X}. \tag{A1}$$

请注意，这一假设较为宽松，且仅为简化可实现性相关表述而提出。事实上，给定任意模型 $f(\cdot;\boldsymbol{\theta})$ 和初始化 $\boldsymbol{\theta}_0$，我们可以定义中心化模型

$$\widetilde{f}(\mathbf{x};\boldsymbol{\theta}) := f(\mathbf{x};\boldsymbol{\theta}) - f(\mathbf{x};\boldsymbol{\theta}_0),$$

˜它满足：对于所有$\mathbf{x}$，$f(\mathbf{x};\boldsymbol{\theta}_0) = 0$，并且在初始化时具有相同的切向特征，即 $\nabla_{\boldsymbol{\theta}}\widetilde{f}(\mathbf{x};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \nabla_{\boldsymbol{\theta}} f(\mathbf{x};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$。

A.1 可实现性

引理 A.1 对于任意 $\varepsilon > 0$ 和 $\delta \in (0,1)$，都存在一定数量的量子比特 m_0 以及一种量子神经网络结构，使得对于所有 $m \geq m_0$，以至少 $1-\delta$ 的概率（对 $\boldsymbol{\theta}_0$ 的随机初始化而言），有：

$$\left|\hat{\mathbf{K}}(\mathbf{x},\mathbf{x}') - \bar{\mathbf{K}}(\mathbf{x},\mathbf{x}')\right| \leq \varepsilon$$

对于任意一对输入 $\mathbf{x}$，$\mathbf{x}' \in \mathcal{X}_{1:TK}$。

Proof 任取 $\mathbf{x},\mathbf{x}' \in \mathcal{X}_{1:TK}$。我们需要以期望经验核为桥梁，根据三角不等式，

$$\left|\hat{\mathbf{K}}(\mathbf{x},\mathbf{x}') - \bar{\mathbf{K}}(\mathbf{x},\mathbf{x}')\right| \leq \underbrace{\left|\hat{\mathbf{K}}(\mathbf{x},\mathbf{x}') - \mathbf{K}(\mathbf{x},\mathbf{x}')\right|}_{\text{Stochastic Part}} + \underbrace{\left|\mathbf{K}(\mathbf{x},\mathbf{x}') - \bar{\mathbf{K}}(\mathbf{x},\mathbf{x}')\right|}_{\text{Deterministic Part}}.$$

首先界定随机部分：对于任意给定的 $\varepsilon > 0$，根据假设 3 和定理 1，存在某个常数 c 使得

$$\mathbb{P}\left(\left|\hat{\mathbf{K}}(\mathbf{x},\mathbf{x}') - \mathbf{K}(\mathbf{x},\mathbf{x}')\right| \geq \frac{\varepsilon}{2}\right) \leq \exp\left[-c\,\varepsilon^2 N_K(m)^2 \frac{N(m)^4}{Lm|\mathcal{M}|^4|\mathcal{N}|^2}\right]. \tag{A2}$$

此处，m 是量子比特的数量，而 $\mathcal{M}$，$\mathcal{N}$，$N_K(m)$ $N(m)$是量子神经网络结构相关参数和归一化因子，定义见 [27]。注意，对于满足假设3的量子神经网络结构，(A2)右侧的表达式随 m递减；更多细节见引理A.3。

接下来，我们对确定性部分进行界定。根据假设2，有

$$\lim_{m\to\infty}\sup_{\mathbf{x},\mathbf{x}'\in\mathcal{X}}\left|\mathbf{K}(\mathbf{x},\mathbf{x}')-\bar{\mathbf{K}}(\mathbf{x},\mathbf{x}')\right|=0.$$

这意味着，对于任意给定的 $\varepsilon>0$，都存在一个量子比特数 m_2，使得对所有$m\geq m_2$:

$$\left|\mathbf{K}(\mathbf{x},\mathbf{x}')-\bar{\mathbf{K}}(\mathbf{x},\mathbf{x}')\right|\leq\frac{\varepsilon}{2}. \tag{A3}$$

结合（A2）和（A3）即可得到所需结果。 □

接下来，我们证明核矩阵在弗罗贝尼乌斯范数下的集中性。

引理 A.2 对于任意 $\varepsilon>0$ 和 $\delta\in(0,1)$，若量子神经网络架构满足缩放条件

$$\frac{N_K(m)^2N(m)^4}{Lm|\mathcal{M}|^4|\mathcal{N}|^2}=\Omega\left(\frac{1}{\varepsilon^2}\log\left(\frac{(TK)^2}{\delta}\right)\right),$$

and

$$m\geq C_\varepsilon\quad\text{such that}\quad\sup_{\mathbf{x},\mathbf{x}'\in\mathcal{X}}\left|\mathbf{K}(\mathbf{x},\mathbf{x}')-\bar{\mathbf{K}}(\mathbf{x},\mathbf{x}')\right|\leq\frac{\varepsilon}{2}.$$

那么，以至少 $1-\delta$ 的概率，在对 $\boldsymbol{\theta}_0$ 进行随机初始化后，我们有

$$\left\|\hat{\mathbf{K}}-\bar{\mathbf{K}}\right\|_{\mathrm{F}}\leq TK\varepsilon.$$

Proof 引理 A.1 表明，对于任意 $\varepsilon'>0$ 和 $\delta'\in(0,1)$，如果量子比特数 m足够大，那么

$$\mathbb{P}\left(\left|\hat{\mathbf{K}}_{ij}-\bar{\mathbf{K}}_{ij}\right|\geq\varepsilon'\right)\leq\delta',$$

这里$\hat{\mathbf{K}}_{ij}:=\hat{\mathbf{K}}(\mathbf{x}^i,\mathbf{x}^j)$，$\bar{\mathbf{K}}$的情况也类似。我们希望此界对矩阵中所有$(TK)^2$个条目同时成立，且总失败概率至多为 δ。我们使用并集界：

$$\mathbb{P}\left(\exists(i,j)\in[TK]^2:\left|\hat{\mathbf{K}}_{ij}-\bar{\mathbf{K}}_{ij}\right|\geq\varepsilon\right)\leq\sum_{i=1}^{TK}\sum_{j=1}^{TK}\mathbb{P}\left(\left|\hat{\mathbf{K}}_{ij}-\bar{\mathbf{K}}_{ij}\right|\geq\varepsilon\right).$$

设单个条目的失败概率为 $\delta'=\delta/(TK)^2$。根据集中界（引理 A2 中的式 A.1），我们需要满足：

$$\exp\left(-c\frac{N_K(m)^2N(m)^4}{Lm|\mathcal{M}|^4|\mathcal{N}|^2}\varepsilon^2\right)\leq\frac{\delta}{(TK)^2}.$$

取对数并重新整理，可得架构所需满足的缩放条件：

$$\frac{N_K(m)^2N(m)^4}{Lm|\mathcal{M}|^4|\mathcal{N}|^2}\geq\frac{1}{c\varepsilon^2}\log\left(\frac{(TK)^2}{\delta}\right),$$

这正是本引理中陈述的第一个条件。结合第二个条件，我们可以得出，以至少 $1-\delta$ 的概率，对所有 $(i,\ j)$:

$$\left|\hat{\mathbf{K}}_{ij}-\bar{\mathbf{K}}_{ij}\right|\leq\varepsilon.$$

现在，我们可以对差值的弗罗贝尼乌斯范数进行界定

$$\left\|\hat{\mathbf{K}}-\bar{\mathbf{K}}\right\|_{\mathrm{F}}^{2}=\sum_{i=1}^{TK}\sum_{j=1}^{TK}\left|\hat{\mathbf{K}}_{ij}-\bar{\mathbf{K}}_{ij}\right|^{2}\leq\sum_{i=1}^{TK}\sum_{j=1}^{TK}\varepsilon^{2}=(TK)^{2}\varepsilon^{2}.$$

因此

$$\left\|\hat{\mathbf{K}}-\bar{\mathbf{K}}\right\|_{\mathrm{F}}\leq TK\varepsilon,$$

如所期望的。 □

引理 A.3 引理 A.2中的假设可以通过某种量子神经网络结构实现，其中 $m=\Omega\left(\frac{1}{\varepsilon^{2}}\log\left(\frac{(TK)^{2}}{\delta}\right)\right)$。

Proof 我们在图 1中给出了一个量子神经网络架构示例。第 2.5 节 [27] 证明了其 $N(m)=\sqrt{m}$, $|\mathcal{M}|=O(L)$, $|\mathcal{N}|=O(L^{2})$, 其值为 $L=O(\log m)$ 或常数（即 $L=O(1)$）。将这些代入引理 A.2 中的第一个条件，即可得到所需结果。 □

引理 A.4 假设引理 A.2 的条件成立。以至少 $1-\delta$的概率，存在一个参数向量 $\boldsymbol{\theta}^{*}\in\mathbb{R}^{p}$ ($p=\dim(\boldsymbol{\theta})$)，使得对于所有上下文 $\mathbf{x}^{i}\in\mathcal{X}_{1:TK}$:

1. 奖励函数可以由线性模型完美表示:

$$h(\mathbf{x}^{i})=\langle\nabla_{\boldsymbol{\theta}}f(\mathbf{x}^{i};\boldsymbol{\theta}_{0}),\boldsymbol{\theta}^{*}-\boldsymbol{\theta}_{0}\rangle.$$

2. 解向量的范数以初始参数 $\boldsymbol{\theta}_{0}$ 为中心有界:

$$N_{K}(m)\left\|\boldsymbol{\theta}^{*}-\boldsymbol{\theta}_{0}\right\|_{2}^{2}\leq 2\mathbf{h}^{\top}\bar{\mathbf{K}}^{-1}\mathbf{h}.$$

这里 $\mathbf{h}=[h(\mathbf{x}^{1})\ldots,h(\mathbf{x}^{TK})]^{\top}$ 是由真实奖励构成的向量。

Proof 根据引理 A.2 和并集界，选择 $\varepsilon'=\lambda_{0}/(2TK)$，则对于足够大的 m，以至少 $1-\delta$ 的概率有:

$$\|\hat{\mathbf{K}}-\bar{\mathbf{K}}\|_{\mathrm{F}}\leq TK\,\varepsilon'=\frac{\lambda_{0}}{2}.$$

因此,

$$\hat{\mathbf{K}}\succeq\bar{\mathbf{K}}-\|\hat{\mathbf{K}}-\bar{\mathbf{K}}\|_{2}\,\mathbf{I}\succeq\bar{\mathbf{K}}-\|\hat{\mathbf{K}}-\bar{\mathbf{K}}\|_{\mathrm{F}}\,\mathbf{I}\succeq\bar{\mathbf{K}}-\frac{\lambda_{0}}{2}\mathbf{I}\succeq\frac{\lambda_{0}}{2}\mathbf{I}\succ 0,$$

第二个不等式由 $\|\mathbf{A}\|_{2}\leq\|\mathbf{A}\|_{\mathrm{F}}$ 得出，第四个不等式由 $\bar{\mathbf{K}}\succeq\lambda_{0}\mathbf{I}$ 得出。因此，$\hat{\mathbf{K}}$ 是正定的。

令 $\mathbf{J}=\mathbf{J}_{0}=\mathbf{J}(\boldsymbol{\theta}_{0})\in\mathbb{R}^{p\times TK}$ 表示初始化时的雅可比矩阵（其列为梯度向量 $\nabla_{\boldsymbol{\theta}}f(\mathbf{x}^{i};\boldsymbol{\theta}_{0})$）。根据定义,

$$\hat{\mathbf{K}}=\frac{1}{N_{K}(m)}\mathbf{J}^{\top}\mathbf{J}.$$

由于 $\hat{\mathbf{K}}\succ 0$，矩阵 $\mathbf{J}^{\top}\mathbf{J}$ 可逆，且 $\mathbf{J}$ 具有满秩。定义

$$\boldsymbol{\theta}^{*}=\boldsymbol{\theta}_{0}+\mathbf{J}\,(\mathbf{J}^{\top}\mathbf{J})^{-1}\mathbf{h}.$$

Then

$$\mathbf{J}^\top(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0) = \mathbf{J}^\top\mathbf{J}\,(\mathbf{J}^\top\mathbf{J})^{-1}\mathbf{h} \;=\; \mathbf{h},$$

即，对于每个 i, $\langle \nabla_{\boldsymbol{\theta}} f(\mathbf{x}^i;\boldsymbol{\theta}_0)\ \ \boldsymbol{\theta}^* - \boldsymbol{\theta}_0\rangle = h(\mathbf{x}^i)$，这证明了引理的第 1 部分。

对于引理的第 2 部分，根据构造可知，

$$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\|_2^2 \;=\; \mathbf{h}^\top(\mathbf{J}^\top\mathbf{J})^{-1}\mathbf{h} \;=\; \frac{1}{N_K(m)}\,\mathbf{h}^\top\hat{\mathbf{K}}^{-1}\mathbf{h}.$$

回顾 $\hat{\mathbf{K}} \succeq \bar{\mathbf{K}} - \frac{\lambda_0}{2}\mathbf{I} \succeq \frac{1}{2}\bar{\mathbf{K}}$，因此 $\hat{\mathbf{K}}^{-1} \preceq 2\,\bar{\mathbf{K}}^{-1}$。于是

$$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\|_2^2 \;\le\; \frac{1}{N_K(m)}\,\mathbf{h}^\top(2\,\bar{\mathbf{K}}^{-1})\mathbf{h} \;=\; \frac{2}{N_K(m)}\,\mathbf{h}^\top\bar{\mathbf{K}}^{-1}\mathbf{h},$$

□

A.2 置信界与瞬时遗憾

引理 A.5 设 $\lambda > 0$ 且 $\delta \in (0,1)$。我们采用算法 1中的记号。以至少 $1-\delta$的概率，对所有 t,

$$\left\|\sqrt{N_K(m)}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0) - (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0)\right\|_{\mathbf{Z}_t} \le \beta_t \quad \text{where} \quad \beta_t = \nu\sqrt{\log\frac{\det(\mathbf{Z}_t)}{\det(\lambda\mathbf{I})} + 2\log\left(\frac{1}{\delta}\right)} + \sqrt{\lambda}S, \tag{A4}$$

where $\mathbf{Z}_t = \lambda\mathbf{I} + \sum_{s=1}^{t}\boldsymbol{\phi}(\mathbf{x}_{s,a_s})\boldsymbol{\phi}(\mathbf{x}_{s,a_s})^\top$ and $S \ge \sqrt{2\mathbf{h}^\top\bar{\mathbf{K}}^{-1}\mathbf{h}}$.

Proof 由于我们已在引理 A.4 中证明奖励函数可由线性模型完美表示，即,

$$h(\mathbf{x}^i) = \langle \nabla_{\boldsymbol{\theta}} f(\mathbf{x}^i;\boldsymbol{\theta}_0)/\sqrt{N_K(m)}, \sqrt{N_K(m)}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)\rangle = \langle \boldsymbol{\phi}(\mathbf{x}^i), \sqrt{N_K(m)}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)\rangle,$$

直接应用线性老虎机的自归一化鞅不等式（参见 [6,定理 2]）可得

$$\left\|\sqrt{N_K(m)}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0) - \mathbf{Z}_t^{-1}\mathbf{b}_t\right\|_{\mathbf{Z}_t} = \left\|\sqrt{N_K(m)}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0) - (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0)\right\|_{\mathbf{Z}_t} \le \beta_t$$

这就是式（A4）中的界。此外，解向量的范数 $\boldsymbol{\theta}^*$ 有界。更确切地说,

$$N_K(m)\left\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\right\|_2^2 \le 2\,\mathbf{h}^\top\bar{\mathbf{K}}^{-1}\mathbf{h},$$

这证明了引理陈述中对 S 的条件是合理的 . □

引理 A.6 假设式（A4）中的置信事件在轮次 t 成立，即

$$\left\|\sqrt{N_K(m)}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0) - (\hat{\boldsymbol{\theta}}_{t-1} - \boldsymbol{\theta}_0)\right\|_{\mathbf{Z}_{t-1}} \le \beta_{t-1},$$

此外，假设 $\lambda \ge \max\{1, S^{-2}\}$。那么定义 $\widetilde{r}_t = h(\mathbf{x}_{t,a_t^*}) - h(\mathbf{x}_{t,a_t})$,

$$\widetilde{r}_t \le 2\,\beta_{t-1}\min\left\{\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|_{\mathbf{Z}_{t-1}^{-1}},\, 1\right\}.$$

Proof 对于任意上下文 $\mathbf{x}$，定义（时间 t）预测（后验）均值和方差

$$\hat{\mu}_t(\mathbf{x}) := \boldsymbol{\phi}(\mathbf{x})^\top(\hat{\boldsymbol{\theta}}_{t-1} - \boldsymbol{\theta}_0), \qquad s_t(\mathbf{x}) := \|\boldsymbol{\phi}(\mathbf{x})\|_{\mathbf{Z}_{t-1}^{-1}},$$

以及相应的乐观和悲观指标

$$U_t(\mathbf{x}) := \hat{\mu}_t(\mathbf{x}) + \beta_{t-1}s_t(\mathbf{x}), \qquad L_t(\mathbf{x}) := \hat{\mu}_t(\mathbf{x}) - \beta_{t-1}s_t(\mathbf{x}).$$

在事件 (A4) 上，由$\mathbf{Z}_{t-1}$-范数下的柯西–施瓦茨不等式可得，对任意$\mathbf{x}$,

$$\begin{aligned}
\left|h(\mathbf{x}) - \hat{\mu}_t(\mathbf{x})\right| &= \left|\boldsymbol{\phi}(\mathbf{x})^\top(\sqrt{N_K(m)}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0) - (\hat{\boldsymbol{\theta}}_{t-1} - \boldsymbol{\theta}_0))\right| \\
&\le \|\boldsymbol{\phi}(\mathbf{x})\|_{\mathbf{Z}_{t-1}^{-1}} \|\sqrt{N_K(m)}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0) - (\hat{\boldsymbol{\theta}}_{t-1} - \boldsymbol{\theta}_0)\|_{\mathbf{Z}_{t-1}} \\
&\le \beta_{t-1}s_t(\mathbf{x}),
\end{aligned}$$

因此

$$L_t(\mathbf{x}) \le h(\mathbf{x}) \le U_t(\mathbf{x}) \qquad \text{for all } \mathbf{x}. \tag{A5}$$

令 $a_t \in \operatorname{argmax}_{a\in[K]} U_t(\mathbf{x}_{t,a})$ 为算法选择的动作，并令 $a_t^* \in \arg\max_{a\in[K]} h(\mathbf{x}_{t,a})$ 为一个最优动作。根据乐观原则和 (A5),

$$h(\mathbf{x}_{t,a_t^*}) \le U_t(\mathbf{x}_{t,a_t^*}) \le U_t(\mathbf{x}_{t,a_t}), \qquad L_t(\mathbf{x}_{t,a_t}) \le h(\mathbf{x}_{t,a_t}).$$

用最左侧的不等式减去最右侧的不等式，得到

$$\tilde{r}_t = h(\mathbf{x}_{t,a_t^*}) - h(\mathbf{x}_{t,a_t}) \le U_t(\mathbf{x}_{t,a_t}) - L_t(\mathbf{x}_{t,a_t}) = 2\beta_{t-1}s_t(\mathbf{x}_{t,a_t}) = 2\beta_{t-1}\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|_{\mathbf{Z}_{t-1}^{-1}}.$$

这证明了 $\tilde{r}_t \le 2\beta_{t-1}\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|_{\mathbf{Z}_{t-1}^{-1}}$。

此外，由于奖励在 $[0,1]$，中有界，我们还有 $\tilde{r}_t \le 1$。结合这两个界可得

$$\tilde{r}_t \le \min\left\{2\beta_{t-1}\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|_{\mathbf{Z}_{t-1}^{-1}}, 1\right\}.$$

最后，由于 $\beta_{t-1} \ge \sqrt{\lambda}S \ge 1$，我们得到了所述结果。□

下面的论证与椭圆势引理 [6]类似。

引理 A.7 以下结论成立。

$$\sum_{t=1}^{T} \min\left\{\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|^2_{\mathbf{Z}_{t-1}^{-1}}, 1\right\} \le 2\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})}.$$

Proof 我们遵循 [6, 引理 11]的证明。

首先，根据初等矩阵恒等式可得，

$$\begin{aligned}
\det(\mathbf{Z}_t) &= \det(\mathbf{Z}_{t-1} + \boldsymbol{\phi}(\mathbf{x}_{t,a_t})\boldsymbol{\phi}(\mathbf{x}_{t,a_t})^\top) \\
&= \det(\mathbf{Z}_{t-1})\left(1 + \boldsymbol{\phi}(\mathbf{x}_{t,a_t})^\top\mathbf{Z}_{t-1}^{-1}\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\right) \\
&= \det(\mathbf{Z}_{t-1})\left(1 + \|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|^2_{\mathbf{Z}_{t-1}^{-1}}\right) \\
&= \det(\lambda\mathbf{I})\prod_{s=1}^{t}\left(1 + \|\boldsymbol{\phi}(\mathbf{x}_{s,a_s})\|^2_{\mathbf{Z}_{s-1}^{-1}}\right)
\end{aligned}$$

对 $t = 1,\ldots,T$ 进行望远镜求和并取对数，得到

$$\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})} = \sum_{t=1}^{T}\log\left(1 + \|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|^2_{\mathbf{Z}_{t-1}^{-1}}\right). \tag{A6}$$

注意，对所有 $x \in [0,1]$, $x \le 2\log(1+x)$ 均成立。因此，我们有 $\min\{x,1\} \le 2\log(1+x)$。因此

$$\sum_{t=1}^{T} \min\{\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|^2_{\mathbf{Z}_{t-1}^{-1}}, 1\} \le 2\sum_{t=1}^{T} \log\big(1+\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|^2_{\mathbf{Z}_{t-1}^{-1}}\big) = 2\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})}.$$

这证明了引理 . □

引理 A.8 令 $\hat{\mathbf{K}} \in \mathbb{R}^{TK\times TK}$ 表示所有上下文 $\mathcal{X}_{1:TK} = \{\mathbf{x}_{t,a}\}_{t\in[T],a\in[K]}$ 上的经验 QNTK 格拉姆矩阵，并令 $\bar{\mathbf{K}}$ 表示同一集合上的极限 QNTK 格拉姆矩阵。定义谱失配 $\boldsymbol{\Delta} := \hat{\mathbf{K}} - \bar{\mathbf{K}}$。则

$$\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})} \le \log\det\Big(\mathbf{I}+\frac{1}{\lambda}\bar{\mathbf{K}}\Big) + \frac{\sqrt{TK}}{\lambda}\|\boldsymbol{\Delta}\|_F. \tag{A7}$$

因此，根据量子有效维数的定义

$$\widetilde{d}_{\mathrm{q}} := \frac{\log\det\big(\mathbf{I}+\bar{\mathbf{K}}/\lambda\big)}{\log\big(1+TK/\lambda\big)},$$

我们有

$$\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})} \le \widetilde{d}_{\mathrm{q}}\log\Big(1+\frac{TK}{\lambda}\Big) + \frac{\sqrt{TK}}{\lambda}\|\boldsymbol{\Delta}\|_F.$$

Proof 考虑,

$$\begin{aligned}
\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})} &= \log\det\Big(\mathbf{I}+\frac{1}{\lambda}\sum_{t=1}^{T}\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\,\boldsymbol{\phi}(\mathbf{x}_{t,a_t})^\top\Big)\\
&\le \log\det\Big(\mathbf{I}+\frac{1}{\lambda}\sum_{i=1}^{TK}\boldsymbol{\phi}(\mathbf{x}^i)\boldsymbol{\phi}(\mathbf{x}^i)^\top\Big)\\
&= \log\det\Big(\mathbf{I}+\frac{1}{\lambda N_K(m)}\mathbf{J}\mathbf{J}^\top\Big) \qquad \text{(A8)}\\
&= \log\det\Big(\mathbf{I}+\frac{1}{\lambda N_K(m)}\mathbf{J}^\top\mathbf{J}\Big). \qquad \text{(A9)}
\end{aligned}$$

现在 $\frac{1}{N_K(m)}\mathbf{J}^\top\mathbf{J}$ 恰好是 $\hat{\mathbf{K}}$，记 $\hat{\mathbf{K}} = \bar{\mathbf{K}} + \boldsymbol{\Delta}$，我们有

$$\begin{aligned}
\log\det\Big(\mathbf{I}+\frac{1}{\lambda}\hat{\mathbf{K}}\Big) &= \log\det\Big(\mathbf{I}+\frac{1}{\lambda}(\bar{\mathbf{K}}+\boldsymbol{\Delta})\Big)\\
&\le \log\det\Big(\mathbf{I}+\frac{1}{\lambda}\bar{\mathbf{K}}\Big) + \Big\langle\frac{1}{\lambda}\Big(\mathbf{I}+\frac{1}{\lambda}\bar{\mathbf{K}}\Big)^{-1}, \boldsymbol{\Delta}\Big\rangle\\
&\le \log\det\Big(\mathbf{I}+\frac{1}{\lambda}\bar{\mathbf{K}}\Big) + \frac{1}{\lambda}\big\|\big(\mathbf{I}+\frac{1}{\lambda}\bar{\mathbf{K}}\big)^{-1}\big\|_{\mathrm{F}}\|\boldsymbol{\Delta}\|_{\mathrm{F}}\\
&\le \log\det\Big(\mathbf{I}+\frac{1}{\lambda}\bar{\mathbf{K}}\Big) + \frac{\sqrt{TK}}{\lambda}\|\boldsymbol{\Delta}\|_{\mathrm{F}},
\end{aligned}$$

其中我们使用了对数行列式的凹性(·)， $|\langle\mathbf{A},\ \mathbf{B}\rangle| \le \|\mathbf{A}\|_{\mathrm{F}}\|\mathbf{B}\|_{\mathrm{F}}$ 以及 $\|\mathbf{A}\|_{\mathrm{F}} \le \sqrt{TK}\|\mathbf{A}\|_2$ 对于任意 $\mathbf{A} \in \mathbb{R}^{TK\times TK}$。

因此我们有

$$\begin{aligned}
\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})} &\le \log\det\Big(\mathbf{I}+\frac{1}{\lambda}\bar{\mathbf{K}}\Big) + \frac{\sqrt{TK}}{\lambda}\|\boldsymbol{\Delta}\|_{\mathrm{F}}\\
&\le \widetilde{d}_{\mathrm{q}}\log\Big(1+\frac{TK}{\lambda}\Big) + \frac{\sqrt{TK}}{\lambda}\|\boldsymbol{\Delta}\|_{\mathrm{F}}.
\end{aligned}$$

□

A.3 定理2的证明

定理 2 *Fix any $\delta \in (0,1)$. Let $m = \Omega\left(\frac{(TK)^3}{\lambda^2}\log\left(\frac{(TK)^2}{\delta}\right)\right)$. Then, with probability at least $1-\delta$, the cumulative regret of QNTK-UCB (Algorithm 1) satisfies*

$$
\begin{aligned}
R_T &:= \sum_{t=1}^{T}\left[h(\mathbf{x}_{t,a_t^*}) - h(\mathbf{x}_{t,a_t})\right] \\
&\le 3\sqrt{T}\sqrt{\widetilde{d}_{\mathrm{q}}\log\left(1+\frac{TK}{\lambda}\right)+1}\left(\nu\sqrt{\widetilde{d}_{\mathrm{q}}\log\left(1+\frac{TK}{\lambda}\right)+1+2\log\left(\frac{1}{\delta}\right)}+\sqrt{\lambda}S\right),
\end{aligned}
$$

where $S \ge \sqrt{2\mathbf{h}^\top \mathbf{K}^{-1}\mathbf{h}}$, $\mathbf{h} = [h(\mathbf{x}^1)\dots, h(\mathbf{x}^{TK})]^\top$, and $\lambda \ge$最大值$\{1, S^{-2}\}$. Ignoring logarithmic terms and constants, this bound simplifies to

$$
R_T = \tilde{\mathcal{O}}\left(\widetilde{d}_{\mathrm{q}}\sqrt{T}\right)
$$

Proof 考虑,

$$
\begin{aligned}
R_T &= \sum_{t=1}^{T}\left(h(\mathbf{x}_{t,a_t^*}) - h(\mathbf{x}_{t,a_t})\right) \\
&\le \sum_{t=1}^{T} 2\beta_{t-1}\min\left\{\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|_{\mathbf{Z}_{t-1}^{-1}},\ 1\right\} \\
&\le 2\beta_T\sum_{t=1}^{T}\min\left\{\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|_{\mathbf{Z}_{t-1}^{-1}},\ 1\right\} \\
&\le 2\beta_T\sqrt{T}\sqrt{\sum_{t=1}^{T}\min\left\{\|\boldsymbol{\phi}(\mathbf{x}_{t,a_t})\|^2_{\mathbf{Z}_{t-1}^{-1}},\ 1\right\}} \\
&\le 2\,\beta_T\sqrt{2T\,\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})}},
\end{aligned}
$$

其中第三个不等式由柯西–施瓦茨不等式得出。回顾一下，谱失配为. $\boldsymbol{\Delta} := \hat{\mathbf{K}} - \bar{\mathbf{K}}$。代入

$$
\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})} \le \widetilde{d}_{\mathrm{q}}\log\left(1+\frac{TK}{\lambda}\right)+\frac{\sqrt{TK}}{\lambda}\|\boldsymbol{\Delta}\|_{\mathrm{F}}
$$

and

$$
\beta_T = \nu\sqrt{\log\frac{\det(\mathbf{Z}_T)}{\det(\lambda\mathbf{I})}+2\log\left(\frac{1}{\delta}\right)}+\sqrt{\lambda}S
$$

得到以下界

$$
\begin{aligned}
R_T \le 2\sqrt{2T\left(\widetilde{d}_{\mathrm{q}}\log\left(1+\frac{TK}{\lambda}\right)+\frac{\sqrt{TK}}{\lambda}\|\boldsymbol{\Delta}\|_{\mathrm{F}}\right)} \\
\times\left(\nu\sqrt{\left(\widetilde{d}_{\mathrm{q}}\log\left(1+\frac{TK}{\lambda}\right)+\frac{\sqrt{TK}}{\lambda}\|\boldsymbol{\Delta}\|_{\mathrm{F}}\right)+2\log\left(\frac{1}{\delta}\right)}+\sqrt{\lambda}S\right).
\end{aligned}
$$

由引理 A.2和A.3, $m = \Omega\left(\frac{T^3K^3}{\lambda^2}\log\left(\frac{(TK)^2}{\delta}\right)\right)$得到 $\|\boldsymbol{\Delta}\|_{\mathrm{F}} = \left\|\hat{\mathbf{K}} - \bar{\mathbf{K}}\right\|_{\mathrm{F}} \le TK\varepsilon \le \frac{\lambda}{\sqrt{TK}}$。由此得到所述界。□