arXiv:2611.32004v1 [quant-ph] 24 Nov 2026

超越经典极小极大速率：基于梯度的仿真优化中的量子采样

曹浩、胡建强、胡明杰、王启新复旦大学管理学院，中国上海 200433,
hcao21@m.fudan.edu.cn,hujq@fudan.edu.cn, 23110690009@m.fudan.edu.cn, 23110690014@m.fudan.edu.cn

我们研究一类凸仿真优化问题中的基于梯度的搜索，其梯度由含噪输出估计得到。对于独立输出，期望目标间隙的最优经典收敛速率为 $\mathcal{O}(N^{-2/3})$ 阶，其中 N 为查询预算。我们将量子均值估计应用于配对有限差分，并随着扰动缩小逐步提高其估计精度，从而将收敛速率提升至 $\mathcal{O}(N^{-4/3})$ 阶。此外，我们证明，对邻近评估进行耦合可分别将经典和量子收敛速率提升至 $\mathcal{O}(N^{-2/(3-\kappa)})$ 阶和 $\mathcal{O}(N^{-4/(3-\kappa)})$ 阶，其中 $\kappa \in [0,1]$ 描述了配对输出中的方差缩减。因此，在各种耦合阶数下，量子估计使查询复杂度中的精度指数减半。此外，我们建立了匹配下界，并证明在决策维度固定时，这些精度指数是极小极大最优的，即使查询位置和批量大小可以自适应选择也是如此。当每次仿真查询都必须满足可行性时，基于可行差分的重构仍保留这些精度指数，并且能够处理边界最优解。

Key words: 基于梯度的仿真优化；随机逼近；同时扰动；查询复杂度；量子采样

1. 引言

基于梯度的算法广泛用于优化具有连续决策的随机系统的期望性能（例如，参见 Fu，2002、2008；Fu 和 Hu，2012）。即使期望目标函数及其梯度无法通过解析方法获得，也可以根据含噪仿真输出估计梯度（Golovin 等，2017；Cao 等，2025年a；Hu 等，2025年a）。逐分量有限差分分别扰动各个决策变量（Kiefer 和 Wolfowitz，1952），而同时扰动随机逼近（SPSA）则同时扰动所有变量，并且仅使用两个位置的输出（Spall，1992）。在这两种方法中，减小扰动幅度都会降低导数偏差，但可能增加梯度估计的方差。

量子均值估计有望改善仿真的收敛阶

超越经典方差缩减所带来的优化收益（参见 Kleinman 等人，1999；He 等人，2024）。对于标准差至多为 s的标量输出，经典平均可达到

均方根误差（RMSE）容差 $\tau > 0$，所需样本数为 $\mathcal{O}(1+s^2/\tau^2)$。通过相干访问可逆采样器，量子均值估计可通过 $\mathcal{O}(1+s/\tau)$次查询达到相同容差（Kothari和O'Donnell，2023年）。

然而，均值估计中的二次加速并不会自动延伸到优化中。量子采样器将可能的差商编码在叠加态中，但直接测量其输出只能得到一次带噪声的实现。加速来自相干查询：在固定决策和扰动下，以指定精度估计均值。一个关键挑战仍然是选择估计精度，因为它既影响每次更新的量子查询成本，也影响达到目标期望目标间隙所需的更新次数。在搜索初期采用高精度，可能会耗费查询，却几乎无助于最终解；而在后期精度不足，则可能阻碍收敛。缩小扰动会使这种分配更加复杂，因为这虽然能减小有限差分偏差，却可能增大差商的方差，进而提高其估计成本。要使总优化查询复杂度获得二次改进，就需要联合选择估计容差和扰动大小，并控制由此产生的误差对后续决策的影响。

为应对这一挑战，我们提出耦合评估量子加速（QACE），这是一种基于梯度的算法，通过协调均值估计精度与不断缩小的扰动，使总查询复杂度实现平方级改进。在每次迭代中，我们根据配对输出构造有限差分，并估计其条件均值。我们的常步长分析表明，早期估计误差对目标间隙界的影响会在后续更新中逐渐减小。因此，我们可以在搜索早期使用较粗略的估计，并随着扰动缩小而逐步收紧估计容差。由此得到的调度方案使期望目标间隙呈几何递减，并使每次更新的查询界呈几何递增。在所有更新中对这些查询界求和后，在不同耦合阶数下，总查询复杂度对精度的依赖仍保持平方级改进。

1.1. 相关工作

随机逼近（SA）为通过仿真估计的广泛目标函数提供了基于递归梯度的方法（Kushner和Yin，2003；Borkar，2009）。随机逼近方法已被用于分位数优化（Hu 等人，2022，2025a）、条件风险价值优化（Hu 等人，2025b；Wang et al.，2025）、扭曲风险度量优化（Jiang等人，2025；Cao等人，2026b）、无似然参数估计（Li等人，2026b）以及动态系统分析（Jiang等人，2023；Cao等人，2025b；Chen等人，2026；Li等人，2026a；Che等人，2026；Wang et al.，2026）。最近，Wang和Zhang（2026年）将随机化双点扰动应用于大语言模型驱动的多智能体仿真中的优化。共同随机数

（共同随机数）可以减少配对差值的方差，并提高随机逼近的收敛速率（Kleinman 等人，1999）。He 等人（2024）为随机逼近开发了自适应重要性采样，但其方法需要似然比，而我们的噪声输出预言机无法提供这些似然比。

要确定量子估计带来的收益，必须将各次迭代的收敛界与每次更新所需的仿真开销联系起来。Hu 和 Fu（2025年）推导了有限时间误差界，允许有限差分偏差和噪声随着扰动缩小而增大。他们的界描述了给定偏差和噪声序列时的收敛性，但并未确定如何分配量子估计所需的查询。我们联合选择估计容差和扰动大小，以保持几何收敛性，并将整个搜索过程中不断增加的查询成本纳入考量。

要进一步证明这种改进是最优的，还需要考虑耦合和相干量子访问的下界。对于独立的含噪函数评估，Yu 等人（2024年）利用函数值的批量估计刻画了经典极小极大精度指数。他们的下界针对独立的经典观测，因此既不能约束相干量子查询，也无法确定配对输出存在耦合时的最优精度指数。我们构造了配对模拟器，在相同耦合条件下为经典和量子精度指数给出匹配的下界。这些界确立了最优的二次分离，即使经典方法利用耦合，并自适应地选择查询位置和采样开销，结论仍然成立。

基于梯度或具有一致精度的函数预言机得出的量子优化结果，并不能直接确定如何利用含噪仿真输出进行优化的成本。Sidford and Zhang（2023年）提出的随机梯度方法假设梯度估计无偏。相比之下，根据仿真输出计算的有限差分会引入偏差，而且随着扰动幅度缩小，方差可能递增。相关的零阶方法则假设可以访问具有一致精度的目标函数近似（Kim 等人，2025年；Gong 等人，2025年）。配对差值方差的上界并不能提供这些方法所需的一致近似误差，而达到规定精度本身也会消耗仿真查询。为在含噪输出访问条件下建立查询复杂度界，我们分析配对仿真差值，并核算其在所有更新中的估计成本。

不连续的仿真响应带来了另一项障碍：光滑的期望目标函数未必具有 Lipschitz 样本响应，也未必具有无偏样本梯度。Kleinman 等人（1999）的 CRN 收敛分析假设样本路径导数有界，而这一假设在我们的设置中并非必需。Liu 等人（2024）在逐样本 Lipschitz 连续性的条件下，将量子均值估计应用于共享输入下的两点差分。该条件可确保扰动缩小时差商仍有界。然而，对于不连续事件输出，差商的方差可能会随着扰动缩小而增大。

差商的方差可能会随着扰动缩小而增大。Ozgul 等人（2026）的零阶梯度估计结果假设样本路径梯度无偏且方差一致有界。但对于不连续事件响应，即使期望目标函数的梯度非零，样本路径梯度也可能几乎必然（a.s.）为零。我们直接研究配对差商的方差，并协调其增长与有限差分偏差和估计精度之间的关系。由此，我们得以证明，在不同耦合阶数下，总查询复杂度均可获得二次改进，其中也包括具有不连续输出的仿真。

当仿真仅能在可行决策处进行时，投影迭代点并不能确保有限差分查询仍然可行。Lam 和 Zhang（2025年）针对概率单纯形，使用狄利克雷扰动解决了这一问题。为适应更一般的凸可行集，我们通过可行线段上的差值重构梯度。我们的方法适用于低维可行集和边界最优解，同时保持查询复杂度上的二次量子加速。

1.2. 主要贡献

我们的主要贡献是证明，在耦合阶数的整个范围内，优化查询复杂度对精度的依赖均可实现二次改进。对于独立输出，经典方法经过 N 次查询后，可达到量级为 $\mathcal{O}(N^{-2/3})$的期望目标间隙；QACE则将这一速率提升至 $\mathcal{O}(N^{-4/3})$。

即使经典估计已受益于强耦合，这种二次改进依然成立。我们用耦合阶数$\kappa \in [0,1]$刻画两种方法共有的方差缩减，其中 $\kappa = 0$ 对应独立输出，较大的取值则表示更强的方差缩减。作为一个特例， $\kappa = 1$ 涵盖了对配对差值具有二次方差界的共同随机数耦合。要使期望目标间隙不超过 ε，采用中心化扰动的经典配对平均需要 $\mathcal{O}(d^2\varepsilon^{-(3-\kappa)/2})$次查询，其中 d 为决策维度。QACE将这一界降至 $\mathcal{O}(d^{3/2}\varepsilon^{-(3-\kappa)/4})$。两种界均由相同的 κ 决定，而量子均值估计在整个范围 $\kappa \in [0,1]$内将 ε^{-1} 的幂次减半。等价地，对于固定的 d，这些界分别给出目标间隙速率 $\mathcal{O}(N^{-2/(3-\kappa)})$和 $\mathcal{O}(N^{-4/(3-\kappa)})$。

我们的第二项贡献是证明，在决策维度固定时，这些精度指数具有极小极大最优性，表明自适应经典采样无法弥补量子带来的改进。我们在相同的耦合条件下证明了匹配的经典下界和量子下界，并允许采用自适应查询位置和批量大小。表 1 汇总了这些界。

我们的第三项贡献是在每次仿真查询都必须可行时（包括最优解位于边界的情况），仍能保持相同的二次改进。我们通过重构

表 1 在相同耦合阶数 κ 下，达到目标期望目标间隙 $\varepsilon > 0$ 所需的仿真查询界。

方法	上界	下界
经典配对平均	$\mathcal{O}(d^2\varepsilon^{-(3-\kappa)/2})$	$\Omega(d^{(1+\kappa)/2}\varepsilon^{-(3-\kappa)/2})$
QACE	$\mathcal{O}(d^{3/2}\varepsilon^{-(3-\kappa)/4})$	$\Omega(d^{(3+\kappa)/4}\varepsilon^{-(3-\kappa)/4})$

可行差分来估计梯度。我们还证明了在递减步长下的收敛性，并给出了确定性查询预算下的保证。

本文其余部分安排如下。第 2 节介绍优化问题和量子均值估计，并构建基于梯度的算法。第 3 节和第 4 节分别建立收敛性界和查询复杂度界。第 5 节给出数值结果，第 6 节作出总结。

2. 基于梯度的优化与量子采样

我们研究如下仿真优化问题 $\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$，其中 $f(\mathbf{x}) := \mathbb{E}[Y(\mathbf{x}\ \xi)]$ 表示目标函数，$\mathcal{X} \subseteq \mathbb{R}^d$ 表示一个非空、闭且凸的可行域。$\mathbf{x}$表示决策，ξ 包含一次仿真实验中的不确定性，并假设$\mathbb{E}[|Y(\mathbf{x}\ \xi)|] < \infty$。为简化记号，我们记 $\mathbf{x}_* := \operatorname{argmin}_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$，并记 $f_* := f(\mathbf{x}_*)$。

基于梯度的搜索利用 f 的局部敏感性来确定如何改进当前决策。当梯度无法通过解析方式获得时，我们考虑如下更新

$$\mathbf{x}_{k+1} = \Pi_{\mathcal{X}}(\mathbf{x}_k - \eta_k \mathbf{g}_k), \quad k = 0, 1, \ldots,$$

其中，$\mathbf{g}_k$ 表示 $\nabla f(\mathbf{x}_k)$ 的估计值，$\eta_k > 0$ 为步长，$\Pi_{\mathcal{X}}$ 表示欧几里得投影，可防止迭代点离开可行集 $\mathcal{X}$。主要任务是根据仿真输出构造$\mathbf{g}_k$。有限差分通过计算相邻决策 $\mathbf{x}$ 和 $\mathbf{x}'$处的差值 $f(\mathbf{x}) - f(\mathbf{x}')$ 来近似梯度，因此，构造$\mathbf{g}_k$ 需要估计配对仿真差值的均值，而量子采样可以降低其查询成本。

2.1. 量子均值估计

为了估计差值 $f(\mathbf{x}) - f(\mathbf{x}')$，我们生成具有指定边缘分布的配对输出 $Y(\mathbf{x}\ \xi)$ 和 $Y(\mathbf{x}',\ \xi')$，使得 $\mathbb{E}[Y(\mathbf{x}\ \xi) - Y(\mathbf{x}',\ \xi')] = f(\mathbf{x}) - f(\mathbf{x}')$。对输入进行耦合可以降低该差值的方差。例如，CRN 施加 $\xi' = \xi$。即使采用耦合，经典平均也可能需要进行许多次查询才能达到较小的估计误差。为了降低这一成本，我们对相同的配对差值采用量子均值估计。

具体而言，量子均值估计使用一个可逆标量采样器 $\mathcal{A}$，其输出为 Y，并给定一个用于界定其标准差的常数 $s \geq 0$，以及一个均方根误差容限 $\tau > 0$。我们假设可以相干

访问可逆采样器 $\mathcal{A}$、其逆操作以及受控查询。我们用 QME$[\mathcal{A}, s, \tau]$ 表示对该采样器输出均值的估计。经典平均使用 $\max\{1, \lceil s^2/\tau^2 \rceil\}$ 个独立输出即可达到相同的容差。我们用 $\mathsf{q}(\widehat{\mu})$ 表示生成 $\widehat{\mu}$ 所用的采样器查询次数。下面的引理 1 给出了梯度估计所需的均方根误差控制。其证明将具有常数成功概率的均值估计转换为均方根误差保证，并通过随机望远镜求和实现无偏性（Kothari和O'Donnell，2023年；Rhee和Glynn，2015年；Sidford and Zhang，2023年）。随后，我们在第2.2节和第2.3节中使用这些量子均值估计来构造梯度更新。

LEMMA 1. 存在一个满足 $\widehat{\mu}\ \mathbb{E}[(\widehat{\mu}-\mathbb{E}[Y])^2] \le \tau^2$ 的估计量。对于某个常数 $C_1 > 0$，无偏实现满足 $\mathbb{E}[\widehat{\mu}] = \mathbb{E}[Y]$，且 $\mathbb{E}[\mathsf{q}(\widehat{\mu})] \le C_1(1 + s/\tau)$。采用确定性查询上限的实现满足 $\mathsf{q}(\widehat{\mu}) \le C_1(1 + s/\tau)$ *a.s.*。

2.2. 中心化扰动

为了仅使用两个仿真位置估计所有梯度分量，我们采用同步扰动。我们首先考虑仿真查询可能位于 $\mathcal{X}$ 之外的情形，此时模拟器定义在扰动后的位置。我们假设 $\mathcal{X}$ 具有非空内部，且 $\mathbf{x}_* \in \operatorname{int}\mathcal{X}$，如 Hu 和 Fu（2025年）所述。在 $\mathbf{x} \in \mathcal{X}$ 处，我们抽取一个随机方向 $\boldsymbol{\Delta} := (\Delta_1, \ldots\ldots, \Delta_d)^\top$，该方向独立于 $\mathbf{x}$ 和过往观测，其分量相互独立并满足 $\mathbb{P}(\Delta_j = 1) = \mathbb{P}(\Delta_j = -1) = 1/2$。我们使用这些随机方向的独立序列 $\{\boldsymbol{\Delta}_k\}$，其生成过程独立于模拟器的随机化。沿 $\boldsymbol{\Delta}/\sqrt{d}$ 的方向导数通过中心化仿真差分的均值进行近似，

$$\begin{aligned} \mu_c(\mathbf{x}, \boldsymbol{\Delta}) &:= \mathbb{E}\Big[\frac{Y(\mathbf{x} + c\boldsymbol{\Delta}/\sqrt{d}, \xi) - Y(\mathbf{x} - c\boldsymbol{\Delta}/\sqrt{d}, \xi')}{2c}\,\Big|\,\mathbf{x}, \boldsymbol{\Delta}\Big] \\ &= \frac{f(\mathbf{x} + c\boldsymbol{\Delta}/\sqrt{d}) - f(\mathbf{x} - c\boldsymbol{\Delta}/\sqrt{d})}{2c}, \end{aligned} \tag{1}$$

其中，$c \in (0, 1]$ 表示扰动大小，$\mathbf{x} \pm c\boldsymbol{\Delta}/\sqrt{d}$ 表示经过扰动的仿真输入。独立符号确保当 c 减小时，$\mathbb{E}[\sqrt{d}\mu_c(\mathbf{x}\ \boldsymbol{\Delta})\boldsymbol{\Delta}]$ 能够逼近 $\nabla f(\mathbf{x})$。为估计 $\mu_c(\mathbf{x}\ \boldsymbol{\Delta})$，令 $\mathcal{A}(\mathbf{x}\ \boldsymbol{\Delta}, c)$ 使用一次成对的模拟器查询和可逆算术，在期望式（1）中制备随机商。先形成差值再进行均值估计，可以保留耦合带来的方差缩减。估计值 $\widehat{\mu} =$ 量子均值估计$[\mathcal{A}(\mathbf{x}\ \boldsymbol{\Delta}, c)\ s, \tau]$ 的均方根误差至多为 τ。在下方的算法 1 中，我们说明如何应用这些估计来更新决策并求解优化问题。

如果用一个观测到的对来代替均值估计，就会得到 $\mathbf{g}_k = \sqrt{d}[Y(\mathbf{x}_k + c_k\boldsymbol{\Delta}_k/\sqrt{d},\ \xi) - Y(\mathbf{x}_k - c_k\boldsymbol{\Delta}_k/\sqrt{d},\ \xi')]\boldsymbol{\Delta}_k/(2c_k)$，也就是通常的双边SPSA估计。在算法 1 中，QACE 则将 $\mu_{c_k}(\mathbf{x}_k, \boldsymbol{\Delta}_k)$ 估计到容差 τ_k，其中减小 τ_k 可能会增加仿真

算法 1 采用中心化扰动的 QACE。

1: **输入** 初始决策 $\mathbf{x}_0 \in \mathcal{X}$；步长和扰动幅度序列 $\{\eta_k\}$ 以及 $\{c_k\}$；估计容差 $\{\tau_k\}$；标准差界 $\{s_k\}$；随机方向 $\{\mathbf{\Delta}_k\}$。

2: **迭代点** 对于 $k = 1$、2、.....、K。

3: **均值估计**

$$\widehat{\mu}_{k-1} = \mathrm{QME}[\mathcal{A}(\mathbf{x}_{k-1}, \mathbf{\Delta}_{k-1}, c_{k-1}), s_{k-1}, \tau_{k-1}].$$

4: **梯度估计和决策更新**

$$\mathbf{g}_{k-1} = \sqrt{d}\widehat{\mu}_{k-1}\mathbf{\Delta}_{k-1} \quad \text{and} \quad \mathbf{x}_k = \Pi_{\mathcal{X}}(\mathbf{x}_{k-1} - \eta_{k-1}\mathbf{g}_{k-1}). \tag{2}$$

5: **返回** $\mathbf{x}_K$。

每次迭代所需的工作量。经典平均需要逐渐增大的成对输出批次，而量子均值估计只需较少的相干查询即可达到相同的容差。我们在第 3.1 节证明，计入所有更新中不断增加的估计成本后，均值估计的二次量子优势仍然适用于总查询复杂度。

2.3. 可行重构

当模拟器只能在 $\mathcal{X}$ 上接受查询时，式 (1) 中的中心化扰动在边界附近可能无法使用。为确保每次查询都可行，我们将当前决策与可行线段上的点进行比较，并根据这些方向估计重构梯度。我们令 $\mathcal{X}$ 为紧集，并允许存在低维集合和边界最优解。定义 $\mathcal{T} := \mathrm{span}\{\mathbf{x} - \mathbf{x}' : \mathbf{x}, \mathbf{x}' \in \mathcal{X}\}$，其中维数为 $\mathcal{T} \geq 1$；令 $\mathbf{P}$ 为到 $\mathcal{T}$ 的正交投影。我们分别用 $\nabla_{\mathcal{T}} f$ 和 $\nabla^2_{\mathcal{T}} f$ 表示沿此空间的梯度和海森矩阵。

我们首先选取可用于重构梯度的可行方向。在$\mathbf{x}$处，我们选取锚点$\mathbf{x}^{[1]}, \ldots, \mathbf{x}^{[m]} \in \mathcal{X}$，其中 $m \geq$维数为 $\mathcal{T}$，并定义$\mathbf{w}_i := \mathbf{x}^{[i]} - \mathbf{x}$和$\mathbf{M}_{\mathbf{x}} := \sum_{i=1}^{m} \mathbf{w}_i \mathbf{w}_i^\top$。对于某个统一常数 $\lambda > 0$，我们要求在 $\mathcal{T}$ 上满足$\mathbf{M}_{\mathbf{x}} \succeq \lambda \mathbf{P}$。重构向量$\mathbf{s}_i := \mathbf{M}_{\mathbf{x}}^{\dagger} \mathbf{w}_i$，其中 $\dagger$表示摩尔–彭若斯逆，则满足 $\sum_i \mathbf{s}_i \mathbf{w}_i^\top = \mathbf{P}$ 和 $\sum_i \mathbf{s}_i \langle \nabla_{\mathcal{T}} f(\mathbf{x}), \mathbf{w}_i \rangle = \nabla_{\mathcal{T}} f(\mathbf{x})$。若锚点固定且张成$\{\mathbf{x}^{[i]} - \mathbf{x}^{[1]} : i = 2, \ldots, m\} = \mathcal{T}$，则对某个 $\lambda > 0$，条件 $\mathbf{M}\mathbf{x} \succeq \lambda \mathbf{P}$ 对所有 $\mathbf{x} \in \mathcal{X}$ 一致地成立。作为一个具体例子，考虑概率单纯形 $\mathcal{S}_d := \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{x} \geq \mathbf{0},\ \mathbf{1}^\top \mathbf{x} = 1\}$，其中 $d \geq 2$，而 $\mathbf{1}$ 表示全 1 向量。我们可以使用坐标单位向量 $\mathbf{e}_i$ 作为锚点，从而得到 $m = d$, $\dim \mathcal{T} = d - 1$ 和 $\lambda = 1$。

接下来，我们仅使用可行查询来估计这些方向导数，其偏差阶为 c^2。对于 $0 < t \leq 1$，单侧展开 $[f(\mathbf{x} + t\mathbf{w}_i) - f(\mathbf{x})]/t = \langle \nabla_{\mathcal{T}} f(\mathbf{x}),\ \mathbf{w}_i \rangle + (t/2)\mathbf{w}_i^\top \nabla^2_{\mathcal{T}} f(\mathbf{x}) \mathbf{w}_i +$

$\mathcal{O}(t^2)$ 包含一个关于 t 的线性项，我们通过理查森外推将其消除。对于 $\mathbf{w}_i \neq \mathbf{0}$，我们令 $\mathcal{A}^F(\mathbf{x}, \mathbf{w}_i, c)$ 进行准备

$$\frac{4Y(\mathbf{x}+c\mathbf{w}_i/2,\xi_1) - 4Y(\mathbf{x},\xi_1') - Y(\mathbf{x}+c\mathbf{w}_i,\xi_2) + Y(\mathbf{x},\xi_2')}{c}, \tag{3}$$

其中，数对 (ξ_1,ξ_1') 和 (ξ_2,ξ_2') 相互独立，且每对内部保留耦合。该采样器使用两次成对查询和可逆算术。使用无偏实现和这些采样器共同的标准差界 s^F，我们独立计算 $\widehat{\mu}_i^F$ =量子均值估计$[\mathcal{A}^F(\mathbf{x}, \mathbf{w}_i, c)$ $s^F, \tau\sqrt{\lambda}/$维。满足以下条件：$\mathbb{E}[\widehat{\mu}_i^F] = [4f(\mathbf{x}+c\mathbf{w}_i/2) - f(\mathbf{x}+c\mathbf{w}_i) - 3f(\mathbf{x})]/c$以及 $\mathbb{E}[\widehat{\mu}_i^F] - \langle \nabla_{\mathcal{T}} f(\mathbf{x}), \mathbf{w}_i\rangle = \mathcal{O}(c^2)$。对于截断估计，我们使用标量容差$\tau\sqrt{\lambda}/m$。如果 $\mathbf{w}_i = \mathbf{0}$，则无需查询模拟器即可设定 $\widehat{\mu}_i^F = 0$。

——

通过将算法 1 中的中心差分替换为沿可行线段的方向估计，我们得到式（4）中的可行重构更新：

$$\widetilde{\mathbf{g}}_{k-1} = \sum\nolimits_{i=1}^{m} \mathbf{s}_{k-1,i}\widehat{\mu}_{k-1,i}^F \quad \text{and} \quad \mathbf{x}_k = \Pi_{\mathcal{X}}(\mathbf{x}_{k-1} - \eta_{k-1}\widetilde{\mathbf{g}}_{k-1}), \tag{4}$$

其中，下标 $k-1$ within $\widehat{\mu}_{k-1,i}^F$ 表示在 $\mathbf{x}_{k-1}$ 处利用 c_{k-1} 和 τ_{k-1}形成的量。投影和凸性意味着 $\mathbf{x}_k \in \mathcal{X}$ 以及 $\mathbf{x}_k + t\mathbf{w}_{k,i} = (1-t)\mathbf{x}_k + t\mathbf{x}_k^{[i]} \in \mathcal{X}$ 对每个 $k \geq 0$、$i = 1, \ldots, m$和 $t \in \{0, c_k/2, c_k\}$均成立。因此，每次仿真查询都是可行的，包括由边界决策发起的查询。

3. 收敛速率分析

我们利用引理 1 中的均值估计保证，说明提高精度如何在相同耦合条件下保持常步长收敛性并降低总查询成本。全文中，对于正标量函数 a 和 b，我们使用 $a = \mathcal{O}(b)$ 和 $a = \Omega(b)$ 分别表示 $a \leq Cb$ 和 $a \geq cb$，其中常数 $C, c > 0$ 与所述渐近状态下变化的参数无关。

Assumption 1. 在包含 $\mathcal{X}$ 以及该过程所使用的所有仿真输入的开凸邻域上，有 $f \in C^2$，并且对于所有$\mathbf{x}$、$\mathbf{x}'$，

$$f(\mathbf{x}') \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{x}' - \mathbf{x}\rangle + \frac{\rho}{2}\|\mathbf{x}' - \mathbf{x}\|^2,$$
$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}')\| \leq L_1\|\mathbf{x} - \mathbf{x}'\|, \quad \textit{and} \quad \|\nabla^2 f(\mathbf{x}) - \nabla^2 f(\mathbf{x}')\| \leq L_2\|\mathbf{x} - \mathbf{x}'\|,$$

对于常数 $0 < \rho \leq L_1 < \infty$ 和 $0 < L_2 < \infty$。对于低维集合上的可行重构(4)，邻域和导数均取自其仿射包。

Assumption 2. 对于不同的仿真输入$\mathbf{x}$、$\mathbf{x}'$，成对输入具有固定的耦合，并满足$\mathrm{Var}[Y(\mathbf{x}\ \xi) - Y(\mathbf{x}'、\ \xi')] \leq \sigma^2\|\mathbf{x} - \mathbf{x}'\|^{2\kappa}$，其中常数为 $\sigma \geq 0$和 $\kappa \in [0,1]$。

耦合阶数 κ 是配对仿真实验的一项属性，决定了经典方法和量子均值估计可利用的方差。具有 $\sup_{\mathbf{x}} \mathrm{Var}[Y(\mathbf{x}\ \xi)] \le \sigma^2/2$ 的独立输出满足假设 2，其中 $\kappa = 0$。满足均方 Lipschitz 条件的中心化输出满足 $\mathbb{E}[|(Y(\mathbf{x}\ \xi) - f(\mathbf{x})) - (Y(\mathbf{x}', \xi') - f\mathbf{x}'))|^2] \le \sigma^2\|\mathbf{x} - \mathbf{x}'\|^2$，对应于$\kappa = 1$。

即使对单个仿真响应求导无法恢复其期望的梯度，耦合差分仍能保留有用的方差缩减。例如，在 $\{\mathbf{x} : \|\mathbf{x}\| \le 1\}$ 上，令 $p(\mathbf{x}) := 1/4 + \|\mathbf{x}\|^2/8 + x_1^3/48$ 和 $Y(\mathbf{x}\ \xi) := \mathbf{1}\{\xi \le p(\mathbf{x})\}$，其中 $\xi \sim \mathrm{Unif}[0,1]$。于是，$f = p \in C^\infty$ 且 $\nabla^2 f(\mathbf{x}) \succeq \mathbf{I}/8$，其中 $\mathbf{I}$ 表示单位矩阵。采用共享输入时，有 $\mathrm{Var}[Y(\mathbf{x}\ \xi) - Y(\mathbf{x}', \xi)] \le |p(\mathbf{x}) - p(\mathbf{x}')| \le 5\|\mathbf{x} - \mathbf{x}'\|/16$，因此假设 2 在 $\kappa = 1/2$ 下成立。尽管对于每个固定的 $\mathbf{x}$，都有 $\nabla_{\mathbf{x}} Y(\mathbf{x}\ \xi) = \mathbf{0}$ a.s.，但当 $\mathbf{x} \ne \mathbf{x}_*$ 时，我们仍有 $\nabla f(\mathbf{x}) \ne \mathbf{0}$。因此，在此例中，对路径导数取平均无法恢复目标梯度，而配对差值仍能从耦合中获益。

在假设 2 下，我们可以在算法 1 中使用 $s_k = 2^{\kappa-1}\sigma c_k^{\kappa-1}$。对于可行重构 (4)，令 $R \ge \sup_{\mathbf{x}, \mathbf{x}' \in \mathcal{X}} \|\mathbf{x} - \mathbf{x}'\|$ 并使用 $s^F = \sigma R^\kappa \sqrt{1 + 2^{4-2\kappa}}\, c^{\kappa-1}$。对于固定的初始决策 $\mathbf{x}_0$，令$\mathcal{F}_k$ 包含前 k 个已完成更新所生成的信息，其中

$$\mathcal{F}_k := \begin{cases} \sigma\{\mathbf{x}_0, \mathbf{\Delta}_j, \widehat{\mu}_j : 0 \le j < k\}, & \text{centered perturbations,} \\ \sigma\{\mathbf{x}_0, \mathbf{x}_j^{[i]}, \widehat{\mu}_{j,i}^F : 0 \le j < k,\ 1 \le i \le m\}, & \text{feasible reconstruction.} \end{cases}$$

两种量子实现都满足 $\mathbb{E}[(\widehat{\mu}_k - \mu_{c_k}(\mathbf{x}_k,\ \mathbf{\Delta}_k))^2 \mid \mathcal{F}_k,\ \mathbf{\Delta}_k] \le \tau_k^2$。对于可行重构 (4)，给定 $\mathcal{F}_k$，$\mathbf{x}_k^{[1]}$，……，$\mathbf{x}_k^{[m]}$，我们使用条件独立的标量估计。

3.1. 常步长

令 N_k^Q 和 N_k^C 分别统计在 k 个已完成更新中使用的所有量子和经典模拟器查询。我们采用几何递减的扰动大小和估计容差，使得常步长能够持续减小目标差距。令 $\varepsilon_0 > 0$ 为常数，并令 $\mathcal{C}_2$、$\mathcal{C}_3$ 表示附录 A 中引理3的常数。在第 3.1 节中，对于中心化扰动 (2)，取 $\eta = 1/(L_1 d)$，选择 ε_0 足够大，使得 $\varepsilon_0 \ge (L_1/\rho)\mathbb{E}[f(\mathbf{x}_0) - f_*]$，并设 $c_k = \min\{1, [\rho\varepsilon_0(1 - \rho\eta/2)^k/(8\mathcal{C}_2 L_2^2)]^{1/4}\}$ 和 $\tau_k = \sqrt{\rho\varepsilon_0}(1 - \rho\eta/2)^k/(8\mathcal{C}_3 d)$。对于可行重构 (4)，取 $\eta = 1/(2L_1)$、$\varepsilon_0 \ge \mathbb{E}[f(\mathbf{x}_0) - f_*]$、$c_k = \min\{1, [\rho\lambda\varepsilon_0(1 - \rho\eta/2)^k/(8\mathcal{C}_2 m L_2^2 R^6)]^{1/4}\}$，以及 $\tau_k = \sqrt{\rho\varepsilon_0}(1 - \rho\eta/2)^k/(8\mathcal{C}_3)$。

THEOREM 1. 在假设 *1* 和 *2* 下，过程 (2) 或 (4) 均满足 $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}_*$几乎必然。对于每个 $0 < \varepsilon < \varepsilon_0$，都存在一个整数 k_ε，使得 $\mathbb{E}[f(\mathbf{x}^{k_\varepsilon}) - f_*] \le \varepsilon$。中心化扰动(2)达到

$$\mathbb{E}[N_{k_\varepsilon}^Q] = \mathcal{O}(d^{3/2}\varepsilon^{-(3-\kappa)/4}) \quad \textit{and} \quad N_{k_\varepsilon}^C = \mathcal{O}(d^2\varepsilon^{-(3-\kappa)/2}) \quad \textit{a.s.} \tag{5}$$

可行重构(4) 保留了这些精度指数。在概率单纯形 $\mathcal{S}_d$上，顶点锚点得到

$$\mathbb{E}[N_{k_\varepsilon}^{Q}] = \mathcal{O}(d^{(7-\kappa)/4}\varepsilon^{-(3-\kappa)/4}) \quad and \quad N_{k_\varepsilon}^{C} = \mathcal{O}(d^{(5-\kappa)/2}\varepsilon^{-(3-\kappa)/2}) \quad a.s. \tag{6}$$

采用确定性查询上限的实现中，量子界几乎必然成立。

定理*1*的证明。令 $\mathcal{E}_k := \mathbb{E}[f(\mathbf{x}_k) - f_*]$。附录 A 中的引理2适用于中心化扰动 (2) 下的 $L_1\|\mathbf{x}_k - \mathbf{x}_*\|^2/2$，也适用于可行重构 (4) 下的 $f(\mathbf{x}_k) - f_*$。这两个量的初值均至多为 ε_0，且调度序列将期望递推中的加性误差界定为 $\rho\eta\varepsilon_0(1-\rho\eta/2)^k/4$。由于

$$(1-\rho\eta)\varepsilon_0(1-\rho\eta/2)^k + \rho\eta\varepsilon_0(1-\rho\eta/2)^k/4 \le \varepsilon_0(1-\rho\eta/2)^{k+1},$$

由归纳法和光滑性可得 $\mathcal{E}_k \le \varepsilon_0(1-\rho\eta/2)^k$。定义 $k_\varepsilon := \lceil \log(2\varepsilon_0/\varepsilon)/\log[1/(1-\rho\eta/2)] \rceil$；则 $\mathcal{E}_{k_\varepsilon} \le \varepsilon$。此外，$\sum_{k=0}^{\infty} \eta\varepsilon_0(1-\rho\eta/2)^k = 2\varepsilon_0/\rho < \infty$，因此引理3推出 $\lim_{k\to\infty}\mathbf{x}_k = \mathbf{x}_*$ 几乎必然成立。

对于 $p \in \{1,2\}$，我们使用 $c_{j+1} \le c_j$和$\tau_{j+1}/\tau_j = \sqrt{1-\rho\eta/2}$来得到

$$\sum_{j=0}^{k-1} \frac{1}{(c_j^{1-\kappa}\tau_j)^p} \le \frac{1}{[1-(1-\rho\eta/2)^{p/2}](c_{k-1}^{1-\kappa}\tau_{k-1})^p} \le \frac{4}{\rho\eta(c_{k-1}^{1-\kappa}\tau_{k-1})^p}.$$

在 $k = k_\varepsilon$ 处，我们有 $\varepsilon_0(1-\rho\eta/2)^{k-1} > \varepsilon/2$ 和 $k_\varepsilon = \mathcal{O}(\eta^{-1}\log(2\varepsilon_0/\varepsilon))$。对于中心化扰动 (2)，$\eta^{-1} = L_1 d$、$\tau_{k-1}^{-1} = \mathcal{O}(\sqrt{d}/\varepsilon)$ 和 $c_{k-1}^{-(1-\kappa)} = \mathcal{O}(\varepsilon^{-(1-\kappa)/4})$。代入附录 A 中的 (12) 即可证明 (5)。

对于可行重构 (4)，我们有 $\eta^{-1} = 2L_1$、$\tau_{k-1}^{-1} = \mathcal{O}(1/\sqrt{\varepsilon})$ 和 $c_{k-1}^{-(1-\kappa)} = \mathcal{O}(\max\{1, mR^6/(\lambda\varepsilon)\}^{(1-\kappa)/4})$。因此，命题1确立了相同的精度指数。在$\mathcal{S}_d$上，我们使用顶点锚点 $(m, \dim\ \mathcal{T}, R, \lambda) = (d, d-1, \sqrt{2}, 1)$，从而得到 (6)。对于截断估计器，将 dim–$\mathcal{T}$替换为 m 不会改变这些阶。□

改进源于提高精度所需的成本。对于固定维数，最精细的阶段使用数量级为 $\varepsilon^{1/4}$ 的 c，以及数量级为 $\varepsilon^{1/2}$ 的 τ。在相同的耦合下，两种估计器的标准差界均为 $c^{\kappa-1}$ 的数量级。因此，量子估计的成本为 $\mathcal{O}(\varepsilon^{-(3-\kappa)/4})$，而经典平均会使这种依赖关系平方化。对所有先前更新中几何级数增长的成本求和后，这些指数仍保持不变。因此，耦合改进了两种速率，而提高精度则为每个 $\kappa \in [0,1]$ 保留了额外的二次量子改进。

相同的界定还规定了在固定查询预算 N 下何时停止。我们选择一个足够大的 $\mathcal{C}_4 > 0$。对于带截断的量子估计器，选择 $k = k_{\mathcal{C}_4 N^{-4/(3-\kappa)}}$ 可确保 $N_k^Q \le N$ 几乎处处成立，并且对于所有足够大的 N，$\mathbb{E}[f(\mathbf{x}_k) - f_*] = \mathcal{O}(N^{-4/(3-\kappa)})$。对于经典配对平均，选择 $k = k_{\mathcal{C}_4 N^{-2/(3-\kappa)}}$ 可确保 $N_k^C \le N$ 几乎处处成立，并且 $\mathbb{E}[f(\mathbf{x}_k) - f_*] = \mathcal{O}(N^{-2/(3-\kappa)})$。

3.2. 递减步长

或者，我们可以设置递减的步长，以控制标量估计误差。在第 3.2 节中，令 $\eta_k = \bar{\eta}(k+k_0)^{-\alpha}$ 和 $c_k = \bar{c}(k+k_0)^{-\alpha/4}$，其中 $\alpha \in (1/2,1)$、$\bar{\eta} > 0$ 和 $0 < \bar{c} \leq 1$。这种多项式形式的步长在随机逼近文献中很常见（例如参见 Hu 和 Fu，2025年；Hu 等人，2025a；Cao等人，2026a）。选择一个足够大的整数 $k_0 \geq 1$，使得对于中心化扰动 (2)，有 $\alpha k_0^{\alpha-1} \leq \rho\bar{\eta}/2$ 和 $\bar{\eta}k_0^{-\alpha} \leq 1/(L_1 d)$；或者对于可行重构 (4)，有 $\bar{\eta}k_0^{-\alpha} \leq 1/L_1$。对于中心化扰动 (2)，我们使用引理 1 的无偏实现，并令 $\tau_k = \bar{\tau} > 0$。对于可行重构 (4)，我们令 $\tau_k = \bar{\tau}(k+k_0)^{-\alpha/2}$。

THEOREM 2. 在假设 *1* 和 *2* 下，过程(2) 或(4) 均满足 $\mathbb{E}[f(\mathbf{x}_k) - f_*] = \mathcal{O}((k+k_0)^{-\alpha})$，$\mathbb{E}[\|\mathbf{x}_k - \mathbf{x}_*\|^2] = \mathcal{O}((k+k_0)^{-\alpha})$，并且 $\lim_{k\to\infty}\mathbf{x}_k = \mathbf{x}_*$ 几乎必然成立。对于中心化扰动(2)，我们有

$$\mathbb{E}[N_k^Q] = \mathcal{O}(k^{1+\alpha(1-\kappa)/4}) \quad and \quad N_k^C = \mathcal{O}(k^{1+\alpha(1-\kappa)/2}) \quad a.s.$$

对于可行重构(4)，相应的阶分别为 $\mathcal{O}(k^{1+\alpha(3-\kappa)/4})$ 和 $\mathcal{O}(k^{1+\alpha(3-\kappa)/2})$。

作为比较，在独立噪声下，使用一对扰动的标准 SPSA 的标量误差容限为 $\tau_k = \sigma/(2c_k)$。通过 $c_k = \bar{c}(k+k_0)^{-\alpha/6}$ 平衡偏差与噪声，可得 $\sqrt{\mathbb{E}[\|\mathbf{x}_k - \mathbf{x}_*\|^2]} = \mathcal{O}((k+k_0)^{-\alpha/3})$（Hu 和 Fu，2025年）；而定理 2 对 QACE 给出了 $\sqrt{\mathbb{E}[\|\mathbf{x}_k - \mathbf{x}_*\|^2]} = \mathcal{O}((k+k_0)^{-\alpha/2})$。为计入这一改进所需的估计工作量，我们选取一个足够小的常数 $\mathcal{C}_5 > 0$。对于采用量子估计的中心化扰动 (2)，取 $k = \lfloor \mathcal{C}_5 N^{4/[4+\alpha(1-\kappa)]} \rfloor$。对于所有充分大的 N，有 $\mathbb{E}[N_k^Q] \leq N$ 且 $\mathbb{E}[f(\mathbf{x}_k) - f_*] = \mathcal{O}(N^{-4\alpha/[4+\alpha(1-\kappa)]})$。对于经典平均，取 $k = \lfloor \mathcal{C}_5 N^{2/[2+\alpha(1-\kappa)]} \rfloor$；于是，我们有 $N_k^C \leq N$ 几乎必然成立，且 $\mathbb{E}[f(\mathbf{x}_k) - f_*] = \mathcal{O}(N^{-2\alpha/[2+\alpha(1-\kappa)]})$。

4. 查询下界

现在，我们为固定的 d 建立与定理1中的精度指数相匹配的下界。现有的量子下界采用随机梯度预言机的多元均值恢复（Sidford and Zhang，2023），以及用于最佳臂识别的对手论证（Wang et al.，2021）。我们则构造耦合模拟器，其最小化器编码了未知的坐标符号。这些模拟器的空间尺度将优化误差与成对查询中的信息联系起来，从而在相同的耦合条件下，既可采用经典检验，也可运用基于态重叠的量子论证。

对于一个实例 P，我们分别用 f_P 和 $\mathbf{x}_{P,*}$ 表示其目标函数和最小化器，并记 $f_{P,*} := f_P(\mathbf{x}_{P,*})$。令 $\mathcal{N}^C$ 和 $\mathcal{N}^Q$ 分别统计任意经典方法或量子算法在返回决策前使用的所有查询次数。

THEOREM 3. 对于每个整数 $d \geq 1$ 和 $\kappa \in [0,1]$，都存在一个固定的类 $\mathcal{P}_{d,\kappa}$ 满足假设 *1* 和 *2*，且有 $\rho = 1$、$L_1 = 3$、$L_2 = 1$、$\sigma = 2$ 和 $\sup^{P \in \mathcal{P}_{d,\kappa}} [f_P(\mathbf{0}) - f_{P,*}] \leq 1$。任何返回满足 $\sup^{P \in \mathcal{P}_{d,\kappa}} \mathbb{E}_P[f_P(\hat{\mathbf{x}}) - f_{P,*}] \leq \varepsilon$ 的 $\hat{\mathbf{x}}$ 的算法，均满足

$$\sup_{P \in \mathcal{P}_{d,\kappa}} \mathbb{E}_P[\mathcal{N}^C] = \Omega(d^{(1+\kappa)/2} \varepsilon^{-(3-\kappa)/2}) \quad and \quad \sup_{P \in \mathcal{P}_{d,\kappa}} \mathbb{E}_P[\mathcal{N}^Q] = \Omega(d^{(3+\kappa)/4} \varepsilon^{-(3-\kappa)/4}),$$

对于 $0 < \varepsilon \leq \mathcal{C}_6 d^{-(1-\kappa)/(3-\kappa)}$，其中 $\mathcal{C}_6 > 0$ 是一个常数。当 $\kappa = 0$时，可以选择类别 $\mathcal{P}_{d,0}$ 在每个输出对中使用独立输入。还有一个关于 $\mathcal{S}_3$ 的类别 $\mathcal{P}_\kappa^\partial$，满足假设 *1* 和 *2*，其中 $(\rho, L_1, L_2, \sigma) = (1, 3, 1, 2)$ 和 $(\mathbf{x}_{P,*})_3 = 0$。对于 $0 < \varepsilon \leq \mathcal{C}_6$，准则 $\sup_{P \in \mathcal{P}_\kappa^\partial} \mathbb{E}_P[f_P(\hat{\mathbf{x}}) - f_{P,*}] \leq \varepsilon$ 意味着

$$\sup_{P \in \mathcal{P}_\kappa^\partial} \mathbb{E}_P[\mathcal{N}^C] = \Omega(\varepsilon^{-(3-\kappa)/2}) \quad and \quad \sup_{P \in \mathcal{P}_\kappa^\partial} \mathbb{E}_P[\mathcal{N}^Q] = \Omega(\varepsilon^{-(3-\kappa)/4}).$$

在 $\kappa = 0$，一个固定类别 $\mathcal{P}_d^G$，其参数为 $(\rho, L_1, L_2, \sigma) = (1, 3, 1, 2)$，且具有独立加性N(0，1)噪声，满足

$$\sup_{P \in \mathcal{P}_d^G} \mathbb{E}_P[\mathcal{N}^C] = \Omega(d^{3/2} \varepsilon^{-3/2}) \quad for \quad 0 < \varepsilon \leq \mathcal{C}_6 d^{-1/3},$$

当$\sup_{P \in \mathcal{P}_d^G} \mathbb{E}_P[f_P(\hat{\mathbf{x}}) - f_{P,*}] \leq \varepsilon$时。

定理3的证明。对于每个 d 和 κ，令$\mathcal{P}_{d,\kappa}$ 包含所有由已知尺度$0 < \ell \leq \min\{1/2, d^{-1/4}, (4d^{1-\kappa/2})^{-1/(3-\kappa)}\}$和未知 $\boldsymbol{\nu} \in \{-1,1\}^d$索引的实例。定义 $\gamma_\ell := d^{1-\kappa/2} \ell^{3-\kappa}$，并根据 $\mathbb{P}_{\boldsymbol{\nu}}(I = i, Z = z) = (1 + z\nu_i\gamma_\ell) / (2d)$ 绘制 $\xi = (I, Z)$，$i \in \{1, \ldots, d\}$，$z \in \{-1, 1\}$。在每个对内共享 ξ，我们设 $Y(\mathbf{x}, \xi) := \|\mathbf{x}\|^2 - \ell^\kappa d^{\kappa/2} \sin(x_I/\ell) Z$。因此，$f_{\ell,\boldsymbol{\nu}}(\mathbf{x}) = \|\mathbf{x}\|^2 - \ell^3 \sum_{i=1}^d \nu_i \sin(x_i/\ell)$。

由于 $\nabla^2 f_{\ell,\boldsymbol{\nu}}(\mathbf{x}) = \mathrm{diag}(2 + \nu_i \ell \sin(x_i/\ell))_{i=1}^d$，我们有 $3\mathbf{I}/2 \preceq \nabla^2 f_{\ell,\boldsymbol{\nu}}(\mathbf{x}) \preceq 5\mathbf{I}/2$ 以及 $\|\nabla^2 f_{\ell,\boldsymbol{\nu}}(\mathbf{x}) - \nabla^2 f_{\ell,\boldsymbol{\nu}}(\mathbf{x}')\| \leq \max_i |x_i - x_i'| \leq \|\mathbf{x} - \mathbf{x}'\|$。因此，$\rho = 1$、$L_1 = 3$ 和 $L_2 = 1$ 一致地成立。利用 $\min\{4, t^2\} \leq 4^{1-\kappa} t^{2\kappa}$, $t > 0$，我们得到

$$\begin{aligned} \mathrm{Var}[Y(\mathbf{x}, \xi) - Y(\mathbf{x}', \xi)] &\leq \ell^{2\kappa} d^{\kappa-1} \sum\nolimits_i \min\{4, (x_i - x_i')^2/\ell^2\} \\ &\leq \ell^{2\kappa} d^\kappa \min\{4, \|\mathbf{x} - \mathbf{x}'\|^2/(d\ell^2)\} \leq 4\|\mathbf{x} - \mathbf{x}'\|^{2\kappa}. \end{aligned}$$

此外，对于每个 $\mathbf{x}$，$|\mathrm{supp}\, Y(\mathbf{x}, \xi)| \leq 2d$。

令 $x_{\ell,*}$ 求解 $2x_{\ell,*} = \ell^2 \cos(x_{\ell,*}/\ell)$。对于一维目标函数，$f'_{\ell,1}(\ell^2/4) = \ell^2[1/2 - \cos(\ell/4)] < 0$ 以及 $f'_{\ell,1}(\ell^2/2) = \ell^2[1 - \cos(\ell/2)] \geq 0$。因此，严格凸性和可分离性意味着$\ell^2/4 \leq x_{\ell,*} \leq \ell^2/2$ 以及 $\mathbf{x}_* = x_{\ell,*}\boldsymbol{\nu}$。此外，$\ell^3|\sin(t/\ell)| - t^2 \leq \ell^2|t| - t^2 \leq \ell^4/4$，所以 $f(\mathbf{0}) - f_* \leq d\ell^4/4 \leq 1$。

为将优化误差与符号恢复联系起来，我们定义 $\widehat{\nu}_i := 2\mathbf{1}\{\widehat{x}_i \geq 0\} - 1$ 和 $\mathcal{H}(\widehat{\boldsymbol{\nu}}, \boldsymbol{\nu}) := \sum_i \mathbf{1}\{\widehat{\nu}_i \neq \nu_i\}$。然后，$\|\hat{\mathbf{x}} - \mathbf{x}_*\|^2 \geq x_{\ell,*}^2 \mathcal{H}(\widehat{\boldsymbol{\nu}}, \boldsymbol{\nu}) \geq \ell^4 \mathcal{H}(\widehat{\boldsymbol{\nu}}, \boldsymbol{\nu})/16$。由强凸性，我们可以选择一个常数 $C_2 > 0$，使得当 $\ell = (C_3\varepsilon/d)^{1/4}$ 且 $C_3 \geq 8/C_2$ 时，

$$f(\widehat{\mathbf{x}}) - f_* \geq C_2 \ell^4 \mathcal{H}(\widehat{\boldsymbol{\nu}}, \boldsymbol{\nu}) \quad \text{and} \quad \mathbb{E}_{\boldsymbol{\nu}}[\mathcal{H}(\widehat{\boldsymbol{\nu}}, \boldsymbol{\nu})] \leq d/8. \tag{7}$$

对于充分小的 $\mathcal{C}_6$，我们有 $(C_3\varepsilon/d)^{1/4} \le \min\{1/2, d^{-1/4}, (4d^{1-\kappa/2})^{-1/(3-\kappa)}\}$ 以及 $\gamma_\ell = (C_3\varepsilon)^{(3-\kappa)/4} d^{(1-\kappa)/4} \le 1/4$。

经典查询。 我们允许算法在每次抽样时获得 I，Z 由此它可以计算任意决策下的配对输出。定义 $\boldsymbol{\nu}^{(i)} := \boldsymbol{\nu} - 2\nu_i \mathbf{e}_i$ 和 $p_{\boldsymbol{\nu},\,i} := \mathbb{P}_{\boldsymbol{\nu}}(\widehat{\nu}_i \neq \nu_i)$。记 $\mathcal{L}_{\boldsymbol{\nu}}$ 为 $\boldsymbol{\nu}$ 下的分布律，则单次抽样的相对熵为 $\mathrm{KL}(\mathcal{L}_{\boldsymbol{\nu}}(I,\ Z) \| \mathcal{L}_{\boldsymbol{\nu}^{(i)}}(I,\ Z)) = \gamma_\ell \log[(1+\gamma_\ell)/(1-\gamma_\ell)]/d \le 4\gamma_\ell^2/d$。对于至多 n 次自适应抽样，链式法则和Pinsker不等式推出 $p_{\boldsymbol{\nu},\,i} + p_{\boldsymbol{\nu}^{(i)},\,i} \ge 1 - \sqrt{2n\gamma_\ell^2/d}$。对该不等式取平均并利用(7)，可得 $n = \Omega(d/\gamma_\ell^2)$。

量子查询。 对于 $\gamma \in [-1/4, 1/4]$，定义平面旋转和未知预言机如下

$$\mathbf{R}_\gamma := \begin{pmatrix} \sqrt{(1+\gamma)/2} & -\sqrt{(1-\gamma)/2} \\ \sqrt{(1-\gamma)/2} & \sqrt{(1+\gamma)/2} \end{pmatrix} \quad \text{and} \quad \mathcal{U}_{\boldsymbol{\nu}} := \sum\nolimits_{i=1}^{d} |i\rangle\langle i| \otimes \mathbf{R}_{\nu_i \gamma_\ell}.$$

其中 $|i\rangle$ 存储索引 i。已知操作会制备均匀索引态 $(1/\sqrt{d}) \sum_i |i\rangle$，并在应用 $\mathcal{U}_{\boldsymbol{\nu}}$ 后计算配对输出。因此，每次模拟器查询（包括逆查询和受控查询）都使用一次对应的 $\mathcal{U}_{\boldsymbol{\nu}}$ 查询。为纳入正向、逆向和非活动分支，我们授予对 $\mathcal{V}_{\boldsymbol{\nu}}$ 的访问权限：$= \mathrm{diag}(\mathbf{I}\ \mathcal{U}_{\nu}, \mathcal{U}_{\nu}^{\dagger})$。

通过保留辅助寄存器中的测量结果和随机化信息，我们用单位向量 $|\Psi_{\boldsymbol{\nu}}^{(t)}\rangle$ 表示经过 t 次查询后的算法，其中包括下一次查询之前的已知操作。定义 $\mathcal{W}_t := \sum_{\boldsymbol{\nu},\,i} \mathrm{Re}\langle \Psi_{\boldsymbol{\nu}}^{(t)} | \Psi_{\boldsymbol{\nu}^{(i)}}^{(t)} \rangle$。初始态与 $\boldsymbol{\nu}$ 无关，因此 $\mathcal{W}_0 = d2^d$。对于索引投影 $\mathbf{P}_i^{\mathrm{idx}} := \mathbf{I}_{\mathrm{control}} \otimes |i\rangle\langle i| \otimes \mathbf{I}_{\mathrm{work}}$，

$$(\mathcal{V}_{\boldsymbol{\nu}}^{\dagger} \mathcal{V}_{\boldsymbol{\nu}^{(i)}} - \mathbf{I})(\mathbf{I} - \mathbf{P}_i^{\mathrm{idx}}) = 0 \quad \text{and} \quad \|\mathcal{V}_{\boldsymbol{\nu}}^{\dagger} \mathcal{V}_{\boldsymbol{\nu}^{(i)}} - \mathbf{I}\| \le \arcsin \gamma_\ell \le 2\gamma_\ell.$$

由于 $\sum_{\boldsymbol{\nu},\,i} \|\mathbf{P}_i^{\mathrm{idx}} \Psi_{\boldsymbol{\nu}}^{(t)}\|^2 = 2^d$，柯西–施瓦茨不等式给出

$$|\mathcal{W}_{t+1} - \mathcal{W}_t| \le 2\gamma_\ell \sum\nolimits_{\boldsymbol{\nu},i} \|\mathbf{P}_i^{\mathrm{idx}} \Psi_{\boldsymbol{\nu}}^{(t)}\| \|\mathbf{P}_i^{\mathrm{idx}} \Psi_{\boldsymbol{\nu}^{(i)}}^{(t)}\| \le 2\gamma_\ell 2^d.$$

Kn自身的操作保持重叠，因为 $\langle V\Psi | V\Psi' \rangle = \langle \Psi | \Psi' \rangle$ 对于酉算子 V .

在至多进行 n 次查询后终止时，对解码比特对应子空间的投影以及柯西–施瓦茨不等式推出 $|\langle \Psi_{\boldsymbol{\nu}}^{(n)} | \Psi_{\boldsymbol{\nu}^{(i)}}^{(n)} \rangle| \le \sqrt{p_{\boldsymbol{\nu},\,i}} + \sqrt{p_{\boldsymbol{\nu}^{(i)},\,i}}$。平方根的凹性以及 (7) 推出 $\mathcal{W}_n \le 2d2^d \sqrt{(d2^d)^{-1} \sum_{\boldsymbol{\nu},\,i} p_{\boldsymbol{\nu},\,i}} \le d2^d/\sqrt{2}$。因此，$(1 - 1/\sqrt{2}) d2^d \le \mathcal{W}_0 - \mathcal{W}_n \le 2n\gamma_\ell 2^d$，且 $n = \Omega(d/\gamma_\ell)$ 。

对于随机查询次数 $\mathcal{N}$ 且上确界为 $\sup_{\boldsymbol{\nu}} \mathbb{E}_{\boldsymbol{\nu}}[\mathcal{N}] \le n$ 的情况，在 $\lceil C_4 n \rceil$ 次查询后截断，其中 $C_4 > 8$，并在截断时返回任意符号向量。马尔可夫不等式意味着 $\sup_{\boldsymbol{\nu}} \mathbb{P}_{\boldsymbol{\nu}}(\mathcal{N} > \lceil C_4 n \rceil) \le 1/C_4$，因此截断输出 $\widetilde{\boldsymbol{\nu}}$ 满足 $\sup_{\boldsymbol{\nu}} \mathbb{E}_{\boldsymbol{\nu}}[\mathcal{H}(\widetilde{\boldsymbol{\nu}}, \boldsymbol{\nu})] \le d(1/8 + 1/C_4) < d/4$。测试和重叠论证仍然适用，因为 $1 - 2(1/8 + 1/C_4) > 0$ 且 $2\sqrt{1/8 + 1/C_4} < 1$。在 $\boldsymbol{\nu}$ 上采用均匀先验时，$\mathbb{E}[\mathcal{H}(\widehat{\boldsymbol{\nu}}, \boldsymbol{\nu}) \mathbf{1}\{\mathcal{N} = 0\}] = (d/2)\mathbb{P}(\mathcal{N} = 0)$。因此，由 (7) 可得 $\mathbb{P}(\mathcal{N} \ge 1) \ge 3/4$、$n \ge 3/4$，以及 $\lceil C_4 n \rceil \le (C_4 + 4/3)n$。我们得到经典地 $n = \Omega(d/\gamma_\ell^2)$，以量子方式 $n = \Omega(d/\gamma_\ell)$。代入所选的 γ_ℓ 即可证明前两个下界。

在 $\kappa = 0$ 处，将共享输入替换为独立副本 ξ，ξ' 可保持 $f_{\ell,\nu}$ 并确保 $\mathrm{Var}[Y(\mathbf{x},\xi) - Y(\mathbf{x}',\xi')] \le 2$；随后，每个成对查询在经典情形下使用两次抽样，在量子情形下则对 $\mathcal{U}_\nu$、其逆运算或其受控版本进行两次对应查询，因此前述下界和截断论证仅会改变常数因子。

独立高斯响应。对于 $\mathcal{P}_d^G$，保留相同的目标函数，并使用 $Y(\mathbf{x},\xi) = f_{\ell,\nu}(\mathbf{x}) + \xi$，其中 $\{(\xi_t,\xi_t')\}_{t\ge1} \overset{\text{iid}}{\sim} \mathrm{N}(0,\mathrm{I}_2)$。高斯输入仍保留在模拟器内部，每次经典查询仅返回输出对。于是，$\mathrm{Var}[Y(\mathbf{x},\xi) - Y(\mathbf{x}',\xi')] = 2$ 且 $\sup_{\mathbf{x}} |f_{\ell,\nu}(\mathbf{x}) - f_{\ell,\nu^{(i)}}(\mathbf{x})| \le 2\ell^3$。对于输出对 $\mathbf{Y} = (Y(\mathbf{x},\xi),\, Y(\mathbf{x}',\xi'))$，有 $\mathrm{KL}(\mathcal{L}_\nu(\mathbf{Y}) \| \mathcal{L}_{\nu^{(i)}}(\mathbf{Y})) \le 4\ell^6$。应用链式法则和Pinsker不等式，可得 $p_{\nu,i} + p_{\nu^{(i)},i} \ge 1 - \sqrt{2n\ell^6}$，而(7)意味着 $n = \Omega(\ell^{-6}) = \Omega(d^{3/2}\varepsilon^{-3/2})$。因此，截断论证给出 $\sup_{P\in\mathcal{P}_d^G} \mathbb{E}_P[\mathcal{N}^C] = \Omega(d^{3/2}\varepsilon^{-3/2})$。

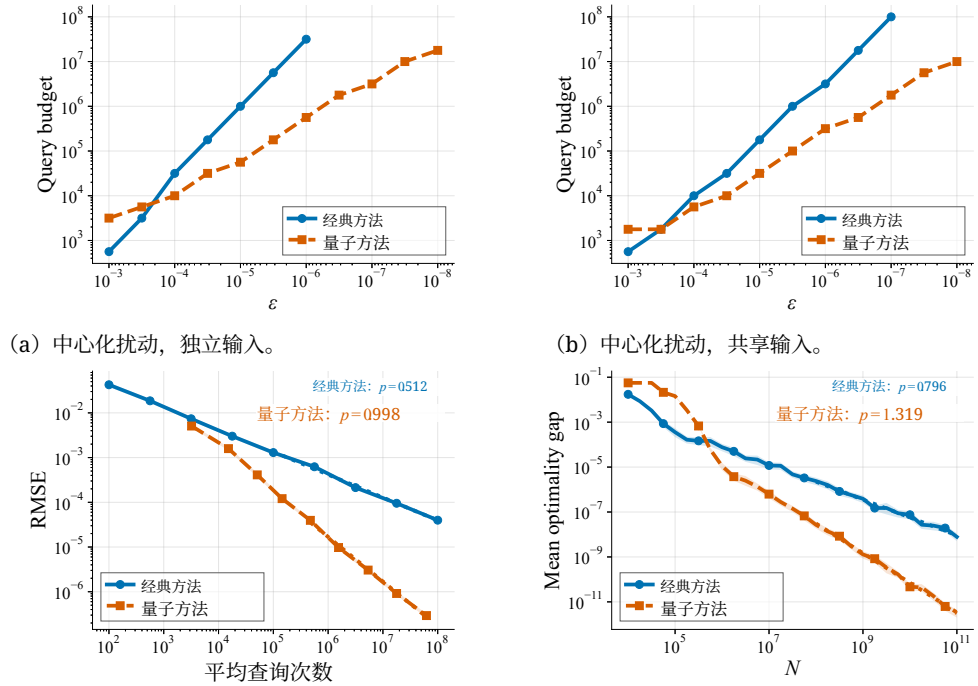
边界单纯形。对于 $\mathcal{P}_\kappa^\partial$ 在 $\mathcal{S}_3$ 上，定义 $z(\mathbf{x}) := (x_1 - x_2)/\sqrt{2}$ 和 $Y(\mathbf{x},\xi) := \|\mathbf{x}\|^2 + 2x_3 - \ell^\kappa \sin(z(\mathbf{x})/\ell)\xi$，其中 $\mathbb{P}_\nu(\xi = a) = (1 + a\nu\ell^{3-\kappa})/2$，$a,\nu \in \{-1,1\}$。由于 $|z(\mathbf{x}) - z(\mathbf{x}')| \le \|\mathbf{x} - \mathbf{x}'\|$，我们得到 $\mathrm{Var}[Y(\mathbf{x},\xi) - Y(\mathbf{x}',\xi)] \le 4\|\mathbf{x} - \mathbf{x}'\|^{2\kappa}$。此外，$f(\mathbf{x}) = 1/2 + f_{\ell,\nu}(z(\mathbf{x})) + x_3 + 3x_3^2/2$，因此 $\mathbf{x}_* = (1/2 + \nu x_{\ell,*}/\sqrt{2},\ 1/2 - \nu x_{\ell,*}/\sqrt{2},\ 0)$ 且 $f(\hat{\mathbf{x}}) - f_* \ge f_{\ell,\nu}(z(\hat{\mathbf{x}})) - f_{\ell,\nu}(\nu x_{\ell,*})$。每次单纯形查询都会在相应的 z 值处使用一个标量对以及已知算术运算。应用 $d = 1$ 界，我们得到 $\sup_{P\in\mathcal{P}_\kappa^\partial} \mathbb{E}_P[\mathcal{N}^C] = \Omega(\varepsilon^{-(3-\kappa)/2})$ 和 $\sup_{P\in\mathcal{P}_\kappa^\partial} \mathbb{E}_P[\mathcal{N}^Q] = \Omega(\varepsilon^{-(3-\kappa)/4})$。□

定理1和定理3共同确立了在决策维度固定时，经典和量子查询复杂度的极小极大最优指数分别为 $(3-\kappa)/2$ 和 $(3-\kappa)/4$。经典下界允许使用自适应评估点和批量大小。因此，在所述预言机模型中，量子优势无法通过重新分配经典仿真工作量或更改搜索规则来实现。维数界有所不同；当 $\kappa = 0$ 时，具有独立噪声的经典下界为 $\Omega(d^{3/2}\varepsilon^{-3/2})$，而上界为 $\mathcal{O}(d^2\varepsilon^{-3/2})$。

5. 数值实验

本节将考察配对差均值的估计查询成本，随后研究这种节省如何影响优化。对于 $\mathbb{R}^2$ 上的中心化扰动，我们在一个包含二次项和正弦项的目标函数上，对比 QACE 及其经典对应方法；该目标函数通过独立输入和共享输入下的不连续指示响应进行观测。对于可行性重构，我们考虑在三个通道之间分配一个单位的资源，这些通道具有指数需求、短缺事件惩罚以及二次分配成本。最优分配不会向第三个通道分配资源。量子实现采用范围归一化的迭代量子振幅估计（IQAE）（Grinko等，2021年），而其经典对应方法则在相同的决策更新中以求平均替代均值估计。电子附录验证了 IQAE 估计器满足规定的标量均方根误差容差（RMSEtolerance），这是引理2所要求的。

图 1 经典平均与量子均值估计的仿真开销和精度。



（a）中心化扰动，独立输入。

（b）中心化扰动，共享输入。

（c）配对差分均值估计。

(d) 边界最优解下的资源分配。

理论查询界使用引理1中的方差敏感估计器。量子测量结果通过经典方式进行模拟，并以预言机查询次数衡量计算工作量。

图 1(a)–1(b) 展示了达到目标均值最优性差距所需的查询次数，而图 1(c)–1(d) 分别报告了均方根误差和均值最优性差距随预言机查询次数的变化。实验设置、置信区间和查询计量详见电子附录。在均值估计实验中，经验均方根误差随着 $\bar{Q}$ （即平均查询次数）的增加而降低，其趋势近似为：经典平均下按 $\bar{Q}^{-0.512}$ 变化，IQAE 下按 $\bar{Q}^{-0.998}$ 变化。这些速率分别接近平方根倒数关系和与查询次数成反比的关系。优化实验展示了这种改进如何延续到决策更新中。对于具有独立输入的 $\mathbb{R}^2$ 上的中心化扰动 (2)，首次达到经验目标 $\varepsilon = 10^{-6}$ 的测试预算分别为 3.16×10^7 次经典查询和 5.62×10^5 次量子查询，比例约为 56。2。对于具有边界最优解的资源分配，在 $\varepsilon = 10^{-7}$ 处相应的预算分别为5.62×10^9 次经典查询和 5.62×10^7 次量子查询，实现了 100 倍的查询效率提升。

6. 结论

量子均值估计降低了光滑强凸期望目标函数的基于梯度的搜索的查询复杂度。通过将估计精度与不断缩小的扰动相协调，QACE 将每次梯度估计中的节省贯穿整个决策更新

过程。耦合在保持经典方法与量子速率之间二次分离的同时，改善了二者的速率；匹配下界则为固定决策维度确立了极小极大最优精度指数。这些下界说明了自适应经典分批为何无法复现量子增益。当每次仿真查询都必须可行时，包括在边界最优解处，可行重构仍能保持这些精度指数。

附录 A：辅助结果的证明

引理 *1* 的证明。对于有限编码，记为 $\mathcal{A}|0\rangle=\sum_z \sqrt{\mathbb{P}[Y=z]}|z\rangle|\text{辅助态}_z\rangle$，其中 $|z\rangle$ 存储编码输出，而 $|\text{辅助态}_z\rangle$ 是为保持可逆性而保留的归一化辅助态。对于 $\mathrm{Var}(Y)\le s^2$ 和 $a>0$，量子均值估计返回 $\widehat{\mu}$，满足 $\mathbb{P}(|\widehat{\mu}-\mathbb{E}[Y]|\le a)\ge 2/3$ 且 $\mathsf{q}(\widehat{\mu})\le C_1(1+s/a)$ 几乎必然成立，其中 $C_1>0$为常数。Kothari和O'Donnell（2023年）的定理 1.1 适用于采用精确可逆数值运算的有限编码输出；对于连续分布，我们使用具有此保证的理想实数运算模型。

若 $\tau\ge s$，则取 $\widehat{\mu}=Y$；于是 $\mathbb{E}[\widehat{\mu}]=\mathbb{E}[Y]$, $\mathbb{E}[(\widehat{\mu}-\mathbb{E}[Y])^2]\le\tau^2$，且 $\mathsf{q}(\widehat{\mu})=1$。因此我们考虑 $0<\tau<s$。固定一个整数 $J\ge 1$，并令 $\widehat{\mu}^{(0)}$ 为 $4J+3$ 个 Y的独立副本的中位数。对于$t\ge 4s$，切比雪夫不等式和并集界可推出 $\mathbb{P}[|\widehat{\mu}^{(0)}-\mathbb{E}[Y]|>t]\le 2^{4J+3}(s/t)^{4J+4}$。对该尾界进行积分，我们得到：对于某个足够大的常数 $C_1>0$，$\mathbb{E}[(\widehat{\mu}^{(0)}-\mathbb{E}[Y])^2\mathbf{1}\{|\widehat{\mu}^{(0)}-\mathbb{E}[Y]|>4s\}]\le C_1 s^2 2^{-4J}$。

对于 $j=1,\dots,J$，令 T_j 为独立量子估计的中位数，并选择这些估计使得 $\mathbb{P}[|T_j-\mathbb{E}[Y]|>4s\,2^{-j}]\le 2^{-4(J-j+1)}$ 且 $\mathsf{q}(T_j)\le C_1(1+2^j)(J-j+1)$。定义 $\widehat{\mu}^{(j)}=\Pi_{[\widehat{\mu}^{(j-1)}-4s\,2^{1-j},\widehat{\mu}^{(j-1)}+4s\,2^{1-j}]}(T_j)$。若 $|\widehat{\mu}^{(j-1)}-\mathbb{E}[Y]|\le 4s\,2^{1-j}$ 且 $|T_j-\mathbb{E}[Y]|\le 4s\,2^{-j}$，则 $|\widehat{\mu}^{(j)}-\mathbb{E}[Y]|\le 4s\,2^{-j}$。若 $|\widehat{\mu}^{(0)}-\mathbb{E}[Y]|\le 4s$，则第一个满足 $|T_j-\mathbb{E}[Y]|>4s\,2^{-j}$ 的 j 意味着 $|\widehat{\mu}^{(J)}-\mathbb{E}[Y]|\le 4s\,2^{1-j}+\sum_{t=j}^{J}4s\,2^{1-t}<12s\,2^{1-j}$；否则，$|\widehat{\mu}^{(J)}-\mathbb{E}[Y]|\le|\widehat{\mu}^{(0)}-\mathbb{E}[Y]|+8s\le 3|\widehat{\mu}^{(0)}-\mathbb{E}[Y]|$。按照首次失败的情况进行划分，可得

$$\mathbb{E}[(\widehat{\mu}^{(J)}-\mathbb{E}[Y])^2]\le 16s^2 2^{-2J}+9\sum_{j=1}^{J}(4s\,2^{1-j})^2\,2^{-4(J-j+1)}+C_1 s^2 2^{-4J}\le C_1 s^2 2^{-2J}.$$

此外，$\mathsf{q}(\widehat{\mu}^{(J)})\le C_1[J+\sum_{j=1}^{J}(1+2^j)(J-j+1)]\le C_2 2^J$。因此，我们设定 $J=\lceil\log_2(\sqrt{C_1}s/\zeta)\rceil$ 和 $\widehat{\mu}_\zeta=\widehat{\mu}^{(J)}$，从而得到 $\mathbb{E}[(\widehat{\mu}_\zeta-\mathbb{E}[Y])^2]\le\zeta^2$ 以及 $\mathsf{q}(\widehat{\mu}_\zeta)\le C_2 s/\zeta$，对于 $0<\zeta<s$。

对于待选取的常数 $C_2\ge 1$，独立地抽取 $J_*\ge 1$，满足 $\mathbb{P}[J_*=j]=2/3^j$。利用独立估计，令 $W_j=\widehat{\mu}_{\tau 2^{-j}/C_2}-\widehat{\mu}'_{\tau 2^{1-j}/C_2}$并返回 $\widehat{\mu}=\widehat{\mu}_{\tau/C_2}+3^{J_*}W_{J_*}/2$。矩界和 $\sum_{j\ge1}3^j4^{-j}/2=3/2$ 推出 $\sum_{j\ge1}3^j\mathbb{E}[W_j^2]/2\le C_1\tau^2/C_2^2$和 $\sum_{j\ge1}\mathbb{E}[|W_j|]<\infty$。因此，$\mathbb{E}[\widehat{\mu}]=\mathbb{E}[\widehat{\mu}_{\tau/C_2}]+\sum_{j\ge1}\mathbb{E}[W_j]=\mathbb{E}[Y]$。此外，对于 $C_2^2\ge 2(1+C_1)$，$\mathbb{E}[(\widehat{\mu}-\mathbb{E}[Y])^2]\le 2\tau^2/C_2^2+\sum_{j\ge1}3^j\mathbb{E}[W_j^2]\le 2(1+C_1)\tau^2/C_2^2\le\tau^2$。由于 $\sum_{j\ge1}(2/3)^j=2$，我们还有 $\mathbb{E}[\mathsf{q}(\widehat{\mu})]\le C_1(s/\tau)[1+\sum_{j\ge1}(2/3)^j]=\mathcal{O}(s/\tau)$。

~ ~ ~对于与 Y耦合的编码输出 Y，满足 $\mathbb{E}[(Y-Y)^2]\le\tau^2/4$，我们有 $|\mathbb{E}[Y]-\mathbb{E}[Y]|\le$~ ~$\tau/2$，且 $\mathrm{sd}(Y)\le s+\tau/2$。将 $\widehat{\mu}_{\tau/2}$应用于满足此标准差界的 Y，可得$\sqrt{\mathbb{E}[(\widehat{\mu}_{\tau/2}-\mathbb{E}[Y])^2]}\le\sqrt{\mathbb{E}[(\widehat{\mu}_{\tau/2}-\mathbb{E}[\widetilde{Y}])^2]}+|\mathbb{E}[\widetilde{Y}]-\mathbb{E}[Y]|\le\tau$，以及 $\mathsf{q}(\widehat{\mu}_{\tau/2})\le C_1(1+s/\tau)$。

对于假设 2 后的事件示例，令 $b_1\ge 1$ 为表示 ~ ~输入和阈值所用的位数。取 $\xi=2^{-b_1}\lfloor 2^{b_1}\xi\rfloor$ 和 $|\widetilde{p}(\mathbf{x})-p(\mathbf{x})|\le 2^{-b_1}$。那么，$\mathbb{P}(\mathbf{1}\{\xi\le\widetilde{p}(\mathbf{x})\}\ne Y(\mathbf{x}\ \xi))\le C_1 2^{-b_1}$。我们将每个标量 z舍入为 $\widetilde{z}$，其中 $|\widetilde{z}-z|\le\tau/4$。对于 Richardson 采样器的完整标量输出Y及其编码值 $\widetilde{Y}$，当 $2^{b_1}\ge C_2/(c^2\tau^2)$时，可得 $\sqrt{\mathbb{E}[(\widetilde{Y}-Y)^2]}\le C_1 2^{-b_1/2}/c+\tau/4\le\tau/2$。对于固定的 d，第 3.1 节中的调度序列意味着 b_1，$b_2=\mathcal{O}(\log(1/\varepsilon))$，其中 b_2 是存储舍入后标量所用的位数。□

LEMMA 2. 定义 $V_k:=L_1\|\mathbf{x}_k-\mathbf{x}_*\|^2/2$ 作为中心化扰动 (2)，并令 $V_k:=f(\mathbf{x}_k)-f_*$作为可行重构 (4)。在假设 *1* 和 *2* 下，存在仅依赖于 L_1/ρ 的常数 $\mathcal{C}_2$、$\mathcal{C}_3>0$，使得采用无偏标量估计且满足 $\eta_k\le 1/(L_1 d)$ 的中心化更新 (2) 满足

$$\mathbb{E}[V_{k+1}\mid\mathcal{F}_k]\le(1-\rho\eta_k)V_k+\mathcal{C}_2L_2^2\eta_k c_k^4+\mathcal{C}_3L_1d^2\eta_k^2\tau_k^2. \tag{8}$$

对于可能有偏的标量估计，有 $\mathbb{E}[V_{k+1}\mid\mathcal{F}_k]\le(1-\rho\eta_k)V_k+\mathcal{C}_2L_2^2\eta_k c_k^4+\mathcal{C}_3 d\eta_k\tau_k^2$. 采用 $\eta_k\le 1/L_1$ 的可行更新 (4) 满足

$$\mathbb{E}[V_{k+1}\mid\mathcal{F}_k]\le(1-\rho\eta_k)V_k+\mathcal{C}_2\frac{mL_2^2R^6}{\lambda}\eta_k c_k^4+\mathcal{C}_3\eta_k\tau_k^2, \tag{9}$$

对于无偏估计使用标量容差 $\overline{\tau_k\sqrt{\lambda/}}$维度 $\mathcal{T}$，对于可能有偏的估计使用 $\overline{\tau_k}\sqrt{\lambda/m}$。

引理*2*的证明。中心化扰动 (2)。 在给定 $\mathcal{F}_k$ 的条件下，记 $\mathbf{x}=\mathbf{x}_k$、$c=c_k$、$\eta=\eta_k$ 和 $\tau=\tau_k$。对于 $\mathbf{\Delta}\in[-1,1]^d$，定义

$$\mathcal{R}_c(\mathbf{\Delta}):=\frac{f(\mathbf{x}+c\mathbf{\Delta}/\sqrt{d})-f(\mathbf{x}-c\mathbf{\Delta}/\sqrt{d})}{2c}-\frac{\langle\nabla f(\mathbf{x}),\mathbf{\Delta}\rangle}{\sqrt{d}}.$$

泰勒公式和假设 1 蕴含 $|\mathcal{R}_c(\mathbf{\Delta})|\le L_2c^2\|\mathbf{\Delta}\|^3/(6d^{3/2})$、$\|\nabla_{\mathbf{\Delta}}\mathcal{R}_c(\mathbf{\Delta})\|\le L_2c^2\|\mathbf{\Delta}\|^2/(2d^{3/2})$ 以及 $\|\nabla^2_{\mathbf{\Delta}}\mathcal{R}_c(\mathbf{\Delta})\|\le L_2c^2\|\mathbf{\Delta}\|/d^{3/2}$。对于采样方向 $\mathbf{\Delta}=\mathbf{\Delta}_k$，$\mathcal{R}_c(-\mathbf{\Delta})=-\mathcal{R}_c(\mathbf{\Delta})$ 蕴含 $\mathbb{E}[\mathcal{R}_c(\mathbf{\Delta})\mid\mathcal{F}_k]=0$。

令 $\mathbf{\Delta}^{(i)}$ 表示符号 i 翻转后的 $\mathbf{\Delta}$。由于对于 $t\in[0,1]$，$\|t\mathbf{\Delta}+(1-t)\mathbf{\Delta}^{(i)}\|\le\sqrt{d}$，根据泰勒公式可得 $|\mathcal{R}_c(\mathbf{\Delta})-\mathcal{R}_c(\mathbf{\Delta}^{(i)})|\le 2|\partial_i\mathcal{R}_c(\mathbf{\Delta})|+2L_2c^2/d$。因此，根据埃夫隆–斯坦不等式，我们有

$$\begin{aligned}\mathbb{E}[\mathcal{R}_c(\mathbf{\Delta})^2\mid\mathcal{F}_k]&\le\frac{1}{4}\sum\nolimits_i\mathbb{E}[(\mathcal{R}_c(\mathbf{\Delta})-\mathcal{R}_c(\mathbf{\Delta}^{(i)}))^2\mid\mathcal{F}_k]\\&\le 2\mathbb{E}[\|\nabla_{\mathbf{\Delta}}\mathcal{R}_c(\mathbf{\Delta})\|^2\mid\mathcal{F}_k]+2L_2^2c^4/d\le\frac{5L_2^2c^4}{2d}.\end{aligned}$$

由于 $\mathbb{E}[\boldsymbol{\Delta}\boldsymbol{\Delta}^\top \mid \mathcal{F}_k] = \mathbf{I}$，$\|\mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})\boldsymbol{\Delta} \mid \mathcal{F}_k]\|^2 \le \mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})^2 \mid \mathcal{F}_k]$。对于无偏标量估计，$\mathbf{g}_k$ 的条件均值和二阶矩分别为 $\mathbb{E}[\mathbf{g}_k \mid \mathcal{F}_k] = \nabla f(\mathbf{x}) + \sqrt{d}\,\mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})\boldsymbol{\Delta} \mid \mathcal{F}_k]$ 和 $\mathbb{E}[\|\mathbf{g}_k\|^2 \mid \mathcal{F}_k] \le d\|\nabla f(\mathbf{x})\|^2 + 2d^{3/2}\langle \nabla f(\mathbf{x}),\, \mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})\boldsymbol{\Delta} \mid \mathcal{F}_k]\rangle + d^2\big(\mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})^2 \mid \mathcal{F}_k] + \tau^2\big)$。

由于 $\nabla f(\mathbf{x}_*) = \mathbf{0}$ 且 $\eta d \le 1/L_1$，海森矩阵界意味着

$$\|\mathbf{x} - \mathbf{x}_* - \eta d \nabla f(\mathbf{x})\| = \Big\| \Big[\mathbf{I} - \eta d \int_0^1 \nabla^2 f(\mathbf{x}_* + t(\mathbf{x}-\mathbf{x}_*))\, dt\Big](\mathbf{x}-\mathbf{x}_*)\Big\| \le (1 - \rho\eta d)\|\mathbf{x}-\mathbf{x}_*\|.$$

由投影的非扩张性和两个矩界可得

$$\begin{aligned}\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{x}_*\|^2 \mid \mathcal{F}_k] &\le (1 - 1/d)\|\mathbf{x}-\mathbf{x}_*\|^2 + \|\mathbf{x} - \mathbf{x}_* - \eta d \nabla f(\mathbf{x})\|^2/d \\ &\quad - 2\eta\sqrt{d}\,\langle \mathbf{x} - \mathbf{x}_* - \eta d \nabla f(\mathbf{x}), \mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})\boldsymbol{\Delta} \mid \mathcal{F}_k]\rangle + \eta^2 d^2\big(\mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})^2 \mid \mathcal{F}_k] + \tau^2\big).\end{aligned}$$

内积项的绝对值至多为

$$2\eta(1-\rho\eta d)\|\mathbf{x}-\mathbf{x}_*\|\sqrt{d\mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})^2 \mid \mathcal{F}_k]} \le \rho\eta(1-\rho\eta d)\|\mathbf{x}-\mathbf{x}_*\|^2 + \frac{\eta d(1-\rho\eta d)}{\rho}\mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})^2 \mid \mathcal{F}_k].$$

这里，平方距离系数可简化为 $1 - 1/d + (1-\rho\eta d)^2/d + \rho\eta(1-\rho\eta d) = 1 - \rho\eta$，而 $\eta d(1-\rho\eta d)/\rho + \eta^2 d^2 = \eta d/\rho$。因此，我们有

$$\begin{aligned}\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{x}_*\|^2 \mid \mathcal{F}_k] &\le (1-\rho\eta)\|\mathbf{x}-\mathbf{x}_*\|^2 + \frac{\eta d}{\rho}\mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})^2 \mid \mathcal{F}_k] + \eta^2 d^2 \tau^2 \\ &\le (1-\rho\eta)\|\mathbf{x}-\mathbf{x}_*\|^2 + \frac{5L_2^2}{2\rho}\eta c^4 + \eta^2 d^2 \tau^2.\end{aligned}$$

乘以 $L_1/2$ 即可证明 (8)，其中 $\mathcal{C}_2 = 5L_1/(4\rho)$ 和 $\mathcal{C}_3 = 1/2$ 足以满足此情形。

对于可能有偏的标量估计，将前述计算中的 $\mathcal{R}_c(\boldsymbol{\Delta})$ 替换为 $\mathcal{R}_c(\boldsymbol{\Delta}) + \widehat{\mu}_k - \mu_c(\mathbf{x}\ \boldsymbol{\Delta})$。其二阶矩至多为 $2\mathbb{E}[\mathcal{R}_c(\boldsymbol{\Delta})^2 \mid \mathcal{F}_k] + 2\tau^2$。应用相同的投影和 Young界，我们得到 $\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \le (1-\rho\eta)V_k + 5L_1L_2^2\eta c^4/(2\rho) + L_1 d\eta\tau^2/\rho$. 由于 $\|\nabla f(\mathbf{x})\| \le L_1\|\mathbf{x}-\mathbf{x}_*\|$，条件的矩界和投影的非扩张性通过从确定性初始化 $\mathbf{x}_0$开始的归纳法给出 $\mathbb{E}[\|\mathbf{x}_k - \mathbf{x}_*\|^2 + V_k + \|\mathbf{g}_k\|^2] < \infty$。

可行重构 (4)。</style>在给定 $,\mathbf{x}_k^{[1]},\ldots,\mathbf{x}_k^{[m]}$ 的条件下，在下面的期望和方差中省略迭代指标及条件。通过对 $\mathbf{M}_\mathbf{x}$ 进行谱分解，我们有 $\sum_i \mathbf{s}_i\mathbf{w}_i^\top = \mathbf{P}$，$\sum_i \|\mathbf{s}_i\|^2 = \mathrm{tr}(\mathbf{M}_\mathbf{x}^\dagger) \le \dim\ \mathcal{T}/\lambda$，以及对于 $(a_1,\ldots,a_m) \in \mathbb{R}^m$ 的 $\|\sum_i \mathbf{s}_i a_i\|^2 \le \lambda^{-1}\sum_i a_i^2$。

令 $b_i^F = [4f(\mathbf{x} + c\mathbf{w}_i/2) - f(\mathbf{x} + c\mathbf{w}_i) - 3f(\mathbf{x})]/c - \langle \nabla_\mathcal{T} f(\mathbf{x}),\ \mathbf{w}_i\rangle$。式 (3) 的泰勒展开中的二次项相互抵消，并且 $|b_i^F| \le L_2\|\mathbf{w}_i\|^3[4(c/2)^3 + c^3]/(6c) \le C_1 L_2 R^3 c^2$。对于 $\mathbf{w}_i \neq \mathbf{0}$，两组成对实验的独立性以及假设 2 意味着

$$\frac{16}{c^2}\mathrm{Var}[Y(\mathbf{x} + c\mathbf{w}_i/2, \xi_1) - Y(\mathbf{x}, \xi_1')] + \frac{1}{c^2}\mathrm{Var}[Y(\mathbf{x} + c\mathbf{w}_i, \xi_2) - Y(\mathbf{x}, \xi_2')] \le \sigma^2 R^{2\kappa}(1 + 2^{4-2\kappa})c^{2\kappa-2}.$$

令 $e_i^F = \widehat{\mu}_i^F - [4f(\mathbf{x} + c\mathbf{w}_i/2) - f(\mathbf{x} + c\mathbf{w}_i) - 3f(\mathbf{x})]/c$。无偏性和条件独立性意味着 $\mathbb{E}[e_i^F] = 0$，$\mathbb{E}[e_i^F e_j^F] = 0$ 对于 $i \neq j$，且 $\mathbb{E}[(e_i^F)^2] \le \tau^2\lambda/$ 维。因此，

$$\mathbb{E}[\|\widetilde{\mathbf{g}} - \nabla_\mathcal{T} f(\mathbf{x})\|^2] = \big\|\sum\nolimits_i \mathbf{s}_i b_i^F\big\|^2 + \sum\nolimits_i \|\mathbf{s}_i\|^2 \mathbb{E}[(e_i^F)^2] \le C_1 m L_2^2 R^6 c^4/\lambda + \tau^2. \tag{10}$$

记作 $\mathbf{x}^+ = \Pi_{\mathcal{X}}(\mathbf{x} - \eta\widetilde{\mathbf{g}})$。对于每个 $\mathbf{x}' \in \mathcal{X}$，投影不等式为 $\langle \widetilde{\mathbf{g}}, \mathbf{x}^+ - \mathbf{x}' \rangle \le [\|\mathbf{x}' - \mathbf{x}\|^2 - \|\mathbf{x}^+ - \mathbf{x}\|^2 - \|\mathbf{x}' - \mathbf{x}^+\|^2]/(2\eta)$。结合光滑性、 $L_1\eta \le 1$和杨氏不等式，我们有

$$\begin{aligned} f(\mathbf{x}^+) &\le f(\mathbf{x}) + \langle \nabla_{\mathcal{T}} f(\mathbf{x}), \mathbf{x}' - \mathbf{x} \rangle + \|\mathbf{x}' - \mathbf{x}\|^2/(2\eta) + \langle \widetilde{\mathbf{g}} - \nabla_{\mathcal{T}} f(\mathbf{x}), \mathbf{x}' - \mathbf{x}^+ \rangle - \|\mathbf{x}' - \mathbf{x}^+\|^2/(2\eta) \\ &\le f(\mathbf{x}) + \langle \nabla_{\mathcal{T}} f(\mathbf{x}), \mathbf{x}' - \mathbf{x} \rangle + \|\mathbf{x}' - \mathbf{x}\|^2/(2\eta) + \eta\|\widetilde{\mathbf{g}} - \nabla_{\mathcal{T}} f(\mathbf{x})\|^2/2. \end{aligned}$$

取$\mathbf{x}' = (1-\rho\eta)\mathbf{x} + \rho\eta\mathbf{x}_* \in \mathcal{X}$。强凸性意味着 $\langle \nabla_{\mathcal{T}} f(\mathbf{x}),\ \mathbf{x}_* - \mathbf{x} \rangle \le f_* - f(\mathbf{x}) - \rho\|\mathbf{x}_* - \mathbf{x}\|^2/2$，因此平方距离项相互抵消，并且 $f(\mathbf{x}^+) - f_* \le (1-\rho\eta)(f(\mathbf{x}) - f_*) + \eta\|\widetilde{\mathbf{g}} - \nabla_{\mathcal{T}} f(\mathbf{x})\|^2/2$。先对锚点取条件期望，再对 $\mathcal{F}_k$取条件期望，并应用(10)，即可证明(9)。此外，由紧致性，$\sup_{\mathbf{x}\in\mathcal{X}} |f(\mathbf{x})| < \infty$，因此 $\mathbb{E}[V_k] < \infty$。

对于容差为 $\tau\sqrt{\lambda/m}$的可能有偏标量估计， $\mathbb{E}[\|\sum_i \mathbf{s}_i e_i^F\|^2] \le \lambda^{-1}\sum_i \mathbb{E}[(e_i^F)^2] \le \tau^2$。因此，$\mathbb{E}[\|\widetilde{\mathbf{g}} - \nabla_{\mathcal{T}} f(\mathbf{x})\|^2] \le C_1 m L_2^2 R^6 c^4/\lambda + 2\tau^2$，而投影界证明了(9)。

对于张成 $\mathcal{X}$ 仿射包的固定锚点，令 $\mathbf{x} = m^{-1}\sum_i \mathbf{x}^{[i]}$。对于 $\mathbf{v} \in \mathcal{T}$， $\mathbf{v}^\top \mathbf{M}_{\mathbf{x}} \mathbf{v} = \sum_i \langle \mathbf{v},\ \mathbf{x}^{[i]} - \mathbf{x} \rangle^2 + m\langle \mathbf{v},\ \underline{\mathbf{x}} - \mathbf{x} \rangle^2 \ge \sum_i \langle \mathbf{v},\ \mathbf{x}^{[i]} - \mathbf{x} \rangle^2$。仿射张成意味着 $\inf_{\mathbf{x}\in\mathcal{X}} \lambda_{\min}(\mathbf{M}_{\mathbf{x}}|_{\mathcal{T}}) \ge \min_{\mathbf{v}\in\mathcal{T}, \|\mathbf{v}\|=1} \sum_i \langle \mathbf{v},\ \mathbf{x}^{[i]} - \mathbf{x} \rangle^2 > 0$。对于单纯形顶点和 $\mathbf{v} \in \mathbf{1}^\perp$， $\mathbf{v}^\top \mathbf{M}_{\mathbf{x}} \mathbf{v} = \sum_i (v_i - \mathbf{x}^\top \mathbf{v})^2 = \|\mathbf{v}\|^2 + d(\mathbf{x}^\top \mathbf{v})^2 \ge \|\mathbf{v}\|^2$。 □

LEMMA 3. 假设假设 *1* 和 *2*成立。令 $\eta_k \le 1/(L_1 d)$ 表示中心化扰动(2)，并令$\eta_k \le 1/L_1$ 表示可行重构(4)。利用引理*2*中的常数，定义

$$\psi_k := \begin{cases} \mathcal{C}_2 L_2^2 c_k^4 + \mathcal{C}_3 L_1 d^2 \eta_k \tau_k^2, & \textit{centered perturbations, unbiased,} \\ \mathcal{C}_2 L_2^2 c_k^4 + \mathcal{C}_3 d \tau_k^2, & \textit{centered perturbations, possibly biased,} \\ \mathcal{C}_2 m L_2^2 R^6 c_k^4/\lambda + \mathcal{C}_3 \tau_k^2, & \textit{feasible reconstruction.} \end{cases}$$

于是，

$$\mathbb{E}[f(\mathbf{x}_k) - f_*] \le \mathbb{E}[V_k] \le V_0 \prod_{j=0}^{k-1}(1-\rho\eta_j) + \sum_{j=0}^{k-1} \eta_j \psi_j \prod_{t=j+1}^{k-1}(1-\rho\eta_t). \tag{11}$$

此外， $\sum_{k=0}^{\infty} \eta_k = \infty$ 和 $\sum_{k=0}^{\infty} \eta_k \psi_k < \infty$ 推出$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}_*$ 几乎必然成立。

引理*3*的证明。 对于中心化扰动，由光滑性和 $\nabla f(\mathbf{x}_*) = \mathbf{0}$可得 $f(\mathbf{x}_k) - f_* \le L_1\|\mathbf{x}_k - \mathbf{x}_*\|^2/2 = V_k$；对于可行重构， $V_k = f(\mathbf{x}_k) - f_*$。对引理2中的期望递推进行迭代，即可证明式 (11)。令 $\mathcal{E}_0 := \mathbb{E}[f(\mathbf{x}_0) - f_*]$，由强凸性可得，对于中心化扰动， $V_0 \le (L_1/\rho)\mathcal{E}_0$；对于可行重构，则有 $V_0 = \mathcal{E}_0$。

对于几乎必然收敛，我们应用引理2得到 $\mathbb{E}[V_{k+1} + \sum_{j=k+1}^{\infty} \eta_j\psi_j \,|\, \mathcal{F}_k] \le V_k + \sum_{j=k}^{\infty} \eta_j\psi_j - \rho\eta_k V_k$. 非负超鞅收敛定理推出 $\lim_{k\to\infty}(V_k + \sum_{j=k}^{\infty} \eta_j\psi_j) \in [0,\infty)$ 几乎必然成立。对漂移界求和并应用托内利定理，我们得到 $\mathbb{E}[\sum_k \eta_k V_k] \le (V_0 + \sum_k \eta_k\psi_k)/\rho < \infty$。由于 $\lim_{k\to\infty} \sum_{j=k}^{\infty} \eta_j\psi_j = 0$ 且 $\sum_k \eta_k = \infty$，我们得到 $\lim_{k\to\infty} V_k = 0$ 几乎必然成立。最后，注意到 $\|\mathbf{x}_k - \mathbf{x}_*\|^2 \le 2V_k/\rho$ 对两种更新都成立。因此，$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}_*$ 几乎必然成立。□

PROPOSITION 1. 在假设 *1* 和 *2* 下，对于中心化扰动 (2)，无偏量子均值估计和经典配对平均满足

$$\mathbb{E}[N_k^Q] = \mathcal{O}\Big(\sum_{j=0}^{k-1}\Big[1+\frac{2^{\kappa-1}\sigma}{c_j^{1-\kappa}\tau_j}\Big]\Big) \quad and \quad N_k^C = \mathcal{O}\Big(\sum_{j=0}^{k-1}\Big[1+\frac{2^{2\kappa-2}\sigma^2}{c_j^{2(1-\kappa)}\tau_j^2}\Big]\Big) \quad a.s. \tag{12}$$

对于在标量容差 $\tau_k\sqrt{\lambda}$/维度 $\mathcal{T}$下具有无偏估计的可行重构 (4)，

$$\mathbb{E}[N_k^Q] = \mathcal{O}\Big(m\sum_{j=0}^{k-1}\Big[1+\frac{\sigma R^\kappa\sqrt{(1+2^{4-2\kappa})\dim\mathcal{T}/\lambda}}{c_j^{1-\kappa}\tau_j}\Big]\Big)$$
$$and \quad N_k^C = \mathcal{O}\Big(m\sum_{j=0}^{k-1}\Big[1+\frac{\sigma^2R^{2\kappa}(1+2^{4-2\kappa})\dim\mathcal{T}}{\lambda c_j^{2(1-\kappa)}\tau_j^2}\Big]\Big) \quad a.s.$$

对于具有确定性查询上限的估计，同样的量子界几乎处处成立，其中将维度$\mathcal{T}$替换为 m ，用于可行重构 (4) 和标量容差 $\tau_k\sqrt{\lambda}/m$。

命题*1*的证明。对于中心化差分采样器， $s_j/\tau_j = 2^{\kappa-1}\sigma/(c_j^{1-\kappa}\tau_j)$。引理 1 和经典批量大小 $\max\{1,\lceil s_j^2/\tau_j^2\rceil\}$，再对 $j<k$ 求和，证明了 (12)。对于可行重构 (4)，在 $c=c_j$ 处代入 s^F 以及给定的标量容差，使用相同的界并对 $i\le m$ 和 $j<k$求和。□

定理 2 的证明。定义 $\mathcal{E}_k := \mathbb{E}[f(\mathbf{x}_k)-f_*]$。这些调度序列意味着，对于某个常数 $C_1>0$, $\psi_k \le C_1(k+k_0)^{-\alpha}$。选择充分大的 $C_2>0$，使得 $C_2 \ge \max\{V_0k_0^\alpha, 2C_1/\rho\}$。于是， $C_2[(k+k_0)^{-\alpha} - (k+k_0+1)^{-\alpha}] \le C_2\alpha(k+k_0)^{-\alpha-1} \le \bar{\eta}(\rho C_2 - C_1)(k+k_0)^{-2\alpha}$。结合 $V_0 \le C_2k_0^{-\alpha}$ 和引理2，通过归纳法可得 $\mathcal{E}_k \le \mathbb{E}[V_k] \le C_2(k+k_0)^{-\alpha}$。由于 $\sum_k \eta_k\psi_k \le C_1\bar{\eta}\sum_k(k+k_0)^{-2\alpha} < \infty$ 且 $\sum_k \eta_k = \infty$，引理3推出 $\lim_{k\to\infty}\mathbf{x}_k = \mathbf{x}_*$ 几乎处处成立；此外， $\mathbb{E}[\|\mathbf{x}_k-\mathbf{x}_*\|^2] \le 2\mathbb{E}[V_k]/\rho = \mathcal{O}((k+k_0)^{-\alpha})$。

根据单调性，我们有 $\sum_{j=0}^{k-1}(c_j^{1-\kappa}\tau_j)^{-p} \le k(c_{k-1}^{1-\kappa}\tau_{k-1})^{-p}$ 对于 $p\in\{1,2\}$。因此，由 (12) 可得，对于采用中心化扰动的 QACE， $\mathbb{E}[N_k^Q] = \mathcal{O}(k^{1+\alpha(1-\kappa)/4})$ 以及 $N_k^C = \mathcal{O}(k^{1+\alpha(1-\kappa)/2})$ 几乎处处成立。结合 $\mathcal{E}_k = \mathcal{O}(k^{-\alpha})$，这些界在期望预算下推出所述速率。

对于可行重构 (4)，命题1中的附加因子 $\tau_k^{-1} = \mathcal{O}(k^{\alpha/2})$ 给出了所述的查询阶数。

对于使用一个独立经典对的 SPSA， $\kappa=0$ 以及 $\tau_k = \sigma/(2c_k)$。取 $c_k = \bar{c}(k+k_0)^{-\alpha/6}$，可得 (8) 中的 $\psi_k = \mathcal{O}((k+k_0)^{-2\alpha/3})$。对指数 $2\alpha/3$ 进行相同的比较，可得 $\mathcal{E}_k \le \mathbb{E}[V_k] = \mathcal{O}((k+k_0)^{-2\alpha/3})$ 以及 $\sqrt{\mathbb{E}[\|\mathbf{x}_k-\mathbf{x}_*\|^2]} = \mathcal{O}((k+k_0)^{-\alpha/3})$。□

参考文献

Borkar VS (2009) 随机逼近：动力系统视角，volume 48(Springer).

Cao H、Hu JQ、Hu J（2026a）黑箱CoVaR及其梯度估计。*INFORMS JournalonComputing* 38(5):1503–1520。

Cao H、Hu JQ、Hu J（2026b）通过仿真进行扭曲风险度量的估计、敏感性分析与优化。*arXiv* 预印本 *arXiv:2610.13751* 。Cao H、Hu JQ、Lian T（2025a）黑箱仿真优化。中国运筹学学会学报 13(3):723–749。Cao H、Hu JQ、Lian T、Yang X（2025b）具有未知参数的无穷小扰动分析（IPA）导数估计。*Automatica* 174:112140。Che E、Dong J、Tong XT（2026）使用自适应数据的随机梯度下降。《运筹学》 74(4):2113–2145。Chen X、Hong G、Liu Y（2026）针对具有未知到达率和服务分布的队列的在线学习与优化。《运筹学》 74(4):2053–2074。Fu MC（2002）仿真优化：理论与实践。*INFORMS Journal on Computing*14(3):192–215。Fu MC（2008）关于仿真与导数，你应该了解的内容。*Naval Research Logistics*（*NRL*） 55(8):723–736。Fu MC、Hu JQ（2012）条件蒙特卡罗：梯度估计与优化应用，第392卷（Springer Science & Business Media）。Golovin D、Solnik B、Moitra S、Kochanski G、Karro J、Sculley D（2017）Google Vizier：一种用于黑箱优化的服务。第*23*届*ACM SIGKDD*知识发现与数据挖掘国际会议论文集，1487–1495。Gong W、Zhang C、Li T（2025）非凸优化量子算法的鲁棒性。*International Conference on Learning Representations*。Grinko D、Gacon J、Zoufal C、Woerner S（2021）迭代量子振幅估计。*npjQuantumInformation* 7(1):52。He S、Jiang G、Lam H、Fu MC（2024）用于高效随机求根和分位数估计的自适应重要性采样。《运筹学》 72(6):2612–2630。Hu J、Fu MC（2025）技术注释——关于基于梯度的随机优化中随机逼近的收敛速率。《运筹学》 73(2):1143–1150。Hu J、Peng Y、Zhang G、Zhang Q（2022）一种基于仿真的分位数优化随机逼近方法。*INFORMS Journal on Computing* 34(6):2889–2907。Hu J、Song M、Fu MC（2025a）通过多时间尺度局部搜索进行黑箱函数的分位数优化。《运筹学》 73(3):1535–1557。Hu J、Song M、Fu MC、Peng Y（2025b）条件风险价值的仿真优化。*IISETransactions* 57(10):1167–1181。Jiang J、Heidergott B、Hu J、Peng Y（2025）扭曲风险度量优化的随机逼近方法。*arXiv*预印本 *arXiv:2510.04563* 。

Jiang J, Hu J, Peng Y（2023）使用双时间尺度策略梯度算法的基于分位数的深度强化学习。*arXiv* 预印本 *arXiv:2305.07248* 。Kiefer J, Wolfowitz J（1952）回归函数最大值的随机估计。《数学统计年刊》23(3):462–466。Kim JL, Augustino B, Herman D, Fontana E, Watkins J, Pistoia M, Chakrabarti S（2025）利用量子梯度方法实现快速零阶凸优化。《神经信息处理系统进展》，第38卷。Kleinman NL, Spall JC, Naiman DQ（1999）使用共同随机数和随机逼近进行基于仿真的优化。《管理科学》45(11):1570–1578。Kothari R, O'Donnell R（2023）已知源代码时的均值估计；或量子蒙特卡罗方法。*2023*年 *ACM-SIAM* 离散算法年度研讨会论文集，1186–1215（工业与应用数学学会）。Kushner HJ, Yin GG（2003）随机逼近与递归算法及其应用（Springer）。Lam H, Zhang J（2025）分布约束黑箱随机梯度估计与优化。《运筹学》73(5):2680–2694。Li X, Liang J, Chen X, Zhang Z（2026a）流式随机梯度下降的收敛性与推断及其在排队系统和库存控制中的应用。《运筹学》74(4):2034–2052。Li Z, Lin Z, Peng Y（2026b）超越似然比偏差：用于无似然参数估计的嵌套多时间尺度随机逼近。《机器学习》115(11):250。Liu C, Guan C, He J, Lui JCS（2024）非光滑非凸优化的量子算法。《神经信息处理系统进展》，第37卷。Ozgul G, Li X, Mahdavi M, Wang C（2026）利用随机预言机加速采样与非凸优化。第*21*届量子计算、通信与密码学理论会议（*TQC 2026*），《莱布尼茨国际信息学论文集》第389卷，8:1–8:25（Schloss Dagstuhl – 莱布尼茨信息学中心）。Rhee CH, Glynn P W（2015）随机微分方程模型的无偏估计与平方根收敛。《运筹学》63(5):1026–1043。Sidford A, Zhang C（2023）随机优化的量子加速。《神经信息处理系统进展》，第36卷，35300–35330。Spall JC（1992）使用同时扰动梯度近似的多元随机逼近。《*IEEE* 自动控制汇刊》37(3):332–341。WangD, You X, Li T, Childs AM（2021）多臂老虎机的量子探索算法。*AAAI* 人工智能会议论文集，第35卷，10102–10110。

Wang Q, Cao H, Hu JQ（2025）条件风险价值优化中的统计推断。可见 SSRN 5168942。
Wang Q, Cao H, Hu JQ, Hu M, Xia L（2026）长期条件风险价值强化学习。*arXiv* 预印本 *arXiv:2603.09734* 。Wang Y, Zhang X（2026）利用大语言模型驱动的多智能体仿真优化服务运营。*arXiv* 预印本 *arXiv:2604.04383* 。Yu Q, Wang Y, Huang B, Lei Q, Lee JD（2024）强凸性与 Lipschitz Hessian 条件下的随机零阶优化：极小极大样本复杂度。《神经信息处理系统进展》，第37卷。