arXiv:2302.04649v2 [quant-ph] 7 Sep 2023

变分量子电路可训练性的高效估计

Valentin Heyraud,[1] Zejian Li,[1,2] Kaelan Donatella,[1] Alexandre Le Boit´e,[1] 和 Cristiano Ciuti[1]

[1] *Université Paris Cité, CNRS, Matériaux et Phénomènes Quantiques (MPQ), F-75013 Paris, France*

[2] *The Abdus Salam International Center for Theoretical Physics (ICTP), Strada Costiera 11, I-34151 Trieste, Italy*

（日期：2023年9月8日）

用作变分 ans¨atzes 的参数化量子电路，正逐渐成为解决复杂问题的有力工具，应用范围涵盖量子化学到组合优化。这些变分量子电路可能遭遇众所周知的贫瘠高原诅咒，其特征是代价函数梯度随系统规模增大而呈指数级消失，导致实际应用中难以训练。由于一般的量子线路无法被高效模拟，确定其可训练性是一项重要问题。本文提出一种高效方法，可计算广泛一类变分量子电路的代价函数梯度及其方差。我们的方法基于一项证明：随机初始化电路与一组 Clifford 电路之间存在精确映射；根据著名的戈特斯曼–克尼尔定理，这些电路可在经典计算机上高效模拟。该方法具有可扩展性，可用于验证变分量子电路的可训练性，并探索能够克服贫瘠高原问题的设计策略。作为示例，我们展示了多达 100 个量子比特的结果。

I. 引言

受机器学习方法成功的启发，变分量子算法 [1–3]已成为一种前景广阔的方法，可在从量子化学 [4–6] 到组合优化问题 [7–9]等多个领域中发挥量子计算的优势。这些算法将参数化量子线路的输出用作变分拟¨设，其参数通过基于梯度的方法进行经典优化。

变分量子电路可能会受到贫瘠高原所导致的可训练性问题的影响 [10],；这一局限已在近期文献中得到广泛研究 [11–38]。其特征是代价函数的梯度随系统规模增大而呈指数级消失，导致在量子比特数量较多时无法训练变分量子电路。贫瘠高原可能源于多种本质不同的现象。文献 [10]首次指出，对于 2-设计（随机酉变换在二阶矩意义上与 Haar 分布相匹配），会出现贫瘠高原。近期研究将贫瘠高原与拟设的可表达性 [11] 以及噪声 [12] 和纠缠联系起来。具体而言，文献 [13]的作者表明，对于可拆分为隐藏子系统和可见子系统的架构（例如量子玻尔兹曼机或前馈量子神经网络），两个子系统之间过度纠缠会导致可见子系统处于高度混合态。这可能使代价函数的景观变得平坦。其他研究[14, 15], 也考察了代价函数的结构对贫瘠高原出现的影响，并发现全局代价函数更容易出现贫瘠高原。值得注意的是，量子核机器 [39–43] 和储备池计算模型 [44–47], 等浅层模型虽然通常比变分量子算法更易训练，但也可能受到类似性质的可训练性问题影响 [48]。

大量研究提出了应对贫瘠高原问题的策略。在纠缠诱导的贫瘠高原背景下，大多数策略都依赖于限制纠缠程度 [16–20]。其他方法则采用针对性设计的初始电路参数分布和精心设计的电路架构 [21–29]。然而，只有少数配置能够提供可训练性保证，并具备抵御贫瘠高原的鲁棒性 [30, 31]。解决这一根本问题仍是一项重要的理论挑战。

本文针对这一问题提出了一种替代方法：为广泛的一类变分量子算法提供一种高效方法，以估计平均梯度及其方差。通过研究与随机单量子比特旋转相关联的量子信道，我们证明，在一些简单条件下，一阶矩和二阶矩可以表示为由 Clifford 门 [50]组成的混合酉通道 [49] 。在对随机角度分布作出一些额外的一般性假设后，我们证明，这使我们能够将由 Clifford 门和参数化旋转组成的随机初始化电路精确映射为一个 Clifford 电路系综。此外，我们证明，所得系综可以通过高效采样来计算所关注的量，例如梯度方差，或对初始随机参数取平均后的代价函数。借助著名的 Gottesman-Knill 定理 [51, 52], ，我们从解析上证明了该方法的高效性；该方法可在经典计算机上实现，其复杂度随变分参数数量和系统规模呈多项式增长。此外，我们还展示了一些数值实验，以随机电路为例说明该方法，并以多项式资源忠实地复现了文献 [10] 中首次发现的方差指数级抑制现象，适用于作用于多达 100 个量子比特的电路。

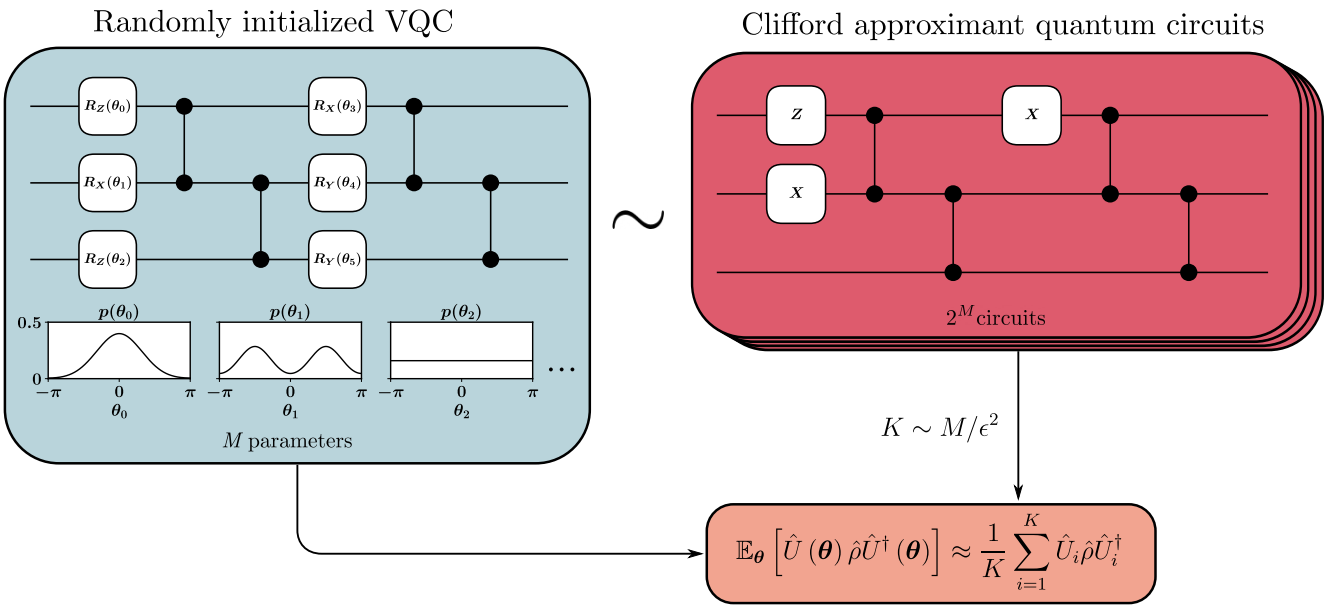


图 1. 示意图展示了如何将参数随机的参数化量子电路映射为用于计算一阶量的 Clifford 近似电路（一阶量的计算仅需了解旋转单重通道，见附录 B）。对于含有 M 个参数的量子线路，样本量达到 M/ϵ^2 量级便足以对初始电路上的平均值进行近似，且可观测量均值的精度为 ϵ （见附录 C）。

II. 理论框架

A. 变分问题

在变分量子算法中，作用于 n 个量子比特的参数化酉变换 $\hat{U}(\boldsymbol{\theta})$ 被用作变分拟设，以实现一个表述为最小化代价函数的任务

$$C(\boldsymbol{\theta}) = \mathrm{Tr}\left[\hat{U}(\boldsymbol{\theta})\hat{\rho}\hat{U}^\dagger(\boldsymbol{\theta})\hat{O}\right] \tag{1}$$

$\hat{}$ 对于某个可观测量 O 以及某个初始的 n量子比特态 $\hat{}\rho$。这一表述具有一般性，涵盖了典型任务，例如制备目标态 $|\psi\rangle\hat{}$（设定 $O = -|\psi\rangle\langle\psi|\hat{}$）或搜索某个哈密顿量 $\hat{}\mathcal{H}$ 的基态（设定 $\hat{}O = \hat{}\mathcal{H}\hat{}$）。所考虑的参数化酉算子通常由一系列参数化门和固定层组成。这里我们考虑如下形式的一般拟设

$$\hat{U}(\boldsymbol{\theta}) = \prod_{i=1}^{M}\hat{U}_i(\theta_i)\hat{W}_i, \tag{2}$$

其中，每个酉算子 $\hat{}U_i(\theta_i) = e^{-\mathrm{i}\frac{\theta_i}{2}\hat{}P_i}$ 都是与给定泡利算符 $\hat{}P_i \in \{\hat{}X、\hat{}Y、\hat{}Z\}$相关联的单量子比特旋转，而 W_k 是由一系列未参数化门组成的固定层，这些门可作用于多个量子比特。将固定层中的 Clifford 门吸收后，我们将考虑一类相当一般的、基于 Z 旋转 [53]的量子线路。酉变换 $U(\boldsymbol{\theta})$取决于 M 个连续参数，这些参数汇集在向量 $\boldsymbol{\theta} = (\theta_0、\ldots\ldots、\theta_{M-1})$中。可以使用经典梯度下降技术优化这些旋转参数。利用参数移位规则 [54, 55]，可以方便地估算代价函数关于第 k个参数的梯度：

$$\partial_k C(\boldsymbol{\theta}) = \frac{1}{2}\left(C(\boldsymbol{\theta} + \frac{\pi}{2}\boldsymbol{e}_k) - C(\boldsymbol{\theta} - \frac{\pi}{2}\boldsymbol{e}_k)\right), \tag{3}$$

其中， $\boldsymbol{e}_k$ 是沿分量 k方向的标准基向量。值得注意的是，参数θ_k 中的 $\pm\pi/2$ 平移可以提取出来，并视为添加到固定层 W_k中的一个额外 Clifford 门 $\hat{}$。事实上，注意到相位门 $\hat{}S$ 可以写成 $\hat{}S = e^{\mathrm{i}\frac{\pi}{4}}e^{-\mathrm{i}\frac{\pi}{2}\hat{}Z}$，–并假设 $\hat{}\hat{}P_k = Z$，我们有：

$$\begin{aligned}\hat{U}_k(\theta_k + \pi/2)\hat{W}_k &= e^{-\mathrm{i}\frac{\theta_k}{2}\hat{Z}}e^{-\mathrm{i}\frac{\pi}{2}\hat{Z}}\hat{W}_k \\ &= e^{-\mathrm{i}\frac{\pi}{4}}\hat{U}_k(\theta_k)\hat{S}\hat{W}_k\,.\end{aligned} \tag{4}$$

我们定义 $\hat{}W_{k,\pm} = e^{\mp\mathrm{i}\frac{\pi}{4}}\hat{}S\,\hat{}W_k$，使得我们可以写成

$$\hat{U}_k(\theta_k \pm \pi/2)\hat{W}_k = \hat{U}_k(\theta_k)\hat{W}_{k,\pm}\,. \tag{5}$$

我们将其记为

$$\hat{U}_\pm(\boldsymbol{\theta}) = \hat{U}\left(\boldsymbol{\theta} \pm \frac{\pi}{2}\boldsymbol{e}_k\right) \tag{6}$$

参数移位规则中出现的移位酉算子。由前文可知

$$\hat{U}_{\pm}(\boldsymbol{\theta}) = \prod_{i=1}^{M} \hat{U}_i(\theta_i)\hat{V}_{i,\pm}\,, \tag{7}$$

with

$$\hat{V}_{i,\pm} = \begin{cases} \hat{W}_{k,\pm} & \text{if } i = k\ , \\ \hat{W}_i & \text{otherwise}\ . \end{cases} \tag{8}$$

B. 酉系综与 t 重折通道

开始优化过程时，按照某个概率分布 $p(\boldsymbol{\theta})$随机初始化旋转角。随后，初始化后的量子线路可表示为酉系综 $\mathbb{U} = \{U, \mathbb{P}(U)\}$，其中 $\mathbb{P}$是定义在 $\mathbb{U}$上的概率测度。人们通常关注计算关于 U各元素的某个给定阶数 t 的多项式量的平均值。这类量完全由对 t 重通道 [56]的了解所决定

$$\Phi_{\mathbb{U}}^{(t)}(\hat{\rho}) = \int_{\mathbb{U}} \hat{U}^{\otimes t}\hat{\rho}\hat{U}^{\dagger\otimes t}\mathrm{d}\mathbb{P}(\hat{U}), \tag{9}$$

其中，$\hat{\rho}$ 是原始 n量子比特系统的 t 个副本的初始态。举例来说，初始化时代价函数平方的期望值可通过 $\mathbb{E}_{\boldsymbol{\theta}}\left[C(\boldsymbol{\theta})^2\right] = \mathrm{Tr}\left[\Phi_{\boldsymbol{\theta}}^{(2)}(\hat{\rho}^{\otimes 2})\hat{O}^{\otimes 2}\right]$ 求得，其中我们将酉系综 $\{\hat{U}(\boldsymbol{\theta}) \in \mathbb{R}^M \ni \boldsymbol{\theta} \sim p(\boldsymbol{\theta})\}$ 所关联的 t重通道记为 $\Phi_{\boldsymbol{\theta}}^{(t)}$。在附录 B 中，我们将可由 2 重通道（分别为 1 重通道）的信息求得的量定义为二阶量（分别为一阶量）。举一个具体例子，$\mathbb{E}_{\boldsymbol{\theta}}\left[C(\boldsymbol{\theta})^2\right]$ 是一个二阶量。

更一般地，可以通过比较给定拟设的 t 重折叠通道与在整个酉群 [11, 57, 58] 上的 Haar（均匀）分布所得到的通道，来刻画其表达能力。t 重折叠通道与 Haar 测度下的 t 重折叠通道相匹配的酉系综，即所谓的 t-设计 [59, 60], ，在最初发现贫瘠高原现象 [10] 的过程中发挥了关键作用。此外，在多种情况下，随机量子线路都是近似的 t-设计 [61–63]。例如，文献 [61] 的作者证明，由多项式数量的量子门构成的量子线路，若这些门从一组通用双量子比特门中随机抽取，并作用于随机的量子比特对，则该线路是近似的 2-设计。这一结果已扩展到量子门作用于最近邻量子比特的情形，见 [62] 和 [63]。

C. 贫瘠高原

对于描述具有随机连续参数 $\boldsymbol{\theta}$ 且架构可能随机的参数化拟 ¨设 $U(\boldsymbol{\theta})$ 的酉系综 $\mathbb{U}$，若代价函数 $\hat{C}(U(\boldsymbol{\theta}))$ 获得的梯度偏离零值至少 $\epsilon > 0$ 的概率随系统规模 n 增大而呈指数级趋于零，则称该代价函数表现出贫瘠高原。更准确地说，对于某个 $\alpha > 0$ [11]，有 $\mathbb{P}_{\mathbb{U}}(|\partial_k C| > \epsilon) \leq \mathcal{O}(\exp(-\alpha n))$。如前所述，贫瘠高原最初是在形成 2-设计 [10] 的酉系综中发现的；后来，人们还发现了它与表达能力 [11], 、噪声 [12], 、纠缠 [13, 16–20] 以及代价函数局域性程度 [14, 15] 之间的联系。在许多情况下，梯度的平均值恰好为零，例如，当旋转参数在 $[-\pi, \pi]$ 上均匀初始化时。然而，这并不意味着梯度幅度的平均值为零，因此也无法充分说明模型的可训练性。在这种无偏情形下，方差是一个相关的量。根据切比雪夫不等式，有 $\mathbb{P}_{\mathbb{U}}(|\partial_k C| > \epsilon) \leq \mathrm{Var}\left[\partial_k C\right]/\epsilon^2$，因此方差趋于零意味着存在贫瘠高原。另一方面，方差不为零则保证初始梯度存在较大波动，从而初始可训练性良好，而与梯度偏差无关。

对于变分量子算法，梯度需要通过在硬件平台上进行的测量来估计。正如文献 [64], 所指出的，探测指数级小的梯度需要测量具有指数级精度，因此需要进行指数级数量的实验，这在实际中难以实现。贫瘠高原通常被视为经典机器学习中梯度消失现象的量子版本。文献 [65] 讨论了梯度消失对经典深度神经网络可训练性的影响。文献 [66] 的作者在量子机器学习背景下，采用最小二乘损失函数并依托神经切线核形式，进一步研究了贫瘠高原的确切性质。具体而言，作者讨论了贫瘠高原与经典梯度消失之间的根本区别，并指出，在某种过参数化区间内，训练过程可能对噪声具有鲁棒性。还需注意，变分量子算法除了会遭遇贫瘠高原之外，也可能面临其他问题，例如与损失景观中的局部极小值数量有关的问题 [67]，或与经典优化过程复杂度有关的问题 [68]。此外，高阶矩或许有助于诊断变分量子算法的可训练性 [69]。

III. 结果与讨论

这项理论工作的主要发现是，在对旋转参数 θ 的分布作出相当一般的假设时，可以将一重——

折叠通道和随机旋转 $\hat{} R_Z(\theta)$ 的二重通道映射为由 Clifford 门构成的有限酉系综。此外，我们证明，这种映射使我们能够仅使用 Clifford 电路来估计梯度方差等感兴趣的量。最后，我们通过数值实验展示了严格的证明。详细的数学证明见附录。

A. 精确映射与高效采样

如前所述，我们将重点研究由固定 Clifford 门与沿以下方向的单量子比特参数化旋转交替组成的变分量子电路 X、Y 或 Z；图1展示了一个例子。如II A 节所述，我们将研究范围限制在沿 Z方向的旋转，因为只需在所考虑拟设的不同固定层中添加额外的 Clifford 门，便可得到沿 Y 和 X 方向旋转的情形。考虑沿 Z轴的旋转，其概率分布关于 $\theta = 0$角 [70]对称。附录A 证明，与一阶平均对应的一重通道可以写成恒等算符和 Pauli Z 门所对应酉通道的凸和，如图2上半部分所示。值得注意的是，针对 [71]服从均匀概率分布的情形，这一结果已在 θ 中推导并用于通过 ZX演算来分析变分试探态。

要计算具有独立旋转参数、形式如式（2）所示的随机初始化拟设的一重通道，只需将与各旋转相关的一重通道组合起来，并在其中交错插入与固定门相关的酉通道 W_k。我们发现，该拟设的一重通道是 2^M 个 Clifford 酉通道的凸和，其中 M 为旋转的数量。可以将这个凸和视为对一个有限的 Clifford 近似电路集合取平均。附录 F 中给出了这类电路的示例，其架构简单，类似于图1所示。尽管该集合中 Clifford 近似电路的数量随参数数量呈指数增长，但我们在附录 C2 中证明，只需数量关于 M/ϵ^2 呈多项式增长的样本，就足以将可观测量期望值（或更一般地，任意一阶量）的平均值近似到任意所需的精度 ϵ。这一结果基于经典集中性论证，并在图 1 中以一个一阶简单电路作了示意。由此可以估计梯度的期望值，因为只需在一重通道的定义中用 $U(\boldsymbol{\theta})$ $\hat{}$替换 $U_\pm(\boldsymbol{\theta})$（定义见第 IIA 节），即可得到 $C(\boldsymbol{\theta} \pm \pi/2)$ 的期望值。根据参数移位规则，这就得到了梯度的期望值。

我们在附录A 中证明，与随机 Z旋转相关联的二重通道也是克利福德通道的线性组合，前提是该概率分布是 θ的偶函数。该结果如图 2(b) 所示。为得到这一映射，我们使用量子通道的 Choi 表示 [49]。由 Z 旋转门的张量积构成的酉量子通道，其 Choi 算子为对角矩阵。因此，可以用相应 Choi 算子的对角系数来表示这些通道。利用这种表示以及期望的线性性，我们得到了 Z 旋转的平均二重通道的一种易于处理的表示。随后，通过求解一个线性方程组来推导该分解；该方程组是通过将前述表示中不同通道的系数对应起来得到的。当不等式 $\mathbb{E}_\theta[f_+(\theta)] \geq 0$ 和 $\mathbb{E}_\theta[f_-(\theta)] \geq 0$ 与$f_\pm(\theta) = \cos\theta(\cos\theta \pm 1)$ 成立时，前述线性组合实际上是一个凸和。等价地，如果 Z具有偶角度概率分布，则其二重通道是克利福德混合酉通道

$$\mathbb{E}_\theta\left[\cos^2\theta\right] \geq \left|\mathbb{E}_\theta\left[\cos\theta\right]\right|. \tag{10}$$

f_+、f_- 的零点是角度 $\{k\pi/2, k \in \mathbb{Z}\}$，在这些角度下，$R_Z(\theta)$ 与一个克利福德门相符（见附录A）。事实上，如果 θ 的分布是这些角度处狄拉克分布的凸和，那么对 θ 的平均就变成了相应克利福德幺正算子的离散平均。因此，关联的二重通道确实是克利福德通道的凸和。还可以验证，对于关于角度 θ 为偶且以 π为周期的分布，前述条件也成立。例如，均匀分布就属于此类。对于中心高斯分布，当且仅当相应的宽度足够大时，前述条件才成立。

只要旋转角的分布满足上述条件，该方案就可以扩展到二阶，从而利用一组克利福德近似量子线路来近似二阶量，例如代价函数平方的平均值 $\mathbb{E}_{\boldsymbol{\theta}}\left[C(\boldsymbol{\theta})^2\right]$。根据参数移位规则，平方梯度的期望值可以通过以下四个量计算得到：$\mathbb{E}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta} \pm (\pi/2)\boldsymbol{e}_k)C(\boldsymbol{\theta} \pm(\pi/2)\boldsymbol{e}_k)]$。通过将二重通道定义中的$U^{\otimes 2}$ 项替换为$U_\pm \otimes U_\pm$，就可以用克利福德近似来估计后者。因此，该方案涵盖了梯度方差的估计。请注意，在二阶情况下，近似量子线路是通过将旋转二重通道替换为图2所示的四个双量子比特 Clifford 门之一而得到的，由此产生 4^M 个可能的克利福德量子线路。与一阶量相同，样本数随 M 线性增长即可保证收敛。以下定理总结了这些严格的结果，详细证明见附录A和C。

定理 III.1。 对于由固定的 *Clifford* 门和沿 X、Y 或 Z 方向的 M 参数化旋转构成的变分试探态，如果随机变分参数 $(\theta_1, \ldots\ldots, \theta_M)$ 彼此独立，且关于某个 *Clifford*角对称，即 $\in$

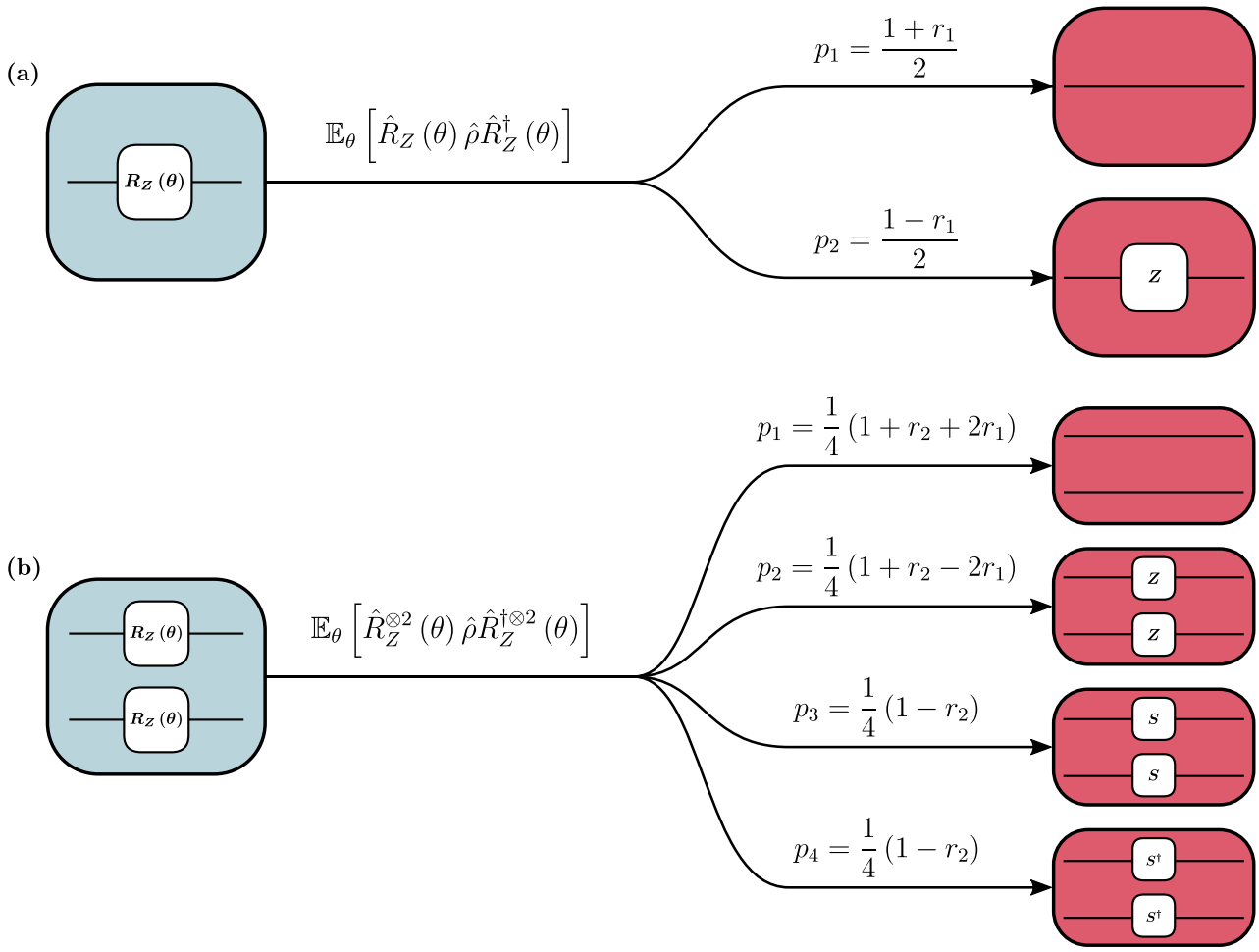


图 2. 从带有随机参数的 Z-旋转到由 Clifford 门组成的酉系综的映射规则示意图。图中 (a) 和 (b) 分别对应一阶和二阶平均。此处的映射适用于关于 θ 为偶函数的概率分布：我们记 $r_1=\mathbb{E}_\theta[\cos(\theta)]$ 和 $r_2=\mathbb{E}_\theta[\cos(2\theta)]$。系数 p_i 是决定如何对相应 Clifford 电路进行采样的概率。

$\{0,\frac{\pi}{2},\pi,\frac{3\pi}{2}\}$，那么对于任意误差 $\epsilon>0$ 以及达到此精度的概率 $1-\delta$，任意一阶量都可以通过以下方式计算

$$K\geq O\left(\frac{M}{\epsilon^2}\log\left(\frac{2}{\delta}\right)\right)$$

Clifford 近似电路。此外，如果 θ_i 的分布满足不等式

$$\mathbb{E}_{\theta_i}\left[\cos^2(\theta_i-\mathbb{E}_{\theta_i}[\theta_i])\right]\geq\left|\mathbb{E}_{\theta_i}[\cos(\theta_i-\mathbb{E}_{\theta_i}[\theta_i])]\right|,\quad\forall i\in\{1,\ldots,M\}$$

那么任何二阶量也满足同样的条件。

最后利用 Gottesman-Knill 定理，该定理指出，对于 Clifford 酉算符 U 和一个在 N_O 个量子比特上非平凡作用的可观测量 O，期望值 $\mathrm{Tr}\left[|0\rangle\langle 0|^{\otimes n}\hat{U}^\dagger\hat{O}\hat{U}\right]$ 可以通过经典方法计算，其复杂度关于 n 和 N_O 均为多项式复杂度。我们的方法继承了这一复杂度，特别是，我们可以通过经典方法估计梯度期望值和方差，其复杂度关于 n、N_O 和 M 均为多项式复杂度。

在附录 D 中，我们将上述针对二重通道提出的方案扩展到随机角度的分布不满足式（10）凸条件的情形。在这种情况下，仍然可以使用 Clifford 近似电路来估计二阶量，但代价是关于变分参数数量的指数复杂度 M。该结果基于 [72] 的作者提出的一种采样方案。在该背景下，该方法可以将关于系统规模的指数复杂度转换为关于变分参数数量的指数复杂度M。

在附录 E中，我们还将方案扩展到一般情形下的 N 重折叠通道。我们证明，与随机 Z 旋转相关的 N 重折叠通道可分解为 Clifford 酉通道的实系数和。基于这一分解，我们推导出以下条件：

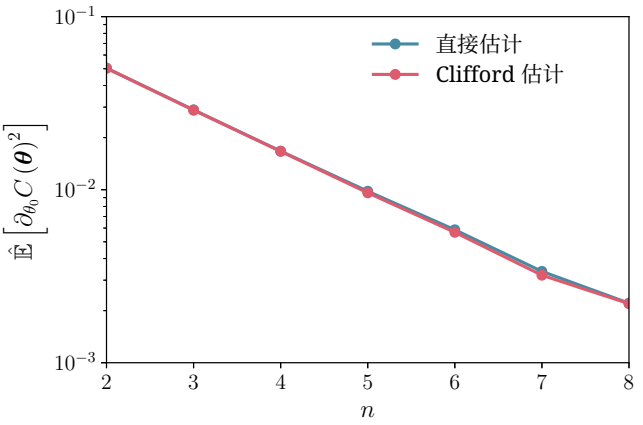


图 3. 代价函数关于第一个变分参数的梯度平方的估计平均值与比特数的关系 n。我们强调，关于其他角度的导数 θ_k 也会得出类似结果（未显示）。结果针对由单层量子门组成的随机电路，每个量子比特有一次旋转。这些旋转从 R_X、R_Y、R_Z中随机选取。旋转层后接一层交替的 CZ 量子门（注意，这与图1所示的架构类型相同）。随机旋转角相互独立，并服从区间 $[0, 2\pi]$ 上的均匀概率分布。为进行估计，我们随机抽取了 500 种不同的电路架构。对于每种量子门架构，我们假设旋转角服从均匀分布，并分别使用直接估计和基于映射到 Clifford 近似电路的方法，计算梯度平方的平均值。具体而言，对于直接估计，我们抽取了 500 个角度参数向量；对于我们的方法，则抽取了 500 个 Clifford 电路。注意，对于均匀分布，平均梯度为零，因此估计梯度平方等同于估计梯度方差。

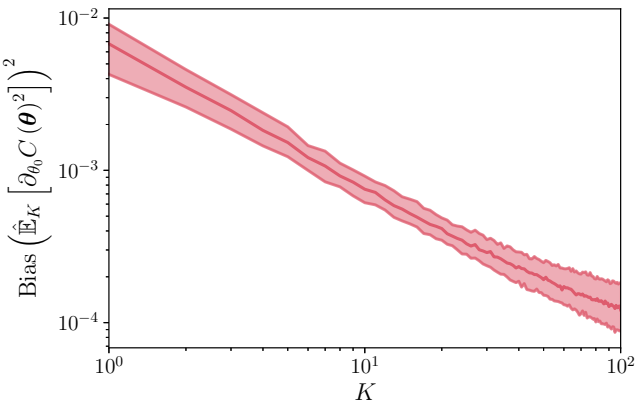


图 4. 图 3 所示估计量的统计偏差平方，针对含 $n = 5$ 个量子比特的随机电路，随 K 个 Clifford 近似电路数量的变化。结果基于 500 种随机抽取的电路架构得到。

对于每个样本量 K，我们考虑一个包含 100 个估计量的自助法批次（每个估计量均通过以下方式得到：针对每种旋转方向选择，从一组 2000 个 Clifford 近似电路中抽取 K 个电路）。随后，对于每个 K，我们根据自助法批次计算统计偏差。估计量的真实期望值通过使用 4000 个样本直接估计梯度平方的平均值获得。阴影区域表示 500 种随机架构下估计偏差的第 20 至第 80 百分位数区间。

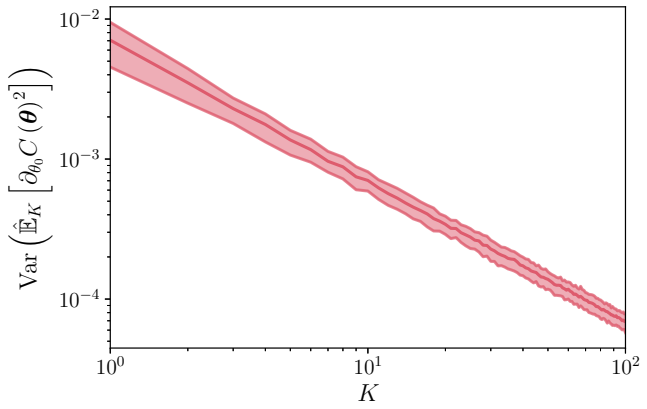


图5. 关于第一个参数的期望平方梯度估计量的方差 θ_0 随 K Clifford 近似电路数量的变化。随机电路类型与图3 中相同，包含 $n = 5$ 个量子比特。我们使用了与图4中相同的自助法程序。阴影区域对应于 500 个随机架构的估计偏差的第 20 至第 80 百分位数之间的区间。

要使 N 重折叠通道成为 Clifford 酉算子的凸和，组合中的系数必须为正。然而，我们证明所得分解并不唯一，因此推导出的条件是充分但非必要的。如何找到关于随机角度分布的充要条件，以保证相应的 N 重折叠通道是 Clifford 混合酉信道，仍是一个开放问题。

B. 数值模拟

为说明精确映射及由此得到的估计方法的应用，我们针对具体示例进行了数值实验。考虑一种简单的变分量子电路，它由若干层单量子比特旋转组成，旋转方向沿 X、Y 或Z 轴，各层之间交替排列着由受控 Z 门构成的固定层。图1展示了这种拟设在三个量子比特上的形式。我们进一步假设，旋转角彼此独立且服从相同的均匀分布

在 $[0, 2\pi]$之上。此外，我们假设代价函数具有方程（1）中的形式，其中 $\hat{O} = |0\rangle\langle 0|^{\otimes n}$。

我们考虑这些架构，其中旋转门的方向是随机的。除了第一层固定但不同之外，已有研究表明此类随机电路会出现贫瘠高原，见文献 [64]。请注意，在这个特定情况下，求平均时不仅对旋转角

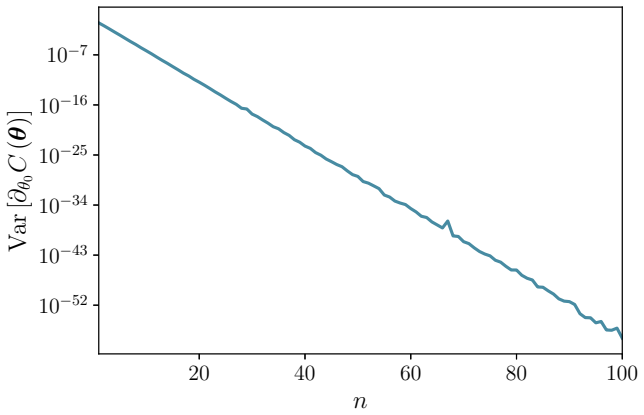


图 6. 相对于第一个变分参数的代价函数梯度的估计方差与量子比特数量的关系 n 。每个方差均使用 10^5 Clifford 电路进行估计。所使用的量子线路和代价函数与图3相同。

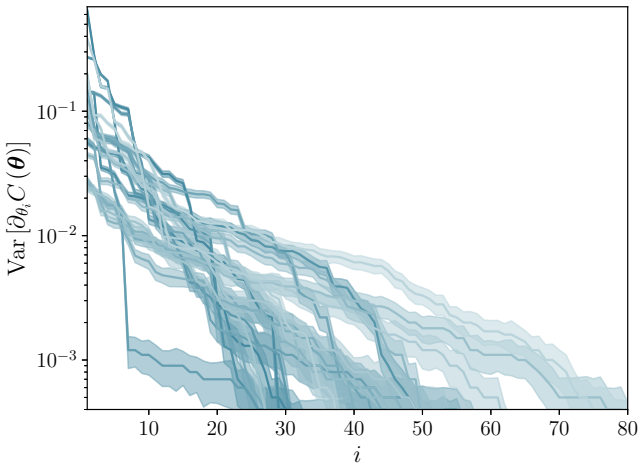


图 7. 20 个随机电路中，代价函数相对于变分参数的梯度方差，按降序排列。这些电路作用于 $n = 40$ 个量子比特。每个电路由 10 层组成。对于每一层 l，在区间 $[0, n]$内随机选择一个数 m_l ，然后先对随机选取的量子比特施加 m_l 个随机单量子比特旋转门，再施加一系列纠缠门；这些纠缠门的类型和排列方式与图 3中所考虑的相同。对于每个电路及其每个参数 θ_i，使用 10^4 Clifford 近似电路估算梯度。代价函数的哈密顿量是 10 个随机泡利串之和。请注意，每条曲线对应不同的架构。

，也对旋转方向进行了平均。这里我们使用 Clifford 近似重现这一结果。为此，我们对精确的电路架构进行采样：从 $\{X, Y, Z\}$中均匀随机选取旋转方向；然后，我们直接对旋转角进行采样，或对 Clifford 近似电路进行采样。对于均匀分布，有 $\forall k \in \mathbb{Z},\ \ \mathbb{E}_\theta\left[\text{余弦}\, k\theta\right] = 0$ 因此，替代用 Clifford 门的采样是均匀的（如图2 所示 $r_1 = r_2 = 0$）。此外，根据参数移位规则（式3），显然，对于均匀分布的旋转，平均梯度在解析上为零，因此只需估计梯度平方的平均值，即 $\mathrm{Var}_{\boldsymbol{\theta}}\left[\partial_k C(\boldsymbol{\theta})\right] = \mathbb{E}_{\boldsymbol{\theta}}\left[\partial_k C(\boldsymbol{\theta})^2\right]$。

在图3中展示了通过直接评估或采样 Clifford 近似得到的平均梯度平方估计值。请注意，平均值同时取自随机旋转角和可变架构（即旋转门的随机方向）。由 Clifford 近似得到的估计值与直接估计值准确吻合，并且平均梯度平方随量子比特数量增加呈指数衰减，正如预期。此外，Clifford 估计的偏差随近似量子线路数量的变化 K 如图4所示。偏差随 K呈多项式下降。如图5所示，Clifford 估计量的方差也呈现相同趋势。这些结果与附录C中推导出的解析标度一致。

使用我们的方法，我们还复现了参考文献中展示的结果，即梯度方差呈指数级抑制 [64] ，且最多适用于 100 个量子比特。结果如图6所示。

最后，我们通过评估一组随机生成的拟设以及给定代价函数哈密顿量对应的梯度方差，说明拟设架构对其可训练性的影响。我们考虑作用于 40 个量子比特的随机电路，以及由 10 个随机选取的泡利串之和构成的哈密顿量。与图 3所示结果类似，所考虑的变分量子线路由旋转层与固定的纠缠层交替组成。此处我们考虑具有 10 层的量子线路。对于每个旋转层，随机选取一部分旋转，并将其替换为恒等门（详情见图 7 ）。代价函数对拟设参数的偏导数方差如图 7所示；图 8则展示了不同随机电路中梯度的平均方差与代价函数均值之间的关系。这些结果表明，对于给定的代价函数，改变拟设架构会对可训练性产生显著影响。因此，我们认为，该方法可用于以合理的成本系统地考察大型系统中的此类影响，进而指导设计更优的变分拟设。

IV. 结论与展望

本文提出了一种经典高效的方法，用于估计一大类随机初始化的变分量子线路的一阶和二阶期望值。这包括估计代价函数的平均梯度及其方差，而这些量可用于评估可训练性。我们的方法适用

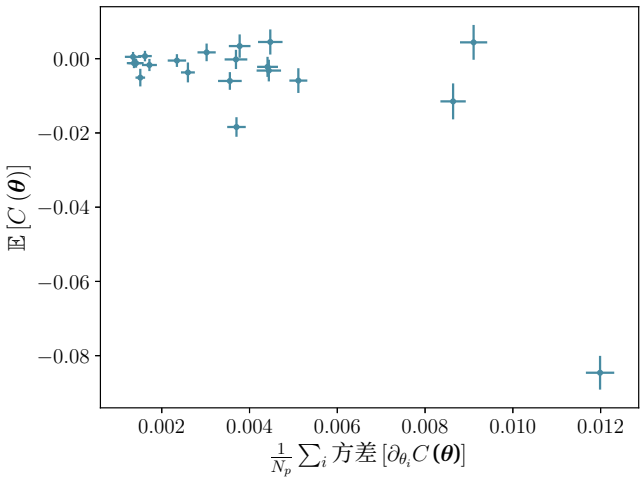


图 8. 20 个随机电路和 40 个量子比特的代价函数均值与代价函数梯度方差的关系。每个点代表一个电路，所考虑的电路与图 7中的电路相同。x 轴表示相对于 N_p 电路旋转参数 θ_i，$i = 1, \ldots\ldots, N_p$。

于一大类量子线路，其架构由固定的 Clifford 门和单量子比特参数化旋转构成，前提是旋转角彼此独立，且其分布关于角度 $\theta_0 \in \{k\pi/2,\ k \in \mathbb{Z}\}$ 对称，并满足 $\mathbb{E}_\theta\left[\cos^2(\theta-\theta_0)\right] \geq |\mathbb{E}_\theta[\cos(\theta-\theta_0)]|$。该方法依赖于将随机初始化的变分量子线路精确映射为 Clifford 电路系综，并运用 Gottesman-Knill 定理。我们给出了严格的收敛保证，特别是证明该方法的复杂度随系统规模和所研究拟设的参数数量均呈多项式增长。我们研究了将所提方案推广至 N 重通道的情形，并证明随机 Z 旋转的 N 重平均可表示为 Clifford 酉算子的实线性组合。然而，这种分解并不唯一；要为所研究的 N 重通道找出成为 Clifford 混合酉通道的充分必要条件，仍是一个开放问题。解决这一问题意义重大，因为这或许能够将本文提出的方案推广到具有相关变分参数的拟设。

我们相信，这种工具在未来的应用中将非常有用，因为它可用于对大规模变分量子电路的架构和初始化进行经典优化。由于无论潜在贫瘠高原的确切成因是什么，只要梯度的方差足够大，就能保证不存在贫瘠高原，因此这种方法可用于认证具有大量量子比特的系统的可训练性。

致谢

ˆ 本研究得到法兰西岛大区的支持，资助框架为法兰西岛量子技术科学与工程重点领域（DIM SIRTEQ）。本研究获准使用法国国家高性能计算设备（GENCI）提供的资源，分配编号为 2022-A0120512462，计算资源来自特雷斯大型计算中心（TGCC）的高性能计算（HPC）资源。我们感谢 Zakari Denis 在本研究早期阶段提供的有益讨论。

附录 A：随机 Z 旋转的一重通道和二重通道

1. 一重通道

这里给出绕 Z 轴的单量子比特旋转的一重通道表达式。绕 X 轴和 Y 轴的旋转可通过与哈达玛门和相位门组合得到。定义 ˆ ˆ$\Pi_0 := |0\rangle\langle 0|$ 和 $\Pi_1 := |1\rangle\langle 1|$:

$$\begin{aligned}\hat{R}_Z(\theta) &= e^{-\mathrm{i}\frac{\theta}{2}}\hat{\Pi}_0 + e^{\mathrm{i}\frac{\theta}{2}}\hat{\Pi}_1 \\ \hat{R}_Z(\theta)\hat{\rho}\hat{R}_Z^\dagger(\theta) &= \hat{\Pi}_0\hat{\rho}\hat{\Pi}_0 + \hat{\Pi}_1\hat{\rho}\hat{\Pi}_1 \\ &\quad + e^{\mathrm{i}\theta}\hat{\Pi}_1\hat{\rho}\hat{\Pi}_0 + e^{-\mathrm{i}\theta}\hat{\Pi}_0\hat{\rho}\hat{\Pi}_1.\end{aligned} \tag{A1}$$

Thus

$$\begin{aligned}\mathbb{E}_\theta\left[\hat{R}_Z(\theta)\hat{\rho}\hat{R}_Z^\dagger(\theta)\right] &= \hat{\Pi}_0\hat{\rho}\hat{\Pi}_0 + \hat{\Pi}_1\hat{\rho}\hat{\Pi}_1 \\ &\quad + \mathbb{E}_\theta\left[e^{\mathrm{i}\theta}\right]\hat{\Pi}_1\hat{\rho}\hat{\Pi}_0 \\ &\quad + \mathbb{E}_\theta\left[e^{-\mathrm{i}\theta}\right]\hat{\Pi}_0\hat{\rho}\hat{\Pi}_1.\end{aligned} \tag{A2}$$

我们认出这是 θ 分布的特征函数，即

$$\phi(t) := \mathbb{E}_\theta\left[e^{\mathrm{i}t\theta}\right].$$

假设这个概率分布关于 θ 为偶分布，则有 $\phi(t) \in \mathbb{R}, \forall t$，并且我们可以定义 $r_1 = \phi(1) = \phi(1)^* =$ˆ ˆ ˆ ˆ ˆ$\phi(-1)$。由于我们有 $\mathbb{1} = \Pi_0 + \Pi_1$ 和 $Z = \Pi_0 - \Pi_1$，因此得到

$$\begin{aligned}\hat{\rho} &= \left(\hat{\Pi}_0\hat{\rho}\hat{\Pi}_0 + \hat{\Pi}_1\hat{\rho}\hat{\Pi}_1\right) + \left(\hat{\Pi}_1\hat{\rho}\hat{\Pi}_0 + \hat{\Pi}_0\hat{\rho}\hat{\Pi}_1\right), \\ \hat{Z}\hat{\rho}\hat{Z} &= \left(\hat{\Pi}_0\hat{\rho}\hat{\Pi}_0 + \hat{\Pi}_1\hat{\rho}\hat{\Pi}_1\right) - \left(\hat{\Pi}_1\hat{\rho}\hat{\Pi}_0 + \hat{\Pi}_0\hat{\rho}\hat{\Pi}_1\right),\end{aligned} \tag{A3}$$

因此

$$\frac{1+r_1}{2}\hat{\rho} + \frac{1-r_1}{2}\hat{Z}\hat{\rho}\hat{Z} = \mathbb{E}_\theta\left[R_Z(\theta)\hat{\rho}R_Z^\dagger(\theta)\right]. \tag{A4}$$

这确实是克利福德通道的凸和，条件是 $r_1 \in [-1, 1]$，而该条件始终成立。

对于关于 Clifford角 $\in \{k\pi/2, k \in \{0,1,2,3\}\}$ 对称的分布，我们可以提出相应的旋转（除一个相位外，它是一个克利福德门）。这样，我们就可以回到无偏偶分布的情形，即关于零角度对称的分布。注意，对于 $[0,2\pi]$ 上的均匀分布这一特殊情形，我们有 $r_1 = 0$。

2. 二重通道

本节将使用量子通道的 Choi 表示，它可以用 16×16 矩阵表示作用于双量子比特态的通道。对于量子信道（即完全正且保迹映射，或 CPTP 映射） $\mathcal{E}$，Choi算子定义如下：

$$\Lambda(\mathcal{E}) = \sum_{i,j,k,l=0}^{1} |ij\rangle\langle kl| \otimes \mathcal{E}\left(|ij\rangle\langle kl|\right). \tag{A5}$$

其对应的矩阵元素为：

$$\begin{aligned}\Lambda(\mathcal{E})_{(ijkl),(mnpq)} =& \mathrm{Tr}\left[\Lambda(\mathcal{E})^\dagger \left(|ij\rangle\langle kl| \otimes |mn\rangle\langle pq|\right)\right] \\ =& \mathrm{Tr}\left[\mathcal{E}\left(|ij\rangle\langle kl|\right)^\dagger |mn\rangle\langle pq|\right]\end{aligned} \tag{A6}$$

下文中，我们将用

$$\mathcal{E}[\hat{U}](\hat{\rho}) := \hat{U}\hat{\rho}\hat{U}^\dagger \tag{A7}$$

表示与酉变换 U相关联的量子信道。我们假设 U 在计算基下是对角的，因此可以写成

$$\hat{U} = \sum_{i,j=0}^{1} \lambda_{ij}\hat{\Pi}_{ij}, \tag{A8}$$

其中，我们定义投影算子 $\Pi_{ij} := \Pi_i \otimes \Pi_j$。对于 U 酉算子，有 $\hat{U}\hat{U}^\dagger = \mathbb{1} = \sum_{i,\ j} \lambda_{ij}\lambda_{ij}^* \hat{\Pi}_{ij}$，因此 $\lambda_{ij} = e^{\mathrm{i}\theta_{ij}}, \quad \forall i,\ j$。由此可得

$$\begin{aligned}\mathcal{E}[\hat{U}](|ij\rangle\langle kl|) &= \sum_{m,n,p,q} \lambda_{mn}\lambda_{pq}^* \hat{\Pi}_{mn}\,|ij\rangle\langle kl|\,\hat{\Pi}_{pq} \\ &= \lambda_{ij}\lambda_{kl}^*\,|ij\rangle\langle kl| \\ &= e^{\mathrm{i}(\theta_{ij}-\theta_{kl})}\,|ij\rangle\langle kl|\,.\end{aligned} \tag{A9}$$

因此， $\mathcal{E}[\hat{U}]$ 的 Choi矩阵U 在 A8式所示的形式时是对角矩阵。我们可以用一个4×4 矩阵 M来表示，其元素定义如下：

$$M_{(ij),(kl)} := \Lambda_{(ijkl),(ijkl)}. \tag{A10}$$

注意，矩阵 M 是厄米矩阵，并且由于式（A9），其对角元素始终等于 1。下文中，我们将用其对应的矩阵 M 在基 (00)、(01)、(10)、(11) 表示每个通道。

与前文一样，我们将聚焦于绕 Z 轴的旋转。我们有

$$\Phi_Z^{(2)}(\hat{\rho}) := \mathbb{E}_\theta\left[(\hat{R}_Z(\theta)\otimes\hat{R}_Z(\theta))\hat{\rho}(\hat{R}_Z^\dagger(\theta)\otimes\hat{R}_Z^\dagger(\theta))\right] \tag{A11}$$

and

$$\begin{aligned}\hat{R}_Z(\theta)\otimes\hat{R}_Z(\theta) =& \left(e^{-\mathrm{i}\theta}\hat{\Pi}_0\otimes\hat{\Pi}_0 + e^{\mathrm{i}\theta}\hat{\Pi}_1\otimes\hat{\Pi}_1\right) \\ &+ \left(\hat{\Pi}_0\otimes\hat{\Pi}_1 + \hat{\Pi}_1\otimes\hat{\Pi}_0\right).\end{aligned} \tag{A12}$$

定义

$$\begin{aligned}\Gamma_\theta &= \left(e^{-\mathrm{i}\theta}\hat{\Pi}_{00} + e^{\mathrm{i}\theta}\hat{\Pi}_{11}\right), \\ \Xi &= \hat{\Pi}_{01} + \hat{\Pi}_{10},\end{aligned} \tag{A13}$$

我们可以写成

$$\begin{aligned}\Phi_Z^{(2)}(\hat{\rho}) =& \mathbb{E}_\theta\left[\Xi\hat{\rho}\Xi^\dagger\right] + \mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Gamma_\theta^\dagger\right] \\ &+ \mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Xi^\dagger\right] + \mathbb{E}_\theta\left[\Xi\hat{\rho}\Gamma_\theta^\dagger\right].\end{aligned} \tag{A14}$$

a. *Uniform distribution*

对于 $[0,2\pi]$中 θ 的均匀分布，我们有 $\mathbb{E}_\theta\left[e^{\pm\mathrm{i}\theta}\right] = \mathbb{E}_\theta\left[e^{\pm2\mathrm{i}\theta}\right] = 0$，因此：

$$\begin{aligned}\mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Xi^\dagger\right] &= 0, \\ \mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Gamma_\theta^\dagger\right] &= \hat{\Pi}_{00}\hat{\rho}\hat{\Pi}_{00} + \hat{\Pi}_{11}\hat{\rho}\hat{\Pi}_{11}, \\ \mathbb{E}_\theta\left[\Xi\hat{\rho}\Xi^\dagger\right] &= \hat{\Pi}_{01}\hat{\rho}\hat{\Pi}_{01} + \hat{\Pi}_{10}\hat{\rho}\hat{\Pi}_{10}, \\ &\quad + \hat{\Pi}_{01}\hat{\rho}\hat{\Pi}_{10} + \hat{\Pi}_{10}\hat{\rho}\hat{\Pi}_{01}.\end{aligned} \tag{A15}$$

最后，我们得到

$$\begin{aligned}\Phi_Z^{(2)}(\hat{\rho}) =& \hat{\Pi}_{00}\hat{\rho}\hat{\Pi}_{00} + \hat{\Pi}_{11}\hat{\rho}\hat{\Pi}_{11} + \hat{\Pi}_{01}\hat{\rho}\hat{\Pi}_{01} \\ &+ \hat{\Pi}_{10}\hat{\rho}\hat{\Pi}_{10} + \hat{\Pi}_{01}\hat{\rho}\hat{\Pi}_{10} + \hat{\Pi}_{10}\hat{\rho}\hat{\Pi}_{01}.\end{aligned} \tag{A16}$$

我们可以用其对应的矩阵表示 $\Phi_Z^{(2)}$

$$M(\Phi_Z^{(2)}) = \begin{pmatrix}1&0&0&0\\0&1&1&0\\0&1&1&0\\0&0&0&1\end{pmatrix}. \tag{A17}$$

可以验证，以下通道也具有对角 Choi矩阵，我们可以使用相同的表示法——

对角线的……，得到

$$
\begin{aligned}
M(\mathcal{E}[\mathbb{1}]) &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \\
M(\mathcal{E}[\hat{Z}\otimes\hat{Z}]) &= \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}, \\
M(\mathcal{E}[\hat{S}\otimes\hat{S}]) &= \begin{pmatrix} 1 & \mathrm{i} & \mathrm{i} & -1 \\ -\mathrm{i} & 1 & 1 & \mathrm{i} \\ -\mathrm{i} & 1 & 1 & \mathrm{i} \\ -1 & -\mathrm{i} & -\mathrm{i} & 1 \end{pmatrix}, \\
M(\mathcal{E}[\hat{S}^\dagger\otimes\hat{S}^\dagger]) &= \begin{pmatrix} 1 & -\mathrm{i} & -\mathrm{i} & -1 \\ \mathrm{i} & 1 & 1 & -\mathrm{i} \\ \mathrm{i} & 1 & 1 & -\mathrm{i} \\ -1 & \mathrm{i} & \mathrm{i} & 1 \end{pmatrix},
\end{aligned}
\tag{A18}
$$

ˆ与 $S=\Pi_0+\mathrm{i}\Pi_1$ 相位门。综合以上所有结果，我们得到

$$
\begin{aligned}
4M(\Phi_Z^{(2)}) = & M(\mathcal{E}[\mathbb{1}]) + M(\mathcal{E}[\hat{Z}\otimes\hat{Z}]) \\
& + M(\mathcal{E}[\hat{S}\otimes\hat{S}]) + M(\mathcal{E}[\hat{S}^\dagger\otimes\hat{S}^\dagger]).
\end{aligned}
\tag{A19}
$$

正文中的最终结果由 Choi矩阵的线性性和唯一性得出。

b. Even distribution

考虑 θ的偶概率分布（即满足 θ 与$-\theta$同分布的分布）。对于此类分布，我们同样有：对所有 $t\in\mathbb{R}$， $\phi_\theta(t)=\phi_\theta(-t)\in[-1,1]\subset\mathbb{R}$，因此

$$
\phi_\theta(t) = \frac{1}{2}\left(\phi_\theta(t)+\phi_\theta(-t)\right) = \mathbb{E}_\theta\left[\cos(t\theta)\right].
$$

定义 $r_1=\phi_\theta(1)$和 $r_2=\phi_\theta(2)$，我们可以写成

$$
\begin{aligned}
\mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Xi^\dagger\right] &= r_1\left(\hat{\Pi}_{00}+\hat{\Pi}_{11}\right)\hat{\rho}\left(\hat{\Pi}_{01}+\hat{\Pi}_{10}\right), \\
\mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Gamma_\theta^\dagger\right] &= \hat{\Pi}_{00}\hat{\rho}\hat{\Pi}_{00}+\hat{\Pi}_{11}\hat{\rho}\hat{\Pi}_{11} \\
&\quad + r_2\left(\hat{\Pi}_{00}\hat{\rho}\hat{\Pi}_{11}+\hat{\Pi}_{11}\hat{\rho}\hat{\Pi}_{00}\right), \\
\mathbb{E}_\theta\left[\Xi\hat{\rho}\Xi^\dagger\right] &= \hat{\Pi}_{01}\hat{\rho}\hat{\Pi}_{01}+\hat{\Pi}_{10}\hat{\rho}\hat{\Pi}_{10} \\
&\quad + \hat{\Pi}_{01}\hat{\rho}\hat{\Pi}_{10}+\hat{\Pi}_{10}\hat{\rho}\hat{\Pi}_{01}.
\end{aligned}
\tag{A20}
$$

因此我们得到:

$$
M(\Phi_Z^{(2)}) = \begin{pmatrix} 1 & r_1 & r_1 & r_2 \\ r_1 & 1 & 1 & r_1 \\ r_1 & 1 & 1 & r_1 \\ r_2 & r_1 & r_1 & 1 \end{pmatrix}. \tag{A21}
$$

我们可以将 $M(\Phi_Z^{(2)})$ 表示为式（A18）中矩阵的线性组合，得到

$$
\begin{aligned}
M(\Phi_Z^{(2)}) = & aM(\mathcal{E}[\mathbb{1}]) + bM(\mathcal{E}[\hat{Z}\otimes\hat{Z}]) \\
& + \frac{c}{2}\left(M(\mathcal{E}[\hat{S}\otimes\hat{S}]) + M(\mathcal{E}[\hat{S}^\dagger\otimes\hat{S}^\dagger])\right).
\end{aligned}
\tag{A22}
$$

系数 a,b,c 可通过求解线性方程组得到

$$
\begin{cases} a+b+c=1 \\ a-b=r_1 \\ a+b-c=r_2 \end{cases}, \tag{A23}
$$

从而可得

$$
\begin{aligned}
M(\Phi_Z^{(2)}) = & \frac{1}{4}\left(1+r_2+2r_1\right)M(\mathcal{E}[\mathbb{1}]) \\
& + \frac{1}{4}\left(1+r_2-2r_1\right)M(\mathcal{E}[\hat{Z}\otimes\hat{Z}]) \\
& + \frac{1}{4}\left(1-r_2\right)M(\mathcal{E}[\hat{S}\otimes\hat{S}]) \\
& + \frac{1}{4}\left(1-r_2\right)M(\mathcal{E}[\hat{S}^\dagger\otimes\hat{S}^\dagger]).
\end{aligned}
\tag{A24}
$$

因此，相应的通道为

$$
\begin{aligned}
\Phi_Z^{(2)}(\hat{\rho}) = & \frac{1}{4}\left(1+r_2+2r_1\right)\hat{\rho} \\
& + \frac{1}{4}\left(1+r_2-2r_1\right)\left(\hat{Z}\otimes\hat{Z}\right)\hat{\rho}\left(\hat{Z}\otimes\hat{Z}\right) \\
& + \frac{1}{4}\left(1-r_2\right)\left(\hat{S}\otimes\hat{S}\right)\hat{\rho}\left(\hat{S}^\dagger\otimes\hat{S}^\dagger\right) \\
& + \frac{1}{4}\left(1-r_2\right)\left(\hat{S}^\dagger\otimes\hat{S}^\dagger\right)\hat{\rho}\left(\hat{S}\otimes\hat{S}\right).
\end{aligned}
\tag{A25}
$$

ˆˆˆ注。 定义 $CZ=\Pi_0\otimes\mathbb{1}+\Pi_1\otimes Z$ 控制-Zˆˆˆˆ门和 $CZ_X=(X\otimes X)CZ(X\otimes X)$，我们有

$$
\begin{aligned}
M(\mathcal{E}[CZ]) &= \begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{pmatrix}, \\
M(\mathcal{E}[CZ_X]) &= \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{pmatrix},
\end{aligned}
\tag{A26}
$$

因此

$$
\mathcal{E}[\hat{S}\otimes\hat{S}] + \mathcal{E}[\hat{S}^\dagger\otimes\hat{S}^\dagger] = \mathcal{E}[CZ] + \mathcal{E}[CZ_X]. \tag{A27}
$$

因此，将 $\Phi_Z^{(2)}$ 分解为式（A25）中 Clifford 通道的凸和并非唯一。

式（A25）中得到的分解是凸和，当且仅当假设（$1+r_2-2r_1$）$\geq$ 0 且（$1+r_2+2r_1$）$\geq$0时成立。该条件当且仅当

$$
\mathbb{E}_\theta\left[\frac{1}{2}(1+\cos 2\theta)\pm\cos\theta\right]\geq 0,
$$

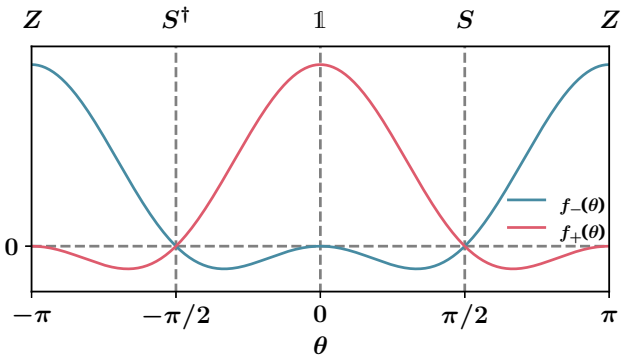


图 9。绘制 $f_{\pm}(\theta) = \cos\theta(\cos\theta \pm 1)$ 与 θ的关系图。当且仅当 $\mathbb{E}_\theta[f_+(\theta)] \geq 0$且 $\mathbb{E}_\theta[f_-(\theta)] \geq 0$. $\hat{}S$ 是相位门时，公式（A28）中的条件才得以满足。

也就是说，当且仅当

$$\mathbb{E}_\theta\left[\cos^2\theta\right] \geq |\mathbb{E}_\theta[\cos\theta]|. \tag{A28}$$

对于π-周期分布，该条件得以满足，因为在这种情况下我们有 $\mathbb{E}_\theta[\cos\theta] = 0$。另一个满足此约束的分布示例是方差足够大的高斯分布。事实上，对于方差为σ^2的中心高斯分布，我们有 $r_1 = e^{-\sigma^2/2}$ 和 $r_2 = e^{-2\sigma^2}$，因此该条件变为

$$1 + e^{-2\sigma^2} - 2e^{-\sigma^2/2} \geq 0. \tag{A29}$$

可以证明，对于某个特定的 $\sigma_0 \in \mathbb{R}$，该条件等价于 $\sigma^2 \geq \sigma_0^2$，从而对高斯分布的宽度提出了要求。

附录 B：一阶和二阶量

在本附录中，我们将一阶量和二阶量定义为：对于给定拟设中出现的每个随机旋转，只要分别知道其一重通道和二重通道，便可以得到的量。我们还将证明，平均代价函数和平均梯度是一阶量，而代价函数平方和梯度平方的平均值是二阶量。

1. 一阶量

考虑由下式定义的拟设

$$\hat{U}(\boldsymbol{\theta}) = \prod_{i=1}^{M} \hat{U}_i(\theta_i)\hat{W}_i, \tag{B1}$$

并记作

$$\begin{aligned} \mathcal{U}_i(\theta_i)(\hat{\rho}) &= \hat{U}_i(\theta_i)\hat{\rho}\hat{U}_i^\dagger(\theta_i) \\ \mathcal{W}_i(\hat{\rho}) &= \hat{W}_i\hat{\rho}\hat{W}_i^\dagger \end{aligned} \tag{B2}$$

与量子线路不同层相关联的酉通道。整个量子线路的幺正变换则可写为

$$\begin{aligned} \mathcal{U}(\boldsymbol{\theta})(\hat{\rho}) &= \mathcal{U}_M(\theta_M) \circ \cdots \circ \mathcal{W}_1(\hat{\rho}) \\ &= \bigcirc_{i=1}^{M} (\mathcal{U}_i(\theta_i) \circ \mathcal{W}_i)(\hat{\rho}). \end{aligned} \tag{B3}$$

代价函数为

$$C(\boldsymbol{\theta}) = \mathrm{Tr}\left[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})\hat{O}\right] \tag{B4}$$

其关于 $\boldsymbol{\theta}$ 的期望值为

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta})] &= \mathbb{E}_{\boldsymbol{\theta}}\left[\mathrm{Tr}\left[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})\hat{O}\right]\right] \\ &= \mathrm{Tr}\left[\mathbb{E}_{\boldsymbol{\theta}}[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})]\hat{O}\right] \\ &= \mathrm{Tr}\left[\mathbb{E}_{\boldsymbol{\theta}}\left[\hat{U}(\boldsymbol{\theta})\hat{\rho}\hat{U}^\dagger(\boldsymbol{\theta})\right]\hat{O}\right] \\ &= \mathrm{Tr}\left[\int_{\mathbb{R}^M} \hat{U}(\boldsymbol{\theta})\hat{\rho}\hat{U}^\dagger(\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}\hat{O}\right] \\ &= \mathrm{Tr}\left[\Phi_{\boldsymbol{\theta}}^{(1)}(\hat{\rho})\hat{O}\right]. \end{aligned} \tag{B5}$$

这里，我们同时使用了期望值的线性性质以及公式（9）中一重通道的定义。因此，只要知道完整的一重通道 $\Phi_{\boldsymbol{\theta}}^{(1)}$，便可以得到代价函数的期望值。假设角度 $\{\theta_i\}$ 彼此独立，则关于 $\boldsymbol{\theta}$ 的期望值可以分解为关于各个 θ_i 的期望值，从而可以写成：

$$\begin{aligned} \Phi_{\boldsymbol{\theta}}^{(1)}(\hat{\rho}) &= \mathbb{E}_{\boldsymbol{\theta}}[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})] \\ &= \bigcirc_{i=1}^{M} (\mathbb{E}_{\theta_i}[\mathcal{U}_i(\theta_i)] \circ \mathcal{W}_i)(\hat{\rho}). \end{aligned} \tag{B6}$$

如正文所述，不失一般性，我们可以认为所有旋转都是 Z 旋转。于是，通道 $\mathbb{E}_{\theta_i}[\mathcal{U}_i(\theta_i)]$ 恰好是与作用于单量子比特的 Z 旋转相关联的一重通道，因此可以根据附录 A中的结果计算得出。如前所述，我们将可通过了解拟设中每个旋转所关联的一重通道而获得的量称为一阶量。因此，平均代价函数 $\mathbb{E}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta})]$ 是一阶量。

另一个有趣的一阶量示例是梯度的平均值。根据参数移位规则，并利用期望值的线性性，我们有：

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}}[\partial_k C(\boldsymbol{\theta})] = &\frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}}\left[C\left(\boldsymbol{\theta} + \frac{\pi}{2}\boldsymbol{e}_k\right)\right] \\ &- \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}}\left[C\left(\boldsymbol{\theta} - \frac{\pi}{2}\boldsymbol{e}_k\right)\right]. \end{aligned} \tag{B7}$$

这里，$\boldsymbol{e}_k$ 是沿着第 k 个分量的单位向量。参数 θ_k 中的$\pm\pi/2$位移可以提取出来，并视为添加到固定层中的额外 Clifford 门

$\hat{W}_k$。事实上，假设 $P_k = Z$ 并记 S 为相位门，我们有 $\hat{U}_k(\theta_k + \pi/2)\hat{W}_k = e^{-i\frac{\theta_k}{2}P_k}e^{-i\frac{\pi}{2}Z}\hat{W}_k = e^{-i\frac{\pi}{4}}\hat{U}_k(\theta_k)\hat{S}\hat{W}_k$。定义 $\hat{W}_{k,\pm} = e^{\mp i\frac{\pi}{4}}\hat{S}\hat{W}_k$，可得 $\hat{U}_k(\theta_k + \pi/2)W_k = U_k(\theta_k)W_{k,+}$。我们可以类似地继续定义 $\hat{W}_{k,-}$。接下来，我们可以写作 $\hat{V}_{i,\pm} = W_i$ 和 $V_{k,\pm} = W_{k,\pm}$，它们是包含所考虑位移的修改后的固定层。我们有

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}\left[\hat{U}_{\pm}(\boldsymbol{\theta})\hat{\rho}\hat{U}_{\pm}^{\dagger}(\boldsymbol{\theta})\right] &= \mathbb{E}_{\boldsymbol{\theta}}\left[\bigcirc_{i=1}^{M}(\mathcal{U}_i(\theta_i)\circ\mathcal{V}_{i,\pm})(\hat{\rho})\right] \\ &= \bigcirc_{i=1}^{M}(\mathbb{E}_{\theta_i}\left[\mathcal{U}_i(\theta_i)\right]\circ\mathcal{V}_{i,\pm})(\hat{\rho}),\end{aligned} \tag{B8}$$

其中 $\mathcal{V}_{i,\pm}(\hat{\rho}) = \hat{V}_{i,\pm}\hat{\rho}\hat{V}_{i,\pm}^{\dagger}$。因此，平均梯度是一阶量，也就是说，它只依赖于 1- 折叠通道。

2. 二阶量

现在我们来考察代价函数平方的平均值。其定义如下：

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}\left[C(\boldsymbol{\theta})^2\right] &= \mathbb{E}_{\boldsymbol{\theta}}\left[\mathrm{Tr}\left[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})\hat{O}\right]^2\right] \\ &= \mathbb{E}_{\boldsymbol{\theta}}\left[\mathrm{Tr}\left[\left(\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})\hat{O}\right)^{\otimes 2}\right]\right] \\ &= \mathbb{E}_{\boldsymbol{\theta}}\left[\mathrm{Tr}\left[\mathcal{U}^{(2)}(\boldsymbol{\theta})(\hat{\rho}^{\otimes 2})\hat{O}^{\otimes 2}\right]\right].\end{aligned} \tag{B9}$$

对于系统的 $2n$ 个量子比特的每个状态 $\hat{\rho}$（即原始系统的双倍版本，其中副本与原始量子线路之间没有通过量子门连接），我们定义

$$\mathcal{U}^{(2)}(\boldsymbol{\theta})(\hat{\rho}) = \hat{U}^{\otimes 2}(\boldsymbol{\theta})\hat{\rho}\hat{U}^{\dagger\otimes 2}(\boldsymbol{\theta}). \tag{B10}$$

同样，我们可以将电路层的双倍版本定义为

$$\begin{aligned}\mathcal{U}_i^{(2)}(\theta_i)(\hat{\rho}) &= \hat{U}_i^{\otimes 2}(\theta_i)\hat{\rho}\hat{U}_i^{\otimes 2}(\theta_i), \\ \mathcal{W}_i^{(2)}(\hat{\rho}) &= \hat{W}_i^{\otimes 2}\hat{\rho}\hat{W}_i^{\otimes 2},\end{aligned} \tag{B11}$$

给出

$$\mathcal{U}^{(2)}(\boldsymbol{\theta})(\hat{\rho}) = \bigcirc_{i=1}^{M}(\mathcal{U}_i^{(2)}(\theta_i)\circ\mathcal{W}_i^{(2)})(\hat{\rho}). \tag{B12}$$

因此，对于独立旋转，我们有

$$\begin{aligned}\Phi_{\boldsymbol{\theta}}^{(2)}(\hat{\rho}) &= \mathbb{E}_{\boldsymbol{\theta}}\left[\mathcal{U}^{(2)}(\boldsymbol{\theta})(\hat{\rho})\right] \\ &= \bigcirc_{i=1}^{M}(\mathbb{E}_{\theta_i}\left[\mathcal{U}_i^{(2)}(\theta_i)\right]\circ\mathcal{W}_i^{(2)})(\hat{\rho}).\end{aligned} \tag{B13}$$

至于一阶量，我们将那些可通过了解旋转层 $\mathbb{E}_{\theta_i}\left[\mathcal{U}_i^{(2)}(\theta_i)\right]$ 的平均二折叠通道而得到的量称为二阶量。

因此，代价函数平方的平均值是一个二阶量；与一阶情形类似，我们可以证明梯度平方也是一个二阶量。事实上，利用参数移位规则可知，要得到梯度平方的平均值，我们必须计算以下四项

$$\mathbb{E}_{\boldsymbol{\theta}}\left[C(\boldsymbol{\theta} + a_1\boldsymbol{e}_k)C(\boldsymbol{\theta} + a_2\boldsymbol{e}_k)\right], \tag{B14}$$

其中 $a_1, a_2 \in \{\frac{\pi}{2}, -\frac{\pi}{2}\}$。与之前一样，只需将式（B13）中的 $\mathcal{W}_i^{(2)}$ 替换为

$$\mathcal{V}_{i,a_1,a_2}^{(2)}(\hat{\rho}) = \left(\hat{V}_{i,a_1}\otimes\hat{V}_{i,a_2}\right)\hat{\rho}(\hat{V}_{i,a_1}^{\dagger}\otimes\hat{V}_{i,a_2}^{\dagger}). \tag{B15}$$

最后，梯度方差可计算为

$$\mathrm{Var}_{\boldsymbol{\theta}}\left[\partial_k C(\boldsymbol{\theta})\right] = \mathbb{E}_{\boldsymbol{\theta}}\left[\partial_k C(\boldsymbol{\theta})^2\right] - \mathbb{E}_{\boldsymbol{\theta}}\left[\partial_k C(\boldsymbol{\theta})\right]^2, \tag{B16}$$

它是一个一阶量与一个二阶量之和。

附录 C：采样效率的证明

在本附录中，我们证明，对于给定的拟设，要将任意一阶或二阶量估计到精度 ϵ，且以概率 $\delta \in [0,1]$ 达到该精度，只需采样数量为 $K \sim \log(2\delta)M/\epsilon^2$ 的 Clifford 近似电路。利用 Gottesman-Knill 定理，我们可以估计上述任意量，且其复杂度关于系统规模和所考虑拟设的变分参数数量均为多项式。

1. 关于映射的详细信息

这里我们详细介绍如何将随机初始化的参数化量子线路映射到 Clifford 近似线路。

注。我们采用适用于一阶量的记号。推广到二阶量和移位版本十分直接，只需将每个通道替换为其二重和/或移位版本，如附录 B 中所做的那样。

假设 θ_i 彼此独立，则对 $\boldsymbol{\theta}$ 上的 $\mathcal{U}(\boldsymbol{\theta})$取平均，相当于将每个旋转通道 $\mathcal{U}_i(\theta_i)$替换为 m 个 Clifford 酉通道的凸和 $\mathcal{U}_{ij}$，其对应权重为 p_{ij}。因此，$\mathbb{E}_{\boldsymbol{\theta}}\left[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})\right]$ 被替换为对 m^M个 Clifford 酉通道的离散平均（对于一重通道，$m = 2$；对于二重通道，$m = 4$）：

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})\right] = \bigcirc_{i=1}^{M}(\sum_{j=1}^{m}p_{ij}\mathcal{U}_{ij}\circ\mathcal{W}_i)(\hat{\rho}). \tag{C1}$$

由于我们希望从该和中采样，因此可以为每个 i 定义一个取值于以下集合的离散随机变量 X_i

$\{1,\ldots,m\}$，使得 $\mathbb{P}(X_i = j) = p_{ij}$。这表示从前述凸和中选取一个给定酉算子的过程。汇集所有 k 对应的变量后，我们得到一个随机向量 $\boldsymbol{X} = (X_1,\ldots,X_M) \in \{1,\ldots,m\}^M$，它通过以下方式唯一确定一个酉算子 $\mathcal{U}(\boldsymbol{X})$：

$$\mathcal{U}(j_1,\ldots,j_M) = \bigcirc_{i=1}^{M} \mathcal{U}_{ij_i} \circ \mathcal{W}_i. \tag{C2}$$

因此，我们有：

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}\left[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})\right] &= \mathbb{E}_{\boldsymbol{X}}\left[\mathcal{U}(\boldsymbol{X})(\hat{\rho})\right] \\ &= \bigcirc_{i=1}^{M}\Big(\sum_{j=1}^{m} p_{ij}\mathcal{U}_{ij}\circ\mathcal{W}_i\Big)(\hat{\rho}).\end{aligned} \tag{C3}$$

核心思路是通过对先前的 Clifford 电路进行 K 次采样并求经验平均，来近似 k 重折叠通道，即：

$$\hat{\Phi}(\hat{\rho}) := \frac{1}{K}\sum_{i=1}^{K}\mathcal{U}(\boldsymbol{X}_i)(\hat{\rho}). \tag{C4}$$

2. 采样效率

我们的结果依赖于经典论证：利用 McDiarmid 集中不等式，对一组随机变量所决定的有界函数进行采样 [73, 74], 。下面我们回顾该不等式。

定义 C.1（有界差分性质）。函数 $f: \mathcal{X}^M \to \mathbb{R}$ 满足有界差分性质，当且仅当存在 $\exists\{c_1,\ldots,c_M\}$ 使得 $\forall i \in\{1,\ldots,M\}, \forall(x_1,\ldots,x_M)$：

$$\sup_{x_i^{'}\in\mathcal{X}}\left|f(x_1,..,x_i,..,x_M) - f(x_1,..,x_i^{'},..,x_M)\right| < c_i\,.$$

定理 C.1（McDiarmid 不等式）。设 $f: \mathcal{X}^M \to\mathbb{R}$ 满足有界差分性质，其界为 $\{c_1,\ldots,c_M\}$；随机向量 $\boldsymbol{X} = (X_1,\ldots,X_M)$ 取值于 $\mathcal{X}^M$，则 $\forall\epsilon > 0$

$$\mathbb{P}\left(|f(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{X}}\left[f(\boldsymbol{X})\right]| \geq \epsilon\right) \leq 2\exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^{M} c_i^2}\right).$$

我们将证明，我们要估计的量满足有界差分性质，并应用 McDiarmid 不等式证明先前的采样是高效的。下面我们定义

$$f(\boldsymbol{x}) = \mathrm{Tr}\left[\mathcal{U}(\boldsymbol{x})(\hat{\rho})\hat{O}\right], \tag{C5}$$

ˆ其中 O 是正文中定义的代价函数可观测量，并且与上一节一样，$\mathcal{U}(\boldsymbol{x})$ 是与给定 Clifford 近似电路相关联的酉通道，该电路由一个离散向量完全指定 $\boldsymbol{x} = (x_1,\ldots,x_i,\ldots,x_M) \in \{1,\ldots,m\}^M$。根据 H¨older 不等式 [49, 75]，f 有如下上界：

$$|f(\boldsymbol{x})| \leq \|\hat{\rho}\|_1\|\hat{O}\|_\infty \tag{C6}$$

其中 $\|A\|_1$，$\|A\|_\infty$ 分别是 Schatten-1 范数和谱范数 [49]。注意，对于 ˆρ，$\|\hat{\rho}\|_1 = 1$是一个密度算子。定义另一个向量，只有第 i 个分量发生变化 $\boldsymbol{x'} = (x_1, \ldots\ldots, x_i^{'}, \ldots\ldots, x_M)$，利用三角不等式可得

$$\begin{aligned}\left|f(\boldsymbol{x}) - f(\boldsymbol{x}^{'})\right| &\leq |f(\boldsymbol{x})| + |f(\boldsymbol{x'})| \\ &\leq 2\|\hat{O}\|_\infty.\end{aligned} \tag{C7}$$

因此，f 满足有界差分性质，其界为 ˆ$c_i = c = 2\|O\|_\infty$。我们可以应用 McDiarmid 不等式，从而得到几乎是所需的结果。为了进一步推导，我们定义

$$\begin{aligned}f_K(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_K) &= \sum_{j=1}^{K} f(x_{j1},\ldots,x_{jM}) \\ &= \sum_{j=1}^{K}\mathrm{Tr}\left[\mathcal{U}(\boldsymbol{x}_j)(\hat{\rho})\hat{O}\right] \\ &= K\mathrm{Tr}\left[\hat{\Phi}(\hat{\rho})\hat{O}\right].\end{aligned} \tag{C8}$$

显然，f_K 满足有界差分性质，且界仍为 c [。为说明这一点，我们令除 x_{kl} 之外的所有 x_{ij}都相等，由此可知，差值 $f_K(\boldsymbol{x}_1, \ldots\ldots, \boldsymbol{x}_K) - f_K(\boldsymbol{x}_1^{'}, \ldots\ldots, \boldsymbol{x}_K^{'})$ 就是 $f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k^{'})$]。因此，McDiarmid 不等式适用于 f_K，它是一个由 KM 个参数构成的函数：

$$\begin{aligned}&\mathbb{P}(|f_K(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{X}}\left[f_K(\boldsymbol{X})\right]| \geq K\epsilon) \\ &= \mathbb{P}\left(\left|\frac{1}{K}f_K(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{X}}\left[\frac{1}{K}f_K(\boldsymbol{X})\right]\right| \geq \epsilon\right) \\ &= \mathbb{P}\left(\left|\mathrm{Tr}\left[\hat{\Phi}(\hat{\rho})\hat{O}\right] - \mathbb{E}_{\boldsymbol{\theta}}\left[\mathrm{Tr}\left[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})\hat{O}\right]\right]\right| \geq \epsilon\right) \\ &\leq 2\exp\left(-\frac{2K^2\epsilon^2}{KMc^2}\right) \\ &= 2\exp\left(-\frac{K\epsilon^2}{2M\|\hat{O}\|_\infty^2}\right).\end{aligned} \tag{C9}$$

因此，选择精度 $\epsilon > 0$ 以及满足该精度的概率 $1-\delta \in [0,1]$，我们得到

$$\begin{aligned}&\mathbb{P}\left(\left|\mathrm{Tr}\left[\hat{\Phi}(\hat{\rho})\hat{O}\right] - \mathbb{E}_{\boldsymbol{\theta}}\left[\mathrm{Tr}\left[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})\hat{O}\right]\right]\right| \leq \epsilon\right) \\ &\geq 1-\delta\end{aligned} \tag{C10}$$

只要抽样的 Clifford 电路数量 K 为

$$K \geq \frac{2}{\epsilon^2}\log\left(\frac{2}{\delta}\right)M\|\hat{O}\|_\infty^2 = O\left(M\right). \tag{C11}$$

请注意，在式 (C9) 中，将可观测量 $\hat{O}$ 替换为其归一化形式 $\hat{O}/\|O\|$，并赋予相应精度 $\tilde{\epsilon}$，对于 K会得到相同的尺度关系，因为在这种情况下 $\hat{\tilde{\epsilon}} = \epsilon/\|O\|$。因此，我们始终可以使用归一化可观测量。不过，如果关注其随系统规模 n 的变化关系，就必须考虑一系列可观测量 $\hat{O}_n$，它们的范数可能会随 n呈现特定的尺度关系，因此式 (C11) 中包含 O 的范数，能够追踪这一影响。在许多值得关注的情况下，所考虑的可观测量随系统规模呈多项式变化，K也如此。最后，可以利用 Gottesman-Knill 定理：对于作用于 N_O个量子比特上的 Clifford 酉算子 $\hat{U}$ 和可观测量 O，期望值 Tr $\left[|0\rangle\langle 0|^{\otimes n}\hat{U}^\dagger\hat{O}\hat{U}\right]$可通过经典计算，以关于 N_O 和量子比特数量 n [24]的多项式复杂度计算。我们的方法继承了这一尺度关系，因此对于每个 k，都能用经典计算机以关于 $O(n^p N_O^q M)$ 的复杂度估计梯度方差 $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_k C(\boldsymbol{\theta})]$，其中 M 为变分量子电路中的参数数量。

附录 D：一般情形下的采样效率

本节将前述方案推广到更一般的分布。首先，我们在附录 D1 中讨论放宽凸性条件后采样复杂度的变化，即不再要求二重通道的分解 [式（A25）]为凸和，而只假设 θ的分布是偶分布。随后，我们在附录D2 中研究旋转角服从任意分布的情形，该分布不一定关于原点对称。最后，我们证明，先前的方案仍然适用，但要付出的代价是，所需采样的 Clifford 近似电路数量会随变分参数 M 的数量呈指数增长。与蛮力模拟相比，该方法可以用系统规模上的指数级复杂度换取变分参数数量上的指数级复杂度。

1. 非凸情形下的采样效率

在此，我们考虑旋转角分布 θ 为偶分布，但不满足式（A28）中的凸性条件的情况。在这种情况下，我们对一重通道的分解仍然是凸的，而二重通道则变成非凸和，因此，克利福德通道的系数不再能解释为概率。我们首先说明，如何仍然通过概率采样来估计这类非凸和 [72]。记

$$\mathbb{E}_{\boldsymbol{\theta}}[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})] = \bigcirc_{k=1}^{M}\left(\sum_{j=1}^{m} q_{kj}\mathcal{U}_{kj}\circ\mathcal{W}_k\right)(\hat{\rho})\,, \tag{D1}$$

我们在此假设

$$q_{kj}\in\mathbb{R}\,,\quad \sum_{j=1}^{M} q_{kj} = 1\,,\quad \forall k\,. \tag{D2}$$

定义

$$\gamma_k := \Sigma_{j=1}^{M}|q_{kj}|\,,\quad \tilde{p}_{kj} := |q_{kj}|/\gamma_k\,, \tag{D3}$$

式（D1）可以用凸和重写为：

$$\mathbb{E}_{\boldsymbol{\theta}}[\mathcal{U}(\boldsymbol{\theta})] = \bigcirc_{k=1}^{M}\sum_{j=1}^{m}\tilde{p}_{kj}\left[\gamma_k\,\mathrm{sgn}(q_{kj})\right]\mathcal{U}_{kj}\circ\mathcal{W}_k\,. \tag{D4}$$

与附录 C 1类似，我们现在定义随机向量 $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_M) \in \{1,\ldots,m\}^M$，其概率为$\mathbb{P}(X_k = j) = \tilde{p}_{kj}$，并通过以下方式定义重新缩放的随机酉算子通道 $\tilde{\mathcal{U}}(\boldsymbol{X})$：

$$\tilde{\mathcal{U}}(j_1,\ldots,j_M) = \bigcirc_{k=1}^{M}\left[\gamma_k\,\mathrm{sgn}(q_{kj_k})\right]\mathcal{U}_{kj_k}\circ\mathcal{W}_k\,. \tag{D5}$$

因此，我们得到了类似于式（C3）的期望值形式：

$$\mathbb{E}_{\boldsymbol{\theta}}[\mathcal{U}(\boldsymbol{\theta})(\hat{\rho})] = \mathbb{E}_{\tilde{\boldsymbol{X}}}\left[\tilde{\mathcal{U}}(\tilde{\boldsymbol{X}})(\hat{\rho})\right]\,. \tag{D6}$$

这使我们能够通过考虑函数，应用与附录 C2 中相同的论证

$$\tilde{f}(\boldsymbol{x}) = \mathrm{Tr}\left[\tilde{\mathcal{U}}(\boldsymbol{x})(\hat{\rho})\hat{O}\right] \tag{D7}$$

而不是式（ f（$\boldsymbol{x}$）中定义的 fC5）。函数界（C6）也应相应地重新缩放：

$$\left|\tilde{f}(\boldsymbol{x})\right| \le \gamma\|\hat{O}\|_\infty\,, \tag{D8}$$

其中，缩放因子定义为

$$\gamma := \prod_{k=1}^{M}\gamma_k\,. \tag{D9}$$

因此，先前在公式（C11）中推导出的采样量子线路数也应乘以相同的缩放因子：

$$K \ge \frac{2}{\epsilon^2}\log\left(\frac{2}{\delta}\right)\gamma M\|\hat{O}\|_\infty^2\,. \tag{D10}$$

请注意，因子 $\gamma_k \ge 1$ 可以视为第 k 个通道分解中“非凸性”的度量。在凸和的情况下，即 $q_{kj} > 0,\ \forall k,\ j$，缩放因子 simply 为 $\gamma = 1^M = 1$，由此我们可以得到之前的结果。

现在我们证明 γ_k 具有上界。根据附录 A 中的讨论，我们只需考虑单量子比特 Z 旋转的 2 重通道，其中分解可能是非凸的。不失一般性地，将公式（A25）重写为

$$\begin{aligned}\Phi_Z^{(2)}(\hat{\rho}) = {} & q_{k1}\hat{\rho} + q_{k2}(Z\otimes Z)\,\hat{\rho}\,(Z\otimes Z)\\ & + q_{k3}CZ\hat{\rho}CZ + q_{k4}CZ_X\hat{\rho}CZ_X\end{aligned} \tag{D11}$$

对于某个 k，其中

$$
\begin{aligned}
q_{k1} &= \mathbb{E}_\theta\left[\frac{1}{4}\left(1+\cos 2\theta+2\cos\theta\right)\right], \\
q_{k2} &= \mathbb{E}_\theta\left[\frac{1}{4}\left(1+\cos 2\theta-2\cos\theta\right)\right], \\
q_{k3} &= q_{k4} = \mathbb{E}_\theta\left[\frac{1}{4}\left(1-\cos 2\theta\right)\right].
\end{aligned}
\tag{D12}
$$

定义非负函数

$$
\begin{aligned}
\varphi(\theta) := &\left|\frac{1}{4}\left(1+\cos 2\theta+2\cos\theta\right)\right| \\
&+\left|\frac{1}{4}\left(1+\cos 2\theta-2\cos\theta\right)\right| \\
&+2\times\left|\frac{1}{4}\left(1-\cos 2\theta\right)\right|,
\end{aligned}
\tag{D13}
$$

于是得到

$$
\begin{aligned}
\gamma_k &= |q_{k1}|+|q_{k2}|+|q_{k3}|+|q_{k4}| \\
&\le \mathbb{E}_\theta\left[\varphi(\theta)\right] \\
&\le \mathbb{E}_\theta\left[\sup_{\theta'}\varphi(\theta')\right] \\
&= \sup_{\theta'}\varphi(\theta') \\
&= \frac{5}{4}.
\end{aligned}
\tag{D14}
$$

此处，函数 $\varphi(\theta)$ 在 $\theta = \pm\frac{\pi}{3}$，$\pm\frac{2\pi}{3}$ 处达到最大值。因此，当 γ_k 达到其上界 $\frac{5}{4}$ 时，θ 的分布是以 $\theta = \pm\frac{\pi}{3}$ 和/或 $\theta = \pm\frac{2\pi}{3}$ 为峰值的狄拉克δ分布之和。在这种情况下，采样的 Clifford 电路数量（D10）呈现最坏情况的尺度：

$$
\begin{aligned}
K &\ge \frac{2}{\epsilon^2}\log\left(\frac{2}{\delta}\right)\gamma M\|\hat{O}\|_\infty^2 = O(\gamma M), \\
\gamma &\le \left(\frac{5}{4}\right)^M.
\end{aligned}
\tag{D15}
$$

结合上述结果与 Gottesman-Knill 定理，对于一个非平凡地作用于 N_O 个量子比特上的代价函数可观测量 O，在一般情形下，我们的方案意味着在经典计算机上估计每个 k 的梯度方差 $\mathrm{Var}_{\boldsymbol{\theta}}\left[\partial_k C(\boldsymbol{\theta})\right]$ 的复杂度至多为 $O\left(n^p N_O^q (\frac{5}{4})^M M\right)$，其中 n 为量子比特数，M 为变分量子电路中的参数数量，而 p，q 是由 Gottesman-Knill 定理导出的常数。

2. 一般情形下的采样效率

在本节中，我们通过考虑旋转角 $\boldsymbol{\theta}$的任意概率分布，将方案扩展到最一般的情形，并推导相应的采样复杂度。与之前一样，只需考虑单量子比特 $\boldsymbol{\theta}$ 旋转门的一重和二重信道。

下文中，记 $\mathbb{E}_\theta\left[\mathrm{e}^{\mathrm{i}\theta}\right] := r_1 + \mathrm{i}s_1$ and $\mathbb{E}_\theta\left[\mathrm{e}^{2\mathrm{i}\theta}\right] := r_2 + \mathrm{i}s_2$。注意，$r_1 = \mathbb{E}_\theta[\cos\theta]$ 和 $r_2 = \mathbb{E}_\theta[\cos 2\theta]$ 的定义与之前对称情形中的定义相同；而由于我们不再假设 θ 的分布为偶分布，$s_1 = \mathbb{E}_\theta[\sin\theta]$ 和 $s_2 = \mathbb{E}_\theta[\sin 2\theta]$ 通常不为零。

a. 1-fold channel

单量子比特的“一重通道”表达式Z旋转可由公式（A2）给出，下面将推导该表达式，且不假设 θ中的分布均匀。我们得到：

$$
\begin{aligned}
&\mathbb{E}_\theta\left[\hat{R}_Z(\theta)\hat{\rho}\hat{R}_Z^\dagger(\theta)\right] \\
&= \hat{\Pi}_0\hat{\rho}\hat{\Pi}_0 + \hat{\Pi}_1\hat{\rho}\hat{\Pi}_1 \\
&\quad + \mathbb{E}_\theta\left[e^{\mathrm{i}\theta}\right]\hat{\Pi}_1\hat{\rho}\hat{\Pi}_0 + \mathbb{E}_\theta\left[e^{-\mathrm{i}\theta}\right]\hat{\Pi}_0\hat{\rho}\hat{\Pi}_1 \\
&= \hat{\Pi}_0\hat{\rho}\hat{\Pi}_0 + \hat{\Pi}_1\hat{\rho}\hat{\Pi}_1 \\
&\quad + (r_1+\mathrm{i}s_1)\hat{\Pi}_1\hat{\rho}\hat{\Pi}_0 + (r_1-\mathrm{i}s_1)\hat{\Pi}_0\hat{\rho}\hat{\Pi}_1 \\
&= \frac{1+r_1}{2}\mathcal{E}[\mathbb{1}](\hat{\rho}) + \frac{1-r_1}{2}\mathcal{E}[\hat{Z}](\hat{\rho}) \\
&\quad + \frac{s_1}{2}\mathcal{E}[\hat{S}^\dagger](\hat{\rho}) - \frac{s_1}{2}\mathcal{E}[\hat{S}](\hat{\rho}),
\end{aligned}
\tag{D16}
$$

^^^其中 $S = \Pi_0 + \mathrm{i}\Pi_1$ 是相位门，并且可以结合公式（A3）和该定义来验证上述公式。

这里，参数 s_1 可以理解为 θ概率分布不对称性的一种度量。在对称情形下，我们有 $s_1 = 0$，此时该总和简化为式（A4）给出的凸组合。按照附录D1中的相同步骤，这种克利福德通道的（可能非凸的）线性组合可以通过采样来估计；所需样本数应根据该总和的非凸性，乘以一个缩放因子 $\gamma = \Pi_{k=1}^M \gamma_k$ [（定义见式（D1）–（D3）和式（D9））]。现在，我们推导 $\gamma_k^{(1)}$ 的一个上界；该量是与一般可按式（k第 Z）分解的单个（第 Z）一重旋转通道相关联的缩放因子。我们沿用式（D11）–（D14）中的相同论证：

$$
\begin{aligned}
\gamma_k^{(1)} &= \left|\frac{1+r_1}{2}\right| + \left|\frac{1-r_1}{2}\right| + \left|\frac{s_1}{2}\right| + \left|-\frac{s_1}{2}\right| \\
&= \left|\mathbb{E}_\theta\left[\frac{1+\cos\theta}{2}\right]\right| + \left|\mathbb{E}_\theta\left[\frac{1-\cos\theta}{2}\right]\right| \\
&\quad + \left|\mathbb{E}_\theta\left[\sin\theta\right]\right| \\
&\le \mathbb{E}_\theta\left[\left|\frac{1+\cos\theta}{2}\right| + \left|\frac{1-\cos\theta}{2}\right| + \left|\sin\theta\right|\right] \\
&\le \sup_\theta\left\{\left|\frac{1+\cos\theta}{2}\right| + \left|\frac{1-\cos\theta}{2}\right| + \left|\sin\theta\right|\right\} \\
&= 2.
\end{aligned}
\tag{D17}
$$

这意味着，估计一般的一重通道 $K^{(1)}$ 所需的样本数 [（见式（D15））] 按如下方式缩放

$$\begin{aligned} K^{(1)} &\sim O(\gamma^{(1)} M)\,, \\ \gamma^{(1)} &= \prod_{k=1}^{M} \gamma_k^{(1)} \leq 2^M\,. \end{aligned} \tag{D18}$$

请注意，上述推导出的上界取决于分解中所选的具体 Clifford 量子信道。由于 Clifford 群并非线性无关集，因此应该可以找到另一种分解，从而得到不同的上界，并进一步优化复杂度。

b. 2-fold channel

单量子比特的二重通道 Z旋转可由方程（A14）给出：

$$\begin{aligned} \Phi_Z^{(2)}(\hat{\rho}) = \mathbb{E}_\theta\left[\Xi\hat{\rho}\Xi^\dagger\right] + \mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Gamma_\theta^\dagger\right] \\ + \mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Xi^\dagger\right] + \mathbb{E}_\theta\left[\Xi\hat{\rho}\Gamma_\theta^\dagger\right]\,. \end{aligned} \tag{D19}$$

对于 θ 的一般概率分布，我们有

$$\begin{aligned} \mathbb{E}_\theta\left[\Xi\hat{\rho}\Xi^\dagger\right] &= \hat{\Pi}_{01}\hat{\rho}\hat{\Pi}_{01} + \hat{\Pi}_{10}\hat{\rho}\hat{\Pi}_{10} \\ &\quad + \hat{\Pi}_{01}\hat{\rho}\hat{\Pi}_{10} + \hat{\Pi}_{10}\hat{\rho}\hat{\Pi}_{01}\,, \\ \mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Gamma_\theta^\dagger\right] &= \hat{\Pi}_{00}\hat{\rho}\hat{\Pi}_{00} + \hat{\Pi}_{11}\hat{\rho}\hat{\Pi}_{11} \\ &\quad + r_2\left(\hat{\Pi}_{00}\hat{\rho}\hat{\Pi}_{11} + \hat{\Pi}_{11}\hat{\rho}\hat{\Pi}_{00}\right) \\ &\quad + \mathrm{i}s_2\left(\hat{\Pi}_{11}\hat{\rho}\hat{\Pi}_{00} - \hat{\Pi}_{00}\hat{\rho}\hat{\Pi}_{11}\right)\,, \\ \mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Xi^\dagger\right] &= r_1\left(\hat{\Pi}_{00} + \hat{\Pi}_{11}\right)\hat{\rho}\left(\hat{\Pi}_{01} + \hat{\Pi}_{10}\right) \\ &\quad + \mathrm{i}s_1\left(\hat{\Pi}_{11} - \hat{\Pi}_{00}\right)\hat{\rho}\left(\hat{\Pi}_{01} + \hat{\Pi}_{10}\right)\,, \\ \mathbb{E}_\theta\left[\Xi\hat{\rho}\Gamma_\theta^\dagger\right] &= \mathbb{E}_\theta\left[\Gamma_\theta\hat{\rho}\Xi^\dagger\right]^\dagger\,. \end{aligned} \tag{D20}$$

可以验证，上述各项的 Choi 表示都是对角的，因此它们的和可以像之前一样通过 M 矩阵表示：

$$\begin{aligned} &M(\Phi_Z^{(2)}) \\ &= \begin{pmatrix} 1 & r_1 + \mathrm{i}s_1 & r_1 + \mathrm{i}s_1 & r_2 + \mathrm{i}s_2 \\ r_1 - \mathrm{i}s_1 & 1 & 1 & r_1 + \mathrm{i}s_1 \\ r_1 - \mathrm{i}s_1 & 1 & 1 & r_1 + \mathrm{i}s_1 \\ r_2 - \mathrm{i}s_2 & r_1 - \mathrm{i}s_1 & r_1 - \mathrm{i}s_1 & 1 \end{pmatrix}\,. \end{aligned} \tag{D21}$$

这又可以分解为以下通道的加权和：$\mathcal{E}[\mathbb{1}]$、$\mathcal{E}[\hat{Z}\otimes\hat{Z}]$、$\mathcal{E}[\hat{S}\otimes\hat{S}]$ 和 $\mathcal{E}[\hat{S}^\dagger\otimes\hat{S}^\dagger]$。

见式（A18）以及以下克利福德通道：

$$\begin{aligned} M\left(\mathcal{E}\left[\mathbb{1}\otimes\hat{S}\right]\right) &= \begin{pmatrix} 1 & \mathrm{i} & 1 & \mathrm{i} \\ -\mathrm{i} & 1 & -\mathrm{i} & 1 \\ 1 & \mathrm{i} & 1 & \mathrm{i} \\ -\mathrm{i} & 1 & -\mathrm{i} & 1 \end{pmatrix}\,, \\ M\left(\mathcal{E}\left[\hat{S}\otimes\mathbb{1}\right]\right) &= \begin{pmatrix} 1 & 1 & \mathrm{i} & \mathrm{i} \\ 1 & 1 & \mathrm{i} & \mathrm{i} \\ -\mathrm{i} & -\mathrm{i} & 1 & 1 \\ -\mathrm{i} & -\mathrm{i} & 1 & 1 \end{pmatrix}\,, \\ M\left(\mathcal{E}\left[\mathbb{1}\otimes\hat{S}^\dagger\right]\right) &= \begin{pmatrix} 1 & -\mathrm{i} & 1 & -\mathrm{i} \\ \mathrm{i} & 1 & \mathrm{i} & 1 \\ 1 & -\mathrm{i} & 1 & -\mathrm{i} \\ \mathrm{i} & 1 & \mathrm{i} & 1 \end{pmatrix}\,, \\ M\left(\mathcal{E}\left[\hat{S}^\dagger\otimes\mathbb{1}\right]\right) &= \begin{pmatrix} 1 & 1 & -\mathrm{i} & -\mathrm{i} \\ 1 & 1 & -\mathrm{i} & -\mathrm{i} \\ \mathrm{i} & \mathrm{i} & 1 & 1 \\ \mathrm{i} & \mathrm{i} & 1 & 1 \end{pmatrix}\,, \\ M\left(\mathcal{E}\left[\hat{Z}\otimes\hat{S}\right]\right) &= \begin{pmatrix} 1 & \mathrm{i} & -1 & -\mathrm{i} \\ -\mathrm{i} & 1 & \mathrm{i} & -1 \\ -1 & -\mathrm{i} & 1 & \mathrm{i} \\ \mathrm{i} & -1 & -\mathrm{i} & 1 \end{pmatrix}\,, \\ M\left(\mathcal{E}\left[\hat{S}\otimes\hat{Z}\right]\right) &= \begin{pmatrix} 1 & -1 & \mathrm{i} & -\mathrm{i} \\ -1 & 1 & -\mathrm{i} & \mathrm{i} \\ -\mathrm{i} & \mathrm{i} & 1 & -1 \\ \mathrm{i} & -\mathrm{i} & -1 & 1 \end{pmatrix}\,, \\ M\left(\mathcal{E}\left[\hat{Z}\otimes\hat{S}^\dagger\right]\right) &= \begin{pmatrix} 1 & -\mathrm{i} & -1 & \mathrm{i} \\ \mathrm{i} & 1 & -\mathrm{i} & -1 \\ -1 & \mathrm{i} & 1 & -\mathrm{i} \\ -\mathrm{i} & -1 & \mathrm{i} & 1 \end{pmatrix}\,, \\ M\left(\mathcal{E}\left[\hat{S}^\dagger\otimes\hat{Z}\right]\right) &= \begin{pmatrix} 1 & -1 & -\mathrm{i} & \mathrm{i} \\ -1 & 1 & \mathrm{i} & -\mathrm{i} \\ \mathrm{i} & -\mathrm{i} & 1 & -1 \\ -\mathrm{i} & \mathrm{i} & -1 & 1 \end{pmatrix}\,. \end{aligned} \tag{D22}$$

请注意，上述量子信道在 Choi 表示中均为对角形式，因此 M 矩阵包含了它们所有非零元素。按照附录A中的相同推理，我们求解一个线性方程组，得到以下分解：

$$\begin{aligned} \Phi_Z^{(2)}(\hat{\rho}) &= \frac{s_2}{8}\left(\mathcal{E}\left[\hat{S}\otimes\mathbb{1}\right](\hat{\rho}) + \mathcal{E}\left[\mathbb{1}\otimes\hat{S}\right](\hat{\rho})\right) \\ &\quad + \frac{s_2}{8}\left(\mathcal{E}\left[\hat{Z}\otimes\hat{S}^\dagger\right](\hat{\rho}) + \mathcal{E}\left[\hat{S}^\dagger\otimes\hat{Z}\right](\hat{\rho})\right) \\ &\quad - \frac{s_2}{8}\left(\mathcal{E}\left[\hat{S}^\dagger\otimes\mathbb{1}\right](\hat{\rho}) + \mathcal{E}\left[\mathbb{1}\otimes\hat{S}^\dagger\right](\hat{\rho})\right) \\ &\quad - \frac{s_2}{8}\left(\mathcal{E}\left[\hat{Z}\otimes\hat{S}\right](\hat{\rho}) + \mathcal{E}\left[\hat{S}\otimes\hat{Z}\right](\hat{\rho})\right) \\ &\quad + \frac{1 + r_2 + 2r_1}{4}\mathcal{E}\left[\mathbb{1}\right](\hat{\rho}) \\ &\quad + \frac{1 + r_2 - 2r_1}{4}\mathcal{E}\left[\hat{Z}\otimes\hat{Z}\right](\hat{\rho}) \\ &\quad + \frac{1 - r_2 + 2s_1}{4}\mathcal{E}\left[\hat{S}\otimes\hat{S}\right](\hat{\rho}) \\ &\quad + \frac{1 - r_2 - 2s_1}{4}\mathcal{E}\left[\hat{S}^\dagger\otimes\hat{S}^\dagger\right](\hat{\rho})\,. \end{aligned} \tag{D23}$$

$\widehat{CNOT}$ 注。记 $CNOT = \Pi_0 \otimes \mathbb{1} + \Pi_1 \otimes X$ 以及 $CNOT_X := (X \otimes X)\, CNOT\, (X \otimes X)$ 为其由 $X \otimes X$ 门诱导的共轭变换，则有

$$
\begin{aligned}
&M\left(\mathcal{E}\left[\widehat{CNOT}(\hat{S}\otimes\hat{S})\widehat{CNOT}\right]\right) \\
&= \begin{pmatrix} 1 & \mathrm{i} & -1 & \mathrm{i} \\ -\mathrm{i} & 1 & \mathrm{i} & 1 \\ -1 & -\mathrm{i} & 1 & -\mathrm{i} \\ -\mathrm{i} & 1 & \mathrm{i} & 1 \end{pmatrix}, \\
&M\left(\mathcal{E}\left[\widehat{CNOT}_X(\hat{S}\otimes\hat{S})\widehat{CNOT}_X\right]\right) \\
&= \begin{pmatrix} 1 & -\mathrm{i} & 1 & \mathrm{i} \\ \mathrm{i} & 1 & \mathrm{i} & -1 \\ 1 & -\mathrm{i} & 1 & \mathrm{i} \\ -\mathrm{i} & -1 & -\mathrm{i} & 1 \end{pmatrix}.
\end{aligned}
\tag{D24}
$$

再次，通过求解一个线性方程组，可以得到二重信道的另一种分解，其中涉及上述量子信道，即：

$$
\begin{aligned}
\Phi_Z^{(2)}(\hat{\rho}) = &\frac{s_2}{4}\mathcal{E}\left[\widehat{CNOT}(\hat{S}\otimes\hat{S})\widehat{CNOT}\right](\hat{\rho}) \\
&- \frac{s_2}{4}\mathcal{E}\left[\hat{Z}\otimes\hat{S}\right](\hat{\rho}) \\
&+ \frac{s_2}{4}\mathcal{E}\left[\widehat{CNOT}_X(\hat{S}\otimes\hat{S})\widehat{CNOT}_X\right](\hat{\rho}) \\
&- \frac{s_2}{4}\mathcal{E}\left[\mathbb{1}\otimes\hat{S}^\dagger\right](\hat{\rho}) \\
&+ \frac{1+r_2+2r_1}{4}\mathcal{E}\left[\mathbb{1}\right](\hat{\rho}) \\
&+ \frac{1+r_2-2r_1}{4}\mathcal{E}\left[\hat{Z}\otimes\hat{Z}\right](\hat{\rho}) \\
&+ \frac{1-r_2+2s_1}{4}\mathcal{E}\left[\hat{S}\otimes\hat{S}\right](\hat{\rho}) \\
&+ \frac{1-r_2-2s_1}{4}\mathcal{E}\left[\hat{S}^\dagger\otimes\hat{S}^\dagger\right](\hat{\rho}).
\end{aligned}
\tag{D25}
$$

与我们对一重通道的处理类似，让我们推导 $\gamma_k^{(2)}$ 的上界；$\gamma_k^{(2)}$ 是估计所需样本数的缩放因子，用于估计一般的二重 k 阶 Z 旋转通道：

$$
\begin{aligned}
\gamma_k^{(2)} &= 4\left|\frac{s_2}{8}\right| + 4\left|-\frac{s_2}{8}\right| \\
&\quad + \left|\frac{1+r_2+2r_1}{4}\right| + \left|\frac{1+r_2-2r_1}{4}\right| \\
&\quad + \left|\frac{1-r_2+2s_1}{4}\right| + \left|\frac{1-r_2-2s_1}{4}\right| \\
&= \left|\mathbb{E}_\theta\left[\sin 2\theta\right]\right| \\
&\quad + \left|\mathbb{E}_\theta\left[\frac{1+\cos 2\theta+2\cos\theta}{4}\right]\right| \\
&\quad + \left|\mathbb{E}_\theta\left[\frac{1+\cos 2\theta-2\cos\theta}{4}\right]\right| \\
&\quad + \left|\mathbb{E}_\theta\left[\frac{1-\cos 2\theta+2\sin\theta}{4}\right]\right| \\
&\quad + \left|\mathbb{E}_\theta\left[\frac{1-\cos 2\theta-2\sin\theta}{4}\right]\right|. \\
&\le \sup_\theta \left\{\left|\sin 2\theta\right|\right. \\
&\quad + \left|\frac{1+\cos 2\theta+2\cos\theta}{4}\right| \\
&\quad + \left|\frac{1+\cos 2\theta-2\cos\theta}{4}\right| \\
&\quad + \left|\frac{1-\cos 2\theta+2\sin\theta}{4}\right| \\
&\quad \left. + \left|\frac{1-\cos 2\theta-2\sin\theta}{4}\right|\right\} \\
&= 1+\sqrt{2}\,.
\end{aligned}
\tag{D26}
$$

这意味着，估计一般的二重信道所需的样本数 $K^{(2)}$ 按以下方式增长：

$$
\begin{aligned}
K^{(2)} &\sim O(\gamma^{(2)} M)\,, \\
\gamma^{(2)} &= \prod_{k=1}^{M} \gamma_k^{(2)} \le (1+\sqrt{2})^M\,,
\end{aligned}
\tag{D27}
$$

这高于估计一重通道的复杂度 [式（D18）] 因为 $1+\sqrt{2} > 2$。

再次，将上述结果与 Gottesman-Knill 定理结合，对于一个非平凡地作用于 N_O 个量子比特上的代价函数可观测量 $\hat{O}$，在最一般的情况下，我们的方案意味着，在经典计算机上，对每个 k 估计梯度方差 $\mathrm{Var}_{\boldsymbol{\theta}}\left[\partial_k C(\boldsymbol{\theta})\right]$ 的复杂度不超过 $O\left(n^p N_O^q (1+\sqrt{2})^M M\right)$，其中 n 是量子比特数，M 是变分试探态中的参数数目，而 p, q 是由 Gottesman-Knill 定理导出的一些常数。

附录 E：N 重通道：随机 Z 旋转

本节将 N 重次通道分解为 Clifford 酉通道的实数和。这样

我们便能将该方案扩展到估计 $N-th$ 阶量，其复杂度在分解为凸分解时，随变分参数数量 M 和系统规模 n 呈多项式增长，否则则随 M 呈指数增长。我们给出了随机角度 θ 的分布满足何种充分条件时，分解为凸分解。

回想一下，对于任意酉算符 $\hat{U}$，我们定义了 $\mathcal{E}\left[\hat{U}\right](\hat{\rho})$: $=\hat{U}\hat{\rho}\hat{U}^{\dagger}$。在式（D16）中，对于随机角度的一般分布，我们将 Z 旋转的一重通道分解为 Clifford 酉通道的线性组合，即

$$\begin{aligned}&\mathbb{E}_{\theta}\left[\hat{R}_{Z}(\theta)\hat{\rho}\hat{R}_{Z}^{\dagger}(\theta)\right]\\=&\frac{1+r_1}{2}\mathcal{E}[\mathbb{1}](\hat{\rho})+\frac{1-r_1}{2}\mathcal{E}[\hat{Z}](\hat{\rho})\\&+\frac{s_1}{2}\mathcal{E}[\hat{S}^{\dagger}](\hat{\rho})-\frac{s_1}{2}\mathcal{E}[\hat{S}](\hat{\rho}).\end{aligned}\tag{E1}$$

更一般地，有

$$\begin{aligned}\hat{R}_{Z}(\theta)\hat{\rho}\hat{R}_{Z}^{\dagger}(\theta)=&\frac{1+\cos\theta}{2}\mathcal{E}[\mathbb{1}](\hat{\rho})\\&+\frac{1-\cos\theta}{2}\mathcal{E}[\hat{Z}](\hat{\rho})\\&+\frac{\sin\theta}{2}\mathcal{E}[\hat{S}^{\dagger}](\hat{\rho})\\&-\frac{\sin\theta}{2}\mathcal{E}[\hat{S}](\hat{\rho}),\end{aligned}\tag{E2}$$

对于任意 $\theta\in\mathbb{R}$。这可以看作式（E1）在以 θ为中心的狄拉克概率测度下的推论。我们可以直接将该式推广，得到 N重通道表示为 Clifford 酉通道实数和的形式，即

$$\hat{R}_{Z}^{\otimes N}(\theta)\hat{\rho}\hat{R}_{Z}^{\otimes N\dagger}(\theta)=\sum_{I=(i_1,\ldots,i_n)}\lambda_{I}(\theta)\mathcal{E}\left[\otimes_{j=1}^{N}\hat{U}_{i_j}\right](\hat{\rho})\tag{E3}$$

其中求和遍历所有多重指标 $I=\hat{}\hat{}\hat{}\hat{}\hat{}(i_1、\ldots\ldots、i_N)\in\{0,1,2,3\}$，以及 $U_0=\mathbb{1}$、$U_1=Z$、$U_2=S$ 和 $\hat{}U_3=\hat{}S^{\dagger}$。对于表示由 m_i 个 U_i 量子门的乘积的多重指标 $I\hat{}$，系数 $\lambda_I(\theta)_I$ 为

$$\begin{aligned}\lambda_{I}(\theta)=&\frac{1}{2^{N}}\left(1+\cos\theta\right)^{m_0}\left(1-\cos\theta\right)^{m_1}\\&\sin^{m_2}\left(-\theta\right)\sin^{m_3}\left(\theta\right),\end{aligned}\tag{E4}$$

其中 $m_0+m_1+m_2+m_3=N$。因此，N 重通道可表示为 4^N 个克利福德酉通道的实线性组合，这些通道由量子门 1、$\hat{}Z$、$\hat{}S$ 和 $\hat{}S^{\dagger}$ 的乘积构成。这为 N 重通道是克利福德酉通道的凸和提供了一个充分条件，即所有系数 $\mathbb{E}_{\theta}\left[\lambda_I(\theta)\right]$ 的期望值均为正。

尽管这一条件是充分条件，但并非必要。特别是在二重通道的情形下，

与多重指标 (2，3) 和 (3，2) 相关的系数的期望值为 $\mathbb{E}_{\theta}\left[-\sin^2\theta\right]$，其始终为负。然而，我们证明了均匀分布存在凸分解。这是因为式 (E3) 的分解并不唯一。事实上，通道族

$$\mathcal{P}:=\{\mathcal{E}[\hat{U}\otimes\hat{V}]:\hat{U},\hat{V}\in\{\mathbb{1},\hat{Z},\hat{S},\hat{S}^{\dagger}\}\}\tag{E5}$$

不线性无关。考虑两个在计算基下为对角形式的单量子比特酉算符 $\hat{}\hat{}$ U 和 V。由于我们可以自由选择这些酉算符的全局相位$\hat{}$，因此总能将它们写成 $U=e^{\mathrm{i}\theta_U/2}\Pi_0+e^{-\mathrm{i}\theta_U/2}\Pi_1$ $\quad V=e^{\mathrm{i}\theta_V/2}\Pi_0+e^{-\mathrm{i}\theta_V/2}\Pi_1$ 和 $\hat{}\hat{}\hat{}$。我们 $\hat{}\hat{}$ 在附录 A 中看到，乘积酉算符 $U\otimes V$可以用相应 Choi矩阵的对角元素表示，写成一个 4×4 矩阵 M：

$$M(\mathcal{E}[\hat{U}\otimes\hat{V}])=\begin{pmatrix}1&e^{-\mathrm{i}\theta_V}&e^{-\mathrm{i}\theta_U}&e^{-\mathrm{i}(\theta_U+\theta_V)}\\e^{\mathrm{i}\theta_V}&1&e^{-\mathrm{i}(\theta_U-\theta_V)}&e^{-\mathrm{i}\theta_U}\\e^{\mathrm{i}\theta_U}&e^{\mathrm{i}(\theta_U-\theta_V)}&1&e^{-\mathrm{i}\theta_V}\\e^{\mathrm{i}(\theta_U+\theta_V)}&e^{\mathrm{i}\theta_U}&e^{\mathrm{i}\theta_V}&1\end{pmatrix}.\tag{E6}$$

这表明，对于单量子比特幺正算子的张量积，在基 ((00)，(01)，(10)，(11)) 下，矩阵 M 关于反对角线转置对称。因此，$\mathcal{P}$ 中的量子信道属于一个维数为 9 的实向量空间（对角线占 1 维，第一行的复指数占 2×3 维，第二行的第三项占 2 维）。由于 $\mathcal{P}$ 中有 16 个量子信道，这个族并非线性无关。要求所有 $\mathbb{E}_{\theta}\left[\lambda_I(\theta)\right]$ 均为正显然过于严格。将其推广以找回我们先前推导出的条件的一种方法，是利用以下事实：

$$\begin{aligned}&\mathcal{E}[\mathbb{1}\otimes\mathbb{1}]+\mathcal{E}[\hat{Z}\otimes\hat{Z}]\\&+\mathcal{E}[\mathbb{1}\otimes\hat{Z}]+\mathcal{E}[\hat{Z}\otimes\mathbb{1}]\\&=\mathcal{E}[\hat{S}\otimes\hat{S}]+\mathcal{E}[\hat{S}^{\dagger}\otimes\hat{S}^{\dagger}]\\&+\mathcal{E}[\hat{S}\otimes\hat{S}^{\dagger}]+\mathcal{E}[\hat{S}^{\dagger}\otimes\hat{S}]\end{aligned}\tag{E7}$$

将 $\mathbb{E}_{\theta}\left[-\sin^2\theta\right]$ 因子吸收到与其他量子信道相关联的系数中。

注。 为了得到前述关系，我们使用了

以下量子信道:

$$
\begin{aligned}
M\left(\mathcal{E}\left[\mathbb{1}\otimes\hat{Z}\right]\right) &= \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix}, \\
M\left(\mathcal{E}\left[\hat{Z}\otimes\mathbb{1}\right]\right) &= \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}, \\
M\left(\mathcal{E}\left[\hat{S}\otimes\hat{S}^{\dagger}\right]\right) &= \begin{pmatrix} 1 & -\mathrm{i} & \mathrm{i} & 1 \\ \mathrm{i} & 1 & -1 & \mathrm{i} \\ -\mathrm{i} & -1 & 1 & -\mathrm{i} \\ 1 & -\mathrm{i} & \mathrm{i} & 1 \end{pmatrix}, \\
M\left(\mathcal{E}\left[\hat{S}^{\dagger}\otimes\hat{S}\right]\right) &= \begin{pmatrix} 1 & \mathrm{i} & -\mathrm{i} & 1 \\ -\mathrm{i} & 1 & -1 & -\mathrm{i} \\ \mathrm{i} & -1 & 1 & \mathrm{i} \\ 1 & \mathrm{i} & -\mathrm{i} & 1 \end{pmatrix}.
\end{aligned}
\tag{E8}
$$

我们证明，与 Z 旋转相关的 N 重通道总能分解为 Clifford 酉通道的实线性组合。然而，找到使 N 重通道能够分解为 Clifford 酉算子的凸组合的必要和/或充分条件，仍是一个未解决的问题；也就是说，找到使 N 重通道成为 Clifford 混合酉通道的条件。了解这些条件或许能够将本文提出的方案扩展到具有相关旋转参数的 ans¨atze。

附录 F：简单拟设的一阶和二阶 Clifford 近似电路示例

在本附录中，我们给出一个 Clifford 近似电路示例，用于估计简单量子线路（如图10所示）中的 $\mathbb{E}_{\boldsymbol{\theta}}\left[C(\boldsymbol{\theta})\right]$ 和 $\mathbb{E}_{\boldsymbol{\theta}}\left[C(\boldsymbol{\theta})^2\right]$。对于其他量，例如梯度平方的期望值，也可以从该示例推导出相应的 Clifford 近似；只需在固定层中引入适当的 Clifford 门，即可得到正确的估计量（参见第 II A 和 B 节）。该量子线路作用于三个量子比特，由两层旋转组成，各层之间交替排列着固定的双量子比特受控 Z 门。要为这些量子线路构造一阶近似，只需将每个旋转随机替换为恒等门（即一条连线）或与相应旋转门方向对应的 Pauli 门。图11展示了三个一阶 Clifford 近似示例。构造二阶近似时，首先利用恒等式 $\hat{X} = \hat{H}^{\dagger}\hat{Z}\hat{H}$ 和 $\hat{Y} = (\hat{S}\hat{H})\hat{Z}(\hat{S}\hat{H})^{\dagger}$，将每个绕 X 或 Y 轴的旋转映射为绕 Z 轴的旋转，其中 $\hat{H}$ 和 $\hat{S}$ 分别表示 Hadamard 门和相位门。因此，我们得到由 Z 旋转层与固定的 Clifford 门层交替组成的拟设，如图12所示。随后将该量子线路沿垂直方向复制一份，以得到

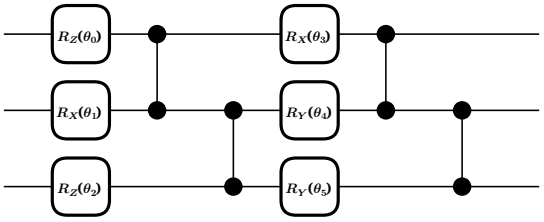


图 10. 初始变分电路，包含随机旋转角。

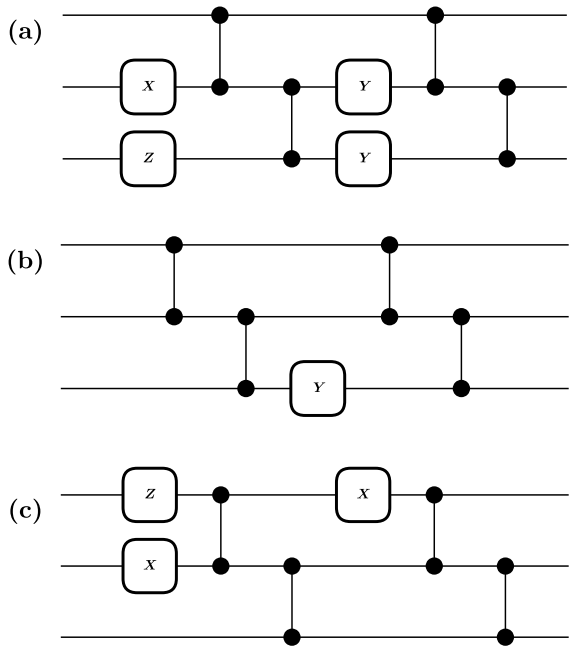


图 11. 图10所示拟设的一阶 Clifford 近似量子线路示例。假设角度的概率分布为偶分布，我们将每个旋转替换为一个根据公式（A4）采样得到的 Clifford 门。

一个作用于六个量子比特的量子线路。最后，根据图2所示方案，将共享相同角度的每对旋转随机替换为两个单量子比特门。

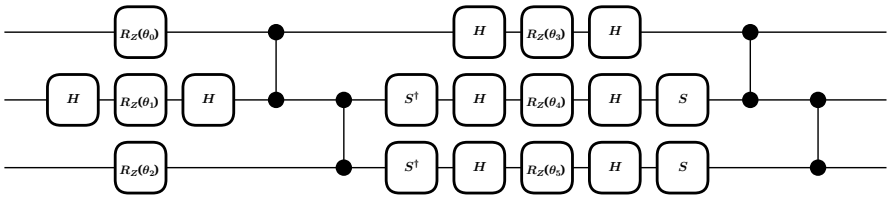
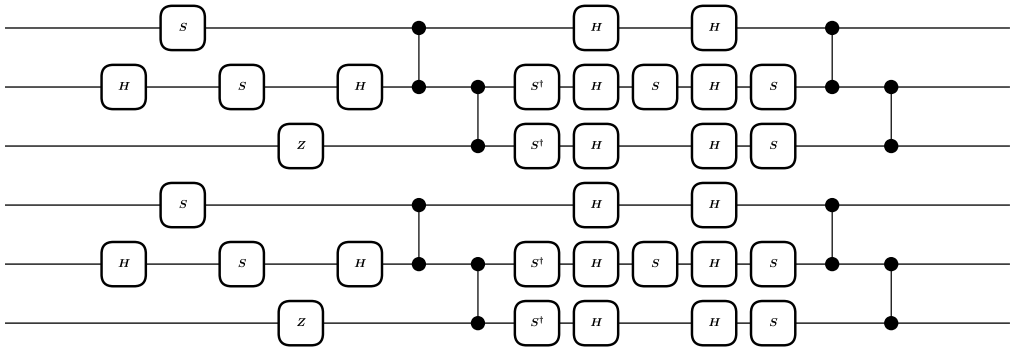
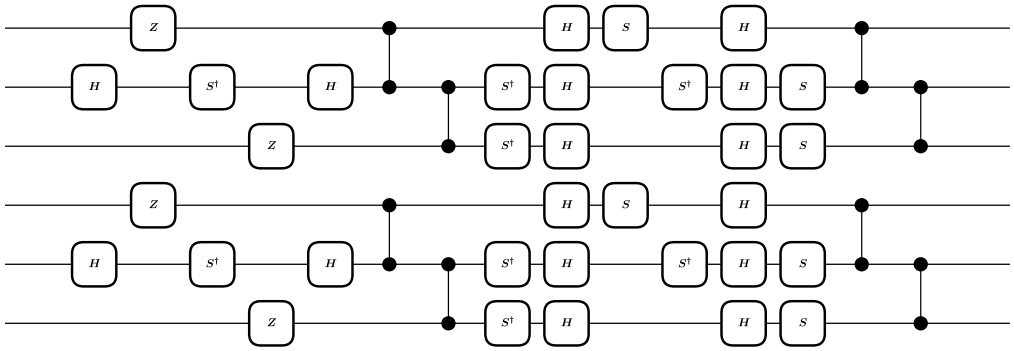


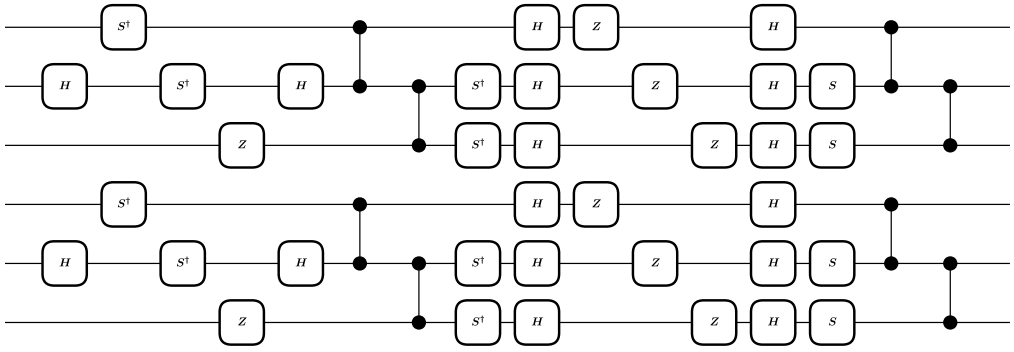
图12. 等　　价形式的初始量子线路，仅含 Z 旋　　转。



(a)



(b)



(c)

图13. 图中拟设的二阶Clifford近似电路示例。1　　0.

[1] M. Cerezo、A. Arrasmith、R. Babbush、S. C. Benjamin、S. Endo、K. Fujii、J. R. McClean、K. Mitarai、X. Yuan、L. Cincio 和 P. J. Coles，变分量子算法，《自然评论：物理学》 3，625 （ 2021 ）。[2] G. Carleo、I. Cirac、K. Cranmer、L. Daudet、M. Schuld、N. Tishby、L. Vogt-Maranto 和 L. Zdeborov´a，机器学习与物理科学，《现代物理评论》91，045002（2019）。[3] M. Cerezo、G. Verdon、H.-Y. Huang、L. Cincio 和 P. J. Coles，量子机器学习中的挑战与机遇，《自然：计算科学》2，567（2022）。[4] A. Peruzzo、J. McClean、P. Shadbolt、M.-H. Yung、X.-Q. Zhou、P. J. Love、A. Aspuru-Guzik 和 J. L. O'Brien，光子量子处理器上的变分本征值求解器，《自然·通讯》5，4213（2014）。[5] A. Kandala、A. Mezzacapo、K. Temme、M. Takita、M. Brink、J. M. Chow 和 J. M. Gambetta，适用于小分子和量子磁体的硬件高效变分量子本征值求解器，《自然》549，242（2017）。[6] Google AI Quantum及其合作者、F. Arute、K. Arya、R. Babbush、D. Bacon、J. C. Bardin、R. Barends、S. Boixo、M. Broughton、B. B. Buckley、D. A. Buell、B. Burkett、N. Bushnell、Y. Chen、Z. Chen、B. Chiaro、R. Collins、W. Courtney、S. Demura、A. Dunsworth、E. Farhi、A. Fowler、B. Foxen、C. Gidney、M. Giustina、R. Graff、S. Habegger、M. P. Harrigan、A. Ho、S. Hong、T. Huang、W. J. Huggins、L. Ioffe、S. V. Isakov、E. Jeffrey、Z. Jiang、C. Jones、D. Kafri、K. Kechedzhi、J. Kelly、S. Kim、P. V. Klimov、A. Korotkov、F. Kostritsa、D. Landhuis、P. Laptev、M. Lind mark、E. Lucero、O. Martin、J. M. Martinis、J. R. McClean、M. McEwen、A. Megrant、X. Mi、M. Mohseni、W. Mruczkiewicz、J. Mutus、O. Naaman、M. Neeley、C. Neill、H. Neven、M. Y. Niu、T. E. O'Brien、E. Ostby、A. Petukhov、H. Putterman、C. Quintana、P. Roushan、N. C. Rubin、D. Sank、K. J. Satzinger、V. Smelyanskiy、D. Strain、K. J. Sung、M. Szalay、T. Y. Takeshita、A. Vainsencher、T. White、N. Wiebe、Z. J. Yao、P. Yeh 和 A. Zalcman，超导量子比特量子计算机上的 Hartree-Fock 方法，《科学》369，1084（2020）。[7] E. Farhi、J. Goldstone 和 S. Gutmann，量子近似优化算法（2014），arXiv:1411.4028。[8] N. Lacroix、C. Hellings、C. K. Andersen、A. Di Paolo、A. Remm、S. Lazar、S. Krinner、G. J. Norris、M. Gabureac、J. Heinsoo、A. Blais、C. Eichler 和 A. Wallraff，利用连续门集合提升深度量子优化算法的性能，《PRX Quantum》1，020304（2020）。[9] M. P. Harrigan、K. J. Sung、M. Neeley、K. J. Satzinger、F. Arute、K. Arya、J. Atalaya、J. C. Bardin、R. Barends、S. Boixo、M. Broughton、B. B. Buckley、D. A. Buell、B. Burkett、N. Bushnell、Y. Chen、Z. Chen、Ben Chiaro、R. Collins、W. Courtney、S. Demura、D. Eppens、A. Fowler、B. Foxen、C. Gidney、M. Giustina、R. Graff、S. Habegger、A. Ho、S. Hong、T. Huang、L. B. Ioffe、S. V. Isakov、E. Jeffrey、Z. Jiang、C. Jones、D. Kafri、K. Kechedzhi、J. Kelly、S. Kim、P. V. Klimov、A. N. Korotkov、F. Kostritsa、D. Landhuis、P. Laptev、M. Lindmark、M. Leib、O. Martin、J. M. Martinis、J. R. McClean, M. McEwen, A. Megrant, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naa- man, C. Neill, F. Neukart, M. Y. Niu, T. E. O'Brien, B. O'Gorman, E. Ostby, A. Petukhov, H. Putterman, C. Quintana, P. Roushan, N. C. Rubin, D. Sank, A. Sko- lik, V. S melyanskiy, D. Strain, M. Streif, M. Szalay, A. Vainsencher, T. White, Z. J. Yao, P. Yeh, A. Zal- cman, L. Zhou, H. Neven, D. Bacon, E. Lucero, E. Farhi, and R. Babbush，平面超导处理器上对非平面图问题进行量子近似优化，《自然·物理》17, 332 (2021)。[10] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Bab- bush, and H. Neven，量子神经网络训练景观中的贫瘠高原，《自然·通讯》9, 4812 2018( )。[11] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles，将拟设可表达性与梯度幅值及贫瘠高原联系起来，PRX Quantum 3, 010313 (2022)。[12] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles，变分量子算法中的噪声诱导贫瘠高原，《自然·通讯》12, 6961 (2021)。[13] C. Ortiz Marrero, M. Kieferov´a, and N. Wiebe，纠缠诱导的贫瘠高原，PRX Quantum 2, 040316 (2021)。[14] A. V. Uvarov and J. D. Biamonte，论变分量子算法中的贫瘠高原与代价函数局域性，Journal of Physics A: Mathematical and Theoretical 54, 245301 (2021)。[15] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles，浅层参数化量子电路中依赖代价函数的贫瘠高原，《自然·通讯》12, 1791 (2021)。[16] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin，纠缠设计的贫瘠高原缓解方法，Physical Review Research 3, 033090 (2021)。[17] R. Wiersema, C. Zhou, J. F. Carrasquilla, and Y. B. Kim，变分量子电路中的测量诱导纠缠相变 (2021)，arXiv:2111.08035。[18] J. Kim and Y. Oz，高效量子计算的纠缠诊断，Journal of Statistical Mechanics: Theory and Experiment 2022, 073101 (2022), arXiv:2102.12534。[19] J. Kim and Y. Oz，量子能量景观与电路优化，《物理评论A》106, 052424 (2022)。[20] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng, and M. Serbyn，利用经典影子避免贫瘠高原，PRX Quantum 3, 020365 (2022)。[21] L. Friedrich and J. Maziero，利用经典深度神经网络避免贫瘠高原，《物理评论A》106, 042433 (2022)。[22] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti，一种用于解决参数化量子电路中贫瘠高原问题的初始化策略，《Quantum》3, 214 (2019)。[23]H.-Y. Liu, Z.-Y. Chen, T.-P. Sun, Y.-C. Wu, Y.- J. Han, and G.-P. Guo，受迁移学习启发的参数初始化以缓解贫瘠高原 (2022)，arXiv:2112.10952。[24] K. Mitarai, Y. Suzuki, W. Mizukami, Y. O. Nakagawa, and K. Fujii，用于高效

变分量子算法基准测试与初始化的二次克利福德展开，Physical Review Research 4, 033012 (2022)。[25] G. S. Ravi, P. Gokhale, Y. Ding, W. M. Kirby, K. N. Smith, J. M. Baker, P. J. Love, H. Hoffmann, K. R. Brown, and F. T. Chong, CAFQA：变分量子算法的经典模拟自举 (2022), arXiv:2202.12924。[26] J. Kim, J. Kim, and D. Rosa，变分本征问题中高深度电路的普遍有效性，Physical Review Research 3, 023203 (2021)。[27] J. Kim, Y. Oz, and D. Rosa，量子混沌与电路参数优化 (2022)，arXiv:2201.01452。[28] M. H. Cheng, K. E. Khosla, C. N. Self, M. Lin, B. X. Li, A. C. Medina, and M. S. Kim，变分量子算法的克利福德电路初始化 (2022), arXiv:2207.01539。[29] J. Dborin, F. Barratt, V. Wimalaweera, L. Wright, and A. G. Green，面向量子机器学习的矩阵乘积态预训练，Quantum Science and Technology 7, 035014 (2022)。[30] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sorn- borger, and P. J. Coles，量子卷积神经网络中不存在贫瘠高原，Physical Review X 11, 041011 (2021)。[31] L. Schatzki, M. Larocca, Q. T. Nguyen, F. Sauvage, and M. Cerezo，置换等变量子神经网络的理论保证 (2022), arXiv:2210.09974。[32]Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Al- brecht, and A. T. Sornborger，贫瘠高原阻碍搅乱器的学习，《物理评论快报》126, 190501 2021( )。[33] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, 贫瘠高原对无梯度优化的影响，《Quantum》5, 558 (2021)。[34] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles，量子贫瘠高原等价于代价集中与狭窄峡谷，Quantum Science and Technology 7, 045015 (2022)。[35] S. Wang, P. Czarnik, A. Arrasmith, M. Cerezo, L. Cincio, and P. J. Coles，误差缓解能否提升含噪变分量子算法的可训练性？(2021), arXiv:2109.01051。[36]Y. Du, T. Huang, S. You, M.-H. Hsieh, and D. Tao，变分量子算法的量子电路架构搜索，npj Quantum Information 8, 1 (2022)。[37] K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles，耗散型感知机量子神经网络的可训练性，《物理评论快报》128, 180505 (2022)。[38] G. De Palma, M. Marvian, C. Rouz´e, and D. S. Fran˙ca，变分量子算法的局限性：量子最优传输方法，PRX Quantum 4, 010309 (2023)。[39] V. Heyraud, Z. Li, Z. Denis, A. Le Boit´e, and C. Ciuti，含噪量子核机器，《物理评论A》106, 052421 2022( )。[40] S. Jerbi, L. J. Fiderer, H. Poulsen Nautrup, J. M. K¨ubler, H. J. Briegel, and V. Dunjko，超越核方法的量子机器学习，《自然·通讯》14, 517 (2023)。[41] Z. Li, V. Heyraud, K. Donatella, Z. Denis, and C. Ciuti，基于相对论启发量子动力学的机器学习，《物理评论A》106, 032413 (2022)。[42] M. Schuld，《监督量子机器学习模型是核方法》（2021），arXiv:2101.11020。[43] P. Rebentrost、M. Mohseni 和 S. Lloyd，《用于大数据分类的量子支持向量机》，《物理评论快报》113，130503（2014）。[44] P. Mujal、R. Mart´ınez-Pe˜na、J. Nokkala、J. Garc´ıa-Beni、G. L. Giorgi、M. C. Soriano 和 R. Zambrini，《量子储备池计算与极限学习机的机遇》，《先进量子技术》4，2100027 2021( )。[45] Z. Denis、I. Favero 和 C. Ciuti，《用于超快光谱分析的光子核机器学习》，《物理评论应用》17，034077（2022）。[46] G. Marcucci、D. Pierangeli 和 C. Conti，《波的神经形态计算理论：利用怪波、色散激波和孤子进行机器学习》，《物理评论快报》125，093901（2020）。[47] D. Pierangeli、G. Marcucci 和 C. Conti，《通过自由空间光传播实现的光子极限学习机》，《光子学研究》9，1446（2021）。[48] S. Thanasilp、S. Wang、M. Cerezo 和 Z. Holmes，《量子核方法中的指数集中与不可训练性》（2022），arXiv:2208.11060。[49] J. Watrous, *The Theory of Quantum Information*（剑桥大学出版社，2018）。[50] M. A. Nielsen 和 I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* 2010（剑桥大学出版社，）。[51] D. Gottesman，《量子计算机的海森堡表象》（1998），arXiv:quant-ph/9807006。[52] S. Aaronson 和 D. Gottesman，《稳定子电路的改进模拟》，《物理评论A》70，052328（2004）。ˆ ˆ[53]我们将 H 称为哈达玛门，将 S 称为相位门；二者都属于 Clifford 群。对于 X-旋转，有 $\hat{}X = \hat{}H^\dagger \hat{}Z \hat{}H$，因此 $e^{-i\frac{\theta_i}{2}\hat{}X} = \hat{}H^\dagger e^{-i\frac{\theta_i}{2}\hat{}Z} \hat{}H$，我们可以分别用 HW_i 和$W_{i+1}H$ 替换 $\hat{}W_i$ 和 $\hat{}W_{i+1}$，从而得到另一个形式与原拟设相同、且仅包含 Y和Z 旋转的拟设。对于 Y-旋转，我们同样利用 $\hat{}Y = (\hat{}S\hat{}H)\hat{}Z(\hat{}S\hat{}H)^\dagger$这一事实进行处理。请注意，在最后一层的情况下，必须将额外量子门中的一个并入代价函数可观测量，才能得到相同的拟设结构。[54] K. Mitarai、M. Negoro、M. Kitagawa 和 K. Fujii，《量子线路学习》，《物理评论A》98，032309（2018），arXiv:1803.00745。[55] M. Schuld、V. Bergholm、C. Gogolin、J. Izaac 和 N. Killoran，《在量子硬件上评估解析梯度》，《物理评论A》99，032331（2019）。[56] D. A. Roberts 和 B. Yoshida，《设计产生的混沌与复杂性》，《高能物理学杂志》2017，121（2017）。[57] S. Sim、P. D. Johnson 和 A. Aspuru-Guzik，《用于量子-经典混合算法的参数化量子电路的可表达性与纠缠能力》，《先进量子技术》2，1900070（2019）。[58] K. Nakaji 和 N. Yamamoto，《用于量子计算的交替分层拟设的可表达性》，《Quantum》5，434（2021）。[59] D. Gross、K. Audenaert 和 J. Eisert，《均匀分布的酉算子：论酉设计的结构》，《数学物理杂志》48，052104（2007）。[60] J. T. Iosue、K. Sharma、M. J. Gullans 和 V. V. Albert，《连续变量量子态设计：理论与

应用》（2022），arXiv:2211.05127。[61] A. W. Harrow 和 R. A. Low，《随机量子电路是近似 2- 设计》，《数学物理通讯》291，257（2009）。[62] F. G. S. L. Brand˜ao、A. W. Harrow 和 M. Horodecki，《局域随机量子电路是近似多项式设计》，《数学物理通讯》346，397（2016）。[63] J. Haferkamp，《随机量子电路在深度 $O\left(nt^{5+o(1)}\right)$下是近似酉 t-设计》，《Quantum》6，795（2022）。[64] J. R. McClean、J. Romero、R. Babbush 和 A. Aspuru-Guzik，《变分量子-经典混合算法理论》，《新物理学杂志》18，023023（2016）。[65] I. Goodfellow、Y. Bengio 和 A. Courville，*Deep Learning*，F. Bach 编，《自适应计算与机器学习丛书》（麻省理工学院出版社，美国马萨诸塞州剑桥，2016）。[66] J. Liu、Z. Lin 和 L. Jiang，《机器学习中的惰性、贫瘠高原与噪声》（2022），arxiv:2206.09313。

[67] E. R. Anschuetz 和 B. T. Kiani，《量子变分算法深陷陷阱》，《自然·通讯》13，7760（2022）。[68] L. Bittel 和 M. Kliesch，《训练变分量子算法》是 NP-困难，《物理评论快报》127，120502（2021）。[69] M. Cerezo 和 P. J. Coles，量子神经网络在贫瘠高原上的高阶导数，《量子科学与技术》6，035006（2021）。[70] 这涵盖了关于角度 $k\pi/2$ 对称的分布，其中 $k \in \mathbb{Z}$。在这种情况下，偏差可以作为一个额外的固定 Clifford 门因子提出。[71] C. Zhao 和 X.-S. Gao，使用 ZX演算分析量子神经网络训练中的贫瘠高原现象，《Quantum》5，466（2021）。[72] C. Piveteau、D. Sutter 和 S. Woerner，降低采样开销的准概率分解，npj Quantum Information 8，1（2022）。[73] C. McDiarmid *et al.*，论有界差分方法，《组合学综述》141，148（1989）。[74] M. Mohri、A. Rostamizadeh 和 A. Talwalkar，*Foundations of Machine Learning, Second Edition*（麻省理工学院出版社，2018）。[75] B. Baumgartner，关于矩阵乘积迹的一个不等式，使用绝对值（2011），arxiv:1106.6189。