用于量子层析的朗之万采样器

Tameem Adel*、Abhishek Agarwal*、St´ephaneChr´etien†、Estelle Massart‡§¶、Danila Mokeev‡、Ivan Rungger*‖、Andrew Thompson*

2026年1月14日

arXiv:2601.08775v1 [math.ST] 13 Jan 2026

摘要：量子层析旨在根据实验结果获取制备量子态的完整经典描述。我们提出一种用于量子层析的朗之万采样器，该采样器依赖于贝叶斯量子层析的一种新形式，此形式利用厄米半正定矩阵的Burer-Monteiro分解。若已知目标密度矩阵的秩，该形式使我们能够定义一个后验分布，其支撑集仅包含秩不超过目标密度矩阵秩的矩阵。反之，若目标秩未知，我们的算法可使用任意秩上界；此外，通过采用低秩促进先验密度，所得后验均值估计量的秩还能进一步降低。该先验密度是对 [《Annales de l'Institut Henri Poincar´e Probability and Statistics, 56(2):1465–1483,2020]中提出方法的复数域扩展。我们为所提出的估计量推导出一个PAC-贝叶斯界，其效果与文献中已知的最佳界相当；数值结果还表明，当密度矩阵的秩较小时，与现有技术相比，该方法在可扩展性方面有显著提升。

1 引言

自 90 年代中期以来，制备、控制和维持量子态的能力取得了长足进步，如今已研发出拥有数百个量子比特的设备 [BEG+24, Goo24]。随着这些进步而来的一项任务，是验证已制备量子态的正确性，这也有助于表征设备的不同组件 [LAZ+25]。一种广泛采用的表征方法，是根据实验结果获取所制备量子态的完整经典描述，这种方法称为量子态层析，或简称量子层析 [VR89]。量子层析方法广泛用于表征量子计算机 [PYBBK25], 、量子传感器 [FKuRS22], 和量子通信网络 [BHK+19]。

一般而言，由于希尔伯特空间的维数呈指数级增长，量子层析所需的实验测量次数也会随量子系统的规模呈指数级增长。因此，它通常仅用于表征量子比特数较少的量子态。单个量子比特是一个二能级量子系统，而一个 n 量子比特量子系统的完整描述由其密度矩阵 ρ^0 给出，其中 $\rho^0 \in \mathbb{C}^{d\times d}$，且

*英国国家物理实验室，英国泰丁顿，TW11 0LW。
†ERIC实验室，里昂第二大学，法国布龙，Pierre Mend`es France 大道 5 号，邮编 69500。
‡ICTEAM，鲁汶大学，Georges Lemaître 大道 4 号，1348 新鲁汶，比利时。§本研究是在作者于英国牛津大学数学研究所从事博士后研究期间启动并部分完成的。
¶通讯作者：estelle.massart@uclouvain.be。
‖伦敦大学皇家霍洛威学院计算机科学系，英国埃格姆，TW20 0EX。

$d = 2^n$。密度矩阵可通过对 ρ^0 的多个副本构成的系综进行测量来估计；这些副本通常是在每次测量后反复重新初始化系统并制备量子态而获得的。因此，量子层析就是根据重复态制备的一组测量结果，求得一个能够表征二能级 n 量子比特系统未知量子态的密度矩阵 ρ^0 。

目前已有多种方法可根据实验结果重构密度矩阵；近期综述见 [GSG+23]。虽然最大似然估计量备受关注 [GKKT20], ，但也有多项研究探讨了用于量子层析的贝叶斯形式，从而能够量化状态的不确定性[BK10, HH12, GCC16, MA17, LLJL20, Mai22]。量化这种不确定性十分重要，因为它有助于考虑数据中有限的统计精度，并能将量子态的真实特征与有限统计量引入的伪特征区分开来[SRmG17, ALL+24, ALL+25]。给定适当选取的先验密度和似然函数，贝叶斯量子层析会在密度矩阵集合上定义一个后验分布，其期望值为系统的密度矩阵提供了自然的估计。然而在实践中，计算该期望值需要求解高维积分，这在计算上颇具挑战。现有的量子层析方法依赖马尔可夫链蒙特卡罗（MCMC）采样，通常结合梅特罗波利斯-黑斯廷斯算法，也可能采用预条件克兰克-尼科尔森方法加以改进，见 [Mai22]。本文提出一种基于朗之万采样的替代方法。

朗之万动力学被视为一类 MCMC 技术 [Nea11]。当直接采样并不简单时，通常会使用朗之万动力学从概率分布中获取一系列随机样本。基于朗之万动力学的一般算法会通过计算目标概率密度函数的对数后验梯度来提出新的状态。

1.1 相关工作

量子层析中最早且最直接的方法，就是对由玻恩规则得出的方程进行求逆；玻恩规则将测量结果的概率与由密度矩阵表示的量子态的经典描述联系起来 ρ^0。可以说，这种在文献中称为线性反演或直接反演的方法，主要缺点在于它可能无法得到具有密度矩阵结构性质的估计量，因此需要额外执行投影步骤，将结果投影到可行密度矩阵集合上 [GKKT20]。因此，最大似然估计——即为给定的一组测量结果寻找最可能的密度矩阵 [H ˇRFJ04], ——成为量子层析的首选方法。

受捕捉估计不确定性的可能性启发，越来越多的研究采用贝叶斯形式进行量子层析 [Jon91, BDAK98, BK10, KSR+13, KF15]。例如， [BK10]的作者提出了一种量子层析贝叶斯框架，其中估计得到的密度矩阵取后验分布的均值，并使用梅特罗波利斯-黑斯廷斯算法进行近似。 [HH12],将这一方法加以扩展，提出了一种自适应贝叶斯量子层析方法，为贝叶斯量子层析引入序贯重要性采样策略，以自适应地确定实验中要采用的新测量。此外，在 [GCC16] 中，作者处理了时变状态，并提出了颇具启发性的先验分布，即由默认（无信息）先验分布构建而成、且具有指定均值的先验分布族；这些均值可根据实验设计或先前的实验估计等信息来选取。

在量子计算领域，人们通常制备由秩-1密度矩阵 [NC10]表征的纯态。然而，量子计算机中存在的非相干噪声会导致系统由混态 [Pre98]描述。在某些情况下，例如

噪声仅影响局部量子比特时，所得量子态具有低秩 [GLF+10]，这一低秩特性可用于简化量子层析。例如，[ABH+13] 进一步将线性投影估计量投影到固定秩密度矩阵集合上，并证明了其所提估计量的一致性。另一种方法是， [GKD12] 构造一系列秩受限的最大似然模型，并从中选出在风险与模型复杂度之间实现最佳权衡的模型；模型复杂度则使用著名的赤池信息准则（AIC）或贝叶斯信息准则（BIC）来衡量。压缩感知等其他方法也已用于低秩状态的量子层析 [GLF+10]。

对于贝叶斯量子层析，可以通过采用合适的先验密度来考虑秩信息；例如， [GCC16] 中使用的先验就基于 Ginibre 系综。另一种做法是， [MA17] 的作者将密度矩阵改写为若干秩一项之和，并以根据狄利克雷分布采样的因子进行缩放，从而促进其稀疏性。他们提出了两种沿用贝叶斯量子层析形式体系的估计量：密度估计量 （其中似然刻画了与密度矩阵最小二乘估计之间的距离）和概率估计器 （其中似然刻画了理论状态概率向量与经验状态概率向量之间的距离，这与本文采用的方法一致）。作者使用梅特罗波利斯-黑斯廷斯算法计算这些估计量，并推导出相应的 PAC-贝叶斯界。由于梅特罗波利斯-黑斯廷斯算法在高维系统中速度较慢， [Mai22], 提出了另一种高效的自适应梅特罗波利斯-黑斯廷斯算法实现，用于概率估计器，并采用 Crank-Nicholson 预条件。 [LLJL20]也将类似方法用于密度估计量。最后，我们提及较新的工作 [QFN21] ，该工作提出了一种自适应量子态层析算法，与 [HH12], 的做法一致，即使用循环神经网络执行标准的贝叶斯更新。

1.2 主要贡献

利用厄米半正定矩阵著名的 Burer-Monteiro分解，我们将采样问题表述为关于密度矩阵低秩因子的问题，并为由此得到的后验分布推导出一种朗之万采样器。为此，我们提出了一种新的低秩促进先验密度，它是 [Dal20]提出的先验的一种复数推广。对于我们提出的模型，我们在完备测量设置下推导出 PAC-贝叶斯界，得到与 [MA17, Mai22], 所得结果相同的主导收敛速率（相差一个常数因子或对数因子），并将我们的方法与 [MA17]提出的马尔可夫链蒙特卡罗采样器进行数值比较。我们表明，尽管两种估计量的精度相当，但我们提出的朗之万采样器比 [MA17]的马尔可夫链蒙特卡罗采样器所需的迭代次数更少。此外，数值结果还表明，当已知密度矩阵的秩较小（因而可以采用维数更小的 Burer-Monteiro因子）时，朗之万采样算法的每次迭代计算成本远低于 [MA17]所提出采样器的每次迭代计算成本。

2 贝叶斯量子层析

在这里，我们讨论对一个 2 级 n量子比特系统的密度矩阵 $\rho^0 \in \mathbb{C}^{d\times d}$ 进行估计，这意味着 $d = 2^n$。密度矩阵 ρ^0 可以分解为纯本征态的凸组合，

$$\rho^0 = \sum_{i=1}^{d} \lambda_i v_i v_i^*,$$

其中，$\{v_i v_i^*\}$ 是纯本征态，$\{\lambda_i\}$ 是相应的特征值。由此可知，ρ^0 是一个迹等于一的复厄米半正定矩阵。因此，一个 n量子比特系统的密度矩阵集合对应于复谱体：

$$\mathcal{CS} = \{\rho \in \mathbb{C}^{d\times d} : \rho = \rho^*, \rho \succeq 0, \mathrm{tr}(\rho) = 1\}, \tag{1}$$

假设先验密度 μ可用。

类似于 [MA17], ，我们假设，对于每个量子比特，都可以测量三个泡利可观测量 σ_x、σ_y、σ_z 中的一个。因此，共有 3^n 个可能的实验，其结果向量属于 $\{-1,1\}^n$。令 R^a 为实验 $a \in \{1,\ldots,3^n\}$ 定义的随机向量，令 $\mathcal{P}_s^a$ 为与实验 $a \in \{1,\ldots,3^n\}$ 和结果向量 $o(s) \in \{-1,1\}^n$ 相关联的测量算子，其中 $s \in \{1,\ldots,2^n\}$。记

$$p_{a,s}^0 := \mathrm{prob}[R^a = o(s)],$$

实验 a 得到结果 s 的概率，根据玻恩规则，该概率为

$$p_{a,s}^0 = \mathrm{trace}(\mathcal{P}_s^a \cdot \rho^0).$$

虽然真实概率 $p_{a,\ s}^0$ 通常无法获得，但可以利用与系统多次重复试验相关联的测量结果对其进行近似。假设实验 a 重复 m_a 次，则真实概率 $p_{a,\ s}^0$ 可以近似为经验概率

$$\hat{p}_{a,s} := \frac{1}{m_a}\sum_{i=1}^{m_a} 1_{R_i^a = s}, \tag{2}$$

其中，R_i^a是 R^a的第 i次重复实验，$i = 1$、.....、m_a。这些经验频率可用于推导似然函数[1], ，在此我们将其选为

$$L : \mathbb{C}^{d\times d} \to \mathbb{R} : \rho \mapsto \sum_{s=1}^{2^n}\sum_{a=1}^{3^n}(\hat{p}_{a,s} - \mathrm{tr}(\mathcal{P}_s^a \cdot \rho))^2. \tag{3}$$

给定适当选取的先验密度 μ，贝叶斯量子层析定义了一个后验分布[2]

$$\hat{\mu}_\lambda(\rho) \propto \exp\left(-\lambda L(\rho)\right)\mu(\rho)$$

在密度矩阵集合上，其中 $\lambda > 0$ 是一个经过适当选择、用于平衡似然与先验信息的参数。随后可以通过例如后验密度[3]的经验均值来估计密度矩阵 ρ^0 ：

$$\hat{\rho}_\lambda = \int_{\rho\in\mathcal{CS}} \rho\hat{\mu}_\lambda(\rho)\mathrm{d}\rho. \tag{4}$$

一种计算该估计量的可能策略是采用朗之万采样，通过生成马尔可夫链来对后验分布 $\hat{}\mu_\lambda$ 进行采样

$$\rho_{k+1} = \rho_k - \eta_k\nabla f(\rho_k) + \sqrt{\frac{2\eta_k}{\beta_k}}w_k, \tag{5}$$

[1]这里我们承认在术语使用上存在一点不严谨之处，因为该 *likelihood* 并非来自测量的噪声模型，因此在 [MA17] 中被称为伪似然（相应得到的后验分布同样被称为伪后验）。[2]由于该似然不一定来自测量的噪声模型，因此在 [MA17]中，后验分布被称为伪后验。[3]这一选择与 [MA17、Mai22、LLJL20]中的选择一致。请注意，选择多峰先验可能会促使人们改用最大 *a posteriori* 估计量。

该链从某个迭代点 ρ_0开始，其中 $f(\rho) := -\log(\hat{\mu}_\lambda(\rho))$ 是（相反的）对数后验， η_k 是步长， β_k 是温度超参数[4], ，而 w_k 是一个复随机矩阵，其实部和虚部均服从标准高斯分布。

3 用于贝叶斯量子层析的 Burer-Monteiro 公式

请注意，式（5）在 $d \times d$ 密度矩阵空间中生成一个马尔可夫链，但没有考虑到所求的密度通常是低秩的，因此采样空间的维数可能可以降低。我们转而采用著名的 Burer-Monteiro 方法 [BM03], ，并将任意密度矩阵 $\rho \in \mathbb{C}^{d\times d}$表示为一个因子 $Y \in \mathbb{C}^{d\times r}$ ，其中假设其秩为 r，且满足 $\rho = YY^*$。这种改写有两方面的优势。首先，该公式使参数空间的维数从 d^2 个复数降至 dr 个复数，进而降低了待采样后验密度的维数。其次，该方法自然满足密度矩阵的厄米半正定性，因为对于任意 $Y \in \mathbb{C}^{d\times r}$，矩阵 YY^* 都是秩至多为 r的厄米半正定矩阵。根据迹和弗罗贝尼乌斯范数的定义，对于所有 $\rho = YY^*$ ，均有

$$\mathrm{tr}(\rho) = \mathrm{tr}(YY^*) = \|Y\|_{\mathrm{F}}^2,$$

密度矩阵的单位迹约束等价于要求这些因子属于复超球面

$$\mathcal{CS}^{d\times r} = \{Y \in \mathbb{C}^{d\times r} : \|Y\|_{\mathrm{F}} = 1\},$$

似然（3）变为

$$L^{\mathrm{BM}} : \mathbb{C}^{d\times r} \to \mathbb{R} : Y \mapsto \sum_{s=1}^{2^n}\sum_{a=1}^{3^n}(\hat{p}_{a,s} - \mathrm{tr}(\mathcal{P}_s^a \cdot YY^*))^2, \tag{6}$$

其中，BM 下标代表“Burer-Monteiro”。因此，在 $\mathbb{C}^{d\times r}$ （更准确地说，在复超球面 $\mathcal{CS}^{d\times r}$）上定义的任何先验分布，都会自然地导出关于 $\mathbb{C}^{d\times r}$ 的后验分布，进而可用于推导目标密度矩阵 $\rho^0 = Y^0Y^{0*}$ 的估计量。不过需要注意的是，使用 Burer-Monteiro分解依赖于已知目标密度矩阵的秩这一假设。当后者未知时，可以使用目标密度矩阵秩的任何上界（包括在缺乏相关信息时选择 $r = d$）。在这种情况下，我们认为，通过选择合适的先验密度，可以进一步提高低秩性。

一种能够诱导密度矩阵低秩性的先验已在 [MA17], 中提出，其中密度矩阵表示为乘积

$$\rho = V\Gamma V^*,$$

其中 V 是列向量范数为 1（不一定正交）的矩阵， Γ 是元素非负的对角矩阵[5] γ_1，……， γ_d。随后将先验选为（γ_1，……， γ_d） $\sim \mathcal{D}ir$（α_1，……， α_d），

[4]尽管在朗之万采样文献中，通常将 β_k 设为 1（例如，参见 [Dal20]），但此处引入这一温度参数是为了提高灵活性。在本文中，我们假设 β_k 在所有迭代中保持固定，并通过交叉验证进行学习。这相当于仅对噪声方差进行缩放，或者等价地，对数后验密度 f进行缩放。[5]由于 V 的列并非正交归一，因此不能将 Γ 的对角元素解释为 ρ的特征值；不过，采用这种方法的原因在于，对于作者使用的均匀密度，采样单位范数向量比采样具有正交归一列的矩阵成本更低。

其中 $\mathcal{D}ir(\alpha_1, \ldots\ldots, \alpha_d)$ 是参数为 $\alpha_1, \ldots\ldots, \alpha_d > 0$的狄利克雷分布，且$V_{:,i} \sim Unif(\mathcal{S}^{d-1})$，其中$V_{:,i}$是 V的第 i列，$Unif(\mathcal{S}^{d-1})$ 是单位复超球面上的均匀先验。为促进稀疏性，[MA17] 的作者选择 $\alpha_1 = \cdots = \alpha_d = \alpha$，其中某个 α接近于零（例如 $\alpha_1 = \cdots = \alpha_d = 1/d$）。

这里我们采用另一种先验，它是文献 [Dal20], 中所述谱缩放 Student t 分布的复数扩展，其表达式使其成为我们 Burer-Monteiro 公式的自然选择：

$$\nu_\theta(Y) = C_\theta \det(\theta^2 I_d + YY^*)^{-(2d+r+2)/2}, \tag{7}$$

对于 $Y \in \mathbb{C}^{d\times r}$，其中 $\theta > 0$是分布参数，且 $C_\theta = (\int_{\mathbb{C}^{d\times r}} \det(\theta^2 I_d + YY^*)^{-(2d+r+2)/2} \mathrm{d}Y)^{-1}$是归一化常数。与文献 [Dal20],类似，注意到

$$\nu_\theta(Y) \propto \Pi_{i=1}^r (\theta^2 + \sigma_i^2)^{-(2d+r+2)/2},$$

其中 $\sigma_1, \ldots, \sigma_r$是 Y的奇异值。因此，我们提出的先验相当于假设 Y 的奇异值服从缩放 Student t 分布，从而促进其稀疏性；例如，参见 [DT12]。由此可知，该先验促进矩阵 YY^* 的低秩性。

如下一个结果所述，这一结果是对 [Dal20, 引理 1], 的直接推广： Y 的列边缘服从 d元复缩放 Student t 分布。

引理 1. 若 Y是一个随机的 $d \times r$ 复矩阵，其密度由函数 ν_θ给出，那么对于 $i = 1, \ldots, r$， Y的列向量 $y_i \in \mathbb{C}^d$服从 d元复缩放 *Student*t分布 $(\sqrt{2}/3\theta) t_{3,d}$。因此，对于所有 i，均有 $\int_{\mathbb{C}^{d\times r}} \|y_i\|^2 \nu_\theta(Y)\mathrm{d}Y = 2\theta^2 d$ 。

在本文中，我们采用这种谱缩放 Student t 先验分布 $\nu_\theta(Y)$，它与似然（6）结合后给出后验密度

$$\hat{\nu}_{\lambda,\theta}^{\mathrm{BM}}(Y) \propto \exp(-\lambda L^{\mathrm{BM}}(Y)) \nu_\theta(Y), \tag{8}$$

我们的密度矩阵估计器为

$$\hat{\rho}_{\lambda,\theta}^{\mathrm{BM}} := \int_{\mathbb{C}^{d\times r}} YY^* \hat{\nu}_{\lambda,\theta}^{\mathrm{BM}}(Y)\mathrm{d}Y. \tag{9}$$

[MA17]的作者为其估计量推导了一个 PAC-贝叶斯界。我们在此将他们的分析扩展到我们的另一种先验，证明这一新先验保留了 [MA17]中达到的收敛速率。与 [MA17], 类似，我们假设采用完备测量设置，即每个实验 $a \in \{1, \ldots, 3^n\}$ 均重复进行 $m_a = m$ 次，每次采用 m 给定的数量，从而总共得到 $N_{\mathrm{tot}} = m3^n$ 次测量。以下结果给出了我们提出的估计器 $\hat{\rho}_{\lambda,\theta}^{\mathrm{BM}}$（定义于式（9））的 PAC-贝叶斯界。

定理1 (PAC-贝叶斯界)。 设 $\lambda = 3m/8$，并设 $\bar{Y} \in \mathbb{C}^{d\times r}$ 满足 $\|Y^0 - \bar{Y}\|_{\mathrm{F}} \|Y^0 + \bar{Y}\|_{\mathrm{F}} \leq 3^{-3n/2} 2^{-n/2}/m$。令 p 为 $\bar{Y}$的秩。对于任意 $\epsilon \in]0, 1[$，以概率$1-\epsilon$成立

$$\begin{aligned}\|\hat{\rho}_{\lambda,\theta}^{\mathrm{BM}} - \rho^0\|_{\mathrm{F}}^2 \leq \frac{3}{N_{\mathrm{tot}}} \left(3^{3n/4} 2^{(n+6)/4} (r + \sqrt{r}\|\bar{Y}\|_{\mathrm{F}}) + \frac{2r}{m} + 1\right) \\ + \frac{3^n 8}{2^n N_{\mathrm{tot}}} \left(\log\left(\frac{2}{\epsilon}\right) + 2p(2^{n+1} + r + 2) \log\left(1 + \frac{\|\bar{Y}\|_2}{\theta}\right)\right).\end{aligned}$$

虽然定理1中的上界取决于任意矩阵 Y，但请注意，选择 $\bar{Y} = Y^0$ 会得到 $p = \mathrm{rank}(\rho^0)$，在这种情况下，当 n 取较大值时，我们可以恢复 [MA17], 中得到的收敛速率$3^n \mathrm{rank}(\rho_0)/N_{\mathrm{tot}}$，相差一个常数因子或对数因子。

4 实现与数值结果

4.1 算法实现

为避免在实现中处理复数，我们引入变量替换，这依赖于一个众所周知的向量空间同构： $\mathbb{C}^{d\times r}$ 与 $\mathbb{R}^{2d\times 2r}$ 的一个子集之间存在向量空间同构（例如，参见 [OTKP12]）：

$$\psi : M^R + iM^I \mapsto \frac{1}{\sqrt{2}}\begin{pmatrix} M^R & -M^I \\ M^I & M^R \end{pmatrix}. \tag{10}$$

由此可直接验证，对于任意 $Y \in \mathbb{C}^{d\times r}$,

$$\psi(YY^*) = \sqrt{2}\psi(Y)\psi(Y)^\top. \tag{11}$$

下一个引理的证明见附录，它使我们能够将后验密度完全改写为实值矩阵的形式。

引理 2. 对于所有 $M, N \in \mathbb{C}^{d\times d}$ 厄米半正定矩阵，均有 $\mathrm{tr}(MN) = \mathrm{tr}(\psi(M)\psi(N))$ 且 $\det(M) = \sqrt{2^d \det(\psi(M))}$。

我们不再从复数 $d \times r$ 矩阵集合上的后验密度 $\nu^{\mathrm{BM}}_{\lambda,\theta}$ 中采样，而是从后验分布 $\tilde{\nu}^{\mathrm{BM}}_{\lambda,\theta}(\tilde{Y}) := \nu^{\mathrm{BM}}_{\lambda,\theta}(\psi^{-1}(\tilde{Y}))$ 中采样，其中对于 $\tilde{Y} = \psi(Y) \in \mathbb{R}^{2d\times 2r}$，它是通过同构 ψ 得到的 Y 的实值表示。这个新的后验分布还可以进一步写成

$$\tilde{\nu}^{\mathrm{BM}}_{\lambda,\theta}(\tilde{Y}) = \exp\left(-\lambda \sum_{s=1}^{2^n}\sum_{a=1}^{3^n}\left(\hat{p}_{a,s} - \sqrt{2}\mathrm{tr}(\tilde{\mathcal{P}}^a_s \cdot \tilde{Y}\tilde{Y}^\top)\right)^2\right)\tilde{\nu}_\theta(\tilde{Y}),$$

其中，附加因子 $\sqrt{2}$ 由式 (11) 得出，而 $\tilde{\nu}_\theta(\tilde{Y}) := \nu_\theta(\psi^{-1}(\tilde{Y}))$。后验分布的负对数密度为

$$\begin{aligned}\tilde{f}_{\lambda,\theta}(\tilde{Y}) &= -\log(\tilde{\nu}^{\mathrm{BM}}_{\lambda,\theta}(\tilde{Y})) \\ &= \lambda \sum_{s=1}^{2^n}\sum_{a=1}^{3^n}\left(\hat{p}_{a,s} - \sqrt{2}\mathrm{tr}(\tilde{\mathcal{P}}^a_s \cdot \tilde{Y}\tilde{Y}^\top)\right)^2 + \frac{2d+r+2}{4}\log\det\left(\frac{\theta^2}{\sqrt{2}}I_{2d} + \sqrt{2}\tilde{Y}\tilde{Y}^\top\right) + \hat{C},\end{aligned}$$

其中 $C = -\log(C_\theta) + (2d+r+2)d\log(2)/4$，其中 C_θ 是先验 (7) 的归一化常数。该函数的梯度为：

$$\nabla\tilde{f}_{\lambda,\theta}(\tilde{Y}) = -2\sqrt{2}\lambda\sum_{s=1}^{2^n}\sum_{a=1}^{3^n}(\hat{p}_{a,s} - \sqrt{2}\mathrm{tr}(\tilde{\mathcal{P}}^s_a \cdot \tilde{Y}\tilde{Y}^\top))(\tilde{\mathcal{P}}^s_a + \tilde{\mathcal{P}}^{s\top}_a)\tilde{Y} + \frac{2d+r+2}{\theta^2}\left(I_{2d} + \frac{2}{\theta^2}\tilde{Y}\tilde{Y}^\top\right)^{-1}\tilde{Y},$$

其中，行列式的导数是根据 [Lut96, Result 10.3.3 (1)] 计算的。根据 Sherman-Morrison-Woodbury 公式 [Hig08, Sec B.10], ，矩阵逆可写为

$$\left(I_{2d} + \frac{2}{\theta^2}\tilde{Y}\tilde{Y}^\top\right)^{-1} = I_{2d} - \tilde{Y}\left(\frac{\theta^2}{2}I_{2r} + \tilde{Y}^\top\tilde{Y}\right)^{-1}\tilde{Y}^\top,$$

在 $r \ll d$ 情况下，计算该式的代价要低得多（回想一下， $r = 1$ 用于纯态的恢复）。

算法1总结了我们所提出的*Burer-Monteiro* 估计量（简称 BM-估计量）的计算过程。请注意，该算法在采样过程中不会对估计量施加单位迹约束；相反，仅在终止前对迹进行归一化。这一做法

的依据是观察到，密度估计量的迹在采样过程中通常会保持接近1。

算法1 量子层析的 BM估计量

1：输入：秩估计 r，初始点 $Y_0 \in \mathbb{C}^{d\times r}$，超参数 $\eta, \beta, \theta, \lambda, n_{\text{burnin}}$。2：计算 $Y_0 := \psi(Y_0)$ 3：对于 $k \geq 1$，直到满足终止准则 4：按照标准高斯分布随机独立采样 $w_k^R, w_k^I \in \mathbb{R}^{d\times r}$ 的元素，并令 $w_k := w_k^R + iw_k^I$，以及 $\tilde{w}_k := \psi(w_k)$。5：$\tilde{Y}_k = \tilde{Y}_{k-1} - \eta\nabla \tilde{f}_{\lambda,\theta}(\tilde{Y}_{k-1}) + \sqrt{\frac{2\eta}{\beta}}\tilde{w}_k$6：结束循环 7：令 $\hat{\rho} = \frac{1}{k-n_{\text{burnin}}} \sum_{i=1}^{k-n_{\text{burnin}}} \psi^{-1}(\tilde{Y}_{k-i+1})\psi^{-1}(\tilde{Y}_{k-i+1})^*$8：返回：估计量 $\frac{1}{\text{tr}(\hat{\rho})}\hat{\rho}$。

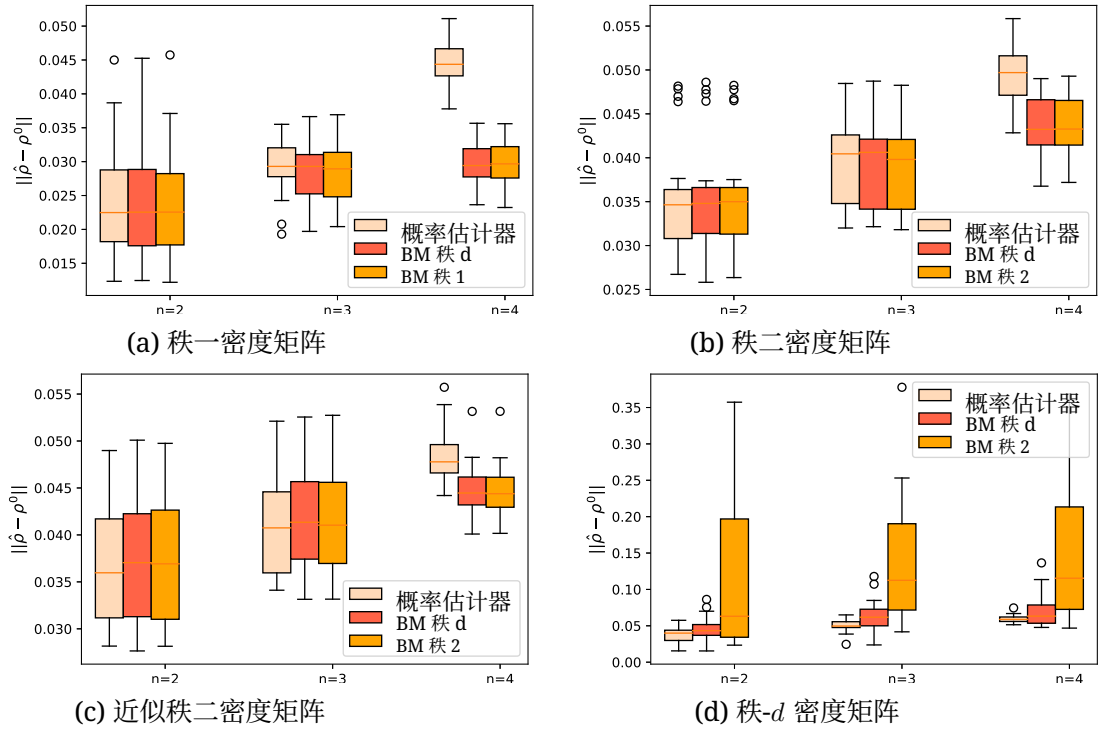
接下来，我们在一组合成数据上，将所提出的朗之万采样算法与 [MA17] 中提出的概率估计器进行比较。

数据生成：与 [MA17], 类似，我们假设目标矩阵 ρ 属于以下形式之一。

1. 某个 $v \in \mathbb{C}^d$ 的纯态密度矩阵 $\rho_{\text{rank 1}}^0 = vv^*$。2. 秩-2密度矩阵 $\rho_{\text{rank 2}}^0 = \frac{1}{2}v_1v_1^* + \frac{1}{2}v_2v_2^*$，其中 v_1，v_2 是 $\mathbb{C}^d$ 中两个归一化正交向量。3. 近似秩-2密度矩阵 $\rho_{\text{approx rank 2}}^0 = w\rho_{\text{rank 2}}^0 + (1-w)I_d/d$，$w = 0.98$，其中 $\rho_{\text{rank 2}}^0$ 按上述方式生成。4. 最大混合态（秩 d）。

我们重点考察完备测量设置，其中实验次数为 4^n；对每个量子比特，这对应于三个泡利可观测量和恒等算子。不同于前文所述的逐量子比特测量设置，我们考虑更一般的设置，即将系统作为整体进行测量：对于实验 $a \in \{1, \ldots, 4^n\}$，只能获得一个标量，而非 n 维向量；后者见我们在 [MA17] 所用设置基础上得到的理论结果。因此，实验结果通过二项随机变量 $b \sim B(m, p)$ 进行模拟，其中 p 是观测到所考虑的（系统）状态的概率，该概率由玻恩规则从理论上确定；m 是该实验的（虚拟）重复次数。

所考察的算法：我们将 [MA17]中介绍的概率估计器与所提出的 BM 估计器的两种变体进行比较。这两种变体旨在分别适用于已知和未知目标密度矩阵秩的情形。对于第一种情形，我们通过运行算法1得到 BM 估计器，其中 $r = 1$、2、2 分别对应秩为 1、秩为 2 和近似秩为 2 的目标密度矩阵。对于第二种情形，为反映目标密度矩阵的秩未知的情况，我们针对每种类型的目标密度都令 $r = d$。除非另有说明，我们采用以下超参数值：$\eta = 10^{-5}$、$\beta = 10^3$、$\lambda = m/2$、$\theta \in \{0.1, 100\}$。当秩未知时，选择最小值以提高先验密度的低秩性；当预期 r 为目标秩时，则选择第二个值，此时先验密度在不同秩上的分布更接近均匀分布。每种算法均运行 10^4 次迭代。按照 [MA17], 中的建议，每种算法都采用预热阶段（默认



(a) 秩一密度矩阵　(b) 秩二密度矩阵

(c) 近似秩二密度矩阵　(d) 秩-d 密度矩阵

图1：比较所考虑的不同估计量的最终精度。

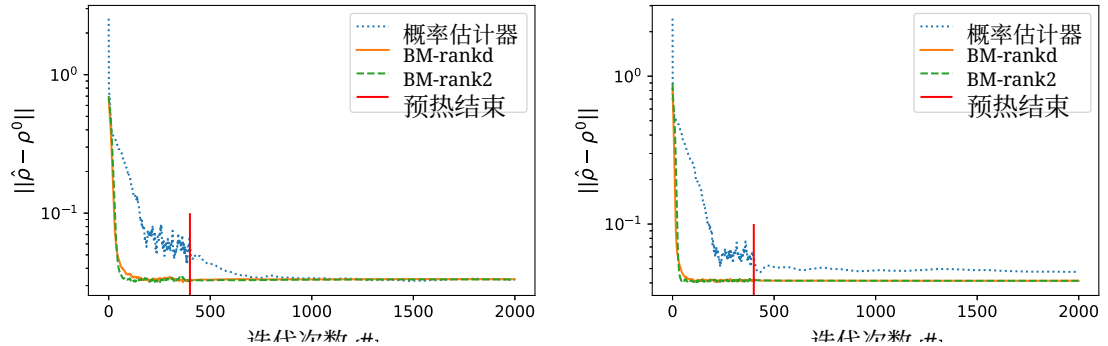
进行 2000 次迭代）。我们将 BM 估计器初始化于某个初始点 $Y_0 = VD^{1/2}$，其中 $V \in \mathbb{C}^{d\times r}$ 是从斯蒂费尔流形上随机均匀采样得到的酉矩阵（服从哈尔分布），而 $D \in \mathbb{R}^{r\times r}$ 是一个对角矩阵，其对角元素服从参数为 $1/r$的狄利克雷分布。

误差度量：我们使用所得估计量 $\hat{\rho}$ 与真实密度矩阵 ρ^0 之间的距离（以弗罗贝尼乌斯范数衡量）。

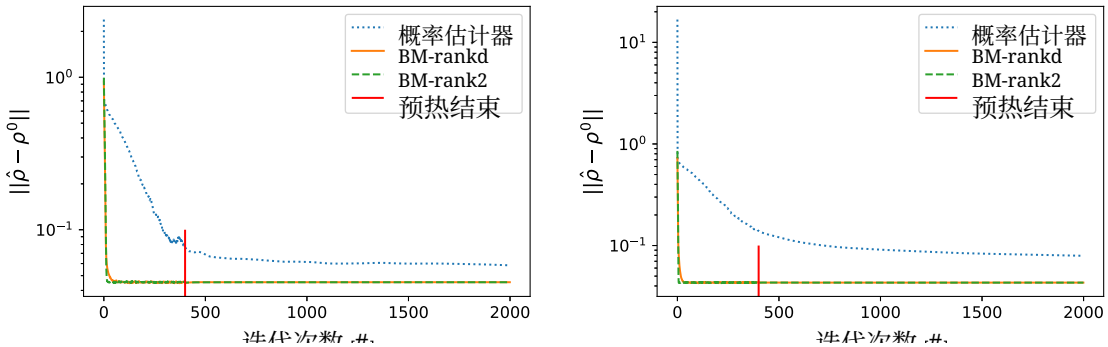
4.2　我们提出的 BM 估计器与概率估计器的比较　or

图1比较了我们提出的 BM 估计量的两种变体与 [MA17], 中推导出的概率估计器，在不同问题维数和目标密度矩阵低秩性下的准确度。结果显示，[MA17] 中提出的概率估计器的最终准确度通常与我们的 BM 估计器相当。对于大型矩阵（$n = 4$），无论目标矩阵的秩已知与否，我们提出的估计器都显著优于概率估计器。我们的 BM 估计器无论是否知道目标密度矩阵的秩，都能达到相同的准确度，这一事实支持了我们提出的先验能够促进低秩性的结论。作为合理性检验，我们还展示了使用 r 远低于目标密度矩阵秩的参数时，BM 估计器所得的结果；如预期，准确度较低。见图1右下图，其中针对满秩目标密度矩阵计算了一个使用 $r = 2$ 的 BM 估计器。

除了最终精度之外，我们还在图2中比较了各算法生成的马尔可夫链的收敛性：在图2中，我们展示了典型算法运行期间各估计量与目标密度矩阵之间距离的变化（我们仅展示每种算法的2000个步骤，此处预热阶段的持续时间设为800）。该图表明，与 [MA17]中提出的概率估计器相关的马尔可夫链相比，与我们的BM估计量相关的马尔可夫链收敛得快得多。如表1所示，计算成本



(a) $n = 2$ (b) $n = 3$



(c) $n = 4$ (d) $n = 5$

图2：不同维数下，各采样算法生成的马尔可夫链的收敛情况，目标密度矩阵的秩为 2。

	概率估计器	BM-rank 2	BM-rank d
$n = 2$	5.7	1.3	1.6
$n = 3$	12	1.7	3
$n = 4$	124.8	37	198.7
$n = 5$	2040	319	3593

表1：每次迭代的计算时间（$\times 10^{-4}$s）

生成马尔可夫链的新样本在不同算法之间也存在很大差异。对于较小的维数（$n = 2$或 $n = 3$），BM 方法比 [MA17]中的概率估计器更快。另一方面，对于更高维的问题，不使用秩信息的BM 方法（即 $r = d$）比概率估计器的成本更高；这可能是因为需要计算先验的对数的梯度，而该计算涉及矩阵逆。对于较小的 r值，通过利用上一节所述的 Sherman-Morrison-Woodbury 公式，可以缓解这一问题。该公式用一个 $r \times r$ 矩阵的逆替代了对单位算子进行秩为r扰动所得的 $d \times d$ 矩阵的逆。使用该公式后，BM估计量的可扩展性在很大程度上得以保证；对于 $n = 5$，生成一个马尔可夫链样本的计算成本约为概率估计器的十分之一。

最后，图 3 展示了在改变每次实验的重复试验次数 m 时，我们的 BM估计量的表现，其中 $n = 3$，目标密度矩阵 ρ^0_{rank2}, $\lambda = m/2$, $\theta = 10^2$。图 3 中的每个点都是算法1在 10 个随机种子下所得结果的平均值。我们从图 3 中可以看到，误差随着采样规模的增大而减小，收敛速率为 $1/m$，这与上一节的理论分析一致；事实上，图 3 中的橙色线

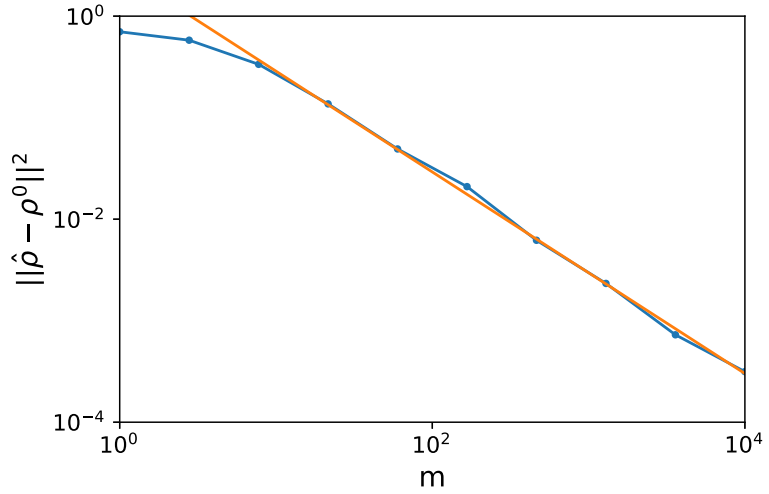


图 3：增加 m（每次实验的重复次数）时，算法1取得的相对误差（蓝色）。橙色曲线表示对曲线上所有点（省略前两个点）进行线性回归的结果。

是对曲线上所有点（前两个点除外）进行线性回归得到的直线，其斜率为 -0.99。

5 结论

我们提出了一种用于量子层析的新估计量，该估计量利用相关对数后验密度的可微性，通过朗之万采样实现高效计算。我们的估计量基于密度矩阵的Burer-Monteiro分解，能够自然地纳入先验秩知识。当目标密度的秩已知时（例如，测量纯态时），与文献中的其他估计量相比，这种分解极大地提升了所提出估计量的可扩展性；鉴于系统中量子比特数量增加会导致问题维数呈指数增长，这一特性尤其具有吸引力。当目标密度矩阵的秩未知、但仅已知其上界时，可以利用该上界降低我们的朗之万采样算法每次迭代的计算负担。同时，我们提出的先验密度是 [Dal20]，中所提出先验在复数情形下的直接扩展，进一步促进了迭代点的低秩性。需要注意的是，由于先验的选择，后验分布并非对数凹，因此我们无法对朗之万采样器的收敛性给出理论保证。然而，我们的数值实验表明，朗之万采样器在实践中表现良好，尤其是在已知密度矩阵的秩较小、对应于低维Burer-Monteiro因子的情况下；与 [MA17] 中提出的方法相比，我们还观察到可扩展性得到了显著提升。在 [Dal20] 中也曾在不同情形下使用过一种类似的先验，该先验会导致后验密度不是对数凹的，数值结果同样表现良好。

致谢

本研究部分由英国政府科学、创新与技术部（DSIT）通过英国国家测量系统（NMS）计划和国家量子技术计划资助。本研究的启动和部分工作

是在 E. Massart 于牛津大学数学研究所担任博士后研究员期间开展的，其经费由英国国家物理实验室提供。本文数值结果所用的部分实现由 Danila Mokeev 在鲁汶大学攻读理学硕士论文期间完成。

参考文献

[ABH+13] Pierre Alquier、Cristina Butucea、Mohamed Hebiri、Katia Meziani 和 Tomoyuki Morimae。量子系统的秩惩罚估计。 *Physical Review A*，88(032113)，2013。[ALL+24] Abhishek Agarwal、Lachlan P Lindoy、Deep Lall、Fran.cois Jamet 和 Ivan Rungger。驱动超导量子比特中的非马尔可夫噪声建模。 *Quantum Sci. Technol.*，9(3):035017，2024。[ALL+25] Abhishek Agarwal、Lachlan P. Lindoy、Deep Lall、Sebastian E. de Graaf、Tobias Lindstr¨om 和 Ivan Rungger。利用最小数据平均和分层离散涨落自动分段，快速追踪并解耦量子比特噪声涨落。 *arXiv:2505.23622*，2025。[BDAK98] V. Bu˘zek、R. Derka、G. Adam 和 P.L. Knight。自旋系统量子态的重构：从量子贝叶斯推断到量子层析。 *Annals of Physics*，266(2):454–496，1998。[BEG+24] Dolev Bluvstein、Simon J. Evered、Alexandra A. Geim、Sophie H. Li、Hengyun Zhou、Tom Manovitz、Sepehr Ebadi、Madelyn Cain、Marcin Kalinowski、Dominik Hangleiter、J. Pablo Bonilla Ataides、Nishad Maskara、Iris Cong、Xun Gao、Pedro Sales Rodriguez、Thomas Karolyshyn、Giulia Semeghini、Michael J. Gullans、Markus Greiner、Vladan Vuleti´c 和 Mikhail D. Lukin。基于可重构原子阵列的逻辑量子处理器。 *Nature*，626(7997):58，2024。[BHK+19] Fr´ed´eric Bouchard、Felix Hufnagel、Dominik Koutn´y、Aazad Abbas、Alicia Sit、Khabat Heshami、Robert Fickler 和 Ebrahim Karimi。高维量子通信信道的量子过程层析。 *Quantum*，3:138，2019。[BK10] R. Blume-Kohout。量子态的最优可靠估计。 *New Journal of Physics*，12 043034 2010 ( )，。[BM03] S. Burer 和 R.D.C. Monteiro。通过低秩分解求解半定规划的非线性规划算法。 *Mathematical Programming*，95(2):329–357，2003。[Cat07] Olivier Catoni。 *Pac-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning* 56 2007 。编号 。

IMS Lecture Notes Monogr，。[Dal20] Arnak S. Dalalyan。多元回归中的指数权重与偏好低秩性的先验。 *Ann. Inst. H. Poincaré Probab. Statist.*，56(2):1465–1483，2020。[DT12] A.S. Dalalyan 和 A.B. Tsybakov。通过聚合和 Langevin Monte Carlo 进行稀疏回归学习。 *Journal of Computer and System Sciences*，78(5):1423–1443，2012。[FKuRS22] Ahmad Farooq、Uman Khalid、Junaid ur Rehman 和 Hyundong Shin。用于量子传感的稳健量子态层析方法。 *Sensors*，22(7):2669，2022。[GCC16] C. Granade、J. Combes 和 D. G. Cory。实用贝叶斯层析。 *New Journal of Physics*，18 133024 259 296，2016 ( )：–。[GKD12] M˘ad˘alin Guţ˘a、Theodore Kypraios 和 Ian Dryden。多离子量子层析的基于秩的模型选择。 *New Journal of Physics*，14(10):105002，2012。[GKKT20] M. Guţ˘a、J. Kahn、R. Kueng 和 J. A. Tropp。具有最优误差界的快速态层析。 *Journal of Physics A: Mathematical and Theoretical*，53(20):204001，2020。[GLF+10] David Gross、Yi-Kai Liu、Steven T. Flammia、Stephen Becker 和 Jens Eisert。基于压缩感知的量子态层析。 *Phys. Rev. Lett.*，105(15)，2010。[Goo24] Google Quantum AI 和合作者。表面码阈值以下的量子纠错。*Nature*，638(8052):920，2024。

[GSG+23] Valentin Gebhart、Raffaele Santagati、Antonio Andrea Gentile、Erik M. Gauger、David Craig、Natalia Ares、Leonardo Banchi、Florian Marquardt、Luca Pezze 和 Cristian Bonato。量子系统学习。 *Nature Reviews Physics*，5(3):141–156，2023。[Gue19] Benjamin Guedj。PAC 贝叶斯学习入门。arXiv:1901.05353，2019。[HH12] F. Huszar 和 N. M. T. Houlsby。自适应贝叶斯量子层析。 *Physical Review A*，85 052120 2012 ( )，。[Hig08] Nicholas J. Higham。 *Functions of Matrices*。SIAM，2008。[HJ13] Roger A. Horn 和 Charles R. Johnson。 *Matrix Analysis (Second Edition)*。剑桥大学出版社，2013。[HˇRFJ04] Zden ˇ ˇ ˇ ˇ ˇ *3 Maximum-Likelihood* ek Hradil、*Methods in Quantum Mechanics* 59 112 2004 Jaroslav Rehacek、Jaromır Fiurasek 和 Miroslav Jezek。，页码 – 。施普林格柏林海德堡，。[Jon91] K.R.W. Jones。量子推断原理。 *Annals of Physics*，207(1):140–170，1991。[KF15] Richard Kueng 和 Christopher Ferrie。近乎最优的量子层析：估计量与界。*New Journal of Physics*，17(12):123013，2015。[KSR+13] K. S. Kravtsov、S. S. Straupe、I. V. Radchenko、N. M. T. Houlsby、F. Husz´ar 和 S. P. Kulik。实验自适应贝叶斯层析。 *Physical Review A*，87(6):062122，2013。[LAZ+25] Deep Lall、Abhishek Agarwal、Weixi Zhang、Lachlan Lindoy、Tobias Lindstr¨om、Stephanie Web- ster、Simon Hall、Nicholas Chancellor、Petros Wallden、Raul Garcia-Patron、Elham Kashefi、Viv Kendon、Jonathan Pritchard、Alessandro Rossi、Animesh Datta、Theodoros Kapourniotis、Kon- stantinos Georgopoulos 和 Ivan Rungger。量子计算机指标与基准综述及汇编：定义、方法与软件，2025。[LLJL20] Joseph M. Lukens、Kody J. H. Law、Ajay Jasra 和 Pavel Lougovski。贝叶斯量子态估计的一种实用高效方法。 *New Journal of Physics*，22(6):063038，2020。[Lut96] H. Lutkepohl。 *Handbook of matrices*。约翰·威立父子公司，1996。[MA17] The Tien Mai 和 Pierre Alquier。秩自适应伪贝叶斯量子层析。*Journal of Statistical Planning and Inference*，184:62–76，2017。[Mai22] The Tien Mai。用于伪贝叶斯量子层析的高效自适应 MCMC 算法，2022。[NC10] Michael A. Nielsen 和 Isaac L. Chuang。 *Quantum Computation and Quantum Information: 10th Anniversary Edition* 2010 。剑桥大学出版社，。[Nea11] Radford M. Neal。利用哈密顿动力学的 MCMC。 *Handbook of Markov Chain Monte Carlo*，2011。[OTKP12] Esa Ollila、David E. Tyler、Visa Koivunen 和 H. Vincent Poor。复椭圆对称分布：综述、新结果及应用。 *IEEE Transactions on Signal Processing*，60 11 5597 5625，2012 ( )：–。[Pre98] John Preskill。物理学 229 课程讲义：量子信息与计算，1998。

[P YBBK25] Timothy Proctor、Kevin Young、Andrew D Baczewski 和 Robin Blume-Kohout。量子计算机基准测试。 *Nature Reviews Physics*，第 1–14 页，2025 年。[QFN21] Yihui Quek、Stanislav Fort 和 Hui Khoon Ng。利用神经网络进行自适应量子态层析。 *npj Quantum Information*，7(1)，2021 年。[SRmG17] Daniel Suess、 -Lukasz Rudnicki、Thiago O maciel 和 David Gross。量子态层析中的误差区域：量子态几何导致的计算复杂性。 *New Journal of Physics* 19 9 093013, 2017 ， ( ) ： 。

[VR89] K. Vogel 和 H. Risken。根据旋转正交相位的概率分布确定准概率分布。 *Physical Review A*，40：2847–2849，1989 年。

6 附录

6.1 引理1的证明

证明。 这个证明是对 [Dal20，引理1] 中给出的证明在复数情形下的直接推广。对于任意有界可测函数 $h\colon \mathbb{C}^d \to \mathbb{R}$，有

$$\begin{aligned}\int_{\mathbb{C}^{d\times r}} h(y_1)\nu_\theta(Y)\mathrm{d}Y &= C_\theta \int_{\mathbb{C}^{d\times r}} \frac{h(y_1)}{\det(\theta^2 I_d + YY^*)^{(2d+r+2)/2}}\mathrm{d}Y \\ &= \tilde{C}_\theta \int_{\mathbb{C}^{d\times r}} \frac{h(\theta z_1)}{\det(I_d + ZZ^*)^{(2d+r+2)/2}}\mathrm{d}Z,\end{aligned}$$

其中， $y_1 \in \mathbb{C}^d$ 是 Y的第一列；对于某个 $\tilde{C}_\theta$，第二个等号是通过变量替换 $Z := Y/\theta$得到的。接着，我们进行第二次变量替换，并将x 记为 Z的第一列，且 $Z = [x, Z_{2:r}] = [x,(I_d + xx^*)^{1/2}\bar{Z}_{2:r}]$。由此可得

$$\begin{aligned}\mathrm{d}Z = \mathrm{d}x\mathrm{d}Z_{2:r} &= \mathrm{d}x\det(I_d + xx^*)^{(r-1)/2}\mathrm{d}\bar{Z}_{2:r} \\ &= (1+\|x\|^2)^{(r-1)/2}\mathrm{d}x\mathrm{d}\bar{Z}_{2:r}\end{aligned}$$

and

$$\begin{aligned}\det(I_d + ZZ^*) &= \det(I_d + xx^* + Z_{2:r}Z_{2:r}^*) \\ &= \det(I_d + xx^* + (I_d + xx^*)^{1/2}\bar{Z}_{2:r}\bar{Z}_{2:r}^*(I_d + xx^*)^{1/2}) \\ &= \det((I_d + xx^*)^{1/2}(I_d + \bar{Z}_{2:r}\bar{Z}_{2:r}^*)(I_d + xx^*)^{1/2}) \\ &= \det(I_d + xx^*)\det(I_d + \bar{Z}_{2:r}\bar{Z}_{2:r}^*) \\ &= (1+\|x\|^2)\det(I_d + \bar{Z}_{2:r}\bar{Z}_{2:r}^*).\end{aligned}$$

接着得到

$$\begin{aligned}\int_{\mathbb{C}^{d\times r}} h(y_1)\nu_\theta(Y)\mathrm{d}d\mathrm{d}Y &= \tilde{C}_\theta \int_{\mathbb{C}^{d\times r}} h(\theta x)(1+\|x\|^2)^{(r-1)/2}[(1+\|x\|^2)\det(I_d + \bar{Z}_{2:r}\bar{Z}_{2:r}^*)]^{-(2d+r+2)/2}\mathrm{d}d\mathrm{d}x\mathrm{d}\bar{Z}_{2:r} \\ &= \tilde{C}_\theta \int_{\mathbb{C}^{d\times r}} h(\theta x)(1+\|x\|^2)^{-(2d+3)/2}\det(I_d + \bar{Z}_{2:r}\bar{Z}_{2:r}^*)^{-(2d+r+2)/2}\mathrm{d}x\mathrm{d}\bar{Z}_{2:r}.\end{aligned}$$

将归一化常数写为

$$\tilde{C}_\theta = \left(\int_{\mathbb{C}^{d\times r}} (1+\|x\|^2)^{-(2d+3)/2}\det(I_d + \bar{Z}_{2:r}\bar{Z}_{2:r}^*)^{-(2d+r+2)/2}\mathrm{d}x\mathrm{d}\bar{Z}_{2:r}\right)^{-1}$$

可得

$$\begin{aligned}\int_{\mathbb{C}^{d\times r}} h(y_1)\nu_\theta(Y)\mathrm{d}Y &= \frac{\int_{\mathbb{C}^d} h(\theta x)(1+\|x\|^2)^{-(2d+3)/2}\mathrm{d}x}{\int_{\mathbb{C}^d}(1+\|x\|^2)^{-(2d+3)/2}\mathrm{d}x} \\ &= \frac{\int_{\mathbb{C}^d} h(\sqrt{2/3}\theta y)(1+2\|y\|^2/3)^{-(2d+3)/2}\mathrm{d}y}{\int_{\mathbb{C}^d}(1+2\|y\|^2/3)^{-(2d+3)/2}\mathrm{d}y},\end{aligned}$$

其中，我们可以从最后一个表达式中看出，这是一个位置参数为 0、尺度矩阵为 $\Sigma = I_d$ [OTKP 12]，的复多元 t_3 分布，其概率密度函数 μ 与 $\mu(y) \propto (1+2\|y\|^2/3)^{-(2d+3)/2}\mathrm{d}y$ 成正比。

由于复多元 t_p 分布的协方差矩阵为 $\frac{p}{p-2}I_d$，因此 —

$$\int_{\mathbb{C}^{d\times r}} \|Y\|_{\mathrm{F}}^2\nu_\theta(Y)\mathrm{d}Y = \int_{\mathbb{C}^{d\times r}} \left(\sum_{i=1}^r \|y_i\|^2\right)\nu_\theta(Y)\mathrm{d}Y = \sum_{i=1}^r \int_{\mathbb{C}^{d\times r}} \|y_i\|^2\nu_\theta(Y)\mathrm{d}Y = \frac{2r}{3}\theta^2\int_{\mathbb{C}^d}\|y\|^2\mu(y)\mathrm{d}y,$$

因此

$$\int_{\mathbb{C}^{d\times r}} \|Y\|_{\mathrm{F}}^2 \nu_\theta(Y)\mathrm{d}Y = 2rd\theta^2.$$

□

6.2 定理 1 的证明

我们将 P^0, $P(\rho)$ 和 $\hat{}P \in [0,1]^{3^n\times 2^n}$ 记为如下定义的矩阵:

$$[P^0]_{a,s} = p^0_{a,s} = \mathrm{tr}(\mathcal{P}^a_s \cdot Y^0 Y^{0*}),$$

对于与目标密度 $\rho^0 = Y^0Y^{0*}$ 相关联的系统，该矩阵表示实验 a 产生结果向量 $o(s)$ 的概率（参见第 2 节引入的记号），

$$[P(YY^*)]_{a,s} = \mathrm{tr}(\mathcal{P}^a_s \cdot YY^*),$$

对于与任意密度密度 $\rho = YY^*$ 相关联的系统，该矩阵表示实验 a 产生结果 $o(s)$ 的概率，并且

$$[\hat{P}]_{a,s} = \hat{p}_{a,s},$$

经验频率矩阵 (2)，令所有 $m_a = m$ 对每个 a 均成立即可得到，即每个实验 $a \in \{1,\dots,3^n\}$ 使用相同数量的重复试验。采用此记号，我们的似然 (6) 可写为

$$L^{\mathrm{BM}} : \mathbb{C}^{d\times r} \to \mathbb{R} : Y \mapsto \|\hat{P} - P(YY^*)\|_{\mathrm{F}}^2,$$

相应的后验分布 (8) 可写为 $\hat{}\nu^{\mathrm{BM}}_{\lambda\,\theta,}(Y) = e^{-f_{\lambda,\ \theta}(Y)}$，根据先验定义 (7),

$$f_{\lambda,\theta}(Y) = \lambda L^{\mathrm{BM}}(Y) + \frac{2d+r+2}{2}\log\det(\theta^2 I_d + YY^*) + C, \tag{12}$$

以及 $C = -\log C_\theta$，其中 C_θ 是谱缩放 Student t分布的归一化常数。我们先回顾以下引理。

引理 3. *[MA17，引理 3]* 对于 $\lambda > 0$，对所有 $Y \in \mathbb{C}^{d\times r}$均有

$$\mathbb{E}\exp\left\{\lambda(\|P(YY^*) - \hat{P}\|_{\mathrm{F}}^2 - \|P^0 - \hat{P}\|_{\mathrm{F}}^2) - \lambda\left(1 + \frac{\lambda}{m}\right)\|P^0 - P(YY^*)\|_{\mathrm{F}}^2\right\} \le 1 \tag{13}$$

$$\mathbb{E}\exp\left\{\lambda\left(1 - \frac{\lambda}{m}\right)\|P^0 - P(YY^*)\|_{\mathrm{F}}^2 - \lambda(\|P(YY^*) - \hat{P}\|_{\mathrm{F}}^2 - \|P^0 - \hat{P}\|_{\mathrm{F}}^2)\right\} \le 1, \tag{14}$$

ˆ其中，期望是对 P中的随机变量 R^a_i 取的，也就是说，是对每个实验进行 m次重复试验所得结果取的期望。

以下引理是对 [MA17，引理 4]的直接推广。

引理 4. 对于满足 $\frac{\lambda}{m} < 1$的 $\lambda > 0$，以概率 $1-\epsilon$， $\epsilon \in]0,\ 1[$, 有

$$\int \|P^0 - P(YY^*)\|_{\mathrm{F}}^2 \hat{\nu}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta}(Y)\mathrm{d}Y \le \inf_{\nu\in\Pi(\mathbb{C}^{d\times r})} \frac{\left(1+\frac{\lambda}{m}\right)\int \|P^0 - P(YY^*)\|_{\mathrm{F}}^2 \nu(Y)\mathrm{d}Y + 2\frac{\log\left(\frac{2}{\epsilon}\right) + D_{\mathrm{KL}}(\nu,\nu_\theta)}{\lambda}}{1-\frac{\lambda}{m}}, \tag{15}$$

其中 $\Pi(\mathbb{C}^{d\times r})$ 是 $\mathbb{C}^{d\times r}$ 上的概率测度集合，而 $D_{\mathrm{KL}}(\nu,\nu_\theta)$ 是任意概率测度 $\nu \in \Pi(\mathbb{C}^{d\times r})$ 与我们的先验密度 ν_θ 之间的 *KL* 散度，并且 $\lambda_{\mathrm{new}} := \frac{\lambda}{2}(1+\frac{\lambda}{m})$。

证明。 由引理 3，可得

$$\int \mathbb{E}\exp\left\{\lambda(\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2-\|P^0-\hat{P}\|_{\mathrm{F}}^2)-\lambda\left(1+\frac{\lambda}{m}\right)\|P^0-P(YY^*)\|_{\mathrm{F}}^2\right\}\nu_\theta(Y)\mathrm{d}Y\le 1,$$
$$\int \mathbb{E}\exp\left\{\lambda\left(1-\frac{\lambda}{m}\right)\|P^0-P(YY^*)\|_{\mathrm{F}}^2-\lambda(\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2-\|P^0-\hat{P}\|_{\mathrm{F}}^2)\right\}\nu_\theta(Y)\mathrm{d}Y\le 1.$$

由富比尼定理可得

$$\mathbb{E}\int \exp\left\{\lambda(\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2-\|P^0-\hat{P}\|_{\mathrm{F}}^2)-\lambda\left(1+\frac{\lambda}{m}\right)\|P^0-P(YY^*)\|_{\mathrm{F}}^2\right\}\nu_\theta(Y)\mathrm{d}Y\le 1$$
$$\mathbb{E}\int \exp\left\{\lambda\left(1-\frac{\lambda}{m}\right)\|P^0-P(YY^*)\|_{\mathrm{F}}^2-\lambda(\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2-\|P^0-\hat{P}\|_{\mathrm{F}}^2)\right\}\nu_\theta(Y)\mathrm{d}Y\le 1.$$

利用 [Cat07、引理 1.1.3], , 对任意 $\epsilon>0$，都有

$$\mathbb{E}\exp\sup_{\nu\in\Pi(\mathbb{C}^{d\times r})}\left\{\lambda\left(\int\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y-\|P^0-\hat{P}\|_{\mathrm{F}}^2\right)-\log\left(\frac{2}{\epsilon}\right)-D_{\mathrm{KL}}(\nu,\nu_\theta)\right.$$
$$\left.-\lambda\left(1+\frac{\lambda}{m}\right)\int\|P^0-P(YY^*)\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y\right\}\le\frac{\epsilon}{2}$$
$$\mathbb{E}\exp\sup_{\nu\in\Pi(\mathbb{C}^{d\times r})}\left\{\lambda\left(1-\frac{\lambda}{m}\right)\int\|P^0-P(YY^*)\|_{\mathrm{F}}^2\nu(Y)dY-\log\left(\frac{2}{\epsilon}\right)-D_{\mathrm{KL}}(\nu,\nu_\theta)\right.$$
$$\left.-\lambda\left(\int\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2\nu(Y)dY-\|P^0-\hat{P}\|_{\mathrm{F}}^2\right)\right\}\le\frac{\epsilon}{2}.$$

现在，利用 $\mathbb{1}_{x\ge 0}(x)\le\exp(x)$，可得

$$\mathbb{P}\left\{\sup_{\nu\in\Pi(\mathbb{C}^{d\times r})}\left[\lambda\left(\int\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y-\|P^0-\hat{P}\|_{\mathrm{F}}^2\right)-\log\left(\frac{2}{\epsilon}\right)-D_{\mathrm{KL}}(\nu,\nu_\theta)\right.\right.$$
$$\left.\left.-\lambda\left(1+\frac{\lambda}{m}\right)\int\|P^0-P(YY^*)\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y\right]\ge 0\right\}\le\frac{\epsilon}{2}$$
$$\mathbb{P}\left\{\sup_{\nu\in\Pi(\mathbb{C}^{d\times r})}\left[\lambda\left(1-\frac{\lambda}{m}\right)\int\|P^0-P(YY^*)\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y-\log\left(\frac{2}{\epsilon}\right)-D_{\mathrm{KL}}(\nu,\nu_\theta)\right.\right.$$
$$\left.\left.-\lambda\left(\int\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y-\|P^0-\hat{P}\|_{\mathrm{F}}^2\right)\right]\ge 0\right\}\le\frac{\epsilon}{2}.$$

我们推得，对于任意密度 $\nu\in\Pi(\mathbb{C}^{d\times r})$,

$$\mathbb{P}\left\{\int\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y\le\|P^0-\hat{P}\|_{\mathrm{F}}^2+\frac{\log\left(\frac{2}{\epsilon}\right)+D_{\mathrm{KL}}(\nu,\nu_\theta)}{\lambda}\right.$$
$$\left.+\left(1+\frac{\lambda}{m}\right)\int\|P^0-P(YY^*)\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y\right\}\ge 1-\frac{\epsilon}{2}$$
$$\mathbb{P}\left\{\int\|P^0-P(YY^*)\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y\le\frac{\int\|P(YY^*)-\hat{P}\|_{\mathrm{F}}^2\nu(Y)\mathrm{d}Y-\|P^0-\hat{P}\|_{\mathrm{F}}^2+\frac{\log\left(\frac{2}{\epsilon}\right)+D_{\mathrm{KL}}(\nu,\nu_\theta)}{\lambda}}{1-\frac{\lambda}{m}}\right\}\ge 1-\frac{\epsilon}{2}.$$

下一个结果通过使用并集论证得到（即，将第二个不等式右侧的第一项替换为第一个不等式给出的上界）：对于

任意密度 $\nu \in \Pi(\mathbb{C}^{d\times r})$，在数据上的概率至少为 $1-\epsilon$

$$\int \|P^0 - P(YY^*)\|_{\mathrm{F}}^2 \nu(Y)\mathrm{d}Y \le \frac{\left(1+\frac{\lambda}{m}\right)\int \|P^0 - P(YY^*)\|_{\mathrm{F}}^2 \nu(Y)\mathrm{d}Y + 2\frac{\log\left(\frac{2}{\epsilon}\right)+D_{\mathrm{KL}}(\nu,\nu_\theta)}{\lambda}}{1-\frac{\lambda}{m}}.$$

将这个表达式改写为对应于右侧下确界的密度，可得：

$$\int \|P^0 - P(YY^*)\|_{\mathrm{F}}^2 \hat{\nu}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta}(Y)\mathrm{d}Y \le \inf_{\nu\in\Pi(\mathbb{C}^{d\times r})} \frac{\left(1+\frac{\lambda}{m}\right)\int \|P^0 - P(YY^*)\|_{\mathrm{F}}^2 \nu(Y)\mathrm{d}Y + 2\frac{\log\left(\frac{2}{\epsilon}\right)+D_{\mathrm{KL}}(\nu,\nu_\theta)}{\lambda}}{1-\frac{\lambda}{m}},$$

左侧的后验密度 $\hat{\nu}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta}$ 对于 $\lambda_{\mathrm{new}} := \frac{\lambda}{2}(1+\frac{\lambda}{m})$，是由[Gue19，式 10]得到的。

□

接下来，我们需要给出式 (15) 右侧的上界。为此，我们借鉴 [Dal20],的做法，将下确界限制在一族先验分布上；这些先验分布由我们的先验 ν_θ平移得到，记作 $\bar{\nu}_\theta(Y) = \nu_\theta(Y-\bar{Y})$，其中 $\bar{Y}\in\mathbb{C}^{d\times r}$为某个参数。我们推导出以下结果，这是对 [Dal20，引理 2] 在复数情形下的直接推广。

引理 5. 对于任意秩为 p 的矩阵 $\bar{Y}\in\mathbb{C}^{d\times r}$，令 $\bar{\nu}_\theta$ 为将先验分布 ν_θ 平移后得到的概率密度函数，即 $\bar{\nu}_\theta(Y) = \nu_\theta(Y-Y)$。则有

$$D_{\mathrm{KL}}(\bar{\nu}_\theta|\nu_\theta) \le 2p(2d+r+2)\log\left(1+\frac{\|\bar{Y}\|_2}{\theta}\right).$$

证明。该结果是 [Dal20，引理 2]的直接推广。为完整起见，我们在此回顾从 [Dal20], 中摘录的证明。根据 KL 散度的定义，有

$$\begin{aligned} D_{\mathrm{KL}}(\bar{\nu}_\theta|\nu_\theta) &= \int_{\mathbb{C}^{d\times r}} \log\left(\frac{\nu_\theta(Y)}{\bar{\nu}_\theta(Y)}\right)\nu_\theta(Y)\mathrm{d}Y \\ &= \int_{\mathbb{C}^{d\times r}} \log\left(\frac{\nu_\theta(Y)}{\nu_\theta(Y-\bar{Y})}\right)\nu_\theta(Y)\mathrm{d}Y. \end{aligned}$$

首先注意，若 $p=0$，结论显然成立。因此从现在起，我们假设 $p>0$。我们定义 $A := (\theta^2 I_d + YY^*)^{-1/2}$，以及 $B := \theta^2 I_d + (Y-\bar{Y})(Y-\bar{Y})^*$，它们都是厄米正定矩阵[6]。由此可得

$$\begin{aligned} 2\log\left(\frac{\nu_\theta(Y)}{\nu_\theta(Y-\bar{Y})}\right) &= 2\log\left(\frac{\det(\theta^2 I_d + YY^*)^{-(2d+r+2)/2}}{\det(\theta^2 I_d + (Y-\bar{Y})(Y-\bar{Y})^*)^{-(2d+r+2)/2}}\right) && (16)\\ &= (2d+r+2)\log\left(\frac{\det(B)}{\det(A^{-2})}\right) && (17)\\ &= (2d+r+2)\log\left(\det(ABA)\right) && (18)\\ &= (2d+r+2)\sum_{i=1}^{d}\log\lambda_i, && (19) \end{aligned}$$

[6]对于厄米正定矩阵 M，我们分别用 $\log(M)$ 和 $M^{1/2}$ 表示主矩阵对数和主矩阵平方根。设 $M = U\Lambda U^*$ 为特征值分解，其中 $U\in\mathbb{C}^{d\times d}$ 为酉矩阵，$\Lambda\in\mathbb{R}^{d\times d}$ 为对角矩阵且对角元素严格为正。则 $\log(M) = U\log(\Lambda)U^*$ 以及 $M^{1/2} = U\Lambda^{1/2}U^*$，其中对数和平方根作用于 Λ 的对角元素。

其中，我们用 λ_i 表示矩阵 ABA 的第 i 大特征值，其对应的特征向量为$u_i \in \mathbb{C}^d$。接下来我们写作

$$\begin{aligned} ABA &= I_d + A\bar{Y}\bar{Y}^*A - A\bar{Y}Y^*A - AY\bar{Y}^*A, \\ &= I_d + A\bar{Y}(\bar{Y}-Y)^*A - AY\bar{Y}^*A, \end{aligned}$$

并证明 $ABA - I_d$ 的秩至多为 $2p$ 与 d中的较小者，简写为 $\min(2p, d)$。事实上，首先注意到，$\mathrm{range}(A\bar{Y}(\bar{Y}-Y)^*A) \subseteq A\,\mathrm{range}(\bar{Y})$[7]，这意味着 $\mathrm{rank}(AY(Y-Y)^*A)$ 的维数至多为 p。然后，注意到 $\mathrm{rank}(AYY^*A) = \mathrm{rank}(AYY^*A) \leq p$，同样使用上述论证即可。由此可知，$ABA - I_d$ 的秩至多为 $\min(2p, d)$，矩阵 ABA 至多有 $\min(2p, d)$ 个不等于 1 的特征值，因此式 (19) 中的求和至多包含 $\min(2p, d)$ 个非零项。不妨设 $ABA - I_d$ 的非零特征值为 λ_1、...、λ_K，其中 $K \leq \min(2p, d)$，如前所述。对于所有 $i = 1$、...、K，有

$$\begin{aligned} \lambda_i &= u_i^*(ABA)u_i \\ &= u_i^*\left(I_d + A\bar{Y}(\bar{Y}-Y)^*A - AY\bar{Y}^*A\right)u_i \\ &= 1 + u_i^*\left(A\bar{Y}(\bar{Y}-Y)^*A - AY\bar{Y}^*A\right)u_i. \end{aligned}$$

由于 $\|(\bar{Y}-Y)^*Au_i\|^2 = u_i^*A(\bar{Y}-Y)(\bar{Y}-Y)^*Au_i = u_i^*A\left(\bar{Y}\bar{Y}^* - \bar{Y}Y^* - Y\bar{Y}^* + YY^*\right)Au_i = u_i^*\left(A\bar{Y}(\bar{Y}-Y)^*A - AY\bar{Y}^*A\right)u_i + \|Y^*Au_i\|^2$，因此可得：

$$\begin{aligned} \lambda_i &= 1 + \|(\bar{Y}-Y)^*Au_i\|^2 - \|Y^*Au_i\|^2 \\ &\leq 1 + (\|\bar{Y}^*Au_i\| + \|Y^*Au_i\|)^2 - \|Y^*Au_i\|^2 \\ &= 1 + \|\bar{Y}^*Au_i\|^2 + \|Y^*Au_i\|^2 + 2\|\bar{Y}^*Au_i\|\|Y^*Au_i\| - \|Y^*Au_i\|^2 \\ &= 1 + \|\bar{Y}^*Au_i\|^2 + 2\|\bar{Y}^*Au_i\|\|Y^*Au_i\| \\ &\leq (1 + \|\bar{Y}^*Au_i\|)^2, \end{aligned}$$

其中，在最后一个不等式中，我们用到了这一事实：$\|Y^*Au_i\| \leq 1$。事实上，请注意 $1 = u_i^*u_i = u_i^*(\theta^2 I_d + YY^*)^{-1/2}(\theta^2 I_d + YY^*)(\theta^2 I_d + YY^*)^{-1/2}u_i = \theta^2\|Au_i\|^2 + \|Y^*Au_i\|^2 \geq \|Y^*Au_i\|^2$. 利用函数 $\log(1 + x^{1/2})$ 在 $(0, \infty)$ 上的凹性，由 (19) 可得

$$\begin{aligned} 2\log\left(\frac{\nu_\theta(Y)}{\nu_\theta(Y-\bar{Y})}\right) &= (2d + r + 2)\sum_{i=1}^{K}\log\lambda_i \\ &\leq 2(2d + r + 2)\sum_{i=1}^{K}\log(1 + \|\bar{Y}^*Au_i\|) \\ &= 2(2d + r + 2)K\sum_{i=1}^{K}\frac{1}{K}\log(1 + (\|\bar{Y}^*Au_i\|^2)^{1/2}) \\ &\leq 2K(2d + r + 2)\log\left(1 + \left(\frac{1}{K}\sum_{i=1}^{K}\|\bar{Y}^*Au_i\|^2\right)^{1/2}\right). \end{aligned}$$

最后，利用特征向量 $u_1, \ldots u_K$ 是正交归一的这一事实，以及 $A \preceq \theta^{-1}I_d$[8]，并记 $U = [u_1, \ldots, u_K]$，可得：

$$\left(\sum_{i=1}^{K}\|\bar{Y}^*Au_i\|^2\right)^{1/2} = \|\bar{Y}^*AU\|_{\mathrm{F}} \leq \sqrt{K}\|\bar{Y}^*AU\|_2 \leq \sqrt{K}\|\bar{Y}\|_2\|AU\|_2 \leq \frac{\sqrt{K}}{\theta}\|\bar{Y}\|_2.$$

[7] $A\mathcal{S}$ 表示在映射 A下 $\mathcal{S}$ 的像。[8]注意，A 的特征值为 $(\theta^2 + \sigma_i(Y)^2)^{-1/2}$ $i=1, \ldots, r$ 其中 $\sigma_i(Y)$ 是 Y 的第 i个奇异值。由于对所有 $\sigma_i(Y) > 0$，均有 $(\theta^2 + \sigma_i(Y)^2)^{-1/2} \leq \theta$，因此可得 $A \preceq \theta^{-1}I_d$。

第一个不等式由 [Lut96，结果 8.5.2(4e)]，得出；第二个不等式源于谱范数的次乘性；最后一个不等式则由谱范数的酉不变性以及矩阵不等式 $A \preceq \theta^{-1} I_d$ 得出。因此，由于 $K = \min(2p, d)$，可得

$$\log\left(\frac{\nu_\theta(Y)}{\nu_\theta(Y-\bar{Y})}\right) \leq 2p(2d+r+2)\log\left(1+\frac{\|\bar{Y}\|_2}{\theta}\right).$$

□

引理 6. 对于任意秩为 p 的 $\bar{Y} \in \mathbb{C}^{d\times r}$，令 $\bar{\nu}_\theta$ 为平移先验 $\bar{\nu}_\theta(Y) = \nu_\theta(Y-\bar{Y})$。则有：

$$\int \|P(YY^*) - P^0\|_{\mathrm{F}}^2 \bar{\nu}_\theta(Y)\mathrm{d}Y \leq d6^{n/2}(a + \sqrt{2dr}\theta b + 2dr\theta^2),$$

其中 $a := \|Y^0 - \bar{Y}\|_{\mathrm{F}}\|Y^0 + \bar{Y}\|_{\mathrm{F}}$，且 $b := \|Y^0 - \bar{Y}\|_{\mathrm{F}} + \|Y^0 + \bar{Y}\|_{\mathrm{F}}$。

证明。回顾一下，对于所有 $Y \in \mathbb{C}^{d\times r}$，矩阵 $P(YY^*)$ 和 P^0 的元素都是概率，因此位于区间 $[0,1]$ 内。因此，矩阵 $M := P(YY^*) - P^0$ 满足 $M_{i,\,j} \in [-1,1]$，于是 $\|M\|_{\mathrm{F}}^2 = \sum_{i,\,j=1}^{d} M_{i,\,j}^2 \leq d^2$，这意味着 $\|M\|_{\mathrm{F}} \leq d$。由此可得，对于所有 $Y \in \mathbb{C}^{d\times r}$，

$$\|P(YY^*) - P^0\|_{\mathrm{F}}^2 \leq d\|P(YY^*) - P^0\|_{\mathrm{F}}. \tag{20}$$

根据 [ABH+13，引理 5], 的证明表明，对所有 $Y \in \mathbb{C}^{d\times r}$

$$\|P(YY^* - Y^0Y^{0*})\|_{\mathrm{F}} \leq 6^{n/2}\|YY^* - Y^0Y^{0*}\|_{\mathrm{F}}. \tag{21}$$

接着，写作

$$YY^* - Y^0Y^{0*} = \frac{1}{2}\left((Y-Y^0)(Y+Y^0)^* + (Y+Y^0)(Y-Y^0)^*\right),$$

并注意到，根据三角不等式以及弗罗贝尼乌斯范数的次乘性 [HJ13, p.342]，有：

$$\|YY^* - Y^0Y^{0*}\|_{\mathrm{F}} \leq \|Y-Y^0\|_{\mathrm{F}}\|Y+Y^0\|_{\mathrm{F}}. \tag{22}$$

再次利用三角不等式，

$$\begin{aligned}\|Y^0 - Y\|_{\mathrm{F}} &\leq \|Y^0 - \bar{Y}\|_{\mathrm{F}} + \|\bar{Y} - Y\|_{\mathrm{F}} \\ \|Y^0 + Y\|_{\mathrm{F}} &\leq \|Y^0 + \bar{Y}\|_{\mathrm{F}} + \|Y - \bar{Y}\|_{\mathrm{F}}.\end{aligned} \tag{23}$$

结合 (20)、(21)、(22) 和 (23) 得到：

$$\begin{aligned}\|P(YY^*) - P^0\|_{\mathrm{F}}^2 &\leq 6^{n/2}d\left(\|Y^0 - \bar{Y}\|_{\mathrm{F}} + \|\bar{Y} - Y\|_{\mathrm{F}}\right)\left(\|Y^0 + \bar{Y}\|_{\mathrm{F}} + \|Y - \bar{Y}\|_{\mathrm{F}}\right) \\ &\leq 6^{n/2}d\left(\|Y^0 - \bar{Y}\|_{\mathrm{F}}\|Y^0 + \bar{Y}\|_{\mathrm{F}} + \|Y - \bar{Y}\|_{\mathrm{F}}(\|Y^0 - \bar{Y}\|_{\mathrm{F}} + \|Y^0 + \bar{Y}\|_{\mathrm{F}}) + \|Y - \bar{Y}\|_{\mathrm{F}}^2\right),\end{aligned}$$

因此

$$\begin{aligned}\int \|P(YY^*) - P^0\|_{\mathrm{F}}^2 \nu_\theta(Y-\bar{Y})\mathrm{d}Y \leq& 6^{n/2}d(\|Y^0 - \bar{Y}\|_{\mathrm{F}}\|Y^0 + \bar{Y}\|_{\mathrm{F}} \\ &+ (\|Y^0 - \bar{Y}\|_{\mathrm{F}} + \|Y^0 + \bar{Y}\|_{\mathrm{F}})\left(\int \|Y - \bar{Y}\|_{\mathrm{F}}\nu_\theta(Y-\bar{Y})\mathrm{d}Y\right) \\ &+ \int \|Y - \bar{Y}\|_{\mathrm{F}}^2 \nu_\theta(Y-\bar{Y})\mathrm{d}Y).\end{aligned}$$

我们计算

$$\int \|Y\|_{\mathrm{F}}\nu_\theta(Y)\mathrm{d}Y = \left(\left(\int \|Y\|_{\mathrm{F}}\nu_\theta(Y)\mathrm{d}Y\right)^2\right)^{1/2} \leq \left(\int \|Y\|_{\mathrm{F}}^2\nu_\theta(Y)\mathrm{d}Y\right)^{1/2} = \sqrt{2rd}\theta,$$

其中，不等式为詹森不等式（该不等式指出，对于任意凸函数 ϕ，都有$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$，这里的凸函数为 $x \mapsto x^2$），而最后一个等号则由引理1得出。同理，

$$\int \|Y\|_{\mathrm{F}}^2 \nu_\theta(Y)\mathrm{d}Y = 2dr\theta^2.$$

该结论得证。 □

由此我们可以证明定理1，下面回顾该定理。

定理1： 令 $\lambda = 3m/8$，并令 $\bar{Y} \in \mathbb{C}^{d\times r}$ 满足 $\|Y^0 - \bar{Y}\|_{\mathrm{F}}\|Y^0 + \bar{Y}\|_{\mathrm{F}} \leq 3^{-3n/2}2^{-n/2}/m$。 ¯令 p 表示 Y的秩。对于任意 $\epsilon \in]0,\ 1[$, 有如下结论以概率 $1-\epsilon$

$$\begin{aligned}\|\hat{\rho}^{\mathrm{BM}}_{\lambda,\theta} - \rho^0\|_{\mathrm{F}}^2 \leq \frac{3}{N_{\mathrm{tot}}}\left(3^{3n/4}2^{(n+6)/4}(r+\sqrt{r}\|\bar{Y}\|_{\mathrm{F}}) + \frac{2r}{m} + 1\right)\\ + \frac{3^n 8}{2^n N_{\mathrm{tot}}}\left(\log\left(\frac{2}{\epsilon}\right) + 2p(2^{n+1}+r+2)\log\left(1+\frac{\|\bar{Y}\|_2}{\theta}\right)\right).\end{aligned}$$

证明。 首先注意到，根据詹森不等式，有:

$$\|\hat{\rho}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta} - \rho^0\|_{\mathrm{F}}^2 \leq \int \|YY^* - Y^0Y^{0*}\|_{\mathrm{F}}^2 \hat{\nu}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta}(Y)\mathrm{d}Y \leq \frac{1}{2^n}\int \|P(YY^*) - P^0\|_{\mathrm{F}}^2 \hat{\nu}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta}(Y)\mathrm{d}Y,$$

其中 $\lambda_{\mathrm{new}} = \frac{\lambda}{2}(1+\frac{\lambda}{m})$（参见引理 4 的证明），且第二个不等式来自[MA17，引理 5] 的证明。引理 4 和引理 6 保证，以概率 $1-\epsilon$ 对于$\epsilon \in]0,\ 1[$,

$$\begin{aligned}\|\hat{\rho}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta} - \rho^0\|_{\mathrm{F}}^2 &\leq \inf_{\bar{Y}\in\mathbb{C}^{d\times r}} \frac{6^{n/2}d\left(1+\frac{\lambda}{m}\right)(a+\sqrt{2dr}\theta b + 2dr\theta^2) + 2\frac{\log\left(\frac{2}{\epsilon}\right)+2p(2d+r+2)\log\left(1+\frac{\|\bar{Y}\|_2}{\theta}\right)}{\lambda}}{(1-\frac{\lambda}{m})2^n}\\ &\leq \inf_{\bar{Y}\in\mathbb{C}^{d\times r}} N_{\mathrm{tot}}\frac{6^{n/2}d\left(1+\frac{\lambda}{m}\right)(a+\sqrt{2dr}\theta b + 2dr\theta^2) + 2\frac{\log\left(\frac{2}{\epsilon}\right)+2p(2d+r+2)\log\left(1+\frac{\|\bar{Y}\|_2}{\theta}\right)}{\lambda}}{(1-\frac{\lambda}{m})N_{\mathrm{tot}}2^n}.\end{aligned}$$

让我们回顾ll 那么 $d = 2^n$ 和 $N_{\mathrm{tot}} = 3^n m$。 选择 $\theta \propto 6^{-3n/4}/m$， 以概率 $1-\epsilon$

$$\begin{aligned}\|\hat{\rho}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta} - \rho^0\|_{\mathrm{F}}^2 \leq \inf_{\bar{Y}\in\mathbb{C}^{d\times r}} \frac{m\left(1+\frac{\lambda}{m}\right)}{N_{\mathrm{tot}}\left(1-\frac{\lambda}{m}\right)}\left(a3^{3n/2}2^{n/2} + \frac{\sqrt{2r}}{m}b3^{3n/4}2^{n/4} + \frac{2r}{m^2}\right)\\ + \frac{2m3^n}{2^n\lambda(1-\frac{\lambda}{m})N_{\mathrm{tot}}}\left(\log\left(\frac{2}{\epsilon}\right) + 2p(2^{n+1}+r+2)\log\left(1+\frac{\|\bar{Y}\|_2}{\theta}\right)\right).\end{aligned}$$

回顾一下 $a = \|Y^0 - \bar{Y}\|_{\mathrm{F}}\|Y^0 + \bar{Y}\|_{\mathrm{F}}$，并注意到

$$b = \|Y^0 - \bar{Y}\|_{\mathrm{F}} + \|Y^0 + \bar{Y}\|_{\mathrm{F}} \leq 2(\|Y^0\|_{\mathrm{F}} + \|\bar{Y}\|_{\mathrm{F}}) \leq (2\sqrt{r} + 2\|\bar{Y}\|_{\mathrm{F}}).$$

选择 $\bar{Y}$，使得 $a = 3^{-3n/2}2^{-n/2}/m$ （请注意，通过选择足够接近 Y^0 的 $\bar{Y}$，可以使 a 任意小），由此可知，以概率 $1-\epsilon$,

$$\begin{aligned}\|\hat{\rho}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta} - \rho^0\|_{\mathrm{F}}^2 \leq \frac{m\left(1+\frac{\lambda}{m}\right)}{N_{\mathrm{tot}}\left(1-\frac{\lambda}{m}\right)}\left(\frac{1}{m} + \frac{\sqrt{2r}}{m}(2\sqrt{r}+2\|\bar{Y}\|_{\mathrm{F}})3^{3n/4}2^{n/4} + \frac{2r}{m^2}\right)\\ + \frac{2m3^n}{2^n\lambda(1-\frac{\lambda}{m})N_{\mathrm{tot}}}\left(\log\left(\frac{2}{\epsilon}\right) + 2p(2^{n+1}+r+2)\log\left(1+\frac{\|\bar{Y}\|_2}{\theta}\right)\right).\end{aligned}$$

对于较大的 n 值，右侧由第二项主导（请注意，由于 N_{tot} 中的指数 n 依赖关系，右侧第一项在 n 较大时可以忽略不计），该式在 $\lambda = m/2$ 处取得最小值（对应于 $\lambda_{\mathrm{new}} = 3m/8$），从而得到

$$\|\hat{\rho}^{\mathrm{BM}}_{\lambda_{\mathrm{new}},\theta} - \rho^0\|_{\mathrm{F}}^2 \leq \frac{3}{N_{\mathrm{tot}}}\left(3^{3n/4}2^{(n+6)/4}(r+\sqrt{r}\|\bar{Y}\|_{\mathrm{F}}) + \frac{2r}{m} + 1\right) + \frac{3^n 8}{2^n N_{\mathrm{tot}}}\left(\log\left(\frac{2}{\epsilon}\right) + 2p(2^{n+1}+r+2)\log\left(1+\frac{\|\bar{Y}\|_2}{\theta}\right)\right)$$

概率为 $1-\epsilon$。结果得证。 □

6.3 引理 2 的证明

引理 2： 对于所有 M，$N \in \mathbb{C}^{d\times d}$厄米半正定矩阵，均有 $\mathrm{tr}(MN) = \mathrm{tr}(\psi(M)\psi(N))$，且 $\det(M) = \sqrt{2^d}\det(\psi(M))$。

证明。首先注意到，由于 M 是厄米半正定的，$M^* = M^{R\top} - iM^{I\top} = M$，这意味着 $M^R = M^{R\top}$ 和 $M^I = -M^{I\top}$。换言之，M^R（以及 N^R）是对称的，而 M^I（以及 N^I）是反对称的。我们计算

$$\mathrm{tr}(\psi(M)\psi(N)) = \frac{1}{2}(\mathrm{tr}(M^R N^R - M^I N^I) + \mathrm{tr}(-M^I N^I + M^R N^R)) = \mathrm{tr}(M^R N^R) - \mathrm{tr}(M^I N^I).$$

另一方面，

$$\begin{aligned}\mathrm{tr}(MN) &= \mathrm{tr}((M^R + iM^I)(N^R + iN^I) = \mathrm{tr}(M^R N^R) - \mathrm{tr}(M^I N^I) + i\mathrm{tr}(M^R N^I + M^I N^R)\\ &= \mathrm{tr}(M^R N^R) - \mathrm{tr}(M^I N^I),\end{aligned}$$

其中，第二个等号是因为对称矩阵与反对称矩阵的内积为零；这证明了第一个结论。

对于第二个断言，注意到厄米矩阵可通过酉变换对角化，且其行列式等于其特征值的乘积。我们证明，对于 M 的任意特征值 λ（重数为 m_λ），$\psi(M)$ 存在一个特征值 $\lambda/\sqrt{2}$（重数为 $2m_\lambda$）。由于 M 是厄米矩阵，它有一组由特征向量组成的正交基。设 $v \in \mathbb{C}^d$ 是 M 的一个特征向量，对应特征值 λ，即 $Mv = (M^R v^R - M^I v^I) + i(M^R v^I + M^I v^R) = \lambda(v^R + iv^I)$。于是，

$$\frac{1}{\sqrt{2}}\begin{pmatrix} M^R & -M^I \\ M^I & M^R \end{pmatrix}\begin{pmatrix} v^R \\ v^I \end{pmatrix} = \frac{\lambda}{\sqrt{2}}\begin{pmatrix} v^R \\ v^I \end{pmatrix},$$

and

$$\frac{1}{\sqrt{2}}\begin{pmatrix} M^R & -M^I \\ M^I & M^R \end{pmatrix}\begin{pmatrix} v^I \\ -v^R \end{pmatrix} = \frac{\lambda}{\sqrt{2}}\begin{pmatrix} v^I \\ -v^R \end{pmatrix}.$$

因此，两个向量 $(v^R,\ v^I)$ 和 $(v^I,\ -v^R)$ 是 $\psi(M)$ 的特征向量，对应特征值 $\lambda/\sqrt{2}$。由于这两个向量正交，结合行列式等于特征值乘积的定义（并计入重数），结论便显然成立。

□