arXiv:2603.04922v1 [math.NA] 5 Mar 2026

用于量子态层析的量子相对熵正则化*

Florian Oberender 和 Thorsten Hohage[1]

[1]格奥尔格-奥古斯特-哥廷根大学数值与应用数学研究所

2026年3月6日

摘要

密度矩阵是迹为 1 的半正定算子，用于表征量子系统的状态。我们研究如何从间接测量中重建此类密度矩阵这一反问题，这也称为量子态层析。为求解高维或无限维情形下的此类反问题，我们研究以量子相对熵作为惩罚泛函的变分正则化。量子相对熵是著名的最大熵泛函的一种类似形式，其中函数的复合由谱函数演算取代。本文的主要目标是确立该方案的正则化性质。作为一项关键的中间步骤，我们确立了惩罚泛函关于弱-∗-拓扑的下半紧性。此外，我们还计算了有限维空间上量子相对熵的次梯度、近端算子和共轭泛函。这使我们能够应用凸优化中的迭代算法，以数值方式求解正则化问题。为展示我们结果的有效性和实际价值，我们将理论应用于光子诱导近场电子显微术（PINEM）和光学同相层析的实例。

数学主题分类： 65J20, 81P18,81P17, 65K10

关键词： 正则化理论、量子态层析、量子相对熵、近端算子

1 引言

量子态层析旨在通过一系列不同的测量重构量子系统的密度矩阵，从而确定其状态。这是一种成熟的量子光学技术，可利用异频或零差测量，并通过不同的重构技术获得密度矩阵 [19, 8]。近来，研究人员利用光子诱导近场电子显微术（PINEM）实现了电子的量子态层析[27]。

密度矩阵是底层复可分希尔伯特空间 $\mathcal{H}$ 上迹为 1 的半正定算子 $\rho\colon \mathcal{H} \to \mathcal{H}$ 。我们将迹类半正定算子 $\rho\colon \mathcal{H} \to \mathcal{H}$ 的集合记为 $\tilde{\mathcal{B}}_1^+$。只有形如 $(e_l,\ \rho e_l)$ 的量可以直接测量。这里，$e_l \in \mathcal{H}$ 是与可观测量对应的（自伴）算子的特征向量。这些量可视为与特征基对应的 ρ 的矩阵表示中的对角元。非对角元素只能通过测量 ρ 的演化间接观测

*感谢德国研究联合会（DFG）通过 CRC 1456 项目 432680300 提供的支持。 lly ack（承接上文）

来观测。密度矩阵的时间演化由冯·诺依曼-刘维尔方程描述，该方程以初值 ρ 为条件的解具有如下形式： $U_\theta \rho U_\theta^*$，其中 U_θ： $\mathcal{H} \to \mathcal{H}$ 是某个酉算子。我们假设 U_θ 由某个相互作用参数 θ 参数化，该参数取值于有界子集 $I \subset$R 中。这将得到如下形式的前向算子

$$T : \tilde{\mathcal{B}}_1^+ \to L^2(I, \ell^p(\mathbb{Z})), \qquad (T\rho)(\theta, l) = \langle e_l, U_\theta \rho U_\theta^* e_l \rangle \tag{1}$$

有时使用形式略有不同的算子 U_θ 可能会有所帮助。从测量数据 $T\rho$ 重建密度矩阵 ρ 的反问题通常是不适定的[35, 1]。此外，估计量还必须是物理密度矩阵，即半正定且迹为 1，这也带来了额外挑战。虽然计算给定估计量在该集合上的度量投影很容易，但这种做法可能无法充分利用半正定性和迹约束所带来的强稳定作用。对于许多问题实例，这些约束的正则化效果已经足够，因此在实践中常常成功采用满足这些约束、且不使用任何额外正则化（早停除外）的迭代方案 [23]。 [27] 中提出的 SQUIRRELS算法采用希尔伯特-施密特范数（或 Frobenius 范数）进行二次正则化，并结合半正定规划。

我们提出使用量子相对熵作为一种在物理意义上更合理的惩罚泛函，以取代主要出于算法便利性而选择的希尔伯特-施密特范数。将密度矩阵视为概率分布的非交换版本，量子相对熵可以理解为库尔贝克-莱布勒散度的对应形式。对于算子 ρ， ρ_0： $\mathcal{H} \to \mathcal{H}$ 其定义为

$$\mathrm{QKL}(\rho, \rho_0) := \begin{cases} \infty \quad \text{if } \rho \notin \tilde{\mathcal{B}}_1^+ \text{ or } \rho_0 \notin \tilde{\mathcal{B}}_1^+ \text{ or } \ker \rho_0 \not\subseteq \ker \rho \\ \mathrm{tr}(\rho_0 - \rho + \rho \ln \rho - \rho \ln \rho_0) \quad \text{else,} \end{cases} \tag{2}$$

见 [21]。这里， $\rho \ln \rho - \rho \ln \rho_0$在 ρ的核上定义为零，其中 ker $\rho_0 \subseteq$ker ρ。有些作者在此定义中省略线性项。更多细节请参见第 2.1 节。

本文将考虑如下形式的广义 Tikhonov 正则化

$$\rho_\alpha \in \operatorname*{argmin}_{\rho \in \tilde{\mathcal{B}}_1} \mathcal{S}_{g^{\mathrm{obs}}}(T\rho) + \alpha \, \mathrm{QKL}(\rho, \rho_0) \tag{3}$$

其中正则化参数为 $\alpha > 0$，含噪数据为 $g^{\mathrm{obs}} \approx T\rho^\dagger$。作为数据保真泛函$\mathcal{S}_{g^{\mathrm{obs}}}$，我们考虑范数 L^2 或库尔贝克-莱布勒散度，因为数据通常是概率密度。

在 [25]中，量子相对熵已用于一种更具实践性的场景：当测量数据无法支持唯一重构时，将先验信息纳入考量。在有限维统计场景中，这类估计量已在 [16], 中得到分析，并且已推导出误差估计。

我们将证明，在适当的假设下，最小化问题 (1) 存在唯一解ρ_α；并且对于正则化参数的适当选择，该方法具有正则化性质，即当噪声水平趋于零时， $\rho_\alpha \to \rho^\dagger$ 在迹范数下以及相对于量子相对熵均成立。

在重述正则化性质的条件后，我们总结了量子相对熵的许多重要性质，并证明它是弱-$*$ 下半紧的。我们还刻画了相应的布雷格曼散度，它同样是量子相对熵。在证明 T 是弱-$*$ 到弱-$*$ 连续之后，我们便能够得出其具有正则化性质，并证明当噪声水平递减且正则化参数选择适当时，解序列弱-$*$ 收敛，甚至在迹范数下收敛。在第 3 章中，我们转向有限维情形，并推导量子相对熵的次梯度、共轭泛函和近端算子。借助它们，可以应用 FISTA 和原始-对偶混合梯度算法等迭代求解算法。本节基本上自成一体，并给出了最终的求解算法，使其

能够轻松重新实现并用于实践。为验证理论结果的有效性，我们将其应用于同相层析成像和 PINEM，并给出了相应的数值示例。

2 量子相对熵的正则化性质

本节旨在说明，在广义 Tikhonov 正则化中将量子相对熵用作惩罚泛函，确实定义了一种有效的正则化方法，即估计误差会随着噪声水平趋于 0。我们首先回顾关于这一性质的经典且适用范围广泛的充分条件。对于范数幂型数据保真项，相应结果可追溯至 [34]。更一般的数据保真项则在博士论文 [26] 中进行了研究，以下定理即据此改编而来。

定理 2.1。设 $T:\operatorname{dom}T\subset\mathcal{X}\to\mathcal{Y}$ 是巴拿赫空间之间的一个算子，并设 $\tau_{\mathcal{X}}$ 和$\tau_{\mathcal{Y}}$ 分别是 $\mathcal{X}$ 和 $\mathcal{Y}$上的拓扑。考虑一个精确解 $\rho^{\dagger}\in\operatorname{dom}T$，其对应的精确数据为 $g^{\dagger}:=T\rho^{\dagger}$；一列观测数据 $(g_k^{obs})_{k\in\mathbb{N}}$，其对应的数据保真泛函为$\mathcal{S}_{g_k^{obs}}\colon\ \mathcal{Y}\to[0,\infty]$；以及一个惩罚泛函 $\mathcal{R}\colon\ \mathcal{X}\to\mathbb{R}\cup\{\infty\}$，满足以下条件：

(C.1) $(\mathcal{X},\tau_{\mathcal{X}})$ 和$(\mathcal{Y},\tau_{\mathcal{Y}})$ 是实局部凸向量空间。

(C.2) $\mathcal{R}$ 是序列下半连续的。

(C.3) $\mathcal{R}$ 是序列下半紧的，即对于所有 $C\in\mathbb{R}$，$\{\rho\in\mathcal{X}\colon\mathcal{R}(\rho)\ \le C\}$ 都是紧的。

(C.4) $\operatorname{dom}T\cap\operatorname{dom}\mathcal{R}\neq\emptyset$ 和 T在$\operatorname{dom}T\cap\operatorname{dom}\mathcal{R}$上序列连续。

(C.5) 对所有数据 g_k^{obs}, $\mathcal{S}_{g_k^{obs}}(g)\ =0\Leftrightarrow g=g_k^{obs}$，$\mathcal{S}_{g_k^{obs}}$都是序列下半连续的，并且$\lim_{k\to\infty}\mathcal{S}_{g_k^{obs}}(g^{\dagger})\ =0$。

(C.6) 对于精确数据 $g^{\dagger}$，存在唯一的 $\mathcal{R}$-最小化解 $\rho^{\dagger}\in\operatorname{argmin}_{T\rho=g^{\dagger}}\mathcal{R}(\rho\qquad\qquad\qquad\qquad)$.

(C.7) $\partial\mathcal{R}(\rho^{\dagger})\neq\emptyset$.

在假设 *(C.1)-(C.5)* 下，对任意正则化参数均存在一个解

$$\rho_k\in\operatorname{argmin}_{\rho}\mathcal{S}_{g_k^{obs}}(T\rho)+\alpha_k\mathcal{R}(\rho)\quad for\ allk\in\mathbb{N}$$

。此外，如果满足 *(C.6)*，并且正则化参数的选取满足

$$\lim_{k\to\infty}\alpha_k=0\ and\ \lim_{k\to\infty}\frac{\mathcal{S}_{g_k^{obs}}(g^{\dagger})}{\alpha_k}=0,$$

则序列 ρ_k 在 $\tau_{\mathcal{X}}$ 拓扑中收敛于 $\rho^{\dagger}$。此外，

$$\lim_{k\to\infty}\mathcal{S}_{g^{\dagger}}(T\rho_k)=0\ and\ \lim_{k\to\infty}\mathcal{R}(\rho_k)=\mathcal{R}(\rho^{\dagger}).$$

最后，如果 *(C.7)* 也成立，则

$$\lim_{k\to\infty}\Delta_{\mathcal{R}}^{\sigma}(\rho_k,\rho^{\dagger})=0\quad for\ \sigma\in\partial\mathcal{R}(\rho^{\dagger})$$

其中 $\Delta_{\mathcal{R}}^{\sigma}$ 表示布雷格曼散度。

2.1 量子相对熵的性质

由于数学物理文献中对量子相对熵所采用的定义和设定各不相同，我们首先总结所需的主要性质。我们的设定包含线性项，并不要求算子作用于有限维空间，也不要求其迹为 1，而物理学文献通常会作此类要求。大多数结果仍可迁移到这一设定中，但需要格外谨慎。

我们先介绍所使用的记号。与引言中一样， $\mathcal{H}$ 表示一个复可分希尔伯特空间。令 $\{f_l\}$ 表示 $\mathcal{H}$的任意标准正交基。按照 [28, 第六章]，我们定义如下算子空间：

$$\begin{aligned}
&\mathcal{B} := \mathcal{L}(\mathcal{H}) && \text{（有界算子）}\\
&\mathcal{K} := \mathcal{K}(\mathcal{H}) && \text{（紧算子）}\\
&\mathcal{B}_2 : \left\{\rho \in \mathcal{K} : \sum\nolimits_l \|\rho f_l\|_{\mathcal{H}}^2 < \infty\right\} && \text{（希尔伯特-施密特算子）}\\
&\mathcal{B}_1 := \{\rho \in \mathcal{B}_2 : \mathrm{tr}(|\rho|) < \infty\} && \text{（迹类算子）。}
\end{aligned}$$

这里可以证明， $\mathcal{B}_2$ 与基的选择无关。我们用 $\mathcal{B}$、$\mathcal{K}$、$\mathcal{B}_1$ 和 $\mathcal{B}_2$ 表示厄米（等价地，自伴）算子对应的子空间。根据谱定理，任意 $\rho \in \mathcal{K}$ 都有一个由特征向量 $\{e_l\}$ 构成的标准正交基，其对应的特征值为 ρ_l。对于定义在 ρ的谱上的实值函数 φ ，泛函演算 $\varphi(\rho)$定义为 $\varphi(\rho) f := \sum_l \varphi(\rho_l)(f, e_l)e_l$ ，对所有级数收敛的 f均成立；若 f 有界，则 $\varphi(\rho) \in \mathcal{B}$ [28, 定理 VII.2]。这既用于量子相对熵的定义（2），也用于 $\mathcal{B}_1$的定义，其中 $|\rho| := \sqrt{\rho^*\rho}$ ，且 $\mathrm{tr}(|\rho|) := \sum_l (f_l, |\rho| f_l)$ 同样与 $\{f_l\}$的选择无关。

我们在定理 2.1 中为 $\mathcal{K}$ 选择为弱-∗拓扑，以确保 $\mathcal{B}_1$ 的定义这一点，回顾一下，$\mathcal{B}_1$ 可通过双线性对偶配对 $\langle \rho, \sigma \rangle := \mathrm{tr}(\rho\sigma)$ 识别为 $\mathcal{K}$ 的对偶空间，其中 $\rho \in \mathcal{B}_1$ 和 $\sigma \in \mathcal{K}^*$，例如参见 [28]。换言之， $\mathcal{B}_1$ 具有一个前对偶

$$\mathcal{B}_1^{\#} := \mathcal{K},$$

并且由 $\mathcal{K}$ 在 $\mathcal{B}_1$ 上诱导的弱拓扑称为弱-∗-拓扑。

在定理 2.1 中我们需要实向量空间，因此不能直接使用 $\mathcal{B}_1$ 。然而，我们可以将 $\mathcal{B}_1$ 视为实向量空间，并刻画相应的弱-∗-拓扑。

引理 2.2。 将 $\mathcal{B}_1$ 和 $\mathcal{K}$ 视为实向量空间，我们有 $\mathcal{B}_1 = \mathcal{K}^*$；相对于 $\mathcal{K}$ 在 $\mathcal{B}_1$ 上定义的弱拓扑（下文称为 $\mathcal{B}_1$ 上的弱-∗-拓扑）与由 ∗-拓扑在 $\mathcal{B}_1$上诱导的相对拓扑一致。

证明。通过几步简单计算即可验证，对每个 $\rho \in \mathcal{B}_1$ ，泛函 $f_\rho : \mathcal{K} \to \mathbb{R}$ $\sigma \mapsto \mathrm{tr}(\sigma\rho)$是有界、实线性且良定义的。另一方面，假设 $f \in \mathcal{K}^*$，则由下式定义的 $f : \mathcal{K} \to \mathbb{C}$

$$\tilde{f}(\sigma) = f\left(\frac{1}{2}(\sigma + \sigma^*)\right) - if\left(\frac{i}{2}(\sigma - \sigma^*)\right)$$

是良定义的、复线性的且有界的。于是 $f \in \mathcal{K}^*$ ，且存在 $\rho \in \mathcal{B}_1$ 使得 $f(\sigma) = \mathrm{tr}(\sigma\rho)$。于是对所有 $\sigma \in \mathcal{K}$ ，我们有

$$\mathrm{tr}(\sigma\rho) = \tilde{f}(\sigma) = f\left(\frac{1}{2}(\sigma + \sigma^*)\right) - if\left(\frac{i}{2}(\sigma - \sigma^*)\right) = f(\sigma) \in \mathbb{R}.$$

现在，对每个单位向量 $v \in \mathcal{H}$ ，都存在一个投影 $P_v \in \mathcal{K}$ ，投影到由v张成的子空间上。于是

$$0 = 2\,\mathrm{Im}(\mathrm{tr}(P_v\rho)) = \mathrm{tr}(P_v\rho) - \overline{\mathrm{tr}(P_v\rho)} = \mathrm{tr}(P_v(\rho - \rho^*)) = \langle v, (\rho - \rho^*)v\rangle$$

˜这意味着 $\rho - \rho^* = 0$，因此 $\rho \in \mathcal{B}_{1}$。

˜为了证明这些拓扑一致，我们取一个网 $(\rho_n)_{n\in\mathcal{N}} \subset \mathcal{B}_{1}$。若它相对于 $\mathcal{K}$ 的弱-$*$-拓扑收敛，则它也相对于 $\mathcal{K}$ 的弱-$*$-拓扑收敛。反过来，假设

$$\lim_{n\in\mathcal{N}} \operatorname{tr}(\sigma\rho_n) = \operatorname{tr}(\sigma\rho) \quad \text{for all } \sigma \in \tilde{\mathcal{K}}.$$

那么对于 $\sigma \in \mathcal{K}$

$$\begin{aligned}\lim_{n\in\mathcal{N}} \operatorname{tr}(\sigma\rho_n) &= \lim_{n\in\mathcal{N}} \operatorname{tr}\left(\frac{1}{2}(\sigma+\sigma^*)\rho_n\right) + \operatorname{tr}\left(\frac{1}{2}(\sigma-\sigma^*)\rho_n\right) \\ &= \operatorname{tr}\left(\frac{1}{2}(\sigma+\sigma^*)\rho\right) + \lim_{n\in\mathcal{N}}(-i)\operatorname{tr}\left(\frac{i}{2}(\sigma-\sigma^*)\rho_n\right) \\ &= \operatorname{tr}\left(\frac{1}{2}(\sigma+\sigma^*)\rho\right) - i\operatorname{tr}\left(\frac{i}{2}(\sigma-\sigma^*)\rho\right) = \operatorname{tr}(\sigma\rho).\end{aligned}$$

因此，它也相对于由 $\mathcal{K}$ 导出的弱-$*$-拓扑收敛。 □

˜这个引理使我们能够采用 $\mathcal{B}_1$ 上弱-$*$-拓扑的通常定义。此外，由于巴拿赫空间 $\mathcal{B}_1$ 的对偶空间相对于弱-$*$-拓扑是局部凸的，因此 **(C.1)** 成立。关于连续性，我们有以下最近在 [17] 中证明的引理。

引理 2.3 ((C.2),[17])**.** 量子相对熵 量子相对熵 在两个自变量上均为弱-$*$ 下半连续。

让我们汇总 [21]中的一些进一步结果:

Lem命题 2.4。 定义在可分离赋范空间上的迹类算子上的量子相对熵量子相对熵 *able Hilb*具有以下性质。

1. 量子相对熵$(\rho,\ \sigma) \geq 0$ 以及 量子相对熵$(\rho,\ \sigma) = 0 \Leftrightarrow \rho = \sigma$

2. 量子相对熵 是联合凸的。

3. 量子相对熵 具有单调性，即量子相对熵$(\Phi(\rho)\ \Phi(\sigma)) \leq$量子相对熵$(\rho, \sigma)$ 对于 Φ 保持迹的期望（有关更多细节，请参见 *[21,* 定理 *1]* ）。特别地，对于投影 P 和 $\Phi(\rho) = P\rho P + (I-P)\rho(I-P)$

$$\mathrm{QKL}(P\rho P + (I-P)\rho(I-P), P\sigma P + (I-P)\sigma(I-P)) \leq \mathrm{QKL}(\rho,\sigma)$$

量子相对熵还满足一个将其与迹范数联系起来的不等式，其形式与例如 [11] 中针对库尔贝克-莱布勒散度和 ℓ^1-范数所展示的不等式相同。

引理 2.5。对于 $\rho,\ \sigma \in \tilde{\mathcal{B}}_1^+$

$$\|\rho-\sigma\|_{\mathcal{B}_1}^2 \leq \left(\frac{2}{3}\|\rho\|_{\mathcal{B}_1} + \frac{4}{3}\|\sigma\|_{\mathcal{B}_1}\right)\mathrm{QKL}(\rho,\sigma).$$

证明。 对于迹为 1 的算子， [13] 中给出了一种证明，其中表明

$$\mathrm{QKL}(\rho,\sigma) \geq \frac{1}{2}\|\rho-\sigma\|_{\mathcal{B}_1}^2$$

对于 $\operatorname{tr}\rho = \operatorname{tr}\ \sigma\ = 1$。我们将沿用这一证明，并修改相关部分，以得到更˜一般的形式。为方便起见，我们定义相应的泛函 $f_\rho\colon\ \mathcal{B} \to \mathrm{R}\ a \mapsto \operatorname{tr}(a\rho)$ 和˜$f_\sigma\colon\ \mathcal{B} \to \mathrm{R}\ a \mapsto \operatorname{tr}(a\sigma)$。与 [13] 中一样，我们可以在 $\mathcal{H}$ 上找到一个正交投影 $P = 1_{[0,\infty)}(\sigma-\rho)$，使得

$$\|\rho-\sigma\|_{\mathcal{B}_1} = (f_\sigma - f_\rho)(P) - (f_\sigma - f_\rho)(I-P) \leq |(f_\sigma - f_\rho)(P)| + |(f_\sigma - f_\rho)(I-P)|.$$

文中还证明了，根据单调性

$$\mathrm{QKL}(\rho,\sigma) \geq \left[f_\sigma(P) - f_\rho(P) + f_\rho(P)\ln\frac{f_\rho(P)}{f_\sigma(P)}\right] + \left[f_\sigma(I-P) - f_\rho(I-P) + f_\rho(I-P)\ln\frac{f_\rho(I-P)}{f_\sigma(I-P)}\right].$$

由 [4] 我们得到不等式

$$|x-y| \leq \sqrt{\left(\frac{2}{3}x+\frac{4}{3}y\right)\left(y-x+x\ln\frac{x}{y}\right)} \quad \text{for all } x\geq 0, y>0.$$

结合柯西-施瓦茨不等式可得

$$(|x-y|+|a-b|)^2 \leq \left(\frac{2}{3}(x+a)+\frac{4}{3}(y+b)\right)\left(y-x+x\ln\frac{x}{y}+b-a+a\ln\frac{a}{b}\right) \quad \text{for all } x,a\geq 0, y,b>0.$$

最后，令 $x = f_\rho(P)$ $y = f_\sigma(P)$ $a = f_\rho(I-P)$ 和 $b = f_\sigma(I-P)$，即可得到所需的不等式。 □

我们还可以用量子相对熵对迹作如下界定：

引理 2.6. 对于欧拉数 e，对所有 $\rho, \rho_0 \in \mathcal{B}_1$，以下不等式成立：

$$\mathrm{QKL}(\rho,\rho_0) \geq \mathrm{tr}(\rho) + (1-\mathrm{e})\,\mathrm{tr}(\rho_0).$$

证明. 若量子相对熵(ρ,ρ_0)为无穷大，则该不等式显然成立。特别地，我们可以假设 $\rho,\rho_0 \in \tilde{\mathcal{B}}_1^+$。如有必要，我们可以转而考虑一个子空间，从而进一步假设 ρ_0 的离散谱不包含零。于是， ρ_0的谱分解具有形式 $\rho_0 = \sum_j d_j P_j$，其特征值为 $d_j > 0$，对应的特征投影为 P_j，满足 $\sum_j P_j = I$。根据凸函数 $\eta, [0,\infty) \to \mathrm{R}\ t \mapsto t\ln t$ 的詹森迹不等式（见 [5]），我们得到

$$\mathrm{tr}\left(\eta\left(\sum\nolimits_j P_j\rho P_j\right)\right) \leq \mathrm{tr}\left(\sum\nolimits_j P_j\eta(\rho)P_j\right) = \sum\nolimits_j \mathrm{tr}(P_j\eta(\rho)P_j) = \sum\nolimits_j \mathrm{tr}\left(P_j\eta(\rho)\right) = \mathrm{tr}(\eta(\rho))\,.$$

现在假设特征值 d_j 按其重数计出现，即对所有 j，秩 $P_j = 1$，则对于某些 ρ 的"对角元" $\rho_{jj} \geq 0$，有 $P_j\rho P_j = \rho_{jj}P_j$。由此得到

$$\sum\nolimits_j \eta(\rho_{jj}) \leq \mathrm{tr}(\eta(\rho)).$$

由此我们现在得到

$$\begin{aligned}\mathrm{QKL}(\rho,\rho_0) &= \mathrm{tr}(\rho_0) - \mathrm{tr}(\rho) + \mathrm{tr}(\eta(\rho)) - \mathrm{tr}(\rho\ln\rho_0)\\ &\geq \sum\nolimits_j (d_j - \rho_{jj} + \eta(\rho_{jj})) - \mathrm{tr}\left(\rho\sum\nolimits_j \ln(d_j)P_j\right)\\ &= \sum\nolimits_j (d_j - \rho_{jj} + \rho_{jj}\ln\rho_{jj} - \rho_{jj}\ln d_j)\,.\end{aligned}$$

函数 $x \mapsto x\ln x - x\ln y - x$ 在 $[0,\infty)$ 上是凸的，因此大于其在 $\mathrm{e}y$ 处的线性近似。对于 $x \in [0,\infty)$ 和 $y \in (0,\infty)$，这意味着不等式

$$\begin{aligned}x\ln x - x\ln y - x &\geq \mathrm{e}y\ln(\mathrm{e}y) - (\mathrm{e}y)\ln y - \mathrm{e}y + (\ln(\mathrm{e}y) - \ln y)(x - \mathrm{e}y)\\ &= 0 + 1\cdot(x-\mathrm{e}y) = x - \mathrm{e}y\,.\end{aligned}$$

因此， $d_j - \rho_{jj} + \rho_{jj}\ln\rho_{jj} - \rho_{jj}\ln d_j \geq d_j + \rho_{jj} - \mathrm{e}d_j$，并且

$$\mathrm{QKL}(\rho,\rho_0) \geq \sum\nolimits_j (d_j + \rho_{jj} - \mathrm{e}d_j) = \mathrm{tr}(\rho) + (1-\mathrm{e})\,\mathrm{tr}(\rho_0).$$ □

量子相对熵通常只针对迹为 1 的矩阵来考虑，因此这个不等式并不太有用。然而在我们的情形下，它使我们无需使用额外的迹约束即可证明下半紧性。

定理 2.7 ((C.3))。 泛函 量子相对熵$(\cdot, \rho_0) : \mathcal{B}_1 \to \mathbb{R} \cup \{\infty\}$ 对于 $\rho_0 \in \tilde{\mathcal{B}}_1^+$是序列弱-$*$下半紧的。

证明。 令 $C > 0$ ，并令 $\rho \in S_C := \{$量子相对熵$(\cdot,\ \rho_0) \le C\}$。由于量子相对熵$(\rho,\ \rho_0)$ 有限，我们有$\rho \in \tilde{\mathcal{B}}_1^+$且 $\mathrm{tr}(\rho) = \|\rho\|_{\mathcal{B}_1}$。由引理 2.6 可得

$$C \ge \mathrm{QKL}(\rho, \rho_0) \ge \mathrm{tr}(\rho) + (1 - \mathrm{e})\,\mathrm{tr}(\rho_0) \quad \Rightarrow \quad \|\rho\|_{\mathcal{B}_1} \le C + (\mathrm{e} - 1)\,\mathrm{tr}(\rho_0).$$

由此我们证明了 S_C 包含在一个范数有界集中。由于 $\mathcal{K}$ 是有限秩算子空间的范数闭包，因此它是可分的 [28, 定理 VI.13]。接着，巴拿赫-阿劳格鲁定理 [33, 定理 3.17] 的序列版本蕴含 S_C 是序列弱 $*$ 预紧。由于 S_C 由引理 2.3 可知是序列弱$*$ 闭的，因此它是序列弱$*$紧的。

□

我们还可以进一步考察相应的次微分和布雷格曼散度。次微分与克莱因不等式密切相关，例如可见于 [18, 10]。我们仍重现该证明，因为其中包含了布雷格曼散度的计算。

引理 2.8. 对于 $\rho \in \tilde{\mathcal{B}}_1^+$，如果$\ln \rho - \ln \rho_0 \in \tilde{\mathcal{B}}$ 良定义，则

$$\ln \rho - \ln \rho_0 \in \partial(\mathrm{QKL}(\cdot, \rho_0))(\rho).$$

˜证明。 我们取 $\sigma \in \mathcal{B}_1$ 并计算

$$\begin{aligned}
&\mathrm{QKL}(\sigma, \rho_0) - \mathrm{QKL}(\rho, \rho_0) \\
&\qquad - \mathrm{tr}((\ln \rho - \ln \rho_0)(\sigma - \rho)) = \mathrm{tr}(\rho_0 - \sigma + \sigma \ln \sigma - \sigma \ln \rho_0) \\
&\qquad\qquad - \mathrm{tr}(\rho_0 - \rho + \rho \ln \rho - \rho \ln \rho_0) \\
&\qquad\qquad - \mathrm{tr}(\sigma \ln \rho - \rho \ln \rho - \sigma \ln \rho_0 + \rho \ln \rho_0) \\
&\qquad\qquad = \mathrm{tr}(\rho - \sigma + \sigma \ln \sigma - \sigma \ln \rho) = \mathrm{QKL}(\sigma, \rho) \ge 0.
\end{aligned}$$

最后一个不等式由引理 2.4 得出 . □

注 2.9. 在此语境下，良定义是指我们将$\ln \rho - \ln\ \rho_0$ 视为单个算子，而不是两个都必须良定义的算子之差。这意味着我们将$\ln \rho - \ln\ \rho_0$ 在 ρ 和 ρ_0的核的交集上定义为0。此外，对于 v，若其属于 ρ 和 ρ_0 对应于 λ 和 λ_0 的公共特征子空间，则有 $(\ln \rho - \ln\ \rho_0)v = (\ln\ \lambda - \ln\ \lambda_0)v$，据此再通过线性延拓，在公共特征子空间的并集上定义$\ln\ \rho - \ln\ \rho_0$。这样，即使 ρ 和 ρ_0 的谱总是以0 为聚点，我们仍然可以使用算子对数，尽管$\ln\ \rho$ 和$\ln \rho_0$ 分别并非良定义的有界算子。

注 2.10 ((C.7))。 为满足条件 *(C.7)*，必须将次微分视为相对于弱-$*$ 拓扑的对偶空间 $\mathcal{B}_1$ 的子集˜。这与通常的对偶˜空间 $\mathcal{B}$ 不同；根据弱-$*$ 拓扑的性质，第*3*节*.14] [33,* 实际上与其前对偶

˜ $\mathcal{K}$ 重合。因此，*(C.7)* 需要更强的假设 $\ln \rho - \ln \rho_0 \in \mathcal{K}$。

利用这个次微分，我们可以计算相应的布雷格曼散度。

推论 2.11. 对于 $\rho \in \tilde{\mathcal{B}}_1^+$，如果 $\ln \rho - \ln\ \rho_0 \in \mathcal{B}$ 是良定义的，那么相对于 $\ln \rho - \ln \rho_0 \in \partial$(量子相对熵$(\cdot,$ $\rho_0))(\rho)$ 的布雷格曼散度为

$$\Delta^{\ln \rho - \ln \rho_0}_{\mathrm{QKL}(\cdot, \rho_0)}(\sigma, \rho) = \mathrm{QKL}(\sigma, \rho).$$

证明。 在前一个定理的证明中，我们已经计算出

$$\mathrm{QKL}(\sigma, \rho_0) - \mathrm{QKL}(\rho, \rho_0) - \mathrm{tr}((\ln \rho - \ln \rho_0)(\sigma - \rho)) = \mathrm{QKL}(\sigma, \rho)$$

而左侧恰好是布雷格曼散度。 □

2.2 算子与数据保真泛函的弱-$*$-连续性

我们将实数列空间记为 $\ell^p(\mathbb{Z})$，将复数列空间记为 $\ell^p(\mathbb{Z},\mathbb{C})$。算子 T 的陪域也有一个前对偶，由此得到弱-$*$-拓扑，其定义为

$$L^2(I,\ell^p(\mathbb{Z}))^{\#} := \begin{cases} L^2(I,c_0(\mathbb{Z})), & p=1 \\ L^2(I,\ell^{p'}(\mathbb{Z})), & p\in(1,\infty) \quad \frac{1}{p}+\frac{1}{p'}=1. \end{cases}$$

$L^2(I,c_0(\mathbb{Z}))^* = L^2(I,\ell^1(\mathbb{Z}))$ 和 $L^2(I,\ell^{p'}(\mathbb{Z}))^* = L^2(I,\ell^p(\mathbb{Z}))$ 这些事实可由[14, 第 1 章3] 利用 $p\in[1,\infty)$ 的 $\ell^p(\mathbb{Z})$ 的可分性推导得出。

我们首先证明算子 T 是从弱-$*$ 到弱-$*$ 的序列连续映射。

引理 2.12 ((C.4))。形式为 (1) 的算子 T，若 $\{U_\theta\in\mathcal{B}:\ \theta\in I\}$ 一致有界，则是从弱-$*$ 到弱-$*$ 的序列连续映射（特别地，若所有算子 U_θ 都是酉算子）。

证明。我们从一个序列 $\rho_n\in\tilde{\mathcal{B}}_1^+$ 开始，该序列在弱-$*$ 拓扑下收敛到 ρ，即 e.,

$$\lim_{n\to\infty}\operatorname{tr}(\sigma\rho_n)=\operatorname{tr}(\sigma\rho)\quad\text{for all }\sigma\in\mathcal{K}.$$

我们首先固定 θ，证明逐点收敛。对于 $(a_l)_{l\in\mathbb{Z}}\in(\ell^{p'})$，有

$$\begin{aligned}&\langle(T\rho)_\theta,a\rangle=\sum_{l\in\mathbb{Z}}(T\rho)(\theta,l)a_l=\sum_{l\in\mathbb{Z}}\langle a_le_l,U_\theta\rho U_\theta^*e_l\rangle=\operatorname{tr}(D_aU_\theta\rho U_\theta^*)\\&\text{with }D_av:=\sum_l a_l(e_l,v)e_l.\end{aligned}$$

然后 D_a 是紧的，因为 $a\in(\ell^{p'})\subset c_0$ 且

$$\begin{aligned}\lim_{n\to\infty}\operatorname{tr}(D_aU_\theta\rho_nU_\theta^*)&=\lim_{n\to\infty}\operatorname{tr}(\rho_nU_\theta^*D_aU_\theta)=\operatorname{tr}(\rho U_\theta^*D_aU_\theta)\\&=\operatorname{tr}(D_aU_\theta\rho U_\theta^*)=\langle(T\rho)_\theta,a\rangle.\end{aligned}$$

由于每个弱-$*$-收敛序列都是范数有界的，我们可以假设 ρ_n 的范数被某个常数 C所界。现在取 $a(\theta)\in L^2(I,(\ell^{p'}(\mathbb{Z})))$，得到

$$\begin{aligned}|\langle(T\rho_n)_\theta,a(\theta)\rangle|&\le\|a(\theta)\|_{(\ell^{p'}(\mathbb{Z}))}\|(T\rho_n)_\theta\|_{\ell_p(\mathbb{Z})}\le\|a(\theta)\|_{(\ell^{p'}(\mathbb{Z}))}\|(T\rho_n)_\theta\|_{\ell_1(\mathbb{Z})}\\&=\|a(\theta)\|_{(\ell^{p'}(\mathbb{Z}))}\operatorname{tr}(U_\theta\rho_nU_\theta^*)\le\|a(\theta)\|_{(\ell^{p'}(\mathbb{Z}))}\|U_\theta^*U_\theta\|\operatorname{tr}(\rho_n)\\&=\|a(\theta)\|_{(\ell^{p'}(\mathbb{Z}))}\|U_\theta\|^2\|\rho_n\|_1\le\|a(\theta)\|_{(\ell^{p'}(\mathbb{Z}))}\|U_\theta\|^2C.\end{aligned}$$

因此，$|\langle(T\rho_n)_\theta,\ a(\theta)\rangle|$ 逐点一致有界于 $L^2(I)$中的某个函数。由控制收敛定理，$\langle(T\rho_n)_\theta,\ a(\theta)\rangle$ 在 L^2 中收敛到 $\langle(T\rho)_\theta,\ a(\theta)\rangle$，从而完成证明。 □

为了同时分析 L^2 和 Kullback–Leibler 数据保真项，并允许使用不同的混合 L^p-范数，我们定义 $\mathcal{S}_{g^{\mathrm{obs}}}:\ L^2(I,\ \ell^p(\mathbb{Z}))\to[0,\infty]$，如下:

$$\mathcal{S}_{g^{\mathrm{obs}}}(f):=\int_I\sum_{l\in\mathbb{Z}}s(g^{\mathrm{obs}}(\theta,l),f(\theta,l))\mathrm{d}\theta$$

其中 $s:\mathrm{R}\times\mathrm{R}\to[0,\infty]$ 和 $p\in[1,\infty)$。我们得到通常的 L^2-范数，其中 $p=2$ 和

$$s_2(x,y):=\frac{1}{2}|x-y|^2.$$

此外，通过设置

$$s_{KL}(x,y) := \begin{cases} \infty & \text{if } x<0 \text{ or } y<0 \text{ or } (x>0 \text{ and } y=0) \\ y & \text{if } x=0 \text{ and } y\geq 0 \\ y-x+x\ln\frac{x}{y} & \text{else} \end{cases}$$

并且 $p=1$ 我们得到库尔贝克-莱布勒散度。由于陪域是一个混合 L^p-空间，数据保真泛函的弱-$*$下半连续性并非显而易见。我们将看到，在本例中，可以将问题归结为研究 $L^2(I)$ 上的下半连续性。

引理 2.13。若 $\mathcal{F}_{s,\,l}\colon\ L^2(I)\to[0,\infty],\ f\mapsto\int_I s(g^{obs}(\theta,l),\ f(\theta))\mathrm{d}\theta$ 对于所有 $l\in\mathbb{Z}$ 关于范数拓扑都是凸且下半连续的，则 $\mathcal{S}_{g^{obs}}$ 是弱-$*$ 下半连续的。此外，若 $s(x,y)=0\Leftrightarrow x=y$，则 $\mathcal{S}_{g^{obs}}(g)=0\Leftrightarrow g=g^{obs}$。

证明。我们将使用下半连续泛函的上确界仍然是下半连续泛函这一事实。于是，对于 $n\in\mathbb{N}$，我们定义

$$\mathcal{F}_n(f) := \int_I \sum_{|l|<n} s(g^{\mathrm{obs}}(\theta,l), f(\theta,l))\mathrm{d}\theta$$

并且由于被加项的非负性，我们得到 $\mathcal{F}(f)=\sup_n \mathcal{F}_n(f)$。现在考虑 $(f_k)_k\subset L^2(I,\ell^p(\mathbb{Z}))$，它在弱-$*$-拓扑中收敛到 f。这意味着

$$\lim_{k\to\infty}\langle f_k,a\rangle=\langle f,a\rangle \quad \text{for all } a\in L^2(I,(\ell^{p'}(\mathbb{Z})))$$

现在令 $P_n\colon\ L^2(I,\ell^p(\mathbb{Z}))\to L^2(I,\mathbb{R}^{2n-1})$ 为 f_k 的逐点限制，使得

$$(P_nf)(\theta,l)=f(\theta,l) \quad \text{for all } |l|<n$$

此外，将嵌入 $\iota_n\colon\ L^2(I,\mathbb{R}^{2n-1})\to L^2(I,(\ell^{p'}(\mathbb{Z})))$ 定义为

$$(\iota_n a)_j(\theta)=\begin{cases} a(\theta,l) & |l|<n \\ 0 & \text{else} \end{cases}$$

对所有 $a^{(n)}\in L^2(I,\mathbb{R}^{2n-1})$，我们得到

$$\lim_{k\to\infty}\langle P_nf_k,a^{(n)}\rangle=\lim_{k\to\infty}\langle f_k,\iota_na^{(n)}\rangle=\langle f,\iota_na^{(n)}\rangle=\langle P_nf,a^{(n)}\rangle$$

这表明 $(P_nf_k)_k$ 在 $L^2(I,\mathbb{R}^{2n-1})$ 中弱收敛。对于凸泛函，弱下半连续性等价于强下半连续性。由于 $\mathcal{F}_{s,l}$ 在 $L^2(I)$ 上下半连续，这一性质可转移到 $\mathcal{F}_n$ 在 $L^2(I,\mathbb{R}^{2n-1})$ 上的限制，而该限制按定义与 $\mathcal{F}_n$ 重合。因此，$\mathcal{F}_n$ 是弱下半连续的，我们得到

$$\lim_{k\to\infty}\mathcal{F}_n(f_k)=\lim_{k\to\infty}\mathcal{F}_n(P_nf_k)=\mathcal{F}_n(P_nf)=\mathcal{F}_n(f).$$

这表明 $\mathcal{F}_n$ 是弱-$*$ 下半连续的，并且取上确界后，$\mathcal{F}$ 也弱-$*$下半连续。

对于命题的第二部分，我们利用如下事实：若 $g=g^{\mathrm{obs}}$，则 $s(g^{\mathrm{obs}}(\theta,\ l),\ g(\theta,\ l))=0$，这意味着 $\mathcal{S}_{g^{\mathrm{obs}}}(g)=0$。反过来，若 $\mathcal{S}_{g^{\mathrm{obs}}}(g)=0$，由 s 的正性可知，$s(g^{\mathrm{obs}}(\theta,\ l),\ g(\theta,\ l))=0$ 几乎处处成立。这意味着 $g=g^{\mathrm{obs}}$ 几乎处处成立。 □

由此可推得 KL 和 $\|\ \|_2$具有所需性质。

推论 2.14 ((C.5))。泛函 $\frac{1}{2}\|\cdot-g^{obs}\|^2_{L^2(I,\ \ell^2(\mathbb{Z}))}$ 在 $L^2(I,\ell^2(\mathbb{Z}))$ 上，以及 $\mathrm{KL}(g^{obs},\cdot)$ 在 $L^2(I,\ell^1(\mathbb{Z}))$ 上均为弱-* 下半连续，并满足 *(C.5)*。

证明。 根据引理 2.13，我们只需证明对所有 $f \in L^2(I)$, $\|\cdot - f\|^2_{L^2(I)}$ 都是下半连续的，这显然成立。对于 KL，我们需要证明 KL 在 $L^2(I)$ 上下半连续；这可由 KL 在 $L^1(I)$ 上的下半连续性推出，该性质已在 [11]中证明。 □

如果愿意对未知密度矩阵 ρ施加进一步的先验假设，那么通常可以通过以下简单引理建立 $*$- T 连续性:

引理 2.15。 设线性算子 T： 定义域$T \subset \mathcal{B}_1 \to \mathcal{Y}$ 取值于某个巴拿赫空间 $\mathcal{Y}$，其定义域为 定义域T，且该定义域在某个巴拿赫空间ˆ ˆ$\mathcal{B}$中是弱序列紧的，并且 $\mathcal{B}$ 关于范数拓扑连续嵌入于 $\mathcal{B}_1$ 。进一步假设ˆ T 关于 $\mathcal{B}$ 和 $\mathcal{Y}$的范数拓扑连续。则 T 从 $\mathcal{B}_1$ 到 $\mathcal{Y}$是弱-$*$ 到弱的序列连续映射。

证明。 考虑定义域 $\operatorname{dom} T$ 中的一个序列 $(\rho^{(k)})$，它关于由 $\mathcal{K}$诱导的 $\mathcal{B}_1$上的弱-$*$-拓扑收敛于 ρ 。由 $\operatorname{dom} T$的弱紧性可知，$(\rho^{(k)})$具有一个子序列，该子序列弱收敛ˆ ˆ于 $\mathcal{B}$中的某个 $\tilde{\rho}$ 。由于嵌入 $\mathcal{B} \hookrightarrow \mathcal{B}_1$ 也是弱连续的，该子序列也弱收敛，并且弱-$*$ 收敛于 $\mathcal{B}_1$中的 $\tilde{\rho}$ 。因此 $\tilde{\rho} = \rho$; 又由于每个子序列都有一个弱收敛于 $\mathcal{B}$中的 ρ 的子序列，所以整个序列弱收敛于 $\mathcal{B}$中的 ρ 。由于 T的强连续性蕴含从ˆ$\mathcal{B}$ 到 $\mathcal{Y}$的弱连续性，因此可得 $T\rho^{(k)} \rightharpoonup T\rho$ in $\mathcal{Y}$。 □

请注意，在这种情况下，可以选择 $\tau_{\mathcal{Y}}$ 作为弱拓扑，而不是弱-$*$-拓扑，这有助于验证 $\mathcal{S}_{g^{\text{obs}}}$的序列下半连续性。

2.3 主要结果

经过这些准备，我们能够证明量子相对熵的正则化性质。

定理 2.16. 假设 T 满足引理 *2.12* 的假设，且 $\mathcal{S}_{g^{obs}}$ 为2-范数或库尔贝克-莱布勒散度。此外，假设条件 *(C.6)* 得到满足，并且我们有满足与定理 *2.1* 相同假设的观测数据序列$(g_k^{obs})_{k\in\mathbb{N}}$ 和参数序列$(\alpha_k)_{k\in\mathbb{N}}$ 。那么，正则化问题的解

$$\rho_k := \operatorname*{argmin}_{\rho} \mathcal{S}_{g_k^{obs}}(T\rho) + \alpha_k \operatorname{QKL}(\rho, \rho_0)$$

存在，收敛到 $\rho^\dagger$ 的弱-$*$-拓扑，并且

$$\operatorname{QKL}(\rho_k, \rho_0) \overset{k\to\infty}{\to} \operatorname{QKL}(\rho^\dagger, \rho_0) \ \textit{and} \ \lim_{k\to\infty} \mathcal{S}_{g^\dagger}(T\rho_k) = 0.$$

如果此外$\ln\rho^\dagger - \ln\rho_0$ 有定义且是紧的，则有

$$\lim_{k\to\infty} \operatorname{QKL}(\rho_k, \rho^\dagger) = 0 \ \textit{and} \ \lim_{k\to\infty} \|\rho_k - \rho^\dagger\|_{\mathcal{B}_1} = 0.$$

证明。 本定理中的假设与本章前面的结果共同保证条件 (C.1)–(C.6) 均成立，因此前三个收敛结果可直接由定理 2.1 得出。假设 $\ln\ \rho^\dagger - \ln\rho_0$ 良定义且紧，还可根据引理 2.8 得到 (C.7)。这一性质保证了布雷格曼散度意义下的收敛，而我们已在推论 2.11 中再次证明该散度就是量子相对熵。最后，对于范数收敛，我们利用 ρ_k弱-$*$ 收敛，因此范数有界。由引理 2.5 可得

$$\lim_{k\to\infty} \|\rho_k - \rho^\dagger\|^2_{\mathcal{B}_1} \le \lim_{k\to\infty} \left(\frac{2}{3}\|\rho_k\|_{\mathcal{B}_1} + \frac{4}{3}\|\rho^\dagger\|_{\mathcal{B}_1}\right) \operatorname{QKL}(\rho_k, \rho^\dagger) \le \lim_{k\to\infty} C \operatorname{QKL}(\rho_k, \rho^\dagger) = 0. \quad \square$$

注 2.17. 到目前为止，我们唯一尚未讨论的条件是 *(C.6)*。例如，当 T 为单射时，该条件便成立。即使不满足该条件，我们至少仍能得到 ρ_k 的某个子序列收敛到某个 $\mathcal{R}$-极小化解 *[26]* 的所有收敛结果。

3 算法方法

本节将探讨求解正则化问题的方法。为此，我们首先定义相应的离散问题，然后刻画量子相对熵的次微分、共轭泛函和近端算子。随后即可采用凸优化中的标准算法求解该问题。

3.1 离散设置的描述

˜我们通过选取合适的有限正交归一基 $V := \{v_n \in \mathcal{H} : 1 \leq n \leq N\}$ 来离散化空间 $\mathcal{B}_1$，并配以相应的正交归一投影 $P_V\colon \mathcal{H} \to \operatorname{span} V \cong \mathbb{C}^N$。于是，$\rho \in \mathcal{B}_1$ 的离散版本 ˜为 ˆ$\rho :=$ $P_V \rho P_V^*$，可将其视为厄米矩阵 ˆ$\rho \in \operatorname{Herm}(N)$。基 V 的具体选择取决于实验设置，并应确保解的离散化误差 $\|\rho^\dagger - P_V^* P_V \rho^\dagger P_V^* P_V\|_{\mathcal{B}_1}$ 较小。此外，我们假设 ˆρ_0 是满秩的，因为对于零空间未包含于 ρ_0零空间的算子，$\operatorname{QKL}(\rho, \rho_0)$ 为无穷大。这意味着，将研究范围限制在 ˆρ_0 满秩的子空间内，并不会缩小搜索空间。为了将凸分析的方法应用于 $\operatorname{Herm}(N)$，将其视为实向量空间，并赋予内积 $\langle \hat{\sigma}, \hat{\rho} \rangle_{\operatorname{Herm}(N)} := \operatorname{tr}(\hat{\sigma}\hat{\rho}) \in \mathbb{R}$，是有益的（参见引理 2.2）。

陪域的离散化通常由物理设置决定，由此得到算子 $Q\colon L^2(I, \ell^p(\mathbb{Z})) \to \mathbb{R}^{M_\theta \cdot M_l}$，该算子通过对 θ 和 l 的有限个取值附近进行局部平均来定义。现在，离散前向算子定义为 ˆ$T\colon \operatorname{Herm}(N) \to \mathbb{R}^{M_\theta \cdot M_l}$

$$\hat{T}(\hat{\rho}) := QT(P_V^* \hat{\rho} P_V).$$

如后文针对具体应用所讨论的，有时使用稍作修改的离散算子会更有用，该算子在离散情形下模拟连续算子的性质。本节中，我们只假设存在某种算子的离散形式，并且也可以计算其伴随算子 ˆT^*。最后，与连续情形一样，我们可以定义 $\widehat{\mathcal{S}}_{g^{\text{obs}}} : \mathbb{R}^{M_\theta \cdot M_l} \to \mathbb{R} \cup \{\infty\}$，其定义如下：

$$\widehat{\mathcal{S}}_{g^{\text{obs}}}(f) := \sum_{l,\theta} s(g^{\text{obs}}(\theta, l), f(\theta, l))$$

其中 s 的定义与第 1 节相同。这引出了离散问题

$$\hat{\rho}_\alpha = \operatorname*{argmin}_{\hat{\rho} \in \operatorname{Herm}(N)} \widehat{\mathcal{S}}_{g^{\text{obs}}}(\hat{T}\hat{\rho}) + \alpha \widehat{\operatorname{QKL}}(\hat{\rho}, \hat{\rho}_0). \tag{4}$$

⌢与其他量一样，我们用量子相对熵表示所处理的是有限维空间。

3.2 矩阵空间上量子相对熵的凸分析

现在，我们进一步考察有限维情形下量子相对熵的性质。请注意，无限维版本的所有结果仍然成立，而现在我们还可以应用 [20] 中发展的厄米矩阵凸分析，从而精确计算次微分和共轭泛函。

⌢**引理 3.1.** 对于处于 ˆρ 的本质定义域内的量子相对熵 $(\cdot, \hat{\rho}_0)$，次微分为

$$\partial\left(\widehat{\operatorname{QKL}}(\cdot, \hat{\rho}_0)\right)(\hat{\rho}) = \begin{cases} \{\ln \hat{\rho} - \ln \hat{\rho}_0\} & if\ \ker \hat{\rho} = \{0\} \\ \emptyset & else \end{cases}$$

证明。 我们首先假设 ˆρ 的特征值包含于 $(0, \infty)$ 中，因此可以计算 $\ln \hat{\rho}$ 和 $\ln \hat{\rho}_0$。此外，函数 $f\colon (0, \infty) \to \mathbb{R}\ t \mapsto t \ln t - t$ 是凸的，而函数 $\tilde{f}\colon (0, \infty)^N \to \mathbb{R}, (t_1, \ldots, t_N) \mapsto \sum_j f(t_j)$ 是凸的，且不依赖其参数的顺序

。这意味着 [20, 定理 3.2] 适用，相应的谱函数$f_{\mathrm{spec}}\colon \mathrm{Herm}(N) \to [0,\infty)$，$\hat{\rho} \mapsto \mathrm{tr}(\hat{\rho}\ln\hat{\rho}-\hat{\rho})$ 的次微分为

$$\partial f_{\mathrm{spec}}(\hat{\rho}) = \{f'(\hat{\rho})\} = \{\ln\hat{\rho}\}.$$

为了求得量子相对熵的次微分，我们利用 $-\mathrm{tr}(\cdot\ln\hat{\rho}_0)$ 是线性的，而 $\mathrm{tr}(\hat{\rho}_0)$ 是常数。因此，我们可以将相应的次微分相加，得到

$$\partial\left(\widehat{\mathrm{QKL}}(\cdot,\hat{\rho}_0)\right)(\hat{\rho}) = \partial f_{\mathrm{spec}}(\hat{\rho}) + \{-\ln\hat{\rho}_0\} + \{0\} = \{\ln\hat{\rho}-\ln\hat{\rho}_0\}.$$

现在我们只需证明，当 $\hat{\rho}$ 的核非平凡时，次微分为空。令 P 为投影到 $\hat{\rho}$ 的零空间中的一维子空间上的投影。与引理的陈述相矛盾的是，我们现在假设存在 $\hat{\sigma}$，使得 $-I+\ln\hat{\rho}_0+\hat{\sigma} \in \partial(\text{量子相对熵}(\cdot,\hat{\rho}_0))(\hat{\rho})$。根据次微分的定义，对于所有 $t>0$，我们有

$$\begin{aligned}
&\widehat{\mathrm{QKL}}(\hat{\rho}+tP,\hat{\rho}_0) - \widehat{\mathrm{QKL}}(\hat{\rho},\hat{\rho}_0) \geq \mathrm{tr}((-I+\ln\hat{\rho}_0+\hat{\sigma})(\hat{\rho}+tP-\hat{\rho}))\\
\Leftrightarrow\quad & \mathrm{tr}(-tP+(\hat{\rho}+tP)\ln(\hat{\rho}+tP)-\hat{\rho}\ln\hat{\rho}+tP\ln\hat{\rho}_0) \geq \mathrm{tr}((-I+\ln\hat{\rho}_0+\hat{\sigma})tP)\\
\Leftrightarrow\quad & \mathrm{tr}((\hat{\rho}+tP)\ln(\hat{\rho}+tP)-\hat{\rho}\ln\hat{\rho}) \geq t\,\mathrm{tr}(\hat{\sigma}P)\\
\Leftrightarrow\quad & \ln t \geq \mathrm{tr}(\hat{\sigma}P).
\end{aligned}$$

由于最后一个不等式对所有 $t>0$ 都成立，且左侧在 $t \searrow 0$ 时趋于 $-\infty$，我们便得出了矛盾。 □

现在我们可以利用杨氏等式推导共轭泛函。

引理 3.2. 对于厄米矩阵 σ，$\mathrm{QKL}(\cdot,\hat{\rho}_0)$ 的共轭泛函为

$$\left(\widehat{\mathrm{QKL}}(\cdot,\hat{\rho}_0)\right)^*(\hat{\sigma}) = \mathrm{tr}(\exp(\hat{\sigma}+\ln\hat{\rho}_0)-\hat{\rho}_0)$$

证明。 取厄米矩阵 $\hat{\sigma}$，并定义 $\hat{\nu} := \exp(\hat{\sigma}+\ln\hat{\rho}_0)$。则 $\hat{\nu}$ 满秩，且

$$\partial\left(\widehat{\mathrm{QKL}}(\cdot,\hat{\rho}_0)\right)(\hat{\nu}) = \{\ln\hat{\nu}-\ln\hat{\rho}_0\} = \{\hat{\sigma}+\ln\hat{\rho}_0-\ln\hat{\rho}_0\} = \{\hat{\sigma}\}.$$

由杨氏等式可得

$$\begin{aligned}
\left(\widehat{\mathrm{QKL}}(\cdot,\hat{\rho}_0)\right)^*(\hat{\sigma}) &= \mathrm{tr}\,(\hat{\sigma}\hat{\nu}) - \widehat{\mathrm{QKL}}(\hat{\nu},\hat{\rho}_0)\\
&= \mathrm{tr}\,(\hat{\sigma}\hat{\nu}) - \mathrm{tr}\,(\hat{\rho}_0-\hat{\nu}+\hat{\nu}(\hat{\sigma}+\ln\hat{\rho}_0-\ln\hat{\rho}_0))\\
&= \mathrm{tr}\,(\hat{\nu}-\hat{\rho}_0) = \mathrm{tr}\,(\exp(\hat{\sigma}+\ln\hat{\rho}_0)-\hat{\rho}_0)\,.
\end{aligned}$$
□

利用次微分，我们还可以计算近端映射

$$\mathrm{prox}_{\tau\widehat{\mathrm{QKL}}(\cdot,\rho_0)}(\hat{\sigma}) := \mathrm{argmin}_{\hat{\rho}}\,\widehat{\mathrm{QKL}}(\hat{\rho},\hat{\rho}_0) + \frac{1}{2\tau}\|\hat{\rho}-\hat{\sigma}\|_2^2.$$

这与库尔贝克-莱布勒散度的情形基本相同，其计算需要求双射函数 $g\colon (0,\infty) \to \mathrm{R}\ t \mapsto \ln t + t$ 的逆函数，该函数可以用 Lambert W 函数表示为 $g^{-1}(t) = W(e^t)$。对于较大的t，使用这一表达式在数值上不稳定，因此应改用牛顿法。

引理 3.3. 对于$\hat{\sigma}$ 厄米

$$\mathrm{prox}_{\tau\widehat{\mathrm{QKL}}(\cdot,\rho_0)}(\hat{\sigma}) = \tau g^{-1}\left(\frac{\hat{\sigma}}{\tau}+\ln\hat{\rho}_0-\ln\tau\right) = \tau W\left(\frac{1}{\tau}\exp\left(\frac{\hat{\sigma}}{\tau}+\ln\hat{\rho}_0\right)\right).$$

证明. 近端映射也可以由以下方式刻画

$$\operatorname{prox}_{\tau\widehat{\mathrm{QKL}}(\cdot,\rho_0)}(\hat{\sigma}) = \hat{\rho} \Leftrightarrow \hat{\sigma} - \hat{\rho} \in \tau\partial\left(\widehat{\mathrm{QKL}}(\cdot,\rho_0)\right)(\hat{\rho}).$$

由于 g^{-1} 的像为正 $\tau g^{-1}\left(\tau^{-1}\hat{\sigma} + \ln\hat{\rho}_0 - \ln\tau\right) =: \tau g^{-1}(\hat{\omega})$ 满秩，因此我们有

$$\begin{aligned}
\tau\partial\left(\widehat{\mathrm{QKL}}(\cdot,\rho_0)\right)(\tau g^{-1}(\hat{\omega})) &= \tau(\ln(\tau g^{-1}(\hat{\omega})) - \ln\hat{\rho}_0) \\
&= \tau(\ln\tau + \ln g^{-1}(\hat{\omega}) + g^{-1}(\hat{\omega}) - g^{-1}(\hat{\omega}) - \ln\hat{\rho}_0) \\
&= \tau(\ln\tau + \hat{\omega} - g^{-1}(\hat{\omega}) - \ln\hat{\rho}_0) \\
&= \tau(\tau^{-1}\hat{\sigma} - g^{-1}(\hat{\omega})) = \hat{\sigma} - \tau g^{-1}(\hat{\omega}).
\end{aligned}$$

□

采用相同的方法，我们也可以计算共轭泛函的梯度，并利用莫罗恒等式得到共轭泛函的近端算子表达式。

引理 3.4. 对于$\hat{\sigma}$ 厄米

$$\partial\left(\widehat{\mathrm{QKL}}(\cdot,\hat{\rho}_0)\right)^*(\hat{\sigma}) = \{\exp(\hat{\sigma} + \ln\hat{\rho}_0)\}$$

and

$$\operatorname{prox}_{\tau(\widehat{\mathrm{QKL}}(\cdot,\rho_0))^*}(\hat{\sigma}) = \hat{\sigma} - W\left(\tau\exp(\hat{\sigma} + \ln\hat{\rho}_0)\right) = \hat{\sigma} - g^{-1}\left(\hat{\sigma} + \ln\hat{\rho}_0 + \ln\tau\right).$$

证明. 共轭泛函只是 $\mathrm{tr}(\exp(\cdot))$ 的平移版本，根据前面提到的理论，其次微分为 $\{\exp(\cdot)\}$。代入平移项即可得到所述公式。对于共轭泛函的近端算子，我们可以利用莫罗恒等式得到

$$\begin{aligned}
\operatorname{prox}_{\tau(\widehat{\mathrm{QKL}}(\cdot,\rho_0))^*}(\hat{\sigma}) &= \hat{\sigma} - \tau\operatorname{prox}_{\tau^{-1}\widehat{\mathrm{QKL}}(\cdot,\rho_0)}(\tau^{-1}\hat{\sigma}) \\
&= \hat{\sigma} - \tau\tau^{-1}W\left(\tau\exp\left(\tau\frac{\hat{\sigma}}{\tau} + \ln\hat{\rho}_0\right)\right) \\
&= \hat{\sigma} - W\left(\tau\exp(\hat{\sigma} + \ln\hat{\rho}_0)\right).
\end{aligned}$$

□

现在我们已经计算出量子相对熵的这些量，可以使用凸优化中的标准算法来求解正则化问题。这种方法的主要缺点是，许多计算都需要通过特征值分解求得矩阵对数，或将函数 g^{-1} 应用于谱上。幸运的是，所需的密度矩阵通常不太大，因此这些计算仍然可行。

如果某个待最小化的泛函$\widehat{}$具有已知的凸性参数 μ且严格凸，那么许多迭代算法都能显著加速。虽然量子相对熵本身并非严格凸，但当矩阵值自变量的特征值有上界时，它就是严格凸的。由于我们通常要寻找特征值有界于 1 的密度矩阵，因此下面的引理很有用：

引理 3.5. $\hat{}$ $QKL(\cdot,\hat{\rho}_0)$ 在集合上以参数 μ 严格凸

$$M_\mu := \{\hat{\rho} \in Herm(N) : eigenvalues\ of\ \hat{\rho}\ not\ bigger\ than\ \mu^{-1}\}.$$

证明。 我们需要证明，在所述集合上，$\hat{\rho} \mapsto \hat{}QKL(\hat{\rho},\ \hat{\rho}_0) - \frac{\mu}{2}\mathrm{tr}(\hat{\rho}^2)$ 是凸的。为此，我们可以忽略关于 $\hat{\rho}$ 的常数项或线性项。因此，只需证明 $\hat{\rho} \mapsto \mathrm{tr}(\hat{\rho}\ln\hat{\rho} - \frac{\mu}{2}\hat{\rho}^2)$ 在 M_μ 上是凸的。由于该函数只依赖于 $\hat{\rho}$ 的特征值，而与它们的顺序无关，因此，当且仅当相应的实值函数$x \mapsto x\ln x - \frac{\mu}{2}x^2$是凸函数时，该函数才是凸的（见 [9]）。计算二阶导数可知，对于 $x \le \frac{1}{\mu}$，确实如此。 –

□

3.3 迭代求解算法

现在我们将介绍一些凸优化算法，它们能够高效且可靠地计算离散正则化问题（4）的解。我们先从使用平方 L^2-范数作为数据保真项的简单情形开始。在这种情况下，我们可以使用 [2], 中提出的 FISTA 版本；该算法此前已针对无正则化的这一问题在 [3] 中提出。与通常的前向后向分裂算法一样，我们交替执行数据保真泛函的梯度步和近端算子的应用。在迭代过程中会改变步长，以加速收敛。由于我们通常重构的密度矩阵具有不大于 1 的$\widehat{\quad}$个特征值，因此在实际中可以假设 QKL 关于参数μ 严格凸，并使用相应算法的更快版本；该版本以指数速度收敛 [7]。在我们考虑的设置中，选择 $\mu = 0.5$ 效果良好；根据引理 3.5，这对应于小于 2 的特征值。

算法 1 广义加速 FISTA

Require: $\mu > 0$; $\tau \in (0, \|T^*T\|^{-1})$; $\rho^{(0)}$;
  $t_0 := 0$; $\rho^{(-1)} := \rho^{(0)}$; $q := \frac{\tau\mu}{1+\tau\mu}$;
  for $l = 0, 1, 2, \ldots$ **do**
    $t_{l+1} = \frac{1}{2}\left(1 - qt_l^2 + \sqrt{(1-qt_l)^2 + 4t_l^2}\right)$
    $\beta_l = \frac{t_l - 1}{t_{l+1}}(1 + (1 - t_{l+1})\tau\mu)$
    $\tilde{\rho}^{(l)} = \rho^{(l)} + \beta_l(\rho^{(l)} - \rho^{(l-1)})$
    $\sigma_l = \tilde{\rho}^{(l)} - \tau T^*(T\tilde{\rho}^{(l)} - g^{\mathrm{obs}})$ ▷ 梯度步
    $\rho^{(l+1)} = \mathrm{prox}_{\alpha\tau\,\mathrm{QKL}(\cdot,\rho_0)}(\sigma_l)$ ▷ 近端步
  end for

如果使用库尔贝克-莱布勒散度，前向后向分裂算法往往会变得数值不稳定，因为当迭代点不在本质定义域内时，可能无法对数据保真泛函执行梯度步；或者当数值接近零时，梯度可能变得极大。作为替代方案，可以采用原始-对偶混合梯度方法，例如 Chambolle-Pock 算法 [6]。在此方法中，我们交替执行数据保真泛函共轭的近端步和惩罚泛函的近端步。通过假设严格凸性，我们可以再次优化参数选择，并显著加快收敛速度，使其达到二次收敛 [7]。

算法 2 加速 Chambolle-Pock 算法

要求： $\mu > 0$; $\alpha > 0$; $\tau_0 > 0$; $\nu_0 > 0$; $\tau_0\nu_0\|T^*T\| \le 1$; $\rho^{(0)}$; $g^{(0)}$;
  $\tilde{\rho}^{(0)} = \rho^{(0)}$
  对于 $l = 0, 1, 2, \ldots$ **执行**
    $\tilde{g}^{(l)} = g^{(l)} + \nu_l T\tilde{\rho}^{(l)}$
    $g^{(l+1)} = \frac{1+\tilde{g}^{(l)}}{2} - \sqrt{\left(\frac{1+\tilde{g}^{(l)}}{2}\right)^2 - \tilde{g}^{(l)} + \nu_l g^{\mathrm{obs}}}$ ▷ 共轭数据保真泛函的近端算子
    $\rho^{(l+1)} = \text{近端算子}_{\tau_l\alpha\text{量子相对熵}(\cdot,\ \rho_0)}(\rho^{(l)} - \tau_l T^* g^{l+1})$ ▷ 惩罚泛函的近端步
    $\beta_l = (1 + 2\mu\tau_l)^{-2}$; $\tau_{l+1} = \beta_l\tau_l$; $\nu_{l+1} = \frac{\nu_l}{\beta_l}$;
    $\tilde{\rho}^{(l+1)} = \rho^{(l+1)} + \beta_l(\rho^{(l+1)} - \rho^{(l)})$
  结束循环

到目前为止，我们尚未讨论何时停止算法。有多种可能的方法，其中最简单的方法之一是用某个范数检查相邻迭代之间的变化，并在该变化小于预定义阈值时停止。更复杂的方法还可以估计当前迭代点与精确极小化子之间的距离，该方法依赖于对偶间隙。

定理 3.6 (对偶间隙)。 设 $\hat{}\rho_\alpha$ 是问题 (4) 的满秩解，则对所有 $\hat{}\ \hat{}\hat{}\rho \in$ 定义域 T和 $\hat{}g \in$ 值域T，有

$$\widehat{\mathrm{QKL}}(\hat{\rho}, \hat{\rho}_\alpha) \leq \frac{1}{\alpha}\widehat{\mathcal{S}}_{g^{obs}}(\hat{T}\hat{\rho}) + \widehat{\mathrm{QKL}}(\hat{\rho}, \hat{\rho}_0) + \frac{1}{\alpha}\widehat{\mathcal{S}}^*_{g^{obs}}(-\alpha\hat{g}) + \left(\widehat{\mathrm{QKL}}(\cdot, \hat{\rho}_0)\right)^* (\hat{T}^*\hat{g}).$$

这一结论的证明基于 Rockafellar-Fenchel 对偶性 [32, 40], ，详见附录。根据引理 3.1，秩假设是保证次微分存在所必需的。

这个不等式的右侧，即对偶间隙，无需知道解 $\hat{}\rho_\alpha$ 便可计算，因此可用于给出当前迭代点 $\hat{}\rho$ 与解之间距离的上界。为使该上界尽可能紧，我们选择 $\hat{}g \in -\frac{1}{\alpha}\partial\widehat{\mathcal{S}}_{g^{\mathrm{obs}}}(T\hat{}\rho)$。如果对偶间隙——进而误差 QKL($\hat{}\rho$, $\hat{}\rho_\alpha$)——小于预设阈值，我们就停止$\widehat{\quad}$迭代。与迄今为止量子层析中使用的其他迭代求解方法相比，这种方法的一项独特优势是能够控制与解之间的距离。

尽管所介绍的两种算法在数值上都相对稳定，但由于特征值分解过程中存在误差，特征值仍可能略小于零，从而导致计算对偶间隙时出现问题。因此，我们在进一步计算前给特征值加上约 10^{-10} 的小偏移量。

4 量子态重构的应用

现在我们将探讨量子态重构的两种不同应用，并展示如何将抽象理论应用于这些场景，同时给出数值示例，以说明我们的方法在实践中的效果。所有实验的代码均可在[24]获取。

4.1 基于 PINEM 的电子断层成像

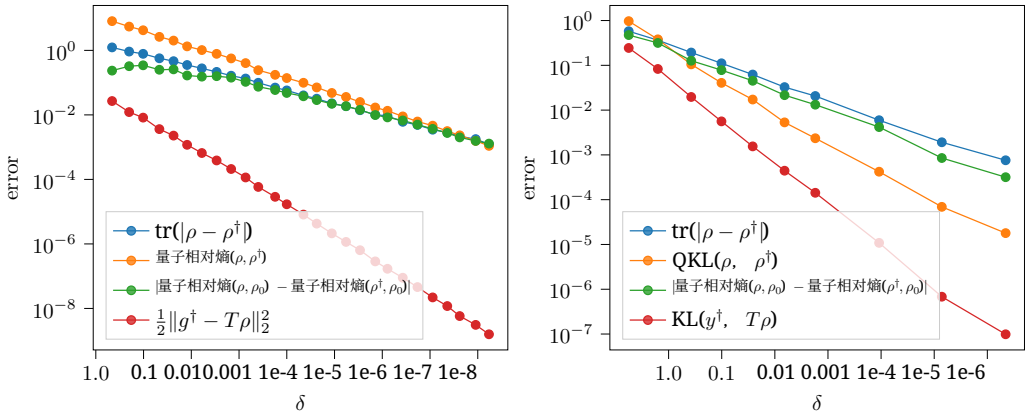
在 [27], 中，研究表明，光子诱导近场电子显微镜（PINEM）能够通过让电子与不同频率的激光光子相互作用，并精确控制它们的相对相位，来重构自由电子束的量子态。针对如何从数据中重构密度矩阵，研究人员提出了一种名为 SQUIRRELS（自由电子态正则化重构的谱量子干涉）的新方法。该方法采用希尔伯特–施密特范数正则化，同时纳入半正定性约束和迹约束。此外， [15] 还提出并比较了贝叶斯方法、最大似然技术和神经网络。

若改用量子相对熵正则化，则无需进一步的约束，并且我们可以通过选择 ρ_0将更多先验知识纳入重构。我们将考虑与 [35]中所分析的相同实验设置。该参考文献中的前向算子$T\colon\ \tilde{\mathcal{B}_1} \to L^2($ $[-\pi, \pi)$ ℓ^1(Z))由下式给出

$$(T\rho)(\theta, l) = \langle e_l, U_\theta \rho U_\theta^* e_l\rangle \quad \text{with } U_\theta e_l = \sum_{k\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}(k-l)\theta} J_{k-l}(2|g|)e_k$$

其中 $\{e_l\}$ 是 ℓ^2(Z)的标准基。算子 U_θ 是酉的，因此引理 2.12适用，并且对于其他基于PINEM的前向算子也同样成立，其中算子 U_θ 略有不同。对于这一具体例子， [35] 中已经证明 T 是单射的。这意味着定理 2.16可直接适用，并给出量子相对熵正则化在该例子中的正则化性质。

对于数值实验，我们采用与 [27]SQUIRRELS 算法补充信息中所述相同的通用设置。这意味着我们根据耦合强度为 $g_{\mathrm{pump}} = 1.73$ 的泵浦相互作用，生成一个大小为 41×41 的真实密度矩阵 $\rho^\dagger$，同时也加入轻微的相位抖动。无噪声数据通过应用前向算子计算得到



(a) L^2-范数　　(b) 库尔贝克-莱布勒散度

图1：不同数据保真泛函的收敛行为

并令 $g = 3g_{\text{pump}}$ 取 θ 个等间距值，由此生成泊松分布合成数据。为了展示我们方法的收敛性，我们采用不同的强度，以获得一系列不同的噪声水平 $\delta = \mathcal{S}_{g^{\text{obs}}}(T\rho^\dagger)$。然后，我们直接令 α 为 $\alpha_0\sqrt{\delta}$，这可保证 $\frac{\delta}{\alpha}$ 趋于零，但对于任何特定噪声水平而言，这很可能都不是最优选择。我们使用平方 L^2 范数和 FISTA，或使用库尔贝克-莱布勒散度和 Chambolle-Pock 算法，如第3节所述。作为参考解，我们直接选取单位矩阵的归一化版本。对于 FISTA，当对偶间隙小于 10^{-6} 时停止；对于 Chambolle-Pock，当对偶间隙小于 10^{-5} 时停止；若迭代次数达到上限 2,000,000 次，也会停止。就我们的目的而言，这两个值都足够小，因为它们远小于无噪声情况下与真实解之间的最终距离。Chambolle-Pock 算法的阈值设得稍高，以补偿其较长的运行时间。只有 Chambolle-Pock 的最后两个数据点受到了迭代次数上限的影响。图1展示了结果；我们可以看到，在这两种情况下，数据和重构密度矩阵都如定理 2.16 所述趋于收敛。

4.2 同相层析成像

同相层析成像是一种成熟的技术，通过使用平衡零拍探测器来重构光量子态。该技术由 Vogel 和 Risken 在 [39] 中提出，随后在 [36] 中通过实验得到验证。测量数据可以看作维格纳函数的拉东变换，而维格纳函数又通过傅里叶变换与密度矩阵相关。因此，最初人们采用了经典层析中的拉东变换逆变换方法（如滤波反投影），根据实验数据重构量子态[19]。另一种逆变换方法是推导模式函数，将其与数据相乘后进行积分，从而直接得到密度矩阵的任意元素[31]。为了保证结果具有物理意义，还可以使用基于最大似然估计的不动点算法 [29]。此外，近年来还采用了凸优化方法，以纳入迹和半正定约束 [38]。

我们将考虑具有完美探测效率的零差层析，其中前向算子 T 可定义为

$$(T\rho)(\theta, x) = \sum_{m,n\in\mathbb{N}} \rho_{m,n} \mathrm{e}^{\mathrm{i}(n-m)\theta} u_m(x) u_n(x) \qquad \text{with } u_m(x) = \frac{1}{\sqrt{\sqrt{\pi} m! 2^m}} H_m(x) \mathrm{e}^{-\frac{x^2}{2}}. \tag{5}$$

Here H_m 表示第 m 个厄米特多项式 [31]。注意，函数 $u_m \mathrm{e}^{\mathrm{i}m\theta}$ 构成 an

L^2(R,C) 的标准正交基，我们可以定义 U_θ: ℓ^2(N, C) $\to L^2$(R, C), $e_n \mapsto \mathrm{e}^{\mathrm{i}n\theta}u_n$ 以及 ℓ^2(N,C) 的标准基 $\{e_n : n \in \mathrm{N}\}$。

利用Gelfand三元组和狄拉克δ函数，人们便可以形式地写出

$$(T\rho)(\theta, x) = \langle \delta_x, U_\theta \rho U_\theta^* \delta_x \rangle$$

从而得到与我们迄今所假设的类似结构。遗憾的是，前向算子并不具备所需的连续性性质，如下引理所示:

Lem引理 4.1. 算子 T 在(5) 中不是弱-$*$ 到弱-$*$ 序列连续的，从 $\mathcal{B}_1$ 到L^2。

证明。 我们仅对单个固定的 θ证明这一点，但该论证可以很容易地推广到 θ的某个区间。取一个有界实值函数 $\tilde{a} \in (L^2(\mathrm{R}) \setminus \{0\})$。则乘法算子 $M_{\tilde{a}}$： L^2(R) $\to L^2$(R)是有界且自伴的，但由于它不恒为0，故不是紧的。于是算子 $U_\theta^* M_{\tilde{a}} U_\theta$ 也不是紧的。

对于 $\mathcal{B}_1$ 上的线性泛函 F，弱-$*$ 连续性等价于单位球上的弱-$*$ 连续性，以及存在元素 $K \in \mathcal{K}$ 使得 $F =$tr$(K\cdot)$ [12, 推论4.46]。因此，线性泛函 $\rho \mapsto$ tr$(U_\theta^* M_{\tilde{a}} U_\theta \rho)$不能是弱-$*$ 连续的，也不能在单位球上是弱-$*$ 连续的。由于 $\mathcal{K}$ 是可分的， $\mathcal{B}_1$ 的单位球在弱-$*$拓扑下是可度量化的，并且该泛函也不是序列弱-$*$ 连续的 [12, 命题3.24]。这意味着存在一个弱-$*$ 收敛序列$(\rho^{(k)})_{k\in\mathbb{N}} \subset \mathcal{B}_1$ 收敛到 ρ，使得tr$(U_\theta^* M_{\tilde{a}} U_\theta \rho^{(k)})$不收敛到tr$(U_\theta^* M_{\tilde{a}} U_\theta \rho)$。现在我们观察到

$$\begin{aligned}
\mathrm{tr}(U_\theta^* M_{\tilde{a}} U_\theta \rho) &= \sum_{m,n} \langle e_m, U_\theta^* M_{\tilde{a}} U_\theta e_n \rangle \langle e_m, \rho e_n \rangle = \sum_{m,n} \langle \mathrm{e}^{\mathrm{i}m\theta} u_m, M_{\tilde{a}} \mathrm{e}^{\mathrm{i}n\theta} u_n \rangle \rho_{m,n} \\
&= \sum_{m,n} \int_{\mathbb{R}} \tilde{a}(x) \mathrm{e}^{-\mathrm{i}m\theta} u_m(x) \mathrm{e}^{\mathrm{i}n\theta} u_n(x) \mathrm{d}x \rho_{m,n} \\
&= \int_{\mathbb{R}} \tilde{a}(x) \sum_{m,n} \rho_{m,n} \mathrm{e}^{-\mathrm{i}(n-m)\theta} u_m(x) u_n(x) \mathrm{d}x = \langle \tilde{a}, (T\rho)(\theta) \rangle_{L^2(\mathbb{R})}.
\end{aligned}$$

因此，$(T\rho^{(k)})(\theta)$不弱-$*$收敛到$(T\rho)(\theta)$。 □

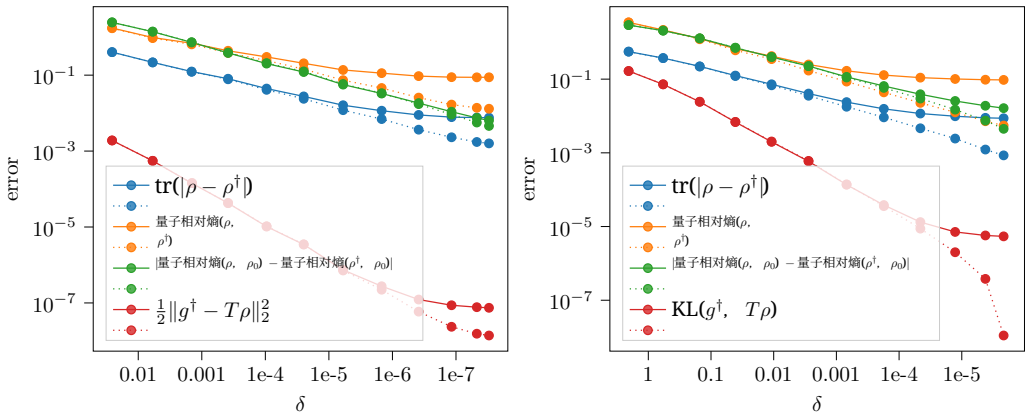
注4.2。 这个问题也无法通过考虑半离散版本，并通过定义 T_{semi}： $\tilde{\mathcal{B}}_1 \to L^2([0,\pi)\ \ell^1(\mathbb{Z}))$转而采用序列空间来解决。

$$(T_{\mathrm{semi}}\rho)(\theta, l) := \int_{lh}^{(l+1)h} (T\rho)(\theta, x) \mathrm{d}x$$

对于某个 $h > 0$。其证明实质上相同，但现在考虑的不是 $\tilde{a} \in L^2(\mathbb{R})$ ，而是对一个序列 $a \in c_0(\mathbb{Z})$ 考察函数 $\tilde{a}(x) := \sum_l a_l 1_{[lh,(l+1)h)}(x)$；它同样不会生成紧乘法算子，并满足 $\langle \tilde{a}, (T\rho)(\theta) \rangle_{L^2(\mathbb{R})} = h\langle a, (T_{\mathrm{semi}}\rho)(\theta) \rangle_{\ell^1(\mathbb{Z})}$。

注4.3。 据我们所知，关于该算子是否在范数拓扑下连续的问题似乎仍未解决。它与拉东变换*[39]* 在 $L^2(\mathbb{R}^2)$ 上的关系表明，该算子在 $\mathcal{B}_2$ 上是单射 *[22]*，但以 L^2 为陪域时并非有界*[37]*。

仍能应用我们理论的一种方法，是对矩阵元$\rho_{m、n}$ 使用衰减条件，并应用引理 2.15。例如，类似于 [1], ，我们可以假设 ρ 属于一个加权 $\mathcal{B}_2$ 空间，其范数为$\|\rho\|$ ^$_{\mathcal{B}} :=$ $\|(\rho_{m,n}$ exp$(B(m+n)^{r/2}))_{m,n}\|_{\mathcal{B}_2}$，其中 $B > 0$ 以及 $r \in$(0, 2)取某些值，并将^ ^ dom T定义为 $\mathcal{B}$中的一个闭球。由于 $\mathcal{B}$是可分希尔伯特空间，dom T是弱序列紧的，并且由 $\|\rho\|_{\mathcal{B}_1} =$ sup$\{|$tr$(\sigma\rho)| : \|\sigma\|_{\mathcal{B}} \le 1, \sigma \in \mathcal{K}\} \le \sum_{m,n} |\rho_{m,n}|$，根据^柯西-施瓦茨不等式可知 $\mathcal{B}$ 连续嵌入于 $\mathcal{B}_1$。为证明 T在(5)中的连续性，我们可以按照 [1] 中的方法进行，并利用 $T\rho$ 是维格纳



(a) L^2-范数　　(b) 库尔贝克-莱布勒散度

图 2：不同数据保真泛函的收敛行为。虚线表示采用精确前向算子时的结果。

函数 W_ρ，它满足$\|W_\rho\|^2_{L^2(\mathbb{R}^2)} = \|\rho\|^2_{\mathcal{B}_2}/(2\pi)$ [1, 式(5)]，并表现出由$\sup_{m,n} |\rho_{m,n}|\exp(B'(m+n)^{r/2})$ [1, 命题 1]所控制的指数衰减。对于 $B' < B$，后一范数可利用柯西-施瓦茨不等式由 $\|\rho\|_{\hat{\mathcal{B}}}$ 界定。

原则上，对于噪声水平趋于零的一系列数据，必须在数值上对 $\|\rho\|_{\hat{\mathcal{B}}}$ 施加统一的硬约束，才能保证相应估计量收敛，但这在数值计算上并不方便。

在我们的框架中，为了获得良定义的算子，同时避免对 ρ作出强先验假设，另一种方法是用任意标准正交基 $\{v_l, l \in \mathbb{Z}\}$ 替换狄拉克δ函数，该基属于 $L^2(\mathbb{R}, \mathbb{C})$，并得到 $\tilde{T}\colon \tilde{\mathcal{B}}_1 \to L^2([0,\pi)\ \ell^1(\mathbb{Z}))$，其形式为

$$(\tilde{T}\rho)(\theta, l) = \langle v_l, U_\theta \rho U_\theta^* v_l \rangle.$$

注意，这涉及对 $U_\theta \rho U_\theta^*$ 的施瓦茨核的平方进行积分，而不仅仅是沿对角线积分。为了仍然拟合测量数据， v_l 的选择应当具有良好的局部性。例如，可以使用适当缩放的哈尔小波基，最终得到一个接近 T_{semi}的算子，尤其是在转入箱宽足够小的完全离散设定时。

经过这些修改后，我们可以很容易地看出，$\tilde{T}$的单射性取决于实际基 v_l 的选择，但正如我们之前所述，即使在非单射情形下，我们仍至少对某个子列得到定理 2.16 的结果。

作为一个例子，我们考虑重构薛定谔的猫态¨，其中 $a = 3$ 是两个相干态 $|a\rangle + |-a\rangle$的叠加。这是展现非经典行为 [19]的状态的一个标准例子。我们考虑一个大小为 21×21 的密度矩阵，使用 60 个不同的相位，并选择 v_l 为区间$[-5 + \frac{l}{12}, -5 + \frac{l+1}{12}]$ 的示性函数，其中$0 \le l < 120$，以覆盖包含相关数据部分的区间 $[-5, 5]$ 。我们使用算子 T_{semi}生成泊松分布合成数据。停止准则 δ 和 α 的选择方式与 PINEM 示例相同。对迭代次数的限制，同样只对最后三个采用库尔贝克-莱布勒散度作为数据保真泛函的数据点有影响。我们使用不同的数据保真泛函˜进行重构，并分别采用 T_{semi} 或 T，对结果进行比较。图 2 显示，当噪声不太小时，两种算子得到的结果非常相似。然而，正如预期的那样，随着噪声减小到一定程度，两者的差异会变得显著；只有使用 T_{semi} 才能继续收敛。

5 结论

作为一项主要的理论结果，我们证明了量子相对熵泛函作为变分正则化方法中惩罚项的正则化性质。与一些竞争方法不同，它不依赖于基的选择。我们还在有限维情形下确立了量子相对熵的所有性质，这些性质是应用凸优化算法所必需的。对于描述 PINEM 实验的前向算子，我们验证了设定中的所有假设，并通过数值实验展示了相应密度矩阵估计量在噪声水平趋于 0 时的稳定性和收敛性。对于同相层析成像，自然的精确前向算子不符合我们的框架，但适当的半离散近似仍可得出部分结果。在实践中，使用精确前向算子效果良好，量子相对熵正则化能够可靠地重构出正确的解。

本研究提出了若干新的、且可能富有成果的研究方向：一个自然的问题是，泊松数据的统计一致性是否可以类比于确定性噪声模型中的正则化性质。这似乎需要对前向算子和/或可能解的集合作出更严格的假设。另一个潜在的研究方向是，在密度矩阵元满足衰减条件的情况下，量子相对熵惩罚估计量的收敛速率；关于这一点， [1]中给出了同相层析成像投影估计量的结果， [35]中给出了 PINEM 的 SQUIRRELS 估计量的结果。此外，建立一个不对密度矩阵集合施加人为约束的同相层析成像恰当泛函分析框架也颇具意义；进一步研究其在其他层析成像场景中的应用同样如此，例如探测效率不完美的同相层析成像（见 [19, 第 5 节3]）或外差层析（见 [30]）。

最后，我们强调，半正定矩阵和算子广泛存在于量子态重构之外的许多场景中，例如协方差算子。本文的结果为反问题提供了一个全新的通用正则化框架，其中未知量是半正定算子。

数据可用性声明

与本文相关的代码可在‘GRO.data’中获取，参见 https://doi.org/10.25625/YMGX29。

A 附录

定理 A.1（一般对偶间隙）。 设 X、Y 为巴拿赫空间， $T \in \mathcal{L}(X、Y)$，并设 $\mathcal{R}\colon X \to \mathbb{R} \cup \{\infty\}$ 与 $\mathcal{S}\colon Y \to \mathbb{R} \cup \{\infty\}$ 为凸泛函。此外，对 $\alpha > 0$ 令

$$f^\dagger \in \operatorname*{argmin}_f \frac{1}{\alpha}\mathcal{S}(Tf) + \mathcal{R}(f)$$

使得 $\partial\mathcal{R}(f^\dagger)$与 $\partial\mathcal{S}(Tf^\dagger)$非空。则对所有 $f \in X$、 $p \in Y^*$ 以及 $g \in \partial\mathcal{R}(f^\dagger)$，不等式

$$\Delta^g_{\mathcal{R}}(f, f^\dagger) \leq \frac{1}{\alpha}\mathcal{S}(Tf) + \mathcal{R}(f) + \mathcal{R}^*(T^*p) + \frac{1}{\alpha}\mathcal{S}^*(-\alpha p)$$

成立。

证明。 将原始和对偶目标泛函定义为

$$\mathcal{J}(f) = \frac{1}{\alpha}\mathcal{S}(Tf) + \mathcal{R}(f), \qquad \mathcal{J}_*(p) = \mathcal{R}^*(T^*p) + \frac{1}{\alpha}\mathcal{S}^*(-\alpha p)$$

并回顾对所有 $f \in X$ 和 $p \in Y^*$， $\mathcal{J}(f) \geq -\mathcal{J}_*(p)$，根据弱 Rockafellar-Fenchel 对偶性，参见 [32]。那么对所有 f， p，由杨氏不等式可得

$$\begin{aligned}\mathcal{J}(f) + \mathcal{J}_*(p) &= \frac{1}{\alpha}\mathcal{S}(Tf) + \mathcal{R}(f) + \mathcal{R}^*(T^*p) + \frac{1}{\alpha}\mathcal{S}^*(-\alpha p) \\ &\geq \frac{1}{\alpha}\langle Tf, -\alpha p\rangle + \langle f, T^*p\rangle = 0\end{aligned}$$

现在假设 $f^\dagger$ 是该优化问题的一个解。那么

$$0 \in \partial J(f^\dagger) = \partial\mathcal{R}(f^\dagger) + \frac{1}{\alpha}T^*\partial\mathcal{S}(Tf^\dagger)$$

那么，根据我们的假设，存在 $g \in \partial\mathcal{R}(f^\dagger)$ 和 $q \in \partial\mathcal{S}(Tf^\dagger)$，使得 $-\frac{1}{\alpha}T^*q = g$这意味着

$$\begin{aligned}&\Delta^g_{\mathcal{R}}(f, f^\dagger) + \frac{1}{\alpha}\Delta^q_{\mathcal{S}}(Tf, Tf^\dagger) \\ =&\mathcal{R}(f) - \mathcal{R}(f^\dagger) - \langle g, f - f^\dagger\rangle + \frac{1}{\alpha}\left(\mathcal{S}(Tf) - \mathcal{S}(Tf^\dagger) - \langle q, Tf - Tf^\dagger\rangle\right) \\ =&\mathcal{R}(f) - \mathcal{R}(f^\dagger) - \langle g, f - f^\dagger\rangle + \frac{1}{\alpha}\left(\mathcal{S}(Tf) - \mathcal{S}(Tf^\dagger)\right) - \frac{1}{\alpha}\langle T^*q, f - f^\dagger\rangle \\ =&\mathcal{J}(f) - \mathcal{J}(f^\dagger).\end{aligned}$$

现在，由于布雷格曼散度非负，我们得到，对于所有 f， p

$$\Delta^g_{\mathcal{R}}(f, f^\dagger) \leq \Delta^g_{\mathcal{R}}(f, f^\dagger) + \frac{1}{\alpha}\Delta^q_{\mathcal{S}}(Tf, Tf^\dagger) \leq \mathcal{J}(f) - \mathcal{J}(f^\dagger) \leq \mathcal{J}(f) + \mathcal{J}_*(p)$$

在最后一个不等式中使用了弱对偶。 □

参考文献

[1] J.-M. Aubry、C. Butucea 和 K. Meziani。含噪数据下量子同频层析中的状态估计。《反问题》， 25(1):015003，2008。[2] A. Beck 和 M. Teboulle。用于线性反问题的快速迭代收缩阈值算法。《*SIAM* 成像科学杂志》， 2(1):183–202，2009。[3] E. Bolduc、G. C. Knee、E. M. Gauger 和 J. Leach。量子态层析的投影梯度下降算法。《*npj* 量子信息》， 3(1):44，2017。[4] J. M. Borwein 和 A. S. Lewis。最佳熵估计的收敛性。《*SIAM* 优化杂志》， 1(2):191–205，1991。[5] L. G. Brown 和 H. Kosaki。半有限冯·诺伊曼代数中的詹森不等式。《算子理论杂志》，第 3–19 页，1990。[6] A. Chambolle 和 T. Pock。一种用于凸问题的一阶原始-对偶算法及其在成像中的应用。《数学成像与视觉杂志》， 40:120–145，2011。[7] A. Chambolle 和 T. Pock。成像连续优化导论。《数值学报》， 25:161–319，2016。[8] G. M. D'Ariano、M. G. Paris 和 M. F. Sacchi。量子层析。《成像与电子物理进展》， 128(10.1016):S1076–5670，2003。[9] C. Davis。厄米矩阵的所有凸不变函数。《数学档案》， 8(4):276–278，1957。

[10] Z. Duan 和 J. Hou。无限维系统的熵交换。《科学报告》，7(1):41692，2017。[11] P. P. Eggermont。第一类弗雷德霍姆积分方程的最大熵正则化。《SIAM 数学分析杂志》，24(6):1557–1576，1993。[12] M. Fabian、P. Habala、P. H´ajek、V. Montesinos Santaluc´ıa、J. Pelant 和 V. Zizler。《泛函分析与无限维几何》，CMS Books Math./Ouvrages Math. SMC 第 8 卷。纽约：施普林格，2001。[13] F. Hiai、M. Ohya 和 M. Tsukada。冯·诺伊曼代数中的充分性、KMS 条件与相对熵。《太平洋数学杂志》，96(1):99–109，1981。[14] T. Hyt¨onen、J. Van Neerven、M. Veraar 和 L. Weis。《巴拿赫空间分析》，第 12 卷。施普林格，2016。[15] H. Jeng 和 C. Ropers。自由电子量子态的断层重建，2025。[16] V. Koltchinskii。冯·诺依曼熵惩罚与低秩矩阵估计。《统计年刊》，2011。[17] L. Lami 和 M. E. Shirokov。纠缠相对熵的可达性与下半连续性及相关主题。在《亨利·庞加莱年刊》，第 24 卷，第 4069–4137 页。施普林格，2023。[18] O. E. I. Lanford 和 D. W. Robinson。量子统计力学中态的平均熵。《数学物理杂志》，9:1120–1125，1968。[19] U. Leonhardt 和 H. Paul。测量光量子态。《量子电子学进展》，19(2):89–130，1995。[20] A. S. Lewis。厄米矩阵上的凸分析。《SIAM 优化杂志》，6(1):164–177，1996。[21] G. Lindblad。有限量子系统的期望值与熵不等式。《数学物理通讯》，39:111–119，1974。[22] D. Ludwig。欧几里得空间上的拉东变换。《纯粹与应用数学通讯》，19(1):49–81，1966。[23] A. I. Lvovsky 和 M. G. Raymer。连续变量光学量子态层析。《现代物理评论》，81(1):299–332，2009。[24] F. Oberender。《复现数据：量子态层析的量子相对熵正则化》。https://doi.org/10.25625/YMGX29，2026。[25] S. Olivares 和 M. G. Paris。基于最小 Kullback 熵原理的量子估计。《物理评论 A：原子、分子与光物理》，76(4):042120，2007。[26] C. P¨oschl。具有一般残差项的吉洪诺夫正则化。na，2008。[27] K. E. Priebe、C. Rathje、S. V. Yalunin、T. Hohage、A. Feist、S. Sch¨afer 和 C. Ropers。超快透射电子显微镜中的阿秒电子脉冲列与量子态重构。《自然·光子学》，11(12):793–797，2017。[28] M. Reed 和 B. Simon。I：泛函分析，第 1 卷。学术出版社，1981。[29] J. ˇReh´aˇcek、Z. Hradil、E. Knill 和 A. Lvovsky。用于量子层析的稀释最大似然算法。《物理评论 A：原子、分子与光物理》，75(4):042108，2007。

[30] T. Richter. 通过双路同相检测测量确定光子统计和密度矩阵。《现代光学杂志》，45(8):1735–1749，1998.[31] T. Richter. 用于光学同相层析和特定期望值确定的真实模式函数。《物理评论*A*》，61(6):063819，2000.[32] R. Rockafellar. 涉及凸函数的极值问题中的对偶性与稳定性。《太平洋数学杂志》，21(1):167–187，1967.[33] W. Rudin. 泛函分析。纽约：McGraw-Hill，第2版，1991.[34] T. I. Seidman and C. R. Vogel. 某些非线性病态问题正则化方法的适定性与收敛。《反问题》，5(2):227，1989.[35] C. Shi, C. Ropers, and T. Hohage. 超快透射电子显微镜中的密度矩阵重构：唯一性、稳定性和收敛速率。《反问题》，36(2):025005，2020.[36] D. Smithey, M. Beck, J. Cooper, and M. Raymer. 光场数相不确定性关系的测量。《物理评论*A*》，48(4):3159，1993.[37] D. C. Solmon. X射线变换。*J. Math. Anal. Appl.*，56:61–83，1976.[38] I. Strandberg. 通过凸优化实现简单、可靠且抗噪的连续变量量子态层析成像。*Physical Review Applied*，18(4):044041，2022.[39] K. Vogel and H. Risken. 根据旋转正交相位的概率分布确定准概率分布。《物理评论*A*》，40(5):2847，1989.[40] E. Zeidler. 非线性泛函分析及其应用：*III*：变分方法与优化。Springer Science & Business Media，2013。