Journal of Statistical Planning and Inference 184 (2017) 62–76

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Pseudo-Bayesian quantum tomography with rank-adaptation

The Tien Mai *, Pierre Alquier

CREST, ENSAE, Université Paris Saclay, 3 av. Pierre Larousse, 92245 Malakoff CEDEX, France

ARTICLE INFO

Article history:
Received 31 May 2016
Received in revised form 10 October 2016
Accepted 24 November 2016
Available online 10 December 2016

Keywords:
Quantum statistics
Bayesian statistics
PAC-Bayesian bounds
Oracle inequalities
MCMC

ABSTRACT

Quantum state tomography, an important task in quantum information processing, aims at reconstructing a state from prepared measurement data. Bayesian methods are recognized to be one of the good and reliable choices in estimating quantum states (Blume-Kohout, 2010). Several numerical works showed that Bayesian estimations are comparable to, and even better than other methods in the problem of 1-qubit state recovery. However, the problem of choosing prior distribution in the general case of n qubits is not straightforward. More importantly, the statistical performance of Bayesian type estimators has not been studied from a theoretical perspective yet. In this paper, we propose a novel prior for quantum states (density matrices), and we define pseudo-Bayesian estimators of the density matrix. Then, using PAC-Bayesian theorems (Catoni, 2007), we derive rates of convergence for the posterior mean. The numerical performance of these estimators is tested on simulated and real datasets.

1. Introduction

Playing a vital role in quantum information processing, as well as being fundamental for characterizing quantum objects, quantum state tomography focuses on reconstructing the (unknown) state of a physical quantum system (Paris and Řeháček, 2004), usually represented by the so-called density matrix ρ (the exact definition of a density matrix is given in Section 2). This task is done by using outcomes of measurements performed on many independent systems identically prepared in the same state.

The 'tomographic' method, also named as linear/direct inversion (Vogel and Risken, 1989; Řeháček et al., 2010), is the simplest and oldest estimation procedure. It is actually the analogous of the least-square estimator in the quantum setting. Although easy in computation and providing unbiased estimate (Schwemmer et al., 2015), it does not generate a physical density matrix as an output (Shang et al., 2014). Maximum likelihood estimation (Hradil et al., 2004) is the current procedure of choice. Unfortunately, it has some critical flaws detailed in Blume-Kohout (2010), including a huge computational complexity. Furthermore, both these methods are not adaptive to the case where a system is in a state ρ for which some additional information is available. Note especially that, physicists focus on so-called pure states, for which $\mathrm{rank}(\rho) = 1$.

The problem of rank-adaptivity was tackled thanks to adequate penalization. Rank-penalized maximum likelihood (BIC) was introduced in Guţă et al. (2012) while a rank-penalized least-square estimator $\hat{\rho}_{\text{rank-pen}}$ was proposed in Alquier et al. (2013), together with a proof of its consistency. More specifically, when the density matrix of the system is ρ^0 with $r = \mathrm{rank}(\rho^0)$, the authors of Alquier et al. (2013) proved that the Frobenius norm of the estimation error satisfies $\|\hat{\rho}_{\text{rank-pen}} - \rho^0\|_F^2 = \mathcal{O}(r4^n/N)$ where N is the number of quantum measurements. The rate was improved to $\mathcal{O}(r3^n/N)$ by

* Corresponding author.
E-mail addresses: thetien.mai@ensae.fr (T.T. Mai), pierre.alquier@ensae.fr (P. Alquier).

http://dx.doi.org/10.1016/j.jspi.2016.11.003
0378-3758/

Butucea et al. (2015), using a thresholding method. Note that the rate $\mathcal{O}(r2^n/N)$ was first claimed in the paper, but in the Corrigendum (Butucea et al., 2016), the authors acknowledge that this is not the case. The paper however contains a proof that no method can reach a rate smaller than $r2^n/N$. So, the minimax-optimal rate is somewhere in between $r2^n/N$ and $r3^n/N$.

Note that all the aforementioned papers only cover the complete measurement case (the definition is given in Section 2, basically it means that we have observations for all the observables given by the Pauli basis). The statistical relationship between matrix completion and quantum tomography with incomplete measurements (in the Le Cam paradigm) has been investigated in Wang (2013). Thus compressed sensing ideas have been successfully proposed in estimating a density state from incomplete measurements (Gross et al., 2010; Gross, 2011; Flammia et al., 2012; Koltchinskii, 2011).

On the other hand, Bayesian estimation has been considered in this context. The papers (Bužek et al., 1998; Baier et al., 2007) compare Bayesian methods to other methods on simulated data. More recently, Kravtsov et al. (2013), Ferrie (2014), Kueng and Ferrie (2015) and Schmied (2016) discuss efficient algorithms for computing Bayesian estimators. Importantly, Blume-Kohout (2010) showed that Bayesian method comes with natural error bars and is the most accurate scheme w.r.t. the expected error (operational divergence) (even) with finite samples. However, there is no theoretical guarantee on the convergence of these estimators.

More works on quantum state tomography in various settings include Audenaert and Scheel (2009), Carlen (2010), Rau (2011), Rau (2014) and Ferrie and Granade (2014).

In this paper, we consider a pseudo-Bayesian estimation, where the likelihood is replaced by pseudo-likelihoods based on various moments (two estimators, corresponding to two different pseudo-likelihood, are actually proposed). Using PAC-Bayesian theory (Shawe-Taylor and Williamson, 1997; McAllester, 1998; Catoni, 2004, 2007; Dalalyan and Tsybakov, 2008; Suzuki, 2012), we derive oracle inequalities for the pseudo-posterior mean. We obtain rates of convergence for these estimators in the complete measurement setting. One of them has a rate as good as the best known rate up to date $\mathcal{O}(\text{rank}(\rho^0)3^n/N)$ (still, the other one is interesting for computational reasons that are discussed in the paper).

The rest of the paper is organized as follows. We recall the standard notations and basics about quantum theory in Section 2. Then the definition of the prior and of the estimators are presented in Section 3. The statistical analysis of the estimators is given in Section 4, while all the proofs are delayed to the Appendix. Some numerical experiments on simulated and real datasets are given in Section 5.

2. Preliminaries

2.1. Notations

A very good introduction to the notations and problems of quantum statistics is given in Artiles et al. (2005). Here, we only provide the basic definitions required for the paper.

In quantum physics, all the information on the physical state of a system can be encoded in its *density matrix* ρ. Depending on the system in hand, this matrix can have a finite or infinite number of entries. A two-level system of n-qubits is represented by a $2^n \times 2^n$ density matrix ρ, with coefficients in $\mathbb{C}$. For the sake of simplicity, the notation $d = 2^n$ is used in Butucea et al. (2015), so note that ρ is a $d \times d$ matrix. This matrix is Hermitian $\rho^\dagger = \rho$ (i.e. self-adjoint), semidefinite positive $\rho \geq 0$ and has $\text{Trace}(\rho) = 1$. Additionally, it often makes sense to assume that the rank of ρ is small (Gross et al., 2010; Gross, 2011). In theory, the rank can be any integer between 1 and 2^n, but physicists are especially interested in pure states and a pure state ρ can be defined by $\text{rank}(\rho) = 1$.

The objective of quantum tomography is to estimate ρ on the basis of experimental observations of many independent and identical systems prepared in the state ρ by the same experimental device.

For each particle (qubit), one can measure one of the three Pauli observables $\sigma_x,\ \sigma_y,\ \sigma_z$. The outcome for each will be 1, or -1, randomly (the corresponding probability depends on the state ρ and will be given in (1)). Thus for a n-qubits system, we consider 3^n possible experimental observables. The set of all possible performed observables is

$$\{\sigma_{\mathbf{a}} = \sigma_{a_1} \otimes \cdots \otimes \sigma_{a_n};\ \mathbf{a} = (a_1, \dots, a_n) \in \mathcal{E}^n := \{x, y, z\}^n\},$$

where vector $\mathbf{a}$ identifies the experiment. The outcome for each fixed observable setting will be a random vector $\mathbf{s} = (s_1, \dots, s_n) \in \mathcal{R}^n := \{-1, 1\}^n$, thus there are 2^n outcomes in total.

Let us denote $R^{\mathbf{a}}$ a $\mathcal{R}^n$-valued random vector that is the outcome of an experiment indexed by $\mathbf{a}$. From the basic principles of quantum mechanics (Born's rule), its probability distribution is given by

$$\forall \mathbf{s} \in \mathcal{R}^n, \quad p_{\mathbf{a},\mathbf{s}} := \mathbb{P}(R^{\mathbf{a}} = \mathbf{s}) = \text{Trace}\left(\rho \cdot P_{\mathbf{s}}^{\mathbf{a}}\right), \tag{1}$$

where $P_{\mathbf{s}}^{\mathbf{a}} := P_{s_1}^{a_1} \otimes \cdots \otimes P_{s_n}^{a_n}$ and $P_{s_i}^{a_i}$ is the orthogonal projection associated to the eigenvalue s_i in the diagonalization of σ_{a_i} for $a_i \in \{x, y, z\}$ and $s_i \in \{-1, 1\}$ – that is $\sigma_{a_i} = -1P_{-1}^{a_i} + 1P_{+1}^{a_i}$.

The quantum state tomography problem is as follows: a physicist has access to an experimental device that produces n-qubits in a state ρ^0, and ρ^0 is assumed to be unknown. He/she can produce a large number of replications of the n-qubits and wants to infer ρ^0 from this.

In the complete measurement case, for *each* experiment setting $\mathbf{a} \in \mathcal{E}^n$, the experimenter repeats m times the experiment corresponding to $\mathbf{a}$ and thus collects m independent random copies of $R^{\mathbf{a}}$, say $R_1^{\mathbf{a}}, \ldots, R_m^{\mathbf{a}}$. As there are 3^n possible experiment settings $\mathbf{a}$, we define the *quantum sample* size as $N := m \cdot 3^n$. We will refer to $(R_i^{\mathbf{a}})_{i\in\{1,\ldots,m\},\mathbf{a}\in\mathcal{E}^n}$ as $\mathcal{D}$ (for data).

Note that the case where we would only have access to experiments $\mathbf{a} \in \mathcal{A}$ where $\mathcal{A}$ is some proper subset of $\mathcal{E}^n$ ($A \subsetneq \mathcal{E}^n$) is referred to as the incomplete measurement case. In this paper, we focus on the complete measurement case, but the extension to the incomplete case is discussed in Section 6.

2.2. Popular estimation methods

A natural idea is to define the empirical frequencies

$$\hat{p}_{\mathbf{a},\mathbf{s}} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}_{\{R_i^{\mathbf{a}}=\mathbf{s}\}}.$$

Note that $\hat{p}_{\mathbf{a},\mathbf{s}}$ is an unbiased estimator of the probability $p_{\mathbf{a},\mathbf{s}}$. The inversion method is based on solving the linear system of equations

$$\begin{cases} \hat{p}_{\mathbf{a},\mathbf{s}} = \text{Trace}\left(\hat{\rho}\cdot P_{\mathbf{s}}^{\mathbf{a}}\right), \\ \mathbf{a}\in\mathcal{E}^n, \\ \mathbf{s}\in\mathcal{R}^n. \end{cases} \tag{2}$$

As mentioned above, the computation of $\hat{\rho}$ is quite straightforward. Explicit formulas are classical, see e.g. Alquier et al. (2013).

Another commonly used method is maximum likelihood (ML) estimation, where the likelihood is

$$\mathcal{L}(\rho;\mathcal{D}) \propto \prod_{\mathbf{a}\in\mathcal{E}^n}\prod_{\mathbf{s}\in\mathcal{R}^n}[\text{Trace}\left(\rho\cdot P_{\mathbf{s}}^{\mathbf{a}}\right)]^{n_{\mathbf{a},\mathbf{s}}},$$

where $n_{\mathbf{a},\mathbf{s}} = m\hat{p}_{\mathbf{a},\mathbf{s}}$ is the number of times we observed output $\mathbf{s}$ in experiment $\mathbf{a}$ (obviously, $\sum_{\mathbf{s}} n_{\mathbf{a},\mathbf{s}} = m$). As mentioned in the introduction, both methods suffer many drawbacks. The inversion method returns a matrix $\hat{\rho}$ that usually does not satisfy the axioms of a density matrix. ML becomes expensive (impractical) for $n \geq 10$. Moreover, these two methods cannot take advantage of a prior knowledge (e.x. low-rank state).

Considering the expansion of the density matrix ρ in the n–Pauli basis, i.e. $\mathcal{B} = \{\sigma_b = \sigma_{b_1}\otimes\cdots\otimes\sigma_{b_n}, b \in \{I,x,y,z\}^n\}$, $\sigma_I = I$,

$$\rho = \sum_{b\in\{I,x,y,z\}^n}\rho_b\sigma_b. \tag{3}$$

One can also estimate the density matrix via estimating the coefficients in the Pauli expansion. This was studied in Cai et al. (2016) where the authors also make a sparsity assumption: that is, most of ρ_b are small or very close to 0. Note that, this is not related to the setting we explore (low-rank assumption).

We now turn to the definition of a prior distribution on density matrices that will allow to perform (pseudo-)Bayesian estimation.

3. Pseudo-Bayesian estimation and prior distribution on density matrices

3.1. Pseudo-Bayesian estimation

We remind that the idea of Bayesian statistics is to encode the prior information on density matrices through a prior distribution $\pi(d\rho)$. Inference is then done through the posterior distribution $\pi(d\rho|\mathcal{D}) \propto \mathcal{L}(\rho)\pi(d\rho)$. Here, for computational reasons, we replace the likelihood by a pseudo-likelihood. This is an increasingly popular method in Bayesian statistics (Bissiri et al., 2016) and in machine learning (Catoni, 2007; Alquier et al., 2015; Bégin et al., 2016). We define the pseudo-posterior by

$$\tilde{\pi}_\lambda(d\nu) \propto \exp\left[-\lambda\ell(\nu,\mathcal{D})\right]\pi(d\nu), \tag{4}$$

the pseudo-likelihood being $\exp\left[-\lambda\ell(\nu,\mathcal{D})\right]$. The term $\ell(\nu,\mathcal{D})$ can be specified by the user. Two examples are provided in Section 4. As a replacement of the likelihood, this term plays the role of the empirical evidence. More specially

- the role of $\exp\left[-\lambda\ell(\nu,\mathcal{D})\right]$ is to give more weight to the density ν when it fits the data well;
- the role of $\pi(d\nu)$, the prior, is to restrict the posterior to the space of densities (and even give more weight to low-rank matrices if needed);
- $\lambda > 0$ is a free parameter that allows to tune the balance between evidence from the data and prior information.

We finally define the pseudo-posterior mean (also referred to as Gibbs estimator, PAC-Bayesian estimator or EWA, for exponentially weighted aggregate (Catoni, 2007; Dalalyan and Tsybakov, 2008)):

$$\tilde{\rho}_\lambda = \int \nu \tilde{\pi}_\lambda(d\nu).$$

The definition of the estimator $\tilde{\rho}_\lambda$ based on the pseudo-posterior $\tilde{\pi}_\lambda$ is actually validated by the theoretical results from Section 4.

3.2. Definition of the prior

In the single qubit state estimation $n = 1$, the representation of the quantum constraints is explicit (Baier et al., 2007; Schmied, 2016). Thus, one can place a prior distribution on the polar reparametrization of the density. Up to our knowledge, this has not been extended to the case $n > 1$, and this extension seems not straightforward. For general n-qubit densities, uninformative priors (e.g the Haar measure) are put on $\psi_{d\times K}$ matrices ($K \geq d$) and the density state is built by $\rho = \psi_{d\times K}\psi^\dagger_{d\times K}$ (Struchalin et al., 2016; Granade et al., 2016; Huszár and Houlsby, 2012; Kueng and Ferrie, 2015; Życzkowski et al., 2011). One could also define a prior on the coefficients $\{\rho_b\}$ of ρ on the Pauli basis. Nevertheless, none of these approaches seem helpful for rank adaptation.

The idea for our prior is inspired by the priors used for low-rank matrix estimation in machine learning, e.g. Mai and Alquier (2015) and Cottet and Alquier (2016) and the references therein. Hereafter, we describe in details the prior construction.

Let V be a vector in $\mathbb{C}^{d\times 1} \setminus \{\mathbf{0}\}$ ($d = 2^n$ in our model), then $VV^\dagger$ is a Hermitian, semi-definite positive matrix in $\mathbb{C}^{d\times d}$ with $\operatorname{rank}(VV^\dagger) = 1$. Additionally, we can normalize V (that is replace V by $V/\|V\|$), this lead to $\operatorname{Trace}(VV^\dagger) = 1$. So, $VV^\dagger$ satisfies the conditions of a density matrix (with rank-1).

Now, let $V_1, \dots, V_d$ be d normalized vectors in $\mathbb{C}^{d\times 1} \setminus \{\mathbf{0}\}$ and $\gamma_1, \dots, \gamma_d$ be non-negative weights with $\sum_{j=1}^d \gamma_j = 1$. Put

$$\nu = \sum_{i=1}^d \gamma_i V_i V_i^\dagger. \tag{5}$$

Then ν is clearly a density matrix: it is Hermitian (as a sum of Hermitian matrices), it is semi-definite positive (same reason) and

$$\operatorname{Tr}(\nu) = \sum_{i=1}^d \gamma_i \operatorname{Tr}(V_i V_i^\dagger) = 1.$$

Moreover, note that any density matrix can be written in such way, as we know that for any density matrix ρ,

$$\rho = U\Lambda U^\dagger \tag{6}$$

and just write $U = (U_1|\cdots|U_d)$ with the U_i's being *orthogonal*, where $\Lambda = \operatorname{diag}(\Lambda_1, \dots, \Lambda_n) : \Lambda_1 \geq \cdots \geq \Lambda_n \geq 0, \sum_{i=1}^d \Lambda_i = 1$.

The only difference in (5) is that we do not require that the V_i's are orthogonal. Thus, it is easier to simulate a matrix ρ by simulating the V_i's and $\gamma_i' s$ in (5) than by simulating U and Λ in (6). Also, note that the γ_i's are not necessarily the eigenvalues of ρ.

Definition 1. We define the prior definition on ρ, $\pi(\mathrm{d}\rho)$, by

$$V_1, \dots, V_d \sim \text{ i.i.d uniform distribution on the unit sphere,}$$
$$(\gamma_1, \dots, \gamma_d) \sim \mathcal{D}ir(\alpha_1, \dots, \alpha_d),$$
$$\rho = \sum_{i=1}^d \gamma_i V_i V_i^\dagger$$

where $\mathcal{D}ir(\alpha_1, \dots, \alpha_d)$ is the Dirichlet distribution with parameters $\alpha_1, \dots, \alpha_d > 0$.

Remark 1. To get an approximate rank-1 matrix ρ, one can take all parameters of the Dirichlet distribution equal to a constant that is very closed to 0 (e.g $\alpha_1 = \cdots = \alpha_d = \frac{1}{d}$). And a typical drawing will lead to one of the $\gamma_i' s$ close to 1 and the others close to 0. See Wallach et al. (2009) for more discussion on choosing the parameters for Dirichlet distribution. Theoretical recommendations for the α_i's are given in Section 4.

Remark 2. We could impose the V_i's to be orthogonal in practice. The theoretical results would be unchanged, however, the implementation of our method would become trickier. Note that to sample from the uniform distribution on the sphere is rather easy. We can for example simulate $\tilde{V}_i$ from any isotropic distribution, e.g. $\mathcal{N}(0, \mathbb{I})$ and define $V_i := \tilde{V}_i/\|\tilde{V}_i\|$.

4. PAC-Bayesian estimation and analysis

4.1. Pseudo-likelihoods

Here, we consider two natural ways to compare a theoretical density ρ and the observations: first $p_{\mathbf{a},\mathbf{s}}$ should be close to the empirical part $\hat{p}_{\mathbf{a},\mathbf{s}}$; second ρ should be close to the least square (invert) estimator $\hat{\rho}$. As we have no reason to prefer one in advance, we define and study 2 estimators.

(a) Distance between the probabilities: prob-estimator

We consider

$$\ell^{prob}(\nu, \mathcal{D}) = \sum_{\mathbf{a}\in\mathcal{E}^n} \sum_{\mathbf{s}\in\mathcal{R}^n} \left[\mathrm{Tr}(\nu P_{\mathbf{s}}^{\mathbf{a}}) - \hat{p}_{\mathbf{a},\mathbf{s}}\right]^2$$

and

$$\tilde{\rho}_{\lambda}^{prob} = \int \nu \tilde{\pi}_{\lambda}^{prob}(d\nu),$$

$$\tilde{\pi}_{\lambda}^{prob}(\mathrm{d}\nu) \propto \exp\left[-\lambda \ell^{prob}(\nu, \mathcal{D})\right] \pi(\mathrm{d}\nu).$$

Note that if we use the shortened notation $p_\nu = [\mathrm{Tr}(\nu P_{\mathbf{s}}^{\mathbf{a}})]_{\mathbf{a},\mathbf{s}}$ and $\hat{p} = [\hat{p}_{\mathbf{a},\mathbf{s}}]_{\mathbf{a},\mathbf{s}}$ then

$$\ell^{prob}(\nu, \mathcal{D}) = \|p_\nu - \hat{p}\|_F^2$$

(Frobenius norm). This distance quantifies how far the probabilities and the empirical frequencies in the sample are.

(b) Distance between the density matrices: dens-estimator

Now, let us take:

$$\ell^{dens}(\nu, \mathcal{D}) = \|\nu - \hat{\rho}\|_F^2$$

and

$$\tilde{\rho}_{\lambda}^{dens} = \int \nu \tilde{\pi}_{\lambda}^{dens}(d\nu),$$

$$\tilde{\pi}_{\lambda}^{dens}(\mathrm{d}\nu) \propto \exp\left[-\lambda \ell^{dens}(\nu, \mathcal{D})\right] \pi(\mathrm{d}\nu).$$

In another words, this estimator finds a balance between prior information and closeness to the least square estimate $\hat{\rho}$. From a computational point of view, this estimator is easier to implement than the previous estimator.

4.2. Statistical properties of the estimators

Assumption 1. Fix some constants $D_1 > 0$ and $D_2 > 0$ (that do not depend on m nor n). We assume that the parameters of the Dirichlet prior distribution $\mathcal{D}ir(\alpha_1, \ldots, \alpha_d)$ satisfy

- $\forall i = 1, \ldots, d : \alpha_i \leq 1$,
- $\sum_{i=1}^d \alpha_i = D_1$,
- $\prod_{i=1}^d \alpha_i \geq e^{-D_2 d \log(d)}$.

Note that this assumption is satisfied for $\alpha_1 = \cdots = \alpha_d = 1/d$ with $D_1 = D_2 = 1$.

The first theorem provides the concentration bound on the square error of the first estimator $\tilde{\rho}_{\lambda}^{prob}$. The proof of this theorem is left to the Appendix.

Theorem 1. *Fix a small $\epsilon \in (0, 1)$. Under Assumption 1, for $\lambda = \lambda^* := m/2$, with probability at least $1 - \epsilon$, one has*

$$\|\tilde{\rho}_{\lambda^*}^{prob} - \rho^0\|_F^2 \leq C_{D_1,D_2}^{prob} \frac{3^n \mathrm{rank}(\rho^0) \log\left(\frac{\mathrm{rank}(\rho^0) N}{2^n}\right) + (1.5)^n \log(2/\epsilon)}{N},$$

where C_{D_1,D_2}^{prob} is a constant that depends only on D_1, D_2.

Remark 3. As said in the introduction, the best known rate up-to-date in this problem is $\frac{3^n \mathrm{rank}(\rho^0)}{N}$, so our estimator $\tilde{\rho}_{\lambda^*}^{prob}$ reaches this rate (up to log terms). This rate is actually $\left(\frac{3}{2}\right)^n \frac{rd}{N}$ and the best lower bound known in this case is $\frac{rd}{N}$ (Butucea et al., 2015) (we remind that $d = 2^n$).

The next theorem presents the square error bound of the second estimator $\tilde{\rho}_{\lambda}^{dens}$. Here again, see the appendix for the proof.

Theorem 2. *Fix a small $\epsilon \in (0,1)$. Under Assumption 1, for $\lambda = \lambda^* := \frac{N}{5^n 4}$, with probability at least $1-\epsilon$,*

$$\|\tilde{\rho}^{dens}_{\lambda^*} - \rho^0\|_F^2 \leq C^{dens}_{D_1,D_2} \frac{10^n \text{rank}(\rho^0) \log\left(\frac{\text{rank}(\rho^0)N}{2^n}\right) + 5^n \log(2/\varepsilon)}{N} \tag{7}$$

where $C^{dens}_{D_1,D_2}$ is a constant that depends only on D_1, D_2.

Remark 4. The guarantee for $\tilde{\rho}^{dens}_{\lambda^*}$ is far less satisfactory. However, as this estimator is easier to compute, we think it is interesting to provide a convergence rate, even if it is far from optimal: note that for a fixed d, the bound goes to 0 when $m \to \infty$.

Remark 5. Experiments show that $\lambda = \lambda^* := \frac{N}{5^n 4}$ is actually not the best choice for dens-estimator. The choice $\lambda = \frac{N}{4}$ (heuristically motivated by Dalalyan and Tsybakov, 2008) leads to results comparable to the prob-estimator in Section 5. This leads to the conjecture that the rate of $\tilde{\rho}^{dens}_{N/4}$ is much better than $\frac{10^n \text{rank}(\rho^0)}{N}$ but this is still an open question.

5. Numerical experiments

5.1. Metropolis–Hastings implementation

We implement the two proposed estimators via the Metropolis–Hasting (MH) algorithm (Robert and Casella, 2013). Note that to draw $(\gamma_1,\ldots,\gamma_d) \sim \mathcal{D}ir(\alpha,\ldots,\alpha)$ is equivalent to draw $\gamma_i = Y_i/(Y_1+\cdots+Y_d)$ with $Y_i \overset{i.i.d}{\sim} Gamma(\alpha,1), \forall i = 1,\ldots,d$. Thus, instead of $\gamma_i's$, we conduct a MH updating for $Y_i's$. So the objective is to produce a Markov chain $(Y_1^{(t)},\ldots,Y_d^{(t)},V_1^{(t)},\ldots,V_d^{(t)})$. From this, we deduce obviously the sequence $(\gamma_1^{(t)},\ldots,\gamma_d^{(t)},V_1^{(t)},\ldots,V_d^{(t)})$ and use the following empirical mean as the Monte-Carlo approximation of our estimator:

$$\hat{\rho}^{\text{MH}} := \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{d}\gamma_i^{(t)}V_i^{(t)}(V_i^{(t)})^\dagger\right).$$

Algorithm 1 MH implementation

For t from 1 to T, we iteratively update through the following steps:

updating for Y_i's : for i from 1 to d,

Sample $\tilde{Y}_i \sim h(y|Y_i^{(t-1)})$ where h is a proposal distribution given explicitly below.

Calculate $\tilde{\gamma}_i = \tilde{Y}_i/(\sum_{i=1}^d \tilde{Y}_i)$.

Set

$$Y_i^{(t)} = \begin{cases} \tilde{Y}_i & \text{with probability } \min\left\{1, R(\tilde{Y}, Y^{(t-1)})\right\}, \\ Y_i^{(t-1)} & \text{otherwise} \end{cases}$$

where $R(\tilde{Y}, Y^{(t-1)})$ is the acceptance ratio given below.

Put $\gamma_i^{(t)} = Y_i^{(t)}/(\sum_{j=1}^d Y_j^{(t)}),\ i = 1,\ldots,d$.

updating for V_i's : for i from 1 to d,

Sample $\tilde{V}_i$ from the uniform distribution on the unit sphere.

Set

$$V_i^{(t)} = \begin{cases} \tilde{V}_i & \text{with probability } \min\{1, A(V^{(t-1)}, \tilde{V})\}, \\ V_i^{(t-1)} & \text{otherwise,} \end{cases}$$

where $A(V^{(t-1)}, \tilde{V})$ is the acceptance ratio given below.

Let us now give precisely h, R and A. We define $h(\cdot|Y_i^{(t-1)})$ as the probability distribution of $U = Y_i^{(t-1)}\exp(y)$ where $y \sim \mathcal{U}(-0.5, 0.5)$. Following Robert and Casella (2013) the acceptance ratios are then given by:

$$\begin{aligned}\log(R(\tilde{Y}, Y^{(t-1)})) = {}& \lambda\ell\left(\sum_{i=1}^d \tilde{\gamma}_i V_i V_i^\dagger, \mathcal{D}\right) - \lambda\ell\left(\sum_{i=1}^d \gamma_i^{(t-1)} V_i V_i^\dagger, \mathcal{D}\right) \\ & + \sum_{i=1}^d((\alpha-1)\log(\tilde{Y}_i) - \tilde{Y}_i) - \sum_{i=1}^d((\alpha-1)\log(Y_i^{(t-1)}) - Y_i^{(t-1)}) + \sum_{i=1}^d \tilde{Y}_i - \sum_{i=1}^d Y_i^{(t-1)}\end{aligned}$$

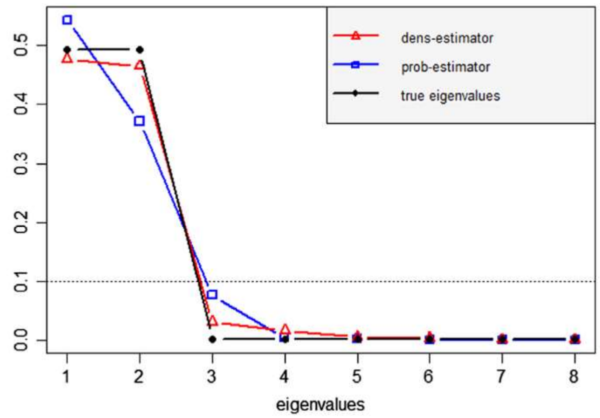

Fig. 1. Eigenvalues of estimates for an "approximate rank-2" density with $d = 2^3$, $m = 200$.

and

$$\log(A(V^{(t-1)}, \tilde{V})) = \lambda\ell\left(\sum_{i=1}^{d}\gamma_i\tilde{V}_i\tilde{V}_i^\dagger, \mathcal{D}\right) - \lambda\ell\left(\sum_{i=1}^{d}\gamma_i V_i^{(t-1)}(V_i^{(t-1)})^\dagger, \mathcal{D}\right)$$

where $\ell(\cdot, \mathcal{D})$ stands for $\ell^{dens}(\cdot, \mathcal{D})$ or $\ell^{prob}(\nu, \mathcal{D})$ depending on the estimator we are computing.

5.2. Experiments

We study the numerical performance of the prob-estimators with $\lambda = m/2$, i.e. $\tilde{\rho}^{prob}_{m/2}$ and the dens-estimator with $\lambda = \frac{N}{4}$, i.e. $\tilde{\rho}^{dens}_{N/4}$ on the following settings, all with $n = 2, 3, 4$ $(d = 4, 8, 16)$:

- a pure state density (rank-1) $\rho = \psi\psi^\dagger$ with $\psi \in \mathbb{C}^{d\times 1}$,
- a rank-2 density matrix that $\rho_{rank-2} = \frac{1}{2}\psi_1\psi_1^\dagger + \frac{1}{2}\psi_2\psi_2^\dagger$ with ψ_1, ψ_2 being two normalized orthogonal vectors in $\mathbb{C}^{d\times 1}$,
- an "approximate rank-2" density matrix: $\rho = w\rho_{rank-2} + (1-w)\frac{\mathbb{I}_d}{d}$, $w = 0.98$. Note that by "approximate rank-2", we mean that ρ is very well approximated by a rank-2 matrix ρ_{rank-2} (in the sense that $\|\rho - \rho_{rank-2}\|_F^2$ is small), but in general ρ itself is full rank,
- a maximal mixed state (rank-d).

The experiments are done for $m = 20$; 200; 1000; 2000. The parameter for $\mathcal{D}ir(\alpha, \ldots, \alpha)$ is $\alpha = 0.5$. We repeat each experiment 10 times, and compute the mean of the square error, MSE, $\|\hat{\rho} - \rho\|_F^2$ for each estimator, together with the associated standard deviation (between brackets in Tables 1–3).

5.3. Results

We compare the prob- and dens-estimator to the simple inversion procedure and to the thresholding estimator of Butucea et al. (2015). The results are given in Tables 1–3 (outputs from the **R** software). The conclusions are:

- The prob-estimator seems to be the most accurate but also comes with a larger standard deviation. This might be due to slow convergence of the MCMC procedure. Indeed each step is computationally highly expensive.
- The dens-estimator is easier to compute and while it is less accurate than the prob-estimator, it still shows better results than the direct inversion method.
- The thresholding estimator of Butucea et al. (2015) works well for rank-1 states but seems to bring too much bias for other states.

Besides the square error, the eigenvalues of the estimates are also important when reconstructing density matrices. In Fig. 1, the dens-estimator returns with eigenvalues similar to the true eigenvalues of the true density matrix, while the prob-estimator seems not to shrink enough.

5.4. Real data tests

The experiments performed to produce the data are explained in Barreiro et al. (2010). The data was kindly provided by M. Guţă and T. Monz. It had been used in Alquier et al. (2013) and Guţă et al. (2012). We apply two proposed estimators to the real dataset of a system of 4 ions which is Smolin state further manipulated. In Fig. 2 we plot the eigenvalues of the inversion estimator and our ones. Note that the distribution of the eigenvalues of the three estimators are rather different. Still, it seems that all estimators return results compatible with a rank-2 state.

Table 1
MSEs for $n = 4$ (together with standard deviations).

	$m = 20$	$m = 200$	$m = 1000$	$m = 2000$
Pure state, MSEs $\times 10^5$				
Inversion	175 (4e−4)	14.8 (2e−5)	2.71 (8e−6)	1.55 (5e−6)
Thresholding	93.5 (3e−4)	**12.6** (3e−5)	**.596** (2e−6)	**.412** (2e−6)
prob	86.3 (6e−4)	22.4 (2e−4)	10.5 (6e−5)	5.13 (2e−5)
dens	**51.5** (2e−4)	21.7 (7e−5)	13.1 (3e−5)	13.2 (2e−5)
rank-2 state, MSEs $\times 10^3$				
Inversion	16.8 (8e−4)	15.9 (3e−4)	15.9 (1e−4)	15.8 (7e−5)
Thresholding	14.9 (3e−4)	15.5 (7e−5)	15.5 (9e−6)	15.5 (7e−6)
prob	**9.29** (2e−3)	**7.90** (1e−3)	**8.46** (1e−3)	**7.84** (8e−4)
dens	14.5 (3e−4)	14.6 (3e−4)	14.4 (3e−4)	14.5 (4e−4)
Approximate rank-2 state, MSEs $\times 10^3$				
Inversion	15.9 (8e−4)	15.4 (2e−4)	15.3 (1e−4)	15.2 (4e−5)
Thresholding	14.3 (2e−4)	14.2 (3e−4)	15.0 (1e−5)	15.0 (6e−6)
prob	**8.88** (9e−4)	**7.68** (2e−3)	**8.11** (1e−3)	**7.39** (1e−3)
dens	13.9 (4e−4)	15.1 (2e−4)	14.2 (3e−4)	14.2 (2e−4)
Maximal mixed state, MSEs $\times 10^4$				
Inversion	15.9 (4e−4)	6.57 (7e−5)	5.09 (5e−5)	4.76 (2e−5)
Thresholding	**4.67** (9e−5)	5.59 (5e−5)	5.34 (8e−5)	6.06 (8e−5)
prob	5.44 (2e−4)	**3.37** (8e−5)	**3.31** (8e−5)	**3.20** (8e−5)
dens	5.72 (9e−5)	4.47 (6e−5)	4.56 (4e−5)	4.24 (2e−5)

Table 2
MSEs for $n = 3$ (together with standard deviations).

	$m = 20$	$m = 200$	$m = 1000$	$m = 2000$
Pure state, MSEs $\times 10^4$				
Inversion	39.5 (9e−4)	3.17 (9e−5)	.559 (1e−5)	.343 (1e−5)
Thresholding	21.4 (6e−4)	**2.26** (1e−4)	**.196** (1e−5)	**.152** (1e−5)
prob	40.3 (2e−2)	5.79 (4e−4)	2.95 (2e−4)	1.78 (1e−4)
dens	**12.8** (5e−4)	2.73 (2e−4)	1.24 (4e−5)	1.07 (4e−5)
rank-2 state, MSEs $\times 10^2$				
Inversion	3.69 (3e−3)	3.35 (6e−4)	3.32 (4e−4)	3.31 (2e−4)
Thresholding	2.94 (1e−3)	3.05 (2e−4)	3.04 (6e−5)	3.05 (5e−5)
prob	**1.91** (5e−3)	**1.17** (3e−3)	**1.18** (3e−3)	**1.14** (2e−3)
dens	2.83 (8e−4)	2.89 (3e−4)	2.89 (3e−4)	3.00 (1e−4)
Approximate rank-2 state, MSEs $\times 10^2$				
Inversion	3.33 (2e−4)	3.22 (8e−4)	3.19 (3e−4)	3.18 (2e−4)
Thresholding	2.81 (1e−3)	2.96 (1e−4)	2.97 (8e−5)	2.97 (9e−5)
prob	**1.10** (5e−3)	**.551** (5e−3)	**.189** (2e−3)	**.113** (1e−3)
dens	2.74 (6e−4)	2.88 (3e−4)	2.91 (3e−4)	2.91 (2e−4)
Maximal mixed state, MSEs $\times 10^3$				
Inversion	6.98 (2e−3)	3.19 (4e−4)	2.88 (2e−4)	3.01 (1e−4)
Thresholding	4.41 (6e−4)	3.26 (6e−4)	3.19 (2e−4)	3.29 (1e−4)
prob	3.63 (1e−3)	**2.70** (7e−4)	**2.28** (7e−4)	**2.29** (1e−3)
dens	**3.18** (6e−4)	2.99 (4e−4)	2.90 (2e−4)	3.04 (1e−4)

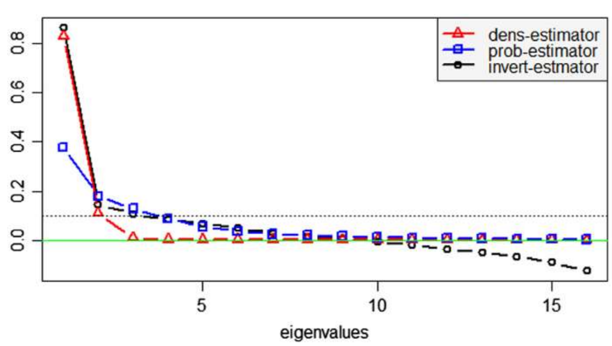

Fig. 2. Eigenvalues plots for real data test with $n = 4$.

Table 3
MSEs for $n = 2$ (together with standard deviations).

	$m = 20$	$m = 200$	$m = 1000$	$m = 2000$
Pure state, MSEs $\times 10^4$				
Inversion	61.9 (3e−3)	9.22 (5e−4)	.802 (4e−5)	.772 (6e−5)
Thresholding	**49.4** (3e−3)	**4.06** (3e−4)	**.737** (4e−5)	**.356** (2e−5)
prob	102 (8e−3)	39.7 (2e−3)	9.37 (8e−4)	7.19 (5e−4)
dens	52.2 (3e−3)	7.57 (5e−4)	1.91 (9e−5)	1.08 (2e−5)
rank-2 state, MSEs $\times 10^2$				
Inversion	8.24 (2e−2)	7.91 (3.2e−3)	7.81 (2e−3)	7.74 (7e−4)
Thresholding	5.13 (3e−3)	5.34 (1.1e−3)	5.32 (5e−4)	5.33 (4e−4)
prob	**2.62** (2e−2)	**1.77** (7.4e−3)	**1.79** (8e−3)	**1.73** (5e−3)
dens	4.53 (3e−3)	5.20 (1.5e−3)	5.24 (9e−4)	5.24 (9e−4)
Approximate rank-2 state, MSEs $\times 10^2$				
Inversion	8.12 (2e−2)	7.54 (4e−3)	7.54 (1.2e−3)	7.56 (6e−4)
Thresholding	4.95 (4e−3)	5.19 (8e−4)	5.23 (5e−4)	5.22 (4e−4)
prob	**2.69** (2e−2)	**1.82** (1.1e−2)	**1.52** (6e−3)	**1.58** (6e−3)
dens	4.40 (4e−3)	5.02 (1.3e−3)	5.11 (1e−3)	5.15 (6e−4)
Maximal state, MSEs $\times 10^2$				
Inversion	3.03 (9e−3)	2.12 (2e−3)	2.11 (2e−3)	2.11 (1e−3)
Thresholding	2.78 (8e−3)	2.36 (2e−3)	2.21 (2e−3)	2.25 (1e−3)
prob	2.32 (2e−2)	**1.15** (5e−3)	**1.19** (5e−3)	**1.07** (4e−3)
dens	**2.30** (6e−3)	2.11 (2e−3)	2.06 (2e−3)	2.09 (1e−3)

6. Discussion and conclusion

We propose a novel prior and introduce two pseudo-Bayesian estimators for the density matrix: the dens-estimator and the prob-estimator. The prob-estimator reaches the best up-to-date rate of convergence in the low-rank case. On the other hand, computation of the dens-estimator is an easier task. In practice, we recommend the prob-estimator. However, in cases where the MCMC shows activities of lacking of convergence, the dens-estimator can be used as a reasonable alternative.

Note also that the prob-estimator can be extended to the incomplete measurement case. We consider the (incomplete) pseudo-likelihood as

$$\ell^{prob-incomplete}(\nu, \mathcal{D}) = \sum_{\mathbf{a}\in\mathcal{A}} \sum_{\mathbf{s}\in\mathcal{R}^n} \left[\mathrm{Tr}(\nu P_{\mathbf{s}}^{\mathbf{a}}) - \hat{p}_{\mathbf{a},\mathbf{s}}\right]^2,$$

where $\mathcal{A} \subsetneq \mathcal{E}^n$. The study in this case will be the object of future works.

Open questions include faster algorithms based on optimization (in the spirit of Alquier et al., 2015). Also, from a theoretical perspective, the most important question is the minimax lower bound.

Acknowledgments

We thank the Associate Editor and the Referees for their constructive and valuable comments. Both authors gratefully acknowledge financial support from GENES and by the French National Research Agency (ANR) under the grant Labex Ecodec (ANR-11-LABEX-0047). P. Alquier gratefully acknowledges financial support from the research programme *New Challenges for New Data* from LCL and GENES, hosted by the *Fondation du Risque*.

Appendix. Proofs

We first remind here a version of Hoeffding's inequality for bounded random variables.

Lemma 1. *Let Y_i, $i = 1, \ldots, n$ be n independent random variables with $|Y_i| \leq b$ a.s., and $\mathbb{E}(Y_i) = 0$. Then, for any $\lambda > 0$,*

$$\mathbb{E}\exp\left(\frac{\lambda}{n}\sum_{i=1}^{n} Y_i\right) \leq \exp\left(\frac{\lambda^2 b^2}{8n}\right).$$

A.1. Preliminary lemmas for the proof of Theorem 1

Lemma 2. *For any $\lambda > 0$, we have*

$$\mathbb{E}\exp\left(\lambda\left\langle p_\nu - p^0, p^0 - \hat{p}\right\rangle_F\right) \leq \exp\left[\frac{\lambda^2}{4m}\|p^0 - p_\nu\|_F^2\right],$$

$$\mathbb{E}\exp\left(-\lambda\left\langle p_\nu - p^0, p^0 - \hat{p}\right\rangle_F\right) \leq \exp\left[\frac{\lambda^2}{4m}\|p^0 - p_\nu\|_F^2\right].$$

Proof. First inequality:

$$\begin{aligned}
\mathbb{E}\exp\left(\lambda\left\langle p_\nu - p^0, p^0 - \hat{p}\right\rangle_F\right) &= \mathbb{E}\exp\left(\lambda\sum_{a\in\mathcal{E}^n}\sum_{s\in\mathcal{R}^n}\underbrace{[\mathrm{Tr}(\nu P_s^a) - p_{a,s}^0]}_{=:c(a,s)}[p_{a,s}^0 - \hat{p}_{a,s}]\right)\\
&= \prod_{a\in\mathcal{E}^n}\mathbb{E}\exp\left(\lambda\sum_{s\in\mathcal{R}^n}c(a,s)\left[p_{a,s}^0 - \frac{1}{m}\sum_{i=1}^m \mathbf{1}(R_i^a = s)\right]\right)\\
&= \prod_{a\in\mathcal{E}^n}\mathbb{E}\exp\Bigg(\frac{\lambda}{m}\sum_{i=1}^m\underbrace{\left[\sum_{s\in\mathcal{R}^n}c(a,s)\{p_{a,s}^0 - \mathbf{1}(R_i^a = s)\}\right]}_{=:Y_{i,a}}\Bigg)
\end{aligned}$$

We have that $\mathbb{E}(Y_{i,a}) = 0$. Then, using Cauchy–Schwarz inequality

$$\begin{aligned}
Y_{i,a}^2 &\leq \left(\sum_{s\in\mathcal{R}^n}c(a,s)^2\right)\left(\sum_{s\in\mathcal{R}^n}[p_{a,s}^0 - \mathbf{1}(R_i^a = s)]^2\right)\\
&\leq \left(\sum_{s\in\mathcal{R}^n}c(a,s)^2\right)\left(\sum_{s\in\mathcal{R}^n}|p_{a,s}^0 - \mathbf{1}(R_i^a = s)|\right) \leq 2\left(\sum_{s\in\mathcal{R}^n}c(a,s)^2\right).
\end{aligned}$$

So we can apply Hoeffding's inequality (Lemma 1):

$$\begin{aligned}
\prod_{a\in\mathcal{E}^n}\mathbb{E}\exp\left(\frac{\lambda}{m}\sum_{i=1}^m Y_{i,a}\right) &\leq \prod_{a\in\mathcal{E}^n}\exp\left[\frac{2\lambda^2}{8m}\left(\sum_{s\in\mathcal{R}^n}c(a,s)^2\right)\right]\\
&\leq \exp\left[\frac{\lambda^2}{4m}\|p - p_\nu\|_F^2\right].
\end{aligned}$$

Second inequality: same proof, just replace $Y_{i,a}$ by $-Y_{i,a}$. □

Lemma 3. *For $\lambda > 0$, we have*

$$\mathbb{E}\exp\left\{\lambda\left(\|p_\nu - \hat{p}\|_F^2 - \|p^0 - \hat{p}\|_F^2\right) - \lambda\left[1 + \frac{\lambda}{m}\right]\|p^0 - p_\nu\|_F^2\right\} \leq 1, \tag{8}$$

$$\mathbb{E}\exp\left\{\lambda\left[1 - \frac{\lambda}{m}\right]\|p^0 - p_\nu\|_F^2 - \lambda\left(\|p_\nu - \hat{p}\|_F^2 - \|p^0 - \hat{p}\|_F^2\right)\right\} \leq 1. \tag{9}$$

Proof. Proof of the first inequality:

$$\begin{aligned}
\mathbb{E}\exp\left\{\lambda\left(\|p_\nu - \hat{p}\|_F^2 - \|p - \hat{p}\|_F^2\right)\right\} &= \mathbb{E}\exp\left\{\lambda\left\langle p_\nu - p^0, p_\nu + p^0 - 2\hat{p}\right\rangle_F\right\}\\
&= \mathbb{E}\exp\left\{\lambda\|p_\nu - p^0\|_F^2 + 2\lambda\left\langle p_\nu - p^0, p^0 - \hat{p}\right\rangle_F\right\}\\
&= \exp\left(\lambda\|p_\nu - p^0\|_F^2\right)\mathbb{E}\exp\left\{2\lambda\left\langle p_\nu - p^0, p^0 - \hat{p}\right\rangle_F\right\}\\
&\leq \exp\left(\lambda\|p_\nu - p^0\|_F^2\right)\exp\left\{\frac{\lambda^2}{m}\|p_\nu - p^0\|_F^2\right\}
\end{aligned}$$

thanks to Lemma 2. The proof of the second inequality is similar. □

Using Lemma 3, we derive an empirical PAC-Bayes bound for the estimator.

Lemma 4. *For $\lambda > 0$ s.t. $\frac{\lambda}{m} < 1$, with prob. $1 - \epsilon/2$, $\epsilon \in (0, 1)$, for any distribution $\hat{\pi}$, we have:*

$$\int \|p_\nu - p^0\|_F^2 \tilde{\pi}_\lambda(\mathrm{d}\nu) \leq \frac{\int \|p_\nu - \hat{p}\|_F^2 \hat{\pi}(\mathrm{d}\nu) - \|p^0 - \hat{p}\|_F^2 + \frac{\mathcal{K}(\tilde{\pi}_\lambda, \pi) + \log\left(\frac{2}{\epsilon}\right)}{\lambda}}{1 - \frac{\lambda}{m}}.$$

Proof. We rewrite (9) in Lemma 3 as follows

$$\int \mathbb{E} \exp\left\{\lambda \left[1-\frac{\lambda}{m}\right] \|p^0 - p_\nu\|_F^2 - \lambda \left(\|p_\nu - \hat{p}\|_F^2 - \|p^0 - \hat{p}\|_F^2\right)\right\}\pi(d\nu) \leq 1.$$

By using Fubini's theorem

$$\mathbb{E} \int \exp\left\{\lambda \left[1-\frac{\lambda}{m}\right] \|p^0 - p_\nu\|_F^2 - \lambda \left(\|p_\nu - \hat{p}\|_F^2 - \|p^0 - \hat{p}\|_F^2\right)\right\}\pi(d\nu) \leq 1.$$

Now, using Catoni (2007, Lemma 1.1.3), for any distribution $\hat{\pi}$, we have

$$\mathbb{E} \exp \sup_{\hat{\pi}}\left\{\lambda \left[1-\frac{\lambda}{m}\right] \int \|p^0 - p_\nu\|_F^2 \hat{\pi}(d\nu) - \log(2/\epsilon) - \mathcal{K}(\hat{\pi},\pi) - \lambda\left(\int \|p_\nu - \hat{p}\|_F^2 \hat{\pi}(d\nu) - \|p^0 - \hat{\rho}\|_F^2\right)\right\} \leq \frac{\epsilon}{2}$$

and with $\mathbf{1}_{\mathbf{R}_+}(x) \leq \exp(x)$, one has

$$\mathbb{P}\left\{\sup_{\hat{\pi}}\left[\lambda \left[1-\frac{\lambda}{m}\right] \int \|p^0 - p_\nu\|_F^2 \hat{\pi}(d\nu) - \log(2/\epsilon) - \mathcal{K}(\hat{\pi},\pi) - \lambda\left(\int \|p_\nu - \hat{p}\|_F^2 \hat{\pi}(d\nu) - \|p^0 - \hat{\rho}\|_F^2\right)\right] \geq 0\right\} \leq \frac{\epsilon}{2}.$$

Taking the complementary yields successfully the results. □

The following lemma gives a theoretical PAC-Bayes bound for the estimator.

Lemma 5. *For $\lambda > 0$ s.t $\frac{\lambda}{m} < 1$, with probability $1-\epsilon$ we have:*

$$\int \|p_\nu - p^0\|_F^2 \hat{\pi}_\lambda^{prob}(\mathrm{d}\nu) \leq \inf_{\hat{\pi}} \frac{\left[1+\frac{\lambda}{m}\right]\int \|p_\nu - p^0\|_F^2 \tilde{\pi}(\mathrm{d}\nu) + \frac{2\mathcal{K}(\hat{\pi},\pi)+2\log\left(\frac{2}{\epsilon}\right)}{\lambda}}{1-\frac{\lambda}{m}} \tag{10}$$

and

$$\int \|\nu - \rho^0\|_F^2 \hat{\pi}_\lambda^{prob}(\mathrm{d}\nu) \leq \inf_{\hat{\pi}} \frac{3^n\left[1+\frac{\lambda}{m}\right]\int \|\nu - \rho^0\|_F^2 \hat{\pi}(\mathrm{d}\nu) + \frac{2\mathcal{K}(\hat{\pi},\pi)+2\log\left(\frac{2}{\epsilon}\right)}{2^n\lambda}}{1-\frac{\lambda}{m}}. \tag{11}$$

Proof. Using the same proof of Lemma 4 for inequality (8) in Lemma 3, we obtain with probability at least $1-\epsilon/2$, $\epsilon \in (0,1)$, for any distribution $\hat{\pi}$ that

$$\int \|p^0 - \hat{p}\|_F^2 \hat{\pi}(d\nu) \leq \left[1+\frac{\lambda}{m}\right]\int \|p_\nu - p^0\|_F^2 \hat{\pi}(d\nu) + \|p^0 - \hat{p}\|_F^2 + \frac{\mathcal{K}(\hat{\pi},\pi)+\log(\frac{2}{\epsilon})}{\lambda}.$$

With a union argument, combining Lemma 4 and the above inequality yields the following inequality with probability at least $1-\epsilon$, $\epsilon \in (0,1)$, for any $\hat{\pi}$

$$\int \|p_\nu - p^0\|_F^2 \hat{\pi}(\mathrm{d}\nu) \leq \frac{\left[1+\frac{\lambda}{m}\right]\int \|p_\nu - p^0\|_F^2 \hat{\pi}(\mathrm{d}\nu) + \frac{2\mathcal{K}(\hat{\pi},\pi)+2\log(2/\epsilon)}{\lambda}}{1-\frac{\lambda}{m}}.$$

Taking $\tilde{\pi}_\lambda^{prob}$ (once again, Catoni, 2007, Lemma 1.1.3) be the minimizer of the right hand side of the above inequality, we obtain (10).

Moreover, in Alquier et al. (2013, equation (5)) states that, for any ν:

$$p_\nu = \mathbf{P}\nu$$

for some operator $\mathbf{P}$. Therefore

$$\|p_\nu - p^0\|_F^2 = \|\mathbf{P}(\nu - \rho^0)\|_F^2.$$

The eigenvalues of $\mathbf{P}^T\mathbf{P}$ are known, they range between 2^n and $3^n 2^n$ according to Alquier et al. (2013, Proposition 1). Thus, for any ν,

$$2^n\|\nu - \rho^0\|_F^2 \leq \|p_\nu - p^0\|_F^2 \leq 6^n\|\nu - \rho^0\|_F^2$$

and so we obtain (11). □

In the following, we will consider $\hat{\pi}$ as a restriction of the prior to a local set around the true density matrix ρ^0. This allows us to obtain an explicit bound of the left hand side of (11). Let $\rho^0 = U\Lambda U^\dagger$ be the spectral decomposition of ρ^0.

Definition 2. Let $r = \#\{i : \Lambda_i > \delta\}$, with small $\delta \in [0, 1)$. Take

$$\tilde{\pi}_c(du, dv) \propto \mathbf{1}(\forall i : |v_i - \Lambda_i| \leq \delta; \forall i = 1, \ldots, r : \|u_i - U_i\|_F \leq c)\pi(du, dv)$$

Note that we have $r \leq \mathrm{rank}(\rho^0)$.

Lemma 6. *We have*

$$\int \|u^\dagger v u - \rho^0\|_F^2 \tilde{\pi}_c(du, dv) \leq (3d\delta + 2rc)^2. \tag{12}$$

And under Assumption 1

$$\mathcal{K}(\tilde{\pi}_c, \pi) \leq ard \log\left(\frac{1}{c}\right) + C_{D_1,D_2} d\left(\log(d) + \log\left(\frac{1}{\delta}\right)\right) \tag{13}$$

where a is a universal constant and where C_{D_1,D_2} depends only on D_1 and D_2.

Proof. Firstly

$$\|uvu^\dagger - \rho^0\|_F^2 \leq \left(\|uvu^\dagger - u\Lambda u^\dagger\|_F + \|u\Lambda u^\dagger - U\Lambda U^\dagger\|_F\right)^2$$

and

$$\begin{aligned}
\|uvu^\dagger - u\Lambda u^\dagger\|_F &\leq \sum_i |v_i - \Lambda_i| \|u_i u_i^\dagger\|_F \leq d\delta, \\
\|u\Lambda u^\dagger - U\Lambda U^\dagger\|_F &\leq \sum_i \Lambda_i \|u_i u_i^\dagger - U_i U_i^\dagger\|_F \\
&\leq \sum_{i:\Lambda_i > \delta} (\|u_i u_i^\dagger - u_i U_i^\dagger\|_F + \|u_i U_i^\dagger - U_i U_i^\dagger\|_F) + \delta \sum_{i:\Lambda_i \leq \delta} (\|u_i u_i^\dagger\|_F + \|U_i U_i^\dagger\|_F) \\
&\leq 2rc + 2\delta(d - r) \leq 2rc + 2\delta d,
\end{aligned}$$

so we obtain (12).

Now, the Kullback–Leibler term

$$\begin{aligned}
\mathcal{K}(\tilde{\pi}_c, \pi) &= \log \frac{1}{\pi(\{u, v : \forall i : |v_i - \Lambda_i| \leq c; \forall i = 1, r : \|u_i - U_i\|_F \leq \delta\})} \\
&= \log \frac{1}{\pi(\{\forall i : |v_i - \Lambda_i| \leq \delta\})} + \log \frac{1}{\pi(\{\forall i = 1, r : \|u_{i.} - U_{i.}\|_F \leq c\})}.
\end{aligned}$$

The first log term

$$\begin{aligned}
\pi(\{\forall i = 1, r : \|u_{i.} - U_{i.}\|_F \leq c\}) &\geq \prod_{i=1}^{r} \left[\frac{\pi^{(d-1)/2}(c/2)^{d-1}}{\Gamma\left(\frac{d-1}{2} + 1\right)} \Bigg/ \frac{2\pi^{(d+1)/2}}{\Gamma\left(\frac{d+1}{2}\right)}\right], \quad d = 2^n \\
&\geq \left[\frac{c^{d-1}}{2^d \pi}\right]^r \geq \frac{c^{r(d-1)}}{2^{4rd}}.
\end{aligned}$$

Note for the above calculation: it is greater or equal to the volume of the (d-1)-“circle” with radius $c/2$ over the surface area of the d-“unit-sphere”.

The second log term in the Kullback–Leibler term

$$\begin{aligned}
\pi(\{\forall i : |v_i - \Lambda_i| \leq \delta\}) &= \frac{\Gamma(D_1)}{\prod_{i=1}^{d} \Gamma(\alpha_i)} \prod_{i=1}^{d} \int_{\max(\Lambda_i - \delta, 0)}^{\min(\Lambda_i + \delta, 1)} v_i^{\alpha_i - 1} dv_i \\
&\geq \Gamma(D_1)\delta^d \prod_{i=1}^{d} \alpha_i \geq C_{D_1} \delta^d \, e^{-D_2 d \log(d)}
\end{aligned}$$

for some constant C_{D_1} that depends only on D_1. Since $\alpha_i \leq 1$ for every i, we can lower bound the integrand by 1 and also $\alpha_i \Gamma(\alpha_i) = \Gamma(\alpha_i + 1) \leq 1$. The interval of integration contains at least an interval of length δ. This trick was presented in Ghosal et al. (2000, Lemma 6.1, page 518)

Thus, we obtain

$$\begin{aligned}\mathcal{K}(\tilde{\pi}_c,\pi) &\leq \log\frac{2^{4rd}}{c^{r(d-1)}} + \log\left(\frac{e^{D_2 d\log(d)}}{C_{D_1}\delta^d}\right)\\ &\leq ard\log\left(\frac{1}{c}\right) + C_{D_1,D_2}d\left(\log(d)+\log\left(\frac{1}{\delta}\right)\right)\end{aligned}$$

for some absolute constant a and where C'_{D_1,D_2} depends only on D_1 and D_2. □

A.2. Proof of Theorem 1

Proof of Theorem 1. Substituting (13), (12) into (11), we obtain

$$\int\|\nu-\rho^0\|_F^2\tilde{\pi}_\lambda(\mathrm{d}\nu) \leq \inf_c\left\{\frac{3^n\left[1+\frac{\lambda}{m}\right](3d\delta+2rc)^2}{1-\frac{\lambda}{m}} + \frac{ard\log\left(\frac{1}{c}\right)+C_{D_1,D_2}d\left(\log(d)+\log\left(\frac{1}{\delta}\right)\right)+2\log(2/\epsilon)}{\lambda 2^n[1-\frac{\lambda}{m}]}\right\}.$$

By taking $\delta=\frac{1}{d\sqrt{N}}$, $c=\sqrt{\frac{d}{rm9^n}}$, $\lambda=m/2$ leads to

$$\int\|\nu-\rho^0\|_F^2\tilde{\pi}_\lambda(\mathrm{d}\nu)\leq A\left(\frac{1}{m}+\frac{rd}{m3^n}\right)+C'_{D_1,D_2}\frac{r\log(rm3^n/d)+\log(m3^n)+\log(2/\epsilon)/2^n}{m}$$

for some absolute constant A. Finally, by Jensen inequality, one has

$$\|\hat{\rho}_\lambda-\rho^0\|_F^2\leq\int\|\nu-\rho^0\|_F^2\hat{\pi}_\lambda(\mathrm{d}\nu).$$

This completes the proof of the theorem. □

A.3. Preliminary results for the proof of Theorem 2

Rewriting Eq. (1), by plugging (3) in, as follows

$$p_{\mathbf{a},\mathbf{s}}=\sum_{b\in\{I,x,y,z\}^n}\rho_b\mathrm{Trace}\left(\sigma_b\cdot P_{\mathbf{s}}^{\mathbf{a}}\right)=\sum_{b\in\{I,x,y,z\}^n}\rho_b\mathbf{P}_{(s,a),b}.$$

where $\mathbf{P}_{(s,a),b}=\prod_{j\neq E_b}s_j\mathbf{1}(a_j=b_j)$ and $E_b=\{j\in\{1,\ldots,n\}:b_j=I\}$, see Alquier et al. (2013) for technical details. We are now ready to handle with the proofs.

Lemma 7. *For any $\lambda>0$, we have*

$$\mathbb{E}\exp\left(\lambda\left\langle\rho^0-\nu,\rho^0-\hat{\rho}\right\rangle_F\right)\leq\exp\left[\frac{4\lambda^2}{m}\left(\frac{5}{3}\right)^n\|\nu-\rho^0\|_F^2\right]$$
$$\mathbb{E}\exp\left(-\lambda\left\langle\rho^0-\nu,\rho^0-\hat{\rho}\right\rangle_F\right)\leq\exp\left[\frac{4\lambda^2}{m}\left(\frac{5}{3}\right)^n\|\nu-\rho^0\|_F^2\right].$$

Proof. First inequality

$$\begin{aligned}&\mathbb{E}\exp\left(\lambda\left\langle\rho^0-\nu,\rho^0-\hat{\rho}\right\rangle_F\right)\\ &=\mathbb{E}\exp\left[\lambda\sum_b(\rho_b^0-\nu_b)\rho_b^0-\hat{\rho}_b Trace(\sigma_b\sigma_b^\dagger)\right]\\ &=\mathbb{E}\exp\left[d\lambda\sum_b(\rho_b^0-\nu_b)\sum_s\sum_a\frac{\mathbf{P}_{(s,a),b}}{3^{d(b)}2^n}(p_{a,s}^0-\hat{p}_{a,s})\right]\\ &=\prod_a\mathbb{E}\exp\left[\lambda\sum_b(\rho_b^0-\nu_b)\sum_s\frac{1}{m}\sum_{i=1}^m\frac{\mathbf{P}_{(s,a),b}}{3^{d(b)}}(p_{a,s}^0-\mathbf{1}_{R_i^a=s})\right]\\ &=\prod_a\prod_i\mathbb{E}\exp\Bigg[\underbrace{\frac{\lambda}{m}\sum_b(\rho_b^0-\nu_b)\sum_s\frac{\mathbf{P}_{(s,a),b}}{3^{d(b)}}(p_{a,s}^0-\mathbf{1}_{R_i^a=s})}_{:=Y_{i,a}}\Bigg].\end{aligned}$$

Remark that $\mathbb{E}(Y_{i,a}) = 0$. Also, from the definitions above, the absolute value $|\mathbf{P}_{(s,a),b}|$ does not depend on s so

$$\begin{aligned}
|Y_{i,a}| &\leq \sum_b |\rho_b^0 - \nu_b| \left|\frac{\mathbf{P}_{(s,a),b}}{3^{d(b)}}\right| \sum_s |p_{a,s}^0 - \mathbf{1}_{R_i^a = s}| \\
&\leq 2\sum_b |\rho_b^0 - \nu_b| \left|\frac{\mathbf{P}_{(s,a),b}}{3^{d(b)}}\right| \leq \frac{2}{2^{n/2}} \sqrt{\sum_b (\rho_b^0 - \nu_b)^2 d \sum_b \left(\frac{\mathbf{P}_{(s,a),b}}{3^{d(b)}}\right)^2} \\
&\leq \frac{2\|\nu - \rho^0\|_F}{2^{n/2}} \left(\sum_b \frac{1}{3^{2d(b)}} \prod_{j \notin E_b} \mathbf{1}_{a_j = b_j}\right)^{1/2} \\
&\leq \frac{2\|\nu - \rho^0\|_F}{2^{n/2}} \left(\sum_{\ell=0}^n \binom{n}{\ell} \frac{1}{3^{2\ell}}\right)^{1/2} \\
&\leq \frac{2\|\nu - \rho^0\|_F}{2^{n/2}} \left(1 + \frac{1}{9}\right)^{n/2} = 2\|\nu - \rho^0\|_F \left(\frac{5}{9}\right)^{n/2}.
\end{aligned}$$

So we can apply Hoeffding's inequality (Lemma 1):

$$\prod_a \mathbb{E} \exp\left(\frac{\lambda}{m} \sum_{i=1}^m Y_{i,a}\right) \leq \exp\left[\frac{\lambda^2}{2m} \left(\frac{5}{3}\right)^n \|\nu - \rho^0\|_F^2\right].$$

Second inequality: same proof, just replace $Y_i(a)$ by $-Y_i(a)$. □

Lemma 8. *We have*

$$\mathbb{E} \exp\left\{\lambda \left[1 - \frac{2\lambda}{m}\left(\frac{5}{3}\right)^n\right] \|\nu - \rho^0\|_F^2 - \lambda\left(\|\nu - \hat{\rho}\|_F^2 - \|\rho^0 - \hat{\rho}\|_F^2\right)\right\} \leq 1,$$

$$\mathbb{E} \exp\left\{\lambda\left(\|\nu - \hat{\rho}\|_F^2 - \|\rho^0 - \hat{\rho}\|_F^2\right) - \lambda \left[1 + \frac{2\lambda}{m}\left(\frac{5}{3}\right)^n\right] \|\nu - \rho^0\|_F^2\right\} \leq 1.$$

Proof. For the second inequality:

$$\begin{aligned}
\mathbb{E} \exp\left\{\lambda\left(\|\nu - \hat{\rho}\|_F^2 - \|\rho^0 - \hat{\rho}\|_F^2\right)\right\} &= \mathbb{E} \exp\left\{\lambda \left\langle \nu - \rho^0, \nu + \rho^0 - 2\hat{\rho}\right\rangle_F\right\} \\
&= \mathbb{E} \exp\left\{\lambda \|\nu - \rho^0\|_F^2 + 2\lambda \left\langle \nu - \rho^0, \rho^0 - \hat{\rho}\right\rangle_F\right\} \\
&= \exp\left(\lambda \|\nu - \rho^0\|_F^2\right) \mathbb{E} \exp\left\{2\lambda \left\langle \nu - \rho^0, \rho^0 - \hat{\rho}\right\rangle_F\right\} \\
&\leq \exp\left(\lambda \|\nu - \rho^0\|_F^2\right) \exp\left\{\frac{2\lambda^2}{m}\left(\frac{5}{3}\right)^n \|\nu - \rho^0\|_F^2\right\}
\end{aligned}$$

thanks to Lemma 7. The proof of the first inequality is similar. □

Lemma 9. *For $\lambda > 0$ s.t $\frac{2\lambda}{m}\left(\frac{5}{3}\right)^n < 1$, with probability at least $1 - \epsilon$, $\epsilon \in (0, 1)$, we have*

$$\int \|\nu - \rho^0\|_F^2 \tilde{\pi}_\lambda^{dens}(d\nu) \leq \inf_{\hat{\pi}} \frac{\left[1 + \frac{2\lambda}{m}\left(\frac{5}{3}\right)^n\right] \int \|\nu - \rho^0\|_F^2 \hat{\pi}(d\nu) + \frac{2\mathcal{K}(\hat{\pi}, \pi) + 2\log(2/\epsilon)}{\lambda}}{1 - \frac{2\lambda}{m}\left(\frac{5}{3}\right)^n}. \tag{14}$$

Proof. By using the results from Lemma 8, the proof is similar to the proof of Lemma 5. □

A.4. Proof of Theorem 2

Proof of Theorem 2. Substituting (13), (12) into (14) we get

$$\int \|\nu - \rho^0\|_F^2 \tilde{\pi}_\lambda(d\nu) \leq \inf_c \left\{\frac{\left[1 + \frac{2\lambda}{m}\left(\frac{5}{3}\right)^n\right](3d\delta + 2rc)^2}{1 - \frac{2\lambda}{m}\left(\frac{5}{3}\right)^n} + \frac{ard\log\left(\frac{1}{c}\right) + C_{D_1,D_2} d\left(\log(d) + \log\left(\frac{1}{\delta}\right)\right) + 2\log(2/\epsilon)}{\lambda[1 - \frac{2\lambda}{m}\left(\frac{5}{3}\right)^n]}\right\}.$$

Taking $\delta = \frac{d}{N}$, $c = \sqrt{\frac{d}{rN}}$, $\lambda = \frac{N}{5^n 4}$ leads to

$$\int \|\nu - \rho^0\|_F^2 \tilde{\pi}_\lambda(d\nu) \leq A' \frac{d^2 r}{N} + C_{D_1,D_2} 5^n \frac{rd\log\left(\frac{Nr}{d}\right) + d\log\left(\frac{N}{d}\right) + 2\log(2/\epsilon)}{N}$$

for some constant $A' > 0$. Simultaneously, by Jensen inequality, one has

$$\|\hat{\rho}_\lambda - \rho^0\|_F^2 \leq \int \|\nu - \rho^0\|_F^2 \hat{\pi}_\lambda(d\nu).$$

This completes the proof of the theorem. □

References

Alquier, P., Butucea, C., Hebiri, M., Meziani, K., Morimae, T., 2013. Rank-penalized estimation of a quantum system. Phys. Rev. A 88 (3), 032113.
Alquier, P., Ridgway, J., Chopin, N., 2015. On the properties of variational approximations of gibbs posteriors. arXiv preprint arXiv:1506.04091.
Artiles, L., Gill, R., Guţă, M., 2005. An invitation to quantum tomography. J. R. Stat. Soc. Ser. B 67, 109–134.
Audenaert, K.M., Scheel, S., 2009. Quantum tomographic reconstruction with error bars: a kalman filter approach. New J. Phys. 11 (2), 023028.
Baier, T., Petz, D., Hangos, K.M., Magyar, A., 2007. Comparison of some methods of quantum state estimation. In: Quantum Probability and Infinite Dimensional Analysis. In: QP–PQ: Quantum Probab. White Noise Anal., vol. 20. World Sci. Publ., Hackensack, NJ, pp. 64–78.
Barreiro, J.T., Schindler, P., Gühne, O., Monz, T., Chwalla, M., Roos, C.F., Hennrich, M., Blatt, R., 2010. Experimental multiparticle entanglement dynamics induced by decoherence. Nat. Phys. 6 (12), 943–946.
Bégin, L., Germain, P., Laviolette, F., Roy, J.-F., 2016. Pac-bayesian bounds based on the rényi divergence. In Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, pp. 435–444.
Bissiri, P.G., Holmes, C.C., Walker, S.G., 2016. A general framework for updating belief distributions. J. R. Stat. Soc. Ser. B Stat. Methodol..
Blume-Kohout, R., 2010. Optimal, reliable estimation of quantum states. New J. Phys. 12 (4), 043034.
Butucea, C., Guţă, M., Kypraios, T., 2015. Spectral thresholding quantum tomography for low rank states. New J. Phys. 17 (11), 113050.
Butucea, C., Guţă, M., Kypraios, T., 2016. Corrigendum: Spectral thresholding quantum tomography for low rank states (2015 New J. Phys. 17 113050). New J. Phys. 18 (6), 069501.
Bužek, V., Derka, R., Adam, G., Knight, P., 1998. Reconstruction of quantum states of spin systems: From quantum bayesian inference to quantum tomography. Ann. Physics 266 (2), 454–496.
Cai, T., Kim, D., Wang, Y., Yuan, M., Zhou, H.H., 2016. Optimal large-scale quantum state tomography with pauli measurements. Ann. Statist. 44 (2), 682–712.
Carlen, E., 2010. Trace inequalities and quantum entropy: an introductory course. In: Entropy and the Quantum, Vol. 529. pp. 73–140.
Catoni, O., 2004. Statistical learning theory and stochastic optimization. In: Picard, Jean (Ed.), Saint-Flour Summer School on Probability Theory 2001, Vol. 1851. In: Lecture Notes in Mathematics, Springer-Verlag, Berlin.
Catoni, O., 2007. PAC-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning. In: IMS Lecture Notes—Monograph Series, vol. 56. Institute of Mathematical Statistics, Beachwood, OH.
Cottet, V., Alquier, P., 2016. 1-bit matrix completion: Pac-bayesian analysis of a variational approximation. arXiv preprint arXiv:1604.04191.
Dalalyan, A., Tsybakov, A.B., 2008. Aggregation by exponential weighting, sharp pac-bayesian bounds and sparsity. Mach. Learn. 72 (1–2), 39–61.
Ferrie, C., 2014. Quantum model averaging. New J. Phys. 16 (9), 093035.
Ferrie, C., Granade, C.E., 2014. Likelihood-free methods for quantum parameter estimation. Phys. Rev. Lett. 112 (13), 130402.
Flammia, S.T., Gross, D., Liu, Y.-K., Eisert, J., 2012. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. New J. Phys. 14 (9), 095022.
Ghosal, S., Ghosh, J.K., van der Vaart, A.W., 2000. Convergence rates of posterior distributions. Ann. Statist. 28 (2), 500–531.
Granade, C., Combes, J., Cory, D.G., 2016. Practical bayesian tomography. New J. Phys. 18 (3), 033024.
Gross, D., 2011. Recovering low-rank matrices from few coefficients in any basis. IEEE Trans. Inform. Theory 57 (3), 1548–1566.
Gross, D., Liu, Y.-K., Flammia, S.T., Becker, S., Eisert, J., 2010. Quantum state tomography via compressed sensing. Phys. Rev. Lett. 105 (15), 150401.
Guţă, M., Kypraios, T., Dryden, I., 2012. Rank-based model selection for multiple ions quantum tomography. New J. Phys. 14 (10), 105002.
Hradil, Z., Řeháček, J., Fiurášek, J., Ježek, M., 2004. 3 maximum-likelihood methodsin quantum mechanics. In: Quantum State Estimation. Springer, pp. 59–112.
Huszár, F., Houlsby, N.M., 2012. Adaptive bayesian quantum tomography. Phys. Rev. A 85 (5), 052120.
Koltchinskii, V., 2011. Von neumann entropy penalization and low-rank matrix estimation. Ann. Statist. 39 (6), 2936–2973.
Kravtsov, K., Straupe, S., Radchenko, I., Houlsby, N., Huszár, F., Kulik, S., 2013. Experimental adaptive bayesian tomography. Phys. Rev. A 87 (6), 062122.
Kueng, R., Ferrie, C., 2015. Near-optimal quantum tomography: estimators and bounds. New J. Phys. 17 (12), 123013.
Mai, T.T., Alquier, P., 2015. A bayesian approach for noisy matrix completion: Optimal rate under general sampling distribution. Electron. J. Stat. 9, 823–841.
McAllester, D., 1998. Some PAC-Bayesian theorems. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory. ACM, New York, pp. 230–234.
Paris, M., Řeháček, J. (Eds.), 2004. Quantum state estimation. In: Lecture Notes in Physics, vol. 649. Springer-Verlag, Berlin.
Rau, J., 2011. Inferring the gibbs state of a small quantum system. Phys. Rev. A 84, 012101.
Rau, J., 2014. Appearance of gibbs states in quantum-state tomography. Phys. Rev. A 90 (6), 062114.
Řeháček, J., Mogilevtsev, D., Hradil, Z., 2010. Operational tomography: fitting of data patterns. Phys. Rev. Lett. 105 (1), 010402.
Robert, C., Casella, G., 2013. Monte Carlo Statistical Methods. Springer Science & Business Media.
Schmied, R., 2016. Quantum state tomography of a single qubit: comparison of methods. J. Modern Opt. 1142018–15.
Schwemmer, C., Knips, L., Richart, D., Weinfurter, H., Moroder, T., Kleinmann, M., Gühne, O., 2015. Systematic errors in current quantum state tomography tools. Phys. Rev. Lett. 114, 080403.
Shang, J., Ng, H.K., Englert, B.-G., 2014. Quantum state tomography: Mean squared error matters, bias does not. arXiv preprint arXiv:1405.5350.
Shawe-Taylor, J., Williamson, R., 1997. A PAC analysis of a Bayes estimator. In: Proceedings of the Tenth Annual Conference on Computational Learning Theory. ACM, New York, pp. 2–9.
Struchalin, G., Pogorelov, I., Straupe, S., Kravtsov, K., Radchenko, I., Kulik, S., 2016. Experimental adaptive quantum tomography of two-qubit states. Phys. Rev. A 93 (1), 012103.
Suzuki, T., 2012. Pac-bayesian bound for gaussian process regression and multiple kernel additive model. In JMLR: Workshop and Conference Proceedings, vol. 23, pp. 8–1.
Vogel, K., Risken, H., 1989. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. Phys. Rev. A 40 (5), 2847.
Wallach, H.M., Mimno, D.M., McCallum, A., 2009. Rethinking LDA: Why priors matter. In: Advances in Neural Information Processing Systems. pp. 1973–1981.
Wang, Y., 2013. Asymptotic equivalence of quantum state tomography and noisy matrix completion. Ann. Statist. 41 (5), 2462–2504.
Życzkowski, K., Penson, K., Nechita, I., Collins, B., 2011. Generating random density matrices. J. Math. Phys. 52 (6), 062201.