



An efficient adaptive MCMC algorithm for Pseudo-Bayesian quantum tomography

The Tien Mai¹

Received: 23 July 2021 / Accepted: 12 July 2022
© The Author(s) 2022

Abstract

We revisit the Pseudo-Bayesian approach to the problem of estimating density matrix in quantum state tomography in this paper. Pseudo-Bayesian inference has been shown to offer a powerful paradigm for quantum tomography with attractive theoretical and empirical results. However, the computation of (Pseudo-)Bayesian estimators, due to sampling from complex and high-dimensional distribution, pose significant challenges that hamper their usages in practical settings. To overcome this problem, we present an efficient adaptive MCMC sampling method for the Pseudo-Bayesian estimator by exploring an adaptive proposal scheme together with subsampling method. We show in simulations that our approach is substantially computationally faster than the previous implementation by at least two orders of magnitude which is significant for practical quantum tomography.

Keywords Quantum tomography · Bayesian inference · Adaptive MCMC · Subsampling · Low-rank matrix

1 Introduction

Quantum state tomography is a fundamental important step in quantum information processing (Nielsen and Chuang 2000; Paris and Řeháček 2004). In general, it aims at finding the underlying density matrix which describes the given state of a physical quantum system. This task is done by utilizing the results of measurements performed on repeated state preparations (Nielsen and Chuang 2000).

Bayesian methods have been recognized as a powerful paradigm for quantum state tomography (Blume-Kohout 2010), that deal with uncertainty in meaningful and informative ways and are the most accurate approach with respect to the expected error (operational divergence) even with finite samples. Several studies

✉ The Tien Mai
the.t.mai@ntnu.no

¹ Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

have been conducted: for example, the papers Bužek et al. (1998) and Baier et al. (2007) performed numerical comparisons between Bayesian estimations with other methods on simulated data; algorithms for computing Bayesian estimators have been discussed in Kravtsov et al. (2013), Ferrie (2014), Kueng and Ferrie (2015), Schmied (2016) and Lukens et al. (2020).

Pseudo-Bayesian method for quantum tomography, introduced in Mai and Alquier (2017), proposes novel approaches for this problem with several attractive features. Importantly, a novel prior distribution for quantum density matrix is introduced based on spectral decomposition parameterization (inspired by the priors used for low-rank matrix estimation, e.g., Mai and Alquier (2015) and Cottet and Alquier (2018)). This prior can be easily used in any dimension and is found to be significantly more efficient to sample from and evaluate than the Cholesky approach in Struchalin et al. (2016), Zyczkowski et al. (2011) and Seah et al. (2015), see Lukens et al. (2020) for more details. By replacing the likelihood with a loss function between a proposed density matrix and experimental data, the paper Mai and Alquier (2017) presents two different estimators: the prob-estimator and the dens-estimator.

However, the reference Mai and Alquier (2017) proposed simply to compute approximately these two Pseudo-Bayesian estimators by naive Metropolis-Hastings (MH) algorithms which are computationally very slow for high-dimensional systems. Recently, a faster and more efficient sampling method has been proposed for the dens-estimator, see Lukens et al. (2020). However, we would like to note that the prob-estimator is shown in Mai and Alquier (2017) to reach the best known up-to-date rate of convergence (Butucea et al. 2015) while the theoretical guarantee for the dens-estimator is far less satisfactory. Moreover, it is also shown in their simulations that the prob-estimator yields better results compared to the dens-estimator.

In this paper, we present a novel efficient adaptive Metropolis-Hastings implementation for the prob-estimator. This adaptive implementation is based on considering the whole density matrix as a parameter that needs to be sampled at a time. Moreover, an adaptive proposal is explored based on the “preconditioned Crank-Nicolson” (Cotter et al. 2013) sampling procedure that can eliminate the “curse of dimensionality”, which is the case for quantum state tomography where the dimension increases exponentially. Further speeding up by using subsampling MCMC approach is also explored in our work.

Through simulations, it is shown that our implementation is computationally significantly faster than the naive MH algorithm in Mai and Alquier (2017). More specifically, for example as in a system of 6 qubits, our algorithm is around 115 times faster than the naive MH algorithm in Mai and Alquier (2017). While in term of accuracy, we show that our algorithms returns similar results with less variation.

The rest of the paper is organized as follow. In Sect. 2, we provide the necessary background and the statistical model for the problem of quantum state tomography. In Sect. 3, we recall the Pseudo-Bayesian approach and the prior distribution. Section 4 presents our novel adaptive MCMC implementation for the Pseudo-Bayesian estimator. Simulation studies are presented in Sect. 5. Conclusions are given in Sect. 6.

2 Background

2.1 The quantum state tomography problem

Hereafter, we only provide the necessary background on quantum state tomography (QST) required for the paper. We would like to remind that a very nice introduction to this problem, from a statistical perspective, can be found in Artiles et al. (2005). Here, we have opted for the notations used in reference Mai and Alquier (2017).

Mathematically speaking, a two-level quantum system of n -qubits is characterized by a $2^n \times 2^n$ density matrix ρ whose entries are complex, i.e. $\rho \in \mathbb{C}^{2^n \times 2^n}$. For the sake of simplicity, put $d = 2^n$, so ρ is a $d \times d$ matrix. This density matrix must be

- Hermitian: $\rho^\dagger = \rho$ (i.e. self-adjoint),
- positive semi-definite: $\rho \succeq 0$,
- normalized: $\text{Trace}(\rho) = 1$.

In addition, physicists are especially interested in pure states and that a pure state ρ can be defined in addition by $\text{rank}(\rho) = 1$. In practice, it often makes sense to assume that the rank of ρ is small (Gross et al. 2010; Gross 2011; Butucea et al. 2015).

The goal of quantum tomography is to estimate the underlying density matrix ρ using measurement outcomes of many independent and identically systems prepared in the state ρ by the same experimental devices.

For a qubit, it is a standard procedure to 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 is given in (1) below). As a consequence, with a n -qubits system, there are 3^n possible experimental observables. The set of all possible performed observables is

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

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

Denote $R^{\mathbf{a}}$ a random vector that is the outcome of an experiment indexed by \mathbf{a} . From the Born's rule (Nielsen and Chuang 2000), its probability distribution is given by

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

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

Statistically, for each experiment $\mathbf{a} \in \mathcal{E}^n$, the experimenter repeats m times the experiment corresponding to \mathbf{a} and thus collects m independent random copies of

$R^{\mathbf{a}}$, say $R_1^{\mathbf{a}}, \dots, R_m^{\mathbf{a}}$. As there are 3^n possible experimental settings \mathbf{a} , we define the **quantum sample** size as $N := m \cdot 3^n$. We will refer to $(R_i^{\mathbf{a}})_{i \in \{1, \dots, m\}, \mathbf{a} \in \mathcal{E}^n}$ as \mathcal{D} (for data). Therefore, quantum state tomography is aiming at estimating the density matrix ρ based on the data \mathcal{D} .

2.2 Popular estimation methods

Here, we briefly recall three classical major approaches that have been adopted to estimate ρ , which are: linear inversion, maximum likelihood and Bayesian inference.

2.2.1 Linear inversion

The first and simplest method considered in quantum information processing is the 'tomographic' method, also known as linear/direct inversion (Vogel and Risken 1989; Řeháček et al. 2010). It is actually the analogue of the least-square estimator in the quantum setting. This method relies on the fact that measurement outcome probabilities are linear functions of the density matrix.

More specifically, let us consider the empirical frequencies

$$\hat{p}_{\mathbf{a},\mathbf{s}} = \frac{1}{m} \sum_{i=1}^m \mathbf{1}_{\{R_i^{\mathbf{a}}=\mathbf{s}\}}. \quad (2)$$

It is noted that $\hat{p}_{\mathbf{a},\mathbf{s}}$ is an unbiased estimator of the underlying probability $p_{\mathbf{a},\mathbf{s}}$ in (1). Therefore, the inversion method is based on solving the linear system of equations

$$\begin{cases} \hat{p}_{\mathbf{a},\mathbf{s}} = \text{Trace}(\hat{\rho} \cdot P_{\mathbf{s}}^{\mathbf{a}}), \\ \mathbf{a} \in \mathcal{E}^n, \quad \mathbf{s} \in \{-1, 1\}^n. \end{cases} \quad (3)$$

As mentioned above, the computation of $\hat{\rho}$ is quite clear and explicit formulas are classical that can be found for example in Alquier et al. (2013). While straightforward and providing unbiased estimate (Schwemmer et al. 2015), it tends to generate a non-physical density matrix as an output (Shang et al. 2014): positive semi-definiteness cannot easily be satisfied and enforced.

2.2.2 Maximum likelihood

A popular approach in QST in recent years is the maximum likelihood estimation (MLE). MLE aims at finding the density matrix which is most likely to have produced the observed data \mathcal{D} :

$$\rho_{MLE} = \arg \max L(\rho; \mathcal{D})$$

where $L(\rho; \mathcal{D})$ is likelihood, the probability of observing the outcomes given state ρ , as defined by some models (Hradil et al. 2004; James et al. 2001; Gonçalves et al. 2018). However, it has some critical problems, detailed in Blume-Kohout (2010), including a huge computational cost. Moreover, it is a point estimate which does not account the level of uncertainty in the result.

Furthermore, these two methods (Linear inversion and MLE) can not take advantage of a prior knowledge where a system is in a state ρ for which some additional information is available. More particularly, it is noted that physicists usually focus on so-called pure states, for which $\text{rank}(\rho) = 1$.

2.2.3 Bayesian inference

Starting receiving attention in recent years, Bayesian QST had been shown as a promising method in this problem (Blume-Kohout 2010; Bužek et al. 1998; Baier et al. 2007; Lukens et al. 2020). Through Bayes' theorem, experimental uncertainty is explicitly accounted in Bayesian estimation. More specifically, suppose a density matrix ρ is parameterized by $\rho(x)$ for some x , Bayesian inference is carried out via the posterior distribution

$$\pi(\rho(x)|\mathcal{D}) \propto L(\rho(x); \mathcal{D})\pi(x),$$

where $L(\rho(x); \mathcal{D})$ is the likelihood (as in MLE) and $\pi(x)$ is the prior distribution. Using the posterior distribution $\pi(\rho(x)|\mathcal{D})$, the expectation value of any function of ρ can be inferred, e.g. the Bayesian mean estimator as $\int \rho(x)\pi(\rho(x)|\mathcal{D})dx$.

Although recognized as a powerful approach, the numerical challenge of sampling from a high-dimensional probability distribution prevents the widespread use of Bayesian methods in the physical problem.

2.2.4 Other approaches

Several other methods have also recently introduced and studied. The reference Cai et al. (2016) proposed a method based on the expansion of the density matrix ρ in the Pauli basis. Some rank-penalized approaches were studied in Guță et al. (2012) and Alquier et al. (2013). A thresholding method is introduced in Butucea et al. (2015).

3 Pseudo-Bayesian quantum state tomography

3.1 Pseudo-Bayesian estimation

Let us consider the pseudo-posterior, studied in Mai and Alquier (2017), defined by

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

where $\exp[-\lambda \ell(\nu, \mathcal{D})]$ is the pseudo-likelihood that plays the role of the empirical evidence to give more weight to the density ν when it fits the data well; $\pi(d\nu)$ is the prior given in Sect. 3.2; and $\lambda > 0$ is a tuning parameter that balances between evidence from the data and prior information.

Taking, with $\hat{\rho}_{\text{a,s}}$ given in (2),

$$\ell(v, \mathcal{D}) := \ell^{prob}(v, \mathcal{D}) = \sum_{\mathbf{a} \in \mathcal{E}^n} \sum_{\mathbf{s} \in \{-1, 1\}^n} [\text{Tr}(v P_{\mathbf{s}}^{\mathbf{a}}) - \hat{p}_{\mathbf{a}, \mathbf{s}}]^2,$$

the “prob-estimator” in Mai and Alquier (2017) is defined as the mean estimator of the pseudo-posterior:

$$\tilde{\rho}_{\lambda}^{prob} = \int v \exp[-\lambda \ell^{prob}(v, \mathcal{D})] \pi(dv). \quad (4)$$

This estimator also referred to, in statistical machine learning, as Gibbs estimator, PAC-Bayesian estimator or EWA (exponentially weighted aggregate) (Catoni 2007; Dalalyan and Tsybakov 2008).

For the sake of simplicity, we use the shortened notation $p_v := [\text{Tr}(v P_{\mathbf{s}}^{\mathbf{a}})]_{\mathbf{a}, \mathbf{s}}$ and $\hat{p} := [\hat{p}_{\mathbf{a}, \mathbf{s}}]_{\mathbf{a}, \mathbf{s}}$ then

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

($\|\cdot\|_F$ is the Frobenius norm). Clearly, we can see that this distance measures the difference between the probabilities given by a density v and the empirical frequencies in the sample. Here, the readers could see that the tuning parameter λ is used to control the difference between the empirical frequencies and the hypothetical one from the prior distribution. We remind the reader that the matrix $[\hat{p}_{\mathbf{a}, \mathbf{s}}]_{\mathbf{a}, \mathbf{s}}$ is of dimension $3^n \times 2^n$.

Remark 1 This kind of pseudo-posterior is an increasingly popular approach in Bayesian statistics and machine learning, see for example Bissiri et al. (2016), Mai (2021b), Grünwald and Van Ommen (2017), Catoni (2007), Mai (2022), Alquier et al. (2016b), Mai (2021a) and Bégin et al. (2016), for models with intractable likelihood or for misspecification models.

3.2 Prior distribution for quantum density matrix

The prior distribution employed in Mai and Alquier (2017) can be expressed as follow: the $d \times d$ density matrix ρ can be parameterized by d non-negative real numbers y_i and d complex column vectors of length d , z_i . Put $x = \{y_1, \dots, y_d, z_1, \dots, z_d\}$, then the density matrix is

$$\rho(x) = \sum_{i=1}^d \frac{y_i}{\sum_{\ell} y_{\ell}} \frac{z_i z_i^{\dagger}}{\|z_i\|^2}, \quad (5)$$

with the prior distribution for x as

$$\pi(x) \propto \prod_{i=1}^d y_i^{\alpha-1} e^{-y_i} e^{-\frac{1}{2} z_i^{\dagger} z_i} \quad (6)$$

where the weights are being treated as Gamma-distributed random variables $Y_i \stackrel{i.i.d.}{\sim} \Gamma(\alpha, 1)$, and the vectors z_i are standard-normal complex Gaussian distributed $Z_i \stackrel{i.i.d.}{\sim} \mathcal{CN}(0, I_d)$.

The tuning parameter α in (6) allows the user to favor low-rank or high-rank density matrices which are corresponding to pure or mixed states, respectively. More particularly, the normalized random variables $Y_i/(\sum Y_j)$ with $Y_i \stackrel{i.i.d.}{\sim} \Gamma(\alpha, 1)$ follows a Dirichlet distribution $\text{Dir}(\alpha)$ which ensures both normalization and non-negativity. An $\alpha < 1$ promotes sparse draws and thus purer states, while $\alpha = 1$ returns a fully uniform prior on all physically realizable states.

Remark 2 It is noted that this parameterization satisfies all physical conditions for the density matrix, details can be found in Mai and Alquier (2017). Moreover, this parameterization have been shown to be significantly more efficient to sample from and to evaluate than the Cholesky approach in references Struchalin et al. (2016), Zyczkowski et al. (2011) and Seah et al. (2015), see Lukens et al. (2020) for details.

Remark 3 The theoretical guarantees for the “prob-estimator” in (4) are validated only for $0 < \alpha \leq 1$. More specifically, the prob-estimator satisfies (up to a multiplicative logarithmic factor) that $\|\hat{\rho}_{\lambda^*}^{prob} - \rho^0\|_F^2 \leq c3^n \text{rank}(\rho^0)/N$ which is the best known up-to-date rate in the problem of quantum state estimation (Butucea et al. 2015), where c is a numerical constant and $\lambda^* = m/2$.

4 A novel efficient adaptive MCMC implementation

Appropriately, the prob-estimator requires an evaluation of the integral (4) which is numerically challenging due to its sophisticated features and high dimensionality. A first attempt has been done in Mai and Alquier (2017) is to use a naive Metropolis-Hastings (MH) algorithm where the authors iterate between a random walk MH for $\log(y_i)$ and an independent MH for z_i . Typically, the approach is designed to obtain T samples $x^{(1)}, \dots, x^{(T)}$ as a consequence the integral (4) can be approximated as

$$\hat{\rho}^{\text{MH}} \approx \frac{1}{T} \sum_{t=1}^T \rho(x^{(t)}).$$

However, as also noted in the reference Mai and Alquier (2017), their proposed algorithm can run into slow convergence and can be arbitrarily slow as the system dimensionality increases. In this paper, we propose a novel efficient MCMC algorithm for the prob-estimator through exploring the adaptive proposal and subsampling scheme.

4.1 A preconditioned Crank-Nicolson adaptive proposal

Borrowing motivation from the recent work in Lukens et al. (2020) that proposes an efficient sampling procedure for Bayesian quantum state estimation (which improve the

computation of the “dens-estimator” in Mai and Alquier (2017) only), we introduce an efficient adaptive Metropolis-Hastings implementation for the prob-estimator in Mai and Alquier (2017). We remind that the prob-estimator shows better performance than the dens-estimator both in theory and simulations.

Specifically, we propose to use a modification of random-walk MH by scaling the previous step before adding a random move and generating the proposal z' . Following Cotter et al. (2013) who introduced an efficient MCMC approach eliminating the “curse of dimensionality”, termed as “preconditioned Crank-Nicolson”, we use the proposal for z_j as

$$z'_j = \sqrt{1 - \beta_z^2} z_j^{(k)} + \beta_z \xi_j, \quad \xi_j \stackrel{\text{i.i.d.}}{\sim} \mathcal{CN}(0, I_d)$$

where $\beta_z \in (0, 1)$ is a tuning parameter. This proposal is a scaled, by the factor $\sqrt{1 - \beta_z^2}$, random walk that results in a slightly simpler acceptance probability. Unlike the independent proposal in Mai and Alquier (2017) (with $\beta_z = 1$) where the acceptance probability can vary substantially, this kind of adaptive proposal allows one to control the acceptance rate efficiently. For $\beta_y \in (0, 1)$, we slightly modify the proposal for y from Mai and Alquier (2017) (with $\beta_y = 1$) as

$$y'_j = y_j^{(j)} e^{\beta_y \eta_j}, \quad \eta_j \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(-0.5, 0.5).$$

The acceptance ratio $\min\{1, A(x'|x^{(k)})\}$ are followed from the standard form for MH (Robert and Casella 2013). Let $p(x'|x^{(k)})$ denote the proposal density, we have

$$A(x'|x^{(k)}) = \frac{\tilde{\pi}(\rho(x')) p(x^{(k)}|x')}{\tilde{\pi}(\rho(x^{(k)})) p(x'|x^{(k)})},$$

where

$$\log A(x'|x^{(k)}) = \log L_D(x') - \log L_D(x^{(k)}) + \sum_{j=1}^d \left[\alpha \log y'_j - y'_j - \alpha \log y_j^{(k)} + y_j^{(k)} \right].$$

4.2 Speeding up by subsampling

We remind that the log (pseudo-)likelihood, $\log L_D(x) = -\lambda \|p_v - \hat{p}\|_F^2$, is the Frobenius norm of a matrix of dimension $3^n \times 2^n$ and thus for large n it will be very costly to evaluate at each iteration. For example, with $n = 7$, this matrix is of dimension 2187×128 . Therefore, we propose to evaluate a random subset of this matrix at each iteration. More precisely, at each iteration, we draw uniformly at random a subset Ω of indices of the $3^n \times 2^n$ matrix. Then, the log pseudo-likelihood, $\log L_D(x)$, is approximated by

$$\log L_\Omega(x) := \sum_{(\mathbf{a}, \mathbf{s}) \in \Omega \subset \mathcal{E}^n \times \{-1, 1\}^n} [\text{Tr}(v P_{\mathbf{s}}^{\mathbf{a}}) - \hat{p}_{\mathbf{a}, \mathbf{s}}]^2,$$

As a consequence, the acceptance rate corresponding with this subsampling is denoted by $A_{\Omega}(x'|x^{(k)})$. It is noted that this kind of using subsampling to speeding up MCMC algorithms is becoming popular in the computational statistics community, see for example Quiroz et al. (2018b), Maire et al. (2019), Quiroz et al. (2018a) and Alquier et al. (2016a).

The details of our novel adaptive MH is given in Algorithm 1.

Algorithm 1 Adaptive MH for Pseudo-Bayesian QST

Input: The tuning parameters $\beta_y, \beta_z \in (0, 1)$.

Parameters: Positive real numbers $\alpha \in (0, 1], T$.

Output: The density matrix $\hat{\rho}$.

Initialize: $x^{(0)}$ drawn from the prior (6), $\hat{\rho} = \mathbf{0}_{d \times d}$.

for $k = 1$ to T **do**

Simulate a new point $x' = \{y'_1, \dots, y'_d; z'_1, \dots, z'_d\}$, according to

$$\begin{aligned} y'_j &= y_j^{(j)} e^{\beta_y \eta_j}, \\ z'_j &= \sqrt{1 - \beta_z^2} z_j^{(k)} + \beta_z \xi_j, \end{aligned} \quad j \in \{1, \dots, d\},$$

where $\eta_j \stackrel{i.i.d.}{\sim} \text{Uniform}(-0.5, 0.5)$ and independently $\xi_j \stackrel{i.i.d.}{\sim} \mathcal{CN}(0, I_d)$.

Draw uniformly at random a subset of indices Ω .

Set $x^{(k+1)} = x'$ with probability $\min\{1, A_{\Omega}(x'|x^{(k)})\}$, otherwise set $x^{(k+1)} = x^{(k)}$.

$\hat{\rho} \leftarrow \hat{\rho} + \rho(x^{(k)})/T$.

end for

5 Numerical studies

5.1 Simulations setups and details

To assess the performance of our new proposed algorithm, a series of experiments were conducted with simulated tomographic data. More particularly, we consider the following setting for choosing the true density matrix, with $n = 2, 3, 4$, ($d = 4, 8, 16$):

- Setting 1: we consider the ideal entangled state which is characterized by 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 $\psi_1 = u/\|u\|$ and $u = (u_1, \dots, u_{d/2}, 0, \dots, 0)$, $u_1 = \dots = u_{d/2} = 1$; $\psi_2 = v/\|v\|$ and $v = (0, \dots, 0, v_{d/2+1}, \dots, v_d)$, $v_{d/2} = \dots = v_d = 1$.

- Setting 2: a maximal mixed state (rank- d) that is

$$\rho_{mixed} = \sum_{i=1}^d \frac{1}{d} \psi_i \psi_i^\dagger,$$

with ψ_i are normalized vectors and independently simulated from $\mathcal{CN}(0, I_d)$.

The experiments are done following Sect. 2 for $m = 1000$. The prob-estimator is employed with $\lambda = m/2$ and a prior with $\alpha = 1$ which are theoretically guaranteed from Theorem 1 in reference Mai and Alquier (2017). We compare our adaptive MH implementation, denoted by “a-MH”, against the (random-walk) in Mai and Alquier (2017), denoted by “MH”; where all algorithms are run with 1000 iterations and 200 burnin steps. We run 50 independent samplers for each algorithm, and compute the mean of the square error (MSE),

$$\text{MSE} := \|\hat{\rho} - \rho\|_F^2/d^2$$

for each method, together with their standard deviations. We also measure the mean absolute error of eigen values (MAEE) by

$$\text{MAEE} := \frac{1}{d} \sum_{i=1}^d |\lambda_i(\hat{\rho}) - \lambda_i(\rho)|,$$

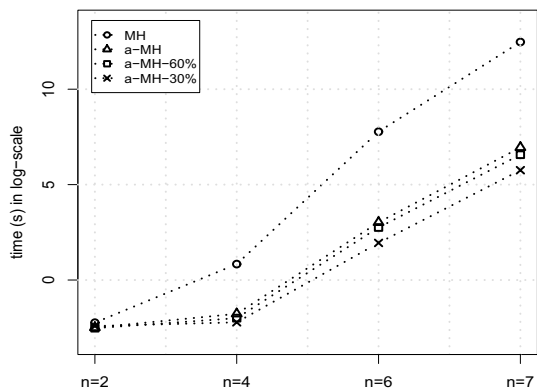
where $\lambda_i(A)$ are the eigen values of the matrix A .

5.2 Significantly speeding up

From Fig. 1, it is clear to see that our adaptive MH implementation is greatly faster than the previous implementation from Mai and Alquier (2017) by at least two orders of magnitude as the number of qubits increase. The data are simulated as in Setting 1 for $n = 2, 4, 6, 7$ for which the dimensions of the density matrix are $d = 4, 16, 64, 128$ and of the empirical frequencies matrices $[\hat{\rho}_{a,s}]$ are $9 \times 4, 81 \times 16, 729 \times 64, 2187 \times 128$. More specifically, for $n = 6$, our adaptive MH gives ~ 115.9 times speedup comparing to the naive “MH” algorithm in Mai and Alquier (2017) and for $n = 7$ the speedup is ~ 251.1 times.

In addition, subsampling approaches also save the computational times respectively with the volume of the subsets, for example “a-MH-30%” will save the computational time by 2/3 while “a-MH-60%” will save the computational time by 1/3 of the full data approach “a-MH”. We note that these improvements are quite

Fig. 1 Plot to compare the running times (s) in log-scale for 10 steps of two algorithms in the setup of Setting 1, for the qubits $n = 2, 4, 6, 7$ ($d = 4, 16, 64, 128$). “MH” is from reference Mai and Alquier (2017); “a-MH” is the Algorithm 1 without subsampling; “a-MH-30%” is the Algorithm 1 with subsampling 30%; “a-MH-60%” is the Algorithm 1 with subsampling 60%



significant for practical quantum tomography where computational time is a precious resource.

5.3 Tuning parameters via acceptance rate

The tuning parameters β_y, β_z are chosen such that the acceptance rate of Algorithm 1 is between 0.15 and 0.3. This interval is chosen to enclose 0.234, the optimum acceptance probability for random-walk Metropolis-Hastings (under assumptions) (Gelman et al. 1997). For example, as in our experiments, for $n = 2$ qubits: $\beta_y = 0.33, \beta_z = 0.2$; for $n = 3$ qubits: $\beta_y = 0.03, \beta_z = 0.03$ and for $n = 4$ qubits: $\beta_y = 0.03, \beta_z = 0.02$ (all are run with $\alpha = 1, \lambda = m/2$). We note that as the number of qubits n increase, these tuning parameters tend to be smaller and smaller to assure that the acceptance rate is between 0.15 and 0.3.

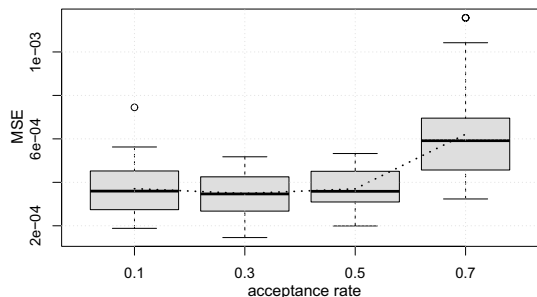
As an illustration, we conduct some simulations with $n = 4$ qubits in Setting 2. It can be seen from Fig. 2 that the acceptance rate between 0.2 and 0.3 would be optimal, as in Gelman et al. (1997). Where as high acceptance rate like 0.7 could make the algorithm be trapped at local points, and very small acceptance rate as 0.1 could make the algorithm converge slower.

5.4 Similar accuracy with less variation

In term of accuracy performance, Fig. 3 compares the performance of our “a-MH” algorithm with the “MH” algorithm in various settings and varying the number of qubits $n = 2, 3, 4, 5$. The results show that both algorithms share similar accuracy in term of both considered errors (MSE and MAEE). However, it shows a clear improvement that our proposed adaptive algorithm yields much stable results (with less variation) compared to the naive MH approach as expected.

Results on subsampling are given in Fig. 4, where we further examine the performance of the “a-MH” algorithm with full data against subsampling the data by 60% and 30%. The outputs show that the subsampling approaches return comparable results. More specifically, in the cae of low-rankness (Setting 1), the subsampling approaches share similar accuracy (with higher variation) with the full data approach. In the case of mixed state (Setting 2), the subsampling approaches seem to return smaller error in term of mean squared errors, however their mean absolute

Fig. 2 Boxplots to examine the effect of the acceptance rate to MSE. The simulations are run within Setting 2 for $n = 4$



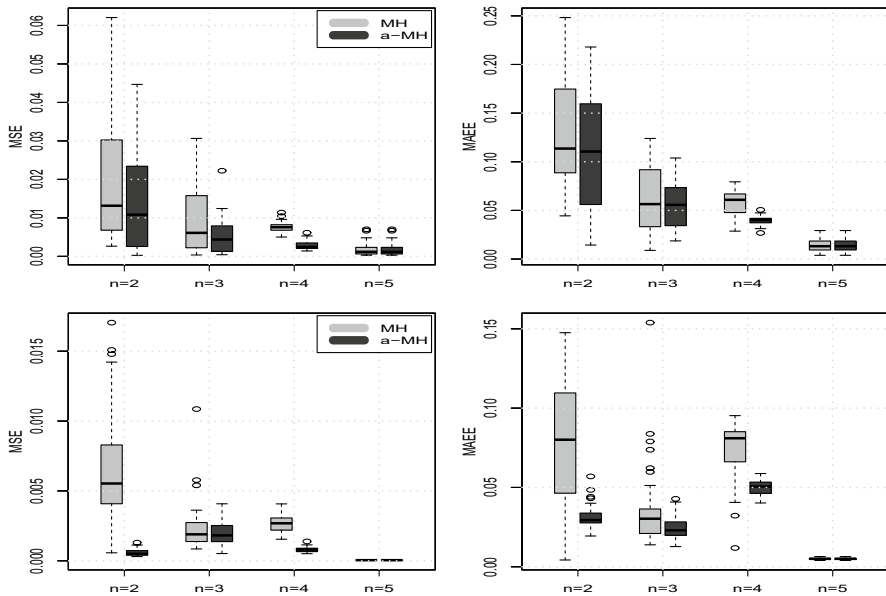


Fig. 3 Plots to compare the errors of two algorithms in different settings and with varying the number of qubits $n = 2, 3, 4, 5$. The top 2 boxplots from left to right are in Setting 1, the bottom 2 boxplots from left to right are in Setting 2. “MH” is from reference Mai and Alquier (2017); “a-MH” is the Algorithm 1 with full data

error of eigen values (MAEE) are slightly higher than the full data approach. This can be explained as the target distribution of the subsampling algorithm is just an approximation of the target distribution in the full setting, thus the mean posterior could be well approximated but higher variation, (Quiroz et al. 2018b; Maire et al. 2019; Quiroz et al. 2018a; Alquier et al. 2016a).

Additional simulation regarding sensitivity analysis for different values of λ and α are given in Fig. 5 in the Appendix.

6 Discussion and conclusion

We have introduced an efficient sampling algorithm for Pseudo-Bayesian quantum tomography, especially for the prob-estimator. Our approach use a preconditioned proposal and subsampling Metropolis-Hasting implementation which shows a clear improvement in convergence, computation and computational time comparing with a naive MH implementation. We would like to mention that such an improvement is significantly important for practical quantum state tomography.

As suggested by one of the anonymous reviewer, in practice, one could change the tuning parameters β_y, β_z in Algorithm 1 dynamically adaptively. For example, one could change the values of these parameter every fixed steps (say 500 or

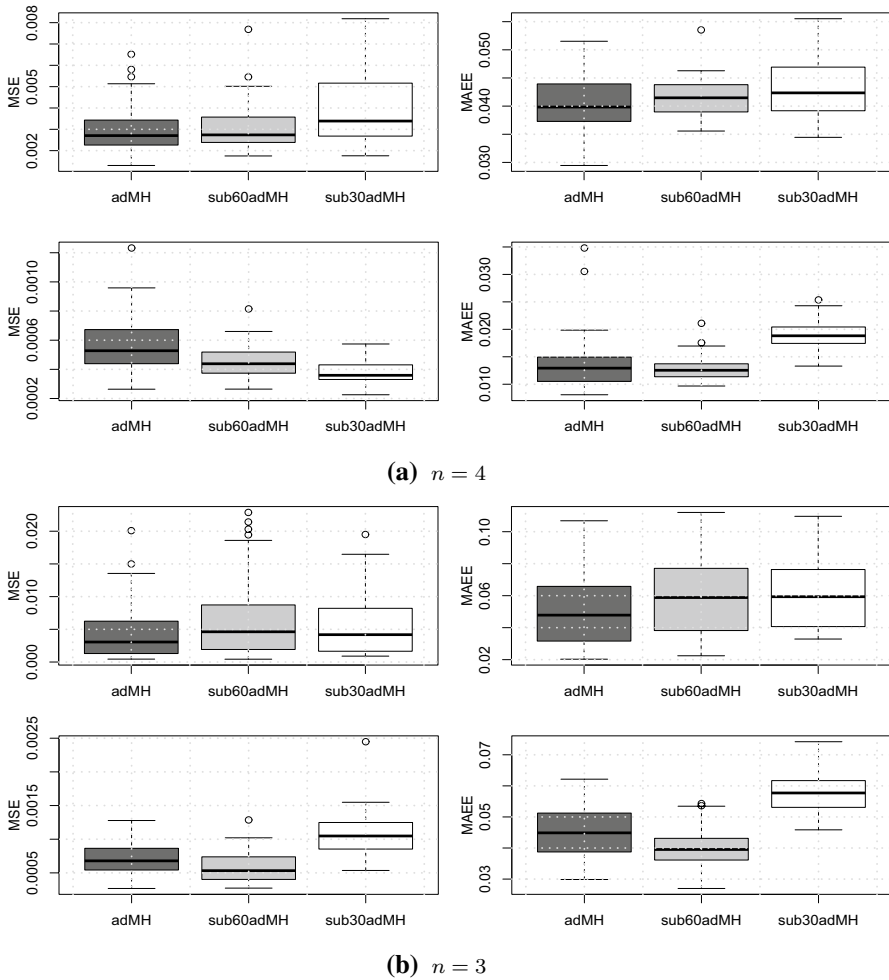


Fig. 4 Plots to compare the error of two algorithms in different settings and with $n = 4$. The top 2 boxplots from left to right are in Setting 1, the bottom 2 boxplots from left to right are in Setting 2. “adMH” is the Algorithm 1 with full data; “sub30adMH” is “adMH” with subsampling 30% the data; “sub60adMH” is “adMH” with subsampling 60% the data

1000 steps) so that the acceptance rate is between 0.15 and 0.3. This could be an important step to obtain a better mixing rate in the chain.

Last but not least, faster algorithms based on optimization, such as Variational inference (Alquier et al. 2016b), for Bayesian quantum tomography would be an interesting research problem. However, it should be noted that the analysis of the uncertainty quantification when using Variational inference is not known up to present, while this matter is an important aspect in the problem of quantum state estimation.

Appendix: Additional simulations for changing λ and α

Here we further present some additional simulations for different values of λ and α . The simulations are carried out as in Setting 2 for 3 qubits. The results are given in Fig. 5. These results show that $\lambda = m/2$ would be an optimal choice while the effect of α is not clearly determined.

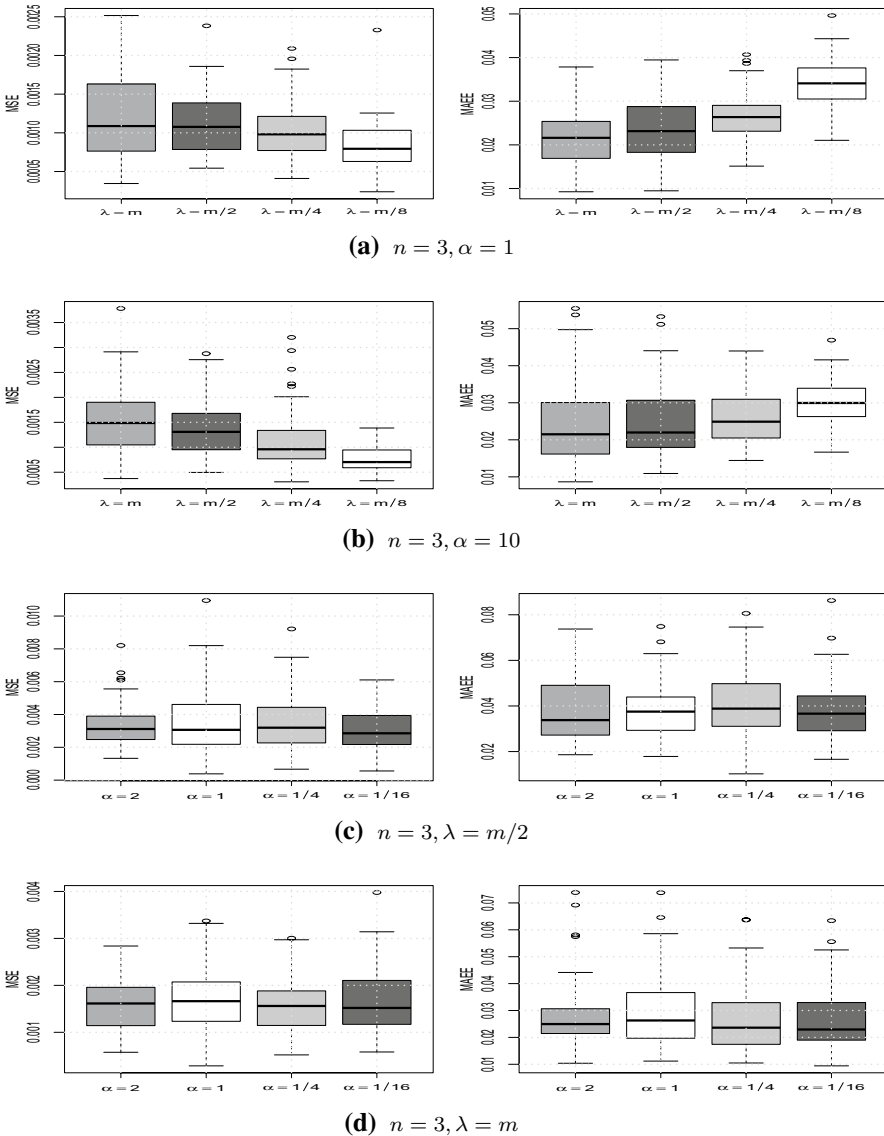


Fig. 5 Plots to compare the errors of Algorithm 1 in Setting 1 and with $n = 3$

Acknowledgements TTM is supported by the Norwegian Research Council grant number 309960 through the Centre for Geophysical Forecasting at NTNU. I would like to warmly thank the anonymous referees who kindly reviewed the earlier version of this manuscript and provided valuable suggestions and enlightening comments that significantly improve the current version of the manuscript.

Funding Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital).

Data availability statement The R codes and data used in the numerical experiments are available at: <https://github.com/tienmt/bqst>.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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, Friel N, Everitt R, Boland A (2016a) Noisy monte carlo: convergence of markov chains with approximate transition kernels. *Stat Comput* 26(1–2):29–47
- Alquier P, Ridgway J, Chopin N (2016b) On the properties of variational approximations of gibbs posteriors. *J Mach Learn Res* 17(1):8374–8414
- Artiles L, Gill R, Guță M (2005) An invitation to quantum tomography. *J R Stat Soc Ser B* 67:109–134
- Baier T, Petz D, Hangos KM, Magyar A (2007) Comparison of some methods of quantum state estimation. In: *Quantum probability and infinite dimensional analysis, QP–PQ: Quantum Probab. White Noise Anal.*, vol 20, World Sci. Publ., Hackensack, pp 64–78, https://doi.org/10.1142/9789812770271_0007
- Bégin L, Germain P, Laviolette F, Roy JF (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 PG, Holmes CC, Walker SG (2016) A general framework for updating belief distributions. *J R Stat Soc Ser B (Stat Methodol)*. <https://doi.org/10.1111/rssb.12158>
- Blume-Kohout R (2010) Optimal, reliable estimation of quantum states. *N J Phys* 12(4):043034
- Butucea C, Guță M, Kypraios T (2015) Spectral thresholding quantum tomography for low rank states. *N J Phys* 17(11):113050. <http://stacks.iop.org/1367-2630/17/i=11/a=113050>
- Bužek V, Derka R, Adam G, Knight P (1998) Reconstruction of quantum states of spin systems: from quantum bayesian inference to quantum tomography. *Ann Phys* 266(2):454–496
- Cai T, Kim D, Wang Y, Yuan M, Zhou HH (2016) Optimal large-scale quantum state tomography with pauli measurements. *Ann Statist* 44(2):682–712. <https://doi.org/10.1214/15-AOS1382>
- Catoni O (2007) PAC-Bayesian supervised classification: the thermodynamics of statistical learning. *IMS Lecture Notes—Monograph Series*, 56, Institute of Mathematical Statistics, Beachwood
- Cotter SL, Roberts GO, Stuart AM, White D (2013) MCMC methods for functions: modifying old algorithms to make them faster. *Stat Sci* 28:424–446

- Cottet V, Alquier P (2018) 1-bit matrix completion: Pac-bayesian analysis of a variational approximation. *Mach Learn* 107(3):579–603
- Dalalyan A, Tsybakov AB (2008) Aggregation by exponential weighting, sharp pac-bayesian bounds and sparsity. *Mach Learn* 72(1–2):39–61
- Ferrie C (2014) Quantum model averaging. *N J Phys* 16(9):093035
- Gelman A, Gilks WR, Roberts GO (1997) Weak convergence and optimal scaling of random walk metropolis algorithms. *Ann Appl Probab* 7(1):110–120
- Gonçalves D, Azevedo C, Lavor C, Gomes-Ruggiero M (2018) Bayesian inference for quantum state tomography. *J Appl Stat* 45(10):1846–1871
- Gross D (2011) Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans Inf Theory* 57(3):1548–1566. <https://doi.org/10.1109/TIT.2011.2104999>
- Gross D, Liu YK, Flammia ST, Becker S, Eisert J (2010) Quantum state tomography via compressed sensing. *Phys Rev Lett* 105(15):150401
- Grünwald P, Van Ommen T et al (2017) Inconsistency of bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Anal* 12(4):1069–1103
- Guță M, Kypraios T, Dryden I (2012) Rank-based model selection for multiple ions quantum tomography. *N J Phys* 14(10):105002
- Hradil Z, Řeháček J, Fiurášek J, Ježek M (2004) 3 maximum-likelihood methods in quantum mechanics. In: *Quantum state estimation*, Springer, pp 59–112
- James DFV, Kwiat PG, Munro WJ, White AG (2001) Measurement of qubits. *Phys Rev A* 64:052312. <https://doi.org/10.1103/PhysRevA.64.052312>
- Kravtsov K, Straupe S, Radchenko I, Houlshby 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. *N J Phys* 17(12):123013. <http://stacks.iop.org/1367-2630/17/i=12/a=123013>
- Lukens JM, Law KJ, Jasra A, Lougovski P (2020) A practical and efficient approach for bayesian quantum state estimation. *N J Phys* 22(6):063038
- Mai TT (2021a) Efficient bayesian reduced rank regression using langevin monte carlo approach. arXiv preprint [arXiv:2102.07579](https://arxiv.org/abs/2102.07579)
- Mai TT (2021b) Numerical comparisons between bayesian and frequentist low-rank matrix completion: estimation accuracy and uncertainty quantification. arXiv preprint [arXiv:2103.11749](https://arxiv.org/abs/2103.11749)
- Mai TT (2022) Pac-bayesian matrix completion with a spectral scaled student prior. In: *The 4th symposium on advances in approximate bayesian inference*
- Mai TT, Alquier P (2015) A bayesian approach for noisy matrix completion: optimal rate under general sampling distribution. *Electron J Stat* 9:823–841. <https://doi.org/10.1214/15-EJS1020>
- Mai TT, Alquier P (2017) Pseudo-bayesian quantum tomography with rank-adaptation. *J Stat Plan Inference* 184:62–76
- Maire F, Friel N, Alquier P (2019) Informed sub-sampling mcmc: approximate bayesian inference for large datasets. *Stat Comput* 29(3):449–482
- Nielsen MA, Chuang IL (2000) *Quantum Comput Quantum Inf*. Cambridge University Press, Cambridge
- Paris M, Řeháček J (2004) *Quantum state estimation*, Lecture Notes in Physics, vol 649. Springer-Verlag, Berlin. <https://doi.org/10.1007/b98673>
- Quiroz M, Kohn R, Villani M, Tran MN (2018a) Speeding up mcmc by efficient data subsampling. *J Am Stat Assoc* 11(526)
- Quiroz M, Villani M, Kohn R, Tran MN, Dang KD (2018b) Subsampling mcmc—an introduction for the survey statistician. *Sankhya A* 80(1):33–69
- 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, New York
- Schmied R (2016) Quantum state tomography of a single qubit: comparison of methods. *J Mod Opt* 1142018:1–15. <https://doi.org/10.1080/09500340.2016.1142018>
- 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. <https://doi.org/10.1103/PhysRevLett.114.080403>
- Seah YL, Shang J, Ng HK, Nott DJ, Englert BG (2015) Monte carlo sampling from the quantum state space. ii. *N J Phys* 17(4):043018
- Shang J, Ng HK, Englert BG (2014) Quantum state tomography: Mean squared error matters, bias does not. arXiv preprint [arXiv:1405.5350](https://arxiv.org/abs/1405.5350)

- 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
- 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
- Zyczkowski K, Penson K, Nechita I, Collins B (2011) Generating random density matrices. *J Math Phys* 52(6):062201

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.