A projected Langevin sampling algorithm for quantum tomography

Tameem Adel, Stéphane Chrétien, Estelle Massart, Andrew Thompson

January 12, 2024

1 Introduction

Quantum tomography amounts to find a density matrix $\rho \in \mathbb{C}^{d\times d}$, with $d = 2^n$, characterizing the unknown quantum state of a 2-level n-qubits system, given a set of measurements of repeated state preparation. While maximum likelihood estimators have attracted much attention [GKKT20], several works explore a Bayesian formulation for quantum tomography, allowing for a quantification of the estimate's uncertainty [BK10, GCC16, MA17, LLJL20, Mai22]. Bayesian estimation amounts to compute a posterior density on the set of density matrices (i.e., the complex spectrahedron):

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

This posterior accounts for prior information on the model and for measurements through the selection of a likelihood/loss function

$$L(\rho) := \sum_{m=1}^{M} (\hat{p}_m - \mathrm{tr}(A_m\rho))^2 \tag{2}$$

with A_m a Hermitian matrix characterizing the m^{th} experiment, and $\hat{p}_m$ a measured probability. Given a suitably selected prior density $\mu(\rho)$, Bayesian quantum tomography defines a posterior[1]

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

over the set of density matrices, the estimate of the density matrix being then

$$\hat{\rho}_\lambda = \int_{\rho\in\mathcal{CS}} \rho \hat{\mu}_\lambda(\rho) \mathrm{d}\rho. \tag{3}$$

In practice, computing the estimator (3) requires evaluating a high-dimensional integral, which is challenging. Existing approaches for quantum tomography rely on Markov Chain Monte-Carlo sampling (typically combined with Metropolis-Hastings and possibly enhanced by preconditioned Crank- Nicolson, see [Mai22]). We propose instead to benefit from the smoothness of the loss (2) by introducing a Langevin sampling algorithm for Bayesian quantum tomography. Exploiting the celebrated Burer-Monteiro factorization, we formulate the sampling problem in the space of rectangular complex matrices (i.e., low-rank factors of the density matrices), and derive a Langevin sampler in this space. For this, we propose a new low-rank promoting prior density, that is a complex generalization of the prior proposed in [Dal20]. We derive PAC-Bayesian bounds in the complete measurement setting for our proposed prior, and obtain identical rates (up to constant/logarithmic factor) as those obtained in the state-of-the-art works [MA17, Mai22]. We finally compare our method numerically to the MCMC sampler proposed in [MA17].

[1] As the likelihood does not necessarily come from a noise model on the measurements, the posterior is referred to as a pseudo-posterior in [MA17].

Related works: () The author of [BK10] proposed a Bayesian framework for quantum tomography; the estimated density matrix is then obtained as the mean of the posterior distribution, approximated using Metropolis-Hastings. This method was extended by [HH12], who proposed an adaptive Bayesian quantum tomography method, endowing Bayesian quantum tomography with a sequential importance sampling strategy allowing to determine adaptively the new measurements to use in the experiment. Further extensions were proposed in [GCC16], including accounting for time-dependency and proposing a new informative prior.

Mai and Alquier propose two estimators: the dens-estimator (where the likelihood captures the distance to the least-squares estimates of the density matrix), and the prob-estimator (where the likelihood captures the distance between the theoretical and empirical state probability vectors, which coincides with the approach used in this work). They compute these estimators using the Metropolis-Hastings algorithm, and derive PAC-Bayesian bounds. As Metropolis-Hastings is slow for high-dimensional systems, Mai subsequently proposed another efficient adaptive Metropolis-Hastings implementation for the prob-estimator, relying on Crank-Nicholson conditioning [Mai22]. A similar approach was used for the dens-estimator in [LLJL20].

2 Proposed prior density

Note that, for a posterior distribution $\hat{\mu}_\lambda(\rho) = \exp(-f_\lambda(\rho))$, (unconstrained) Langevin sampling generates a Markov chain

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

where w_k is a sequence of independent identically distributed Gaussian variables, and η_k a stepsize parameter. This approach does not exploit the fact that the sought density is typically low-rank, leading to a possible reduction of the sampling space dimension.

We propose instead to rely on the well-known Burer-Monteiro approach [BM03], and represent the density matrix $\rho \in \mathbb{C}^{d\times d}$, assumed to be of rank r, by a factor $Y \in \mathbb{C}^{d\times r}$ such that $\rho = YY^*$. By defining a posterior directly on the Y space, the dimension of the parameter space is reduced from $2d^2$ to $2dr$. Note also that, by construction, the density matrices generated by this representation are Hermitian and positive-semidefinite. The unit trace constraint amounts to assume that the factors belong to the complex hypersphere $\mathcal{CS}^{d\times r} = \{Y \in \mathbb{C}^{d\times r} : \|Y\|_{\mathrm{F}} = 1\}$.

We propose to use as prior a complex extension to the spectral scaled Student distribution [Dal20]:

$$\nu_\theta(Y) = C_\theta \det(\theta^2 I_d + YY^*)^{-(2d+r+2)/2}, \tag{4}$$

for $Y \in \mathbb{C}^{d\times r}$, with $C_\theta = (\int_{\mathbb{C}^{d\times r}} \det(\theta^2 I_d + YY^*)^{-(2d+r+2)/2})^{-1}$ a normalizing constant.

With this choice of prior, the posterior writes $\hat{\nu}_{\lambda,\theta}(Y) = e^{-f_{\lambda,\theta}(Y)}$, with

$$f_{\lambda,\theta}(Y) = \lambda \sum_{m=1}^{M} (\hat{p}_m - \mathrm{tr}(A_m YY^*))^2 + \frac{2d+r+2}{2} \log\det(\theta^2 I_d + YY^*) + \tilde{C}. \tag{5}$$

As we will see next, similarly to the real-valued case, this prior promotes low-rankness of the matrix $X = YY^*$. We compare in the next section the theoretical properties of the resulting Gibbs estimator to those obtained in [MA17]. The following result is an immediate extension of [Dal20, Lemma 1].

Lemma 1. *If Y is a random $d \times r$ complex matrix having as density the function ν_θ, then the column vectors $y_i \in \mathbb{C}^d$ of Y, for $i = 1, \ldots, r$, follow the d-variate complex scaled Student distribution $(\sqrt{2/3}\theta)t_{3,d}$. As a consequence, we have $\int_{\mathbb{C}^{d\times r}} \|y_i\|^2 \nu_\theta(Y) dY = 2\theta^2 d$ for all i.*

Proof. For any bounded and measurable function $h : \mathbb{C}^d \to \mathbb{R}$, we have

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

where $y_1 \in \mathbb{C}^d$ is the first column of Y, and where the second equality follows from applying the change of variable $Z := Y/\theta$. We then make a second change of variable and write x the first column of Z, and $Z = [x, Z_{2:r}] = [x, (I_d + xx^*)^{1/2}\bar{Z}_{2:r}]$. This yields

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

and

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

We then get

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

Writing the normalization constant as:

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

gives:

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

where we recognize in the last expression the complex multivariate t_3-distribution with location 0 and scale matrix $\Sigma = I_d$ [OTKP12], whose p.d.f μ is proportional to $\mu(y) \propto (1 + 2\|y\|^2/3)^{-(2d+3)/2} dy$.

As the covariance matrix of the complex multivariate t_p-distribution is $\frac{p}{p-2} I_d$, there follows that

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

so that

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

□

We next provide a PAC Bayesian bound on the Gibbs estimator obtained using this prior, and compare it with the approach of [MA17]. Let us first recall the experimental setting of [MA17]. We assume in this section that, for each qubit, one can measure one of the three Pauli observables σ_x, σ_y, σ_z. There are therefore $N_{\text{exp}} = 3^n$ possible experiments, with outcomes in $\{-1, 1\}^n$. Writing $\mathbf{R^a}$ the random variable associated to experiment $a \in \{1, \dots, N_{\text{exp}}\}$, we have

$$p^0_{a,s} := \text{prob}[\mathbf{R^a} = s] = \text{trace}(Y_0 Y_0^* \cdot P^a_s),$$

where P^a_s is a measurement operator associated with experiment $a \in \{1, \dots, N_{\text{exp}}\}$ and outcome $s \in \{1, \dots, 2^n\}$, and where $\rho_0 \in \mathbb{C}^{d\times d}$ is the true density matrix of the system, with factorization $\rho_0 = Y_0 Y_0^*$ for some $Y_0 \in \mathbb{C}^{d\times r}$. For an arbitrary density matrix $\rho = YY^*$, we write

$$p_{a,s}(\rho) := \text{prob}[\mathbf{R^a} = s] = \text{trace}(YY^* \cdot P^a_s).$$

We assume that we are in the complete measurement setting, meaning that each experiment $a \in \{1, \dots, N_{\text{exp}}\}$ is performed m times, leading to $N_{\text{tot}} = mN_{\text{exp}}$ measurements in total. We write

$$\hat{p}_{a,s} := \frac{1}{m} \sum_{i=1}^{m} 1_{\mathbf{R^a_i}=s}$$

the empirical frequencies, with $\mathbf{R^a_i}$ the random variable associated to the ith replication of random variable $\mathbf{R^a}$. We finally define the matrices P^0, $P(\rho)$ and $\hat{P} \in [0,1]^{N_{\text{exp}}\times 2^n}$, whose (a, s) entries are equal to $p^0_{a,s}$, $p_{a,s}(\rho)$ and $\hat{p}_{a,s}$, respectively.

Following [MA17], we also write the posterior distribution $\hat{\nu}_{\theta,\lambda}$ over $\mathbb{C}^{d\times r}$ defined as

$$\hat{\nu}_{\theta,\lambda}(Y) \propto \exp\left[-\lambda \|P(YY^*) - \hat{P}\|^2_{\text{F}}\right] \nu_\theta(Y), \tag{6}$$

for a prior $\nu_\theta(Y)$, and our density matrix estimator is given by

$$\hat{\rho}_{\theta,\lambda} = \hat{Y}_{\theta,\lambda} \hat{Y}^*_{\theta,\lambda}, \qquad \hat{Y}_{\theta,\lambda} := \int_{\mathbb{C}^{d\times r}} Y \hat{\nu}_{\theta,\lambda}(Y) dY.$$

In this setting, the authors of [MA17] derived a PAC-Bayesian bound for their estimator. We extend here their analysis to our alternative prior, showing that this new prior preserves the rate achieved in [MA17]. We first recall the following lemma.

Lemma 2. *[MA17, Lemma 3] For $\lambda > 0$, we have for all $Y \in \mathbb{C}^{d\times r}$*

$$\mathbb{E} \exp\left\{\lambda(\|P(YY^*) - \hat{P}\|^2_{\text{F}} - \|P^0 - \hat{P}\|^2_{\text{F}}) - \lambda\left(1 + \frac{\lambda}{m}\right)\|P^0 - P(YY^*)\|^2_{\text{F}}\right\} \leq 1 \tag{7}$$

$$\mathbb{E} \exp\left\{\lambda\left(1 - \frac{\lambda}{m}\right)\|P^0 - P(YY^*)\|^2_{\text{F}} - \lambda(\|P(YY^*) - \hat{P}\|^2_{\text{F}} - \|P^0 - \hat{P}\|^2_{\text{F}})\right\} \leq 1, \tag{8}$$

where the expectations are over the random variables $\mathbf{R^a_i}$ (i.e., over the results of the m replications of the experiments).

The following lemma is a straightforward adaptation of [MA17, Lemma 4].

Lemma 3. *For $\lambda > 0$ such that $\frac{\lambda}{m} < 1$, with probability $1 - \epsilon$, $\epsilon \in]0, 1[$, for any distribution ν there holds*

$$\int \|P^0 - P(YY^*)\|^2_{\text{F}} \hat{\nu}_{\theta,\lambda}(Y) dY \leq \inf_\nu \frac{\left(1 + \frac{\lambda}{m}\right) \int \|P^0 - P(YY^*)\|^2_{\text{F}} \nu(Y) dY + 2\frac{\log\left(\frac{2}{\epsilon}\right) + D_{\text{KL}}(\nu, \nu_\theta)}{\lambda}}{1 - \frac{\lambda}{m}}. \tag{9}$$

Proof. Due to Lemma 2, there holds

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

By using Fubini's theorem, there holds:

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

Now, using [Cat07, Lemma 1.1.3], for any distribution ν, we have

$$\begin{aligned}\mathbb{E}\exp\sup_\nu\Bigg\{&\lambda\left(\int \|P(YY^*) - \hat{P}\|_{\mathrm{F}}^2\nu(Y)dY - \|P^0 - \hat{P}\|_{\mathrm{F}}^2\right) - \log\left(\frac{2}{\epsilon}\right) - D_{\mathrm{KL}}(\nu,\nu_\theta)\\ &-\lambda\left(1+\frac{\lambda}{m}\right)\int\|P^0 - P(YY^*)\|_{\mathrm{F}}^2\nu(Y)dY\Bigg\} \leq \frac{\epsilon}{2}\\ \mathbb{E}\exp\sup_\nu\Bigg\{&\lambda\left(1-\frac{\lambda}{m}\right)\int\|P^0 - P(YY^*)\|_{\mathrm{F}}^2\nu(Y)dY - \log\left(\frac{2}{\epsilon}\right) - D_{\mathrm{KL}}(\nu,\nu_\theta)\\ &-\lambda\left(\int \|P(YY^*) - \hat{P}\|_{\mathrm{F}}^2\nu(Y)dY - \|P^0 - \hat{P}\|_{\mathrm{F}}^2\right)\Bigg\} \leq \frac{\epsilon}{2}.\end{aligned}$$

Now, using $\mathbb{1}_{x\geq 0}(x) \leq \exp(x)$, one has

$$\begin{aligned}\mathbb{P}\Bigg\{\sup_\nu\Bigg[&\lambda\left(\int \|P(YY^*) - \hat{P}\|_{\mathrm{F}}^2\nu(Y)dY - \|P^0 - \hat{P}\|_{\mathrm{F}}^2\right) - \log\left(\frac{2}{\epsilon}\right) - D_{\mathrm{KL}}(\nu,\nu_\theta)\\ &-\lambda\left(1+\frac{\lambda}{m}\right)\int\|P^0 - P(YY^*)\|_{\mathrm{F}}^2\nu(Y)dY\Bigg] \geq 0\Bigg\} \leq \frac{\epsilon}{2}\\ \mathbb{P}\Bigg\{\sup_\nu\Bigg[&\lambda\left(1-\frac{\lambda}{m}\right)\int\|P^0 - P(YY^*)\|_{\mathrm{F}}^2\nu(Y)dY - \log\left(\frac{2}{\epsilon}\right) - D_{\mathrm{KL}}(\nu,\nu_\theta)\\ &-\lambda\left(\int \|P(YY^*) - \hat{P}\|_{\mathrm{F}}^2\nu(Y)dY - \|P^0 - \hat{P}\|_{\mathrm{F}}^2\right)\Bigg] \geq 0\Bigg\} \leq \frac{\epsilon}{2}.\end{aligned}$$

We deduce that, with probability at least $1-\epsilon/2$ over the data, for any density ν,

$$\begin{aligned}\int \|P(YY^*) - \hat{P}\|_{\mathrm{F}}^2\nu(Y)dY &\leq \|P^0 - \hat{P}\|_{\mathrm{F}}^2 + \frac{\log\left(\frac{2}{\epsilon}\right) + D_{\mathrm{KL}}(\nu,\nu_\theta)}{\lambda}\\ &\quad+\left(1+\frac{\lambda}{m}\right)\int\|P^0 - P(YY^*)\|_{\mathrm{F}}^2\nu(Y)dY\\ \int\|P^0 - P(YY^*)\|_{\mathrm{F}}^2\nu(Y)dY &\leq \frac{\int \|P(YY^*) - \hat{P}\|_{\mathrm{F}}^2\nu(Y)dY - \|P^0 - \hat{P}\|_{\mathrm{F}}^2 + \frac{\log\left(\frac{2}{\epsilon}\right) + D_{\mathrm{KL}}(\nu,\nu_\theta)}{\lambda}}{1-\frac{\lambda}{m}}.\end{aligned}$$

The next result is obtained by using an union argument (i.e., replacing the first term in the right-hand side of the second inequality by its upper bound given by the first inequality): for any density ν, with probability at least $1-\epsilon$ over the data,

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

□

According to the Donsker-Varadhan variational formula, we get:

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

We next need to upper bound the right-hand side of (9). For this, borrowing on [Dal20], we restrict the infimum to a family of priors obtained as translations of our prior $\nu_\theta(Y)$, that we write $\bar{\nu}_\theta(Y) = \nu_\theta(Y-\bar{Y})$ for some $\bar{Y}$. We derive the following result, which is a direct extension of [Dal20, Lemma 2] to the complex setting.

Lemma 4. *Let $\bar{\nu}_\theta$ be the probability density function obtained from the prior ν_θ by a translation, $\bar{\nu}_\theta(Y) = \nu_\theta(Y-\bar{Y})$. Then, for any matrix $\bar{Y} \in \mathbb{C}^{d\times r}$ of rank at most p, we have*

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

Proof. By definition of the KL-divergence, there holds

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

For a Hermitian positive definite matrix M, we write $\log(M)$ and $M^{1/2}$ the principal matrix logarithm and principal matrix square root, respectively. Let $M = U\Lambda U^*$ be an eigenvalue decomposition, with $U \in \mathbb{C}^{d\times d}$ unitary and $\Lambda \in \mathbb{R}^{d\times d}$ diagonal with strictly positive entries. Then, $\log(M) = U\log(\Lambda)U^*$ and $M^{1/2} = U\Lambda^{1/2}U^*$, where the log and square root are applied to the diagonal entries of Λ.

We define $A := (\theta^2 I_d + YY^*)^{-1/2}$, and $B := \theta^2 I_d + (Y-\bar{Y})(Y-\bar{Y})^*$. Note that A and B are Hermitian and positive definite. There follows that

$$2\log\left(\frac{\nu_\theta(Y)}{\nu_\theta(Y-\bar{Y})}\right) = 2\log\left(\frac{\det(\theta^2 I_d + YY^*)^{-(2d+r+2)/2}}{\det(\theta^2 I_d + (Y-\bar{Y})(Y-\bar{Y})^*)^{-(2d+r+2)/2}}\right) \tag{10}$$

$$= (2d+r+2)\log\left(\frac{\det(B)}{\det(A^{-2})}\right) \tag{11}$$

$$= (2d+r+2)\log\left(\det(ABA)\right) \tag{12}$$

$$= (2d+r+2)\sum_{i=1}^{d}\log\lambda_i, \tag{13}$$

where we noted λ_i the ith largest eigenvalue of ABA, with associated eigenvector $u_i \in \mathbb{C}^d$. We next write

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

and show that the rank of $ABA-I_d$ is at most $2p$. Indeed, note first that $\mathrm{range}(A\bar{Y}(\bar{Y}-Y)^*A) \subseteq A\ \mathrm{range}(\bar{Y})$, which implies that $\mathrm{rank}(A\bar{Y}(\bar{Y}-Y)^*A)$ has dimension at most p. Then, notice that $\mathrm{rank}(AY\bar{Y}^*A) = \mathrm{rank}(A\bar{Y}Y^*A) \leq p$ using the same argument. There follows that the rank of $ABA - I_p$ is at most $2p$, the matrix ABA has at most $2p$ eigenvalues different from one, and consequently the summation in (13) involves at most $2p$ nonzero terms.

There holds:

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

Note that $\|(\bar{Y}-Y)^*Au_i\|^2 = u_i^*A(\bar{Y}-Y)(\bar{Y}-Y)^*Au_i = u_i^*A\left(\bar{Y}\bar{Y}^* - \bar{Y}Y^* - Y\bar{Y}^* + YY^*\right)Au_i = u_i^*\left(A\bar{Y}(\bar{Y}-Y)^*A - AY\bar{Y}^*A\right)u_i + \|Y^*Au_i\|^2$. There follows:

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

where we used in the last inequality the fact that $\|Y^*Au_i\|^2 \leq 1$ by definition of A. Indeed, note that

$$1 = u_i^*u_i = u_i^*(\theta^2 I_d + YY^*)^{-1/2}(\theta^2 I_d + YY^*)(\theta^2 I_d + YY^*)^{-1/2}u_i = \theta^2\|Au_i\|^2 + \|Y^*Au_i\|^2 \geq \|Y^*Au_i\|^2.$$

Using the concavity of the function $\log(1+x^{1/2})$ over $(0,\infty)$, we get

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

Using the fact that u_i's are orthonormal and that $A \preceq \theta^{-1}I_d$, there holds:

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

$\square$

Lemma 5. *For any $\bar{Y} \in \mathbb{C}^{d\times r}$ of rank at most p, let $\bar{\nu}_\theta$ be the shifted prior $\bar{\nu}_\theta(Y) = \nu_\theta(Y-\bar{Y})$. Then, there holds:*

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

with $a := \|Y_0 - \bar{Y}\|_{\mathrm{F}}\|Y_0 + \bar{Y}\|_{\mathrm{F}}$ and $b := \|Y_0 - \bar{Y}\|_{\mathrm{F}} + \|Y_0 + \bar{Y}\|_{\mathrm{F}}$.

Proof. Recall that the entries of the matrices $P(YY^*)$ and P^0 are probabilities, and therefore lie in the interval $[0,1]$. As a result, the matrix $A := P(YY^*) - P^0$ satisfies $A_{i,j} \in [-1,1]$, hence $\|A\|_{\mathrm{F}}^2 = \sum_{i,j=1}^d A_{i,j}^2 \leq d^2$, which implies that $\|A\|_{\mathrm{F}} \leq d$. There follows that

$$\|P(YY^*) - P^0\|_{\mathrm{F}}^2 \leq d\|P(YY^*) - P^0\|_{\mathrm{F}}. \tag{14}$$

Note then that, according to [ABH+13, Proof of Lemma 5], there holds

$$\|P(YY^* - Y_0Y_0^*)\|_{\mathrm{F}} \leq 6^{n/2}\|YY^* - Y_0Y_0^*\|_{\mathrm{F}}. \tag{15}$$

Then, write

$$YY^* - Y_0Y_0^* = \frac{1}{2}\left((Y - Y_0)(Y + Y_0)^* + (Y + Y_0)(Y - Y_0)^*\right),$$

and note that, by submultiplicativity of the Frobenius norm [HJ13, p.342], there holds:

$$\|YY^* - Y_0Y_0^*\|_{\mathrm{F}} \leq \|Y - Y_0\|_{\mathrm{F}}\|Y + Y_0\|_{\mathrm{F}} \tag{16}$$

By the triangle inequality,

$$\begin{aligned} \|Y_0 - Y\|_{\mathrm{F}} &\leq \|Y_0 - \bar{Y}\|_{\mathrm{F}} + \|\bar{Y} - Y\|_{\mathrm{F}} \\ \|Y_0 + Y\|_{\mathrm{F}} &\leq \|Y_0 + \bar{Y}\|_{\mathrm{F}} + \|Y - \bar{Y}\|_{\mathrm{F}}. \end{aligned} \tag{17}$$

Combining (14), (15), (16) and (17) gives:

$$\begin{aligned} \|P(YY^*) - P^0\|_{\mathrm{F}}^2 &\leq 6^{n/2}d\left(\|Y_0 - \bar{Y}\|_{\mathrm{F}} + \|\bar{Y} - Y\|_{\mathrm{F}}\right)\left(\|Y_0 + \bar{Y}\|_{\mathrm{F}} + \|Y - \bar{Y}\|_{\mathrm{F}}\right) \\ &\leq 6^{n/2}d(\|Y_0 - \bar{Y}\|_{\mathrm{F}}\|Y_0 + \bar{Y}\|_{\mathrm{F}} + \|Y - \bar{Y}\|_{\mathrm{F}}(\|Y_0 - \bar{Y}\|_{\mathrm{F}} + \|Y_0 + \bar{Y}\|_{\mathrm{F}}) + \|Y - \bar{Y}\|_{\mathrm{F}}^2), \end{aligned}$$

so that

$$\begin{aligned} \int \|P(YY^*) - P^0\|_{\mathrm{F}}^2 \nu_\theta(Y - \bar{Y})dY \leq & 6^{n/2}d(\|Y_0 - \bar{Y}\|_{\mathrm{F}}\|Y_0 + \bar{Y}\|_{\mathrm{F}} \\ & + (\|Y_0 - \bar{Y}\|_{\mathrm{F}} + \|Y_0 + \bar{Y}\|_{\mathrm{F}})\left(\int \|Y - \bar{Y}\|_{\mathrm{F}}\nu_\theta(Y - \bar{Y})\mathrm{d}Y\right) \\ & + \int \|Y - \bar{Y}\|_{\mathrm{F}}^2\nu_\theta(Y - \bar{Y})\mathrm{d}Y). \end{aligned}$$

We compute

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

where the inequality is Jensen's inequality (stating that for any convex function ϕ, there holds $\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$, and the convexity of the function $x \mapsto x^2$, and where the last equality comes from Lemma 1. Similarly,

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

The claim follows.

□

Lemma 6. *For any $\bar{Y}$ of rank at most p, let $\bar{\nu}_\theta$ be the shifted prior $\bar{\nu}_\theta(Y) = \nu_\theta(Y - \bar{Y})$. Then, there holds:*

$$\int \|Y_0Y_0^* - YY^*\|_{\mathrm{F}}^2\hat{\nu}_{\theta,\lambda}(Y)dY \leq \inf_{\bar{Y}} \frac{3^n\left(1 + \frac{\lambda}{m}\right)(a + \sqrt{2}\theta drb + 2\theta^2dr\sqrt{dr}) + 2\frac{\log\left(\frac{2}{\epsilon}\right) + 2p(2d+r+2)\log\left(1 + \frac{\|\bar{Y}\|_{\mathrm{F}}^2}{\sqrt{2p}\theta}\right)}{2^n\lambda}}{1 - \frac{\lambda}{m}}.$$

Proof. Note first that, according to Jensen's inequality, there holds:

$$\|\hat{\rho}_{\theta,\lambda} - \rho_0\|_{\mathrm{F}}^2 \leq \int \|Y_0Y_0^* - YY^*\|_{\mathrm{F}}^2 \hat{\nu}_{\theta,\lambda}(Y)dY \leq \frac{1}{2^n} \int \|P(YY^*) - P^0\|_{\mathrm{F}}^2 \hat{\nu}_{\theta,\lambda}(Y)dY,$$

where the second inequality comes from [MA17]. Now, using Lemma 3 and Lemma 5 yields

$$\begin{aligned}\|\hat{\rho}_{\theta,\lambda} - \rho_0\|_{\mathrm{F}}^2 &\leq \inf_{\bar{Y}} \frac{6^{n/2}d\left(1+\frac{\lambda}{m}\right)(a + \sqrt{2dr}\theta b + 2dr\theta^2) + 2\frac{\log\left(\frac{2}{\epsilon}\right)+2p(2d+r+2)\log\left(1+\frac{\|\bar{Y}\|_{\mathrm{F}}}{\sqrt{2p}\theta}\right)}{\lambda}}{(1-\frac{\lambda}{m})2^n}. \\ &\leq \inf_{\bar{Y}} N_{\mathrm{tot}} \frac{6^{n/2}d\left(1+\frac{\lambda}{m}\right)(a + \sqrt{2dr}\theta b + 2dr\theta^2) + 2\frac{\log\left(\frac{2}{\epsilon}\right)+2p(2d+r+2)\log\left(1+\frac{\|\bar{Y}\|_{\mathrm{F}}}{\sqrt{2p}\theta}\right)}{\lambda}}{(1-\frac{\lambda}{m})N_{\mathrm{tot}}2^n}\end{aligned}$$

Now, let us recall that $d = 2^n$, $N_{\mathrm{tot}} = N_{\mathrm{exp}}m$, $N_{\mathrm{exp}} = 3^n$. Choosing $\theta \propto 6^{-3n/4}$ gives:

$$\begin{aligned}\|\hat{\rho}_{\theta,\lambda} - \rho_0\|_{\mathrm{F}}^2 \leq \inf_{\bar{Y}} &\frac{m\left(1+\frac{\lambda}{m}\right)}{N_{\mathrm{tot}}\left(1-\frac{\lambda}{m}\right)}\left(a3^{3n/2}2^{n/2} + \sqrt{2r}b3^{3n/4}2^{-n/2} + 2r\right) \\ &+ \frac{2m3^n}{2^n\lambda(1-\frac{\lambda}{m})N_{\mathrm{tot}}}\left(\log\left(\frac{2}{\epsilon}\right) + 2p(2d+r+2)\log\left(1+\frac{\|\bar{Y}\|_{\mathrm{F}}}{\sqrt{2p}\theta}\right)\right).\end{aligned}$$

Letting $\bar{Y} = \bar{Y}_{\mathrm{chos}}$, a chosen value satisfying $a = 3^{-3n/2}2^{-n/2}$ and $b = 3^{-3n/4}2^{n/2}$ leads to a first term that is independent on n apart from the dataset size N_{tot}: the larger the amount of available data, the lower this contribution to this error bound. Letting $C := \frac{m\left(1+\frac{\lambda}{m}\right)}{\left(1-\frac{\lambda}{m}\right)}\left(1+\sqrt{2r}+2r\right)$, we get:

$$\|\hat{\rho}_{\theta,\lambda} - \rho_0\|_{\mathrm{F}}^2 \leq \frac{C}{N_{\mathrm{tot}}} + \frac{2m3^n}{2^n\lambda(1-\frac{\lambda}{m})N_{\mathrm{tot}}}\left(\log\left(\frac{2}{\epsilon}\right) + 2p(2d+r+2)\log\left(1+\frac{\|\bar{Y}_{\mathrm{chos}}\|_{\mathrm{F}}}{\sqrt{2p}\theta}\right)\right).$$

The right-hand side is minimized for $\lambda = m/2$, leading to

$$\|\hat{\rho}_{\theta,\lambda} - \rho_0\|_{\mathrm{F}}^2 \leq \frac{C + 8(3/2)^n\log\left(\frac{2}{\epsilon}\right) + 16p3^n(2+2^{-n}r+2^{1-n})\log\left(1+\frac{\|\bar{Y}_{\mathrm{chos}}\|_{\mathrm{F}}}{\sqrt{2p}\theta}\right)}{N_{\mathrm{tot}}}.$$

Taking $p = \mathrm{rank}(\rho_0)$, we retrieve the rate $3^n\mathrm{rank}(\rho_0)/N_{\mathrm{tot}}$ obtained in [MA17], up to logarithmic factors. □

3 A Langevin sampler for quantum tomography

With this choice of prior, the posterior writes $\mu_{\lambda,\theta}(Y) = e^{-f_{\lambda,\theta}(Y)}$, with

$$f_{\lambda,\theta}(Y) = \lambda\sum_{m=1}^{M}(\hat{p}_m - \mathrm{tr}(A_mYY^*))^2 + \frac{2d+r+2}{2}\log\det(\theta^2 I_d + YY^*) + \tilde{C}. \tag{18}$$

To avoid the need to manipulate complex numbers in our implementations, we operate a change of variables, relying on the well-known vector space isomorphism between $\mathbb{C}^{d\times r}$ and a subset of $\mathbb{R}^{2d\times 2r}$ (see, e.g., [OTKP12]):

$$\psi : M^R + iM^I \mapsto \frac{1}{\sqrt{2}}\begin{pmatrix} M^R & -M^I \\ M^I & M^R \end{pmatrix}. \tag{19}$$

The next lemma allows us to rewrite the posterior in terms of real-valued matrices.

Lemma 7. *For all $M, N \in \mathbb{C}^{d\times d}$ Hermitian positive semidefinite, there holds* $\mathrm{tr}(MN) = \mathrm{tr}(\psi(M)\psi(N))$ *and* $\det(M) = \sqrt{2^d \det(\psi(M))}$.

Proof. Note first that, since M is Hermitian positive definite, then, since $M^* = M^{R\top} - iM^{I\top} = M$, there holds $M^R = M^{R\top}$ and $M_I = -M^{I\top}$. In other words, M^R (and N^R) are symmetric, while M^I (and N^I) are skew-symmetric. We compute

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

On the other hand,

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

where the second equality comes from the fact that the inner product between a symmetric and a skew-symmetric matrix is zero.

For the second claim, we prove that, for any eigenvalue λ of M with multiplicity m_λ, there exists an eigenvalue $\lambda/\sqrt{2}$ of $\psi(M)$ with multiplicity $2m_\lambda$ (note that algebraic and geometric multiplicities coincide for Hermitian/symmetric matrices). Since M is Hermitian, it admits an orthogonal basis of eigenvectors. Let $v \in \mathbb{C}^d$ be an eigenvector of M, with eigenvalue λ, i.e., $Mv = (M^R v^R - M^I v^I) + i(M^R v^I + M^I v^R) = \lambda(v^R + iv^I)$. Then,

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

and

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

Thus, the two vectors (v^R, v^I) and $(v^I, -v^R)$ are eigenvectors of $\psi(M)$ with eigenvalue $\lambda/\sqrt{2}$. Note that these vectors are distinct, as otherwise $v^I = v^R = -v^I = 0$ so that $v = v^R + iv^I = 0$ which conflicts with the definition of v. The result then simply follows from the definition of the determinant as the product of the eigenvalues (taking into account multiplicities).

$\square$

Note that there holds, for any $Y \in \mathbb{C}^{d\times r}$, $\psi(YY^*) = \sqrt{2}\psi(Y)\psi(Y)^\top$. In terms of $\tilde{Y} := \psi(Y)$, the posterior becomes $\tilde{\mu}_{\lambda,\theta}(\tilde{Y}) = \exp(-\tilde{f}_{\lambda,\theta}(\tilde{Y}))$, with:

$$\tilde{f}_{\lambda,\theta}(\tilde{Y}) = \lambda \sum_{m=1}^{M} (\hat{p}_m - \sqrt{2}\mathrm{tr}(\tilde{A}_m \tilde{Y}\tilde{Y}^\top))^2 + \frac{2d+r+2}{4} \log\det\left(\frac{\theta^2}{\sqrt{2}} I_{2d} + \sqrt{2}\tilde{Y}\tilde{Y}^\top\right) + \hat{C}, \quad (20)$$

with $\hat{C} = \tilde{C} + (2d + r + 2)d\log(2)/4$. The gradient of this function is obtained as:

$$\nabla\tilde{f}_{\lambda,\theta}(\tilde{Y}) = -2\sqrt{2}\lambda \sum_{m=1}^{M} (\hat{p}_m - \sqrt{2}\mathrm{tr}(\tilde{A}_m \tilde{Y}\tilde{Y}^\top))(\tilde{A}_m + \tilde{A}_m^\top)\tilde{Y} + \frac{2d+r+2}{\theta^2}\left(I_{2d} + \frac{2}{\theta^2}\tilde{Y}\tilde{Y}^\top\right)^{-1}\tilde{Y}.$$

According to the Sherman-Morrison-Woodbury formula [Hig08], the matrix inverse can be written as

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

which is substantially cheaper to evaluate in the case $r \ll d$ (remember that $r = 1$ for the recovery of a pure state).

Algorithm 1 A Langevin sampling algorithm for quantum tomography

1: Initialize $Y_0 \in \mathbb{C}^{d\times r}$, η_k a sequence of stepsize, and compute $\tilde{Y}_0 := \psi(Y_0)$
2: **for** $k \geq 1$ until the termination criterion is satisfied **do**
3: Sample the entries of $w_k^R, w_k^I \in \mathbb{R}^{d\times r}$ i.i.d. at random according to the standard Gaussian distribution, and let $w_k := w_k^R + iw_k^I$, and $\tilde{w}_k := \psi(w_k)$.
4: $\tilde{Y}_k = \tilde{Y}_{k-1} - \eta_k \nabla f(\tilde{Y}_{k-1}) + \frac{\sqrt{2\eta_k}}{\beta} w_k$
5: **end for**
6: Return $\tilde{Y} = \frac{1}{N_{\text{est}}} \sum_{k=1}^{N_{\text{est}}} \tilde{Y}_{\text{end}-k+1}$

4 Numerical results

The performance obtained by our Langevin sampler, for various values of m (the number of repetitions of measurements), is illustrated on Figure 1. This experiment was obtained with $n = 3$, $d = 2^n = 8$, $r = 5$, $\lambda = m/2$, $\eta = 0.1/m$, $\beta = 100$, $\theta = 1$. Each point on Figure 1 was obtained as an average of the result of Algorithm 1 for 5 random seeds (impacting here the initialization and Langevin noise). Algorithm 1 was run for 5000 iterations, and N_{est} was set to 100 (in other words, the estimate was computed as the average of the 100 last points of the Markov chain generated by Algorithm 1.

The true density matrix for this experiment is a rank-2 density matrix generated according to [MA17]: $\rho_0 = \frac{1}{2}\psi_1\psi_1^* + \frac{1}{2}\psi_2\psi_2^*$, where ψ_1 and ψ_2 are two normalized orthogonal vectors in $\mathbb{C}^{d\times 1}$.

We see on Figure 1 that the error decreases with the sampling size, with a rate $1/m$, in accordance with the theoretical analysis of the last section.

This experiment was run in a complete measurement setting, but unlike [MA17], we only assume one measurement value per experiment (so we do not measure each qubit separately).

References

[ABH+13] Pierre Alquier, Cristina Butucea, Mohamed Hebiri, Katia Meziani, and Tomoyuki Morimae. Rank-penalized estimation of a quantum system. *Physical Review A*, 88(032113), 2013.

[BK10] R. Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12(043034), 2010.

[BM03] S. Burer and R.D.C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via lowrank factorization. *Mathematical Programming*, 95(2):329–357, 2003.

[Cat07] Olivier Catoni. *Pac-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning.* Number 56. IMS Lecture Notes Monogr, 2007.

[Dal20] Arnak S. Dalalyan. Exponential weights in multivariate regression and a low-rankness favoring prior. *Ann. Inst. H. Poincaré Probab. Statist.*, 56(2):1465–1483, 2020.

[GCC16] C. Granade, J. Combes, and D. G. Cory. Practical bayesian tomography. *New Journal of Physics*, 18(133024):259–296, 2016.

[GKKT20] M. Guţă, J. Kahn, R. Kueng, and J. A. Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, 2020.

[HH12] F. Husz'ar and N. M. T. Houlsby. Adaptive bayesian quantum tomography. *Physical Review A*, 85(052120), 2012.

[Hig08] Nicholas J. Higham. *Functions of Matrices.* SIAM, 2008.

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix Analysis (Second Edition).* Cambridge University Press, 2013.

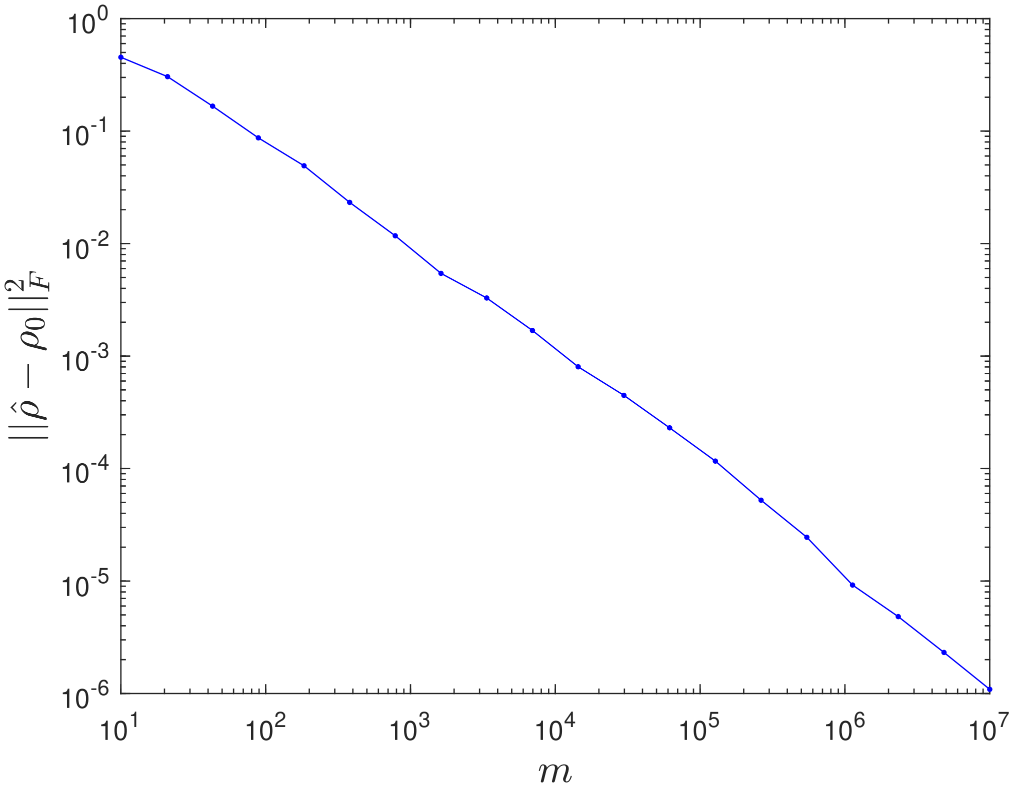

Figure 1: Relative error achieved by Algorithm 1 when increasing m, the number of repetition of each experiment.

[LLJL20] Joseph M. Lukens, Kody J. H. Law, Ajay Jasra, and Pavel Lougovski. A practical and efficient approach for bayesian quantum state estimation. *New Journal of Physics*, 22(6):063038, 2020.

[MA17] The Tien Mai and Pierre Alquier. Pseudo-bayesian quantum tomography with rank-adaptation. *Journal of Statistical Planning and Inference*, 184:62–76, 2017.

[Mai22] The Tien Mai. An efficient adaptive MCMC algorithm for Pseudo-Bayesian quantum tomography, 2022.

[OTKP12] Esa Ollila, David E. Tyler, Visa Koivunen, and H. Vincent Poor. Complex elliptically symmetric distributions: Survey, new results and applications. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 60(11):5597–5625, 2012.