Master's thesis: Numerical comparison of MCMC methods for Quantum Tomography

Danila Mokeev
Supervisors: Estelle Massart and Andrew Thompson
21st of June 2024

Ecole Polytechnique de Louvain

Plan of this thesis

- This thesis main problem is Quantum Tomography
- We talk about algorithms that solve that problem, in particular MCMC
- In our experiments, we numerically compare 2 methods in different experimental setups
- We also introduce the 2 new algorithms to understand why one algorithm might work better than the other

Table of Contents

Table of Contents

Problem: quantum state reconstruction

Goal: Reconstitute a quantum state

Unfortunately, there are some challenges:

- Quantum systems are inherently probabilistic
- A measurement can ony be made once
- We can only measure the position or momentum, but not both

Quantum Tomography

Quantum tomography provides a solution to this problem.

Key steps:

1. Replicate the initial state of the system multiple times
2. Measure each clone once
3. Calculate the empirical probabilities
4. Estimate the quantum state with any appropriate method

Quantum Tomography: mathematical description (1)

The Born rule states that

$$p(m) = \mathsf{tr}(\rho P_m) \tag{1}$$

with

- P_m the projector matrix associated to the eigenvalue m of an *observable* O
- $p(m)$ the probability of occurence of m
- ρ the *density matrix* representing the quantum state
  - positive semi-definite
  - Hermitian ($\rho = \rho^\dagger$)
  - trace$(\rho) = 1$

Quantum Tomography: mathematical description (2)

If we flatten the matrices

$$A = \begin{bmatrix} \vec{P}_1 \\ \vec{P}_2 \\ \vec{P}_3 \\ \vdots \end{bmatrix} \quad \vec{\rho} = \begin{bmatrix} \rho_{11} \\ \rho_{12} \\ \rho_{13} \\ \vdots \end{bmatrix} \tag{2}$$

then we can estimate ρ by solving the resulting system of equations

$$A\vec{\rho} = \hat{p} \tag{3}$$

Existing methods

- Direct methods:

$$\hat{\rho} = (A^T A)^{-1} A^T \hat{p} \tag{4}$$

- Optimization-based methods:

$$\hat{\rho} = \text{argmin}_{\vec{\rho}} ||A\vec{\rho} - \hat{p}|| \tag{5}$$

- Pauli basis expansion:

$$\hat{\rho} = \sum_{b \in \{I,x,y,z\}^n} \rho_b \sigma_b \tag{6}$$

- Bayesian methods, and in particular MCMC methods

$$\hat{\rho} = \frac{1}{N} \sum_{i=1}^{N} \rho_i \quad \text{with } \rho_i \sim \pi(\rho|\mathbf{D}) \tag{7}$$

Existing methods: our focus in this thesis

- Direct methods:

$$\hat{\rho} = (A^T A)^{-1} A^T \hat{p} \tag{8}$$

- Optimization-based methods:

$$\hat{\rho} = \text{argmin}_{\vec{\rho}} ||A\vec{\rho} - \hat{p}|| \tag{9}$$

- Pauli basis expansion:

$$\hat{\rho} = \sum_{b \in \{I,x,y,z\}^n} \rho_b \sigma_b \tag{10}$$

- Bayesian methods, and in particular MCMC methods

$$\hat{\rho} = \frac{1}{N} \sum_{i=1}^{N} \rho_i \quad \text{with } \rho_i \sim \pi(\rho|\mathbf{D}) \tag{11}$$

Table of Contents

Bayesian inference

Context: We are working in the Bayesian framework:

$$\pi(\rho|\mathbf{D}) \propto \mathcal{L}(\mathbf{D}|\rho)\pi(\rho) \tag{12}$$

In the context of Quantum Tomography:

- Likelihood $\mathcal{L}(\mathbf{D}|\rho) = ||A\vec{\rho} - \hat{p}||$
- Prior $\pi(\rho)$ is method specific
- Posterior $\pi(\rho|\mathbf{D})$ corresponds to a distribution over density matrices ρ

Markov chain Monte Carlo methods

- Markov chain Monte Carlo (MCMC) methods *sample* from $\pi(\rho|\mathbf{D})$.
- They build a Markov chain of samples $\rho_1, \rho_2, \ldots$ such that

$$f(x) = \pi(\rho|\mathbf{D}) \tag{13}$$

with the equilibrium distribution $f(x)$ of the chain

- The density matrix is then approximated as

$$\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N} \rho_i \quad \text{with } \rho_i \sim \pi(\rho|\mathbf{D}) \tag{14}$$

The Metropolis-Hastings algorithm

Algorithm 1: Metropolis-Hastings algorithm

for $t \leftarrow 1 : T$ **do**

1. Generate a candidate $\rho^* \sim \overbrace{q(\rho|\rho^{(t-1)})}^{\text{proposal}}$
2. Set $\rho^{(t)} = \begin{cases} \rho^* & \text{with prob. } \alpha(\rho^*, \rho^{(t-1)}) \\ \rho^{(t-1)} & \text{with prob. } 1 - \alpha(\rho^*, \rho^{(t-1)}) \end{cases}$

with

$$\underbrace{\alpha(\rho^*, \rho^{(t-1)})}_{\text{acceptance ratio}} = \frac{\pi(\rho^*|\mathbf{D})q(\rho^{(t-1)}|\rho^*)}{\pi(\rho^{(t-1)}|\mathbf{D})q(\rho^*|\rho^{(t-1)})} \tag{15}$$

end

Illustration of the Metropolis-Hastings algorithm

mcmc.gif

Advantages of MCMC algorithms

Why are we interested in MCMC methods?

- Prior $\pi(\rho)$: additional information about the density matrix - low-rank for example
- Uncertainty quantification: working with distributions instead of point estimates

Prob-estimator (1)

Introduced in [**MA17**], it combines Metropolis-within-Gibbs sampling with a low-rank prior.

- Analogous to eigenvector factorization: $\rho = \sum_{i=1}^{d} \gamma_i V_i V_i^\dagger$
- $\pi_1(\gamma_1 \ldots \gamma_d)$ is a Dirichlet distribution with a small, constant parameter, leading to sparse values
- $\pi_2(V_1 \ldots V_d)$ is a unit sphere distribution

Prob-estimator (2)

Algorithm: combination between Metropolis-Hastings and Gibbs sampling

Algorithm 2: Prob-estimator algorithm

for $t \leftarrow 1 : T$ **do**

  for $i \leftarrow 1 : d$ **do**

    1. Sample γ_i^* from $\pi_1(\gamma_1, \ldots, \gamma_d)$
    2. Update $\gamma^{(t)}$ with accept/reject step

  end

  for $i \leftarrow 1 : d$ **do**

    1. Sample V_i^* from $\pi_2(V_1, \ldots, V_d)$
    2. Update $V^{(t)}$ with an accept/reject step

  end

end

Projected Langevin algorithm (1)

Introduced in [**ACMT2024**], it combines the Unadjusted Langevin algorithm with a different low-rank prior.

- Burer-Monteiro factorization: $\rho = YY^\dagger$, with rank$(Y) = r$
- Low-rank prior: spectral scaled Student-t distribution

$$\pi(Y) = C_\theta \det(\theta^2 I_d + YY^\dagger)^{-(2d+r+2)/2} \tag{16}$$

equivalent to

$$\pi(Y) = \prod_{j=1}^{r} (\theta^2 + s_j(Y)^2)^{-(2d+r+2)/2} \tag{17}$$

with s_j the jth largest eigenvalue

Projected Langevin algorithm (2)

Note that there is no accept/reject step!

Algorithm 3: Projected Langevin algorithm

1 **for** $t \leftarrow 1 : T$ **do**

1. Sample $\tilde{w}^{(t)} \sim N(\mathbf{0}, \mathbf{I})$
2. $\tilde{Y}^{(t)} \leftarrow \tilde{Y}^{(t-1)} - \eta^{(t)} \nabla f(\tilde{Y}^{(t-1)}, \mathbf{D}) + \frac{\sqrt{2\eta^{(t)}}}{\beta} \tilde{w}^{(t)}$

   with $\pi(Y|\mathbf{D}) = \exp(-f(Y, \mathbf{D}))$

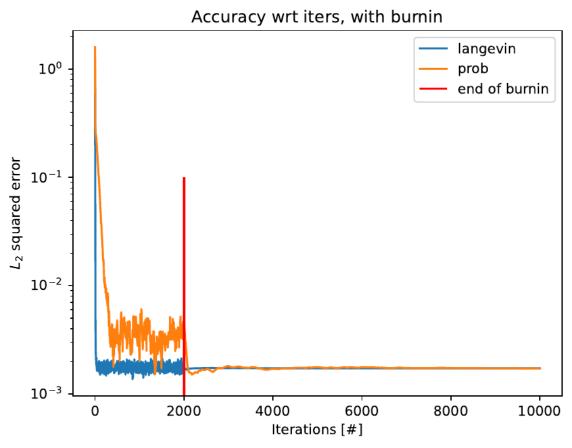
2 **end**

Table of Contents

Thesis contributions

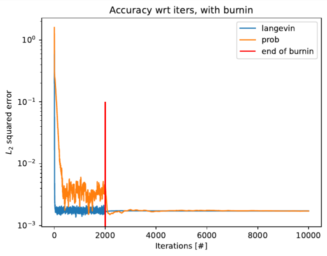
1. Numerically compare the prob-estimator and the Projected Langevin algorithm
2. Propose 2 new algorithms to understand the impact of the prior vs the algorithm on the accuracy

Numerical comparison: convergence

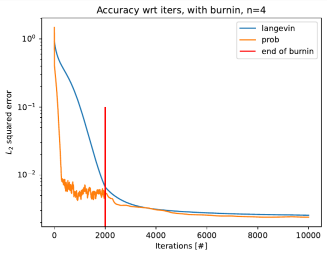


$\Longrightarrow$ Projected Langevin converges faster

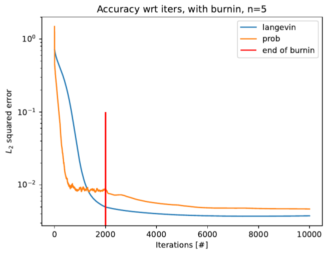
Numerical comparison: convergence across qubits (1)



(a) $n = 3$

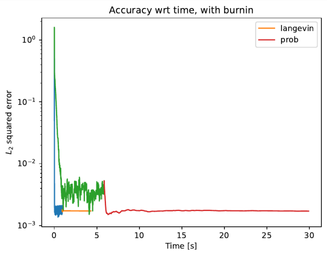


(b) $n = 4$

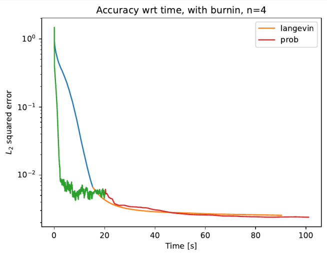


(c) $n = 5$

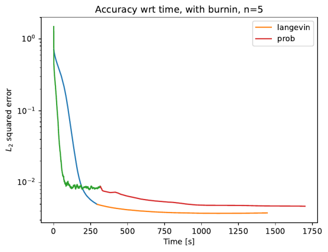
Numerical comparison: computation time across qubits (2)



(a) $n = 3$



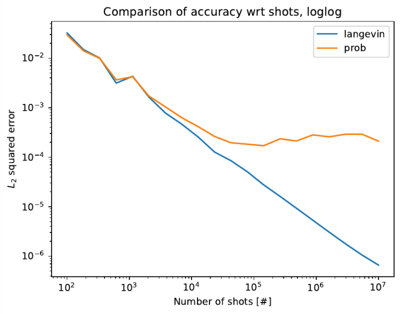
(b) $n = 4$



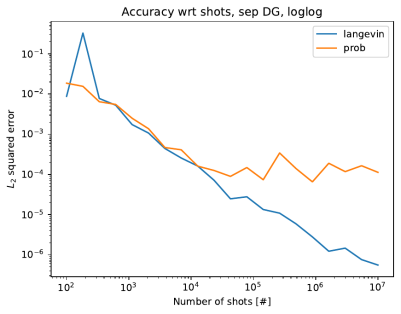
(c) $n = 5$

Numerical comparison: number of shots

Shot: measurement we perform on a clone of the state



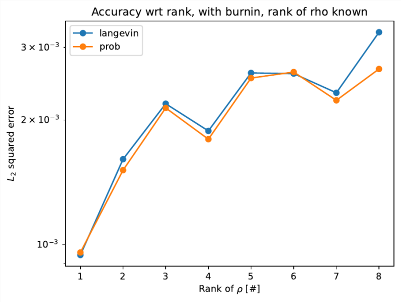
(a) Mixed qubit



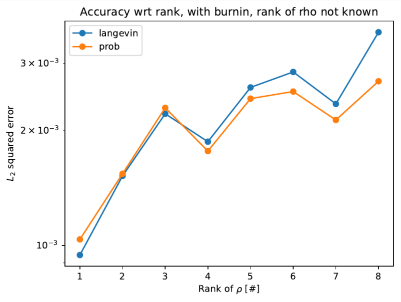
(b) Separate qubit

$\Longrightarrow$ The prob-estimator does not scale!

Numerical experiments: impact of knowledge of rank



(a) Rank of ρ known



(b) Rank of ρ not known

Figure 4: Rank knowledge plot for $n = 3$

$\implies$ For Projected Langevin, the information about the rank only marginally affects the accuracy

References