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

Danila Mokeev
Supervisors: Estelle Massart, Andrew Thompson and Tameem Adel

21st of June 2024

Ecole Polytechnique de Louvain

Plan of this thesis

Topic: Markov chain Monte Carlo (MCMC) methods in Quantum tomography

Research questions:

1. How do these methods perform in different experimental setups?
2. Why do some methods perform better than others?

Purpose:

- Enable new directions of research
- Help researchers make an informed choice for their use case

Thesis contributions

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

Table of Contents

Table of Contents

Motivation behind Quantum tomography

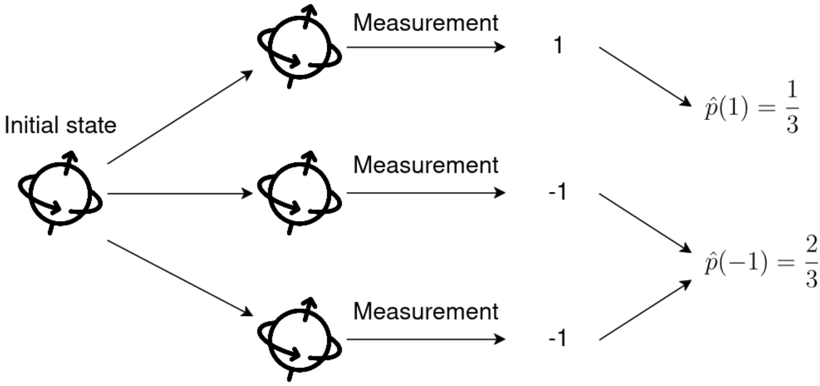
Quantum tomography is a process to reconstruct the quantum state of a system.

There are some challenges to consider:

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

Quantum tomography: a diagram

Quantum tomography allows to address the existing challenges



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$(\rho) = 1$
  - size $2^n \times 2^n$ with n the number of qubits

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

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

Most common methods

- Direct methods:

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

- Optimization-based methods:

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

- 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

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

Each term is a distribution

In the context of Quantum tomography:

- Likelihood $\mathcal{L}(\mathbf{D}|\rho) = \exp(-||A\vec{\rho} - \hat{p}||)$
- Prior $\pi(\rho)$ is method specific

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 calculated as

$$\rho = \int \nu \pi(\nu|\mathbf{D}) d\nu \tag{14}$$

$$\Leftrightarrow \hat{\rho} = \frac{1}{N} \sum_{i=1}^{N} \rho_i \quad \text{with } \rho_i \sim \pi(\rho|\mathbf{D}) \tag{15}$$

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

Table of Contents

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 1: Prob-estimator algorithm

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

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

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

  end

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

    1. Sample V_i^* from $\pi_2(V_1, \dots, V_d)$
    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 $\mathrm{rank}(Y) = r$
- 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)

Algorithm 2: Projected Langevin algorithm

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

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

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

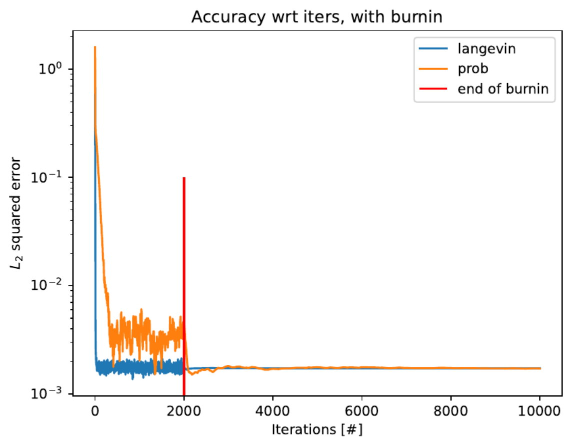
end

Observe that

- There is no accept/reject step
- We use the gradient of the posterior

Table of Contents

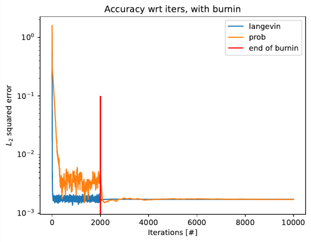
Convergence plot



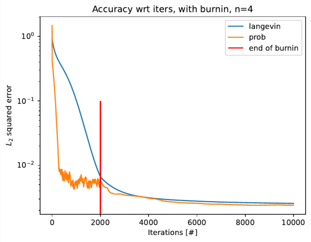
$\Longrightarrow$ Projected Langevin converges faster

Convergence across qubits (1)

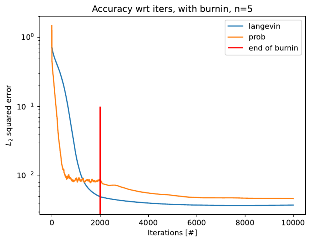
Reminder: n is the number of qubits



(a) $n = 3$



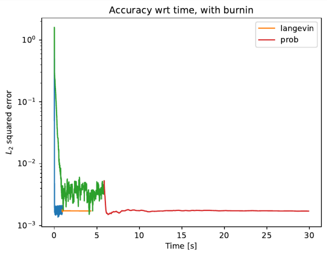
(b) $n = 4$



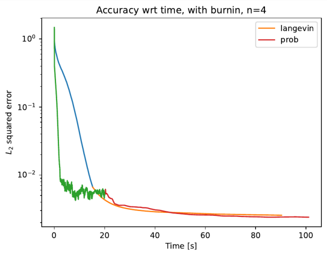
(c) $n = 5$

Projected Langevin converges faster and is more accurate for higher n! But..

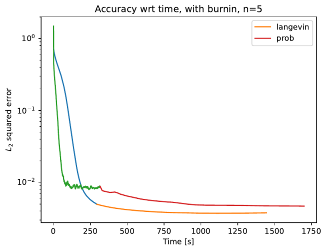
Computation time across qubits (2)



(a) $n = 3$



(b) $n = 4$



(c) $n = 5$

When n increases, the computation time does too!

Introducing 2 new methods

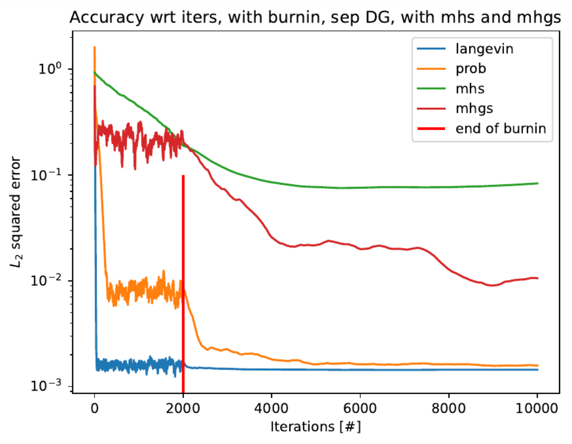
What makes Projected Langevin perform better ?

To answer this question, we introduce 2 new algorithms, Metropolis-Hastings with Student-t prior (MHS) and Metropolis-Hastings with Gibbs with Student-t prior (MHGS)

They combine:

- The algorithm from the prob-estimator
- The prior from the Projected Langevin algorithm

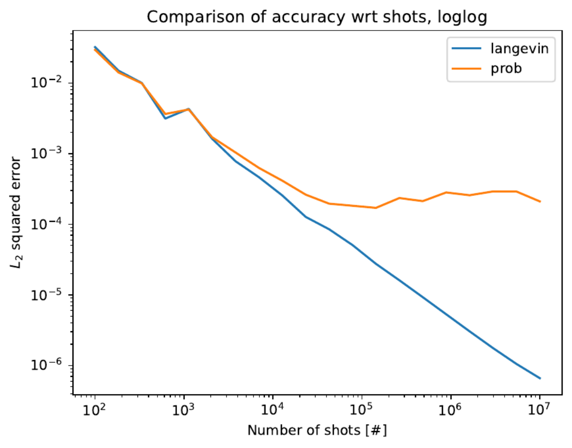
Convergence comparison



$\Longrightarrow$ The prior itself is not a solution, and must be paired with a fast algorithm

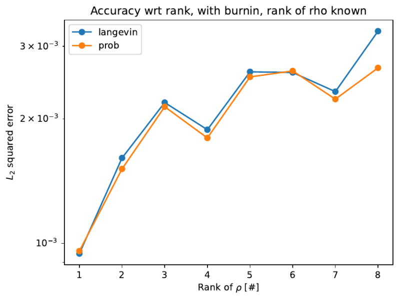
Numerical comparison: number of shots

Shot: measurement we perform on a clone of the state

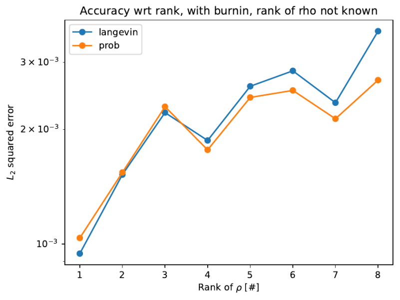


$\Longrightarrow$ The prob-estimator does not scale!

Numerical experiments: impact of knowledge of rank



(a) Rank of ρ known



(b) Rank of ρ not known

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

Summary and future work

- Quantum tomography is not yet a solved problem, especially for large systems
- MCMC methods are a promising direction of research, thanks to uncertainty quantification and prior information
- The choice of the algorithm might have more impact on the scalability of a method than the prior
- More experiments are needed to investigate the exact reasons of performance and scalability (for example Hamiltonian Monte Carlo)

References