Final Project Report: A Bayesian Probabilistic Model for Quantum State Tomography

Ryan Anselm (raa2218), Aunika Zheng (az2640)

December 21, 2023

Abstract:

We introduce a probabilistic model that captures the data-generating process for performing arbitrary, pre-specified measurements on quantum states. A key feature of our model is that performing approximate posterior inference with measurements from many identical copies of a quantum state allows one to find an approximate reconstruction of the state, a task known as quantum state tomography. We perform numerical experiments on simulated data to benchmark variational inference on our probabilistic model as a Bayesian approach to quantum state tomography. Furthermore, we investigate the use of measurement data collected from random positive operator-valued measure (POVM) sets in order to boost the performance of our model.

1 Introduction

Quantum state tomography (QST) is the task of approximately reconstructing a quantum state by making measurements on N identical copies of that state. QST is a fundamental primitive in the field of quantum information processing, where the ability to identify quantum states from measurements is a key step to benchmarking quantum algorithms and other quantum processes. For example, suppose we have a device like a quantum computer or a laser, and we wish to characterize how it deviates from its manufacturer's specifications, such as what distribution of noise it produces. Reconstructing the noisy quantum state allows for us to characterize the exact noise being produced and quantify its extent and probability distribution. A major challenge of QST is that it inherently scales poorly with system size because the size of the desired output, a density matrix, grows exponentially in the size of the system; however, QST remains a widely used technique in the engineering of small quantum devices.

Most protocols for quantum state tomography aim at conducting maximum likelihood estimation (MLE) of the density matrix by producing a point estimate of the density matrix most likely to give rise to the observed data. However, MLE is known to typically produce estimates that have one or more zero eigenvalues, making them implausible [2, 4], and furthermore only offers a point estimate of a single best state rather than a distribution over candidate states. We propose a Bayesian probabilistic model for quantum measurement that remedies some of the difficulties with maximum likelihood estimation. The framework of graphical models naturally allows us to reason with and design probabilistic models that incorporate various assumptions about the true density matrix through the dependency structure of the graphical model and the priors assigned to our latent variables. Furthermore, as measurement of a quantum state gives outcomes probabilistically, measuring a quantum state naturally draws parallels to sampling from a probabilistic model.

We provide a brief overview of relevant concepts from quantum information/quantum computing in Appendix A.

2 Methods

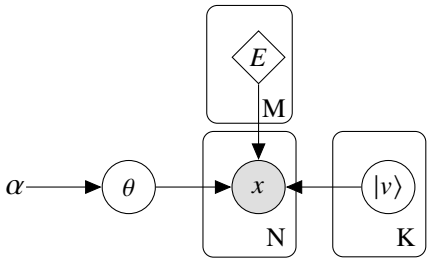


Figure 1: Graphical model for generating measurements of quantum states

2.1 Generative Process

Refer to Figure 1 for a picture of our graphical model. We consider data generated by performing the same measurement on N identical copies of a quantum state ρ of dimension d. Before any data is observed, our graphical model is given knowledge of what the possible outcomes of the measurement are in the form of a set of positive operator-valued measure (POVM) matrices corresponding to each of the outcomes (for example, a set of possible outcomes could be observing an electron to be in either the first energy level, the second energy level, or the third energy level of an atom). The i-th outcome can in general be represented by a $d \times d$ matrix E_i (detailed further in Section 2.2). We denote M to be the number of possible outcomes for a measurement and $E = \{E_1, E_2, \ldots, E_M\}$ to represent the set of possible outcomes. The matrices $\{E_1, E_2, \ldots, E_M\}$ act as hyperparameters of our model, and if we wish to investigate doing a different type of measurement we must update the set E accordingly.

To draw a sample from our observed variable x, we first sample K unit-length d-dimensional complex vectors, $\left|v^{(1)}\right\rangle, \ldots, \left|v^{(K)}\right\rangle$, uniformly at random. The value of K is a hyperparameter chosen to specify the maximum rank of our reconstruction estimate. In practice, we generate unit-length d-dimensional complex vectors by sampling from two d-dimensional multivariate standard normal distributions and letting one draw be the real components and one draw be the imaginary components of a d-dimensional complex vector. We then normalize the resulting complex vector to get a unit-length vector. In all, this resulting vector is drawn uniformly over the d-dimensional complex unit sphere.

Next, our model samples $\theta \in \Delta^K$ such that $\theta \sim \text{Dirichlet}(\alpha)$. θ is the *weighting parameter* that specifies how much weight should be assigned to each of the vectors $\left|v^{(1)}\right\rangle, \ldots, \left|v^{(K)}\right\rangle$ where the total weight must sum up to 1. The hyperparameter α is used as a prior on the weight distribution over vectors – choosing a low value of α favors mixtures concentrated on a few vectors while a high value of α favors more uniform mixtures. Finally, using the drawn values of θ and $\left|v^{(1)}\right\rangle, \ldots, \left|v^{(K)}\right\rangle$ we obtain our estimated 'density matrix'

$$\tilde{\rho} := \sum_{k=1}^{K} \theta_k \left|v^{(k)}\right\rangle \left\langle v^{(k)}\right|.$$

Then, we sample each observation x_i for $1 \leq i \leq N$ as

$$x_i \sim \text{Categorical}(\text{Tr}[E_1\tilde{\rho}], \ldots, \text{Tr}[E_M\tilde{\rho}])$$

where $\text{Tr}[AB]$ is the trace of the product of matrices A and B and $x \in \{1, \ldots, M\}$ is the number of the outcome that was observed. The structure of our probabilistic model is motivated by the rules of quantum mechanics, particularly **Born's rule** [6] which in its generalized version states that

$$Pr[E_k|\rho] = \text{Tr}[E_k\rho].$$

This means that the probability of seeing outcome k from a state ρ is the trace of the 'measurement matrix' E_k times the state's 'density matrix' ρ. At a high level, our model assumes a particular distribution over possible density matrices which it samples from to get a $\tilde{\rho}$, and then it applies Born's rule to this density matrix in order to get a probabilistic measurement outcome.

2.2 Positive Operator-Valued Measures (Generalized Measurements)

Typically, measurement of a d-dimensional quantum state probabilistically yields one of d possible outcomes – these are called standard or projective measurements. However, infinitely many distinct quantum states could yield the same probability distribution over the set of standard measurements. This is a major problem for our model because it means that there is a very spread-out high-dimensional distribution of density matrices that all match the observed data equally well. This makes approximating the posterior distribution more challenging or even intractable when doing any variety of approximate posterior inference.

To alleviate this challenge, we use positive operator-valued measures (POVMs), which are a generalization of standard measurements that can take on an arbitrarily *larger* number of outcomes than d. Mathematically, POVMs are complex-valued, positive semidefinite $d \times d$ matrices such that the sum over a valid set of POVMs equals the $d \times d$ identity matrix $\mathbb{I}_d$. A set of POVMs is informationally complete if every possible state corresponds to a unique probabillity distribution over the set of POVM outcomes – in other words, knowing the probability distribution theoretically should give all the information needed to uniquely determine the state. An informationally-complete (IC) set of POVMs for a d-dimensional system requires having at least d^2 POVM matrices in the set [8].

However, we need not always use an informationally complete set. We experimented with randomly generated sets of POVMs of various sizes (different values of M) ranging from d, the size of a projective measurement set, up to d^2, the IC threshold. We generated random sets of POVMs following a technique originally proposed in [10] and described as follows:

1. Generate M unit d-dimensional complex vectors uniformly at random: $v_1, \ldots, v_M \in \mathbb{C}^d$.
2. Compute the value of $G := \sum_{i=1}^{M} vv^*$.
3. Designate the new POVM set to be $E := \{G^{-\frac{1}{2}} v_1 v_1^* G^{-\frac{1}{2}}, \ldots, G^{-\frac{1}{2}} v_M v_M^* G^{-\frac{1}{2}}\}$.

If we pick M random complex vectors, the operators associated with them will almost surely be linearly independent from one another. Applying the transformation $a \rightarrow G^{-\frac{1}{2}} a G^{-\frac{1}{2}}$ ensures that each element of the POVM set will be a positive semidefinite matrix and that their sum is the identity matrix $\mathbb{I}$.

2.3 Simulation Procedure

We generated simulated measurement observations on several quantum states, in particular Greenberger-Horne-Zeilinger (GHZ) states and random states using quantum circuits made with Qiskit [7] to refine and validate our model (Figure 2). When using POVMs, alongside each of our states, we randomly generated a set of POVMs following the procedure in Section 2.2 and produced a set of 10,000 observations on copies of that state by doing a type of measurement specified by the POVM set. We then saved our POVM set and the observed data for use during variational inference.

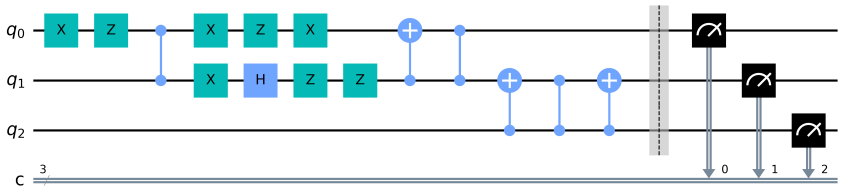
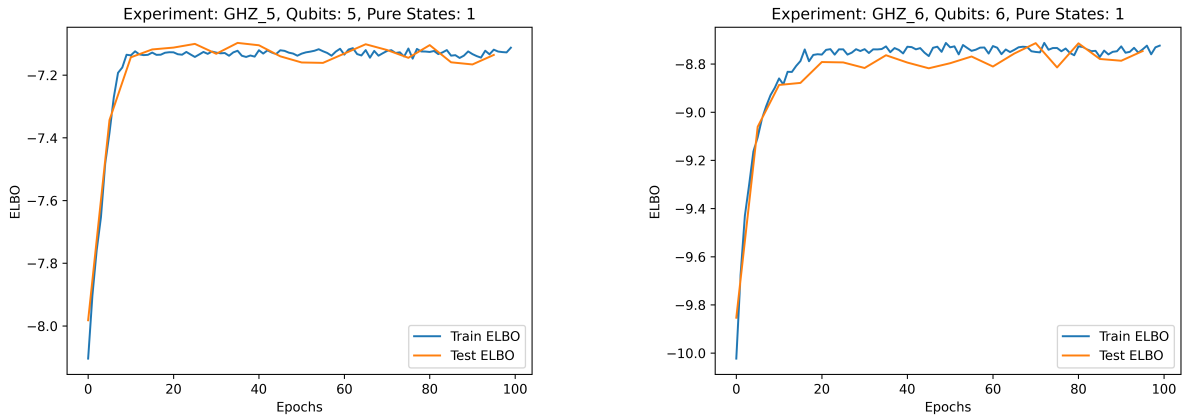


Figure 2: Example of a state preparation quantum circuit generated with Qiskit.

2.4 Inference Procedure

We used variational inference [1] in order to approximate the posterior distribution with a mean-field variational family of distributions. Our variational family parameterized θ with a vector of K weights with a simplex constraint, and parameterized each of entries of the vectors $|v\rangle$ as drawn from a multivariate normal distribution with diagonal covariance matrix with the mean and diagonal entries of the covariance matrix as parameters. In order to optimize the variational parameters to approximate the posterior, we maximized the evidence lower bound (ELBO), with representative examples of the ELBO being optimized shown in figure 3.



(a) 5-qubit GHZ state

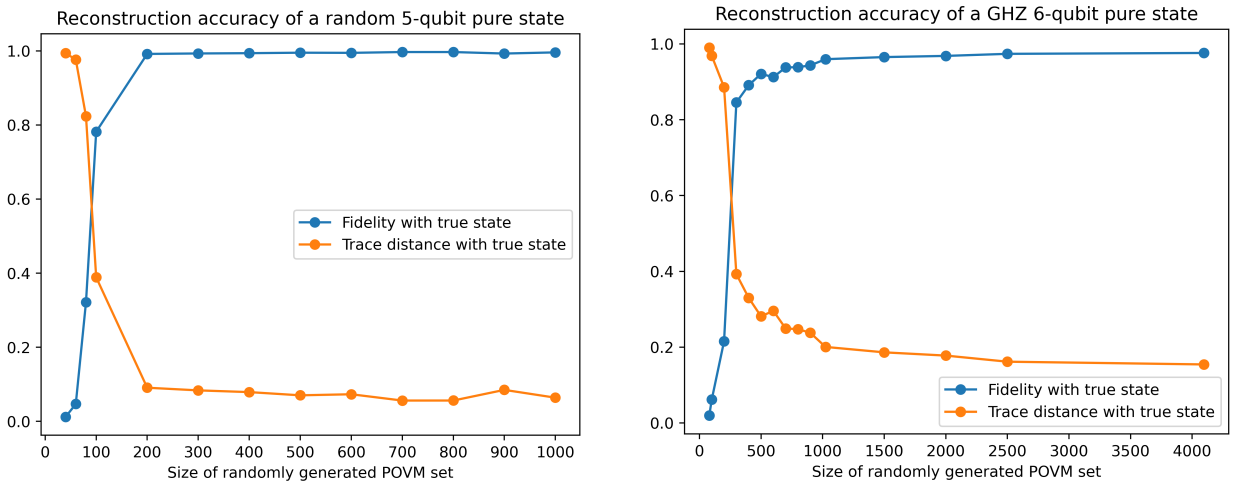
(b) 6-qubit GHZ state

Figure 3: Representative examples of the evidence lower bound (ELBO) of the inference model as a function of iterations for informationally-complete (IC) GHZ states

3 Results

We assessed the performance of our model by fitting it on simulated data where the true state was known to us. Using estimates of latent variables θ and $\left|v^{(1)}\right\rangle, \ldots, \left|v^{(K)}\right\rangle$ after variational inference, we reconstruct an estimate $\tilde{\rho}$ of the density matrix. Running our model on simulated data where we knew the true state enabled us to accurately measure how well the model performed through two measures, trace distance and fidelity. For a full description of these measures, see Appendix B. We varied parameters such as the number of qubits of a state, the size of POVM set used, and the weighting of a mixed state to see how these parameters affected performance. To interpret the plots easily, note that *higher fidelity and lower trace distance are better* – they imply that the model is doing better in reconstructing the original state.

We experimented with varying the size of the POVM set (the hyperparameter M) that our model was given. For a 5-qubit state of dimension $2^5 = 32$, at least $2^{2 \cdot 5} = 1024$ POVMs are needed for informational-completeness (IC); however, our model achieved >99% reconstruction fidelity on POVM sets of size around only 200 (Figure 4a). This was much lower than expected based on the informational-completeness threshold. We see a similar result around POVM sets of size around 500 for a 6-qubit state (Figure 4b), despite the IC-threshold being $2^{2 \cdot 6} = 4096$.



(a) 5-qubit state

(b) 6-qubit state

Figure 4: How increasing the size of POVM sets affects performance of the model. The performance of the model decreases with an increased number of qubits and the same number of POVMs, and more POVMs are required for larger quantum systems.

In Figure 5, we observe that our graphical model reconstructs density matrices of quantum states very well for small states with less than 6 qubits. However, performance rapidly degraded when attempting to reconstruct states on more qubits. Note that the dimension of a state is 2^n where n is the number of qubits and that we corresponding scaled up the size of the POVM set to stay at a fixed ratio of $\frac{1}{2}$ with the informationally-complete threshold of 2^{2n}. Even with scaled up POVM sets, our model begins to do worse with reconstructing larger states due to the exponentially increasing number of parameters to optimize over.

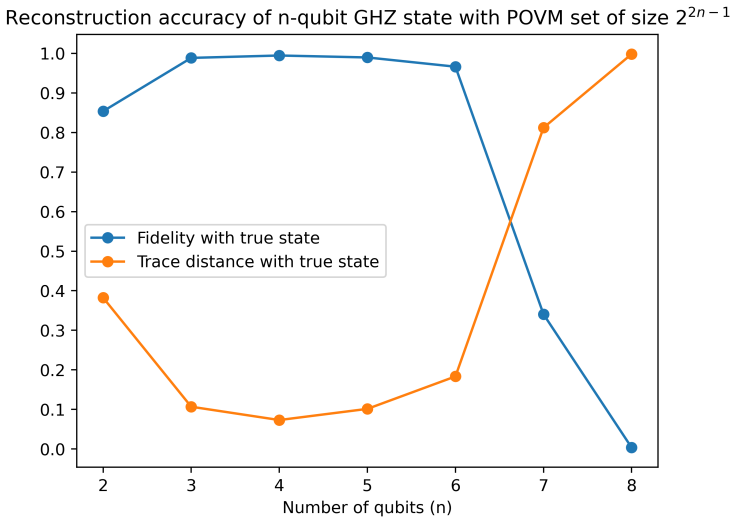
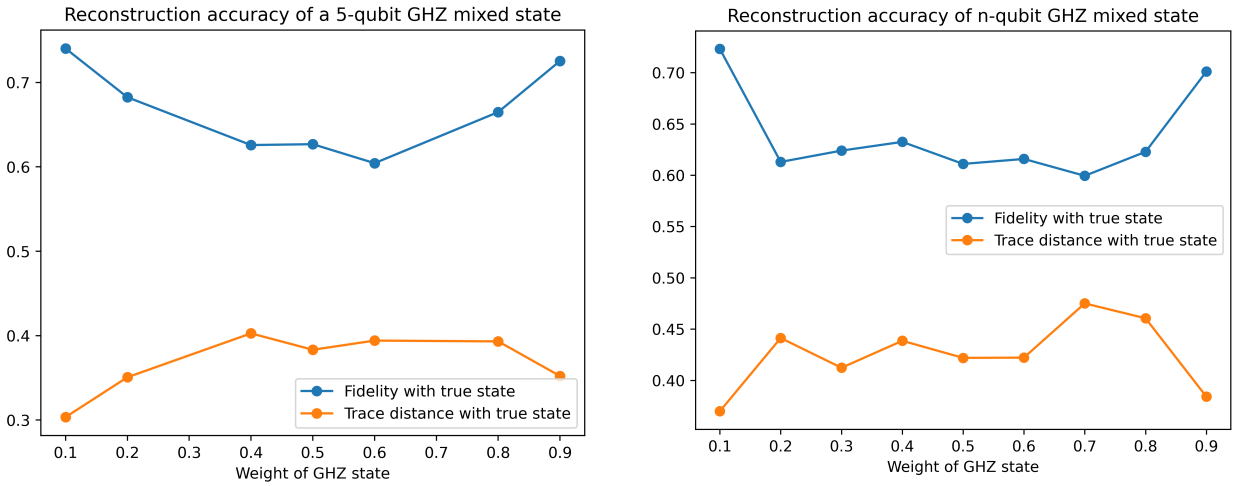


Figure 5: How increasing the number of qubits n affects performance of the model, with 2^{2n-1} POVMs.

In Figure 6, we observe the influence of our Dirichlet prior on the reconstruction accuracy of a mixed state with 2 pure states (K = 2). With $\alpha = 0.1$, the model performs better on mixed states that are concentrated mostly on one of the two states, while it does the worst on uniform mixtures of states. With $\alpha = 10$, there is not a clear inversion of the trend like we would expect due to the influence of the prior, as it still does the best on the mixed states that are mostly concentrated on one state. This could possibly be due to other factors such as a mixed state being more difficult to optimize the parameters for than a pure state, rather than due to the influence of the prior.



(a) $\alpha = 0.1$ (b) $\alpha = 10$

Figure 6: Reconstruction of a mixed state of a 5-qubit pure GHZ state and a pure all zeros state by model described by 200 POVMs with different settings of α, the Dirichlet hyperparameter. Weights were varied for the GHZ state in the mixed state.

4 Discussion

We have introduced a simple, interpretable Bayesian probabilistic model that is able to successfully reconstruct the density matrices of small (number of qubits < 6) quantum states with $> 99\%$ fidelity using variational inference. However, it falls short when reconstructing states from larger quantum systems because of the exponential scaling in the number of parameters required to describe quantum states, and it struggles somewhat more with reconstructing mixed states than pure states. In addition to this, the time of computation needed grows exponentially as we apply it on larger and larger systems.

One potential direction to pursue to remedy these issues of scalability is to explore the use of neural networks to amortize some aspects of the computation for quantum state tomography, a topic explored in [9, 3]. Additionally, more sophisticated priors can be introduced which could potentially have a beneficial effect for the posterior distribution. For example, rather than a uniform random prior over all unit-length complex vectors, a prior that favors sparse complex vectors could be used to lead to improved performance in some contexts.

An interesting takeaway from our experiments is that full d^2-sized informationally-complete POVM sets were not necessary to attain good performance, though having a POVM set larger than the number of dimensions was consistently essential for our model to attain good performance. The ability of our model to reconstruct states successfully even when underspecified by the POVMs provided corroborates with work about quantum state tomography via compressed sensing [5], which showed that for rank r and dimension d, using $O(rd \log^2 d)$ measurement outcomes rather than d^2 was sufficient to reconstruct the density matrix of a state, provided that the state was close to pure. Through our probabilistic model, we create a prior for close-to-pure density matrices, surprisingly enabling us to also achieve results comparable to compressed sensing using variational inference.

References

[1] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518 (Apr. 2017), pp. 859–877. ISSN: 1537-274X. DOI: 10.1080/01621459.2017.1285773. URL: http://dx.doi.org/10.1080/01621459.2017.1285773.

[2] Robin Blume-Kohout. "Optimal, reliable estimation of quantum states". In: *New Journal of Physics* 12.4 (Apr. 2010), p. 043034. ISSN: 1367-2630. DOI: 10.1088/1367-2630/12/4/043034. URL: http://dx.doi.org/10.1088/1367-2630/12/4/043034.

[3] Juan Carrasquilla et al. "Reconstructing quantum states with generative models". In: *Nature Machine Intelligence* 1.3 (Mar. 2019), pp. 155–161. ISSN: 2522-5839. DOI: 10.1038/s42256-019-0028-1. URL: http://dx.doi.org/10.1038/s42256-019-0028-1.

[4] Christopher Granade, Joshua Combes, and D G Cory. "Practical Bayesian tomography". In: *New Journal of Physics* 18.3 (Mar. 2016), p. 033024. ISSN: 1367-2630. DOI: 10.1088/1367-2630/18/3/033024. URL: http://dx.doi.org/10.1088/1367-2630/18/3/033024.

[5] David Gross et al. "Quantum State Tomography via Compressed Sensing". In: *Physical Review Letters* 105.15 (Oct. 2010). ISSN: 1079-7114. DOI: 10.1103/physrevlett.105.150401. URL: http://dx.doi.org/10.1103/PhysRevLett.105.150401.

[6] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[7] Qiskit contributors. *Qiskit: An Open-source Framework for Quantum Computing*. 2023. DOI: 10.5281/zenodo.2573505.

[8] Joseph M. Renes et al. "Symmetric informationally complete quantum measurements". In: *Journal of Mathematical Physics* 45.6 (May 2004), pp. 2171–2180. ISSN: 1089-7658. DOI: 10.1063/1.1737053. URL: http://dx.doi.org/10.1063/1.1737053.

[9] Giacomo Torlai et al. "Neural-network quantum state tomography". In: *Nature Physics* 14.5 (May 2018), pp. 447–450. ISSN: 1745-2481. DOI: 10.1038/s41567-018-0048-5. URL: https://doi.org/10.1038/s41567-018-0048-5.

[10] Stefan Weigert. "Simple Minimal Informationally Complete Measurements for Qudits". In: *International Journal of Modern Physics B* 20.11n13 (May 2006), pp. 1942–1955. ISSN: 1793-6578. DOI: 10.1142/s021797920603442x. URL: http://dx.doi.org/10.1142/S021797920603442X.

Appendix A: Crash Course on Quantum Information Jargon

In quantum state tomography, measuring many identical states is necessary to perform a reconstruction because a quantum state is a complex number weighted *superposition* of outcomes until it is measured, upon which point only one of the possible outcomes is observed with a probability distribution over outcomes determined by the complex weights.

For the purposes of our model, we look at quantum states acting on n qubits. These states can be described by *statevectors* of $d = 2^n$ complex numbers. For a state $v \in \mathbb{C}^d$, its density matr ix is $\rho_v := vv^*$ where v^* denotes the transpose of the complex conjugate of v. ρ_v is a $d \times d$ complex matrix and is considered a *pure state* because it comes from a single statevector. Density matrices can also be probabilistic mixtures of other density matrices. For example, we could have $\rho = \frac{1}{3}vv^* + \frac{2}{3}ww^*$ for $v, w \in \mathbb{C}^d$, which would be considered a mixed state.

Suppose we have a mixed quantum state on n qubits that is described by a density matrix $\rho \in \mathbb{C}^{2^n \times 2^n}$. We define a set of K measurement outcomes $E = \{E_k \in \mathbb{C}^{2^n \times 2^n}\}_{k=1}^K$ that can be obtained from measurement on this quantum state. Born's rule states the probability of seeing outcome E_k from state ρ is $Pr[E_k|\rho] = \mathrm{Tr}[E_k\rho]$. Therefore, the likelihood of observing a list of N measurement outcomes $\mathscr{M} = \{M_1, \ldots M_N\}$ from ρ can be written

$$\mathscr{L}(\rho) = \prod_{i=1}^{N} Pr[M_i|\rho] = \prod_{i=1}^{N} \mathrm{Tr}[M_i\rho] = \mathrm{Tr}[E_1\rho]^{n_1}\mathrm{Tr}[E_2\rho]^{n_2} \ldots \mathrm{Tr}[E_K\rho]^{n_K} \tag{1}$$

where $n_k := \sum_{i=1}^{N} \mathbb{1}[M_i = E_k]$, the number of times measurement outcome E_k was observed. In principle, we can find a maximum likelihood estimate $\hat{\rho}_{\mathrm{MLE}} := \arg\max_\rho \log \mathscr{L}(\rho)$ which is the density matrix that would maximize the probability of observing the measurements seen.

Appendix B: Measures of Reconstruction Accuracy

Assume that $\rho, \sigma \in \mathbb{C}^{d \times d}$.

Trace Distance:

$$T(\rho, \sigma) := \frac{1}{2}\mathrm{Tr}\left[\sqrt{(\rho - \sigma)^\dagger(\rho - \sigma)}\right]$$

The trace distance is a measure of distinguishability between two states that ranges between 0 and 1, with a higher trace distance corresponding to farther apart states and a trace distance of 0 for $\rho = \sigma$.

Fidelity:

$$F(\rho, \sigma) := \left(\mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2$$

Fidelity also ranges between 0 and 1, with a higher fidelity corresponding to closer states, with a fidelity of 1 for $\rho = \sigma$. [6]