Maximum Likelihood Estimation of Quantum State Tomography Data

Brian Shoffeitt and Jenna Zaidspiner
18 April 2022
MATH 3236 – Statistical Theory
Georgia Institute of Technology

We replicate the maximum likelihood estimation performed by [1] to determine the quantum state sent through their elementary quantum network. While the estimated density matrix agrees in magnitude with the estimate from [1], the phases of the off-diagonal components vary considerably. This discrepancy is investigated through the use of parametric bootstrapping.

I. INTRODUCTION

The ability to distribute entangled quantum states over long distances is the crux of a quantum network, with applications ranging from quantum cryptography to distributed quantum computing. Much effort has been placed into improving the rate at which entanglement can be transported in order to build faster and more error-tolerant quantum networks. These protocols are prone to noise, so it is important to assess how much the states that are actually sent deviate from the target states. Directly measuring a quantum state inevitably destroys some of the information about the initial state, so we are left to use more subtle methods. One common choice is a procedure known as quantum state tomography (QST), which involves performing repeated measurements on an ensemble of similarly prepared particles to estimate the average state of the ensemble. Since the measurement outcomes are probabilistic, this amounts to regarding the state of the ensemble as a parameter, which we would like to estimate from the QST data.

In this work, we replicate the maximum likelihood estimation performed by [1] to estimate the quantum state sent through their elementary quantum network. Additionally, we employ parametric bootstrapping in order to investigate the discrepencies between our results and those of [1]. We provide a brief background in quantum information theory, assuming only knowledge from linear algebra.

II. BACKGROUND

A. Two-Level Quantum System

The state of a two-level quantum system, known as a **qubit**, may be described with a vector $|\psi\rangle$ in the inner product space $\mathbb{C}^2$. Physically, a qubit may consist of anything from two different atomic orbitals to the polarization states of a photon. The two different states are commonly labeled $|0\rangle$ and $|1\rangle$, and together they are known as the **computational basis**. We commonly represent states in matrix form with respect to this basis, for example

$$|a\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}.$$

The row vector equal to the conjugate transpose of $|\psi\rangle$ is denoted $\langle\psi|$. Given $|\psi\rangle, |\varphi\rangle \in \mathbb{C}^2$, the inner product $\langle|\psi\rangle, |\varphi\rangle\rangle$ is denoted[1] $\langle\varphi|\psi\rangle$ and may be calculated via matrix multiplication of $\langle\varphi|$ and $|\psi\rangle$. For example, if $|a\rangle$ is as before and $|b\rangle = |1\rangle$ (these two vectors will be used for examples throughout),

$$\langle a|b\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -i\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix} = -\frac{i}{\sqrt{2}}.$$

We cannot directly observe the full state of a qubit, as performing an observation on the qubit alters its state. Instead, we can perform measurements with respect to orthogonal bases to gain some information about the initial state (e.g. detecting the polarization of a photon). Suppose a qubit is initially in the state $|\psi\rangle$ and we perform a measurement with respect to some basis $\{|\varphi_1\rangle, |\varphi_2\rangle\}$. The probability of observing the qubit to be in the state $|\varphi_1\rangle$ is $|\langle\psi|\varphi_1\rangle|^2$. After such an observation, the qubit is left in the state $|\varphi\rangle$. In order to represent a valid quantum state, a vector $|\psi\rangle$ must therefore be normalized so that $\langle\psi|\psi\rangle = 1$. For example, if we measure a qubit initially in the state $|a\rangle$ with respect to the computational basis $\{|0\rangle, |1\rangle\}$, we obtain the state $|1\rangle = |b\rangle$ with probability $|\langle a|b\rangle|^2 = |-i/\sqrt{2}|^2 = 1/2$.

B. Two Qubits and Entangled States

In order to describe the interaction between two qubits, we need to introduce the concept of a tensor product of vector spaces. The essential behavior of a tensor product in matrix form can be captured with the Kronecker product. If A and B are $m \times n$ and $p \times q$ matrices, respectively, the **Kronecker product** of A and B is the $mp \times nq$ block matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}.$$

If we have two qubits in the states $|\psi\rangle$ and $|\varphi\rangle$, the state of the composite system is described by the vector $|\psi\rangle \otimes$

[1] Note that the order is switched to ensure linearity in the second argument of the inner product, as is convention in quantum mechanics. That is, $\langle\varphi|\,(a\,|\psi\rangle) = a\,\langle\varphi|\psi\rangle = (\bar{a}\,\langle\varphi|)\,|\psi\rangle$ for $a \in \mathbb{C}$.

$|\varphi\rangle \in \mathbb{C}^4$. For example, with one qubit in state $|a\rangle$ and the other in state $|b\rangle$, their combined state is

$$|a\rangle \otimes |b\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \cdot |b\rangle \\ i \cdot |b\rangle \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 0 \\ i \end{pmatrix}.$$

This vector is expressed in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, where $|ij\rangle$ is shorthand for $|i\rangle \otimes |j\rangle$. The state $|01\rangle$, for example, corresponds to the first qubit in the $|0\rangle$ state and the second in the $|1\rangle$ state.

Not every vector in $\mathbb{C}^4$ may be written as the tensor product of two vectors, for example the Bell state

$$|\Psi^+\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

Nonetheless, these are still valid states for two interacting qubits to have; any normalized vector in $\mathbb{C}^4$ represents a possible state. Any two-qubit state that cannot be factored into the tensor product of two one-qubit states is said to be **entangled**.

C. Statistical Ensembles

Often, we are interested not in the state of one individual qubit but in the states of a collection of similarly prepared qubits. Suppose we have a group of particles, called an **ensemble**, in which each of the N states $|\psi_i\rangle$ appears with probability p_i. This ensemble is described by the **density matrix**

$$\varrho = \sum_{j=1}^{N} p_j |\psi_j\rangle\langle\psi_j|.$$

Density matrices are necessarily positive semi-definite (and hence Hermitian) and they have trace one; in fact, they are characterized by these two properties. We define $\mathcal{S}(\mathcal{H})$ to be the set of positive-semi definite operators on inner product space $\mathcal{H}$ with trace one.

An equal mixture of the computational basis states $|0\rangle$ and $|1\rangle$ has density matrix

$$\varrho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

in the computational basis. Note that this is different from an equal superposition $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ of the two states, which has density matrix

$$|\psi\rangle\langle\psi| = \frac{1}{2}(|0\rangle + |1\rangle)(\langle 0| + \langle 1|) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The latter is definitively in one state, while the former is an even mixture of two distinct quantum states; the two can be differentiated with measurement in a basis containing $|\psi\rangle$. Density matrices of the form $\varrho = |\psi\rangle\langle\psi|$ (or equivalently density matrices of rank one) are said to represent a **pure state**.

The actual target state for the protocol in [1] is $|\Psi_\phi\rangle = (|01\rangle + e^{i\phi}|10\rangle)/\sqrt{2}$, where the phase ϕ is stable between each measurement. This has the density matrix

$$\varrho_\phi = |\Psi_\phi\rangle\langle\Psi_\phi| = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & e^{-i\phi} & 0 \\ 0 & e^{i\phi} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Suppose we perform a measurement on a random state in an ensemble with density matrix $\varrho \in \mathcal{S}(\mathbb{C}^m)$ in the orthonormal basis $\{\varphi_1, \ldots, \varphi_m\}$. The probability of measuring state $|\varphi_j\rangle$ is $\mathrm{tr}(\Pi_j \varrho)$, where $\Pi_j = |\varphi_j\rangle\langle\varphi_j|$ is the projector onto $|\varphi_j\rangle$. The post-measurement state is

$$\varrho' = \frac{\Pi_j \varrho \Pi_j}{\mathrm{tr}(\Pi_j \varrho \Pi_j)}.$$

D. Quantum State Tomography

Even though we cannot directly measure the state of a quantum system, we can still measure the density operator of an ensemble of identically prepared states with a procedure known as **quantum state tomography** (QST). This involves taking measurements in several different bases and estimating the ensemble state.

In a 4×4 Hermitian matrix, we have 16 degrees of freedom: the 12 real and imaginary components of the six elements of the upper triangular block and the 4 elements on the main diagonal (which must be real). Requiring a trace of one removes one degree of freedom, so we are left with 15 degrees of freedom for an arbitrary $\varrho \in \mathcal{S}(\mathbb{C}^4)$. We therefore must perform measurements with 15 different projectors in order to fully specify ϱ. In practice, however, it is better to use an overcomplete set of measurements to help protect against measurement errors and better estimate small components of ϱ.

A common choice of basis for QST measurements is given by the orthogonal eigenvectors of the Pauli matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

as these are easy to implement experimentally. The Pauli matrices have eigenvalues 1 and -1, so their eigenvectors are labeled $|+\rangle_A$ and $|-\rangle_A$ for $A \in \{X, Y, Z\}$. In [1], measurements are performed with respect to the basis consisting of the four orthogonal eigenvalues of the matrices $A \otimes B$. This is done for each of the nine pairs $(A, B) \in \{X, Y, Z\}^2$ for a total of 36 projectors. These eigenvectors are simply the tensor products of the eigenvectors of the individual Pauli matrices: $A \otimes B$ has the four eigenvectors $|\sigma_1\rangle_A \otimes |\sigma_2\rangle_B$ where $\sigma_1, \sigma_2 \in \{+, -\}$. The 36 projectors used are therefore of the form

$$\Pi^{AB}_{\sigma_1\sigma_2} = (|\sigma_1\rangle_A \otimes |\sigma_2\rangle_B)(\langle\sigma_1|_A \otimes \langle\sigma_2|_B)$$

for $A, B \in \{X, Y, Z\}$ and $\sigma_1, \sigma_2 \in \{+, -\}$. For simplicity, we index all the measurement outcomes with the index set J, so that $\{\Pi_\alpha : \alpha \in J\}$ is the set of 36 projectors used.

Table 1. QST data from [1]

		$\sigma_1\sigma_2$			
A	B	$++$	$-+$	$+-$	$--$
Z	Z	4	424	564	10
Z	X	255	217	297	194
Z	Y	283	218	316	193
X	Z	212	222	271	263
X	X	259	236	227	289
X	Y	28	463	477	26
Y	Z	178	213	288	285
Y	X	424	32	32	524
Y	Y	249	223	193	295

III. MLE OF DENSITY MATRICES

In statistical terms, measurement with respect to a basis corresponds to a categorical distribution with four outcomes and the initial state ϱ as a parameter of the distribution. We can form a likelihood function for each of these bases and maximize each of them simultaneously by maximizing their product. Suppose we take N measurements and observe each outcome α a total of n_α times. The combined log-likelihood function is

$$\ell(\varrho) = \ln \prod_{\alpha\in J} [\mathrm{tr}(\Pi_\alpha \varrho)]^{n_\alpha} = \sum_{\alpha\in J} n_\alpha \ln[\mathrm{tr}(\Pi_\alpha \varrho)].$$

Optimizing this by differentiating with respect to the components of ϱ is impractical, but fortunately there is a common algorithm for finding maximum likelihood estimates for density matrices [2]. This is the same algorithm used in [1]. We define the operator

$$R(\varrho) = \frac{1}{N} \sum_{\alpha\in J} \frac{n_\alpha}{\mathrm{tr}(\Pi_\alpha \varrho)} \Pi_\alpha.$$

It can be shown that the state $\hat{\varrho}$ maximizing $\ell(\varrho)$ satisfies $R(\hat{\varrho})\hat{\varrho} = \hat{\varrho}$. This allows us to use a common fixed-point algorithm for Hermitian matrices that iteratively sets

$$\varrho_{k+1} = \frac{R(\varrho_k)\varrho_k R(\varrho_k)}{\mathrm{tr}[R(\varrho_k)\varrho_k R(\varrho_k)]}.$$

There is no guarantee that this algorithm converges as is. However, by diluting $R(\varrho)$ to

$$R'(\varrho) = \frac{I + \varepsilon R(\varrho)}{1+\varepsilon},$$

for some sufficiently small $\varepsilon < 1$, we can ensure that $\ell(\varrho_k)$ increases with each iteration. As $\ell(\varrho)$ is concave, this guarantees convergence [2].

IV. RESULTS

Although [1] analyzes four similar QST datasets, we only consider the first for simplicity.[2] These data are given in Table 1.

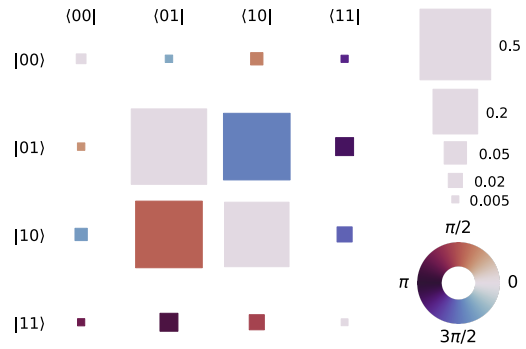

Figure 1. Magnitude (size) and complex phase (color) of the components of our estimate $\hat{\varrho}$.

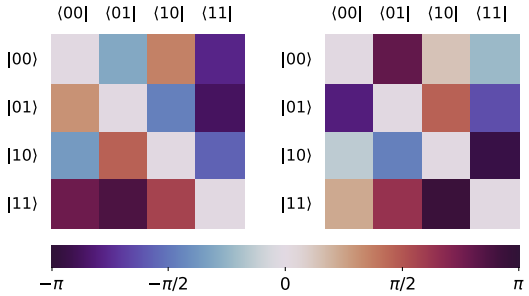

Figure 2. Comparison of the phase between our estimated density matrix (left) and that of [1].

The MLE $\hat{\varrho}$ that we obtained is shown in Figure 1. The magnitudes of the components of $\hat{\varrho}$ are identical to those reported by [1] within the reported three decimal places. However, the phases of the off-diagonal components differ considerably, as is evident in Figure 2. We expect the off-diagonal components other than $\hat{\varrho}_{32}$ and $\hat{\varrho}_{23}$ (these ten components are hereby referred to as off-center components) to vary in phase considerably, as they are small in magnitude. This means that the measurement outcomes that provide information on these phases are uncommon.

However, the discrepancy in the stable phase ϕ of $\hat{\varrho}_{32}$ is unexpected. We obtain a value of $\phi = 1.450$, while [1] obtains a value of -1.448. These differ in absolute value by only 0.1%. In order to investigate whether this is coincidental, we perform parametric bootstrapping. This entails using our estimate $\hat{\varrho}$ to simulate the outcomes of performing the same QST measurements many times, then finding maximum likelihood estimates of the density matrices for each simulated dataset. While this does not directly tell us anything about the differences between our results and those of [1], any discrepancies larger than those observed from bootstrapping suggest a systematic error occurred.

An empirical cdf of ϕ obtained from bootstrapping is plotted in Figure 3(a). The empirical distribution appears to be normal.[3] Indeed, performing a χ^2 test of

[2] The QST data and statistical analysis performed in [1] may be found in the supplemental material on arXiv.

[3] While the distribution cannot be truly normal (being supported on $[0, 2\pi)$), for small variance the difference is negligible.

(a)

(b)

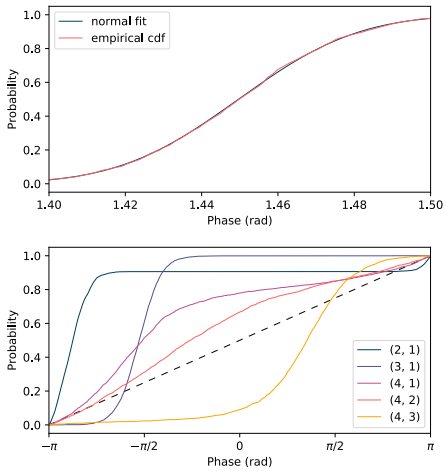

Figure 3. (a) Empirical cdf of the phase of the ϱ_{32} component of the bootstrap density matrices, with a standard normal fit of mean 1.450 rad and standard deviation 0.025 rad. (b) Empirical cdfs of the off-center phases, with a dashed line indicating a uniform distribution.

normality, we obtain a p-value of 0.785. The observed variance in ϕ is 0.025 rad. This is small, especially compared to the variation in the off-center phases (plotted in Figure 3(b)), indicating that a systematic error is almost certainly responsible for the discrepancy.

After carefully checking our work, we suspect that this error is on the part of [1]. This is supported by computing the log-likelihood function for our estimate and that of [1]; the latter has a log-likelihood that is 37% larger than the former. One possible cause for this is incorrectly computing the outer product with the complex conjugate on the column vector. Doing so, we obtain the same stable phase ϕ as [1] within 0.1%. While this discrepancy in the phase ϕ may seem immaterial, a difference like this can have disastrous effects for common quantum computing algorithms. Nonetheless, we were mostly successful in replicating the results of [1], as we obtained precisely the same magnitudes of the components, and the only statistically significant phase differed negligibly except in sign.

[1] L. J. Stephenson, D. P. Nadlinger, B. C. Nichol, S. An, P. Drmota, T. G. Ballance, K. Thirumalai, J. F. Goodwin, D. M. Lucas, and C. J. Ballance, Phys. Rev. Lett. **124**, 110501 (2020). (Supplemental material on arXiv).

[2] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky, Phys. Rev. A **75**, 042108 (2007).