CHAPTER 4

Algorithms for State Tomography

One of the most fundamental objects in quantum theory is the quantum state, or density matrix. Therefore, one of the most fundamental problems in quantum learning theory is that of learning a density matrix ρ. More formally, suppose we are given access to a device that prepares some unknown state ρ. We can make measurements on ρ, and then ask for a new copy of ρ. How many copies of ρ do we need to measure such that we can learn ρ to some fixed precision? This is the problem of **quantum state tomography**. The word tomography comes from the Greek 'tomos', meaning 'slice' or 'section', and 'grapho', meaning 'to write' or 'to describe'. In this sense, quantum state tomography results in a description of an unknown quantum state, attained by one measurement or 'slice' at a time.

At first, we will be concerned with the problem of **query complexity**; that is, *how many* ρ's do we need to learn a description of the unknown state to fixed precision? This will depend on exactly what we mean by *fixed precision* (we will see that there are multiple plausible definitions with different features), as well as what kinds of measurements we allow ourselves. As for the latter, we will start by considering **single-copy measurements**, namely where we can only measure one ρ at a time. More generally, we can make entangled measurements on multiple copies of ρ; this will give more optimal query complexity bounds, and we will pursue this later.

It is worth emphasizing that query complexity is distinct from **gate complexity**, which quantifies the number of operations required to perform each measurement. That is, by initially only quantifying query complexity, we are only counting the number of total measurements required, and will not initially be attentive to the difficulty of making each measurement. (We will, however, comment on the difficulties of performing certain kinds of measurements as we go along.) We remark that the gate complexity of a protocol is always lower bounded by the query complexity. Roughly speaking, if we need to query a state k times, then we require *at least* k gate operations. Thus the query complexity at the very least gives us a lower bound on the absolute difficulty of implementing a given protocol in practice.

1. Basic State Tomography

We will consider density matrices ρ on a Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$, as usual. The standard orthonormal basis of $\mathbb{C}^d$ is $\{|i\rangle\}_{i=0}^{d-1}$, and we can consider the matrices $E_{ij} := |i\rangle\langle j|$ where $\{E_{ij}\}_{i,j=0}^{d-1}$ is a complete orthonormal (with respect to Hilbert-Schmidt) basis for $\mathbb{C}^{d\times d}$. Indeed $\mathrm{tr}(E_{ij}^\dagger E_{k\ell}) = \delta_{ik}\delta_{j\ell}$, and we can write any density matrix ρ as

$$\rho = \sum_{i,j=0}^{d-1} \rho_{ij}\, E_{ij}\,.$$

Our first goal is to provide a protocol for measuring all of the ρ_{ij}'s. Now the E_{ij}'s for $i \neq j$ are not Hermitian operators, we cannot measure $\rho_{ij} = \mathrm{tr}(E_{ij}^\dagger \rho) = \langle i|\rho|j\rangle$ directly. So it is natural to separate our analysis into measuring the diagonal components of ρ, namely ρ_{ii}, and measuring the off-diagonal components of ρ, namely ρ_{ij} for $i \neq j$.

Measuring the diagonal components. To measure the diagonal components of ρ, we need to get good estimates for $\rho_{ii} = \mathrm{tr}(E_{ii}\rho) = \langle i|\rho|i\rangle$ for $i = 0, ..., d-1$. To do so, we can simply measure ρ in the $\{|i\rangle\}_{i=0}^{d-1}$ basis. Upon making a measurement, our apparatus will output 'i' with probability $\langle i|\rho|i\rangle$. So suppose we make N_{diag} total measurements, requiring as many copies of ρ. If N_i is the number of times our measurement apparatus outputted 'i' during those trials, then we can estimate ρ_{ii} by

$$\widehat{\rho}_{ii} := \frac{N_i}{N_{\text{diag}}}\,,$$

which is just the fraction of the times that the apparatus outputted 'i'. We will return soon to estimating the size of N_{diag} such that we can guarantee that e.g. $|\rho_{ii} - \widehat{\rho}_{ii}| \leq \varepsilon$ for all i.

Measuring the off-diagonal components. While the E_{ij} are not Hermitian for $i \neq j$, we can instead consider

$$\begin{aligned} E_{ij}^{+} &= |i\rangle\langle j| + |j\rangle\langle i| \\ E_{ij}^{-} &= -\mathrm{i}\,(|i\rangle\langle j| - |j\rangle\langle i|) \end{aligned}$$

which are Hermitian (in E_{ij}^{-} the i out front is the imaginary number, not an index label), and satisfy

$$E_{ij} = \frac{1}{2}\left(E_{ij}^{+} + \mathrm{i}\, E_{ij}^{-}\right).$$

Thus if we can estimate $\mathrm{Re}(\rho_{ij}) = \frac{1}{2}\,\mathrm{tr}(E_{ij}^{+}\rho)$ and $\mathrm{Im}(\rho_{ij}) = -\frac{1}{2}\,\mathrm{tr}(E_{ij}^{-}\rho)$, then we can estimate $\mathrm{tr}(E_{ij}\rho)$.

To organize these two–outcome blocks into a small number of measurement settings, we group disjoint pairs of indices so that many (i,j)'s are probed simultaneously. To this end, let us fix a decomposition of the complete graph on vertices $\{0, 1, \ldots, d-1\}$ into disjoint perfect matchings $M_1, \ldots, M_T$. For even d one may take $T = d-1$; for odd d one can take $T = d$ where in each round exactly one index is left unpaired (the unpaired index varies from round to round). For each matching M_t we use two measurement settings:

$$\mathsf{R}_t := \left\{\, |\psi_{ij,\pm}^{(R)}\rangle = \tfrac{|i\rangle \pm |j\rangle}{\sqrt{2}} : \{i,j\} \in M_t \,\right\}, \quad \mathsf{I}_t := \left\{\, |\psi_{ij,\pm}^{(I)}\rangle = \tfrac{|i\rangle \pm \mathrm{i}|j\rangle}{\sqrt{2}} : \{i,j\} \in M_t \,\right\},$$

implemented as a block-diagonal projective measurement whose 2×2 blocks are the two-outcome projectors on each pair in M_t (and, if d is odd, singleton projectors for the unpaired index). Now let $N_{\text{off},t}^{\mathrm{R}}$ (respectively $N_{\text{off},t}^{\mathrm{I}}$) denote the number of copies of ρ we measure in setting R_t (respectively I_t). The total number of copies used for off-diagonals is therefore

$$N_{\text{off}} = \sum_{t=1}^{T} \big(N^{\text{R}}_{\text{off},t} + N^{\text{I}}_{\text{off},t}\big).$$

A single outcome in a given setting contributes information to *all* pairs in that matching. Note that we do not dedicate separate copies to each pair; instead, the copies from round t are shared statistically across the pairs $\{i,j\} \in M_t$.

Fix a round t of the R ("real") setting. For each pair $\{i,j\}$ that appears in round t, write $m^{\text{R},t}_{ij,+}$ and $m^{\text{R},t}_{ij,-}$ for the observed counts of the $+$ and $-$ outcomes in the $\{i,j\}$ two-dimensional block. Then define

$$\widehat{\text{Re}(\rho_{ij})} := \frac{m^{\text{R},t}_{ij,+} - m^{\text{R},t}_{ij,-}}{2\,N^{\text{R}}_{\text{off},t}}\,, \tag{20}$$

which is our estimator for $\text{Re}(\rho_{ij})$.

Analogously, in the I ("imaginary") setting for the same round t, with total shots $N^{\text{I}}_{\text{off},t}$ and counts $m^{\text{I},t}_{ij,\pm}$, we set

$$\widehat{\text{Im}(\rho_{ij})} := \frac{m^{\text{I},t}_{ij,-} - m^{\text{I},t}_{ij,+}}{2\,N^{\text{I}}_{\text{off},t}}. \tag{21}$$

Then combining the above two estimators, we have

$$\widehat{\rho}_{ij} := \frac{m^{\text{R},t}_{ij,+} - m^{\text{R},t}_{ij,-}}{2\,N^{\text{R}}_{\text{off},t}} + \mathrm{i}\,\frac{m^{\text{I},t}_{ij,-} - m^{\text{I},t}_{ij,+}}{2\,N^{\text{I}}_{\text{off},t}}\,.$$

This amounts to an unbiased estimator for ρ_{ij}.

So far we have explained the procedure for estimating the ρ_{ij}'s, but have not specified how many measurements we need to perform so that $|\rho_{ij} - \widehat{\rho}_{ij}| \le \varepsilon$ for all i,j. For this, we require a standard but highly useful concentration inequality:

Theorem 47 (Hoeffding inequality)**.** *Let $Z_1, \ldots, Z_n$ be independent random variables with $Z_k \in [a_k, b_k]$ almost surely and mean $\mu = \mathbb{E}[\frac{1}{n}\sum_{k=1}^n Z_k]$. Then for any $\eta > 0$,*

$$\Pr\left[\left|\frac{1}{n}\sum_{k=1}^{n} Z_k - \mu\right| \ge \eta\right] \le 2\,\exp\!\left(-\frac{2n^2\eta^2}{\sum_{k=1}^n (b_k - a_k)^2}\right).$$

The Hoeffding inequality says that the empirical mean of independent, bounded trials is sharply concentrated around its expectation, with a tail that decays like $\exp(-\text{const} \times n\,\eta^2)$. In our setting each measurement outcome is in $\{+1,-1\}$ or $\{0,1\}$, so a simple instance of the Hoeffding inequality applies directly to each estimated quantity.

We can use our estimators in tandem with the Hoeffding inequality to get our first bound on the sample complexity of quantum state tomography:

Theorem 48 (Basic tomography)**.** *Fix accuracy $\varepsilon \in (0,1)$ and confidence $\delta \in (0,1)$. Then with probability at least $1-\delta$, we can obtain $\widehat{\rho}_{ij}$ such that*

$$|\rho_{ij} - \widehat{\rho}_{ij}| \le \varepsilon \quad \textit{for all } i,j = 0,...,d-1\,,$$

with at most $O(\frac{d}{\varepsilon^2}\log\frac{d^2}{\delta})$ copies of ρ and as many measurements.

PROOF. We use the Hoeffding inequality stated above and a union bound over all matrix entries.

First we consider obtaining the diagonal entries of ρ. Measuring in the computational basis yields indicators $X_s^{(i)} \in \{0,1\}$ for the event "outcome i," with $\mathbb{E}[X_s^{(i)}] = \rho_{ii}$. The estimator $\widehat{\rho}_{ii} = N_i/N_{\mathrm{diag}} = \frac{1}{N_{\mathrm{diag}}}\sum_{s=1}^{N_{\mathrm{diag}}} X_s^{(i)}$ is therefore an empirical mean of $[0,1]$–bounded variables. Hoeffding's inequality gives us

$$\Pr\big[\,|\widehat{\rho}_{ii} - \rho_{ii}| \geq \eta\,\big] \leq 2\exp\big(-2N_{\mathrm{diag}}\eta^2\big).$$

Choosing, for example, $N_{\mathrm{diag}} \geq \frac{1}{2\varepsilon^2}\log\frac{2d^2}{\delta}$ ensures $\Pr[\,|\widehat{\rho}_{ii} - \rho_{ii}| \geq \varepsilon\,] \leq \delta/d^2$ for each i.

Next we turn to obtaining the off-diagonal entries of ρ. Fix a round t and a pair $\{i,j\} \in M_t$. To obtain the real part of ρ_{ij} in the R_t setting, define per–shot variables

$$X_s^{\mathrm{R},t,(ij)} = \begin{cases} +\frac{1}{2}, & \text{if the outcome is the } + \text{ projector in the } \{i,j\} \text{ block},\\ -\frac{1}{2}, & \text{if the outcome is the } - \text{ projector in the } \{i,j\} \text{ block},\\ 0, & \text{if the outcome lies in a different block,}\end{cases}$$

where $X_s^{\mathrm{R},t,(ij)} \in \big[-\frac{1}{2}, \frac{1}{2}\big]$. Then $\widehat{\mathrm{Re}(\rho_{ij})} = \frac{1}{N^{\mathrm{R}}_{\mathrm{off},t}}\sum_{s=1}^{N^{\mathrm{R}}_{\mathrm{off},t}} X_s^{\mathrm{R},t,(ij)}$ (compare with (20)) and thus

$$\mathbb{E}\Big[\widehat{\mathrm{Re}(\rho_{ij})}\Big] = \tfrac{1}{2}\Big(\Pr[+] - \Pr[-]\Big) = \tfrac{1}{2}\,\mathrm{tr}(E_{ij}^{+}\rho) = \mathrm{Re}(\rho_{ij}),$$

and so the estimator is unbiased. Hoeffding's inequality with range length 1 implies

$$\Pr\Big[\,\big|\widehat{\mathrm{Re}(\rho_{ij})} - \mathrm{Re}\,\rho_{ij}\big| \geq \eta\Big] \leq 2\exp\big(-2N^{\mathrm{R}}_{\mathrm{off},t}\eta^2\big).$$

To obtain the imaginary part of ρ_{ij} in the I_t setting, define

$$X_s^{\mathrm{I},t,(ij)} = \begin{cases} +\frac{1}{2}, & \text{if the outcome is the } - \text{ projector in the } \{i,j\} \text{ block},\\ -\frac{1}{2}, & \text{if the outcome is the } + \text{ projector in the } \{i,j\} \text{ block},\\ 0, & \text{otherwise,}\end{cases}$$

so that $\widehat{\mathrm{Im}(\rho_{ij})} = \frac{1}{N^{\mathrm{I}}_{\mathrm{off},t}}\sum_{s=1}^{N^{\mathrm{I}}_{\mathrm{off},t}} X_s^{\mathrm{I},t,(ij)}$ (compare with (21)). Using $\mathrm{tr}(E_{ij}^{-}\rho) = -2\,\mathrm{Im}\,\rho_{ij}$, we get

$$\mathbb{E}\Big[\widehat{\mathrm{Im}(\rho_{ij})}\Big] = \tfrac{1}{2}\Big(\Pr[-] - \Pr[+]\Big) = -\frac{1}{2}\,\mathrm{tr}(E_{ij}^{-}\rho) = \mathrm{Im}\,\rho_{ij},$$

which is evidently an unbiased estimator, and Hoeffding's inequality gives

$$\Pr\Big[\,\big|\widehat{\mathrm{Im}(\rho_{ij})} - \mathrm{Im}\,\rho_{ij}\big| \geq \eta\Big] \leq 2\exp\big(-2N^{\mathrm{I}}_{\mathrm{off},t}\eta^2\big).$$

Now impose $N^{\mathrm{R}}_{\mathrm{off},t} = N^{\mathrm{I}}_{\mathrm{off},t} =: N_{\mathrm{off},t}$ for all t and set $\eta = \varepsilon/\sqrt{2}$. Then for each $i \neq j$,

$$\Pr\Big[\,\big|\widehat{\mathrm{Re}(\rho_{ij})} - \mathrm{Re}\,\rho_{ij}\big| \geq \tfrac{\varepsilon}{\sqrt{2}}\,\Big] \leq 2e^{-N_{\mathrm{off},t}\varepsilon^2}, \quad \Pr\Big[\,\big|\widehat{\mathrm{Im}(\rho_{ij})} - \mathrm{Im}\,\rho_{ij}\big| \geq \tfrac{\varepsilon}{\sqrt{2}}\,\Big] \leq 2e^{-N_{\mathrm{off},t}\varepsilon^2}.$$

By a union bound,

$$\Pr[\,|\widehat{\rho}_{ij}-\rho_{ij}|\geq\varepsilon\,]\leq\Pr\Big[\,\big|\widehat{\mathrm{Re}(\rho_{ij})}-\mathrm{Re}\,\rho_{ij}\big|\geq\tfrac{\varepsilon}{\sqrt{2}}\,\Big]+\Pr\Big[\,\big|\widehat{\mathrm{Im}(\rho_{ij})}-\mathrm{Im}\,\rho_{ij}\big|\geq\tfrac{\varepsilon}{\sqrt{2}}\,\Big]$$

$$\leq\ 4\,e^{-N_{\mathrm{off},t}\varepsilon^2}.$$

Now there are d diagonal events and $2\binom{d}{2}$ off-diagonal (real or imaginary) events, giving in total d^2 events. Choosing, for example,

$$N_{\mathrm{diag}}\ \geq\ \frac{1}{2\varepsilon^2}\log\frac{2d}{\delta},\quad N_{\mathrm{off},t}\ \geq\ \frac{1}{\varepsilon^2}\log\frac{4d^2}{\delta},$$

makes each event fail with probability at most δ/d^2, whence by a union bound all hold simultaneously with probability at least $1-\delta$. This means that with probability at least $1-\delta$ we have

$$|\widehat{\rho}_{ii}-\rho_{ii}|\leq\varepsilon\quad\text{for all } i=0,\ldots,d-1\,,$$

and with probability at least $1-\delta$ we have

$$|\widehat{\rho}_{ij}-\rho_{ij}|\leq\varepsilon\quad\text{for all } \{i,j\}\in M_t\,.$$

Note that each shot consumes one fresh copy of ρ. We use one diagonal setting with N_{diag} shots and two settings per matching round (R_t and I_t) with $N^{\mathrm{R}}_{\mathrm{off},t}$ and $N^{\mathrm{I}}_{\mathrm{off},t}$ shots per each per round. With $T=d-1$ if d is even and $T=d$ if d is odd, the total is

$$N_{\mathrm{tot}}=N_{\mathrm{diag}}+\sum_{t=1}^{T}\big(N^{\mathrm{R}}_{\mathrm{off},t}+N^{\mathrm{I}}_{\mathrm{off},t}\big)=N_{\mathrm{diag}}+2\sum_{t=1}^{T}N_{\mathrm{off},t}=O\bigg(\frac{d}{\varepsilon^2}\log\frac{d^2}{\delta}\bigg)\,.$$

This equals the number of measurements performed. The theorem follows. □

There are several ways to quantify the error in approximating ρ. To articulate another way, consider the following definition:

Definition 49 (Frobenius norm)**.** *The* ***Frobenius norm*** *on* $\mathbb{C}^{d\times d}$ *is defined by*

$$\|A\|_F:=\sqrt{\langle A,A\rangle_{\mathrm{HS}}}=\sqrt{\mathrm{tr}(A^\dagger A)}\,,$$

for all $A\in\mathbb{C}^{d\times d}$*.*

Then we have the following corollary of our basic tomography theorem:

Corollary 50 (Frobenius error)**.** *Under the conditions of Theorem 48, with probability at least* $1-\delta$*,*

$$\|\rho-\widehat{\rho}\|_F\leq\sqrt{\sum_{i,j}|\widehat{\rho}_{ij}-\rho_{ij}|^2}\leq\sqrt{d^2\,\varepsilon^2}=d\,\varepsilon.$$

Above, if we choose $\varepsilon=\varepsilon'/d$, then we can get $\|\rho-\widehat{\rho}\|_F\leq\varepsilon'$ with probability at least $1-\delta$. But then this will require $O(\frac{d^3}{\varepsilon'^2}\log\frac{d^2}{\delta})$ copies of ρ and as many measurements.

Remark 51 (Projection to the density-matrix cone preserves and possibly improves Frobenius error)**.** *The empirical matrix* $\widehat{\rho}$ *constructed entrywise need not be positive semidefinite nor have unit trace. Let* $\mathcal{D}:=\{X\succeq 0:\mathrm{tr}(X)=1\}$ *denote the set of* *density matrices, and let $\Pi_{\mathcal{D}}$ be the Euclidean (Frobenius) projection onto $\mathcal{D}$. Since $\mathcal{D}$ is closed and convex, $\Pi_{\mathcal{D}}$ is nonexpansive:*

$$\|\Pi_{\mathcal{D}}(A) - \Pi_{\mathcal{D}}(B)\|_F \leq \|A - B\|_F \quad \text{for all } A, B.$$

In particular, because $\rho \in \mathcal{D}$,

$$\|\rho - \Pi_{\mathcal{D}}(\widehat{\rho})\|_F \leq \|\rho - \widehat{\rho}\|_F.$$

Thus the Frobenius bound from Corollary 50 continues to hold (and can only improve) after projecting $\widehat{\rho}$ onto $\mathcal{D}$. Operationally, if $\widehat{\rho} = U\mathrm{diag}(\lambda)U^\dagger$ is an eigendecomposition, then $\Pi_{\mathcal{D}}(\widehat{\rho})$ is obtained by projecting the eigenvalue vector λ onto the probability simplex $\{\mu \geq 0 : \sum_i \mu_i = 1\}$ (via the usual simplex projection) and setting $\widetilde{\rho} = U\mathrm{diag}(\mu)U^\dagger$.

Before moving on to fancier versions of quantum state tomography, we comment here on how practical it is to perform it. Consider a system of 8 qubits, corresponding to $d = 2^8 = 256$; therefore a ρ is described by $65,535$ real numbers. Such a quantum state tomography (with a slightly different method than the one presented above) was performed in [**HHR**$^+$**05**] with about 10 hours of data acquisition. Note that each individual qubit increases the number of parameters and data acquisition time exponentially. As such, full quantum state tomography is often totally impractical even for modest system sizes.

Nonetheless, we study full quantum state tomography since it is a fundamental problem in quantum learning theory, and allows us to build tools for more pragmatic, 'partial' forms of state tomography which are highly practical and often used.

2. Learning a state in the operator norm

In many applications we do not need to reconstruct ρ entrywise; rather, we want to *predict expectation values* of observables with respect to ρ. If $\widehat{\rho}$ is an estimate, the prediction error for an observable O is

$$\big|\langle O\rangle_\rho - \langle O\rangle_{\widehat{\rho}}\big| = \big|\mathrm{tr}\big(O(\rho - \widehat{\rho})\big)\big|.$$

Different matrix norms control this quantity for different classes of O. We next introduce the operator norm and explain when it is the right notion of accuracy.

Definition 52 (Operator (spectral) norm)**.** *The* ***operator norm*** *of a matrix $A \in \mathbb{C}^{d\times d}$ is*

$$\|A\|_{\mathrm{op}} := \sup_{\|v\|_2=1} \|Av\|_2 = \sigma_{\max}(A)\,,$$

which is the largest singular value of A. If A is Hermitian, then $\|A\|_{\mathrm{op}} = \max_k |\lambda_k(A)|$, the largest eigenvalue magnitude. We will sometimes also write $\|\cdot\|_\infty$ for $\|\cdot\|_{\mathrm{op}}$.

Two elementary inequalities will be useful: the Hilbert–Schmidt Cauchy–Schwarz bound

$$|\mathrm{tr}(X^\dagger Y)| \leq \|X\|_{\mathrm{F}}\|Y\|_{\mathrm{F}}\,,$$

and the trace–operator Hölder bound (duality of $\|\cdot\|_1$ and $\|\cdot\|_\infty$)

$$|\mathrm{tr}(XY)| \leq \|X\|_1\|Y\|_\infty\,.$$

Using these inequalities we have the following: