Unentangled spectrum estimation

Faster than full-state learning

Angelos Pelecanos
UC Berkeley

Norah Tan
MIT

Ewin Tang
UC Berkeley

John Wright
UC Berkeley

1. Spectrum estimation

Spectrum estimation

Problem definition

Given a d-dimensional quantum state ρ.

Model $\rho \in \mathbb{C}^{d \times d}$ as a density matrix: Hermitian, PSD, and has trace 1.

Eigenvalues (spectrum) $\alpha_1 \geq \ldots \geq \alpha_d \geq 0$ sum to $1 \Longrightarrow$ form a probability distribution.

Goal: Estimate the spectrum of ρ to total variation distance ε.

Output $\widehat{\alpha}_1 \geq \ldots \geq \widehat{\alpha}_d$ such that $d_{\mathrm{TV}}(\widehat{\alpha}, \alpha) \leq \varepsilon$ with constant probability.

Spectrum estimation

Problem definition: Input model and measurements

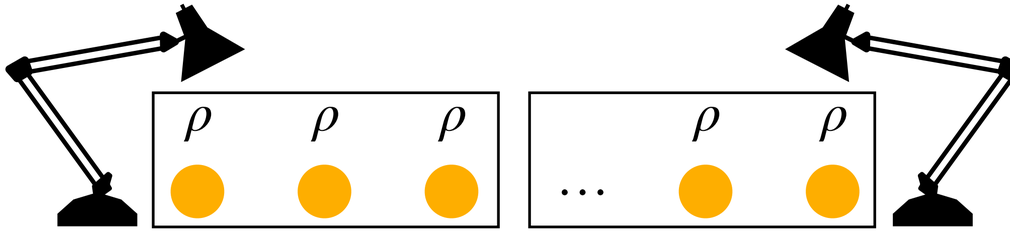
Goal: Estimate the spectrum of ρ to total variation distance ε.

Output $\hat{\alpha}_1 \geq \ldots \geq \hat{\alpha}_d$ such that $d_{\mathrm{TV}}(\hat{\alpha}, \alpha) \leq \varepsilon$ with constant probability.

Given n copies of the state ρ.

Unentangled measurements: Measure each copy of ρ separately.

Entangled measurements: Measure multiple copies at once.

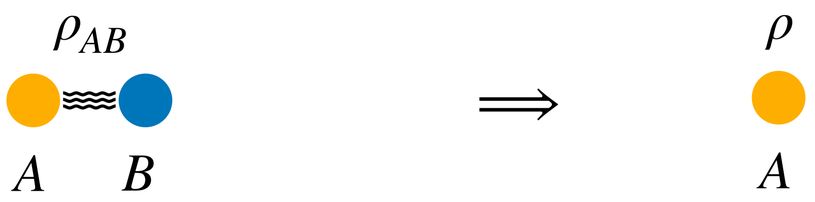


Spectrum estimation

Why estimate the spectrum?

Encodes fundamental properties of the state ρ, such as:

- Purity of the state.
- Information about entanglement, e.g. when ρ is the first part of bipartite state.



Spectrum learning is tightly related (and often necessary) for state learning algorithms.

Spectrum estimation

Prior work: sample complexity

Upper bound using full-state learning.

- Entangled: $n = O(d^2/\varepsilon^2)$ [OW16, HHJ+16]
- Unentangled: $n = O(d^3/\varepsilon^2)$ [KRT14]

Lower bound via uniformity testing.

- Entangled: $n = \Omega(d/\varepsilon^2)$ [OW15]
- Unentangled: $n = \Omega(d^{1.5}/\varepsilon^2)$ [CHLL22]

Question. Is spectrum estimation **easier**, or **as hard as** full-state learning (tomography)?

Spectrum estimation

Main result: unentangled spectrum estimation easier than state learning

Theorem. There is an algorithm that solves spectrum estimation using

$$n = O\left(\frac{d^3}{\varepsilon^6} \cdot \left(\frac{\log\log d}{\log d}\right)^4\right)$$

samples and unentangled measurements.

Remark. Recall unentangled state learning uses $n = \Theta(d^3/\varepsilon^2)$ copies.

Better than unentangled state learning for ε "not too small".

2. Classical: Learning sorted distribution

Classical: Learning sorted distribution

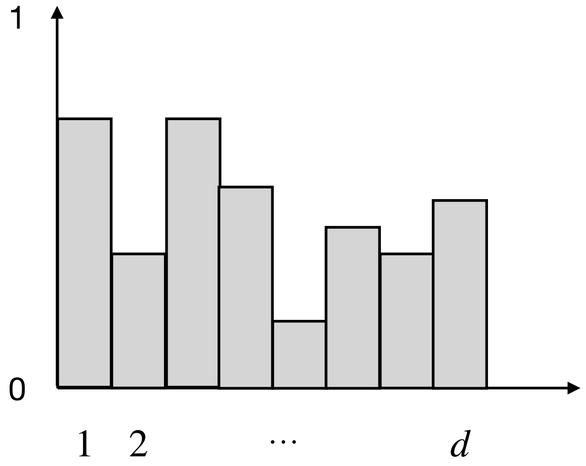
Symmetric properties

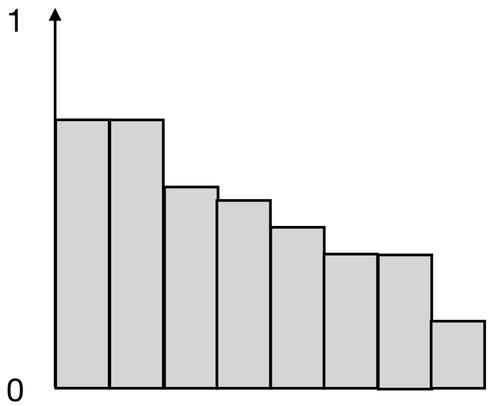
Classically: spectrum estimation becomes learning the sorted distribution.

Given samples from a probability distribution $\beta = (\beta_1, \ldots, \beta_d)$.

Goal: Estimate its sorted histogram to total variation distance ε.

Output $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_d) \in \mathbb{R}^d$ such that $d_{\mathrm{TV}}(\widehat{\beta}, \mathrm{sort}(\beta)) \leq \varepsilon$ with constant probability.





Classical: Learning sorted distribution

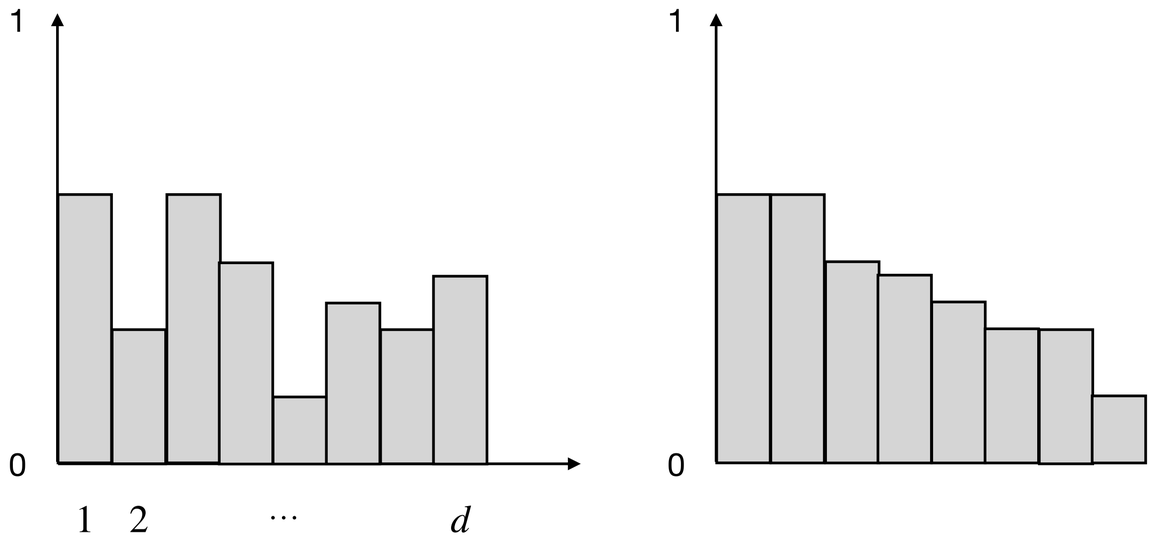
Why estimate the sorted histogram?

Goal: Estimate its sorted histogram to total variation distance ε.

Output $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_d) \in \mathbb{R}^d$ such that $d_{\mathrm{TV}}(\widehat{\beta}, \mathrm{sort}(\beta)) \leq \varepsilon$ with constant probability.

Similar to the quantum case, the sorted histogram encodes fundamental properties of β.

- Support size
- Entropy



Classical: Learning sorted distribution

Prior work

Goal: Estimate sorted histogram of β to total variation distance ε.

Idea 1. Compute estimate of distribution $\widehat{\beta}$ using $n = O(d/\varepsilon^2)$ samples and then sort.

[VV11]. Valiant and Valiant showed that $n = O(d/\log d)$ samples suffice for ε constant.

[HJW18]. Han, Jiao, Weissman obtained the optimal sample complexity.

- Use $O(d/\varepsilon^2)$ samples for $\varepsilon \leq 1/d$ and $n = O(d/(\log d \cdot \varepsilon^2))$ otherwise.
- Introduced local moment matching—what we see next.

Classical: Learning sorted distribution

Idea 1: Moment matching

The k-th moment of a distribution is $p_k(\beta) = \sum \beta_i^k$.

- Captures the probability of k samples being equal (k-way collision).

Observation: The moments $p_1, \ldots, p_d$, determine the sorted distribution $\mathrm{sort}(\beta)$.

- Suffices to know the moments of β.

Moment matching: Use samples from β to estimate $p_k(\beta)$ and determine $\mathrm{sort}(\beta)$.

Classical: Learning sorted distribution

Idea 1: Moment matching

Given samples $x_1, \ldots, x_n$ from β, estimate k-th moment using the estimator:

$$\widehat{p}_k(x) = \binom{n}{k}^{-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \mathbf{1}[x_{i_1} = \ldots = x_{i_k}]$$

Counts the empirical number of k-wise collisions.

This estimator has great variance.

Unbiased estimator for $p_k(\beta) = \sum_i \beta_i^k$.

$$\mathbb{E}[\widehat{p}_k(x)]$$

$$= \binom{n}{k}^{-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \mathbb{E}[\mathbf{1}[x_{i_1} = \ldots = x_{i_k}]]$$

$$= \binom{n}{k}^{-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \Pr[x_{i_1} = \ldots = x_{i_k}]$$

$$= \binom{n}{k}^{-1} \sum_{1 \le i_1 < \ldots < i_k \le n} p_k(\beta) = p_k(\beta)$$

Classical: Learning sorted distribution

Idea 1: Moment matching

Given samples $x_1, \ldots, x_n$ from β, estimate k-th moment using the estimator:

$$\widehat{p}_k(x) = \binom{n}{k}^{-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \mathbf{1}[x_{i_1} = \ldots = x_{i_k}]$$

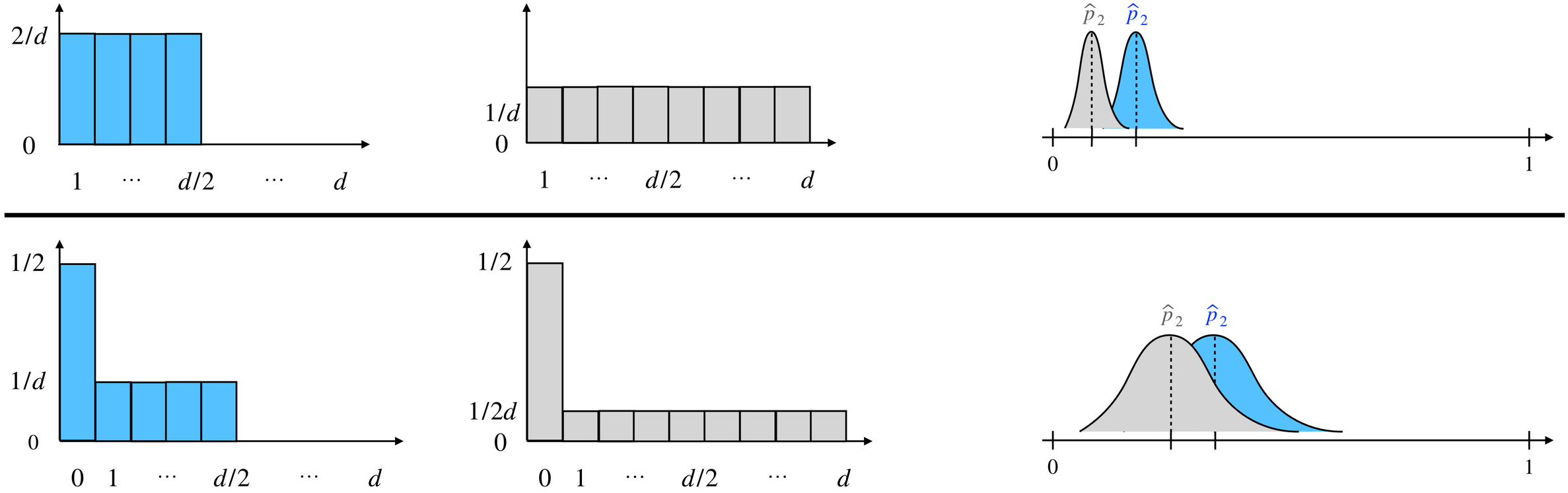
Moment matching algorithm:

1. Estimate $\widehat{p}_1(x), \ldots, \widehat{p}_d(x)$ using samples.
2. Find the sorted histogram $\widehat{\beta}$ such that $p_k(\widehat{\beta}) \approx \widehat{p}_k(x)$.

Classical: Learning sorted distribution

Idea 1: Moment matching — Issue

Issue: The large items hide the moment contribution of the small items.



Classical: Learning sorted distribution

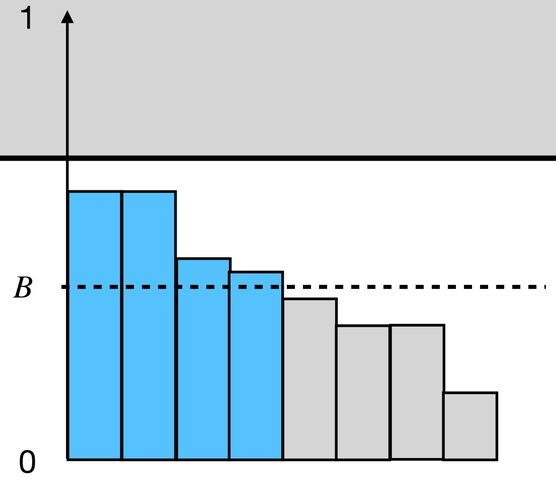
Idea 2: Local moment matching

Solution: Easy to distinguish large and small elements and deal with them separately.

Local moment matching algorithm:

1. Use $O(n)$ samples to learn which items are **Large** and **Small**.
2. Use $O(n)$ samples to obtain $\widehat{\beta}_{\text{Large}}$ and $\widehat{\beta}_{\text{Small}}$ by moment matching on **Large** and **Small**.
3. Stitch the two estimators and output $(\widehat{\beta}_{\text{Large}}, \widehat{\beta}_{\text{Small}})$.

Large contains probabilities at least B.



Note. [HJW18] use more than two buckets for the optimal bound.

Two buckets suffice to do better than $O(d)$.

3. Making it quantum

Making it quantum

Local moment matching blueprint

We will perform local moment matching in the quantum case.

Steps

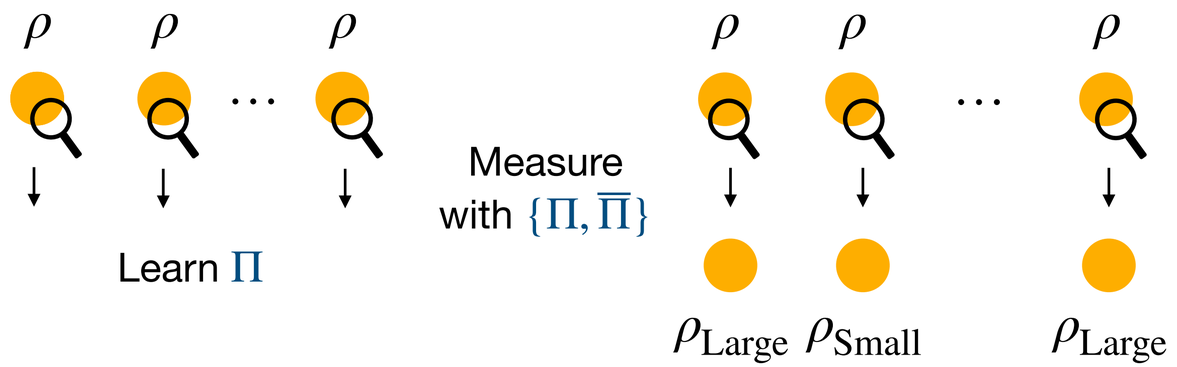
1. (Bucketing) Use $O(n)$ samples to divide the eigenvalues of ρ into **Large** and **Small**.

2. (Moment estimation) Use $O(n)$ samples to estimate the moments of **Large** and **Small**.

3. (Local moment matching) Convert moments estimates to eigenvalues and output together.

Making it quantum

Step 1: Bucketing



Goal. Divide the eigenvalues of ρ into **Large** and **Small**.

Given copies of state ρ, we want projector Π onto the **Large** eigenvalues.

Measure remaining copies of ρ using $\{\Pi, \overline{\Pi}\}$ with results $\rho_{\text{Large}} = \Pi\rho\Pi$ and $\rho_{\text{Small}} = \overline{\Pi}\rho\overline{\Pi}$.

Properties:

1. (Classification) $\Pi\rho\Pi$ must contain the **Large** eigenvalues and $\overline{\Pi}\rho\overline{\Pi}$ the **Small**.
2. (Alignment) The spectrum of $\Pi\rho\Pi + \overline{\Pi}\rho\overline{\Pi}$ must be ε-close to the spectrum of ρ.

The alignment property is quantum-specific!

Making it quantum

Step 1: Bucketing

Large contains eigenvalues at least B.

Properties:

1. (Classification) $\Pi\rho\Pi$ must contain the **Large** eigenvalues and $\overline{\Pi}\rho\overline{\Pi}$ the **Small**.
2. (Alignment) The spectrum of $\Pi\rho\Pi + \overline{\Pi}\rho\overline{\Pi}$ must be ε-close to the spectrum of ρ.

How to construct Π:

1. Run state learning on ρ with $o(d^3/\varepsilon^2)$ samples and obtain estimate $\widehat{\rho}$.
2. Set Π as the projector on the large eigenvalues of $\widehat{\rho}$.

Which state learning algorithm gives the desired properties?

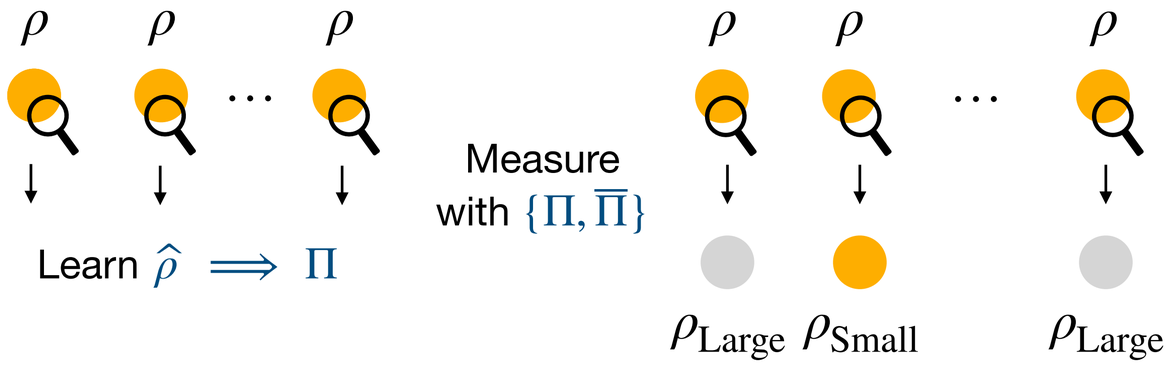
Making it quantum

Step 1: Bucketing

For classification and alignment, it suffices to use an estimator such that $\|\rho - \widehat{\rho}\|_\infty \leq \varepsilon$.

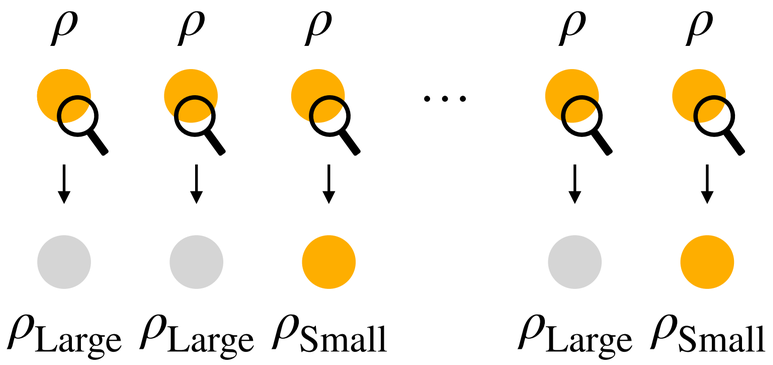
The unentangled algorithm of [GKKT18] provides this guarantee:

- Other algorithms also work — [GKKT18] is the simplest that works.
- No entangled tomography algorithms are known (yet) to have this guarantee.
- **Bonus:** Also learns the **Large** eigenvalues $\widehat{\alpha}_{\text{Large}}$ for free!



Making it quantum

Step 2: Moment estimation



After bucketing: Estimated $\widehat{\alpha}_{\text{Large}}$ and have samples of $\rho_{\text{Small}} = \overline{\Pi}\rho\overline{\Pi}$.

Want to estimate $\mathrm{tr}(\sigma^k) = \sum \alpha_i^k \cdot \mathbf{1}[i \in \text{Small}]$.

Let $\sigma = \rho_{\text{Small}}$

1. Let $\mathscr{A}$ be unentangled unbiased state learning algorithm.
2. Given copies of σ, use $\mathscr{A}$ to get $\widehat{\sigma}_1, \ldots, \widehat{\sigma}_n$ with $\mathbb{E}\,\widehat{\sigma}_i = \sigma$.
3. Estimate the k-th moment to be:

$$Z_k = N^{-1} \sum_{\text{distinct } i_1,\ldots,i_k} \mathrm{tr}(\widehat{\sigma}_{i_1} \ldots \widehat{\sigma}_{i_k})$$

Generalization of the k-wise collision estimator

$$\begin{aligned}
&\mathbb{E}[Z_k] \\
&= N^{-1} \sum_{\text{distinct } i_1,\ldots,i_k} \mathrm{tr}(\mathbb{E}[\widehat{\sigma}_{i_1} \ldots \widehat{\sigma}_{i_k}]) \\
&= N^{-1} \sum_{\text{distinct } i_1,\ldots,i_k} \mathrm{tr}(\mathbb{E}[\widehat{\sigma}_{i_1}] \ldots \mathbb{E}[\widehat{\sigma}_{i_k}]) \\
&= N^{-1} \sum_{\text{distinct } i_1,\ldots,i_k} \mathrm{tr}(\sigma \cdots \sigma) \\
&= N^{-1} \sum_{\text{distinct } i_1,\ldots,i_k} \mathrm{tr}(\sigma^k) = \mathrm{tr}(\sigma^k)
\end{aligned}$$

Making it quantum

Step 3: Local moment matching

After bucketing: Estimate $\widehat{\alpha}_{\text{Large}}$.

After moment estimation: Moment estimates $Z_1, Z_2, \ldots, Z_K$ of ρ_{Small} up to the K-th moment.

Given the first K moment estimates, local moment matching outputs the estimate $\widehat{\alpha}_{\text{Small}}$.

- Construct a linear program and [HJW18] shows how to round its solution and bound the error.
- The error of $\widehat{\alpha}_{\text{Small}}$ scales with the number of moments K and the accuracy of the moments.

The same approach works in the quantum case.

Remark: Depending on the error, we do not always need d moments.

Making it quantum

Putting it all together

We have seen all the ingredients of local moments matching in the quantum setting.

1. (Bucketing) Use $O(n)$ samples to divide the eigenvalues of ρ into **Large** and **Small**.

2. (Moment estimation) Use $O(n)$ samples to estimate the moments of **Large** and **Small**.

3. (Local moment matching) Convert moments estimates to eigenvalues and output together.

Theorem. There is an algorithm that solves spectrum estimation using

$$n = O\left(\frac{d^3}{\varepsilon^6} \cdot \left(\frac{\log\log d}{\log d}\right)^4\right)$$

samples and unentangled measurements.

4. Additional results

Additional results

Result 2: Unentangled moment estimation

Key ingredient: Estimate $\mathrm{tr}(\sigma^k)$ with multiplicative error using unentangled measurements.

$$(1-\delta)\cdot \mathrm{tr}(\sigma^k) \leq Z_k \leq (1+\delta)\cdot \mathrm{tr}(\sigma^k)$$

The multiplicative error implies an additive error approximation of quantum Renyi entropy.

$$S_k(\sigma) = \frac{1}{1-k}\log \mathrm{tr}(\sigma^k)$$

The classical Renyi entropy of the spectrum of σ.

Theorem. For any $k \geq 2$, exists an algorithm that estimates $S_k(\sigma)$ to additive error δ using

$$n = O\left(\max\left\{\frac{d^{2-2/k}}{\delta^2}, \frac{d^{3-2/k}}{\delta^{2/k}}\right\}\right)$$

samples and unentangled measurements.

Related to the Renyi entanglement entropy – information about the entanglement of states.

Estimated in experimental settings [IMP+15,KTL+16].

Additional results

Result 3: Numerical lower bound for entangled measurements

We also study the entangled measurement setting for constant ε.

Full-state learning: $n = O(d^2/\varepsilon^2)$

Uniformity testing: $n = \Omega(d/\varepsilon^2)$

Provide numerical evidence that the $\Omega(d)$ lower bound is not tight.

Our evidence suggests a $\Omega(d^{2-\gamma})$ for all $\gamma > 0$ lower bound.

Very close to what the classical bound suggests.

See the paper for more details.

5. Discussion

Discussion

Spectrum estimation algorithm better than full-state learning when ε is not too small.

Similar to the classical case, the improvement is only polylogarithmic.

Theorem. There is an algorithm that solves spectrum estimation using

$$n = O\left(\frac{d^3}{\varepsilon^6} \cdot \left(\frac{\log\log d}{\log d}\right)^4\right)$$

samples and unentangled measurements.

It would be surprising if the current bound is optimal:

- The ε^6 dependence seems too steep.
- The $\log\log(d)$ factors also look suspicious.

Also surprising if local moment matching is not optimal quantumly!

Future directions

Improved algorithm in the unentangled setting.

Faster-than-tomography spectrum estimation using entangled measurements.

Lower bounds for spectrum estimation.

Moment estimation using unentangled measurements.

- Lower bounds or improved upper bounds.

Thank you

Questions?