Beating full state tomography for unentangled spectrum estimation

Angelos Pelecanos
UC Berkeley

Xinyu (Norah) Tan
MIT

Ewin Tang
UC Berkeley

John Wright
UC Berkeley

MIT Quantum CS Seminar
April 2, 2025

Two fundamental quantum learning theoretic tasks

Given n samples/copies of an unknown d-dim mixed state ρ with spectrum $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)$ (sorted eigenvalues)

State tomography	Spectrum estimation
• obtain an estimate $\hat{\boldsymbol{\rho}}$ such that $\|\rho - \hat{\boldsymbol{\rho}}\|_1 \leq \epsilon$	• obtain an estimate $\hat{\boldsymbol{\alpha}}$ such that $d_{\mathrm{TV}}(\alpha, \hat{\boldsymbol{\alpha}}) \leq \epsilon$

Why do we care about the spectrum?

- Unitarily invariant properties:
  - Purity of a state $\mathrm{tr}(\rho^2)$
  - Shannon and Rényi Entropy
  - Entanglement entropy of a bipartite pure state
  - …

Sample complexity

How big can n be? $g(d,\epsilon) \leq n \leq f(d,\epsilon)$

- **Upper bound**: give an algorithm that uses $n = f(d,\epsilon)$ samples
  - Quantum circuits + measurements
  - Classical post-processing using the measurement outcomes

- **Lower bound**: prove that no algorithm can learn the task with fewer than $n = g(d,\epsilon)$ samples

Measurements

- Fully entangled measurements
- Unentangled (single-copy) measurements
  - Adaptive vs non-adaptive algorithms
  - **Uniform POVM**: $\{d|u\rangle\langle u| \cdot \mathrm{d}u\}$ where $\mathrm{d}u$ is the Haar measure over pure states $|u\rangle \in \mathbb{C}_d$

$$d \cdot \mathbb{E}_{|u\rangle\sim\mathrm{Haar}}|u\rangle\langle u| = I$$

Uniform POVM tomography algorithm

- Measure each copy of ρ with the uniform POVM $\{d|u\rangle\langle u| \cdot \mathrm{d}u\}$
- Set $\rho_i = (d+1)\cdot|u_i\rangle\langle u_i| - I$ where $|u_i\rangle$ is the i-th measurement outcome
- Output $\widehat{\boldsymbol{\rho}} = \frac{1}{n}(\rho_1 + \cdots + \rho_n)$

Theorem [Krishnamurthy and Wright, GKKT20]: $\|\rho - \widehat{\boldsymbol{\rho}}\|_1 \le \epsilon$ if $n = O\left(\frac{d^3}{\epsilon^2}\right)$.

What's known so far

- We can always use a state tomography algorithm to learn its spectrum:
  - $\|\rho - \hat{\boldsymbol{\rho}}\|_1 \leq \epsilon \Rightarrow d_{\mathrm{TV}}\big(\alpha, \hat{\boldsymbol{\alpha}} = \text{eigenvalues}(\hat{\boldsymbol{\rho}})\big) \leq \epsilon$
- Can we do better than full state tomography?

	Unentangled	Fully entangled
State tomography	$\Theta\left(\frac{d^3}{\epsilon^2}\right)$ [KRT14,CHL+23]	$\Theta\left(\frac{d^2}{\epsilon^2}\right)$ [OW15,16]
Spectrum estimation	$\Omega\left(\frac{d^{3/2}}{\epsilon^2}\right)$ [CHLL22]	$\Omega\left(\frac{d}{\epsilon^2}\right)$ [OW15]

Mixedness testing: Test if a state is maximally mixed, i.e. with spectrum $\left(\frac{1}{d}, \cdots, \frac{1}{d}\right)$, or ϵ-far from I_d/d

Main results on spectrum estimation

	Unentangled	**Fully entangled**
State tomography	$\Theta\left(\frac{d^3}{\epsilon^2}\right)$	$\Theta\left(\frac{d^2}{\epsilon^2}\right)$
Spectrum estimation (previously known)	$\Omega\left(\frac{d^{3/2}}{\epsilon^2}\right)$	$\Omega\left(\frac{d}{\epsilon^2}\right)$
Spectrum estimation **(our results)**	$O\left(d^3 \cdot \left(\frac{\log\log(d)}{\log(d)}\right)^4 \cdot \frac{1}{\epsilon^6}\right)$	$\Omega(d^{2-\gamma})$ for any constant $\gamma > 0$ (we give numerical evidence for constant ϵ)

An adaptive algorithm that uses asymptotically fewer samples in large ϵ-regime

Spectrum learning can only improve over full state tomography by a sub-polynomial factor

Next

- Part I: Motivation and main results
- **Part II: The classical analogue**
  - The idea of (global) moment matching and why it fails
  - Local moment matching
- Part III: The quantum case

The classical analogues

- The spectrum α of ρ can be viewed as **a probability distribution**
- Let $\alpha = (\alpha_1, \ldots, \alpha_d)$ be an unknown (unsorted) distribution
  - Item i is sampled with probability α_i

Learn the distribution α	Learn the **sorted** distribution $\alpha^{\geq} = \mathrm{sort}(\alpha)$
• obtain an estimate $\widehat{\boldsymbol{\alpha}}$ such that $d_{\mathrm{TV}}(\alpha, \widehat{\boldsymbol{\alpha}}) \leq \epsilon$ • State tomography	• obtain an estimate $\widehat{\boldsymbol{\alpha}}^{\geq}$ such that $d_{\mathrm{TV}}(\alpha^{\geq}, \widehat{\boldsymbol{\alpha}}^{\geq}) \leq \epsilon$ • Spectrum estimation

Idea: Matching moments

- Let $\alpha = (\alpha_1, \ldots, \alpha_d)$ be an unknown (unsorted) distribution.

- The k-th moment of α: $p_k(\alpha) := \sum_{i=1}^{d} \alpha_i^k$

Moment matching

1. Obtain estimates $\widehat{\boldsymbol{p}}_{\boldsymbol{k}}$ for the first K moments:

$$\widehat{\boldsymbol{p}}_{\boldsymbol{k}} \approx p_k(\alpha), \qquad \text{for all } k \in [K]$$

2. Solve for a distribution $\widehat{\boldsymbol{\alpha}}$ with matching moments:

$$p_k(\widehat{\boldsymbol{\alpha}}) \approx \widehat{\boldsymbol{p}}_{\boldsymbol{k}}, \qquad \text{for all } k \in [K]$$

- Formally: a feasibility linear program

Why global moment matching fails

- Uniformity testing: distinguish $\alpha = \left(\frac{1}{d}, \ldots, \frac{1}{d}\right)$ or $\beta = \left(\frac{1}{2d}, \ldots, \frac{1}{2d}, 0, \ldots, 0\right)$
  - Given n samples $\{x_i\}_{i\in[n]}$ drawn from γ, output either $\gamma = \alpha$ or $\gamma = \beta$
- Standard algorithm: estimate $p_2(\gamma)$
- Standard unbiased estimator $c_2 := \frac{1}{\binom{n}{2}} \sum_{i<j} \mathbf{1}[x_i = x_j]$ (collision statistics)
- For both $\gamma = \alpha$ and $\gamma = \beta$:
  - $p_2(\gamma) = \Theta\left(\frac{1}{d}\right)$
  - $|p_2(\alpha) - p_2(\beta)| = \Theta\left(\frac{1}{d}\right)$
  - $\mathrm{Var}(c_2) = \Theta\left(\frac{1}{n^2 d}\right)$ is small
- Hence $n = O(\sqrt{d})$

New Task:

- $\alpha = \left(\frac{1}{2}, \frac{1}{2d}, \ldots, \frac{1}{2d}\right)$ vs $\beta = \left(\frac{1}{2}, \frac{1}{d}, \ldots, \frac{1}{d}, 0, \ldots, 0\right)$
- $p_2(\gamma) = \Theta(1)$ and $\mathrm{Var}(c_2) = \Theta\left(\frac{1}{n}\right)$ is much bigger.
- $|p_2(\alpha) - p_2(\beta)| = \Theta\left(\frac{1}{d}\right)$
- Hence $n = O(d^2)$

Why global moment matching fails

- The moments are dominated by large values, and they tend to "wash out" the small values.
- But capturing low-probability elements is important for estimating in TV distance.

- Alternatively, use $O(1)$ samples to learn the index $i \in [d]$ for which $\gamma_i = \frac{1}{2}$.
- Then, use $O(\sqrt{d})$ samples to test if γ is uniform on $d/\{i\}$.

New Task:

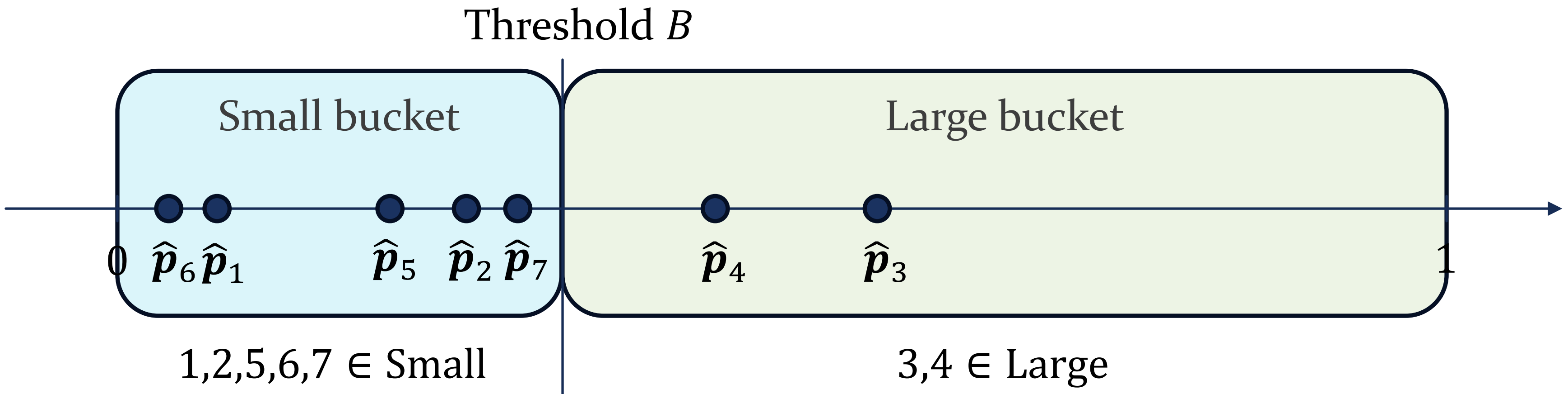
- $\alpha = \left(\frac{1}{2}, \frac{1}{2d}, \ldots, \frac{1}{2d}\right)$ vs $\beta = \left(\frac{1}{2}, \frac{1}{d}, \ldots, \frac{1}{d}, 0, \ldots, 0\right)$
- $p_2(\gamma) = \Theta(1)$ and $\mathrm{Var}(c_2) = \Theta\left(\frac{1}{n}\right)$ is much bigger.
- $|p_2(\alpha) - p_2(\beta)| = \Theta\left(\frac{1}{d}\right)$
- Hence $n = O(d^2)$

Solution: local moment matching [HJW18]

- **Step 1:** Bucketing
  - Draw n samples according to α. Denote the empirical distribution as $\widehat{\boldsymbol{p}}$.
  - Assign items to two buckets based on $\hat{p}$:

$$\text{Small} \coloneqq \{i \in [d]: \widehat{\boldsymbol{p}}_i \in [0, B)\} \quad \text{and} \quad \text{Large} \coloneqq \{i \in [d]: \widehat{\boldsymbol{p}}_i \in [B, 1]\}$$

  - Directly output $\widehat{\boldsymbol{p}}_i$ for $i \in$ Large



Requirements

The small eigenvalues are classified into the small bucket:

$$\text{Small} = \{i \in [d]: p_i \in [0, (1+\epsilon)B)\}$$

The large eigenvalues are accurately learned:

$$\sum_{i \in \text{Large}} |p_i - \widehat{\boldsymbol{p}}_i| \le \epsilon$$

Solution: local moment matching [HJW18]

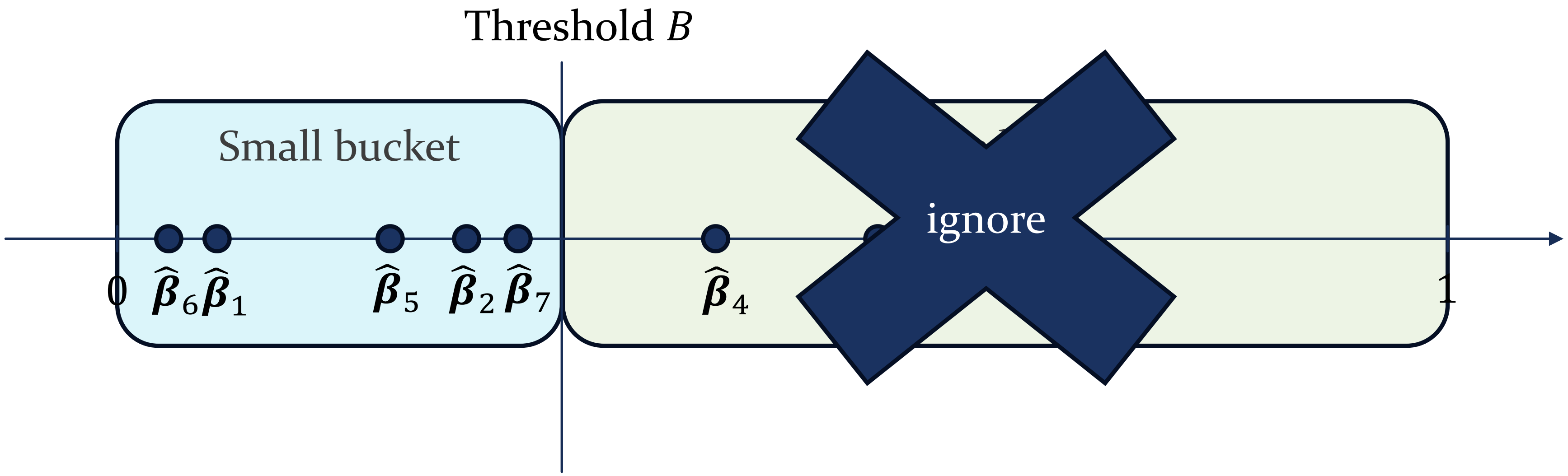
- **Step 2:** local moment matching on the small bucket
  - Draw another n samples according to α. Denote the empirical distribution as $\widehat{\boldsymbol{\beta}}$.
  - Use $\{\widehat{\boldsymbol{\beta}}_i : i \in \text{Small}\}$ to obtain moment estimates

$$\widehat{\boldsymbol{p}}_k \approx \sum_{i \in \text{Small}} \alpha_i^k \text{ on the small bucket}$$

  - Find a (subnormalized) distribution on $[0, B)$ that matches the moment estimates $\widehat{\boldsymbol{p}}_k$

Overall, after setting K and B appropriately, the sample complexity:

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



Next

- Part I: Motivation and main results
- Part II: The classical analogue
- **Part III: The quantum case**
  - Bucketing
  - Multiplicative-error moment estimation

Step 1: Bucketing

1. Use n copies of ρ to learn a projective measurement $\{\Pi, \overline{\Pi} = I - \Pi\}$
   - Π ($\overline{\Pi}$) projects onto the large (small) eigenvalues of ρ
2. Measure <u>another</u> n copies with $\{\Pi, \overline{\Pi}\}$
   - Receiving Π ($\overline{\Pi}$) outcome is as if we are sampling from the large (small) part of spectrum α
   - i.e. $\rho \mapsto \Pi\rho\Pi + \overline{\Pi}\rho\overline{\Pi}$

Uniform POVM tomography algorithm

- Measure each copy of ρ with the uniform POVM $\{d|u\rangle\langle u| \cdot \mathrm{d}u\}$
- Set $\rho_i = (d+1) \cdot |u_i\rangle\langle u_i| - I$ where $|u_i\rangle$ is the i-th measurement outcome
- Output $\widehat{\boldsymbol{\rho}} = \frac{1}{n}(\rho_1 + \cdots + \rho_n)$

Bucketing algorithm

- Perform the uniform POVM tomography algorithm using n copies of ρ and obtain $\widehat{\boldsymbol{\rho}}$
- Write $\widehat{\boldsymbol{\rho}} = U \cdot \hat{\alpha} \cdot U^\dagger$
  - $\hat{\alpha}_1 \geq \cdots \geq \hat{\alpha}_k \geq B > \hat{\alpha}_{k+1} \geq \cdots \geq \hat{\alpha}_d$
- Output $\Pi = U \cdot (|1\rangle\langle 1| + \cdots + |k\rangle\langle k|) \cdot U^\dagger$

Step 1: Bucketing

1. Use n copies of ρ to learn a projective measurement $\{\Pi, \overline{\Pi} = I - \Pi\}$
   - Π ($\overline{\Pi}$) projects onto the large (small) eigenvalues of ρ

The large eigenvalues are accurately learned

- $d_{\mathrm{TV}}(\mathrm{spec}(\rho)_{\leq k}, \mathrm{spec}(\hat{\rho})_{\leq k}) \leq \epsilon$

The small eigenvalues are classified into the small bucket

- $\|\overline{\Pi}\rho\overline{\Pi}\|_{\mathrm{op}} \leq (1+\epsilon)B$

The full spectrum is not much disturbed

- $d_{\mathrm{TV}}(\mathrm{spec}(\rho), \mathrm{spec}(\Pi\rho\Pi + \overline{\Pi}\rho\overline{\Pi})) \leq \epsilon$

Eventually we set $B = O\left(\frac{\epsilon^2}{d} \cdot \left(\frac{\log(d)}{\mathrm{loglog}(d)}\right)^2\right)$

Bucketing algorithm

- Perform the uniform POVM tomography algorithm using n copies of ρ and obtain $\hat{\boldsymbol{\rho}}$
- Write $\hat{\boldsymbol{\rho}} = U \cdot \hat{\alpha} \cdot U^\dagger$
  - $\hat{\alpha}_1 \geq \cdots \geq \hat{\alpha}_k \geq B > \hat{\alpha}_{k+1} \geq \cdots \geq \hat{\alpha}_d$
- Output $\Pi = U \cdot (|0\rangle\langle 0| + \cdots + |k\rangle\langle k|) \cdot U^\dagger$

Theorem: The bucketing algorithm satisfies the green requirements when

$$n = O(dB^{-2}\epsilon^{-2})$$

Step 2: local moment matching on the small bucket

1. Use n copies of ρ to learn a projective measurement $\{\Pi, \overline{\Pi} = I - \Pi\}$
   - Π ($\overline{\Pi}$) projects onto the large (small) eigenvalues of ρ
2. Measure <u>another</u> n copies with $\{\Pi, \overline{\Pi}\}$
   - Receiving Π ($\overline{\Pi}$) outcome is as if we are sampling from the large (small) part of spectrum α
   - i.e. $\rho \mapsto \Pi\rho\Pi + \overline{\Pi}\rho\overline{\Pi}$ $=: \sigma$ (a subnormalized quantum state)
   - Bucketing $\Rightarrow \|\sigma\|_{\mathrm{op}} \leq (1+\epsilon)B$
3. Use copies of σ to obtain "good" estimates for the moments $\mathrm{tr}(\sigma^k)$ for $k \in [K]$
4. Perform local moment matching using the moment estimates

Side results on moment estimation

k-th moment estimator

- Measure each copy of ρ with the uniform POVM $\{d|u\rangle\langle u| \cdot \mathrm{d}u\}$
- Set $\rho_i = (d+1) \cdot |u_i\rangle\langle u_i| - I$ where $|u_i\rangle$ is the i-th measurement outcome
- Output $Z_k \coloneqq \frac{1}{n(n-1)\cdots(n-k+1)} \sum_{\text{distinct } i_1, i_2, \ldots, i_k \in [n]} \mathrm{tr}(\rho_{i_1}\rho_{i_2}\cdots\rho_{i_k})$

- Additive error: $\mathrm{tr}(\rho^k) \pm \delta$
- Multiplicative error: $(1 \pm \delta) \cdot \mathrm{tr}(\rho^k)$

Theorem: Z_k can estimate $\mathrm{tr}(\rho^k)$ to multiplicative error δ using

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

copies of ρ.

Additive-error Rényi entropy estimation

- The quantum Rényi entropy of order k: $S_k(\rho) = \frac{1}{1-k}\log \mathrm{tr}(\rho^k)$

δ-**multiplicative** error approximation for $\mathrm{tr}(\rho^k)$ ⟺ δ-**additive** error approximation for $S_k(\rho)$

Fully entangled	Unentangled (our results)	
$\Theta\left(\max\left\{\frac{d^{1-1/k}}{\delta^2}, \frac{d^{2-2/k}}{\delta^{2/k}}\right\}\right)$	$O\left(\max\left\{\frac{d^{2-2/k}}{\delta^2}, \frac{d^{3-2/k}}{\delta^{2/k}}\right\}\right)$	Same trade off point at $\delta = d^{\frac{-k}{2k-2}}$
$\Theta(d^{2-2/k})$	$O(d^{3-2/k})$	When δ is constant

[AISW17]

Parameters

- The highest estimated moment: $K = O\left(\frac{\log(d)}{\log\log(d)}\right)$
- The bucketing threshold hold: $B = O\left(\frac{\epsilon^2 K^2}{d}\right)$
- The overall sample complexity:

$$n = O\left(\frac{d}{B^2\epsilon^2}\right) = O\left(d^3 \cdot K^4 \cdot \frac{1}{\epsilon^6}\right) = O\left(d^3 \cdot \left(\frac{\log\log(d)}{\log(d)}\right)^4 \cdot \frac{1}{\epsilon^6}\right)$$

Summary

- Spectrum is important because we care about multiple unitarily invariant properties.
- We give an adaptive spectrum estimation algorithm using unentangled measurements (in fact, only uniform POVM) and a sub-polynomial factor fewer samples than the full state tomography.
  - The idea is to first split the eigenvalues into small and large buckets without disturbing the spectrum by too much.
  - Then perform local moment matching on the small bucket.
- We provide numerical evidence in the setting of fully entangled measurements that spectrum estimation can only improve over full state tomography by a sub-polynomial factor.