Mixed state tomography reduces to pure state tomography

Angelos Pelecanos
UC Berkeley

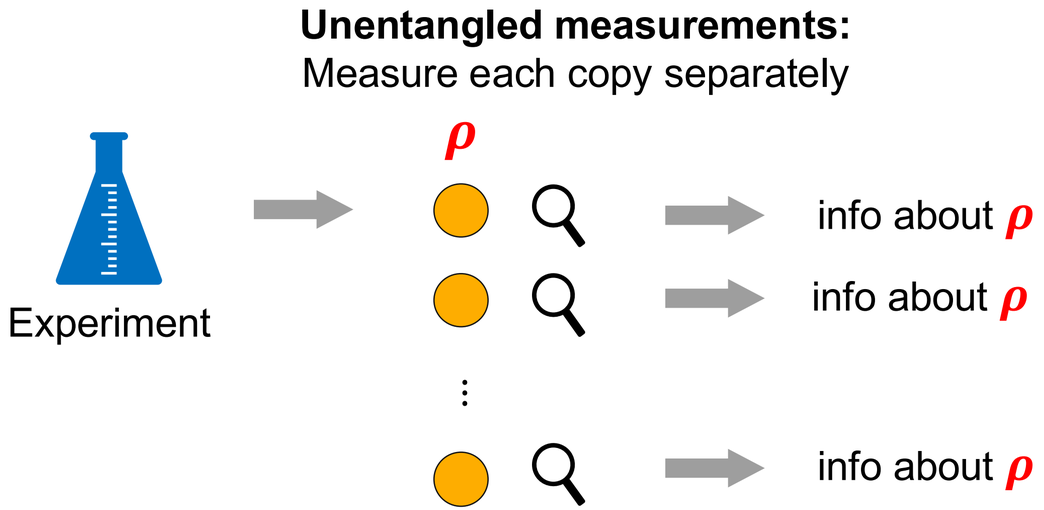
Jack Spilecki
UC Berkeley

Ewin Tang
UC Berkeley

John Wright
UC Berkeley

Learning Pure vs Mixed States

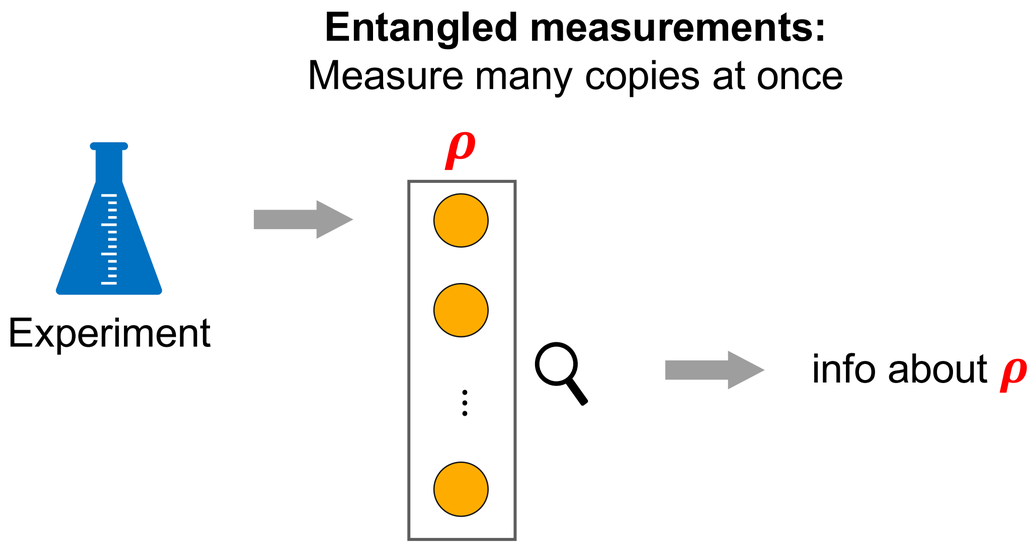
Quantum State Tomography



Given: n copies of a quantum state ρ.

Goal: Learn something about ρ.

Quantum State Tomography



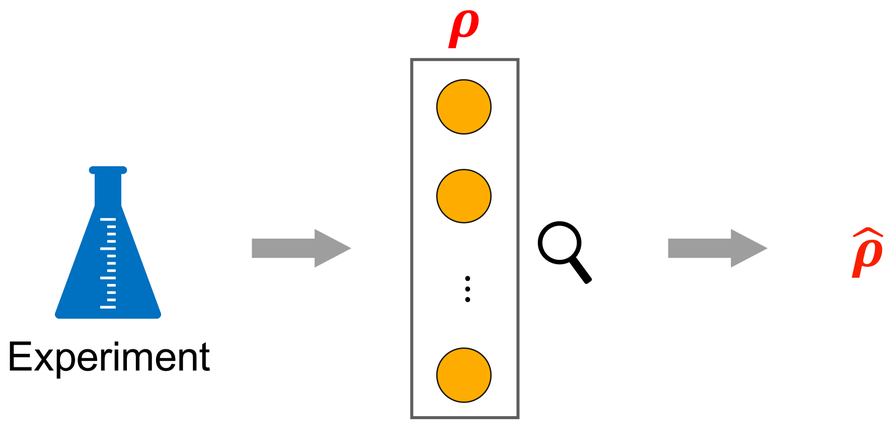
Given: n copies of a quantum state ρ.

Goal: Learn something about ρ.

Quantum state tomography: Learn the entire state ρ.

More tomography problems later on.

Quantum State Tomography



Given: n copies of a rank-r state $\rho \in \mathbb{C}^{d \times d}$.

Goal: Output $\hat{\rho}$ such that $D(\hat{\rho}, \rho) \leq \varepsilon$ with probability $1 - \delta$.

D will be **trace distance** D_{tr} or **infidelity** $1 - F$.

Samples: Find the minimum number of copies n.

Implementation: Circuits with $\mathrm{poly}(n)$ gates.

Pure states: $r = 1$.

Infidelity:
more stringent than trace

Pure state tomography
when $\boldsymbol{\rho}$ is pure ($\boldsymbol{r = 1}$)

- Optimal algorithm from the 90's.
- Uses $\boldsymbol{n = \Theta(d)}$ samples.

[Hay98,SSW25]

Learn $\boldsymbol{\Theta(d)}$ parameters

Mixed state tomography
when $\boldsymbol{\rho}$ is mixed ($\boldsymbol{1 \leq r \leq d}$)

- Optimal algorithms in 2016.
- Uses $\boldsymbol{n = \Theta(rd)}$ samples.

[HHJ+16,OW16,Yue23,SSW25]

Pure state tomography
when $\boldsymbol{\rho}$ is pure ($\boldsymbol{r = 1}$)

- Optimal algorithm from the 90's.
- Uses $\boldsymbol{n = \Theta(d)}$ samples.

  [Hay98,SSW25]
- Unentangled measurements suffice.

  [KRT14,GKKT20]
- Computationally **efficient**.
- Conceptually **simple**.

Mixed state tomography
when $\boldsymbol{\rho}$ is mixed ($\boldsymbol{1 \leq r \leq d}$)

- Optimal algorithms in 2016.
- Uses $\boldsymbol{n = \Theta(rd)}$ samples.

  [HHJ+16,OW16,Yue23,SSW25]
- Requires **entangled** measurements.

  [CHL+23,CLL24b]
- Don't know how to implement efficiently.
- Heavy use of representation theory.

  [HHJ+16,OW16,**P**SW25]

Currently: Pure and mixed state tomography seem very different!

Pure vs Mixed State Tomography

Main result: Mixed state tomography reduces to pure state tomography.

Main ingredient: Efficient procedure to generate random purifications [TWZ25].

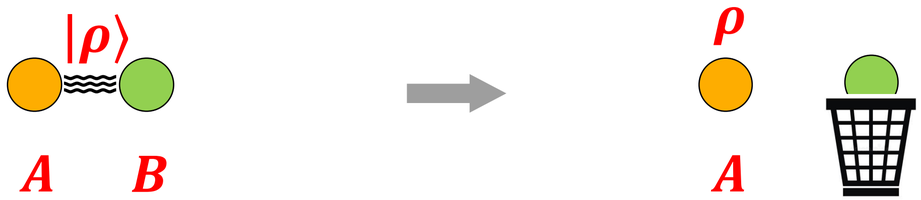
This talk: Use of purification in Mixed State Tomography (not just “vanilla” tomography).

Recover (or improves upon) essentially all results on mixed state tomography.

Establish that purification is the object to study for mixed state tomography.

Purification

Definition: The state $|\rho\rangle_{AB}$ is a **purification** of ρ_A if $\mathrm{tr_B}(|\rho\rangle\langle\rho|_{AB}) = \rho_A$.



Register B is $\mathbb{C}^r$

$\rho = \sum_{i=1}^{r} \alpha_i |\psi_i\rangle\langle\psi_i|$ → Natural purification: $|\rho_0\rangle = \sum_{i=1}^{r} \sqrt{\alpha_i}|\psi_i\rangle_A \otimes |i\rangle_B$

→ General purification: $|\rho\rangle = \sum_{i=1}^{r} \sqrt{\alpha_i}|\psi_i\rangle_A \otimes |u_i\rangle_B$

For any orthonormal basis $|u_1\rangle, \dots, |u_r\rangle$

Question: Given copies of state ρ, can we generate a purification $|\rho\rangle$?

Generating fixed purification is impossible

Assume: exists $\boldsymbol{\Phi}$ that maps a mixed state $\boldsymbol{\rho}$ to some **fixed** purification $|\boldsymbol{\rho}\rangle$.

Then it holds that

$$\boldsymbol{\Phi}(\boldsymbol{\rho}) = |\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|, \qquad \boldsymbol{\Phi}(\boldsymbol{\sigma}) = |\boldsymbol{\sigma}\rangle\langle\boldsymbol{\sigma}|$$

Linearity implies:

$$\boldsymbol{\Phi}\left(\frac{\mathbf{1}}{\mathbf{2}}\cdot\boldsymbol{\rho} + \frac{\mathbf{1}}{\mathbf{2}}\cdot\boldsymbol{\sigma}\right) = \frac{\mathbf{1}}{\mathbf{2}}\cdot\boldsymbol{\Phi}(\boldsymbol{\rho}) + \frac{\mathbf{1}}{\mathbf{2}}\cdot\boldsymbol{\Phi}(\boldsymbol{\sigma}) = \boxed{\frac{\mathbf{1}}{\mathbf{2}}\left(|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}| + |\boldsymbol{\sigma}\rangle\langle\boldsymbol{\sigma}|\right)}$$

not a pure state!!!

Thus: no channel maps one copy of $\boldsymbol{\rho}$ to any **fixed** purification.

Extended to many copies and approximate purifications [LDCL25].

Efficiently generating random purifications

Surprisingly, [TWZ25] show how to efficiently construct a random purification.

Works for all values of n.

Theorem. Channel $\mathbf{purify}(\cdot)$ that takes n copies of a rank-r mixed state ρ and outputs

$$\mathbf{purify}(\rho^{\otimes n}) = \mathop{E}_{|\rho\rangle}[|\rho\rangle\langle\rho|^{\otimes n}].$$

Can implement to diamond distance δ in time $\mathrm{poly}(n, \log(d), \log(1/\delta))$.

Efficiently generating random purifications

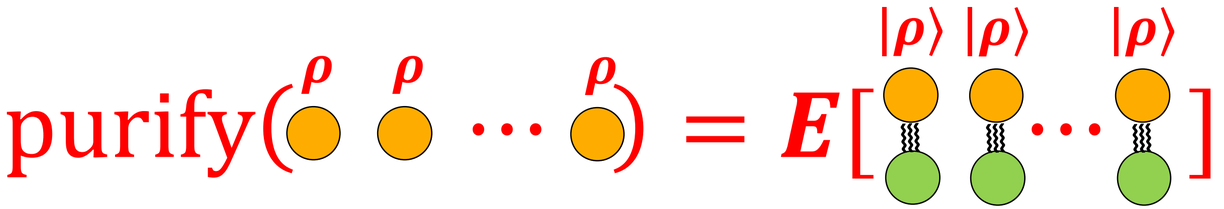
Surprisingly, [TWZ25] show how to efficiently construct a random purification.

Theorem. Channel $\mathbf{purify}(\cdot)$ that takes $\boldsymbol{n}$ copies of a rank-$\boldsymbol{r}$ mixed state $\boldsymbol{\rho}$ and outputs

$$\mathbf{purify}(\boldsymbol{\rho}^{\otimes n}) = \mathop{\boldsymbol{E}}_{|\rho\rangle}[|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|^{\otimes n}].$$

Can implement to diamond distance $\boldsymbol{\delta}$ in time $\mathrm{poly}(\boldsymbol{n}, \log(\boldsymbol{d}), \log(\mathbf{1}/\boldsymbol{\delta}))$.

$$|\boldsymbol{\rho}\rangle = \sum_{i=1}^{r} \sqrt{\boldsymbol{\alpha_i}}|\boldsymbol{\psi_i}\rangle_A \otimes |\boldsymbol{u_i}\rangle_B \text{ for random orthonormal basis } |\boldsymbol{u_1}\rangle, \ldots, |\boldsymbol{u_r}\rangle.$$



Efficiently generating random purifications

Surprisingly, [TWZ25] show how to efficiently construct a random purification.

Theorem. Channel $\mathbf{purify}(\cdot)$ that takes $\boldsymbol{n}$ copies of a rank-$\boldsymbol{r}$ mixed state $\boldsymbol{\rho}$ and outputs

$$\mathbf{purify}(\boldsymbol{\rho}^{\otimes n}) = \mathop{\boldsymbol{E}}_{|\boldsymbol{\rho}\rangle}[|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|^{\otimes n}].$$

Can implement to diamond distance $\boldsymbol{\delta}$ in time $\mathrm{poly}(\boldsymbol{n}, \log(\boldsymbol{d}), \log(\mathbf{1}/\boldsymbol{\delta}))$.

Can generate efficiently.

Efficiently generating random purifications

Surprisingly, [TWZ25] show how to efficiently construct a random purification.

Theorem. Channel $\mathbf{purify}(\cdot)$ that takes $\boldsymbol{n}$ copies of a rank-$\boldsymbol{r}$ mixed state $\boldsymbol{\rho}$ and outputs

$$\mathbf{purify}(\boldsymbol{\rho}^{\otimes \boldsymbol{n}}) = \mathop{\boldsymbol{E}}_{|\boldsymbol{\rho}\rangle}[|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|^{\otimes \boldsymbol{n}}].$$

Can implement to diamond distance $\boldsymbol{\delta}$ in time $\mathrm{poly}(\boldsymbol{n}, \log(\boldsymbol{d}), \log(\mathbf{1}/\boldsymbol{\delta}))$.

Random purification channel

Let's try to understand the expectation: $\mathop{E}_{|\rho\rangle}[|\rho\rangle\langle\rho|^{\otimes n}]$.

Illuminating example is one copy: $\text{purify}(\rho) = E[|\rho\rangle\langle\rho|]$

$$|\rho\rangle = \sum_{i=1}^{r} \sqrt{\alpha_i}|\psi_i\rangle_A \otimes |u_i\rangle_B$$
for random basis $|u_1\rangle, \ldots, |u_r\rangle$.

$$E[|\rho\rangle\langle\rho|] = E\left(\sum_{i=1}^{r} \sqrt{\alpha_i}\,|\psi_i\rangle \otimes |u_i\rangle\right)\left(\sum_{j=1}^{r} \sqrt{\alpha_j}\,\langle\psi_j| \otimes \langle u_j|\right)$$

$$= \sum_{i,j=1}^{r} \sqrt{\alpha_i}\sqrt{\alpha_j}\,|\psi_i\rangle\langle\psi_j| \otimes \boxed{E[|u_i\rangle\langle u_j|]} = \begin{cases} I/r, & i = j \\ 0, & i \neq j \end{cases}$$

$$= \sum_{i=1}^{r} \alpha_i\,|\psi_i\rangle\langle\psi_i| \otimes I/r \qquad = \rho \otimes I/r$$

Purifying one copy amounts to appending a maximally-mixed state.

Very easy!
General case includes representation theory.

Mixed state tomography:

Simpler, faster, with high probability

Mixed-to-pure state reduction

Introduce reduction from mixed to pure state tomography via the random purification channel.

Let $\boldsymbol{A}$ be your favorite pure state tomography algorithm.

The following is a mixed state tomography algorithm $\mathrm{Mix}(\boldsymbol{A})$.

Algorithm $\mathrm{Mix}(\boldsymbol{A})$**.** Given $\boldsymbol{n}$ copies of $\boldsymbol{d}$-dimensional rank-$\boldsymbol{r}$ state $\boldsymbol{\rho}$:

1. Run $\mathrm{purify}(\cdot)$ to produce $\boldsymbol{n}$ copies of random purification $|\boldsymbol{\rho}\rangle \in \mathbb{C}^{\boldsymbol{d}} \otimes \mathbb{C}^{\boldsymbol{r}}$.
2. Run $\boldsymbol{A}$ and obtain estimate $|\widehat{\boldsymbol{\rho}}\rangle$ of $|\boldsymbol{\rho}\rangle$.
3. Compute an estimate of $\boldsymbol{\rho}$ using $\widehat{\boldsymbol{\rho}} = \mathrm{tr}_{\boldsymbol{B}}(|\widehat{\boldsymbol{\rho}}\rangle\langle\widehat{\boldsymbol{\rho}}|)$.

Mixed-to-pure state reduction

Algorithm $\mathrm{Mix}(\boldsymbol{A})$**.** Given $\boldsymbol{n}$ copies of $\boldsymbol{d}$-dimensional rank-$\boldsymbol{r}$ state $\boldsymbol{\rho}$:

1. Run $\mathrm{purify}(\cdot)$ to produce $\boldsymbol{n}$ copies of random purification $|\boldsymbol{\rho}\rangle \in \mathbb{C}^{\boldsymbol{d}} \otimes \mathbb{C}^{\boldsymbol{r}}$.
2. Run $\boldsymbol{A}$ and obtain estimate $|\widehat{\boldsymbol{\rho}}\rangle$ of $|\boldsymbol{\rho}\rangle$.
3. Compute an estimate of $\boldsymbol{\rho}$ using $\widehat{\boldsymbol{\rho}} = \mathrm{tr}_{\boldsymbol{B}}(|\widehat{\boldsymbol{\rho}}\rangle\langle\widehat{\boldsymbol{\rho}}|)$.

Recover $\boldsymbol{\Theta(rd)}$ sample complexity for mixed states!

Quick check:

1. Map $\boldsymbol{\rho}$ to pure state $|\boldsymbol{\rho}\rangle$ in $\boldsymbol{rd}$-dimensional space.
2. Pure state tomography requires $\boldsymbol{\Theta}(\#\mathrm{params}) = \boldsymbol{\Theta(rd)}$ samples to learn $|\widehat{\boldsymbol{\rho}}\rangle \approx |\boldsymbol{\rho}\rangle$.
3. Tracing out gives

$$\widehat{\boldsymbol{\rho}} = \mathrm{tr}_{\boldsymbol{B}}(|\widehat{\boldsymbol{\rho}}\rangle\langle\widehat{\boldsymbol{\rho}}|) \approx \mathrm{tr}_{\boldsymbol{B}}(|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|) = \boldsymbol{\rho}.$$

Mixed-to-pure state reduction

Algorithm $\mathrm{Mix}(\boldsymbol{A})$**.** Given $\boldsymbol{n}$ copies of $\boldsymbol{d}$-dimensional rank-$\boldsymbol{r}$ state $\boldsymbol{\rho}$:

1. Run $\mathrm{purify}(\cdot)$ to produce $\boldsymbol{n}$ copies of random purification $|\boldsymbol{\rho}\rangle \in \mathbb{C}^{\boldsymbol{d}} \otimes \mathbb{C}^{\boldsymbol{r}}$.
2. Run $\boldsymbol{A}$ and obtain estimate $|\widehat{\boldsymbol{\rho}}\rangle$ of $|\boldsymbol{\rho}\rangle$.
3. Compute an estimate of $\boldsymbol{\rho}$ using $\widehat{\boldsymbol{\rho}} = \mathrm{tr}_{\boldsymbol{B}}(|\widehat{\boldsymbol{\rho}}\rangle\langle\widehat{\boldsymbol{\rho}}|)$.

Let's do this more carefully…

Which pure state tomography algorithm $\boldsymbol{A}$ should we use?

Two sample-optimal algorithms.

Pure state algorithm 1: A_{GKKT}

Uses **unentangled** measurements.

Performs the uniform POVM.

Know how to implement efficiently.

Theorem. Given $n = O((d + \log(1/\delta))/\varepsilon)$ copies of a pure state $|\psi\rangle$,

A_{GKKT} outputs a pure state $|\widehat{\psi}\rangle$ such that with probability $1 - \delta$,

$$|\langle\psi|\widehat{\psi}\rangle|^2 \geq 1 - \varepsilon.$$

Moreover, A_{GKKT} can be implemented with a $\mathrm{poly}(n)$-size circuit.

Use because it has efficient implementation (also simpler).

It does not suffice to achieve all known tomography applications.

Pure state algorithm 2: A_{Hayashi}

Due to Hayashi [Hay98].

Uses **entangled** measurements.

The optimal algorithm for pure state tomography.

Do not know how to implement yet.

Will use later to recover the remaining state learning applications.

The $\mathrm{Mix}(A_{\mathrm{GKKT}})$ mixed state algorithm

Algorithm $\mathrm{Mix}(A_{\mathrm{GKKT}})$**.** Given n copies of d-dimensional rank-r state ρ:

1. Run $\mathrm{purify}(\rho^{\otimes n})$ to produce n copies of random purification $|\rho\rangle \in \mathbb{C}^d \otimes \mathbb{C}^r$.
2. Run A_{GKKT}.

   With probability $1-\delta$, outputs $|\hat{\rho}\rangle$ such that

   $$|\langle\rho|\hat{\rho}\rangle|^2 \geq 1-\varepsilon$$

3. Output $\hat{\rho} = \mathrm{tr}_B(|\hat{\rho}\rangle\langle\hat{\rho}|)$. Then

   $$F(\rho,\hat{\rho}) \geq F(|\rho\rangle\langle\rho|, |\hat{\rho}\rangle\langle\hat{\rho}|) = |\langle\rho|\hat{\rho}\rangle|^2 \geq 1-\varepsilon$$

Let n be sample complexity of A_{GKKT} for pure state in $\mathbb{C}^{rd}$

The $\mathrm{Mix}(A_{\mathrm{GKKT}})$ mixed state algorithm

Algorithm $\mathrm{Mix}(A_{\mathrm{GKKT}})$**.** Given n copies of d-dimensional rank-r state ρ:

1. Run $\mathrm{purify}(\rho^{\otimes n})$ to produce n copies of random purification $|\rho\rangle \in \mathbb{C}^d \otimes \mathbb{C}^r$.
2. Run A_{GKKT}.

   With probability $1-\delta$, outputs $|\widehat{\rho}\rangle$ such that

   $$|\langle\rho|\widehat{\rho}\rangle|^2 \geq 1-\varepsilon$$

3. Output $\widehat{\rho} = \mathrm{tr}_B(|\widehat{\rho}\rangle\langle\widehat{\rho}|)$. Then

   $$F(\rho,\widehat{\rho}) \geq F(|\rho\rangle\langle\rho|, |\widehat{\rho}\rangle\langle\widehat{\rho}|) = |\langle\rho|\widehat{\rho}\rangle|^2 \geq 1-\varepsilon$$

Let n be sample complexity of A_{GKKT} for pure state in $\mathbb{C}^{rd}$

Theorem. Given n copies of rank-r state ρ, $\mathrm{Mix}\,(A_{\mathrm{GKKT}})$ outputs $\widehat{\rho}$ such that

$$F(\rho,\widehat{\rho}) \geq 1-\varepsilon$$

with probability $1-\delta$, when $n = O((rd + \log(1/\delta))/\varepsilon)$.

Moreover, $\mathrm{Mix}(A_{\mathrm{GKKT}})$ can be implemented with a $\mathrm{poly}(n)$-size circuit.

Sample complexity of mixed state tomography

Upper bounds $n = \widetilde{O}\left(\frac{rd + \log(1/\delta)}{\varepsilon}\right)$ [HHJWY16] $n = O\left(\frac{rd \cdot \log(1/\delta)}{\varepsilon}\right)$ [PSW25]

Lower bounds $n = \Omega\left(\frac{rd + \log(1/\delta)}{\varepsilon}\right)$ [Yue23,SSW25]

Lower bound [Yue23] also uses a mixed-to-pure strategy.

Our bound $n = O\left(\frac{rd + \log(1/\delta)}{\varepsilon}\right)$

Short remark: Confidence parameter δ

Very important parameter.

Appears in “with 99% probability” or “95% confidence interval”.

Can always get multiplicative dependence on $\log(1/\delta)$ via amplification.

Remarkable: samples for $\delta = 0.01$ $\approx$ samples for $\delta = \exp(-rd)$

The $\mathrm{Mix}(A_{\mathrm{GKKT}})$ mixed state algorithm

Theorem. Given n copies of rank-r state ρ, $\mathrm{Mix}\,(A_{\mathrm{GKKT}})$ outputs $\hat{\rho}$ such that

$$F(\rho, \hat{\rho}) \geq 1 - \varepsilon$$

with probability $1 - \delta$, when $n = O((rd + \log(1/\delta))/\varepsilon)$.

Moreover, $\mathrm{Mix}(A_{\mathrm{GKKT}})$ can be implemented with a $\mathrm{poly}(n)$-size circuit.

Simple and intuitive algorithm. ✓

Can be implemented efficiently. ✓

Optimal sample complexity. ✓

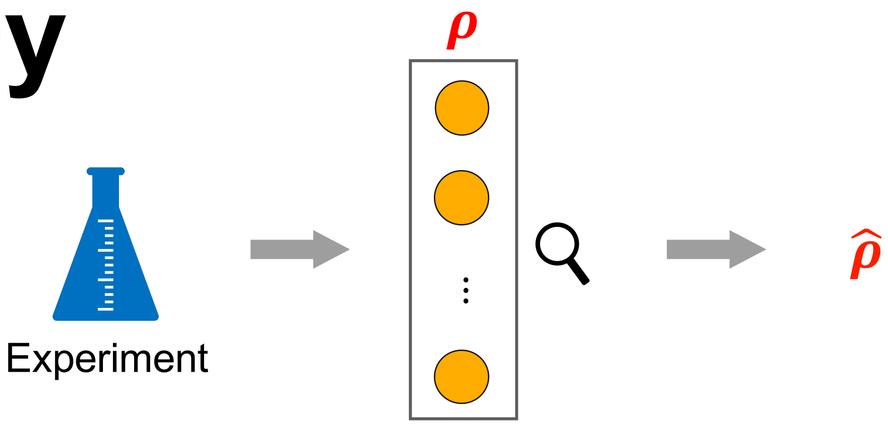
Remark: Only needed entangled measurements to produce the purification.

New Unbiased Estimators

Quantum State Tomography

Goal: Learn something about $\boldsymbol{\rho}$.

Quantum state tomography: Learn the entire state $\boldsymbol{\rho}$.



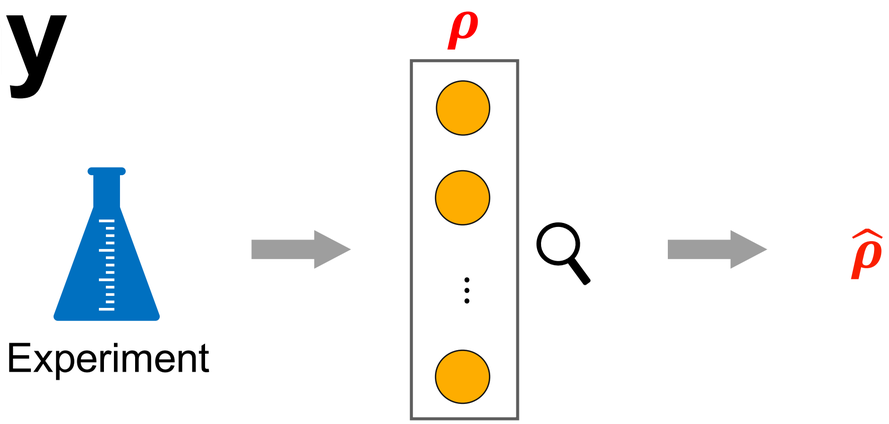
$\boldsymbol{k}$**-entangled tomography:** Estimate $\boldsymbol{\rho}$ to distance $\boldsymbol{\varepsilon}$ by measuring $\boldsymbol{k}$ copies at a time.

Near-term devices may only implement smaller entangled measurements.

Interpolates between unentangled ($\boldsymbol{k = 1}$) and entangled ($\boldsymbol{k = n}$) measurements.

Help us understand the power of bigger measurements.

Quantum State Tomography



Goal: Learn something about $\boldsymbol{\rho}$.

Quantum state tomography: Learn the entire state $\boldsymbol{\rho}$.

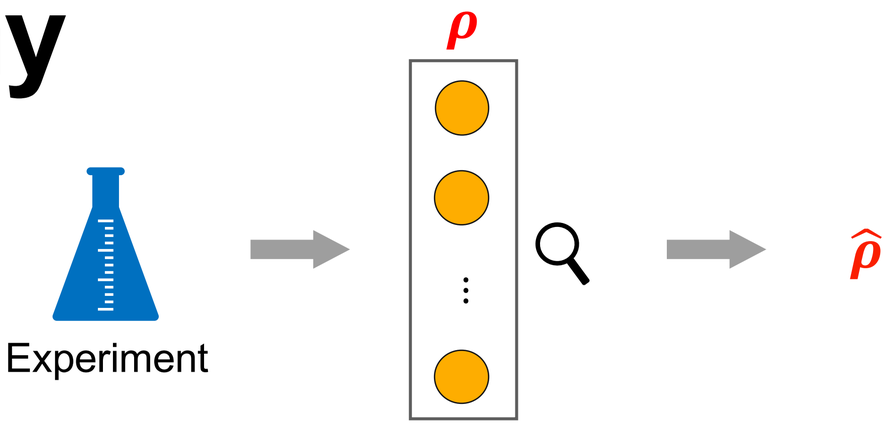
$\boldsymbol{k}$**-entangled tomography:** Estimate $\boldsymbol{\rho}$ to distance $\boldsymbol{\varepsilon}$ by measuring $\boldsymbol{k}$ copies at a time.

Shadow tomography: Observables $\boldsymbol{O_1, \ldots, O_m}$,estimate $\mathrm{tr}(\boldsymbol{O_1} \cdot \boldsymbol{\rho}), \ldots, \mathrm{tr}(\boldsymbol{O_m} \cdot \boldsymbol{\rho})$ to error $\boldsymbol{\varepsilon}$.

Many applications to learning, testing among others.

Often much more efficient than full-state tomography.

Quantum State Tomography



Goal: Learn something about $\boldsymbol{\rho}$.

Quantum state tomography: Learn the entire state $\boldsymbol{\rho}$.

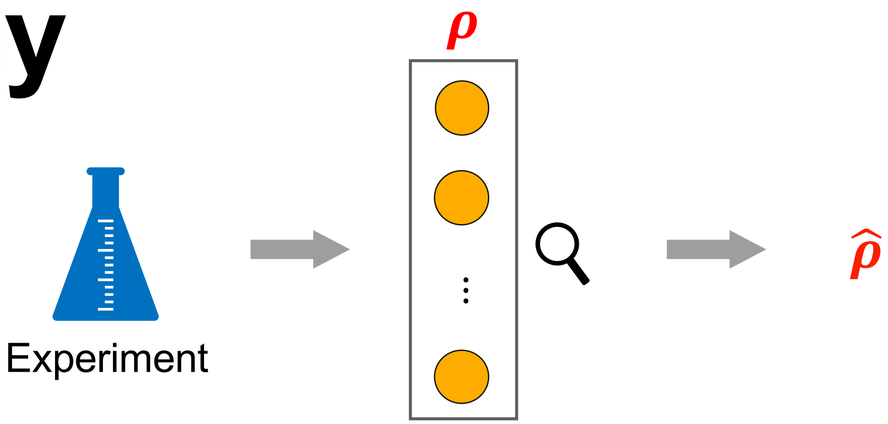
$\boldsymbol{k}$**-entangled tomography:** Estimate $\boldsymbol{\rho}$ to distance $\boldsymbol{\varepsilon}$ by measuring $\boldsymbol{k}$ copies at a time.

Shadow tomography: Observables $\boldsymbol{O_1, \ldots, O_m}$,estimate $\mathrm{tr}(\boldsymbol{O_1} \cdot \boldsymbol{\rho}), \ldots, \mathrm{tr}(\boldsymbol{O_m} \cdot \boldsymbol{\rho})$ to error $\boldsymbol{\varepsilon}$.

Quantum metrology: Given parametrized state $\boldsymbol{\rho_\theta}$,estimate $\boldsymbol{\theta} \in \mathbb{R}^{\boldsymbol{m}}$.

Applications to fundamental physics tests.

Quantum State Tomography



Goal: Learn something about $\boldsymbol{\rho}$.

Quantum state tomography: Learn the entire state $\boldsymbol{\rho}$.

$\boldsymbol{k}$**-entangled tomography:** Estimate $\boldsymbol{\rho}$ to distance $\boldsymbol{\varepsilon}$ by measuring $\boldsymbol{k}$ copies at a time.

Shadow tomography: Observables $\boldsymbol{O_1, \ldots, O_m}$,estimate $\mathrm{tr}(\boldsymbol{O_1} \cdot \boldsymbol{\rho}), \ldots, \mathrm{tr}(\boldsymbol{O_m} \cdot \boldsymbol{\rho})$ to error $\boldsymbol{\varepsilon}$.

Quantum metrology: Given parametrized state $\boldsymbol{\rho_\theta}$,estimate $\boldsymbol{\theta} \in \mathbb{R}^{\boldsymbol{m}}$.

First two focus of recent works by Chen, Li, Liu [CLL24a,CLL24b].

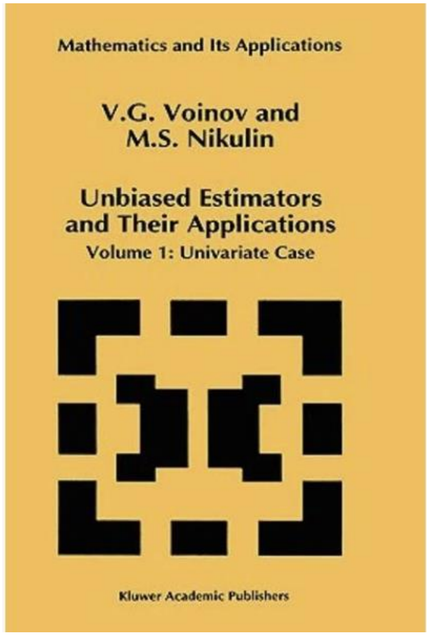
Bottleneck: no **unbiased estimators** for entangled tomography.

Unbiased estimators

Definition: An estimator is unbiased if $\boldsymbol{E[\hat{\rho}] = \rho}$.

Unbiased estimators have many desirable features.

> "Unbiasedness is one of the most important properties of statistical estimators."
>
> Voinov and Nikulin



Collect ≈ 1000 unbiased estimators for different statistical parameters.

Theoretical motivation: Construct unbiased estimators for fundamental problems.

Also: Unbiased estimators give simple algorithms for quantum state tomography tasks.

Unbiasedness gives simple algorithms

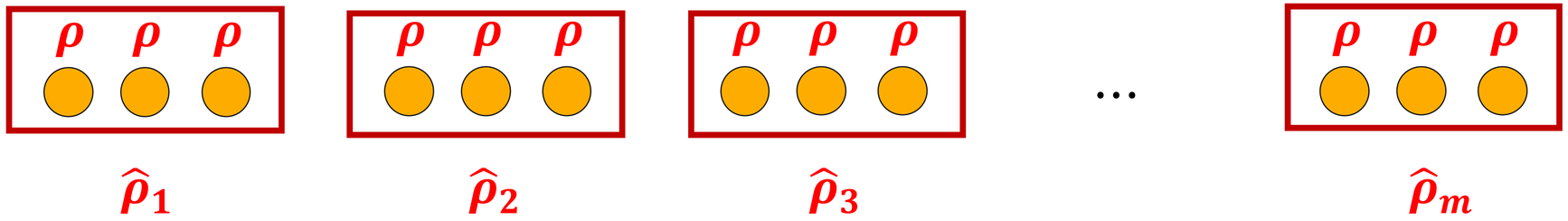
Definition: An estimator is unbiased if $\boldsymbol{E[\hat{\rho}] = \rho}$.

Quality of an unbiased estimator is given by its "variance".

Application 1: $\boldsymbol{k}$-entangled tomography.

Estimate $\boldsymbol{\rho}$ to distance $\boldsymbol{\varepsilon}$ by measuring $\boldsymbol{k}$ copies at a time.

1. Divide the $\boldsymbol{n}$ copies of $\boldsymbol{\rho}$ into $\boldsymbol{m = n/k}$ groups.



2. Run unbiased tomography on each group.

3. Output $\boldsymbol{\hat{\rho} = \frac{1}{m} \cdot (\hat{\rho}_1 + \cdots + \hat{\rho}_m)}$. also unbiased average decreases variance

Unbiasedness gives simple algorithms

Definition: An estimator is unbiased if $\boldsymbol{E[\hat{\rho}] = \rho}$.

Quality of an unbiased estimator is given by its "variance".

Application 2: Shadow tomography.

Observables $\boldsymbol{O_1, \dots, O_m}$,estimate $\mathrm{tr}(\boldsymbol{O_1 \cdot \rho}), \dots, \mathrm{tr}(\boldsymbol{O_m \cdot \rho})$ to error $\boldsymbol{\varepsilon}$.

1. Run tomography on $\boldsymbol{n}$ copies. Compute an unbiased estimator $\boldsymbol{\hat{\rho}}$.
2. For each $\boldsymbol{i}$, output plug-in estimator: $\boldsymbol{\hat{o}_i = \mathrm{tr}(O_i \cdot \hat{\rho})}$.

Note: The observables are unbiased, because $\hat{\rho}$ is unbiased.

Observables have lower variance than $\hat{\rho}$.

Classical shadows (oblivious shadow tomography)**:** Measurement oblivious to observables.

Unbiasedness gives simple algorithms

Definition: An estimator is unbiased if $\boldsymbol{E[\hat{\rho}] = \rho}$.

Quality of an unbiased estimator is given by its “variance”.

Application 3: Multi-parameter quantum metrology.

Given copies of a parametrized state $\boldsymbol{\rho_\theta}$, with $\boldsymbol{\theta \in \mathbb{R}^m}$, output an estimate of $\boldsymbol{\theta}$.

Outside the scope of this talk, see paper for more details.

Unbiased estimators

Definition: An estimator is unbiased if $\boldsymbol{E[\hat{\rho}] = \rho}$.

---

k-entangled tomography: Estimate $\boldsymbol{\rho}$ to distance $\boldsymbol{\varepsilon}$ by measuring $\boldsymbol{k}$ copies at a time.

Shadow tomography: Observables $\boldsymbol{O_1, \dots, O_m}$,estimate $\mathrm{tr}(\boldsymbol{O_1 \cdot \rho}), \dots, \mathrm{tr}(\boldsymbol{O_m \cdot \rho})$ to error $\boldsymbol{\varepsilon}$.

Quantum metrology: Given parametrized state $\boldsymbol{\rho_\theta}$,estimate $\boldsymbol{\theta \in \mathbb{R}^m}$.

Debiased Keyl's algorithm [PSW25]**:** first unbiased estimator for entangled tomography

- Carefully modified one of the sample-optimal algorithms [Key06,OW16] to remove bias.
- Obtained improved or optimal bounds for these problems.
- Involved analysis with heavy use of representation theory.

Random purification channel: New, simpler unbiased estimator for entangled tomography.

New unbiased estimator

Recall the second pure state tomography algorithm $\boldsymbol{A}_{\mathrm{Hayashi}}$.

Unfortunately, it is biased.

Fortunately, simple modification suffices to remove bias [GPS24] — obtain $\boldsymbol{A}_{\mathrm{GPS}}$.

Algorithm $\mathrm{Mix}(\boldsymbol{A}_{\mathrm{GPS}})$**.** Given $\boldsymbol{n}$ copies of $\boldsymbol{d}$-dimensional rank-$\boldsymbol{r}$ state $\boldsymbol{\rho}$:

1. Run $\mathrm{purify}(\cdot)$ to produce $\boldsymbol{n}$ copies of random purification $|\boldsymbol{\rho}\rangle \in \mathbb{C}^{\boldsymbol{d}} \otimes \mathbb{C}^{\boldsymbol{r}}$.
2. Run $\boldsymbol{A}_{\mathrm{GPS}}$ and obtain estimate $\widehat{\boldsymbol{\sigma}}$ of $|\boldsymbol{\rho}\rangle$.
3. Compute an estimate of $\boldsymbol{\rho}$ using $\widehat{\boldsymbol{\rho}} = \mathrm{tr}_{\boldsymbol{B}}(\widehat{\boldsymbol{\sigma}})$.

New unbiased estimator

Algorithm $\mathrm{Mix}(A_{\mathrm{GPS}})$**.** Given n copies of d-dimensional rank-r state ρ:

1. Run $\mathrm{purify}(\cdot)$ to produce n copies of random purification $|\rho\rangle \in \mathbb{C}^d \otimes \mathbb{C}^r$.
2. Run A_{GPS} and obtain estimate $\hat{\sigma}$ of $|\rho\rangle$.
3. Compute an estimate of ρ using $\hat{\rho} = \mathrm{tr}_B(\hat{\sigma})$.

Inherits the **unbiasedness** of A_{GPS}:

$$E[\hat{\rho}] = E[\mathrm{tr}_B(\hat{\sigma})] = \mathrm{tr}_B(E[\hat{\sigma}]) = \mathrm{tr}_B(|\rho\rangle\langle\rho|) = \rho.$$

Matches the **variance** of the debiased Keyl's algorithm:

$$E[\hat{\rho} \otimes \hat{\rho}] = \frac{n-1}{n} \cdot \rho \otimes \rho + \frac{1}{n} \cdot (\rho \otimes I + I \otimes \rho) \cdot \mathrm{SWAP} + \frac{E[\ell(\lambda)]}{n^2} \cdot \mathrm{SWAP} - \mathrm{Lower}_{\rho}$$

What is $\ell(\lambda)$?
Coming up soon…

New unbiased estimator

Inherits the **unbiasedness** of $\boldsymbol{A}_{\mathrm{GPS}}$:

$$\boldsymbol{E}[\hat{\boldsymbol{\rho}}] = \boldsymbol{E}[\mathrm{tr}_{\boldsymbol{B}}(\hat{\boldsymbol{\sigma}})] = \mathrm{tr}_{\boldsymbol{B}}(\boldsymbol{E}[\hat{\boldsymbol{\sigma}}]) = \mathrm{tr}_{\boldsymbol{B}}(|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|) = \boldsymbol{\rho}.$$

Matches the **variance** of the debiased Keyl's algorithm:

$$\boldsymbol{E}[\hat{\boldsymbol{\rho}} \otimes \hat{\boldsymbol{\rho}}] = \frac{\boldsymbol{n-1}}{\boldsymbol{n}} \cdot \boldsymbol{\rho} \otimes \boldsymbol{\rho} + \frac{\boldsymbol{1}}{\boldsymbol{n}} \cdot (\boldsymbol{\rho} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{\rho}) \cdot \mathrm{SWAP} + \frac{\boldsymbol{E}[\ell(\boldsymbol{\lambda})]}{\boldsymbol{n^2}} \cdot \mathrm{SWAP} - \mathrm{Lower}_{\boldsymbol{\rho}}$$

May seem intimidating at first.

Actually very easy to use!

Involves nice matrices.

A new unbiased tomography estimator

Inherits the **unbiasedness** of A_{GPS}:

$$E[\hat{\rho}] = E[\mathrm{tr}_B(\hat{\sigma})] = \mathrm{tr}_B(E[\hat{\sigma}]) = \mathrm{tr}_B(|\rho\rangle\langle\rho|) = \rho.$$

Matches the **variance** of the debiased Keyl's algorithm:

$$E[\hat{\rho}\otimes\hat{\rho}] = \frac{n-1}{n}\cdot\rho\otimes\rho + \frac{1}{n}\cdot(\rho\otimes I + I\otimes\rho)\cdot\mathrm{SWAP} + \frac{E[\ell(\lambda)]}{n^2}\cdot\mathrm{SWAP} - \mathrm{Lower}_\rho$$

Main term: ideally $E[\hat{\rho}\otimes\hat{\rho}] = \rho\otimes\rho$

A new unbiased tomography estimator

Inherits the **unbiasedness** of A_{GPS}:

$$E[\hat{\rho}] = E[\mathrm{tr}_B(\hat{\sigma})] = \mathrm{tr}_B(E[\hat{\sigma}]) = \mathrm{tr}_B(|\rho\rangle\langle\rho|) = \rho.$$

Matches the **variance** of the debiased Keyl's algorithm:

$$E[\hat{\rho} \otimes \hat{\rho}] = \frac{n-1}{n} \cdot \rho \otimes \rho + \frac{1}{n} \cdot (\rho \otimes I + I \otimes \rho) \cdot \mathrm{SWAP} + \frac{E[\ell(\lambda)]}{n^2} \cdot \mathrm{SWAP} - \mathrm{Lower}_\rho$$

Main error term:
Scales with $1/n$ like variance.

A new unbiased tomography estimator

Inherits the **unbiasedness** of $\boldsymbol{A}_{\mathrm{GPS}}$:

$$\boldsymbol{E}[\hat{\boldsymbol{\rho}}] = \boldsymbol{E}[\mathrm{tr}_{\boldsymbol{B}}(\hat{\boldsymbol{\sigma}})] = \mathrm{tr}_{\boldsymbol{B}}(\boldsymbol{E}[\hat{\boldsymbol{\sigma}}]) = \mathrm{tr}_{\boldsymbol{B}}(|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|) = \boldsymbol{\rho}.$$

Matches the **variance** of the debiased Keyl's algorithm:

$$\boldsymbol{E}[\hat{\boldsymbol{\rho}} \otimes \hat{\boldsymbol{\rho}}] = \frac{\boldsymbol{n-1}}{\boldsymbol{n}} \cdot \boldsymbol{\rho} \otimes \boldsymbol{\rho} + \frac{\boldsymbol{1}}{\boldsymbol{n}} \cdot (\boldsymbol{\rho} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{\rho}) \cdot \mathrm{SWAP} + \boxed{\frac{\boldsymbol{E}[\ell(\boldsymbol{\lambda})]}{\boldsymbol{n^2}} \cdot \mathrm{SWAP}} - \mathrm{Lower}_{\boldsymbol{\rho}}$$

Second-order error term:
Matters only for small $\boldsymbol{n}$.

A new unbiased tomography estimator

Inherits the **unbiasedness** of $\boldsymbol{A_{\mathrm{GPS}}}$:

$$\boldsymbol{E}[\hat{\boldsymbol{\rho}}] = \boldsymbol{E}[\mathrm{tr}_{\boldsymbol{B}}(\hat{\boldsymbol{\sigma}})] = \mathrm{tr}_{\boldsymbol{B}}(\boldsymbol{E}[\hat{\boldsymbol{\sigma}}]) = \mathrm{tr}_{\boldsymbol{B}}(|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|) = \boldsymbol{\rho}.$$

Matches the **variance** of the debiased Keyl's algorithm:

$$\boldsymbol{E}[\hat{\boldsymbol{\rho}} \otimes \hat{\boldsymbol{\rho}}] = \frac{\boldsymbol{n-1}}{\boldsymbol{n}} \cdot \boldsymbol{\rho} \otimes \boldsymbol{\rho} + \frac{\boldsymbol{1}}{\boldsymbol{n}} \cdot (\boldsymbol{\rho} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{\rho}) \cdot \mathrm{SWAP} + \frac{\boldsymbol{E}[\ell(\boldsymbol{\lambda})]}{\boldsymbol{n^2}} \cdot \mathrm{SWAP} - \boxed{\mathrm{Lower}_{\boldsymbol{\rho}}}$$

Lower-order error term:
Vanishes for large $\boldsymbol{d}$.
Negative for our applications.

A new unbiased tomography estimator

Inherits the **unbiasedness** of $\boldsymbol{A}_{\mathrm{GPS}}$:

$$\boldsymbol{E}[\hat{\boldsymbol{\rho}}] = \boldsymbol{E}[\mathrm{tr}_{\boldsymbol{B}}(\hat{\boldsymbol{\sigma}})] = \mathrm{tr}_{\boldsymbol{B}}(\boldsymbol{E}[\hat{\boldsymbol{\sigma}}]) = \mathrm{tr}_{\boldsymbol{B}}(|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|) = \boldsymbol{\rho}.$$

Matches the **variance** of the debiased Keyl's algorithm:

$$\boldsymbol{E}[\hat{\boldsymbol{\rho}} \otimes \hat{\boldsymbol{\rho}}] = \frac{\boldsymbol{n-1}}{\boldsymbol{n}} \cdot \boldsymbol{\rho} \otimes \boldsymbol{\rho} + \frac{\boldsymbol{1}}{\boldsymbol{n}} \cdot (\boldsymbol{\rho} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{\rho}) \cdot \mathrm{SWAP} + \frac{\boldsymbol{E}[\ell(\boldsymbol{\lambda})]}{\boldsymbol{n^2}} \cdot \mathrm{SWAP} - \mathrm{Lower}_{\boldsymbol{\rho}}$$

Very clean and easy to use.

All learning applications follow from the two properties above.

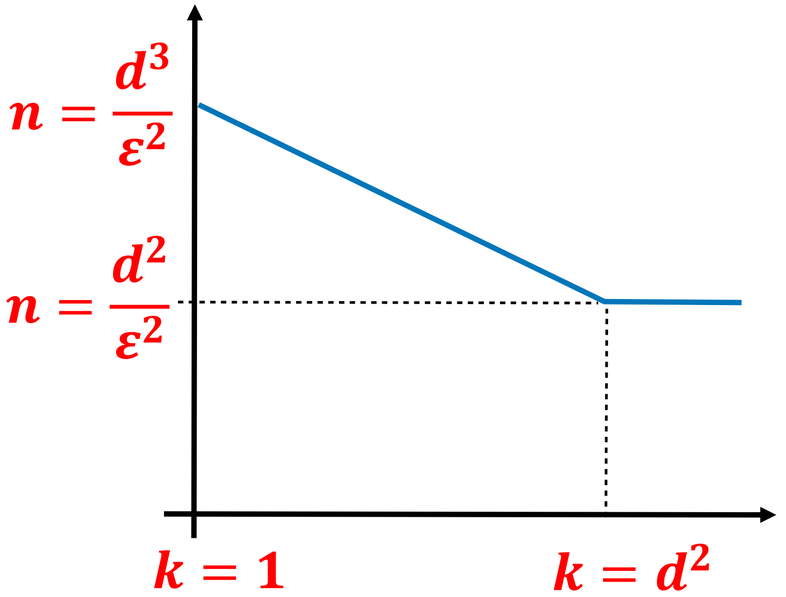
Remark. Requires a crucial modification to the channel, called **quasi-purification**.

Application 1: k-entangled tomography

Goal: Estimate ρ to distance ε by measuring k copies at a time.

Theorem. Samples $n = O\left(\frac{d^3}{\sqrt{k}\varepsilon^2} + \frac{d^2}{\varepsilon^2}\right)$ suffice to trace distance ε.

Matches bounds for unentangled and entangled measurements.



Previous work: $n = \widetilde{O}\left(\frac{d^3}{\sqrt{k}\varepsilon^2}\right)$ only for "small" k [CLL24a].

$n = \Omega\left(\frac{d^3}{\sqrt{k}\varepsilon^2}\right)$ only for "small" k [CLL24a].

Application 2: Shadow tomography

Goal: Observables $\boldsymbol{O_1, \ldots, O_m}$,estimate $\mathrm{tr}(\boldsymbol{O_1} \cdot \boldsymbol{\rho}), \ldots, \mathrm{tr}(\boldsymbol{O_m} \cdot \boldsymbol{\rho})$ to error $\boldsymbol{\varepsilon}$.

Theorem. Samples $\boldsymbol{n = O(\log(m)/\varepsilon^2)}$ suffice when $\boldsymbol{\varepsilon = O(1/d)}$.

- Sample-optimal in high-precision regime.

Previous work: Samples $\boldsymbol{n = O(\log(m)/\varepsilon^2)}$ suffice when $\boldsymbol{\varepsilon = O(1/d^{12})}$]CLL24b].

Classical shadows: Obtain a more general bound.

- Improves upon prior work [GLM24].
- Disproves a conjecture by [CLL24b].

Application 3: Multi-parameter quantum metrology

Goal: Given parametrized state $\boldsymbol{\rho_\theta}$,estimate $\boldsymbol{\theta} \in \mathbb{R}^{\boldsymbol{m}}$.

Well-known bound called the **quantum Cramér-Rao bound**.

We attain twice this bound, which turns out to be optimal.

The bound we obtain has been stated before, but without proof.

Interesting: our techniques are very different from prior work.

See paper for more details.

Quasi-purification

The need for quasi-purification

Recall the random purification channel for one copy:

$$\text{purify}(\boldsymbol{\rho}) = \boldsymbol{E}[|\boldsymbol{\rho}\rangle\langle\boldsymbol{\rho}|] = \boldsymbol{\rho} \otimes \boldsymbol{I}/\boldsymbol{r}$$

If you purify one copy of $\boldsymbol{\rho}$, second register $\mathbb{C}^{\boldsymbol{r}}$ does not contain useful information.

Performing pure-state tomography on both registers $\mathbb{C}^{\boldsymbol{d}} \otimes \mathbb{C}^{\boldsymbol{r}}$ feels wasteful.

Using "standard" random purification channel in $\text{Mix}(\boldsymbol{A}_{GPS})$:

$$\boldsymbol{E}[\hat{\boldsymbol{\rho}} \otimes \hat{\boldsymbol{\rho}}] = \frac{\boldsymbol{n-1}}{\boldsymbol{n}} \cdot \boldsymbol{\rho} \otimes \boldsymbol{\rho} + \frac{\boldsymbol{1}}{\boldsymbol{n}} \cdot (\boldsymbol{\rho} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{\rho}) \cdot \text{SWAP} + \boxed{\frac{\boldsymbol{r}}{\boldsymbol{n^2}} \cdot \text{SWAP}} - \text{Lower}_{\boldsymbol{\rho}}$$

Worse bounds for $\boldsymbol{k}$-entangled and shadow tomography.

$$\boxed{\frac{\boldsymbol{E}[\ell(\boldsymbol{\lambda})]}{\boldsymbol{n^2}} \cdot \text{SWAP}}$$

Debiased Keyl's

The random purification channel

Schur-Weyl duality: $$\rho_A^{\otimes n} = \sum_{\lambda \vdash n} |\lambda\rangle\langle\lambda|_A \otimes I_A \otimes \nu_\lambda(\rho)_A$$

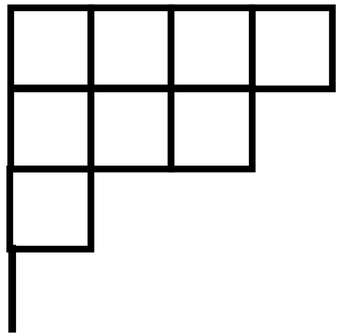
Young diagrams:

consist of n boxes

in r rows (rows possibly empty)

with rows (weakly) decreasing

$\ell(\lambda)$ is # non-empty rows



Young diagram λ

$\lambda_1 = 4$
$\lambda_2 = 3$
$\lambda_3 = 1$
$\lambda_4 = 0$

$\lambda = (4, 3, 1, 0)$
$n = 8$
$r = 4$
$\ell(\lambda) = 3$

$\mathbf{E}[\ell(\lambda)] \leq \min\{r, \sqrt{n}\}$ and thus $\frac{r}{n^2} \cdot \mathrm{SWAP}$ is worse than $\frac{E[\ell(\lambda)]}{n^2} \cdot \mathrm{SWAP}$

$\mathrm{Mix}(A_{GPS})$ without quasi-purification

Debiased Keyl's

The random purification channel

Schur-Weyl duality: $$\rho_A^{\otimes n} = \sum_{\lambda \vdash n} |\lambda\rangle\langle\lambda|_A \otimes I_A \otimes \nu_\lambda(\rho)_A$$

Random purification channel. Given n copies of d-dimensional rank-r state ρ:

1. Measure in the Schur basis to obtain some λ.

   State collapses to $|\lambda\rangle\langle\lambda|_A \otimes I_A \otimes \nu_\lambda(\rho)_A$.

2. ~~Append B registers to every factor so that the system lives in $\mathbb{C}^d \otimes \mathbb{C}^r$.~~

   Append B registers to every factor so that the system lives in $\mathbb{C}^d \otimes \mathbb{C}^{\ell(\lambda)}$.

Strange… The dimension of your output is decided after you measure.

Type check: resulting state is in symmetric subspace of dimension $d \cdot \ell(\lambda)$.

Still makes sense to run pure-state tomography.

Conclusion

Conclusion

This talk: Thorough study of the use of purification in mixed state tomography.

Recover essentially all known results on mixed state tomography.

Establish that the random purification channel is the object to study for state tomography.

Since then, the random purification channel has been…

- characterized analytically and applied to quantum Shannon theory [GML25],
- applied to channel tomography [MB25, CYZ25],
- extended to a random dilation super-channel [GMZFL25, YNM25],
- applied to fermions and bosons [WW25, MGCFL25],
- …

Future directions

1. More applications of the random purification channel?
2. Efficient implementation for Hayashi's measurement.
3. Debiased Keyl's algorithm.

   Related to purifications?

Felix Leditzky 🏳️‍🌈🏳️‍⚧️ @felixled.bsky.social · 2mo

shut up and purify

arxiv.org/abs/2511.15806

Mixed state tomography reduces to pure state tomography

A longstanding belief in quantum tomography is that estimating a mixed state is far harder than estimating a pure state. This is borne out in the mathematic...

arxiv.org