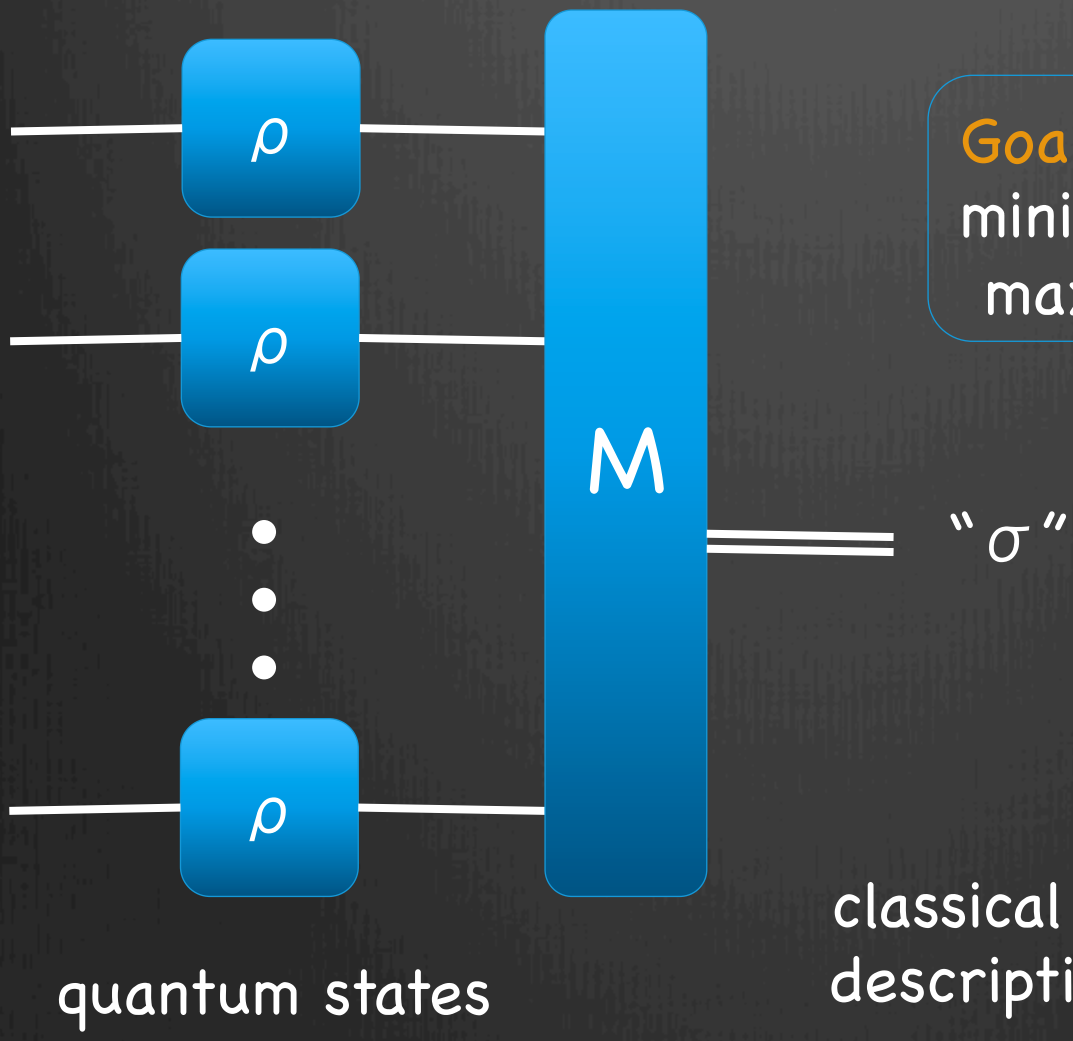
Sample-optimal tomography of quantum states

Jeongwan Haah (MIT)
Aram Harrow (MIT)
Zhengfeng Ji (ISCAS/Waterloo)
Xiaodi Wu (MIT -> Oregon)
Nengkun Yu (Guelph/Waterloo)

QIPC Leeds
2015.9.28

arXiv:1508.01797

state tomography



Goal:
minimize loss, i.e.
$\max_\rho E_\sigma \, \mathrm{dist}(\rho, \sigma)$

how many copies?

Distance measures

Trace distance: $\varepsilon = \|\rho - \sigma\|_1 / 2$

Infidelity: $\delta = 1 - F(\rho, \sigma)$

$= 1 - \|\rho^{1/2} \sigma^{1/2}\|_1$

$\varepsilon^2 \leq \delta \leq \varepsilon$

States

d dimensions

Assume rank $\leq$r.

$n = f(\varepsilon, d, r)$

$n = g(\delta, d, r)$

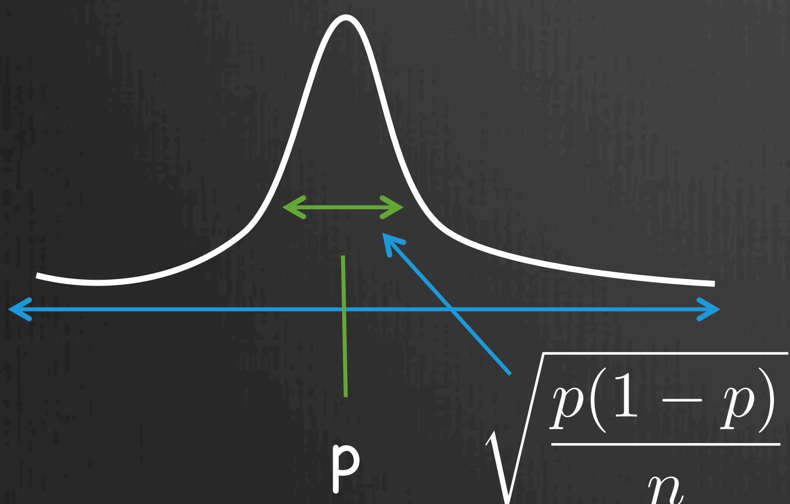
copies necessary/sufficient.

How does n scale with d, r, ε, δ?

boundary case: d=2

suppose $\rho = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}$

distribution of q



Estimation protocol

- Measure in {|0〉, |1〉} basis
- Let q := (#0's) / n.

output

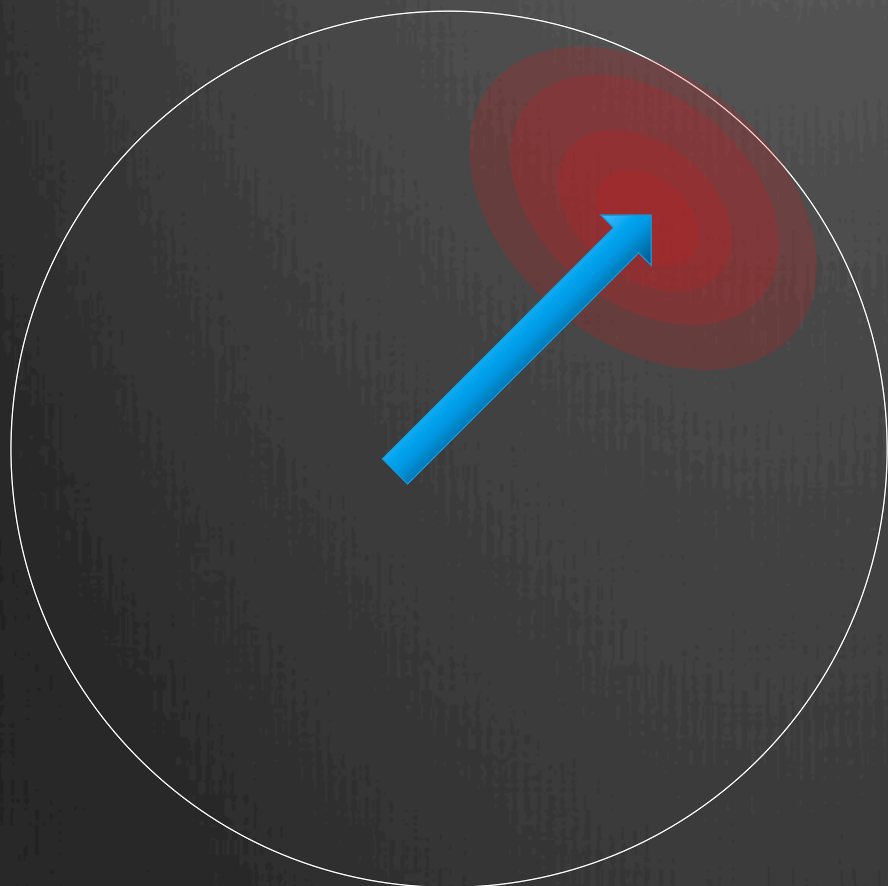
$$\sigma = \begin{pmatrix} q & 0 \\ 0 & 1-q \end{pmatrix}$$

$$\epsilon = |p - q| \sim \sqrt{\frac{p(1-p)}{n}}$$

$$\delta = 1 - F(p, q) \sim \frac{1}{8n}$$

n ~ 1 / ε^2 ~ 1 / δ

Local Asymptotic Normality



Bloch ball

Kahn-Guta
0804.3876

ρ -dependent
covariance matrix

implies optimal
$n \approx f(d) / \delta$
$= g(d) / \varepsilon^2$

for unknown f, g

boundary case: constant error

Intuition

ρ has $d^2 - 1$ real parameters $\rightarrow$ $n \sim d^2$

bounded rank: $\approx rd$ parameters $\rightarrow$ $n \sim rd$

# of copies $\sim$ # of parameters?

plausible, but not a proof.

boundary case: r=1

Symmetry determines optimal measurement

$$M_\varphi = \binom{d+n-1}{n} (|\varphi\rangle\langle\varphi|)^{\otimes n} \mathrm{d}\varphi$$

Fidelity and Trace distance are equivalent: $F^2 + \varepsilon^2 = 1$

Explicit formula for all moments of F

$$\mathbb{E}[F^k] = \frac{\binom{d+n-1}{n}}{\binom{d+n+k-1}{n+k}}$$

[Chiribella, 1010.1875]

$$n \sim \frac{d}{\delta} \sim \frac{d}{\epsilon^2}$$

our results

$$\frac{rd}{\delta} \leq \frac{rd}{\epsilon^2} \lesssim n \lesssim \frac{rd}{\delta} \log\left(\frac{d}{\delta}\right) \leq \frac{rd}{\epsilon^2} \log\left(\frac{d}{\epsilon}\right)$$

related work [O'Donnell-Wright, 1508.01907]

$$n \lesssim \frac{d}{\gamma^2} \leq \frac{rd}{\epsilon^2} \qquad \gamma = \mathbb{E}[\|\rho - \sigma\|_2]$$

product measurements
[Kueng, Rauhut, Terstiege 1410.6913]

$$n \lesssim \frac{rd}{\gamma^2} \leq \frac{r^2 d}{\epsilon^2}$$

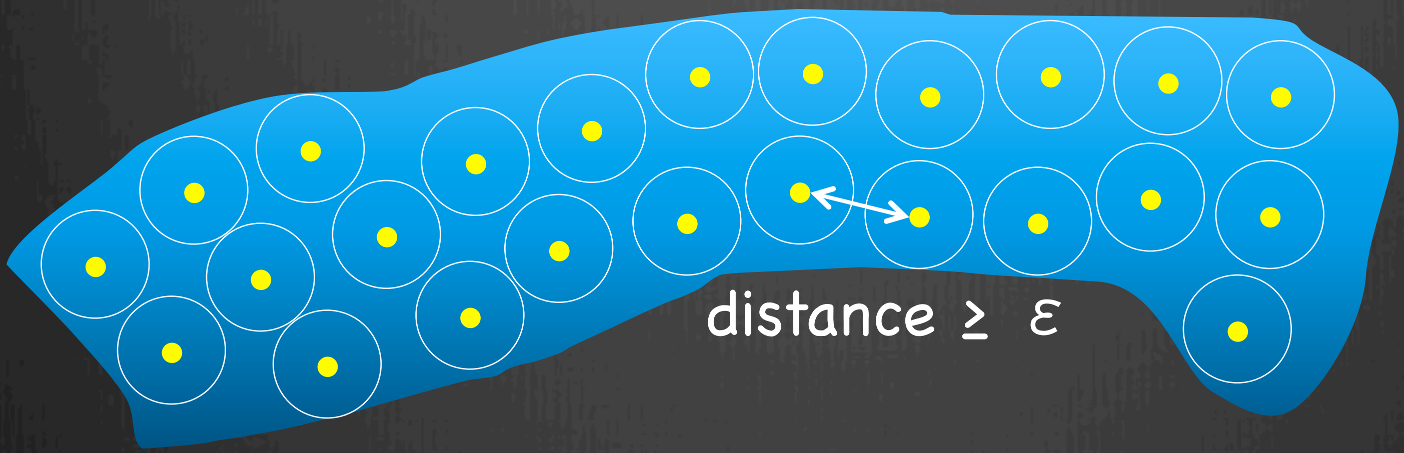
ongoing work
(speculative)

$$n \gtrsim \frac{r^2 d}{\epsilon^2}$$

product measurements

lower bound

rank-r d-dim states form a manifold of dimension ≈rd



ε =0.1 packing net has size exp(rd) [Szarek '81]

→ state estimation can transmit O(rd) bits

Lower bound (for r=d)

An ensemble $\{p_i, \sigma_i\}$ can transmit at most
$\chi = S(\Sigma_i p_i \sigma_i) - \Sigma_i p_i S(\sigma_i)$ bits per copy

[Holevo '73]

Given an ε-net $\rho_1, \ldots, \rho_M$, choose $p_i = 1/M$, $\sigma_i = \rho_i^{\otimes n}$

Choose an $\varepsilon/10$-net of states of the form

$$\rho_i = U_i \begin{pmatrix} \frac{1+\epsilon}{d} & & & \\ & \ddots & & \\ & & \frac{1-\epsilon}{d} & \\ & & & \ddots \end{pmatrix} U_i^\dagger$$

$S(\rho_i) = \log(d) - O(\varepsilon^2)$

$\chi \leq O(n\varepsilon^2)$

$\chi \geq O(\log M) \approx d^2$
(from last slide)

$$\Rightarrow \quad n \gtrsim \frac{d^2}{\epsilon^2}$$

Upper bound inspiration

1. Use symmetry,
   cf. spectrum estimation [Keyl-Werner '01]
   and rank-1 case [Holevo '79]

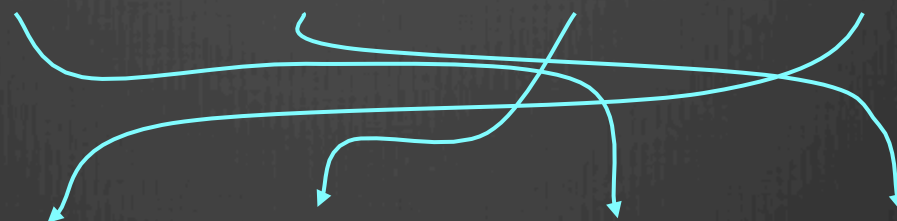
2. Use pretty-good measurement (PGM)
   [Belavkin '75] [Hausladen-Wootters '94]

symmetries of $(\mathbb{C}^d)^{\otimes n}$

$U \in \mathcal{U}_d \to U \otimes U \otimes U \otimes U$

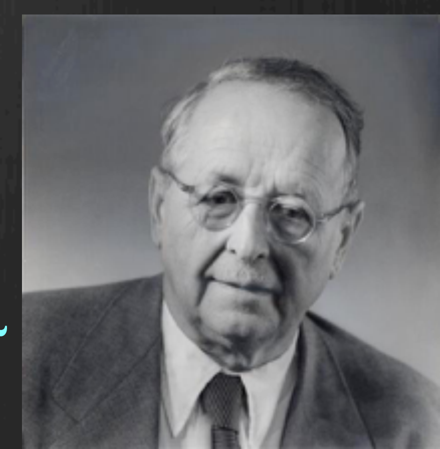
$(\mathbb{C}^d)^{\otimes 4} = \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^d$

$(1324) \in \mathcal{S}_4 \to$



Schur-Weyl duality

$$(\mathbb{C}^d)^{\otimes n} \cong \bigoplus_{\lambda \in \mathrm{Par}(n,d)} Q_\lambda^d \otimes \mathcal{P}_\lambda$$



spectrum estimation

$\rho^{\otimes n} = \oplus_\lambda \, q_\lambda(\rho) \otimes I_{m_\lambda}$

q_λ irrep of GL(d)

- For d=2, λ analogous to J (total angular momentum).
- In general, $\lambda \approx \text{spec}(\rho)$
- Measuring λ causes no disturbance.

Thm: [Keyl-Werner, quant-ph/0102027]
$m_\lambda \, \text{tr} \, q_\lambda(\rho) \leq \exp(-n \, D(\lambda \,||\, \text{spec}(\rho)) \, n^{d^2}$

$n \leq O(d^2 \log(d/\varepsilon) / \varepsilon^2)$ for spectrum estimation

substantial improvements by O'Donnell-Wright, 1501.05028

pretty-good measurement

[Belavkin '75] [Hausladen-Wootters '94]

Given an ensemble $\{p_i, \sigma_i\}$, define

$M_i = \sigma^{-1/2} p_i \sigma_i \sigma^{-1/2}$ with $\sigma = \Sigma_i p_i \sigma_i$

Classical analogue

Given underlying distribution $p(i)$, and observed $j \sim p(j|i)$, guess i' with probability $p(i'|j)$ using Bayes' rule.

Thm: [Barnum-Knill, quant-ph/0004088]

Pr[PGM correct] ≥ Pr[optimal measurement is correct]^2

Thm: [Harrow-Winter, quant-ph/0606131]

Given a set of M states with pairwise infidelity $\geq \delta$,

PGM requires $\leq O(\log(M)/\delta)$ copies to distinguish w.h.p.

putting it together

1. First estimate spectrum using Keyl-Werner. Measurement yields estimate λ.

2. Do PGM with $\{\sigma = U \lambda U^\dagger : U \text{ uniform}\}$

lemma: $m_\lambda^2 \operatorname{tr} q_\lambda (U \lambda U^\dagger \rho) \leq F(\rho, U \lambda U^\dagger)^{2n} n^{rd}$

...a little more algebra...

thm: $\Pr[\text{guess } \sigma \mid \rho] \leq F(\rho, \sigma)^{2n} n^{O(rd)}$

Proof. Consider a positive semi-definite matrix X and a number $k \geq 0$. The largest term in the Schur polynomial $s_\lambda(X^k)$ at eigenvalues $x_1 \geq \cdots \geq x_d \geq 0$ of X is

$$x_1^{k\lambda_1} \cdots x_d^{k\lambda_d} = e^{-nkH(\bar{\lambda})} e^{-nkD(\bar{\lambda}\|\bar{x})} (\operatorname{tr} X)^{kn}$$

where $\bar{x} = (x_1, \ldots, x_d)/\operatorname{tr}(X)$, and $D(p\|q) = \sum_i p_i \ln(p_i/q_i)$ is the relative entropy. This is because majorization implies that

$$\max_{\nu \prec \lambda} x^\nu = x^\lambda,$$

i.e. the maximum is attained by putting the largest number x_1 with the largest possible exponent $\nu_1 = \lambda_1$ and the second largest x_2 with $\nu_2 = \lambda_2$ and so on, subject to the majorization condition $\nu \prec \lambda$.

It follows that

$$s_\lambda(X^k) \leq \dim \mathcal{Q}_\lambda \cdot e^{-nkH(\bar{\lambda})} e^{-nkD(\bar{\lambda}\|\bar{x})} (\operatorname{tr} X)^{kn}. \quad (7)$$

Now, we set $X = \sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}$ and observe $s_\lambda(\rho\sigma) = s_\lambda(X^2)$. Using the fact that $D(\bar{\lambda}\|\bar{x})$ is always non-negative and $= +\infty$ when the rank of $\bar{\lambda}$ is larger than that of $\bar{x}$, we arrive at Eq. (5) □

things we don't know

- Efficiency? Not even known for pure states.
- Process tomography
- Other prior distributions / assumptions about ρ
- Adaptive measurements
- Continuous-variable tomography