University of Colorado **Boulder**

Optimization for statistical estimators: Applications to quantum fidelity estimation

Stephen Becker (CU Boulder, Applied Math)

Conference on the Mathematics of Complex Data

KTH Royal Institute of Technology, Stockholm. June 13-16 2022

Joint work with Akshay Seshadri, Martin Ringbauer, Rainer Blatt, Thomas Monz

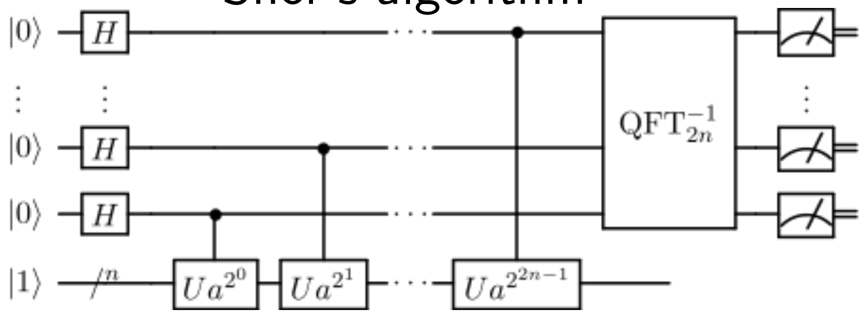
- Versatile fidelity estimation with confidence, https://arxiv.org/abs/2112.07925
- Theory of versatile fidelity estimation with confidence, https://arxiv.org/abs/2112.07947
- code: https://github.com/akshayseshadri/minimax-fidelity-estimation

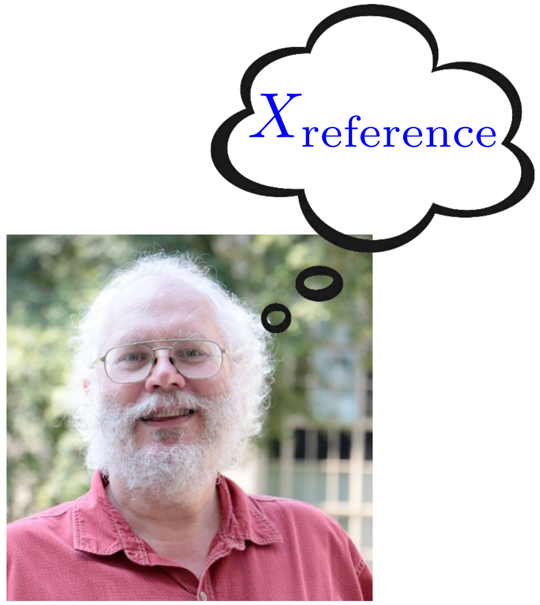
The opinions, findings, and conclusions or recommendations expressed are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

Why Quantum Tomography?

Theory

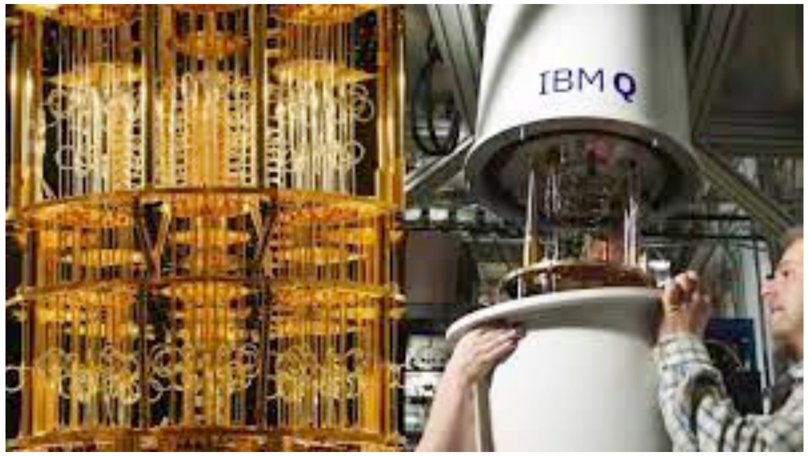
Shor's algorithm





What is X_{actual}?

Reality



X_{actual}

Why?

- diagnosing hardware issues
- verifying if quantum error-correction will apply
- verifying entanglement, etc.

(strictly speaking, full tomography may not be required, as we'll see. Also related to randomized benchmarking)

Quantum Tomography (simplified)

A quantum state

$$d = 2^{\text{number of qubits}}$$

$$X \in \mathcal{X} \stackrel{\text{def}}{=} \{X \in \mathbb{C}^{d\times d} : \underline{X = X^*, X \succeq 0}, \underline{\mathrm{tr}(X) = 1}\}$$

(note: *pure* states correspond to rank 1 matrices)

$$X = |\psi\rangle\langle\psi|$$

column vector row vector

Mathematical structure:

$\{X \in \mathbb{C}^{d\times d} : X = X^*\}$ a real Hilbert space,

$$\langle X, W\rangle = \mathcal{R}e[\mathrm{tr}(X^*W)] = \mathrm{tr}(XW)$$

warning: mix of optimization, stat & physics notation

Capital X means matrix, not random variable

Quantum Tomography (simplified)

A quantum state

$$X \in \mathcal{X} \stackrel{\text{def}}{=} \{X \in \mathbb{C}^{d\times d} : X = X^*, X \succeq 0, \operatorname{tr}(X) = 1\}$$

Collect noisy linear measurements. Informally,

$$\mathbf{y} = \underbrace{\mathcal{A}(X)}_{\mathbf{p}_X} + \mathbf{z}$$

Quantum Tomography (simplified)

A quantum state

$$X \in \mathcal{X} \stackrel{\text{def}}{=} \{X \in \mathbb{C}^{d\times d} : X = X^*, X \succeq 0, \operatorname{tr}(X) = 1\}$$

Collect noisy linear measurements. Informally,

$$\mathbf{y} = \underbrace{\mathcal{A}(X)}_{\mathbf{p}_X} + \mathbf{z}$$

More precisely, use **Born's rule**:

$$p_i = \operatorname{tr}(A_i^* X),\ i = 1, \ldots, N \quad \mathbf{p} = \mathbf{p}_X$$

$$\mathbf{y} \sim \operatorname{multinomial}(\mathbf{p}_X, n)$$

POVM (Pos. Operator-Valued Measure)

[easy to generalize to multiple POVM too]

$$\{A_i\}_{i=1}^N : A_i \succeq 0, \sum_{i=1}^N A_i = I$$

Simple case: state is pure, POVM is observable w/ discrete spectrum

$$A_i = |\lambda_i\rangle\langle\lambda_i|$$

$$X = |\psi\rangle\langle\psi|$$

$$\operatorname{tr}(A_i^* X) = |\langle \lambda_i \mid \psi\rangle|^2 \in [0, 1]$$

Fidelity estimation

Often, our goal is simpler: just estimate the fidelity with a reference state

$$F(X, W) \stackrel{\text{def}}{=} \left(\operatorname{tr}(\sqrt{X^{1/2} W X^{1/2}}\right)^2 \in [0,1] \text{ if } X, W \in \mathcal{X}$$
$$= \operatorname{tr}(XW) \text{ if either } X, W \text{ is pure}$$

... so we just want a linear functional

$g(X) \stackrel{\text{def}}{=} F(X, X_{\text{reference}})$ since almost always $X_{\text{reference}}$ is a pure state.

The quantity we want to know is $F(X_{\text{actual}}, X_{\text{reference}})$

Fidelity estimation

Often, our goal is simpler: just estimate the fidelity with a reference state

$$F(X, W) \stackrel{\text{def}}{=} \left(\operatorname{tr}(\sqrt{X^{1/2} W X^{1/2}}\right)^2 \in [0,1] \text{ if } X, W \in \mathcal{X}$$
$$= \operatorname{tr}(XW) \text{ if either } X, W \text{ is pure}$$

... so we just want a linear functional

$g(X) \stackrel{\text{def}}{=} F(X, X_{\text{reference}})$ since almost always $X_{\text{reference}}$ is a pure state.

Why? Used for diagnosing quantum systems

Goals:

1. As few POVMs and repetitions as possible
2. High accuracy: low bias and low variance
3. Confidence intervals... especially for error correction.

One approach

(or MLE, etc.)

$$\widehat{X} = \operatorname{argmin}_{X \in \mathcal{X}} \|X\|_* \quad \text{s.t.} \quad \mathcal{A}(X) \approx \mathbf{y}$$

Gross, Liu, Flammia, B., Eisert; PRL '10

nuclear norm

then "plug-in" estimator into fidelity:

$F(\widehat{X}, X_{\text{reference}})$ is our estimate of $F(X_{\text{actual}}, X_{\text{reference}})$

One approach

(or MLE, etc.)

$$\widehat{X} = \text{argmin}_{X \in \mathcal{X}} \; \|X\|_* \quad \text{s.t.} \quad \mathcal{A}(X) \approx \mathbf{y}$$

Gross, Liu, Flammia, B., Eisert; PRL '10

then "plug-in" estimator into fidelity:

$F(\widehat{X}, X_{\text{reference}})$ is our estimate of $F(X_{\text{actual}}, X_{\text{reference}})$

This works best if actual state is low-rank

Sometimes state is also sparse... could exploit this prior information too!

... in fact, we expect actual state to be close to reference state, so why not $\widehat{X} = X_{\text{reference}}$?
But then $F(\widehat{X}, X_{\text{reference}}) = 1$ is very biased!

Cutting out the middle-man

Observations/Data

$\mathbf{y} \in \mathbb{N}^N, n$

(or simplify and take $n \to \infty$)

$\mathbf{p}_X \in [0,1]^N$

Number of parameters

$\lesssim N$

Intermediate estimate

(often via optimization)

$\widehat{X}$

usual path

Number of parameters

$d^2 = 2^{2\times \#qubits}$

(complex-valued)

Final Estimate

$\hat{g} = g(\widehat{X})$ Ex: MLE

another path

$\hat{g} = \texttt{algo}(\mathbf{y})$

intermediate estimator (often via optimization); independent of *data*

Number of parameters

1

[Reminder: $n =$ repetitions/shots, $N =$ size of POVM]

Is this possible?

Instead of POVM, take orthonormal basis for $\mathbb{C}^{d\times d}$; e.g., the tensor product of all Paulis

$$\{V_i\}_{i=1}^{N=d^2}$$

Ignore "noise" for now, i.e., take # of repetitions/shots $n \to \infty$

(each basis element induces its own POVM, so this is a multi-POVM setting)

Take measurements using this rule: $$y = \left\{ \mathrm{tr}(V_i X)/\mathrm{tr}(V_i X_{\mathrm{ref}}) \quad \text{w.p. } \mathrm{tr}(V_i X_{\mathrm{ref}})^2 \right.$$

(i.e., this tells us which i to measure)

observation

computable

Is this possible? **yes**

Instead of POVM, take orthonormal basis for $\mathbb{C}^{d\times d}$; e.g., the tensor product of all Paulis

$$\{V_i\}_{i=1}^{N=d^2}$$

Ignore "noise" for now, i.e., take # of repetitions/shots $n \to \infty$ (each basis element induces its own POVM, so this is a multi-POVM setting)

Take measurements using this rule: $y = \left\{ \mathrm{tr}(V_i X)/\mathrm{tr}(V_i X_{\mathrm{ref}}) \quad \text{w.p. } \mathrm{tr}(V_i X_{\mathrm{ref}})^2 \right.$

then:

$$\begin{aligned}\mathbb{E}[y] &= \sum_{i=1}^{d^2} \mathrm{tr}(V_i X)\mathrm{tr}(V_i X_{\mathrm{ref}}) \\ &= \mathrm{tr}(X X_{\mathrm{ref}}) \quad \text{(inner prod. independent of o.n.b.)} \\ &= g(X) \checkmark\end{aligned}$$

Take repeated measurements and use concentration inequalities:

PRL **106,** 230501 (2011) PHYSICAL REVIEW LETTERS

Direct Fidelity Estimation from Few Pauli Measurements

Steven T. Flammia[1] and Yi-Kai Liu[2]

PRL **107,** 210404 (2011) PHYSICAL REVIEW LETTERS 18

Practical Characterization of Quantum Devices without Tomography

Marcus P. da Silva,[1,2] Olivier Landon-Cardinal,[2] and David Poulin[2]

Other approaches... with a focus on confidence intervals

Maximum Likelihood Estimation (MLE)

negative log-likelihood

$$\mathcal{L}(X) = -\log \mathbb{P}[\mathbf{Y} = \mathbf{y} \mid X]$$

$$\widehat{X}_{\mathrm{MLE}} \stackrel{\mathrm{def}}{=} \operatorname{argmin}_{X \in \mathcal{X}} \mathcal{L}(X)$$

Other approaches... with a focus on confidence intervals

Maximum Likelihood Estimation (MLE)

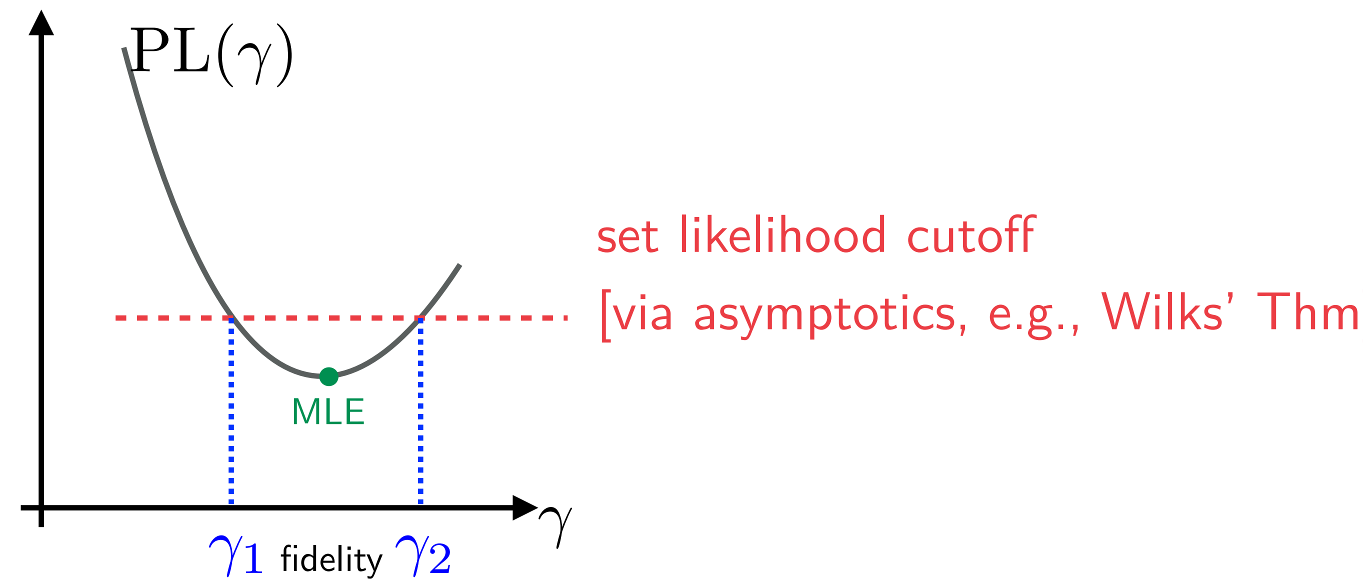
negative log-likelihood

$$\mathcal{L}(X) = -\log \mathbb{P}[\mathbf{Y} = \mathbf{y} \mid X]$$

$$\widehat{X}_{\text{MLE}} \stackrel{\text{def}}{=} \text{argmin}_{X \in \mathcal{X}} \mathcal{L}(X)$$

Better: **Profile Likelihood (PL)**

$$\text{PL}(\gamma) \stackrel{\text{def}}{=} \min_{\substack{X \in \mathcal{X} \\ g(X) = \gamma}} \mathcal{L}(X)$$



... get confidence interval: $g \in [\gamma_1, \gamma_2]$ with some probability

Other approaches... with a focus on confidence intervals

Maximum Likelihood Estimation (MLE)

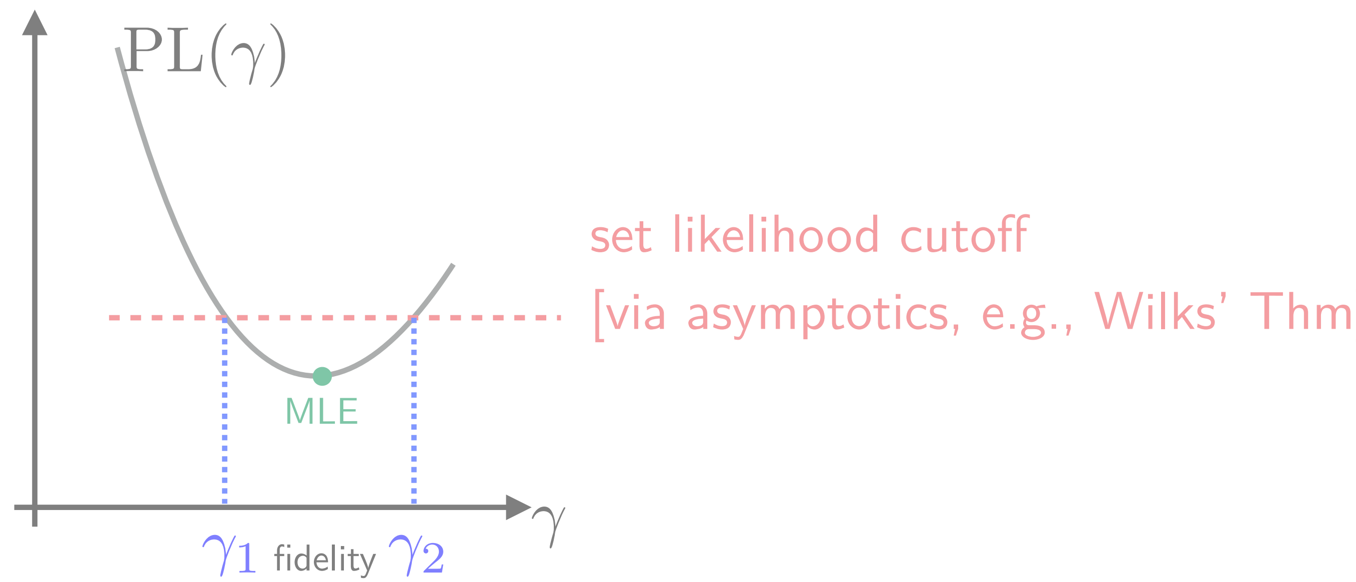
negative log-likelihood

$$\mathcal{L}(X) = -\log \mathbb{P}[\mathbf{Y} = \mathbf{y} \mid X]$$

$$\widehat{X}_{\mathrm{MLE}} \stackrel{\mathrm{def}}{=} \operatorname{argmin}_{X \in \mathcal{X}} \mathcal{L}(X)$$

Better: **Profile Likelihood (PL)**

$$\mathrm{PL}(\gamma) \stackrel{\mathrm{def}}{=} \min_{\substack{X \in \mathcal{X} \\ g(X)=\gamma}} \mathcal{L}(X)$$



... get confidence interval: $g \in [\gamma_1, \gamma_2]$ with some probability

Other approaches:

- **SDP / matrix completion**

$$\gamma_1 = \min_{\substack{X \in \mathcal{X} \\ \|\mathcal{A}(X)-\mathbf{y}\| \le \epsilon}} g(X), \quad \gamma_2 = \max_{\substack{X \in \mathcal{X} \\ \|\mathcal{A}(X)-\mathbf{y}\| \le \epsilon}} g(X)$$

(same drawback as PL: how to choose parameter?)

Other approaches... with a focus on confidence intervals

Maximum Likelihood Estimation (MLE)

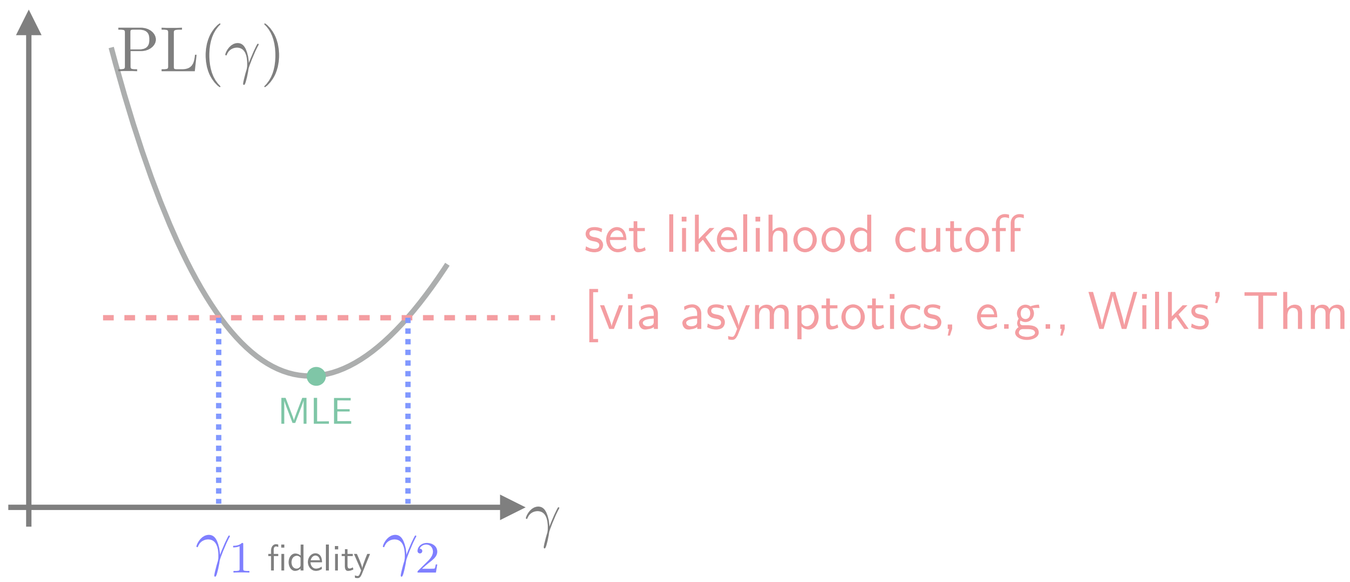
negative log-likelihood

$$\mathcal{L}(X) = -\log \mathbb{P}[\mathbf{Y} = \mathbf{y} \mid X]$$

$$\widehat{X}_{\mathrm{MLE}} \stackrel{\mathrm{def}}{=} \mathrm{argmin}_{X \in \mathcal{X}}\, \mathcal{L}(X)$$

Better: **Profile Likelihood (PL)**

$$\mathrm{PL}(\gamma) \stackrel{\mathrm{def}}{=} \min_{\substack{X \in \mathcal{X} \\ g(X) = \gamma}} \mathcal{L}(X)$$



... get confidence interval: $g \in [\gamma_1, \gamma_2]$ with some probability

Other approaches:

- **SDP / matrix completion**
- **2 step: Least squares, then project**

Guţă, Kahn, Kueng, Tropp; J. Phys. A: Math. Theor. 2020

Shortcomings of other approaches

	MLE	Profile likelihood	SDP	Direct Fidelity Estimation	(2-step) Projected Least-Squares	**Proposed minimax estimate**
Rigorous?	✗	✗	✗	✓ * rigorous version	✓	✓
Doesn't assume $n \to \infty$	✓	✓	✗	isn't tight ✓ *	✗	✓
Avoids unknown parameters	✓	✗	✗	✓	✓	✓
Computable before seeing data	✗	✗	✗	✓	✗	✓
Applies to any measurement setting	✓	✓	✓	✗	✓	✓
Computational speed	Bad	Bad	Bad	Great	Ok	Offline

some methods solvable by hand under certain settings

Minimax approach

The Annals of Statistics
2009, Vol. 37, No. 5A, 2278–2300
DOI: 10.1214/08-AOS654
© Institute of Mathematical Statistics, 2009

NONPARAMETRIC ESTIMATION BY CONVEX PROGRAMMING

BY ANATOLI B. JUDITSKY AND ARKADI S. NEMIROVSKI[1]

Université Grenoble I and Georgia Institute of Technology

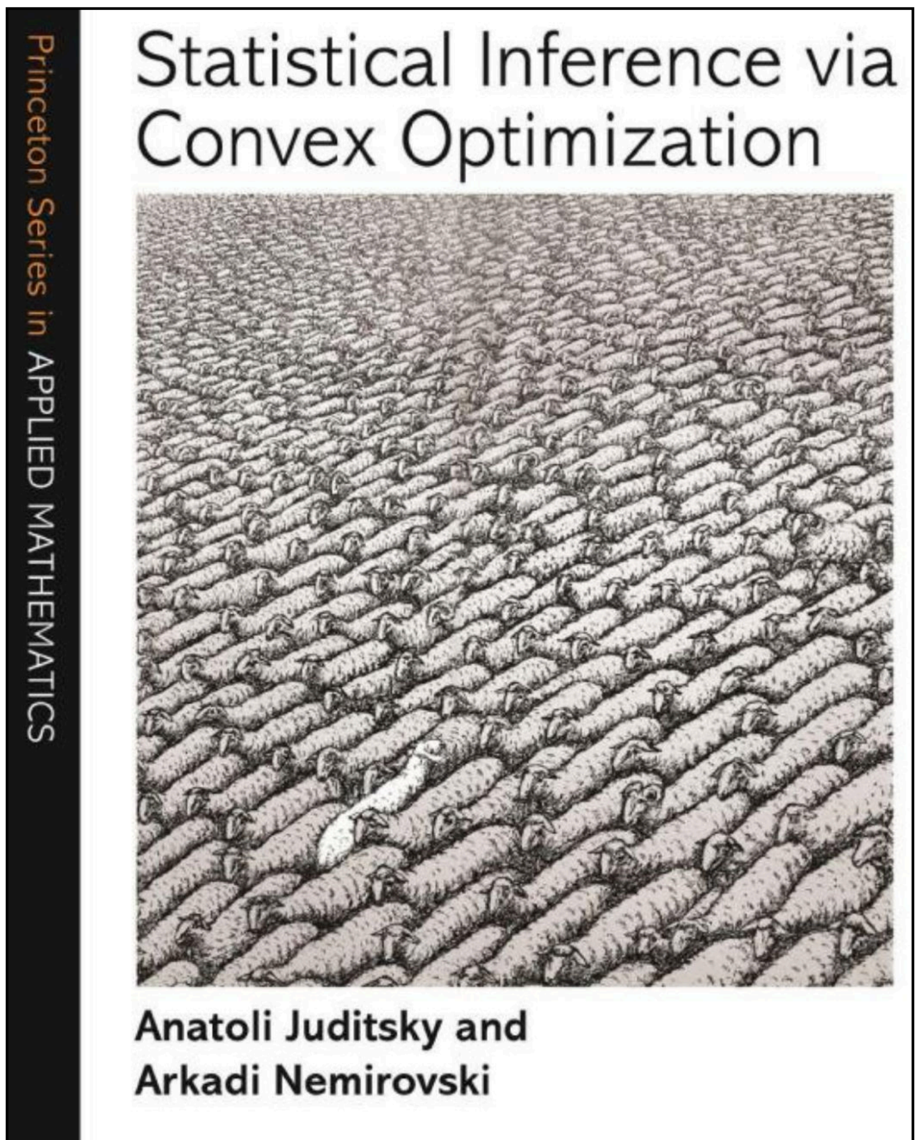
The problem we concentrate on is as follows: given (1) a convex compact set X in $\mathbb{R}^n$, an affine mapping $x \mapsto A(x)$, a parametric family $\{p_\mu(\cdot)\}$ of probability densities and (2) N i.i.d. observations of the random variable ω, distributed with the density $p_{A(x)}(\cdot)$ for some (unknown) $x \in X$, estimate the value $g^T x$ of a given linear form at x.

For several families $\{p_\mu(\cdot)\}$ with no additional assumptions on X and A, we develop computationally efficient estimation routines which are minimax optimal, within an absolute constant factor. We then apply these routines to recovering x itself in the Euclidean norm.

Used in our paper

What I'll present today

to convey main idea



ch. 3

The Annals of Statistics
1994, Vol. 22, No. 1, 238–270

STATISTICAL ESTIMATION AND OPTIMAL RECOVERY[1]

BY DAVID L. DONOHO

University of California, Berkeley

New formulas are given for the minimax linear risk in estimating a linear functional of an unknown object from indirect data contaminated with random Gaussian noise. The formulas cover a variety of loss functions and do not require the symmetry of the convex a priori class. It is shown that affine minimax rules are within a few percent of minimax even among nonlinear rules, for a variety of loss functions. It is also shown that difficulty of estimation is measured by the modulus of continuity of the functional to be estimated.

The method of proof exposes a correspondence between minimax affine estimates in the statistical estimation problem and optimal algorithms in the theory of optimal recovery.

See also ch 7.4 in Boyd & Vandenberghe
for more interesting applications
(designing Chernoff bounds)

Setup

$$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{z}$$
$$\mathbf{Y} = \mathcal{A}(\mathbf{x}) + \mathbf{Z} \qquad \mathbf{Z} \sim \mathcal{N}(0, \sigma^2 I)$$

$\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^m$ is linear

(if noise not iid, then whiten and do change-of-variables)

Prior knowledge: $\mathbf{x} \in \mathcal{X}$ (always convex; often compact)

Setup

$$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{z}$$
$$\mathbf{Y} = \mathcal{A}(\mathbf{x}) + \mathbf{Z} \qquad \mathbf{Z} \sim \mathcal{N}(0, \sigma^2 I)$$

$\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^m$ is linear

(if noise not iid, then whiten and do change-of-variables)

Prior knowledge: $\mathbf{x} \in \mathcal{X}$ (always convex; often compact)

Goal: design an **estimator** $\hat{g}$ with small risk $r(\hat{g}, \mathbf{x})$

Focus on this for exposition

Ex.

$$r(\hat{g}, \mathbf{x}) = \mathbb{E}[(\hat{g}(\mathbf{Y}) - g(\mathbf{x}))^2]$$

$$r(\hat{g}, \mathbf{x}) = \mathbb{E}[|\hat{g}(\mathbf{Y}) - g(\mathbf{x})|]$$

$$r_\alpha(\hat{g}, \mathbf{x}) = \inf \delta \text{ s.t. } \mathbb{P}[|\hat{g}(\mathbf{Y}) - g(\mathbf{x})| \le \delta] \ge 1 - \alpha$$

i.e., confidence intervals. Use this for our quantum setting.

(as before, g is a linear functional)

Setup

$$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{z}$$
$$\mathbf{Y} = \mathcal{A}(\mathbf{x}) + \mathbf{Z} \quad \mathbf{Z} \sim \mathcal{N}(0, \sigma^2 I)$$

$\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^m$ is linear

(if noise not iid, then whiten and do change-of-variables)

Prior knowledge: $\mathbf{x} \in \mathcal{X}$ (always convex; often compact)

Goal: design an **estimator** $\hat{g}$ with small risk $r(\hat{g}, \mathbf{x})$

Ex.

$$r(\hat{g}, \mathbf{x}) = \mathbb{E}[(\hat{g}(\mathbf{Y}) - g(\mathbf{x}))^2]$$

$$r(\hat{g}, \mathbf{x}) = \mathbb{E}[|\hat{g}(\mathbf{Y}) - g(\mathbf{x})|]$$

$$r_\alpha(\hat{g}, \mathbf{x}) = \inf \delta \text{ s.t. } \mathbb{P}[|\hat{g}(\mathbf{Y}) - g(\mathbf{x})| \le \delta] \ge 1 - \alpha$$

In particular, look at minimax risk

$$R^\star = \inf_{\hat{g}} \sup_{\mathbf{x} \in \mathcal{X}} r(\hat{g}, \mathbf{x})$$

and minimax **affine** risk

$$R^\star_{\text{affine}} = \inf_{\hat{g} \text{ affine}} \sup_{\mathbf{x} \in \mathcal{X}} r(\hat{g}, \mathbf{x})$$

Univariate case

multivariate

$$\mathbf{Y} = \mathcal{A}(\mathbf{x}) + \mathbf{Z}$$
$$\mathbf{Z} \sim \mathcal{N}(0, \sigma^2 I)$$
$$\mathbf{x} \in \mathcal{X}$$

$$r(\hat{g}, \mathbf{x}) = \mathbb{E}[(\hat{g}(\mathbf{y}) - g(\mathbf{x}))^2]$$

$$R^\star = \inf_{\hat{g}} \sup_{\mathbf{x} \in \mathcal{X}} r(\hat{g}, \mathbf{x})$$

$$R^\star_{\text{affine}} = \inf_{\hat{g}\ \text{affine}} \sup_{\mathbf{x} \in \mathcal{X}} r(\hat{g}, \mathbf{x})$$

[linear operators and closed convex sets in 1D are very simple!]

univariate

$$Y = x + Z$$
$$Z \sim \mathcal{N}(0, \sigma^2)$$
$$x \in [-\tau, \tau]$$

$$\rho^\star(\tau) = \inf_{\hat{g}} \max_{x \in [-\tau,\tau]} \mathbb{E}[(\hat{g}(Y) - x)^2]$$ (no closed form)

$$\rho^\star_{\text{affine}}(\tau) = \min_{c,d} \max_{x \in [-\tau,\tau]} \mathbb{E}[(cY + d - x)^2] = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}$$

$$d^\star = 0, c^\star = \frac{\tau^2}{\sigma^2\tau^2}$$

(closed form) ✔

✔ Theorem (Feldman, Brown '89; Donoho et al. '90)

$$\rho^\star(\tau) \le \rho^\star_{\text{affine}}(\tau) \le \frac{5}{4}\rho^\star(\tau)$$

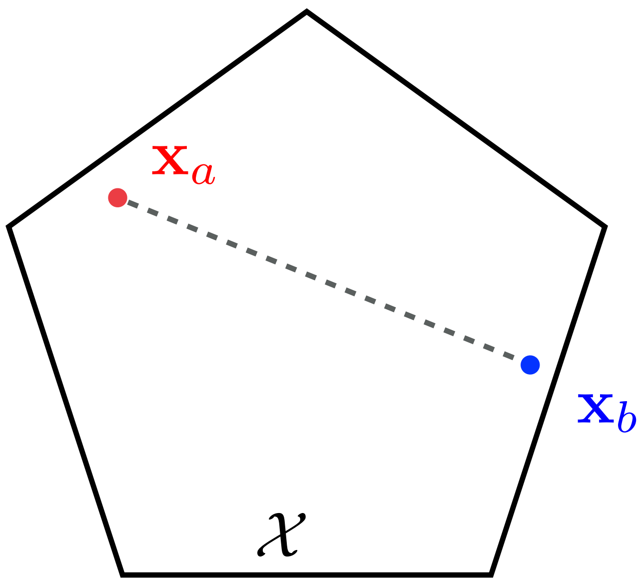
Univariate case is well-understood, and little penalty for restricting to affine estimators

Reducing multivariate to univariate

$$R^{\star}_{\text{affine}} = \inf_{\hat{g} \text{ affine}} \sup_{\mathbf{x} \in \mathcal{X}} r(\hat{g}, \mathbf{x})$$

Reducing multivariate to univariate

$$
\begin{aligned}
R^\star_{\text{affine}} &= \inf_{\hat{g}\ \text{affine}} \ \sup_{\mathbf{x}\in\mathcal{X}} \ r(\hat{g},\mathbf{x}) \\
&= \inf_{\hat{g}\ \text{affine}} \ \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \ \sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} \ r(\hat{g},\mathbf{x})
\end{aligned}
$$



Reducing multivariate to univariate

$$
\begin{aligned}
R^\star_{\text{affine}} &= \inf_{\hat{g} \text{ affine}} \sup_{\mathbf{x}\in\mathcal{X}} r(\hat{g},\mathbf{x}) \\
&= \inf_{\hat{g} \text{ affine}} \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat{g},\mathbf{x}) \\
&= \inf_{\hat{g} \text{ affine}} \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \left(\sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat{g},\mathbf{x}) \right)
\end{aligned}
$$

Reducing multivariate to univariate

$$
\begin{aligned}
R^\star_{\text{affine}} &= \inf_{\hat{g}\text{ affine}} \sup_{\mathbf{x}\in\mathcal{X}} r(\hat{g},\mathbf{x}) \\
&= \inf_{\hat{g}\text{ affine}} \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat{g},\mathbf{x}) \\
&= \inf_{\hat{g}\text{ affine}} \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \left(\sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat{g},\mathbf{x}) \right) \\
&= \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \inf_{\hat{g}\text{ affine}} \left(\sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat{g},\mathbf{x}) \right)
\end{aligned}
$$

via a saddle point theorem
(like strong duality)

($\geq$ always true via weak duality)

Reducing multivariate to univariate

$$\begin{aligned}
R^\star_{\text{affine}} &= \inf_{\hat g \text{ affine}} \sup_{\mathbf{x}\in\mathcal{X}} r(\hat g, \mathbf{x}) \\
&= \inf_{\hat g \text{ affine}} \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat g, \mathbf{x}) \\
&= \inf_{\hat g \text{ affine}} \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \left(\sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat g, \mathbf{x}) \right) \\
&= \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \inf_{\hat g \text{ affine}} \left(\sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat g, \mathbf{x}) \right) \\
&= \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \left(\frac{g(\mathbf{x}_a)-g(\mathbf{x}_b)}{\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)} \right)^2 \rho^\star_{\text{affine}}(\underbrace{\frac{1}{2}\|\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)\|}_{\tau})
\end{aligned}$$

Inner problem is 1D

Change of variables $\quad \mathbf{x} \in \{\alpha\mathbf{x}_a + (1-\alpha)\mathbf{x}_b \mid \alpha\in[0,1]\}$

$$\mathbf{x}_0 = \frac{\mathbf{x}_a+\mathbf{x}_b}{2} \qquad \tau = \|\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)\|/2$$

$$\mathbf{w}_0 = \mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)/\|\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)\|$$

$$x = \langle \mathbf{w}_0, \mathcal{A}(\mathbf{x}-\mathbf{x}_0)\rangle \in [-\tau,\tau]$$

$$Y = \langle \mathbf{w}_0, \mathbf{y}-\mathcal{A}(\mathbf{x}_0)\rangle \sim \mathcal{N}(x,\sigma^2)$$

Reducing multivariate to univariate

$$
\begin{aligned}
R^\star_{\text{affine}} &= \inf_{\hat{g}\text{ affine}} \sup_{\mathbf{x}\in\mathcal{X}} r(\hat{g},\mathbf{x}) \\
&= \inf_{\hat{g}\text{ affine}} \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat{g},\mathbf{x}) \\
&= \inf_{\hat{g}\text{ affine}} \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \left(\sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat{g},\mathbf{x}) \right) \\
&= \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \inf_{\hat{g}\text{ affine}} \left(\sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat{g},\mathbf{x}) \right) \\
&= \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \left(\frac{g(\mathbf{x}_a)-g(\mathbf{x}_b)}{\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)} \right)^2 \rho^\star_{\text{affine}} \underbrace{\left(\frac{1}{2}\|\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)\|\right)}_{\tau} \\
&= \sup_{\epsilon\geq 0} \sup_{\substack{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X} \\ \|\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)\|=\epsilon}} \left(\frac{g(\mathbf{x}_a)-g(\mathbf{x}_b)}{\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)} \right)^2 \rho^\star_{\text{affine}}\left(\frac{1}{2}\|\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)\|\right)
\end{aligned}
$$

Reducing multivariate to univariate

$$R^\star_{\text{affine}} = \inf_{\hat{g}\text{ affine}} \sup_{\mathbf{x}\in\mathcal{X}} r(\hat{g},\mathbf{x})$$

$$\omega(\epsilon) \stackrel{\text{def}}{=} \sup_{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}} \{|g(\mathbf{x}_a)-g(\mathbf{x}_b)| : \|\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)\|\le\epsilon\}$$

Example: $\mathcal{X}$ is ball of diameter D

$$\omega(\epsilon) = \|\mathcal{A}\|\min(\epsilon, D) \qquad \frac{\omega(\epsilon)}{\epsilon} = \|\mathcal{A}\|\min\left(1,\frac{D}{\epsilon}\right)$$

$$= \sup_{\epsilon\ge0} \sup_{\substack{\mathbf{x}_a,\mathbf{x}_b\in\mathcal{X}\\ \|\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)\|=\epsilon}} \left(\frac{g(\mathbf{x}_a)-g(\mathbf{x}_b)}{\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)}\right)^2 \rho^\star_{\text{affine}}\left(\frac{1}{2}\|\mathcal{A}(\mathbf{x}_a-\mathbf{x}_b)\|\right)$$

$$= \sup_{\epsilon\ge0}\left(\frac{\omega(\epsilon)}{\epsilon}\right)^2 \rho^\star_{\text{affine}}(\epsilon/2)$$

Reducing multivariate to univariate

Caveat: I'm not being careful with details

Donoho doesn't present it exactly like this

$$
\begin{aligned}
R^\star_{\text{affine}} &= \inf_{\hat g \text{ affine}} \sup_{\mathbf{x}\in\mathcal{X}} r(\hat g, \mathbf{x}) \\
&= \inf_{\hat g \text{ affine}} \sup_{\mathbf{x}_a, \mathbf{x}_b \in \mathcal{X}} \sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat g, \mathbf{x}) \\
&= \inf_{\hat g \text{ affine}} \sup_{\mathbf{x}_a, \mathbf{x}_b \in \mathcal{X}} \left(\sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat g, \mathbf{x}) \right) \\
&= \sup_{\mathbf{x}_a, \mathbf{x}_b \in \mathcal{X}} \inf_{\hat g \text{ affine}} \left(\sup_{\mathbf{x}\in\overline{\mathbf{x}_a\mathbf{x}_b}} r(\hat g, \mathbf{x}) \right) \\
&= \sup_{\mathbf{x}_a, \mathbf{x}_b \in \mathcal{X}} \left(\frac{g(\mathbf{x}_a) - g(\mathbf{x}_b)}{\mathcal{A}(\mathbf{x}_a - \mathbf{x}_b)} \right)^2 \rho^\star_{\text{affine}} \underbrace{\left(\frac{1}{2}\|\mathcal{A}(\mathbf{x}_a - \mathbf{x}_b)\|\right)}_{\tau} \\
&= \sup_{\epsilon \ge 0} \sup_{\substack{\mathbf{x}_a, \mathbf{x}_b \in \mathcal{X} \\ \|\mathcal{A}(\mathbf{x}_a - \mathbf{x}_b)\| = \epsilon}} \left(\frac{g(\mathbf{x}_a) - g(\mathbf{x}_b)}{\mathcal{A}(\mathbf{x}_a - \mathbf{x}_b)} \right)^2 \rho^\star_{\text{affine}} \left(\frac{1}{2}\|\mathcal{A}(\mathbf{x}_a - \mathbf{x}_b)\|\right) \\
&= \sup_{\epsilon \ge 0} \left(\frac{\omega(\epsilon)}{\epsilon} \right)^2 \rho^\star_{\text{affine}}(\epsilon/2)
\end{aligned}
$$

... a 1D problem, easy to solve!

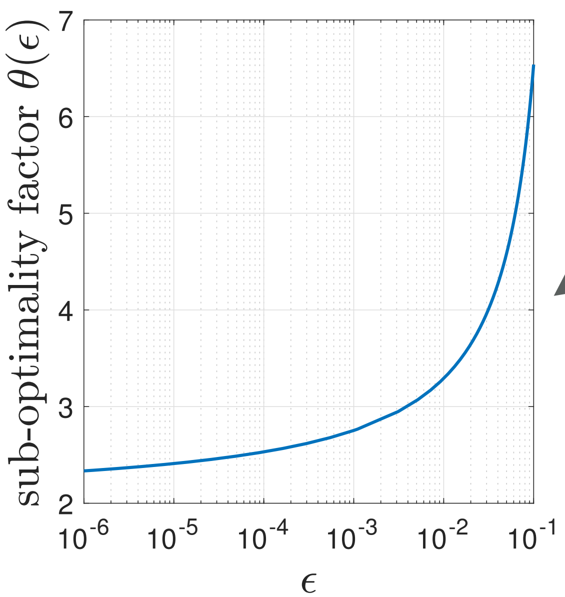
Furthermore:

1. sup=max
2. argmax computable
3. *Affine* sub-optimality carries over from 1D (and invoke Brown-Cohen-Strawderman)

… back to Juditsky and Nemirovski setting

Theorem Solve saddle-point problem to find (optimal) affine estimator and its confidence interval

Theorem Risk of affine estimator is within $\theta(\epsilon)$ of the optimal risk



Differences from Donoho:

- Generalized setting
- Chernoff bound style argument (with scalar "TBD")
    - Turns into a perspective function (still convex)
- Doesn't require Gaussian
- Only for "confidence interval" risk, not MSE risk

$$\cancel{r(\hat{g}, \mathbf{x}) = \mathbb{E}[(\hat{g}(\mathbf{Y}) - g(\mathbf{x}))^2]}$$

$$R^\star_{\text{affine}} = \inf_{\hat{g} \text{ affine}} \sup_{\mathbf{x} \in \mathcal{X}} r(\hat{g}, \mathbf{x}) \qquad \cancel{r(\hat{g}, \mathbf{x}) = \mathbb{E}[|\hat{g}(\mathbf{Y}) - g(\mathbf{x})|]}$$

$$r_\epsilon(\hat{g}, \mathbf{x}) = \inf \delta \text{ s.t. } \mathbb{P}[|\hat{g}(\mathbf{Y}) - g(\mathbf{x})| \le \delta] \ge 1 - \epsilon$$

$\widehat{R}_\star$ denotes the affine minimax risk, e.g., half the width of the confidence interval

Thanks for listening

- "Versatile fidelity estimation with confidence", https://arxiv.org/abs/2112.07925
- "Theory of versatile fidelity estimation with confidence", https://arxiv.org/abs/2112.07947

Extensions

- Optimal design (+ more efficient optimization solvers)
- Quantum *channels*

More details in slides appendix

- how to solve saddle point problem
- sample complexity bounds
- empirical demonstrations of tightness
- applied to real quantum data
- comparisons with other methods

This material is based upon work supported by the National Science Foundation under grant no. 1819251.This work utilized the Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University. The opinions, findings, and conclusions or recommendations expressed are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

We gratefully acknowledge support by the Austrian Science Fund (FWF), through the SFB BeyondC (FWF Project No. F7109) and the Institut fur Quanteninformation GmbH. We also acknowledge funding from the EU H2020-FETFLAG-2018-03 under Grant Agreement no. 820495, by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via US Army Research Office (ARO) grant no. W911NF-16-1-0070 and W911NF-20-1-0007, and the US Air Force Office of Scientific Research (AFOSR) via IOE Grant No. FA9550-19-1-7044 LASCEM. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 840450.

Saddle point problem

Here's the particular saddle point problem to solve:

$$\inf_{\alpha>0,\phi} \sup_{X_a, X_b \in \mathcal{X}} \Phi(X_a, X_b;\ \phi, \alpha) \stackrel{\text{def}}{=} g(X_a) - g(X_b) + 2\alpha \ln(2/\epsilon) + \alpha \cdot n \cdot h(X_a, X_b; \phi/\alpha)$$

$$h(X_a, X_b; \phi) \stackrel{\text{def}}{=} \ln\left(\sum_{i=1}^{N} \exp\left(-\phi_i\right) \operatorname{tr}(A_i X_a)\right) + \ln\left(\sum_{i=1}^{N} \exp\left(\phi_i\right) \operatorname{tr}(A_i X_b)\right)$$

(also generalized to more than 1 POVM)

To solve:

1. For fixed $\alpha > 0, X_a, X_b \in \mathcal{X}$, *inf* over ϕ has closed form expression (and reduces to Hellinger affinity). So make this inner part
2. For fixed $\alpha > 0$, solve for $X_a, X_b \in \mathcal{X}$ via Nesterov's second method
3. Minimize $\alpha > 0$ using any reasonable 1D method, e.g., scipy's `minimize_scalar`

Physics theorems

from Seshadri, Ringbauer, Monz, B. 2021; arXiv 2112.07947

Theorem For a $1-\epsilon$ confidence interval of width $2\widehat{R}_\star$, must take $n \gtrsim \dfrac{\ln(2/\epsilon)}{2\widehat{R}_\star^2}$ shots/repetitions.

Theorem If target state is a stabilizer state, suffices to take $n \approx 4\dfrac{\ln(2/\epsilon)}{2\widehat{R}_\star^2}$ shots (using special Pauli-based POVM)

Theorem For any n-qubit pure target state, suffices to take $n \approx 2^{n+2}\dfrac{\ln(2/\epsilon)}{2\widehat{R}_\star^2}$ shots (using special Pauli-based POVM).

Theorem Scheme is robust against noise/imperfections.
(Due to linearity)

Binomial example

POVM:

$$A_0 = |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$A_1 = |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Let $N = 2, n = 100$

y_0 is $\#$ of times (of 100) we measure $|0\rangle$

y_1 is $\#$ of times (of 100) we measure $|1\rangle$

$X_{\text{reference}} = |1\rangle\langle 1|$

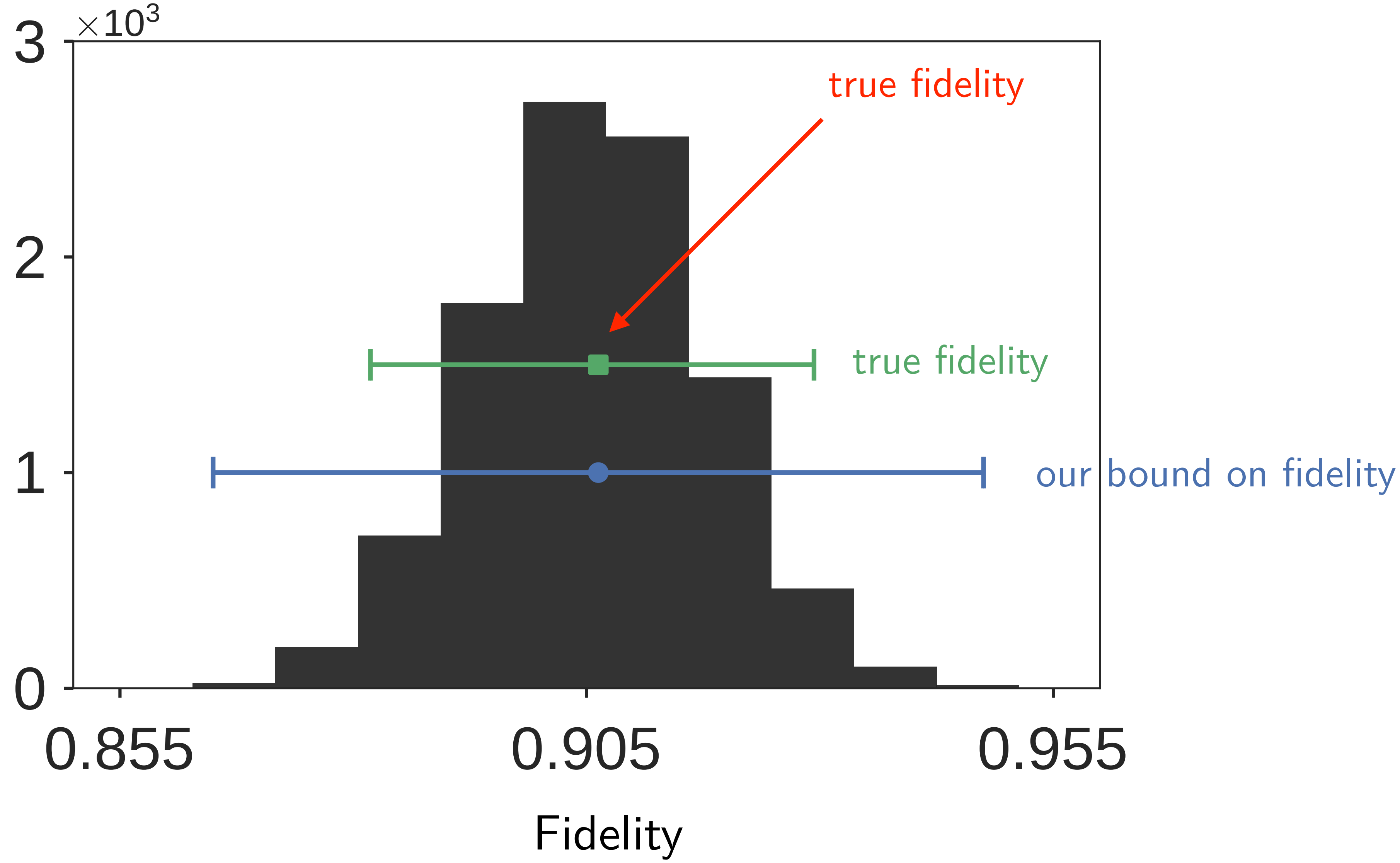
Then our computed estimator, for a 95% confidence interval, is:

$$\widehat{g}(\{y_i\}_{i=0}^{1}) = 0.952\frac{y_1}{100} + 0.024$$

Tightness of risk

4 qubit state, use 3/4 of all possible Pauli POVM, 100 repetitions

Empirically compute "true" risk (=width of confidence interval) over 10k simulations
Our guaranteed risk is only 1.74x larger



Comparison to MLE

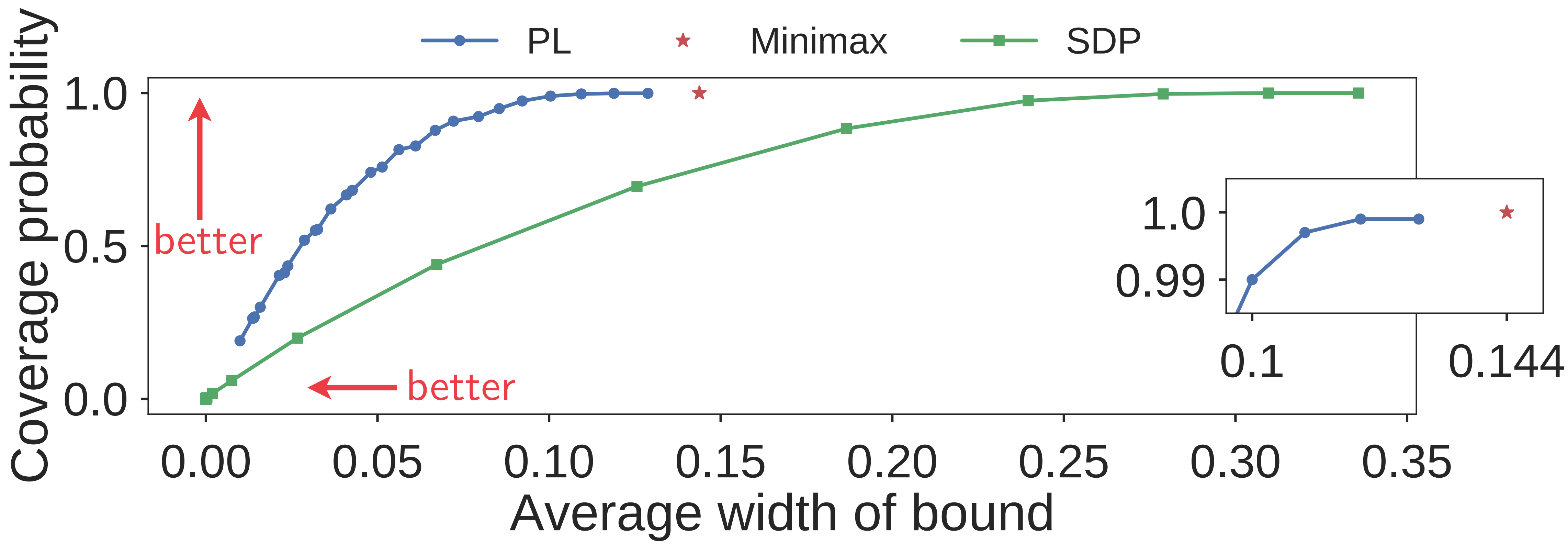
Experimental data: 3 different 4 qubit states, 81 POVM, 100 shots

	Minimax method		MLE	
	F Estimate	Risk	F Estimate	MC risk
GHZ	0.84	0.053	0.84	0.023
W	0.89	0.049	0.88	0.019
Cluster	0.79	0.048	0.79	0.020

heuristic estimate via bootstrap;
sometimes "hedge" away from 0

We can construct states for which MLE + MC/boostrap risk is **overconfident**
e.g., POVM is uninformative and MLE returns overconfident risk

Comparisons to SDP and Profile Likelihood



Profile likelihood (PL) isn't bad, except **you don't know** the coverage probability $1-\epsilon$ (we could calculate it here only because we setup a synthetic simulation)

Computational time

State	n	1	2	3	4	5
Random	L	3	12	48	192	768
	Time	1 min	2.6 min	13.6 min	1.3 hr	13.1 hr
GHZ	L	2	4	8	16	32
	Time	23.4 s	2 min	3.8 min	10.9 min	2.2 hr

TABLE II. Time taken to construct the estimator for a random n-qubit target state and an n-qubit GHZ state (average of 3 simulations). We use $L = 0.75 \times 4^n$ Pauli measurements for the random state, while $L = 2^n$ Pauli measurements for the GHZ state. Total memory (for constructing the estimator and the data for testing it) for all qubits put together is approximately 1.2 GB for the random state and 112 MB for the GHZ state. The computations were performed on a 2.5GHz CPU without parallelization.