From Fisher Information to Adaptive Experimental Design: Statistical Optimality in Quantum State Tomography

Hongru Zhao

IRSA Faragher Distinguished Postdoctoral Fellow
School of Statistics
University of Minnesota

July 9, 2026

Outline

1. Sequential estimation and adaptive experimental design
2. Fisher information and optimal design
3. Quantum state tomography as a statistical problem
4. Adaptive measurement design for quantum tomography
5. Optimality theory and simulation results

A simple example: estimating a Gaussian mean

Assume

$$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma_0^2),$$

where σ_0^2 is known.
The estimator is

$$\widehat{\mu}_n = \bar{X}_n.$$

A simple example: estimating a Gaussian mean

Assume

$$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma_0^2),$$

where σ_0^2 is known.
The estimator is

$$\widehat{\mu}_n = \bar{X}_n.$$

We have

$$\sqrt{n}(\bar{X}_n - \mu) \sim N(0, \sigma_0^2).$$

The variance of the estimator is controlled by the information contained in the data.

From Gaussian models to general parametric models

For a regular parametric model

$$X \sim f_\theta, \qquad \theta \in \mathbb{R}^p,$$

the MLE satisfies

$$\sqrt{n}(\widehat{\theta}_n - \theta) \Rightarrow N(0, \mathcal{I}(\theta)^{-1}).$$

The Fisher information matrix is

$$\mathcal{I}(\theta) = \mathbb{E}_\theta[\nabla_\theta \log f_\theta(X) \nabla_\theta \log f_\theta(X)^T].$$

MLE under sequential experimental design

Now suppose the experiment can be chosen adaptively.
At each step, $X_t | a_t \sim f_{\theta, a_t}$, where $a_t \in \{1, \ldots, k\}$ is the selected experiment.

MLE under sequential experimental design

Now suppose the experiment can be chosen adaptively.
At each step, $X_t | a_t \sim f_{\theta, a_t}$, where $a_t \in \{1, \ldots, k\}$ is the selected experiment.
Let $\bar{\pi}_n(a) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}\{a_t = a\}$ be the empirical proportion of using experiment $a \in [k]$.

MLE under sequential experimental design

Now suppose the experiment can be chosen adaptively.
At each step, $X_t | a_t \sim f_{\theta, a_t}$, where $a_t \in \{1, \ldots, k\}$ is the selected experiment.
Let $\bar{\pi}_n(a) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}\{a_t = a\}$ be the empirical proportion of using experiment $a \in [k]$.
If $\bar{\pi}_n(a) \to \pi_a, a = 1, \ldots, k$, then the accumulated information satisfies

$$\frac{1}{n} J_n(\theta) \to \mathcal{I}^{\pi}(\theta) := \sum_{a=1}^{k} \pi_a \mathcal{I}_a(\theta), \text{ where } J_n(\theta) = \sum_{t=1}^{n} \mathcal{I}_{a_t}(\theta).$$

MLE under sequential experimental design

Now suppose the experiment can be chosen adaptively.
At each step, $X_t | a_t \sim f_{\theta, a_t}$, where $a_t \in \{1, \ldots, k\}$ is the selected experiment.
Let $\bar{\pi}_n(a) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}\{a_t = a\}$ be the empirical proportion of using experiment $a \in [k]$.
If $\bar{\pi}_n(a) \to \pi_a$, $a = 1, \ldots, k$, then the accumulated information satisfies

$$\frac{1}{n} J_n(\theta) \to \mathcal{I}^{\pi}(\theta) := \sum_{a=1}^{k} \pi_a \mathcal{I}_a(\theta), \text{ where } J_n(\theta) = \sum_{t=1}^{n} \mathcal{I}_{a_t}(\theta).$$

Asymptotic behavior of the MLE

$$\sqrt{n}(\widehat{\theta}_n - \theta) \Rightarrow N\left(0, [\mathcal{I}^{\pi}(\theta)]^{-1}\right).$$

MLE under sequential experimental design

Now suppose the experiment can be chosen adaptively.
At each step, $X_t | a_t \sim f_{\theta, a_t}$, where $a_t \in \{1, \ldots, k\}$ is the selected experiment.
Let $\bar{\pi}_n(a) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}\{a_t = a\}$ be the empirical proportion of using experiment $a \in [k]$.
If $\bar{\pi}_n(a) \to \pi_a, a = 1, \ldots, k$, then the accumulated information satisfies

$$\frac{1}{n} J_n(\theta) \to \mathcal{I}^{\pi}(\theta) := \sum_{a=1}^{k} \pi_a \mathcal{I}_a(\theta), \text{ where } J_n(\theta) = \sum_{t=1}^{n} \mathcal{I}_{a_t}(\theta).$$

Asymptotic behavior of the MLE

$$\sqrt{n}(\widehat{\theta}_n - \theta) \Rightarrow N\left(0, [\mathcal{I}^{\pi}(\theta)]^{-1}\right).$$

The central question:

How should we collect data to maximize the information about θ?

Why adaptive experimental design?

In many experiments, we can choose the next experiment.
Examples:

- medical experiments: choose the next treatment;
- quantum experiments: choose the next measurement.

The choice should depend on previous observations.

$$(a_1, X_1), \ldots, (a_n, X_n) \longrightarrow a_{n+1}.$$

Sequential experimental design

At time t:

1. Choose action

$$a_t = \phi_t(H_{t-1}),$$

   where

$$H_{t-1} = \{(a_i, X_i) : i < t\}.$$

2. Observe

$$X_t | a_t \sim f_{\theta, a_t}.$$

Sequential experimental design

At time t:

1. Choose action
$$a_t = \phi_t(H_{t-1}),$$
where
$$H_{t-1} = \{(a_i, X_i) : i < t\}.$$
2. Observe
$$X_t | a_t \sim f_{\theta, a_t}.$$

Goal:

Design a sequential policy whose estimator is optimal.

Static design and averaged Fisher information

Suppose actions are selected with proportions

$$\pi = (\pi_1, \ldots, \pi_k) \in \Delta_k.$$

The averaged Fisher information is

$$\mathcal{I}^\pi(\theta) = \sum_{a=1}^{k} \pi_a \mathcal{I}_a(\theta).$$

Different choices of π lead to different estimation accuracy.

Why optimal design is nontrivial

For multidimensional parameters,

$$\mathrm{Cov}(\widehat{\theta}) \approx \frac{1}{n}\mathcal{I}^{\pi}(\theta)^{-1}.$$

Matrices are partially ordered:

$$A \succeq B$$

does not imply every pair is comparable.

Why optimal design is nontrivial

For multidimensional parameters,

$$\mathrm{Cov}(\widehat{\theta}) \approx \frac{1}{n}\mathcal{I}^{\pi}(\theta)^{-1}.$$

Matrices are partially ordered:

$$A \succeq B$$

does not imply every pair is comparable.
Therefore we need scalar criteria preserving the Loewner ordering.

A-optimal design

The classical A-optimal criterion minimizes average variance:

$$\pi_A^\star = \arg\min_{\pi \in \Delta_k} \operatorname{tr}\{\mathcal{I}^\pi(\theta)^{-1}\}.$$

Interpretation:

- smaller total estimation variance;
- optimal average precision.

Classical optimal design, see Pukelsheim (2006).

Loss induced optimal design

Consider a loss satisfying

$$d_D^2(\theta, \theta + \Delta\theta) = \Delta\theta^T G_D(\theta)\Delta\theta + o(\|\Delta\theta\|^2).$$

Combining with

$$\sqrt{n}(\widehat{\theta}_n - \theta) \approx N(0, \mathcal{I}^{\bar{\pi}_n}(\theta)^{-1}),$$

gives

$$\mathbb{E}d_D^2(\hat{\theta}_n, \theta) \approx \frac{1}{n}\operatorname{tr}\left(G_D(\theta)[\mathcal{I}^{\bar{\pi}_n}(\theta)]^{-1}\right).$$

See Li and Zhao (2025) for the sequential optimality theory.

Loss optimal design criterion

The optimal allocation minimizes the asymptotic estimation risk

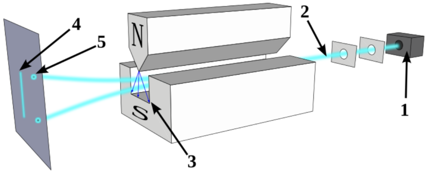
$$\Phi_D(\pi;\theta) = \operatorname{tr}\left(G_D(\theta)[\mathcal{I}^\pi(\theta)]^{-1}\right).$$

The criterion balances two components:

- $\mathcal{I}^\pi(\theta)$: information obtained from the experimental design;
- $G_D(\theta)$: local geometry of the loss function.

The optimal design depends on the statistical objective and the uncertainty structure of the parameter.

Stern–Gerlach experiment: measuring quantum spin



Stern–Gerlach experiment: Silver atoms travelling through an inhomogeneous magnetic field, and being deflected up or down depending on their spin; (1) furnace, (2) beam of silver atoms, (3) inhomogeneous magnetic field, (4) classically expected result, (5) observed result

A magnetic field gradient separates spin-1/2 particles into two outcomes:

spin up (+1) and spin down (-1)

along the measurement direction.

From spin measurement to Born's rule

A spin-1/2 state is represented by a Bloch vector

$$\rho = \frac{1}{2}(I + \theta^\top \sigma), \qquad \theta \in \mathbb{R}^3, \quad \|\theta\| \leq 1.$$

A Stern–Gerlach apparatus with magnetic-field gradient direction

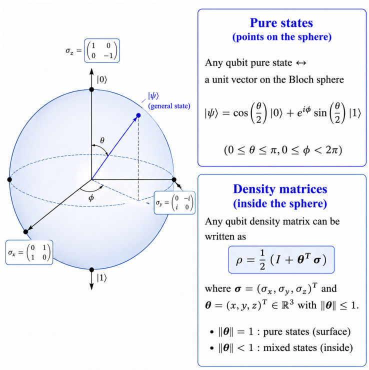
$$u = (u_x, u_y, u_z), \qquad \|u\| = 1$$

measures the spin observable

$$\sigma_u = u_x \sigma_X + u_y \sigma_Y + u_z \sigma_Z.$$

The Born rule gives

$$P(+1 \mid u, \rho) = \mathrm{tr}\left(\rho \frac{I + \sigma_u}{2}\right) = \frac{1 + \theta^\top u}{2}.$$



The angle between the Bloch vector and the magnetic-field gradient determines the measurement probability.

Quantum state estimation

The unknown object is the density matrix

$$\rho \in \mathbb{C}^{2\times 2}.$$

Constraints:

$$\rho \succeq 0, \qquad \mathrm{tr}(\rho) = 1.$$

The task:

$$\mathbb{E} d_D^2(\hat{\rho}, \rho).$$

What does tomography mean?

Tomography reconstructs an object from projections.
Examples:

- medical CT reconstructs a body from X-ray projections;
- quantum tomography reconstructs ρ from measurement probabilities.

Measurements as statistical experiments

A quantum measurement setting a is described by a POVM

$$\{Q_{a,b} : b \in [r_a]\},$$

where

$$Q_{a,b} \succeq 0, \qquad \sum_{b=1}^{r_a} Q_{a,b} = I.$$

Here

$$a \in [k], \qquad r_a \geq 2.$$

For a state $\rho(\theta)$, Born's rule gives

$$p_{a,b}(\theta) = P(X = b|a) = \mathrm{tr}(\rho(\theta) Q_{a,b}).$$

A quantum experiment produces classical random outcomes.

Local geometry of quantum state loss

We measure estimation error through a distance between states:

$$d_D(\rho(\theta + \Delta\theta), \rho(\theta)), \qquad D \in \{F, B\}.$$

For full-rank states $\rho(\theta) \succ 0$, both Frobenius and Bures distances admit the local expansion

$$d_D^2(\theta, \theta + \Delta\theta) = \Delta\theta^T G_D(\theta)\Delta\theta + o(\|\Delta\theta\|^2).$$

- Frobenius loss:

$$d_F(\rho, \sigma) = \|\rho - \sigma\|_F, \qquad G_F = \frac{1}{2} I_d.$$

- Bures loss:

$$d_B(\rho, \sigma)^2 = 2 - 2\,\mathrm{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}},$$

  where $G_B(\theta)$ is the positive definite Bures metric matrix.

Likelihood and maximum likelihood estimation

Given observations (a_t, b_t), define counts

$$N_{a,b} = \sum_{t=1}^{n} 1\{a_t = a, b_t = b\}.$$

The likelihood is

$$L(\rho) = \prod_{a,b} \{\mathrm{tr}(\rho Q_{a,b})\}^{N_{a,b}}.$$

The constrained MLE is

$$\widehat{\rho}_n = \arg \max_{\rho \succeq 0, \mathrm{tr}\, \rho = 1} L(\rho).$$

The design problem is to choose future measurements to improve this estimator.

Fisher information for quantum measurements

Use a parametrization

$$\rho(\theta), \qquad \theta \in \mathbb{R}^d.$$

Define

$$s_i(a,b) = \mathrm{tr}(\sigma_i Q_{a,b}),$$

and

$$p_{a,b}(\theta) = \mathrm{tr}(\rho(\theta) Q_{a,b}).$$

For one measurement setting a,

$$[\mathcal{I}_a(\theta)]_{ij} = \frac{1}{4} \sum_b \frac{s_i(a,b) s_j(a,b)}{p_{a,b}(\theta)}.$$

Loss-optimal experimental design

For a static design

$$\pi \in \Delta_k,$$

the averaged Fisher information is

$$\mathcal{I}^\pi(\theta) = \sum_{a=1}^{k} \pi_a \mathcal{I}_a(\theta).$$

The optimal allocation solves

$$\pi_D^\star(\theta) = \arg\min_{\pi \in \Delta_k} \operatorname{tr}\left(G_D(\theta)\{\mathcal{I}^\pi(\theta)\}^{-1}\right).$$

The objective depends on both:

- information provided by measurements;
- geometry of the loss.

Adaptive measurement selection

For adaptive experiments,

$$J_n = \sum_{t=1}^{n} \mathcal{I}_{a_t}(\widehat{\theta}_n)$$

is the accumulated Fisher information.
If

$$\bar{\pi}_n(a) = \frac{1}{n}\sum_{t=1}^{n} 1\{a_t = a\},$$

then

$$\frac{1}{n} J_n \approx \mathcal{I}^{\bar{\pi}_n}(\theta).$$

For an efficient estimator,

$$\mathbb{E} d_D^2(\widehat{\theta}_n, \theta) \approx \frac{1}{n} \operatorname{tr}\left(G_D(\theta)[\mathcal{I}^{\bar{\pi}_n}(\theta)]^{-1} \right).$$

Greedy information design

The goal is to approximate the optimal allocation $\pi_D^\star$ sequentially.

$$a_{n+1} = \arg\min_a \operatorname{tr}\left[G_D(\widehat{\theta}_n)(J_n + \mathcal{I}_a)^{-1}\right].$$

This is the GI0 rule.
Using

$$(J + \mathcal{I}_a)^{-1} \approx J^{-1} - J^{-1}\mathcal{I}_a J^{-1},$$

we obtain GI1:

$$a_{n+1} = \arg\max_a \operatorname{tr}(G_D(\widehat{\theta}_n)J_n^{-1}\mathcal{I}_a J_n^{-1}).$$

Convergence to the loss-optimal measurement allocation

Theorem: asymptotic optimality of adaptive design

Fix a loss function $D \in \{\mathrm{A}, \mathrm{B}\}$, and run the corresponding D-GI0 or D-GI1 adaptive measurement rule.
Let $\bar{\pi}_n$ be the empirical measurement frequency after n single-shot measurements.
Define the optimal static design

$$\pi_D^\star(\theta) = \arg\min_{\pi \in \Delta_k} \Phi_D(\pi; \theta),$$

where

$$\Phi_D(\pi; \theta) = \mathrm{tr}\left(G_D(\theta) \{\mathcal{I}^\pi(\theta)\}^{-1} \right).$$

If the minimizer is unique, then

$$\bar{\pi}_n \longrightarrow \pi_D^\star(\theta) \qquad \mathbb{P}_\theta\text{-a.s.}$$

Local asymptotic minimax optimality

Theorem: local asymptotic minimax lower bound

Consider any regular adaptive estimation procedure T_n.
For the local alternatives

$$\theta_n = \theta + \frac{h}{\sqrt{n}},$$

the asymptotic risk satisfies

$$\sup_{|F|<\infty} \liminf_{n\to\infty} \sup_{h\in F} \mathbb{E}_{\theta+h/\sqrt{n}} \left[n d_D^2 \left(\rho(T_n), \rho\left(\theta + \frac{h}{\sqrt{n}}\right)\right)\right]$$
$$\geq \inf_{\pi\in\Delta_k} \operatorname{tr}\left(G_D(\theta)[\mathcal{I}^\pi(\theta)]^{-1}\right).$$

- The right-hand side is the optimal risk achievable by any measurement allocation.
- No regular adaptive procedure can outperform this bound.

Optimality of adaptive GI design

Theorem: asymptotic efficiency of adaptive MLE

Let $\widehat{\theta}_n$ be the MLE obtained under the loss-optimal GI0 or GI1 measurement strategy. The empirical measurement frequencies converge to the optimal allocation: $\bar{\pi}_n \longrightarrow \pi^\star_D(\theta)$, where

$$\pi^\star_D(\theta) = \arg\min_{\pi \in \Delta_k} \operatorname{tr}\left(G_D(\theta)[\mathcal{I}^\pi(\theta)]^{-1}\right).$$

Consequently,

$$\lim_{n\to\infty} n\, \mathbb{E}_\theta d_D^2\left(\rho(\widehat{\theta}_n), \rho(\theta)\right) = \inf_{\pi \in \Delta_k} \operatorname{tr}\left(G_D(\theta)[\mathcal{I}^\pi(\theta)]^{-1}\right).$$

Adaptive design learns the optimal experiment allocation and achieves the local minimax lower bound.

Simulation: one-qubit Pauli library

We compare measurement strategies:

- uniform allocation;
- GI0 adaptive design;
- GI1 adaptive design;
- oracle optimal design.

For the one-qubit Pauli library, the adaptive procedures approach the oracle performance.

Simulation: One-qubit measurement libraries

For a unit measurement direction $u \in \mathbb{R}^3$,

$$\sigma_u = u_x\sigma_x + u_y\sigma_y + u_z\sigma_z, \qquad \|u\|_2 = 1.$$

The corresponding projective measurement has POVM elements $Q_{u,\pm} = \frac{I \pm \sigma_u}{2}$, corresponding to spin outcomes $+1$ and -1.

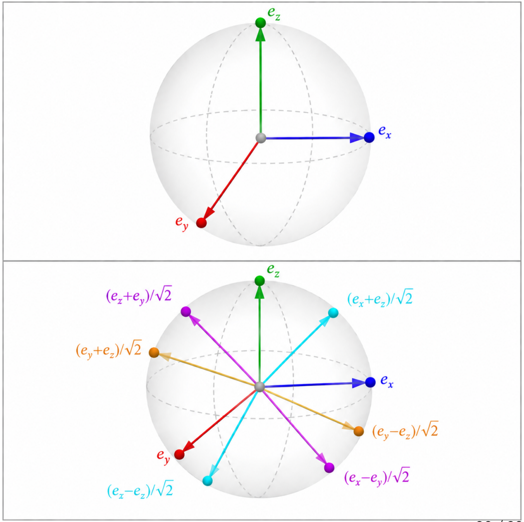
We compare two one-qubit measurement libraries:

- **Pauli library**

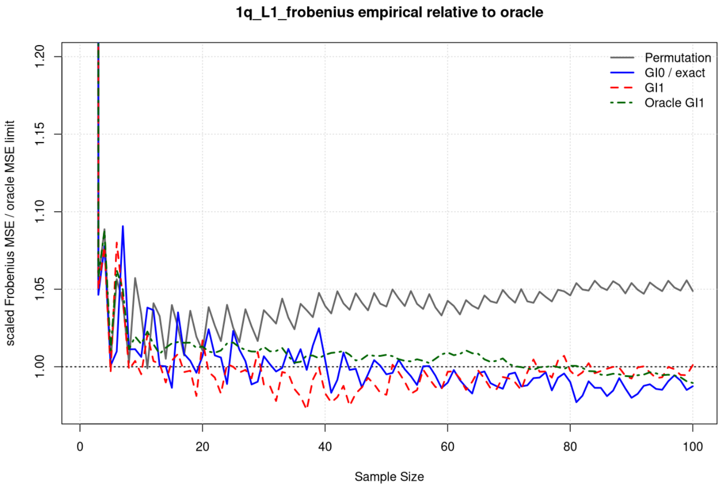
$$u \in \{e_x, e_y, e_z\}.$$

- **Nine-setting projective library**

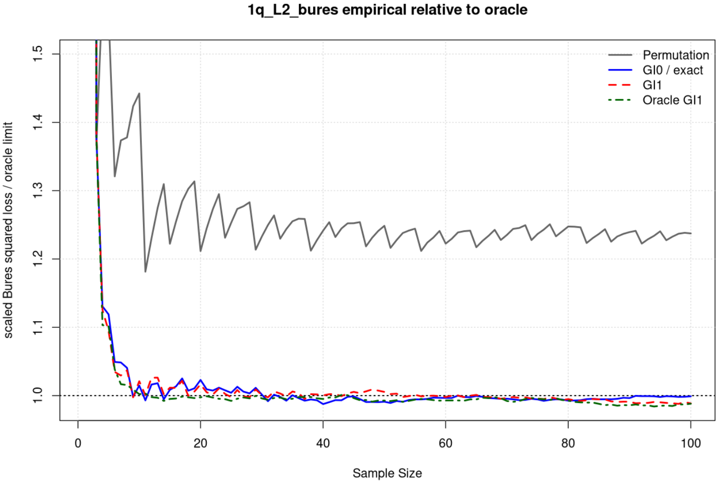
$$u \in \left\{ e_x, e_y, e_z, \frac{e_x \pm e_y}{\sqrt{2}}, \frac{e_x \pm e_z}{\sqrt{2}}, \frac{e_y \pm e_z}{\sqrt{2}} \right\}.$$



Simulation results



Simulation results: nine-setting measurements



Conclusion

- Quantum tomography is a statistical estimation problem.
- Measurements are experimental choices controlled by the researcher.
- Fisher information connects measurements and estimation accuracy.
- Loss geometry determines the optimal design criterion.
- Adaptive GI methods achieve asymptotic optimality.

References

F. Pukelsheim.
Optimal Design of Experiments.
SIAM, 2006.

X. Li and H. Zhao.
Globally-Optimal Greedy Active Sequential Estimation.
IEEE Transactions on Information Theory, 71(5):3871–3924, 2025.