# Introduction to Optimal Design of Experiments (DoE) and its Applications to Quantum Estimation

Jun Suzuki
Graduate School of Informatics and Engineering
Institute for Advanced Science
The University of Electro-Communications, Tokyo

2023.9.20 @ QIMG2023

# Introduction

- **Parameter estimation problem for quantum states**
  - Given a family of quantum states $\{\rho_\theta | \theta = (\theta_1, \theta_2, \ldots, \theta_n) \in \Theta\}$
  - To infer the true parameter value $\theta$ from a "measurement"

  Q. 1) Optimal estimation strategy?
  2) Estimation error bounds?

- Importance of *quantum estimation theory*:

  Metrology, sensing, imaging, noise characterization, etc.

- Interplay between classical statistics and quantum statistics

- An approach: **Optimal design of experiment** (DoE) in statistics

  Any experiment = designing experiment + data analysis

  A useful statistical method for designing good experiments

  (Cf. Active learning or Bayesian optimization in machine learning)

# Introduction

- Design of Experiment (DoE) in classical statistics:

  Mainly for linear regression models, or its variants

- Applications of the theory of DoE to quantum systems

  Hayashi, Matsumoto IEICE Transsactions (2000)
  Kosut, Walmsley, Rabitz, arXiv: 0411093 (2004).
  Nunn, Smith, Puentes, Walmsley, Lundeen, PRA (2010)
  Balló, Hangos, arXiv: 1004.5209 (2010), 1107.0890 (2011)
  Balló, Hangos, Petz, IEEE Trans.Autom.Control (2012)
  Ruppert, Virosztek, Hangos, J.Phys.A (2012)

  - Discrete design problem (Combinatorial optimization)

  - Additional constraints and/or a specific figure of merit.

1) General formulation of the problem
2) Optimal designs for popular optimality criteria in qubit models

# Classical vs Quantum

| | Classical case | Quantum case |
|---|---|---|
| Model | Linear regression<br>Nonlinear response | General<br>Cf. Helstrom(67) |
| Constraint | Additional | by default |
| Optimality | A, c, D, E, G, $\gamma$, ... | A, c, D, E, $\gamma$ |
| Equivalence Thm. | Kiefer-Wolfowitz (60)<br>General (70s) | qubit<br>$D = A_G$ |
| Adaptive design | Various schemes | Nagaoka (89) |
| Comparison | Not popular | A, D, E (qubit) |
| Information Geometry | Few | Petz, Nagaoka |

Gazit, Ng, JS (2019), JS, Yang, Hayashi (2020), JS (2021)

# Outline of Talk

- Introduction to DoE
- General formulation of DoE
- Analytically solvable examples (qubit)
- Comparison of different optimal designs
- Equivalence theorem
- Application to estimation of Pauli noise

What is Design of Experiments (DoE)?

# A Toy Example: Laboratory assignment

- Students are asked to measure the positions of a particle on a linear air truck.
- In the ideal limit, it obeys the Newton's law without friction.

$$x(t) = v_0 t + x(0)$$

- The objective of the project is to determine the speed $v_0$.

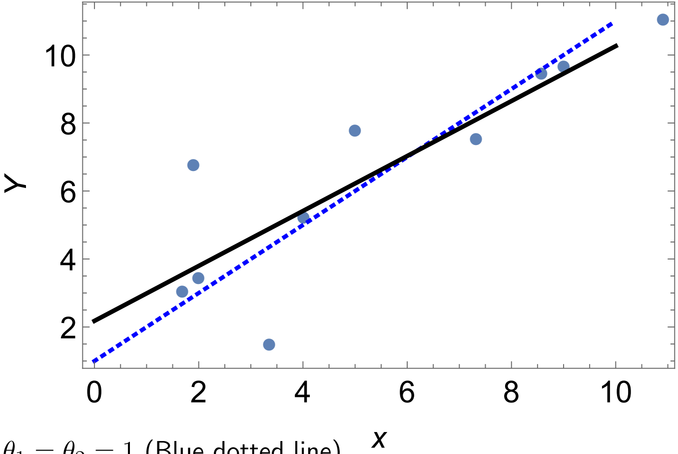
- Question: When and how many times should we measure positions?
- In the language of linear regression, this is to estimate $\theta_1, \theta_2$:

$$Y_i = \theta_1 x_i + \theta_2 + \epsilon$$

where $\epsilon$: Noise, error, uncertainty (Random variable!)

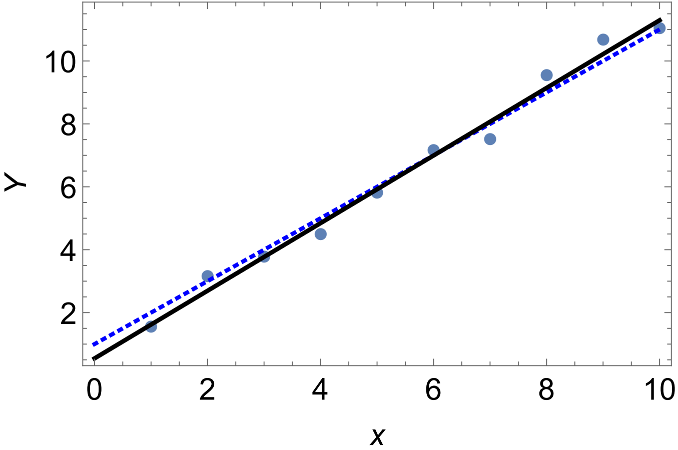
| DoE: to choose the set of $\{x_i\}$ to minimize estimation error. |
|---|

# Simple Linear Regression: $Y_i = \theta_1 x_i + \theta_2 + \epsilon$



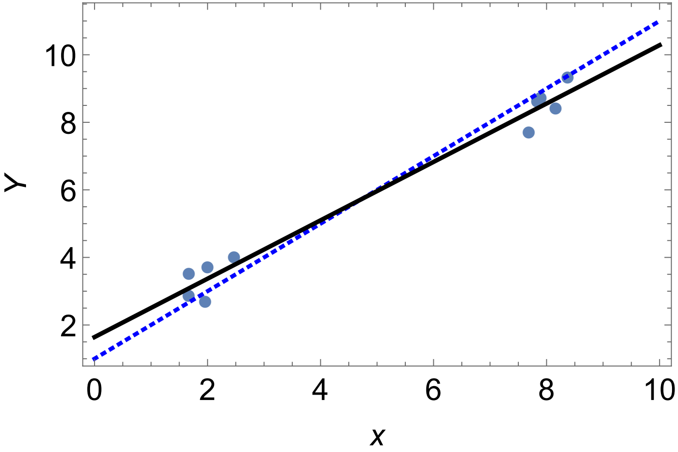

True parameter: $\theta_1 = \theta_2 = 1$ (Blue dotted line)
In the usual setting, points are given.
In the theory of optimal DoE, we can choose $x_i$ $(i = 1, 2, \ldots, 10)$!
Design $e = (x_1, x_2, \ldots, x_{10})$ with $x \in [0, 11]$

# Simple Linear Regression: $Y_i = \theta_1 x_i + \theta_2 + \epsilon$ ($\theta_1 = 1$, $\theta_2 = 1$)




**Student A:** $x_1 = 1, x_2 = 2, \ldots, x_{10} = 10$ (Equal space points)

# Simple Linear Regression: $Y_i = \theta_1 x_i + \theta_2 + \epsilon$ $(\theta_1 = 1,\ \theta_2 = 1)$




**Student B:** $x_1, \ldots, x_5 \sim 2$ and $x_6, \ldots, x_{10} \sim 8$

# What can DoE do?

e.g., simple linear regression: $Y_i = \theta_1 x_i + \theta_2 + \epsilon$

- How should we choose the set of points $x_i$ to minimize the error? $(i = 1, 2, \ldots, N)$
- How many points (samples) do we need to guarantee the aimed error? $(N = ?)$

- We need to identify the risk. (estimation error)

  e.g., mean squared error, mean absolute error, Bayes error, maximum error, etc.
- DoE $=$ optimization over measurement variables $x_i$

  Goal: to design the best statistical model for good experimental data

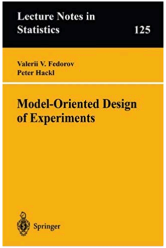
# Optimal Design of Experiments (DoE)

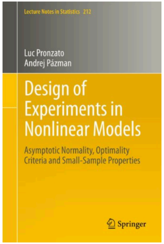
Classic books:

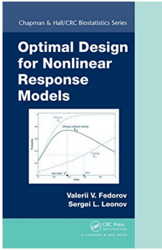
**Fedorov**, *Theory of Optimal Experiments* (1972)
**Pukelsheim**, *Optimal Design of Experiments* (2006)

DoE for Nonlinear models

**Fedorov&Hackl**, *Model-Oriented Design of Experiments* (1997)
**Pronzato&Pázman**, *Design of Experiments in Nonlinear Models: Asymptotic Normality, Optimality Criteria and Small-Sample Properties* (2013)
**Fedorov&Leonov**, *Optimal Design for Nonlinear Response Models* (2014)



# Point Estimation of Parametric Model

- In the standard setting, a parametric model is **fixed**.

$$\mathcal{M} = \{p_\theta \,|\, \theta = (\theta_1, \theta_2, \ldots, \theta_n) \in \Theta\}$$

E.g. $p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\quad$ ($n = 2$ with $\theta_1 = \mu, \theta_2 = \sigma^2$)

Objectives:

- Given a datum: $X^N = X_1 X_2 \ldots X_N \overset{\text{i.i.d.}}{\sim} p_\theta$ (sample size $N$)

1) To find a good estimator $\hat{\theta} : \mathcal{X}^N \to \Theta$.
2) Estimation error bounds.
3) To analyze statistical properties of random variable $\hat{\theta}(X^N)$.

# Basic Setting of DoE

In contrast, a parametric model is **not** fixed in DoE.

① **Model**: $s \in \mathcal{S} \subset \mathbb{R}^d$ (parametrized model: $s_\theta$ with $\theta \in \Theta \subset \mathbb{R}^n$: Model parameter set)

② **Design**: $e \in \mathcal{E}$: Design set (experiment setting)

③ **Model function**: $f : \mathcal{S} \times \mathcal{E} \to \mathcal{P}(\mathcal{X})$

$\mathcal{P}(\mathcal{X})$: Set of all probability distributions on $\mathcal{X}$

$f : (s, e) \mapsto p_s(\cdot|e)$ conditional probability distribution on $\mathcal{X}$.

E.g Linear model: $Y_\theta(e) = \sum_i \theta_i g_i(e) + \epsilon$. ($g_i$: polynominal, $\epsilon \sim \mathcal{N}(0, \sigma)$: noise)

- A model parameter $\theta \in \Theta$, design $e \in \mathcal{E}$, and model function $f$

  specify a parametric model: $\mathcal{M}[e] = \{p_\theta(\cdot|e) \,|\, \theta \in \Theta\}$

$$\boxed{① + ② + ③ \Rightarrow p_\theta(x|e)}$$

(Cf. GPT)

# Objective of DoE

- To find an optimal experimental setting $e_{\mathrm{opt}}$ that minimizes the risk.
- Parameter estimation: the standard risk = Mean-Square-Error (MSE) matrix.

$$V_\theta[\hat{\theta}|e] := \left[E_\theta\left[(\hat{\theta}_i(X)-\theta_i)(\hat{\theta}_j(X)-\theta_j)\middle|e\right]\right]_{i,j\in\{1,2,\dots,n\}}$$

$E_\theta[X|e]$: Conditional expectation value w.r.t. $p_\theta(x|e)$

- Cramér-Rao inequality: $V_\theta[\hat{\theta}|e] \geq \left(J_\theta[e]\right)^{-1}$

for any **locally unbiased estimators**.

$J_\theta[e]$: Fisher information matrix about $\mathcal{M}[e] = \{p_\theta(\cdot|e)\,|\,\theta\in\Theta\}$

Optimization problem: $\min_{e\in\mathcal{E}} \left(J_\theta[e]\right)^{-1} \Leftrightarrow$ Optimal DoE

# Optimality Criteria 1

- Minimization of $D_\theta(e) = \big(J_\theta[e]\big)^{-1}$ (Design matrix)

  Not uniquely defined in general! (Matrix quantity?)

- Popular optimality criteria in DoE:

  - Löwner optimality: as a matrix inequality $D(e_*) \leq D(e)$ $(\forall e \in \mathcal{E})$

  - A-optimality: $\mathrm{Tr}\,\{D(e)\}$ or weighted trace $\mathrm{Tr}\,\{W D(e)\}$ $(W > 0)$

  - D-optimality: $\mathrm{Det}\,\{D(e)\}$

  - E-optimality: Maximum eigenvalue of $D(e)$

  - c-optimality: $c^t D(e) c$ (scalar) for a given vector $c \in \mathbb{R}^n$

  - Min-max optimality: $\max_i [D(e)]_{ii}$

  In fact; A, c, D, E, G, $\gamma$, I, K, L, M, T, V, $\mathcal{X}$,... (Alphabetic criteria!)

# Optimality Criteria 2

- One-parameter family of optimality criteria ($\gamma$-optimal):

$$\Psi_\gamma(D) := \left(\frac{1}{n}\mathrm{Tr}\left\{D^\gamma\right\}\right)^{\frac{1}{\gamma}} \quad (\gamma \in \mathbb{R})$$

To minimize $\Psi_\gamma[e] := \Psi_\gamma\big(D(e)\big) = \left(\frac{1}{n}\mathrm{Tr}\left\{J_\theta[e]^{-\gamma}\right\}\right)^{\frac{1}{\gamma}}$

- General formulation

$\Psi$-optimality: $\Psi\big(D\big)$ (Scalar function of the design matrix $D \geq 0$), $D_\theta(e) = \big(J_\theta[e]\big)^{-1}$

- Remarks:

- $\gamma = 1 \Rightarrow$ A-optimality: $\Psi_A\big(D\big) = \mathrm{Tr}\left\{D\right\}$
  $\gamma \to 0 \Rightarrow$ D-optimality: $\Psi_D\big(D\big) = \mathrm{Det}\left\{D\right\}$
  $\gamma \to \infty \Rightarrow$ E-optimality: Maximum eigenvalue of $D(e)$
- An optimal design depends on $\theta$ in general. (Local optimality)

# General Design Problem

- i.i.d. design: Sample size $N$

  Additivity of Fisher information $\Rightarrow$ Estimation error scales as $\frac{1}{N}$
- Discrete design: $e(m) = (\boldsymbol{\nu}, \boldsymbol{e}) = \big((\nu_1, e_1), \dots, (\nu_m, e_m)\big)$ (Combinatorial optimization)

$$J_\theta[e(m)] = \sum_{i=1}^{m} \nu_i J_\theta[e_i] \quad \text{with } \nu_i = \frac{N_i}{N} (N_i \in \mathbb{N}), \sum_{i=1}^{m} N_i = N$$

- Continuous design: $e(m) = (\boldsymbol{p}, \boldsymbol{e}) = \big((p_1, e_1), \dots, (p_m, e_m)\big)$ (Good approx. $N \geq 1$)

$$J_\theta[e(m)] = \sum_{i=1}^{m} p_i J_\theta[e_i] \quad \text{with } p_i \in (0,1), \sum_{i=1}^{m} p_i = 1$$

- Randomized designs perform better in general. (Cf. Randomized measurement)

# Incompatibility in DoE

- Löwner optimality: $D(e_*) \leq D(e)$ does not exist in general
- c-optimal design: $e_* = \arg\min c^t D(e) c$ depends on the choice of $c \in \mathbb{R}^n$

  E.g. $c_1 = [1, 0, \ldots, 0]^t$, $c_2 = [0, 1, 0, \ldots, 0]^t$

  Optimal designs for estimating $\theta_1$ and $\theta_2$ are incompatible!

  They are compatible, iff $e_{*,i} = \arg\min c_i^t D(e) c_i$ are independent on $i$.

- Incompatibility of estimating multiple parameters is common phenomenon in classical estimation.

# Parameter Estimation of Quantum States as DoE Problem

① model, ② design, and ③ model function

# Terminology & Notation in Quantum System

- NND$(d)$ := set of all positive semidefinite matrices on $\mathbb{C}^d$.
- **Quantum system** $\Leftrightarrow \mathcal{H} = \mathbb{C}^d$ ($d$-dim. vector space)
- **Quantum state** $\Leftrightarrow \rho \geq 0$: $\mathrm{tr}\,(\rho) = 1$
- **Measurement** $\Leftrightarrow \Pi = \{\Pi_x\}_{x \in \mathcal{X}}$: $\sum_{x \in \mathcal{X}} \Pi_x = I_d$

  ↑ design $(e)$ ⋒ NND$(d)$ ↑ $\frac{1}{2}d(d+1)$ constraints

- Born's rule $\Leftrightarrow$ $(\rho, \Pi) \to p_\rho(\cdot|\Pi)$

  ↑ model function $p_\rho(x|\Pi) = \mathrm{tr}\,(\rho \Pi_x)$

Optimal design problem on positive matrices with constrains

Analytically hard to solve $\to$ numerical search

# Parameter Estimation of Quantum State

- Given an $n$-parameter family of states: $\mathcal{M}^Q := \{\rho_\theta \,|\, \theta = (\theta_1, \ldots, \theta_n) \in \Theta\}$
- Objectives of a quantum-state estimation:
  - To find a "good" measurement (design $\Pi$) and estimator ($\hat{\theta}$). (Optimal DoE!)
- Estimation error ⇔ Mean square error (MSE):

$$V_\theta[(\Pi, \hat{\theta})] = \Big[ \sum_{x \in \mathcal{X}} (\hat{\theta}_i(x) - \theta_i)(\hat{\theta}_j(x) - \theta_j) \mathrm{tr}\,(\rho_\theta \Pi_x) \Big]$$

- Common choice for the figure of merit (A-optimality)

⇔ Weighted trace of MSE (Weight matrix $W > 0$):

To find

$$\boxed{C_\theta[W, \rho_\theta] := \min_{(\Pi,\hat{\theta})\,:\mathrm{l.u.at}\theta} \mathrm{Tr}\left\{ W V_\theta[(\Pi, \hat{\theta})] \right\}}.$$

# Parameter Estimation of Quantum State as DoE

- A-optimality

  ⇔ Weighted trace of MSE (Weight matrix $W > 0$):

  To find $$\boxed{C_\theta[W, \rho_\theta] = \min_{\Pi:\mathrm{POVM}} \mathrm{Tr}\left\{W\left(J_\theta[\Pi]\right)^{-1}\right\}}.$$

  $J_\theta[\Pi]$: Classical Fisher information matrix $\mathcal{M}[\Pi] = \{p_\theta(x|\Pi) = \mathrm{tr}\,(\rho_\theta \Pi_x) \mid \theta \in \Theta\}$

  $$J_{\theta,jk}[\Pi] = \sum_{x\in\mathcal{X}} \frac{\mathrm{tr}\,((\partial_j \rho_\theta)\Pi_x)\,\mathrm{tr}\,((\partial_k \rho_\theta)\Pi_x)}{\mathrm{tr}\,(\rho_\theta \Pi_x)}$$

  Nagaoka, IEICE Technical Report, **IT 89-42**, 9 (1989).
  Cf. Braunstein&Caves, Phys. Rev. Lett. **72**, 3439 (1994).

- This is a convex optimization with constrains.

  Efficient algorithm: QestOptPOVM (Matlab code) (Zhang, JS, in preparation)

  https://github.com/ZHANGJianchao97/Qest.git

# Analytically Solvable Cases

c-optimality and qubit-state estimation

# c-optimality: To minimize $c^t D(e) c$

## Theorem

*Given an $n$-parameter model $M^Q$, for each $n$-dimensional (column) vector $c = (c_1, c_2, \ldots, c_n)^t \in \mathbb{R}^n$, the infimum of the MSE matrix in the direction of $\boldsymbol{c}$ is*

$$\inf_{\Pi} c^t J_\theta[\Pi]^{-1} c = c^t (J_\theta^{\mathrm{SLD}})^{-1} c. \tag{1}$$

*An optimal measurement is given by a set of projectors about the operator:*

$$L_{\theta,c} = \sum_{i,j=1}^{n} c_i \left(J_\theta^{\mathrm{SLD}}\right)^{-1}_{ij} L_{\theta,j}, \tag{2}$$

*with $J_\theta^{\mathrm{SLD}}$ the SLD Fisher information matrix and $L_{\theta,j}$ the SLD operator.*

Operational meaning of the SLD Fisher information

# Qubit-State Estimation 1

- When $\dim \mathcal{H} = 2$, $\mathcal{M}^Q = \{\rho_\theta \,|\, \theta \in \Theta\}$

  Two- and three-parameter models $\Rightarrow$ Nontrivial case

- Recall:
  Löwner optimality: as a matrix inequality ($J_\theta[e_*]^{-1} \leq J_\theta[e]^{-1}$)
  A-optimality: $\mathrm{Tr}\left\{J_\theta[e]^{-1}\right\}$
  D-optimality: $\mathrm{Det}\left\{J_\theta[e]^{-1}\right\}$
  E-optimality: Maximum eigenvalue of $J_\theta[e]^{-1}$

  $\gamma$-optimality: $\Psi_\gamma[e] := \left(\frac{1}{n}\mathrm{Tr}\left\{J_\theta[e]^{-\gamma}\right\}\right)^{\frac{1}{\gamma}}$ ($\gamma \in \mathbb{R}$)

# Qubit-State Estimation 2

## Lemma (Gill-Massar (00), Yamagata (11))

*For a qubit model* $\{\rho_\theta | \theta \in \Theta\}$, *the Fisher information matrix about a measurement* $\Pi$ *takes the form of* $J_\theta[\Pi] = \sqrt{J_\theta^{\mathrm{SLD}}} J \sqrt{J_\theta^{\mathrm{SLD}}}$.

- $J_\theta^{\mathrm{SLD}}$: *the SLD Fisher information matrix.*
- $J$: *a nonnegative matrix with* $\mathrm{Tr}\{J\} \leq 1$.

**Known Bounds (A-optimality):**

- Nagaoka bound (89) for two-parameter qubit-state model
  $$\min_{\Pi} \mathrm{Tr}\left\{W J_\theta[\Pi]^{-1}\right\} = \mathrm{Tr}\left\{W (J_\theta^{\mathrm{SLD}})^{-1}\right\} + 2\sqrt{\mathrm{Det}\left\{W (J_\theta^{\mathrm{SLD}})^{-1}\right\}}$$
- Hayashi(97)-Gill-Massar(00) bound for three-parameter model
  $$\min_{\Pi} \mathrm{Tr}\left\{W J_\theta[\Pi]^{-1}\right\} = \left(\mathrm{Tr}\left\{\sqrt{\sqrt{W} (J_\theta^{\mathrm{SLD}})^{-1} \sqrt{W}}\right\}\right)^2$$

# Qubit-State Estimation as DoE

- It suffices to look for $e(n) = \big((p_1, \ldots, p_n), (\Pi(u_1), \ldots, \Pi(u_n))\big)$

  $n$: Number of model parameters ($n = 2, 3$)

  - $\boldsymbol{p} = (p_1, \ldots, p_n) \in \mathcal{P}(n)$

  - $\Pi(u)$: PVM about $L_u = \sum_{i,j} u_i (J_\theta^{\mathrm{SLD}})_{ij}^{-\frac{1}{2}} L_{\theta,j}$

  - $\boldsymbol{u} := \{u_1, u_2, \ldots, u_n\}$ are ONB for $\mathbb{R}^n$.

To find $\gamma$-optimal design:

$$\min_{e(n)=(\boldsymbol{p},\boldsymbol{u})} \left( \frac{1}{n} \mathrm{Tr} \left\{ (J_\theta[\boldsymbol{p},\boldsymbol{u}])^{-\gamma} \right\} \right)^{\frac{1}{\gamma}}$$

$$J_\theta[\boldsymbol{p},\boldsymbol{u}]^{-1} := \sum_{i=1}^{n} \frac{1}{p_i} (J_\theta^{\mathrm{SLD}})^{-\frac{1}{2}} u_i {u_i}^t (J_\theta^{\mathrm{SLD}})^{-\frac{1}{2}}$$

# $\gamma$-Optimal Design ($\gamma \geq 1$)

## Theorem (JS 17)

*Given an $n$-parameter qubit model ($n = 2, 3$), an optimal design $e_*(n)$ and the minimum $\gamma$-optimality function ($\gamma \geq 1$) are given by*

$$\min_{e(n)} \Psi_\gamma\Big[J_\theta[e(n)]\Big] = \frac{1}{n^\gamma}\left(\mathrm{Tr}\left\{\left(J_\theta^{\mathrm{SLD}}\right)^{-\frac{\gamma}{1+\gamma}}\right\}\right)^{\frac{1+\gamma}{\gamma}}$$

$$\boldsymbol{p}_* = (p_i) \text{ with } p_i = \frac{(\lambda_i^{\mathrm{SLD}})^{-\frac{\gamma}{1+\gamma}}}{\sum_j (\lambda_j^{\mathrm{SLD}})^{-\frac{\gamma}{1+\gamma}}}$$

$$\boldsymbol{u}_* = (u_1^{\mathrm{SLD}}, \ldots, u_n^{\mathrm{SLD}})$$

*where $\lambda_i^{\mathrm{SLD}}$ and $u_j^{\mathrm{SLD}}$ are the eigenvalues and eigenvectors of the SLD Fisher information matrix.*

## Proof for $\gamma$-Optimal Design ($\gamma \geq 1$)

Converse part:

$$\begin{aligned}
\Psi_\gamma[J_\theta[e(n)]] &\overset{(1)}{\geq} \Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i^\gamma}[\boldsymbol{u}^t(J_\theta^{\mathrm{SLD}})^{-1}\boldsymbol{u}]_{ii}^\gamma\Big)^{\frac{1}{\gamma}} \\
&= \Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i^\gamma}[{u_i}^t(J_\theta^{\mathrm{SLD}})^{-1}u_i]^\gamma\Big)^{\frac{1}{\gamma}} \\
&\overset{(2)}{\geq} \frac{1}{n^\gamma}\Big(\sum_{i=1}^{n}[{u_i}^t(J_\theta^{\mathrm{SLD}})^{-1}u_i]^{\frac{\gamma}{1+\gamma}}\Big)^{\frac{1+\gamma}{\gamma}} \\
&\overset{(3)}{\geq} \frac{1}{n^\gamma}\Big(\sum_{i=1}^{n}\{\lambda_i[(J_\theta^{\mathrm{SLD}})^{-1})]\}^{\frac{\gamma}{1+\gamma}}\Big)^{\frac{1+\gamma}{\gamma}} \\
&= \frac{1}{n^\gamma}\Big(\mathrm{Tr}\{(J_\theta^{\mathrm{SLD}})^{-\frac{\gamma}{1+\gamma}}\}\Big)^{\frac{1+\gamma}{\gamma}},
\end{aligned}$$

(1): Schur's theorem of majorization + Convexity of $\ell_p$ norm ($p \geq 1$)
(2): Jensen's inequality + Convexity of $f(x) = x^{1/(1+\gamma)}$ ($\gamma > 0$)
(3): Doubly stochastic matrix + concavity of $g(x_1, \dots, x_n) = \sum_i x_i^{\gamma/(1+\gamma)}$ ($\gamma > 0$)

# A-Optimal Design

- $\gamma \to 1 \Leftrightarrow$ A-optimal design:

  $\min \ \mathrm{Tr}\left\{J_\theta[e(n)]^{-1}\right\}$

  - $\boldsymbol{p}_* = (p_i)$ with

  $$p_i = \frac{\lambda_i^{\mathrm{SLD}\,-1/2}}{\sum_{j=1}^n \lambda_j^{\mathrm{SLD}\,-1/2}},$$

  - $\boldsymbol{u}_* = (u_1^{\mathrm{SLD}}, \ldots, u_n^{\mathrm{SLD}})$ (eigenvectors of the SLD Fisher information matrix)

  PVM about $L_u = \sum_{i,j} u_i (J_\theta^{\mathrm{SLD}})_{ij}^{-1/2} L_{\theta,j}$

- $J_\theta[e_A^*] = \frac{1}{\mathrm{Tr}\left\{(J_\theta^{\mathrm{SLD}\,-1/2})\right\}}(J_\theta^{\mathrm{SLD}})^{1/2}$

# D-Optimal Design

- $\gamma \to 0 \Leftrightarrow$ D-optimal design:

  $\min\ \mathrm{Det}\left\{J_\theta[e(n)]^{-1}\right\} \Leftrightarrow \max\ \mathrm{Det}\left\{J_\theta[e(n)]\right\}$

- D-optimal design:

  - $\boldsymbol{p}_* = (\frac{1}{n}, \dots, \frac{1}{n})$ (uniform distribution)

  - $\boldsymbol{u}_* = (u_1, \dots, u_n)$: any ONB for $\mathbb{R}^n$

  PVM about $L_u = \sum_{i,j} u_i (J_\theta^{\mathrm{SLD}})_{ij}^{-1/2} L_{\theta,j}$

- $J_\theta[e_D^*] = \frac{1}{n} J_\theta^{\mathrm{SLD}}$

- Large degrees of freedoms in optimal designs.

# E-Optimal Design

- Min. of maximum eigenvalue of $J_\theta[(\boldsymbol{p}, \boldsymbol{u})]^{-1}$

  $\Leftrightarrow$ Max. of minimum eigenvalue of $J_\theta[(\boldsymbol{p}, \boldsymbol{u})]$
- E-optimal design:

  - $\boldsymbol{p}_* = (p_i)$ with

$$p_i = \frac{\lambda_i^{\mathrm{SLD}\,-1}}{\sum_{j=1}^{n} \lambda_j^{\mathrm{SLD}\,-1}},$$

  - $\boldsymbol{u}_* = (u_1^{\mathrm{SLD}}, \ldots, u_n^{\mathrm{SLD}})$
- $J_\theta[e_E^*] = \frac{1}{\mathrm{Tr}\left\{(J_\theta^{\mathrm{SLD}\,-1})\right\}} \mathbb{I}_n$

## Comparison of Optimal Designs

A-optimal design: $\min \mathrm{Tr}\left\{J^{-1}\right\}$

D-optimal design: $\min \mathrm{Det}\left\{J^{-1}\right\}$

E-optimal design: $\min \lambda_{\max}(J^{-1})$

# Efficiency 1

- Given two optimality functions $\Psi_1, \Psi_2$ (e.g., $\Psi_1$: A-opt, $\Psi_2$: D-opt)

$$\boldsymbol{e}_1^* = \arg\min \Psi_1(J(\boldsymbol{e}))$$
$$\boldsymbol{e}_2^* = \arg\min \Psi_2(J(\boldsymbol{e}))$$

- $\boldsymbol{e}_1^*$ may not be optimal for $\Psi_2$.
- $\boldsymbol{e}_2^*$ may not be optimal for $\Psi_1$.

- Efficiency of a design $\boldsymbol{e}$ with respect to $\Psi$ optimality:

$$\eta_\Psi[\boldsymbol{e}] := \frac{\Psi\big(J(\boldsymbol{e}^*)\big)}{\Psi\big(J(\boldsymbol{e})\big)} \quad \text{with } \boldsymbol{e}^* = \arg\min \Psi(J(\boldsymbol{e}))$$

- Comparison of different optimal designs.

$$0 \le \eta_\Psi[\boldsymbol{e}] \le 1 \quad \big(\text{Note } \eta_\Psi = 1 \not\Rightarrow J(\boldsymbol{e}) = J(\boldsymbol{e}^*)\big)$$

# Comparison of Optimal Designs

General qubit model: $\rho_\theta = \frac{1}{2}\begin{pmatrix} 1+s_{\theta,3} & s_{\theta,1}-\mathrm{i}s_{\theta,2} \\ s_{\theta,1}+\mathrm{i}s_{\theta,2} & 1-s_{\theta,3} \end{pmatrix}$

Example (three parameter family):

$$\rho_\theta = \frac{1}{2}\begin{pmatrix} 1+\theta_3 & \theta_1-\mathrm{i}\theta_2 \\ \theta_1+\mathrm{i}\theta_2 & 1-\theta_3 \end{pmatrix}$$

$$\left(J_\theta^{\mathrm{SLD}}\right)^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \end{pmatrix}$$

$$|\theta|^2 = \sum_k (\theta_k)^2 < 1$$

# Comparison of Optimal Designs 2

- Optimal designs and their information matrices:

$$J_\theta[e_A^*] = \frac{1}{\mathrm{Tr}\left\{\left(J_\theta^{\mathrm{SLD}}\right)^{-1/2}\right\}}\left(J_\theta^{\mathrm{SLD}}\right)^{1/2}$$

$$J_\theta[e_D^*] = \frac{1}{n} J_\theta^{\mathrm{SLD}}$$

$$J_\theta[e_E^*] = \frac{1}{\mathrm{Tr}\left\{\left(J_\theta^{\mathrm{SLD}\,-1}\right)\right\}}\mathbb{I}_n$$

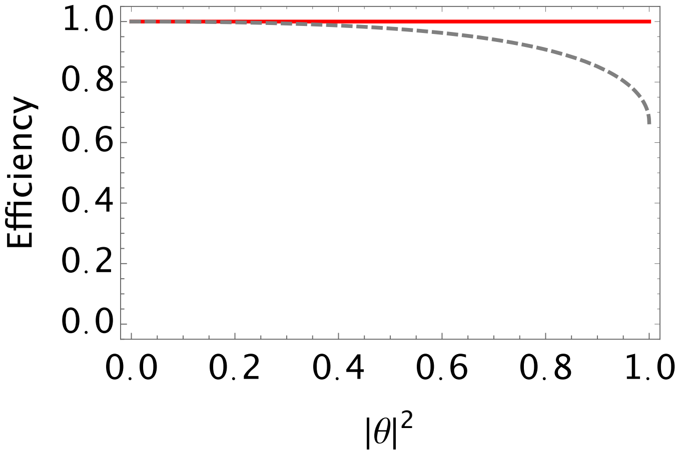
$$J_\theta[e_{ST}] = \left[\frac{1}{3}\sum_{k=1,2,3}\frac{1}{1-s_{\theta,k}^2}\frac{\partial s_{\theta,k}}{\partial \theta_i}\frac{\partial s_{\theta,k}}{\partial \theta_j}\right]$$

$n$: Number of parameters, $\mathbb{I}_n$: $n \times n$ identity matrix

- $e_{ST}$: Design for the standard quantum state tomography

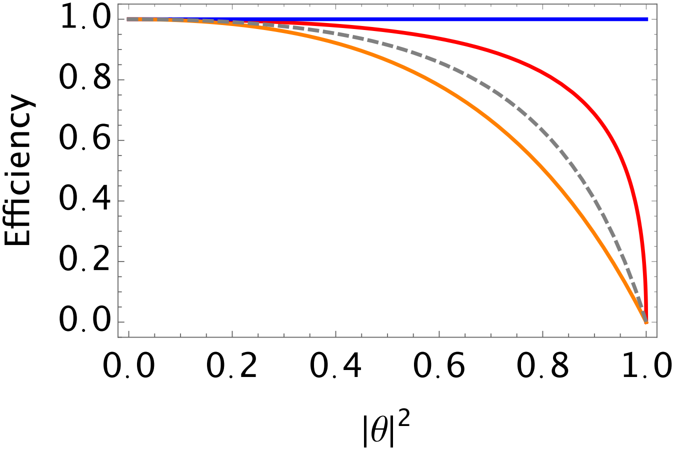
# Efficiency: A-optimality ($\eta_A(e_A^*) = 1$)

$$\eta_A(e_D^*) = \eta_A(e_E^*) = \eta_A(e_{ST}) = \frac{(2+\sqrt{1-|\theta|^2})^2}{3\left(3-|\theta|^2\right)}$$



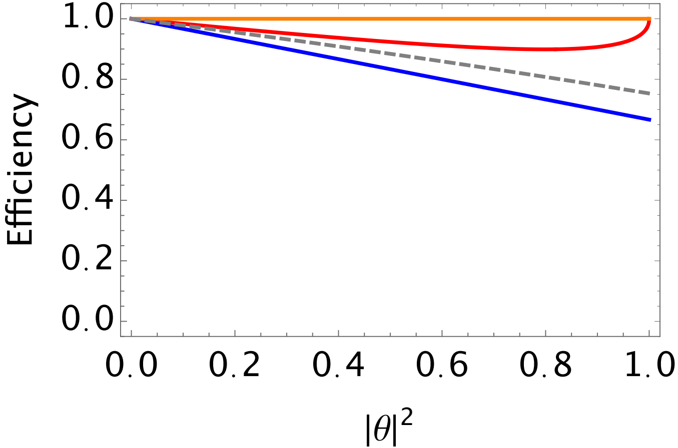
# Efficiency: D-optimality ($\eta_D(e_D^*) = 1$)

$$\eta_D(e_A^*) = \frac{3^3\sqrt{1-|\theta|^2}}{(2+\sqrt{1-|\theta|^2})^3}, \eta_D(e_E^*) = \frac{3^3(1-|\theta|^2)}{\left(3-|\theta|^2\right)^3}, \eta_D(e_{ST}) = \frac{\left(1-|\theta|^2\right)}{\prod_k \left(1-(\theta_k)^2\right)}$$

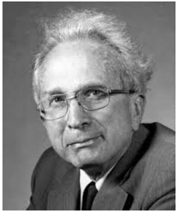


# Efficiency: E-optimality ($\eta_E(e_E^*) = 1$)

$$\eta_E(e_A^*) = \frac{3 - |\theta|^2}{2 + \sqrt{1 - |\theta|^2}}, \eta_E(e_D^*) = \frac{1}{3}(3 - |\theta|^2), \eta_E(e_{ST}) = \frac{3 - |\theta|^2}{3(1 - \min\{(\theta_i)^2\})}$$



# Equivalence Theorem (1960)



Kiefer Wolfowitz

A, c, D, E, G, $\gamma$, I, K, L, M, T, V, $\mathcal{X}$,... (Alphabetic criteria!)

$$\boxed{\text{Equivalence Theorem} \Leftrightarrow \text{X-optimality} = \text{Z-optimality}}$$

# Classical Equivalence Theorem 1

- Given an optimality function $\Psi$
- Directional (Gâteaux) derivative at $\xi_0$ in the direction $\xi$:

$$\Psi'(\xi_0;\xi) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big[\Psi\big((1-\epsilon)J(\xi_0)+\epsilon J(\xi)\big) - J(\xi_0)\Big]$$

assume $\Psi'(\xi_0;\xi) = \int \xi(de)\psi(e,\xi_0)$ "Sensitivity" function

$\xi$: Design measure on the design set $\mathcal{E}$

- General design:

  - $\xi$: probability measure on the design space $\mathcal{E}$

  - Fisher information: $J_\theta[\xi] = \int \xi(de) J_\theta[e]$

- Necessary and sufficient condition:

$$\xi_* = \arg\min_{\xi\in\Xi}\Psi(\xi) \;\Leftrightarrow\; \forall\, e\in\mathcal{E}\;\psi(e,\xi_*)\geq 0 \Leftrightarrow\; \min_{e\in\mathcal{E}}\psi(e,\xi_*)\geq 0$$

# Classical Equivalence Theorem 2

- **Sensitivity function** $\varphi$ for the optimality function $\Psi$:

$$\varphi(e,\xi) := -\psi(e,\xi) + C(\xi),$$

- Kiefer-Wolfowitz theorem (1960):

  D-opt $=$ G-opt for linear regression models

## Theorem (Fedorov)

*The following design problems are equivalent.*

$$\begin{aligned} &1) \quad \min_{\xi\in\Xi} \Psi(\xi),\\ &2) \quad \min_{\xi\in\Xi} \max_{e\in\mathcal{E}} \varphi(e,\xi),\\ &3) \quad \max_{e\in\mathcal{E}} \varphi(e,\xi) = C(\xi). \end{aligned}$$

# Quantum Equivalence Theorem

## Theorem (JS 19)

*For qubit models, the following design problems are equivalent.*

$$1) \quad \min_{\xi \in \Xi} \mathrm{Det}\left\{ J(\xi)^{-1} \right\}$$

$$2) \quad \min_{\xi \in \Xi} \mathrm{Tr}\left\{ J_\theta^{\mathrm{SLD}} J(\xi)^{-1} \right\}$$

- D-optimality = A-optimality with a specific weight matrix

$$\Psi_D(J) = \mathrm{Det}\left\{ J^{-1} \right\}$$
$$\Psi_A(J) = \mathrm{Tr}\left\{ J_\theta^{\mathrm{SLD}} J^{-1} \right\}$$

# Optimal design of experimentsfor quantum channel estimation

- Detection of noise asymmetry in Pauli channel-

Gazit, Ng, JS, *Quantum process tomography via optimal design of experiments* (2019)

# Quantum Process Tomography as DoE

- Quantum channel estimation problem can be phrased as DoE.

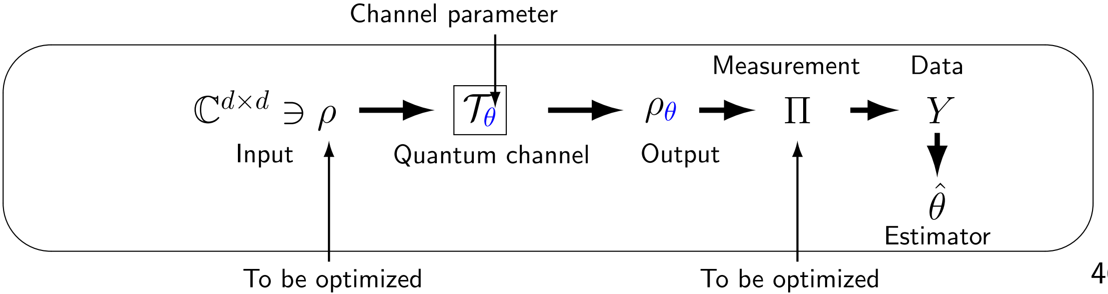
① Model ⇔ Family of quantum channels $\mathcal{M}^Q = \{\mathcal{T}_\theta\}$

② Design $e$ ⇔ Input+measurement $[e = (\rho, \Pi)]$

③ Model function $f$ ⇔ Born's rule: $(\theta, e) \mapsto p_\theta(x|e) = \mathrm{tr}\left(\mathcal{T}_\theta(\rho)\Pi_x\right)$

- Optimization over input+meas. about $D(\rho, \Pi) = \Big[J_\theta[\rho, \Pi]\Big]^{-1}$

  $J_\theta[\rho, \Pi] = J_\theta[p_\theta(\cdot|\Pi)]$: Fisher information matrix about $p_\theta(\cdot|\Pi)$



# Qubit Pauli Channel

- $\mathcal{T}_\theta(\rho) = (1 - \sum_{i=1}^{3} \theta_i)\rho + \sum_{i=1}^{3} \theta_i \sigma_i \rho \sigma_i$
- To simplify, consider a two-parameter subfamily with $\theta_3 = 0$
- Noise asymmetry: $\vartheta_1 := \theta_1 - \theta_2$ (Parameter of interest)
- Noise strength: $\vartheta_2 := 1 - (\theta_1 + \theta_2)$ (Nuisance parameter)
- Objective: to find an optimal design $e(m) = (\boldsymbol{\nu}, \boldsymbol{\rho}, \boldsymbol{\Pi})$
  $m = 1, 2$ designs to get analytical expressions.
- $e_{\mathrm{PT}}(i) = \big(\rho(i), \Pi(i)\big)$: Standard Pauli tomography $(i = 1, 2, 3)$

# Optimal i.i.d. strategy ($m=1$: fixed probe state)

- Result:

$$\min_{s_1,s_2:\sum_i s_i^2=1} \left(J_\vartheta^{SLD}[e(1)]^{-1}\right)_{11} = \min\{f_1^2\,,\ f_2^2\} \text{ with } f_{1,2} = \frac{1}{2}\sqrt{1-(\vartheta_1 \pm \vartheta_2)^2}$$

$$\text{Optimal design: } e_* = \begin{cases} e_{\mathrm{PT}}(1) & \vartheta_1 > 0 \\ e_{\mathrm{PT}}(2) & \vartheta_1 \le 0 \end{cases}$$

$e_{\mathrm{PT}}(i) = \big(\rho(i), \Pi(i)\big)$: Standard Pauli tomography ($i = 1, 2, 3$)

- Optimal design $e_*$ depends on $\vartheta_1$!

$\Rightarrow$ **Local optimal** design and **singular** design problem

JS, IJQI (2021)

# Optimal mixed strategy ($m = 2$: two probe states)

- Relative frequency: $\boldsymbol{\nu} = \left(\frac{1}{2}(1+\lambda), \frac{1}{2}(1-\lambda)\right)$, $\qquad \lambda \in [-1, 1]$

  Optimal frequency: $\lambda^* = \frac{f_1 - f_2}{f_1 + f_2} \Rightarrow \boldsymbol{\nu}_* = \left(\frac{f_1}{f_1+f_2}\ \frac{f_2}{f_1+f_2}\right)$, $\quad f_{1,2} = \frac{1}{2}\sqrt{1 - (\vartheta_1 \pm \vartheta_2)^2}$
- Optimal design $e_*(2) = (\boldsymbol{\nu}_*, \boldsymbol{e}_*)$ with $\boldsymbol{e}_* = \Big(e_{\mathrm{PT}}(1)\,,\, e_{\mathrm{PT}}(2)\Big)$

  $e_{\mathrm{PT}}(i) = \big(\rho(i), \Pi(i)\big)$: Standard Pauli tomography ($i = 1, 2, 3$)

- $\boldsymbol{\nu}_*$ depends on $\vartheta_1$ and $\vartheta_2$!!
- $\vartheta_1$ is the parameter of interest.
- But, nuisance parameter $\vartheta_2$ needs to be estimated as well.

# Adaptive Design

- Adaptive (sequential) design:
  A common remedy for the local optimal design problem.
- Is it useful for our problem in the presence of a nuisance parameter?

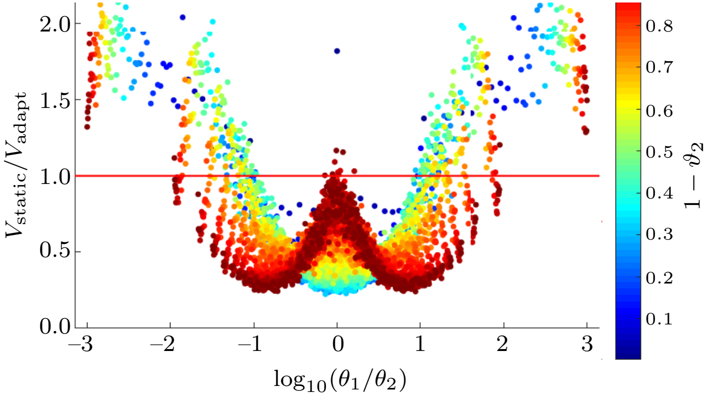
- To implement the next design based on the previous estimates.
- Various methods, mathematical proofs...

E.g. Two-step (two-stage) method
(Hayashi-Matsumoto 1998, Barndorff-Nielsen-Gill 2000, Gill-Massar 2000)
Step-wise update: Nagaoka's method based on MLE (Nagaoka 1989, Fujiwara 2006)

# An Algorithm: $K$-step Adaptation

1) Split $N$ channel use into $K$ adaptive steps. $(N = \sum_{k=1}^{K} N_k)$
2) Decide on the relative proportion of $e_1$ and $e_2$ at each step.
   $N_{k,1} \equiv \frac{1}{2}(1+\lambda_k)N_k$ use for $e_1$, $N_{k,2} \equiv \frac{1}{2}(1-\lambda_k)N_k$ for $e_2$.
3) Make estimate $\hat{\vartheta}_{1,2} \Leftrightarrow \hat{f}_{1,2} \Rightarrow$ To update $\lambda_k$
   $$\widehat{f}_{k,i} = \left[\frac{n_{1:k,i}}{N_{1:k,i}}\left(1 - \frac{n_{1:k,i}}{N_{1:k,i}}\right)\right]^{1/2} \qquad (n_{1:k,i}\text{: Cumulative counts})$$
4) Update the knowledge on $\lambda_k$ based on the all outcomes.
   $$N_{k+1,1} = \left[\frac{\widehat{f}_{k,1}}{\widehat{f}_{k,1}+\widehat{f}_{k,2}} N_{1:k+1} - N_{1:k,1}\right]_+ , \ ([x]_+\text{: positive part})$$
   $$\lambda_{k+1} = 2\left[\frac{\widehat{f}_{k,1}+\widehat{f}_{k,1}\frac{N_{1:k,1}}{N_{k+1}}-\widehat{f}_{k,2}\frac{N_{1:k,2}}{N_{k+1}}}{\widehat{f}_{k,1}+\widehat{f}_{k,2}}\right]_+ - 1$$

# Numerical Result ($N = 200, K = 10$)




MSE ratios: $\frac{V_{\mathrm{static}}}{V_{\mathrm{adapt}}} = (1 - \lambda_{\mathrm{eff}}^2)\left[-\lambda_{\mathrm{eff}} \frac{f_1^2 - f_2^2}{f_1^2 + f_2^2} + 1\right]^{-1}$ $(\lambda_{\mathrm{eff}} = \frac{2N_{1:K,1}}{N} - 1)$, $V_{\mathrm{static}}$: Standard Pauli tomography

Benefit from adaptation $\Leftrightarrow$ $V_{\mathrm{static}} > V_{\mathrm{adapt}}$ $\Leftrightarrow$ High asymmetry region (large $|\log_{10}(\theta_1/\theta_2)|$)

# Runway Method

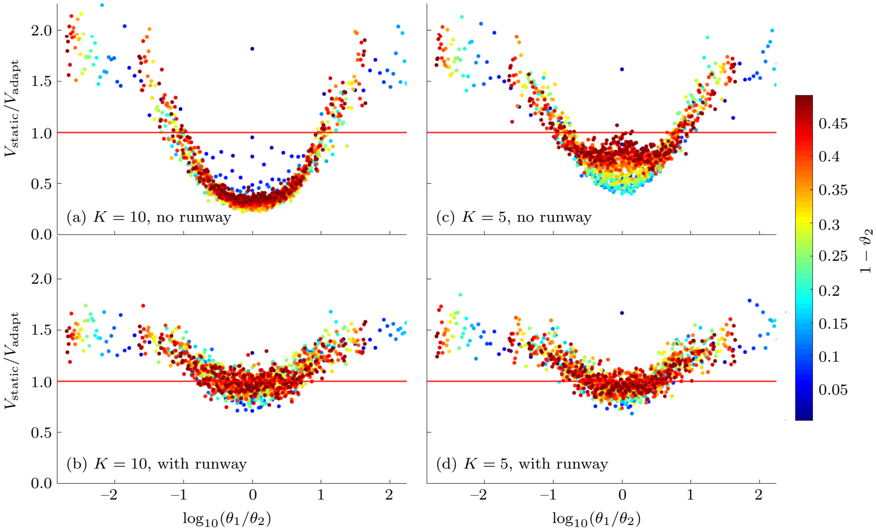
- Modify our adaptive scheme to include an initial "runway":
  With fixed equal weight between the $e_1$ and $e_2$ designs.

- The MSE ratios for

(a) $K = 10$, no runway
(b) $K = 10$, a runway with $N/2 = 100$ + Adaptation for $N/2 = 100$
(c) $K = 5$, no runway
(d) $K = 5$, a runway with $N/2 = 100$ + Adaptation for $N/2 = 100$

Small noise is of interest. $\Rightarrow$ Only the $1 - \vartheta_2 \leq 0.5$ data

# Comparison ($N = 200$)



## Observations

- $\dfrac{V_{\text{static}}}{V_{\text{adapt}}} > 1 \Leftrightarrow$ Gain from adaptation
- Benefit is visible only for high asymmetry.
- Different structures for strong and weak noise cases.
- Adaptation performs bad for low asymmetry...
- Runway method works!

- High asymmetry $\Leftrightarrow$ Near boundaries $(\vartheta_2 \simeq 1 - |\vartheta_1|)$
  Low asymmetry $\Leftrightarrow$ Interior regions $(\vartheta_1 \simeq 0)$

# Recent Results

- The theory of DoE = optimization about the classical Fisher information
- An alternative approach: to minimize the MSE directly

  Known lower bounds:

| | |
|---|---|
| 1967 | Helstrom, SLD bound |
| 1972 | Yuen-Lax, RLD bound |
| 1976 | Holevo bound |
| 1989 | Nagaoka bound (two parameter only) |
| 1996 | Petz, monotone metric |
| 2021 | Generalized Nagaoka bound (Tightest bound for A-optimal design so far) — Conlon, JS, Lam, Assad, npj QI (2021) |

Cf. Bayesian setting: JS, 2302.14223 (2023)

# Summary & Outlook

- General formulation optimal design of experiments in quantum system
- A systematic and powerful tool for designing quantum process & state tomography
  1. Popular optimal designs (A, c, D, E, $\gamma$)
  2. Quantum equivalence theorem (Qubit case only)
  3. Detection of noise asymmetry in the qubit Pauli channel
     - Nuisance parameter problem
     - Adaptation does not work well for small samples
- DoE also applies to collective POVM and entangled probe states.

- More on DoE for quantum systems? YES
- Information geometry, Bayesian design, numerical algorithm, etc