Fisher Information in NISQ Applications

JOHANNES JAKOB MEYER, FU BERLIN & QMATH

@jj_xyz

Based on arXiv:2103.15191

Parametrized Quantum States

Distances Between Parametrized States

We want to understand parametrized quantum states in *parameter space*. One way to do so is via *distances*

→ Perform a *pullback* of a distance between quantum states

$$d(\boldsymbol{\theta}, \boldsymbol{\theta}') = d(\rho(\boldsymbol{\theta}), \rho(\boldsymbol{\theta}'))$$

→ or between the output probability distributions

$$d_{\mathcal{M}}(\boldsymbol{\theta}, \boldsymbol{\theta}') = d(p_{\mathcal{M}}(\boldsymbol{\theta}), p_{\mathcal{M}}(\boldsymbol{\theta}'))$$

We require $d(\boldsymbol{\theta}, \boldsymbol{\theta}') \geq 0$, $d(\boldsymbol{\theta}, \boldsymbol{\theta}) = 0$, *differentiability* and *monotonicity*.

Small Changes

What happens in the local neighbourhood of the parameter $\boldsymbol{\theta}$?

→ Let's look at perturbation of pullback distance – a Taylor expansion gives

$$d(\boldsymbol{\theta}, \boldsymbol{\theta} + \boldsymbol{\delta}) = \frac{1}{2}\boldsymbol{\delta}^T M(\boldsymbol{\theta})\boldsymbol{\delta} + O(\|\boldsymbol{\delta}\|^3)$$

with the *Hessian*

$$M(\boldsymbol{\theta})_{ij} = \frac{\partial^2}{\partial \delta_i \partial \delta_j} d(\boldsymbol{\theta}, \boldsymbol{\theta} + \boldsymbol{\delta})\bigg|_{\boldsymbol{\delta}=0}$$

Information Matrices

The Hessian induces an inner product $\langle \boldsymbol{\delta}, \boldsymbol{\delta}' \rangle_M = \boldsymbol{\delta}^T M \boldsymbol{\delta}'$

With it we can measure

lengths $\quad \|\boldsymbol{\delta}\| = \sqrt{\langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle}$

distances $\quad d(\boldsymbol{\delta}, \boldsymbol{\delta}') = \|\boldsymbol{\delta} - \boldsymbol{\delta}'\| = \sqrt{\langle \boldsymbol{\delta} - \boldsymbol{\delta}', \boldsymbol{\delta} - \boldsymbol{\delta}' \rangle}$

angles $\quad \sphericalangle(\boldsymbol{\delta}, \boldsymbol{\delta}') = \arccos(\langle \boldsymbol{\delta}, \boldsymbol{\delta}' \rangle / \|\boldsymbol{\delta}\| \|\boldsymbol{\delta}'\|)$

The Hessian contains information about the underlying quantum state, by nature an information-theoretic object. We thus call it *information matrix*.

Classical Fisher Information Matrix

Use the KL divergence (relative entropy) for the pullback

$$d_{\mathrm{KL}}(p(\boldsymbol{\theta}), p(\boldsymbol{\theta}')) = \sum_l p_l(\boldsymbol{\theta}) \log \frac{p_l(\boldsymbol{\theta})}{p_l(\boldsymbol{\theta}')}$$

And find the *classical Fisher information matrix (CFIM)* with entries

$$I_{ij}(\boldsymbol{\theta}) = \sum_l \frac{1}{p_l(\boldsymbol{\theta})} \frac{\partial p_l(\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial p_l(\boldsymbol{\theta})}{\partial \theta_j}$$

Intuition: the classical Fisher information quantifies how much a change of parameters changes the underlying probability distribution

Uniqueness

What happens if we perform the same with the Bhattacharyya distance?

$$d_\mathrm{B}(p(\boldsymbol{\theta}), p(\boldsymbol{\theta}')) = 1 - \sum_l \sqrt{p_l(\boldsymbol{\theta})p_l(\boldsymbol{\theta}')}$$

We again find the classical Fisher information matrix times a constant!

Theorem (Morozova/Chentsov):

The information matrix associated to any monotonic distance measure between probability distributions will be a positive scalar multiple of the classical Fisher information matrix

Monotonicity: $d(T[p], T[q]) \leq d(p, q)$ for all stochastic maps T

Calculation

To calculate the classical Fisher information matrix, we need

output probabilities
$$p_l(\boldsymbol{\theta})$$

→ Output of experiment (histogram)

→ Can also use more sophisticated techniques: Bayesian approach, Machine Learning

their derivatives
$$\partial_i p_l(\boldsymbol{\theta})$$

→ Finite differences

→ Parameter shift rule

Quantum Fisher Information Matrix (Pure States)

Use twice the fidelity distance for the pullback

$$2d_f(|\psi(\boldsymbol{\theta})\rangle, |\psi(\boldsymbol{\theta}')\rangle) = 2 - 2|\langle\psi(\boldsymbol{\theta})|\psi(\boldsymbol{\theta}')\rangle|^2$$

And find the *quantum Fisher information matrix (QFIM)* with entries

$$\mathcal{F}_{ij}(\boldsymbol{\theta}) = 4\,\mathrm{Re}[\langle\partial_i\psi(\boldsymbol{\theta})|\partial_j\psi(\boldsymbol{\theta})\rangle - \langle\partial_i\psi(\boldsymbol{\theta})|\psi(\boldsymbol{\theta})\rangle\langle\psi(\boldsymbol{\theta})|\partial_j\psi(\boldsymbol{\theta})\rangle]$$

Intuition: the quantum Fisher information quantifies how much a change of parameters changes the underlying quantum state

Non-Uniqueness

Theorem (Petz):

There are infinitely many information matrices arising from monotonic distances between quantum states.

Monotonicity: $d(\Phi[\rho], \Phi[\sigma]) \leq d(\rho, \sigma)$ for all quantum channels Φ

„Our" quantum Fisher information matrix is also known as *SLD-QFIM,* because it can also be defined via the *symmetric logarithmic derivative* operators

$$\mathcal{F}_{ij} = \frac{1}{2} \text{Tr}\{\rho(L_i L_j + L_j L_i)\} \qquad \frac{\partial \rho}{\partial \theta_i} = \frac{1}{2}(L_i \rho + \rho L_i)$$

Calculation (Pure States)

For parameters are encoded via parallel Hamiltonian evolution

$$\mathcal{F}_{ij}(\boldsymbol{\theta}) = 4[\langle\psi(\boldsymbol{\theta})|\frac{\{H_i, H_j\}}{2}|\psi(\boldsymbol{\theta})\rangle - \langle\psi(\boldsymbol{\theta})|H_i|\psi(\boldsymbol{\theta})\rangle\langle\psi(\boldsymbol{\theta})|H_j|\psi(\boldsymbol{\theta})\rangle]$$

We can also use perturbations of the fidelity distance itself and tricks using the parameter-shift rule. To get approximation on the matrix level, an analogue of SPSA can be used.

Classical and Quantum Fisher Information

Monotonicity of the underlying distance measure implies that

$$I(T[p(\boldsymbol{\theta})]) \leq I(p(\boldsymbol{\theta})) \quad \text{for all stochastic maps } T$$

$$\mathcal{F}(\Phi[\rho(\boldsymbol{\theta})]) \leq \mathcal{F}(\rho(\boldsymbol{\theta})) \quad \text{for all quantum channels } \Phi$$

Classical and Quantum Fisher Information

Monotonicity of the underlying distance measure implies that

$$I(T[p(\boldsymbol{\theta})]) \leq I(p(\boldsymbol{\theta})) \quad \text{for all stochastic maps } T$$
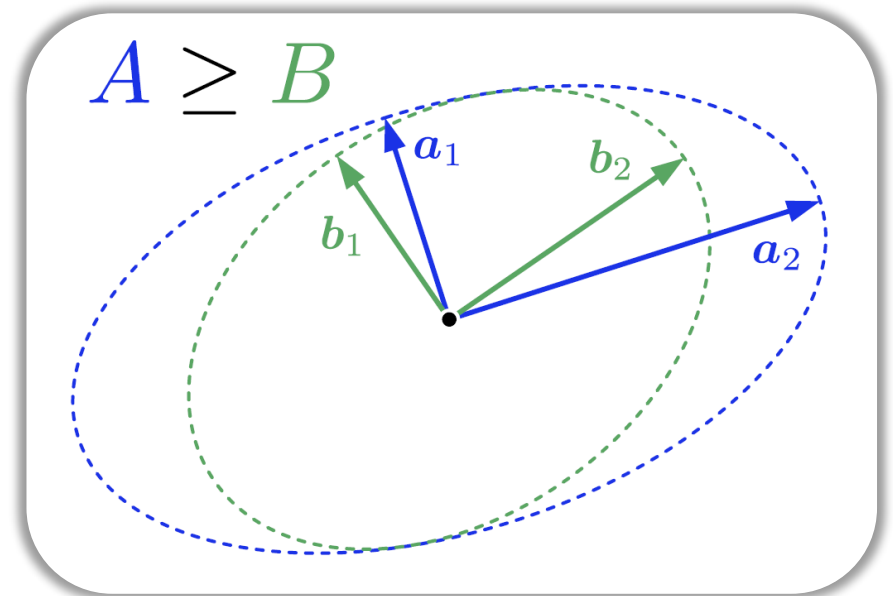
$$\mathcal{F}(\Phi[\rho(\boldsymbol{\theta})]) \leq \mathcal{F}(\rho(\boldsymbol{\theta})) \quad \text{for all quantum channels } \Phi$$

Classical and Quantum Fisher Information

Monotonicity of the underlying distance measure implies that

$$I(T[p(\boldsymbol{\theta})]) \leq I(p(\boldsymbol{\theta})) \quad \text{for all stochastic maps } T$$

$$\mathcal{F}(\Phi[\rho(\boldsymbol{\theta})]) \leq \mathcal{F}(\rho(\boldsymbol{\theta})) \quad \text{for all quantum channels } \Phi$$

But measurements are also channels and the quantum Fisher information for a classical state is equal to the classical Fisher information. Therefore

$$\mathcal{F}(\mathcal{M}[\rho(\boldsymbol{\theta})]) = I(\mathcal{M}[\rho(\boldsymbol{\theta})]) \leq \mathcal{F}(\rho(\boldsymbol{\theta}))$$

The quantum Fisher information is therefore an upper bound for any classical Fisher information matrix arising from a measurement.

Roles of Classical and Quantum Fisher Information

$$\mathcal{F}(\boldsymbol{\theta})$$

QUANTUM FISHER INFORMATION

Quantifies ultimate limits for a specific underlying state

Can tell us a lot about the quantum effects influencing our experiments

$$I(\boldsymbol{\theta})$$

CLASSICAL FISHER INFORMATION

Quantifies behaviour for a fixed measurement

Extremely relevant in practice, as we always have to fix some sort of measurement

Quantum Fisher Information (Noisy States)

Noise decreases the quantum Fisher information due to monotonicity.

For mixed states, we use the *Bures distance*

$$2d_B(\rho(\boldsymbol{\theta}), \rho(\boldsymbol{\theta}')) = 2 - 2\operatorname{Tr}\left\{\sqrt{\sqrt{\rho(\boldsymbol{\theta})}\rho(\boldsymbol{\theta}')\sqrt{\rho(\boldsymbol{\theta})}}\right\}^2$$

The resulting quantum Fisher information for a state $\rho = \sum_k \lambda_k |\lambda_k\rangle\langle\lambda_k|$ is

$$\mathcal{F}_{ij} = \sum_{\substack{kl \\ \lambda_k + \lambda_l \neq 0}} \frac{2\operatorname{Re}(\langle\lambda_k|\partial_i\rho|\lambda_l\rangle\langle\lambda_l|\partial_j\rho|\lambda_k\rangle)}{\lambda_k + \lambda_l}$$

Quantum Fisher Information (Noisy States)

We can „simplify" this further

$$\mathcal{F}_{ij} = \sum_{\substack{k \\ \lambda_k \neq 0}} \frac{(\partial_i \lambda_k)(\partial_j \lambda_k)}{\lambda_k}$$

Classical Part

Quantum Part

$$+ 4\lambda_k \, \mathrm{Re}(\langle \partial_i \lambda_k | \partial_j \lambda_k \rangle)$$

$$- \sum_{\substack{kl \\ \lambda_k, \lambda_l \neq 0}} \frac{8\lambda_k \lambda_l}{\lambda_k + \lambda_l} \mathrm{Re}(\langle \partial_i \lambda_k | \lambda_l \rangle \langle \lambda_l | \partial_j \lambda_k \rangle)$$

Calculation (Noisy States)

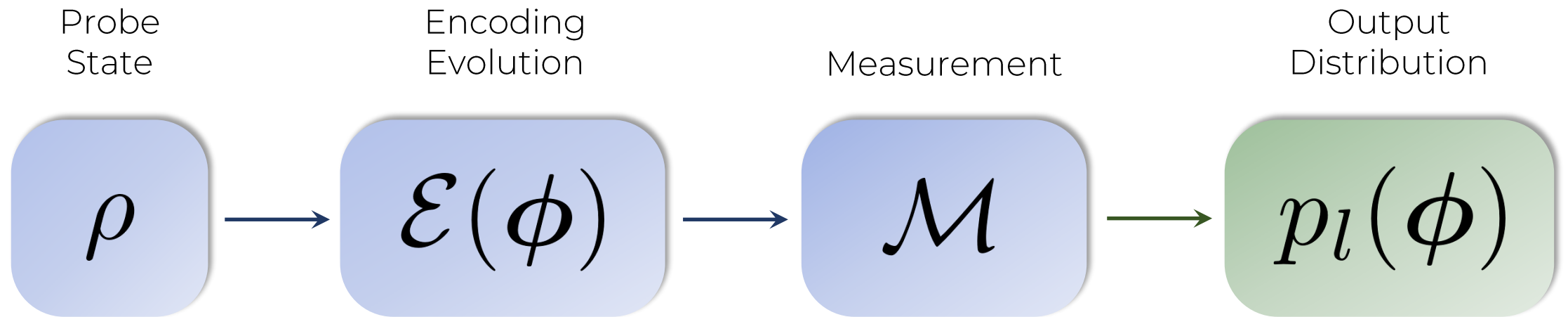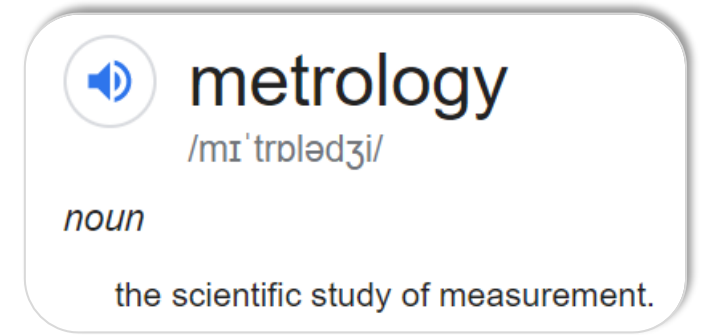In principle, full tomography is necessary to calculate the quantum Fisher information for noisy states

Variational methods have been proposed, but with high overheads and relying on the success of variational subroutines

An alternative is to approximate the quantum Fisher information, *e.g.* via the *truncated* quantum Fisher information
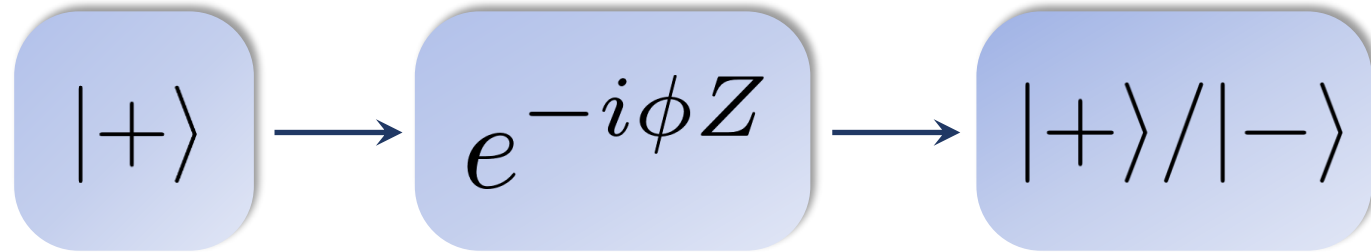
Application 1: Quantum Metrology

Physical quantities (magnetic fields, energies, ...) need to be **measured** accurately

Study how **quantum effects** can help
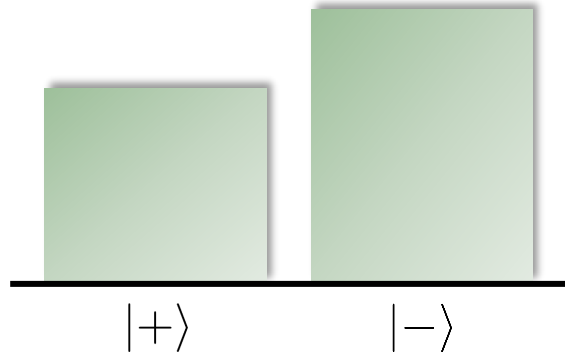
metrology
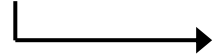/mɪˈtrɒlədʒi/
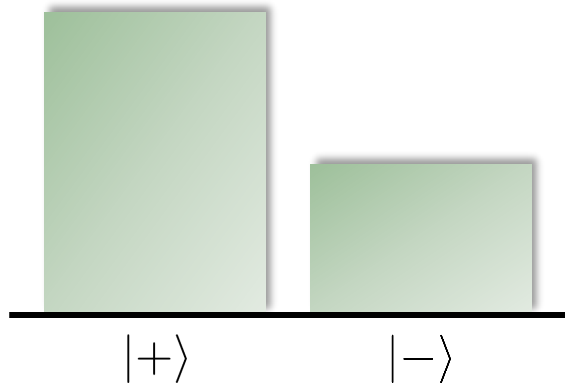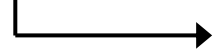
noun

the scientific study of measurement.

Probe State	Encoding Evolution	Measurement	Output Distribution
ρ	$\mathcal{E}(\phi)$	$\mathcal{M}$	$p_l(\phi)$

Gathering Intuition
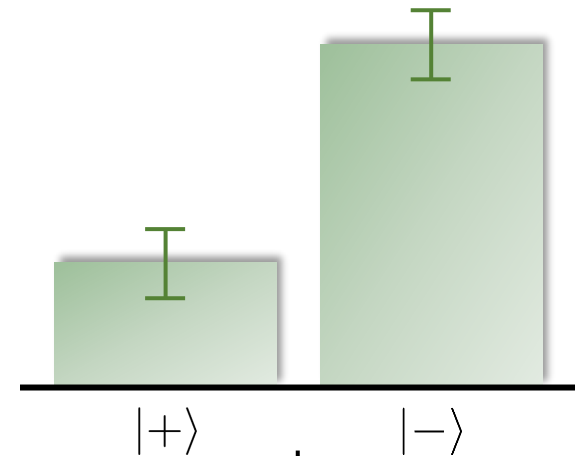
Cramér-Rao Bound

Formally, we construct an *estimator* for the physical quantity from the output probability distribution

$$p(\boldsymbol{\theta}) \mapsto \hat{\boldsymbol{\varphi}} \quad \text{unbiased if} \quad \mathbb{E}\{\hat{\boldsymbol{\varphi}}\} = \boldsymbol{\phi}$$

The *Cramér-Rao bound* limits the precision of any unbiased estimator

$$\text{Cov}[\hat{\boldsymbol{\varphi}}] \geq \frac{1}{n} I_{\mathcal{M}}(\boldsymbol{\phi})^{-1} \geq \frac{1}{n}\mathcal{F}(\boldsymbol{\phi})^{-1}$$

$$\text{Tr}\{\text{Cov}[\hat{\boldsymbol{\varphi}}]\} = \text{MSE}[\hat{\boldsymbol{\varphi}}]$$

Standard Quantum and Heisenberg Limit

Repeating the same experiment n times results in a scaling

$$\mathrm{Cov}[\hat{\varphi}] \propto \frac{1}{n} \quad \textbf{STANDARD QUANTUM LIMIT (SQL)}$$

With entanglement, we can reach

$$\mathrm{Cov}[\hat{\varphi}] \propto \frac{1}{n^2} \quad \textbf{HEISENBERG LIMIT (HL)}$$

This advantage is not resistant to noise. But it can be recovered using *metrological codes.*

Application 2: Quantum Natural Gradient

Gradient descent is a general-purpose method to minimize a cost function

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta \nabla C(\boldsymbol{\theta}^{(t)})$$

We can reformulate the gradient update as an optimization problem

$$\boldsymbol{\theta}^{(t+1)} = \underset{\boldsymbol{\vartheta}}{\mathrm{argmin}} \left\{ \boxed{\langle \boldsymbol{\vartheta} - \boldsymbol{\theta}^{(t)}, \nabla C(\boldsymbol{\theta}^{(t)}) \rangle} + \boxed{\frac{1}{2\eta} \|\boldsymbol{\vartheta} - \boldsymbol{\theta}^{(t)}\|_2^2} \right\}$$

Minimal for update opposite to gradient

Regularisation to avoid overstepping

Quantum Natural Gradient Update

But we started this talk realizing that measuring distances between parmeters makes more sense if we take the pullback of a distance between quantum states! Replacing
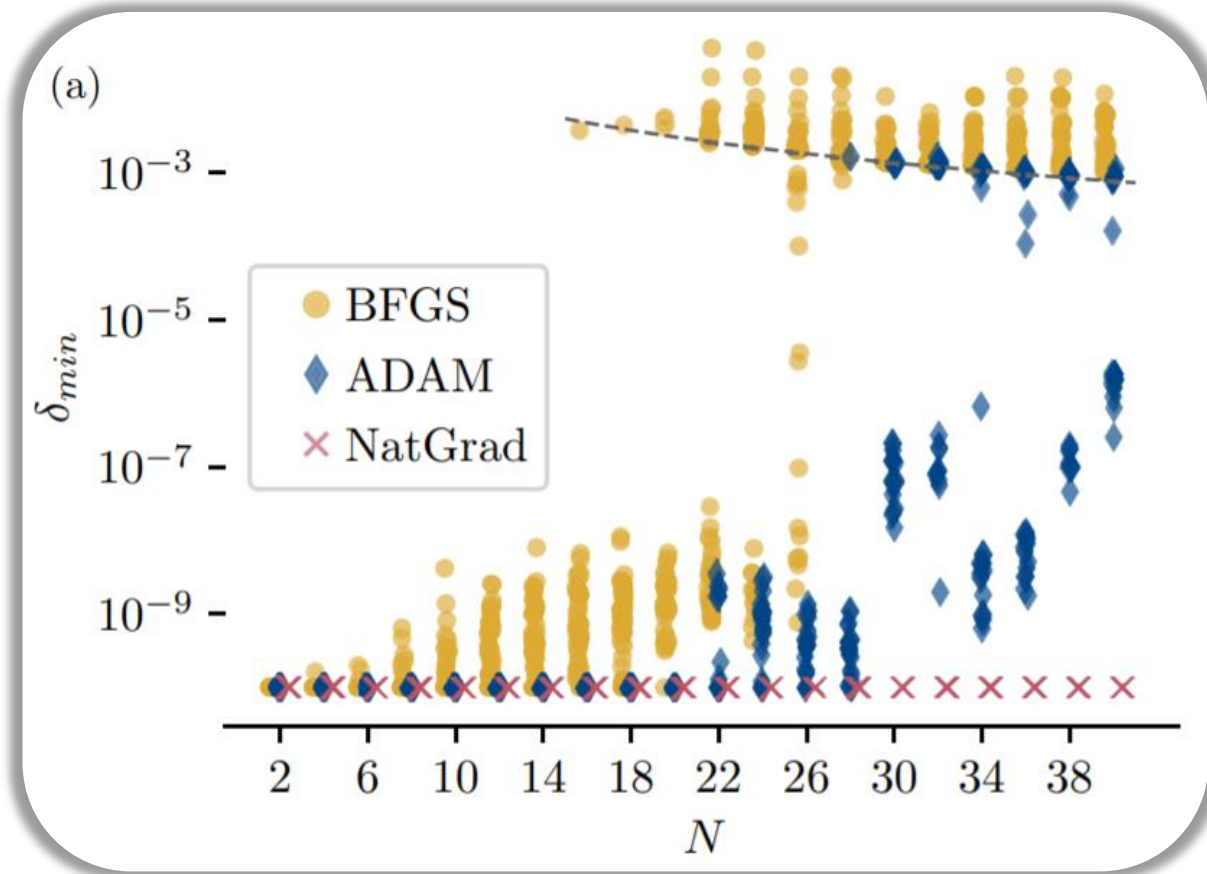
$$\|\boldsymbol{\vartheta} - \boldsymbol{\theta}^{(t)}\|_2^2 \quad \text{with} \quad d_f(\boldsymbol{\vartheta}, \boldsymbol{\theta}^{(t)}) \approx (\boldsymbol{\vartheta} - \boldsymbol{\theta}^{(t)})^T \mathcal{F}(\boldsymbol{\theta}^{(t)})(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{(t)})$$

in the optimization yields the *quantum natural gradient* update step

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta \mathcal{F}(\boldsymbol{\theta}^{(t)})^{-1} \nabla C(\boldsymbol{\theta}^{(t)})$$

This update now takes the underlying geometry of the parametrized quantum state into account

Quantum Natural Gradient Helps

Further Reading

There are many great reviews about quantum Fisher information:

J. Liu, H. Yuan, X.-M. Lu, and X. Wang
Quantum fisher information matrix and multiparameter estimation
Journal of Physics A: Mathematical and Theoretical **53**, 023001 (2020)

J. S. Sidhu and P. Kok
Geometric perspective on quantum parameter estimation
AVS Quantum Science **2**, 014701 (2020)

V. Katariya and M. M. Wilde
Geometric distinguishability measures limit quantum channel estimation and discrimination
Quantum Information Processing **20**, 78 (2021)

Outlook

- Fisher information is a very versatile tool that has found a lot of cool applications in various fields, for example theoretical quantum information, error correction and resource theories

- Parametrized quantum states are fundamental to NISQ applications, so I expect many more interesting applications of Fisher information

- Many open questions remain and much more work is to be done!

Some Open Questions

* Analysis of estimators for the classical Fisher information

* A (more) efficient way to calculate the quantum Fisher information in the noisy setting

* Can we use neural network quantum states or other computational approximations to compute the quantum Fisher information?

* Can we reproduce proofs of important theorems using Fisher information?

* Can we use (quantum) Fisher information to understand ansätze and layers for variational quantum algorithms?

* Can we leverage (quantum) Fisher information to better quantify the capabilities of learning models based on parametrized quantum circuits?

Thank you for your attention!

Paper

Slides