On quantum backpropagation, information reuse and cheating measurement collapse

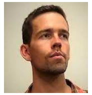
Amira Abbas

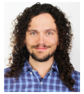
University of Amsterdam, QuSoft, CWI

arXiv 2305.13362, NeurIPS 2023

QTML 2023





Robbie King   Hsin-Yuan Huang   Bill Huggins   Ramis Movassagh   Dar Gilboa   Jarrod McClean

QuSoft
Research Center for Quantum Software

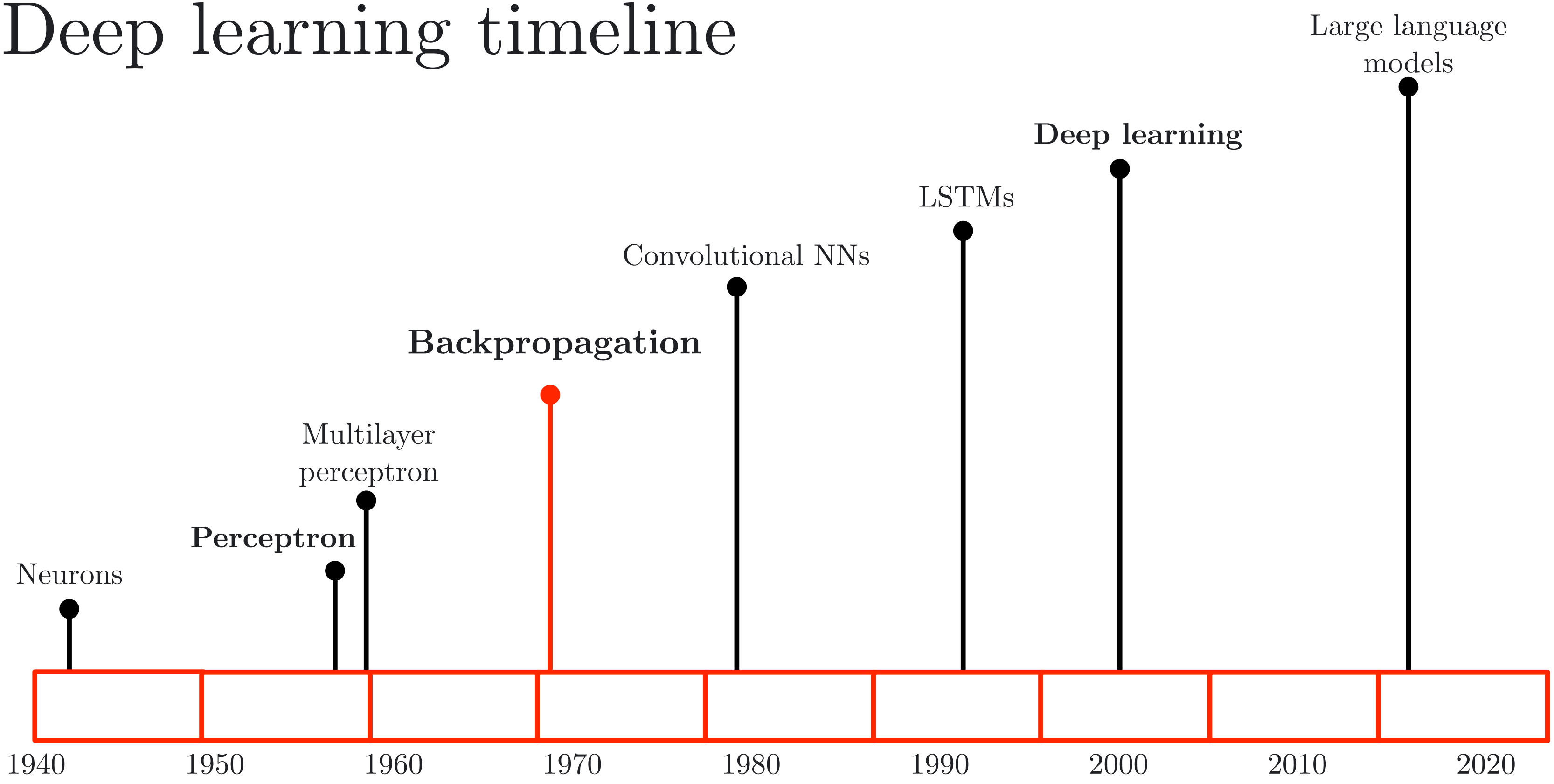
Universiteit van Amsterdam

CWI
Research institute for mathematics & computer science in the Netherlands

Google AI Quantum

What is the key to scaling neural networks in practice?

Deep learning timeline



Types of quantum models

Parameterized quantum models

Parameterized quantum models

Parameterized operations acting on an initial state, measure some Hermitian observable

$$F(\theta) =$$

Parameterized quantum models

Parameterized operations acting on an initial state, measure some Hermitian observable

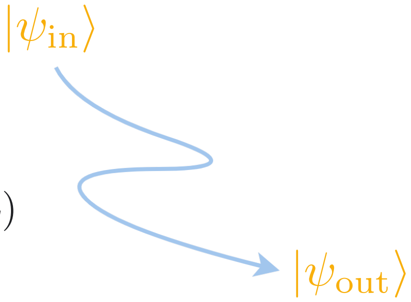
$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

Parameterized quantum models

Parameterized operations acting on an initial state, measure some Hermitian observable

$$F(\theta) = \langle\psi(\theta)|O|\psi(\theta)\rangle$$

- Quantum simulation
  - Simulate the dynamics of some target system
  - Approximate ground state energy/learn ground states (VQE)
  - Chemistry and material science

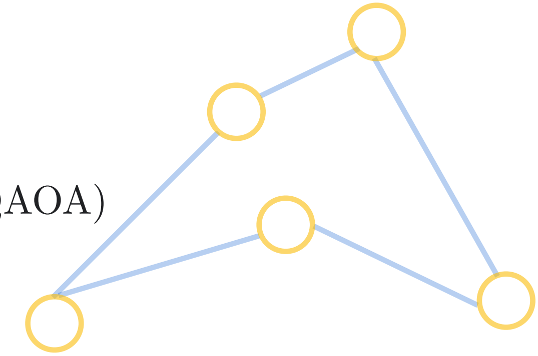


Parameterized quantum models

Parameterized operations acting on an initial state, measure some Hermitian observable

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

- Quantum optimization
  - Approximate solutions to large combinatorial problems (QAOA)
  - Logistics and finance



Parameterized quantum models

Parameterized operations acting on an initial state, measure some Hermitian observable

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

- Quantum machine learning
  - Output label or loss function
  - More expressive/interesting?

Parameterized quantum models

Parameterized operations acting on an initial state, measure some Hermitian observable

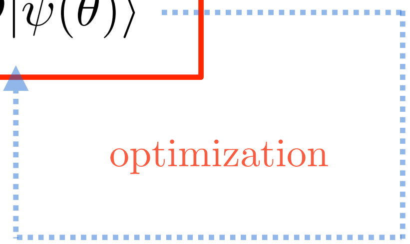
$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

- In short, people care about these parameterized models

Parameterized quantum models

Parameterized operations acting on an initial state, measure some Hermitian observable

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$



- In short, people care about these models
- Need to be able to optimize them

Parameterized quantum models

Parameterized operations acting on an initial state, measure some Hermitian observable

$$F(\theta) = \langle \psi(\theta) | O | \psi(\theta) \rangle$$

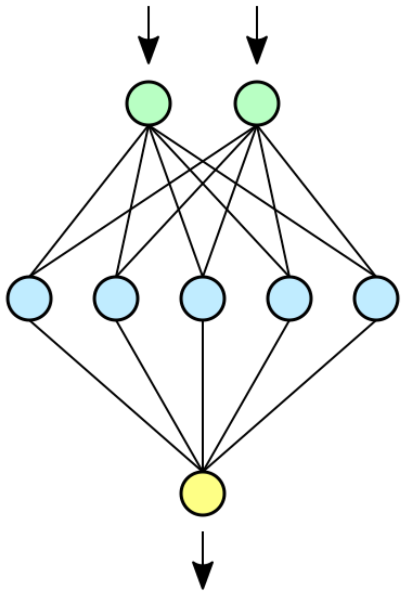
optimization

- In short, people care about these models
- Need to be able to optimize them
- Gradient-based methods

Gradient-based optimization in machine learning

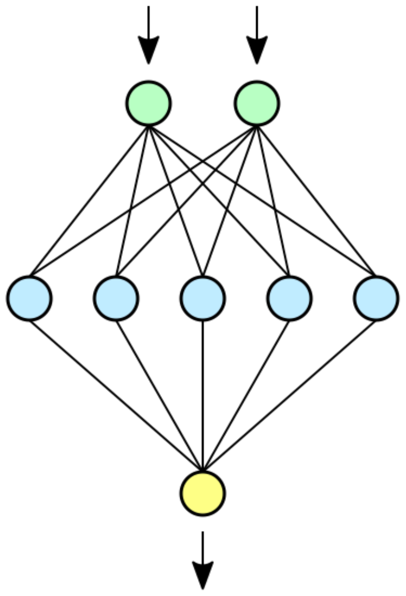
Backpropagation

- Recipe to compute gradients of a function



Backpropagation

- Recipe to compute gradients of a function
- The first computationally efficient method to update parameters of a neural network

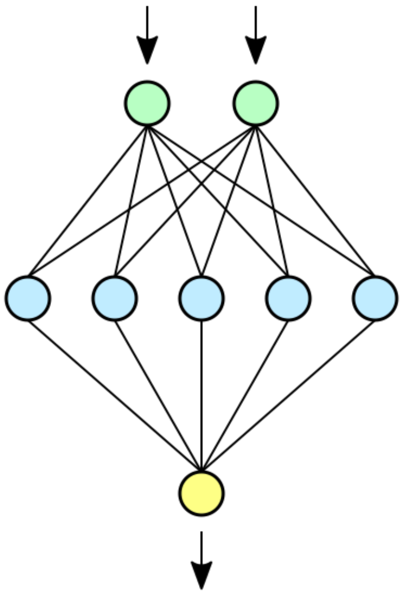


Backpropagation

- Recipe to compute gradients of a function
- The first computationally efficient method to update parameters of a neural network

$$F(\boldsymbol{\theta}, x) = \sigma(\theta_M(\sigma(\theta_{M-1}...\theta_1(x))))$$

- As a neural network function is being computed, intermediate information is cleverly stored and reused for gradient computation
- Not just "the chain rule"



Backpropagation scaling

Backpropagation scaling

Given a parameterized function, $F(\theta),\ \theta \in \mathbb{R}^M$

Griewank, A., & Walther, A. (2008). *Evaluating derivatives: principles and techniques of algorithmic differentiation*. Society for industrial and applied mathematics.

Backpropagation scaling

Given a parameterized function, $F(\theta),\ \theta \in \mathbb{R}^M$

Let $F'(\theta)$ be an estimate of the gradient vector, to within constant ϵ in the infinity norm

Griewank, A., & Walther, A. (2008). *Evaluating derivatives: principles and techniques of algorithmic differentiation*. Society for industrial and applied mathematics.

Backpropagation scaling

Given a parameterized function, $F(\theta),\ \theta \in \mathbb{R}^M$

Let $F'(\theta)$ be an estimate of the gradient vector, to within constant ϵ in the infinity norm

$$\text{TIME}(F'(\theta)) \leq c_1\ \text{TIME}(F(\theta))$$

$$\text{MEMORY}(F'(\theta)) \leq c_2\ \text{MEMORY}(F(\theta))$$

Griewank, A., & Walther, A. (2008). *Evaluating derivatives: principles and techniques of algorithmic differentiation*. Society for industrial and applied mathematics.

Backpropagation scaling

Given a parameterized function, $F(\theta),\ \theta \in \mathbb{R}^M$

Let $F'(\theta)$ be an estimate of the gradient vector, to within constant ϵ in the infinity norm

$$\text{TIME}(F'(\theta)) \leq c_1\ \text{TIME}(F(\theta))$$

$$\text{MEMORY}(F'(\theta)) \leq c_2\ \text{MEMORY}(F(\theta))$$

For neural networks, $c_1, c_2 \in [2, 5]$

Griewank, A., & Walther, A. (2008). *Evaluating derivatives: principles and techniques of algorithmic differentiation*. Society for industrial and applied mathematics.

"Quantum backpropagation" scaling

Given a parameterized function, $F(\theta), \ \theta \in \mathbb{R}^M$

Let $F'(\theta)$ be an estimate of the gradient vector, to within constant ϵ in the infinity norm

$$\mathrm{TIME}(F'(\theta)) \leq c_1 \ \mathrm{TIME}(F(\theta))$$

$$\mathrm{MEMORY}(F'(\theta)) \leq c_2 \ \mathrm{MEMORY}(F(\theta))$$

For quantum models?

“Quantum backpropagation” scaling

Given a parameterized function, $F(\theta),\ \theta \in \mathbb{R}^M$

Let $F'(\theta)$ be an estimate of the gradient vector, to within constant ϵ in the infinity norm

$$\mathrm{TIME}(F'(\theta)) \leq c_1\ \mathrm{TIME}(F(\theta))$$

$$\mathrm{MEMORY}(F'(\theta)) \leq c_2\ \mathrm{MEMORY}(F(\theta))$$

For quantum models? $c_1, c_2 = \mathrm{polylog}(M)$

"Quantum backpropagation" scaling

Given a parameterized function, $F(\theta),\ \theta \in \mathbb{R}^M$

Let $F'(\theta)$ be an estimate of the gradient vector, to within constant ϵ in the infinity norm

$$\mathrm{TIME}(F'(\theta)) \leq c_1\ \mathrm{TIME}(F(\theta))$$

$$\mathrm{MEMORY}(F'(\theta)) \leq c_2\ \mathrm{MEMORY}(F(\theta))$$

For quantum models? $c_1, c_2 = \mathrm{polylog}(M)$

All quantum gradient methods in literature **do not achieve this**

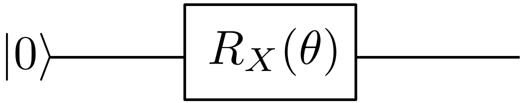
Nuance of quantum gradients

The nuance of quantum gradients

$|\psi(\theta)\rangle = e^{-i\theta X}|0\rangle$

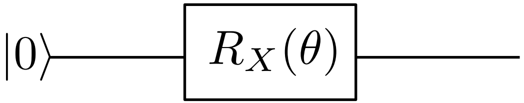
The nuance of quantum gradients

$|\psi(\theta)\rangle = e^{-i\theta X}|0\rangle$



The nuance of quantum gradients

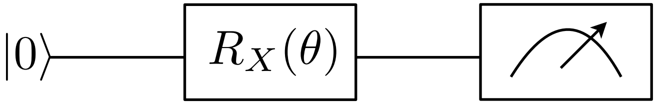
$|\psi(\theta)\rangle = e^{-i\theta X}\,|0\rangle$



$\nabla_\theta\,|\psi(\theta)\rangle = (-iX)e^{-i\theta X}\,|0\rangle$

The nuance of quantum gradients

$$|\psi(\theta)\rangle = e^{-i\theta X}|0\rangle$$

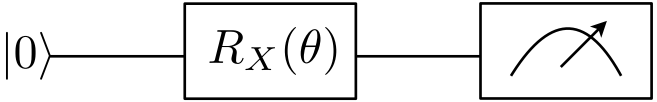


$$\nabla_\theta |\psi(\theta)\rangle = (-iX)e^{-i\theta X}|0\rangle$$

Let our observable of interest be Z, then:

The nuance of quantum gradients

$$|\psi(\theta)\rangle = e^{-i\theta X}|0\rangle$$

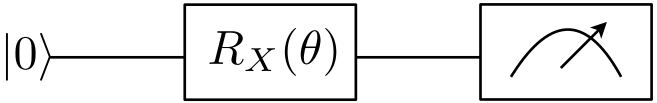


$$\nabla_\theta |\psi(\theta)\rangle = (-iX)e^{-i\theta X}|0\rangle$$

Let our observable of interest be Z, then:

$$F(\theta) = \langle 0| e^{i\theta X} Z e^{-i\theta X}|0\rangle$$

The nuance of quantum gradients



$$|\psi(\theta)\rangle = e^{-i\theta X}|0\rangle$$

$$\nabla_\theta |\psi(\theta)\rangle = (-iX)e^{-i\theta X}|0\rangle$$

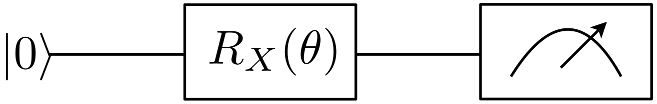
Let our observable of interest be Z, then:

$$F(\theta) = \langle 0| e^{i\theta X} Z e^{-i\theta X}|0\rangle$$

$$\nabla_\theta F(\theta) = \langle 0| e^{i\theta X} Z(-iX) e^{-i\theta X}|0\rangle + \langle 0| (iX) e^{i\theta X} Z e^{-i\theta X}|0\rangle$$

The nuance of quantum gradients

$$|\psi(\theta)\rangle = e^{-i\theta X}|0\rangle$$



$$\nabla_\theta |\psi(\theta)\rangle = (-iX)e^{-i\theta X}|0\rangle$$

Let our observable of interest be Z, then:

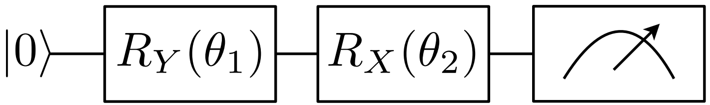
$$F(\theta) = \langle 0| e^{i\theta X} Z e^{-i\theta X}|0\rangle$$

$$\nabla_\theta F(\theta) = \langle 0| e^{i\theta X} Z(-iX) e^{-i\theta X}|0\rangle + \langle 0|(iX)e^{i\theta X} Z e^{-i\theta X}|0\rangle$$

$$\nabla_{\theta=0} F(0) = 2\ \mathrm{Im}(\langle 0| ZX |0\rangle)$$

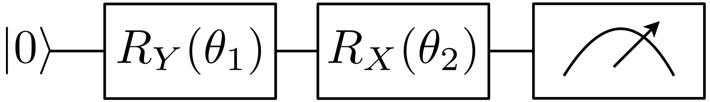
The nuance of quantum gradients

$|\psi(\theta)\rangle = e^{-i\theta_2 X} e^{-i\theta_1 Y} |0\rangle$



The nuance of quantum gradients

$$|\psi(\theta)\rangle = e^{-i\theta_2 X} e^{-i\theta_1 Y} |0\rangle$$



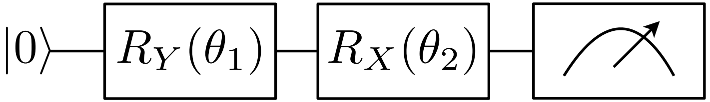
Let our observable of interest be Z, then:

$$\nabla_{\theta_1} F(\theta) = 2\ \mathrm{Re}(\langle 0| e^{i\theta_1 Y} e^{i\theta_2 X} Z e^{-i\theta_2 X} (-iY) e^{-i\theta_1 Y} |0\rangle)$$

$$\nabla_{\theta_2} F(\theta) = 2\ \mathrm{Re}(\langle 0| e^{i\theta_1 Y} e^{i\theta_2 X} Z (-iX) e^{-i\theta_2 X} e^{-i\theta_1 Y} |0\rangle)$$

The nuance of quantum gradients

$|\psi(\theta)\rangle = e^{-i\theta_2 X} e^{-i\theta_1 Y} |0\rangle$



Let our observable of interest be Z, then:

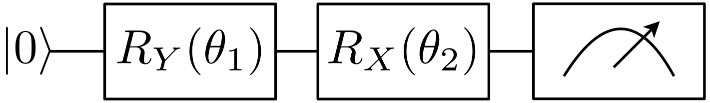
Setting $\theta_1, \theta_2 = 0$

$$\nabla_{\theta_1} F(\theta) = 2\ \mathrm{Re}(\langle 0| e^{i\theta_1 Y} e^{i\theta_2 X} Z e^{-i\theta_2 X} (-iY) e^{-i\theta_1 Y} |0\rangle) \ = 2\ \mathrm{Im} \langle 0| ZY |0\rangle$$

$$\nabla_{\theta_2} F(\theta) = 2\ \mathrm{Re}(\langle 0| e^{i\theta_1 Y} e^{i\theta_2 X} Z (-iX) e^{-i\theta_2 X} e^{-i\theta_1 Y} |0\rangle) \ = 2\ \mathrm{Im} \langle 0| ZX |0\rangle$$

The nuance of quantum gradients

$$|\psi(\theta)\rangle = e^{-i\theta_2 X} e^{-i\theta_1 Y} |0\rangle$$



Let our observable of interest be Z, then:

Setting $\theta_1, \theta_2 = 0$

$$\nabla_{\theta_1} F(\theta) = 2\ \mathrm{Re}(\langle 0| e^{i\theta_1 Y} e^{i\theta_2 X} Z e^{-i\theta_2 X} (-iY) e^{-i\theta_1 Y} |0\rangle) \ = 2\ \mathrm{Im} \langle 0| ZY |0\rangle$$

$$\nabla_{\theta_2} F(\theta) = 2\ \mathrm{Re}(\langle 0| e^{i\theta_1 Y} e^{i\theta_2 X} Z (-iX) e^{-i\theta_2 X} e^{-i\theta_1 Y} |0\rangle) \ = 2\ \mathrm{Im} \langle 0| ZX |0\rangle$$

Estimating M gradient components of a model with M parameters, corresponds to estimating M expected values

Cost model

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$ $\sim \ M$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)|0\rangle \qquad \sim \quad M$

$F(\theta) = \langle 0| U(\theta)^\dagger O \, U(\theta) |0\rangle$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle \quad \sim \quad M$

$F(\theta) = \langle 0|\, U(\theta)^\dagger O\, U(\theta)\,|0\rangle \quad \sim \quad M$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)|0\rangle \quad \sim \quad M$

$F(\theta) = \langle 0| U(\theta)^\dagger O \, U(\theta) |0\rangle \quad \sim \quad M/\epsilon^2$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)|0\rangle \quad \sim \quad M$

$F(\theta) = \langle 0| U(\theta)^\dagger O\, U(\theta) |0\rangle \quad \sim \quad M/\epsilon^2$

$[F'(\theta)]_{\theta_k} =$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)|0\rangle \quad \sim \quad M$

$F(\theta) = \langle 0| U(\theta)^\dagger O\, U(\theta) |0\rangle \quad \sim \quad M/\epsilon^2$

$[F'(\theta)]_{\theta_k} = 2\ \mathrm{Re}[\langle 0| U(\theta)^\dagger O \partial_{\theta_k} U(\theta) |0\rangle]$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle \quad \sim \quad M$

$F(\theta) = \langle 0|\, U(\theta)^\dagger O\, U(\theta)\,|0\rangle \quad \sim \quad M/\epsilon^2$

$[F'(\theta)]_{\theta_k} = 2\ \mathrm{Re}[\langle 0|\, U(\theta)^\dagger O \partial_{\theta_k} U(\theta)\,|0\rangle] \quad \sim \quad M/\epsilon^2$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle \quad \sim \quad M$

$F(\theta) = \langle 0|\, U(\theta)^\dagger O\, U(\theta)\,|0\rangle \quad \sim \quad M/\epsilon^2$

$[F'(\theta)]_{\theta_k} = 2\,\mathrm{Re}[\langle 0|\, U(\theta)^\dagger O \partial_{\theta_k} U(\theta)\,|0\rangle] \quad \sim \quad M/\epsilon^2$ for a single component!

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle \qquad \sim \quad M$

$F(\theta) = \langle 0|\, U(\theta)^\dagger O\, U(\theta)\,|0\rangle \quad \sim \quad M/\epsilon^2$

$[F'(\theta)]_{\theta_k} = 2\,\mathrm{Re}[\langle 0|\, U(\theta)^\dagger O \partial_{\theta_k} U(\theta)\,|0\rangle] \quad \sim \quad M/\epsilon^2$ for a single component!

$\sim \quad M^2/\epsilon^2$ for full gradient

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle \quad \sim \quad M$

$F(\theta) = \langle 0|\, U(\theta)^\dagger O\, U(\theta)\,|0\rangle \quad \sim \quad M/\epsilon^2$

$[F'(\theta)]_{\theta_k} = 2\,\mathrm{Re}[\langle 0|\, U(\theta)^\dagger O \partial_{\theta_k} U(\theta)\,|0\rangle] \quad \sim \quad M/\epsilon^2$ for a single component!

$\sim \quad M^2/\epsilon^2$ for full gradient

$\mathrm{TIME}(F'(\theta)) = M^2$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle \quad \sim \quad M$

$F(\theta) = \langle 0|\, U(\theta)^\dagger O\, U(\theta)\,|0\rangle \quad \sim \quad M/\epsilon^2$

$[F'(\theta)]_{\theta_k} = 2\ \mathrm{Re}[\langle 0|\, U(\theta)^\dagger O \partial_{\theta_k} U(\theta)\,|0\rangle] \quad \sim \quad M/\epsilon^2$ for a single component!

$\sim \quad M^2/\epsilon^2$ for full gradient

$$\mathrm{TIME}(F'(\theta)) = M^2 = M\ \mathrm{TIME}(F(\theta))$$

Cost model

Each parameterized operation and its inverse = unit cost

$U(\theta) = \prod_{j=1}^{M} U_j(\theta_j)$

$|\psi(\theta)\rangle = U(\theta)|0\rangle \quad \sim \quad M$

$F(\theta) = \langle 0| U(\theta)^\dagger O\, U(\theta) |0\rangle \quad \sim \quad M/\epsilon^2$

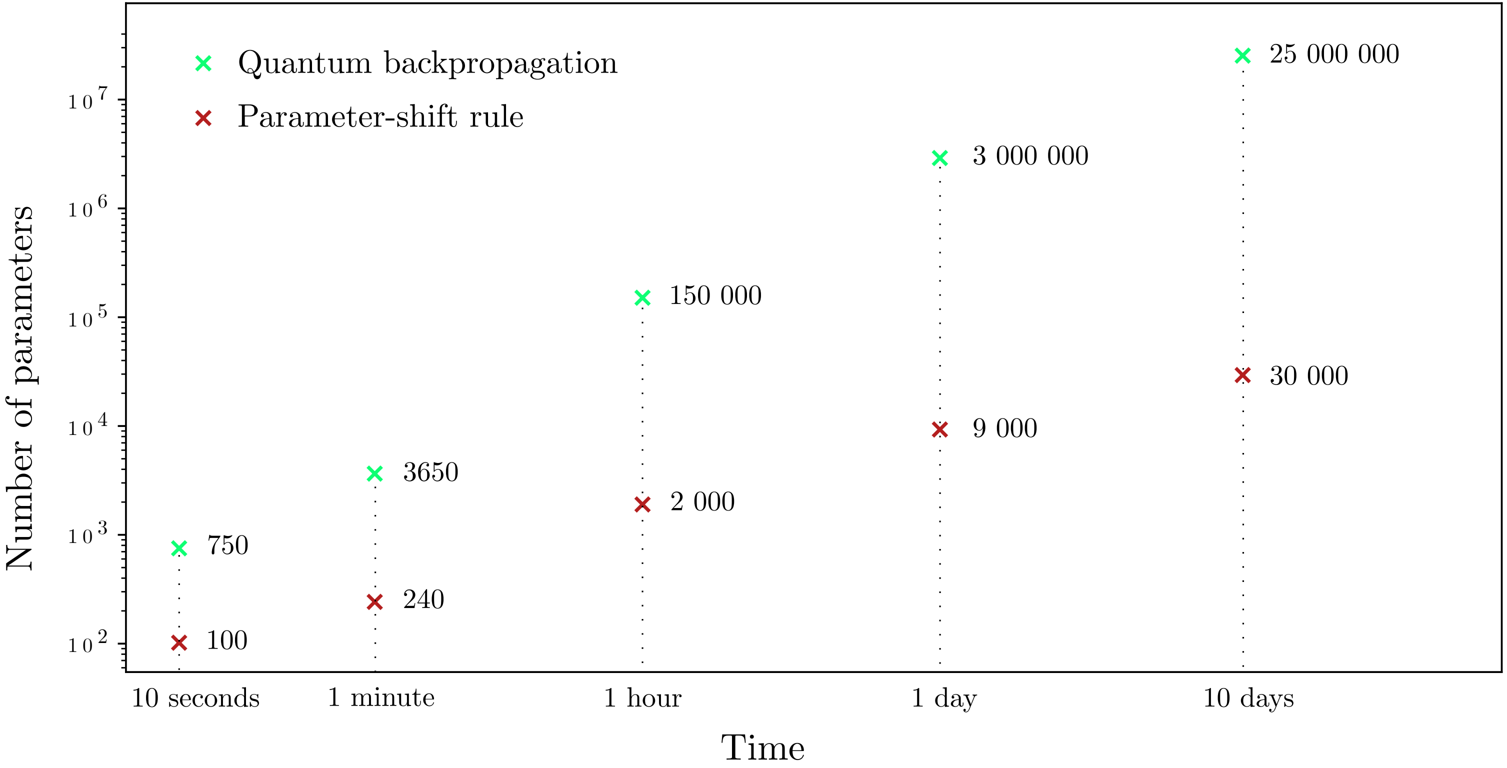
$[F'(\theta)]_{\theta_k} = 2\ \mathrm{Re}[\langle 0| U(\theta)^\dagger O \partial_{\theta_k} U(\theta) |0\rangle] \quad \sim \quad M/\epsilon^2$ for a single component!

$\sim \quad M^2/\epsilon^2$ for full gradient

$$\mathrm{TIME}(F'(\theta)) \approx \mathrm{TIME}(F(\theta))$$

Classical

It *is* a big deal



Babbush, Ryan, Jarrod R. McClean, Michael Newman, Craig Gidney, Sergio Boixo, and Hartmut Neven. "Focus beyond quadratic speedups for error-corrected quantum advantage." *PRX Quantum* 2, no. 1 (2021): 010103.

Quantum model settings

Quantum model settings

1. Input state is known

Quantum model settings

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

Quantum model settings

1. Input state is known	2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

Quantum model settings

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$$

Quantum model settings

1. Input state is known

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

$$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

Quantum model settings

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)|0\rangle$$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)|\psi\rangle$$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

b) Multi-copy access

$|\psi\rangle|\psi\rangle|\psi\rangle\cdots|\psi\rangle$

Quantum model settings

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)|0\rangle$$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)|\psi\rangle$$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

b) Multi-copy access

$|\psi\rangle|\psi\rangle|\psi\rangle\cdots|\psi\rangle$

Backpropagation scaling is not possible

1. Input state is known

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$

2. Input state is unknown

$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$

a) Single copy access

$|\psi\rangle \quad |\psi\rangle \quad |\psi\rangle \cdots |\psi\rangle$

b) Multi-copy access

$|\psi\rangle\,|\psi\rangle\,|\psi\rangle \cdots |\psi\rangle$

Backpropagation scaling is not possible

1. Input state is known

$|\psi(\theta)\rangle = U(\theta)|0\rangle$

2. Input state is unknown

$|\psi(\theta)\rangle = U(\theta)|\psi\rangle$

a) Single copy access

b) Multi-copy access

$U(\theta) = \prod_{j=1}^{M} e^{-i\theta_j P_j}$

and then set $\theta = 0$

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

$|\psi\rangle |\psi\rangle |\psi\rangle \cdots |\psi\rangle$

S. Chen, J. Cotler, H. -Y. Huang and J. Li, "Exponential Separations Between Learning With and Without Quantum Memory," *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, Denver, CO, USA, 2022, pp. 574-585, doi: 10.1109/FOCS52979.2021.00063.

Backpropagation scaling is not possible

1. Input state is known

2. Input state is unknown

$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$

$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

Corollary 5.9 (Shadow tomography lower bound for Pauli observables). *Any learning algorithm without quantum memory requires*

$$T \geq \Omega\left(2^n/\varepsilon^2\right) \tag{117}$$

copies of ρ *to predict expectation values of* $\mathrm{tr}(P_i\rho)$ *to at most* ε*-error for all* $i = 1, \ldots, 2(4^n - 1)$ *with at least a probability of* $2/3$.

$|\psi\rangle\,|\psi\rangle\,|\psi\rangle \cdots |\psi\rangle$

S. Chen, J. Cotler, H. -Y. Huang and J. Li, "Exponential Separations Between Learning With and Without Quantum Memory," *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, Denver, CO, USA, 2022, pp. 574-585, doi: 10.1109/FOCS52979.2021.00063.

Backpropagation scaling is <u>not</u> possible

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$$

a) Single copy access

$$\text{TIME}(F'(\theta)) = M^2 = M\ \text{TIME}(F(\theta))$$

$$|\psi\rangle \quad |\psi\rangle \quad |\psi\rangle \cdots |\psi\rangle$$

$$|\psi\rangle\,|\psi\rangle\,|\psi\rangle \cdots |\psi\rangle$$

S. Chen, J. Cotler, H. -Y. Huang and J. Li, "Exponential Separations Between Learning With and Without Quantum Memory," *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, Denver, CO, USA, 2022, pp. 574-585, doi: 10.1109/FOCS52979.2021.00063.

Backpropagation scaling is not possible

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

b) Multi-copy access

$|\psi\rangle\,|\psi\rangle\,|\psi\rangle\cdots|\psi\rangle$

Quantum model settings

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

b) Multi-copy access

$|\psi\rangle\,|\psi\rangle\,|\psi\rangle\cdots|\psi\rangle$

Quantum model settings

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)|0\rangle$$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)|\psi\rangle$$

a) Single copy access

b) Multi-copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

$|\psi\rangle|\psi\rangle|\psi\rangle\cdots|\psi\rangle$

Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." *Nature Physics* 16.10 (2020): 1050-1057.

Ji, Zhengfeng, Yi-Kai Liu, and Fang Song. "Pseudorandom quantum states." *Advances in Cryptology–CRYPTO 2018: 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19–23, 2018, Proceedings, Part III 38.* Springer International Publishing, 2018.

Single copies, pure state of poly complexity

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

1. Can learn the circuit efficiently info-theoretically (using classical shadows)

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

1. Can learn the circuit efficiently info-theoretically (using classical shadows)

2. Cannot determine said state efficiently, computationally (in general)

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

$K \approx n^{p(n)}$ possible circuits.

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

$K \approx n^{p(n)}$ possible circuits. Denote the states by $|\phi_1\rangle, |\phi_2\rangle, ..., |\phi_K\rangle$

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

$K \approx n^{p(n)}$ possible circuits. Denote the states by $|\phi_1\rangle, |\phi_2\rangle, ..., |\phi_K\rangle$

Classical shadows method to estimate the fidelity w.r.t. all K states using a "shadow" of ρ

$$|\langle\phi_1|\psi\rangle|^2$$
$$|\langle\phi_2|\psi\rangle|^2$$
$$\vdots$$
$$|\langle\phi_K|\psi\rangle|^2$$

Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." *Nature Physics* 16.10 (2020): 1050-1057.

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

$K \approx n^{p(n)}$ possible circuits. Denote the states by $|\phi_1\rangle, |\phi_2\rangle, ..., |\phi_K\rangle$

Classical shadows method to estimate the fidelity w.r.t. all K states using a "shadow" of ρ

$$\left.\begin{matrix} |\langle\phi_1|\psi\rangle|^2 \\ |\langle\phi_2|\psi\rangle|^2 \\ \vdots \\ |\langle\phi_K|\psi\rangle|^2 \end{matrix}\right\} \quad \Omega(\log(K)/\epsilon^2) \text{ measurements}$$

Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." *Nature Physics* 16.10 (2020): 1050-1057.

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

$K \approx n^{p(n)}$ possible circuits. Denote the states by $|\phi_1\rangle, |\phi_2\rangle, ..., |\phi_K\rangle$

Classical shadows method to estimate the fidelity w.r.t. all K states using a "shadow" of ρ

$$\left.\begin{matrix} |\langle\phi_1|\psi\rangle|^2 \\ |\langle\phi_2|\psi\rangle|^2 \\ \vdots \\ |\langle\phi_K|\psi\rangle|^2 \end{matrix}\right\} \quad \Omega(\log(K)/\epsilon^2) \text{ measurements}$$

Will find fidelity $= 1$, w.h.p.

Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." *Nature Physics* 16.10 (2020): 1050-1057.

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

$K \approx n^{p(n)}$ possible circuits. Denote the states by $|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_K\rangle$

Classical shadows method to estimate the fidelity w.r.t. all K states using a "shadow" of ρ

$$\left.\begin{matrix} |\langle\phi_1|\psi\rangle|^2 \\ |\langle\phi_2|\psi\rangle|^2 \\ \vdots \\ |\langle\phi_K|\psi\rangle|^2 \end{matrix}\right\} \quad \Omega(\log(K)/\epsilon^2) \text{ measurements}$$

Obtaining the maximum fidelity involves storing K values and searching over them

Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." *Nature Physics* 16.10 (2020): 1050-1057.

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

Proposition: Under standard cryptographic assumptions, no efficient computational procedure exists to identify a pure state of polynomial complexity to trace distance ϵ

Single copies, pure state of poly complexity

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a state generated from a polynomial complexity circuit

Proposition: Under standard cryptographic assumptions, no efficient computational procedure exists to identify a pure state of polynomial complexity to trace distance ϵ

Proof:

1. A pseudo-random quantum state is defined to be a pure state of polynomial complexity
2. No efficient computational algorithm given a polynomial number of copies of the state can distinguish from the Haar random state
3. Classical shadows + classical search procedure recreates the state to trace distance ϵ using a polynomial number copies of the state
4. If this is computationally efficient, then the state can be cloned efficiently, violating the no-cloning theorem for pseudo-random states which rests upon standard cryptographic assumptions

Ji, Zhengfeng, Yi-Kai Liu, and Fang Song. "Pseudorandom quantum states." *Advances in Cryptology–CRYPTO 2018: 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19–23, 2018, Proceedings, Part III 38*. Springer International Publishing, 2018.

Quantum model settings

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$$

b) Multi-copy access

$|\psi\rangle\,|\psi\rangle\,|\psi\rangle \cdots |\psi\rangle$

Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." *Nature Physics* 16.10 (2020): 1050-1057.

Ji, Zhengfeng, Yi-Kai Liu, and Fang Song. "Pseudorandom quantum states." *Advances in Cryptology–CRYPTO 2018: 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19–23, 2018, Proceedings, Part III 38.* Springer International Publishing, 2018.

Achieving backprop scaling seems unlikely with single copies

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

b) Multi-copy access

$|\psi\rangle\,|\psi\rangle\,|\psi\rangle \cdots |\psi\rangle$

Why is it not so straightforward?

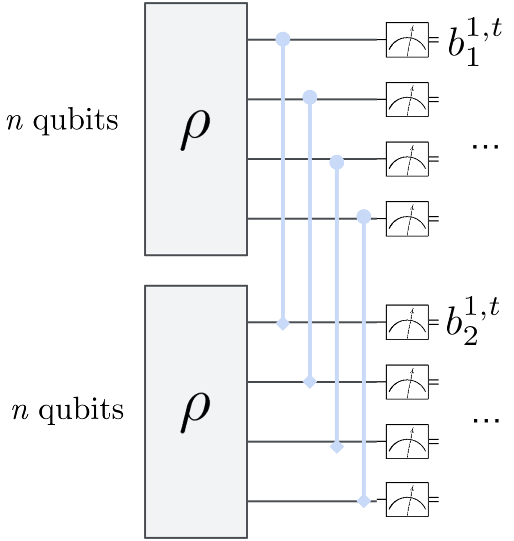
Why is it not so straightforward?

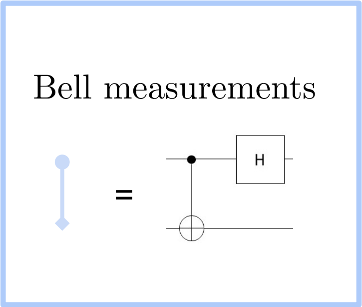
$$F(\theta) = \langle 0| U(\theta)^\dagger O\, U(\theta) |0\rangle$$

$$[F'(\theta)]_{\theta_k} = 2\ \mathrm{Re}[\langle 0| U(\theta)^\dagger O \partial_{\theta_k} U(\theta) |0\rangle]$$

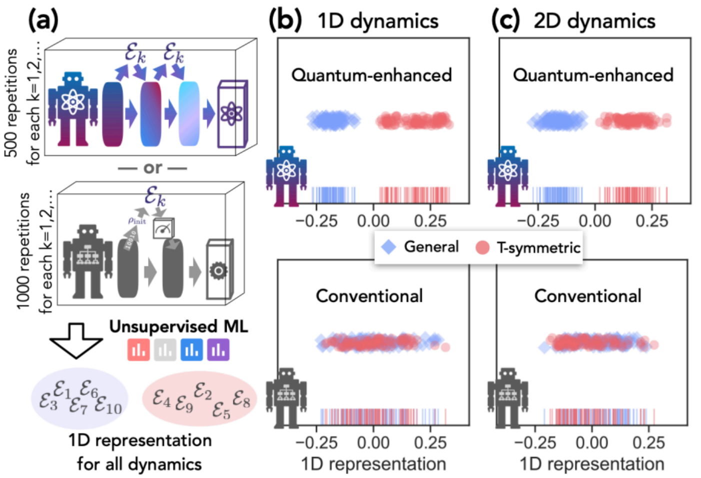
Allowing multi-copy access
(Intuition)

Multi-copy measurements





$2n$ classical bits for N rounds creates Bell sketch of state $\{b_i^{k,t}\}$



Huang, H.Y., Broughton, M., Cotler, J., Chen, S., Li, J., Mohseni, M., Neven, H., Babbush, R., Kueng, R., Preskill, J. and McClean, J.R., 2022. Quantum advantage in learning from experiments. Science, 376(6598), pp.1182-1186.

Multi-copy measurements (restricted setting)

$U(\theta) = \prod_{j=1}^{M} e^{-i\theta_j P_j}$ and then set parameters to zero

Multi-copy measurements (restricted setting)

$U(\theta) = \prod_{j=1}^{M} e^{-i\theta_j P_j}$ and then set parameters to zero

All M gradient components can be estimated to within a fixed precision ϵ using $O(\log(M)/\epsilon^4)$ function calls

Multi-copy measurements (restricted setting)

$U(\theta) = \prod_{j=1}^{M} e^{-i\theta_j P_j}$ and then set parameters to zero

All M gradient components can be estimated to within a fixed precision ϵ using $O(\log(M)/\epsilon^4)$ function calls

$$\text{TIME}(F'(\theta)) = O(\log(M))\text{TIME}(F(\theta))$$

which is in line with our definition of backpropagation scaling! :)

Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Information-theoretic bounds on quantum advantage in machine learning." *Physical Review Letters* 126.19 (2021): 190505.

Move away from restricted setting?

Multiple copies (general setting)

Multiple copies (general setting)

Shadow tomography:

Aaronson, Scott. "Shadow tomography of quantum states." *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*. 2018.

Multiple copies (general setting)

Shadow tomography:

Let $\mathcal{E}$ be a class of two-outcome measurements with outcomes in $\{\pm 1\}$

Aaronson, Scott. "Shadow tomography of quantum states." *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*. 2018.

Multiple copies (general setting)

Shadow tomography:

Let $\mathcal{E}$ be a class of two-outcome measurements with outcomes in $\{\pm 1\}$

Given m copies of an unknown n-qubit quantum state $|\psi\rangle$,

Aaronson, Scott. "Shadow tomography of quantum states." *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*. 2018.

Multiple copies (general setting)

Shadow tomography:

Let $\mathcal{E}$ be a class of two-outcome measurements with outcomes in $\{\pm 1\}$

Given m copies of an unknown n-qubit quantum state $|\psi\rangle$,

and known measurements $E_1, \ldots, E_M \in \mathcal{E}$

Aaronson, Scott. "Shadow tomography of quantum states." *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*. 2018.

Multiple copies (general setting)

Shadow tomography:

Let $\mathcal{E}$ be a class of two-outcome measurements with outcomes in $\{\pm 1\}$

Given m copies of an unknown n-qubit quantum state $|\psi\rangle$,

and known measurements $E_1, \ldots, E_M \in \mathcal{E}$

Task:

Aaronson, Scott. "Shadow tomography of quantum states." *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*. 2018.

Multiple copies (general setting)

Shadow tomography:

Let $\mathcal{E}$ be a class of two-outcome measurements with outcomes in $\{\pm 1\}$

Given m copies of an unknown n-qubit quantum state $|\psi\rangle$,

and known measurements $E_1, \ldots, E_M \in \mathcal{E}$

Task:

Output estimates $b_1, \ldots, b_M \in [-1, 1]$ such that $|b_k - \langle\psi|E_k|\psi\rangle| \leq \varepsilon \; \forall k$

Aaronson, Scott. "Shadow tomography of quantum states." *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*. 2018.

Multiple copies (general setting)

Shadow tomography:

Let $\mathcal{E}$ be a class of two-outcome measurements with outcomes in $\{\pm 1\}$

Given m copies of an unknown n-qubit quantum state $|\psi\rangle$,

and known measurements $E_1, \ldots, E_M \in \mathcal{E}$

Task:

Output estimates $b_1, \ldots, b_M \in [-1, 1]$ such that $|b_k - \langle\psi|E_k|\psi\rangle| \leq \varepsilon \ \forall k$

In particular, do this via a measurement of $|\psi\rangle^{\otimes m}$ where m is as small as possible

Aaronson, Scott. "Shadow tomography of quantum states." *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*. 2018.

Multiple copies (general setting)

Shadow tomography reduction:

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

Shadow tomography reduction:

Suppose there is an algorithm which can estimate all gradient components to precision ε, with m copies and with runtime T.

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

Shadow tomography reduction:

Suppose there is an algorithm which can estimate all gradient components to precision ε, with m copies and with runtime T.

Then, this gives an algorithm for shadow tomography of poly-time observables, to precision $\frac{\varepsilon}{2}$, with m copies of $|\psi\rangle$ and runtime T

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

Shadow tomography reduction:

Suppose there is an algorithm which can estimate all gradient components to precision ε, with m copies and with runtime T.

Then, this gives an algorithm for shadow tomography of poly-time observables, to precision $\frac{\varepsilon}{2}$, with m copies of $|\psi\rangle$ and runtime T

Quantum neural network:

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

Shadow tomography reduction:

Suppose there is an algorithm which can estimate all gradient components to precision ε, with m copies of $|\psi$
and with runtime T.

Then, this gives an algorithm for shadow tomography of poly-time observables, to precision $\frac{\varepsilon}{2}$,
with m copies of $|\psi\rangle$ and runtime T

Quantum neural network:

$$\mathrm{QNN}_{\vec{\theta}}(|\psi\rangle) = \langle 0|\langle\psi|\mathcal{U}^\dagger(\vec{\theta})\, Z_0\, \mathcal{U}(\vec{\theta})|0\rangle|\psi\rangle$$

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

“Quantum-efficient” backpropagation:

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

“Quantum-efficient” backpropagation:

There exists an explicit algorithm which produces estimates b_k for all $k = 1, ..., M$ such that

$$|b_k - \tfrac{1}{2}\partial_{\theta_k}\mathrm{QNN}_{\vec{\theta}}(|\psi\rangle)| \leq \varepsilon \ \forall k$$

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

“Quantum-efficient” backpropagation:

There exists an explicit algorithm which produces estimates b_k for all $k = 1, ..., M$ such that

$$|b_k - \frac{1}{2}\partial_{\theta_k}\mathrm{QNN}_{\vec{\theta}}(|\psi\rangle)| \leq \varepsilon \ \forall k$$

Using only

$$m = O\left(\frac{n \log^2 M}{\varepsilon^4}\right)$$

copies of $|\psi\rangle$. Moreover, the required number of quantum operations for the proposed algorithm is $\tilde{O}(mM)$

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

“Quantum-efficient” backpropagation:

There exists an explicit algorithm which produces estimates b_k for all $k = 1, ..., M$ such that

$$|b_k - \frac{1}{2}\partial_{\theta_k} \mathrm{QNN}_{\vec{\theta}}(|\psi\rangle)| \leq \varepsilon \ \forall k$$

Using only

$$m = O\left(\frac{n \log^2 M}{\varepsilon^4}\right)$$

copies of $|\psi\rangle$. Moreover, the required number of quantum operations for the proposed algorithm is $\tilde{O}(mM)$

$$\mathrm{TIME}(F'(\theta)) \approx \mathrm{polylog}(M)\ \mathrm{TIME}(F(\theta))$$

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

“Quantum-efficient” backpropagation:

There exists an explicit algorithm which produces estimates b_k for all $k = 1, \ldots, M$ such that

$$|b_k - \tfrac{1}{2}\partial_{\theta_k}\mathrm{QNN}_{\vec{\theta}}(|\psi\rangle)| \leq \varepsilon \ \forall k$$

Using only

$$m = O\left(\frac{n \log^2 M}{\varepsilon^4}\right)$$

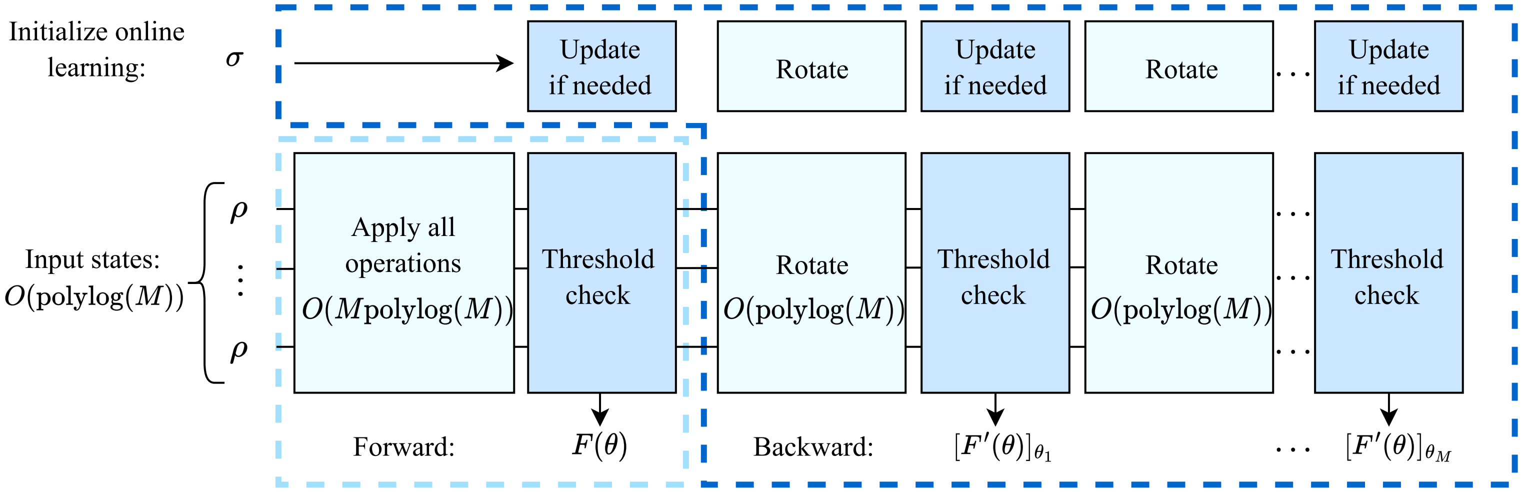
copies of $|\psi\rangle$. Moreover, the required number of quantum operations for the proposed algorithm is $\tilde{O}(mM)$

Classical resources: $M \cdot 2^{\tilde{O}(n)}$

Abbas, Amira, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh, Dar Gilboa, and Jarrod R. McClean. "On quantum backpropagation, information reuse, and cheating measurement collapse." *arXiv preprint arXiv:2305.13362* (2023).

Multiple copies (general setting)

“Quantum-efficient” backpropagation:



[1] Aaronson, Scott, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. "Online learning of quantum states." *Advances in neural information processing systems* 31 (2018).

[2] Bădescu, Costin, and Ryan O'Donnell. "Improved quantum data analysis." Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (2021).

Takeaways

1. Input state is known

$$|\psi(\theta)\rangle = U(\theta)\,|0\rangle$$

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

b) Multi-copy access

$|\psi\rangle|\psi\rangle|\psi\rangle\cdots|\psi\rangle$

Takeaways

1. Input state is known

Computational arguments

2. Input state is unknown

$$|\psi(\theta)\rangle = U(\theta)\,|\psi\rangle$$

a) Single copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

b) Multi-copy access

$|\psi\rangle|\psi\rangle|\psi\rangle\cdots|\psi\rangle$

Takeaways

1. Input state is known

2. Input state is unknown

Computational arguments

Info-theoretic lower bounds

a) Single copy access

b) Multi-copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

$|\psi\rangle|\psi\rangle|\psi\rangle\cdots|\psi\rangle$

Takeaways

1. Input state is known

2. Input state is unknown

a) Single copy access

b) Multi-copy access

$|\psi\rangle$ $|\psi\rangle$ $|\psi\rangle$ $\cdots$ $|\psi\rangle$

$|\psi\rangle|\psi\rangle|\psi\rangle\cdots|\psi\rangle$

Time efficient in quantum resources

Memory fails

Open questions

- Is there an efficient computational scheme for quantum gradients?
- Special cases of parameterized models that scale and train well?
- Other models types?
- Different methods for optimization?

Open questions

- Is there an efficient computational scheme for quantum gradients?
    - Special cases of parameterized models that scale and train well?
        - Other models types?
            - Different methods for optimization?

Open questions

- Is there an efficient computational scheme for quantum gradients?
  - Special cases of parameterized models that scale and train well?
    - Other models types?
      - Different methods for optimization?

Open questions

- Is there an efficient computational scheme for quantum gradients?
  - Special cases of parameterized models that scale and train well?
    - Other models types?
      - Different methods for optimization?

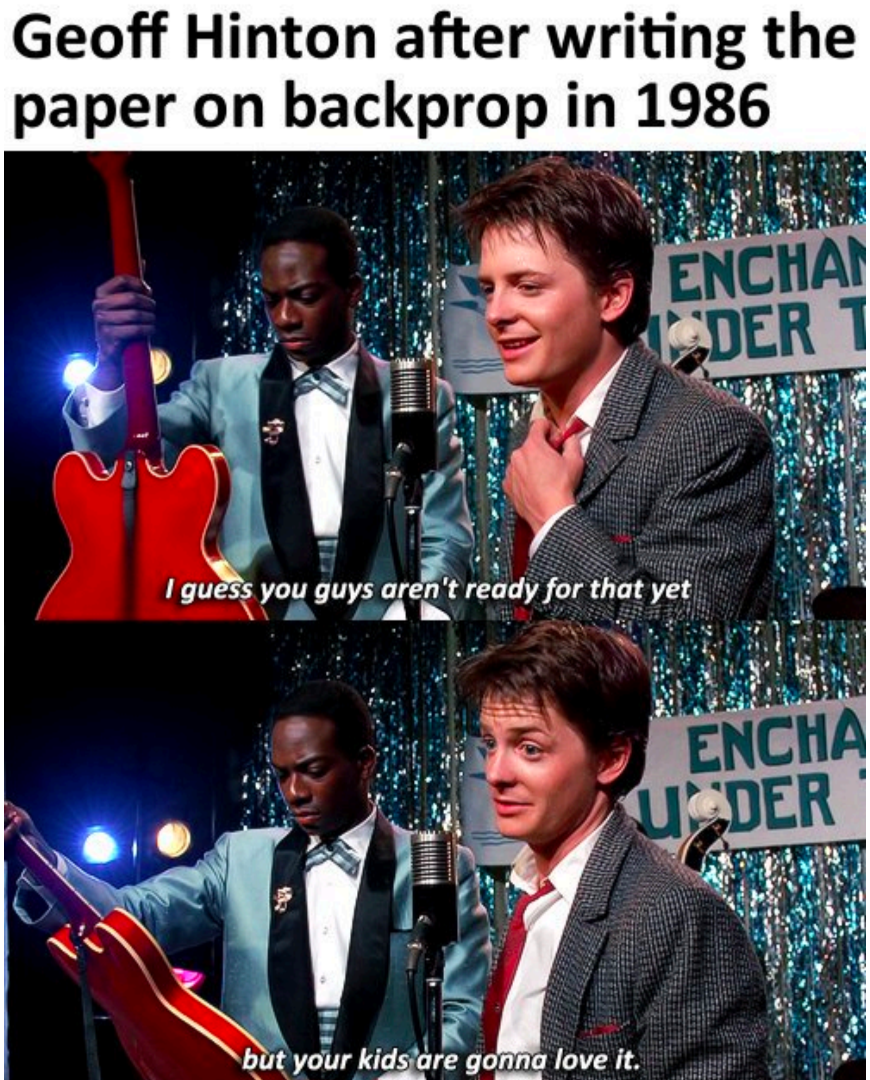
Open questions

- Is there an efficient computational scheme for quantum gradients?
- Special cases of parameterized models that scale and train well?
- Other models types?
- Different methods for optimization?

Open questions

- Is there an efficient computational scheme for quantum gradients?
- Special cases of parameterized models that scale and train well?
- Other models types?
- Different methods for optimization?

Thank you!


Geoff Hinton after writing the paper on backprop in 1986
ENCHAN
I guess you guys aren't ready for that yet
but your kids are gonna love it.