Exponential Hardness of Optimization in Variational Quantum Algorithms

Based on: arXiv: 2205.05056

Hao-Kai Zhang, Chengkai Zhu, Geng Liu, Xin Wang

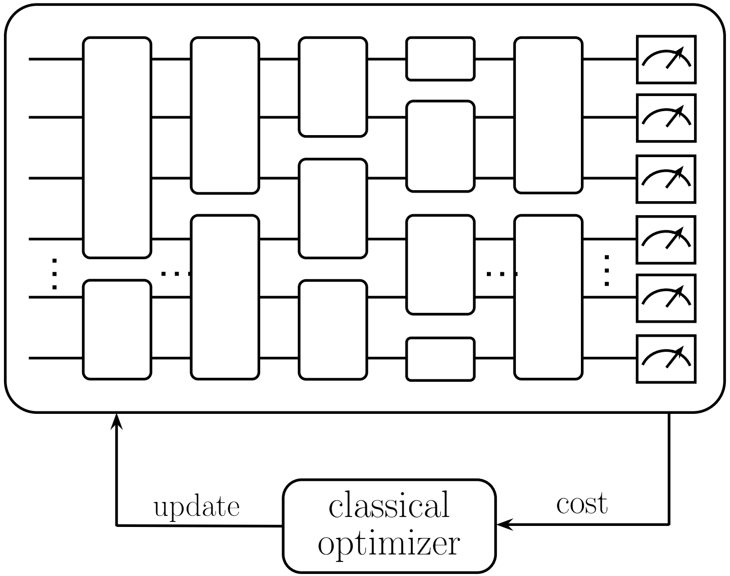
Outline

- Background
  - VQA setting
  - Barren plateaus: what & why
- Main results
  - Theorem & proof
  - Case study
  - Implication: relation with BP
- Summary

Variational Quantum Algorithm (VQA)

- VQA - use a classical optimizer to train a quantum circuit

1. Initialize a circuit with an input state
2. Run & measure to get the cost
3. Update circuit parameters
4. Converge and get the desired circuit



- VQA cost function:

input state

$$C_{H,\rho}(\mathbf{U}) = \mathrm{tr}(H\mathbf{U}\rho\mathbf{U}^{\dagger})$$

a task-dependent observable

circuit (ansatz)

What is Barren Plateaus (BP) ?

- Barren plateau = exponentially vanishing gradients (in the number of qubits)

arXiv: 1803.11173

$$\mathbb{E}[\partial_\mu C] = 0,\ \mathrm{Var}[\partial_\mu C] \in \mathcal{O}(b^{-n}),\ b > 1$$

randomness from where?
- random initialization

- → Exponential small probability to get non-zero gradients (to a fixed precision)

$$\Pr[|\partial_\mu C| \geq \epsilon] \leq \frac{1}{\epsilon^2}\mathrm{Var}[\partial_\mu C]$$

(Chebyshev's inequality)

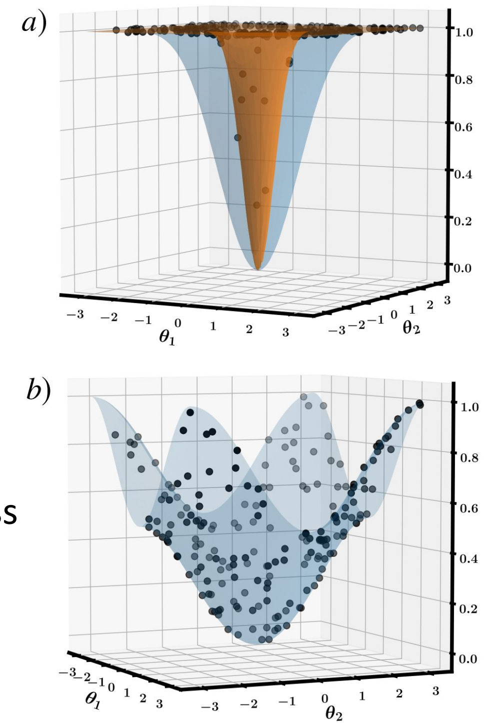
- → need exponential precision on quantum measurement to make progress

$$\theta_\mu^{(t)} = \theta_\mu^{(t-1)} - \eta \cdot \partial_\mu C$$

$$\text{resource} \in \mathcal{O}(1/\epsilon^\alpha),\ \alpha > 0$$

arXiv: quant-ph/0607019

- Note that quantum advantage is realized only for a large number of qubits



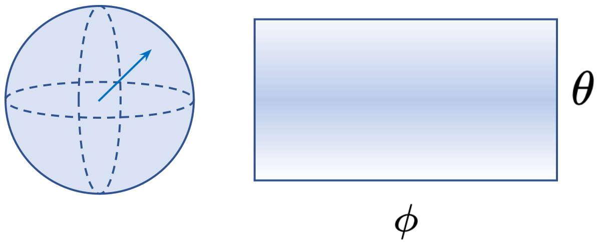
arXiv: 2001.00550

Why there is BP ?

- Intuition: concentration of measure from Haar

(Levy's lemma)

$$d\Omega = \sin\theta \, d\theta d\phi$$



Dimension ↑ Concentration ↑ Flatness ↑

- McClean's BP theorem says 2-design is enough

t-design: $$\frac{1}{|\mathbb{V}|}\sum_{V\in\mathbb{V}} P_{t,t}(V) = \int_{\mathcal{U}(d)} d\mu(V) P_{t,t}(V)$$

"pseudo-Haar"

(match Haar up to the 2nd moment)

- One line proof (exact 2-design is exactly integrable just using formula)

$$e^{-i\theta_\mu \Omega_\mu} \qquad C_{H,\rho}(\mathbf{U}) = \mathrm{tr}(H\mathbf{U}\rho\mathbf{U}^\dagger)$$

$$\mathrm{Var}[\partial_\mu C] = 2\,\mathrm{tr}(H^2)\,\mathrm{tr}(\rho^2)\left(\frac{\mathrm{tr}(\Omega_\mu^2)}{2^{3n}} - \frac{\mathrm{tr}(\Omega_\mu)^2}{2^{4n}}\right)$$

$$\in \mathcal{O}(2^{-n})$$

Cost: 1-degree

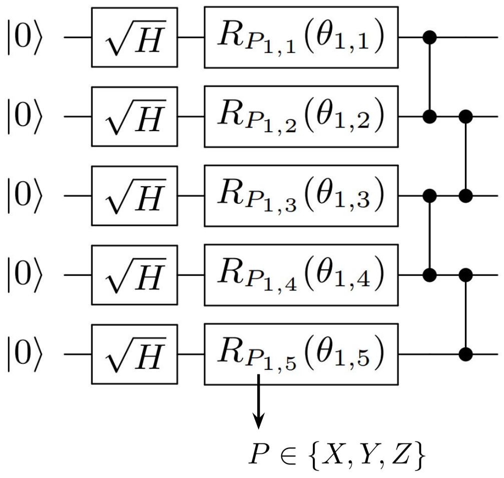
Gradient: 1-degree

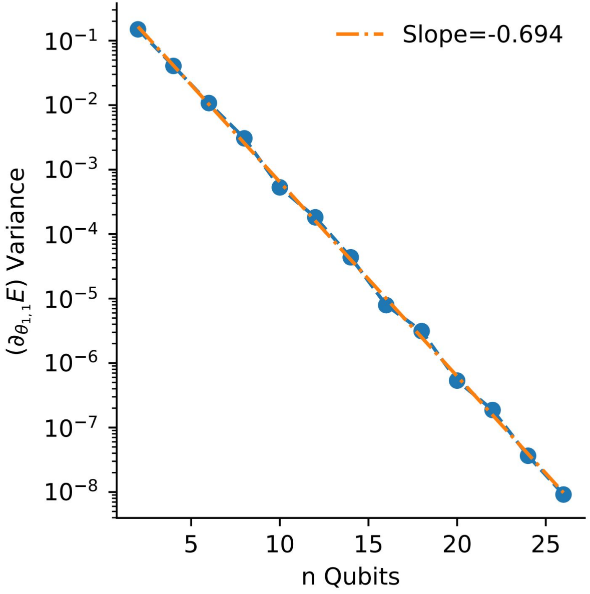
Variance: 2-degree

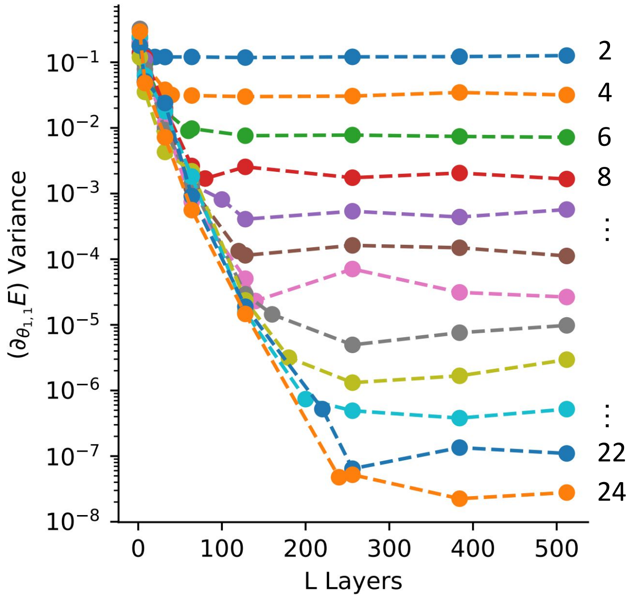
An example circuit showing BP

- Hardware efficient ansatz

arXiv: 1803.11173





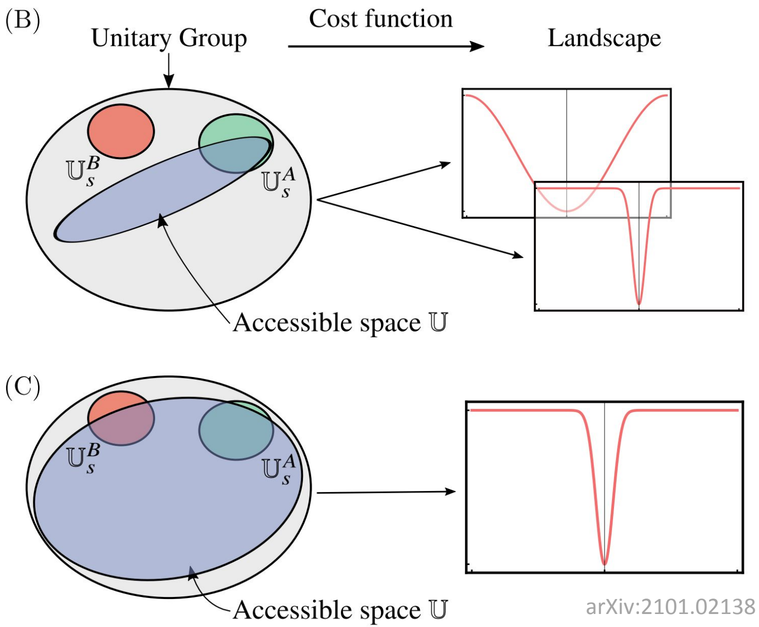


- Many global-repeated-layer-type ansatzes are 2-designs when the number of layers is large

e.g., 10×n layers of Ry-CNOT or U3-CNOT.

How to avoid BP ?

- Shallower? But we also want sufficient expressibility
- Natural gradient descent?
- Gradient-free method ?
- Gate-by-gate optimization ?
- Reparameterization ?
- Clever initialization?
- Designed architecture ?
- Adaptive method ?
- …

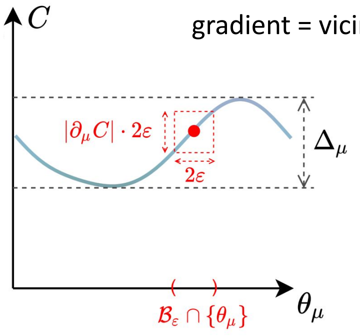


arXiv:2101.02138

→ We need more information to guide us !

Beyond gradients

- Variation range of cost function



Variation range

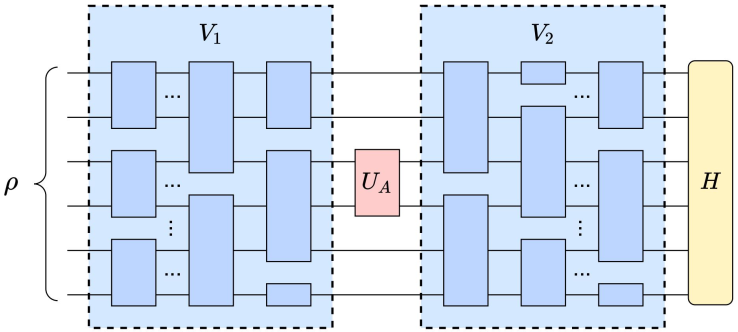
via adjusting a local unitary

Whole system: n qubits

Subsystem A: m qubits

Subsystem B: n-m qubits

- Locality of quantum circuits



here

local = acting on few qubits

e.g. Rx,Ry,Rz +

CNOT

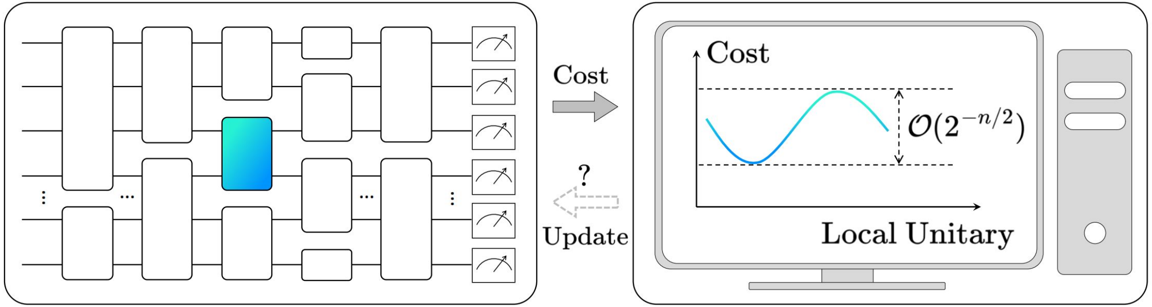
Definition 1 *For a generic VQA cost function $C_{H,\rho}(\mathbf{U})$ in Eq. (1), we define its variation range with given V_1, V_2 as*

$$\Delta_{H,\rho}(V_1, V_2) := \max_{U_A} C_{H,\rho}(\mathbf{U}) - \min_{U_A} C_{H,\rho}(\mathbf{U}), \qquad (2)$$

where the maximum and minimum with respect to U_A are taken over the unitary group $\mathcal{U}(2^m)$ of degree 2^m.

Main theorem

- Variation range is exponentially small !



Theorem 1 *Suppose $\mathbb{V}_1, \mathbb{V}_2$ are ensembles from which V_1, V_2 are sampled, respectively. If either $\mathbb{V}_1$ or $\mathbb{V}_2$, or both form unitary 2-designs, then for arbitrary H, ρ, the following inequality holds*

$$\mathbb{E}_{V_1,V_2}[\Delta_{H,\rho}(V_1, V_2)] \leq \frac{w(H)}{2^{n/2-3m-2}}, \quad (3)$$

where $\mathbb{E}_{V_1,V_2}$ denotes the expectation over $\mathbb{V}_1, \mathbb{V}_2$ independently. $w(H) = \lambda_{\max}(H) - \lambda_{\min}(H)$ denotes the spectral width of H, where $\lambda_{\max}(H)$ is the maximum eigenvalue of H and $\lambda_{\min}(H)$ is the minimum.

+ non-negativity & boundness→

$$\mathrm{Var}_{V_1,V_2}[\Delta_{H,\rho}(V_1, V_2)] \leq \frac{w^2(H)}{2^{n/2-3m-2}}$$

+ Markov's inequality →

$$\Pr[\Delta_{H,\rho}(V_1, V_2) \geq \epsilon] \leq \frac{1}{\epsilon} \cdot \frac{w(H)}{2^{n/2-3m-2}}$$

+ design unitary preservation →

global gate obeying parameter-shift rule

Sketch proof

$$\mathbb{E}_{V_1,V_2}[\Delta_{H,\rho}(V_1,V_2)] \leq \frac{w(H)}{2^{n/2-3m-2}}$$

1. reduce to traceless *H*

$$H \to H + cI,\ c \in \mathbb{R}.$$

2. reduce to max

$$C_{H,\rho}(\mathbf{U}) = \mathrm{tr}(H\mathbf{U}\rho\mathbf{U}^\dagger)$$

$$\Delta_{H,\rho}(V_1,V_2) := \max_{U_A} C_{H,\rho}(\mathbf{U}) - \min_{U_A} C_{H,\rho}(\mathbf{U})$$

$$H \to -H,\quad -\min \to \max$$

3. If V_1 is 2-design

$$\mathbb{E}_{V_1} \max_{U_A}\left[\mathrm{tr}\left(\tilde{H}(U_A \otimes I_B)V_1\rho V_1^\dagger(U_A^\dagger \otimes I_B)\right)\right]$$

$$\tilde{H} = V_2^\dagger H V_2$$

3.(a) Pauli decomposition on A

$$\tilde{H} = \mathrm{tr}_B(\tilde{H}) \otimes \frac{I_B}{2^{n-m}} + \frac{I_A}{2^m} \otimes \mathrm{tr}_A(\tilde{H}) + \sum_{j=1}^{4^m-1} \hat{\sigma}_j^A \otimes O_j^B$$

3.(b) Holder's inequality to relax U_A

$$\left|\mathrm{tr}\left[\left(U_A^\dagger O_A U_A\right)\mathrm{tr}_B\left((I_A \otimes O_B)V\rho V^\dagger\right)\right]\right| \leq \left\|U_A^\dagger O_A U_A\right\|_2 \left\|\mathrm{tr}_B\left((I_A \otimes O_B)V\rho V^\dagger\right)\right\|_2$$

3.(c) 2-design integral & minor relaxation to w(H)

$$\mathbb{E}[\|X\|_2] \leq 2^{m/2}\sqrt{\mathbb{E}[\|X\|_2^2]}$$

(Jensen's inequality)

2-degree

4. If V_2 is 2-design, similar spirit

See Appendix in arXiv: 2205.05056 for detailed proof

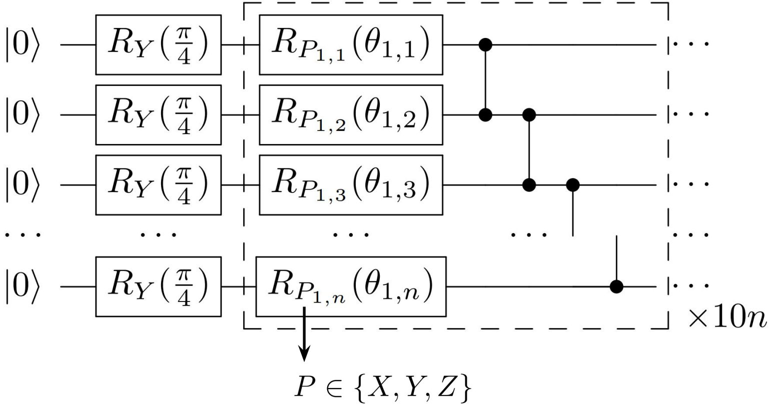
Case study 1: VQE

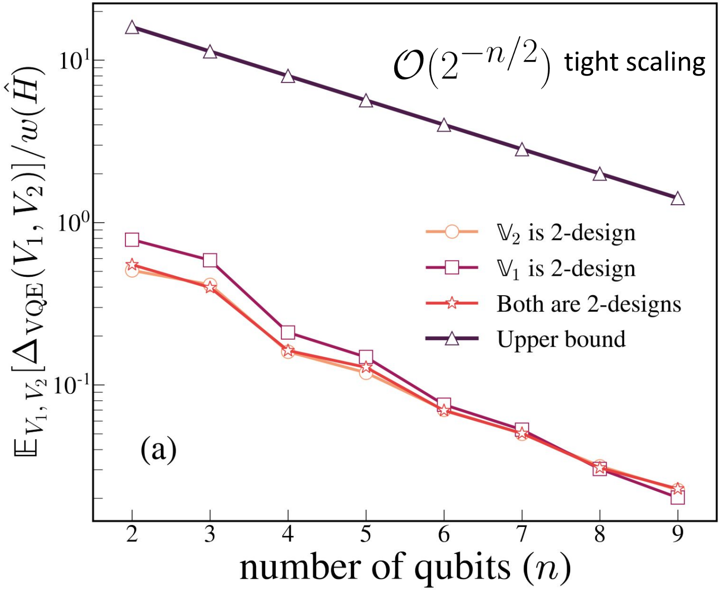
- Variational quantum eigensolver (VQE)

H: Hamiltonian of a physical system, *ρ*: zero state

1-d antiferromagnetic Heisenberg model

$$\hat{H} = \sum_{i=1}^{n} (X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1})$$





Case study 2: autoencoder

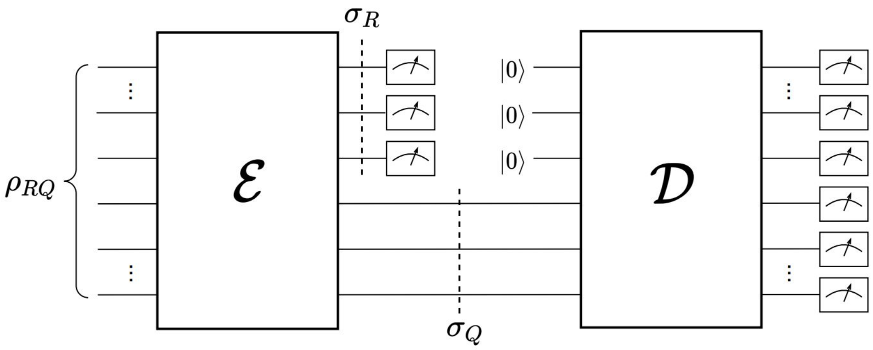
- Quantum autoencoder (QAE)

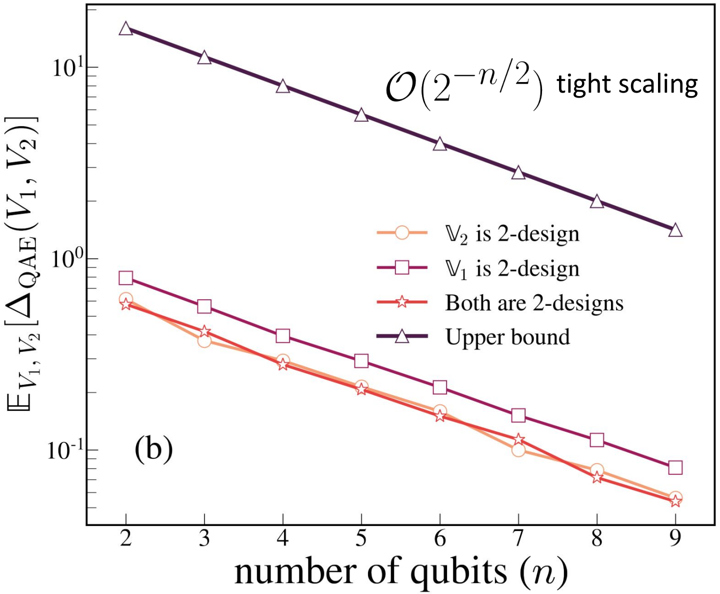
H: zero state of discarded qubits , ρ: given state

1-qubit compression encoder

$$H = -|0\rangle\langle 0|_R \otimes I_Q$$

Expression from relaxation of fidelity





Case study 3: state learning

- Quantum state learning (QSL)

$$H_{\mathrm{QSL}} = -|0\rangle\langle 0|$$

$$C_{\mathrm{QSL}}(\mathbf{U}) = -F(\sigma, \mathbf{U}\rho\mathbf{U}^{\dagger})$$ (generally)

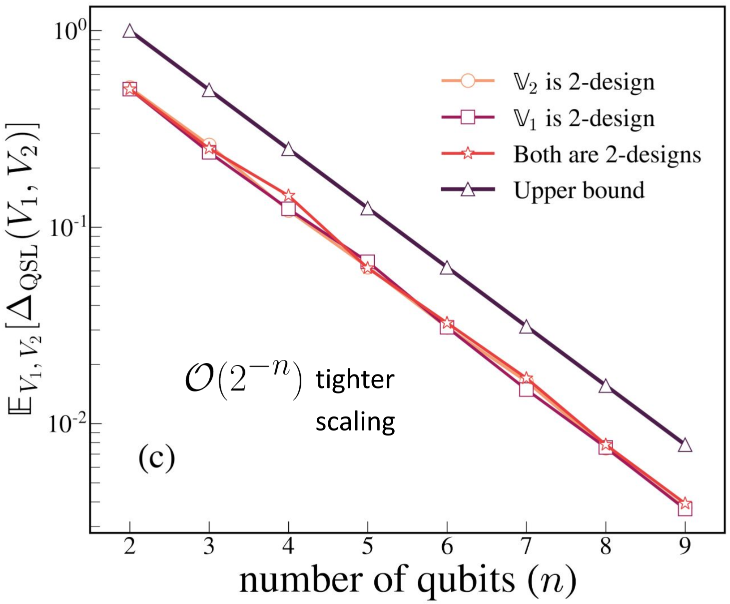
Proposition 2 *Let C_{QSL} be the cost function defined in (16) on an n-qubit system. Suppose $\mathbb{V}_1, \mathbb{V}_2$ are ensembles from which V_1, V_2 are sampled, respectively. If either $\mathbb{V}_1$ or $\mathbb{V}_2$, or both form unitary 1-designs, then the following inequality holds*

$$\mathbb{E}_{V_1,V_2}\left[\Delta_{\mathrm{QSL}}(V_1, V_2)\right] \leq \frac{1}{2^{n-2m}}, \quad (17)$$

where $\mathbb{E}_{V_1,V_2}$ denotes the expectation over $\mathbb{V}_1, \mathbb{V}_2$ independently.

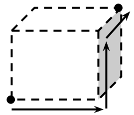
(even a single U3 layer is 1-design)

H: target state, *ρ*: zero state



Beyond BP?

$$\boldsymbol{\theta}^{(\mu)} = \boldsymbol{\theta} + \sum_{\nu=1}^{\mu} (\theta'_\nu - \theta_\nu)\, \mathbf{e}_\nu$$

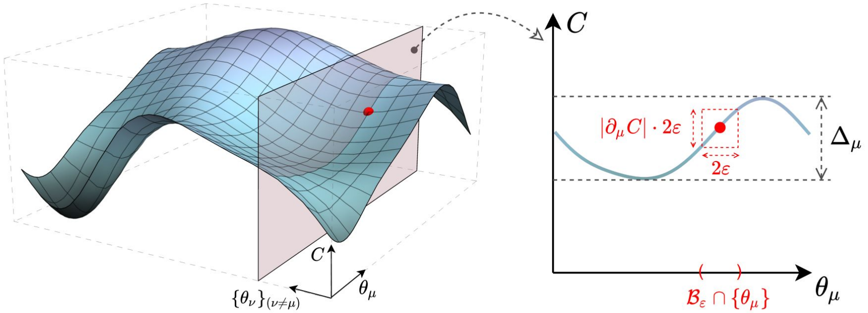


1. Independence with optimizer

(gradient-free methods are based on cost difference)

Unify the restrictions of gradient-based & -free naturally

$$\mathbb{E}\left[|C(\boldsymbol{\theta}') - C(\boldsymbol{\theta})|\right] \leq \mathbb{E}\left[\sum_{\mu=1}^{M} \left|C\left(\boldsymbol{\theta}^{(\mu)}\right) - C\left(\boldsymbol{\theta}^{(\mu-1)}\right)\right|\right] \leq \sum_{\mu=1}^{M} \mathbb{E}\left[|\Delta_\mu|\right] \in \mathcal{O}(M 2^{-n/2}),$$



2. Independence with parameterization

The proof has nothing to do with how θ enters a gate

$$\mathbb{E}\left[|\partial_\mu C|\right] \leq \mathbb{E}\left[\frac{\Delta_\mu}{2\varepsilon}\right] \in \mathcal{O}(2^{-n/2}\frac{1}{\varepsilon})$$

(general)

3. The whole unitary

Useless to replace Ry with U3 when encountering BP

$$\mathbb{E}[|\partial_\mu C|] = \mathbb{E}\left[\left|C\left(\boldsymbol{\theta} + \frac{\pi}{4}\mathbf{e}_\mu\right) - C\left(\boldsymbol{\theta} - \frac{\pi}{4}\mathbf{e}_\mu\right)\right|\right] \leq \mathbb{E}[\Delta_\mu] \in \mathcal{O}(2^{-n/2}),$$

(parameter-shift)

4. state learning suppressed for 1-design

Fidelity is a poor choice for random circuit training

Guidance from this work

- Natural gradient descent ✘
- Gradient-free method ✘
- Gate-by-gate optimization ✘
- Reparameterization ✘
- Clever initialization ?
- Designed architecture ?
- Adaptive method ?
- …

Theorem 1 *Suppose $\mathbb{V}_1, \mathbb{V}_2$ are ensembles from which V_1, V_2 are sampled, respectively. If either $\mathbb{V}_1$ or $\mathbb{V}_2$, or both form unitary 2-designs, then for arbitrary H, ρ, the following inequality holds*

$$\mathbb{E}_{V_1,V_2}[\Delta_{H,\rho}(V_1, V_2)] \leq \frac{w(H)}{2^{n/2-3m-2}}, \tag{3}$$

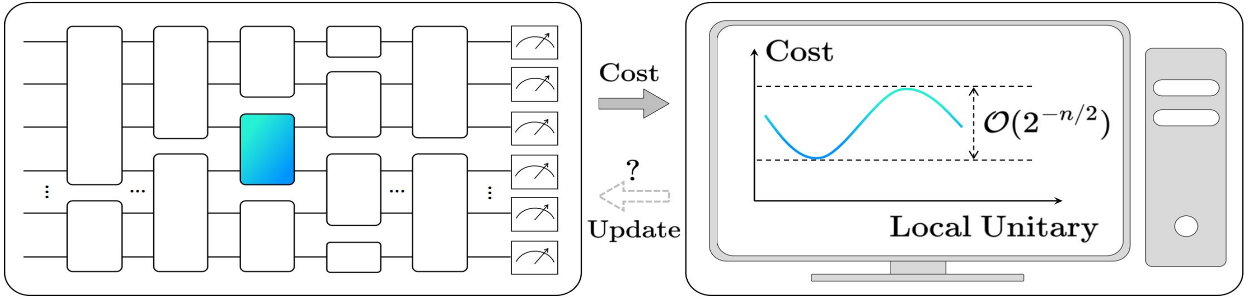
where $\mathbb{E}_{V_1,V_2}$ denotes the expectation over $\mathbb{V}_1, \mathbb{V}_2$ independently. $w(H) = \lambda_{\max}(H) - \lambda_{\min}(H)$ denotes the spectral width of H, where $\lambda_{\max}(H)$ is the maximum eigenvalue of H and $\lambda_{\min}(H)$ is the minimum.

Proposition 2 *Let C_{QSL} be the cost function defined in (16) on an n-qubit system. Suppose $\mathbb{V}_1, \mathbb{V}_2$ are ensembles from which V_1, V_2 are sampled, respectively. If either $\mathbb{V}_1$ or $\mathbb{V}_2$, or both form unitary 1-designs, then the following inequality holds*

$$\mathbb{E}_{V_1,V_2}[\Delta_{\mathrm{QSL}}(V_1, V_2)] \leq \frac{1}{2^{n-2m}}, \tag{17}$$

where $\mathbb{E}_{V_1,V_2}$ denotes the expectation over $\mathbb{V}_1, \mathbb{V}_2$ independently.

Summary



- Barren Plateaus ⇨ Our theorem (variation range)
- Case study: VQE, autoencoder, state learning (tighter bound)
- Implication: reproducing BP, guidance for strategies, …
- Explore the potential solutions
- Go beyond local optimization