PAPER • **OPEN ACCESS**

Equivalence of cost concentration and gradient vanishing for quantum circuits: an elementary proof in the Riemannian formulation

To cite this article: Qiang Miao and Thomas Barthel 2024 *Quantum Sci. Technol.* **9** 045039

IOP Publishing Quantum Sci. Technol. 9 (2024) 045039 https://doi.org/10.1088/2058-9565/ad6fca

Quantum Science and Technology

OPEN ACCESS

RECEIVED
27 March 2024

REVISED
7 August 2024

ACCEPTED FOR PUBLICATION
15 August 2024

PUBLISHED
9 September 2024

PAPER

Equivalence of cost concentration and gradient vanishing for quantum circuits: an elementary proof in the Riemannian formulation

Qiang Miao[1] and Thomas Barthel[1,2,3,*]

[1] Duke Quantum Center, Duke University, Durham, NC 27701, United States of America
[2] Department of Physics, Duke University, Durham, NC 27708, United States of America
[3] Tensor Center, Auf dem Dresch 15, 52152 Simmerath, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: thomas.barthel@duke.edu

Keywords: variational quantum algorithms, barren plateaus, cost-function concentration, quantum circuits, Riemannian optimization, tensor networks

Abstract

The optimization of quantum circuits can be hampered by a decay of average gradient amplitudes with increasing system size. When the decay is exponential, this is called the barren plateau problem. Considering explicit circuit parametrizations (in terms of rotation angles), it has been shown in Arrasmith *et al* (2022 *Quantum Sci. Technol.* **7** 045015) that barren plateaus are equivalent to an exponential decay of the variance of cost-function differences. We show that the issue is particularly simple in the (parametrization-free) Riemannian formulation of such optimization problems and obtain a tighter bound for the cost-function variance. An elementary derivation shows that the single-gate variance of the cost function is *strictly equal* to half the variance of the Riemannian single-gate gradient, where we sample variable gates according to the uniform Haar measure. The total variances of the cost function and its gradient are then both bounded from above by the sum of single-gate variances and, conversely, bound single-gate variances from above. So, decays of gradients and cost-function variations go hand in hand, and barren plateau problems cannot be resolved by avoiding gradient-based in favor of gradient-free optimization methods.

1. Introduction

Recent rapid advancements in quantum computing hardware enable the implementation of large and deep quantum circuits, reaching regimes beyond the simulation capabilities of classical computers. A promising scheme to harness this potential before the advent of practical fault-tolerance are variational quantum algorithms (VQAs) [1]: quantum circuits are executed on quantum computers and the quantum-gate parameters are optimized through a classical backend to minimize a given cost function. A critical challenge in such hybrid quantum–classical optimizations consists in noise and the probabilistic nature of quantum measurements. In generic variational quantum circuits, average gradient amplitudes tend to decrease *exponentially* in the system size (number of qudits). This phenomenon is known as *barren plateaus* [2, 3]. Unless one already has a very good guess for the optimal circuit, the barren plateau problem implies that we would need an exponential number of measurement shots for a sufficiently accurate determination of cost-function gradients, prohibiting the application for large problem sizes. Otherwise, we would very likely end up with random walks in flat regions of the cost landscape. Numerous works investigate how to avoid an exponential decay of gradient amplitudes [4–18], and the absence of barren plateaus might imply classical simulability [19].

The quantum circuits in VQA can comprise fixed unitary gates $\{\hat{W}_1, \hat{W}_2, \ldots\}$ and variable unitary gates $\{\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_K\}$. For example, the former could be controlled NOT (CNOT) gates and the latter single-qubit gates. The variable gates are typically parametrized by rotation angles, $\hat{U}_i = \hat{U}_i(\boldsymbol{\theta}_i) \in \mathrm{U}(N_i)$, and the optimization is based on the Euclidean metric in the angle space $\{\boldsymbol{\theta}_i\}$. Using this framework and

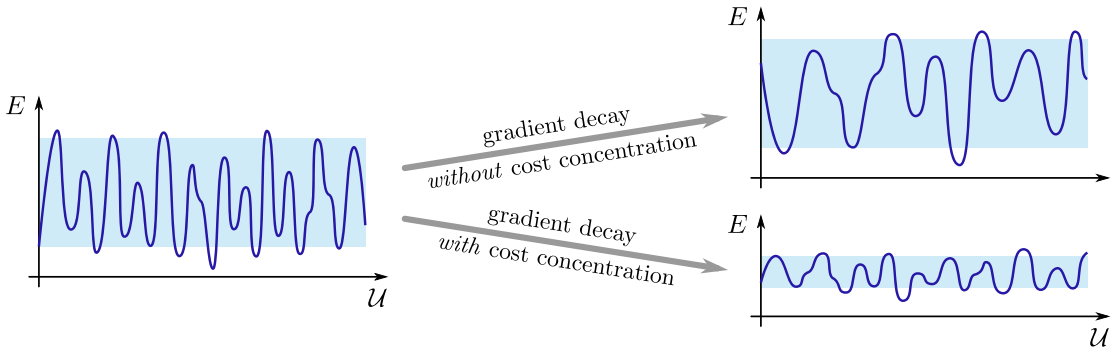

Figure 1. An increase in dimensionality (scaling up the system size n) can lead to a decay of gradients. The situation where the average gradient amplitude decays exponentially in n is the so-called barren-plateau problem. In general, gradient decay may or may not be accompanied by concentration of the cost function. As discussed here and in [20] the two phenomena go hand in hand for VQA.

assuming that the variables gates are composed of rotations $\mathrm{e}^{-\mathrm{i}\theta_{i,k}\hat{\sigma}_k}$ with involutory generators $\hat{\sigma}_k = \hat{\sigma}_k^\dagger = \hat{\sigma}_k^{-1}$, Arrasmith *et al* established the equivalence of barren plateaus and an exponential decay of the variance of cost-function differences with respect to increasing system size [20].

Alternatively, one can formulate the circuit optimization problem directly over the manifold

$$\mathcal{M} = \mathrm{U}(N_1) \times \mathrm{U}(N_2) \times \cdots \times \mathrm{U}(N_K) \tag{1}$$

formed by the direct product of the gates' unitary groups in a representation-free form. In this Riemannian approach [21, 22], gradients are elements of the tangent space of $\mathcal{M}$, and one can implement line searches and Riemannian quasi-Newton methods through retractions and vector transport on $\mathcal{M}$ as discussed in recent works [23, 24]. Riemannian optimization has some advantages over the Euclidean optimization of parametrized quantum circuits. For example, it avoids cost-function saddle points that are introduced when employing a global parametrization $\{\boldsymbol{\theta}_i\}$ of the manifold $\mathcal{M}$ (consider, e.g. sitting at the north pole of a sphere and rotating around the z axis). Furthermore, the Riemannian formulation can simplify analytical considerations, e.g. concerning average gradient amplitudes [16, 17] and cost-function variances as discussed in the following.

In this report, we establish a direct connection between cost-function concentration and the decay of Riemannian gradient amplitudes in the optimization of quantum circuits (see figure 1). The proof in the Riemannian formulation is surprisingly simple and, compared to [20], yields tighter bounds. We will show that when the gates are sampled according to the uniform Haar measure, the single-gate cost-function variance is exactly half the single-gate variance of the Riemannian gradient. The corresponding total variances, where all gates are varied simultaneously, are both bounded from above by sums of the single-gate variances. Furthermore, the total variances bound all individual single-gate variances. As a consequence, the barren plateau problem can be equivalently diagnosed through the analysis of cost function concentration and cannot be resolved by switching from gradient-based optimization to a gradient-free optimization [20, 25].

2. Cost function and Riemannian gradient

Consider a generic quantum circuit $\hat{\mathcal{U}}$ composed of some fixed unitary gates $\{\hat{W}_1, \hat{W}_2, \ldots\}$ and variable unitary gates $\{\hat{U}_1, \hat{U}_2, \ldots\}$ over which we optimize. Starting from a reference state $\hat{\rho}_0$, the circuit prepares the state $\hat{\rho} = \hat{\mathcal{U}}\hat{\rho}_0\hat{\mathcal{U}}^\dagger$. With an observable $\hat{O}$, the cost function takes the form

$$E(\{\hat{U}_i\}) = \mathrm{Tr}\left(\hat{\mathcal{U}}\hat{\rho}_0\hat{\mathcal{U}}^\dagger\hat{O}\right) \tag{2}$$

With $\hat{\rho}_0 = \sum_{s=1}^S \hat{\rho}_s \otimes |s\rangle\langle s|$ and $\hat{O} = \sum_{s=1}^S \hat{O}_s \otimes |s\rangle\langle s|$, this setup also covers the more general case $E(\{\hat{U}_i\}) = \sum_{s=1}^S \mathrm{Tr}\left(\hat{\mathcal{U}}\hat{\rho}_s\hat{\mathcal{U}}^\dagger\hat{O}_s\right)$ with a training set $\{(\hat{\rho}_s, \hat{O}_s)\}$ of S initial states $\hat{\rho}_s$ and associated measurement operators $\hat{O}_s$ [20].

Considering the dependence on one of the variable gates, $\hat{U} \in \mathrm{U}(N)$, we can write the cost function in the compact form

$$E(\hat{U}) = \mathrm{Tr}\left(\hat{Y}\tilde{U}\hat{X}\tilde{U}^\dagger\right) \quad \text{with} \quad \tilde{U} := \hat{U} \otimes \mathbb{1}_M \quad \text{and} \quad \hat{X}, \hat{Y} \in \mathrm{End}\left(\mathbb{C}^N \otimes \mathbb{C}^M\right) \tag{3}$$

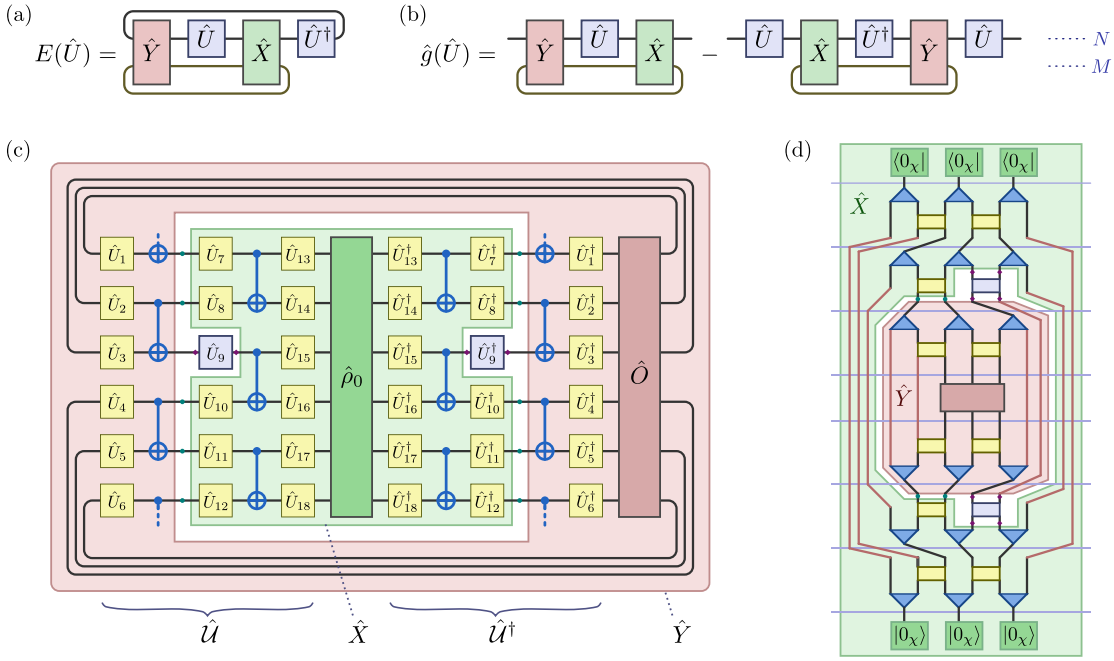

Figure 2. Diagrammatic representations for (a) cost function $E(\hat{U})$ and (b) Riemannian gradient $\hat{g}(\hat{U})$, where we consider the dependence on a single gate $\hat{U} \in \mathrm{U}(N)$ from the variable gates $\{\hat{U}_i \in \mathrm{U}(N_i)\}$ that compose the quantum circuit $\hat{\mathcal{U}}$; cf equations (3) and (4). Cost functions of this form arise in various applications like quantum machine learning and variational quantum algorithms for the investigation of many-body ground states. Panel (c) shows how a cost function (2) for a quantum circuit consisting of single-qubit and CNOT gates attains the form (3). Panel (d) shows an example for the variational optimization of a multiscale entanglement renormalization ansatz (MERA) [26, 27]—a hierarchical tensor network state that features unitary disentanglers (yellow squares) and isometries (blue triangles). See [16, 17, 23] for more context.

as illustrated in figure 2(a), where the Hermitian operator $\hat{X}$ on $\mathbb{C}^N \otimes \mathbb{C}^M$ comprises $\hat{\rho}_0$, $\hat{Y}$ comprises $\hat{O}$, and both comprise further circuit gates except $\hat{U}$. See figure 2(c) for an example. As discussed in [16, 17], expectation values $\langle\Psi|\hat{H}|\Psi\rangle$ of a Hamiltonian $\hat{H}$ with respect to isometric tensor network states (TNS) $|\Psi\rangle = \hat{\mathcal{U}}|0\rangle$ can also be written in the form (3). In this case the TNS are generated from a pure reference state $|0\rangle$ by application of a quantum circuit, and $\hat{U}$ corresponds to one tensor of the TNS. The example of a multiscale entanglement renormalization ansatz (MERA) [26, 27] is illustrated in figure 2(d).

Here and in section 3, we consider variation of one specific unitary gate $\hat{U}$ such that the Riemannian manifold is just $\mathrm{U}(N)$; this is referred to as a *'single-gate'* variation. The extension to variation of all gates (*'total'* variation) on the full manifold (1) will be discussed in section 4. Projecting the gradient $\hat{d} = \partial_{\hat{U}} E(\hat{U}) = 2\,\mathrm{Tr}_M(\hat{Y}\tilde{U}\hat{X})$ of the cost function (3) onto the tangent space of the unitary group $\mathrm{U}(N)$ at $\hat{U}$, we obtain the Riemannian gradient

$$\hat{g}(\hat{U}) = \mathrm{Tr}_M\left(\hat{Y}\tilde{U}\hat{X} - \tilde{U}\hat{X}\tilde{U}^\dagger\hat{Y}\tilde{U}\right) \tag{4}$$

as illustrated in figure 2(b). Given that we need to stay on the manifold $\mathrm{U}(N)$ during the optimization, $\hat{g}$ is the relevant direction of change. As discussed in [23, 24], it can be efficiently measured on quantum computers. Here and in the following, Tr_N and Tr_M denote the partial traces over the first and second components of $\mathbb{C}^N \otimes \mathbb{C}^M$, respectively.

Let us summarize the derivation of equation (4): The $N \times N$ unitary gates are embedded in the $2N^2$ real Euclidean space $\mathcal{E} = \mathrm{End}(\mathbb{C}^N) \simeq \mathbb{R}^{2N^2}$. The gradient in this embedding space is $\hat{d}$. Using the (Euclidean) metric $(\hat{U}, \hat{U}') := \mathrm{Re}\,\mathrm{Tr}(\hat{U}^\dagger\hat{U}')$ for the embedding space $\mathcal{E}$, $\hat{d}$ fulfills $\partial_\varepsilon E(\hat{U}+\varepsilon\hat{V})|_{\varepsilon=0} = (\hat{d}, \hat{V})$ for all $\hat{V}$. For Riemannian optimization algorithms, one needs to project $\hat{d}$ onto the tangent space $\mathcal{T}_{\hat{U}}$ of $\mathrm{U}(N)$ at $\hat{U}$, and then construct retractions for line search, and vector transport to form linear combinations of gradient vectors from different points on the manifold [21–23]. An element $\hat{V}$ of the tangent space $\mathcal{T}_{\hat{U}}$ needs to obey $(\hat{U}+\varepsilon\hat{V})^\dagger(\hat{U}+\varepsilon\hat{V}) = \mathbb{1} + \mathcal{O}(\varepsilon^2)$ such that

$$\mathcal{T}_{\hat{U}} = \left\{\mathrm{i}\hat{U}\hat{\eta}\,|\,\hat{\eta} = \hat{\eta}^\dagger \in \mathrm{End}\left(\mathbb{C}^N\right)\right\}. \tag{5}$$

The projection $\hat{g}$ of $\hat{d}$ onto this tangent space obeys $(\hat{V}, \hat{g}) = (\hat{V}, \hat{d})$ for all $\hat{V} \in \mathcal{T}_{\hat{U}}$. This gives the Riemannian gradient $\hat{g} = (\hat{d} - \hat{U}\hat{d}^\dagger\hat{U})/2$ which results in equation (4).

3. Single-gate Haar-measure variances

To evaluate averages and variances over $\mathrm{U}(N)$ (or, more generally the manifold $\mathcal{M}$), we employ Haar-measure integrals. The average of the Riemannian gradient (4) is zero,

$$\mathrm{Avg}_{\hat{U}}\hat{g} := \int_{\mathrm{U}(N)} \mathrm{d}U \hat{g}(\hat{U}) = \frac{1}{2}\int_{\mathrm{U}(N)} \mathrm{d}U \left[\hat{g}(\hat{U}) + \hat{g}(-\hat{U})\right] = 0, \tag{6}$$

because $\hat{g}$ is an odd function in $\hat{U}$. For the evaluation of $\mathrm{Avg}_{\hat{U}}E$ and the variances, we only need the first and second-moment Haar-measure integrals over the unitary group. From the Weingarten formulas [28, 29] for the first and second moments, one obtains [17]

$$\mathrm{Avg}_{\hat{U}}\, \hat{U}^\dagger \otimes \hat{U} = \tfrac{1}{N}\,\mathrm{Swap} \quad \text{and} \tag{7a}$$

$$\mathrm{Avg}_{\hat{U}}\, \hat{U}^\dagger \otimes \hat{U} \otimes \hat{U}^\dagger \otimes \hat{U} = \tfrac{1}{N^2-1}\left(1 - \tfrac{1}{N}\,\mathrm{Swap}_{2,4}\right)\left(\mathrm{Swap}_{1,2}\,\mathrm{Swap}_{3,4} + \mathrm{Swap}_{1,4}\,\mathrm{Swap}_{2,3}\right) \tag{7b}$$

with $\mathrm{Swap} = \sum_{i,j=1}^{N} |i,j\rangle\langle j,i|$ and $\mathrm{Swap}_{k,\ell}$ swaps the kth and ℓth components of $\mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N$. Graphical representations of equations (7*a*) and (7*b*) are shown in figures 5(a) and (c). Weingarten formulas can be proven using the Schur-Weyl duality and the double centralizer theorem [29]. An illustrating proof for equation (7*a*) is given in appendix A.

Applying the Weingarten formulas (7), we find a simple linear relation between the single-gate cost-function variance

$$\mathrm{Var}_{\hat{U}}E = \mathrm{Avg}_{\hat{U}}E^2 - \left(\mathrm{Avg}_{\hat{U}}E\right)^2 \quad \text{over} \quad \mathrm{U}(N) \tag{8}$$

and the single-gate gradient variance $\mathrm{Var}_{\hat{U}}\hat{g}$. With the average gradient (6) being zero, we can quantify the gradient variance by

$$\mathrm{Var}_{\hat{U}}\hat{g} := \mathrm{Avg}_{\hat{U}}\frac{1}{N}\mathrm{Tr}\left(\hat{g}^\dagger\hat{g}\right) \quad \text{over} \quad \mathrm{U}(N). \tag{9}$$

This definition can be motivated as follows [17]: As any element of the tangent space (5), the gradient $\hat{g}$ can be expanded in an orthonormal basis of involutory Hermitian operators $\{\hat{\sigma}_k \,|\, \hat{\sigma}_k = \hat{\sigma}_k^\dagger = \hat{\sigma}_k^{-1}\}$ for $\mathrm{End}(\mathbb{C}^N)$ with $\mathrm{Tr}(\hat{\sigma}_k\hat{\sigma}_{k'}) = N\delta_{k,k'}$. This gives the gradient in the form $\hat{g} = \mathrm{i}\hat{U}\sum_{k=1}^{N^2}\alpha_k\hat{\sigma}_k/N$, where each α_k corresponds to the derivative of one rotation angle. Hence, $\mathrm{Tr}(\hat{g}^\dagger\hat{g})/N = \sum_k \alpha_k^2/N^2$, i.e. equation (9) coincides with the average variance of the rotation-angle derivatives [30].

An elementary proof given in appendix B establishes a linear relation between the single-gate cost-function variance (8) and gradient variance (9).

Theorem 1 (Exact equivalence of single-gate cost-function and gradient variances). *In the Riemannian formulation, the variance of the cost function (2) is exactly half the variance of the Riemannian gradient (4) when considering the dependence on one of the unitary gates of the quantum circuit ($\hat{U}_{j\neq i}$ fixed), i.e.*

$$\mathrm{Var}_{\hat{U}_i} E\left(\{\hat{U}_j\}\right) = \frac{1}{2}\mathrm{Var}_{\hat{U}_i}\hat{g}_i\left(\{\hat{U}_j\}\right) \qquad \forall\ \{\hat{U}_{j\neq i}\}. \tag{10}$$

Of course, the proportionality of these *conditional* single-gate variances translates directly into a proportionality of the averaged single-gate variances (the conditional variances (10) averaged over all $\hat{U}_{j\neq i}$),

$$V_i := \mathrm{Avg}_{\{\hat{U}_{j\neq i}\}}\mathrm{Var}_{\hat{U}_i}E\left(\{\hat{U}_j\}\right) \overset{(10)}{=} \frac{1}{2}\mathrm{Avg}_{\{\hat{U}_{j\neq i}\}}\mathrm{Var}_{\hat{U}_i}\hat{g}_i\left(\{\hat{U}_j\}\right). \tag{11}$$

4. Total Haar-measure variances

In section 3, we only considered the dependence of the cost function (2) on one of the unitary gates ($\hat{U}$) in the circuit as well as the single-gate gradient (4). In this section, we consider the dependence on all variable unitary gates $(\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_K) \in \mathcal{M}$ with $\hat{U}_i \in \mathrm{U}(N_i)$ and the corresponding total variances like

$$\mathrm{Var}_{\{\hat{U}_j\}}E \equiv \mathrm{Avg}_{\{\hat{U}_j\}}\left(E^2\right) - \left(\mathrm{Avg}_{\{\hat{U}_j\}}E\right)^2 \tag{12}$$

for the cost function.

The full Riemannian gradient of the cost function (2) with respect to all variable gates is simply the direct sum of the individual gradients, i.e.

$$\hat{g}_{\text{full}} = \hat{g}_1 \oplus \hat{g}_2 \oplus \cdots \oplus \hat{g}_k \quad \text{with} \quad \hat{g}_i = \text{Tr}_{M_i}\left(\hat{Y}_i \tilde{U}_i \hat{X}_i - \tilde{U}_i \hat{X}_i \tilde{U}_i^\dagger \hat{Y}_i \tilde{U}_i\right). \tag{13}$$

Here $\tilde{U}_i := \hat{U}_i \otimes \mathbb{1}_{M_i}$ with $\hat{X}_i$ and $\hat{Y}_i$ depending on the remaining gates of the circuit, and $\hat{\rho}_0$ as well as $\hat{O}$ as in equation (3). In extension of equation (9), we define the total variance of $\hat{g}_{\text{full}}$ as

$$\text{Var}_{\{\hat{U}_j\}} \hat{g}_{\text{full}} := \frac{1}{K}\sum_{i=1}^{K} \text{Avg}_{\{\hat{U}_j\}} \frac{1}{N_i} \text{Tr}\left(\hat{g}_i^\dagger \hat{g}_i\right) \overset{(9)}{=} \frac{1}{K}\sum_{i=1}^{K} \text{Avg}_{\{\hat{U}_{j\neq i}\}} \text{Var}_{\hat{U}_i} \hat{g}_i. \tag{14}$$

The following central result as proven in appendix C is based on an analysis of covariances, the law of total variance, and theorem 1.

Theorem 2 (equivalence of circuit cost-function concentration and gradient vanishing). *When averaging over the variable unitaries $\{\hat{U}_i\}$ of the quantum circuit $\hat{\mathcal{U}}$ according to the Haar measure, the total variance of the cost function (2) and the total variance of the full Riemannian gradient (14) are both bounded from below by single-gate variances V_i (equation (11)), and they are bounded from above by or proportional to the sum $\sum_i V_i$,*

$$V_j \leqslant \quad \text{Var}_{\hat{U}_1,\ldots,\hat{U}_K} E\left(\hat{U}_1,\ldots,\hat{U}_K\right) \leqslant \sum_{i=1}^{K} V_i \quad \forall\, j \quad \textit{and} \tag{15a}$$

$$V_j \overset{(14)}{\leqslant} \frac{K}{2} \text{Var}_{\hat{U}_1,\ldots,\hat{U}_K} \hat{g}_{\text{full}}\left(\hat{U}_1,\ldots,\hat{U}_K\right) \overset{(14)}{=} \sum_{i=1}^{K} V_i \quad \forall\, j. \tag{15b}$$

In particular, if all single-gate variances V_i of polynomial-depth circuits ($K = \text{poly}\, n$) decay exponentially in the system size (number of qudits) n, then both total variances (12) and (14) decay exponentially in n. Conversely, if one of the total variances decays exponentially in n, then all single-gate variances also decay exponentially. So, the barren-plateau problem and exponential cost-function concentration always appear simultaneously.

Note that the conclusions below equation (15*b*) remain valid if we choose a different weighting in the definition of the full gradient variance (14). For example, we could also define it as $\frac{1}{\sum_{i=1}^{K} N_i^2}\sum_{i=1}^{K} N_i \,\text{Avg}_{\{\hat{U}_j\}} \text{Tr}(\hat{g}_i^\dagger \hat{g}_i)$, corresponding to an equal weighting of all rotation-angle derivatives in the parametrization discussed below equation (9).

5. Numerical verification

For an illustration of the general bounds on the cost-function variance in theorem 2, consider a one-dimensional binary MERA $|\Psi\rangle$ for spin-1/2 chains of length $L = 3 \cdot 2^T$, where T is the number of layers in the MERA [26, 27]. The cost function is given by the energy (density) expectation value

$$E = \langle\Psi|\hat{H}|\Psi\rangle, \quad \text{where the Hamiltonian} \quad \hat{H} = \frac{1}{\sqrt{24}\,L}\sum_{i=1}^{L}\sum_{a=x,y,z} \hat{\sigma}_i^a \hat{\sigma}_{i+1}^a \hat{\sigma}_{i+2}^a \tag{16}$$

is a sum of three-site interaction terms with Pauli operators $\hat{\sigma}_i^x, \hat{\sigma}_i^y, \hat{\sigma}_i^z$ acting on site i. In the evaluation of the variances, Haar-averages are executed by numerical sampling, and we denote the single-gate variances (11) by $V_{\tau,k}$, where the position (τ, k) indicates the kth tensor in layer τ.

As shown in figure 3 for MERA with bond dimension $\chi = 2$, the total cost-function variance $\text{Var}(E)$ (equation (12)) is, in accordance with equation (15*a*), bounded from above by the sum of all single-gate variances $\sum_{\tau,k} V_{\tau,k}$, and the single-gate variances $V_{\tau,k}$ provide lower bounds. Within a given layer τ, the single-gate variances $V_{\tau,k}$ are approximately constant and, as shown in [16, 17], they decrease exponentially as

$$V_{\tau,k} \propto (324/625)^\tau. \tag{17}$$

Hence, the best lower bound in figure 3 is given by the maximum of $V_{1,k}$.

Note that the energy optimization problems for MERA, tree tensor networks states [31–33], and matrix product states [34, 35] for Hamiltonian with finite-range interactions are actually free of barren plateaus [16, 17]. Nevertheless, the general variance relations from theorem 2 apply.

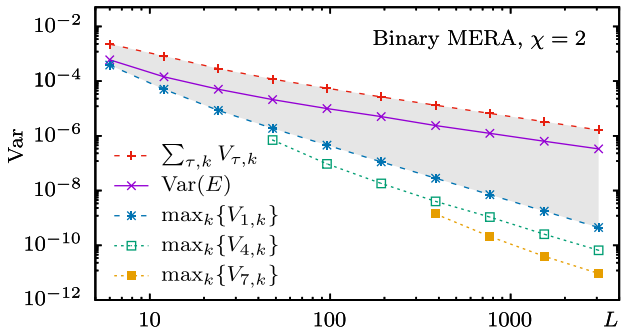

Figure 3. Numerical verification of theorem 2 for binary one-dimensional MERA [26, 27] with bond dimension $\chi = 2$ and the cost function E given by the energy expectation value (16). In accordance with equation (15*a*), the cost-function variance is bounded from below by single-gate variances $V_{\tau,k}$ and from above by $\sum_{\tau,k} V_{\tau,k}$.

6. Discussion

Given the equivalence of cost-function concentration and gradient vanishing on both the single-gate as well as full-circuit levels (theorems 1 and 2, respectively), we can assess gradient vanishing and, especially, barren plateaus more easily through the scalar cost function. In fact, this route has already been pursued in recent analytic works on the trainability of VQAs [19, 36–40].

Inspired by the work of Arrasmith *et al* [20] on the parametrized circuits and Euclidean gradients, we studied the question in the Riemannian formulation which makes the proofs rather simple and yields additional insights: (a) The single-gate variances of gradients and of the cost function turn out to be strictly proportional. (b) In the Euclidean formulation, Arrasmith *et al* obtained results for the variance of cost-function *differences* like $\mathrm{Var}_{\boldsymbol{\theta}}\left(E(\boldsymbol{\theta}') - E(\boldsymbol{\theta})\right) \leqslant \mathcal{O}\left(K^2(n)V(n)\right)$, where $\boldsymbol{\theta}'$ is a random reference point, $K(n)$ is the number of variable gates as a function of the system size n, and $V(n)$ is a common upper bound for all single-gate (gradient) variances V_i. This difference construction turned out to be unnecessary in the Riemannian formulation and we could access $\mathrm{Var}_{\hat{U}_1,\ldots,\hat{U}_K} E$ directly. (c) Furthermore, we obtained the tighter bound $\mathrm{Var}_{\hat{U}_1,\ldots,\hat{U}_K} E \leqslant K(n)V(n)$. This result aligns with our experience in numerical simulations and could probably be further tightened.

While quantifying the total cost-function variance is easier than studying single-gate gradient variances or, equivalently, single-gate cost-function variances, the latter provide more detailed trainability information. For example, the single-gate variances in MERA tensor networks vary strongly from layer to layer. Gates in lower layers have a more substantial impact on the cost function landscape than those in upper layers. This can be taken into account to improve optimization schemes [41].

As a specific example, consider the optimization of the quantum circuit that defines the MERA $|\Psi\rangle$ to minimize the energy expectation value $\langle\Psi|\hat{H}|\Psi\rangle$ for the spin-1/2 transverse-field Ising chain

$$\hat{H} = -\sum_{i=1}^{L} \left(\hat{\sigma}_i^x \hat{\sigma}_{i+1}^x + h\hat{\sigma}_i^z\right). \tag{18}$$

The Ising chain has a critical point at $|h| = 1$, where the groundstate is particularly strongly entangled, featuring the entanglement log-area law. It follows from the analysis in [16, 17] that the total cost-function variance is (up to finite-size corrections) independent of the system size L and decays algebraically with increasing MERA bond dimension χ. This means that there is no barren-plateau problem and the optimization is in general possible. The single-gate variances provide more detailed information: MERA are hierarchical tensor networks with a layer structure. It turns out that the single-gate variances decay exponentially in the layer index $\tau = 1, \ldots, T$, where T is the number of MERA layers [16, 17]. See equation (17) for binary MERA with $\chi = 2$. As demonstrated in figure 4, this suggests a more efficient optimization scheme [41], where we start by setting the gates of layers $\tau \geqslant 2$ to $\hat{U}_{\tau,k} = \mathbb{1}$ and, initially, only optimize those of layer $\tau = 1$. After a suitable number of iterations, we proceed by optimizing the gates of layers $\tau \leqslant 2$, then those of layers $\tau \leqslant 3$ and continue in this way, building up the MERA circuit layer by layer. The numerical results close to the critical point of the model confirm that this scheme is more efficient than the traditional approach of, right away, optimizing all layers simultaneously. On average, one achieves a higher energy accuracy and less circuits remain stuck in local minima.

While single-gate variances provide considerably more information than the total cost-function variance alone, they still give limited insight about trainability and convergence properties. They can show that

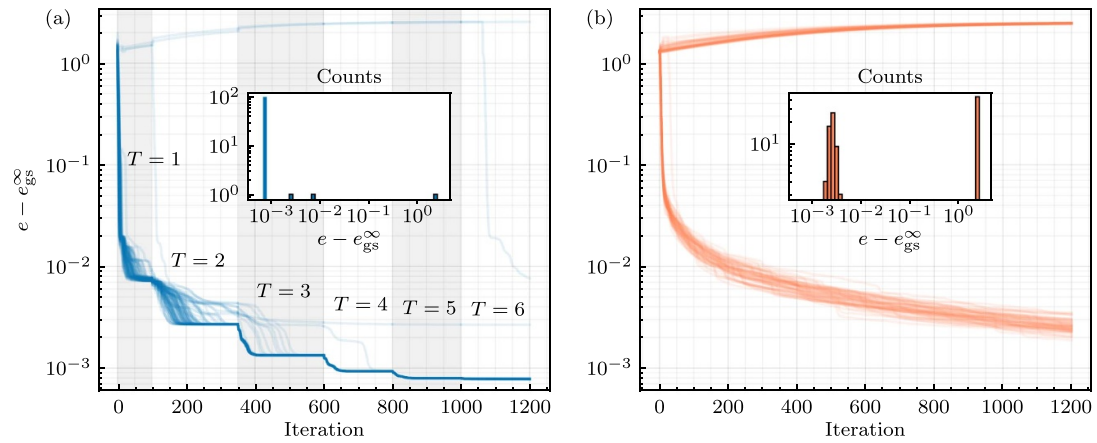

Figure 4. Optimization of randomly initialized binary-MERA quantum circuits with bond dimension $\chi = 2$ for the minimization of the energy expectation value of the spin-1/2 transverse-field Ising chain (18) at $h = 1.03$. The plots show the optimization history for the deviation of the energy density $e = \langle\Psi|\hat{H}|\Psi\rangle/L$ from the exact groundstate energy density e_{gs}^{∞}. Insets show histograms for the accuracy of 100 randomly initialized MERA with six layers after 1200 iterations. (a) The single-gate variances (11) turn out to decay exponentially in the layer index τ [16, 17]. As indicated by the gray regions, we hence, first optimize layer $\tau = 1$ only, then, after 100 iterations, all tensors of layers $\tau \leqslant 2$, then all tensors of layers $\tau \leqslant 3$ etc (b) In contrast, using the traditional apporach of simultaneously optimizing all layers from the very beginning leads to considerably slower convergence and more circuits stuck in local minima.

gradient amplitudes are *on average* above or below certain thresholds, but they are certainly not a measure for the complexity of the cost function landscape, the importance of local minima, or specifics of optimization trajectories.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

We gratefully acknowledge helpful discussions with Baoyou Qu and support by the National Science Foundation (NSF) Quantum Leap Challenge Institute for Robust Quantum Simulation (Award No. OMA-2120757).

Appendix A. Proof of the first Weingarten formula (7*a*)

Proof. Consider the Haar-measure integral

$$\hat{B} := \int_{\mathrm{U}(N)} \mathrm{d}U\, \hat{U}^{\dagger}\hat{A}\hat{U} \tag{A1}$$

over the unitary group $\mathrm{U}(N)$, where $\hat{A}, \hat{B} \in \mathrm{End}(\mathbb{C}^N)$ are linear operators. Due to the invariance of the Haar measure, we have

$$\hat{V}^{\dagger}\hat{B}\hat{V} = \int_{\mathrm{U}(N)} \mathrm{d}U\, (\hat{U}\hat{V})^{\dagger}\hat{A}(\hat{U}\hat{V}) = \hat{B} \quad \text{for all} \quad \hat{V} \in \mathrm{U}(N)\,. \tag{A2}$$

According to Schur's lemma, a linear operator $\hat{B}$ that commutes with all elements $\hat{V}$ of the unitary group is a scalar multiple of the identity, i.e. $\hat{B} = \lambda \mathbb{1}_N$. Taking the trace, we have

$$N\lambda = \mathrm{Tr}\,\hat{B} \overset{(A1)}{=} \int_{\mathrm{U}(N)} \mathrm{d}U\, \mathrm{Tr}\left(\hat{U}^{\dagger}\hat{A}\hat{U}\right) = \int_{\mathrm{U}(N)} \mathrm{d}U\, \mathrm{Tr}(\hat{A}) = \mathrm{Tr}(\hat{A}). \tag{A3}$$

Now, choosing $\hat{A} = |i\rangle\langle j|$ for an orthonormal basis $\langle i|j\rangle = \delta_{i,j}$, we can conclude that

$$\frac{\mathrm{Tr}(\hat{A})}{N}\mathbb{1}_N = \int_{\mathrm{U}(N)} \mathrm{d}U\, \hat{U}^\dagger \hat{A} \hat{U} \quad \Rightarrow \quad \int_{\mathrm{U}(N)} \mathrm{d}U\, U^*_{i,k} U_{j,\ell} = \frac{1}{N}\delta_{i,j}\delta_{k,\ell}, \tag{A4}$$

which is the first Weingarten formula and equivalent to equation (7*a*). See also figure 5(a). □

Appendix B. Proof of theorem 1

Proof. Applying the first Weingarten formula (7*a*), illustrated in figure 5(a), the Haar-measure average of the cost-function (3) evaluates to

$$\mathrm{Avg}_{\hat{U}} E \overset{(7a)}{=} \frac{1}{N}\mathrm{Tr}\left(\mathrm{Tr}_N(\hat{X})\,\mathrm{Tr}_N(\hat{Y})\right) = \frac{1}{N}\mathrm{Tr}\,\hat{Z} \tag{B1}$$

as shown in figure 5(b), where $\hat{Z} \in \mathrm{End}(\mathbb{C}^N \otimes \mathbb{C}^N)$ with

$$\langle i_1, i_2|\hat{Z}|j_1, j_2\rangle = \sum_{m,m'=1}^{M} \langle i_1, m|\hat{X}|j_1, m'\rangle \langle i_2, m'|\hat{Y}|j_2, m\rangle. \tag{B2}$$

Applying the second Weingarten formula (7*b*), illustrated in figure 5(c), the second moment of the cost function evaluates to

$$\mathrm{Avg}_{\hat{U}} E^2 \overset{(7b)}{=} \frac{1}{N^2-1}\left((\mathrm{Tr}\,\hat{Z})^2 - \frac{\mathrm{Tr}\left((\mathrm{Tr}_1\hat{Z})^2 + (\mathrm{Tr}_2\hat{Z})^2\right)}{N} + \mathrm{Tr}\,\hat{Z}^2\right), \tag{B3}$$

where $\mathrm{Tr}_1\hat{Z}$ and $\mathrm{Tr}_2\hat{Z}$ denote the partial traces of $\hat{Z}$ over the first and second components of $\mathbb{C}^N \otimes \mathbb{C}^N$, respectively. See figure 5(d). Hence, the single-gate cost-function variance $\mathrm{Var}_{\hat{U}} E = \mathrm{Avg}_{\hat{U}} E^2 - (\mathrm{Avg}_{\hat{U}} E)^2$ over $\mathrm{U}(N)$ is

$$\mathrm{Var}_{\hat{U}} E \overset{(B1),(B3)}{=} \frac{1}{N^2-1}\left(\frac{(\mathrm{Tr}\,\hat{Z})^2}{N^2} - \frac{\mathrm{Tr}\left((\mathrm{Tr}_1\hat{Z})^2 + (\mathrm{Tr}_2\hat{Z})^2\right)}{N} + \mathrm{Tr}\,\hat{Z}^2\right). \tag{B4}$$

As discussed in [17] and shown diagrammatically in figure 6, the single-gate gradient variance (9) valuates to

$$\mathrm{Var}_{\hat{U}} \hat{g} \overset{(4),(7)}{=} \frac{2}{N^2-1}\left(\frac{(\mathrm{Tr}\,\hat{Z})^2}{N^2} - \frac{\mathrm{Tr}\left((\mathrm{Tr}_1\hat{Z})^2 + (\mathrm{Tr}_2\hat{Z})^2\right)}{N} + \mathrm{Tr}\,\hat{Z}^2\right). \tag{B5}$$

So, the cost-function variance (B4) is *exactly half* of the Riemannian gradient variance (B5). □

Appendix C. Proof of theorem 2

Proof. **(a)** Let us first consider a circuit with only two variable unitaries $\hat{U}_1$ and $\hat{U}_2$. In this case, the cost function (2) can be written in the form

$$E = E\left(\hat{U}_1, \hat{U}_2\right) = \sum_a f_a\left(\hat{U}_1\right) g_a\left(\hat{U}_2\right), \tag{C1}$$

where f_a and g_a are continuous functions which only depend on $\hat{U}_1$ and $\hat{U}_2$, respectively: Analogously to figure 2(c), we can always bipartition the tensor network for $E(\hat{U}_1, \hat{U}_2)$ into two parts f and g with f containing $\hat{U}_1, \hat{U}_1^\dagger$ and g containing $\hat{U}_2, \hat{U}_2^\dagger$. The contraction of the two parts (operator products and trace to obtain the scalar E) then corresponds to the sum over a in equation (C1).

The single-gate cost variance for $\hat{U}_1$ at fixed $\hat{U}_2$ then is ($\mathrm{Avg}_i \equiv \mathrm{Avg}_{\hat{U}_i}$ and $\mathrm{Var}_i \equiv \mathrm{Var}_{\hat{U}_i}$)

$$\begin{aligned}\mathrm{Var}_1 E\left(\hat{U}_1, \hat{U}_2\right) &= \mathrm{Avg}_1\left(E^2\right) - \left(\mathrm{Avg}_1 E\right)^2 \\ &= \sum_{a,b} \underbrace{\left[\mathrm{Avg}_1(f_a f_b) - \mathrm{Avg}_1(f_a)\,\mathrm{Avg}_1(f_b)\right]}_{\equiv \mathrm{Cov}(f_a, f_b)} g_a g_b\end{aligned} \tag{C2}$$

and, similarly $\mathrm{Var}_2 E = \sum_{a,b} f_a f_b\,\mathrm{Cov}(g_a, g_b)$.

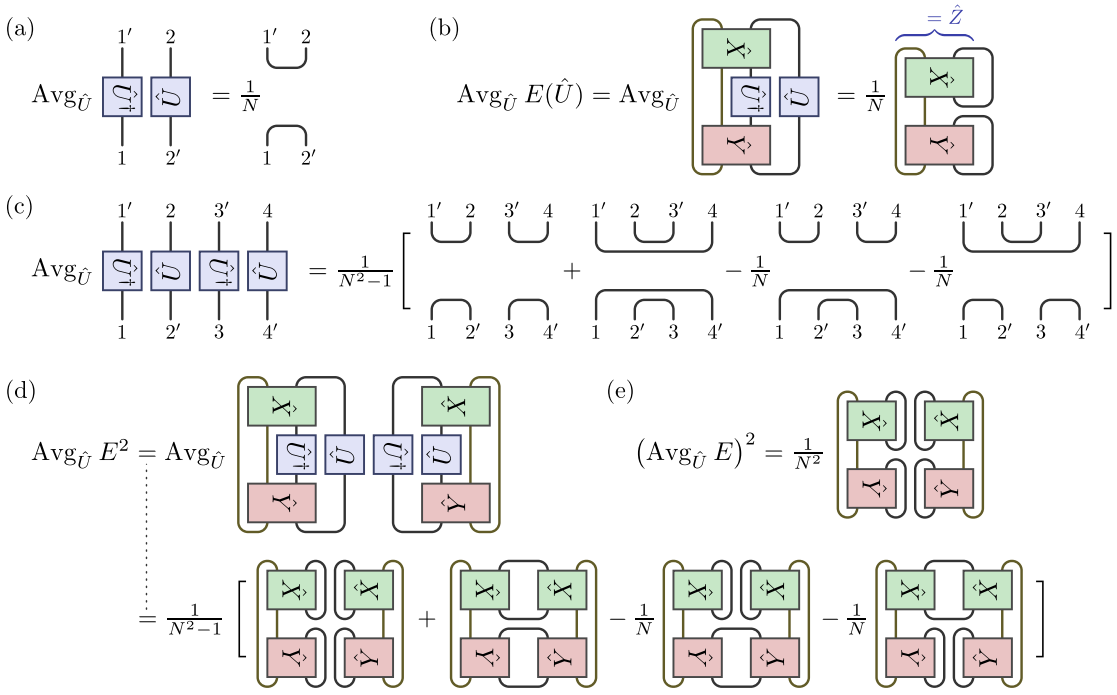

Figure 5. (a) Diagrammatic representation for the first-moment Haar-measure integral (7*a*) over the unitary group U(N). (b) The single-gate Haar average (B1) of the cost function (3). (c) Second-moment Haar-measure integral (7*b*) over the unitary group U(N). (d) Average (B3) of the squared cost function. (e) The single-gate Haar variance (B4) is obtained by subtracting the squared cost-function average.

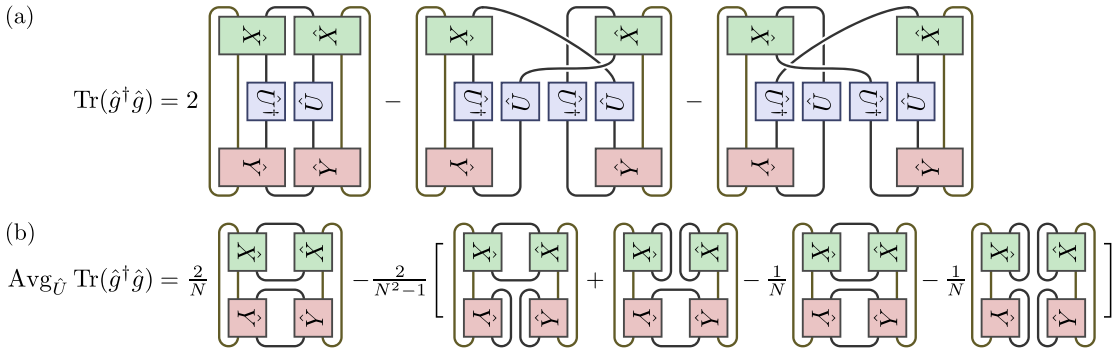

Figure 6. (a) Diagrammatic representation of the central quantity $\mathrm{Tr}(\hat{g}^\dagger\hat{g})$ in the definition (9) of the Riemannian gradient variance. (b) The single-gate variance (B5) of the Riemannian gradient, obtained by applying the second-moment Weingarten formula (7*b*) as illustrated in figure 5(c).

Using that $\mathrm{Avg}_{1,2}(f_af_bg_ag_b) = \mathrm{Avg}_1(f_af_b)\,\mathrm{Avg}_2(g_ag_b)$ due to the independence of $\hat{U}_1$ and $\hat{U}_2$ in the Haar-measure average, the global cost-function variance is ($\mathrm{Avg}_{1,2} \equiv \mathrm{Avg}_{\hat{U}_1,\hat{U}_2} \equiv \mathrm{Avg}_{\hat{U}_1}\mathrm{Avg}_{\hat{U}_2}$)

$$\begin{aligned}
\mathrm{Var}_{1,2}\,E &= \mathrm{Avg}_{1,2}\left(E^2\right) - \left(\mathrm{Avg}_{1,2}\,E\right)^2 \\
&= \sum_{a,b}\left[\mathrm{Avg}_1\left(f_af_b\right)\mathrm{Avg}_2\left(g_ag_b\right) - \mathrm{Avg}_1\left(f_a\right)\mathrm{Avg}_1\left(f_b\right)\mathrm{Avg}_2\left(g_a\right)\mathrm{Avg}_2\left(g_b\right)\right] \\
&= \mathrm{Avg}_2\,\mathrm{Var}_1\,E + \mathrm{Avg}_1\,\mathrm{Var}_2\,E - \sum_{a,b}\mathrm{Cov}\left(f_a,f_b\right)\mathrm{Cov}\left(g_a,g_b\right) \\
&\leqslant \mathrm{Avg}_2\,\mathrm{Var}_1\,E + \mathrm{Avg}_1\,\mathrm{Var}_2\,E. \qquad \text{(C3)}
\end{aligned}$$

This is the right inequality in equation (15*a*) for the case of two variable unitaries. In the last step, we have used that the covariance matrices

$$\mathrm{Cov}\left(f_a,f_b\right) = \mathrm{Avg}\left(f_af_b\right) - \mathrm{Avg}\left(f_a\right)\mathrm{Avg}\left(f_b\right) = \mathrm{Avg}\left(\left[f_a - \mathrm{Avg}f_a\right]\left[f_b - \mathrm{Avg}f_b\right]\right) \qquad \text{(C4)}$$

and $\mathrm{Cov}(g_a, g_b)$ are positive semidefinite such that the trace $\sum_{a,b}\mathrm{Cov}(f_a, f_b)\,\mathrm{Cov}(g_a, g_b)$ of their product is non-negative.

(b) The generalization to a circuit with K variable unitaries follows by iterating equation (C3). Decomposing the tensor network as before into K parts, each containing only one of the variable unitaries and its adjoint, we can write the cost function in the form

$$E = E\left(\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_K\right) = \sum_a f_a^{(1)}\left(\hat{U}_1\right) f_a^{(2)}\left(\hat{U}_2\right)\cdots f_a^{(K)}\left(\hat{U}_K\right). \tag{C5}$$

Now, iterating equation (C3), we find

$$\begin{aligned}\mathrm{Var}_{1,2,\ldots,K} E &\leqslant \mathrm{Avg}_{2,\ldots,K}\mathrm{Var}_1 E + \mathrm{Avg}_1\mathrm{Var}_{2,\ldots,K} E\\ &\leqslant \mathrm{Avg}_{2,\ldots,K}\mathrm{Var}_1 E + \mathrm{Avg}_1\left(\mathrm{Avg}_{3,\ldots,K}\mathrm{Var}_2 E + \mathrm{Avg}_2\mathrm{Var}_{3,\ldots,K} E\right)\\ &\leqslant \cdots \leqslant \sum_i \mathrm{Avg}_{\{j\neq i\}}\mathrm{Var}_i E \equiv \sum_i V_i,\end{aligned}$$

where $\mathrm{Avg}_{i_1,\ldots,i_n} h \equiv \mathrm{Avg}_{i_1}\ldots\mathrm{Avg}_{i_n} h$ and $\mathrm{Var}_{i_1,\ldots,i_n} h \equiv \mathrm{Avg}_{i_1,\ldots,i_n} h^2 - (\mathrm{Avg}_{i_1,\ldots,i_n} h)^2$.

(c) The right inequality in equation (15*a*) follows by applying the law of total variance,

$$\mathrm{Var}_{1,2,\ldots,K} E = \mathrm{Avg}_{\{j\neq i\}}\mathrm{Var}_i E + \mathrm{Var}_{\{j\neq i\}}\mathrm{Avg}_i E \geqslant \mathrm{Avg}_{\{j\neq i\}}\mathrm{Var}_i E \overset{(11)}{\equiv} V_i, \tag{C6}$$

which holds for all i. Recall that, given two random variables E and U_i on the same probably space $(\mathcal{M})$, the law of total variance states that $\mathrm{Var}(E) = \mathrm{Avg}\left(\mathrm{Var}(E|U_i)\right) + \mathrm{Var}\left(\mathrm{Avg}(E|U_i)\right)$. This corresponds to the first step in equation (C6). In the second, step, we have used the nonegativity of the variance. □

ORCID iDs

Qiang Miao https://orcid.org/0000-0002-3480-9284
Thomas Barthel https://orcid.org/0000-0001-5351-4011

References

[1] Cerezo M *et al* 2021 Variational quantum algorithms *Nat. Rev. Phys.* **3** 625
[2] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 Barren plateaus in quantum neural network training landscapes *Nat. Commun.* **9** 4812
[3] Cerezo M, Sone A, Volkoff T, Cincio L and Coles P J 2021 Cost function dependent barren plateaus in shallow parametrized quantum circuits *Nat. Commun.* **12** 1791
[4] Grant E, Wossnig L, Ostaszewski M and Benedetti M 2019 An initialization strategy for addressing barren plateaus in parametrized quantum circuits *Quantum* **3** 214
[5] Zhang K, Hsieh M-H, Liu L and Tao D 2022 Escaping from the barren plateau via Gaussian initializations in deep variational quantum circuits (arXiv:2203.09376)
[6] Mele A A, Mbeng G B, Santoro G E, Collura M and Torta P 2022 Avoiding barren plateaus via transferability of smooth solutions in Hamiltonian Variational Ansatz (arXiv:2206.01982)
[7] Kulshrestha A and Safro I 2022 BEINIT: avoiding barren plateaus in variational quantum algorithms (arXiv:2204.13751)
[8] Dborin J, Barratt F, Wimalaweera V, Wright L and Green A G 2022 Matrix product state pre-training for quantum machine learning *Quantum Sci. Technol.* **7** 035014
[9] Skolik A, McClean J R, Mohseni M, van der Smagt P and Leib M 2021 Layerwise learning for quantum neural networks *Quantum Mach. Intell.* **3** 5
[10] Slattery L, Villalonga B and Clark B K 2022 Unitary block optimization for variational quantum algorithms *Phys. Rev. Res.* **4** 023072
[11] Haug T and Kim M 2021 Optimal training of variational quantum algorithms without barren plateaus (arXiv:2104.14543)
[12] Sack S H, Medina R A, Michailidis A A, Kueng R and Serbyn M 2022 Avoiding barren plateaus using classical shadows *PRX Quantum* **3** 020365
[13] Rad A, Seif A and Linke N M 2022 Surviving the barren plateau in variational quantum circuits with Bayesian learning initialization (arXiv:2203.02464)
[14] Tao Z, Wu J, Xia Q and Li Q 2022 LAWS: look around and warm-start natural gradient descent for quantum neural networks (arXiv:2205.02666)
[15] Wang Y, Qi B, Ferrie C and Dong D 2023 Trainability enhancement of parameterized quantum circuits via reduced-domain parameter initialization (arXiv:2302.06858)
[16] Miao Q and Barthel T 2024 Isometric tensor network optimization for extensive Hamiltonians is free of barren plateaus *Phys. Rev. A* **109** L050402
[17] Barthel T and Miao Q 2023 Absence of barren plateaus and scaling of gradients in the energy optimization of isometric tensor network states (arXiv:2304.00161)
[18] Zhang H-K, Liu S and Zhang S-X 2024 Absence of barren plateaus in finite local-depth circuits with long-range entanglement *Phys. Rev. Lett.* **132** 150603
[19] Cerezo M et al 2023 Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing (arXiv:2312.09121)

[20] Arrasmith A, Holmes Z, Cerezo M and Coles P J 2022 Equivalence of quantum barren plateaus to cost concentration and narrow gorges *Quantum Sci. Technol.* **7** 045015
[21] Smith S T 1994 Chap. Optimization techniques on Riemannian manifolds *Hamiltonian and Gradient Flows, Algorithms and Control* (*Fields Institute Communications)* vol 3 p 113
[22] Huang W, Gallivan K A and Absil P-A 2015 A Broyden class of quasi-Newton methods for Riemannian optimization *SIAM J. Optim.* **25** 1660
[23] Miao Q and Barthel T 2023 Quantum-classical eigensolver using multiscale entanglement renormalization *Phys. Rev. Res.* **5** 033141
[24] Wiersema R and Killoran N 2023 Optimizing quantum circuits with Riemannian gradient flow *Phys. Rev.* A **107** 062421
[25] Arrasmith A, Cerezo M, Czarnik P, Cincio L and Coles P J 2021 Effect of barren plateaus on gradient-free optimization *Quantum* **5** 558
[26] Vidal G 2007 Entanglement renormalization *Phys. Rev. Lett.* **99** 220405
[27] Vidal G 2008 Class of quantum many-body states that can be efficiently simulated *Phys. Rev. Lett.* **101** 110501
[28] Weingarten D 1978 Asymptotic behavior of group integrals in the limit of infinite rank *J. Math. Phys.* **19** 999
[29] Collins B and Śniady P 2006 Integration with respect to the Haar measure on unitary, orthogonal and symplectic group *Commun. Math. Phys.* **264** 773
[30] We ignore the heterogeneity of $\mathrm{Avg}_{\hat{U}}(\alpha_k^2)$ for different $k = 1, \ldots, N^2$ of a single gate, because the gate Hilbert-space dimension N is usually system-size independent
[31] Fannes M, Nachtergaele B and Werner R F 1992 Ground states of VBS models on cayley trees *J. Stat. Phys.* **66** 939
[32] Otsuka H 1996 Density-matrix renormalization-group study of the spin-1/2XXZ antiferromagnet on the Bethe lattice *Phys. Rev.* B **53** 14004
[33] Shi Y-Y, Duan L-M and Vidal G 2006 Classical simulation of quantum many-body systems with a tree tensor network *Phys. Rev.* A **74** 022320
[34] Fannes M, Nachtergaele B and Werner R F 1992 Finitely correlated states on quantum spin chains *Commun. Math. Phys.* **144** 443
[35] Schollwöck U 2011 The density-matrix renormalization group in the age of matrix product states *Ann. Phys., NY* **326** 96–192
[36] Thanasilp S, Wang S, Cerezo M and Holmes Z 2022 Exponential concentration and untrainability in quantum kernel methods (arXiv:2208.11060)
[37] Rudolph M S, Lerch S, Thanasilp S, Kiss O, Vallecorsa S, Grossi M and Holmes Z 2023 Trainability barriers and opportunities in quantum generative modeling (arXiv:2305.02881)
[38] Ragone M, Bakalov B N, Sauvage F, Kemper A F, Marrero C O, Larocca M and Cerezo M 2023 A unified theory of barren plateaus for deep parametrized quantum circuits (arXiv:2309.09342)
[39] Diaz N L, García-Martín D, Kazi S, Larocca M, and Cerezo M 2023 Showcasing a barren plateau theory beyond the dynamical Lie algebra (arXiv:2310.11505)
[40] Xiong W, Facelli G, Sahebi M, Agnel O, Chotibut T, Thanasilp S and Holmes Z 2023 On fundamental aspects of quantum extreme learning machines (arXiv:2312.15124)
[41] Miao Q and Barthel T 2023 Convergence and quantum advantage of Trotterized MERA for strongly-correlated systems (arXiv:2303.08910)