Diagnosing barren plateaus with tools from quantum optimal control

Martín Larocca[1,2], Piotr Czarnik[2], Kunal Sharma[3,2], Gopikrishnan Muraleedharan[2], Patrick J. Coles[2], and M. Cerezo[4,5]

[1]Departamento de Física "J. J. Giambiagi" and IFIBA, FCEyN, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina

[2]Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[3]Hearne Institute for Theoretical Physics and Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA USA

[4]Information Sciences, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

[5]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

arXiv:2105.14377v3 [quant-ph] 27 Sep 2022

Variational Quantum Algorithms (VQAs) have received considerable attention due to their potential for achieving near-term quantum advantage. However, more work is needed to understand their scalability. One known scaling result for VQAs is barren plateaus, where certain circumstances lead to exponentially vanishing gradients. It is common folklore that problem-inspired ansatzes avoid barren plateaus, but in fact, very little is known about their gradient scaling. In this work we employ tools from quantum optimal control to develop a framework that can diagnose the presence or absence of barren plateaus for problem-inspired ansatzes. Such ansatzes include the Quantum Alternating Operator Ansatz (QAOA), the Hamiltonian Variational Ansatz (HVA), and others. With our framework, we prove that avoiding barren plateaus for these ansatzes is not always guaranteed. Specifically, we show that the gradient scaling of the VQA depends on the degree of controllability of the system, and hence can be diagnosed through the dynamical Lie algebra $\mathfrak{g}$ obtained from the generators of the ansatz. We analyze the existence of barren plateaus in QAOA and HVA ansatzes, and we highlight the role of the input state, as different initial states can lead to the presence or absence of barren plateaus. Taken together, our results provide a framework for trainability-aware ansatz design strategies that do not come at the cost of extra quantum resources. Moreover, we prove no-go results for obtaining ground states with variational ansatzes for controllable system such as spin glasses. Our work establishes a link between the existence of barren plateaus and the scaling of the dimension of $\mathfrak{g}$.

1 INTRODUCTION

Quantum computers hold the promise to achieve computational speed-ups over classical supercomputers for certain tasks [1, 2, 3, 4]. However, despite recent tremendous progress in quantum technologies, present-day quantum devices (known as Noisy Intermediate-Scale Quantum (NISQ) devices) are constrained by the limited number of qubits, connectivity, and by the presence of quantum noise [5]. Hence, it becomes crucial to determine what are the capabilities and limitations of NISQ computers to achieving a quantum advantage.

One of the most promising computational models for making use of near-term quantum computers are Variational Quantum Algorithms (VQAs) [6]. Here, a task of interest is encoded into a parametrized cost function $C(\boldsymbol{\theta})$ that is efficiently computable on a noisy quantum computer. Part of the computational complexity is pushed onto classical computers by leveraging the power of classical optimizer that train the parameters $\boldsymbol{\theta}$ and minimize the cost. VQAs have been proposed for tasks such as solving linear systems of equations [7, 8, 9] or performing dynamical quantum simulations [10, 11, 12, 13, 14, 15, 16, 17], as well as for many others relevant applications [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30].

Despite the wide application range of VQAs, their widespread use is still limited by several challenges that can hinder their success. For instance, it has been shown that the optimization task associated with minimizing the VQA cost function is in general an NP-hard non-convex optimization problem [31]. Moreover, despite the typical difficulties encountered in classical non-convex optimization tasks, there are new challenges that arise when training the parameters of VQAs such as hardware noise, or the limited precision arising from a limited number of shots. These difficulties have then led to several quantum-aware optimizers being developed [32, 33, 34, 10, 35, 36].

In addition, certain VQAs have been shown to exhibit the so-called barren plateau phenomenon, where

the cost function becomes untrainable due to gradients that vanish, on average, exponentially with the system size [37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48]. Thus, barren plateaus have then been recognized as one of the main limitations to overcome in order to preserve the hope of achieving quantum advantage with VQAs. Recently, a great deal of effort has been put forward to developing methods that can mitigate the effect of barren plateaus [49, 50, 51, 52, 53, 54, 55], but ideally one would like to devise and employ VQA ansatzes which do not exhibit barren plateaus altogether.

For instance, it is known that one should avoid problem-agnostic ansatzes such as deep hardware efficient ansatzes, as these can exhibit barren plateaus due to their high expressibility [37, 38, 47]. Hence, so-called problem-inspired ansatzes have been speculated to be able to overcome barren plateaus by encoding information about the problem at hand in the ansatz. Here, the intuition is that problem-inspired ansatzes constrain the space explored during the optimization to a space that either contains the solution to the problem, or that at least contains a good approximation to the solution, while maintaining a low expressibility.

In this work we employ tools from Quantum Optimal Control (QOC) to diagnose the presence or absence of barren plateaus in certain families of problem-inspired ansatzes with a periodic structure. QOC theory is a long standing theoretical framework developed to provide tools for the manipulation of quantum dynamical processes. As shown in Fig. 1, we here make use of the fact that periodic VQAs and QOC systems can be considered as different level formulations of a common variational problem [56], as both aim at driving a quantum system with a classical optimization loop. Most importantly, this connection allows us to understand and forecast the presence or absence of barren plateaus in problem-inspired variational ansatzes like the Quantum Alternating Operator Ansatz (QAOA) [19, 57] and the Hamiltonian Variational Ansatz (HVA) [58, 59]. We note that the procedure is perfectly suitable for other periodic ansatzes like the adaptive QAOA [57, 60] and the quantum optimal control ansatz [61]. Moreover, our results also extend to quantum neural network architectures used in the quantum machine learning literature [62]. Our results indicate that problem-inspired ansatzes are not immune to barren plateaus, and hence that certain ansatz strategies in the literature need to be revised.

Our main results are organized into propositions and theorems that show that one can diagnose the existence of barren plateaus by analyzing the controllability of the system, i.e., by studying the Dynamical Lie Algebra (DLA) of the system. The DLA is the subspace of operator space spanned by the nested commutators of the elements in the set of generators of the ansatz (e.g., see [63] for an introduction to quantum control theory). In an effort to give a comprehensive picture, our results follow the different controllability scenarios shown in Fig. 2.

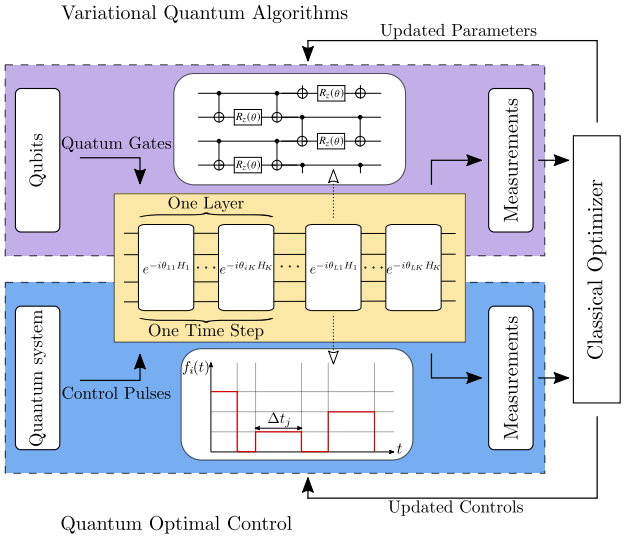

Figure 1: **Framework for Variational Quantum Algorithms (VQA) and Quantum Optimal Control (QOC).** VQAs and QOCs can be regarded as two different levels of a theory that manipulate the evolution of a quantum system by training sets of parameters governing the system's dynamical evolution [56]. In VQAs (QOC) one applies a series of parametrized quantum gates (control pulses) to an input state. By gathering knowledge on the evolution via measurements on the resulting evolved state, the set of parameters (controls) are trained using a classical optimizer until a given task is completed. In this work we consider VQAs and QOC systems that have periodic structure ansatzes as in Eq. (2).

The manuscript is organized as follows. In Section 2 we present the theoretical framework for VQAs, which includes a description of the type of ansatz considered, as well as a basic review of concepts related to barren plateaus and ansatz expressibility. Then, in Section 3 we introduce the framework of QOC, and recall how in QOC theory the DLA of the ansatz generators is used to study the controllability of the system. Section 4 contains the main results of this work, while in Section 5 we present our numerical simulations. Finally, in Section 6 we present our discussions and conclusions.

2 VARIATIONAL QUANTUM ALGORITHMS

In this section we review the basic framework of Variational Quantum Algorithms (VQAs). In particular, we discuss a general form for ansatzes that have a periodic structure, which we consider throughout this work. Since our goal is to analyze the gradient scaling, we additionally provide an overview of the barren plateau phenomenon.

2.1 General framework

We consider an optimization task where the goal is to minimize a cost function of the form

$$C(\boldsymbol{\theta}) = \mathrm{Tr}[OU(\boldsymbol{\theta})\rho U^{\dagger}(\boldsymbol{\theta})]\,. \tag{1}$$

Here, ρ is an input state on n qubits in a d-dimensional Hilbert space with $d = 2^n$, $U(\boldsymbol{\theta})$ a parametrized quantum circuit, and O is a Hermitian operator that defines the task at hand.

Throughout this work we consider layered parametrized quantum circuits that, as shown in Fig. 1, have a periodic structure of the form

$$U(\boldsymbol{\theta}) = \prod_{l=1}^{L} U_l(\boldsymbol{\theta}_l), \quad U_l(\boldsymbol{\theta}_l) = \prod_{k=0}^{K} e^{-iH_k\theta_{lk}}\,. \tag{2}$$

Here, the index l indicates the layer, $\boldsymbol{\theta}_l = (\theta_{l1}, \ldots, \theta_{lK})$ contains the parameters of such layer (such that $\boldsymbol{\theta} = \{\boldsymbol{\theta}_l\}_{l=1}^{L}$), and H_k are Hermitian traceless operators that generate the unitaries in the ansatz. For generality, one can also allow for certain layers to be unparametrized, in which case one would simply set certain θ_{lk} to be constant. In what follows we refer to this type of ansatz as a Periodic Structure Ansatz (PSA). We refer the reader to Appendix B for a detailed discussion of several ansatzes from the literature that are PSAs.

2.2 Barren plateaus

Recently, it has been shown that the choice of ansatz can hinder the trainability of the parameters $\boldsymbol{\theta}$ for large problem sizes due to the existence of the so-called barren plateau phenomenon. In this context, deep unstructured problem-agnostic ansatz are known to exhibit barren plateaus [37, 49, 47]. Hence, the design of ansatzes that overcome barren plateaus has been recognized as one of the most important challenges to guarantee the success of VQAs [6], and problem-inspired ansatzes have been proposed as one of the most promising strategies. However, despite their promise, little is known about the existence of barren plateaus in problem-inspired ansatzes.

Let us now briefly recall that when the cost exhibits a barren plateau, the gradients are exponentially suppressed (on average) across the optimization landscape. This implies that an exponentially large precision is needed to navigate trough the flat landscape and determine a cost minimizing direction [37, 49, 42]. Hence, consider the following definition.

Definition 1 (Barren Plateau)**.** *A cost function $C(\boldsymbol{\theta})$ as in Eq. (1) is said to have a barren plateau when training $\theta_\mu \equiv \theta_{pq} \in \boldsymbol{\theta}$, if the cost function partial derivative $\partial C(\boldsymbol{\theta})/\partial\theta_\mu \equiv \partial_\mu C(\boldsymbol{\theta})$ is such that*

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] \leqslant F(n)\,, \quad \textit{with} \quad F(n) \in \mathcal{O}\left(\frac{1}{b^n}\right)\,, \tag{3}$$

for some $b > 1$. Here the variance is taken with respect to the set of parameters $\boldsymbol{\theta}$.

We refer the reader to Appendix C for additional details on barren plateaus.

It is worth remarking that the barren plateau phenomenon has been linked to the expressibility of the ansatz, as it has been shown that circuits with large expressibility will exhibit small gradients [47]. In this context, one can quantify the expressibility of an ansatz by comparing the distribution of unitaries obtained from $U(\boldsymbol{\theta})$ to the maximally expressive uniform (Haar) distribution U_H [64]. Defining the t-th moment superoperator of the distribution generated by the ansatz $U(\boldsymbol{\theta})$,

$$M^{(t)}_{U(\boldsymbol{\theta})} = \int_{\boldsymbol{\theta}} dU(\boldsymbol{\theta})\, U(\boldsymbol{\theta})^{\otimes t} \otimes (U^*(\boldsymbol{\theta}))^{\otimes t}\,, \tag{4}$$

we recall that its ordinary action on a given operator can be obtained by placing said operator into the center of the representation of $M^{(t)}_{U(\boldsymbol{\theta})}$ as [65, 66]

$$M^{(t)}_{U(\boldsymbol{\theta})}(\cdot) = \int_{\boldsymbol{\theta}} dU(\boldsymbol{\theta})\, U(\boldsymbol{\theta})^{\otimes t}(\cdot)(U(\boldsymbol{\theta})^{\dagger})^{\otimes t}\,. \tag{5}$$

In our case, we will only be interested in second moments. For that reason, we will focus on the deviation of the second moments of the distribution generated by the ansatz $M^{(2)}_{U(\boldsymbol{\theta})}$ from $M^{(2)}_{U_H}$ the second moments of the Haar distribution, via the norm of the superoperator

$$\mathcal{A}^{(t)}_{U(\boldsymbol{\theta})} = M^{(t)}_{U_H} - M^{(t)}_{U(\boldsymbol{\theta})}\,. \tag{6}$$

For our purposes here, we find it convenient to define the expressibility as the infinity norm, $\|A\|_\infty = \lambda_{\max}(A)$, with $\lambda_{\max}(A)$ the largest singular value of A [1]. Thus, the more expressible the ansatz, the smaller the norm $\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty$, and the smaller the gradients of the cost partial derivatives [47]. The limit $\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty = 0$ is reached when $U(\boldsymbol{\theta})$ forms a 2-design, in which case the cost exhibits a barren plateau according to Definition 1 [37, 49].

3 QUANTUM OPTIMAL CONTROL

Quantum Optimal Control (QOC) is a theoretical framework that provides tools for the systematic manipulation of quantum dynamical systems. The connection between VQAs and QOC has been previously established showing that one can use QOC tools to

[1] The expressibility can also be defined in terms of other matrix norms such as the diamond norm or the Schatten p-norms. However, due to the matrix norm equivalence, if $\|\mathcal{A}^{(t)}_{U(\boldsymbol{\theta})}\| = \epsilon$ for our definition, there always exists an ϵ' for other expressibility definitions such that the expressibility is equal to ϵ' and such that ϵ and ϵ' are related via a dimensionallity factor [67, 68].

specify the parameters $\boldsymbol{\theta}$ at a device-level [69, 56, 70] and to analyze VQA landscapes [71]. Conversely, tools from VQAs have been employed to determine optimal control sequences [72]. In particular, Ref. [56] notes that VQAs and QOC can be unified as formulations of variational optimization at the circuit level and pulse level, respectively. In addition, the framework of QOC has been employed to analyze the computational universality of quantum circuits [73, 74, 75], as well as their reachability [76].

In QOC one is interested in controlling the dynamical evolution of a quantum state $|\psi\rangle$ in a complex d-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^d$ (where $d = 2^n$) [63]. In the typical setting, one has a Hamiltonian

$$H(\{f_k(t)\}) = H_0 + \sum_{k=1}^{K} f_k(t) H_k \tag{7}$$

that is tunable through some time-dependent functions $\{f_k(t)\}$, know as control fields or protocols. The fixed Hamiltonian H_0, usually called the drift, represents the natural or free evolution of the system, whereas the control Hamiltonians $\{H_i\}$ are associated with interactions with external degrees of freedom (usually electromagnetic radiation). Thus, $|\psi\rangle$ evolves through the parametrized propagator $U(t)$ as $|\psi\rangle(t) = |\psi(t)\rangle$. In turn, $U(t)$ is the solution to the Schrödinger equation

$$\frac{dU(t)}{dt} = -iH(\{f_k(t)\})U(t), \quad \text{with} \quad U(0) = \mathbb{1}\,. \tag{8}$$

As shown in Appendix D, under standard assumptions, the Trotrerized QOC propagator of Eq. (8) is a PSA as in Eq. (2).

The variety of different dynamics a quantum control system in the form of Eq. (7) can undergo, upon variation of the control fields, is well understood though group theory. Since the Hamiltonian is Hermitian and traceless, $U(\boldsymbol{\theta})$ belongs to $\mathcal{SU}(d)$, the Lie group of unitary $d \times d$ complex matrices that preserves the standard inner product on $\mathcal{H}$. Surprisingly, the set of all unitaries $U(\boldsymbol{\theta})$ that can be accessed by such a control system forms itself a Lie group, known as the *dynamical Lie group* $\mathbb{G} \subseteq \mathcal{U}(d)$. Hence, a natural question which arises is: how can this group be determined?

First, let us define the set of generators.

Definition 2 (Set of generators)**.** *Given a parametrized quantum circuit of the form in Eq. (2) we define the set of generators $\mathcal{G} = \{H_k\}_{k=0}^{K}$ as the set (of size $|\mathcal{G}| = K + 1$) of the (traceless) Hermitian operators that generate the unitaries in a single layer of $U(\boldsymbol{\theta})$.*

Naturally, the group $\mathbb{G}$ depends on the set of generators $\mathcal{G}$, yet it is not sufficient to look at the individual elements of $\mathcal{G}$. Instead, one must consider the Lie algebra that emerges from their nested commutators. Hence, consider the following definition [77].

Definition 3 (Dynamical Lie Algebra)**.** *Given a control system with generators $\mathcal{G}$ (see Definition 2), the Dynamical Lie Algebra (DLA) $\mathfrak{g}$ is the subalgebra of $\mathfrak{su}(d)$ spanned by the repeated nested commutators of the elements in $\mathcal{G}$, i.e.,*

$$\mathfrak{g} = \operatorname{span} \langle iH_0, \ldots, iH_K \rangle_{\text{Lie}} \subseteq \mathfrak{su}(d), \tag{9}$$

where $\langle S \rangle_{\text{Lie}}$ denotes the Lie closure, i.e., the set obtained by repeatedly taking the nested commutators between the elements in S.

Here, $\mathfrak{su}(d)$ is the special unitary algebra of degree d, the Lie algebra formed by the set of $d \times d$ skew-Hermitian, traceless matrices. In Appendix E, we lay down the basic procedure to build DLAs (see Algorithm 1) and provide some discussion on the complexity of such construction.

Once the DLA is obtained from the set of generators, one can determine the set of unitaries that are expressible by the control system. Specifically, one can now properly define the dynamical Lie group as follows.

Definition 4 (dynamical Lie group)**.** *The set unitaries $\mathbb{G}$ that can be generated by a control system is determined by its DLA (see Definition 3) through* [2]

$$\mathbb{G} = e^{\mathfrak{g}} := \{e^V, \quad V \in \mathfrak{g}\}\,. \tag{10}$$

The dynamical Lie group in turn determines the set of states $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi\rangle$ that can be reached by evolving an initial state $|\psi\rangle$. Specifically, here $U(\boldsymbol{\theta})$ can attain values in the Lie group $\mathbb{G}$. In addition, Definition 4 crucially shows that one can study the expressibility of a control system (i.e., the unitaries that can be generated, or the set of states that can be reached) via the DLA obtained from the set of generators. As shown in Fig. 2, when computing $\mathfrak{g}$ there are several cases of interest that can arise and which we here consider. For the sake of clarity, in what follows we briefly recall several key concepts that will be useful throughout the manuscript. We refer the reader to [63] for additional details.

First and foremost, we recall the concept of *controllability*. A control system is said to be *controllable* if its DLA is full rank, i.e., $\mathfrak{g} = \mathfrak{su}(d)$. This implies that $\mathbb{G} = \mathcal{SU}(d)$ and hence every unitary (up to a phase) can be obtained by appropriately choosing control parameters in Eq. (45). In particular, this means that for any two states $|\psi\rangle$ and $|\phi\rangle$, there always exists a unitary $U(\boldsymbol{\theta}) \in \mathbb{G}$ such that $U(\boldsymbol{\theta})|\psi\rangle = |\phi\rangle$.

If the DLA is not full rank, then the system is said to be *uncontrollable*. In this case $\mathfrak{g}$ is a *proper* subalgebra of $\mathfrak{su}(d)$, and only a *proper* subgroup $\mathbb{G} \subset \mathcal{SU}(d)$

[2] This is grounded in the fact that every possible DLA is a subalgebra of the special unitary algebra $\mathfrak{su}(d)$ and therefore compact. The exponential map is a function $e : \mathfrak{g} \to \mathbb{G}$. If $\mathfrak{g}$ is compact then the exponential map is surjective: every element of $\mathbb{G}$ is the image of at least one element of $\mathfrak{g}$.

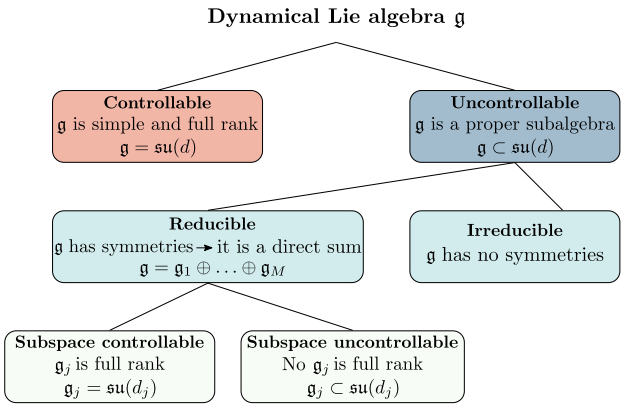

Figure 2: **Cases of interest for the Dynamical Lie Algebra.** The Dynamical Lie Algebra (DLA) $\mathfrak{g}$ determines the set of unitaries expressible, and concomitantly, the set of states reachable. In this figure we show different scenarios that can arise when computing $\mathfrak{g}$. Our main results (on gradient scaling of VQAs and QOC systems) pertain to these different scenarios.

is available to the control system, meaning that the set of reachable states $\{U(\boldsymbol{\theta})|\psi\rangle, \quad \forall U(\boldsymbol{\theta}) \in \mathbb{G}\}$ is not the whole state space. As depicted in Fig. 2, there are two sources of uncontrollability [78].

On one hand, if the generators in $\mathcal{G}$ share one or more common symmetries, i.e., there is at least one Hermitian operator Σ that commutes with every element in $\mathcal{G}$, then every $H \in \mathfrak{g}$ is block diagonal in the eigenbasis of Σ. This causes the state space to break into subspaces that are invariant under the action of $\mathfrak{g}$, in which case controllability is clearly disrupted. Here, the DLA is a *reducible* representation of some Lie algebra. On the other hand, even in the absence of symmetries, that is, when the DLA is *irreducible*, uncontrollability can arise simply because the Lie algebra is a proper subalgebra of $\mathfrak{su}(d)$.

Let us finally remark that, as shown in Fig. 2, even though a reducible system cannot be controllable on the entire Hilbert space it may still be controllable on some (or all) of the invariant subspaces. Given a DLA that is a direct sum of irreducible representations, i.e., $\mathfrak{g} = \bigoplus_j \mathfrak{g}_j$, then the Hilbert space can be expressed as $\mathcal{H} = \bigoplus_j \mathcal{H}_j$, with $\mathcal{H}_j$ being invariant under the action of $\mathfrak{g}$. A system is said to be *subspace controllable* on subspace $\mathcal{H}_j$ if $\mathfrak{g}_j$ is full rank, i.e., $\mathfrak{g}_j = \mathfrak{u}(d_j)$, where $d_j = \dim(\mathcal{H}_j))$ and $\mathfrak{u}(d_j)$ denotes the unitary algebra of degree d_j.

4 MAIN RESULTS

As previously discussed, VQAs and QOC can be considered as two formulations of a common variational optimization problem that optimizes parameters controlling the dynamical evolution of a quantum system. In this section we present our main results, where we basically leverage tools from QOC to analyze the trainability and the existence of barren plateaus in VQAs. Specifically, we organize our results in term of the different controllability settings shown in Fig. 2. In all cases, the proofs are presented in the Appendix. The main idea behind our results is that, given a PSA $U(\boldsymbol{\theta})$ as in Eq. (2), the study of the DLA of the ansatz can diagnose the presence (or absence) of barren plateaus in the VQA landscapes.

4.1 Controllable systems

First, let us consider controllable systems. It is well known that the distribution of unitaries generated by controllable systems converges to a 2-design in the long-time (i.e., for sufficiently deep circuits) [79]. However, the rate of convergence actually depends on the specific choice of generators. Hence, our first result analyzes the depth at which the expressibility $\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty$ of a controllable system is ε small.

Theorem 1. *Consider a controllable system. Then, the PSA $U(\boldsymbol{\theta})$ will form an ε-approximate 2-design, i.e. $\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty = \varepsilon$ with $\epsilon > 0$, when the number of layers L in the circuit is*

$$L = \frac{\log(1/\varepsilon)}{\log\left(1/\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty\right)}. \tag{11}$$

Here $\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty$ denotes the expressibility of a single layer $U_1(\boldsymbol{\theta}_1)$ of the ansatz according to Eqs. (2) and (6).

See Appendix F for a proof of Theorem 1.

We note that Theorem 1 arises from the following expression that connects the expressibility of an L-layered PSA to the expressibility of a single layer of the ansatz to the L-th power as

$$\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty = \left(\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty\right)^L. \tag{12}$$

Here we can see that $\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty = 0$ if and only if $\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty = 0$. Hence, as expected, PSAs that have more expressible layers require less depth to have an ε-expressibility (to be ε-approximate 2-designs). Conversely, one can also see that ansatzes with less expressible layers require more depth to become ε-approximate 2-design.

The following corollary analyzes the scaling of L.

Corollary 1. *Let the single layer expressibility of a controllable system be $\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty = 1 - \delta(n)$, with $\delta(n)$ being at most polynomially vanishing with n, i.e., with $\delta(n) \in \Omega(1/\text{poly}(n))$. Then, if $L(n) \in \Omega(n/\delta(n))$, $U(\boldsymbol{\theta})$ will be no worse than an $\varepsilon(n)$-approximate 2-design (i.e., $\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty \leqslant \varepsilon(n)$) with $\varepsilon(n) \in \mathcal{O}(1/2^n)$, where we have added the n-dependence in L and ε for clarity.*

See Appendix G for a proof of Corollary 1.

From Corollary 1 we have that when the single layer expressibility is (at most) polynomially vanishing with n, then a polynomial number of layers suffice to make the PSA $U(\boldsymbol{\theta})$ exponentially close to being a 2-design. We note, however, that in the case where the single layer expressibility is exponentially close to 1, one requires an exponential number of layers to form an ε-approximate 2-design. In all the aforementioned cases it is worth remarking that an exponential number of layers will always lead to ε-approximate 2-designs with $\varepsilon \in \mathcal{O}(1/2^n)$, independently of the value of $\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty$.

Once the depth of the ansatz is sufficient for the controllable system to be an ε-approximate 2-design, then a barren plateau will arise. Hence, one can prove the following proposition from Theorem 1 and Corollary 1.

Proposition 1 (Controllable)**.** *There exists a scaling of the depth for which controllable systems form ε-approximate 2-designs with $\varepsilon \in \mathcal{O}(1/2^n)$, and hence the system exhibits a barren plateau according to Definition 1.*

See Appendix H for a proof of Proposition 1.

Proposition 1 rephrases the well known barren plateau results of [37, 47] in terms of controllability. Specifically, it has been shown that when an ansatz forms a 2-design, such randomness leads to a barren plateau. Hence, the proof of Proposition 1 simply follows the proof in [37], with the addition that the convergence to a 2-design comes from the fact that the system is controllable.

Evidently, it becomes relevant to determine systems that are controllable as these can exhibit barren plateaus. In this work we prove that two relevant sets of generators lead to full rank DLAs, and hence to controllable systems.

Proposition 2. *The following two sets of generators generate full rank DLAs, and concomitantly lead to controllable systems:*

- $\mathcal{G}_{\text{HEA}} = \left\{X_i, Y_i\right\}_{i=1}^{n} \bigcup \left\{\sum_{i=1}^{n-1} Z_i Z_{i+1}\right\}$,
- $\mathcal{G}_{\text{SG}} = \left\{\sum_{i=1}^{n} X_i, \sum_{i<j} \left(h_i Z_i + J_{ij} Z_i Z_j\right)\right\}$, *with $h_i, J_{ij} \in \mathbb{R}$ sampled from a Gaussian distribution.*

See Appendix I for a proof of Proposition 2.

The first case in Proposition 2 corresponds to the generators of PSA layered Hardware Efficient Ansatz [80], and hence Proposition 1 indicates that this system can exhibit barren plateaus. While it is known that the layered Hardware Efficient Ansatz converges to a 2-design for sufficient depth [81, 82, 83, 37, 49], the proof of existence of barren plateaus for this ansatz presented here is novel in that we show that the system is controllable.

The second result in Proposition 2 pertains to determining the ground state energy of quantum spin glasses (usually configured to encode solution to combinatorial optimization problems) [84, 85] with a PSA generated by $\mathcal{G}_{\text{SG}}$. Hence, since the system is controllable, according to Proposition 1, such an ansatz will also exhibit a barren plateau. This provides a no-go theorem for determining the ground state of certain spin glasses with Hamiltonians using deep PSA variational ansatzes .

4.2 Subspace controllable systems

Let us now consider the case of reducible DLAs, i.e., control systems with symmetries. Here we recall that in this case the DLA is a direct sum of the form $\mathfrak{g} = \bigoplus_j \mathfrak{g}_j$, such that any unitary $U(\boldsymbol{\theta})$ in the dynamical group $\mathbb{G}$ preserves the subspaces $U(\boldsymbol{\theta})\mathcal{H}_j \subset \mathcal{H}_j$. Then, similarly to the fully controllable case, if a system is subspace controllable in a given subspace there exists a depth at which the unitaries $U(\boldsymbol{\theta})$ form 2-designs in that subspace. In such a case, we can derive the following theorem for the variance of the cost function partial derivative with respect to a parameter $\theta_\mu \equiv \theta_{pq}$ associated to layer p and generator H_q (see Eq. (2) for a definition of the ansatz). In the following, will slightly abuse notation and denote $H_\mu = H_q$.

Theorem 2 (Subspace controllable)**.** *Consider a system that is reducible, i.e. so that the Hilbert space is $\mathcal{H} = \bigoplus_j \mathcal{H}_j$ with each $\mathcal{H}_j$ invariant under the action of the dynamical Lie group $\mathbb{G}$ (see Def. 4), and controllable on some $\mathcal{H}_k$ of dimension d_k (i.e. $\mathfrak{g}_k = \mathfrak{u}(d_k)$ or $\mathfrak{su}(d_k)$). Consider a cost function $C(\boldsymbol{\theta})$ in the form of Eq. (1) and suppose that the number of layers L in the circuit is enough to allow the distribution of unitaries $U(\boldsymbol{\theta})$ to be ε close to a 2-design in $\mathcal{H}_k$. Then, if the initial state is such that $\rho \in \mathcal{H}_k$, the variance of the cost function partial derivative with respect to parameter θ_μ is given by*

$$\text{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] = \frac{2d_k}{(d_k^2-1)^2} \Delta(H_\mu^{(k)}) \Delta(O^{(k)}) \Delta(\rho^{(k)}) \,, \quad (13)$$

where O is the operator whose expectation value is being minimized and H_μ is the generator of the corresponding gate. Here $\Delta(A) = D_{HS}\left(A, \text{Tr}[A]\frac{\mathbb{1}_d}{d}\right)$, with $D_{HS}(A,B) = \text{Tr}[(A-B)^2]$ the Hilbert-Schmidt distance, and where we defined $A^{(k)}$ as the reduction of operator A onto the subspace of $\mathcal{H}_k$. Explicitly, $A^{(k)} = Q_k A Q_k^\dagger$ with Q_k a matrix (of dimension $d_k \times 2^n$) with columns corresponding to a basis of $\mathcal{H}_k$.

See Appendix J for a proof of Theorem 2.

Theorem 2 shows that the input state ρ can actually play a crucial role in determining the gradient scaling of the cost function. Specifically, if ρ belongs to an invariant subspace where the system is controllable, then the scaling of the cost function partial derivative variance is determined by the dimension of the invariant subspace rather than by the dimension $d = 2^n$

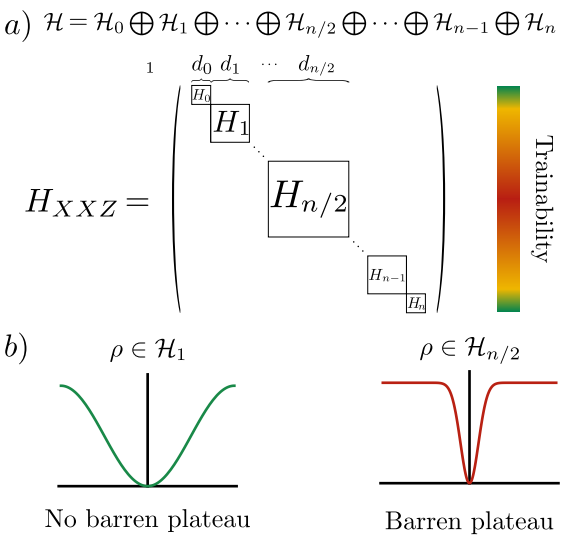

Figure 3: **Trainability analysis for a PSA used to prepare the ground state of a Heisenberg XXZ Hamiltonian.** a) Due to the symmetries in the XXZ Hamiltonian of Eq. (15), the Hilbert space spanned by the PSA can be expressed as a direct sum of invariant subspaces $\mathcal{H}_k$ (of dimension $d_k = \binom{n}{k}$) composed of states with k excitations. Hence both H_{XXZ} and $U(\boldsymbol{\theta})$ generated by Eq. (16) will be block diagonal. Since the system is subspace controllable in each invariant subspace, the gradient scaling of the PSA can be analyzed via Theorem 2. b) The existence or absence of barren plateaus is directly determined by the dimension of the invariant subspace to which the input state ρ belongs. For instance, the cost can be trainable for $\rho \in \mathcal{H}_1$, but will exhibit a barren plateau if $\rho \in \mathcal{H}_{n/2}$, as in the latter case the dimension $d_{n/2}$ is exponentially large.

of the Hilbert space. Hence, the cost function $C(\boldsymbol{\theta})$ might exhibit barren plateaus in some subspaces but not in others. This is formalized in the following corollary.

Corollary 2. *Consider an ansatz of the form in (2) giving rise to a reducible DLA, and let $\rho \in \mathcal{H}_k$, with $\mathcal{H}_k$ some invariant subspace that is controllable (i.e. the DLA reduced to such subspace is full rank). The following bound holds*

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] \leqslant \frac{4d_k^2}{(d_k^2-1)^2}\sqrt{\mathrm{Tr}[H_\mu^4]}\sqrt{\mathrm{Tr}[O^4]}\,. \quad (14)$$

That is, provided $\mathrm{Tr}[(H_\mu)^4], \mathrm{Tr}[O^4] \in \mathcal{O}(2^n)$, the cost function will exhibit a barren plateau for any subspace such that $d_k \in \mathcal{O}(2^n)$.

See Appendix K for a proof of Corollary 2. In addition, in this appendix, we also note relevant cases for which $\mathrm{Tr}[(H_\mu)^4], \mathrm{Tr}[O^4] \in \mathcal{O}(2^n)$ holds.

For example, let us consider the problem of preparing the ground state of a Heisenberg XXZ spin chain

$$H_{XXZ} = \sum_{i=1}^{n-1}(X_iX_{i+1} + Y_iY_{i+1} + JZ_iZ_{i+1}) \quad (15)$$

with a PSA generated by $\mathcal{G} = \mathcal{G}_{XXZ_U} \bigcup \{Z_1\}$. Here, the XXZ generators

$$\mathcal{G}_{XXZ_U} = \left\{\sum_i X_iX_{i+1} + Y_iY_{i+1}, \sum_i Z_iZ_{i+1}\right\}_{i:even,odd} \quad (16)$$

are accompanied by a *control* generator Z_1, which is introduced precisely to make the system (subspace) controllable. We remark that we here employ a U subindex in $\mathcal{G}_{XXZ_U}$ to indicate that this set of generators is uncontrollable. Since all elements in $\mathcal{G}_{XXZ_U}$ commute with $M_z = \sum_{i=1}^n Z_i$, the Hilbert space fractures into $n+1$ invariant subspaces of fixed excitation [3] $\mathcal{H} = \bigoplus_{m=0}^n \mathcal{H}_m$, of dimension $d_m = \dim(\mathcal{H}_m) = \binom{n}{m}$ [86].

Because the example set $\mathcal{G}$ has a DLA that is full rank on every subspace [87], we can analyze the trainability of such a PSA using Theorem 2. The implications of Corollary 2 for such a VQA are schematically shown in Fig. 3. Here we find that the presence, or absence, of barren plateaus for the PSA $U(\boldsymbol{\theta})$ generated by Eq. (16) is completely determined by the scaling of the invariant subspace to which the input state belongs. For instance, the cost may not exhibit a barren plateau if ρ has a number of excitations that does not scale with n, while it will have a barren plateau for $k = n/2$ (as in this case the dimension $d_{n/2}$ scales exponentially with the number of qubits).

4.3 Uncontrollable and reducible systems

Analyzing the scaling of the gradients in the case of uncontrollable systems becomes much more intricate than in the controllable or subspace controllable cases, mainly because integrating over the Haar measure of proper subgroups of the unitary group is not so straightforward [88]. As shown in this (and the next) section, one can still obtain a few analytical results for these cases. In particular, one can derive an upper bound for the variance of partial derivatives in terms of the degree of expressibility on the invariant subspaces, in a spirit similar to that of [47].

Before presenting our main results for uncontrollable and reducible systems, it is convenient to introduce some notation. We will use U_B and U_A, respectively, to address the portions of the circuit that come before and after a given parameter $\theta_\mu \equiv \theta_{pq} \in \boldsymbol{\theta}$. That is,

$$U_B = \prod_{m=0}^{q} e^{-iH_m\theta_{pm}} \left(\prod_{l=1}^{p-1}\prod_{k=0}^{K} e^{-iH_k\theta_{lk}}\right), \quad (17)$$

and

$$U_A = \prod_{l=p+1}^{L}\prod_{k=0}^{K} e^{-iH_k\theta_{pk}} \left(\prod_{k=q+1}^{K} e^{-iH_k\theta_{pk}}\right), \quad (18)$$

[3] We recall that a state $|\psi\rangle$ has m excitations if it can be expressed as a linear combination of computational basis states with Hamming weight m

where we have omitted the $\boldsymbol{\theta}$ dependency for simplicity.

Then, the following theorem holds.

Theorem 3. *Consider a system that is reducible and let $\rho \in \mathcal{H}_k$ with $\mathcal{H}_k$ an invariant subspace of dimension d_k. Then, the variance of the cost function partial derivative is upper bounded by*

$$\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] \leqslant \min\{G_A(\rho^{(k)}), G_B(O^{(k)})\}\,, \tag{19}$$

with

$$G_B(\rho^{(k)}) = \left(\left\| \mathcal{A}_{U_B^{(k)}}\left((\rho^{(k)})^{\otimes 2}\right)\right\|_2 - \frac{\Delta(\rho^{(k)})}{d_k^2-1}\right) \mathrm{Tr}\left[\left\langle X^2\right\rangle_{U_A^{(k)}}\right]$$

$$G_A(O^{(k)}) = \left(\left\| \mathcal{A}_{U_A^{(k)}}\left((O^{(k)})^{\otimes 2}\right)\right\|_2 - \frac{\Delta(O^{(k)})}{d_k^2-1}\right) \mathrm{Tr}\left[\left\langle Y^2\right\rangle_{U_B^{(k)}}\right].$$

Here we defined $X = [H_\mu^{(k)}, (U_A^{(k)})^\dagger O\, U_A^{(k)}]$ and $Y = [H_\mu^{(k)}, (U_B^{(k)})\rho^{(k)}(U_B^{(k)})^\dagger]$. For simplicity, we here employed the short-hand notation $\langle\cdot\rangle_{U_x^{(k)}}$ (with $x = A, B$) indicates the expectation value over the distribution of unitaries obtained from $U_x^{(k)}$ in the k-th subsystem. Finally, $\|M\|_2 = \sqrt{\mathrm{Tr}[M^\dagger M]}$ is the Frobenius norm, and $\Delta(\cdot)$ was defined in Theorem 2.

See Appendix L for a proof of Theorem 3.

Theorem 3 generalizes the expressibility results in [47] to invariant subspaces. More specifically, Theorem 3 provides a bound on the variance of the cost function partial derivative $\partial_\mu C(\boldsymbol{\theta})$ as a function of the ansatz expressibility on the relevant invariant subspace. Hence, similar to the results observed in [47], the more expressible an ansatz is in a subspace, the smaller the gradients will be. Moreover, ansatzes that are very expressible in subspaces with exponentially large dimensions can exhibit barren plateaus as the variance of the cost function partial derivative will vanish exponentially, according to Eq. (19).

4.4 Uncontrollable and irreducible systems

Here, we analyze a case where the DLA is an irreducible representation of some proper subalgebra of $\mathfrak{su}(d)$. Specifically, we consider a toy model ansatz $U(\boldsymbol{\theta}) = \prod_{l=1}^{L} e^{-i\theta_{lx}S_y} e^{-i\theta_{ly}S_x}$ with generators $\mathcal{G} = \{S_x, S_y\}$, where $\mathfrak{g} = \{iS_x, iS_y, iS_z\}$ is the spin $S = (d-1)/2$ irreducible representation of $\mathfrak{su}(2)$. That is, $[S_j, S_k] = 2i\epsilon_{jkl}S_l$ with ϵ the Levi-Civita symbol and $j,k,l \in \{x,y,z\}$. We address the task of minimizing a cost function of the form

$$C(\boldsymbol{\theta}) = \langle m|U^\dagger(\boldsymbol{\theta})(S_x+S_y+S_z)U(\boldsymbol{\theta})|m\rangle\,, \tag{20}$$

where $|m\rangle$ is an eigenstate of S_z, i.e., $S_z|m\rangle = m|m\rangle$ with $m \in \{-S, -S+1, \ldots, S-1, S\}$.

Let us analyze the variance of partial derivative with respect to parameter $\theta_\mu = \theta_{jx}$, i.e., the one corresponding to the generator S_x on the j-th layer. Assuming a depth p such that that the distributions $U_A(\boldsymbol{\theta})$ and $U_B(\boldsymbol{\theta})$ converge to ε-approximate 2-designs on the dynamical Lie group $\mathbb{G}$ (which in this case is the d-dimensional irreducible representation of $\mathcal{SU}(2)$), we are able to explicitly integrate over the Haar measure on $\mathbb{G}$ and find the following proposition to hold.

Proposition 3. *Consider the cost function of Eq. (20). Let $\theta_\mu = \theta_{j,x}$, and let us assume that the circuit is deep enough to allow for the distribution of unitaries U_A and U_B to converge to 2-designs on $\mathbb{G} = \mathcal{SU}(2)$. Then variance of the cost function partial derivative $\partial_\mu C(\boldsymbol{\theta}) = \partial C(\boldsymbol{\theta})/\partial\theta_\mu$ is*

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] = \frac{2m^2}{3}\,. \tag{21}$$

See Appendix M for a proof of Proposition 3.

Proposition 3 shows that the variance of the cost function again depends on the input state $|m\rangle$, which is a similar result to the one obtained in Theorem 2. Moreover, here $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ can in fact be as large as d^2. This is due to the fact that the "size" (the difference between maxima and minima) of the landscape also grows with d. One can get rid of this effect by considering a normalized cost instead, $\widetilde{C}(\boldsymbol{\theta}) = C(\boldsymbol{\theta})/S$, where the *ad-hoc* factor $1/S$ guarantees that the landscape is $|C(\boldsymbol{\theta})| \leqslant 1$ for all values of d. The variance of such normalized landscape is $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu \widetilde{C}(\boldsymbol{\theta})] = \frac{2}{3}\frac{m^2}{S^2} = \frac{8}{3}\frac{m^2}{(d-1)^2}$, that is, vanishes exponentially for initial states with $|m| \in \mathcal{O}(\mathrm{poly}(\log(d)))$. Similar to the subspace controllable results in Corollary 2, here the choice of initial state is again crucial as it can lead to the cost function exhibiting barren plateaus.

4.5 General case: linking gradient scaling to the dimension of the Lie algebra

In this section we note that the dimension of the DLA can be linked to the scaling of the variance of the cost function partial derivatives. This opens up the possibility of diagnosing the existence of barren plateaus of uncontrollable systems by analyzing the scaling of their DLAs. First, let us remark that a key aspect of the toy model in Section 4.4 is that the dimension of the DLA is $\dim(\mathfrak{g}) = 3$. This is independent of the dimension d of the Hilbert space it acts on. Moreover, as shown in Eq. (21) the variance is also independent of d as it does not present the typical dimensional-dependent factor in the denominator that one usually obtains when integrating over unitary 2-designs (see Eq. (13) in Theorem 2).

For instance, when the system is controllable, $\dim(\mathfrak{g}) = d^2 - 1 = 2^{2n} - 1$, and thus the dimension of the DLA is exponentially growing with the system size n. Concomitantly, one finds that $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] = \frac{2d}{(d^2-1)^2}\, g(H_\mu, O, \rho)$ [37, 49], and hence the variance is exponentially vanishing with the system size. A similar result is obtained in the subspace controllable case (see Theorem 2) where the variance is of the form $\frac{2d_k}{(d_k^2-1)^2} g(H_\mu^{(k)}, O^{(k)}, \rho^{(k)})$.

These facts have led us to conjecture that the dimension of the DLA plays a key role in determining the presence or absence of barren plateaus in the cost function landscape. More specifically, for PSAs with sufficient depth (i.e., with a depth such that the distribution of unitaries generated by $U(\boldsymbol{\theta})$ has converged to the Haar measure in the Lie group $\mathbb{G}$), we have noted that the following conjecture appears to hold.

Conjecture 1. *Let the state ρ belong to a subspace $\mathcal{H}_k$ associated with a subspace DLA $\mathfrak{g}_k$ (or sub-DLA, the subrepresentation in $\mathfrak{g}$ where ρ has support on). Then, the scaling of the variance of the cost function partial derivative is inversely proportional to the scaling of the dimension of the DLA, i.e.*

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] \in \mathcal{O}\left(\frac{1}{\mathrm{poly}(\dim(\mathfrak{g}_k))}\right). \tag{22}$$

The implications of Conjecture 1 are as follows. First, it means that systems with a sub-DLA $\mathfrak{g}_k$ [4] that is polynomially growing with the system size can exhibit gradients that vanish only polynomially, and hence may not exhibit barren plateaus. Conversely, systems with a sub-DLA that is exponentially growing with the system size would exhibit gradients that vanish exponentially with the system size, hence exhibiting barren plateaus. Here, we remark that systems with sub-DLAs that are not exponentially growing may still have barren plateaus which are not related to the dimension of the DLA. For instance, if the cost function is global, the system can still exhibit barren plateaus even with trivial ansatzes that do not a have exponentially growing dimension of the DLA [49].

Using Conjecture 1, one could diagnose gradient scalings by determining the size of the Lie algebra of a given ansatz $U(\boldsymbol{\theta})$. This comes at the cost of taking the set of generators $\mathcal{G}$ and computing the DLA. While numerical methods (as in Algorithm 1) can prove valuable insights for small system sizes, these algorithms will generally scale poorly in the number of qubits. Hence, performing a theoretical analysis of the DLA (similar to the one performed in Proposition 2) is a preferable method.

Here we remark that there are simple (yet pathological) cases that show that Eq. (22) does not preclude the possibility that systems with algebras that grow polynomial with the system size may still exhibit barren plateaus. For instance, consider Eq. (13), where ρ belongs to a subspace with polynomially growing algebra: $d_k \in \mathcal{O}(\mathrm{poly}(n))$. Then, note that if the input state is exponentially close to being maximally mixed on $\mathcal{H}_k$ (i.e., if $\Delta(\rho^{(k)}) \in \mathcal{O}(1/2^n)$) one can easily verify that the system will exhibit a barren plateau according to Definition 1 as the cost function partial derivative will be exponentially vanishing. Here, the barren plateau arises not from the dimension of the DLA being exponentially large but rather from trying to train a VQA on an input state that is exponentially close to being maximally mixed. A similar result can be found if $H_\mu^{(k)}$ is exponentially close to the identity. Hence, we remark that Conjecture 1 does not imply that systems with polynomially growing algebras are exempt from having barren plateaus, as cases where ρ ($H_\mu^{(k)}$) is exponentially close to being maximally mixed (the identity) will naturally be hard to train from the definition of the cost function in Eq. (1).

We finally note that to further support the claim in Conjecture 1, we present in the following section results obtained from numerically computing the scaling of the variance of the cost function partial derivatives for systems with DLAs having several different dependencies on the number of qubits. As discussed in Section 5, we see that the the result in Conjecture 1 holds true for all cases considered, as in these cases the scaling of the variance of the cost function partial derivative is inversely proportional to the scaling of the dimension of the DLA. In addition, based on our conjecture one can accurately make predictions regarding whether a given modification to an ansatz (adding a new generator to $\mathcal{G}$ by introducing a new unitary in each layer) might improve or be detrimental to the trainability of the parameters.

[4] Let us note that one should look at the sub-DLA instead of the full DLA, since when there are symmetries and the initial state belongs to one or multiple invariant subspaces, the dynamics is contained in those.

5 NUMERICAL SIMULATIONS

In this section we present results obtained by numerically computing the variance of the cost function partial derivatives for systems with different PSAs, and with DLAs of dimensions with different scaling. In particular, we consider systems that are controllable, subspace controllable, and subspace uncontrollable. As we show, in all cases Conjecture 1 is verified. Finally, we refer the reader to Appendix M for a numerical study of the toy model in Section 4.4, where $\mathfrak{g}$ and $\mathbb{G}$ are, respectively, the d-dimensional irreducible representations of $\mathfrak{su}(2)$ and $\mathcal{SU}(2)$.

5.1 Controllable systems

First, let us remind that when the system is controllable, $U(\boldsymbol{\theta})$ forms a 2-design (see Proposition 1). In this case the scaling of $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ has been widely analyzed in the literature (see for instance [37, 47]). Controllable systems, as previously discussed, satisfy Conjecture 1. In Fig. 4 we show the variance of cost function partial derivatives as a function of $1/\dim(\mathfrak{g})$ for the cost function

$$C(\boldsymbol{\theta}) = \langle \mathbf{0}|U^\dagger(\boldsymbol{\theta})(Z_1 \otimes Z_2)U(\boldsymbol{\theta})|\mathbf{0}\rangle. \tag{23}$$

Here, $U(\boldsymbol{\theta})$ is a layered Hardware Efficient ansatz (see the circuit in the inset of Fig. 4) with 200 layers and

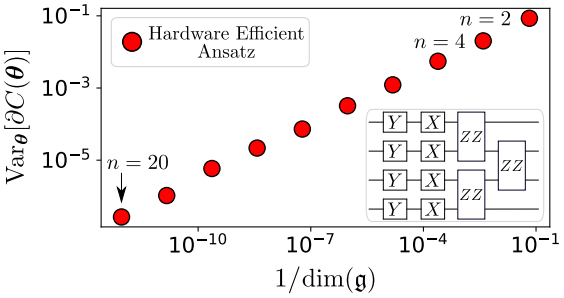

Figure 4: **Variance of cost function partial derivatives versus inverse of the DLA dimension for a controllable system.** The layered Hardware Efficient Ansatz (shown in the inset for $n = 4$) is a controllable system with generators given in Proposition 2. Then, as shown in Proposition 1, the cost function of Eq. (23) exhibits a barren plateau and hence $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] \in \mathcal{O}(1/2^n)$. Moreover, since the system is controllable one finds that $\dim(\mathfrak{g}) = 4^n - 1$. Hence, as shown in the plot, Conjecture 1 holds for controllable systems, since the dependence of $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ versus $1/\dim(\mathfrak{g})$ is linear on a log-log scale.

where $|\mathbf{0}\rangle = |0\rangle^{\otimes n}$. For each value of $n = 2, 4, \ldots, 20$, the variance was computed by randomly initializing 1000 sets of parameters. Since this system is controllable (as proved in Proposition 1), then $\dim(\mathfrak{g}) = d^2 - 1 = 4^n - 1$. In Fig. 4 we see that, as expected, the variance is a polynomial function of $1/\dim(\mathfrak{g})$ (indicated by a straight line in a log-log scale).

5.2 Reducible systems

5.2.1 The XXZ model

Let us first consider the task of finding the ground state energy of the XXZ Hamiltonian H_{XXZ} of Eq. (15). First, let us notice that $\mathcal{G}_{XXZ_U}$, the uncontrollable set of generators of Eq. (16), has two symmetries: magnetization and parity. Hence, the DLA is reducible, i.e. a sum of irreducible sub-representations $\mathfrak{g}_{XXZ_U} = \bigoplus_{\substack{m=0 \\ \sigma=\pm}}^{n} \mathfrak{g}_{m,\sigma} \subseteq \mathfrak{u}(d_{m,\sigma})$, where the indices m and σ indicate number of excitations and parity, respectively (see Appendix N for details). Notably, the system can be rendered subspace controllable (while preserving the invariant subspace structure) by introducing an additional generator consisting of local fields at the ends of the chain [89, 90]

$$\mathcal{G}_{XXZ} = \mathcal{G}_{XXZ_U} \cup \{\mathrm{Z}_1 + \mathrm{Z}_N\} \,. \tag{24}$$

The new set $\mathcal{G}_{XXZ}$ generates a DLA that is full rank on each of the invariant subspaces, i.e. $\mathfrak{g}_{\mathrm{XXZ}} = \bigoplus_{\substack{m=0 \\ \sigma=\pm}}^{n} \mathfrak{u}(d_{m,\sigma})$. In Figure 5(a), we sketch a single layer of the ansatz generated by $\mathcal{G}_{XXZ}$. Note that upon the removal of the unitary generated by $Z_1 + Z_N$ (indicated by a shaded area), one recovers the HVA ansatz with generators $\mathcal{G}_{XXZ_U}$ proposed in Ref. [59].

Figure 5(b) shows numerical results obtained by computing $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ for the cost function

$$C(\boldsymbol{\theta}) = \langle \psi_{m,+} | U^\dagger(\boldsymbol{\theta}) H_{XXZ} U(\boldsymbol{\theta}) | \psi_{m,+} \rangle / n \,, \tag{25}$$

with $J = 1$. Here, $U(\boldsymbol{\theta})$ is the HVA ansatz generated by $\mathcal{G}_{XXZ}$ (see Eq. (24)) with $L = 6n$ layers, and $|\psi_{m,+}\rangle$ is an initial state with m excitations and even parity $\sigma = +$ (see Appendix O for details). For each system system size $n = 2, 4, \ldots, 20$, and for each value of m, we computed the variance with respect to $\theta_{\frac{L}{2},2}$ by randomly initializing each $\boldsymbol{\theta}_{pq} \in [0, 2\pi]$ and averaging over 9500 sets of parameters (for $n = 20$ we averaged over 2700 sets of parameters).

In Figure 5(b, left) we see that for $m = 1, 2, \ldots, 5$ the variance of the cost function partial derivative is polynomially decreasing with n, indicating that the cost function does not exhibit a barren plateau for initial states with fixed number of excitations. However, in the case $m = n/2$ (see Figure 5(b, right)), one can observe that $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ vanishes exponentially. In addition, in Figure 5(b) we also show the curves for $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ obtained from the analytical result in Eq. (13) of Theorem 2. The agreement between theoretical and numerical results indicates that, already for the linear depths used in the experiments, the ansatz is well converged to a 2-design. Hence, the results in Theorem 2 suggest that the system will exhibit a barren plateau when initialized on any subspace where $d_m \in \mathcal{O}(2^n)$, for example, in the case of $m = n/2$ excitations.

In addition, Figure 5(b, right) shows the scaling of the variance for the PSA generated by $\mathcal{G}_{XXZ_U}$, with an initial state with $m = n/2$. As previously noted, this case is not controllable and hence Theorem 2 does not hold. However, the gradient scaling of the cost function can still be diagnosed using the expressibility result of Theorem 3. First, we note that the variance values for the uncontrollable case are larger than the ones for the controllable case. This result is in accordance with the fact that the smallest variances are reached with the higher expressibilities. Still, despite the system not being controllable, we find that the cost function still exhibits a barren plateau as the cost vanishes exponentially with n.

In Figure 5(c) we show that for all subspace controllable cases considered, Conjecture 1 holds. Specifically, we have shown $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ as a function of $1/\dim(\mathfrak{g})$, and we see a linear dependence in a log-log scale. This is true both for the exponentially growing algebras ($m = n/2$) as well as for the polynomially growing algebras ($m = 1, 2, \ldots, 5$). Moreover, we see that the Conjecture is verified on the subspace uncontrollable case of $\mathcal{G}_{XXZ_U}$ (pink stars), where the dimension of the DLA is exponentially growing, and concomitantly, the variance of the cost function partial derivative is exponentially suppressed.

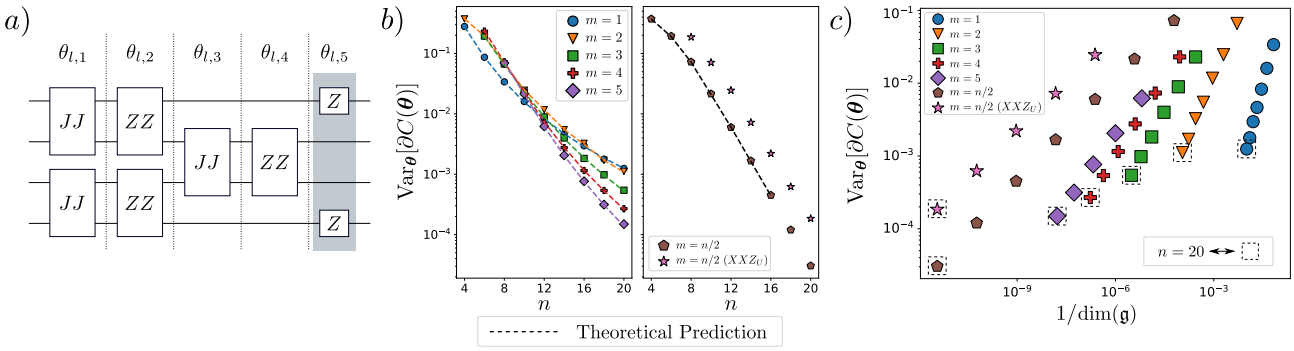

Figure 5: **Numerical results for the XXZ model.** a) Schematic illustration of a single layer of the subspace controllable periodic structure ansatz for the set of generators $\mathcal{G}_{XXZ}$ in Eq. (24) for $n = 4$ qubits. Removing the unitary in the shaded area leads to the uncontrollable periodic structure HVA generated by $\mathcal{G}_{XXZ_U}$. b) Left panel: Variance of the partial derivative of the cost in Eq. (25) versus the number of qubits n. The different markers correspond to initial states with a number of excitations $m = 1, 2, \ldots, 5$ (left) with an ansatz generated by $\mathcal{G}_{XXZ}$, and to $m = n/2$ (right) with an ansatz generated by $\mathcal{G}_{XXZ}$ and by $\mathcal{G}_{XXZ_U}$. Here we recall that a state $|\psi\rangle$ has m excitations if it can be expressed as a linear combination of computational basis states with Hamming weight m. The dashed lines represent the theoretical prediction of Eq. (13). In both cases the plot is shown in a log-linear scale. For $m = 1, 2, \ldots, 5$ the variance is polynomially vanishing with n, while for $m = n/2$ the variance is exponentially vanishing with n. c) Variance of the cost function partial derivative versus $1/\dim(\mathfrak{g})$. The plot is shown in a log-log scale. For each value of m we see a linear dependence, which verifies Conjecture 1.

5.2.2 The Ising Model

In this section we present results obtained for numerically simulating the use of a PSA to find the ground state of the Ising model. Specifically, consider the Hamiltonian of the one-dimensional Transverse Field Ising Model (TFIM)

$$H_{\mathrm{TFIM}} = \sum_{i=1}^{n_f} Z_i Z_{i+1} + h_x \sum_{i=1}^{n} X_i \,, \tag{26}$$

where $n_f = n - 1$ in the case of open boundary conditions, and $n_f = n$ in the periodic boundary conditions case (where $Z_{n+1} \equiv Z_1$). Then, as shown in Figure 6(a, left), the ansatz is generated by the set

$$\mathcal{G}_{\mathrm{TFIM}} = \left\{ \sum_{i=1}^{n_f} Z_i Z_{i+1}, \sum_{i=1}^{n} X_i \right\} . \tag{27}$$

Note that the PSA generated by $\mathcal{G}_{\mathrm{TFIM}}$ is in fact the QAOA employed for solving the MAXCUT problem on a 2-regular graph [19, 57].

As discussed in Appendix N, the generators in $\mathcal{G}_{\mathrm{TFIM}}$ (with open boundary conditions) have two symmetries: parity symmetry Π, and the so-called $\mathbb{Z}_2$ symmetry $\Pi_{\mathbb{Z}_2}$ (representing an invariance under a global flip in the qubits). The Hilbert space is broken into four invariant subspaces, $\mathcal{H} = \bigoplus_{\sigma,\sigma'} \mathcal{H}_{\sigma,\sigma'}$, where $\sigma, \sigma' = \pm 1$ respectively spanning the eigenvalues of Π and $\Pi_{\mathbb{Z}_2}$, and where $\dim(\mathcal{H}_{\sigma,\sigma'})$ is exponentially growing, i.e., $\dim(\mathcal{H}_{\sigma,\sigma'}) \in \mathcal{O}(2^n)$. In turn, the DLA decomposes as $\mathfrak{g}_{\mathrm{TFIM}} = \bigoplus_{\sigma,\sigma'} \mathfrak{g}_{\sigma,\sigma'} \subseteq \mathfrak{u}(d_{\sigma,\sigma'})$. However, employing Algorithm 1 we computed the dimension of the DLA generated by $\mathcal{G}_{\mathrm{TFIM}}$ and we found that it only grows polynomially with n. That is, we obtain that

$$\dim(\mathfrak{g}_{\mathrm{TFIM}}) = n^2 \,. \tag{28}$$

Clearly, this implies that $\dim(\mathfrak{g}_{\sigma,\sigma'}) \leqslant n^2$ for all σ, σ'.

Note that the set $\{\mathbb{1}, \Pi\}$ constitutes a representation of S_2, the symmetric group of two elements, under which the open-boundary-condition TFIM generators are invariant. Instead, the TFIM generators with closed boundary conditions are invariant under a representation of C_n, the cyclic group of n elements. As discussed in Appendix N, the dimension of the DLA now grows linearly instead of quadratically.

Similarly to what happened in the XXZ case, we can turn the TFIM model subspace controllable upon the introduction of an extra generator. Consider the set

$$\mathcal{G}_{\mathrm{LTFIM}} = \mathcal{G}_{\mathrm{TFIM}} \bigcup \left\{ \sum_{i=1}^{n} Z_i \right\} , \tag{29}$$

leading to the PSA in Figure 6(a, right). The set $\mathcal{G}_{\mathrm{LTFIM}}$ can also be regarded as being constituted by the individual terms in the one-dimensional Longitudinal and Transverse Field Ising Model (LTFIM) Hamiltonian

$$H_{\mathrm{LTFIM}} = \sum_{i=1}^{n_f} Z_i Z_{i+1} + h_x \sum_{i=1}^{n} X_i + h_z \sum_{i=1}^{n} Z_i \,. \tag{30}$$

In addition, the ansatz generated by $\mathcal{G}_{\mathrm{LTFIM}}$ is also a QAOA-type ansatz where an additional mixer has been added.

In the case of open boundary conditions, the $\sum_{i=1}^{n} Z_i$ term breaks the $\mathbb{Z}_2$ symmetry, and thus the set $\mathcal{G}_{\mathrm{LTFIM}}$ only conserves the parity symmetry, $\mathfrak{g}_{\mathrm{LTFIM}} = \bigoplus_{\sigma} \mathfrak{g}_{\sigma}$. Using Algorithm 1 we find that the DLA is full rank on both $\sigma = \pm 1$ parity subspaces, and hence

$$\dim(\mathfrak{g}_{\sigma}) \in \mathcal{O}(2^{2n}) \,. \tag{31}$$

Similarly, in the closed boundary condition case, one can also find that the dimension of the DLA grows

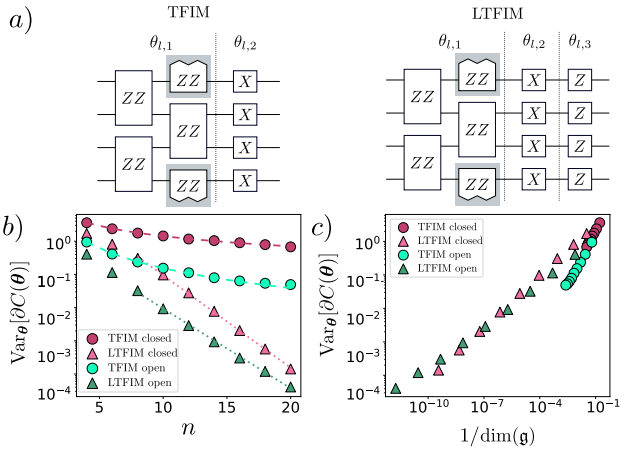

Figure 6: **Numerical results for the TFIM and LTFIM models.** a) Schematic illustration of a single layer of the PSA for the sets of generators $\mathcal{G}_{\text{TFIM}}$ in Eq. (27) (left), and $\mathcal{G}_{\text{LTFIM}}$ in Eq. (29) (right) for $n = 4$ qubits. By adding (removing) the gates in the shaded one obtains the ansatz in Eq. (27) with periodic (open) boundary conditions. b) Variance of the cost function partial derivative of the cost function in Eq. (32) versus the number of qubits n for each ansatz. The dashed (dotted) lines indicate the best polynomial (exponential) fit. The plot is shown in a log-linear scale. c) Variance of the cost function partial derivative versus $1/\dim(\mathfrak{g})$. The plot is shown in a log-log scale.

exponentially with n. This is an example where we show how a simple modification to the ansatz (adding a layer generated by $\sum_{i=1}^{n} Z_i$) can greatly change the dimension of the DLA, and, as discussed below, such a small change can greatly affect the trainability of the cost function.

In Figure 6(b) we show results for numerically computing $\text{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ for the cost function

$$C(\boldsymbol{\theta}) = \langle +|^{\otimes n} U^\dagger(\boldsymbol{\theta}) H_{\text{TFIM}} U(\boldsymbol{\theta}) |+\rangle^{\otimes n}/n\,, \qquad (32)$$

where $U(\boldsymbol{\theta})$ is the PSA generated by the set $\mathcal{G}_{\text{TFIM}}$ of Eq. (27) with $L = 12n$ layers for open boundary conditions, and $L = 6n$ for closed boundary conditions. For each value of $n = 4, 6, \ldots, 18$ we computed the variance by picking 4400 random sets of parameters, while for $n = 20$ we picked 1000 random intializations. In all cases the partial derivative was taken with respect to $\theta_{\frac{L}{2},2}$. We see from Figure 6(b) that the variance of the cost partial derivative vanishes polynomially with n for both open and closed boundary conditions, and hence the system does not exhibit a barren plateau. Then, as shown in Figure 6(c), once again, Conjecture 1 holds for both open and closed boundary conditions: $\text{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ and $\dim(\mathfrak{g}_{\text{TFIM}})$ respectively vanish, and grow, polynomially with n.

Moreover, in Figure 6(b) we also depict results obtained by computing $\text{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ for the LTFIM ansatz, using the same cost function of Eq. (32). Now, $U(\boldsymbol{\theta})$ is the PSA generated by the set $\mathcal{G}_{\text{LTFIM}}$ in Eq. (27) with $L = 6n$ layers. Using the same number of samples than for the TFIM case, we find

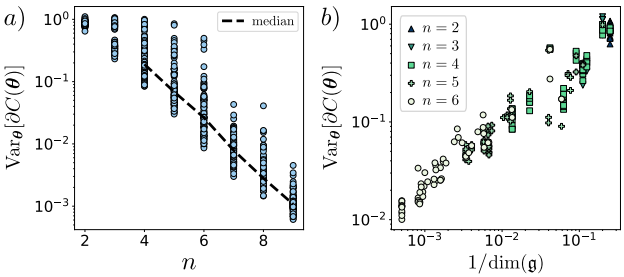

Figure 7: **Numerical results for the Erdös–Rényi model.** a) Variance of the cost function partial derivative of the cost function in Eq. (35) versus the number of qubits n. The dashed line indicates the medians across graphs computed for each value of n. The plot is shown in a log-linear scale. b) Variance of the cost function partial derivative versus $1/\dim(\mathfrak{g})$. The plot is shown in a log-log scale.

that $\text{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ vanishes exponentially with n for both open and closed boundary conditions, and hence the cost exhibits a barren plateau. We see in Figure 6(c) that Conjecture 1 also holds for the LTFIM ansatzes with open and closed boundary conditions, as this time $\text{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ vanishes exponentially with n, while $\dim(\mathfrak{g}_{\text{TFIM}})$ grows exponentially with n.

It is worth noting that, as discussed before, and as shown in Figure 6(a), the difference between the TFIM and the LTFIM ansatz is given by an additional unitary in each layer (parametrized by a single angle). However, despite this simple difference, we find the variance of the cost function have different scaling, as one cost exhibits a barren plateau while the other one does not exhibit a barren plateau.

5.2.3 Erdös–Rényi model

Let us now consider the task of solving MAXCUT problems with a QAOA ansatz. Here, we recall that MAXCUT is specified by a graph $G = (V, E)$ of nodes V and edges E, such that one seeks to determine a partition of the nodes of G into two sets that maximize the number of edges connecting nodes between sets. The MAXCUT Hamiltonian is given by

$$H_{\text{ER}} = -\frac{1}{2}\sum_{ij\in E}(\mathbb{1} - Z_i Z_j)\,, \qquad (33)$$

and we consider the standard QAOA ansatz generated by

$$\mathcal{G} = \left\{\sum_{i=1}^{n} X_i, H_{\text{ER}}\right\}. \qquad (34)$$

Let us analyze the variance of the partial derivative of the cost

$$C(\boldsymbol{\theta}) = \langle +|^{\otimes n} U^\dagger(\boldsymbol{\theta}) H_{\text{ER}} U(\boldsymbol{\theta}) |+\rangle^{\otimes n}/|E|\,, \qquad (35)$$

where we use $|E|$ (the number of edges in the graph) to normalize the cost function. For each value of $n = 2, 3, \ldots, 9$ we generated 90 graphs according to the Erdös–Rényi model [91]. That is, each graph G

was chosen uniformly at random from the set of all graphs of n nodes. Then, for each graph we sampled 3000 random initializations with $L = 12n$ layers and we took the partial derivative with respect to the angle in the $L/2$-th layer associated to the (mixer) Hamiltonian $\sum_{i=1}^{n} X_i$. In Fig. 7(a) we show results of $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ versus the number of qubits. Here we can see that, as expected, even for fixed n different graphs will have different value of the variance. However, by computing the median variance for each system size we found that the scaling of the median is exponentially decaying with the system size. While this result does not preclude the possibility of generating graphs that will not have a barren plateau, it suggests that uniform sampling of graphs from the Erdös–Rényi model will lead to the landscape for a typical graph having a barren plateau. Then, as shown in Fig. 7(b), we compute the dimension of the DLA for each graph, and find that Conjecture 1 is confirmed, as the relation between $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ and $\dim(\mathfrak{g})$ is linear in a log-log-scale.

6 DISCUSSION

In this work, we have explored a fundamental connection between VQAs and the theory of QOC with the purpose of analyzing the existence of barren plateaus in a family of periodic-structured ansatzes which contain, as special cases, the QAOA and the HVA, among other widely used ansatzes in variational quantum algorithms and quantum machine learning. Our results show that one can diagnose the presence of barren plateaus in the cost function landscape by analyzing the degree of controllability of the system, characterized by the dimension of the dynamical Lie algebra (DLA) obtained from the set of generators of the ansatz.

Our main results are the following. First, we show that if the DLA is full rank, i.e. if the system is controllable, then the cost function exhibits a barren plateau. This follows from the fact that, as we show, controllable systems converge to 2-designs. Here, we also derive an expression relating the depth required for a given ansatz to become an ε-approximate two-design with the expressibility of one of its layers.

We then consider systems with symmetries, where the Hilbert space partitions into invariant subspaces associated with the different eigenspaces. In this context, we show that when the system is subspace controllable, the existence of barren plateaus crucially depends on the input state to the VQA. For example, the cost might be trainable for certain input states, but might exhibit a barren plateau for others. Specifically, our results connect the scaling of the variance of cost function partial derivatives to that of the dimension of the subspace in which the input state has support on. Instead, when the system is subspace uncontrollable, we show that one can still upper bound the variance of the cost function partial derivative using the expressibility of the ansatz in the relevant subspace. This indicates that larger subspace expressibilities leads to smaller gradients.

Finally, we present an conjecture that shows that one can directly study the scaling of the cost function partial derivative variance by computing the dimension of the subspace DLA to which the input state belongs. This conjecture implies that ansatzes with polynomially growing DLAs can exhibit polynomially vanishing gradients, while ansatzes with exponentially growing DLAs should exhibit exponentially vanishing gradients.

In addition, we performed numerical simulations of VQAs with the hardware efficient ansatz, QAOA, and HVA, for problems such as preparing ground states of the XXZ model and of the Ising model, or solving MAXCUT problems on graphs generated from the Erdös–Rényi model. The numerical results match our theoretical predictions and hence verify our analytical results for controllable and subspace controllable systems. Moreover, in all cases considered we verify that our conjecture holds, further providing evidence that the scaling of the cost function partial derivative variance may be directly linked to the dimension of the subspace DLA.

Implications of our results to ansatz design

The broader implication of our results is that the framework introduced here can be used to design ansatzes, as one could potentially predict if an ansatz, or a modification to the ansatz, will lead to the cost function exhibiting a barren plateau. Hence, our work can be considered as paving the way towards trainability-aware ansatz design.

For instance, we have shown how a simple change in the ansatz structure, such as adding an additional parametrized unitary per layer, can greatly affect the gradient scaling of the cost by changing the controllability of the system. This means that one should be careful when employing schemes such as the Adaptive QAOA or quantum optimal control ansatz as the addition of an operator H to the set $\mathcal{G}$ of generators of the ansatz can lead to barren plateaus if the system becomes controllable (or subspace controllable in an exponentially growing subspace). In particular, if H does not commute with the elements in $\mathcal{G}$, one should analyze how the DLA changes by such addition before proceeding to change the ansatz.

Here, we crucially remark that one of the main advantages of the aforementioned theoretical analysis is that it can be performed classically (either analytically or numerically) as it just requires the evaluation of the DLA. Hence, our methods save precious quantum resources as one does not need to run the quantum algorithm, or even access a quantum computer, to test the trainability of the ansatzes.

Finally, we remark that if our conjecture holds more generally, then one can use this additional tool to directly study the trainability of an ansatz by estimating the scaling of the variance of the cost function partial derivative through the scaling of the dimension of the DLA. For example, such results can be used to show that certain ansatzes might not have exponentially vanishing gradients. For instance, when considering a QAOA ansatz for solving MAXCUT on 2-regular graphs, a straightforward computation of the DLA reveals its scaling is only linear in n. Hence, we expect (and we find) no barren plateaus. Similarly, one can use our conjecture to analyze ansatz proposals in the literature. For example, Ref. [71] recently proposed an ansatz generated by the set of products up to K-body Pauli X operators, i.e., $\mathcal{G} = \{X_i\}_i \cup \{X_i X_j\}_{i>j} \cup \{X_i X_j X_k\}_{i>j>k} \cup \cdots$. Since the ansatz is abelian, the dimension of DLA is just the number of generators. Thus, we expect that when using a poly number of layers the ansatz should be rid barren plateaus.

Outlook

In the present work, we have established a novel framework for diagnosing the presence of barren plateaus in VQAs. While here we mainly focus on the trainability of ansatzes for near-term quantum computing, our results should also be considered as useful in the broader context of QOC. For instance, while the barren plateau phenomenon has been recently widely studied in VQAs, it is clear from our manuscript that barren plateaus can (and will) also arise in QOC schemes (see also [92]). Hence, we leave for future work to study how some of the results derived for the trainability of VQAs be used to analyze the trainability of QOC control pulses.

In addition, we note that since our work studies the trainability of certain families of ansatzes, we also leave for future work to show how the tools here presented can be employed to study more general ansatzes (e.g., ansatzes for quantum machine learning applications) which do not necessarily have a periodic structure. In addition, we leave as an open question how the results in our conjecture can be generalized and formally proved.

7 ACKNOWLEDGMENTS

We thank Marco Farinati and Robert Zeier for useful discussions on Lie-algebras, and we also thank Zoe Holmes and Pablo Poggi for helpful discussions. ML acknowledges partial support by CONICET (PIP 112201 50100493CO), UBA-CyT (20020130100406BA) and ANPCyT (PICT-2016-1056)). ML was also supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under the Quantum Computing Applications Team (QCAT) program and also under the Accelerated Research in Quantum Computing (ARQC) program. Piotr C. was supported by the Laboratory Directed Research and Development (LDRD) program of Los Alamos National Laboratory (LANL) under project numbers 20190659PRD4. KS acknowledges support from the National Science Foundation under Grant No. 2014010. GM was supported by the US Department of Energy, Office of Advanced Scientific Computing Research, under the ARQC program and by the LANL ASC Beyond Moore's Law project. PJC acknowledges initial support from the LANL ASC Beyond Moore's Law project. MC acknowledges initial support from the Center for Nonlinear Studies at LANL. KS, PJC, and MC also acknowledge support from LDRD program of LANL under project number 20190065DR.

References

[1] Peter W Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994. DOI: 10.1109/SFCS.1994.365700. URL https://ieeexplore.ieee.org/document/365700.

[2] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009. DOI: 10.1103/PhysRevLett.103.150502. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.103.150502.

[3] Dominic W Berry, Andrew M Childs, Richard Cleve, Robin Kothari, and Rolando D Somma. Simulating hamiltonian dynamics with a truncated taylor series. *Physical Review Letters*, 114(9):090502, 2015. DOI: 10.1103/PhysRevLett.114.090502. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.114.090502.

[4] Iulia M Georgescu, Sahel Ashhab, and Franco Nori. Quantum simulation. *Reviews of Modern Physics*, 86(1):153, 2014. DOI: 10.1103/RevModPhys.86.153. URL https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.86.153.

[5] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2: 79, 2018. DOI: 10.22331/q-2018-08-06-79. URL https://quantum-journal.org/papers/q-2018-08-06-79/.

[6] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational

quantum algorithms. *Nature Reviews Physics*, 3(1):625–644, 2021. DOI: 10.1038/s42254-021-00348-9. URL https://www.nature.com/articles/s42254-021-00348-9.

[7] Carlos Bravo-Prieto, Ryan LaRose, M. Cerezo, Yigit Subasi, Lukasz Cincio, and Patrick Coles. Variational quantum linear solver. *arXiv preprint arXiv:1909.05820*, 2019. URL https://arxiv.org/abs/1909.05820.

[8] Hsin-Yuan Huang, Kishor Bharti, and Patrick Rebentrost. Near-term quantum algorithms for linear systems of equations. *arXiv preprint arXiv:1909.07344*, 2019. URL https://arxiv.org/abs/1909.07344.

[9] Xiaosi Xu, Jinzhao Sun, Suguru Endo, Ying Li, Simon C Benjamin, and Xiao Yuan. Variational algorithms for linear algebra. *Science Bulletin*, 66(21):2181–2188, 2021. DOI: 10.1016/j.scib.2021.06.023. URL https://www.sciencedirect.com/science/article/pii/S2095927321004631.

[10] Sam McArdle, Tyson Jones, Suguru Endo, Ying Li, Simon C Benjamin, and Xiao Yuan. Variational ansatz-based quantum simulation of imaginary time evolution. *npj Quantum Information*, 5(1):1–6, 2019. DOI: 10.1038/s41534-019-0187-2. URL https://www.nature.com/articles/s41534-019-0187-2.

[11] Harper R Grimsley, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications*, 10(1):1–9, 2019. DOI: 10.1038/s41467-019-10988-2. URL https://www.nature.com/articles/s41467-019-10988-2.

[12] Cristina Cirstoiu, Zoe Holmes, Joseph Iosue, Lukasz Cincio, Patrick J. Coles, and Andrew Sornborger. Variational fast forwarding for quantum simulation beyond the coherence time. *npj Quantum Information*, 6(1):1–10, 2020. DOI: 10.1038/s41534-020-00302-0. URL https://www.nature.com/articles/s41534-020-00302-0.

[13] Benjamin Commeau, M. Cerezo, Zoë Holmes, Lukasz Cincio, Patrick J. Coles, and Andrew Sornborger. Variational hamiltonian diagonalization for dynamical quantum simulation. *arXiv preprint arXiv:2009.02559*, 2020. URL https://arxiv.org/abs/2009.02559.

[14] Joe Gibbs, Kaitlin Gili, Zoë Holmes, Benjamin Commeau, Andrew Arrasmith, Lukasz Cincio, Patrick J. Coles, and Andrew Sornborger. Long-time simulations with high fidelity on quantum hardware. *arXiv preprint arXiv:2102.04313*, 2021. URL https://arxiv.org/abs/2102.04313.

[15] Yong-Xin Yao, Niladri Gomes, Feng Zhang, Cai-Zhuang Wang, Kai-Ming Ho, Thomas Iadecola, and Peter P Orth. Adaptive variational quantum dynamics simulations. *PRX Quantum*, 2(3):030307, 2021. DOI: 10.1103/PRXQuantum.2.030307. URL https://journals.aps.org/prxquantum/abstract/10.1103/PRXQuantum.2.030307.

[16] Suguru Endo, Jinzhao Sun, Ying Li, Simon C Benjamin, and Xiao Yuan. Variational quantum simulation of general processes. *Physical Review Letters*, 125(1):010501, 2020. DOI: 10.1103/PhysRevLett.125.010501. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.125.010501.

[17] Jonathan Wei Zhong Lau, Kishor Bharti, Tobias Haug, and Leong Chuan Kwek. Quantum assisted simulation of time dependent hamiltonians. *arXiv preprint arXiv:2101.07677*, 2021. URL https://arxiv.org/abs/2101.07677.

[18] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):1–7, 2014. DOI: doi.org/10.1038/ncomms5213. URL https://www.nature.com/articles/ncomms5213#citeas.

[19] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014. URL https://arxiv.org/abs/1411.4028.

[20] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016. DOI: 10.1007/978-94-015-8330-5_4. URL https://iopscience.iop.org/article/10.1088/1367-2630/18/2/023023.

[21] Sumeet Khatri, Ryan LaRose, Alexander Poremba, Lukasz Cincio, Andrew T Sornborger, and Patrick J Coles. Quantum-assisted quantum compiling. *Quantum*, 3:140, 2019. DOI: 10.22331/q-2019-05-13-140. URL https://quantum-journal.org/papers/q-2019-05-13-140/.

[22] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology*, 2(4):045001, 2017. DOI: 10.1088/2058-9565/aa8072. URL https://iopscience.iop.org/article/10.1088/2058-9565/aa8072.

[23] Ryan LaRose, Arkin Tikku, Étude O'Neel-Judy, Lukasz Cincio, and Patrick J Coles. Variational quantum state diagonalization. *npj Quantum Information*, 5(1):1–10, 2019. DOI: 10.1038/s41534-019-0167-6.

URL https://www.nature.com/articles/s41534-019-0167-6.

[24] Andrew Arrasmith, Lukasz Cincio, Andrew T Sornborger, Wojciech H Zurek, and Patrick J Coles. Variational consistent histories as a hybrid algorithm for quantum foundations. *Nature Communications*, 10(1):1–7, 2019. DOI: 10.1038/s41467-019-11417-0. URL https://www.nature.com/articles/s41467-019-11417-0.

[25] M. Cerezo, Alexander Poremba, Lukasz Cincio, and Patrick J Coles. Variational quantum fidelity estimation. *Quantum*, 4:248, 2020. DOI: 10.22331/q-2020-03-26-248. URL https://quantum-journal.org/papers/q-2020-03-26-248/.

[26] Y. Li and S. C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X*, 7:021050, Jun 2017. DOI: 10.1103/PhysRevX.7.021050. URL https://link.aps.org/doi/10.1103/PhysRevX.7.021050.

[27] Kentaro Heya, Ken M Nakanishi, Kosuke Mitarai, and Keisuke Fujii. Subspace variational quantum simulator. *arXiv preprint arXiv:1904.08566*, 2019. URL https://arxiv.org/abs/1904.08566.

[28] Kishor Bharti and Tobias Haug. Quantum-assisted simulator. *Physical Review A*, 104(4):042418, 2021. DOI: 10.1103/PhysRevA.104.042418. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.104.042418.

[29] M. Cerezo, Kunal Sharma, Andrew Arrasmith, and Patrick J Coles. Variational quantum state eigensolver. *npj Quantum Information*, 8(1):1–11, 2022. DOI: 10.1038/s41534-022-00611-6. URL https://doi.org/10.1038/s41534-022-00611-6.

[30] Jacob L Beckey, M. Cerezo, Akira Sone, and Patrick J Coles. Variational quantum algorithm for estimating the quantum Fisher information. *Physical Review Research*, 4(1):013083, 2022. DOI: 10.1103/PhysRevResearch.4.013083. URL https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.4.013083.

[31] Lennart Bittel and Martin Kliesch. Training variational quantum algorithms is np-hard. *Phys. Rev. Lett.*, 127:120502, Sep 2021. DOI: 10.1103/PhysRevLett.127.120502. URL https://link.aps.org/doi/10.1103/PhysRevLett.127.120502.

[32] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018. DOI: 10.1103/PhysRevA.98.032309. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.98.032309.

[33] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, 2019. DOI: 10.1103/PhysRevA.99.032331. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.99.032331.

[34] Jonas M Kübler, Andrew Arrasmith, Lukasz Cincio, and Patrick J Coles. An adaptive optimizer for measurement-frugal variational algorithms. *Quantum*, 4:263, 2020. DOI: 10.22331/q-2020-05-11-263. URL https://quantum-journal.org/papers/q-2020-05-11-263/.

[35] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum natural gradient. *Quantum*, 4:269, 2020. DOI: 10.22331/q-2020-05-25-269. URL https://quantum-journal.org/papers/q-2020-05-25-269/.

[36] Andrew Arrasmith, Lukasz Cincio, Rolando D Somma, and Patrick J Coles. Operator sampling for shot-frugal optimization in variational algorithms. *arXiv preprint arXiv:2004.06252*, 2020. URL https://arxiv.org/abs/2004.06252.

[37] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):1–6, 2018. DOI: 10.1038/s41467-018-07090-4. URL https://www.nature.com/articles/s41467-018-07090-4.

[38] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1):1–12, 2021. DOI: 10.1038/s41467-021-21728-w. URL https://www.nature.com/articles/s41467-021-21728-w.

[39] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, 12(1):1–11, 2021. DOI: 10.1038/s41467-021-27045-6. URL https://www.nature.com/articles/s41467-021-27045-6.

[40] M. Cerezo and Patrick J Coles. Higher order derivatives of quantum neural networks with barren plateaus. *Quantum Science and Technology*, 6(2):035006, 2021. DOI: 10.1088/2058-9565/abf51a. URL https://iopscience.iop.org/article/10.1088/2058-9565/abf51a.

[41] Kunal Sharma, M. Cerezo, Lukasz Cincio, and Patrick J Coles. Trainability of dissi-

pative perceptron-based quantum neural networks. *Physical Review Letters*, 128(18):180505, 2022. DOI: 10.1103/PhysRevLett.128.180505.

[42] Andrew Arrasmith, M. Cerezo, Piotr Czarnik, Lukasz Cincio, and Patrick J Coles. Effect of barren plateaus on gradient-free optimization. *Quantum*, 5:558, 2021. DOI: 10.22331/q-2021-10-05-558. URL https://quantum-journal.org/papers/q-2021-10-05-558/.

[43] Zoë Holmes, Andrew Arrasmith, Bin Yan, Patrick J. Coles, Andreas Albrecht, and Andrew T Sornborger. Barren plateaus preclude learning scramblers. *Physical Review Letters*, 126(19):190501, 2021. DOI: 10.1103/PhysRevLett.126.190501. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.126.190501.

[44] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. Entanglement-induced barren plateaus. *PRX Quantum*, 2(4):040316, 2021. DOI: 10.1103/PRXQuantum.2.040316. URL https://journals.aps.org/prxquantum/abstract/10.1103/PRXQuantum.2.040316.

[45] Taylor L Patti, Khadijeh Najafi, Xun Gao, and Susanne F Yelin. Entanglement devised barren plateau mitigation. *Physical Review Research*, 3(3):033090, 2021. DOI: 10.1103/PhysRevResearch.3.033090. URL https://par.nsf.gov/servlets/purl/10328786.

[46] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T Sornborger, and Patrick J Coles. Absence of barren plateaus in quantum convolutional neural networks. *Physical Review X*, 11(4):041011, 2021. DOI: 10.1103/PhysRevX.11.041011. URL https://journals.aps.org/prx/abstract/10.1103/PhysRevX.11.041011.

[47] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3:010313, Jan 2022. DOI: 10.1103/PRXQuantum.3.010313. URL https://link.aps.org/doi/10.1103/PRXQuantum.3.010313.

[48] Andrew Arrasmith, Zoë Holmes, Marco Cerezo, and Patrick J Coles. Equivalence of quantum barren plateaus to cost concentration and narrow gorges. *Quantum Science and Technology*, 7(4):045015, 2022. DOI: 10.1088/2058-9565/ac7d06. URL https://iopscience.iop.org/article/10.1088/2058-9565/ac7d06.

[49] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1):1–12, 2021. DOI: 10.1038/s41467-021-21728-w. URL https://www.nature.com/articles/s41467-021-21728-w.

[50] AV Uvarov and Jacob D Biamonte. On barren plateaus and cost function locality in variational quantum algorithms. *Journal of Physics A: Mathematical and Theoretical*, 54(24):245301, 2021. DOI: 10.1088/1751-8121/abfac7. URL https://doi.org/10.1088/1751-8121/abfac7.

[51] Tyler Volkoff and Patrick J Coles. Large gradients via correlation in random parameterized quantum circuits. *Quantum Science and Technology*, 6(2):025008, 2021. DOI: 10.1088/2058-9565/abd89. URL https://iopscience.iop.org/article/10.1088/2058-9565/abd891.

[52] Guillaume Verdon, Michael Broughton, Jarrod R McClean, Kevin J Sung, Ryan Babbush, Zhang Jiang, Hartmut Neven, and Masoud Mohseni. Learning to learn with quantum neural networks via classical neural networks. *arXiv preprint arXiv:1907.05415*, 2019. URL https://arxiv.org/abs/1907.05415.

[53] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, 2019. DOI: 10.22331/q-2019-12-09-214. URL https://quantum-journal.org/papers/q-2019-12-09-214/.

[54] Andrea Skolik, Jarrod R McClean, Masoud Mohseni, Patrick van der Smagt, and Martin Leib. Layerwise learning for quantum neural networks. *Quantum Machine Intelligence*, 3(1):1–11, 2021. DOI: 10.1007/s42484-020-00036-4. URL https://doi.org/10.1007/s42484-020-00036-4.

[55] M Bilkis, M. Cerezo, Guillaume Verdon, Patrick J. Coles, and Lukasz Cincio. A semi-agnostic ansatz with variable structure for quantum machine learning. *arXiv preprint arXiv:2103.06712*, 2021. URL https://arxiv.org/abs/2103.06712.

[56] Alicia B Magann, Christian Arenz, Matthew D Grace, Tak-San Ho, Robert L Kosut, Jarrod R McClean, Herschel A Rabitz, and Mohan Sarovar. From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms. *PRX Quantum*, 2(1):010101, 2021. DOI: https://doi.org/10.1103/PRXQuantum.2.010101. URL https://journals.aps.org/prxquantum/abstract/10.1103/PRXQuantum.2.010101.

[57] Stuart Hadfield, Zhihui Wang, Bryan O'Gorman, Eleanor G Rieffel, Davide Venturelli, and Rupak Biswas. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms*,

12(2):34, 2019. DOI: 10.3390/a12020034. URL https://www.mdpi.com/1999-4893/12/2/34.

[58] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Physical Review A*, 92:042303, Oct 2015. DOI: 10.1103/PhysRevA.92.042303. URL https://link.aps.org/doi/10.1103/PhysRevA.92.042303.

[59] Roeland Wiersema, Cunlu Zhou, Yvette de Sereville, Juan Felipe Carrasquilla, Yong Baek Kim, and Henry Yuen. Exploring entanglement and optimization within the hamiltonian variational ansatz. *PRX Quantum*, 1(2):020319, 2020. DOI: 10.1103/PRXQuantum.1.020319. URL https://journals.aps.org/prxquantum/abstract/10.1103/PRXQuantum.1.020319.

[60] Linghua Zhu, Ho Lun Tang, George S Barron, Nicholas J Mayhall, Edwin Barnes, and Sophia E Economou. Adaptive quantum approximate optimization algorithm for solving combinatorial problems on a quantum computer. *Physical Review Research*, 4(3):033029, 2022. DOI: 10.1103/PhysRevResearch.4.033029. URL https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.4.033029.

[61] Alexandre Choquette, Agustin Di Paolo, Panagiotis Kl Barkoutsos, David Sénéchal, Ivano Tavernelli, and Alexandre Blais. Quantum-optimal-control-inspired ansatz for variational quantum algorithms. *Physical Review Research*, 3(2):023092, 2021. DOI: 10.1103/PhysRevResearch.3.023092. URL https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.3.023092.

[62] Supanut Thanasilp, Samson Wang, Nhat A Nghiem, Patrick J. Coles, and M. Cerezo. Subtleties in the trainability of quantum machine learning models. *arXiv preprint arXiv:2110.14753*, 2021. URL https://arxiv.org/abs/2110.14753.

[63] D. D'Alessandro. *Introduction to Quantum Control and Dynamics*. Chapman & Hall/CRC Applied Mathematics & Nonlinear Science. Taylor & Francis, 2007. ISBN 9781584888840. URL https://books.google.sm/books?id=HbMYmAEACAAJ.

[64] Sukin Sim, Peter D Johnson, and Alán Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies*, 2(12):1900070, 2019. DOI: 10.1002/qute.201900070. URL https://onlinelibrary.wiley.com/doi/full/10.1002/qute.201900070.

[65] Carlton M Caves. Quantum error correction and reversible operations. *Journal of Superconductivity*, 12(6):707–718, 1999. DOI: 10.1023/A:1007720606911. URL https://link.springer.com/article/10.1023/A:1007720606911.

[66] P Rungta, WJ Munro, K Nemoto, P Deuar, Gerard J Milburn, and CM Caves. Qudit entanglement. In *Directions in Quantum Optics*, pages 149–164. Springer, 2001. DOI: 10.1007/3-540-40894-0_14. URL https://link.springer.com/chapter/10.1007%2F3-540-40894-0_14.

[67] Nicholas Hunter-Jones. Unitary designs from statistical mechanics in random quantum circuits. *arXiv preprint arXiv:1905.12053*, 2019. URL https://arxiv.org/abs/1905.12053.

[68] Yoshifumi Nakata, Masato Koashi, and Mio Murao. Generating a state t-design by diagonal quantum circuits. *New Journal of Physics*, 16(5):053043, 2014. DOI: 10.1088/1367-2630/16/5/053043. URL https://iopscience.iop.org/article/10.1088/1367-2630/16/5/053043.

[69] Zhi-Cheng Yang, Armin Rahmani, Alireza Shabani, Hartmut Neven, and Claudio Chamon. Optimizing variational quantum algorithms using pontryagin's minimum principle. *Physical Review X*, 7(2):021027, 2017. DOI: 10.1103/PhysRevX.7.021027. URL https://journals.aps.org/prx/abstract/10.1103/PhysRevX.7.021027.

[70] Oinam Romesh Meitei, Bryan T Gard, George S Barron, David P Pappas, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. Gate-free state preparation for fast variational quantum eigensolver simulations: ctrl-vqe. *arXiv preprint arXiv:2008.04302*, 2020. URL https://arxiv.org/abs/2008.04302.

[71] Juneseo Lee, Alicia B Magann, Herschel A Rabitz, and Christian Arenz. Progress toward favorable landscapes in quantum combinatorial optimization. *Physical Review A*, 104(3):032401, 2021. DOI: 10.1103/PhysRevA.104.032401. URL https://link.aps.org/doi/10.1103/PhysRevA.104.032401.

[72] Jun Li, Xiaodong Yang, Xinhua Peng, and Chang-Pu Sun. Hybrid quantum-classical approach to quantum optimal control. *Physical Review Letters*, 118(15):150503, 2017. DOI: 10.1103/PhysRevLett.118.150503. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.118.150503.

[73] Viswanath Ramakrishna and Herschel Rabitz. Relation between quantum computing and quantum controllability. *Physical Review A*, 54(2):1715, 1996. DOI: 10.1103/PhysRevA.54.1715. URL

https://journals.aps.org/pra/abstract/10.1103/PhysRevA.54.1715.

[74] Seth Lloyd. Quantum approximate optimization is computationally universal. *arXiv preprint arXiv:1812.11075*, 2018. URL https://arxiv.org/abs/1812.11075.

[75] Mauro ES Morales, JD Biamonte, and Zoltán Zimborás. On the universality of the quantum approximate optimization algorithm. *Quantum Information Processing*, 19(9): 1–26, 2020. DOI: 10.1007/s11128-020-02748-9. URL https://link.springer.com/article/10.1007/s11128-020-02748-9.

[76] V Akshay, H Philathong, Mauro ES Morales, and Jacob D Biamonte. Reachability deficits in quantum approximate optimization. *Physical Review Letters*, 124(9):090504, 2020. DOI: 10.1103/PhysRevLett.124.090504. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.124.090504.

[77] Robert Zeier and Thomas Schulte-Herbrüggen. Symmetry principles in quantum systems theory. *Journal of mathematical physics*, 52(11):113510, 2011. DOI: https://doi.org/10.1063/1.3657939. URL https://aip.scitation.org/doi/pdf/10.1063/1.3657939.

[78] Thomas Polack, Haim Suchowski, and David J Tannor. Uncontrollable quantum systems: A classification scheme based on lie subalgebras. *Physical Review A*, 79(5):053403, 2009. DOI: https://doi.org/10.1103/PhysRevA.79.053403. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.79.053403.

[79] Leonardo Banchi, Daniel Burgarth, and Michael J Kastoryano. Driven quantum dynamics: will it blend? *Physical Review X*, 7(4):041015, 2017. DOI: 10.1103/PhysRevX.7.041015. URL https://journals.aps.org/prx/abstract/10.1103/PhysRevX.7.041015.

[80] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, Sep 2017. ISSN 1476-4687. DOI: 10.1038/nature23879. URL https://doi.org/10.1038/nature23879.

[81] Aram W Harrow and Richard A Low. Random quantum circuits are approximate 2-designs. *Communications in Mathematical Physics*, 291 (1):257–302, 2009. DOI: 10.1007/s00220-009-0873-6. URL https://link.springer.com/article/10.1007%2Fs00220-009-0873-6.

[82] Fernando GSL Brandao, Aram W Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346 (2):397–434, 2016. DOI: 10.1007/s00220-016-2706-8. URL https://link.springer.com/article/10.1007%2Fs00220-016-2706-8.

[83] Aram Harrow and Saeed Mehraban. Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates. *arXiv preprint arXiv:1809.06957*, 2018. URL https://arxiv.org/abs/1809.06957.

[84] Andrew Lucas. Ising formulations of many np problems. *Frontiers in Physics*, 2:5, 2014. DOI: 10.3389/fphy.2014.00005. URL https://www.frontiersin.org/articles/10.3389/fphy.2014.00005/full.

[85] Michael Streif and Martin Leib. Training the quantum approximate optimization algorithm without access to a quantum processing unit. *Quantum Science and Technology*, 5(3):034008, 2020. DOI: 10.1088/2058-9565/ab8c2b. URL https://iopscience.iop.org/article/10.1088/2058-9565/ab8c2b.

[86] M. Cerezo, Raúl Rossignoli, N Canosa, and E Ríos. Factorization and criticality in finite xxz systems of arbitrary spin. *Physical Review Letters*, 119(22):220605, 2017. DOI: 10.1103/PhysRevLett.119.220605. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.119.220605.

[87] Xiaoting Wang, Daniel Burgarth, and S Schirmer. Subspace controllability of spin-1 2 chains with symmetries. *Physical Review A*, 94(5):052319, 2016. DOI: https://doi.org/10.1103/PhysRevA.94.052319. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.94.052319.

[88] Benoît Collins and Piotr Śniady. Integration with respect to the haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006. DOI: 10.1007/s00220-006-1554-3. URL https://link.springer.com/article/10.1007%2Fs00220-006-1554-3.

[89] PM Poggi and Diego Ariel Wisniacki. Optimal control of many-body quantum dynamics: Chaos and complexity. *Physical Review A*, 94(3):033406, 2016. DOI: 10.1103/PhysRevA.94.033406. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.94.033406.

[90] Martín Larocca and Diego Wisniacki. Krylov-subspace approach for the efficient control of quantum many-body dynamics. *Physical Review A*, 103(2):023107, 2021. DOI: 10.1103/PhysRevA.103.023107. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.103.023107.

[91] P Erdos and A Renyi. On random graphs

i. *Publ. math. debrecen*, 6(290-297):18, 1959. URL http://snap.stanford.edu/class/cs224w-readings/erdos59random.pdf.

[92] Christian Arenz and Herschel Rabitz. Drawing together control landscape and tomography principles. *Physical Review A*, 102(4):042207, 2020. DOI: 10.1103/PhysRevA.102.042207. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.102.042207.

[93] Zbigniew Puchala and Jaroslaw Adam Miszczak. Symbolic integration with respect to the haar measure on the unitary groups. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 65(1):21–27, 2017. DOI: 10.1515/bpasts-2017-0003. URL http://journals.pan.pl/dlibra/publication/121307/edition/105697/content.

[94] Bryan T Gard, Linghua Zhu, George S Barron, Nicholas J Mayhall, Sophia E Economou, and Edwin Barnes. Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. *npj Quantum Information*, 6(1): 1–9, 2020. DOI: 10.1038/s41534-019-0240-1. URL https://www.nature.com/articles/s41534-019-0240-1.

[95] Christian Kokail, Christine Maier, Rick van Bijnen, Tiff Brydges, Manoj K Joshi, Petar Jurcevic, Christine A Muschik, Pietro Silvi, Rainer Blatt, Christian F Roos, et al. Self-verifying variational quantum simulation of lattice models. *Nature*, 569(7756):355–360, 2019. DOI: 10.1038/s41586-019-1177-4. URL https://www.nature.com/articles/s41586-019-1177-4.

[96] Kunal Sharma, Sumeet Khatri, M. Cerezo, and Patrick J Coles. Noise resilience of variational quantum compiling. *New Journal of Physics*, 22(4):043006, 2020. DOI: 10.1088/1367-2630/ab784c. URL https://iopscience.iop.org/article/10.1088/1367-2630/ab784c.

[97] Nikolay V Tkachenko, James Sud, Yu Zhang, Sergei Tretiak, Petr M Anisimov, Andrew T Arrasmith, Patrick J. Coles, Lukasz Cincio, and Pavel A Dub. Correlation-informed permutation of qubits for reducing ansatz depth in vqe. *PRX Quantum*, 2(2):020337, 2021. DOI: 10.1103/PRXQuantum.2.020337. URL https://journals.aps.org/prxquantum/abstract/10.1103/PRXQuantum.2.020337.

[98] Bobak Toussi Kiani, Seth Lloyd, and Reevu Maity. Learning unitaries by gradient descent. *arXiv preprint arXiv:2001.11897*, 2020. URL https://arxiv.org/abs/2001.11897.

[99] Zhihui Wang, Nicholas C Rubin, Jason M Dominy, and Eleanor G Rieffel. *XY* mixers: Analytical and numerical results for the quantum alternating operator ansatz. *Physical Review A*, 101(1):012320, 2020. DOI: 10.1103/PhysRevA.101.012320. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.101.012320.

[100] Andreas Bärtschi and Stephan Eidenbenz. Grover mixers for qaoa: Shifting complexity from mixer design to state preparation. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 72–82. IEEE, 2020. DOI: 10.1109/QCE49297.2020.00020. URL https://www.computer.org/csdl/proceedings-article/qce/2020/896900a072/1p2VnUCmpYA.

[101] Wen Wei Ho and Timothy H. Hsieh. Efficient variational simulation of non-trivial quantum states. *SciPost Phys.*, 6:29, 2019. DOI: 10.21468/SciPostPhys.6.3.029. URL https://scipost.org/10.21468/SciPostPhys.6.3.029.

[102] Chris Cade, Lana Mineh, Ashley Montanaro, and Stasja Stanisic. Strategies for solving the fermi-hubbard model on near-term quantum computers. *Physical Review B*, 102(23):235122, 2020. DOI: 10.1103/PhysRevB.102.235122. URL https://journals.aps.org/prb/abstract/10.1103/PhysRevB.102.235122.

[103] Chen Zhao and Xiao-Shan Gao. Analyzing the barren plateau phenomenon in training quantum neural networks with the ZX-calculus. *Quantum*, 5:466, June 2021. ISSN 2521-327X. DOI: 10.22331/q-2021-06-04-466. URL https://doi.org/10.22331/q-2021-06-04-466.

[104] Kaining Zhang, Min-Hsiu Hsieh, Liu Liu, and Dacheng Tao. Toward trainability of quantum neural networks. *arXiv preprint arXiv:2011.06258*, 2020. URL https://arxiv.org/abs/2011.06258.

[105] Frederic Sauvage, Sukin Sim, Alexander A Kunitsa, William A Simon, Marta Mauri, and Alejandro Perdomo-Ortiz. Flip: A flexible initializer for arbitrarily-sized parametrized quantum circuits. *arXiv preprint arXiv:2103.08572*, 2021. URL https://arxiv.org/abs/2103.08572.

[106] Yidong Liao, Min-Hsiu Hsieh, and Chris Ferrie. Quantum optimization for training quantum neural networks. *arXiv preprint arXiv:2103.17047*, 2021. URL https://arxiv.org/abs/2103.17047.

[107] Raj Chakrabarti and Herschel Rabitz. Quantum control landscapes. *International Reviews in Physical Chemistry*, 26(4):671–735, 2007. DOI: 10.1080/01442350701633300. URL https://www.tandfonline.com/doi/abs/10.1080/01442350701633300.

[108] Martín Larocca, Pablo M Poggi, and Diego A Wisniacki. Quantum control landscape for a two-level system near the quantum speed

limit. *Journal of Physics A: Mathematical and Theoretical*, 51(38):385305, aug 2018. DOI: 10.1088/1751-8121/aad657. URL https://doi.org/10.1088/1751-8121/aad657.

[109] Martín Larocca, Esteban Calzetta, and Diego A. Wisniacki. Exploiting landscape geometry to enhance quantum optimal control. *Physical Review A*, 101:023410, Feb 2020. DOI: 10.1103/PhysRevA.101.023410. URL https://link.aps.org/doi/10.1103/PhysRevA.101.023410.

[110] Winton G. Brown and Lorenza Viola. Convergence rates for arbitrary statistical moments of random quantum circuits. *Phys. Rev. Lett.*, 104:250501, Jun 2010. DOI: 10.1103/PhysRevLett.104.250501. URL https://link.aps.org/doi/10.1103/PhysRevLett.104.250501.

[111] Domenico D'Alessandro and Jonas T Hartwig. Dynamical decomposition of bilinear control systems subject to symmetries. *Journal of Dynamical and Control Systems*, 27(1):1–30, 2021. DOI: https://doi.org/10.1007/s10883-020-09488-0.

Appendices

In the following appendices we present additional information and derive proofs for the main results in the manuscript. In Appendix A we introduce preliminary notation and definitions that will be relevant for the rest of the appendices. Then, in Appendix B we provide additional details on different widely known ansatzes that are Periodic Structure Ansatz (PSA). In Appendix C we provide a brief review of barren plateaus. Appendices F–M contain the proofs of our main Theorems, Corollaries and Propositions. Finally, in Appendix N we discuss the symmetries in the XXZ and Ising spin models considered in the main text, and in Appendix O we provide additional details on the initial state used for the numerical simulations of the XXZ model.

A Preliminaries

Let us first review some definitions and prior results that will be relevant for the rest of the appendices.

Properties of the Haar measure. Let $\mathcal{U}(d)$ denote the group of $d \times d$ unitary matrices. Let $d\mu_H(V) = d\mu(V)$ be the volume element of the Haar measure, where $V \in \mathcal{U}(d)$. Then, the Haar measure has the following properties: (1) The volume of the Haar measure is finite: $\int_{\mathcal{U}(d)} d\mu(V) < \infty$. (2) The Haar measure is uniquely defined up to a multiplicative constant factor. (3) Let $d\zeta(V)$ be an invariant measure. Then there exists a constant c such that $d\zeta(V) = c \cdot d\mu(V)$. (4) The Haar measure is left- and right-invariant under the action of the unitary group of degree d, i.e., for any integrable function $g(V)$, the following holds:

$$\int_{\mathcal{U}(d)} d\mu(V) g(WV) = \int_{\mathcal{U}(d)} d\mu(V) g(VW) = \int_{\mathcal{U}(d)} d\mu(V) g(V), \tag{36}$$

where $W \in \mathcal{U}(d)$.

Symbolic integration. Let us present formulas that allow for the symbolical integration with respect to the Haar measure on a unitary group [93]. For any $V \in \mathcal{U}(d)$ the following expressions are valid for the first two moments:

$$\begin{aligned}
\int_{\mathcal{U}(d)} d\mu(U) u_{\boldsymbol{ij}} u^*_{\boldsymbol{pk}} &= \frac{\delta_{\boldsymbol{ip}}\delta_{\boldsymbol{jk}}}{d}\,, \\
\int_{\mathcal{U}(d)} d\mu(U) u_{\boldsymbol{i}_1\boldsymbol{j}_1} u_{\boldsymbol{i}_2\boldsymbol{j}_2} u^*_{\boldsymbol{i}'_1\boldsymbol{j}'_1} u^*_{\boldsymbol{i}'_2\boldsymbol{j}'_2} &= \frac{\delta_{i_1 i'_1}\delta_{i_2 i'_2}\delta_{j_1 j'_1}\delta_{j_2 j'_2} + \delta_{i_1 i'_2}\delta_{i_2 i'_1}\delta_{j_1 j'_2}\delta_{j_2 j'_1}}{d^2-1} - \frac{\delta_{i_1 i'_1}\delta_{i_2 i'_2}\delta_{j_1 j'_2}\delta_{j_2 j'_1} + \delta_{i_1 i'_2}\delta_{i_2 i'_1}\delta_{j_1 j'_1}\delta_{j_2 j'_2}}{d(d^2-1)}\,,
\end{aligned} \tag{37}$$

where $u_{\boldsymbol{ij}}$ are the matrix elements of U. Assuming $d = 2^n$, we use the notation $\boldsymbol{i} = (i_1, \dots i_n)$ to denote a bitstring of length n such that $i_1, i_2, \dots, i_n \in \{0, 1\}$.

Useful Identities. We introduce the following identities, which can be derived using Eq. (37) (see [49] for a review):

$$\int_{\mathcal{U}(d)} d\mu(U) \operatorname{Tr}\left[UAU^\dagger B\right] = \frac{\operatorname{Tr}[A]\operatorname{Tr}[B]}{d}, \tag{38}$$

$$\int_{\mathcal{U}(d)} d\mu(U) \operatorname{Tr}\left[UAU^\dagger BUCU^\dagger D\right] = \frac{\operatorname{Tr}[A]\operatorname{Tr}[C]\operatorname{Tr}[BD] + \operatorname{Tr}[AC]\operatorname{Tr}[B]\operatorname{Tr}[D]}{d^2-1} - \frac{\operatorname{Tr}[AC]\operatorname{Tr}[BD] + \operatorname{Tr}[A]\operatorname{Tr}[B]\operatorname{Tr}[C]\operatorname{Tr}[D]}{d\,(d^2-1)}, \tag{39}$$

$$\int_{\mathcal{U}(d)} d\mu(U) \operatorname{Tr}\left[UAU^\dagger B\right] \operatorname{Tr}\left[UCU^\dagger D\right] = \frac{\operatorname{Tr}[A]\operatorname{Tr}[B]\operatorname{Tr}[C]\operatorname{Tr}[D] + \operatorname{Tr}[AC] + \operatorname{Tr}[BD]}{d^2-1} - \frac{\operatorname{Tr}[AC]\operatorname{Tr}[B]\operatorname{Tr}[D] + \operatorname{Tr}[A]\operatorname{Tr}[C]\operatorname{Tr}[BD]}{d\,(d^2-1)}, \tag{40}$$

where A, B, C, and D are linear operators on a d-dimensional Hilbert space.

Integration over parameter space: In the next sections we will derive analytical expressions for the variance of the partial derivatives of cost functions $C(\boldsymbol{\theta})$ over parametrized circuits $U(\boldsymbol{\theta})$. In such derivations, we will have to deal with integration over the parameter space. A key step in the following analysis will be to relate the integration over parameters with integration over the ensemble of unitaries arising from different parameter choices. In this sense, we recall that given a set of parameters $\{\boldsymbol{\theta}\}$ one can obtain an associated set of unitaries generated by the quantum circuit $\{U(\boldsymbol{\theta})\}$. Then, consider the integration of some function $f(U(\boldsymbol{\theta}))$ over $\boldsymbol{\theta}$. Defining $\mathbb{U}$ as the distribution of unitaries generated by $U(\boldsymbol{\theta})$, the following identity holds

$$\int_{\boldsymbol{\theta}} d\boldsymbol{\theta} f(U(\boldsymbol{\theta})) = \int_{\mathbb{U}} dU f(U)\,. \tag{41}$$

In addition, if the distribution of unitaries $\mathbb{U}$ can be shown to converge to a 2-design, the integration over the distibution can be further converted into an integration over the Haar measure

$$\int_{\mathbb{U}} dU \to \int_{\mathcal{U}(d)} d\mu(U) \tag{42}$$

allowing the use of identities (38), (39), and (40).

B Ansatzes

In general, ansatzes for parametrized quantum circuits can be divided into two primary categories: problem-agnostic and problem-inspired ansatzes. In a problem agnostic ansatz one does not have any information about the problem, or its solution, that one can encode in the ansatz. Such is the case for instance in a task of estimating the spectrum of an unknown density operator [29]. On the other hand, problem-inspired ansatzes employ prior information about a given problem or task. For example, for the problem of estimating the ground state energy of a particular Hamiltonian, one can design ansatzes that preserve the symmetry of the problem Hamiltonian [94].

Here we remark that several well known problem-agnostic and problem-inspired ansatzes in the literature are PSAs of the form in Eq. (2). In particular, our framework allows us to study the hardware-efficient ansatz (HEA) [80], quantum alternating operator ansatz (QAOA) [19], Adaptive QAOA [57, 60], Hamiltonian variational ansatz (HVA) [58], and Quantum Optimal Control Ansatz (QAOC) [61].

Below we provide several examples of problem-inspired and problem-agnostic PSA and highlight their advantages in different problems.

Hardware efficient ansatz. The Hardware Efficient Ansatz (HEA) is a problem-agnostic ansatz, which relies on gates native to a quantum hardware. In particular, an ansatz can be designed based on a gate alphabet, which depends on the architecture and the connectivity of a given quantum hardware. This procedure helps in avoiding the overhead associated with transpiling an arbitrary unitary into a sequence of native gates. For example, one can consider native gates, such as single qubit rotations $e^{-i\theta/2Z}e^{-i\gamma/2Y}$ and CNOTs, where Y and Z denote Pauli matrices, and a CNOT between the control qubit i and the target qubit j is given by: $e^{-i\pi/2(|1\rangle\langle 1|_i\otimes(X_j-\mathbb{1}_j))}$. Then an ansatz of the form in (2) can be generated as follows: one layer consists of parametrized single qubit rotations on each qubit, followed by unparametrized CNOTs acting on neighboring qubits.

The HEA has been employed to prepare the ground state of molecules [80], to study Hamiltonians that are similar to the device's interactions [95], and in several other variational quantum algorithms [21, 96, 97, 49]. The HEA is also suitable in the near-term implementations of VQAs due to its low-depth structure which results into a lower-noise circuit in comparison to other ansatze [97, 39].

Quantum alternating operator ansatz. The Quantum Alternating Operator Ansatz (QAOA) is a problem-inspired ansatz that simulates the discretized adiabatic transformations [19]. Consider a goal of preparing the ground state of a problem Hamiltonian H_P. Let H_M denote a mixer Hamiltonian, with corresponding ground state $|\psi\rangle$. Then the QAOA maps $|\psi\rangle$ to the ground state of H_P by sequentially applying the problem unitary $e^{-i\gamma_l H_P}$, followed by the mixer unitary $e^{-i\beta_l H_M}$. Let $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta})$. Then the QAOA is given by $U(\boldsymbol{\theta}) = \prod_{l=1}^{L} e^{-i\beta_l H_M} e^{-i\gamma_l H_P}$, which follows the general form of the ansatz defined in (2). Here, p is the order of the discretized adiabatic transformation and it determines the precision of the solution [19]. The QAOA was originally introduced for finding approximate solutions to combinatorial optimization problems [19]. The QAOA has been generalized as a standalone ansatz [57] and its performance has been investigated in several tasks, including the task of learning a unitary [98]. Moreover, the QAOA has been shown to be computationally universal [74, 75], and the choice of optimal mixer is still an open debate [99, 100].

Adaptive QAOA. As a consequence of the adiabatic theorem, the QAOA should lead to good solutions for high values of p [19]. However, for small values of p, the QAOA is an *ad-hoc* ansatz, which is not necessarily an optimal strategy to approximate the ground state of the problem Hamiltonian. A way to improve such an *ad-hoc* ansatz is to employ a variable mixer instead of a fixed mixer at each layer [57]. Let $\{G_k\}_{k=1}^{q}$ denote a set of mixer Hamiltonians. Then an adaptive QAOA can be defined as follows: $U(\boldsymbol{\theta}) = \prod_{l=1}^{L} e^{-i\beta_l G_l} e^{-i\gamma_l H_p}$, where each G_l can be adaptively picked from $\{G_k\}_{k=1}^{q}$.

One particular adaptive approach was introduced in [60], where at each layer, G_l is picked based on the largest gradient of the cost function among all $\{G_l\}$. Moreover, [60] observed that adaptive entangling mixers can improve performance and reduce the number of parameters and CNOTs to achieve a desired accuracy in comparison to the non-adaptive QAOA. We note that the adaptive QAOA follows the form in

(2) if adaptive mixers are learned up to a fixed layer and then the whole structure is repeated. That is, $U(\boldsymbol{\theta}) = \prod_{m=1}^{p}\prod_{l=1}^{r} e^{-i\beta_{l,m}G_{l,m}}e^{-i\gamma_{l,m}H_P}$, where $G_{l,1}$ are learned adaptively for each $l \in \{1,\ldots,r\}$, and $G_{l,1} = G_{l,m}$ for all $m \in \{1,\ldots,m\}$.

Hamiltonian variational ansatz. The Hamiltonian variational ansatz is another problem-inspired ansatz, which implements time evolution under problem Hamiltonian via Trotterization [58]. It can be understood as a generalization of the QAOA to more than two non-commuting Hamiltonians. Let $H_P = \sum_l H_l$ denote a problem Hamiltonian, such that $[H_l, H_{l'}] \neq 0$. Then the HVA of order p is given by $U(\boldsymbol{\theta}) = \prod_{k=1}^{p}\prod_l e^{-i\theta_{l,k}H_l}$, which is in the form of (2). The HVA has been investigated in studying one- and two-dimensional quantum many-body models [101, 102].

A simple example where the HVA can be employed is the XXZ model to study magnetism. For a one-dimensional chain, the Hamiltonian for the XXZ model is given by $H_{XXZ} = -\sum_{l=1}^{n} X_lX_{l+1}+Y_lY_{l+1}+gZ_lZ_{l+1}$, where g determines the phase of magnetisation. Let $H_A = \sum_{l=1}^{n} A_lA_{l+1}$, where $A \in \{X,Y,Z\}$. Then, one way to parametrize a HVA of order p is as follows: $U(\boldsymbol{\theta}) = \prod_{l=1}^{p} e^{-i\beta_l H_X}e^{-i\gamma_l H_Y}e^{-i\delta_l H_Z}$ for $g = 1$, and where $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta})$. Another way to parametrize a HVA is as follows: $U(\boldsymbol{\theta}) = \prod_{l=1}^{p} e^{-i\beta_l(H_X+H_Y)}e^{-i\delta_l H_Z}$, where we redefined $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\delta})$.

Quantum optimal control ansatz. The HVA discussed above helps constraining the variational search to a relevant symmetric subspace of the the total Hilbert space. In general, this approach might require high values of p to achieve a desired accuracy in approximating the ground state of many-body Hamiltonians. One way to avoid high values of p is to introduce drive terms in addition to the problem Hamiltonian, which break the symmetry of the problem Hamiltonian H_p. This approach falls under the framework of quantum optimal control [61]. In particular, let $\{\hat{H}_k\}$ denote a set of drive terms. Then the update time-dependent Hamiltonian is given by $\bar{H}(t) = H_P + \sum_k c_k(t)H_k$, where drive terms H_k are picked such that $[H_P, H_k] \neq 0$ for all k. Here, $c_k(t)$ are time-dependent control parameters. Let $H_P = \sum_q H_q$ and let $\boldsymbol{\theta} = (\boldsymbol{\gamma},\boldsymbol{\beta})$. Then the Quantum Optimal Control Ansatz (QOCA) of order L is given by $U(\boldsymbol{\theta}) = \prod_{l=1}^{L}\prod_q e^{-i\beta_{l,q}H_q}\prod_k e^{-i\gamma_{l,k}H_k}$, where $\gamma_{l,k}$ denote the discrete drive amplitudes of the control parameter $c_k(t)$.

In general, finding an optimal drive Hamiltonian terms $\{H_k\}$ is a computationally challenging problem. One can employ an adaptive approach to pick drive Hamiltonians from a fixed set of Hamiltonian, similar to the adaptive QAOA [60]. In [61], the QOCA was shown to outperform other ansatze, including the HEA and the HVA for the task of preparation of the ground state of the half-filled Fermi Hubbard model.

C Barren Plateaus

As mentioned in the main text, the barren plateau phenomenon has been recognized as one of the most important challenges to overcome to guarantee the success of VQAs. When a cost function exhibits a barren plateau, its gradients are exponentially suppressed (in average) across the optimization landscape. Consider the following mathematical definition.

Definition 1 (Barren Plateau)**.** *A cost function $C(\boldsymbol{\theta})$ as in Eq. (1) is said to have a barren plateau when training $\theta_\mu \equiv \theta_{pq} \in \boldsymbol{\theta}$, if the cost function partial derivative $\partial C(\boldsymbol{\theta})/\partial\theta_\mu \equiv \partial_\mu C(\boldsymbol{\theta})$ is such that*

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] \leqslant F(n)\,, \quad \textit{with} \quad F(n) \in \mathcal{O}\left(\frac{1}{b^n}\right)\,, \tag{43}$$

for some $b > 1$. Here the variance is taken with respect to the set of parameters $\boldsymbol{\theta}$.

From Chebyshev inequality we know that $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ bounds the probability that $\partial_\mu C(\boldsymbol{\theta})$ diverges from its average (of zero) as $P(|(\partial_\mu C(\boldsymbol{\theta})| > c) \leqslant \mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]/c^2$ for any $c > 0$.

Equation (43) implies that one requires a precision (i.e., a number of shots) that grows exponentially with n to navigate trough the flat landscape and determine a cost minimizing direction when optimizing the cost function. Moreover, as shown in [40, 42], barren plateaus affect both gradient-based and gradient-free methods meaning that simply changing the optimization strategy does not mitigate or solve the barren plateau issues. Since the goal of VQAs is to have computational complexities that scale polynomial with n, such exponential scaling in the required precision destroys the hope of achieving a computational advantage with the VQA over classical methods (which usually scale exponentially with n).

The first result for barren plateaus was obtained in [37], where it was shown that deep unstructured ansatz that form 2-designs have barren plateaus. This phenomenon was then generalized to layered Hardware Efficient Ansatzes in [49] were it was proven that the locality of the cost function is connected to the existence of barren plateaus. That is, global cost functions (i.e., cost functions where O in (1) acts non-trivially in all qubits) exhibit

barren plateaus even for shallow depths, whereas local cost functions (i.e., cost functions where O in (1) acts non-trivially in a small number of neighboring qubits) do not exhibit barren plateaus for short-depth ansatzes.

The barren plateaus phenomenon has also been studied in the context of quantum neural networks [41, 46, 103], and to the problem of learning scramblers [43]. In addition, it has been shown that circuits that generate large amounts of entanglement [41, 45, 44] are prone to suffer from barren plateaus. To circumvent or mitigate the effect of barren plateaus, several strategies have been developed [52, 51, 54, 53, 46, 104, 28, 29, 105, 106].

D Quantum Optimal Control

Here we recall for convenience that in a standard QOC setting one is interested in controlling the dynamical evolution of a quantum state $|\psi\rangle$ in ad-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^d$ (where $d = 2^n$) [63]. Here, the system dynamics are determined by a Hamiltonian

$$H(\{f_k(t)\}) = H_0 + \sum_{k=1}^{K} f_k(t) H_k \tag{44}$$

that is tunable through some time-dependent control fields functions $\{f_k(t)\}$.

At its core, the problem in QOC is to determine how to shape the control fields such that the system evolves in a desired manner. A specific set of optimal fields is usually constructed by imposing a parametrization on the functions and applying standard numerical optimization routines. The success of such optimization process depends on the structure of the underling optimization spaces, the so-called quantum control landscapes [107, 108, 109].

For instance, a common choice is to consider piece-wise constant fields where the protocol duration T is divided in L intervals $\Delta t_j = t_j - t_{j-1}$ (such that $T = \sum_{j=1}^{L} \Delta t_j$) at each of which the fields take a constant value, e.g., $f_k(t) = f_{k,j}$ if $t_{j-1} < t < t_j$. In this case, the propagator factorizes into a product of individual sub-propagators, each of which is generated by a constant Hamiltonian and thus leads to the simple matrix exponential form

$$U(\{f_k(t)\}) = \prod_{l=1}^{L} e^{-i\tilde{H}_l \Delta t_l}, \quad \tilde{H}_l = H_0 + \sum_{k=1}^{K} f_{l,k} H_k \,. \tag{45}$$

By Trotterizing Eq. (45) one finds

$$U(\{f_k(t)\}) \approx \prod_{l=1}^{L} \prod_{k=0}^{K} e^{-iH_k \Delta t_l f_{l,k}} \,, \tag{46}$$

where $f_{l,0} = 1$ for all l. Note that (46) is a PSA of the form of Eq. (2), through the identification

$$\theta_{l,0} = \Delta t_0 \,, \qquad \theta_{l,k} = \Delta t_l f_{l,k}, \quad k \geqslant 1 \,. \tag{47}$$

We remark that in the limit $\Delta t_l \to 0$, Eq. (46) becomes exact. In the general case, the exact and Trotterized ansatzes approximately coincide, and a nontrivial correction of Eq. (47) is needed to make the correspondence exact. In any case, Eq. (47) allows us to henceforth use the notation $U(\{f_k(t)\}) = U(\boldsymbol{\theta})$ and indicate with $\boldsymbol{\theta}$ the trainable parameters in a QOC setting.

E Dynamical Lie Algebra computation

First, let us recall that a Lie algebra is a vector space $\mathfrak{g}$ together with an operation $[\cdot,\cdot] : \mathfrak{g} \to \mathfrak{g}$ called *Lie bracket* that is bilinear, *alternating* (the output is zero if the inputs are linearly dependent) and satisfies the identity $[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0$, known as the Jacobi identity. Lie algebras are vector spaces that are *closed* under such Lie bracket, i.e. $[x,y] \in \mathfrak{g}$ for all $x, y \in \mathfrak{g}$. Note that this operation is not necessarily associative, i.e. $[x[y,z] \neq [[x,y],z]$. In fact, this is precisely what the Jacobi identity captures: how the order of evaluation affects the result of the operation.

In our quantum context, Lie algebras manifest as matrix Lie algebras. For example, the space of quantum observables $\mathfrak{u}(d)$ is a subspace of the vector space of $d \times d$ complex matrices that is closed under matrix commutator (playing the role of a Lie bracket). More generally, we will encounter ourselves with Lie algebras that form Lie subalgebras of $\mathfrak{u}(d)$. A subalgebra is a subspace of an algebra that is itself closed under the Lie bracket. For example, in a n qubit quantum system, the subspace $\Omega = \mathrm{span}\{X, Y, Z\}$, where $X = \sum_i X_i$,

Algorithm 1: Basis for the Dynamical Lie Algebra (DLA).

Input: Set of generators $\mathcal{G}$ of the ansatz.
Output: Basis S of the algebra $\mathfrak{g}$ obtained from $\mathcal{G}$.
Init: $S \leftarrow \mathcal{G}$, $S_{prev} \leftarrow \mathcal{G}$, $S_{new} \leftarrow \{\}$, $C \leftarrow \mathbb{1}$, $\texttt{new} \leftarrow 1$.

```
while new>0 do
    S_new ← {}.
    for h_0 ∈ G do
        for h ∈ S_prev do
            C ← [h_0, h]
            if S ∪ {C} is linearly independent
             then
                S ← S ∪ {C}.
                S_new ← S_new ∪ {C}.
    new ← |S_new|.
    S_prev ← S_new.
return S
```

$Y = \sum_i Y_i$ and $Z = \sum_i Z_i$ is closed under commutation and thus constitutes a 3-dimensional subalgebra of the 4^n-dimensional operator space.

Specifically, we will be interested in the so-called *dynamical Lie algebra* (DLA). This Lie algebra is the subspace of $\mathfrak{u}(d)$ generated by the Lie closure of the generators of a paramterized quantum circuit (see Definition 3). In general, computing the DLA is a highly nontrivial task. One possible approach to the DLA is direct construction, i.e. start with a set of generators defining a subspace of operator space (but not a subalgebra), and start commuting them, finding new elements until one obtains a basis of the DLA (see Algorithm 1). The complexity of such approach is, in general, $\mathcal{O}(\text{poly(d)})$ with $d = 2^n$, that is, exponential in n the number of qubits. For example, a naive approach (representing operators as dense $d \times d$ matrices) yields roughly $\mathcal{O}(d^2 d^6)$, since, in general, one has to check linear independence $\mathcal{O}(d^2)$, for example by implementing LU or QR decompositions (whose cost is $\mathcal{O}(N^3)$ for $N \times N$ matrices) on such $d \times d$ matrices. Although such complexity can be reduced, for example, using more intelligent representations of operators, in essence direct construction is attempting to build a basis for a subalgebra of $\mathfrak{su}(d^2)$, i.e. a basis with potentially as many as d^2 elements, and therefore it cannot generally avoid exponentiality.

Despite being exponential, direct construction of the DLA (either numerically or analitically) on small system sizes can constitute a remarkably useful tool to later extrapolate or prove the scaling (e.g. by induction) of the DLA beyond those small 'afforable' system sizes. Moreover, note that in many cases one may only be interested in checking whether the dimension of the DLA is above a certain threshold, a task with complexity linear in the size of such threshold. Of course, this is neglecting the complexity of computing new DLA elements, which, as mentioned above, can be substantially diminished by choosing efficient representations for those operators.

F Proof of Theorem 1: Convergence of controllable systems to 2-designs

In the following we provide a proof for Theorem 1, which we recall for convenience.

Theorem 1. *Consider a controllable system. Then, the PSA $U(\boldsymbol{\theta})$ will form an ε-approximate 2-design, i.e. $\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty = \varepsilon$ with $\epsilon > 0$, when the number of layers L in the circuit is*

$$L = \frac{\log(1/\varepsilon)}{\log\left(1/\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty\right)}. \tag{48}$$

Here $\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty$ denotes the expressibility of a single layer $U_1(\boldsymbol{\theta}_1)$ of the ansatz according to Eqs. (2) and (6).

Proof. To study the convergence of the PSA $U(\boldsymbol{\theta})$ to an approximate 2-design we employ the tools of Harmonic analysis. The following arguments are based on Ref. [110]. This is similar to using Fourier analysis to study the convergence of a probability distribution on a real line to the normal distribution, which is also known as the central limit theorem.

The second moment operator corresponding to the distribution $\mathbb{U}$ over unitaries U can be defined as follows

$$M_U^{(2)} = \int_{\mathbb{U}} dU U \otimes U \otimes U^* \otimes U^* \,. \tag{49}$$

Let $M_{U_H}^{(2)}$ denote the second moment operator corresponding to the Haar distribution. Our goal is to evaluate the difference between $M_U^{(2)}$ and $M_{U_H}^{(2)}$. Let us review some properties of $M_{U_H}^{(2)}$ before we calculate this distance operator. An important property of $M_{U_H}^{(2)}$ is that it is a projector onto a two-dimensional subspace, that is, $M_{U_H}^{(2)}$ has eigenvalues 0 or 1. We show this by noting that the following equations hold:

$$\begin{aligned}
(M_{U_H}^{(2)})^2 &= \left(\int_{\mathcal{U}(d)} d\mu(U) U \otimes U \otimes U^* \otimes U^*\right)\left(\int_{\mathcal{U}(d)} d\mu(V) V \otimes V \otimes V^* \otimes V^*\right) \\
&= \int_{\mathcal{U}(d)} d\mu(U) \int_{\mathcal{U}(d)} d\mu(V) \, (UV \otimes UV \otimes (UV)^* \otimes (UV)^*) \qquad (50) \\
&= \int_{\mathcal{U}(d)} dU dW \, (W \otimes W \otimes W^* \otimes W^*) \\
&= \int_{\mathcal{U}(d)} dW W \otimes W \otimes W^* \otimes W^* \\
&= M_{U_H}^{(2)} \,, \qquad (51)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Tr}\left[M_{U_H}^{(2)}\right] &= \mathrm{Tr}\left[\int_{\mathcal{U}(d)} d\mu(U) U \otimes U \otimes U^* \otimes U^*\right] \\
&= \sum_{i,j,k,l=1}^{d} \int_{\mathcal{U}(d)} d\mu(U) U_{i,i} U_{j,j} U^*_{k,k} U^*_{l,l} \qquad (52) \\
&= \frac{2}{d^2-1} \sum_{i,j=1}^{d} 1 - \frac{2}{d(d^2-1)} \sum_{i=1}^{d} 1 \\
&= 2 \,. \qquad (53)
\end{aligned}$$

In Eq. (50), we used the left invariance of the Haar measure, and in (52) we used the Weingarten function to explicitly evaluate the integral. The first property puts in evidence that $M_{U_H}^{(2)}$ is a projector and the second property shows that the eigenspace with eigenvalue 1 is a two-dimensional subspace.

Let $V = U_L \cdots U_2 U_1$, be an L-layered PSA, where each unitary U_j is sampled from the same distribution $d\mu = P(U)dU$. Then, the probability distribution and moment operator of V are respectively given by

$$\mu_L[V] = \int_{\mathcal{U}(d)} dU_1 \int_{\mathcal{U}(d)} dU_2 \cdots \int_{\mathcal{U}(d)} dU_L P(U_1)P(U_2)\cdots P(U_L)\delta(V - U_L \cdots U_2 U_1) \,, \tag{54}$$

$$\begin{aligned}
M_V^{(2)} &= \int_{\mathcal{U}(d)} d\mu_L[V] \quad V \otimes V \otimes V^* \otimes V^* \\
&= \int_{\mathcal{U}(d)} dV \int_{\mathcal{U}(d)} dU_1 \int_{\mathcal{U}(d)} dU_2 \cdots \int_{\mathcal{U}(d)} dU_L P(U_1)P(U_2)\cdots P(U_L)\delta(V - U_L \cdots U_2 U_1) \quad V \otimes V \otimes V^* \otimes V^* \\
&= \int_{\mathcal{U}(d)} dU_1 \int_{\mathcal{U}(d)} dU_2 \cdots \int_{\mathcal{U}(d)} dU_L P(U_1)P(U_2)\cdots P(U_L) \quad (\prod_{l=1}^{L} U_l) \otimes (\prod_{l=1}^{L} U_l) \otimes (\prod_{l=1}^{L} U_l^*) \otimes (\prod_{l=1}^{L} U_l^*) \\
&= \prod_{l=1}^{L} \int_{\mathcal{U}(d)} dU_l P(U_l) U_l \otimes U_l \otimes U_l^* \otimes U_l^* \\
&= \left(M_{U_1}^{(2)}\right)^L \,. \qquad (55)
\end{aligned}$$

Equation (55) shows that the moment operator of an L-layered ansatz is equal to the L-th power of the moment operator of a single layer. We can also calculate this formally. In our case each U_l is given by $U_l = \prod_{k=1}^{K} e^{-i\theta_{lk} H_k}$,

where $\mathcal{G} = \{H_k\}_{k=1}^K$ are the set of generators, and where the θ_{lk} are sampled from the uniform distribution. Then, let us note that

$$\begin{aligned}
M_{U_1}^{(2)} &= \left(\frac{1}{2\pi}\right)^n \int \prod_{k=1}^K d\theta_k \prod_{k=1}^K e^{-i\theta_k H_k} \otimes \prod_{k=1}^K e^{-i\theta_k H_k} \otimes \prod_{k=1}^K e^{i\theta_k H_k^*} \otimes \prod_{k=1}^K e^{i\theta_k H_k^*} \\
&= \left(\frac{1}{2\pi}\right)^n \prod_{k=1}^K \int d\theta_k e^{-i\theta_k H_k} \otimes e^{-i\theta_k H_k} \otimes e^{i\theta_k H_k^*} \otimes e^{i\theta_k H_k^*} \\
&= \left(\frac{1}{2\pi}\right)^n \prod_{k=1}^K \left(V_k^{\otimes 2} \otimes V_k^{*\otimes 2}\right) \int d\theta_k \sum_{\substack{l_{1,k}, l_{2,k} \\ l_{3,k}, l_{4,k}}} e^{-i\theta_k (l_{1,k} + l_{2,k} - l_{3,k} - l_{4,k})} |l_{1,k}, l_{2,k}, l_{2,k}, l_{4,k}\rangle\langle l_{1,k}, l_{2,k}, l_{2,k}, l_{4,k}| \left(W_k^{\otimes 2} \otimes W_k^{*\otimes 2}\right)^\dagger \\
&= \prod_{k=1}^K \left(W_k^{\otimes 2} \otimes W_k^{*\otimes 2}\right) \sum_{\substack{l_{1,k}, l_{2,k} \\ l_{3,k}, l_{4,k}}} \delta(l_{1,k} + l_{2,k} - l_{3,k} - l_{4,k}) |l_{1,k}, l_{2,k}, l_{3,k}, l_{4,k}\rangle\langle l_{1,k}, l_{2,k}, l_{3,k}, l_{4,k}| \left(W_k^{\otimes 2} \otimes W_k^{*\otimes 2}\right)^\dagger .
\end{aligned} \tag{56}$$

Here $H_k|l_{i,k}\rangle = l_{i,k}|l_{i,k}\rangle$, and W_k is the unitary matrix that diagonalizes H_k. To calculate the distance to a 2-design we need to prove some properties of eigenvalues and eigenvectors of $M_{U_1}^{(2)}$. Let $|\phi\rangle$ be an eigenvector of $M_{U_1}^{(2)}$ with associated eigenvalue λ. Then,

$$|\langle\phi|M_{U_1}^{(2)}|\phi\rangle| = |\lambda| = \left| \int_{\mathcal{U}(d)} dU_1 \; P(U_1) \; \langle\phi|U_1 \otimes U_1 \otimes U_1^* \otimes U_1^*|\phi\rangle \right| \tag{57}$$

$$\leqslant \int_{\mathcal{U}(d)} dU_1 \; P(U_1) \; \left|\langle\phi|U_1 \otimes U_1 \otimes U_1^* \otimes U_1^*|\phi\rangle\right| \tag{58}$$

$$\leqslant 1 . \tag{59}$$

The equality holds if and only if $|\phi\rangle$ is an eigenvector of $U \otimes U_1 \otimes U_1^* \otimes U_1^* \; \forall U_1$, such that $P(U_1) \neq 0$. For the specific case of Haar measure, this means that $|\phi\rangle$ is an eigenvector of $U \otimes U \otimes U^* \otimes U^* \; \forall U \in \mathcal{SU}(d)$. We already showed that there are two such eigenvectors for Haar measure with eigenvalue 1.

Now, using the following argument we hope to show that those two are also the only eigenvectors of $M_{U_1}^{(2)}$ with eigenvalue 1, given that the set $\mathcal{G}$ is controllable. Let $|\phi\rangle$ be an eigenvector of $M_{U_1}^{(2)}$ with eigenvalue 1. One can now also see that $|\phi\rangle$ is also an eigenvector of $M_V^{(2)}$ (since $M_V^{(2)} = (M_{U_1}^{(2)})^L$) $\forall L \in \mathbb{Z}^+$. Then, $\mathcal{G}$ being a controllable set implies that for all $U \in \mathcal{SU}(d)$, there exists an L for which $V = \prod_{l=1}^L \prod_{k=1}^K e^{-i\theta_{lk} H_k} = U$ so that $P(U) \neq 0$. That, is one can obtain any unitary in $\mathcal{U}(d)$ by tuning the parameters in V. It also implies that $|\phi\rangle$ has to be an eigenvector of $U \otimes U \otimes U^* \otimes U^*$ with eigenvalue 1. But this means that $|\phi\rangle$ is an eigenvector of $U \otimes U \otimes U^* \otimes U^*$ with eigenvalue 1 $\forall U \in \mathcal{SU}(d)$. So $|\phi\rangle$ is also an eigenvector of $M_2[\mu_H]$ with eigenvalue 1, and there are two such eigenvectors. Let us call them $|\phi_1\rangle$, and $|\phi_2\rangle$. Hence, we have

$$M_{U_H}^{(2)} = |\phi_1\rangle\langle\phi_1| + |\phi_2\rangle\langle\phi_2| \tag{60}$$

$$M_{U_1}^{(2)} = |\phi_1\rangle\langle\phi_1| + |\phi_2\rangle\langle\phi_2| + \sum_{i=1}^{d^4-2} \lambda_i |\psi_i\rangle\langle\psi_i| \tag{61}$$

$$M_V^{(2)} = |\phi_1\rangle\langle\phi_1| + |\phi_2\rangle\langle\phi_2| + \sum_{i=1}^{d^4-2} \lambda_i^L |\psi_i\rangle\langle\psi_i| . \tag{62}$$

Here $\{\lambda_i\}$ is the set of the remaining eigenvalues such that $0 \leqslant |\lambda_i| < 1$. Let $\lambda_{\max}$ be the eigenvalue with maximum modulus. Thus, we can now show that $\|\mathcal{A}_{U(\boldsymbol{\theta})}^{(2)}\|_\infty = \|M_{U_H}^{(2)} - M_V^{(2)}\|_\infty = \|\sum_{i=1}^{N^4-2} \lambda_i^k |\psi_i\rangle\langle\psi_i|\|_\infty = |\lambda_{\max}^L|$. Then, recalling that $|\lambda_{\max}| = \|\mathcal{A}_{U_1(\boldsymbol{\theta})}^{(2)}\|_\infty$ is also the expressibility of one layer, one can find that

$$\|\mathcal{A}_{U(\boldsymbol{\theta})}^{(2)}\|_\infty = \left(\|\mathcal{A}_{U_1(\boldsymbol{\theta})}^{(2)}\|_\infty\right)^L . \tag{63}$$

Solving for L and denoting $\varepsilon = \|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty$ leads to

$$L = \frac{\log(1/\varepsilon)}{\log\left(1/\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty\right)} \,. \tag{64}$$

□

G Proof of Corollary 1: Rate of convergence of controllable systems to 2-designs

Here we prove Corollary 1.

Corollary 1. *Let the single layer expressibility of a controllable system be $\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty = 1 - \delta(n)$, with $\delta(n)$ being at most polynomially vanishing with n, i.e., with $\delta(n) \in \Omega(1/\operatorname{poly}(n))$. Then, if $L(n) \in \Omega(n/\delta(n))$, $U(\boldsymbol{\theta})$ will be no worse than an $\varepsilon(n)$-approximate 2-design (i.e., $\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty \leqslant \varepsilon(n)$) with $\varepsilon(n) \in \mathcal{O}(1/2^n)$, where we have added the n-dependence in L and ε for clarity.*

Proof. Let us recall from the main text, and from Eq. (63), that

$$\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty = \left(\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty\right)^L \,. \tag{65}$$

Replacing $\|\mathcal{A}^{(2)}_{U_1(\boldsymbol{\theta})}\|_\infty = 1 - \delta(n)$, one finds

$$\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty = (1 - \delta(n))^L \,. \tag{66}$$

Then, assuming that $L(n) \in \Omega(n/\delta(n))$ we have, by definition, that there exists $c \geqslant 0$ and n_0 such that $L(n) \geqslant cn/\delta(n)$ for all $n \geqslant n_0$. From the previous, we find that for $n \geqslant n_0$

$$\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty = (1 - \delta(n))^L \leqslant (1 - \delta(n))^{\frac{cn}{\delta(n)}} \,, \tag{67}$$

where we have used the fact that $(1 - \delta(n)) \leqslant 1$. Recalling that $\delta(n) \in \Omega(1/\operatorname{poly}(n))$, i.e., $\delta(n)$ vanishes no faster than a polynomial function of n, we find to first order that

$$(1 - \delta(n))^{\frac{c'n}{\delta(n)}} \approx e^{-n + \mathcal{O}(\frac{1}{n})} \,. \tag{68}$$

Then, defining $\varepsilon(n) = (1 - \delta(n))^{\frac{cn}{\delta(n)}}$ we prove the result in Corollary 1 as

$$\|\mathcal{A}^{(2)}_{U(\boldsymbol{\theta})}\|_\infty \leqslant \varepsilon(n) \,, \quad \text{with} \quad \varepsilon(n) \in \mathcal{O}(\frac{1}{2^n}). \tag{69}$$

□

H Proof of Proposition 1: Controllability leads to barren plateaus

Let us now provide a proof for Proposition 1.

Proposition 1 (Controllable)**.** *There exists a scaling of the depth for which controllable systems form ε-approximate 2-designs with $\varepsilon \in \mathcal{O}(1/2^n)$, and hence the system exhibits a barren plateau according to Definition 1.*

Proof. Let us start by noting that for all controllable systems one can form a ε-approximate 2-designs with $\varepsilon \in \mathcal{O}(1/2^n)$ with a depth scaling obtained from Eq. (64). Then, let us recall that we have defined the expressibility superoperator as

$$\mathcal{A}^{(t)}_{U(\boldsymbol{\theta})} = M^{(t)}_{U_H} - M^{(t)}_{U(\boldsymbol{\theta})} \,. \tag{70}$$

with its ordinary action given by

$$\mathcal{A}^{(t)}_{U(\boldsymbol{\theta})}(\cdot) = M^{(t)}_{U_H}(\cdot) - M^{(t)}_{U(\boldsymbol{\theta})}(\cdot) \,. \tag{71}$$

We then have that for any quantum state ρ

$$\|\mathcal{A}^{(t)}_{U(\boldsymbol{\theta})}(\rho)\|_\infty \leqslant \|\mathcal{A}^{(t)}_{U(\boldsymbol{\theta})}\|_\infty \,, \tag{72}$$

which follows from the normalization of ρ. Hence, if $\|\mathcal{A}^{(t)}_{U(\boldsymbol{\theta})}\|_\infty \leqslant \varepsilon$ we find from Eq. (72) that for any quantum state ρ, the inequality $\|\mathcal{A}^{(t)}_{U(\boldsymbol{\theta})}(\rho)\|_\infty \leqslant \varepsilon$ holds, which is precisely the definition of an ε-approximate state 2-design. Finally, we can use the results from [37], which imply that since $\varepsilon \in \mathcal{O}(1/2^n)$, then the variance of the cost function partial derivative is given by

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] = \widehat{F}(n) \,, \quad \text{with} \quad \widehat{F}(n) \in \mathcal{O}(1/2^n) \,. \tag{73}$$

Thus, the cost function exhibit a barren plateau.

□

I Proof of Proposition 2: Controllability of the HEA and the Spin Glass model

Let us now prove Proposition 2.

Proposition 2. *The following two sets of generators generate full rank DLAs, and concomitantly lead to controllable systems:*

- $\mathcal{G}_{\mathrm{HEA}} = \left\{X_i, Y_i\right\}_{i=1}^{n} \bigcup \left\{\sum_{i=1}^{n-1} Z_i Z_{i+1}\right\}$,
- $\mathcal{G}_{\mathrm{SG}} = \left\{\sum_{i=1}^{n} X_i, \sum_{i<j} \left(h_i Z_i + J_{ij} Z_i Z_j\right)\right\}$, *with* $h_i, J_{ij} \in \mathbb{R}$ *sampled from a Gaussian distribution.*

In the following we find that by repeated nested commutators between the elements of the sets $\mathcal{G}$ in Proposition 2. one can obtain all $2^{2n}-1$ Pauli strings, and hence, that the DLA $\mathfrak{g}$ obtained is full rank.

Proof. We divide the proof into HEA and GS models.

I.0.1 Generators of the Hardware Efficient Ansatz (HEA)

We first start with the set $\mathcal{G}_{\mathrm{HEA}}$ corresponding to the set of generators of a HEA.

$$\mathcal{G}_{\mathrm{HEA}} = \left\{X_i, Y_i\right\}_{i=1}^{n} \bigcup \left\{\sum_{i=1}^{n-1} Z_i Z_{i+1}\right\} \,. \tag{74}$$

First, let us note that from the commutation of X_i with Y_i, we get every Z_i. Meaning that one can already obtain all single qubit Pauli operators. Then, the commutation of X_i and Y_i with $\sum_{i=1}^{n-1} Z_i Z_{i+1}$, respectively, gives

$$A_i = Y_i(Z_{i+1} + Z_{i-1}) \,, \quad \text{and} \quad B_i = X_i(Z_{i+1} + Z_{i-1}) \,. \tag{75}$$

It then follows that the commutator of X_{i+1} with B_i is

$$C_i = X_i Y_{i+1}, \ \forall i \in \{1, \ldots, n-1\} \,. \tag{76}$$

Then by computing the commutators of $\{X_i\}_{i=1}^n$, $\{Y_i\}_{i=1}^n$, and $\{Z_i\}_{i=1}^n$ with $\{C_i\}_{i=1}^{n-1}$ we get all nearest-neighbour two-body Pauli operators

$$D_i = \{X_i X_{i+1}, Y_i Y_{i+1}, Z_i Z_{i+1}, X_i Y_{i+1}, Y_i X_{i+1}, X_i Z_{i+1}, Z_i X_{i+1}, Y_i Z_{i+1}, Z_i Y_{i+1}\}, \ \forall i \in \{1, \ldots, n-1\} \,. \tag{77}$$

Similarly, it can be readily verified that the commutators between $\{X_i\}_{i=1}^n$, $\{Y_i\}_{i=1}^n$, $\{Z_i\}_{i=1}^n$, and operators in $\{D_i\}_{i=1}^{n-1}$ yield all "nearest-neighbour" three-body Pauli operators.

Now, let us show that $\mathfrak{g}$ also contains the remaining non-nearest-neighbour two body operators. Consider the commutator between the three-body nearest-neighbour operators

$$[M_i N_{i+1} O_{i+2}, P_i N_{i+1} Q_{i+2}] = [M_i, P_i][O_{i+2}, Q_{i+2}] \tag{78}$$

where $N, M, O, P, Q \in \{X, Y, Z\}$. Clearly, the different choices of M, O, P and Q will generate all next-nearest-neighbour two-body operators. Iterating this procedure we obtain all $9\binom{n}{2}$ two-body terms. Then, once we have all two-bodies we can use one-bodies $\{X_i\}_{i=1}^n$, $\{Y_i\}_{i=1}^n$ and $\{Z_i\}_{i=1}^n$ to get all three-bodies. Three-bodies with one-bodies will give four-bodies, and so on. We will get all n-body operators. Thus the DLA of the HEA is full rank, which implies that the HEA is a controllable ansatz.

I.0.2 Generators of the Spin Glass (SG)

The set of generators for a spin glass system is given by

$$\mathcal{G}_{\text{SG}} = \left\{ \sum_{i=1}^{n} X_i, \sum_{i<j} (h_i Z_i + J_{ij} Z_i Z_j) \right\} , \tag{79}$$

with $h_i, J_{ij} \in \mathbb{R}$. For convenience, we define the following two operators

$$H_p = \sum_{ij} h_i Z_i + J_{ij} Z_i Z_j, \quad H_m = \sum_i X_i \ . \tag{80}$$

The commutator of H_p and H_m gives

$$A_0 = [H_p, H_m] = \sum_{ij} h_i Y_i + J_{ij} Y_i Z_j \ . \tag{81}$$

We then compute the commutator of A_0 and H_m:

$$[H_m, A_0] = -H_p + \sum_{ij} J_{ij} Y_i Y_j \ . \tag{82}$$

Combining H_p and $[H_m, A_0]$ gives

$$A_1 = \sum_{ij} J_{ij} Y_i Y_j \ . \tag{83}$$

We now compute the commutator between A_1 and H_m as follows:

$$A_3 := [A_1, H_m] = \sum_{rs} J_{rs} Z_r Y_s \ . \tag{84}$$

Combining A_0 and A_3, we get

$$A_4 := \sum_i h_i Y_i \ . \tag{85}$$

Similarly, combining H_m and A_4 gives

$$A_5 := \sum_i h_i Z_i \ . \tag{86}$$

Finally, combining A_5 and H_p gives

$$A_6 := \sum_{ij} J_{ij} Z_i Z_j \ . \tag{87}$$

From the commutator of A_4 and A_5 we get

$$A_7 := [A_4, A_5] = \sum_i h_i^2 X_i \ . \tag{88}$$

Moreover, the commutators of A_7 with A_4 and A_5 lead to

$$A_8 := [A_7, A_4] = \sum_i h_i^3 Z_i \ , \quad A_9 := [A_5, A_7] = \sum_i h_i^3 Y_i \ . \tag{89}$$

By repeating this procedure, we get that the set

$$\bar{\mathcal{S}} = \{ \sum_i h_i^{2m} X_i, \ \sum_i h_i^m Y_i \ , \sum_i h_i^m Z_i \}_{m=1}^{n} \tag{90}$$

also belongs to the Lie algebra. Now, because the h_i's are sampled from a Gaussian distribution, we can safely assume them to be non-zero and different from each other. Then, using a Vandermonde determinant type of argument one can show that the $3n$ elements in $\bar{\mathcal{S}}$ are linearly independent and span the same subspace as $\mathcal{S} = \{X_i, Y_i, Z_i\}_{i=1}^n$. Thus $\mathcal{S}$ belongs to $\mathfrak{g}_{\text{SG}}$. Combining this with A_6, we essentially get the generators of $\mathcal{G}_{\text{HEA}}$ and hence one can again generate all n-body Pauli operators. Thus the DLA of the spin-glass system is also of full rank, which implies that the spin-glass system is controllable.

□

J Proof of Theorem 2: Variance in subspace controllable systems

In the following, we provide a proof for Theorem 2 by explicitly computing the variance of the cost function partial derivative in a subspace controllable setting. Consider a set of generators that share a symmetry (for simplicity we assume only one symmetry, although generalization to multiple symmetries is straightforward), i.e., there is a Hermitian operator Σ such that $[\Sigma, g] = 0 \quad \forall g \in \mathfrak{g}$. Assuming Σ has N distinct eigenvalues, the DLA has the form $\mathfrak{g} = \bigoplus_{m=1}^{N} \mathfrak{g}_m$. This imposes a partition of $\mathcal{H} = \bigoplus_{m=1}^{N} \mathcal{H}_m$ where each subspace $\mathcal{H}_m$ of dimension d_m is invariant under $\mathfrak{g}$.

Let us introduce some notation. Consider the $d \times d_m$ matrix that results from horizontally stacking the eigenvectors of Σ associated with the m-th eigenvalue (of degeneracy g_m)

$$Q_m^\dagger = \begin{bmatrix} \vdots & \vdots & \vdots \\ |v_1\rangle, & |v_2\rangle, & , |v_{g_m}\rangle \\ \vdots & \vdots & \vdots \end{bmatrix}, \tag{91}$$

such that Q_m maps vectors from $\mathcal{H}$ to $\mathcal{H}_m$. These satisfy

$$Q_m Q_n^\dagger = \mathbb{1}_{d_m} \delta_{mn}, \quad Q_m^\dagger Q_m = \mathbb{P}_m \,, \tag{92}$$

where $\mathbb{P}_m$ are projectors onto the subspace, such that $\sum_{m=1}^{N} \mathbb{P}_m = \mathbb{1}_d$. Let us now use the notation

$$|\psi\rangle^{(m)} = Q_m |\psi\rangle, \quad A^{(m)} = Q_m A Q_m^\dagger \,, \tag{93}$$

to denote the d_m-dimensional reduced states and operators, respectively. Recall that, since any unitary $U \in \mathbb{G}$ produced by such a system is block diagonal, we can write $U = \sum_m \mathbb{P}_m U \mathbb{P}_m$. Also, let us note that if $A = A^\dagger$ then $(A^{(k)})^\dagger = A^{(k)}$.

We are ready to prove of Theorem 2, which we here recall for convenience.

Theorem 2 (Subspace controllable)**.** *Consider a system that is reducible (so that the Hilbert space is $\mathcal{H} = \bigoplus_j \mathcal{H}_j$ with each $\mathcal{H}_j$ invariant under $\mathbb{G}$), and controllable on some $\mathcal{H}_k$ of dimension d_k. Then, if the initial state is such that $\rho \in \mathcal{H}_k$, the variance of the cost function partial derivative is given by*

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] = \frac{2d_k}{(d_k^2-1)^2} \Delta(H_\mu^{(k)}) \Delta(O^{(k)}) \Delta(\rho^{(k)}) \,. \tag{94}$$

Here

$$\Delta(A) = D_{HS}\left(A, \mathrm{Tr}[A]\frac{\mathbb{1}_d}{d}\right), \quad \textit{with } D_{HS}(A,B) = \mathrm{Tr}[(A-B)^2] \tag{95}$$

is the Hilbert-Schmidt distance, and $A^{(k)}$ the reduction of operator A onto the subspace of $\mathcal{H}_k$ as defined in Eq (93).

Proof. Consider the partial derivative of the cost function $C(\boldsymbol{\theta})$ with respect to the parameter θ_{pq} $(= \theta_\mu)$, i.e. the one associated with layer p and generator H_q $(= H_\mu)$. We have

$$\partial_\mu C(\boldsymbol{\theta}) = i\mathrm{Tr}[U_B \rho U_B^\dagger [H_\mu, O_A]] \tag{96}$$

$$\begin{aligned} \partial_{pq} C(\boldsymbol{\theta}) &= \partial_{pq}\left(\mathrm{Tr}[U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger O]\right) \\ &= \partial_{pq}\left(\mathrm{Tr}[U_B \rho U_B^\dagger O_A]\right) \\ &= i\mathrm{Tr}[U_B \rho U_B^\dagger [H_q, O_A]] \end{aligned} \tag{97}$$

where in the second line we have expanded $U(\boldsymbol{\theta}) = U_A U_B$, with

$$U_B = \prod_{m=0}^{q} e^{-iH_m\theta_{pm}} \left(\prod_{l=1}^{p-1}\prod_{k=0}^{K} e^{-iH_k\theta_{lk}}\right) \quad \text{and} \quad U_A = \prod_{l=p+1}^{L}\prod_{k=0}^{K} e^{-iH_k\theta_{pk}} \left(\prod_{k=q+1}^{K} e^{-iH_k\theta_{pk}}\right), \tag{98}$$

corresponding to the unitaries before and after the parameter θ_{pq}, and $O_A = U_A^\dagger O U_A$. Then, the variance of the partial derivative is

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_{pq} C] = -\int_{\mathbb{U}_A} dU_A I(U_A) \,, \quad \text{with} \quad I(U_A) = \int_{\mathbb{U}_B} dU_B \Big(F(U_B, U_A)\Big)^2 \,, \tag{99}$$

where $\mathbb{U}_A$ and $\mathbb{U}_B$ denote the distribution of before and after unitaries, respectively, and $F(U_B, U_A) = \mathrm{Tr}[U_B \rho U_B^\dagger [H_q, U_A^\dagger O U_A]]$. Now, expanding the block-diagonal unitaries in this equation and assuming that the initial state belongs to a particular invariant subspace, i.e., $\rho^{(k)} = \mathbb{P}_k \rho \mathbb{P}_k = \rho$, we find

$$F(U_B, U_A) = \sum_{mn} \mathrm{Tr}[\mathbb{P}_m U_B \mathbb{P}_m \rho \mathbb{P}_n U_B^\dagger \mathbb{P}_n X] = \mathrm{Tr}[U_B^{(k)} \rho^{(k)} U_B^{\dagger(k)} X] \tag{100}$$

where we defined $X = [H_q^{(k)}, (U_A^{(k)})^\dagger O^{(k)} U_A^{(k)}]$, and thus

$$I(U_A) = \int_{\mathbb{U}_B} dU_B \left(\mathrm{Tr}[U_B \rho U_B^\dagger [H_q, U_A^\dagger O U_A]] \right)^2 = \int_{\mathbb{U}_B^{(k)}} dU_B^{(k)} \left(\mathrm{Tr}[U_B^{(k)} \rho^{(k)} U_B^{\dagger(k)} X] \right)^2 . \tag{101}$$

Note that here we have used the fact that, owing to Eq. (100), $F(U_B, U_A)$ actually depends only on the action of $U_B^{(k)}$, i.e., on the action of U_B projected onto the k-invariant subspace, and not on the action of the entire unitary U_B,. Hence, the integration over $\mathbb{U}_B$ can be replaced by an integration over $\mathbb{U}_B^{(k)}$.

At this point, we introduce the assumption of subspace controllability on $\mathcal{H}_k$. By virtue of Theorem 1 and Corollary 1, we know that the distribution of unitaries produced by a subspace controllable PSA constitute a ε-approximate 2-design in the subspace. Therefore, we can use Eq. (39) to integrate and get

$$I(U_A) = \int_{\mathcal{U}(d_k)} dU^{(k)} \left(\mathrm{Tr}[U^{(k)} \rho^{(k)} (U^{(k)})^\dagger X] \right)^2 = \frac{\mathrm{Tr}[XX]\, \Delta(\rho^{(k)})}{d_k^2 - 1} , \tag{102}$$

where $\Delta(A)$ is the Hilbert-Schmidt distance defined in Theorem 2. Here, we used that $\mathrm{Tr}[X] = 0$. This is grounded in the fact that, because the generator V shares the symmetry, the commutator X in the subspace is still a commutator, i.e., $X = [H_q^{(k)}, \tilde{O}^{(k)}]$ with $\tilde{O} = U_A^\dagger O U_A$. If the initial state was spread across two (or more) subspaces, neither $\rho^{(k)}$ would be a density matrix, nor X would be a commutator and one would have to be more careful in the derivation.

Finally, we proceed to integrate Eq. (102) over $\mathbb{U}_A$. Notice again that $\tilde{O}^{(k)} = \mathbb{P}_k \tilde{O} \mathbb{P}_k = U_A^{\dagger(k)} O^{(k)} U_A^{(k)}$ so X is actually only a function of $U_A^{(k)}$ (and not of the entire U_A). Consequently we can integrate over the reduced distribution $\mathbb{U}_A^{(k)}$, that, according to the subspace controllability assumption, forms a ε-approximate 2-design over $\mathcal{U}(d_k)$. This leads to

$$\mathrm{Var}_{A,B}[\partial_{pq} C] = -\frac{\Delta(\rho^{(k)})}{d_k^2 - 1} \int_{\mathcal{U}(d_k)} d\mu(U^{(k)})\, \mathrm{Tr}[XX] \tag{103}$$

$$= -2\frac{\Delta(\rho^{(k)})}{d_k^2 - 1} \int_{\mathcal{U}(d_k)} d\mu(U^{(k)}) \Big(\mathrm{Tr}[H_q^{(k)} \tilde{O}^{(k)} H_q^{(k)} \tilde{O}^{(k)}] - \mathrm{Tr}[H_q^{(k)} H_q^{(k)} \tilde{O}^{(k)} \tilde{O}^{(k)}] \Big) \tag{104}$$

$$\begin{aligned} = -2\frac{\Delta(\rho^{(k)})}{d_k^2 - 1} \Big(& \frac{(\mathrm{Tr}[H_q^{(k)}])^2 \mathrm{Tr}[O^{(k)} O^{(k)}] + \mathrm{Tr}[H_q^{(k)} H_q^{(k)}] (\mathrm{Tr}[O^{(k)}])^2}{d_k^2 - 1} \\ & - \frac{\mathrm{Tr}[H_q^{(k)} H_q^{(k)}] \mathrm{Tr}[O^{(k)} O^{(k)}] + (\mathrm{Tr}[H_q^{(k)}] \mathrm{Tr}[O^{(k)}])^2}{d_k (d_k^2 - 1)} - \frac{\mathrm{Tr}[H_q^{(k)} H_q^{(k)}] \mathrm{Tr}[O^{(k)} O^{(k)}]}{d_k} \Big) \end{aligned}$$

$$= \frac{2 d_k \Delta(\rho^{(k)})}{d_k (d_k + 1)(d_k^2 - 1)} \left(\mathrm{Tr}[H_q^{(k)} H_q^{(k)}] - \frac{1}{d_k} \mathrm{Tr}[H_q^{(k)}]^2 \right) \left(\mathrm{Tr}[O^{(k)} O^{(k)}] - \frac{1}{d_k} \mathrm{Tr}[O^{(k)}])^2 \right) \tag{105}$$

$$= \frac{2 d_k}{(d_k^2 - 1)^2} \Delta(H_q^{(k)}) \Delta(O^{(k)}) \Delta(\rho^{(k)}) . \tag{106}$$

Here, we used the notation $\mathrm{Var}_{\boldsymbol{\theta}} \to \mathrm{Var}_{A,B}$ to make explicit the assumption that both $\mathbb{U}_A^{(k)}$ and $\mathbb{U}_B^{(k)}$ form 2-designs in $\mathcal{U}(d_k)$. Note that we first expanded $\mathrm{Tr}[XX]$ in Eq. (104), and then used identities (39) and (40) on each integrand respectively to arrive at Eq. (105). Note that, even though $\mathrm{Tr}[H_q] = 0$ (and, in most applications, $\mathrm{Tr}[O] = 0$), this is not necessarily true for their reduced analogs $H_q^{(k)}$ and $O^{(k)}$.

□

K Proof of Corollary 2: Exponentially growing subspaces have barren plateaus

Let us prove Corollary 2, which we here recall.

Corollary 2. *Consider a PSA of the form in (2) giving rise to a reducible DLA, and let $\rho \in \mathcal{H}_k$, with $\mathcal{H}_k$ some invariant subspace that is controllable (i.e. the DLA reduced to such subspace is full rank). If, $\mathrm{Tr}[(H_\mu)^4], \mathrm{Tr}[O^4] \in \mathcal{O}(2^n)$, the cost function will exhibit a barren plateau for any subspace such that $d_k \in \mathcal{O}(2^n)$.*

Proof. We consider the situation in which we have a reducible system and an initial state ρ that belongs to an invariant subspace $\mathcal{H}_k$ which, by assumption, is controllable. Notice that, using Eqs. (J) and (93), we find

$$\mathrm{Tr}[A^{(k)}A^{(k)}] = \mathrm{Tr}[A\mathbb{P}_k A\mathbb{P}_k]\,. \tag{107}$$

Then, since the operator $M = A\mathbb{P}_k A$ is a positive semi-definite operator, and since $\mathbb{P}_k \leqslant \mathbb{1}$, one can write

$$\mathrm{Tr}[A^{(k)}A^{(k)}] \leqslant \mathrm{Tr}[A^2\mathbb{P}_k]\,. \tag{108}$$

Using this expression, the following bound on the Hilbert-Schmidt distance holds

$$\begin{aligned}
\Delta(A^{(k)}) &= \mathrm{Tr}[A^{(k)}A^{(k)}] - \frac{1}{d_k}(\mathrm{Tr}[A^{(k)}])^2 && (109)\\
&\leqslant \mathrm{Tr}[A^2\mathbb{P}_k] && (110)\\
&\leqslant \sqrt{\mathrm{Tr}[A^4]\mathrm{Tr}[\mathbb{P}_k^2]} && (111)\\
&\leqslant \sqrt{d_k}\sqrt{\mathrm{Tr}[A^4]}\,. && (112)
\end{aligned}$$

Using Eq. (112) and the equation for the variance in (106) we find

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] \leqslant \frac{4d_k^2}{(d_k^2-1)^2}\sqrt{\mathrm{Tr}[H_\mu^4]}\sqrt{\mathrm{Tr}[O^4]}\,, \tag{113}$$

where we have additionally used the fact that $\Delta(\rho^{(k)}) \leqslant 2\ \forall \rho$. Recalling that we are interested in the case when $d_k \in \mathcal{O}(2^n)$ we find that, assuming $\mathrm{Tr}[V^4], \mathrm{Tr}[O^4] \in \mathcal{O}(2^n)$, then the function

$$T(n) = \frac{4d_k^2}{(d_k^2-1)^2}\sqrt{\mathrm{Tr}[H_\mu^4]}\sqrt{\mathrm{Tr}[O^4]} \tag{114}$$

is such that

$$T(n) \in \mathcal{O}(1/2^n)\,. \tag{115}$$

Hence, we finally find that $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] \leqslant T(n)$, and the cost exhibits a barren plateau from the fact that the variance of the cost function partial derivative is upper bounded by a function that vanishes exponentially with n.

□

Let us finally denote that the proof of Corollary 2 follow from the fact that $\mathrm{Tr}[(H_\mu)^4], \mathrm{Tr}[O^4] \in \mathcal{O}(2^n)$. We here note that the following relevant cases satisfy this assumption:

- H_μ, O are projectors of arbitrary rank.
- H_μ, O have a decomposition in the Pauli string basis of the form $\sum_i c_i \boldsymbol{\sigma}_i$ (with c_i real coefficients, and $\boldsymbol{\sigma}_i \in \{\mathbb{1}, X, Y, Z\}^{\otimes n}$) with up to $\mathcal{O}(\mathrm{poly}(n))$ terms such that $\sum_i c_i^4 \in \mathcal{O}(\mathrm{poly}(n))$.

L Proof of Theorem 3: Expressibility in the subspace

Here we prove Theorem 3, which we now recall.

Theorem 3. *Consider a system that is reducible and let $\rho \in \mathcal{H}_k$ with $\mathcal{H}_k$ an invariant subspace of dimension d_k. Then, the variance of the cost function partial derivative is upper bounded by*

$$\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] \leqslant \min\{G_A(\rho^{(k)}), G_B(O^{(k)})\}\,, \tag{116}$$

with

$$G_B(\rho^{(k)}) = \left(\left\|\mathcal{A}_{U_B^{(k)}}\left((\rho^{(k)})^{\otimes 2}\right)\right\|_2 - \frac{\Delta(\rho^{(k)})}{d_k^2-1}\right)\mathrm{Tr}\left[\langle X^2\rangle_{U_A^{(k)}}\right] \tag{117}$$

$$G_A(O^{(k)}) = \left(\left\|\mathcal{A}_{U_A^{(k)}}\left((O^{(k)})^{\otimes 2}\right)\right\|_2 - \frac{\Delta(O^{(k)})}{d_k^2-1}\right)\mathrm{Tr}\left[\langle Y^2\rangle_{U_B^{(k)}}\right]. \tag{118}$$

Here we define $X = [H_\mu^{(k)}, (U_A^{(k)})^\dagger O\, U_A^{(k)}]$ *and* $Y = [H_\mu^{(k)}, U_B^{(k)} \rho^{(k)} (U_B^{(k)})^\dagger]$. *For simplicity, we here employed the short-hand notation* $\langle\cdot\rangle_{U_x^{(k)}}$ *(with* $x = A, B$*) to indicate the expectation value over the distribution of unitaries obtained from* $U_x^{(k)}$ *in the k-th subsystem. Finally,* $\|M\|_2 = \sqrt{\mathrm{Tr}[M^\dagger M]}$ *is the Frobenius norm, and* $\Delta(\cdot)$ *was defined in Theorem 2.*

Proof. Let us first note that the variance of the cost function partial derivative can be expressed as

$$\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] = -\int_{\mathbb{U}_A^{(k)}} dU_A^{(k)} \int_{\mathbb{U}_B^{(k)}} dU_B^{(k)} \mathrm{Tr}[U_B^{\otimes 2} (\rho^{(k)})^{\otimes 2} (U_B^\dagger)^{\otimes 2} X^{\otimes 2}] \tag{119}$$

$$= \int_{\mathbb{U}_A} dU_A \int_{\mathbb{U}_B} dU_B \mathrm{Tr}[(U_A^\dagger)^{\otimes 2} (O^{(k)})^{\otimes 2} U_A{}^{\otimes 2} Y^{\otimes 2}] \tag{120}$$

where we defined

$$X = [H_\mu^{(k)}, (U_A^{(k)})^\dagger O^{(k)} U_A^{(k)}], \quad \text{and} \quad Y = [H_\mu^{(k)}, (U_B^{(k)}) \rho^{(k)} (U_B^{(k)})^\dagger]. \tag{121}$$

Then, we recall that $\mathcal{A}_x(\cdot)$ denotes the expressibility superoperator for the second moment in the k-th subspace

$$\mathcal{A}_\omega(\cdot) = \int_{\mathcal{U}(d_k)} d\mu(U)\, U^{\otimes 2}(\cdot)(U^\dagger)^{\otimes 2} - \int_{\mathbb{U}_\omega^{(k)}} dU\, (U)^{\otimes t}(\cdot)(U^\dagger)^{\otimes t}, \tag{122}$$

where $\omega = A, B$ indicates that one evaluates the expressibility of $U_A^{(k)}(\boldsymbol{\theta})$ or $U_B^{(k)}(\boldsymbol{\theta})$ in the k-th subspace.

First, let us derive Eq. (117). Replacing $\mathcal{A}_B((\rho^{(k)})^{\otimes 2})$ into (119) leads to

$$\begin{aligned}\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] &= -\int_{\mathbb{U}_A^{(k)}} dU_A^{(k)} \int_{\mathcal{U}(d_k)} d\mu(U) \mathrm{Tr}[U_H^{\otimes 2} (\rho^{(k)})^{\otimes 2} U_H^{\dagger \otimes 2} X^{\otimes 2}] + \int_{\mathbb{U}_A^{(k)}} dU_A^{(k)} \mathrm{Tr}[\mathcal{A}_B((\rho^{(k)})^{\otimes 2}) X^{\otimes 2}] \\ &= \mathrm{Var}_B[\partial_\mu C(\boldsymbol{\theta})] + \int_{\mathbb{U}_A^{(k)}} dU_A^{(k)} \mathrm{Tr}[\mathcal{A}_B((\rho^{(k)})^{\otimes 2}) X^{\otimes 2}].\end{aligned} \tag{123}$$

Which leads to

$$|\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] - \mathrm{Var}_B[\partial_\mu C(\boldsymbol{\theta})]| \leqslant \left| \int_{\mathbb{U}_A^{(k)}} dU_A^{(k)} \mathrm{Tr}[\mathcal{A}_B((\rho^{(k)})^{\otimes 2}) X^{\otimes 2}] \right|. \tag{124}$$

Using the triangle inequality and then the Cauchy-Schwarz inequality one finds

$$|\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] - \mathrm{Var}_B[\partial_\mu C(\boldsymbol{\theta})]| \leqslant \int_{\mathbb{U}_A^{(k)}} dU_A^{(k)} |\mathrm{Tr}[\mathcal{A}_B((\rho^{(k)})^{\otimes 2}) X^{\otimes 2}]| \tag{125}$$

$$\leqslant \|\mathcal{A}_B((\rho^{(k)})^{\otimes 2})\|_2 \int_{\mathbb{U}_A^{(k)}} dU_A^{(k)} ||X^{\otimes 2}||_2 \tag{126}$$

$$\leqslant \|\mathcal{A}_B((\rho^{(k)})^{\otimes 2})\|_2 \left\langle ||X^{\otimes 2}||_2 \right\rangle_{U_A^{(k)}}. \tag{127}$$

Then, using the fact that

$$||X^{\otimes 2}||_2 = \sqrt{\mathrm{Tr}[X^{\otimes 2} X^{\otimes 2}]} = \sqrt{\mathrm{Tr}[X^2 \otimes X^2]} = |\mathrm{Tr}[X^2]| = |\mathrm{Tr}[[H_\mu^{(k)}, (U_A^{(k)})^\dagger O^{(k)} U_A^{(k)}]^2]|, \tag{128}$$

one finds

$$|\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] - \mathrm{Var}_B[\partial_\mu C(\boldsymbol{\theta})]| \leqslant \|\mathcal{A}_B((\rho^{(k)})^{\otimes 2})\|_2 \mathrm{Tr}\left[\left\langle X^2 \right\rangle_{U_A^{(k)}}\right]. \tag{129}$$

Note that the term $\mathrm{Var}_B[\partial_\mu C(\boldsymbol{\theta})]$ was explicitly computed in Eq. (102), i.e., in Eq. (102) one evaluates the variance when $U_B^{(k)}(\boldsymbol{\theta})$ forms a 2-design. Hence, replacing $\mathrm{Var}_B[\partial_\mu C(\boldsymbol{\theta})] = -\frac{\Delta(\rho^{(k)})}{d_k^2-1} \mathrm{Tr}\left[\left\langle X^2 \right\rangle_{\mathbb{U}_A^{(k)}}\right]$ in (129) we finally obtain

$$\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] \leqslant \left(\|\mathcal{A}_B((\rho^{(k)})^{\otimes 2})\|_2 - \frac{\Delta(\rho^{(k)})}{d_k^2 - 1} \right) \mathrm{Tr}\left[\left\langle X^2 \right\rangle_{U_A^{(k)}}\right]. \tag{130}$$

Similarly, one can derive Eq. (118) by replacing $\mathcal{A}_A((O^{(k)})^{\otimes 2})$ into (119), which leads to

$$\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] = -\int_{U_B^{(k)}} dU_B^{(k)} \int_{\mathcal{U}(d_k)} d\mu(U) \mathrm{Tr}[U_H^{\otimes 2} (O^{(k)})^{\otimes 2} U_H^{\dagger \otimes 2} Y^{\otimes 2}] + \int_{U_B^{(k)}} dU_B^{(k)} \mathrm{Tr}[\mathcal{A}_A((O^{(k)})^{\otimes 2}) Y^{\otimes 2}]. \tag{131}$$

Following a similar derivation to the one used in obtaining (125), one finds

$$\mathrm{Var}_\theta[\partial_\mu C(\boldsymbol{\theta})] \leqslant \left(\|\mathcal{A}_A((O^{(k)})^{\otimes 2})\|_2 - \frac{\Delta(O^{(k)})}{d_k^2 - 1} \right) \mathrm{Tr}\left[\left\langle Y^2 \right\rangle_{U_B^{(k)}}\right]. \tag{132}$$

□

M Proof of Proposition 3: Variance on the irreducible representations of $\mathcal{SU}(2)$

In this section we derive a proof for Proposition 3 for the variance of the irreducible representations of $\mathcal{SU}(2)$. Proposition 3 reads.

Proposition 3. *Consider the cost function of Eq. (20). Let $\theta_\mu = \theta_{j,x}$, and let us assume that the circuit is deep enough to allow for the distribution of unitaries U_A and U_B to converge to 2-designs on $\mathbb{G} = \mathcal{SU}(2)$. Then variance of the cost function partial derivative $\partial_\mu C(\boldsymbol{\theta}) = \partial C(\boldsymbol{\theta})/\partial\theta_\mu$ is*

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] = \frac{2m^2}{3}\,. \tag{133}$$

Proof. First, let us bring the reader into context. We consider a toy model ansatz $U(\boldsymbol{\theta}) = \prod_{l=1}^{L} e^{-i\theta_{lx}S_y} e^{-i\theta_{ly}S_x}$ with generators $\mathcal{G} = \{S_x, S_y\}$, where $\{iS_x, iS_y, iS_z\}$, with $S_\nu \in \mathbb{C}^{d\times d}$ $(\nu = x, y, z)$, form a basis of the spin $S = (d-1)/2$ irreducible representation of $\mathfrak{su}(2)$. Here, it is convenient to use as basis of the Hilbert space the set $\{|m\rangle\}$, $m = -S, -S+1, \ldots, S-1, S$, of eigenvectors of S_z. That is, we have

$$S_z|m\rangle = m|m\rangle\,, \qquad S_\pm|m\rangle = \lambda_\pm(m)|m\pm 1\rangle \tag{134}$$

with $\lambda_\pm = \sqrt{S(S+1) - m(m\pm 1)}$, and where S_+ and S_- are the spin ladder operators such that $S_x = \frac{1}{2}(S_+ + S_-)$ and $S_- = \frac{1}{2i}(S_+ - S_-)$.

The dynamical group $\mathbb{G}$ is the d-dimensional representation of $\mathcal{SU}(2)$ and we will be interested in the partial derivative w.r.t $\theta_\mu = \theta_{j,x}$, i.e., the parameter associated with generator $V = S_x$ on the j-th layer of the circuit. We will assume that the circuit is deep enough to allow for the distribution of unitaries U_A and U_B to converge to 2-designs on $\mathbb{G} = \mathcal{SU}(2)$. In consequence, when computing the variance of the different gradient components, we will be allowed to replace the integration over the angles in the PSA with an integration over the Haar measure on the group $\mathcal{SU}(2)$. Moreover, because irreducible representations of $\mathcal{SU}(2)$ and $\mathcal{SO}(3)$ are isomorphic, we can choose to integrate over the latter

$$\int_{\boldsymbol{\theta}} dU(\boldsymbol{\theta}) \to \int_{\mathcal{SU}(2)} d\mu(U) \to \int_{\mathcal{SO}(3)} d\mu(U)\,. \tag{135}$$

That is, we choose a parameterization of $\mathcal{SO}(3)$, e.g. in terms of Euler angles

$$U(\alpha,\beta,\gamma) = e^{-i\alpha S_z} e^{-i\beta S_y} e^{-i\gamma S_z}\,, \tag{136}$$

and compute the mean value of a function of the form $F_M(\boldsymbol{\theta}) = \mathrm{Tr}[U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger M]$, with M Hermitian, over the cost landscape as

$$\begin{aligned}\langle F\rangle &= \int_{\boldsymbol{\theta}} dU(\boldsymbol{\theta})\mathrm{Tr}[U(\boldsymbol{\theta})\rho U(\boldsymbol{\theta})^\dagger M]\\ &= \int_{\mathcal{SO}(3)} d\mu(U)\mathrm{Tr}[U\rho U^\dagger M]\\ &= \frac{1}{8\pi^2}\int_0^{2\pi} d\alpha \int_0^{\pi} d\beta \sin\beta \int_0^{2\pi} d\gamma\, \mathrm{Tr}[U(\alpha,\beta,\gamma)\rho U(\alpha,\beta,\gamma)^\dagger M]\,.\end{aligned} \tag{137}$$

where $\frac{1}{8\pi^2} d\alpha\, d\beta \sin\beta\, d\gamma$ is the normalized Haar measure for this parametrization of $\mathcal{SO}(3)$. Considering an initial state that is an eigenstate of J_z, i.e., $\rho = |m\rangle\langle m|$, its evolution can be conveniently expressed in terms of Wigner's small d-matrices

$$U(\alpha.\beta,\gamma)\rho U(\alpha.\beta,\gamma)^\dagger = \sum_{m'm''} e^{-i\alpha(m'-m'')} d_{m'm}(\beta) d_{mm''}(-\beta)|m'\rangle\langle m''| \tag{138}$$

and using that $d^j_{rk}(\beta) = d^j_{kr}(-\beta)$ and the orthonormality relations

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta e^{-i\theta(m-m')} = \delta_{mm'}\,, \qquad \frac{2S+1}{2}\int_0^{\pi} d\beta\, \sin\beta\, d^S_{rk}(\beta) d^{S'}_{rk}(\beta) = \delta_{S,S'} \tag{139}$$

we arrive at

$$\langle F\rangle = \frac{\mathrm{Tr}[M]}{2S+1}\,. \tag{140}$$

M.0.1 An unbiased gradient

Now, the computation of the mean value of $\partial_\mu C(\boldsymbol{\theta})$ amounts to making the choice $M = X$, with $X = [H_\mu, U_A^\dagger O U_A]$. We readily find that

$$\langle \partial_\mu C \rangle = \int_{\mathbb{U}_A} dU_A W(U_A) = 0 \tag{141}$$

since

$$W(U_A) = \int_{\mathbb{U}_B} dU_B \mathrm{Tr}[U_B \rho U_B^\dagger X(U_A)] = \int_{\mathcal{SO}(3)} d\mu(U) \mathrm{Tr}[U\rho U^\dagger X] = \frac{\mathrm{Tr}[X(U_A)]}{2S+1} = 0\,, \tag{142}$$

which follows from the fact that commutators are traceless.

M.0.2 The variance

Let us now compute the variance of $\partial_\mu C(\boldsymbol{\theta})$. Having shown that $\langle \partial_\mu C \rangle = 0$, the variance is

$$\mathrm{Var}[\partial_\mu C] = \langle (\partial_\mu C)^2 \rangle = -\int_{\mathbb{U}_A} dU_A W(U_A)\,, \tag{143}$$

where

$$W(U_A) = \int_{\mathbb{U}_B} dU_B \mathrm{Tr}[U_B^{\otimes 2} \rho^{\otimes 2} (U_B^{\otimes 2})^\dagger X^{\otimes 2}] = \frac{1}{8\pi^2} \int_0^{2\pi} d\alpha \int_0^{\pi} d\beta \sin\beta \int_0^{2\pi} d\gamma\, Y(\alpha,\beta,\gamma)\,, \tag{144}$$

and

$$\begin{aligned} Y(\alpha,\beta,\gamma) &= \mathrm{Tr}[U^{\otimes 2}(\alpha,\beta,\gamma)\rho^{\otimes 2}(U^{\otimes 2}(\alpha,\beta,\gamma))^\dagger X] \\ &= \sum_{m'm''n'n''} e^{-i\alpha(m'-m'')} e^{-i\alpha(n'-n'')} d_{m'm}(\beta) d_{m''m}(\beta) \langle m''|X|m'\rangle d_{n'n}(\beta) d_{n''n}(\beta) \langle n''|X|n'\rangle \end{aligned} \tag{145}$$

Here and henceforth, $d_{m,m'} \equiv d^S_{m,m'}$. Now, again, integration over γ is trivial, and integrating over α we get $\delta_{n'',m'+n'-m''}$, so that

$$W(U_A) = \sum_{m'm''n'} \langle m''|X|m'\rangle \langle m'+n'-m''|X|n'\rangle \frac{1}{2} \int_0^{\pi} d\beta\, \sin\beta\, d_{m'm}(\beta)\, d_{m''m}(\beta)\, d_{n'n}(\beta)\, d_{m'+n'-m'',n}(\beta)\,. \tag{146}$$

In the following we will consider that the cost function minimize the expectation value of an operator O that is an element of the dynamical algebra, i.e., $O \in \mathfrak{g}$. It is easy to verify that any unitary $U(\alpha,\beta,\gamma)$ of the form of Eq. (136), associated with some 3-dimensional Euler rotation matrix $R(\alpha,\beta,\gamma)$, transforms an arbitrary element $S_{\hat{n}} = (S_x, S_y, S_z)\cdot\hat{n}$ (with $\hat{n}$ a unit vector $\in \mathbb{R}^3$) of the algebra in the following manner

$$U(\alpha,\beta,\gamma) S_{\hat{n}} U^\dagger(\alpha,\beta,\gamma) = S_{\hat{n}'}, \quad \text{with } \hat{n}' = R(\alpha,\beta,\gamma)\hat{n} \tag{147}$$

Using the shorthand $\sin(x) \to s_x$ and $\cos(x) \to c_x$, the rotation matrix reads

$$R(\alpha,\beta,\gamma) = \begin{bmatrix} c_\alpha c_\beta c_\gamma - s_\alpha s_\gamma & -c_\gamma s_\alpha - c_\alpha c_\beta s_\gamma & c_\alpha s_\beta \\ c_\alpha s_\gamma + c_\beta c_\gamma s_\alpha & c_\alpha c_\gamma - c_\beta s_\alpha s_\gamma & s_\alpha s_\beta \\ -c_\gamma s_\beta & s_\beta s_\gamma & c_\beta \end{bmatrix}. \tag{148}$$

Hence, considering $O = S_x + S_y + S_z$ and, calling $X^{(\iota)} = [J_x, U S_\nu U^\dagger]$, we get

$$\begin{aligned} X^{(x)} &= i s_\beta c_\gamma S_y + i(c_\alpha s_\gamma + s_\alpha c_\beta c_\gamma) S_z \\ X^{(y)} &= -i s_\beta s_\gamma S_y + i(c_\alpha c_\gamma - s_\alpha c_\beta s_\gamma) S_z \\ X^{(z)} &= -i c_\beta S_y + i s_\alpha s_\beta S_z\,. \end{aligned} \tag{149}$$

Thus, denoting $X = \sum_{\nu=\{x,y,z\}} X^{(\nu)} = (a_x, a_y, a_z)\cdot(S_x, S_y, S_z)^T$ we find

$$\begin{aligned} a_x &= 0 \\ a_y &= -i(s_\beta(s_\gamma - c_\alpha) + c_\beta) \\ a_z &= i(s_\alpha(s_\beta + c_\beta(c_\gamma - s_\gamma)) + c_\alpha(s_\gamma + c_\gamma))\,. \end{aligned} \tag{150}$$

For the moment, let us neglect the explicit dependence of a_i on the Euler angles and let us compute the matrix elements of X

$$\begin{aligned}\langle m''|X|m'\rangle &= a_+\langle m''|S_+|m'\rangle + a_-\langle m''|S_-|m'\rangle + a_z\langle m''|S_z|m'\rangle \\ &= a_+\lambda_+(m')\delta_{m'',m'+1} + a_-\lambda_-(m')\delta_{m'',m'-1} + a_z m'\delta_{m'',m'}\end{aligned} \tag{151}$$

where $a_\pm = \frac{a_x \mp i a_y}{2}$, and where we here define $S_\pm|m\rangle = \lambda_\pm(m)|m\pm 1\rangle$. Similarly,

$$\begin{aligned}\langle m'+n'-m''|X|n'\rangle &= a_+\lambda_+(n')\delta_{m'+n'-m'',n'+1} + a_-\lambda_-(n')\delta_{m'+n'-m'',n'-1} + a_z n'\delta_{m'+n'-m'',n'} \\ &= a_+\lambda_+(n')\delta_{m'',m'-1} + a_-\lambda_-(n')\delta_{m'',m'+1} + a_z n'\delta_{m'',m'}\,.\end{aligned} \tag{152}$$

Now, out of the nine terms in the product $\langle m''|X|m'\rangle\langle m'+n'-m''|X|n'\rangle$ appearing in Eq. (146), only three terms are non-zero. That is, one only gets

$$\begin{aligned}T_1 &= a_+a_-\lambda_+(m')\lambda_-(n')\delta_{m'',m'+1} \\ T_2 &= a_+a_-\lambda_-(m')\lambda_+(n')\delta_{m'',m'-1} \\ T_3 &= a_z^2 m'n'\delta_{m'',m'}\,.\end{aligned} \tag{153}$$

which give rise to three terms on Y, namely Y_1, Y_2 and Y_3. Consider the first of these

$$Y_1 = \sum_{m'n'} a_+a_-\lambda_+(m')\lambda_-(n')\frac{1}{2}\int_0^\pi d\beta\,\sin\beta\, d_{m'm}(\beta)\, d_{m'+1,m}(\beta)\, d_{n'n}(\beta)\, d_{n'-1,n}(\beta)\,, \tag{154}$$

where, we recall that the coefficients $a_i = a_i(U_A)$ depend on U_A. Now, we can use the identity

$$d^S_{rk}(\beta)d^{S'}_{r'k'}(\beta) = \sum_{J=|S-S'|}^{S+S'} \langle S'rSr'\,|\,J, r+r'\rangle\langle SkS'k'\,|\,J, k+k'\rangle d^J_{r+r',k+k'}(\beta)\,. \tag{155}$$

where we henceforth employ the more general notation $|m\rangle \to |S,m\rangle$. In our context, $S = S'$, and we will reference the Clebsh-Gordan coefficients as $g_{r,k} = \langle SrSk\,|\,J, r+k\rangle$. We get

$$\begin{aligned}Y_1 = \sum_{m'n'JJ'} & a_+a_-\lambda_+(m')\lambda_-(n')\frac{1}{2}\int_0^\pi d\beta\,\sin\beta\, d^J_{m'+n',m+n}(\beta)d^{J'}_{m'+n',m+n}(\beta) \\ & \times \langle Sm'Sn'\,|\,J, m'+n'\rangle\,\langle S, m'+1, S, n'-1\,|\,J, m'+n'\rangle\,\langle SmSn\,|\,J, m+n\rangle\,\langle SmSn\,|\,J', m+n\rangle,\end{aligned} \tag{156}$$

and using the orthogonality relations yields

$$\begin{aligned}Y_1 &= \sum_{m'n'J} \frac{a_+a_-\lambda_+(m')\lambda_-(n')}{2J+1}\langle Sm'Sn'\,|\,J, m'+n'\rangle\,\langle S, m'+1, S, n'-1\,|\,J, m'+n'\rangle(\langle SmSn\,|\,J, m+n\rangle)^2 \\ &= \sum_{m'n'J} \frac{a_+a_-\lambda_+(m')\lambda_-(n')}{2J+1} g_{m',n'}g_{m'+1,n'-1}g^2_{m,n}\,.\end{aligned} \tag{157}$$

Similarly

$$\begin{aligned}Y_2 &= \sum_{m'n'J} \frac{a_+a_-\lambda_-(m')\lambda_+(n')}{2J+1} g_{m',n'}g_{m'-1,n'+1}g^2_{m,n}\,, \\ Y_3 &= \sum_{m'n'J} \frac{a_z^2 m'n'}{2J+1} g^2_{m',n'}g^2_{m,n}\,.\end{aligned} \tag{158}$$

These can be restated as $Y_1 = Y_2 = c_1\, a_+a_-$ and $Y_3 = c_3\, a_z^2$ with

$$\begin{aligned}c_1 &= \sum_{m'n'J} \frac{\lambda_-(m')\lambda_+(n')}{2J+1} g_{m',n'}g_{m'-1,n'+1}g^2_{m,n} = \frac{2}{3}m^2 \\ c_3 &= \sum_{m'n'J} \frac{a_z^2 m'n'}{2J+1} g^2_{m',n'}g^2_{m,n} = \frac{1}{3}m^2\end{aligned} \tag{159}$$

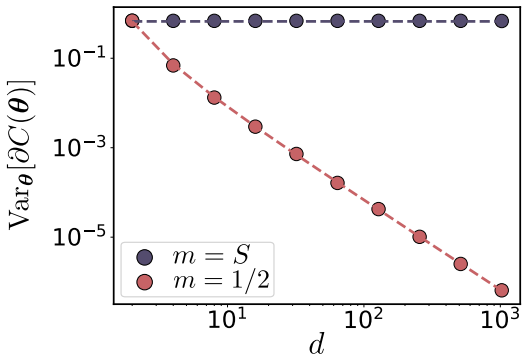

Figure 1: **Numerical results for the $\mathcal{SU}(2)$ toy model algebra.** Variance of the cost function partial derivative of the cost function in (164) as a function of $d = \dim(\mathcal{H})$. Dashed lines indicate the theoretical prediction of (162).

We can proceed to integrate over U_A, assuming again we can replace $\mathbb{U}_A$ with an integration over the Haar measure on $\mathbb{G}$. Let us call v_i each contribution to the variance, i.e. $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] = \sum_{i=1}^{3} v_i$. First, one finds by explicit integration that let us consider

$$\begin{aligned} v_3 &= -c_3 \int_{U_A} dU_A a_z^2(U_A) \\ &= -c_3 \frac{1}{8\pi^2} \int_0^{2\pi} d\tilde{\alpha} \int_0^{\pi} d\tilde{\beta} \sin\tilde{\beta} \int_0^{2\pi} d\tilde{\gamma} a_z^2(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \\ &= c_3 \,. \end{aligned} \tag{160}$$

Similarly, it is easy to see that

$$v_1 = v_2 = \frac{c_1}{4} \,. \tag{161}$$

Altogether, we finally find

$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})] = \mathrm{Var}_{A,B}[\partial_\mu C(\boldsymbol{\theta})] = \sum_{i=1}^{3} v_i = \frac{2}{3} m^2 \tag{162}$$

where we have introduced $\mathrm{Var}_{A,B}$ to make explicit the assumption made, that is, that the before and after distributions of unitaries form 2-designs.

□

M.0.3 Numerical simulations of the $\mathcal{SU}(2)$ toy model algebra.

Here we numerically simulate the model employing a PSA of the form

$$U(\boldsymbol{\theta}) = \prod_{l=1,\ldots,L} e^{-i\theta_{l1} S_y} e^{-i\theta_{l2} S_x} \,, \tag{163}$$

for minimizing the normalized cost function

$$C(\boldsymbol{\theta}) = \langle m | U^\dagger(\boldsymbol{\theta}) \left(S_x + S_y + S_z\right) U(\boldsymbol{\theta}) | m \rangle / S \,. \tag{164}$$

We compute $\mathrm{Var}_{\boldsymbol{\theta}}[\partial_\mu C(\boldsymbol{\theta})]$ for the initial states $|m\rangle = |S\rangle$ and $|m\rangle = |1/2\rangle$ for $L = 100$ using 1200 random initialization. As shown in Figure 1, the theoretical prediction of Eq. (162) matches the numerical results, indicating that an initial state $|m\rangle = |1/2\rangle$ is trainable, while an initial state $|m\rangle = |S\rangle$ leads to a barren plateau in the cost function.

N Symmetries in the XXZ and TFIM systems

N.1 The XXZ model

Consider the generators for an HVA ansatz to the XXZ model in Eq. (16) of section 5.2.1. The elements in $\mathcal{G}_{XXZ_U}$, and thus in its associated DLA, share two common symmetries: they commute both with the

magnetization operator $M = \sum_{i=1}^{n} Z_i$ and with the parity operator

$$\Pi = \prod_{m=1}^{[n/2]} P_{m,n-m+1}, \quad \text{where } P_{ij} = \frac{\mathbb{1} + \boldsymbol{\sigma}_i \boldsymbol{\sigma}_j}{2} \quad \text{and } \boldsymbol{\sigma}_i = (X_i, Y_i, Z_i). \tag{165}$$

Here, Π is the unitary d-dimensional irreducible representation of the element of the symmetric group $\mathcal{S}_n$ that corresponds to a reflection over the central constituent. As a result of these two symmetries, the state space is broken into invariant subspaces with well defined parity and excitation number

$$\mathcal{H} = \bigoplus_{\substack{m=0 \\ \sigma=\pm}}^{n} \mathcal{H}_{m,\sigma} \tag{166}$$

The dimension of each excitation subspace is $\dim(\mathcal{H}_m) = \binom{n}{m}$, whereas, the joint parity-excitation subspaces have dimension $\dim(\mathcal{H}_{m,\sigma}) \approx \frac{1}{2}\dim(\mathcal{H}_m)$. The DLA is, accordingly, a direct sum of simple algebras on each invariant subspace

$$\mathfrak{g}_{XXZ_U} = \bigoplus_{\substack{m=0 \\ \sigma=\pm}}^{n} \mathfrak{g}_{m,\sigma} \subseteq \mathfrak{u}(d_{m,\sigma})\,. \tag{167}$$

Notably, upon the addition of a generator consisting of local fields on either end of the chain, $\mathcal{G}_{XXZ} = \mathcal{G}_{XXZ_U} \cup \{\mathrm{Z}_1 + \mathrm{Z}_N\}$, the DLA can be shown to become full rank on each subspace $\mathfrak{g}_{\mathrm{XXZ}} = \bigoplus_{\substack{m=0 \\ \sigma=\pm}}^{n} \mathfrak{u}(d_{m,\sigma})$ [87].

N.2 The TFIM and LTFIM models

In this section we review the symmetries of the different variants of the Transverse Field Ising Model (TFIM) presented in Section 5.2.2.

N.2.1 Open boundary condition

Let us first consider open boundary conditions on the TFIM model. In this case, the generators $\mathcal{G}_{\mathrm{TFIM}} = \left\{\sum_{i=1}^{n-1} Z_i Z_{i+1}, \sum_{i=1}^{n} X_i\right\}$ have two symmetries. On one hand, they commute with the parity symmetry Π defined in Eq. (165). On the other hand, they commute with the so-called $\mathbb{Z}_2$ operator

$$\Pi_{\mathbb{Z}_2} = \prod_{i=1}^{n} X_i \tag{168}$$

that amounts to a global flip of the qubits. Consequently, $\mathcal{H}$ is broken into four subspaces $\mathcal{H}_{\sigma,\sigma'}$ with $\sigma, \sigma' \in \{1, -1\}$. Because the initial state $|+\rangle^{\otimes n}$ for the cost function in Eq. (32) is an eigenstate of both symmetry operators with eigenvalues $\sigma = \sigma' = +1$, the dynamics under such a PQC is constrained to the $\mathcal{H}_{+1,+1}$ subspace. We find, using Algorithm 1, that the dimension of the DLA scales polynomially, $\dim(\mathfrak{g}) = n^2$ (and so does the restriction to the $\sigma = \sigma' = +1$ subspace.

In turn, consider open boundary conditions on the *Longitudinal* Transverse Field Ising Model (LTFIM), given by generators $\mathcal{G}_{\mathrm{LTFIM}} = \mathcal{G}_{\mathrm{TFIM}} \bigcup \{\sum_{i=1}^{n} Z_i\}$. While parity symmetry is conserved, the introduction of this new global longitudinal field breaks the $\mathbb{Z}_2$ symmetry. Thus, we are left with only two subspaces, $\mathcal{H} = \bigoplus_{\sigma=\pm 1} \mathcal{H}_\sigma$, of dimensions $\dim(\mathcal{H}_\sigma) \approx \frac{2^n}{2}$. As expected, the DLA breaks into two corresponding subspace DLAs, each of which we find to be full rank on the corresponding subspaces, i.e., both subspaces are controllable.

N.2.2 Closed boundary conditions

Let us now consider closed boundary conditions. This case is slightly more involved since the parity symmetry is replaced by C_n, the cyclic group of n elements. Hereafter, we follow Ref. [111]. Consider the operator R whose action is to cycle the qubits in a state, i.e., $R|a_1, \ldots, a_n\rangle = |a_n, a_1, \ldots, a_{n-1}\rangle$. Clearly, $R^n = \mathbb{1}$. Now, as a consequence of the invariance of the generators under the action of this group of symmetries, the state space is broken into n invariant subspaces $\mathcal{H} = \bigoplus_{k=0}^{n-1} \mathcal{H}_k$, where the projector onto each subspace is given by

$$\mathbb{P}_k = \frac{1}{n} \sum_{j=0}^{n-1} \varepsilon^{kj} R^j\,. \tag{169}$$

In particular, the state $|+\rangle^{\otimes n} \in \mathcal{H}_0$ and for that reason we will only focus on this subspace. Let us note that a general formula for the dimension of this subspace is, as far as we know, not known, but it is possible to derive a closed expression in the case of n prime [111]

$$\dim(\mathcal{H}_0) = 2 + \frac{(2^n - 2)}{n}, \tag{170}$$

which makes evident that the subspace is exponentially large. Furthermore, we compute the DLA reduced to this subspace in the LTFIM case and see an exponential behaviour. In contrast, for the TFIM model we find a DLA that reduced to the $k = 0$ subspace and the $\sigma' = +1$ (also has $\mathbb{Z}_2$ symmetry), and that has a linear scaling, i.e., $\dim(\mathfrak{g}_{0,+}) \in \mathcal{O}(n)$.

O Initial state for the numerical simulations of the XXZ spin chain model

The initial states $|\psi_{m,+}\rangle$ in the cost function of Eq. (25) are chosen to be eigenstates of the symmetries of the generators, namely $M = \sum_{i=1}^{n} Z_i$ and Π (the reflective spatial symmetry w.r.t. the chain's center) defined in Eq. (165). That is, the initial state is

$$|\psi_{m,+}\rangle = \frac{1}{\sqrt{2}} \left(\bigotimes_{i=1}^{m} |0\rangle_i \bigotimes_{i=m+1}^{n} |1\rangle_i + \bigotimes_{i=1}^{n-m} |1\rangle_i \bigotimes_{i=n-m+1}^{n} |0\rangle_i \right). \tag{171}$$

Here m denotes the number of excitations in the state, and $+$ the fact that it is an even parity eigenstate.