Dynamical Lie Algebras and Barren Plateaus

arxiv:2309.07902

The Adjoint Is All You Need: Characterizing Barren Plateaus in Quantum Ansätze

Enrico Fontana,[1,2] Dylan Herman,[1,*] Shouvanik Chakrabarti,[1] Niraj Kumar,[1] Romina Yalovetzky,[1] Jamie Heredge,[1,3] Shree Hari Sureshbabu,[1] and Marco Pistoia[1]

[1] *Global Technology Applied Research, JPMorgan Chase*
[2] *Computer and Information Sciences, University of Strathclyde*
[3] *School of Physics, The University of Melbourne*

Using tools from the representation theory of compact Lie groups, we formulate a theory of Barren Plateaus for parameterized quantum circuits whose observables lie in their dynamical Lie algebra (DLA), a setting that we term Lie algebra Supported Ansatz (LASA). A large variety of commonly used ansätze such as the Hamiltonian Variational Ansatz, Quantum Alternating Operator Ansatz, and many equivariant quantum neural networks are LASAs. In particular, our theory provides for the first time the ability to compute the variance of the gradient of the cost function for a non-trivial, subspace uncontrollable family of quantum circuits, the quantum compound ansätze. We rigorously prove that the variance of the gradient of the cost function, under Haar initialization, scales inversely with the dimension of the DLA, which agrees with existing numerical observations. Lastly, we include potential extensions for handling cases when the observable lies outside of the DLA and the implications of our results.

I. INTRODUCTION

Variational quantum algorithms (VQAs) are a popular class of quantum computing heuristics due to their low circuit cost and ability to be trained in a hybrid quantum-classical fashion [1]. VQAs have found a variety of applications in the areas of optimization [2–7] and machine learning [8–12]. Unfortunately, the optimization of VQAs can be a computationally challenging task due to (1) exponentially many parameters being required to ensure convergence [13–17], and (2) exponentially many samples being required to estimate gradients, known as the barren plateau (BP) problem [18–21]. It has been observed that both of these obstacles to VQA optimization can be mitigated when the chosen parameterized quantum circuit (PQC) obeys symmetries [14, 22]. The symmetries of the ansatz cause its action to break into invariant subspaces, and in each invariant subspace the quantities controlling trainability and convergence only depend on characteristics of the subspace, e.g., its dimension.

The existing theoretical results on the trainability of ansätze with symmetries have been restricted to the subspace controllable setting [18, 22, 23]. Subspace controllability occurs when the circuit can express any unitary transformation between states in an invariant subspace and it has been observed that it results in training landscapes that are essentially trap-free [24, 25]. In addition, if the invariant subspaces have small dimension, i.e. scale polynomially in system size, it can be easily shown that BPs are not present. However, it is challenging in general to determine if subspace controllability can be attained, and there are commonly used PQCs that do not attain it, such as the quantum compound ansatz [26, 27]. With respect to the BPs problem, existing work has observed a desirable feature of subspace uncontrollable circuits [22]. In this setting, it appears that the trainability of the ansatz depends on the

arxiv:2309.09342

A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits

Michael Ragone,[1,*] Bojko N. Bakalov,[2,*] Frédéric Sauvage,[3] Alexander F. Kemper,[4] Carlos Ortiz Marrero,[5,6] Martín Larocca,[3,7] and M. Cerezo[8,†]

[1] *Department of Mathematics, University of California Davis, Davis, California 95616, USA*
[2] *Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695, USA*
[3] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[4] *Department of Physics, North Carolina State University, Raleigh, North Carolina 27695, USA*
[5] *AI & Data Analytics Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA*
[6] *Department of Electrical & Computer Engineering, North Carolina State University, Raleigh, North Carolina 27607, USA*
[7] *Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[8] *Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

Variational quantum computing schemes have received considerable attention due to their high versatility and potential to make practical use of near-term quantum devices. At their core, these models train a loss function by sending an initial state through a parametrized quantum circuit, and evaluating the expectation value of some operator at the circuit's output. Despite their promise, the trainablity of these algorithms is hindered by barren plateaus induced by the expressiveness of the parametrized quantum circuit, the entanglement of the input data, the locality of the observable, or the presence of hardware noise. Up to this point, these sources of barren plateaus have been regarded as independent and have been studied only for specific circuit architectures. In this work, we present a general Lie algebraic theory that provides an exact expression for the variance of the loss function of sufficiently deep parametrized quantum circuits, even in the presence of certain noise models. Our results unify under one single framework all aforementioned sources of barren plateaus by leveraging generalized (and subsystem independent) notions of entanglement and operator locality, as well as generalized notions of algebraic decoherence due to noise. This theoretical leap resolves a standing conjecture about a connection between loss concentration and the dimension of the Lie algebra of the generators of the parametrized circuit.

I. INTRODUCTION

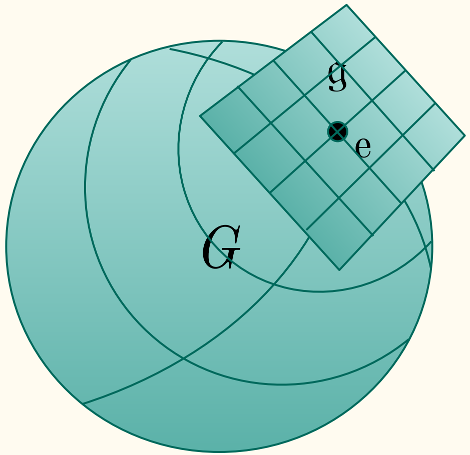
Variational quantum computing schemes, such as variational quantum algorithms [1–6] or quantum machine learning models [7–10], share a common structure in which quantum and classical resources are used to solve a given task. In a nutshell, these algorithms send some initial state through a parametrized quantum circuit, and then perform (a polynomial number of) measurements to estimate the expectation value of some observable that encodes the loss function (also called cost function) appropriate for the

of the main obstacles towards trainability is the presen of Barren Plateaus (BPs) [13] in the loss function. In t presence of BPs, this loss function (and its gradients) e ponentially concentrates in parameter space as the size the problem increases. Therefore, unless an exponenti number of measurement shots are employed, the model b comes untrainable, as one does not have enough precisi to find a loss-minimizing direction and navigate the lo landscape.

Due to the tremendous limitations that BPs place on t potential to scale variational quantum computing schem

Representation Theory of VQAs

- Symmetric ansätze appear in quantum machine learning, problem-aware ansätzes (eg. QAOA).
- The Lie group of allowed unitaries in a VQA is a subgroup of the unitary group, the Dynamical Lie Group G.
- By taking the tangent space at the identity, we find the **Dynamical Lie Algebra $\mathfrak{g}$ (DLA).**



Barren Plateaus (the Rigorous Way)

- The **exponential decay of the variance** (under uniform initialization or Haar measure) of an observable's expectation value with system size:

$$\mathrm{Var}_{\boldsymbol{\theta}\sim\mu}[\partial_i \ell_{\boldsymbol{\theta}}(\rho, O)] \in \mathcal{O}\left(\frac{1}{b^n}\right)$$

- **Strongly limits the usefulness** of certain ansatzes.
- We need to compute a second-moment integral under a group's Haar measure:

$$\mathcal{T}_2(O) := \int_G (U_g O U_g^\dagger)^{\otimes 2}\, d\mu_g$$

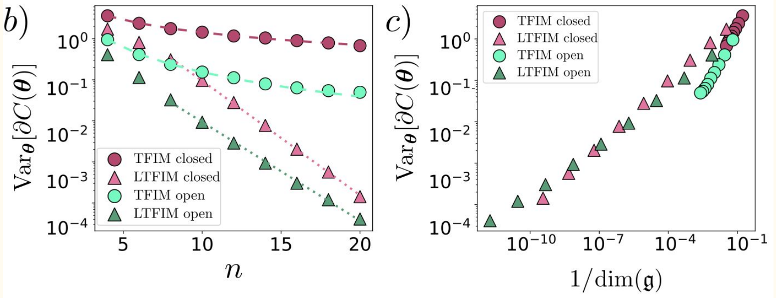
- Easy for classical Lie groups (Schur-Weyl and variants), essentially impossible with arbitrary Lie groups.

Conjecture

BP-DLA Conjecture [Larocca *et al.*, 2021]
Given an ansatz with DLA $\mathfrak{g}$, the variance of the gradient over random initialization of the parameters scales as

$$\mathrm{Var}_{\boldsymbol{\theta}\sim\mu}[\partial_i \ell_{\boldsymbol{\theta}}(\rho, O)] \in \mathcal{O}\left(\frac{1}{\mathrm{poly}(\dim(\mathfrak{g}))}\right)$$

- Observationally true, and in some cases analytically.
- No clear indication of why it should hold – **where is the DLA coming from?**



Diagnosing barren plateaus with tools from quantum optimal control, Larocca *et al. (2022)*.

Proof of Conjecture

- Key realization: all examples had an observable (or state) that was also in the DLA. We call this a **Lie Algebra Supported Ansatz (LASA)**.
- We need to look at an object, constructed from the DLA, invariant under the second moment integral. Only choice is the **quadratic Casimir**:

$$K := \sum_i E_i \otimes E_i$$

Lemma

The second moment of an operator in the Lie algebra over the Haar measure of a Lie group is

$$\mathcal{T}_2(O) = \frac{\mathrm{Tr}[O^{\otimes 2} K]\, K}{\dim(\mathfrak{g})}$$

Main Theorems

- Define the **g-purity** $\mathcal{P}_{\mathfrak{g}}(O) := \mathrm{Tr}[O^{\otimes 2} K]$ (ie. **purity on DLA subspace**).
- Define the Killing norm $\|O\|_K^2 := \mathrm{Tr}_{Ad}[O^2]$ as the Frobenius norm in the adjoint representation.

Theorem (Simple group, cost variance)

$$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \frac{\mathcal{P}_{\mathfrak{g}}(\rho)\mathcal{P}_{\mathfrak{g}}(O)}{\dim(\mathfrak{g})}$$

Theorem (Simple group, gradient variance)

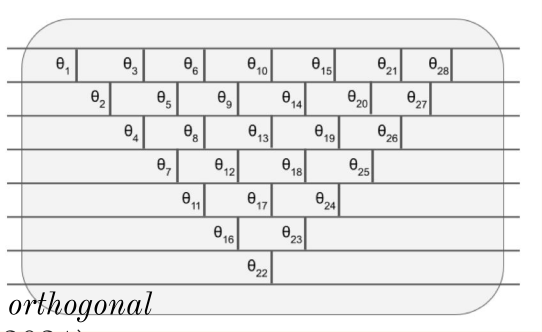
$$\mathrm{Var}_{\boldsymbol{\theta}}[\partial\ell_{\boldsymbol{\theta}}(\rho, O)] = \frac{\mathcal{P}_{\mathfrak{g}}(\rho)\mathcal{P}_{\mathfrak{g}}(O)\|H\|_K^2}{\dim(\mathfrak{g})^2}$$

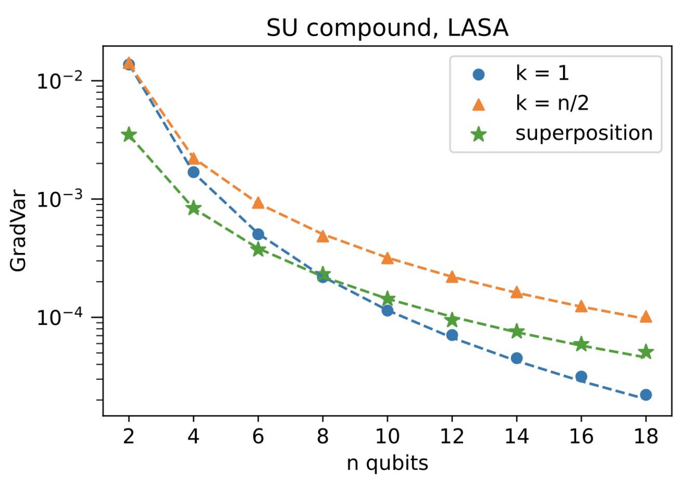
- Can be extended to general compact Lie groups (so *any* dynamical Lie group of a circuit): sum of contributions from subalgebras/ideals.

Theory in Practice: Quantum Compound Layers

- Quantum compound layers or orthogonal layers are composed by fermionic beam splitter (FBS) or reconfigurable beam splitter (RBS) gates. They represent the **orthogonal group (alternating rep)**.
- We find that compound layers **do not** have BPs (for LASA).
- **Excellent agreement** with numerics.
- Convergence to Haar is fast in practice.

$$RBS(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$





[Images] *Classical and quantum algorithms for orthogonal neural networks*, Kerenidis, Landman, Mathur (2021).

No Barren Plateau

Barren Plateau

$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] \in \Omega(1/\mathrm{poly}(n))$

$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] \in \mathcal{O}(1/b^n)$ with $b > 1$.

Image credit: Samson Wang

A "Unified" Theory of Barren Plateaus

The theorem (for simple $\mathfrak{g}$):*

Dynamical Lie algebra $\mathfrak{g}$

$\mathfrak{g}$-purity of input state ϱ and measurement observable O

$$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \frac{1}{\dim(\mathfrak{g})} \mathcal{P}_{\mathfrak{g}}(\rho) \mathcal{P}_{\mathfrak{g}}(O)$$

*when either ϱ or O in $i\mathfrak{g}$

How can we see existing knowledge of barren plateaus here?
Let's first revisit $\mathfrak{g}$-purity.

Another Way to Think About $\mathfrak{g}$-purity

Let $\mathfrak{g} \subseteq \mathfrak{su}(2^n)$ any subalgebra with orthonormal basis $\{B_j\}$, $1 \leq j \leq \dim(\mathfrak{g})$.

Equivalent definition of $\mathfrak{g}$-purity of Hermitian operator H :

$$\mathcal{P}_{\mathfrak{g}}(H) = \mathrm{Tr}\left[H_{\mathfrak{g}}^2\right] = \sum_{j=1}^{\dim(\mathfrak{g})} \left|\mathrm{Tr}\left[B_j^\dagger H\right]\right|^2$$

where $H_{\mathfrak{g}}$ is the projection of H onto the vector space $\mathfrak{g}$.

Circuit Expressivity (or, no-free-lunch for trainability)

Original insight: when $U(\boldsymbol{\theta})$ a 2-design (so $\mathfrak{g}$ is $\mathfrak{u}(2^n)$), we have a barren plateau for any $\boldsymbol{\varrho}$, O

Lie unification:

Highly expressive circuit

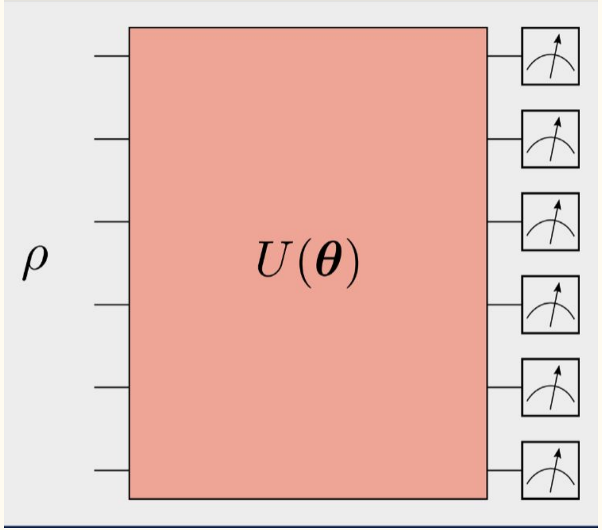
||

Large $\dim(\mathfrak{g})$

||

Small variance

$$\frac{1}{\dim(\mathfrak{g})}\mathcal{P}_{\mathfrak{g}}(\rho)\mathcal{P}_{\mathfrak{g}}(O)$$



Barren Plateaus in Quantum Neural Networks, McClean et al, (2018)

Entanglement of Input State ϱ

Original insight: highly entangled input states lead to barren plateaus

Lie unification:

Highly entangled ϱ

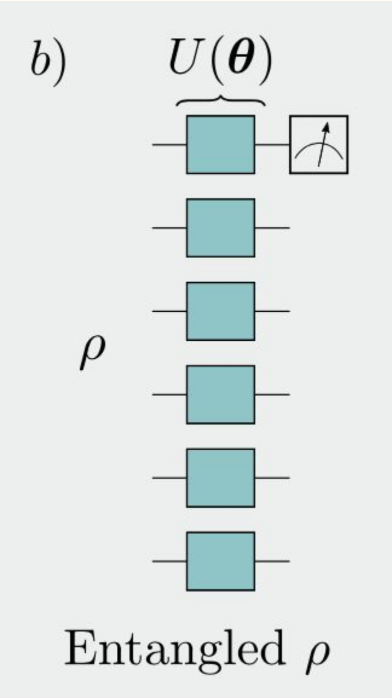
$\parallel$

Small $\mathfrak{g}$-purity

$\parallel$

Small variance

$$\frac{1}{\dim(\mathfrak{g})}\mathcal{P}_{\mathfrak{g}}(\rho)\mathcal{P}_{\mathfrak{g}}(O)$$



pure highest weight state ϱ = maximal $\mathfrak{g}$-purity

Entanglement-Induced Barren Plateaus, Marrero et al (2021)

Global Measurement Observables O

Original insight: global measurements lead to barren plateaus

Lie unification:

Global measurement O

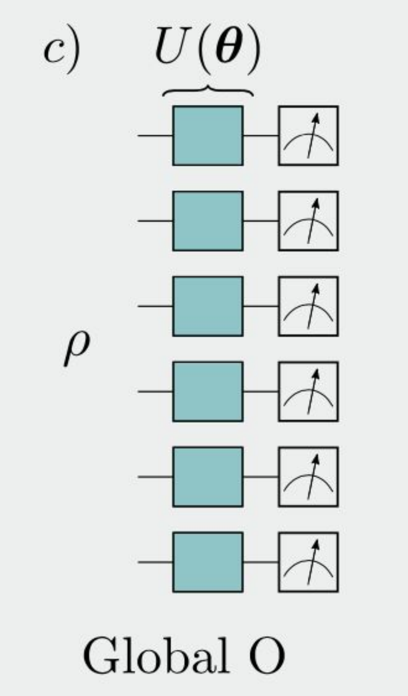
||

Small $\mathfrak{g}$-purity

||

Small variance

$$\frac{1}{\dim(\mathfrak{g})}\mathcal{P}_{\mathfrak{g}}(\rho)\mathcal{P}_{\mathfrak{g}}(O)$$



O overlaps more with DLA = larger $\mathfrak{g}$-purity

Cost function dependent barren plateaus in shallow parameterized quantum circuits, Cerezo et al (2021)

Noise-Induced Barren Plateaus

Original insight: state preparation and measurement (SPAM) and depolarizing noise cause barren plateaus

Lie unification:

Noise affects $\boldsymbol{\varrho}$, O

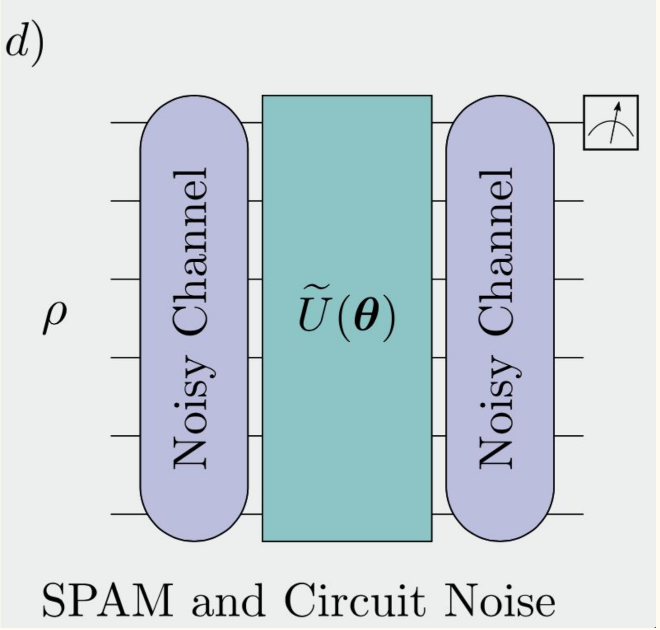
||

Small $\mathfrak{g}$-purity

||

Small variance

$$\frac{1}{\dim(\mathfrak{g})}\mathcal{P}_{\mathfrak{g}}(\rho)\mathcal{P}_{\mathfrak{g}}(O)$$



Unlike purity, local qubit unitary noise can change $\mathfrak{g}$-purity $\Rightarrow$ "*algebraic decoherence*"

Noise-Induced Barren Plateaus in Variational Quantum Algorithms Wang 2021

Open questions

- So far: every trainable model we know is also classically simulable!
  - e.g. matchgate circuits grow quadratically (orthogonal group), so trainable, but also classically simulable [1]
- Are there interesting dynamical Lie algebras in between quadratic and exponential scaling? [2]
- Relaxing the ϱ or O in $i\mathfrak{g}$ condition? [1]
- Shallow depth circuits and mixing time?

[1] *Showcasing a Barren Plateau Theory Beyond Dynamical Lie Algebras*, Diaz et al (2023)
[2] *Classification of dynamical Lie algebras for translation-invariant 2-local spin systems in one dimension* Wiersema et al (2023)