Optimization of Quantum Algorithm Protocols without Barren Plateaus

Lukas Broers\textsuperscript{1,2} and Ludwig Mathey\textsuperscript{1,2,3}

\textsuperscript{1}Center for Optical Quantum Technologies, University of Hamburg, 22761 Hamburg, Germany
\textsuperscript{2}Institute for Laser Physics, University of Hamburg, 22761 Hamburg, Germany
\textsuperscript{3}The Hamburg Center for Ultrafast Imaging, 22761 Hamburg, Germany

Abstract. Quantum machine learning has emerged as a promising method to improve near-term quantum computation devices. However, algorithmic classes such as variational quantum algorithms have been shown to suffer from barren plateaus due to vanishing gradients in their parameter spaces. We present an approach to quantum algorithm optimization that is based on trainable Fourier coefficients of Hamiltonian system parameters. Our ansatz applies to the extension of discrete quantum variational algorithms to analogue quantum optimal control schemes and is non-local in time. We demonstrate the viability of our ansatz on several objective functions using quantum natural gradient descent. In comparison to the temporally local discretization ansätze in quantum optimal control and parametrized circuits, our ansatz exhibits faster and more consistent convergence with a distinct lack of barren plateaus. We propose our ansatz as a viable parametrization candidate for near-term quantum machine learning.

INTRODUCTION

Quantum machine learning (QML) connects classical machine learning and quantum information processing. This emergent field promises new methods that advance quantum computation and has brought forth a class of approaches referred to as variational quantum algorithms (VQA) \cite{1,2,3}. In particular, noisy intermediate-scale quantum (NISQ) devices \cite{4,5,6} are predicted to benefit from the synergies with machine learning found in VQA. These approaches optimize parameters in a sequence of unitary operations, the product of which describes the time-evolution of the system. The optimization is performed with respect to a chosen observable. Examples include quantum approximate optimization algorithms \cite{7,8,9}, quantum neural networks \cite{10,11,12,13,14,15}, quantum circuit learning \cite{16,17} and quantum assisted quantum-compiling \cite{18,19,20}.

Similarly, quantum optimal control (QOC) aims to optimize the time-dependent system parameters of a quantum system to attain a given objective \cite{21,22,23}. QOC has been connected to VQA approaches, and advantages of moving from the discrete circuit picture to the underlying physical system parameters have been demonstrated \cite{22,23}. Such analogue VQA approaches commonly utilize piece-wise constant, or step-wise, parametrization ansätze \cite{24,25,26,27,28}, which behave like the Trotterized limit of very deep parametrized quantum circuits with very small actions per gate.

A major obstacle of VQAs is the existence of barren plateaus of the error surfaces, i.e. increasingly large parameter regimes with exponentially vanishing gradients \cite{29,30,31,32}. Recent publications have investigated the dependence of barren plateaus on the locality of objective functions, circuit depth, spatial and temporal locality, and expressibility of the parametrization ansätze \cite{33,34}. The relationship between the possibility of universal quantum computing in a system and its controllability has been explained \cite{35,36}. In particular, the emergence of barren plateaus is proven in parametrized quantum circuits, where the generated unitaries approximate 2-designs \cite{37,38}. However, the exact scaling behavior and emergence of barren plateaus in ansatz-agnostic VQA is unknown and overcoming this obstacle is crucial for the...
success of near-term QML technologies.

In this paper, we propose a Fourier mode parametrization ansatz for quantum information processing using generalized analogue protocols. In our ansatz we control the Fourier coefficients of the system parameters of a Hamiltonian, which constitutes a method that is non-local in time. The resulting protocols are parallel and continuous by construction, while the change in the resulting unitary transformations is smooth with respect to the Fourier coefficients. The amount of parameters, i.e. non-zero Fourier coefficients, can be restricted to reduce computational complexity, without losing the aforementioned properties. The key advantage of our ansatz is that it shows a lack barren plateaus, in contrast to the step-wise protocol ansatz.

In discrete, gate-based quantum circuits the protocols of the gate implementations are sequential, rather than parallel, and therefore contain long idling times in the Hamiltonian parameters. This is a consequence of deconstructing unitary transformations into algorithmic sequences of logical gates. Fig. 1 illustrates the levels of abstraction of quantum algorithm parametrization, for the example of the quantum Fourier transform (QFT). Our departure from the conventional quantum circuit paradigm towards a larger and more intricate space of solutions of quantum algorithms enables a computational speed-up due to parallelized Hamiltonian operations. At the same time our ansatz eliminates complications that can occur due to the discrete nature of other ansätze. Our ansatz is exclusive to analogue quantum protocols and does not translate into discrete circuit parametrizations, conventionally found in VQA.

We use measurement based quantum natural gradient (QNG) descent in a hybrid learning scheme in order to find solutions to a given objective function. Fig. 2 shows an overview of quantum hybrid learning methods. In our ansatz, we give a quantum processing unit (QPU) an input state and a set of Fourier coefficients of the parameters of the QPU’s underlying Hamiltonian, as an initial guess. We evolve the input state according to the Hamiltonian, that contains the time-dependent system. We restrict the time-evolution of \( t \in [0, 1] \) and use units in which \( \hbar = 1 \), for simplicity. In our ansatz we parametrize the Hamiltonian in terms of \( n_f \) Fourier coefficients \( \theta_{j,k} \) such that

\[
\theta_j(t) = \sum_{k=1}^{n_k} \theta_{j,k} \sin(\pi kt).
\]

In comparison, the step-wise ansatz uses the common discretization in terms of piece-wise constant system parameters

\[
\theta_j(t) = \theta_{j,k}, \quad \frac{k}{n_f} \leq t < \frac{k + 1}{n_f},
\]

with \( k = 0, \ldots, n_f - 1 \). This is reminiscent of parametrized sequential circuits. In either ansatz, we

\[
L = \sum_{\theta} (f(\theta) - g)^2.
\]

We write a general time-dependent Hamiltonian as

\[
H(t) = \sum_j \theta_j(t) H_j,
\]

where \( H_j \) are Hermitian matrices that define the system. \( \theta_j(t) \) are the parameters that determine the time-dependence of the system. We restrict the time-evolution to \( t \in [0, 1] \) and use units in which \( \hbar = 1 \), for simplicity. In our ansatz we parametrize the Hamiltonian in terms of \( n_f \) Fourier coefficients \( \theta_{j,k} \) such that

\[
\theta_j(t) = \sum_{k=1}^{n_k} \theta_{j,k} \sin(\pi kt).
\]

In comparison, the step-wise ansatz uses the common discretization in terms of piece-wise constant system parameters

\[
\theta_j(t) = \theta_{j,k}, \quad \frac{k}{n_f} \leq t < \frac{k + 1}{n_f},
\]

with \( k = 0, \ldots, n_f - 1 \). This is reminiscent of parametrized sequential circuits. In either ansatz, we
optimizing the parameters $\theta = \{\theta_{jk}\}$ with respect to some objective function $L_0$ that is constructed from observables and depends on the time-evolution operator

$$U_\theta = \hat{T} e^{-i \int_0^t \sum_j \theta_j(t) H_j dt},$$

where $\hat{T}$ indicates time-ordering. We perform the optimization via stochastic quantum natural gradient descent as detailed in the Methods section.

In this work we use the transverse Ising Hamiltonian for $n_q$ qubits

$$H(t) = \sum_{j=1}^{n_q} B^z_j(t) \sigma^z_j + \sum_{j=1}^{n_q-1} J_j(t) \sigma^x_j \sigma^x_{j+1},$$

with controllable parameters $B^z_j(t)$ and $J_j(t)$, and with the constraint $B^z_j = 0$. $\vec{\sigma}$ is a vector containing the Pauli matrices $\sigma_x$, $\sigma_y$ and $\sigma_z$ acting on the $j$th qubit.

We consider open boundary conditions, such that $J_j(t)$ takes non-zero values for $j = 1, \ldots, n_q - 1$. In total this gives $(3n_q - 1) n_q$ trainable parameters.

We first demonstrate the performance of our ansatz for the example of learning implementations of the QFT represented by the unitary operation $V$, operating on $n_q$ qubits. The matrix elements of $V$ are

$$V_{k,l} = 2^{-\frac{n_q}{2}} \exp(i \pi k l 2^{-n_q}),$$

where $k, l = 0, \ldots, 2^{n_q} - 1$. For generating unitary transformations, we utilize the objective function

$$L^U_{\theta} = 1 - \frac{1}{|\{r\}|} \sum_r |\langle r | U^\dagger_\theta V |r \rangle|^2,$$

where $\{r\}$ is a set of randomized unentangled input states, such that this objective function estimates the implementation error $\epsilon = 1 - |\text{Tr}(U^\dagger_\theta V) 2^{-n_q}|^2$ between the unitaries $U_\theta$ and $V$.

In Fig. 3 we show the estimated implementation error $\epsilon$ during training, as a function of $n_f$ for $n_q \leq 4$. We observe that both implementations converge to the target transformation for sufficiently large $n_f$. For smaller $n_f$ the accessible unitary transformations generated from the ansätze Eqs. 3 and 4 are insufficient and presumably do not contain the QFT on $n_q$ qubits.

We emphasize that our Fourier based ansatz is consistently outperforming the step-wise ansatz in terms of convergence speed. We show in Figs. 3 (a,b,c) that our ansatz tends to converge after roughly 50, 100 and 200 training iterations for $n_q = 2$, 3 and 4, respectively.

Figs. 3 (d,e,f) show that the step-wise protocol ansatz tends to converge after roughly 100, 300 and 1800 episodes for $n_q = 2, 3, 4$, respectively. For $n_q = 4$ in Fig. 3 (f), the convergence behavior of the step-wise ansatz is increasingly inconsistent. The step-wise ansatz has the tendency to linger on plateaus of suboptimal fidelity from which it only moves away slowly, due to vanishing gradients. This behavior becomes more prominent

with increasing $n_q$ and is a consequence of the error surface that follows from the parametrization in Eq. 3. Our ansatz does not show this behavior, but rather exhibits faster and more direct convergence. This indicates the absence of barren plateaus, as is apparent when comparing Figs. 3 (c,f).

In order to further evaluate the quality of the converged solutions, we show the minimal errors after training $\epsilon_{\text{opt}}$ with respect to the hyperparameter $n_f$ for both ansätze in Figs. 4 (a,b). We find the minimal $n_f$ that is necessary for convergence during training to be approximately $n_{f,\text{min}} \approx 4, 6$ and 8 for $n_q = 2, 3$ and 4, respectively. The minimal $n_f$ necessary for convergence appears to be the same for both ansätze in this example. For larger $n_f$, the minimal error converges to very small values that show no strong dependence on $n_f$. For the cases of $n_q = 3$ and $n_q = 4$, the resulting minimal error tends to approach $\epsilon_{\text{opt}} \approx 10^{-5}$. We note that for a concrete experimental realization, additional considerations, e.g. what dissipative processes are present and how well a specific parameter can be tuned dynamically, determine

![FIG. 3. Implementation errors during training of the quantum Fourier transform. The errors $\epsilon$ during training as a function of the hyperparameter $n_f$ for $n_q = 2$ (a,d), 3 (b,e) and 4 (c,f) for our Fourier based ansatz (a,b,c) and the step-wise protocol ansatz (d,e,f). For sufficiently large $n_f \geq n_{f,\text{min}}$ both ansätze converge to very small errors. Our Fourier based ansatz outperforms the step-wise ansatz in terms of convergence speed and consistency.](image-url)
As a second optimization task, we consider the energy expectation value of a problem Hamiltonian \( H \), and its minimization. Specifically, we consider the objective function

\[
\mathcal{L}_\theta^E = \langle E \rangle_\theta = \langle 0 | U_\theta^\dagger H_\theta U_\theta | 0 \rangle ,
\]

where \( U_\theta \) is the time-evolution operator of the Hamiltonian given in Eq. 4, which we use to construct the trial state \( U_\theta | 0 \rangle \). We perform this ground state search for random problem Hamiltonians for both our ansatz and the step-wise ansatz with \( n_f = 16 \). In this example we do not apply the QNG, i.e., we set the metric \( g = 1 \), for simplicity. Fig. 3 shows the energy differences to the ground state energies \( \Delta E = \langle E \rangle_\theta - E_0 \) for the training trajectories of three randomized problem Hamiltonians for up to six qubits. We again see that our ansatz outperforms the step-wise ansatz in terms of convergence speed. There is an increasing tendency of gradients to flatten out in the step-wise ansatz. This behavior is not present in our ansatz and indicates the onset of barren plateaus in the optimization of ground state preparation for step-wise protocols.

In the presence of barren plateaus the change of the objective function from one iteration to the next is suppressed, because the error surface exhibits increasingly...
large regimes of vanishing gradients. To quantify this property, we consider the variance of the gradients of the objective function with respect to the parameters $\theta_{j,k}$ for both our ansatz and the step-wise ansatz. The variance is given by

$$\text{Var}[\partial_{\theta_{j,k}} \mathcal{L}_\theta^E] = \langle (\partial_{\theta_{j,k}} \mathcal{L}_\theta^E)^2 \rangle - \langle (\partial_{\theta_{j,k}} \mathcal{L}_\theta^E)^2 \rangle,$$  \hspace{1cm} (10)

where the gradients are sampled over the parameter space of $\theta$. As a specific Hamiltonian we consider

$$H_p = \sigma_z^1 \sigma_z^2 \prod_{j=3}^{n_q} 1^j.$$

(11)

Fig. 6 (a) shows the variance $\text{Var}[\partial_{\theta_{j,k}} \mathcal{L}_\theta^E]$ with respect to the first parameter $\theta_{1,1}$. For the step-wise ansatz, the variance quickly decays for increasing $n_t$. An increasing $n_t$ represents an increased circuit depth and therefore an increasingly random unitary that follows the first timestep. In sufficiently randomized methods, the sampled transformations approach uniform distributions up to the second moment in the space of unitaries. Such asymptotes of the variance with respect to the circuit depth have been related to 2-designs of random parametrized circuits [29].

In our ansatz, increasing the hyperparameter $n_d$ does not appear to affect the variance of the gradient with respect to $\theta_{1,1}$. An increasing $n_d$ represents further Fourier modes of the system parameters that can be optimized. This does, however, not influence the gradients with respect to the lowest mode, which explains the independence with respect to $n_d$.

In Fig. 6 (b), we show the variances $\text{Var}[\partial_{\theta_{j,k}} \mathcal{L}_\theta^E]$ with respect to all $\theta_{j,k}$, $k \leq 20$, for $n_q = 3$ and $n_t = 50$, for our Fourier based ansatz. We see that for small $k$, the variance does not vanish. Note that the variances with respect to parameters that control the third qubit ($j = 5, 6$) all vanish, as a consequence of the problem Hamiltonian in Eq. 11 that acts trivially on the third qubit. Similarly, the gradients with respect to the Ising interaction between the first and second qubit ($j = 7$) show a larger variance compared to the interaction between the second and third qubit ($j = 8$). The variances with respect to all parameters $\theta_{j,k}$ in the step-wise ansatz are not depicted, since they all vanish uniformly.

CONCLUSION

We have proposed a system-agnostic ansatz of analogue variational quantum algorithms rooted in quantum optimal control. The central feature of our ansatz is that it treats the Fourier coefficients of the time-controlled system parameters of a given Hamiltonian as trainable. Therefore our ansatz is non-local in time and has no direct analogue in discretized parametrized quantum circuits. By restricting the modes to low-end frequencies we keep the amount of trainable parameters low, while also ensuring smooth quantum protocols and sufficient controllability by construction. We have applied a measurement based stochastic quantum natural gradient optimization scheme to our ansatz to generate protocols for the quantum Fourier transform for up to four qubits. Additionally, we have optimized ground state preparation processes for random problem Hamiltonians for up to six qubits. We compared the results to optimizations of the more commonly utilized step-wise parametrization ansatz.

We have demonstrated that the convergence behavior of our ansatz outperforms the step-wise protocols in speed and consistency. In particular our ansatz indicates a distinct lack of barren plateaus, which hinders the convergence of step-wise protocols and which are a known obstacle of variational quantum algorithms. At the same time, we have found the effective implementation time to be comparable in both ansätze. We have analyzed the gradient along the error surface for both ansätze, and we have shown that our ansatz shows non-vanishing vari-
ances for low frequency modes, indicating an absence of barren plateaus. These results, such as the scaling behavior associated with the absence of barren plateaus, will be elaborated on elsewhere.

In conclusion, our ansatz is a promising candidate for overcoming barren plateaus in quantum algorithm optimization and presents an alternative to parametrizations that are discrete or local in time. This approach is of direct relevance for current efforts of implementing quantum computing, as it provides realistic and efficient access to optimal quantum algorithm protocols.

METHODS

Parametrization Ansatz. Our ansatz works with any Hamiltonian that has controllable system parameters which we express in terms of its Fourier coefficients as

\[ H(t) = \sum_{j,k} \theta_{j,k} \sin(\pi kt) H_j, \]

where for simplicity and without loss of generality it is \( t \in [0,1] \). The sum over \( j \) goes over the system parameters of the Hamiltonian associated with the Hermitian operators \( H_j \). The sum over \( k \) goes up to the amount of accessible Fourier modes \( n_f \), which is a hyperparameter that is preferably tuned to be small, while still providing an accessible space of unitaries that approaches computational universality. A small choice of \( n_f \) reduces the computational demand of the algorithm and leads to more slowly varying protocols which are potentially easier to implement in real devices. We initialize the \( k \)th Fourier coefficients of any system parameter randomly between \( \pm \pi/k \), such that slow modes are emphasized from the start. The possible time-evolutions are universal, i.e. they encapsulate the entirety of unitary matrices, in the limit of large \( n_f \). In this limit the accessible unitaries of our ansatz and the step-wise ansatz converge. In the step-wise ansatz (Eq. 3) we initialize all parameters randomly between \( \pm \pi \).

Random Input States. Given a set of pairs of random angles \( \{r\} = \{\{\phi_i, \psi_i\} \in [0,2\pi)^2\}_{i=1}^{n_q} \) we define the random and unentangled quantum state

\[ |r\rangle = \otimes_{i=1}^{n_q} \cos(\frac{\phi_i}{2}) |0\rangle + e^{i\psi_i} \sin(\frac{\phi_i}{2}) |1\rangle = U_r |0\rangle \otimes_{i=1}^{n_q}. \]

We also use this as the definition of \( U_r \). Sampling the overlap \( \langle r|U^\dagger \phi U_r|r\rangle \) over such a set estimates the transformation fidelity \( |\text{Tr}(U^\dagger \phi V)|^2 \). Restricting this method to the subset of unentangled states is a choice motivated by experimental feasibility.

Quantum Natural Gradient Descent. In order to estimate the gradient of \( L_\theta \), we modify a single component \( \theta_{j,k} \) by a small amount \( \delta = 10^{-7} \). This results in slightly altered time-evolution operators \( U_{\theta + \delta \theta_{j,k}} \) and values for the objective \( L_{\theta + \delta \theta_{j,k}} \). This gives access to the finite difference estimate

\[ \frac{\partial L_\theta}{\partial \theta_{j,k}} \approx \frac{L_{\theta + \delta \theta_{j,k}} - L_\theta}{\delta}. \]

We do this for all possible \( j \) and \( k \) and write

\[ \nabla L_\theta = \sum_{j,k} \frac{\partial L_\theta}{\partial \theta_{j,k}} \theta_{j,k}. \]

The quantum natural gradient update \( \Delta \theta \) is then given by [37]

\[ g(\Delta \theta) = -\eta \nabla L_\theta \]

where \( \eta \) is a dynamical learning rate following the ADAM algorithm with standard parameters and a step-size of 0.01 [38]. The quantum natural gradient considers the underlying geometry of the parametrized states using the Fubini-Study metric \( g \) which has the components

\[ g_{(i,k)} = \text{Re}[\langle \partial_{\theta_{i,q}}\psi|\partial_{\theta_{j,l}}\psi \rangle - \langle \partial_{\theta_{i,q}}\psi|\partial_{\theta_{j,l}}\psi \rangle] \approx \text{Re}[\langle r|U^\dagger U_{\theta + \delta \theta} r\rangle - \langle r|U^\dagger U_{\theta} r\rangle \langle r|U^\dagger U_{\theta} r\rangle]. \]

The corresponding operator products are naturally expressed as longer time-evolution operators of the same form as Eq. 4 with the given parameters \( \theta \) as

\[ U^\dagger U_{\theta} = \mathcal{T}[e^{-i\sum_{j,k} \theta_{j,k} \hat{e}_{j,k} \psi(t-1)} \sin(\pi kt) H_j dt], \]

and analogously \( U_{\theta}^\dagger U_{\theta} \). \( \Theta \) is the Heaviside-function such that the parameter \( \theta_{i,q} \) is slightly altered by \( \delta \) at \( t = 1 \). The Fubini-Study metric \( g \) with respect to \( |r\rangle \) is \( U_r |0\rangle \otimes_{n_q} \) can be measured by evaluating \( \langle 0| U^\dagger U_{\theta} U_{r}^\dagger U_{\theta} U_r |0\rangle \otimes_{n_q} \). Solving the linear system of Eq. 10 yields the quantum natural gradient descent step. For very large experimental setups, determining the curvature with respect to only a select subset of \( \theta \) can be a beneficial compromise in terms of time-efficiency.

ACKNOWLEDGMENTS

This work is funded by the Deutsches Forschungszentrum (DFG, German Research Foundation) – SFB-925 – project 170620586, and the Cluster of Excellence ‘Advanced Imaging of Matter’ (EXC 2056), Project No. 390715994.
References

[1] 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, 2020.

[2] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical variational quantum algorithms. Phys. Rev. A, 92:042303, Oct 2015.

[3] 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):4213, 2014.

[4] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alpern-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heinonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. Noisy intermediate-scale quantum (nisq) algorithms, 2021.

[5] John Preskill. Quantum Computing in the NISQ era and beyond. Quantum, 2:79, August 2018.

[6] 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, Feb 2016.

[7] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. 11 2014.

[8] Leo Zhou, Sheng-Tao Wang, SooWon Choi, Hannes Pichler, and Mikhail D. Lukin. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. Phys. Rev. X, 10:021067, Jun 2020.

[9] 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), 2019.

[10] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. The quest for a quantum neural network. Quantum Information Processing, 13(11):2567–2586, 2014.

[11] Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. Nature Computational Science, 1(6):403–409, 2021.

[12] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J. Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf. Training deep quantum neural networks. Nature Communications, 11(1):808, 2020.

[13] Kunal Sharma, M. Cerezo, Lukasz Cincio, and Patrick J. Coles. Trainability of dissipative perceptron-based quantum neural networks, 2020.

[14] Iris Cong, SooWon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. Nature Physics, 15(12):1273–1278, 2019.

[15] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T. Sornborger, and Patrick J. Coles. Absence of barren plateaus in quantum convolutional neural networks, 2020.

[16] M. Cerezo and Patrick J. Coles. Higher order derivatives of quantum neural networks with barren plateaus. Quantum Science and Technology, 6(3):035006, Jun 2021.

[17] T. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. Quantum circuit learning. Phys. Rev. A, 98:032309, Sep 2018.

[18] Sumeet Khatri, Ryan LaRose, Alexander Poremba, Lukasz Cincio, Andrew T. Sornborger, and Patrick J. Coles. Quantum-assisted quantum compiling. Quantum, 3:140, May 2019.

[19] M. E. S. Morales, J. D. Biamonte, and Z. Zimborás. On the universality of the quantum approximate optimization algorithm. Quantum Information Processing, 19(9):291, 2020.

[20] Jacob Biamonte. Universal variational quantum computation. Phys. Rev. A, 103:L030401, Mar 2021.

[21] Steffen J. Glaser, Ugo Boscain, Tommaso Calarco, Christiane P. Koch, Walter Köckenberger, Ronnie Kosloff, Illya Kuprov, Burkhard Luy, Sophie Schirmer, Thomas Schulte-Herbrüggen, Dominique Sugny, and Frank K. Wilhelm. Training schrödinger’s cat: quantum optimal control. The European Physical Journal D, 69(12):279, 2015.

[22] 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, 2:010101, Jan 2021.

[23] Alexandre Choquette, Agustin Di Paolo, Panagiotis Kl. Barkoutsos, David Sénéchal, Ivano Tavernelli, and Alexandre Blais. Quantum-optimal-control-inspired ansatz for variational quantum algorithms. Phys. Rev. Research, 3:023002, May 2021.

[24] Navin Khaneja, Timo Reiss, Cindie Kehlet, Thomas Schulte-Herbrüggen, and Steffen J. Glaser. Optimal control of coupled spin dynamics: design of nmr pulse sequences by gradient ascent algorithms. Journal of Magnetic Resonance, 172(2):296–305, 2005.

[25] Marin Bukov, Alexandre G. R. Day, Dries Sels, Philip Weinberg, Anatoli Polkovnikov, and Pankaj Mehta. Reinforcement learning in different phases of quantum control. Phys. Rev. X, 8:031086, Sep 2018.

[26] Murphy Yuezhun Niu, Sergio Boixo, Vadim N. Smelyanskiy, and Hartmut Neven. Universal quantum control through deep reinforcement learning. npj Quantum Information, 5(1):33, 2019.

[27] Zhi-Cheng Yang, Armin Rahmani, Alireza Shabani, Hartmut Neven, and Claudio Chamon. Optimizing variational quantum algorithms using pontryagin’s minimum principle. Phys. Rev. X, 7:021027, May 2017.

[28] V.G. Boltzynski, R.V. Garmkulidzhe, E.F. Mishchenko, and L.S. Pontryagin. The maximum principle in the theory of optimal processes of control. IFAC Proceedings Volumes, 1(1):464–469, 1960. 1st International IFAC Congress on Automatic and Remote Control, Moscow, USSR, 1960.

[29] 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):4812, 2018.

[30] Martin Larocca, Piotr Czarnik, Kunal Sharma, Gopikrishnan Muraleedharan, Patrick J. Coles, and M. Cerezo. Diagnosing barren plateaus with tools from quantum optimal control, 2021.

[31] Andrew Arrasmith, Zoë Holmes, M. Cerezo, and...
Patrick J. Coles. Equivalence of quantum barren plateaus to cost concentration and narrow gorges, 2021.

[32] M. Cerezo, Akira Sone, Tyler Volkoff, Łukasz Cincio, and Patrick J. Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1):1791, 2021.

[33] A V Uvarov and J D Biamonte. On barren plateaus and cost function locality in variational quantum algorithms. *Journal of Physics A: Mathematical and Theoretical*, 54(24):245301, May 2021.

[34] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus, 2021.

[35] Viswanath Ramakrishna and Herschel Rabitz. Relation between quantum computing and quantum controllability. *Phys. Rev. A*, 54:1715–1716, Aug 1996.

[36] S. G. Schirmer, H. Fu, and A. I. Solomon. Complete controllability of quantum systems. *Phys. Rev. A*, 63:063410, May 2001.

[37] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum Natural Gradient. *Quantum*, 4:269, May 2020.

[38] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.