Mitigated barren plateaus in the time-nonlocal optimization of analog quantum-algorithm protocols

Lukas Broers [1,2] and Ludwig Mathey [1,2,3]

[1] *Center for Optical Quantum Technologies, University of Hamburg, 22761 Hamburg, Germany*
[2] *Institute for Quantum Physics, University of Hamburg, 22761 Hamburg, Germany*
[3] *The Hamburg Center for Ultrafast Imaging, 22761 Hamburg, Germany*

(Received 26 April 2023; accepted 15 November 2023; published 22 January 2024)

Quantum machine learning has emerged as a promising utilization of 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 parameters spaces. We present an approach to quantum algorithm optimization that is based on trainable Fourier coefficients of Hamiltonian system parameters. Our *ansatz* is exclusive to the extension of discrete quantum variational algorithms to analog quantum optimal control schemes and is nonlocal in time. We demonstrate the viability of our *ansatz* on the objectives of compiling the quantum Fourier transform and preparing ground states of random problem Hamiltonians. In comparison to the temporally local discretization *ansätze* in quantum optimal control and parametrized circuits, our *ansatz* exhibits faster and more consistent convergence. We uniformly sample objective gradients across the parameter space and find that in our *ansatz* the variance decays at a nonexponential rate with the number of qubits, while it decays at an exponential rate in the temporally local benchmark *ansatz*. This indicates the mitigation of barren plateaus in our *ansatz*. We propose our *ansatz* as a viable candidate for near-term quantum machine learning.

DOI: 10.1103/PhysRevResearch.6.013076

I. INTRODUCTION

Quantum machine learning (QML) connects classical machine learning and quantum information processing. This emergent field promises new methods that advance quantum computation [1,2] and has brought forth a class of approaches referred to as variational quantum algorithms (VQA) [3–6]. In particular, noisy intermediate-scale quantum (NISQ) devices [7–9] 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 (QAOA) [10,11], quantum neural networks [12–18], quantum circuit learning [19], and quantum-assisted quantum compiling [20–22].

Similarly, quantum optimal control (QOC) aims to optimize the time-dependent system parameters of a quantum system to attain a given objective [23–29]. 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 [30,31]. Such analog VQA approaches commonly utilize piecewise constant, or stepwise, parametrization *ansätze* [32–36], which behave like the Trotterized limit of very deep parametrized quantum circuits with very small actions per gate.

A major obstacle of VQA is the existence of barren plateaus in the loss landscapes, i.e., increasingly large regimes in the parameter space with exponentially vanishing gradients, which hinder training [37–44]. The general scaling behavior and emergence of barren plateaus is largely not understood and the dependence of barren plateaus on the details of VQA has been an active field of research in recent years. The comparison of local to global objective functions, the dependence on circuit depth, and the effects of spatial and temporal locality of parametrizations have been studied in connection to barren plateaus [40–42,45–47]. In particular, the emergence of barren plateaus has been proven in time-locally parameterized quantum circuits for global objective functions and for local objective functions in the case of nonshallow circuits [40,42,45]. Limiting the controllability of such *ansätze* can reduce the onset of barren plateaus [47–50], which constitutes a tradeoff in expressibility [51,52] in favor of trainability. This includes *ansätze* that are tailored to a given problem, such as the variational Hamiltonian *ansatz* [4,53], the unitary coupled cluster *ansatz* [54], and QAOA [11]. These results suggest that nonlocal *ansätze* that depart from the parametrized circuit paradigm may mitigate barren plateaus without the loss of generality. Overcoming the obstacle of barren plateaus is crucial for the success of near-term QML technologies.

In this paper, we propose a parametrization *ansatz* for quantum algorithm optimization using generalized analog protocols. In this *ansatz* we directly control the Fourier coefficients of the system parameters of a Hamiltonian. This constitutes a method that is nonlocal in time and is exclusive

to analog quantum protocols as it does not translate into discrete circuit parametrizations which are conventionally found in VQA. We compare our *ansatz* to the common optimal control *ansatz* of stepwise parametrizations for the example objectives of compiling the quantum Fourier transform as well as minimizing the energy of random problem Hamiltonians. This comparison shows that this Fourier-based *ansatz* results in solutions with higher fidelity and, in particular, superior convergence behavior. Note that the optimization of Fourier coefficients has been proposed for the control of molecular dynamics [55]. It has also been used in a mixed approach that optimizes in the basis of piecewise constant functions [56], as well as in phase-modulated gradient-free optimization [57]. The Fourier basis has also been used with tuned frequencies in the CRAB algorithm [23,58]. However, studies on this particular *ansatz* in the context of analog quantum computing as a natural extension of VQA appear to be lacking. We demonstrate that our *ansatz* exhibits nonexponential scaling behavior with respect to the number of qubits in the objective gradient variance, which suggests the absence of barren plateaus. We conclude that our *ansatz* is a promising candidate for efficient training and avoiding barren plateaus in VQA.

II. METHODS

In quantum circuits, the time-dependent Hamiltonian parameters that implement the gates are sequential, rather than parallel, and therefore contain long idling times. This is a consequence of deconstructing unitary transformations into algorithmic sequences of logical gates. Figure 1 illustrates different levels of abstraction of quantum algorithms. The departure from the conventional quantum circuit paradigm towards a larger and more intricate space of solutions of quantum protocols enables a computational speed-up due to parallelized Hamiltonian operations.

We write a general time-dependent Hamiltonian as

$$H(t) = \sum_j \theta_j(t) H_j, \tag{1}$$

where H_j are Hermitian matrices that define the system. $\theta_j(t)$ are the parameters that determine the time dependence of the system. The resulting time-evolution operator is formally written as

$$U_{\boldsymbol{\theta}} = \hat{T}\left[\exp\left\{-i\int_0^1 \sum_j \theta_j(t) H_j dt\right\}\right], \tag{2}$$

where $\hat{T}$ indicates time ordering. We restrict the time evolution to $t \in [0, 1]$ and use units in which $\hbar = 1$, for simplicity. The unitary transformation $U_{\boldsymbol{\theta}}$ is explicitly a function of the protocols $\theta_j(t)$. In order to perform numerical optimization, it is necessary to choose a particular parametrization for the $\theta_j(t)$.

In the *ansatz* which we highlight in this work, we parametrize the $\theta_j(t)$ in terms of the first n_f real-valued Fourier coefficients $\theta_{j,k}$ such that

$$\theta_j(t) = \sum_{k=1}^{n_\mathrm{f}} \theta_{j,k} \sin(\pi k t). \tag{3}$$

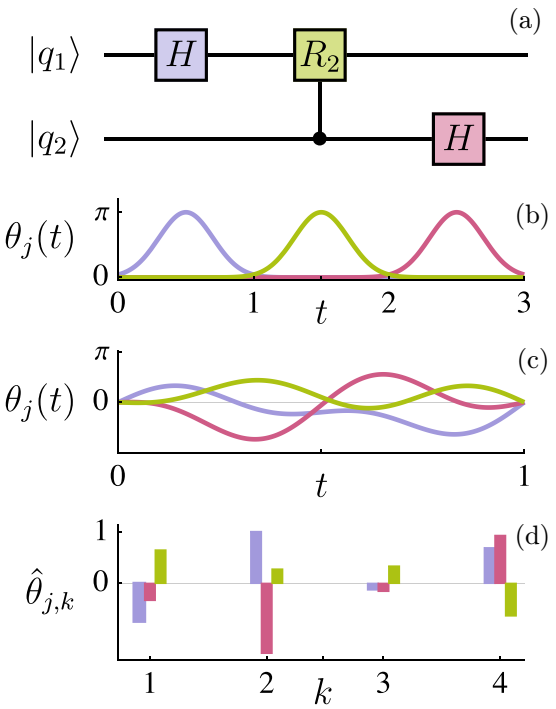

FIG. 1. Levels of abstraction of quantum algorithms. A common formulation of quantum circuits consists of a set of discrete gates (a). The physical realization of these gates consists of temporally isolated control protocols of the system parameters. These are denoted as $\theta_j(t)$ for the different parameters (b). A more efficient realization utilizes the full space of temporal evolutions of the parameters $\theta_j(t)$. This includes fully parallel protocols which take less time to complete the task (c). Any such protocol can be expressed via its Fourier coefficients $\hat{\theta}_{j,k}$, which we specifically treat as trainable parameters in our *ansatz* (d).

This *ansatz* is motivated by its inherent temporal nonlocality, as varying a single parameter $\theta_{j,k}$ changes the protocol $\theta_j(t)$ at all points in time. It presents a natural choice for a time-nonlocal parametrization that results in protocols that are smooth and slowly varying by construction, which is advantageous experimentally. We initialize the parameters $\theta_{j,k}$ randomly between $\pm\pi/k$, such that slow modes are emphasized.

In addition to our *ansatz*, we consider the stepwise *ansatz* that uses the common discretization in terms of piecewise constant system parameters,

$$\theta_j(t) = \theta_{j,k}, \frac{k}{n_\mathrm{f}} \leqslant t < \frac{k+1}{n_\mathrm{f}}, \tag{4}$$

with $k = 0, \ldots, n_\mathrm{f} - 1$. We initialize the $\theta_{j,k}$ randomly between $\pm\pi$. This *ansatz* is time local and generates discontinuous step functions with n_f steps with values $\theta_{j,k}$. These steps are reminiscent of the sequences of parametrized gates in quantum circuits as they are conventionally found in VQA. Due to its connection to conventional parametrized variational circuits, this *ansatz* serves as a benchmark to which we compare our *ansatz* of Eq. (3).

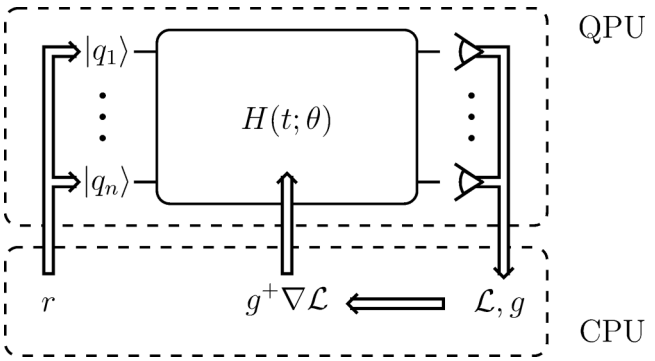

FIG. 2. Illustration of hybrid quantum optimization. A quantum processing unit (QPU) is assumed to have controllable parameters θ. Problem-specific input r is mapped onto the initial state of the qubits, which the QPU evolves in time according to the parameters θ and its underlying Hamiltonian H. The final qubit state is measured to determine the value of an objective function $\mathcal{L}_{\boldsymbol{\theta}}$ and the Fubini-Study metric g. These quantities are used on a classical machine to approximate the quantum natural gradient step to update the parameters θ and improve $\mathcal{L}_{\boldsymbol{\theta}}$.

In either *ansatz*, we optimize the parameters

$$\boldsymbol{\theta} = \sum_{j,k} \theta_{j,k} \hat{e}_{j,k} \tag{5}$$

with respect to a given objective function $\mathcal{L}_{\boldsymbol{\theta}}$, which encodes a target transformation. The exact expression of any objective $\mathcal{L}_\theta$ depends on the details of the problem it describes. The $\hat{e}_{j,k}$ are formally constructed unit vectors that collect the trainable parameters $\theta_{j,k}$ in the vector $\boldsymbol{\theta}$. Successful optimization corresponds to a time-evolution operator $U_{\boldsymbol{\theta}}$ which implements the target transformation. For a single optimization iteration, we vary the individual parameters $\theta_{j,k}$ by a small δ and evaluate the objective function to estimate the respective derivatives

$$\frac{\partial \mathcal{L}_{\boldsymbol{\theta}}}{\partial \theta_{j,k}} \approx \frac{\mathcal{L}_{\boldsymbol{\theta}+\delta\hat{e}_{j,k}} - \mathcal{L}_{\boldsymbol{\theta}}}{\delta} \tag{6}$$

such that we obtain the gradient

$$\nabla \mathcal{L}_{\boldsymbol{\theta}} = \sum_{j,k} \frac{\partial \mathcal{L}_{\boldsymbol{\theta}}}{\partial \theta_{j,k}}. \tag{7}$$

We then update the parameters as

$$\boldsymbol{\theta}_{\text{old}} \to \boldsymbol{\theta}_{\text{new}} = \boldsymbol{\theta}_{\text{old}} - \eta g^+ \nabla \mathcal{L}_{\boldsymbol{\theta}}, \tag{8}$$

where η is the learning rate, which we update dynamically using the ADAM [59] algorithm. g is the Fubini-Study metric, which contains information on the quantum geometry of the system in order to improve training behavior and makes this approach a quantum natural gradient descent method [60]. This method is illustrated in Fig. 2. For more details see the Appendix.

Note that in a physical realization, the parameters $\theta_j(t)$ cannot become arbitrarily large, and are limited by physical constraints or features of the realization. In our numerical approach, these parameters are unbounded. However, we find that these parameters remain reasonably small throughout learning, as we show below.

III. RESULTS

We compare our Fourier *ansatz* to the stepwise *ansatz* for the objectives of quantum compiling and energy minimization. Further, we evaluate the scaling behavior of the variances of objective gradients with respect to the number of qubits. Throughout this work we use the Ising Hamiltonian [61] with a two-component transverse field for n_{q} qubits as the controllable system that generates the variational unitary $U_{\boldsymbol{\theta}}$. It is

$$H(t) = \sum_{j=1}^{n_{\text{q}}} \left(B_x^j(t)\sigma_x^j + B_y^j(t)\sigma_y^j \right) + \sum_{j=1}^{n_{\text{q}}-1} J_j(t)\sigma_z^j\sigma_z^{j+1}, \tag{9}$$

with controllable parameters $B_x^j(t)$, $B_y^j(t)$ and $J_j(t)$. We consider open boundary conditions, such that the index of $J_j(t)$ goes up to $j = n_{\text{q}} - 1$. In total this gives $(3n_{\text{q}} - 1)n_{\text{f}}$ trainable parameters in $\boldsymbol{\theta}$, as the $B_x^j(t)$, $B_y^j(t)$ and $J_j(t)$ take the role of the $\theta_j(t)$ in Eq. (1). Our *ansatz* in Eq. (3) presents a general parametrization of system parameters and therefore the particular choice of the Hamiltonian is not essential. In particular, neither the Fourier *ansatz* nor the choice of the Hamiltonian are informed *a priori* by any objective at hand. They are agnostic to the optimizational tasks we utilize them for.

A. Quantum compiling

We first demonstrate the performance of our *ansatz* for the example of learning implementations of the quantum Fourier transform (QFT) represented by the unitary operation V, operating on n_{q} qubits. The matrix elements of V are

$$V_{k,l} = 2^{-\frac{n_{\text{q}}}{2}} \exp\{i2\pi kl2^{-n_{\text{q}}}\}, \tag{10}$$

where $k, l = 1, \ldots, 2^{n_{\text{q}}}$. For compiling unitary transformations, we utilize the objective function

$$\mathcal{L}_{\boldsymbol{\theta}}^U = 1 - \frac{1}{|\{r\}|} \sum_r |\langle r|U_{\boldsymbol{\theta}}^\dagger V|r\rangle|^2, \tag{11}$$

where $\{r\}$ is a set of randomized unentangled input states

$$|r\rangle = \otimes_{i=1}^{n_{\text{q}}} \left[\cos\left(\frac{\phi_i}{2}\right)|0\rangle + e^{i\psi_i} \sin\left(\frac{\phi_i}{2}\right)|1\rangle \right], \tag{12}$$

which is similar to recent methods [62]. This objective function estimates the implementation error $\epsilon = 1 - |\text{Tr}(U_\theta^\dagger V)2^{-n_{\text{q}}}|^2$ between the unitaries $U_{\boldsymbol{\theta}}$ and V. Note that there exist state estimation and tomography methods [63–67] that are experimentally favorable over the overlap in Eq. (11). Here, we use this overlap due to its straightforward numerical accessibility.

In Fig. 3, we show the estimated implementation error ϵ during training, as a function of n_{f} for $n_{\text{q}} \leqslant 4$. 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 stepwise *ansatz* in terms of convergence speed. We show in Figs. 3(a)–3(c) that our *ansatz* tends to converge after roughly 50, 100, and 200 training iterations for

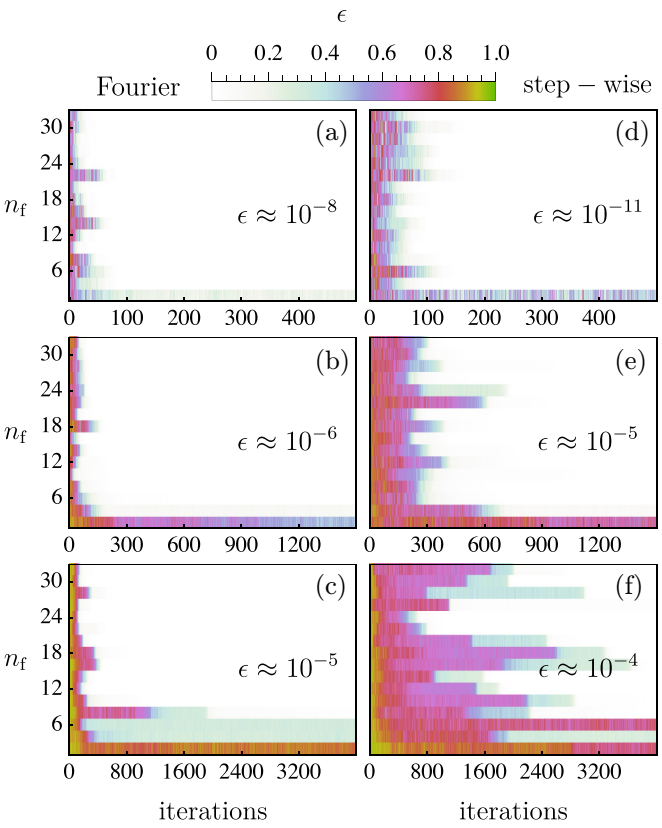

FIG. 3. Implementation errors during training of the quantum Fourier transform. The errors ϵ during training as a function of the hyperparameter n_f for the QFT for $n_q = 2$ (a and d), 3 (b and e), and 4 (c and f) for our Fourier based *ansatz* (a)–(c) and the stepwise protocol *ansatz* (d)–(f). For sufficiently large $n_f \geqslant n_{f,\min}$ both *ansätze* converge to very small errors. Our Fourier-based *ansatz* outperforms the stepwise *ansatz* in terms of convergence speed and consistency.

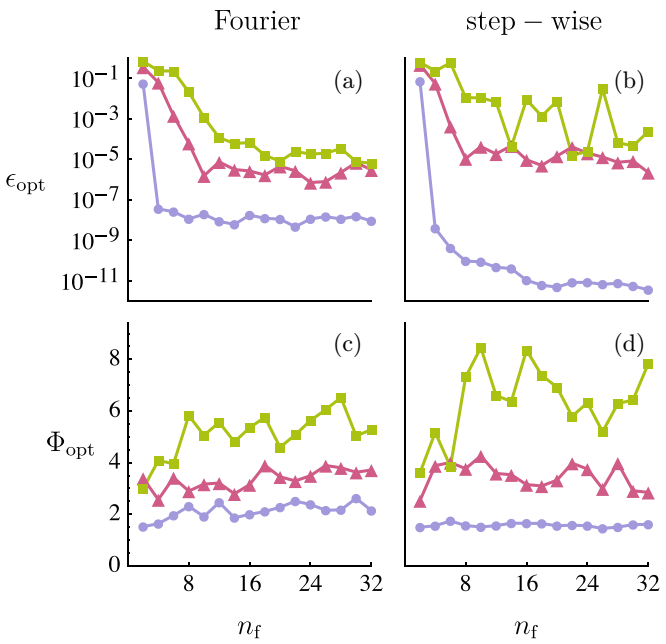

FIG. 4. Minimal errors and effective actions for training the quantum Fourier transform. The minimal errors ϵ_{opt} (a and b) found during training and the corresponding effective protocol actions Φ_{opt} (c and d) of both *ansätze*. The training results are for the QFT for $n_q = 2$ (blue circles), 3 (red triangles), and 4 (green squares). The inconsistent ϵ_{opt} in the stepwise *ansatz* for $n_q = 4$ (b) is a consequence of the suboptimal convergence behavior, related to the emergence of barren plateaus.

$n_q = 2$, 3, and 4, respectively. Figures 3(d)–3(f) show that the stepwise protocol *ansatz* tends to converge after roughly 100, 300, and 1800 iterations for $n_q = 2, 3, 4$, respectively. For $n_q = 4$ in Fig. 3(f), the convergence behavior of the stepwise *ansatz* is increasingly inconsistent. The stepwise *ansatz* has the tendency to linger at suboptimal fidelities from which it only moves away very slowly. This behavior becomes more prominent with increasing n_q and is a consequence of the loss landscape that follows from the parametrization in Eq. (4). Our *ansatz* does not show this behavior, but rather exhibits faster and more direct convergence. This is an indication for the absence of vanishing gradients, as is apparent when comparing Figs. 3(c) and 3(f).

In order to further evaluate the quality of the converged solutions, we show the minimal errors after training ϵ_{opt} with respect to the hyperparameter n_f for both *ansätze* in Figs. 4(a) and 4(b). We find the minimal n_f that is necessary for convergence during training to be approximately $n_{f,\min} \approx 4$, 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 $\epsilon_{opt} \approx 10^{-5}$. 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 the overall success of these approaches, which will be explored elsewhere.

As a second figure of merit we consider the effective implementation action, which we quantify with the integrated magnitude of the vector of system parameters $\boldsymbol{\theta}(t)$, such that

$$\Phi = \int_0^1 |\boldsymbol{\theta}(t)| \mathrm{d}t. \tag{13}$$

Given that the parameters $\theta_j(t)$ have the units of energy, this quantity is an overall measure of the phase or action that is accumulated during the time evolution. It therefore quantifies an estimate of both the energy that is required to implement a protocol in a given time, as well as the time that is required given a bound to the magnitude of the parameters $\theta_j(t)$. This figure of merit allows us to determine whether a solution with improved fidelity in our Fourier *ansatz* merely emerges due to decreased time efficiency. In Figs. 4(c) and 4(d) we show the effective actions Φ_{opt} of the same optimal solutions of Figs. 4(a) and 4(b), with respect to the hyperparameter n_f. We find the two *ansätze* to be very similar in terms of necessary action and therefore time efficiency. In both *ansätze*, there is no strong dependence on the hyperparameter n_f past $n_{f,\min}$. While the implementation actions consistently remain reasonably small, there is a clear and expected tendency of implementations to require larger effective actions with increasing amounts of qubits.

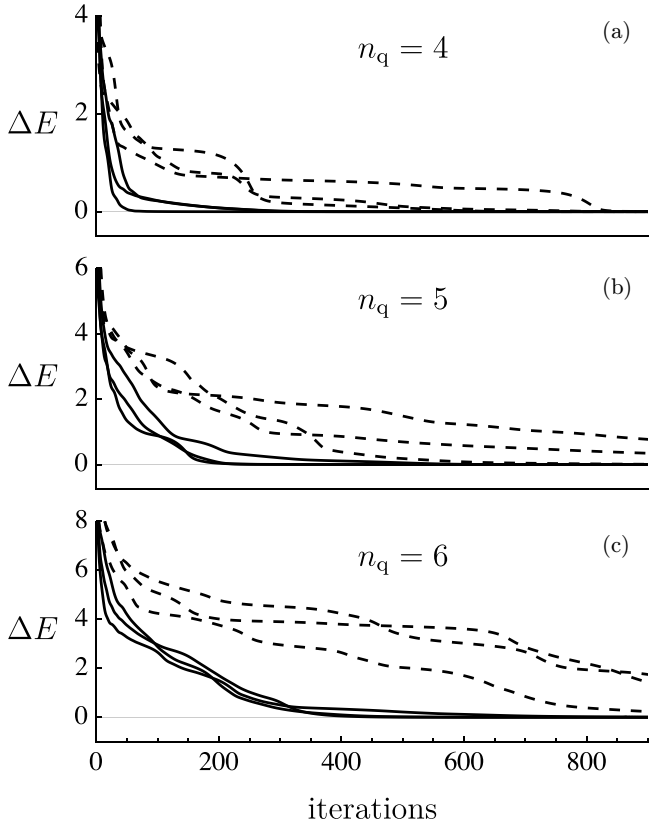

FIG. 5. Training trajectories for energy minimization. Learning trajectories for the ground state preparation of three randomly generated problem Hamiltonians for $n_q = 4, 5, 6$ qubits for our *ansatz* (solid lines) and the stepwise *ansatz* (dashed lines). $\Delta E = \langle E \rangle_{\boldsymbol{\theta}} - E_0$ is the expected energy of the prepared states relative to the ground state energy. In all cases $n_f = 16$.

B. Energy minimization

As a second optimization task, we consider the energy expectation value of a problem Hamiltonian H_p and its minimization. Specifically, we consider the objective function

$$\mathcal{L}^E_{\boldsymbol{\theta}} = \langle E \rangle_{\boldsymbol{\theta}} = \langle 0 | U^\dagger_{\boldsymbol{\theta}} H_p U_{\boldsymbol{\theta}} | 0 \rangle \,, \tag{14}$$

where $U_{\boldsymbol{\theta}}$ is the time-evolution operator of the Hamiltonian given in Eq. (2), which we use to construct the trial state $U_{\boldsymbol{\theta}} |0\rangle$. We use the shortened notation $|0\rangle = |0\rangle^{\otimes n_q}$ of the state where all qubits are in the logical zero state. We perform this ground state search for random problem Hamiltonians for both our *ansatz* and the stepwise *ansatz* with $n_f = 16$. In this example we do not apply the quantum natural gradient (QNG), i.e., we set the metric $g = \mathbb{1}$, for simplicity. Figure 5 shows the energy differences to the ground state energies $\Delta E = \langle E \rangle_{\boldsymbol{\theta}} - E_0$ for the training trajectories of three randomized problem Hamiltonians for up to six qubits. We again see that our *ansatz* outperforms the stepwise *ansatz* in terms of convergence speed. There is an increasing tendency of gradients to flatten out in the stepwise *ansatz*. This behavior is not present in our *ansatz* and indicates the onset of barren plateaus in the optimization of ground state preparation for stepwise protocols.

C. Objective gradient variances

In order to quantify the presence of barren plateaus, we consider the variance of the gradients of the objective function for both our *ansatz* and the stepwise *ansatz*. In random parametrized quantum circuits this amounts to uniformly sampling possible initializations in the parameter space of $\boldsymbol{\theta}$ [40]. In analog parametrizations of quantum algorithms, the parameter space is aperiodic and noncompact such that sampling is more intricate. We consider uniformly sampled vectors $\boldsymbol{\theta}$ inside a $(3n_q - 1)$-dimensional ball with radius $|\boldsymbol{\theta}|_{\max}$ for each time step in the stepwise *ansatz*, and $|\boldsymbol{\theta}|_{\max}/k$ for each kth Fourier mode in our *ansatz*. The value of $|\boldsymbol{\theta}|_{\max}$ determines the set of reachable states of a given *ansatz*. We consider the variance of the gradient with respect to the first parameter

$$\mathrm{Var}\big[\partial_{\theta_{1,1}} \mathcal{L}^E_{\boldsymbol{\theta}}\big] = \big\langle \big(\partial_{\theta_{1,1}} \mathcal{L}^E_{\boldsymbol{\theta}}\big)^2 \big\rangle - \big\langle \partial_{\theta_{1,1}} \mathcal{L}^E_{\boldsymbol{\theta}} \big\rangle^2 \tag{15}$$

for the specific problem Hamiltonian

$$H_p = \sigma^1_z \sigma^2_z \prod_{j=3}^{n_q} \mathbb{1}^j. \tag{16}$$

We calculate the variance as a function of $|\boldsymbol{\theta}|_{\max}$ for up to eight qubits for $n_f = 128$. Analytical arguments on the existence of barren plateaus in randomly parametrized quantum circuits (RPQCs) [40] rely on time-local expressions of the gradient of a loss function such as Eq. (14). This also applies to the stepwise *ansatz*. However, in our *ansatz* given by Eq. (3), the expression is

$$\partial_{\boldsymbol{\theta}_{1,1}} \mathcal{L}^E_{\boldsymbol{\theta}} = i \int_0^1 \sin(\pi t') \, \langle 0 | U^0_{t'} \big[U^{t'}_1 H_p U^1_{t'}, H_1 \big] U^{t'}_0 | 0 \rangle \, dt', \tag{17}$$

where U^b_a is the time-evolution operator from the time a to the time $b \geqslant a$. For $a \geqslant b$ it is $U^b_a = (U^a_b)^\dagger$. The variance of this expression includes all possible covariances of time-local changes to the protocols $\boldsymbol{\theta}(t)$, which differs substantially from the variances in RPQCs. Further, in the parameter space of $\boldsymbol{\theta}(t)$, the unitaries $U^{t'}_0$ and $U^1_{t'}$ are neither necessarily independent in the sense of the Haar measure nor guaranteed to be two designs. Therefore, the analytical argument for RPQCs [40] does not apply to our *ansatz*. In particular, the argument generates no statement about the scaling behavior.

In Fig. 6(a), we show the results of the stepwise *ansatz*. We find that the variance is independent of the amount of qubits n_q for small $|\boldsymbol{\theta}|_{\max}$. For increasing $|\boldsymbol{\theta}|_{\max}$, the variance decays exponentially with $|\boldsymbol{\theta}|_{\max}$ with slopes that are independent of n_q. More importantly, the variance decays exponentially as a function of n_q with a log-scale slope of roughly $\ln(\frac{1}{2})$, as indicated by the equally spaced lines. The stepwise *ansatz* is reminiscent of a continuous Trotterized limit of parametrized circuits and therefore these results are consistent with barren plateau studies on RPQCs [40].

In Fig. 6(b), we show the results for our *ansatz*. The variances show asymptotic behavior as functions of $|\boldsymbol{\theta}|_{\max}$. They converge at increasingly large values of $|\boldsymbol{\theta}|_{\max}$, which vastly exceed implementation actions that are necessary for highly entangling unitaries such as the QFT, as we show in Fig. 4(c). Thus, in our *ansatz* $|\boldsymbol{\theta}|_{\max}$ provides a useful hyperparameter for initialization that can be tuned to comparatively small values where the scaling with n_q is very favorable. Further,

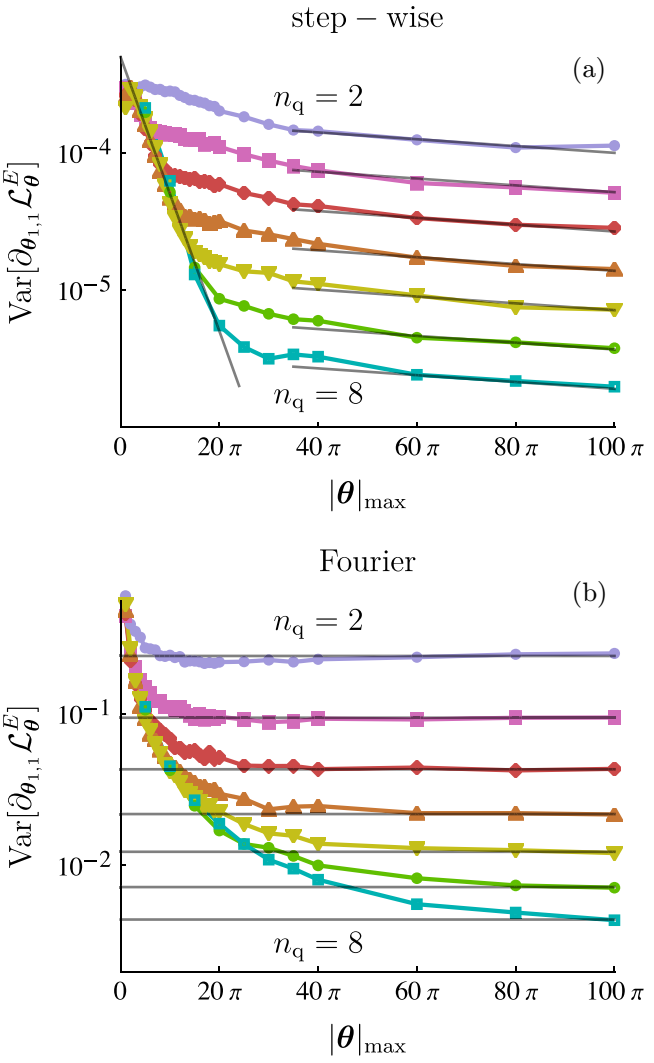

FIG. 6. Variances of the energy objective gradient. The variance of the gradient $\partial_{\theta_{1,1}} \mathcal{L}_\theta^E$ of the loss function $\mathcal{L}_\theta^E = \langle 0|U_\theta^\dagger[\sigma_z^1\sigma_z^2]U_\theta|0\rangle$ for up to eight qubits for the stepwise *ansatz* (a) and our Fourier-based *ansatz* (b) on a logarithmic scale. The parameters are sampled uniformly within a radius of $|\boldsymbol{\theta}|_{\max}$ for $n_f = 128$. The lines are visual guides.

we find that the variance decreases as a function of n_q at a decreasing and nonexponential rate. This nonexponential scaling behavior indicates the reduction of barren plateaus in our *ansatz*, in particular during initialization.

IV. CONCLUSION

We have proposed a system-agnostic *ansatz* of analog 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 nonlocal in time and has no direct analog 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 stepwise parametrization *ansatz*. The results we have presented here are limited to few-qubit systems, as the numerical simulations on the native Hamiltonian level are computationally more demanding than the circuit-based counterparts of conventional VQA. This does not translate into a lack of scalability in a true hybrid realization of the proposed method.

We have demonstrated that the convergence behavior of our *ansatz* outperforms the stepwise protocols in speed and consistency. We have found the effective implementation action to be comparable and to remain reasonably small in both *ansätze*. We have analyzed the gradient along the loss landscape for both *ansätze*, and have shown that our *ansatz* shows nonexponentially decreasing variances with respect to the amount of qubits, indicating an absence of barren plateaus. The stepwise *ansatz* shows a characteristic exponential decay with the amount of qubits that is consistent with barren plateau studies on random parametrized quantum circuits. The scaling behavior of objective gradient variances for larger systems, as well as tuning the sampling range for initialization and its relation to expressibility, will be elaborated on elsewhere.

In conclusion, our *ansatz* is a promising candidate for mitigating 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.

ACKNOWLEDGMENTS

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—SFB-925—Project No. 170620586, and the Cluster of Excellence "Advanced Imaging of Matter" (EXC 2056), Project No. 390715994.

APPENDIX: QUANTUM NATURAL GRADIENT

In order to estimate the gradient of $\mathcal{L}_{\boldsymbol{\theta}}$, we modify a single component $\theta_{j,k}$ by a small amount $\delta = 10^{-7}$. This results in slightly altered time-evolution operators $U_{\boldsymbol{\theta}}^{j,k} = U_{\boldsymbol{\theta}+\delta\hat{e}_{j,k}}$ and values for the objective $\mathcal{L}_{\boldsymbol{\theta}+\delta\hat{e}_{j,k}}$. This gives access to the finite difference estimate

$$\frac{\partial \mathcal{L}_{\boldsymbol{\theta}}}{\partial \theta_{j,k}} \approx \frac{\mathcal{L}_{\boldsymbol{\theta}+\delta\hat{e}_{j,k}} - \mathcal{L}_{\boldsymbol{\theta}}}{\delta}. \tag{A1}$$

We do this for all possible j and k and write

$$\vec{\nabla}\mathcal{L}_{\boldsymbol{\theta}} = \sum_{j,k} \frac{\partial \mathcal{L}_{\boldsymbol{\theta}}}{\partial \theta_{j,k}} \hat{e}_{j,k}. \tag{A2}$$

The quantum natural gradient update $\Delta\boldsymbol{\theta}$ is then given by [60]

$$g(\Delta\boldsymbol{\theta}) = -\eta\vec{\nabla}\mathcal{L}_{\boldsymbol{\theta}}, \tag{A3}$$

where η is a dynamical learning rate following the ADAM algorithm with standard parameters and a step size of 0.01 [59]. The quantum natural gradient considers the underlying geometry of the parametrized states using the Fubini-Study metric g, which has the components

$$g^{(j,l)}_{(i,k)} = \mathrm{Re}\big[\langle\partial_{\boldsymbol{\theta}_{i,q}}\psi|\partial_{\boldsymbol{\theta}_{j,l}}\rangle - \langle\partial_{\boldsymbol{\theta}_{i,q}}\psi|\psi\rangle\langle\psi|\partial_{\boldsymbol{\theta}_{j,l}}\psi\rangle\big] \approx \mathrm{Re}\big[\langle r|U^{\dagger,i,q}_{\boldsymbol{\theta}}U^{j,l}_{\boldsymbol{\theta}}|r\rangle - \langle r|U^{\dagger,i,q}_{\boldsymbol{\theta}}U_{\boldsymbol{\theta}}|r\rangle\langle r|U^{\dagger}_{\boldsymbol{\theta}}U^{j,l}_{\boldsymbol{\theta}}|r\rangle\big]. \tag{A4}$$

The corresponding operator products are naturally expressed as longer time-evolution operators of the same form as Eq. (2) with the given parameters $\boldsymbol{\theta}$ as

$$U^{\dagger,i,q}_{\boldsymbol{\theta}}U_{\boldsymbol{\theta}} = \hat{T}\big[e^{-i\int_0^2\sum_{j,k}(\boldsymbol{\theta}_{j,k}+\delta\hat{e}_{j,k}\hat{e}_{i,q}\Theta(t-1))\sin(\pi kt)H_j dt}\big], \tag{A5}$$

and analogously $U^{\dagger,i,q}_{\boldsymbol{\theta}}U^{j,l}_{\boldsymbol{\theta}}$ and $U^{\dagger}_{\boldsymbol{\theta}}U^{j,l}_{\boldsymbol{\theta}}$. Θ is the Heaviside-function such that the parameter $\theta_{i,q}$ is slightly altered by δ at $t = 1$. The Fubini-Study metric g with respect to $|r\rangle = U_r|0\rangle^{\otimes n}$ can be measured by evaluating $\langle 0|^{\otimes n_q}U^{\dagger}_r U^{\dagger,i,q}_{\boldsymbol{\theta}}U^{j,l}_{\boldsymbol{\theta}}U_{\boldsymbol{\theta}}U_r|0\rangle^{\otimes n_q}$. Solving the linear system of Eq. (A3) yields the quantum natural gradient descent step. For very large experimental setups, determining the curvature with respect to only a select subset of $\boldsymbol{\theta}$ can be a beneficial compromise in terms of time efficiency.

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[2] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, Challenges and opportunities in quantum machine learning, Nat. Comput. Sci. **2**, 567 (2022).

[3] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[4] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Phys. Rev. A **92**, 042303 (2015.

[5] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[6] X. Bonet-Monroig, H. Wang, D. Vermetten, B. Senjean, C. Moussa, T. Bäck, V. Dunjko, and T. E. O'Brien, Performance comparison of optimization methods on variational quantum algorithms, Phys. Rev. A **107**, 032407 (2023).

[7] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[8] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[9] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).

[10] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices, Phys. Rev. X **10**, 021067 (2020).

[11] S. Hadfield, Z. Wang, B. O'Gorman, E. G. Rieffel, D. Venturelli, and R. Biswas, From the quantum approximate optimization algorithm to a quantum alternating operator ansatz, Algorithms **12**, 34 (2019).

[12] M. Schuld, I. Sinayskiy, and F. Petruccione, The quest for a quantum neural network, Quantum Inf. Proc. **13**, 2567 (2014).

[13] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, The power of quantum neural networks, Nat. Comput. Sci. **1**, 403 (2021).

[14] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, Nat. Commun. **11**, 808 (2020).

[15] K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles, Trainability of dissipative perceptron-based quantum neural networks, Phys. Rev. Lett. **128**, 180505 (2022).

[16] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, Nat. Phys. **15**, 1273 (2019).

[17] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Absence of barren plateaus in quantum convolutional neural networks, Phys. Rev. X **11**, 041011 (2021).

[18] M. Cerezo and P. J. Coles, Higher order derivatives of quantum neural networks with barren plateaus, Quantum Sci. Technol. **6**, 035006 (2021).

[19] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018).

[20] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum-assisted quantum compiling, Quantum **3**, 140 (2019).

[21] M. E. S. Morales, J. D. Biamonte, and Z. Zimborás, On the universality of the quantum approximate optimization algorithm, Quantum Inf. Proc. **19**, 291 (2020).

[22] J. Biamonte, Universal variational quantum computation, Phys. Rev. A **103**, L030401 (2021).

[23] P. Doria, T. Calarco, and S. Montangero, Optimal control technique for many-body quantum dynamics, Phys. Rev. Lett. **106**, 190501 (2011).

[24] T. Caneva, T. Calarco, and S. Montangero, Chopped random-basis quantum optimization, Phys. Rev. A **84**, 022326 (2011).

[25] S. Lloyd and S. Montangero, Information theoretical analysis of quantum optimal control, Phys. Rev. Lett. **113**, 010502 (2014).

[26] S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, Training Schrödinger's cat: Quantum optimal control, Eur. Phys. J. D **69**, 279 (2015).

[27] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control, Phys. Rev. Lett. **118**, 150503 (2017).

[28] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm, Tunable, flexible, and efficient optimization of control pulses for practical qubits, Phys. Rev. Lett. **120**, 150401 (2018).

[29] C. P. Koch, U. Boscain, T. Calarco, G. Dirr, S. Filipp, S. J. Glaser, R. Kosloff, S. Montangero, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, Quantum optimal control in quantum technologies. strategic report on current status, visions and goals for research in Europe, EPJ Quantum Technol. **9**, 19 (2022).

[30] A. B. Magann, C. Arenz, M. D. Grace, T.-S. Ho, R. L. Kosut, J. R. McClean, H. A. Rabitz, and M. Sarovar, From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms, PRX Quantum **2**, 010101 (2021).

[31] A. Choquette, A. Di Paolo, P. Kl. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, Quantum-optimal-control-inspired ansatz for variational quantum algorithms, Phys. Rev. Res. **3**, 023092 (2021).

[32] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms, J. Magn. Reson. **172**, 296 (2005).

[33] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, Reinforcement learning in different phases of quantum control, Phys. Rev. X **8**, 031086 (2018).

[34] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, npj Quantum Inf. **5**, 33 (2019).

[35] Z.-C. Yang, A. Rahmani, A. Shabani, H. Neven, and C. Chamon, Optimizing variational quantum algorithms using Pontryagin's minimum principle, Phys. Rev. X **7**, 021027 (2017).

[36] V. G. Boltyanski, R. V. Gamkrelidze, E. F. Mishchenko, and L. S. Pontryagin, The maximum principle in the theory of optimal processes of control, IFAC Proc. Vol. **1**, 464 (1960).

[37] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).

[38] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, PRX Quantum **2**, 040316 (2021).

[39] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin, Entanglement devised barren plateau mitigation, Phys. Rev. Res. **3**, 033090 (2021).

[40] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[41] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Diagnosing barren plateaus with tools from quantum optimal control, Quantum **6**, 824 (2022).

[42] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, Equivalence of quantum barren plateaus to cost concentration and narrow gorges, Quantum Sci. Technol. **7**, 045015 (2022).

[43] T. Volkoff and P. J. Coles, Large gradients via correlation in random parameterized quantum circuits, Quantum Sci. Technol. **6**, 025008 (2021).

[44] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, Nat. Commun. **13**, 7760 (2022).

[45] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[46] A. V. Uvarov and J. D. Biamonte, On barren plateaus and cost function locality in variational quantum algorithms, J. Phys. A: Math. Theor. **54**, 245301 (2021).

[47] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, Commun. Math. Phys. **346**, 397 (2016).

[48] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX Quantum **3**, 010313 (2022).

[49] E. R. Anschuetz, Critical points in quantum generative models, arXiv:2109.06957v3.

[50] A. Pérez-Salinas, R. Draškić, J. Tura, and V. Dunjko, Reduce&chop: Shallow circuits for deeper problems, arXiv:2212.11862v2.

[51] V. Ramakrishna and H. Rabitz, Relation between quantum computing and quantum controllability, Phys. Rev. A **54**, 1715 (1996).

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

[53] C.-Y. Park and N. Killoran, Hamiltonian variational ansatz without barren plateaus, arXiv:2302.08529v1.

[54] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, J. Chem. Theory Comput. **15**, 311 (2019).

[55] K. Kormann, S. Holmgren, and H. O. Karlsson, A Fourier-coefficient based solution of an optimal control problem in quantum chemistry, J. Optim. Theory Appl. **147**, 491 (2010).

[56] Y. Song, J. Li, Y.-J. Hai, Q. Guo, and X.-H. Deng, Optimizing quantum control pulses with complex constraints and few variables through autodifferentiation, Phys. Rev. A **105**, 012616 (2022).

[57] J. Tian, H. Liu, Y. Liu, P. Yang, R. Betzholz, R. S. Said, F. Jelezko, and J. Cai, Quantum optimal control using phase-modulated driving fields, Phys. Rev. A **102**, 043707 (2020).

[58] J. Scheuer, X. Kong, R. S. Said, J. Chen, A. Kurz, L. Marseglia, J. Du, P. R. Hemmer, S. Montangero, T. Calarco, B. Naydenov, and F. Jelezko, Precise qubit control beyond the rotating wave approximation, New J. Phys. **16**, 093022 (2014).

[59] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *CoRR*, abs/1412.6980, 2015.

[60] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4**, 269 (2020).

[61] R. B. Stinchcombe, Ising model in a transverse field. I. Basic theory, J. Phys. C: Solid State Phys. **6**, 2459 (1973).

[62] M. C. Caro, H.-Y. Huang, N. Ezzell, J. Gibbs, A. T. Sornborger, L. Cincio, P. J. Coles, and Z. Holmes, Out-of-distribution generalization for learning quantum dynamics, Nat. Commun. **14**, 3751 (2023).

[63] R. Levy, D. Luo, and B. K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, arXiv:2110.02965v2.

[64] M. Neugebauer, L. Fischer, A. Jäger, S. Czischek, S. Jochim, M. Weidemüller, and M. Gärttner, Neural-network quantum state tomography in a two-qubit experiment, Phys. Rev. A **102**, 042604 (2020).

[65] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Permutationally invariant quantum tomography, Phys. Rev. Lett. **105**, 250403 (2010).

[66] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Efficient quantum state tomography, Nat. Commun. **1**, 149 (2010).

[67] R. LaRose, A. Tikku, É. O'Neel-Judy, L. Cincio, and P. J. Coles, Variational quantum state diagonalization, npj Quantum Inf. **5**, 57 (2019).