Regularizing quantum loss landscapes by noise injection

Daniil S. Bagaev,[1, 2, *] Maxim A. Gavreev,[1] Alena S. Mastiukova,[1] Aleksey K. Fedorov,[1, †] and Nikita A. Nemkov[1, ‡]

[1]*National University of Science and Technology "MISIS", Moscow 119049, Russia*
[2]*Lomonosov Moscow State University, Leninskie Gory 1 building 35, Moscow 119991, Russia*

The difficulty of training variational quantum algorithms and quantum machine learning models is well established. In particular, quantum loss landscapes are often highly non-convex and dominated by poor local minima. While this renders their training NP-hard in general, efficient heuristics that work well for typical instances may still exist. Here, we propose a protocol that uses a targeted noise injection to smooth and regularize quantum loss landscapes. It works by exponentially suppressing the high-frequency components in the Fourier expansion of the quantum loss function. The protocol can be efficiently implemented both in hardware and in simulations. We observe significant and robust improvements of solution quality across various problem types. Our method can be combined with existing techniques mitigating the local minima, such as the quantum natural gradient optimizer, and adds to the toolbox of methods for optimizing quantum loss functions.

I. INTRODUCTION AND RESULTS

Variational quantum algorithms (VQA) [1] and quantum machine learning (QML) models [2, 3] are generalizations of classical optimization methods and machine learning (ML) with the potential to harness quantum effects. The range of problems that can be formulated as VQA or QML is extremely broad. However, similarly to classical deep learning, studying VQA and QML remains largely an empirical effort, while firm theoretical performance guarantees are scarce.

One of the key problems facing VQA and QML is trainability. Two main issues here are the onset of barren plateaus (BPs) [4, 5] and proliferation of poor local minima. The BPs essentially manifest the curse of dimensionality, and most often arise in sufficiently deep circuits. Local minima, on the other hand, are typically associated with shallow quantum circuits, which may be free of BPs [6]. (Note, however, that the absence of BPs is often associated with classical simulability [7].) This work focuses on the problem of local minima in quantum loss landscapes.

On the one hand, results such as [8–10] show that the local minima problem in VQA and QML is intractable in the general case. On the other hand, there may still exist heuristics allowing for efficient training of typical instances despite the no-go results for a general worst-case. After all, training deep neural networks can be very efficient, even though the associated loss functions are typically highly non-convex [11].

In this work, we propose a heuristic regularization procedure that smoothes the quantum loss landscapes and alleviates the problem of local minima. It is based on the intuition that high-frequency terms in the Fourier expansion of the loss function are primarily responsible for the superfluous local minima, see illustration in Fig. 1a. Note
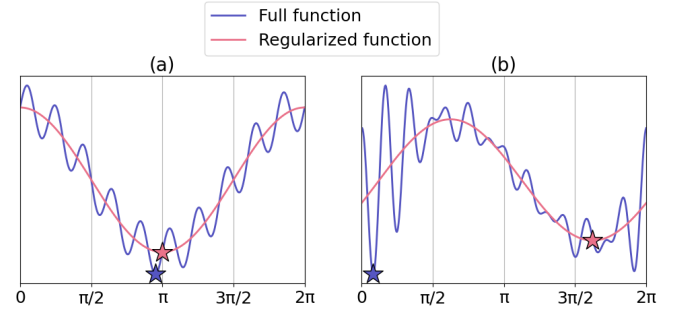
FIG. 1: A sketch showing how high-frequency terms may affect the optimization landscape. Stars mark the position of global minima. Both in (a) and (b) the high-frequency terms produce many local minima. However, in (a) the global minimum of the regularized loss function (with high frequency modes discarded) is close to the true global minimum, and can be efficiently used to warm start the optimization. In contrast, in (b) the global minimum of the full function is far from the regularized one.

that this is merely a sketch and cannot faithfully represent high-dimensional VQA landscapes. Also, Fig. 1b depicts an opposite situation, where the high-frequency modes conspire to produce a global minimum far from the global minimum of the regularized function. Essentially, our conjecture is that the latter scenario is less likely in typical quantum landscapes, and hence regularization of the high-frequency modes is beneficial for optimization.

One analogy to this intuitive picture the use of Gaussian filtering in image processing, which effectively removes high-frequency components from a signal denoising it. Another related concept is that of graduated optimization [12], where the loss landscape during the optimization is first highly smoothed, and then gradually restored to its original form. Graduated optimization have been related to optimization techniques such as simulated annealing [13] and also explored in the context of classical deep learning [14–16]. Finally, we note that the

Fourier expansion is non-locally related to the original loss function, and can thus give a window into the global properties of the landscape.

A. Results

In general, the Fourier expansion of quantum loss functions contains exponentially many terms (both in the number of qubits and in the number of parameters; see e.g. [17]), and hence is hard to access or manipulate directly.

The key technical contribution of this work is a protocol that exponentially suppresses high-frequency terms in the Fourier expansion of quantum loss functions. This regularization amounts to injecting a certain amount of noise for each parameterized gate in the quantum circuit, and can be implemented efficiently both in hardware and in simulation. In hardware, a single extra qubit is in principle sufficient to implement the necessary Pauli noise channels, though allowing for more ancillary qubits can reduce the depth of the overall circuit. In software, to simulate our protocol it is sufficient to switch to the density-matrix based simulation, though there is an associated computational cost.

In addition, we show that the regularized and original loss functions are related by the heat equation, which can give both qualitative and quantitative insight to when and why our protocol should make optimization more efficient.

For numerical experiments, we choose two types of models. The first is based on random Wishart fields, which serve as a statistical model for a large class of VQA with local minima [8, 9]. The second is the quantum convolutional neural network, which is a well-known QML model that has been shown to suffer from local minima already in small circuit sizes [9]. Our empirical results show a consistent and significant improvement in optimization metrics across all instances studied. For example, the probability to match or improve on the best solution found by the non-regularized optimization typically increases several-fold.

B. Related work

The problem of local minima in VQA and QML models has long been recognized empirically, see e.g. [18–20]. As the problem is generally NP-hard [10], all existing techniques attempting to solve it are heuristics.

A number of studies systematically explored how the use of different gradient-based and gradient-free optimizers affects the quality of VQA and QML solutions [21, 22]. Notably, quantum natural gradient [23] has been claimed to significantly alleviate the problem of local minima in variational quantum eigensolver [24], though to our knowledge it was only tested in a very limited setting and furthermore incurs significant computational overhead.

Other proposals include using classical heuristics such as differential evolution [25] and classical neural networks to warm-start the optimization [26] or regularize the loss landscape [27].

Furthermore, the potential benefits of noise to optimization of quantum loss functions have been noted previously. Refs [28, 29] argued that quantum noise can improve generalization of quantum neural networks, while [30] made a similar observation for quantum kernel methods. In [19] it was observed that a small amount of noise may prevent layer-wise training saturation in quantum approximate optimization algorithms. Ref. [31] argued that statistical sampling noise can help avoid saddle points. We also note that our protocol was inspired by work [32], which showed that quantum noise typically exponentially suppresses high-frequency terms in the Fourier expansion of quantum loss functions.

Importantly, our algorithm is complementary to most techniques mentioned above. The use of specific optimizers or warm-starts can be thought of as changing the trajectory of the basic gradient descent to help it avoid local minima and improve solution quality. In contrast, our method alters the loss landscape itself to help smooth or eliminate the superfluous local minima. As such, it can be integrated with most other mitigation methods, such as the quantum natural gradient optimizer.

II. THEORY

A. VQA and QML

A VQA is defined by a parameterized quantum circuit $U(\phi)$ (PQC) and a Hamiltonian operator H. For simplicity, we assume the initial state to be $|0\rangle$, so that the loss function is

$$L(\phi) = \langle 0|U^\dagger(\phi)HU(\phi)|0\rangle \ . \tag{1}$$

There is a variety of techniques that use PQCs to formulate QML models. We will consider the simplest case of a supervised ML problem, where the data features x_i are encoded into input vectors $|x_i\rangle$ and the model predictions are given by expectation values

$$\widehat{y}_i = \langle x_i|U^\dagger(\phi)HU(\phi)|x_i\rangle \ . \tag{2}$$

The loss function is then computed from the ground labels y_i and predictions $\widehat{y}_i$ by standard rules, e.g.

$$L(\phi) = \frac{1}{D}\sum_{i=1}^{D} l(\widehat{y}_i, y_i) \ , \tag{3}$$

where D is the dataset size and l is a classical loss function (e.g. cross-entropy).

B. Fourier expansion

We will assume that the PQC $U(\phi)$ consists only of constant Clifford gates C_k and parameterized Pauli rota-

tions $U_P(\phi_k) = e^{i\frac{\phi_k P_k}{2}} = \cos\frac{\phi_k}{2}\mathbb{I} + i\sin\frac{\phi_k}{2}P_k$

$$U(\phi) = C_{m+1}\prod_{k=1}^{m} U_{P_k}(\phi_k)C_k . \tag{4}$$

Virtually all common parameterized quantum circuits studied in modern literature have this form.

For clarity of exposition, we will first focus on the simplest VQA model (1). This loss function admits a natural Fourier-series expansion

$$L(\phi_k) = \sum_{m=0} L_m(\cos\phi_k, \sin\phi_k) . \tag{5}$$

(Note that the arguments are $\cos\phi_k$ and $\sin\phi_k$, not $\cos\frac{\phi_k}{2}$ and $\sin\frac{\phi_k}{2}$). Here L_m represent all Fourier modes of order m. They are homogenous polynomials of degree m, meaning that if all arguments of F_m are rescaled by some factor λ, the function is rescaled by λ^m

$$L_m(\lambda\cos\phi_k, \lambda\cos\phi_k) = \lambda^m L_m(\cos\phi_k, \cos\phi_k) . \tag{6}$$

Note that while the number of terms in (5) is in general exponentially large, the maximal degree is always bound by the total number of parameters in the circuit, and thus the Fourier series truncates to a Fourier polynomial.

Our main conjecture is that the highly oscillatory higher-order Fourier terms are responsible for the majority of poor local minima, and suppressing them should effectively regularize the loss landscape and increase the quality of solutions. Given this intuition, one approach is to simply discard the high-frequency terms. While straightforward conceptually, this possibility seems nontrivial to implement either in simulations or in hardware. As the number of terms in the Fourier expansion is exponentially large, directly accessing and manipulating each term is not possible beyond small scales.

C. Regularizing Fourier expansion by noise injection

Remarkably, there is a protocol that allows to exponentially suppress high-frequency modes, while admitting efficient implementation both in simulation and in hardware.

The Fourier expansion of the loss function can be computed recursively, using the following simple rule (see e.g. [17])

$$U_{P_k}(\phi_k) \circ H_\alpha = \begin{cases} H_\alpha, & (+) \\ H_\alpha + i\sin\phi_k P_k H_\alpha, & (-) \end{cases} \tag{7}$$

Here $U_{P_k}(\phi_k) \circ H_\alpha = e^{-\frac{iP_k\phi_k}{2}}H_\alpha e^{\frac{iP_k\phi_k}{2}}$ is the Heisenberg action of $U_{P_k}(\phi_k)$ and we assume that the Hamiltonian is decomposed into a sum of Pauli strings $H = \sum c_\alpha H_\alpha$, so that each H_α either commutes $(+)$ or anti-commutes $(-)$ with P_k. The Fourier expansion of the loss function

for the full Hamiltonian H is then simply a sum of the Fourier expansions for individual Pauli terms H_α.

Now, to a Pauli operator P let us associate a noise channel $\mathcal{E}_P(\mu)$ with Kraus operators $\{\sqrt{1-\frac{\mu}{2}}\mathbb{I}, \sqrt{\frac{\mu}{2}}P\}$, which acts in Heisenberg picture as

$$C_P(\mu) \circ O = \left(1 - \frac{\mu}{2}\right)O + \frac{\mu}{2}POP . \tag{8}$$

Assuming operator O either commutes or anti-commutes with P leads to

$$\mathcal{E}_P \circ O = \begin{cases} O, & OP = +PO \\ (1-\mu)O, & OP = -PO \end{cases} . \tag{9}$$

Combining (9) and (7) yields

$$\mathcal{E}_{P_k}(\mu) \circ U_{P_k}(\phi_k) \circ H_\alpha =$$
$$\begin{cases} H_\alpha, & (+) \\ (1-\mu)\left(\cos\phi_k H_\alpha + i\sin\phi_k P_k H_\alpha\right), & (-) \end{cases} . \tag{10}$$

Thus, adding to each Pauli rotation $U_P(\phi)$ the associated noise channel $\mathcal{E}_P(\mu)$ effectively rescales $\cos\phi_k$ and $\sin\phi_k$ by $1-\mu$. The homogeneity (6) then implies

$$L(\mu, \phi) = \sum_{m=0}(1-\mu)^m L_m(\phi) . \tag{11}$$

Here $L(\mu, \phi) = \langle 0|U(\mu, \phi)^\dagger H U(\mu, \phi)|0\rangle$ is the loss function of the quantum circuit $U(\mu, \phi)$ with the noise channels, see Fig. (2a).

Therefore, this simple noise-injection protocol results in a loss function with exponentially suppressed high-frequency terms, and the regularization factor $1-\mu$ is directly controlled by the noise strength.
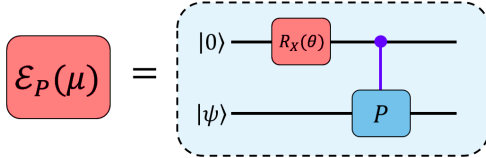
D. Implementation

In principle, a single ancilla qubit is sufficient to implement our noise injection protocol in hardware. Indeed, the Pauli noise channel $\mathcal{E}_P(\mu)$ is equivalent to applying Pauli gate P, controlled by an ancilla qubit initialized in the state $|\psi\rangle = R_X(\theta)$ with $\mu = 2\sin^2\theta/2$, see Fig. 2b. Immediately after, the ancilla qubit can be reset and reused to implement the subsequent noise channel. Using additional ancilla qubits allows applying several noise channels in parallel. The depth overhead of our protocol depends on how many ancilla qubits are available, as well as how efficient the controlled Pauli gates can be compiled given hardware restrictions.

Note that so far our discussion assumed that the original quantum circuits are noiseless, and the decoherence only happens due to the controlled entanglement with the ancilla qubits. It may be possible to use the actual quantum noise present in real hardware to achieve a similar effect [29, 32].
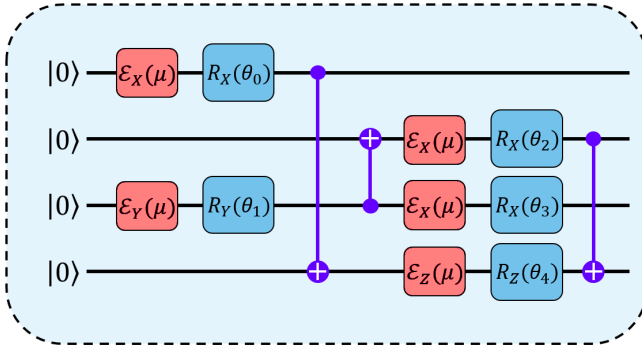
(a) A generic quantum circuit modified by injecting the Pauli noise channels according to our protocol.

(b) An possible implementation of the Pauli noise channel $\mathcal{E}(\mu)$. Here $\mu = 2\sin^2\theta/2$.

In software, a density-matrix simulation allows implementing the required noise channels directly. The downside is that the direct density-matrix simulation effectively doubles the system size with the associated increase in computational costs. More efficient approaches, e.g. based on tensor networks [33, 34] or Pauli propagation [17, 32, 35] may be preferable.

E. Heat equation

There is an interesting physical interpretation of the regularized loss function (11). Introduce a "time" parameter t so that $1 - \mu = e^{-t}$. Then, (11) satisfies the heat equation

$$\partial_t L(t,\phi) = \Delta_\phi L(t,\phi) . \quad (12)$$

Moreover, $L(0,\phi) = L(\phi)$ is the original loss function. Hence, if the initial loss function is visualized as a temperature distribution, our regularization is equivalent to letting this temperature distribution to evolve and equilibrate. The stronger the noise, the longer the effective evolution time. In the limit of $t \to \infty$ ($\mu = 1$) only the constant term in the Fourier series survives, leading to the flat landscape.

This picture can provide intuitive guidance to when our regularization prescription could be useful. For instance, if the majority of local minima are shallow, and are washed away by thermalization before the interesting solutions are erased, we expect our prescription to be effective.

III. NUMERICAL EXPERIMENTS

A. Hyperparameters

Since our method implies gradient-based optimization over a regularized landscape it requires the usual data such an optimizer, learning rate, number of iterations etc. There is an additional important ingredient, though, which we refer to as the regularization schedule. While at the beginning the landscape is strongly regularized, we need to lift the regularization at some point, since ultimately we are interested in the minima of the original loss. Removing the regularization can be performed gradually, by making the noise strength μ a function of the iteration number $\mu(i)$, which interpolates between some maximum value $\mu_{\max}$ and $\mu = 0$.

A schedule $\mu(i)$ should not fall off to zero too fast, because the optimizer needs to spend enough time in the regularized regime. The schedule should also allow enough steps for the optimizer to navigate the original non-regularized loss landscape, to reach better convergence and solution quality. Finally, the transition between the two regimes should not be too abrupt, to allow the optimizer to track the position of local minima as the landscape is deformed. Through numerical experiments, we found that most schedules satisfying these basic requirements perform similarly, see details in App. A.

For all experiments reported below, we use the ADAM optimizer [36] with learning rate 0.5×10^{-2}. The initial conditions for multi-start optimization are chosen uniformly at random $\phi \in (0, 2\pi)^m$. The noise schedule is that of an exponential decay

$$\mu(i) = \mu_{\max} e^{-a\frac{i}{i_{\max}}} \quad , \quad (13)$$

with $\mu_{\max} = 0.9$ and $a = 10$. The total number of iterations $i_{\max}$ varies and will be specified for each model. Our code is available at [37].

B. Toy example

As a warm-up example, we consider a single-layer QAOA [38] with a 5-qubit Hamiltonian engineered specifically to feature many local minima in its landscape. Since there are only two parameters in the corresponding PQC, the loss landscape can be visualized directly, as shown in Fig. 3(a). In comparison, Fig. 3(b) plots the regularized landscape ($\mu = 1/3$), which clearly has less pronounced local minima, while keeping the global minimum mostly intact. Loss landscapes are plotted with the help of ORQVIZ [39] package. We note that the connection between the QAOA data and the Fourier expansion of the loss has been studied in [40].

Fig. 4 depicts several optimization trajectories of regularized vs. non-regularized optimization starting from the same initial conditions. The background heatmap corresponds to the original loss function. We see that the
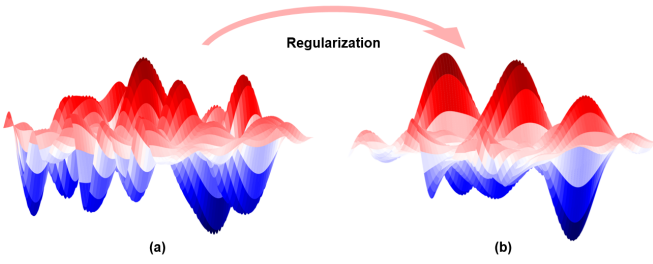
FIG. 3: A single-layer QAOA loss landscape without (a) and with (b) regularization.
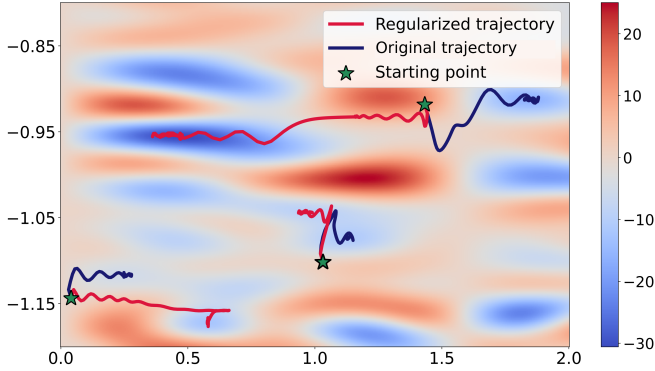

FIG. 4: Several optimization trajectories traced by the same optimizer navigating the regularized and non-regularized landscapes in our single-layer QAOA example. Only the relevant patch of the landscape is shown.

trajectories traced by the gradient descent in the regularized case are clearly different, and tend to lead to better solutions. The total number of iterations for each trajectory is $i_{max} = 2000$.

C. Statistical model: random Wishart fields

As shown in [8, 9], a large class of quantum landscapes can be described by Wishart hypertoroidal random fields (WHRFs)

$$L_{WHRF}(\phi) = w^T(\phi)Ww(\phi) . \qquad (14)$$

Here $w(\phi) = \otimes_{k=0}^{m}(\cos\frac{\theta_k}{2}, \sin\frac{\theta_k}{2})$ is a vector of dimension 2^m, and W is a matrix drawn from a Wishart ensemble, m is the number of parameters in the PQC (see App. B for details). A Wishart matrix W with d degrees of freedom can be sampled using a Gaussian matrix X with 2^m rows and d columns

$$W = \frac{1}{d}XX^\dagger . \qquad (15)$$

The statistical model (14) allows us to study generic features of quantum landscapes, without specifying exact Hamiltonian, PQC structure, or even the number of

qubits. For our purposes, the key parameter of WHRFs is the overparametrization ratio

$$\gamma = \frac{m}{2d} , \qquad (16)$$

which controls the distribution of local minima [8, 9]. Namely, for $\gamma \geq 1$ most local minima are close in value to the global minimum, while the opposite is true for $\gamma \ll 1$. Hence, small γ is the regime of interest for us here.

Since the model does not rely on an explicit quantum circuit, our noise-injection protocol does not apply directly. Instead, it is possible to use a different trick achieving the same exponential regularization of the high-frequency modes in the WHRF, see App. B for details.

To compare optimization over regularized vs non-regularized landscapes, at each value of γ we collect extensive statistics by generating 100 different Wishart matrices (15), then running 2000 optimizations with $i_{max} = 4000$ iterations, starting from random initial conditions for each landscape.

Representative histograms of the final loss values are shown in Fig. 5. Clearly, optimizations using regularization tend to converge to better solutions on average, and reach the best solutions more often. Note that the true values of the global minima are not available for WHRFs, and by the best solution we mean the best found for a given landscape.

To quantify the performance of the regularized optimization systematically, we compare the percentiles of the final loss distributions in Fig. 6. For instance, we compute the first percentile of the non-regularized distribution and see what percent of regularized loss values are below this value. Our show that the probability of finding good solutions (say from the first, or fifth percentile of the non-regularized distribution) is increased 2x to 5x on average, though there is a significant variation across different WHRF instances. Also, we note that this improvement is almost constant as we vary γ. Also, for sufficiently large γ the effect of regularization seems to diminish, probably because the optimization becomes easier for the non-regularized landscapes as well. (Fig. 9 from App. B shows how the quality of non-regularized optimization changes as we cross the $\gamma = 1$ threshold).

D. Quantum convolutional neural network

The quantum convolutional neural network (QCNN) [41] is a popular QML model that was shown to be free of BPs [6]. Although QCNN were recently shown to be classically simulable [7, 42], which undermines their practical utility, they are still an excellent test bed for our optimization technique as they were shown to feature numerous local minima already at small qubit counts.

Another appealing feature of QCNN is that the solution quality can be quantified by a single number – the
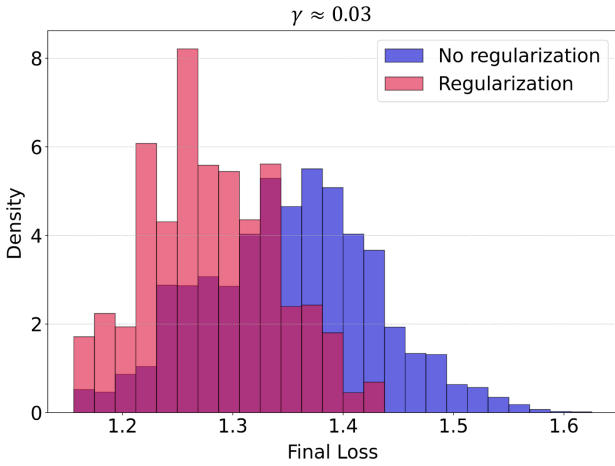
FIG. 5: An example pair of histograms showing the distribution of final loss values achieved by regularized and non-regularized optimizations. All loss values correpond to a particular WHRF with $\gamma \approx 0.03$.
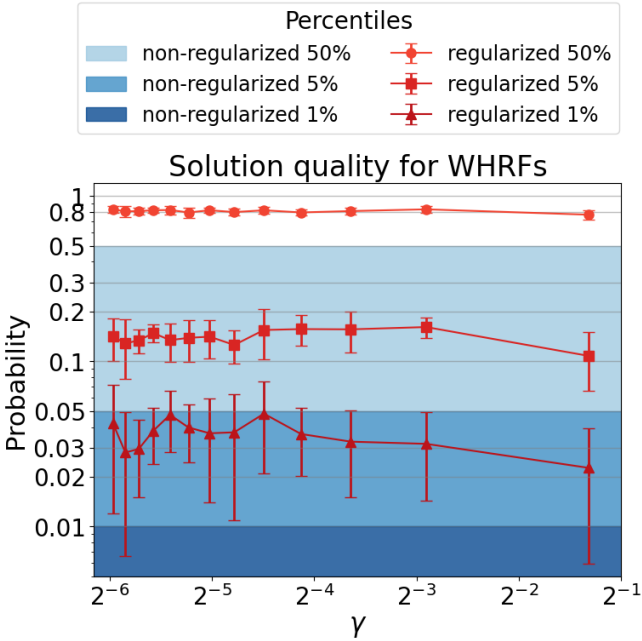

FIG. 6: Quality of the regularized optimization as a function of the overparameterization ratio γ, as measured by the value of $1\%, 5\%$ and 50% percentiles. Background colors correspond the baseline quality obtained by non-regularized optimization. Red lines show percentiles of the regularized value

accuracy of classification. The training data is generated by the teacher network, an instance of the QCNN with randomly chosen parameters ϕ_*

$$y_i = \langle x_i | U^\dagger(\phi_*) H U(\phi_*) | x_i \rangle \ , \qquad (17)$$

where the input states $|x_i\rangle$ are random computational
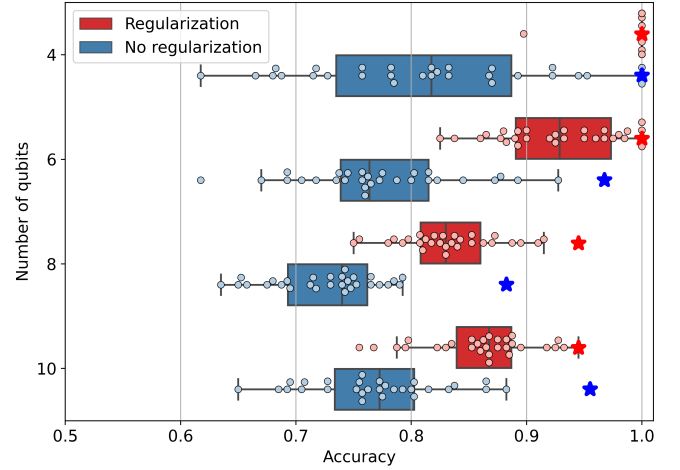

FIG. 7: Optimization results for QCNN. The highest accuracy values obtained are indicated by stars. To avoid cluttering, the regularized markers are slightly shifted up. In the four-qubit case, the regularized box plot has nearly zero width.

basis states. Then, a version of the same network, the student network $U(\phi)$, is trained to replicate the teacher network, starting from an unrelated random parameter configuration. The accuracy is evaluated on the test set, and hence the perfect accuracy can always be reached at $\phi_* = \phi$. Therefore, the train set accuracy directly reflects the difficulty of optimizing the corresponding loss function, which we take to be an MSE $l(\widehat{y}_i, y_i) = \frac{1}{D} \sum_i (\widehat{y}_i - y_i)^2$ following [9].

We perform numerical experiments for up to $n = 10$ qubits, each featuring a teacher circuit and 30 student circuits, each optimizer over $i_{\max} = 2000$ iterations. Results are shown in Fig. 7. Overall, the trend closely mirrors the metrics reported for WRHFs. Both average and best accuracies achieved by the regularized optimization are consistently higher (with the only exception being the best solution for $n = 10$ qubits, which is likely an artifact of insufficient statistics). We also note that the results for the non-regularized optimization are consistent with [9]. Our implementation is an extension of [43] supporting noisy quantum channels.

IV. DISCUSSION

We presented a protocol that uses targeted noise injection to suppress high-frequency Fourier modes, which effectively smoothes and regularizes quantum landscapes. Numerical experiments show that the regularization consistently and significantly improves the quality of optimization solutions. Moreover, as our approach deforms the landscape itself, it is complementary to most other local minima mitigation strategies, and hence can be used jointly with them. However, a number of important caveats need to be addressed to ensure that this

technique can be useful in practice.

Noise

Our procedure assumes the ability to inject variable-strength noise, reducing its value to zero as the optimization proceeds. The actual noise levels on NISQ devices [44, 45] may set a threshold that is too high, while the noise mitigation techniques may be too costly [46, 47]. Nevertheless, Ref. [29] suggested that the currently achievable noise levels may already be sufficient to enhance optimization.

Moreover, there are several scenarios where the quantum loss functions are free of the inherent noise. For instance, small-scale circuits which are easily simulable can still have extremely rough loss landscapes, which is a hindrance e.g. to variational compiling [18, 20]. In fact, many large-scale VQA and QML loss functions are amenable to efficient classical simulation as well. Some proposals imply that the loss function computation and optimization may be performed entirely classically, while the role of the quantum computer is restricted to the initial data collection or sampling from the final states [48]. Also, VQA and QML are not necessarily a NISQ concept, and may see fault-tolerant implementations in the future. In these cases, the loss landscapes can be assumed noiseless, and our technique should be directly applicable.

It is worth mentioning that the regularization by noise injection may also lead to unwanted side effects. Namely, the noise is known to be one of the causes of barren plateaus [5]. For some circuits, the noise strength that provides sufficient regularization of local minima may simultaneously induce barren plateaus. This potential problem is absent in our small-sized numerical experiments, but should be kept in mind when scaling.

Overhead

Though our regularization robustly improves optimization results, it can not guarantee reaching the desired solution quality every time, and requires an additional resource overhead to implement. Therefore, the extra cost should be compared to e.g. the cost of optimizing the non-regularized loss function multiple times starting with different initial conditions. The results will depend on a particular problem, simulation technique or hardware specs, and we leave a careful analysis of potential trade-offs for future work. We also note that this question is relevant for any other mitigation technique, though rarely discussed explicitly.

Note that in this paper we assumed a very specific noise channel for each of the parameterized gates, which leads to an exact exponential decay of the high-frequency modes. However, it is known [32] that quantum noise leads to suppression of the high-frequency modes quite

generally. Therefore, different noise injection protocols could result in similar regularization, including those that might be much more efficient to simulate or implement in hardware.

Scaling

The principle question is whether our mitigation strategy can successfully scale to large problem sizes. Ideally, this should be probed through simulations or actual experiments with practical circuits, which is currently very challenging. Though the small-scale numerical experiments reported here showed that the probability of finding good solutions can be increased several-fold, it is by no means clear that this is sufficient to solve useful problems. After all, the optimization at hand is NP-hard and its difficulty can increase exponentially with the problem size.

We note that the framework of random Wishart fields [8, 9] may provide a very useful testing ground. The WHRFs are much simpler to simulate at scale, yet should capture the universal behavior of many quantum loss functions. Remarkably, they might even enable some analytic progress, e.g. by studying how the distribution of local minima in WHRFs changes subject to the deformation through heat equation (Sec. II E).

Even if it is not possible to overcome the exponential scaling asymptotically, our and other mitigation techniques might still enable solving useful problems more efficiently. However, to the best of our knowledge, no well-defined targets have been established in this space yet; meaning no problem was shown to become feasible to solve if only the probability of finding good solutions could be increased by, say, a hundredfold. Without clear metrics to target, we felt that fine-tuning such as tailoring the learning rate schedule to the regularization schedule is premature. Nevertheless, we expect that the protocol developed in this work will be a useful addition to the quantum loss landscape optimization suite.
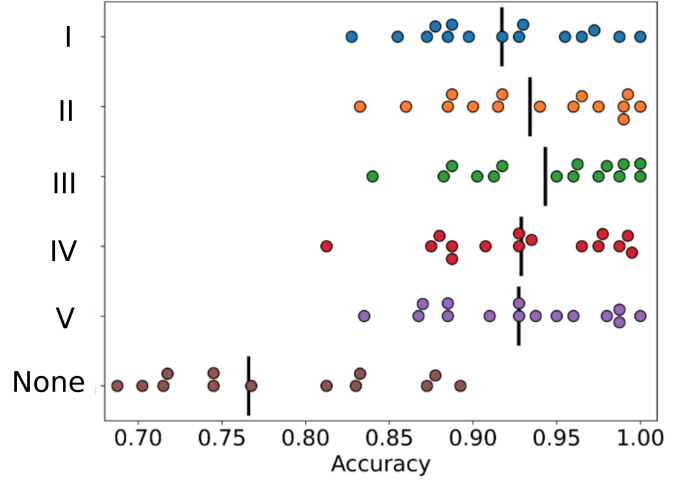
ACKNOWLEDGMENTS

Appendix A: Schedules

We tested several different regularization schedules, some examples are presented in Fig. 8a. All schedules seem to perform rather similarly on the instanced we studied, as is illustrated in Fig. 8b for the QCNN case.
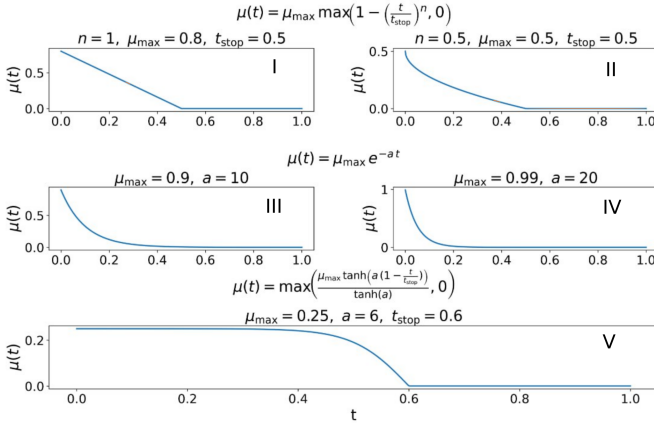
(a) Regularization schedules

(b) Effect of regularization schedule on accuracy for $n = 6$ QCNN model.

Appendix B: Wishart random fields

Loss functions of generic VQAs are expected to be well approximated by Wishart ensembles [9]

$$L(\phi) = \sum_{i,j} W_{IJ} w_I(\boldsymbol{\phi}) w_J(\boldsymbol{\phi}) . \quad (B1)$$

Here I, J are multi-indices $I = (i_1, \ldots, i_p), J = (j_1, \ldots, j_p)$. Vectors $w_I(\phi)$ are defined by

$$w_I(\boldsymbol{\phi}) = \prod_{i_k \in I} w_{i_k}(\phi_k), \qquad w_i(\phi) = \begin{cases} \cos \frac{\phi}{2}, i = 0 \\ \sin \frac{\phi}{2}, i = 1 \end{cases} \quad (B2)$$

Exponential suppression of the high-frequency modes in the Fourier expansion of (B1) can be achieved as follows. Replace

$$w_I(\boldsymbol{\phi}) w_J(\boldsymbol{\phi}) \to w_{IJ}(\lambda, \boldsymbol{\phi}) = \prod_k w_{i_k j_k}(\lambda, \phi_k) , \quad (B3)$$

where

$$w_{ij}(\lambda, \phi) = \begin{cases} \frac{1}{2}(1 + \lambda \cos \phi), & i = j = 0, \\ \frac{1}{2} \lambda \sin \phi, & i + j = 1 \\ \frac{1}{2}(1 - \lambda \cos \phi), & i = j = 1 \end{cases} . \quad (B4)$$

Note that at $\lambda = 1$ we get back the original matrix $w_I w_J$

$$w_{ij}(\lambda = 1, \phi) = \begin{cases} \cos^2 \frac{\phi}{2}, & i = j = 0, \\ \sin \frac{\phi}{2} \cos \frac{\phi}{2}, & i + j = 1 \\ \sin^2 \frac{\phi}{2}, & i = j = 1 \end{cases} , \quad (B5)$$

while at generic λ the modes of order m are multiplied by $\mu^m = (1 - \lambda)^m$.

[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". Nature Reviews Physics 2021 3:9 3, 625–644 (2021). arXiv:2012.09265.

[2] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. "Quantum machine learning". Nature 549, 195–202 (2017).

[3] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. "The effect of data encoding on the expressive power of variational quantum machine learning models". Physical Review A103 (2020). arXiv:2008.08605v2.

[4] Jarrod R. McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. "The theory of variational hybrid quantum-classical algorithms". New Journal of Physics 18, 1–20 (2016). arXiv:1509.04279.

[5] Martin Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J. Coles, Lukasz Cincio, Jarrod R. McClean, Zoë Holmes, and M. Cerezo. "A Review of Barren Plateaus in Variational Quantum Computing" (2024). arXiv:2405.00781.

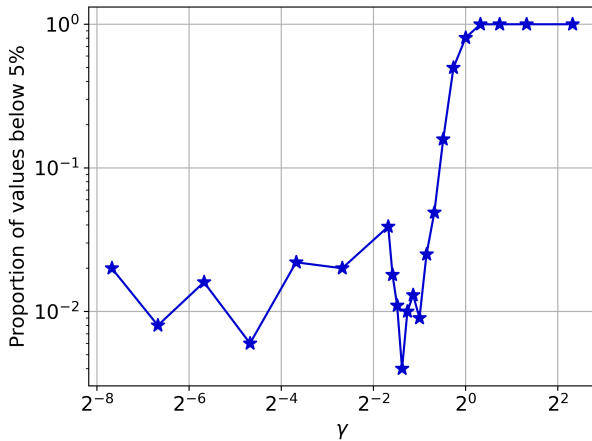[6] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T. Sornborger, and Patrick J. Coles. "Absence

FIG. 9: Fraction of loss values in the fifth percentile as a function of γ computed under discretization in 100 bins. Past $\gamma = 1$ local minima disappear from the landscape.

of barren plateaus in quantum convolutional neural networks". Physical Review X **11** (2021).

[7] M. Cerezo, Martin Larocca, Diego García-Martín, N. L. Diaz, Paolo Braccia, Enrico Fontana, Manuel S. Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, Eric R. Anschuetz, and Zoë Holmes. "Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing" (2023). arXiv:2312.09121.

[8] Eric R. Anschuetz. "Critical Points in Quantum Generative Models" (2021). arXiv:2109.06957.

[9] Eric R. Anschuetz and Bobak T. Kiani. "Beyond Barren Plateaus: Quantum Variational Algorithms Are Swamped With Traps" (2022). arXiv:2205.05786.

[10] Lennart Bittel and Martin Kliesch. "Training variational quantum algorithms is NP-hard – even for logarithmically many qubits and free fermionic systems" (2021). arXiv:2101.07267.

[11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. "Deep learning". MIT Press. (2016). url: http://www.deeplearningbook.org.

[12] Andrew Blake and Andrew Zisserman. "Visual reconstruction". The MIT Press. (1987).

[13] Hossein Mobahi and John W. Fisher. "On the Link between Gaussian Homotopy Continuation and Convex Envelopes". In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Volume 8932, pages 43–56. Springer, Cham (2015).

[14] Elad Hazan, Kfir Y. Levy, and Shai Shalev-Shwartz. "On Graduated Optimization for Stochastic Non-Convex Problems". 33rd International Conference on Machine Learning, ICML 2016 **4**, 2726–2739 (2015). arXiv:1503.03712.

[15] Da Li, Jingjing Wu, and Qingrun Zhang. "Stochastic Gradient Descent in the Viewpoint of Graduated Optimization". 33rd International Conference on Machine Learning, ICML 2016 **4**, 2726–2739 (2023). arXiv:2308.06775.

[16] Naoki Sato and Hideaki Iiduka. "Explicit and Im-

plicit Graduated Optimization in Deep Neural Networks" (2024). arXiv:2412.11501.

[17] Nikita A. Nemkov, Evgeniy O. Kiktenko, and Aleksey K. Fedorov. "Fourier expansion in variational quantum algorithms". Phys. Rev. A **108**, 032406 (2023).

[18] Bobak Toussi Kiani, Seth Lloyd, and Reevu Maity. "Learning Unitaries by Gradient Descent" (2020). arXiv:2001.11897.

[19] E. Campos, D. Rabinovich, V. Akshay, and J. Biamonte. "Training Saturation in Layerwise Quantum Approximate Optimisation" (2021). arXiv:2106.13814.

[20] Nikita A. Nemkov, Evgeniy O. Kiktenko, Ilia A. Luchnikov, and Aleksey K. Fedorov. "Efficient variational synthesis of quantum circuits with coherent multi-start optimization". Quantum **7**, 993 (2023). arXiv:2205.01121.

[21] Owen Lockwood. "An Empirical Review of Optimization Techniques for Quantum Variational Circuits" (2022). arXiv:2202.01389.

[22] Xavier Bonet-Monroig, Hao Wang, Diederick Vermetten, Bruno Senjean, Charles Moussa, Thomas Bäck, Vedran Dunjko, and Thomas E. O'Brien. "Performance comparison of optimization methods on variational quantum algorithms". Physical Review A **107**, 032407 (2023). arXiv:2111.13454v3.

[23] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. "Quantum Natural Gradient". Quantum **4** (2020). arXiv:1909.02108.

[24] David Wierichs, Christian Gogolin, and Michael Kastoryano. "Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer". Physical Review Research **2** (2020). arXiv:2004.14666.

[25] Daniel Faílde, José Daniel Viqueira, Mariamo Mussa Juane, and Andrés Gómez. "Using Differential Evolution to avoid local minima in Variational Quantum Algorithms". Scientific Reports 2023 13:1 **13**, 1–10 (2023). arXiv:2303.12186.

[26] Shikun Zhang, Zheng Qin, Yongyou Zhang, Yang Zhou, Rui Li, Chunxiao Du, and Zhisong Xiao. "Diffusion-Enhanced Optimization of Variational Quantum Eigensolver for General Hamiltonians" (2025). arXiv:2501.05666.

[27] Javier Rivera-Dean, Patrick Huembeli, Antonio Acín, and Joseph Bowles. "Avoiding local minima in Variational Quantum Algorithms with Neural Networks" (2021). arXiv:2104.02955.

[28] Nam H. Nguyen, Elizabeth C. Behrman, and James E. Steck. "Quantum Learning with Noise and Decoherence: A Robust Quantum Neural Network". Quantum Machine Intelligence **2**, 1 (2016). arXiv:1612.07593.

[29] Wilfrid Somogyi, Ekaterina Pankovets, Viacheslav Kuzmin, and Alexey Melnikov. "Method for noise-induced regularization in quantum neural networks" (2024). arXiv:2410.19921.

[30] Valentin Heyraud, Zejian Li, Zakari Denis, Alexandre Le Boité, and Cristiano Ciuti. "Noisy quantum kernel machines". Physical Review A **106**, 052421 (2022). arXiv:2204.12192.

[31] Junyu Liu, Frederik Wilde, Antonio Anna Mele, Liang Jiang, and Jens Eisert. "Stochastic noise can be helpful for variational quantum algorithms" (2022). arXiv:2210.06723.

[32] Enrico Fontana, Manuel S Rudolph, Ross Duncan, Ivan Rungger, and Cristina Cîrstoiu. "Classical sim-

ulations of noisy variational quantum circuits" (2023). arXiv:2306.05400.

[33] Yi-Ting Chen, Collin Farquhar, and Robert M. Parrish. "Low-rank density-matrix evolution for noisy quantum circuits". npj Quantum Information **7**, 61 (2021). arXiv:2009.06657.

[34] Anthony P. Thompson, Arie Soeteman, Chris Cade, and Ido Niesen. "Non-zero noise extrapolation: accurately simulating noisy quantum circuits with tensor networks" (2025). arXiv:2501.13237.

[35] Tomislav Begušić, Kasra Hejazi, and Garnet Kin-Lic Chan. "Simulating quantum circuit expectation values by Clifford perturbation theory" (2023). arXiv:2306.04797.

[36] Diederik P. Kingma and Jimmy Lei Ba. "Adam: A method for stochastic optimization". 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track ProceedingsPages 1–15 (2015). arXiv:1412.6980.

[37] D. Bagaev (2024). url: github.com/quantumoon/noise-induced-optimization.

[38] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. "A quantum approximate optimization algorithm" (2014). arXiv:1411.4028.

[39] Manuel S. Rudolph, Sukin Sim, Asad Raza, Michal Stechly, Jarrod R. McClean, Eric R. Anschuetz, Luis Serrano, and Alejandro Perdomo-Ortiz. "Orqviz: Visualizing high-dimensional landscapes in variational quantum algorithms" (2021). arXiv:2111.04695.

[40] Michał Stęchły, Lanruo Gao, Boniface Yogendran, Enrico Fontana, and Manuel Rudolph. "Connecting the Hamiltonian structure to the QAOA energy and Fourier landscape structure" (2023). arXiv:2305.13594.

[41] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. "Quantum convolutional neural networks". Nature Physics **15**, 1273–1278 (2019).

[42] Pablo Bermejo, Paolo Braccia, Manuel S. Rudolph, Zoë Holmes, Lukasz Cincio, and M. Cerezo. "Quantum Convolutional Neural Networks are (Effectively) Classically Simulable" (2024). arXiv:2408.12739.

[43] B. Kiani (2024). url: github.com/bkiani/Beyond-Barren-Plateaus.

[44] John Preskill. "Quantum computing in the NISQ era and beyond". Quantum **2**, 1–20 (2018). arXiv:1801.00862.

[45] A. K. Fedorov, N. Gisin, S. M. Beloussov, and A. I. Lvovsky. "Quantum computing at the quantum advantage threshold: a down-to-business review" (2022). arXiv:2203.17181.

[46] Alexey Uvarov, Daniil Rabinovich, Olga Lakhmanskaya, Kirill Lakhmanskiy, Jacob Biamonte, and Soumik Adhikary. "Mitigating Quantum Gate Errors for Variational Eigensolvers Using Hardware-Inspired Zero-Noise Extrapolation". Physical Review A**110** (2023). arXiv:2307.11156v3.

[47] Wanqi Sun, Jungang Xu, and Chenghua Duan. "Noise-Mitigated Variational Quantum Eigensolver with Pre-training and Zero-Noise Extrapolation" (2025). arXiv:2501.01646.

[48] Zoltán Zimborás, Bálint Koczor, Zoë Holmes, Elsi-Mari Borrelli, András Gilyén, Hsin-Yuan Huang, Zhenyu Cai, Antonio Acín, Leandro Aolita, Leonardo Banchi, Fernando G. S. L. Brandão, Daniel Cavalcanti, Toby Cubitt, Sergey N. Filippov, Guillermo García-Pérez, John Goold, Orsolya Kálmán, Elica Kyoseva, Matteo A. C. Rossi, Boris Sokolov, Ivano Tavernelli, and Sabrina Maniscalco. "Myths around quantum computation before full fault tolerance: What no-go theorems rule out and what they don't" (2025). arXiv:2501.05694.