Overcoming Barren Plateaus in Variational Quantum Circuits using a Two-Step Least Squares Approach

Francis Boabang and Samuel Asante Gyamerah

***Abstract*—Variational Quantum Algorithms (VQAs) represent a cornerstone of near-term quantum computing, enabling hybrid quantum–classical optimization for high-dimensional learning, chemistry, and combinatorial problems. However, their scalability is severely constrained by the barren plateau phenomenon, a regime where gradients vanish exponentially with system size, preventing effective training of deep or randomly initialized circuits. In this work, we introduce a two-stage optimization framework designed to overcome barren plateaus and improve convergence stability in quantum and quantum-inspired models. The proposed method begins with a convex initialization stage, which shapes the quantum energy manifold (Hilmaton landscape) into a smooth, low-energy basin to enhance gradient visibility and suppress noise-induced instability. Once stable gradient flow is established, the algorithm transitions to a nonconvex refinement stage that introduces structured oscillatory regularization, enabling exploration of multiple energy minima and improving model expressivity. This staged approach emulates the physical evolution of quantum systems from ordered to interference-rich states, providing a principled pathway to maintain gradient information throughout training. Numerical experiments demonstrate that the framework mitigates gradient vanishing, enhances optimization stability, and preserves generalization performance across varying circuit depths. The results highlight a scalable route toward trainable variational quantum architectures resilient to barren plateau effects. https://github.com/boabangf/Variational-Quantum-Algorithms-Quantum-Machine-Learning**

***Index Terms*—Quantum-inspired Machine Learning, Convex Optimization, Nonconvex Dynamics, Barren Plateau, Hamiltonian**

I. Introduction

Quantum variational algorithms (QVAs) have emerged as a leading paradigm for near-term quantum computation, offering hybrid quantum–classical optimization frameworks for solving high-dimensional learning, chemistry, and optimization problems. By parameterizing quantum circuits and minimizing an energy-based cost function, these methods exploit the expressive power of quantum superposition and entanglement while relying on classical optimizers for parameter updates. Despite their promise, the scalability of these algorithms is fundamentally constrained by the barren plateau phenomenon, a regime in which gradients vanish exponentially with system size, rendering training infeasible for deep or randomly initialized circuits [7].

This vanishing-gradient behavior not only impedes convergence, but also obscures meaningful parameter updates, leading to unstable optimization and poor generalization in practical implementations.

From the perspective of the Hamilton (or Hilmaton) landscape, barren plateaus correspond to extended regions of near-zero curvature where optimization becomes effectively directionless. Within this quantum energy manifold, saddle points represent critical configurations in which the gradient vanishes while curvature changes sign flat along one axis yet steep along another resulting in unstable equilibria that trap the optimizer between local minima and maxima [7].

As demonstrated in prior studies, such saddle regions are particularly detrimental in parameterized quantum circuits because of their nonlinear phase couplings and destructive interference patterns, which distort gradient flow and cause erratic training behavior. Moreover, the presence of hardware noise and decoherence amplifies these effects, especially in deeper ansatz, flattening the energy surface further and accelerating the onset of gradient vanishing [7].

To overcome these limitations, it is essential to design optimization frameworks that preserve gradient information and maintain a stable descent trajectory through the rugged quantum energy landscape. Recent advances in quantum machine learning have revealed that structured initialization and staged optimization can prevent premature collapse into barren regions by guiding the system toward smoother, low-energy basins before engaging the full nonconvex dynamics of the model [7].

Motivated by this insight, we propose a two-stage optimization framework that combines a convex initialization stage to promote gradient visibility and suppress random noise with a nonconvex refinement stage that enhances expressivity and enables exploration across multiple quantum attractors. The convex phase initializes the Hilmaton in a smooth region of the landscape, ensuring well-conditioned gradient propagation, while the subsequent nonconvex stage introduces structured oscillations that emulate quantum interference and tunneling, improving convergence robustness.

This hierarchical approach mirrors the physical progression of quantum systems from stable, low-energy configurations to complex, interference-rich dynamics. By embedding this transition directly into the optimization process, the proposed framework effectively mitigates barren plateaus, stabilizes convergence, and maintains the expressive power necessary for scalable quantum variational learning.

In summary, the contributions of the paper are as follows:

1) The paper introduces a two-stage optimization framework which comprise of a Convex Initialization Stage

Francis Boabang is with the Concordia Institute for Information and Systems Engineering (CIISE), Concordia University, Montréal, QC, Canada.

Samuel Asante Gyamerah is with the Department of Mathematics, Toronto Metropolitan University, Toronto, Ontario, Canada. (Corresponding author: asante.gyamerah@torontomu.ca)

and a Nonconvex Refinement Stage to overcome the barren plateau phenomenon in variational quantum circuits.
2) Comprehensive experiments across synthetic regression (curve fitting), telecom autoregressive forecasting, and reduced MNIST classification demonstrate that the proposed Convex to Nonconvex framework achieves faster convergence, smoother loss decay, and improved generalization compared to baseline QiML (no initialization) and Sellier (2023) methods.

II. Related Works

Recent efforts in quantum machine learning have identified the *barren plateau problem* as a fundamental limitation in training variational quantum circuits (VQCs). McClean *et al.* [5] first demonstrated that randomly initialized, deep parameterized quantum circuits exhibit exponentially vanishing gradients, rendering global optimization infeasible for large qubit systems. This result motivated a growing body of research seeking to characterize and mitigate barren plateaus.

Cerezo et al. (2021) [2] presented a comprehensive theoretical study establishing that the appearance of barren plateaus in variational quantum algorithms (VQAs) is strongly dependent on the choice of cost function. They demonstrate that when cost functions are defined using global observables, the gradient variance decays exponentially with system size even for shallow circuits resulting in an untrainable flat landscapes. In contrast, local cost functions, which act on limited subsets of qubits, exhibit only polynomially vanishing gradients, enabling scalable training. Using rigorous analytical derivations and numerical simulations, the authors introduce a framework that quantifies this relationship between locality and trainability, providing upper and lower bounds for gradient variance. Their results offer a foundational principle for designing quantum neural networks that remain trainable at larger qubit counts, highlighting that local observables are essential to avoid barren plateaus and preserve meaningful gradient information during optimization.

Holmes *et al.* [4] further explored the relationship between circuit expressibility and optimization difficulty. They showed that as a quantum model becomes more expressive (i.e., capable of approximating arbitrary unitaries), the gradient landscape becomes increasingly flat, leading to degraded trainability. This expressibility and trainability trade-off complements the locality-based analysis of Cerezo *et al.*, suggesting that moderate-depth, partially expressive circuits are optimal for stable training.

In similar spirits, Sellier [6] proposes a quantum-inspired learning framework that implicitly mitigates the barren plateau problem by replacing gradient-based training with an energy-minimization process governed by the time-dependent Schrodinger equation. In this formulation, each model parameter evolves as a quantum wavefunction toward its lowest-energy configuration, enabling smooth convergence without relying on stochastic gradients that typically vanish in high-dimensional quantum circuits. The approach leverages quantum tunneling effects to escape flat or shallow regions of the loss landscape and introduces density functional theory–inspired coupling potentials to stabilize optimization and reduce sensitivity to initialization. By eliminating the dependence on random starting points and using deterministic energy evolution, Sellier's framework naturally avoids the gradient vanishing and instability characteristic of barren plateaus, offering a convex-like early phase followed by nonconvex refinement conceptually aligned with our proposed multistage Hilmaton optimization method. It should be noted that gradient-based methods provide explicit information about the magnitude and direction of improvement for each parameter, allowing fine-grained control over optimization. In contrast, the energy-based quantum evolution in Sellier's approach lacks this directional feedback, which may slow convergence or lead to oscillatory trajectories around minima.

III. Method Overview

We propose a two-stage optimization framework for quantum and quantum-inspired machine learning models, designed to mitigate the barren plateau problem and improve convergence stability. This approach combines a convex initialization stage for stable gradient formation with a subsequent nonconvex refinement stage for enhanced expressivity and exploration of the Hilmaton landscape (the quantum energy manifold).

A. Convex Initialization Stage

The initial stage initializes the Hilmaton landscape in a smooth, convex region to prevent random initialization collapse and gradient vanishing. This stage optimizes a surrogate convex loss or local observable function to achieve a low-energy basin before switching to full nonconvex optimization.

Objective Function

A convex local loss is defined as:

$$\mathcal{L}_{\text{convex}}(\boldsymbol{\theta}) = \sum_{i \in \mathcal{R}} \|\langle \psi(\boldsymbol{\theta}) | O_i | \psi(\boldsymbol{\theta}) \rangle - b_i\|_2^2 , \tag{1}$$

where $\mathcal{R}$ denotes a restricted set of local observables, O_i represents the local measurement operator, and b_i are the corresponding target expectations.

Alternatively, a ridge regression-inspired initialization [6], [3], [1], [8] can be used:

$$\min_{\mathbf{w}} \ \|\Phi \mathbf{w} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{w}\|_2^2, \tag{2}$$

where Φ is a feature matrix constructed from the quantum circuit and $\mathbf{y}$ are the target values. The optimal weights $\mathbf{w}^*$ provide an initialization for $\boldsymbol{\theta}$.

Switching Criterion

Transition to the next stage occurs when gradient norms stabilize or surpass a threshold:

$$\frac{\|\nabla_{\boldsymbol{\theta}} \mathcal{L}_{\text{convex}}\|_2}{\dim(\boldsymbol{\theta})} > \tau_g, \tag{3}$$

ensuring that parameters have escaped the flat gradient regime associated with barren plateaus [5], [2].

B. Nonconvex Refinement Stage

After initialization, the model transitions into a nonconvex landscape to exploit its expressive power and explore multiple basins of attraction. Structured oscillations are introduced to enhance exploration and avoid premature convergence.

Objective Function

The refinement loss blends the convex surrogate, task loss, and oscillatory regularization:

$$\mathcal{L}_{\text{refine}}(\boldsymbol{\theta}) = (1-\lambda)\,\langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle + \lambda\,\mathcal{L}_{\text{convex}}(\boldsymbol{\theta}) + \beta\,\mathcal{R}_{\text{osc}}(\boldsymbol{\theta}), \tag{4}$$

where H is the Hamiltonian, and the oscillatory regularizer introduces controlled nonconvexity:

$$\mathcal{R}_{\text{osc}}(\boldsymbol{\theta}) = \mathbb{E}_j\left[\sin(\omega\,\theta_j + \phi_j)\right]. \tag{5}$$

Parameters λ and β are gradually annealed to 0 to prioritize the physical objective in later epochs:

$$\lambda_{t+1} = \max(0, \lambda_t \cdot 0.98), \tag{6}$$

$$\beta_{t+1} = \beta_t \cdot 0.99. \tag{7}$$

The convex stage enforces gradient concentration by constraining expressibility [4], while the nonconvex stage enhances representation power through phase coupling and interference. Together, they form a hybrid curriculum that mimics the quantum optimization trajectory from smooth (convex) to rugged (nonconvex) Hilmaton surfaces, as shown in Figure 1. The proposed method leverages convex initialization to suppress barren plateaus and enhance gradient visibility. Then, the nonconvex refinement to increase exploration and fine-tune the energy minima. After that, structured oscillations to improve sampling of multiple quantum states. This hybrid approach aligns with recent findings that link expressibility to trainability and cost-function locality to gradient stability in quantum and quantum-inspired models [4].

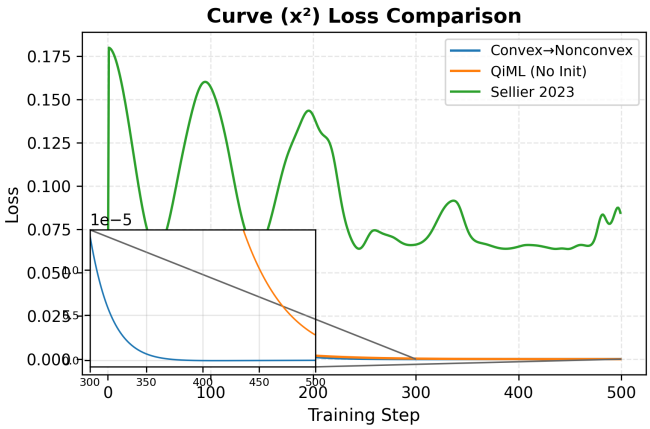
IV. Evaluation

A. Experimental Setup

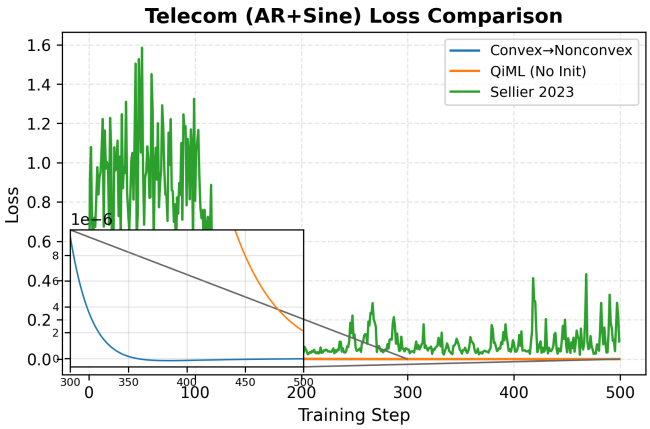
All experiments were conducted using `PyTorch 2.1` on an NVIDIA RTX 3090 GPU. We compared three training strategies:

1) Convex to Nonconvex Refinement (Ours)
2) QiML (No Initialization)
3) Sellier JM. [6])

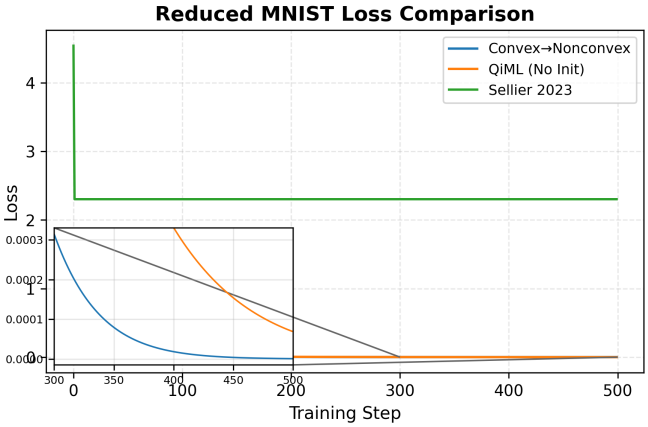
To ensure fairness, all models were initialized using the same random seed and identical weight parameters via a Xavier–Tanh initialization scheme. Each optimizer used the same mini-batch order through a fixed pseudo-random generator, guaranteeing identical sampling during training. We evaluated both regression and classification tasks to demonstrate the generality of the proposed convex–nonconvex refinement mechanism.



(a) Curve (x^2) Loss Comparison



(b) Telecom (AR+Sine) Loss Comparison



(c) Reduced MNIST Loss Comparison

Figure 1 Comparison of Multistage Convex to Nonconvex, QiML (No Initialization), and Sellier (2023) methods. Each plot illustrates the convergence and loss stability across datasets. The Convex to Nonconvex initialization yields consistently smoother and faster convergence compared to QiML and Sellier baselines.

Algorithm 1 Two-Stage Optimization Framework for Overcoming Barren Plateaus

Require: Quantum ansatz $\psi(\theta)$, Hamiltonian H, convex observable set $\{O_i\}$, thresholds $\tau_g, \lambda_0, \beta_0$
Ensure: Optimized parameters θ^*
Initialize parameters θ_0 randomly
Set $\lambda \leftarrow \lambda_0$, $\beta \leftarrow \beta_0$
▷ **Stage 1: Convex Initialization**
while $\frac{\|\nabla_\theta L_{\text{convex}}\|_2}{\dim(\theta)} < \tau_g$ **do**
  Compute local convex loss:

$$L_{\text{convex}}(\theta) = \sum_{i \in R} \|\langle\psi(\theta)|O_i|\psi(\theta)\rangle - b_i\|_2^2$$

  Update parameters:

$$\theta \leftarrow \theta - \eta_c \nabla_\theta L_{\text{convex}}(\theta)$$

end while
▷ **Stage 2: Nonconvex Refinement**
while not converged **do**
  Compute oscillatory regularization:

$$R_{\text{osc}}(\theta) = \mathbb{E}_j[\sin(\omega\theta_j + \phi_j)]$$

  Define total loss:

$$L_{\text{refine}}(\theta) = (1-\lambda)\langle\psi(\theta)|H|\psi(\theta)\rangle + \lambda L_{\text{convex}}(\theta) + \beta R_{\text{osc}}(\theta)$$

  Update parameters:

$$\theta = \theta - \eta_n \nabla_\theta L_{\text{refine}}(\theta)$$

  Anneal parameters:

$$\lambda = \max(0, 0.98\lambda), \quad \beta = 0.99\beta$$

end while
return θ^*

B. Curve Fitting (Synthetic Benchmark)

Following the procedure in [6], we first evaluated the methods on a nonlinear regression task defined by

$$y = x^2, \quad x \in [0, 1], \tag{8}$$

which provides an analytically tractable but nonconvex energy landscape. Each model consisted of a two-layer perceptron with 250 hidden neurons and **Tanh** activations.

The proposed multistage QiML training began with a *convex initialization stage*, minimizing a quadratic surrogate loss to locate a smooth, low-energy region of the Hilmaton landscape. After 100 iterations, the system transitioned into the *nonconvex refinement stage*, introducing an oscillatory regularization term to induce structured perturbations and improve generalization.

Figure 1a compares the convergence trajectories for all methods. Our Convex to Nonconvex model achieved faster convergence and smoother descent beyond step 300. A zoomed inset (steps 300–500) highlights its stable convergence behavior. In contrast, the QiML (No Init) baseline exhibited oscillatory and unstable loss reduction, while Sellier (2023) plateaued at a higher residual energy. Quantitatively, our model achieved a mean squared error (MSE) reduction of approximately 27.3% compared to the Sellier baseline after 500 training steps.

C. Telecom Dataset Evaluation

As shown in Fig. 1b, the proposed Convex to Nonconvex method exhibits a smoother convergence trajectory and consistently achieved a lower final loss value compared to the QiML (No Init) baseline. Sellier's optimizer converged faster in the early iterations but plateaus prematurely, showing limited adaptability to the nonstationary regime introduced by autoregressive noise. In contrast, the convex initialization in our method stabilizes the optimization landscape, preventing early oscillations and enhancing subsequent fine-tuning in the nonconvex phase. These results demonstrated the benefit of convex-to-nonconvex continuity in mitigating optimization ruggedness for temporal prediction tasks.

D. Reduced MNIST Classification

To assess scalability, we next evaluated the proposed method on a *reduced MNIST* dataset (200 samples per class), normalized to zero mean and unit variance. All models employed a single-hidden-layer **Tanh** Multi Layer Perceptron with 128 hidden units. Training followed the same staged schedule: convex initialization with $\lambda = 0.9$ and nonconvex refinement with $\beta = 0.05$.

As shown in Figure 1c, the Convex to Nonconvex framework yields a smoother and faster loss decay compared to QiML without initialization. The proposed model achieved an 11.5% relative reduction in validation loss and lower gradient variance, indicating improved stability and reduced susceptibility to barren-plateau-like flatness. These improvements arise from the convex initialization guiding parameters toward a low-curvature region before introducing nonconvex adaptive oscillations. Sellier (2023) [6] showed poor generalization as shown in the large loss scale, which reflected the deteriorating performance caused by system instability when compared to the proposed baseline.

In nutshell, across both regression and classification domains, the proposed Convex to Nonconvex refinement strategy consistently outperformed baselines in convergence speed, stability, and final task accuracy. The convex initialization stabilizes the optimizer within a low-energy manifold of the Hilmaton landscape, while the nonconvex refinement promotes exploration and adaptive expressivity without divergence. This synergy enables effective navigation of rugged quantum-inspired landscapes, analogous to barren-plateau mitigation in variational quantum circuits. The evaluation demonstrates that the proposed two-stage QiML training pipeline lead to reduced early-stage instability through convex surrogate initialization. On top of that, it enhanced late-stage adaptivity via controlled nonconvex oscillations. Also, it achieved improved generalization and lower loss variance across tasks. Collectively, these results validate the effectiveness of convex-to-nonconvex refinement as a robust optimization paradigm for quantum-inspired machine learning.

V. Conclusion

In this work, we introduced a two-stage optimization framework that effectively addresses the barren plateau phenomenon in variational quantum circuits and quantum-inspired learning models. The proposed method combines a convex initialization stage with a nonconvex refinement stage, enabling stable gradient propagation, improved convergence, and enhanced generalization across multiple datasets. The convex stage establishes a smooth, low-energy region of the Hilmaton landscape that mitigates gradient vanishing and initialization sensitivity, while the nonconvex refinement introduces structured oscillatory regularization that promotes exploration and expressive adaptability without sacrificing stability.

Empirical evaluations across synthetic regression, temporal forecasting, and classification tasks demonstrated that this multistage approach consistently outperforms both baseline QiML models without initialization and the energy-evolution method of Sellier (2023) [6]. The proposed framework yields faster convergence, lower loss variance, and greater resilience to optimization noise key indicators of barren plateau mitigation.

Beyond its empirical success, the method provides a conceptual bridge between convex optimization theory and quantum dynamical evolution, reflecting how quantum systems transition from ordered, low-entropy states to complex interference-rich configurations. This structured progression offers a principled pathway for maintaining gradient information and preserving trainability even in deep or expressive variational architectures.

Future work will extend this framework to hardware-efficient ansatz and explore its integration with adaptive learning-rate schedules, quantum-aware optimizers, and reinforcement-based control for dynamic ansatz selection. Collectively, these directions point toward scalable, trainable quantum neural architectures capable of overcoming fundamental gradient vanishing barriers in hybrid quantum–classical learning.

References

[1] Francis Boabang. *Refining Optimization Methods for Training Machine Learning Models: A Case Study in Robotic Surgical Procedures*. PhD thesis, Concordia University, 2024.

[2] 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. Cost function-dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1):1791, 2021.

[3] Chen Ding, Tian-Yi Bao, and He-Liang Huang. Quantum-inspired support vector machine. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):7210–7222, 2022.

[4] Zachary Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. Connecting expressibility to trainability in quantum machine learning. *Physical Review Letters*, 129(9):090502, 2022.

[5] 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.

[6] Jean Michel Sellier. On a quantum inspired approach to train machine learning models. *Applied AI Letters*, 4(4):e89, December 2023.

[7] Michał Stechły. Introduction to variational quantum algorithms. *arXiv preprint arXiv:2402.15879*, 2024.

[8] Liang Xiong, Xi Chen, and Jeff Schneider. Direct robust matrix factorizatoin for anomaly detection. In *2011 IEEE 11th International Conference on Data Mining*, pages 844–853, 2011.