QTN-VQC: an end-to-end learning framework for quantum neural networks

Jun Qi, Chao-Han Yang and Pin-Yu Chen

Abstract

This work focuses on investigating an end-to-end learning approach for quantum neural networks (QNN) on noisy intermediate-scale quantum devices. The proposed model combines a quantum tensor network (QTN) with a variational quantum circuit (VQC), resulting in a QTN-VQC architecture. This architecture integrates a QTN with a horizontal or vertical structure related to the implementation of quantum circuits for a tensor-train network. The study provides theoretical insights into the quantum advantages of the end-to-end learning pipeline based on QTN-VQC from two perspectives. The first perspective refers to the theoretical understanding of QTN-VQC with upper bounds on the empirical error, examining its learnability and generalization powers; The second perspective focuses on using the QTN-VQC architecture to alleviate the Barren Plateau problem in the training stage. Our experimental simulation on CPU/GPUs is performed on a handwritten digit classification dataset to corroborate our proposed methods in this work.

1. Introduction

State-of-the-art machine learning technologies, particularly deep neural networks (DNN), have revolutionized various fields ranging from speech recognition [1] to computer vision [2], and scientific research in synthetic biology [3]. However, as new scientific applications, such as drug discovery [4] and material science [5], emerge, the computational cost of training DNN models becomes challenging, exceeding classical hardware's limits [6–8]. The promising solution on the horizon is quantum computing. The advent of quantum computing holds great potential for enhancing the computational efficiency of machine learning algorithms through quantum machine learning (QML) [9–11]. In particular, quantum computers leverage the principles of quantum mechanics to perform certain computations much faster than classical computers. By harnessing the power of quantum computers, QML can potentially accelerate the training and inference processes of machine learning models, making them more efficient and cost-effective for computationally demanding scientific applications [12–14].

Although QML is still in its early stages of development, researchers are actively exploring and designing quantum algorithms and architectures conducted on noisy intermediate-scale quantum (NISQ) devices [15–18]. Although some methods like Bosonic quantum encoding allow for single-mode quantum error correction [19, 20], NISQ devices have a small number of qubits and are hard to implement quantum error correction. Variational quantum circuits (VQC) have been proposed as a way to compose QML algorithms in the NISQ era [21–23], and the VQC admits many machine learning applications in NISQ devices [24–31]. However, VQC does face two significant challenges in machine learning tasks: (1) The VQC’s linearly unitary transformation limits its ability to approximate complex target functions, which means that certain types of functions may be difficult for VQC to be represented accurately [32, 33]; (2) VQC often encounters Barren Plateau landscapes, where the gradients of the cost function diminish exponentially as the depth of the VQC increases [34–36]. This can make it challenging to optimize the circuit and find good solutions.
To address the above issues, in our previous work [37, 38], we put forth a TTN-VQC architecture that is composed of a tensor-train network (TTN) [39] with a VQC. The TTN introduces non-linearity to the quantum input features to enhance the VQC representation power, and many works [37, 40, 41] have demonstrated that TTN introduces an inductive bias such that it can assist the VQC in overcoming the Barren Plateau problem in the training process. Since the TTN is a classical simulation of quantum circuits that could be implemented using one and two-qubit quantum circuits [42], we claim that the TTN is associated with a quantum tensor network (QTN) and simulates the wave function of the corresponding quantum circuits classically. Thus, the TTN-VQC is still referred to as a hybrid quantum–classical model where data preprocessing is done classically relying on the TTN, and the parameters of both TTN and VQC are jointly adjusted in a supervised learning fashion on the training dataset. However, an actual end-to-end quantum learning paradigm of QTN-VQC is expected, and in this work, we focus on implementing the end-to-end QTN-VQC architecture in which every component, including the QTN, is built upon quantum circuits. Furthermore, we introduce the non-linear property in the feature space by using the Sigmoid function to the QTN features before they are encoded into the quantum states for the parametric component in VQC.

More specifically, we propose QTN models as a means to implement a TTN using quantum circuits. The QTN is composed of parametric quantum circuits that can be adjustable during the training process, similar to how the TTN operates. By combining the QTN with the VQC, we establish an end-to-end learning paradigm called QTN-VQC. This architecture allows for the seamless integration of quantum circuits in the entire learning process, enabling the training and optimization of both the QTN and VQC in a unified manner. As shown in figure 1, we construct an end-to-end QTN-VQC model, utilizing the QTN for feature abstraction and the VQC model for making predictions. Using a back-propagation algorithm, we evaluate the cost function’s gradients to train the model and jointly fine-tune the QTN and VQC model parameters. This enables the optimization of the model parameters based on the desired output, enhancing the overall performance of the QTN-VQC.

In our analysis of the QTN-VQC model, we focus on the upper bound of the estimation error associated with empirical risk minimization (ERM) [43–45], which seeks to minimize the error on the training sample. More specifically, given a set of QTN-VQC operators \( \mathcal{H}_{QV} \) and a training dataset \( S = \{x_1, x_2, \ldots, x_N\} \), we define an ERM \( f^*_S \) as:

\[
  f^*_S = \arg \min_{f \in \mathcal{H}_{QV}} \mathcal{L}_S(f),
\]

where \( \mathcal{L}_S(f) \) denotes an empirical loss that is a numerical approximation to an expected one \( \mathcal{L}_P(h) \). Given a target operator \( h_P^* \), a distribution \( \mathcal{D} \) and a classified loss function \( \ell \), both \( \mathcal{L}_P(f) \) and \( \mathcal{L}_S(f) \) are separately defined as:

\[
  \mathcal{L}_P(f) := \mathbb{E}_{x \sim \mathcal{D}}[\ell(h_P^*(x), f(x))],
\]

and

\[
  \mathcal{L}_S(f) := \frac{1}{N} \sum_{n=1}^{N} \ell(h_P^*(x_n), f(x_n)).
\]

The estimation error \( \epsilon_{est} \) is defined as the difference between \( \mathcal{L}_P(f^*_S) \) and \( \inf_{f \in \mathcal{H}_{QV}} \mathcal{L}_P(f) \), and it can be upper bounded as:

![Figure 1. An illustration of an end-to-end QTN-VQC architecture. Given the inputs \( x_1, x_2, \ldots, x_N \), the QTN transforms the inputs \( x_i \) into QTN-VQC feature vectors \( \phi(x_i) \), and the VQC classifier \( \varphi \) is set to make predictions. To train the model parameters, we compute gradients of the cost function and then fine-tune the model parameters of both QTN and VQC based on the back-propagation algorithm.](image-url)
In this work, we provide two end-to-end quantum learning architectures based on the QTN-VQC. The QTN is substituted for the TTN, resulting in a modification of the hybrid quantum–classical model based on TTN-VQC into a truly quantum-circuit-based architecture. The transition from TTN to QTN enables a more comprehensive integration of quantum principles within the learning paradigm. By leveraging the QTN-VQC architecture, the model can harness the inherent quantum properties offered by quantum circuits even to improve the empirical performance of machine learning tasks.

- In this study, we demonstrate that the QTN-VQC achieves an upper bound on the estimation error, comparable to the TTN-VQC approach. More specifically, we show that the estimation error $\epsilon_{est}$ given a set of $N$ training data, is upper bounded by $O(1/\sqrt{N})$. The finding highlights the effectiveness of QTN-VQC in managing the generalization capability in the quantum learning process.

- We justify the end-to-end quantum learning paradigm based on QTN-VQC by demonstrating that the PL condition incorporated in the QTN-VQC is sufficient to guarantee an exponential convergence rate during training. This highlights the advantages of the QTN-VQC architectures in addressing challenges such as the Barren Plateau problem and ensuring efficient training in the quantum learning pipeline.

2. Results

2.1. Preliminaries

Before exploiting the theoretical understanding of QTN-VQC, we first introduce the VQC that has been largely employed in quantum machine learning on NISQ devices. As shown in Figure 2, a VQC consists of three components: (a) Tensor Product Encoding (TPE); (b) Parametric Quantum Circuit (PQC); (c) Measurement.

The TPE framework was originally shown in [46], where a classical data $x$ is converted into a quantum state $|x\rangle$ by adopting a one-to-one mapping as:

$$|x\rangle = \left(\bigotimes_{i=1}^{U} R_y\left(\frac{\pi x_i}{2}\right)\right)|0\rangle^\otimes U = \begin{bmatrix} \cos \left(\frac{\pi x_1}{2}\right) & \cdots & \cos \left(\frac{\pi x_U}{2}\right) \\ \sin \left(\frac{\pi x_1}{2}\right) & \cdots & \sin \left(\frac{\pi x_U}{2}\right) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} \cos \left(\frac{\pi x_U}{2}\right) \\ \sin \left(\frac{\pi x_U}{2}\right) \end{bmatrix},$$

(5)

where $\hat{R}_\phi$ refers to the empirical Rademacher complexity of $\mathcal{H}_{QV}$ and $\epsilon_{opt}$ represents an optimization bias. Our work aims to provide upper bounds on the empirical Rademacher complexity $\hat{R}(S)$ while minimizing the optimization bias. The contribution to this work can be summarized as follows:

- In this work, we provide two end-to-end quantum learning architectures based on the QTN-VQC architectures, with all model components comprising quantum circuits. The performance of machine learning tasks.
where $\mathbf{x}_i$ is imposed to lie in the domain of $[0, 1]$ such that the conversion between $\mathbf{x}$ and $|\mathbf{x}\rangle$ is a reversely one-to-one correspondence.

The PQC comprises $U$ quantum channels associated with $U$ qubits. We realize the quantum entanglement by using two-qubit controlled-NOT (CNOT) gates imposed upon mutually two quantum channels, and we build up the parametric PQC model with single-qubit rotation gates $R_{x_i}$, $R_{y_i}$, $R_{z_i}$ with adjustable parameters $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_U\}$, $\beta = \{\beta_1, \beta_2, ..., \beta_U\}$, and $\gamma = \{\gamma_1, \gamma_2, ..., \gamma_U\}$. The PQC is taken as a unitary linear operator $U_{\theta} = \{\alpha\} \cup \{\beta\} \cup \{\gamma\}$ that can help QNN to better deal with the issue of the Barren Plateaus problem.

### 2.2. Theoretical results

In this part, we first introduce two QTN architectures and then analyze their theoretical understanding using error performance analysis.

The two QTN architectures, as depicted in figure 3, consist of an encoder with non-trainable quantum gates $R_{\chi}(\mathbf{x}_i)$ for the classical input data $\mathbf{x}_i$, and a parametric quantum model with trainable quantum gates $R_{\chi}(\theta_i)$ and $R_{\chi}(\theta_i)$ associated with the parameter $\theta_i$. Two quantum circuit-implemented QTN architectures are built into this work: one denotes convolutional matrix product state (ConvMPS); and the other refers to convolutional tree tensor network (ConvTTN). Since a matrix product state [47] factorizes a tensor with multiple indices into a product of three-index tensors, the ConvMPS is in line with this architecture by entangling the current quantum state with the previous one. On the other hand, the ConvTTN is arranged in a hierarchical tree structure, where each quantum state represents a local region of the system. The quantum states at the leaves of the tree correspond to the system’s components, while the quantum states at higher levels capture the correlations between these components. Besides, the QTN based on both ConvMPS and ConvTTN with the convolutional mechanism possesses an inductive bias from owning a prior over the space of architectures that are sharper than deep hardware-efficient architectures [48], which can help QNN to better deal with the issue of the Barren Plateaus problem.

By combining the QTN with the VQC, we obtain the QTN-VQC architecture, as illustrated in figure 1. Starting with the classical input data $\mathbf{x}_i$, the QTN operator generates the QTN feature $\phi(\mathbf{x}_i)$. These QTN features serve as inputs to the TPE component within the VQC framework. To introduce the non-linearity property, we apply the sigmoid function to the QTN features $\phi(\mathbf{x}_i)$, which is

$$|\mathbf{x}\rangle = \bigotimes_{i=1}^{U} R_{\chi} (\frac{\pi}{2}\text{sigm}(\phi(\mathbf{x}_i))) |0\rangle^\otimes U. \quad (6)$$

The introduction of the Sigmoid function brings in non-linearity to QTN features $\phi(\mathbf{x})$ in feature space, such that the end-to-end quantum learning architecture QTN-VQC owns a more powerful representation capability to approximate target functions.
This transformation enhances the expressiveness of the model and allows for capturing complex relationships between the QTN and the target output. Thus, the QTN-VQC architecture leverages the strengths of both QTN and VQC, combining their respective capabilities to enhance the learning and generalization capability of the overall model.

Next, we first focus on characterizing the generalization power of the QTN-VQC architecture. We derive an upper bound on the estimation error, which provides insights into the model’s ability to generalize well to unseen data. Then, the optimization performance of the QTN-VQC architecture is analyzed. We justify that QTN-VQC satisfies the PL condition during the training process. This condition ensures that the optimization algorithm can effectively converge to a desirable solution by avoiding the Barren Plateau problem. By demonstrating that the QTN-VQC architecture meets the PL condition, we argue that it can achieve a small optimization bias, which means that the model can efficiently optimize its parameters during training, leading to better performance and learning outcomes.

\textbf{Theorem 1.} Given the training dataset \(S = \{x_1, x_2, ..., x_N\}\) and a set of QTN-VQC operators \(\mathcal{H}_{\text{QV}}\), we define \(\mathcal{H}_Q\) and \(\mathcal{H}_V\) refer to the sets of QTN and QTT operators, respectively, then the estimation error \(\epsilon_{\text{est}}\) is upper bounded using the empirical Rademacher complexity \(\hat{\mathcal{R}}_{\mathcal{S}}(\mathcal{H}_{\text{QV}})\) as:

\[
\epsilon_{\text{est}} \leq 2\hat{\mathcal{R}}_{\mathcal{S}}(\mathcal{H}_{\text{QV}}) \leq 2\hat{\mathcal{R}}_{\mathcal{S}}(\mathcal{H}_Q) + 2\hat{\mathcal{R}}_{\mathcal{S}}(\mathcal{H}_V) \leq \frac{2P}{\sqrt{N}}(\Lambda_Q + \Lambda_V)
\]

subject to \(\langle x_i | x_n \rangle \leq P, \quad n \in [N],\)

\[
\|W_Q\|_F \leq \Lambda_Q, \quad \|W_V\|_F \leq \Lambda_V.
\]

(7)

where \(\Lambda_Q\) and \(\Lambda_V\) separately denote the constant constants for the Frobenius norm of QTN and VQC parameter matrices \(W_Q\) and \(W_V\).

The upper bound on the estimation error \(\epsilon_{\text{est}}\) in equation (7) in theorem 1 suggests that a sufficient amount of training data guarantees a smaller \(\epsilon_{\text{est}}\) and the given constants \(\Lambda_Q\) and \(\Lambda_V\) determine the model complexity of QTN and VQC, respectively. In particular, \(\hat{\mathcal{R}}(\mathcal{H}_{\text{QV}})\) is upper bounded by \(\hat{\mathcal{R}}_{\mathcal{S}}(\mathcal{H}_Q) + \hat{\mathcal{R}}_{\mathcal{S}}(\mathcal{H}_V)\), which is the sum of empirical Rademacher complexities of \(\mathcal{H}_Q\) and \(\mathcal{H}_V\).

The optimization bias of QTN-VQC is related to the Barren Plateau problem that occurs when optimizing a non-convex objective function, and the gradients of the cost function vanish exponentially, making it challenging to optimize the parameters of the QTN-VQC model. To mitigate this issue, we propose the PL condition for the setup of QTN-VQC parameters, which was previously introduced in our work [37] for the model setup of TTN-VQC parameters. In this work, we generalize the PL condition to the model setup of QTN-VQC and justify that the QTN-VQC architecture with the PL condition can also ensure an exponential convergence rate during training.

We assume the parameter of QTN-VQC as \(\theta_{\text{QV}} = \{W_Q, W_V\}\). The PL condition means that for a given constant \(\mu\), an empirical loss function \(\mathcal{L}_S\) concerning the QTN-VQC parameter \(\theta_{\text{QV}}\) should satisfy the following inequality as:

\[
\frac{1}{2}\|\mathcal{L}_S(\theta_{\text{QV}})\|_2^2 \geq \mu \mathcal{L}_S(\theta_{\text{QV}}).
\]

(8)

Next, we employ the technique of tangent kernel as shown in proposition 1 to verify if the QTN-VQC’s parameters can satisfy the PL condition. Then, in theorem 2, we show that the gradient descent algorithm can ensure an exponential convergence rate to reduce the loss value to small values.

\textbf{Proposition 1.} For a QTN-VQC operator \(f \in \mathcal{H}_{\text{QV}}\) with the parameter \(\theta_{\text{QV}}\), we define the tangent kernel \(K_f\) as

\[
K_f = \nabla f(\theta_{\text{QV}}) \nabla f(\theta_{\text{QV}})^T.
\]

(9)

Then, if the empirical loss function \(\mathcal{L}_S(\theta_{\text{QV}})\) satisfies the \(\mu\)-PL condition, the smallest eigenvalue \(\lambda_{\text{min}}(K_f)\) for \(K_f\) meets the condition as:

\[
\lambda_{\text{min}}(K_f) \geq \mu.
\]

(10)

\textbf{Theorem 2.} The QTN-VQC parameter \(\theta_{\text{QV}}\) ensures the empirical loss function \(\mathcal{L}_S(\theta_{\text{QV}})\) satisfying the PL condition, then the gradient descent algorithm can lead to an exponential convergence rate during the training process, which is

\[
\mathcal{L}_S(\theta_{\text{QV}}(T)) \leq \exp(-\mu T)\mathcal{L}_S(\theta_{\text{QV}}(0)),
\]

(11)

where \(\theta_{\text{QV}}(T)\) and \(\theta_{\text{QV}}(0)\) separately refer to the QTN-VQC’s parameters at the initial time and the epoch \(T\).
Equation (11) in theorem 2 suggests that with the setup of PL for QTN-VQC parameters, for a given large epoch $T$, the optimization bias $\epsilon_{opt}$ can be reduced to a small value. This highlights the advantage of QTN-VQC with the PL condition for the model training, efficiently alleviating the Barren Plateau landscape.

2.3. Empirical results
The experimental section aims to assess the generalization capability and learnability of the QTN-VQC architecture. To validate its performance, we conduct several comparisons and evaluations as follows:

1. The QTN-VQC models based on ConvMPS-VQC and ConvTTN-VQC are tested and compared against the baseline linear PCA-VQC model. The PCA-VQC model utilizes principal component analysis (PCA) [49] for unsupervised dimensionality reduction, while the QTN-VQC model leverages the QTN for abstract feature extraction. Our experiments aim to demonstrate that the QTN-VQC models outperform the linear PCA-VQC model and achieve similar results compared to the TTN-VQC counterpart, which was previously proposed in [37].

2. The convergence rates during the training process are examined for both ConvMPS-VQC and ConvTTN-VQC models. The goal is to show that these models exhibit exponential convergence rates, which validate the PL condition that ensures efficient training.

To conduct the experiments, the MNSIT dataset [50] is used, consisting of handwritten digit images. The dataset comprises 60,000 for training and 10,000 for testing. In the experiments, a subset of 10,000 training data samples and 2,000 test data samples are randomly selected. The empirical performance is measured using the cross-entropy loss function. The SGD algorithm with a learning rate of 0.01 and a mini-batch size of 50 is employed during the training process. For each 28 $\times$ 28 image, we first use the 2D average pooling with a kernel of 6 to downsample the images to the shape of 4 $\times$ 4, then the QTN is based on 4 quantum channels and a quantum convolutional kernel of 2 $\times$ 2 to obtain tensors of 4 $\times$ 2 $\times$ 2.

Additionally, the PCA-VQC and TTN-VQC models are built as a baseline and their setups follow our previous work in [37]. The model combines PCA and TTN as feature extractors, followed by the VQC for making predictions. The PCA-VQC represents a linear VQC, in contrast to the non-linear QTN-VQC models including ConvMPS-VQC and ConvTTN-VQC. Moreover, we assign 4 qubits to compose the QTN and 8 qubits for the composition of the VQC model. These experiments aim to provide empirical evidence to support the superiority of the QTN-VQC architecture, showcasing its potential in the end-to-end learning paradigm for quantum neural networks. Besides, the experimental simulations on CPU/GPUs based on a total of 12-qubits are used to corroborate our theoretical analysis, and the experimental scaling-up on real quantum computers is not our major consideration in this work.

Figure 4 illustrates the empirical results of binary classification on the MNIST dataset to compare the performance of QTN-VQC with the PCA-VQC and TTN-VQC counterparts. Although the PCA-VQC suffers from the Barren Plateau problem, the two QTN-VQC models and the TTN-VQC exhibit exponential convergence rates and they finally reduce the loss to a small value. In particular, both ConvMPS-VQC and ConvTTN-VQC models can maintain the empirical performance of TTN-VQC during the training process, while they perform a little worse than the TTN-VQC in the evaluation stage. Table 1 provides the evaluation
results, where both ConvTTN-VQC and ConvMPS-VQC can attain a performance close to the TTN-VQC in terms of the CE scores and accuracies. The experimental simulation also corroborates that the QTN-VQC including ConvMPS-VQC and ConvTTN-VQC satisfying the PL condition ensures an exponential convergence rate during training.

Moreover, as shown in figure 5 and table 2, we conduct the ternary classification on the MNIST dataset to compare the performance of PCA-VQC, TTN-VQC, and QTN-VQC models. Similar to the binary classification, the loss of PCA-VQC lies in high values due to the Barren Plateau problem. However, both ConvMPS-VQC and ConvTTN-VQC can even achieve lower losses than the TTN-VQC counterpart in both training and test datasets, which empirically justifies our theorem that QTN-VQC can inherently ensure exponential convergence rates during the training process because of the PL condition. Besides, the results demonstrate that it can generalize well in the evaluation stage if sufficient training data is available, which corresponds to our analysis in theorem 1.

3. Discussion

This work focuses on investigating QTN models to construct an end-to-end quantum learning architecture called QTN-VQC, which replaces our previously proposed quantum–classical learning system based on the TTN-VQC model. We put forth the ConvMPS and ConvTTN models as components of QTN, resulting in the

![Figure 5. Experiments of ternary classification on the MNIST dataset to evaluate the empirical performance of the PCA-VQC, TTN-VQC, and QTN-VQC models. a) empirical results on the training dataset. b) empirical results on the test dataset. The QTN-VQC models include ConvMPS-VQC and ConvTTN-VQC, where 4 qubits are used to construct the QTN and 8 qubits are employed to compose the VQC.](image-url)

| Models       | CE     | Accuracy (%) |
|--------------|--------|--------------|
| PCA-VQC     | 0.3301 | 87.7         |
| TTN-VQC     | 0.0634 | 98.6         |
| ConvTTN-VQC | 0.0805 | 98.5         |
| ConvMPS-VQC | 0.0917 | 98.3         |

| Models       | CE     | Accuracy (%) |
|--------------|--------|--------------|
| PCA-VQC     | 1.0900 | 64.7         |
| TTN-VQC     | 0.3781 | 96.2         |
| ConvTTN-VQC | 0.2653 | 96.0         |
| ConvMPS-VQC | 0.2698 | 95.9         |
structures ConvMPS-VQC and ConvTTN-VQC. The theoretical understanding of QTN-VQC is explored using the paradigm of Empirical Risk Minimization to derive an upper bound on the estimation error, which determines its generalization power and the learnability of VQC. Our derived theorem indicates that QTN-VQC satisfying the PL condition exhibits an exponential convergence rate during training, and our derived upper bound on the empirical Rademacher complexity also suggests the need for a sufficient amount of training data to generalize well to unseen data.

Our experiments are conducted on the MNIST dataset to validate the derived theoretical results. The empirical performance of QTN-VQC, including ConvMPS-VQC and ConvTTN-VQC, is compared with PCA-VQC and TTN-VQC counterparts. Both ConvMPS-VQC and ConvTTN-VQC models demonstrate their ability to overcome the Barren Plateau and exhibit exponential convergence rates. Notably, the end-to-end quantum learning architecture QTN-VQC, based on ConvMPS-VQC and ConvTTN-VQC models, maintains the empirical performance of the TTN-VQC counterpart.

While previous work has shown the effectiveness of QTN like TTN in facilitating efficient training of VQC, this new work justifies the use of QTN implemented by quantum gates. It emphasizes the significance of the proposed end-to-end quantum learning paradigm represented by QTN-VQC in the context of machine learning tasks, particularly in the NISQ era. By demonstrating that QTN implemented with quantum gates can assist VQC during training, we underscore the potential of the QTN-VQC approach for machine learning approaches in the current quantum computing landscape.

4. Method

This section aims at providing proof for our theoretical results. We first explain how to derive equation (4) associated with the ERM, and then we show the procedures to attain the theoretical results in theorem 1 and theorem 2.

4.1. Validation of equation (4)

By the definition \( \inf_{f \in \mathcal{H}_{QV}} \mathcal{L}_{D}(f) \), for any \( \epsilon_{\text{opt}} > 0 \), there exists \( f_{S_{\epsilon}} \) such that \( \mathcal{L}_{D}(f_{S_{\epsilon}}) \leq \inf_{f \in \mathcal{H}_{QV}} \mathcal{L}_{D}(f) + \epsilon_{\text{opt}} \).

Thus, using \( \mathcal{L}_{S}(f_{S_{\epsilon}}) \leq \mathcal{L}_{S}(f_{S_{\epsilon}}) \), which holds by the definition of the algorithm, we can upper bound the estimation error \( \epsilon_{\text{est}} \) as:

\[
\epsilon_{\text{est}} = \mathcal{L}_{D}(f_{S_{\epsilon}}) - \inf_{f \in \mathcal{H}_{QV}} \mathcal{L}_{D}(f) \\
\leq \mathcal{L}_{D}(f_{S_{\epsilon}}) - \mathcal{L}_{D}(f_{S_{\epsilon}}) + \mathcal{L}_{D}(f_{S_{\epsilon}}) - \inf_{f \in \mathcal{H}_{QV}} \mathcal{L}_{D}(f) + \epsilon_{\text{opt}} \\
\leq \mathcal{L}_{D}(f_{S_{\epsilon}}) - \mathcal{L}_{D}(f_{S_{\epsilon}}) + \mathcal{L}_{D}(f_{S_{\epsilon}}) - \mathcal{L}_{D}(f_{S_{\epsilon}}) + \epsilon_{\text{opt}} \\
\leq 2 \sup_{f \in \mathcal{H}_{QV}} |\mathcal{L}_{D}(f) - \mathcal{L}_{S}(f)| + \epsilon_{\text{opt}}.
\]

(12)

Furthermore, we employ the empirical Rademacher complexity \( \hat{\mathcal{R}}_{S}(\mathcal{H}_{QV}) \) to upper bound the term \( \sup_{f \in \mathcal{H}_{QV}} |\mathcal{L}_{D}(f) - \mathcal{L}_{S}(f)| \), which is based on the fact that for the set \( S \) with \( N \) training samples, the following inequality holds for \( f_{S_{\epsilon}} \) as:

\[
P\left[ \mathcal{L}_{D}(f_{S_{\epsilon}}) - \inf_{f \in \mathcal{H}_{QV}} \mathcal{L}_{D}(f) > \epsilon_{\text{opt}} \right] \leq P\left[ \sup_{f \in \mathcal{H}_{QV}} |\mathcal{L}_{D}(f) - \mathcal{L}_{S}(f)| > \frac{\epsilon_{\text{opt}}}{2} \right] \\
\leq \exp(-2N(\epsilon_{\text{opt}} - \hat{\mathcal{R}}_{S}(\mathcal{H}_{QV}))),
\]

(13)

where given the random variables \( \sigma_{n} \) taking values in \( \{-1, 1\} \), the empirical Rademacher complexity \( \hat{\mathcal{R}}_{S}(\mathcal{H}_{QV}) \) is defined as:

\[
\hat{\mathcal{R}}_{S}(\mathcal{H}_{QV}) = \frac{1}{N} \mathbb{E}_{q} \left[ \sum_{n=1}^{N} \sigma_{n} f_{S_{\epsilon}}(x_{n}) \right].
\]

(14)

From the fact in equation (13), we can finally derive the theoretical result:

\[
\epsilon_{\text{est}} \leq 2 \sup_{f \in \mathcal{H}_{QV}} |\mathcal{L}_{D}(f) - \mathcal{L}_{S}(f)| + \epsilon_{\text{opt}} \leq 2\hat{\mathcal{R}}_{S}(\mathcal{H}_{QV}) + \epsilon_{\text{opt}}.
\]

(15)
4.2. Proof of theorem 1

Since \( \mathcal{H}_{QV} = \{ f = g*h: g \in \mathcal{H}_Q, h \in \mathcal{H}_V \} \), based on the property of the Rademacher identities [51], we have

\[
\hat{R}_s(\mathcal{H}_{QV}) \leq \hat{R}_s(\mathcal{H}_Q) + \hat{R}_s(\mathcal{H}_V),
\]

which means that \( \hat{R}_s(\mathcal{H}_{QV}) \) should be characterized by upper bounding \( \hat{R}_s(\mathcal{H}_Q) \) and \( \hat{R}_s(\mathcal{H}_V) \), respectively.

\[
\hat{R}_s(\mathcal{H}_Q) = \frac{1}{N} \mathbb{E}_\sigma \left[ \sup_{g \in \mathcal{H}_Q} \sum_{n=1}^{N} \sigma_n g(x_n) \right]
\]

\[
= \frac{1}{N} \mathbb{E}_\sigma \left[ \sup_{\|w\|_2 \leq \Lambda} \sum_{n=1}^{N} \sigma_n w_n \cdot |x_n| \right]
\]

\[
= \frac{1}{N} \mathbb{E}_\sigma \left[ \left\| \sum_{n=1}^{N} \sigma_n x_n \right\|_2 \right]
\]

\[
\leq \frac{\Lambda Q}{N} \left\| \sum_{n=1}^{N} \sigma_n x_n \right\|_2
\]

\[
= \frac{\Lambda Q}{N} \left( \sum_{i,j=1}^{N} \mathbb{E}_\sigma[\sigma_i \sigma_j] \langle x_i | x_j \rangle \right).
\]

Since \( \sigma_i \) are identically and independently distributed, we have

\[
\mathbb{E}_\sigma[\sigma_i \sigma_j] = \begin{cases} 1, & \text{if } i = j \\ -1, & \text{if } i \neq j. \end{cases}
\]

Then, we can further upper bound the \( \hat{R}_s(\mathcal{H}_Q) \) as:

\[
\hat{R}_s(\mathcal{H}_Q) \leq \frac{\Lambda Q}{N} \left( \sum_{i,j=1}^{N} \langle x_i | x_j \rangle \right)
\]

\[
\leq \frac{\Lambda Q}{N} \left( \sum_{n=1}^{N} \langle x_n | x_n \rangle \right) \quad (\langle x_n | x_n \rangle \leq P^2)
\]

\[
\leq \frac{\Lambda Q}{N} P
\]

Similarly, we can attain the theoretical result \( \hat{R}_s(\mathcal{H}_V) \leq \frac{\Lambda Q P}{N} \). Thus, we attain the theoretical result:

\[
\epsilon_{out} \leq 2\hat{R}_s(\mathcal{H}_{QV}) \leq 2\hat{R}_s(\mathcal{H}_Q) + 2\hat{R}_s(\mathcal{H}_V) \leq \frac{2P}{\sqrt{N}} (\Lambda_Q + \Lambda_V).
\]

4.3. Proof of theorem 2

At the iteration \( t + 1 \), the loss function \( \mathcal{L}_s(\theta_{QV}^{(t+1)}) \) can be represented using the Taylor expansion as:

\[
\mathcal{L}_s(\theta_{QV}^{(t+1)}) = \mathcal{L}_s(\theta_{QV}^{(t)}) + \nabla \mathcal{L}_s(\theta_{QV}^{(t)}) \cdot (\theta_{QV}^{(t+1)} - \theta_{QV}^{(t)}) + \frac{1}{2} (\theta_{QV}^{(t+1)} - \theta_{QV}^{(t)})^T H(\theta_{QV}^{(t)}) (\theta_{QV}^{(t+1)} - \theta_{QV}^{(t)})
\]

\[
\leq \mathcal{L}_s(\theta_{QV}^{(t)}) - \eta \nabla \mathcal{L}_s(\theta_{QV}^{(t)}) \cdot (\theta_{QV}^{(t+1)} - \theta_{QV}^{(t)}) + \frac{\eta^2}{2} \nabla \mathcal{L}_s(\theta_{QV}^{(t)})^T H(\theta_{QV}^{(t)}) \nabla \mathcal{L}_s(\theta_{QV}^{(t)})
\]

\[
\leq \mathcal{L}_s(\theta_{QV}^{(t)}) - \eta (1 - \frac{\eta}{2}) \nabla \mathcal{L}_s(\theta_{QV}^{(t)})^2
\]

\[
\leq \mathcal{L}_s(\theta_{QV}^{(t)}) - \eta (2 - \eta) \mu \mathcal{L}_s(\theta_{QV}^{(t)}) \quad \text{(by } \mu \text{ - PLAssumption)}
\]

\[
= (1 - 2\eta \mu + \eta^2 \mu) \mathcal{L}_s(\theta_{QV}^{(t)})
\]

\[
\leq (1 - 2\eta \mu + \eta^2 \mu)^{t+1} \mathcal{L}_s(\theta_{QV}^{(0)}) \quad \text{(by setting } \eta = 1)
\]

\[
\leq \exp(-\mu (t + 1)) \mathcal{L}_s(\theta_{QV}^{(0)}),
\]
where \( H(\theta_{QV}) \) refers to the Hessian matrix associated with the second derivative of \( L_3(\theta_{QV}) \), and we assume that \( \| H(\theta_{QV}) \|_2 \leq 1 \). When \( t + 1 = T \), we attain
\[
L_3(\theta_{QV}^{t+1}) \leq \exp(-\mu T) L_3(\theta_{QV}^t).
\]

(22)

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Jun Qi \( \text{https://orcid.org/0000-0001-7533-2630} \)
Chao-Han Yang \( \text{https://orcid.org/0000-0003-2879-8811} \)
Pin-Yu Chen \( \text{https://orcid.org/0000-0003-1039-8369} \)

References

[1] Deng L et al 2013 Recent advances in deep learning for speech research at microsoft IEEE International Conference on Acoustics, Speech and Signal Processing 8604–8
[2] Sermanet P et al 2014 Overfeat: integrated recognition, localization and detection using convolutional networks International Conference on Learning Representations
[3] Jumper J et al 2021 Highly accurate protein structure prediction with alphafold Nature 596 583–9
[4] Smalley E 2017 AI-powered drug discovery captures pharma interest Nat. Biotechnol. 35 604–6
[5] DeCost B L, Hattrick-Simpers J, Trautt Z, Kusne A G, Campo E and Green M L 2020 Scientific AI in materials science: a path to a sustainable and scalable paradigm Machine Learning: Science and Technology 1 033001
[6] Freedman D H 2019 Hunting for New Drugs with AI Nature 576 549–53
[7] LeCun Y 2019 1.1 Deep learning hardware: past, present, and future IEEE International Solid-State Circuits Conference (Piscataway, NJ: IEEE) 12–9
[8] Potok T E et al 2018 A study of complex deep learning networks on high-performance, neuromorphic, and quantum computers ACM Journal on Computing Systems 14 1–21
[9] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2017 Quantum machine learning Nature 549 195–202
[10] Schuld M, SinAYSkiy I and Petruccione F 2015 An introduction to quantum machine learning Contemp. Phys. 56 172–85
[11] Schuld M and Killoran N 2019 Quantum machine learning in feature hilbert spaces Phys. Rev. Lett. 122 040504
[12] Cao Y, Romero J and Aspuru-Guzik A 2018 Potential of quantum computing for drug discovery J. Chem. Inf. Model. 61 2641–7
[13] Bauer B, Bavyi S, Motta M and Kin-Lic Chan G 2020 Quantum algorithms for quantum chemistry and quantum materials science Chem. Rev. 120 12685–717
[15] Preskill J 2018 Quantum computing in the NISQ Era and beyond Quantum vol. 2 79
[16] Huggins W, Patil P, Mitchell B, Whaley K B and Stoudenmire E M 2019 Towards quantum machine learning with tensor networks Quantum Science and Technology 4 024001
[17] Huang H-Y et al 2021 Power of data in quantum Nat. Commun. 12 1–9
[18] Kandala A et al 2017 Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets Nature 549 242–6
[19] Zeng Y, Zhou Z-Y, Rinaldi E, Gneiting C and Nori F 2020 Approximate autonomous quantum error correction with reinforcement learning Phys. Rev. Lett. 131 050601
[20] Chen Y-H, Stassi R, Qin W, Miranowicz A and Nori F 2022 Fault-tolerant multiqudit geometric entangling gates using photonic cat-state qubits Physical Review Applied 18 044076
[21] Benedetti M, Lloyd E, Sack S and Fiorentini M 2019 Parameterized quantum circuits as machine learning models Quantum Science and Technology 4 044001
[22] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 Quantum circuit model Physical Review A 98 032309
[23] Chen S-e, Huang C-M, Hsing C-W and Kao Y-J 2021 An end-to-end trainable hybrid classical-quantum classifier Machine Learning: Science and Technology 2 045021
[24] Chen S-e, Yang H, Qi J, Chen P-Y, Ma X and Goan H-S 2020 Variational circuits for deep reinforcement learning IEEE Access 8 141007–24
[25] Yang H, Qi J, Chen S-e, Chen P-Y, Siniscalchi SM, Ma X and Lee C-H 2021 Decentralizing feature extraction with quantum convolutional neural network for automatic speech recognition IEEE International Conference on Acoustics, Speech and Signal Processing 6523–7
[26] Yang H, Qi J, Chen S-e, Tsao Y and Chen P-Y 2022 When BERT meets quantum temporal convolutional learning for text classification in heterogeneous computing IEEE International Conference on Acoustics, Speech and Signal Processing (https://doi.org/10.1109/ICASSP54922.2022.9746412)
[27] Huang H-L et al 2021 Experimental quantum generative adversarial networks for image generation Physical Review Applied 16 024051
[28] Skolik A, Jerbi S and Dunjko V 2022 Quantum agents in the gym: a variational quantum algorithm for deep q-learning Quantum 6 720
[29] Jun Qi J 2022 Classical-to-quantum transfer learning for spoken command recognition based on quantum neural networks IEEE International Conference on Acoustics, Speech and Signal Processing 8627–31
[30] Mohammadreza H and Wang H 2022 Matching point sets with quantum circuit learning IEEE International Conference on Acoustics, Speech and Signal Processing 8607–11
[31] Mari A, Bromley T R, Izac J, Schuld M and Killoran N 2020 Transfer learning in hybrid classical-quantum Quantum 4 340
[32] Du Y, Hsieh M-H, Liu T and Tao D 2020 Expressive power of parametrized quantum circuits Physical Review Research 2 033125
Du Y, Hsieh M-H, Liu T, Tao D and Liu N 2021 Quantum noise protects quantum classifiers against adversaries Physical Review Research 3 023153
Arrasmith A, Cerezo M, Czarnik P, Cincio L and Coles P J 2021 Effect of barren plateaus on gradient-free optimization Quantum 5 558
Wang S et al 2021 Noise-induced barren plateaus in variational quantum algorithms Nat. Commun. 12 6961
Larocca M et al 2022 Diagnosing barren plateaus with tools from quantum optimal control Quantum 6 824
Qi J, Yang H H, Chen P-Y and Hsieh M-H 2023 Theoretical error performance analysis for variational quantum circuit based functional regression npj Quantum Information 9 4
Qi J, Yang H H, Chen P-Y and Hsieh M-H 2023 Pre-training Tensor-Train Networks Facilitates Machine Learning with Variational Quantum Circuits arXiv:2306.03741
Oseledets I V 2011 Tensor-train decomposition SIAM J. Sci. Comput. 33 2295–317
Martin E C, Plekhanov K and Lubasch M 2023 Barren plateaus in quantum tensor network optimization Quantum 7 974
Huang J, He W, Zhang Y, Wu Y, Wu B and Yuan X 2022 Tensor Network Assisted Variational Quantum Algorithm arXiv:2212.10421
Ran S J 2020 Encoding of matrix product states into quantum circuits of one- and two-qubit gates Physical Review A 101 032310
Montanari A and Saeed B N 2022 Universality of empirical risk minimization Proc. Conference on Learning Theory (PMLR) 4310–2
Qi J, Du J, Siniscalchi S M, Ma X and Lee C-H 2020 Analyzing upper bounds on mean absolute errors for deep neural network-based vector-to-vector regression IEEE Trans. Signal Process. 68 3411–22
Zhang H, Cisse M, Dauphin Y and Lopez-Paz D 2018 Mixup: beyond empirical risk minimization International Conference on Learning Representation
Stoudenmire E and Schwab D 2016 Supervised learning with tensor networks Advances in Neural Information Processing Systems 29
Cirac J I, Perez-Garcia D, Schuch N and Verstraete F 2021 Matrix product states and projected entangled pair states: concepts, symmetries, theorems Rev. Mod. Phys. 93 045003
Cerezo M, Verdon G, Huang H-Y, Cincio L and Coles P J 2022 Challenges and opportunities in quantum Nature Computational Science 2 567–76
Abdi H and Williams I J 2010 Principal component analysis Wiley Interdiscip. Rev. Comput. Stat. 2 433–59
Li D 2012 The MNIST database of handwritten digit images for machine learning research IEEE Signal Process Mag. 29 141–2
Mohri M, Rostamizadeh A and Talwalkar A 2018 Foundations of Machine Learning (Cambridge, MA: MIT Press)