RESEARCH ARTICLE

Expressive quantum supervised machine learning using Kerr-nonlinear parametric oscillators

Yuichiro Mori[1] · Kouhei Nakaji[1,2,3] · Yuichiro Matsuzaki[1,4] · Shiro Kawabata[1,4]

Abstract

Quantum machine learning with variational quantum algorithms (VQA) has been actively investigated as a practical algorithm in the noisy intermediate-scale quantum (NISQ) era. Recent researches reveal that the data reuploading, which repeatedly encode classical data into quantum circuit, is necessary for obtaining the expressive quantum machine learning model in the conventional quantum computing architecture. However, the data reuploading tends to require a large amount of quantum resources, which motivates us to find an alternative strategy for realizing the expressive quantum machine learning efficiently. In this paper, we propose quantum machine learning with Kerr-nonlinear parametric oscillators (KPOs), as another promising quantum computing device. We use not only the ground state and first excited state but also higher excited states, which allows us to use a large Hilbert space even if we have a single KPO. Our numerical simulations show that the expressibility of our method with only one mode of the KPO is much higher than that of the conventional method with six qubits. Our results pave the way towards resource-efficient quantum machine learning, which is essential for the practical applications in the NISQ era.

Keywords Machine learning · KPO · Expressibility · NISQ · Variational quantum algorithms

1 Introduction

The quantum computers have attracted much attention due to its potential impact on quantum chemistry (Aspuru-Guzik et al. 2005; Cao et al. 2019; McArdle et al. 2020; Armaos et al. 2020), machine learning (Schuld et al. 2015; Biamonte et al. 2017; Schuld and Killoran 2022), cryptography (Shor 1994, 1997; Lenstra 2000), search problems (Grover 1996), and so on. With advancements in quantum technology, commercially available quantum computers have become a reality. In principle, we could realize a fault-tolerant quantum computer, if the number of qubits is more than 10 million with a fidelity around 0.999 (Jones et al. 2012; Devitt et al. 2013; Gidney and Ekerå 2021). However, in the current device, the available number of qubits is an order of 500 or less, which is much smaller than that required for the fault-tolerant quantum computation. A more feasible scenario to be realized in the near future is the so-called NISQ regime (Preskill 2018; Bharti et al. 2022).

Numerous quantum algorithms have been designed for execution on NISQ devices. Among these, VQAs are considered some of the most promising applications for NISQ devices (Bharti et al. 2022; Endo et al. 2021). Specifically, quantum machine learning has emerged as an appealing use case for VQAs. As a NISQ algorithm, quantum machine learning has been predominantly investigated in the context of qubit-based systems. Recent studies have shown that data reuploading, the process of repeatedly encoding classical

✉ Yuichiro Mori
  mori-yuichiro.9302@aist.go.jp

  Kouhei Nakaji
  kohei.nakaji@utoronto.ca

  Yuichiro Matsuzaki
  ymatsuzaki872@g.chuo-u.ac.jp

  Shiro Kawabata
  s-kawabata@aist.go.jp

1   Global Research and Development Center for Business by Quantum-AI Technology (G-QuAT), National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568, Japan

2   Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, Ontario, Canada

3   Quantum Computing Center, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan

4   NEC-AIST Quantum Technology Cooperative Research Laboratory, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8568, Japan

data into quantum circuits, is essential for achieving expressive quantum machine learning models within traditional quantum computing frameworks (Pérez-Salinas et al. 2020; Gil Vidal and Theis 2020; Schuld et al. 2021). However, data reuploading often demands much quantum resources. This encourages us to seek alternative approaches to achieve expressive quantum machine learning. We could adopt a photonic device where fock states can be used and the necessary quantum resources for the data embedding with this device could be different from that with the conventional approach using qubits (Killoran et al. 2019; Steinbrecher et al. 2019; Volkoff 2021; Gan et al. 2020; Liu et al. 2023).

On the other hand, the KPO is one of the candidates to realize quantum computation (Milburn and Holmes 1991; Wielinga and Milburn 1993; Cochrane et al. 1999). The KPO is a parametric oscillator with large Kerr nonlinearity. This Kerr nonlinearity can be used to generate cat states. We can realize the KPO by using superconducting resonators with Josephson junctions (Bourassa et al. 2012; Meaney et al. 2014). The KPO is one of the candidates to perform gate-type quantum computation (Cochrane et al. 1999; Goto 2016a; Puri et al. 2017) and quantum annealing (Goto 2016b; Puri et al. 2017), and the KPO qubit is realized experimentally (Grimm et al. 2020). It is known that the KPO qubit is highly tolerant to bit-flip errors, and we can exploit this property to reduce the overhead for fault-tolerant quantum computation (Puri et al. 2017; Masuda et al. 2022).

In this paper, we propose to use the KPO for the supervised machine learning with a variational algorithm. KPO is a bosonic system, and we can in principle use the infinitely large Hilbert space with the single KPO. Also, unlike the conventional approach to use parametrized gates, we use a natural Hamiltonian dynamics where we change the Hamiltonian parameter to implement the variational algorithm. We numerically study the performance of our method to use the KPO with that of the conventional method with qubits.

In our method, we start from a coherent state with an amplitude of α. Importantly, we numerically find that, by changing the amplitude, we can tune the expressibility. Since we encode the input classical data by using the detuning of the KPO, we can include a higher frequency as we increase the amplitude of the coherent state. We expect that the high-frequency terms will improve the expressibility, and we confirm this point by using numerical simulations. As the expressivity increases, on the other hand, more often, overfitting occurs, and so our method allows us to optimize the expressibility by tuning the amplitude of the coherent state.

This paper is organized as follows. In Section 2, we review the physics of single and multiple KPO systems. The latter is called KPO network. In Section 3, we explain a standard supervised machine learning algorithm as a NISQ algorithm, and a supervised machine learning algorithm for KPO is proposed based on the ideas in Section 4. We performed

numerical simulations to validate our proposed method. In Section 5, we explain the simulations and the results precisely. Finally, we conclude with some final thoughts in Section 6.

2 KPO

KPO is a bosonic system with a nonlinear effect called Kerr nonlinearity. Here, we first describe a single KPO and next explain a network of KPOs that have been used for a gate-type quantum computer or quantum annealing.

First, in a frame rotating at half the pump frequency of the parametric drive and in the rotating wave approximation, the Hamiltonian of the single KPO is written as Goto (2016b, 2019)

$$\hat{H} = \chi \hat{a}^{\dagger 2}\hat{a}^2 + \Delta \hat{a}^\dagger \hat{a} \\ - p(\hat{a}^2 + \hat{a}^{\dagger 2}) + r(\hat{a} + \hat{a}^\dagger), \tag{1}$$

where χ, Δ, p, and r are the Kerr nonlinearity, the detuning, the pump amplitude of the parametric drive, and the strength of the coherent drive, respectively.

We can easily tune Δ, p, and r during the experiment by changing the parameters of the external driving fields. Although we can tune χ by changing magnetic flux penetrating the superconducting loop of the KPO, the dynamic range is typically small, and therefore, we assume that χ is fixed at a specific value.

The coherent state is defined by

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_k \frac{\alpha^k}{\sqrt{k!}}|k\rangle, \tag{2}$$

where $|k\rangle$ are the fock states. The system is initially prepared in the coherent state in our method. For a linear resonator, we can prepare the coherent state by adding the coherent driving term $r(\hat{a}+\hat{a}^\dagger)$. However, due to the term $\chi \hat{a}^{\dagger 2}\hat{a}^2$ in Eq. 1, we cannot prepare the coherent state just by adding the coherent drive. Instead, we can prepare the coherent state by using the KPO as follows. By setting $p = r = 0$, the Hamiltonian Eq. 1 becomes

$$\hat{H} = \chi \hat{a}^{\dagger 2}\hat{a}^2 + \Delta \hat{a}^\dagger \hat{a}. \tag{3}$$

If $\Delta > \chi$ is satisfied, the ground state of this Hamiltonian becomes the vacuum state $|0\rangle$. On the other hand, when Δ and r are zero, Eq. 1 can be rewritten as

$$\hat{H} = \chi \left(\hat{a}^{\dagger 2} - \frac{p}{\chi}\right)\left(\hat{a}^2 - \frac{p}{\chi}\right) - \frac{p^2}{\chi}, \tag{4}$$

and the ground state is in the eigenspace that is spanned by two coherent states $|\sqrt{p/\chi}\rangle$ and $|-\sqrt{p/\chi}\rangle$. By adding a coherent drive as a perturbation, we can solve the degeneracy, and the ground state becomes approximately $|\sqrt{p/\chi}\rangle$ with a negative value of r. If we prepare a vacuum state with the Hamiltonian of Eq. 3, the system is in the ground state. By adiabatically changing the Hamiltonian from Eq. 3 to Eq. 4, we obtain the coherent state $|\sqrt{p/\chi}\rangle$ due to the adiabatic theorem. This operation is frequently used in quantum annealing with KPOs (Goto 2016a, b, 2019; Puri et al. 2017).

Next, the Hamiltonian of multiple KPOs, which is called a KPO network, is written as

$$
\begin{aligned}
\hat{H} = &\sum_{j=1}^{K} \chi_j \hat{a}_j^{\dagger 2} \hat{a}_j^2 + \Delta_j \hat{a}_j^{\dagger} \hat{a}_j \\
&- p_j(\hat{a}_j^2 + \hat{a}_j^{\dagger 2}) + r_j(\hat{a}_j + \hat{a}_j^{\dagger}) \\
&+ \sum_{j>j'}^{K} \left(J_{jj'} \hat{a}_j^{\dagger} \hat{a}_{j'} + J_{jj'}^* \hat{a}_{j'}^{\dagger} \hat{a}_j \right).
\end{aligned}
\tag{5}
$$

where K denotes the number of KPOs and $J_{jj'}$ denotes the coupling strength between KPOs. Here, we assume that we fix the values of χ_j and $J_{jj'}$ during the experiment, while we can control the values of Δ_j, p_j, and r_j.

If $J_{jj'}$ is zero, we can independently perform the adiabatic state preparation described above and prepare the following state:

$$
\bigotimes_{j=1}^{K} |\alpha_j\rangle.
\tag{6}
$$

Here, each α_j is the eigenvalue of $|\alpha_j\rangle$ with the annihilation operator on the j-th KPO $\hat{a}_j$.

It is worth mentioning that even when $J_{jj'}$ is nonzero, we can prepare the product of the coherent state as follows. Let us assume that Δ_j, r_j, and J_{ij} are much smaller than p_j and χ_j. In this case, the last terms of the Hamiltonian Eq. 5 can be interpreted as the longitudinal-field Ising Hamiltonian in a coherent state basis. If J_{ij} is negative, we have a ferromagnetic Hamiltonian. Moreover, by setting J_{ij} to be much smaller than r_j, the state in Eq. 6 can be a ground state, and so we can prepare this state in an adiabatic way. Also, a coupling scheme of KPOs with high fidelity has already been proposed theoretically (Goto 2019; Masuda et al. 2022; Aoki et al. 2024).

supervised learning task, a number of training set $\{(\boldsymbol{x}_m, \boldsymbol{y}_m)\}_{m=1}^{N}$ are given. Here, all input data $\boldsymbol{x}_m$ (output data $\boldsymbol{y}_m$) are d_x (d_y) dimensional arrays. Suppose that there is a hidden relationship between an input data $\boldsymbol{x}$ and the output data $\boldsymbol{y}$ as $\boldsymbol{y} = \tilde{f}(\boldsymbol{x})$ with a function $\tilde{f}$. The objective of the task is to find the hidden relationship $\tilde{f}$ from the training data. More specifically, we define the model function f and optimize it so that it becomes close to $\tilde{f}$ by using the training data.

In most of the quantum machine learning with near-term devices, we use a parameterized quantum circuit to construct a model function. More precisely, by tuning a parameter, we try to minimize a cost function. In usual cases, we choose the mean squared error

$$
L(\boldsymbol{\theta}) = \frac{1}{N} \sum_{m=1}^{N} \left| f(\boldsymbol{x}_m; \boldsymbol{\theta}) - \boldsymbol{y}_m \right|^2,
\tag{7}
$$

for the cost function. Here, N is the number of data sets, $f(\boldsymbol{x}; \boldsymbol{\theta})$ is an array as an output of the parameterized quantum circuit, and $\boldsymbol{\theta}$ is the corresponding parameter. Let us summarize such a quantum machine learning as follows.

1. Prepare an initial state $|\psi\rangle$, and apply an input gate $\hat{U}(\boldsymbol{x})$ to encode the input data $\{\boldsymbol{x}_i\}$.
2. Apply a parameterized unitary $\hat{V}(\boldsymbol{\theta})$ to the state.
3. Measure the expectation values of an observable $\hat{M}$, and we define the function as $f(\boldsymbol{x}; \boldsymbol{\theta}) = \langle \hat{M} \rangle$.
4. By repeating the above three steps, minimize the cost function L by tuning the parameter $\boldsymbol{\theta}$ iteratively.

The function $f(\boldsymbol{x})$ is represented as

$$
f(\boldsymbol{x}; \boldsymbol{\theta}) = \langle \psi | \hat{U}^{\dagger}(\boldsymbol{x}) \hat{V}^{\dagger}(\boldsymbol{\theta}) \hat{M} \hat{V}(\boldsymbol{\theta}) \hat{U}(\boldsymbol{x}) | \psi \rangle.
\tag{8}
$$

According to a previous study (Schuld et al. 2021), we may not expect high expressibility with a parametrized quantum circuit using single-qubit rotations in the NISQ era. In fact, the study shows that only a sinusoidal curve can be obtained as Eq. 8 with using a qubit and single-qubit rotations, and if we want different functions as an output, we need to prepare more qubits or obtain other outputs than Eq. 8 with adding another operation called data reuploading. However, neither increasing the number of qubits nor increasing the number of gate operations that cause noise is desirable for the NISQ algorithm.

3 Quantum supervised machine learning as a NISQ algorithm

In this section, let us review a quantum supervised machine learning as a preparation for introducing our model. In a

4 Quantum supervised machine learning with KPO

We introduce our method to use the KPO for supervised quantum machine learning. We begin by describing a simplified

scenario with $d_x = d_y = 1$ by using the single KPO. Next, we explain how to use the KPO network for supervised quantum machine learning with $d_x = d_y = 1$. Finally, we describe a scenario to implement supervised quantum machine learning with $d_x > 1$ and/or $d_y > 1$ by using the KPO network.

4.1 $d_x = d_y = 1$ case

4.1.1 Single KPO

In our paper, the initial state is set to be a coherent state. Also, to upload the classical data, we could adopt $\hat{U}(x) = e^{-i\pi x \hat{n}}$, where $\hat{n}$ is the number operator defined by $\hat{n} = \hat{a}^\dagger \hat{a}$. However, if we use the KPO, it is difficult to realize in situ tunability of the nonlinearity χ. Assuming that we fix the value of χ during the experiment, we will adopt the following operator to upload the classical data

$$\hat{U}(x) = e^{-i\tilde{\chi}\hat{n}^2 - i\pi x \hat{n}}, \tag{9}$$

where we have,

$$\tilde{\chi} = t_d \chi, \tag{10}$$
$$\pi x = t_d (\Delta - \chi). \tag{11}$$

In the actual experiment, we can easily tune the time duration t_d and detuning Δ. Throughout this paper, we fix the value of $\tilde{\chi}$.

Let us define a set of unitary operators $\hat{V}_i(\Delta_i, p_i, r_i)$

$$\hat{V}_i(\Delta_i, p_i, r_i) = e^{-i\tau \hat{H}}. \tag{12}$$

where $\hat{H}$ denotes the Hamiltonian of the KPO Eq. 1 and τ denotes an evolution time by the Hamiltonian. By turning on and off the parameters of the Hamiltonian, we can construct a unitary operator

$$\hat{V}(\boldsymbol{\theta}) = \prod_i^D \hat{V}_i(\Delta_i, p_i, r_i), \tag{13}$$

where D is the number of combinations of (Δ_i, p_i, r_i). Here, $\boldsymbol{\theta}$ corresponds to a set of parameters $\{\Delta_j, p_j, r_j\}_{j=1}^D$. For simplicity, we define

$$\theta_k := \begin{cases} \Delta_i & k = 3i - 2, \\ p_i & k = 3i - 1, \\ r_i & k = 3i, \end{cases} \tag{14}$$

with $i = 1, \ldots, d$. We choose $\hat{a} + \hat{a}^\dagger = \hat{M}$ as the observable to be measured. Since a bosonic system has an infinite dimensional Fock space, even a single KPO may have the

ability to approximate the target function, while the previous approach required multiple qubits to represent the target function.

To minimize the cost function, we need to tune the parameter $\boldsymbol{\theta}$. For this purpose, we should adopt a classical algorithm to show how we should update the parameters based on the expectation value of $\hat{M}$.

Several types of classical algorithms are used to update $\boldsymbol{\theta}$. One of them is the gradient descent method to use the gradient of the cost function. If we construct the unitary operator $\hat{V}(\boldsymbol{\theta})$ by using a sequence of parameterized gates, we can use the so-called parameter shift rule (Mitarai et al. 2018; Wierichs et al. 2022) to determine the gradient. On the other hand, since we use the Hamiltonian dynamics to realize the unitary operator $\hat{V}(\boldsymbol{\theta})$, it is not straightforward to use the parameter shift rule. We could use a numerical differentiation where we measure small changes in $f(x; \boldsymbol{\theta})$ when we incrementally increase $\boldsymbol{\theta}$ by changing $\boldsymbol{\theta}$ in small increments and detecting the resulting small changes in the output $f(x; \boldsymbol{\theta})$. However, to detect the small changes, this method requires a large number of measurements.

If we cannot use a sufficient number of shots, we could adopt an optimization using the Nelder-Mead or Powell method, which does not use the information of gradients. Throughout this paper, we use the Nelder-Mead method (Nelder and Mead 1965) for our simulation.

Our method to use the single KPO needs to access higher excited states in the Fock space, which may cause experimental difficulties. This problem may be circumvented by using the KPO network.

4.1.2 KPO network

Next, we consider a case using a KPO network. We prepare the product state of the coherent state (6) as the initial state. To upload the classical data, we apply the following operator on the j-th KPO,

$$\hat{U}_j(x) = e^{-i\tilde{\chi}\hat{n}_j^2 - i\pi x \hat{n}_j}, \tag{15}$$

where we have $\tilde{\chi}_j = t_d \chi_j$ and $\pi x = t_d \Delta_j$. We define a unitary operator with $3K$ parameters,

$$\hat{V}(\vec{\Delta}, \vec{p}, \vec{r}) = e^{-i t_d \hat{H}}, \tag{16}$$

where $\hat{H}$ is given by Eq. 5. Here, $\vec{\Delta} = (\Delta_1, \Delta_2, \cdots, \Delta_K)$, $\vec{p} = (p_1, p_2, \cdots, p_K)$ and $\vec{r} = (r_1, r_2, \cdots, r_K)$ are K dimensional arrays.

If we need more than $3K$ adjustable parameters, we could consider a different combination of $\vec{\Delta}$, $\vec{p}$, and $\vec{r}$. Let us define a set of such a combination as $\{\vec{\Delta}_i, \vec{p}_i, \vec{r}_i\}_{i=1}^D$ where D is the number of the combination. Thus, we can generate D

different unitary operators based on Eq. 16. When we sequentially implement these, the unitary operator is given as

$$\hat{V}(\boldsymbol{\theta}) = \prod_{i}^{D} \hat{V}_i(\vec{\Delta}_i, \vec{p}_i, \vec{r}_i). \tag{17}$$

Here, $\boldsymbol{\theta}$ corresponds to a set of parameters as described below.

$$(\vec{\Delta}_1, \vec{p}_1, \vec{r}_1, ..., \vec{\Delta}_D, \vec{p}_D, \vec{r}_D)$$
$$= (\Delta_{11}, \Delta_{12}, ..., \Delta_{1K}, p_{11}, p_{12}, ..., p_{1K}, ..., r_{DK}).$$

After applying $\hat{V}(\boldsymbol{\theta})$ given by Eq. 17, we measure an observable $\hat{M}$. For this $\hat{M}$, we can choose an observable of $\hat{a}_1 + \hat{a}_1^{\dagger}$, for example.

4.2 $d_x > 1$ and/or $d_y > 1$ case

We describe our method to implement supervised quantum machine learning with $d_x > 1$ and/or $d_y > 1$ by using the KPO network. Let us assume $K \geq d_x$. For $j = 1, 2, \cdots, d_x$, we define

$$\hat{U}_j(x_j) = e^{-i\tilde{\chi}\hat{n}_j^2 - i\pi x_j \hat{n}_j}. \tag{18}$$

To upload the classical data, we use a unitary operator of $\prod_{j=1}^{d_x} \hat{U}_j(x_j)$.

Subsequently, we apply $\hat{V}(\boldsymbol{\theta})$ in Eq. 17 and measure a set of observable $\{\hat{M}_k\}_{k=1}^{d_y}$. The expectation value of $\hat{M}_k$ corresponds to the k-th component of $\boldsymbol{y}$. By repeating these steps, we update the parameter $\boldsymbol{\theta}$ to minimize the cost function. In principle, we could use the single KPO with $d_x > 1$ and $d_y > 1$, and we discuss such an example in Appendix C.

4.3 Potential advantage to use KPOs

Even if we can use only a single KPO, the function obtained as Eq. 8 is expected to exhibit a large expressibility. Similar to the previous study (Schuld et al. 2021), we construct the Fourier spectrum of Eq. 8 in the case of a single KPO as $d_x = d_y = 1$.

When χ is negligibly small, we obtain

$$f(x; \boldsymbol{\theta}) = \langle \alpha | e^{i\pi x\hat{n}} \hat{V}^{\dagger}(\boldsymbol{\theta}) \hat{M} \hat{V}(\boldsymbol{\theta}) e^{-i\pi x\hat{n}} | \alpha \rangle$$
$$= e^{-|\alpha|^2} \sum_{k,l=0}^{\infty} \langle k | \hat{V}^{\dagger}(\boldsymbol{\theta}) \hat{M} \hat{V}(\boldsymbol{\theta}) | l \rangle \frac{\alpha^l \alpha^{*k}}{\sqrt{k! l!}} e^{i\pi x(k-l)}, \tag{19}$$

which is the Fourier series. Importantly, there are high-frequency terms in this form, and the number of terms is infinite. This would improve expressibility. A similar discussion has been made by Gan et al. in the context of a multi-mode photonic device, which supports our claim (Gan et al. 2020).

If we can provide an appropriate $\hat{V}(\boldsymbol{\theta})$ and $\hat{M}$, we could represent any function that can be represented by the Fourier series. Moreover, previous research shows that the Kerr nonlinearity could enhance the performance of a specific scheme of quantum machine learning (Liu et al. 2023), and so our method to utilize the Kerr nonlinearity might improve the expressibility.

On the other hand, if we use ordinary qubits, the number of high-frequency terms is limited by the finite number of qubits. This could limit the expressibility, as suggested in Schuld et al. (2021). To improve the expressibility, we could increase the number of qubits (Schuld et al. 2021) or circuit depth. However, it is difficult to increase the number of qubits or circuit depth in the NISQ device.

5 Simulations and results

To evaluate the performance of our proposed method, we perform numerical simulations for $d_x = d_y = 1$ and compare the results of our method with that of the conventional one (Mitarai et al. 2018). Specifically, we perform the fitting of $\tilde{f}(x) = e^{-36x^2}$ (Gaussian), $|x|$, and $0.4 \sin(4\pi x) + 0.5 \sin(6\pi x)$. Also, we perform the fitting of the square wave defined as

$$\tilde{f}(x) = \begin{cases} 1 & (|x| < 0.4) \\ 0 & (|x| \geq 0.4). \end{cases} \tag{20}$$

We create the training set as follows. We set $N = 100$. First, we randomly choose a value between -1 and 1 and adopt these values as x_m. Next, for each x_m, we calculate $\tilde{f}(x_m)$ by using the given function $\tilde{f}$ and assign this value as y_m.

For our method to use a single KPO, we choose $\chi = 0.1$, $t_d = \tau = 0.7$, $\hat{M} = \hat{a} + \hat{a}^{\dagger}$, and $D = 12$. Also, we set the cutoff of the Hilbert space dimension as 25.

For the conventional method (Mitarai et al. 2018), we set depth $D = 2$, the number of qubit $K = 6$, time step $\tau = 10$, and $\hat{M} = 2Z^{(1)}$. Precise setups of the conventional method are given in Appendix B. Here, for a fair comparison, we set the number of parameter $\boldsymbol{\theta}$ as 36, which is equal to that of our method.

We show the results of the fitting in Fig. 1. Our method approximates all functions better than the conventional method. In order to compare the expressibility more clearly, we define a Fourier transform as

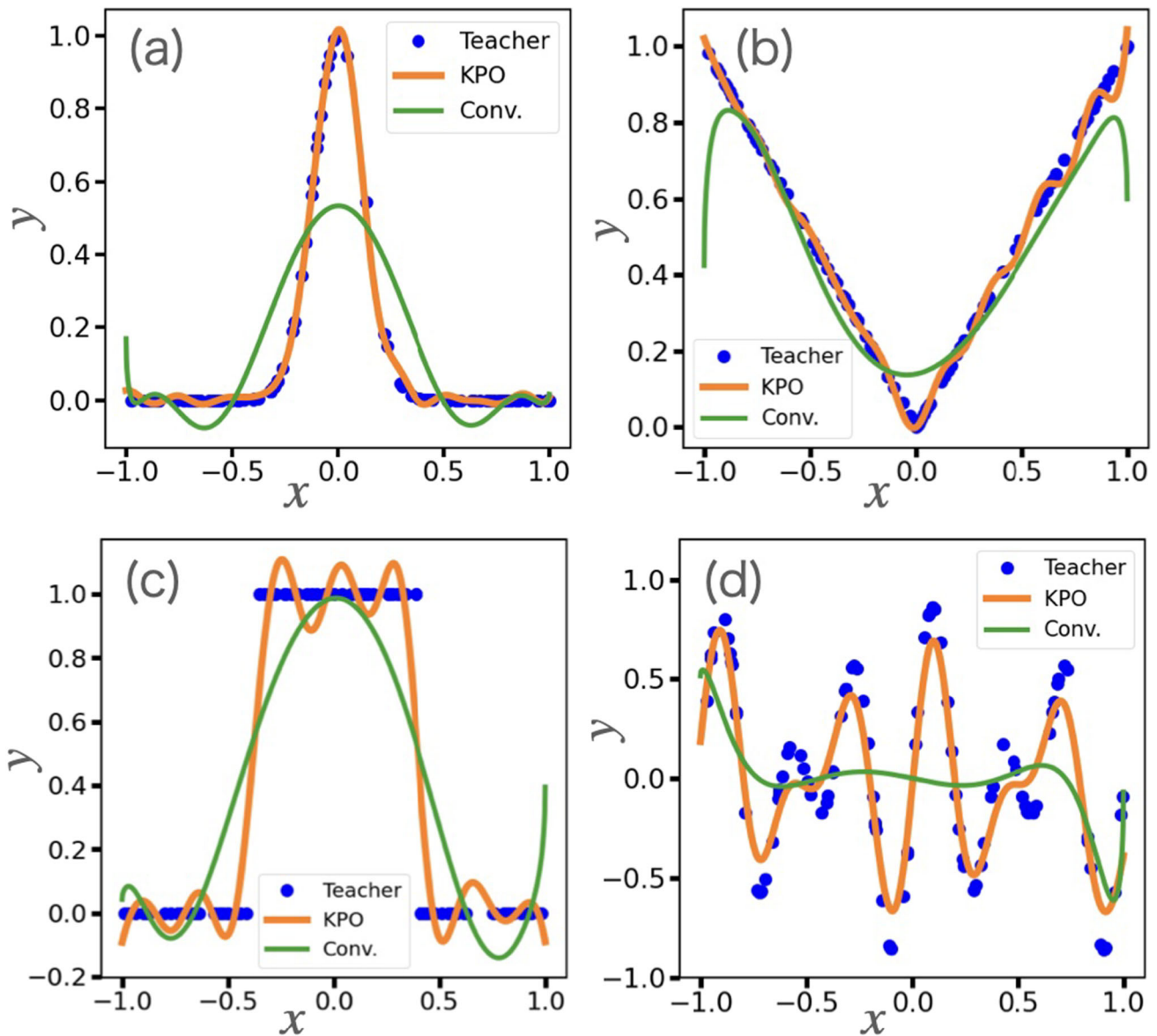$$\hat{F}(\nu) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} dx \, F(x) e^{-2\pi i \nu x}, \tag{21}$$

Fig. 1 Demonstration of quantum machine learning to represent functions. Blue dots indicate the teacher data. KPO (conventional) indicates the output by our (conventional) method after the optimization. We fit **a** e^{-36x^2}, **b** $|x|$, **c** square wave, and **d** $0.4\sin(4\pi x) + 0.5\sin(6\pi x)$, respectively

for any function $F(x)$, and we plot the absolute value of $\hat{f}(\nu)$ in Fig. 2. As can be easily seen in (b) and (d), the results by our method contain more Fourier components than that by the conventional method.

Also, we plot the value of the cost function after the optimization by our method and compare this with the conventional method in Table 1.

Next, let us discuss the case of the KPO network for $d_x = d_y = 1$. Here, we use $\chi_1 = \chi_2 = 1$, $J_{12} = -0.1$, $K = 2$, and $t_d = \tau = 1$. Also, by choosing $D = 6$, we set the total number of parameters as 36, which is equal to that of the single KPO.

We could choose $\hat{M} = (\hat{a}_1 + \hat{a}_1^\dagger) \otimes (\hat{a}_2 + \hat{a}_2^\dagger)$ for our numerical simulations. However, it is not straightforward to measure such a non-local observable with the KPO. So, instead, we consider two observables $\hat{M}_1 = \hat{a}_1 + \hat{a}_1^\dagger$ and $\hat{M}_2 = \hat{a}_2 + \hat{a}_2^\dagger$. Also, we represent the function as $f(x; \boldsymbol{\theta}) = \langle \hat{M}_1 \rangle \langle \hat{M}_2 \rangle$. We perform the fitting of the Gaussian and square wave, which we used in the case of the single KPO. Finally, we set the Hilbert space cutoff dimension of each KPO as 10.

We plot the results in Fig. 3 and compare the performance of our method to use the KPO network with that to use the single KPO. The cost functions after the optimization for
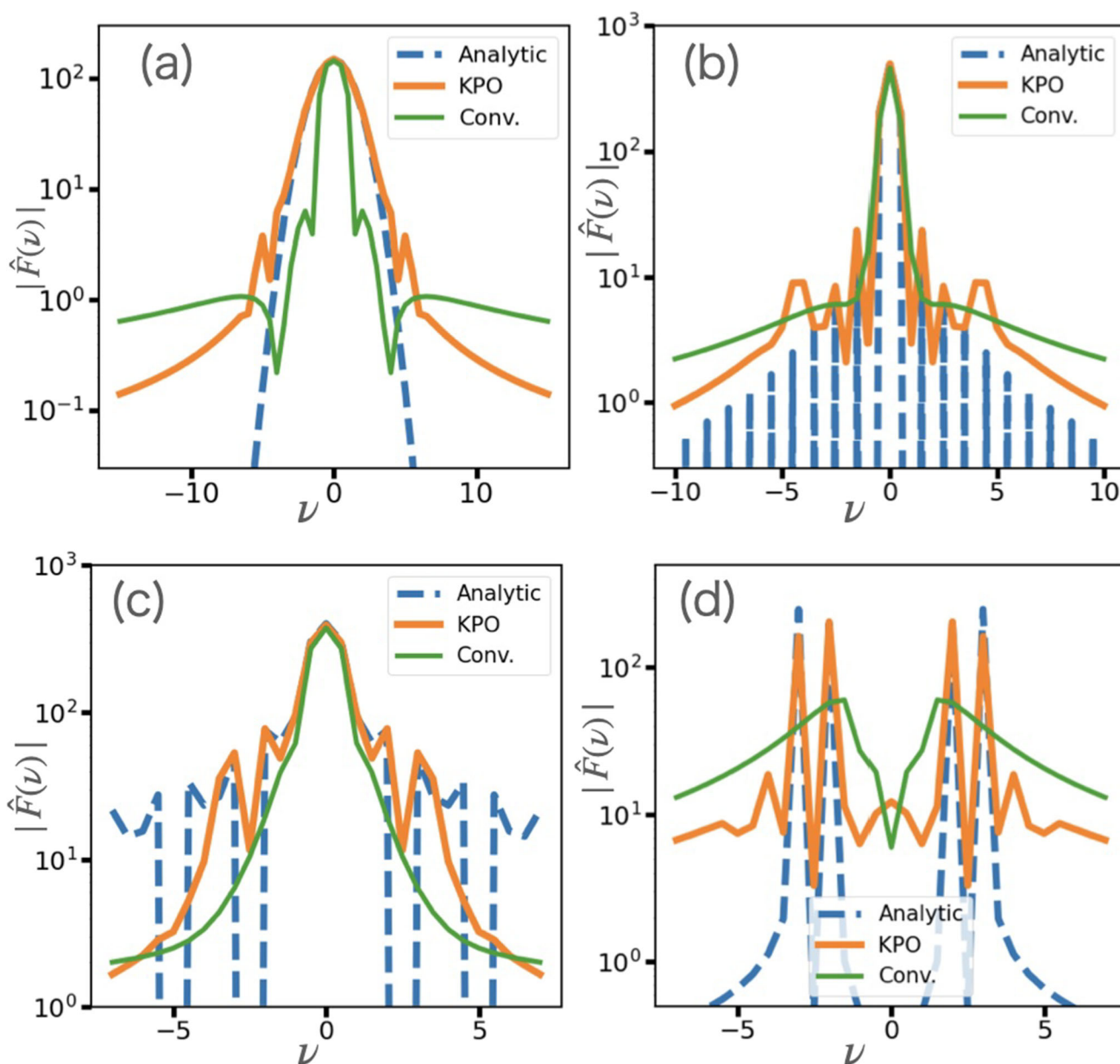
Fig. 2 Plot of the absolute value of the Fourier transform of the function against the frequency. The blue dotted line denotes the function to be fitted. The orange (green) line denotes the output by our (conventional) method after the optimization. The functions are **a** e^{-36x^2}, **b** $|x|$, **c** square wave, and **d** $0.4\sin(4\pi x) + 0.5\sin(6\pi x)$, respectively

1KPO (2KPO) are 1.016×10^{-4} (9.711×10^{-5}) for the Gaussian (e^{-36x^2}) and 1.344×10^{-2} (2.119×10^{-2}) for the

Table 1 Finally obtained values of the cost function

$\tilde{f}(x)$	KPO	Conventional
e^{-36x^2}	1.016×10^{-4}	2.220×10^{-2}
$|x|$	3.388×10^{-4}	7.923×10^{-3}
Square wave	1.344×10^{-2}	4.661×10^{-2}
$0.4\sin(4\pi x) + 0.5\sin(6\pi x)$	1.693×10^{-2}	1.562×10^{-1}

square wave. The performance of our method using the KPO network is similar to that using the single KPO. However, we need to access higher excited states for the case of the single KPO than that of the KPO network, and therefore, we could avoid the experimental difficulties by using the KPO network.

Let us explain the runtime of our scheme. During our simulations, we employed a maximum iteration count of 7200, the default setting provided by Scipy.optimize.minimize (Virtanen et al. 2020), when dealing with 36 variables. The optimization process terminates when the cost function either
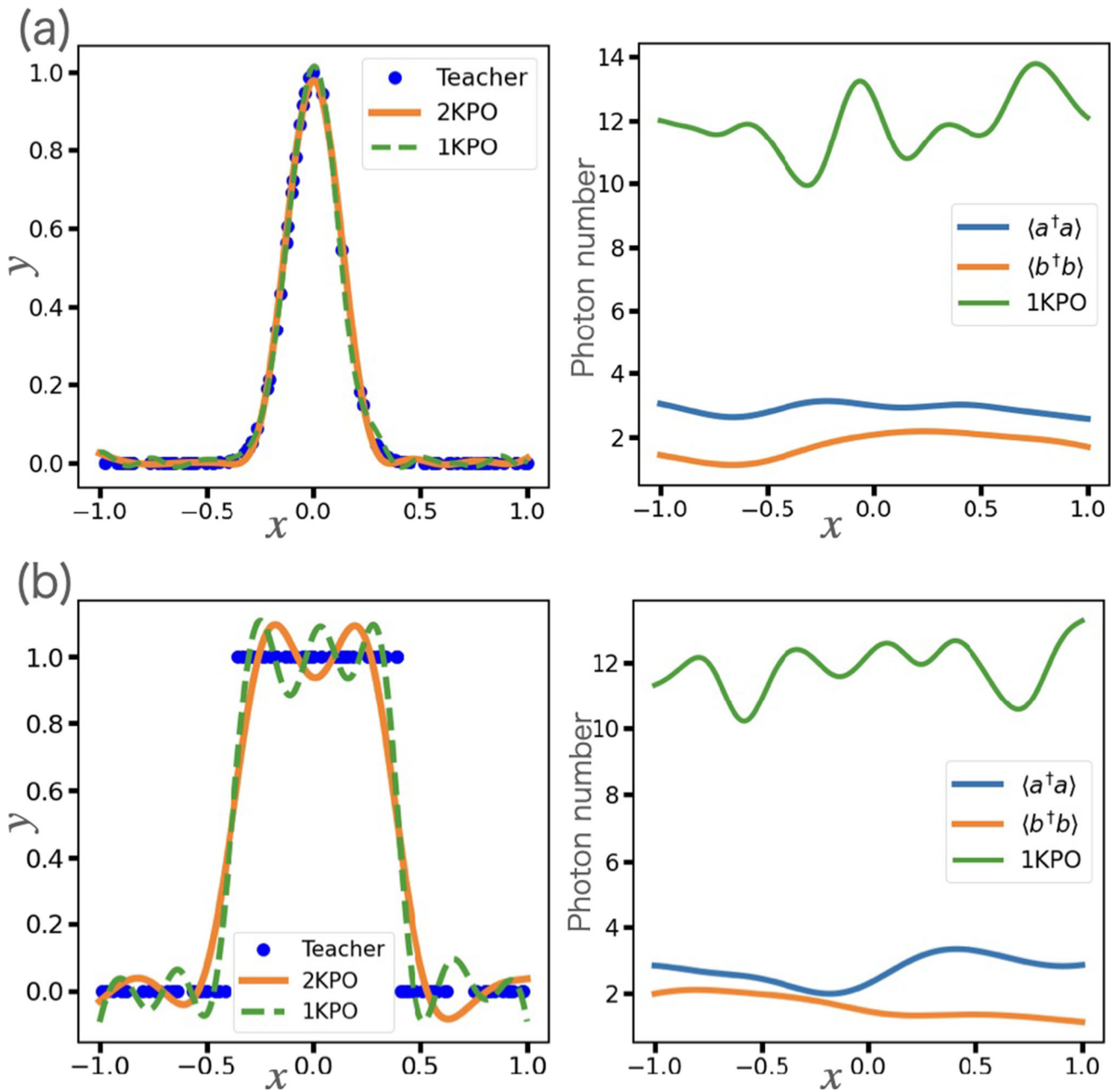
Fig. 3 Demonstration results of our quantum machine learning for e^{-36x^2} (**a**) and square wave (**b**) for the 1KPO and 2KPO cases. Left: the teacher data and the training results. Right: the average photon number (since we evaluated $\langle a^\dagger a \rangle$, the result depends on the variable x)

meets the predefined tolerance level (default value, 10^{-4}) or when the maximum allowed number of iterations is reached. In either case, both the parameter set to minimize the cost function and the iteration number are outputted in Table 2.

In the KPO cases, the number of iterations is equal to or less than that in the conventional cases. The most time-consuming part of the practical runtime of the superconducting circuit is the execution time of two-qubit gates. Importantly, coupling strength between KPOs, as demonstrated in previous work (Yamaji et al. 2022), is approximately 10MHz, which is similar to that of superconducting

transmon qubits (Stehlik et al. 2021). Consequently, these findings highlight that the runtime of our method using KPOs

Table 2 Numbers of iterations

$\tilde{f}(x)$	KPO	Conventional
e^{-36x^2}	5982	6062
$\lvert x \rvert$	5981	6129
Square wave	6015	6083
$0.4\sin(4\pi x) + 0.5\sin(6\pi x)$	6041	6115
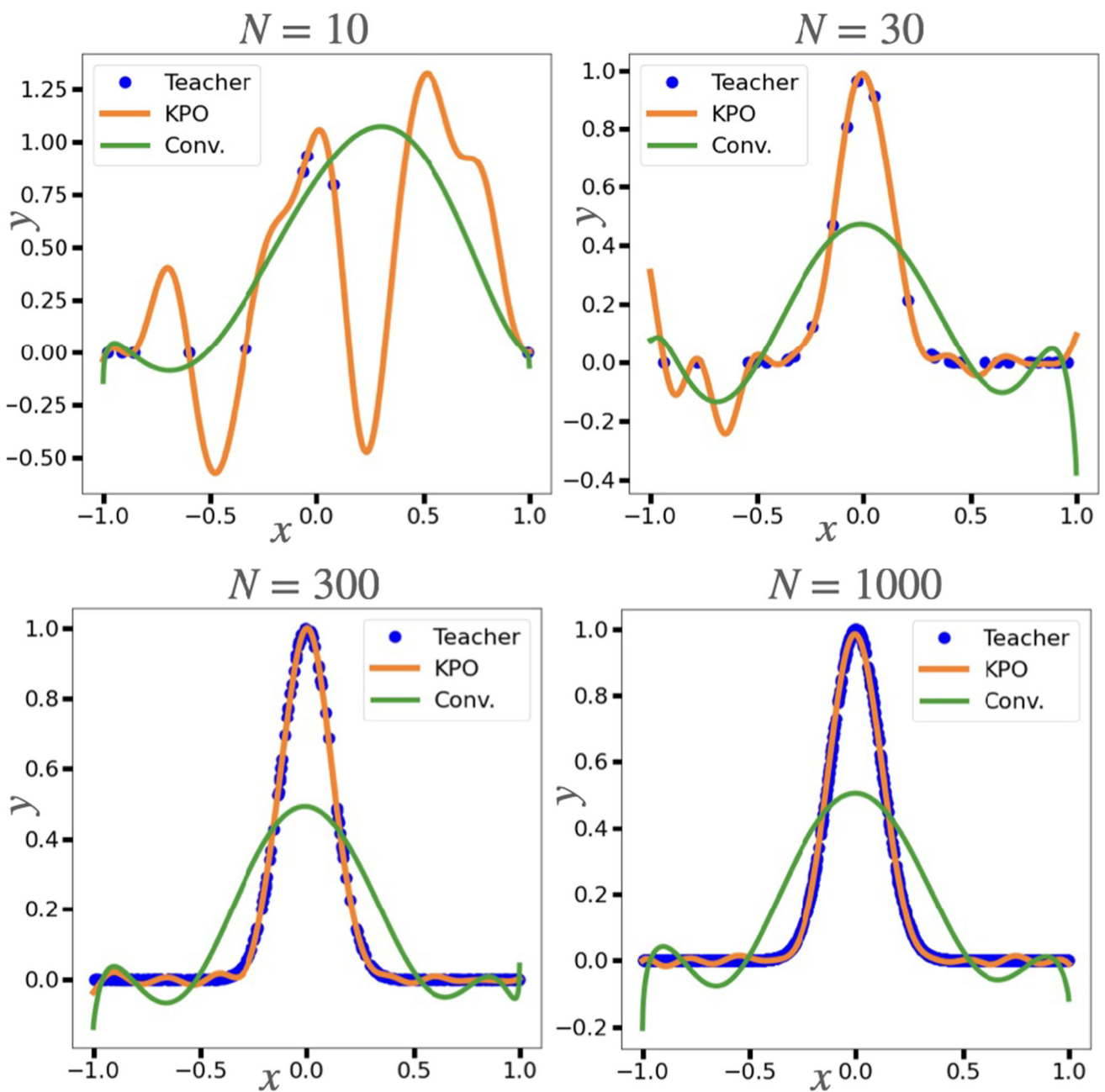
Fig. 4 Results of our quantum machine learning for Gaussian described as e^{-36x^2} with $N = 10, 30, 300, 1000$ training data. When N is small, overfitting occurs

Table 3 Variation of the number of iterations with the number of training data

N	KPO(e^{-36x^2})	Conventional (e^{-36x^2})	KPO (square wave)	Conventional (square wave)
10	4741	6050	1821	6084
30	6009	6051	5977	6106
100	5982	6062	6015	6083
300	6000	6042	6009	6080
1000	6008	6062	6080	6058

is comparable with that of the conventional approach using transmon qubits.

We show how our fitting results depend on the number of training data N in Figs. 4 and 5 and the variation of the number of iterations in Table 3. For small N, our method seems to be susceptible to overfitting due to its inherent high expressiveness. Fortunately, to reduce the impact of overfitting, we can regulate this expressiveness by adjusting the photon number of the initial coherent state, as we will show in Section 5.1.

5.1 α and expressive power

From Eq. 19, we find a tendency that, as we increase (decrease) α, more (less) high-frequency terms are added. Therefore, it is expected that we can control the expressibility by tuning the size of the coherent state prepared as the initial state.

We confirmed this point by numerical simulations. In Fig. 6, we perform numerical simulations for $\alpha = 1, 3$, and 5 with the use of supervised data generated from two func-
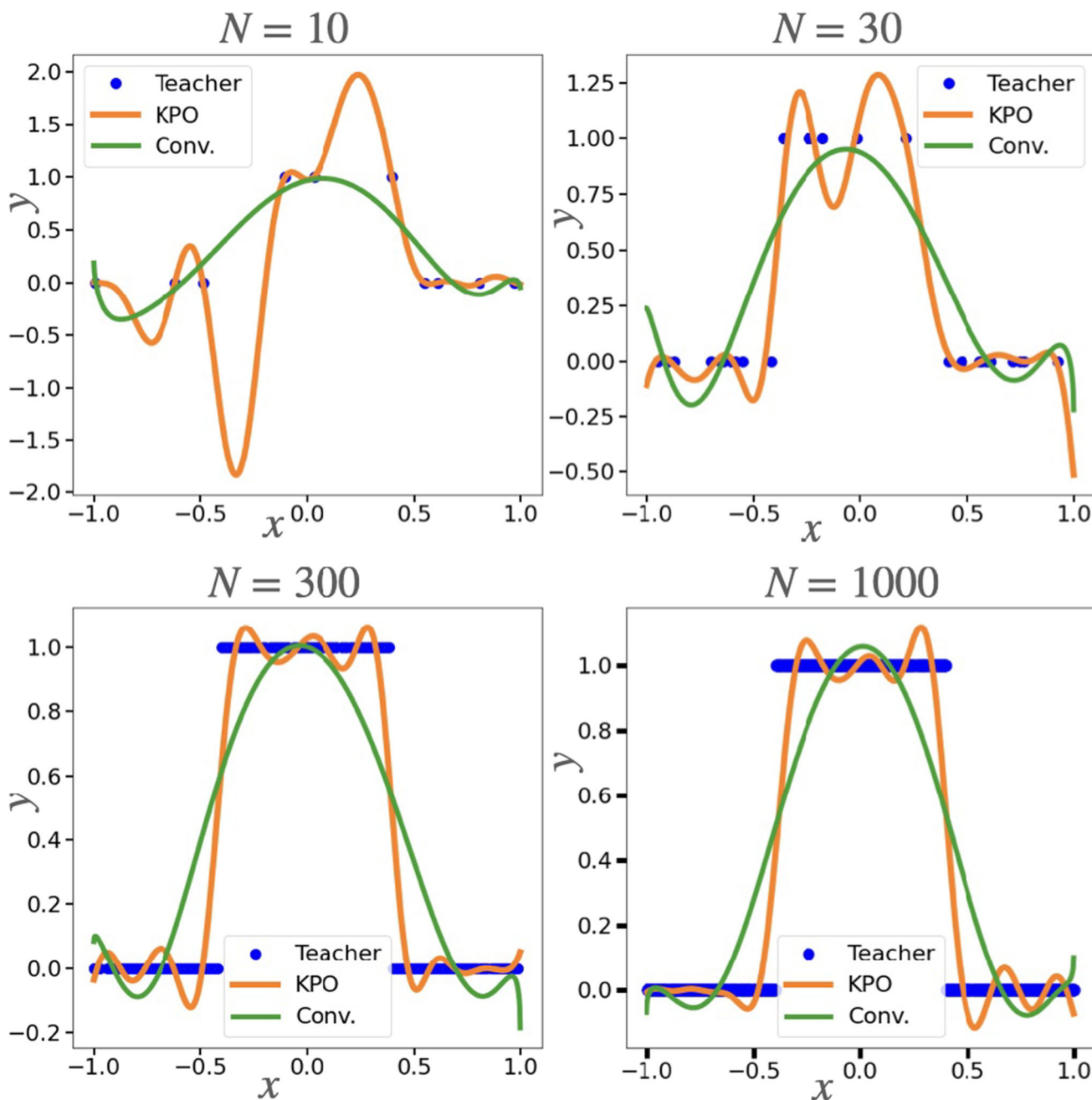
Fig. 5 Results of our quantum machine learning for a square wave function with $N = 10, 30, 300, 1000$ training data. When N is small, overfitting occurs
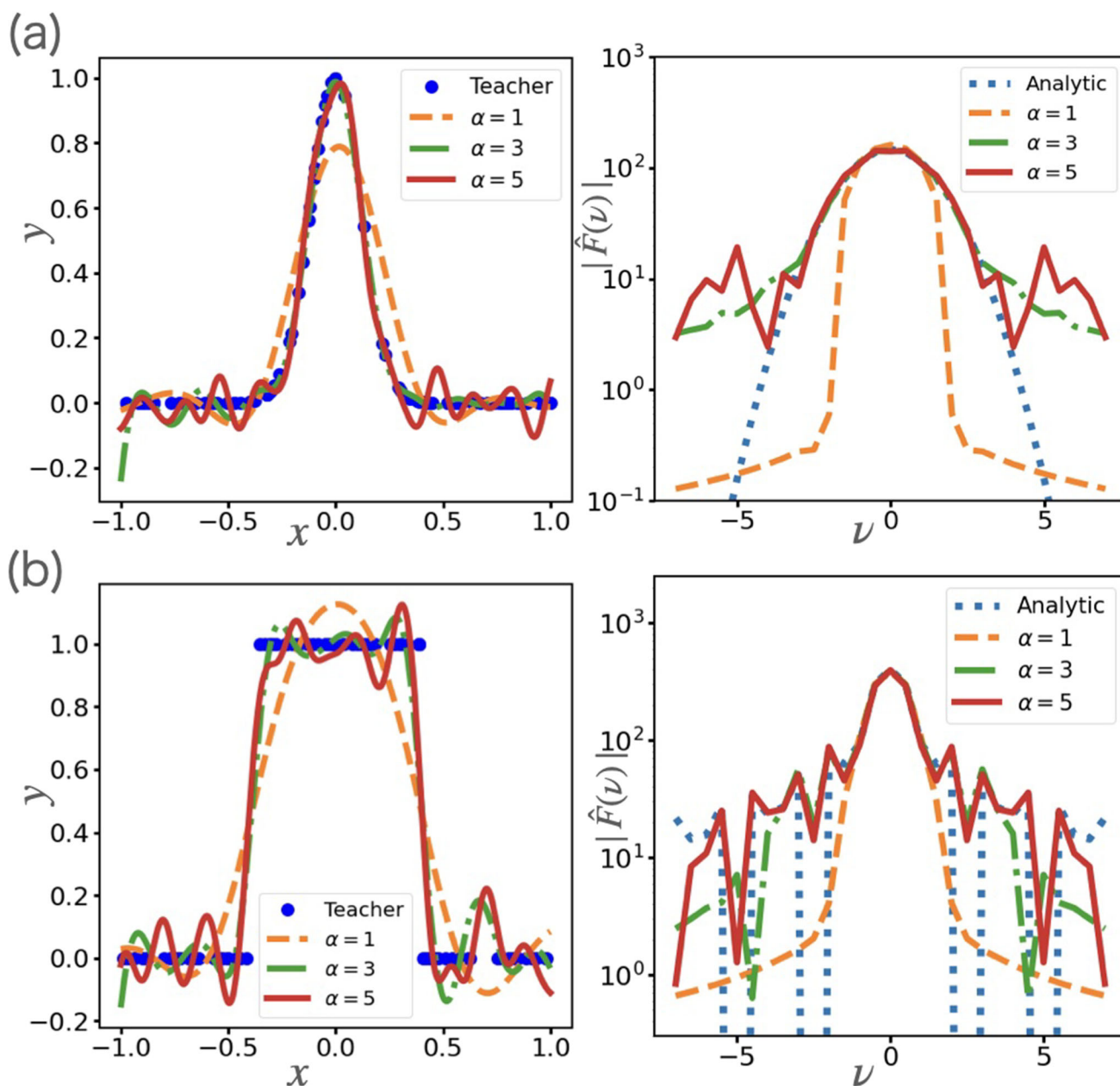
Fig. 6 Demonstration results of our quantum machine learning for e^{-36x^2} (**a**) and the square wave (**b**) for the different $\alpha = 1, 3, 5$ cases. Left: the teacher data and the training results. Right: the Fourier spectrum of the training results

tions, a Gaussian and a square wave. Only for $\alpha = 5$, we change the cutoff dimension of the Hilbert space from 25 to 100 because the average photon number is 25.

In machine learning, there is a trade-off between increasing expressive power and overfitting. This means that, as we increase the expressibility, the problem of the overfitting becomes more severe. In our method, we could tune the parameter α to choose the best point for the fitting.

To illustrate our concept, we performed numerical simulations in which we varied the number of photons in the initial coherent state. As we mentioned before, in Fig. 4, overfitting occurs for a smaller number of the training data N. We apply our method to tune the expressibility to this case. In Fig. 7, we present the results, highlighting that reducing the photon number of the initial coherent state effectively mitigates the impact of overfitting.
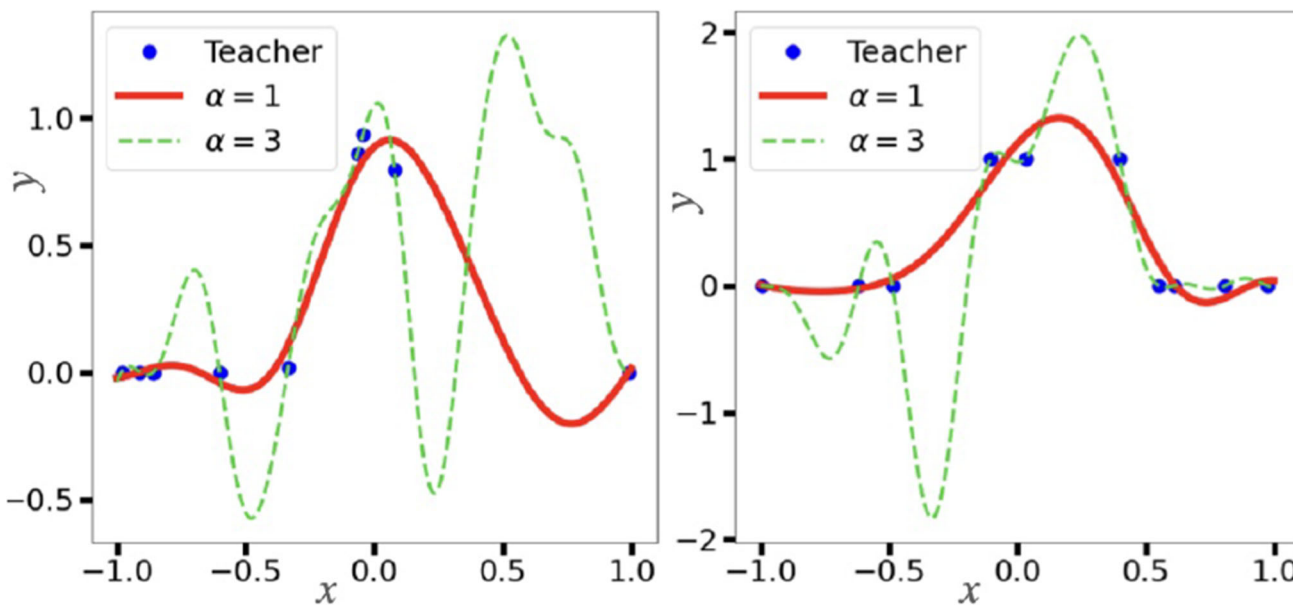
Fig. 7 Results of our quantum machine learning for Gaussian (left side) and the square wave function (right) for $N = 10$ where we change the photon number of the initial coherent state. We successfully mitigate the impact of the overfitting by reducing the photon number of the coherent state

6 Conclusions and discussion

In conclusion, we propose to use the KPO for the quantum supervised machine learning with variational quantum circuits. We numerically show that, although we use a single KPO, the expressibility of our method is higher than the conventional method with six qubits. In our method, we can tune an amplitude of the initial coherent state, and we numerically show that the expressibility increases as we increase the amplitude.

In this paper, we provide proof of concept using a regression problem as an example. Our method also offers advantages due to its expressive nature for other machine learning problems, including classification, generation, reinforcement learning, and sequential learning. Furthermore, we acknowledge that the quantum kernel method (Havlíček et al. 2019) could be another promising application of our approach, as our data encoding methodology into quantum states introduces new types of quantum kernels. Exploring these applications is a promising direction for future research.

In the NISQ era, it is crucial to implement the algorithm with a fewer resource, and our results to use the KPO will contribute to reduce resource. KPO network may be used as a variant of continuous variable neural network (Killoran et al. 2019). There are many potential applications to use the continuous degrees of freedom of the KPO. We hope that our research will help to expand the range of applications of the KPO.

Appendix A: A derivation of a Hamiltonian of a KPO

We explain a derivation of the effective Hamiltonian of the KPO composed of an array of superconducting quantum interference devices (SQUIDs) (Wang et al. 2019; Gao et al. 2021; Masuda et al. 2022). Let us denote the overall phase across the junction array by $\hat{\varphi}$. Also, we denote $\hat{n}_C$ by a difference in the number of charges on the two plates of the capacitor in units $2e$, which corresponds to the charge of a Cooper pair.
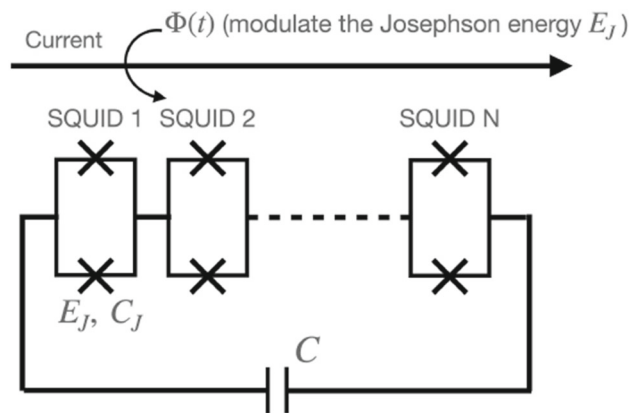

Fig. 8 Circuit model of a KPO with N SQUIDs and a shunt capacitor C

These will obey the following canonical commutation relation.

$$[\hat{\varphi}, \hat{n}_C] = i. \tag{A1}$$

The Hamiltonian of the system is given as

$$\hat{H} = 4E_C \hat{n}_C^2 - N_S E_J[\Phi(t)] \cos \frac{\hat{\varphi}}{N_S}, \tag{A2}$$

where N_S, E_J, and E_C are the number of SQUIDs, the Josephson energy of a SQUID, and the charging energy of the resonator.

By using a time-dependent magnetic flux $\Phi(t)$ penetrating the SQUID loops, we can modulate the Josephson energy as $E_J(t) = E_J + \delta E_J \cos \omega_p t$. For simplicity, we set the phase of the pump field to be zero, $\theta = 0$.

By performing a Taylor expansion and taking into account up to the fourth order of φ/N_S, we approximate Hamiltonian as

$$\hat{H} = 4E_C \hat{n}_C^2 + \frac{E_J}{2N_S} \hat{\varphi}^2 - \frac{E_J}{24N_S^3} \varphi^4$$
$$- N_S \delta E_J \cos \omega_p t \left(1 - \frac{1}{2} \frac{\hat{\varphi}^2}{N_S^2} + \frac{1}{24} \frac{\hat{\varphi}^4}{N_S^4} \right). \tag{A3}$$

Here, we assume $|\langle \varphi/N_S \rangle| \ll 1$.

We can define the creation and annihilation operator as follows.

$$\hat{a} = \frac{1}{2} \left(\frac{E_J}{2N_S E_C} \right)^{\frac{1}{4}} \hat{\varphi} + i \left(\frac{2N_S E_C}{E_J} \right)^{\frac{1}{4}} \hat{n}_C, \tag{A4}$$

$$\hat{a}^\dagger = \frac{1}{2} \left(\frac{E_J}{2N_S E_C} \right)^{\frac{1}{4}} \hat{\varphi} - i \left(\frac{2N_S E_C}{E_J} \right)^{\frac{1}{4}} \hat{n}_C. \tag{A5}$$

Then, the Hamiltonian Eq. A3 becomes

$$\hat{H} = \omega \left(\hat{a}^\dagger \hat{a} + \frac{1}{2} \right) - \frac{K}{12} (\hat{a} + \hat{a}^\dagger)^4$$
$$+ \left[-N_S \delta E_J + p(a + a^\dagger)^2 - \frac{Kp}{3\omega} (a + a^\dagger)^4 \right]$$
$$\times \cos \omega_p t, \tag{A6}$$

where $\omega = \sqrt{8E_C E_J/N_S}$, $K = E_C/N_S^2$, and $p = 2\omega \delta E_J/8E_J$.

We assume $Kp \ll \omega$, and we drop the last term in Eq. A6. Then, we obtain

$$\hat{H} = \omega a^\dagger a - \frac{K}{12}(a + a^\dagger)^4 + p(a + a^\dagger)^2 \cos \omega_p t. \tag{A7}$$

Moving to the rotating frame at the frequency of $\omega_p/2$ and using the rotating wave approximation, we obtain

$$\hat{H} = \Delta a^\dagger a - \frac{K}{12} a^{\dagger 2} a^2 + \frac{p}{2}(a^2 + a^{\dagger 2}), \tag{A8}$$

where $\Delta = \omega - K - \omega_p/2$.

Appendix B: A conventional scheme for our simulation for $d_x = d_y = 1$

Let us review a conventional scheme for quantum circuit learning with qubits (Mitarai et al. 2018). The unitary gate to encode input data $U(x)$ is chosen as

$$\hat{U}(x) = \prod_{j=1}^{M} R_j^Z(\arccos(x^2)) R_j^Y(\arcsin(x)), \tag{B1}$$

where $R_j^Z(\phi)$ is the rotation of the j-th qubit around the z axis with an angle of ϕ, $R_j^Y(\phi)$ is the rotation of the j-th qubit around the y axis with an angle of ϕ, respectively, and M is the number of the qubits.

The parameterized unitary $\hat{V}(\boldsymbol{\theta})$ is composed of two parts. The first part is a single-qubit rotation on the j-th qubit $\hat{U}(\theta_j^i)$, which can be generally decomposed into the following form:

$$\hat{U}(\theta_j^i) = R_j^X(\theta_{j1}^i) R_j^Z(\theta_{j2}^i) R_j^X(\theta_{j3}^i). \tag{B2}$$

This means that the single-qubit rotation gate contains three free parameters. The other part is a unitary operation induced by the following Hamiltonian:

$$\hat{H} = \sum_{j=1}^{M} a_j X_j + \sum_{j=1}^{M} \sum_{k=1}^{j-1} J_{jk} Z_j Z_k, \tag{B3}$$

where the coefficients a_j and J_{jk} are taken randomly from a uniform distribution on $[-1, 1]$. This means that the unitary operator becomes

$$\hat{V}(\boldsymbol{\theta}) = \prod_{i=1}^{D} e^{-i\tau\hat{H}} \prod_{j=1}^{K} \hat{U}(\theta_j^i). \tag{B4}$$

We use $\tau = 10$ in our simulation, which is the same as that used in Mitarai et al. (2018).

Appendix C: A method for supervised machine learning with $d_x > 1$ by using a single KPO

We could use the single KPO with $d_x > 1$ and $d_y > 1$. For $d_x > 1$, we need to encode the variable into the initial state. If we encode a different data, different outputs should be generated. So, in this case, $|\psi(x_1, ...x_{d_x})\rangle$ should be different from $|\psi(x'_1, ...x'_{d_x})\rangle$ unless $x_j = x'_j$ for $j = 1, 2, \cdots, d_x$ is satisfied. For $d_y > 1$, we have to measure independent observables $\hat{M}_1$, $\hat{M}_2$,..., $\hat{M}_{d_y}$ at the step 3.

Let us explain such an example. Firstly, we prepare the coherent state $|\alpha\rangle$ where we set $\alpha = r = \sqrt{x_1^2 + x_2^2}$. Second, we perform a unitary operation $e^{-i\tilde{\chi}\hat{n}^2 - i\varphi\hat{n}}$, where φ is defined by

$$\varphi = \begin{cases} \arccos \frac{x_1}{r}, & (x_2 > 0) \\ -\arccos \frac{x_1}{r}, & (x_2 \le 0) \\ 0, & (r = 0). \end{cases} \tag{C1}$$

It is notable that φ is defined on the interval $[-\pi, \pi)$. We obtain

$$|\psi(x_1, x_2)\rangle = e^{-i\tilde{\chi}\hat{n}^2 - i\varphi\hat{n}}|r\rangle. \tag{C2}$$

The overlap $\langle\psi(x'_1, x'_2)|\psi(x_1, x_2)\rangle$ is calculated as follows:

$$\begin{aligned}
&\langle\psi(x'_1, x'_2)|\psi(x_1, x_2)\rangle \\
&= \langle r'|e^{i\tilde{\chi}\hat{n}^2 + i\varphi'\hat{n}} e^{-i\tilde{\chi}\hat{n}^2 - i\varphi\hat{n}}|r\rangle \\
&= \langle r'|e^{i(\varphi'-\varphi)\hat{n}}|r\rangle \\
&= \langle r'|re^{i(\varphi'-\varphi)}\rangle \\
&= e^{-\frac{1}{2}(|r'|^2 + |r|^2 - 2r're^{i(\varphi'-\varphi)})}.
\end{aligned} \tag{C3}$$

If the overlap Eq. C3 is 1, the exponent of the overlap is zero, and we obtain

$$r'^2 + r^2 - 2r're^{i(\varphi'-\varphi)} = 0. \tag{C4}$$

We obtain the solution as follows.

$$\varphi' = \varphi, \tag{C5}$$
$$r' = r. \tag{C6}$$

Therefore, the overlap is not unity unless $x_1 = x'_1$ and $x_2 = x'_2$ are satisfied. Also, as the observable, we can adopt $\hat{M}_1 = \hat{a} + \hat{a}^\dagger$ and $\hat{M}_2 = \hat{a}^\dagger\hat{a}$. So we can use this initial state for our method with $d_x = 2$ and $d_y = 2$.

Declarations

References

Aoki T, Kanao T, Goto H, Kawabata S, Masuda S (2023) Control of the ZZ coupling between Kerr-cat qubits via transmon couplers. Phys Rev Applied 21:014030. https://journals.aps.org/prapplied/abstract/10.1103/PhysRevApplied.21.014030

Armaos V, Badounas DA, Deligiannis P, Lianos K (2020) Computational chemistry on quantum computers. Appl Phys A 126:625

Aspuru-Guzik A, Dutoi AD, Love PJ, Head-Gordon M (2005) Simulated quantum computation of molecular energies. Science 309:1704

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

Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S (2017) Quantum machine learning. Nature 549:212–219

Bourassa J, Beaudoin F, Gambetta JM, Blais A (2012) Josephson-junction-embedded transmission-line resonators: from Kerr medium to in-line transmon. Phys Rev A 86:013814

Cao Y, Romero J, Olson JP, Degroote M, Johnson PD, Kieferová M, Kivlichan ID, Menke T, Peropadre B, Sawaya NPD, Sim S, Veis L, Aspuru-Guzik A (2019) Quantum chemistry in the age of quantum computing. Chem Rev 119:10856. pMID:31469277

Cochrane PT, Milburn GJ, Munro WJ (1999) Macroscopically distinct quantum-superposition states as a bosonic code for amplitude damping. Phys Rev A 59:2631

Devitt SJ, Stephens AM, Munro WJ, Nemoto K (2013) Requirements for fault-tolerant factoring on an atom-optics quantum computer. Nat Commun 4:2524

Endo S, Cai Z, Benjamin SC, Yuan X (2021) Hybrid quantum-classical algorithms and quantum error mitigation. J Phys Soc Jpn 90:032001

Gan BY, Leykam D, Angelakis DG (2022) Fock stateenhanced expressivity of quantum machine learning models. EPJ Quantum Techno 9:16

Gao YY, Rol MA, Touzard S, Wang C (2021) Practical guide for building superconducting quantum devices. PRX Quantum 2:040202

Gidney C, Ekerå M (2021) How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. Quantum 5:433

Gil Vidal FJ, Theis DO (2020) Input redundancy for parameterized quantum circuits. Front Phys 8:297

Goto H (2016a) Universal quantum computation with a nonlinear oscillator network. Phys Rev A 93:050301

Goto H (2016b) Bifurcation-based adiabatic quantum computation with a nonlinear oscillator network. Sci Rep 6:21686

Goto H (2019) Quantum computation based on quantum adiabatic bifurcations of Kerr-nonlinear parametric oscillators. J Phys Soc Jpn 88:061015

Grimm A, Frattini NE, Puri S, Mundhada SO, Touzard S, Mirrahimi M, Girvin SM, Shankar S, Devoret MH (2020) Stabilization and operation of a Kerr-cat qubit. Nature 584:205

Grover, LK (1996) A fast quantum mechanical algorithm for database search, in proceedings of the twenty-eighth annual acm symposium on theory of computing. STOC '96 Association for Computing Machinery, New York, USA, p 212–219

Havlíček V, Córcoles AD, Temme K, Harrow AW, Kandala A, Chow JM, Gambetta JM (2019) Supervised learning with quantum-enhanced feature spaces. Nature 567:209

Johansson JR, Nation P, Nori F (2013) QuTiP 2: a python framework for the dynamics of open quantum systems. Comput Phys Commun 184:1234

Jones NC, Van Meter R, Fowler AG, McMahon PL, Kim J, Ladd TD, Yamamoto Y (2012) Layered architecture for quantum computing. Phys Rev X 2:031007

Killoran N, Bromley TR, Arrazola JM, Schuld M, Quesada N, Lloyd S (2019) Continuous-variable quantum neural networks. Phys Rev Res 1:033063

Lenstra AK (2000) Integer Factoring. Des Codes Crypt 19:101

Liu J, Zhong C, Otten M, Chandra A, Cortes CL, Ti C, Gray SK, Han X (2023) Quantum Kerr learning. Mach Learn Sci Technol 4:025003

Masuda S, Kanao T, Goto H, Matsuzaki Y, Ishikawa T, Kawabata S (2022) Fast tunable coupling scheme of Kerr parametric oscillators based on shortcuts to adiabaticity. Phys Rev Appl 18:034076

McArdle S, Endo S, Aspuru-Guzik A, Benjamin SC, Yuan X (2020) Quantum computational chemistry. Rev Mod Phys Rev Mod Phys 92:015003

Meaney CH, Nha H, Duty T, Milburn GJ (2014) Quantum and classical nonlinear dynamics in a microwave cavity. EPJ Quantum Technol 1:7

Milburn G, Holmes C (1991) Quantum coherence and classical chaos in a pulsed parametric oscillator with a Kerr nonlinearity. Phys Rev A 44:4704

Mitarai K, Negoro M, Kitagawa M, Fujii K (2018) Quantum circuit learning. Phys Rev A 98:032309

Nelder JA, Mead R (1965) A simplex method for function minimization. J Comput 7:308

Pérez-Salinas A, Cervera-Lierta A, Gil-Fuster E, Latorre JI (2020) Data re-uploading for a universal quantum classifier. Quantum 4:226

Preskill J (2018) Quantum computing in the NISQ era and beyond. Quantum 2:79

Puri S, Boutin S, Blais A (2017) Engineering the quantum states of light in a Kerr-nonlinear resonator by twophoton driving. Npj Quantum Inf 3:18

Puri S, Andersen CK, Grimsmo AL, Blais A (2017) Quantum annealing with all-to-all connected nonlinear oscillators. Nat Commun 8:15785

Schuld M, Killoran N (2022) Is quantum advantage the right goal for quantum machine learning? PRX Quantum 3:030101

Schuld M, Sinayskiy I, Petruccione F (2015) An introduction to quantum machine learning. Contemp Phys 56:172

Schuld M, Sweke R, Meyer JJ (2021) Effect of data encoding on the expressive power of variational quantummachine- learning models. Phys Rev A 103:032430

Shor P. (1994) Algorithms for quantum computation: discrete logarithms and factoring. In: Proceedings 35th annual symposium on foundations of computer science, pp 124–134

Shor PW (1997) Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J Comput 26:1484

Stehlik J, Zajac D, Underwood D, Phung T, Blair J, Carnevale S, Klaus D, Keefe G, Carniol A, Kumph M et al (2021) Tunable coupling architecture for fixed-frequency transmon superconducting qubits. Phys Rev Lett 127:080505

Steinbrecher GR, Olson JP, Englund D, Carolan J (2019) Quantum optical neural networks. Npj Quantum Information 5:60

Suzuki Y, Kawase Y, Masumura Y, Hiraga Y, Nakadai M, Chen J, Nakanishi KM, Mitarai K, Imai R, Tamiya S, Yamamoto T, Yan T, Kawakubo T, Nakagawa YO, Ibe Y, Zhang Y, Yamashita H, Yoshimura H, Hayashi A, Fujii K (2021) Qulacs: a fast and versatile quantum circuit simulator for research purpose. Quantum 5:559

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski D, Peterson P, Weckesser W, Bright W, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P (2020) Fundamental algorithms for scientific computing in python and SciPy 10. contributors SciPy 10. Nat Methods 17:261

Volkoff TJ (2021) Efficient trainability of linear optical modules in quantum optical neural networks. J Russ Laser Res 42:250

Wang Z, Pechal M, Wollack EA, Arrangoiz-Arriola P, Gao M, Lee NR, Safavi-Naeini AH (2019) Quantum dynamics of a few-photon parametric oscillator. Phys Rev X 9:021049

Wielinga B, Milburn G (1993) Quantum tunneling in a Kerr medium with parametric pumping. Phys Rev A 48:2494

Wierichs D, Izaac J, Wang C, Lin CY-Y (2022) General parameter-shift rules for quantum gradients. Quantum 6:677

Yamaji T, Kagami S, Yamaguchi A, Satoh T, Koshino K, Goto H, Lin Z, Nakamura Y, Yamamoto T (2022) Spectroscopic observation of the crossover from a classical Duffing oscillator to a Kerr parametric oscillator. Phys Rev A 105:023519