Quantum Kerr learning

To cite this article: Junyu Liu *et al* 2023 *Mach. Learn.: Sci. Technol.* **4** 025003

View the article online for updates and enhancements.

You may also like

MACHINE LEARNING
Science and Technology

PAPER

OPEN ACCESS

Quantum Kerr learning

Junyu Liu[1,2,3,4,*] , Changchun Zhong[1], Matthew Otten[5], Anirban Chandra[6,7], Cristian L Cortes[6], Chaoyang Ti[6] , Stephen K Gray[6] and Xu Han[6,*]

1 Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, United States of America
2 Chicago Quantum Exchange, Chicago, IL 60637, United States of America
3 Kadanoff Center for Theoretical Physics, The University of Chicago, Chicago, IL 60637, United States of America
4 qBraid Co., Harper Court 5235, Chicago, IL 60615, United States of America
5 HRL Laboratories, LLC, Malibu, CA 90265, United States of America
6 Center for Nanoscale Materials, Argonne National Laboratory, Lemont, IL 60439, United States of America
7 University of Illinois, Chicago, IL 60607, United States of America
* Authors to whom any correspondence should be addressed.

E-mail: junyuliu@uchicago.edu and xu.han@anl.gov

Abstract

Quantum machine learning is a rapidly evolving field of research that could facilitate important applications for quantum computing and also significantly impact data-driven sciences. In our work, based on various arguments from complexity theory and physics, we demonstrate that a single Kerr mode can provide some 'quantum enhancements' when dealing with kernel-based methods. Using kernel properties, neural tangent kernel theory, first-order perturbation theory of the Kerr non-linearity, and non-perturbative numerical simulations, we show that quantum enhancements could happen in terms of convergence time and generalization error. Furthermore, we make explicit indications on how higher-dimensional input data could be considered. Finally, we propose an experimental protocol, that we call *quantum Kerr learning*, based on circuit QED.

1. Introduction

Quantum machine learning, i.e. combining machine learning with the computational power of quantum devices, is an exciting and emerging direction for modern information technology [1–4]. In the present noisy intermediate-scale quantum era [5], there has been significant progress on quantum machine learning based on variational quantum circuits [6–15]. However, it is still not completely clear, both in theory and practice, if and how a true quantum advantage relative to completely classical-computer based approaches will be achieved [15, 16].

Over several years in classical machine learning, kernel methods have been developed and applied to address numerous problems [17]. A non-linear machine learning problem could ideally be described by linear models in a sufficiently high dimensional *Hilbert space* using these kernel methods. In the context of classical machine learning, the corresponding Hilbert space of a given kernel is abstract, whereas in the quantum computing setup, the Hilbert space could be *physical*—such as the state vector space of a quantum device [18]. Complex quantum systems could produce complicated enough kernels that are hard to evaluate using classical computers, providing a potential regime of quantum advantage in quantum machine learning [3].

Now, how can we produce complicated kernels in a quantum system? In the quantum kernel method, the kernel is evaluated from a quantum measurement. For instance, if we consider a Hamiltonian with the real time evolution, $H(t)$, one could construct the kernel as

$$K(\mathbf{x}, \mathbf{x}') = \left| \left\langle \psi \left| \bar{\mathbb{T}} \exp\left[\frac{i}{\hbar} \int_0^{T'} H(t') \mathrm{d}t' \right] \times \mathbb{T} \exp\left[-\frac{i}{\hbar} \int_0^T H(t) \mathrm{d}t \right] \right| \psi \right\rangle \right|^2 . \quad (1)$$

Here, starting from $|\psi\rangle$ and evolving with the Hamiltonian $H(t)$ during over times T and T' respectively, the inner product of two quantum states is measured. $\mathbb{T}$ denotes time ordering in quantum mechanics. The more complicated the Hamiltonian is, the harder it is for a classical computer to simulate. In equation (1), the kernel K is a matrix where the matrix elements are specified by the input data points, $\mathbf{x}$ and $\mathbf{x}'$. The vectors $\mathbf{x}$ and $\mathbf{x}'$ could be constructed from parameters of the Hamiltonian or the evolution time.

2. Model setup

In this paper, we consider a single-mode quantum system with Kerr non-linearity that could potentially provide computational benefits compared to its classical counterparts. Such systems can be easily realized in circuit quantum electrodynamics (QED) experiments where the Kerr non-linearity can be obtained from Josephson junctions or the kinetic inductance of a superconducting resonator [19–21]. We propose and describe an experimental realization in the supplemental material (SM). The effective Hamiltonian of the considered system can be described as,

$$H = H_{\mathrm{f}} + H_{\mathrm{I}} \, , H_{\mathrm{f}} = H_0 + H_{\mathrm{t}} \frac{H_0}{\hbar} = \omega_{\mathrm{m}} \left(b^\dagger b + \frac{1}{2} \right) \, , \frac{H_{\mathrm{t}}}{\hbar} = \Omega \left(b e^{i\omega_{\mathrm{L}} t} + b^\dagger e^{-i\omega_{\mathrm{L}} t} \right) \, , \frac{H_{\mathrm{I}}}{\hbar} = -K_{\mathrm{err}} b^\dagger b^\dagger b b \, . \quad (2)$$

For our demonstrative examples, we restrict ourselves to three-dimensional inputs and treat ω_{m} and K_{err} as constants. The input data is thus defined as a real vector $\mathbf{x} = (\mathbf{x})_{i=1}^3 = (\Omega, \omega_{\mathrm{L}}, T)$. Ω and ω_{L} standard for the energy and the frequency of the laser, and T means the evolution time. All of them are tunable in the experimental devices.

Depending on the magnitude of K_{err} a few regimes exist that warrant some discussion. When the Kerr non-linearity is turned off, the system Hamiltonian reduces to a quadratic form in the rotating frame (The precise procedure of the rotating frames is summarized in the supplementary materials, where we transform the original Hamiltonian into the rotating basis, and the Hamiltonian only has terms as $b^\dagger ger$ and b, the quadratic Hamiltonian, which could be equivalently represented in q (location) and p (momentum).). If we turn on the Kerr non-linearity and when K_{err} is much smaller than Ω, the kernel could be predicted by perturbation theory (see SM for detailed calculations). On the other hand, when K_{err} is much larger than Ω, the non-linear resonator can be approximated as a qubit with negligible excitations at energy levels higher than the first excited state (see SM). However, for a general K_{err} in the regime of $K_{\mathrm{err}}/\Omega \sim \mathcal{O}(1)$, we do not have analytic predictions and the system must be evaluated on a quantum system or simulated using a classical computer.
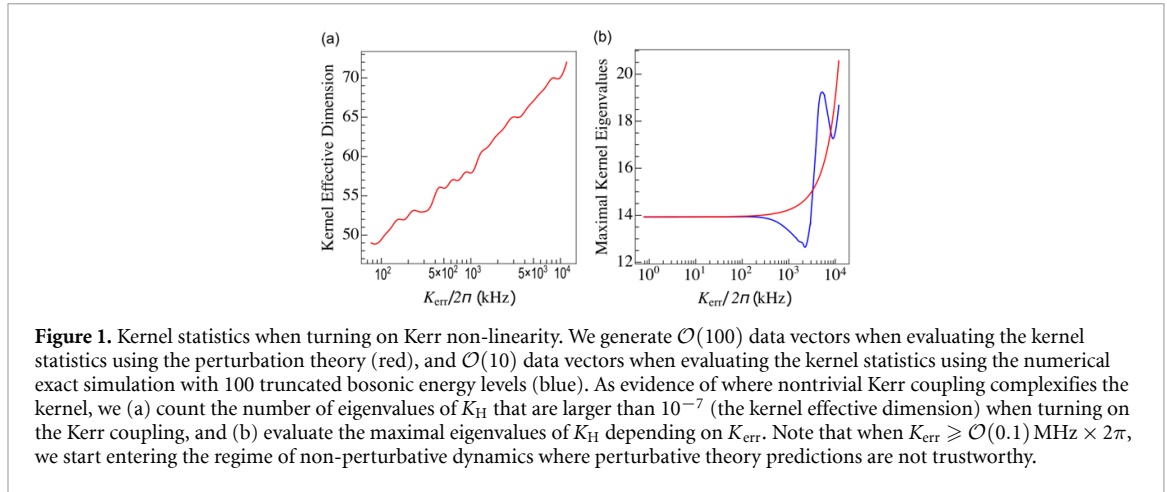
3. Quantum enhancement

The Kerr non-linear quantum system considered in this work has some inherent advantages that imply quantum enhancement. Here, we use the word 'quantum enhancement' to indicate benefits in computation provided by the Kerr non-linearity; this is not the same as 'quantum advantage' determined by the rigorous complexity theory statements and practical demonstrations. However, a few complexity-related arguments are presented in SM. A few possible inherent advantages of our Kerr system are discussed next.

First, although in the bulk of our paper, we treat Ω as a constant, we could make it time-dependent. For the time-dependent $\Omega(t)$, it is shown that the Hamiltonian in equation (2) could simulate universal quantum computing in polynomial time, and the whole class **BQP** [22], even for small Kerr couplings. Thus, we expect that a combination of time-dependent $\Omega(t)$ and K_{err} could provide further quantum enhancements.

Second, although the bulk of our paper is about single modes, we briefly introduce and discuss multimodes. A multimode Hamiltonian in our circuit QED setup will include several free bosonic modes that are interacting with another Kerr mode. Through the coupling, those free modes will have more complicated dynamics, and we expect that it might present stronger quantum enhancements. An analog of the multiple mode system will be a zero-dimensional $\lambda\phi^4$ quantum field theory on the lattice with the system size N (note that the Kerr term is also quartic), where we have the theoretical predictions when the bare coupling λ is either zero or infinity. The non-perturbative corrections will happen around the critical point, and the strong coupling regime, where we have a $\mathbb{Z}_2$ symmetry which is spontaneously broken, and we expect a quantum speedup to simulate the non-perturbative dynamics [23–25].

4. Kernel properties and statistics

The property of the kernel function K, or more precisely the neural tangent kernel (NTK) $K_{\mathrm{H}} = K^2$ of the linear model (see SM), plays an important role in the performance of the machine learning algorithm. To discuss and demonstrate the quantum enhancement achieved by Kerr non-linearity, we first create an

Figure 1. Kernel statistics when turning on Kerr non-linearity. We generate $\mathcal{O}(100)$ data vectors when evaluating the kernel statistics using the perturbation theory (red), and $\mathcal{O}(10)$ data vectors when evaluating the kernel statistics using the numerical exact simulation with 100 truncated bosonic energy levels (blue). As evidence of where nontrivial Kerr coupling complexifies the kernel, we (a) count the number of eigenvalues of K_H that are larger than 10^{-7} (the kernel effective dimension) when turning on the Kerr coupling, and (b) evaluate the maximal eigenvalues of K_H depending on K_{err}. Note that when $K_{err} \geqslant \mathcal{O}(0.1)\,\text{MHz} \times 2\pi$, we start entering the regime of non-perturbative dynamics where perturbative theory predictions are not trustworthy.

arbitrary kernel matrix by specifying the associated parameters – Ω, ω_L, T, ω_m, K_{err}. We randomly generate the input data $\mathbf{x} = (\Omega, \omega_L, T)$ through uniform distributions within experimentally feasible data range. We set the data range from $(0, 0, 0)$ to $\left(\Omega^{range}, \omega_L^{range}, T^{range}\right) \sim (300\,\text{MHz} \times 2\pi, 10\,\text{GHz} \times 2\pi, 0.05\,\mu\text{s})$. Moreover, we fix ω_m at around $10\,\text{GHz} \times 2\pi$ which is typical for a superconducting microwave resonator. K_{err} is treated as a control parameter to understand potential quantum enhancement.
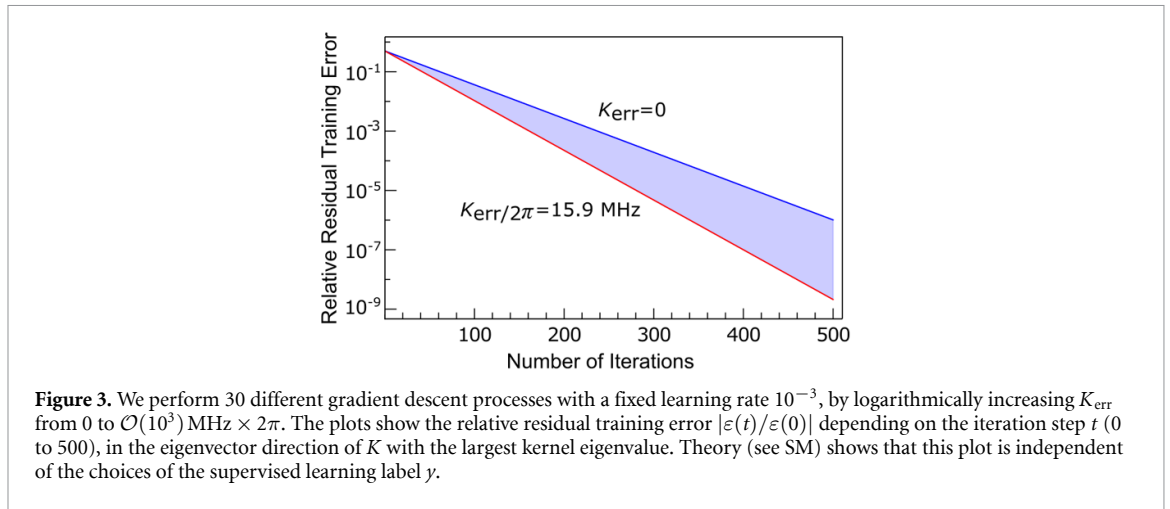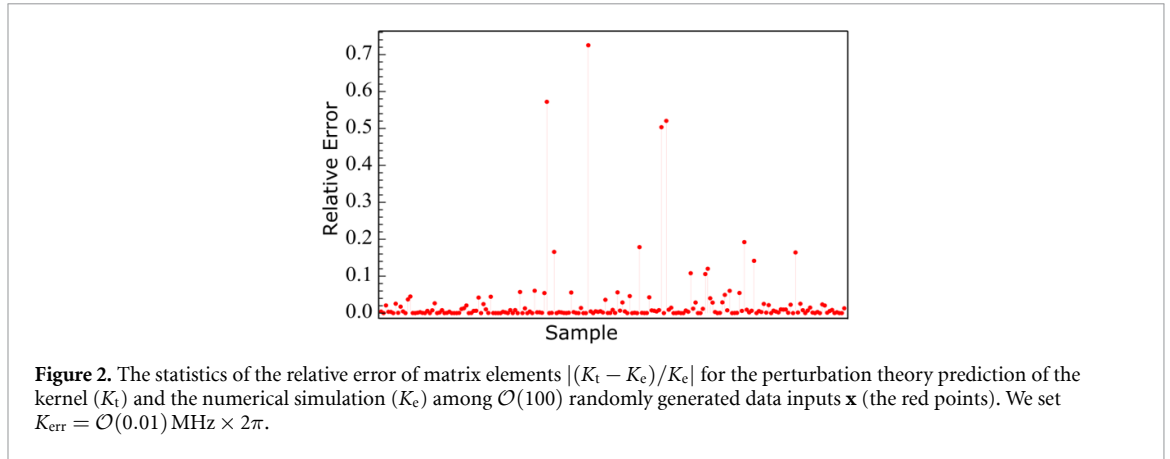
We assess the performance of our proposed K_{err} kernel based on the theory of the kernel method [17] and the NTK theory for linear models [15, 26–33]. Note that, we use the traditional quantum kernel methods to solve the problem that is similar to the support vector machine with continuous variables, but we use NTK theory as a tool to make predictions about the gradient descent dynamics in linear regression. NTK theory is a theory for explaining wide neural networks, which could be deep but the ratio of depth to width should be small. In the large-width limit, the gradient descent dynamics of the neural networks will coincide with the kernel method, so it is also called the kernel limit. In our work, we directly use a linear model (namely, the kernel method), so the square of the NTK is equal to the traditional quantum kernel. Moreover, better kernels will have flatter eigenspectra with more non-trivial kernel eigenvalues, which will lead to faster convergence speed [33] (see discussions in SM) and less generalization error for good enough alignments [34–38].

In figure 1, we study the change of the kernel eigenspectra when turning on the Kerr coupling. We measure the complexity of the kernel by looking at the kernel effective dimension (the number of eigenvalues that are not small, where we set the criterion to be $>10^{-7}$). Since the kernel is normalized by the diagonal matrix element to be no larger than 1, we define the criterion to be 10^{-7} with respect to the norm of the kernel, which is also the numerical accuracy of our calculations. As shown in figure 1, we find that when we turn on the Kerr coupling toward the non-perturbative regime, the kernel spectra have the tendency to be flatter, inferred from the increasing kernel effective dimension. This suggests that better performance can be achieved in generic numerical optimization experiments for larger Kerr non-linearity. In fact, a flatter spectrum means that the distribution has longer tails toward the higher values of the NTK eigenvalue parameters, which might indicate faster convergence and less generalization error for sufficient kernel alignments, the inner product between kernel predictions and the target functions [34–38].

Analyses of kernel properties suggest that a non-zero K_{err} might generically lead to better performance in supervised learning tasks. Our detailed perturbation analysis for leading order of K_{err}, shown in supplementary materials, is well-tested through numerics and provides reasonable predictions in the perturbative regime of K_{err}. The perturbation analysis corroborates our findings from numerical simulations—analytic formulas obtained from the analysis match with numerical simulations of our bosonic quantum systems up to a given truncation. In figure 2, we verify the validity of our perturbation theory prediction by comparing it with the numerical simulation for truncated Hilbert space dimensions. K_t and K_e represent the value of matrix elements obtained from theory and simulations respectively. The relative error is negligible in most cases.

5. Supervised learning and gradient descent

We next focus on an actual machine learning task to discuss and demonstrate the quantum enhancement achieved by Kerr non-linearity. Consider a simple support vector machine model where $(\mathbf{x}_{\tilde{\alpha}}, y_{\tilde{\alpha}})$ represent the corresponding input-output pairs. $\tilde{\alpha}$ denotes the index of data from the training set $\mathcal{A}$. In a supervised

Figure 2. The statistics of the relative error of matrix elements $|(K_t - K_e)/K_e|$ for the perturbation theory prediction of the kernel (K_t) and the numerical simulation (K_e) among $\mathcal{O}(100)$ randomly generated data inputs $\mathbf{x}$ (the red points). We set $K_{err} = \mathcal{O}(0.01)\,\mathrm{MHz} \times 2\pi$.

Figure 3. We perform 30 different gradient descent processes with a fixed learning rate 10^{-3}, by logarithmically increasing K_{err} from 0 to $\mathcal{O}(10^3)\,\mathrm{MHz} \times 2\pi$. The plots show the relative residual training error $|\varepsilon(t)/\varepsilon(0)|$ depending on the iteration step t (0 to 500), in the eigenvector direction of K with the largest kernel eigenvalue. Theory (see SM) shows that this plot is independent of the choices of the supervised learning label y.

learning setting, while training, a predictive model is developed using $(\mathbf{x}_{\tilde{\alpha}}, y_{\tilde{\alpha}})$. Subsequently, predictions (z_δ) can be obtained using this model:

$$z_\delta = z(\mathbf{x}_\delta) = \sum_{\tilde{\alpha} \in \mathcal{A}} \theta_{\tilde{\alpha}} K(\mathbf{x}_{\tilde{\alpha}}, \mathbf{x}_\delta) , \tag{3}$$
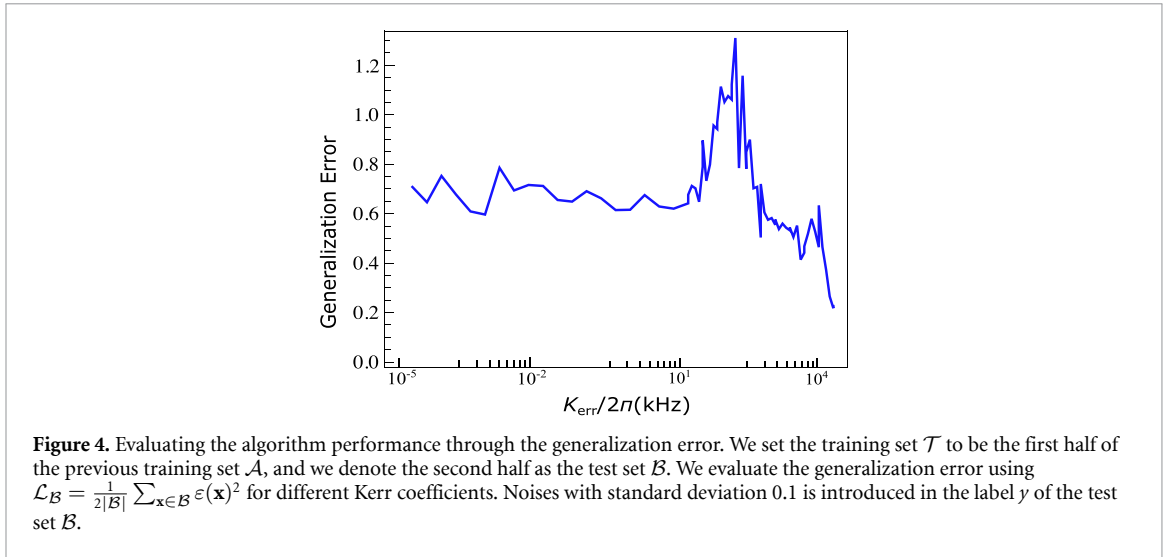
where K is the kernel, $\mathbf{x}_\delta$ is a general input and δ denotes the index of data from the whole input data set $\mathcal{D}$ (so we have $\mathcal{A} \subset \mathcal{D}$), and $\theta_{\tilde{\alpha}}$ is a trainable variable. Ideally, in a quantum kernel method [18], K is evaluated using quantum measurements, but here we will utilize numerical simulations and theoretical arguments to demonstrate quantum enhancement achieved by the Kerr kernel.

For our randomly generated $\mathbf{x}$, we assign them with one-dimensional outputs $y(\mathbf{x})$. The residual training error $\varepsilon(\mathbf{x}) = z(\mathbf{x}) - y(\mathbf{x})$, where $z(\mathbf{x})$ is the kernel method prediction, enters in the formula of the mean-square loss function $\mathcal{L} = \frac{1}{2} \sum_{\mathbf{x}} \varepsilon^2(\mathbf{x})$. We use the gradient descent algorithm to minimize $\mathcal{L}$. In the kernel method, the optimization process is exactly solvable, and the decay of the residual training error is exponential when the learning rate is small, and the decay rate can be predicted by kernel eigenvalues (see SM).

In figure 3, we perform different gradient descent processes for increasing K_{err}. We find that the increasing Kerr non-linearity will generically accelerate the gradient descent dynamics, consistent with our findings of kernel eigenspectra (see SM).

6. Generalization error

Another important metric for quantifying the performance of machine learning models is the generalization error. As an example, we split half of the data as the training set, and we evaluate the loss function on the remaining half as the test set after we train the model. We consider learning a function $y(\mathbf{x}) = \sum_{i=1}^{3} \sin^2(\mathbf{x}_i^2)$ instead, where our time T is rescaled to be $\mathcal{O}(1)$ and other components in T are set by dimensional analysis accordingly. Noise with the standard deviation 0.1 is introduced in all the instances of the task labels in $\mathcal{B}$. In figure 4, we find that generalization error has non-trivial behavior for growing Kerr non-linearity, and for

Figure 4. Evaluating the algorithm performance through the generalization error. We set the training set $\mathcal{T}$ to be the first half of the previous training set $\mathcal{A}$, and we denote the second half as the test set $\mathcal{B}$. We evaluate the generalization error using $\mathcal{L}_{\mathcal{B}} = \frac{1}{2|\mathcal{B}|} \sum_{\mathbf{x} \in \mathcal{B}} \varepsilon(\mathbf{x})^2$ for different Kerr coefficients. Noises with standard deviation 0.1 is introduced in the label y of the test set $\mathcal{B}$.

large Kerr values in the non-perturbative regime, it decays significantly. According to references [34–38], generalization errors could be related to NTK eigenvalues, and the kernel method is generalized well from good enough alignments. Thus, our finding provides good evidence that quantum Kerr learning will provide extra enhancement in algorithm performances.

Moreover, it might be worth noticing that other studies [39, 40] show related quantitative methods for predicting the generalization error for quantum machine learning methods. In [39], it is shown that higher entropies in the kernel spectra might lead to worse generalizations in quantum machine learning. Our results are indeed consistent with such a theory, since the distribution of the kernel eigenvalues will lead to better generalization errors only when one has a good alignment between the target and the data. The kernel alignment [41] is an inner product between the kernel in our assumption and the target function we want to fit and higher alignment will indicate better performance in generalization [34, 41]. Intuitively, one could understand it as the following. If we have bad alignments between the target function we want to fit and the data, then the loss function landscape will be sharp for large kernel eigenvalues (especially when we consider the mean square error, and in this case, the (non-zero) NTK eigenspectra are the same as the (non-zero) Hessian spectra, as shown by the singular value decomposition). Otherwise, flatter higher eigenvalues will lead to better generalizations, since there is more space to move around the local minima. Perhaps the condition among NTKs, alignments, and generalization errors could be refined by entropies, which is closely related to a potential information bottleneck theory in quantum machine learning [42]. Finally, we emphasize that in our simulation, we do not consider quantum noises during sampling. However, one can still speak about the noise due to the randomness of the sampling from the input data distribution. More details are presented in the SM.

7. Higher dimensions

Regression in higher dimensions can be performed by using two quantum systems instead of one, each encoding a subset of the data,

$$\mathcal{K} = \left| \left\langle \psi \otimes \phi \left| \bar{\mathbb{T}} \exp\left[\frac{i}{\hbar} \int_0^{T'} H(t') dt' \right] \times \mathbb{T} \exp\left[-\frac{i}{\hbar} \int_0^{T} H(t) dt \right] \right| \psi \otimes \phi \right\rangle \right|^2, \tag{4}$$

where we have assumed that we initially start in a product state. $H(t)$ could, again, be a general Hamiltonian. If $H(t)$ includes terms that couple the two subsystems, $|\psi\rangle$ and $|\phi\rangle$, this could potentially create entanglement between the two systems, increasing the complexity and making it harder for a classical computer to compute. It is also interesting to look at the case when $H(t)$ does not couple the two subsystems. For example, the Hamiltonian could represent two uncoupled Kerr resonators as

$$H(t) = H \otimes I + I \otimes H, \tag{5}$$

where H is the Hamiltonian described in equation (2) and I is identity. In this initially unentangled and uncoupled example, our kernel reduces to a Hadamard product of the two uncoupled kernels,

$$\mathcal{K} = K_\alpha \circ K_\beta, \tag{6}$$

where K_α (K_β) is the single-system Kerr kernel discussed in previous sections, and the subscript (α, β) represents two independent subsystems. Because kernels are symmetric, positive definite matrices, we can bound the properties of the eigenvalues of $\mathcal{K}$ using the spectra of K. For example, we have the following bound on the spectral radius, ρ, of the product kernel [43]

$$\rho(\mathcal{K}) \leqslant \rho(K_\alpha)\rho(K_\beta). \tag{7}$$

This bound is known to be rather loose, and tighter bounds have been derived [43]. As shown in figure 1, the maximum kernel eigenvalues of the single system grow with increasing K_{err}. For multiple unentangled and uncoupled systems, we can expect that the maximal eigenvalues of this product kernel grow faster, leading to increased performance in higher-dimensional data sets compared with the non-Kerr kernel. This can be extended to the product of many kernels, allowing for the learning of arbitrary dimensional data with collections of single oscillators. It is likely that the addition of entanglement, through a Hamiltonian term that couples the various systems, will further increase the performance [11].

8. Conclusion and outlooks

Our paper opens up a novel direction by exploring the potential of quantum machine learning through circuit QED devices with non-trivial Kerr non-linearity. We find theoretical and numerical evidence where non-trivial Kerr coupling could significantly enhance the performance of the quantum kernel method, based on solid evaluations of kernel statistics and numerical optimization experiments. We strengthen our claims by applying the NTK theory in machine learning, arguments from the theory of quantum complexity, and generalizations toward higher dimensions. Here, we suggest the following directions for future research.

(a) Experimental implementation. It will be interesting to implement the proposal in our paper in the laboratory. When working with actual hardware, measurement noise and errors need to be considered. A combined design between theory and experiments could be obtained based on the theory of the kernel method, including trade-offs among the learning rate, the experimental precision for estimating the residual training error ε, and the number of gradient descent steps we could perform in the laboratory. In addition, the effect of dissipation and decoherence of the non-linear mode deserves future investigation. In experiment, the evolution time of the quantum state should be controlled to be much shorter than the coherence time of the Kerr mode in order to maintain good quantum coherence.

(b) Extensions to multiple modes. One could generalize the above work to multiple modes [21, 44], where we expect the coupling between different bosonic modes will strengthen the complexity of the kernel. It will be interesting to see if our approach will realize the power of multimode devices toward hard problems in quantum machine learning.

(c) Theoretical considerations. It will be interesting to explore further the complexity foundations of our claims. The argument about the Kerr non-linearity and complexity, although promising, is not proven rigorously. A more solid statement about complexity might deepen our understanding of the algorithmic potential of the Kerr non-linearity.

Data availability statement

The data cannot be made publicly available upon publication because the cost of preparing, depositing and hosting the data would be prohibitive within the terms of this research project. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ORCID iDs

Junyu Liu ● https://orcid.org/0000-0003-1669-8039
Chaoyang Ti ● https://orcid.org/0000-0001-6590-9133

References

[1] Huang H Y, Broughton M, Mohseni M, Babbush R, Boixo S, Neven H and McClean J R 2020 arXiv:2011.01938
[2] Abbas A, Sutter D, Zoufal C, Lucchi A, Figalli A and Woerner S 2021 *Nat. Comput. Sci.* **1** 403–9
[3] Liu Y, Arunachalam S and Temme K 2021 *Nat. Phys.* **17** 1–5
[4] Aharonov D, Cotler J and Qi X L 2022 *Nat. Commun.* **13** 1–9
[5] Preskill J 2018 *Quantum* **2** 79
[6] Schuld M, Sinayskiy I and Petruccione F 2015 *Contemp. Phys.* **56** 172–85
[7] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2017 *Nature* **549** 195–202
[8] Dunjko V and Briegel H J 2018 *Rep. Prog. Phys.* **81** 074001
[9] Farhi E and Neven H 2018 arXiv:1802.06002
[10] Cong I, Choi S and Lukin M D 2019 *Nat. Phys.* **15** 1273–8
[11] Otten M, Goumiri I R, Priest B W, Chapline G F and Schneider M D 2020 arXiv:2004.11280
[12] Bausch J 2020 *Advances in Neural Information Processing Systems* vol 33 pp 1368–79
[13] Beer K, Bondarenko D, Farrelly T, Osborne T J, Salzmann R, Scheiermann D and Wolf R 2020 *Nat. Commun.* **11** 1–6
[14] Mangini S, Tacchino F, Gerace D, Bajoni D and Macchiavello C 2021 *Europhys. Lett.* **134** 10002
[15] Liu J, Tacchino F, Glick J R, Jiang L and Mezzacapo A 2021 arXiv:2111.04225
[16] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 *Nat. Commun.* **9** 1–6
[17] Mohri M, Rostamizadeh A and Talwalkar A 2018 *Foundations of Machine Learning* (Cambridge, MA: MIT Press)
[18] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 *Nature* **567** 209–12
[19] Xu M, Han X, Fu W, Zou C L, Devoret M H and Tang H X 2019 *Appl. Phys. Lett.* **114** 192601
[20] Han X *et al* 2020 *Nat. Commun.* **11** 3237
[21] Han X, Zou C L, Fu W, Xu M, Xu Y and Tang H X 2022 *Phys. Rev. Lett.* **129** 107701
[22] Lloyd S and Braunstein S L 1999 Quantum computation over continuous variables *Quantum Information With Continuous Variables* (Berlin: Springer) pp 9–17
[23] Jordan S P, Lee K S M and Preskill J 2012 *Science* **336** 1130–3
[24] Jordan S P, Lee K S M and Preskill J 2014 *Quantum Inf. Comput.* **14** 1014–80
[25] Preskill J 2018 *PoS* LATTICE2018 p 024
[26] Lee J, Bahri Y, Novak R, Schoenholz S S, Pennington J and Sohl-Dickstein J 2017 arXiv:1711.00165
[27] Jacot A, Gabriel F and Hongler C 2018 arXiv:1806.07572
[28] Lee J, Xiao L, Schoenholz S, Bahri Y, Novak R, Sohl-Dickstein J and Pennington J 2019 *Advances in Neural Information Processing Systems* vol 32 pp 8572–83
[29] Arora S, Du S S, Hu W, Li Z, Salakhutdinov R and Wang R 2019 arXiv:1904.11955
[30] Sohl-Dickstein J, Novak R, Schoenholz S S and Lee J 2020 arXiv:2001.07301
[31] Yang G and Hu E J 2020 arXiv:2011.14522
[32] Yaida S 2020 Non-gaussian processes and neural networks at finite widths *Mathematical and Scientific Machine Learning (PMLR)* pp 165–92
[33] Roberts D A, Yaida S and Hanin B 2021 arXiv:2106.10165
[34] Bordelon B, Canatar A and Pehlevan C 2020 Spectrum dependent learning curves in kernel regression and wide neural networks *Int. Conf. on Machine Learning (PMLR)* pp 1024–34
[35] Canatar A, Bordelon B and Pehlevan C 2021 *Nat. Commun.* **12** 1–12
[36] Simon J B, Dickens M and DeWeese M R 2021 arXiv:2110.03922
[37] Bahri Y, Dyer E, Kaplan J, Lee J and Sharma U 2021 arXiv:2102.06701
[38] Atanasov A, Bordelon B and Pehlevan C 2021 arXiv:2111.00034
[39] Banchi L, Pereira J and Pirandola S 2021 *PRX Quantum* **2** 040321
[40] Caro M C, Huang H Y, Cerezo M, Sharma K, Sornborger A, Cincio L and Coles P J 2022 *Nat. Commun.* **13** 1–11
[41] Cristianini N, Shawe-Taylor J, Elisseeff A and Kandola J 2001 *Advances in Neural Information Processing Systems* vol 14
[42] Meltzer D, Zheng H, Du Y H, Roberts D R and Liu J
[43] Guo Q, Leng J, Li H and Cattani C 2019 *Mathematics* **7** 147
[44] Han X, Zou C L and Tang H X 2016 *Phys. Rev. Lett.* **117** 123603