Quantum kernel machine learning with continuous variables

Laura J. Henderson[1,2], Rishi Goel[1], and Sally Shrapnel[1,2]
[1] *School of Mathematics and Physics, The University of Queensland, QLD 4072, Australia and*
[2] *ARC Centre for Engineered Quantum Systems, The University of Queensland, QLD, 4072, Australia.*

Machine learning (ML) is ubiquitous and it is natural to ask if quantum resources can be used to provide some advantage. One natural fit for these resources is kernel methods [1]. Kernels, which essentially provide a similarity metric between data points, appear as filters in convolutional neural networks [2], can represent attention matrices in transformer networks [3], are used as training signals for generative networks [4], and can provide a key mechanism for causal discovery [5]. Although there has been much work on studying quantum kernels in the discrete (qubit/qudit) case, there are only a few works which have evaluated specific, continuous variable (CV), infinite dimensional quantum encodings [1, 6–9].

This work presents the first unifying framework for CV quantum kernels, which we use to understand and analyse quantum kernels. This framework leverages the holomorphic representation of CV quantum computing [10], which represents CV quantum states through the quasi-probability distribution of holomorphic functions. Importantly, this representation provides a natural hierarchy of "quantumness" through the *stellar rank* – the number of zeros in the polynomial. More precisely, the evaluation of CV quantum kernels can be shown to be equivalent to CV sampling computation. Such computations can be simulated in $\mathcal{O}(2^n)$ provided there are two or more modes [11, 12]. States of infinite stellar rank, such as the cat state or the GKP state are outside of the stellar hierarchy, but can be approximated arbitrarily well by states of finite stellar rank. Using the holomorphic representation, we derive analytic expressions for four CV kernels: (i) a single-mode displaced Fock state kernel, (ii) a single-mode displaced and phase-shifted Fock state kernel, (iii) the general form for any CV multi-mode kernel of finite stellar rank, and (iv) a qudit kernel.

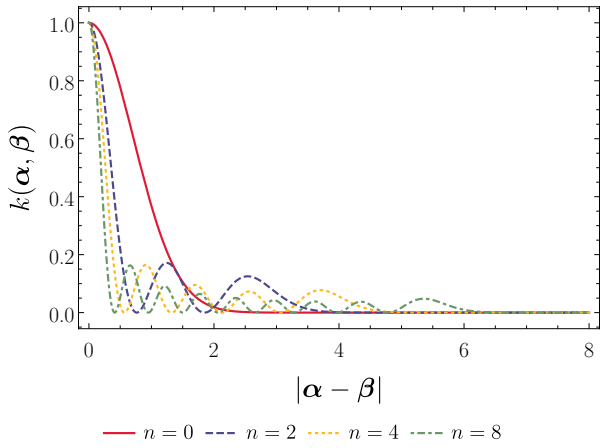

FIG. 1. The kernel function in the case of displacement encoding (Eq. (2)) as a function of $|\boldsymbol{\alpha}-\boldsymbol{\beta}|$, the distance in the data space, for various values of the initial Fock state, n.

First, we present the example of a single-mode displaced Fock state kernel, where a real vector is encoded into the displacement operator applied to a general Fock state

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\intercal \to \hat{D}(\alpha_1 + \mathrm{i}\alpha_2)\,|n\rangle \tag{1}$$

for $\alpha_1, \alpha_2 \in \mathbb{R}$, resulting in the encoded state having a stellar rank of n. We calculate the closed form of this kernel

$$k(\boldsymbol{\alpha},\boldsymbol{\beta}) = \left|\left\langle n \right| \hat{D}^\dagger(\alpha_1 + \mathrm{i}\alpha_2)\hat{D}(\beta_1 + \mathrm{i}\beta_2) \left| n \right\rangle\right|^2 \tag{2}$$

which is the product of a Gaussian and a polynomial of degree $4n$. While for a single mode encoding, we do not have exponential growth in simulablity with stellar rank, we will use stellar rank to generate intuition for the multi-mode encoding, which we present later. We prove that this particular kernel is radial and characteristic, meaning it can be used in probabilistic machine learning applications [13]. In figure 1, we plot the displaced Fock state kernel function, $k(\boldsymbol{\alpha},\boldsymbol{\beta})$ for various values of the stellar rank, n. As the stellar rank increases, the kernel's ability to distinguish between

large distances in the original data space, $|\boldsymbol{\alpha} - \boldsymbol{\beta}|$, improves, however the amplitude of each maxima decreases, leading to increasing concentration at low values (exponential concentration).

We construct a supervised learning classification task using the single-mode displaced Fock state kernel to exploit the underlying structure of the kernel. The task to classify an annular data set of three concentric circles with binary labels generated with multiple instances of the Scikit-learn dataset method, *make circles*. The results of the first task are shown in figure 2, where we compare the decision boundaries of five kernels. We find that accuracy increases with increasing stellar rank. Additionally, the displaced Fock state kernel out-performs the default Scikit-learn Gaussian kernel. However, when the Gaussian kernel has been tuned via Bayesian optimisation, which takes a significant amount of computational time, it is also able to classify the data to a high degree of accuracy. The displaced Fock state kernels require no hyperparameter tuning, suggesting that these quantum kernels are better suited for this task.

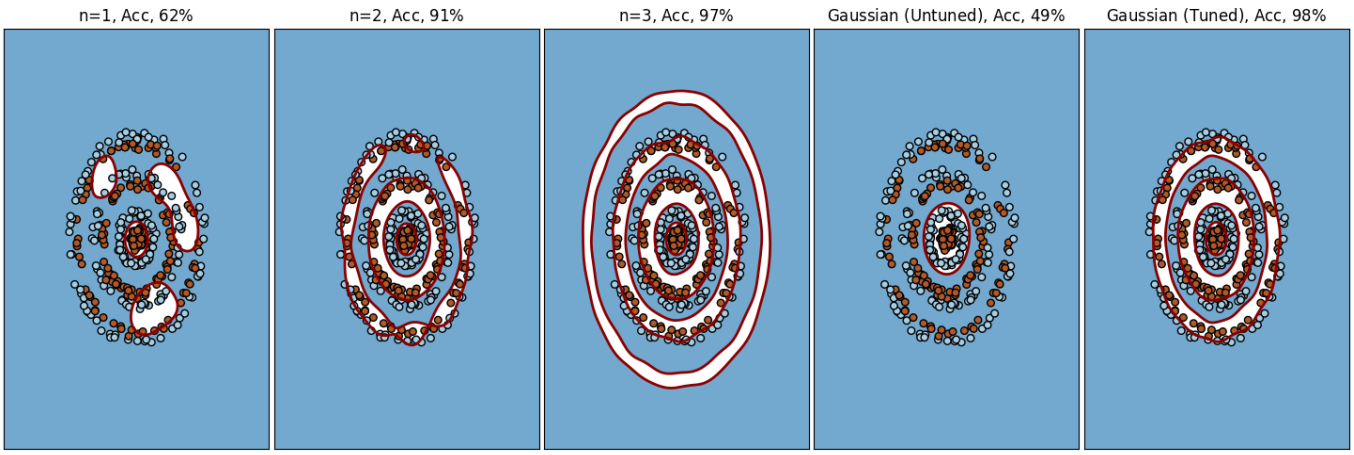

FIG. 2. Plots of the decision boundaries for $n = 1, 2, 3$ displaced Fock state kernels (left three panels) with no hyperparameter tuning, benchmarked against a classical Gaussian kernel (right two panels)

In figure 3, we present the effect of an additional hyperparameter $c > 0$ which rescales the data as $\boldsymbol{\alpha} \to c\boldsymbol{\alpha}$. Physically this corresponds to rescaling the value of the displacement applied to the initial Fock state, and is similar to the bandwidth parameter presented in [14]. By increasing this bandwidth hyperparameter, we are able to improve learning in the $n = 1$ kernel. However, when the hyperparameter is too large, the kernel becomes overfit.

In a separate learning experiment, we also find that when the data is very noisy, decreasing the bandwidth improves learnability, but at the cost of the effective stellar rank of the kernel.

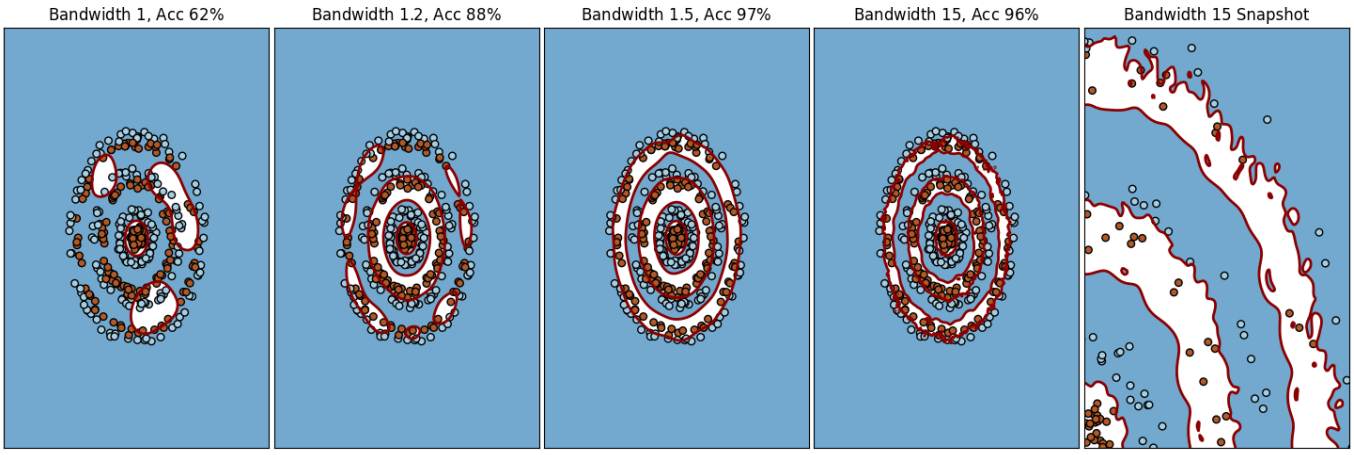

FIG. 3. Here we plot the performance of the $n = 1$ displaced Fock state kernel with bandwidth hyperparameter tuning.

Next we present the the single-mode displaced Fock state kernel with and added a phase shift, which allows for the

encoding of three-dimensional data with the addition of the compact real variable $\varphi \in [0, 2\pi)$ as:

$$(\alpha_1, \alpha_2, \varphi)^\intercal \to \hat{R}(\varphi)\hat{D}(\alpha_1 + \mathrm{i}\alpha_2)\,|n\rangle \tag{3}$$

where $\alpha_1, \alpha_2 \in \mathbb{R}$. This again results in a kernel that is the product of a Gaussian and a polynomial of degree $4n$ and prove it is no longer a radial kernel. We find similar functional behaviour as the single-mode displaced Fock state case, i.e. the Gaussian term suppresses kernel values representing large distances in the data.

We then present a closed form equation representing a general finite stellar rank multi-mode CV state. In this case, the kernel is a product of a Gaussian and an algebraic function. Specifically, the algebraic function is a solution to the polynomial equation $P_0(\boldsymbol{x}) = P_2(\boldsymbol{x})f(\boldsymbol{x})^2$. For this case, the classical simuablity will scale exponentially with stellar rank. From this, we expect that for all finite rank CV kernels, there will be kernel concentration beyond a threshold value due to the Gaussian term. We also show that in the case where data is encoded in a CV state of infinite stellar rank, the resulting kernel can be approximated arbitrarily well by a kernel of finite stellar rank.

Finally, we consider the case of qudit encoding, which can also be represented as a holomorphic function. In this case, the kernel is a polynomial of degree $4d$, where d is the dimension of the qudit, which acts as the stellar rank. With this we have shown that all quantum kernels are of the form of, or can be approximated by, a product of a Gaussian and an algebraic function.

[1] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, Phys. Rev. Lett. **122**, 040504 (2019).
[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
[3] Y.-H. H. Tsai, S. Bai, M. Yamada, L.-P. Morency, and R. Salakhutdinov, Transformer dissection: An unified understanding for transformer's attention via the lens of kernel, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, edited by K. Inui, J. Jiang, V. Ng, and X. Wan (Association for Computational Linguistics, Hong Kong, China, 2019) pp. 4344–4353.
[4] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, Demystifying MMD GANs, in *International Conference on Learning Representations* (2018).
[5] J. Mitrovic, D. Sejdinovic, and Y. W. Teh, Causal inference via kernel deviance measures, Advances in neural information processing systems **31**, https://doi.org/10.48550/arXiv.1804.04622 (2018).
[6] R. Ghobadi, Nonclassical kernels in continuous-variable systems, Phys. Rev. A **104**, 052403 (2021).
[7] P. Tiwari, S. Dehdashti, A. K. Obeid, P. Marttinen, and P. Bruza, Kernel method based on non-linear coherent states in quantum feature space, Journal of Physics A: Mathematical and Theoretical **55**, 355301 (2022).
[8] J. Liu, C. Zhong, M. Otten, A. Chandra, C. L. Cortes, C. Ti, S. K. Gray, and X. Han, Quantum Kerr learning, Machine Learning: Science and Technology **4**, 025003 (2023).
[9] C. Bowie, S. Shrapnel, and M. J. Kewming, Quantum kernel evaluation via Hong–Ou–Mandel interference, Quantum Science and Technology **9**, 015001 (2023).
[10] U. Chabaud and S. Mehraban, Holomorphic representation of quantum computations, Quantum **6**, 831 (2022).
[11] U. Chabaud, G. Ferrini, F. Grosshans, and D. Markham, Classical simulation of gaussian quantum circuits with non-gaussian input states, Phys. Rev. Res. **3**, 033018 (2021).
[12] U. Chabaud and M. Walschaers, Resources for bosonic quantum computational advantage, Physical Review Letters **130**, 10.1103/physrevlett.130.090602 (2023).
[13] B. K. Sriperumbudur, K. Fukumizu, and G. R. Lanckriet, Universality, characteristic kernels and RKHS embedding of measures, Journal of Machine Learning Research **12**, 2389 (2011).
[14] A. Canatar, E. Peters, C. Pehlevan, S. M. Wild, and R. Shaydulin, Bandwidth enables generalization in quantum kernel models, Transactions on Machine Learning Research (2023).