Kernel Method based on Non-Linear Coherent States in Quantum Feature Space

Prayag Tiwari [1],* Shahram Dehdashti [2],† Abdul Karim Obeid [2], Pekka Marttinen [1],‡ and Peter Bruza [2§]
[1] *Department of Computer Science, Aalto University, Finland. and*
[2] *School of Information Systems, Queensland University of Technology, Brisbane 4000, Australia.*¶
(Dated: July 28, 2022)

In this paper, by mapping datasets to a set of non-linear coherent states, the process of encoding inputs in quantum states as a non-linear feature map is re-interpreted. As a result of the fact that the Radial Basis Function is recovered when data is mapped to a complex Hilbert state represented by coherent states, non-linear coherent states can be considered as a natural generalisation of the associated kernels. In this paper, as an example of kernels based on non-linear coherent states, we propose kernel functions based on generalized hypergeometric functions, as orthogonal polynomial functions. The suggested kernel is implemented with the support vector machine on two well known datasets (make_circles, and make_moons) and outperforms the baselines, even when the level of noise is high. In addition, we study the impact of the geometrical properties of the feature space, obtained by the non-linear coherent states, on the SVM classification task, by considering the Fubini-Study metric of the associated coherent states.

I. INTRODUCTION

Quantum machine learning is a rapidly growing field of investigation. It can be argued that developments are being driven from two directions. Firstly, quantum computers offer a promise of massive improvement in the speed of computational processing [1–3]. Secondly, the mathematical framework of quantum mechanics is increasingly being seen as suitable for designing algorithms that aren't constrained by Boolean algebra and logic [4, 5]. The reasons that support the latter claim are many, e.g., the linearity of the Schrödinger equation [6, 7], which leads to the definition of superposed states in complex Hilbert spaces with an 'interference' term affecting probabilities. Consider modelling the dependence that measurement outcomes have on the preparation of states: duly named 'contextual scenarios' [8]. There is also the novelty of correlations observed through entanglement, discord, *etc* in quantum mechanics that go beyond classically correlated structures [9, 10], as well as quasi-distributions, which occur in phase space, so-called Wigner distributions [11, 12], and achieve negativity - this is not possible in Kolmogorovian probability theory. All of the preceding introduce the potential for access to an information space greater than that of classical alternatives [13–16]. This is encouraging for scientists wishing to apply quantum formalism within machine learning (ML) as a generalisation of probability theory.

In ML, kernel methods [17–22] are a class of categorization algorithms. Used within a wide range of methods and algorithms, they include the support vector machine (SVM) [23–25], kernel operators with principal components analysis (PCA) [26], spectral clustering [27], canonical correlation analysis [28], linear adaptive filters [29], and ridge regression [30]. Indeed, kernel methods are proving to assist also in deep neural networks, for which there are many recently published works [31, 32]. There exist vast prospects of kernel methods in ML due to the non-linear nature of the underlying data. Within application settings, data is not linearly separable, in most of the cases, and the requirement of the kernel then becomes to transform (the data) into higher dimensions where it may be (linearly) separable as can be seen in Figure 1. One of the well-known kernel functions in ML, the Radial Basis function (RBF), is defined by $K(x, x') = \exp\left[-|x - x'|^2/2\sigma^2\right]$ where x and x' are two sample elements, and σ controls the decision boundary [33–35]. It is worthwhile to mention that the RBF can be understood as an inner product of the linear coherent state, see [36], which is defined as an eigenstate of the annihilation operator of a harmonic oscillator. The idea of using a quantum mechanics formalism in kernel methods was suggested by Schuld and Killoran, who introduced squeezed kernels in feature spaces [37]. In fact, they defined the feature space as a set of squeezed states so that the kernel is obtained by inner products of the squeezed states [38, 39].

In this paper, we express non-linear coherent states [40–44] as a quantum feature space, such that kernel functions are defined as their inner products. We show that the mathematical structure of non-linear coherent states leads to

*Electronic address: prayag.tiwari@aalto.fi
†Electronic address: shahram.dehdashti@qut.edu.au
‡Electronic address: pekka.marttinen@aalto.fi
§Electronic address: p.bruza@qut.edu.au
¶The first two authors contributed equally in this research.

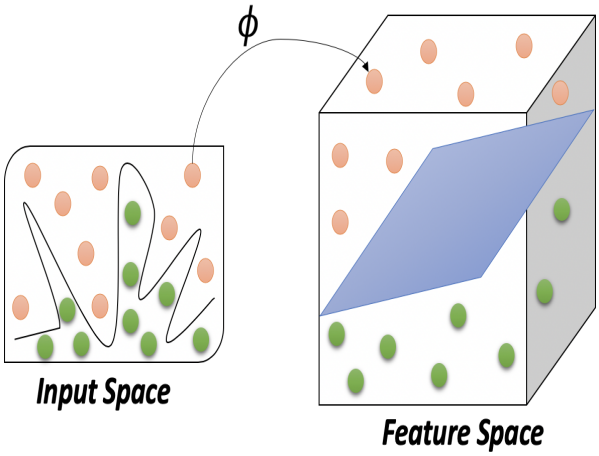

FIG. 1: Kernel visualization from low dimension input space to high dimension feature space.

infinite kernels. As an example of a non-linear feature space, we investigate coherent states constructed by several deformation functions. Generalized hypergeometric functions, as orthogonal polynomials, are identified to provide the associated kernel. Our proposed kernels based on non-linear coherent states are applied in an SVM classifier, along with the RBF and squeezed kernels as baselines, on two-well-known datasets (make_moons, make_circles). Kernels based on non-linear coherent states are shown to outperform the baselines (squeezed and RBF kernels) even as we increase the level of noise in the dataset (which increases difficulty of generalisation).

On the other hand, manifold machine learning is a novel and prospecting approach motivated by the belief that data should be mapped on an associated curved manifold for appropriate classification due to its nonlinear nature, geometric intuition, and computational feasibility [45, 46]. By considering this motivation, we study the geometrical properties of the feature space, by obtaining the Fubini-Study metric of non-linear coherent states. We show that the feature spaces of the non-linear coherent states correspond to surfaces with non-zero curvature, which opens a new line of investigation on how the feature space's curvature affects the accuracy of the SVM classifier.

The rest of the paper is organised as follows: in Section II, we briefly review the Kernel Method by using formalism of coherent states. Section III provides an overview on non-linear coherent states and introduces the coherent states of a quantum oscillator with variable mass. This allows the main contribution of the paper to be defined: a kernel function based on generalized hypergeometric functions. Section IV details the experimental design which allows the proposed kernel function to be empirically evaluated against two baselines: RBF and squeezed kernel. Section V discusses the results of the empirical evaluation. This section also examines the geometrical properties of a non-linear coherent state. Finally, Section VI concludes the article.

II. KERNEL METHOD

Traditional ML begins with a dataset of inputs $\mathcal{D} = \{x_1, \ldots, x_M\}$ drawn from a set $\mathcal{X}$. The goal is to produce a predictive model that allows patterns to be discovered in yet to be observed data. Kernel methods use the inner product $K(x, x')$ between any given two points $x, x' \in \mathcal{X}$ in feature space as a similarity measure, to build models that assist with predicting characteristics of novel data points.

Let $\mathcal{X}$ be a nonempty set, called the index set, and $\mathcal{H}$ a Hilbert space of functions $\phi : \mathcal{X} \to \mathbb{R}$. Then $\mathcal{H}$ is called a reproducing kernel Hilbert space endowed with the dot product, $\langle \cdot | \cdot \rangle$, if there exists a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with the two following properties: (i) k has the reproducing property, i.e., $\langle \phi | k(x, \cdot) \rangle = \phi(x)$ for all $\phi \in \mathcal{H}$; (ii) k spans $\mathcal{H}$ [47]. Therefore, a kernel function can be defined as an inner product of two inputs mapped to some feature space, as $k(x, x') = \langle \phi(x), \phi(x') \rangle$ [20, 36]. Note that in particular $\langle k(x, \cdot) | k(x', \cdot) \rangle = k(x, x') = k(x', x)$ guarantees the symmetry of the arguments of k.

A. Kernel Method by Using Coherent States

A coherent state is the specific quantum state of the quantum harmonic oscillator, often described as a state which has dynamics most closely resembling the oscillatory behavior of a classical harmonic oscillator [48]. A coherent state is defined as superposition of number state $|n\rangle$ as following:

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \,, \tag{1}$$

where number states satisfy $\langle n|m\rangle = \delta_{n,m}$, where $\delta_{n,m}$ is Kronecker delta. It follows that the inner product of two coherent states is given by:

$$\langle \alpha|\beta\rangle = e^{|\alpha-\beta|^2/2} \tag{2}$$

A coherent state adheres the following:

1. It is obtained by applying the displacement operator $D(\alpha) = \exp\{(\alpha^* a^\dagger - \alpha a)\}$ on a reference state: $|\alpha\rangle = D(\alpha)|0\rangle$, in which $|\alpha\rangle$ is given by the relation (1).
2. It is an eigenvector of the annihilation operator, $a|\alpha\rangle = \alpha|\alpha\rangle$.
3. It fulfills the minimum uncertainty relation, i.e., $\Delta(\mathbf{x}) = \Delta(\mathbf{p}) = 1/\sqrt{2}$, where $\Delta(\mathbf{x}) = \sqrt{\langle \mathbf{x}^2\rangle - \langle \mathbf{x}\rangle^2}$, where $\langle \mathbf{x}\rangle = \langle \alpha| \mathbf{x} |\alpha\rangle$.
4. It is over-complete, $K(\alpha, \alpha') = \langle \alpha|\alpha'\rangle \neq \delta(\alpha - \alpha')$, i.e., coherent states fulfill the resolution of the identity, $\int d\mu(\alpha)|\alpha\rangle\langle\alpha| = \mathbb{I}$, which leads to the following relation:

$$\int d\mu(\mathbf{x})\langle\phi|\mathbf{x}\rangle\langle\mathbf{x}|\psi\rangle = \langle\phi|\psi\rangle. \tag{3}$$

The latter, i.e., item 4, implies an arbitrary function can be expressible as a linear combination of kernel functions in a "reproducing Hilbert space" [49]. We should mention that the first above-mentioned property leads to define a displacement-type coherent states [48], for generalized annihilation and creation operators. Moreover a Gazeau-Klauder coherent state is defined by the second property and fulfils the third property [48]. It should be noted that the latter, i.e., resolution of the identity, fulfils all types of coherent states. For methods of implementations of non-linear coherent states see [50] and references therein.

Recently Schuld and Killoran published a paper in which data is mapped into a feature space established by squeezed states [37]. Squeezed states are states that saturate the Heisenberg uncertainty principle; additionally, the quadrature variance of position and momentum depend on a parameter, so-called squeezing parameter. The squeezing parameter causes the uncertainty to be squeezed into one of its quadrature components, while stretching the uncertainty for the other component, i.e. $\Delta(\mathbf{x}) = \exp\{2\zeta\}/\sqrt{2}$ and $\Delta(\mathbf{p}) = \exp\{-2\zeta\}/\sqrt{2}$, where ζ is the squeezing parameter. Therefore, the squeezing parameter controls uncertainty via a quadrature component, while the third and fourth properties of coherent states are preserved. However, as we mentioned, one of the well-known kernel functions in ML, the Radial Basis function (RBF), is defined by

$$K(\mathbf{x}, \mathbf{x}'; \sigma) = \exp\{(-||\mathbf{x} - \mathbf{x}'||^2/2\sigma^2)\}, \tag{4}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are two sample elements, and σ is understood as a free parameter. However, drawing a comparison between relations (2) and (4), inspires someone to interpret the RBF as an inner product of two coherent states. This interpretation opens the door to define new kernels and consequently to improve the kernel method [36].

III. NON-LINEAR COHERENT STATE

As mentioned before, some efforts have been devoted to study possible generalization of the quantum harmonic oscillator algebra [41]. A deformed algebra is a non-trivial generalization of a given algebra through the introduction of one or more deformation parameters, such that, in a certain limit of the parameters, the non-deformed algebra can be recovered.

A particular deformation of the Weyl-Heisenberg algebra led to the notion of f-deformed oscillator [41]. An f-deformed oscillator is a non-harmonic system where its dynamical variables (creation and annihilation operators) are constructed from a non-canonical transformation through

$$\hat{A} = \hat{a} f(\hat{n}) = f(\hat{n}+1)\hat{a}, \tag{5}$$

$$\hat{A}^\dagger = f^\dagger(\hat{n})\hat{a}^\dagger = \hat{a}^\dagger f^\dagger(\hat{n}+1). \tag{6}$$

where $f(\hat{n})$ is called a deformation function by which non-linear properties of this system are governed. It is worth mentioning that when the deformation function approaches 1, i.e., $f(n) \to 1$, the non-deformed commutation relation $[\hat{a}, \hat{a}^\dagger] = 1$ is recovered. Hence, a non-linear coherent state is defined as the right-hand eigenstates of the deformed annihilation operator $\hat{A}$ as follows [48]:

$$\hat{A}|\alpha\rangle_f = \alpha|\alpha\rangle_f. \tag{7}$$

From equation (7) one can obtain an explicit form of the non-linear coherent states in the number state representation as,

$$|\alpha\rangle_f = \frac{1}{\mathcal{N}} \sum_n^M \frac{\alpha^n}{\sqrt{n![f(n)]!}} |n\rangle, \tag{8}$$

where M can be finite, or infinite (corresponding to finite or infinite dimensional Hilbert spaces), α is a complex number and $[f(n)]! = \prod_{i=0}^{n} f(i)$, with $[f(0)]! = 1$. The normalization factor $\mathcal{N}$ is given by

$$\mathcal{N} = \left(\sum_{n=0}^{M} \frac{|\alpha|^{2n}}{n![f(n)]!} \right)^{-1/2}. \tag{9}$$

Therefore, by mapping multi-dimensional input set $\mathbf{x} = (x_1, \cdots, x_D)^T \in \mathbb{R}^D$ into non-linear coherent states, defined by the relation (8), a feature space can be defined as

$$\phi : (x_1, \cdots, x_D) \to |x_1\rangle_f \otimes |x_2\rangle_f \otimes \cdots \otimes |x_D\rangle_f.$$

In addition, the associated kernel is obtained by the inner product of non-linear coherent states as the following:

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{D} {}_f\langle x_i | x_i' \rangle_f, \tag{10}$$

in which

$${}_f\langle x_i | x_i' \rangle_f = \frac{1}{\mathcal{N}^2} \sum_{n=0}^{M} \frac{(x_i x_i')^n}{n![f(n)]!}. \tag{11}$$

A. Geometrical properties of the Hilbert Space constructed by the Non-linear coherent state

For understanding the role of the nonlinear function f, we will study the geometrical properties of above-mentioned feature spaces. We can define the line element of the feature space by using the Fubini-Study metric. A suitable metric between two Hilbert space vectors, e.g, $|\psi\rangle$ and $|\phi\rangle$, is defined by [51],

$$d(|\psi\rangle, |\phi\rangle) = \min || \, |\psi\rangle - e^{i\delta} \, |\phi\rangle \, ||, \; 0 \leq \delta \leq 2\pi. \tag{12}$$

The infinitesimal form of this metric is given by the Fubini-Study metric:

$$ds^2 = ||d \, |\psi\rangle \, ||^2 - || \, \langle\psi| \, d \, |\psi\rangle \, ||^2. \tag{13}$$

in which $|\psi\rangle$ is a typical point in the complex Hilbert space. Therefore, by using the definition (13), the Fubini-Study metric for the Hilbert space constructed by f-deformed coherent states $\left|\psi = re^{i\phi}\right\rangle$ is given by

$$ds^2 = \Omega(r) \left(dr^2 + r^2 d\phi^2 \right), \tag{14}$$

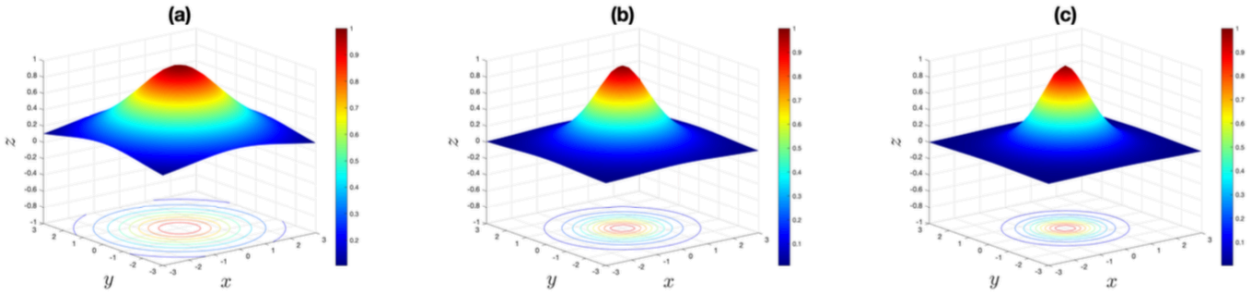

FIG. 2: Schematic shape of the kernel function (26) for $\beta = -0.5$, $\beta = -0.1$ and $\beta = -0.01$ in plots (a), (b) and (c), respectively. The input x is fixed at $(0, 0)$ and x' is varied.

which is a surface with non-zero curvature. The conformal factor Ω is given by

$$\Omega(r) = \frac{\partial_r^2 \mathcal{N}(r)}{\mathcal{N}(r)} + \frac{\partial_r \mathcal{N}(r)}{r\mathcal{N}(r)} - \left(\frac{\partial_r \mathcal{N}(r)}{r\mathcal{N}(r)}\right)^2. \tag{15}$$

with $\partial_r \equiv \frac{\partial}{\partial r}$. By applying the Einstein summation rule, the non-zero Christoffel symbols are given by

$$\Gamma^r_{rr} = \frac{1}{2}\partial_r \ln \Omega(r), \tag{16}$$

$$\Gamma^r_{\phi\phi} = -\left(r + \frac{r^2}{2}\partial_r \ln \Omega(r)\right), \tag{17}$$

$$\Gamma^\phi_{\phi r} = \frac{1}{r} + \frac{1}{2}\partial_r \ln \Omega(r). \tag{18}$$

Hence, the non-zero Ricci tensors are given by,

$$R_{rr} = -\frac{1}{2}\left(\partial_r^2 \ln \Omega(r) + \frac{1}{r}\partial_r \ln \Omega(r)\right) \tag{19}$$

$$R_{\phi\phi} = -\frac{r}{2}\left(\partial_r \ln \Omega(r) + r\partial_r^2 \ln \Omega(r)\right) \tag{20}$$

Finally, the Ricci scalar, $R = g^{ab}R_{ab}$, is obtained as

$$R = -\Omega(r)^{-1}\left(\partial_r^2 \ln \Omega(r) + \frac{1}{r}\partial_r \ln \Omega(r)\right) \tag{21}$$

It is worthwhile to note that the metric (14), which describes the feature space, is a surface that is conformal with the flat space, i.e., conformally preserves angles, while lengths can be changed. In fact, this metric represents a two-dimensional curved space, depending on the conformal function $\Omega(r)$. In the following, we consider two examples of non-linear coherent states.

1. *Example One*

By considering the following deformation function:

$$f(n) = (-\beta)^n (2 - 1/\beta)_n, \tag{22}$$

where $(u)_n = u(u-1)\cdots(u+n-1)$ represents the Pochhammer symbol, one can define a non-linear coherent state as follows [52]:

$$|x, \beta\rangle = \frac{1}{\mathcal{N}(x)} \sum_{n=0}^{\infty} \frac{x^n}{\rho_n} |n\rangle \tag{23}$$

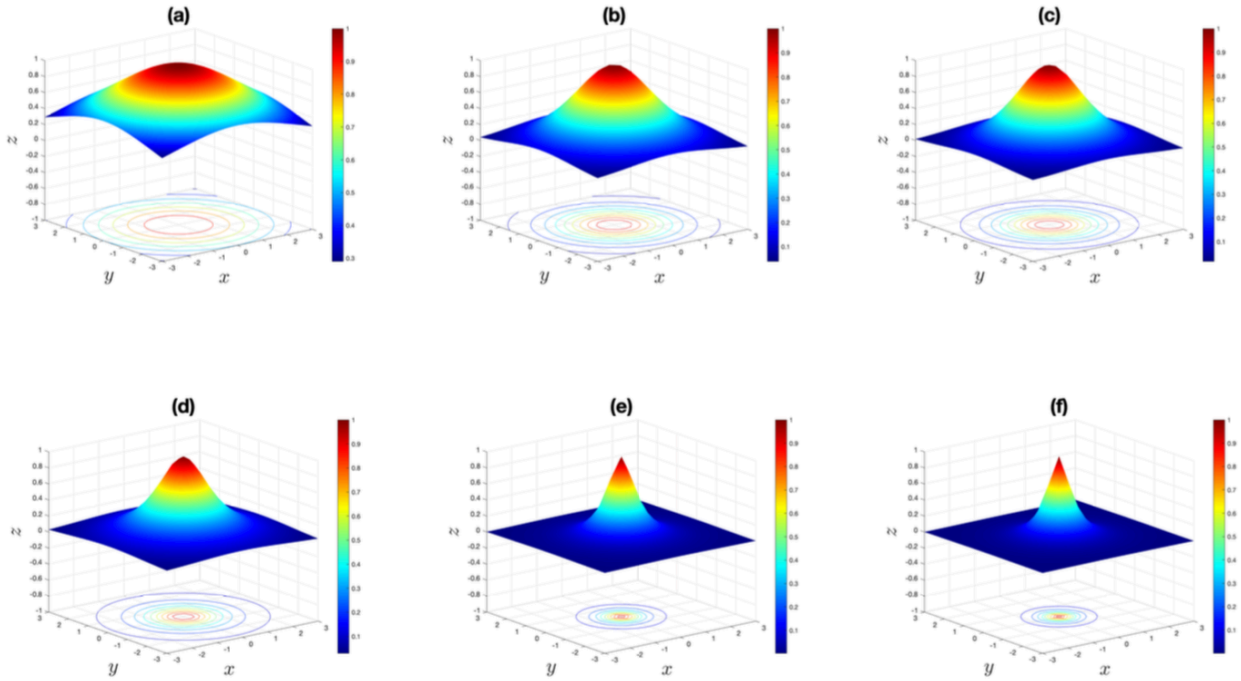

FIG. 3: Schematic shape of the kernel function (25) for $\beta = 0.01$, and $\gamma = 0.1$, $\gamma = 0.5$ and $\gamma = 0.9$, respectively in plots (a), (b) and (c); plots (d), (e) and (f) indicate the shape of the same kernel for $\beta = -0.9$ and $\gamma = 0.1$, $\gamma = 0.5$ and $\gamma = 0.9$.

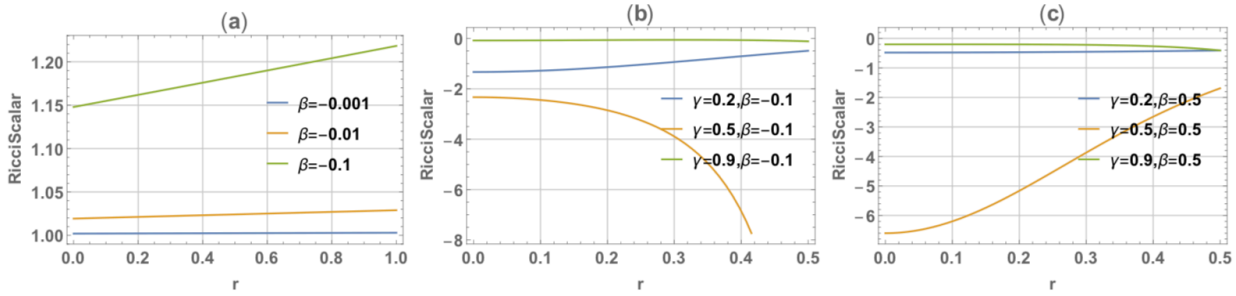

FIG. 4: Ricci scalar of the Fubini-Study metric of a feature space constructed by deformation coherent states i.e., the equations (26) and (31) as function of r, for different values of parameters β and γ.

in which $\rho_n = n!(-\beta)^n(2-1/\beta)_n$. The normalization factor $\mathcal{N}(x)$ is given by

$$\begin{aligned}\mathcal{N}^2(x) &= \sum_{n=0}^{\infty}\left(\frac{(1/\beta)^n}{n!(2-1/\beta)_n}\right)^2|x|^{2n}\\ &= {}_0F_3(;1,2-1/\beta,2-1/\beta;|x/\beta|^2),\end{aligned} \tag{24}$$

where ${}_0F_3(;1,2-1/\beta,2-1/\beta;|x/\beta|^2)$ is the generalized hypergeometric function.

By considering a multi-dimensional input data set of vectors $\mathbf{x} = (x_1,\cdots,x_D)^T \in \mathbb{R}^D$, one can define the joint state of D deformed coherent states,

$$\phi : (x_1,\cdots,x_D) \to |x_1,\beta\rangle \otimes |x_2,\beta\rangle \otimes \cdots \otimes |x_D,\beta\rangle.$$

Therefore, the kernel is defined as the following:

$$K(\mathbf{x},\mathbf{x}') = \prod_{i=1}^{D} \langle x_i;\beta|x_i';\beta\rangle. \tag{25}$$

Note that data can map into real part of the associated coherent state as the following:

$$\langle x_i;\beta|x_i';\beta\rangle = \frac{{}_0F_3(;1,2-1/\beta,2-1/\beta;x_ix_i'/\beta^2)}{\mathcal{N}(x)\mathcal{N}(x')} \tag{26}$$

In addition, data can be mapped into the phase part of the coherent state, i.e.,

$$\langle x_i;\beta|x_i';\beta\rangle = \frac{{}_0F_3(;1,2-1/\beta,2-1/\beta;y^2e^{i(x_i-x_i')}/\beta^2)}{{}_0F_3(;1,2-1/\beta,2-1/\beta;y^2/\beta^2)}, \qquad y=\text{Const.} \tag{27}$$

Figures 2(a), 2(b) and 2(c) schematically illustrate the kernel function (26), for $\beta=-0.5$, $\beta=-0.1$ and $\beta=-0.01$. The plots indicate that increasing the value of the parameter β increases the sharpness of the curve.

In the table I, the kernel is denoted by KMNCS1. By considering the non-linear coherent state (23) and the normalization coefficient (24), Fig. 4(a) shows the Ricci scalar for different values of β. This plot indicates that the feature space is a complex Hilbert space with non-zero curvature.

2. Example two

As another example, we can consider the following deformation function $f(n)$:

$$f(n) = \sqrt{\frac{\gamma}{1-\gamma}}\ \frac{1}{\beta+n} \tag{28}$$

where $0<\gamma<1$ and $-1<\beta<1$. The associated non-linear coherent state is defined as follows:

$$|x\rangle = \frac{1}{\mathcal{N}(x)}\sum_{n=0}^{\infty}\frac{x^n}{\rho_n}|n\rangle\,, \tag{29}$$

in which $\rho_n = \left[\left(\frac{\gamma}{1-\gamma}\right)^n n!(1+\beta)_n\right]^{1/2}$ and normalization factor $\mathcal{N}(x)$ is given by

$$\begin{aligned}\mathcal{N}^2(x) &= \sum_{n=0}^{\infty}\frac{1}{n!(1+\beta)_n}\left|\sqrt{\frac{\gamma}{1-\gamma}}x\right|^{2n}\\ &= {}_0F_1(;1+\beta;\frac{\gamma}{1-\gamma}|x|^2),\end{aligned} \tag{30}$$

where ${}_0F_1(;1+\beta;|x|^2)$ is the generalized hypergeometric function. Therefore, the kernel can be defined as follows:

$$K(\mathbf{x},\mathbf{x}') = \prod_{i=1}^{N}\langle x_i';\beta;\gamma|x_i;\beta;\gamma\rangle = \prod_{i=1}^{N}\frac{{}_0F_1(;1+\beta;\frac{\gamma}{1-\gamma}x_ix_i')}{\mathcal{N}(x)\mathcal{N}(x')}. \tag{31}$$

when data mapped into the real part of associated coherent states; moreover, data can be mapped into the phase part of coherent state:

$$K(\mathbf{x},\mathbf{x}') = \prod_{i=1}^{N}\langle x_i';\beta;\gamma|x_i;\beta;\gamma\rangle = \prod_{i=1}^{N}\frac{{}_0F_1(;1+\beta;\frac{\gamma}{1-\gamma}y^2e^{i(x_i-x_i')})}{{}_0F_1(;1+\beta;\frac{\gamma}{1-\gamma}y^2)}, \qquad y=\text{Const.} \tag{32}$$

In Table I, this kernel is denoted by KMNCS2 and KMNCS3, for negative and positive values of β, respectively. Figure 3, panels (a)-(c), schematically illustrate the kernel function (31) for $\beta=0.1$ and $\gamma=0.1$, $\gamma=0.5$, and $\gamma=0.9$; $\beta=-0.9$ with same γ are illustrated in plots (d)-(f), respectively. The figure illustrates that decreasing β and increasing γ increases the sharpness of the curve. In addition, Fig. 4(b) and 4(c) represent the Ricci scalar for different values of β and γ. Decreasing γ and increasing β cause the value of the Ricci curvature to increase. Despite the fact that the feature space constructed using a non-linear coherent state has a non-zero curvature, the RBF, constructed from the linear coherent state, is a kernel on a flat space with Ricci scalar equal to zero.

TABLE I: Kernels, Constructed by Non-Linear Coherent States, Normalization Factor, and Parameters

Non-linear Coherent State's Kernel				
	Non-Linear Function	Normalization Factor	Kernel	Parameters
RBF	$f(n) = 1$	$\exp\left[\lvert x\rvert^2/2\right]$	$\exp\left[\lvert x - x'\rvert^2/2\sigma^2\right]$	σ
Squeeze	$f(n) = \sqrt{\frac{c}{2}(1+n)}$	$\operatorname{sech} c$	$\sqrt{\frac{\operatorname{sech}^2 c}{1-\tanh^2 ce^{i(x-x')}}}$	c
KMNCS1	$f(n) = \frac{\beta}{\beta-n}$	${}_0F_3(;1,2-1/\beta,2-1/\beta;x^2/\beta^2)$	$\frac{{}_0F_3(;1,2-1/\beta,2-1/\beta;x_i x_i'/\beta^2)}{\mathcal{N}(x)\mathcal{N}(x')}$	$-1 < \beta < 1$
KMNCS2	$f(n) = \frac{\gamma}{\gamma-1}\frac{1}{\beta+n}$	${}_0F_1(1+\beta, \frac{\gamma}{1+\gamma}x^2)$	$\frac{{}_0F_1(1+\beta, \frac{\gamma}{1+\gamma}xx')}{\mathcal{N}(x)\mathcal{N}(x')}$	$0 < \gamma < 1, -1 < \beta < 0$
KMNCS3	$f(n) = \frac{\gamma}{\gamma-1}\frac{1}{\beta+n}$	${}_0F_1(;1+\beta, \frac{\gamma}{1-\gamma}x^2)$	$\frac{{}_0F_1(;1+\beta, \frac{\gamma}{1-\gamma}xx')}{\mathcal{N}(x)\mathcal{N}(x')}$	$0 < \gamma < 1, 0 < \beta < 1$
KMNCS	$f(n) = \frac{\beta}{\beta-n}$	${}_0F_3(;1,2-1/\beta,2-1/\beta;\exp\lvert x/\beta\rvert^2)$	$\frac{{}_0F_3(;1,2-1/\beta,2-1/\beta;\exp\left[\lvert x-x'\rvert^2/\beta^2\right])}{\mathcal{N}(x)\mathcal{N}(x')}$	$-1 < \beta < 1$

TABLE II: Kernels, constructed by the phase of Non-Linear Coherent States, Normalization Factor, and Parameters

Non-linear Coherent State's Kernel				
	Non-Linear Function	Normalization Factor	Kernel	Parameters
Squeeze	$f(n) = \sqrt{\frac{c}{2}(1+n)}$	$\operatorname{sech} c$	$\sqrt{\frac{\operatorname{sech}^2 c}{1-\tanh^2 ce^{i(x-x')}}}$	c
KMNCS1_Phase	$f(n) = \frac{\beta}{\beta-n}$	${}_0F_3(;1,2-1/\beta,2-1/\beta;y^2/\beta^2)$	$\frac{{}_0F_3(;1,2-1/\beta,2-1/\beta;y^2e^{x-x'}/\beta^2)}{\mathcal{N}(y)}$	$-1 < \beta < 1,$
KMNCS2_Phase	$f(n) = \frac{\gamma}{\gamma-1}\frac{1}{\beta+n}$	${}_0F_1(1+\beta, \frac{\gamma}{1+\gamma}y^2)$	$\frac{{}_0F_1(1+\beta, \frac{\gamma}{1+\gamma}y^2e^{x-x'})}{\mathcal{N}(y)}$	$0 < \gamma < 1, -1 < \beta < 0$
KMNCS3_Phase	$f(n) = \frac{\gamma}{\gamma-1}\frac{1}{\beta+n}$	${}_0F_1(;1+\beta, \frac{\gamma}{1-\gamma}y^2)$	$\frac{{}_0F_1(;1+\beta, \frac{\gamma}{1-\gamma}y^2e^{x-x'})}{\mathcal{N}(y)}$	$0 < \gamma < 1, 0 < \beta < 1$

IV. EXPERIMENTAL DESIGN

In the following, we empirically evaluate the different kernels constructed by linear and non-linear coherent states, i.e. the RBF kernel, the Squeeze state, KMNCS1 defined in Equation (26), KMNCS2 and KMNCS3 defined in Equation (31) for $\beta < 0$ and $\beta > 0$, respectively, and finally KMNCS obtained by combining RBF and KMNCS1. All kernels are summarized in Table I. In addition, we evaluate kernels built by using the phase part of coherent states, i.e., kernels (27) and (32). Table II summarizes kernels that is defined based on the phase in this paper. In our comparison, we use the SVM classifier, which uses a kernel to define the decision boundary for classifying the data points. Generally, the hyper-parameter σ in the Gaussian kernel is used to optimize the performance of the SVM, using a cost function based on mis-classification of data points in the training set. We kept the hyperparameter $\sigma = 1$ as the optimal value.

We used two synthetic toy datasets, which are commonly used to compare the performance of kernels. The two datasets (*make_moons*[82] and *make_circles*[83]) are taken from sklearn. There is some flexibility in regard to each dataset. Random noise can be introduced by adjusting different parameters for each dataset: the **Moons** dataset generates two half circles with the noise parameters affecting 'interleave', the **Circles** dataset generates concentric circles also affected by 'interleave' via the noise parameter. To comprehensively study the performance of the different kernels, we varied the difficulty of the classification task systematically by adding noise to the target variable. For example, if the noise parameter was 0.2, we added Gaussian noise with a standard deviation of 0.2 to the output. When the different classes become inseparable due to noise, classifying the data points accurately becomes more challenging.

The decision region for class 1 is color-coded 'red', and 'blue' for class 0. We have also recorded differing values of $flip\_y$ which is an inbuilt parameter in $make\_classification$ [84] also from sklearn. A large $flip\_y$ supplements the effect of noise, making accurate classification even more challenging. The $factor$ and $random\_state$ in the setting have also been recorded. We divided the dataset into training set and test set. We used 5-fold cross-validation during training to avoid over-fitting. We examined using different values for the parameters of the different kernels to evaluate how they affect the classification performance and understand how the decision boundaries are formed.

V. RESULTS

The sklearn package includes a "fit" method that is used for training the model using the training set. To compute the score by cross-validation of SVM, "cross_val_score" is used also from sklearn with a "5"-fold cross-validation.

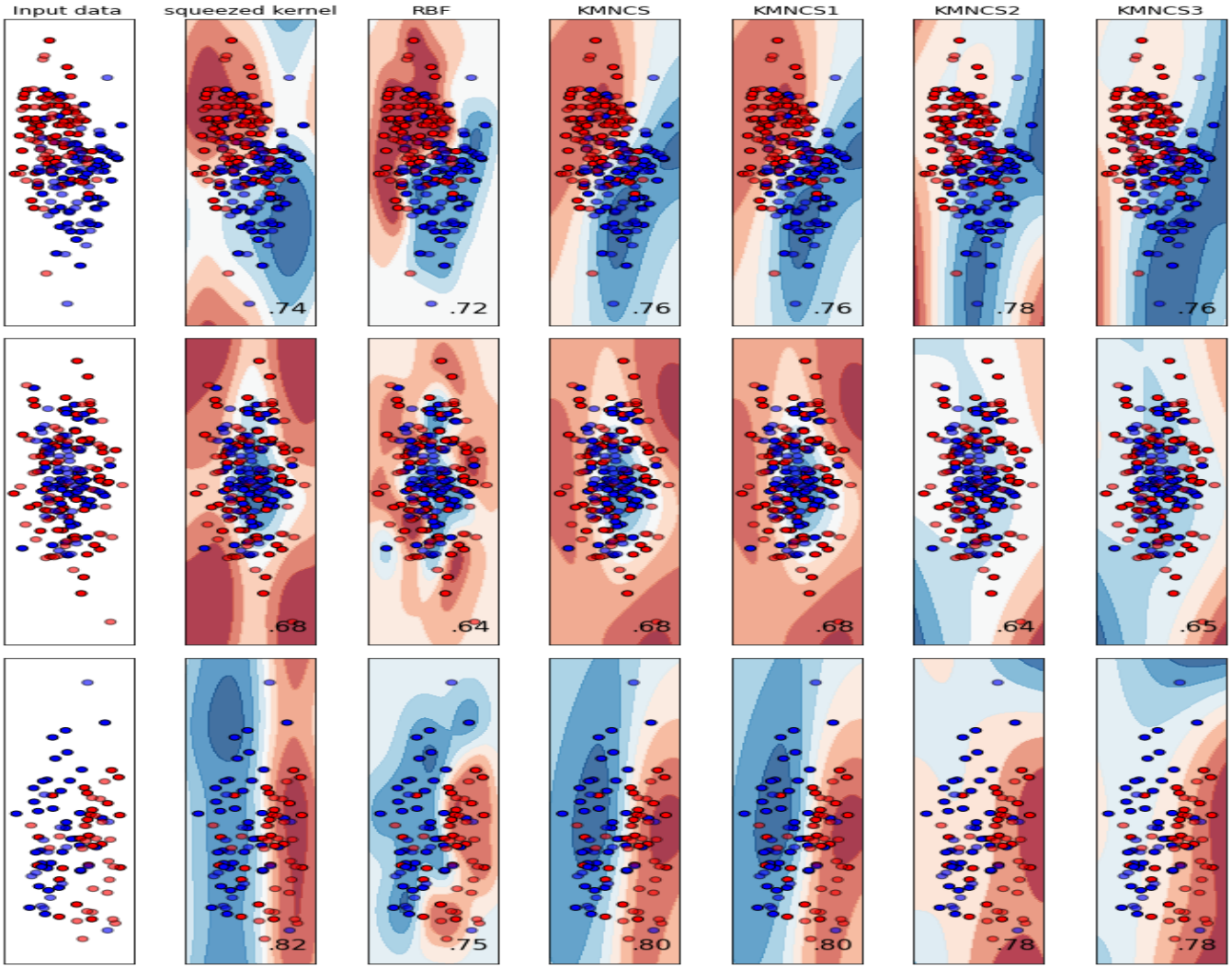

FIG. 5: Kernel visualization (Squeezed, RBF, KMNCS, KMNCS1, KMNCS2, and KMNCS3) when noise is 0.5 on the make_moons dataset, and make_circles dataset. We kept $\beta = -0.01$ for KMNCS and KMNCS1, $\beta = -0.5$ and $\gamma = 0.2$ for KMNCS2, and $\beta = 0.5$ and $\gamma = 0.2$ for KMNCS3

As we can see from Table III, we have checked the performance of KMNCS and KMNCS1 with different values of β. We have extra parameter γ along with β for KMNCS2 and KMNCS3 (see a comparison in Table III, IV, and V). The performance of all kernels is pretty good with very low noise, i.e., = 0.1, and this happens because of the fact that input data is separable, and it is an easy task for the classifier to classify them. However, when the noise ratio starts increasing, then the performance of the RBF deteriorates. For instance, when noise is 0.5 or 0.7, non-linear coherent states' kernels outperform the squeezed and RBF kernels as evident in Table III. By changing the metric of the feature space, we can adjust the separability of the data in the feature space. Hence, kernels based on the non-linear coherent states have a better performance than either baseline (Squeezed kernel and the RBF) for both datasets in presence of a high level of noise (see a comparison in Table III, IV, and V). This performance in the presence of such noise demonstrates the effectiveness of kernels based on the non-linear coherent states. The decision boundary is depicted in Figure 5. The performance of kernels based on the non-linear coherent states can be improved by proper tuning of the hyperparamaters β and γ for the datasets. The decision boundaries of the kernels can be also adjusted by proper tuning.

We also studied the impact of the size of the training set, by increasing the number of training samples to 3000 and the results are shown in Tables IV and V. We also considered different noise ratios and also tuned hyperparameters. The results illustrate that kernels based on non-linear coherent states outperform both baselines. Note that there is a trade-off between accuracy and the sharpness of the decision boundaries. It is possible to improve accuracy by modifying β and γ, but this may obscure the decision boundaries. This fact could be related to the bias-variance

TABLE III: Accuracy on Squeezed kernel, RBF, KMNCS, KMNCS1, KMNCS2 and KMNCS3 kernels on the make_moon and make_circles datasets with the following parameters: n_samples=200, random_state=50, y_flip=0.2, factor=0.5, n_informative=2, 5-fold cross validation, 60% training set and 40% test set. Noise effects are introduced in the data to increase difficulty of accurate classification.

		Squeezed kernel	RBF	KMNCS	KMNCS1	KMNCS2	KMNCS3
moon	Noise=0.1	0.90	0.99	0.97 (β=-0.001)	0.88 (β=-0.001) 0.88 (β=-0.01)	0.91(β=-0.5,γ=0.2) 0.97(β=-0.5,γ=0.9)	0.90(β=0.5, γ=0.2) 0.89(β=0.5, γ=0.9)
	Noise=0.2	0.85	0.95	0.95 (β=-0.001) 0.95 (β=-0.01) 0.94 (β=-0.1)	0.89 (β=-0.001) 0.89(β=-0.01) 0.84 (β=-0.1)	0.90 (β=-0.5,γ=0.2) 0.90 (β=-0.1,γ=0.2) 0.89 (β=-0.1,γ=0.9)	0.91 (β=0.5,γ=0.2) 0.90 (β=0.1,γ=0.2) 0.86 (β=1, γ=0.2) 0.90 (β=0.5, γ=0.9)
	Noise=0.3	0.82	0.89	**0.90** (β=-0.001) **0.90** (β=-0.01)	0.85(β=-0.01) 0.84 (β=-0.1)	0.88(β=-0.5, γ=0.2) 0.88 (β=-0.5, γ=0.9)	0.88 (β=0.5, γ=0.2) 0.80 (β=0.5, γ=0.9)
	Noise=0.4	0.79	0.82	**0.84** (β=-0.001) **0.84** (β=-0.01)	0.82(β=-0.001) 0.81 (β=-0.1)	**0.85** (β=-0.5, γ=0.2) 0.82 (β=-0.5, γ=0.9)	**0.84** (β=0.5, γ=0.2) 0.76 (β=0.5, γ=0.9)
	Noise=0.5	0.74	0.72	**0.75** (β=-0.001) **0.76** (β=-0.01) **0.76** (β=-0.1)	**0.76** (β=-0.001) **0.76** (β=-0.01) 0.74 (β=-0.1)	**0.78** (β=-0.5, γ=0.2) **0.75** (β=-0.5, γ=0.9)	**0.76** (β=0.5, γ=0.2) 0.55 (β=0.5, γ=0.9)
	Noise=0.7	0.62	0.65	0.64 (β=-0.001) 0.62 (β=-0.01) 0.65 (β=-0.1)	0.65 (β=-0.001) **0.68** (β=-0.01) **0.69** (β=-0.1)	**0.69** (β=-0.5, γ=0.2) 0.64 (β=-0.5, γ=0.9)	**0.68** (β=0.5, γ=0.2) 0.50 (β=0.5, γ=0.9) 0.62 (β=0.9, γ=0.9) **0.69**(β=0.9, γ=0.2)
Circle	Noise=0.1	0.95	1.0	1.0 (β=-0.001)	1.0 (β=-0.001) 1.0 (β=-0.01)	0.99 (β=-0.5,γ=0.2) 1.0 (β=-0.5,γ=0.9)	0.99 (β=0.5, γ=0.2) 0.85 (β=0.5, γ=0.9)
	Noise=0.2	0.86	0.90	0.88 (β=-0.001) 0.89 (β=-0.01) **0.95** (β=-0.1)	0.89 (β=-0.001) 0.89 (β=-0.01) **0.94** (β=-0.1)	0.89 (β=-0.5,γ=0.2) 0.78 (β=-0.1,γ=0.2) 0.76 (β=-0.1,γ=0.9)	0.90 (β=0.5,γ=0.2) 0.76 (β=0.1,γ=0.2) 0.76 (β=1,γ=0.2) 0.74 (β=0.5, γ=0.9)
	Noise=0.3	0.80	0.80	0.79 (β=-0.001) 0.79 (β=-0.01)	0.79 (β=-0.01) 0.80 (β=-0.1)	0.78 (β=-0.5, γ=0.2) 0.76 (β=-0.5, γ=0.9)	**0.81** (β=0.5, γ=0.2) 0.76 (β=0.5, γ=0.9)
	Noise=0.4	0.71	0.69	**0.72** (β=-0.001) **0.72** (β=-0.01)	**0.74** (β=-0.001) **0.74** (β=-0.1)	0.68 (β=-0.5, γ=0.2) 0.62 (β=-0.5, γ=0.9)	0.68 (β=0.5, γ=0.2) 0.59 (β=0.5, γ=0.9)
	Noise=0.5	0.68	0.64	0.68 (β=-0.001) **0.69** (β=-0.01) **0.70** (β=-0.1)	0.68 (β=-0.001) 0.68 (β=-0.01) **0.69** (β=-0.1)	0.64 (β=-0.5, γ=0.2) 0.57 (β=-0.5, γ=0.9)	0.65 (β=0.5, γ=0.2) 0.57 (β=0.5, γ=0.9)
	Noise=0.7	0.61	0.62	0.62 (β=-0.001) 0.62 (β=-0.01) **0.66** (β=-0.1)	**0.64** (β=-0.001) **0.64**(β=-0.01) **0.65** (β=-0.1)	0.61 (β=-0.5, γ=0.2) 0.56 (β=-0.5, γ=0.9)	0.59 (β=0.5, γ=0.2) 0.61 (β=0.5, γ=0.9) 0.61 (β=0.9, γ=0.9) 0.60 (β=0.9, γ=0.2)

TABLE IV: Accuracy on 3000 samples with same parameters from Table III

		Squeezed kernel	RBF	KMNCS	KMNCS1	KMNCS2	KMNCS3
moon	Noise=0.5	0.81	0.81	**0.82** (β=-0.1) 0.81 (β=-0.01)	**0.82** (β=-0.01)	**0.82** (β=-0.5, γ=0.2)	**0.82** (β=0.5, γ=0.2)
	Noise=0.8	0.72	0.73	0.73 (β=-0.1)	0.73 (β=-0.1)	0.73 (β=-0.5, γ=0.2)	0.73 (β=0.5, γ=0.2)
	Noise=1.0	0.68	0.68	**0.69** (β=-0.1)	**0.69** (β=-0.1)	**0.69** (β=-0.5, γ=0.2)	**0.70** (β=0.5, γ=0.2)
Circle	Noise=0.5	0.65	0.64	0.65 (β=-0.1) 0.65 (β=-0.01)	0.64 (β=-0.01)	0.65 (β=-0.5, γ=0.2)	0.65 (β=0.5, γ=0.2)
	Noise=0.8	0.55	0.54	**0.56** (β=-0.1)	0.55 (β=-0.1)	0.55 (β=-0.5, γ=0.2)	0.55 (β=0.5, γ=0.2)
	Noise=1.0	0.52	0.51	**0.53** (β=-0.1)	0.51 (β=-0.1)	0.52 (β=-0.5, γ=0.2)	0.52 (β=0.5, γ=0.2)

trade-off, and it might be possible to use this to improve the uncertainty of the classification decision, which is a current aspiration in the machine learning community [53], but this is left for future work.

In addition, Table VI shows the performance of squeezed kernel, the relations (27) and (32) with positive and negative β, denoted with KMNCS1_phase, KMNCS2_phase and KMNCS3_phase, respectively. This table gives the chance to draw a comparison between kernels constructed on phase parts. In addition, we can compare the result of kernels constructed by phase and real parts of non-linear coherent states by using tables IV and VI. As evident from Table VI, the performance is almost the same with very low noise 0.1. However, when the noise ratio starts increasing, the performances of the kernels constructed by the phase parts of non-linear coherent states increases.

TABLE V: Accuracy on 3000 samples with same parameters from Table IV, but 80% training set and 20% test set.

		Squeezed kernel	RBF	KMNCS	KMNCS1	KMNCS2	KMNCS3
moon	Noise=0.5	0.79	0.78	**0.80** (β=-0.01)	**0.80** (β=-0.01)	0.79 (β=-0.5, γ=0.9)	0.79 (β=0.9, γ=0.2)
	Noise=0.8	0.69	0.70	**0.71** (β=-0.01)	**0.71** (β=-0.01)	0.70 (β=-0.5, γ=0.9)	**0.71** (β=0.9, γ=0.2)
	Noise=1.0	0.66	0.66	**0.67**(β=-0.01)	**0.67** (β=-0.01)	**0.67** (β=-0.5, γ=0.9)	**0.67** (β=0.9, γ=0.2)
Circle	Noise=0.5	0.63	0.62	0.63 (β=-0.1)	0.63 (β=-0.1)	**0.64** (β=-0.5, γ=0.1)	0.62 (β=0.9, γ=0.2)
	Noise=0.8	0.55	0.54	0.55 (β=-0.01)	0.55 (β=-0.01)	0.53 (β=-0.5, γ=0.9)	0.55 (β=0.9, γ=0.2)
	Noise=1.0	0.51	0.50	0.51 (β=-0.01)	0.51 (β=-0.01)	**0.52** (β=-0.5, γ=0.1)	**0.53** (β=0.9, γ=0.2)

TABLE VI: Accuracy on Squeezed kernel, KMNCS1_Phase, KMNCS2_Phase, and KMNCS3_Phase kernels on the make_moon and make_circles datasets with the following parameters: n_samples=100, random_state=50, y_flip=0.2, factor=0.5, n_informative=2, 5-fold cross validation, 60% training set and 40% test set. Noise effects are introduced in the data to increase difficulty of accurate classification.

		Squeezed kernel	KMNCS1_Phase	KMNCS2_Phase	KMNCS3_Phase
moon	Noise=0.5	0.83	0.83 (β=-0.001)	0.80 (β=-0.5, γ=0.9)	0.78 (β=0.5, γ=0.9)
	Noise=0.8	0.65	**0.70** (β=-0.1)	**0.68** (β=-0.5, γ=0.9)	0.65 (β=0.5, γ=0.2)
	Noise=1.0	0.58	**0.63** (β=-0.1)	**0.65** (β=-0.5, γ=0.9)	**0.62** (β=0.5, γ=0.2)
Circle	Noise=0.5	0.55	0.53 (β=-0.1)	**0.60** (β=-0.9, γ=0.9)	0.55 (β=0.5, γ=0.2)
	Noise=0.8	0.55	0.48 (β=-0.1)	0.50 (β=-0.9, γ=0.9)	0.55 (β=0.5, γ=0.2)
	Noise=1.0	0.55	0.47 (β=-0.1)	0.53 (β=-0.5, γ=0.9)	0.55 (β=0.5, γ=0.2)

VI. DISCUSSION

In general, the kernels based on the non-linear coherent states cause the margin between decision boundaries to become large, which is a major point of contrast compared with the RBF kernel, as evident in Figure 5. As previously mentioned, the hyper-parameter σ can be used to optimise the regular SVM classifier, based on the cost associated with mis-classification of elements in the 'training' set. Accordingly, the maximisation of σ leads to tight decision boundaries (so called 'hard' margins), introduced by Boser *et al.* [54]. In later work, 'hard' margins were found to fail on even slightly inseparable data. As a solution, [55] introduced a meaningful technique for the minimisation of σ that was found to enlarge the space covered by decision boundaries ('soft' margins). While this has improved the competitiveness of the SVM against other classification techniques, the trade-off is that significantly 'soft' margins fail to classify data entirely. Formally known as over-generalisation, such large decision boundaries (as produced by the Squeezed kernel) become non-representative of the data. As a result, the objective is to minimise the hyper-parameter σ while maintaining the highest possible classification accuracy. Relating this to the findings of our work, the topological structure of the decision boundaries corresponding to the kernels based on the non-linear coherent states are reminiscent of [54], and surprisingly in cases of sparse data (e.g., the sparse *make_moons* dataset), without a concern for σ.

VII. CONCLUSIONS AND REMARKS

In this paper, we mapped datasets into non-linear coherent states, as a non-linear feature space, constructed by a complex Hilbert space. We showed that the RBF kernel is recovered when data is mapped to a complex Hilbert space represented by coherent states. Therefore, non-linear coherent states can be considered as natural candidates for formulating new kernels. This idea suggests a method for obtaining a generalized kernel function which can be expressed by orthogonal polynomial functions. In addition, we experimentally indicated that kernels based on non-linear coherent states perform better, even when noise is added to the data.

It is well-known fact that mapping data into curved and pseudo-Riemannian manifolds characterizes learned latent spaces better than an Euclidean space for machine learning, see [56] and references therein. In this paper, we also studied the geometrical properties of the surface in which the kernel lives. We indicated that the feature space of a non-linear coherent state is a non-zero curved space, despite the fact that the RBF kernel lives on feature space which is flat. Therefore, inseparability of data, due to noise, can be compensated by using such kernels based on non-linear coherent states in the feature space. An experimental verification of this idea is pointed out in Tables III, IV, and V. As a result, non-linear coherent states can be considered as potential candidates for the mapping into curved and pseudo Riemannian manifolds, to harness the benefit of the flexibility of a general curved space on the machine learning methods.

More generally, this research demonstrated how quantum approaches to machine learning may prove beneficial. In fact, by using the phase part of non-linear coherent states, we introduced a wide range spectrum of feature spaces for the kernel introduced by the Refs. [37, 57]. In practical usage, machine learning applications of quantum theory have begun involving developments of physical circuitry [57]. These contributions have begun to realise the quantum processing components required to build quantum computational devices designed solely for feature-based classification. It is inspiring to think that eventually, SVM classification may (with the assistance of quantum theory) be computed substantially faster than ever before.

Acknowledgments

This work was supported by the Academy of Finland (grants 336033, 315896), Business Finland (grant 884/31/2018), and EU H2020 (grant 101016775). It was additionally supported by the Asian Office of Aerospace Research and Development (AOARD) grant: FA2386-17-1-4016 and by the German Federal Ministry for Education and Research (BMBF) under grant number 16KISQ039.

[1] S. Aaronson and A. Arkhipov, arXiv preprint arXiv:1309.7460 (2013).
[2] M. Tillmann, B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walther, Nature photonics **7**, 540 (2013).
[3] D. J. Brod, E. F. Galvão, A. Crespi, R. Osellame, N. Spagnolo, and F. Sciarrino, Advanced Photonics **1**, 034001 (2019).
[4] I. Pitowsky, The British Journal for the Philosophy of Science **45**, 95 (1994).
[5] A. Vourdas, Journal of Physics A: Mathematical and Theoretical (2019).
[6] E. Schrödinger, *Schrödinger: Centenary celebration of a polymath* (CUP Archive, 1987).
[7] Y. Tsutsumi, Funkcialaj Ekvacioj **30**, 115 (1987).
[8] S. Kochen and E. P. Specker, in *The logico-algebraic approach to quantum mechanics* (Springer, 1975), pp. 293–328.
[9] W. H. Zurek, Annalen der Physik **9**, 855 (2000).
[10] D. Girolami, A. M. Souza, V. Giovannetti, T. Tufarelli, J. G. Filgueiras, R. S. Sarthour, D. O. Soares-Pinto, I. S. Oliveira, and G. Adesso, Physical Review Letters **112**, 210401 (2014).
[11] R. Simon, E. Sudarshan, and N. Mukunda, Physical Review A **36**, 3868 (1987).
[12] C. Lorce and B. Pasquini, Physical Review D **84**, 014015 (2011).
[13] K. T. Goh, J. Kaniewski, E. Wolfe, T. Vértesi, X. Wu, Y. Cai, Y.-C. Liang, and V. Scarani, Physical Review A **97**, 022104 (2018).
[14] M. F. Pusey, L. Del Rio, and B. Meyer, arXiv preprint arXiv:1904.08699 (2019).
[15] S. Dehdashti, L. Fell, and P. Bruza, Entropy **22**, 174 (2020).
[16] S. Uprety, P. Tiwari, S. Dehdashti, L. Fell, D. Song, P. Bruza, and M. Melucci, Advances in Information Retrieval **12035**, 728 (2020).
[17] J. Shawe-Taylor, N. Cristianini, et al., *Kernel methods for pattern analysis* (Cambridge university press, 2004).
[18] D. Zelenko, C. Aone, and A. Richardella, Journal of machine learning research **3**, 1083 (2003).
[19] R. Soentpiet, *Advances in kernel methods: support vector learning* (MIT press, 1999).
[20] T. Hofmann, B. Schölkopf, and A. J. Smola, The annals of statistics pp. 1171–1220 (2008).
[21] T. Evgeniou, C. A. Micchelli, and M. Pontil, Journal of Machine Learning Research **6**, 615 (2005).
[22] C. Campbell, Neurocomputing **48**, 63 (2002).
[23] S.-i. Amari and S. Wu, Neural Networks **12**, 783 (1999).
[24] L. Wang, *Support vector machines: theory and applications*, vol. 177 (Springer Science & Business Media, 2005).
[25] W. S. Noble, Nature biotechnology **24**, 1565 (2006).
[26] B. Schölkopf, A. Smola, and K.-R. Müller, in *International conference on artificial neural networks* (Springer, 1997), pp. 583–588.
[27] I. S. Dhillon, Y. Guan, and B. Kulis, in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2004), pp. 551–556.
[28] S. Akaho, arXiv preprint cs/0609071 (2006).
[29] W. Liu, I. Park, and J. C. Principe, IEEE transactions on neural networks **20**, 1950 (2009).
[30] S. An, W. Liu, and S. Venkatesh, in *2007 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2007), pp. 1–7.
[31] Y. Cho and L. K. Saul, in *Advances in neural information processing systems* (2009), pp. 342–350.
[32] M. Belkin, S. Ma, and S. Mandal, arXiv preprint arXiv:1802.01396 (2018).
[33] M. T. Musavi, W. Ahmed, K. H. Chan, K. B. Faris, and D. M. Hummels, Neural networks **5**, 595 (1992).
[34] M. D. Buhmann, Acta numerica **9**, 1 (2000).
[35] M. J. Orr et al., *Introduction to radial basis function networks* (1996).
[36] J. M. Kübler, K. Muandet, and B. Schölkopf, Physical Review Research **1**, 033159 (2019).
[37] M. Schuld and N. Killoran, Physical review letters **122**, 040504 (2019).

[38] R. Datko, Journal of Mathematical analysis and applications **32**, 610 (1970).
[39] A. M. Gleason, Journal of mathematics and mechanics pp. 885–893 (1957).
[40] R. de Matos Filho and W. Vogel, Physical Review A **54**, 4560 (1996).
[41] V. Man'ko, G. Marmo, E. Sudarshan, and F. Zaccaria, Physica Scripta **55**, 528 (1997).
[42] S. Mancini, Physics Letters A **233**, 291 (1997).
[43] B. Roy and P. Roy, Journal of Optics B: Quantum and Semiclassical Optics **2**, 65 (2000).
[44] S. Sivakumar, Journal of Optics B: Quantum and Semiclassical Optics **2**, R61 (2000).
[45] Y. Kim, Y. Choi, D. Widemann, and T. Zohdi, arXiv preprint arXiv:2011.07727 (2020).
[46] L. K. Saul and S. T. Roweis, unpublished. Available at: http://www. cs. toronto. edu/˜ roweis/lle/publications. html (2000).
[47] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond* (MIT press, 2001).
[48] S. T. Ali, J.-P. Antoine, J.-P. Gazeau, et al., *Coherent states, wavelets and their generalizations*, vol. 3 (Springer, 2000).
[49] M. Combescure and D. Robert, *Coherent states and applications in mathematical physics* (Springer Science & Business Media, 2012).
[50] A. Mahdifar, S. Dehdashti, R. Roknizadeh, and H. Chen, Quantum Information Processing **14**, 2895 (2015).
[51] I. Bengtsson and K. Życzkowski, *Geometry of quantum states: an introduction to quantum entanglement* (Cambridge university press, 2017).
[52] M. Tchoffo, F. Migueu, M. Vubangsi, and L. Fai, Heliyon **5**, e02395 (2019).
[53] A. G. Wilson and P. Izmailov, arXiv preprint arXiv:2002.08791 (2020).
[54] B. E. Boser, I. M. Guyon, and V. N. Vapnik, in *Proceedings of the fifth annual workshop on Computational learning theory* (ACM, 1992), pp. 144–152.
[55] P. Schiilkop, C. Burgest, and V. Vapnik, in *Proceedings of the 1st international conference on knowledge discovery & data mining* (1995), pp. 252–257.
[56] A. Vlontzos, H. B. Rocha, D. Rueckert, and B. Kainz, arXiv preprint arXiv:2008.09154 (2020).
[57] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019).
[58] A. J. Smola, S. Vishwanathan, and T. Hofmann, in *AISTATS* (Citeseer, 2005).
[59] J. Zhuang, I. W. Tsang, and S. C. Hoi, in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (2011), pp. 909–917.
[60] M. Planck, Nobel lecture **2**, 1 (1920).
[61] A. K. Ekert, Physical review letters **67**, 661 (1991).
[62] A. Steane, Reports on Progress in Physics **61**, 117 (1998).
[63] E. O. Kiktenko, N. O. Pozhar, M. N. Anufriev, A. S. Trushechkin, R. R. Yunusov, Y. V. Kurochkin, A. Lvovsky, and A. Fedorov, Quantum Science and Technology **3**, 035004 (2018).
[64] L. Mandel and E. Wolf, *Optical coherence and quantum optics* (Cambridge university press, 1995).
[65] M. Harouni, Z. Harsij, J. Shen, H. Wang, Z. Xu, B. Mirza, and H. Chen, Quantum Information & Computation **16**, 1365 (2016).
[66] E. P. Wigner, in *Part I: Physical Chemistry. Part II: Solid State Physics* (Springer, 1997), pp. 110–120.
[67] C. K. Zachos, D. B. Fairlie, and T. L. Curtright, *Quantum mechanics in phase space: an overview with selected papers* (World Scientific, 2005).
[68] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, Physics Reports **506**, 1 (2011).
[69] R. Rosipal and L. J. Trejo, Journal of machine learning research **2**, 97 (2001).
[70] S. Dehdashti, M. B. Harouni, B. Mirza, and H. Chen, Physical Review A **91**, 022116 (2015).
[71] S. Dehdashti, A. Mahdifar, M. B. Harouni, and R. Roknizadeh, Annals of Physics **334**, 321 (2013).
[72] S. Dehdashti, R. Li, J. Liu, F. Yu, and H. Chen, AIP Advances **5**, 067165 (2015).
[73] A. W. Harrow, A. Hassidim, and S. Lloyd, Physical review letters **103**, 150502 (2009).
[74] O. Chapelle, B. Scholkopf, and A. Zien, IEEE Transactions on Neural Networks **20**, 542 (2009).
[75] P. Rebentrost, M. Mohseni, and S. Lloyd, Physical review letters **113**, 130503 (2014).
[76] S. Aaronson, Nature Physics **11**, 291 (2015).
[77] J. J. Sakurai and E. D. Commins, *Modern quantum mechanics, revised edition* (1995).
[78] S. K. Zhou and R. Chellappa, IEEE transactions on pattern analysis and machine intelligence **28**, 917 (2006).
[79] G. Wahba, Proceedings of the National Academy of Sciences **99**, 16524 (2002).
[80] S. Dehdashti, A. Mahdifar, and R. Roknizadeh, International Journal of Geometric Methods in Modern Physics **10**, 1350014 (2013).
[81] S. Dehdashti, C. Moreira, A. K. Obeid, and P. Bruza, arXiv preprint arXiv:2006.02904 (2020).
[82] `https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_moons.html#sklearn.datasets.make_moons`
[83] `https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_circles.html#sklearn.datasets.make_circles`
[84] `https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_classification.html#sklearn.datasets.make_classification`