Enhancing Quantum Machine Learning: The Power of Non-Linear Optical Reproducing Kernels

Anonymous authors
Paper under double-blind review

Abstract

Amidst the array of quantum machine learning algorithms, the quantum kernel method has emerged as a focal point, primarily owing to its compatibility with noisy intermediate-scale quantum devices and its promise to achieve quantum advantage. This method operates by nonlinearly transforming data into feature space constructed with quantum states, enabling classification and regression tasks. In this study, we present a novel feature space constructed using Kerr coherent states, which generalize $su(2)$, $su(1,1)$ coherent states, and squeezed states. Notably, the feature space exhibits constant curvature, comprising both spherical and hyperbolic geometries, depending on the sign of the Kerr parameter. Remarkably, the physical parameters associated with the coherent states, enable control over the curvature of the feature space. Our study employs Kerr kernels derived from encoding data into the phase and amplitude of Kerr coherent states. We analyze various datasets ranging from Moon to breast cancer diagnostics. Our findings demonstrate the robustness of Kerr coherent states, attributed to their flexibility in accommodating different hyperparameters, thereby offering superior performance across noisy datasets and hardware setups.

1 Introduction

Quantum machine learning aims to exploit the unique capabilities of quantum systems to enhance machine learning tasks, either through computational speedups or improved expressive power Biamonte et al. (2017). Early developments in the field focused primarily on algorithms based on the Harrow–Hassidim–Lloyd (HHL) matrix inversion algorithm Harrow et al. (2009), which serves as a fundamental primitive for approaches such as quantum support vector machines (SVMs) Rebentrost et al. (2014), quantum principal component analysis (PCA) Lloyd et al. (2014), and quantum causal inference Goel et al. (2026). Additional efforts explored the extraction of topological properties from data using quantum resources Lloyd et al. (2016).

However, many of these early proposals relied on strong assumptions regarding data structure, including sparsity conditions and access to Quantum Random Access Memory (QRAM) Giovannetti et al. (2008), while also requiring fault-tolerant quantum hardware beyond current technological capabilities. More recent research has therefore shifted toward algorithms compatible with Noisy Intermediate-Scale Quantum (NISQ) devices Preskill (2018). Prominent approaches in this setting include variational quantum algorithms such as the Variational Quantum Eigensolver (VQE) Peters & Schuld (2023) and the Quantum Approximate Optimization Algorithm (QAOA) Farhi et al. (2014); Zhou et al. (2020), which combine quantum and classical resources to address problems in materials science and combinatorial optimization. Furthermore, Quantum Neural Networks have emerged as promising candidates for supervised learning tasks Abbas et al. (2021). Among these developments, quantum kernels have attracted particular attention as a flexible and experimentally accessible framework for quantum-enhanced learning Havlíček et al. (2019).

Quantum kernels constitute a promising class of nonlinear data-analysis techniques, as they enable the embedding of classical data into quantum feature spaces. Classical kernel methods already play a central role across numerous scientific disciplines, including biology, chemistry, and physics Shawe-Taylor & Cristianini (2004); Hofmann et al. (2008); Schölkopf et al. (2004); Ralaivola et al. (2005); Karniadakis et al. (2021). Many kernel-based algorithms, such as Ridge Regression and kernel Principal Component Analysis (KPCA), rely

on convex optimization and can therefore be efficiently implemented on classical hardware. Kernel methods employ positive-definite symmetric functions to map data into a Reproducing Kernel Hilbert Space (RKHS), in which functions can be represented and reconstructed Perelomov (1977). Quantum kernels have also been theoretically shown to provide exponential speedups in algorithms for causal estimands. Such advantages are particularly relevant in quantum algorithms where the data remain encoded within the quantum system, thereby avoiding the bottleneck associated with repeated classical information readout extractions Goel et al. (2026). Additionally, group-symmetric quantum kernels have been shown to avoid exponential concentration phenomena while retaining trainability independently of system size Henderson et al. (2025). Such properties make them particularly attractive for scalable near-term quantum machine learning.

Quantum kernels are based on the observation that classical data $\mathbf{x} \in \mathbb{R}^n$ can be embedded into quantum states through a feature map of the form

$$\rho(\mathbf{x}) = \hat{U}(\mathbf{x}) \, |0\rangle\langle 0| \, \hat{U}^\dagger(\mathbf{x}),$$

where $\hat{U}(\mathbf{x})$ denotes a data-dependent quantum operation. Thus, encoding two data points $\mathbf{x}$ and $\mathbf{w}$ into the states $\rho(\mathbf{x})$ and $\rho(\mathbf{w})$ allows to quantify the similarity between the data points by calculating the overlap between their corresponding quantum states

$$K(\mathbf{x}, \mathbf{w}) = \mathrm{Tr}[\rho(\mathbf{x})\rho(\mathbf{w})],$$

yielding a kernel function $K(\mathbf{x}, \mathbf{w})$. A paradigmatic example in classical machine learning is the radial basis function (RBF) kernel,

$$K(x, w) = \exp\left(-\frac{|x - w|^2}{2\sigma^2}\right),$$

where the parameter σ plays a crucial role in determining the decision boundary Orr et al. (1996); Musavi et al. (1992); Buhmann (2000). In machine learning, the degree of freedom associated with σ is usually referred to as the bandwidth. Notably, the RBF kernel can also be interpreted as the overlap between coherent quantum states Kübler et al. (2019). In this construction, the feature map, $U(\mathbf{x})$, is generated through displacement operators

$$\hat{D}(x) = \exp\left(x\hat{a}^\dagger - x\hat{a}\right),$$

where $\hat{a}^\dagger$ and $\hat{a}$ denote the bosonic creation and annihilation operators.

Coherent states satisfy the resolution of the identity

$$\int d\mu(\alpha) \, |\alpha\rangle\langle\alpha| = \mathbb{I},$$

with $\alpha \in \mathbb{C}$. Consequently, any quantum state $|\psi\rangle$ can be expressed as a superposition of coherent states:

$$\int d\mu(\alpha) \, |\alpha\rangle \, \langle\alpha|\psi\rangle = |\psi\rangle \, .$$

Defining the coherent-state wavefunction as

$$\psi(\alpha^*) \equiv \langle\alpha|\psi\rangle \, ,$$

establishes a one-to-one correspondence between quantum states and analytic functions on the complex plane Perelomov (1977). This immediately leads to the reproducing-kernel property

$$\psi(\beta) = \int d\mu(\alpha) \, K(\beta, \alpha)\psi(\alpha),$$

where the kernel is given by

$$K(\alpha, \beta) = \langle\alpha|\beta\rangle \, .$$

Several proposals have explored embedding classical data into coherent and squeezed states through either phase or amplitude encodings Schuld & Killoran (2019); Kübler et al. (2019); Tiwari et al. (2022). In

particular, phase encodings in squeezed states have demonstrated improved classification performance in certain settings, despite the absence of tunable kernel hyperparameters Schuld & Killoran (2019).

In supervised learning, one is given labeled pairs $(\mathbf{x}, y) \in \mathcal{D} \times \{0, 1\}$ and seeks to learn a predictive function capable of generalizing to unseen data. A fundamental example is binary classification, where the dataset consists of feature vectors $\mathbf{x}_i \in \mathcal{D}$ together with binary labels y_i. Unless explicitly stated otherwise, we assume throughout this work that $\mathcal{D} = \mathbb{R}^N$ for some $N \in \mathbb{N}$.

One of the most widely used supervised-learning models is the Support Vector Machine (SVM), which constructs classification boundaries from inner products between data points Cortes & Vapnik (1995); Boser et al. (1992). When the data are not linearly separable, the kernel trick allows the data to be implicitly mapped into a higher-dimensional feature space through a nonlinear feature map

$$\Phi : \mathbf{x} \to \Phi(\mathbf{x}),$$

such that similarities are computed through

$$K_{ij} = \langle \Phi(x_i) | \Phi(x_j) \rangle \, .$$

Different kernels therefore correspond to different geometries and similarity measures in feature space Aizerman et al. (1964); Aronszajn (1950). Curved feature spaces can reshape similarity relations between samples, potentially enabling simpler linear decision boundaries in the induced Hilbert space Viladomat Jasso et al. (2023).

In this work, we introduce quantum kernels based on Kerr coherent states (KCSs), which generalize both $su(2)$ and $su(1,1)$ coherent states and recover squeezed states for specific values of the Kerr parameter. Unlike conventional squeezed kernels, the proposed Kerr kernels possess tunable physical hyperparameters that directly control the geometry of the induced feature space and the resulting classification boundary. In particular, the Kerr parameter acts as a bandwidth, while together with the effective Hilbert-space dimension it determines the curvature and structure of the decision boundary (see Appendix B for details).

A central result of this work is that Kerr coherent states naturally induce curved feature-space geometries. For positive Kerr parameters, the associated manifold acquires a hyperbolic geometry, whereas negative Kerr parameters generate spherical feature spaces. In this context, the Ricci scalar becomes directly related to experimentally controllable physical parameters of the Kerr coherent states. This connection is particularly relevant from a machine-learning perspective, since embeddings into curved and pseudo-Riemannian manifolds have been shown to provide more expressive latent representations than Euclidean embeddings in several learning scenarios Arvanitidis et al. (2018). Motivated by these advantages, a wide range of geometric learning approaches such as Lie group machine learning Lu & Li (2020), manifold optimization Boumal (2023), and machine learning in hyperbolic spaces Prakash et al. (2024); Hong et al. (2023); Qu & Zou (2022); Fang et al. (2023); Mettes et al. (2024); Gao et al. (2022); Atigh et al. (2022); Yang et al. (2024; 2023); Mauran et al. (2023); Fang et al. (2021) have been recently explored.

Kerr-inspired learning protocols can be realized across multiple physical platforms. For instance, Kerr kernels have previously been studied in superconducting quantum-circuit architectures, where the nonlinear response arises from the nonlinear inductance of SQUID-based elements. In such approaches, the kernel is implemented by engineering tunable Kerr nonlinearities in coplanar microwave resonators. The tuning is achieved through the flux bias of a DC SQUID Wood et al. (2024); Yamaji et al. (2022).

By contrast, the framework introduced in this work is fully applicable in current photonic platforms, relying on amplitude and phase embedding of the classical data being fed directly at the state-preparation protocol. Moreover, the kernel evaluation is recovered through direct projective measurements, eliminating the need for QRAM. This NISQ-compatible method provides a comparatively simple and scalable route towards experimental implementation. Additionally, this perspective opens new possibilities for quantum machine learning on curved manifolds using experimentally accessible optical systems.

To evaluate the practical performance of the proposed kernels, we benchmark several Kerr-coherent-state embeddings within support vector machine classifiers across a variety of synthetic and real-world datasets. The synthetic benchmarks include the Moons, Circles, and periodic classification datasets, while the

real-world tasks comprise the Framingham Chronic Heart Disease (CHD) dataset Janosi et al. (1989) and Fashion-MNIST Xiao et al. (2017).

Our results demonstrate that the optical phase and amplitudes and the nonlinearity Kerr parameter play a central role in determining classification performance. Remarkably, this same parameter directly controls the curvature of the induced feature-space manifold, thereby establishing a direct connection between geometric properties of the embedding and learning performance.

More broadly, the proposed framework enables a continuous interpolation between quantum and classical regimes through physically controllable parameters alone. This perspective naturally connects Kerr-inspired quantum kernels with analog and physics-inspired computation, where engineered physical interactions are directly exploited to perform computational learning tasks.

2 Related Work

Building upon the principle of encoding classical data into quantum states, Schuld and Killoran Schuld & Killoran (2019) introduced the notion of a *quantum feature map*. They observed that any quantum encoding of the form $x \mapsto |\psi(x)\rangle$, where $|\psi(x)\rangle$ is a quantum state in a complex Hilbert space $\mathcal{H}$, inherently defines a feature map. As such, it naturally induces a kernel function through the inner product $K(x, x') = \langle\psi(x)|\psi(x')\rangle$. This conceptual bridge enables the integration of established kernel methods with quantum computing frameworks, thereby enriching the toolbox for quantum-enhanced machine learning.

Data Encoding Schemes: Qubit Approach

There are several methods by which we can encode data using Qubits, which summarize as follows:

Computational Basis Encoding. By directly mapping classical bits into qubits using the computational basis states $|0\rangle$ and $|1\rangle$, the information is stored in the amplitudes of these states. Therefore, the feature encode data $\varphi : \mathcal{B}^n \to \mathcal{H}$ is defined as

$$\varphi : (b_1, b_2, \cdots, b_n) \to |i\rangle = \bigotimes_{k=0}^{n} |b_k\rangle\,,\ b_k \in \mathcal{B}$$

where $\mathcal{B} = \{0, 1\}$ is the Boolean set and $\mathcal{H}$ is the 2^n-dimensional Hilbert space; the associated kernel is also given by

$$K(i, j) = \langle i|j\rangle = \delta_{ij}$$

where δ_{ij} is the Kronecker delta.

Superposition Encoding: Information can be encoded by putting qubits in a superposition of classical states, thereby encoding multiple possibilities simultaneously. Similar to the previous method, the kernel is the Kronecker delta.

Angle Encoding. Angle encoding utilizes rotation gates to represent classical information. The classical data is mapped by defining the rotation angles gates,

$$|\mathbf{x}\rangle = \bigotimes_{i=1}^{n} R(\mathbf{x}_i)\,|0^n\rangle$$

where R can be one of the rotation matrices R_x, R_y, or R_z. The associated kernel is given by

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{n} \cos\left(\frac{x_i - x_i'}{2}\right)$$

Unlike the above-mentioned encodings in which the data is restricted to binary values, angle encoding allows data to take the form of real, floating point numbers.

Amplitude Encoding. Amplitude encoding maps a complex vector of length N into the amplitudes of an n-qubit quantum state, with $n = \log_2 N$:

$$|\mathbf{x}\rangle = \sum_{i=1}^{N} x_i \, |i\rangle$$

As amplitude encoding enables an exponential reduction in the number of required bits, the encoding method potentially allows for computational speed-up. However, it may also demand a significant number of gates. The associated kernel is given by

$$K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$$

As suggested in Ref. Swaminathan & Akgün (2025), by encoding the data in the following manner, one can recover a *radial basis function* (RBF) and *Laplacian* kernels using qubits. Given classical data $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T$, one can define the quantum feature maps

$$\begin{aligned}
|\psi(\mathbf{x})\rangle &= \frac{1}{\sqrt{D}} \sum_{j=1}^{D} \left[\cos\left(w_j^T \mathbf{x}\right) |2j-2\rangle + \sin\left(w_j^T \mathbf{x}\right) |2j-1\rangle \right], \\
|\psi(\mathbf{x})\rangle &= \frac{1}{\sqrt{D}} \sum_{j=1}^{D} \Big[\cos\left(w_j^T \mathbf{x} + \alpha_j\right) |2j-2\rangle \\
&\quad + \sin\left(w_j^T \mathbf{x} + \alpha_j\right) |2j-1\rangle \Big],
\end{aligned}$$

where $\lceil \log_2(2D) \rceil$ determines the number of qubits required. For recovering RBF and Laplace kernels, respectively, the vectors w_j are drawn independently from the normal distribution $\mathcal{N}\left(0, \sigma^{-2} I\right)$ and Cauchy distribution $\mathcal{C}\left(0, \alpha^{-1} I\right)$. Parameters α_j are independent samples from the uniform distribution $\mathcal{U}(0, 2\pi)$.

Data Encoding Schemes: CV approch

Coherent state encoding. An important example of a quantum state that naturally defines such a kernel is the *coherent state*, which is defined as a superposition of number states $|n\rangle$:

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

where the number states satisfy the orthonormality condition $\langle n|m\rangle = \delta_{n,m}$, with $\delta_{n,m}$ denoting the Kronecker delta. The RBF kernel is recovered as an inner product between two coherent states Kübler et al. (2019); Chatterjee & Yu (2017).

Holomorphic functions encoding. Rather than representing quantum states as infinite countable vectors, as in the Fock state formalism, they can be described as holomorphic functions via a transformation Henderson et al. (2024),

$$|n\rangle \longleftrightarrow \left(z \mapsto \frac{z^n}{\sqrt{n!}} \right)$$

and annihilation operators acting on the Hilbert space of a quantum state are mapped to differential operators, as

$$\hat{a}^\dagger \longleftrightarrow z\times, \qquad \hat{a} \longleftrightarrow \partial_z$$

where $z\times$ acts on a holomorphic function by multiplying it by z and ∂_z takes the partial derivative of it with respect to z. Hence a particular quantum state, decomposed into its Fock basis as $|\psi\rangle = \sum_n c_n |n\rangle$ can be transformed as,

$$|\psi\rangle = \sum_n c_n |n\rangle \longleftrightarrow F_\psi^\star(z) = \sum_n \frac{c_n z^n}{\sqrt{n!}}$$

which is called the stellar function of the state $|\psi\rangle$. Stellar functions live in the Segal-Bargmann (SB) space. The stellar functions of finite rank n can be written as a Gaussian and polynomial function,

$$F_{\psi}^{\star}(z) = G(z)P(z) = e^{-a/z^2+bz+c} \sum_{j}^{n} \beta^j z^j$$

for $a, b, c, \beta_j \in \mathbb{C}$. Common examples of functions with zero stellar rank include vacuum states, coherent states, squeezed states, and two-mode squeezed states. Fock states with particle number n possess a stellar rank of n. Notably, stellar rank serves as a measure of non-Gaussianity, with an important property being its invariance under Gaussian operations. This approach enables data to be encoded into continuous variable quantum states using holomorphic functions, thereby establishing a method for data encoding using the following map:

$$x_i \mapsto \Phi(x_i) := F_{x_i}^{\star}(z)$$

Therefore, a Kernel can be defined as:

$$\left\langle F_i^{\star} \middle| F_j^{\star} \right\rangle = \frac{1}{\pi^m} \int_{\mathbf{z}\in\mathbb{C}^m} d^2\mathbf{z}\ e^{-|\mathbf{z}|} F_i^{\star}(\mathbf{z}) F_j^{\star}(\mathbf{z})$$

Squeezed Vacuum State Encoding. Several studies (see, e.g., Kübler et al. (2019); Zhong et al. (2020; 2021); Xia et al. (2021); Zhou et al. (2022); Li et al. (2022)) suggest that data can be encoded in either the amplitude or the phase of a squeezed vacuum state, defined as

$$|z\rangle = \frac{1}{\sqrt{\cosh r}} \sum_{n=0}^{\infty} \frac{\sqrt{(2n)!}}{2^n n!} \left[-e^{i\varphi} \tanh r\right]^n |2n\rangle\,,$$

where $z = r\, e^{i\varphi}$ with $r \in \mathbb{R}^+$ the squeezing parameter and φ the squeezing phase.

For *phase encoding*, the corresponding kernel takes the form

$$K_{\text{phase}}(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{N} \frac{\operatorname{sech} r}{\sqrt{1 - e^{i(x_i' - x_i)} \tanh^2 r}},$$

where x_i and x_i' denote the phase-encoded features.

For *amplitude encoding*, the kernel is given by

$$K_{\text{amp}}(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{N} \cosh^{-1/2}(x_i - x_i')\,,$$

where x_i and x_i' now denote the amplitude-encoded features.

Barut–Girardello Coherent State Encoding. Consider the f-deformed annihilation and creation operators

$$\hat{A} = \hat{a}\ f(\hat{n}), \qquad \hat{A}^\dagger = f^\dagger(\hat{n})\ \hat{a}^\dagger,$$

where $f(\hat{n})$ is known as the *deformation function*, governing the non-linear properties of the quantum system.

A Barut–Girardello (BG) coherent state is defined as the right-hand eigenstate of the deformed annihilation operator Ali et al. (2000); Combescure & Robert (2012):

$$\hat{A}\ |\alpha\rangle_f = \alpha\ |\alpha\rangle_f\,, \tag{1}$$

where $\alpha \in \mathbb{C}$ is the eigenvalue. The explicit form of the BG coherent states in the Fock basis is given by

$$|\alpha\rangle_f = \frac{1}{\sqrt{\mathcal{N}}} \sum_{n=0}^{M} \frac{\alpha^n}{\sqrt{n!\,[f(n)]!}} \; |n\rangle \,,$$

where M may be finite or infinite (corresponding to finite- or infinite-dimensional Hilbert spaces), and $[f(n)]! = \prod_{i=1}^{n} f(i)$ with $[f(0)]! = 1$. The normalization factor is

$$\mathcal{N} = \sum_{n=0}^{M} \frac{|\alpha|^{2n}}{n!\,[f(n)]!}. \tag{2}$$

As suggested in Tiwari et al. (2022), both the amplitude and phase encoding schemes can be applied in the BG coherent states. By mapping classical data into the BG coherent states, the associated kernel function is obtained from the inner product of two non-linear coherent states:

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

where

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

3 Theoretical aspects of Kerr Coherent state's Kernels

In the following, we introduce the Kerr coherent states as the feature space and study their mathematical characters.

3.1 Kerr Feature Space

A Kerr coherent state is defined as

$$|\alpha; \lambda, j\rangle = e^{\alpha \hat{A}^\dagger - \alpha^* \hat{A}} \, |0\rangle \,, \;\; \alpha = re^{i\phi} \in \mathbb{C} \tag{3}$$

The Kerr annihilation operators are respectively defined by $\hat{A} = \sqrt{\frac{\lambda}{2}} \hat{a} \sqrt{2j - 1 + \hat{n}}$ and $\hat{A} = \sqrt{\frac{|\lambda|}{2}} \hat{a} \sqrt{2j + 1 - \hat{n}}$, for positive and negative sign of the Kerr parameter $\lambda \in \mathbb{R}$; $\hat{a}$ and $\hat{a}^\dagger$ are annihilation and creation operator, and $\hat{n} = \hat{a}^\dagger \hat{a}$ is the number operator; the Kerr creation operator $\hat{A}^\dagger$ is the conjugate transpose of the Kerr annihilation operator $\hat{A}$. For the positive value of the Kerr parameter, the Kerr coherent state can be expanded in terms of the Fock state basis as

$$\begin{aligned} \left|\alpha; \lambda^+, j\right\rangle &= \cosh^{-2j}\left[\sqrt{\frac{\lambda}{2}}|\alpha|\right] \\ &\times \sum_{n=0}^{\infty} \sqrt{\frac{\Gamma(2j+n)}{\Gamma(2j) n!}} e^{-in\phi} \tanh^n\left[\sqrt{\frac{\lambda}{2}}|\alpha|\right] |n\rangle \,, \end{aligned} \tag{4}$$

where j can be an integer or half-integer Dehdashti et al. (2025). It's important to note that when λ is set to 2, the $su(1,1)$-coherent state is obtained. Furthermore, when λ is 2 and j is $1/2$, the result is a squeezed state.

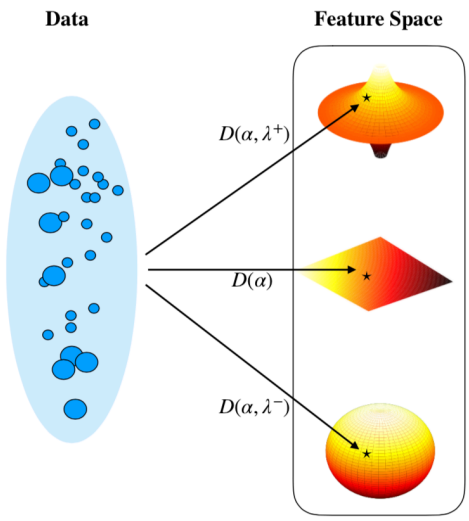

Figure 1: Data by using the displacement operator is transferred in the feature space, which can be pseudosphere, flat and sphere according to the sign of the Kerr parameter.

For negative values of the Kerr parameter λ, the Kerr coherent state $|\alpha; \lambda^-, j\rangle$ is given by

$$\begin{aligned} |\alpha; \lambda^-, j\rangle &= \cos^{2j}\left[\sqrt{\frac{|\lambda|}{2}}|\alpha|\right] \\ &\times \sum_{n=0}^{2j} \sqrt{\frac{(2j)!}{(2j-n)!n!}} e^{-in\phi} \tan^n\left[\sqrt{\frac{|\lambda|}{2}}|\alpha|\right] |n\rangle\,, \end{aligned} \tag{5}$$

Note that for the $\lambda = -2$, the above Kerr coherent state represents a $su(2)$-coherent state.

It's important to note that even though the Hilbert space representation of the Kerr coherent states, whether in the negative or positive case, involves more than two dimensions, these states exist on a surface with distinct topological characteristics that depend on the sign of the Kerr parameter. In fact, the feature space created by a positive Kerr coherent state forms a pseudosphere with consistently negative curvature as Fig 1 schematically illustrates. For mathematical detail see Appendix A.

The inner product between two Kerr coherent states, for $\lambda \in \mathbb{R}^+$, is given by

$$K(\alpha_1, \alpha_2) = \frac{\left(\operatorname{sech}^2\left[\sqrt{\frac{\lambda}{2}}r_1\right]\operatorname{sech}^2\left[\sqrt{\frac{\lambda}{2}}r_2\right]\right)^j}{\left[1 - e^{i(\phi_1-\phi_2)}\tanh\left(\sqrt{\frac{\lambda}{2}}r_1\right)\tanh\left(\sqrt{\frac{\lambda}{2}}r_2\right)\right]^{2j}}. \tag{6}$$

where $\alpha_j = r_j \exp(i\phi_j)$, with $j = 1, 2$. For $\lambda \in \mathbb{R}^-$, the inner product is given by

$$K(\alpha_1, \alpha_2) = \frac{\left[1 + e^{i(\phi_1-\phi_2)}\tan\left[\sqrt{\frac{|\lambda|}{2}}r_1\right]\tan\left[\sqrt{\frac{|\lambda|}{2}}r_2\right]\right]^{2j}}{\left(\sec^2\left[\sqrt{\frac{|\lambda|}{2}}r\right]\sec^2\left[\sqrt{\frac{|\lambda|}{2}}r'\right]\right)^j}. \tag{7}$$

When it comes to encoding data within quantum states, we have two distinct methods at our disposal. The first method involves encoding the data into the phase of the quantum states, which is commonly

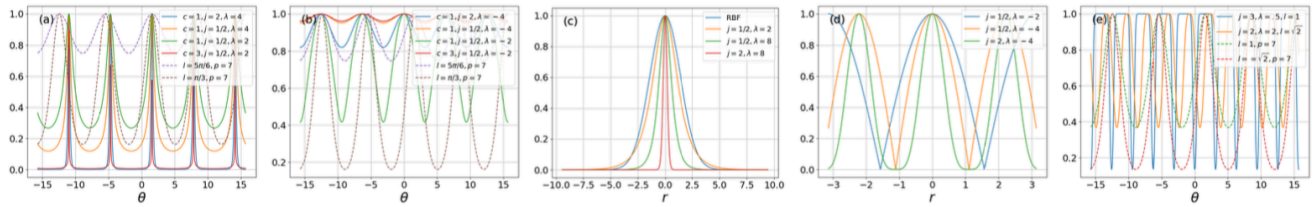

Figure 2: The Kerr phase kernel with positive Kerr parameters, represented by Eq. (6, is depicted in the plot (a) across various parameter values and compared with the exponential sine-squared kernel (Eq. (8)). Plot (b) illustrates the Kerr phase kernel with negative Kerr parameters, i.e., Eq. (7). Plots (c) and (d) showcase the positive and negative amplitude Kernels, specifically Eqs. (9) and (10), respectively. Finally, plot (e) demonstrates the exponential sine-squared kernel (Eq. (8)) and the exponential cosine kernel (Eq. (11)) for different parameters.

referred to as phase encoding. Alternatively, the second method allows us to map the data to the amplitude of the quantum state, known as amplitude encoding. In the following, we consider both cases in detail.

Phase encoding: In the context of phase encoding, we can take an N-dimensional dataset $\{x_m\}$ and represent the data by utilizing the phase attributes of N Kerr coherent states. The specific choices for the amplitude $|\alpha| = c$, Kerr parameter λ, and j defined as integer and half-integer serve as hyperparameters for this encoding process, i.e., $K(x_m, x_{m'}) = \langle c, x_m; \lambda | c, x_{m'}; \lambda \rangle$ using relations (6) and (7). We can call it the Kerr phase kernel. It's worth highlighting that the Kerr phase kernel can be considered an extension of the well-known squeezed phase kernel, a widely employed concept in quantum machine learning Schuld & Petruccione (2021); Li et al. (2022) and exhibits analogous characteristics to the exponential sine squared (ESS) kernel commonly encountered in machine learning:

$$K_{ess}(x_m - x_{m'}) = \exp\left[-\frac{2}{l^2}\sin^2\left(\frac{\pi}{p}|x_m - x_{m'}|\right)\right] \tag{8}$$

where l and p are hyperparameters of the length scale and the period, respectively Li et al. (2022). These kinds of kernels are suitable for data that has features of periodicity, such as weather forecasting data and the stock market data.

Amplitude encoding: Another way to encode into the Kerr coherent state is to encode the data into the amplitude, i.e., $\phi : x_m \to |x_m, \theta; \lambda\rangle$. In such case, θ, λ and j can be hyperparameters. The direct calculation shows that whenever we set $\theta = 0$, we can write the Kernel for the positive Kerr parameters as follows:

$$K(x_m, x_{m'}) = \cosh^{-2j}\left[\sqrt{\frac{\lambda}{2}}|x_m - x_{m'}|\right], \tag{9}$$

which is generalization of the squeezed amplitude kernel, that is $K_{sa} = \cosh^{-1/2}[|x_m - x_{m'}|]$. For the negative Kerr parameters, the Kernel is given by

$$K(x_m, x_{m'}) = \cos^{2j}\left[\sqrt{\frac{\lambda}{2}}|x_m - x_{m'}|\right]. \tag{10}$$

It's worth noting that the positive Kerr kernel represents a broader concept encompassing the hyperbolic kernels Vedaldi & Zisserman (2012); Wilson et al. (2014), while the negative Kerr kernel extends the idea of the squared cosine kernel Schuld (2021); Fang et al. (2021).

In addition, following the construction principles of Kernels Hofmann et al. (2008), we propose the corresponding quantum counterpart of the aforementioned Kernel (8) as follows:

$$K_{qecs}(x_m - x_{m'}) = \exp\left[-\frac{2}{l^2}\cos^{2j}\left(\sqrt{\frac{|\lambda|}{2}}|x_m - x_{m'}|\right)\right] \tag{11}$$

which we refer to as the quantum exponential cosine kernel. Figure 2 presents a comparative analysis of the Kerr phase kernel, Kerr amplitude kernels with varying parameters j and λ, and exponential sine-squared and exponential cosine kernels. Upon comparison with the squeezing phase, RBF, and exponential sine-squared kernels, it is evident that the hyperparameters of the Kerr phase kernel offer greater control over the distribution boundaries. This heightened flexibility renders the Kerr phase kernel a more versatile option for data classification.

4 Kernels implementation

The Kerr coherent states are categorized as a type of coherent states known as "non-linear coherent states." A non-linear coherent state is characterized by the utilization of a deformed displacement operator, wherein deformed annihilation and creation operators, denoted as $\hat{A} = f(\hat{n})\hat{a}$ and $\hat{A} = \hat{a}^\dagger f^\dagger(\hat{n})$, respectively, substitute the conventional ones Dodonov & Man'ko (2003). A non-linear coherent state can be produced through the application of laser cooling and ion trapping techniques, as discussed in de Matos Filho & Vogel (1996). When a trapped ion is subjected to laser fields, its internal and external degrees of freedom become interconnected through the interaction. Therefore, through careful adjustment of the laser fields, one can manipulate the quantized vibrational motion of the ion within the trap potential. The notable advantage of the exceedingly weak coupling between the vibrational modes and the external environment lies in its potential to facilitate the preparation and observation of non-linear coherent states with a remarkable level of stability, see appendix de Matos Filho & Vogel (1996); Vogel & de Matos Filho (1995).

Furthermore, we propose an approach for generating Kerr coherent states. This method involves the interaction of a two-level atom with a single-mode quantized cavity field, utilizing an intensity-dependent Jaynes–Cummings model, and concurrently exposing it to a strong external classical field. The temporal evolution of the system initially results in the formation of a superposition of Kerr coherent states. Depending on the initial states of the atom and the field, which can be suitably prepared, and considering the conditions under which the atom is detected (whether in the excited or ground state) following the interaction, the Kerr coherent state will collapse to the desired class of Kerr coherent states, see appendix B.1 for details.

Moreover, a proposed optical system facilitates the direct observation of a classical analog for the displacement of Fock states Perez-Leija et al. (2010). This system involves a photonic lattice comprising evanescently coupled waveguides Christodoulides et al. (2003), with a well-designed distribution of coupling between adjacent guides. Within these Glauber-Fock photonic lattices, each excited waveguide serves as a representation of a Fock state, and the spatial evolution of the light field corresponds to the probability amplitudes of the Kerr coherent state. Consequently, the emergence of these fundamental states and the underlying displacement process can be visually apprehended, see appendix B.2.

5 Benchmark Results

This section presents the benchmark protocol used to characterize the kernels in Eqs. (6), (7), (9), and (10). We first describe the evaluation procedure in Section 5.1. We then characterize the behavior of the kernels as a function of their hyperparameters (λ, j) in Section 5.1.1. The resulting classification performance on synthetic and real-world benchmarks is discussed in Section 5.2. Additional comparisons with the kernels of Eqs. (11) and (8) are presented in Section 5.2.3.

5.1 Benchmark procedure

To assess the classification capabilities of the proposed kernels, a systematic benchmarking procedure was carried out on both artificial and real-world data sets. For every benchmark instance, an SVM classifier was employed together with a five-fold cross-validation scheme to obtain robust performance estimates. Each kernel was independently evaluated while varying the hyperparameters λ and j, such that every combination of kernel type and parameter values corresponds to a new training and validation cycle.

The encoding parameters were fixed throughout the benchmarks. For phase-encoding kernels in Eqs. (6) and (7), the parameter α was set to $\alpha = 1$, and all data points were rescaled to the interval $[0, 3]$ so that the encoded values remained within the relevant phase periodicity. For amplitude-encoding kernels, the parameter θ was fixed to $\theta = 0$.

5.1.1 Kernel behavior characterization.

Synthetic data sets were generated using the utilities available in the Python package `scikit-learn`, namely `make_blobs`, `make_circles`, `make_classification`, and `make_moons`. These benchmarks were selected to probe different classification challenges such as non-linearly separable classes, clustered distributions, and higher-dimensional structured data. For more details about the synthetic datasets construction see Appendix C.

The accuracy heatmaps obtained for the synthetic benchmarks are shown in Fig. 3. Columns (a)–(d) correspond to the artificially generated data sets: Circles, Double Layers, Hypercubes, and Moons, respectively. Rows (I) and (II) display the λ^- kernels using amplitude and phase embeddings, respectively, while rows (III) and (IV) show the corresponding λ^+ kernels for amplitude and phase embeddings.

A first notable result is that each kernel exhibits a characteristic performance pattern that remains largely consistent across all artificial data sets. While the detailed location of maxima and minima varies depending on the geometry and noise realization of each benchmark, the global behavior of each kernel type is preserved. This indicates that the dependence on the hyperparameters λ and j is primarily determined by the kernel structure itself rather than by the specific data set.

For the λ^- amplitude-encoding kernel, the highest accuracies are generally obtained for small values of $|\lambda|$. As $|\lambda|$ increases beyond approximately 5, a rapid deterioration in classification accuracy is observed. This trend remains stable for all tested values of j and for every synthetic benchmark considered. Such behavior suggests that excessively strong negative deformation reduces the kernel's ability to resolve relevant feature similarities.

In contrast, the λ^- phase-encoding kernel displays a markedly periodic structure in the accuracy landscape. Alternating regions of high and low performance appear as a function of λ, revealing an oscillatory dependence induced by the phase embedding. The sharpness and contrast of these valleys depend on the chosen data set. In particular, for the Double Layers benchmark, the relative position of maxima and minima appears inverted when compared with the remaining data sets, yet the periodic behavior remains the same as with the other datasets.

The λ^+ amplitude-encoding kernel shows a clear preference for lower values of j. In particular, for $j = 1$, comparatively large values of $|\lambda|$ can still maintain strong classification accuracy before performance substantially degrades. As j increases, this tolerance progressively diminishes, restricting the parameter region of high performance. The same qualitative tendency is observed across all synthetic benchmarks.

Similarly, the λ^+ phase-encoding kernel follows the same overall structure as the positive amplitude-encoding case, with optimal performance concentrated at low values of j. However, the loss in performance becomes more pronounced as $|\lambda|$ increases, indicating a stronger sensitivity of the phase kernel to large positive deformations. This qualitative pattern is again independent of the selected benchmark data set.

Despite the differences in parameter sensitivity, the best accuracies achieved by all kernel families are remarkably similar for a given artificial data set. The dominant factor determining the maximum reachable accuracy is therefore not the kernel family itself, but rather the intrinsic difficulty of the generated classifica-

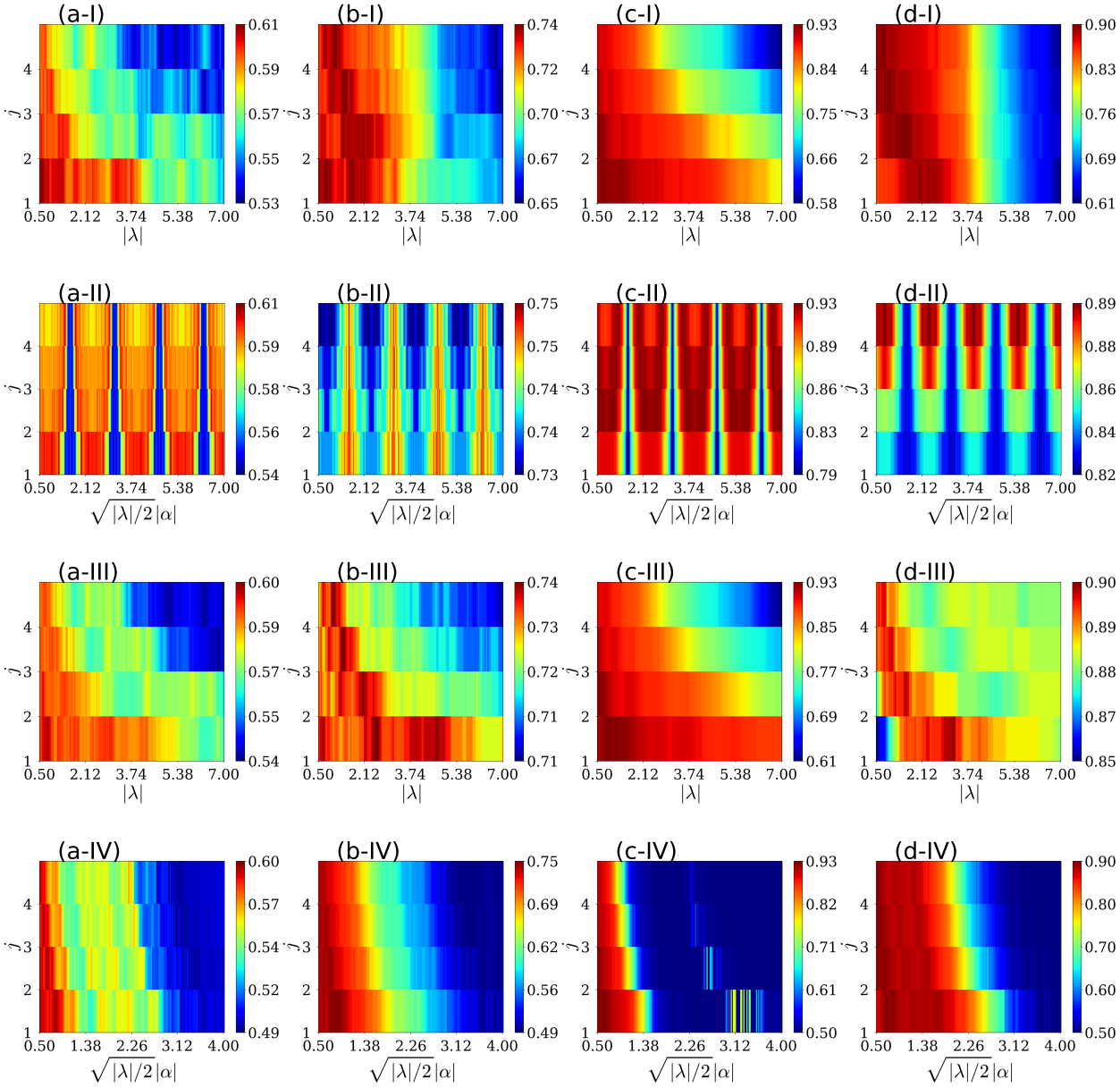

Figure 3: Accuracy heatmaps for synthetic datasets. Columns (a)–(d) correspond to Circles, Double Layers, Hypercubes, and Moons, respectively. Rows (I) and (II) correspond to $\lambda < 0$ kernels with amplitude and phase embeddings, respectively. Rows (III) and (IV) correspond to $\lambda > 0$ kernels with amplitude and phase embeddings, respectively.

tion task, including the imposed noise level and point distribution. In this sense, all kernels appear capable of reaching a common saturation limit set by the data structure, while differing mainly in how efficiently they access that optimum region in parameter space.

5.2 Optimized benchmark comparisons

In this section, we present the classification performance obtained with the proposed Kerr-inspired kernels after hyperparameter optimization over (λ, j). We first analyze optimized results on synthetic benchmark datasets in order to compare the kernels against previously reported reference methods under controlled geometric classification tasks. Finally, we evaluate the kernels on real-world benchmark datasets, where fixed feature distributions and experimental noise offer a more representative assessment of practical classification capabilities. More details about the datasets, synthetic and real-world, are described in Appendix C.

Beyond the global characterization provided by the heatmap analysis of Section 5.1.1, we next compare the optimized classification performance of the proposed kernels against previously reported benchmark results. For every kernel family, a full hyperparameter sweep over (λ, j) was performed and the best-performing configuration under five-fold cross-validation was retained. The optimal hyperparameters together with their corresponding classification accuracies are reported in the benchmark tables throughout this section.

To provide meaningful comparisons with previous literature, the benchmark methodology was selected to closely reproduce the evaluation procedures employed in the corresponding reference works whenever possible. Reported baseline values therefore correspond to results obtained under comparable dataset constructions and validation protocols.

5.2.1 Synthetic datasets benchmark

We first consider several synthetic binary-classification benchmarks commonly employed in the kernel-learning literature. The datasets were generated using `scikit-learn` utilities with a total of 2000 samples per dataset. Further details of the datasets used is shown in Appendix C. These benchmarks were selected to probe different geometric decision boundaries, including periodic structures, disconnected regions, and highly non-linear manifolds.

Table 1 summarizes the best classification accuracies obtained for each synthetic benchmark after hyperparameter optimization over (λ, j). Across all considered datasets, the Kerr-inspired kernels achieve high classification accuracies, indicating that the proposed embeddings are able to capture a wide range of non-linear decision-boundary geometries.

For the Annular Circles and Spiral datasets, all kernel families achieve accuracies above 0.95, showing that both amplitude and phase embeddings can effectively represent radial and winding decision structures. The Spiral benchmark reaches accuracies up to 0.994, suggesting that this dataset is close to saturation for the considered kernel families. In this regime, differences between kernels are more strongly reflected in the location and stability of the optimal hyperparameter region than in the peak accuracy itself.

The Corners benchmark exhibits a clearer distinction between embedding strategies. The λ^- amplitude kernel achieves the highest accuracy among the Kerr-inspired kernels, while the phase-based kernels show lower performance. This behavior is consistent with the heatmap analysis, where phase embeddings displayed stronger oscillatory sensitivity to the geometric arrangement of samples.

For the Sinus dataset, the λ^- amplitude kernel and the λ^+ phase kernel both reach the highest observed accuracy among the Kerr-inspired kernels. This indicates that different Kerr embeddings can access comparable classification performance while relying on distinct regions of hyperparameter space.

Overall, the synthetic benchmarks show that the proposed kernels can reach high-performance regimes across several controlled nonlinear classification tasks. However, because artificially generated datasets depend strongly on the sampling procedure, imposed noise level, and geometric construction, these results should be interpreted primarily as qualitative probes of kernel expressivity and hyperparameter behavior rather than as definitive measures of general performance.

Table 1: Best classification accuracy obtained on synthetic benchmark datasets. For each kernel family, the reported values correspond to the optimal hyperparameter pair (λ, j) found during the grid search. Best result for each dataset is highlighted in bold. Baseline values correspond to the best reported results from comparable dataset constructions and evaluation protocols.

Dataset	Method	Accuracy ↑	Best (λ, j)
Annular Circles	λ^- amplitude kernel	0.968	(2.34, 3)
	λ^- phase kernel	0.954	(0.70, 4)
	λ^+ amplitude kernel	0.964	(2.68, 4)
	λ^+ phase kernel	0.962	(0.78, 4)
	Baseline Rodriguez-Grasa et al. (2025)	**0.980**	–
Corners	λ^- amplitude kernel	**0.978**	(4.82, 2)
	λ^- phase kernel	0.942	(0.58, 4)
	λ^+ amplitude kernel	0.964	(1.70, 4)
	λ^+ phase kernel	0.966	(1.36, 1)
	Baseline Rodriguez-Grasa et al. (2025)	**0.978**	–
Sinus	λ^- amplitude kernel	**0.972**	(2.10, 2)
	λ^- phase kernel	0.954	(0.66, 4)
	λ^+ amplitude kernel	0.968	(4.54, 4)
	λ^+ phase kernel	**0.972**	(1.28, 1)
	Baseline Rodriguez-Grasa et al. (2025)	**0.972**	–
Spiral	λ^- amplitude kernel	0.992	(2.24, 4)
	λ^- phase kernel	0.988	(0.78, 4)
	λ^+ amplitude kernel	0.994	(3.94, 2)
	λ^+ phase kernel	0.994	(0.88, 2)
	Baseline Rodriguez-Grasa et al. (2025)	**1.000**	–

Table 2: Accuracy on the Framingham CHD benchmark. Best result in bold. Baseline values correspond to the best reported results obtained under the same dataset and evaluation protocol.

Method	Accuracy ↑	Best (λ, j)
λ^- phase kernel	0.7141	(0.12, 2)
λ^+ phase kernel	**0.7190**	(0.15, 1)
λ^- amplitude kernel	0.7072	(0.33, 2)
λ^+ amplitude kernel	0.7078	(0.27, 1)
Baseline Swaminathan & Akgün (2025)	0.6788	–

5.2.2 Real-world datasets benchmark

While synthetic datasets are useful for probing geometric inductive biases and kernel expressivity, real-world benchmarks provide a more stringent evaluation scenario due to fixed feature distributions, experimental noise, and reduced controllability of the data-generation process. We therefore complement the synthetic analyses with two real-world classification tasks involving medical tabular data and high-dimensional image classification.

The results obtained for the Framingham CHD and Fashion-MNIST benchmarks are summarized in Tables 2 and 3, respectively.

For the Framingham CHD dataset, the best performance is obtained with the λ^+ phase-encoding kernel, reaching an accuracy of 0.7190, followed closely by the λ^- phase-encoding kernel with 0.7141. The amplitude-encoding kernels achieve slightly lower, but still competitive, accuracies above 0.707. These results suggest

Table 3: Accuracy on the Fashion-MNIST binary classification benchmark (dress vs. shirt). Best result in bold. Baseline values correspond to the best reported results obtained under the same dataset and evaluation protocol.

Method	Accuracy ↑	Best (λ, j)
λ^- phase kernel	**0.9300**	(20.10, 3)
λ^+ phase kernel	0.9200	(0.01, 1)
λ^- amplitude kernel	0.9120	(0.01, 1)
λ^+ amplitude kernel	0.9120	(0.01, 1)
Baseline Rodriguez-Grasa et al. (2025)	≈ 0.91	–
Baseline Placidi et al. (2023)	0.8678	–

Table 4: Best classification accuracy obtained on synthetic benchmark datasets for comparison kernels. For each method, the reported values correspond to the optimal hyperparameters found during the grid search. For the exponential sine squared kernel, parameters are reported as (p, l), where p is the period and l the length scale. For the quantum exponential cosine kernel, parameters are reported as (λ, l). Best result for each dataset is highlighted in bold. Baseline values correspond to the best reported results from comparable dataset constructions and evaluation protocols.

Dataset	Method	Accuracy ↑	Best parameters
Moons	Exponential sine squared kernel	**1.000**	$(p, l) = (3.20, 0.10)$
	Quantum exponential cosine kernel	**1.000**	$(\lambda, l) = (0.50, 0.10)$
	Baseline Li et al. (2022)	**1.000**	–
Sinus	Exponential sine squared kernel	0.974	$(p, l) = (3.00, 0.30)$
	Quantum exponential cosine kernel	**0.976**	$(\lambda, l) = (0.50, 0.10)$
	Baseline Li et al. (2022)	0.96	–
Spiral	Exponential sine squared kernel	0.994	$(p, l) = (5.10, 0.20)$
	Quantum exponential cosine kernel	0.994	$(\lambda, l) = (2.50, 0.30)$
	Baseline Li et al. (2022)	**1.000**	–

that phase-based embeddings provide a modest advantage for this structured medical tabular dataset, likely due to their richer nonlinear feature representation.

For the Fashion-MNIST binary classification task, corresponding to dress-versus-shirt classification, the best result is obtained with the λ^- phase-encoding kernel, reaching an accuracy of 0.9300. The remaining kernels achieve accuracies between 0.9120 and 0.9200, confirming that all four Kerr-inspired constructions remain competitive in a high-dimensional image-classification setting.

A consistent observation across both real-world benchmarks is the strong performance of phase-encoding kernels, particularly in the λ^- case for Fashion-MNIST and the λ^+ case for Framingham CHD. This trend is consistent with the oscillatory expressivity observed in the heatmap analysis, where phase embeddings exhibited richer hyperparameter landscapes. Such flexibility may be advantageous when learning structured or high-dimensional decision boundaries.

Overall, the real-world benchmarks support the practical relevance of the proposed kernels. The results indicate that Kerr-inspired embeddings can be integrated into a simple SVM-based pipeline while maintaining competitive performance across both tabular and image-based classification tasks.

5.2.3 Additional kernel comparisons

We further compare the observed trends with the alternative kernel constructions in Eqs. (8) and (11). These additional benchmarks help contextualize the saturation behavior observed in the synthetic datasets.

Table 4 reports optimized results for the exponential sine squared kernel and the quantum exponential cosine kernel. Both kernels achieve near-perfect accuracies on the considered synthetic datasets, with several cases reaching complete or near-complete saturation.

In particular, the Moons dataset reaches perfect classification accuracy for both comparison kernels, while the Spiral dataset saturates at approximately 0.994. This behavior is consistent with the saturation observed for the Kerr-inspired kernels and reinforces the conclusion that, for low-noise synthetic tasks, the peak accuracy can be largely determined by the dataset construction itself.

Under these conditions, the most informative differences between kernels are not necessarily reflected in the maximum accuracy, but rather in the structure of their hyperparameter landscapes, the width of high-performance regions, and their robustness to changes in the data distribution.

6 Conclusions

In this work, we introduced quantum kernels based on Kerr coherent states and investigated their geometric and machine-learning properties. We showed that Kerr coherent states generalize both $su(1,1)$ and $su(2)$ coherent states depending on the sign of the Kerr parameter, naturally inducing hyperbolic and spherical feature-space geometries with curvature determined by physically controllable parameters such as the optical frequency and Kerr nonlinearity.

Using these coherent states, we constructed both phase-encoding and amplitude-encoding quantum feature maps. The resulting kernels define structured feature spaces corresponding to pseudo-Riemannian and Riemannian manifolds of constant curvature. This establishes a direct connection between experimentally tunable physical parameters and the geometry of the induced machine-learning feature space, opening potential applications in quantum manifold learning and hyperbolic machine learning using photonic platforms.

We further analyzed the performance of the proposed Kerr-inspired kernels within Support Vector Machine classifiers. For phase encoding, the resulting kernels recover and generalize squeezed-state kernels, while the amplitude-encoding constructions generalize cosine and hyperbolic-cosine kernel families. In addition, we introduced the quantum exponential cosine kernel as a periodic-kernel construction derived from Kerr coherent states.

Extensive benchmark analyses were performed on both synthetic and real-world datasets in order to characterize the dependence of the kernels on the hyperparameters (λ, j). The results show that the proposed kernels achieve competitive classification performance across a variety of nonlinear classification tasks while exhibiting characteristic and robust hyperparameter landscapes. In particular, phase-encoding Kerr kernels demonstrated strong performance on real-world datasets involving medical and image-classification tasks.

Overall, the proposed framework establishes Kerr coherent states as a flexible and physically motivated platform for quantum kernel methods, connecting quantum optics, curved feature-space geometry, and kernel-based machine learning within a unified approach.

References

Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1:403–409, 2021.

Mark A. Aizerman, Emmanuel M. Braverman, and Lev I. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2623–2631, 2019.

Syed Twareque Ali, Jean-Pierre Antoine, and Jean-Pierre Gazeau. *Coherent States, Wavelets and Their Generalizations*. Springer, 2000.

Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. Latent space oddity: On the curvature of deep generative models. In *International Conference on Learning Representations*, 2018.

Mina Ghadimi Atigh, Joris Schoep, Erman Acar, Nanne van Noord, and Pascal Mettes. Hyperbolic image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4453–4462, 2022.

Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549:195–202, 2017.

Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152, 1992.

Nicolas Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023.

Martin D. Buhmann. Radial basis functions. *Acta Numerica*, 9:1–38, 2000.

Ranjan Chatterjee and Ting Yu. Generalized coherent states, reproducing kernels, and quantum machine learning. *Quantum Information and Computation*, 17:1292–1306, 2017.

Demetrios N. Christodoulides, Falk Lederer, and Yaron Silberberg. Discretizing light behaviour in linear and nonlinear waveguide lattices. *Nature*, 424:817–823, 2003.

Monique Combescure and Didier Robert. *Coherent States and Applications in Mathematical Physics*. Springer, 2012.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995.

R. L. de Matos Filho and W. Vogel. Nonlinear coherent states. *Physical Review A*, 54:4560–4563, 1996.

Shahram Dehdashti, Janis Nötzel, and Peter van Loock. Quantum capacity of a deformed bosonic dephasing channel. *Optical Switching and Networking*, pp. 100814, 2025.

Shahram Dehdashti et al. Realization of non-linear coherent states by photonic lattices. *AIP Advances*, 5 (6):067165, 2015.

Victor V. Dodonov and Vladimir I. Man'ko. Nonclassical states in quantum physics: A brief historical review. *Theory of Nonclassical States of Light*, pp. 1–102, 2003.

Peng Fang, Mehrtash Harandi, and Lars Petersson. Kernel methods in hyperbolic spaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10665–10674, 2021.

Peng Fang, Mehrtash Harandi, Zhen Lan, and Lars Petersson. Poincaré kernels for hyperbolic representations. *International Journal of Computer Vision*, 2023.

Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.

Zhi Gao, Yuwei Wu, Yunde Jia, and Mehrtash Harandi. Hyperbolic feature augmentation via distribution estimation and infinite sampling on manifolds. In *Advances in Neural Information Processing Systems*, volume 35, pp. 34421–34433, 2022.

Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum random access memory. *Physical Review Letters*, 100:160501, 2008.

Rishi Goel, Casey R Myers, and Sally Shrapnel. Quantum algorithms for causal estimands. *Quantum Science and Technology*, 11(1):015024, 2026.

Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103:150502, 2009.

Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567: 209–212, 2019.

Laura J Henderson, Kerstin Beer, Salini Karuvade, Riddhi Gupta, Angela White, and Sally Shrapnel. Quantum advantage without exponential concentration: Trainable kernels for symmetry-structured data. *arXiv preprint arXiv:2509.14337*, 2025.

Lewis J. Henderson, Rishabh Goel, and Sally Shrapnel. Quantum kernel machine learning with continuous variables. *Quantum*, 8:1570, 2024.

Thomas Hofmann, Bernhard Schölkopf, and Alexander J. Smola. Kernel methods in machine learning. *The Annals of Statistics*, 36(3):1171 – 1220, 2008. doi: 10.1214/009053607000000677. URL `https://doi.org/10.1214/009053607000000677`.

Junjie Hong, Peng Fang, Wei Li, Jungong Han, Lars Petersson, and Mehrtash Harandi. Curved geometric networks for visual anomaly recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

Andras Janosi, William Steinbrunn, Matthias Pfisterer, and Robert Detrano. Heart disease dataset. *UCI Machine Learning Repository*, 1989. doi: 10.24432/C52P4X.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021.

Robert Keil, Armando Perez-Leija, Felix Dreisow, Matthias Heinrich, Héctor Moya-Cessa, Stefan Nolte, Demetrios N. Christodoulides, and Alexander Szameit. Classical analogue of displaced Fock states and quantum correlations in Glauber–Fock photonic lattices. *Physical Review Letters*, 107:103601, 2011.

Jonas M. Kübler, Krikamol Muandet, and Bernhard Schölkopf. Quantum mean embedding of probability distributions. *Physical Review Research*, 1:033159, 2019.

Long Hin Li, Dan-Bo Zhang, and Z.D. Wang. Quantum kernels with gaussian state encoding for machine learning. *Physics Letters A*, 436:128088, 2022. ISSN 0375-9601. doi: https://doi.org/10.1016/j.physleta.2022.128088. URL `https://www.sciencedirect.com/science/article/pii/S0375960122001700`.

Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10:631–633, 2014.

Seth Lloyd, Silvano Garnerone, and Paolo Zanardi. Quantum algorithms for topological and geometric analysis of data. *Nature Communications*, 7:10138, 2016.

Min Lu and Fei Li. Survey on Lie group machine learning. *Big Data Mining and Analytics*, 3(3):235–258, 2020.

Samuel Mauran, Sylvain Mouysset, Emmanuel Simon, and Laurent Bertino. A kernel extension of the ensemble transform Kalman filter. In *International Conference on Computational Science*, pp. 438–452. Springer, 2023.

Pascal Mettes, Mina Ghadimi Atigh, Martin Keller-Ressel, Jeffrey Gu, and Serena Yeung. Hyperbolic deep learning in computer vision: A survey. *International Journal of Computer Vision*, 132(9):3484–3508, 2024.

S. R. Miry and M. K. Tavassoly. Generation of a class of SU(1,1) coherent states of the Gilmore–Perelomov type and a class of SU(2) coherent states and their superposition. *Physica Scripta*, 85:035404, 2012.

Mohamad T. Musavi, Waseem Ahmed, King Ho Chan, Khalid B. Faris, and Donald M. Hummels. On the training of radial basis function classifiers. *Neural Networks*, 5(4):595–603, 1992.

Mark JL Orr et al. Introduction to radial basis function networks, 1996.

A. M. Perelomov. Generalized coherent states and some of their applications. *Soviet Physics Uspekhi*, 20(9): 703–720, 1977.

Armando Perez-Leija, Héctor Moya-Cessa, Alexander Szameit, and Demetrios N. Christodoulides. Glauber-fock photonic lattices. *Optics Letters*, 35(14):2409–2411, 2010.

Evan Peters and Maria Schuld. Generalization despite overfitting in quantum machine learning models. *Quantum*, 7:1210, December 2023. ISSN 2521-327X. doi: 10.22331/q-2023-12-20-1210. URL `https://doi.org/10.22331/q-2023-12-20-1210`.

Leonardo Placidi et al. MnisQ: A large-scale quantum circuit dataset for machine learning on/for quantum computers in the NISQ era. *arXiv preprint arXiv:2306.16627*, 2023.

Saurav Prakash, Jin Sima, Chao Pan, Eli Chien, and Olgica Milenkovic. Federated classification in hyperbolic spaces via secure aggregation of convex hulls. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL `https://openreview.net/forum?id=umggDfMHha`.

John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018. ISSN 2521-327X. doi: 10.22331/q-2018-08-06-79. URL `https://doi.org/10.22331/q-2018-08-06-79`.

Enyan Qu and Dongmian Zou. Lorentzian fully hyperbolic generative adversarial network. *arXiv preprint arXiv:2201.12825*, 2022.

Liva Ralaivola, S. Joshua Swamidass, Hiroto Saigo, and Pierre Baldi. Graph kernels for chemical informatics. *Neural Networks*, 18(8):1093–1110, 2005.

Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Physical Review Letters*, 113:130503, 2014.

Pablo Rodriguez-Grasa, Yue Ban, and Mikel Sanz. Neural quantum kernels: Training quantum kernels with quantum neural networks. *Phys. Rev. Res.*, 7:023269, Jun 2025. doi: 10.1103/xphb-x2g4. URL `https://link.aps.org/doi/10.1103/xphb-x2g4`.

Bernhard Schölkopf, Koji Tsuda, and Jean-Philippe Vert. *Kernel Methods in Computational Biology*. MIT Press, 2004.

Maria Schuld. Supervised quantum machine learning models are kernel methods. *arXiv preprint arXiv:2101.11020*, 2021.

Maria Schuld and Nathan Killoran. Quantum machine learning in feature Hilbert spaces. *Physical Review Letters*, 122:040504, 2019.

Maria Schuld and Francesco Petruccione. Representing data on a quantum computer. In *Machine Learning with Quantum Computers*, pp. 147–176. Springer, 2021.

Marlan O. Scully and M. Suhail Zubairy. *Quantum Optics*. Cambridge University Press, 1999.

John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

E. Solano, G. S. Agarwal, and H. Walther. Strong-driving-assisted multipartite entanglement in cavity QED. *Physical Review Letters*, 90:027903, 2003.

Arjhun Swaminathan and Mete Akgün. Distributed and secure kernel-based quantum machine learning. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL `https://openreview.net/forum?id=3jdIOaEW3k`.

Prayag Tiwari et al. Kernel method based on non-linear coherent states in quantum feature space. *Journal of Physics A: Mathematical and Theoretical*, 55(35):355301, 2022.

Andrea Vedaldi and Andrew Zisserman. Efficient additive kernels via explicit feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):480–492, 2012.

Alonso Viladomat Jasso, Ark Modi, Roberto Ferrara, Christian Deppe, Janis Nötzel, Fred Fung, and Maximilian Schädler. Quantum and quantum-inspired stereographic k nearest-neighbour clustering. *Entropy*, 25(9), 2023. ISSN 1099-4300. doi: 10.3390/e25091361. URL `https://www.mdpi.com/1099-4300/25/9/1361`.

W. Vogel and R. L. de Matos Filho. Nonlinear Jaynes-Cummings dynamics of a trapped ion. *Physical Review A*, 52:4214–4217, 1995.

Shuhei Watanabe. Tree-structured Parzen estimator: Understanding its algorithm components and their roles for better empirical performance. *arXiv preprint arXiv:2304.11127*, 2023.

Shuhei Watanabe, Arpit Bansal, and Frank Hutter. PED-ANOVA: Efficiently quantifying hyperparameter importance in arbitrary subspaces. *arXiv preprint arXiv:2304.10255*, 2023.

Richard C. Wilson, Edwin R. Hancock, Elzbieta Pekalska, and Robert P. W. Duin. Spherical and hyperbolic embeddings of data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2255–2269, 2014.

Carolyn Wood, Sally Shrapnel, and G J Milburn. A kerr kernel quantum learning machine, 2024. URL `https://arxiv.org/abs/2404.01787`.

Yi Xia, Wei Li, Quntao Zhuang, and Zheshen Zhang. Quantum-enhanced data classification with a variational entangled sensor network. *Physical Review X*, 11:021047, 2021.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017.

T. Yamaji, S. Kagami, A. Yamaguchi, T. Satoh, K. Koshino, H. Goto, Z. R. Lin, Y. Nakamura, and T. Yamamoto. Spectroscopic observation of the crossover from a classical duffing oscillator to a kerr parametric oscillator. *Phys. Rev. A*, 105:023519, Feb 2022. doi: 10.1103/PhysRevA.105.023519. URL `https://link.aps.org/doi/10.1103/PhysRevA.105.023519`.

Ming Yang, Peng Fang, and Hui Xue. Expanding the hyperbolic kernels: A curvature-aware isometric embedding view, 2023.

Ming Yang, Qing Liu, Ximeng Sun, Ning Shi, and Hui Xue. Towards kernelizing the classifier for hyperbolic data. *Frontiers of Computer Science*, 18:181301, 2024.

Han-Sen Zhong et al. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, 2020.

Han-Sen Zhong et al. Phase-programmable Gaussian boson sampling using stimulated squeezed light. *Physical Review Letters*, 127:180502, 2021.

Jian Zhou, Shuang Zhu, Yan Qiu, Danial Jahed Armaghani, Aiguo Zhou, and Wan Yong. Predicting tunnel squeezing using support vector machine optimized by whale optimization algorithm. *Acta Geotechnica*, 17 (4):1343–1366, 2022.

Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D. Lukin. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. *Physical Review X*, 10:021067, 2020.

X. Zou, K. Pahlke, and W. Mathis. Scheme for direct measurement of the Wigner characteristic function in cavity QED. *Physical Review A*, 69:015802, 2004.

A Kerr Coherent states

A Kerr Hamiltonian can be defined as follows:

$$\hat{H} = \omega\hat{n} + \frac{\lambda}{2}\hat{n}^2 \tag{12}$$

in which $\hat{n} = \hat{a}^\dagger\hat{a}$ is the number operator and λ is called Kerr parameter Scully & Zubairy (1999). The Hamiltonian represents an anharmonic oscillator in which the spacing between energy levels is anharmonic as schematically plotted in Fig. 4. The Kerr Hamiltonian is widely used in different quantum systems, from Josephson-Junction qubits, fiber optics, Kerr cavity, *etc.*

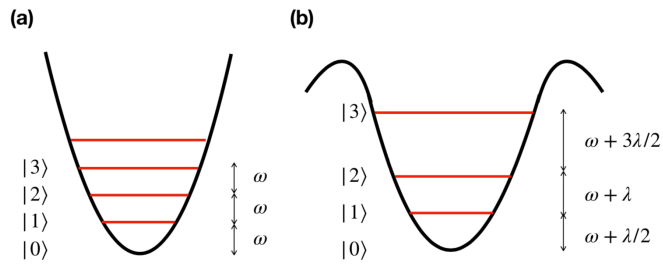

Figure 4: Schematically the spectrum of energy for a harmonic and anharmonic oscillator respectively in the plot (a) and (b).

Definition A.1. *A Kerr coherent state is defined by applying the Kerr displacement operator*

$$D(\alpha) = \exp\left[\alpha\hat{A}^\dagger - \alpha^*\hat{A}\right] \tag{13}$$

on the reference state $|0\rangle$*. For the positive Kerr parameter, the ladder operators are defined by*

$$\begin{aligned} \hat{A} &= \sqrt{\frac{\lambda}{2}}\hat{a}\sqrt{2j-1+\hat{n}} \\ \hat{A}^\dagger &= \sqrt{\frac{\lambda}{2}}\sqrt{2j-1+\hat{n}}\hat{a}^\dagger \end{aligned} \tag{14}$$

For the negative Kerr parameter, the ladder operators are given by

$$\begin{aligned} \hat{A} &= \sqrt{\frac{|\lambda|}{2}}\hat{a}\sqrt{2j+1-\hat{n}} \\ \hat{A}^\dagger &= \sqrt{\frac{|\lambda|}{2}}\sqrt{2j+1-\hat{n}}\ \hat{a}^\dagger \end{aligned} \tag{15}$$

Lemma 1. *Let us define* $[\hat{A}, \hat{A}^\dagger] = 2\hat{K}_0$*. Then the operators* $\hat{A}$*,* $\hat{A}^\dagger$ *and* $\hat{K}_0$ *satisfy the following commutation relations:*

$$[\hat{K}_0, \hat{A}] = -\frac{\lambda}{2}\hat{A},\ [\hat{K}_0, \hat{A}^\dagger] = \frac{\lambda}{2}\hat{A}^\dagger \tag{16}$$

Then the Gaussian decomposition of the displacement operator $D(\alpha, \lambda) = \exp\left[\alpha\hat{A}^\dagger - \alpha^*\hat{A}\right]$ *is given by*

$$D(\alpha, \lambda) = e^{\alpha\hat{A}^\dagger - \alpha^*\hat{A}} = e^{\zeta\hat{A}^\dagger}e^{\ln[\zeta_0]\hat{K}_0}e^{-\zeta^*\hat{A}} \tag{17}$$

where in case that $\lambda > 0$

$$\zeta = \frac{\alpha}{|\alpha|}\sqrt{\frac{2}{\lambda}}\tanh\left(\sqrt{\frac{\lambda}{2}}|\alpha|\right), \tag{18}$$

$$\zeta_0 = \cosh^{-j}\left(\sqrt{\frac{\lambda}{2}}|\alpha|\right) \tag{19}$$

and in the case of negative Kerr parameter, we have

$$\zeta = \frac{\alpha}{|\alpha|}\sqrt{\frac{2}{|\lambda|}}\tan\left(\sqrt{\frac{|\lambda|}{2}}|\alpha|\right), \tag{20}$$

$$\zeta_0 = \cos^j\left(\sqrt{\frac{|\lambda|}{2}}|\alpha|\right). \tag{21}$$

Proof. Let us consider operator $F(t)$, with parameter t defined by

$$F(t) = e^{\left[\beta A^\dagger - \beta^* A\right]t} = e^{\zeta_+(t)A^\dagger} e^{\ln[\zeta_0(t)]K_0} e^{-\zeta_-(t)A} \tag{22}$$

where $\zeta_\pm(t)$ and $\zeta_0(t)$ are c-number functions of parameter t to be determined under the following conditions: $\zeta_\pm(0) = 0$ and $\zeta_0(0) = 1$. Once $\zeta_\pm(t)$ and $\zeta_0(t)$ are determined, $\zeta_\pm$ and ζ_0 are given by $\zeta_\pm(1)$ and $\zeta_0(1)$, respectively.
Now we have

$$\begin{aligned}
\frac{d}{dt}F &= \left[\alpha A^\dagger - \alpha^* A\right] F \\
&= \dot{\zeta}_+ \hat{A}^\dagger F \\
&\quad + e^{\zeta_+(t)A^\dagger}\frac{\dot{\zeta}_0}{\zeta_0}K_0 e^{\ln[\zeta_0(t)]K_0} e^{-\zeta_-(t)A} \\
&\quad + e^{\zeta_+(t)A^\dagger} e^{\ln[\zeta_0(t)]K_0}\dot{\zeta}_- A e^{-\zeta_-(t)A} \\
&= \dot{\zeta}_+ \hat{A}^\dagger F \\
&+ \frac{\dot{\zeta}_0}{\zeta_0}(K_0 - \frac{\lambda}{2}\zeta_+ A^\dagger)F \\
&- \dot{\zeta}_- e^{-\lambda \ln[\zeta_0]/2}(A - 2\zeta_+ K_0 + \frac{\zeta_+^2\lambda}{2}A^\dagger)F
\end{aligned}$$

Hence, three coupling differential equations are derived, i.e., Dehdashti et al. (2025):

$$\begin{aligned}
& e^{-\lambda \ln[\zeta_0]/2}\frac{d\zeta_-}{dt} = \alpha^*, \\
& \zeta_0^{-1}\frac{d\zeta_0}{dt} + 2e^{-\lambda \ln[\zeta_0]/2}\zeta_+\frac{d\zeta_-}{dt} = 0 \\
& \frac{d\zeta_+}{dt} - \frac{\lambda\zeta_+}{2\zeta_0}\frac{d\zeta_0}{dt} - e^{-\lambda \ln[\zeta_0]/2}\frac{\lambda\zeta_+^2}{2}\frac{d\zeta_-}{dt} = \alpha
\end{aligned} \tag{23}$$

These can be reduced to

$$\begin{aligned}
& \zeta_0^{-1}\frac{d\zeta_0}{dt} + 2\zeta_+\beta^* = 0 \\
& \frac{d\zeta_+}{dt} - \frac{\lambda\zeta_+}{2\zeta_0}\frac{d\zeta_0}{dt} - \frac{\lambda\zeta_+^2}{2}\beta^* = \beta
\end{aligned}$$

The simple calculations gives the differential equation:

$$\frac{d}{dt}\zeta_+ + \frac{\lambda}{2}\beta^*\zeta_+^2 = \beta \tag{24}$$

which, under the condition $\zeta_+(0) = 0$ and $\lambda > 0$, has the following solution:

$$\begin{aligned}
\zeta_+ &= \sqrt{\frac{\beta}{\beta^*}}\sqrt{\frac{2}{\lambda}}\tanh\left[t\sqrt{\frac{\lambda}{2}}|\beta|\right] \\
&= e^{i\phi}\sqrt{\frac{2}{\lambda}}\tanh\left[t\sqrt{\frac{\lambda}{2}}|\beta|\right]
\end{aligned} \tag{25}$$

in which $\beta = |\beta|e^{i\phi}$. Now by using the equation (23), we write

$$\zeta_0 = \cosh^{-4/\lambda}\left[t\sqrt{\frac{\lambda}{2}}|\beta|\right]. \tag{26}$$

Now, by considering $t = 1$ in the relations (25) and (26), the desired results are obtained.

Note that Baker-Hausdorff lemma and the commutation relations (16) can be used to achieve the following relations:

$$\begin{aligned} e^{\zeta_+(t)A^\dagger}K_0 e^{\zeta_-(t)A} &= K_0 - \frac{\lambda}{2}\zeta_+ A^\dagger \\ e^{\zeta_+(t)A^\dagger}e^{\ln[\zeta_0(t)]K_0}Ae^{\ln[\zeta_0(t)]K_0}e^{\zeta_-(t)A} &= \zeta_0^{-\lambda/2} \\ &\times\left(A - 2\zeta_+K_0 + \frac{\zeta_+^2\lambda}{2}A^\dagger\right) \end{aligned}$$

In the case $\lambda < 0$ then we can use the fact that $\cosh ix = \cos x$ and $\sinh ix = i\sin x$ to rewrite the relations (25) and (26) as follows:

$$\begin{aligned} \zeta_+ &= \sqrt{\frac{\beta}{\beta^*}}\sqrt{\frac{2}{|\lambda|}}\tan\left[t\sqrt{\frac{|\lambda|}{2}}|\beta|\right] \\ &= e^{i\phi}\sqrt{\frac{2}{|\lambda|}}\tan\left[t\sqrt{\frac{|\lambda|}{2}}|\beta|\right] \end{aligned} \tag{27}$$

and

$$\zeta_0 = \cos^{4/|\lambda|}\left[t\sqrt{\frac{|\lambda|}{2}}|\beta|\right]. \tag{28}$$

■

Lemma 2. *The displacement Kerr coherent states are given by*

$$\begin{aligned} \left|\alpha;\lambda^+,j\right\rangle &= \cosh^{-2j}\left[\sqrt{\frac{\lambda}{2}}|\alpha|\right]\sum_{n=0}^{\infty}\sqrt{\frac{\Gamma(2j+n)}{\Gamma(2j)n!}}e^{-in\phi} \\ &\times \tanh^n\left[\sqrt{\frac{\lambda}{2}}|\alpha|\right]|n\rangle,\ \lambda > 0 \end{aligned} \tag{29}$$

and

$$\begin{aligned} \left|\alpha;\lambda^-,j\right\rangle &= \cos^{2j}\left[\sqrt{\frac{|\lambda|}{2}}|\alpha|\right]\sum_{n=0}^{2j}\sqrt{\frac{(2j)!}{(2j-n)!n!}}e^{-in\phi} \\ &\times \tan^n\left[\sqrt{\frac{|\lambda|}{2}}|\alpha|\right]|n\rangle,\ \lambda < 0 \end{aligned} \tag{30}$$

Proof. Using the Gaussian decomposition introduced in the lemma 1 gives the result directly. ■

Proposition 1. *The Kerr coherent states given by equations (29) and (30) correspondingly simplify to the $su(1,1)$ and $su(2)$ coherent states when the values of λ are set to 2 and -2, respectively.*

The approach used to explore the structure of the Kerr coherent state manifolds involves examining their Fubini-Study metric, which is defined by the infinitesimal distance ds between two neighboring Kerr coherent states.

Definition A.2. *A Fubini-Study metric is defined by considering the differential distance ds between two neighboring pure quantum states $|\psi(x^\mu)\rangle$ and $|\psi(x^\mu + dx^\mu)\rangle$:*

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{31}$$

where

$$g_{\mu\nu} = \frac{\langle \psi_\mu | \psi_\nu \rangle}{\langle \psi | \psi \rangle} - \frac{\langle \psi_\mu | \psi \rangle \langle \psi | \psi_\nu \rangle}{| \langle \psi | \psi \rangle |^2} \tag{32}$$

in which x^μ is a set of real parameters which define the state $|\psi\rangle$ and $|\psi_\mu\rangle = \frac{\partial}{\partial x^\mu} |\psi\rangle$.

Lemma 3. *The Fubini-Study metrics of the Kerr coherent state, for the positive and negative Kerr parameter, are respectively given by*

$$ds^2 = j\lambda dr^2 + \frac{j}{2} \sinh^2 \left[\sqrt{2\lambda} r \right] d\phi^2 \tag{33}$$

$$ds^2 = j\lambda dr^2 + \frac{j}{2} \sin^2 \left[\sqrt{2\lambda} r \right] d\phi^2 \tag{34}$$

while the following stands for the coherent state one:

$$ds^2 = dr^2 + r^2 d\phi^2 \tag{35}$$

Proof. Let us define the non-normalized Kerr coherent state as

$$|z\rangle = \sum_{n=0}^{\infty} \sqrt{\frac{\Gamma(2j+n)}{\Gamma(2j) n!}} z^n |n\rangle , \tag{36}$$

where $z = \tanh \left[\sqrt{\frac{\lambda}{2}} r \right] e^{-i\phi}$. By replacing the following relations

$$\begin{aligned} \langle z | z \rangle &= (1 - |z|^2)^{-2j} \\ \langle z | z_z \rangle &= 2j \bar{z} (1 - |z|^2)^{-2j-1} dz \\ \langle z_z | z_z \rangle &= 2j (1 - |z|^2)^{-2j-2} (1 + 2j|z|^2) d\bar{z} dz \end{aligned}$$

in the equation (32), we obtain:

$$ds^2 = 2j \frac{dz d\bar{z}}{(1 - |z|^2)^2} \tag{37}$$

Now, by considering the fact that

$$dz = e^{-i\phi} \cosh^{-2} \left[\sqrt{\frac{\lambda}{2}} r \right] \left(\sqrt{\frac{\lambda}{2}} dr - \frac{\sinh \left[\sqrt{2\lambda} r \right]}{2} d\phi \right) ,$$

the above metric can be rewritten as

$$ds^2 = j\lambda dr^2 + \frac{j}{2} \sinh^2 \left[\sqrt{2\lambda} r \right] d\phi^2 \tag{38}$$

In the case of negative value of the Kerr parameters, the non-normalized coherent state is defined as

$$|z\rangle = \sum_{n=0}^{2j} \sqrt{\frac{(2j)!}{(2j-n)! n!}} z^n |n\rangle \tag{39}$$

in which $z = e^{-in\phi} \tan\left[\sqrt{\frac{|\lambda|}{2}}|\alpha|\right]$. Following the above method leads into the relations:

$$\begin{aligned}
\langle z|z\rangle &= (1+|z|^2)^{2j} \\
\langle z|z_z\rangle &= 2j\bar{z}(1+|z|^2)^{2j-1}dz \\
\langle z_z|z_z\rangle &= 2j(1+|z|^2)^{2j-2}(1+2j|z|^2)d\bar{z}dz
\end{aligned}$$

which gives the Fubini-Study metric,

$$ds^2 = \frac{2jdzd\bar{z}}{(1+|z|^2)^2} \tag{40}$$

In the case of normal coherent states, the non-normilzed coherent state is defined by $|z\rangle = \sum_{n=0}^{\infty} z^n/n!\,|n\rangle$, in which $z = re^{i\phi}$. Using the following relations

$$\begin{aligned}
\langle z|z\rangle &= e^{|z|^2} \\
\langle z|z_z\rangle &= e^{|z|^2}\bar{z}dz \\
\langle z_z|z_z\rangle &= (1+|z|^2)e^{|z|^2}d\bar{z}dz,
\end{aligned}$$

leads to the Fubini-Study metric as $ds^2 = dzd\bar{z}$. ■

Lemma 4. *The feature space derived from encoding data into Kerr coherent states is a two-dimensional manifold with uniform curvature. Specifically, the feature space created by utilizing positive Kerr coherent states forms a Hyperbolic plane $\mathbb{H}$, whereas the feature space constructed with negative Kerr coherent states constitutes a spherical plane $\mathbb{S}$.*

Proof. The non-zero Christoffel symbols, defined as $2\Gamma^i_{jk} = g^{il}\left(\partial_k g_{lj} + \partial_j g_{lk} - \partial_l g_{jk}\right)$, where g^{il} is the inverse metric g_{il} and $\partial_i = \frac{\partial}{\partial x_i}$, of the Fubini-Study metric are given by

$$\begin{aligned}
\Gamma^{\phi}_{r\phi} &= \sqrt{2\lambda}\coth\left[\sqrt{2\lambda}r\right], \\
\Gamma^{r}_{\phi\phi} &= -\sqrt{\frac{\lambda}{2}}\sinh\left[2\sqrt{2\lambda}r\right]
\end{aligned}$$

The non-zero element of the Riemann tensor defined

$$R^l_{ijk} = \partial_i\Gamma^l_{jk} - \partial_j\Gamma^l_{ik} - \Gamma^l_{jm}\Gamma^m_{ik} + \Gamma^l_{im}\Gamma^m_{jk} \tag{41}$$

is obtained by

$$R^r_{r\phi r\phi} = -2\lambda\sinh^2\left[\sqrt{2\lambda}r\right] \tag{42}$$

We find the non-zero Ricci tensor as the following:

$$\begin{aligned}
R_{rr} &= -4\lambda^2, \\
R_{\phi\phi} &= -2\lambda\sinh^2\left[\sqrt{2\lambda}r\right].
\end{aligned} \tag{43}$$

which leads to the constant Ricci scalar, i.e.,

$$R = g^{rr}R_{rr} + g^{\phi\phi}R_{\phi\phi} = -\frac{2\lambda}{j} \tag{44}$$

which indicates the metric has the negative constant. In addition, we can define the following parametrical equations as follows:

$$x_0 = \sqrt{\frac{j}{2}}\cosh\left[\sqrt{2\lambda}r\right] \tag{45}$$

$$x_1 = \sqrt{\frac{j}{2}}\sinh\left[\sqrt{2\lambda}r\right]\cos\phi \tag{46}$$

$$x_2 = \sqrt{\frac{j}{2}}\sinh\left[\sqrt{2\lambda}r\right]\sin\phi \tag{47}$$

which is defined as a conformal Minkowski space,

$$ds^2 = dx_0^2 - dx_1^2 - dx_2^2 \tag{48}$$

and defines the following hyperbolic sheet:

$$x_0^2 - x_1^2 - x_2^2 = \frac{j^2}{4} \tag{49}$$

For the metric (34), the non-zero Christoffel symbols are given by

$$\begin{aligned} \Gamma^{\phi}_{r\phi} &= \sqrt{2\lambda}\cot\left[\sqrt{2\lambda}r\right], \\ \Gamma^{r}_{\phi\phi} &= -\sqrt{\frac{\lambda}{2}}\sin\left[2\sqrt{2\lambda}r\right] \end{aligned}$$

which leads to the following Ricci scalar

$$R = \frac{2|\lambda|}{j} \tag{50}$$

which indicates the positive constant curvature. Moreover, we define the following parametric equations as follows

$$x_0 = \sqrt{\frac{j}{2}}\cos\left[\sqrt{2\lambda}r\right] \tag{51}$$

$$x_1 = \sqrt{\frac{j}{2}}\sin\left[\sqrt{2\lambda}r\right]\cos\phi \tag{52}$$

$$x_2 = \sqrt{\frac{j}{2}}\sin\left[\sqrt{2\lambda}r\right]\sin\phi \tag{53}$$

which gives a spherical space embedded in the three dimension, i.e.,

$$x_0^2 + x_1^2 + x_2^2 = \frac{j}{2} \tag{54}$$

■

Lemma 5. *For each fixed λ, the Kerr coherent states give rise to a resolution of unity,*

$$\int d\mu\, |\alpha, \lambda\rangle \langle\alpha, \lambda| = \mathbb{1} \tag{55}$$

with the corresponding invariant measure i.e.,

$$d\mu = \frac{2j+1}{2\pi}\sqrt{\frac{|\lambda|}{2}}\sin\left[\sqrt{2|\lambda|}r\right] dr d\phi, \lambda < 0, \tag{56}$$

$$d\mu = \frac{2j-1}{2\pi}\sqrt{\frac{\lambda}{2}}\sinh\left[\sqrt{2\lambda}r\right] dr d\phi, \lambda > 0, \tag{57}$$

$$\tag{58}$$

Proof. By substituting the relation (30) into the right side of the above equation the above equation and the following identity

$$\delta_{n,m} = \frac{1}{2\pi}\int_0^{2\pi} d\phi e^{-i\phi(n-m)} \tag{59}$$

where $\delta_{n,m}$ denotes Kronecker-Delta, we have

$$\begin{aligned}
& \sqrt{2\lambda}\sum_{n=0}^{2j}\frac{(2j+1)!}{(2j-n)!n!}\int_0^{\frac{\pi}{\sqrt{2\lambda}}} dr\sin\left[\sqrt{2|\lambda|}r\right] \\
\times\quad & \cos^{4j}\left[\sqrt{\frac{\lambda}{2}}r\right]\tan^{2n}\left[\sqrt{\frac{\lambda}{2}}r\right]|n\rangle\langle n| \\
= \quad & \sum_{n=0}^{2j}\frac{(2j+1)!}{(2j-n)!n!}\int_0^{\pi} d(\cos\vartheta)\left(\frac{1+\cos\vartheta}{2}\right)^{2j-n} \\
\times\quad & \left(\frac{1-\cos\vartheta}{2}\right)^{n}|n\rangle\langle n| \\
= \quad & \sum_{n=0}^{2j}\frac{(2j+1)!}{(2j-n)!n!}\int_0^{1} dx x^{2j-n}(1-x)^n|n\rangle\langle n| \\
= \quad & \sum_{n=0}^{2j}|n\rangle\langle n| = \mathbb{1}.
\end{aligned}$$

In the last step of the proof, we applied:

$$\int_0^1 dx x^{2j-n}(1-x)^n = \frac{(2j-n)!n!}{(2j+1)!} \tag{60}$$

By substituting the relation (29) into the right side of the relation (55), we obtain

$$\begin{aligned}
& \sqrt{2\lambda}\sum_{n=0}^{\infty}\frac{\Gamma(2j+n)}{\Gamma(2j-1)n!}\int_0^{\infty} dr\sinh\left[\sqrt{2|\lambda|}r\right] \\
\times\quad & \cosh^{-4j}\left[\sqrt{\frac{\lambda}{2}}r\right]\tanh^{2n}\left[\sqrt{\frac{\lambda}{2}}r\right]|n\rangle\langle n| \\
= \quad & \sum_{n=0}^{\infty}\frac{\Gamma(2j+n)}{\Gamma(2j-1)!n!}\int_1^{\infty} d(\cosh\vartheta)\left(\frac{\cosh\vartheta+1}{2}\right)^{-2j-n} \\
\times\quad & \left(\frac{\cosh\vartheta-1}{2}\right)^{n}|n\rangle\langle n| \\
= \quad & \sum_{n=0}^{2j}\frac{\Gamma(2j+n)}{\Gamma(2j-1)n!}\int_1^{\infty} dx x^{-2j-n}(1-x)^n|n\rangle\langle n| \\
= \quad & \sum_{n=0}^{2j}|n\rangle\langle n| = \mathbb{1}.
\end{aligned}$$

in which the following identity is applied:

$$\int_1^{\infty} dx x^{-2j-n}(1-x)^n = \frac{\Gamma(2j-1)n!}{\Gamma(2j+n)} \tag{61}$$

■

Definition A.3. *A Kerr kernel can be defined as the overlap of two Kerr coherent states:*

$$K(\alpha_1,\alpha_2) = \langle\alpha;\lambda|\alpha;\lambda\rangle \tag{62}$$

where direct calculation gives the following result for the positive Kerr parameter

$$K_{\lambda^+}(\alpha_1,\alpha_2)=\frac{\left(\operatorname{sech}^2\left[\sqrt{\frac{\lambda}{2}}r_1\right]\operatorname{sech}^2\left[\sqrt{\frac{\lambda}{2}}r_2\right]\right)^j}{\left[1-e^{i(\theta_1-\theta_2)}\tanh\left(\sqrt{\frac{\lambda}{2}}r_1\right)\tanh\left(\sqrt{\frac{\lambda}{2}}r_2\right)\right]^{2j}}. \tag{63}$$

and for $\lambda\in\mathbb{R}^-$, is given by

$$K_{\lambda^-}(\alpha_1,\alpha_2)=\frac{\left[1+e^{i(\theta_1-\theta_2)}\tan\left[\sqrt{\frac{|\lambda|}{2}}r_1\right]\tan\left[\sqrt{\frac{|\lambda|}{2}}r_2\right]\right]^{2j}}{\left(\sec^2\left[\sqrt{\frac{|\lambda|}{2}}r_1\right]\sec^2\left[\sqrt{\frac{|\lambda|}{2}}r_2\right]\right)^j}. \tag{64}$$

Lemma 6. *The Kerr kernels exhibit the property of reproducibility, i.e.,*

$$K(\alpha_1,\alpha_2)=\int d\mu(\alpha)K(\alpha_1,\alpha)K(\alpha,\alpha_2) \tag{65}$$

Proof. Using 5 leads to the result directly. ■

B Kerr Coherent States' Implementation

In the following, we develop two methods to implement Kerr coherent states. The first method involves a quantum optics setup, while the second method is based on simulating the Kerr coherent states in a set of coupled waveguides.

B.1 Generation of Kerr coherent states

In our analysis, we examine a Hamiltonian that describes the interaction between a two-level atom and a single-mode cavity field. This interaction is subject to an intensity-dependent coupling, along with the influence of an external classical field. By taking the account the resonance conditions and apply the rotating-wave approximation, the Hamiltonian can be expressed in the following Solano et al. (2003); Zou et al. (2004); Miry & Tavassoly (2012):

$$\begin{aligned}\hat{H} &= g\left(e^{i\varphi}\sigma_- + e^{-i\varphi}\sigma_+\right)\\ &+ \Omega\left(\sigma_-\hat{A}^\dagger+\sigma_+\hat{A}\right)\end{aligned} \tag{66}$$

where parameters g and Ω are the coupling coefficients of the atom with classical and quantized cavity fields, $\sigma_\pm$ is the ladder operator, i.e., $\sigma_-=|e\rangle\langle g|$ and $\sigma_+=|g\rangle\langle e|$ in which $|e\rangle$ and $|g\rangle$ are respectively ground and excited states of the atom; $\hat{A}$ and $\hat{A}^\dagger$ are the Kerr annihilation and creation operators and φ is the phase of classical field. By using the following transformation:

$$R=\exp\left[\frac{\pi}{4}(\sigma_+-\sigma_-)\right]\exp\left[\frac{i\varphi}{2}\sigma_z\right] \tag{67}$$

by which

$$R\sigma_\pm R^\dagger=e^{\pm i\varphi}2\left[\sigma_z\pm(\sigma_+-\sigma_-)\right] \tag{68}$$

the Hamiltonian (66) becomes

$$\begin{aligned}H_{tran} &= RHR^\dagger=g\sigma_z\\ &+ \frac{\Omega}{2}\Big[e^{-i\varphi}\hat{A}^\dagger\left(\sigma_z-(\sigma_+-\sigma_-)\right)\\ &+ e^{i\varphi}\hat{A}\left(\sigma_z+(\sigma_+-\sigma_-)\right)\Big]\end{aligned} \tag{69}$$

Hence, in the interaction picture, by using the unitary operator $\hat{T}(t) = \exp[ig\sigma_z t]$ the time evolution of H_{tran} is given by

$$\begin{aligned} H_{int} &= e^{ig\sigma_z t} H_{tran} e^{-ig\sigma_z t} \\ &= \frac{\Omega}{2}\Big[e^{-i\varphi}\hat{A}^\dagger\left(\sigma_z - \left(\sigma_+ e^{2igt} - \sigma_- e^{-2igt}\right)\right) \\ &+ e^{i\varphi}\hat{A}\left(\sigma_z + \left(\sigma_+ e^{2igt} - \sigma_- e^{-2igt}\right)\right)\Big] \end{aligned} \tag{70}$$

In the strong classical field regime, i.e., $g \ll \Omega$, e the term that oscillates with high frequencies in the Hamiltonian is eliminated

$$H_{eff} = \frac{\Omega}{2}\left[e^{-i\varphi}\hat{A}^\dagger + e^{i\varphi}\hat{A}\right]\sigma_z \tag{71}$$

The time evolution operator of Hamiltonian (71) is given by

$$U_{eff} = \exp\left[\frac{-i\Omega t}{2}\left[e^{-i\varphi}\hat{A}^\dagger + e^{i\varphi}\hat{A}\right]\sigma_z\right] \tag{72}$$

which is treated as a Kerr displacement operator. We are able to come back to the standard time evolution, i.e., the time evolution is corresponded to the the Hamiltonian (66), by using following transformations:

$$\begin{aligned} U(t) &= \hat{R}^\dagger\hat{T}^\dagger(t)\hat{U}_{eff}\hat{T}(0)\hat{R} \\ &= \exp\left[-igt\left(\sigma_- e^{i\varphi} + \sigma_+ e^{-i\varphi}\right)\right] \\ &\times \exp\left[\frac{-i\Omega t}{2}\left(e^{-i\varphi}\hat{A}^\dagger + e^{i\varphi}\hat{A}\right)\left(\sigma_- e^{i\varphi} + \sigma_+ e^{-i\varphi}\right)\right] \end{aligned} \tag{73}$$

Now, by assuming that the cavity is initially prepared in the vacuum of the field and the atoms are in a superposition of excited and ground states, i.e.,

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|g\rangle + |e\rangle\right)|0\rangle \tag{74}$$

Using relation (67), we have

$$\hat{R}|\psi\rangle = \left[\cos\frac{\varphi}{2}|e\rangle - i\sin\frac{\varphi}{2}|g\rangle\right]|0\rangle \tag{75}$$

Hence, the simple calculations leads into the final result:

$$|\psi(t)\rangle = \frac{1}{\sqrt{2}}\Big[e^{-i\varphi/2}e^{-igt}\cos\frac{\varphi}{2}|\alpha;\lambda,j\rangle \tag{76}$$

$$+ ie^{-i\varphi/2}e^{igt}\sin\frac{\varphi}{2}|-\alpha;\lambda,j\rangle\Big]|e\rangle$$

$$+ \frac{1}{\sqrt{2}}\Big[e^{-i\varphi/2}e^{-igt}\cos\frac{\varphi}{2}|\alpha;\lambda,j\rangle \tag{77}$$

$$- ie^{-i\varphi/2}e^{igt}\sin\frac{\varphi}{2}|-\alpha;\lambda,j\rangle\Big]|g\rangle \tag{78}$$

in which $\alpha = \frac{-i\Omega t}{2}e^{-i\varphi}$. Therefore, by setting $\varphi = 0$, and measuring the atom state, the Kerr coherent state is obtained.

B.2 Photonic Lattice Realization of Kerr Coherent States

Optics typically involves the study of electromagnetic waves as continuous functions of space and time. However, there are instances where the evolution of an optical field can be treated as a discrete problem. A notable example of this is found in the study of light in a coupled 1D waveguide array. In such an array,

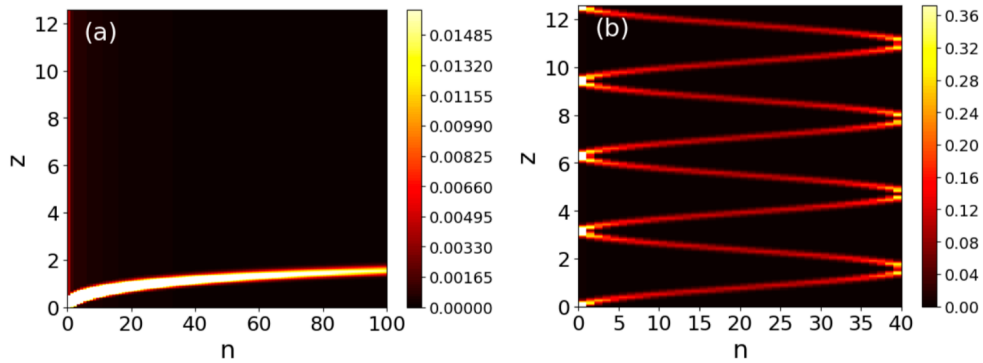

Figure 5: The intensity evolution in a waveguide lattice for $|\lambda| = 2$ and $j = 20$ during n waveguides is depicted in plots (a) and (b) for the positive and negative Kerr coherent states, respectively.

numerous (potentially infinite) single-mode channel waveguides are arranged side by side, causing their individual modes to overlap. This arrangement leads to coupling between each waveguide and its nearest neighbors, significantly influencing the propagation of light in these waveguide arrays. In this scenario, the evolution of the optical mode electric field amplitude, denoted as E_n, at the nth site of the array, follows a discrete linear Schrödinger-like equation is given by:

$$i\frac{dE_n}{dz} + C_n E_{n-1} + C_{n+1} E_{n+1} = 0 \tag{79}$$

where in the weak coupling regime, C_n depends exponentially on the distance between the guides, i.e., $C_n = C_1 \exp\left[-(d_n - d_0/\kappa)\right]$, in which d_1 and κ are fitting parameter Keil et al. (2011); Dehdashti et al. (2015); Christodoulides et al. (2003). By choosing the following distance between the first site and nth site:

$$d_n = d_0 - \kappa \ln\left[\frac{|\lambda|\sqrt{2j \mp 1 \pm n}}{2C_0}\right] \tag{80}$$

the coefficient C_n is given by

$$C_n = \frac{|\lambda|}{2}\sqrt{2j \mp 1 \pm n} \tag{81}$$

Now, we can write the differential equation (79) as the follows:

$$i\frac{d\Psi}{dz} = -(\hat{A} + \hat{A}^\dagger)\Psi \tag{82}$$

where $\Psi = \sum_{n=0}^{s} E_n \left|n\right\rangle$. Therefore, we have

$$\Psi(z) = e^{i(\hat{A}+\hat{A}^\dagger)z}\left|0\right\rangle \tag{83}$$

the sign $\pm$ implies the sign of the Kerr parameter. Therefore, By selecting specific values for z corresponding to each data point, we can effectively encode the data into the photonic lattice. Fig 5, plots (a) and (b), illustrate light propagation in the waveguide lattice for the positive and negative Kerr parameters.

C Datasets description

In this Appendix section we describe in detail the construction of the synthetic and real-world datasets used for the benchmarking of the Kernels covered in Section 5.

C.1 Synthetic datasets used for the Kernel behavior characterization

This are the details of the datasets used in Section 5.1.1.

Circles This data set was generated using `make_circles`, producing two concentric circular classes with additive noise. It provides a standard benchmark for nonlinear decision boundaries. The data set consists of 500 samples, with noise level 0.7 and inner-circle scaling factor 0.5.

Double layers This custom data set consists of two horizontal layers, each containing samples from both classes. Four Gaussian clusters were centered at the edges of the unitary square, where the left clusters belong to class 0 and the right clusters belong to class 1. This arrangement creates two vertically stacked classification layers. Each cluster has a Gaussian spread of 1.5 standard deviation consisting of 500 samples in total.

Hypercubes This data set was generated using `make_classification` in an eight-dimensional feature space. All features were informative, with no redundant or repeated variables. Samples were distributed according to a hypercube structure, producing a higher-dimensional benchmark with moderate label noise. The data set consists of 500 samples with 8 features, all of them informative, class separation 1.2, and label noise of 2%.

Moons The moons data set was generated using `make_moons`, producing two interleaved crescent-shaped classes. This benchmark is widely used to test nonlinear classification performance. The data set consists of 500 samples with additive noise level 0.3.

C.2 Synthetic datasets used for the Kernel performance benchmark

In this Appendix section we detail the synthetic datasets used for the benchmarking of Section 5.2.1. Representative examples of the generated datasets are shown in Appendix D.1

Annular Circles This synthetic data set consists of two classes distributed according to concentric radial regions in a two-dimensional plane. Samples were uniformly generated within the square domain $[-1, 1] \times [-1, 1]$. The negative class corresponds to points lying inside an annular region bounded by the radii $r_{\text{in}} = 0.5\sqrt{\frac{2}{\pi}}, r_{\text{out}} = \sqrt{\frac{2}{\pi}}$, while points outside this annulus belong to the positive class. This benchmark probes the ability of the kernels to resolve disconnected radial decision boundaries and nonlinearly separable structures. The data set consists of 2000 samples in a two-dimensional feature space.

Corners This synthetic data set was constructed from four circular corner regions centered at the vertices of a square. Samples were uniformly generated within the square domain $[-1.5, 1.5] \times [-1.5, 1.5]$. Four circular regions of radius 0.75 were centered at the coordinates $(-1, -1), \quad (-1, 1), \quad (1, -1), \quad (1, 1)$. Points located inside any of these corner regions were assigned to the negative class, while all remaining points were assigned to the positive class. This benchmark generates a disconnected classification geometry with multiple localized regions and sharp transitions between classes. The data set consists of 2000 samples in two dimensions.

Sinus The sinus data set was generated by sampling points uniformly in the square domain $[-1, 1] \times [-1, 1]$. Class labels were assigned according to a sinusoidal decision boundary defined by $y = -0.8 \sin(\pi x)$, where points below the curve belong to the positive class and points above the curve belong to the negative class. This benchmark probes the ability of the kernels to learn smooth periodic nonlinear decision boundaries. The data set consists of 2000 samples in a two-dimensional feature space.

Spiral This synthetic benchmark consists of two interlaced spiral-shaped classes embedded in a two-dimensional plane. Each class was generated from a parametric spiral of the form $r = 0.5 + 0.18\,\theta$, with angular parameter $\theta \in [0, 2\pi]$. The second class was constructed by reflecting the first spiral through the

origin, producing two intertwined structures with opposite orientations. Additive Gaussian noise with standard deviation 0.1 was applied independently to both spatial coordinates. This benchmark is particularly challenging for kernel methods due to the highly nonlinearly separable winding geometry of the decision boundary. The data set consists of 2000 samples equally distributed between both classes.

C.3 Real world datasets used for the Kernel performance benchmark

The following real-world datasets were used for benchmarking of Section 5.2.2:

Framingham CHD data set The Framingham cardiovascular data set Janosi et al. (1989) was used to predict the sex of the patient from medical attributes selected from the available clinical features (14 in total) related to the diagnosis of chronic heart disease. The benchmarking methodology chosen in this work to closely resembles that of Ref. Swaminathan & Akgün (2025), allowing a direct and meaningful comparison of classification performance. All of the 4240 data points were taken. Points with incomplete features were encoded accordingly to the kernel in the matter of study with only their available features.

Fashion-MNIST The Fashion-MNIST data set Xiao et al. (2017) was used as a high-dimensional image-classification benchmark. Each grayscale image of size 28×28 pixels was flattened into a feature vector of dimension 784. Only the classes corresponding to dresses and shirts were retained, resulting in a binary classification task. The benchmarking procedure was selected to be comparable to that employed in Ref. Rodriguez-Grasa et al. (2025).

C.4 Datasets description for additional kernels

The datasets setup used for the benchmark in Section 5.2.3 remains the same as the ones used for previous benchmarks with synthetic datasets of Sections 5.1.1 and 5.2.1.

D Further Experiments on Additional Datasets

To assess the practical effectiveness of Kerr kernels, we conduct a systematic benchmark across *nine* binary classification tasks and compare their performance against a range of established kernel methods. Six of the tasks involve analytically tractable two-dimensional geometries, namely, Sinus, Corners, Circles, Disks-More, Disks-3, and Moons, while the remaining three evaluate performance in more realistic settings, i.e., a medium-scale tabular classification task Classifications and two low-dimensional image projection tasks Fashion-MNIST and MNIST. Further details of the datasets used in this Appendix section are described in Appendix D.1.

The raw samples $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ are split once into stratified training and test partitions, $\mathcal{D}_{\text{tr}}$ and $\mathcal{D}_{\text{te}}$, such that class priors are preserved exactly. Additionally, the features in each dataset are linearly rescaled to the interval $[-1, 1]$ for non-periodic datasets. However, for periodic datasets, they are rescaled to $[-\pi/2, \pi/2]$.

Model selection is performed by a fixed five-fold stratified cross-validation on D_{tr}. Within each fold, the full Kerr-kernel Gram matrix is computed once, and the associated support-vector problem is solved to convergence. Balanced accuracy, averaged over the five folds, serves as the sole selection metric and provides an unbiased proxy for generalization performance. Moreover, we consider two different noise; a deterministic case and a stochastic case obtained by injecting zero-mean Gaussian fluctuations into the amplitude parameter $|\alpha|$.

The tunable parameters of the Kerr kernel, $\theta = \{\lambda,\, j,\, \alpha\}$, together with the SVM soft margin parameter C, are jointly optimized using a Tree-structured Parzen Estimator (TPE) algorithm augmented with *Hyperband* early stopping. The search space is defined as $|\lambda| \in \mathbb{R}^+$, j restricted to half-integer values, $\alpha \in (10^{-3}, 5.5)$, and $C \in [10^{-3}, 10]$. For the squeezing and RBF kernel baselines, the corresponding hyperparameter domains are appropriately restricted to reflect their reduced parameterization. In Appendix D.2 the algorithm and procedure of the experiment is described in detail, then experimental the results are shown in Appendix D.3 followed by a discussion at the same Appendix subsection.

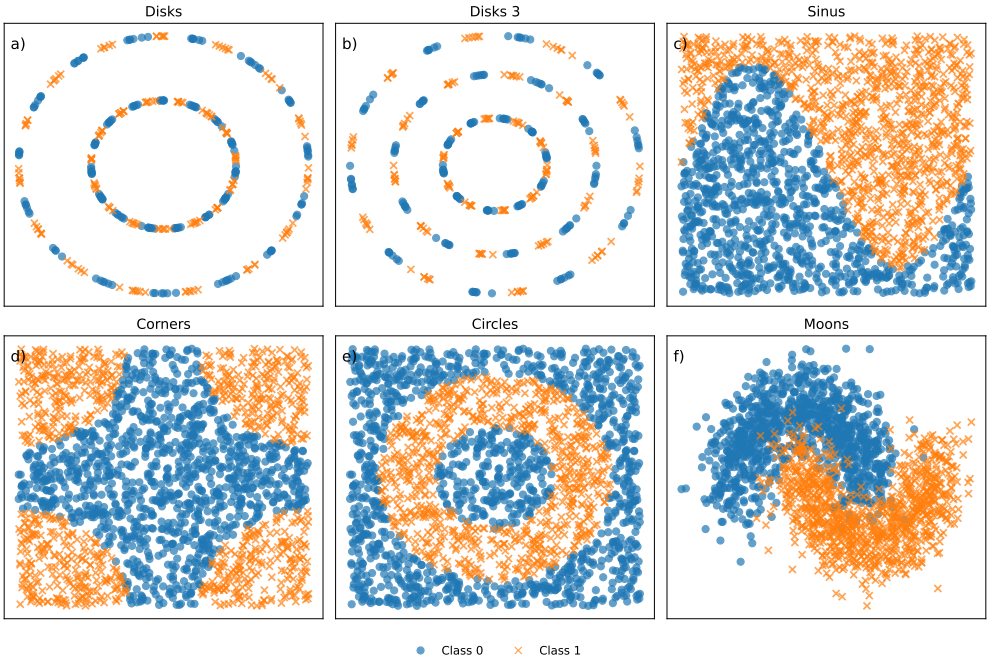

Figure 6: Visual representation of some of the datasets used in the Appendix D

D.1 Additional datasets

The nine benchmark datasets, summarized in Table 5, cover a broad spectrum of complexity, from analytically defined 2-D geometries to realistic image-based tasks, thus forming a rigorous and diverse testbed for evaluating Kerr–coherent kernels. Figure 6 presents schematic visualizations of the first six datasets which are synthetic, hand-crafted problems designed to highlight distinct geometric properties, including global periodicity, angular discontinuities, radial symmetry, and structured noise at varying spatial scales. These toy datasets exhibit perfectly balanced class priors and controlled noise, allowing performance differences to be attributed directly to the expressivity and regularization properties of the kernels. The remaining three datasets introduce increasing degrees of real-world complexity.

We now outline the characteristics of each dataset:

Table 5: Experimental data sets. $N_{\text{train}}/N_{\text{test}}$ lists the split sizes used in this work, d is the feature dimension fed to the kernel, "bal." states whether the classes are perfectly balanced (B) or label-noisy (LN), "geometry / structure" summarises the decision-boundary type, and the last column specifies the sole source of stochastic noise, if any.

Name	N_{train} / N_{test}	d	classes	bal.	geometry / structure	noise source
Sinus	1500 / 500	2	2	B	periodic sine curve	none
Corners	1500 / 500	2	2	B	4 quarter-circles	none
Circles	1500 / 500	2	2	B	thin annulus	none
Disks-More	1500 / 500	2	2	B	two concentric sector rings	angle $\mathcal{N}(0, 0.05^2)$
Disks-3	270 / 90	2	2	B	three concentric sector rings	angle $\mathcal{N}(0, 0.05^2)$
Moons	1500 / 500	2	2	B	two crescents	coord. $\mathcal{N}(0, 0.25^2)$
Classifications	1500 / 500	20	2	LN	mixed informative / redundant	10% label flip
Fashion-MNIST (3 vs 6)	800 / 200	784	2	B	PCA-3 image (dress/shirt)	none
MNIST (0 vs 3)	500 / 300	3	2	B	PCA-3 image (digits)	none

Sinus. A total of 2×10^3 points are uniformly sampled from $[-1, 1]^2$, with positive labels assigned to those below the curve $y = -0.8 \sin(\pi x)$, yielding a smooth, periodic decision boundary with zero Bayes risk. The boundary contains a single spatial frequency, making the dataset ideal for testing kernels that encode global phase information. The class prior is exactly balanced, and no extrinsic noise is added, ensuring that performance degradation stems purely from the kernel's ability to model curvature rather than stochastic label variation.

Corners. This dataset consists of four quarter-circles of radius $r = 0.75$ placed at the corners of the square domain $[-1, 1]^2$. Points within these arcs are labelled positive, while those in the central diamond region ($|x_1| + |x_2| < r$) receive the negative label. The resulting piecewise-linear decision boundary includes four right-angle junctions, posing a challenge to kernels biased toward smooth curvature, while favouring those that can capture abrupt geometric transitions. The class distribution is exactly balanced and no noise is introduced, ensuring zero Bayes risk and isolating the kernel's expressivity as the sole performance factor.

Circles. The negative class is defined as a narrow annulus with inner and outer radii $r_{\text{in}} = \frac{1}{2}\sqrt{2/\pi}$ and $r_{\text{out}} = \sqrt{2/\pi}$, such that its area matches that of the complementary positive region. Sampling 2×10^3 points uniformly from $[-1, 1]^2$ yields a perfectly balanced binary task with a ground-truth boundary of constant radial curvature. This setup serves as a benchmark for rotational invariance: kernels whose geometry aligns with polar symmetries should achieve near Bayes-optimal performance, while those biased toward Cartesian axes typically degrade as the ring width nears the kernel bandwidth. No label noise is added, ensuring that all accuracy loss reflects geometric limitations of the kernel.

Disks-More. A noisy double-ring configuration composed of two concentric circles with radii 0.9 and 0.45, each subdivided into 30 angular sectors. Per sector, one anchor point and five jittered samples (drawn from $\mathcal{N}(0, 0.05)$ radians) are generated, yielding 6×30 samples per ring and 360 total points. Class labels alternate by sector on the outer ring and are inverted on the inner ring, forming a balanced binary task. The Bayes boundary is piecewise-constant in angle, with structured label noise. All coordinates are rescaled to $[-\pi/2, \pi/2]^2$ to preserve the radial structure during kernel evaluation.

Disks-3. This benchmark generalizes the radial sector setup to three concentric rings with radii 1, 0.7, and 0.36, each divided into 24 angular sectors. Points are sampled at sector centers and perturbed with $\mathcal{N}(0, 0.05)$ radian noise to smooth transitions. Labels alternate on the outer ring, invert on the middle ring, and receive a fixed offset -0.5 on the inner ring. A final binary label merges the outer and inner rings as positive and assigns the middle ring as negative, forming a balanced, non-convex classification task. Coordinates are rescaled to $[-\pi/2, \pi/2]^2$ to preserve radial geometry. This setup evaluates a kernel's ability to model multi-scale curvature and nested label heterogeneity.

Moons. This benchmark consists of 2000 samples drawn from two interleaved crescents, with additive Gaussian noise of standard deviation 0.25. A stratified split allocates 25% to the test set, leaving 1500 training points. The non-convex structure and noise-induced class overlap test a kernel's ability to trade off margin maximization and variance control. Features are linearly rescaled to $[-\pi/2, \pi/2]$ to ensure full phase coverage in the Kerr kernel, and class priors are perfectly balanced by design.

Classifications. This synthetic tabular task is generated using `make_classification` from Scikit-Learn with $N_{\text{tot}} = 2000$ samples and $d = 20$ features. Of these, 8 are informative, 4 are redundant, and the remaining 8 are pure noise. A 10% label-flip rate induces a Bayes error of roughly 5%, probing robustness to mislabeling. Data are stratified into a 75/25% train-test split, and features are affinely rescaled to $[-\pi/2, \pi/2]$ to equalize phase contributions in the Kerr map. The weakly informative structure ensures that improvements on low-dimensional geometric sets do not degrade performance in high-dimensional, noisy tabular settings.

Fashion-MNIST (classes 3 vs 6). Dataset comprises 70,000 grayscale 28×28 images across ten apparel classes. We extract a balanced binary subset of 500 training and 300 test samples from class 3, that is "dress", and class 6, that is "shirt", which requires distinguishing fine-grained shape differences. Images

are flattened into 784-dimensional vectors, min–max normalized to $[0, 1]$, and projected onto the top three principal components to maintain nonlinear structure while aligning with the study's low-dimensional toy setups. The resulting 3D embeddings are rescaled to $[-1, 1]^3$ prior to kernel evaluation, ensuring uniform contribution to the Kerr phase term.

MNIST (digits 0 vs 3). From the original MNIST dataset, we extract 800 grayscale 28×28 images corresponding to digits 0 and 3, allocating 500 for training and 300 for testing via a stratified split. Each image is flattened into a 784-dimensional vector, normalized to $[0, 1]$, and reduced to three principal components that predominantly capture loop vs. arc morphology. These 3D embeddings are then affinely scaled to $[-1, 1]^3$ prior to kernel evaluation to maintain consistency with the toy datasets. This benchmark assesses whether the curvature-aware Kerr kernels can approximate near-optimal decision boundaries despite aggressive dimensionality reduction on a standard vision task.

D.2 Pipeline Flow and Dataset processing descriptions

Algorithm 1 outlines the full pipeline. Details of the simulation setup and methodological protocol, including preprocessing procedures, and search space configurations, are provided in Appendix D.2.1 and D.2.2.

D.2.1 Data partitions and global workflow

Every experiment begins with a single stratified split of the raw dataset $\mathcal{D}$ into a *training* set $\mathcal{D}_{\text{tr}}$ and a *held-out test* set $\mathcal{D}_{\text{te}}$, ensuring that class priors are preserved exactly. No dedicated validation set is carved out; instead, model selection is conducted via *five-fold stratified cross-validation* (CV) on $\mathcal{D}_{\text{tr}}$. Specifically, the training data are partitioned once into five disjoint folds $\mathcal{F}_k \subset \mathcal{D}_{\text{tr}}$ for $k = 1, \ldots, 5$, with each fold serving in turn as a temporary validation slice while the remaining four form the corresponding training slice.

Feature scaling is applied prior to hyperparameter tuning using the affine transformation:

$$\begin{aligned} \widetilde{x}_j &= a_j x_j + b_j, \\ a_j &= \frac{L_{\max} - L_{\min}}{\max_j - \min_j}, \\ b_j &= L_{\min} - a_j \min_j, \end{aligned}$$

where $\min_j$ and $\max_j$ denote the minimum and maximum values of the j-th feature, computed on the *training* set only. The range $[L_{\min}, L_{\max}]$ defines the desired target interval, which is set to either $[-1, 1]$ for unbounded features or $[-\pi/2, \pi/2]$ for periodic inputs. The computed coefficients a_j and b_j are then applied unchanged to the corresponding validation and test samples to prevent data leakage.

Since the class distribution is balanced and nearly identical across all datasets, we adopt the *balanced accuracy* (BA) as the primary performance metric. It is defined as

$$\text{BA} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}\right), \tag{84}$$

The five-fold cross-validation score used for model selection is

$$\widehat{R}(\theta) = \frac{1}{K}\sum_{k=1}^{K} \text{BA}_k(\theta), \tag{85}$$

where $K = 5$ is the number of folds, and TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively. Balanced accuracy serves as the main objective to be maximized during the hyperparameter optimization phase. The next subsection provides a detailed description of the fine-tuning procedure.

Algorithm 1: Quantum-enhanced SVM pipeline using Kerr-evolved coherent states

Input : Classical dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$;
hyperparameter search space Θ;
optional amplitude noise level $\sigma \geq 0$.

Output: Optimal hyperparameters $\theta^\star \in \Theta$;
trained SVM parameters $(\boldsymbol{\alpha}^\star, b^\star)$.

Step 1: Quantum State Encoding & Preprocessing
`// Optional dimensionality reduction for high-dimensional input`
if *image task* **then**
  Apply PCA and retain top d components
`// Normalize features to a target quantum encoding range`
Linearly rescale each input dimension to a phase or amplitude range (e.g., $[-\pi/2, \pi/2]$ or $[-1, 1]$)
`// Classical-to-quantum encoding`
For each data point $\mathbf{x}_i$, encode in the amplitude or phase of a Kerr coherent state;

Step 2: Quantum Kernel Hyperparameter Optimization
for $k \leftarrow 1, \ldots, K$ **(cross-validation folds) do**
  Split $\mathcal{D}$ into $\mathcal{D}_k^{\text{tr}}$ and $\mathcal{D}_k^{\text{val}}$
foreach $\theta \in \Theta$ **(proposed by TPE + Hyperband) do**
  for $k \leftarrow 1, \ldots, K$ **do**
    `// Optional amplitude noise to simulate experimental imperfections`
    if $\sigma > 0$ **then**
      Draw $\delta \sim \mathcal{N}(0, \sigma^2)$
      Perturb amplitude: $\alpha \leftarrow |\alpha + \delta|$
    `// Quantum kernel evaluation`
    Compute Gram matrix: $K_{ij}^{(k)} = \langle \psi(\mathbf{x}_i^{\text{tr}}; \theta) | \psi(\mathbf{x}_j^{\text{tr}}; \theta) \rangle$
    Train soft-margin SVM on $\mathcal{D}_k^{\text{tr}}$
    Evaluate balanced accuracy $\text{BA}_k(\theta)$ on $\mathcal{D}_k^{\text{val}}$
  Aggregate: $\widehat{R}(\theta) = \frac{1}{K} \sum_{k=1}^K \text{BA}_k(\theta)$
Select optimal hyperparameters: $\theta^\star = \arg\max_{\theta \in \Theta} \widehat{R}(\theta)$

Step 3: Final Quantum Kernel Training and Inference
Compute final Gram matrix: $K_{ij} = \langle \psi(\mathbf{x}_i; \theta^\star) | \psi(\mathbf{x}_j; \theta^\star) \rangle$
Train SVM on full dataset $\mathcal{D}$ to obtain $(\boldsymbol{\alpha}^\star, b^\star)$
`// Prediction for a new sample x`
Encode $\mathbf{x}$ as $|\psi(\mathbf{x}; \theta^\star)\rangle$

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^N \alpha_i^\star \, y_i \, \langle \psi(\mathbf{x}_i; \theta^\star) | \psi(\mathbf{x}; \theta^\star) \rangle + b^\star \right)$$

D.2.2 Hyperparameter Tuning

Search Space. We optimize over the curvature branch λ, amplitude α, half-integer index j, and soft-margin penalty C, with reduced spaces $\{C, \texttt{sq_mag}\}$ and $\{C, \gamma\}$ for squeezing and RBF baselines. All variables are continuous except j.

Bayesian Optimization. Hyperparameters are selected via the Tree-structured Parzen Estimator (TPE) Watanabe (2023) in Optuna Akiba et al. (2019), which maintains density models for high- and low-performing configurations. The acquisition rule favors regions with high likelihood under the "good" set and low likelihood under the "bad" set, balancing exploration and exploitation without unimodality assumptions.

Validation Metric. The objective is the mean balanced-accuracy over five fixed stratified folds, reducing variance, ensuring unbiased generalization estimates, and maintaining identical train/validation splits for all trials.

Early Stopping and Convergence. Hyperband pruning terminates underperforming trials early, reallocating computation to promising candidates. Each TPE suggestion has a 5-min budget with $R = 25$ epochs and $\eta = 3$ down-sampling; $\sim$70% of trials are pruned. Search stops when the budget expires or no improvement beyond 10^{-3} is observed in 50 consecutive trials, yielding 5×10^2–1.2×10^3 evaluations.

Kernel Freezing and Final Model. The optimal parameters define a fixed Gram matrix for all subsequent steps, preventing train/test leakage. The final soft-margin SVM is trained in the induced feature space via the dual formulation, and predictions use the optimized kernel.

D.3 Results and discussions

Across nine benchmarks the Kerr family matches or exceeds all baselines while retaining markedly lower variance. In table 6, the cross-validation (CV) statistics and the held-out test balanced-accuracy scores obtained with the family of Kerr–coherent kernels and with two widely used baselines, namely, RBF and squeezing, are introduced. The first number always represents the mean CV score during the fine-tuning process, either in noisy or ideal simulation. Unless stated otherwise, all numbers refer to the *frozen best hyper-parameters* selected by the five-fold CV procedure described above.

Overall performance trends. Across the nine benchmark datasets, the proposed kernels attain or match the best test accuracy in six cases, while either the RBF or the squeezing kernel leads in the remaining three. The superiority of the Kerr family is most pronounced on geometry-rich synthetic tasks: Circles, Corners, Disks_3, and Sinus all exhibit perfect or near-perfect performance ($>$99%) with at least one Kerr variant, whereas the classical RBF lags by up to 1.3 pp on Circles and fails catastrophically ($\sim$ 53%) on the multi-scale Disks_3 configuration. On real-world images the gap narrows but remains favourable: for the MNIST task the negative-curvature branch reaches 95.2% and statistically ties with squeezing, while on Fashion-MNIST Dress/Shirt it delivers 91.6% against 91.8% for squeezing and 91.6% for RBF. The sole data set where the RBF baseline edges ahead is the noisy two-crescent *Moons*; even there the difference is marginal ($\leq$ 0.3 pp) and well inside one standard deviation of the CV estimate.

Stability and hyper-parameter efficiency. Whenever a baseline kernel outperforms the Kerr kernel family, the improvement is accompanied by a notably higher variance. For instance, on the Moons dataset, the RBF kernel exhibits a cross-validation standard deviation of ±0.007, which is comparable to that of the Kerr models. However, its mean cross-validation accuracy of 94.53 % remains below the Kerr peak performance of 94.80 %, indicating a tendency toward mild overfitting to the test set. Furthermore, for the squeezing kernel, derived from the Kerr kernel by fixing $\lambda = 2$ and $j = 1/2$, the observed standard deviation on the Corners dataset is ±0.006, exceeding that of the best-performing Kerr kernel configuration, which achieves a lower deviation of ±0.004.

Robustness to amplitude noise. To examine the impact of hardware noise, which is inevitable in quantum devices, we introduce a single-shot amplitude perturbation with standard deviation $\sigma = 0.10$. Remarkably, the influence of this noise is negligible in the case of the Kerr kernels. For instance, in the Sinus dataset, the noisy negative-branch performance improves slightly from 99.41% to 99.47%. Similarly, for the Corners and Circles datasets, the test accuracy remains unchanged at the third decimal place. The largest observed degradation is 0.30 percentage points on the Disks_more dataset; however, even in this case, the noisy model retains a margin of 2.2 percentage points over the RBF baseline. These results demonstrate that the hyperparameters j and λ, which govern the curvature and bandwidth of the feature space, confer not only strong expressivity but also numerical robustness—even under realistic experimental amplitude fluctuations in the encoding of Kerr coherent states. The error-bar analysis presents the mean and standard deviation of the scores obtained by the proposed kernels. After selecting the optimal hyperparameters through training, the datasets are partitioned as previously described, followed by the injection of random Gaussian noise

Table 6: Comparison of standard and noisy positive/negative Kerr kernels with RBF and squeezing baselines. Values are given as *mean* ± *standard error* with the cross-validated mean in parentheses; the best score per row is in bold.

Dataset	negative_phase	negative_amp	positive_phase	positive_amp	RBF	Squeezing	negative_noisy	positive_noisy
CIRCLES	99.005% ± 0.006 (98.129%)	98.952% ± 0.006 (98.129%)	99.005% ± 0.006 (98.288%)	**99.005% ± 0.006 (98.447%)**	98.152% ± 0.005 (97.969%)	98.362% ± 0.011 (98.178%)	99.005% ± 0.004 (97.700%)	99.005% ± 0.005 (98.397%)
CLASSIFICATIONS	86.257% ± 0.009 (88.591%)	85.193% ± 0.009 (88.390%)	**86.533% ± 0.009 (88.607%)**	86.057% ± 0.009 (86.982%)	85.657% ± 0.006 (87.791%)	86.926% ± 0.011 (87.385%)	86.056% ± 0.008 (87.993%)	86.536% ± 0.013 (88.393%)
CORNERS	98.876% ± 0.005 (99.231%)	**99.171% ± 0.004 (99.596%)**	98.802% ± 0.005 (99.231%)	98.876% ± 0.004 (99.231%)	98.716% ± 0.002 (98.529%)	98.789% ± 0.006 (98.750%)	98.876% ± 0.006 (99.414%)	98.764% ± 0.005 (99.231%)
DISKS_3	**99.630% ± 0.025 (100.000%)**	**99.630% ± 0.025 (100.000%)**	**99.630% ± 0.025 (100.000%)**	**99.630% ± 0.025 (100.000%)**	53.333% ± 0.045 (43.333%)	99.630% ± 0.025 (97.778%)	**99.630% ± 0.025 (100.000%)**	**99.630% ± 0.028 (100.000%)**
DISKS_MORE	98.148% ± 0.019 (95.556%)	98.148% ± 0.019 (95.556%)	97.778% ± 0.019 (95.556%)	98.148% ± 0.019 (95.556%)	48.889% ± 0.035 (44.444%)	**97.778% ± 0.030 (97.778%)**	98.148% ± 0.051 (55.556%)	98.148% ± 0.025 (94.444%)
FASHION_MNIST	91.600% ± 0.021 (90.333%)	91.600% ± 0.021 (90.000%)	91.400% ± 0.021 (90.000%)	91.400% ± 0.021 (90.333%)	91.600% ± 0.019 (89.000%)	**91.800% ± 0.031 (90.999%)**	91.600% ± 0.016 (90.000%)	91.600% ± 0.029 (90.000%)
MNIST_DIGITS	**95.195% ± 0.015 (98.693%)**	**95.211% ± 0.013 (98.693%)**	**95.195% ± 0.015 (98.693%)**	**94.999% ± 0.014 (98.693%)**	**94.999% ± 0.013 (98.693%)**	**95.195% ± 0.015 (98.693%)**	**95.195% ± 0.016 (98.693%)**	**95.195% ± 0.013 (98.693%)**
MOONS	94.733% ± 0.006 (94.200%)	94.733% ± 0.008 (94.200%)	94.800% ± 0.006 (94.000%)	94.733% ± 0.008 (94.200%)	**94.533% ± 0.007 (94.600%)**	94.933% ± 0.006 (94.000%)	94.733% ± 0.008 (94.200%)	94.733% ± 0.010 (94.200%)
SINUS	99.406% ± 0.003 (98.409%)	99.406% ± 0.004 (98.409%)	99.406% ± 0.003 (98.409%)	99.406% ± 0.004 (98.409%)	98.805% ± 0.008 (98.805%)	99.068% ± 0.006 (98.612%)	**99.471% ± 0.004 (99.012%)**	99.471% ± 0.003 (98.612%)

to assess the sensitivity of the results. As illustrated, the proposed kernels successfully retain the overall performance trend and yield well-fitted models. This behavior is consistent across all other datasets. Despite potential discrepancies between the hyperparameters that achieved high mean cross-validation (CV) scores during fine-tuning and the fixed parameters used to report standard deviation and test scores, the Kerr kernels consistently produce well-generalized models. It is also noteworthy that during the primary fine-tuning phase, the presence of noise resulted in longer convergence times compared to the idealized noise-free simulations.

In addition, we draw a comparison between embedding quantum kernel (EQK) Rodriguez-Grasa et al. (2025) and the Kerr Kernels family. Table 7 reveals a clear accuracy advantage for the Kerr–SVM across all three analytic benchmarks. On Sinus and Corners the deterministic Kerr model improves the balanced accuracy reported for the EQK by 1.6% and 2.2%, respectively, while matching the state-of-the-art on Circles. Introducing a $\sigma = 0.10$ single-shot amplitude perturbation leaves the Kerr accuracies virtually unchanged. In fact, the noisy variant attains the overall best score on Sinus, highlighting the intrinsic resilience of our photonic feature map. Beyond peak test numbers, the cross-validated means and standard

Table 7: Balanced accuracy on the three analytic toy data sets. The first two columns reproduce the best embedded Quantum Kernel (EQK) results reported in Ref. Rodriguez-Grasa et al. (2025) for the deterministic setting ($\sigma = 0$); the last column lists the Kerr–SVM accuracy when a single Gaussian amplitude noise with $\sigma = 0.10$ is injected during training.

Dataset	Paper EQK ($\sigma = 0$)	Kerr ($\sigma = 0$)	Kerr ($\sigma = 0.10$)
Sinus	0.972	0.988	**0.990**
Corners	0.974	**0.996**	0.994
Circles	0.980	**0.984**	**0.984**

deviations indicate a markedly tighter performance envelope for the Kerr models. For every dataset, the inter-fold deviation remains at or below 6×10^{-2}, substantially lower than the spread observed for the EQK baselines Rodriguez-Grasa et al. (2025).

Noise-robust Learning Amplitude fluctuations, modeled via noise injection, lead to either a negligible drop in accuracy or, in the case of Sinus, a slight improvement. This observation confirms that Kerr kernels retain their discriminative power even in the presence of realistic experimental imperfections. Consequently, the proposed approach not only matches or surpasses qubit-based kernels in ideal and noiseless simulations but also offers a practical advantage in near-term photonic implementations.

Analysing Hyperparameters. To quantify the influence of hyperparameters, we adopt the PED-ANOVA framework Watanabe et al. (2023) and analyze the resulting hyperparameter importances. Across all six benchmarks, PED-ANOVA attributes between 60% and 75% of the explained variance to the index j in scenarios where the decision boundary plays a dominant role, as observed in datasets such as Sinus and Corners. In these cases, the remaining 25%–40% of the variance is accounted for by the bandwidth parameter λ. However, this pattern reverses in datasets such as Fashion-MNIST and the multi-scale Disks 3 task, where λ contributes more than 55% to the explained variance. In these settings, the model benefits from precise bandwidth calibration, which helps avoid over-smoothing high-frequency texture in Fashion-MNIST, and allows alignment with the nested ring structure in Disks 3. Finally, in the Moons dataset, the variance is more evenly split, indicating a balanced contribution from both hyperparameters j and λ.

We conducted an extensive analysis across multiple datasets to evaluate the efficacy of the Kerr kernels and to elucidate the influence of hyperparameters. Recognizing that noise can inherently accompany data encoded in both amplitude and phase, we deliberately introduced noise levels of up to 10%. Our findings demonstrate the robustness of the Kerr Kernels in the face of such perturbations, reaffirming their viability in real-world applications where data integrity may be compromised.