Final version :
Phys. Rev. A 107, 062612 – Published 29 June 2023
https://journals.aps.org/pra/abstract/10.1103/PhysRevA.107.062612

Multidimensional Fourier series with quantum circuits

Berta Casas[1,*] and Alba Cervera-Lierta[1,†]

[1] *Barcelona Supercomputing Center, Plaça Eusebi Güell, 1-3, 08034 Barcelona, Spain.*

(Dated: September 13, 2023)

Quantum machine learning is the field that aims to integrate machine learning with quantum computation. In recent years, the field has emerged as an active research area with the potential to bring new insights to classical machine learning problems. One of the challenges in the field is to explore the expressibility of parametrized quantum circuits and their ability to be universal function approximators, as classical neural networks are. Recent works have shown that with a quantum supervised learning model, we can fit any one-dimensional Fourier series, proving their universality. However, models for multidimensional functions have not been explored in the same level of detail. In this work, we study the expressibility of various types of circuit ansatzes that generate multidimensional Fourier series. We found that, for some ansatzes, the degrees of freedom required for fitting such functions grow faster than the available degrees in the Hilbert space generated by the circuits. For example, single-qudit models have limited power to represent arbitrary multidimensional Fourier series. Despite this, we show that we can enlarge the Hilbert space of the circuit by using more qudits or higher local dimensions to meet the degrees of freedom requirements, thus ensuring the universality of the models.

I. INTRODUCTION

Machine learning (ML) is a well-established field that aims to develop the necessary tools to extract knowledge from big data batches by drawing inferences from its patterns. The development of computational paradigms such as quantum computation opens the path to explore the use of quantum devices to perform ML tasks, which raises the question of whether quantum machine learning (QML) algorithms can offer an advantage compared to classical ones.

QML explores the use of quantum computing devices to implement ML algorithms [1]. In some of these algorithms, in particular, in the supervised learning ones, data stored in a classical register needs to be mapped into a quantum state to be later processed by the quantum circuit. The parameters of the circuit are optimized by minimizing a cost function that compares some expectation values obtained from the circuit with the true data labels. Several proposals exist to embed data into quantum circuits [2–4]. Particularly, the re-uploading strategy has been applied to several problems, such as classification [2] or function fitting [5].

It has been shown that partial Fourier series emerge as the output of these models when using the re-uploading protocol [6]. Fourier series are universal to represent any square-integrable function in a given interval. This result has attracted the attention of recent QML works that use classical data [7–11]. Nevertheless, many applications that may require the use of QML rely on multidimensional datasets [12]. Although several works discuss the generalization of this model to multidimensional data, there needs to be a more thoughtful analysis of the implementation and scaling of the model with the data dimensions.

The current state of the art in experimental quantum computation, the so-called noisy intermediate-scale quantum (NISQ) era [13], presents a few qubit devices with limited coherence times that impose restrictions on the circuit depth and the fidelity of quantum operations. A family of algorithms suitable for these devices are variational quantum algorithms (VQA) [14, 15], from which the aforementioned supervised QML algorithms are part of. Variational algorithms present certain noise resilience due to their hybridization with classical optimization subroutines that fine-tune the circuit parameters. However, the low number of qubits in these NISQ devices limits the size of the Hilbert space available to perform quantum computational tasks. A way to circumnavigate this issue is to increase the Hilbert space by exploiting the higher local dimensions of each quantum information unit. Indeed, all quantum systems naturally contain more than two levels, which makes translating a qubit quantum device into a qudit system with d dimensions technically feasible, despite being experimentally challenging. Thus, increasing the Hilbert space from 2^n to d^n for n quantum information units might prove valuable in the near term [16–21].

This work analyzes the multidimensional Fourier series representation of QML circuits fed with classical data using a general formalism for qubits and beyond (qudits). We show how the quantum circuit requirements scale with the dimension of qudits and data when using the well-established data re-uploading strategy. Furthermore, we provide four types of circuit ansatzes to generate these functions and examine the constraints to ensure a proper fitting of a general series. The study of this problem sheds light on the expressibility of QML models, and in turn, provides insights into their capacity and limitations. It also addresses the question of whether these models can be considered universal approximators.

* berta.casas@bsc.es
† alba.cervera@bsc.es

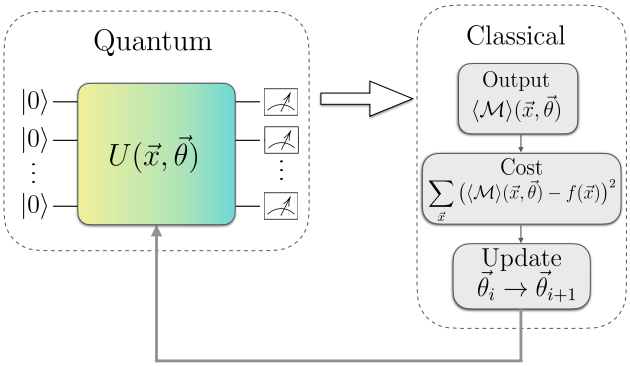

FIG. 1. Schematic representation of a quantum supervised learning model, which pertains to the family of Variational Quantum Algorithm (VQA). The picture divides the quantum and the classical part (dashed line boxes). The quantum part is composed of a quantum circuit with a given unitary U that depends on the data point $\vec{x}$ and the trainable parameters $\vec{\theta}$. The classical part consists of building an expectation value of a given observable $\mathcal{M}$, introducing it into a cost function (in our case, tailored for the classification of a target function $f(\vec{x})$) and optimize it with respect to the parameters via a classical subroutine.

We show that, while the proposed encoding strategies generate multidimensional Fourier series, the scaling of the degrees of freedom required to fit a general function grows rapidly, in some cases faster than the Hilbert space of the Parameterized Quantum Circuit (PQC). However, we present strategies that can overcome such limitations by increasing the dimension of the local space or employing a greater number of qudits. These strategies demonstrate the capability of generating arbitrary Fourier series through re-uploading methods, providing evidence of their universality.

This work is organized as follows: In Section II, we review the formalism of the one-dimensional Fourier series, and in Section III, we extend it to the multidimensional case. Furthermore, we discuss the degrees of freedom requirements, implementations and analysis of the models in Section IV. Finally, the conclusions are presented in Section V.

II. ONE-DIMENSIONAL FOURIER SERIES WITH QUANTUM CIRCUITS

This section reviews the quantum supervised learning model for fitting one-dimensional functions with the re-uploading strategy. We use a VQA fed with classical data, as we present in Fig. 1. Using a training dataset $\{x\}$ with corresponding function images $f(x)$, we generate a quantum circuit for each data point by using encoding gates $S(x)$ and parameterized gates $A(\vec{\theta})$, which act as trainable gates. By measuring the expectation value of some operator $\mathcal{M}$, we obtain a Fourier series in the $\{x\}$ domain. After optimizing the parameters $\vec{\theta}$, we introduce test data points into the circuit to obtain a function prediction.

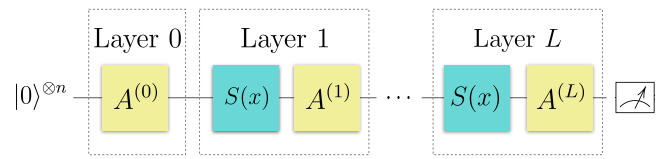

FIG. 2. General structure of a data re-uploading procedure. The 0th layer is used to generate an initial superposition. The other layers contain an encoding gate $S(x)$ and a processing gate $A(\vec{\theta_l})$, containing parameters which are optimized.

A widely used technique to define a PQC ansatz consists of defining a layer: a subcircuit composed of encoding and processing quantum gates. This structure is repeated L times along the circuit [2, 22] (see Fig. 2). We define the general circuit layer l with the data re-uploading encoding as

$$U_0 \equiv A(\vec{\theta_0}), \quad U_l \equiv A(\vec{\theta_l})S(x), \tag{1}$$

where $A(\vec{\theta_l})$ and $A(\vec{\theta_0})$ are general n qudit unitary gates, with free parameters $\vec{\theta}$, that act as processing step. For the encoding gate, a unitary gate $S(x) = e^{ixH}$ is used, where H is an arbitrary encoding Hamiltonian, and x is the data point considered. We assume that H is diagonal since its single value decomposition is $S(x) = V^\dagger e^{ix\Sigma}V$, where V is a unitary gate that can be re-absorbed by the processing gate $A(\vec{\theta_l})$, and Σ is a diagonal matrix with the eigenvalues of H (see App. A for the detailed derivation). The circuit layer can be interpreted as a quantum analogy of a neuron from a classical neural network because both structures are repeatedly fed with data. This is why these models are sometimes called quantum neural networks. The aforementioned quantum circuit ansatz has been used in prior works to prove that we obtain a one-dimensional truncated Fourier series as the output of these models [5, 6].

Definition 1. *One-dimensional truncated Fourier series.* A truncated Fourier series is an expansion of a periodic real-valued function that can be expressed as a sum of sines and cosines. In the exponential form, it takes the following structure:

$$f(x) = \sum_{\omega=-D}^{D} c_\omega e^{ix\frac{\pi}{K}\omega}, \tag{2}$$

where K is a half of the function period, $D = \max(\Omega)$ is the Fourier series degree and $\omega \in \mathbb{N}$ are the multiples of the fundamental frequency $\frac{\pi}{T}$. The frequency spectrum is given by $\Omega = \{\omega\frac{\pi}{K}\}_\omega$. The coefficients $c_\omega \in \mathbb{C}$, fulfill the condition $c_\omega = c^*_{-\omega}$. A Fourier series with enough degree D can approximate any continuous, square-integrable function.

By computing the expectation value of an observable $\mathcal{M}$ in the output state of the PQC, we obtain an approximation of $f(x)$. For simplicity, when we use more than one qudit in the PQC, we measure only one of them. Then, $\mathcal{M} = \mathbf{M} \otimes \mathbb{I} \otimes ...\mathbb{I}$, where $\mathbf{M}$ is a single-qudit observable. We obtain (see App. A for the details)

$$\langle \mathcal{M}(x) \rangle = \sum_{\omega \in \Omega_{\mathbf{kk'}}} c_\omega \; e^{ix\omega}, \tag{3}$$

with

$$c_\omega = \sum_{\substack{\mathbf{k},\mathbf{k}'=1 \\ \Omega_{\mathbf{kk'}}=\omega}}^{N} \left(\sum_{i=1}^{N} \mathcal{M}_i A^{(L)}_{ik_L} A^{*(L)}_{ik'_L} \right) \times A^{(0)}_{k_1 1} A^{*(0)}_{k'_1 1} \prod_{p=2}^{L} A^{(p-1)}_{k_p k_{p-1}} A^{*(p-1)}_{k'_p k'_{p-1}}, \tag{4}$$

where $N = d^n$ is the dimension of the Hilbert space of n qudits with local dimension d, L is the number of layers, and $\mathcal{M}_i$ are the eigenvalues of the observable, which we assume to be diagonal. We use the multi-index notation $\mathbf{k} \equiv \{k_1, k_2, ...k_L\} \in [N]^L$, with $k_i \in \{1, ..N\}$. Also, we introduce the multi-index sum defined by $\Lambda_{\mathbf{k}} \equiv (\lambda_{k_1} + \lambda_{k_2} + ...\lambda_{k_L})$, where λ_{k_i} is an eigenvalue of the encoding Hamiltonian H in the layer i. With this, the frequency spectrum of the model is given by

$$\begin{aligned} &\Omega_{\mathbf{kk'}} \equiv \{\Lambda_{\mathbf{k}} - \Lambda_{\mathbf{k'}}\} = \\ &\{\lambda_{k_1} + \lambda_{k_2} + \cdots + \lambda_{k_L} - \left(\lambda_{k'_1} + \lambda_{k'_2} + \cdots + \lambda_{k'_L}\right)\}. \end{aligned} \tag{5}$$

The frequency spectrum is fully characterized by the eigenvalues of the encoding Hamiltonian, and the coefficients rely on the trainable gates and their parameters. Notice that all combinations of the multi-indices $\mathbf{k}$ and $\mathbf{k'}$ that generate a frequency $\Omega_{\mathbf{kk'}} = \omega$ give us a different contribution to the coefficient c_ω.

So far, this model accepts a general encoding Hamiltonian H, but let us take the following choice for practical purposes: using a Hamiltonian with the N-dimensional spin S_z eigenvalues, which we name the "spin-like" encoding. For one qubit we use $H = \frac{1}{2}\sigma_z$, with eigenvalues $\pm 1/2$; for one qutrit, the analogous Hamiltonian of a spin-1 Hamiltonian, which eigenvalues are $\pm 1, 0$, etc. We use a global S_z Hamiltonian tailored to the dimension of the quantum circuit $N = d^n$. For example, for a system of two qubits ($N = 2^2 = 4$), H is the spin-3/2 Hamiltonian, which acts on the whole circuit instead of using $H = \frac{1}{2}\sigma_z \otimes \frac{1}{2}\sigma_z$. This encoding will be the same used with a ququart, a qudit of $d = 4$ (with the same value of N as for two qubits). Using this particular encoding, the positive frequency spectrum, emerging from Eq. (5), is $\Omega = \{0, \cdots, (N-1)L - 1, (N-1)L\}$, where L is the number of layers in the circuit. The negative frequencies are also included in the spectrum, which is symmetric by its construction. For simplicity, from now on, we drop the label for multi-indices in the frequency spectrum: $\Omega \equiv \Omega_{\mathbf{kk'}}$. The degree of the Fourier series, with this particular encoding, is given by $D = (N-1)L$.

The spectrum generated with the "spin-like" encoding only contains integer frequencies. For instance, if the function we are fitting requires semi-integer or real frequencies a proper approximation cannot be achieved, regardless of the number of layers used. We can tackle this issue by introducing a re-scaling parameter η into the encoding gate: $S(x) = e^{ix\eta H}$. This parameter is optimized together with the rest of the free parameters of the circuit, as we explain in App. C. With this re-scaling factor, the whole Fourier series spectrum is multiplied by η, and the degree of the Fourier series becomes $D = \eta(N-1)L$. We can introduce more fine-tuning in the frequency spectrum by using a different η_i for each encoding gate in the circuit layers. This extension is closer to the original idea from the data re-uploading work [2] but should be treated carefully, because methods with re-scaling parameters may lead to good expressibility but overfitting and poor generalization bounds for more complex tasks [23]. Nevertheless, when applied appropriately, the re-scaling factor can serve as an effective hyperparameter for promoting generalization in quantum kernel models [24]. It is important to note that the encoding strategy plays a significant role in determining the characteristics of the frequency spectrum. For instance, in Ref. [8], the authors propose a technique that utilizes re-scaling factors to generate exponentially more frequencies.

III. MULTIDIMENSIONAL FOURIER SERIES WITH QUANTUM CIRCUITS

In this section, we present different ansatzes that generate multidimensional Fourier series and explore the scaling of the QML models' performance with the dimensions of the input data. To do it, we study the expressibility of the multidimensional Fourier model, meaning the type of functions that the model can generate. Hence, we expand the function fitting formalism to multidimensional data, opening the possibility to explore more complex problems with these quantum models.

Definition 2. *Multidimensional truncated Fourier series.* The generalization of a one-dimensional truncated Fourier series to M-dimensional data is given by

$$f(\vec{x}) = \sum_{\omega_1, \omega_2, \ldots, \omega_M = -D}^{D} c_{\vec{\omega}} e^{i\vec{x}\cdot\vec{\omega}}, \tag{6}$$

where $D = \max(\omega_1, \omega_2, ..., \omega_M)$ is the degree of the Fourier series. The data $\vec{x} = (x_1, x_2, ..., x_M) \in \mathbb{R}^M$ and the frequencies $\vec{\omega} = (\omega_1, \omega_2, ..., \omega_M) \in \mathbb{Z}^M$ are represented by $M-$dimensional vectors, and $\vec{x} \cdot \vec{\omega}$ is the scalar product. The coefficients $c_{\omega_1, \omega_1, \ldots, \omega_M} \in \mathbb{C}$ fulfill the relation $c_{\omega_1, \omega_1, \ldots, \omega_M} = c^*_{-\omega_1, -\omega_1, \ldots, -\omega_M}$.

The M-dimensional Fourier series contains substantially more coefficients than a series with only one di-

mension. For a given degree $D \in \mathbb{N}$, the number of independent coefficients $c_{\vec{\omega}}$ is

$$N_c = \frac{(2D+1)^M - 1}{2} + 1. \tag{7}$$

As an example, for degree $D = 1$ and data-dimension $M = 2$, according to Eq. (7), we have 5 coefficients: $c_{00}, c_{01}, c_{10}, c_{11}, c_{-11}$, while the other ones are constrained by $c_{\omega_1\omega_2} = c^*_{-\omega_1-\omega_2}$. The degrees of freedom ν (abbreviated as DOF) of the Fourier series are the number of independent variables needed to fully characterize a set of coefficients of a series with a given degree D. They account for the real and imaginary parts of each coefficient, except the one associated with the zero frequency, which only contains a real part. Therefore, the DOF of an M-dimensional Fourier series are given by

$$\nu \equiv 2N_c - 1 = \big(2D+1\big)^M. \tag{8}$$

Depending on the circuit ansatz, the number of layers required to achieve enough freedom to represent an arbitrary series will vary. In the following subsection, we propose four strategies: the Line, Parallel, Mixed, and Super-parallel Ansatzes. We assume that the processing gates are general unitary transformations in all models. Hence, the number of parameters they contain is $N^2 - 1$, determined by the dimension of $\mathrm{SU}(N)$, with $N = d^n$ being the dimension of the quantum circuit. We acknowledge that this approach may not be practical in terms of trainability. However, we utilize this strategy to explore the limits of the models.

A. Line ansatz

The Line Ansatz (LA) encodes all data dimensions in a single qudit. The structure of the model is shown in Fig. 3 (**a**). Each layer $L^{(l)}$ of the LA encodes M data features as

$$L^{(l)}(\vec{x}, \vec{\theta})_{\mathrm{LA}} \equiv \prod_{m=1}^{M} S(x_m) A_m^{(l)}(\vec{\theta}_{l,m}), \tag{9}$$

where M is the dimension of the dataset, x_m the data features and $\vec{\theta}_{l,m}$ are the processing parameters corresponding to feature m from layer l. After the encoding step, the processing gate is applied to avoid the collapse of the data in a single variable (see App. D for details).

Another strategy to encode multidimensional data in one qudit is to use non-commuting gates, meaning $[S_1(x), S_2(x)] \neq 0$. For example, in the two-dimensional qubit case, we could use $S_1(x_1) = R_y(x_1)$ and $S_2(x_2) = R_z(x_2)$. Given that S_1 and S_2 do not commute, the single value decomposition of the encoding gate is given by

$$S_1(x_1)S_2(x_2) = V_1^\dagger \Sigma(x_1) V_1 V_2^\dagger \Sigma(x_2) V_2. \tag{10}$$

This can be interpreted as adding an extra layer $V_1 V_2^\dagger$ in between the two encoding gates that does not contain trainable parameters and also avoids the model from interpreting the data as one-dimensional.

The number of parameters to be optimized in the LA ansatz is given by

$$N_p^{(LA)} = (ML+1)(N^2-1) \sim MLN^2. \tag{11}$$

As mentioned above, $N = d$ is the dimension of the circuit and the dimension of the encoding gate $S(x)$. Thus, the parameters grow linearly with the number of layers L and the data dimension M, and quadratically with the circuit dimension N.

B. Parallel ansatz

The parallel ansatz (PA) encodes each data feature in a different qudit with a single-qudit gate, therefore, we require $n = M$ qudits to encode all M data features. The encoding is followed by a processing $M-$qudit gate, as shown in Fig. 3 (**b**). We define each layer as

$$L^{(l)}(\vec{x}, \vec{\theta})_{\mathrm{PA}} \equiv \left(\bigotimes_{m=1}^{M} S(x_m) \right) A^{(l)}(\vec{\theta}_l), \tag{12}$$

where $A^{(l)}(\vec{\theta}_l)$ is a $N \times N$ general processing unitary, with $N = d^M$. This ansatz contains a total number of parameters

$$N_p^{(PA)} = (d^{2M} - 1)(L+1) \sim d^{2M} L, \tag{13}$$

which grows exponentially with the number of features M.

If single-qudit gates are used in the processing steps instead of general multi-qudit gates, then a product of M one-dimensional Fourier series is generated. Therefore, entanglement must be included in the processing gates to obtain a genuine multidimensional series (see App. E for more details).

C. Mixed ansatz

Now we consider a mix between the two ansatzes previously discussed, named the Mixed Ansatz (MA). It divides the data features into different batches and uses p qudits for different sets (see Fig. 3 (**c**)). Taking $p \leq M$ qudits, the data features are distributed in p qudits, as in the PA, and more encoding layers can be used for each qudit if required, as in the LA. For $p = 1$, MA behaves like LA, and for $p = M$ it behaves like PA. More formally, we define a layer of the MA as

$$L^{(l)}(\vec{x}, \vec{\theta})_{\mathrm{MA}} \equiv \prod_{k=1}^{\lceil M/p \rceil} \left(\bigotimes_{m=1}^{p} S(x_m) \right) A_k^{(l)}(\vec{\theta_{l,k}}), \tag{14}$$

and the number of parameters in this model is given by

$$N_p^{(MA)} = (d^{2p} - 1)(\lceil M/p \rceil L + 1). \tag{15}$$

If M/p is not an integer, we use fewer encoding gates on the layers' last encoding block.

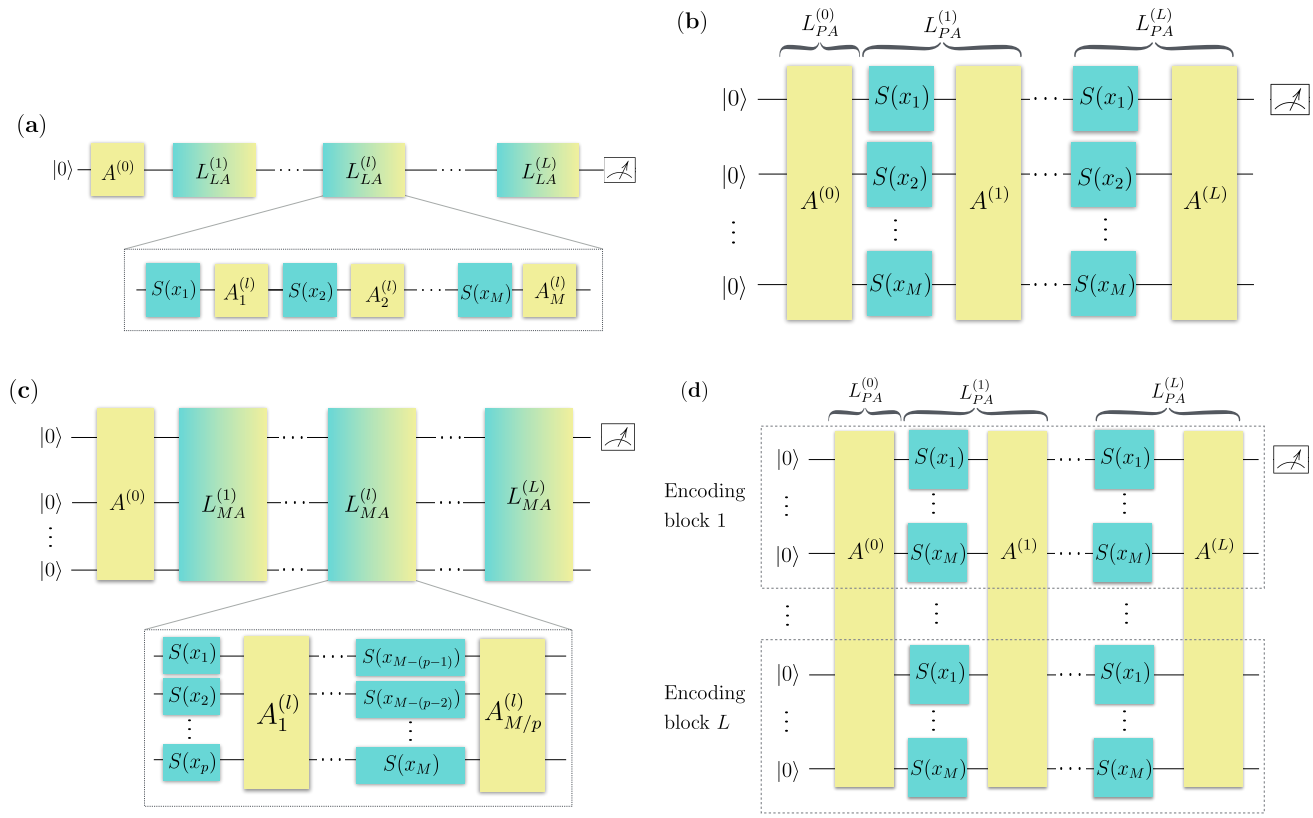

FIG. 3. Quantum circuit ansatzes of the models. The Line ansatz (**a**) encodes each data feature in a single qudit. Thus, the circuit depth grows linearly with the total number of features M. The Parallel ansatz (**b**) encodes the M features in M qudits instead. The Mixed ansatz (**c**) combines the two approaches by distributing the features encoding between $p < M$ qudits and uses more gates to introduce several features in each layer. Finally, the Super-Parallel ansatz (**d**) uses L layers and L encoding blocks per qudit, therefore requiring $d = ML$ qudits.

D. Super-Parallel ansatz

Finally, we provide an ansatz that includes layers in depth and width directions of the quantum circuit. In other words, it has the same layer structure of re-uploading models but with L encoding blocks per layer. Each block contains M single-qubit encoding gates (one for each data feature). As shown in Fig. 3 (**d**), a total of $n = LM$ qudits are required for this ansatz. Each layer is given by

$$L^l(\vec{x}, \vec{\theta})_{SP} \equiv \left(\bigotimes_{i=1}^{L} \left(\bigotimes_{m=1}^{M} S(x_m) \right) \right) A^{(l)}(\vec{\theta}_l), \tag{16}$$

and the number of parameters in this model is

$$N_p^{(SP)} = (L+1)(d^{2ML} - 1) \sim L d^{2ML}, \tag{17}$$

which grows exponentially with the input data dimension and the number of layers.

IV. RESULTS AND DISCUSSION

To compare the models, we assume that all use singe-qudit encoding gates, although their extension to multi-qudit gates is straightforward (equivalent to finding an information unit of dimension $\bar{d} = d^n$). The degree of the Fourier series generated by the first three ansatzes is determined by

$$D = (d-1)L, \tag{18}$$

where d is the dimension of the qudit(s) used in the model. The super parallel model outputs a Fourier series of degree

$$D^{(SP)} = (d-1)L^2, \tag{19}$$

which grows quadratic with the number of layers due to the use of multiple encoding blocks per layer. For detailed derivation and further information on the models, refer to App. D, E, F, and G. By specifying the qudit dimension d and the number of layers L, we determine the Fourier series frequency spectrum Ω and its degree D. This allows us to calculate the degrees of freedom ν

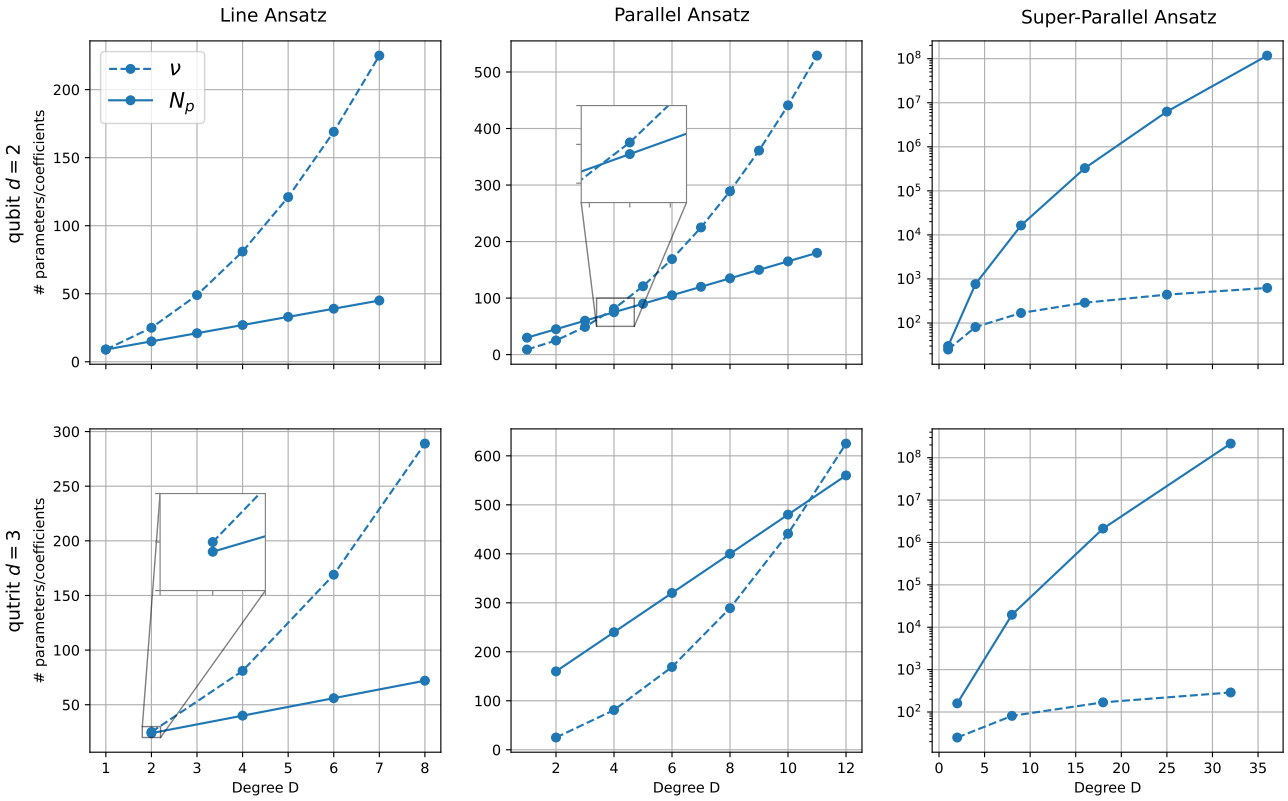

FIG. 4. Comparison of the degrees of freedom condition for a two-dimensional Fourier series for the Line (first column), Parallel (second column), and Super-parallel (third column) ansatzes using qubits (first row) and qutrits (second row). The dashed lines represent the degrees of freedom ν of the Fourier series produced by the models plotted against its degree D, while the solid line depicts the number of trainable parameters N_p in the circuit ansatzes that generate a Fourier series of degree D with L layers. The Parallel Ansatz fulfils the degrees of freedom condition for higher-degree Fourier series compared to the Line Ansatz, being the gap wider when using qutrits. In particular, the condition for the Parallel Ansatz is fulfilled until $D = 3$ and $D = 10$ for qubits and qutrits, respectively. In the Super-Parallel ansatz, plotted on a logarithmic y-axis scale, we observe a faster increase in the number of parameters compared to the degrees of freedom.

of the output series by plugging Eq. (18) and (19) in Eq. (8). Our next step is to investigate whether we can fully represent the Fourier series coefficients by comparing the degrees of freedom ν with the number of independent parameters N_p in the quantum circuit. The ansatzes are required to contain, at least, ν free parameters to generate any series coefficient, resulting in a condition $N_p \geq \nu$. When this condition is not accomplished, the model is not general enough to approximate all possible Fourier series. However, having $\nu > N_p$ does not necessarily guarantee that we can fit a general series since the coefficient equations are highly coupled and non-linear (see Eq. (4)), although one might expect that, in general, $\nu > N_p$ is enough to approximate a given series. Our goal is to establish a lower bound in the worst-case scenario in terms of the circuit ansatz requirements to fit a general multidimensional Fourier series.

The one-dimensional feature model, introduced in section II, possesses $\nu = 2(D+1) = 2(L(d-1)+1)$ degrees of freedom, and $N_p = (L+1)(d^2-1)$ free parameters. Both these quantities increase linearly with the number of layers and $N_p > N_c$ holds for all values of L and d, which implies that the degrees of freedom requirements are satisfied in all cases.

For the line, parallel and mixed ansatz with a fixed qudit dimension, the degrees of freedom grow polynomially with the number of layers, $\nu \sim L^M$, while the number of parameters grows proportionally to the layers used $N_p \sim L$, regardless of the data dimension. Therefore, at some point, the DOF will exceed the number of free parameters of the model.

In Fig. 4 we represent the DOF condition for the Line, Parallel and Super-parallel ansatzes for a case of two-dimensional data with qubits and qutrits models. In the first place, we see that the Line Ansatz only fulfils the condition for the qubit model with $D = 1$. Taking into account the asymptotics in the degrees of freedom condition $N_p \geq \nu$ and assuming that $\nu \sim (2DL)^M$ and $N_p^{(LA)} \sim MLd^2$, we obtain the following condition:

$$d \lesssim \left(\frac{M}{2^M L^{M-1}}\right)^{1/(M-2)} \xrightarrow[M\to\infty]{} \frac{1}{2L}. \tag{20}$$

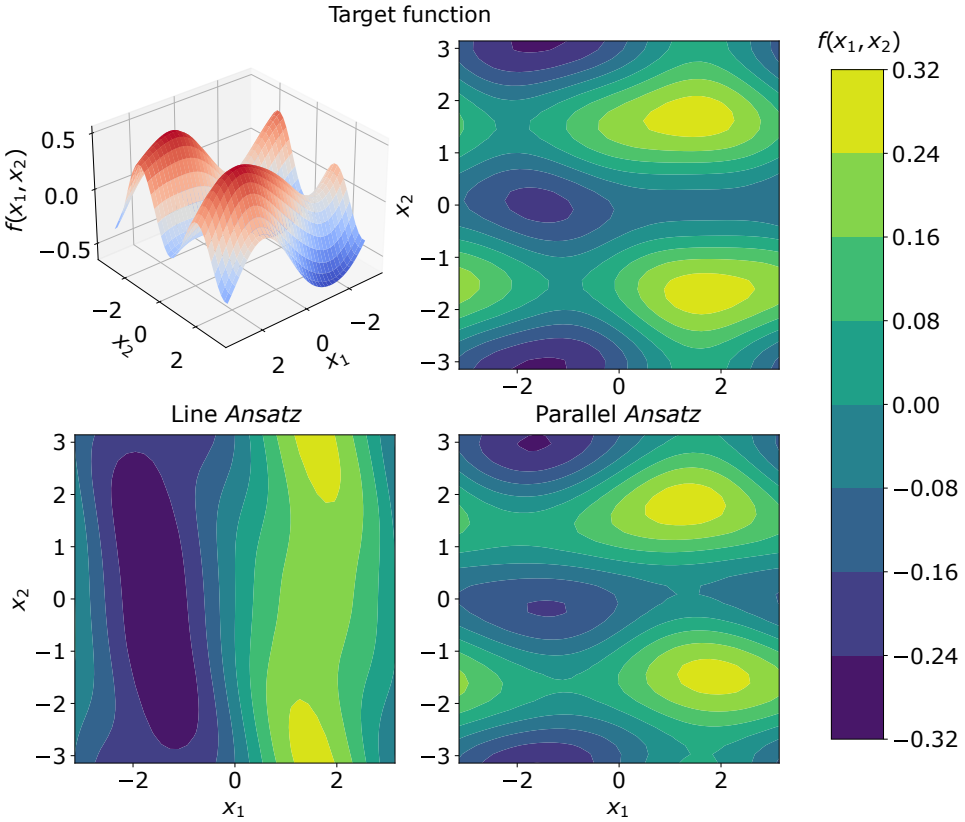

FIG. 5. Simulation the Line and Parallel ansatzes for fitting a Fourier series of degree $D = 2$. The target function in the trigonometric form is given by $f(x_1, x_2) = -0.02 + 0.04\cos(2x_1 + x_2) + 0.25\sin x_1 - 0.3\cos 2x_2 - 0.1\sin(x_1 - x_2)$. We use the Nelder-Mead method as a classical optimization subroutine with 500 training and 1500 testing data points. In the Line ansatz, we use one qubit and $L = 2$. The accuracy obtained is 38.53% and, as we can see, the model does not capture the structure of the target function. In the Parallel ansatz, we use two qubits and $L = 2$. The accuracy obtained is 95.63%, and the predicted and target functions have a similar structure. In this regime, the Line ansatz model does not fulfil the DOF condition, while the Parallel does. This is reflected in the accuracy of the simulations.

This suggests that simply increasing the qudit dimension is insufficient to achieve the desired number of parameters. In particular, for large data dimensions, the condition can not be fulfilled because $d \leq 1/2L$ is an impossible condition for $L > 1$. Therefore, one-qudit models have limited power for fitting multidimensional functions.

For the Parallel ansatz depicted in the second row of Fig. 4, we see that by increasing the qudit dimension, we arrive at higher Fourier series Degrees. Indeed, by running the asymptotics in the DOF condition considering $N_p^{(PA)} \sim Ld^{2M}$, we achieve

$$d \gtrsim 2L^{\frac{M-1}{M}} \xrightarrow[M\to\infty]{} 2L, \tag{21}$$

which implies that in order to satisfy the inequality, we would need to use qudit dimensions that increase with the number of layers. In particular, in the limit of large dimensional datasets, d has to grow proportional to L. This indicates that the model remains universal, in the sense that can generate any arbitrary Fourier series, as long as d can grow with L. In that case, the complexity of the problem would shift to finding systems with arbitrarily large qudit dimensions to fit generic Fourier series, which can be resource-demanding for some technologies.

Fig. 5 depicts the results of simulation for models fitting a two-dimensional Fourier series of degree $D = 2$ with the Line and Parallel ansatzes. Since for two-dimensional data, the LA does not accomplish the DOF condition, the simulation does not find suitable parameters to fit the coefficients. On the contrary, the parallel model approximates with high accuracy the target function with only two layers, because meets the DOF requirements. For more details on two and three-dimensional models see Ref. [25], where expressibility is analyzed in detail.

The limitation that we exhibit for the Line and Parallel ansatzes is that we do not have enough free parametrization in the Hilbert space of the PQC to accomplish the DOF that multidimensional Fourier series require. Therefore, we can only fit functions up to a certain degree. One might assume that the problem could be solved

by adding more layers, which introduces more trainable parameters in the PQC to match the DOF required for the desired set of coefficients. However, the addition of more layers also increases the output Fourier series' degree D (see Eq. (18)), which requires more coefficients and thus more degrees of freedom. Therefore, the problem cannot be resolved by merely increasing the number of layers because this also raises the required degrees of freedom of the output model.

The Super-parallel ansatz has a number of qudits that grows with the number of layers and the data dimension. In this scenario, the degrees of freedom scale as $\nu \sim (2dL^2)^M$. Substituting this into the inequality $N_p \geq \nu$ with the parameter count given in Eq. (17) yields

$$d \geq 2^{\frac{1}{2L-1}} L^{\frac{2}{2L-1}}. \tag{22}$$

It can be easily seen that this inequality is always satisfied. An example of this can be observed in the third row of Fig. 4, where we can notice that the number of parameters grows more rapidly than the degrees of freedom of the resulting model function, making the model capable of fitting any arbitrary Fourier series. There is an increasing gap between the number of parameters and degrees of freedom, which opens the possibility of finding more sophisticated Super-parallel encodings that can employ lesser variables while remaining universal, for instance, by using non-general unitary gates or fewer qudits.

V. CONCLUSIONS

In this work, we have explored how to generate multidimensional Fourier series with parameterized quantum circuits. These series emerge naturally from the expectation values of quantum operators with a particular data encoding. We have compared the degrees of freedom of a general Fourier series of a given degree to the number of free parameters in the circuit, which we refer to as the DOF condition and provides insight into the model expressibility. We provide a trade-off between the number of qudits, circuit depth (measured with the number of layers of the circuit), data dimension, and local qudit dimension.

Current quantum computers can use higher energy states to perform high-dimensional quantum computation. Apart from this, the use of a general formalism for qudits is motivated by the possibility of exploring larger Hilbert spaces (with more extensive parametrization), which gives more freedom when fitting the desired set of coefficients.

For one-dimensional data, the DOF requirement is always accomplished. However, for higher-dimensional data, the degrees of freedom grow exponentially with the data dimensions, which can be problematic for some models that may not be able to keep up with this rapid growth. For example, single-qudit models have limited power in the expressivity of multidimensional data, because they lack the appropriate parametrization. However, multi-qudit models can approximate functions up to a higher degree, which can be used as an appropriate approximation for some problems.

We can always find a model that satisfies the DOF condition by using large qudit dimensions or a substantial number of qudits, then being fully expressive. Therefore, multidimensional quantum learning models can be considered universal, as with sufficient parametrization, they can fit any arbitrary Fourier series. Potential issues may appear with trainability and generalization resulting from using such a large parametrization. However, we speculate that for most problems, such an extensive parametrization would not be necessary nor practical in terms of trainability.

The Line, Parallel and Mixed ansatzes exhibit an inductive bias towards limited-band functions, which contain low frequencies. Further work needs to be done for studying how the inductive bias of quantum learning models [26, 27] varies with the PQC used. It would be interesting to benchmark the performance, trainability, and generalization capabilities of the ansatzes against a classical surrogate model [28, 29]. Another open question is how to determine the level of redundancy necessary in the output Fourier series degree when the target function is not known beforehand.

We aim for our work to contribute to the understanding of QML with classical multidimensional data and the further exploration of more sophisticated embedding strategies.

CODE AND DATA AVAILABILITY

All code used in this work can be found on the GitHub repository https://github.com/bsc-quantic/fourier.

ACKNOWLEDGEMENTS

This work has been financially supported by the Ministry of Economic Affairs and Digital Transformation of the Spanish Government through the QUANTUM ENIA project call – Quantum Spain project, and by the European Union through the Recovery, Transformation and Resilience Plan – NextGenerationEU within the framework of the Digital Spain 2026 Agenda. Also, we acknowledge the BSC's Quantic group members for the suggestions and discussions provided.

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).

[2] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Data re-uploading for a universal quantum classifier, Quantum **4**, 226 (2020).

[3] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, Quantum embeddings for machine learning, arXiv 10.48550/ARXIV.2001.03622 (2020).

[4] T. Goto, Q. H. Tran, and K. Nakajima, Universal approximation property of quantum machine learning models in quantum-enhanced feature spaces, Physical Review Letters **127**, 10.1103/physrevlett.127.090506 (2021).

[5] F. J. G. Vidal and D. O. Theis, Input redundancy for parameterized quantum circuits, Frontiers in Physics **8**, 10.3389/fphy.2020.00297 (2020).

[6] M. Schuld, R. Sweke, and J. J. Meyer, Effect of data encoding on the expressive power of variational quantum-machine-learning models, Physical Review A **103**, 10.1103/physreva.103.032430 (2021).

[7] D. Heimann, G. Schönhoff, and F. Kirchner, Learning capability of parametrized quantum circuits, arXiv 10.48550/ARXIV.2209.10345 (2022).

[8] S. Shin, Y. S. Teo, and H. Jeong, Exponential data encoding for quantum supervised learning, Physical Review A **107**, 10.1103/physreva.107.012422 (2023).

[9] A. Daskin, A walk through of time series analysis on quantum computers, arXiv 10.48550/ARXIV.2205.00986 (2022).

[10] Z. Yu, H. Yao, M. Li, and X. Wang, Power and limitations of single-qubit native quantum neural networks, arXiv 10.48550/ARXIV.2205.07848 (2022).

[11] B. Y. Gan, D. Leykam, and D. G. Angelakis, Fock state-enhanced expressivity of quantum machine learning models, EPJ Quantum Technology **9**, 16 (2022).

[12] K. A. Woźniak, V. Belis, E. Puljak, P. Barkoutsos, G. Dissertori, and *et. al.*, Quantum anomaly detection in the latent space of proton collision events at the lhc (2023).

[13] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[14] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, and *et al.* , Noisy intermediate-scale quantum algorithms, Reviews of Modern Physics **94**, 10.1103/revmodphys.94.015004 (2022).

[15] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, and **et al.**, Variational quantum algorithms, Nature Reviews Physics **3**, 625 (2021).

[16] M. Blok, V. Ramasesh, T. Schuster, K. O'Brien, J. Kreikebaum, and *et. al.*, Quantum information scrambling on a superconducting qutrit processor, Physical Review X **11**, 10.1103/physrevx.11.021010 (2021).

[17] A. Cervera-Lierta, M. Krenn, A. Aspuru-Guzik, and A. Galda, Experimental high-dimensional Greenberger-Horne-Zeilinger entanglement with superconducting transmon qutrits, Physical Review Applied **17**, 10.1103/physrevapplied.17.024062 (2022).

[18] M. Erhard, M. Krenn, and A. Zeilinger, Advances in high-dimensional quantum entanglement, Nature Reviews Physics **2**, 365 (2020).

[19] A. Castro, A. G. Carrizo, S. Roca, D. Zueco, and F. Luis, Optimal control of molecular spin qudits, Physical Review Applied **17**, 10.1103/physrevapplied.17.064028 (2022).

[20] M. D. Jenkins, Y. Duan, B. Diosdado, J. J. García-Ripoll, A. Gaita-Ariño, and *et al.*, Coherent manipulation of three-qubit states in a molecular single-ion magnet, Phys. Rev. B **95**, 064423 (2017).

[21] C. M. Caves and G. J. Milburn, Qutrit entanglement, Optics Communications **179**, 439 (2000).

[22] A. Pérez-Salinas, D. López-Núñez, A. García-Sáez, P. Forn-Díaz, and J. I. Latorre, One qubit as a universal approximant, Physical Review A **104**, 10.1103/physreva.104.012405 (2021).

[23] V. Vapnik, E. Levin, and Y. L. Cun, Measuring the vc-dimension of a learning machine, Neural Computation **6**, 851 (1994).

[24] A. Canatar, E. Peters, C. Pehlevan, S. M. Wild, and R. Shaydulin, Bandwidth enables generalization in quantum kernel models, arXiv 10.48550/arXiv.2206.06686 (2022).

[25] A. Gratsea and P. Huembeli, The effect of the processing and measurement operators on the expressive power of quantum models, arXiv 10.48550/ARXIV.2211.03101 (2022).

[26] J. Kübler, S. Buchholz, and B. Schölkopf, The inductive bias of quantum kernels, Advances in Neural Information Processing Systems **34**, 12661 (2021).

[27] E. Peters and M. Schuld, Generalization despite overfitting in quantum machine learning models, arXiv 10.48550/arXiv.2209.05523 (2022).

[28] F. J. Schreiber, J. Eisert, and J. J. Meyer, Classical surrogates for quantum learning models, arXiv 10.48550/arXiv.2206.11740 (2022).

[29] J. Landman, S. Thabet, C. Dalyac, H. Mhiri, and E. Kashefi, Classically approximating variational quantum machine learning with random fourier features, arXiv 0.48550/arXiv.2210.13200 (2022).

Appendix A: Data re-uploading for one-dimensional Fourier series

This section explains how we generate one-dimensional Fourier series with a circuit of n qudits of dimension d. For processing gate a general unitary is taken, of total dimension $N = d^n$ (n qudits of dimension d):

$$A^{(l)} = \begin{pmatrix} A^{(l)}_{11} & \cdots & A^{(l)}_{1N} \\ \vdots & \ddots & \vdots \\ A^{(l)}_{N1} & \cdots & A^{(l)}_{NN} \end{pmatrix}. \tag{A1}$$

In contrast to the processing, the encoding gates are the same in all layers. Let's assume the following single-qudit encoding gate:

$$S(x) = e^{ixH}, \tag{A2}$$

where H is a $N \times N$ Hermitian operator and x is a one-dimensional data point from the function we want to represent. We assume that the encoding Hamiltonian H is diagonal because, when taking the Singular Value Decomposition, $H = V^\dagger \Sigma V$, where V and $V^\dagger$ are unitary matrices and Σ is a diagonal matrix formed with the eigenvalues of H, the encoding gate becomes

$$S(x) = e^{-ixV\Sigma V^\dagger} = \sum_{m=0}^{\infty} \frac{1}{m!}(-ixV\Sigma V^\dagger)^m = 1 + \sum_{m=1}^{\infty} \frac{1}{m!} V(-ix\Sigma)^m V^\dagger = V e^{-ix\Sigma} V^\dagger = V R(x) V^\dagger. \tag{A3}$$

where we take into consideration that $V^\dagger V = \mathbb{I}$ and $\Sigma = \mathrm{diag}(\lambda_1, ..., \lambda_N)$. The resulting diagonal encoding gate is

$$R(x) = \mathrm{diag}\left(e^{ix\lambda_1}, \cdots, e^{ix\lambda_N}\right). \tag{A4}$$

Therefore, we assume, without loss of generality, a diagonal encoding matrix, since V and $V^\dagger$ gates are re-absorbed in the definition of the general processing gates $A^{(l)}$. With this assumption, each layer l is composed of the product

$$L^{(l)} = A^{(l)} R(x) \qquad L^{(0)} = A^{(0)}, \tag{A5}$$

which matrix elements are

$$L^{(l)}_{ij} = \sum_{k=1}^{N} A^{(l)}_{ik} R_{kj} = A^{(l)}_{ij} e^{ix\lambda_j}, \qquad L^{(0)}_{ij} = A^{(0)}_{ij}. \tag{A6}$$

In general, for L layers, the unitary transformation of the whole circuit is expressed as

$$U_{ij} = \sum_{k_1,\cdots,k_L=1}^{N} A^{(L)}_{ik_L} e^{-ix\lambda_{k_L}} A^{(L-1)}_{k_L k_{L-1}} ... A^{(1)}_{k_2 k_1} e^{-ix\lambda_{k_1}} A^{(0)}_{k_1 k_0}. \tag{A7}$$

The initial state of the circuit is the zero state of dimension N. Therefore $|0\rangle = (1, 0, \cdots, 0)^T$. Thus, the state generated by the circuit becomes:

$$|\psi\rangle = U|0\rangle^{\otimes n}, \tag{A8}$$

$$\psi_i = U_{ij}\delta_{j1} = U_{i1}. \tag{A9}$$

Putting it all together, we obtain

$$\psi_i = \sum_{k_1,\cdots,k_L=1}^{N} e^{-ix(\lambda_{k_1}+...+\lambda_{k_L})} A^{(L)}_{ik_L} A^{(L-1)}_{k_L k_{L-1}} ... A^{(1)}_{k_2 k_1} A^{(0)}_{k_1 1}, \tag{A10}$$

By introducing the multi-index notation $\mathbf{k} \equiv \{k_1, ..., k_L\} \in [N]^L$ and the multi-index sum $\Lambda_{\mathbf{k}} = \lambda_{k_1} + ... + \lambda_{k_L}$,we re-express the i-vector state:

$$\psi_i = \sum_{\mathbf{k}\in[N]^L} e^{-ix\Lambda_{\mathbf{k}}} A^{(L)}_{ik_L} A^{(L-1)}_{k_L k_{L-1}} ... A^{(1)}_{k_2 k_1} A^{(0)}_{k_1 1}. \tag{A11}$$

Each multi-index $\mathbf{k}$ is a possible combination of L indices and each run from 1 to N. The multi-index sum $\Lambda_{\mathbf{k}} = \lambda_{k_1} + ... + \lambda_{k_L}$ is a sum that has $|\mathbf{k}| = N^L$ possible values, given by all the possible combinations of λ_{k_i}.

Now we compute the expectation value of a given observable $\mathcal{M}$ in this state. The observable can be diagonal by the same argument used for the encoding gate $S(x)$, explained above. The eigenvalues of any observable are real, so $\mathcal{M}^*_{ii} = \mathcal{M}_{ii} \equiv \mathcal{M}_i$. All together,

$$\langle \mathcal{M} \rangle = \langle \psi | \mathcal{M} | \psi \rangle = \sum_{i=1}^{N} U^*_{i1} M_i U_{i1} = \sum_{\substack{k_1,\cdots,k_L=1 \\ k'_1,\cdots,k'_L=1}}^{N} e^{ix(\Lambda_{\mathbf{k}} - \Lambda_{\mathbf{k}'})} \left(\sum_{i=1}^{N} M_i A^{(L)}_{ik_L} A^{*(L)}_{ik'_L} \right) A^{(0)}_{k_1 1} A^{*(0)}_{k'_1 1} \prod_{p=2}^{L} A^{(p-1)}_{k_p k_{p-1}} A^{*(p-1)}_{k'_p k'_{p-1}}. \quad \text{(A12)}$$

For each set of $\mathbf{k}$ and $\mathbf{k}'$ parameters, we generate a particular frequency $\Omega_{kk'} = \Lambda_{\mathbf{k}} - \Lambda_{\mathbf{k}'}$. Notice that different $\mathbf{k}, \mathbf{k}'$ choices can give the same frequency. Also, to obtain the opposite sign value $-\Omega_{kk'}$, one needs to exchange the $\mathbf{k}$ and $\mathbf{k}'$ indices. Therefore, we can group the coefficients that generate the same frequencies from the above expression:

$$\langle \mathcal{M} \rangle = \sum_{\omega \geq 0} \sum_{\substack{\mathbf{k},\mathbf{k}'=1 \\ \Omega_{kk'}=\omega}}^{N} e^{ix\omega} \left(\sum_{i=1}^{N} M_i A^{(L)}_{ik_L} A^{*(L)}_{ik'_L} \right) A^{(0)}_{k_1 1} A^{*(0)}_{k'_1 1} \prod_{p=2}^{L} A^{(p-1)}_{k_p k_{p-1}} A^{*(p-1)}_{k'_p k'_{p-1}}$$
$$+ e^{-ix\omega} \left(\sum_{i=1}^{N} M_i A^{(L)}_{ik'_L} A^{*(L)}_{ik_L} \right) A^{(0)}_{k'_1 1} A^{*(0)}_{k_1 1} \prod_{p=2}^{L} A^{(p-1)}_{k'_p k'_{p-1}} A^{*(p-1)}_{k_p k_{p-1}}. \quad \text{(A13)}$$

Thus, we generate a Fourier series with coefficients

$$c_\omega = \sum_{\substack{\mathbf{k},\mathbf{k}'=1 \\ \Omega_{kk'}=\omega}}^{N} \left(\sum_{i=1}^{N} M_i A^{(L)}_{ik_L} A^{*(L)}_{ik'_L} \right) A^{(0)}_{k_1 1} A^{*(0)}_{k'_1 1} \prod_{p=2}^{L} A^{(p-1)}_{k_p k_{p-1}} A^{*(p-1)}_{k'_p k'_{p-1}}, \quad \text{(A14)}$$

and $c_{-\omega} = c^*_\omega$, with a frequency spectrum

$$\Omega_{\mathbf{kk'}} = \{\Lambda_{\mathbf{k}} - \Lambda_{\mathbf{k'}}\} = \{(\lambda_{k_1} + \cdots + \lambda_{k_L}) - (\lambda_{k'_1} + \cdots + \lambda_{k'_L})\}. \quad \text{(A15)}$$

Notice that the frequency spectrum is directly related to the eigenvalues of the encoding Hamiltonian, while the coefficients depend on the elements of the trainable parameters.

Appendix B: Matrix elements combinations to generate Fourier coefficients

All the terms that constitute a given coefficient with associated frequency ω have in common that $\Lambda_{\mathbf{k}} - \Lambda_{\mathbf{k'}} = \omega$ (see Eq. (A14)). In this appendix, we explore how many combinations of the multi-indices **k**, **k'** give us the same frequency ω, what we call the number of contributions to the coefficient (s_ω). In some occasions, this is also called the degeneracy of the frequency. Having more terms that contribute to the same coefficient may be beneficial since we can have more parameters yielding the same coefficient. The number of contributions does not affect the discussion about the degrees of freedom.

The frequencies are generated according to Eq. (A15). For one qudit model and with the "spin-like" encoding discussed in Sec. II, the frequencies are generated by subtracting L eigenvalues to L eigenvalues, and each of them can take N values. In total we have $|\omega| = N^{2L}$ possible eigenvalues combinations. However, some subtraction results with the same frequency ω. The number of combinations that gives rise to the same frequency is given by

$$s_\omega = \binom{2L}{L-\omega}_{N-1}. \quad \text{(B1)}$$

As expected, the number of combinations depends on the number of layers, the frequency considered and the dimension of the model. The symmetry in the generation of the frequencies is reflected by $s_\omega = s_{-\omega}$. The sum of all combinations gives back $|\omega|$, the eigenvalues combinations. $\sum_{\omega \in \Omega} s_\omega = d^{2L} = |\omega|$. For qubits ($N = 2$), the distribution of combinations becomes a binomial distribution:

$$s^{N=2}_\omega = \binom{2L}{L-\omega}_1 = \frac{2L!}{(L-\omega)!(L+\omega)!}. \quad \text{(B2)}$$

$\omega = 0$	$\omega = 1$	$\omega = 2$
$\frac{1}{2}+\frac{1}{2}-\left(\frac{1}{2}+\frac{1}{2}\right)$	$\frac{1}{2}+\frac{1}{2}-\left(-\frac{1}{2}+\frac{1}{2}\right)$	$\frac{1}{2}+\frac{1}{2}-\left(-\frac{1}{2}-\frac{1}{2}\right)$
$\frac{1}{2}-\frac{1}{2}-\left(\frac{1}{2}-\frac{1}{2}\right)$	$\frac{1}{2}+\frac{1}{2}-\left(+\frac{1}{2}-\frac{1}{2}\right)$	
$\frac{1}{2}-\frac{1}{2}-\left(-\frac{1}{2}+\frac{1}{2}\right)$	$\frac{1}{2}-\frac{1}{2}-\left(-\frac{1}{2}-\frac{1}{2}\right)$	
$-\frac{1}{2}+\frac{1}{2}-\left(-\frac{1}{2}+\frac{1}{2}\right)$	$-\frac{1}{2}+\frac{1}{2}-\left(-\frac{1}{2}-\frac{1}{2}\right)$	
$-\frac{1}{2}+\frac{1}{2}-\left(+\frac{1}{2}-\frac{1}{2}\right)$		
$-\frac{1}{2}-\frac{1}{2}-\left(-\frac{1}{2}-\frac{1}{2}\right)$		
$\boldsymbol{s_0 = 6}$	$\boldsymbol{s_1 = 4}$	$\boldsymbol{s_2 = 1}$

TABLE I. Example of the number of combinations of the positive frequencies in the single-qubit model with 2 layers and encoding gate $S(x) = e^{ixH}$ with $H = \sigma_z/2$. The combinations of the negative frequencies are obtained by swapping the eigenvalues, giving rise to the same combination value ($s_\omega = s_{-\omega}$).

For qutrits ($N = 3$), we have a trinomial distribution:

$$s_\omega^{N=3} = \binom{2L}{\omega}_2 = \sum_{\substack{0\leq\mu,\nu\leq 2L \\ \mu+2\nu=2L+\omega}} \frac{2L!}{\mu!\nu!(2L-\mu-\nu)!}. \tag{B3}$$

For ququarts, a quadrinomial distribution, etc. By using higher-dimensional systems, the number of total combinations grows exponentially with the number of layers used.

For example, let us consider a model with one qubit and $L = 2$. The frequency spectrum is given by $\Omega = \{\lambda_{k_1} + \lambda_{k_2} - (\lambda_{k_1'} + \lambda_{k_2'})\}$. The eigenvalues are $\lambda_{k_i} = \pm 1/2$ and we can combine them in $|\omega| = d^{2L} = 2^4 = 16$ ways to obtain one of the frequencies $\Omega = \{-2, -1, 0, 1, 2\}$. In Tab. I, we show the distribution of combinations to obtain these frequencies. As we can see, it fulfils $s_{-2} + s_{-1} + s_0 + s_1 + s_2 = 16$.

We can easily generalize this for the $M-$dimensional models. The discussion is valid for the Line, Parallel, and Mixed ansatz. We have coefficients associated with M frequencies in such models: $c_{\omega_1,...,\omega_M}$. Hence, the possible combinations of the coefficients depend on the combinations of every single frequency ω_i,

$$s_{\vec{\omega}} = s_{\omega_1} s_{\omega_2} ... s_{\omega_M}, \tag{B4}$$

where s_{ω_i} with $i \in \{1, ..., M\}$ is the degeneracy of the one-dimensional model given in Eq. (B1). The more terms are contributing to the sum of a coefficient, the more likely is to exist more than one set of parameters that contributes to the same coefficient.

Appendix C: Re-scaling factor

In this section, we discuss why it is relevant to introduce a re-scaling factor in quantum methods for function fitting, which can be extended to other methods. For simplicity, we first discuss the re-scaling strategy for the one-dimensional case and then generalize it to the multidimensional model.

The non-negative frequency spectrum generated by the one-qudit models with the "spin-like" encoding (without the re-scaling factor) is given by

$$\Omega = \{(d-1)L, (d-1)L - 1, ..., 0\}. \tag{C1}$$

Suppose we want to fit a $M-$dimensional function that can be approximated with a Fourier series of degree D'. We fit the function with at least $L = D'/(d-1)$ layers, under the condition of the degrees of freedom (see section III). Nevertheless, we can find cases in which there is no number of layers such that $D' = (d-1)L$. For instance, with the "spin-like" encoding, we generate integer frequencies without the possibility of generating semi-integer or float frequencies, regardless of the number of layers used. If the target function contains a non-integer frequency, the model cannot generate the proper functions and the model's training fails. We overcome this inconvenience by introducing a re-scaling η.

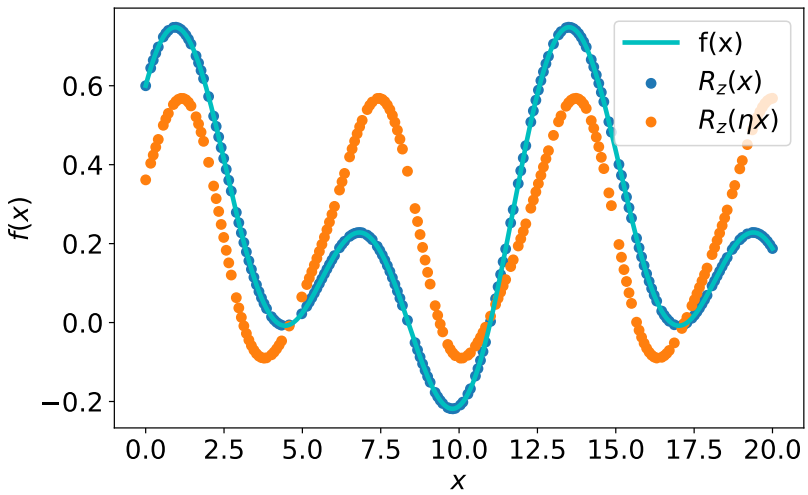

FIG. 6. Simulations of a one-dimensional Fourier serie with one qubit and two layers. The target function is $f(x) = 0.2(1 + \cos\frac{x}{2} + \sin\frac{x}{2} + \cos x + \sin x)$ and the trainable gates are $A^{(i)}(\vec{\theta}_i) = R_y(\theta_i^{(1)})R_z(\theta_i^{(2)})$ in both cases. The solid line is the target function, while the dots correspond to the output of different quantum models. The blue dots use an encoding gate $R_z(x)$ and the orange ones use $R_z(\eta x)$, where η is the re-scaling factor. The optimization subroutine finds $\eta \simeq \frac{1}{2}$.

Let's study the case of fitting a function $f(x) = 0.2(1 + \cos\frac{x}{2} + \sin\frac{x}{2} + \cos x + \sin x)$, which is decomposed in a Fourier series with frequencies $\Omega' = \{0, \pm 1/2, \pm 1\}$. With one layer, the qubit model generates the frequency $\omega = 0$ and $\omega = \pm 1$, but it cannot generate the semi-integer ones. Consequently, it fails to fit this function. A solution to avoid this is introducing a re-scaling factor η in the Hamiltonian H of the encoding gates, such that $S(x) = e^{ix\eta H} = e^{ixH'}$. Now, all the eigenvalues of the Hamiltonian become $\eta\lambda_i$, leading us to a different frequency spectrum

$$\Omega_\eta = \eta\Omega. \tag{C2}$$

The re-scaling factor η is optimized by the classical subroutine and all the other trainable parameters. In the example considered, the optimization subroutine ideally finds $\eta = \pm\frac{1}{2}$. With this, we change the eigenvalues of the encoding Hamiltonian to $\lambda' = \{-1/4, 1/4\}$. Consequently, the frequency spectrum becomes $\Omega_{\eta=1/2} = \{\pm L/2, \pm(L-1)/2, ..., 0\}$. With two layers, we have the desired frequencies present in the target function. A simulation with this example is shown in Fig. 6.

To have a more flexible model in terms of the frequencies, one can introduce different re-scaling parameters in all L processing gates to be optimized together with the rest of the model's parameters (see Ref.[2] for a similar proposal for classification problems). By doing this, we can match the target function frequencies with more freedom. The accessible frequencies for the model become

$$\Omega_{\vec{\eta}} = \{\eta_1(\lambda_{k_1} - \lambda_{k'_1}) + ... + \eta_L(\lambda_{k_L} - \lambda_{k'_L})\}. \tag{C3}$$

The Fourier series now has non-equispaced and real frequencies that can capture better the target function structure. We obtain any desired frequency present in the target with enough layers by optimizing the right parameters.

The re-scaling proposal can be easily generalized to $M-$dimensional models by introducing more re-scaling parameters: $S(\vec{\eta}\cdot\vec{x})$, where $\vec{\eta} = (\eta_1, \eta_2, ..., \eta_M)$ and $\vec{x} = (x_1, x_2, ..., x_M)$ are $M-$dimensional vectors. In this way, we modify the frequency associated with each dimension.

This method requires a small amount of classical pre-processing. Instead of generating a unitary transformation with parameter $\theta = x$, we have to generate a $\theta = \eta_i x$ transformation. In exchange for this, we obtain flexibility to fit functions with unknown frequencies. However, this approach should be treated carefully, because methods with re-scaling parameters may lead to good expressibility but overfitting and poor generalization bounds for more complex tasks.

Appendix D: Fourier series with the Line Ansatz

In this appendix, we explore the Line Ansatz (LA) for a qudit of arbitrary dimension d. We show the exact form of the parameterized states after the quantum circuit and the expectation value of an observable in this state (the

output of the quantum model). We also include a slight variation of the LA that does not contain the processing gates between each data dimension. We show why this formalism does not fit multidimensional functions in general.

We study the state after the quantum circuit of the LA model (see Fig. 3 (**a**) for the quantum circuit). We consider the "spin-like" encoding discussed in the main article for all data features, having the same frequency spectrum in all dimensions. The quantum state is given by

$$\psi_i = \sum_{\substack{j_1,\dots,j_L \\ k_1,\dots,k_L \\ t_1,\dots,t_L \\ \vdots}}^{d} A^{(L)}_{M_{it_L}} e^{ix_M\lambda t_L} ... A^{(L)}_{1_{k_L j_L}} e^{ix_1\lambda j_L} ... A^{(1)}_{M_{j_2 t_1}} e^{ix_M\lambda t_1} ... A^{(1)}_{1_{k_1 j_1}} e^{ix_1\lambda j_1} A^{(0)}_{j_1 1}, \tag{D1}$$

where λ_{j_i} are the eigenvalues of the single-qudit encoding Hamiltonian H, with $i \in \{1, \cdots, d\}$. In the symbol for a trainable gate $A^{(l)}_{m_{i,j}}$, l indicates the layer, m the position of the gate in the layer, and i, j are the indices of the matrices. Note that each eigenvalue on the exponentials, introduced by the encoding gates $S(x_i)$, has a different index. This is crucial for having a non-related dependency in each dimension and it occurs because of the intermediate trainable gates between each encoding gate. For the sake of simplicity, from now on, we assume that the M processing gates of any layer l have the same structure but taking into consideration that they have different parameters: $A^{(l)}_1 \sim A^{(l)}_2 \sim ... \sim A^{(l)}_M \equiv A^{(l)}$. With the multi-index notation $\mathbf{j} = \{j_1, j_2, ..., j_L\}$ and re-grouping terms we write

$$\psi_i = \sum_{\mathbf{j},\mathbf{k},\dots\mathbf{t}}^{d} e^{ix_1(\lambda_{j_L}+\dots+\lambda_{j_1})} ... e^{ix_M(\lambda_{t_L}+\dots+\lambda_{t_1})} A^{(L)}_{it_L} ... A^{(L)}_{k_L j_L} .. A^{(1)}_{j_2 t_1} ... A^{(1)}_{k_1 j_1} A^{(0)}_{j_1 1}. \tag{D2}$$

Finally, we define the multi-index sum: $\Lambda_{\mathbf{j}} = \lambda_{j_1} + \lambda_{j_2} + ... + \lambda_{j_L}$, simplifying the previous expression

$$\psi_i = \sum_{\mathbf{j},\mathbf{k},\dots\mathbf{t}}^{d} e^{i(x_1\Lambda_{\mathbf{j}}+\dots+x_M\Lambda_{\mathbf{t}})} A^{(L)}_{it_L} ... A^{(L)}_{k_L j_L} .. A^{(1)}_{j_2 t_1} ... A^{(1)}_{k_1 j_1} A^{(0)}_{j_1 1}. \tag{D3}$$

After this, we compute the expectation value of the observable $\mathcal{M}$. Without loss of generality, we assume that $\mathcal{M}_{ij} = \mathcal{M}_{ii} = \mathcal{M}_i$

$$\begin{aligned} \langle \mathcal{M} \rangle = \sum_{\substack{\mathbf{j},\ \mathbf{j'},\dots\mathbf{t},\ \mathbf{t'} \in [N]^L \\ \omega_1=\omega'_1,\dots,\omega_M=\omega'_M}}^{d} \sum_{i}^{N} & e^{ix_1(\Lambda_{\mathbf{j}}-\Lambda_{\mathbf{j'}})} ... e^{ix_M(\Lambda_{\mathbf{t}}-\Lambda_{\mathbf{t'}})} A^{(0)*}_{j'_1 1} A^{(1)*}_{k'_1 j'_1} ... A^{(1)*}_{j'_2 t'_1} ... A^{(L)*}_{k'_L j'_L} ... A^{(L)*}_{it'_L} \mathcal{M}_i \\ & \times A^{(L)}_{it_L} ... A^{(L)}_{k_L j_L} ... A^{(1)}_{j_2 t_1} ... A^{(1)}_{k_1 j_1} A^{(0)}_{j_1 1}. \end{aligned} \tag{D4}$$

As we can observe, the expectation value has the structure of a multidimensional Fourier series:

$$\langle \mathcal{M} \rangle = \sum_{\vec{\omega} \in \vec{\Omega}} c_{\vec{\omega}} e^{i\vec{x}\cdot\vec{\omega}}. \tag{D5}$$

Each data dimension has its frequency spectrum, resulting in a $M-$dimensional vector of frequencies:

$$\vec{\Omega}^{(LA)} = \big(\{\Lambda_{\mathbf{j}} - \Lambda_{\mathbf{j'}}\}, ..., \{\Lambda_{\mathbf{t}} - \Lambda_{\mathbf{t'}}\}\big). \tag{D6}$$

The set of coefficients

$$c_{\vec{\omega}} = \sum_{\substack{\Lambda_{\mathbf{j}}-\Lambda_{\mathbf{j'}}=\omega_1 \\ \vdots \\ \Lambda_{\mathbf{t}}-\Lambda_{\mathbf{t'}}=\omega_M}} A^{(0)*}_{j'_1 1} A^{(1)*}_{k'_1 j'_1} ... A^{(1)*}_{j'_2 t'_1} ... A^{(L)*}_{k'_L j'_L} ... A^{(L)*}_{it'_L} O_i A^{(L)}_{it_L} ... A^{(L)}_{k_L j_L} ... A^{(1)}_{j_2 t_1} ... A^{(1)}_{k_1 j_1} A^{(0)}_{j_1 1}, \tag{D7}$$

also fulfill that $c_{\omega_1,\omega_2,\cdots,\omega_M} = c^*_{-\omega_1,-\omega_2,\cdots,-\omega_M}$.

Now, we provide an alternative ansatz in which data of all dimensions is uploaded without a processing gate separating the different dimensions. Considering this, the ansatz is composed of the following layers

$$L_0 = A^{(0)}, \qquad L_i = S(x_1)S(x_2)...S(x_M)A^{(i)}. \tag{D8}$$

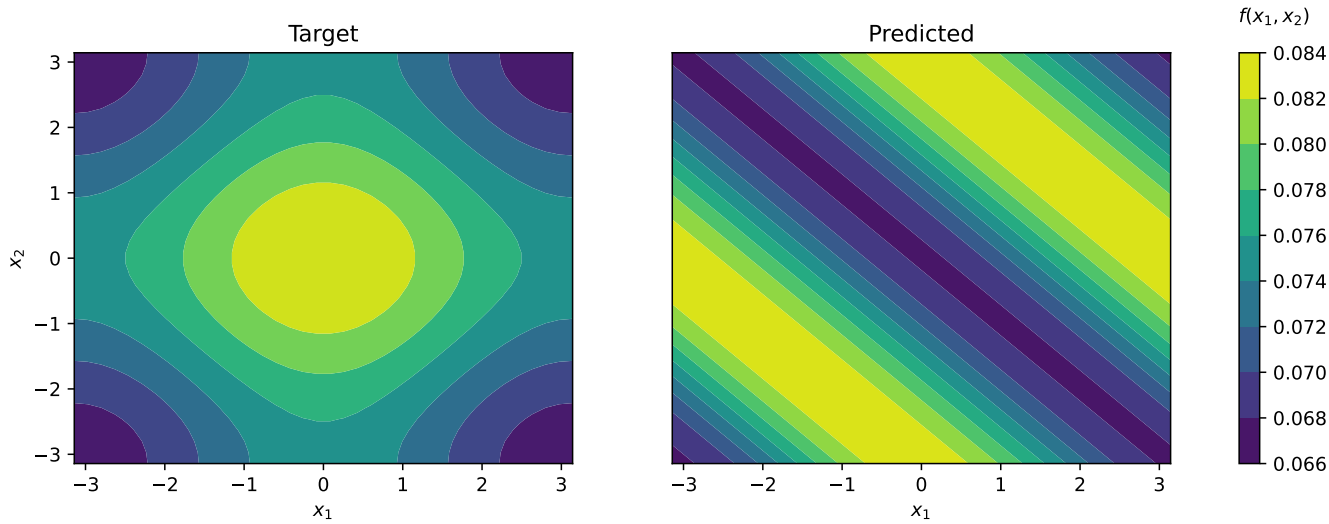

FIG. 7. Fitting of a Fourier two-dimensional series with $S(\vec{x}) = R_z(x_1)R_z(x_2)$ and 10 layers. The target function is $f(x_1, x_2) = \frac{1}{12}(1+\cos x_1+\cos x_2)$. The classical optimization method is the Nelder-Mead. We use 500 training points and 1500 test points. The predicted functions do not result to be like the target. Instead, it presents a structure similar to $f(x_1 + x_2)$.

The problem with this ansatz is that, by using the same $S(x_i)$ for all encoding gates, these gates become $S(x_1)S(x_2)\cdots S(x_M) = S(x_1 + x_2 + ... + x_M)$. Hence, all data dimensions are mapped to a single dimension in the following way: $\bar{x} = x_1 + ... + x_M$. In other words, the $M-$dimensional data collapses in a one-dimensional space. For example, the data we introduce for $M = 2$ with a qubit circuit is $\bar{x} = x_1 + x_2$. The model cannot distinguish the data points $x_1 = 1$, $x_2 = 0$, and $x_1 = 0$, $x_2 = 1$. Consequently, the results obtained by the model for fitting a two-dimensional function have a similar structure to the function $f(x_1 + x_2)$ (depicted in Fig. 7). For this reason, we need to introduce a trainable gate in the middle of the encoding gate to separate the data dependency.

Appendix E: Fourier series with the Parallel Ansatz

This section explains in more detail the Parallel Ansatz (PA) and studies its performance. With this ansatz, we take advantage of larger Hilbert spaces for operations with the trainable gates, which determine the coefficients of the Fourier series. First, each data dimension is encoded in a different qudit. Then, we present the resulting state after the parameterized quantum circuit and the expectation value of an arbitrary observable $\mathcal{M}$ in this state. We also provide a variation of this ansatz that does not use entanglement in the trainable gates and we discuss the differences between the two models.

Following the circuit of the PA in Fig. 3 (**b**), the $M-$qudit state after the circuit is given by

$$\psi_i = \sum_{j_1,\dots,j_L=1}^{N=d^M} A^{(L)}_{ij_L} e^{i(x_1\lambda^{(1)}_{j_L}+\dots+x_M\lambda^{(M)}_{j_L})} \dots A^{(1)}_{j_2j_1} e^{i(x_1\lambda^{(1)}_{j_1}+\dots+x_M\lambda^{(M)}_{j_1})} A^{(0)}_{j_11}. \tag{E1}$$

Now the indices run from 1 to $N = d^M$, being $M = n$, the number of qudits, and d its dimension. The encoding gate is $S(\vec{x}) = S(x_1) \otimes S(x_2) \otimes ... \otimes S(x_M) = e^{ix_1H} \otimes e^{ix_2H} ... e^{ix_MH}$, which can be rewritten as $S(\vec{x}) = e^{i(x_1H^{(1)}+x_2H^{(2)}+\dots+x_MH^{(M)})}$, with

$$H^{(i)} = \underbrace{\mathbb{I}}_{1} \otimes \underbrace{\mathbb{I}}_{2} \otimes \cdots \otimes \underbrace{H}_{i} \otimes \cdots \otimes \underbrace{\mathbb{I}}_{M}. \tag{E2}$$

All the Hamiltonians are built by the tensor product of $n-1$ identities except in the i position where we have the single-qudit Hamiltonian. This gives a Hamiltonian of dimension $N = d^M$, which has the d eigenvalues λ_i of the single-qudit Hamiltonian distributed along the $N = d^M$ possible positions on the diagonal. For example, a qubit model with two-dimensional data ($M = 2$) and the "spin-like" encoding ($H = \sigma_z/2$) has the following encoding gate: $S(x_1, x_2) = e^{ix_1H} \otimes e^{ix_2H} = e^{i(x_1H^{(1)}+x_2H^{(2)})}$. The eigenvalues of the two Hamiltonians are: $\lambda^{(1)} =$

$\{-\frac{1}{2},-\frac{1}{2},+\frac{1}{2},+\frac{1}{2}\}$, and $\lambda^{(2)}=\{-\frac{1}{2},+\frac{1}{2},-\frac{1}{2},+\frac{1}{2}\}$. Therefore, the eigenvalues of $\sigma_z/2$ are maintained but distributed differently. Continuing with the quantum state, we group the different eigenvalues with the multi-index sum notation, i.e., $\Lambda^{(i)}_{\mathbf{j}}=\lambda^{(i)}_{j_1}+\lambda^{(i)}_{j_2}+...+\lambda^{(i)}_{j_L}$, we rewrite the final state as

$$\psi_i=\sum_{\mathbf{j}}^{N=d^n} e^{i(x_1\Lambda^{(1)}_{\mathbf{j}}+...+x_M\Lambda^{(M)}_{\mathbf{j}})}A^{(L)}_{ij_L}...A^{(1)}_{j_2j_1}A^{(0)}_{j_11}. \tag{E3}$$

All data dimensions in the exponential are multiplied by a multi-index sum with the same index, which may cut some freedom in fitting a Fourier series. Although having the same index, they are non-equally distributed. With this, the model has enough freedom to have an independent frequency spectrum in each dimension. Finally, the expectation value of the observable is given by

$$\langle\mathcal{M}\rangle=\sum_{\mathbf{j},\mathbf{j'}=1}^{N}\sum_{i=1}^{N} e^{i[x_1(\Lambda^{(1)}_{\mathbf{j}}-\Lambda^{(1)}_{\mathbf{j'}})+...+x_M(\Lambda^{(M)}_{\mathbf{j}}-\Lambda^{(M)}_{\mathbf{j'}})]}A^{(0)*}_{j'_11}A^{(1)*}_{j'_2j'_1}...A^{(L)*}_{ij'_L}M_iA^{(L)*}_{ij_L}...A^{(1)}_{j_2j_1}A^{(0)}_{j_11}, \tag{E4}$$

which has the structure of a multidimensional Fourier series. The observable $\mathcal{M}$ acts on n qudits, but we propose to measure only one qudit, for example, the first one. Then the observable becomes $\mathcal{M}=\mathbf{M}\otimes\mathbb{I}\otimes...\otimes\mathbb{I}$, where $\mathbf{M}$ is a single-qudit observable. This does not affect the models' structure. The frequency spectrum obtained is

$$\vec{\Omega}^{(PA)}=(\{\Lambda^{(1)}_{\mathbf{j}}-\Lambda^{(1)}_{\mathbf{j'}}\},...,\{\Lambda^{(M)}_{\mathbf{j}}-\Lambda^{(M)}_{\mathbf{j'}}\}). \tag{E5}$$

Therefore, the PA's output has a multidimensional Fourier series structure.

Now we explore a variation of the PA. The circuit ansatz we consider is the same, but instead of using processing gates $A^{(i)}$ as multi-qudit gates, we only use single-qudit gates, meaning that we need to eliminate any entangling gate. In this way, we explore the outcome of the model working with each qudit separately. Now we measure an observable acting in all qudits because they do not interact and consequently do not share any correlation. The state after the circuit with the ansatz considered is

$$|\psi\rangle=A^{(0)}S(\vec{x})A^{(1)}...S(\vec{x})A^{(L)}=|\psi_1\rangle\otimes|\psi_2\rangle\otimes\cdots\otimes|\psi_M\rangle\,. \tag{E6}$$

where $A^{(i)}=A^{(i)}_1\otimes A^{(i)}\otimes...\otimes A^{(i)}_M$, and $S(\vec{x})=S(x_1)\otimes S(x_2)\otimes...\otimes S(x_M)$. Hence, if we choose to measure an observable $\mathcal{M}=\sigma_z\otimes\sigma_z\otimes....\otimes\sigma_z$, the result is

$$\begin{aligned}\langle\mathcal{M}(\vec{x})\rangle=\langle\psi_1|\,\sigma_z\,|\psi_1\rangle\,\langle\psi_2|\,\sigma_z\,|\psi_2\rangle...\,\langle\psi_M|\,\sigma_z\,|\psi_M\rangle&=\sum_{\omega_1}c^{(1)}_{\omega_1}e^{ix_1\omega_1}\sum_{\omega_2}c^{(2)}_{\omega_2}e^{ix_1\omega_2}...\sum_{\omega_M}c^{(M)}_{\omega_M}e^{ix_1\omega_M}\\&=\sum_{\vec{\omega}}c^{(1)}_{\omega_1}c^{(2)}_{\omega_2}...c^{(M)}_{\omega_M}e^{i(x_1\omega_1+x_2\omega_2+...+x_M\omega_M)},\end{aligned} \tag{E7}$$

where we have used the results of the one-dimensional Fourier series discussed in App. A. We obtain a multiplication of M one-dimensional Fourier series. In this case, however, we do not have all the free coefficients because $c_{\vec{\omega}}\neq c^{(1)}_{\omega_1}c^{(2)}_{\omega_2}...c^{(M)}_{\omega_M}$. For example, for $M=2$ we have 5 free coefficients: $c_{00},c_{01},c_{10},c_{11},\ c_{1-1}$ but in this variation of the model we have 4 free coefficients: $c^{(1)}_0,c^{(2)}_0,c^{(1)}_1,c^{(2)}_1$, since the other ones are constrained by $c^{(i)}_{\omega_i}=c^{(i)*}_{-\omega_i}$. The output of this model is still a Fourier series, but not the most general one. Regardless of this, the possibility of finding applications for some problems is not discarded, as the question remains open.

Appendix F: Fourier series with the Mixed Ansatz

In this section, we explore the last ansatz considered. The circuit in Fig. 3 (**c**) is called the Mixed Ansatz (MA) because it combines elements of the LA and the PA: various data dimensions are encoded in the same qudit as in the LA, but also other qudits are used to encode them, like in the PA. The idea of this model is to take advantage of the hardware requirements of the current quantum devices. For instance, if we have a chip with 4 qubits and 8 data features, we introduce 2 dimensions in each qubit.

Let us consider a MA circuit of p qudits and a $M-$dimensional dataset. In the ideal case, we encode M/p data dimensions in every qudit. Of course, M/p will not always be an integer. In such cases, we use fewer encoding gates in the last layer. Since we are encoding different data features in the same qudit, we need extra processing gates

between these dimensions, like in the LA. All the trainable gates are p-dimensional qudit gates since we encode data in this number of qudits. The quantum state after the circuit is given by

$$\begin{aligned}\psi_i = \sum_{\substack{j_1,...,j_L\\k_1,...,k_L\\t_1,...,t_L}}^{N=d^p} & A^{(L)}_{M/p_{it_L}} e^{i(x_{M-(p-1)}\lambda^{(1)}_{t_L}+...+x_M\lambda^{(p)}_{t_L})}...A^{(L)}_{1_{k_L}j_L} e^{i(x_1\lambda^{(1)}_{j_L}+...x_p\lambda^{(p)}_{j_L})}...\\ &\vdots\\ &\times A^{(1)}_{M/p_{j_2t_1}} e^{i(x_{M-(p-1)}\lambda^{(1)}_{t_1}+...+x_M\lambda^{(p)}_{t_1})}...A^{(1)}_{1_{k_1}j_1} e^{i(x_1\lambda^{(1)}_{j_1}+...+x_p\lambda^{(p)}_{j_1})}A^{(0)}_{j_11}.\end{aligned} \tag{F1}$$

This equation has a different distribution of eigenvalues: $\lambda^{(i)}$ with $i \in \{1, ..., p\}$ because we have p qudits. The different distribution comes from the tensor product of the identity in all the p qudits and the encoding Hamiltonian in the $i-$th qudit (see Eq. (E2)). By introducing the multi-index and multi-sum notation that we have explored in this work, we reduce the expression to

$$\psi_i = \sum_{\mathbf{j},\mathbf{k},\mathbf{t},...}^{N=d^p} e^{i(x_1\Lambda^{(1)}_{\mathbf{j}}+...+x_p\Lambda^{(p)}_{\mathbf{j}}+...+x_{M-(p-1)}\Lambda^{(1)}_{\mathbf{t}}+...+x_M\Lambda^{(p)}_{\mathbf{t}})} A^{(L)}_{M/p_{it_L}}...A^{(L)}_{1_{k_L}j_L}...A^{(1)}_{M/p_{j_2t_1}}...A^{(1)}_{1_{k_1}j_1}A^{(0)}_{j_11}, \tag{F2}$$

where $\mathbf{j} = \{j_1, j_2, ..., j_L\}$ and $\Lambda^{(i)}_{\mathbf{j}} = \lambda^{(i)}_{j_1} + \lambda^{(i)}_{j_2} + ... + \lambda^{(i)}_{j_L}$. The expectation value of an observable $\mathcal{M}$ in this quantum state is given by

$$\begin{aligned}\langle\mathcal{M}\rangle = \sum_{\mathbf{j},\mathbf{j'},\mathbf{k},\mathbf{k'},\mathbf{t},\mathbf{t'}...}^{N=d^p} \sum_{i=1}^{N} & e^{i[x_1(\Lambda^{(1)}_{\mathbf{j}}-\Lambda^{(1)}_{\mathbf{j'}})+...+x_p(\Lambda^{(p)}_{\mathbf{j}}-\Lambda^{(p)}_{\mathbf{j'}})+...+x_{M-(p-1)}(\Lambda^{(1)}_{\mathbf{t}}-\Lambda^{(1)}_{\mathbf{t'}})+...+x_M(\Lambda^{(p)}_{\mathbf{t}}-\Lambda^{(p)}_{\mathbf{t'}})]} A^{(0)*}_{j'_11}A^{(1)*}_{1_{k'_1}j'_1}...\\ &\times A^{(1)*}_{M/p_{j'_2t'_1}}...A^{(L)*}_{1_{k'_L}j'_L}...A^{(L)*}_{M/p_{it'_L}}\mathcal{M}_i A^{(L)}_{M/p_{it_L}}...A^{(L)}_{1_{k_L}j_L}...A^{(1)}_{M/p_{j_2t_1}}...A^{(1)}_{1_{k_1}j_1}A^{(0)}_{j_11}.\end{aligned} \tag{F3}$$

In terms of the eigenvalues, we see combined features of the two models: having a different distribution of eigenvalues with the same index (in the data dimensions processed at the same level of depth) and having a different index with the same distribution of eigenvalues (the data-features processed in the same qudit). Therefore, this model also generates a multidimensional Fourier series.

Appendix G: Fourier series with the Super-parallel Ansatz

In this section, we introduce the Super-parallel Ansatz. It contains $n = ML$ qudits. The ansatz in question is depicted in Fig. 3 (**d**). It can be seen that now the layers grow in two dimensions: in width and depth. Each layer comprises L encoding blocks, with each block consisting of M single-qudit encoding gates. Similar to the previous ansatz, the Super-parallel Ansatz also generates multidimensional Fourier series, and the derivation is similar to that of the Parallel Ansatz. Thus, we will not provide an explicit derivation here. However, the encoding gates are significantly different in this ansatz. For instance, consider an encoding block of the Super-parallel Ansatz with $L = 2$ for M-dimensional data:

$$\begin{aligned}S(\vec{x}) =& S(x_1) \otimes ... \otimes S(x_M) \otimes S(x_1) \otimes ... \otimes S(x_M) = e^{ix_1H} \otimes ... \otimes e^{ix_MH} \otimes e^{ix_1H} \otimes ... \otimes e^{ix_MH}\\ \sum_{j_1=1}^{d^{LM}} & e^{i(x_1\lambda^{(1)}_{j_1}+...+x_M\lambda^{(M)}_{j_1}+x_1\lambda^{(M+1)}_{j_1}+...+x_M\lambda^{(2M)}_{j_1})} = \sum_{j_1=1}^{d^{LM}} e^{i[x_1(\lambda^{(1)}_{j_1}+\lambda^{(M+1)}_{j_1})+...+x_M(\lambda^{(M)}_{j_1}+\lambda^{(2M)}_{j_1})]},\end{aligned} \tag{G1}$$

where $\lambda^{(1)}_{j_1}$ is given in Eq. (E2). We see that each data feature in a single encoding block is multiplied by a sum of L eigenvalues (in this case $L = 2$). Consequently, the output Fourier series degree D becomes additionally dependent on the number of layers in depth (the number of encoding blocks for a data feature per layer), yielding $D = (d - 1)L^2$. The rest of the derivations can be deduced using a similar approach as in Appendix E.

Appendix H: Practical case: fitting 4-dimensional data

In this appendix, we fit a 4-dimensional Fourier series of degree D with the three ansatzes described in this work: the LA, PA, and MA. We compare which models are more convenient for fitting higher-degree Fourier series. The

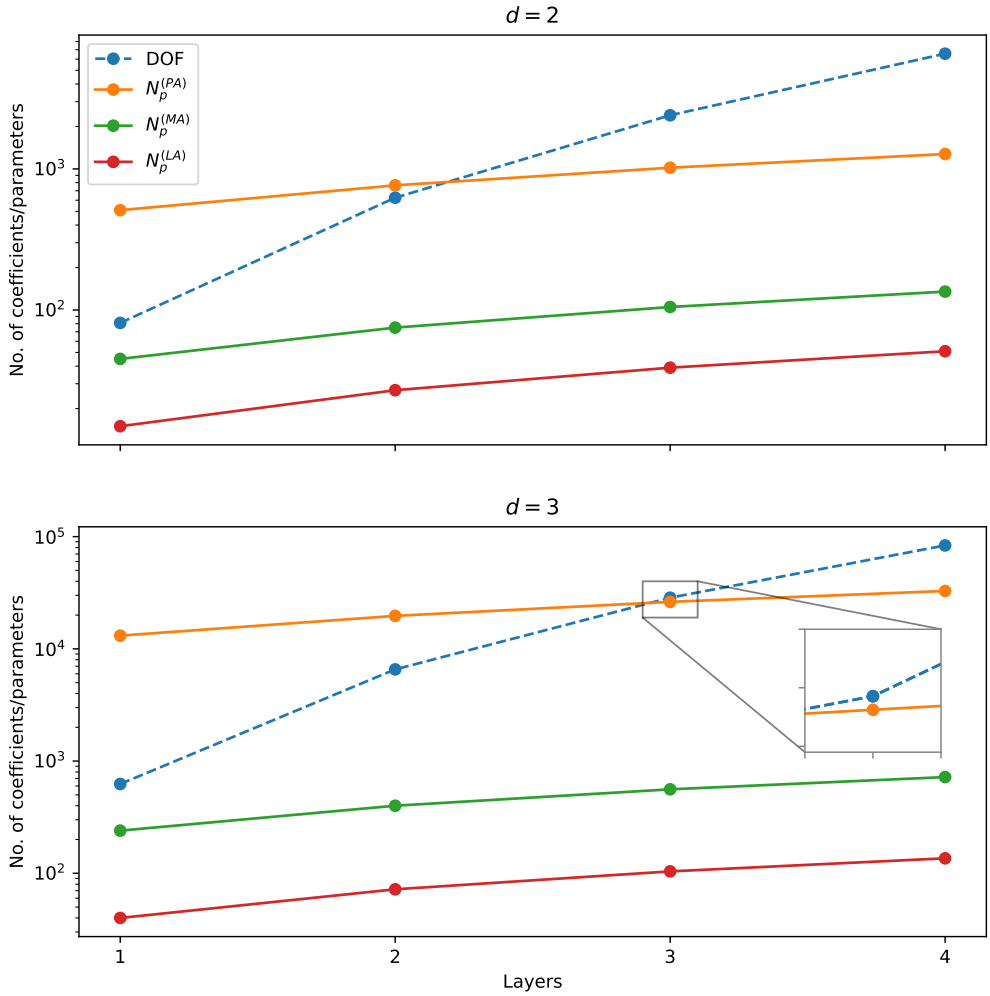

FIG. 8. Graphic representation of the degrees of freedom and the free parameters with respect to the number of layers for the Line, Parallel, and Mixed ansatzes models with 4−dimensional data. The upper plot depicts results for qubits and the lower for qutrits. The y-axis is logarithmic and represents the number of coefficients or parameters and the x-axis is the number of layers in the models. The only model that accomplishes the DOF condition is the PA and the results are improved when we use qutrits. With qubits, the condition is satisfied until the Fourier series of degree $D = (d-1)L = 2$ and $D = 4$ for qutrits.

general target function is given by

$$f(x_1, x_2, x_3, x_4) = f(\vec{x}) = \sum_{\vec{\omega}=-D}^{D} c_{\vec{\omega}} e^{i\vec{x}\cdot\vec{\omega}}, \tag{H1}$$

where $\vec{\omega} = (\omega_1, \omega_2, \omega_3, \omega_4)$ describes the frequency in each dimension. The number of free coefficients N_c in the 4-dimensional Fourier series is determined by the degree of the series:

$$N_c = \frac{(2D+1)^4 - 1}{2} + 1, \tag{H2}$$

since half of them are constrained by $c_{\vec{\omega}} = c^*_{-\vec{\omega}}$, except $c_{\vec{0}}$. At the same time, the degree of the Fourier series generated by the model is determined by the qudit used and the number of layers of the model: $D = (d-1)L$. Hence, N_c depends on d, and L. We know that $\nu \equiv 2N_c - 1$ is the number of degrees of freedom required for the model to fit

the general Fourier series. The number of free parameters of the different models for 4−dimensional data is given by

$$\begin{aligned} N_p^{(LA)} &= (4L+1)(d^2-1), \\ N_p^{(PA)} &= (L+1)(d^8-1), \\ N_p^{(MA)} &= (2L+1)(d^4-1), \end{aligned} \tag{H3}$$

where for the MA we use two qudits, therefore $p = 2$. With this, we determine graphically the Fourier series degree that accomplishes the DOF condition. This is shown in Fig. 8 for qubit and qutrit models. As we can notice, for 4−dimensional data, both the LA nor the MA do not accomplish the DOF condition for any number of layers. This means we do not have enough free parameters in the model for fitting an arbitrary Fourier series. With the PA model, the condition is satisfied until $L = 2$, meaning that we achieve Fourier series of degree $D = 2(d-1)$, having degree 2 for the qubit model and 4 for the qutrit. Hence, this model can give limited approximations to some functions.