Quantum algorithms for approximate function loading

Gabriel Marin-Sanchez [1], Javier Gonzalez-Conde [1,2,*] and Mikel Sanz [1,2,3,4,†]

[1]*Department of Physical Chemistry, University of the Basque Country UPV/EHU, Apartado 644, 48080 Bilbao, Spain*
[2]*EHU Quantum Center, University of the Basque Country UPV/EHU, Apartado 644, 48080 Bilbao, Spain*
[3]*IKERBASQUE, Basque Foundation for Science, Plaza Euskadi 5, 48009 Bilbao, Spain*
[4]*Basque Center for Applied Mathematics (BCAM), Alameda de Mazarredo, 14, 48009 Bilbao, Spain*

(Received 21 December 2021; revised 18 April 2023; accepted 12 July 2023; published 18 August 2023)

Loading classical data into quantum computers represents an essential stage in many relevant quantum algorithms, especially in the field of quantum machine learning. Therefore, the inefficiency of this loading process means a major bottleneck for the application of these algorithms. Here, we introduce two approximate quantum-state preparation methods for the noisy intermediate-scale quantum era inspired by the Grover-Rudolph algorithm, which partially solve the problem of loading real functions. Indeed, by allowing for an infidelity ϵ and under certain smoothness conditions, we prove that the complexity of the implementation of the Grover-Rudolph algorithm without ancillary qubits, first introduced by Möttönen *et al.*, results into $O(2^{k_0(\epsilon)})$, with n the number of qubits and $k_0(\epsilon)$ asymptotically independent of n. This leads to a dramatic reduction in the number of required two-qubit gates. Aroused by this result, we also propose a variational algorithm capable of loading functions beyond the aforementioned smoothness conditions. Our variational *Ansatz* is explicitly tailored to the landscape of the function, leading to a quasioptimized number of hyperparameters. This allows us to achieve high fidelity in the loaded state with high speed convergence for the studied examples.

DOI: 10.1103/PhysRevResearch.5.033114

I. INTRODUCTION

Quantum computing has triggered a great interest in the past decades due to its theoretical capability to outperform classical information processing. Even though noise and decoherence are major drawbacks for the computational capacity of current quantum computers, quantum advantage has been experimentally achieved [1–3]. Unfortunately, these accomplishments lack any industrial or scientific relevance, so the search of a useful application still remains. In this sense, the realistic experimental implementation of many promising quantum algorithms in several fields like solving systems of linear equations [4,5], performing data fitting [6], computing scattering cross sections [7,8], pricing financial derivatives [9–11], or initial conditions in differential equations [12–14] is constrained by the assumption that data can be efficiently loaded into a quantum device. In this context, the efficient loading of classical data into quantum computers is a particularly important problem and represents a major bottleneck of the practical application of quantum computation in the noisy intermediate-scale quantum (NISQ) era, especially with the emergence of the quantum machine learning field [15–20].

There exist different quantum embedding techniques transforming classical data into quantum information [18,21]. In particular, we can distinguish two main embedding protocols depending on how the information is encoded. On the one hand, there is the basis embedding, in which each bit value "0" or "1" is mapped to computational qubit state $|0\rangle$ or $|1\rangle$, respectively [22]. In this way, the embedded quantum state corresponds to a uniform superposition of the bitwise translations of binary strings. On the other hand, the amplitude-embedding technique encodes the normalized vector of classical data, which is now not necessarily binary, into the amplitudes of a quantum state [23–35]. In particular, these feature maps have been proposed to load discretized real valued functions [24–27] with relevant applications in loading initial conditions for solving partial derivatives equations [11–14], computing Monte Carlo integrations [9,10,36], and quantum field theory [37,38]. However, a practical implementation of these approaches generally incurs into an overhead of resources, which can be reflected into either an exponential number of entangling gates [26–29,39,40] or the employment of a huge number of ancillary qubits [31–33]. A rather different approach sustained by the Solovay–Kitaev theorem [41,42] is based on the application of quantum generative models to efficiently accomplish an approximate amplitude encoding of discretized real valued functions [43,44]. Nonetheless, in these generic *Ansätze*, increasing the number of hyperparameters does not necessarily reflect in improving the expressability of the *Ansatz* to capture the function details [45]. Additionally, these variational methods usually suffer from training problems such as local minima and barren plateaus [46].

*Corresponding author: javier.gonzalezc@ehu.eus
†Corresponding author: mikel.sanz@ehu.eus

In this article, we present two approximate quantum algorithms to load real functions into quantum computers for the NISQ era. Our first protocol, inspired by the Grover-Rudolph algorithm and its implementation proposed by Möttönen *et al.* [26,27], implements the algorithm without ancillary qubits with complexity $O(2^{k_0(\epsilon)})$ for functions whose second logarithm derivative is upper bounded by a constant, with n the number of qubits, ϵ the infidelity with respect to the exact state, and $k_0(\epsilon)$ asymptotically independent of n. This leads to a dramatic reduction in the number of required two-qubit gates. Inspired by this result, we also introduce and benchmark a variational quantum circuit with applications in a broader family of functions. Our proposed *Ansatz* is adapted to the structure of the function, which intuitively correlates hyperparameters and expressibility. Moreover, by taking the angles provided by the Grover-Rudolph protocol, we can define a suitable initial training angle set, which considerably improves the training process, avoiding barren plateaus and local minima. Finally we have numerically proven resilience of our algorithm against several noise sources.

II. GROVER-RUDOLPH ALGORITHM

The method of Grover and Rudolph, originally proposed in Ref. [26], describes a constructive protocol to load into an n-qubit quantum state the discretized version $\{f_i\}$ of certain integrable density function $f : [x_{\min}, x_{\max}] \subset \mathrm{IR} \to \mathrm{IR}^+$ as

$$|\Psi(f)\rangle_n = \sum_{i=0}^{2^n-1} \sqrt{f_i}|i\rangle. \tag{1}$$

A. Grover and Rudolph algorithm without ancillas

The first proposal in the literature that provided an explicit circuit implementation of the Grover-Rudolph idea without using ancillary qubits was proposed by Möttönen *et al.* [27]. Without using ancillas, this protocol provides a constructive algorithm which applies a sequence of operation blocks, $F_k^{(k-1)}(\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$, to the initial state $|0\rangle^{\otimes n}$. For $k \in \{1, \ldots, n\}$, each k-qubit block, $F_k^{(k-1)}(\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}) = \sum_{l=0}^{2^{k-1}-1} |l\rangle\langle l| \otimes R_y(\theta_l^{(k-1)})$, corresponds to a uniformly controlled y-axis rotation, where the lth component of the angle vector $\boldsymbol{\theta}^{(k-1)}$ is calculated as

$$\theta_l^{(k-1)}(l) = 2 \arccos\left(\sqrt{\frac{\int_{x_{\min}+l\delta_k}^{x_{\min}+(l+1/2)\delta_k} f(x)dx}{\int_{x_{\min}+l\delta_k}^{x_{\min}+(l+1)\delta_k} f(x)dx}}\right). \tag{2}$$

Here, $\delta_k = \frac{x_{\max}-x_{\min}}{2^{k-1}}$ and $l \in \{0, \ldots, 2^{k-1}-1\}$ is the index corresponding to the $(l+1)$th subinterval of the 2^{k-1} partition of the interval $[x_{\min}, x_{\max}]$. Each multicontrolled gate comprising $F_k^{(k-1)}(\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$ is denoted by $\wedge_{k-1}[R_y(\theta_l^{(k-1)})]$ and bisects the $(l+1)$th partition interval of the function by using the conditional probability of being in the right or left side of the interval, as depicted in Fig. 1.

In the complete algorithm, the total number of angles needed scales as $\sum_{m=1}^{n-1} 2^{m-1} = 2^n - 1$, which is exponential in the number of qubits and requires an exponential number of multicontrolled CNOT gates. Therefore, without using ancillary qubits, the required number of two qubit gates scales exponentially as the number of angles required do, and therefore the circuit complexity of implementing the Grover-Rudolph algorithm without ancillas is $O(2^n)$.

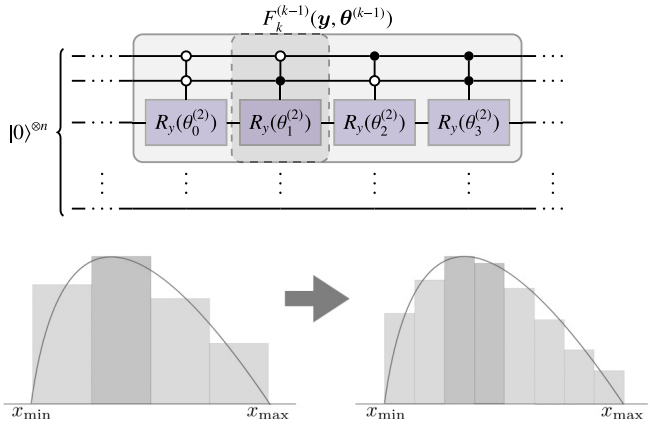

FIG. 1. Effect of the uniformly controlled y-axis rotation $F_k^{(k-1)}(\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$ for $k = 3$ and n qubits. Each multicontrolled gate comprising $F_k^{(k-1)}(\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$ bisects the $(l+1)$th partition interval of the function by using the conditional probability of being in the right or left side of the interval.

We can conclude that this protocol without ancillas is theoretically capable of loading a discretized density function at the cost of an exponential overhead of resources [39,40] to prepare the state in Eq. (1) using blocks $F_k^{(k-1)}(\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)})$.

B. Grover and Rudolph algorithm with ancillas

According to the original Grover-Rudolph algorithm [26], for each step k, we can efficiently prepare $|\Psi_k\rangle = \sum_{l=0}^{2^k-1} \sqrt{f_l^{(k)}}|l\rangle$ by first loading the rotation angles, Eq. (2), with bit precision m into a bit string of m ancillary registers and then performing k controlled rotations of angle $2\pi/2^j$, $j = 1, \ldots, m$ controlled by the ancillary registers. Indeed, the encoding of the rotation angles can be efficiently achieved if we have access to the oracle:

$$\begin{aligned} |\Psi_k\rangle &= \sum_{l=0}^{2^k-1} \sqrt{f_l^{(k)}}|l\rangle|0\rangle^{\otimes m} \to |\Psi_k\rangle \\ &= \sum_{l=0}^{2^k-1} \sqrt{f_l^{(k)}}|i\rangle \underbrace{|\theta_l^{(k-1)}(l)\rangle}_{\text{m bit precision}}. \end{aligned} \tag{3}$$

An example case when this operation can be efficiently performed is when the function $\theta_l^{(k-1)}(l)$ can be well approximated by a polynomial and then implemented by employing a polynomial amount of classical half adder operations, i.e., NAND gates. Lastly, the NAND gates are efficiently mapped into Toffoli quantum gates by making use of at most three qubits per classical bit [47]. As this map is input dependent, it enables us to compute it simultaneously for all $\theta_l^{(k-1)}(l)$ when the input state is a quantum superposition. However, this oracle is not explicitly provided in the original manuscript and the efficient circuit to which the authors refer is not presented, remaining as an oracle. Additionally, the original manuscript of Grover and Rudolph does not provide the

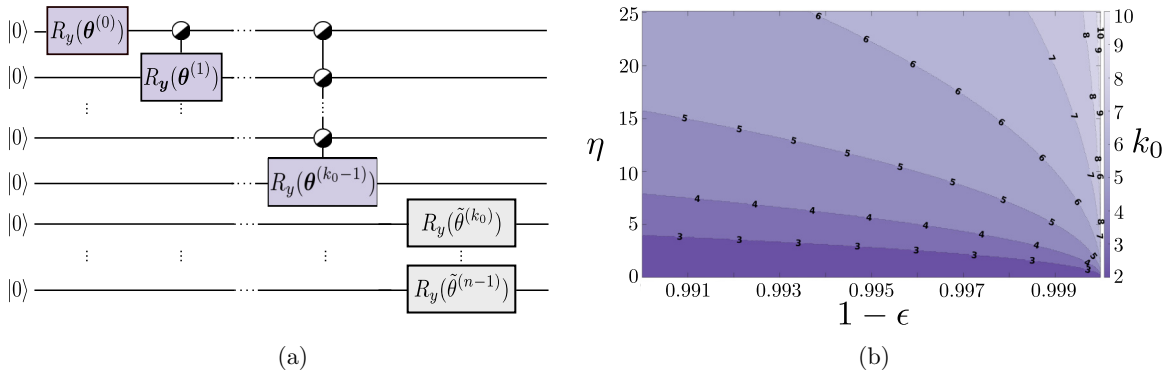

FIG. 2. (a) Quantum circuit performing the protocol presented in Theorem 4, based on the Grover-Rudolph method for a system of n qubits. We cluster the angles of the blocks for $k \geqslant k_0 + 1 > 2$, leading to a drastic reduction in the number of gates needed. (b) Values of k_0 for $\eta \in [0, 8\pi]$, $\epsilon \in [0.0001, 0.01]$, and $n \to \infty$, following Eq. (10), where the dotted lines correspond to the contour lines and denote the change of values. We can appreciate that $k_0 \leqslant 10$ for most of the cases.

analytical bounds of these approximations (m bit precision; oracle implementation), as well as implicitly makes use of additional ancillary qubits which incurs into an important cost for the NISQ era. Last but not least, some recent works have risen criticism about the feasibility of this original proposal [14,48,49]. Moreover, to our best knowledge, there is no explicit efficient implementation of this oracle in terms of gates without employing ancillary qubits.

III. FIRST ALGORITHM

Inspired by the Grover-Rudolph algorithm [26], we present an efficient method to encode discretized density functions into quantum states. By permitting an error in the final state and assuming certain smoothness conditions, an angle clustering significantly reduces the required entangling gates.

Definition 1. Let $f : [0, 1] \to \mathrm{IR}^+$ be a positive function in $L^2([0, 1])$. We define the n-qubit normalized representative state of $f(x)$ as the n-qubit state $|f(x)\rangle_n = \sum_{l=0}^{2^n-1} f(l\delta_n)|l\rangle$, with $\delta_n = \frac{1}{2^n-1}$ and $\sum_{l=0}^{2^n-1} f^2(l\delta_n) = 1$.

According to this definition, we encode the discretized function into the amplitude of a quantum state, in contrast to the Grover-Rudolph algorithm, which does it in the probability, Eq. (1). We also consider our function defined in the interval $[0, 1]$ as a standardization criterion.

Theorem 1. Let $f : [0, 1] \to \mathrm{IR}^+$ be a positive integrable function in $L^2([0, 1])$ and $0 \leqslant \eta \leqslant 8\pi$ a constant such that $\eta = \sup_{x\in[0,1]} |\partial_x^2 \log f^2(x)|$. Then, it is possible to approximate the n-qubit representative state of $f(x)$, $|f(x)\rangle_n$, by a quantum state $|\Psi(f)\rangle_n$ such that the fidelity $|\langle\Psi(f)|f(x)\rangle_n|^2 \geqslant 1 - \epsilon$ with at most $2^{k_0(\epsilon)} - 1$ two-qubit gates, with

$$k_0(\epsilon) = \max\left\{\left\lceil -\frac{1}{2}\log_2\left(4^{-n} - \frac{96}{\eta^2}\log(1-\epsilon)\right)\right\rceil, 2\right\}, \quad (4)$$

and the circuit to perform it is provided in Fig. 2(a).

Let us analyze each part of the algorithm. First, we provide a sufficient condition on the target density function to guarantee an upper bound over the difference between two contiguous angles.

Lemma 1. Let f be a continuous function such that $f : [0, 1] \to \mathrm{IR}^+$ and consider a block comprising a uniformly controlled rotation of k qubits. Then, the difference between two contiguous angles is bounded by the second derivative of the logarithm of f in the following way:

$$\left|\theta_{l+1}^{(k-1)} - \theta_l^{(k-1)}\right| \leqslant \frac{\delta_k^2}{4} \max_{y'\in[l\delta_k,(l+1)\delta_k]} \left|\left(\partial_y^2 \log f(y)\right)\big|_{y=y'}\right|, \quad (5)$$

where $\delta_k = \frac{1}{2^{k-1}}$, $l \in \{0, \ldots, 2^{k-1} - 2\}$.

Consequently, if $|\partial_y^2 \log f(y)| \leqslant \eta$, then for each block comprising a uniformly controlled rotation of k qubits, the difference between any two angles is bounded by

$$\left|\theta_l^{(k-1)} - \theta_{l'}^{(k-1)}\right| \leqslant \frac{\delta_k}{4}\eta, \quad \forall l, l' \in \{0, \ldots, 2^k - 1\}. \quad (6)$$

This result allows us to cluster angles of each block in which Eq. (6) is fulfilled. Thus we define a cluster representative angle, $\tilde{\theta}^{(k-1)}$, as

$$\left|\tilde{\theta}^{(k-1)} - \theta_l^{(k-1)}\right| \leqslant \frac{\delta_k}{8}\eta := \eta_k, \quad \forall l \in \{0, \ldots, 2^k - 1\}, \quad (7)$$

$\forall k \geqslant k_0$, with $k_0 + 1$ the index of the first block in which Eq. (6) is fulfilled (successive blocks also verify it). Note that $|\partial_y^2 \log f(y)| \leqslant \eta$ is a sufficient condition for the bound in Eq. (5); however, if in a singularity point it grew slower than $\frac{1}{\delta_k^2}$, the difference between consecutive angles still vanishes as we will analyze later.

We now analyze how clustering the angles according to Eq. (7) affects the final error of the process, measured by means of the fidelity with respect to the exact discretized state $|f(x)\rangle_n$. Considering an n-qubit system, the unitary gate to prepare the quantum state representing the target density function can be written in terms of the blocks as $\mathfrak{U}_n = \boldsymbol{\mathcal{U}}_{n-1}(\boldsymbol{\theta}^{(n-1)}) \ldots \boldsymbol{\mathcal{U}}_0(\boldsymbol{\theta}^{(0)})$, where we define $\boldsymbol{\mathcal{U}}_{k-1}(\boldsymbol{\theta}^{(k-1)}) := F_k^{k-1}(\mathbf{y}, \boldsymbol{\theta}^{(k-1)}) \otimes \mathbb{1}^{\otimes(n-k)}$.

Let $\tilde{\mathfrak{U}}_n$ denote the operation $\mathfrak{U}_n$ when the rotation angles corresponding to each block k are replaced by a representative, $\tilde{\theta}^{(k-1)}$, such that its difference with any angle of the block is at most η_k, i.e., $|\theta_l^{(k-1)} - \tilde{\theta}^{(k-1)}| \leqslant \eta_k$ for $l = 0, \ldots, 2^{k-1} - 1$ and $k = 1, \ldots, n$. Then, the following lemma can be proven.

Lemma 2. Consider a system of n qubits and an error η_k between any angle of the kth block and its representative such that $\eta_k \leqslant \pi$, with $k = 1, \ldots, n$. Then, the fidelity

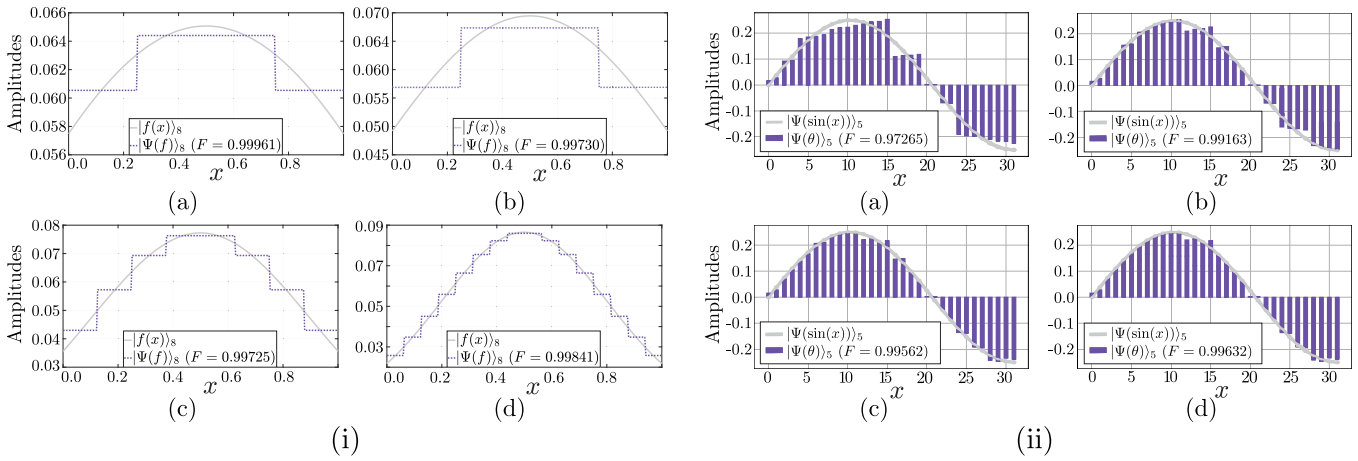

FIG. 3. (i) Simulations of $|f(x)\rangle_8$ and $|\Psi(f)\rangle_8$ resulting of loading the normal distribution according to Theorem 4, $n = 8$, $\epsilon = 0.05$, and different values of σ: (a) $\sigma = 1$, (b) $\sigma = 0.6$, (c) $\sigma = 0.4$, and (d) $\sigma = 0.3$. The numerical results for each experiment are given in Table I. (ii) Simulations of $|\Psi[\sin(x)]\rangle_5$ and $|\Psi(\boldsymbol{\theta})\rangle_5$ resulting of loading the normalized sine function by training the variational *Ansatz*, learning rate $\gamma = 1.5$, $n = 5$, $k_0 = 2$, and different values of $p(k)$: (a) $p(k) = 1$, (b) $p(k) = 2$, (c) $p(k) = 3$, and (d) $p(k) = k$.

between the final states with and without clustering, $F = |\langle 0|^{\otimes n}\mathfrak{U}_n^\dagger\tilde{\mathfrak{U}}_n|0\rangle^{\otimes n}|^2$, satisfies

$$F \geqslant \prod_{k=1}^{n} \cos^2(\eta_k/2). \tag{8}$$

Assuming that we cluster angles of the blocks for $k \geqslant k_0 + 1 > 2$, then

$$F \geqslant \prod_{k=k_0+1}^{n} \cos^2\left(\frac{\eta}{8^{2^k}}\right) \geqslant e^{-\frac{\eta^2}{96}(4^{-k_0}-4^{-n})} := F_{k_0}, \tag{9}$$

since $\cos(x) \geqslant e^{-x^2}$ for $x \lessapprox \pi/2$. Therefore, if an infidelity $\epsilon = 1 - F_{k_0}$ is allowed and the angles of all blocks comprised of more than $k_0(\epsilon)$ qubits are clustered, with k_0 given by Eq. (4), then the fidelity satisfies $F \geqslant 1 - \epsilon$. In the asymptotic limit of $n \to \infty$, we have that $k_0(\epsilon)$ tends to

$$k_0(\epsilon) \to \max\left\{\left\lceil -\frac{1}{2}\log_2\left(-\frac{96}{\eta^2}\log(1-\epsilon)\right)\right\rceil, 2\right\}, \tag{10}$$

which is independent of the system size, n. In Fig. 2(b), we have depicted the values of k_0 at this limit, for $\eta \in [0, 8\pi]$ and $\epsilon \in [0.0001, 0.01]$.

We finally study the implementation cost of the proposed protocol. We only take into account the latter type of gates and ignore single-qubit operations [50]. Using the result of [39,40], which illustrates how a uniformly controlled rotation of k qubits can be implemented with 2^{k-1} CNOTs, the complexity of the circuit described in Theorem 4 is $O(2^{k_0(\epsilon)})$.

A. Normal distribution

We apply the algorithm given by Theorem 4 to a normal distribution with a mean value $\mu = 0.5$ and for different values of the variance σ. We numerically benchmark the fidelity attained by our first protocol using the value of k_0 resulting of Eq. (4) when we assume a maximum infidelity of $\epsilon = 0.05$ and a system of eight qubits. Notice that, for this distribution, $\eta = 2/\sigma^2$. In Fig. 3(i) and Table I, we have depicted the results from the simulations of $|f(x)\rangle_8$ and $|\Psi(f)\rangle_8$, for different values of σ. We appreciate that the condition of $F \geqslant 0.95 = 1 - \epsilon$ is not only satisfied in all cases, but the fidelities obtained are considerably better. Furthermore, the significant reduction in the number of two-qubit gates required to achieve these results is noteworthy. In the worst case, for $\sigma = 0.3$, the quantity of gates needed represents the 12.16% of the original set, while the fidelity of the experiment reaches 0.99841.

B. Generalization: Singular points

In this section we study the generalization of Theorem 4 when the functions are allowed to have singularities on the boundary that grow slower than the size of the grid.

Consider a function $f : [0, 1] \to \mathbb{R}^+$ with

$$\left|\partial_x^2 \log f(x)\right| \underset{x\to 0}{\longrightarrow} \infty. \tag{11}$$

An example is the beta density function, defined as

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \tag{12}$$

with $\alpha, \beta > 0$ and $B(\alpha, \beta)$ the beta function. Then, the second derivative of its logarithm is

$$\partial_x^2 \log f(x) = \frac{1-\alpha}{x^2} + \frac{1-\beta}{(1-x)^2}. \tag{13}$$

TABLE I. Numerical data for the simulations depicted in Fig. 3(i). The values of k_0 have been computed using Eq. (4). In addition, the number of two-qubit gates is given by $2^{k_0} - 1$ and the last column is the percentage between the required gates and the total given by the Grover-Rudolph algorithm, which for $n = 8$ are 255.

σ	η	k_0	Fidelity	No. TQG	%TQG
1.0	2.00	2	0.99961	3	1.18
0.6	5.56	2	0.99730	7	2.75
0.4	12.50	3	0.99725	15	5.88
0.3	22.22	4	0.99841	31	12.16

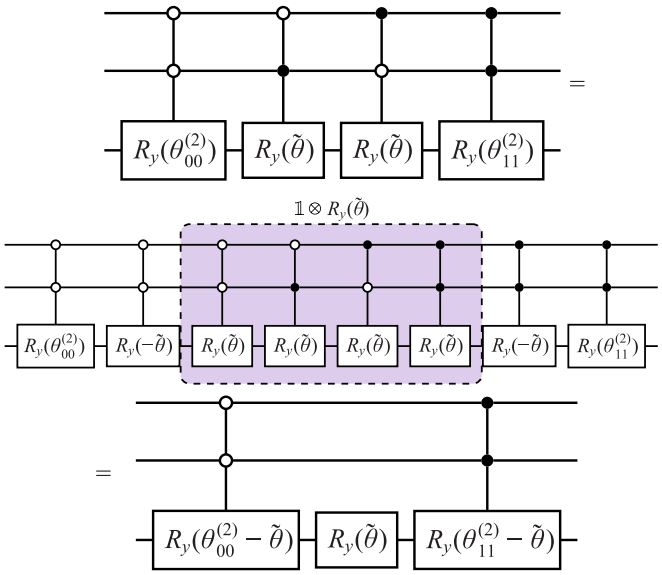

FIG. 4. Example of the reduction of gates for the block of three qubits $F_3^2(\hat{\mathbf{y}}, \boldsymbol{\theta}^{(2)})$, where it is assumed that both $\theta^{(2)}_{01}$ and $\theta^{(2)}_{10}$ can be approximated by $\tilde{\theta}$.

Thus, for $\alpha \neq 1$, we have a singularity at $x = 0$. Also, we see that, if $\beta \neq 1$, there is a singularity at $x = 1$.

In this situation, $|\partial_x^2 \log f(x)|$ cannot be bounded by a finite factor η in the whole interval and, hence, Theorem 4 cannot be applied. However, under certain circumstances, this issue can be solved. Recall that, in Lemma 1, we obtained that the difference between two consecutive angles in a block of k qubits satisfies

$$\left|\theta_{l+1}^{(k-1)} - \theta_l^{(k-1)}\right| \leqslant \frac{\delta_k^2}{4}\left|\partial_x^2 \log f(x)\right|, \tag{14}$$

with $\delta_k = \frac{1}{2^{k-1}}$, $l \in \{0, 1, \ldots, 2^{k-1} - 2\}$, and $x = l\delta_k$. Since the singularity is found in $x = 0$, there exists a $k_{\max}$ from which the maximum of $|\partial_x^2 \log f(x)|$ in $[2^{-k_{\max}+1}, 1]$ is found in $x = 2^{-k_{\max}+1}$. Then,

$$\left|\theta_{l+1}^{(k_{\max}-1)} - \theta_l^{(k_{\max}-1)}\right| \leqslant \frac{1}{4}\frac{\left|\partial_x^2 \log f(x)|_{x=2^{-k_{\max}+1}}\right|}{2^{2k_{\max}-2}}. \tag{15}$$

Next, if for the limit $k \to \infty$ the term $|\partial_x^2 \log f(x)|_{x=2^{-k+1}}|2^{-2k+2} \to 0$, this value is decreasing and there exists a $k^* \geqslant k_{\max}$ from which an $\eta = |\partial_x^2 \log f(x)|_{x=2^{-k^*+1}}|$ can be set so the clustering conditions $\frac{\delta_k}{8}\eta \leqslant \pi \Rightarrow \eta \leqslant 2^{k-1}8\pi$ are satisfied. This k^* acts as the k_0 of Theorem 4 and represents the last block without clustering. Then, for $k > k^*$, the angles corresponding to the interval $[2^{-k+1}, 1]$ can be clustered, following our protocol. The case for a singularity in $x = 1$ is analogous and in Fig. 4 the resulting gates corresponding to the process of clustering the inner angles for a block of three qubits are depicted. Given that $R_y(\tilde{\theta})R_y(-\tilde{\theta}) = \mathbb{1}$, we can complete the identity of the clusterized block by subtracting $\tilde{\theta}$ from the rest of the angles.

In this situation, if we consider that singularities exist in $x = \{0, 1\}$, the total number of two-qubit gates required to load f into a quantum state is

$$\begin{aligned}\text{No. TQGs} &= \sum_{k=1}^{k^*} 2^{k-1} + 2\sum_{k=k^*+1}^{n}(80k - 398)\\ &= 2^{k^*} - 1 + 2(n - k^*)(20k^* + 20n - 179).\end{aligned} \tag{16}$$

See the Appendix A for further details in gate decomposition. Notice that this result is only valid for $k^* > 6$, but since we are interested in the asymptotical behavior of the protocol, it is not an issue. All in all, we obtain that the complexity of this process is exponentially dependent on k^* with an extra polynomial term.

On the other hand, if a singularity is found in (0, 1), we end up with multiple clusters of angles in each block. The reason behind this is that the clusters are formed with contiguous angles. Therefore, the reduction of gates is not significant and the protocol cannot be performed efficiently (polynomial). However, if the representatives of the disjointed clusters are equal, then we can create a single cluster, so the reduction is doable. In this sense, we have numerically observed that, far from the singularities, all angles converge to a value of $\pi/2$, but we have not found an analytical proof yet.

Let us see an example of a function that meets the previous description and analyze the outcome of the protocol. Consider the function

$$f(x) = \frac{1}{N}e^{x^{3/2}} \tag{17}$$

in [0, 1], where N is the normalization factor, and a system of $n = 10$ qubits. The second derivative of its logarithm is

$$\log f(x) = x^{\frac{3}{2}} - \log N \Rightarrow \partial_x \log f(x) = \frac{3}{2}\sqrt{x} \tag{18}$$

$$\Rightarrow \partial_x^2 \log f(x) = \frac{3}{4\sqrt{x}}. \tag{19}$$

Hence $\partial_x^2 \log f(x)$ has a singularity at $x = 0$. First of all, since $\partial_x^2 \log f(x)$ is a monotonic decreasing function, its maximum in the interval $[2^{-k+1}, 1]$ is found in $x = 2^{-k+1}$, for any $k = 1, \ldots, n$. Then, we can set $k_{\max} = 1$. Next, we need to compute the limit of $|\partial_x^2 \log f(x)|_{x=2^{-k+1}}|2^{-2k+2}$:

$$\begin{aligned}&\left|\partial_x^2 \log f(x)\right|_{x=2^{-k+1}}|2^{-2k+2}\\ &= \left.\frac{3}{4\sqrt{x}}\right|_{x=2^{-k+1}} 2^{-2k+2} = \frac{3}{4}2^{\frac{1}{2}(k-1)}2^{-2k+2}\\ &= \frac{3}{4}2^{-\frac{3}{2}k+\frac{3}{2}} \xrightarrow[k\to\infty]{} 0.\end{aligned} \tag{20}$$

Therefore, the difference in the angles is bounded and we can select a $k^* \geqslant 1$ for which the conditions of Theorem 4 are satisfied in $[2^{-k^*+1}, 1]$. The first condition we need to check is the inequality given by

$$\frac{\delta_k}{8}\eta \leqslant \pi, \tag{21}$$

with $\eta = |\partial_x^2 \log f(x)|_{x=2^{-k+1}}| = \frac{3}{4}2^{\frac{1}{2}(k-1)}$. Then,

$$\tfrac{3}{32}2^{\frac{1}{2}(k-1)}2^{-k+1} = \tfrac{3}{32}2^{\frac{1}{2}(1-k)}. \tag{22}$$

Since this term is decreasing, its maximum is found when $k = 1$, with a value of $\frac{3}{32} < \pi$. Thus this condition is met

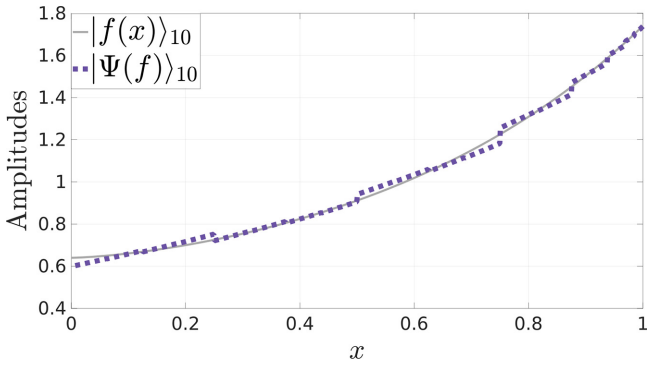

FIG. 5. Simulation of $|f(x)\rangle_{10}$ and $|\Psi(f)\rangle_{10}$ for the function defined in Eq. (17), $n = 10$, $\epsilon = 0.01$, and the clustering process performed for $k = 3, \ldots, n$.

for any $k = 1, \ldots, n$, so we can select $k^* = 1$. Additionally, if we compute the value of k_0 with $\epsilon = 0.01$ and $\eta = |\partial_x^2 \log f(x)|_{x=2^{-k^*+1}}| = 0.75$ using Eq. (4), we obtain $k_0 = 2$. Then, the last block to remain unclustered must be the maximum between k_0 and k^*, which in this case is 2.

Now, in Fig. 5, we have depicted the result of the experiment of the considered function for a system of $n = 10$ qubits and the clustering starting with the three-qubit block, following the protocol described in this section. With a fidelity of 0.99975, larger than the one required, we have that the final state $|\Psi(f)\rangle_{10}$ successfully captures the features of f with a reduction of the complexity of two-qubit gates from $O(2^{10})$ to $O(2^2)$.

C. Analysis of resilience to noise in NISQ era

In this subsection, we present a theoretical and numerical analysis of how experimental errors affect different clustering levels in our first protocol and their impact on the final fidelity. The crucial point here is that, when digital accuracy increases, the number of gates requested grows exponentially. As the introduction of these gates implies a growing experimental error in NISQ quantum processors, we observe a trade-off, see Fig. 6, between the clusterization error and the experimental error, quite similar to the trade-off observed in digital quantum simulations between the number of Trotter steps and the experimental error in Refs. [51,52]. This balance is crucial for algorithms in noisy quantum processors. Reproducing a similar reasoning as the one in the aforementioned reference and references thereof, we first propose an approximated model of how the experimental error combined with the clustering error of our algorithm scales as a function of the number of nonparallel two-qubit gates. Then, we perform some numerical simulations introducing multiple realistic noises in our algorithm to support our theoretical predictions.

In order to establish a theoretical framework to understand the behavior of our system when clusterization and experimental errors are considered, we make the assumption that the main source of experimental noise comes from the two-qubit gates, while ignoring the noise arising from single-qubit rotations. Additionally, we consider that the application of each of these gates onto a quantum state, denoted as ρ, is modeled in

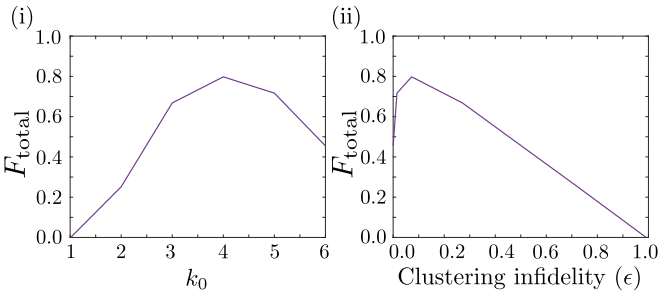

FIG. 6. Total fidelity, F_{total}, in terms of the clustering infidelity assumed in the protocol according to the expression $\epsilon(k_0) \sim 1 - e^{-\eta^2/24(4^{-k_0}-4^{-n})}$, where $\eta = 22.22$, $k_0 = 1, \ldots, 6$, and $n = 6$. As we can appreciate, the expected fidelity of our protocol is above the fidelity resulting of implementing the protocol without clustering in noisy devices, which dramatically tends to zero. Actually, the fidelity reaches a maximum for $k_0 = 4$, which corresponds to a clustering error $\epsilon_0 = 0.0726$.

the following form:

$$\rho \to (1-\xi)U^\dagger \rho U + \xi \tilde{U}^\dagger \rho \tilde{U}, \tag{23}$$

where U corresponds to the desired dynamics while the second term introduces a certain Taylor expansion of the deviation from the exact evolution, with $\xi \ll 1$. After m nonparallel gates, ignoring quadratic terms in ξ, the infidelity of the state has approximately evolved according to

$$1 - F_{\text{exp}} \sim m\, \xi\, \|\tilde{U}\| \tag{24}$$

as the leading term. Note that the argument could be extended by replacing $\tilde{U} \to T(\cdot)$, an arbitrary quantum channel, but the calculation is more complicated to produce a similar argument. In a rough approximation, for a certain value of the clustering index k_0, the total fidelity of the protocol run on a NISQ device has a contribution coming from this experimental noise together with the digital infidelity due to the clustering procedure. We can approximate this quantity by

$$F_{\text{total}} \sim e^{-\eta^2/24(4^{-k_0}-4^{-n})} - 2\xi\|\tilde{U}\|(-1+2^{k_0}), \tag{25}$$

which, in terms of the clustering error ϵ and by using that $1 - \epsilon \sim e^{-\eta^2/24(4^{-k_0}-4^{-n})}$ and hence $2^{k_0} - 1 \sim (4^{-n} - \frac{96}{\eta^2}\log(1-\epsilon))^{-\frac{1}{2}} - 1$, can be expressed as

$$F_{\text{total}} \sim 1 - \epsilon - 2\xi\|\tilde{U}\|\left[\left(4^{-n} - \frac{96}{\eta^2}\log(1-\epsilon)\right)^{-\frac{1}{2}} - 1\right]. \tag{26}$$

From the equation above, we can find the value of clustering error ϵ_0 for which F_{total} achieves its maximum by deriving the function. This condition holds as

$$4\,(1-\epsilon_0)^2\,[4^{-n} - \mu\,\log(1-\epsilon_0)]^3 = \alpha^2\mu^2, \tag{27}$$

with $\alpha = 2\xi\|\tilde{U}\|$ and $\mu = 96/\eta^2$, which is a transcendent equality, so it cannot be analytically solved. However, for the range of parameters in which we are interested and for sufficiently small, we can approximate $\epsilon_0 \sim \frac{\alpha\, 2^{n-\frac{3}{2}}}{\sqrt{3}}$.

In order to provide a numerical example for this expression, we analyze a normal distribution with $\sigma = 0.3$ encoded

TABLE II. Noise parameters' description and their value. We have estimated the numerical values from the calibration data provided for the IBM device "ibm_jakarta."

Parameter	Description	Value
SQG time	Single qubit gate time (ns)	35
CX time	CX gate time (ns)	540
rD	Deviation ratio for the single qubit gates	2457×10^{-4}
Pbf	Bit-flip error during the rz gate	2457×10^{-4}
CNOTerror	Deviation ratio for CX gate	8328×10^{-3}
pmeas	Readout error	223×10^{-1}
pth	Thermal population of the ground state	0.01
T1	Decoherence time (us)	114.84
T2	Dephasing time (us)	38.65

in $n = 6$ qubits. For the experimental error, we consider $\alpha = 2\xi||\tilde{U}|| = 0.0087$. We depict the results of this analysis in Fig. 1. If we compare results from this analysis, we can see that the maximum fidelity is achieved for a value of $\epsilon_0 = 0.0726$, while the predicted value reads $\epsilon_0 = 0.1131$.

Once the theoretical framework has been established, let us carry out some numerical simulations analyzing the robustness (in terms of fidelity) against different noises. The main objective is to support the aforementioned analytical findings. The guidelines of the numerical experiment are described as follows. The circuit is transpiled to a native set of gates given by *CNOT, Id, Rz* (θ)*, X* and *Sx*. The noise quantum channels considered are (i) bit flip (Pbf), (ii) amplitude damping (T1), (iii) dephasing (T2), (iv) gate errors (rD, CNOT error), and (v) measurement error (pmeas).

For a realistic scenario, we have taken the value of these errors from the IBM Jakarta quantum processor, which are summarized in Table II. Considering this setup, we have studied again the normal distribution with $\sigma = 0.3$ encoded in $n = 6$ qubits, focusing on the fidelity and l_2 norm (normalized with the system size such that it converges to the L_2 norm in the continuous limit) with respect to the exact discretized state. The numerical results are depicted in Fig. 7, which also shows a trade-off between the clustering and the experimental error. More explicitly, the maximal fidelity, respectively the minimal error measure with the norm L_2, is achieved for a clusterization level $k_0 = 2$, which reproduces the structure showing a trade-off between errors predicted by our theoretical model. However, the maximum is reached for a smaller value of k_0, compared to the predictions, as the theoretical model is a first order simplification which becomes not that accurate when the presence of more kinds of noises is assumed. This means that our idea based on clustering can be implemented with shallow but not trivial circuits and presumably offers a robust performance for the NISQ era.

Consequently, although our protocol introduces a controllable error, it significantly reduces the depth required for the computation as well, resulting in a more balanced and reliable outcome. In the presence of experimental noise, our algorithm achieves a good balance between fidelity (experimental + clustering errors) and feasibility (realistic depth), which is crucial for this stage of quantum computing, where computers are characterized by high error rates and limited coherence times.

IV. VARIATIONAL QUANTUM CIRCUIT

A. Ansatz

Based on the previous protocol, we propose a variational *Ansatz* for loading functions beyond the conditions required by Theorem 1. We consider $(z + s + 1)p(k)$ hyperparameters for each k-qubit block satisfying $k > k_0$, where z and s are respectively the number of zeros and singular points in the function, $p(k)$ is the polynomial number in k denoting the number of hyperparameters allowed per singular point/zero for the kth block, and $k_0 = \max\{\ k \mid z + s + 1 \geqslant 2^k\}$. Assuming $p(k) \geqslant 1$, the minimum number of hyperparameters needed to capture the singular behavior of the function is $z + s + 1$. Therefore, for each k-qubit block $F_k^{(k-1)}(\hat{\boldsymbol{y}}, \boldsymbol{\theta})$ comprising more than $z + s + 1$ parameters, we cluster the angles which do not correspond to the position of zeros or singularities. In this form, this proposal establishes an intuitive correlation between the number of hyperparameters $p(k)$ and the expressability of the circuit to capture the details of the target functions in the most relevant points. By using

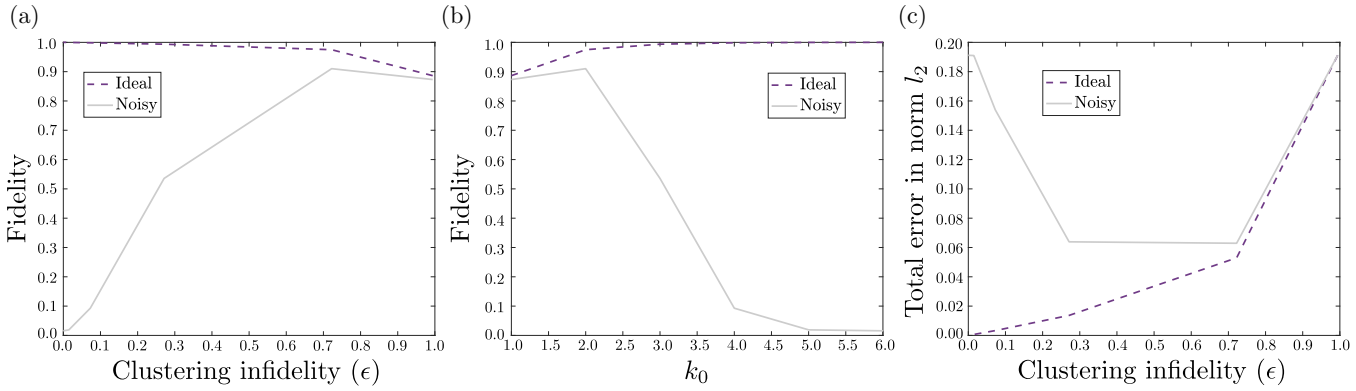

FIG. 7. (a), (b) Total fidelity, F_{total}, in terms of the clustering infidelity assumed and k_0 in the protocol according to the expression $\epsilon(k_0) \sim e^{-\eta^2/24(4^{-k_0}-4^{-n})}$, where $\eta = 22.22$, $k_0 = 1, \ldots, 6$, and $n = 6$ for the ideal and noisy cases. (c) l_2 normalized error in terms of the clustering in the same conditions. The maximal fidelity, respectively the minimal error measure with the l_2 norm, is achieved for a clusterization level $k_0 = 2$, which corresponds to $\epsilon_0 = 0.7222$.

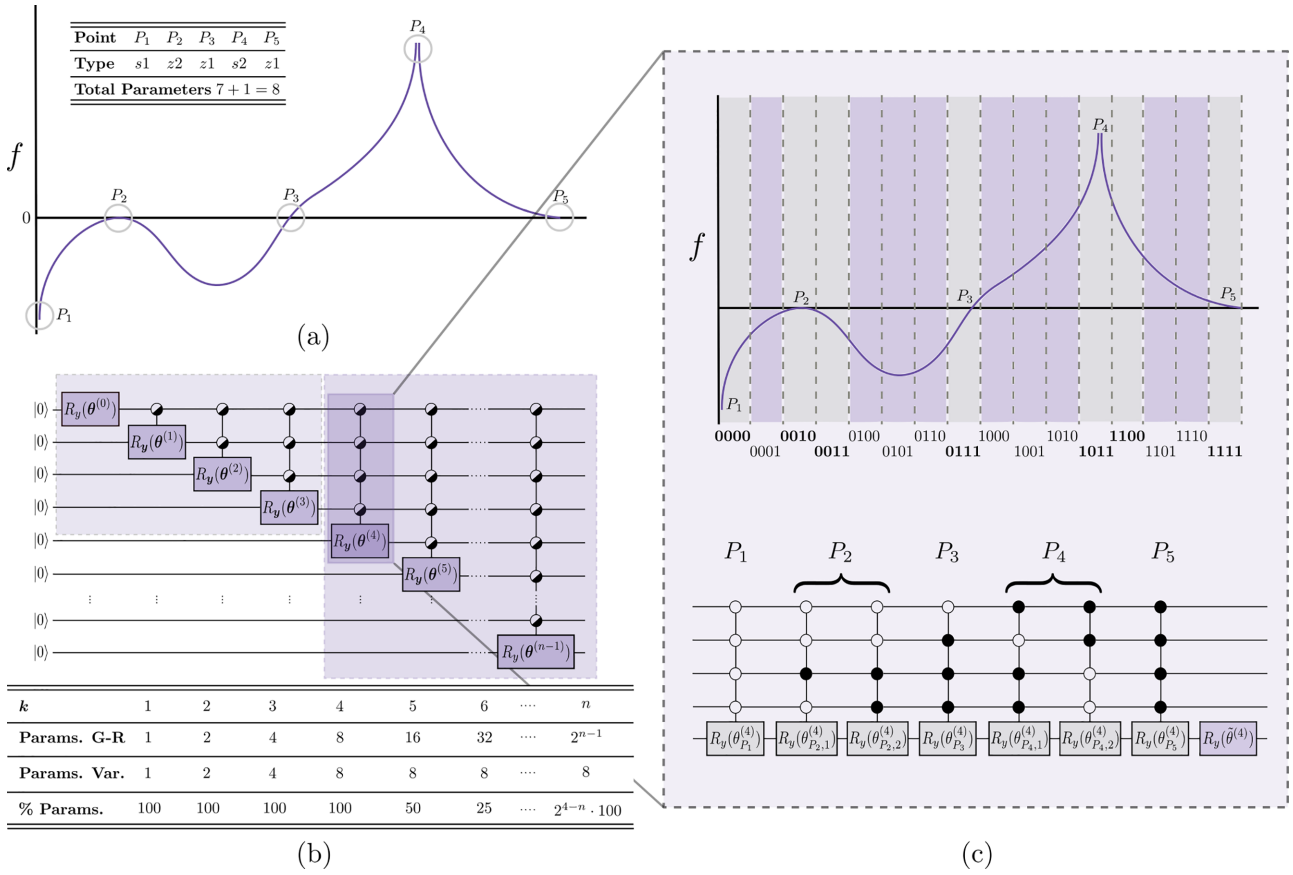

FIG. 8. Illustration of the proposed variational circuit for a general function. (a) Target function. The proposed landscape generalizes the possible cases of special points that lead to singularities in the second logarithmic derivative. The special points are classified in the legend of the graph, where z denotes the zeros, s the singularities, and the index denotes whether the slope in the contiguous points has the same sign, index $= 1$, or not, index $= 2$. This is translated to a minimum amount of seven necessary parameters required to capture the local behavior of the function in these points, one per index $= 1$ and 2 per index $= 2$. We add an additional parameter which represents the clusterization. (b) Variational circuit. The blocks which have a number of angles lower than or equal to the necessary parameters will remain unclustered. From the first block which has more angles, we proceed to reduce the number of angles to the number of parameters by clustering the ones that do not correspond to the special points. This leads to a significant reduction in the number of gates. (c) Zoom in of the $F_5^4(y,\boldsymbol{\theta})$ block clusterization. As we only need eight parameters, the 16 angles of this block are reduced to the half. We keep one local rotation for the clustering parameter and seven multicontrolled gates corresponding to the intervals where the singularity is located.

the decomposition of the multicontrolled rotations [39,40], the number of two-qubit gates which must be included in the variational circuit sums up to

$$\begin{aligned}\text{No. TQGs} &= \sum_{k=1}^{k_0} 2^{k-1} + (s+z)\sum_{k=k_0+1}^{n} p(k)(80k-398)\\ &= 2^{k_0} - 1 + r_{k_0+1}(n, s+z),\end{aligned} \tag{28}$$

with m the number of zeros/singularities and $r_{k_0+1}(n, s+z)$ growing polynomially with the system size n. A remarkable advantage of this *Ansatz* is the training procedure, since the Grover-Rudolph algorithm provides a suitable set of initial training angles which considerably enhances the convergence of the protocol with respect to a random initialization. This fact, together with a scaling in the number of hyperparameters $p(k)$ substantially slower than the system size, allows us to avoid the training procedure to get stuck into both local minima and barren plateaus. In Fig. 8, we illustrate how to select the hyperparameters of the variational circuit for a general function containing multiple zeros and singular points.

B. Training process

We proceed to illustrate the training method for the variational circuit proposed in this article. This process consists of iterative steps in which a loss function that measures how far the outcome is from the desired state is recursively minimized to obtain the optimal parameters.

Consider a function f and n qubits. The desired quantum state is

$$|\Psi(f)\rangle_n = \sum_{l=0}^{2^n-1} f_l |l\rangle, \tag{29}$$

where the f_l terms are discrete approximations to the objective function, as presented in Definition 1. Now, given a set of angles $\boldsymbol{\theta}$, our *Ansatz* $U(\boldsymbol{\theta})$ returns the state

$$|\Psi(\boldsymbol{\theta})\rangle_n = U(\boldsymbol{\theta})|0\rangle, \tag{30}$$

where $|0\rangle \equiv |0\rangle^{\otimes n}$, and the components of the obtained state are products of sines and cosines.

Let us now introduce the mean squared error loss function, defined as

$$L(n, f, \boldsymbol{\theta}) = \frac{1}{2^n}\sum_{l=0}^{2^n-1}[f_l - \Psi_l(\boldsymbol{\theta})]^2 = \frac{1}{2^n}\sum_{l=0}^{2^n-1}\left[f_l^2 + \Psi_l(\boldsymbol{\theta})^2 - 2f_l\Psi_l(\boldsymbol{\theta})\right], \tag{31}$$

where $\Psi_l(\boldsymbol{\theta}) := \langle l|\Psi(\boldsymbol{\theta})\rangle_n$.

This process aims to find the optimal parameters $\boldsymbol{\theta}$ for which $|\Psi(\boldsymbol{\theta})\rangle_n$ approximates $|\Psi(f)\rangle_n$, which is equivalent to minimizing the loss function. Here, we use the gradient descent method [53] to do so.

Then, given any angle $\theta_l^{(k-1)}$, with $k \in \{1, \ldots, n\}$ and $l \in \{0, 2^{k-1} - 1\}$, its value gets updated after each training step in the following way:

$$\theta_l^{(k-1)} = \theta_l^{(k-1)} - \gamma\frac{\partial L(n, f, \boldsymbol{\theta})}{\partial\theta_l^{(k-1)}}, \tag{32}$$

where γ is the learning rate. Let us now compute the expression of the derivative:

$$\frac{\partial L(n, f, \boldsymbol{\theta})}{\partial\theta_j^{(k-1)}} = \frac{1}{2^n}\sum_{l=0}^{2^n-1}2[\Psi_l(\boldsymbol{\theta}) - f_l]\frac{\partial\Psi_l(\boldsymbol{\theta})}{\partial\theta_j^{(k-1)}}. \tag{33}$$

Since $\Psi_l(\boldsymbol{\theta})$ is a product of sines and cosines, its partial derivative with respect to a given angle is

$$\frac{\partial\Psi_l(\boldsymbol{\theta})}{\partial\theta_j^{(k-1)}} = \begin{cases} -\frac{1}{2}\frac{\sin\left(\theta_j^{(k-1)}/2\right)}{\cos\left(\theta_j^{(k-1)}/2\right)}\Psi_l(\boldsymbol{\theta}) & \text{if } j2^{n-k+1} \leqslant l < (j+1)2^{n-k}, \\ \frac{1}{2}\frac{\cos\left(\theta_j^{(k-1)}/2\right)}{\sin\left(\theta_j^{(k-1)}/2\right)}\Psi_l(\boldsymbol{\theta}) & \text{if } (j+1)2^{n-k} \leqslant l < (j+1)2^{n-k+1}, \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

Finally, different stopping criteria exist to put an end to the training process. A simple example is to fix a number of training steps. Another criteria, which is the one we consider for our numerical simulations, is to set a tolerance for the loss function. Therefore, when the difference between the cost function of two consecutive steps is less than the given tolerance, the process is considered satisfactory.

As a summary, in Algorithm 1 we have depicted the pseudocode corresponding to the training process.

C. Examples

Let us now illustrate the behavior of the variational circuit different density functions.

1. Sine function

As a first example, we have tested the variational circuit with the normalized sine function $\sin x$ in the domain $[0, \frac{3}{2}\pi]$. This example contains zeros in $x = 0$ and $x = \pi$ and clearly does not satisfy the conditions in Theorem 4 or a possible generalization, since it is not even positive. In order to train the parametrized quantum circuit, we use the mean squared error loss function and the gradient descent as optimization method, with a learning rate $\gamma = 1.5$. In Fig. 3(ii), we illustrate the arising loaded states of our trained circuits for different values of $p(k) = 1, 2, 3$ and k (the last case means the introduction of k hyperparameters per zero of the function) and a system comprising $n = 5$ qubits, with $k_0 = 2$. For the studied cases, our trained *Ansatz* is able to load the sine function with a fidelity larger than 0.97. Additionally, in Fig. 9, we show the resulting fidelity of the different combinations of the number of qubits and $p(k)$, for $n \in \{5, \ldots, 10\}$ and $k_0 = 3$.

Algorithm 1. Algorithm for the training process.

Require: n qubits, normalized function f, objective state $|\Psi(f)\rangle_n$, circuit $U(\boldsymbol{\theta})$ with hyperparameters, stopping criteria, and learning rate γ.

1: Initialize parameters $\boldsymbol{\theta}$
2: **while** stopping criteria is not met **do**
3: $\quad|\Psi(\boldsymbol{\theta})\rangle_n \leftarrow U(\boldsymbol{\theta})|0\rangle$
4: $\quad$Compute loss function $L(n, f, \boldsymbol{\theta})$
5: $\quad\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \gamma\frac{\partial L(n,f,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$
6: **end while**

2. Black–Scholes distribution

As a second example, let us now apply the proposed variational circuit to the Black-Scholes distribution, which is given

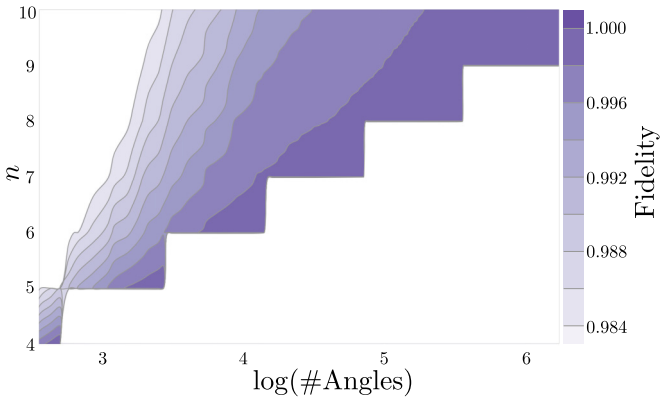

FIG. 9. Fidelity of loading $f(x) = \sin x$, $x \in [0, 3\pi/2]$, for different combinations of $n \in \{5, \ldots, 10\}$, learning rate $\gamma = 1.5$ and $p(k)$, with $k_0 = 3$. The axis corresponding to the total number of angles is in logarithmic scale.

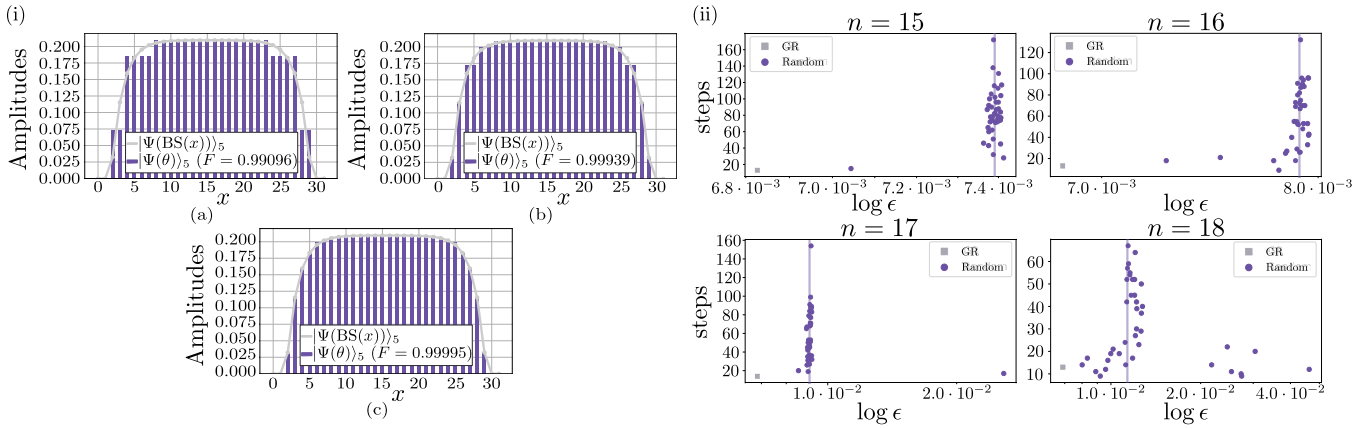

FIG. 10. (i) Simulations of $|\Psi[\mathrm{BS}(x)]\rangle_5$ and $|\Psi(\boldsymbol{\theta})\rangle_5$ for the Black-Scholes distribution, $n = 5$, $k_0 = 2$, and different values of $p(k)$: (a) $p(k) = 1$, (b) $p(k) = 2$, and (c) $p(k) = 3$. Parameters of the distribution: $K = 45$ and $c = 3$. Training parameters: $k_0 = 2$, learning rate $\gamma = 1.5$, and tolerance of 10^{-9}. We can appreciate that in all cases the circuit is able to capture the main features of the target state; consequently all fidelities are close to 1. (ii) Results of training the variational *Ansätze* for a different number of qubits ($n = 15, 16, 17, 18$) to load the Black-Scholes distribution. We compare the training results (numbers of step and final infidelity) of 40 randomly initialized parameters versus the initialization with Grover-Rudolph method angles. These randomly initialized parameters are generated by drawing samples from a uniform distribution between the values of zero and π. As we can appreciate the initialization with the Grover-Rudolph method always achieves the best performance, while most random initializations get stuck around a fidelity value (vertical lines). Training parameters: $k_0 = 2$, $p(k) = 1$, learning rate $\gamma = 1.5$, and tolerance of 10^{-9}.

by

$$\mathrm{BS}(x) = \begin{cases} K - e^{-x}/s & \text{if } -\log(Ks) \leqslant x < 0, \\ K - e^{x}/s & \text{if } 0 < x \leqslant \log(Ks), \end{cases} \tag{35}$$

with $s = Kc$, where K and c are the parameters. The zeros are found in $x = \pm\log(Ks)$.

We benchmark the performance of our *Ansatz* for loading the Black-Scholes distribution, Eq. (35), into a five-qubit system. We choose the parameters of the distribution $K = 45$ and $c = 3$ and, for the training parameters, we consider $k_0 = 2$, a learning rate of $\gamma = 1.5$, and a tolerance of 10^{-9}. In Fig. 10(i), we depict the results from the numerical simulations of the discretized state $|\Psi[\mathrm{BS}(x)]\rangle_5$ and the trained *Ansatz* $|\Psi(\boldsymbol{\theta})\rangle_5$, for different values of $p(k)$ and $n = 5$ qubits. We can appreciate that in all cases the circuit is able to capture the main features of the target state; consequently all fidelities are close to 1. We highlight that for $p(k) = 3$ the resulting state and the objective one are almost equal, $\epsilon = 5\mathrm{e}{-5}$.

We also provide an analysis for larger systems. In Table III, we depict the numerical results for $|\Psi[\mathrm{BS}(x)]\rangle_{12}$ and $|\Psi(\boldsymbol{\theta})\rangle_{12}$ in a 12-qubit system, with $k_0 = 2$, and different values of $p(k)$. We are able to obtain large fidelities with only a few percentage of the original angles, which implies that, with only a small portion of the initial parameter space, the circuit can approximately simulate the target state efficiently.

Finally, in Fig. 10(ii) and Table IV we analyze the performance of initializing the value of the parameters with the angles provided by the Grover-Rudolph method for different sizes of the system. We present the training results of the *Ansatz* initialized with several random angles sets and with the angles provided by the Grover-Rudolph method. We analyze the performance in terms of infidelity and number of steps of the training. We can observe how this initialization reduces drastically the number of steps, as well as enables us to achieve the largest fidelities.

V. CONCLUSIONS

In this article, we have considered the problem of loading real valued functions into a quantum computer, which is a major bottleneck for solving partial derivatives equations [11–14], computing Monte Carlo integrations [9,10,36], and quantum field theory [37,38] and quantum machine learning [15,17–20]. First, inspired by the Grover-Rudolph algorithm [26] without ancillas, we have analytically proven that the complexity for implementing our method on an n-qubit system scales as $O(2^{k_0(\epsilon)})$, with ϵ the infidelity with respect to the exact state and $k_0(\epsilon)$ asymptotically independent of n. This reduction of two-qubit gates leads to a significant speedup, which allows us to implement quantum protocols involving data embeddings in large qubit systems. Additionally, we have generalized this method for functions containing a certain type of singularities, obtaining promising results for

TABLE III. Numerical data for the simulations of $|\Psi[\mathrm{BS}(x)]\rangle_{12}$ and $|\Psi(\boldsymbol{\theta})\rangle_{12}$, $n = 12$, and different values of $p(k)$, with $k_0 = 2$. The value No. Angles is the number of independent angles necessary to generate $|\Psi(\boldsymbol{\theta})\rangle_{12}$ and %Angles is in comparison with the full circuit. We are able to obtain large fidelities with only a few percentage of the original angles, which implies that, with only a small portion of the initial parameter space, the circuit can approximately simulate the target state efficiently.

$p(k)$	Fidelity	No. Angles	%Angles
1	0.99303	33	0.806
2	0.99838	52	1.269
3	0.99890	70	1.709
k	0.99913	142	3.468

TABLE IV. Numerical results of training the variational *Ansätze* for a different number of qubits ($n = 15, 16, 17, 18$) to load the Black-Scholes distribution. We compare the training results (numbers of step and final infidelity) of 40 randomly initialized parameters versus the initialization with Grover-Rudolph method angles. These randomly initialized parameters are generated by drawing samples from a uniform distribution between the values of zero and π. Training parameters: $k_0 = 2$, $p(k) = 1$, learning rate $\gamma = 1.5$, and tolerance of 10^{-9}.

n	Fidelity (GR)	Steps (GR)	Avg. fidelity (random)	Avg. steps (random)	Max fidelity (random)	Dif. fid. GR vs avg. random
15	0.99317	13	0.99261	82	0.99295	0.00056
16	0.99316	13	0.99211	59	0.99271	0.00105
17	0.99314	14	0.99052	57	0.99145	0.00262
18	0.99309	13	0.98566	31	0.99199	0.00743

density functions with singularities that satisfy the expanded theorem conditions. Furthermore, we have proposed a variational *Ansatz* inspired by our previous protocol. We have observed that it can efficiently and accurately load functions with zeros and singularities. Our proposed *Ansatz* is tailored to the landscape of the function, providing an intuitive correlation between hyperparameters and expressability. Moreover, our previous protocol allows us to define a suitable initial training angle set, which considerably improves the training process, avoiding barren plateaus and local minima. As a future work, tensor networks could be used to prove the quasioptimality in the minimal number of hyperparameters in the variational *Ansatz*.

ACKNOWLEDGMENTS

We thank J. J. García-Ripoll for the useful discussions regarding the quasi optimality of the variational Ansatz, T. Nguyen for the useful discussions regarding the Grover Rudolph algorithm, and M. Garcia-de-Andoin for the discussion on the noise impact. The authors acknowledge financial support from OpenSuperQ+100 (Grant No. 101113946) of the EU Flagship on Quantum Technologies, as well as from the EU FET-Open project EPIQUS (Grant No. 899368), also from Project Grant No. PID2021-125823NA-I00 and Spanish Ramón y Cajal Grant No. RYC-2020-030503-I funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe" and "ERDF Invest in your Future", Basque Government through Grant No. IT1470-22 and the IKUR Strategy under the collaboration agreement between Ikerbasque Foundation and BCAM on behalf of the Department of Education of the Basque Government, as well as from and UPV/EHU Ph.D. Grant No. PIF20/276.

APPENDIX A: GATES DECOMPOSITION AND COMPLEXITY

1. Uniformly controlled rotation

Let us consider $n-1$ control qubits, a target nth qubit, an axis of rotation $\boldsymbol{u}$, and a vector of 2^{n-1} angles $\boldsymbol{\theta}$. Then, a uniformly controlled rotation $F_n^{n-1}(\boldsymbol{u}, \boldsymbol{\theta})$, depicted in Fig. 11, is a sequence of multicontrolled rotations $\wedge_{n-1}[R_{\boldsymbol{u}}(\theta_i)]$ comprising the 2^{n-1} combinations of control bits.

Analytically this gate can be expressed as

$$F_n^{n-1}(\boldsymbol{u}, \boldsymbol{\theta}) = \prod_{i=0}^{2^{n-1}-1} [|i\rangle\langle i| \otimes R_{\boldsymbol{u}}(\theta_i) + (\mathbb{1}_{n-1\times n-1} - |i\rangle\langle i|) \otimes \mathbb{1}] \tag{A1}$$

and, since $\sum_{i=0}^{2^n-1} |i\rangle\langle i| = \mathbb{1}_{n\times n}$ and $\langle i|j\rangle = \delta_{ij}$, we can simplify it to

$$F_n^{n-1}(\boldsymbol{u}, \boldsymbol{\theta}) = \sum_{i=0}^{2^{n-1}-1} |i\rangle\langle i| \otimes R_{\boldsymbol{u}}(\theta_i). \tag{A2}$$

2. Cost of multicontrolled rotations

Let us now analyze the cost of implementing a multicontrolled rotation in terms of single- and two-qubit gates.

Theorem 1. Let us consider an n-qubit system. Then, the multicontrolled rotation $\wedge_{n-1}[R_y(\theta)]$ can be decomposed employing two one-qubit controlled rotations $\wedge_1[R_y(\pm\theta/2)]$ and two $\wedge_{n-2}(X)$, as illustrated in Fig. 12

Proof. First, let us see that $R_y(\theta) = XR_y(-\theta/2)XR_y(\theta/2)$:

$$\begin{aligned} XR_y(-\theta/2)X &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos(\theta/4) & \sin(\theta/4) \\ -\sin(\theta/4) & \cos(\theta/4) \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sin(\theta/4) & \cos(\theta/4) \\ \cos(\theta/4) & -\sin(\theta/4) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta/4) & \sin(\theta/4) \\ -\sin(\theta/4) & \cos(\theta/4) \end{pmatrix} \\ &= R_y(\theta/2). \end{aligned} \tag{A3}$$

Then,

$$XR_y(-\theta/2)XR_y(\theta/2) = R_y(\theta/2)R_y(\theta/2) = R_y(\theta). \tag{A4}$$

The intuition behind this decomposition is that, when all the control states are in $|1\rangle$, we obtain the desired rotation, as illustrated in Eq. (A4). Also, when the controlled rotations are activated but the $\wedge_{n-2}(X)$ is not, we obtain the identity, since $R_y(-\theta/2)R_y(\theta/2) = \mathbb{1}$, as expected. The same occurs when the $\wedge_{n-2}(X)$ gates are activated but the controlled rotations are not. Finally, when none of the controls are triggered, the circuits in both sides are equal to $\mathbb{1}^{\otimes n}$.

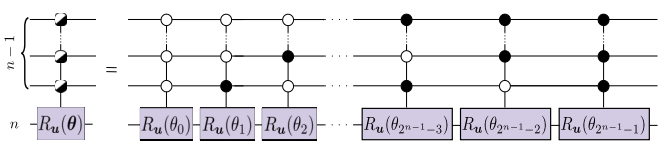

FIG. 11. Circuit corresponding to the uniformly controlled rotation gate $F_n^{n-1}(\boldsymbol{u}, \boldsymbol{\theta})$, with $n-1$ control qubits, a target at the nth qubit, an axis of rotation $\boldsymbol{u}$, and a set of angles $\boldsymbol{\theta}$.

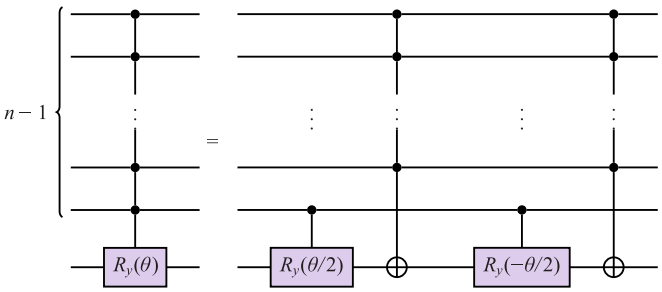

FIG. 12. Circuit showing the decomposition of $\wedge_{n-1}[R_y(\theta)]$ employing two controlled rotations $\wedge_1[R_y(\pm\theta/2)]$ and two $\wedge_{n-2}(X)$ gates.

Now, let us check that the circuits are equivalent. On the left-hand side, we have

$$\wedge_{n-1}[R_y(\theta)] = |1\rangle^{\otimes(n-1)}\langle 1|^{\otimes(n-1)} \otimes R_y(\theta) + (\mathbb{1}^{\otimes(n-1)} - |1\rangle^{\otimes(n-1)}\langle 1|^{\otimes(n-1)}) \otimes \mathbb{1}. \quad \text{(A5)}$$

On the right-hand side

$$\begin{aligned}
&\wedge_{n-2}(X)\wedge_1[R_y(-\theta/2)]\wedge_{n-2}(X)\wedge_1[R_y(\theta/2)] \\
&= [|1\rangle^{\otimes(n-2)}\langle 1|^{\otimes(n-2)} \otimes \mathbb{1} \otimes X && \text{(A6)}\\
&\quad + (\mathbb{1}^{\otimes(n-1)} - |1\rangle^{\otimes(n-2)}\langle 1|^{\otimes(n-2)} \otimes \mathbb{1}) \otimes \mathbb{1}] \\
&\quad \times [\mathbb{1}^{\otimes(n-2)} \otimes (|1\rangle\langle 1| \otimes R_y(-\theta/2) + |0\rangle\langle 0| \otimes \mathbb{1})] \\
&\quad \times [|1\rangle^{\otimes(n-2)}\langle 1|^{\otimes(n-2)} \otimes \mathbb{1} \otimes X && \text{(A7)}\\
&\quad + (\mathbb{1}^{\otimes(n-1)} - |1\rangle^{\otimes(n-2)}\langle 1|^{\otimes(n-2)} \otimes \mathbb{1}) \otimes \mathbb{1}] \\
&\quad \times [\mathbb{1}^{\otimes(n-2)} \otimes (|1\rangle\langle 1| \otimes R_y(\theta/2) + |0\rangle\langle 0| \otimes \mathbb{1})], && \text{(A8)}
\end{aligned}$$

where we have ignored the identities on the left-hand side of the equation for the sake of simplicity. If we expand the previous expression, we obtain the following:

$$\begin{aligned}
&[|1\cdots 1\rangle\langle 1\cdots 1| \otimes XR_y(-\theta/2) \\
&\quad + |1\cdots 10\rangle\langle 1\cdots 10| \otimes X + F(-\theta/2)] && \text{(A9)}\\
&\quad \times [|1\cdots 1\rangle\langle 1\cdots 1| \otimes XR_y(\theta/2) && \text{(A10)}\\
&\quad + |1\cdots 10\rangle\langle 1\cdots 10| \otimes X + F(\theta/2)], && \text{(A11)}
\end{aligned}$$

where we have defined the projector

$$\begin{aligned}
F(\alpha) \equiv\ & [(\mathbb{1}^{\otimes(n-1)} - |1\cdots 1\rangle\langle 1\cdots 1| \\
& - |1\cdots 10\rangle\langle 1\cdots 10|) \otimes \mathbb{1}] \\
& \times [\mathbb{1}^{\otimes(n-2)} \otimes (|1\rangle\langle 1| \otimes R_y(\alpha) + |0\rangle\langle 0| \otimes \mathbb{1})] \\
=\ & \mathbb{1}^{\otimes(n-2)} \otimes |1\rangle\langle 1| \otimes R_y(\alpha) + \mathbb{1}^{\otimes(n-2)} \otimes |0\rangle\langle 0| \otimes \mathbb{1} \\
& - |1\cdots 10\rangle\langle 1\cdots 10| \otimes \mathbb{1} - |1\cdots 1\rangle\langle 1\cdots 1| \otimes R_y(\alpha). \quad \text{(A12)}
\end{aligned}$$

It is straightforward to see that both terms $|1\cdots 1\rangle\langle 1\cdots 1| \otimes XR_y(\alpha)\cdot F(\alpha)$ and $|1\cdots 10\rangle\langle 1\cdots 10| \otimes X \cdot F(\alpha)$ are equal to zero, due to the projector $F(\alpha)$. Then, we have

$$\begin{aligned}
&\wedge_{n-2}(X)\wedge_1[R_y(-\theta/2)]\wedge_{n-2}(X)\wedge_1[R_y(\theta/2)] \\
&= |1\cdots 1\rangle\langle 1\cdots 1| \otimes R_y(\theta) + |1\cdots 10\rangle\langle 1\cdots 10| \otimes \mathbb{1} && \text{(A13)}\\
&\quad + F(-\theta/2)F(\theta/2), && \text{(A14)}
\end{aligned}$$

where we have used $XR_y(-\theta/2)XR_y(\theta/2) = R_y(\theta)$ and $XX = \mathbb{1}$. Let us now compute the last term:

$$\begin{aligned}
&F(-\theta/2)F(\theta/2) \\
&= \mathbb{1}^{\otimes(n-2)} \otimes |1\rangle\langle 1| \otimes R_y(-\theta/2)R_y(\theta/2) && \text{(A15)}\\
&\quad - |1\cdots 1\rangle\langle 1\cdots 1| \otimes R_y(-\theta/2)R_y(\theta/2) \\
&\quad + \mathbb{1}^{\otimes(n-2)} \otimes |1\rangle\langle 1| \otimes \mathbb{1} - |1\cdots 10\rangle\langle 1\cdots 10| \otimes \mathbb{1} && \text{(A16)}\\
&\quad - |1\cdots 10\rangle\langle 1\cdots 10| \otimes \mathbb{1} + |1\cdots 10\rangle\langle 1\cdots 10| \otimes \mathbb{1} \\
&\quad - |1\cdots 1\rangle\langle 1\cdots 1| \otimes R_y(-\theta/2)R_y(\theta/2) \\
&\quad + |1\cdots 1\rangle\langle 1\cdots 1| \otimes R_y(-\theta/2)R_y(\theta/2) \\
&= \left(\mathbb{1}^{\otimes(n-1)} - |1\cdots 1\rangle\langle 1\cdots 1| - |1\cdots 10\rangle\langle 1\cdots 10|\right) \otimes \mathbb{1}, && \text{(A17)}
\end{aligned}$$

where we have used $R_y(-\theta/2)R_y(\theta/2) = \mathbb{1}$ and

$$\mathbb{1}^{\otimes(n-2)} \otimes |1\rangle\langle 1| \otimes \mathbb{1} + \mathbb{1}^{\otimes(n-2)} \otimes |0\rangle\langle 0| \otimes \mathbb{1} = \mathbb{1}^{\otimes(n-1)} \otimes \mathbb{1}. \quad \text{(A18)}$$

Finally,

$$\begin{aligned}
&\wedge_{n-2}(X)\wedge_1[R_y(-\theta/2)]\wedge_{n-2}(X)\wedge_1[R_y(\theta/2)] \\
&= |1\cdots 1\rangle\langle 1\cdots 1| \otimes R_y(\theta) + |1\cdots 10\rangle\langle 1\cdots 10| \otimes \mathbb{1} \\
&\quad + (\mathbb{1}^{\otimes(n-1)} - |1\cdots 1\rangle\langle 1\cdots 1| - |1\cdots 10\rangle\langle 1\cdots 10|) \otimes \mathbb{1} \\
&= |1\cdots 1\rangle\langle 1\cdots 1| \otimes R_y(\theta) \\
&\quad + (\mathbb{1}^{\otimes(n-1)} - |1\cdots 1\rangle\langle 1\cdots 1|) \otimes \mathbb{1} \\
&= \wedge_{n-1}[R_y(\theta)], && \text{(A19)}
\end{aligned}$$

as we wanted to prove. ■

Corollary 1. The order of complexity of implementing the $\wedge_{n-1}[R_y(\theta)]$ gate is linear in n, for $n > 6$. In fact, $\wedge_{n-1}[R_y(\theta)]$ can be decomposed with $80n - 398$ two-qubit gates.

proof. Following Theorem 1, we know that the multicontrolled rotation $\wedge_{n-1}[R_y(\theta)]$ can be decomposed into two one-qubit controlled rotations, $\wedge_1[R_y(\theta/2)]$ and $\wedge_1[R_y(-\theta/2)]$, and two $\wedge_{n-2}(X)$. Therefore, we need to study the number of two-qubit gates necessary to implement the last two gates $\wedge_{n-2}(X)$.

Corollary 7.4 from [54] states that, for $n > 6$, $\wedge_{n-2}(X)$ can be realized with $8(n-5)$ $\wedge_2(X)$. Then, we reduce the problem to obtain the cost of the Toffoli gate, $\wedge_2(X)$.

Lemma 6.1 from the previous reference [54] states that any one-qubit unitary, $\wedge_2(U)$, can be implemented with five two-qubit gates comprising two CNOTs, two $\wedge_1(V)$, and one $\wedge_1(V^\dagger)$, with $V^2 = U$. In our case, we have $U = X$. Then, we can choose $V = \frac{1-i}{2}(\mathbb{1} + iX)$. We provide the exact decomposition of the Toffoli gate using the previous description in Fig. 13.

Finally, if $n > 6$, then the number of gates required for decomposing $\wedge_{n-1}[R_y(\theta)]$, considering only two-qubit operators, is

$$\begin{aligned}
\text{No. gates}(\wedge_{n-1}[R_y(\theta)]) &= 2 + 2 \times [8(n-5) \times 5] \\
&= 80n - 398. && \text{(A20)}
\end{aligned}$$

■

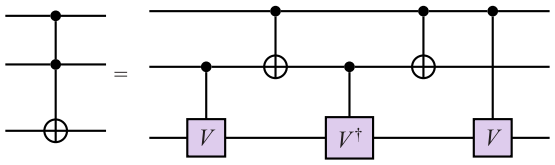

FIG. 13. Circuit of the decomposition of the Toffoli gate, $\wedge_{n-2}(X)$, using five two-qubit controlled operations, where $V^2 = X$.

APPENDIX B: THEOREM PROOFS

1. Relation between the difference in the angles and the target function

Theorem 2. Let f be a continuous function such that $f : [0, 1] \to \mathbb{R}^+$ and consider a block comprising a uniformly controlled rotation of k qubits. Then, the difference between two contiguous angles is bounded by the second derivative of the logarithm of f in the following way:

$$\left|\theta_{l+1}^{(k-1)} - \theta_l^{(k-1)}\right| \leqslant \frac{\delta_k^2}{4} \max_{y' \in [l\delta_k, (l+1)\delta_k]} \left|\left[\partial_y^2 \log f(y)\right]\right|_{y=y'}\Big|, \tag{B1}$$

where $\delta_k = \frac{1}{2^{k-1}}$, $l \in \{1, \ldots, 2^{k-1} - 1\}$.

Proof. The discrete expression of the angles for a block comprised of uniformly controlled rotation with k qubits, first introduced in [55], is given by

$$\theta_l^{(k-1)} = 2 \arccos\left(\sqrt{\frac{\int_{l\delta_k}^{(l+1/2)\delta_k} f(x)dx}{\int_{l\delta_k}^{(l+1)\delta_k} f(x)dx}}\right), \tag{B2}$$

with $\delta_k = \frac{1}{2^{k-1}}$ and $l \in \{1, \ldots, 2^{k-1} - 1\}$. We can define the continuous extension of the previous function as

$$\theta^{(k-1)}(y) := 2 \arccos\left(\sqrt{\frac{\int_y^{y+\delta_k/2} f(x)dx}{\int_y^{y+\delta_k} f(x)dx}}\right), \tag{B3}$$

where the original expression can be recovered replacing $y = l\delta_k$.

Now, given a function g and a displacement μ, we define the numerical derivative of g with respect to μ as

$$\partial_y^{(\mu)} g(y) := \frac{g(y+\mu) - g(y)}{\mu}. \tag{B4}$$

Notice that for $\mu \to 0$, $\partial_y^{(\mu)} g(y) \to \partial_y g(y)$, the usual derivative. Then, for two consecutive angles, we have

$$\left|\theta_{l+1}^{(k-1)} - \theta_l^{(k-1)}\right| = \left|\delta_k \left(\partial_y^{(\delta_k)} \theta^{(k-1)}(y)\right)\right|_{y=l\delta_k}\Big|. \tag{B5}$$

Now, we require a connection between the difference in the angles and the partial derivative of Eq. (B3). The mean value theorem guarantees that the numerical derivative, which corresponds to the slope of the straight line connecting $\theta_l^{(k-1)}$ and $\theta_{l+1}^{(k-1)}$, is constrained by the maximum absolute value of the exact derivative in that interval. Then, we have

$$\left|\theta_{l+1}^{(k-1)} - \theta_l^{(k-1)}\right| \leqslant \delta_k \max_{y' \in [l\delta_k, (l+1)\delta_k]} \left|\left(\partial_y \theta^{(k-1)}(y)\right)\right|_{y=y'}\Big|. \tag{B6}$$

Let us now develop the term of the exact derivative of the function $\theta^{(k-1)}(y)$:

$$\partial_y \theta^{(k-1)}(y) = \frac{[f(y+\delta_k) - f(y)] \int_y^{y+\delta_k/2} f(x)dx - [f(y+\delta_k/2) - f(y)] \int_y^{y+\delta_k} f(x)dx}{\int_y^{y+\delta_k} f(x)dx \sqrt{\int_y^{y+\delta_k/2} f(x)dx \int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}}. \tag{B7}$$

By introducing $[f(y+\delta_k/2) - f(y+\delta_k/2)]$ in the first term and reorganizing the expression, we have

$$\begin{aligned}
\partial_y \theta^{(k-1)}(y) &= \frac{[f(y+\delta_k) - f(y+\delta_k/2)] \int_y^{y+\delta_k/2} f(x)dx}{\int_y^{y+\delta_k} f(x)dx \sqrt{\int_y^{y+\delta_k/2} f(x)dx \int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}} - \frac{[f(y+\delta_k/2) - f(y)]\left(\int_y^{y+\delta_k} f(x)dx - \int_y^{y+\delta_k/2} f(x)dx\right)}{\int_y^{y+\delta_k} f(x)dx \sqrt{\int_y^{y+\delta_k/2} f(x)dx \int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}} \\
&= \frac{[f(y+\delta_k) - f(y+\delta_k/2)] \int_y^{y+\delta_k/2} f(x)dx}{\int_y^{y+\delta_k} f(x)dx \sqrt{\int_y^{y+\delta_k/2} f(x)dx \int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}} - \frac{[f(y+\delta_k/2) - f(y)] \int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}{\int_y^{y+\delta_k} f(x)dx \sqrt{\int_y^{y+\delta_k/2} f(x)dx \int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}} \\
&= \frac{\sqrt{\int_y^{y+\delta_k/2} f(x)dx \int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}}{\int_y^{y+\delta_k} f(x)dx} \left(\frac{f(y+\delta_k) - f(y+\delta_k/2)}{\int_{y+\delta_k/2}^{y+\delta_k} f(x)dx} - \frac{f(y+\delta_k/2) - f(y)}{\int_y^{y+\delta_k/2} f(x)dx}\right),
\end{aligned} \tag{B8}$$

where the first term is positive. Now, since $(\sqrt{a} - \sqrt{b})^2 = a + b - 2\sqrt{ab} \geqslant 0$, we have the well-known inequality for the geometric and arithmetic means $\sqrt{ab} \leqslant \frac{a+b}{2}$. Then, if we select $a = \frac{\int_y^{y+\delta_k/2} f(x)dx}{\int_y^{y+\delta_k} f(x)dx}$ and $b = \frac{\int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}{\int_y^{y+\delta_k} f(x)dx}$, it leads to

$$\frac{\sqrt{\int_y^{y+\delta_k/2} f(x)dx \int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}}{\int_y^{y+\delta_k} f(x)dx} \leqslant \frac{1}{2} \frac{\int_y^{y+\delta_k/2} f(x)dx + \int_{y+\delta_k/2}^{y+\delta_k} f(x)dx}{\int_y^{y+\delta_k} f(x)dx} = \frac{1}{2}. \tag{B9}$$

With this result, we get the following inequality for the absolute value of the derivative of $\theta^{(k-1)}(y)$:

$$\left|\partial_y \theta^{(k-1)}(y)\right| \leqslant \frac{1}{2}\left|\frac{f(y+\delta_k)-f(y+\delta_k/2)}{\int_{y+\delta_k/2}^{y+\delta_k} f(x)dx} - \frac{f(y+\delta_k/2)-f(y)}{\int_{y}^{y+\delta_k/2} f(x)dx}\right|. \tag{B10}$$

By using again the numerical derivative defined in Eq. (B4), we can simplify the previous expression in the following way:

$$\frac{f(y+\delta_k)-f(y+\delta_k/2)}{\int_{y+\delta_k/2}^{y+\delta_k} f(x)dx} - \underbrace{\frac{f(y+\delta_k/2)-f(y)}{\int_{y}^{y+\delta_k/2} f(x)dx}}_{:=h(y)} \tag{B11}$$

$$= h(y+\delta_k/2) - h(y) = \frac{\delta_k}{2}\partial_y^{(\delta_k/2)} h(y). \tag{B12}$$

Also, since f must be integrable, its primitive F exists. Then,

$$h(y) = \frac{f(y+\delta_k/2)-f(y)}{\int_{y}^{y+\delta_k/2} f(x)dx} = \frac{f(y+\delta_k/2)-f(y)}{F(y+\delta_k/2)-F(y)} = \frac{\partial_y^{(\delta_k/2)} f(y)}{\partial_y^{(\delta_k/2)} F(y)}. \tag{B13}$$

Let us now plug it into the inequality

$$\left|\partial_y \theta^{(k-1)}(y)\right| \leqslant \frac{1}{2}\left|\frac{\delta_k}{2}\cdot \partial_y^{(\delta_k/2)}\left(\frac{\partial_y^{(\delta_k/2)} f(y)}{\partial_y^{(\delta_k/2)} F(y)}\right)\right| = \frac{\delta_k}{4}\left|\partial_y^{(\delta_k/2)}\left(\frac{\partial_y^{(\delta_k/2)} f(y)}{\partial_y^{(\delta_k/2)} F(y)}\right)\right|. \tag{B14}$$

If we follow the same argument as in Eq. (B6), we have that the term of the numerical derivative is upper bounded by the exact derivative, in the absolute value, which corresponds to the second derivative of the function's logarithm. Hence the following inequality holds:

$$|\partial_y \theta^{(k-1)}(y)|_{[l\delta_k,(l+1)\delta_k]} \leqslant \frac{\delta_k}{4}\max_{y'\in[l\delta_k,(l+1)\delta_k]}\left|\left[\partial_y^2 \log f(y)\right]\right|_{y=y'}|. \tag{B15}$$

Finally, by plugging this result in Eq. (B6), we obtain

$$\left|\theta_{l+1}^{(k-1)} - \theta_l^{(k-1)}\right| \leqslant \max_{y'\in[l\delta_k,(l+1)\delta_k]}\left|\frac{\delta_k^2}{4}\max_{y'\in[l\delta_k,(l+1)\delta_k]}\left|\left[\partial_y^2 \log f(y)\right]\right|_{y=y'}\right| = \frac{\delta_k^2}{4}\max_{y'\in[l\delta_k,(l+1)\delta_k]}\left|\left[\partial_y^2 \log f(y)\right]\right|_{y=y'}|, \tag{B16}$$

as we wanted to prove. ∎

Corollary 2. Let f be a continuous function such that $f : [0, 1] \to \mathbb{R}^+$ and consider a block of uniformly controlled rotations of k qubits. Then, if $\exists \eta \geqslant 0$ such that $|\partial_y^2 \log f(y)| \leqslant \eta\ \forall y \in [0, 1]$,

$$\left|\theta_{l+1}^{(k-1)} - \theta_l^{(k-1)}\right| \leqslant \frac{\delta_k^2}{4}\eta. \tag{B17}$$

Corollary 3. For a nonstandardized function $f : [x_{\min}, x_{\max}] \to \mathbb{R}^+$, the result in Corollary 2 holds with the modification in the bound

$$\left|\theta_{l+1}^{(k-1)} - \theta_l^{(k-1)}\right| \leqslant \frac{\delta_k^2 \cdot \eta}{4L^2}, \tag{B18}$$

where $L = x_{\max} - x_{\min}$ and $\delta_k = \frac{L}{2^{k-1}}$.

Proof. The change of variables that map the $x' \in [0, 1]$ with $x \in [x_{\max} - x_{\min}]$ is given by

$$x = x_{\min} + x'L. \tag{B19}$$

Now, by using the chain rule, we obtain

$$\frac{\partial f}{\partial y'} = \frac{\partial f}{\partial y}\frac{\partial y'}{\partial y'} = \frac{\partial f}{\partial y}L \Rightarrow \left|\partial_y^2 \log f(y)\right| \leqslant \frac{\tilde{\eta}}{L^2}. \tag{B20}$$

∎

Corollary 4. Let η be such that $|\partial_y^2 \log f(y)| \leqslant \eta\ \forall y \in [0, 1]$. Then, the difference between any two angles is

$$\left|\theta_l^{(k-1)} - \theta_{l'}^{(k-1)}\right| \leqslant \frac{\delta_k}{4}\eta, \tag{B21}$$

$\forall l, l' \in \{1, \ldots, 2^{k-1}\}$.

Proof. The worst scenario is when $l = 1$ and $l' = 2^{k-1}$. In this case, by using the triangular inequality, we have

$$\left|\theta_1^{(k-1)} - \theta_{2^{k-1}}^{(k-1)}\right| \leqslant \left|\theta_1^{(k-1)} - \theta_2^{(k-1)}\right| + \cdots + \left|\theta_{2^{k-1}-1}^{(k-1)} - \theta_{2^{k-1}}^{(k-1)}\right| \leqslant \frac{\delta_k}{4}\eta, \tag{B22}$$

since there are less than δ_k^{-1} elements in the sum. ∎

In this situation, we can define the representative angle $\tilde{\theta}^{(k-1)}$ for the clustering process as the one corresponding to the middle part of the interval, satisfying

$$\left|\tilde{\theta}^{(k-1)} - \theta_l^{(k-1)}\right| \leqslant \frac{\delta_k}{8}\eta, \quad \forall l \in \{0, 1, \ldots, 2^{k-1} - 1\}. \tag{B23}$$

2. Error bound of the algorithm

a. Relation between the difference in the angles and the fidelity

Let us consider a system with n qubits; thus the unitary gate to prepare the quantum state representing the target density function can be written in terms of the blocks as

$$\mathfrak{U}_n = \mathcal{U}_{n-1}(\boldsymbol{\theta}^{(n-1)}) \cdots \mathcal{U}_0(\boldsymbol{\theta}^{(0)}), \tag{B24}$$

where we define

$$\mathcal{U}_{k-1}(\boldsymbol{\theta}^{(k-1)}) := F_k^{k-1}(\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}) \otimes \mathbb{1}^{\otimes(n-k)}. \tag{B25}$$

Let $\tilde{\mathfrak{U}}_n$ denote the operation $\mathfrak{U}_n$ given a representative with an error η_k between the angles for each block, i.e., $|\theta_l^{(k-1)} - \tilde{\theta}^{(k-1)}| \leqslant \eta_k$ for $l = 0, \ldots, 2^{k-1} - 1$ and $k = 1, \ldots, n$.

Theorem 3. Consider a system of n qubits and an error η_k between any angle of the kth block and its representative such that $\eta_k \leqslant \pi$, with $k = 1, \ldots, n$. Then, the fidelity

between the final states with and without clustering, $F = |\langle 0|^{\otimes n}\mathfrak{U}_n^\dagger\tilde{\mathfrak{U}}_n|0\rangle^{\otimes n}|^2$, satisfies

$$F \geqslant \prod_{k=1}^{n}\cos^2(\eta_k/2). \tag{B26}$$

Proof. We use induction to prove the inequality. Hence we start with the elemental case of $n = 1$ and later proceed assuming it is satisfied for $n-1$ and check if it holds for n.

$n = 1$:

$$\begin{aligned}\langle 0|\mathfrak{U}_1^\dagger\tilde{\mathfrak{U}}_1|0\rangle =& \langle 0|R_y^\dagger(\theta^{(0)})R_y(\tilde{\theta}^{(0)})|0\rangle \\ =& \langle 0|R_y^\dagger(\theta^{(0)}-\tilde{\theta}^{(0)})|0\rangle = \cos\left(\frac{\theta^{(0)}-\tilde{\theta}^{(0)}}{2}\right).\end{aligned} \tag{B27}$$

Notice that the angles, given by

$$\theta_l^{(n-1)} = 2\arccos\left(\sqrt{\frac{\int_{x_{\min}+l\delta_n}^{x_{\min}+(l+1/2)\delta_n} f(x)dx}{\int_{x_{\min}+l\delta_n}^{x_{\min}+(l+1)\delta_n} f(x)dx}}\right), \tag{B28}$$

take values between $2\arccos(1) = 0$ and $2\arccos(0) = \pi$. Then, since the cosine is a decreasing function in the interval $[0, \pi/2]$ and $\eta \leqslant \pi$, we have that

$$\begin{aligned}|\theta^{(0)}-\tilde{\theta}^{(0)}| \leqslant \eta_1 \Rightarrow \cos\left(\frac{\theta^{(0)}-\tilde{\theta}^{(0)}}{2}\right) \geqslant \cos(\eta_1/2) \\ \Rightarrow F \geqslant \cos^2(\eta_1/2).\end{aligned} \tag{B29}$$

$n > 1$. Assume the condition holds for $n-1$. We aim to recover the expression of $\langle 0|^{\otimes n-1}\mathfrak{U}_{n-1}^\dagger\tilde{\mathfrak{U}}_{n-1}|0\rangle^{\otimes n-1}$ so we can use the induction hypothesis. Therefore, as we did for $n = 1$, we sandwich the operator

$$\left(F_n^{n-1}\left(\boldsymbol{y},\boldsymbol{\theta}^{(n-1)}\right)\right)^\dagger F_n^{n-1}\left(\boldsymbol{y},\tilde{\boldsymbol{\theta}}^{(n-1)}\right) \tag{B30}$$

with the state $|0\rangle$ at each side. From now on during this proof, to simplify the notation, we denote the gate $F_k^{k-1}(\boldsymbol{y},\tilde{\boldsymbol{\theta}}^{(k-1)})$ as $U_{k-1}(\boldsymbol{\theta}^{(k-1)})$. Then, for n,

$$\begin{aligned}\langle 0|^{\otimes n}\mathfrak{U}_n^\dagger\tilde{\mathfrak{U}}_n|0\rangle^{\otimes n} &= \langle 0|^{\otimes n}\mathcal{U}_0^\dagger\left(\theta^{(0)}\right)\cdots\mathcal{U}_{n-1}^\dagger\left(\boldsymbol{\theta}^{(n-1)}\right)\mathcal{U}_{n-1}\left(\tilde{\boldsymbol{\theta}}^{(n-1)}\right)\cdots\mathcal{U}_0\left(\tilde{\theta}^{(0)}\right)|0\rangle^{\otimes n} \\ &= \langle 0|^{\otimes(n-1)}\otimes\langle 0|\left(U_0^\dagger\left(\theta^{(0)}\right)\otimes\mathbb{1}^{\otimes(n-1)}\right)\cdots\left(U_{n-2}^\dagger\left(\boldsymbol{\theta}^{(n-2)}\right)\otimes\mathbb{1}\right)\left(U_{n-1}^\dagger\left(\boldsymbol{\theta}^{(n-1)}\right)\right) \\ &\quad\times\left(U_{n-1}\left(\tilde{\boldsymbol{\theta}}^{(n-1)}\right)\right)\left(U_{n-2}\left(\tilde{\boldsymbol{\theta}}^{(n-2)}\right)\otimes\mathbb{1}\right)\cdots\left(U_0\left(\tilde{\theta}^{(0)}\right)\otimes\mathbb{1}^{\otimes(n-1)}\right)|0\rangle^{\otimes(n-1)}\otimes|0\rangle \\ &= \langle 0|^{\otimes(n-1)}\left(U_0^\dagger\left(\theta^{(0)}\right)\otimes\mathbb{1}^{\otimes(n-2)}\right)\cdots\left(U_{n-2}^\dagger\left(\boldsymbol{\theta}^{(n-2)}\right)\right)\left(\sum_{i_1,\ldots,i_{n-1}=0}^{1}|i_1\ldots i_{n-1}\rangle\langle i_1\ldots i_{n-1}|\langle 0|R_y^\dagger\left(\theta_{i_1\ldots i_{n-1}}^{(n-1)}-\tilde{\theta}^{(n-1)}\right)|0\rangle\right) \\ &\quad\times\left(U_{n-2}\left(\tilde{\boldsymbol{\theta}}^{(n-2)}\right)\right)\cdots\left(U_0\left(\tilde{\theta}^{(0)}\right)\otimes\mathbb{1}^{\otimes(n-2)}\right)|0\rangle^{\otimes(n-1)}.\end{aligned} \tag{B31}$$

Notice that we now we can substitute the rotation terms in the following way:

$$\langle 0|R_y^\dagger\left(\theta_{i_1\ldots i_{n-1}}^{(n-1)}-\tilde{\theta}^{(n-1)}\right)|0\rangle = \cos\left(\frac{\theta_{i_1\ldots i_{n-1}}^{(n-1)}-\tilde{\theta}^{(n-1)}}{2}\right) \geqslant \cos(\eta_n/2), \tag{B32}$$

since $\theta_{i_1\ldots i_{n-1}}^{(n-1)}, \tilde{\theta}^{(n-1)}, \eta_n \in [0,\pi]\ \forall i_1,\ldots i_{n-1} \in \{0,1\}$. However, before introducing the inequality, we need to check that the rest of the terms have the same sign. To do so, we first sandwich the terms corresponding to the block $n-2$ with the state $|0\rangle$ as well:

$$\begin{aligned}\langle 0|^{\otimes n}\mathfrak{U}_n^\dagger\tilde{\mathfrak{U}}_n|0\rangle^{\otimes n} &= \langle 0|^{\otimes(n-2)}\left(U_0^\dagger\left(\theta^{(0)}\right)\otimes\mathbb{1}^{\otimes(n-3)}\right)\cdots\left(U_{n-3}^\dagger\left(\boldsymbol{\theta}^{(n-3)}\right)\right) \\ &\quad\times\left[\sum_{i_1,\ldots,i_{n-1}=0}^{1}|i_1\ldots i_{n-2}\rangle\langle i_1\ldots i_{n-2}|\langle 0|R_y^\dagger\left(\theta_{i_1\ldots i_{n-2}}^{(n-2)}\right)|i_{n-1}\rangle\langle i_{n-1}|R_y\left(\tilde{\theta}^{(n-2)}\right)|0\rangle\right. \\ &\quad\left.\times\cos\left(\frac{\theta_{i_1\ldots i_{n-1}}^{(n-1)}-\tilde{\theta}_{i_1\ldots i_{n-1}}^{(n-1)}}{2}\right)\right]\left(U_{n-3}\left(\tilde{\boldsymbol{\theta}}^{(n-3)}\right)\right)\cdots\left(U_0\left(\tilde{\theta}^{(0)}\right)\otimes\mathbb{1}^{\otimes(n-3)}\right)|0\rangle^{\otimes(n-2)}.\end{aligned} \tag{B33}$$

Here, since all the angles are between 0 and π, we have that

$$\langle 0|R_y^\dagger\left(\theta_{i_1\ldots i_{n-2}}^{(n-2)}\right)|0\rangle\langle 0|R_y\left(\tilde{\theta}^{(n-2)}\right)|0\rangle = \cos\frac{\theta_{i_1\ldots i_{n-2}}^{(n-2)}}{2}\cos\frac{\tilde{\theta}^{(n-2)}}{2} \geqslant 0, \tag{B34}$$

$$\langle 0|R_y^\dagger\left(\theta_{i_1\ldots i_{n-2}}^{(n-2)}\right)|1\rangle\langle 1|R_y\left(\tilde{\theta}^{(n-2)}\right)|0\rangle = \sin\frac{\theta_{i_1\ldots i_{n-2}}^{(n-2)}}{2}\sin\frac{\tilde{\theta}^{(n-2)}}{2} \geqslant 0, \tag{B35}$$

$\forall i_1, \dots i_{n-2} \in \{0, 1\}$. Then, by proceeding analogously with the rest of the blocks, we obtain that all terms are positive. Hence we can apply the inequality:

$$\begin{aligned}
\langle 0|^{\otimes n}\mathfrak{U}_n^\dagger\tilde{\mathfrak{U}}_n|0\rangle^{\otimes n} &\geqslant \langle 0|^{\otimes(n-1)}\big(U_0^\dagger\big(\theta^{(0)}\big)\otimes\mathbb{1}^{\otimes(n-2)}\big)\cdots\big(U_{n-3}^\dagger\big(\boldsymbol{\theta}^{(n-3)}\big)\otimes\mathbb{1}\big)\\
&\times\left(\sum_{i_1,\dots,i_{n-2}=0}^{1}|i_1\dots i_{n-2}\rangle\langle i_1\dots i_{n-2}|\otimes\sum_{i_{n-1}=0}^{1}R_y^\dagger\big(\theta_{i_1\dots i_{n-2}}^{(n-2)}\big)|i_{n-1}\rangle\langle i_{n-1}|R_y\big(\tilde{\theta}^{(n-2)}\big)\right)\\
&\times\big(U_{n-3}\big(\tilde{\boldsymbol{\theta}}^{(n-3)}\otimes\mathbb{1}\big)\big)\cdots\big(U_0\big(\tilde{\theta}^{(0)}\big)\otimes\mathbb{1}^{\otimes(n-2)}\big)|0\rangle^{\otimes(n-1)}\cos(\eta_n/2)\\
&=\langle 0|^{\otimes(n-1)}\big(U_0^\dagger(\theta^{(0)})\otimes\mathbb{1}^{\otimes(n-2)}\big)\cdots\big(U_{n-3}^\dagger\big(\boldsymbol{\theta}^{(n-3)}\big)\otimes\mathbb{1}\big)\\
&\times\left(\sum_{i_1,\dots,i_{n-2}=0}^{1}|i_1\dots i_{n-2}\rangle\langle i_1\dots i_{n-2}|\otimes R_y^\dagger\big(\theta_{i_1\dots i_{n-2}}^{(n-2)}\big)R_y\big(\tilde{\theta}^{(n-2)}\big)\right)\\
&\times\big(U_{n-3}\big(\tilde{\boldsymbol{\theta}}^{(n-3)}\big)\otimes\mathbb{1}\big)\cdots\big(U_0(\tilde{\theta}^{(0)})\otimes\mathbb{1}^{\otimes(n-2)}\big)|0\rangle^{\otimes(n-1)}\cos(\eta_n/2)\\
&=\langle 0|^{\otimes(n-1)}\mathfrak{U}_{n-1}^\dagger\tilde{\mathfrak{U}}_{n-1}|0\rangle^{\otimes(n-1)}\cos(\eta_n/2).
\end{aligned}\tag{B36}$$

Finally, using the induction hypothesis, we conclude that

$$\begin{aligned}
\langle 0|^{\otimes n}\mathfrak{U}_n^\dagger\tilde{\mathfrak{U}}_n|0\rangle^{\otimes n} &\geqslant \prod_{k=1}^{n-1}\cos(\eta_k/2)\cos(\eta_n/2)=\prod_{k=1}^{n}\cos(\eta_k/2)\\
&\Rightarrow F\geqslant\prod_{k=1}^{n}\cos^2(\eta_k/2).
\end{aligned}\tag{B37}$$

■

Notice that one of the conditions of the previous Theorem is that $\eta_k \leqslant \pi\ \ \forall k \in \{1, \dots, n\}$. Recall that, as seen in Eq. (B23), $\eta_k \equiv \frac{\delta_k}{8}\eta$. Therefore, we obtain the following condition for the value of η for Theorem 3 to be satisfied:

$$\frac{\delta_k}{8}\eta\leqslant\pi\Rightarrow\eta\leqslant 2^{k-1}8\pi.\tag{B38}$$

In particular, if this holds for the smallest k, with value 1, it will be satisfied for the rest of the blocks. Then, we can write the previous condition simply as $\eta \leqslant 8\pi$.

b. Expression of k_0

Assume only clusterization for $k > k_0$:

$$\begin{aligned}
F&\geqslant\prod_{k=k_0+1}^{n}\cos^2\left(\frac{\eta}{4\times 2^k}\right)\geqslant\prod_{k=k_0+1}^{n}e^{-2(\frac{\eta}{4\times 2^k})^2}\\
&=e^{-2\frac{\eta^2}{16}\sum_{k=k_0+1}^{n}4^{-k}},
\end{aligned}\tag{B39}$$

since $\cos(x) \geqslant e^{-x^2}$ for $x \lessapprox \pi/2$. Also, the sum is a geometric series with value

$$\sum_{k=k_0+1}^{n}4^{-k}=\frac{1}{3}(4^{-k_0}-4^{-n}).\tag{B40}$$

Then,

$$F\geqslant e^{-\frac{2}{3}\frac{\eta^2}{16}(4^{-k_0}-4^{-n})}:=F_{k_0}.\tag{B41}$$

Let us finally find an expression for k_0 given minimum fidelity $F_{k_0} = 1 - \epsilon$:

$$-\frac{2}{3}\frac{\eta^2}{16}(4^{-k_0}-4^{-n})=\log F_{k_0},\tag{B42}$$

$$(4^{-k_0}-4^{-n})=-\frac{3}{2}\frac{16}{\eta^2}\log F_{k_0},\tag{B43}$$

$$4^{-k_0}=4^{-n}-\frac{3}{2}\frac{16}{\eta^2}\log F_{k_0},\tag{B44}$$

$$-k_0\log 4=\log\left(4^{-n}-\frac{3}{2}\frac{16}{\eta^2}\log F_{k_0}\right),\tag{B45}$$

$$k_0=-\frac{\log\left(4^{-n}-\frac{3}{2}\frac{16}{\eta^2}\log 1-\epsilon\right)}{\log 4}.\tag{B46}$$

Note that, in the limit for both $F_0 \to 1$ and $\eta \to \infty$, we have that

$$k_0\to-\log(4^{-n})/\log(4)=n,\tag{B47}$$

as expected.

Also, since k_0 must be an integer, we take the ceiling of its value. In addition, we consider the minimum of this parameter to be 2, so the condition of $\eta \leqslant 8\pi$ is satisfied. Hence the final expression for k_0 is

$$k_0=\max\left\{\left\lceil-\frac{1}{2}\log_2\left(4^{-n}-\frac{96}{\eta^2}\log(1-\epsilon)\right)\right\rceil,2\right\},\tag{B48}$$

where we have defined $\epsilon \equiv 1 - F_0$.

Once we have a fixed number of uniformly controlled rotation blocks, k_0, for which there is no clustering, we count the number of necessary gates. The gate $F_k^{k-1}(\hat{\mathbf{y}}, \boldsymbol{\theta})$ can be implemented with 2^{k-1} CNOTs and 2^{k-1} single-qubit gates [40]. Thus the total number of gates is

$$\text{No. CNOTs}=\sum_{k=1}^{k_0}2^{k-1}=2^{k_0}-1,\tag{B49}$$

$$\text{No. SQGs} \quad = \sum_{k=1}^{k_0} 2^{k-1} + \sum_{k=k_0+1}^{n} 1 = 2^{k_0} - 1 + n - k_0. \tag{B50}$$

3. Final protocol

Definition 1. Let $f : [x_{\min}, x_{\max}] \to \mathrm{IR}^+$ be a positive function in $L^2([x_{\min}, x_{\max}])$. We define the n-qubit normalized representative state of $f(x)$ as the n-qubit state $|f(x)\rangle_n = \sum_{j=0}^{2^n-1} f(x_{\min} + j\delta_n)|j\rangle$, with $\delta_n = \frac{x_{\max}-x_{\min}}{2^n-1}$ and $\sum_{j=0}^{2^n-1} f^2(x_{\min} + j\delta_n) = 1$.

Theorem 4. Let $f : [0, 1] \to \mathrm{IR}^+$ be a positive integrable function in $L^2([0, 1])$ and $0 \leqslant \eta \leqslant 8\pi$ a constant such that $\eta = \sup_{x\in[0,1]} |\partial_x^2 \log f^2(x)|$. Then, it is possible to approximate the n-qubit representative state of $f(x)$, $|f(x)\rangle_n$, by a quantum state $|\Psi(f)\rangle_n$ such that the fidelity $|\langle\Psi(f)|f(x)\rangle_n|^2 \geqslant 1 - \epsilon$ with at most $2^{k_0(\epsilon)} - 1$ two-qubit gates, with

$$k_0(\epsilon) = \max\left\{\left\lceil -\frac{1}{2}\log_2\left(4^{-n} - \frac{96}{\eta^2}\log(1-\epsilon)\right)\right\rceil, 2\right\}, \tag{B51}$$

and the circuit to perform it is given in Fig. 14.

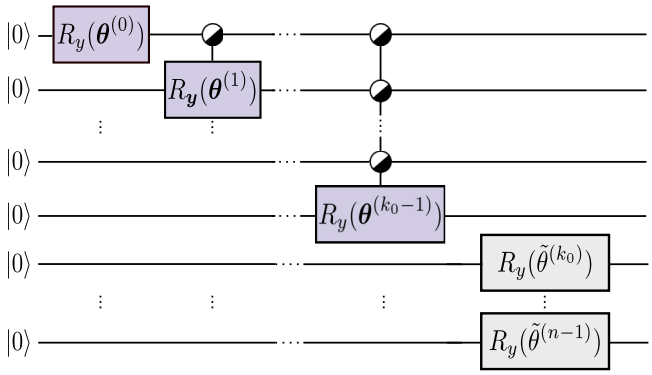

FIG. 14. Quantum circuit performing the protocol presented in this section, based on the Grover-Rudolph method for a system of n qubits.

Proof. Using Theorem 2 and Corollary 4, we can define the representative angle for each block as in Eq. (B23). Then, by applying Theorem 3 with the development between Eqs. (B39) and (B46), it is clear that the fidelity will be larger than or equal to $1 - \epsilon$. Finally, Eq. (B49) states that the number of two-qubit gates necessary to realize this protocol is $2^{k_0} - 1$, as we wanted to prove. ■

[1] F. Arute, K. Arya, R. Babbush *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[2] Y. Wu*et al.*, Strong Quantum Computational Advantage Using a Superconducting Quantum Processor, Phys. Rev. Lett. **127**, 180501 (2021).

[3] F. Zhong, H. Wang, Y. Deng *et al.*, Quantum computational advantage using photons, Science **370**, 1460 (2020).

[4] A. W. Harrow, A. Hassidim and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, Phys. Rev. Lett. **103**, 150502 (2009).

[5] A. M. Childs, R. Kothari and R. D. Somma, Quantum algorithm for systems of linear equations with exponentially improved dependence on precision, SIAM J. Comput. **46**, 1920 (2017).

[6] N. Wiebe, D. Braun, and S. Lloyd, Quantum Algorithm for Data Fitting, Phys. Rev. Lett. **109**, 050505 (2012).

[7] B. D. Clader, B. C. Jacobs, and C. R. Sprouse, Preconditioned Quantum Linear System Algorithm, Phys. Rev. Lett. **110**, 250504 (2013).

[8] A. Scherer, B. Valiron, S.-C. Mau, S. Alexander, E. van den Berg, and T. E. Chapuran, Concrete resource analysis of the quantum linear-system algorithm used to compute the electromagnetic scattering cross section of a 2D target, Quantum Inf. Process. **16**, 60 (2017).

[9] P. Rebentrost, B. Gupt, and T. R. Bromley, Quantum computational finance: Monte Carlo pricing of financial derivatives, Phys. Rev. A **98**, 022321 (2018).

[10] N. Stamatopoulos, D. J. Egger, Y. Sun, C. Zoufal, R. Iten, N. Shen, and S. Woerner, Option pricing using quantum computers, Quantum **4**, 291 (2020).

[11] J. Gonzalez-Conde, A. Rodríguez-Rozas, E. Solano, and M. Sanz, Simulating option price dynamics with exponential quantum speedup, arXiv:2101.04023.

[12] J. Liu, H. Ø. Kolden, H. K. Krovi, N. F. Loureiro, K. Trivisa, and A. M. Childs, Efficient quantum algorithm for dissipative nonlinear differential equations, Proc. Natl. Acad. Sci. USA **118**, e2026805118 (2021).

[13] B. Zanger, C. B. Mendl, M. Schulz, and M. Schreiber, Quantum algorithms for solving ordinary differential equations via classical integration methods, Quantum **5**, 502 (2021).

[14] J. J. García-Ripoll, Quantum-inspired algorithms for multivariate analysis: from interpolation to partial differential equations, Quantum **5**, 431 (2021).

[15] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum algorithms for supervised and unsupervised machine learning, arXiv:1307.0411 v2.

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

[17] H. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Power of data in quantum machine learning, Nat. Commun. **12**, 2631 (2021).

[18] M. Schuld and N. Killoran, Quantum Machine Learning in Feature Hilbert Spaces, Phys. Rev. Lett. **122**, 040504 (2019).

[19] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, Quantum embeddings for machine learning, arXiv:2001.03622.

[20] N. Wiebe, Key questions for the quantum machine learner to ask themselves, New J. Phys. **22**, 091001 (2020).

[21] V. Havlíček *et al.*, Supervised learning with quantum enhanced feature spaces, Nature (London) **567**, 209 (2019).

[22] J. A. Cortese and T. M. Braje, Loading classical data into a quantum computer, arXiv:1803.01958.

[23] M. Schuld and F. Petruccione, *Supervised Learning with Quantum Computers* (Springer, Berlin, 2018).

[24] L. K. Grover, Synthesis of Quantum Superpositions by Quantum Computation, Phys. Rev. Lett. **85**, 1334 (2000).

[25] Y. R. Sanders, G. H. Low, A. Scherer, and D. W. Berry, Black-Box Quantum State Preparation without Arithmetic, Phys. Rev. Lett. **122**, 020502 (2019).
[26] L. Grover and T. Rudolph, Creating superpositions that correspond to efficiently integrable probability distributions, arXiv:quant-ph/0208112.
[27] M. Möttönen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, Transformation of quantum states using uniformly controlled rotations, Quantum Inf. Comput. **5**, 467 (2005).
[28] K. Ghosh, Encoding classical data into quantum data, arXiv:2107.09155.
[29] M. Plesch and Č. Brukner, Quantum-state preparation with universal gate decompositions, Phys. Rev. A **83**, 032302 (2011).
[30] J. Bausch, Fast black-box quantum state preparation, Quantum **6**, 773 (2022).
[31] I. F. Araujo, D. K. Park, F. Petruccione, and A. J. da Silva, A divide-and-conquer algorithm for quantum state preparation, Sci. Rep. **11**, 6329 (2021).
[32] X.-M. Zhang, M.-H. Yung, and X. Yuan, Low-depth quantum state preparation, arXiv:2102.07533.
[33] J. Zhao, Y. Wu, G. Guo, and G. Guo, State preparation based on quantum phase estimation, arXiv:1912.05335.
[34] C. W. Bauer, P. Deliyannis, M. Freytsis, and B. Nachman, Practical considerations for the preparation of multivariate Gaussian states on quantum computers, arXiv:2109.10918.
[35] S. Anwer, A. Younes, I. Elkabani, and A. Elsayed, Preparation of quantum superposition using partial negation, arXiv:2109.14369.
[36] A. Montanaro, Quantum speedup of Monte Carlo methods, Proc. R. Soc. Ser. A **471**, 20150301 (2015).
[37] N. Klco and M. J. Savage, Digitization of scalar fields for quantum computing, Phys. Rev. A **99**, 052335 (2019).
[38] S. P. Jordan, K. S. M. Lee, and J. Preskill, Quantum algorithms for quantum field theories, Science **336**, 1130 (2012).
[39] V. Bergholm, J. J. Vartiainen, M. Möttönen, and M. M. Salomaa, Quantum circuits with uniformly controlled one-qubit gates, Phys. Rev. A **71**, 052330 (2005).
[40] M. Möttönen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, Quantum Circuits for General Multiqubit Gates, Phys. Rev. Lett. **93**, 130502 (2004).
[41] A. Y. Kitaev, Quantum computations: Algorithms and error correction, Russian Math. Surveys **52**, 1191 (1997).
[42] C. M. Dawson and M. Nielsen, The Solovay-Kitaev algorithm, Quantum Inf. Comput. **6**, 81 (2006).
[43] C. Zoufal, A. Lucchi, and S. Woerner, Quantum generative adversarial networks for learning and loading random distributions, npj Quantum Inf. **5**, 103 (2019).
[44] K. Nakaji *et al.*, Approximate amplitude encoding in shallow parameterized quantum circuits and its application to financial market indicator, Phys. Rev. Res. **4**, 023136 (2022).
[45] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).
[46] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).
[47] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
[48] S. Chakrabarti, R. K. Nakumar, G. Mazzola, N. Stamatopoulos, S. Woerner, and W. J. Zeng, A threshold for quantum advantage in derivative pricing, Quantum **5**, 463 (2021).
[49] S. Herbert, The problem with Grover-Rudolph state preparation for quantum Monte Carlo, Phys. Rev. E **103**, 063302 (2021).
[50] M. C. Collodo, J. Herrmann, N. Lacroix, C. K. Andersen, A. Remm, S. Lazar, J.-C. Besse, T. Walter, A. Wallraff, and C. Eichler, Implementation of Conditional Phase Gates Based on Tunable ZZ Interactions, Phys. Rev. Lett. **125**, 240502 (2020).
[51] U. L. Heras, A. Mezzacapo, L. Lamata, S. Filipp, A. Wallraff, and E. Solano, Digital Quantum Simulation of Spin Systems in Superconducting Circuits, Phys. Rev. Lett. **112**, 200501 (2014).
[52] U. Las Heras, L. García-Álvarez, A. Mezzacapo, L. Lamata, and E. Solano, in *Mathematics for Industry 11*, edited by R. S. Anderssen *et al.* (Springer, Japan, 2015).
[53] A. Cauchy, Methode generale pour la resolution des systemes d'equations simultanees, C. R. Acad. Bulg. Sci. **25**, 536 (1847).
[54] A. Barenco, C. H. Bennett, R. Cleve, D. P. Di Vincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Elementary gates for quantum computation, Phys. Rev. A **52**, 3457 (1995).
[55] L. Grover and T. Rudolph, Creating superpositions that correspond to efficiently integrable probability distributions, arXiv:quant-ph/0208112.