Parameter shift rule with optimal phase selection

Markovich, L.; Malikis, S.; Polla, S.; Tura Brugués, J.

Parameter shift rule with optimal phase selection

Liubov Markovich *

Instituut-Lorentz, Universiteit Leiden, P.O. Box 9506, 2300 RA Leiden, The Netherlands Applied Quantum Algorithms—Leiden, The Netherlands and Russian Quantum Center, Skolkovo, Moscow 121205, Russia

Savvas Malikis, Stefano Polla, and Jordi Tura

Instituut-Lorentz, Universiteit Leiden, P.O. Box 9506, 2300 RA Leiden, The Netherlands and Applied Quantum Algorithms—Leiden, The Netherlands

(Received 25 March 2024; accepted 30 May 2024; published 21 June 2024)

Parameter shift rules enable the estimation of the derivatives of expectation values with respect to the dynamical evolution of a state. Thus, they provide a valuable tool in variational optimization and insight into the dynamical behavior of quantum systems. Constructing parameter shift rules applicable to complex dynamical generators is useful towards improving the control of analog quantum simulation devices and extending the applicability of variational quantum simulating methods. However, parameter shift rules are typically designed only for dynamical generators with highly degenerate or equidistant eigenvalues, such as Pauli operators, commonplace in the construction of parametrized quantum circuits. For generators with irregularly spaced eigenvalues, effective rules have not been established. We provide insights about the optimal design of a parameter shift rule tailored to various sorts of spectral information that may be available. The proposed method targets the calculation of any linear combination of k th derivatives with respect to the conjugate variable of a generator with a known spectrum, regardless of how close the eigenvalues are to each other. We discuss the optimal choice of parameter shifts, minimizing their number and the variance of the target estimator. Furthermore, we discuss the construction of parameter shift rules when only partial information about the generator spectrum is known.

DOI: 10.1103/PhysRevA.109.062429

I. INTRODUCTION

Many near-term quantum computing methods are based on parametrized quantum circuits (PQCs) [1,2], sequences of quantum gates specified by classical parameters, which are tuned iteratively in order to address specific tasks. Examples of these variational quantum algorithms include the quantum approximate optimization algorithm [3] and the variational quantum eigensolver [1,4], along with many proposed approaches to quantum machine learning. Variational algorithms often rely on derivative estimation to guide the parameter optimization process, leading to improved efficiency and better outcomes [5,6]. Derivative estimation is also crucial in scientific and engineering fields for solving differential equations, numerical integration, and precise modeling of complex systems. It plays a key role in quantum machine learning algorithms, like quantum neural networks and support vector machines, enhancing their learning capabilities.

The output of a PQC provides a probabilistic result, with the expectation value of an observable serving as an estimate. While mean values of simple variables are found by averaging measurement results, the estimation of expectation values of multiqubit observables can be optimized through various approaches [1,7–16]. Therefore, it is desirable to formally approximate the derivatives of these averages as a linear combination of expectation value estimates on states generated by different choices of the variational circuit parameters. Various methods, such as finite differences or robust polynomial interpolation techniques [17,18], can be used to approximate these derivatives. However, exact derivative calculations are challenging to implement on hardware because traditional mathematical derivatives cannot be directly applied to quantum gates, which must be unitary. This leads to the necessity of calculating analytical derivatives using combinations of quantum implementable operations.

Parameter shift rules (PSRs), introduced in quantum machine learning [2,5], provide a method for obtaining exact derivatives of expectation values by evaluating parameter shifted instances of a PQC. These rules relate derivatives of the expectation value $f(t) = \langle\psi(t)| O |\psi(t)\rangle$ with respect to some parameter t to evaluations of the function at a set of different points $\{t_j\}$. The functional form of these rules depends on the form of the generator H defining the dependence of the state on the conjugate variable t, i.e., $|\psi(t)\rangle = e^{-iHt} |\psi_0\rangle$. From here on, we call the generator H a Hamiltonian. However, it should be noted that H is a generic Hermitian generator (not necessarily representing the energy of a physical system), and that in the definition of a variational state multiple parametrized operations governed by different generators can be applied in sequence.

The original two-term PSR is based on gates with two distinct eigenvalues [5,6,19] and the PSR variations [20–22] maintain the restriction on the number of Hamiltonian eigenvalues. The stochastic PSR [23], as well as the generalized PSR [24], allow to calculate derivatives for and generator

*Contact author: markovich@mail.lorentz.leidenuniv.nl

with equidistant eigenvalues. The mentioned rules are limited to evenly spaced parameter shifts and do not take into account Hamilton's eigenvalue structure. While some exploration of different shifts exists [25] for standard PSRs, including symmetric and distinct shifts, to our best knowledge, no study has been conducted on their optimal selection. However, for quasidegenerate Hamiltonian eigenvalue structures, existing PSRs may yield poor results, indicating the need for versatile parameter shift selection.

In this paper, we introduce the PSR with the selection of the shift parameters and the vector of coefficients to derive any order derivative of the observables' mean value and their linear combinations. By expressing the parametrized unitary $U(t) = e^{iHt}$ as a sum of finite powers of the Hamiltonian H, we reduce the problem to solving an operator equation depending on the differences between the pairs of eigenvalues of H. The problem of seeking a solution to the operator equation is considered well posed by Hadamard if the solution is existing, unique, and stable [26]. If the solution lacks at least one of these three conditions, it is deemed ill posed [27]. Generally, determining the optimal parameter shifts can become an ill-posed problem, i.e., when some eigenvalues are close to each other, leading to quasidegeneracies in the distances between various pairs of eigenvalues.

In the case of a well-posed problem, we provide the explicit solution yielding the best estimate for $f'(t)$ and any of its higher derivatives $f^{(p)}(t)$ and their linear combinations. In the challenging scenario of an ill-posed problem, we begin by considering the ideal situation of perfectly equidistant eigenvalues [see Fig. 1(a)]. We demonstrate that it is possible to reduce the dimension of the system of linear equations to transform it into a well-posed and solvable problem. The exact solution to the problem and the optimal set of parameter shifts are provided. Notably, this solution matches the one intuitively introduced in [24]; we add to this result by establishing that this is the unique solution for such a problem. In the scenario where the eigenvalues are not perfectly equidistant but are slightly perturbed [see Fig. 1(b)], the provided solution remains applicable.

This situation reflects a realistic scenario where eigenvalues are not precisely known and their values may be affected by estimation errors and measurement noise. We quantify the distance between the obtained solution and the optimal one as a function of the magnitude of the perturbation. We examine the scenario of equidistant sets of eigenvalues [see Fig. 1(c)]. This situation corresponds to cases where different sets of experiments are conducted to estimate the eigenvalues. Finally, we consider the ill-posed problem case with no structure in the eigenvalues [see Fig. 1(d)]. We demonstrate how to solve such an ill-posed problem using the regularization method [27]. Introducing a regularization parameter allows us to find an approximate solution for the operator equation which can be asymptotically reconducted to the true one. This offers a recipe to determine the best coefficients and parameter shifts numerically even for complex, ill-posed problems. Hence, our paper comprehensively covers all Hamiltonian eigenvalue structures.

The paper is organized as follows: In Sec. II we define the problem of determining general parameter shift rules. We defer to Appendix A for a brief overview of established PSRs.

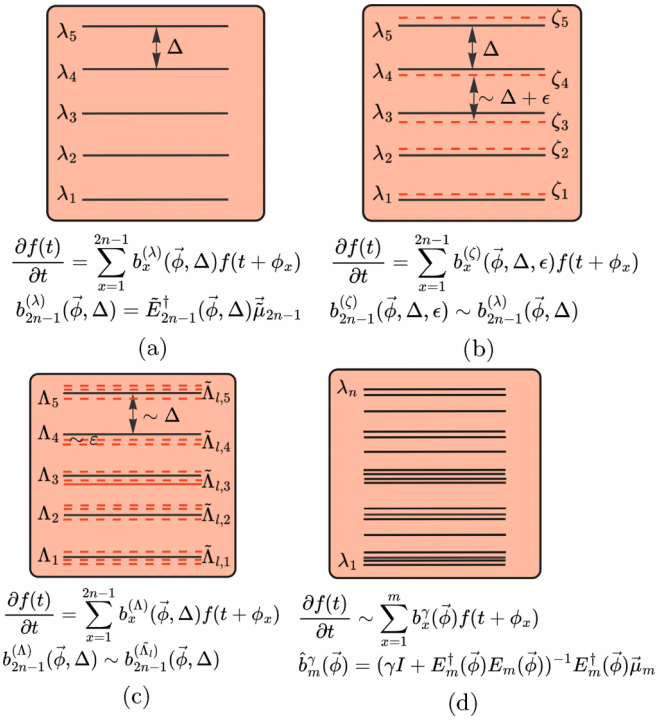

FIG. 1. Different cases of the eigenvalue structures: (a) the equidistant eigenvalues ($n = 5$), (b) the equidistant eigenvalues (solid lines) and perturbed eigenvalues (dashed lines), (c) the equidistant eigenvalues (solid lines) and sets of eigenvalues close to the equidistant positions (dashed lines), and (d) n eigenvalues with no structure.

In Sec. III, we delve into the general parameter shift rule and derive the optimal coefficient for a well-posed problem. In Sec. IV, we explore the ill-posed problem, examining cases such as equidistant eigenvalues of the Hamiltonian, equidistant eigenvalues except for one, slightly perturbed equidistant eigenvalues, and highly nonequidistant eigenvalues forming equidistant sets. In Sec. V, we present a phase-shift selection method for an ill-posed problem as defined by Hadamard, specifically for eigenvalues with no discernible structure. We conclude the main text in Sec. VI with a discussion. Additionally, in the Appendixes we provide technical details for some derivations.

II. BACKGROUND

Let $|\psi\rangle$ denote the quantum state in the Hilbert space. Consider the unitary operator $U(t)$, defined by a Hamiltonian H and a parameter t. The eigenvalues of $U(t)$ are expressed as $\{\exp(i\lambda_j t)\}_{j=1}^n$ with real-valued $\{\lambda_j\}_{j=1}^n$, we have arranged the indices j in nondecreasing order. We aim to determine the mean value of a measurable observable C defined as follows:

$$f(t) \equiv \langle\psi|\, U(t)^{\dagger} C U(t)\, |\psi\rangle\,. \tag{1}$$

The expectation value $f(t)$ can be expressed as a finite-term Fourier series

$$f(t) = a_0 + \sum_{l=1}^{m} a_l \cos(\Omega_l t) + b_l \sin(\Omega_l t), \tag{2}$$

where we have m unique positive differences $\{\Omega_l\}_{l\in m} = \{|\lambda_j - \lambda_k|, j, k \in [1, n], \lambda_j > \lambda_k\}$.

For functions $f(t)$ with a single frequency $\Omega_1 = \Omega$ (i.e., U has two eigenvalues), the derivative can be computed via the parameter shift rule [5,6,19]:

$$\left.\frac{\partial f(t)}{\partial t}\right|_{t=0} = \frac{\Omega}{2\sin(\Omega\phi_1)}[f(\phi_1) - f(-\phi_1)], \tag{3}$$

where $\phi_1 \in (0, \pi)$. An overview of some variations of the PSRs is provided in Appendix A.

In [24], the general parameter shift rules are introduced for the scenario of evenly spaced parameter shifts $\phi_j = (2j-1)\pi/2n$ ($\phi_j = j\pi/n$), where $j \in [1, n]$ is considered to reconstruct odd (even) functions:

$$f'(0) = \sum_{j=1}^{2n} f\left(\frac{(2j-1)\pi}{2n}\right)\frac{(-1)^{j-1}}{4n\sin^2\left(\frac{(2j-1)\pi}{4n}\right)},$$
$$f''(0) = -f(0)\frac{2n^2+1}{6} + \sum_{j=1}^{2n-1} f\left(\frac{j\pi}{n}\right)\frac{(-1)^{j-1}}{2\sin^2\left(\frac{j\pi}{2n}\right)}. \tag{4}$$

The selection of parameter shifts depends on various factors, including the specific problem being solved, the available resources, and the desired accuracy. Parameter shifts may be chosen based on mathematical considerations or analytical insights into the problem structure. In other cases, numerical methods or optimization techniques can be employed to find optimal parameter shift values that minimize errors or enhance efficiency. Further, we present an optimal parameter shift selection method designed to accommodate any structure within the Hamiltonian system.

III. PSR FOR A WELL-POSED PROBLEM

Any function from H can be expressed as a finite sum of H powers (see Appendix B), namely,

$$e^{\mathbb{i}Ht} = \sum_{k=0}^{n-1} a_k(t)H^k, \tag{5}$$

where the coefficients are $|a(t)\rangle = \Lambda^{-1}|e(t)\rangle$, with Λ being the $n\times n$ Vandermonde matrix containing the vector $\vec{\lambda}$, and $\langle k|e(t)\rangle = e^{\mathbb{i}\lambda_k t}$. Subsequently, we can reformulate (1) as follows:

$$f(t) = \mathrm{Tr}[|e(t)\rangle\langle e(t)|(\Lambda^{-1})^\dagger \tilde{C}\Lambda^{-1}], \tag{6}$$

where $(\tilde{C})_{k,l} \equiv \langle\psi|H^k C H^l|\psi\rangle$.

The PSRs establish connections between derivatives of a quantum function and the function's evaluations at distinct points. In this paper, we introduce the PSR:

$$\frac{\partial f(t)}{\partial t} = \sum_{x=1}^{m} b_x(\vec{\phi})f(t+\phi_x), \tag{7}$$

with the shift parameters $\vec{\phi} = \{\phi_x\}_{x=1}^m$ and the vector of coefficients $b(\vec{\phi}) = \{b_x(\vec{\phi})\}_{x=1}^m$. Utilizing (6), we can rewrite the PSR (7) as follows:

$$|e'(t)\rangle\langle e(t)| - |e(t)\rangle\langle e'(t)| = \sum_{x=1}^{m} b_x(\vec{\phi})\,|e(t+\phi_x)\rangle\langle e(t+\phi_x)|\,. \tag{8}$$

This is equivalent to solving the following system of equations $\forall k, l \in \overline{1,n}$:

$$\mathbb{i}(\lambda_k - \lambda_l)e^{\mathbb{i}(\lambda_k-\lambda_l)t} = \sum_{x=1}^{m} b_x(\vec{\phi})e^{\mathbb{i}(\lambda_k-\lambda_l)t}e^{\mathbb{i}(\lambda_k-\lambda_l)\phi_x}, \tag{9}$$

where we used the notation $\langle(k,l)|e(t)\rangle = e^{\mathbb{i}(\lambda_k-\lambda_l)t}$. Since the latter equation must be satisfied for every t, the compatibility equation reads as

$$\sum_{x=1}^{m} b_x(\vec{\phi})e^{\mathbb{i}(\lambda_k-\lambda_l)\phi_x} = \mathbb{i}(\lambda_k - \lambda_l), \qquad \forall k, l \in \overline{1,n}. \tag{10}$$

This system is highly nonlinear in the $\vec{\phi}$ but it is nevertheless linear in the b's.

Consider Hermitian metric spaces U and V. The continuous one-to-one operator E from U to V corresponds to the $m\times m$ matrix $E_m(\vec{\phi})$ with elements given by $E_{(k,l),x} = e^{\mathbb{i}\mu_{(k,l)}\phi_x}$, $k, l \in \overline{1,n}$. Here, the distance between two eigenvalues is defined as $\mu_{(k,l)} \equiv \lambda_k - \lambda_l$, $k, l = \overline{1,n}$. The function $\vec{\mu}_m$ represents the $m\times 1$ vector with elements $\mathbb{i}(\lambda_k - \lambda_l)$. Thus, (10) in operator form can be expressed as

$$E_m(\vec{\phi})b_m(\vec{\phi}) = \mu_m, \quad b \in U, \quad \mu \in V. \tag{11}$$

In this section, we assume that the phases are selected in such a way that the problem is well posed by Hadamard. The constraints on them we observe further. Then we can determine the vector of solutions of (11) as follows:

$$b_m(\vec{\phi}) = E_m^{-1}(\vec{\phi})\mu_m. \tag{12}$$

By applying Cramer's rule, we can express each element of $b_m(\vec{\phi})$ in closed form:

$$b_x(\vec{\phi}) = \det[E_m(\vec{\phi}/\phi_x)]\cdot[\det E_m(\vec{\phi})]^{-1}, \tag{13}$$

where we introduce the notation

$$E_m(\vec{\phi}/\phi_x) \equiv \begin{bmatrix} 1 & 1 & \dots & 0 & \dots & 1 \\ e^{i\mu_{(1,2)}\phi_1} & e^{i\mu_{(1,2)}\phi_2} & \dots & i\mu_{(1,2)} & \dots & e^{i\mu_{(1,2)}\phi_m} \\ e^{-i\mu_{(1,2)}\phi_1} & e^{-i\mu_{(1,2)}\phi_2} & \dots & -i\mu_{(1,2)} & \dots & e^{-i\mu_{(1,2)}\phi_m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{i\mu_{(n-1,n)}\phi_1} & e^{i\mu_{(n-1,n)}\phi_2} & \dots & i\mu_{(n-1,n)} & \dots & e^{i\mu_{(n-1,n)}\phi_m} \\ e^{-i\mu_{(n-1,n)}\phi_1} & e^{-i\mu_{(n-1,n)}\phi_2} & \dots & -i\mu_{(n-1,n)} & \dots & e^{-i\mu_{(n-1,n)}\phi_m} \end{bmatrix} = \left.\frac{\partial E_m(\vec{\phi})}{\partial\phi_x}\right|_{\phi_x=0}, \tag{14}$$

that is, the matrix $E_m(\vec{\phi})$ with the x column substituted by the μ_m vector and not depending on ϕ_x. The determinants are not equal to zero if the matrix does not have identical columns or rows or if it lacks entirely zero columns or rows. In this regard, we establish that $\phi_i \neq \pm\phi_j + 2\pi c$, where $c \in \mathbb{Z}$, and all distances $\mu_{(k,l)}$ must be distinct. This implies that the equidistant n-eigenvalue problem is ill posed according to Hadamard. The resolution to this issue will be presented in the next section.

The solution (13) is exact when $m = n(n-1)+1$. This number is obtained by counting every $\mu_{(k,l)}$, $k \neq l$. The $k = l$ case always yields the same equation $\sum_x b_x(\vec{\phi}) = 0$. This automatically yields a PSR for the derivatives of arbitrary order without increasing the number of function evaluations. The following theorem holds.

Theorem III.1. Let n be the number of distinct, not equidistant, eigenvalues of H. Let $\vec{\phi} \in \mathbb{R}^m$ with $m = n(n-1)+1$ and $\phi_i \neq \pm\phi_j + 2\pi c, c \in Z, \forall i, j \in \overline{1,m}$. Then, the following parameter shift rule,

$$\frac{\partial^p f(t)}{\partial t^p} = \sum_{x=1}^{m} b_x^{(p)}(\vec{\phi}) f(t+\phi_x), \quad p \geqslant 1, \tag{15}$$

holds if, and only if, the vector $b_x(\vec{\phi})$ satisfies

$$b_x^{(p)}(\vec{\phi}) = \frac{\begin{vmatrix} \vec{v}(\phi_0) & \cdots & \frac{\partial^p \vec{v}(\phi_x)}{\partial \phi_x^p}|_{\phi_x=0} & \cdots & \vec{v}(\phi_m) \end{vmatrix}}{\begin{vmatrix} \vec{v}(\phi_0) & \cdots & \vec{v}(\phi_x) & \cdots & \vec{v}(\phi_m) \end{vmatrix}}, \tag{16}$$

where the vectors forming $E_m(\vec{\phi})$ are denoted as

$$\vec{v}(\phi_i) = (1 \quad e^{\mathbb{i}\mu_{(12)}\phi_i} \quad e^{-\mathbb{i}\mu_{(12)}\phi_i} \quad \cdots \quad e^{-\mathbb{i}\mu_{(n-1,n)}\phi_i}). \tag{17}$$

Proof. The proof relies on the ability to express $\frac{\partial^p f(t)}{\partial t^p}$ similarly to the decomposition in (7). An equation akin to (9) can be formulated, with the left-hand side featuring the term $[i(\lambda_k - \lambda_l)]^p$. Solving this equation leads to the deduction of (16).

The latter statement can be generalized. Specifically, any linear combination of high-order derivatives can be similarly expressed as follows:

$$\sum_i a_i f^{(i)}(t) = \sum_{x=0}^{m-1} \tilde{b}_x(\vec{\phi}) f(t+\phi_x), \quad \text{where}$$

$$\tilde{b}_x(\vec{\phi}) = \frac{\begin{vmatrix} \vec{v}(\phi_0) & \cdots & \sum_i a_i \frac{\partial^i \vec{v}(\phi_x)|_{\phi_x=0}}{\partial \phi_x^i} & \cdots & \vec{v}(\phi_m) \end{vmatrix}}{\begin{vmatrix} \vec{v}(\phi_0) & \cdots & \vec{v}(\phi_x) & \cdots & \vec{v}(\phi_m) \end{vmatrix}} \tag{18}$$

holds. Even though it is not generalizable for any other algebraic expression $F = F(f, f', f'' \dots)$ involving nonlinear terms, we can always rewrite it as products of functions we can compute.

Theorem III.1 works for any shift vector $\vec{\phi}$ such that the problem (12) is well posed.

The variance of the estimate of the derivative $\widehat{\frac{\partial f(t,\vec{\phi})}{\partial t}}$ is

$$\sigma^2\left(\widehat{\frac{\partial f(t,\vec{\phi})}{\partial t}}\right) = \sum_{x=1}^{m} b_x^2(\vec{\phi}) \sigma^2[\hat{f}(t)], \tag{19}$$

where we assumed that the variances $\sigma^2[\hat{f}(t+\phi_x)]$ are not dependent on the phase and are equal to $\sigma^2[\hat{f}(t)]$. Chebyshev's inequality can be written

$$\mathbb{P}\left(\left|\frac{\partial f(t)}{\partial t} - \widehat{\frac{\partial f(t,\vec{\phi})}{\partial t}}\right| \geqslant \nu\right) \leqslant \frac{\sigma^2[\hat{f}(t)]}{\nu^2} \sum_{x=1}^{m} b_x^2(\vec{\phi}), \tag{20}$$

where $\nu > 0$ is a real number. For the probability $\eta \in (0,1)$, we get

$$\nu = \left(\frac{\sigma^2[\hat{f}(t)]}{\eta} \sum_{x=1}^{m} b_x^2(\vec{\phi})\right)^{\frac{1}{2}} \tag{21}$$

and the confidence interval for the estimate of the derivative is

$$\frac{\partial f(t)}{\partial t} - \nu \leqslant \widehat{\frac{\partial f(t,\vec{\phi})}{\partial t}} \leqslant \frac{\partial f(t)}{\partial t} + \nu. \tag{22}$$

To minimize the variance (19), we need to minimize the sum of squared coefficients $b_x(\vec{\phi})$. The latter problem is equivalent to solving the following system of equations:

$$\sum_{x=1}^{m} \det[E_m(\vec{\phi}/\phi_x)] \left(\det[E_y(\vec{\phi}/\phi_x)] - \frac{\det[E_m(\vec{\phi}/\phi_x)]}{\det[E_m(\vec{\phi})]} \det[E_y(\vec{\phi})] \right) = \mathbf{0}, \tag{23}$$

where we introduce the notations $\det[E_y(\vec{\phi}/\phi_x)] \equiv \frac{\partial \det[E_m(\vec{\phi}/\phi_x)]}{\partial \phi_y}$ and $\det[E_y(\vec{\phi})] \equiv \frac{\partial \det[E_m(\vec{\phi})]}{\partial \phi_y}$. Solving this system of equations with respect to all ϕ_y, $y \in \overline{1,m}$, one can find the optimal $\vec{\phi}$ minimizing the variance (19). ■

IV. PSR FOR SPECIFIC EIGENVALUE STRUCTURES

If the problem (11) is ill posed, the solution (12) is unstable or even may not exist. Therefore, it is necessary to establish conditions on the phases and eigenvalues to ensure the solvability of the problem.

Let us first look for a set of parameter shifts such that the vectors (17) forming the matrix $E_m(\vec{\phi})$ would be orthogonal to each other. In this case, the inversion of $E_m(\vec{\phi})$ is equal to its Hermitian conjugation.

Further, we use the fact that $\mu_{(k,l)} = -\mu_{(l,k)}$ holds, so we introduce the notation $\mu_{(t,p)}$, $t < p$, $\forall t, p \in \overline{1,n}$. We impose the orthogonality condition on the columns of $E_m(\vec{\phi})$, namely, $\vec{v}^\star(\phi_j)\vec{v}(\phi_i) = 0$. We get

$$1 + 2 \sum_{t,p=1, t<p}^{n} \cos(\mu_{(t,p)}\Phi_{ij}) = 0, \quad \forall \phi_i, \phi_j \in \vec{\phi}, \tag{24}$$

where we use the notation $\Phi_{ij} \equiv \phi_i - \phi_j$. The latter condition can be rewritten as follows:

$$1 + 2\sum_{p=2}^{n} \cos\left(\mu_{(1,p)}\Phi_{ij}\right) + 2\sum_{p=3}^{n} \cos(\mu_{(2,p)}\Phi_{ij}) + \ldots\ldots + 2\sum_{p=m-1}^{n} \cos\left(\mu_{(n-2,p)}\Phi_{ij}\right) + 2\cos(\mu_{(n-1,n)}\Phi_{ij}) = 0. \tag{25}$$

To solve the latter equation, we need to make some assumptions on the eigenvalue distances $\mu_{(t,p)}$. Below, we first discuss the equidistant Hamiltonian eigenvalues case, moving on to the perturbed case in the following subsection.

A. Equidistant eigenvalues

Observing that $\mu(1,2) = 1\Delta$, $\mu_{(1,3)} = 2\Delta$, and $\mu_{(1,n)} = (n-1)\Delta$, as well as $\mu_{(2,4)} = 2\Delta$ and $\mu_{(2,n)} = (n-2)\Delta$, it is evident that $\mu_{(t,p)} = (p-t)\Delta$ for $t < p$. In this scenario, some rows of the matrix E_m coincide, rendering it singular. Consequently, the problem (11) becomes ill posed.

Since we are interested in the inversion of the matrix $E_m(\vec{\phi})$, we exclude all the similar rows, reducing the matrix to a smaller size $E_{2n-1}(\vec{\phi})$ one, which is nonsingular:

$$E_{2n-1}(\vec{\phi}) = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ e^{\mathbb{i}1\Delta\phi_1} & e^{\mathbb{i}1\Delta\phi_2} & \ldots & e^{\mathbb{i}1\Delta\phi_{2n-1}} \\ e^{-\mathbb{i}1\Delta\phi_1} & e^{-\mathbb{i}1\Delta\phi_2} & \cdot & e^{-\mathbb{i}1\Delta\phi_{2n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\mathbb{i}(n-1)\Delta\phi_1} & e^{-\mathbb{i}(n-1)\Delta\phi_2} & \ldots & e^{-\mathbb{i}(n-1)\Delta\phi_{2n-1}} \end{bmatrix}, \quad \vec{\mu}_{2n-1} = \mathbb{i}\Delta \begin{bmatrix} 0 \\ 1 \\ -1 \\ \vdots \\ -(n-1) \end{bmatrix}. \tag{26}$$

Here $\vec{\mu}_{2n-1}$ is a vector of all unique distances $\mu_{(1,i)}$, $i = \overline{1,n}$. In this case, the condition (25) can be reduced to

$$1 + 2\sum_{k=1}^{n-1} \cos(k\Delta\Phi_{ij}) = 0. \tag{27}$$

The Dirichlet kernel is defined as follows:

$$D_n(x) = 1 + 2\sum_{k=1}^{n} \cos(kx) = \frac{\sin\left[\left(n+\frac{1}{2}\right)x\right]}{\sin\left(\frac{1}{2}x\right)}, \tag{28}$$

where its zeros are at the points $x_t = \frac{2\pi t}{2n+1}$, $t \in \mathbb{Z}$. Hence, the condition (27) can be rewritten as

$$D_{n-1}(\Delta\Phi_{ij}) = 0. \tag{29}$$

Then $\forall i,\ j = [1, 2n-1]$, $i < j$ and we have a system of equations

$$\Phi_{ij} = \frac{2\pi}{(2n-1)\Delta} t_{ij}, \quad t_{ij} \in Z. \tag{30}$$

To solve it we first consider the case of equidistant parameter shifts ϕ_j, $\forall j$. From (30) we can conclude

$$\phi_j = -\frac{2\pi j}{(2n-1)\Delta}, \quad \forall j = [1, 2n-1]. \tag{31}$$

One can see that if $j = 2n-1$ holds, then $\phi_{2n-1} = -2\pi/\Delta$, and $\exp j\Delta\phi_{2n-1} = 1$. Then the matrix (26) reduces to

$$E_{2n-1} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ e^{-i\tau} & e^{-2i\tau} & e^{-3i\tau} & \ldots & 1 \\ e^{i\tau} & e^{2i\tau} & e^{3i\tau} & \ldots & 1 \\ e^{-2i\tau} & e^{-4i\tau} & e^{-6i\tau} & \ldots & 1 \\ e^{2i\tau} & e^{4i\tau} & e^{6i\tau} & \ldots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-(n-1)i\tau} & e^{-2(n-1)i\tau} & e^{-3(n-1)i\tau} & \ldots & 1 \\ e^{(n-1)i\tau} & e^{2(n-1)i\tau} & e^{3(n-1)i\tau} & \ldots & 1 \end{bmatrix}, \tag{32}$$

where we used the notation $\tau = \frac{2\pi}{(2n-1)}$. Let us normalize the latter matrix, introducing

$$\tilde{E}_{2n-1} \equiv E_{2n-1}/\sqrt{2n-1}, \quad \tilde{\vec{\mu}}_{2n-1} \equiv \vec{\mu}_{2n-1}/\sqrt{2n-1}. \tag{33}$$

This matrix is unitary since its rows and columns are orthonormal. Then its inverse matrix is $\tilde{E}^{\dagger}_{2n-1}$ and the solution of (12) is the following:

$$b_{2n-1}(\tau, \Delta) = \tilde{E}^{\dagger}_{2n-1}\tilde{\vec{\mu}}_{2n-1}. \tag{34}$$

Since we know the form of the matrix (32) explicitly, we can write the solution:

$$b_{2n-1}(\tau, \Delta) = -\frac{2\Delta}{2n-1} \begin{bmatrix} \sum_{j=0}^{n-1} 2^j \sin\left((j+1)\tau\right) \\ \sum_{j=0}^{n-1} 2^j \sin\left((j+1)2\tau\right) \\ \sum_{j=0}^{n-1} 2^j \sin\left((j+1)3\tau\right) \\ \vdots \\ \sum_{j=0}^{n-1} 2^j \sin((j+1)(2n-2)\tau) \\ 0 \end{bmatrix}. \tag{35}$$

Therefore, we have derived the explicit solution to our problem in the case of equidistant eigenvalues and parameter shifts. However, can the system (30) be solved without imposing the latter constraint? The general solution is presented in Appendix C. We have shown that due to the periodicity of the complex exponent, using this solution renders the matrix E_{2n-1} singular in all cases except for equidistant phases.

In the scenario of equidistant eigenvalues, the number of unique distances $\mu_{(t,p)}$ reduces to $2n-1$, and only equidistant parameter shifts, as defined by (31), ensure the nonsingularity of E_{2n-1}. Under these conditions, the system (11) possesses a

unique solution (35). To calculate the derivative of the function, one requires $2n-2$ parameter shifts, where n represents the number of eigenvalues.

B. Equidistant eigenvalues except one

Let us assume that all eigenvalues $\{\lambda_i\}_{i=2}^n$ are equidistant, with the distance between every neighboring one denoted by Δ, such that $\mu(2,3)=1\Delta$, $\mu_{(2,4)}=2\Delta$, and $\mu_{(2,n)}=(n-2)\Delta$. The first eigenvalue is distinct from all the others, where $\mu_{(1,2)}=\Delta_1$, $\mu_{(1,3)}=\Delta_1+\Delta$, $\mu_{(1,4)}=\Delta_1+2\Delta$, and $\mu_{(1,n)}=\Delta_1+(n-2)\Delta$. Then, (25) is reduced to

$$1+2\cos(\Delta_1\Phi_{ij})+2\sum_{k=1}^{n-2}\{\cos(k\Delta\Phi_{ij}) + \cos[(\Delta_1+k\Delta)\Phi_{ij}]\}=0. \tag{36}$$

According to the definition of the Dirichlet and the conjugate Dirichlet kernels, the above expression can be reformulated as

$$[1+\cos(\Delta_1\Phi_{ij})]D_{n-2}(\Delta\Phi_{ij})+\tfrac{1}{2}\cos(\Delta_1\Phi_{ij}) \times \sin(\Delta_1\Phi_{ij})\tilde{D}_{n-2}(\Delta\Phi_{ij}). \tag{37}$$

The general solution to the latter expression can be found in the form $\Delta_1\Phi_{ij}=f(n,\Delta\Phi_{ij})$, where $f(\cdot)$ is a combination of trigonometric functions. One can see that in this case Δ_1 is different for every Φ_{ij}; however, the distance must be the same $\forall i,j$ pairs of phases.

For example, one of the solutions to the latter equation has the following form:

$$\Phi_{ij}=\frac{2\pi t_{ij}}{(2n-3)\Delta},\quad t_{ij}\in\mathbb{Z},$$
$$\Delta_1=\frac{(2n-3)}{2t_{ij}}\left(\cot^{-1}\left(\frac{1}{2}\left(\cos\left(\frac{2\pi t_{ij}}{3-2n}\right)+(-1)^{t_{ij}+1}\right)\times\csc\left(\frac{\pi t_{ij}}{2n-3}\right)\right)+\pi c\right)\Delta,\quad c\in\mathbb{Z}, \tag{38}$$

meaning equidistant phases and the distance of the out eigenvalue scale with $n>2$. However, it is not possible to find all the parameter shifts since Δ_1 is dependent on t_{ij} which is different for every Φ_{ij}. This is contradictory to the fact that Δ_1 must be constant.

We can conclude that in the case of all equidistant eigenvalues except one, the orthogonality condition on the vectors forming the matrix E_m is not fulfilled. That means that the orthogonality property is a specific feature of equidistant eigenvalue systems. This result can be readily extended to the case of multiple eigenvalues that do not form an equidistant set.

C. Perturbed equidistant eigenvalues

An equidistant eigenvalue scenario is a theoretical assumption that does not align with any realistic scenarios. Nevertheless, the eigenvalues can be in proximity to the ideal equidistant positions. The number of parameter shifts remains equal to m, and the exact solution in the well-posed case is given by (13). However, we may question whether it is still possible to utilize the equidistant solution as a good approximation for (13) when the eigenvalues are close to an equidistant configuration. This issue can be addressed using perturbation theory [28].

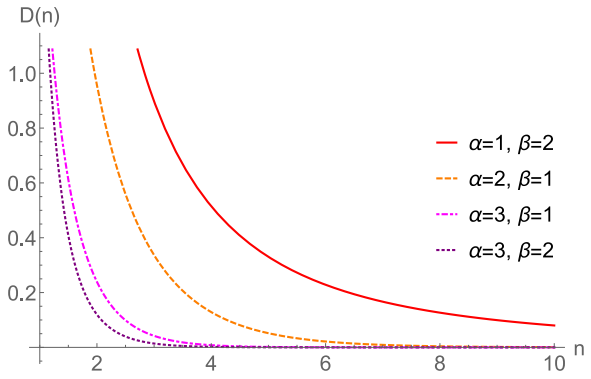

FIG. 2. The distance $D(n)$ defined in (41) from $n\in[1,10]$ for different parameters α and β.

We consider another set of eigenvalues $\{\zeta_i\}_{i=1}^n$, that are not equidistant, such that

$$\zeta_i-\lambda_i=\Delta_i,\quad \forall i\in\overline{1,n}, \tag{39}$$

defining the shift of $\{\zeta_i\}_{i=1}^n$ eigenvalues from the equidistant $\{\lambda_i\}_{i=1}^n$. Then the distances between the new set of eigenvalues are defined as

$$\tilde{\mu}_{(i,j)}=\Delta+\Delta_{i,j}, \tag{40}$$

where we used the notation $\Delta_{i,j}\equiv\Delta_i-\Delta_j$. Similarly to the previous section, we define the $E_m^{(\zeta)}(\vec{\phi})$ matrix with elements given by $E_{(k,l),x}=e^{i\tilde{\mu}_{(k,l)}\phi_x}$, $k,l\in[1,n]$ and the $\mu_m^{(\zeta)}$ vector of differences between the eigenvalues. We can always assume that $\Delta_{i,j}<\epsilon$, $\epsilon\equiv\max_{i,j}(\Delta_{i,j})$, $\forall i,j\in\overline{1,m}$. If $\epsilon\sim\Delta$ the set $\{\zeta_i\}_{i=1}^n$ is highly distant from the equidistant set $\{\lambda_i\}_{i=1}^n$. If $\Delta_{i,j}$ are different $\forall i,j\in[1,m]$, then all the rows of $E_m^{(\zeta)}$ are different and the matrix is nonsingular. If the problem is well posed, we use the results of Sec. III and, in general, we cannot connect it to the solution of the equidistant set $\{\lambda_i\}_{i=1}^n$.

Let us consider the case when $\epsilon<<\Delta$. We can assume that $\{\zeta_i\}_{i=1}^n$ is ϵ perturbed from $\{\lambda_i\}_{i=1}^n$. We use the same technique as in the previous section, to reduce the dimension of the problem: $E_{2n-1}^{(\zeta)}(\vec{\phi})b_{2n-1}^{(\zeta)}(\vec{\phi})=\mu_{2n-1}^{(\zeta)}$, considering the same set of phases $\vec{\phi}$. We approximate $E_{2n-1}^{(\zeta)}(\vec{\phi})$ with a nonsingular matrix $\hat{E}_{2n-1}=E_{2n-1}+\epsilon R_{2n-1}$, that is, an ϵ perturbation of E_{2n-1}. Selecting $\epsilon\sim O(\frac{1}{2^{\alpha n}n^{\beta}})$, $\alpha,\beta\geqslant 1$, the distance between the solution $b_{2n-1}^{(\zeta)}(\vec{\phi})$ for the $\{\zeta_i\}_{i=1}^n$ system and the solution (35) for equidistant $\{\lambda_i\}_{i=1}^n$ is (see the detailed analyses in Appendix D)

$$D(n)\equiv\|b^{(\zeta)}-b_{2n-1}(\tau,\Delta)\|_2\sim O\left(\frac{1}{2^{5\alpha n}n^{5\beta-1}}\right). \tag{41}$$

In Fig. 2, the distance (41) between the equidistant solution and the perturbed one is shown from $n\in[1,10]$. We assume that $\Delta=1$. For different selections of parameters α and β, the decay is slightly changing. However, for any n, one can select the minimal numbers to maximize ϵ, which increases the robustness of the method. For example, for $n=3$, we can select $\alpha=3$, $\beta=1$, to get for $\epsilon\approx 6.5\times10^{-4}$ the distance

(41) equal to 4.2×10^{-2}. For $n = 5$ we can get the distance 5.1×10^{-2} for $\epsilon \approx 1.9 \times 10^{-4}$ selecting $\alpha = 2$ and $\beta = 1$. For $n = 10$ the distance 2.4×10^{-2} can be achieved only for small perturbations $\epsilon \approx \times 3 \times 10^{-6}$. We observe that with the increase in the dimensionality of the system, the size of the perturbation that can be tolerated to use the solution for the unperturbed system decreases. This is an expected result, as the number of differences in eigenvalues increases with the dimensionality growth. We see that the choice of solution according to the structure of the system's eigenvalues significantly influences the estimation result. It is important to consider the eigenvalue structure and take into account the possible statistical errors in the estimation of eigenvalues, even if the system is theoretically equidistant.

D. Eigenvalues forming equidistant sets

In the case of having L realizations of the Hamiltonian, the sets of eigenvalues $(\{\lambda_i\}_{i=1}^n)_L$ can be regarded as perturbations of each other. Let us postulate that it is possible to organize all eigenvalues from the L realizations into n equidistant sets, with median values denoted as Λ_i, where $i \in [1, n]$. The distance between the median values of any two neighboring clusters is given by

$$\mu_{(i,i+1)} \equiv |\Lambda_i - \Lambda_{i+1}| \approx \Delta, \quad \forall i \in \overline{1, n}. \tag{42}$$

We demand the width of every set to be $\varepsilon_i << \Delta, \forall i \in [1, n]$. Let us consider the median eigenvalues. For the set of equidistant Λ_i, $i = [1, n]$ the solution of the problem (11) is given by (35).

We sort the eigenvalues into equidistant groups in such a way that from every set, only one eigenvalue is picked. The eigenvalues in every group are denoted as $\tilde{\Lambda}_{l,i}$, $\forall l \in \overline{1, L}$, $i \in [1, n]$, where the first index is the number of the group and the second is the number of the eigenvalue. The distance between any eigenvalue in one set and its median is

$$|\tilde{\Lambda}_{l,i} - \Lambda_i| = \Delta_{l,i} < \varepsilon_i. \tag{43}$$

Then the collection of eigenvalues $\{\tilde{\Lambda}_{l,i}\}_{i=1}^n$ is slightly shifted from the central set $\{\Lambda_i\}_{i=1}^n$, but not more than the width ε_i according to (43).

First, we consider the case when we picked the shift from the median collection of eigenvalues in such a way that the new collection $\{\tilde{\Lambda}_{l,i}\}_{i=1}^n$ is equidistant too. That means that

$$|\tilde{\Lambda}_{l,i} - \tilde{\Lambda}_{l,i-1}| = \tilde{\Delta}_l, \quad \forall i \in [1, n]. \tag{44}$$

For these n eigenvalues, the solution of the reduced problem (11) is $b_{2n-1}(\tau, \tilde{\Delta}_l)$ and is given by (35). If $\Delta = \tilde{\Delta}$ the solutions from the median set and from the shifted set are coincident.

However, if we sort the real data, we see that the eigenvalues are slightly not equidistant, corresponding to the case of perturbed equidistant eigenvalues, considered in the previous section. One has to do the same analysis, taking into account the amount of perturbation from the equidistant positions, corresponding to L realizations.

V. GENERAL SOLUTION OF THE ILL POSED PSR PROBLEM

In this section, we solve the problem (11) being ill posed by Hadamard. As we mentioned, it can happen for multiple reasons. First, the eigenvalues of the Hamiltonian can be close to each other, such that different $\mu_{(k,p)}$ would be equal. This leads to singularity in the matrix E_m (hereafter, we omit the m index, assuming all matrices in this section are of size $m \times m$). Secondly, we explore the solution to (11) in the scenario where the operators E and the functions μ are not precisely known, but rather approximations $\hat{E}_l$ and $\hat{\mu}_l$ are available. In this context, the index $l \in \overline{1, L}$, where $L > 0$, denotes the realization number. The approximates $\hat{E}_l$ and $\hat{\mu}_l$ are defined on a probability space $(\Omega, \mathcal{A}, P)$ and are close to E and μ in some probabilistic sense. Here, $\hat{\mu}_l \in V$, and the operator $\hat{E}_l$ is continuous for all $\omega \in \Omega$.

Due to the sensitivity of $\hat{b}(\vec{\phi}) = \hat{E}_l^{-1}(\vec{\phi})\hat{\mu}_l$ to fluctuations in empirical data, it cannot serve as a reliable approximation of $b(\vec{\phi})$. To elaborate further, minor deviations in the values of $\hat{\mu}_l$ from μ have the capacity to lead to substantial variations in $\hat{b}$. This indicates that the inverse operator $\hat{E}_l^{-1}$ may not exhibit continuity, rendering the problem ill posed.

In our specific case, μ is a $(m \times 1)$ vector. If E is an $m \times m$ matrix and $\det E \neq 0$ [or rank $(E) = m$] then E^{-1} exists. However, the problem can still be ill posed. One can define an orthogonal transformation $b = Vb^\star$ and $\mu = V\mu^\star$ such that E will be represented in a diagonal form $(l_1, \ldots, l_m)$, where $\{l_i\}_{i=1}^m$ are the eigenvalues of E. When some differences $\mu_{(k,p)}$ between the eigenvalues of the Hamiltonian are equal, the rank $(E) = r < m$ and then the $m - r$ eigenvalues l_i of the matrix E are zero. Then the matrix is not invertible. Let $l_i = 0$, $i = \overline{1, r}$ and $l_i \neq 0$ for $i \in \overline{r+1, m}$. For given approximations $\hat{E}_l$ and $\hat{\mu}_l$ such that

$$\begin{aligned} &\|\hat{E}_l - E\| \leqslant \xi, \quad \xi > 0, \\ &\|\hat{\mu}_l - \mu\| \leqslant \delta, \quad \delta > 0, \end{aligned} \tag{45}$$

the eigenvalues $\tilde{l}_i$, $i \in \overline{r+1, m}$ of $\hat{E}_l$ may be close to zero for a sufficiently small ξ. Then $\tilde{b}_i^\star = \tilde{\lambda}_i^\star / \tilde{l}_i$ may be large for a small perturbation of $\hat{E}_l$ and $\hat{\mu}_l$. This implies that the solution of the system of linear equations (11) is unstable.

In this case, we use the regularization technique introduced in [27] that entails the stabilization of solutions by limiting the set of feasible solutions $\mathcal{D} \in U$ to a compact set $\mathcal{D}^\star$, due to the subsequent lemma.

Lemma V.1. The inverse operator E^{-1} is continuous on the set $N^\star = E\mathcal{D}^\star$ if the continuous one-to-one operator E is defined on the compact $\mathcal{D}^\star \in \mathcal{D} \subseteq U$.

The reduction of solutions is facilitated by the stabilizing functional, defined on $\mathcal{D}$. It is worth noting that the regularization method bears similarity to the Lagrange method, as we seek a solution $\hat{b}$ that minimizes a functional $\Omega(\hat{b})$: $|\hat{E}_l\hat{b} - \hat{\mu}_l| \leqslant \xi$, where $\xi > 0$.

In this paper, to find the solution of (11) we propose to use the extension of the regularization method from a deterministic operator equation to the case of stochastic ill-posed problems. The function that minimizes the functional

$$R_\gamma(\hat{\mu}_l, b) = \|\hat{E}_l b - \hat{\mu}_l\|_V^2 + \gamma\Omega(b) \tag{46}$$

in a set $\mathcal{D}$ of functions $b \in U$ is taken as an approximate solution of (11). The parameter $\gamma > 0$ is called the regularization parameter and $\Omega(b)$ is a stabilizing functional that satisfies the following conditions.

(1) $\Omega(b)$ is defined on the set $\mathcal{D}$.

(2) $\Omega(b)$ assumes real non-negative values and is lower semicontinuous on $\mathcal{D}$.

(3) All sets $M_c = \{b : \Omega(b) \leqslant c\}$ are compact in U.

Theorems V.1 and V.2 [29,30] provide the theoretical background of the statistical regularization method for the case of an accurately given operator E, and Theorem V.3 [31] provide the background for the case of an inaccurately given operator E.

Theorem V.1. If, for each l, a positive $\gamma = \gamma(l)$ is chosen such that $\gamma \to 0$ as $l \to \infty$, then for any positive α and β there will be a number $N = N(\alpha, \beta)$ such that, for all $l > N$, the elements $\hat{b}^\gamma$ that minimize the functional (46) satisfy the inequality

$$P\{\rho_U(\hat{b}^\gamma, b) > \alpha\} \leqslant P\{\rho_V^2(\hat{\mu}_l, \mu) > \beta\gamma\}, \tag{47}$$

where b is the precise solution of (11) with the right-hand side μ, and $\rho(f, g) = \|f - g\|$.

For our concrete case, all spaces are Hilbert ones.

Theorem V.2. Let U be a Hilbert space, let E be a linear operator, and let $\Omega(b) = \|b\|_U^2$. Then, $\forall\xi$, there exists a number $l(\xi)$ such that $\forall k > k(\xi)$, and the inequality

$$P\{\|\hat{b}^\gamma - b\|_U^2 > \xi\} \leqslant 2P\{\rho_V^2(\hat{\mu}_l, \mu) > (\xi/2)\gamma\} \tag{48}$$

holds.

Theorem V.3. Let U and V be normed spaces. For any $\xi > 0$ and any constants $c_1, c_2 > 0$, there exists a number $\gamma_0 > 0$ such that $\forall\gamma \geqslant \gamma_0$:

$$P\{\omega : \|\hat{b}^\gamma - b\|_U > \xi\} \leqslant P\left\{\omega : \frac{\|\hat{\mu}_l - \mu\|_V}{\sqrt{\gamma}} > c_1\right\} + P\left\{\omega : \frac{\|\hat{E}_l - E\|}{\sqrt{\gamma}} > c_2\right\}, \tag{49}$$

where

$$\|\hat{E}_l - E\| = \sup_{g\in\mathcal{D}} \frac{\|\hat{E}_l b - Eb\|_V}{\sqrt{\Omega(b)}}. \tag{50}$$

These theorems imply that the minimization of (46) is a stable problem, i.e., close functions $\hat{\mu}_l$ and μ (and close operators $\hat{E}_l$ and E) correspond to close (in a probabilistic sense) regularized solutions $\hat{b}^\gamma$ and b that minimize the functionals $R_\gamma(\hat{\mu}_l, b)$ and $R_\gamma(\mu, b)$, respectively.

For the Hilbert spaces U and V, the solution of (11) with $\Omega(b) = \|b\|_U^2$ has a simple form

$$\hat{b}^\gamma = (\gamma I + \hat{E}_l^\dagger \hat{E}_l)^{-1} \hat{E}_l^\dagger \hat{\mu}_l, \tag{51}$$

where I is an identity operator.

The stability of the approximation $\hat{b}^\gamma$ to b is ensured by an appropriate choice of γ. For selection of the regularization parameter see [32–34]. For example, the mismatch method [32] determines γ from the equality

$$\|\hat{E}_l \hat{b}^\gamma - \hat{\mu}_l\|_V = \xi(l) + \eta(l, b),$$
$$\|\hat{\mu}_l - \mu\|_V \leqslant \xi(l), \quad \|\hat{E}_l b - Eb\|_V \leqslant \eta(l, b), \tag{52}$$

where $\xi(l)$ and $\eta(l, b)$ are known estimates of the data error. The stochastic analog of the mismatch method is the discrepancy method [35,36]. If the operator is defined precisely ($\eta(l, b) = 0$), then the choice of γ from (52) provides a rate of convergence of the regularized estimate $\hat{b}^\gamma$ to b that is no better than $O(\xi^{1/2})$ (see [33]).

Using the form of the regularized solution (51), we can write

$$\hat{b}_x^\gamma(\vec{\phi}) = \sum_{j,i=1}^{m} (\gamma I + \hat{E}^\dagger(\vec{\phi})\hat{E}(\vec{\phi}))^{-1}_{xj} \hat{E}^\dagger_{ji}(\vec{\phi})\hat{\mu}_i, \tag{53}$$

an approximate solution of (11). Here we omit the index l, meaning we treat one experimental realization of $\hat{E}$ and $\hat{\mu}$.

Similarly to Sec. III, we minimize the variance

$$\sigma^2\left(\widehat{\frac{\partial f(t, \vec{\phi})}{\partial t}}\right) = \sigma^2[\hat{f}(t)] \sum_{x=1}^{m} (\hat{b}_x^{(\gamma)}(\vec{\phi}))^2, \tag{54}$$

solving the problem $\min_{\vec{\phi}} \sum_{x=1}^{m} (\hat{b}_x^{(\gamma)}(\vec{\phi}))^2$. Using the expression (53) for the regularized solution, we get the system of equations

$$\sum_{x=1}^{m}\sum_{j,i=1}^{m} (\gamma I + \hat{E}^\dagger\hat{E})^{-1}_{xj} \hat{E}^\dagger_{ji}\lambda_i \sum_{l=1}^{m} (\gamma I + \hat{E}^\dagger\hat{E})^{-1}_{xl} \times \sum_{s=1}^{m} \left(\left(\frac{\partial \hat{E}^\dagger}{\partial \phi_y}\right)_{ls} \delta_{l,y} - \sum_{p,v,t=1}^{m} \left(\left(\frac{\partial \hat{E}^\dagger}{\partial \phi_y}\right)_{lv} \hat{E}_{vp}\delta_{l,y} + \hat{E}^\dagger_{lv}\left(\frac{\partial \hat{E}}{\partial \phi_y}\right)_{vp} \delta_{p,y}\right)(\gamma I + \hat{E}^\dagger\hat{E})^{-1}_{pt} \hat{E}^\dagger_{ts}\right)\lambda_s = \mathbf{0}. \tag{55}$$

Solving this system of equations with respect to all ϕ_y, $y \in \overline{1, m}$, one can find the optimal $\vec{\phi}$ minimizing the variance (54).

VI. DISCUSSION AND CONCLUSION

We propose a method for optimal parameter selection in parameter shift rules, which takes into account the eigenvalue structure of the dynamical generator. Our method is suitable for big Hamiltonian systems with known eigenvalues. Dependent on the distance between the eigenvalues and the precision of the eigenvalue estimates, the problem can be well or ill posed by Hadamard, which makes it nontrivial for optimization in the case when some distances are close to each other.

In the case of a well-posed problem, an explicit solution is proposed, and the recipe for finding the optimal phases is provided. For the ill-posed problem arising, for example, when the eigenvalues of the Hamiltonian are close to each other and the distances between them can coincide, we find the explicit solution as well. We show that it is unique and that the phases must be picked equidistantly. We observe the realistic case of slightly perturbed equidistant eigenvalues arising in practice and the case of equidistant clusters formed by the different realizations of the Hamiltonian. We provide the regularized solution for the ill-posed problem that does not have a particular eigenvalue structure.

In addition to a full reconstruction of the derivative, the presented approach offers parameter shift rules for derivatives of arbitrary order and any linear combination of them.

ACKNOWLEDGMENTS

L.M. was supported by the Netherlands Organisation for Scientific Research, as part of the Quantum Software Consortium program (Project No. 024.003.037/3368). This work has received support from the European Union's Horizon Europe research and innovation program through the European Research Council StG FINE-TEA-SQUAD (Grant No. 101040729). This work is supported by the Dutch National Growth Fund, as part of the Quantum Delta NL programme. This research work was partly supported by the Roadmap for the Development of Quantum 400 Technologies in Russian Federation, Contract No. 868-1.3-15/15-2021.

APPENDIX A: OVERVIEW OF KNOWN PARAMETER SHIFT RULES

In [21], the latter rule is generalized to gates with eigenvalues $\{-1, 0, 1\}$, resulting in $m = 2$ frequencies:

$$\left.\frac{\partial f(t)}{\partial t}\right|_{t=0} = y_1(f(\phi_1) - f(-\phi_1)) - y_2(f(\phi_2) - f(-\phi_2)), \tag{A1}$$

where $\phi_1, \phi_2 \in (0, \pi)$, and $y_{1,2}$ are the corresponding coefficients. In [21] $\phi_{1,2} = \pi/2 \mp \pi/4$ and $y_{1,2} = (\sqrt{2} \pm 1)/2\sqrt{2}$ are studied. On the other hand, in [37], for the same eigenvalues, the following rule is introduced:

$$\left.\frac{\partial f(t)}{\partial t}\right|_{t=0} = \frac{1}{4}(f_+^+ - f_-^+ + f_+^- - f_-^-), \tag{A2}$$

where $f_\pm^\alpha$ is the measured energy when replacing the gate $U(t)$ in question by $U(t \pm \pi/2)\exp(\mp\alpha i\pi/4\Pi_0)$, and Π_0 is the projector onto the zero eigenspace of the generator of U. In the context of the perturbed quantum evolution denoted as $U_F(t) = \exp[i(tH + F)]$, the stochastic parameter shift rule is presented in [23] as follows:

$$\left.\frac{\partial f(t)}{\partial t}\right|_{t=\phi_0} = \frac{\Omega}{2\sin(\Omega\phi_1)}\int_0^1 (f_+(r) - f_-(r))dr. \tag{A3}$$

Here, $f_\pm(t)$ represents the measured energy in the state prepared by a modified circuit that divides $U_F(r_0)$ into $U_F(r\phi_0)$ and $U_F[(1-r)\phi_0]$, interleaving these two gates with $U_{F=0}(\pm\phi_0)$. These results were further developed to introduce the Nyquist shift rule in [38]. Additionally, a PSR for higher-order derivatives, which is based on iteratively applying the original rule, has been proposed in [20].

APPENDIX B: HAMILTONIAN WITH FINITE NUMBER OF EIGENVALUES

Suppose a Hamiltonian H has n distinct eigenvalues $\vec{\lambda} = \{\lambda_i\}_{i=1}^n$. Then, one can write the characteristic polynomial of its matrix as

$$p(H) = \prod_{i=1}^{n}(H - \lambda_i\mathbb{I}) = 0. \tag{B1}$$

This means that there exist only $n-1$ nontrivial powers of H, since one can write

$$H^n = -\sum_{i=1}^{n-1}(-1)^i S_i(\vec{\lambda})H^{n-i}, \tag{B2}$$

where the term $S_i(\vec{\lambda})$ is the symmetric polynomial of degree i that contains all the possible combinations of elements of $\vec{\lambda}$ (of degree 1). For example, if $n = 3$ and $i = 2$:

$$S_2 = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3. \tag{B3}$$

The fact that there are only finite number of powers means that any function of H can be written as follows:

$$f(H) = \sum_{i=0}^{n-1} c_i H^i := \langle H\rangle\, c. \tag{B4}$$

To determine the form of the vector c_i, one can project (B4) onto the eigenstates of H. Schematically, we can write it as follows:

$$|f(\vec{\lambda})\rangle = \Lambda(\vec{\lambda})\,|c\rangle \Rightarrow |c\rangle = \Lambda^{-1}(\vec{\lambda})\,|f(\vec{\lambda})\rangle\,. \tag{B5}$$

In the above expression, the vector $|f(\vec{\lambda})\rangle$ contains the $f(\lambda_i), \forall\lambda_i \in \vec{\lambda}$ and Λ is the $n \times n$ Vandermonde matrix containing the $\vec{\lambda}$. The explicit form of this square matrix is

$$(\Lambda^{-1})_{ij} = (-1)^{i+1}\frac{S_{n-1-i}(\vec{\lambda}_j)}{\prod_{k\in\vec{\lambda}_j}(k - \lambda_j)}. \tag{B6}$$

This is the element that belongs to ith row and jth column. The $\vec{\lambda}_j$ denotes the set of $\vec{\lambda}$ that does not include the single element λ_j. Then the coefficients in (B4) are the following:

$$c_i(f) = (-1)^{i+1}\sum_{j=1}^{n}\frac{S_{n-1-i}(\vec{\lambda}_j)}{\prod_{k\in\vec{\lambda}_j}(k - \lambda_j)}f(\lambda_j). \tag{B7}$$

Let us consider the function $f(H) = \exp(-\mathbb{i}Ht)$. Based on the previous results, the coefficients in the decomposition (B4) are the following:

$$c_p(e) = (-1)^{p+1}\sum_{j=1}^{n}\frac{S_{n-1-p}(\vec{\lambda}_j)}{\prod_{k\in\vec{\lambda}_j}(k - \lambda_j)}e^{-\mathbb{i}\lambda_j t}. \tag{B8}$$

This implies that the coefficients of the derivative of $f(H)$ with respect to t can be written in a closed form as follows:

$$\tilde{c}_p(e) = (-1)^{p+1}\sum_{j=1}^{n}\frac{S_{n-1-p}(\vec{\lambda}_j)\lambda_j}{\prod_{k\in\vec{\lambda}_j}(k - \lambda_j)}e^{-\mathbb{i}(\lambda_j t + \frac{\pi}{2})}. \tag{B9}$$

APPENDIX C: NONEQUIDISTANT PHASES

We want to solve the system of equations

$$\phi_i - \phi_j = \frac{\tau}{\Delta}t_{ij}, \quad t_{ij} \in Z, \quad \forall i, j = [1, 2n-1], \quad i < j. \tag{C1}$$

Let us observe an example of $n = 3$. Then $m = 5$ and we can always select $\phi_5 = 0$ without loss of generality. From (C1) we

can immediately conclude

$$\phi_1 = \frac{\tau}{\Delta} t_{15}, \quad \phi_2 = \frac{\tau}{\Delta}(t_{15} + c_1\epsilon),$$
$$\phi_3 = \frac{\tau}{\Delta}(t_{15} + c_2\epsilon), \quad \phi_4 = \frac{\tau}{\Delta}(t_{15} + c_3\epsilon). \tag{C2}$$

However, we can also see that

$$\phi_2 = \frac{\tau}{\Delta} t_{25}, \quad \phi_3 = \frac{\tau}{\Delta}(t_{25} + b_1\epsilon), \quad \phi_4 = \frac{\tau}{\Delta}(t_{25} + b_2\epsilon) \tag{C3}$$

and

$$\phi_3 = \frac{\tau}{\Delta} t_{35}, \quad \phi_4 = \frac{\tau}{\Delta}(t_{35} + a_1\epsilon), \quad \phi_4 = \frac{\tau}{\Delta} t_{45} \tag{C4}$$

hold. Here we denoted $t_{1,2} = c_1\epsilon, t_{1,3} = c_2\epsilon, t_{1,4} = c_3\epsilon, t_{2,3} = b_1\epsilon$, $t_{2,4} = b_2\epsilon$, $t_{3,4} = a_1\epsilon$, and $\epsilon > 0$ is a constant. The constants can be easily found:

$$c_1 = \frac{1}{\epsilon}(t_{25} - t_{15}), \quad c_2 = \frac{1}{\epsilon}(t_{35} - t_{15}),$$
$$c_3 = \frac{1}{\epsilon}(t_{45} - t_{15}), \quad b_1 = \frac{1}{\epsilon}(t_{35} - t_{25}),$$
$$b_2 = \frac{1}{\epsilon}(t_{45} - t_{25}), \quad a_1 = \frac{1}{\epsilon}(t_{45} - t_{35}). \tag{C5}$$

Selecting any $t_{15} < t_{25} < t_{35} < t_{45}$, one gets the following solutions:

$$\phi_1 = \frac{t}{\Delta} t_{15}, \quad \phi_2 = \frac{t}{\Delta} t_{25}, \quad \phi_3 = \frac{t}{\Delta} t_{35},$$
$$\phi_4 = \frac{t}{\Delta} t_{45}, \quad \phi_5 = 0. \tag{C6}$$

One can see that Eq. (C1) is solvable. The solution is

$$\phi_i = \frac{t}{\Delta} t_{i,2n-1}, \quad \phi_{2n-1} = 0, \quad \forall i \in [1, 2n-2]. \tag{C7}$$

However, due to periodicity of the complex exponent, we get singular matrix $E_{2n-1}(\vec{\phi})$ for all choices of (C7) except the equidistant phases one. The latter example can be easily generalized to any value of n.

APPENDIX D: PERTURBED EQUIDISTANT EIGENVALUES

Let $\{\lambda_i\}_{i=1}^n$ be equidistant eigenvalues with $\mu_{ij} = \Delta$. We consider another set of eigenvalues $\{\zeta_i\}_{i=1}^n$, that are not equidistant, such that

$$\zeta_i - \lambda_i = \Delta_i, \quad \forall i \in \overline{1, n}. \tag{D1}$$

Here Δ_i defines the shift of the new set of eigenvalues from the equidistant one. Then the distances between the new set of eigenvalues are defined as

$$\tilde{\mu}_{(i,j)} = \zeta_i - \zeta_j = \lambda_i - \lambda_j + \Delta_i - \Delta_j = \Delta + \Delta_{i,j}, \tag{D2}$$

where we used the notation $\Delta_{i,j} \equiv \Delta_i - \Delta_j$. We can define the matrix:

$$E_m^{(\zeta)}(\vec{\phi}) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{\mathbb{i}\Delta\phi_1} e^{\mathbb{i}\Delta_{1,2}\phi_1} & e^{\mathbb{i}\Delta\phi_2} e^{\mathbb{i}\Delta_{1,2}\phi_2} & \dots & e^{\mathbb{i}\Delta\phi_m} e^{\mathbb{i}\Delta_{1,2}\phi_m} \\ e^{-\mathbb{i}\Delta\phi_1} e^{-\mathbb{i}\Delta_{1,2}\phi_1} & e^{-\mathbb{i}\Delta\phi_2} e^{-\mathbb{i}\Delta_{1,2}\phi_2} & . & e^{-\mathbb{i}\Delta\phi_m} e^{-\mathbb{i}\Delta_{1,2}\phi_m} \\ \vdots & \vdots & \ddots & \vdots \\ e^{(n-1)\mathbb{i}\Delta\phi_1} e^{\mathbb{i}\Delta_{n,n-1}\phi_1} & e^{(n-1)\mathbb{i}\Delta\phi_2} e^{\mathbb{i}\Delta_{n,n-1}\phi_2} & \dots & e^{(n-1)\mathbb{i}\Delta\phi_m} e^{\mathbb{i}\Delta_{n,n-1}\phi_m} \\ e^{-(n-1)\mathbb{i}\Delta\phi_1} e^{-\mathbb{i}\Delta_{n,n-1}\phi_1} & e^{-(n-1)\mathbb{i}\Delta\phi_2} e^{-\mathbb{i}\Delta_{n,n-1}\phi_2} & \dots & e^{-(n-1)\mathbb{i}\Delta\phi_m} e^{-\mathbb{i}\Delta_{n,n-1}\phi_m} \end{bmatrix}. \tag{D3}$$

The vector of distances is defined as follows:

$$\mu_m^{(\zeta)} = \mu_m^{(\lambda)} + \mathbb{i} \begin{bmatrix} 0 \\ \Delta_{1,2} \\ -\Delta_{1,2} \\ \vdots \\ -\Delta_{n,n-1} \end{bmatrix} \equiv \mu_m^{(\lambda)} + \Delta_m. \tag{D4}$$

We can always assume that $\Delta_{i,j} < \epsilon$, $\epsilon \equiv \max_{i,j}(\Delta_{i,j})$, $\forall i, j \in \overline{1, m}$. If $\epsilon \sim \Delta$ the set $\{\zeta_i\}_{i=1}^n$ is highly distant from the equidistant set $\{\lambda_i\}_{i=1}^n$. If $\Delta_{i,j}$ are different $\forall i, j \in \overline{1, m}$, then all the rows of $E_m^{(\zeta)}$ are different and the matrix is nonsingular. If the problem is well posed, according to (12) the solution for the system $\{\zeta_i\}_{i=1}^n$ is

$$b_m^{(\zeta)}(\vec{\phi}) = \left(E_m^{(\zeta)}\right)^{-1}(\vec{\phi})\mu_m^{(\zeta)} \tag{D5}$$

and, in general, we cannot connect it to the solution of the equidistant set $\{\lambda_i\}_{i=1}^n$.

Let us consider the case when $\epsilon \ll \Delta$. In this case, we can assume that $\{\zeta_i\}_{i=1}^n$ is ϵ perturbed from $\{\lambda_i\}_{i=1}^n$.

Using the small angle approximation, the truncation error for the exponential function in the complex plane is

$$|\exp(\mathbb{i}\Delta_{i,j}\phi_k) - (1 + \mathbb{i}\Delta_{i,j}\phi_k)| \leqslant \frac{\Delta_{i,j}^2\phi_k^2}{2}. \tag{D6}$$

Then, one can write

$$|\exp(\mathbb{i}\Delta_{i,j}\phi_k) - (1 + \mathbb{i}\epsilon\phi_k)|$$
$$\leqslant |\exp(\mathbb{i}\epsilon\phi_k) - (1 + \mathbb{i}\epsilon\phi_k)| \leqslant \frac{\epsilon^2\phi_k^2}{2}. \tag{D7}$$

We can conclude that for sufficiently small ϵ the matrix $E_m^{(\zeta)}$ can be singular. The singularity arises due to the closeness of the distances to each other. So we reduce $E_m^{(\zeta)}$ to

$E^{(\zeta)}_{2n-1}$ (that is, nonsingular) in the same manner as we did for $E^{(\lambda)}_m$ corresponding to the perfectly equidistant set of eigenvalues and take the upper bound for all Δ_{ij} to be ϵ:

$$E^{(\zeta)}_{2n-1}(\vec{\phi}) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{\mathbb{i}\Delta\phi_1}e^{\mathbb{i}\epsilon\phi_1} & e^{\mathbb{i}\Delta\phi_2}e^{\mathbb{i}\epsilon\phi_2} & \dots & e^{\mathbb{i}\Delta\phi_m}e^{\mathbb{i}\epsilon\phi_m} \\ e^{-\mathbb{i}\Delta\phi_1}e^{-\mathbb{i}\epsilon\phi_1} & e^{-\mathbb{i}\Delta\phi_2}e^{-\mathbb{i}\epsilon\phi_2} & \cdot & e^{-\mathbb{i}\Delta\phi_m}e^{-\mathbb{i}\epsilon\phi_m} \\ \vdots & \vdots & \ddots & \vdots \\ e^{(n-1)\mathbb{i}\Delta\phi_1}e^{\mathbb{i}\epsilon\phi_1} & e^{(n-1)\mathbb{i}\Delta\phi_2}e^{\mathbb{i}\epsilon\phi_2} & \dots & e^{(n-1)\mathbb{i}\Delta\phi_m}e^{\mathbb{i}\epsilon\phi_m} \\ e^{-(n-1)\mathbb{i}\Delta\phi_1}e^{-\mathbb{i}\epsilon\phi_1} & e^{-(n-1)\mathbb{i}\Delta\phi_2}e^{-\mathbb{i}\epsilon\phi_2} & \dots & e^{-(n-1)\mathbb{i}\Delta\phi_m}e^{-\mathbb{i}\epsilon\phi_m} \end{bmatrix}. \tag{D8}$$

Let us define a shifted nonsingular matrix $\hat{E}_{2n-1}$, where we assume equidistant phases $\phi_j = -\frac{\tau j}{\Delta}$, $j \in [1, 2n-1]$ defined in (31):

$$\hat{E}_{2n-1} \equiv E_{2n-1} + \epsilon R_{2n-1}, \tag{D9}$$

$$R_{2n-1} = \mathbb{i}\begin{bmatrix} 0 & 0 & \dots & 0 \\ \phi_1 e^{\mathbb{i}\Delta\phi_1} & \phi_2 e^{\mathbb{i}\Delta\phi_1} & \dots & \phi_{2n-1}e^{\mathbb{i}\Delta\phi_{2n-1}} \\ \phi_1 e^{-\mathbb{i}\Delta\phi_1} & \phi_2 e^{-\mathbb{i}\Delta\phi_1} & \cdot & \phi_{2n-1}e^{-\mathbb{i}\Delta\phi_{2n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1 e^{-\mathbb{i}(n-1)\Delta\phi_1} & \phi_2 e^{-\mathbb{i}(n-1)\Delta\phi_1} & \dots & \phi_{2n-1}e^{\mathbb{i}(n-1)\Delta\phi_{2n-1}} \end{bmatrix}, \quad r_{2n-1} = \mathbb{i}\begin{bmatrix} 0 \\ 1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}.$$

Then the perturbation of the equidistant problem $\tilde{E}^{(\lambda)}_{2n-1} b_{2n-1}(0) = \tilde{\mu}^{(\lambda)}_{2n-1}$ is the following:

$$\begin{aligned} \left(\tilde{E}^{(\lambda)}_{2n-1} + \epsilon\tilde{R}_{2n-1}\right) b_{2n-1}(\epsilon) &= \left(\tilde{\mu}^{(\lambda)}_{2n-1} + \epsilon\tilde{r}\right), \\ \tilde{R}_{2n-1} &\equiv \frac{R_{2n-1}}{\sqrt{2n-1}}, \\ \tilde{r}_{2n-1} &\equiv \frac{r_{2n-1}}{\sqrt{2n-1}}, \end{aligned} \tag{D10}$$

and the matrices $\tilde{E}^{(\lambda)}_{2n-1}$, $\vec{\tilde{\mu}}^{(\lambda)}_{2n-1}$ are defined in (33). Further we omit the indices $2n-1$ and (λ) for briefness.

Differentiating by ϵ (we suppose that this derivative exists), one can derive

$$\tilde{E}\dot{b}(\epsilon) + \tilde{R}b(\epsilon) + \epsilon\tilde{R}\dot{b}(\epsilon) = \tilde{r}. \tag{D11}$$

For $\epsilon = 0$ we get the following expression:

$$\tilde{E}\dot{b}(0) + \tilde{R}b(0) = \tilde{r} \longrightarrow \dot{b}(0) = \tilde{E}^{-1}(\tilde{r} - \tilde{R}b(0)), \tag{D12}$$

where $b(0)$ is the solution (35) of the nonperturbed equidistant problem. Using the Taylor expansion

$$b(\varepsilon) = b(0) + \varepsilon\dot{b}(0) + o(\varepsilon), \tag{D13}$$

we can write

$$\frac{\|b(\varepsilon) - b(0)\|}{\|b(0)\|} \leqslant k(\tilde{E})\left(\frac{\|\epsilon\tilde{r}\|}{\|\tilde{\mu}\|} + \frac{\|\epsilon\tilde{R}\|}{\|\tilde{E}\|}\right) + o(\epsilon), \tag{D14}$$

where $k(E) \equiv \|\tilde{E}^{-1}\|\|\tilde{E}\| \geqslant 1$ is the condition number. In the case of equidistant eigenvalues and phases the matrix $\tilde{E}$ is unitary and its norm is equal to 1. Then $k(\tilde{E}) = 1$ holds.

Finally, we can conclude

$$\|b(\varepsilon) - b(0)\| \approx \epsilon(\|\tilde{r}\| + \|\tilde{R}\|)\|b(0)\|. \tag{D15}$$

Let us calculate the latter distance for the case of the l_2- norm. It is known that

$$\|\tilde{R}\|_2 \leqslant \|\tilde{R}\|_F = \left(\sum_{i,j=1}^{2n-1} |\tilde{R}_{ij}|^2\right)^{\frac{1}{2}}, \tag{D16}$$

where $\|\tilde{R}\|_F$ is the Frobenius norm. Since we know the form of the matrix, we can immediately write

$$\begin{aligned} \|\tilde{R}\|_F &= \left(\sum_{j=1}^{2n-1} |\phi_j|^2\right)^{\frac{1}{2}} = \frac{\tau}{\Delta}\left(\sum_{j=1}^{2n-1} j^2\right)^{\frac{1}{2}} \\ &= \frac{2\pi}{\Delta}\frac{\sqrt{n(4n-1)}}{\sqrt{3(2n-1)}}. \end{aligned} \tag{D17}$$

Moreover, we can write

$$\|\tilde{r}\|_2 = \sqrt{\sum_{i=1}^{2n-1} |\tilde{r}_i|^2} = 1. \tag{D18}$$

The l_2 norm of (35) is

$$\|b(0)\|_2 = \frac{2\Delta}{2n-1}\sqrt{\sum_{k=1}^{2n-2}\left(\sum_{j=0}^{n-1} 2^j \sin\left((j+1)kt\right)\right)^2}. \tag{D19}$$

We can upper bound it as

$$\|b(0)\|_2 \leqslant \frac{2(2^n - 1)\Delta}{\sqrt{2n-1}}. \tag{D20}$$

Then (D15) can be bounded by

$$\|b(\varepsilon) - b(0)\|_2 \leqslant \epsilon\left(1 + \frac{2\pi}{\Delta}\frac{\sqrt{n(4n-1)}}{\sqrt{3(2n-1)}}\right)\frac{2(2^n-1)\Delta}{\sqrt{2n-1}}. \tag{D21}$$

Now we know the deviation of the solution $b(0)$ for the equidistant problem $\{\lambda_i\}_{i=1}^n$ from its perturbed version solution $b(\epsilon)$. However, to what extent is the perturbed model (D9) close to the true one (D3) for a nonequidistant $\{\zeta_i\}_{i=1}^n$ but close-to-equidistant case?

The distance between the solution is

$$\|b^{(\zeta)} - b(0)\| \leqslant \|b^{(\zeta)} - b(\epsilon)\| + \|b(\epsilon) - b(0)\|, \tag{D22}$$

where $b^{(\zeta)}$ is the solution for the system $\{\zeta_i\}_{i=1}^n$. We can write

$$\|b^{(\zeta)} - b(\epsilon)\| = \left\|\left(\tilde{E}_{2n-1}^{(\zeta)}\right)^{-1}\left(\tilde{\mu}_{2n-1}^{(\lambda)} + \Delta_{2n-1}\right) - (\hat{E}_{2n-1})^{-1}\left(\tilde{\mu}_{2n-1}^{(\lambda)} + \epsilon\tilde{r}\right)\right\| \leqslant \left\|\left(\tilde{E}_{2n-1}^{(\zeta)}\right)^{-1} - (\hat{E}_{2n-1})^{-1}\right\|\left(\left\|\tilde{\mu}_{2n-1}^{(\lambda)}\right\| + \epsilon\|\tilde{r}\|\right). \tag{D23}$$

Using the inequality $\|A^{-1} - B^{-1}\| \leqslant \|B - A\|\|A^{-1}\|\|B^{-1}\|$, we can write

$$\left\|\left(E_{2n-1}^{(\zeta)}\right)^{-1} - (\hat{E}_{2n-1})^{-1}\right\| \leqslant \left\|E_{2n-1}^{(\zeta)} - \hat{E}_{2n-1}\right\|\left\|\left(E_{2n-1}^{(\zeta)}\right)^{-1}\right\|\|(\hat{E}_{2n-1})^{-1}\|. \tag{D24}$$

The l_2 norms of the inverse matrices are

$$\begin{aligned}\|(E_{2n-1}^{(\zeta)})^{-1}\|_2 &= \sigma_{\max}\left(\left(E_{2n-1}^{(\zeta)}\right)^{-1}\right),\\ \|(\hat{E}_{2n-1})^{-1}\|_2 &= \sigma_{\max}((\hat{E}_{2n-1})^{-1}),\end{aligned} \tag{D25}$$

where $\sigma_{\max}(\cdot)$ represents the largest singular value of the corresponding matrices. The l_2-norm distance between $E_{2n-1}^{(\zeta)}$ and $\hat{E}_{2n-1}$ is

$$\begin{aligned}\left\|E_{2n-1}^{(\zeta)} - \hat{E}_{2n-1})\right\|_2 &= \sqrt{\sum_{i,j=1}^{2n-1}\left(\left(E_{2n-1}^{(\zeta)}\right)_{ij} - (\hat{E}_{2n-1})_{ij}\right)^2}\\ &\leqslant \sqrt{\sum_{j,k=1}^{2n-2} e^{2\mathbb{i}(j\Delta)(k\tau)}\left(e^{\mathbb{i}\frac{k\tau}{\Delta}\epsilon} - \left(1 + \mathbb{i}\frac{k\tau}{\Delta}\epsilon\right)\right)^2 + e^{-2\mathbb{i}(j\Delta)(k\tau)}\left(e^{-\mathbb{i}\frac{k\tau}{\Delta}\epsilon} - \left(1 - \mathbb{i}\frac{k\tau}{\Delta}\epsilon\right)\right)^2}\\ &\leqslant \frac{2\epsilon^2\tau^2}{\Delta^2}\sqrt{\sum_{j,k=1}^{2n-2} k^2\cos\left(2(j\Delta)(k\tau)\right)} \leqslant \frac{4\epsilon^2\tau^2}{3\Delta^2}(n-1)^2(2n-1)(4n-3) = \frac{16\pi^2\epsilon^2}{3\Delta^2}\frac{(n-1)^2(4n-3)}{(2n-1)}.\end{aligned} \tag{D26}$$

Then (D24) can be rewritten as

$$\left\|\left(E_{2n-1}^{(\zeta)}\right)^{-1} - (\hat{E}_{2n-1})^{-1}\right\|_2 \leqslant \frac{16\pi^2\epsilon^2}{3\Delta^2}\frac{(n-1)^2(4n-3)}{(2n-1)}\sigma_{\max}\left(\left(E_{2n-1}^{(\zeta)}\right)^{-1}\right)\sigma_{\max}((\hat{E}_{2n-1})^{-1}). \tag{D27}$$

The distance between the solution is

$$\begin{aligned}\|b^{(\zeta)} - b(0)\|_2 \leqslant{}& \frac{16\pi^2\epsilon^2}{3\Delta^2}\frac{(n-1)^2(4n-3)}{(2n-1)}\sigma_{\max}\left(\left(E_{2n-1}^{(\zeta)}\right)^{-1}\right)\sigma_{\max}((\hat{E}_{2n-1})^{-1})\left(\frac{\Delta}{3}\sqrt{n(n-1)} + \epsilon\right)\\ &\times \epsilon\left(1 + \frac{2\pi}{\Delta}\frac{\sqrt{n(4n-1)}}{\sqrt{3(2n-1)}}\right)\frac{2(2^n-1)\Delta}{\sqrt{2n-1}}.\end{aligned} \tag{D28}$$

Since $\sigma_{\max}[(E_{2n-1}^{(\zeta)})^{-1}] \sim \sigma_{\max}[(\hat{E}_{2n-1})^{-1}] \sim \epsilon(\sqrt{2n-1})^{-1}$, hold, and we select ϵ as

$$\epsilon \sim O\left(\frac{1}{2^{\alpha n}n^{\beta}}\right), \quad \alpha, \beta \geqslant 1, \tag{D29}$$

we can write

$$\|b^{(\zeta)} - b(0)\|_2 \sim O\left(\frac{1}{2^{5\alpha n}n^{5\beta-1}}\right), \quad \alpha, \beta \geqslant 1, \tag{D30}$$

that tends to zero while $n \to \infty$.

[1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[2] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Circuit-centric quantum classifiers, Phys. Rev. A **101**, 032308 (2020).

[3] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[4] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

[5] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018).

[6] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Phys. Rev. A **99**, 032331 (2019).

[7] A. Yu. Kitaev, Quantum measurements and the Abelian stabilizer problem, arXiv preprint arXiv:quant-ph/9511026.

[8] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate set via robust phase estimation, Phys. Rev. A **92**, 062315 (2015).

[9] E. van den Berg, Iterative quantum phase estimation with optimized sample complexity, in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, New York, 2020), pp. 1–10.

[10] N. Wiebe and C. Granade, Efficient Bayesian phase estimation, Phys. Rev. Lett. **117**, 010503 (2016).

[11] T. E. O'Brien, B. Tarasinski, and B. M. Terhal, Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments, New J. Phys. **21**, 023022 (2019).

[12] D. Wang, O. Higgott, and S. Brierley, Accelerated variational quantum eigensolver, Phys. Rev. Lett. **122**, 140504 (2019).

[13] I. Hamamura and T. Imamichi, Efficient evaluation of quantum observables using entangled measurements, npj Quantum Inf. **6**, 56 (2020).

[14] O. Crawford, B. Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, Efficient quantum measurement of Pauli operators in the presence of finite sampling error, Quantum **5**, 385 (2021).

[15] L. A. Markovich, A. Almasi, S. Zeytinoğlu, and J. Borregaard, Quantum memory assisted observable estimation, arXiv:2212.07710.

[16] S. Polla, G.-L. R. Anselmetti, and T. E. O'Brien, Optimizing the information extracted by a single qubit measurement, Phys. Rev. A **108**, 012403 (2023).

[17] D. Kane, S. Karmalkar, and E. Price, Robust polynomial regression up to the information theoretic limit, in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2017), pp. 391–402.

[18] D. Stilck Franca, L. A. Markovich, V. V. Dobrovitski, A. H. Werner, and J. Borregaard, Efficient and robust estimation of many-qubit Hamiltonians, Nat. Commun. **15**, 311 (2024).

[19] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control, Phys. Rev. Lett. **118**, 150503 (2017).

[20] A. Mari, T. R. Bromley, and N. Killoran, Estimating the gradient and higher-order derivatives on quantum hardware, Phys. Rev. A **103**, 012405 (2021).

[21] G.-L. R. Anselmetti, D. Wierichs, C. Gogolin, and R. M. Parrish, Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems, New J. Phys. **23**, 113010 (2021).

[22] J. G. Vidal and D. O. Theis, Calculus on parameterized quantum circuits, arXiv:1812.06323.

[23] L. Banchi and G. E. Crooks, Measuring analytic gradients of general quantum evolution with the stochastic parameter shift rule, Quantum **5**, 386 (2021).

[24] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, General parameter-shift rules for quantum gradients, Quantum **6**, 677 (2022).

[25] O. Kyriienko and V. E. Elfving, Generalized quantum circuit differentiation rules, Phys. Rev. A **104**, 052417 (2021).

[26] J. Hadamard, Sur les problèmes aux dérivées partielles et leur signification physique, Princeton University Bulletin **13**, 49 (1902).

[27] A. N. Tikhonov and V. I. Arsenin, *Solutions of Ill-Posed Problems*, edited by J. Fritz (John Wiley & Sons, New York, 1977).

[28] G. H. Golub and C. F. Van Loan, *Matrix Computations* (Johns Hopkins University, Baltimore, MD, 2013).

[29] V. N. Vapnik and A. R. Stephanyuk, Nonparametric methods for probability density reconstruction, Autom. Remote Control **39**, 1127 (1979).

[30] V. Vapnik, *Estimation of Dependences Based on Empirical Data* (Springer Science & Business Media, New York, 2006).

[31] A. R. Stefanyuk, Estimating the likelihood ratio function in the problem of failure of a stochastic-process, Autom. Remote Control **47**, 1210 (1986).

[32] V. A. Morozov, Regular methods for solving linear and nonlinear ill-posed problems, in *Methods for Solving Incorrectly Posed Problems* (Springer, New York, 1984), pp. 65–122.

[33] H. W. Engl and H. Gfrerer, A posteriori parameter choice for general regularization methods for solving linear ill-posed problems, Appl. Numer. Mathem. **4**, 395 (1988).

[34] V. N. Vapnik, N. M. Markovich, and A. R. Stefanyuk, Rate of convergence in l2 of the projection estimator of the distribution density, Autom. Remote Control **53**, 677 (1992).

[35] J. Cheng and B. Hofmann, Regularization methods for ill-posed problems, Handbook of Mathematical Methods in Imaging **1**, 2 (2015).

[36] S. Lu, S. V. Pereverzev, Y. Shao, and U. Tautenhahn, On the generalized discrepancy principle for Tikhonov regularization in Hilbert scales, J. Inte. Equations Appl. **22**, 483 (2010).

[37] J. S. Kottmann, A. Anand, and A. Aspuru-Guzik, A feasible approach for automatically differentiable unitary coupled-cluster on quantum computers, Chem. Sci. **12**, 3497 (2021).

[38] D. O. Theis, Proper shift rules for derivatives of perturbed-parametric quantum evolutions, Quantum **7**, 1052 (2023).