Spectral estimation for Hamiltonians

A comparison between classical imaginary-time evolution and quantum real-time evolution

Stroeks, M. E.; Helsen, J.; Terhal, B. M.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

PAPER • OPEN ACCESS

Spectral estimation for Hamiltonians: a comparison between classical imaginary-time evolution and quantum real-time evolution

To cite this article: M E Stroeks et al 2022 New J. Phys. **24** 103024

New Journal of Physics
The open access journal at the forefront of physics

PAPER

Spectral estimation for Hamiltonians: a comparison between classical imaginary-time evolution and quantum real-time evolution

M E Stroeks[1,2,*] 🔾, J Helsen[3] 🔾 and B M Terhal[1,2,4] 🔾

1 QuTech, Delft University of Technology, PO Box 5046, 2600 GA Delft, The Netherlands
2 Faculty of EEMCS, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
3 CWI & QuSoft, Science Park 123, 1098 XG Amsterdam, The Netherlands
4 JARA Institute for Quantum Information, Forschungszentrum Jülich, D-52425 Jülich, Germany
* Author to whom any correspondence should be addressed.

E-mail: m.e.h.m.stroeks@tudelft.nl, Jonas.Helsen@cwi.nl and b.m.terhal@tudelft.nl

Abstract

We consider the task of spectral estimation of local quantum Hamiltonians. The spectral estimation is performed by estimating the oscillation frequencies or decay rates of signals representing the time evolution of states. We present a classical Monte Carlo (MC) scheme which efficiently estimates an imaginary-time, decaying signal for stoquastic (i.e. sign-problem-free) local Hamiltonians. The decay rates in this signal correspond to Hamiltonian eigenvalues (with associated eigenstates present in an input state) and can be extracted using a classical signal processing method like ESPRIT. We compare the efficiency of this MC scheme to its quantum counterpart in which one extracts eigenvalues of a general local Hamiltonian from a real-time, oscillatory signal obtained through quantum phase estimation circuits, again using the ESPRIT method. We prove that the ESPRIT method can resolve $S = \mathrm{poly}(n)$ eigenvalues, assuming a $1/\mathrm{poly}(n)$ gap between them, with $\mathrm{poly}(n)$ quantum and classical effort through the quantum phase estimation (QPE) circuits, assuming efficient preparation of the input state. We prove that our MC scheme plus the ESPRIT method can resolve $S = O(1)$ eigenvalues, assuming a $1/\mathrm{poly}(n)$ gap between them, with $\mathrm{poly}(n)$ purely classical effort for stoquastic Hamiltonians, requiring some access structure to the input state. However, we also show that under these assumptions, i.e. $S = O(1)$ eigenvalues, assuming a $1/\mathrm{poly}(n)$ gap between them and some access structure to the input state, one can achieve this with $\mathrm{poly}(n)$ purely classical effort for *general* local Hamiltonians. These results thus quantify some opportunities and limitations of MC methods for spectral estimation of Hamiltonians. We numerically compare the MC eigenvalue estimation scheme (for stoquastic Hamiltonians) and the quantum-phase-estimation-based eigenvalue estimation scheme by implementing them for an archetypal stoquastic Hamiltonian system: the transverse field Ising chain.

Contents

1. Introduction

In general, it is a computationally intractable task to obtain, by classical or quantum means, the eigenvalues of a Hamiltonian H associated with a many-body quantum system. However, more restricted tasks related to estimating the spectrum of H can be executed on a quantum computer by means of quantum phase estimation (QPE) algorithms [31, 32, 38, 43], using the ability to simulate the real-time dynamics $e^{-iHt/\hbar}$ efficiently on a quantum computer via Trotterization [27].

Classical alternatives are provided by quantum Monte Carlo (MC) methods [11, 16]. The efficiency of quantum MC methods when used to simulate many-body systems is generally limited by the *sign problem*. This can cause the variance of the estimator in the MC algorithm to grow exponentially in the system size n, necessitating an exponential number of runs of the MC algorithm.

A (ubiquitous) class of Hamiltonians that is sign-problem-free has been formalized under the name *stoquastic* Hamiltonians [6]. Roughly speaking, a (real-valued) Hamiltonian is stoquastic (in a particular basis $\mathcal{B}$) if its off-diagonal elements are non-positive: $\langle x|H|y\rangle \leqslant 0$, for $x \neq y$ (with $|x\rangle, |y\rangle$ being elements of $\mathcal{B}$). As a consequence, its associated Gibbs density matrix $e^{-\tau H}$ is an element-wise non-negative matrix (for $\tau \in \mathbb{R}_+$). This property makes it particularly suitable for MC sampling as complexity results [1, 6, 23] and various algorithmic results [4, 5, 9, 10] have demonstrated.

Since stoquastic Hamiltonians are sign-problem-free, it is of interest to see if one can indeed prove that (part of its) spectrum can be efficiently estimated through classical MC methods. Conversely, can quantum algorithms, *even* for stoquastic Hamiltonians, provide an advantage over MC algorithms in carrying out this task? In this work, we address these questions by making a direct comparison between the task of estimating the spectral content of a stoquastic local Hamiltonian in an input state via a quantum circuit versus via a classical MC scheme. In addition, we investigate to what extent this task can be efficiently carried out classically for a general local Hamiltonian.

Central in our study is, first of all, the real-time *signal*

$$g_{\mathrm{R}}(k) = \langle \Phi | e^{-iHk\Delta t} | \Phi \rangle = \sum_{j=1}^{2^n} \left| \langle \psi_j | \Phi \rangle \right|^2 \left(e^{-iE_j \Delta t} \right)^k, \tag{1}$$

for $k = 0, 1\ldots, K$ and where $|\Phi\rangle$ is some pre-specified n-qubit input state. The estimation of $g_{\mathrm{R}}(k)$ for various k is a crucial step in the QPE algorithm. In what follows we will fix Δt so that the eigenstates $|\psi_j\rangle$ with nonzero or substantial overlap $|\langle \psi_j | \Phi \rangle|^2 > 0$, showing up in the signal, have the property that $E_j \Delta t \in [0, 2\pi]$. Thus, from now on, we assume that these E_j are shifted and rescaled to lie in $[0, 2\pi]$. We will assume that there are at most S eigenvectors with nonzero $|\langle \psi_j | \Phi \rangle|^2$, where S is desired to be poly(n) or less for overall efficiency. Identifying a state $|\Phi\rangle$ which has non-zero overlap on only a few ($S = \text{poly}(n)$ or $S = O(1)$) eigenstates and which obeys the assumptions in the following theorems is not so simple, and can be considered one of the bottlenecks in using QPE or other MC methods to determine spectral information of the Hamiltonian.

Besides the real-time signal, one can define the imaginary-time *signal*

$$g_{\mathrm{I}}(k) = \langle \Phi | e^{-Hk} | \Phi \rangle = \sum_{j=1}^{2^n} \left| \langle \psi_j | \Phi \rangle \right|^2 \left(e^{-E_j} \right)^k, \tag{2}$$

where again we can assume that $E_j \in [0, 2\pi]$. We will prove, *for local stoquastic Hamiltonians*, that the quantum cost of estimating $g_{\mathrm{R}}(k)$ and the classical MC cost of estimating $g_{\mathrm{I}}(k)$ within error ϵ are *approximately identical*, although the assumptions on our knowledge/preparation costs of $|\Phi\rangle$ are slightly different in the two cases. We present the MC scheme that estimates the signal in equation (2) in section 2. We stress that this MC scheme does *not* rely on the Metropolis–Hastings algorithm.

The two statements are as follows (the proof of these statements is given in section 2):

Theorem 1.1. *For a local Hamiltonian acting on n qubits, one can estimate $g_R(k)$ in equation (1) with probability at least $1 - \delta$ with sampling error ϵ and Trotter error ϵ_{trot} (and total error $\epsilon_{\text{tot}} = \epsilon + \epsilon_{\text{trot}}$), using quantum circuits acting on $n + 1$ qubits, where the depth of the quantum circuit scales as $\mathcal{O}(k^{1+o(1)})\mathcal{O}\left(\epsilon_{\text{trot}}^{-o(1)}\right) \times \text{poly}(n)$ and the number of times one executes the circuit is $\Theta(\epsilon^{-2}\log(4\delta^{-1}))$, under the assumption that $|\Phi\rangle$ is a state of n qubits which can be generated by a $\text{poly}(n)$-size quantum circuit. Hence to obtain $g_R(k)$ for $k = 0, \ldots, K$, with error at most $\epsilon_{\text{tot}} = \epsilon + \epsilon_{\text{trot}}$ for all k, with probability $1 - \delta$ requires using quantum circuits for $k = 0, \ldots, K$, each acting on $n + 1$ qubits, where the depth of the quantum circuit scales as $\mathcal{O}(k^{1+o(1)})\mathcal{O}\left(\epsilon_{\text{trot}}^{-o(1)}\right) \times \text{poly}(n)$ and each circuit is repeated $\Theta(\epsilon^{-2}\left[\log(4\delta^{-1}) + \log(K)\right])$ times.*

Theorem 1.2. *For a local stoquastic Hamiltonian acting on n qubits, one can estimate $g_I(k)$ in equation (2) with probability at least $1 - \delta$ with total error $\epsilon_{\text{tot}} = \epsilon + \epsilon_{\text{trot}}$, using a classical MC algorithm on n-bit strings where the depth of the algorithm scales as $\mathcal{O}(k^{1+o(1)})\mathcal{O}\left(\epsilon_{\text{trot}}^{-o(1)}\right) \times \text{poly}(n)$ and the number of times one runs the algorithm is $\Theta(\epsilon^{-2}\log(\delta^{-1}))$, under the assumption that $|\Phi\rangle = \sum_{x=1}^{2^n}\Phi(x)|x\rangle$ is a normalized state of n qubits such that (1) $\frac{\Phi(y)}{\Phi(x)}$ can be efficiently ($\text{poly}(n)$) calculated for a given x and y and (2) we can efficiently draw samples from the probability distribution $P(x) = |\Phi(x)|^2$. Hence to obtain $g_I(k)$ for all $k = 0, \ldots, K$, with error at most ϵ_{tot} for each k, with probability $1 - \delta$ requires using a classical MC algorithm on n-bit strings for $k = 0, \ldots, K$, where the depth of each algorithm scales as $\mathcal{O}(k^{1+o(1)})\mathcal{O}\left(\epsilon_{\text{trot}}^{-o(1)}\right) \times \text{poly}(n)$ and the number of times one runs the algorithm (for each k) is $\Theta(\epsilon^{-2}\left[\log(\delta^{-1}) + \log(K)\right])$.*

We then ask, given knowledge of either the real-time signal $g_R(k)$ or imaginary-time signal $g_I(k)$, what can be learnt about those eigenvalues E_j, whose associated eigenstates have nonzero overlap with the input state $|\Phi\rangle$? The signals $g_I(k)$ and $g_R(k)$ respectively correspond to a probabilistic sum of decaying components and a sum of oscillating components with decay rates and oscillation frequencies E_j as a function of discrete 'time' $k = 0, \ldots, K$. Hence a method which extracts those decay and oscillation rates from knowing $g_I(k)$ or $g_R(k)$ at various k is needed. A method of choice which has already been used in quantum information theory is the matrix pencil method [22, 35, 37] (with equivalent methods known as ESPRIT and MUSIC). This method has been used for processing randomized benchmarking data [19, 33], QPE [32], spectral tomography of superoperators [18], for processing experimental time-series data to identify Hamiltonian parameters [17] or generally in processing discretely-sampled decaying Ramsey signals. In this work, we employ specifically the ESPRIT method and investigate its ability to extract the E_j's from the signals in equations (1) and (2). More details on the ESPRIT method are given in section 3 and the algorithm implementing the ESPRIT method is given explicitly in algorithm 2.

Using this method, it is known that if *either* $g_R(k)$ or $g_I(k)$ is known *exactly* for $k = 0, \ldots, K$ where $K + 1 \geqslant 2S$, one can learn those eigenvalues E_j and probabilities $|\langle\psi_j|\Phi\rangle|^2$ *exactly*. However, in the presence of sampling and Trotter noise, the resolving power also depends on the gap between the eigenvalues E_j, the number S of eigenvalues and whether we extract them from an oscillating or decaying signal. Our work is thus focused on understanding whether there are fundamental advantages in learning $g_R(k)$ with noise versus learning $g_I(k)$ with noise, as this quantifies the benefit of a quantum algorithm versus a classical algorithm for spectral estimation of (stoquastic) Hamiltonians.

Not surprisingly, there are drawbacks to processing data from the imaginary-time evolution. As the signal decays exponentially, k cannot be chosen too large otherwise the signal becomes smaller than the noise. Our goal is to quantify this precisely and show that, at least theoretically, a regime exists in which the MC method may be competitive.

The first statement we make can be viewed as a summary of previous work, namely it combines lemma 2.1 via theorem 1.1 (which state the computational effort of estimating the real-time evolution signal up to a given error and with a given confidence) and the performance of the ESPRIT method in theorem 3.1 (which gives an error bound for the eigenvalue estimates obtained from application of the ESPRIT method to the real-time evolution signal) in the presence of a gap:

Theorem 1.3. *Given a local Hamiltonian on n qubits. Let the number of eigenvectors supported in some (efficient-to-prepare) input state $|\Phi\rangle$ be $S = p_1(n)$ (with $p_1(n)$ some polynomial in n), and each occurs with nonzero probability at least $1/\text{poly}(n)$. Furthermore, assume that the S eigenvalues $\{E_i\}$ with $E_i \in [0, 2\pi)$ are sufficiently well-separated, i.e. at least by a gap $\Delta \geqslant C/K$ with constant C and $K = \Theta(p_1(n))$. Then using Hadamard test (QPE) quantum circuits plus signal post-processing via ESPRIT, each requiring a $\text{poly}(n)$ effort, one can resolve the eigenvalues $\{E_j\}$ with distance $d(\{E_i\}, \{\tilde{E}_j\})$ (defined in equation (32)) at most $1/\text{poly}(n)$.*

For local stoquastic Hamiltonians the combination of lemma 2.3 via theorem 1.2 (which state the computational effort of estimating the imaginary-time evolution signal up to a given error and with a given

confidence) and the performance of the ESPRIT method in theorem 3.2 (which gives an error bound for the eigenvalue estimates obtained from application of the ESPRIT method to the imaginary-time evolution signal) in the presence of a gap leads to:

Theorem 1.4. *Given a local stoquastic Hamiltonian on n qubits. Let the number of eigenvectors supported in some efficient-to-sample (i.e. with* poly(n) *effort) input state* $|\Phi\rangle$ *be* $S = O(1)$, *and each occurs with nonzero probability at least* $1/\text{poly}(n)$. *In addition, assume that for a fixed* x, y *it is efficient to compute* $\frac{\Phi(y)}{\Phi(x)}$. *Furthermore, assume that the S eigenvalues* $\{E_i\}, E_i \in [0, 2\pi)$ *are sufficiently well-separated, i.e. at least by* $\Delta \geqslant 1/\text{poly}(n)$ *with some* poly(n). *Then using a MC algorithm plus signal post-processing via ESPRIT, each requiring (some)* poly(n) *effort, one can resolve the eigenvalues* $\{E_j\}$ *with distance* $d(\{E_i\}, \{\tilde{E}_j\})$ *at most* $1/\text{poly}(n)$.

Theorem 1.4 immediately begs the question whether such a result could hold for general local Hamiltonians as well: the assumptions that there are only $S = O(1)$ eigenstates in the initial state, as well as the assumption of efficient access to the initial state appear rather strong. To address this question, we define another real-valued, decaying signal as

$$g_D(k) = \langle\Phi|(I - H/2\pi)^k|\Phi\rangle = \sum_{j=1}^{2^n}|\langle\psi_j|\Phi\rangle|^2(1 - E_j/2\pi)^k. \tag{3}$$

If $S = O(1)$ and if $g_D(k)$ can be estimated with some accuracy for $k = 1, \ldots, K = O(1)$, we can also apply the ESPRIT method to extract these S eigenvalues. We note that this requires that the eigenvalues E_j are bounded away from 2π. Hence if we use $g_D(k)$ we assume that we have shifted and rescaled the eigenvalues so that, say, the E_j's lie in $[0, \pi]$.

One can prove that for general local Hamiltonians, assuming $S = O(1)$ eigenvalues in $|\Phi\rangle$, one can estimate $g_D(k)$ with ϵ accuracy, under an assumption about the access to $|\Phi\rangle$ which is identical to the MC case for stoquastic Hamiltonians (theorem 1.4). In fact, this result shows that theorem 1.4 is not particular to local stoquastic Hamiltonians at all, if we only care about 'nominally poly(n)' algorithms. However, the computational cost of estimating $g_D(k)$ for general local Hamiltonians is significantly higher in practice compared to the MC method for stoquastic Hamiltonians. The result expressed in lemma 2.4 can be viewed as 'dequantization' as it is similar in spirit to the singular value transformation (SVT) tool (theorem 3 in [13]). Theorem 3 in [13] is used to construct an algorithm that estimates the ground state energy of a Hamiltonian to $O(1)$ (in n) precision, given an initial state with only some constant overlap with the ground state.

Applying the ESPRIT analysis to lemma 2.4, we will obtain the following theorem:

Theorem 1.5. *Given a local Hamiltonian on n qubits. Let the number of eigenvectors supported in some efficient-to-sample* (poly(n) *effort) input state* $|\Phi\rangle$ *be* $S = O(1)$, *and each occurs with nonzero probability at least* $1/\text{poly}(n)$. *In addition, assume that for a fixed* x, y *it is efficient to compute* $\frac{\Phi(y)}{\Phi(x)}$. *Furthermore, assume that the S eigenvalues* $\{E_i\}, E_i \in [0, \pi]$ *are sufficiently well-separated, i.e. at least by* $\Delta \geqslant 1/\text{poly}(n)$ *with some* poly(n). *Then using lemma 2.4 plus signal post-processing via ESPRIT, each requiring (some)* poly(n) *classical effort, one can resolve the eigenvalues* $\{E_j\}$ *with distance* $d(\{E_i\}, \{\tilde{E}_j\})$ *at most* $1/\text{poly}(n)$.

In [13], it was additionally shown that estimating the smallest eigenvalue of a local Hamiltonian (with some restrictions on its locality) with inverse polynomial precision, even when provided that the guiding (input) state (which is a state with access structure similar to the one considered in this work) has close to a constant $(1/2 - \Omega(1/\text{poly}(n)))$ overlap with the ground state, is BQP-complete. Improving on this result, the authors of [7] have shown that this problem is BQP-complete even when the overlap of the guiding state with the ground state is $1 - \Omega(1/\text{poly}(n))$ (a result which was also obtained in [14]) and that equivalent results hold for the task of estimating excited state eigenvalues of the Hamiltonian. As also mentioned in [7], the ability to classically estimate $S = O(1)$ eigenvalues up to inverse polynomial error (provided an at least inverse polynomial spectral gap)—as described in theorems 1.4 and 1.5—thus depends strongly on the number of eigenvectors supported in the input state being *only* $S = O(1)$. Note that in theorems 1.4 and 1.5 we furthermore require that the $S = O(1)$ eigenvalues are separated by an at least $1/\text{poly}(n)$ gap.

To investigate practical aspects of the MC scheme for stoquastic Hamiltonians and compare it to the quantum scheme, we numerically study the one-dimensional Ising chain in a transverse field g [36] in a *proof-of-principle* setting. We numerically study, amongst several other aspects, the recovery of the ground-state and first-excited-state eigenvalues in the $(g > 1)$-regime from the signals $g_R(k)$ and $g_I(k)$ (in the presence of sampling noise and Trotter error) using the ESPRIT method.

An overview of the paper is as follows. In section 2, we review the *Hadamard* or *overlap* quantum subroutine (lemma 2.1) and we present the MC algorithm (lemma 2.3) for stoquastic Hamiltonians with its

proof, as well as stating a straightforward lemma 2.4 on 'dequantization'. Section 3 reviews the ESPRIT method and has an extensive appendix D in which we prove the performance of the ESPRIT method for imaginary-time decaying signals using many lemmas also needed in the real-time signal case. The arguments for theorem 1.5 are presented in section 3 as well. In section 4, we numerically compare the quantum scheme and the MC scheme (for stoquastic Hamiltonians) for determining part of the spectrum of a transverse field Ising chain. In section 5, we discuss our work and propose some directions for future study. Several appendices give additional background information and details.

We note that very extensive literature exists on the MC power method [16] in which one applies a sequences of steps which gradually project an initial input state onto the ground state. In this method, unlike in our MC scheme of lemma 2.3, one renormalizes the state after each iteration, so that the signal does not die out. In our approach, we do not renormalize, but study the decay rates themselves. In terms of other previous work, we note that in [4] the ground state energy of a stoquastic Hamiltonian was efficiently estimated by means of a projector MC scheme, under an additional 'guiding state' promise. In [30] the authors consider the implementation of the imaginary-time evolution $\exp(-\tau H)$ on a quantum computer in order to prepare a ground state of any local Hamiltonian. Note that our goal is not to prepare any ground or excited state but rather only learn some eigenvalues.

In the remainder of this section, we will review a few definitions which are used in this paper.

Definition 1. Stoquastic Hamiltonians. A (real-valued) Hamiltonian H is (globally) stoquastic [6] in a basis $\mathcal{B}$ if all its off-diagonal elements are non-positive: $\langle x|H|y\rangle \leqslant 0$, for $x \neq y$ (and states $|x\rangle, |y\rangle$ being elements of basis $\mathcal{B}$).

In this work, we are interested in Hamiltonians that are *local* and *stoquastic*:

Definition 2. Local Hamiltonians. A Hamiltonian H associated with a system consisting of n degrees of freedom (e.g. spins/qubits) is local if it admits a decomposition into a set of Hermitian operators $\{H_i\}$—i.e. $\sum_i^N H_i$—such that each H_i acts non-trivially on $O(1)$ (not growing with n) degrees of freedom of the system.

We denote the maximum number of degrees of freedom on which each H_i acts non-trivially (i.e. its *locality*) by k and note that the number of terms in a local Hamiltonian is $N = O(n^k)$.

For local Hamiltonians there is a slightly stronger notion of stoquasticity, called termwise stoquasticity, which can differ from the definition of stoquasticity given above, see [6, 23].

Definition 3. Termwise stoquastic Hamiltonians. A (real-valued) k-local Hamiltonian H is m-termwise stoquastic in a basis $\mathcal{B}$ if it admits a decomposition into (real-valued) $m(\geqslant k)$-local terms $\{H_a\}$ such that each H_a is stoquastic: $\forall\, a, \langle x|H_a|y\rangle \leqslant 0$, for $x \neq y$ (and states $|x\rangle, |y\rangle$ being elements of basis $\mathcal{B}$).

Most many-body Hamiltonians considered in physics which are stoquastic are $O(1)$-termwise stoquastic. The results in this paper apply to both termwise stoquastic as well as globally stoquastic Hamiltonians (using some small adaptions employing results in [23]), and we will refer to them simply as 'stoquastic'. For a matrix X we will use the operator or spectral norm $\|X\| = \sqrt{\lambda_{\max}(X^\dagger X)} = \sigma_{\max}(X)$, where $\sigma_{\max}(X)$ is the largest singular value of X. We also refer to the Frobenius norm $\|X\|_F = \sqrt{\text{Tr}(X^\dagger X)}$ and the induced $-\infty$ norm $\|X\|_\infty = \max_i \sum_j |X_{ij}|$. For an $m \times n$ matrix X, we use $\|X\| \leqslant \sqrt{m}\|X\|_\infty$ and $\|X\| \leqslant \|X\|_F$.

2. Quantum scheme versus Monte Carlo scheme for spectral estimation

In this section we show how to estimate $g_R(k)$ on a quantum computer, and $g_I(k)$ for stoquastic Hamiltonians via a MC algorithm, as well as how to estimate $g_D(k)$ inefficiently (in k) via a classical algorithm for general local Hamiltonians.

Lemma 2.1 states a well-known quantum subroutine, namely the Hadamard or overlap test, while a new result, a MC version of the routine, is proved in lemma 2.3. After these lemmas, the proofs of theorems 1.1 and 1.2 are given. Then we give lemma 2.4 for general local Hamiltonians, using similar tools as in lemma 2.3.

We note that the overlap test is used in versions of QPE which do not aim at preparing an energy eigenstate of the Hamiltonian, but rather only learn the spectral content in its input state, as in references [26, 32, 38]. Here we basically follow this approach for the real-time quantum evolution, which can in addition be randomized to save on implementation costs, see [44].

Lemma 2.1. (Hadamard or overlap test). *Let $\mathcal{F} \equiv \langle\Phi|G_1 G_2 \ldots G_L|\Phi\rangle$, where:*

(a) $|\Phi\rangle = \sum_{x=1}^{2^n} \Phi(x)|x\rangle$ *is a state of n qubits which can be generated by a* poly(n)*-size quantum circuit.*
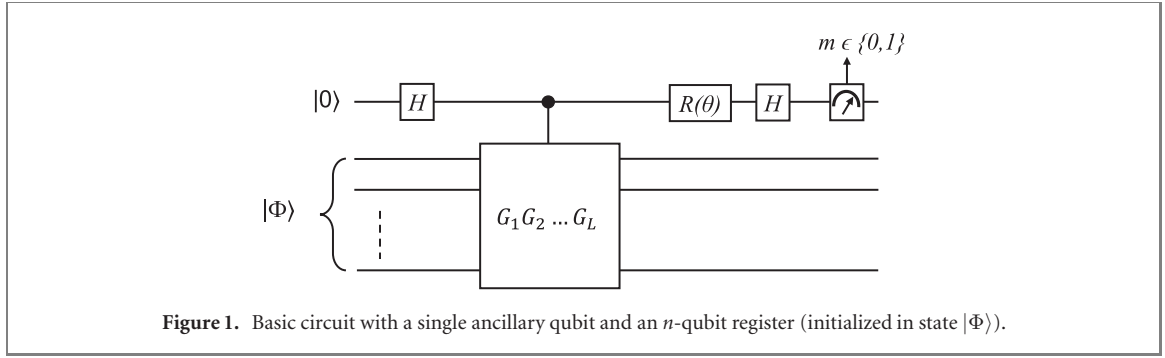*(b)* *Each G_l is a k-local unitary matrix.*

Figure 1. Basic circuit with a single ancillary qubit and an *n*-qubit register (initialized in state $|\Phi\rangle$).

$\mathcal{F}$ *can be estimated within error ϵ with probability at least $1 - \delta$ with a quantum circuit with* $\Theta(\epsilon^{-2} \log(4\delta^{-1})) \times \left[\Theta(L) + \mathrm{poly(n)}\right]$ *single and two-qubit gates.*

Proof. Figure 1 depicts the quantum circuit which is used. It involves an *n*-qubit register and a single ancillary qubit. The state of the composite system can be tracked through the circuit and the final state can be found to be (where $R(\theta) \equiv e^{-i\theta Z/2}$):

$$\frac{1}{2}\left(\left(e^{-i\theta/2}I + e^{i\theta/2}G_1 G_2 \dots G_L\right)|0\rangle_a \otimes |\Phi\rangle + \left(e^{-i\theta/2}I - e^{i\theta/2}G_1 G_2 \dots G_L\right)|1\rangle_a \otimes |\Phi\rangle\right). \quad (4)$$

A *Z*-measurement is now performed on the ancillary qubit, measuring either $|0\rangle$ or $|1\rangle$ with associated outcomes resp. $m = 0$ or $m = 1$. The probability to measure state $|0\rangle$ ($m = 0$) on the ancillary qubit after application of the depicted gates is then given by:

$$\Pr(m = 0|\theta) = \frac{1}{2} + \frac{1}{4}\left(e^{i\theta}\langle\Phi|G_1 G_2 \dots G_L|\Phi\rangle + e^{-i\theta}\left(\langle\Phi|G_1 G_2 \dots G_L|\Phi\rangle\right)^*\right)$$

$$= \begin{cases} \dfrac{1}{2} + \dfrac{1}{2}\,\mathrm{Re}\left(\langle\Phi|G_1 G_2 \dots G_L|\Phi\rangle\right), & \text{for } \theta = 0, \\[2mm] \dfrac{1}{2} - \dfrac{1}{2}\,\mathrm{Im}\left(\langle\Phi|G_1 G_2 \dots G_L|\Phi\rangle\right), & \text{for } \theta = \dfrac{\pi}{2}. \end{cases} \quad (5)$$

In the final expression, we have restricted ourselves to $\theta = 0$ and $\theta = \frac{\pi}{2}$, which are the θ values of interest. Suppose that for $\theta = 0$ and $\theta = \pi/2$, the quantum circuits are repeated $|\Sigma|$ times to obtain a set $2|\Sigma|$ of independent realizations of the ancillary-qubit state to be measured and let $\left|\Sigma_0^{\theta=0}\right|$ and $\left|\Sigma_0^{\theta=\pi/2}\right|$ be the number of times the ancilla measurement returns 0 so that

$$\tilde{\mathcal{F}} = \left(2\frac{\left|\Sigma_0^{\theta=0}\right|}{|\Sigma|} - 1\right) - i\left(2\frac{\left|\Sigma_0^{\theta=\pi/2}\right|}{|\Sigma|} - 1\right) \quad (6)$$

is our (unbiased) estimator, i.e. $\mathbb{E}(\tilde{\mathcal{F}}(t)) = \mathcal{F}(t)$. Then by means of the Chernoff bound we have

$$\Pr\left(\left|\tilde{\mathcal{F}} - \mathcal{F}\right| \leqslant \epsilon\right) \geqslant \Pr\left(\left|\mathrm{Re}(\tilde{\mathcal{F}} - \mathcal{F})\right| \leqslant \epsilon/\sqrt{2}\right)\Pr\left(\left|\mathrm{Im}(\tilde{\mathcal{F}} - \mathcal{F})\right| \leqslant \epsilon/\sqrt{2}\right)$$

$$= \left(1 - \Pr\left(\left|\mathrm{Re}(\tilde{\mathcal{F}} - \mathcal{F})\right| \leqslant \epsilon/\sqrt{2}\right)\right)\left(1 - \Pr\left(\left|\mathrm{Im}(\tilde{\mathcal{F}} - \mathcal{F})\right| \leqslant \epsilon/\sqrt{2}\right)\right)$$

$$\geqslant \left[\max\left(0, \left(1 - 2\exp(-|\Sigma|\epsilon^2/4)\right)\right)\right]^2 \quad (7)$$

$$\geqslant 1 - 4\exp\left(-|\Sigma|\epsilon^2/4\right) = 1 - \delta,$$

where the number of samples is chosen as $|\Sigma| = \Theta(\epsilon^{-2}\log(4\delta^{-1}))$. $\square$

Next, we will consider a classical MC version of the quantum routine given above. The key result here is lemma 2.3. However, before we can state it, we collect a few facts about matrices G_i which will be useful in the proof of lemma 2.3. The matrices G_i that we will consider now can be seen as analogous to the (unitary) local real-time propagation operators G_i considered earlier in lemma 2.1, but are now local *imaginary-time* propagation operators. The G_i are no longer unitary but, for local stoquastic Hamiltonians, are elementwise nonnegative. They are of the form $G_i = e^{-a_i/M\, kH_i}$, where a_i/M is a positive parameter set by the

Trotterization scheme and $k = 0, 1, \ldots, K$ denotes imaginary-time coordinate. We have the following proposition on further properties of these operators:

Proposition 2.2. *Let $G_i = e^{-a_i/M kH_i}$, where H_i is a stoquastic Hermitian matrix (a term in $H = \sum_i H_i$) which acts nontrivially on some subset of $O(1)$ qubits. Let the smallest eigenvalue of H_i be 0, i.e. $\lambda_{\min}(H_i) = 0$. We have:*

- *The matrix G_i is an elementwise nonnegative and positive definite matrix, with eigenvalues in the interval $(0, 1]$, acting nontrivially only on the same $O(1)$ qubits as H_i.*
- *If G_i is reducible, then we can write $G_i = \oplus_{b=1}^{B_i} G_i^b$ with B_i irreducible sub-matrices G_i^b. The set of bit string basis states on which the irreducible sub-matrix G_i^b acts is denoted by S_i^b, where $\cup_b S_i^b \subseteq \{0, 1\}^n$.*
- *From the Perron–Frobenius theorem (theorem 8.4.4 in [20]) it follows that for each nonnegative and irreducible sub-matrix there exists a unique and strictly positive eigenstate associated with its largest eigenvalue, i.e.*

$$|\phi_i^b\rangle = \sum_{x \in S_i^b} \phi_i^b(x)|x\rangle, \quad G_i^b|\phi_i^b\rangle = \lambda_i^b|\phi_i^b\rangle, \tag{8}$$

  where $\phi_i^b(x) > 0, \forall\, x \in S_i^b$. Since the spectrum of G_i is the union of spectra of the submatrices G_i^b, the spectrum of each G_i^b also lies in the interval $(0, 1]$ and one of the blocks b will contain the largest eigenvalue of G_i equal to 1. In case G_i is irreducible itself, there is a largest nonnegative eigenvector as in equation (8) which has support $\phi_i(x) > 0$ for all x. In this case, the corresponding eigenvalue will be $\lambda_i = 1$.
- *Naturally, since G_i acts nontrivially only on a subset of $O(1)$ qubits (and acts as I on other qubits) one can efficiently compute the blocks G_i^b, its largest eigenvalue λ_i^b and associated eigenstate $|\phi_i^b\rangle$ in each block b.*

We prove the following:

Lemma 2.3. *Let $\mathcal{F} \equiv \langle\Phi|G_1 G_2 \ldots G_L|\Phi\rangle$, where:*

(a) *$|\Phi\rangle = \sum_{x=1}^{2^n} \Phi(x)|x\rangle$ is a normalized state of n qubits where $\Phi(x) \in \mathbb{C}\ (\forall\, x)$ and $\sum_x |\Phi(x)|^2 = 1$. We assume that (1) $\frac{\Phi(y)}{\Phi(x)}$ can be efficiently (poly(n)) calculated for a given x and y and (2) we can efficiently draw samples from the probability distribution $P(x) = |\Phi(x)|^2$.*

(b) *Each $G_l = G_l$ is a k-local, positive-definite, (elementwise) nonnegative matrix with eigenvalues in $(0, 1]$.*

$\mathcal{F}$ can be estimated within error ϵ with probability at least $1 - \delta$ with a classical MC algorithm with runtime $\Theta(\epsilon^{-2} \log(\delta^{-1})) \times \text{poly}(n) \times \Theta(L)$.

Proof. The proof of lemma 2.3 consists of two steps: to construct an estimator for $\mathcal{F}(\tau)$ and to show that the error of this estimator can be bounded according to the lemma.

We rewrite the quantity of interest $\mathcal{F}$ as follows (where $L - 1$ complete sets of basis states are inserted in between the G_l operators in the final equality):

$$\mathcal{F} = \sum_{x_0, x_1, \ldots, x_L} |\Phi(x_0)|^2 \frac{\Phi(x_L)}{\Phi(x_0)} \langle x_0|G_1|x_1\rangle\langle x_1|G_2|x_2\rangle \ldots \langle x_{L-1}|G_L|x_L\rangle, \tag{9}$$

where we have set $|x\rangle = |x_0\rangle$ and $|y\rangle = |x_L\rangle$. $\mathcal{F}$ thus corresponds to the sum of an exponential number of products of (non-negative) matrix elements of $G_1, \ldots, G_L$, weighted by amplitudes in the state $|\Phi\rangle$. Evidently, only terms for which all the matrix elements in the product are non-zero contribute to the sum.

We now consider the string of basis states $|x_0\rangle, \ldots, |x_L\rangle$ and associate with each step $|x_{l-1}\rangle$ to $|x_l\rangle$ in this string a probability

$$P_l(x_{l-1} \to x_l) = \frac{1}{\lambda_l^b} \langle x_{l-1}|G_l|x_l\rangle \frac{\phi_l^b(x_l)}{\phi_l^b(x_{l-1})}, \tag{10}$$

where b labels the sub-block in $G_l = \oplus_b G_l^b$ which contains the strings x_{l-1} and x_l. Here $\phi_l^b(x) \equiv \langle x_{l-1}|\phi_l^b\rangle$ with $|\phi_l^b\rangle$ defined in proposition 2.2.

The probability distribution P_l is thus non-negative as $\lambda_l^b \in (0, 1]$, G_l is element-wise non-negative and $\phi_l^b(x_l) > 0$ and $\phi_l^b(x_{l-1}) > 0$. It can be shown to be normalized:

$$\sum_{x_l} P_l(x_{l-1} \to x_l) = \sum_{x_l} \frac{1}{\lambda_l^b} \langle x_{l-1}|G_l|x_l\rangle \frac{\phi_l^b(x_l)}{\phi_l^b(x_{l-1})} = \sum_{x_l \in S_l^b} \frac{1}{\lambda_l^b} \langle x_{l-1}|G_l^b|x_l\rangle \frac{\phi_l^b(x_l)}{\phi_l^b(x_{l-1})}$$

$$= \frac{1}{\phi_l^b(x_{l-1})} \frac{1}{\lambda_l^b} \langle x_{l-1}|G_l^b|\phi_l^b\rangle = \frac{1}{\phi_l^b(x_{l-1})} \frac{1}{\lambda_l^b} \langle x_{l-1}|\lambda_l^b|\phi_l^b\rangle = \frac{\langle x_{l-1}|\phi_l^b\rangle}{\phi_l^b(x_{l-1})} = 1. \tag{11}$$

We use $P_l(x_{l-1} \to x_l)$ to rewrite $\mathcal{F}(\tau)$ as:

$$\mathcal{F}(\tau) = \sum_{x_0,x_1,\dots,x_L} \underbrace{|\Phi(x_0)|^2 P_1(x_0 \to x_1) P_2(x_1 \to x_2) \dots P_L(x_{L-1} \to x_L)}_{\equiv \Pi(\boldsymbol{x}).} \times \underbrace{\frac{\Phi(x_L)}{\Phi(x_0)} \prod_{l=1}^{L} \lambda_l^{b(l)} \frac{\phi_l^{b(l)}(x_{l-1})}{\phi_l^{b(l)}(x_l)}}_{\equiv \mathcal{R}(\boldsymbol{x}).}, \qquad (12)$$

where $\boldsymbol{x} \equiv (x_0, x_1, \dots, x_L)$ and we have defined the quantities $\Pi(\boldsymbol{x})$ and $\mathcal{R}(\boldsymbol{x})$. Since $|\Phi(x_0)|^2$ and each P_l are probability distributions, $\Pi(\boldsymbol{x})$ is a probability distribution as well, i.e.

$$\sum_{x_0,x_1,\dots,x_L} \Pi(\boldsymbol{x}) = \sum_{x_0} \left(|\Phi(x_0)|^2 \sum_{x_1} \left(P_1(x_0 \to x_1) \dots \sum_{x_L} \left(P_L(x_{L-1} \to x_L) \right) \dots \right) \right) = 1. \qquad (13)$$

Clearly, one can sample from $\Pi(\boldsymbol{x})$ by first sampling from $|\Phi(x_0)|^2$, then sampling from $P_1(x_0 \to x_1)$ to generate x_1 etc until x_L.

By thus sampling from the probability distribution $\Pi(\boldsymbol{x})$ and obtaining a mean estimator for $\mathcal{F}(\tau)$ using the samples $\mathcal{R}(\boldsymbol{x})$, we can estimate $\mathcal{F}(\tau)$. We note that $\mathcal{F}(\tau) = \mathbb{E}(\mathcal{R}(\boldsymbol{x}))$. Since $\mathcal{R}(\boldsymbol{x}) \in \mathbb{C}$, a mean estimator over a finite number of samples will generally be complex-valued. Since $\mathcal{F}(\tau) \in \mathbb{R}$, we will instead obtain a mean estimator using samples $\mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$. The mean estimator that we shall use to estimate $\mathcal{F}(\tau)$ is the median-of-means estimator [28]. Using a set Σ of samples $\{\boldsymbol{x}\}$ (distributed according to $\Pi(\boldsymbol{x})$), the median-of-means estimator is defined as follows: divide the set Σ into q subsets $s_1, \dots, s_q$ of size approximately $|\Sigma|/q$. Calculate the empirical mean of $\mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ over the samples in each subset: $f_j = \frac{1}{|s_j|} \sum_{\boldsymbol{x} \in s_j} \mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ for $j \in \{1, \dots, q\}$ (each f_j is an unbiased estimator of $\mathcal{F}(\tau)$). Now the median-of-means estimator is given by the median of these empirical means: $\hat{\mathcal{F}} = \mathrm{M}(f_1, \dots, f_q)$. See appendix C for more details.

The algorithm that efficiently produces $\hat{\mathcal{F}} = \mathrm{M}(f_1, \dots, f_q)$ is explicitly given in algorithm 1 below. Note that when $\Phi(x_0)$ is small for some x_0, the probability of drawing this x_0, $|\Phi(x_0)|^2$, is very small, but the ratio $\frac{\Phi(x_L)}{\Phi(x_0)}$ in the estimator could get very large.

Algorithm 1 thus efficiently provides an estimate of $\mathcal{F}(\tau)$ (albeit biased). To complete the proof, we will show that the variance of $\mathrm{Re}(\mathcal{R}(\boldsymbol{x})) \in \mathbb{R}$ can be bounded which in turn is used to bound the number of samples to get an estimate close to the mean, leading to lemma 2.3.

For a complex random variable $Z = \mathcal{R}(\boldsymbol{x})$, $\mathbb{E}(Z) \equiv \mathbb{E}(\mathrm{Re}(Z)) + i\mathbb{E}(\mathrm{Im}(Z))$ and $\mathrm{Var}(Z) = \mathrm{Var}(\mathrm{Re}(Z)) + \mathrm{Var}(\mathrm{Im}(Z))) \geqslant \mathrm{Var}(\mathrm{Re}(Z))$. Hence we can bound the variance of random variable $\mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ by bounding the variance of the random variable $\mathcal{R}(\boldsymbol{x})$. This variance is given by:

$$\mathrm{Var}(\mathcal{R}(\boldsymbol{x})) = \mathbb{E}\left(|\mathcal{R}(\boldsymbol{x})|^2\right) - \underbrace{|\mathbb{E}(\mathcal{R}(\boldsymbol{x}))|^2}_{=|\mathcal{F}(\tau)|^2 = \mathcal{F}(\tau)^2.} \leqslant \mathbb{E}\left(|\mathcal{R}(\boldsymbol{x})|^2\right), \qquad (14)$$

where the inequality holds because $\mathcal{F}^2 \geqslant 0$ (since $\mathcal{F} \in \mathbb{R}$). To obtain an upper bound on the variance, we shall investigate this expression in more detail:

$$\mathbb{E}\left(|\mathcal{R}(\boldsymbol{x})|^2\right) = \sum_{\boldsymbol{x}} \Pi(\boldsymbol{x}) |\mathcal{R}(\boldsymbol{x})|^2$$

$$= \sum_{\boldsymbol{x}} |\Phi(x_L)|^2 \langle x_0|G_1|x_1\rangle \langle x_1|G_2|x_2\rangle \dots \langle x_{L-1}|G_L|x_L\rangle \prod_{l=1}^{L} \lambda_l^{b(l)} \frac{\phi_l^{b(l)}(x_{l-1})}{\phi_l^{b(l)}(x_l)} \qquad (15)$$

$$= \sum_{\boldsymbol{x}} |\Phi(x_L)|^2 Q_1(x_0, x_1) Q_2(x_1, x_2) \dots Q_L(x_{L-1}, x_L),$$

where in the last equality we defined the non-negative quantity $Q_l(x, y) \equiv \langle x|G_l|y\rangle \lambda_l^{b(l)} \frac{\phi_l^{b(l)}(x)}{\phi_l^{b(l)}(y)}$. Exploiting the Hermiticity of G_l^b, $Q_l(x, y)$ can be shown to have the following property:

$$\sum_x Q_l(x, y) = \sum_x \langle x|G_l|y\rangle \lambda_l^b \frac{\phi_l^b(x)}{\phi_l^b(y)} = \sum_{x \in S_l^b} \langle x|G_l^b|y\rangle \lambda_l^b \frac{\phi_l^b(x)}{\phi_l^b(y)} = (\lambda_l^b)^2 \frac{\langle \phi_l^b|y\rangle}{\phi_l^b(y)} = (\lambda_l^b)^2 \leqslant 1. \qquad (16)$$

$Q_l(x, y)$ thus satisfies $0 \leqslant Q_l(x, y) \leqslant 1, \forall x, y$ and $\forall l \in \{1, 2, \dots, L\}$. By consecutively exploiting the property in equation (16) for all Q_l's and the normalization property of state $|\Phi\rangle$ in the expression in equation (15), we obtain

$$\mathrm{Var}(\mathcal{R}(\mathbf{x})) \leqslant \mathbb{E}\left(|\mathcal{R}(\mathbf{x})|^2\right) \leqslant 1. \Rightarrow \mathrm{Var}(\mathrm{Re}(\mathcal{R}(\mathbf{x}))) \leqslant 1. \qquad (17)$$

Algorithm 1. Efficiently obtaining a median-of-means estimate of $\mathcal{F}(\tau)$ through sampling of the probability distribution $\Pi(\boldsymbol{x})$.

Input: Initial state $|\Phi\rangle$. Local propagation operators $\{G_l\}_{l=1}^L$. Sample size $|\Sigma|$. Number of subsets q.
Output: Median of means estimate of $\mathcal{F}(\tau)$.

for $\sigma \in \{1, 2, ..., |\Sigma|\}$ **do**
> Sample an initial basis state $|x_0\rangle$ from the probability distribution $|\Phi(x_0)|^2$. The state $|x_0\rangle$ is part of $S_1^{b(1)}$.
> **for** $l \in \{1, ..., L\}$ **do**
> > Pick a state $|x_l\rangle \in S_l^{b(l)}$ with probability $P_l(x_{l-1} \to x_l) = \frac{1}{\lambda_l^{b(l)}} \langle x_{l-1}| G_l |x_l\rangle \frac{\phi_l^{b(l)}(x_l)}{\phi_l^{b(l)}(x_{l-1})}$. The state $|x_l\rangle$ is part of $S_{l+1}^{b(l+1)}$ (for $l < L$).
>
> **end**
>
> Given $\{x_l\}_{l=1}^L$ (sampled from $\Pi(\boldsymbol{x})$), calculate $\mathcal{R}_\sigma(\boldsymbol{x}) = \frac{\Phi(x_L)}{\Phi(x_0)} \prod_{l=1}^L \lambda_l^{b(l)} \frac{\phi_l^{b(l)}(x_{l-1})}{\phi_l^{b(l)}(x_l)}$.

end
Divide the $|\Sigma|$ samples into q subsets $s_1, \ldots, s_q$, such that $|s_j| \approx |\Sigma|/q$, $\forall j$.
for $j \in \{1, 2, \ldots, q\}$ **do**
> Calculate $f_j = \frac{1}{|s_j|} \sum_{\boldsymbol{x} \in s_j} \mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$.

end
Output $\hat{\mathcal{F}} = \mathrm{M}(f_1, \ldots, f_q)$.

If we take the number of samples $|\Sigma|$, and divide them into q subsets $s_1, \ldots, s_q$ of size approximately $|\Sigma|/q$, then (by means of Chebyshev's inequality) each f_j obeys $|f_j - \mathcal{F}| \leqslant \sqrt{\mathrm{Var}(\mathrm{Re}(\mathcal{R}(\boldsymbol{x})))}\sqrt{4q/|\Sigma|} \leqslant \sqrt{4q/|\Sigma|}$ with probability at least $3/4$. Using Hoeffding's inequality and the definition of the mean, one can show that (see appendix C):

$$\Pr\left(\left|\hat{\mathcal{F}} - \mathcal{F}\right| \leqslant \sqrt{4q/|\Sigma|}\right) \geqslant 1 - e^{-q/8}. \tag{18}$$

Hence $\mathcal{F}$ can be estimated with error ϵ with probability at least $1 - \delta$ (with $q = 8\log(\delta^{-1})$) for $|\Sigma| = \Theta(\log(\delta^{-1})\epsilon^{-2})$, where obtaining each sample takes a number of operations that scales linearly in L and poly(n). This completes the proof of lemma 2.3. $\square$

Remark. Note that if one would have chosen the empirical mean $\tilde{\mathcal{F}} = \frac{1}{|\Sigma|}\sum_{\boldsymbol{x} \in \Sigma} \mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ as a mean estimator for $\mathcal{F}$ (instead of the median-of-means estimator), then using equation (17) and Chebyshev's inequality, we obtain:

$$\Pr\left(\left|\tilde{\mathcal{F}} - \mathcal{F}\right| \leqslant \epsilon\right) \geqslant 1 - \frac{\mathrm{Var}(\mathrm{Re}(\tilde{\mathcal{F}}))}{\epsilon^2} \geqslant 1 - \frac{1}{|\Sigma|\epsilon^2}. \tag{19}$$

Hence $\mathcal{F}$ can be estimated using $\tilde{\mathcal{F}}$ with error ϵ with probability at least $1 - \delta$, for $|\Sigma| = \Theta(\delta^{-1}\epsilon^{-2})$. Using the median-of-means estimator thus provides an exponential improvement in the required scaling of $|\Sigma|$ with δ^{-1}. Note that if we could upper and lower bound the range of $\mathrm{Re}(\mathcal{R}(\mathbf{x}))$ by some constants, then we could have used a Chernoff–Hoeffding bound for the empirical mean $\tilde{\mathcal{F}}$ which gives the aforementioned (exponentially) better dependence of the run-time of the algorithm with δ^{-1} (as in lemma 2.1 where we do use a Chernoff–Hoeffding bound).

Remark. Note that the lemma also applies to estimating $\langle x|G_1 G_2 \ldots G_L|x'\rangle$ (with $1/\mathrm{poly}(n)$ accuracy) as one simply starts the process at $x_0 = x$ and $\mathcal{R}(\mathbf{x})$ is only nonzero when one arrives at $x_L = x'$. Similarly, one can estimate $\langle \Phi_1|G_1 G_2 \ldots G_L|\Phi_2\rangle$ with $1/\mathrm{poly}(n)$ accuracy, assuming one can sample from $|\Phi_1(x)|^2$ (or $|\Phi_2(x)|^2$) *and* compute for a given x and y, the ratio $\frac{\Phi_2(y)}{\Phi_1(x)}$. In addition, one can extend the lemma to the case where the local propagation operators G_l are not Hermitian, but are still nonnegative matrices, see appendix B.

We stress that lemma 2.3 provides an efficient classical algorithm provided that: for a given $x, y \in \{0, 1\}^n$, one can efficiently determine $\frac{\Phi(y)}{\Phi(x)}$ and the state $|\Phi\rangle$ is such that one can efficiently draw samples from $P(x) = |\Phi(x)|^2$. In many practical settings, $|\Phi\rangle$ is such that one can define a function $f : \{0, 1\}^n \to \mathbb{C}$ which takes as input the n-bit string x, and efficiently outputs the corresponding coefficient $\Phi(x)$. This is e.g. the case for (matrix) product states or for other ansatz classes of states. Then, given x and y, the fraction $\frac{\Phi(y)}{\Phi(x)}$ can be efficiently obtained. Note that under this assumption one can set-up a MC scheme based on the Metropolis algorithm to sample from $|\Phi(x)|^2$, although this scheme is only a heuristic strategy and its efficient convergence would have to be proved. A good class of states to which both lemmas 2.3 and 2.1 apply are of course product states. Note that even when running the overlap test is too costly (as quantum circuits are noisy), but preparing the state $|\Phi\rangle$ is feasible, one could use this preparation to sample from $|\Phi(x)|^2$ for

the application of the MC method. Of course the requirement of being able to compute $\frac{\Phi(y)}{\Phi(x)}$ remains. For the transverse field Ising model, an example of a $|\Phi\rangle$ which obeys these conditions will be given in section 4.

Proof of theorems 1.1 and 1.2. We require the Trotterization of e^{-ikH} resp. e^{-kH} into a string of local propagation operators G_i which are unitary (in lemma 2.1) resp. Hermitian and non-negative (in lemma 2.3). This non-unique decomposition of e^{-ikH} and e^{-kH} into an ordered string of local propagation operators depends on the Trotterization scheme and is discussed in appendix A (and more extensively in [8]). The Trotterization gives an error ϵ_{trot} (in addition to the sampling error ϵ in lemmas 2.1 and 2.3) and the number of local propagation operators (for each sample) L in lemmas 2.1 and 2.3 will be $L = \text{poly}(n)\mathcal{O}\left(\Upsilon k^{1+1/p}\epsilon_{\text{trot}}^{-1/p}\right)$ (for real time) and $L = \text{poly}(n)\mathcal{O}\left(\Upsilon k^{1+1/p}\epsilon_{\text{trot}}^{-1/p}\right)$ (for imaginary time, provided that $M \geqslant 4\tau\Upsilon\left(\sum_\gamma \|H_\gamma\|\right)$), where M is the Trotter variable. Υ denotes the number of *stages* in the Trotterization scheme of order p, and typically scales exponentially in p (but p is chosen a constant). For given order $p = O(1)$ of the Trotterization scheme, L in lemma 2.3 thus scales with the length of the time interval over which the system is simulated as $k^{1+o(1)}$ and $k^{1+o(1)}$, and with the imposed Trotter error as $\epsilon_{\text{trot}}^{-o(1)}$. Then, if we wish to estimate $g_R(k)$ and $g_I(k)$ at multiple $k = 0, \ldots, K$, we use that the probability that all K estimates are up to uncertainty ϵ equals unity minus the probability that at least one of the estimates is beyond ϵ (which, by the union bound, is at most $K\delta$).

Finally, before we move on to extracting eigenenergy estimates from the (real-time and imaginary-time) signals using the ESPRIT method, we prove the lemma related to the signal $g_D(k)$ in equation (3).

Lemma 2.4. *Let $g_D(k)$ be defined as in equation (3) for a local n-qubit Hamiltonian H, with E_j in $[0, \pi]$, and assume that (1) one can efficiently (i.e. with $\text{poly}(n)$ effort) sample from $|\Phi(x)|^2$, and (2) given x and y, one can compute $\Phi(y)/\Phi(x)$ efficiently. Then, $g_D(k)$ can be classically estimated within error ϵ with probability at least $1 - \delta$ with $[\text{poly}(n)]^k \times \Theta(\epsilon^{-2}\log(\delta^{-1}))$ classical computational effort.*

Proof. By definition of $g_D(k)$, we can write

$$g_D(k) = \sum_{x,y} |\Phi(x)|^2 \frac{\Phi(y)}{\Phi(x)}\langle x|(I - H/2\pi)^k|y\rangle. \tag{20}$$

To estimate $g_D(k)$, one first draws an x from $P(x) = |\Phi(x)|^2$, and then one collects all y which are obtained after the application of $(I - H/2\pi)^k$ to $\langle x|$. Each application of $I - H/2\pi$ maps the input string onto at most $\text{poly}(n)$ new output strings, hence one obtains at most $[\text{poly}(n)]^k$ such y's after k applications. Let $\boldsymbol{x} = (x_k = y, x_{k-1}, \ldots, x_1, x_0 = x)$ be a particular path of strings and let

$$\mathcal{R}(x) \equiv \sum_y \frac{\Phi(y)}{\Phi(x)}\langle x|(I - H/2\pi)^k|y\rangle$$

$$= \sum_{x_1,\ldots,x_{k-1},y} \frac{\Phi(y)}{\Phi(x)}\langle x|I - H/2\pi|x_1\rangle\langle x_1|I - H/2\pi|x_2\rangle \ldots \langle x_{k-1}|I - H/2\pi|y\rangle, \tag{21}$$

so that $g_D(k) = \sum_x |\Phi(x)|^2 \mathcal{R}(x)$. For each x that is sampled from $P(x)$, one thus computes and outputs $\text{Re}(\mathcal{R}(x))$ by summing over the contributions from *all* paths $\boldsymbol{x}$ that start at string x. As in lemma 2.3, we need to establish how many samples $|\Sigma|$ we need to draw from $P(x)$ to obtain $g_D(k)$ within error ϵ with probability at least $1 - \delta$. This analysis depends on the variance of the complex variable $\mathcal{R}(x)$ through equation (14), requiring us to upper bound

$$\mathbb{E}\big(|\mathcal{R}(x)|^2\big) = \sum_x |\Phi(x)|^2 \left(\sum_y \frac{\Phi^*(y)}{\Phi^*(x)}\langle y|(I - H/2\pi)^k|x\rangle\right)\left(\sum_{y'} \frac{\Phi(y')}{\Phi(x)}\langle x|(I - H/2\pi)^k|y'\rangle\right)$$

$$= \sum_x \langle\Phi|(I - H/2\pi)^k|x\rangle\langle x|(I - H/2\pi)^k|\Phi\rangle$$

$$= \langle\Phi|(I - H/2\pi)^{2k}|\Phi\rangle \leqslant 1, \tag{22}$$

where in the final line we have used that the eigenvalues of H lie in $[0, \pi]$. As in the proof of lemma 2.3, this establishes that $\text{Var}(\text{Re}(\mathcal{R}(x))) \leqslant 1$. Then we can use the median-of-means estimator as in the proof of lemma 2.3 and appendix C to establish that with probability at least $1 - \delta$, $g_D(k)$ can be estimated with error at most ϵ, taking $|\Sigma| = \Theta(\epsilon^{-2}\log(\delta^{-1}))$ samples from $P(x) = |\Phi(x)|^2$, and with $[\text{poly}(n)]^k$ computational effort per sample. $\square$

It is important to note that unlike in lemma 2.3, here we only sample x and compute the rest as the estimator $\mathcal{R}(x)$, while in lemma 2.3 we sample the whole path of length L. This is why the computational effort in lemma 2.3 is efficient (linear) in L and thus polynomial in k, while in lemma 2.4 the computational effort is exponential in k. This is thus the difference between the stoquastic Hamiltonian case versus the general Hamiltonian case. Note also that one can take each G_i in lemma 2.3 to be $G = I - H/2\pi$ in principle, as it obeys condition (b) when H is stoquastic.

We note that in [13] the sampling-access assumption is formulated slightly differently, that is, one gets access to $\Phi(x)$ for a given x, which can be stronger than only knowing the ratio $\Phi(x)/\Phi(y)$ for a given x and y. In addition, reference [13] allows an additional error in the sampling access whereas we gloss over this here and assume perfect sampling-access (similar to the exact assumptions in the other lemmas).

3. Classically processing the signal: the ESPRIT method

We turn to discussing the ESPRIT method [25] which is a method like the matrix pencil method [21, 22, 37] for processing a signal as in equations (1) and (2) consisting of S components. Indeed, suppose a set of values for the signal $g(k)$,

$$g(k) = \sum_{j=1}^{S} c_j z_j^k, \tag{23}$$

where $|z_j| \leqslant 1$, for $k \in \{0, 1, \ldots, K\}$, K even. The goal is to determine the z_j and the coefficients $c_j > 0$[5] using $g(k)$ for sufficiently many k. In case of the real-time signal $g_R(k)$, we have $z_j \equiv e^{-iE_j}$, in case of a purely-decaying imaginary-time signal $g_I(k)$, we have $z_j \equiv e^{-E_j} \in (e^{-2\pi}, 1]$ and for the purely-decaying signal $g_D(k)$ we have $z_j \equiv (1 - E_j/(2\pi)) \in [\frac{1}{2}, 1]$.

Due to sampling and Trotter noise, one is effectively given a noisy signal $y(k)$ (for $k \in \{0, 1, \ldots, K\}$), which is related to the original signal $g(k)$ by:

$$y(k) := g(k) + \eta(k) = \sum_{j=1}^{S} c_j z_j^k + \eta(k), \tag{24}$$

where $\eta(k)$ denotes e.g. the sampling and Trotter noise, and we have $|\eta(k)| \leqslant \epsilon_{\text{tot}}$ in theorems 1.1 and 1.2 with high probability.

It is well-known that for a noiseless signal ($\eta(k) = 0$), the z_j's and the c_j's can be resolved perfectly via ESPRIT and the matrix pencil method if we take $K + 1 \geqslant 2S$. Importantly, this result does not depend on whether the signal is oscillatory or decaying. For illustration, figure 2 depicts the results of application of the matrix pencil method to a noiseless signal. We consider separately a decaying signal and an oscillating signal, and for both cases we depict respectively the estimates of the decay rates and oscillation frequencies as a function of K. When $K + 1 \geqslant 2S$, the eigenvalues are indeed resolved both for the decaying and the oscillating signal. When $K + 1 < 2S$, the eigenvalues are not resolved. We will see however, that in the presence of noise a decaying or oscillatory signal fares very differently.

Let us consider in more detail the task of obtaining the z_j's from the signal $y(k)$ in equation (24). The key object of study here is the Hankel matrix $H(y) := H(g) + H(\eta)$, containing all K data points of the noisy signal $y(k)$ and a positive integer 'matrix pencil' parameter L:

$$H(y) = \begin{pmatrix} y(0) & y(1) & \ldots & y(K-L) \\ y(1) & y(2) & \ldots & y(K-L+1) \\ \vdots & \vdots & & \vdots \\ y(L) & y(L+1) & \ldots & y(K) \end{pmatrix}_{(L+1)\times(K-L+1)}$$

$$= \sum_{j=1}^{S} c_j \begin{pmatrix} 1 & z_j & \ldots & z_j^{K-L} \\ z_j & z_j^2 & \ldots & z_j^{K-L+1} \\ \vdots & \vdots & & \vdots \\ z_j^L & z_j^{L+1} & \ldots & z_j^K \end{pmatrix} + H(\eta), \tag{25}$$

---

[5] Here we focus on determining the z_j, but given the z_j one can determine the c_j as well and methods for analyzing the performance also exist for this [29].

Figure 2. Estimates of the decay rates (of a decaying signal) and oscillation frequencies (of an oscillating signal) as a function of K. The estimates are obtained from applying the matrix pencil method [21, 22, 37] to the noiseless signals $g(k) = \sum_{j=1}^{S} c_j z_j^k$, where $z_j = e^{-E_j}$ (in the case of the decaying signal) and $z_j = e^{-iE_j}$ (in the case of the oscillating signal) for $k = 0, 1, \ldots, K$ where $E_j \in [0, 2\pi]$. All c_j's are set equal to $1/S$ and the E_j's have been randomly produced. The eigenvalues are recovered for $K + 1 \geqslant 2S$.

where $H(\eta)$ is purely due to the noise and has norm $\|H(\eta)\|$. We can decompose the Hankel matrix $H(g)$ of the noiseless signal in terms of Vandermonde matrices V_L:

$$H(g) = \sum_{j=1}^{S} c_j \begin{pmatrix} 1 & z_j & \cdots & z_j^{K-L} \\ z_j & z_j^2 & \cdots & z_j^{K-L+1} \\ \vdots & \vdots & & \vdots \\ z_j^L & z_j^{L+1} & \cdots & z_j^K \end{pmatrix}_{(L+1)\times(K-L+1)} = V_L C V_{K-L}^T, \tag{26}$$

where

$$C \equiv \mathrm{diag}(c_1, c_2, \ldots, c_S), \tag{27}$$

and V_L is

$$V_L = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_S \\ \vdots & \vdots & & \vdots \\ z_1^L & z_2^L & \cdots & z_S^L \end{pmatrix}_{(L+1)\times S}. \tag{28}$$

In general, methods such as ESPRIT (see the ESPRIT algorithm 2) rely on the parameter L and for convenience we will keep it general in some of the analysis (specifically in appendix D). Our results will, however, focus on the choice $L = K/2$. For $L = K/2$, we have

$$H(y) = \begin{pmatrix} y(0) & y(1) & \cdots & y(K/2) \\ y(1) & y(2) & \cdots & y(K/2+1) \\ \vdots & \vdots & & \vdots \\ y(K/2) & y(K/2+1) & \cdots & y(K) \end{pmatrix}_{(K/2+1)\times(K/2+1)} \tag{29}$$

and

$$H(g) = V_{K/2} C V_{K/2}^T \in \mathbb{C}^{(K/2+1)\times(K/2+1)}. \tag{30}$$

Making contact with error bounds in the previous section, we see that (for $L = K/2$)

$$\forall \, k, \eta(k) \leqslant \epsilon_{\mathrm{tot}} \Rightarrow \|H(\eta)\| = \sigma_{\max}(H(\eta)) \leqslant \|H(\eta)\|_{\mathrm{F}} \leqslant K\epsilon_{\mathrm{tot}}. \tag{31}$$

From the 'Vandermonde decomposition' in equation (26) of the Hankel matrix encoding a real-time or imaginary-time signal, one can develop numerical algorithms to extract the decay rates z_i. One such

Algorithm 2. ESPRIT algorithm.

Data: Time signal y, number of decay rates or oscillation frequencies S
Result: List $\tilde{z}_1, \dots, \tilde{z}_S$
$K \leftarrow \text{length}(y)$;   `/* We will assume K is even for simplicity */`
$L \leftarrow K/2$;   `/* Not the most general choice, however it works well in practice */`
$H(y) \leftarrow$ Hankel matrix built from y;
$\tilde{U}, \tilde{\Sigma}, \tilde{W} \leftarrow \text{SVD}(H(y))$;   `/* Make sure $\tilde{\Sigma}$ is decreasingly ordered */`
$\tilde{U}_S \leftarrow$ First S columns of $\tilde{U}$;   `/* Remember $\tilde{U}$ is a$(L+1) \times (L+1)$unitary matrix */`
$\tilde{U}_0 \leftarrow$ First L rows of $\tilde{U}_S$;
$\tilde{U}_1 \leftarrow$ Last L rows of $\tilde{U}_S$;
$\tilde{\Psi} \leftarrow \tilde{U}_0^+ \tilde{U}_1$;   `/* Make $S \times S$ signal matrix $\tilde{\Psi}$,+denotes Moore-Penrose inverse */`
$\tilde{z}_1, \dots, \tilde{z}_S \leftarrow$ eigenvalues of signal matrix $\tilde{\Psi}$

---

algorithm is ESPRIT (given in algorithm 2), which specifically exploits the relation between the Vandermonde decomposition of $H(y)$ and its singular value decomposition.

We will see that this algorithm comes with recovery guarantees on the parameters $z_1, \dots, z_S$, in both the real-time and imaginary-time signal case, provided the noise vector η is small enough. The strength of these guarantees differs significantly between the two types of signal, and we will discuss them separately in the next sections. From the $\tilde{z}_j$'s we can then (for both the real-time and imaginary-time signal) extract $\tilde{E}_j$'s, which denote the S estimates for $\{E_i \in [0, 2\pi)\}_{i=1}^S$ returned by the classical post-processing algorithm. The error in the energy estimates is set as the optimal matching distance [3]

$$d(\{E_i\}, \{\tilde{E}_j\}) = \frac{1}{2\pi} \min_{\pi \in \text{Perm}_S} \max_j |\tilde{E}_{\pi(j)} - E_j|, \tag{32}$$

i.e. the returned list is optimally matched with the actual eigenvalues and the error is set by the largest mismatch.

3.1. Real-time (oscillatory) signal

In this section we discuss the performance of ESPRIT on real-time (oscillatory) signals. This performance has been well studied in the signal processing literature. Here, we will follow the analysis of [25], which provides theorem 3.1 relating $\|H(\eta)\|$ in equation (31) and the energy matching error defined in equation (32).

The performance of ESPRIT in the oscillatory signal case relies on lower bounding the smallest nonzero singular value of the Vandermonde matrix $V_{L=K/2}$ in equation (28), (or similarly upperbounding the condition number $\kappa(V_{K/2}) = \sigma_{\max}(V_{K/2})/\sigma_{\min}(V_{K/2})$). The smallest nonzero singular value of the Vandermonde matrix $V_{K/2}$ will depend on K, S and the location of the poles z_j. For the real-time signal, the z_j lie on the unit circle whereas for the imaginary-time signal the z_j lie in the interval $(e^{-2\pi}, 1]$. Let the minimal gap between the E_i be defined as

$$\Delta = \frac{1}{2\pi} \min_{j \neq k} |E_j - E_k|. \tag{33}$$

It has been proved [29] for $z_j = e^{-iE_j}$ that

$$\Delta \geqslant \frac{C}{K} \Rightarrow \sigma_{\min}^2(V_{K/2}) \geqslant \frac{C-1}{C}K, \tag{34}$$

for some constant $C > 1$. Note that if there are S eigenvalues $E_j \in [0, 2\pi)$ in the signal, it is clear that the minimal gap $\Delta \leqslant 1/S$, hence one should at least take $K \geqslant CS$. Based on this bound, theorem 4 in [25] says:

Theorem 3.1. ([25]). *Let* $(g + \eta)(k)$ *be a real-time signal with* $k = 0, \dots, K$, *and with* $g(k) = \sum_{i=1}^S c_i z_i^k$, $c_i > 0 \forall i$, $c_{\min} = \min_i c_i$ *and* $\eta(k)$ *a small noise vector. Let* $z_j = e^{-iE_j}$ *with* $j = 1, \dots, S$ *and* $E_j \in [0, 2\pi) \forall j$, *and* $K \geqslant 2C/\Delta$ *for some constant* $C > 2$ *with gap* Δ, *and* $K + 1 \geqslant 2S$. *If*

$$\|H(\eta)\| \leqslant c_{\min} K h_1(S, C, K), \tag{35}$$

with

$$h_1(S, C, K) = \frac{C-1}{8\sqrt{2S}C} \sqrt{1 - \frac{2CS}{(C-1)K}}, \tag{36}$$

then the ESPRIT algorithm outputs energy estimates $\{\tilde{E}_j\}$ *with distance*

$$d(\{E_i\}), \{\tilde{E}_j\}) \leqslant \|H(\eta)\| c_{\min}^{-1} K^{-1} h_2(S, C, K), \tag{37}$$

with

$$h_2(S, C, K) = 40\sqrt{2}S^2 \left(\frac{C}{C-1}\right)^{3/2} \left(1 - \frac{2CS}{(C-1)K}\right)^{-1}. \tag{38}$$

By equation (31) we have $\|H(\eta)\| \leqslant K\epsilon_{\text{tot}}$ and if we choose $K \sim S$, ϵ_{tot} can be chosen sufficiently small, inversely polynomial with S, such that at least equation (35) holds. Then $d(\{E_i\}, \{\tilde{E}_j\})$ will be $\Theta(\|H(\eta)\|S)$, hence decreasing like $S^2\epsilon_{\text{tot}}$.

If we combine this theorem with the quantum results of theorem 1.1, then we obtain theorem 1.3. These results thus form the theoretical underpinning of the ideas and numerical work in [32] in which QPE was replaced by the repeated execution of a circuit applying controlled-U^k (conditioned on an ancilla qubit state) which gets Trotterized to the overlap test circuit in figure 1.

Remark. It is noteworthy that even when the eigenvalues E_j are not well-separated but occur in 'clumps', results exist [25] which bound the performance of ESPRIT.

3.2. Imaginary-time (decaying) signal

Let us now discuss what information can be extracted from the imaginary-time signal in the presence of sampling and Trotter noise and compare this to the known theorem 3.1 for the real-time signal.

In appendix D we discuss in detail the recovery guarantees for ESPRIT for imaginary-time signals. This analysis is an adaptation of the work done in [25] for real-time signals, with the only true novelty being lemma D.7. However, since no rigorous analysis for imaginary-time signals exists in the literature we go through all the steps in considerable detail. The analysis will again depend on the condition number of the Vandermonde matrix $V_{L=K/2}$ in equation (28).

This condition number is much worse behaved, i.e. much larger, in case the z_i's all lie on the real axis—which is the case for the imaginary-time signal—but bounds on this condition number do exist [2]. Based on the work of Gautschi [12], we derive our own upper bounds on this condition number, which are asymptotically sub-optimal but have a clearer dependence on the choice of K and the given S than previous bounds in [2]. We then use the gap Δ to fill in the upper bound.

In analogy to theorem 3.1, we then obtain the following:

Theorem 3.2. *Let* $(g + \eta)(k)$ *be an imaginary-time decaying signal with* $k = 0, \ldots, K$, *and with* $g(k) = \sum_{i=1}^{S} c_i z_i^k$, $c_i > 0, \forall\, i$, $c_{\min} = \min_i c_i$, *and* $\eta(k)$ *a small noise vector. Let* $z_i = e^{-E_i}$ *with* $E_i \in [0, 2\pi)$ *and given eigenvalue gap* $\Delta < 1$ *in equation* (33), *and* $\{\tilde{E}_i\}$ *the energy estimates of ESPRIT with* $L = K/2$. *Let* $K + 1 \geqslant 2S$, K *even and* $K = TS$ *for some positive integer* T. *If we have*

$$\|H(\eta)\| \leqslant \frac{c_{\min}}{\sqrt{K}} g_1(S, \Delta), \tag{39}$$

with

$$g_1(S, \Delta) = \frac{1}{32S^2}(e^{-2\pi}\pi\Delta)^{3(S-1)}, \tag{40}$$

then

$$d(\{\tilde{E}_i\}, \{E_j\}) \leqslant \|H(\eta)\|c_{\min}^{-1}K\sqrt{K}g_2(S, \Delta), \tag{41}$$

with

$$g_2(S, \Delta) = e^{2\pi}640\sqrt{2}S^{5.5}(e^{-2\pi}\pi\Delta)^{-5(S-1)}. \tag{42}$$

Since the dependence on S is exponential in equation (41), one cannot make the distance $d(\{\tilde{E}_i\}, \{E_j\})$ small when the number of eigenvalues $S = \text{poly}(n)$, no matter what the gap. This is a crucial difference with the oscillatory real-time case. However, for $S = O(1)$, with sufficient, poly(n), effort one can make $\|H(\eta)\|$ sufficiently small to obey equation (39) and then reduce the error on the found eigenvalues to $1/\text{poly}(n)$. This assumes that the gap between the $O(1)$ rescaled eigenvalues present in the initial state is at least $1/\text{poly}(n)$ (and not exponentially small in n).

Furthermore, given that $\|H(\eta)\|$ should decrease at least as $\sim 1/\sqrt{K}$ through equation (39) but the upper bound in equation (41) scales as $\|H(\eta)\|K^{3/2}$, one obtains the optimal bound by choosing the *minimal K*, namely $K = 2S$, so that $L = K/2 = S$. In this case the Vandermonde matrix $V_{L-1} = V_{S-1}$ is square[6]. This expresses the intuitive fact that increasing K will not help beyond a point, as for larger K the signal simply dies out. This is unlike the oscillatory case of theorem 3.1 in which the optimal K is required to grow with $1/\Delta$. Here the bound does not require that K grows with $1/\Delta$, so there is no 'super-resolution'. We note that the upper bounds may have a sub-optimal dependence on K and S, which is due to the proof techniques.

---

[6] Hence, strictly speaking lemma D.11 is not much of a help.

Practically (roughly) speaking, whenever the condition number of the Vandermonde matrix $V_{L=K/2}$ grows by choosing a larger K, choosing that larger K can be beneficial.

For the other decaying signal ($g_D(k)$), a rather small change from $z_i = \exp(-E_i)$ to $z_i = 1 - E_i/2\pi$ gives:

Theorem 3.3. *Let* $(g + \eta)(k)$ *be a decaying signal with* $k = 0, \ldots, K$, *and with* $g(k) = \sum_{i=1}^{S} c_i z_i^k$, $c_i > 0, \forall\, i$, $c_{\min} = \min_i c_i$, *and* $\eta(k)$ *a small noise vector. Let* $z_i = 1 - E_i/2\pi$ *with* $E_i \in [0, \pi]$ *and given eigenvalue gap* $\Delta < 1$ *in equation* (33), *and* $\{\tilde{E}_i\}$ *the energy estimates of ESPRIT with* $L = K/2$. *Let* $K + 1 \geqslant 2S$, K *even and* $K = TS$ *for some positive integer T. If we have*

$$\|H(\eta)\| \leqslant \frac{c_{\min}}{\sqrt{K}}\tilde{g}_1(S, \Delta), \tag{43}$$

with

$$\tilde{g}_1(S, \Delta) = \frac{1}{32S^2}\Delta^{3(S-1)}, \tag{44}$$

then

$$d(\{\tilde{E}_i\}, \{E_j\}) \leqslant \|H(\eta)\| c_{\min}^{-1} K\sqrt{K}\tilde{g}_2(S, \Delta), \tag{45}$$

with

$$\tilde{g}_2(S, \Delta) = 640\sqrt{2}S^{5.5}\Delta^{-5(S-1)}. \tag{46}$$

Now to argue theorem 1.5 from theorem 3.3, we simply choose the minimal $K = 2S$, and since $S = O(1)$, it implies that the classical algorithm which estimates $g_D(k)$ for $k = 0, \ldots, K(= O(1))$ within error ϵ using lemma 2.4 requires poly(n) effort.

4. Spectral estimation for a transverse-field Ising chain

In this section, we numerically investigate the methods described thus far by applying them to an archetypal stoquastic Hamiltonian: the transverse field Ising chain. This system has been extensively studied [36] and will serve as a proof-of-principle test. The system consists of qubits on a one-dimensional lattice, which interact via an Ising interaction and are exposed to an external magnetic field in the transverse direction. The Hamiltonian associated with this system is:

$$H = -J\left(\sum_i Z_i Z_{i+1} + g\sum_i X_i\right), \tag{47}$$

where X, Y, Z denote the Pauli matrices, $J > 0$ (for a ferromagnetic interaction) and $g \geqslant 0$, so that H is term-wise stoquastic in the standard basis. We take the field to be pointing in the x-direction without loss of generality[7].

The system exhibits an abrupt change in the ground state of the system as a function of g at $g = 1$ (for $n \to \infty$). On either side of the phase transition, one has:

- **Strong-coupling limit** ($g \gg 1$). In this limit, the Hamiltonian is dominated by the magnetic field terms and the ground state is given by $|\psi_0\rangle \approx |+\rangle^{\otimes n}$. The p-particle excitations correspond to states $|-\rangle_{q_1}|-\rangle_{q_2}\cdots|-\rangle_{q_p}\prod_{i \neq q_1, q_2, \ldots, q_p}|+\rangle_i$, i.e., the ground state with spin flips at p sites $q_1, \ldots, q_p$ along the chain. These p-particle excited states are $\binom{n}{p}$-fold degenerate.
- **Weak-coupling limit** ($g \ll 1$). In this limit, the Hamiltonian is dominated by the Ising interaction terms and the (degenerate) ground state is given by either $|\psi_0\rangle \approx |0\rangle^{\otimes n}$ or $|\psi_0\rangle \approx |1\rangle^{\otimes n}$ (ferromagnetic phase). The excitations w.r.t. the ground state correspond to domain walls separating ferromagnetic regions of opposite spin.

To run the MC scheme described in lemma 2.3, the imaginary-time propagation operator e^{-kH} must be decomposed (by means of Trotterization) in terms of the local propagation operators $e^{-a_l k/M H_i}$ (where a_l and M are set by the Trotterization scheme)[8]. The local propagation operators acting on a subset of two qubits on

---

[7] The Hamiltonian can be transformed to $\tilde{H} = UHU^\dagger$ by the unitary transformation $U = \bigotimes_i \exp\left(\frac{i\theta Z_i}{2}\right)$, which alters the direction of the field in the transverse plane while preserving the spectrum.

[8] We note that the numerical results presented in this section are obtained using a first-order Trotter decomposition.

the chain are given by:

$$
e^{-\tilde{k}H_i} = \begin{pmatrix} \dfrac{\sinh(\lambda\tilde{k})}{\sqrt{1+g^2}} + \cosh(\lambda\tilde{k}) & 0 & \dfrac{g\,\sinh(\lambda\tilde{k})}{\sqrt{1+g^2}} & 0 \\[2mm] 0 & \dfrac{-\sinh(\lambda\tilde{k})}{\sqrt{1+g^2}} + \cosh(\lambda\tilde{k}) & 0 & \dfrac{g\,\sinh(\lambda\tilde{k})}{\sqrt{1+g^2}} \\[2mm] \dfrac{g\,\sinh(\lambda\tilde{k})}{\sqrt{1+g^2}} & 0 & \dfrac{-\sinh(\lambda\tilde{k})}{\sqrt{1+g^2}} + \cosh(\lambda\tilde{k}) & 0 \\[2mm] 0 & \dfrac{g\,\sinh(\lambda\tilde{k})}{\sqrt{1+g^2}} & 0 & \dfrac{\sinh(\lambda\tilde{k})}{\sqrt{1+g^2}} + \cosh(\lambda\tilde{k}) \end{pmatrix}, \quad (48)
$$

where $\lambda = J\sqrt{1+g^2}$ and $\tilde{k} = a_l k/M$. This operator is element-wise non-negative and can be efficiently brought to bock-diagonal form (with each block being irreducible).

Since the choice of $|\Phi\rangle$ directly governs which eigenvalues can be obtained from the real-time and imaginary-time evolution signals, it is a point of particular importance. In addition, the ability of ESPRIT to extract eigenvalues from the imaginary-time and real-time signals depends very strongly on the spectral gap between the eigenvalues in the signal. We consider a state $|\Phi\rangle$ which has considerable overlap with the ground state and the (n-fold degenerate) first excited state in the ($g > 1$)-regime. Since the gap between their associated eigenvalues increases monotonically as a function of g in this regime, this allows us to present the aforementioned gap dependence numerically. We shall call the state $|\Phi_{\text{optimal}}\rangle$ since in the ($g \gg 1$)-regime it optimally overlaps with the eigenstates of interest, i.e. $|\langle +^{\otimes n}|\psi_{p=0}\rangle|^2 = \sum_{q=1}^{n}|\langle +^{\otimes n}|\psi_{p=1,q}\rangle|^2 = \frac{1}{2}$. This state is given by:
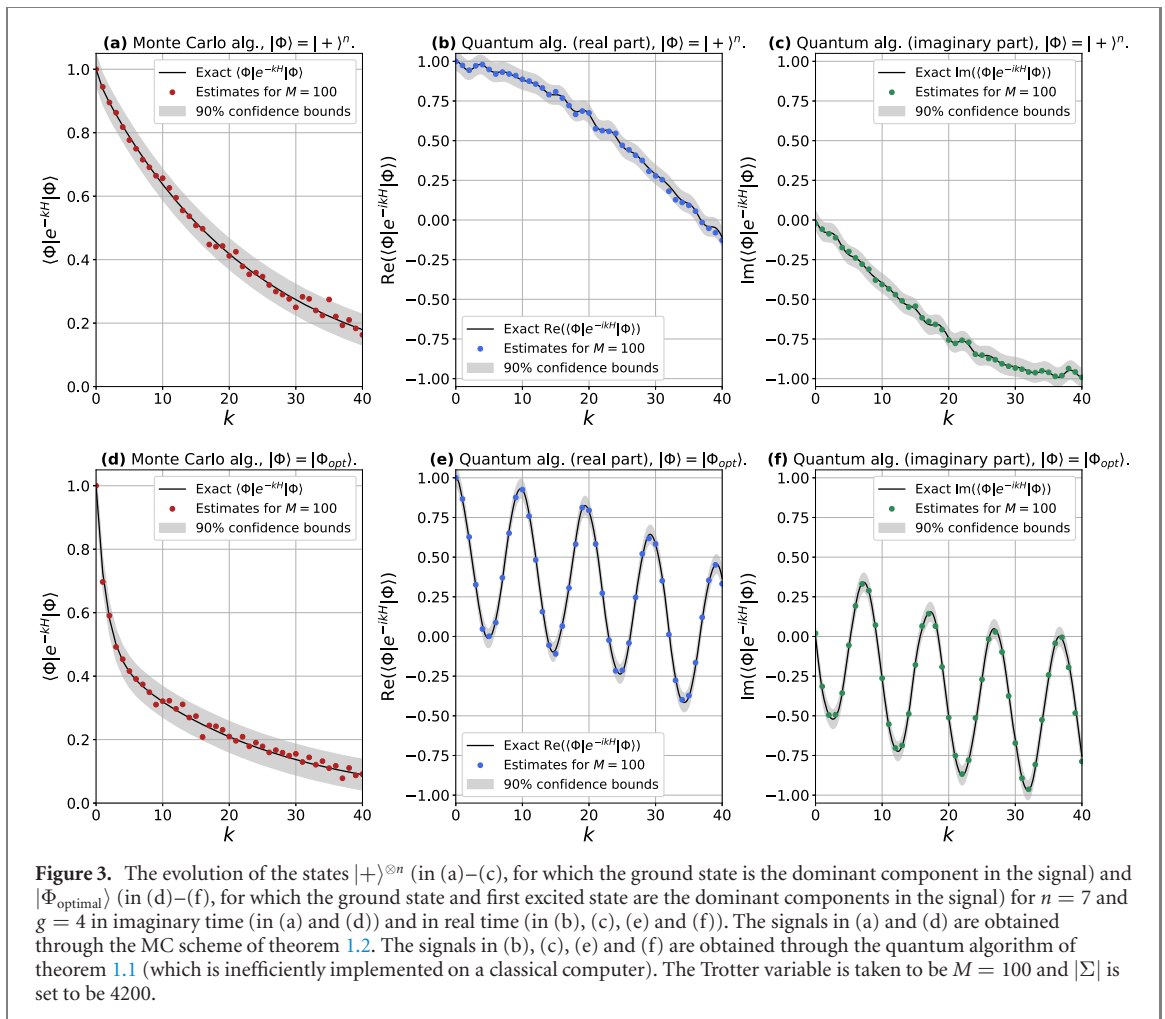
$$
|\Phi_{\text{optimal}}\rangle = \frac{1}{\sqrt{2}}\left(\underbrace{\prod_{i=1}^{n}|+\rangle_i}_{|\psi_{p=0}\rangle} + \sum_{q=1}^{n}\frac{1}{\sqrt{n}}\underbrace{|-\rangle_q\prod_{i\neq q}|+\rangle_i}_{|\psi_{p=1,q}\rangle} \right)
$$

$$
= \frac{1}{2^{(n+1)/2}}\sum_{q=1}^{n}\left(\left(\left(\frac{1}{n}+\frac{1}{\sqrt{n}}\right)|0\rangle_q + \left(\frac{1}{n}-\frac{1}{\sqrt{n}}\right)|1\rangle_q \right)\sum_{x\in\{0,1\}^{n-1}}|x\rangle \right), \quad (49)
$$

where $\sum_{x\in\{0,1\}^{n-1}}|x\rangle$ denotes an equal superposition of $(n-1)$-bit strings that exclude the bit in register q.

We note that for $|\Phi_{\text{optimal}}\rangle$, one can efficiently obtain $\frac{\Phi(y)}{\Phi(x)}$ for a given $x, y \in \{0,1\}^n$ and one can efficiently sample from $|\Phi(x)|^2$: from equation (49), one can infer a function $\Phi(x)$ ($\{0,1\}^n \to \mathbb{R}$) that (efficiently) gives the coefficient of the state $|\Phi_{\text{optimal}}\rangle$ associated with an n-bit string x: $\Phi(x) = 1/2^{(n+1)/2}$ $\left(\left(\frac{1}{n}+\frac{1}{\sqrt{n}}\right)(n-|x|) + \left(\frac{1}{n}-\frac{1}{\sqrt{n}}\right)|x| \right)$, so $\Phi(x)$ only depends on the Hamming weight $|x|$ of bit string x, i.e. the quantity $\frac{\Phi(y)}{\Phi(x)}$ can be efficiently determined. Furthermore, since $\Phi(x)$ only depends on n and $|x|$, the distribution $|\Phi(x)|^2$ also depends solely on these quantities. This implies that one can indeed efficiently sample from this distribution: first, one draws a Hamming weight $|x|$ from the distribution $|\Phi(x)|^2 = |\Phi(|x|)|^2$. Then, given $|x|$, one constructs at random an n-bit string with this Hamming weight. This latter step can be efficiently implemented by starting from some n-bit string with Hamming weight $|x|$ (such as $\{1\}^{|x|}\{0\}^{n-|x|}$) and then applying a random permutation.

4.1. Numerical method and results

We briefly discuss the details of the numerical analysis that is used to obtain the results presented in this section. We use the MC and quantum algorithms (where the latter is inefficiently implemented on a classical computer), which are presented in section 2 and summarized in theorems 1.2 and 1.1, to obtain resp. the imaginary-time and real-time evolution signals for the transverse-field Ising chain. We note that here we estimate the imaginary-time evolution signal using the empirical mean estimator, instead of the (asymptotically superior) median-of-means estimator. Having obtained these signals, we obtain estimates of the eigenvalues using the filtered ESPRIT method: this method corresponds to algorithm 2 in combination with an additional filtering step. This additional step is required since in principle the number of components in the signal S is not known *a priori* in the current setting. Therefore, we construct the matrix $\tilde{U}_S$ (in algorithm 2) by taking the first S columns of $\tilde{U}$, where S is now the number of singular values in the SVD of the Hankel matrix $H(y)$ that exceed $\mathsf{TF}\sigma_{\max}$. TF denotes what we call a truncation factor, and $\sigma_{\max}$ denotes the largest singular value of $H(y)$. In this way, the number of components in the signal emerges from the analysis of its Hankel matrix, rather than being a quantity that is known beforehand. By implementing

Figure 3. The evolution of the states $|+\rangle^{\otimes n}$ (in (a)–(c), for which the ground state is the dominant component in the signal) and $|\Phi_{\text{optimal}}\rangle$ (in (d)–(f), for which the ground state and first excited state are the dominant components in the signal) for $n = 7$ and $g = 4$ in imaginary time (in (a) and (d)) and in real time (in (b), (c), (e) and (f)). The signals in (a) and (d) are obtained through the MC scheme of theorem 1.2. The signals in (b), (c), (e) and (f) are obtained through the quantum algorithm of theorem 1.1 (which is inefficiently implemented on a classical computer). The Trotter variable is taken to be $M = 100$ and $|\Sigma|$ is set to be 4200.

the remainder of algorithm 2 as usual, we obtain estimates of the z_j's. From these estimates of the z_j's, we obtain the *spectral estimates* $\tilde{E}_j$ for the quantum algorithm and for the MC algorithm.

Note that this approach of including a filtering step—which often resembles more closely the practically encountered scenario when running the algorithms from lemmas 2.1 and 2.3—differs from that considered in theorems 3.1 and 3.2, where the number of components S in signals is known beforehand. Here, S is a quantity emerging in the analysis and it can even generally occur that components of the signal with very small coefficients—corresponding to eigenstates with very small overlap with $|\Phi\rangle$—are filtered out.

In the results presented in this section, note that the real-time and imaginary-time increments have been chosen such that all E_j that are present in the signals lie in $[0, 2\pi)$. This does *not* mean that the whole spectrum of the Hamiltonian lies in $[0, 2\pi)$, as the majority of its eigenvalues will not be present in the signals.
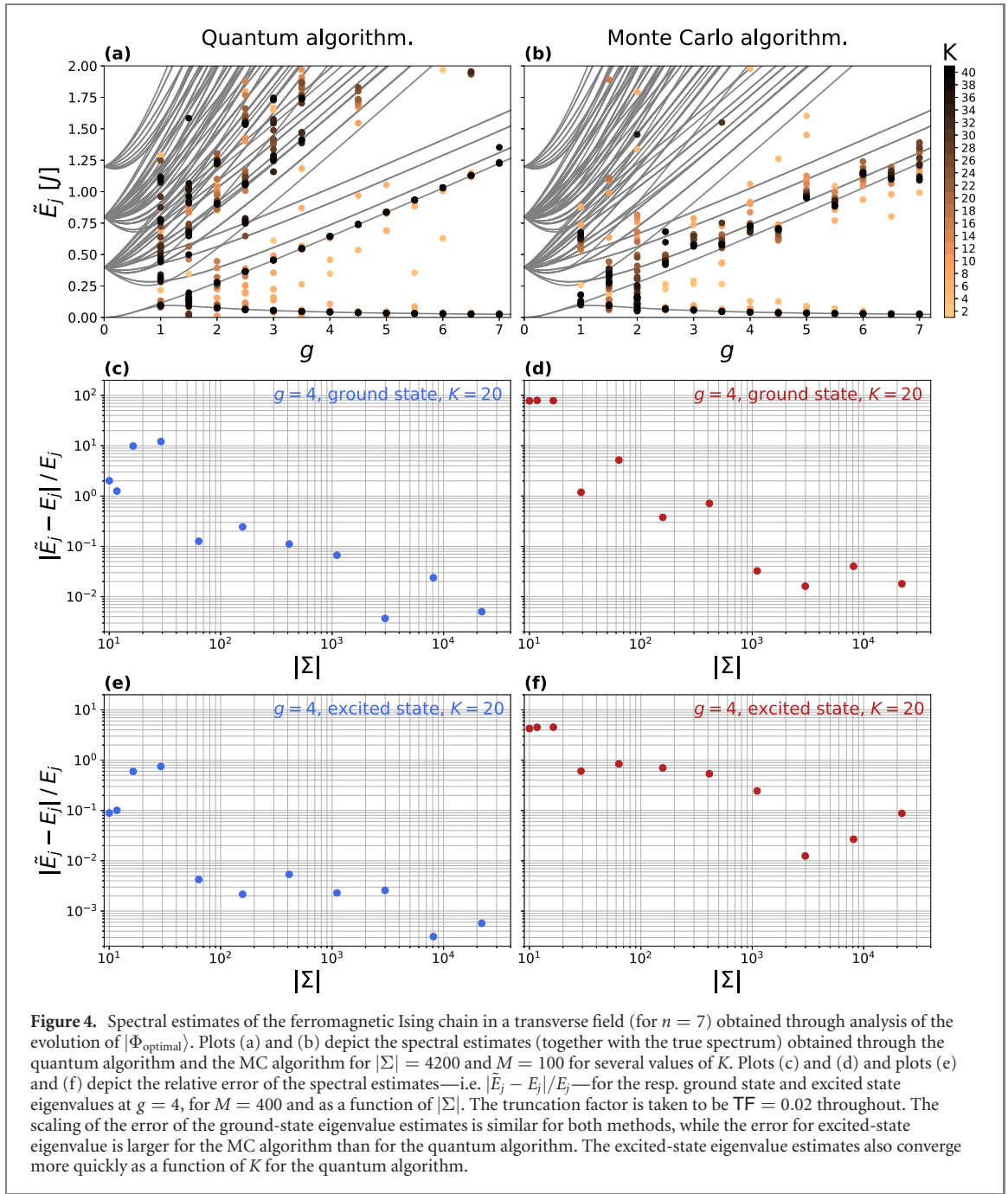
We note that for the quantum algorithm, the parameters $\{z_j\}$ have unit norm. However, due to finite sampling, one determines a noisy version of the signal $g_R(k)$, resulting in estimated eigenvalues of the Trotterized unitary having norms that slightly deviate from unity. To ensure that the estimates $\tilde{E}_j$ are real-valued, we take them to be the real parts of $i \log(\tilde{z}_j)$.

The code that is used to obtain the numerical results presented in this work can be found at [41].

In figure 3, the MC signals $\langle\Phi|e^{-kH}|\Phi\rangle$ and the real and imaginary parts of the quantum algorithm signals $\langle\Phi|e^{-ikH}|\Phi\rangle$ for $|\Phi\rangle = |+\rangle^{\otimes n}$ and $|\Phi_{\text{optimal}}\rangle$ are depicted.

The upper three figures correspond to $|\Phi\rangle = |+\rangle^{\otimes n}$. For this choice of $|\Phi\rangle$, the signals are clearly dominated by a single eigenvalue (the ground state eigenvalue): the MC signal decays with a single decay rate and the quantum algorithm signals oscillate with a single frequency. For the quantum algorithm signals, there are also higher-frequency components visible (due to $|+\rangle^{\otimes n}$ not having overlap with *only* the ground state).

For the lower three figures, we take $|\Phi\rangle = |\Phi_{\text{optimal}}\rangle$. For this choice of $|\Phi\rangle$, there are two eigenvalues present in the signals (the ground state and first excited state eigenvalues). For the MC signal, the excited state eigenvalue can be seen to die out within a few units of time, after which only the ground state component is left. The quantum algorithm signals can be seen to be composed of a high-frequency

Figure 4. Spectral estimates of the ferromagnetic Ising chain in a transverse field (for $n = 7$) obtained through analysis of the evolution of $|\Phi_{\text{optimal}}\rangle$. Plots (a) and (b) depict the spectral estimates (together with the true spectrum) obtained through the quantum algorithm and the MC algorithm for $|\Sigma| = 4200$ and $M = 100$ for several values of K. Plots (c) and (d) and plots (e) and (f) depict the relative error of the spectral estimates—i.e. $|\tilde{E}_j - E_j|/E_j$—for the resp. ground state and excited state eigenvalues at $g = 4$, for $M = 400$ and as a function of $|\Sigma|$. The truncation factor is taken to be $\mathsf{TF} = 0.02$ throughout. The scaling of the error of the ground-state eigenvalue estimates is similar for both methods, while the error for excited-state eigenvalue is larger for the MC algorithm than for the quantum algorithm. The excited-state eigenvalue estimates also converge more quickly as a function of K for the quantum algorithm.

(excited-state) component superposed on the ground-state component, where the excited-state component now obviously does *not* die out.

We now consider the spectral estimates that are obtained by applying ESPRIT to the evolution signals that are produced by the quantum algorithm (from theorem 1.1) and MC algorithm (from theorem 1.2). In particular, we determine both time evolution signals at a given total number of measurement points in real/imaginary time. We then determine the spectral estimates from both signals for increasing K, by including step-by-step more of the total number of measurement points in the analysis[9]. The truncation factor TF is taken to be equal to 0.02 throughout.

The top two plots in figure 4 depict, for a given $|\Sigma|$, the eigenvalue estimates as a function of g and for several values of K. For both the quantum algorithm and MC algorithm estimates, it is clear that a smaller spectral gap indeed requires a larger K for the eigenvalues to be obtained accurately. Furthermore, for a given $|\Sigma|$ and K, it is clear that the error of the estimate for the excited-state eigenvalue obtained from the imaginary-time signal is larger than that obtained from the real-time signal. We conclude furthermore that,

---

[9] For $K = 2$; $k = 0, 1, 2$. For $K = 4$; $k = 0, 1, 2, 3, 4$ etc.

in line with theorems 3.1 and 3.2, increasing K beyond a certain threshold does not necessarily reduce the error of the eigenvalue estimates.

It is apparent that as one approaches the $g = 1$ point, more higher-lying eigenvalues emerge from the ESPRIT analysis. This is especially true for the quantum algorithm (note that for the MC signal, the larger the eigenvalues are, the quicker the associated components in the signal die out). The appearance of these higher-lying eigenvalues can be attributed to the fact that (for finite n) the state $|\Phi_{\text{optimal}}\rangle$ starts to have significant overlap with states other than the two lowest-energy eigenstates in this regime.

The middle two and bottom two plots in figure 4 depict the relative error of the spectral estimates—i.e. $|\tilde{E}_j - E_j|/E_j$—for resp. the ground-state eigenvalue and excited-state eigenvalue (at fixed $g = 4$). We consider a range of values for $|\Sigma|$. For the ground-state eigenvalue, the scaling of the relative errors as a function of $|\Sigma|$ is similar for the quantum algorithm and the MC algorithm. Clearly, the relative errors of the excited-state eigenvalue estimates for the quantum algorithm are smaller than those for the MC algorithm.

We have also implemented the matrix pencil method in [18, 21] to estimate the eigenvalues from the real-time and imaginary-time signals. The only significant difference that was found between the estimates obtained through the ESPRIT method and through this matrix pencil method is that—in the $(K < 2S)$-regime—the matrix pencil method outputs estimates which resemble an average of the eigenvalues in the signal (as can be seen in figure 2 in a noiseless setting), while this is not the case generally for the ESPRIT method.

5. Discussion

We have considered the problem of obtaining (some) eigenvalues of local stoquastic—i.e. sign-problem-free—Hamiltonians and general local Hamiltonians H by means of tracking the evolution of the system state, differentiating between the evolution of the system state in real time and imaginary time. In both cases, we examine the use of the matrix pencil ESPRIT method in extracting eigenvalues of H from the state evolution signal. The real-time (oscillating) evolution signal is obtained through running quantum circuits, while the imaginary-time (decaying) signal for local stoquastic Hamiltonians is obtained through a MC scheme (developed in this work) that is implemented in a computationally tractable manner classically. Another type of decaying evolution signal—from which the ESPRIT method can extract eigenvalues of H—is obtained through a classical method for general local Hamiltonians that is similar in spirit to 'dequantization'.

We have invoked some known performance bounds of the ESPRIT method for the real-time signal and applied and extended bounds for the imaginary-time signal. Our bounds suggest that the ESPRIT method (or matrix pencil methods more generally) performs—not surprisingly—worse in extracting (multiple) eigenvalues from an imaginary-time decaying (MC algorithm) signal than from a real-time oscillating (quantum algorithm) signal in the presence of noise. However, we show that if the input state contains $S = O(1)$ eigenstates and the spectral gap is at least $1/\text{poly}(n)$, and the right access to the input state is available, the associated eigenvalues can be resolved efficiently (with $\text{poly}(n)$ classical effort) for local stoquastic as well as for general local Hamiltonians. Even though for $S = O(1)$, the classical effort for stoquastic as well as general Hamiltonians is $\text{poly}(n)$, the 'brute-force' algorithm for general Hamiltonians (in lemma 2.4) incurs an exponential cost in k in estimating the signal $g_{\text{D}}(k)$, while for stoquastic Hamiltonians the cost is polynomial in k. Despite this difference in cost, the error bounds for the eigenvalue estimates obtained here through analysis of the ESPRIT method applied to a decaying signal ($g_{\text{D}}(k)$ or $g_{\text{I}}(k)$) suggests that letting k grow as some function of n will generally not help.

Even though our results show that for these Hamiltonians, for an input state supported on $S = O(1)$ eigenvalues (separated by an at least $1/\text{poly}(n)$ gap), these eigenvalues can be estimated with $\text{poly}(n)$ classical effort, it remains to be better understood how *practical* this MC method for stoquastic Hamiltonians or the 'dequantization' method in lemma 2.4 are. The upper bounds for the errors on the eigenvalue estimates in theorem 3.2 grow rather fast with S (and the computational effort grows fast with k in lemma 2.4 for general local Hamiltonians), and it is not clear how much one can improve, say, the ESPRIT bounds.

Indeed, it would be interesting to show that the current bounds of ESPRIT for the imaginary-time decaying signal cannot be improved upon. There are definitely known negative results on the condition number of Vandermonde matrices [34], but there might be signal extraction algorithms that have better practical performance on decaying signals, or have looser requirements (such as the requirement that all data is evenly spaced). However, we suspect that the difficulty gap we observe between real-time and imaginary-time signal is universal. One possible way to argue this is through the Cramer–Rao bound (which has been analysed for real-time signals [40] but not for imaginary-time signals), which is a question we leave for further research.

In terms of numerical results, we find that: for a given spectral gap and sample size, the ability to distinguish between two eigenvalues indeed depends on the number of measurement points K at which the real-time and imaginary-time evolution signals are evaluated. The MC algorithm for stoquastic Hamiltonians and the quantum algorithm (in combination with the ESPRIT method) lead to a similar scaling of the relative error of the ground-state eigenvalue as a function of the sample size. However, for an excited-state eigenvalue, the quantum algorithm leads to significantly smaller relative errors than the MC algorithm. More extensive numerical studies, also of models other than the transverse-field Ising chain, may shed further light on whether the MC + ESPRIT method is useful in practice. For frustrated stoquastic Hamiltonians, even the smallest eigenvalue may lead to a fast decaying signal, requiring small sampling error and Trotter error in practice.

As for other directions of further research, one can ask whether a hybrid approach in which imaginary-time data from an error-free MC algorithm can strengthen the use of real-time data from a quantum algorithm obtained from a noisy quantum circuit. This approach requires combining the data where the poles/nodes $z_j = \mathrm{e}^{-iE_j}$ on the unit circle each have a partner pole $z'_j = \mathrm{e}^{-E_j}$ (or $z'_j = I - E_j/2\pi$) on the real axis. If the effect of noise can be modeled $z_j = \mathrm{e}^{-iE_j} \rightarrow \mathrm{e}^{iE_j - \gamma}$ [32], then the imaginary-time data may help in extracting the values for E_j. It may also be of interest to consider the case of sampling k for both the quantum circuit and MC method at random (instead of picking $k = 0, 1, \ldots, K$). Another direction of further research is the following. Suppose the input state has overlap with S (here not necessarily $O(1)$) eigenstates of the Hamiltonian, one could assess how well the ESPRIT methods succeeds in extracting e.g. the ground-state eigenvalue by filtering out all other components in the real-time or imaginary-time evolution signals. Another rather different direction of further research is the estimation of low-lying eigenvalues of Laplacian matrices of graphs (which relate to the connectivity of the graph), where we note that Laplacian matrices are stoquastic.

Acknowledgments

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/MStroeks/ClassQuantSimStoqHam.

Appendix A. Trotterization

Suppose $H = \sum_{i=1}^{N} H_i$ (where $N = \mathcal{O}(\mathrm{poly}(n))$) represents a k-local Hamiltonian of a quantum system. $\{H_i\}_{i=1}^{N}$ is generally a set of non-commuting terms but can be divided into subsets, such that within each subset all terms commute. For a given set $\{H_i\}_{i=1}^{N}$, we denote the minimum possible number of these subsets by Γ. This number of subsets is at most N and equals 1 in the trivial case where all H_i's commute with each other. The Hamiltonian H can thus be decomposed as $H = \sum_{\gamma=1}^{\Gamma} H_\gamma$, where all H_γ do not commute with each other, but the terms of which each individual H_γ is composed do commute. Choosing a decomposition into the minimum number of subsets brings about an additional advantage of parallelizability when implementing the evolution of the systems in imaginary or real time.

The following lemma (adaptation from [8]) upper bounds the errors of implementing imaginary-time and real-time state evolution through a first-order Trotter decomposition.

Lemma A.1. First-order Trotter decomposition. *Given a k-local Hamiltonian $H = \sum_{i}^{N} H_i$. Furthermore, suppose the set $\{H_i\}_{i=1}^{N}$ can be divided into a minimum of Γ subsets $\{H_\gamma\}_{\gamma=1}^{\Gamma}$, such that within each individual subset all H_i's commute. Then the quantities $\left| \langle \Phi | \mathrm{e}^{-itH} | \Phi \rangle - \langle \Phi | \left(\prod_\gamma \mathrm{e}^{-\mathrm{i}t\,H_\gamma/M} \right)^{M} | \Phi \rangle \right|$ and*

Figure A1. Absolute Trotter error (imposed on the signal estimate) as a function of the Trotter variable M for the imaginary-time and real-time signals. The noisy ($|\Sigma| = 200$) and Trotterized versions of $\langle\Phi|e^{-\tau H}|\Phi\rangle$ and $\langle\Phi|e^{-itH}|\Phi\rangle$ for a ferromagnetic Ising chain in a transverse field (for $g = 4$ and $n = 8$) are evaluated at $\tau = t = 3$ and several values of M. The Trotterization schemes are first-order N-term and first-order $\Gamma (= 2)$-term schemes. The associated error bounds are included in matching colors.

$\left| \langle\Phi|e^{-\tau H}|\Phi\rangle - \langle\Phi|\left(\prod_\gamma e^{-\tau H_\gamma/M}\right)^M|\Phi\rangle \right|$ *(where* $|\Phi\rangle$ *is a normalized state and* $t, \tau \in \mathbb{R}_+$*) are bounded as follows:*

$$\left| \langle\Phi|e^{-itH}|\Phi\rangle - \langle\Phi|\left(\prod_\gamma e^{-itH_\gamma/M}\right)^M|\Phi\rangle \right| \leqslant \sum_{\gamma'=1}^{\Gamma-1}\sum_{\gamma>\gamma'} \|[H_{\gamma'}, H_\gamma]\| \frac{t^2}{2M}, \tag{A.1a}$$

$$\left| \langle\Phi|e^{-\tau H}|\Phi\rangle - \langle\Phi|\left(\prod_\gamma e^{-\tau H_\gamma/M}\right)^M|\Phi\rangle \right| \leqslant 3\,e^2 \sum_{\gamma'=1}^{\Gamma-1}\sum_{\gamma>\gamma'} \|[H_{\gamma'}, H_\gamma]\| \frac{\tau^2}{2M}, \tag{A.1b}$$

where the second inequality holds provided that $\left\|e^{-\tau H/M}\right\| \leqslant 1$, $\left\|e^{-\tau H_\gamma/M}\right\| \leqslant 1$ $(\forall\,\gamma)$ *and* $\frac{\tau\left(\sum_\gamma \|H_\gamma\|\right)}{M} \leqslant 1$, *and* M *denotes the Trotter variable.*

To obtain a better scaling of the errors as a function of the Trotter variable M, one can employ higher-order Trotter decompositions. We denote the pth-order approximants of $e^{-itH/M}$ and $e^{-\tau H/M}$ by $\mathcal{T}_M(p, t)$ and $\mathcal{T}_M(p, \tau)$, respectively. We denote $\left| \langle\Phi|e^{-itH}|\Phi\rangle - \langle\Phi|\mathcal{T}_M(p, t)^M|\Phi\rangle \right|$ and $\left| \langle\Phi|e^{-\tau H}|\Phi\rangle - \langle\Phi|\mathcal{T}_M(p, \tau)^M|\Phi\rangle \right|$ by ϵ_{trot}. In [8], it was shown that, for general p, ϵ_{trot} is upper bounded as

follows:

$$\epsilon_{\text{trot}} \leqslant \mathcal{O}\big(\alpha t^{p+1}/M^p\big), \quad \text{for real-time evolution,} \tag{A.2a}$$

$$\epsilon_{\text{trot}} \leqslant \mathcal{O}\big(\alpha \tau^{p+1}/M^p\big), \quad \text{for imaginary-time evolution,} \tag{A.2b}$$

where $\alpha = \sum_{\gamma_1,\gamma_2,\dots,\gamma_{p+1}=1}^{\Gamma} \|[H_{\gamma_{p+1}},\dots,[H_{\gamma_2},H_{\gamma_1}]\dots]\|$ ($\alpha^{1/p}$ is typically poly(n)) and equation (A.2b) holds provided that $4\tau\Upsilon\big(\sum_\gamma \|H_\gamma\|\big)/M \leqslant 1$ (where Υ corresponds to the number of *stages* of the Trotter decomposition and typically scales exponentially in p)[10]. In [42], a widely used scheme is discussed for constructing pth-order approximants.

It is important to consider the total number of k-local propagation operators L required to simulate e^{-itH} and $e^{-\tau H}$ (for a given order p and Trotter variable M). For the scheme in [42], the number of these k-local propagation operators required to be implemented for the simulation of e^{-itH} and $e^{-\tau H}$ for $p > 1$ is $L = 2MN5^{\frac{p}{2}-1}$ (and for $p = 1$ is MN). If one wishes to obtain a given ϵ_{trot}, the number of k-local propagation operators into which the evolutions are decomposed scales as $L = \text{poly}(n)\mathcal{O}\big(\Upsilon t^{1+1/p}\epsilon_{\text{trot}}^{-1/p}\big)$ (for real time) and $L = \text{poly}(n)\mathcal{O}\big(\Upsilon \tau^{1+1/p}\epsilon_{\text{trot}}^{-1/p}\big)$ (for imaginary time). We thus conclude that for large p (i.e. high-order decompositions), L scales approximately linearly in the evolution time of the system under consideration (for real-time and imaginary-time evolution).

In figure A1, we have depicted the absolute error of noisy MC (imaginary-time) and QPE (real-time) signals at fixed $\tau = t$ as a function of M, obtained through first-order N-term and $\Gamma (= 2)$-term Trotterization schemes. We have included the first-order Trotter error bounds. We note that the apparent drastic increase in noise magnitude as a function of M is primarily due to the fact that the absolute error decreases as a function of M and is plotted on a logarithmic scale.

Appendix B. Extension to non-Hermitian propagation operators

In this Appendix we prove the following lemma, extending lemma 2.3:

Lemma B.1. *Let $\mathcal{F} \equiv \langle\Phi|G_1 G_2 \dots G_L|\Phi\rangle$, where:*

(a) *$|\Phi\rangle = \sum_{x=1}^{2^n} \Phi(x)|x\rangle$ is a normalized state of n qubits where $\Phi(x) \in \mathbb{C}$ ($\forall\, x$) and $\sum_x |\Phi(x)|^2 = 1$. We assume that (1) $\frac{\Phi(y)}{\Phi(x)}$ can be efficiently (poly(n)) calculated for a given x and y and (2) we can efficiently draw samples from the probability distribution $P(x) = |\Phi(x)|^2$.*

(b) *Each G_l is a k-local (possibly non-Hermitian) element-wise nonnegative matrix with singular values in $(0, 1]$.*

$\mathcal{F}$ can be estimated within error ϵ with probability at least $1 - \delta$ with a classical MC algorithm with runtime $\text{poly}(n) \times \Theta(\epsilon^{-2}\delta^{-1}) \times \Theta(L)$.

Proof. In addition to the n-qubit register, we exploit a single ancillary qubit. The matrices G_l are still element-wise non-negative. The state $|a\rangle$ denotes the state of the single ancillary qubit. By making use of the single ancillary qubit, the propagation operators can be symmetrized as follows:

$$F_l \equiv \begin{cases} G_l \otimes |0\rangle\langle 1| + G_l^\dagger \otimes |1\rangle\langle 0|, & \text{if } l \text{ is odd} \\[2mm] G_l \otimes |1\rangle\langle 0| + G_l^\dagger \otimes |0\rangle\langle 1|, & \text{if } l \text{ is even.} \end{cases} \tag{B.1}$$

In this form, F_l (the 'new' propagation operator) is element-wise non-negative, $k + 1$-local and Hermitian and hence one can apply lemma 2.3 to $\langle\Phi|F_1 F_2 \dots F_L|\Phi\rangle$, provided that its eigenvalues lie in $(0, 1]$. The eigenvalues λ of F_l (for l odd) can be found by solving:

$$\det\begin{pmatrix} -\lambda\mathbb{1} & G_l \\ G_l^\dagger & -\lambda\mathbb{1} \end{pmatrix} = \det\big(\lambda^2\mathbb{1} - G_l G_l^\dagger\big) = \det\big(G_l G_l^\dagger - \lambda^2\mathbb{1}\big) = 0, \tag{B.2}$$

where we have used that $G_l^{(\dagger)}$ commutes with $\mathbb{1}$ and that $G_l^{(\dagger)}$ is of even dimensionality. The eigenvalues of the Hermitian and positive semi-definite matrix $G_l G_l^\dagger$ are thus equal to λ^2. Since the singular values of G_l are equal to the square root of the eigenvalues of $G_l G_l^\dagger$, the eigenvalues of F_l will lie in $(0, 1]$ if the singular values of G_l lie in $(0, 1]$. This can be similarly shown for l even and this statement thus holds for all l.

---

[10] In the remainder of this discussion it is assumed that this condition is satisfied.

What is left to prove is that estimating the signal for the string of F_l's is equivalent to estimating the signal for the string of G_l's. Specifically, we want to prove the following identity: $G_1 G_2 \ldots G_L = \langle 0|F_1 F_2 \ldots F_L|L \bmod 2\rangle$, for $L \in \mathbb{Z}_+$. This is done below by means of induction.

- For $L = 1$:

$$\langle 0|F_1|1\rangle = \langle 0|\left(G_1 \otimes |0\rangle\langle 1| + G_1^\dagger \otimes |1\rangle\langle 0|\right)|1\rangle$$

$$= G_1 \langle 0|0\rangle\langle 1|1\rangle + G_1^\dagger\langle 0|1\rangle\langle 0|1\rangle \tag{B.3}$$

$$= G_1,$$

- Assuming $G_1 G_2 \ldots G_L = \langle 0|F_1 F_2 \ldots F_L|L \bmod 2\rangle$ holds for L, it holds for $L + 1$ as well:
  Making use of the definition in equation (B.1), we write F_{L+1} as follows:

$$F_{L+1} = G_{L+1} \otimes |L \bmod 2\rangle\langle L+1 \bmod 2| + G_{L+1}^\dagger \otimes |L+1 \bmod 2\rangle\langle L \bmod 2|. \tag{B.4}$$

The quantity of interest—in the case of the length of the operator string being $L + 1$—can now be rewritten as follows:

$$\langle 0|F_1 F_2 \ldots F_L F_{L+1}|L+1 \bmod 2\rangle = \langle 0|F_1 F_2 \ldots F_L|L \bmod 2\rangle G_{L+1} = G_1 G_2 \ldots G_L G_{L+1}, \tag{B.5}$$

which finishes the proof.

□

Appendix C. Median-of-means estimator

The MC scheme described in section 2 produces a set of $|\Sigma|$ samples $\{\boldsymbol{x}\}$ which are distributed according to $\Pi(\boldsymbol{x})$. For each sample, $\mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ can be evaluated and subsequently an estimate of $\mathcal{F}$ can be obtained. Only the first and second moments of the random variable $\mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ can be upper bounded in general. Therefore, if one would use the empirical mean $\mathrm{Re}(\tilde{\mathcal{F}}) = \frac{1}{|\Sigma|}\sum_{\boldsymbol{x} \in \Sigma} \mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ as a mean estimator for $\mathcal{F}$, then the best achievable scaling of $|\Sigma|$ such that

$$\mathrm{Pr}\left(\left|\mathrm{Re}(\tilde{\mathcal{F}}) - \mathcal{F}\right| \leqslant \epsilon\right) \geqslant 1 - \delta, \tag{C.1}$$

is $|\Sigma| = \Theta(\epsilon^{-2}\delta^{-1})$ (by means of Chebyshev's inequality).

Taking the median-of-means estimator [28] as estimator (instead of the empirical mean), one can obtain a more convenient scaling of $|\Sigma|$ w.r.t. δ (despite the fact that only the first two moments of $\mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ can be upper bounded). The median-of-means estimator can be constructed as follows: partition the set of MC samples Σ into q groups $s_1, \ldots, s_q$ of size approximately $|\Sigma|/q$. One then computes the empirical mean of $\mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ over the samples in each group separately (giving q unbiased estimators of $\mathcal{F}$) and takes the median of these empirical means. We denote the empirical mean for each group by $f_j = \frac{1}{|s_j|}\sum_{\boldsymbol{x} \in s_j} \mathrm{Re}(\mathcal{R}(\boldsymbol{x}))$ (for $j \in \{1, \ldots, q\}$) and denote the median of these empirical means by $\hat{\mathcal{F}} = \mathrm{M}(f_1, \ldots, f_q)$. The estimator $\hat{\mathcal{F}}$ is the median-of-means estimator.

We define the median of q real numbers $a_1, \ldots, a_q$ as $\mathrm{M}(a_1, \ldots, a_q) = a_i$ with a_i such that

$$|\{j : a_j \leqslant a_i\}| \geqslant q/2 \quad \wedge \quad |\{j : a_j \geqslant a_i\}| \geqslant q/2, \tag{C.2}$$

where we take the smallest i if multiple is obey this condition.

$\{\mathrm{Re}(\mathcal{R}(\boldsymbol{x}))\}$ are i.i.d. random variables with mean $\mathcal{F}$ and variance $\mathrm{Var}(\mathrm{Re}(\mathcal{R}(\boldsymbol{x}))) \leqslant 1$. Let q and $|\Sigma|/q$ be positive integers, then

$$\mathrm{Pr}\left(\left|\hat{\mathcal{F}} - \mathcal{F}\right| \leqslant \sqrt{4q/|\Sigma|}\right) \geqslant 1 - \mathrm{e}^{-q/8}. \tag{C.3}$$

So for $q = 8\log(\delta^{-1})$ and $|\Sigma| = 4q\epsilon^{-2} = 32\log(\delta^{-1})\epsilon^{-2}$, we have:

$$\mathrm{Pr}\left(\left|\hat{\mathcal{F}} - \mathcal{F}\right| \leqslant \epsilon\right) \geqslant 1 - \delta. \tag{C.4}$$

Note that the estimator $\hat{\mathcal{F}} = \mathrm{M}(f_1, \ldots, f_q)$ depends explicitly on the confidence since q scales with δ. Given that indeed $q = \Theta(\log(\delta^{-1}))$, the number of samples required to obtain equation (C.4) is $|\Sigma| = \Theta(\log(\delta^{-1})\epsilon^{-2})$ (which is an exponentially better scaling w.r.t. δ compared to that for the empirical mean estimator).

To see why equation (C.3) is true, see [28], note that one can apply Chebyshev's inequality to each of the empirical means f_j: with probability at least $3/4$, we have $|f_j - \mathcal{F}| \leqslant \sqrt{4q/|\Sigma|}$. If $\left|\hat{\mathcal{F}} - \mathcal{F}\right| \geqslant \sqrt{4q/|\Sigma|}$,

then, by definition of $\hat{\mathcal{F}}$, at least $q/2$ of the empirical means f_j satisfy $\left| f_j - \mathcal{F} \right| \geqslant \sqrt{4q/|\Sigma|}$. Hence the probability that $\left| \hat{\mathcal{F}} - \mathcal{F} \right| \geqslant \sqrt{4q/|\Sigma|}$ is upper bounded by the probability that a binomially distributed random variable with q draws and success probability $1/4$ exceeds $q/2$:

$$\Pr\left(\left| \hat{\mathcal{F}} - \mathcal{F} \right| \geqslant \sqrt{4q/|\Sigma|} \right) \leqslant \Pr\left(\mathrm{Bin}(q, 1/4) \geqslant q/2 \right) = \Pr\left(\mathrm{Bin}(q, 1/4) - \mathbb{E}\left(\mathrm{Bin}(q, 1/4) \right) \geqslant q/4 \right) \leqslant \mathrm{e}^{-q/8},$$

(C.5)

where we have used $\mathbb{E}\left(\mathrm{Bin}(q, 1/4) \right) = q/4$ and Hoeffding's inequality.

Appendix D. Performance of ESPRIT on the imaginary-time (decaying) signal

In this section we prove a series of lemmas that characterize the behaviour of the ESPRIT algorithm (algorithm 2) on an imaginary-time signal obtained with finite error. They are direct generalisations of the work done in [25], which leads up to theorem 3.1 for oscillatory signals, to signals composed of real exponential decays. We will see that the guarantees on the algorithm will be substantially weaker in this case. The end goal of this section is theorem 3.2 in the main text.

The argument decomposes roughly into two halves. In the first half we argue that the behaviour of ESPRIT is controlled by the smallest non-zero singular value of the Vandermonde matrix V_L. In the second half we argue that this smallest nonzero singular value can be controlled in terms of a gap condition on the energy eigenvalues of the imaginary-time signal.

We start by proving a short result on the smallest nonzero singular values of products of matrices.

Lemma D.1. *Let the smallest nonzero singular value of a matrix X be $\sigma_{\min}(X)$. For any matrix, X we have $\sigma_{\min}(X) := \|X^+\|^{-1}$, where X^+ is the Moore–Penrose pseudo-inverse of X, i.e. through the SVD, we have $\sigma_{\min}^{-1}(X) = \|X^+\|$, where $\|X\|$ is the operator norm (the largest singular value). Let A, B be (non-square) matrices such that $(AB)^+ = B^+ A^+$. Then we have that*

$$\sigma_{\min}(AB) \geqslant \sigma_{\min}(A)\sigma_{\min}(B). \tag{D.1}$$

Proof. By sub-multiplicativity of the operator norm, we have that

$$\sigma_{\min}(AB) = \left(\|(AB)^+\| \right)^{-1} = \left(\|B^+ A^+\| \right)^{-1} \geqslant \left(\|B^+\| \|A^+\| \right)^{-1} = \sigma_{\min}(B)\sigma_{\min}(A). \tag{D.2}$$

$\square$

We note that the product property on the Moore–Penrose pseudo-inverse does not hold for all matrices (unlike for the regular inverse). We will make use of the following sufficient condition:

Lemma D.2. ([15]). *Let A, B be matrices and let A have full column rank, and B have full row rank. Then we have $(AB)^+ = B^+ A^+$.*

Next, we argue that a small perturbation in the imaginary-time signal does not impact the space spanned by the first S left singular vectors of the Hankel matrix $H(g)$ too strongly, see the ESPRIT algorithm 2. It is a compressed version of lemmas 4 and 5 in [25] (which are formulated for real-time signals only, but hold more generally). To state this lemma we need to consider a freedom of choice in U_S and $\tilde{U}_S$ with $\tilde{U}_S$ as defined in the ESPRIT algorithm 2 and U_S its noise-free version. It is possible that U_S and $\tilde{U}_S$ are far apart as operators, even if the spaces they span are close together.

We solve this by not considering U_S proper, but rather a rotated version of U_S. As we will see, this rotation will not impact the actual output of ESPRIT which are the eigenvalues of the signal matrix $\tilde{\Psi}$. The rotated version of U_S is given through the $S \times S$ unitary operator $(O_2 O_1)^\dagger$, which is defined via the singular value decomposition of $U_S^\dagger \tilde{U}_S$, i.e.

$$O_1 U_S^\dagger \tilde{U}_S O_2 = D, \tag{D.3}$$

with $I_S \geqslant D \geqslant 0$ and D diagonal. The diagonal elements of the matrix D are cosines of the so-called *canonical angles*. We note that this internal rotation is performed implicitly in [25], whereas we have chosen to make it explicit at all times.

Lemma D.3. *Let $(g + \eta)(k)$ be an imaginary-time signal with $g(k) = \sum_{i=1}^{S} c_i z_i^k$ and $\eta(k)$ a small noise vector. Consider the associated Hankel matrices $H(g)$ and $H(g + \eta)$, with singular value decompositions $H(g) = U\Sigma W$ and $H(g + \eta) = \tilde{U}\tilde{\Sigma}\tilde{W}$, and label the matrix of the first S columns of U (resp. $\tilde{U}$) as U_S (resp. $\tilde{U}_S$). Finally, let $O_1 U_S^\dagger \tilde{U}_S O_2 = D$ with $I_S \geqslant D \geqslant 0$ be the singular value decomposition of $U_S^\dagger \tilde{U}_S$. If*

$$\|H(\eta)\| \leqslant \sigma_{\min}(H(g))/2, \tag{D.4}$$

then

$$\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\| \leqslant \frac{2\sqrt{2S}\|H(\eta)\|}{\sigma_{\min}(H(g))}. \tag{D.5}$$

Proof. First, we can observe that indeed $I_S \geqslant D$ as $\left\| O_1 U_S^\dagger \tilde{U}_S O_2 \right\| \leqslant \|O_1\| \left\| U_S^\dagger \right\| \|\tilde{U}_S\| \|O_2\| \leqslant 1$.

The proof follows from Wedin's $\sin\Theta$ theorem for perturbations of singular subspaces as well as Weyl's perturbation theorem for singular values, see e.g. [3, 39]. From this latter theorem we know that $|\sigma_i(H(g+\eta)) - \sigma_i(H(g))| \leqslant \|H(\eta)\| \leqslant \sigma_{\min}(H(g))/2$ where σ_i is the ith singular value (in order and some singular values can be zero). Let $\sigma_{\min}(H(g+\eta)) > 0$ be the kth singular value, and thus

$$\sigma_{\min}(H(g+\eta)) \geqslant \sigma_k(H(g)) - \sigma_{\min}(H(g))/2 \geqslant \sigma_{\min}(H(g))/2, \tag{D.6}$$

where the last inequality holds as $\sigma_k(H(g)) > 0$ and hence is at least $\sigma_{\min}(H(g))$. Hence we can use Wedin's theorem on singular values (theorem 3.4 in [24], setting $\delta = \alpha = \sigma_{\min}(H(g))/2$) to conclude that

$$\left\| (U_S^\perp)^\dagger \tilde{U}_S \right\| \leqslant \frac{2\|H(\eta)\|}{\sigma_{\min}(H(g))}, \tag{D.7}$$

where $U_S^\perp$ is the matrix formed from the $L+1-S$ other (besides U_S) columns of the noiseless U. To connect this to $U_S(O_2O_1)^\dagger - \tilde{U}_S$ we can make the following long calculation:

$$\begin{aligned}
\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\| &\leqslant \left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\|_{\mathrm{F}} \\
&= \Big[\mathrm{tr}(U_S(O_2O_1)^\dagger(O_2O_1)U_S^\dagger) + \mathrm{tr}(\tilde{U}_S\tilde{U}_S^\dagger) \\
&\quad - \mathrm{tr}\big((U_S(O_2O_1)^\dagger\tilde{U}_S^\dagger + \tilde{U}_S(O_2O_1)U_S^\dagger) \big) \Big]^{1/2} \\
&= \Big[\mathrm{tr}(U_S U_S^\dagger) + \mathrm{tr}(\tilde{U}_S\tilde{U}_S^\dagger) - 2\,\mathrm{tr}(D) \Big]^{1/2} \\
&\leqslant \Big[2\,\mathrm{tr}(\tilde{U}_S\tilde{U}_S^\dagger) - 2\,\mathrm{tr}(DD^\dagger) \Big]^{1/2} \\
&= \sqrt{2}\Big[\mathrm{tr}(\tilde{U}_S\tilde{U}_S^\dagger) - \mathrm{tr}\big((O_1 U_S^\dagger\tilde{U}_S O_2)(O_1 U_S^\dagger\tilde{U}_S O_2)^\dagger \big) \Big]^{1/2} \\
&= \sqrt{2}\Big[\mathrm{tr}(\tilde{U}_S\tilde{U}_S^\dagger) - \mathrm{tr}\big(U_S^\dagger\tilde{U}_S\tilde{U}_S^\dagger U_S \big) \Big]^{1/2} \\
&= \sqrt{2}\Big[\mathrm{tr}(\tilde{U}_S\tilde{U}_S^\dagger) - \mathrm{tr}\big(U_S U_S^\dagger\tilde{U}_S\tilde{U}_S^\dagger \big) \Big]^{1/2} \\
&= \sqrt{2}\Big[\mathrm{tr}\big(\tilde{U}_S\tilde{U}_S^\dagger \big) - \mathrm{tr}(\tilde{U}_S\tilde{U}_S^\dagger) + \mathrm{tr}\big(U_S^\perp(U_S^\perp)^\dagger\tilde{U}_S\tilde{U}_S^\dagger \big) \Big]^{1/2} \\
&= \sqrt{2}\left\| (U_S^\perp)^\dagger\tilde{U}_S \right\|_{\mathrm{F}} \leqslant \sqrt{2S}\left\| (U_S^\perp)^\dagger\tilde{U}_S \right\|. \tag{D.8}
\end{aligned}$$

In the second inequality we used that $\mathrm{tr}(U_S U_S^\dagger) = \mathrm{tr}(\tilde{U}_S\tilde{U}_S^\dagger) = S$ since U_S as well as $\tilde{U}_S$ consist of S orthonormal columns, and $D \geqslant D^2 = DD^\dagger$, since $I \geqslant D \geqslant 0$. In addition, at the end we use that $UU^\dagger = U_S U_S^\dagger + U_S^\perp(U_S^\perp)^\dagger = I$ as U is unitary. $\quad\square$

The next step is to bound the deviation of the ESPRIT signal matrix $\tilde{\Psi} = \tilde{U}_0^+ \tilde{U}_1$ from the rotated version of its noiseless variant $(O_2O_1)\Psi(O_2O_1)^\dagger = (O_2O_1)U_0^+ U_1(O_2O_1)^\dagger$ in terms of $\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\|$. Recall that U_0 (resp. U_1) are constructed by removing respectively the first or last row from the matrix U_S. Note also that only the eigenvalues of the signal matrix Ψ matter in the ESPRIT algorithm 2 and the additional unitary rotations O_2O_1 do not alter these eigenvalues. We first establish some intermediate result:

Lemma D.4. *Let A, B be matrices such that* $\mathrm{rank}(A) = \mathrm{rank}(B)$. *If* $\|A - B\| \leqslant \sigma_{\min}(A)/2$ *then*

$$\left\| A^+ - B^+ \right\| \leqslant \frac{1+\sqrt{5}}{2}\|A - B\|\left\| A^+ \right\|^2 = \frac{1+\sqrt{5}}{2}\frac{\|A - B\|}{\sigma_{\min}^2(A)}. \tag{D.9}$$

Proof. From theorem 4.1 in [45] we get that

$$\left\| A^+ - B^+ \right\| \leqslant \frac{1+\sqrt{5}}{2}\|A - B\|\left\| A^+ \right\|\left\| B^+ \right\|. \tag{D.10}$$

Furthermore, since $\|A - B\| \leqslant \frac{\sigma_{\min}(A)}{2} < \frac{1}{\|A^+\|}$, we have by lemma 3.1 in [45] that

$$\|B^+\| \leqslant \frac{\|A^+\|}{1 - \|A^+\|\|A - B\|} \leqslant \|A^+\|, \tag{D.11}$$

leading to the first inequality in equation (D.9) and the lemma follows. □

The next lemma establishes that if a (non-square) matrix is full rank, a sufficiently small perturbation does not *decrease* the rank (and hence rank is preserved). Note that full-rankness is really required, as an arbitrarily small perturbation can always *increase* the rank.

Lemma D.5. *Let A be an $m \times n$ ($m \leqslant n$) matrix of rank m and let B an $m \times n$ matrix s.t. $\|A - B\| \leqslant \sigma_{\min}(A)/2$. Then $\mathrm{rank}(A) = \mathrm{rank}(B)$.*

Proof. By construction, we have $\mathrm{rank}(A) \geqslant \mathrm{rank}(B)$. Moreover we have that the smallest singular value of B is at least $\sigma_{\min}(A) - \|A - B\|$, by Weyl's singular value perturbation theorem [39]. Thus by construction the smallest singular value of B is at least $\sigma_{\min}(A)/2$ which is strictly larger than 0 as A is full rank and thus B is also full rank, and thus $\mathrm{rank}(A) = \mathrm{rank}(B)$. □

Finally, we will require a bound on the smallest nonzero singular value of U_0. This is the only lemma where we deviate significantly from the work done in [25], where the corresponding result, lemma 3 in [25], makes explicit use of the fact that in their scenario all poles z_j lie on the unit circle (what we call the real-time, oscillatory signal). The bound we present here is simpler and more general and thus applies to both imaginary (decaying) as well as real-time (oscillatory) signals, but leads to a suboptimal dependence on the condition number of the Vandermonde matrix V_L defined in equation (28) (in particular $\sigma_{\min}(V_L)$). However, it is sufficient for our purpose. The lemma will use some essential properties of how the ESPRIT algorithm 2 works which we review first. Key to the functioning of ESPRIT is the fact that $H(g)$ has two decompositions

$$H(g) = U\Sigma W = V_L C V_{K-L}^T, \tag{D.12}$$

where V_L is the $(L+1) \times S$ Vandermonde matrix defined in equation (28) and the coefficient matrix C is given in equation (27). When $S \leqslant L \leqslant K + 1 - S$ (requiring $K + 1 \geqslant 2S$), V_L and V_{K-L} are full rank. Then V_L and U_S have an image of the same dimension, which means there exists an invertible matrix A such that

$$U_S = V_L A, \tag{D.13}$$

and thus

$$U_0 = V_{L-1}A, \quad U_1 = V_{L-1}ZA, \tag{D.14}$$

with $Z = \mathrm{diag}(z_1, \ldots, z_S)$. This implies that

$$\Psi = U_0^+ U_1 = A^{-1} V_{L-1}^+ V_{L-1} Z A = A^{-1} Z A, \tag{D.15}$$

and hence the eigenvalues of Ψ are the poles z_i.

Lemma D.6. *Let U_0 be the $L \times S$ matrix obtained from U_S by removing the last row. If the associated Vandermonde matrix V_{L-1} is of (full) rank S then so is U_0, and moreover*

$$\sigma_{\min}(U_0) \geqslant \frac{\sigma_{\min}(V_{L-1})}{\|V_L\|}. \tag{D.16}$$

Proof. We have

$$I_S = U_S^\dagger U_S = (V_L A)^\dagger V_L A = A^\dagger V_L^\dagger V_L A, \tag{D.17}$$

which means that the singular values of A are precisely inverse to those of V_L, or equivalently that A^+ has the same singular spectrum as V_L. Moreover, by assumption V_{L-1} has full column rank, and A is invertible so

$$\sigma_{\min}^{-1}(U_0) = \|U_0^+\| = \|(V_{L-1}A)^+\| = \|A^{-1} V_{L-1}^+\| \leqslant \|A^{-1}\|\|V_{L-1}^+\| = \frac{\|V_L\|}{\sigma_{\min}(V_{L-1})}, \tag{D.18}$$

which is the inverse of the lemma statement. □

With lemmas D.4–D.6 in hand, we can give a perturbation bound for the signal matrix $\tilde{\Psi}$. We will show that if $\tilde{U}_0$ does not deviate strongly from the rotated version of U_0, namely $U_0(O_2O_1)^\dagger$, then the noisy signal matrix is also close to the ideal (rotated) version.

Lemma D.7. *Let* $\Psi := U_0^+ U_1$, $\tilde{\Psi} = \tilde{U}_0^+ \tilde{U}_1$ *be the ideal and perturbed version of the signal matrix, respectively. Now assume that* $\left\| U_0(O_2O_1)^\dagger - \tilde{U}_0 \right\| \leqslant \sigma_{\min}(U_0)/2$, *where* O_2O_1 *is defined through the singular value decomposition of* $U_S^\dagger \tilde{U}_S$ *(i.e.* $O_1 U_S^\dagger \tilde{U}_S O_2 = D$*). With this assumption we have*

$$\left\| (O_2O_1)\Psi(O_2O_1)^\dagger - \tilde{\Psi} \right\| \leqslant 5 \left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\| \frac{\|V_L\|^2}{\sigma_{\min}^2(V_{L-1})}. \tag{D.19}$$

Proof. Following [25] we have

$$\left\| (O_2O_1)\Psi(O_2O_1)^\dagger - \tilde{\Psi} \right\| \leqslant \left\| (O_2O_1)U_0^+ \right\| \left\| U_1(O_2O_1)^\dagger - \tilde{U}_1 \right\| + \left\| (O_2O_1)U_0^+ - \tilde{U}_0^+ \right\| \left\| \tilde{U}_1 \right\|. \tag{D.20}$$

We have $\left\| \tilde{U}_1 \right\| \leqslant \left\| \tilde{U}_S \right\| = 1$, since $\tilde{U}_S^\dagger \tilde{U}_S = I_S$ and removing a row vector decreases the operator norm. Similarly we have $\left\| U_1(O_2O_1)^\dagger - \tilde{U}_1 \right\| \leqslant \left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\|$. Now note that by our initial assumption and lemma D.5 (with $A = U_0(O_2O_1)^\dagger$ and $B = \tilde{U}_0$) we have $\mathrm{Rank}(U_0(O_2O_1)^\dagger) = \mathrm{Rank}(\tilde{U}_0)$. This means that we can use lemma D.4 to conclude that

$$\left\| (O_2O_1)U_0^+ - \tilde{U}_0^+ \right\| \leqslant \frac{1+\sqrt{5}}{2} \frac{\left\| U_0(O_2O_1)^\dagger - \tilde{U}_0 \right\|}{\sigma_{\min}^2(U_0)} \leqslant \frac{1+\sqrt{5}}{2} \frac{\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\|}{\sigma_{\min}^2(U_0)}. \tag{D.21}$$

Hence we get

$$\left\| (O_2O_1)\Psi(O_2O_1)^\dagger - \tilde{\Psi} \right\| \leqslant \frac{\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\|}{\sigma_{\min}(U_0)} + \frac{1+\sqrt{5}}{2} \frac{\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\|}{\sigma_{\min}^2(U_0)}$$

$$\leqslant (\sigma_{\min}(U_0) + (1+\sqrt{5})/2) \frac{\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\|}{\sigma_{\min}^2(U_0)}$$

$$\leqslant \left(\frac{3}{2} + \sqrt{5} \right) \frac{\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\|}{\sigma_{\min}^2(U_0)}, \tag{D.22}$$

where we used $\sigma_{\min}(U_0) \leqslant \|U_0\| \leqslant 1$. Plugging in the lower bound on $\sigma_{\min}(U_0)$ (lemma D.6) and noting that $\frac{3}{2} + \sqrt{5} \leqslant 5$ we obtain the lemma statement. $\qquad\square$

Combining all of this we get the following general theorem. From now on we restrict ourselves to the case $L = K/2$:

Theorem D.8. *Let* $(g+\eta)(k)$ *be the signal with* $g(k) = \sum_{i=1}^S c_i z_i^k$ *and* $\eta(k)$ *a small noise vector to which we apply the ESPRIT algorithm* 2. *Consider then the associated Hankel matrices* $H(g)$ *and* $H(g+\eta)$, *with singular value decompositions* $H(g) = U\Sigma W$ *and* $H(g+\eta) = \tilde{U}\tilde{\Sigma}\tilde{W}$, *and label the matrix of the first S columns of U (resp. $\tilde{U}$) as U_S (resp. as $\tilde{U}_S$). Denote by U_0 (resp. U_1) the submatrix of U_S with the last row (resp. first row) removed and define the signal matrix $\Psi = U_0^+ U_1$ (similarly for $\tilde{\Psi}$). Let $L = K/2$ and $K + 1 \geqslant 2S$. Now, if*

$$4\|H(\eta)\| \leqslant c_{\min} \sigma_{\min}^2(V_{K/2}) \sigma_{\min}(V_{K/2-1}) \|V_{K/2}\|^{-1}, \tag{D.23}$$

then

$$\left\| (O_2O_1)\Psi(O_2O_1)^\dagger - \tilde{\Psi} \right\| \leqslant \frac{10\sqrt{2S}\|H(\eta)\| \|V_{K/2}\|^2}{\sigma_{\min}^4(V_{K/2-1})} c_{\min}^{-1}, \tag{D.24}$$

where O_2O_1 *is defined through the singular value decomposition of* $U_S^\dagger \tilde{U}_S$, *equation* (D.3).

Proof. We start from the requirement in lemma D.7 that $\left\| U_0(O_2O_1)^\dagger - \tilde{U}_0 \right\| \leqslant \left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\| \leqslant \sigma_{\min}(U_0)/2$. By lemma D.6, this is certainly satisfied if $\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\| \leqslant \sigma_{\min}(V_{K/2-1}) \|V_{K/2}\|^{-1}/2$. Moreover, from lemma D.3 we know that $\left\| U_S(O_2O_1)^\dagger - \tilde{U}_S \right\| \leqslant 2\sqrt{2S}\|H(\eta)\|\sigma_{\min}^{-1}(H(g))$ so now let's upperbound $\sigma_{\min}^{-1}(H(g))$ in terms of the smallest singular value of $V_{K/2}$. We have

$$\sigma_{\min}^{-1}(H(g)) = \left\| H(g)^+ \right\| = \left\| (V_{K/2}^T)^+ C^{-1} V_{K/2}^+ \right\| \leqslant \sigma_{\min}^{-2}(V_{K/2}) c_{\min}^{-1}, \tag{D.25}$$

where we used $c_{\min} = \min_i c_i$. This means that the condition

$$4\|H(\eta)\| \leqslant c_{\min} \sigma_{\min}^2(V_{K/2}) \sigma_{\min}(V_{K/2-1}) \|V_{K/2}\|^{-1}, \tag{D.26}$$

allows us to use lemma D.7

$$\left\|(O_2O_1)\Psi(O_2O_1)^\dagger - \tilde\Psi\right\| \leqslant 10\sqrt{2S}\frac{\|H(\eta)\|\|V_{K/2}\|^2}{\sigma_{\min}^2(V_{K/2})\sigma_{\min}^2(V_{K/2-1})}c_{\min}^{-1}$$

$$\leqslant \frac{10\sqrt{2S}\|H(\eta)\|\|V_{K/2}\|^2}{\sigma_{\min}^4(V_{K/2-1})}c_{\min}^{-1}, \tag{D.27}$$

where we also used the general fact about Vandermonde matrices [2, theorem 1] that $\sigma_{\min}(V_{K/2}) \geqslant \sigma_{\min}$ $(V_{K/2-1})$ (i.e. the smallest non-zero singular value of V_L grows monotonically with L). $\qquad\square$

We wish to translate the bound in theorem D.8 to a theorem on the distance between the inferred eigenvalues z_i and $\tilde z_i$. The argument is as follows. We know from the Bauer–Fike theorem and the fact that $(O_2O_1)\Psi(O_2O_1)^\dagger$ is diagonalizable (see [3, exercise VIII.3.2]) that the operator norm bound on $(O_2O_1)\Psi(O_2O_1)^\dagger$ implies a matching bound distance on its eigenvalues z_i, using equation (D.15). That is, we have

$$d(\{z_i\}, \{\tilde z_j\}) := \max_{\pi\in\mathrm{Perm}_S}\min_i|z_{\pi(i)} - \tilde z_i| \leqslant (2S-1)\kappa(A(O_2O_1)^\dagger)\left\|(O_2O_1)\Psi(O_2O_1)^\dagger - \tilde\Psi\right\|, \tag{D.28}$$

where $\kappa(A(O_2O_1)^\dagger) = \kappa(A) := \|A\|\|A^{-1}\|$ is the condition number of the invertible matrix A in equation (D.13). We have $A = V_{K/2}^+ U_S$ and since U_S is an isometry we know that

$$\kappa(A) = \|A\|\|A^+\| = \|V_{K/2}\|\left\|V_{K/2}^+\right\| = \frac{\|V_{K/2}\|}{\sigma_{\min}(V_{K/2})}, \tag{D.29}$$

and hence we get a bound on the matching distance of the eigenvalues in terms of known quantities, as expressed in the following theorem:

Theorem D.9. *Let $y(k) = (g+\eta)(k)$ $(k = 0,\ldots,K)$ be the signal with $g(k) = \sum_{i=1}^S c_iz_i$, let $\eta(k)$ a small noise vector, and $K+1 \geqslant 2S$ $(L = K/2)$. Let $\tilde z_i$ be the output of the ESPRIT algorithm. Then under the noise condition:*

$$4\|H(\eta)\| \leqslant c_{\min}\sigma_{\min}^2(V_{K/2})\sigma_{\min}(V_{K/2-1})\|V_{K/2}\|^{-1}, \tag{D.30}$$

we have

$$d(\{z_i\}, \{\tilde z_j\}) \leqslant (2S-1)\frac{10\sqrt{2S}\|H(\eta)\|\|V_{K/2}\|^3}{\sigma_{\min}^5(V_{K/2-1})}c_{\min}^{-1}. \tag{D.31}$$

This theorem thus holds for both the decaying as well as oscillatory signal.

In order to use the theorem, one has to lower bound $\sigma_{\min}(V_{L-1})$ (and more trivially, upper bound $\|V_L\|$ as well) for $L = K/2$. For complex poles z_i on the unit circle one can get very good bounds, assuming a gap, see equation (34).

To lower bound $\sigma_{\min}(V_{L-1})$ for a purely decaying signal, we start with the following characterization of *square* Vandermonde matrices with real poles due to Gautschi:

Theorem D.10. (Theorem 1 in [12]). *Let V_{S-1} be a square $S \times S$ Vandermonde matrix with S (unequal) real positive poles $z_1,\ldots,z_S$. Then ∞ norm of V_{S-1}^{-1} is*

$$\left\|V_{S-1}^{-1}\right\|_\infty := \max_i\sum_j\left|\left(V_{S-1}^{-1}\right)_{ij}\right| = \max_{i\in\{1,\ldots,S\}}\prod_{j=1,j\neq i}^S\frac{1+z_i}{|z_j - z_i|}. \tag{D.32}$$

Based on this theorem we can work out a very similar statement for non-square Vandermonde matrices V_{L-1} where L is a multiple of S. Note that this lemma does not depend on any gap.

Lemma D.11. *Let V_{ST-1} be an $ST \times S$ Vandermonde matrix (where T is a positive integer) with S (unequal) real positive poles $z_1,\ldots,z_S \leqslant 1$. Then we have*

$$\left\|V_{ST-1}^+\right\|_\infty \leqslant 2\left\|V_{S-1}^{-1}\right\|_\infty. \tag{D.33}$$

Proof. Note that

$$V_{ST-1} = \left(V_{S-1}^T \quad Z^S V_{S-1}^T \quad Z^{2S} V_{S-1}^T \quad \cdots \quad Z^{(T-1)S} V_{S-1}^T\right)^T, \tag{D.34}$$

with $Z = \mathrm{diag}(z_1,\ldots,z_S)$, using $(V_{S-1})_{ij} = z_j^{i-1}$ and $(Z^S V_{S-1}^T)_{ij} = z_i^{s+j-1}$.

The pseudo-inverse V_{ST-1}^+ of size $S \times ST$ can be directly calculated as

$$V_{ST-1}^+ = (I - Z^{2S})(I - Z^{2ST})^{-1}\left(V_{S-1}^{-1} \quad Z^S V_{S-1}^{-1} \quad Z^{2S} V_{S-1}^{-1} \quad \cdots \quad Z^{(T-1)S} V_{S-1}^{-1}\right), \tag{D.35}$$

using a geometric series. Hence $\left\|V_{ST-1}^+\right\|_\infty$ can be calculated to be

$$
\max_{i\in\{1,\ldots,S\}}\frac{1-z_i^{2S}}{1-z_i^{2ST}}\sum_{p=0}^{T-1}z_i^{pS}\prod_{j=1,j\neq i}^{S}\frac{1+z_i}{|z_j-z_i|}\leqslant\max_i\left[\frac{1-z_i^{2S}}{1-z_i^{2ST}}\frac{1-z_i^{ST}}{1-z_i^{S}}\right]\left\|V_{S-1}^{-1}\right\|_\infty
$$

$$
=\max_i\left[\frac{1+z_i^S}{1+z_i^{ST}}\right]\left\|V_{S-1}^{-1}\right\|_\infty\leqslant 2\left\|V_{S-1}^{-1}\right\|_\infty, \tag{D.36}
$$

which gives the lemma statement.   □

To apply this lemma, we use that $\sigma_{\min}^{-1}(V_{ST-1})=\left\|V_{ST-1}^+\right\|\leqslant\sqrt{S}\left\|V_{ST-1}^+\right\|_\infty$ so that

$$
\sigma_{\min}(V_{ST-1})\geqslant(2\sqrt{S})^{-1}\left\|V_{S-1}^{-1}\right\|_\infty^{-1}, \tag{D.37}
$$

for any T. Unlike the lower bound for the real-time signal which explicitly uses the gap Δ in equation (34), this bound does not improve with T. We will now use the gap to upper bound $\left\|V_{S-1}^{-1}\right\|_\infty$ given in equation (D.32), thus lower bounding $\sigma_{\min}(V_{ST-1})$ for any T. This is done in the proof of our final theorem 3.2 (and its slight adaptation theorem 3.3) restated here:

Theorem. *Let $(g+\eta)(k)$ be an imaginary-time decaying signal (of length K) with $g(k)=\sum_{i=1}^{S}c_iz_i^k$, $c_i>0\,\forall\,i$, $c_{\min}=\min_i c_i$ and $\eta(k)$ a small noise vector. Let $z_i=\mathrm{e}^{-E_i}$ with $E_i\in[0,2\pi)$ and given eigenvalue gap $\Delta<1$ in equation (33), and $\{\tilde{E}_i\}$ the energy estimates of ESPRIT with $L=K/2$. Let $K+1\geqslant 2S$, K even and $K=TS$ for some positive integer T. If we have*

$$
\|H(\eta)\|\leqslant\frac{c_{\min}}{\sqrt{K}}g_1(S,\Delta), \tag{D.38}
$$

with

$$
g_1(S,\Delta)=\frac{1}{32S^2}(\mathrm{e}^{-2\pi}\pi\Delta)^{3(S-1)}, \tag{D.39}
$$

then

$$
d(\{\tilde{E}_i\},\{E_j\})\leqslant\|H(\eta)\|c_{\min}^{-1}K\sqrt{K}g_2(S,\Delta), \tag{D.40}
$$

with

$$
g_2(S,\Delta)=\mathrm{e}^{2\pi}\sqrt{2}640S^{5.5}(\mathrm{e}^{-2\pi}\pi\Delta)^{-5(S-1)}. \tag{D.41}
$$

Proof. First, we can lift the bound on the eigenvalue distance $d(\{z_i\},\{\tilde{z}_j\})$ to one on energies $d(\{E_i\},\{\tilde{E}_j\})$ defined in equation (32), with $\tilde{E}_i:=-\log(\tilde{z}_i)$ and $E_i\in[0,2\pi)$ by noting that

$$
\frac{1}{2\pi}|E_{\pi(i)}-\tilde{E}_i|=\frac{1}{2\pi}|\log(z_{\pi(i)})-\log(\tilde{z}_i)|=\frac{1}{2\pi}|\log\big(1-(\tilde{z}_i-z_{\pi(i)})/\tilde{z}_i\big)|
$$

$$
\leqslant|(\tilde{z}_i-z_{\pi(i)})/\tilde{z}_i|\leqslant\mathrm{e}^{2\pi}|(\tilde{z}_i-z_{\pi(i)})|, \tag{D.42}
$$

using that $z_i\in(\mathrm{e}^{-2\pi},1]$ and $\tilde{z}_i\in(\mathrm{e}^{-2\pi},1]$. In particular, for the first inequality, let $x=(\tilde{z}_i-z_{\pi(i)})/\tilde{z}_i$. If $x<0$, $|\log(1-x)|\leqslant|x|$. If $x>0$, since $z_i\in(\mathrm{e}^{-2\pi},1]$, we have $x\leqslant 1-\mathrm{e}^{-2\pi}$, so that $|\log(1-x)|\leqslant 2\pi|x|$.

Second, let us now use the gap condition $|E_i-E_j|\geqslant 2\pi\Delta$ in equation (33). This leads to a gap condition on the z_i themselves through (assuming w.l.o.g. that $z_i\geqslant z_j$):

$$
2\pi\Delta\leqslant|E_i-E_j|=|\log(z_i/z_j)|, \tag{D.43}
$$

and thus

$$
\mathrm{e}^{2\pi\Delta}\leqslant\frac{z_i}{z_j}=\frac{z_i-z_j}{z_j}+1, \tag{D.44}
$$

which gives

$$
|z_i-z_j|=z_i-z_j\geqslant z_j(\mathrm{e}^{2\pi\Delta}-1)\geqslant\mathrm{e}^{-2\pi}2\pi\Delta. \tag{D.45}
$$

This implies through equation (D.32) that

$$
\left\|V_{S-1}^{-1}\right\|_\infty\leqslant(\pi\Delta)^{-(S-1)}\,\mathrm{e}^{2\pi(S-1)}, \tag{D.46}
$$

so that

$$
\sigma_{\min}(V_{ST-1})\geqslant(2\sqrt{S})^{-1}(\pi\Delta\mathrm{e}^{-2\pi})^{(S-1)}, \tag{D.47}
$$

for any integer T. We note that this lower bound on $\sigma_{\min}$ is exponentially small in S as $\Delta<1$.

Third, we need an upper bound on $\|V_L\|$ in order to use theorem D.9. For $z_j\in(0,1]$, we have

$$
\|V_L\|\leqslant\|V_L\|_\mathrm{F}=\left(\sum_{i,j}z_j^{2(i-1)}\right)^{1/2}\leqslant S^{1/2}\left(\sum_{i=1}^{L+1}z_\mathrm{max}^{2(i-1)}\right)^{1/2}=S^{1/2}\left(\frac{1-z_\mathrm{max}^{2(L+1)}}{1-z_\mathrm{max}^2}\right)^{1/2}\text{ which can tend to}
$$

$(S(L + 1))^{1/2}$ when $z_{\max} \to 1$, so we use the simple upper bound $(S(L + 1))^{1/2} \leqslant (SK)^{1/2}$. Putting all this together allows to translate equation (D.31) to equation (D.40). The condition on $\|H(\eta)\|$ in equation (D.30) then translates to the sufficient condition in equation (D.38) using that $\sigma_{\min}(V_{K/2}) \geqslant \sigma_{\min}(V_{K/2-1})$, the lower bound on $\sigma_{\min}(V_{K/2-1})$, and the upper bound on $\|V_{K/2}\|$. $\qquad\square$

The adapted version, theorem 3.3, is proved almost identically (but requires that all E_i are in principle bounded away from 2π):

Theorem. *Let $(g + \eta)(k)$ be an imaginary-time decaying signal (of length K) with $g(k) = \sum_{i=1}^{S} c_i z_i^k$, $c_i > 0, \forall\, i$, $c_{\min} = \min_i c_i$, and $\eta(k)$ a small noise vector. Let $z_i = 1 - E_i/2\pi$ with $E_i \in [0, \pi]$ and given eigenvalue gap $\Delta < 1$ in equation (33), and $\{\tilde{E}_i\}$ the energy estimates of ESPRIT with $L = K/2$. Let $K + 1 \geqslant 2S$, K even and $K = TS$ for some positive integer T. If we have*

$$\|H(\eta)\| \leqslant \frac{c_{\min}}{\sqrt{K}}\tilde{g}_1(S, \Delta), \tag{D.48}$$

with

$$\tilde{g}_1(S, \Delta) = \frac{1}{32S^2}\Delta^{3(S-1)}, \tag{D.49}$$

then

$$d(\{\tilde{E}_i\}, \{E_j\}) \leqslant \|H(\eta)\| c_{\min}^{-1} K \sqrt{K} \tilde{g}_2(S, \Delta), \tag{D.50}$$

with

$$\tilde{g}_2(S, \Delta) = 640\sqrt{2}S^{5.5}\Delta^{-5(S-1)}. \tag{D.51}$$

Proof. First, we convert the eigenvalue distance $d(\{z_i\}, \{\tilde{z}_j\})$ to one on energies $d(\{E_i\}, \{\tilde{E}_j\})$ defined in equation (32), with $\tilde{E}_i := 2\pi(1 - \tilde{z}_i)$, so

$$\frac{1}{2\pi}|E_{\pi(i)} - \tilde{E}_i| = |z_{\pi(i)} - \tilde{z}_i|. \tag{D.52}$$

Second, let us now use the gap condition $|E_i - E_j| \geqslant 2\pi\Delta$ in equation (33). This leads to a gap condition on the z_i themselves through:

$$|z_i - z_j| \geqslant \Delta. \tag{D.53}$$

This implies through equation (D.32) that

$$\left\|V_{S-1}^{-1}\right\|_\infty \leqslant \Delta^{-(S-1)} \tag{D.54}$$

so that

$$\sigma_{\min}(V_{ST-1}) \geqslant (2\sqrt{S})^{-1}\Delta^{(S-1)}, \tag{D.55}$$

for any integer T. Following identical steps as in the proof of the previous theorem then leads to the final statements. $\qquad\square$

ORCID iDs

M E Stroeks ⓘ https://orcid.org/0000-0001-8558-5914
J Helsen ⓘ https://orcid.org/0000-0001-7218-2585
B M Terhal ⓘ https://orcid.org/0000-0003-0218-6614

References

[1] Aharonov D and Grilo A B 2019 Stoquastic PCP vs randomness *2019 IEEE 60th Annual Symp. Foundations of Computer Science (FOCS)* pp 1000–23
[2] Bazán F S V 1999 Conditioning of rectangular Vandermonde matrices with nodes in the unit disk *SIAM J. Matrix Anal. Appl.* **21** 679–93
[3] Bhatia R 1997 *Matrix Analysis Graduate Texts in Mathematics vol 169* (Berlin: Springer)
[4] Bravyi S 2015 Monte Carlo simulation of stoquastic Hamiltonians *Quantum Inf. Comput.* **15** 1122–40
[5] Bravyi S and Gosset D 2017 Polynomial-time classical simulation of quantum ferromagnets *Phys. Rev. Lett.* **119** 100503
[6] Bravyi S, DiVincenzo D P, Oliveira R and Terhal B M 2008 The complexity of stoquastic local Hamiltonian problems *Quantum Inf. Comput.* **8** 361–85
[7] Cade C, Folkertsma M and Weggemans J 2022 Complexity of the guided local Hamiltonian problem: improved parameters and extension to excited states (arXiv:2207.10097 [quant-ph])
[8] Childs A M, Su Y, Tran M C, Wiebe N and Zhu S 2021 Theory of Trotter error with commutator scaling *Phys. Rev.* X **11** 011020
[9] Crosson E and Harrow A W 2021 Rapid mixing of path integral Monte Carlo for 1D stoquastic Hamiltonians *Quantum* **5** 395

[10] Crosson E and Slezak S 2020 Classical simulation of high temperature quantum Ising models (arXiv:2002.02232v1 [quant-ph])

[11] Foulkes W M C, Mitas L, Needs R J and Rajagopal G 2001 Quantum Monte Carlo simulations of solids *Rev. Mod. Phys.* **73** 33–83

[12] Gautschi W 1962 On inverses of Vandermonde and confluent Vandermonde matrices *Numer. Math.* **4** 117–23

[13] Gharibian S and Le Gall F 2021 Dequantizing the quantum singular value transformation: hardness and applications to quantum chemistry and the quantum PCP conjecture (arXiv:2111.09079 [quant-ph])

[14] Gharibian S, Hayakawa R, Le Gall F and Morimae T 2022 Improved hardness results for the guided local Hamiltonian problem (arXiv:2207.10250 [quant-ph])

[15] Greville T N E 1966 Note on the generalized inverse of a matrix product *SIAM Rev.* **8** 518–21

[16] Gubernatis J, Kawashima N and Werner P 2016 *Quantum Monte Carlo Methods: Algorithms for Lattice Models* (Cambridge: Cambridge University Press)

[17] Hangleiter D, Roth I, Eisert J and Roushan P 2021 Precise Hamiltonian identification of a superconducting quantum processor (arXiv:2108.08319 [quant-ph])

[18] Helsen J, Battistel F and Terhal B M 2019 Spectral quantum tomography *npj Quantum Inf.* **5** 74

[19] Helsen J, Roth I, Onrati E, Werner A H and Eisert J 2020 A general framework for randomized benchmarking (arXiv:2010.07974 [quant-ph])

[20] Horn R A and Johnson C R 2012 *Matrix Analysis* 2nd edn (Cambridge: Cambridge University Press)

[21] Hua Y and Sarkar T K 1990 Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise *IEEE Trans. Acoust. Speech Signal Process.* **38** 814–24

[22] Hua Y and Sarkar T K 1991 On SVD for estimating generalized eigenvalues of singular matrix pencil in noise *IEEE Trans. Signal Process.* **39** 892–900

[23] Ioannou M, Piddock S, Marvian M, Klassen J and Terhal B M 2020 Sign-curing local Hamiltonians: termwise versus global stoquasticity and the use of Clifford transformations (arXiv:2007.11964 [quant-ph])

[24] Li R C 1998 Relative perturbation theory: II. Eigenspace and singular subspace variations *SIAM J. Matrix Anal. Appl.* **20** 471–92

[25] Li W, Liao W and Fannjiang A 2020 Super-resolution limit of the ESPRIT algorithm *IEEE Trans. Inf. Theory* **66** 4593–608

[26] Lin L and Tong Y 2022 Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers *PRX Quantum* **3** 010318

[27] Lloyd S 1996 Universal quantum simulators *Science* **273** 1073–8

[28] Lugosi G and Mendelson S 2019 Mean estimation and regression under heavy-tailed distributions: a survey *Found. Comput. Math.* **19** 1145–90

[29] Moitra A 2015 Super-resolution, extremal functions and the condition number of Vandermonde matrices *Proc. 47th Annual ACM Symp. Theory of Computing* pp 821–30

[30] Motta M, Sun C, Tan A T K, O'Rourke M J, Ye E, Minnich A J, Brandão F G S L and Chan G K-L 2019 Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution *Nat. Phys.* **16** 205–10

[31] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge: Cambridge University Press)

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

[33] Onorati E, Werner A H and Eisert J 2019 Randomized benchmarking for individual quantum gates *Phys. Rev. Lett.* **123** 060501

[34] Pan V Y 2016 How bad are Vandermonde matrices? *SIAM J. Matrix Anal. Appl.* **37** 676–94

[35] Potts D and Tasche M 2013 Parameter estimation for nonincreasing exponential sums by Prony-like methods *Linear Algebr. Appl.* **439** 1024–39

[36] Sachdev S 2011 *Quantum Phase Transitions* 2nd edn (Cambridge: Cambridge University Press)

[37] Sarkar T K and Pereira O 1995 Using the matrix pencil method to estimate the parameters of a sum of complex exponentials *IEEE Antennas Propag. Mag.* **37** 48–55

[38] Somma R D 2019 Quantum eigenvalue estimation via time series analysis *New J. Phys.* **21** 123025

[39] Stewart G W 1991 Perturbation theory for the singular value decomposition *SVD and Signal Processing, II: Algorithms, Analysis and Applications* (Amsterdam: Elsevier) pp 99–109 https://users.math.msu.edu/users/iwenmark/Teaching/MTH995/Papers/SVD_Stewart.pdf

[40] Stoica P and Nehorai A 1989 MUSIC, maximum likelihood, and Cramer–Rao bound *IEEE Trans. Acoust. Speech Signal Process.* **37** 720–41

[41] Stroeks M E H M 2022 ClassQuantSimStoqHam https://github.com/MStroeks/ClassQuantSimStoqHam

[42] Suzuki M 1991 General theory of fractal path integrals with applications to many-body theories and statistical physics *J. Math. Phys.* **32** 400–7

[43] Svore K M, Hastings M B and Freedman M H 2014 Faster phase estimation *Quantum Inf. Comput.* **14** 306–28

[44] Wan K, Berta M and Campbell E T 2021 A randomized quantum algorithm for statistical phase estimation (arXiv:2110.12071 [quant-ph])

[45] Wedin P Å 1973 Perturbation theory for pseudo-inverses *BIT Numer. Math.* **13** 217–32