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

ME Stroeks\textsuperscript{1,2}, J Helsen\textsuperscript{3} and BM Terhal\textsuperscript{1,2,4}

\textsuperscript{1} QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands
\textsuperscript{2} Faculty of EEMCS, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
\textsuperscript{3} CWI & QuSoft, Science Park 123, 1098 XG Amsterdam, The Netherlands
\textsuperscript{4} JARA Institute for Quantum Information, Forschungszentrum Juelich, D-52425 Juelich, Germany

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

Abstract. 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 initial state) and can be classically 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 $S = \text{poly}(n)$ eigenvalues, assuming a $1/\text{poly}(n)$ gap between them, with poly$(n)$ quantum and classical effort through the quantum phase estimation circuits. We prove that our Monte Carlo scheme plus the ESPRIT method can resolve $S = O(1)$ eigenvalues, assuming a $1/\text{poly}(n)$ gap between them, with poly$(n)$ purely classical effort for stoquastic Hamiltonians. These results thus quantify some opportunities and limitations of classical Monte Carlo methods for spectral estimation of stoquastic Hamiltonians. We numerically compare these MC and QPE eigenvalue estimation schemes by implementing them for an archetypal stoquastic Hamiltonian system: the transverse field Ising chain.

Keywords: quantum information, stoquastic Hamiltonian, spectral estimation, Monte Carlo algorithm, quantum phase estimation, ESPRIT.
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 algorithms [29, 41, 30, 36], using the ability to simulate the real-time dynamics $e^{-iHt/\hbar}$ efficiently on a quantum computer via Trotterization [25].

Classical alternatives are provided by quantum Monte Carlo methods [14, 10]. The efficiency of quantum Monte Carlo 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 Monte Carlo algorithm to grow exponentially in the system size $n$, necessitating an exponential number of runs of the Monte Carlo 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: $\langle x | H | y \rangle \leq 0$, 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 Monte Carlo sampling as complexity results [6, 21, 1] and various algorithmic results [5, 4, 8, 9] 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 Monte Carlo methods. Conversely, can quantum algorithms, even for stoquastic Hamiltonians, provide an advantage over Monte Carlo 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 Hamiltonian in an input state via a quantum circuit versus via a classical Monte Carlo scheme.

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

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

for $k = 0, 1 \ldots, K$ and where $|\Phi\rangle$ is some pre-specified $n$-qubit input state. The estimation of $g_R(k)$ for various $k$ is a crucial step in the quantum phase estimation algorithm (QPE). In what follows we will fix $\Delta t$ so that the eigenstates $|\psi_j\rangle$ with nonzero or substantial overlap $\langle \psi_j | \Phi \rangle^2 > 0 \dagger$, showing up in the signal, have the

\dagger Of course, 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

---

2. Quantum scheme versus Monte Carlo scheme for spectral estimation

3. Classically processing the signal: the ESPRIT method

3.1 Real-time (oscillatory) signal

3.2 Imaginary-time (decaying) signal

4. Spectral estimation for a transverse-field Ising chain

4.1 Numerical method and results

5. Discussion

Appendix A Trotterization

Appendix B Extension to non-Hermitian propagation operators

Appendix C Median-of-means estimator

Appendix D Performance of ESPRIT on the imaginary-time (decaying) signal
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$(n)$ for overall efficiency.

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

$$g_I(k) = \langle \Phi | e^{-i H k} | \Phi \rangle = \sum_{j=1}^{2^n} |\langle \psi_j | \Phi \rangle|^2 \left(e^{-E_j} \right)^k,$$

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 $g_R(k)$ and the classical Monte Carlo cost of estimating $g_I(k)$ within error $\epsilon$ are approximately identical, although the assumptions on our knowledge/preparation costs of $|\Phi\rangle$ are slightly different in the two cases. Here are the two statements:

**Theorem 1.1.** For a local Hamiltonian acting on $n$ qubits, one can estimate $g_R(k)$ in Eq. (1) with probability at least $1 - \delta$ with sampling error $\epsilon$ and Trotter error $\epsilon_{\text{trot}}$ (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 $O(k^{1+o(1)} \delta^{o(1)}) \times \text{poly}(n)$ 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 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 $O(k^{1+o(1)} \delta^{o(1)}) \times \text{poly}(n)$ and each circuit is repeated $\Theta(\epsilon^{-2} \log(4\delta^{-1}) + \log(K))$ times.

**Theorem 1.2.** For a local stoquastic Hamiltonian acting on $n$ qubits, one can estimate $g_I(k)$ in Eq. (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 $O(k^{1+o(1)} \delta^{o(1)}) \times \text{poly}(n)$ 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 $|\langle \Phi | \Phi \rangle|^2$ can be efficiently 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 $\epsilon_{\text{tot}}$ 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 $O(k^{1+o(1)} \delta^{o(1)}) \times \text{poly}(n)$ and the number of times one runs the algorithm (for each $k$) is $\Theta(\epsilon^{-2} \log(\delta^{-1}) + \log(K))$.

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 eigenvectors 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 [35, 20, 33] (with equivalent methods known as ESPRIT and MUSIC). This method has been used for processing randomized benchmarking data [31, 17], quantum phase estimation [30], spectral tomography of superoperators [16], for processing experimental time-series data to identify Hamiltonian parameters [15] or generally in processing discretely-sampled detecting Ramsey signals.

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 $K + 1 \geq 2S$, 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 $E_j$. 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 obeys the assumptions in Lemmas 2.1 and 2.3 is not simple and is one of the known bottlenecks in using quantum phase estimation (and our Monte Carlo scheme) to determine spectral information in $H$. 

3
The goal is to quantify this precisely and show that, at least theoretically, a regime exists in which the Monte Carlo method is 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 and the performance of the ESPRIT method in Theorem 3.1 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 $1/poly(n)$. 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 $\Delta \geq C/K$ 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 poly$(n)$ effort, one can resolve the eigenvalues $\{E_j\}$ with some poly$(n)$ effort, at most $1/poly(n)$.

For local stoquastic Hamiltonians the combination of Lemma 2.3, via Theorem 1.2 and the performance of the ESPRIT method in Theorem 3.2 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) input state $|\Phi\rangle$ be $S = O(1)$, and each occurs with nonzero probability at least $1/poly(n)$. Furthermore, assume that the $S$ eigenvalues $\{E_i\}$, $E_i \in [0,2\pi)$ are sufficiently well-separated, i.e. at least by $\Delta \geq 1/poly(n)$ with some poly$(n)$. Then using a Monte Carlo 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 $1/poly(n)$.

An immediate open question is whether a result such as Theorem 1.4 would also hold for general local Hamiltonians under similar conditions. This would require dequantization of QPE in this setting and replacing the quantum state preparation assumption. Partial, $O(1)$ precision for the ground state, results in this direction have been obtained in [12].

To investigate practical aspects of the MC scheme and compare it to the quantum scheme, we numerically study the one-dimensional Ising chain in a transverse field $g$ [34] 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 $g_K(k)$ 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 Monte Carlo algorithm (Lemma 2.3) for stoquastic Hamiltonians with its proof. 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. In Section 4, we numerically compare the quantum scheme and the Monte Carlo scheme 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.

Very extensive literature exist on the Monte Carlo power method [14] 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 Monte Carlo scheme, under an additional ‘guiding state’ promise. In [28] 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 $B$ if all its off-diagonal elements are non-positive: $\langle x | H | y \rangle \leq 0$, for $x \neq y$ (and states $|x\rangle$, $|y\rangle$ being elements of basis $B$).
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, 21].

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

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 adaptations employing results in [21]), and we will refer to them simply as ‘stoquastic’.

For a matrix $X$ we will use the operator or spectral norm $\|X\| = \sqrt{\lambda_{\text{max}}(X^\dagger X)} = \sigma_{\text{max}}(X)$, where $\sigma_{\text{max}}(X)$ 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 $\|X\| \leq \sqrt{m}\|X\|_\infty$ and $\|X\| \leq \|X\|_F$.

## 2. Quantum scheme versus Monte Carlo scheme for spectral estimation

In this section we show how to estimate $g_k(n)$ on a quantum computer and $g_k(n)$ for stoquastic Hamiltonians via a Monte Carlo algorithm.

Lemma 2.1 states a well-known quantum subroutine, namely the Hadamard or overlap test, while a new result, a Monte Carlo version of routine, is proved in Lemma 2.3. After these Lemmas, the proofs of Theorems 1.1 and 1.2 are given at the end of this section.

We note that the overlap test is used in versions of quantum phase estimation 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 Refs. [30, 36, 24]. Here we basically follow this approach for the real-time quantum evolution, which can in addition be randomized to save on implementation costs, see [42].

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

(i) $| \Phi \rangle = \sum_{x=1}^{2^n} \Phi(x) | x \rangle$ is a state of $n$ qubits which can be generated by a poly($n$)-size quantum circuit.

(ii) Each $G_i$ is a $k$-local unitary matrix.

$\mathcal{F}$ can be estimated within error $\epsilon$ with probability at least $1 - \delta$ with a quantum circuit with $\Theta(\epsilon^{-2} \log(4\delta^{-1})) \times \Theta(L) + \text{poly}(n)$ 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 $R(\theta) = e^{-i\theta Z/2}$):

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

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:

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

$$= \begin{cases} \frac{1}{2} + \frac{1}{2} \text{Re} \langle \Phi | G_1 G_2 \ldots G_L | \Phi \rangle, & \text{for } \theta = 0, \\ \frac{1}{2} - \frac{1}{2} \text{Im} \langle \Phi | G_1 G_2 \ldots G_L | \Phi \rangle, & \text{for } \theta = \frac{\pi}{2}. \end{cases}$$

(4)
In the final expression, we have restricted ourselves to \( \theta = 0 \) and \( \theta = \frac{\pi}{2} \), which are the \( \theta \) 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 \( |\Sigma^0_0 = 0| \) and \( |\Sigma^0_0 = \pi/2| \) be the number of times the ancilla measurement returns 0 so that

\[
\mathcal{F} = \left( 2\frac{|\Sigma^0_0 = 0|}{|\Sigma|} - 1 \right) - i \left( 2\frac{|\Sigma^0_0 = \pi/2|}{|\Sigma|} - 1 \right)
\]

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

\[
\Pr\left( |\mathcal{F} - \mathcal{F}| \leq \epsilon \right) \geq \Pr\left( |\text{Re}(\mathcal{F} - \mathcal{F})| \leq \epsilon / \sqrt{2} \right) \Pr\left( |\text{Im}(\mathcal{F} - \mathcal{F})| \leq \epsilon / \sqrt{2} \right)
\]

\[
= \left( 1 - \Pr\left( |\text{Re}(\mathcal{F} - \mathcal{F})| \leq \epsilon / \sqrt{2} \right) \right) \left( 1 - \Pr\left( |\text{Im}(\mathcal{F} - \mathcal{F})| \leq \epsilon / \sqrt{2} \right) \right)
\]

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

\[
\geq 1 - 4 \exp\left( - |\Sigma| \epsilon^2 / 4 \right) = 1 - \delta,
\]

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

Next, we will consider a classical Monte Carlo 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 on only the same \( O(1) \) qubits as \( H_i \).
- If \( G_i \) is reducible, then we can write \( G_i = \oplus_{b=1}^B G_i^b \) with \( B \) 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 \( S_b^i \), where \( \cup_b S_b^i \subseteq \{0,1\}^n \).
- From the Perron-Frobenius Theorem (Theorem 8.4.4 in [18]) 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.

\[
|\phi_i^b\rangle = \sum_{x \in S_i^b} \phi_i^b(x) |x\rangle, \quad G_i^b |\phi_i^b\rangle = \lambda_i^b |\phi_i^b\rangle,
\]

Figure 1: Basic circuit with a single ancillary qubit and an \( n \)-qubit register (initialized in state \( |\Phi\rangle \)).
Lemma 2.3. We prove the following: only terms for which all the matrix elements in the product are non-zero contribute to the sum.

\[ \text{products of (non-negative) matrix elements of } G \text{ with } b \]

\[ \text{where } b \mid \]

\[ \text{acts nontrivially only on a subset of } O(1) \text{ qubits (and acts as } f \text{ on other qubits)} \text{ one can efficiently compute the blocks } G_i^b, \text{ its largest eigenvalue } \lambda_i^b \text{ and associated eigenstate } |\phi_i^b\rangle \text{ in each block } b. \]

We prove the following:

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

(i) \( |\Phi\rangle = \sum_{x=1}^{2^n} \Phi(x) |x\rangle \) 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) \( \frac{\Phi(x)}{\Phi(y)} \) can be efficiently (poly(n)) calculated for a given \( x \) and \( y \) and (2) we can efficiently draw samples from the probability distribution \( P(x) = |\Phi(x)|^2 \).

(ii) Each \( G_i = G_i \) is a \( k \)-local, positive-definite, (elementwise) nonnegative matrix with eigenvalues in \( (0, 1] \). \( \mathcal{F} \) can be estimated within error \( \epsilon \) 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 \( G_i \) operators in the final equality):

\[ \mathcal{F} = \sum_{x_0, x_1, \ldots, x_L} |\Phi(x_0)|^2 \frac{\Phi(x_L)}{\Phi(x_0)} \langle x_0| G_1 |x_1\rangle \langle x_1| G_2 |x_2\rangle \ldots \langle x_{L-1}| G_L |x_L\rangle, \tag{8} \]

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 \( |x_{i-1}\rangle \) to \( |x_i\rangle \) in this string a probability

\[ P_i(x_{i-1} \rightarrow x_i) = \frac{1}{\lambda_i^b} \langle x_{i-1}| G_i |x_i\rangle \frac{\phi_i^b(x_i)}{\phi_i^b(x_{i-1})}, \tag{9} \]

where \( b \) labels the sub-block in \( G_i = \oplus_b G_i^b \) which contains the strings \( x_{i-1} \) and \( x_i \). Here \( \phi_i^b(x) = \langle x_{i-1}| \phi_i^b \rangle \) with \( |\phi_i^b\rangle \) defined in Proposition 2.2.

The probability distribution \( P_i \) is thus non-negative as \( \lambda_i^b \in (0, 1], G_i \) is element-wise non-negative and \( \phi_i^b(x_i) > 0 \) and \( \phi_i^b(x_{i-1}) > 0 \). It can be shown to be normalized:

\[ \sum_{x_i} P_i(x_{i-1} \rightarrow x_i) = \sum_{x_i} \frac{1}{\lambda_i^b} \langle x_{i-1}| G_i |x_i\rangle \frac{\phi_i^b(x_i)}{\phi_i^b(x_{i-1})} = \sum_{x_i \in S_i^b} \frac{1}{\lambda_i^b} \langle x_{i-1}| G_i^b |x_i\rangle \frac{\phi_i^b(x_i)}{\phi_i^b(x_{i-1})} = \frac{1}{\phi_i^b(x_{i-1})} \frac{1}{\lambda_i^b} \langle x_{i-1}| G_i^b |\phi_i^b\rangle = \frac{1}{\phi_i^b(x_{i-1})} \frac{1}{\lambda_i^b} \langle x_{i-1}| \lambda_i^b |\phi_i^b\rangle = \frac{\langle x_{i-1}| \phi_i^b \rangle}{\phi_i^b(x_{i-1})} = 1. \tag{10} \]

We use \( P_i(x_{i-1} \rightarrow x_i) \) to rewrite \( \mathcal{F}(\tau) \) as:

\[ \mathcal{F}(\tau) = \sum_{x_0, x_1, \ldots, x_L} |\Phi(x_0)|^2 P_1(x_0 \rightarrow x_1) P_2(x_1 \rightarrow x_2) \ldots P_L(x_{L-1} \rightarrow x_L) \times \frac{|\Phi(x_L)|}{|\Phi(x_0)|} \prod_{i=1}^{L} \lambda_i^b \phi_i^b(x_i) \prod_{i=1}^{L} \phi_i^b(x_i), \tag{11} \]

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

\[
\sum_{x_0, x_1, \ldots, x_L} \Pi(x) = \sum_{x_0} |\Phi(x_0)|^2 \sum_{x_1} P_1(x_0 \rightarrow x_1) \sum_{x_L} P_L(x_{L-1} \rightarrow x_L) = 1. \tag{12}
\]

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

By thus sampling from the probability distribution \( \Pi(x) \) and obtaining a mean estimator for \( F(\tau) \) using the samples \( R(x) \), we can estimate \( F(\tau) \). We note that \( F(\tau) = E(R(x)) \). Since \( R(x) \in \mathbb{C} \), a mean estimator over a finite number of samples will generally be complex-valued. Since \( F(\tau) \in \mathbb{R} \), we will instead obtain a mean estimator using samples \( \text{Re}(R(x)) \). The mean estimator that we shall use to estimate \( F(\tau) \) is the median-of-means estimator \[26\]. Using a set \( \Sigma \) of samples \( \{x\} \) (distributed according to \( \Pi(x) \)), the median-of-means estimator is defined as follows: Divide the set \( \Sigma \) into \( q \) subsets \( s_1, \ldots, s_q \) of size approximately \( |\Sigma|/q \). Calculate the empirical mean of \( \text{Re}(R(x)) \) over the samples in each subset: \( f_j = \frac{1}{|s_j|} \sum_{x \in s_j} \text{Re}(R(x)) \) for \( j \in \{1, \ldots, q\} \) (each \( f_j \) is an unbiased estimator of \( F(\tau) \)). Now the median-of-means estimator is given by the median of these empirical means: \( \hat{F} = M(f_1, \ldots, f_q) \). See Appendix C for more details.

The algorithm that efficiently produces \( \hat{F} = M(f_1, \ldots, f_q) \) 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:** Efficiently obtaining a median-of-means estimate of \( F(\tau) \) through sampling of the probability distribution \( \Pi(x) \).

| **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 \( F(\tau) \). |

for \( \sigma \in \{1, 2, \ldots, |\Sigma|\} \) 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 \( S_{1(1)}^{b(1)} \).

for \( l \in \{1, \ldots, L\} \) do

Pick a state \( |x_l\rangle \in S_{l(1)}^{b(l)} \) with probability \( P_l(x_{l-1} \rightarrow x_l) = \frac{1}{\lambda_l^{b(l)}} \langle x_{l-1} | G_l | x_l \rangle \frac{\phi_{l(1)}^{b(l)}(x_l)}{\phi_{l(1)}^{b(l)}(x_{l-1})} \). The state \( |x_l\rangle \) is part of \( S_{l+1(1)}^{b(l+1)} \) (for \( l < L \)).

end

Given \( \{x_l\}_{l=1}^L \) (sampled from \( \Pi(x) \)), calculate \( R_\sigma(x) = \frac{\Phi(x_L)}{\Phi(x_0)} \prod_{l=1}^L \lambda_l^{b(l)} \frac{\phi_{l(1)}^{b(l)}(x_{l-1})}{\phi_{l(1)}^{b(l)}(x_l)} \).

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 \( f_j = \frac{1}{|s_j|} \sum_{x \in s_j} \text{Re}(R(x)) \).

end

Output \( \hat{F} = M(f_1, \ldots, f_q) \).

Algorithm 1 thus efficiently provides an estimate of \( F(\tau) \) (albeit biased). To complete the proof, we will show that the variance of \( \text{Re}(R(x)) \in \mathbb{R} \) 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 \( Z = R(x) \), \( \mathbb{E}(Z) = \mathbb{E}(\text{Re}(Z)) + i \mathbb{E}(\text{Im}(Z)) \) and \( \text{Var}(Z) = \text{Var}(\text{Re}(Z)) + \text{Var}(\text{Im}(Z)) \geq \text{Var}(\text{Re}(Z)) \). Hence we can bound the variance of random variable \( \text{Re}(R(x)) \) by bounding the variance of the random variable \( R(x) \). This variance is given by:

\[
\text{Var}(R(x)) = \mathbb{E} \left( |R(x)|^2 \right) - \mathbb{E} \left( |R(x)| \right)^2 
\leq \mathbb{E} \left( |R(x)|^2 \right) - \mathbb{E} \left( |R(x)| \right)^2, \tag{13}
\]

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

\[
\mathbb{E}
\left(
\left| \mathcal{R}(x) \right|^2
\right) = \sum_x \Pi(x) \left| \mathcal{R}(x) \right|^2
\]

\[
= \sum_x \left| \Phi(x) \right|^2 \left( \langle x_0 | G_1 | x_1 \rangle \langle x_1 | G_2 | x_2 \rangle \cdots \langle x_{L-1} | G_L | x_L \rangle \prod_{l=1}^{L} \frac{\lambda_l^{(l)} \phi_l^{(l)}(x_l-1)}{\phi_l^{(l)}(x_l)} \right)
\]

(14)

\[
= \sum_x \left| \Phi(x) \right|^2 Q_1(x_0, x_1) Q_2(x_1, x_2) \cdots Q_L(x_{L-1}, x_L),
\]

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

\[
\sum_x Q_l(x, y) = \sum_x \langle x | G_l | y \rangle \frac{\phi_l^*(x)}{\phi_l^*(y)} = \sum_{x \in S_\| \Sigma \|} \left( \frac{\lambda_l^{(l)} \phi_l^*(x)}{\phi_l^*(y)} \right) = \left( \lambda_l^{(l)} \right)^2 \leq 1.
\]

(15)

\( Q_l(x, y) \) thus satisfies \( 0 \leq Q_l(x, y) \leq 1, \forall x, y \) and \( \forall l \in \{1, 2, ..., L\} \). By consecutively exploiting the property in Eq. (15) for all \( Q_l \)'s and the normalization property of state \( |\Phi\rangle \) in the expression in Eq. (14), we obtain

\[
\text{Var}(\mathcal{R}(x)) \leq \mathbb{E}
\left( \left| \mathcal{R}(x) \right|^2 \right) \leq 1. \quad \Rightarrow \quad \text{Var}(\text{Re}(\mathcal{R}(x))) \leq 1.
\]

(16)

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 \( |f_j - \mathcal{F}| \leq \sqrt{\text{Var}(\text{Re}(\mathcal{R}(x)))}/\sqrt{4q} \leq \sqrt{\| \Sigma \|/4q} \) with probability at least \( 3/4 \). Using Hoeffding's inequality and the definition of the mean, one can show that (see Appendix C):

\[
\text{Pr}
\left( |\bar{f} - \mathcal{F}| \leq \sqrt{4q/\| \Sigma \|} \right) \geq 1 - e^{-q/8}.
\]

(17)

Hence \( \mathcal{F} \) can be estimated with error \( \epsilon \) 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 \( \text{poly}(n) \). This completes the proof of Lemma 2.3. \( \blacksquare \)

**Remark.** Note that if one would have chosen the empirical mean \( \bar{f} = \frac{1}{\| \Sigma \|} \sum_{x \in \Sigma} \text{Re}(\mathcal{R}(x)) \) as a mean estimator for \( \mathcal{F} \) (instead of the median-of-means estimator), then using Eq. (16) and Chebyshev's inequality, we obtain:

\[
\text{Pr}
\left( |\bar{f} - \mathcal{F}| \leq \epsilon \right) \geq 1 - \frac{\text{Var}(\text{Re}(\bar{f}))}{\epsilon^2} \geq 1 - \frac{1}{\| \Sigma \| \epsilon^2}.
\]

(18)

Hence \( \mathcal{F} \) can be estimated using \( \bar{f} \) with error \( \epsilon \) 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 \( \delta^{-1} \).

Note that if we could upper and lower bound the range of \( \text{Re}(\mathcal{R}(x)) \) by some constants, then we could have used a Chernoff-Hoeffding bound for the empirical mean \( \bar{f} \) which gives the aforementioned (exponentially) better dependence of the run-time of the algorithm with \( \delta^{-1} \) (as in Lemma 2.1 where we do use a Chernoff-Hoeffding bound).

**Remark.** Note that the Lemma also applies to estimating \( \langle x | G_1 G_2 \cdots G_L | x' \rangle \) (with \( 1/\text{poly}(n) \) accuracy) as one simply starts the process at \( x_0 = x \) and \( \mathcal{R}(x) \) is only nonzero when one arrives at \( x_L = x' \). Similarly, one can estimate \( \langle \Phi_1 | G_1 G_2 \cdots G_L | \Phi_2 \rangle \) with \( 1/\text{poly}(n) \) accuracy, assuming one can sample from \( \left| \Phi_1(x) \right|^2 \) (or \( \left| \Phi_2(x) \right|^2 \) and compute for a given \( x \) and \( y \), the ratio \( \frac{\Phi_2(y)}{\Phi_2(x)} \). 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 \( P(x) = \left| \Phi(x) \right|^2 \). In many practical settings, \( |\Phi\rangle \) is such that one can define a function \( f : \{0, 1\}^n \rightarrow \mathbb{C} \)
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 $\frac{\Phi(y) - \Phi(x)}{\Phi(x)}$ can be efficiently obtained. Note that under this assumption one can set-up a Monte Carlo 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 $\frac{\Phi(y) - \Phi(x)}{\Phi(x)}$ 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 [7]). The Trotterization gives an error $\epsilon_{\text{trot}}$ (in addition to the sampling error $\epsilon$ 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 $L = \text{poly}(n)O\left(\Upsilon k^{1+1/p}e^{-1/p_{\text{trot}}}\right)$ (for real time) and $L = \text{poly}(n) O\left(\Upsilon k^{1+1/p}e^{-1/p_{\text{trot}}}\right)$ (for imaginary time, provided that $M \geq 4\Upsilon\sum_i |H_i|$, where $M$ is the Trotter variable). $\Upsilon$ 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 $\epsilon$ equals unity minus the probability that at least one of the estimates is beyond $\epsilon$ (which, by the union bound, is at most $K\delta$).

3. Classically processing the signal: the ESPRIT method

We turn to discussing the ESPRIT method [23] which is a method like the matrix pencil method [35, 20, 19] for processing a signal as in Eqs. (1) and (2) consisting of $S$ components. Indeed, suppose a set of values for the signal $g(k)$,

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

(19)

where $|z_j| \leq 1$, for $k \in \{0, 1, \ldots, K\}$, $K$ even. The goal is to determine the $z_j$ and the coefficients $c_j > 0$ \S using $g(k)$ for sufficiently many $k$. In case of the real-time signal $z_j \equiv e^{-iE_j}$, and in case of a purely-decaying imaginary-time signal $z_j \equiv e^{-E_j} \in (e^{-2\pi}, 1)$.

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:

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

(20)

where $\eta(k)$ denotes e.g. the sampling and Trotter noise, and we have $|\eta(k)| \leq \epsilon_{\text{trot}}$ in Theorem 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 $K + 1 \geq 2S$. 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 $K + 1 \geq 2S$, the eigenvalues are indeed resolved both for the decaying and the oscillating

\S 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 [27].
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 [35, 20, 19] 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 $K + 1 \geq 2S$.

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 Eq. (20). 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 $L$:

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

$$= \sum_{j=1}^{S} c_j \begin{pmatrix} 1 & z_j & \cdots & z_j^{K-L} \\ z_j & z_j^2 & \cdots & z_j^{K-L+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_j^L & z_j^{L+1} & \cdots & z_j^K \end{pmatrix}_{(L+1) \times (K-L+1)} + H(\eta), \quad (21)$$

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 $V_L$:

$$H(g) = \sum_{j=1}^{S} c_j \begin{pmatrix} 1 & z_j & \cdots & z_j^{K-L} \\ z_j & z_j^2 & \cdots & z_j^{K-L+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_j^L & z_j^{L+1} & \cdots & z_j^K \end{pmatrix}_{(L+1) \times (K-L+1)} = V_L C V_{K-L}^T, \quad (22)$$

where

$$C \equiv \text{diag}(c_1, c_2, \ldots, c_S), \quad (23)$$
and $V_L$ is
\[
V_L = \begin{pmatrix}
1 & 1 & \ldots & 1 \\
z_1 & z_2 & \ldots & z_S \\
\vdots & \vdots & \ddots & \vdots \\
\bar{z}_1^L & \bar{z}_2^L & \ldots & \bar{z}_S^L
\end{pmatrix}_{(L+1)\times S}. \tag{24}
\]

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
\[
H(y) = \begin{pmatrix}
y(0) & y(1) & \ldots & y(K/2) \\
y(1) & y(2) & \ldots & y(K/2 + 1) \\
\vdots & \vdots & \ddots & \vdots \\
y(K/2) & y(K/2 + 1) & \ldots & y(K)
\end{pmatrix}_{(K/2+1)\times(K/2+1)} \tag{25}
\]
and
\[
H(g) = V_{K/2}C V_{K/2}^T \in \mathbb{C}^{(K/2+1)\times(K/2+1)} \tag{26}
\]

Making contact with error bounds in the previous section, we see that (for $L = K/2$)
\[
\forall k, \eta(k) \leq \epsilon_{\text{tot}} \Rightarrow \|H(\eta)\| = \sigma_{\text{max}}(H(\eta)) \leq \|H(\eta)\|_F \leq K\epsilon_{\text{tot}}. \tag{27}
\]

From the ‘Vandermonde decomposition’ in Eq. (22) 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 is ESPRIT (given in Algorithm 2), which specifically exploits the relation between the Vandermonde decomposition of $H(y)$ and its singular value decomposition.

**Algorithm 2: ESPRIT algorithm.**

**Data:** Time signal $y$, number of decay rates or oscillation frequencies $S$.

**Result:** List $\tilde{z}_1, \ldots, \tilde{z}_S$.

1. $K \leftarrow \text{length}(y)$; /* We will assume $K$ is even for simplicity. */
2. $L \leftarrow K/2$; /* Not the most general choice, however it works well in practice. */
3. $H(y) \leftarrow$ Hankel matrix built from $y$;
4. $\tilde{U}, \tilde{\Sigma}, \tilde{W} \leftarrow \text{SVD}(H(y))$; /* Make sure $\tilde{\Sigma}$ is decreasingly ordered. */
5. $\tilde{U}_S \leftarrow$ First $S$ columns of $\tilde{U}$; /* Remember $\tilde{U}$ is a $(L+1)\times(L+1)$ unitary matrix */
6. $\tilde{U}_0 \leftarrow$ First $L$ rows of $\tilde{U}_S$;
7. $\tilde{U}_1 \leftarrow$ Last $L$ rows of $\tilde{U}_S$;
8. $\tilde{\Psi} \leftarrow \tilde{U}_0^T \tilde{U}_1$; /* Make $S \times S$ signal matrix $\tilde{\Psi}$, + denotes Moore–Penrose inverse. */
9. $\tilde{z}_1, \ldots, \tilde{z}_S \leftarrow$ eigenvalues of signal matrix $\tilde{\Psi}$.

We will see that this algorithm comes with recovery guarantees on the parameters $z_1, \ldots, z_S$, in both the real-time and imaginary-time signal case, provided the noise vector $\eta$ 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]
\[
d((E_i), (\tilde{E}_j)) = \frac{1}{2\pi} \min_{\pi \in \text{Perm}_S} \max_j |\tilde{E}_{\pi(j)} - E_j|, \tag{28}
\]
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 [23], which provides Theorem 3.1 relating \( \|H(\eta)\| \) in Eq. (27) and the energy matching error defined in Eq. (28).

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 Eq. (24), (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

\[
\Delta = \frac{1}{2\pi} \min_{j \neq k} |E_j - E_k|.
\]  

(29)

It has been proved [27] for \( z_j = e^{-iE_j} \) that

\[
\Delta \geq C \Rightarrow \sigma^2_{\min}(V_{K/2}) \geq \frac{C - 1}{C} K,
\]  

(30)

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 \( \Delta \leq 1/S \), hence one should at least take \( K \geq CS \). Based on this bound, Theorem 4 in [23] says:

**Theorem 3.1** ([23]). Let \( (g + \eta)(k) \) be a real-time signal with \( k = 0, \ldots, K \), and with \( g(k) = \sum_{i=1}^{S} c_i z_i^k \), \( c_i > 0 \forall i \), \( c_{\text{min}} = \min_i c_i \) and \( \eta(k) \) a small noise vector. Let \( z_j = e^{-iE_j} \) with \( j = 1, \ldots, S \) and \( E_j \in [0, 2\pi) \) \( \forall j \), and \( K \geq 2C/\Delta \) for some constant \( C > 2 \) with gap \( \Delta \), and \( K + 1 \geq 2S \). If

\[
\|H(\eta)\| \leq c_{\text{min}} K h_1(S, C, K),
\]  

(31)

with

\[
h_1(S, C, K) = \frac{C - 1}{8\sqrt{2SC}} \sqrt{1 - \frac{2CS}{(C - 1)K}},
\]  

(32)

then the ESPRIT algorithm outputs energy estimates \( \{\tilde{E}_j\} \) with distance

\[
d(\{E_i\}, \{\tilde{E}_j\}) \leq \|H(\eta)\| \frac{1}{c_{\text{min}} K^{-1}} h_2(S, C, K),
\]  

(33)

with

\[
h_2(S, C, K) = 40\sqrt{2S} \left( \frac{C}{C - 1} \right)^{3/2} \left( 1 - \frac{2CS}{(C - 1)K} \right)^{-1},
\]  

(34)

By Eq. (27) we have \( \|H(\eta)\| \leq K \epsilon_{\text{tot}} \) and if we choose \( K \sim S \), \( \epsilon_{\text{tot}} \) can be chosen sufficiently small, inversely polynomial with \( S \), such that at least Eq. (31) 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 [30] in which quantum phase estimation 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 Fig. 1.

**Remark.** It is noteworthy that even when the eigenvalues \( E_j \) are not well-separated but occur in ‘clumps’, results exist [23] which bound the performance of ESPRIT.
3. Classical Processing The Signal: The ESPRIT Method

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 [23] 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 Eq. (24).

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—than bounds on this condition number do exist [2]. Based on the work of Gautschi [11], 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 \( \Delta \) 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 (of length \( K \)) with \( g(k) = \sum_{i=1}^{S} c_i z_i^k \), \( c_i > 0, \forall i, c_{\text{min}} = \min c_i \), and \( \eta(k) \) a small noise vector. Let \( z_j = e^{-E_j} \) with \( E_j \in [0, 2\pi) \) and given eigenvalue gap \( \Delta < 1 \) in Eq. (29), and \( \{ E_i \} \) the energy estimates of ESPRIT with \( L = K/2 \). Let \( K + 1 \geq 2S \), \( K \) even and \( K = TS \) for some positive integer \( T \). If we have

\[
\|H(\eta)\| \leq \frac{c_{\text{min}}}{\sqrt{K}} g_1(S, \Delta),
\]  

(35)

with

\[
g_1(S, \Delta) = \frac{1}{32S^2} (e^{-2\pi \Delta})^{3(S-1)},
\]  

(36)

then

\[
d(\{ E_i \}, \{ E_j \}) \leq \|H(\eta)\| c_{\text{min}}^{-1} K \sqrt{K} g_2(S, \Delta),
\]  

(37)

with

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

(38)

Since the dependence on \( S \) is exponential in Eq. (37), one cannot make the distance \( d(\{ E_i \}, \{ E_j \}) \) 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, \( \text{poly}(n) \), effort one can make \( \|H(\eta)\| \) sufficiently small to obey Eq. (35) and then reduce the error on the found eigenvalues to 1/poly(\( n \)). This assumes that the gap between the \( O(1) \) rescaled eigenvalues present in the initial state is at least 1/poly(\( n \)) (and not exponentially small in \( n \)).

Furthermore, given that \( \|H(\eta)\| \) should decrease at least as \( \sim 1/\sqrt{K} \) through Eq. (35) but the upper bound in Eq. (37) 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 \( V_{L=K/2} = V_{S-1} \) is square \( \|
\). 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. 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.

\[\|\text{Hence, strictly speaking Lemma D.11 is not much of a help.}\]
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 [34] 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:

\[ H = -J \left( \sum_i Z_i Z_{i+1} + g \sum_i X_i \right), \tag{39} \]

where \( X, Y, Z \) denote the Pauli matrices, \( J > 0 \) (for a ferromagnetic interaction) and \( g \geq 0 \), 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.\footnote{The Hamiltonian can be transformed to \( \tilde{H} = UHU^\dagger \) by the unitary transformation \( U = \bigotimes_i \exp(\frac{\sqrt{g} Z_i}{2}) \), which alters the direction of the field in the transverse plane while preserving the spectrum.}

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 \(|-\rangle_{q_1} |-\rangle_{q_2} \cdots |-\rangle_{q_p} \prod_{\ell \neq q_1, q_2, \ldots, q_p} |+\rangle_\ell \rangle \), 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 Monte Carlo 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 \( e^{-a_i k/M} H_i \), (where \( a_i \) and \( M \) are set by the Trotterization scheme)\footnote{We note that the numerical results presented in this section are obtained using a first-order Trotter decomposition.}. The local propagation operators acting on a subset of two qubits on the chain are given by:

\[
e^{-\tilde{k}H_i} = \begin{pmatrix}
\frac{\sinh(\tilde{\lambda} \tilde{k})}{\sqrt{1+g^2}} + \cosh(\tilde{\lambda} \tilde{k}) & 0 & \frac{g \sinh(\tilde{\lambda} \tilde{k})}{\sqrt{1+g^2}} & 0 \\
0 & -\frac{\sinh(\tilde{\lambda} \tilde{k})}{\sqrt{1+g^2}} + \cosh(\tilde{\lambda} \tilde{k}) & 0 & \frac{g \sinh(\tilde{\lambda} \tilde{k})}{\sqrt{1+g^2}} \\
\frac{g \sinh(\tilde{\lambda} \tilde{k})}{\sqrt{1+g^2}} & 0 & -\frac{\sinh(\tilde{\lambda} \tilde{k})}{\sqrt{1+g^2}} + \cosh(\tilde{\lambda} \tilde{k}) & 0 \\
0 & \frac{g \sinh(\tilde{\lambda} \tilde{k})}{\sqrt{1+g^2}} & 0 & -\frac{\sinh(\tilde{\lambda} \tilde{k})}{\sqrt{1+g^2}} + \cosh(\tilde{\lambda} \tilde{k})
\end{pmatrix},
\tag{40}
\]

where \( \tilde{\lambda} = \sqrt{J^2 + g^2} \) and \( \tilde{k} = a_i k / M \). This operator is element-wise non-negative and can be efficiently brought to block-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:

\[
|\Phi_{\text{optimal}}\rangle = \frac{1}{\sqrt{2}} \left( \prod_{i=1}^{n} |+\rangle_i + \sum_{q=1}^{\frac{n}{2}} \frac{1}{\sqrt{2}} \prod_{i \neq q} |\gamma\rangle_i |+\rangle_i \right)|\psi_{p=0}\rangle |\psi_{p=1,q}\rangle
\]

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

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 \(\Phi(y)\) for a given \(x, y \in \{0,1\}^n\) and one can efficiently sample from \(|\Phi(x)|^2\): From Eq. (41), 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 \(x\): \(\Phi(x) = 1/2^{(n+1)/2} \left( \frac{1}{n} + \frac{1}{\sqrt{n}} \right) (n - |x|) + \left( \frac{1}{n} - \frac{1}{\sqrt{n}} \right) |x| \), so \(\Phi(x)\) only depends on the Hamming weight \(|x|\) of bit string \(x\), i.e. the quantity \(\Phi(y)/\Phi(x)\) 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 Monte Carlo 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 \(\hat{U}_S\) (in Algorithm 2) by taking the first \(S\) columns of \(\hat{U}\), where \(S\) is now the number of singular values in the SVD of the Hankel matrix \(H(y)\) that exceed \(TF \sigma_{\text{max}}\). \(TF\) denotes what we call a truncation factor, and \(\sigma_{\text{max}}\) 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 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 \(\hat{E}_j\) for the quantum algorithm and for the Monte Carlo 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 \(\hat{E}_j\) are real-valued, we take them to be the real parts of \(i \log(\hat{E}_j^\ast)\).

The code that is used to obtain the numerical results presented in this work can be found at [39].

In Figure 3, the Monte Carlo signals \(\langle \Phi | e^{-ikH} | \Phi \rangle\) 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.
Figure 3: The evolution of the states $|+\rangle^\otimes n$ (in (a), (b) and (c), for which the ground state is the dominant component in the signal) and $|\Phi_{\text{optimal}}\rangle$ (in (d), (e) and (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 Monte Carlo 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 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 Monte Carlo 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 Monte Carlo 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 (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 Monte Carlo 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. 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 Monte Carlo 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

* For $K = 2; k = 0, 1, 2$. For $K = 4; k = 0, 1, 2, 3, 4$. Etc.
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 Monte Carlo algorithm for $|\Sigma| = 4200$ and $M = 100$ for several values of $K$. Plots (c),(d) and plots (e),(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 $\text{TF} = 0.02$ 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.

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 Monte Carlo 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 Monte Carlo algorithm. Clearly, the relative errors of the excited-state eigenvalue estimates for the quantum algorithm are smaller than those for the Monte Carlo algorithm.

We have also implemented the matrix pencil method in [19, 16] 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 a stoquastic – i.e. a sign-problem-free – local Hamiltonian $H$ 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 evolution signal is obtained through running quantum circuits, while the imaginary-time signal is obtained through a Monte Carlo scheme (developed in this work) that is implemented in a computationally tractable manner classically.

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 worse in extracting (multiple) eigenvalues from a MC algorithm signal than from a quantum algorithm signal in the presence of noise. Even though our results show that for stoquastic Hamiltonians, for an input state supported on $S = O(1)$ eigenvalues (separated by an at least $1/poly(n)$ gap), these eigenvalues can be estimated with poly$(n)$ classical effort, it remains to be better understood how practical this method is. The upper bounds for the errors of the eigenvalue estimates in Theorem 3.2 grow rather fast with $S$, and it is not clear how much one can improve these bounds.

We find numerically 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 and 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 Monte Carlo + 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 directions of further research one can ask whether a hybrid approach in which imaginary-time data from a noiseless Monte Carlo 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 $z_j = e^{-iE_j}$ on the unit circle each have a partner pole $z'_j = e^{-E_j}$ on the real axis. If the effect of noise can be modeled $z_j = e^{-iE_j} \rightarrow e^{iE_j-\gamma}$ [30], 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 Monte Carlo method at random (instead of picking $k = 0, 1, \ldots, K$).

It would also be interesting to show whether 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 [32], 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 [38] but not for imaginary-time signals), which is a question we leave for further research.
Acknowledgements

This work is supported by QuTech NWO funding 2020-2024 – Part I “Fundamental Research”, project number 601.QT.001-1, financed by the Dutch Research Council (NWO). JH is supported by the Quantum Software Consortium (NWO Gravitation Grant, project number 024.003.037). MS and BMT developed the MC method based on unpublished results of Sergey Bravyi, MS implemented the numerics on the transverse field Ising model, JH performed the analysis of the ESPRIT algorithm for decaying signals, BMT supervised the whole project and all authors contributed to the writing. We thank Ingo Roth for pointing out the use of median-of-means estimators for observables whose higher-order moments cannot be upper bounded by a constant.

Appendix A. Trotterization

Suppose $H = \sum_{i=1}^{N} H_i$ (where $N = \mathcal{O}(\text{poly}(n))$) represents a $k$-local Hamiltonian of a quantum system. Each individual subset all

\[ ||H_i|| = 1 \]

Suppose $H = \sum_{i=1}^{N} H_i$, 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 [7]) 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 $H = \sum_{i=1}^{N} H_i$. Furthermore, suppose the set $\{H_i\}_{i=1}^{N}$ can be divided into a minimum of $\Gamma$ subsets $\{H_\gamma\}_{\gamma=1}^{\Gamma}$, such that within each individual subset all $H_i$’s commute. Then the quantities $|\langle \Phi | e^{-itH} | \Phi \rangle - \langle \Phi | \prod_{\gamma} e^{-itH_\gamma/M} | \Phi \rangle |$ and $|\langle \Phi | e^{-\tau H} | \Phi \rangle - \langle \Phi | \prod_{\gamma} e^{-\tau H_\gamma/M} | \Phi \rangle |$ (where $|\Phi\rangle$ is a normalized state and $t, \tau \in \mathbb{R}_+$) are bounded as follows:

\[ |\langle \Phi | e^{-itH} | \Phi \rangle - \langle \Phi | \prod_{\gamma} e^{-itH_\gamma/M} | \Phi \rangle | \leq \sum_{\gamma=1}^{\Gamma-1} \| [H_\gamma, H_{\gamma'}] \| \frac{t^2}{2M}, \quad \text{(A.1a)} \]

\[ |\langle \Phi | e^{-\tau H} | \Phi \rangle - \langle \Phi | \prod_{\gamma} e^{-\tau H_\gamma/M} | \Phi \rangle | \leq \sum_{\gamma=1}^{\Gamma-1} \sum_{\gamma > \gamma'} \| [H_\gamma, H_{\gamma'}] \| \frac{\tau^2}{2M}, \quad \text{(A.1b)} \]

where the second inequality holds provided that $|e^{-\tau H/M}| \leq 1$, $|e^{-\tau H_\gamma/M}| \leq 1$ (for $\forall \gamma$) and $\frac{\tau (\sum_{M} \| H_{\gamma} \|)}{M} \leq 1$, 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 $p$th-order approximants of $e^{-itH/M}$ and $e^{-\tau H/M}$ by $T_M(p, t)$ and $T_M(p, \tau)$, respectively. We denote $|\langle \Phi | e^{-itH} | \Phi \rangle - \langle \Phi | T_M(p, t) | \Phi \rangle |$ and $|\langle \Phi | e^{-\tau H} | \Phi \rangle - \langle \Phi | T_M(p, \tau) | \Phi \rangle |$ by $\epsilon_{\text{trot}}$. In [7], it was shown that, for general $p$, $\epsilon_{\text{trot}}$ is upper bounded as follows:

\[ \epsilon_{\text{trot}} \leq \mathcal{O} \left( \alpha \frac{t^{p+1}/M^p}{p^p} \right), \quad \text{for real-time evolution,} \quad \text{(A.2a)} \]

\[ \epsilon_{\text{trot}} \leq \mathcal{O} \left( \alpha \frac{\tau^{p+1}/M^p}{p^p} \right), \quad \text{for imaginary-time evolution,} \quad \text{(A.2b)} \]

where $\alpha = \sum_{p_1, p_2, \ldots, p_{\Gamma-1}} \| [H_{\gamma_{p+1}}, \ldots, H_{\gamma_{p+1}}] \| (\alpha^{1/p} \text{ is typically poly}(n))$ and Eq. (A.2b) holds provided that $4\mathcal{Y} (\sum_{\gamma} \| H_{\gamma} \|) / M \leq 1$ (where $\mathcal{Y}$ corresponds to the number of stages of the Trotter decomposition and typically scales exponentially in $p$). In [40], a widely used scheme for constructing $p$th-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 [40], 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
If one wishes to obtain a given $\epsilon_{\text{trot}}$, the number of $k$-local propagation operators into which the evolutions are decomposed scales as $L = \text{poly}(n) O(T t^{1+1/p \epsilon_{\text{trot}}^{-1/p}})$ (for real time) and $L = \text{poly}(n) O(T \tau^{1+1/p \epsilon_{\text{trot}}^{-1/p}})$ (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.

**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.
Appendix B. Extension to non-Hermitian propagation operators

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

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

(i) $|\Phi\rangle = \sum_{x=1}^{2^n} \Phi(x) |x\rangle$ 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) $\Phi(y)/\Phi(x)$ can be efficiently (poly($n$)) calculated for a given $x$ and $y$ and (2) we can efficiently draw samples from the probability distribution $P(x) = |\Phi(x)|^2$.

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

$\mathcal{F}$ can be estimated within error $\epsilon$ 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:

$$F_l \equiv \begin{cases} G_l \otimes |0\rangle\langle 1| + G_l^\dagger \otimes |1\rangle\langle 0|, & \text{if } l \text{ is odd} \\ G_l \otimes |1\rangle\langle 0| + G_l^\dagger \otimes |0\rangle\langle 1|, & \text{if } l \text{ is even}. \end{cases} \quad (B.1)$$

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 $\langle \Phi | F_1 F_2 \ldots F_L | \Phi \rangle$, provided that its eigenvalues lie in $(0,1]$. The eigenvalues $\lambda$ of $F_l$ (for $l$ odd) can be found by solving:

$$\det \begin{pmatrix} -\lambda & G_l \\ G_l^\dagger & -\lambda^* \end{pmatrix} = \det \left( \lambda^2 I - G_l G_l^\dagger \right) = \det \left( G_l G_l^\dagger - \lambda^2 I \right) = 0, \quad (B.2)$$

where we have used that $G_l^{(t)}$ commutes with $1$ and that $G_l^{(t)}$ is of even dimensionality. The eigenvalues of the Hermitian and positive semi-definite matrix $G_l G_l^\dagger$ are thus equal to $\lambda^2$. 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 $l$.

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: $G_1 G_2 \ldots G_L = (0 | F_1 F_2 \ldots F_L | L \mod 2)$, for $L \in \mathbb{Z}_+$. This is done below by means of induction.

- For $L = 1$:

$$\langle 0 | F_1 | 1 \rangle = \langle 0 | \left( G_1 \otimes |0\rangle\langle 1| + G_1^\dagger \otimes |1\rangle\langle 0| \right) | 1 \rangle = G_1 \langle 0 | 0 \rangle \langle 1 | 1 \rangle + G_1^\dagger \langle 0 | 1 \rangle \langle 0 | 1 \rangle = G_1, \quad (B.3)$$

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

$$F_{L+1} = G_{L+1} \otimes |L \mod 2 \rangle \langle L + 1 \mod 2 | + G_{L+1}^\dagger \otimes |L + 1 \mod 2 \rangle \langle L \mod 2 |. \quad (B.4)$$

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

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

which finishes the proof.
Appendix C. Median-of-means estimator

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

\[
\text{Pr}\left( |\text{Re}(\hat{\mathcal{F}}) - \mathcal{F}| \leq \epsilon \right) \geq 1 - \delta,
\]

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

Taking the median-of-means estimator [26] as estimator (instead of the empirical mean), one can obtain a more convenient scaling of \( |\Sigma| \) w.r.t. \( \delta \) (despite the fact that only the first two moments of \( \text{Re}(\mathcal{R}(x)) \) can be upper bounded). The median-of-means estimator can be constructed as follows: Partition the set of MC samples \( \Sigma \) into \( q \) groups \( s_1, \ldots, s_q \) of size approximately \( |\Sigma|/q \). One then computes the empirical mean of \( \text{Re}(\mathcal{R}(x)) \) 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 \( f_j = \frac{1}{|s_j|} \sum_{x \in s_j} \text{Re}(\mathcal{R}(x)) \) (for \( j \in \{1, \ldots, q\} \)) and denote the median of these empirical means by \( \hat{\mathcal{F}} = M(f_1, \ldots, f_q) \). 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 \( \text{M}(a_1, \ldots, a_q) = a_i \) with \( a_i \) such that

\[
|\{j: a_j \leq a_i\}| \geq q/2 \quad \land \quad |\{j: a_j \geq a_i\}| \geq q/2,
\]

where we take the smallest \( i \) if multiple \( i \) obey this condition.

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

\[
\text{Pr}\left( |\hat{\mathcal{F}} - \mathcal{F}| \leq \sqrt{4q/|\Sigma|} \right) \geq 1 - e^{-q/8}.
\]

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

\[
\text{Pr}\left( |\hat{\mathcal{F}} - \mathcal{F}| \leq \epsilon \right) \geq 1 - \delta.
\]

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

To see why Eq. (C.3) is true, see [26], note that one can apply Chebyshev’s inequality to each of the empirical means \( f_j \): with probability at least \( 3/4 \), we have \( |f_j - \mathcal{F}| \leq \sqrt{4q/|\Sigma|} \). If \( |\hat{\mathcal{F}} - \mathcal{F}| \geq \sqrt{4q/|\Sigma|} \), then, by definition of \( \hat{\mathcal{F}} \), at least \( q/2 \) of the empirical means \( f_j \) satisfy \( |f_j - \mathcal{F}| \geq \sqrt{4q/|\Sigma|} \). Hence the probability that \( |\hat{\mathcal{F}} - \mathcal{F}| \geq \sqrt{4q/|\Sigma|} \) is upper bounded by the probability that a binomially distributed random variable with \( q \) draws and success probability \( 1/4 \) exceeds \( q/2 \):

\[
\text{Pr}\left( |\hat{\mathcal{F}} - \mathcal{F}| \geq \sqrt{4q/|\Sigma|} \right) \leq \text{Pr}\left( \text{Bin}(q, 1/4) \geq q/2 \right) = \text{Pr}\left( \text{Bin}(q, 1/4) - \mathbb{E}(\text{Bin}(q, 1/4)) \geq q/4 \right) \leq e^{-q/8},
\]

where we have used \( \mathbb{E}(\text{Bin}(q, 1/4)) = q/4 \) 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 [23], 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
APPENDIX D PERFORMANCE OF ESPRIT ON THE IMAGINARY-TIME (DECAYING) SIGNAL

defined in the ESPRIT Algorithm 2 and hold more generally). To state this Lemma we need to consider a freedom of choice in $U$. It is a compressed version of Lemmas 4 and 5 in [23] (which are formulated for real-time signals only, but hold more generally). We will make use of the following sufficient condition:

**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 $\sigma_{\min}(X) = \|X\|$, 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

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

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

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

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 ([13]).** 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 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 [23] (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 $\bar{U}_S$ with $\bar{U}_S$ as defined in the ESPRIT Algorithm 2 and $U_S$ its noise-free version. It is possible that $U_S$ and $\bar{U}_S$ 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 $\bar{\Psi}$. The rotated version of $U_S$ is given through the $S \times S$ unitary operator $(O_2O_1)^\dagger$, which is defined via the singular value decomposition of $U_S^\dagger\bar{U}_S$, i.e.

$$O_1U_S^\dagger\bar{U}_SO_2 = D, \tag{D.3}$$

with $I_S \geq D \geq 0$ 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 [23], whereas we have chosen to make it explicit at all times.

**Lemma D.3.** Let $(g + \eta)(k)$ be an imaginary-time signal with $g(k) = \sum_{k=1}^S c_i z_i^k$ and $\eta(k)$ a small noise vector. Consider the associated Hankel matrices $H(g)$ and $H(g + \eta)$, with singular value decompositions $H(g) = USW$ and $H(g + \eta) = \bar{U}\bar{S}\bar{W}$, and label the matrix of the first $S$ columns of $U$ (resp. $\bar{U}$) as $U_S$ (resp. $\bar{U}_S$). Finally, let $O_1U_S^\dagger\bar{U}_SO_2 = D$ with $I_S \geq D \geq 0$ be the singular value decomposition of $U_S^\dagger\bar{U}_S$. If

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

then

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

**Proof.** First, we can observe that indeed $I_S \geq D$ as $\|O_1U_S^\dagger\bar{U}_SO_2\| \leq \|O_1\| \|U_S^\dagger\| \|\bar{U}_S\| \|O_2\| \leq 1$.

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, 37]. From this latter theorem we know that $|\sigma_i(H(g + \eta)) - \sigma_i(H(g))| \leq \|H(\eta)\| \leq \sigma_{\min}(H(g))/2$ where $\sigma_i$ is the $i$th singular value (in order and some singular values can be zero). Let $\sigma_{\min}(H(g + \eta)) > 0$ be the $k$th singular value, and thus

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

24
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 [22], setting $\delta = \alpha = \sigma_{\min}(H(g))/2$) to conclude that
\[
\left\| (U^\perp_S)^\dagger \hat{U}_S \right\| \leq \frac{2 \| H(\eta) \|}{\sigma_{\min}(H(g))}, \tag{D.7}
\]
where $U^\perp_S$ is the matrix formed from the $L + 1 - S$ other (besides $U_S$) columns of the noiseless $U$. To connect this to $U_S(O_2O_1)^\dagger - \hat{U}_S$ we can make the following long calculation:
\[
\| U_S(O_2O_1)^\dagger - \hat{U}_S \| \leq \left\| U_S(O_2O_1)^\dagger - U_S \right\|_F
\]
\[
= \left[ \text{tr} \left( U_S(O_2O_1)^\dagger (O_2O_1) U^\perp_S \right) + \text{tr} \left( \hat{U}_S \hat{U}^\perp_S \right) - \text{tr} \left( (U_S(O_2O_1)^\dagger \hat{U}^\perp_S + U_S(O_2O_1) U^\perp_S) \right) \right]^{1/2}
\]
\[
= \left[ 2 \text{tr} \left( \hat{U}_S \hat{U}^\perp_S \right) - 2 \text{tr} (D) \right]^{1/2}
\]
\[
= \sqrt{2} \left[ \text{tr} \left( \hat{U}_S \hat{U}^\perp_S \right) - \text{tr} \left( (O_1 U^\perp_S O_2)(O_1 U^\perp_S O_2)^\dagger \right) \right]^{1/2}
\]
\[
= \sqrt{2} \left[ \text{tr} \left( \hat{U}_S \hat{U}^\perp_S \right) - \text{tr} (U^\perp_S \hat{U}^\perp_S U_S) \right]^{1/2}
\]
\[
= \sqrt{2} \left[ \text{tr} \left( \hat{U}_S \hat{U}^\perp_S \right) - \text{tr} (U^\perp_S \hat{U}^\perp_S \hat{U}_S) \right]^{1/2}
\]
\[
= \sqrt{2} \left[ \text{tr} \left( \hat{U}_S \hat{U}^\perp_S \right) + \text{tr} (U^\perp_S (U^\perp_S)^\dagger \hat{U}^\perp_S) \right]^{1/2}
\]
\[
= \sqrt{2} \left\| (U^\perp_S)^\dagger \hat{U}_S \right\|_F \leq 2\sqrt{5} \left\| (U^\perp_S)^\dagger \hat{U}_S \right\| \tag{D.8}
\]
In the second inequality we used that $\text{tr} \left( U^\perp_S \hat{U}^\perp_S \right) = \text{tr} \left( \hat{U}_S \hat{U}^\perp_S \right) = S$ since $U_S$ as well as $\hat{U}_S$ consist of $S$ orthonormal columns, and $D \geq D^2 = DD^\dagger$, since $I \geq D \geq 0$. In addition, at the end we use that $UU^\dagger = U_S U^\perp_S + U^\perp_S (U^\perp_S)^\dagger = I$ as $U$ is unitary.

The next step is to bound the deviation of the ESPRIT signal matrix $\hat{\Psi} = \hat{U}_S \hat{U}_I$ from the rotated version of its noiseless variant $O_2O_1^\dagger \Psi (O_2O_1)^\dagger = (O_2O_1^\dagger) U^\perp_S U_1 (O_2O_1)^\dagger$ in terms of $\left\| U_S(O_2O_1)^\dagger - \hat{U}_S \right\|$. 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 $\Psi$ 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 rank($A$) = rank($B$). If $\| A - B \| \leq \sigma_{\min}(A)/2$ then
\[
\| A^\perp - B^\perp \| \leq \frac{1 + \sqrt{5}}{2} \| A - B \| \| A^\perp \|^2 = \frac{1 + \sqrt{5}}{2} \| A - B \| \tag{D.9}
\]

**Proof.** From Theorem 4.1 in [43] we get that
\[
\| A^\perp - B^\perp \| \leq \frac{1 + \sqrt{5}}{2} \| A - B \| \| A^\perp \| \| B^\perp \|, \tag{D.10}
\]
Furthermore, since $\| A - B \| \leq \frac{\sigma_{\min}(A)}{2} < \frac{1}{\| A^\perp \|}$, we have by Lemma 3.1 in [43] that
\[
\| B^\perp \| \leq \frac{\| A^\perp \|}{\| A^\perp \| - \| A - B \|} \leq \| A^\perp \|, \tag{D.11}
\]
leading to the first inequality in Eq. (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.

APPENDIX D PERFORMANCE OF ESPRIT ON THE IMAGINARY-TIME (DECAYING) SIGNAL
Lemma D.5. Let \( A \) be an \( m \times n \) \((m \leq n)\) matrix of rank \( m \) and let \( B \) an \( m \times n \) matrix s.t. \( \| A - B \| \leq \sigma_{\min}(A)/2 \). Then \( \text{rank}(A) = \text{rank}(B) \).

Proof. By construction, we have \( \text{rank}(A) \geq \text{rank}(B) \). 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 [37]. 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 \( \text{rank}(A) = \text{rank}(B) \). \qed

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 [23], where the corresponding result, Lemma 3 in [23], makes explicit use of the fact that in their scenario all poles \( z_i \) 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 Eq. (24) (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

\[
H(g) = U \Sigma W = V_L CV_K^T,
\]

where \( V_L \) is the \((L + 1) \times S\) Vandermonde matrix defined in Eq. (24) and the coefficient matrix \( C \) is given in Eq. (23). When \( S \leq L \leq K + 1 - S \) (requiring \( K + 1 \geq 2S \)), \( V_L \) 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

\[
U_S = V_L A,
\]

and thus

\[
U_0 = V_{L-1} A, U_1 = V_{L-1} ZA.
\]

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

\[
\Psi = U_0^+ U_1 = A^{-1} V_{L-1}^+ V_{L-1} ZA = A^{-1} ZA.
\]

and hence the eigenvalues of \( \Psi \) 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

\[
\sigma_{\min}(U_0) \geq \sigma_{\min}(V_{L-1}) \| V_L \|^{-1}.
\]

Proof. We have

\[
I_S = U_0^+ U_S = (V_L A)^+ V_L A = A^+ V_L^+ V_L A,
\]

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

\[
\sigma_{\min}^{-1}(U_0) = \| U_0^+ \| = \| (V_{L-1} A)^+ \| = \| A^{-1} V_{L-1}^+ \| \leq \| A^{-1} \| \| V_{L-1}^+ \| = \frac{\| V_L \|}{\sigma_{\min}(V_{L-1})},
\]

which is the inverse of the Lemma statement. \qed

With Lemmas D.4, D.5 and 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 \( U_0 (O_2 O_1)^\dagger \), 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

\[
\| U_0 (O_2 O_1)^\dagger - \tilde{U}_0 \| \leq \sigma_{\min}(U_0)/2,
\]

where \( O_2 O_1 \) is defined through the singular value decomposition of \( U_0^+ U_S \) (i.e. \( O_1 U_0^+ U_S O_2 = D \)). With this assumption we have

\[
\| (O_2 O_1) \Psi (O_2 O_1)^\dagger - \tilde{\Psi} \| \leq 5 \| U_S (O_2 O_1)^\dagger - \tilde{U}_S \| \frac{\| V_L \|}{\sigma_{\min}(V_{L-1})},
\]

which is a simple bound that is sufficient for our purpose.
Proof. Following [23] we have
\[
\left\| (O_2O_1) \Psi (O_2O_1)^\dagger - \tilde{\Psi} \right\| \leq \left\| (O_2O_1) U_0^{\dagger} \right\| \left\| U_1 (O_2O_1)^\dagger - \tilde{U}_1 \right\| + \left\| (O_2O_1) U_0^{\dagger} - \tilde{U}_0^{\dagger} \right\| \left\| \tilde{U}_1 \right\|. \tag{D.20}
\]
We have \( \left\| \tilde{U}_1 \right\| \leq \left\| \tilde{U}_S \right\| = 1 \), since \( \tilde{U}_S^{\dagger} \tilde{U}_S = I_S \) and removing a row vector decreases the operator norm. Similarly we have \( \left\| U_1 (O_2O_1)^\dagger - \tilde{U}_1 \right\| \leq \left\| U_S (O_2O_1)^\dagger - \tilde{U}_S \right\| \). 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 \( \text{Rank}(U_0 (O_2O_1)^\dagger) = \text{Rank}(\tilde{U}_0) \). This means that we can use Lemma D.4 to conclude that
\[
\left\| (O_2O_1) U_0^{\dagger} - \tilde{U}_0^{\dagger} \right\| \leq \frac{2}{1 + \sqrt{\frac{5}{\sigma_{\text{min}}^2(U_0)}}} \left\| U_0 (O_2O_1)^\dagger - \tilde{U}_0 \right\| \leq \frac{2}{\sigma_{\text{min}}(U_0)} \left\| U_0 (O_2O_1)^\dagger - \tilde{U}_0 \right\|. \tag{D.21}
\]
Hence we get
\[
\left\| (O_2O_1) \Psi (O_2O_1)^\dagger - \tilde{\Psi} \right\| \leq \frac{2}{\sigma_{\text{min}}(U_0)} \left( \sigma_{\text{min}}(U_0) + (1 + \sqrt{\frac{5}{\sigma_{\text{min}}^2(U_0)}}) \right) \left\| U_0 (O_2O_1)^\dagger - \tilde{U}_0 \right\| \leq \frac{3}{2} \left\| U_0 (O_2O_1)^\dagger - \tilde{U}_0 \right\|. \tag{D.22}
\]
where we used \( \sigma_{\text{min}}(U_0) \leq \left\| U_0 \right\| \leq 1 \). Plugging in the lower bound on \( \sigma_{\text{min}}(U_0) \) (Lemma D.6) and noting that \( \frac{3}{2} + \sqrt{5} \leq 5 \) we obtain the Lemma statement. \( \blacksquare \)

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 \( g(k) = \sum_{i=1}^{S} c_i x_i^k \) 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 \( H(g) = U S \Sigma W \) and \( H(g + \eta) = \tilde{U} \tilde{S} \tilde{W} \), 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. \( \tilde{U}_0 \)) the submatrix of \( U_S \) with the last row (resp. first row) removed and define the signal matrix \( \Psi = U_0^{\dagger} U_1 \) (similarly for \( \tilde{\Psi} \)). Let \( L = K/2 \) and \( K + 1 \geq 2S \). Now, if
\[
4 \left\| H(\eta) \right\| \leq c_{\text{min}}^2 \sigma_{\text{min}}^2(V_{K/2}) \sigma_{\text{min}}(V_{K/2-1}) \left\| V_{K/2} \right\|^{-1}, \tag{D.23}
\]
then
\[
\left\| (O_2O_1) \Psi (O_2O_1)^\dagger - \tilde{\Psi} \right\| \leq \frac{10 \sqrt{25} \left\| H(\eta) \right\| \left\| V_{K/2} \right\|^2}{c_{\text{min}}^4 \sigma_{\text{min}}^4(V_{K/2-1})} c_{\text{min}}^{-1}. \tag{D.24}
\]
where \( O_2O_1 \) is defined through the singular value decomposition of \( U_0^{\dagger} \tilde{U}_S \), Eq. (D.3).

**Proof.** We start from the requirement in Lemma D.7 that \( \left\| U_0 (O_2O_1)^\dagger - \tilde{U}_0 \right\| \leq \left\| U_S (O_2O_1)^\dagger - \tilde{U}_S \right\| \leq \sigma_{\text{min}}(U_0)/2 \). By Lemma D.6, this is certainly satisfied if \( \left\| U_S (O_2O_1)^\dagger - \tilde{U}_S \right\| \leq \sigma_{\text{min}}(V_{K/2-1}) \left\| V_{K/2} \right\|^{-1}/2 \). Moreover, from Lemma D.3 we know that \( \left\| U_S (O_2O_1)^\dagger - \tilde{U}_S \right\| \leq 2 \sqrt{25} \left\| H(\eta) \right\| \sigma_{\text{min}}^{-1}(H(g)) \) so now let’s upperbound \( \sigma_{\text{min}}^{-1}(H(g)) \) in terms of the smallest singular value of \( V_{K/2} \). We have
\[
\sigma_{\text{min}}^{-1}(H(g)) = \left\| H(\eta) \right\| ^{+} = \left\| (V_{K/2}^T)^{-1} C^{-1} V_{K/2}^+ \right\| \leq \sigma_{\text{min}}^{-2}(V_{K/2}) c_{\text{min}}^{-1}, \tag{D.25}
\]
where we used \( c_{\text{min}} = \min_i c_i \). This means that the condition
\[
4 \left\| H(\eta) \right\| \leq c_{\text{min}}^2 \sigma_{\text{min}}^2(V_{K/2}) \sigma_{\text{min}}(V_{K/2-1}) \left\| V_{K/2} \right\|^{-1}, \tag{D.26}
\]

allows us to use Lemma D.7

\[ \left\| (O_2 O_1) \Psi (O_2 O_1)^T - \hat{\Psi} \right\| \leq 10 \sqrt{2S} \frac{\| H(\eta) \|}{\sigma_{\min}(V_{K/2})} \frac{\| V_{K/2} \|}{\sigma_{\min}(V_{K/2-1})} \frac{c_{\min}^2}{c_{\min}^2} \leq \frac{10 \sqrt{2S} \| H(\eta) \|}{\sigma_{\min}(V_{K/2-1})} \frac{\| V_{K/2} \|}{\sigma_{\min}(V_{K/2})} c_{\min}^{-1}, \]  

(D.27)

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

We wish to translate the bound in Theorem D.8 to a theorem on the distance between the inferred eigenvalues \( z_i \) and \( \hat{z}_i \). The argument is as follows. We know from the Bauer-Fike theorem and the fact that \( (O_2 O_1) \Psi (O_2 O_1)^T \) is diagonalizable (see [3, Exercise VIII.3.2]) that the operator norm bound on \( (O_2 O_1) \Psi (O_2 O_1)^T \) implies a matching bound distance on its eigenvalues \( z_i \), using Eq. (D.15). That is, we have

\[ d(\{z_i\}, \{\hat{z}_j\}) := \max_{\pi \in \text{Perms}} \min_i |z_{\pi(i)} - \hat{z}_j| \leq (2S - 1) \kappa(A(O_2 O_1)^T) \left\| (O_2 O_1) \Psi (O_2 O_1)^T - \hat{\Psi} \right\|, \]  

(D.28)

where \( \kappa(A(O_2 O_1)^T) = \kappa(A) := \|A\| \|A^{-1}\| \) is the condition number of the invertible matrix \( A \) in Eq. (D.13). We have \( A = V_{K/2}^+ U_S \) and since \( U_S \) is an isometry we know that

\[ \kappa(A) = \|A\| \|A^+\| = \|V_{K/2}\| \left\| V_{K/2}^+ \right\| = \frac{\|V_{K/2}\|}{\sigma_{\min}(V_{K/2})}, \]  

(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 \( g(k) = \sum_{i=1}^{S} c_i z_i \), let \( \eta(k) \) a small noise vector, and \( K + 1 \geq 2S \) \((L = K/2)\). Let \( \hat{z}_i \) be the output of the ESPRIT algorithm. Then under the noise condition:

\[ 4 \| H(\eta) \| \leq c_{\min} \sigma_{\min}(V_{K/2}) \sigma_{\min}(V_{K/2-1}) \| V_{K/2} \|^{-1}, \]  

(D.30)

we have

\[ d(\{z_i\}, \{\hat{z}_j\}) \leq (2S - 1) \frac{10 \sqrt{2S} \| H(\eta) \|}{\sigma_{\min}(V_{K/2-1})} \frac{\| V_{K/2} \|}{\sigma_{\min}(V_{K/2})} c_{\min}^{-1}. \]  

(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 Eq. (30).

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 [11]).** Let \( V_{S-1} \) be a square \( S \times S \) Vandermonde matrix with \( S \) (unequal) real positive poles \( z_1, \ldots, z_S \). Then \( \infty \) norm of \( V_{S-1} \) is

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

(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 \( z_1, \ldots, z_S \leq 1 \). Then we have

\[ \| V_{ST-1}^+ \|_{\infty} \leq 2 \| V_{S-1}^{-1} \|_{\infty}, \]  

(D.33)
Proof. Note that
\[ V_{ST-1} = (V^T_{S-1} \quad Z^S V^T_{S-1} \quad Z^{2S} V^T_{S-1} \quad \cdots \quad Z^{(T-1)S} V^T_{S-1})^T, \]
with \( Z = \text{diag}(z_1, \ldots, z_S) \), using \((V_{S-1})_{ij} = z^{i-1}_j\) and \((Z^S V^T_{S-1})_{ij} = z^{i+j-1}_i\).

The pseudo-inverse \(V^+_{ST-1}\) of size \( S \times ST \) can be directly calculated as
\[ V^+_{ST-1} = (I - Z^{2S})(I - Z^{2ST})^{-1} (V^T_{S-1} \quad Z^S V^T_{S-1} \quad Z^{2S} V^T_{S-1} \quad \cdots \quad Z^{(T-1)S} V^T_{S-1}), \]
using a geometric series. Hence \( \|V^+_{ST-1}\|_\infty \) can be calculated to be
\[
\max_{i \in \{1, \ldots, S\}} \frac{1 - z^{2S}_i}{1 - z^{2ST}_i} \sum_{p=0}^{T-1} \sum_{j=1, j \neq i}^S \frac{1 + z_i}{|z_j - z_i|} \leq \max_i \left[ \frac{1 - z^{2S}_i}{1 - z^{2ST}_i} \right] \|V^{-1}_{S-1}\|_\infty = \max_i \left[ \frac{1 + z^S_i}{1 + z^{2ST}_i} \right] \|V^{-1}_{S-1}\|_\infty \leq 2\|V^{-1}_{S-1}\|_\infty, \]
which gives the Lemma statement. \( \square \)

To apply this Lemma, we use that \( \sigma^{-1}_{\min}(V_{ST-1}) = \|V^+_{ST-1}\| \leq \sqrt{S} \|V^+_{ST-1}\|_\infty \) so that
\[
\sigma_{\min}(V_{ST-1}) \geq (2\sqrt{S})^{-1} \|V^{-1}_{S-1}\|_\infty^{-1}, \]
for any \( T \). Unlike the lower bound for the real-time signal which explicitly uses the gap \( \Delta \) in Eq. (30), this bound does not improve with \( T \).

We will now use the gap to upper bound \( \|V^{-1}_{S-1}\|_\infty \) given in Eq. (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 restated here:

**Theorem.** Let \((g+\eta)(k)\) be an imaginary-time decaying signal (of length \( K \)) with \( g(k) = \sum_{i=1}^S c_i z^k_i, c_i > 0 \forall i, c_{\min} = \min_i c_i \) and \( \eta(k) \) a small noise vector. Let \( z_i = e^{-\tilde{E}_i} \) with \( E_i \in [0, 2\pi) \) and given eigenvalue gap \( \Delta < 1 \) in Eq. (29), and \( \{\tilde{E}_i\} \) the energy estimates of ESPRIT with \( L = K/2 \). Let \( K + 1 \geq 2S, K \) even and \( K = TS \) for some positive integer \( T \). If we have
\[
\|H(\eta)\| \leq \frac{c_{\min}}{\sqrt{K}} g_1(S, \Delta), \]
with
\[
g_1(S, \Delta) = \frac{1}{32S^2} (e^{-2\pi \Delta})^{3(S-1)}, \]
then
\[
d(\{\tilde{E}_i\}, \{E_j\}) \leq \|H(\eta)\| c^{-1}_{\min} K \sqrt{K} g_2(S, \Delta), \]
with
\[
g_2(S, \Delta) = e^{2\pi \sqrt{2640.855} (e^{-2\pi \Delta})^{-5(S-1)}}. \]

**Proof.** First, we can lift the bound on the eigenvalue distance \( d(\{z_i\}, \{\tilde{z}_i\}) \) to one on energies \( d(\{E_i\}, \{\tilde{E}_j\}) \) defined in Eq. (28), with \( \tilde{E}_i := -\log(\tilde{z}_i) \) and \( E_i \in [0, 2\pi) \) by noting that
\[
\frac{1}{2\pi} |E_{\pi(i)} - \tilde{E}_i| = \frac{1}{2\pi} |\log(z_{\pi(i)}) - \log(\tilde{z}_i)| = \frac{1}{2\pi} |\log (1 - (\tilde{z}_i - z_{\pi(i)})/\tilde{z}_i)| \leq |(\tilde{z}_i - z_{\pi(i)})/\tilde{z}_i| \leq e^{2\pi |(\tilde{z}_i - z_{\pi(i)})|}, \]
using that \( z_i \in (e^{-2\pi}, 1) \) and \( \tilde{z}_i \in (e^{-2\pi}, 1) \). In particular, for the first inequality, let \( x = (\tilde{z}_i - z_{\pi(i)})/\tilde{z}_i \). If \( x < 0, |\log(1 - x)| \leq |x| \). If \( x > 0, \) since \( z_i \in (e^{-2\pi}, 1) \), we have \( x \leq 1 - e^{-2\pi} \), so that \( |\log(1 - x)| \leq 2\pi |x| \).

Second, let us now use the gap condition \( |E_i - E_j| \geq 2\pi \Delta \) in Eq. (29). This leads to a gap condition on the \( z_i \) themselves through (assuming w.l.o.g. that \( z_i \geq z_j \)):
\[
2\pi \Delta \leq |E_i - E_j| = |\log(z_i/z_j)|, \]

29
and thus
\[ e^{2\pi \Delta} \leq \frac{z_i}{z_j} = \frac{z_i - z_j}{z_j} + 1, \]
which gives
\[ |z_i - z_j| = z_i - z_j \geq z_j (e^{2\pi \Delta} - 1) \geq e^{-2\pi} 2\pi \Delta. \]
This implies through Eq. (D.32) that
\[ \|V^{-1}_{S-1}\|_{\infty} \leq (\pi \Delta)^{-(S-1)} e^{2\pi (S-1)}, \]
so that
\[ \sigma_{\min}(V_{ST-1}) \geq (2\sqrt{S})^{-1} (\pi \Delta e^{-2\pi})^{(S-1)}, \]
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
\[ \|V_L\| \leq \|V_L\|_F = \left( \sum_{i,j} z_j^{2(i-1)} \right)^{1/2} \leq S^{1/2} \left( \sum_{i=1}^{L+1} z_{\max}^{2(i-1)} \right)^{1/2} = S^{1/2} \left( \frac{1-S^{(L+1)}}{1-z_{\max}^{2}} \right)^{1/2}, \]
which can tend to \((S(L + 1))^{1/2} \) when \( z_{\max} \to 1 \), so we use the simple upper bound \((S(L + 1))^{1/2} \leq (SK)^{1/2} \). Putting all this together allows to translate Eq. (D.31) to Eq. (D.40). The condition on \( \|H(\eta)\| \) in Eq. (D.30) then translates to the sufficient condition in Eq. (D.38) using that \( \sigma_{\min}(V_{K/2}) \geq \sigma_{\min}(V_{K/2-1}) \), the lower bound on \( \sigma_{\min}(V_{K/2-1}) \), and the upper bound on \( \|V_{K/2}\| \).

References

[1] D Aharonov and A Bredariol Grilo. “Stoquastic PCP vs. randomness”. In: 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS). 2019, pp. 1000–1023. DOI: 10.1109/FOCS.2019.00065.

[2] FSV Bazán. “Conditioning of rectangular Vandermonde matrices with nodes in the unit disk”. In: SIAM J. Matrix Anal. Appl. 21.2 (Oct. 1999), pp. 679–693. ISSN: 0895-4798. DOI: 10.1137/S0895479899336021.

[3] R Bhatia. Graduate Texts in Mathematics: Matrix analysis. Vol. 169. Springer Science & Business Media, 1997. DOI: 10.1007/978-1-4612-0653-8.

[4] S Bravyi. “Monte Carlo simulation of stoquastic Hamiltonians”. In: Quantum Inf. Comput. 15.13-14 (Oct. 2015), pp. 1122–1140. DOI: 10.5555/2871363.2871366.

[5] S Bravyi and D Gosset. “Polynomial-Time classical simulation of quantum ferromagnets”. In: Physical Review Letters 119.10 (Sept. 2017). DOI: 10.1103/PhysRevLett.119.100503.

[6] S Bravyi et al. “The complexity of stoquastic local Hamiltonian problems”. In: Quantum Inf. Comput. 8.5 (May 2008), pp. 361–385. DOI: 10.5555/2011772.2011773.

[7] AM Childs et al. “Theory of Trotter error with commutator scaling”. In: Phys. Rev. X 11 (1 Feb. 2021), p. 011020. DOI: 10.1103/PhysRevX.11.011020.

[8] E Crosson and AW Harrow. “Rapid mixing of path integral Monte Carlo for 1D stoquastic Hamiltonians”. In: Quantum 5 (Feb. 2021), p. 395. ISSN: 2521-327X. DOI: 10.22331/q-2021-02-11-395.

[9] E Crosson and S Sleczak. “Classical simulation of high temperature quantum Ising models”. Feb. 2020. arXiv: 2002.02232v1 [quant-ph].

[10] WMC Foulkes et al. “Quantum Monte Carlo simulations of solids”. In: Reviews of Modern Physics 73.1 (Jan. 2001). DOI: 10.1103/RevModPhys.73.33.

[11] W Gantschi. “On the inverses of Vandermonde and confluent Vandermonde matrices I.” In: Numer. Math 4 (1962), pp. 117–123. URL: https://www.cs.purdue.edu/homes/wxg/selected_works/section_01/016.pdf.

[12] SGharibian and F Le Gall. “Dequantizing the Quantum Singular Value Transformation: Hardness and Applications to Quantum Chemistry and the Quantum PCP Conjecture”. Nov. 2021. arXiv: 2111.09079 [quant-ph].
REFERENCES

[13] TNE Greville. “Note on the generalized inverse of a matrix product”. In: SIAM Review 8.4 (Oct. 1966), pp. 518–521. doi: 10.1137/1008107.

[14] J Gubernatis, N Kawashima, and P Werner. Quantum Monte Carlo Methods. Algorithms for lattice models. Cambridge University Press, 2016. doi: 10.1017/CB09780511902581.

[15] D Hangleiter et al. “Precise Hamiltonian identification of a superconducting quantum processor”. Aug. 2021. arXiv: 2108.08319 [quant-ph].

[16] J Helsen, F Battistel, and BM Terhal. “Spectral quantum tomography”. In: npj Quantum Information 5.74 (Sept. 2019). doi: 10.1038/s41534-019-0189-0.

[17] J Helsen et al. “A general framework for randomized benchmarking”. Oct. 2020. arXiv: 2010.07974 [quant-ph].

[18] RA Horn and CR Johnson. Matrix analysis. 2nd ed. Cambridge University Press, 2012. doi: 10.1017/9781139020411.

[19] Y Hua and TK Sarkar. “Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise”. In: IEEE Transactions on Acoustics, Speech and Signal Processing 38.5 (May 1990), pp. 814–824. doi: 10.1109/29.56027.

[20] Y Hua and TK Sarkar. “On SVD for estimating generalized eigenvalues of singular matrix pencil in noise”. In: IEEE Transactions on Signal Processing 39.4 (Apr. 1991), pp. 892–900. doi: 10.1109/78.80911.

[21] M Ioannou et al. ”Sign-curing local Hamiltonians: termwise versus global stoquasticity and the use of Clifford transformations”. July 2020. arXiv: 2007.11964 [quant-ph].

[22] RC Li. “Relative perturbation theory: II. Eigenspace and singular subspace variations”. In: SIAM Journal on Matrix Analysis and Applications 20.2 (1998), pp. 471–492. url: https://doi.org/10.1137/S0895479896298506.

[23] W Li, W Liao, and A Fannjiang. “Super-resolution limit of the ESPRIT algorithm”. In: SIAM Journal on Matrix Analysis and Applications 20.2 (1998), pp. 471–492. doi: 10.1137/S0895479896298506.

[24] L Lin and Y Tong. “Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers”. In: Proceedings of the forty-seventh annual ACM symposium on Theory of Computing (June 2015), pp. 821–830. doi: 10.1145/2746539.2746561.

[25] M Motta et al. “Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution”. In: Nature Physics 16.2 (Nov. 2019), pp. 205–210. issn: 1745-2481. doi: 10.1038/s41567-019-0704-4.

[26] MA Nielsen and IL Chuang. Quantum computation and quantum information: 10th Anniversary Edition. Cambridge University Press, 2010. doi: 10.1017/CBO9780511976667.

[27] E Onorati, AH Werner, and J Eisert. “Randomized benchmarking for individual quantum gates”. In: Physical Review Letters 123.6 (Aug. 2019). issn: 1079-7114. doi: 10.1103/physrevlett.123.060501.

[28] VY Pan. “How Bad Are Vandermonde Matrices?” In: SIAM Journal on Matrix Analysis and Applications 37.2 (2016), pp. 676–694. doi: 10.1137/15M1030170.

[29] D Potts and M Tasche. “Parameter estimation for nonincreasing exponential sums by Prony-like methods”. In: Linear Algebra and its Applications 439.4 (Aug. 2013), pp. 1024–1039. doi: 10.1016/j.laa.2012.10.036.
[34] S Sachdev. *Quantum phase transitions*. 2nd ed. Cambridge University Press, 2011. DOI: 10.1017/CBO9780511973765.

[35] TK Sarkar and O Pereira. “Using the matrix pencil method to estimate the parameters of a sum of complex exponentials”. In: *IEEE Antennas and Propagation Magazine* 37.1 (Feb. 1995), pp. 48–55. DOI: 10.1109/74.370583.

[36] RD Somma. “Quantum eigenvalue estimation via time series analysis”. In: *New Journal of Physics* 21.12 (Dec. 2019), p. 123025. DOI: 10.1088/1367-2630/ab5c60.

[37] G. W. Stewart. “Perturbation theory for the singular value decomposition”. In: *SVD and signal processing, II: algorithms, analysis and applications* (1991), pp. 99–109. URL: https://users.math.msu.edu/users/iwenmark/Teaching/MTH995/Papers/SVD_Stewart.pdf.

[38] P Stoica and A Nehorai. “MUSIC, maximum likelihood, and Cramer-Rao bound”. In: *IEEE Transactions on Acoustics, speech, and signal processing* 37.5 (May 1989), pp. 720–741. DOI: 10.1109/29.17564.

[39] MEHM Stroeks. “ClassQuantSimStoqHam”. https://github.com/MStroeks/ClassQuantSimStoqHam. 2022.

[40] M Suzuki. “General theory of fractal path integrals with applications to many-body theories and statistical physics”. In: *Journal of Mathematical Physics* 32.2 (Feb. 1991), pp. 400–407. DOI: 10.1063/1.529425.

[41] KM Svore, MB Hastings, and MH Freedman. “Faster phase estimation”. In: *Quantum Inf. Comput.* 14.3-4 (Mar. 2014), pp. 306–328. DOI: 10.26421/QIC14.3-4-7.

[42] K Wan, M Berta, and ET Campbell. “A randomized quantum algorithm for statistical phase estimation”. Oct. 2021. arXiv: 2110.12071 [quant-ph].

[43] PÅ Wedin. “Perturbation theory for pseudo-inverses”. In: *BIT Numerical Mathematics* 13.2 (June 1973), pp. 217–232. URL: https://doi.org/10.1007/BF01933494.