

UC Berkeley

UC Berkeley Previously Published Works

Title

Simultaneous estimation of multiple eigenvalues with short-depth quantum circuit on early fault-tolerant quantum computers

Permalink

<https://escholarship.org/uc/item/5sb3s83c>

Authors

Ding, Zhiyan

Lin, Lin

Publication Date

2023

DOI

10.22331/q-2023-10-11-1136

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Simultaneous estimation of multiple eigenvalues with short-depth quantum circuit on early fault-tolerant quantum computers

Zhiyan Ding¹ and Lin Lin^{1,2,3}

¹Department of Mathematics, University of California, Berkeley, CA 94720, USA

²Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

³Challenge Institute of Quantum Computation, University of California, Berkeley, CA 94720, USA

We introduce a multi-modal, multi-level quantum complex exponential least squares (MM-QCELS) method to simultaneously estimate multiple eigenvalues of a quantum Hamiltonian on early fault-tolerant quantum computers. Our theoretical analysis demonstrates that the algorithm exhibits Heisenberg-limited scaling in terms of circuit depth and total cost. Notably, the proposed quantum circuit utilizes just one ancilla qubit, and with appropriate initial state conditions, it achieves significantly shorter circuit depths compared to circuits based on quantum phase estimation (QPE). Numerical results suggest that compared to QPE, the circuit depth can be reduced by around two orders of magnitude under several settings for estimating ground-state and excited-state energies of certain quantum systems.

1 Introduction

The estimation of ground-state energies and excited-state energies of a Hamiltonian is a fundamental problem in quantum physics with numerous practical applications, such as in the design of new materials, drug discovery, and optimization problems. While the ground-state is often the state of interest for many quantum systems, excited-state energies also provide crucial information for understanding the electronic and optical properties of materials. Classical computers are limited in their ability to accurately estimate these energies for large-scale systems, and quantum computers have the potential to provide a significant speedup in solving such problems. Therefore, developing efficient and accurate methods for estimating ground-state and excited-state energies on quantum computers has become a major area of research in quantum information science.

When estimating multiple eigenvalues on quantum computers, there are two different strategies to consider. The first method involves preparing a variety of initial states that approximate different target eigenstates, and then estimating each eigenvalue one by one. The second method is to prepare a *single* initial state that has nontrivial overlap with all eigenstates of interest and estimate the eigenvalues simultaneously. The effectiveness of each approach depends on the assumptions and qualities of the initial state. This paper

Zhiyan Ding: zding.m@math.berkeley.edu

Lin Lin: linlin@math.berkeley.edu

concerns the second approach. Given a quantum Hamiltonian $H \in \mathbb{C}^{M \times M}$, we assume that we can prepare an initial quantum state $|\psi\rangle$ with K *dominant* modes. Specifically, let $\{(\lambda_m, |\psi_m\rangle)\}_{m=1}^M$ denote the eigenvalue and eigenvector pairs of H , and define $p_m = |\langle \psi_m | \psi \rangle|^2$ as the overlap between the initial state and the m -th eigenvector. We assume that we can choose a set $\mathcal{D} \subset \{1, 2, \dots, M\}$ satisfying $|\mathcal{D}| = K$, and $p_m = \Omega(R^{(K)})$ for any $m \in \mathcal{D}$. Here $R^{(K)} = \sum_{m' \in \mathcal{D}^c} p_{m'}$ is called the residual overlap, and $\mathcal{D}^c = \{1, 2, \dots, M\} \setminus \mathcal{D}$. The eigenvalues $\{\lambda_m\}_{m \in \mathcal{D}}$ are called the dominant eigenvalues of H with respect to the initial state $|\psi\rangle$ (or dominant eigenvalues for short), and the associated eigenvectors are called the dominant eigenvectors (or dominant modes). We assume $\{\lambda_m\}_{m \in \mathcal{D}} \subset [-\pi, \pi]$ for simplicity. Using an oracle access to the Hamiltonian evolution operators e^{-itH} for any $t \in \mathbb{R}$, we introduce an efficient quantum algorithm to estimate these dominant eigenvalues. We quantify the efficiency of a quantum algorithm by means of the maximal runtime denoted by T_{\max} , and the total runtime T_{total} . Assuming an algorithm needs to run a set of Hamiltonian evolution operators $\{\exp(-it_n H)\}_{n=1}^N$, then the maximal runtime is $T_{\max} = \max_{1 \leq n \leq N} |t_n|$ and the total runtime is $T_{\text{total}} = \sum_{n=1}^N |t_n|$. Here, T_{\max} and T_{total} approximately measure the circuit depth and the total cost of the algorithm, respectively, in a way that is oblivious to the details in implementing the Hamiltonian evolution operator e^{-itH} .

Our algorithm satisfies the following properties:

- (1) Allow a nonzero residual overlap: $R^{(K)} = \sum_{m' \in \mathcal{D}^c} p_{m'} > 0$.
- (2) Maintain Heisenberg-limited scaling [11],[41],[42]: To estimate all dominant eigenvalues to precision ϵ with probability $1 - \eta$, the total cost is $\mathcal{O}(\epsilon^{-1} \text{poly} \log(\epsilon^{-1} \eta^{-1}))$;
- (3) Use one ancilla qubit.
- (4) Reduce the circuit depth: the maximal runtime can be (much) lower than π/ϵ , especially when the ratio of residual overlap to the minimum overlap in the dominant set $R^{(K)}/(\min_{m \in \mathcal{D}} p_m)$ approaches zero.
- (5) Use the information of the spectral gap to further reduce maximal runtime: in the presence of a spectral gap Δ , when $\epsilon \ll \Delta$, the maximal runtime can be independent of ϵ as $\tilde{\mathcal{O}}(\Delta^{-1})$, and in this case, the total runtime is $\tilde{\mathcal{O}}(\epsilon^{-2})$ (see detail in Remark 2).

Our algorithm is designed to perform only classical postprocessing on the quantum data that is generated by the Hadamard test circuit, which uses one ancilla qubit. It is particularly well-suited for early fault-tolerant quantum devices that may be limited in the number of ancilla qubits and maximal coherence times.

The structure of this paper is organized as follows: In Section 2, we present the main idea of our method and provide a brief summary of related works. Next, in Section 3, we describe the main algorithm and provide the complexity results. The proof is included in the Appendix. Finally, we showcase the efficiency of our algorithm through numerical examples in Section 4.

2 Main idea and related work

In order to illustrate the main idea of the algorithm, we assume that we can use a quantum circuit (see Section 3.1) to accurately estimate

$$\langle \psi | \exp(-itH) | \psi \rangle = \sum_{m=1}^M p_m \exp(-i\lambda_m t) \quad (1)$$

for any $t \in \mathbb{R}$, where t is drawn from a probability distribution with a symmetric density $a(t)$, i.e., $a(t) \geq 0$, $a(t) = a(-t)$, and $\|a(t)\|_{L^1} = \int_{-\infty}^{\infty} a(t) dt = 1$. In our implementation, to estimate (1), we first use a classical computer to randomly sample t from distribution $a(t)$. Then, we execute the Hadamard test circuit (Fig. 2) on a quantum computer multiple times and average the output of measurement to estimate (1). We emphasize that the random selection of t plays an important role in improving the efficiency of our algorithm (see Section 2.1).

Define the Fourier transform of the probability density

$$F(x) = \int_{-\infty}^{\infty} a(t) \exp(ixt) dt, \quad (2)$$

which is a real-valued even function. As detailed in the following explanation, our approach requires carefully selecting an appropriate distribution $a(t)$ to ensure that the resulting filter function $F(x)$ reaches its maximum at $x = 0$ and decay rapidly as $|x|$ increases. There are several choices of $a(t)$ such that $F(x)$ concentrates around $x = 0$. In this paper, we choose $a(t)$ as the density function of a truncated Gaussian function:

$$a(t) = \frac{1}{A_\gamma \sqrt{2\pi T}} \exp\left(-\frac{t^2}{2T^2}\right) \mathbf{1}_{[-\gamma T, \gamma T]}(t), \quad \text{for } T, \gamma > 0, \quad (3)$$

where the normalization constant $A_\gamma = \int_{-\gamma}^{\gamma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$. The selection of $a(t)$ is guided by the observation that the Fourier transform of a Gaussian function remains a Gaussian function. More specifically, when $\gamma = \infty$, we have $F(x) = \exp\left(-\frac{T^2 x^2}{2}\right)$, which attains its maximal value at $x = 0$ and exponentially decays to zero with respect to $T|x|$. In addition, the use of the truncated Gaussian ensures that the maximal runtime never exceeds γT . We may also use the notation $a_T(t) = a(t)$, $F_T(x) = F(x)$ to emphasize the dependence on T .

From this we can define an *ideal* loss function as

$$\mathcal{L}_K \left(\{r_k\}_{k=1}^K, \{\theta_k\}_{k=1}^K \right) = \int_{-\infty}^{\infty} a(t) \left| \sum_{m=1}^M p_m \exp(-i\lambda_m t) - \sum_{k=1}^K r_k \exp(-i\theta_k t) \right|^2 dt. \quad (4)$$

Then we can estimate the dominant eigenvalues $\{\lambda_m\}_{m \in \mathcal{D}}$ by solving the optimization problem

$$\left(\{r_k^*\}_{k=1}^K, \{\theta_k^*\}_{k=1}^K \right) = \arg \min_{r_k \in \mathbb{C}, \theta_k \in \mathbb{R}} \mathcal{L}_K \left(\{r_k\}_{k=1}^K, \{\theta_k\}_{k=1}^K \right). \quad (5)$$

Intuitively, if the residual overlap $R^{(K)} = \sum_{m' \in \mathcal{D}^c} p_{m'}$ is small enough, to minimize the loss function, $\{(r_k^*, \theta_k^*)\}_{k=1}^K$ should be close to $\{(p_m, \lambda_m)\}_{m \in \mathcal{D}}$ so that the optimizer $\sum_{k=1}^K r_k^* \exp(-i\theta_k^* t)$ can eliminate the dominant term $\sum_{m \in \mathcal{D}} p_m \exp(-i\lambda_m t)$. For example, consider the case $K = 1$, we can obtain a closed form expression for the r_1 variable, and the optimization problem in Eq. (5) can be rewritten as

$$r_1^* = \sum_{m=1}^M p_m \int_{-\infty}^{\infty} a(t) \exp(i(\theta_1^* - \lambda_m)t) dt, \quad (6)$$

and

$$\theta_1^* = \arg \max_{\theta \in \mathbb{R}} \left| \sum_{m=1}^M p_m \int_{-\infty}^{\infty} a(t) \exp(i(\theta - \lambda_m)t) dt \right|^2 = \arg \max_{\theta \in \mathbb{R}} \left| \sum_{m=1}^M p_m F(\theta - \lambda_m) \right|. \quad (7)$$

Since $F(x)$ is a function that concentrates around $x = 0$, we expect that $\theta_1^* \approx \lambda_{m^*}$, where $m^* = \arg \max_m p_m$.

In practice, we can only generate an approximate value of $\langle \psi | \exp(-itH) | \psi \rangle$ up to statistical errors for a finite number of time steps t . As a result, the loss function we use in our optimization problem is only an approximation of the ideal loss function presented in Eq. (4). The main objective of this work is to generalize this idea to the estimation of multiple eigenvalues, and to rigorously implement this idea and control the complexity.

Using the concentration property of $F(x)$, we can demonstrate that the solution of the optimization problem is within δ/T of the dominant eigenvalues λ_m . Therefore by choosing $T_{\max} = \delta/\epsilon$, the parameter δ directly affects the circuit depth, and can be selected to be arbitrarily small if $R^{(K)}$ is small enough. Additionally, the solution of the optimization problem presented in Eq. (5) is robust to the measurement noise, which means that we do not need to generate a large number of data points and samples to construct the approximated optimization problem. This ensures that the algorithm can achieve the Heisenberg limit, i.e., $T_{\text{total}} = \mathcal{O}(1/\epsilon)$ (see Section 3.3).

2.1 Related work

If the initial quantum state $|\psi\rangle$ is a precise eigenvector of H with a single dominant eigenvalue, such as λ_1 , the Hadamard test algorithm is the simplest algorithm for estimating this eigenvalue with ϵ -accuracy. Given any real number $T \neq 0$, we can run the Hadamard test circuits (Fig. 2) several times to generate an estimate $Z_T \approx \langle \psi_1 | \exp(-iTH) | \psi_1 \rangle = \exp(-iT\lambda_1)$. We then define the approximation $\lambda_1^* = -\frac{\arg Z_T}{T}$. Because we assume $\lambda_1 \in [-\pi, \pi]$, it suffices to choose $T = 1$ and λ_1^* can be arbitrarily close to λ_1 by increasing the number of measurements. This observation implies that the maximal running time $T_{\max} = 1$ suffices to achieve arbitrary precision when employing the Hadamard test algorithm. Although the Hadamard test algorithm only utilizes a short circuit depth, it has several limitations: (1) the initial state $|\psi\rangle$ must be an exact eigenstate, (2) the total runtime T_{total} does not satisfy the Heisenberg-limited scaling and is $\Omega(\epsilon^{-2})$, and (3) it cannot estimate multiple eigenvalues. The first two limitations can be addressed by using quantum phase estimation (QPE) and its variants [17, 19, 26, 31, 1, 13, 12]. The textbook version of QPE [28] uses multiple ancilla qubits to store the phase and relies on inverse quantum Fourier transform (QFT). However, it can be difficult to employ the textbook version of QPE for multiple eigenvalue phase estimation due to the difficulty in handling spurious eigenvalues. In addition, although QPE can start from an imperfect eigenstate and saturates the Heisenberg-limited scaling, it requires the maximal runtime T_{\max} at least π/ϵ to achieve ϵ accuracy [28, Section 5.2.1].

Ref. [9] recently proposed a multiple eigenvalue phase estimation method that satisfies the Heisenberg limit scaling without any spectral gap assumptions. However, the theoretical analysis assumes that all non-dominant modes vanish [9, Definition 3.1 and Theorem 4.5], i.e., $p_{m'} = 0$ for any $m' \in \mathcal{D}^c$. Additionally, it has not yet been demonstrated that the method satisfies the short-depth property (4) described at the beginning of the paper.

An alternative way to understand the phase estimation problem is to view it as a sparse Fourier transform problem in classical signal processing. In recent years, the sparse Fourier transform problem has been extensively studied in both the discrete [38, 3, 40, 8, 2, 16] and continuous [32, 4, 15] settings. In particular, Ref. [32] concerns the recovery of the Fourier frequency of a continuous signal in the low noise setting. Our Theorem 1 shares similarities with [32, Theorem 1.1]. However, in our approach, each data point is generated by running the Hadamard test *once*, resulting in significantly higher noise level than that in [32]. Our optimization based method may also be easier to implement in practice. Very recently,

Ref. [21] introduces a robust multiple-phase estimation (RMPE) algorithm that can be considered as an extension of the multi-level robust phase estimation method (RPE) proposed in an earlier work [27]. By leveraging a multi-level signal processing technique [20], the RMPE algorithm can estimate multiple eigenvalues with Heisenberg-limited scaling without spectral gap assumptions. Under an additional spectral gap assumption, and with a different technique called ESPRIT [22] originally developed for super-resolution of signals and capable of efficiently estimating multiple eigenvalues in polynomial time [36], the resulting algorithm can satisfy the short-circuit depth property when the residual overlap $R^{(K)}/(\min_{m \in \mathcal{D}} p_m)$ is small.

Quantum subspace diagonalization (QSD) methods, quantum Krylov methods, and matrix pencil methods [5, 14, 18, 24, 25, 30, 33, 35, 29] estimates the eigenvalues by solving certain projected eigenvalue problems or singular value problems, and have been used to estimate ground-state and excited-state energies in a number of scenarios. Classical perturbation theories suggest that such projected problems can be highly ill-conditioned and the solutions are sensitive to noise. However, it has been observed that these methods can perform better than the pessimistic theoretical predictions. Recently, Ref. [10] provided a theoretical analysis explaining this phenomenon in the context of quantum subspace diagonalization methods used for computing the smallest eigenvalue. Nevertheless, the error of these methods generally increase with respect to the number of data points, which is related to the dimension of the projected matrix problem. In contrast, the error of our algorithm *decreases* with respect to the number of data points, making it much more robust to measurement noise. It is worth noting that the sharp complexity analysis of these methods (especially, with respect to the dimension of the projected matrix problem) remains open.

The loss function (4) utilized in this work is inspired by a recently developed subroutine called quantum complex exponential least squares (QCELS) [6]. QCELS employs the same quantum circuit as depicted in Fig. 2 to generate data and uses an optimization-based approach for estimating a single dominant eigenvalue. Although our method is also based on solving an optimization problem, it is important for us to emphasize that directly extending QCELS to estimate multiple eigenvalues **does not** yield satisfactory results. We detail the differences in the following:

- QCELS samples the time steps t on a uniform grid as $t_n = \frac{n}{N}T$ for $0 \leq n \leq N - 1$. In order to satisfy the Heisenberg-limited scaling, Ref. [6] proposed a multi-level QCELS algorithm that gradually increases the step size $\tau = T/N$ as one refines the estimate to the eigenvalue. However, such an approach can result in aliasing problems for multiple eigenvalue estimation since it may not be possible to distinguish $\exp(-i\lambda_{m_1}t_n)$ and $\exp(-i\lambda_{m_2}t_n)$ in the loss function if $(\lambda_{m_1} - \lambda_{m_2})T/N = 2\pi q$ for some $m_1 \neq m_2$ and $q \in \mathbb{Z}$. To overcome this difficulty, we combine the random sampling strategy [23, 37] with QCELS and sample t_n 's randomly from a probability density $a(t)$.
- For estimating a single dominant eigenvalue, our method also achieves improved theoretical results. In Ref. [6], QCELS requires $p_1 > 0.71$ and the circuit depth satisfies $T_{\max} = \delta/\epsilon$, where $\delta = \Theta(\sqrt{1-p_1})$. By applying our method to the single eigenvalue estimation problem, our analysis demonstrates that the method works for any $p_1 > 0.5$, and the constant δ can be improved to $\delta = \Theta(1-p_1)$. This improvement tightens the bound for the circuit depth. For clarity, the disparities in the constant dependence can be observed in Fig. 1.

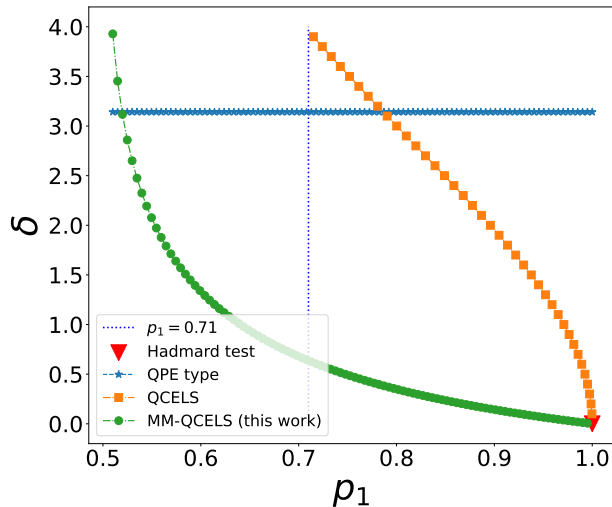


Figure 1: A comparison of the theoretical upper bound of $\delta = T_{\max}\epsilon$ for QCELS, QPE-type algorithm, and MM-QCELS for ground state energy estimation. The Hadamard test is only applicable when $p_1 = 1$ and the theoretical result in [6] requires $p_1 > 0.71$ for QCELS.

- As previously noted, QCELS [6] is formulated as a multi-level algorithm to attain Heisenberg-limited scaling and mitigate the aliasing problem. In our work, we also develop MM-QCELS as a multi-level method to help us control the random noise throughout the entire optimization domain. However, unlike QCELS [6], our approach can be simplified to solve the optimization problem with suitable parameters only once to approximate the dominant eigenvalues, resulting in a one-level algorithm. We put detailed discussion in the first comment of Remark 2.
- In the original QCELS [6], it was demonstrated that the maximum runtime can be further decreased to $T_{\max} = \tilde{\Theta}(1/(\lambda_2 - \lambda_1))$ by initially employing a ground state filter [23] to enhance p_1 . Our approach extends this characteristic to the estimation of multiple eigenvalues without requiring the use of a ground-state filter. This significantly simplifies the post-processing procedure. See the second comment of Remark 2.

We include a summary of the comparison among various methods for estimating multiple eigenvalues in Table 1.

3 Main method and informal complexity result

This section begins by presenting a quantum circuit for data generation. Subsequently, we use these data points to formulate an approximated optimization problem and propose the main algorithm. Finally, we provide an intuitive complexity analysis for our algorithm, followed by its rigorous statement.

Algorithms	Requirement				
	(1) Nonzero Residual	(2) Heisenberg limit	(3) Single ancilla	(4) Short depth	(5) Use gap
QSD type ([5]-[10])	✓	?	?	?	?
QEEA [34]	✓	✗	✓	?	?
ESPRIT [36]	?	?	✓	?	?
Ref. [9]	?	✓	✓	?	✗
Ref. [21, Theorem III.5]	✓	✓	✓	✗	✗
Ref. [21, Theorem IV.2]	✓	✓	✓	✓	?
MM-QCELS (this work)	✓	✓	✓	✓	✓

Table 1: Comparison of various methods for estimating multiple eigenvalues.

3.1 Generating data from quantum circuit

The quantum circuit used in this paper is the same as that used in the Hadamard test. In Fig. 2, we may

- (1) Set $W = I$, measure the ancilla qubit and define a random variable X_n such that $X_n = 1$ if the outcome is 0 and $X_n = -1$ if the outcome is 1. Then

$$\mathbb{E}(X_n) = \text{Re}(\langle \psi | \exp(-itH) | \psi \rangle). \quad (8)$$

- (2) Set $W = S^\dagger$ (S is the phase gate), measure the ancilla qubit and define a random variable Y_n such that $Y_n = 1$ if the outcome is 0 and $Y_n = -1$ if the outcome is 1. Then

$$\mathbb{E}(Y_n) = \text{Im}(\langle \psi | \exp(-itH) | \psi \rangle). \quad (9)$$

We assume an oracle access to the Hamiltonian evolution operator e^{-itH} for any $t \in \mathbb{R}$. This assumption is not overly restrictive. For instance, if we can approximate $e^{-i\tau H}$ for some desired $\tau > 0$ using the Trotter method of a certain order (the choice of τ and the order of the Trotter splitting is problem dependent and should balance between efficiency and accuracy), we may express $e^{-itH} = (e^{-i\tau H})^p e^{-i\tau' H}$, where $t = p\tau + \tau'$, $p \in \mathbb{Z}$, and $|\tau'| < \tau$. The cost for implementing $e^{-i\tau' H}$ should be no larger than that of implementing $e^{-i\tau H}$. The Trotter error can be systematically controlled and this is an orthogonal issue to the type of error considered in this paper (see, e.g., [23, Appendix D] for how to factor such errors into the analysis). Therefore, throughout the paper, we assume that the implementation of e^{-itH} for any $t \in \mathbb{R}$ is exact.

Given a set of time points $\{t_n\}_{n=1}^N$ drawn from the probability density $a(t)$, we use the quantum circuit in Fig. 2 to prepare the following data set:

$$\mathcal{D}_H = \{(t_n, Z_n)\}_{n=1}^N. \quad (10)$$

By running the quantum circuit (Fig. 2) for each t_n *once*, we define the output

$$Z_n = X_n + iY_n. \quad (11)$$

Here X_n, Y_n are independently generated by the quantum circuit (Fig. 2) with different W and satisfy (8), (9) respectively. Hence, we have

$$\mathbb{E}(Z_n) = \langle \psi | \exp(-it_n H) | \psi \rangle, \quad |Z_n| \leq 2. \quad (12)$$

Therefore Z_n is an unbiased estimation $\langle \psi | \exp(-it_n H) | \psi \rangle = \sum_{m=1}^M p_m \exp(-i\lambda_m t_n)$. We also note that if we use the above method to prepare the data set in Eq. (10), the maximal simulation time is $T_{\max} = \max_{1 \leq n \leq N} |t_n|$ and the total simulation time is $\sum_{n=1}^N |t_n|$.

Given t_n , we can sample the circuit N_s times such that $|Z_n - \langle \psi | \exp(-it_n H) | \psi \rangle| = \mathcal{O}(1/\sqrt{N_s})$. Most algorithms, including the QCELS algorithm in Ref. [6], require this repetition step. In this aspect, our algorithm is innovative in that it achieves convergence with just a single sample per t_n , i.e., $N_s = 1$. For simplicity, we assume $N_s = 1$ throughout the paper. Increasing N_s reduces the statistical noise in Z_n and can further decrease the error in the eigenvalue estimate. This also increases T_{total} by a factor of N_s without affecting T_{\max} .

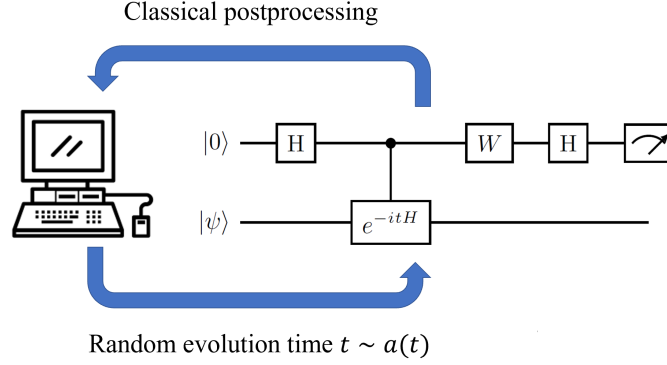


Figure 2: Choosing $W = I$ or $W = S^\dagger$ (S is the phase gate), the Hadamard test circuit allows us to estimate the real or the imaginary part of $\langle \psi | \exp(-itH) | \psi \rangle$ on quantum computers. H is the Hadamard gate. The classical computer provides the evolution time t according to a probability density $a(t)$, and performs postprocessing on the quantum data for eigenvalue estimation.

Algorithm 1 describes the data generating process.

Algorithm 1 Data generator

- 1: **Preparation:** Number of data pairs: N ; probability density: $a(t)$;
 - 2: $n \leftarrow 1$;
 - 3: **while** $n \leq N$ **do**
 - 4: Generate a random variable t_n with the probability density $a(t)$.
 - 5: Run the quantum circuit (Figure 2) with $t = t_n$ and $W = I$ to obtain X_n .
 - 6: Run the quantum circuit (Figure 2) with $t = t_n$ and $W = S^\dagger$ to obtain Y_n .
 - 7: $Z_n \leftarrow X_n + iY_n$.
 - 8: $n \leftarrow n + 1$
 - 9: **end while**
 - 10: **Output:** $\{(t_n, Z_n)\}_{n=1}^N$
-

3.2 Main algorithm

After generating the data set, we define the numerical loss function (referred to as the loss function for short) as

$$L_K \left(\{r_k\}_{k=1}^K, \{\theta_k\}_{k=1}^K \right) = \frac{1}{N} \sum_{n=1}^N \left| Z_n - \sum_{k=1}^K r_k \exp(-i\theta_k t_n) \right|^2 \quad (13)$$

and the optimization problem

$$\left(\{r_k^*\}_{k=1}^K, \{\theta_k^*\}_{k=1}^K\right) = \arg \min_{r_k \in \mathbb{C}, \theta_k \in \mathbb{R}} L_K \left(\{r_k\}_{k=1}^K, \{\theta_k\}_{k=1}^K\right). \quad (14)$$

Compared to the ideal loss function described in Eq. (4), the current loss function is considerably noisier. Define the expectation error $E_n = Z_n - \sum_{m=1}^M p_m \exp(-i\lambda_m t_n)$. We note that $|E_n|$ is bounded by 3 but not small since Z_n is generated by running the quantum circuit only once. On the other hand, the expectation of E_n is zero and $\{E_n\}_{n=1}^N$ are independent. This implies $\left|\frac{1}{N} \sum_{n=1}^N E_n\right| = \mathcal{O}\left(1/\sqrt{N}\right)$. Define the vector $\mathbf{r} = (r_1, r_2, \dots, r_K)$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K)$, and plug this into (13), we have

$$\begin{aligned} & \arg \min_{\mathbf{r}, \boldsymbol{\theta}} L_K(\mathbf{r}, \boldsymbol{\theta}) \\ &= \arg \min_{\mathbf{r}, \boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^N \left| Z_n - \sum_{k=1}^K r_k \exp(-i\theta_k t_n) \right|^2 \\ &= \arg \min_{\mathbf{r}, \boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^N \left| \sum_{m=1}^M p_m \exp(-i\lambda_m t_n) - \sum_{k=1}^K r_k \exp(-i\theta_k t_n) \right|^2 \\ &\quad - \frac{2}{N} \sum_{n=1}^N \operatorname{Re} \left\langle E_n, \sum_{k=1}^K r_k \exp(-i\theta_k t_n) \right\rangle \\ &\approx \arg \min_{\mathbf{r}, \boldsymbol{\theta}} \int_{-\infty}^{\infty} a(t) \left| \sum_{m=1}^M p_m \exp(-i\lambda_m t) - \sum_{k=1}^K r_k \exp(-i\theta_k t) \right|^2 dt \\ &= \arg \min_{\mathbf{r}, \boldsymbol{\theta}} \mathcal{L}_K(\mathbf{r}, \boldsymbol{\theta}), \end{aligned} \quad (15)$$

where $\langle a, b \rangle = \bar{a} \cdot b$ for any $a, b \in \mathbb{C}$. In the second equality, we omit terms that are independent of $(\mathbf{r}, \boldsymbol{\theta})$. In the approximation step, we use $\frac{2}{N} \sum_{n=1}^N \operatorname{Re} \left\langle E_n, \sum_{k=1}^K r_k \exp(-i\theta_k t_n) \right\rangle \approx 0$ when $N \gg 1$. While the loss function L_K may not converge to the ideal loss function as N approaches infinity due to the statistical noise in Z_n , we find that the optimization problem in Eq. (14) can still yield a solution that approaches that in Eq. (5) with the ideal loss.

After formulating the loss function (13) and the optimization problem (14), we are ready to introduce our main algorithm. Inspired by the multi-level QCELS [6], the algorithm contains two steps:

- **Step 1:** Choose a relative small T_0 . Set $T = T_0$, $a(t) = a_{T_0}(t)$ and solve the optimization problem (14) to give a rough estimation for each dominant eigenvalues $\{\lambda_m\}_{m \in \mathcal{D}}$.
- **Step 2:** Gradually increase T so that:
 1. The solution with T_j gives a good initial guess for the optimization problem with T_{j+1} .
 2. The total running time still satisfies the Heisenberg-limit.

The detailed algorithm is summarized in Algorithm 2, which is referred to as the multi-modal, multi-level quantum complex exponential least squares (MM-QCELS) algorithm. We note that the optimization problems (16) in Algorithm 2 have the technical constraint $\|\mathbf{r}\|_1 \leq 1$. This constraint is natural since each r_k should approximate some p_{m_k} , and

the sum of the absolute value thus should exceed 1. This constraint is added to simplify the complexity analysis for a uniform upper bound of the expectation error (see Lemma 8 in Appendix E for detail). In practice, we find that the performance of the algorithm is independent of this constraint (see the numerical tests in Section 4).

Algorithm 2 Multi-modal, multi-level quantum complex exponential least squares (MM-QCELS)

- 1: **Preparation:** Number of data pairs: $\{N_j\}_{j=0}^l$; number of iterations: l ; sequence of time steps: $\{T_j\}_{j=0}^l$; sequence of probability distributions $\{a_{T_j}(t)\}_{j=0}^l$; number of dominant eigenvalues: K
 - 2: **Running:**
 - 3: $k \leftarrow 1$; ▷ Step 1 starts
 - 4: **while** $k \leq K$ **do**
 - 5: $\lambda_{\min,k} \leftarrow -\pi$; $\lambda_{\max,k} \leftarrow \pi$; ▷ $[\lambda_{\min,k}, \lambda_{\max,k}]$ is the estimation interval.
 - 6: $k \leftarrow k + 1$
 - 7: **end while**
 - 8: $j \leftarrow 0$;
 - 9: **while** $j \leq l$ **do**
 - 10: Generate a data set $\{(t_n, Z_n)\}_{n=1}^{N_j}$ using Algorithm 1 with $a_{T_j}(t)$.
 - 11: Define the loss function $L_K(r, \theta)$ in Eq. (13).
 - 12: Minimizing loss function:

$$(\mathbf{r}^*, \boldsymbol{\theta}^*) \leftarrow \arg \min_{\|\mathbf{r}\|_1 \leq 1, \theta_k \in [\lambda_{\min,k}, \lambda_{\max,k}]} L_K(\mathbf{r}, \boldsymbol{\theta}) . \quad (16)$$
 - 13: $\lambda_{\min,k} \leftarrow \theta_k^* - \frac{\pi}{T_j}$; $\lambda_{\max,k} \leftarrow \theta_k^* + \frac{\pi}{T_j}$ ▷ Shrink the search interval.
 - 14: $j \leftarrow j + 1$
 - 15: **end while** ▷ Step 2 ends
 - 16: **Output:** $\{(\mathbf{r}^*, \boldsymbol{\theta}^*)\}_{k=1}^K$
-

3.3 Complexity analysis of Algorithm 2

In this section, we study the complexity of Algorithm 2. First, the cost of the algorithm depends on a number of parameters used throughout the analysis, including

1. The minimal dominant overlap:

$$p_{\min}^{(K)} = \min_{m \in \mathcal{D}} p_m ; \quad (17)$$

2. The minimal dominant spectral gap:

$$\Delta_{\lambda}^{(K)} = \min_{m, m' \in \mathcal{D}, m \neq m'} |\lambda_{m'} - \lambda_m| ; \quad (18)$$

3. The residual overlap:

$$R^{(K)} = \sum_{m' \in \mathcal{D}^c} p_{m'} . \quad (19)$$

Now, we are ready to introduce the complexity result of Algorithm 2, which is summarized in the following theorem.

Theorem 1. *Given the failure probability $0 < \eta < 1/2$, error $\epsilon > 0$, and any $\zeta > 0$. Assume $p_{\min}^{(K)} > 3R^{(K)}$ and define $q = \Theta(R^{(K)}/p_{\min}^{(K)})$. In Algorithm 2, we choose the following parameters:*

- Set $\gamma = \Theta\left(\log\left(1/\min\left\{p_{\min}^{(K)}q, \left(p_{\min}^{(K)}\right)\left(p_{\min}^{(K)} - 3R^{(K)}\right)\right\}\right)\right)$.
- For Step 1, set

$$T_0 = \tilde{\Theta}\left(\left(\Delta_\lambda^{(K)}\right)^{-1} \log(q^{-1})\right), \quad N_0 = \tilde{\Theta}\left(T_0^2 \left(p_{\min}^{(K)}\left(p_{\min}^{(K)} - 3R^{(K)}\right)\right)^{-2} \text{poly log}(\eta^{-1})\right). \quad (20)$$

- For Step 2, set $l = \max\{\lceil \log_2(q/(\epsilon T_0)) \rceil, 1\}$, $T_j = 2^j T_0$, and

$$N_j = \tilde{\Theta}\left(\left(p_{\min}^{(K)}\right)^{-4} q^{-2-\zeta} \text{poly log}(\log(\zeta^{-1})\eta^{-1})\right) \quad (21)$$

for $1 \leq j \leq l$.

This gives

$$T_{\max} = \frac{\delta}{\epsilon}, \quad T_{\text{total}} = \tilde{\Theta}\left(\frac{1}{\left(p_{\min}^{(K)}\right)^4 \delta^{1+\zeta\epsilon}} \text{poly log}(\log(\zeta^{-1})\delta^{-1}\eta^{-1})\right). \quad (22)$$

where $\delta = \tilde{\Theta}(q \log(q^{-1}))$. In the above equations, the constant in $\tilde{\Theta}$ depends at most polynomially on $\log\left(\left(p_{\min}^{(K)}\left(p_{\min}^{(K)} - 3R^{(K)}\right)\right)^{-1} K\right)$. Then with probability $1 - \eta$, for any $m \in \mathcal{D}$, there exists a unique $1 \leq k_m \leq K$ such that

$$|\lambda_m - \theta_{k_m}^*| \leq \epsilon. \quad (23)$$

Here, ζ can be chosen arbitrarily close to 0 and the constant in $\tilde{\Theta}$ depends on ζ .

We put the proof of the above theorem in the Appendix. To provide a clear exposition of the core concept, we first present the intuitive idea of the proof in Appendix B. The key step in our proof is to demonstrate that solving the ideal optimization problem (5) could yield a precise approximation to dominant eigenvalues with a short maximal running time. The rigorous proof of the theorem is then given in Appendix C. In particular, we rigorously bound the approximation error in (15), and optimizing the numerical loss function (13) gives us an accurate approximation of dominant eigenvalues with low cost.

Remark 2. *The results of Theorem 1 can be extended as follows:*

1. In the above theorem, we could relax the condition that $T_0 = \tilde{\Theta}((\Delta_\lambda^K)^{-1} \log(q^{-1}))$. To be more precise, if we are given a lower bound Δ_{low} for Δ_λ^K , we could set $T_0 = \tilde{\Theta}((\Delta_{\text{low}})^{-1} \log(q^{-1}))$, and the result of the theorem still holds.
2. In Theorem 1 and Algorithm 2, we utilize a sequence of T_n to approximate the dominant eigenvalues. This is due to a technical consideration in our theoretical analysis. Specifically, in order to ensure the feasibility of optimization problem (16), our proof requires that the random discrepancy between (16) and the ideal optimization problem (5) remains uniformly small when $\theta_k \in [\lambda_{\min,k}, \lambda_{\max,k}]$ (see Appendix C Lemma 4 for detail). Achieving this requires a uniform bound for infinitely many continuous random variables, which necessitates constraining $\lambda_{\max,k} - \lambda_{\min,k}$ to be sufficiently

small, thus guaranteeing the manageability of random noise with a reasonable number of samples. Detailed information can be found in Appendix E.

In practical calculations, Algorithm 2 may be simplified into a one-level algorithm by directly choosing sufficiently large values for T_{\max} and N . Moreover, if we optimize our function only over a finite number of grid intervals such that $\theta_k \in -\pi + k\epsilon, 0 \leq k < \frac{2\pi}{\epsilon}$, it is theoretically possible to show that by minimizing (16) once with $T_{\max} = \tilde{\Theta}(q/\epsilon)$ and $N = \tilde{\Theta}(q^{-2-o(1)})$, we can achieve ϵ -accuracy for all dominant eigenvalues. This approach also offers benefits such as short circuit depth and Heisenberg-limited scaling.

3. When a spectral gap exists between the dominant eigenvalues and the non-dominant ones (represented as Δ_λ), it may be feasible to further reduce the maximum runtime to $T_{\max} = \tilde{\Theta}(1/\min\{\Delta_\lambda, \Delta_\lambda^{(K)}\})$. The reason is given in the first point of Remark 3.

4 Numerical results

In this section, we numerically demonstrate the efficiency of our method using two different models. In Section 4.1, we compare the performance of Algorithm 2 with QPE (textbook version [28]) for a transverse-field Ising model. In Section 4.2, we compare the performance of Algorithm 2 with QPE for a Hubbard model. In both cases, we assume there are two dominant eigenvalues (λ_1, λ_2), meaning $K = 2$. We share the code on Github (<https://github.com/zhiyanding/MM-QCELS>).

In our numerical experiments, we normalize the spectrum of the original Hamiltonian H so that the eigenvalues belong to $[-\pi/4, \pi/4]$. Given a Hamiltonian H , we use the normalized Hamiltonian in the experiment:

$$\tilde{H} = \frac{\pi H}{4\|H\|_2}. \quad (24)$$

Recall that the textbook version of QPE [28] relies on a quantum circuit that involves the inverse Quantum Fourier Transform (QFT). This circuit serves to encode the information of approximate eigenvalues using the ancilla register. By measuring the ancilla qubits, we could acquire an output θ that closely approximates λ_k , one of the eigenvalues of Hamiltonian H . To find an approximation to the smallest eigenvalue, we repeat the quantum circuit for N times and defines the approximation $\theta_1^* = \min_{1 \leq n \leq N} \theta_n$, where θ_n is the output of n -th repetition. The analysis of this method can be found in e.g., [23, Section I.A]. However, QPE can also produce spurious eigenvalues which lead to failures, and it is not straightforward to generalize the procedure above for estimating multiple eigenvalues.

Consequently, in our experiment, we first utilize QPE to estimate the smallest eigenvalue λ_1 and measure the error accordingly. We then use Algorithm 2 to estimate the two dominant eigenvalues and measure the error by (assuming $\theta_1^* \leq \theta_2^*$)

$$\text{error} = \max\{|\theta_1^* - \lambda_1|, |\theta_2^* - \lambda_2|\}. \quad (25)$$

For simplicity, in this section, Algorithm 2 is implemented without the constraint $\|\mathbf{r}\|_1 < 1$ in (16). While QPE's error is gauged based on a single eigenvalue, the error of Algorithm 2 is evaluated using the maximum error across two eigenvalues. This testing design intrinsically gives QPE a head start. Even with this bias, we demonstrate that Algorithm 2 can outperform QPE.

4.1 Ising model

Consider the one-dimensional transverse field Ising model (TFIM) model defined on L sites with periodic boundary conditions:

$$H = - \left(\sum_{i=1}^{L-1} Z_i Z_{i+1} + Z_L Z_1 \right) - g \sum_{i=1}^L X_i \quad (26)$$

where g is the coupling coefficient, Z_i, X_i are Pauli operators for the i -th site and the dimension of H is 2^L . We choose $L = 8, g = 4$. In the test, we set $p_1 = p_2 = 0.4$. In Algorithm 2, the parameters are set to be $K = 2, T_0 = 2(\lambda_2 - \lambda_1)^{-1}, N_0 = 3 \times 10^3, N_{j \geq 1} = 2 \times 10^3$ and $\gamma = 1$. An illustrative example, Fig 3 demonstrates the landscape of the loss function in Eq. (13). As T increases, the landscape of the loss function becomes increasingly complex. However, the value of the loss function around the true eigenvalues decreases significantly, which also leads to a reduction in run-to-run variation of the optimizer. As a result, the optimizer concentrates more tightly around the true eigenvalues.

We apply Algorithm 2 and QPE to estimate the dominant eigenvalues (λ_1, λ_2) of the normalized Hamiltonian \tilde{H} according to Eq. (24). We then run Algorithm 2 (with $K = 2, T_0 = 2(\lambda_2 - \lambda_1)^{-1}, N_0 = 3 \times 10^3, N_{j \geq 1} = 2 \times 10^3$ and $\gamma = 1$) to compute the error of both λ_1 and λ_2 . We also run QPE 10 times *only* to estimate λ_1 . The comparison of the results is shown in Fig. 4. We find that the errors of both QPE and Algorithm 2 are proportional to the inverse of the maximal evolution time (T_{\max}). But the constant factor $\delta = T\epsilon$ of Algorithm 2 is much smaller than that of QPE. Fig. 4 shows that Algorithm 2 reduces the maximal evolution time by two orders of magnitude. The error of QPE is observed to scale as $6\pi/T$. Moreover, the total evolution time (T_{total}) of Algorithm 2 is also slightly smaller than that of QPE.

According to Theorem 1, accurate estimation of the dominant eigenvalues with a short circuit depth using MM-QCELS depends on two critical factors: the appropriate selection of the parameter K and the fulfillment of the condition $R^{(K)}/p_{\min}^{(K)} \ll 1$. We would like to emphasize that these criteria are necessary for addressing worst-case scenarios. However, in practical implementations, even if a slightly larger value of K is chosen and the ratio $R^{(K)}/p_{\min}^{(K)}$ approximates 1, Algorithm 2 is still possible to produce a precise approximation of the dominant eigenvalues with short circuit depth. In Appendix A, we test these two cases and demonstrate the robustness of MM-QCELS to the choice of parameters.

4.2 Hubbard model

Consider the one-dimensional Hubbard model defined on L spinful sites with open boundary conditions

$$H = -t \sum_{j=1}^{L-1} \sum_{\sigma \in \{\uparrow, \downarrow\}} c_{j,\sigma}^\dagger c_{j+1,\sigma} + U \sum_{j=1}^L \left(n_{j,\uparrow} - \frac{1}{2} \right) \left(n_{j,\downarrow} - \frac{1}{2} \right).$$

Here $c_{j,\sigma}(c_{j,\sigma}^\dagger)$ denotes the fermionic annihilation (creation) operator on the site j with spin σ . $\langle \cdot, \cdot \rangle$ denotes sites that are adjacent to each other. $n_{j,\sigma} = c_{j,\sigma}^\dagger c_{j,\sigma}$ is the number operator.

We choose $L = 4, 8, t = 1, U = 10$. To implement Algorithm 2 and QPE, we again normalize H according to Eq. (24) and choose overlap $p_1 = 0.4, p_2 = 0.4$. We run Algorithm 2 (with $K = 2, T_0 = 10(\lambda_2 - \lambda_1)^{-1}, N_0 = 4 \times 10^4, N_{j \geq 1} = 2 \times 10^3, \gamma = 1$), and QPE 10 times to compare the errors (using Eq. (25) for Algorithm 2 and only estimating the error

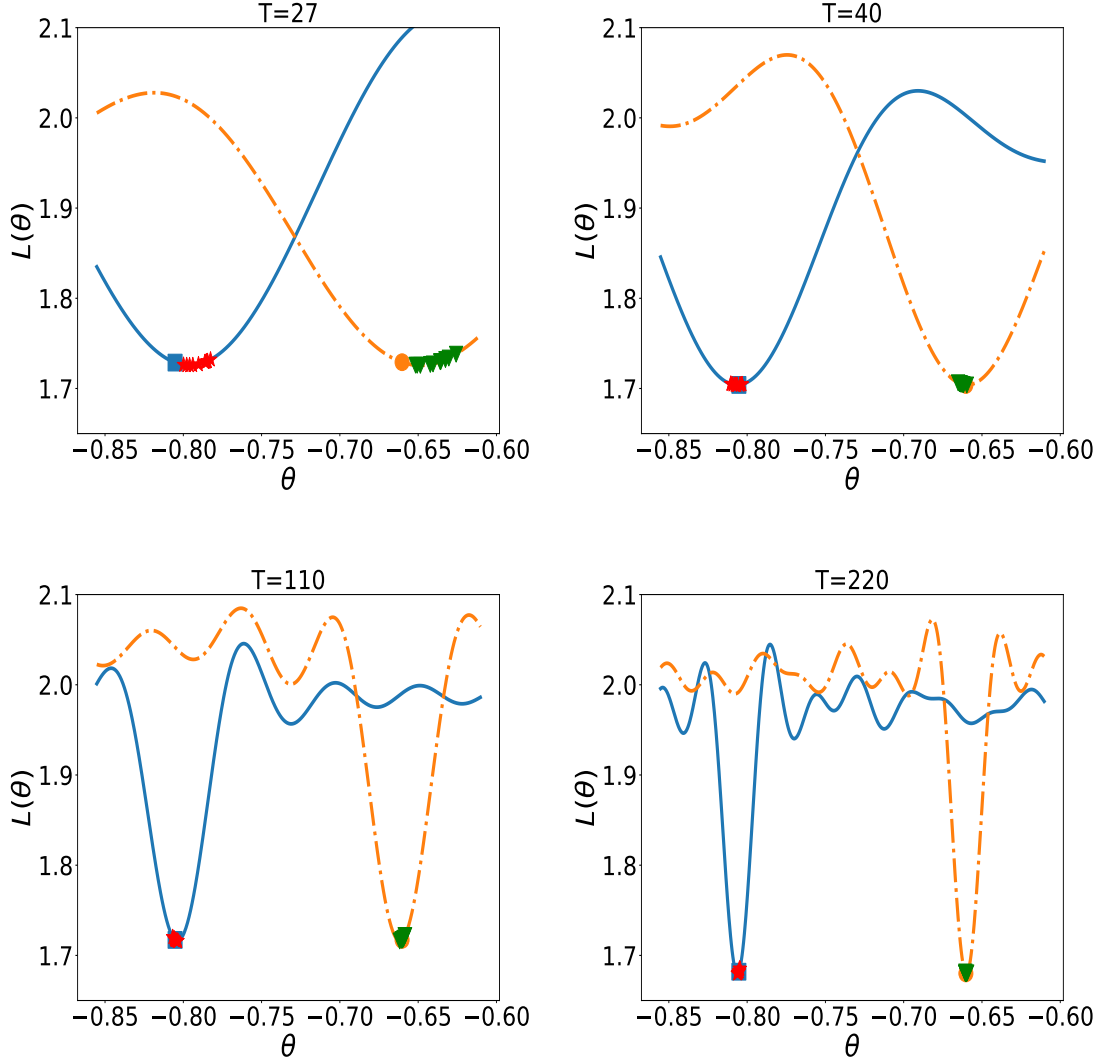


Figure 3: The landscape of the loss function L (13) from the TFIM model at time T_l with $p_1 = 0.4$, $p_2 = 0.4$, $K = 2$, $T_0 = 2(\lambda_2 - \lambda_1)^{-1}$, $N_0 = 3 \times 10^3$, $N_{j \geq 1} = 2 \times 10^3$, $\gamma = 1$. We run the experiment ten times and plot the positions of ten minimizers θ^* using star (θ_1^*) and triangle (θ_2^*) points. The landscape of the loss function is calculated using the data from the last experiment. The blue solid line stands for $L_K(\mathbf{r}^*, \theta_1, \theta_2^*)$ (the variable is θ_1), and the blue square is the true eigenvalue λ_1 . The orange dash-dotted stands for $L_K(\mathbf{r}^*, \theta_1^*, \theta_2)$ (the variable is θ_2), and the orange filled circle is the true eigenvalue λ_2 . When T is large, the minimizer of the loss function concentrates around the dominant eigenvalues λ_1 and λ_2 .

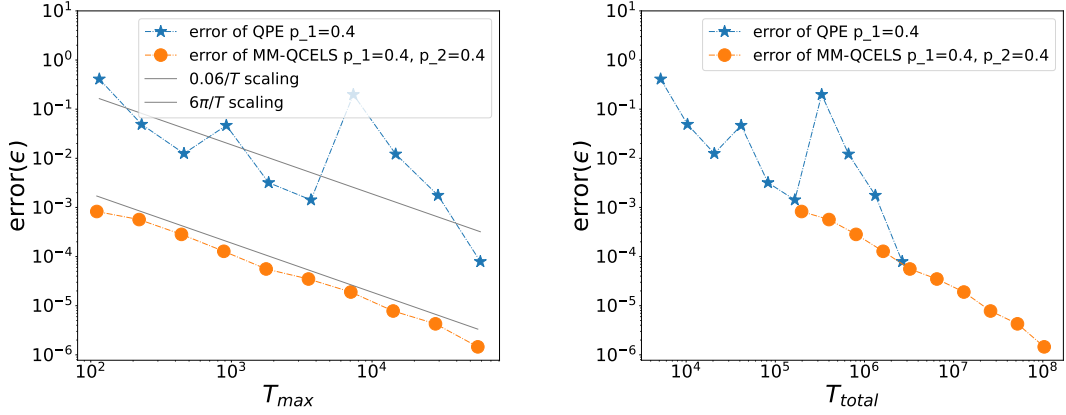


Figure 4: QPE vs Algorithm 2 in TFIM model with 8 sites. $p_1 = p_2 = 0.4$. Left: Depth (T_{\max}); Right: Cost (T_{total}). For Algorithm 2, we choose $K = 2, T_0 = 2(\lambda_2 - \lambda_1)^{-1}, N_0 = 3 \times 10^3, N_{j \geq 1} = 2 \times 10^3, \gamma = 1$. l, T_j are chosen according to Theorem 1. Both methods have the error scales linearly in $1/T_{\max}$. The constant factor $\delta = T\epsilon$ of Algorithm 2 is much smaller than that of QPE.

of λ_1 for QPE). The results are shown in Fig. 5 (4 sites) and Fig. 6 (8 sites). In both cases, it can be seen that the maximal evolution time of Algorithm 2 is almost two orders of magnitude smaller than that of QPE. The total cost of the two methods are comparable.

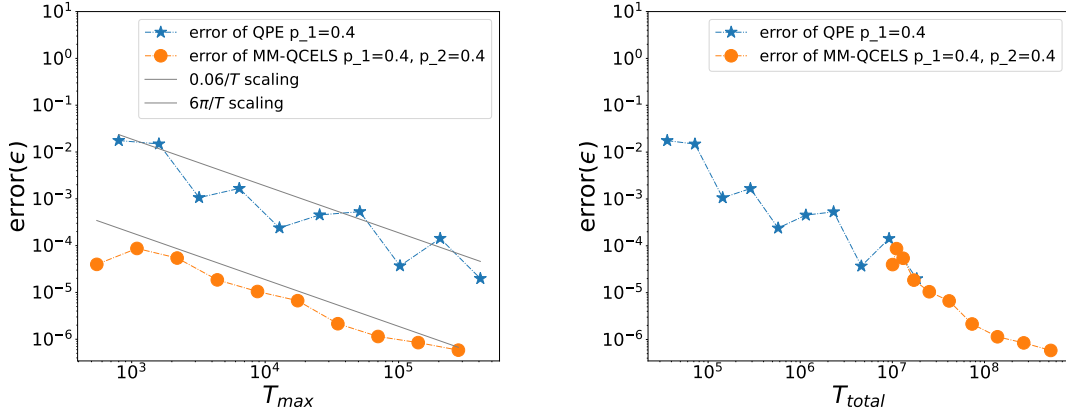


Figure 5: QPE vs Algorithm 2 in Hubbard model with 4 sites. $p_1 = p_2 = 0.4$. Left: Depth (T_{\max}); Right: Cost (T_{total}). For Algorithm 2, we choose $K = 2, T_0 = 10(\lambda_2 - \lambda_1)^{-1}, N_0 = 4 \times 10^4, N_{j \geq 1} = 2 \times 10^3$. l, T_j are chosen according to Theorem 1. Compared with QPE, to achieve the same accuracy, Algorithm 2 requires a much smaller circuit depth.

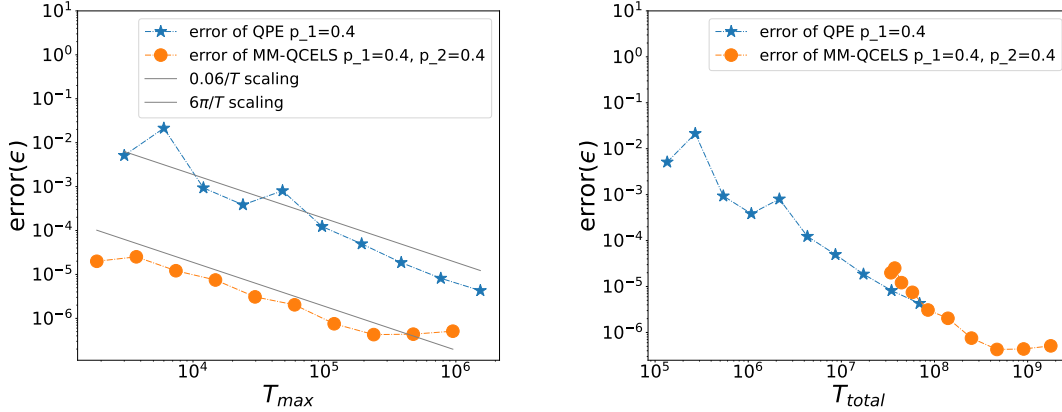


Figure 6: QPE vs Algorithm 2 in Hubbard model with 8 sites. $p_1 = p_2 = 0.4$. Left: Depth (T_{\max}); Right: Cost (T_{total}). For Algorithm 2, we choose $K = 2, T_0 = 10(\lambda_2 - \lambda_1)^{-1}, N_0 = 4 \times 10^4, N_{j \geq 1} = 2 \times 10^3, \gamma = 1$. l, T_j are chosen according to Theorem 1 with $\epsilon' = 0$. Compared with QPE, to achieve the same accuracy, Algorithm 2 requires a much smaller circuit depth.

5 Discussion

In this paper, we have developed a new algorithm to simultaneously estimate multiple eigenvalues using a simple circuit. This optimization-based approach, called multimodal, multi-level quantum complex exponential least squares (MM-QCELS) saturates the Heisenberg-limited scaling, and achieves a relatively short maximal running time when the residual overlap (denoted by $R^{(K)}$) is small. With a proper initial state, this algorithm can be used to efficiently estimate ground-state and excited-state energies of a quantum many-body Hamiltonian on early fault-tolerant quantum computers.

There are a number of directions to extend this work and to strengthen the analysis.

1. If the initial state has significant high energy contributions, it can be implicitly filtered using the circuit in Fig. 2 [23], or explicitly filtered using quantum eigenvalue transformation of unitary matrices (QETU) [7] to satisfy the condition of small residual overlap. Similar to the discussion in Ref. [6], this does not necessarily require a large spectral gap between the dominant eigenvalues and the non-dominant ones, and a *relative overlap* condition (which is a property of the initial state rather than the Hamiltonian) could suffice.
2. Our complexity analysis depends on the minimal dominant spectral gap $\Delta_\lambda^{(K)}$. This is a necessary condition, since the simulation time should be long enough in order to separate the eigenvalues. If two or more eigenvalues are nearly degenerate (i.e., the distance is less than ϵ), we may combine them and view these nearly degenerate eigenvalues as a single eigenvalue. The MM-QCELS method can still be applied without changes. However, compared to the result in Theorem 1, there may not be a one-to-one correspondence between the estimated eigenvalues θ_k and the dominant eigenvalues λ_{m_k} .
3. Due to the complex energy landscape of the loss function, if we use a grid search to find the global minima, the cost of solving the optimization problem in Eq. (14) on a *classical computer* may scale exponentially in K in the worst case. On the other

hand, the signal processing method in [32, 21] can scale polynomially in K , but the implementation can be much more complicated than ours. Therefore it is desirable to improve our algorithm (e.g., using a robust initialization strategy) to reduce the cost on the classical post-processing step for large K .

4. While Algorithm 2 is formulated as a multi-level optimization problem, as discussed in Remark 2, solving (14) with sufficiently large values of T and N once may be enough to approximate the dominant eigenvalues. This approach differs from the signal processing-based method discussed in [21], which requires multi-level signal processing.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator (L.L.). Additional support is acknowledged from the NSF Quantum Leap Challenge Institute (QLCI) program under Grant number OMA-2016245 (Z.D.), and a Google Quantum Research Award (L. L.). L.L. is a Simons Investigator. The authors thank Haoya Li, Hongkang Ni, Lexing Ying, Ruizhe Zhang for helpful discussions.

References

- [1] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman. How to perform the most accurate possible phase measurements. *Phys. Rev. A*, 80(5):052114, 2009. doi:10.1103/PhysRevA.80.052114.
- [2] P. Boufounos, V. Cevher, A. C. Gilbert, Y. Li, and M. J. Strauss. What’s the frequency, kenneth?: Sublinear Fourier sampling off the grid. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 61–72, 2012. doi:10.1007/s00453-014-9918-0.
- [3] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014. doi:10.1002/cpa.21455.
- [4] S. Chen and A. Moitra. Algorithmic foundations for the diffraction limit. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, page 490–503, 2021. doi:10.1145/3406325.3451078.
- [5] C. L. Cortes and S. K. Gray. Quantum Krylov subspace algorithms for ground- and excited-state energy estimation. *Phys. Rev. A*, 105:022417, 2022. doi:10.1103/PhysRevA.105.022417.
- [6] Z. Ding and L. Lin. Even shorter quantum circuit for phase estimation on early fault-tolerant quantum computers with applications to ground-state energy estimation. *PRX Quantum*, 4:020331, May 2023. doi:10.1103/PRXQuantum.4.020331.
- [7] Y. Dong, L. Lin, and Y. Tong. Ground-state preparation and energy estimation on early fault-tolerant quantum computers via quantum eigenvalue transformation of unitary matrices. *PRX Quantum*, 3:040305, 2022. doi:10.1103/PRXQuantum.3.040305.
- [8] M. F. Duarte and R. G. Baraniuk. Spectral compressive sensing. *Applied and Computational Harmonic Analysis*, 35(1):111–129, 2013. doi:10.1016/j.acha.2012.08.003.

- [9] A. Dutkiewicz, B. M. Terhal, and T. E. O’Brien. Heisenberg-limited quantum phase estimation of multiple eigenvalues with few control qubits. *Quantum*, 6:830, 2022. doi:10.22331/q-2022-10-06-830.
- [10] E. N. Epperly, L. Lin, and Y. Nakatsukasa. A theory of quantum subspace diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 43(3):1263–1290, 2022. doi:10.1137/21M145954X.
- [11] V. Giovannetti, S. Lloyd, and L. Maccone. Advances in quantum metrology. *Nature Photonics*, 5(4):222–229, 2011. doi:10.1038/nphoton.2011.35.
- [12] R. B. Griffiths and C.-S. Niu. Semiclassical Fourier transform for quantum computation. *Phys. Rev. Lett.*, 76:3228–3231, 1996. doi:10.1103/PhysRevLett.76.3228.
- [13] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde. Entanglement-free Heisenberg-limited phase estimation. *Nature*, 450(7168):393–396, 2007. doi:10.1038/nature06257.
- [14] W. Huggins, J. Lee, U. Baek, B. O’Gorman, and K. Whaley. A non-orthogonal variational quantum eigensolver. *New Journal of Physics*, 22, 2020. doi:10.1088/1367-2630/ab867b.
- [15] Y. Jin, D. Liu, and Z. Song. Super-resolution and robust sparse continuous Fourier transform in any constant dimension: nearly linear time and sample complexity. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4667–4767, 2023. doi:10.1137/1.9781611977554.ch176.
- [16] M. Kapralov. Sparse Fourier transform in any constant dimension with nearly-optimal sample complexity in sublinear time. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, page 264–277, 2016. doi:10.1145/2897518.2897650.
- [17] A. Y. Kitaev, A. Shen, and M. N. Vyalyi. *Classical and quantum computation*. American Mathematical Soc., 2002. URL: <https://dl.acm.org/doi/10.5555/863284>.
- [18] K. Klymko, C. Mejuto-Zaera, S. J. Cotton, F. Wudarski, M. Urbanek, D. Hait, M. Head-Gordon, K. B. Whaley, J. Moussa, N. Wiebe, W. A. de Jong, and N. M. Tubman. Real-time evolution for ultracompact Hamiltonian eigenstates on quantum hardware. *PRX Quantum*, 3:020323, 2022. doi:10.1103/PRXQuantum.3.020323.
- [19] E. Knill, G. Ortiz, and R. D. Somma. Optimal quantum measurements of expectation values of observables. *Phys. Rev. A*, 75:012328, 2007. doi:10.1103/PhysRevA.75.012328.
- [20] H. Li, H. Ni, and L. Ying. A note on spike localization for line spectrum estimation. *preprint*, 2023. doi:10.48550/arXiv.2303.00946.
- [21] H. Li, H. Ni, and L. Ying. On low-depth quantum algorithms for robust multiple-phase estimation. *preprint*, 2023. doi:10.48550/arXiv.2303.08099.
- [22] W. Li, W. Liao, and A. Fannjiang. Super-resolution limit of the esprit algorithm. *IEEE Transactions on Information Theory*, 66(7):4593–4608, 2020. doi:10.1109/TIT.2020.2974174.
- [23] L. Lin and Y. Tong. Heisenberg-limited ground state energy estimation for early fault-tolerant quantum computers. *PRX Quantum*, 3:010318, 2022. doi:10.1103/PRXQuantum.3.010318.

- [24] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A*, 95:042308, 2017. doi:10.1103/PhysRevA.95.042308.
- [25] M. Motta, C. Sun, A. Tan, M. O’Rourke, E. Ye, A. Minnich, F. Brandão, and G. Chan. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nature Physics*, 16:1–6, 2020. doi:10.1038/s41567-019-0704-4.
- [26] D. Nagaj, P. Wocjan, and Y. Zhang. Fast amplification of QMA. *Quantum Inf. Comput.*, 9(11), 2009. doi:10.5555/2012098.2012106.
- [27] H. Ni, H. Li, and L. Ying. On low-depth algorithms for quantum phase estimation. *preprint*, 2023. doi:10.48550/arXiv.2302.02454.
- [28] M. A. Nielsen and I. Chuang. *Quantum computation and quantum information*. Cambridge Univ. Pr., 2000. doi:10.5555/1972505.
- [29] T. E. O’Brien, B. Tarasinski, and B. M. Terhal. Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments. *New J. Phys.*, 21(2):023022, 2019. doi:10.1088/1367-2630/aafb8e.
- [30] R. M. Parrish and P. L. McMahon. Quantum filter diagonalization: Quantum eigendecomposition without full quantum phase estimation. *preprint*, 2019. doi:10.48550/arXiv.1909.08925.
- [31] D. Poulin and P. Wocjan. Sampling from the thermal quantum Gibbs state and evaluating partition functions with a quantum computer. *Phys. Rev. Lett.*, 103:220502, Nov 2009. doi:10.1103/PhysRevLett.103.220502.
- [32] E. Price and Z. Song. A robust sparse Fourier transform in the continuous setting. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 583–600, 10 2015. doi:10.1109/FOCS.2015.42.
- [33] K. Seki and S. Yunoki. Quantum power method by a superposition of time-evolved states. *PRX Quantum*, 2:010333, 2021. doi:10.1103/PRXQuantum.2.010333.
- [34] R. D. Somma. Quantum eigenvalue estimation via time series analysis. *New J. Phys.*, 21(12):123025, 2019. doi:10.1088/1367-2630/ab5c60.
- [35] N. H. Stair, R. Huang, and F. A. Evangelista. A multireference quantum Krylov algorithm for strongly correlated electrons. *Journal of Chemical Theory and Computation*, 16(4):2236–2245, 2020. doi:10.1021/acs.jctc.9b01125.
- [36] M. E. Stroeks, J. Helsen, and B. M. Terhal. Spectral estimation for hamiltonians: a comparison between classical imaginary-time evolution and quantum real-time evolution. *New Journal of Physics*, 24(10):103024, 2022. doi:10.1088/1367-2630/ac919c.
- [37] Y. Subaşı, R. D. Somma, and D. Orsucci. Quantum algorithms for systems of linear equations inspired by adiabatic quantum computing. *Phys. Rev. Lett.*, 122:060504, 2019. doi:10.1103/PhysRevLett.122.060504.
- [38] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490, 2013. doi:10.1109/TIT.2013.2277451.
- [39] G. Wang, D. Stilck-Franca, R. Zhang, S. Zhu, and P. D. Johnson. Quantum algorithm for ground state energy estimation using circuit depth with exponentially improved dependence on precision. *preprint*, 2022. doi:10.48550/arXiv.2209.06811.

- [40] Z. Yang and L. Xie. Achieving high resolution for super-resolution via reweighted atomic norm minimization. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3646–3650, 2015. doi:10.1109/ICASSP.2015.7178651.
- [41] M. Zwierz, C. A. Pérez-Delgado, and P. Kok. General optimality of the Heisenberg limit for quantum metrology. *Phys. Rev. Lett.*, 105:180402, 2010. doi:10.1103/PhysRevLett.105.180402.
- [42] M. Zwierz, C. A. Pérez-Delgado, and P. Kok. Ultimate limits to quantum metrology and the meaning of the Heisenberg limit. *Phys. Rev. A*, 85:042112, 2012. doi:10.1103/PhysRevA.85.042112.

The appendix is organized as follows:

Appendix A. We give two extra numerical tests to demonstrate the robustness of MM-QCELS in relation to parameter selection.

Appendix B. We give an intuitive proof for Theorem 1 in this section.

Appendix C. We give a rigorous proof for Theorem 1 in this section.

Appendix D. We give the proof of Proposition 5, which is a key in the proof of Theorem 1.

Appendix E. We provide some other technical estimations used for our proof.

A Additional numerical experiments

In this section, we numerically demonstrate the robustness of Algorithm 2 using the one-dimensional transverse field Ising model (TFIM) model with $L = 8$, $g = 4$, as defined in Section 4.1. We also normalize the spectrum of the original Hamiltonian H using (24).

In our first test, we set $p_1 = 0.7$, $p_2 = 0.2$, and $p_k = 1/2540$ for $3 \leq k \leq 256$ (so that $\sum_k p_k = 1$). Therefore according to Theorem 1, we should choose $K = 1$ or $K = 2$. We apply Algorithm 2 (with $K = 2, 3, 4$ ¹, $T_0 = 10(\lambda_2 - \lambda_1)^{-1}$, $N_0 = 3 \times 10^3$, $N_{j \geq 1} = 2 \times 10^3$ and $\gamma = 1$) to estimate (λ_1, λ_2) and compute the maximal error (25). We also run QPE 10 times *only* to estimate λ_1 . The comparison of the results is shown in Fig. 7. Surprisingly, although $R^{(K)}/p_{\min}^{(K)} \gg 1$ when we choose $K = 3, 4$, meaning that it does not satisfy the condition of Theorem 1, Algorithm 2 still estimates (λ_1, λ_2) accurately with short circuit depth (T_{\max}). Moreover, the total evolution time (T_{total}) of Algorithm 2 is similar to that of QPE. One intuitive explanation for this phenomenon can be derived from the proof of Theorem 1 in Appendix B. When $K > 2$, aiming to reach the global minimum leads us to anticipate the existence of a unique pair $(k_1, k_2) \in \{1, 2, \dots, K\}^{\otimes 2}$ with the property that $\theta_{k_1}^* \approx \lambda_1$ and $\theta_{k_2}^* \approx \lambda_2$. Then, analogous to the derivation of (37), we can show $|\theta_{k_1}^* - \lambda_1| = \mathcal{O}\left(\frac{R^{(K)}}{p_1 T}\right)$ and $T|\theta_{k_2}^* - \lambda_2| = \mathcal{O}\left(\frac{R^{(K)}}{p_2 T}\right)$. These equalities imply that, even if a wrong choice of K is chosen, as long as $R^{(K)}/p_1 \ll 1$ and $R^{(K)}/p_2 \ll 1$ hold true, an accurate estimation of the dominant eigenvalues can still be obtained with relatively short circuit depth.

The second test in this section focuses on studying the effect of small $p_{\min}^{(K)}$ (or $R^{(K)}/p_{\min}^{(K)} \approx 1$). We construct the initial state with $p_1 = 0.21$, $p_2 = 0.6$ and set $K = 2$. In this setting, we have $R^{(K)}/p_{\min}^{(K)} = 0.19/0.21 \approx 1$. We apply Algorithm 2 (with $K = 2, T_0 = 10(\lambda_2 - \lambda_1)^{-1}$, $N_0 = 3 \times 10^3$, $N_{j \geq 1} = 2 \times 10^3$ and $\gamma = 1$) to estimate (λ_1, λ_2) and compute the maximal error (25). We also run QPE 10 times to estimate λ_1 *only*. The results is summarized in Fig. 8. Interestingly, although Theorem 1 requires $R^{(K)}/p_{\min}^{(K)} \ll 1$ to achieve short circuit depth, MM-QCELS still demonstrates superior performance in circuit depth compared to QPE in this case. Intuitively, this phenomenon finds its explanation through reasoning analogous to the first point raised in Remark 3. To be more specific, for any given $\delta > 0$ and a small enough ϵ , when $T \geq \delta/\epsilon \gg \Delta_\lambda^{-1}$, where Δ_λ is the spectral gap between the dominant eigenvalues and the non-dominant ones, we can establish that

¹The output of the algorithm comprises K pairs of approximated eigenvalues and corresponding weights denoted as $(r_i, \theta_i^*)_{i=1}^K$. From this set, we identify the two approximated eigenvalues with the greatest absolute weights, considering them to be the approximated dominant eigenvalues.

$|\theta_k^* - \lambda_{m_k}| \leq \epsilon$. Here, θ_k^* is the solution to the ideal loss function (4). Furthermore, we can select the number of data points as $N = \tilde{\Theta}(\delta^{-2-o(1)})$ to effectively control the random measurement error. This approach ultimately enables us to attain Heisenberg-limited scaling and a short circuit depth.

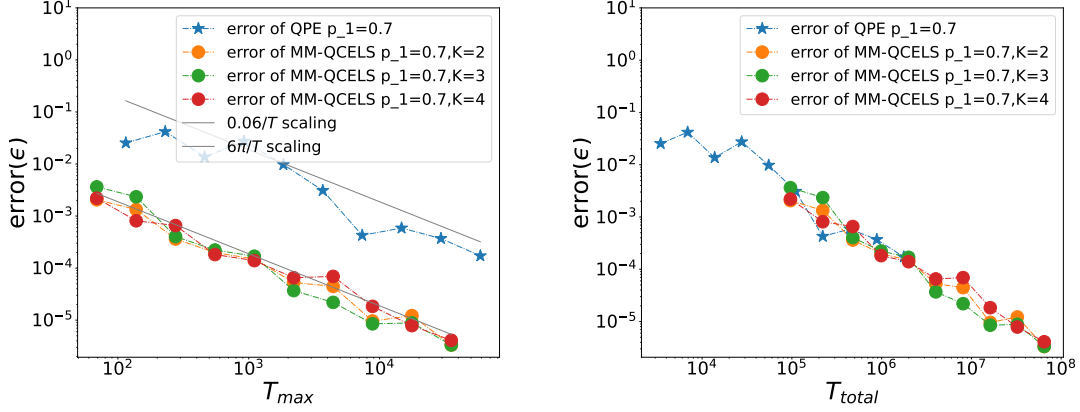


Figure 7: QPE vs Algorithm 2 in TFIM model with 8 sites. $p_1 = 0, 7, p_2 = 0.2, p_k = 1/2540$ for $k \geq 3$. Left: Depth (T_{\max}); Right: Cost (T_{total}). For Algorithm 2, we choose $K = 2, 3, 4, T_0 = 10(\lambda_2 - \lambda_1)^{-1}, N_0 = 3 \times 10^3, N_{j \geq 1} = 2 \times 10^3, \gamma = 1$. l, T_j are chosen according to Theorem 1. Both methods have the error scales linearly in $1/T_{\max}$. The constant factor $\delta = T\epsilon$ of Algorithm 2 is much smaller than that of QPE.

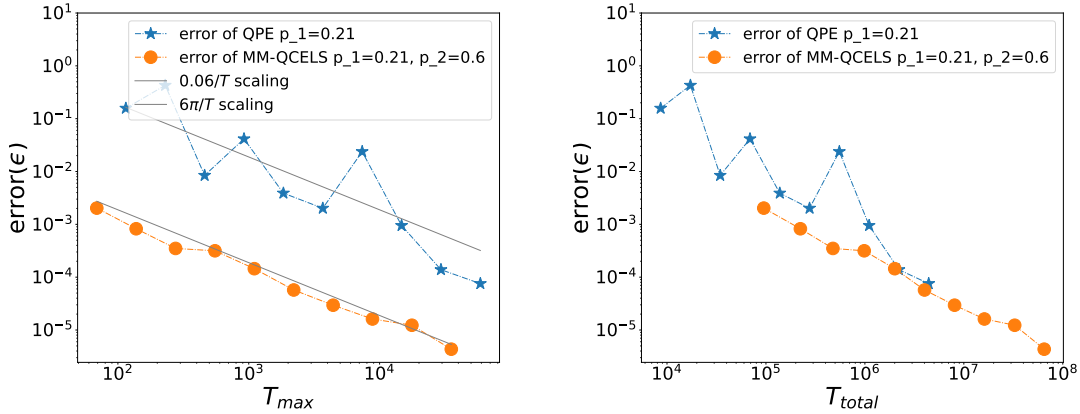


Figure 8: QPE vs Algorithm 2 in TFIM model with 8 sites. $p_1 = 0, 21, p_2 = 0.6$. Left: Depth (T_{\max}); Right: Cost (T_{total}). For Algorithm 2, we choose $K = 2, T_0 = 10(\lambda_2 - \lambda_1)^{-1}, N_0 = 3 \times 10^3, N_{j \geq 1} = 2 \times 10^3, \gamma = 1$. l, T_j are chosen according to Theorem 1. Both methods have the error scales linearly in $1/T_{\max}$. The constant factor $\delta = T\epsilon$ of Algorithm 2 is much smaller than that of QPE.

B An intuitive proof of Theorem 1

In this section, we first give an informal derivation to show that by solving the optimization problem, it is possible to find an accurate approximation to dominant eigenvalues with a

short maximal running time. For simplicity, in the intuitive proof, we only consider the case when

1. The minimal dominant spectral gap $\Delta_\lambda^{(K)}$ is much larger than the precision ϵ : $\Delta_\lambda^{(K)} \gg \epsilon$;
2. The modes in \mathcal{D} are dominant: $p_{\min}^{(K)} > CR^{(K)}$ for some constant $C > 0$;
3. The maximal runtime is sufficient for resolving the dominant eigenvalues: $T_{\max} \geq C'/\Delta_\lambda^{(K)}$ for some constant $C' > 0$.

Here, $p_{\min}^{(K)}$ is defined in (17), $\Delta_\lambda^{(K)}$ is defined in (18), and $R^{(K)}$ is defined in (19).

For now we only focus on the ideal loss function, which can be rewritten as

$$\mathcal{L}_K(\mathbf{r}, \boldsymbol{\theta}) = \mathbf{r}^\dagger U(\boldsymbol{\theta}) \mathbf{r} - \left(V^\dagger(\boldsymbol{\theta}) \mathbf{r} + \mathbf{r}^\dagger V(\boldsymbol{\theta}) \right) + W. \quad (27)$$

Here $U(\boldsymbol{\theta}) \in \mathbb{C}^{K \times K}$, $V(\boldsymbol{\theta}) \in \mathbb{C}^{(K)}$, and $W \in \mathbb{R}$ are defined as

$$U_{k_1, k_2} = F(\theta_{k_1} - \theta_{k_2}), \quad V_k = \sum_{m=1}^M p_m F(\theta_k - \lambda_m), \quad W = \int_{-\infty}^{\infty} a(t) \left| \sum_{m=1}^M p_m \exp(-i\lambda_m t) \right|^2 dt$$

for $1 \leq k, k_1, k_2 \leq K$. For simplicity, in this intuitive analysis we also neglect the difference between the truncated Gaussian distribution and the Gaussian distribution, i.e.,

$$F(x) = \int_{-\infty}^{\infty} a(t) \exp(ixt) dt \approx \exp(-T^2 x^2 / 2). \quad (28)$$

Denote $\mathcal{D} = \{m_1, m_2, \dots, m_K\}$, where $m_1 < m_2 < \dots < m_K$. Without loss of generality, assume the minimizer satisfies $\theta_1^* \leq \theta_2^* \leq \dots \leq \theta_K^*$. We first claim (without giving the proof here) that when $T = \Omega(1/\Delta_\lambda^{(K)})$,

$$|\lambda_{m_k} - \theta_k^*| \leq \frac{\Delta_\lambda^{(K)}}{4}, \quad \forall 1 \leq k \leq K. \quad (29)$$

In other words, each θ_k^* approximates a unique dominant eigenvalue up to a constant proportional to the minimal dominant spectral gap. The next step is to refine the eigenvalue estimate to the target precision ϵ .

When $T = \Omega(1/\Delta_\lambda^{(K)})$, the matrix U is approximately the identity matrix, and $V_k \approx p_{m_k} F(\theta_k - \lambda_{m_k})$. This gives

$$\mathcal{L}_K(\mathbf{r}, \boldsymbol{\theta}) \approx \sum_{k=1}^K \left(|r_k|^2 - 2\text{Re}(p_{m_k} r_k) F(\theta_k - \lambda_{m_k}) - 2 \sum_{m' \in \mathcal{D}^c} \text{Re}(p_{m'} r_k) F(\theta_k - \lambda_{m'}) \right) + W. \quad (30)$$

Hence conceptually, we can solve for each pair (r_k, θ_k) , $1 \leq k \leq K$ independently as

$$(r_k^*, \theta_k^*) = \arg \min_{r, \theta} |r|^2 - 2\text{Re}(p_{m_k} r) F(\theta - \lambda_{m_k}) - 2 \sum_{m' \in \mathcal{D}^c} \text{Re}(p_{m'} r) F(\theta - \lambda_{m'}). \quad (31)$$

Consider the minimization problem on the right-hand side with fixed k . Noticing that this new loss function is quadratic in r , we obtain that

$$\theta_k^* = \arg \max_{\theta} p_{m_k} F(\theta - \lambda_{m_k}) + \sum_{m' \in \mathcal{D}^c} p_{m'} F(\theta - \lambda_{m'}) \quad (32)$$

Plugging $\theta = \lambda_{m_k}$ in the right-hand side, we obtain

$$p_{m_k} F(\theta_k^* - \lambda_{m_k}) + \sum_{m' \in \mathcal{D}^c} p_{m'} F(\theta_k^* - \lambda_{m'}) \geq p_{m_k} + \sum_{m' \in \mathcal{D}^c} p_{m'} F(\lambda_{m_k} - \lambda_{m'}). \quad (33)$$

Using the Gaussian approximation $F(x) = \exp(-T^2 x^2/2)$, we have

$$|F'(x)| = T^2 |x| \exp(-T^2 x^2/2) \leq \left(\sup_{z \in \mathbb{R}} z \exp(-z^2/2) \right) T = \Theta(T),$$

where we view $T|x|$ as z in the inequality. This implies $F(x)$ is a $\mathcal{O}(T)$ -Lipschitz function. Combining this with (33), we obtain

$$\begin{aligned} & \exp\left(-\frac{T^2(\theta_k^* - \lambda_{m_k})^2}{2}\right) - 1 \\ & \geq \sum_{m' \in \mathcal{D}^c} \frac{p_{m'}}{p_{m_k}} (F(\lambda_{m_k} - \lambda_{m'}) - F(\theta_k^* - \lambda_{m'})) \\ & \geq -\frac{R^{(K)}}{p_{m_k}} \min\{\mathcal{O}(|T(\theta_k^* - \lambda_{m_k})|), 1\}. \end{aligned} \quad (34)$$

where we use $0 \leq F \leq 1$ and F is a $\mathcal{O}(T)$ -Lipschitz function in the last inequality. From (34), we first have $\exp\left(-\frac{T^2(\theta_k^* - \lambda_{m_k})^2}{2}\right) \geq 1 - \frac{R^{(K)}}{p_{m_k}}$, which implies

$$T|\theta_k^* - \lambda_{m_k}| = \mathcal{O}\left(\sqrt{\frac{R^{(K)}}{p_{m_k}}}\right). \quad (35)$$

When $R^{(K)}$ is sufficiently small, combining Eqs. (34) and (35), we further obtain

$$1 - \frac{T^2(\theta_k^* - \lambda_{m_k})}{4} \geq \exp\left(-\frac{T^2(\theta_k^* - \lambda_{m_k})^2}{2}\right) \geq 1 - \frac{R^{(K)}}{p_{m_k}} \mathcal{O}(|T(\theta_k^* - \lambda_{m_k})|). \quad (36)$$

This implies

$$T|\theta_k^* - \lambda_{m_k}| = \mathcal{O}\left(\frac{R^{(K)}}{p_{m_k}}\right). \quad (37)$$

In summary, to obtain $|\theta_k^* - \lambda_{m_k}| \leq \epsilon$ for all $1 \leq k \leq K$, we can set

$$T = \Theta\left(\frac{R^{(K)}}{p_{\min}^{(K)} \epsilon}\right). \quad (38)$$

This implies the depth constant of maximal running time $T_{\max} = \Theta\left(\frac{R^{(K)}}{p_{\min}^{(K)} \epsilon}\right)$ is much smaller than $\frac{\pi}{\epsilon}$ when $R^{(K)}/p_{\min}^{(K)}$ is close to 0.

We remark that when $K = 1$, there is only one dominant mode m_k , and $p_{m_k} = 1 - R^{(K)}$ by definition. In this case, the result in Eq. (35) is comparable to the estimate in Ref. [6] for the QCELS method. The analysis in this work provides a tighter bound of the maximal runtime (or the circuit depth). Specifically, Eq. (37) provides a quadratic improvement with respect to the preconstant $R^{(K)}$ for estimating a single dominant eigenvalue, and the same conclusion holds for estimating multiple eigenvalues.

Remark 3. 1. When a spectral gap exists between the dominant eigenvalues and the non-dominant ones (represented as Δ_λ), we can further reduce the maximum runtime to $T_{\max} = \tilde{\Theta}(1/\min\{\Delta_\lambda, \Delta_\lambda^{(K)}\})$. This reduction can be derived in a similar manner to the previous intuitive analysis. Specifically, referring to equation (34), when T is sufficiently large to ensure $|\theta^{(K)} - \lambda_{m_k}| \leq \frac{\Delta_\lambda}{2}$, we have

$$\left| F(\lambda_{m_k} - \lambda_{m'}) - F(\theta^{(K)} - \lambda_{m'}) \right| = \mathcal{O}(\exp(-T^2 \Delta_\lambda^2 / 8)).$$

Following a similar derivation to (35)-(37), we obtain the following expression:

$$T|\theta_k^* - \lambda_{m_k}| = \mathcal{O}\left(\frac{\exp(-T^2 \Delta_\lambda^2 / 8)}{p_{m_k}}\right),$$

This implies that $T_{\max} = \tilde{\Theta}(1/\min\{\Delta_\lambda, \Delta_\lambda^{(K)}\})$ is sufficient to ensure ϵ -accuracy. However, in this scenario, while T_{\max} logarithmically depends on the desired precision ϵ , the number of data points needs to increase to $N = \Omega(\epsilon^{-2})$ to adequately control the random noise. This is similar in flavor to the result in [39] as well as in QCELS [6] for estimating a single dominant eigenvalue.

2. The rigorous proof of Theorem 1 follows a slightly different path from the previous intuitive derivation. The numerical loss function (13) admits a similar quadratic expansion as in Eq. (27) with noisy coefficients $U(\theta)$, $V(\theta)$, and W . Due to the presence of noise and off-diagonal entries in U , the perfect separation assumed in (31) no longer holds, and the analysis of the independent optimization problem cannot be directly applied to show that θ_k^* is close to λ_{m_k} . To overcome this difficulty, we adopt the idea of separation and decompose the numerical loss function after bounding the noise. By comparing the resulting loss function with $L_K(\{p_m\}_{m \in \mathcal{D}}, \{\lambda_m\}_{m \in \mathcal{D}})$, we demonstrate that $\exp\left(-\frac{T^2(\theta_k^* - \lambda_{m_k})^2}{2}\right) - 1 = \mathcal{O}(q)$ and $|\theta_k^* - \lambda_{m_k}| = \mathcal{O}(q/T)$, where $q = \Theta(R^{(K)}/p_{\min}^{(K)})$. This implies, to obtain ϵ -accuracy, the maximal running time $T_{\max} = \gamma T = \delta/\epsilon$, where $\delta = \tilde{\Theta}(q \log(q^{-1}))$.

C Rigorous proof of Theorem 1

To prove Theorem 1, we first rewrite the optimization problem (14) as

$$\begin{aligned} & \arg \min_{\mathbf{r}, \boldsymbol{\theta}} L_K(\mathbf{r}, \boldsymbol{\theta}) \\ &= \arg \min_{\mathbf{r}, \boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^N \left| \sum_{m=1}^M p_m \exp(-i\lambda_m t_n) - \sum_{k=1}^K r_k \exp(-i\theta_k t_n) \right|^2 \\ & \quad - \frac{2}{N} \sum_{n=1}^N \operatorname{Re} \left(\left\langle E_n, \sum_{k=1}^K r_k \exp(-i\theta_k t_n) \right\rangle \right) \\ &= \arg \min_{\mathbf{r}, \boldsymbol{\theta}} \int_{-\infty}^{\infty} a(t) \left| \sum_{m=1}^M p_m \exp(-i\lambda_m t) - \sum_{k=1}^K r_k \exp(-i\theta_k t) \right|^2 dt + E_{p,r} + E_{r,r} + E_{r,z} \end{aligned} \tag{39}$$

where we omit term $\frac{1}{N} \sum_{n=1}^N |E_n|^2$ in the first equality and $p_m p_{m'} \exp(i(\lambda_m - \lambda_{m'})t)$ terms in the second equality because they do not affect the solution of the optimization problem.

Here

$$\begin{aligned}
E_{p,r}(\mathbf{r}, \boldsymbol{\theta}) &= 2\text{Re} \left(\sum_{m=1}^M \sum_{k=1}^K p_m r_k \left(\frac{1}{N} \sum_{n=1}^N \exp(i(\theta_k - \lambda_m)t_n) - F(\theta_k - \lambda_m) \right) \right), \\
E_{r,r}(\mathbf{r}, \boldsymbol{\theta}) &= 2\text{Re} \left(\sum_{k_1 \neq k_2}^K \bar{r}_{k_1} r_{k_2} \left(\frac{1}{N} \sum_{n=1}^N \exp(i(\theta_{k_2} - \theta_{k_1})t_n) - F(\theta_{k_2} - \theta_{k_1}) \right) \right), \\
E_{r,Z}(\mathbf{r}, \boldsymbol{\theta}) &= -2\text{Re} \left(\sum_{k=1}^K \bar{r}_k \left(\frac{1}{N} \sum_{n=1}^N E_n \exp(i\theta_k t_n) \right) \right).
\end{aligned} \tag{40}$$

Roughly speaking, when $N \gg 1$, we have the expectation error $|E_{p,r}|, |E_{r,r}|, |E_{r,Z}| = \mathcal{O}(1/\sqrt{N})$ (see Appendix E for detail).

The proof contains two steps: 1. Find initial estimation intervals for the dominant eigenvalues (using T_0, N_0); 2. In the correct estimation intervals, find more accurate approximations of the dominant eigenvalues (from T_j to T_{j+1}). Now, we introduce a lemma and a proposition to control the complexity of these two steps respectively.

Let $E(\mathbf{r}, \boldsymbol{\theta}) = E_{p,r}(\mathbf{r}, \boldsymbol{\theta}) + E_{r,r}(\mathbf{r}, \boldsymbol{\theta}) + E_{r,Z}(\mathbf{r}, \boldsymbol{\theta})$. We use the following lemma to control the complexity of the first step:

Lemma 4. *Given $0 < q < 1$ such that $q = \Omega(R^{(K)}/p_{\min}^{(K)})$, where $p_{\min}^{(K)}$ and $R^{(K)}$ are defined in (17) and (19), we assume $p_{\min}^{(K)} > 3R^{(K)}$. Define*

$$(\mathbf{r}^*, \boldsymbol{\theta}^*) = \arg \min_{\|\mathbf{r}\|_1 \leq 1, \theta_k \in [-\pi, \pi]} L_K(\mathbf{r}, \boldsymbol{\theta}),$$

where L_K is defined in (13). If

$$\begin{aligned}
\gamma &= \Theta \left(\log \left(1 / \min \left\{ p_{\min}^{(K)} q, \left(p_{\min}^{(K)} \right) \left(p_{\min}^{(K)} - 3R^{(K)} \right) \right\} \right) \right), \\
T &= \Omega \left(\left(\Delta_{\lambda}^{(K)} \right)^{-1} \log \left(\left(p_{\min}^{(K)} \right)^{-1} \max \left\{ \left(p_{\min}^{(K)} - 3R^{(K)} \right)^{-1}, \left(p_{\min}^{(K)} \right)^{-1} q^{-2} \right\} \right) \right), \\
|E| &= \mathcal{O} \left(\min \left\{ \left(p_{\min}^{(K)} \right)^2 q^2, p_{\min}^{(K)} \left(p_{\min}^{(K)} - 3R^{(K)} \right) \right\} \right),
\end{aligned} \tag{41}$$

then, for each $m \in \mathcal{D}$, there must exist a unique $1 \leq k_m \leq K$ such that

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{q}{T}, \quad |p_m - r_{k_m}^*| \leq qp_m. \tag{42}$$

Lemma 4 constitutes a key element in the remaining part of the proof. According to Lemma 4, when the expectation error is small enough, the error of refined dominant eigenvalue estimation is bounded by q/T , where q is a fixed (maybe small) constant and T is the maximal runtime. Combining this lemma with Lemma 8 by setting $q = 1$, we can demonstrate that when $T_0 = \Omega \left(\left(\Delta_{\lambda}^{(K)} \right)^{-1} \right)$ and $N_0 = \Omega(T_0^2)$, solving the optimization problem gives us a reasonable approximation to the refined dominant eigenvalues, meaning: for each $m \in \mathcal{D}$, there must exist a unique $1 \leq k_m \leq K$ such that $|\lambda_m - \theta_{k_m}^*| = \mathcal{O} \left(\Delta_{\lambda}^{(K)} \right)$.

Proof of Lemma 4. First, we define

$$F^*(x) = \exp \left(-\frac{T^2 x^2}{2} \right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}T} \exp \left(-\frac{x^2}{2T^2} \right) \exp(ixt) dt.$$

Notice

$$L_K(\mathbf{r}, \boldsymbol{\theta}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{t^2}{2T^2}\right) \left| \sum_{m=1}^M p_m \exp(-i\lambda_m t) - \sum_{k=1}^K r_k \exp(-i\theta_k t) \right|^2 dt + E_{p,r} + E_{r,r} + E_{r,Z} + E_F,$$

where $E_F = \int_{-\infty}^{\infty} \left(a(t) - \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{t^2}{2T^2}\right) \right) \left| \sum_{m=1}^M p_m \exp(-i\lambda_m t) - \sum_{k=1}^K r_k \exp(-i\theta_k t) \right|^2 dt$. Using the tail bound of a Gaussian and the choice of γ (41), we have

$$|E_F| = \mathcal{O}\left(\exp(-\gamma^2/2)\right) = \mathcal{O}\left(\min\left\{\left(p_{\min}^{(K)} q\right)^2, \left(p_{\min}^{(K)} - 3R^{(K)}\right) p_{\min}^{(K)}\right\}\right). \quad (43)$$

Notice

$$\begin{aligned} & L_K(\{p_m\}_{m \in \mathcal{D}}, \{\lambda_m\}_{m \in \mathcal{D}}) \\ & \leq \int_{-\infty}^{\infty} a(t) \left| \sum_{m \in \mathcal{D}^c} p_m \exp(-i\lambda_m t) \right|^2 dt + \hat{E} + \hat{E}_F \\ & \leq \int_{-\infty}^{\infty} a(t) \left| \sum_{m=1}^M p_m \exp(-i\lambda_m t) - \sum_{m \in \mathcal{D}} p_m \exp(-i\lambda_m t) \right|^2 dt + \hat{E} + \hat{E}_F \\ & \leq (R^{(K)})^2 + \hat{E} + \hat{E}_F, \end{aligned} \quad (44)$$

where we $\hat{E} = E(\{p_m\}_{m \in \mathcal{D}}, \{\lambda_m\}_{m \in \mathcal{D}})$, and $\hat{E}_F = E_F(\{p_m\}_{m \in \mathcal{D}}, \{\lambda_m\}_{m \in \mathcal{D}})$.

We separate the proof into two steps. In the first step, we show that for each $m \in \mathcal{D}$, there must exist a unique $1 \leq k_m \leq K$ such that

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{\Delta_\lambda^{(K)}}{4}. \quad (45)$$

In the second step, we further improve the bound in (45).

Step 1: Show each θ_k^* should be close to one λ_m for $m \in \mathcal{D}$.

Suppose there exists $m^* \in \mathcal{D}$ such that for any $1 \leq k \leq K$,

$$|\lambda_{m^*} - \theta_k^*| > \frac{\Delta_\lambda^{(K)}}{4}.$$

Then, let $E^* = E(\mathbf{r}^*, \boldsymbol{\theta}^*)$ and $E_F^* = E_F(\mathbf{r}^*, \boldsymbol{\theta}^*)$,

$$\begin{aligned} & L_K(\mathbf{r}^*, \boldsymbol{\theta}^*) \\ & = \int_{-\infty}^{\infty} a(t) \left| \left(\sum_{m \in \mathcal{D}, m \neq m^*} p_m \exp(-i\lambda_m t) - \sum_{k=1}^K r_k^* \exp(-i\theta_k^* t) \right) + \sum_{m \in \mathcal{D}^c} p_m \exp(-i\lambda_m t) \right|^2 dt + E^* + E_F^* \\ & \geq p_{m^*}^2 - 2p_{m^*} R^K - 16 \exp\left(-\frac{(T\Delta_\lambda^{(K)})^2}{16}\right) + E^* + E_F^* \\ & > (R^{(K)})^2 + \hat{E} + \hat{E}_F \\ & \geq L_K(\{p_m\}_{m \in \mathcal{D}}, \{\lambda_m\}_{m \in \mathcal{D}}). \end{aligned}$$

In the second inequality, we use $p_{m^*} > 3R^{(K)}$, (41), and (43). In the first inequality, we use the decaying property of $F(x)$ to obtain

$$\left| \int_{-\infty}^{\infty} a(t) \left(\sum_{m \in \mathcal{D}, m \neq m^*} p_m \exp(-i\lambda_m t) - \sum_{k=1}^K r_k^* \exp(-i\theta_k^* t) \right) p_{m^*} \exp(-i\lambda_{m^*} t) dt \right| \leq 8 \exp\left(-\frac{(T\Delta_\lambda^{(K)})^2}{16}\right),$$

$$\left| \int_{-\infty}^{\infty} a(t) \left(\sum_{m \in \mathcal{D}^c} p_m \exp(-i\lambda_m t) \right) p_{m^*} \exp(-i\lambda_{m^*} t) dt \right| \leq p_{m^*} R^{(K)}$$

where we use $|\lambda_m - \lambda_{m^*}| > \Delta_\lambda^{(K)}/2$, $|\theta_k^* - \lambda_{m^*}| > \Delta_\lambda^{(K)}/4$. This contradicts the assumption that $(\{r_k^*\}_{k=1}^K, \{\theta_k^*\}_{k=1}^K)$ is the minimization point. Thus, (45) is true for all $1 \leq k \leq K$.

Step 2: Improve the upper bound.

Define

$$\mathcal{F}^* = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{x^2}{2T^2}\right) \left| \sum_{m \in \mathcal{D}} p_m \exp(-i\lambda_m t) - r_{k_m}^* \exp(-i\theta_{k_m}^* t) \right|^2 dt.$$

We have

$$\begin{aligned} & L_K(\mathbf{r}^*, \boldsymbol{\theta}^*) \\ &= \int_{-\infty}^{\infty} a(t) \left| \left(\sum_{m \in \mathcal{D}} p_m \exp(-i\lambda_m t) - r_{k_m}^* \exp(-i\theta_{k_m}^* t) \right) + \sum_{m \in \mathcal{D}^c} p_m \exp(-i\lambda_m t) \right|^2 + E^* + E_F^* \\ &\geq \mathcal{F}^* - 2(\mathcal{F}^*)^{1/2} R^{(K)} + \int_{-\infty}^{\infty} a(t) \left| \sum_{m \in \mathcal{D}^c} p_m \exp(-i\lambda_m t) \right|^2 dt + E^* + E_F^* \end{aligned}$$

Noticing that $(\mathbf{r}^*, \boldsymbol{\theta}^*)$ is the minimum point and comparing the above inequality with the second inequality of (44), we have

$$\mathcal{F}^* - 2(\mathcal{F}^*)^{1/2} R^{(K)} \leq |\hat{E} - E^*| + |\hat{E}_F - E_F^*| \quad (46)$$

Using (41) and conditions of Theorem 1, we have $|\hat{E}|, |E^*|, |\hat{E}_F|, |E_F^*| = \mathcal{O}\left(\left(p_{\min}^{(K)}\right)^2\right)$.

This implies

$$\mathcal{F}^* - 2(\mathcal{F}^*)^{1/2} R^{(K)} = \mathcal{O}\left(\left(p_{\min}^{(K)}\right)^2\right),$$

Since $q = \Omega(R^{(K)}/p_{\min}^{(K)})$, we further have $\mathcal{F}^* = \mathcal{O}\left(\left(p_{\min}^{(K)}\right)^2\right)$.

Using (45), $\sum_k |r_k^*| \leq 1$, and $\exp\left(-(\Delta_\lambda^{(K)} T)^2/8\right) = \mathcal{O}\left(\left(p_{\min}^{(K)}\right)^2\right)$, we can show

$$\begin{aligned} & \sum_{m, m' \in \mathcal{D}, m \neq m'} \left| \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{x^2}{2T^2}\right) (p_m \exp(-i\lambda_m t) - r_{k_m}^* \exp(-i\theta_{k_m}^* t)) \right. \\ & \quad \left. \cdot (p_{m'} \exp(-i\lambda_{m'} t) - r_{k_{m'}}^* \exp(-i\theta_{k_{m'}}^* t)) dt \right| \\ &\leq \sum_{m, m' \in \mathcal{D}, m \neq m'} p_m p_{m'} \exp\left(-\frac{T^2(\lambda_m - \lambda_{m'})^2}{2}\right) + p_m r_{k_{m'}}^* \exp\left(-\frac{T^2(\lambda_m - \theta_{k_{m'}}^*)^2}{2}\right) \\ & \quad + r_{k_m}^* p_{m'} \exp\left(-\frac{T^2(\theta_{k_m}^* - \lambda_{m'})^2}{2}\right) + r_{k_m}^* r_{k_{m'}}^* \exp\left(-\frac{T^2(\lambda_m - \theta_{k_{m'}}^*)^2}{2}\right) \\ &\leq 4 \exp\left(-\frac{(\Delta_\lambda^{(K)} T)^2}{8}\right) = \mathcal{O}\left(\left(p_{\min}^{(K)}\right)^2\right). \end{aligned} \quad (47)$$

Expanding \mathcal{F}^* , we find that

$$\begin{aligned} \mathcal{F}^* &\geq \sum_{m \in \mathcal{D}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{t^2}{2T^2}\right) |p_m \exp(-i\lambda_m t) - r_{k_m}^* \exp(-i\theta_{k_m}^* t)|^2 dt \\ &\quad - 2 \sum_{m, m' \in \mathcal{D}, m \neq m'} \left| \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{x^2}{2T^2}\right) (p_m \exp(-i\lambda_m t) - r_{k_m}^* \exp(-i\theta_{k_m}^* t)) \right. \\ &\quad \cdot \left. (p_{m'} \exp(-i\lambda_{m'} t) - r_{k_{m'}}^* \exp(-i\theta_{k_{m'}}^* t)) dt \right|. \end{aligned} \quad (48)$$

By utilizing the bound $\mathcal{F}^* = \mathcal{O}\left(\left(p_{\min}^{(K)} q\right)^2\right)$ and (47), we can infer from (48) that

$$\sum_{m \in \mathcal{D}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{t^2}{2T^2}\right) |p_m \exp(-i\lambda_m t) - r_{k_m}^* \exp(-i\theta_{k_m}^* t)|^2 dt = \mathcal{O}\left(\left(p_{\min}^{(K)} q\right)^2\right). \quad (49)$$

This implies

$$\begin{aligned} &\sum_{m \in \mathcal{D}} p_m^2 \left(1 - \exp\left(-T^2 (\theta_{k_m}^* - \lambda_m)^2\right)\right) \\ &\leq \sum_{m \in \mathcal{D}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{t^2}{2T^2}\right) |p_m \exp(i(\theta_{k_m}^* - \lambda_m)t) - r_{k_m}^*|^2 dt \\ &= \sum_{m \in \mathcal{D}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{t^2}{2T^2}\right) |p_m \exp(-i\lambda_m t) - r_{k_m}^* \exp(-i\theta_{k_m}^* t)|^2 dt \\ &= \mathcal{O}\left(\left(p_{\min}^{(K)} q\right)^2\right). \end{aligned}$$

Because $p_m \geq p_{\min}^{(K)}$ for $m \in \mathcal{D}$, we obtain that, for each k , $1 - \exp\left(-T^2 (\lambda_m - \theta_{k_m}^*)^2\right) = \mathcal{O}(q^2)$, which implies the first inequality of (42). For the second inequality of (42), (49) also implies

$$\left| p_m \exp\left(-T^2 (\theta_{k_m}^* - \lambda_m)^2 / 2\right) - r_{k_m}^* \right|^2 = \mathcal{O}\left(\left(p_{\min}^{(K)} q\right)^2\right)$$

Because $\left|\exp\left(-T^2 (\theta_{k_m}^* - \lambda_m)^2 / 2\right) - 1\right| = \mathcal{O}(q^2)$, we have $|p_m - r_{k_m}^*| \leq qp_m$. This concludes the proof of the second inequality of (42). \square

The following proposition controls the complexity of the second step of the algorithm:

Proposition 5. *Given failure probability $0 < \eta < 1/2$, any small constant $\zeta > 0$, $0 < q < 1$ such that $q = \Theta(R^{(K)}/p_{\min}^{(K)})$, and given a sequence of rough intervals $\{I_k\}_{k=1}^K \subset \mathbb{R}$, we assume 1. $|I_k| \leq 4\pi/T$; 2. for any $m \in \mathcal{D}$, there exists a unique $1 \leq k_m \leq K$ such that $\lambda_m \in I_{k_m}$; 3. $p_{\min}^{(K)} > 3R^{(K)}$. Define*

$$(\mathbf{r}^*, \boldsymbol{\theta}^*) = \arg \min_{\|\mathbf{r}\|_1 \leq 1, \theta_k \in I_k} L_K(\mathbf{r}, \boldsymbol{\theta}),$$

where L_K is defined in (13). If

$$\begin{aligned} \gamma &= \Theta\left(\log\left(\left(p_{\min}^{(K)}\right)^{-1} q^{-1}\right)\right), \\ T &= \Omega\left(\left(\Delta_{\lambda}^{(K)}\right)^{-1} \log\left(\left(p_{\min}^{(K)}\right)^{-1} q^{-1}\right)\right), \\ N &= \Theta\left(\gamma^2 \left(p_{\min}^{(K)}\right)^{-4} q^{-(2+\zeta)} \text{polylog}\left(K \log(\zeta^{-1}) \left(p_{\min}^{(K)}\right)^{-1} \eta^{-1}\right)\right), \end{aligned} \quad (50)$$

then with probability $1 - \eta$,

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{q}{T}, \quad |p_m - r_{k_m}^*| \leq qp_m. \quad (51)$$

We give the proof of Proposition 5 in Appendix D. Here, we emphasize that Proposition 5 can not be directly proved by combining Lemma 4 and Lemma 8. According to Lemma 4, to obtain the accuracy q/T , we need the expectation error $|E| = \mathcal{O}(q^2)$. According to Lemma 8, to guarantee $|E| = \mathcal{O}(q^2)$, we need to choose $N = \Omega(q^{-4})$, which is worse than the scaling of N with respect to q^{-1} in the proposition. This means that to achieve the correct scaling of N with respect to q^{-1} , we need to bound the expectation error in a different way to obtain the sharp estimation.

Now, we are ready to use Lemma 4 and Proposition 5 to prove Theorems 1.

Proof of Theorem 1. First, according to the definition of T_0 , we have $T_0 = \tilde{\Theta}((\Delta_K^\lambda)^{-1} \log(q^{-1}))$. Combining (20) with Lemma 4 and Lemma 8 (by setting $q = 1$ in Lemma 4 and $\rho = \pi T_0$ and $\xi = \Theta\left(\left(p_{\min}^{(K)} - 3R^{(K)}\right) p_{\min}^{(K)}\right)$ in Lemma 8 (59)), after step 1, with probability $1 - \eta/(l+1)$, we obtain that, for every $m \in \mathcal{D}$, there exists a unique $1 \leq k_m \leq K$ such that

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{1}{T_0} < \min\left\{\frac{\Delta_K^\lambda}{4}, \frac{\pi}{T_0}\right\}, \quad |p_m - r_{k_m}^*| \leq p_m. \quad (52)$$

Thus, with probability $1 - \eta/(l+1)$, after step 1, $I_{k_m} = [\lambda_{\min, k_m}, \lambda_{\max, k_m}]$ satisfies the condition of Proposition 5 with $T = T_1$.

Next, for $j = 1$ in step 2, using Proposition 5 by setting $T = T_1$, we obtain that, with probability $1 - \eta/(l+1)$,

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{q}{T_1} < \frac{\pi}{T_1}, \quad |p_m - r_{k_m}^*| \leq qp_m. \quad (53)$$

This implies that, after step 2 with $j = 1$, with probability $1 - 2\eta/(l+1)$, $I_{k_m} = [\lambda_{\min, k_m}, \lambda_{\max, k_m}]$ satisfies the condition of Proposition 5 with $T = T_2$. Using this recursively, we finally obtain that, when $j = l$, with probability $1 - \eta$, we have

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{q}{T_l} < \epsilon, \quad |p_m - r_{k_m}^*| \leq qp_m.$$

which implies (23).

Finally, using the choices of γ, l, T_l and N_l , we obtain

$$T_{\max} = T_l \gamma = \Theta\left(\frac{q}{\epsilon} \log\left(\frac{1}{\min\{p_{\min}^{(K)} q, (p_{\min}^{(K)}) (p_{\min}^{(K)} - 3R^{(K)})\}}\right)\right) = \frac{\delta}{\epsilon}$$

and

$$T_{\text{total}} = \tilde{\Theta}\left(\frac{1}{(p_{\min}^{(K)})^4 q^{1+\zeta\epsilon}} \text{poly log}(\log(\zeta^{-1})\eta^{-1})\right) = \tilde{\Theta}\left(\frac{\text{poly log}(\log(\zeta^{-1})\delta^{-1}\eta^{-1})}{(p_{\min}^{(K)})^4 \delta^{1+\zeta\epsilon}}\right),$$

where $\delta = \tilde{\Theta}(q \log(q^{-1}))$. □

D Proof of Proposition 5

In this section, we prove Proposition 5. The proof idea is similar to [6, Appendix B.2]. We first give a rough complexity result that has a $\mathcal{O}(q^{-4})$ scaling in N . Then, we consider the difference of the expectation error more carefully and improve this scaling to $\mathcal{O}(q^{-2-o(1)})$ using an iteration argument.

Applying Eq. (59) from Lemma 8 in Appendix E to bound the expectation error $E(\mathbf{r}, \boldsymbol{\theta})$, we obtain the following lemma:

Lemma 6. *Given failure probability $0 < \eta < 1/2$, any small constant $\zeta > 0$, $0 < q < 1$ such that $q = \Omega(R^{(K)}/p_{\min}^{(K)})$, and given a sequence of rough interval $\{I_k\}_{k=1}^K \subset \mathbb{R}$, we assume 1. $|I_k| \leq 4\pi/T$; 2. for any $m \in \mathcal{D}$, there exists a unique $1 \leq k_m \leq K$ such that $\lambda_m \in I_{k_m}$; 3. $p_{\min}^{(K)} > 3R^{(K)}$. Define*

$$(\mathbf{r}^*, \boldsymbol{\theta}^*) = \arg \min_{\|\mathbf{r}\| \leq 1, \theta_k \in I_k} L_K(\mathbf{r}, \boldsymbol{\theta}),$$

where L_K is defined in (13). If

$$\begin{aligned} \gamma &= \Omega \left(\log \left(\left(p_{\min}^{(K)} \right)^{-1} q^{-1} \right) \right), \\ T &= \Omega \left(\left(\Delta_\lambda^{(K)} \right)^{-1} \log \left(\left(p_{\min}^{(K)} \right)^{-1} q^{-1} \right) \right), \\ N &= \Omega \left(\gamma^2 \left(p_{\min}^{(K)} \right)^{-4} q^{-4} \text{polylog} \left(K \left(p_{\min}^{(K)} \right)^{-1} q^{-1} \eta^{-1} \right) \right), \end{aligned} \quad (54)$$

then with probability $1 - \eta$,

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{q}{T}, \quad |p_m - r_{k_m}^*| \leq qp_m. \quad (55)$$

Proof of Lemma 6. According to the second equality of (54), we have

$$|I_k| \leq \frac{4\pi}{T} \leq \frac{\Delta_\lambda^{(K)}}{4},$$

which implies that for all $m \in \mathcal{D}$

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{\Delta_\lambda^{(K)}}{4}. \quad (56)$$

Using (54) and Lemma 8 (59) by setting $\xi = \mathcal{O} \left(\left(p_{\min}^{(K)} q \right)^2 \right)$, we can conclude that with probability $1 - \eta$, $|E| = \mathcal{O} \left(\left(p_{\min}^{(K)} q \right)^2 \right)$. Finally, since we have obtained a rough bound (56), we can use the same argument as the second step in the proof of Lemma 4 to prove (55). \square

Next, in order to improve the scaling $\mathcal{O}(q^{-4})$ in (54). We propose a different approach to bound the error terms. Instead of bounding $E(\mathbf{r}^*, \boldsymbol{\theta}^*)$ and $E(\{p_m\}, \{\lambda_m\})$ separately, we aim to bound the difference between these two error terms. Intuitively, when $(r_{k_m}^*, \theta_{k_m}^*)$ and (p_m, λ_m) are close to each other, the two error terms are likely to cancel each other out when we compare the difference between $L(\mathbf{r}^*, \boldsymbol{\theta}^*)$ and $L(\{p_m\}, \{\lambda_m\})$. This intuition is supported by Lemma 8 (60). Assuming that we already have $|\theta_{k_m}^* - \lambda_m| < \frac{q}{T}$

and $|r_{k_m} - p_m| \leq qp_m$, then it is sufficient to choose $N = \tilde{\Omega}(q^2\xi^{-2})$ to ensure that $|E(\mathbf{r}^*, \boldsymbol{\theta}^*) - E(\{p_m\}, \{\lambda_m\})| \geq \xi$ with high probability. This requirement, compared with the second inequality of (54), reduces the blow-up rate to $\mathcal{O}(q^{-2})$ as $q \rightarrow 0$, which matches the condition in Proposition 5. However, the above calculation is assuming $|\theta_{k_m}^* - \lambda_m| < \frac{q}{T}$ and $|r_{k_m} - p_m| \leq qp_m$, which is unknown to us in advance. To overcome this difficulty, we need to use an iteration argument to obtain the desired order. This is summarized in the following lemma:

Lemma 7. *Given failure probability $0 < \eta < 1/2$, an integer $S > 1$, any small constant $\zeta > 0$, a sequence of rough interval $\{I_k\}_{k=1}^K \subset \mathbb{R}$, and a decreasing sequence $\{q_s\}_{s=0}^S$ with $0 < q_0 \leq 1$ and $q_S = \Omega(R^{(K)}/p_{\min}^{(K)})$, we assume 1. $|I_k| \leq 4\pi/T$; 2. for any $m \in \mathcal{D}$, there exists a unique $1 \leq k_m \leq K$ such that $\lambda_m \in I_{k_m}$; 3. $p_{\min}^{(K)} > 3R^{(K)}$; 4. (54) holds with $q = q_0$. If*

$$\begin{aligned} \gamma &= \Omega\left(\log\left(\left(p_{\min}^{(K)}\right)^{-1} q_S^{-1}\right)\right), \\ T &= \Omega\left(\left(\Delta_\lambda^{(K)}\right)^{-1} \log\left(\left(p_{\min}^{(K)}\right)^{-1} q_S^{-1}\right)\right), \\ N &= \Omega\left(\max_{0 \leq s \leq S-1} \left\{\left(p_{\min}^{(K)}\right)^{-4} q_{s+1}^{-4} q_s^2 \text{polylog}\left(KQ\left(p_{\min}^{(K)}\right)^{-1} q_{s+1}^{-1} q_s \eta^{-1}\right)\right\}\right), \end{aligned} \quad (57)$$

then with probability $1 - \eta$,

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{q_S}{T}, \quad |p_m - r_{k_m}^*| \leq q_S p_m. \quad (58)$$

Proof of Lemma 7. By utilizing Lemma 6 in conjunction with (54), we can demonstrate that with probability $1 - \eta/(2Q)$,

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{q_0}{T}, \quad |p_m - r_{k_m}^*| \leq q_0 p_m.$$

Using these inequalities with Lemma 8 (60), where we have $\rho = q_0$, $\xi = \left(p_{\min}^{(K)} q_1\right)^2 q_0^{-1}$, we obtain that with probability $1 - \eta/Q$, $|\hat{E} - E^*|, |\hat{E}_F|, |E_F^*| = \mathcal{O}\left(\left(p_{\min}^{(K)} q_1\right)^2\right)$. Then, similar to the second step in the proof of Lemma 4 (plugging the bound of $|\hat{E} - E^*|, |\hat{E}_F|, |E_F^*|$ into (46)), we can show that with probability $1 - \eta/Q$,

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{q_1}{T}, \quad |p_m - r_{k_m}^*| \leq q_1 p_m.$$

Next, similar to previous argument, using these inequalities with Lemma 8 (60), where we have $\rho = q_1$, $\xi = \left(p_{\min}^{(K)} q_2\right)^2 q_1^{-1}$, we obtain that with probability $1 - 3\eta/(2Q)$, $|\hat{E} - E^*|, |\hat{E}_F|, |E_F^*| = \mathcal{O}\left(\left(p_{\min}^{(K)} q_2\right)^2\right)$. This further implies, with probability $1 - 2\eta/Q$,

$$|\lambda_m - \theta_{k_m}^*| \leq \frac{q_2}{T}, \quad |p_m - r_{k_m}^*| \leq q_2 p_m.$$

Doing this repeatedly, we finally obtain (58). \square

Finally, we are ready to prove the proposition 5:

Proof of Proposition 5. Using the given q from the conditions of the proposition, we construct a decreasing positive sequence $\{q_s\}_{s=0}^S$ with $q_s = q^{\frac{2-(1/2)^s}{2-(1/2)^S}}$. With this choice, we have $q_{s+1}^{-4} q_s^2 = q^{-\frac{4}{2-(1/2)^S}}$.

Setting $\zeta = (1/2)^{S-1} > \frac{4}{2-(1/2)^S} - 2$ and using Lemma 7 (58), we prove (51). \square

E Bound of the expectation error

In this section, we bound the expectation error. Recall the definition in (40),

$$\begin{aligned}
E_{p,r}(\mathbf{r}, \boldsymbol{\theta}) &= 2\text{Re} \left(\sum_{m=1}^M \sum_{k=1}^K p_m r_k \left(\frac{1}{N} \sum_{n=1}^N \exp(i(\theta_k - \lambda_m)t_n) - F(\theta_k - \lambda_m) \right) \right), \\
E_{r,r}(\mathbf{r}, \boldsymbol{\theta}) &= 2\text{Re} \left(\sum_{k_1 \neq k_2}^K \bar{r}_{k_1} r_{k_2} \left(\frac{1}{N} \sum_{n=1}^N \exp(i(\theta_{k_2} - \theta_{k_1})t_n) - F(\theta_{k_2} - \theta_{k_1}) \right) \right), \\
E_{r,Z}(\mathbf{r}, \boldsymbol{\theta}) &= -2\text{Re} \left(\sum_{k=1}^K \bar{r}_k \left(\frac{1}{N} \sum_{n=1}^N E_n \exp(i\theta_k t_n) \right) \right).
\end{aligned}$$

Intuitively, when $N \gg 1$, we have $|E_{p,r}|, |E_{r,r}|, |E_{r,Z}| = \mathcal{O}(1/\sqrt{N})$ for fixed $\mathbf{r}, \boldsymbol{\theta}$. However, this result does not directly apply to the optimization problem because the range of $(\mathbf{r}, \boldsymbol{\theta})$ is always an infinite set. To overcome this difficulty, we notice that the Lipschitz constant of the function $\exp(i\theta t)$ is bounded by T if $t \leq T$. First, we use Hoeffding's inequality to obtain a uniform bound for these expectation errors with a finite number of $(\mathbf{r}, \boldsymbol{\theta})$ points. Then, we extend this bound to all other points using the Lipschitz continuity property of $\exp(i\theta t)$. Specifically, we have the following result that gives a uniform bound for these expectation errors:

Lemma 8. *Define $E_{p,r}, E_{r,r}, E_{r,Z}$ as above. Assume $a(t)$ is defined as (3), then*

- *Given $0 < \eta < 1/2$ and a sequence of interval $\{I_k\}_{k=1}^K$ on \mathbb{R} , we define $\rho = \max_k \{T|I_k|/2\}$, if*

$$N = \Omega \left(\max\{\gamma^2 \rho^2, 1\} \xi^{-2} \text{polylog}(K \xi^{-1} \eta^{-1}) \right),$$

then

$$\begin{aligned}
\mathbb{P} \left(\sup_{\|\mathbf{r}\|_1 \leq 1, \theta_k \in I_k} |E_{p,r}(\mathbf{r}, \boldsymbol{\theta})| \geq \xi \right) &\leq \eta \\
\mathbb{P} \left(\sup_{\|\mathbf{r}\|_1 \leq 1, \theta_k \in I_k} |E_{r,r}(\mathbf{r}, \boldsymbol{\theta})| \geq \xi \right) &\leq \eta. \\
\mathbb{P} \left(\sup_{\|\mathbf{r}\|_1 \leq 1, \theta_k \in I_k} |E_Z(\mathbf{r}, \boldsymbol{\theta})| \geq \xi \right) &\leq \eta
\end{aligned} \tag{59}$$

- *Denote $\mathcal{D} = \{m_1, \dots, m_K\}$. Given $0 < \eta < 1/2$, a sequence of intervals $\{I_k\}_{k=1}^K$ on \mathbb{R} , and a sequence of discs $\{R_k\}_{k=1}^K$ on \mathbb{C} , we assume: 1. $\lambda_{m_k} \in I_k$; 2. $p_{m_k} \in R_k$. Define $\rho = \max_k \{T|I_k|/2, \text{radius}(R_k)/p_{m_k}\}$. If*

$$N = \Omega \left(\max\{\gamma^2 \rho^2, 1\} \xi^{-2} \text{polylog}(K \xi^{-1} \eta^{-1}) \right),$$

then

$$\begin{aligned}
\mathbb{P} \left(\sup_{r_k \in R_k, \theta_k \in I_k} |E_{p,r}(\mathbf{r}, \boldsymbol{\theta}) - E_{p,r}(\{(p_{m_k}, \lambda_{m_k})\}_{k=1}^K)| \geq \rho \xi \right) &\leq \eta \\
\mathbb{P} \left(\sup_{r_k \in R_k, \theta_k \in I_k} |E_{r,r}(\mathbf{r}, \boldsymbol{\theta}) - E_{r,r}(\{(p_{m_k}, \lambda_{m_k})\}_{k=1}^K)| \geq \rho \xi \right) &\leq \eta \\
\mathbb{P} \left(\sup_{r_k \in R_k, \theta_k \in I_k} |E_Z(\mathbf{r}, \boldsymbol{\theta}) - E_Z(\{(p_{m_k}, \lambda_{m_k})\}_{k=1}^K)| \geq \rho \xi \right) &\leq \eta
\end{aligned} \tag{60}$$

Proof of Lemma 8. We start by proving (59). We will only prove the inequalities for $E_{p,r}$, but note that the other two can be shown similarly. Define

$$G(\theta) = \frac{1}{N} \sum_{n=1}^N \left(\sum_{m=1}^M p_m \exp(i(\theta - \lambda_m)t_n) \right) - \left(\sum_{m=1}^M p_m F(\theta - \lambda_m) \right).$$

For the truncated filter, since $t_n \leq \gamma T$, we have

$$|G(\theta_1) - G(\theta_2)| \leq \gamma T |\theta_1 - \theta_2|,$$

which implies G is a γT -Lipschitz function. Because $|F(\theta - \lambda_m)| \leq 1$, using Hoeffding's inequality, for fixed θ , we have

$$\mathbb{P}(|G(\theta)| \geq \xi) \leq 4 \exp\left(-\frac{N\xi^2}{32}\right).$$

Combining this with the fact that G is a γT -Lipschitz function, we have

$$\mathbb{P}\left(\sup_{\theta \in I_k} |G(\theta)| \geq \xi + \gamma T \epsilon\right) \leq \frac{8\rho}{T\epsilon} \exp\left(-\frac{N\xi^2}{32}\right).$$

Choosing $\epsilon = \frac{\rho}{T\sqrt{N}}$, $\xi = \frac{4\sqrt{2}}{\sqrt{N}} \log^{1/2}(8\sqrt{N}K/\eta)$,

$$\mathbb{P}\left(\sup_{\theta \in I_k} |G(\theta)| \geq \left(4\sqrt{2} \log^{1/2}(8\sqrt{N}K/\eta) + \gamma\rho\right) \frac{1}{\sqrt{N}}\right) \leq \frac{\eta}{K},$$

which implies the first inequality of (59) using $\sum |r_k| \leq 1$.

We proceed to prove (60). As before, we only show the proof of the first inequality in (60). Fixed $1 \leq k \leq K$, when $r_k \in R_k$ and $\theta_k \in I_k$, we have

$$r_k G(\theta_k) - p_{m_k} G(\lambda_{m_k}) = (r_k - p_{m_k}) G(\theta_k) + p_{m_k} (G(\theta_k) - G(\lambda_{m_k})).$$

For the first term, we have $|r_k - p_{m_k}| \leq 2\rho p_{m_k}$ and

$$\mathbb{P}\left(\sup_{\theta_k \in I_k} |G(\theta_k)| \geq \xi/2\right) \leq \frac{\eta}{2K},$$

according to the previous proof. This implies

$$\mathbb{P}\left(\sup_{r_k \in R_k, \theta_k \in I_k} \sum_{k=1}^K |r_k - p_{m_k}| |G(\theta_k)| \geq \rho\xi/2\right) \leq \frac{\eta}{2}. \quad (61)$$

For the second term, we first notice that

$$\left| \sum_{m=1}^M p_m \exp(i(\theta_k - \lambda_m)t_n) - \sum_{m=1}^M p_m \exp(i(\lambda_{m_k} - \lambda_m)t_n) \right| \leq \gamma T |\theta_k - \lambda_{m_k}| \leq 2\rho$$

and $G(\theta)$ is a γT -Lipschitz function. Then, similar to before, we have

$$\mathbb{P}\left(\sup_{\theta_k \in I_k} \sum_{k=1}^K |G(\theta_k) - G(\lambda_{m_k})| \geq \rho\xi/2\right) \leq \frac{\eta}{2K}.$$

Since $\sum_k p_{m_k} \leq 1$, we have

$$\mathbb{P}\left(\sup_{r_k \in R_k, \theta_k \in I_k} \sum_{k=1}^K |p_k (G(\theta_k) - G(\lambda_{m_k}))| \geq \rho\xi/2\right) \leq \frac{\eta}{2}.$$

Combining this inequality with (61), we prove the first inequality of (60). \square