Infinite quantum signal processing

Yulong Dong[1], Lin Lin[1,2,3], Hongkang Ni[4], and Jiasu Wang[1]

[1]Department of Mathematics, University of California, Berkeley, CA 94720, USA.

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

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

[4] Institute for Computational and Mathematical Engineering (ICME), Stanford University, Stanford, CA 94305, USA.

Quantum signal processing (QSP) represents a real scalar polynomial of degree d using a product of unitary matrices of size 2×2, parameterized by $(d+1)$ real numbers called the phase factors. This innovative representation of polynomials has a wide range of applications in quantum computation. When the polynomial of interest is obtained by truncating an infinite polynomial series, a natural question is whether the phase factors have a well defined limit as the degree $d \to \infty$. While the phase factors are generally not unique, we find that there exists a consistent choice of parameterization so that the limit is well defined in the ℓ^1 space. This generalization of QSP, called the infinite quantum signal processing, can be used to represent a large class of non-polynomial functions. Our analysis reveals a surprising connection between the regularity of the target function and the decay properties of the phase factors. Our analysis also inspires a very simple and efficient algorithm to approximately compute the phase factors in the ℓ^1 space. The algorithm uses only double precision arithmetic operations, and provably converges when the ℓ^1 norm of the Chebyshev coefficients of the target function is upper bounded by a constant that is independent of d. This is also the first numerically stable algorithm for finding phase factors with provable performance guarantees in the limit $d \to \infty$.

Contents

1 Introduction

1.1 Background

The study of the representation of polynomials has a long history, with rich applications in a diverse range of fields. It is therefore exciting that a new way of representing polynomials, called quantum signal processing[1] (QSP) [18, 12], has emerged recently in the context of quantum computation. The motivation for this development can be seen as follows. For simplicity let $H \in \mathbb{C}^{N\times N}$ be a Hermitian matrix with all eigenvalues in the interval $[-1, 1]$, and let $f \in \mathbb{R}[x]$ be a real scalar polynomial of degree d. We would like to efficiently encode a matrix polynomial $f(H)$ using a unitary matrix that can be efficiently implemented on a quantum computer. An inherent difficulty of this task is that a quantum algorithm is given by the product of a sequence of unitary matrices, but in general neither H nor $f(H)$ is unitary. In an extreme scenario, let $H = x \in [-1, 1]$ be a scalar, and we are interested in a representation of the polynomial $f(x)$ in terms of unitary matrices.

Quantum signal processing proposes the following solution to the problem above: Let

$$W(x) = e^{\mathrm{i}\arccos(x)X} = \begin{pmatrix} x & \mathrm{i}\sqrt{1-x^2} \\ \mathrm{i}\sqrt{1-x^2} & x \end{pmatrix}, \text{ with } X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

[1] The term "signal processing" is due to an analogy to digital filter designs on classical computers.

be a 2×2 unitary matrix parameterized by $x \in [-1, 1]$. Then the following expression

$$U(x, \Psi) := e^{\mathrm{i}\psi_0 Z} \prod_{j=1}^{d} \left[W(x) e^{\mathrm{i}\psi_j Z} \right], \text{ with } Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1}$$

is a unitary matrix for any choice of phase factors $\Psi := (\psi_0, \psi_1, \cdots, \psi_d) \in \mathbb{R}^{d+1}$. Here X, Z are Pauli matrices. One can verify that the top-left entry of $U(x, \Psi)$ is a complex polynomial in x. Moreover, for any *target polynomial* $f \in \mathbb{R}[x]$ satisfying (1) $\deg(f) = d$, (2) the parity of f is $d \bmod 2$, (3) $\|f\|_\infty := \max_{x \in [-1,1]} |f(x)| \leq 1$, we can find phase factors $\Psi \in \mathbb{R}^{d+1}$ such that $f(x)$ is equal to the real (or imaginary) part of the top-left entry of $U(x, \Psi)$ for all $x \in [-1, 1]$ [12]. By setting $x = \frac{z+z^{-1}}{2}$ with $z = e^{\mathrm{i}\theta}, \theta \in [0, 2\pi)$, the representation in Eq. (1) can be viewed as a 2×2 matrix Laurent polynomial, and we are interested in its values on the unit circle.

Eq. (1) is an innovative way of encoding the information of a polynomial in terms of 2×2 unitary matrices. It also leads to a very compact quantum algorithm for constructing a unitary matrix that encodes the matrix polynomial $f(H)$, called *quantum singular value transformation (QSVT)* [12]. Assume that H is accessed via its *block encoding*

$$W = \begin{pmatrix} H & * \\ * & * \end{pmatrix},$$

where $W \in \mathbb{C}^{MN \times MN}$ is a unitary matrix ($M \geq 2$), the matrix H is its top-left $N \times N$ matrix subblock, and $*$ indicates matrix entries irrelevant to the current task. When given the phase factors Ψ and the block encoding W, QSVT constructs a unitary $U \in \mathbb{C}^{2MN \times 2MN}$ (by introducing one ancillary qubit) such that

$$U = \begin{pmatrix} f(H) & * \\ * & * \end{pmatrix}.$$

In other words, although $f(H)$ is not a unitary matrix, it can be block encoded by a unitary matrix of a larger size that can be efficiently implemented on quantum computers. This construction has found many applications, such as Hamiltonian simulation [18, 12], linear system of equations [12, 17, 20], eigenvalue problems [16, 7], Gibbs states preparation [12], Petz recovery channel [11], benchmarking quantum systems [5, 9], to name a few. We refer interested readers to Refs. [12, 20].

To implement QSVT, we need to efficiently calculate the phase factors Ψ corresponding to a target polynomial of degree d. Many of the aforementioned applications are formulated as the evaluation of a matrix function $f(H)$, where $f : [-1, 1] \to \mathbb{R}$ is not a polynomial but a smooth function, which can be expressed as an infinite polynomial series (e.g., the Chebyshev polynomial series). To approximate $f(H)$, we need to first truncate the polynomial series to $f^{(d)}$ with a proper degree d so that the difference between $f^{(d)}$ and f is sufficiently small. Then for each $f^{(d)}$ we can find (at least) one set of phase factors $\Psi^{(d)}$. When d is fixed, there has been significant progress in computing the phase factors in the past few years [12, 13, 3, 8, 26]. The questions we would like to answer in this paper are as follows.

1. As $d \to \infty$, can the phase factors $\{\Psi^{(d)}\}$ be chosen to have a well-defined limit $\Psi^\star$ in a properly chosen space?

2. If f is smooth, its Chebyshev coefficients decay rapidly. Does the tail of $\Psi^\star$ exhibit decay properties? If so, how is it related to the smoothness of f?

3. Is there an efficient algorithm to approximately compute $\Psi^\star$?

Our work is to first generalize QSP to represent smooth functions with a set of infinitely long phase factors, and we dub the resulting limit *infinite quantum signal processing* (iQSP).

1.2 Setup of the problem

We follow the bra-ket notation widely used in quantum mechanics. Specifically, we define two "kets" as basis vectors of $\mathbb{C}^2$, namely,

$$|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The "bra" can be viewed as row vectors induced from the corresponding ket by taking Hermitian conjugate. In the bra notation, $\langle 0| = (1,0), \langle 1| = (0,1)$. The inner product is written as $\langle x|y\rangle := (|x\rangle, |y\rangle)$. Using this notation, the upper left element of $U(x,\Psi)$ in Eq. (1) can be written as $\langle 0|U(x,\Psi)|0\rangle$. Direct calculation shows that the real part of $\langle 0|U(x,\Psi)|0\rangle$ can be recovered from the imaginary part by adding $\frac{\pi}{4}$ to both ψ_0 and ψ_d:

$$\mathrm{Re}[\langle 0|U(x,\Psi)|0\rangle] = \mathrm{Im}\left[\langle 0|e^{\mathrm{i}\frac{\pi}{4}Z}U(x,\Psi)e^{\mathrm{i}\frac{\pi}{4}Z}|0\rangle\right]. \tag{2}$$

For convenience, throughout this paper, we focus on the imaginary part of $\langle 0|U(x,\Psi)|0\rangle$, which is denoted by $g(x,\Psi)$, i.e.,

$$g(x,\Psi) := \mathrm{Im}[\langle 0|U(x,\Psi)|0\rangle]. \tag{3}$$

Due to the parity constraint, the number of degrees of freedom in a given target polynomial $f \in \mathbb{R}[x]$ is only $\widetilde{d} := \lceil \frac{d+1}{2} \rceil$. Therefore the phase factors Ψ cannot be uniquely defined. To address this problem, Ref. [8] suggests that phase factors $\Psi := (\psi_0, \psi_1, \ldots, \psi_d)$ can be restricted to be symmetric:

$$\psi_j = \psi_{d-j}, \quad \forall j = 0, 1, \cdots, d. \tag{4}$$

Without loss of generality, we define the *reduced phase factors* as follows,

$$\Phi = (\phi_0, \phi_1, \ldots, \phi_{\widetilde{d}-1}) := \begin{cases} (\frac{1}{2}\psi_{\widetilde{d}-1}, \psi_{\widetilde{d}}, \cdots, \psi_d), & d \text{ is even}, \\ (\psi_{\widetilde{d}}, \psi_{\widetilde{d}+1}, \cdots, \psi_d), & d \text{ is odd}. \end{cases} \tag{5}$$

The number of reduced phase factors is equal to $\widetilde{d} = \lceil \frac{d+1}{2} \rceil$, and matches the number of degrees of freedom in f. With some abuse of notation, we identify $U(x,\Phi)$ with $U(x,\Psi)$ and $g(x,\Phi)$ with $g(x,\Psi)$, and Φ is always referred to as the *reduced phase factors* of a full set of phase factors Ψ. For a given target polynomial, the existence of the symmetric phase factors is proved in [25, Theorem 1], but the choice is still not unique. However, near the trivial phase factors $\Phi = (0, \ldots, 0)$, there exists a unique and consistent choice of symmetric phase factors called the maximal solution [25].

Let ℓ^1 denote the set of all infinite dimensional vectors with finite 1-norm:

$$\ell^1 := \{v = (v_0, v_1, \cdots) : \|v\|_1 < \infty\}, \quad \|v\|_1 := \sum_{k=0}^{\infty} |v_k|, \quad v = (v_0, v_1, \cdots). \tag{6}$$

The vector space ℓ^1 is complete, i.e., every Cauchy sequence of points in ℓ^1 has a limit that is also in ℓ^1. Let $\mathbb{R}^\infty$ be the set of all infinite dimensional vectors with only a finite number of nonzero elements.

Definition 1 (Target function)**.** *A target function $f : \mathbb{R} \to \mathbb{R}$ is an infinite Chebyshev polynomial series with a definite parity*

$$f(x) = \begin{cases} \sum_{j=0}^{\infty} c_j T_{2j}(x), & f \text{ is even}, \\ \sum_{j=0}^{\infty} c_j T_{2j+1}(x), & f \text{ is odd}, \end{cases} \tag{7}$$

The coefficient vector $c = (c_0, c_1, \ldots) \in \ell^1$, and f satisfies the norm constraint

$$\|f\|_\infty = \max_{x\in[-1,1]} |f(x)| \leq 1. \tag{8}$$

In other words, the set of even target functions is

$$S_e = \left\{ f : [-1,1] \to [-1,1] : f(x) = \sum_{j=0}^{\infty} c_j T_{2j}(x), \quad \sum_{j=0}^{\infty} |c_j| < \infty \right\}, \tag{9}$$

and the set of odd target functions is

$$S_o = \left\{ f : [-1,1] \to [-1,1] : f(x) = \sum_{j=0}^{\infty} c_j T_{2j+1}(x), \quad \sum_{j=0}^{\infty} |c_j| < \infty \right\}. \tag{10}$$

If we truncate the Chebyshev coefficients to be $c^{(\widetilde{d})} = (c_0, c_1, \ldots, c_{\widetilde{d}-1}, 0, \ldots) \in \mathbb{R}^\infty$, the corresponding Chebyshev polynomial $f^{(d)}$ is of degree d (recall that $\widetilde{d} = \lceil \frac{d+1}{2} \rceil$ and hence d is determined by $\widetilde{d}$ and the parity of the function). Furthermore, $c \in \ell^1$ implies that $\lim_{d\to\infty} \left\| f^{(d)} - f \right\|_\infty = 0$. Throughout the paper, $f^{(d)}$ will be referred to as a *target polynomial* approximating the target function f.

In order to compare phase factors of different lengths, an important observation is that if we pad $\Phi = (\phi_0, \phi_1, \ldots, \phi_{\widetilde{d}})$ with an arbitrary number of 0's at the right end and obtain $\widetilde{\Phi} = (\phi_0, \phi_1, \ldots, \phi_{\widetilde{d}}, 0, \ldots, 0)$, we have $g(x, \Phi) = g(x, \widetilde{\Phi})$ (see Lemma 10). Therefore $g(x, \cdot)$ is a well defined mapping in $\mathbb{R}^\infty$, and we can identify Φ with $\widetilde{\Phi}$. Let $\mathcal{F}$ be the linear mapping from a target polynomial to its Chebyshev-coefficient vector $c \in \mathbb{R}^\infty$ as defined in Eq. (7). This induces a mapping

$$F : \mathbb{R}^\infty \to \mathbb{R}^\infty, \quad F(\Phi) := \mathcal{F}(g(x, \Phi)), \tag{11}$$

which maps the reduced phase factors $\Phi \in \mathbb{R}^\infty$ to the Chebyshev coefficients of $g(x, \Phi)$.

Note that $\mathbb{R}^\infty$ is dense in ℓ^1, i.e., any point in ℓ^1 is either a point in $\mathbb{R}^\infty$ or a limit point of $\mathbb{R}^\infty$. By exploiting nice properties that F and its Jacobian matrix are Lipschitz continuous, we can define $\overline{F} : \ell^1 \to \ell^1$ to be the extension of F, such that $\overline{F}(\Phi)$ agrees with $F(\Phi)$ for any $\Phi \in \mathbb{R}^\infty$. Then the problem of infinite quantum signal processing asks whether the inverse of the mapping $\overline{F}$ exists.

Problem 2 (Infinite quantum signal processing)**.** *For a target function in Definition 1 given by its Chebyshev coefficients $c \in \ell^1$, does there exist $\Phi^\star \in \ell^1$ such that $\overline{F}(\Phi^\star) = c$?*

1.3 Main results

Theorem 3 (Invertibility of $\overline{F}$)**.** *There exists a universal constant $r_c \approx 0.902$, so that $\overline{F}$ has an inverse map $\overline{F}^{-1} : B(0, r_c) \subset \ell^1 \to \ell^1$, where $B(a, r) := \{v \in \ell^1 : \|v - a\|_1 < r\}$.*

Theorem 3 provides a positive answer to Problem 2 as well as to the first question raised in Section 1.1, when the 1-norm of the Chebyshev coefficients is upper bounded by a constant. Note that for a given target function f, we can always multiply it by a constant C, so that the Cf satisfies the condition of Theorem 3. The main technical tools are a series of vector 1-norm estimates of F, and matrix 1-norm estimates of the Jacobian matrix DF. These estimates do not explicitly depend on the length of phase factors, and can therefore be extended to $\overline{F}$. A more detailed statement of Theorem 3 is Theorem 23, which will be presented in Section 3.4.

Since $\Phi^\star = (\phi_0, \phi_1, \ldots) \in \ell^1$, the tail of $\Phi^\star$ must exhibit decay properties, i.e., $\lim_{n\to\infty} \sum_{k>n} |\phi_k| = 0$. Fig. 4 in Section 6 shows that the tail decay of $\Phi^\star$ closely matches that of the Chebyshev coefficients c. The duality between the smoothness of a function and the decay of its Fourier / Chebyshev coefficients is well studied (see e.g. [24, Chapter 7]). However, it is surprising that the tail decay of the reduced phase factors can be directly related to the smoothness of the target function. Such a behavior was first numerically observed in Ref. [8], in which an explanation of the phenomenon was also given in the perturbative regime. Using the tools developed in proving Theorem 3, we provide a refined and non-perturbative analysis of the tail decay in Theorem 4.

Theorem 4 (Decay properties of reduced phase factors)**.** *Given a target function f with $\|c\|_1 < r_c$, and $\Phi^\star := \overline{F}^{-1}(c) = (\phi_0, \phi_1, \ldots) \in \ell^1$, then there exist constants C, C' such that for any n,*

$$C' \sum_{k>n} |c_k| \leq \sum_{k>n} |\phi_k| \leq C \sum_{k>n} |c_k| \,. \tag{12}$$

The proof is given in Section 4 with an explicit characterization of constants C, C'. Assume the target function f is of C^α smoothness for some $\alpha > 0$, then the Chebyshev coefficients decay algebraically in the sense of $\sum_{k>n} |c_k| = \mathcal{O}\left(n^{-\alpha}\right)$. Then, it induces a decay of the corresponding reduced phase factors, namely, $\sum_{k>n} |\phi_k| = \mathcal{O}\left(n^{-\alpha}\right)$. If f is of C^∞ or C^ω smoothness, then the tail of its Chebyshev-coefficient vector decays super-algebraically or exponentially respectively, and so does the tail of the reduced factors. We summarize this in the following corollary.

Corollary 5. *If f is of C^α smoothness for some $\alpha > 0$, then the tail of the corresponding reduced phase factors decay algebraically, i.e., $\sum_{k>n} |\phi_k| = \mathcal{O}\left(n^{-\alpha}\right)$. Furthermore, if f is of C^∞ or C^ω smoothness, then the tail of the corresponding reduced phase factors decays super-algebraically or exponentially respectively.*

These results are also verified numerically in Fig. 4. These results provide a positive answer to the second question raised in Section 1.1.

Theorem 3 also has algorithmic implications. It has been empirically observed that a quasi-Newton optimization based algorithm is highly effective for finding the phase factors [8]. However, the theoretical justification of the optimization based algorithm has only been shown if the target function satisfies $\|f\|_\infty \leq C_f \widetilde{d}^{-1}$, where $C_f \approx 0.028$ is a universal constant [25, Corollary 7]. Hence as $\widetilde{d}$ increases, existing theoretical results fail to predict the effectiveness of the algorithm, even if the target function is a fixed polynomial of finite degree (in this case, we pad the Chebyshev coefficients with zeros to increase $\widetilde{d}$).

Inspired by our analysis of the Jacobian map DF, we propose a very simple iterative algorithm to find phase factors for a given target polynomial f (Algorithm 1). This algorithm can be viewed as finding the fixed point of the mapping $G(x) := x - \frac{1}{2}(F(x) - c)$ by means of a fixed point iteration $\Phi^{t+1} = G(\Phi^t)$. This can also be viewed as an inexact Newton algorithm [22, Chapter 11], as the inverse of the Jacobian matrix at $0 \in \ell^1$ satisfies $[DF(0)]^{-1} = \frac{1}{2}\mathbb{I}$ (Lemma 17).

Algorithm 1 Fixed-point iteration algorithm for solving reduced phase factors

Input: Chebyshev-coefficient vector c of a target polynomial, and stopping criteria.
Initiate $\Phi^0 = 0$, $t = 0$;
while stopping criterion is not satisfied **do**
  Update $\Phi^{t+1} = \Phi^t - \frac{1}{2}\left(F\left(\Phi^t\right) - c\right)$;
  Set $t = t + 1$;
end while
Output: Reduced phase factors Φ^t.

For a given target function, the Chebyshev-coefficient vector can be efficiently evaluated using the fast Fourier transform (FFT). The fixed-point iteration algorithm (Algorithm 1) is the simplest algorithm thus far to evaluate phase factors. This algorithm provably converges when $\|c\|_1$ is upper bounded by a constant.

Theorem 6 (Convergence of the fixed-point iteration algorithm)**.** *There exists a universal constant $\widetilde{r}_c \approx 0.861$ so that when $\|c\|_1 \leq \widetilde{r}_c$,*

(i) Algorithm 1 converges Q-linearly to $\Phi^\star = \overline{F}^{-1}(c)$, i.e., there exists a constant C and the error satisfies

$$\left\|\Phi^t - \Phi^\star\right\|_1 \leq C\widetilde{\gamma}^t, \quad \widetilde{\gamma} \approx 0.8189, \quad t \geq 1. \tag{13}$$

(ii) The overall time complexity is $\mathcal{O}(d^2 \log \frac{1}{\epsilon})$, where d is the degree of target polynomial and ϵ is the desired precision.

A more accurate characterization about the region where Algorithm 1 converges and the convergence rate is presented in Section 5.1. In Section 6, numerical experiments demonstrate that Algorithm 1 is an efficient algorithm, and we observe that its convergence radius can be much larger than the theoretical prediction. These results provide a positive answer to the third question raised in Section 1.1. We implement Algorithm 1 with more examples as part of QSPPACK, an open-source package for finding phase factors [2].

1.4 Related works

The original QSP paper [18] demonstrated the existence of phase factors without providing a constructive algorithm, and finding phase factors was considered to be a main bottleneck of the approach [4]. In the past few years, there has been significant progresses in computing the phase factors. Refs. [12, 13] developed the factorization based method. For a given target (real) polynomial $f^{(d)}$, one needs to find a complementary (real) polynomial satisfying the requirement of [12, Corollary 5] (also see Theorem 7). This step is based on finding roots of high degree polynomials to high precision, and this is not a numerically stable procedure. Specifically, the algorithm requires $\mathcal{O}(d\log(d/\epsilon))$ bits of precision [13]. There have been two recent improvements of the factorization based method, based on the capitalization method [3], and the Prony method [26], respectively. Although the two methods differ significantly, empirical results indicate that both methods are numerically stable, and are applicable to polynomials of large degrees. Furthermore, both methods take advantage of that the mapping from the Chebyshev coefficients c to the full phase factors Ψ *is not necessarily* well-defined in ℓ^1. For instance, a key step in [3] is to introduce

[2]The examples are available on the website https://qsppack.gitbook.io/qsppack/ and the codes are open-sourced in https://github.com/qsppack/QSPPACK.

a very small perturbation to the high order Chebyshev coefficients, which can nonetheless induce a large change in the phase factors Ψ. As a result, the question raised in Problem 2 cannot be well defined in such factorization based methods, and the tail of the phase factors Ψ does not exhibit decay properties.

The optimization based method developed in Ref. [8] uses a different approach, and computes the symmetric phase factors without explicitly constructing the complementary polynomials. Empirical results show that the quasi-Newton optimization method in [8] is numerically stable and can be applicable to polynomials of large degrees. Ref. [25] analyzes the symmetric QSP, and proves that starting from a fixed initial guess of the reduced phase factors $\Phi = (0, \cdots, 0)$, a simpler optimization method (the projected gradient method) converges linearly to a unique maximal solution, when the target polynomial satisfies $\left\| f^{(d)} \right\|_\infty \leq Cd^{-1}$ for some constant C. The fixed point iteration method in Algorithm 1 is the simplest algorithm thus far for finding phase factors, and is the first provably numerically stable algorithm in the limit $d \to \infty$.

2 Preliminaries on quantum signal processing

The set $[n] := \{0, 1, \cdots, n-1\}$ is referred to as the index set generated by a positive integer n. The row and column indices of a n-by-n matrix run from 0 to $n-1$, namely in the index set $[n]$. For a matrix $A \in \mathbb{C}^{m\times n}$, the transpose, Hermitian conjugate and complex conjugate are denoted by $A^\top$, $A^\dagger$, A^*, respectively. The same notations are also used for the operations on a vector.

For a matrix M of infinite dimension, we equip it with the induced 1-norm, i.e.,

$$\|M\|_1 := \max_{v\in\ell^1, \|v\|_1=1} \|Mv\|_1 .$$

For any function f over $[-1, 1]$, we define its infinity norm as $\|f\|_\infty := \max_{-1\leq x\leq 1} |f(x)|$. The key to quantum signal processing (QSP) is a representation theorem for certain matrices in SU(2):

Theorem 7 (**Quantum signal processing** [12, Theorem 4])**.** *For any $P, Q \in \mathbb{C}[x]$ and a positive integer d such that*

(1) $\deg(P) \leq d, \deg(Q) \leq d-1$,

(2) P has parity $(d \mod 2)$ *and Q has parity* $(d-1 \mod 2)$,

(3) (Normalization condition) $|P(x)|^2 + (1-x^2)|Q(x)|^2 = 1, \forall x \in [-1, 1]$.

Then, there exists a set of phase factors $\Psi := (\psi_0, \cdots, \psi_d) \in [-\pi, \pi)^{d+1}$ such that

$$U(x, \Psi) = e^{\mathrm{i}\psi_0 Z} \prod_{j=1}^{d} \left[W(x) e^{\mathrm{i}\psi_j Z} \right] = \begin{pmatrix} P(x) & \mathrm{i}Q(x)\sqrt{1-x^2} \\ \mathrm{i}Q^*(x)\sqrt{1-x^2} & P^*(x) \end{pmatrix} \tag{14}$$

where

$$W(x) = e^{\mathrm{i}\arccos(x)X} = \begin{pmatrix} x & \mathrm{i}\sqrt{1-x^2} \\ \mathrm{i}\sqrt{1-x^2} & x \end{pmatrix}.$$

Here, the complex conjugate of a complex polynomial is defined by taking complex conjugate on all of its coefficients. X, Z are Pauli matrices. In most applications, we are only interested in using the real part of P. The following corollary is a slight variation of [12, Corollary 5], which states that the condition on the real part of P can be easily satisfied. Due to the relation between the real and imaginary components given in Eq. (2), the conditions on the imaginary part of P are the same.

Corollary 8 (**Quantum signal processing with real target polynomials** [12, Corollary 5]). *Let $f \in \mathbb{R}[x]$ be a degree-d polynomial for some $d \geq 1$ such that*

- *$f(x)$ has parity $(d \mod 2)$,*
- *$|f(x)| \leq 1, \forall x \in [-1, 1]$.*

Then there exists some $P, Q \in \mathbb{C}[x]$ satisfying properties (1)-(3) of Theorem 7 such that $f(x) = \mathrm{Im}[P(x)]$.

Since we are interested in P, we may further restrict $Q \in \mathbb{R}[x]$. In such a case, the phase factors can be restricted to be symmetric. Let D_d denote the domain of the symmetric phase factors:

$$D_d = \begin{cases} [-\frac{\pi}{2}, \frac{\pi}{2})^{\frac{d}{2}} \times [-\pi, \pi) \times [-\frac{\pi}{2}, \frac{\pi}{2})^{\frac{d}{2}}, & d \text{ is even}, \\ [-\frac{\pi}{2}, \frac{\pi}{2})^{d+1}, & d \text{ is odd}. \end{cases} \tag{15}$$

Theorem 9 (**Quantum signal processing with symmetric phase factors** [25, Theorem 1]). *Consider any $P \in \mathbb{C}[x]$ and $Q \in \mathbb{R}[x]$ satisfying the following conditions*

(1) $\deg(P) = d$ *and* $\deg(Q) = d - 1$.

(2) P has parity $(d \bmod 2)$ and Q has parity $(d - 1 \bmod 2)$.

(3) (Normalization condition) $\forall x \in [-1, 1] : |P(x)|^2 + (1 - x^2)|Q(x)|^2 = 1$.

(4) If d is odd, then the leading coefficient of Q is positive.

There exists a unique set of symmetric phase factors $\Psi := (\psi_0, \psi_1, \cdots, \psi_1, \psi_0) \in D_d$ such that

$$U(x, \Psi) = \begin{pmatrix} P(x) & \mathrm{i}Q(x)\sqrt{1 - x^2} \\ \mathrm{i}Q(x)\sqrt{1 - x^2} & P^*(x) \end{pmatrix}. \tag{16}$$

We want to emphasize that the set of symmetric phase factors is unique only if both $P(x)$ and $Q(x)$ are determined. If only $f = \mathrm{Im}[P]$ is given, then the set of symmetric phase factors may not be unique as mentioned above. When we are only interested in $f(x) = \mathrm{Im}[P(x)]$ represented by symmetric phase factors, the conditions on f are the same as those in Corollary 8. This is proved constructively in [25, Theorem 4].

Throughout the paper, unless otherwise specified, we refer to $\Psi := (\psi_0, \cdots, \psi_d)$ as the full set of phase factors and use $\Phi := (\phi_0, \cdots, \phi_{\widetilde{d}-1})$ to denote the set of reduced phase factors after imposing symmetry constraint on Ψ, where $\widetilde{d} := \lceil \frac{d+1}{2} \rceil$.

In this paper, in order to characterize decay properties, we choose the second half of Ψ to be the corresponding reduced phase factors. Specifically, when d is odd, the set of reduced phase factors is

$$\Phi = (\phi_0, \cdots, \phi_{\widetilde{d}-1}) := (\psi_{\widetilde{d}}, \cdots, \psi_d). \tag{17}$$

When d is even, the set of reduced phase factors is

$$\Phi = (\phi_0, \cdots, \phi_{\widetilde{d}-1}) := (\frac{1}{2}\psi_{\widetilde{d}-1}, \psi_{\widetilde{d}}, \cdots, \psi_d). \tag{18}$$

In this way, one has

$$U(x, \Psi) = e^{i\phi_{\widetilde{d}-1}Z} W(x) e^{i\phi_{\widetilde{d}-2}Z} \cdot \ldots \cdot W(x) e^{i\phi_0 Z} W(x) e^{i\phi_0 Z} W(x) \cdot \ldots \cdot e^{i\phi_{\widetilde{d}-2}Z} W(x) e^{i\phi_{\widetilde{d}-1}Z}$$

for the odd case, and

$$U(x, \Psi) = e^{i\phi_{\widetilde{d}-1}Z} W(x) e^{i\phi_{\widetilde{d}-2}Z} \cdot \ldots \cdot W(x) e^{2i\phi_0 Z} W(x) \cdot \ldots \cdot e^{i\phi_{\widetilde{d}-2}Z} W(x) e^{i\phi_{\widetilde{d}-1}Z}$$

for the even case.

Lemma 10 (Phase-factor padding)**.** *For any symmetric phase factors* Ψ*,*

$$\mathrm{Im}\left(\langle 0|U(x,\Psi)|0\rangle\right) = \mathrm{Im}\left(\langle 0|W(x)U(x,\Psi)W(x)|0\rangle\right). \tag{19}$$

Proof. According to the definition, $U(x,\Psi)$ takes the form

$$\begin{pmatrix} P(x) & \mathrm{i}Q(x)\sqrt{1-x^2} \\ \mathrm{i}Q(x)\sqrt{1-x^2} & P^*(x) \end{pmatrix}.$$

Here, $P(x) \in \mathbb{C}[x]$, and $Q(x) \in \mathbb{R}[x]$ due to the symmetry of Ψ. Direct computation shows:

$$\begin{aligned} &W(x)U(x,\Psi)W(x) \\ &= \begin{pmatrix} x & \mathrm{i}\sqrt{1-x^2} \\ \mathrm{i}\sqrt{1-x^2} & x \end{pmatrix} \begin{pmatrix} P(x) & \mathrm{i}Q(x)\sqrt{1-x^2} \\ \mathrm{i}Q(x)\sqrt{1-x^2} & P^*(x) \end{pmatrix} \begin{pmatrix} x & \mathrm{i}\sqrt{1-x^2} \\ \mathrm{i}\sqrt{1-x^2} & x \end{pmatrix} \\ &= \begin{pmatrix} xP-(1-x^2)Q & \mathrm{i}(xQ+P^*)\sqrt{1-x^2} \\ \mathrm{i}(P+xQ)\sqrt{1-x^2} & xP^*-(1-x^2)Q \end{pmatrix} \begin{pmatrix} x & \mathrm{i}\sqrt{1-x^2} \\ \mathrm{i}\sqrt{1-x^2} & x \end{pmatrix}. \end{aligned} \tag{20}$$

The upper-left entry of $W(x)U(x,\Psi)W(x)$ is

$$x^2P - x(1-x^2)Q - (xQ+P^*)(1-x^2) = x^2P + (x^2-1)P^* - 2x(1-x^2)Q. \tag{21}$$

Hence

$$\mathrm{Im}\left(\langle 0|W(x)U(x,\Psi)W(x)|0\rangle\right) = \mathrm{Im}\left(x^2P + (x^2-1)P^*\right) = x^2P_{\mathrm{Im}} + (1-x^2)P_{\mathrm{Im}} = P_{\mathrm{Im}}. \tag{22}$$

Note that $P_{\mathrm{Im}} = \mathrm{Im}\left(\langle 0|U(x,\Psi)|0\rangle\right)$, which completes the proof. □

Recall that we are interested in $g(x,\Psi) := \mathrm{Im}[\langle 0|U(x,\Psi)|0\rangle]$, and $g(x,\Phi)$ is identified with $g(x,\Psi)$. Lemma 10 implies that for reduced phase factors Φ, $g(x,\Phi)$ remains the same if we pad Φ with an arbitrary number of 0's at the right end. In this way, we are able to identify Φ with the infinite dimensional vector $(\phi_0,\cdots,\phi_{\widetilde{d}-1},0,\cdots)$ in $\mathbb{R}^\infty$. Then for any $\Phi^{(1)},\Phi^{(2)} \in \mathbb{R}^\infty$, the distance $\left\|\Phi^{(1)}-\Phi^{(2)}\right\|_1$ is well defined.

Definition 11. *The effective length of* $\Phi \in \mathbb{R}^\infty$ *is the largest index of the nonzero elements of* Φ*. If* $\Phi = (\phi_0,\cdots,\phi_{\widetilde{d}-1},0,\cdots)$ *and* $\phi_{\widetilde{d}-1} \neq 0$*, then its effective length is* $\widetilde{d}$*.*

By viewing reduced phase factors Φ as an infinite dimensional vector in $\mathbb{R}^\infty$, the problem of *symmetric quantum signal processing* is to find reduced phase factors $\Phi \in \mathbb{R}^\infty$ such that

$$F(\Phi) := \mathcal{F}\left(g(x,\Phi)\right) = \mathcal{F}(f) \tag{23}$$

holds for a target polynomial f.

3 Infinite quantum signal processing

We use $DF(\Phi)$ to denote the Jacobian matrix of $F(\Phi)$, which is a matrix of infinite dimension. Following the construction of the mapping F, for any $k \in \mathbb{N}$, the k-th column of $DF(\Phi)$ is

$$\frac{\partial F(\Phi)}{\partial \phi_k} = \mathcal{F}\left(\frac{\partial g(x,\Phi)}{\partial \phi_k}\right). \tag{24}$$

Similarly, for any $r, s \in \mathbb{N}$, the second order derivative is

$$\frac{\partial^2 F(\Phi)}{\partial \phi_r \partial \phi_s} = \mathcal{F}\left(\frac{\partial^2 g(x,\Phi)}{\partial \phi_r \partial \phi_s}\right). \tag{25}$$

As a remark, both $\frac{\partial F(\Phi)}{\partial \phi_k}$ and $\frac{\partial^2 F(\Phi)}{\partial \phi_r \partial \phi_s}$ are infinite dimensional vectors.

The main goal of this section is to prove Theorem 3. We first present a useful estimate of the vector 1-norm of F and $\frac{\partial^2 F(\Phi)}{\partial \phi_r \partial \phi_s}$ in Section 3.1. This allows us to estimate the matrix 1-norm of the Jacobian DF and prove the invertibility of DF in Section 3.2. Based on these technical preparations, we prove the invertibility of the mapping F in $\mathbb{R}^\infty$ in Section 3.3. As a final step, since $\mathbb{R}^\infty$ is dense in ℓ^1 and all derived estimates are independent of the effective length of Φ, we extend the result to the invertibility of $\overline{F}$ in Section 3.4. The analysis leverages some facts about the Banach space, which are summarized in Appendix A for completeness.

Without loss of generality, we consider the case that the target function is even in this section. The analytical results can be similarly generalized to the odd case.

3.1 Estimating the vector 1-norm of F and its second-order derivatives

We first summarize the main goal of this subsection as the following lemmas. To prove them, we consider a more general case where phase factors are not necessarily symmetric. Consequentially, we prove stronger results in Lemma 14 and Corollary 15. Lemma 12 and Lemma 13 are consequences of Lemma 14 and Corollary 15 respectively by restricting to the symmetric phase factors. As a remark, the upper bounds are independent of the effective length of the reduced phase factors, which will enable the generalization to ℓ^1.

Lemma 12. *For any $\Phi \in \mathbb{R}^\infty$, it holds that*

$$\|F(\Phi)\|_1 \leq \sinh\left(2\|\Phi\|_1\right). \tag{26}$$

Lemma 13. *For any $\Phi \in \mathbb{R}^\infty$, and $r, s \in \mathbb{N}$, it holds that*

$$\left\|\frac{\partial^2 F(\Phi)}{\partial \phi_r \partial \phi_s}\right\|_1 \leq 4\sinh\left(2\|\Phi\|_1\right). \tag{27}$$

To prove the previous lemmas, we start from a general setup where the phase factors are not necessarily symmetric.

Lemma 14. *For any full set of phase factors $\Psi := (\psi_0, \psi_1, \cdots, \psi_d)$, it holds that*

$$\|\mathcal{F}\left(g(x,\Psi)\right)\|_1 \leq \sinh\left(\|\Psi\|_1\right). \tag{28}$$

Proof. Let $\Psi = (\psi_0, \psi_1, \cdots, \psi_d)$ be a full set of phase factors. The corresponding QSP matrix can be expanded as

$$\begin{aligned} U(x,\Psi) :&= e^{\mathrm{i}\psi_0 Z}\prod_{k=1}^{d}\left(W(x)e^{\mathrm{i}\psi_k Z}\right) \\ &= \left(\cos(\psi_0)\mathbb{I} + \mathrm{i}\sin(\psi_0)Z\right)\prod_{k=1}^{d}\left(W(x)\left(\cos(\psi_k)\mathbb{I} + \mathrm{i}\sin(\psi_k)Z\right)\right) \\ &= \left(\prod_{k=0}^{d}\cos(\psi_k)\right)\left(\mathbb{I} + \mathrm{i}t_0 Z\right)\prod_{j=1}^{d}\left(W(x)\left(\mathbb{I} + \mathrm{i}t_j Z\right)\right), \end{aligned} \tag{29}$$

where $t_j := \tan\psi_j$. Notice that $W(x)Z = ZW(x)^{-1}$ and

$$W(x)^k = \begin{pmatrix} T_k(x) & \mathrm{i}\sqrt{1-x^2}U_{k-1}(x) \\ \mathrm{i}\sqrt{1-x^2}U_{k-1}(x) & T_k(x) \end{pmatrix}, \tag{30}$$

where $T_k(x)$ and $U_k(x)$ are Chebyshev polynomials of the first and second kind respectively. Then,

$$\begin{aligned} U(x,\Psi) &= \left(\prod_{k=0}^{d}\cos(\psi_k)\right)(\mathbb{I}+\mathrm{i}t_0 Z)\prod_{j=1}^{d}\left(W(x)\left(\mathbb{I}+\mathrm{i}t_j Z\right)\right) \\ &= \left(\prod_{k=0}^{d}\cos(\psi_k)\right)\Big(W^d(x) + \mathrm{i}\sum_{j_1=0}^{d} t_{j_1} Z W^{d-2j_1}(x) - \sum_{j_1<j_2} t_{j_1}t_{j_2} W^{d-2(j_2-j_1)}(x) \\ &\quad - \mathrm{i}\sum_{j_1<j_2<j_3} t_{j_1}t_{j_2}t_{j_3} Z W^{d-2(j_3-j_2+j_1)}(x) + \ldots\Big). \end{aligned} \tag{31}$$

Note that each term is a matrix whose upper left element is of the following form

$$\mathrm{i}^l \cdot t_{j_1}t_{j_2}\cdots t_{j_l}T_k(x),$$

where $j_1 < j_2 < \cdots < j_l$ for some l, and $k = d - 2\sum_{i=1}^{l}(-1)^{l-i}j_i$.

When considering the Chebyshev coefficients of the imaginary part of the upper left element of $U(x,\Psi)$, only those terms with odd number of t_j's matter. Then we have the estimate

$$\begin{aligned} \|\mathcal{F}\left(g(x,\Psi)\right)\|_1 &\le \prod_{k=0}^{d}|\cos(\psi_k)| \sum_{l \text{ is odd}}\sum_{j_1<j_2<\cdots<j_l}\prod_{i=1}^{l}|t_{j_i}| \\ &= \sum_{l \text{ is odd}}\sum_{j_1<j_2<\cdots<j_l}\prod_{i=1}^{l}|\sin(\psi_{j_i})|\prod_{k\neq j_1,\cdots,j_l}|\cos(\psi_k)| \\ &\le \sum_{l \text{ is odd}}\sum_{j_1<j_2<\cdots<j_l}\prod_{i=1}^{l}|\sin(\psi_{j_i})| \\ &\le \sinh\left(\sum_{k=0}^{d}|\sin(\psi_k)|\right) \\ &\le \sinh\left(\|\Psi\|_1\right). \end{aligned} \tag{32}$$

We remark that the last inequality holds because $\sinh(x)$ is monotonic increasing, and the penultimate equality is due to the following observation:

$$\begin{aligned} &\sinh\left(\sum_{k=0}^{d}|\sin(\psi_k)|\right) = \sum_{l \text{ is odd}}\frac{1}{l!}\left(\sum_{k=0}^{d}|\sin(\psi_k)|\right)^l = \sum_{l \text{ is odd}}\frac{1}{l!}\sum_{j_1,j_2,\cdots,j_l}\prod_{i=1}^{l}|\sin(\psi_{j_i})| \\ &\ge \sum_{l \text{ is odd}}\frac{1}{l!}\sum_{\substack{j_1,\cdots,j_l \\ \text{are all different}}}\prod_{i=1}^{l}|\sin(\psi_{j_i})| = \sum_{l \text{ is odd}}\sum_{j_1<j_2<\cdots<j_l}\prod_{i=1}^{l}|\sin(\psi_{j_i})|. \end{aligned} \tag{33}$$

The proof is completed. □

Corollary 15. *For any full set of phase factors* $\Psi := (\psi_0, \psi_1, \cdots, \psi_d)$ *and any* $r, s \in [d+1]$, *it holds that*

$$\left\| \mathcal{F}\left(\frac{\partial^2 g(x,\Psi)}{\partial \psi_r \partial \psi_s} \right) \right\|_1 \leq \sinh(\|\Psi\|_1). \tag{34}$$

Proof. If $r = s$, then $\partial_r^2 U(x,\Psi) = -U(x,\Psi)$, and the desired result can be obtained by directly applying Lemma 14. Then we consider the case $r \neq s$. Note that

$$\partial_r \partial_s U(x,\Psi) = U\left(x, \Psi + \frac{\pi}{2}e_r + \frac{\pi}{2}e_s\right), \tag{35}$$

where $e_r = (0, \ldots, 0, 1, 0, \ldots, 0)$ denotes the r-th standard unit vector. Then

$$\begin{aligned} \frac{\partial^2 g(x,\Psi)}{\partial \psi_r \partial \psi_s} &= \frac{\partial^2 \mathrm{Im}[\langle 0|U(x,\Psi)|0\rangle]}{\partial \psi_r \partial \psi_s} = \mathrm{Im}[\langle 0|\partial_r \partial_s U(x,\Psi)|0\rangle] \\ &= \mathrm{Im}[\langle 0|U(x, \Psi + \frac{\pi}{2}e_r + \frac{\pi}{2}e_s)|0\rangle] = g(x, \Psi + \frac{\pi}{2}e_r + \frac{\pi}{2}e_s). \end{aligned} \tag{36}$$

To simplify the notation, let $\tilde{\Psi} = \Psi + \frac{\pi}{2}e_r + \frac{\pi}{2}e_s$, and $\tilde{\psi}_k$ be the components of $\tilde{\Psi}$. Similar to the proof of Lemma 14, we have

$$\begin{aligned} &\left\| \frac{\partial^2 g(x,\Psi)}{\partial \psi_r \partial \psi_s} \right\|_1 = \left\| g(x,\tilde{\Psi}) \right\|_1 \\ &\leq \sum_{l \text{ is odd}} \sum_{j_1 < j_2 < \cdots < j_l} \prod_{i=1}^{l} \left|\sin(\tilde{\psi}_{j_i})\right| \prod_{k \neq j_1, \cdots, j_l} \left|\cos(\tilde{\psi}_k)\right| \\ &= \frac{1}{2} \prod_k \left(\left|\cos(\tilde{\psi}_k)\right| + \left|\sin(\tilde{\psi}_k)\right| \right) - \frac{1}{2} \prod_k \left(\left|\cos(\tilde{\psi}_k)\right| - \left|\sin(\tilde{\psi}_k)\right| \right) \\ &\overset{(*)}{=} \sum_{l \text{ is odd}} \sum_{j_1 < j_2 < \cdots < j_l} \prod_{i=1}^{l} |\sin(\psi_{j_i})| \prod_{k \neq j_1, \cdots, j_l} |\cos(\psi_k)| \\ &\leq \sum_{l \text{ is odd}} \sum_{j_1 < j_2 < \cdots < j_l} \prod_{i=1}^{l} |\sin(\psi_{j_i})| \\ &\leq \sinh\left(\|\Psi\|_1\right). \end{aligned} \tag{37}$$

To see that equality $(*)$ holds, we note that $|\sin(\tilde{\psi}_k)| = |\cos(\psi_k)|$ and $|\cos(\tilde{\psi}_k)| = |\sin(\psi_k)|$ for $k = r$ or s which interchanges two pairs of sine and cosine. Because this interchange operation leaves the production invariant, we can directly replace $\tilde{\Psi}$ by Ψ in the expression. The proof is completed. □

Lemma 12 is a direct application of Lemma 14. Now we use Corollary 15 to prove Lemma 13.

Proof of Lemma 13. Choose $n \geq \max(r, s)$ such that all elements of Φ with index $> n$ are zero. Then we may view Φ as a vector of length $n+1$, i.e., $(\phi_0, \cdots, \phi_n)$. Since we only consider the even case, we let $\Psi := (\psi_0, \psi_1, \cdots, \psi_{2n})$ be the corresponding full set of phase factors and then

$$\psi_k = \begin{cases} \phi_{n-k} & k < n, \\ 2\phi_0 & k = n, \\ \phi_{k-n} & k > n. \end{cases} \tag{38}$$

For first-order derivative, when $k > 0$,

$$\frac{\partial g(x,\Phi)}{\partial \phi_k} = \frac{\partial g(x,\Psi)}{\partial \psi_{n+k}} + \frac{\partial g(x,\Psi)}{\partial \psi_{n-k}}, \tag{39}$$

and when $k = 0$,

$$\frac{\partial g(x,\Phi)}{\partial \phi_k} = 2\frac{\partial g(x,\Psi)}{\partial \psi_n} = \frac{\partial g(x,\Psi)}{\partial \psi_{n+k}} + \frac{\partial g(x,\Psi)}{\partial \psi_{n-k}}. \tag{40}$$

Similarly, the second-order derivative is

$$\frac{\partial^2 g(x,\Phi)}{\partial \phi_r \partial \phi_s} = \frac{\partial^2 g(x,\Psi)}{\partial \psi_{n+r} \partial \psi_{n+s}} + \frac{\partial^2 g(x,\Psi)}{\partial \psi_{n-r} \partial \psi_{n+s}} + \frac{\partial^2 g(x,\Psi)}{\partial \psi_{n+r} \partial \psi_{n-s}} + \frac{\partial^2 g(x,\Psi)}{\partial \psi_{n-r} \partial \psi_{n-s}}. \tag{41}$$

Invoking the triangle inequality for 1-norm and applying Corollary 15, the results follow

$$\begin{aligned}
\left\| \frac{\partial^2 F(\Phi)}{\partial \phi_r \partial \phi_s} \right\|_1 &= \left\| \mathcal{F}\left(\frac{\partial^2 g(x,\Phi)}{\partial \phi_r \partial \phi_s} \right) \right\|_1 \\
&\leq \left\| \mathcal{F}\left(\frac{\partial^2 g(x,\Psi)}{\partial \psi_{n+r} \partial \psi_{n+s}} \right) \right\|_1 + \left\| \mathcal{F}\left(\frac{\partial^2 g(x,\Psi)}{\partial \psi_{n-r} \partial \psi_{n+s}} \right) \right\|_1 \\
&\quad + \left\| \mathcal{F}\left(\frac{\partial^2 g(x,\Psi)}{\partial \psi_{n+r} \partial \psi_{n-s}} \right) \right\|_1 + \left\| \mathcal{F}\left(\frac{\partial^2 g(x,\Psi)}{\partial \psi_{n-r} \partial \psi_{n-s}} \right) \right\|_1 \\
&\leq 4 \sinh\left(2 \left\| \Phi \right\|_1 \right).
\end{aligned} \tag{42}$$

□

3.2 Matrix 1-norm estimates of DF

The following lemma characterizes the Lipschitz continuity of DF.

Lemma 16 (Lipschitz continuity of DF)**.** *For any $\delta > 0$ and any $\Phi^{(j)} \in \mathbb{R}^\infty$ with $\left\| \Phi^{(j)} \right\|_1 \leq \delta$, $j = 1, 2$, it holds that*

$$\left\| DF(\Phi^{(1)}) - DF(\Phi^{(2)}) \right\|_1 \leq C_2(\delta) \left\| \Phi^{(1)} - \Phi^{(2)} \right\|_1, \tag{43}$$

where $C_2(\delta) = 4\sinh(2\delta)$.

Proof. Using the definition of the 1-norm of infinite dimensional matrix, one has

$$\begin{aligned}
\left\| DF(\Phi^{(1)}) - DF(\Phi^{(2)}) \right\|_1 &= \max_{\|v\|_1 = 1} \left\| \left(DF(\Phi^{(1)}) - DF(\Phi^{(2)}) \right) v \right\|_1 \\
&= \max_{\|v\|_1 = 1} \left\| \sum_{k=0}^{\infty} \left(\frac{\partial F(\Phi^{(1)})}{\partial \phi_k} - \frac{\partial F(\Phi^{(2)})}{\partial \phi_k} \right) v_k \right\|_1 \\
&\leq \max_{\|v\|_1 = 1} \sum_{k=0}^{\infty} \left\| \frac{\partial F(\Phi^{(1)})}{\partial \phi_k} - \frac{\partial F(\Phi^{(2)})}{\partial \phi_k} \right\|_1 |v_k| \\
&= \max_k \left\| \frac{\partial F(\Phi^{(1)})}{\partial \phi_k} - \frac{\partial F(\Phi^{(2)})}{\partial \phi_k} \right\|_1.
\end{aligned} \tag{44}$$

For fixed k, $\Phi^{(1)}$ and $\Phi^{(2)}$, applying mean value inequality to the function

$$y(t) := \frac{\partial F}{\partial \phi_k} \left(\Phi^{(1)} + t \left(\Phi^{(2)} - \Phi^{(1)} \right) \right),$$

we have

$$\begin{aligned}
\left\|\frac{\partial F(\Phi^{(1)})}{\partial \phi_k}-\frac{\partial F(\Phi^{(2)})}{\partial \phi_k}\right\|_1 &\leq \max_{\substack{\Phi'=(1-t)\Phi^{(1)}+t\Phi^{(2)}\\0\leq t\leq 1}} \left\|\nabla\frac{\partial F(\Phi')}{\partial \phi_k}\cdot\left(\Phi^{(1)}-\Phi^{(2)}\right)\right\|_1 \\
&= \max_{\substack{\Phi'=(1-t)\Phi^{(1)}+t\Phi^{(2)}\\0\leq t\leq 1}} \left\|\sum_{l=0}^{n}\frac{\partial^2 F(\Phi')}{\partial \phi_l\partial \phi_k}\left(\Phi_l^{(1)}-\Phi_l^{(2)}\right)\right\|_1 \\
&\leq \max_{\substack{\Phi'=(1-t)\Phi^{(1)}+t\Phi^{(2)}\\0\leq t\leq 1}} \sum_{l=0}^{n}\left\|\frac{\partial^2 F(\Phi')}{\partial \phi_l\partial \phi_k}\right\|_1\left|\Phi_l^{(1)}-\Phi_l^{(2)}\right| \\
&\leq C_2(\delta)\sum_{l=0}^{n}\left|\Phi_l^{(1)}-\Phi_l^{(2)}\right| \\
&= C_2(\delta)\left\|\Phi^{(1)}-\Phi^{(2)}\right\|_1
\end{aligned} \tag{45}$$

where n is the effective length of $\Phi^{(1)}-\Phi^{(2)}$. The last inequality follows Lemma 13, and notice that $\|\Phi'\|_1=\left\|(1-t)\Phi^{(1)}+t\Phi^{(2)}\right\|_1$ is still bounded by δ. Since k can be arbitrary, the proof is completed. □

Lemma 17. *$DF(0)=2\mathbb{I}$, where $0\in\mathbb{R}^\infty$ is the vector with all elements equal to zero, and $\mathbb{I}$ is the identity matrix of infinite dimension.*

Proof. It is equivalent to show that $\frac{\partial F(0)}{\partial \phi_k}=2e_k$, where $e_k\in\mathbb{R}^\infty$ is the vector with all components equal to 0 except for the k-th component which is equal to 1. Recall Eq. (24) and notice that $\mathcal{F}(T_{2k}(x))=e_k$, we can prove that $\frac{\partial g(x,0)}{\partial \phi_k}=2T_{2k}(x)$ instead. Invoking Lemma 10, we only need to show that $\frac{\partial g(x,\Phi)}{\partial \phi_k}=2T_{2k}(x)$, where $\Phi=(0,\cdots,0)\in\mathbb{R}^{k+1}$. We know that $U(x,\Phi)=W^{2k}(x)$ as well as Eq. (30). Then direct computation gives that

$$\begin{aligned}
\frac{\partial g(x,\Phi)}{\partial \phi_k} &= \mathrm{Im}\left(\langle 0|\frac{\partial U(x,\Phi)}{\partial \phi_k}|0\rangle\right)=\mathrm{Im}\left(\langle 0|\mathrm{i}ZU(x,\Phi)+U(x,\Phi)\mathrm{i}Z|0\rangle\right) \\
&= 2\mathrm{Re}\left(\langle 0|U(x,\Phi)|0\rangle\right)=2\mathrm{Re}\left(\langle 0|W^{2k}(x)|0\rangle\right)=2T_{2k}(x).
\end{aligned} \tag{46}$$

□

By choosing $\Phi^{(2)}$ to be 0 in Lemma 16, we obtain a rough estimate about $DF(\Phi)$,

$$\|DF(\Phi)-2\mathbb{I}\|_1\leq C_2(\|\Phi\|_1)\|\Phi\|_1.$$

However, this estimate can be refined, which is given in the following lemma.

Lemma 18. *Define*

$$h(x):=\int_0^x C_2(\delta)\mathrm{d}\delta=2\cosh(2x)-2. \tag{47}$$

For any $\Phi\in\mathbb{R}^\infty$,

$$\|DF(\Phi)-2\mathbb{I}\|_1\leq h(\|\Phi\|_1). \tag{48}$$

Proof. Note that $DF(0)=2\mathbb{I}$. For a fixed Φ, for an arbitrary partition of $[0,1]$

$$0=t_0<t_1<\cdots<t_m=1,$$

the following can be obtained by invoking the triangle inequality of 1-norm and applying Lemma 16

$$\begin{aligned}\|DF(\Phi) - DF(0)\|_1 &= \left\| \sum_{k\in[m]} (DF(t_{k+1}\Phi) - DF(t_k\Phi)) \right\|_1 \\ &\leq \sum_{k\in[m]} \|DF(t_{k+1}\Phi) - DF(t_k\Phi)\|_1 \\ &\leq \sum_{k\in[m]} C_2(\|t_{k+1}\Phi\|_1)\, \|t_{k+1}\Phi - t_k\Phi\|_1 \\ &= \sum_{k\in[m]} C_2(t_{k+1}\, \|\Phi\|_1)\, \|\Phi\|_1\, |t_{k+1} - t_k|\,.\end{aligned} \tag{49}$$

Note that $\sum_{k\in[m]} C_2(\|t_{k+1}\Phi\|_1)\, \|\Phi\|_1\, |t_{k+1} - t_k|$ is a Riemann sum integrating $C_2(t\, \|\Phi\|_1)\, \|\Phi\|_1$ over $t \in [0, 1]$. Since the partition is arbitrary, one gets

$$\|DF(\Phi) - DF(0)\|_1 \leq \int_0^1 C_2(t\, \|\Phi\|_1)\, \|\Phi\|_1 \, \mathrm{d}t = \int_0^{\|\Phi\|_1} C_2(\delta)\mathrm{d}\delta = h(\|\Phi\|_1). \tag{50}$$

□

As an immediate consequence of Lemma 18, for any $\Phi \in \mathbb{R}^\infty$ with bounded norm $\|\Phi\|_1 \leq \delta$, it holds that

$$\|DF(\Phi)\|_1 \leq 2\cosh(2x) - 2 + 2 = 2\cosh(2\delta) := C_1(\delta). \tag{51}$$

We will use this upper bound on $\|DF(\Phi)\|_1$ to prove the Lipschitz continuity of F.

Corollary 19 (Lipschitz continuity of F)**.** *For any $\Phi^{(j)} \in \mathbb{R}^\infty$ with bounded norm $\left\|\Phi^{(j)}\right\|_1 \leq \delta$ where $j = 1, 2$ it holds that*

$$\left\|F(\Phi^{(1)}) - F(\Phi^{(2)})\right\|_1 \leq C_1(\delta) \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1. \tag{52}$$

Proof. Applying the mean value inequality, there exists Φ' which is some convex combination of $\Phi^{(1)}$ and $\Phi^{(2)}$ so that

$$\left\|F(\Phi^{(1)}) - F(\Phi^{(2)})\right\|_1 \leq \|DF(\Phi')\|_1 \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1. \tag{53}$$

Invoking Eq. (51) and recalling the condition $\|\Phi'\|_1 \leq \delta$, one has $\|DF(\Phi')\|_1 \leq C_1(\delta)$, which completes the proof. □

3.3 Invertibility of F in $\mathbb{R}^\infty$

According to the inverse mapping theorem and Lemma 17, we know that F is invertible near 0. We now prove a stronger result about the invertibility of F in a neighborhood of the origin. This is obtained via an upper bound on $\|F^{-1}(c)\|_1$ in terms of $\|c\|_1$. The proof of the following lemma is given in Appendix B. Since F is not an injection globally, the F^{-1} we refer to only means a continuous function whose domain is an open subset of $\mathbb{R}^\infty$ containing 0 and satisfies $F \circ F^{-1} = \mathrm{id}$.

Lemma 20 (Invertibility of F in $\mathbb{R}^\infty$)**.** *Define*

$$H(x) := \int_0^x 2 - h(t)\mathrm{d}t = 4x - \sinh(2x), \tag{54}$$

$$r_\Phi := h^{-1}(2) \approx 0.658, \tag{55}$$

and

$$r_c := H(r_\Phi) \approx 0.902. \tag{56}$$

F has an inverse map $F^{-1} : B(0, r_c) \subset \mathbb{R}^\infty \to \mathbb{R}^\infty$. Moreover, for any $c \in \mathbb{R}^\infty$ with $\|c\|_1 < r_c$, it holds that

$$\left\|F^{-1}(c)\right\|_1 \leq H^{-1}(\|c\|_1). \tag{57}$$

As a remark, the effective length of $F^{-1}(c)$ is always equal to that of c, which is implied by the proof of Lemma 20.

Corollary 21. *For any $c^{(j)} \in \mathbb{R}^\infty$ with $\left\|c^{(j)}\right\|_1 \leq \theta < r_c$ where $j = 1, 2$, define $\Phi^{(j)} := F^{-1}(c^{(j)})$ for $j = 1, 2$. It holds that*

$$\widetilde{C}(\theta) \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1 \leq \left\|c^{(1)} - c^{(2)}\right\|_1, \tag{58}$$

where $\widetilde{C}(\theta) = 2 - h(H^{-1}(\theta))$.

Proof. First, one has

$$\begin{aligned}
&\left\|c^{(1)} - c^{(2)}\right\|_1 = \left\|F(\Phi^{(1)}) - F(\Phi^{(2)})\right\|_1 \\
&= \left\|\int_0^1 DF\left(s\Phi^{(1)} + (1-s)\Phi^{(2)}\right) \cdot \left(\Phi^{(1)} - \Phi^{(2)}\right) \mathrm{d}s\right\|_1 \\
&= \left\|\int_0^1 \left(2\mathbb{I} + DF\left(s\Phi^{(1)} + (1-s)\Phi^{(2)}\right) - 2\mathbb{I}\right) \cdot \left(\Phi^{(1)} - \Phi^{(2)}\right) \mathrm{d}s\right\|_1 \\
&\geq 2\left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1 - \int_0^1 \left\|DF\left(s\Phi^{(1)} + (1-s)\Phi^{(2)}\right) - 2\mathbb{I}\right\|_1 \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1 \mathrm{d}s
\end{aligned} \tag{59}$$

Hence, there exists $t \in (0, 1)$ such that $\Phi' = t\Phi^{(1)} + (1-t)\Phi^{(2)}$ and

$$\left\|c^{(1)} - c^{(2)}\right\|_1 \geq 2\left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1 - \|DF(\Phi') - 2\mathbb{I}\|_1 \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1. \tag{60}$$

From Lemma 20, for $j = 1, 2$, one has

$$\left\|\Phi^{(j)}\right\|_1 = \left\|F^{-1}(c^{(j)})\right\|_1 \leq H^{-1}(\left\|c^{(j)}\right\|_1) \leq H^{-1}(\theta) < r_\Phi. \tag{61}$$

Then $\|\Phi'\|_1 \leq H^{-1}(\theta)$ follows by the convexity of 1-norm ball. According to Lemma 18, one has $\|DF(\Phi') - 2\mathbb{I}\|_1 \leq h(\|\Phi'\|_1) < 1$. Thus, we obtain

$$\widetilde{C}(\theta) \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1 \leq \left\|c^{(1)} - c^{(2)}\right\|_1, \tag{62}$$

where $\widetilde{C}(\theta) := 2 - h(H^{-1}(\theta))$. □

Combining Corollary 19 and Corollary 21, we obtain the following theorem.

Lemma 22 (Equivalence of distance in $\mathbb{R}^\infty$)**.** *For any $c^{(j)} \in \mathbb{R}^\infty$ with $\left\|c^{(j)}\right\|_1 \leq \theta < r_c$, $j = 1, 2$, define $\Phi^{(j)} := F^{-1}(c^{(j)})$, $j = 1, 2$. It holds that*

$$\widetilde{C}(\theta) \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1 \leq \left\|c^{(1)} - c^{(2)}\right\|_1 \leq C(\theta) \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1, \tag{63}$$

where $\widetilde{C}(\theta) = 2 - h(H^{-1}(\theta))$ and $C(\theta) = C_1(H^{-1}(\theta))$.

An immediate consequence of Lemma 22 is that: for any $c \in \mathbb{R}^\infty$ with $\|c\|_1 \leq \theta < r_c$, if we let $\Phi := F^{-1}(c)$, then

$$\widetilde{C}(\theta) \|\Phi\|_1 \leq \|c\|_1 \leq C(\theta) \|\Phi\|_1 . \tag{64}$$

To show the sharpness of r_c, we plot $\widetilde{C}(\theta)$ and $C(\theta)$ as function of θ in Fig. 1.

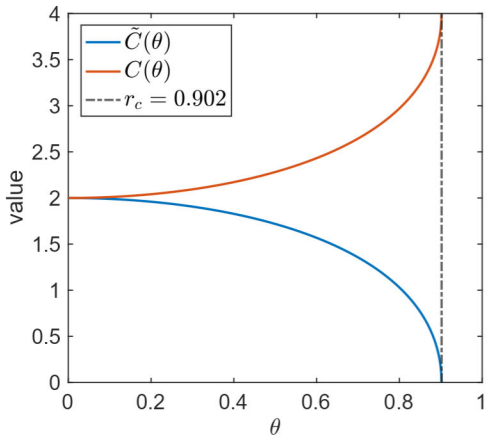


Figure 1: The plot of $\widetilde{C}(\theta)$ and $C(\theta)$ as function of θ.

According to Lemma 22, the map F preserves the distance up to a constant, i.e., F is a *quasi isometry* near the origin.

3.4 Extension to ℓ^1

In this section, we extend Lemma 20 and Lemma 22 from $\mathbb{R}^\infty$ to ℓ^1. The map $\mathcal{F}$ is a well-defined mapping from S_e (or S_o) to ℓ^1 according to

$$\mathcal{F}(f) = (c_0, c_1, \cdots). \tag{65}$$

The following theorem is a more detailed statement of Theorem 3.

Theorem 23 (Invertibility of F in ℓ^1)**.** *The map $F : \mathbb{R}^\infty \to \mathbb{R}^\infty$ can be extended to $\overline{F} : \ell^1 \to \ell^1$. Furthermore, $\overline{F}$ has an inverse map $\overline{F}^{-1} : B(0, r_c) \subset \ell^1 \to \ell^1$. For any $c \in \mathbb{R}^\infty$ with $\|c\|_1 < r_c$, it holds that*

$$\left\|\overline{F}^{-1}(c)\right\|_1 \leq H^{-1}(\|c\|_1). \tag{66}$$

Proof. Lemma 16 and Corollary 19 state that F and DF are both Lipschitz continuous. By Theorem 30 and the fact that $\mathbb{R}^\infty$ is a dense subspace of ℓ^1, F can be extended to the whole ℓ^1. The inverse mapping theorem for Banach spaces (Theorem 31), together with Theorem 30 imply that $\overline{F}$ shares the same property with F within a neighborhood of the origin. This completes the proof due to Lemma 20. □

Following the proof of Theorem 23, we can also extend Lemma 22 to ℓ^1, which states that $\overline{F}$ preserves the distance up to a constant in a neighborhood of the origin.

Theorem 24 (Equivalence of distance in ℓ^1)**.** *For any $c^{(j)} \in \ell^1$ with $\left\|c^{(j)}\right\|_1 \leq \theta < r_c$ where $j = 1, 2$, define $\Phi^{(j)} := \overline{F}^{-1}(c^{(j)})$, $j = 1, 2$. It holds that*

$$\widetilde{C}(\theta) \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1 \leq \left\|c^{(1)} - c^{(2)}\right\|_1 \leq C(\theta) \left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1, \tag{67}$$

where $\widetilde{C}(\theta) = 2 - h(H^{-1}(\theta))$ and $C(\theta) = C_1(H^{-1}(\theta))$.

With the help of the inverse mapping theorem of Banach spaces, the proof of this theorem follows the same idea as Lemma 22.

Now we are ready to give a positive answer to the first question raised in Section 1.1. Let $\Psi^{(d)}$ denote the symmetric phase factors and $\Phi^{(d)}$ denote the corresponding reduced phase factors. Although the solution to $F(\Phi) = \mathcal{F}(f^{(d)})$ may not be unique, Theorem 23 allows us to choose one $\Phi^{(d)}$ such that $\left\|\Phi^{(d)}\right\|_1 < r_\Phi$ as long as $\left\|\mathcal{F}(f^{(d)})\right\|_1 < r_c$. Assume that $c := \mathcal{F}(f)$ satisfies $\|c\|_1 < r_c$ as well. Then $\lim_{d\to\infty} f^{(d)} = f$ implies that $\{\mathcal{F}(f^{(d)})\}$ converges to c with respect to the vector 1-norm. The convergence of $\{\Phi^{(d)}\}$ to $\overline{F}^{-1}(c)$ follows by applying the equivalence of distance in Theorem 24. Theorem 23 provides a sufficient condition to the existence of the solution to $\overline{F}(\Phi) = c$. We would like to emphasize that similar to the case of polynomials, the solution might not be unique in ℓ^1.

3.5 Structure of DF

In this section, we focus on the structure of the Jacobian matrix DF. Given any $\Phi \in \mathbb{R}^\infty$, we let n be its effective length. According to Eq. (24), the k-th column of $DF(\Phi)$, i.e., $\frac{\partial F(\Phi)}{\partial \phi_k}$, is the Chebyshev-coefficient vector of $\frac{\partial g(x,\Phi)}{\partial \phi_k}$. Direct computation shows

$$\frac{\partial F(\Phi)}{\partial \phi_k} = F(\Phi + \frac{\pi}{4} e_k) - F(\Phi - \frac{\pi}{4} e_k), \quad \forall k \in \mathbb{N}, \Phi \in \mathbb{R}^\infty. \tag{68}$$

Here, $e_k \in \mathbb{R}^\infty$ is the vector with all components equal to 0 except for the k-th component which is equal to 1. As a reminder, when we refer to the component of a vector or matrix, the index begins with 0.

We consider the even case for simplicity. For $k < n$, according to Eq. (68), $\frac{\partial g(x,\Phi)}{\partial \phi_k}$ is a polynomial of degree at most $2n - 2$. Thus the components $\left[\frac{\partial F(\Phi)}{\partial \phi_k}\right]_j = 0$ for any $j \geq n$. For $k \geq n$, $\frac{\partial g(x,\Phi)}{\partial \phi_k}$ is a polynomial of degree at most $2k$, and then $\left[\frac{\partial F(\Phi)}{\partial \phi_k}\right]_j = 0$ for any $j > k$. Therefore $DF(\Phi)$ takes the form

$$\begin{pmatrix} D_1 & D_2 \\ 0 & D_3 \end{pmatrix}. \tag{69}$$

Here, D_1 is a matrix of size $n \times n$ and D_3 is an upper triangular matrix of infinite dimension. As for matrix D_2, the number of rows is n, while the number of columns is infinite.

When $DF(\Phi)$ is invertible, the inverse takes the form

$$DF(\Phi)^{-1} = \begin{pmatrix} D_1^{-1} & -D_1^{-1} D_2 D_3^{-1} \\ 0 & D_3^{-1} \end{pmatrix}. \tag{70}$$

After extending F to ℓ^1, we can characterize the invertibility of matrix $D\overline{F}$ in ℓ^1.

Lemma 25. *For any $\Phi \in \ell^1$ with $\|\Phi\|_1 < r_\Phi$, $D\overline{F}(\Phi)$ is invertible and*

$$\left\|D\overline{F}(\Phi)^{-1}\right\|_1 \leq \frac{1}{2 - h(\|\Phi\|_1)}. \tag{71}$$

Proof. When $\|\Phi\|_1 < h^{-1}(2)$, we have $\left\|\frac{1}{2}D\overline{F}(\Phi) - \mathbb{I}\right\|_1 \leq \frac{1}{2}h(\|\Phi\|_1) < 1$. Applying Theorem 29 in the Banach space ℓ^1, $D\overline{F}(\Phi)$ is invertible and

$$\left\|D\overline{F}(\Phi)^{-1}\right\|_1 = \frac{1}{2}\left\|\left(\frac{1}{2}D\overline{F}(\Phi)\right)^{-1}\right\|_1 \leq \frac{\frac{1}{2}}{1 - \frac{1}{2}h(\|\Phi\|_1)} = \frac{1}{2 - h(\|\Phi\|_1)}.$$

□

4 Decay properties

As an immediate consequence of Theorem 24, we now prove the decay properties of the reduced phase factors $\Phi \in \ell^1$ (Theorem 4) with an explicit characterization of the constant C.

Proof of Theorem 4. Let $\Phi^{(n)} = (\phi_0, \ldots, \phi_n, 0, 0, \ldots)$, and $\theta = \|c\|_1$ in Theorem 24, then we get

$$\begin{aligned} C_1(H^{-1}(\|c\|_1)) \sum_{k>n} |\phi_k| &= C_1(H^{-1}(\|c\|_1)) \left\|\Phi - \Phi^{(n)}\right\|_1 \\ &\geq \left\|\overline{F}(\Phi) - \overline{F}\left(\Phi^{(n)}\right)\right\|_1 = \left\|c - \overline{F}\left(\Phi^{(n)}\right)\right\|_1 \\ &\geq \sum_{k>n} |c_k|, \end{aligned} \tag{72}$$

where the last inequality follows the fact that $\overline{F}(\Phi^{(n)})$ is zero after the n-th component. Similarly, we choose $c^{(n)} = (c_0, \ldots, c_n, 0, 0, \ldots)$, and Theorem 24 gives

$$\begin{aligned} \frac{1}{2 - h(H^{-1}(\|c\|_1))} \sum_{k>n} |c_k| &= \frac{1}{2 - h(H^{-1}(\|c\|_1))} \left\|c - c^{(n)}\right\|_1 \\ &\geq \left\|\overline{F}^{-1}(c) - \overline{F}^{-1}\left(c^{(n)}\right)\right\|_1 = \left\|\Phi - \overline{F}^{-1}\left(c^{(n)}\right)\right\|_1 \\ &\geq \sum_{k>n} |\phi_k|, \end{aligned} \tag{73}$$

where the last inequality follows the fact that $\overline{F}^{-1}(c^{(n)})$ is zero after the n-th component. This completes the proof with $C' = \frac{1}{C_1(H^{-1}(\|c\|_1))}$ and $C = \frac{1}{2 - h(H^{-1}(\|c\|_1))}$. □

If $\sum_{k>n} |c_k| = O(n^{-\alpha})$ for some $\alpha > 0$, then $\sum_{k>n} |\phi_k| = O(n^{-\alpha})$. If c_k decays super-algebraically or exponentially, so does the reduced phase factors. Therefore the tail decay of the reduced phase factors is determined by the smoothness of the target function.

5 Fixed-point iteration for finding phase factors

Algorithm 1 is a very simple iterative algorithm based on fixed-point iteration for finding phase factors. Numerical results suggest that this algorithm is quite robust, starting from a *fixed* initial point $\Phi^0 = 0$ (or $\Phi^1 = \frac{1}{2}c$ according to Algorithm 1). We emphasize that the choice of the initial guess is important, and other initial points may make the algorithm diverge. Based on the developments in Section 3, we prove that Algorithm 1 converges linearly in ℓ^1, and describe the computational complexity.

5.1 Convergence

To prove the convergence of Algorithm 1, it is sufficient to prove that

$$G(\Phi) := \Phi - \frac{1}{2}F(\Phi) + \frac{1}{2}F(\Phi^\star) \tag{74}$$

is a contraction map in a neighborhood of $\Phi^\star$. $\Phi^\star$ denotes the desired set of reduced phase factors. The contraction property of G follows the observation that the Jacobian matrix $DG(\Phi^\star) = \mathbb{I} - \frac{1}{2}DF(\Phi^\star) = \frac{1}{2}\left(DF(0) - DF(\Phi^\star)\right)$ would be small when $\|\Phi^\star\|_1$ is sufficiently small according to Lemma 18. We define a function

$$\gamma(r) := \frac{1}{2}\int_0^1 h\left(r + s\left(\frac{1}{2}\sinh(2r) - r\right)\right)\mathrm{d}s, \tag{75}$$

and the following constants

$$\widetilde{r}_\Phi := \frac{1}{2}\operatorname{arcsinh}(\operatorname{arccosh}(2)) \approx 0.544 \tag{76}$$

$$\widetilde{r}_c := H(\widetilde{r}_\Phi) \approx 0.861 \tag{77}$$

$$\widetilde{\gamma} := \gamma(\widetilde{r}_\Phi) \approx 0.8189. \tag{78}$$

We will also use the fact that $\widetilde{r}_\Phi < \frac{1}{2}\operatorname{arccosh}(2) = r_\Phi \approx 0.658$.

Lemma 26. *If $\Phi^\star$ satisfies $c = \overline{F}(\Phi^\star)$ and $\|\Phi^\star\|_1 \leq \widetilde{r}_\Phi$, G is a contraction map in the open ball $B := B\left(\Phi^\star, \left\|\frac{1}{2}F(\Phi^\star) - \Phi^\star\right\|_1\right)$.*

Proof. First, we give an upper bound on the radius of ball B.

$$\begin{aligned}
\left\|\frac{1}{2}F(\Phi^\star) - \Phi^\star\right\|_1 &= \frac{1}{2}\left\|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left(F(t\Phi^\star) - 2t\Phi^\star\right)\mathrm{d}t\right\|_1 \\
&= \frac{1}{2}\left\|\int_0^1 \left(DF(t\Phi^\star) - 2\mathbb{I}\right)\cdot\Phi^\star \mathrm{d}t\right\|_1 \\
&\leq \frac{1}{2}\int_0^1 \left\|DF\left(t\Phi^\star\right) - 2\mathbb{I}\right\|_1 \left\|\Phi^\star\right\|_1 \mathrm{d}t \qquad \text{(use Lemma 18)} \\
&\leq \frac{1}{2}\int_0^{\|\Phi^\star\|_1} h(t)\mathrm{d}t = \frac{1}{2}\sinh\left(2\left\|\Phi^\star\right\|_1\right) - \left\|\Phi^\star\right\|_1.
\end{aligned} \tag{79}$$

Thus, for any $\Phi \in B$, we have the following estimate

$$\begin{aligned}
\|\Phi\|_1 &< \|\Phi^\star\|_1 + \left\|\frac{1}{2}F(\Phi^\star) - \Phi^\star\right\|_1 \leq \frac{1}{2}\sinh\left(2\left\|\Phi^\star\right\|_1\right) \\
&\leq \frac{1}{2}\sinh(2\widetilde{r}_\Phi) = \frac{1}{2}\operatorname{arccosh}(2) = r_\Phi.
\end{aligned} \tag{80}$$

Then we can use Lemma 18 again to conclude

$$\|DG(\Phi)\|_1 = \left\|\mathbb{I} - \frac{1}{2}DF(\Phi)\right\|_1 \leq \frac{1}{2}h(\|\Phi\|_1) < \frac{1}{2}h(r_\Phi) = 1. \tag{81}$$

This means G is a contraction map in the ball B and $\Phi^\star$ is its only fixed point. □

Lemma 27. *If $\Phi^\star$ satisfies $c = \overline{F}(\Phi^\star)$ and $\|\Phi^\star\|_1 \leq \widetilde{r}_\Phi$, Algorithm 1 converges Q-linearly to $\Phi^\star$. The rate of convergence is bounded by $\frac{1}{2}h(\|\Phi^\star\|_1)$, i.e.,*

$$\lim_{t\to\infty}\frac{\left\|\Phi^{t+1} - \Phi^\star\right\|_1}{\left\|\Phi^t - \Phi^\star\right\|_1} \leq \frac{1}{2}h(\|\Phi^\star\|_1). \tag{82}$$

Here Φ^t is the set of reduced phase factors in the t-th iteration step. Furthermore, for every $t \geq 1$,

$$\frac{\left\|\Phi^{t+1} - \Phi^\star\right\|_1}{\left\|\Phi^t - \Phi^\star\right\|_1} \leq \gamma(\left\|\Phi^\star\right\|_1). \tag{83}$$

In particular, $\gamma(\left\|\Phi^\star\right\|_1) \leq \widetilde{\gamma} \approx 0.8189$.

Proof. Lemma 26 guarantees the convergence of Algorithm 1 as long as $\Phi^t \in B$. However, we note that $\Phi^1 = \frac{1}{2}F(\Phi^\star)$ lies on the boundary of B. Hence, we need a finer estimation.

$$\begin{aligned}
\left\|\Phi^{t+1} - \Phi^\star\right\|_1 &= \left\|G(\Phi^t) - G(\Phi^\star)\right\|_1 \\
&= \left\|\int_0^1 DG\left(s\Phi^t + (1-s)\Phi^\star\right) \cdot \left(\Phi^t - \Phi^\star\right) \mathrm{d}s\right\|_1 \\
&\leq \int_0^1 \left\|DG\left(s\Phi^t + (1-s)\Phi^\star\right)\right\|_1 \left\|\Phi^t - \Phi^\star\right\|_1 \mathrm{d}s \\
&\leq \frac{1}{2}\left\|\Phi^t - \Phi^\star\right\|_1 \int_0^1 h\left(\left\|s\Phi^t + (1-s)\Phi^\star\right\|_1\right) \mathrm{d}s \\
&\leq \frac{1}{2}\left\|\Phi^t - \Phi^\star\right\|_1 \int_0^1 h\left(s\left\|\Phi^t\right\|_1 + (1-s)\left\|\Phi^\star\right\|_1\right) \mathrm{d}s
\end{aligned} \tag{84}$$

The last inequality follows that h is monotonic increasing on $[0, \infty)$.

By replacing the "$<$" by "$\leq$" in Eq. (80), we get $\left\|\Phi^1\right\|_1 \leq r_\Phi$. Hence,

$$\begin{aligned}
\left\|\Phi^2 - \Phi^\star\right\|_1 &\leq \frac{1}{2}\left\|\Phi^1 - \Phi^\star\right\|_1 \int_0^1 h\left(sr_\Phi + (1-s)\widetilde{r}_\Phi\right) \mathrm{d}s \\
&\leq \widetilde{\gamma}\left\|\Phi^1 - \Phi^\star\right\|_1.
\end{aligned} \tag{85}$$

Note that $\widetilde{\gamma} \approx 0.8189$ implies that $\Phi^2 \in B$. According to Lemma 26, we know that the rate of convergence is bounded by $\left\|DG(\Phi^\star)\right\|_1 \leq \frac{1}{2}h(\left\|\Phi^\star\right\|_1)$.

Furthermore, for any $t \geq 1$ it holds that

$$\begin{aligned}
\left\|\Phi^{t+1} - \Phi^\star\right\|_1 &\leq \frac{1}{2}\left\|\Phi^t - \Phi^\star\right\|_1 \int_0^1 h\left(\frac{s}{2}\sinh(2\left\|\Phi^\star\right\|_1) + (1-s)\left\|\Phi^\star\right\|_1\right) \mathrm{d}s \\
&= \left\|\Phi^t - \Phi^\star\right\|_1 \gamma(\left\|\Phi^\star\right\|_1).
\end{aligned} \tag{86}$$

Here, we use the result that for any $\Phi \in B$, $\left\|\Phi\right\|_1 \leq \frac{1}{2}\sinh(2\left\|\Phi^\star\right\|_1)$ from Eq. (80). This proves the lemma. □

As a remark, both the theoretical analysis and numerical results suggest that the region in which Algorithm 1 converges should be larger than B. We now prove Theorem 6 (i) with an explicit characterization of the constant C.

Proof of Theorem 6 (i). Eq. (57) implies that $\Phi^\star := F^{-1}(c)$ satisfies $\left\|\Phi^\star\right\|_1 < \widetilde{r}_\Phi$, given $\left\|c\right\|_1 < H(\widetilde{r}_\Phi)$. According to Eq. (79), we get

$$\begin{aligned}
\left\|\Phi^1 - \Phi^\star\right\|_1 = \left\|\frac{1}{2}F(\Phi^\star) - \Phi^\star\right\|_1 &\leq \frac{1}{2}\sinh(2\left\|\Phi^\star\right\|_1) - \left\|\Phi^\star\right\|_1 \\
&\leq \frac{1}{2}\sinh(2\widetilde{r}_\Phi) - \widetilde{r}_\Phi = r_\Phi - \widetilde{r}_\Phi.
\end{aligned} \tag{87}$$

Here we use the fact that $\frac{1}{2}\sinh(2x) - x$ is monotonic increasing on $[0, \infty)$. Applying Lemma 27, we obtain the error estimate

$$\left\|\Phi^t - \Phi^\star\right\|_1 \leq (r_\Phi - \widetilde{r}_\Phi)\widetilde{\gamma}^{t-1}, \quad \widetilde{\gamma} \approx 0.8189, \quad t \geq 1. \tag{88}$$

This finishes the proof of Theorem 6 (i), and the constant C is $(r_\Phi - \widetilde{r}_\Phi)/\widetilde{\gamma} \approx 0.1393$. □

5.2 Complexity

In this subsection, we discuss the complexity of Algorithm 1. For any target function satisfying $\|c\|_1 < \widetilde{r}_c$, to reach the ℓ^1-error tolerance ϵ, the number of iterations is at most

$$\left\lceil \log\left(\frac{1}{(r_\Phi - \widetilde{r}_\Phi)\epsilon}\right) / \log \widetilde{\gamma} \right\rceil. \tag{89}$$

That means the number of iterations is $\mathcal{O}(\log \frac{1}{\epsilon})$, and the upper bound of the number of iterations is independent of the target function and the effective length of Φ. Therefore, we only need to analyze the complexity of implementing F. On a computer we can only perform operations for matrices of finite sizes. For any set of phase reduced phase factors Φ whose effective length is $\widetilde{d}$, let $\mathfrak{G}(\Phi) = (g(x_0, \Phi), \cdots, g(x_{2d}, \Phi))^\top$. Here, $d = 2\widetilde{d} - 2$ if the target polynomial is even, and $d = 2\widetilde{d} - 1$ if odd. The Chebyshev node is $x_j = \cos(\frac{2\pi j}{2d+1})$.

Observe that for $k, l = 0, \cdots, d$,

$$\begin{aligned}
\mathrm{Re}\left(\sum_{j=0}^{2d} T_k(x_j) e^{-\mathrm{i}\frac{2\pi}{2d+1}lj}\right) &= \sum_{j=0}^{2d} \cos\left(\frac{2\pi k}{2d+1}j\right)\cos\left(\frac{2\pi l}{2d+1}j\right) \\
&= \frac{1}{2}\sum_{j=0}^{2d} \cos\left(\frac{2\pi(k+l)}{2d+1}j\right) + \frac{1}{2}\sum_{j=0}^{2d} \cos\left(\frac{2\pi(k-l)}{2d+1}j\right) \\
&= \frac{2d+1}{2}\delta_{kl}(\delta_{k0}+1).
\end{aligned} \tag{90}$$

Hence the Chebyshev-coefficient vector can be evaluated by applying the fast Fourier transform (FFT) to $\mathfrak{G}(\Phi)$. The output from applying FFT to $\mathfrak{G}(\Phi)$ is a vector of length $2d+1$, denoted as v, satisfying

$$\mathrm{Re}(v_0) = (2d+1)\widetilde{c}_0, \quad \mathrm{Re}(v_j) = \frac{2d+1}{2}\widetilde{c}_j, j = 1, \cdots, d \tag{91}$$

where $\widetilde{c}_j$, $j = 0, \cdots, d$, are the Chebyshev coefficients of $g(x, \Phi)$ with respect to T_j. Recall that $F(\Phi)$ is the Chebyshev-coefficient vector of $g(x, \Phi)$ which is either $(\widetilde{c}_0, \widetilde{c}_2, \cdots, \widetilde{c}_d)$ or $(\widetilde{c}_1, \widetilde{c}_3, \cdots, \widetilde{c}_d)$ depending on the parity of d.

For completeness, the procedure for computing $F(\Phi)$ is given in Algorithm 2. The cost of evaluating $\mathfrak{G}(\Phi)$ is $\mathcal{O}(d^2)$, and the cost of FFT is $\mathcal{O}(d \log d)$. Therefore, the overall time complexity is $\mathcal{O}(d^2 \log \frac{1}{\epsilon})$. This concludes the proof of Theorem 6 (ii).

Algorithm 1 is numerically stable, in the sense that the number of bits required in the computation is $\mathcal{O}(\mathrm{polylog}(d/\epsilon))$[3]. To show the numerical stability, we consider the model of finite precision arithmetic (see standard axioms in [14]) and denote by u the unit roundoff error. In each iteration step, the rounding error when evaluating $\mathfrak{G}(\Phi)$ is $\mathcal{O}(d^2 u)$. The rounding error occurred in FFT is $\mathcal{O}(\log(d)u)$ [23]. So the total rounding error accumulated in each iteration step remains $\mathcal{O}(d^2 u)$. Therefore the number of bits required by Algorithm 1 is $\mathcal{O}\left(\log\left(\frac{d\log(1/\epsilon)}{\epsilon}\right)\right)$.

6 Numerical results

We present a number of tests to demonstrate the efficiency of the fixed-point iteration (FPI) algorithm (Algorithm 1). All numerical tests are performed on a 6-core Intel Core

[3]Generically, the number of bits cannot be rigorously bounded by a constant in a proof. Most numerically stable algorithms can be robustly implemented using fixed double precision arithmetic operations in practice.

Algorithm 2 Compute $F(\Phi)$.

Input: Reduced phase factors Φ, parity, and its effective length $\widetilde{d}$.
if parity is even **then**
  Set $d = 2\widetilde{d} - 2$.
else
  Set $d = 2\widetilde{d} - 1$.
end if
Initialize $\mathbf{g} = (0, 0, \cdots) \in \mathbb{R}^{2d+1}$.
Evaluate $\mathbf{g}_j \leftarrow g(x_j, \Phi), x_j = \frac{2\pi j}{2d+1}, j = 0, \cdots, 2d+1$.
Compute $v_l \leftarrow \mathrm{Re}\left(\sum_{j=0}^{2d-1} \mathbf{g}_j e^{-\mathrm{i}\frac{2\pi}{2d+1}lj}\right), l = 0, \ldots, d$ using FFT.
if parity is even **then**
  $F(\Phi) \leftarrow \frac{2}{2d+1}(\frac{v_0}{2}, v_2, v_4, \cdots, v_d)$.
else
  $F(\Phi) \leftarrow \frac{2}{2d+1}(v_1, v_3, v_5, \cdots, v_d)$.
end if
Output: $F(\Phi)$.

i7 processor at 2.60 GHz with 16 GB of RAM. Our method is implemented in MATLAB R2019a.

The target function is $f(x) = e^{-\mathrm{i}\tau x}$, which has applications in Hamiltonian simulation. It can be expanded using the Jacobi-Anger expansion as

$$e^{-\mathrm{i}\tau x} = J_0(\tau) + 2\sum_{k \text{ even}} (-1)^{k/2} J_k(\tau) T_k(x) + 2\mathrm{i}\sum_{k \text{ odd}} (-1)^{(k-1)/2} J_k(\tau) T_k(x), \tag{92}$$

where J_k's are the Bessel functions of the first kind.

We use Algorithm 1 to find the phase factors respectively for the even part

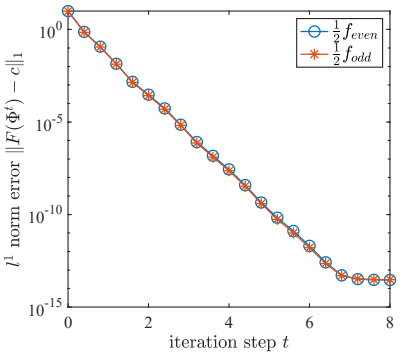
$$f_{\text{even}}(x) := J_0(\tau) + 2\sum_{k \text{ even}, k<d} (-1)^{k/2} J_k(\tau) T_k(x), \tag{93}$$

and the odd part

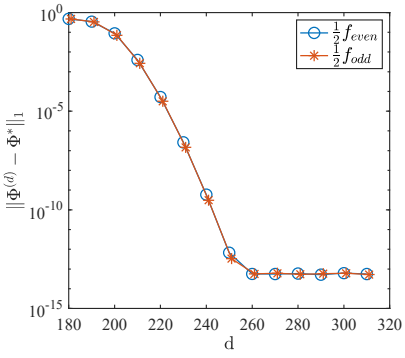
$$f_{\text{odd}}(x) := 2\sum_{k \text{ odd}, k<d} (-1)^{(k-1)/2} J_k(\tau) T_k(x). \tag{94}$$

We choose $d = 1.4\,|\tau| + \log(1/\epsilon_0)$, where $\epsilon_0 = 10^{-14}$. Since the target polynomial should be bounded by 1, to ensure numerical stability, we scale $f_{\text{even}}(x)$ and $f_{\text{odd}}(x)$ by a factor of $\frac{1}{2}$. Then we use Algorithm 1 to find phase factors for target polynomials, $\frac{1}{2}f_{\text{even}}(x)$ and $\frac{1}{2}f_{\text{odd}}(x)$, respectively. Fig. 2a displays the corresponding residual error, $\left\|F(\Phi^t) - c\right\|_1$ with $\tau = 1000$, where Φ^t is the set of reduced phase factors at the t-th iteration. Note that $\|c\|_1$ is very large, and is equal to 9.8609 and 9.7403 for the even and odd case, respectively. Nonetheless, Algorithm 1 converges starting from the fixed initial guess.

In Fig. 2b, we demonstrate that for a fixed target function, as the polynomial degree d increases, the corresponding set of reduced phase factors $\Phi^{(d)}$ indeed converges to some $\Phi^\star$. We still take $\frac{1}{2}f_{\text{even}}(x)$ and $\frac{1}{2}f_{\text{odd}}(x)$ as examples. We choose $\tau = 200$, and $\Phi^\star$ are approximately computed by setting the degrees of $\frac{1}{2}f_{\text{even}}(x)$ and $\frac{1}{2}f_{\text{odd}}(x)$ to 312 and 313 respectively. We truncate $\frac{1}{2}f_{\text{even}}(x)$ to even polynomials with degree $d = 180, 190, \cdots, 310$. Similarly, we also truncate $\frac{1}{2}f_{\text{odd}}(x)$ to odd polynomials with degree $d = 181, 191, \cdots, 311$. Then we use Algorithm 1 to compute the corresponding reduced phase factors, $\Phi^{(d)}$. We would like to emphasize that to our knowledge, only the optimization based method is able to find $\Phi^{(d)}$ that has a well defined limit as $d \to \infty$.



(a) The residual error after each iteration by using Algorithm 1 to determine phase factors for $\frac{1}{2}f_{\text{even}}(x)$ and $\frac{1}{2}f_{\text{odd}}(x)$ respectively, where $\tau = 1000$.



(b) The convergence of $\Phi^{(d)}$ as the degree of polynomial approximating target function increases. The target functions are $\frac{1}{2}f_{\text{even}}(x)$ and $\frac{1}{2}f_{\text{odd}}(x)$ of degree 312 and 313 respectively, where $\tau = 200$.

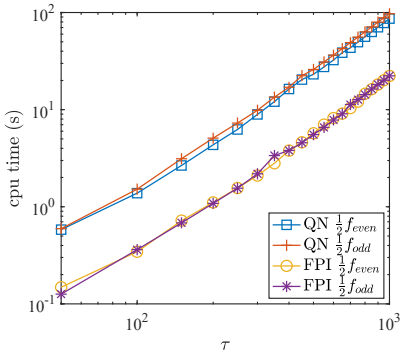
Figure 2: The performance of the fixed-point iteration (FPI) algorithm (Algorithm 1) to find phase factors for $\frac{1}{2}f_{\text{even}}(x)$ and $\frac{1}{2}f_{\text{odd}}(x)$.

We also compare the performance of Algorithm 1 with the quasi-Newton method implemented in [8] on $\frac{1}{2}f_{\text{even}}(x)$ and $\frac{1}{2}f_{\text{odd}}(x)$ with $\tau = 50, 100, \cdots, 1000$. The stopping criteria for Algorithm 1 is $\|F(\Phi^t) - c\|_1 \leq \epsilon$. As for quasi-Newton method, the iteration stops when

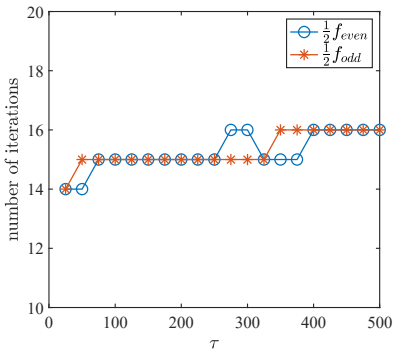
$$L(\Phi) := \max_{j=1,\cdots,\widetilde{d}} \left| g(x_j, \Phi^t) - f_{\text{target}}(x_j) \right| \leq \epsilon. \tag{95}$$

where f_{target} is the target polynomial and $x_j = \cos\left(\frac{(2j-1)\pi}{4\widetilde{d}}\right)$, $j = 1, \cdots, \widetilde{d}$ are the positive roots of the Chebyshev polynomial $T_{2\widetilde{d}}(x)$. For numerical demonstration, we choose $\epsilon = 10^{-12}$. The results of comparison are displayed in Fig. 3. Since $L(\Phi) \leq \|F(\Phi) - c\|_1$ for any Φ, the stopping criteria for Algorithm 1 is actually tighter than that in quasi-Newton method. Thus Fig. 3a implies that Algorithm 1 converges faster than quasi-Newton method in this example. The degree of the target polynomial linearly increases as the value of τ increases, and the CPU time scales asymptotically as $\mathcal{O}(\tau^2) = \mathcal{O}(d^2)$. In Fig. 3b, we present the number of iterations required using FPI to find phase factors for $\frac{1}{2}f_{\text{even}}(x)$ and $\frac{1}{2}f_{\text{odd}}(x)$. Fig. 3b indicates that the number of iterations is almost independent of the degree of the target polynomial.

Finally, we demonstrate the decay of phase factors in Fig. 4. In the first example, we truncate the series expansion of $f(x) = 0.8\,|x|^3$ in terms of Chebyshev polynomials of the first kind up to degree $d = 1000$ and use Algorithm 1 to find the corresponding reduced phase factors. The third order derivative of $f(x)$ is discontinuous. In Fig. 4a, we plot the magnitude of its Chebyshev-coefficient vector, as well as the reduced phase factors obtained by Algorithm 1. Here, the 1-norm of Chebyshev coefficients is about 0.8149, which is bounded by r_c. Fig. 4a shows that the reduced phase factors decay away from the center with an algebraic decay rate around 4, which matches the decay rate of Chebyshev coefficients. This also agrees with our theoretical results in Theorem 4. In the second example, we choose $\frac{1}{2}f_{\text{odd}}(x)$ as target polynomial and present the magnitude of both Chebyshev-coefficient vector and the corresponding reduced phase factors in Fig. 4b. The 1-norm of Chebyshev coefficients is around 3.2332, which exceeds the norm condition in Theorem 4. Nonetheless, the decay of the tail of the phase factors closely matches that of the Chebyshev-coefficient vector.

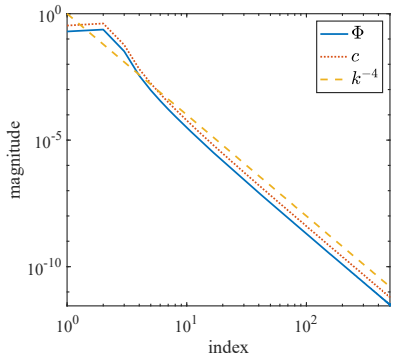


(a) CPU time(s) required using FPI and QSPPACK to find QSP phase factors. The slopes of the red and purple lines are about 2, representing CPU time = const·τ^2.

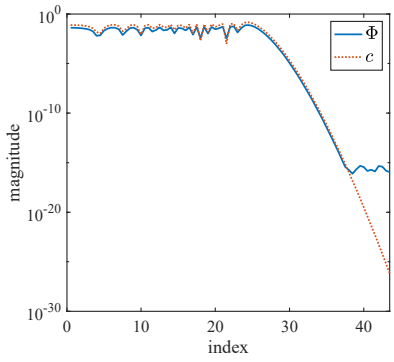


(b) Number of iterations required using FPI to find QSP phase factors. The number of iterations required varies between 14 and 16 and has no direct relation to τ.

Figure 3: Comparison of the performance of the fixed-point iteration (FPI) algorithm (Algorithm 1) with the quasi-Newton (QN) method in [8] to find QSP phase factors for $\frac{1}{2}f_{\text{even}}(x)$ and $\frac{1}{2}f_{\text{odd}}(x)$ with $\tau = 50, 100, 150, \cdots, 1000$. The error tolerance is $\epsilon = 10^{-12}$.



(a) The target polynomial is $\sum_{k=0}^{1000} c_k T_{2k}(x)$, where c_k is the Chebyshev coefficient of $0.8\,|x|^3$ w.r.t. T_{2k}. The slopes of blue and red curves are about -4, representing $|\phi_k| \approx \text{const} \cdot k^{-4}$ and $|c_k| \approx \text{const} \cdot k^{-4}$.



(b) The target polynomial is $\frac{1}{2}f_{\text{odd}}(x)$ with $\tau = 100$ of degree 173.

Figure 4: Magnitude of the Chebyshev-coefficient vector c and the corresponding reduced phase factors Φ.

7 Discussion

The question of infinite quantum processing (Problem 2) asks whether there is a set of phase factors of infinite length in ℓ^1 for representing target polynomials expressed as an infinite polynomial series. Theorem 3 provides a positive answer, when the 1-norm of the Chebyshev coefficients c of the target function is upper bounded by a constant r_c. While it is always possible to rescale the target function to satisfy the constraint, the constraint may be violated for many target functions without rescaling. For instance, in the Hamiltonian simulation problem, we have $\|c\|_1 = \mathcal{O}(\tau)$, where τ is the simulation time and can be arbitrarily large. Numerical results in Section 6 indicate that both the fixed point algorithm and the decay properties persist even for large $\|c\|_1$. Therefore it may be possible to significantly relax the condition $\|c\|_1 \leq r_c$.

Ref. [25] shows that the structure of phase factors using the infinity norm of the target polynomial $f^{(d)}$, and proves the convergence of the projected gradient method when $\left\|f^{(d)}\right\|_\infty \leq Cd^{-1}$. Using the tools developed in this paper, this condition can also be relaxed to $d^{-1/2}$, but the d-dependence may not be removed by the techniques in this paper alone.

It is worth noting that the conditions on $\left\|f^{(d)}\right\|_\infty$ and $\|c\|_1$ are generally unrelated, i.e., neither implies the other. It is of particular interest to develop a method for computing phase factors that provably converges in the limit $\|f\|_\infty \to 1$, which is called the fully coherent limit [19]. This limit is important for the performance of certain amplitude amplification algorithms [12, 10] and Hamiltonian simulation problems [19].

The decay properties of the reduced phase factors may also have some practical implications. For instance, for approximating smooth functions, the phase factors towards both ends of the quantum circuit are very close to being a constant. This may facilitate the compilation and error mitigation of future applications using the quantum singular value transformation.

Note: Since the initial posting of this paper, there have been significant algorithmic and theoretical advancements in this area. We recently proposed an algorithm based on Newton's method [6], which demonstrates rapid and robust numerical convergence for all functions admitting a QSP representation, including those where $\|f\|_\infty \to 1$. [1] provides the first provably stable algorithm for finding phase factors for all polynomials with $\|f\|_\infty < 1$. This is due to the recent theoretical development connecting the iQSP problem with nonlinear Fourier analysis first established in [2]. The analysis of [2] provides a positive solution to Problem 2 for functions with $\|f\|_\infty < 1/\sqrt{2}$, and [1] generalizes this to include nearly all functions (polynomials and beyond) that admit a QSP representation.

Another recent development is the introduction of generalized quantum signal processing (GQSP) [21], which relaxes the parity constraint on the target polynomial. A promising direction for future research is to investigate the potential for extending GQSP to represent smooth functions.

Acknowledgment

This work was supported by the U.S. Department of Energy under the Quantum Systems Accelerator program under Grant No. DE-AC02-05CH11231 (Y.D. and L.L.), and by the Challenge Institute for Quantum Computation (CIQC) funded by National Science Foundation (NSF) through grant number OMA-2016245 (J.W.). L.L. is a Simons Investigator in Mathematics. The authors thank discussions with András Gilyén and Felix Otto.

References

[1] M. Alexis, L. Lin, G. Mnatsakanyan, C. Thiele, and W. Jiasu. Infinite quantum signal processing for arbitrary szegö functions. 2024. arXiv:2407.05634.

[2] M. Alexis, G. Mnatsakanyan, and C. Thiele. Quantum signal processing and nonlinear fourier analysis. *Revista Matemática Complutense*, 37, 06 2024. doi:10.1007/s13163-024-00494-5.

[3] R. Chao, D. Ding, A. Gilyén, C. Huang, and M. Szegedy. Finding Angles for Quantum Signal Processing with Machine Precision. 2020. arXiv:2003.02831.

[4] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su. Toward the first quantum simulation with quantum speedup. *Proc. Nat. Acad. Sci.*, 115(38):9456–9461, 2018. doi:10.1073/pnas.1801723115.

[5] Y. Dong and L. Lin. Random circuit block-encoded matrix and a proposal of quantum linpack benchmark. *Phys. Rev. A*, 103:062412, Jun 2021. doi:10.1103/PhysRevA.103.062412.

[6] Y. Dong, L. Lin, H. Ni, and J. Wang. Robust iterative method for symmetric quantum signal processing in all parameter regimes. *SIAM J. Sci. Comput.*, 46:A2951–A2971, 2024. doi:10.1137/23M1598192.

[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, Oct 2022. doi:10.1103/PRXQuantum.3.040305.

[8] Y. Dong, X. Meng, K. B. Whaley, and L. Lin. Efficient phase factor evaluation in quantum signal processing. *Phys. Rev. A*, 103:042419, Apr 2021. doi:10.1103/PhysRevA.103.042419.

[9] Y. Dong, K. B. Whaley, and L. Lin. A quantum hamiltonian simulation benchmark. *npj Quantum Information*, 8, Nov 2022. doi:10.1038/s41534-022-00636-x.

[10] D. Fang, L. Lin, and Y. Tong. Time-marching based quantum solvers for time-dependent linear differential equations. *Quantum*, 7:955, Mar. 2023. doi:10.22331/q-2023-03-20-955.

[11] A. Gilyén, S. Lloyd, I. Marvian, Y. Quek, and M. M. Wilde. Quantum algorithm for petz recovery channels and pretty good measurements. *Phys. Rev. Lett.*, 128:220502, Jun 2022. doi:10.1103/PhysRevLett.128.220502.

[12] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 193–204, 2019. doi:10.48550/arXiv.1806.01838.

[13] J. Haah. Product decomposition of periodic functions in quantum signal processing. *Quantum*, 3:190, Oct. 2019. doi:10.22331/q-2019-10-07-190.

[14] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, second edition, 2002. doi:10.1137/1.9780898718027.

[15] S. Lang. *Real and functional analysis*, volume 142. Springer Science & Business Media, 2012. doi:10.1007/978-1-4612-0897-6.

[16] L. Lin and Y. Tong. Near-optimal ground state preparation. *Quantum*, 4:372, Dec. 2020. doi:10.22331/q-2020-12-14-372.

[17] L. Lin and Y. Tong. Optimal quantum eigenstate filtering with application to solving quantum linear systems. *Quantum*, 4:361, Nov. 2020. doi:10.22331/q-2020-11-11-361.

[18] G. H. Low and I. L. Chuang. Optimal hamiltonian simulation by quantum signal processing. *Phys. Rev. Lett.*, 118:010501, Jan 2017. doi:10.1103/PhysRevLett.118.010501.

[19] J. M. Martyn, Y. Liu, Z. E. Chin, and I. L. Chuang. Efficient fully-coherent quantum signal processing algorithms for real-time dynamics simulation. *The Journal of Chemical Physics*, 158(2):024106, 01 2023. doi:10.1063/5.0124385.

[20] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang. A grand unification of quantum algorithms. *PRX Quantum*, 2:040203, Dec 2021. doi:10.1103/PRXQuantum.2.040203.

[21] D. Motlagh and N. Wiebe. Generalized quantum signal processing. *PRX Quantum*, 5:020368, Jun 2024. doi:10.1103/PRXQuantum.5.020368.

[22] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Verlag, 1999. doi:10.1007/978-0-387-40065-5.

[23] G. U. Ramos. Roundoff error analysis of the fast fourier transform. *Mathematics of Computation*, 25(116):757–768, 1971. doi:10.2307/2004342.

[24] L. N. Trefethen. *Approximation theory and approximation practice*, volume 164. SIAM, 2019. doi:10.1137/1.9781611975949.

[25] J. Wang, Y. Dong, and L. Lin. On the energy landscape of symmetric quantum signal processing. *Quantum*, 6:850, Nov. 2022. doi:10.22331/q-2022-11-03-850.

[26] L. Ying. Stable factorization for phase factors of quantum signal processing. *Quantum*, 6:842, Oct. 2022. doi:10.22331/q-2022-10-20-842.

A Some useful results related to Banach space

Definition 28. *Let X, Y be normed vector spaces. A map $F : X \to Y$ is called C^1 if for every $x \in X$, there exists a (unique) bounded linear map $DF_x : X \to Y$ such that $\lim_{h\to 0} \frac{\|F(x+h)-F(x)-DF_x(h)\|}{\|h\|} = 0$, and the map $X \to \mathcal{L}(X,Y), x \mapsto DF_x$ is continuous. Here $\mathcal{L}(X,Y)$ is the set of all bounded linear maps from X to Y with norm topology.*

Here is a useful result from functional analysis.

Theorem 29. *Let X be a Banach space equipped with some norm $\|\cdot\|$ and $T : X \to X$ be a bounded linear operator. Suppose $\|T\| < 1$, then $\mathbb{I} - T$ is invertible with*

$$\left\|(\mathbb{I} - T)^{-1}\right\| \leq \frac{1}{1 - \|T\|}, \tag{96}$$

where $\mathbb{I}$ is the identity operator.

Proof. Consider the series $\sum_{k=0}^{\infty} \|T\|^k = \frac{1}{1-\|T\|}$, it converges since $\|T\| < 1$. Hence $\sum_{k=0}^{\infty} T^k$ converges in $\mathcal{L}(X,X)$ and we denote its limit as S. Observe that $S(\mathbb{I} - T) = (\mathbb{I} - T)S = \mathbb{I}$. Hence $\mathbb{I} - T$ is invertible and $\left\|(\mathbb{I} - T)^{-1}\right\| = \|S\| \leq \frac{1}{1-\|T\|}$. □

The terminologies used in the statement of the following theorem are explained in Definition 28.

Theorem 30. *Suppose that X, Y are dense subspaces of Banach spaces $\overline{X}, \overline{Y}$. Let F be a C^1 map, $F : X \to Y$, such that for any $\{x_n\}$ a Cauchy sequence in $X, F(x_n)$ is Cauchy in Y and DF_{x_n} is Cauchy in $\mathcal{L}(X,Y)$. Then there exists a unique C^1 map $\overline{F} : \overline{X} \to \overline{Y}$ such that $\overline{F}\Big|_X = F$ and $D\overline{F}\Big|_X = DF$.*

Proof. $\forall x \in \overline{X}$, there exists $\{x_n\} \subset X$ such that $x = \lim_{n\to\infty} x_n$, then we define $\overline{F}(x) = \lim_{n\to\infty} F(x_n)$. It is a well-defined and unique continuous map from $\overline{X}$ to $\overline{Y}$ extending F.

To show $\overline{F}$ is C^1, fix $x \in X$, then $DF_x : X \to Y$ is a bounded linear map, hence it has a unique extension $G_x : \overline{X} \to \overline{Y}$. Since $F \in C^1$,$\forall \epsilon > 0, \exists \delta > 0$, if $h \in X$, $\|h\| < \delta$, then

$$\|F(x+h) - F(x) - DF_x(h)\| \leq \epsilon \|h\|. \tag{97}$$

By continuity of $\overline{F}$, we have for $\forall h \in \overline{X}$,

$$\left\|\overline{F}(x+h) - \overline{F}(x) - G_x(h)\right\| \leq \epsilon \|h\|. \tag{98}$$

Now if $x \in \overline{X}$, write $x = \lim_{n\to\infty} x_n$ for $x_n \in X$, define $G_x : \overline{X} \to \overline{Y}$ by $G_x = \lim_{n\to\infty} G_{x_n}$, where the limit is in norm topology. Then by assumption, G_x is a well-defined, bounded linear map, and the map $X \to \mathcal{L}(\overline{X}, \overline{Y}), x \mapsto G_x$ is continuous.

By Eq. (98) and the continuity of $\overline{F}(\cdot), G.(h)$, we have $\forall \epsilon > 0$, $\exists \delta > 0$, if $x \in \overline{X}$, $h \in \overline{X}$, $\|h\| < \delta$,

$$\left\|\overline{F}(x+h) - \overline{F}(x) - G_x(h)\right\| \leq \epsilon \|h\|. \tag{99}$$

Hence $\overline{F} : \overline{X} \to \overline{Y}$ is C^1 such that $\overline{F}|_X = F$ and $D\overline{F}_x = G_x$, $\forall x \in \overline{X}$. □

The following inverse mapping theorem can be found in e.g., [15, Theorem 1.2, Chapter XIV].

Theorem 31 (Inverse Mapping Theorem)**.** *Let X, Y be Banach spaces and $F : X \to Y$ be a C^1 map. Let $x_0 \in X$ and assume that DF_{x_0} is invertible as a bounded linear map. Then there exist open sets $U \subset X, V \subset Y$ such that $x_0 \in U, F(x_0) \in V$ and $F : U \to V$ is bijective and F^{-1} is C^1.*

B Proof for Lemma 20

Proof. The existence of F is guaranteed by inverse mapping theorem. Lemma 18 implies that $DF(\Phi)$ is invertible for any Φ with $\|\Phi\|_1 < r_\Phi$. Let $\widetilde{F}$ be the restriction of F on $B(0, r_\Phi)$. The proof is divided into two parts:

1. show that $\widetilde{F}$ is injective. Moreover, given $\|\Phi\|_1 \leq r_\Phi$, the effective length of $\widetilde{F}(\Phi)$ is equal to that of Φ.
2. show that for any $c \in \mathbb{R}^\infty$ with $\|c\|_1 < r_c$, $\widetilde{F}^{-1}(c)$ exists.

For the first part, we prove it by contradiction. Suppose that $\widetilde{F}$ is not injective, then there exists $\Phi^{(1)} \neq \Phi^{(2)}$ such that $\widetilde{F}(\Phi^{(1)}) = \widetilde{F}(\Phi^{(2)})$. Denote $r = \max\left\{\left\|\Phi^{(1)}\right\|_1, \left\|\Phi^{(2)}\right\|_1\right\} < r_\Phi$, and define $\widetilde{G}(\Phi) := \widetilde{F}(\Phi) - 2\Phi$ for $\Phi \in B(0, r)$. Hence $\left\|D\widetilde{G}\right\|_1 = \left\|D\widetilde{F} - 2\mathbb{I}\right\|_1 \leq h(r) < 2$, which derives that

$$\left\|\widetilde{G}(\Phi^{(1)}) - \widetilde{G}(\Phi^{(2)})\right\|_1 \leq h(r)\left\|\Phi^{(1)} - \Phi^{(2)}\right\|_1. \tag{100}$$

Plug in that $\widetilde{F}(\Phi^{(1)}) = \widetilde{F}(\Phi^{(2)})$, and we get $2\left\|\Phi^{(2)} - \Phi^{(1)}\right\|_1 \leq h(r)\left\|\Phi^{(2)} - \Phi^{(1)}\right\|_1$, which is a contradiction. Hence, $\widetilde{F}$ is injective.

Next we show that the effective length of $\widetilde{F}(\Phi)$ is equal to that of Φ given $\|\Phi\|_1 \leq r_\Phi$. We use l_Φ and l_c to denote the effective length of Φ and $\widetilde{F}(\Phi)$, respectively. Hence, $\Phi = (\phi_0, \phi_1, \cdots, \phi_{l_\Phi - 1}, 0, \cdots)$. According to Theorem 7, l_Φ is no less than l_c. Suppose that $l_\Phi > l_c$. We may apply [25, Lemma 10] to compute the Chebyshev coefficient of $\mathrm{Im}[\langle 0|U(x, \Phi(s_n))|0\rangle]$ with respect to $T_{2(l_\Phi - 1)}$ and get

$$0 = \sin(2\phi_{l_\Phi - 1})\cos(2\phi_0)\prod_{i=1}^{l_\Phi - 2}\cos^2(\phi_i). \tag{101}$$

Hence, there exists $i \leq l_\phi - 2$ such that $\phi_i = \frac{\pi}{4}j$ for some nonzero integer j. Then, $\|\Phi\|_1 \geq \frac{\pi}{4} > r_\Phi$, which is a contradiction. Hence, $l_\Phi = l_c$.

For the second part, we also prove it by contradiction. Suppose that there is an $\overline{s} < r_c$ and a $Z \in \mathbb{R}^\infty$ with $\|Z\|_1 = 1$ such that $\overline{s}Z$ does not lie in the range of $\widetilde{F}$, then we can define

$$s^* = \inf\{s \in [0, \overline{s}] : sZ \text{ doesn't lie in the range of } \widetilde{F}\}. \tag{102}$$

From inverse mapping theorem, we know that $\widetilde{F}$ is invertible near 0, hence s^* is well defined. We also define $\Phi(r) := \widetilde{F}^{-1}(rZ)$ for any $r \in [0, s^*)$.

We claim that there exists $r' < r_\Phi$ such that $\|\Phi(s)\|_1 < r'$ for any $s \in (0, s^*)$. That is because

$$\begin{aligned}
s &= \left\|\widetilde{F}(\Phi(s))\right\|_1 = \left\|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{F}(t\Phi(s))\mathrm{d}t\right\|_1 \\
&= \left\|\int_0^1 D\widetilde{F}(t\Phi(s)) \cdot \Phi(s)\mathrm{d}t\right\|_1 \\
&= \left\|\int_0^1 \left(2\mathbb{I} + D\widetilde{F}(t\Phi(s)) - 2\mathbb{I}\right) \cdot \Phi(s)\mathrm{d}t\right\|_1 \\
&\geq \left\|\int_0^1 2\mathbb{I} \cdot \Phi(s)\mathrm{d}t\right\|_1 - \left\|\int_0^1 \left(D\widetilde{F}(t\Phi(s)) - 2\mathbb{I}\right) \cdot \Phi(s)\mathrm{d}t\right\|_1 \\
&\geq 2\|\Phi(s)\|_1 - \|\Phi(s)\|_1 \int_0^1 \left\|D\widetilde{F}(t\Phi(s)) - 2\mathbb{I}\right\|_1 \mathrm{d}t.
\end{aligned} \tag{103}$$

We apply Lemma 18 and get

$$s \geq \|\Phi(s)\|_1 \int_0^1 2 - h(t\,\|\Phi(s)\|_1)\mathrm{d}t = H(\|\Phi(s)\|_1). \tag{104}$$

Note that $H(x)$ is monotonically increasing over $[0, r_\Phi]$ and $r_c := H(r_\Phi) > \overline{s} > s$. We choose $r' \in (0, r_\Phi)$ such that $H(r') = \overline{s}$. It follows that $\|\Phi(s)\|_1 \leq r'$ for any $s \in (0, s^*)$ with $\Phi(s) \in B(0, r_\Phi)$.

Let $\{s_n\}_{n=0}^{\infty}$ be arbitrary series such that $\lim_{n\to\infty} s_n = s^*$ and $s_n < s^*$. Notice that $t\Phi(s_n) + (1-t)\Phi(s_m) \in B_1(0, r')$ for any $t \in (0,1)$. Denote $v := \Phi(s_m) - \Phi(s_n)$ and then one has

$$\begin{aligned}
&|s_n - s_m| = \left\|\widetilde{F}(\Phi(s_n)) - \widetilde{F}(\Phi(s_m))\right\|_1 \\
&= \left\|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{F}(\Phi(s_n) + tv)\mathrm{d}t\right\|_1 \\
&= \left\|\int_0^1 D\widetilde{F}(\Phi(s_n) + tv)\cdot v\mathrm{d}t\right\|_1 \\
&= \left\|\int_0^1 \left(2\mathbb{I} + D\widetilde{F}(\Phi(s_n) + tv) - 2\mathbb{I}\right)\cdot v\mathrm{d}t\right\|_1 \\
&\geq \left\|\int_0^1 2v\mathrm{d}t\right\|_1 - \left\|\int_0^1 \left(D\widetilde{F}(\Phi(s_n) + tv) - 2\mathbb{I}\right)\cdot v\mathrm{d}t\right\|_1 \\
&\geq 2\,\|\Phi(s_n) - \Phi(s_m)\|_1 - \int_0^1 \left\|\left(D\widetilde{F}(\Phi(s_n) + tv) - 2\mathbb{I}\right)\cdot v\right\|_1 \mathrm{d}t \\
&\geq 2\,\|\Phi(s_n) - \Phi(s_m)\|_1 - \int_0^1 h\left(\|(1-t)\Phi(s_n) + t\Phi(s_m)\|_1\right)\|\Phi(s_n) - \Phi(s_m)\|_1\,\mathrm{d}t \\
&\geq 2\,\|\Phi(s_n) - \Phi(s_m)\|_1 - h(r')\,\|(\Phi(s_n) - \Phi(s_m))\|_1 \\
&\geq (2 - h(r'))\,\|\Phi(s_n) - \Phi(s_m)\|_1 .
\end{aligned} \tag{105}$$

Since $2 - h(r') > 0$, we know that

$$\|\Phi(s_n) - \Phi(s_m)\|_1 \to 0, \quad \text{as} \quad n, m \to \infty. \tag{106}$$

According to the proof for the first part, for any n, the effective length of $\Phi(s_n)$ is l_c, where l_c is the effective length of Z. We may view $\Phi(s_n)$ as vectors in $\mathbb{R}^{l_c}$ instead. The limit of $\{\Phi(s_n)\}$ exists in $\mathbb{R}^{l_c}$ and is unique, denoted by Φ^*. Moreover, Φ^* can also be viewed as an element of $\mathbb{R}^\infty$ and $\Phi^* \in B(0, r_\Phi)$. By the continuity of $\widetilde{F}$, we know that

$$\widetilde{F}(\Phi^*) = \lim_{s\to s^*} \widetilde{F}(\Phi(s)) = s^* Z, \tag{107}$$

i.e., $\widetilde{F}^{-1}(s^* Z)$ exists.

Then we can use Theorem 31 at Φ^* and obtain that $\widetilde{F}^{-1}(sZ)$ exists for $s \in [s^*, s^* + \epsilon]$ for some ϵ, which contradicts with the definition of s^*.

When it comes to the proof of Eq. (57), we only need to let $s = \|c\|_1$ and $Z = \frac{c}{s}$, and do the calculation in Eq. (103) again. □