Appl. Comput. Harmon. Anal. 45 (2018) 463–485

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

The discrete sign problem: Uniqueness, recovery algorithms and phase retrieval applications

Ben Leshem [a], Oren Raz [b], Ariel Jaffe [c], Boaz Nadler [c,*]

[a] *Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel*
[b] *Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA*
[c] *Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel*

ARTICLE INFO

Article history:
Received 21 April 2016
Received in revised form 1 December 2016
Accepted 27 December 2016
Available online 3 January 2017
Communicated by Radu Balan

Keywords:
Phase retrieval
Signal reconstruction
Compact support
Sampling theory

ABSTRACT

In this paper we consider the following real-valued and finite dimensional specific instance of the 1-D classical phase retrieval problem. Let $\mathbf{F} \in \mathbb{R}^N$ be an N-dimensional vector, whose discrete Fourier transform has a compact support. The sign problem is to recover $\mathbf{F}$ from its magnitude $|\mathbf{F}|$. First, in contrast to the classical 1-D phase problem which in general has multiple solutions, we prove that with sufficient over-sampling, the sign problem admits a unique solution. Next, we show that the sign problem can be viewed as a special case of a more general piecewise constant phase problem. Relying on this result, we derive a computationally efficient and robust to noise sign recovery algorithm. In the noise-free case and with a sufficiently high sampling rate, our algorithm is guaranteed to recover the true sign pattern. Finally, we present two phase retrieval applications of the sign problem: (i) vectorial phase retrieval with three measurement vectors; and (ii) recovery of two well separated 1-D objects.

1. Introduction

The recovery of a signal from modulus (absolute value) measurements of its Fourier transform, known as *phase retrieval*, is a classical problem with a broad range of applications, including X-ray crystallography [25], astrophysics [18], lensless imaging [24,13], and characterization of ultra-short pulses [36], to name but a few.

From a mathematical perspective, fundamental questions regarding uniqueness of the phase problem and development of reconstruction algorithms have been topics of intense research for several decades. Uniqueness of the phase problem in one and two-dimensions, typically under the assumption that the underlying signal has a compact support, was studied by many authors, see for example [2,9,10,20,30] and

[*] Corresponding author.
E-mail address: boaz.nadler@weizmann.ac.il (B. Nadler).

http://dx.doi.org/10.1016/j.acha.2016.12.003
1063-5203/© 2017 Elsevier Inc. All rights reserved.

many additional references therein. For a recent survey and new results on the uniqueness of the 1-D phase problem, see Beinert and Plonka [6]. From the computational aspect, as the classical phase problem is non-convex, the most commonly used phase-retrieval methods are iterative [5,15,17,23]. These often require careful initialization, may exhibit limited robustness to noise, and in general, even in the absence of noise, are not guaranteed to converge to the correct solution.

In recent years there has been a renewed surge of interest in the phase problem. This was motivated in part by proposals of new measurement schemes coupled with novel convex-optimization approaches that provide strong guarantees on correct recovery. Examples include coded diffraction patterns, polarization type schemes and other methods with multiple illuminations [12,3,1,28,26], semi-definite programs for matrix completion [35,11] and sparsity-based recovery methods [31,32].

Motivated by several phase retrieval applications, here we consider a finite dimensional and real-valued particular instance of the general 1-D phase problem, which we denote as the *sign problem.* As described in Section 2, its formulation is as follows: Let $\mathbf{F}$ be an N-dimensional real-valued vector, whose discrete Fourier transform, $\mathbf{f} = \mathcal{DFT}\{\mathbf{F}\} \in \mathbb{C}^N$, has a support of length $\tau + 1$. The sign problem is to recover the sign pattern $\mathbf{s} = sign(\mathbf{F}) \in \{\pm 1\}^N$ from possibly noisy measurements of $|\mathbf{F}|$.

In this paper we perform a detailed study of this finite dimensional sign problem, including its uniqueness and the development of a stable reconstruction algorithm. We also present its application to two practical phase problems. First, in Section 3, Theorem 1 we prove that if $N > 2\tau$, our discrete sign problem admits a unique solution, up to a global ± 1 sign ambiguity. Our proof, similar to [28,9,6], is based on analyzing the roots of high degree polynomials. Since finding such roots is known to be an ill-conditioned problem [34], our proof does not directly lead to a stable reconstruction algorithm.

To robustly solve the sign problem we take a different approach. First, we study the structure of its solutions, showing in Lemma 1 that the sign pattern $\mathbf{s} = sign(\mathbf{F}) \in \{\pm 1\}^N$ cannot be arbitrary, but rather has at most τ sign changes. Next, we relax the constraint that $\mathbf{s} \in \{\pm 1\}^N$ and allow the sign pattern to be a complex-valued N dimensional vector, which guided by Lemma 1, is piecewise constant over at most $\tau + 1$ intervals. This leads us to study the following two questions: (i) is it possible to detect at least parts of these intervals, where the underlying sign pattern is constant? and (ii) does such an *over-segmentation* of $1, \ldots, N$ to intervals of constant values indeed retains uniqueness of the problem? With respect to question (ii), we prove in Theorem 2 that given an over-segmentation with few segments M, such that $N > 2\tau + M$, this piecewise constant phase problem has a unique solution.

Based on these theoretical results, in Section 4 we address question (i) above and develop methods to find either an exact or an approximate over-segmentation to intervals of constant sign, given only (noisy) measurements of $|\mathbf{F}|$. Given such an over-segmentation, we then develop a computationally efficient algorithm to retrieve the unknown sign pattern. Our approach follows our previous works [28,29], whereby instead of taking the signal $\mathbf{f}$ as our unknown, we work with the unknown phases, and formulate for them a quadratic functional to be minimized. Relaxing the requirement that the solution is a phase vector leads to solving a system of linear equations. In the noise-free case we prove that with a sufficiently high sampling rate, our algorithm is guaranteed to recover the true sign pattern.

Section 5 presents two phase retrieval applications of practical interest where the sign problem arises. The first is vectorial phase retrieval with three measurement vectors. Here the problem is to recover two compactly supported signals $\mathbf{f}_1$ and $\mathbf{f}_2$ from measurements of $|\mathbf{F}_1|$, $|\mathbf{F}_2|$ and their interference $|\mathbf{F}_1 + \mathbf{F}_2|$. With sufficient over-sampling, this problem was proven to admit a unique solution in [6], but no reconstruction algorithm was given. The second application is the recovery of two well separated 1-D objects from a single spectrum, a problem known to have a unique solution [14]. We show how stable recovery for both problems is possible by solving a related sign problem. Finally, in Section 6 we illustrate the performance of our algorithm via several simulations. For an example with real 2-D experimental data (involving a 2-D sign problem), we refer the reader to [21].

This discrete sign problem considered in this paper can be viewed as the finite dimensional analogue of the following continuous problem, recently studied by Thakur [33]: Recover a real-valued function $g(t)$ whose continuous Fourier transform $G(\omega)$ is band-limited, from discrete measurements $|g(t_j)|$. The analogy between the two problems follows by relating the continuous function g to our finite dimensional vector $\mathbf{F}$ and its Fourier transform G to $\mathbf{f}$. In [33], Thakur proved that if the sampling rate is at least twice the Nyquist rate then the continuous sign problem is well posed. He further developed an algorithm to reconstruct the function from a finite number of N measurements. While with noise-free data the reconstruction error of Thakur's method decays exponentially fast in N, as we illustrate in Section 6, it is quite sensitive to even small measurement noise. Our sign recovery algorithm, in contrast, is based on an entirely different approach, in fact relying on a discrete formulation. Yet, it is applicable to this continuous setting, and as shown in Section 6, offers improved robustness to noise. We conclude in Section 7 with a discussion and directions for future research.

2. Problem setup

Notation. We denote $\mathbf{i} = \sqrt{-1}$. For $z \in \mathbb{C}$ we denote by $\mathcal{R}(z)$ and $\mathcal{I}(z)$ its real and imaginary parts and by $\bar{z}$ its complex conjugate. Vectors appear in boldface letters, for example $\mathbf{F} = (F_0, \ldots, F_{N-1})$. We denote the entry-wise multiplication of two vectors $\mathbf{F}$ and $\mathbf{G}$ by $\mathbf{FG}$. We further denote by $\mathbf{f} \star \mathbf{g}$ the cross-correlation between the vectors $\mathbf{f}$ and $\mathbf{g}$.

Measurements. While in this paper we shall mostly study a discrete formulation, it is nonetheless instructive to first briefly review the continuous setting. Let $f_c(t)$ denote a one dimensional continuous signal, whose 1-D Fourier transform $F_c(\omega)$ is given by

$$F_c(\omega) = \int f_c(t) e^{-\mathbf{i}\omega t} dt = |F_c(\omega)| e^{\mathbf{i}\phi(\omega)}. \tag{1}$$

As mentioned in the introduction, in many applications direct measurement of $f_c(t)$ is not possible. Rather, the measured data is typically an equispaced sampling of $|F_c(\omega)|^2$ at $\omega_j = j\Delta\omega$, $j = 0, ..., N-1$. We denote the values of $F_c(\omega)$ at the sampled frequencies by $F_j := F_c(\omega_j)$, and denote $\mathbf{F} = (F_0, \ldots, F_{N-1})$. A discrete approximation of $f_c(t)$ at the N points $t_k = \frac{2\pi k}{N\Delta\omega}$, $k = 0, \ldots, N-1$ can be computed via the discrete Fourier transform $(\mathcal{DFT})$ of $F_c(\omega_j)$

$$f(t_k) = \frac{\Delta\omega}{2\pi} \sum_{j=0}^{N-1} |F_c(w_j)| e^{\mathbf{i}\phi(\omega_j)} e^{\mathbf{i}\omega_j t_k} := \mathcal{DFT}\left\{\mathbf{F}\right\}. \tag{2}$$

For convenience, we rescale time and frequency so that $\Delta\omega = \frac{2\pi}{N}$ and $t_k = k$ for $k = 0, ..., N-1$. In this paper we restrict our attention to a finite dimensional formulation and consider the reconstruction of the N signal values $\mathbf{f} = (f_0, \ldots, f_{N-1})$ as our end goal. We note that, strictly speaking, $f(t_k) \neq f_c(t_k)$. However, for $N \gg 1$ and a sufficiently high sampling rate, $f(t_k) \approx f_c(t_k)$, see for example [19].

For future use, we recall that the two vectors $\mathbf{f}$ and $\mathbf{F}$ are thus related as follows

$$f_k = \mathcal{DFT}\{\mathbf{F}\}(k) = \frac{1}{N} \sum_{j=0}^{N-1} F_j e^{\mathbf{i}\omega_j k} \quad \text{and} \quad F_j = \mathcal{IDFT}\{\mathbf{f}\}(j) = \sum_{k=0}^{N-1} f_k e^{-\mathbf{i}\omega_j k}. \tag{3}$$

The classical phase problem. Let $|\mathbf{F}|^2 \in \mathbb{R}^N$ be the measured spectrum at the N equispaced frequencies as described above and denote the unobserved phase vector by $\boldsymbol{\phi} = arg(\mathbf{F})$. The phase problem is to reconstruct from the intensity measurements $|\mathbf{F}|^2$, the missing phases $\{\phi_j\}_{j=0}^{N-1}$ or equivalently the N values $\mathbf{f} = \mathcal{DFT}\left\{\mathbf{F}\right\}$. Clearly, without additional constraints this problem is ill-posed, as any vector of phase values is a valid solution. A commonly imposed constraint is that $\mathbf{f}$ has a compact support. However, even with this assumption, the 1-D phase retrieval problem is in general still ambiguous. If $N > 2\tau$ and $\mathbf{f}$ has a support of length $\tau + 1$, it still has at most $2^{\tau-1}$ different solutions, see [6,9]. However, as shown in Theorem 3.1 and Corollary 3.2 of [7], additional information on the underlying signal, such as its amplitude at a single point inside its support almost always suffices for uniqueness.

The sign problem. In this paper we consider phase retrieval when the underlying function $F_c(\omega)$ is *real-valued*. Namely, the unobserved phases of $\mathbf{F}$ are restricted to $\phi_j \in \{0, \pi\}$. Given $|\mathbf{F}|^2$ the problem is then to reconstruct this sign pattern. Clearly, without additional constraints this problem is also ill-posed as any sign pattern is a valid solution. In this paper we assume that $\mathbf{f}$ has a support of size $(\tau + 1) < N$ and focus on the following two questions: 1) is there a unique solution? and 2) assuming there is, can one develop a stable and computationally efficient algorithm to recover it.

Trivial ambiguities. If $\mathbf{f}$ is a solution to the classical phase problem, then so are its circular shift, its reflection, conjugation and multiplication by a unimodular factor $e^{\mathbf{i}\theta}$. These transformations do not change the physical nature of the signal and are thus known as trivial ambiguities of the phase problem. In the sign problem, since $\mathbf{F}$ is real-valued it implies that $f_k = f^*_{(-k \bmod N)}$ for all $k = 0, \ldots, N-1$. This eliminates the reflection ambiguity and all circular time shifts ambiguities except for a shift by $N/2$, assuming N is even. Hence, the trivial ambiguities of the sign problem are only a circular shift by $N/2$ and multiplication by a global sign.

Signal support. We say that a signal $\mathbf{f} \in \mathbb{C}^N$ has a support of length $\tau + 1$ if $f(k) = 0$ for all $k \notin CS$ where for some $\alpha \in \mathbb{Z}$

$$CS = \{k : k \in [-\alpha, -\alpha + \tau] \mod N\}. \tag{4}$$

For the sign problem, given the discussion above on trivial ambiguities and assuming for simplicity that τ is even, there are only two options for the set CS. Either it is centered around zero ($\alpha = \tau/2$),

$$CS = \left\{k : k \in [-\frac{\tau}{2}, \frac{\tau}{2}] \mod N\right\}, \tag{5}$$

or it is the above set, circularly shifted by $N/2$. Without loss of generality, we will use Eq. (5) as the assumed support of the sign problem.

3. Mathematical uniqueness

In this section we study the uniqueness of the sign problem. As mentioned in the introduction, there are quite a few results on the number of frame-type measurements required for uniqueness of various phase problems, see for example [3,4,6,8] and references therein. However, our specific problem, involving a discrete unknown sign pattern $\mathbf{s} \in \{-1, 1\}^N$, does not fit into these general formulations, and hence their results do not directly apply to the sign problem. First, we show that in the ideal setting of noise free measurements, if the signal $\mathbf{f}$ has a sufficiently small support compared to N (or equivalently, if $|F_c(\omega)|^2$ is sampled at a sufficiently high rate), then the sign problem admits a unique solution. Next, we show how the sign problem can be viewed as a special case of a more general phase problem, where the (complex-valued) phase is assumed to be piecewise constant. Moreover, we prove that at the expense of a higher sampling rate, this piecewise constant phase problem also admits a unique solution. Finally, in Section 4 we combine the above results and develop a simple, computationally efficient algorithm to solve the original sign problem.

Specifically, we introduce the following assumption:

A1 *Support.* The vector $\mathbf{f} = \mathcal{DFT}\{\mathbf{F}\} \in \mathbb{C}^N$ has a support of length $\tau + 1$. For the classical phase problem the support set is given by Eq. (4), whereas for the sign problem it is given by Eq. (5).

In this section we assume that τ is known. In Section 4 we discuss how τ can be estimated from the measured data. The following theorem proves that, when sampled properly, this sign problem has a unique solution.

Theorem 1. *Let $|\mathbf{F}|^2$ be the observed noise-free spectrum of a signal that satisfies assumption A1, with support given by Eq. (5). If $\mathbf{F} \in \mathbb{R}^N$ and $N > 2\tau$ then the sign problem has a unique solution up to a global ± 1 sign ambiguity.*

Proof. By assumption A1, the support of $\mathbf{f}$ is $[-\tau/2, \ldots, \tau/2] \mod N$. It is convenient to consider its circularly *shifted* signal, $\mathbf{f}_s$, defined as $f_s(k) = f(k + \tau/2 \mod N)$. By definition, the support of $\mathbf{f}_s$ is $[0, 1, \ldots, \tau]$, and thus its $\mathcal{DFT}$ is $F_s(\omega_j) = \sum_{k=0}^{\tau} f_s(k) e^{\mathbf{i}\omega_j k}$. The relation between $\mathbf{F}_s$ and $\mathbf{F}$ is

$$F(\omega_j) = e^{-\mathbf{i}\omega_j \tau/2} F_s(\omega_j) = e^{\mathbf{i}\omega_j (N - \frac{\tau}{2})} F_s(\omega_j)$$

where the last equality follows since $e^{\mathbf{i}\omega_j N} = 1$ for all j.

Next, we make the change of variables $z = e^{\mathbf{i}\omega}$, known as the z-transform. With some abuse of notation we denote the resulting polynomial as $F(z)$, defined for all $z \in \mathbb{C}$,

$$F(z) = z^{N - \frac{\tau}{2}} \sum_{k=0}^{\tau} f_s(k) z^k. \tag{6}$$

By its definition, $F(z)$ is analytic in the complex plane. Further, at the N sampling points, $z_j = e^{\mathbf{i}\omega_j}$, we have that $F(z_j) = F(\omega_j)$.

By the fundamental theorem of algebra, the polynomial $F(z)$ of Eq. (6) can be decomposed as

$$F(z) = c z^{N - \frac{\tau}{2}} \prod_{r=1}^{\tau} (z - z_r) \tag{7}$$

for some $c \in \mathbb{C}$. Since the support of the shifted signal $f_s(k)$ is the set $\{0, 1, ..., \tau\}$ it follows that both $f_s(0) \neq 0$ and $f_s(\tau) \neq 0$. Hence, all $z_r \neq 0$, and the polynomial $F(z)$ has exactly τ non-zero roots.

To guarantee a unique solution to the sign problem, these roots must be uniquely determined from the N measurements $\{|F(\omega_j)|\}_{j=0}^{N-1}$. In the classical 1-D phase problem, where $\mathbf{F} \in \mathbb{C}^N$, these N measurements do not yield a unique solution. In our sign problem, in contrast, $\mathbf{F} \in \mathbb{R}^N$, and as we now show, this leads to uniqueness, up to trivial ambiguities of the sign problem.

To this end, consider the polynomial $F(z)^2$. By Eq. (7) it is given by

$$F(z)^2 = c^2 z^{2N - \tau} \prod_{r=1}^{\tau} (z - z_r)^2. \tag{8}$$

Namely, $F(z)^2$ has the *same* roots as $F(z)$, but each with its multiplicity doubled. Thus, $F(z)^2$ has exactly 2τ non-zero roots and is uniquely determined by its values at any $2\tau + 1$ distinct points. Since $F(z_j)$ is real valued, at the N observed points $F(z_j)^2 = |F(z_j)|^2$. Hence, if $N > 2\tau$ these observations uniquely determine the 2τ roots, which in turn determine $F(z)$ up to a global ± 1 sign ambiguity. □

According to Theorem 1, up to a ± 1 global sign, there is a single sign pattern $\mathbf{s} = (s_0, \ldots, s_{N-1})$, such that $\mathcal{DFT}\{|\mathbf{F}|\mathbf{s}\}$ yields a signal with the correct support. As the following lemma shows, this sign pattern cannot be arbitrary. Rather, it has a limited number of sign changes.

Lemma 1. *Let* $\mathbf{F} \in \mathbb{R}^N$ *be a vector such that* $\mathbf{f} = \mathcal{DFT}\{\mathbf{F}\}$ *satisfies assumption A1 with support given by Eq.* (5). *Then the vector* $\mathbf{s} = sign(\mathbf{F}) \in \{-1, 1\}^N$ *has at most* τ *sign changes.*

Proof. The assumption that $\mathbf{F} \in \mathbb{R}^N$ implies that $f(k) = f^*(-k \mod N)$ for all $k = 0, \dots, N-1$. This, in turn, implies that the function $F(\omega) = \sum_{k=-\tau/2}^{\tau/2} f(k)e^{\mathbf{i}\omega k}$, which extends $F(\omega_j) = F_j$ to all $\omega \in [0, 2\pi]$, is real-valued for all ω. As in the proof of Theorem 1, we consider the z-transform, $z = e^{\mathbf{i}\omega}$, and the polynomial $F(z)$ of Eq. (6). Then, $F(z)$ is real-valued for all $|z| = 1$. As in the proof of Theorem 1, $F(z)$ has τ non-zero roots. In particular, $F(z)$ can vanish in at most τ points on the unit circle. Furthermore, at the sampling points $z_j = e^{\mathbf{i}\omega_j}$ we have $F(z_j) = F_j$. Hence, in order for F_j and F_{j+1} to have different signs, the continuous function $F(z)$ must have a zero crossing somewhere along the arc between z_j and z_{j+1}. Thus, the total number of zeros of $F(z)$ bounds the maximal number of sign changes in the vector $\mathbf{s} = sign(\mathbf{F})$ to be τ. If at some sampling points $F(z_j) = 0$ then $sign(F(z_j))$ is ill-defined. In this case we set $sign(F(z_j)) = sign(F(z_{j-1}))$, and $sign(F(z_0)) = 1$ if $F(z_0) = 0$. □

Remark 1. It is interesting to contrast Lemma 1 with the phenomenon of superoscillations. As studied for example in [16], the Fourier transform of a compactly supported continuous signal may have an arbitrarily large number of sign changes. As a specific finite dimensional analogue, the vector $F_j = (-1)^j$ has $N-1$ sign changes and its $\mathcal{DFT}$ is a delta function at the point $N/2$ which has a support of length one ($\tau = 0$). While at first sight this seems to contradict Lemma 1, it does not, as Lemma 1 requires that the support is centered around the origin. Indeed, shifting the above delta function by $N/2$, so that its support is at the index $k = 0$ yields a $\mathcal{DFT}$ vector $\mathbf{F} = (1, 1, \dots, 1)$ with no sign changes at all, in accordance with Lemma 1.

According to Lemma 1, the sign pattern $\mathbf{s}$ that corresponds to a signal $\mathbf{f}$ with support of length $\tau + 1$, belongs to a set of size $O(N^\tau)$. For a fixed τ this set has size polynomial rather than exponential in N. Unfortunately, it is exponential in τ. Thus, finding the unique solution to a given sign problem by exhaustive search over all possible sign patterns in this set is generally computationally intractable.

To construct a computationally efficient sign retrieval algorithm, we consider the following more general phase problem: We relax the strict assumption of a real-valued phase and instead assume the phase is complex-valued but piecewise constant in M a-priori known segments. Note that according to Lemma 1, the (real-valued) sign problem is a particular instance of this piecewise constant phase problem. The following theorem shows that under suitable conditions, this modified phase problem also admits a unique solution. In section 4 below we then show how this modified problem can be solved computationally efficiently by framing it as the solution to a set of linear equations.

We mathematically formulate the piecewise constant phase property as follows:

A2 *Known segmentation into constant phase intervals.* There is a known division of the N frequencies ω_j to M contiguous segments, in which the unobserved phase vector $\boldsymbol{\phi}$ is piecewise constant. Let $\mathbf{c}$ be a vector of length M containing the first index in each of these segments. The m-th segment of length N_m consists of all frequencies ω_j with $j \in [c(m), c(m)+1, \dots, c(m)+N_m-1]$ and $\phi(\omega_{c(m)}) = \phi(\omega_{c(m)+k}), 1 \leq k \leq N_m - 1$.

If all indices where $\mathbf{F}$ changed its sign were known (e.g., we had a complete segmentation), then this information would directly resolve the sign pattern $\mathbf{s}$, up to its global ±1 sign ambiguity. Unfortunately, in practice it is difficult to determine all zero crossings from the observed $|\mathbf{F}|$, see for example [33].

Assumption A2 is thus interesting when it defines an over-segmentation of $\boldsymbol{\phi}$ to piecewise constant phase intervals, where the number of intervals M of the given segmentation is not necessarily minimal. Indeed, a key result of our paper, stated in Theorem 2 below, is that under suitable conditions even such an over-segmentation can suffice to resolve the sign problem.

Theorem 2. *Let $\mathbf{f} \in \mathbb{C}^N$ satisfy assumption A1 with a support given by Eq. (4), and suppose that $\mathbf{F} = \mathcal{IDFT}\{\mathbf{f}\}$ has a piecewise constant phase and satisfies assumption A2. Assume $|F_j| \neq 0$ at the first and last indices of each of the M segments of the given segmentation. If $N > 2\tau + M$ then the piecewise constant phase problem has a unique solution up to multiplication by a global phase. Namely, the vector $|\mathbf{F}|$ and the known segmentation uniquely determine $\mathbf{f}$.*

Proof. By assumption A1, the signal $\mathbf{f}$ has a support of length $\tau + 1$, given by Eq. (4), with an integer $\alpha \in [0, N-1]$ that is in general unknown. It will be convenient to work with the circularly shifted signal $\mathbf{f}_s$ given by $f_s(k) = f(k + \alpha \mod N)$, whose support consists of the indices $[0, 1, \dots, \tau]$. Its $\mathcal{DFT}$ is $F_s(\omega_j) = \sum_{k=0}^{\tau} f_s(k) e^{\mathbf{i}\omega_j k}$ and it is related to $\mathbf{F}$ via $F_s(\omega_j) = e^{\mathbf{i}\omega_j \alpha} F(\omega_j)$.

As in Theorem 1, consider the polynomials $F_s(z) = \sum_{k=0}^{\tau} f_s(k) z^k$ and $F(z) = z^{N-\alpha} F_s(z)$. In contrast to the sign problem, where $F(z)$ was real-valued for all $|z| = 1$, here the phase is only assumed to be piecewise constant, and so $F(z)$ is in general complex-valued. We thus write it as

$$F(z) = |F(z)| X(z). \tag{9}$$

Assumption A2 can be expressed as $X(z_m \gamma^n) = a_m,\ m = 1, \dots, M,\ n = 0, \dots, N_m - 1$ where $z_m = e^{\mathbf{i}\omega_{c(m)}}$ is the first point in each segment, $\gamma = e^{\mathbf{i}\Delta\omega}$ and a_m are unknown constants of unit modulus. Note that if at some interior point of a segment $|F(z_j)| = 0$, its phase $X(z_j)$ is ill defined. In such a case we define it to be equal to the phase of the left-most point in that segment.

To show that $|\mathbf{F}|^2$ uniquely determines $F(z)$ up to a global phase $e^{\mathbf{i}\phi}$, assume to the contrary that there exists another signal $\mathbf{g} \neq \mathbf{f}$ whose support is of the form $[-\alpha', -\alpha' + \tau] \mod N$, where possibly $\alpha' \neq \alpha$. According to the problem statement, $\mathbf{G} = \mathcal{IDFT}\{\mathbf{g}\}$ satisfies that $|\mathbf{G}| = |\mathbf{F}|$ and it has a piecewise constant phase in the same M segments as $\mathbf{F}$ as defined in assumption A2.

We denote the circular shift of $\mathbf{g}$ by $\mathbf{g}_s$, and the corresponding polynomials by $G(z) = z^{N-\alpha'} G_s(z)$ and $G_s(z) = \sum_{j=0}^{\tau} g_s(k) z^k$ respectively. Similarly to Eq. (9) we write $G(z) = |G(z)| Y(z)$. Assumption A2 implies that $Y(z_m \gamma^n) = b_m$, where b_m are unknown constants of unit modulus.

Next, we define the following polynomial, where $\gamma = e^{\mathbf{i}\Delta\omega}$

$$\begin{aligned} P(z) &= F(z)G(\gamma z) - G(z)F(\gamma z) \\ &= z^{2N-\alpha-\alpha'} \left(\gamma^{N-\alpha'} F_s(z) G_s(\gamma z) - \gamma^{N-\alpha} F_s(\gamma z) G_s(z) \right). \end{aligned} \tag{10}$$

Since both $F_s(z)$ and $G_s(z)$ are polynomials of degree τ, the term inside the brackets in Eq. (10) is a polynomial of degree at most 2τ. Hence $P(z)$ may have at most 2τ non-zero roots.

Now let us study the values $P(z_j)$. In each segment m of length $N_m \geq 2$ we have that $P(z)$ vanishes at all the points in this segment excluding its last one, $z_m \gamma^{N_m - 1}$, since for any $n = 0, \dots, N_m - 2$

$$\begin{aligned} P(z_m \gamma^n) &= F(z_m \gamma^n) G(z_m \gamma^{n+1}) - G(z_m \gamma^n) F(z_m \gamma^{n+1}) \\ &= |F(z_m \gamma^n)| |F(z_m \gamma^{n+1})| \left(X(z_m \gamma^n) Y(z_m \gamma^{n+1}) - Y(z_m \gamma^n) X(z_m \gamma^{n+1}) \right) \\ &= |F(z_m \gamma^n)| |F(z_m \gamma^{n+1})| (a_m b_m - b_m a_m) = 0. \end{aligned} \tag{11}$$

The total number of points where P vanishes in Eq. (11) is $N - M$. The condition that $N - M > 2\tau$ implies that $P(z) = 0$ everywhere.

Hence, at the last point in each segment, upon division by the non-vanishing signal magnitudes

$$X(z_m \gamma^{N_m - 1}) Y(z_m \gamma^{N_m}) = Y(z_m \gamma^{N_m - 1}) X(z_m \gamma^{N_m}),\ m = 1, \dots, M \tag{12}$$

Next, using $z_m\gamma^{N_m} = z_{m+1}$ and the fact that $\mathbf{X}$ and $\mathbf{Y}$ are phase vectors, Eq. (12) can be written as

$$X(z_m\gamma^{N_m-1})X^*(z_{m+1}) = Y(z_m\gamma^{N_m-1})Y^*(z_{m+1}),\ m = 1, \ldots, M. \tag{13}$$

From Eq. (13) it follows that the phase difference between any pair of consecutive segments in $X(z_j)$ is equal to the phase difference between the corresponding segments in $Y(z_j)$. Hence, we have that

$$X(z_j) = e^{\mathbf{i}\phi}Y(z_j),\ j = 0, \ldots, N-1 \tag{14}$$

where ϕ is an arbitrary global phase. Therefore, $F(z) = e^{\mathbf{i}\phi}G(z)$ and since $F_j = F(z_j)$ we conclude that $\mathbf{F}$ is uniquely determined up to an arbitrary global phase. □

Remark 2. Our proof of Theorem 2 breaks down if $F_j = 0$ at the last and/or first indices in one or more segments. To cover this case, the proof can be modified as follows: Redefine the polynomial $P(z)$ of Eq. (10) as $P(z) = F(z)G(\gamma^2 z) - G(z)F(\gamma^2 z)$. By Eq. (11) the modified $P(z)$ vanishes at $N - 2M$ points. Hence, if $N > 2M + 2\tau$ the proof can be completed as above. If $|\mathbf{F}|$ vanishes at a number of consecutive points near the edges of some segments the proof can be modified in a similar manner.

4. Sign reconstruction with known support

We now describe a simple, computationally efficient algorithm to solve the sign problem. First, we study the case of noise-free measurements, given an over-segmentation of $[0, \ldots, N-1]$ into intervals of constant sign. Using Theorem 2, in Section 4.1 we show how the sign pattern can be cast as the solution to an over-determined system of linear equations. Next, in Section 4.2 we describe a method to segment the N indices into intervals of constant sign, given only the vector $|\mathbf{F}|$. An algorithm to recover the sign pattern in the presence of noisy measurements appears in Section 4.3. Finally, estimation of the typically unknown support length τ is addressed in Section 4.4.

4.1. Sign recovery with known over-segmentation

Consider a signal $\mathbf{F} \in \mathbb{R}^N$ satisfying assumptions A1 and A2, and further assume that the length τ of the support of $\mathbf{f}$ is a-priori known. As in our earlier works [27,28], instead of working with the $(\tau+1)$ signal values f_k, $k = -\tau/2, \ldots, \tau/2 \pmod N$, as the unknown variables, we consider the vector of N unknown signs $\mathbf{X} = (X_0, \ldots, X_{N-1})$. Further, rather than dealing with the combinatorial set $\mathbf{X} \in \{-1, 1\}^N$, we relax this constraint and allow all entries X_j to be complex valued, namely $\mathbf{X} \in \mathbb{C}^N$.

As we now show, this allows us to write a system of linear equations for the vector $\mathbf{X}$ over the field $\mathbb{C}$ whose unique real-valued solution is the true sign pattern. The first set of equations captures the support assumption on the vector $\mathbf{f} = \mathcal{DFT}\{\mathbf{F}\}$. Since $\mathbf{f}$ is zero outside the set of indices CS of Eq. (5), the unknown vector $\mathbf{X}$ must satisfy the following set of $N - \tau - 1$ linear equations

$$\mathcal{DFT}\left\{|\mathbf{F}|\mathbf{X}\right\}(k) = 0\ ,\ k \notin CS. \tag{15}$$

The second set of equations imposes the known segmentation into intervals of constant sign, as described in A2. Let $\mathcal{S}$ denote the set of indices within the intervals of constant sign, excluding the last one in each interval. Then,

$$X_j = X_{j+1},\ \forall j \in \mathcal{S}. \tag{16}$$

Given a segmentation to M constant sign intervals the number of equations in (16) is thus $N - M$.

By definition, the true sign pattern $\mathbf{s} = sign(\mathbf{F})$ is a solution of Eqs. (15)–(16). Theorem 3 below shows that under suitable conditions and up to multiplication by a constant, it is the *only* solution to this system of equations.

Theorem 3. *Let $|\mathbf{F}|^2$ be the intensity of a signal that satisfies assumptions A1 and A2, with a support given by Eq. (4). Assume $|F_j| \neq 0$ at the first and last indices of each of the M segments of the given segmentation. If $N > 2\tau + M$ then, in the noise-free case, the only solutions to the set of linear equations (15)–(16) are of the form $\mathbf{X} = c\,\mathbf{s}$ where $c \in \mathbb{C}$.*

Proof. By definition, the true phase vector $\mathbf{s} = \mathbf{F}/|\mathbf{F}|$ is a solution of Eqs. (15)–(16). We now show that all solutions are of the form $\mathbf{X} = c\,\mathbf{s}$ where $c \in \mathbb{C}$. Assume to the contrary that there exists another $\mathbf{Y} \neq c\mathbf{s}$, which satisfies Eqs.(15)–(16). Let $\mathbf{G} = |\mathbf{F}|\mathbf{Y}$ be the signal associated with this solution. By Eq. (15), $\mathbf{g} = \mathcal{DFT}\{\mathbf{G}\}$ has the same (or smaller) support as that of $\mathbf{f}$. As in the proof of Theorem 2, let $\mathbf{g}_s$ be its shifted signal and let its corresponding polynomial be $G(z) = z^{N-\tau/2}G_s(z)$.

Next, as in Theorem 2, consider the polynomial $P(z)$ of Eq. (10) which has at most 2τ non-zero roots. Since by Eq. (16) the vector $\mathbf{Y}$ is piecewise constant in the same segments as $\mathbf{X}$, it follows from Eq. (11) that $P(z)$ vanishes at a total of $N - M$ points. Hence, the condition $N - M > 2\tau$ implies that $P(z) = 0$ everywhere. In particular, $P(z_j) = 0$ for all $j = 0, \ldots, N-1$. Specifically for the last point at each segment, upon division by $|F(z_m\gamma^{N_m-1})F(z_{m+1})|$ which is non-zero by our assumption,

$$\frac{X(z_m\gamma^{N_m-1})}{X(z_{m+1})} = \frac{Y(z_m\gamma^{N_m-1})}{Y(z_{m+1})}, \; m = 1, \ldots, M-1. \tag{17}$$

Eq. (17) implies that the proportionality constants between each pair of consecutive segments in $\mathbf{X}$ and $\mathbf{Y}$ are equal. Hence, $Y_j = cX_j$ for all $j = 0, \ldots, N-1$ for some constant $c \in \mathbb{C}$. □

Remark 3. Any over-segmentation with number of segments $M < N - 2\tau$ suffices for Theorem 3 to hold. Importantly, not all indices where $\mathbf{s}$ is piecewise constant need to be captured by Eq. (16).

Remark 4. While the condition $N > 2\tau + M$ is sufficient to ensure a rank-one solution to the linear system (15)–(16), it is by no means a necessary condition. Empirically, often an over-segmentation with more segments still suffices to reconstruct the correct sign pattern.

4.2. Segmentation to constant sign intervals

By Theorem 3, with a suitable over-segmentation one can retrieve the sign of $\mathbf{F}$ by solving the set of linear equations (15)–(16). To this end, however, one must first determine an (over-)segmentation from $|\mathbf{F}|^2$ alone. The following lemma provides a principled method to do so.

Lemma 2. *Let $\mathbf{F} \in \mathbb{R}^N$ be a signal that satisfies assumption A1 with support set (5). Then, the difference between two consecutive values of $\mathbf{F}$ is bounded by*

$$|F_j - F_{j-1}| \leq \Big(\frac{2}{N}\Big)^{3/2}\pi\mathcal{S}_{\tau/2}\|\mathbf{F}\| \tag{18}$$

where $\|\mathbf{F}\|^2 = \sum_{j=0}^{N-1} F_j^2$ and

$$\mathcal{S}_{\tau/2}^2 = \sum_{k=1}^{\tau/2} k^2 = \frac{\tau(\tau+1)(\tau+2)}{24} = O(\tau^3).$$

Proof. As in the proof of Lemma 1, let $F(\omega)$ be the extension of $F_j = F(\omega_j)$ to all $\omega \in [0, 2\pi]$. Since $\mathbf{f}$ is symmetric-conjugate and non-zero only in (5), $F(\omega)$ can be expressed as

$$\begin{aligned} F(\omega) &= \sum_{k=-\tau/2}^{\tau/2} f(k)e^{-\mathbf{i}\omega k} = f_0 + \sum_{k=1}^{\tau/2} \left(f(k)e^{-\mathbf{i}\omega k} + f^*(k)e^{\mathbf{i}\omega k} \right) \\ &= f_0 + \sum_{k=1}^{\tau/2} 2\mathcal{R}(f(k)e^{-\mathbf{i}\omega k}) = f_0 + 2\sum_{k=1}^{\tau/2} |f(k)| \cos(\omega k + \theta_k), \end{aligned} \tag{19}$$

where $f_0 = f(k = 0)$. Since $\omega_j - \omega_{j-1} = \Delta\omega = \frac{2\pi}{N}$,

$$|F(\omega_j) - F(\omega_{j-1})| \leq \frac{2\pi}{N} \max_{\omega} |\frac{d}{d\omega} F(\omega)| \tag{20}$$

Combining the last two equations gives

$$|F(\omega_j) - F(\omega_{j-1})| \leq \frac{4\pi}{N} \sum_{k=1}^{\tau/2} |f(k)|k$$

Finally, by the Cauchy–Schwartz inequality,

$$|F(\omega_j) - F(\omega_{j-1})| \leq \frac{4\pi}{N} \sqrt{\sum_{k=1}^{\tau/2} |f(k)|^2} \sqrt{\sum_{k=1}^{\tau/2} k^2} \tag{21}$$

Next, using Parseval's theorem

$$\sum_{k=1}^{\tau/2} |f(k)|^2 = \frac{1}{2} \sum_{k=-\tau/2}^{\tau/2} |f(k)|^2 - \frac{1}{2}|f_0|^2 = \frac{\|\mathbf{F}\|^2}{2N} - \frac{1}{2}|f_0|^2 \leq \frac{\|\mathbf{F}\|^2}{2N} \tag{22}$$

Combining Eqs. (21)–(22) yields Eq. (18) which concludes the proof. □

Lemma 2 implies that the sign of $\mathbf{F}$ must be equal at any consecutive indices $\{j-1, j\}$ that satisfy

$$|F_j| + |F_{j-1}| > \left(\frac{2}{N}\right)^{3/2} \pi \mathcal{S}_{\tau/2} \|\mathbf{F}\| \tag{23}$$

Eq. (23) thus provides an over-segmentation to constant sign intervals that depends only on the measured vector $|\mathbf{F}|$. If the number of found segments M is sufficiently small, so that $N > 2\tau + M$, we can then directly recover the sign pattern by solving the system of linear equations (15)–(16).

Remark 5. With a sufficiently high oversampling rate, the number of segments determined by Eq. (23) satisfies $N > 2\tau + M$, which in turn guarantees recovery of the true sign pattern by solving equations (15)–(16). To see this, note that as we increase the number of measurements N while keeping τ fixed, $\|\mathbf{F}\|$ increases as $O(\sqrt{N})$ and hence the threshold on the right hand side of Eq. (23) decreases as $O(1/N)$. Thus, for sufficiently large N, a sufficient number of pairs $(j-1, j)$ satisfy Eq. (23).

Unfortunately, even though Eq. (23) provides a correct over-segmentation, with a finite over-sampling rate the resulting number of segments may be too large, so the corresponding Eqs. (15)–(16) have multiple solutions. To decrease the number of segments, and thus increase the number of linear equations, we

Table 1
Heuristic over-segmentation scheme.

```
Algorithm Heuristic over-segmentation

Input: $|\mathbf{F}|^2$.
Algorithm:
1: Find the local minima of $|\mathbf{F}|$ given by the indices $j$ s.t. $|F_j| < \min\{|F_{j-1}|, |F_{j+1}|\}$.
2. Define all minima indices as single index segments.
3. Define the indices between each pair of consecutive minima indices as a constant sign interval.
4: For each minima index, find the adjacent index with the closest value of $|\mathbf{F}|$ value and exclude it from the constant sign interval, i.e. define it as a single index segment.
Output: Over-segmentation to constant sign intervals.
```

supplement Eq. (23) with a heuristic segmentation scheme which works well in practice, even though it is not theoretically guaranteed to yield a correct over-segmentation. Our heuristic segmentation scheme relies on the fact that for a sign change to occur between F_j and F_{j+1} its underlying continuous function $F(\omega)$ must have a zero crossing at some intermediate $\omega \in [\omega_j, \omega_{j+1}]$. Hence, $|\mathbf{F}|$ is likely to have a local discrete minimum near this zero crossing.

This observation leads to the following segmentation algorithm summarized in Table 1: First, the discrete local minima of $|\mathbf{F}|$ are found and their indices are defined as single index segments. Next, all the indices between each pair of consecutive minima are defined to have the same sign. Finally, for each minimum index, the adjacent index for which $|\mathbf{F}|$ has closer value to the minimum value is also defined as a single index segment. The last step is applied to reduce errors in the resulting segmentation. In the absence of noise, our final segmentation is the merging of both segmentations described above. Fig. 9 in Section 6.6 presents an example of the segmentations produced by both schemes.

Note that this heuristic algorithm assumes that sign changes occur at local minima of the intensity. Unfortunately, this is not always the case – sign flips may occur at points with small intensity which are not a local minimum. This leads to indices with small intensity assigned to wrong segments. Although the reconstructed signs at these indices are wrong, their small amplitude yield a relatively small reconstruction error in the signal $\mathbf{f}$.

4.3. Sign retrieval from noisy measurements

In realistic experimental scenarios the vector of intensities $|\mathbf{F}|^2$ is measured with some noise. In this case, no sign pattern satisfies Eq. (15). To cope with measurement noise, we reformulate the sign problem as the minimization of a suitable quadratic functional. Below, we first describe our assumed noise model, then construct the functional to be minimized and detail our minimization approach.

Let $|\tilde{\mathbf{F}}|^2$ be the vector of noisy measurements. As described in [28], a rather general noise model is

$$|\tilde{\mathbf{F}}|^2 = |\mathbf{F} + \tfrac{\sigma}{\sqrt{N}}\boldsymbol{\eta}^s|^2 + |\mathbf{F} + \tfrac{\sigma}{\sqrt{N}}\boldsymbol{\eta}^s|\boldsymbol{\eta}^{sh} + \boldsymbol{\eta}^d, \tag{24}$$

where $\boldsymbol{\eta}^s$ is background additive noise with noise level σ, $\boldsymbol{\eta}^d$ is dark counts noise, and $|\mathbf{F} + \frac{\sigma}{\sqrt{N}}\boldsymbol{\eta}^s|\boldsymbol{\eta}^{sh}$ is the detector shot noise which is proportional to the signal intensity. In many cases the dark counts $\boldsymbol{\eta}^d$ has small variance and its main effect can be removed by proper calibration. We thus assume for simplicity that $\boldsymbol{\eta}^d = 0$. We assume a classical light experiment, far from the single photon regime and we thus neglect the effect of shot noise. We further assume that $\boldsymbol{\eta}^s = (\eta_0^s, \ldots, \eta_{N-1}^s)$ consists of N independent and identically distributed (i.i.d.) complex-valued Gaussian random variables $\mathcal{N}(0, 1)$. Since the noise level can be accurately estimated in most experimental realizations e.g. by calibration measurements, we assume that σ is known.

To handle noisy measurements, we modify our sign retrieval scheme in two ways: (i) refine the method to find an oversegmentation; (ii) replace the homogeneous linear system by the minimization of a quadratic functional. Given noisy measurements, our sign retrieval consists of the following steps.

Step 1: Find an over-segmentation. First we apply the heuristic segmentation scheme defined in Table 1. We denote by $\mathcal{S}_1$ the set of indices of size S_1 in the resulting segmentation, excluding the last index in each segment. Next, we modify Eq. (23) as detailed in the Appendix to yield,

$$|F_j| + |F_{j-1}| > \left(\frac{2}{N}\right)^{3/2} \pi \mathcal{S}_{\tau/2} \|\tilde{\mathbf{F}}\| + \sigma/\sqrt{N}, \tag{25}$$

where σ is the standard deviation of the noise. We denote by $\mathcal{S}_2$ the set of indices of size S_2 in the segmentation defined by Eq. (25), excluding the last index in each segment.

Step 2: Construct and minimize a quadratic functional. Given the noisy measurements $|\tilde{\mathbf{F}}|^2$ and the support parameter τ, we define the matrix $\mathcal{A}_{cs}$ of size $(N-\tau-1) \times N$ as

$$(\mathcal{A}_{cs})_{k,j} = (DFT \cdot \mathcal{D})_{k,j}, \quad k \notin CS, \ j \in [1, N], \tag{26}$$

where $(DFT)_{j,k} = \frac{1}{N} e^{\mathbf{i}\omega_{j-1}(k-1)}$ with $j, k = 0, \dots, N-1$ is the discrete Fourier transform matrix, $\mathcal{D} = diag(|\tilde{\mathbf{F}}|)$ is a diagonal matrix and the set CS is given by Eq. (5). In the noise free case, $\mathcal{A}_{cs} X = 0$ imposes that $\mathcal{DFT}\{|\tilde{\mathbf{F}}|\mathbf{X}\}$ vanishes for $k \notin CS$ and is precisely Eq. (15).

Next, we impose our two segmentation schemes starting with the heuristic one. We use $\mathcal{S}_1$ to construct the matrix $\mathcal{A}_1$ of size $S_1 \times N$, whose kth row is given by

$$(\mathcal{A}_1)_{k,j} = \begin{cases} W_k & j = l_k \\ -W_k & j = l_k + 1 \qquad l_k \in \mathcal{S}_1 \\ 0 & \text{otherwise}, \end{cases} \tag{27}$$

where $k \in [1, S_1]$, l_k is the k-th index in $\mathcal{S}_1$, and the vector of weights $\mathbf{W} = [W_1, \dots, W_{S_1}]$ is defined as $W(k) = \min\{|\tilde{F}_{l_k}|^2, |\tilde{F}_{l_k+1}|^2\}$. We note that $\mathcal{A}_1 \mathbf{X} = 0$ imposes that $\mathbf{X}$ is piecewise constant in the intervals defined by Table 1. The purpose of $\mathbf{W}$ is to account for the fact that the heuristic algorithm presented in Table 1 does not guarantee a correct over-segmentation and that errors typically occur at indices corresponding to low $|\tilde{\mathbf{F}}|^2$ values where we have less certainty in our segmentation. With our weighting approach, we have that $\mathcal{A}_1 \mathbf{X} = 0$ imposes $W_j(X_j - X_{j+1}) = 0$ for $j \in \mathcal{S}_1$. Hence, when a quadratic functional based on $\mathcal{A}_1$ is minimized, as described below, the constraint that $\mathbf{X}$ is piecewise constant at $j \in \mathcal{S}_1$ is suppressed at indices with low $|\tilde{\mathbf{F}}|^2$ values.

Our second segmentation approach is guaranteed to yield a correct over-segmentation in the noise-free case, and in practice for high SNR, it is unlikely to make errors. Hence, we impose this segmentation without weights, in a way that also reduces the computational cost of our problem by reducing the number of variables. To this end, we sum over the appropriate columns of the $(N + S_1 - \tau - 1) \times N$ matrix $[\mathcal{A}_{cs}; \mathcal{A}_1]$ to construct the matrix $\mathcal{A}$ of size $(N + S_1 - \tau - 1) \times (N - S_2)$, according to

$$\mathcal{A}_{k,m} = \sum_{j \in \text{segment m}} ([\mathcal{A}_{cs}; \mathcal{A}_1])_{k,j}, \ 1 \leq m \leq M, \ k \in [1, N + S_1 - \tau - 1]. \tag{28}$$

Our approach is to find a vector $\mathbf{X}$ which minimizes the following quadratic functional,

$$Q(\mathbf{X}) = \|\mathcal{A}\mathbf{X}\|^2. \tag{29}$$

Since the required output is a sign vector, in principle Eq. (29) should be minimized over $\{-1, 1\}^N$, which results in a non-convex problem. Here we relax this constraint and instead minimize $Q(\mathbf{X})$ over vectors

Table 2
Sign retrieval algorithm.

```
Algorithm Sign retrieval

Input: $|\tilde{\mathbf{F}}|^2$, $\tau$, $\sigma^2$.
Algorithm:
1: Compute the sets of indices $\mathcal{S}_1$ and $\mathcal{S}_2$ according
   to Table 1 and Eq. (25) respectively.
2: Construct the matrix $\mathcal{A}$.
3. Compute the minimizer $\hat{\mathbf{X}}$ of $\|\mathcal{A}_{-\nu} - \mathbf{X}_{-\nu} - a_\nu\|^2$.
4. Project $\hat{\mathbf{X}}$ onto a sign: $\hat{\mathbf{s}} = sign(\mathcal{R}(\hat{\mathbf{X}}))$.
Output: Estimated sign vector: $\hat{\mathbf{s}}$
```

$\mathbf{X} \in \mathbb{C}^N$. To avoid the zero solution and given the ±1 global sign ambiguity, as in [28] we set $X(\nu) = -1$ for some index $\nu \in \{0, \dots, N-1\}$. The functional $Q(\mathbf{X})$ of Eq. (29) then becomes

$$Q(\mathbf{X}) = \|\mathcal{A}_{-\nu}\mathbf{X}_{-\nu} - a_\nu\|^2, \tag{30}$$

where a_ν is the νth column of $\mathcal{A}$, $\mathcal{A}_{-\nu}$ is the matrix $\mathcal{A}$ without a_ν and $\mathbf{X}_{-\nu}$ is the vector $\mathbf{X}$ without its νth entry. We choose ν as the index for which $|\tilde{\mathbf{F}}|^2$ is maximal. Minimizing $Q(\mathbf{X})$ amounts to the convex problem of solving a set of linear equations without constraints. Moreover, in the noise-free case, minimizing $Q(\mathbf{X})$ is equivalent to solving the set of linear equations

$$\mathcal{A}\mathbf{X} = 0, \tag{31}$$

which effectively imposes Eqs. (15)–(16) with $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$. Therefore, under the assumptions of Theorem 3 the minimizer of $Q(\mathbf{X})$ is the true sign vector $\mathbf{s}$ (up to multiplication by a constant). As demonstrated in section 6.2, empirically this approach is robust to noise and, in practice, even tolerates a small number of segmentation errors. Table 2 summarizes our algorithm.

4.4. Estimating the support

In many practical cases, the support parameter τ of $\mathbf{f}$ is only known to be bounded between some a-priori known values τ_{min} and τ_{max}. Here we propose an algorithm to estimate it from the (possibly noisy) measurements $|\tilde{\mathbf{F}}|^2$. To estimate τ, we scan over the possible values $\tau_{min} \le \tau_s \le \tau_{max}$. For each value τ_s we retrieve the sign pattern $\hat{\mathbf{s}}$ according to the algorithm in Table 2, compute the corresponding signal $\hat{\mathbf{f}} = \mathcal{DFT}\{|\tilde{\mathbf{F}}|\hat{\mathbf{s}}\}$ and its average energy outside of the assumed support,

$$E_{out}(\tau_s) = \frac{1}{N - \tau_s - 1} \sum_{k \notin CS(\tau_s)} |\hat{f}_k|^2. \tag{32}$$

As detailed in Table 3, our estimate $\hat{\tau}$ of τ is the smallest τ_s where $E_{out}(\tau_s)$ attains its minimal value.

To justify this approach let us first analyze the noise-free case. Here, for $\tau_s = \tau$, as proven in Section 4.1 our algorithm perfectly recovers the true signal $\mathbf{f}$. Hence, at the correct support, $E_{out}(\tau) = 0$. For $\tau_s < \tau$ there is no signal $\mathbf{f}$ with support parameter τ_s whose DFT has a piecewise constant-phase. Namely, there is no vector $\mathbf{X}$ which gives $Q(\mathbf{X}) = 0$ in Eq. (29). Minimizing this quadratic functional gives as output some signal $\mathbf{f}$ which does not vanish outside the assumed support, as otherwise this would contradict Theorem 3. Hence, for any $\tau_s < \tau$, $E_{out}(\tau_s) > 0$. For $\tau_s > \tau$, the only solution is the true sign vector $\mathbf{s}$. In the noise-free case, our scheme indeed recovers the correct support, $\hat{\tau} = \tau$.

In the presence of noise, even at $\tau_s = \tau$, the recovered signal is noisy and $E_{out}(\tau) > 0$. As τ_s increases above τ the number of linear equations in Eq. (31) decreases. Hence, the sensitivity to noise of the corresponding

Table 3
Compact support estimation.

```
Algorithm Compact support estimation

Input: $|\tilde{\mathbf{F}}|^2$, $\tau_{min}$, $\tau_{max}$, $\sigma^2$.
Algorithm:
1: For $\tau_s = \tau_{min}$ to $\tau_{max}$ do:
        a: Retrieve the sign of $\mathbf{F}$ using Table 2 with support of $\tau_s + 1$.
        b: For the retrieved sign $\hat{\mathbf{s}}$, compute $\hat{\mathbf{f}} = \mathcal{DFT}\{|\tilde{\mathbf{F}}|\hat{\mathbf{s}}\}$.
        c: Compute $E_{out}(\tau_s) = \sum_{k \notin CS} |\hat{f}(k)|^2/(N - \tau_s - 1)$.
2. Estimate $\tau$ by $\hat{\tau} = \underset{\tau_s}{\operatorname{argmin}}\{E_{out}(\tau_s)\}$.
Output: Estimated support: $\hat{\tau} + 1$
```

solution increases, which in turn leads to an increase in $E_{out}(\tau_s)$ as a function of τ_s. At low noise levels, our approach is thus still able to correctly estimate the true support parameter.

5. Phase retrieval applications of the sign problem

We now present two phase retrieval settings of practical interest in which the sign problem plays a key role. The first is vectorial phase retrieval (VPR) [27,28] but with only 3 measurements, and the second is phase retrieval from two sufficiently separated objects [14,21].

5.1. VPR with 3 measurements

VPR consists of a recently suggested family of physically feasible measurement schemes together with computationally efficient methods to recover the phase. The VPR problem is to recover two signals $\mathbf{f}_1, \mathbf{f}_2 \in \mathbb{C}^N$ with corresponding Fourier transforms $\mathbf{F}_1$ and $\mathbf{F}_2$ from the following measurements,

$$|\mathbf{F}_1|, \quad |\mathbf{F}_2|, \quad \mathbf{F}_1\mathbf{F}_2^*. \tag{33}$$

As proven in [28], under suitable conditions, in particular both signals $\mathbf{f}_1, \mathbf{f}_2$ having sufficiently small supports, this phase problem admits a unique solution. Furthermore, the phase vectors $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{C}^N$ that correspond to $\mathbf{F}_1$ and $\mathbf{F}_2$ can be uniquely retrieved by solving the following set of linear equations

$$\begin{aligned} &\mathcal{DFT}\{|\mathbf{F}_1|\mathbf{X}_1\}(k) = 0, \;\; k \notin CS \\ &\mathcal{DFT}\{|\mathbf{F}_2|\mathbf{X}_2\}(k) = 0, \;\; k \notin CS \\ &(\mathbf{F}_1\mathbf{F}_2^*)\mathbf{X}_2 = |\mathbf{F}_1||\mathbf{F}_2||\mathbf{X}_1. \end{aligned} \tag{34}$$

In [28], several physical scenarios were described where the following 4 vectors can be measured, $|\mathbf{F}_1|$, $|\mathbf{F}_2|$, $|\mathbf{F}_1 + \mathbf{F}_2|$ and $|\mathbf{F}_1 + \mathbf{i}\mathbf{F}_2|$. From these measurements the interference term $\mathbf{F}_1\mathbf{F}_2^*$ can be easily computed as $\frac{1}{2}\big(|\mathbf{F}_1 + \mathbf{F}_2|^2 + \mathbf{i}|\mathbf{F}_1 + \mathbf{i}\mathbf{F}_2|^2 - (1 + \mathbf{i})(|\mathbf{F}_1|^2 + |\mathbf{F}_2|^2)\big)$. Then, the phase is retrieved by solving Eq. (34). However, in various physical settings obtaining the fourth measurement, $|\mathbf{F}_1 + \mathbf{i}\mathbf{F}_2|$, is difficult or impossible and only the following three spectra can be measured

$$|\mathbf{F}_1|, \quad |\mathbf{F}_2|, \quad \mathbf{S} = |\mathbf{F}_1 + \mathbf{F}_2|. \tag{35}$$

As we now show, using sign retrieval, these three measurements allow to recover $\mathbf{F}_1\mathbf{F}_2^*$, the interference term required to apply VPR to solve the phase problem.

It was recently proven in [6], that the phase problem corresponding to Eq. (35) admits a unique solution if $\mathbf{f}_1$ and $\mathbf{f}_2$ have a sufficiently small support, but without suggesting a possible algorithm to solve it. We propose the following scheme: Given the three vectors of Eq. (35), compute

$$\mathbf{E}_R = \frac{1}{2}(\mathbf{S}^2 - |\mathbf{F}_1|^2 - |\mathbf{F}_2|^2) = |\mathbf{F}_1||\mathbf{F}_2|\cos(\phi_{12}) = \mathcal{R}(\mathbf{F}_1\mathbf{F}_2^*), \tag{36}$$

where $\phi_{12} = arg(\mathbf{F}_1\mathbf{F}_2^*)$. Since $\mathbf{E}_R = \mathcal{R}\big[\mathbf{F}_1\mathbf{F}_2^*\big]$ its $\mathcal{DFT}$ has support of length at most twice that of $\mathbf{f}_1, \mathbf{f}_2$. This is also true for the unknown corresponding imaginary part $\mathbf{E}_I = \mathcal{I}\big[\mathbf{F}_1\mathbf{F}_2^*\big]$. Moreover, the absolute value of $\mathbf{E}_I$ can be calculated from Eq. (35),

$$|\mathbf{E}_I| = \sqrt{|\mathbf{F}_1|^2|\mathbf{F}_2|^2 - \mathbf{E}_R^2}.$$

To recover $\mathbf{F}_1\mathbf{F}_2^*$ we thus need to solve the following sign problem: Given the absolute value $|\mathbf{E}_I|$ of the imaginary part of the interference term $\mathbf{E}_I$, whose $\mathcal{DFT}$ has a support of length τ, retrieve $sign(\mathbf{E}_I)$. For sufficiently small τ (sufficiently high over-sampling), we can thus use our sign retrieval algorithm to retrieve this sign. This immediately allows computation of $\mathbf{E}_R + \mathbf{i}\mathbf{E}_I = \mathbf{F}_1\mathbf{F}_2^*$ which, together with Eq. (35) completes the required input for VPR depicted in Eq. (33). Applying this scheme, we demonstrate VPR reconstructions with 3 measurements in Section 6.4.

5.2. Phase retrieval from separated objects

A second phase retrieval scenario in which the sign problem arises is the reconstruction of two objects that are well-separated, by more than the length of the larger support. Let $\mathbf{f} \in \mathbb{C}^N$ be of the form $\mathbf{f} = \mathbf{f}_1 + \mathbf{f}_2$ where $\mathbf{f}_1$ and $\mathbf{f}_2$ have small supports and are well-separated from each other. In [14], it was shown that despite this being a 1-D phase problem, under suitable conditions, the signal $\mathbf{f}$ is uniquely determined by the single measurement vector $|\mathbf{F}| = |\mathcal{IDFT}\{\mathbf{f}\}| = |\mathbf{F}_1 + \mathbf{F}_2|$.

Given the above uniqueness result, the goal here is to reconstruct $\mathbf{f}$ from the single measurement $|\mathbf{F}|^2$. Our approach is to use sign retrieval as a step to recover the input required for VPR from $|\mathbf{F}|^2$ and then apply VPR to retrieve $\mathbf{f}$. Since $\mathbf{f}_1$ and $\mathbf{f}_2$ are well-separated, we have that $\mathcal{DFT}\left\{|\mathbf{F}_1 + \mathbf{F}_2|^2\right\}$ yields 3 separated terms. The central term is the sum of the autocorrelations of $\mathbf{f}_1$ and $\mathbf{f}_2$, given by $\mathbf{f}_1 \star \mathbf{f}_1 + \mathbf{f}_2 \star \mathbf{f}_2$. The two other terms are their cross-correlation, $\mathbf{f}_1 \star \mathbf{f}_2$ and its complex conjugate. Performing an inverse $\mathcal{DFT}$ separately on each of these terms, yields the following equations

$$\mathbf{I}_S = |\mathbf{F}_1|^2 + |\mathbf{F}_2|^2, \quad \mathbf{E}_3 = \mathbf{F}_1\mathbf{F}_2^*. \tag{37}$$

In order to apply VPR we first need to resolve the vectors $|\mathbf{F}_1|^2$ and $|\mathbf{F}_2|^2$ from Eq. (37). To this end, consider the (unknown) function

$$\mathbf{I}_D = |\mathbf{F}_1|^2 - |\mathbf{F}_2|^2. \tag{38}$$

Since $\mathbf{I}_D$ is the difference between the Fourier intensities of $\mathbf{f}_1$ and $\mathbf{f}_2$, its Fourier transform has a support of size at most twice that of $\mathbf{f}_1, \mathbf{f}_2$. Also, its absolute value can be computed from Eq. (37) as,

$$|\mathbf{I}_D| = \sqrt{\mathbf{I}_S^2 - 4|\mathbf{E}_3|^2}.$$

This gives rise to the following sign problem: Given $|\mathbf{I}_D|$ with $\mathcal{DFT}\{\mathbf{I}_D\}$ having a small support, retrieve $sign(\mathbf{I}_D)$. Once the sign is retrieved $|\mathbf{F}_1|^2$ and $|\mathbf{F}_2|^2$ can be immediately computed from their sum and difference which together with $\mathbf{F}_1\mathbf{F}_2^*$ gives the required input of Eq. (33). The underlying two signals $\mathbf{f}_1$ and $\mathbf{f}_2$ can be now recovered by applying VPR.

In Section 6.5 below we numerically demonstrate this scheme. For its application in the reconstruction of separated 2D objects from experimental X-ray free electron laser measurements, see [21].

6. Simulations

We illustrate the performance of our algorithm via several simulations. First, we consider noise-free and noisy reconstructions of complex-valued random signals with support length $\tau + 1$. Next, we consider the continuous sign problem, and compare our method to the one suggested by Thakur [33]. Finally, we apply our sign retrieval algorithm to the two phase retrieval applications described in Section 5, VPR with 3 measurements and phase retrieval for two well-separated 1-D objects.

Given the global ±1 ambiguity of the sign problem, we measure the reconstruction quality by the following mean-square-error (MSE),

$$\mathrm{MSE} = \min\big(\sum_k |\hat{f}_k - f_k|^2,\ \sum_k |\hat{f}_k + f_k|^2\big). \tag{39}$$

6.1. Noise-free reconstruction

We demonstrate that with noise-free measurements and sufficient over-sampling, our sign retrieval algorithm perfectly recovers the underlying signal, essentially with machine-precision error. To this end, we generated a complex-valued random signal of length $N = 500$ with support parameter $\tau = 100$, whose Fourier transform is real-valued. We applied our sign retrieval algorithm including a scan over the unknown support, as described in Section 4. The right panel of Fig. 1 shows, on a logarithmic scale, the residual energy outside the assumed support, as described in Section 4.4. In accordance with our theoretical analysis, the residual error sharply drops precisely at the correct support parameter $\tau_s = 100$ and remains at essentially zero value (up to machine error) for all values $\tau_s \geq \tau$. Our scheme thus correctly estimates $\hat{\tau} = \tau = 100$, and as shown in the left panel of Fig. 1, at this estimated support, our algorithm recovers correctly the exact sign pattern, leading to a zero MSE.

In the above example, our algorithm found a correct over-segmentation with not too many segments and thus resulted in perfect reconstruction. Even with noise-free measurements, this is not always the case. The left panel of Fig. 2 shows the number of sign errors, averaged over 100 random realizations as a function of the support parameter τ, with a fixed signal length $N = 500$, whereas the right panel shows the corresponding averaged MSE. In accordance with Remark 5, for small values of τ (e.g., a high oversampling rate), our segmentation scheme almost always obtains a perfect recovery. In contrast, as τ increases, the exact segmentation does not yield a sufficient number of equations, and the heuristic segmentation may make small errors. These, however, typically occur at indices with small values $|F_j|$, which as seen in the right panel lead to small reconstruction errors.

6.2. Reconstruction in the presence of noise

Next, we demonstrate the robustness of our algorithm to noise. First, we show the resulting reconstructions from noisy measurements, where an uncorrelated Gaussian noise with $\sigma = 0.03$ was added to the same signal as in Section 6.1, normalized to $\|\mathbf{F}\| = 1$. As shown in Fig. 3(left) the reconstruction is in very good agreement with the true signal. The scan over E_{out} for different values of the support depicted in Fig. 3(right), shows that in contrast to the noise-free case, E_{out} increases as the support increases above its true value. This occurs because a larger assumed support yields a smaller number of linear equations which in turn increases the sensitivity to noise of the minimization problem described in Section 4.3.

Next, we study the effect of noise on our reconstructions by Monte Carlo simulations for different support lengths, $\tau = 20, 100, 140, 200$, while keeping the oversampling ratio constant at $N = 5\tau$. For each τ, we computed the average MSE from the reconstructions of 10 randomly generated, complex-valued signals, each with 100 noise realizations. The results, presented in Fig. 4, show that the reconstructions are stable

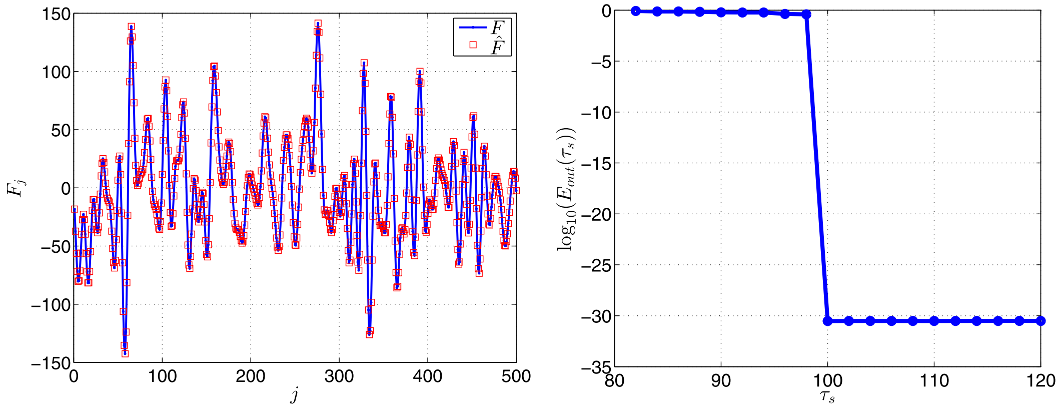

Fig. 1. Reconstruction of a signal with support $\tau = 100$, from $N = 500$ noise-free measurements. (Left) The reconstructed versus true signal; (right) the residual energy outside the support, on a log-scale, as a function of the unknown support. Our algorithm correctly estimates $\hat{\tau} = 100$, and perfectly recovers the unknown sign pattern, leading to an essentially zero MSE.

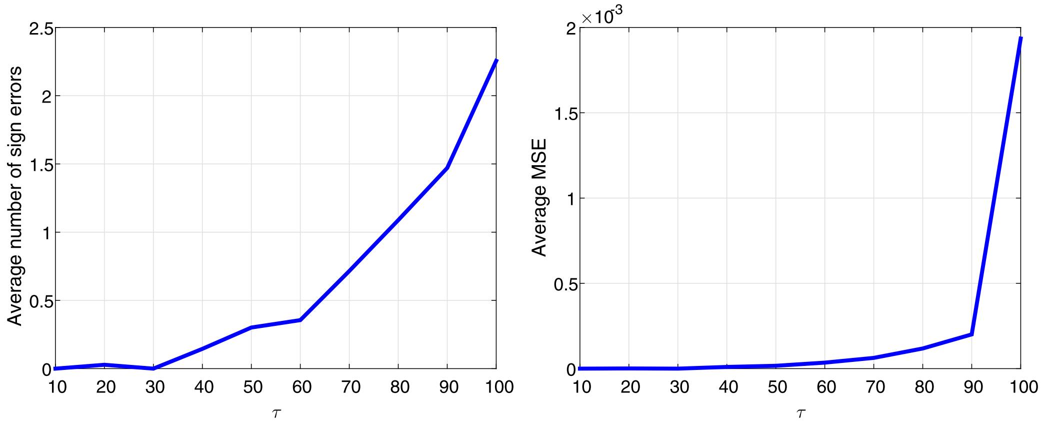

Fig. 2. Reconstruction of a random signal as a function of the support parameter τ with $N = 500$. (Left) Average number of sign errors over 100 realizations. (Right) Average MSE.

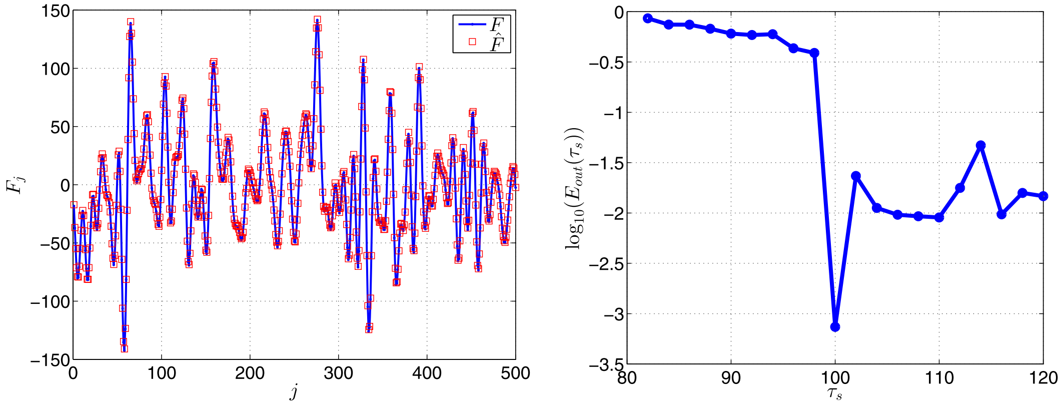

Fig. 3. Reconstruction of the same signal as in Fig. 1 in the presence of noise at level $\sigma = 0.03$. (Left) Signal reconstruction. (Right) Residual error vs. assumed support τ_s.

up to $\sigma \simeq 0.01$ even for $\tau = 200$. Note that for $\tau = 20$, the segmentation scheme makes almost no errors, and thus the MSE increases linearly with σ (on a log–log scale). For higher values of τ, the probability for small segmentation errors increases, leading to nearly constant MSE for small noise levels. Possible routes to improve the noise stability are discussed in Section 7.

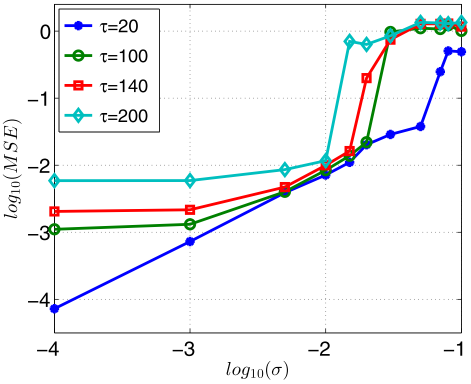

Fig. 4. Reconstruction error versus noise level (log–log scale).

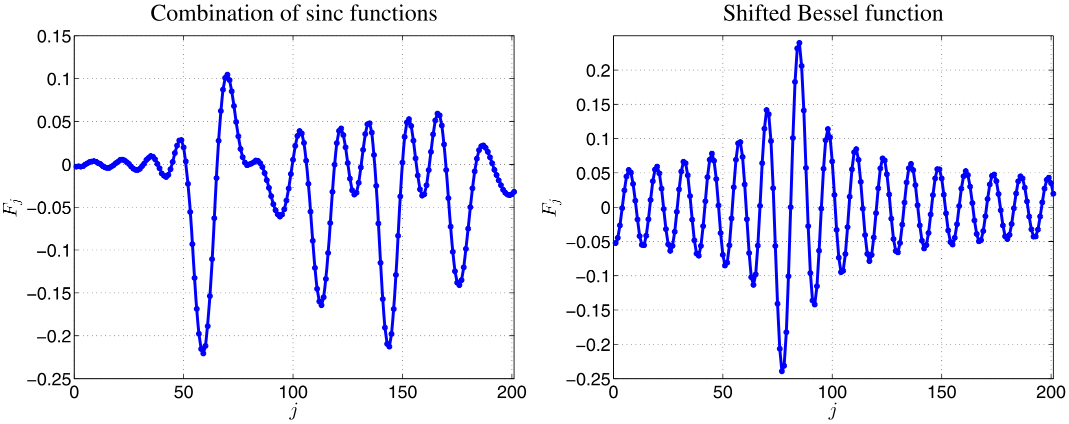

Fig. 5. Two different real-valued continuous signals with band-limited Fourier Transform.

6.3. Reconstruction of a continuous function

As mentioned in the introduction, the discrete sign problem considered in this paper can be viewed as the finite dimensional analogue of the following continuous problem: Recover a real-valued function $g(t)$ whose continuous Fourier transform $G(\omega)$ is band-limited, from discrete measurements $|g(t_j)|$. As in the discrete case, the key challenge in the continuous problem is to locate the zero crossings of the function $g(t)$, from the values $|g(t_j)|$. This problem was recently studied by Thakur [33], who considered the analytic continuation of $g(t)$ to the complex valued plane, and developed a method to reconstruct the function $g(t + \mathbf{i}c)$, for some fixed $c \in \mathbb{R}$. In contrast, our approach, beyond being based on a discrete formulation, instead narrows down the possible locations of these zero crossings via an over-segmentation to intervals of constant sign. This different approach avoids the rather unstable operation of extending the function into the complex plane, and projecting it back to the real line.

We now illustrate the applicability of our discrete-formulation based algorithm to the continuous sign problem. Following Thakur [33], we sample at $N = 200$ equispaced points the absolute $|F_c(\omega_j)|$ of a real-valued continuous function $F_c(\omega)$ whose Fourier transform $f_c(t)$ has a compact support. Note that this scenario is different from our previous investigations, since the discrete Fourier transform of $F_c(\omega_j)$ is *not* compactly supported. It is thus interesting to see if our algorithm can still succeed in recovering the sign pattern.

We compare our algorithm to [33] via Monte Carlo simulations at several noise levels, $\sigma = (0, 1, 2, 3, 4) \cdot 10^{-3}$. We consider two signals, a shifted Bessel function as in [33] and depicted in Fig. 5(right), and a linear combination of 10 randomly shifted, randomly weighted sinc functions, depicted in Fig. 5(left). All signals

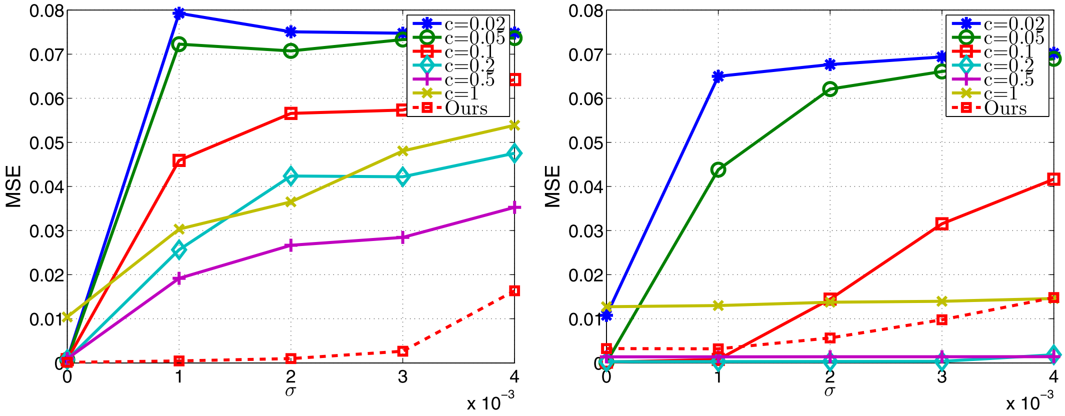

Fig. 6. MSE comparison of Thakur's method with different values of c and ours as a function of noise level. (Left) combination of sinc functions; (right) Shifted Bessel function.

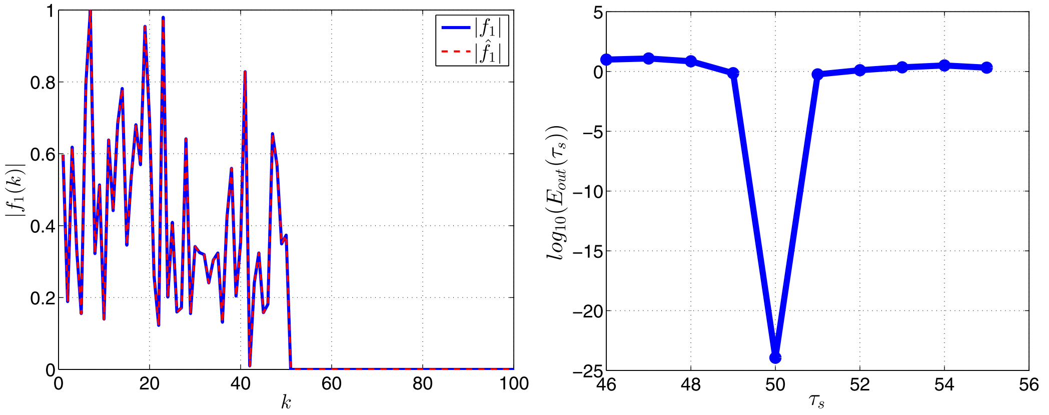

Fig. 7. VPR with 3 measurements. (Left) Normalized magnitude of the reconstructed and true signal. (Right) Scan over averaged energy outside of the support.

were normalized to have unit norm prior to adding Gaussian uncorrelated noise. Since Thakur's method depends on the above mentioned parameter c, following the recommendation in [33], we considered the following six values $c = 0.02, 0.05, 0.1, 0.2, 0.5, 1$. Fig. 6(left) compares the two methods for the randomly shifted sinc functions. In this case, Thakur's method exhibited poor sensitivity to noise and its MSE rapidly increased with noise level for all considered values of c. In contrast, our method achieved a significantly lower error. For the shifted Bessel function, presented in Fig. 6(right), Thakur's method achieved a lower MSE than ours for $c = 0.2, 0.5$. However, we note that the choice of the optimal c value is not a-priori known, and may depend on the signal to be reconstructed. Also, as we verified in additional simulations, with a higher sampling rate, our method achieved a lower MSE than Thakur's.

6.4. VPR with 3 measurements

Next, we demonstrate the use of sign retrieval with VPR to reconstruct two unknown complex-valued random signals, $\mathbf{f}_1$ and $\mathbf{f}_2$ both with $N = 150$ and a support parameter $\tau = 50$, from 3 measurements $|\mathbf{F}_1 + \mathbf{F}_2|^2$, $|\mathbf{F}_1|^2$ and $|\mathbf{F}_2|^2$, as described in Section 5.1. Fig. 7(left) shows the result for a noise-free reconstruction. For simplicity, we present the magnitudes of the complex-valued true and reconstructed signal, $\mathbf{f}_1$ (the results for $\mathbf{f}_2$ are similar). The reconstruction is perfect, up to machine error. In this reconstruction we do not assume that the support size is known. Instead, we scan over its possible values, similarly to Section 4.4 and [28]. A plot of the average energy outside of the support is presented in Fig. 7(right). Note that in contrast to

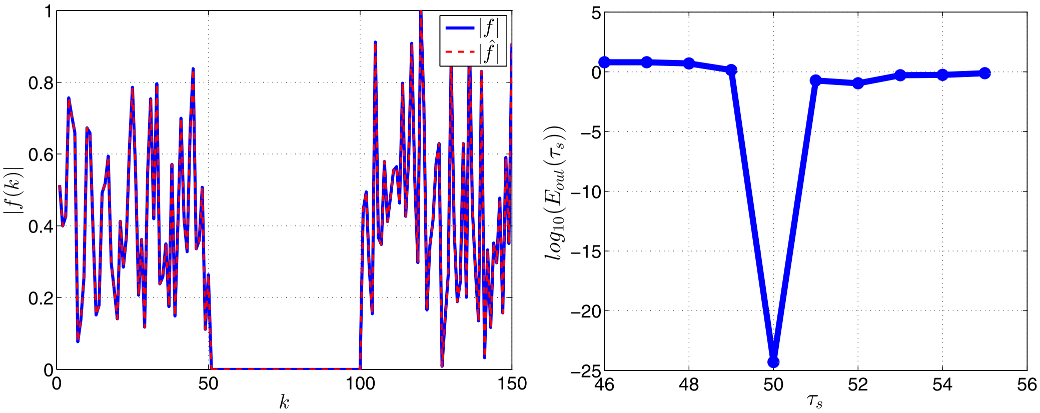

Fig. 8. Phase retrieval with separated objects. (Left) Normalized magnitude of the reconstructed and true signal. (Right) Scan over averaged energy outside of the support.

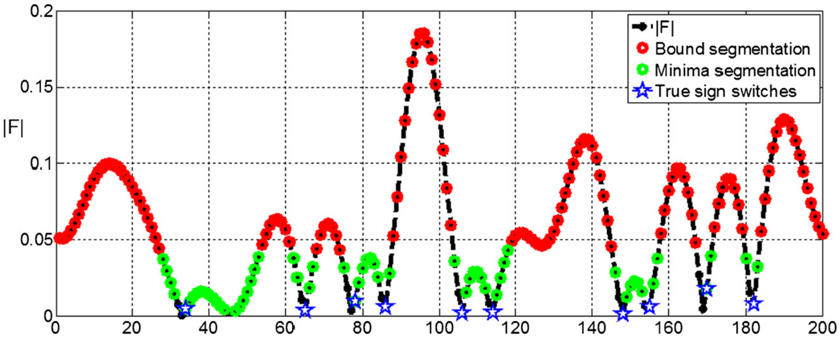

Fig. 9. Illustration of the segmentation scheme. The black line denotes the absolute value of the signal $|F(\omega_j)|$. The red circles mark the regions guaranteed to have a common sign according to Lemma 2. The green circles mark the additional regions that have common sign according to our heuristic segmentation scheme. The blue stars denote the true sign switching indices. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the sign problem here a clear minimum of E_{out} is obtained at the correct value of the support. The reason for this behavior is that for the phase problem, as opposed to the sign problem, shifted solutions are allowed as detailed in [28].

6.5. Phase retrieval of two separated objects

Here we demonstrate 1D phase retrieval of a complex-valued vector consisting of two well separated signals, from a single measurement. This is achieved by combining sign retrieval with VPR as described in Section 5.2. To this end, we generated a signal $\mathbf{f}$ of length $N = 500$ and support length 151, which consists of two random complex-valued vectors of length 50 separated by 51 zero entries. In the absence of noise, the signal is perfectly reconstructed (to within machine error), as demonstrated in Fig. 8(left). As in section 6.4, we did not assume here a known support but rather estimated it as a part of the algorithm. The scan over τ_s for different support values is presented in Fig. 8(right). As in Section 6.4, also here a clear minimum of E_{out} is attained at the true support value.

6.6. Segmentation scheme

Fig. 9 illustrates our segmentation scheme of Section 4.2 in the noise-free case. The indices in which sign changes occur are denoted by blue stars. The red circle denote indices for which Eq. (23) holds. The green circles denote indices for which Eq. (23) does not hold, but are assigned to constant sign intervals according to the heuristic segmentation scheme of Table 1.

7. Discussion

In this paper we presented a theoretical study of the finite dimensional sign problem and developed a novel computationally efficient sign retrieval algorithm. We then showed its applicability to two phase retrieval problems of practical interest, VPR with 3 measurements and reconstruction of two well separated objects.

Our sign problem, which can be formulated as the solution to Eq. (15), is a specific instance of an under-determined system of linear equations with an integer solution (the sign pattern in our case). The question of uniqueness of the solution to such systems, was recently studied by [22]. In their paper, [22] also proposed a linear programming relaxation of this non-convex problem, and proved it to work with overwhelming high probability if the equations are *random* and their number is more than half the number of unknowns. Unfortunately, this relaxation was unable to recover the correct sign pattern in our sign problem, probably due to the fact that our equations are far from being random.

Due to Lemma 1, which states that the sign pattern can have at most τ sign changes, our sign problem can also be viewed as a specific instance of an under-determined set of linear equations whose solution is sparse (for example by a change of variables from $\mathbf{s}$ to its discrete derivative). For this problem, L_1-type minimization schemes have been proposed and proven to recover the correct solution under various conditions. Unfortunately, a straightforward application of L_1 Lasso penalization was again unable to recover the correct sign pattern, even with noise-free measurements, unless the support parameter, τ which also controls the sparsity, was extremely small compared to N. This naturally raises a question on the role of oversampling in the sign problem: in the examples above we used $N = 5\tau$, which works in practice, but the exact dependence of the robustness on the oversampling is yet an open question.

Theoretical understanding of the robustness to noise of our algorithm is still lacking and is an interesting topic for further investigation. Improving the noise robustness is also an interesting route of future research. In particular, coupling the powerful relaxation schemes discussed above and their underlying theoretical guarantees to our segmentation-based relaxation could potentially improve the overall robustness of our sign retrieval algorithm.

Acknowledgments

We thank Dan Oron and Nirit Dudovich for useful discussions. We thank Gaurav Thakur for interesting discussions and for providing us with the code for his algorithm. We also thank Uri Feige for pointing us to Ref. [22]. The research of B.N. was supported in part by grant 892/14 from the Israel Science Foundation. O.R. acknowledges financial support from the James S. McDonnell Foundation.

Appendix A. Segmentation in the presence of noise

Here we describe a simple modification to Eq. (23) to account for noise. As described in Section 4.3, our noise model is given by Eq. (24). After neglecting dark counts and shot noise, we are left with

$$|\tilde{\mathbf{F}}| = |\mathbf{F} + \tfrac{\sigma}{\sqrt{N}}\boldsymbol{\eta}^s|. \tag{40}$$

Hence, assuming a high SNR we have,

$$|\tilde{\mathbf{F}}| = |\mathbf{F}| + \frac{\sigma}{\sqrt{N}}\mathcal{R}(e^{-\mathbf{i}\boldsymbol{\phi}}\boldsymbol{\eta}^s) + O\left(\frac{\sigma^2}{N}\right). \tag{41}$$

Since we assumed $\boldsymbol{\eta}^s \sim \mathcal{N}(0,1)$, we may thus write $\boldsymbol{\eta}^s = \frac{\mathbf{a}+\mathbf{ib}}{\sqrt{2}}$ where $\mathbf{a}$ and $\mathbf{b}$ are i.i.d. $\mathcal{N}(0,1)$. Hence, $\mathcal{R}(e^{-\mathbf{i}\boldsymbol{\phi}}\boldsymbol{\eta}^s) = \frac{1}{\sqrt{2}}(\mathbf{a}\cos\boldsymbol{\phi} + \mathbf{b}\sin\boldsymbol{\phi})$ is zero-mean Gaussian and its variance is $\frac{1}{2}$. In the presence of noise, $|\tilde{F}_j| + |\tilde{F}_{j+1}|$ is thus approximately the correct value $|F_j| + |F_{j+1}|$, perturbed by a Gaussian with variance σ^2/N. To account for noise, we add this standard deviation $\frac{\sigma}{\sqrt{N}}$ to the threshold on the right hand side of Eq. (23). Thus Eq. (25) follows.

References

[1] B. Alexeev, A.S. Afonso, M. Fickus, D.G. Mixon, Phase retrieval with polarization, SIAM J. Imaging Sci. 7 (1) (2014) 35–66.

[2] R. Balan, Reconstruction of signals from magnitudes of redundant representations: the complex case, Found. Comput. Math. 16 (2016) 677–721.

[3] R. Balan, B.G. Bodmann, P.G. Casazza, D. Edidin, Painless reconstruction from magnitudes of frame coefficients, J. Fourier Anal. Appl. 15 (2009) 488–501.

[4] R. Balan, P. Casazza, D. Edidin, On signal reconstruction without phase, Appl. Comput. Harmon. Anal. 20 (3) (2006) 345–356.

[5] H.H. Bauschke, P.L. Combettes, D.R. Luke, Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization, J. Opt. Soc. Amer. A 19 (2002) 1334–1345.

[6] R. Beinert, G. Plonka, Ambiguities in one-dimensional discrete phase retrieval from Fourier magnitudes, J. Fourier Anal. Appl. 21 (2015) 1169–1198.

[7] R. Beinert, G. Plonka, Enforcing uniqueness in one-dimensional phase retrieval by additional signal information in time domain, arXiv preprint, arXiv:1604.04493, 2016.

[8] B.G. Bodmann, Nathaniel Hammen, Stable phase retrieval with low-redundancy frames, Adv. Comput. Math. 41 (2) (2015) 317–331.

[9] Y.M. Bruck, L.G. Sodin, On the ambiguity of the image reconstruction problem, Opt. Commun. 30 (3) (1979) 304–308.

[10] R.E. Burge, M.A. Fiddy, A.H. Greenaway, G. Ross, The phase problem, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 350 (1661) (1976) 191–212.

[11] E.J. Candes, Y. Eldar, T. Strohmer, V. Voroninski, Phase retrieval via matrix completion, SIAM Rev. 57 (2015) 225–251.

[12] E.J. Candes, X. Li, M. Soltanolkotabi, Phase retrieval from coded diffraction patterns, Appl. Comput. Harmon. Anal. 39 (2) (2015) 277–299.

[13] H.N. Chapman, A. Barty, M.J. Bogan, S. Boutet, M. Frank, S.P. Hau-Riege, S. Marchesini, B.W. Woods, S. Bajt, W.H. Benner, et al., Femtosecond diffractive imaging with a soft-X-ray free-electron laser, Nat. Phys. 2 (12) (2006) 839–843.

[14] T.R. Crimmins, J.R. Fienup, Uniqueness of phase retrieval for functions with sufficiently disconnected support, J. Opt. Soc. Amer. 73 (2) (1983) 218–221.

[15] V. Elser, Phase retrieval by iterated projections, J. Opt. Soc. Amer. A 20 (1) (2003) 40–55.

[16] P. Ferreira, A. Kempf, Superoscillations: faster than the Nyquist rate, IEEE Trans. Signal Process. 54 (10) (2006) 3732–3740.

[17] J.R. Fienup, Phase retrieval algorithms: a comparison, Appl. Optim. 21 (15) (1982) 2758–2769.

[18] J.R. Fienup, J. Dainty, Phase retrieval and image reconstruction for astronomy, Image Recovery, Theory Appl. (1987) 231–275.

[19] D.W. Kammler, A First Course in Fourier Analysis, Cambridge University Press, 2007.

[20] M.V. Klibanov, P.E. Sacks, A.V. Tikhonravov, The phase retrieval problem, Inverse Probl. 11 (1995) 1–28.

[21] B. Leshem, R. Xu, Y. Dallal, J. Miao, B. Nadler, D. Oron, N. Dudovich, O. Raz, Direct single-shot phase retrieval from the diffraction pattern of separated objects, Nat. Commun. 7 (2016).

[22] O.L. Mangasarian, B. Recht, Probability of unique integer solution to a system of linear equations, European J. Oper. Res. 214 (1) (2011) 27–30.

[23] S. Marchesini, A unified evaluation of iterative projection algorithms for phase retrieval, Rev. Sci. Instrum. 78 (2007) 011301.

[24] J. Miao, D. Sayre, H.N. Chapman, Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects, J. Opt. Soc. Amer. A 15 (6) (1998) 1662–1669.

[25] R.P. Millane, Phase retrieval in crystallography and optics, J. Opt. Soc. Amer. A 7 (1990) 394–411.

[26] A. Novikov, M. Moscoso, G. Papanicolaou, Illumination strategies for intensity-only imaging, SIAM J. Imaging Sci. 8 (3) (2015) 1547–1573.

[27] O. Raz, B. Leshem, J. Miao, B. Nadler, D. Oron, N. Dudovich, Direct phase retrieval in double blind Fourier holography, Opt. Express 22 (21) (Oct 2014) 24935–24950.

[28] O. Raz, B. Nadler, N. Dudovich, Vectorial phase retrieval for 1-D signals, IEEE Trans. Signal Process. 61 (2013) 1632–1643.

[29] O. Raz, O. Schwartz, D. Austin, A.S. Wyatt, A. Schiavi, O. Smirnova, B. Nadler, I.A. Walmsley, D. Oron, N. Dudovich, Vectorial phase retrieval for linear characterization of attosecond pulses, Phys. Rev. Lett. 107 (13) (2011) 133902.

[30] J.L.C. Sanz, Mathematical considerations for the problem of Fourier transform phase retrieval from magnitude, SIAM J. Appl. Math. 45 (1985) 651–664.

[31] P. Sidorenko, O. Kfir, Y. Shechtman, A. Fleischer, Y.C. Eldar, M. Segev, O. Cohen, Sparsity-based super-resolved coherent diffraction imaging of one-dimensional objects, Nat. Commun. 6 (2015).

[32] A. Szameit, Y. Shechtman, E. Osherovich, E. Bullkich, P. Sidorenko, H. Dana, S. Steiner, E.B. Kley, S. Gazit, T. Cohen-Hyams, S. Shoham, M. Zibulevsky, I. Yavneh, Y.C. Eldar, O. Cohen, M. Segev, Sparsity-based single-shot subwavelength coherent diffractive imaging, Nat. Mater. 11 (2012) 455–459.

[33] G. Thakur, Reconstruction of bandlimited functions from unsigned samples, J. Fourier Anal. Appl. 17 (4) (2011) 720–732.
[34] L.N. Trefethen, D. Bau III, Numerical Linear Algebra, vol. 50, Siam, 1997.
[35] I. Waldspurger, A. d'Aspremont, S. Mallat, Phase recovery, maxcut and complex semidefinite programming, Math. Program. 149 (1–2) (2015) 47–81.
[36] I.A. Walmsley, C. Dorrer, Characterization of ultrashort electromagnetic pulses, Adv. Opt. Photon. 1 (2009) 308–437.