

Blind Phaseless Short-Time Fourier Transform Recovery

Tamir Bendory¹, Dan Edidin, and Yonina C. Eldar²

Abstract—The problem of recovering a pair of signals from their blind phaseless short-time Fourier transform measurements arises in several important phase retrieval applications, including ptychography and ultra-short pulse characterization. In this paper, we prove that in order to determine a pair of generic signals uniquely, up to trivial ambiguities, the number of phaseless measurements one needs to collect is, at most, five times the number of parameters required to describe the signals. This result improves significantly upon previous papers, which required the number of measurements to be quadratic in the number of parameters rather than linear. In addition, we consider the simpler problem of recovering a pair of generic signals from their blind short-time Fourier transform, when the phases are known. In this setting, which can be understood as a special case of the blind deconvolution problem, we show that the number of measurements required to determine the two signals, up to trivial ambiguities, equals exactly the number of parameters to be recovered. As a side result, we study the classical phase retrieval problem—that is, recovering a signal from its Fourier magnitudes—when some entries of the signal are known a priori. We derive a bound on the number of required measurements as a function of the size of the set of known entries. Specifically, we show that if most of the signal’s entries are known, then only a few Fourier magnitudes are necessary to determine a signal uniquely.

Index Terms—Phase retrieval, blind deconvolution, ptychography, short-time Fourier transform, ultra-short pulse characterization, FROG.

I. INTRODUCTION

PHASE retrieval is the problem of recovering a signal from its Fourier magnitudes. This problem arises in a variety of applications and scientific fields, such as X-ray crystallography [1], [2], optical imaging [3], [4], ultra-short pulse characterization [5], astronomy [6] and signal processing [7]–[9]. For recent surveys from a signal processing perspective; see [4], [10], [11].

Manuscript received August 22, 2018; revised April 9, 2019; accepted October 2, 2019. Date of publication October 11, 2019; date of current version April 21, 2020. D. Edidin was supported by Simons Collaboration under Grant 315460. Y. C. Eldar was supported in part by the European Union’s Horizon 2020 Research and Innovation Program under Grant 646804-ERCCOG-5BNYQ, and in part by the Israel Science Foundation under Grant 335/14.

T. Bendory is with the School of Electrical Engineering, Tel Aviv University, Tel Aviv 6997801, Israel (e-mail: bendory@tauex.tau.ac.il).

D. Edidin is with the Department of Mathematics, University of Missouri, Columbia, MO 65211 USA (e-mail: edidind@missouri.edu).

Y. C. Eldar is with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

Communicated by P. Grohs, Associate Editor for Signal Processing.
Digital Object Identifier 10.1109/TIT.2019.2947056

Evidently, there are infinitely many signals with the same Fourier magnitudes. Therefore, to make the problem well-posed—that is, having a unique mapping between the Fourier magnitudes and the signal—some additional information on the signal must be harnessed. In many applications it is common to assume prior knowledge on the structure of the underlying signal. For instance, in crystallography the signal is sparse [1], [2], [12]. If the signal is known to be of minimum phase, then there is a unique mapping between the Fourier magnitudes and the signal [13]. Other useful assumptions are that the signal has nonzero entries only in a known region or that its entries are nonnegative.

We focus in this paper on an alternative strategy to enforce uniqueness by collecting additional phaseless measurements. One important example, which serves as the the main motivation for this paper, is *ptychography* [14]–[17]. In ptychography, the specimen (i.e., signal, image or volume) is scanned by a localized illumination beam and Fourier magnitudes of overlapping windowed measurements are recorded. Another popular technique that collects multiple measurements is Frequency-Resolved Optical Gating (FROG), which is used to characterize ultra-short laser pulses [5]. In FROG, the Fourier magnitudes of the product of the signal with a shifted version of itself are recorded, for several shifts. An extension, called *blind FROG*, can be used to characterize two signals simultaneously by measuring the Fourier magnitudes of the product of one signal with a shifted version of the other.

In ptychography and blind FROG, the collected data can be modeled as the phaseless blind short-time Fourier transform (STFT) with respect to two signals. A detailed mathematical model is provided in Section II. In ptychography, the two signals are the specimen and the illumination beam. In blind FROG, the signals are the two optical pulses to be characterized and the blind phaseless STFT measurements are usually called the blind FROG trace. In those applications, a variety of algorithms have been suggested to estimate both signals simultaneously [5], [18]–[21]. However, these methods lack theoretical analysis. In this paper, we study the question of uniqueness, characterize the trivial ambiguities of the problem and derive bounds on the number of measurements required to determine the two signals uniquely, up to these ambiguities.

We begin by presenting the trivial ambiguities of the phaseless blind STFT problem in Proposition III.4. The main result of this paper is Theorem III.6. It shows that the number of measurements required to uniquely determine a generic pair of signals, up to trivial ambiguities, is (at most) five times the number of parameters that describe the signals. This result

significantly improves upon previous results [22] that required the number of measurements to be quadratic in the number of parameters rather than linear.

We also consider a simpler problem of recovering two signals from their blind STFT, when the phases are assumed to be known. This problem can be understood as a special case of the *blind deconvolution* problem [23]–[27]; see the discussion in Section II. We show that the number of measurements required to determine the two signals uniquely, up to scaling ambiguities, is optimal. That is, the number of measurements equal exactly the number of parameters to be recovered; see Theorem III.3. The dimension of the ambiguity group is inversely proportional to the overlap between adjacent sections. Hence, a small overlap results in a large scaling ambiguities group. The technical details are presented and discussed in Section III, while proofs are provided in Section V and Appendix B. Some of the proofs require basic definitions in group theory, which are summarized in Appendix A.

As a side result of this work, we study in Section IV the classical phase retrieval problem—that is, recovering a signal from its Fourier magnitudes—when some entries of the signal are known. In [28], Beinert and Plonka showed that one entry of the signal together with its Fourier magnitudes determine almost all signals uniquely. In Proposition IV.2, we extend this result by bounding the number of required phaseless Fourier measurements when several entries are known. In particular, we show that if the number of unknown entries is relatively small, then one needs to collect only few Fourier intensity measurements. We successively use this result to determine the pair of signals in a specified section of the phaseless blind STFT measurements based on the Fourier magnitudes of this section and some of the signals' entries. These known entries are shared with an adjacent overlapping section whose entries were determined previously.

II. PROBLEM FORMULATION

The phaseless blind STFT of a signal $x \in \mathbb{C}^N$ with respect to a window $w \in \mathbb{C}^W$, for some $N \geq W$, is given by

$$\begin{aligned} |\hat{y}[k, m]| &= \left| \sum_{n=0}^{N-1} x[n]w[mL - n]e^{-2\pi i kn/N} \right| \\ &= |x[0]w[mL] + \eta_k x[1]w[mL - 1] + \dots \\ &\quad + \eta_k^{N-1} x[N-1]w[mL - (N-1)]|, \end{aligned} \quad (\text{II.1})$$

where $\eta_k := e^{-2\pi i k/N}$. Here, $0 < L < W$ is the step size. We assume that the signals are zero outside their support so that $w[n] = 0$ for all $n \notin [0, \dots, W-1]$ and $x[n] = 0$ for $n \notin [0, \dots, N-1]$. The ratio between W and L determines the number of short-time sections which is given by $M := \lceil (N+W-1)/L \rceil$. The goal is derive the number of frequencies required per window in order to determine x and w from $|\hat{y}[k, m]|$. Note that if $W = L$ then there is no overlap between adjacent sections, and the problem reduces to the standard phase retrieval problem.

The problem of recovering a signal from its phaseless STFT—when the window w is assumed to be known—was studied thoroughly in recent years [17], [29]–[33]. However,

in ptychography, which is the prime motivation of these papers, the precise structure of the window is usually unknown a priori and thus standard algorithms in the field optimize over the signal and the window simultaneously [18]–[20].

The main result of this paper, presented in Theorem III.6, shows that it suffices to consider less than $10L$ Fourier frequencies per window to determine a pair of generic signals (x, w) . Therefore, in total, we require $10LM \approx 10(N+W)$ phaseless measurements. This result is near optimal in the sense that the number of parameters to be recovered is $2(N+W)$: the real and imaginary parts of the signal and the window. We mention that our result does not hold for the special case when $x = w$ as the problem appears in the FROG setup [5]. The latter case was investigated in [34], where it was shown that the number of measurements required to determine the (single) signal is three times its bandwidth. In Section III, we provide a more comprehensive comparison with related results in the literature.

We also explore the simpler case of blind STFT when the phases are assumed to be known. In this problem, the goal is to determine $x \in \mathbb{C}^N$ and $w \in \mathbb{C}^W$ from

$$\hat{y}[k, m] = \sum_{n=0}^{N-1} x[n]w[mL - n]e^{-2\pi i kn/N}. \quad (\text{II.2})$$

For generic (x, w) , we show that if $W > L$ then it suffices to consider only $2L$ Fourier measurements per window to determine x and w , up to unavoidable L scaling ambiguities presented in Proposition III.2. Therefore, $2ML \approx 2(N+W)$ measurements are enough. However, in Theorem III.6 we prove by a more careful examination that the number of measurements can be reduced to only $N+W-L$ Fourier measurements if we choose the measurements properly for each window. This result is optimal as it captures precisely the number of parameters to be determined.

The model of blind STFT (II.2) can be understood as a special case of the *blind deconvolution* problem. In particular, let us denote by \hat{y}_k all entries of \hat{y} for fixed k . Then, we can write (II.2) as

$$\hat{y}_k = x_k * w,$$

where $x_k[n] := x[n]e^{-2\pi i kn/N}$ and $*$ denotes convolution. Blind deconvolution is a fundamental problem in a variety of applications, including astronomy, communication, image deblurring, system identification, optics and structural biology; see [23]–[27], [35], [36] to name a few. Following advances in related fields like compressed sensing and phase retrieval, many papers have focused on establishing theoretical foundations for different settings of the blind deconvolution problem; see for instance [37]–[41]. In particular, [39] provides a thorough analysis of bilinear problems in general and, in particular, blind deconvolution. These works assume a low-dimensional structure of the signals to enforce uniqueness. Inspired by some phase retrieval and channel estimation applications, it was shown in [42], [43] that a pair of signals with known autocorrelations can be recovered from their blind deconvolution by a convex program. Our model deviates from the papers mentioned above as it relies on overlapping windows

to derive uniqueness for generic signals with optimal number of measurements.

III. MAIN RESULTS

We begin by studying the symmetries of the blind STFT map (II.2), often called *trivial ambiguities*. Next, we derive the number of measurements required to determine a pair of *generic* signals (x, w) . By generic signals, we mean the following:

Definition III.1. When saying that a generic signal is uniquely determined by a collection of polynomial measurements we mean that the set of signals which cannot be determined by these measurements lies in the vanishing locus of a nonzero polynomial¹. In particular, this means that we can recover almost all signals with the given measurements.

We remark that while almost all signals are generic, Definition III.1 does not cover several classes of signals which might be important in some applications. A notable example is the class of sparse signals (signals with only few nonzero entries). For instance, the following pair of signals (x_1, w_1) and (x_2, w_2) share the same blind phaseless STFT (II.1) (this example is taken from [44]):

$$\begin{aligned} x_1 &= [1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1], & w_1 &= [1, 0, 0, 0, 1, 0, 0], \\ x_2 &= [1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0], & w_2 &= [1, 0, 1, 0, 1, 0, 1]. \end{aligned}$$

Further ambiguities for sparse signals when the window is known are discussed in [30].

Since the window $w \in \mathbb{C}^W$ is unknown, the group of ambiguities of the map (II.2) is large, although the phases of the blind STFT are known. This group of ambiguities can be thought of as *scaling ambiguities* between the signal and the window as formulated in the following proposition.

Proposition III.2. Let $\lambda := (\lambda_0, \dots, \lambda_{L-1})$ be complex, nonzero, numbers. Then, the following action preserves the blind STFT measurements (II.2):

$$\begin{aligned} \lambda \circ (x, w) & \\ & := ((\lambda_0 x[0], \dots, \lambda_{L-1} x[L-1], \lambda_0 x[L], \dots), \\ & (\lambda_0^{-1} w[0], \lambda_{L-1}^{-1} w[1], \dots, \lambda_1^{-1} w[L-1], \lambda_0^{-1} w[L], \dots)). \end{aligned}$$

Proof. Note that λ acts on x by multiplication by $\lambda_n \bmod L$ and on w by multiplication by $\lambda_{(L-n) \bmod L}^{-1}$. Thus, every term in the sum for $\hat{y}[k, m]$ is preserved under the action of λ since $x[n]w[mL-n]$ is transformed into $(\lambda_n \bmod L)x[n](\lambda_n^{-1} \bmod L)w[mL-n]$. \square

Note that Proposition III.2 can be understood as an action of a group $(\mathbb{C}^\times)^L$ on $\mathbb{C}^N \times \mathbb{C}^W$.

Next, we are ready to present the uniqueness result for the blind STFT case. We show that for each section it suffices to acquire $2L$ Fourier measurements to determine

the signals. Thus, $ML \approx 2(N+W)$ measurements are enough to determine the signals up to the L scaling ambiguities of Proposition III.2. This bound can be improved by choosing the number of Fourier measurements carefully per section. Particularly, we show that only $N+W-L$ measurements suffice. This result is optimal since the number of parameters to be determined is $N+W$ (the length of the signals), while L entries can be fixed arbitrarily by Proposition III.2. Interestingly, the result holds for any $L < W$. Therefore, considering small overlaps between adjacent windows (i.e., large L) increases only the size of the ambiguity group. The proof is given in Section V-A and is based on a recursive argument.

Theorem III.3. For a generic signal $x \in \mathbb{C}^N$ and a generic window $w \in \mathbb{C}^W$ with $W > L$, the pair (x, w) is uniquely determined, modulo the trivial ambiguities of Proposition III.2, from $N+W-L$ measurements from (II.2).

While the model (II.2) has not been studied in the literature, it is instructive to compare Theorem III.3 with [39]. In this paper, Kech and Kraemer considered the uniqueness of bilinear maps and, as an application, blind (circular) convolution maps $\mathbb{C}^N \times \mathbb{C}^N \rightarrow \mathbb{C}^N$. In particular, they assume the two signals lie in generic known subspaces of dimension d_1 and d_2 . In this case, *all signals* are uniquely determined from their convolution provided that $N \geq 2(d_1 + d_2) - 4$, modulo a one-dimensional scaling ambiguity between the two signals. A similar result holds true when the signals are sparse (see also [45], [46]). Therefore, the number of measurements is approximately twice the number of parameters to be estimated. When considering generic sparse signals, it suffices to demand $N \geq s_1 + s_2$, where s_1 and s_2 denote the cardinality of the signals. Comparably, Theorem III.3 states that the number of measurements required to determine a pair of generic signals is exactly the number of parameters to be recovered $N+W-L$. We did not derive a result that holds for all signals. In addition, the dimension of the ambiguity group presented in Proposition III.2 grows with L , whereas the dimension of the ambiguity group in [39] is always one.

We now turn our attention to the problem of determining a pair of signals from their phaseless blind STFT measurements (II.1), which is the focal point of this paper. As expected, in this case the group of trivial ambiguities increases, but only by two real dimensions. These two additional symmetries correspond to multiplication by global phase and continuous modulation. We note that other symmetries that frequently appear in phase retrieval setups, such as conjugate reflection or discrete shifts, do not occur in our setting because of the aperiodicity of the setup; compare with [22, Proposition 1]. In what follows, S^1 denotes the unit circle:

$$S^1 := \{\eta \in \mathbb{C} : |\eta| = 1\}.$$

Proposition III.4. Let $\lambda := (\lambda_0, \dots, \lambda_{L-1})$ be complex, nonzero, numbers and let $\eta_1, \eta_2 \in S^1$. Then, the following action preserves the phaseless blind STFT

¹Given a nonzero polynomial f in T variables (real or complex), the vanishing locus of f is the set $V(f) = \{(x_1, \dots, x_T) \in \mathbb{R}^T \text{ (resp. } \mathbb{C}^T) \mid f(x_1, \dots, x_T) = 0\}$. By a general result in algebraic geometry $\dim V(f) < T$. As a consequence, the complement of $V(f)$ in \mathbb{R}^T (resp. \mathbb{C}^T) is dense.

measurements (II.1):

$$\begin{aligned} & (\eta_1, \eta_2, \lambda) \circ (x, w) \\ &= (\eta_1 \lambda_0 x[0], \eta_1 \eta_2 \lambda_1 x[1], \dots, \\ & \quad \eta_1 \eta_2^{L-1} \lambda_{L-1} x[L-1], \eta_1 \eta_2^L \lambda_0 x[L], \dots, \\ & \quad \lambda_0^{-1} w[0], \eta_2 \lambda_{L-1}^{-1} w[1], \dots, \\ & \quad \eta_2^{L-1} \lambda_1^{-1} w[L-1], \eta_2^L \lambda_0^{-1} w[L], \dots). \end{aligned}$$

If $N = W$, then we have an additional ambiguity because (x, w) and (\bar{w}, \bar{x}) have the same phaseless blind STFT measurements.

Proof. It is straightforward to verify that under this action, the blind STFT measurement $\hat{y}[k, m]$ is translated to $\eta_1 \eta_2^{mL} \hat{y}[k, m]$ which has the same absolute value. \square

As for the blind STFT ambiguities, Proposition III.4 can be understood as an action of the group $S^1 \times S^1 \times (\mathbb{C}^\times)^L$ on $\mathbb{C}^N \times \mathbb{C}^W$. The following proposition uncovers an interesting property of this action. Specifically, it shows that the unique element of a quotient of this group that maps a generic vector to itself is the identity. The proof is provided in Section V-B.

Proposition III.5. *Define an action of the group $S^1 \times S^1 \times (\mathbb{C}^\times)^L$ on $\mathbb{C}^N \times \mathbb{C}^W$ as in Proposition III.4. Then, a quotient of this group by the finite group μ_L of L -th roots of unity acts generically freely².*

We are now ready to present the main result of this paper. As in Theorem III.3, the result holds true for any $W > L$. In particular, for any window it is sufficient to acquire less than $10L$ Fourier intensity measurements. Therefore, less than $10LM \approx 10(N + W)$ are required in total. Increasing L increases the size of the ambiguity group but has only a negligible effect on the number of required measurements. The proof is provided in Section V-C.

Theorem III.6. *For a generic signal $x \in \mathbb{C}^N$ and a generic window $w \in \mathbb{C}^W$, the pair (x, w) is uniquely determined, modulo the trivial ambiguities of Proposition III.4, from fewer than $10(N + W)$ phaseless blind STFT measurements (II.1). Precisely, (x, w) is uniquely determined by the measurements*

$$\{|\hat{y}[0, m]|, \dots, |\hat{y}[Q(m), m]|\}_{m=0, \dots, M-1},$$

where $Q(m) := \min(2mL + 1, 10L - 3)$ and $M := \lceil (N + W - 1)/L \rceil$.

In [22], it was shown that a pair of a signal and a window (which, in contrast to our model, are allowed to be equal to each other) can be uniquely determined when $L = 1$ and all N Fourier intensity measurements are recorded for each window, resulting in order of N^2 phaseless blind STFT measurements. Namely, the number of measurements is quadratic in the number of parameters to be determined. Theorem III.6 improves this result significantly as, for any L , the number of required measurements is linear in N . In [34], we have shown that in the special $x = w$ case, appearing in the FROG technique [5], one needs to acquire only $3B$ measurements, where B is the bandwidth of the signal.

²See Appendix A and in particular Example A.5.

The two-dimensional blind ptychography problem was analyzed in [47]. It is important to note, however, that in contrast to the one-dimensional case, the phase retrieval problem admits a unique solution for generic signals in two dimensions³, up to ambiguities [10], [49].

Finally, we would like to refer to a recent paper [50], considering the recovery of a pair of signals from the Fourier magnitudes of their (circular) blind deconvolution. The underlying assumption is that the two signals lie in low-dimensional random subspaces of dimensions k and m . Thus, it studies a complementary problem to our model which is based on overlapping windows (II.1). The main result of this paper states that the two signals can be recovered by a convex program provided that $N/\log^2 N \gg (k + m)$.

IV. PHASE RETRIEVAL FROM LIMITED MEASUREMENTS FOR PARTIALLY KNOWN SIGNALS

In this section, we study a general question in phase retrieval about recovering a signal from its Fourier magnitudes, where a subset of the signal's entries is already known. This situation occurs in the phaseless blind STFT problem since, if $W > L$, the sections in (II.1) overlap. Thus, the m th section is recovered from its Fourier magnitudes and the knowledge of some of its entries. These known entries are determined by the $(m - 1)$ th section. We use this procedure successively in the proof of Theorem III.6 in Section V-C. However, since the main result of this section, Proposition IV.2, is quite general, we devote this separate section to present it and discuss its ramifications.

Consider the following setup. Let $x \in \mathbb{C}^N$ be a signal and, as usual, we assume $x[i] = 0$ for $i \notin \{0, \dots, N - 1\}$. In this section, we examine the continuous Fourier measurements $S^1 \rightarrow \mathbb{R}_{\geq 0}$ given by

$$A(\eta) = \left| \sum_{n=0}^{N-1} x[n] \eta^n \right|^2, \quad (\text{IV.1})$$

where η in on the unit circle S^1 . Let $S \subset [0, N - 1]$ be a proper subset (i.e., $|S| < N$) and assume that $x[n]$ is known for the complementary set $n \in S^c$.

The question of determining a signal from its Fourier magnitudes, when partial information on the signal's entries are known, has been studied in [28], [51], [52], where the former generalizes the results of the two latter papers. In particular, it was shown in [28] that almost any signal is uniquely determined from its Fourier magnitudes and one entry. Since we use this result repeatedly, we cite it as follows:

Lemma IV.1. [28] *Let ℓ be an arbitrary integer between 0 and $N-1$. Then, almost every $x \in \mathbb{C}^N$ can be uniquely recovered from $A(\eta)$ and $x[\ell]$. If $\ell = (N - 1)/2$, then the reconstruction is up to conjugate reflection.*

In this section, we generalize this result and derive a bound on the number of Fourier intensity measurements required if several entries of the signal are known.

³Interestingly, it was recently shown that this solution might be extremely sensitive to errors [48].

For a generic x , the trigonometric polynomial $A(\eta)$ is of degree $N - 1$ and thus can be recovered from its value at $2N - 1$ distinct frequencies. If in addition one entry of the signal is known, then Lemma IV.1 implies that a generic signal is determined uniquely. The following proposition, which is the main result of this section, provides a bound on the number of intensity measurements required for uniqueness as a function of the number of known entries. In particular, if the size of the set of unknown entries $|S|$ is small relative to N , then x can be recovered from only a few measurements. In what follows, we denote a difference set by

$$S - S = \{m - n \mid m, n \in S\}.$$

The following result is proved in Section V-D.

Proposition IV.2. *If $k \geq 2|S - S| - 1 + 2|S|$ and $|S| < N$, then $A(\eta)$ is uniquely determined by $A(\eta_1), \dots, A(\eta_k)$ for distinct frequencies η_1, \dots, η_k .*

Combining Lemma IV.1 and Proposition IV.2 we conclude:

Corollary IV.3. *If $k \geq 2|S - S| - 1 + 2|S|$ and $|S| < N$ (that is, at least one signal entry is known), then almost every $x \in \mathbb{C}^N$ is determined uniquely by $A(\eta_1), \dots, A(\eta_k)$ for distinct frequencies η_1, \dots, η_k .*

The main application for Proposition IV.2 and Corollary IV.3 in this paper is the proof of Theorem III.6 in Section V-C. In particular, the following situation occurs in phaseless blind STFT. The set of unknown entries is $S = [0, L - 1] \cup [L(m - 1), mL - 1]$ and thus $|S| = 2L$. In this case, the difference set is given by

$$S - S = [0, L - 1] \cup [L(m - 2) + 1, mL - 1],$$

and hence $|S - S| = 3L - 1$. Proposition IV.2 affirms that in this case the signal can be determined by the values of at most $k = 2(3L - 1) - 1 + 2(2L) = 10L - 3$ frequencies.

V. PROOFS

A. Proof of Theorem III.3

Throughout the proof we assume that $w[0], \dots, w[L - 1]$ are all nonzero as w is generic. Because of the scaling ambiguities of Proposition III.2, we can rescale and assume that $w[0] = w[1] \dots w[L - 1] = 1$ and thereby eliminate the ambiguities.

Let x', w' be a solution to the system of bilinear equations (II.2). We will use recursion to show that for generic (x, w) , there is a unique solution under the constraint $w'[0] = w'[1] = \dots = w'[L - 1] = 1$, and therefore, $x = x'$ and $w = w'$.

Step 0: From (II.2) we have $\hat{y}[0, 0] = x'[0]w'[0]$ (for any k). Since $w'[0] = 1$, we can determine $x'[0]$ uniquely.

Step 1: For $m = 1$ and fixed k , we get

$$\hat{y}[k, 1] = x'[0]w'[L] + \eta_k x'[1]w'[L - 1] + \dots + \eta_k^L x'[L]w'[0]. \quad (\text{V.1})$$

Since $x'[0]$ is determined in Step 0 and $w'[0], \dots, w'[L - 1]$ are normalized to 1, we get a linear system of equations with the $L + 1$ unknowns $x'[1], \dots, x'[L], w'[L]$. Hence, if we take

$M_1 = L + 1$ Fourier measurements, we can uniquely determine the unknowns for generic $x[0] = x'[0]$.

Step 2: For $m = 2$ and fixed k , we get

$$\begin{aligned} \hat{y}[k, 2] &= x[0]w'[2L] + \eta_k x[1]w'[2L - 1] \\ &+ \dots + \eta_k^L x'[L]w'[L] + \eta_k^{L+1} x'[L+1]w'[L-1] + \dots \\ &+ \eta_k^{2L} x'[2L]w'[0]. \end{aligned} \quad (\text{V.2})$$

By the previous steps, $w'[0], \dots, w'[L]$ and $x'[0], \dots, x'[L]$ are determined, so this is a linear equation in the $2L$ unknowns $w'[2L], \dots, w'[L + 1], x'[L + 1], \dots, x'[2L]$. Hence we can solve the system with $M_2 = 2L$ Fourier measurements. In total, up to this stage we require $3L + 2$ Fourier measurements.

Step m: In the m th step, we observe

$$\begin{aligned} \hat{y}[k, m] &= x'[0]w'[mL] + \eta_k x'[1]w'[mL - 1] \\ &+ \dots + \eta_k^{L-1} x'[L - 1]w'[mL - L + 1] \\ &+ \eta_k^L x'[L]w'[mL - L] \\ &+ \dots + \eta_k^{2L-1} x'[2L - 1]w'[mL - 2L + 1] \\ &+ \dots \\ &+ \eta_k^{mL-L+1} x'[mL - L + 1]w'[L - 1] \\ &+ \dots + \eta_k^{mL} x'[mL]w'[0]. \end{aligned}$$

By induction, there are at most $2L$ unknowns in this equation, namely, $x'[L(m - 1) + 1], \dots, x'[Lm]$ and $w'[Lm], \dots, w'[Lm - L + 1]$. Hence $M_m = 2L$ Fourier measurements suffice.

The crude estimate above leads to order of $2ML \approx 2(N + W)$ Fourier measurements. This estimate can be significantly refined by taking into account that we know the length of the signals N and W . Starting from the $m = 3$ step, we need one Fourier measurement for each unknown. Therefore, to determine the last $N - (2L + 1)$ entries of the signal we need $N - (2L + 1)$ measurements in total, and similarly for the last $W - (2L + 1)$ entries of the window (if $W \geq 2L + 1$). Together with the $3L + 2$ measurements required in the first three steps, we conclude that merely $N + W - L$ measurements suffice. The same holds true if $2L \geq W > L$.

B. Proof of Proposition III.5

Since x and w are generic, we may assume that $x[i]$ and $w[L - i]$ are nonzero for $0 \leq i \leq L$. Then, if $(\eta_1, \eta_2, \lambda_0, \lambda_{L-1})$ acts trivially on (x, w) , then we see that for all $0 \leq i \leq L$,

$$\eta_1 \eta_2^i \lambda_{(i \bmod L)} = 1 \quad (\text{coefficient of } x[i])$$

and

$$\eta_2^{mL-i} \lambda_{(i \bmod L)}^{-1} = 1 \quad (\text{coefficient of } w[mL - i]).$$

Looking at the coefficients of $x[0]$ and $x[L]$ we see that both $\eta_1 \lambda_0 = 1$ and $\eta_1 \eta_2^L \lambda_0 = 1$ which forces $\eta_2^L = 1$. Looking at the coefficient of $w[L]$, we see that $\eta_2^{mL} \lambda_0^{-1} = 1$ so $\lambda_0 = \eta_2^L = 1$ and also forces $\eta_1 = 1$. Looking at the coefficient of $x[i]$ or $w[i]$ for $i > 0$ we also see that $\lambda_i = \eta_2^{mL-i}$. Therefore, the elements of $S^1 \times S^1 \times (\mathbb{C}^\times)^L$ that fix the generic element of $\mathbb{C}^N \times \mathbb{C}^W$ are of the form

$((1, \omega), (1, \omega, \dots, \omega^{L-1}))$ where $\omega = e^{2\pi ik/L}$. These elements form a subgroup isomorphic to the finite group μ_L of L -th roots of unity. Hence, the group $(S^1 \times S^1 \times (\mathbb{C}^\times)^L) / \mu_L$ acts generically freely. Note that since S^1 / μ_L is isomorphic to S^1 and $\mathbb{C}^\times / \mu_L$ is isomorphic to \mathbb{C}^\times , the quotient group is also isomorphic to $S^1 \times S^1 \times (\mathbb{C}^\times)^L$.

C. Proof of Theorem III.6

Given a pair of signals (x', w') whose STFT intensity measurements are $\{|\hat{y}[k, m]|\}$ as in (II.1), let us define the vector $y_m \in \mathbb{C}^{mL+1}$:

$$y_m[n] := x'[n]w'[mL - n], \quad n = 0, \dots, mL.$$

Note that the length of the vectors changes for different values of m . From Theorem III.3, we know that modulo the $(\mathbb{C}^\times)^L$ group of scaling ambiguities (see Proposition III.2), a generic pair (x', w') can be recovered from y_m for $m = 0, \dots, M - 1$. Therefore, we will prove that the vectors y_m are unique, modulo ambiguities.

The proof takes inspiration from the recursive technique of [34]. In particular, for fixed m , let $A_m(\eta) = |\sum_{n=0}^{mL} y_m[n]\eta^n|^2$ be the Fourier intensity function of y_m , where $\eta \in S^1$. Then, the phaseless STFT measurements are the samples $|\hat{y}[k, m]|^2 = A_m(e^{2\pi ik/N})$. Since $A_m(\eta)$ is a trigonometric polynomial of degree mL , a priori, it can always be recovered from the $2mL + 1$ STFT intensity measurements $|\hat{y}[0, m]|, \dots, |\hat{y}[0, 2mL]|$. However, we will show that if y_0, \dots, y_{m-1} are previously known (modulo ambiguities), then $A_m(\eta)$ can be recovered from at most $10L - 3$ STFT intensity measurements and that the knowledge of y_{m-1} and $A_m(\eta)$ uniquely determines y_m .

The proof follows the recursion below:

Step 0: As in Section V-A, we may use the action of the ambiguity group to assume that $w'[0] = \dots = w'[L - 1] = 1$. Since we know $|\hat{y}[0, 0]| = |x'[0]w'[0]|$ and we assume that $w'[0] = 1$, we know $x'[0]$ up to phase. Using the global phase ambiguity, we can then fix the phase of $x'[0]$ arbitrarily.

Step 1: The trigonometric polynomial $A_1(\eta)$ has degree L , so it is uniquely determined by the $2L + 1$ STFT intensity measurements $|\hat{y}[0, 1]|, \dots, |\hat{y}[2L, 1]|$. However, the vector $y_1 = (x'[0]w'[L], \dots, x'[L]w'[0])$ is not uniquely determined by the Fourier intensity function $A_1(\eta)$. Indeed, there are up to 2^{L-1} vectors, modulo global phase and reflection and conjugation, with Fourier intensity function $A_1(\eta)$; see for instance [10, Section 3.1].

Step 2: Using the second S^1 ambiguity, we fix the phase of $w'[L]$ arbitrarily. For example, we may fix it to be real. In addition, note that for each of the possible solutions from Step 1, $|w'[L]x'[L]|$ has to be the same. The reason is that the last entry of the autocorrelation of y_1 (which is equivalent to $A_1(\eta)$) is given by $w[0]\overline{x[0]}w[L]\overline{x[L]}$ and $x[0]$ and $w[0]$ are already known from Step 0.

Now, $A_2(\eta)$ is a trigonometric polynomial of degree $2L$, hence it is uniquely determined by $4L + 1$ STFT intensity measurements $|\hat{y}[0, 2]|, \dots, |\hat{y}[4L, 2]|$. From Lemma IV.1, we know that the Fourier magnitudes of a signal and the knowledge of one of its entries suffice to determine

almost all signals. Since $|x'[L]w'[L]|$ is determined, for each $(x'[0], \dots, x'[L], w'[0], \dots, w'[L])$ obtained in Step 1, there are two possible vectors

$$(x'[0], \dots, x'[2L], w'[0], \dots, w'[2L]) \text{ and } (\overline{w'[0]}, \dots, \overline{w'[2L]}, \overline{x'[0]}, \dots, \overline{x'[2L]}),$$

with Fourier intensity function determined by the phaseless STFT measurements $|\hat{y}[0, 2]|, \dots, |\hat{y}[4L, 2]|$.

Step 3: After step 2, we have 2^L possible values for the unknowns

$$(x'[0], \dots, x'[2L], w'[0], \dots, w'[2L]),$$

that are consistent with the constraints

$$\left| x'[0]w'[L] + \eta x'[1]w'[L-1] + \dots + \eta^L x'[L]w'[0] \right|^2 = A_1(\eta), \quad (\text{V.3})$$

and

$$\left| x'[0]w'[2L] + \eta x'[1]w'[2L-1] + \dots + \eta^{2L} x'[2L]w'[0] \right|^2 = A_2(\eta). \quad (\text{V.4})$$

In this stage, $A_3(\eta)$ is a trigonometric polynomial of degree $3L$ so it can be determined by the $6L + 1$ STFT intensity measurements $|\hat{y}[0, 3]|, \dots, |\hat{y}[6L, 3]|$. Moreover, for each choice of $(x'[0], x'[1], \dots, x'[2L], w'[0], \dots, w'[2L])$, the values of $w'[2L+1], \dots, w'[3L], x'[2L+1], \dots, x[3L]$ are uniquely determined from $A_3(\eta)$; see Lemma IV.1.

The following result shows that at this step there is only one pair of vectors (modulo ambiguities), out of the 2^L possible vectors of Step 2, that is consistent with the constraints. The proof is somewhat technical so we defer it to Appendix B.

Proposition V.1. *For generic choice of vectors $x \in \mathbb{C}^N$ and $w \in \mathbb{C}^W$, there is a unique choice among the 2^L possible vectors $(x'[0], \dots, x'[2L], w'[0], \dots, w'[2L])$ that is consistent with (V.3) and (V.4) such that the equations*

$$\left\{ \left| \sum_{n=0}^{3L} x'[n]w'[3L-n]e^{2\pi i kn/N} \right| = |\hat{y}[k, m]| \right\}_{k=0, \dots, 2L}$$

have a solution for the unknown $x'[i]$ and $w'[j]$ for $i, j = 0, \dots, 2L$.

From Proposition V.1, we have uniquely determined the vectors y_0, y_1, y_2, y_3 , modulo the trivial ambiguities of Proposition III.4, and we can now proceed recursively.

Step m : Note that knowledge of y_0, \dots, y_{m-1} determines, modulo trivial ambiguities, the entries $(x'[0], \dots, x'[(m-1)L])$ and $(w'[0], \dots, w'[(m-1)L])$. Then, if $m > 3$ and y_0, \dots, y_{m-1} are known, then the entries $y_m[n] = x'[n]w'[mL - n]$ are known if $L \leq n \leq (m-1)L$. Hence, from Lemma IV.1, a pair of generic signals (x, w) is uniquely determined by the Fourier intensity function $A_m(\eta)$. A priori, we need $2mL + 1$ intensity measurements to determine the trigonometric polynomial $A_m(\eta)$. However, the only unknown entries in the vector y_m are the $2L$ coefficients $S = [1, L] \cup [mL - L, mL - 1]$. By proposition IV.2, we can therefore uniquely determine $A_m(\eta)$ from at most $10L - 3$ measurements.

This concludes the proof.

D. Proof of Proposition IV.2

We begin by expressing $A(\eta)$ as a trigonometric polynomial. Let us write $x[n] = u[n] + v[n]$, so

$$A(\eta) = \left(\sum_{n=0}^{N-1} u[n] \cos(n\eta) - v[n] \sin(n\eta) \right)^2 + \left(\sum_{n=0}^{N-1} u[n] \sin(n\eta) + v[n] \cos(n\eta) \right)^2.$$

Now, we notice that the right hand side can be expanded in terms of the $4N^2$ quadratic monomials

$$\{u[n]u[m], u[n]v[m], v[n]v[m]\}_{0 \leq n, m \leq N-1}.$$

Writing these expressions explicitly, we get the following coefficients:

- 1) the coefficient of $u[n]u[m]$ is $\cos(n\eta)\cos(m\eta) + \sin(n\eta)\sin(m\eta) = \cos((n-m)\eta) = \cos(|n-m|\eta)$;
- 2) the coefficient of $v[n]v[m]$ is $\cos(|n-m|\eta)$;
- 3) the coefficient of $u[n]v[m]$ is $\sin((n-m)\eta) = \pm(\sin(|n-m|\eta))$.

Since we assume to know the entries for $n \in S^c$, we denote $u[n] = c_n$, $v[n] = d_n$ and write $x[n] = u[n] + v[n]$ only for the unknown set S . For fixed value of η , we can then write the corresponding coefficients explicitly:

- 1) for $u[n]u[m]$ we have $(|S|^2 + |S|)/2$ coefficients given by $\cos(|n-m|\eta)$ for $m, n \in S$;
- 2) for $v[n]v[m]$ we have $(|S|^2 + |S|)/2$ coefficients given by $\cos(|n-m|\eta)$ for $m, n \in S$;
- 3) for $u[n]v[m]$ we have $|S|^2 - |S|$ coefficients given by $\sin((n-m)\eta)$ for $m, n \in S$ when excluding $m = n$ (since in this case the coefficients are zeros);
- 4) for $u[n]$ we have $|S|$ coefficients given by $\sum_{m \in S^c} c_m \cos(n-m)\eta + d_m \sin((n-m)\eta)$ for $n \in S$;
- 5) for $v[n]$ we have $|S|$ coefficients given by $\sum_{m \in S^c} -c_m \sin((n-m)\eta) + d_m \cos((n-m)\eta)$ for $n \in S$;
- 6) the constant term is $\sum_{n, m \in S^c} c_n d_m \sin(n-m)\eta + (c_n c_m + d_n d_m) \cos((n-m)\eta)$.

Thus, for any $\eta \in S^1$, the value of $A(\eta)$ provides a linear equation in the $2|S|^2 + 2|S|$ unknowns

$$\mathcal{U} = \{u[n]u[m], u[n]v[m], v[n]v[m]\}_{n, m \in S} \cup \{u[n], v[n]\}_{n \in S}.$$

If we have k intensity measurements $A(\eta_1), \dots, A(\eta_k)$, then we obtain k linear equations in the unknowns \mathcal{U} . However, we note that the coefficients of the quadratic terms

$$\{u[n]v[m], v[n]v[m], u[n]v[m]\}_{n, m \in S}$$

are in the set $\{\cos(n-m), \pm \sin(n-m)\}_{n-m \in S-S}$. Therefore, the rank of the linear system is at most $2|S-S| - 1 + 2|S|$, where we subtract one because $\sin(n-m) = 0$ for $m = n$. It follows that if $k \geq 2|S-S| - 1 + 2|S|$ and $A(\eta_1), \dots, A(\eta_k)$ are known for distinct η_1, \dots, η_k , then $A(\eta)$ can be computed for any value of η . Lemma IV.1 implies that in this case, a generic x can be determined uniquely.

To explicitly compute $A(\eta)$ from $A(\eta_1), \dots, A(\eta_k)$, we consider the system of linear equations in the variables \mathcal{U} :

$$\begin{aligned} f(\eta_1) &= A(\eta_1) - D(\eta_1) \\ f(\eta_2) &= A(\eta_2) - D(\eta_2) \\ &\vdots \\ f(\eta_k) &= A(\eta_k) - D(\eta_k), \end{aligned} \quad (\text{V.5})$$

where for any value η

$$\begin{aligned} f(\eta) &:= \sum_{n, m \in S} (\cos \eta(n-m)) (u[n]v[m] + v[n]v[m]) \\ &\quad + \sin(\eta(n-m)) u[n]v[m] \\ &\quad + \sum_{n \in S, m \in S^c} (\cos(\eta(n-m))) u[m] \\ &\quad + \sin(\eta(n-m)) v[m]u[n] \\ &\quad + \sum_{n \in S, m \in S^c} (-\sin(\eta(n-m))) u[m] \\ &\quad + \cos(\eta(n-m)) v[m]v[n] \end{aligned}$$

is a linear combination of the variables in the set \mathcal{U} and

$$D(\eta) := \sum_{n, m \in S^c} c_n d_m (\sin(n-m)\eta) + (c_n c_m + d_n d_m) \cos((n-m)\eta)$$

is known. Accordingly, the right-hand side of (V.5) contains k known scalars, and the left-hand side is composed of a linear combination of the unknown set \mathcal{U} ; the particular linear combination depends on the signal's known entries (the set S^c) and the sampling points η_1, \dots, η_k . Since the rank of the coefficient matrix of this system is at most $2|S-S| - 1 + 2|S|$ we know that it is overdetermined. Hence, for fixed η , there exist (not necessarily unique) scalars $\lambda_1, \dots, \lambda_k$ such that

$$f(\eta) = \lambda_1 f(\eta_1) + \dots + \lambda_k f(\eta_k).$$

We also know that the system is consistent since $A(\eta_1), \dots, A(\eta_k), A(\eta)$ are Fourier intensity measurements from an actual vector. It follows that

$$A(\eta) - D(\eta) = \lambda_1 (A(\eta_1) - D(\eta_1)) + \dots + \lambda_k (A(\eta_k) - D(\eta_k)),$$

which allows us to determine $A(\eta)$ since $D(\eta)$ is known.

An important question from a practical point of view is whether the linear system (V.5) is stable to errors. We leave this analysis for future research.

APPENDIX

A. Necessary background on group theory

For completeness, we briefly present some basic definitions in group theory used throughout the proofs. Most of the definitions are taken from [53], yet they also appear in any standard book on group theory.

We start with the general definition of a group.

Definition A.1. A group is a set G together with an operation \circ such that for any $g_1, g_2, g_3 \in G$ the following properties hold:

- (closure) if $g_1, g_2 \in G$, then $g_1 \circ g_2 \in G$;
- (associativity) $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$;

- (existence of an identity) there exists an element $1 \in G$ such that $1 \circ g = g \circ 1 = g$ for any $g \in G$;
- (existence of an inverse) for every element $g \in G$, there exists an element $g^{-1} \in G$ such that $g^{-1} \circ g = g \circ g^{-1} = 1$.

We next present the notion of subgroups and in particular normal subgroups.

Definition A.2. A subgroup H is a subset of a group $H \subseteq G$ which is itself a group, namely, it is closed under the group operation of G , $1 \in H$ and $h^{-1} \in H$ whenever $h \in H$. A subgroup N is called *normal* if $g^{-1}Ng = N$ for all $g \in G$.

For example, any subgroup of an Abelian group (a group satisfying $g_1 \circ g_2 = g_2 \circ g_1$ for all $g_1, g_2 \in G$) is normal.

Our next definition concerns the notions of cosets and equivalence classes:

Definition A.3. Given a subgroup H of G and any element $g \in G$, the left coset gH (similarly, the right coset Hg) is defined as

$$gH := \{g \circ h \mid h \in H\}.$$

Any group is divided into disjoint left (right) cosets and if g_1 and g_2 are in the same coset we denote $g_1 \sim g_2$. The relation between g_1 and g_2 is called an *equivalence relation*. The set of all left (right) cosets is called *coset space* and is denoted by G/H ($G \setminus H$).

Note that the sets of left and right cosets coincide for normal subgroups. We are now ready to define quotient groups:

Definition A.4. If N is a normal subgroup of G , then the coset space G/N together with the operation $(g_1N)(g_2N) = (g_1 \circ g_2)N$ is a group, called the *quotient group*.

The last notion we would like to mention is of a generically freely action of a group, which is used in Proposition III.4. A group G acts *generically freely* on a vector space V if for a generic vector $v \in V$, $gv = g'v$ for some $g, g' \in G$, if and only if $g = g'$. Equivalently, $gv = v$ if and only if g is the identity element. In other words, the only element that acts trivially on a generic vector is the identity. Below we provide a simple example to illustrate the definition:

Example A.5. Let $G = S^1 \times S^1$ acting on \mathbb{C}^4 , where we view G as consisting of diagonal matrices $\begin{pmatrix} e^{i\theta_1} & 0 \\ 0 & e^{i\theta_2} \end{pmatrix}$ and \mathbb{C}^4 as 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The action of G on \mathbb{C}^4 is by conjugation; that is, $g \circ x$ is defined to be the matrix

$$\begin{pmatrix} e^{i\theta_1} & 0 \\ 0 & e^{i\theta_2} \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e^{-i\theta_1} & 0 \\ 0 & e^{-i\theta_2} \end{pmatrix} \\ = \begin{pmatrix} a & be^{i(\theta_1-\theta_2)} \\ ce^{i(\theta_2-\theta_1)} & d \end{pmatrix}.$$

Now, observe that if $g = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix}$ for any θ , then $g \circ x = x$ for all matrices in \mathbb{C}^4 . Thus, there is a subgroup of $S^1 \times S^1$ which acts trivially on \mathbb{C}^4 , even though it consists of more than the identity element. This is an example of

an action which is not generically free. On the other hand, the quotient of $S^1 \times S^1$ by the subgroup of matrices where $\theta_1 = \theta_2$ acts generically free.

B. Proof of Proposition V.1

Let us define the vectors

$$\begin{aligned} y_0 &:= (x[0]w[0]) \in \mathbb{C}, \\ y_1 &:= (x[0]w[L], \dots, x[L]w[0]) \in \mathbb{C}^{L+1}, \\ y_2 &:= (x[0]w[2L], \dots, x[2L]w[0]) \in \mathbb{C}^{2L+1}, \\ y_3 &:= (x[0]w[3L], \dots, x[3L]w[0]) \in \mathbb{C}^{3L+1}. \end{aligned}$$

The entries of y_0, y_1, y_2, y_3 satisfy the following relations:

$$y_0[0]y_2[L] = y_1[0]y_1[L] \quad (\text{B.1})$$

$$\{y_1[n]y_3[L+n] = y_2[L+n]y_2[n]\}_{n=0, \dots, L}. \quad (\text{B.2})$$

Equation (B.1) implies that we may determine y_0 from the entries of y_1, y_2 provided that $y_2[L] \neq 0$, which is equivalent to assuming that $w[L]$ and $x[L]$ are nonzero. Thus, we may ignore y_0 . Therefore, we consider the three vectors y_1, y_2, y_3 , where y_1, y_2 are unconstrained generic vectors and the entries of y_3 depends on y_1, y_2 through (B.2).

The conclusion of Proposition V.1 follows from the following more general statement. In what follows, we use \tilde{z} to denote the reflected and conjugated version of $z \in \mathbb{C}^P$, that is, $\tilde{z}[n] = \overline{z[P-n]}$.

Proposition B.1. Let $y_1 \in \mathbb{C}^{L+1}, y_2 \in \mathbb{C}^{2L+1}$ and $y_3 \in \mathbb{C}^{3L+1}$ be generic vectors satisfying (B.2). Then, any solution y'_1, y'_2, y'_3 to the system of equations

$$\begin{aligned} \{y'_1[n]y'_3[L+n] = y'_2[n]y'_2[L+n]\}_{n=0, \dots, L} \\ \left| \hat{y}'_1(\eta) \right|^2 = \left| \hat{y}_1(\eta) \right|^2, \\ \left| \hat{y}'_2(\eta) \right|^2 = \left| \hat{y}_2(\eta) \right|^2, \\ \left| \hat{y}'_3(\eta) \right|^2 = \left| \hat{y}_3(\eta) \right|^2, \end{aligned}$$

must obey $y'_1 = e^{i\theta_1}y_1, y'_2 = e^{i\theta_2}y_2, y'_3 = e^{i\theta_3}y_3$ for some $\theta_1, \theta_2, \theta_3$ or $y'_1 = e^{i\theta_1}\tilde{y}_1, y'_2 = e^{i\theta_2}\tilde{y}_2, y'_3 = e^{i\theta_3}\tilde{y}_3$.

Proof. Let $Z \subset \mathbb{C}^{L+1} \times \mathbb{C}^{2L+1} \times \mathbb{C}^{3L+1}$ be the subvariety defined the system of equations (B.2). Note that Z is irreducible because it is the product of $L+1$ irreducible hypersurfaces, where the n th hypersurface satisfies the equation $y_1[n]y_3[n+L] = y_2[n]y_2[n+L]$ in the \mathbb{C}^4 with coordinates $(y_1[n], y_3[n+L], y_2[n], y_2[n+L])$ and $n = 0, \dots, L$.

Notice also that Z is invariant under the action of the group of ambiguities $S^1 \times S^1 \times \mu_2$, where $e^{i\theta_1}, e^{i\theta_2}$ acts on $(y_1, y_2, y_3) \in \mathbb{C}^{L+1} \times \mathbb{C}^{2L+1} \times \mathbb{C}^{3L+1}$ by $(e^{i\theta_1}y_1, e^{i\theta_2}y_2, e^{2i\theta_2-i\theta_1}y_3)$ and $-1 \in \mu_2$ acts by taking (y_1, y_2, y_3) to $(\tilde{y}_1, \tilde{y}_2, \tilde{y}_3)$. The action of $S^1 \times S^1 \times \mu_2$ also preserves the Fourier intensity of each vector.

Let X be the quotient of $\mathbb{C}^L \times \mathbb{C}^{2L} \times \mathbb{C}^{3L}$ and let H be the quotient of Z under the action of this group. An element of X is an equivalence class of triples (y_1, y_2, y_3) modulo this group of ambiguities and H is the subvariety of equivalence classes that satisfy the relations (B.2).

Consider the subvariety $W \subset H \times X$ of pairs of equivalence classes $(y_1, y_2, y_3), (y'_1, y'_2, y'_3)$, where y_ℓ and y'_ℓ have the same Fourier intensity function for $\ell = 1, 2, 3$. Let I_W be the intersection of W with the subvariety $H \times H$ of $H \times X$. Both W and I_W clearly contain the diagonal (that is, the set of pairs of equivalence classes $((y_1, y_2, y_3), (y_1, y_2, y_3))$ with $(y_1, y_2, y_3) \in H$).

The projection $W \rightarrow H$ (of taking the first triplet) is finite because, modulo trivial ambiguities, there are only a finite number of vectors with the same Fourier intensity function of a given vector (For further discussion of the geometry of related varieties, see [54]). Let W_0 be the closure of the complement of the diagonal in W and let I_W^0 be the closure of the complement of the diagonal in I_W .

The projection map $W_0 \rightarrow H$ is also finite and the assertion of the proposition is equivalent to the statement that the image of I_W^0 under the projection is contained in a proper (i.e., its complement is dense) subvariety of H . To see that, note that if the equivalence class of (y_1, y_2, y_3) is in the complement of the image of I_W^0 , then if (y'_1, y'_2, y'_3) is a triplet with same Fourier intensity function which also satisfies the equations

$$y'_1[n]y'_3[n+L] = y'_2[n]y'_2[n+L], \quad n = 0, \dots, L, \quad (\text{B.3})$$

we must have that (y'_1, y'_2, y'_3) is obtained from (y_1, y_2, y_3) by an action of $S^1 \times S^1 \times \mu_2$.

To this end, it suffices to show that the image of I_W^0 is not all of H . We will show this with an explicit example. Let $y_1 = (1, 0, \dots, 0, 2) \in \mathbb{C}^{L+1}$, and let y_2 be the vector with $y_2[0] = y_2[L] = y_2[2L] = 1$ and $y_2[n] = 0$ otherwise. Then, the set of vectors $y_3 \in \mathbb{C}^{3L}$ that satisfy the relations $y_1[n]y_3[n+L] = y_2[n]y_2[n+L]$ is the $(3L-1)$ -dimensional linear subspace of vectors satisfying

$$y_3[L] = 1 \text{ and } y_3[2L] = 1/2. \quad (\text{B.4})$$

The Fourier transform of y_2 is $1 + \eta^L + \eta^{2L}$. The roots of such a polynomial (called cyclotomic polynomial), are on the unit circle and are given by the $3L$ root of unity. Therefore, any vector with the same Fourier intensity must be obtained from y_2 by a trivial ambiguity; i.e., multiplication by $e^{i\theta}$ or conjugation and reflection.

The Fourier transform of y_1 is the function $1 + 2\eta^L$. Thus, its Fourier intensity function is given by $B(\eta) = 2\eta^{-L} + 5 + 2\eta^L$. We also note that if y'_1 is any vector with Fourier intensity function $B(\eta)$, then it must have at least 3 nonzero entries unless it is of the form $e^{i\theta}y_1$ or $e^{i\theta}\tilde{y}_1$. Hence, if y'_1 is not equivalent to y_1 under the group of trivial ambiguities, the solution space to the equations (B.3) must have dimension at most $3L-2$, which is inconsistent with (B.4). This implies that for generic y_3 in the linear subspace above there is no non-equivalent (y'_1, y'_2, y'_3) with y'_3 having the same Fourier intensity function as y_3 and also satisfying the equations (B.3). \square

ACKNOWLEDGMENT

Tamir Bendory is grateful to Ti-Yen Lan for discussions on ptychography.

REFERENCES

- [1] R. W. Harrison, "Phase problem in crystallography," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 10, no. 5, pp. 1046–1055, May 1993.
- [2] C. Giacobozzo, H. L. Monaco, G. Artioli, D. Viterbo, G. Ferraris, and C. Giacobozzo, "Fundamentals of crystallography," Tech. Rep., 2002.
- [3] A. Walther, "The question of phase retrieval in optics," *Opt. Acta*, vol. 10, no. 1, pp. 41–49, Jan. 1963.
- [4] Y. Schechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, Mar. 2015.
- [5] R. Trebino, *Frequency-Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses*. Boston, MA, USA: Springer, 2012.
- [6] C. Fienup and J. Dainty, "Phase retrieval and image reconstruction for astronomy," *Image Recovery, Theory Appl.*, vol. 231, p. 275, Jan. 1987.
- [7] B. Baykal, "Blind channel estimation via combining autocorrelation and blind phase estimation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 6, pp. 1125–1131, Jun. 2004.
- [8] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, vol. 14. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
- [9] T. Bendory, N. Boumal, C. Ma, Z. Zhao, and A. Singer, "Bispectrum inversion with application to multireference alignment," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 1037–1050, Feb. 2018.
- [10] T. Bendory, R. Beinert, and Y. C. Eldar, "Fourier phase retrieval: Uniqueness and algorithms," in *Proc. 2nd Int. MATHEON Conf. Compressed Sens. Appl.*, H. Boche, G. Caire, R. Calderbank, M. März, G. Kutyniok, and R. Mathar, Eds. Cham, Switzerland: Springer, 2017, pp. 55–91. doi: 10.1007/978-3-319-69802-1_2.
- [11] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval: An overview of recent developments," in *Optical Compressive Imaging*, A. Stern, Ed. Boca Raton, FL, USA: CRC Press, 2016.
- [12] V. Elser, T.-Y. Lan, and T. Bendory, "Benchmark problems for phase retrieval," *SIAM J. Imag. Sci.*, vol. 11, no. 4, pp. 2429–2455, 2018.
- [13] K. Huang, Y. C. Eldar, and N. D. Sidiropoulos, "Phase retrieval from 1D Fourier measurements: Convexity, uniqueness, and algorithms," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6105–6117, Dec. 2016.
- [14] P. Nellist, B. McCallum, and J. Rodenburg, "Resolution beyond the 'information limit' in transmission electron microscopy," *Nature*, vol. 374, no. 6523, pp. 630–632, Apr. 1995.
- [15] J. M. Rodenburg, "Ptychography and related diffractive imaging methods," *Adv. Imag. Electron Phys.*, vol. 150, no. 7, pp. 87–184, 2008.
- [16] A. M. Maiden, M. J. Humphry, F. Zhang, and J. M. Rodenburg, "Superresolution imaging via ptychography," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 28, no. 4, pp. 604–612, 2011.
- [17] T. Bendory, Y. C. Eldar, and N. Boumal, "Non-convex phase retrieval from STFT measurements," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 467–484, Jan. 2018.
- [18] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, "Probe retrieval in ptychographic coherent diffractive imaging," *Ultramicroscopy*, vol. 109, no. 4, pp. 338–343, Mar. 2009.
- [19] A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," *Ultramicroscopy*, vol. 109, no. 10, pp. 1256–1262, 2009.
- [20] M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: A nonlinear optimization approach," *Opt. Express*, vol. 16, no. 10, pp. 7264–7278, 2008.
- [21] D. J. Kane, "Principal components generalized projections: A review," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 25, no. 6, pp. A120–A132, 2008.
- [22] T. Bendory, P. Sidorenko, and Y. C. Eldar, "On the uniqueness of FROG methods," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 722–726, May 2017.
- [23] S. M. Jefferies and J. C. Christou, "Restoration of astronomical images by iterative blind deconvolution," *Astrophys. J.*, vol. 415, p. 862, Oct. 1993.
- [24] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312–321, Mar. 1990.
- [25] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.
- [26] G. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," *Opt. Lett.*, vol. 13, no. 7, pp. 547–549, 1988.
- [27] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in *Computer Vision—ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 783–798.
- [28] R. Beinert and G. Plonka, "Enforcing uniqueness in one-dimensional phase retrieval by additional signal information in time domain," *Appl. Comput. Harmon. Anal.*, vol. 45, no. 3, pp. 505–525, Nov. 2018.

- [29] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "STFT phase retrieval: Uniqueness guarantees and recovery algorithms," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 770–781, Jun. 2016.
- [30] Y. C. Eldar, P. Sidorenko, D. G. Mixon, S. Barel, and O. Cohen, "Sparse phase retrieval from short-time Fourier measurements," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 638–642, May 2015.
- [31] G. E. Pfander and P. Salanevich, "Robust phase retrieval algorithm for time-frequency structured measurements," *SIAM J. Imag. Sci.*, vol. 12, no. 2, pp. 736–761, 2019.
- [32] M. A. Iwen, A. Viswanathan, and Y. Wang, "Fast phase retrieval from local correlation measurements," *SIAM J. Imag. Sci.*, vol. 9, no. 4, pp. 1655–1688, 2016.
- [33] R. Xu *et al.*, "Accelerated Wirtinger flow: A fast algorithm for ptychography," Jun. 2018, *arXiv:1806.05546*. [Online]. Available: <https://arxiv.org/abs/1806.05546>
- [34] T. Bendory, D. Edidin, and Y. C. Eldar, "On signal reconstruction from FROG measurements," *Appl. Comput. Harmon. Anal.*, to be published.
- [35] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, "Toward single particle reconstruction without particle picking: Breaking the detection limit," Sep. 2018, *arXiv:1810.00226*. [Online]. Available: <https://arxiv.org/abs/1810.00226>
- [36] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, "Multi-target detection with application to cryo-electron microscopy," *Inverse Problems*, vol. 35, no. 10, Sep. 2019. [Online]. Available: <https://iopscience.iop.org/article/10.1088/1361-6420/ab2aec/meta>
- [37] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1711–1732, Mar. 2014.
- [38] Y. Li, K. Lee, and Y. Bresler, "Identifiability and stability in blind deconvolution under minimal assumptions," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4619–4633, Jul. 2017.
- [39] M. Kech and F. Kraemer, "Optimal injectivity conditions for bilinear inverse problems with applications to identifiability of deconvolution problems," *SIAM J. Appl. Algebra Geometry*, vol. 1, no. 1, pp. 20–37, 2017.
- [40] S. Ling and T. Strohmer, "Blind deconvolution meets blind demixing: Algorithms and performance bounds," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4497–4520, Jul. 2017.
- [41] Y. Zhang, H.-W. Kuo, and J. Wright, "Structured local optima in sparse blind deconvolution," Jun. 2018, *arXiv:1806.00338*. [Online]. Available: <https://arxiv.org/abs/1806.00338>
- [42] P. Walk, P. Jung, G. E. Pfander, and B. Hassibi, "Blind deconvolution with additional autocorrelations via convex programs," Jan. 2017, *arXiv:1701.04890*. [Online]. Available: <https://arxiv.org/abs/1701.04890>
- [43] P. Walk and B. Hassibi, "Stable blind deconvolution over the reals from additional autocorrelations," Oct. 2017, *arXiv:1710.07879*. [Online]. Available: <https://arxiv.org/abs/1710.07879>
- [44] S. Choudhary and U. Mitra, "Fundamental limits of blind deconvolution part i: Ambiguity kernel," Nov. 2014, *arXiv:1411.3810*. [Online]. Available: <https://arxiv.org/abs/1411.3810>
- [45] Y. Li, K. Lee, and Y. Bresler, "Identifiability in blind deconvolution with subspace or sparsity constraints," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4266–4275, Jul. 2016.
- [46] S. Choudhary and U. Mitra, "Sparse blind deconvolution: What cannot be done," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 3002–3006.
- [47] A. Fannjiang and P. Chen, "Blind ptychography: Uniqueness and ambiguities," Jun. 2018, *arXiv:1806.02674*. [Online]. Available: <https://arxiv.org/abs/1806.02674>
- [48] A. Barnett, C. L. Epstein, L. Greengard, and J. Magland, "Geometry of the phase retrieval problem," Aug. 2018, *arXiv:1808.10747*. [Online]. Available: <https://arxiv.org/abs/1808.10747>
- [49] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, no. 2, pp. 140–154, Apr. 1982.
- [50] A. Ahmed, A. Aghasi, and P. Hand, "Blind deconvolutional phase retrieval via convex programming," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 10051–10061.
- [51] R. Beinert and G. Plonka, "Ambiguities in one-dimensional discrete phase retrieval from Fourier magnitudes," *J. Fourier Anal. Appl.*, vol. 21, no. 6, pp. 1169–1198, Dec. 2015.
- [52] L. Xu, P. Yan, and T. Chang, "Almost unique specification of discrete finite length signal: From its end point and Fourier transform magnitude," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 12, Apr. 1987, pp. 2097–2100.
- [53] G. S. Chirikjian and A. B. Kyatkin, *Harmonic Analysis for Engineers and Applied Scientists: Updated and Expanded Edition*. New York, NY, USA: Dover, 2016.
- [54] D. Edidin, "The geometry of ambiguity in one-dimensional phase retrieval," Nov. 2018, *arXiv:1811.06963*. [Online]. Available: <https://arxiv.org/abs/1811.06963>