

# STFT Phase Retrieval: Uniqueness Guarantees and Recovery Algorithms

Kishore Jaganathan, Yonina C. Eldar, *Fellow, IEEE*, and Babak Hassibi, *Member, IEEE*

**Abstract**—The problem of recovering a signal from its Fourier magnitude is of paramount importance in various fields of engineering and applied physics. Due to the absence of Fourier phase information, some form of additional information is required in order to be able to uniquely, efficiently, and robustly identify the underlying signal. Inspired by practical methods in optical imaging, we consider the problem of signal reconstruction from the short-time Fourier transform (STFT) magnitude. We first develop conditions under, which the STFT magnitude is an *almost surely unique signal representation*. We then consider a semidefinite relaxation-based algorithm (STliFT) and provide recovery guarantees. Numerical simulations complement our theoretical analysis and provide directions for future work.

**Index Terms**—Short-Time Fourier Transform (STFT), Phase Retrieval, Super-Resolution, Semidefinite Relaxation.

## I. INTRODUCTION

**I**N MANY physical measurement systems, the measurable quantity is the magnitude-square of the Fourier transform of the underlying signal. The problem of reconstructing a signal from its Fourier magnitude is known as phase retrieval [1], [2]. This reconstruction problem is one with a rich history and occurs in many areas of engineering and applied physics such as optics [3], X-ray crystallography [4], astronomical imaging [5], speech recognition [6], computational biology [7], blind channel estimation [8] and more. We refer the readers to [9]–[11] for a comprehensive survey of classical approaches. Recent reviews can be found in [12], [13].

It is well known that phase retrieval is an ill-posed problem [14]. In order to be able to uniquely identify the underlying signal, various methods have been explored, which can be broadly classified into two categories: (i) *Additional prior information*: common approaches include bounds on the support of the signal [9]–[11] and sparsity constraints [17]–[24]. (ii) *Additional*

*magnitude-only measurements*: popular examples include the use of structured illuminations and masks [24]–[30], and Short-Time Fourier Transform (STFT) magnitude measurements [31]–[39].

We consider STFT phase retrieval, which is the problem of reconstructing a signal from its STFT magnitude. In some applications of phase retrieval, it is easy to obtain such measurements. One example is Frequency Resolved Optical Gating (FROG), which is a general method for measuring ultrashort laser pulses [31]. Fourier ptychography [32]–[34], a technology which has enabled X-ray, optical and electron microscopy with increased spatial resolution without the need for advanced lenses, is another popular example. In applications such as speech processing, it is natural to work with the STFT instead of the Fourier transform as the spectral content of speech changes over time [35]. The key idea, when using STFT measurements, is to introduce redundancy in the magnitude-only measurements by maintaining a substantial overlap between adjacent short-time sections. This mitigates the uniqueness and algorithmic issues of phase retrieval.

In this work, our contribution is two-fold:

(i) *Uniqueness guarantees*: Researchers have previously developed conditions under which the STFT magnitude uniquely identifies signals (up to a global phase). However, either prior information on the signal is assumed in order to provide the guarantees, or the guarantees are limited. For instance, the results provided in [37] require exact knowledge of a small portion of the underlying signal. In [39], the guarantees developed are for the setup in which adjacent short-time sections differ in only one index. These limitations are primarily due to a small number of adversarial signals which cannot be uniquely identified from their STFT magnitude. Here, in contrast, we develop conditions under which the STFT magnitude is an *almost surely unique signal representation*. In particular, we show that, with the exception of a set of signals of measure zero, non-vanishing signals can be uniquely identified (up to a global phase) from their STFT magnitude if adjacent short-time sections overlap (Theorem III.1). We then extend this result to incorporate sparse signals which have a limited number of consecutive zeros (Corollary III.1).

(ii) *Recovery algorithms*: Researchers have previously developed efficient iterative algorithms based on classic optimization frameworks to solve the STFT phase retrieval problem. Examples include the Griffin-Lim (GL) algorithm [38] and STFT-GESPAR for sparse signals [39]. While these techniques work well in practice, they do not have theoretical guarantees. In [36] and [43], a semidefinite relaxation-based STFT phase retrieval algorithm, called STliFT (see Algorithm 1 below),

Manuscript received August 07, 2015; revised January 15, 2016; accepted March 15, 2016. Date of publication March 31, 2016; date of current version May 11, 2016. The work of K. Jaganathan and B. Hassibi was supported in part by the National Science Foundation under Grants CCF-0729203, CNS-0932428, and CIF-1018927, in part by the Office of Naval Research under the MURI Grant N00014-08-1-0747, and in part by a grant from Qualcomm Inc. and the work of Y. C. Eldar was supported in part by the European Union's Horizon 2020 Research and Innovation Program through the ERC-BNYQ Project, and in part by the Israel Science Foundation under Grant 335/14. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Justin Romberg.

K. Jaganathan and B. Hassibi are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: kishore@caltech.edu).

Y. C. Eldar is with the Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTSP.2016.2549507

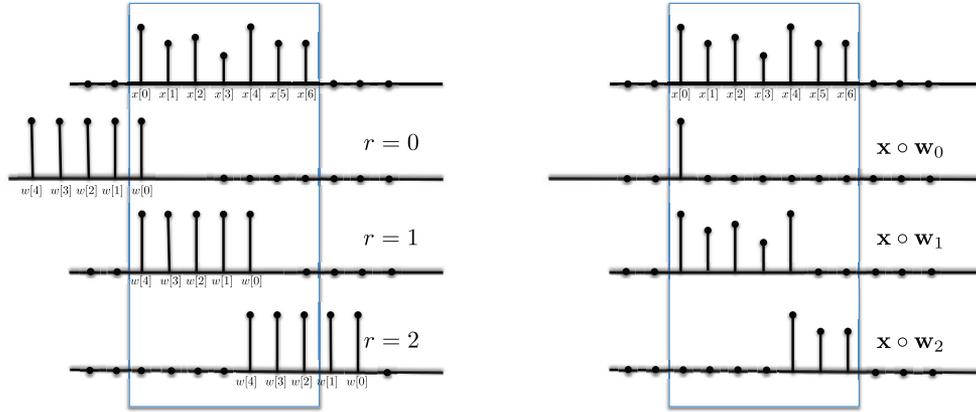


Fig. 1. Sliding window interpretation of the STFT for  $N = 7$ ,  $W = 5$  and  $L = 4$ . The shifted window overlaps with the signal for 3 shifts, and hence  $R = 3$  short-time sections are considered.

was proposed. In this work, we conduct extensive numerical simulations and provide theoretical guarantees for STliFT. In particular, we conjecture that STliFT can recover most non-vanishing signals (up to a global phase) from their STFT magnitude if adjacent short-time sections differ in at most half the indices (Conjecture IV.1). When this condition is satisfied, we argue that one can super-resolve (i.e., discard high frequency measurements) and reduce the number of measurements to  $(4 + o(1))N$ , where  $N$  is the length of the complex signal. Therefore, STliFT recovers most non-vanishing signals uniquely, efficiently and robustly, using an order-wise optimal number of phaseless measurements.

We prove this conjecture for the setup in which the exact knowledge of a small portion of the underlying signal is available (Theorem IV.1). For particular choices of STFT parameters, this portion vanishes asymptotically, due to which this setup is asymptotically reasonable. We also prove this conjecture for the case in which adjacent short-time sections differ in only one index (Theorem IV.2). We then extend these results to incorporate sparse signals which have a limited number of consecutive zeros (Corollary IV.1).

The rest of the paper is organized as follows. In Section 2, we mathematically formulate STFT phase retrieval and establish our notation. We present uniqueness guarantees in Section 3. Section 4 considers the STliFT algorithm and provides recovery guarantees. Numerical simulations are presented in Section 5. Section 6 concludes the paper.

## II. PROBLEM SETUP

Let  $\mathbf{x} = (x[0], x[1], \dots, x[N-1])^T$  be a signal of length  $N$  and  $\mathbf{w} = (w[0], w[1], \dots, w[W-1])^T$  be a window of length  $W$ . The STFT of  $\mathbf{x}$  with respect to  $\mathbf{w}$ , denoted by  $\mathbf{Y}_w$ , is defined as:

$$Y_w[m, r] = \sum_{n=0}^{N-1} x[n] w[rL - n] e^{-i2\pi \frac{mn}{N}} \quad (1)$$

for  $0 \leq m \leq N-1$  and  $0 \leq r \leq R-1$ , where the parameter  $L$  denotes the separation in time between adjacent short-time sections and the parameter  $R = \lceil \frac{N+W-1}{L} \rceil$  denotes the number of short-time sections considered.

The STFT can be interpreted as follows: Suppose  $w_r$  denotes the signal obtained by shifting the flipped window  $\mathbf{w}$  by  $rL$  time units (i.e.,  $w_r[n] = w[rL - n]$ ) and  $\circ$  is the Hadamard (element-wise) product operator. The  $r$ th column of  $\mathbf{Y}_w$ , for  $0 \leq r \leq R-1$ , corresponds to the  $N$  point DFT of  $\mathbf{x} \circ \mathbf{w}_r$ . In essence, the window is flipped and slid across the signal (see Figure 1 for a pictorial representation), and  $\mathbf{Y}_w$  corresponds to the Fourier transform of the windowed signal recorded at regular intervals. This interpretation is known as the *sliding window* interpretation.

Let  $\mathbf{Z}_w$  be the  $N \times R$  measurements corresponding to the magnitude-square of the STFT of  $\mathbf{x}$  with respect to  $\mathbf{w}$  so that  $Z_w[m, r] = |Y_w[m, r]|^2$ . Let  $\mathbf{W}_r$ , for  $0 \leq r \leq R-1$ , be the  $N \times N$  diagonal matrix with diagonal elements  $(w_r[0], w_r[1], \dots, w_r[N-1])$ . STFT phase retrieval can be mathematically stated as:

$$\begin{aligned} &\text{find} && \mathbf{x} \\ &\text{subject to} && Z_w[m, r] = |\langle \mathbf{f}_m, \mathbf{W}_r \mathbf{x} \rangle|^2 \end{aligned} \quad (2)$$

for  $0 \leq m \leq N-1$  and  $0 \leq r \leq R-1$ , where  $\mathbf{f}_m$  is the conjugate of the  $m$ th column of the  $N$  point DFT matrix and  $\langle \cdot, \cdot \rangle$  is the inner product operator. In fact, STFT phase retrieval can be equivalently stated by only considering the measurements corresponding to  $0 \leq m \leq M-1$  and  $0 \leq r \leq R-1$ , for any  $M$  satisfying  $2W \leq M \leq N$  (see Section VII for details). This equivalence significantly reduces the number of measurements considered when  $W \ll N$ , which is typically the case in practical methods.

We use the following definitions: A signal  $\mathbf{x}$  is said to be non-vanishing if  $x[n] \neq 0$  for all  $0 \leq n \leq N-1$ . Similarly, a window  $\mathbf{w}$  is said to be non-vanishing if  $w[n] \neq 0$  for all  $0 \leq n \leq W-1$ . Further, a signal  $\mathbf{x}$  is said to be sparse if it is not non-vanishing, i.e.,  $x[n] = 0$  for at least one  $0 \leq n \leq N-1$ .

## III. UNIQUENESS GUARANTEES

In this section, we review existing results regarding the uniqueness of STFT phase retrieval and present our uniqueness guarantees. The results are summarized in Table I.

In STFT phase retrieval, the global phase of the signal cannot be determined due to the fact that signals  $\mathbf{x}$  and  $e^{i\phi} \mathbf{x}$ , for 3

TABLE I  
UNIQUENESS RESULTS FOR STFT PHASE RETRIEVAL

Non-vanishing signals $\{x[n] \neq 0 \text{ for all } 0 \leq n \leq N - 1\}$	Uniqueness if the first $L$ samples are known a priori, $2L \leq W \leq \frac{N}{2}$ and $\mathbf{w}$ is non-vanishing [37]
	Uniqueness up to a global phase if $L = 1$ , $2 \leq W \leq \frac{N+1}{2}$ , $W - 1$ coprime with $N$ and mild conditions on $\mathbf{w}$ [39]
	Uniqueness up to a global phase for almost all signals if $L < W \leq \frac{N}{2}$ and $\mathbf{w}$ is non-vanishing [This work]
Sparse signals $\{x[n] = 0 \text{ for at least one } 0 \leq n \leq N - 1\}$	Uniqueness for signals with at most $W - 2L$ consecutive zeros if the first $L$ samples, starting from the first non-zero sample, are known a priori, $2L \leq W \leq \frac{N}{2}$ and $\mathbf{w}$ is non-vanishing [37]
	No uniqueness for most signals with $W$ consecutive zeros [39]
	Uniqueness up to a global phase and time-shift for almost all signals with less than $\min\{W - L, L\}$ consecutive zeros if $L < W \leq \frac{N}{2}$ and $\mathbf{w}$ is non-vanishing [This work]

any  $\phi$ , always have the same STFT magnitude regardless of the choice of  $\{\mathbf{w}, L\}$ . In contrast, in classic phase retrieval, signals which differ from each other by a global phase, time-shift and/or conjugate-flip (together called *trivial ambiguities*) cannot be distinguished from each other as they have the same Fourier magnitude [12], [17], [21].

Observe that  $L < W$  is a necessary condition in order to be able to uniquely identify most signals. If  $L > W$ , then the STFT magnitude does not contain any information from some locations of the signal. When  $L = W$ , adjacent short-time sections do not overlap and hence STFT phase retrieval is equivalent to a series of non-overlapping instances of classic phase retrieval. Since there is no way of determining the relative phase, time-shift or conjugate-flip between the windowed signals corresponding to the various short-time sections, most signals cannot be uniquely identified. For example, suppose  $\{\mathbf{w}, L\}$  is chosen such that  $L = W = 2$  and  $w[n] = 1$  for all  $0 \leq n \leq W - 1$ . Consider the signal  $\mathbf{x}_1 = (1, 2, 3)^T$  of length  $N = 3$ . Signals  $\mathbf{x}_1$  and  $\mathbf{x}_2 = (1, -2, -3)^T$  have the same STFT magnitude. In fact, more generally, signals  $\mathbf{x}_1$  and  $(1, e^{i\phi}2, e^{i\phi}3)^T$ , for any  $\phi$ , have the same STFT magnitude.

#### A. Non-vanishing Signals

For some specific choices of  $\{\mathbf{w}, L\}$ , it has been shown that all non-vanishing signals can be uniquely identified from their STFT magnitude up to a global phase. In [39], it is proven that the STFT magnitude uniquely identifies non-vanishing signals up to a global phase for  $L = 1$  if the window  $\mathbf{w}$  is chosen such that the  $N$  point DFT of  $(|w[0]|^2, |w[1]|^2, \dots, |w[N-1]|^2)$  is non-vanishing,  $2 \leq W \leq \frac{N+1}{2}$  and  $W - 1$  is coprime with  $N$ . In [37], the authors prove that if the first  $L$  samples are known a priori, then the STFT magnitude can uniquely identify non-vanishing signals for any  $L$  if the window  $\mathbf{w}$  is chosen such that it is non-vanishing and  $2L \leq W \leq \frac{N}{2}$ .

In this work, we prove the following result for non-vanishing signals:

*Theorem III.1:* Almost all non-vanishing signals can be uniquely identified (up to a global phase) from their STFT magnitude if  $\{\mathbf{w}, L\}$  satisfy

- (i)  $\mathbf{w}$  is non-vanishing

- (ii)  $L < W \leq \frac{N}{2}$ .

*Proof:* The proof is based on a technique commonly known as *dimension counting*. The outline is as follows (see Section VIII for details):

Consider the short-time sections  $r$  and  $r + 1$ . Since adjacent short-time sections overlap (due to  $L < W$ ), there exists at least one index, say  $n_0$ , where both  $\mathbf{x} \circ \mathbf{w}_r$  and  $\mathbf{x} \circ \mathbf{w}_{r+1}$  have non-zero values.

Since  $W \leq \frac{N}{2}$ , there can be at most  $2^W$  distinct windowed signals  $\mathbf{x} \circ \mathbf{w}_r$  (up to a phase) that have the same Fourier magnitude [14]. Consequently,  $|x[n_0]|$  is restricted to  $2^W$  values by the  $r$ th column of the STFT magnitude (let  $\mathcal{S}_r$  denote the set of these values). Similarly,  $|x[n_0]|$  is restricted to  $2^W$  values by the  $r + 1$ th column of the STFT magnitude (denote the set of these values by  $\mathcal{S}_{r+1}$ ).

By construction,  $\mathcal{S}_r \cap \mathcal{S}_{r+1} \neq \emptyset$  as the STFT magnitude is generated by an underlying signal  $\mathbf{x}_0$ , i.e.,  $|x_0[n_0]| \in \mathcal{S}_r \cap \mathcal{S}_{r+1}$ . Using Lemma VIII.1 and Theorem VIII.1, we show that, for almost all non-vanishing signals,  $\mathcal{S}_r \cap \mathcal{S}_{r+1}$  has cardinality one. In other words,  $\mathbf{x} \circ \mathbf{w}_r$  is uniquely identified (up to a phase) almost surely.

Since adjacent short-time sections overlap, non-vanishing signals are uniquely identified up to a *global* phase from the knowledge of  $\mathbf{x} \circ \mathbf{w}_r$  (up to a phase) for  $0 \leq r \leq R - 1$  if  $\mathbf{w}$  is non-vanishing. ■

#### B. Sparse Signals

While the aforementioned results provide guarantees for non-vanishing signals, they do not say anything about sparse signals. Reconstruction of sparse signals involves certain challenges which are not encountered in the reconstruction of non-vanishing signals.

The following example is provided in [39] to show that the time-shift ambiguity cannot be resolved for some classes of sparse signals and some choices of  $\{\mathbf{w}, L\}$ : Suppose  $\{\mathbf{w}, L\}$  is chosen such that  $L \geq 2$ ,  $W$  is a multiple of  $L$  and  $w[n] = 1$  for all  $0 \leq n \leq W - 1$ . Consider a signal  $\mathbf{x}_1$  of length  $N \geq L + 1$  such that it has non-zero values only within an interval of the form  $[(t - 1)L + 1, (t - 1)L + L - p] \subset [0, N - 1]$  for some

integers  $1 \leq p \leq L - 1$  and  $t \geq 1$ . The signal  $\mathbf{x}_2$  obtained by time-shifting  $\mathbf{x}_1$  by  $q \leq p$  units (i.e.,  $x_2[i] = x_1[i - q]$ ) has the same STFT magnitude. The issue with this class of sparse signals is that the STFT magnitude is identical to the Fourier magnitude because of which the time-shift and conjugate-flip ambiguities cannot be resolved.

It is also shown that some sparse signals cannot be uniquely recovered even up to the trivial ambiguities for some choices of  $\{\mathbf{w}, L\}$  using the following example: Consider two non-overlapping intervals  $[u_1, v_1], [u_2, v_2] \subset [0, N - 1]$  such that  $u_2 - v_1 > W$ , and take a signal  $\mathbf{x}_1$  supported on  $[u_1, v_1]$  and  $\mathbf{x}_2$  supported on  $[u_2, v_2]$ . The magnitude-square of the STFT of  $\mathbf{x}_1 + \mathbf{x}_2$  and of  $\mathbf{x}_1 - \mathbf{x}_2$  are equal for any choice of  $L$ . The difficulty with this class of sparse signals is that the two intervals with non-zero values are separated by a distance greater than  $W$  because of which there is no way of establishing relative phase using a window of length  $W$ .

These examples establish the fact that sparse signals are harder to recover than non-vanishing signals in this setup. Since the aforementioned issues are primarily due to a large number of consecutive zeros, the uniqueness guarantees for non-vanishing signals have been extended to incorporate sparse signals with limits on the number of consecutive zeros. In [37], it was shown that if  $L$  consecutive samples, starting from the first non-zero sample, are known a priori, then the STFT magnitude can uniquely identify signals with less than  $W - 2L$  consecutive zeros for any  $L$  if the window  $\mathbf{w}$  is chosen such that it is non-vanishing and  $2L \leq W \leq \frac{N}{2}$ .

Below, we extend Theorem III.1 to prove the following result for sparse signals:

*Corollary III.1:* Almost all sparse signals with less than  $\min\{W - L, L\}$  consecutive zeros can be uniquely identified (up to a global phase and time-shift) from their STFT magnitude if  $\{\mathbf{w}, L\}$  satisfy

- 1)  $\mathbf{w}$  is non-vanishing
- 2)  $L < W \leq \frac{N}{2}$ .

*Proof:* The  $\min\{W - L, L\}$  bound on consecutive zeros ensures the following: For sufficient pairs of adjacent short-time sections, there is at least one index among the overlapping and non-overlapping indices respectively, where the underlying signal has a non-zero value. We refer the readers to Section IX for details. ■

#### IV. RECOVERY ALGORITHMS

The classic alternating projection algorithm to solve phase retrieval [9] has been adapted to solve STFT phase retrieval by Griffin and Lim [38]. To this end, STFT phase retrieval is reformulated as the following least-squares problem:

$$\min_{\mathbf{x}} \sum_{r=0}^{R-1} \sum_{m=0}^{N-1} \left( Z_w[m, r] - |\langle \mathbf{f}_m, \mathbf{W}_r \mathbf{x} \rangle|^2 \right)^2. \quad (3)$$

The Griffin-Lim (GL) algorithm attempts to minimize this objective by starting with a random initialization and imposing the time domain and STFT magnitude constraints alternately using projections. The objective is shown to be monotonically decreasing as the iterations progress. An important feature of

---

#### Algorithm 1. STliFT

---

**Input:** STFT magnitude measurements  $Z_w[m, r]$  for  $1 \leq m \leq M$  and  $0 \leq r \leq R - 1$ ,  $\{\mathbf{w}, L\}$ .

**Output:** Estimate  $\hat{\mathbf{x}}$  of the underlying signal  $\mathbf{x}_0$ .

- Obtain  $\hat{\mathbf{X}}$  by solving:

$$\begin{aligned} & \text{minimize} && \text{trace}(\mathbf{X}) \\ & \text{subject to} && Z_w[m, r] = \text{trace}(\mathbf{W}_r^* \mathbf{f}_m \mathbf{f}_m^* \mathbf{W}_r \mathbf{X}) \\ & && \mathbf{X} \succcurlyeq 0 \end{aligned} \quad (4)$$

for  $1 \leq m \leq M$  and  $0 \leq r \leq R - 1$ .

- Return  $\hat{\mathbf{x}}$ , where  $\hat{\mathbf{x}}\hat{\mathbf{x}}^*$  is the best rank-one approximation of  $\hat{\mathbf{X}}$ .
- 

the GL algorithm is its empirical ability to converge to the global minimum when there is substantial overlap between adjacent short-time sections. However, no theoretical recovery guarantees are available. To establish such guarantees, we rely on a semidefinite relaxation approach.

##### A. Semidefinite Relaxation-Based Algorithm

Semidefinite relaxation has enjoyed considerable success in provably and stably solving several quadratic-constrained problems [40]–[42]. The steps to formulate such problems as a semidefinite program (SDP) are as follows: (i) Embed the problem in a higher dimensional space using the transformation  $\mathbf{X} = \mathbf{x}\mathbf{x}^*$ , a process which converts the problem of recovering a signal with quadratic constraints into a problem of recovering a rank-one matrix with affine constraints. (ii) Relax the rank-one constraint to obtain a convex program.

If the convex program has a unique solution  $\mathbf{X}_0 = \mathbf{x}_0\mathbf{x}_0^*$ , then  $\mathbf{x}_0$  is the unique solution to the quadratic-constrained problem (up to a global phase). Many recent results in related problems like *generalized phase retrieval* [42] and *phase retrieval using random masks* [28], [29] suggest that one can provide conditions, which when satisfied, ensure that the convex program has a unique solution  $\mathbf{X}_0 = \mathbf{x}_0\mathbf{x}_0^*$ .

A semidefinite relaxation-based STFT phase retrieval algorithm, called STliFT, was explored in [36] and [43]. The details of the algorithm are provided in Algorithm 1. In the following, we develop conditions on  $\{\mathbf{x}_0, \mathbf{w}, L\}$  which ensure that the convex program (4) has  $\mathbf{X}_0 = \mathbf{x}_0\mathbf{x}_0^*$  as the unique solution. Consequently, under these conditions, STliFT uniquely recovers the underlying signal up to a global phase.

Based on extensive numerical simulations, we conjecture the following:

*Conjecture IV.1:* The convex program (4) has a unique solution  $\mathbf{X}_0 = \mathbf{x}_0\mathbf{x}_0^*$ , for most non-vanishing signals  $\mathbf{x}_0$ , if

- (i)  $\mathbf{w}$  is non-vanishing
- (ii)  $2L \leq W \leq \frac{N}{2}$
- (iii)  $4L \leq M \leq N$ .

The number of phaseless measurements considered can be calculated as follows: The total number of short-time sections is  $\lceil \frac{N+W-1}{L} \rceil$ . For each short-time section,  $M = 4L$  phaseless measurements are sufficient. Hence, the total number of phaseless measurements is  $\lceil \frac{N+W-1}{L} \rceil \times 4L \leq 4(N + W) +$

$2W$ . Consequently, when  $W = o(N)$ , this number is  $(4 + o(1))N$ , which is order-wise optimal. In fact, in generalized phase retrieval, it is conjectured that  $(4 - o(1))N$  phaseless measurements are necessary [45].

The proof techniques used in [28] and [42] are not applicable in the STFT setup. In [28] and [42], the measurement vectors are chosen from a random distribution such that they satisfy the restricted isometry property. Furthermore, the randomness in the measurement vectors is used to construct approximate dual certificates based on concentration inequalities. In the STFT setup, testing whether the given measurement vectors satisfy the restricted isometry property is difficult. Also, due to the lack of randomness in the measurement vectors, a different approach is required to construct dual certificates.

In the following, we develop a proof technique for the STFT setup, and use it to prove Conjecture IV.1, with additional assumptions.

*Theorem IV.1:* The convex program (4) has a unique feasible matrix  $\mathbf{X}_0 = \mathbf{x}_0 \mathbf{x}_0^*$ , for almost all non-vanishing signals  $\mathbf{x}_0$ , if

- (i)  $\mathbf{w}$  is non-vanishing
- (ii)  $2L \leq W \leq \frac{N}{2}$
- (iii)  $4L \leq M \leq N$
- (iv)  $x_0[n]$  for  $0 \leq n \leq \lfloor \frac{L}{2} \rfloor$  is known a priori.

*Proof:* See Section X. ■

While it is sufficient to show that (4) has a unique solution  $\mathbf{X}_0 = \mathbf{x}_0 \mathbf{x}_0^*$ , observe that Theorem IV.1. ensures that (4) has a unique feasible matrix. This is a stronger condition, and as a consequence, the choice of the objective function does not matter in the noiseless setting. While this might suggest that the requirements of the setup are strong, we argue that it is not the case. In fact, this phenomenon is also observed in generalized phase retrieval (Section 1.3 in [42]) and phase retrieval using random masks (Theorem 1.1 in [28]).

Theorem IV.1 assumes prior knowledge of the first  $\lceil \frac{L}{2} \rceil$  samples, i.e., half of the second short-time section is required to be known a priori. This is not a lot of prior information if  $W \ll N$ , which is typically the case. When  $W = o(N)$ , the fraction of the signal that is required to be known a priori is less than  $\frac{W}{N}$ , which tends to 0 as  $N \rightarrow \infty$ .

*Theorem IV.2:* The convex program (4) has a unique feasible matrix  $\mathbf{X}_0 = \mathbf{x}_0 \mathbf{x}_0^*$ , for almost all non-vanishing signals  $\mathbf{x}_0$ , if

- (i)  $\mathbf{w}$  is non-vanishing
- (ii)  $2 \leq W \leq \frac{N}{2}$
- (iii)  $4 \leq M \leq N$
- (iv)  $L = 1$ .

*Proof:* This is a direct consequence of Theorem IV.1. The value of  $|x_0[0]|$  (and hence  $x_0[0]$ , without loss of generality) can be inferred from the STFT magnitude if  $L = 1$ . ■

When  $L = 1$ , the number of phaseless measurements is  $4(N + W)$ , which is again order-wise optimal. For example, when  $W = 2$ , at most  $4N + 8$  phaseless measurements are considered. Unlike Theorem IV.1, no prior information is necessary.

Theorems IV.1 and IV.2 can be seamlessly extended to incorporate sparse signals:

*Corollary IV.1:* The convex program (4) has a unique feasible matrix  $\mathbf{X}_0 = \mathbf{x}_0 \mathbf{x}_0^*$ , for almost all sparse signals  $\mathbf{x}_0$  which have at most  $W - 2L$  consecutive zeros, if

- (i)  $\mathbf{w}$  is non-vanishing
- (ii)  $2L \leq W \leq \frac{N}{2}$
- (iii)  $4L \leq M \leq N$
- (iv) Either  $L = 1$  or  $x_0[n]$  for  $i_0 \leq n < i_0 + L$  is known a priori, where  $i_0$  is the smallest index such that  $x_0[i_0] \neq 0$ .

*Proof:* See Section X. ■

## B. Noisy Setting

In practice, the measurements are contaminated by additive noise, i.e., the measurements are of the form

$$Z_w[m, r] = \|\langle \mathbf{f}_m, \mathbf{W}_r \mathbf{x} \rangle\|^2 + z[m, r]$$

for  $1 \leq m \leq M$  and  $0 \leq r \leq R - 1$ , where  $\mathbf{z}_r = (z[0, r], z[1, r], \dots, z[M - 1, r])^T$  is the additive noise corresponding to the  $r$ th short-time section and  $4L \leq M \leq N$ . STliFT, in the noisy setting, can be implemented as follows: Suppose  $\|\mathbf{z}_r\|_2 \leq \eta$  for all  $0 \leq r \leq R - 1$ . The constraints in the convex program (4) can be replaced by

$$\sum_{m=1}^M (Z_w[m, r] - \text{trace}(\mathbf{W}_r^* \mathbf{f}_m \mathbf{f}_m^* \mathbf{W}_r \mathbf{X}))^2 \leq \eta^2 \quad (5)$$

for  $0 \leq r \leq R - 1$ . We recommend the use of trace minimization as the objective function. Numerical simulations strongly suggest that STliFT can recover most non-vanishing signals stably in the noisy setting under certain conditions. The details of the simulations are provided in the following section.

## V. NUMERICAL SIMULATIONS

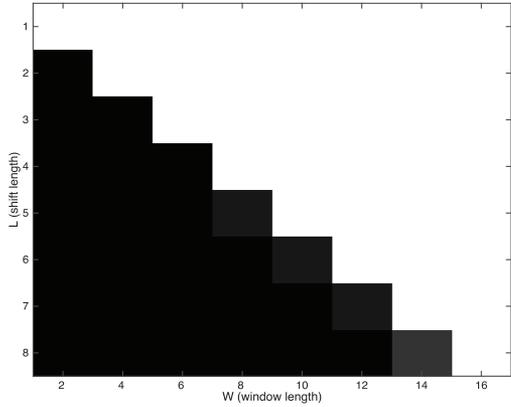
In this section, we demonstrate the empirical abilities of STliFT using numerical simulations.

In the first set of simulations, we evaluate the performance of STliFT as a function of window and shift lengths. We choose  $N = 32$ , and vary  $\{L, W\}$ . For each choice of  $\{L, W\}$ , we consider  $M = 4L$  phaseless measurements and perform 100 trials. In every trial, we choose a random signal such that the values in each location are drawn from an i.i.d. standard complex normal distribution. We select the window  $\mathbf{w}$  such that  $w[n] = 1$  for all  $0 \leq n \leq W - 1$ . The probability of successful recovery as a function of  $\{L, W\}$  is plotted in Fig. 2a.

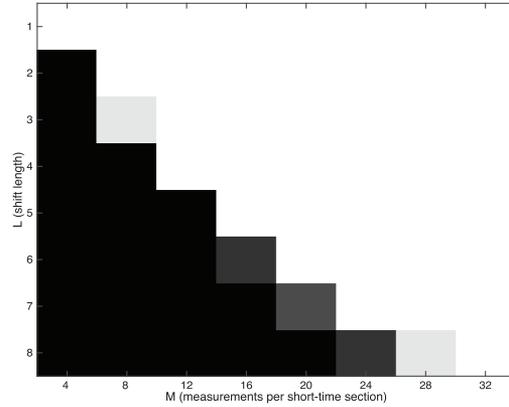
Observe that STliFT successfully recovers the underlying signal with very high probability when  $2L \leq W \leq \frac{N}{2}$  and fails with very high probability when  $2L > W$ . The choice of  $\{L, W\} = \{\frac{N}{4}, \frac{N}{2}\}$  uses only six short-time sections and STliFT recovers the underlying signal with very high probability, which, given the limited success of semidefinite relaxation-based algorithms in the Fourier phase retrieval setup, is very encouraging.

In the second set of simulations, we evaluate the performance of STliFT as a function of shift length and measurements per short-time section. We choose  $N = 32$  and  $W = 16$ , and vary  $\{L, M\}$ . For each choice of  $\{L, M\}$ , we perform 100 trials as before. The probability of successful recovery as a function of  $\{L, M\}$  is plotted in Fig. 2b.

In the third set of simulations, we evaluate the performance of STliFT in the noisy setting. We choose  $M = 2W$ , the rest of



(a)  $N = 32$ ,  $M = 4L$ , and various choices of  $\{L, W\}$ .



(b)  $N = 32$ ,  $W = 16$ , and various choices of  $\{L, M\}$  (demonstrates super-resolution).

Fig. 2. Probability of successful recovery using STliFT in the noiseless setting (white region: success with probability 1, black region: success with probability 0).

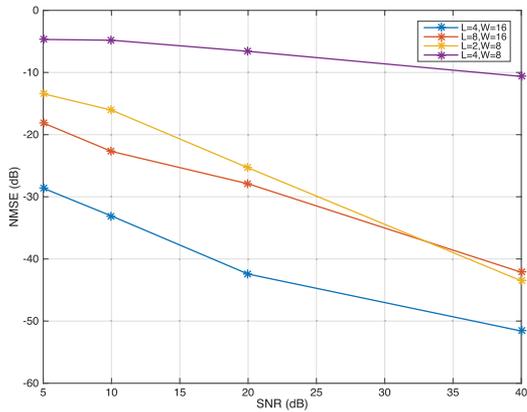


Fig. 3. NMSE (dB) vs SNR (dB) using STliFT in the noisy setting for  $N = 32$ ,  $M = 2W$ .

the parameters are the same as the first set of simulations. The normalized mean-squared error, given by

$$NMSE = \min_{|c|=1} \frac{\|\mathbf{x}_0 - c\hat{\mathbf{x}}\|^2}{\|\mathbf{x}_0\|^2}, \quad (6)$$

is plotted as a function of SNR in Fig. 3. The linear relationship between them shows that STliFT stably recovers the underlying signal in the presence of noise. Also, it can be observed that the choices of  $\{W, L\}$  which correspond to significant overlap between adjacent short-time sections tend to recover signals more stably compared to values of  $\{W, L\}$  which correspond to less overlap, which is not surprising.

### VI. CONCLUSIONS AND FUTURE DIRECTIONS

In this work, we considered the STFT phase retrieval problem. We showed that, if  $L < W \leq \frac{N}{2}$ , then almost all non-vanishing signals can be uniquely identified from their STFT magnitude (up to a global phase), and extended this result to incorporate sparse signals which have less than  $\min\{W - L, L\}$  consecutive zeros.

For  $2L \leq W \leq \frac{N}{2}$ , we conjectured that most non-vanishing signals can be recovered (up to a global phase) by a semidefinite relaxation-based algorithm (STliFT). When  $W = o(N)$ , through super-resolution, we reduced the number of phaseless measurements to  $(4 + o(1))N$ . We proved this conjecture for the setup in which the first  $\lfloor \frac{L}{2} + 1 \rfloor$  samples are known, and for the case in which  $L = 1$ . We argued that the additional assumptions are asymptotically reasonable when  $W \ll N$ , which is typically the case in practical methods. We then extended these results to incorporate sparse signals which have at most  $W - 2L$  consecutive zeros.

Natural directions for future study include a proof of this conjecture without the additional assumptions, and a stability analysis in the noisy setting. Also, a thorough analysis of the phase transition at  $2L = W$  will provide a more complete characterization of STliFT.

### APPENDIX

#### VII. NUMBER OF PHASELESS MEASUREMENTS PER SHORT-TIME SECTION

Throughout the paper, we use an  $N$  point DFT and  $W$  satisfies  $W \leq \frac{N}{2}$ . In this setup, STFT phase retrieval can be equivalently stated in terms of the short-time autocorrelation  $\mathbf{a}_w$ :

$$\begin{aligned} &\text{find} && \mathbf{x} && (7) \\ &\text{subject to} \end{aligned}$$

$$a_w[m, r] = \sum_{n=0}^{N-1-m} x[n]w[rL-n]x^*[n+m]w^*[rL-(n+m)]$$

for  $0 \leq m \leq N - 1$  and  $0 \leq r \leq R - 1$ .

The knowledge of the short-time autocorrelation is sufficient for all the guarantees provided in this paper. Note that the  $r$ th column of  $\mathbf{Z}_w$  and the  $r$ th column of  $\mathbf{a}_w$  are Fourier pairs. Hence, for a particular  $r$ , if  $Z_w[m, r]$  for  $0 \leq m \leq N - 1$  is available, then  $a_w[m, r]$  for  $0 \leq m \leq N - 1$  can be calculated by taking an inverse Fourier transform.

*Lemma VII.1:*  $Z_w[m, r]$  for  $0 \leq m \leq 2W - 2$  is sufficient to calculate  $a_w[m, r]$  for  $0 \leq m \leq N - 1$ .

*Proof:* If the window length is  $W$ , then  $\mathbf{a}_w$  has non-zero values only in the interval  $0 \leq m \leq W - 1$  and  $N - W + 1 \leq m \leq N - 1$ . Let  $\mathbf{b}_w$  be the signal obtained by circularly shifting  $\mathbf{a}_w$  by  $W - 1$  rows, so that  $\mathbf{b}_w$  has non-zero values only in the interval  $0 \leq m \leq 2W - 2$ . Since the submatrix of the  $N$  point DFT matrix obtained by considering the first  $2W - 1$  rows and columns is invertible (the Vandermonde structure is retained),  $Z_w[m, r]$  for  $0 \leq m \leq 2W - 2$  and  $b_w[m, r]$  for  $0 \leq m \leq 2W - 2$  are related by an invertible matrix. Note that  $a_w[m, r]$  for  $0 \leq m \leq N - 1$  can be trivially calculated from  $b_w[m, r]$  for  $0 \leq m \leq 2W - 2$ . ■

Hence, for each short-time section,  $2W$  phaseless measurements are sufficient. Consequently, if the  $N$  point DFT is used and  $2W \leq M \leq N$  is satisfied, the affine constraints in (4) can be rewritten in terms of  $\mathbf{a}_w$  and  $\mathbf{X}$  as:

$$a_w[m, r] = \sum_{n=0}^{N-1-m} X[n, n+m] w[rL - n] w[rL - (n+m)].$$

### VIII. PROOF OF THEOREM III.1

The symbol  $\equiv$  is used to denote equality up to a global phase and time-shift<sup>1</sup>. We say that two signals  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are distinct if  $\mathbf{x}_1 \not\equiv \mathbf{x}_2$ , and equivalent if  $\mathbf{x}_1 \equiv \mathbf{x}_2$ .

Let  $\mathcal{P}$  denote the set of all distinct non-vanishing complex signals of length  $N$ .  $\mathcal{P}$  is a manifold of dimension  $2N - 1$ , i.e.,  $\mathcal{P}$  locally resembles a real  $2N - 1$  dimensional space. This can be seen as follows: In order to discard the global phase of non-vanishing signals, we can assume that  $x[n_0]$  is real and positive at one index  $n_0$ , without loss of generality. Hence,  $x[n_0]$  can take any value in  $\mathbb{R}_+$ , and  $x[n]$ , for each  $0 \leq n \leq N - 1$  not equal to  $n_0$ , can take any value in  $\mathbb{R}^2 \setminus \{0, 0\}$ , due to the one-to-one correspondence between  $\mathbb{C}$  and  $\mathbb{R}^2$ .

Let  $\mathcal{P}_c \subset \mathcal{P}$  be the set of distinct non-vanishing complex signals which cannot be uniquely identified from their STFT magnitude if  $\mathbf{w}$  is chosen such that it is non-vanishing and  $W \leq \frac{N}{2}$ . We show that  $\mathcal{P}_c$  has measure zero in  $\mathcal{P}$ . In order to do so, our strategy is as follows:

We first characterize  $\mathcal{P}_c$  using Lemma VIII.1.. In particular, we show that  $\mathcal{P}_c$  is a finite union of images of continuously differentiable maps from  $\mathbb{R}^{2N-2}$  to  $\mathcal{P}$ . Since  $\mathcal{P}$  is a manifold of dimension  $2N - 1$ , the following result completes the proof:

*Theorem VIII.1 ([16], Chapter 5):* If  $f: \mathbb{R}^{N_0} \rightarrow \mathbb{R}^{N_1}$  is a continuously differentiable map, then the image of  $f$  has measure zero in  $\mathbb{R}^{N_1}$ , provided  $N_0 < N_1$ .

We use the following notation in this section: If  $\mathbf{g}$  is a signal of length  $l_g$ , then  $\mathbf{g} = (g[0], g[1], \dots, g[l_g - 1])^T$  such that  $\{g[0], g[l_g - 1]\} \neq 0$  and  $g[n] = 0$  outside the interval  $[0, l_g - 1]$ . The vector  $\tilde{\mathbf{g}}$  denotes the conjugate-flipped version of  $\mathbf{g}$ , i.e.,  $\tilde{\mathbf{g}} = (g[l_g - 1]^*, g[l_g - 2]^*, \dots, g[0]^*)^T$ . Let  $u_r$  and  $v_r$  denote the smallest and largest index where the windowed signal  $\mathbf{x} \circ \mathbf{w}_r$  has a non-zero value respectively.

*Lemma VIII.1:* Consider two signals  $\mathbf{x}_1 \not\equiv \mathbf{x}_2$  of length  $N$  which have the same STFT magnitude. If the window  $\mathbf{w}$  is chosen such that it is non-vanishing and  $W \leq \frac{N}{2}$ , then, for

<sup>1</sup>For non-vanishing signals, there is no ambiguity due to time-shift.

each  $r$ , there exists signals  $\mathbf{g}_r$  and  $\mathbf{h}_r$ , of lengths  $l_{g_r}$  and  $l_{h_r}$  respectively, such that

- (i)  $\mathbf{x}_1 \circ \mathbf{w}_r \equiv \mathbf{g}_r \star \mathbf{h}_r$ ,  $\mathbf{x}_2 \circ \mathbf{w}_r \equiv \mathbf{g}_r \star \tilde{\mathbf{h}}_r$
- (ii)  $l_{g_r} + l_{h_r} - 1 = v_r - u_r + 1$
- (iii)  $g_r[l_{g_r} - 1] = 1$  and  $\{g_r[0], h_r[0], h_r[l_{h_r} - 1]\} \neq 0$  where  $\star$  is the convolution operator. Further, there exists at least one  $r$  such that
- (iv)  $l_{h_r} \geq 2$ ,  $h_r[0]$  is real and positive.

*Proof:* In Lemma 7.1 of [22], it is shown that if two non-equivalent signals of length  $N$  have the same Fourier magnitude and if the DFT dimension is at least  $2N$  (this would imply that they have the same autocorrelation), then there exists signals  $\mathbf{g}$  and  $\mathbf{h}$ , of lengths  $l_g$  and  $l_h$  respectively, such that one signal can be decomposed as  $\mathbf{g} \star \mathbf{h}$  and the other signal can be decomposed as  $\mathbf{g} \star \tilde{\mathbf{h}}$ . For each  $r$ , the  $r$ th column of the STFT magnitude is equivalent to the Fourier magnitude of the windowed signal  $\mathbf{x} \circ \mathbf{w}_r$ . The DFT dimension is  $N$ , and the windowed signal length is  $v_r - u_r + 1$  (which is less than or equal to  $\frac{N}{2}$ ). Since, for every  $r$ ,  $\mathbf{x}_1 \circ \mathbf{w}_r$  and  $\mathbf{x}_2 \circ \mathbf{w}_r$  have the same Fourier magnitude, the aforementioned result proves (i).

The conditions (ii) and (iii) are properties of convolution (see Lemma 7.1 of [22] for details), and therefore hold for every  $r$ .

Furthermore, if  $l_{h_r} = 1$  for all  $0 \leq r \leq R - 1$ , then  $\mathbf{x}_1 \equiv \mathbf{x}_2$ . Hence,  $l_{h_r} \geq 2$  for at least one  $r$ . For this  $r$ , since  $e^{i\phi_1} \mathbf{x}_1$  and  $e^{i\phi_2} \mathbf{x}_2$  have the same STFT magnitude,  $h_r[0]$  can be assumed to be real and positive without loss of generality. Hence, (iv) holds for at least one  $r$ . ■

Consequently, for each  $\mathbf{x} \in \mathcal{P}_c$ , condition (iv) of Lemma VIII.1 holds for at least one  $r$ . Let  $\mathcal{P}_c^{r l_r l_{r+1}} \subset \mathcal{P}_c$  denote the set of signals for which  $l_{h_r} = l_r \geq 2$  and  $l_{h_{r+1}} = l_{r+1}$ .<sup>2</sup> It suffices to show that for each  $r$ ,  $l_r$  and  $l_{r+1}$ , there exists a set  $\mathcal{Q}_c^{r l_r l_{r+1}} \supseteq \mathcal{P}_c^{r l_r l_{r+1}}$ , which is the image of a continuously differentiable map from  $\mathbb{R}^{2N-2}$  to  $\mathcal{P}$ .

We first show the arguments for the  $L = W - 1$  case as the expressions are simple and provide intuition for the technique. Then, we show the arguments for the  $L < W - 1$  case.

(i)  $L = W - 1$ :

The set  $\mathcal{Q}_c^{r l_r l_{r+1}}$  is constructed as follows: Consider the variables  $\{\mathbf{g}_r, \mathbf{h}_r, \mathbf{g}_{r+1}, \mathbf{h}_{r+1}\}$  satisfying  $l_{h_r} = l_r \geq 2$  and  $l_{h_{r+1}} = l_{r+1}$ , and  $x[n]$  for  $n \in [0, u_r] \cup (v_{r+1}, N - 1]$ . The map  $f = (f_0, f_1, \dots, f_{N-1})^T$  from these variables to  $\mathcal{P}$  is the following:

$$f_n = \begin{cases} x[n] & \text{for } n \in [0, u_r] \cup (v_{r+1}, N - 1] \\ \sum_{m=0}^{n-u_r} g_r[m] h_r[n - u_r - m] & \text{for } n \in [u_r, v_r] \\ \sum_{m=0}^{n-u_{r+1}} g_{r+1}[m] h_{r+1}[n - u_{r+1} - m] & \text{for } n \in [u_{r+1}, v_{r+1}]. \end{cases} \quad (8)$$

Observe that, for  $n = u_{r+1} = v_r$ ,  $f_n$  has two definitions. The variables can admit only those values for which the two definitions have the same value. In the following, we show that there is a one-to-one correspondence between the set of admissible values of the variables and a subset of  $\mathbb{R}^{2N-2}$ .

<sup>2</sup>When  $r = R$ , we consider the short-time section  $r - 1$  instead of  $r + 1$ . We show the detailed calculations for the case when short-time section  $r + 1$  is considered, the arguments are symmetric for  $r - 1$ .

Each  $x[n]$ , for  $n \in [0, u_r] \cup (v_{r+1}, N - 1]$ , can be chosen from  $\mathbb{C} \mathbb{R}^2$ . The set of  $\{\mathbf{g}_{r+1}, \mathbf{h}_{r+1}\}$  is a subset of  $\mathbb{R}^{2(v_{r+1}-u_{r+1}+1)}$ , which can be seen as follows:  $g_{r+1}[l_{g,r+1} - 1] = 1$  is fixed (see Lemma VIII.1), there are  $v_{r+1} - u_{r+1} + 1$  other terms and each can be chosen from  $\mathbb{C} \mathbb{R}^2$ .

For each choice of  $\{\mathbf{g}_{r+1}, \mathbf{h}_{r+1}\}$ , consider the set of  $\{\mathbf{g}_r, \mathbf{h}_r\}$  excluding the terms  $h_r[0]$  and  $h_r[l_{hr} - 1]$ :  $g_r[l_{gr} - 1] = 1$  is fixed, there are  $v_r - u_r - 1$  other terms and each can be chosen from  $\mathbb{C} \mathbb{R}^2$ . Hence, this set is a subset of  $\mathbb{R}^{2(v_r-u_r-1)}$ .

Since the short-time sections  $r$  and  $r + 1$  overlap in the index  $v_r$ ,  $\mathbf{g}_r \star \mathbf{h}_r$  and  $\mathbf{g}_{r+1} \star \mathbf{h}_{r+1}$  must be consistent in this index, i.e.,  $\{\mathbf{g}_r, \mathbf{h}_r\}$  must satisfy:

$$\frac{1}{w[0]} h_r[l_{hr} - 1] = \frac{1}{w[W - 1]} g_{r+1}[0] h_{r+1}[0]. \quad (9)$$

Due to Lemma VIII.1,  $\mathbf{g}_r \star \tilde{\mathbf{h}}_r$  and  $\mathbf{g}_{r+1} \star \tilde{\mathbf{h}}_{r+1}$  must also be consistent in this index up to a phase, i.e.,  $\{\mathbf{g}_r, \mathbf{h}_r\}$  must also satisfy:

$$h_r[0] \equiv \frac{w[0]}{w[W - 1]} g_{r+1}[0] h_{r+1}^*[l_{h,r+1} - 1]. \quad (10)$$

Observe that  $\equiv$  is used in (10), due to the fact that the equality is only up to a phase. However,  $h_r[0]$  is real and positive (see Lemma VIII.1), due to which (10) fixes  $h_r[0]$ .

Consequently, the set of admissible values of the variables, excluding  $h_r[0]$  and  $h_r[l_{hr} - 1]$ , is a subset of  $\mathbb{R}^{2N-2}$ , as  $2(N - v_{r+1} + u_r - 1 + v_{r+1} - u_{r+1} + 1 + v_r - u_r - 1) = 2N - 2$ . For each point in this set,  $h_r[0]$  and  $h_r[l_{hr} - 1]$  are uniquely determined. It is straightforward to check that the map  $f$  from this set to  $\mathcal{P}$  is continuously differentiable. Consequently,  $\mathcal{Q}_c^{r,l_{hr}}$  is the image of a continuously differentiable map from  $\mathbb{R}^{2N-2}$  to  $\mathcal{P}$ .

(ii)  $L < W - 1$ :

Consider the setup for which  $2L \geq W$ . The set of  $\{\mathbf{g}_{r+1}, \mathbf{h}_{r+1}\}$ , as earlier, is a subset of  $\mathbb{R}^{2(v_{r+1}-u_{r+1}+1)}$ .

The short-time sections  $r$  and  $r + 1$  overlap in the interval  $[u_{r+1}, v_r]$ . Let  $v_r - u_{r+1} + 1 = T$  (the number of indices in the overlapping interval). Due to  $2L \geq W$ , we have  $T = W - L \leq \lfloor \frac{W}{2} \rfloor$ . Hence, for each choice of  $\{\mathbf{g}_{r+1}, \mathbf{h}_{r+1}\}$ ,  $\{\mathbf{g}_r, \mathbf{h}_r\}$  must satisfy:

$$\begin{aligned} & \sum_{m=0}^{n+u_{r+1}-u_r} \frac{1}{w_r[u_r + m]} g_r[m] h_r[n + u_{r+1} - u_r - m] \\ &= \sum_{m=0}^n \frac{1}{w_{r+1}[u_{r+1} + m]} g_{r+1}[m] h_{r+1}[n - m] \end{aligned} \quad (11)$$

for  $0 \leq n \leq T - 1$ . In addition,  $\{\mathbf{g}_r, \mathbf{h}_r\}$  must also satisfy:

$$\frac{1}{w[0]} h_r[0] \equiv \sum_{m=0}^{T-1} \frac{1}{w_{r+1}[u_{r+1} + m]} g_{r+1}[m] h_{r+1}[T - 1 - m] \quad (12)$$

If  $l_{hr} \geq \lfloor \frac{W}{2} \rfloor + 1$ , then the  $T$  bilinear equations (11) can be written as  $\mathbf{G} \mathbf{h}_r = \mathbf{c}$ , where  $\mathbf{G}$  has upper triangular structure with unit diagonal entries, due to which  $\text{rank}(\mathbf{G}) = T$ . The set of  $\mathbf{g}_r$  is a subset of  $\mathbb{R}^{2(l_{gr}-1)}$ . For each choice of  $\mathbf{g}_r$ , the

terms  $\{h_r[l_{hr} - T], \dots, h_r[l_{hr} - 1]\}$  are fixed by  $\mathbf{G} \mathbf{h}_r = \mathbf{c}$ . The constraint (12) fixes the value of  $h_r[0]$ , as earlier. Each of the remaining  $(l_{hr} - 1 - T)$  terms of  $\mathbf{h}_r$  may be chosen from  $\mathbb{C} \mathbb{R}^2$ .

Hence, the set of admissible values of the variables, excluding  $\{h_r[l_{hr} - T], h_r[l_{hr} - T + 1], \dots, h_r[l_{hr} - 1]\}$  and  $h_r[0]$ , is a subset of  $\mathbb{R}^{2N-2}$ , due to the fact that  $2(N - v_{r+1} + u_r - 1 + v_{r+1} - u_{r+1} + 1 + v_r - u_r - T) = 2N - 2$  (as  $l_{gr} + l_{hr} - 1 = v_r - u_r + 1$ ). For each point in this set,  $h_r[0]$  and  $\{h_r[l_{hr} - T], \dots, h_r[l_{hr} - 1]\}$  are uniquely determined. The rest of the arguments are identical to those of  $L = W - 1$ .

If  $l_{gr} \geq \lfloor \frac{W}{2} \rfloor + 1$  instead, then the bilinear equations (11) can be equivalently written as  $\mathbf{H} \mathbf{g}_r = \mathbf{c}$ , the same arguments may be applied to draw the same conclusion. For the setup with  $2L > W$ , the same arguments hold for the short-time sections  $r$  and  $r + t$ , where  $t$  is the largest integer such that the short-time sections  $r$  and  $r + t$  overlap (this ensures  $T \leq \lfloor \frac{W}{2} \rfloor$ ).

### IX. Proof of Corollary III.1

We now extend Theorem III.1 to incorporate sparse signals. Let  $\mathcal{P}^S$  denote the set of all distinct complex signals of length  $N$  with a support  $S$ . Here,  $S$  is a binary vector of length  $N$ , such that  $x[n] \neq 0$  whenever  $S[n] = 1$  and  $x[n] = 0$  whenever  $S[n] = 0$ . Further,  $S$  has less than  $\min\{L, W - L\}$  consecutive zeros.

Let  $\mathcal{P}_c^S \subset \mathcal{P}^S$  denote the set of signals which cannot be uniquely identified from their STFT magnitude if  $\mathbf{w}$  is chosen such that it is non-vanishing and  $W \leq \frac{N}{2}$ . We show that  $\mathcal{P}_c^S$  has measure zero in  $\mathcal{P}^S$ .

In the proof of Theorem III.1, in order to show dimension reduction, we used the fact that for sufficient pairs of adjacent short-time sections  $r$  and  $r + 1$ , the following holds:

- (i) There is at least one index in the non-overlapping indices  $[u_r, u_{r+1} - 1]$  or  $[v_r + 1, v_{r+1}]$  where the signals  $\mathbf{x}_1$  and  $\mathbf{x}_2$  have a non-zero value. This ensures that  $h_r[0]$  is not constrained by  $\{\mathbf{g}_{r+1}, \mathbf{h}_{r+1}\}$  in general. This condition can be ensured by imposing the constraint that the sparse signal cannot have  $L$  consecutive zeros.
- (ii) There is at least one index in the overlapping indices  $[u_{r+1}, v_r]$  where the signals  $\mathbf{x}_1$  and  $\mathbf{x}_2$  have a non-zero value. This ensures that  $h_r[0]$  is constrained by  $\{\mathbf{g}_{r+1}, \mathbf{h}_{r+1}\}$  (10) for signals which cannot be uniquely identified by their STFT magnitude. This condition can be ensured by imposing the constraint that the sparse signal cannot have  $W - L$  consecutive zeros.

The only difference in the proof is the following: Unlike in the case of non-vanishing signals, there is time-shift ambiguity. Hence, the constraint (12) is replaced by:

$$\frac{1}{w[0]} h_r[0] \equiv \sum_{m=0}^n \frac{1}{w_{r+1}[u_{r+1} + m]} g_{r+1}[m] h_{r+1}[n - m] \quad (13)$$

for some  $0 \leq n \leq T - 1$ . This fixes the value of  $h_r[0]$  to one of at most  $T$  values, due to which there is a dimension reduction.

### X. Proof of Theorem IV.1

We first show the arguments for the case  $2W \leq M \leq N$  as the expressions are simple and provide intuition. Then, we show the arguments for  $4L \leq M < 2W$ .

(i)  $2W \leq M \leq N$  :

The affine constraints in (4) can be rewritten as (see Section VII):

$$a_w[m, r] = \sum_{n=0}^{N-1-m} X[n, n+m]w[rL-n]w[rL-(n+m)]$$

The proof strategy is as follows: We begin by focusing our attention on short-time section  $r = 1$ . We show that the prior information available, along with the affine autocorrelation measurements corresponding to  $r = 1$  and the positive semidefinite constraint, will ensure that every feasible matrix of (4) satisfies  $X[n, m] = x_0[n]x_0^*[m]$  for  $0 \leq n, m \leq L$ . We then apply this argument incrementally, i.e., we show that the affine measurements corresponding to short-time section  $r$ , along with the entries of  $\mathbf{X}$  uniquely determined and the positive semidefinite constraint, will ensure that  $X[n, m] = x_0[n]x_0^*[m]$  for  $u_r \leq n, m \leq v_r$ , where  $u_r$  and  $v_r$  denote the smallest and largest index where  $\mathbf{w}_r$  has a non-zero value respectively. Consequently, the entries along the diagonal and the first  $W - L$  off-diagonals of every feasible matrix of (4) match the entries along the diagonal and the first  $W - L$  off-diagonals of the matrix  $\mathbf{x}_0\mathbf{x}_0^*$ . Since the entries are sampled from a rank one matrix with non-zero diagonal entries (i.e.,  $\mathbf{x}_0\mathbf{x}_0^*$ ), there is exactly one positive semidefinite completion, which is the rank one completion  $\mathbf{x}_0\mathbf{x}_0^*$  [50].

Let  $\mathbf{s}_0 = (x_0[0], x_0[1], \dots, x_0[L])^T$  be a length  $L + 1$  sub-signal of  $\mathbf{x}_0$ , and  $\mathbf{S}$  be the  $(L + 1) \times (L + 1)$  submatrix of  $\mathbf{X}$  corresponding to the first  $L + 1$  rows and columns. We now show that  $\mathbf{S} = \mathbf{s}_0\mathbf{s}_0^*$  is the only feasible matrix under the constraints of (4).

Since  $x_0[n]$  for  $0 \leq n \leq \lfloor \frac{L}{2} \rfloor$  is known a priori, we have  $S[n, m] = x_0[n]x_0^*[m]$  for  $0 \leq n, m \leq \lfloor \frac{L}{2} \rfloor$ . Let  $\mathcal{A}(\mathbf{S}) = \mathbf{c}$  denote these constraints due to prior information, along with the affine constraints corresponding to  $r = 1$ . In particular,  $\mathcal{A}(\mathbf{S}) = \mathbf{c}$  denotes the following set of constraints:

$$S[n, m] = x_0[n]x_0^*[m] \quad \text{for } 0 \leq n, m \leq \lfloor \frac{L}{2} \rfloor,$$

$$a_w[m, 1] = \sum_{n=0}^{L-m} S[n, n+m]w[L-n]w^*[L-(n+m)].$$

For each feasible matrix  $\mathbf{S}$ , these set of measurements fix (i) the  $\lfloor \frac{L}{2} + 1 \rfloor \times \lfloor \frac{L}{2} + 1 \rfloor$  submatrix, corresponding to the first  $\lfloor \frac{L}{2} + 1 \rfloor$  rows and columns, of  $\mathbf{S}$  (ii) the appropriately weighted sum along the diagonal and each off-diagonal of  $\mathbf{S}$  ( $2L \leq W$  is implicitly used here).

*Lemma X.1:* If  $\mathbf{S}_0 = \mathbf{s}_0\mathbf{s}_0^*$  satisfies  $\mathcal{A}(\mathbf{S}) = \mathbf{c}$ , then it is the only positive semidefinite matrix which satisfies  $\mathcal{A}(\mathbf{S}) = \mathbf{c}$ .

*Proof:* Let  $T$  be the set of Hermitian matrices of the form

$$T = \{\mathbf{S} = \mathbf{s}_0\mathbf{v}^* + \mathbf{v}\mathbf{s}_0^* : \mathbf{v} \in \mathbb{C}^n\}$$

and  $T^\perp$  be its orthogonal complement. The set  $T$  may be interpreted as the tangent space at  $\mathbf{s}_0\mathbf{s}_0^*$  to the manifold of Hermitian

matrices of rank one. Influenced by [42], we use  $\mathbf{S}_T$  and  $\mathbf{S}_{T^\perp}$  to denote the projection of a matrix  $\mathbf{S}$  onto the subspaces  $T$  and  $T^\perp$  respectively.

Standard duality arguments in semidefinite programming show that the following are sufficient conditions for  $\mathbf{S}_0 = \mathbf{s}_0\mathbf{s}_0^*$  to be the unique optimizer of (4):

(i) *Condition 1:*  $\mathbf{S} \in T$  and  $\mathcal{A}(\mathbf{S}) = 0 \Rightarrow \mathbf{S} = 0$ .

(ii) *Condition 2:* There exists a dual certificate  $\mathbf{D}$  in the range space of  $\mathcal{A}^*$  obeying:  $\mathbf{D}\mathbf{s}_0 = 0$   
 $\text{rank}(\mathbf{D}) = L$   
 $\mathbf{D} \succ 0$ .

The proof of this result is based on KKT conditions, and can be found in any standard reference on semidefinite programming (for example, see [51]).

We first show that *Condition 1* is satisfied. The set of constraints in  $\mathcal{A}(\mathbf{S}) = 0$  due to prior information fix the entries of the first  $\lfloor \frac{L}{2} + 1 \rfloor$  rows and columns of  $\mathbf{S}$  to 0. Since  $\mathbf{S} = \mathbf{s}_0\mathbf{v}^* + \mathbf{v}\mathbf{s}_0^*$  for some  $\mathbf{v} = (v[0], v[1], \dots, v[L])^T$  (due to  $\mathbf{S} \in T$ ), we infer that  $v[n] = icx_0[n]$  for  $0 \leq n \leq \lfloor \frac{L}{2} \rfloor$ , for some real constant  $c$ . Indeed, the equations of the form  $s_0[n]v[n]^* + v[n]s_0^*[n] = 0$  imply  $v[n] = ic_nx_0[n]$ , for some real constant  $c_n$ . The equations  $s_0[n]v^*[m] + v[n]s_0^*[m] = 0$  imply  $c_n = c_m$ .

The set of constraints in  $\mathcal{A}(\mathbf{S}) = 0$  due to the measurements corresponding to  $r = 1$ , along with  $v[n] = icx_0[n]$  for  $0 \leq n \leq \lfloor \frac{L}{2} \rfloor$ , imply  $v[n] = icx_0[n]$  for  $\lfloor \frac{L}{2} + 1 \rfloor \leq n \leq L$ . Hence, for  $\mathbf{S} \in T$ ,  $\mathcal{A}(\mathbf{S}) = 0$  implies  $\mathbf{v} = ic\mathbf{s}_0$ , which in turn implies  $\mathbf{S} = -ics_0\mathbf{s}_0^* + ics_0\mathbf{s}_0^* = 0$ .

We next establish *Condition 2*. For simplicity of notation, we consider the case where  $w[n] = 1$  for  $0 \leq n \leq W - 1$ . For a general non-vanishing  $\mathbf{w}$ , the same arguments hold (the Toeplitz matrix considered is appropriately redefined with weights).

The range space of  $\mathcal{A}^*$  is the set of all  $L + 1 \times L + 1$  matrices which are a sum of the following two matrices: The first matrix can have any value in the  $\lfloor \frac{L}{2} + 1 \rfloor \times \lfloor \frac{L}{2} + 1 \rfloor$  submatrix corresponding to the first  $\lfloor \frac{L}{2} + 1 \rfloor$  rows and columns, and has a value zero outside this submatrix (dual of the set of constraints due to prior information). The second matrix has a Toeplitz structure (dual of the measurements corresponding to  $r = 1$ ).

Suppose  $\mathbf{s}_1$  is the vector containing the first  $\lfloor \frac{L}{2} + 1 \rfloor$  entries of  $\mathbf{s}_0$  and  $\mathbf{s}_2$  is the vector containing the remaining entries of  $\mathbf{s}_0$ . Here,  $\mathbf{s}_1$  corresponds to the locations where we have knowledge of the entries and  $\mathbf{s}_2$  corresponds to the locations where the entries are not determined. Let  $\mathbf{L}$  be a lower triangular  $\lceil \frac{L}{2} \rceil \times \lfloor \frac{L}{2} + 1 \rfloor$  Toeplitz matrix satisfying  $\mathbf{L}\mathbf{s}_1 + \mathbf{s}_2 = 0$ . Such an  $\mathbf{L}$  always exists if  $s_1[0]$  is non-zero and the length of  $\mathbf{s}_1$  is greater than or equal to the length of  $\mathbf{s}_2$ . Let  $\Lambda$  be any  $\lfloor \frac{L}{2} + 1 \rfloor \times \lfloor \frac{L}{2} + 1 \rfloor$  positive semidefinite matrix with rank  $\lfloor \frac{L}{2} \rfloor$  satisfying  $\Lambda\mathbf{s}_1 = 0$ . Again, such a  $\Lambda$  always exists (any positive semidefinite matrix with eigenvectors perpendicular to  $\mathbf{s}_1$ ). Consider the following dual certificate:

$$\mathbf{D} = \begin{bmatrix} \mathbf{L}^*\mathbf{L} + \Lambda & \mathbf{L}^* \\ \mathbf{L} & \mathbf{I}_{\lfloor \frac{L}{2} \rfloor} \end{bmatrix}. \quad (14)$$

Clearly,  $\mathbf{D}$  is in the range space of  $\mathcal{A}^*$ . Also,  $\mathbf{D}\mathbf{s}_0 = 0$  by construction. From the Schur complement, it is straightforward to see that  $\text{rank}(\mathbf{D}) = L$  and  $\mathbf{D} \succcurlyeq 0$ . ■

We have shown that  $\mathbf{S}_0 = \mathbf{s}_0\mathbf{s}_0^*$  is the only positive semidefinite matrix which satisfies the prior information and the measurements corresponding to  $r = 1$ . Redefine  $\mathbf{s}_0$  and  $\mathbf{S}$  such that  $\mathbf{s}_0 = (x_0[0], x_0[1], \dots, x_0[2L])^T$  is the  $2L + 1$  length sub-signal of  $\mathbf{x}$  and  $\mathbf{S}$  is the  $(2L + 1) \times (2L + 1)$  submatrix of  $\mathbf{X}$  corresponding to the first  $2L + 1$  rows and columns.

We already have  $S[n, m] = x_0[n]x_0^*[m]$  for  $0 \leq n, m \leq L$  from above. Let  $\mathcal{A}(\mathbf{S}) = \mathbf{c}$  denote these constraints, along with the affine constraints corresponding to  $r = 2$ . Due to  $2L \leq W$ , Lemma X.1 proves that  $\mathbf{S}_0 = \mathbf{s}_0\mathbf{s}_0^*$  is the only psd matrix which satisfies the prior information and the measurements corresponding to  $r = 1, 2$ . Applying this argument incrementally, the entries along the diagonal and the first  $W - L$  off-diagonals of every feasible matrix of (4) match the entries along the diagonal and the first  $W - L$  off-diagonals of the matrix  $\mathbf{x}_0\mathbf{x}_0^*$ .

**Sparse signals:** The arguments can be seamlessly extended to incorporate sparse signals.

(i) The fact that there exists a unique positive semidefinite completion once the diagonal and the first  $W - L$  off-diagonal entries are sampled from  $\mathbf{x}_0\mathbf{x}_0^*$  holds when  $\mathbf{x}_0$  has less than  $W - L$  consecutive zeros.

(ii) Note that the length of  $\mathbf{s}_2$  is at most  $L$ , as it corresponds to the locations in the window where the entries are not determined. Since we know  $x_0[n]$  for  $i_0 \leq n < i_0 + L$  a priori, where  $i_0$  is the smallest index such that  $x_0[i_0] \neq 0$ , the length of  $\mathbf{s}_1$  is  $W - L$ . Redefine  $\mathbf{s}_1$  so that it corresponds to the locations in the window where the entries are determined, starting from the smallest index which has a non-zero value in order to ensure  $\mathbf{s}_1[0] \neq 0$ . If  $\mathbf{x}_0$  has at most  $W - 2L$  consecutive zeros, then the length of  $\mathbf{s}_1$  is at least  $(W - L) - (W - 2L) = L$ . Hence, a lower triangular Toeplitz matrix  $\mathbf{L}$ , satisfying  $\mathbf{L}\mathbf{s}_1 + \mathbf{s}_2 = 0$ , always exists.

(ii)  $4L \leq M < 2W$  (super-resolution case):

Unlike in the previous case, we do not have the knowledge of the autocorrelation of each short-time section. However, by making optimal use of the additional redundancy due to the increased overlap, we show the existence of a dual certificate similar in nature to (14). As earlier, we begin by focusing our attention on short-time section  $r = 1$ .

Let  $\mathbf{l} = (l[0], l[1], \dots, l[L])^T$  be a vector of length  $L + 1$ , such that  $(l[L], l[L - 1], \dots, l[0])$  is the last row in (14). Let  $\hat{\mathbf{l}} = (l[0], l[1], \dots, l[N - 1])^T$  be a vector of length  $N$ , obtained by extending  $\mathbf{l}$ . We show the existence of a vector  $\hat{\mathbf{l}}$ , from which a vector  $\mathbf{l}$  satisfying all the necessary conditions can be obtained.

The range space of the dual certificate is the set of all  $L + 1 \times L + 1$  matrices which are a sum of the following two matrices: The first matrix can have any value in the  $\lfloor \frac{L}{2} + 1 \rfloor \times \lfloor \frac{L}{2} + 1 \rfloor$  submatrix corresponding to the first  $\lfloor \frac{L}{2} + 1 \rfloor$  rows and columns, and has a value zero outside this submatrix (dual of the set of constraints due to prior information). The second matrix has the form  $\sum_{m=1}^M \alpha_m \mathbf{W}_r^* \mathbf{f}_m \mathbf{f}_m^* \mathbf{W}_r$ , where  $\alpha_m$  is real-valued for each  $m$  (dual of the measurements corresponding to  $r = 1$ ). Consider an  $\hat{\mathbf{l}}$  that satisfies the following set of conditions:

- (i)  $l[0] = 1, l[n] = l[N - n] = 0$  for  $1 \leq n \leq \lfloor \frac{L}{2} \rfloor - 1$
- (ii)  $\sum_{n=0}^m x_0[n]l[m - n] = \sum_{n=0}^m x_0[n] \star l[N - m + n] = 0$  for  $\lfloor \frac{L}{2} + 1 \rfloor \leq m \leq L$
- (iii)  $\mathbf{f}_m^* \hat{\mathbf{l}} = 0$  for  $M + 1 \leq m \leq N$ .

These constraints together can be written as  $\mathbf{A}\hat{\mathbf{l}} = \mathbf{b}$ . When  $M \geq 4\lfloor \frac{L}{2} \rfloor$ , the matrix  $\mathbf{A}$  is square or wide, and almost always (pseudo) invertible. This can be seen as follows: the determinant of  $\mathbf{A}$  is a polynomial function of the entries of  $\mathbf{x}_0$ , due to which it is either always zero or almost surely non-zero. By substituting  $x_0[0] = 1$  and  $x_0[n] = 0$  for  $n \neq 0$ , it is straightforward to check that the determinant is non-zero. Hence, such an  $\hat{\mathbf{l}}$  almost always exists.

The vector  $\mathbf{l}$  obtained by choosing the first  $L + 1$  entries of this  $\hat{\mathbf{l}}$  is used to construct the lower left and lower right blocks of the dual certificate: (i) The lower right block is an identity matrix. (ii)  $\mathbf{L}\mathbf{s}_1 + \mathbf{s}_2 = 0$  is satisfied. (iii) Since  $\mathbf{b}$  is a real vector,  $\hat{\mathbf{l}}$  satisfies  $l[n] = l[N - n]^*$ . Therefore,  $\hat{\mathbf{l}}$  is in the range space of  $\sum_{m=1}^M \alpha_m \mathbf{f}_m$  where  $\alpha_m$  is real-valued, due to which the resulting second matrix is in the range space of  $\sum_{m=1}^M \alpha_m \mathbf{W}_r^* \mathbf{f}_m \mathbf{f}_m^* \mathbf{W}_r$ .

The arguments are applied incrementally as earlier (with  $M \geq 4L$ , since  $L \times L$  is the size of the lower right block, for  $r > 1$ ).

#### ACKNOWLEDGMENTS

We would like to thank Mordechai Segev and Oren Cohen for introducing us to the STFT phase retrieval problem, and for many insightful discussions.

#### REFERENCES

- [1] A. L. Patterson, "A Fourier series method for the determination of the components of interatomic distances in crystals," *Phys. Rev.*, vol. 46, no. 5, p. 372, 1934.
- [2] A. L. Patterson, "Ambiguities in the X-ray analysis of crystal structures," *Phys. Rev.*, vol. 65, nos. 5–6, p. 195, 1944.
- [3] A. Walther, "The question of phase retrieval in optics," *J. Modern Opt.*, vol. 10, no. 1, pp. 41–49, 1963.
- [4] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Amer. A*, vol. 7, no. 3, pp. 394–411, 1990.
- [5] J. C. Dainty and J. R. Fienup, "Phase retrieval and image reconstruction for astronomy," in *Image Recovery: Theory and Application*, San Diego, CA, USA: Academic, 1987, pp. 231–275.
- [6] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
- [7] M. Stefik, "Inferring DNA structures from segmentation data," *Artif. Intell.*, vol. 11, no. 1, pp. 85–114, 1978.
- [8] B. Baykal, "Blind channel estimation via combining autocorrelation and blind phase estimation," *IEEE Trans. Circuits Syst.*, vol. 51, no. 6, pp. 1125–1131, Jun. 2004.
- [9] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, p. 237, 1972.
- [10] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, 1982.
- [11] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization," *J. Opt. Soc. Amer. A*, vol. 19, no. 7, pp. 1334–1345, 2002.
- [12] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, May 2015.
- [13] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval: An overview of recent developments," arXiv:1510.07713, 2015.

- [14] E. M. Hofstetter, "Construction of time-limited functions with specified autocorrelation functions," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 2, pp. 119–126, Apr. 1964.
- [15] M. H. Hayes and J. H. McClellan, "Reducible polynomials in more than one variable," *Proc. IEEE*, vol. 70, no. 2, pp. 197–198, Feb. 1982.
- [16] J. M. Ortega and W. C. Rheinboldt, "Iterative solution of nonlinear equations in several variables," *SIAM*, vol. 30, pp. 119–146, 1970.
- [17] Y. M. Lu and M. Vetterli, "Sparse spectral factorization: Unicity and reconstruction algorithms," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2011, pp. 5976–5979.
- [18] A. Szameit *et al.*, "Sparsity-based single-shot subwavelength coherent diffractive imaging," *Nat. Mater.*, vol. 11, pp. 455–459, 2012.
- [19] S. Mukherjee and C. Seelamantula, "An iterative algorithm for phase retrieval with sparsity constraints: Application to frequency domain optical coherence tomography," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2012, pp. 553–556.
- [20] Y. Shechtman, Y. C. Eldar, A. Szameit, and M. Segev, "Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing," *Opt. Exp.*, vol. 19, pp. 14807–14822, 2011.
- [21] K. Jaganathan, S. Oymak, and B. Hassibi, "Recovery of sparse 1-D signals from the magnitudes of their Fourier transform," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 1473–1477.
- [22] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Uniqueness guarantees and recovery algorithms," arXiv:1311.2745, 2015.
- [23] Y. Shechtman, A. Beck, and Y. C. Eldar, "GESPAR: Efficient phase retrieval of sparse signals," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 928–938, Feb. 2014.
- [24] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 199–225, 2013.
- [25] Y. J. Liu *et al.*, "Phase retrieval in X-ray imaging based on using structured illumination," *Phys. Rev. A*, vol. 78, no. 2, p. 023817, 2008.
- [26] I. Johnson *et al.*, "Coherent diffractive imaging using phase front modifications," *Phys. Rev. Lett.*, vol. 100, no. 15, p. 155503, 2008.
- [27] E. G. Loewen and E. Popov, *Diffraction Gratings and Applications*. Boca Raton, FL, USA: CRC Press, 1997.
- [28] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Appl. Comput. Harmonic Anal.*, vol. 39, pp. 277–299, 2014.
- [29] D. Gross, F. Kraemer, and R. Kueng, "Improved recovery guarantees for phase retrieval from coded diffraction patterns," arXiv:1402.6286, 2014.
- [30] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval with masks using convex optimization," in *Proc. IEEE Int. Symp. Inf. Theory*, 2015, pp. 1655–1659.
- [31] R. Trebino, *Frequency-Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses*. New York, NY, USA: Springer, 2002.
- [32] M. J. Humphry, B. Kraus, A. C. Hurst, A. M. Maiden, and J. M. Rodenburg, "Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging," *Nat. Commun.*, vol. 3, p. 730, 2012.
- [33] J. M. Rodenburg, "Ptychography and related diffractive imaging methods," *Adv. Imag. Electron Phys.*, vol. 150, pp. 87–184, 2008.
- [34] G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," *Nat. Photonics*, vol. 7, no. 9, pp. 739–745, 2013.
- [35] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.
- [36] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Recovering signals from the short-time Fourier transform magnitude," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2015, pp. 3277–3281.
- [37] S. H. Nawab, T. F. Quatieri, and J. S. Lim, "Signal reconstruction from short-time Fourier transform magnitude," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-31, no. 4, pp. 986–998, Aug. 1983.
- [38] D. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-32, no. 2, pp. 236–243, Apr. 1984.
- [39] 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.
- [40] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.
- [41] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, MaxCut and complex semidefinite programming," *Math. Program.*, vol. 149, nos. 1–2, pp. 47–81, 2015.
- [42] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.
- [43] D. L. Sun and J. O. Smith, "Estimating a signal from a magnitude spectrogram via convex optimization," arXiv:1209.2076, 2012.
- [44] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1978.
- [45] R. Balan, P. Casazza, and D. Edidin, "On signal reconstruction without phase," *Appl. Comput. Harmonic Anal.*, vol. 20, no. 3, pp. 345–356, 2006.
- [46] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [47] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [48] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
- [49] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Convex algorithms and limitations," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2013, pp. 1022–1026.
- [50] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [51] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, pp. 49–95, 1996.



**Kishore Jaganathan** was born in Tamil Nadu, India, in 1989. He received the B.Tech. degree from the Indian Institute of Technology, Madras, Chennai, India, in 2010, and the M.S. degree from California Institute of Technology, Pasadena, CA, USA, in 2011, both in electrical engineering. He is currently pursuing the Ph.D. degree in electrical engineering at California Institute of Technology. His research interests include signal processing, convex optimization, and statistical machine learning.

He has been with the California Institute of Technology since 2011. At Caltech, he was the recipient of the Atwood Fellowship in 2010. He was also the recipient of the Qualcomm Innovation Fellowship, in 2014.



**Yonina C. Eldar** (S'98–M'02–SM'07–F'12) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering both from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2002. From January 2002 to July 2002, she was a Postdoctoral Fellow with the Digital Signal Processing Group, MIT.

She is currently a Professor with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel and holds the Edwards Chair in Engineering. She is also a Research Affiliate with the Research Laboratory of Electronics, MIT and a Visiting Professor at Stanford University, Stanford, CA, USA. Her research interests include statistical signal processing, sampling theory and compressed sensing, optimization methods, and their applications to biology and optics. She was Co-Chair and Technical Co-Chair of several international conferences and workshops. She is author of the book *Sampling Theory: Beyond Bandlimited Systems* (Cambridge University Press) and co-author of the books *Compressed Sensing and Convex Optimization Methods in Signal Processing and Communications* (Cambridge University Press).

Dr. Eldar is the Editor-in-Chief of *Foundations and Trends in Signal Processing* and a member of the IEEE Sensor Array and Multichannel Technical Committee. In the past, she was a Signal Processing Society Distinguished Lecturer, member of the IEEE SIGNAL PROCESSING THEORY AND METHODS and Bio Imaging Signal Processing Technical Committees, and served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*. She was a Horev Fellow of the Leaders in Science and Technology program at the Technion from 2002 to 2005 and an Alon Fellow. In 2004, she was the recipient of the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005, the Andre and Bella Meyer Lectureship, in 2007, the Henry Taub

Prize for Excellence in Research, in 2008, the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion Outstanding Lecture Award, in 2009, the Technions Award for Excellence in Teaching, in 2010, the Michael Bruno Memorial Award from the Rothschild Foundation, and in 2011, the Weizmann Prize for Exact Sciences. In 2012, she was elected to the Young Israel Academy of Science and to the Israel Committee for Higher Education. In 2013, she was also the recipient of the Technions Award for Excellence in Teaching, the Hershel Rich Innovation Award, and the IEEE Signal Processing Technical Achievement Award, and in 2014, the IEEE/AESS Fred Nathanson Memorial Radar Award. She received several best paper awards together with her research students and colleagues. She was selected as one of the 50 most influential women in Israel.



**Babak Hassibi** (M'08) was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran, Tehran, Iran, in 1989, and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, USA, in 1993 and 1996, respectively, all in electrical engineering.

He has been with the California Institute of Technology since January 2001, where he is currently the Gordon M Binder/Amgen Professor of electrical engineering. From 2008 to 2015, he was an Executive Officer of Electrical Engineering, as well as an

Associate Director of Information Science and Technology. From October 1996 to October 1998, he was a Research Associate with the Information Systems Laboratory, Stanford University, and from November 1998 to December 2000, he was a Member of the Technical Staff with the Mathematical Sciences Research Center, Bell Laboratories, Murray Hill, NJ, USA. He is the coauthor of the books (both with A.H. Sayed and T. Kailath) *Indefinite Quadratic Estimation and Control: A Unified Approach to  $H^2$  and  $H^\infty$  Theories* (SIAM, 1999) and *Linear Estimation* (Prentice Hall, 2000). His research interests include wireless communications and networks, robust estimation and control, adaptive signal processing and linear algebra.

He has been a Guest Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY special issue on "space-time transmission, reception, coding, and signal processing" was an Associate Editor for Communications of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2004 to 2006, and is currently an Editor for the journal *Foundations and Trends in Information and Communication* and for the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING. He is an IEEE Information Theory Society Distinguished Lecturer for 2016–2017. He was the recipient of an Alborz Foundation Fellowship, the 1999 O. Hugo Schuck Best Paper Award of the American Automatic Control Council (with H. Hindi and S.P. Boyd), the 2002 National Science Foundation Career Award, the 2002 Okawa Foundation Research Grant for Information and Telecommunications, the 2003 David and Lucille Packard Fellowship for Science and Engineering, the 2003 Presidential Early Career Award for Scientists and Engineers (PECASE), and the 2009 Al-Marai Award for Innovative Research in Communications, and was a participant in the 2004 National Academy of Engineering "Frontiers in Engineering" program.