

# PHASE RETRIEVAL FROM STFT MEASUREMENTS VIA NON-CONVEX OPTIMIZATION

Tamir Bendory and Yonina C. Eldar, Fellow IEEE

Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, Israel.

## ABSTRACT

The problem of recovering a signal from its phaseless short-time Fourier transform (STFT) measurements arises in several applications, such as ultra-short pulse measurements and ptychography. The redundancy offered by the STFT enables unique recovery under mild conditions. We show that in some cases, the principle eigenvector of a designed matrix recovers the underlying signal. This matrix is constructed as the solution of a simple least-squares problem. When these conditions are not met, we suggest to use this principle eigenvector to initialize a gradient algorithm, minimizing a non-convex loss function. We prove that under appropriate conditions, this initialization results in a good estimate of the underlying signal. We further analyze the geometry of the loss function and show empirically that the gradient algorithm is robust to noise. Our method is both efficient and enjoys theoretical guarantees.

**Index Terms**—phase retrieval, short-time Fourier transform, gradient descent, non-convex optimization, ptychography

## I. INTRODUCTION

Phase retrieval is the process of retrieving the phase information of a signal from its Fourier transform magnitude. This problem arises in many areas of engineering and science, such as optics, X-ray crystallography, speech recognition, blind channel estimation and astronomy (see for instance, [1], [2], [3], [4], [5], [6]). For almost all one-dimensional (1D) signals phase retrieval is an ill-posed problem (see e.g. [7]). To overcome the fundamental ill-posedness, it is common to employ measurement techniques that generate redundancy in the acquired information. Such measurements can be obtained by masks (see e.g. [8], [9]) or by measuring the magnitude of the short-time Fourier transform (STFT). In [10], it was demonstrated that the STFT magnitude leads to better recovery than an over-sampled discrete Fourier transform (DFT) for the same number of measurements.

This paper deals with the problem of recovering a 1D signal from its STFT magnitude. This problem serves as the model for ptychography, in which a moving probe is used to sense multiple diffraction measurements, and laser pulse measurement techniques [11], [12], [13]. It is now known that by choosing windows with appropriate properties, only a small redundancy in the measurements (i.e. short windows) is needed to enforce uniqueness (for details, see [14], [15], [16], [17], [10]). The STFT of a 1D signal  $\mathbf{x} \in \mathbb{C}^N$  with respect to a real sliding window  $\mathbf{g}$  of length  $W$  is defined as

$$\mathbf{X}[m, k] := \sum_{n=0}^{N-1} \mathbf{x}[n] \mathbf{g}[mL - n] e^{-2\pi j kn/N}, \quad (\text{I.1})$$

where  $k = 0, \dots, N-1$ ,  $m = 0, \dots, \lceil \frac{N}{L} \rceil - 1$  and  $L$  determines the separation in time between adjacent sections. Hereafter, all indices should be considered as modulo the signal's length  $N$ . We assume that  $\mathbf{x}$  and  $\mathbf{g}$  are periodically extended over the boundaries in (I.1) and that  $L$  divides  $N$ .

This work was funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 646804-ERCCOG-BNYQ, and by the Israel Science Foundation under Grant no. 335/14. TB was partially funded by the Andrew and Erna Finci Viterbi Fellowship.

Two main approaches exist to recover a signal from STFT magnitudes. The classic method, called Griffin-Lim algorithm (GLA) [18], is a modification of the alternating projection algorithms of Gerchberg and Saxton and Fineup [19], [20] (see also [12]). The properties of this algorithm are not well-understood. Recently, the authors of [16] suggested to relax the problem to a tractable semidefinite program (SDP). While this algorithm works well, solving an SDP requires high computational resources. We suggest a third approach based on minimizing a non-convex loss function by gradient descent. In contrast to previous works, our algorithm reflects a practical setup, is computationally efficient and enjoys theoretical guarantees.

We begin by taking the 1D DFT of the acquired information with respect to the frequency variable. This transformation greatly simplifies the analysis and implies, almost directly, a least-squares (LS) algorithm for  $L = 1$  and long windows  $W \geq \lceil \frac{N+1}{2} \rceil$ . Specifically, under appropriate conditions one can recover the signal by extracting the principle eigenvector of a designed matrix. This matrix approximates the correlation matrix  $\mathbf{X} := \mathbf{x}\mathbf{x}^*$  and is obtained as the solution of a simple LS problem. When the conditions for a closed-form solution are not met, we propose using the principle eigenvector of the designed matrix to initialize a gradient descent (GD) algorithm that minimizes a non-convex loss function. Our approach deviates in two important aspects from the recent line of work in non-convex phase retrieval [21], [22], [23], [24], [25]. First, all these papers focus their attention on the setup of phase retrieval with random sensing vectors and rely heavily on statistical considerations, while we consider a deterministic framework. Second, we approximate  $\mathbf{X}$  by solving a LS problem, whereas the aforementioned papers approximate it by taking a superposition of the measurements.

The properties of a GD algorithm depend on the initialization method and the geometry of the loss function. For  $L = 1$ , we estimate the distance between the proposed initialization and the global minimum, which decays to zero as  $W$  tends to  $\frac{N+1}{2}$ . We also prove the existence of a *basin of attraction* around the global minimum of the loss function for signals with unit module entries. We note that while the theoretical guarantees of the algorithm are limited, the algorithm performs well and is robust to noise.

The paper is organized as follows. We begin in Section II by formulating mathematically the problem of phase retrieval from STFT magnitude measurements. In Section III we present conditions under which it has a closed-form LS solution. Section IV considers the gradient algorithm, its initialization and numerical results. Section V presents our theoretical findings and Section VI concludes the paper.

Throughout the paper we use the following notation. Bold small and capital letters denote vectors and matrices, respectively. We use  $\mathbf{Z}^\dagger$  and  $\text{tr}(\mathbf{Z})$  for the pseudo-inverse and the trace of the matrix  $\mathbf{Z}$ , respectively. The  $\ell$ th circular diagonal of a matrix  $\mathbf{Z}$  is denoted by  $\text{diag}(\mathbf{Z}, \ell)$ . Namely,  $\text{diag}(\mathbf{Z}, \ell)$  is a column vector with entries  $\mathbf{Z}[i, (i + \ell) \bmod N]$  for  $i = 0, \dots, N-1$ . We reserve  $\circ$  and  $\ast$  for the Hadamard (point-wise) product and convolution, respectively. The sign of a complex number  $a$  is defined as  $\text{sign}(a) = \frac{a}{|a|}$ .

## II. PROBLEM FORMULATION

We aim at recovering the underlying signal  $\mathbf{x}$  from the magnitude of its STFT, i.e. from measurements

$$\mathbf{Z}[m, k] = |\mathbf{X}[m, k]|^2. \quad (\text{II.1})$$

Note that the signals  $\mathbf{x}$  and  $\mathbf{x}e^{j\phi}$  yield the same measurements for any *global phase*  $\phi \in \mathbb{R}$  and therefore the phase  $\phi$  cannot be recovered by any method. This *global phase ambiguity* leads naturally to the following definition:

**Definition II.1.** *The distance between two vectors is defined as*

$$d(\mathbf{z}, \mathbf{x}) = \min_{\phi \in [0, 2\pi)} \|\mathbf{z} - \mathbf{x}e^{j\phi}\|_2.$$

We say that  $\mathbf{x}$  and  $\mathbf{z}$  are equal up to global phase if  $d(\mathbf{z}, \mathbf{x}) = 0$ .

Instead of treating the measurements (II.1) directly, we consider the acquired data in a transformed domain by taking its 1D DFT with respect to the frequency variable (normalized by  $\frac{1}{N}$ ). Then, our measurement model reads

$$\begin{aligned} \mathbf{Y}[m, \ell] &= \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{Z}[m, k] e^{-2\pi j k \ell / N} \\ &= \sum_{n=0}^{N-1} \mathbf{x}[n] \mathbf{x}^*[n + \ell] \mathbf{g}[mL - n] \mathbf{g}[mL - n - \ell]. \end{aligned} \quad (\text{II.2})$$

When  $W \leq \ell \leq (N - W)$ , we have  $\mathbf{Y}[m, \ell] = 0$  for all  $m$ . Observe that for fixed  $m$ ,  $\mathbf{Y}[m, \ell]$  is simply the auto-correlation of  $\mathbf{x} \circ \mathbf{g}_{mL}$ , where  $\mathbf{g}_{mL} := \{\mathbf{g}[mL - n]\}_{n=0}^{N-1}$ .

We will make repetitive use of two representations of (II.2). The first is based on a matrix formulation. Let  $\mathbf{D}_{mL} \in \mathbb{R}^{N \times N}$  be a diagonal matrix composed of the entries of  $\mathbf{g}_{mL}$  and let  $\mathbf{P}_\ell$  be a matrix that shifts (circularly) the entries of a vector by  $\ell$  locations. Then,  $\mathbf{X} := \mathbf{x}\mathbf{x}^*$  is mapped linearly to  $\mathbf{Y}[m, \ell]$  by the relation:

$$\mathbf{Y}[m, \ell] = \mathbf{x}^* \mathbf{H}_{m, \ell} \mathbf{x} = \text{tr}(\mathbf{X} \mathbf{H}_{m, \ell}), \quad (\text{II.3})$$

where

$$\mathbf{H}_{m, \ell} := \mathbf{P}_{-\ell} \mathbf{D}_{mL} \mathbf{D}_{mL - \ell}. \quad (\text{II.4})$$

Observe that  $\mathbf{P}_\ell^T = \mathbf{P}_{-\ell}$  and  $\mathbf{H}_{m, \ell} = 0$  for  $W \leq \ell \leq (N - W)$ .

An alternative useful representation of (II.2) is as multiple systems of linear equations. For fixed  $|\ell| \leq W - 1$  we can write

$$\mathbf{y}_\ell = \mathbf{G}_\ell \mathbf{x}_\ell, \quad (\text{II.5})$$

where  $\mathbf{y}_\ell := \{\mathbf{Y}[m, \ell]\}_{m=0}^{N-1}$  and  $\mathbf{x}_\ell := \text{diag}(\mathbf{X}, \ell)$ . The  $(m, n)$ th entry of the matrix  $\mathbf{G}_\ell \in \mathbb{R}^{\frac{N}{L} \times N}$  is given by  $\mathbf{g}[mL - n] \mathbf{g}[mL - n - \ell]$ . Let  $\mathbf{g} := \{\mathbf{g}[n]\}_{n=0}^{N-1}$ . For  $L = 1$ ,  $\mathbf{G}_\ell$  is a circulant matrix and therefore it can be factored as  $\mathbf{G}_\ell = \mathbf{F}^* \mathbf{\Sigma}_\ell \mathbf{F}$ , where  $\mathbf{F}$  is the DFT matrix and  $\mathbf{\Sigma}_\ell$  is a diagonal matrix, whose entries are given by  $\mathbf{F}(\mathbf{g} \circ (\mathbf{P}_{-\ell} \mathbf{g}))$ . Therefore the matrix  $\mathbf{G}_\ell$  is invertible if and only if  $\mathbf{F}(\mathbf{g} \circ (\mathbf{P}_{-\ell} \mathbf{g}))$  is non-vanishing.

Our problem of recovering  $\mathbf{x}$  from the measurements (II.1) can therefore be equivalently posed as

$$\begin{aligned} \min_{\tilde{\mathbf{X}} \in \mathcal{H}^N} \sum_{\ell=-(W-1)}^{W-1} \|\mathbf{y}_\ell - \mathbf{G}_\ell \text{diag}(\tilde{\mathbf{X}}, \ell)\|_2^2 \\ \text{subject to } \tilde{\mathbf{X}} \succeq 0, \quad \text{rank}(\tilde{\mathbf{X}}) = 1, \end{aligned} \quad (\text{II.6})$$

where  $\mathcal{H}^N$  is the set of all Hermitian matrices of size  $N$ . In the spirit of standard phase retrieval techniques, the problem can then be relaxed to a tractable SDP by dropping the rank constraint [26], [27], [28], [29]. Nonetheless, it requires solving the problem in a lifted domain with  $N^2$  variables. We take a different route to reduce the computational load. In the next sections, we show that (II.6) has a closed-form LS solution when the window  $\mathbf{g}$  is sufficiently long. If the conditions for the LS solution are not met, then we suggest applying a GD algorithm. To initialize the GD, we approximate

---

**Algorithm 1** Least-squares algorithm for  $L = 1$

---

**Input:** The measurements  $\mathbf{Z}[m, k]$  as given in (II.1).

**Output:**  $\mathbf{x}_0$ : estimation of  $\mathbf{x}$ .

- 1) Compute  $\mathbf{Y}[m, \ell]$ , the 1D DFT with respect to the second variable of  $\mathbf{Z}[m, k]$  as given in (II.2).
- 2) Construct a matrix  $\mathbf{X}_0$  such that

$$\text{diag}(\mathbf{X}_0, \ell) = \begin{cases} \mathbf{G}_\ell^\dagger \mathbf{y}_\ell & \ell = -(W-1), \dots, (W-1), \\ 0 & \text{otherwise,} \end{cases}$$

where  $\mathbf{G}_\ell \in \mathbb{R}^{N \times N}$  are defined in (II.5).

- 3) Let  $\mathbf{x}_p$  be the principle (unit-norm) eigenvector of  $\mathbf{X}_0$ . Then,

$$\mathbf{x}_0 = \sqrt{\sum_{n \in P} (\mathbf{G}_0^\dagger \mathbf{y}_0)[n] \mathbf{x}_p},$$

where  $P := \{n : (\mathbf{G}_0^\dagger \mathbf{y}_0)[n] > 0\}$ .

---

(II.6) in two stages by first solving the LS objective function and then extracting its principal eigenvector.

## III. LEAST-SQUARES ALGORITHM

The objective of the non-convex problem (II.6) implies that the success of a recovery algorithm is related to the window's length  $W$  and the invertibility of the matrices  $\mathbf{G}_\ell$  for  $|\ell| < W$ . For that reason, we focus our attention to windows for which the associated matrices  $\mathbf{G}_\ell$  for  $|\ell| < W$  are invertible:

**Definition III.1.** *A window  $\mathbf{g}$  is called an admissible window of length  $W$  if for all  $\ell = -(W-1), \dots, (W-1)$  the associated circulant matrices  $\mathbf{G}_\ell$  as given in (II.5) for  $L = 1$  are invertible.*

The family of admissible windows is quite large. For instance, a rectangular window of any length is admissible if  $N$  is a prime number (see Claim 3.3 in [14]).

For  $L = 1$ , we next derive a LS algorithm that stably recovers any complex signal if the window is sufficiently long. In the absence of noise, the recovery is exact (up to global phase). The method, summarized in Algorithm 1, is based on constructing a matrix  $\mathbf{X}_0$  that approximates the correlation matrix  $\mathbf{X} := \mathbf{x}\mathbf{x}^*$ . The  $\ell$ th diagonal of  $\mathbf{X}_0$  is chosen as the solution of the LS problem  $\min_{\tilde{\mathbf{x}} \in \mathbb{C}^N} \|\mathbf{y}_\ell - \mathbf{G}_\ell \tilde{\mathbf{x}}\|_2$  (see (II.5)). If  $\mathbf{G}_\ell$  is invertible, then

$$\text{diag}(\mathbf{X}_0, \ell) = \mathbf{G}_\ell^{-1} \mathbf{y}_\ell = \text{diag}(\mathbf{X}, \ell).$$

Therefore, when all matrices  $\mathbf{G}_\ell$  are invertible,  $\mathbf{X}_0 = \mathbf{X}$ . In order to estimate  $\mathbf{x}$ , the (unit-norm) principle eigenvector of  $\mathbf{X}_0$  is normalized by

$$\alpha = \sqrt{\sum_{n \in P} (\mathbf{G}_0^\dagger \mathbf{y}_0)[n]}, \quad (\text{III.1})$$

where  $P := \{n : (\mathbf{G}_0^\dagger \mathbf{y}_0)[n] > 0\}$ . If  $\mathbf{G}_0$  is invertible then

$$\sum_{n=0}^{N-1} (\mathbf{G}_0^{-1} \mathbf{y}_0)[n] = \sum_{n=0}^{N-1} (\text{diag}(\mathbf{X}, 0))[n] = \|\mathbf{x}\|_2^2 = \lambda_0,$$

where  $\lambda_0$  is the top eigenvector of  $\mathbf{X}$ . If  $\mathbf{G}_0$  is not invertible or in the presence of noise, some terms of  $\mathbf{G}_0^\dagger \mathbf{y}_0$  might be negative. In this case, we estimate  $\|\mathbf{x}\|_2$  by summing only the positive terms. Note that all matrix inversions can be performed efficiently using the FFT due to the circular structure of  $\mathbf{G}_\ell$ .

The following proposition shows that for  $L = 1$ , Algorithm 1 recovers the underlying signal if the window is sufficiently long and satisfies some additional technical conditions (the proof can be found in [14], see also [30], [31]).

**Proposition III.2.** *Let  $L = 1$  and suppose that  $\mathbf{g}$  is an admissible window of length  $W \geq \lceil \frac{N+1}{2} \rceil$ . Then, Algorithm 1 recovers any complex signal uniquely (up to global phase) and efficiently.*

---

**Algorithm 2** Gradient descent algorithm

---

**Input:** The measurements  $\mathbf{Z}[m, k]$  as given in (II.1) and (optional) thresholding parameter  $B > 0$ .

**Output:** Estimation of  $\mathbf{x}$ .

- 1) Initialization by Algorithm 1 (for  $L = 1$ ) or Algorithm 3 (for  $L > 1$ ).
- 2) Apply the update rule until convergence:
  - a) (gradient step)

$$\tilde{\mathbf{x}}_k = \mathbf{x}_{k-1} - \mu \nabla f(\mathbf{x}_{k-1}),$$

for step size  $\mu$  and  $\nabla f$  given in (IV.2).

- b) (optional thresholding)

$$\mathbf{x}_k[n] = \text{sign}(\tilde{\mathbf{x}}_k[n]) \min\{|\tilde{\mathbf{x}}_k[n]|, B\}$$

---

If the signal has unit module entries, then a slight modification of Algorithm 1 recovers the signal exactly for  $W \geq 2$  (see [14]).

In many cases, the window is shorter than  $\lceil \frac{N+1}{2} \rceil$  so that (II.6) does not admit a closed-form LS solution. In these cases, we employ a GD algorithm to minimize a non-convex loss function. To initialize the algorithm, we use the same LS-based algorithm. However, for short windows we cannot estimate  $\text{diag}(\mathbf{X}, \ell)$  for  $\ell = W, \dots, (N - W)$  as the matrices  $\mathbf{G}_\ell$  are simply zero. Nonetheless, we will show by both theoretical results and numerical experiments that under appropriate conditions, the principle eigenvector of  $\mathbf{X}_0$ , with a proper normalization, is a good estimate of  $\mathbf{x}$ .

#### IV. GRADIENT DESCENT ALGORITHM

Recall that by taking the DFT with respect to the frequency variable, the measurement model reads  $\mathbf{Y}[m, \ell] = \mathbf{x}^* \mathbf{H}_{m, \ell} \mathbf{x}$ , where  $\mathbf{H}_{m, \ell}$  is defined in (II.4). It is therefore natural to minimize the following non-convex loss function (also called empirical risk):

$$f(\mathbf{z}) = \frac{1}{2} \sum_{m=0}^{N/L-1} \sum_{\ell=-(W-1)}^{W-1} (\mathbf{z}^* \mathbf{H}_{m, \ell} \mathbf{z} - \mathbf{Y}[m, \ell])^2. \quad (\text{IV.1})$$

To keep the framework simple, we focus here on real signals. If the signal is complex, we replace the inner term of (IV.1) by  $|\mathbf{z}^* \mathbf{H}_{m, \ell} \mathbf{z} - \mathbf{Y}[m, \ell]|^2$ . We suggest minimizing (IV.1) by employing a gradient algorithm, where the  $k$ th iteration takes on the form

$$\mathbf{x}_k = \mathbf{x}_{k-1} - \mu \nabla f(\mathbf{x}_{k-1}),$$

for step size  $\mu$ . Direct computation of the gradient gives

$$\nabla f(\mathbf{z}) = \sum_{m=0}^{N/L-1} \sum_{\ell=-(W-1)}^{W-1} (h(\mathbf{z}) - \mathbf{Y}[m, \ell]) \nabla h(\mathbf{z}), \quad (\text{IV.2})$$

where  $h(\mathbf{z}) := \mathbf{z}^T \mathbf{H}_{m, \ell} \mathbf{z}$  and  $\nabla h(\mathbf{z}) = (\mathbf{H}_{m, \ell} + \mathbf{H}_{m, \ell}^T) \mathbf{z}$ . The algorithm is summarized in Algorithm 2. The code is publicly available on <http://webee.technion.ac.il/Sites/People/YoninaEldar>.

##### IV-A. Initialization for $L = 1$

We note that the function (IV.1) is non-convex. Hence, it is not clear whether an arbitrary initialization will converge to a global minimum. When  $L = 1$ , we propose initializing the GD algorithm by using Algorithm 1. As explained in Section III, for  $W \geq \lceil \frac{N+1}{2} \rceil$  the algorithm returns  $\mathbf{x}$  exactly. However, when  $W < \lceil \frac{N+1}{2} \rceil$ ,  $\mathbf{G}_\ell = 0$  for  $\ell = W, \dots, (N - W)$  so Proposition III.2 does not hold. Nevertheless, in Section V we provide theoretical guarantees establishing that under appropriate conditions, this initialization results in a good estimate of  $\mathbf{x}$ .

##### IV-B. Initialization for $L > 1$

When  $L > 1$ , the representation (II.5) results in an underdetermined system of equations as  $\mathbf{y}_\ell \in \mathbb{R}^{\frac{N}{L}}$ ,  $\mathbf{G}_\ell \in \mathbb{R}^{\frac{N}{L} \times N}$  and  $\mathbf{x}_\ell \in \mathbb{R}^N$ . We notice that  $\mathbf{y}_\ell$  is a downsampled version by a factor  $L$  of the case of maximal overlapping ( $L = 1$ ). Therefore, we suggest upsampling  $\mathbf{y}_\ell$  to approximate the case of  $L = 1$  based on

---

**Algorithm 3** Least-squares initialization for  $L > 1$ 

---

**Input:** The measurements  $\mathbf{Z}[m, k]$  as given in (II.1) and a smooth interpolation filter  $\mathbf{h}_L \in \mathbb{R}^N$  that approximates a low-pass filter with bandwidth  $N/L$ .

**Output:**  $\mathbf{x}_0$ : estimation of  $\mathbf{x}$ .

- 1) Compute  $\mathbf{Y}[m, \ell]$ , the 1D DFT with respect to the second variable of  $\mathbf{Z}[m, k]$  as given in (II.2).
- 2) (upsampling) For each  $\ell \in [-(W-1), \dots, (W-1)]$ :

- a) Let  $\mathbf{y}_\ell[m] := \{\mathbf{Y}[m, \ell]\}_{m=0}^{\frac{N}{L}-1}$  for fixed  $\ell$ .
  - b) (expansion)

$$\tilde{\mathbf{y}}_\ell[n] := \begin{cases} \mathbf{y}_\ell[m], & n = mL, \\ 0, & \text{otherwise.} \end{cases}$$

- c) (interpolation)

$$\bar{\mathbf{y}}_\ell = \tilde{\mathbf{y}}_\ell * \mathbf{h}_L$$

- 3) Construct a matrix  $\mathbf{X}_0$  such that

$$\text{diag}(\mathbf{X}_0, \ell) = \begin{cases} \mathbf{G}_\ell^\dagger \bar{\mathbf{y}}_\ell & \ell = -(W-1), \dots, (W-1), \\ 0 & \text{otherwise,} \end{cases}$$

where  $\mathbf{G}_\ell \in \mathbb{R}^{N \times N}$  are defined as in (II.5) for  $L = 1$ .

- 4) Let  $\mathbf{x}_p$  be the principle (unit-norm) eigenvector of  $\mathbf{X}_0$ . Then,

$$\mathbf{x}_0 = \sqrt{\sum_{n \in P} (\mathbf{G}_0^\dagger \mathbf{y}_0)[n] \mathbf{x}_p},$$

$$\text{where } P := \{n : (\mathbf{G}_0^\dagger \mathbf{y}_0)[n] > 0\}.$$

---

the averaging nature of the window  $\mathbf{g}$ . When  $\mathbf{F}(\mathbf{g} \circ (\mathbf{P}_{-\ell} \mathbf{g}))$  is an ideal low-pass filter with band-width  $N/L_{BW}$ , then no information is lost by taking  $L = L_{BW}$ . In this case, the upsampled (i.e.  $L = 1$ ) signal can be obtained by expansion and low-pass interpolation (see Lemma 4.1 in [14]). In practice we do not use ideal low-pass windows. In the spirit of standard DSP practice (see Section 4.6.1 of [32]) we approximate the low-pass interpolation by a simple smooth interpolation. This gives us better numerical results and reduces the computational complexity. Following the upsampling stage, the algorithm proceeds as for  $L = 1$  by extracting the principle eigenvector (with proper normalization) of a designed matrix. This initialization is summarized in Algorithm 3.

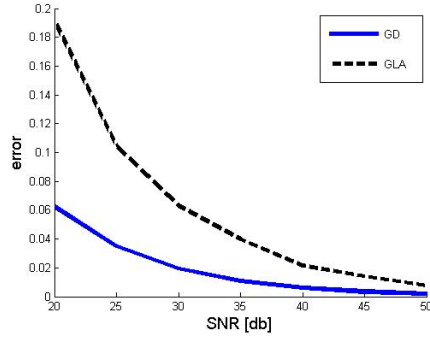
#### IV-C. Numerical Experiments

We present two numerical examples demonstrating the effectiveness of the proposed GD algorithm. In the experiments, the underlying signal is composed of additive white Gaussian entries with variance one. The measurements were contaminated with noise that was drawn from the same distribution with the appropriate variance according to the desired signal to noise (SNR) ratio. Figure 1 shows a representative example for a recovery process. Figure 2 presents the normalized recovery error as a function of the SNR level. We compared the algorithm's performance with the Griffin-Lim Algorithm (GLA) [18]. As can be seen, the GD algorithm outperforms the GLA especially in high noise levels. For more comprehensive numerical results, see [14].

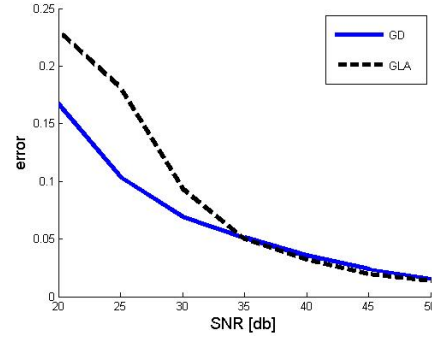
#### V. THEORY

In this section we summarize our main theoretical results. Due to paucity of space, we omit the proofs which can be found in [14]. The success of gradient algorithms stands on two pillars: the accuracy of the initialization and the geometry of the loss function near the global minimum. The following result quantifies the estimation error of the initialization presented in Algorithm 1 for bounded signals and  $L = 1$ . The error reduces to zero as  $W$  approaches  $\frac{N+1}{2}$ .

**Theorem V.1.** Suppose that  $L = 1$ ,  $\mathbf{g}$  is an admissible window of length  $W \geq 2$  and that  $\|\mathbf{x}\|_\infty \leq \sqrt{\frac{B}{N}}$  for some  $0 < B \leq$

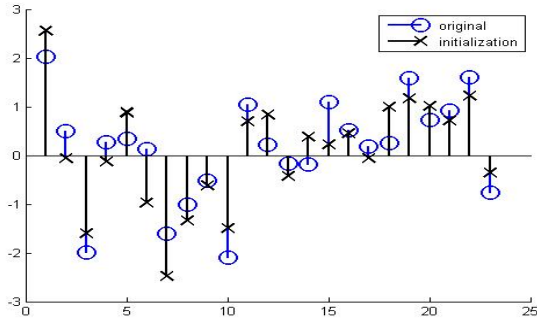


(a) Recovery error with  $L = 2$

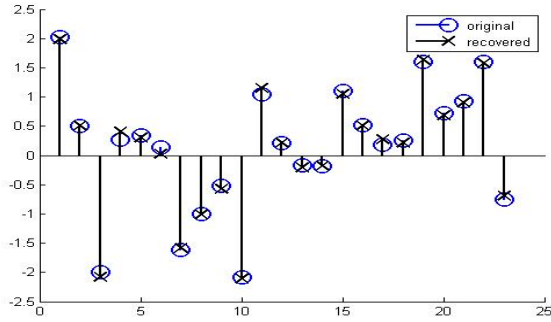


(b) Recovery error with  $L = 4$

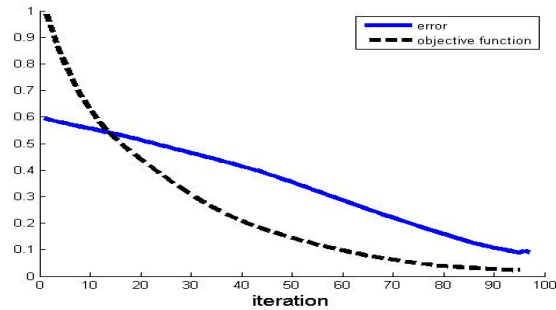
**Fig. 2:** The average recovery error (over 20 experiments) of the GD and GLA algorithms in the presence of noise. The experiments were conducted on signal of length  $N = 53$  with a rectangular window of length  $W = 19$ , step size  $\mu = 5 \times 10^{-3}$  and  $L = 2, 4$ . The error is computed as  $\frac{d(\mathbf{x}, \mathbf{x}_0)}{\|\mathbf{x}\|_2}$ , where  $\mathbf{x}_0$  is the estimated signal.



(a) Initialization by Algorithm 3



(b) Recovery by Algorithm 2



(c) The error and objective function value curves as a function of iterations

**Fig. 1:** Recovery of a signal of length  $N = 23$  with a rectangular window of length  $W = 7$ ,  $L = 2$ ,  $\mu = 5 \times 10^{-2}$  in a noisy environment of SNR= 25 db.

$\frac{N}{2(N-2W+1)}$ . Then, under the measurement model of (II.1), the initialization point as given in Algorithm 1 satisfies

$$d^2(\mathbf{x}_0, \mathbf{x}) \leq \|\mathbf{x}\|_2^2 \left( 1 - \sqrt{1 - 2B \frac{N - 2W + 1}{N}} \right).$$

The second result quantifies the basin of attraction of the loss function (IV.1) for signals with unit module entries. In this area, the gradient algorithm is guaranteed to converge to the global minimum. Numerical experiments (not shown here) indicate that in practice the basin of attraction is quite large and exists for a broad family of signals.

**Theorem V.2.** Let  $L = 1$ , suppose that  $\mathbf{x}[i] \in \{\pm \frac{1}{\sqrt{N}}\}$  for all  $i$  and  $\mathbf{g}$  is a rectangular window of length  $W$ . Additionally, suppose that the initialization point  $\mathbf{x}_0$  obeys  $d(\mathbf{x}_0, \mathbf{x}) \leq \frac{1}{8\sqrt{N}W^2}$  and  $\|\mathbf{x}_0\|_\infty \leq \frac{1}{\sqrt{N}}$ . Then, under the measurement model (II.1), Algorithm 2 with step size  $0 < \mu \leq 2/\beta$  and  $B = \frac{1}{N}$  achieves the following geometrical convergence

$$d^2(\mathbf{x}_k, \mathbf{x}) \leq \left( 1 - \frac{2\mu}{\alpha} \right)^k d^2(\mathbf{x}_{k-1}, \mathbf{x}),$$

where  $\alpha \geq \frac{4N}{W}$  and  $\beta \geq 256N^2W^3$ .

## VI. DISCUSSION

This paper aims at suggesting a practical and efficient phase retrieval algorithm with theoretical guarantees. The algorithm begins by taking the DFT of the measurements (II.1). For sufficiently long windows, we show that the principal eigenvector of a designed matrix recovers the signal. This matrix is constructed as the solution of a LS problem. For general settings, we employ a gradient descent algorithm, initialized by the principle eigenvector of the same designed matrix. While for  $L = 1$  we derived an estimation of the distance between the initialization and the global minimum, the case of  $L > 1$  raises some interesting questions. In this case, we suggest to smoothly interpolate the missing entries. This practice works quite well since the window acts as an averaging operator. A main challenge for future research is analyzing this setting.

The analysis of the non-convex algorithm relies on the geometry of the loss function (IV.1). Theorem V.2 states that for signals with unit module entries, there exists a basin of attraction. Yet, numerical experiments indicate that the actual basin of attraction is larger than the theoretical bound and exists for a broader family of signals. The gap between the actual basin of attraction and the theoretical result is the bottleneck that prevents a full theoretical understanding of the algorithm. Bridging that gap is an additional major goal of a future work. Additionally, the theoretical guarantees for signal with unit module entries implies a potential applicability to angular synchronization [33], [34], [35].

## VII. REFERENCES

- [1] R. Harrison, "Phase problem in crystallography," *JOSA A*, vol. 10, no. 5, pp. 1046–1055, 1993.
- [2] B. Juang and L. Rabiner, "Fundamentals of speech recognition," *Signal Processing Series*. Prentice Hall, Englewood Cliffs, NJ, 1993.
- [3] B. Baykal, "Blind channel estimation via combining autocorrelation and blind phase estimation," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 51, no. 6, pp. 1125–1131, 2004.
- [4] C. Fienup and J. Dainty, "Phase retrieval and image reconstruction for astronomy," *Image Recovery: Theory and Application*, pp. 231–275, 1987.
- [5] Y. Shechtman, Y. Eldar, O. Cohen, H. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," *Signal Processing Magazine, IEEE*, vol. 32, no. 3, pp. 87–109, 2015.
- [6] K. Jaganathan, Y. Eldar, and B. Hassibi, "Phase retrieval: An overview of recent developments," *arXiv preprint arXiv:1510.07713*, 2015.
- [7] K. Huang, Y. Eldar, and N. Sidiropoulos, "Phase retrieval from 1D Fourier measurements: Convexity, uniqueness, and algorithms," *arXiv preprint arXiv:1603.05215*, 2016.
- [8] E. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Applied and Computational Harmonic Analysis*, vol. 39, no. 2, pp. 277–299, 2015.
- [9] D. Gross, F. Kraemer, and R. Kueng, "Improved recovery guarantees for phase retrieval from coded diffraction patterns," *Applied and Computational Harmonic Analysis*, 2015.
- [10] Y. Eldar, P. Sidorenko, D. Mixon, S. Barel, and O. Cohen, "Sparse phase retrieval from short-time fourier measurements," *Signal Processing Letters, IEEE*, vol. 22, no. 5, pp. 638–642, 2015.
- [11] J. Rodenburg, "Ptychography and related diffractive imaging methods," *Advances in Imaging and Electron Physics*, vol. 150, no. 07, pp. 87–184, 2008.
- [12] S. Marchesini, Y. Tu, and H. Wu, "Alternating projection, ptychographic imaging and phase synchronization," *Applied and Computational Harmonic Analysis*, 2015.
- [13] R. Trebino, "Frequency resolved optical gating: The measurement of ultrashort optical pulses," *Kluwer Academic Publishers, Boston*, vol. 5, pp. 70–73, 2002.
- [14] T. Bendory and Y. C. Eldar, "Non-Convex Phase Retrieval from STFT Measurements," *arXiv preprint arXiv:1607.08218*, 2016.
- [15] S. Nawab, T. Quatieri, and J. Lim, "Signal reconstruction from short-time fourier transform magnitude," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 31, no. 4, pp. 986–998, 1983.
- [16] K. Jaganathan, Y. Eldar, and B. Hassibi, "STFT phase retrieval: Uniqueness guarantees and recovery algorithms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 770–781, 2016.
- [17] I. Bojarovska and A. Flinthe, "Phase retrieval from gabor measurements," *Journal of Fourier Analysis and Applications*, vol. 22, no. 3, pp. 542–567, 2016.
- [18] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 2, pp. 236–243, 1984.
- [19] R. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, p. 237, 1972.
- [20] J. Fienup, "Phase retrieval algorithms: a comparison," *Applied optics*, vol. 21, no. 15, pp. 2758–2769, 1982.
- [21] E. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *Information Theory, IEEE Transactions on*, vol. 61, no. 4, pp. 1985–2007, 2015.
- [22] Y. Chen and E. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *To appear in Communications on Pure and Applied Mathematics*, 2015.
- [23] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *Signal Processing, IEEE Transactions on*, vol. 63, no. 18, pp. 4814–4826, 2015.
- [24] C. White, S. Sanghavi, and R. Ward, "The local convexity of solving systems of quadratic equations," *arXiv preprint arXiv:1506.07868*, 2015.
- [25] G. Wang, G. Giannakis, and Y. Eldar, "Solving random systems of quadratic equations via truncated generalized gradient flow," *arXiv preprint arXiv:1605.08285*, 2016.
- [26] M. Goemans and D. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the ACM (JACM)*, vol. 42, no. 6, pp. 1115–1145, 1995.
- [27] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1–2, pp. 47–81, 2015.
- [28] E. Candes, Y. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Review*, vol. 57, no. 2, pp. 225–251, 2015.
- [29] Y. Shechtman, Y. Eldar, A. Szameit, and M. Segev, "Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing," *Optics express*, vol. 19, no. 16, pp. 14807–14822, 2011.
- [30] J. Rodenburg and R. Bates, "The theory of super-resolution electron microscopy via wigner-distribution deconvolution," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 339, no. 1655, pp. 521–553, 1992.
- [31] C. Yang, J. Qian, A. Schirotzek, F. Maia, and S. Marchesini, "Iterative algorithms for ptychographic phase retrieval," *arXiv preprint arXiv:1105.5628*, 2011.
- [32] A. Oppenheim and R. Schaffer, *Discrete-time signal processing*. Pearson Higher Education, 2010.
- [33] A. Singer, "Angular synchronization by eigenvectors and semidefinite programming," *Applied and computational harmonic analysis*, vol. 30, no. 1, pp. 20–36, 2011.
- [34] A. Bandeira, N. Boumal, and A. Singer, "Tightness of the maximum likelihood semidefinite relaxation for angular synchronization," *arXiv preprint arXiv:1411.3272*, 2014.
- [35] N. Boumal, "Nonconvex phase synchronization," *arXiv preprint arXiv:1601.06114*, 2016.