Fast Phase Retrieval from Local Correlation Measurements[*]

Mark A. Iwen[†], Aditya Viswanathan[‡], and Yang Wang[§]

Abstract. We develop a fast phase retrieval method which can utilize a large class of local phaseless correlation-based measurements in order to recover a given signal $\mathbf{x} \in \mathbb{C}^d$ (up to an unknown global phase) in near-linear $\mathcal{O}(d \log^4 d)$-time. Accompanying theoretical analysis proves that the proposed algorithm is guaranteed to deterministically recover all signals $\mathbf{x}$ satisfying a natural flatness (i.e., nonsparsity) condition for a particular choice of deterministic correlation-based measurements. A randomized version of these same measurements is then shown to provide nonuniform probabilistic recovery guarantees for arbitrary signals $\mathbf{x} \in \mathbb{C}^d$. Numerical experiments demonstrate the method's speed, accuracy, and robustness in practice—all code is made publicly available. In its simplest form, our proposed phase retrieval method employs a modified lifting scheme akin to the one utilized by the well-known *PhaseLift* algorithm. In particular, it interprets quadratic magnitude measurements of $\mathbf{x}$ as linear measurements of a restricted set of lifted variables, $x_i \overline{x_j}$, for $|j - i| < \delta \ll d$. This leads to a linear system involving a total of $(2\delta - 1)d$ unknown lifted variables, all of which can then be solved for using only $\mathcal{O}(\delta d)$ measurements. Once these lifted variables, $x_i \overline{x_j}$ for $|j - i| < \delta \ll d$, have been recovered, a fast angular synchronization method can then be used to propagate the local phase difference information they provide across the entire vector in order to estimate the (relative) phases of every entry of $\mathbf{x}$. In addition, the lifted variables corresponding to $x_j \overline{x_j} = |x_j|^2$ automatically provide magnitude estimates for each entry, x_j, of $\mathbf{x}$. The proposed phase retrieval method then approximates $\mathbf{x}$ by carefully combining these entrywise phase and magnitude estimates. Finally, we conclude by developing an extension of the proposed method to the sparse phase retrieval problem; specifically, we demonstrate a sublinear-time compressive phase retrieval algorithm which is guaranteed to recover a given s-sparse vector $\mathbf{x} \in \mathbb{C}^d$ with high probability in just $\mathcal{O}(s \log^5 s \cdot \log d)$-time using only $\mathcal{O}(s \log^4 s \cdot \log d)$ magnitude measurements. In doing so we demonstrate the existence of compressive phase retrieval algorithms with near-optimal linear-in-sparsity runtime complexities.

Key words. phase retrieval, ptychography, angular synchronization, compressive phase retrieval, sublinear-time algorithms

AMS subject classifications. 49N45, 49N30, 42A38, 65T50, 15B52, 68W25

DOI. 10.1137/15M1053761

1. Introduction. We consider the *phase retrieval problem* of recovering a vector $\mathbf{x} \in \mathbb{C}^d$, up to an unknown global phase factor, from squared magnitude measurements

$$(1) \qquad \mathbf{b} := |M\mathbf{x}|^2 + \mathbf{n},$$

[†]Department of Mathematics and Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824 (markiwen@math.msu.edu).
[‡]Department of Mathematics, Michigan State University, East Lansing, MI 48824 (aditya@math.msu.edu).
[§]Department of Mathematics, The Hong Kong University of Science and Technology, Kowloon, Hong Kong (yangwang@ust.hk).
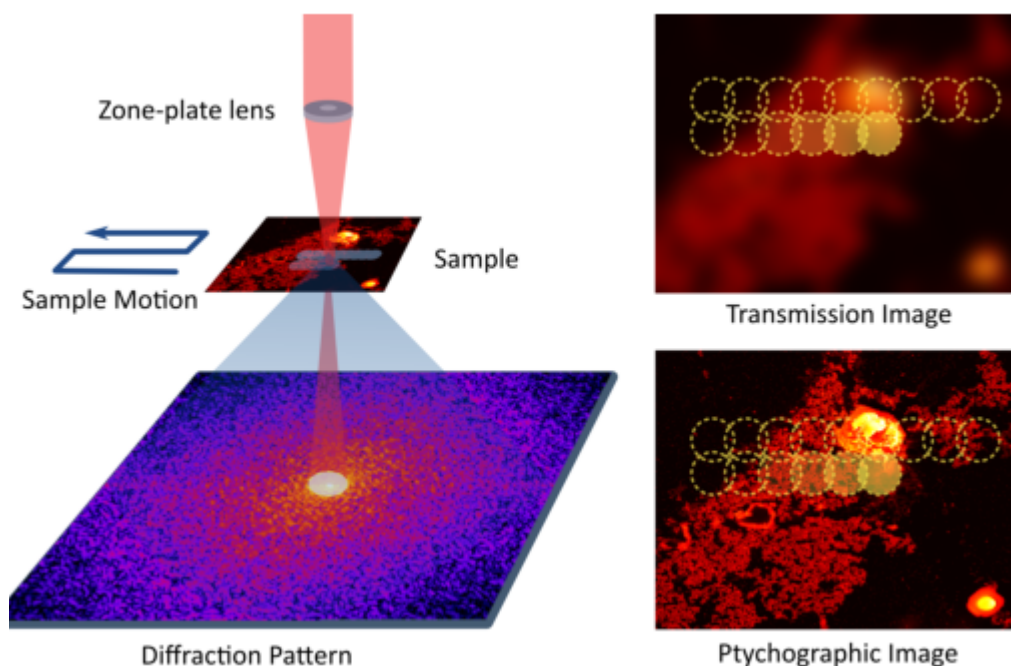
Figure 1. *A typical imaging setup for ptychography. First published in J. Qian et al., Inverse Problems Applications, Contemp. Math. 615, AMS, Providence, RI, 2014, pp. 261–280; used by permission of the American Mathematical Society.*

where $\mathbf{b} \in \mathbb{R}^D$ is the vector of phaseless measurements (with $D \geq d$), $M \in \mathbb{C}^{D \times d}$ is a measurement matrix, $\mathbf{n} \in \mathbb{R}^D$ denotes measurement noise, and, $|\cdot|^2 : \mathbb{C}^D \to \mathbb{R}^D$ computes the componentwise squared magnitude of each vector entry. Our objective is to design a computationally efficient and robust recovery method, $\mathcal{A}_M : \mathbb{R}^D \to \mathbb{C}^d$, which can approximately recover $\mathbf{x}$ using the magnitude measurements $\mathbf{b}$ that result from any member of a relatively large class of local correlation-based measurement matrices $M \in \mathbb{C}^{D \times d}$.

Phase retrieval problems of this form arise naturally in many crystallography and optics applications (see, e.g., [52, 38, 27, 37]). As an illustrative example, Figure 1 presents a typical imaging setup for a popular molecular imaging modality known as ptychography [45]. The figure shows a beam of electromagnetic radiation illuminating a small region of a test specimen. Under certain conditions (for example, if the wavelength of the incident radiation is of the same order as the atomic features in the specimen), the resulting diffraction pattern contains information about the atomic structure of the region being imaged. Therefore, by successively imaging shifts of the specimen, one might hope to deduce the complete atomic structure from such measurements. Unfortunately, it can be shown that the diffraction pattern measured by the detector corresponds to the (squared) magnitude of the Fourier transform of the specimen being imaged, meaning, among other things, that all phase information is lost. This is characteristic of many molecular imaging methods, where, either due to the underlying physics or due to instrumentation challenges, the detectors are not capable of capturing phase information. Therefore, a phase retrieval problem needs to be solved in order to recover the

underlying atomic structure. For ptychography applications, in particular, and restricting our attention to one dimension, the acquired measurements are of the form

$$b_\ell(\omega) = \left| \mathcal{F} \left[\widetilde{m} \ S_\ell x \right] (\omega) \right|^2,$$

where x is the specimen of interest, $\widetilde{m}$ is a localized illumination or window function (which depends on the optical setup), S_ℓ denotes a shift/translation operator, and $\mathcal{F}$ denotes the Fourier transform. Note that the Fourier intensity measurement $b_\ell(\omega)$ corresponds to a specific translate, or shift, ℓ of the unknown specimen x. Furthermore, we may assume that $\text{supp}(\widetilde{m}) \subset \text{supp}(x)$ since the imaging field of view is typically much smaller than the support of the specimen.

Discretizing the problem on a uniform grid, we obtain

$$(b_\omega)_\ell = \left| \sum_{k=0}^{d-1} \mathbb{e}^{\frac{-2\pi\mathbb{i}\,\omega k}{d}} \left(\widetilde{m}_k \, x_{k+\ell} \right) \right|^2, \quad \omega, \ell \in \{0, 1, \ldots, d-1\},$$

where $\mathbf{x}, \widetilde{\mathbf{m}} \in \mathbb{C}^d$ and $(b_\omega)_\ell \in \mathbb{R}$ is the analogous discrete measurement corresponding to a (circular) ℓ-shift of the unknown vector $\mathbf{x}$. Due to the finite field of view of the imaging system, we may also assume that $\widetilde{m}_k = 0, k > \delta$, where $\delta \in \mathbb{Z}^+$ and $\delta < d$. Hence, we may write[1]

$$(2) \qquad (b_\omega)_\ell = \left| \sum_{k=0}^{\delta-1} \left(\mathbb{e}^{\frac{-2\pi\mathbb{i}\,\omega k}{d}} \widetilde{m}_k \right) x_{k+\ell} \right|^2 = \left| \sum_{k=0}^{\delta-1} \overline{(\mathbf{m}_\omega)}_k \, x_{k+\ell} \right|^2, \quad (\mathbf{m}_\omega)_k := \mathbb{e}^{\frac{2\pi\mathbb{i}\,\omega k}{d}} \overline{\widetilde{m}}_k.$$

The imaging process thus involves collecting measurements $\{(b_\omega)_\ell\}$ for various Fourier frequencies and overlapping image space shifts $(\omega, \ell) \subseteq \mathbb{Z}^2 \cap [0, d-1]^2$. Here, in (2), and throughout the remainder of the paper, it is important to keep in mind that the parameter δ always corresponds to the width of each window/mask $\mathbf{m}_\omega$. The fact that this support size, δ, is significantly smaller that d is what leads to the locality of our proposed measurements.

As we will see in section 2, our proposed measurement constructions are identical to (2) for a specific collection of Fourier frequencies and spatial shifts. We also note that (2) can be interpreted as squared magnitude *correlation* measurements of the unknown vector $\mathbf{x}$ with the *local* (since $\delta \ll d$) masks $\mathbf{m}_\omega$. In particular,

$$(3) \qquad \mathbf{b}_\omega = \left| \text{corr}(\mathbf{m}_\omega, \mathbf{x}) \right|^2,$$

where $\mathbf{b}_\omega \in \mathbb{R}^d$ denotes diffraction measurements corresponding to frequency ω and all possible image space shifts of the unknown vector $\mathbf{x}$.

1.1. Survey of previous work. Given the importance of phase retrieval in crystallographic and optical imaging methods, there is a rich history of research on this topic by scientists

---

[1]All indexing is considered mod d; $\overline{x}$ denotes the complex conjugate of x.

and practitioners across diverse fields. With regard to computational algorithms specifically, most popular methods in use today can trace their origins back to the alternating projection algorithms developed in the 1970s by Gerchberg and Saxton [23] and Fienup [20]. These are iterative formulations which work by alternately requiring that the following two sets of constraints be satisfied:

    (i) (in data space) the current iterate has the same magnitude as that of the measured data, and

   (ii) (in image space) the current iterate satisfies certain problem-specific constraints such as positivity or compact support.

These algorithms are conceptually simple, efficient to implement (FFT time for certain measurement constructions) and popular among the practitioners, despite the lack of a rigorous mathematical understanding of their properties or global recovery guarantees. Indeed, given the nonconvex nature of the phase retrieval problem, the lack of global convergence results for these methods is not surprising. The interested reader is referred to [36, 19] for some recent developments and a review of this family of methods, while [44, 46, 35] contain some specific applications of alternating projection algorithms to ptychographic imaging.

More recently, there have been significant efforts devoted towards developing phase retrieval algorithms which are (i) computationally efficient, (ii) robust to measurement noise, and (iii) theoretically guaranteed to reconstruct a given vector up to a global phase error using a near-minimal number of magnitude measurements. For example, it has been shown that robust phase retrieval is possible with $D = \mathcal{O}(d)$ magnitude measurements by solving a semidefinite programming relaxation (*PhaseLift*) of a rank-1 matrix recovery problem [14, 10]. This allows polynomial-time convex optimization methods to be used for phase retrieval. Furthermore, the runtimes of these convexity-based methods can be reduced with the use of $\mathcal{O}(d \log d)$ magnitude measurements [17]. Other phase retrieval approaches include the use of spectral recovery methods together with magnitude measurement ensembles inspired by expander graphs [2]. These methods allow the recovery of $\mathbf{x}$ up to a global phase factor using $\mathcal{O}(d)$ noiseless magnitude measurements or $\mathcal{O}(d \log d)$ noisy magnitude measurements, and run in $\Omega(d^2)$-time in general.[2] Finally, the recently proposed *Wirtinger flow* algorithm [12] and *truncated Wirtinger flow* (TWF) algorithm [15] employ stochastic gradient descent schemes with special eigenvector-based initialization methods to recover $\mathbf{x}$. The TWF algorithm, for example, recovers $\mathbf{x}$ using $\mathcal{O}(d)$ magnitude measurements robustly with computational complexity[3] $\mathcal{O}(Dd \log 1/\epsilon)$, where ϵ is an accuracy parameter. All of these approaches utilize magnitude measurements $|M\mathbf{x}|^2$ resulting from either (i) Gaussian random matrices M, (ii) random masked Fourier-based constructions known as coded diffraction patterns (CDPs) [26], or (iii) unbalanced expander graph constructions, in order to prove their recovery guarantees.

1.2. Main result. In this paper we demonstrate that a relatively general class of local correlation-based magnitude measurements allow for phase retrieval in just $\mathcal{O}(d \log^c d)$-time.[4]

---

[2]Their runtime complexity is dominated by the time required to solve an overdetermined linear system.

[3]The *Wirtinger flow* algorithms are known to empirically work with random masked measurement constructions known as CDPs [11, 15] in essentially FFT time, although no robust recovery guarantees are available for such measurements.

[4]Herein c is a fixed absolute constant.

In particular, we construct a well-conditioned set of Fourier-based measurements which are theoretically guaranteed to allow for the phase retrieval of a given vector $\mathbf{x} \in \mathbb{C}^d$ with high probability in $\mathcal{O}(d \log^4 d)$-time. These measurements are of particular interest given that they are closely related to, e.g., ptychography [45]. In particular, we prove the following theorem.

Main result (fast phase retrieval from correlation-based measurements). Let $\mathbf{x} \in \mathbb{C}^d$ with d sufficiently large have $\|\mathbf{x}\|_2^2 \geq C \ (d \ln d)^2 \ln^3(\ln d) \ \|\mathbf{n}\|_2$.[5] Then, one can select a random measurement matrix $M \in \mathbb{C}^{D \times d}$ such that the following holds with probability at least $1 - \frac{1}{C' \cdot \ln^2(d) \cdot \ln^3(\ln d)}$.[6] The proposed phase retrieval Algorithm 1 (BlockPR) will recover an $\tilde{\mathbf{x}} \in \mathbb{C}^d$ with

$$\min_{\theta \in [0, 2\pi)} \left\| \mathbf{x} - \mathbb{e}^{\mathbb{i}\theta} \tilde{\mathbf{x}} \right\|_2^2 \leq C'' (d \ln d)^2 \ln^3(\ln d) \|\mathbf{n}\|_2$$

when given arbitrarily noisy input measurements $\mathbf{b} = |M\mathbf{x}|^2 + \mathbf{n} \in \mathbb{R}^D$ as per (1). Here D can be chosen to be $\mathcal{O}(d \cdot \ln^2(d) \cdot \ln^3(\ln d))$. Furthermore, the proposed Algorithm 1 will run in $\mathcal{O}(d \cdot \ln^3(d) \cdot \ln^3(\ln d))$-time.

To the best of our knowledge, this result provides the best existing error guarantee for correlation-based measurements. Moreover, this result also guarantees exact recovery, up to a global phase multiple, of $\mathbf{x}$ in the noiseless setting (i.e., when $\mathbf{n} = \mathbf{0}$). Further, for a particular class of flat[7] vectors $\mathbf{x}$, M can be chosen to be a *deterministic* matrix arising from local correlation-based measurements, and $D = 3d$ measurements suffice for recovery in the noiseless setting.

Numerical experiments both verify the speed and accuracy of the proposed phase retrieval approach, as well as indicate that the approach is highly robust to measurement noise. Additionally, after establishing and analyzing our general phase retrieval method, we then utilize it in order to establish the first known linear-in-sparsity compressive phase retrieval method capable of recovering s-sparse vectors $\mathbf{x}$ (up to an unknown phase factor) in only $\mathcal{O}(s \log^c d)$-time.

1.2.1. An overview of the proposed approach for local correlation-based measurements.
The proposed phase retrieval approach works in two stages in the ptychographic setting: during the first lifting (see, e.g., [14, 10]) stage, the quadratic magnitude measurements (3) are viewed as linear measurements of $(2\delta - 1)d$ new lifted variables, the entrywise products $x_i \overline{x_j}$ for $|j - i| < \delta \ll d$.[8] Given the correlation structure of the original measurements (3), the resulting lifted linear system in the new lifted unknowns $x_i \overline{x_j}$ is also highly

---

[5]Herein $C, C', C'' \in \mathbb{R}^+$ are all fixed and absolute constants.

[6]Note that M is selected independently of both $\mathbf{x}$ and $\mathbf{n}$. Furthermore, the nonuniform probability of success can be boosted to $1 - p$ for any $p \in (0, 1)$ at the expense of introducing additional logarithmic factors of $1/p$ into the runtime and measurement bounds. See section 4.3 for additional details.

[7]Here, "flat" simply means that there are no long strings of consecutive entries of $\mathbf{x}$ all of whose magnitudes are less than $\frac{\|\mathbf{x}\|_2}{2\sqrt{d}}$. See section 4.2 for additional details.

[8]Here, $i - j \in \{-d/2, \ldots, -1, 0, 1, \ldots, d/2\}$ is always considered modulo d. Note that these particular $(2\delta - 1)d$ entrywise products are exactly all those that appear when the squared magnitude measurements (2) are multiplied out for all shifts ℓ (treating $\mathbf{x}$ as periodic).

structured. In particular, the resulting coefficient matrix of the lifted linear system turns out to be block circulant [50], which both allows explicit condition number bounds to be derived for particular choices of windows $\mathbf{m}_\omega$, and also allows its fast numerical inversion using Fourier transforms. Thus, we are generally always able to (rapidly) invert the lifted linear system in order to solve for all local entrywise products $x_i \overline{x_j}$ with $|j - i| < \delta$. As an immediate consequence, the magnitude of each entry of $\mathbf{x}$ can also be estimated using, e.g., the products $x_j \overline{x_j} = |x_j|^2$.

During the second angular synchronization (see, e.g., [49]) stage of the algorithm, the local entrywise products $x_i \overline{x_j}$, $i \neq j$, are used in order to estimate the relative phases of each individual entry of $\mathbf{x}$. This is done by noting that the phase of each lifted variable $x_i \overline{x_j}$, $\arg(x_i \overline{x_j})$, provides an estimate of the phase angle difference between the corresponding entries' phases $\arg(x_i)$ and $\arg(x_j)$. Hence, the relative phases between all pairs of entries of $\mathbf{x}$ can be approximated by adding these local relative phase differences, provided by $\arg(x_i \overline{x_j})$ for $|j - i| < \delta$, together in appropriate (telescoping) sums. The only potential difficulty with this simple approach occurs when either x_i or x_j happens to be zero (or, more generally, very small in magnitude). In this case the phase difference given by $\arg(x_i \overline{x_j})$ is unreliable/unusable. In the worst possible case, when, e.g., $\mathbf{x}$ contains $\geq \delta$ consecutive zero entries in a row, the available local phase differences cannot span the corresponding gap in reliable phase difference information, and signal recovery (up to a single global phase ambiguity) becomes impossible. Much of the analysis of this stage focuses on categorizing and circumventing this fundamental difficulty.

Combining both stages above allows one to obtain deterministic algorithms for recovering all $\mathbf{x}$ that do not contain any long contiguous strings of $\geq \delta$ small entries with local correlation measurements. This result, which applies in the ptychographic setting, is ultimately developed in section 4.2 and stated in Theorem 5. Next, in section 4.3, it is shown that the worst case set of signals can still be recovered with high probability if one is allowed to precede one's correlation measurements with an initial randomized global masking operation. Combining this result with Theorem 5 yields the main theorem stated above. Finally, once phase retrieval of sparse signals has been allowed by the employment of randomized masking in section 4.3, it becomes possible to develop sparse phase retrieval algorithms with near-optimal runtimes in section 6.

The remainder of this paper is organized as follows: In section 2 we establish notation and discuss important preliminary results. Next, in section 3, we present our general phase retrieval algorithm and discuss it's runtime complexity. We then analyze our phase retrieval algorithm and prove recovery guarantees for specific types of Fourier-based measurement matrices in section 4. In section 5, we empirically evaluate the proposed phase retrieval method for speed and robustness. Finally, in section 6, we use our general phase retrieval algorithm in order to construct a sublinear-time compressive phase retrieval method which is guaranteed to recover sparse vectors (up to an unknown phase factor) in near-optimal time. Section 7 concludes with several suggestions for future work.

2. Preliminaries: Notation and setup. For any matrix $X \in \mathbb{C}^{D \times d}$ we will denote the jth column of X by $\mathbf{X}_j \in \mathbb{C}^D$. The conjugate transpose of a matrix $X \in \mathbb{R}^{D \times d}$ will be denoted by $X^* \in \mathbb{C}^{d \times D}$, and the singular values of any matrix $X \in \mathbb{C}^{D \times d}$ will always be ordered as

$\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_{\min(D,d)}(X) \geq 0$. Also, the condition number of the matrix X will be denoted by $\kappa(X) := \sigma_1(X)/\sigma_{\min(D,d)}(X)$. We will use the notation $[n] := \{1,\ldots,n\} \subset \mathbb{N}$ for any $n \in \mathbb{N}$. Finally, given any $\mathbf{x} \in \mathbb{C}^d$, the vector $\mathbf{x}_s^{\mathrm{opt}} \in \mathbb{C}^d$ will always denote an optimal s-sparse approximation to $\mathbf{x}$. That is, it preserves the s largest entries in magnitudes of $\mathbf{x}$ while setting the rest of the entries to 0. Note that $\mathbf{x}_s^{\mathrm{opt}} \in \mathbb{C}^d$ may not be unique as there can be ties for the sth largest entry in magnitude.

2.1. An illustrative example. Before we write down the setup for the general case, we present a simple but illustrative example which highlights the structure of our proposed measurements, and provides a general overview of the reconstruction algorithm. For simplicity, let us assume that we are given noiseless measurements in this section so that

$$\mathbf{b} = |M\mathbf{x}|^2$$

with $\mathbf{x} \in \mathbb{C}^4$, $\mathbf{b} \in \mathbb{R}^{12}$, and $M \in \mathbb{C}^{12\times4}$. Further, let us assume that the measurement matrix M has a block circulant structure of the form

$$M = \begin{pmatrix}
\overline{(\mathbf{m}_1)}_1 & \overline{(\mathbf{m}_1)}_2 & 0 & 0 \\
0 & \overline{(\mathbf{m}_1)}_1 & \overline{(\mathbf{m}_1)}_2 & 0 \\
0 & 0 & \overline{(\mathbf{m}_1)}_1 & \overline{(\mathbf{m}_1)}_2 \\
\overline{(\mathbf{m}_1)}_2 & 0 & 0 & \overline{(\mathbf{m}_1)}_1 \\
\hdashline
\overline{(\mathbf{m}_2)}_1 & \overline{(\mathbf{m}_2)}_2 & 0 & 0 \\
0 & \overline{(\mathbf{m}_2)}_1 & \overline{(\mathbf{m}_2)}_2 & 0 \\
0 & 0 & \overline{(\mathbf{m}_2)}_1 & \overline{(\mathbf{m}_2)}_2 \\
\overline{(\mathbf{m}_2)}_2 & 0 & 0 & \overline{(\mathbf{m}_2)}_1 \\
\hdashline
\overline{(\mathbf{m}_3)}_1 & \overline{(\mathbf{m}_3)}_2 & 0 & 0 \\
0 & \overline{(\mathbf{m}_3)}_1 & \overline{(\mathbf{m}_3)}_2 & 0 \\
0 & 0 & \overline{(\mathbf{m}_3)}_1 & \overline{(\mathbf{m}_3)}_2 \\
\overline{(\mathbf{m}_3)}_2 & 0 & 0 & \overline{(\mathbf{m}_3)}_1
\end{pmatrix} = \begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix},$$

where $M_1, M_2, M_3 \in \mathbb{C}^{4\times4}$ are circulant matrices and $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3 \in \mathbb{C}^4$ are *masks* or *window functions* with finite support. In particular,[9] $(\mathbf{m}_\ell)_i = 0$ for $i = 3,4$ and $\ell \in \{1,2,3\}$. The astute reader will note that this construction describes *correlation* measurements of the unknown vector $\mathbf{x}$ with *local* masks $\mathbf{m}_\ell$, i.e., $\mathbf{b} = |\mathrm{corr}(\mathbf{m}_\ell, \mathbf{x})|^2$, $\ell \in \{1,2,3\}$. Writing out the correlation sum explicitly and setting $\delta = 2$, we obtain

$$(4) \qquad (b_\ell)_i = \left| \sum_{k=1}^{\delta} \overline{(\mathbf{m}_\ell)}_k \cdot x_{i+k-1} \right|^2, \qquad (i,\ell) \in \{1,2,3,4\} \times \{1,2,3\}.$$

For a suitable choice of mask such as

$$(\mathbf{m}_\ell)_k = \begin{cases} \dfrac{\mathrm{e}^{-k/a}}{\sqrt[4]{2\delta-1}} \cdot \mathrm{e}^{\frac{2\pi\mathrm{i}\cdot(k-1)\cdot(\ell-1)}{2\delta-1}} & \text{if } k \leq \delta, \\ 0 & \text{if } k > \delta, \end{cases} \quad \text{with } a := \max\left\{4, \frac{\delta-1}{2}\right\},$$

---

[9] The notation $(m_\ell)_i$ denotes the ith entry of the ℓth mask.

we see that (4) is of the form of (2) in section 1 which described ptychographic measurements. These are exactly the measurements analyzed in section 4.1.

Having summarized our measurement construction, we now turn our attention to describing the reconstruction algorithm, which can be conceptually divided into the following two steps:

(i) Estimate (scaled) local phase differences of the form $\{\overline{x}_j x_k \mid j, k \in [d], |j-k \bmod d| < \delta\}$.

(ii) Recover the phases of each entry of $\mathbf{x}$, and consequently $\mathbf{x}$ itself, by using the estimates of the local phase differences from (i).

To obtain the phase differences in step (i) above, we start by rewriting the squared correlation measurements (4) as follows:

$$(b_\ell)_i = \left| \sum_{k=1}^{\delta} \overline{(\mathbf{m}_\ell)}_k \cdot x_{i+k-1} \right|^2 = \sum_{j,k=1}^{\delta} (\mathbf{m}_\ell)_j \, \overline{(\mathbf{m}_\ell)}_k \, \overline{x}_{i+j-1} \, x_{i+k-1} := \sum_{j,k=1}^{\delta} (\mathbf{m}_\ell)_{j,k} \, \overline{x}_{i+j-1} \, x_{i+k-1},$$

where we have used the notation $(\mathbf{m}_\ell)_{j,k} := (\mathbf{m}_\ell)_j \overline{(\mathbf{m}_\ell)}_k$. This is a linear system of equations for $D = (2\delta - 1)d = 12$ phase differences, $\mathbf{y} \in \mathbb{C}^{12}$, with

$$\mathbf{y} = \begin{bmatrix} |x_1|^2 & \overline{x}_1 x_2 & \overline{x}_2 x_1 & |x_2|^2 & \overline{x}_2 x_3 & \overline{x}_3 x_2 & |x_3|^2 & \overline{x}_3 x_4 & \overline{x}_4 x_3 & |x_4|^2 & \overline{x}_4 x_1 & \overline{x}_1 x_4 \end{bmatrix}^T.$$

By defining $\widetilde{\mathbf{b}} \in \mathbb{R}^{12}$ to be the interleaved vector of measurements (obtained by permuting the entries of $\mathbf{b}$)

$$\widetilde{\mathbf{b}} = \begin{bmatrix} (b_1)_1 & (b_2)_1 & (b_3)_1 & (b_1)_2 & (b_2)_2 & (b_3)_2 & (b_1)_3 & (b_2)_3 & (b_3)_3 & (b_1)_4 & (b_2)_4 & (b_3)_4 \end{bmatrix}^T,$$

we may write the resulting linear system as $M'\mathbf{y} = \widetilde{\mathbf{b}}$, where

$$M' = \left(\begin{array}{cccc:cc:cc:cc:cc}
(\mathbf{m_1})_{1,1} & (\mathbf{m_1})_{1,2} & (\mathbf{m_1})_{2,1} & (\mathbf{m_1})_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
(\mathbf{m_2})_{1,1} & (\mathbf{m_2})_{1,2} & (\mathbf{m_2})_{2,1} & (\mathbf{m_2})_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
(\mathbf{m_3})_{1,1} & (\mathbf{m_3})_{1,2} & (\mathbf{m_3})_{2,1} & (\mathbf{m_3})_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline
0 & 0 & 0 & (\mathbf{m_1})_{1,1} & (\mathbf{m_1})_{1,2} & (\mathbf{m_1})_{2,1} & (\mathbf{m_1})_{2,2} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & (\mathbf{m_2})_{1,1} & (\mathbf{m_2})_{1,2} & (\mathbf{m_2})_{2,1} & (\mathbf{m_2})_{2,2} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & (\mathbf{m_3})_{1,1} & (\mathbf{m_3})_{1,2} & (\mathbf{m_3})_{2,1} & (\mathbf{m_3})_{2,2} & 0 & 0 & 0 & 0 & 0 \\ \hdashline
0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{m_1})_{1,1} & (\mathbf{m_1})_{1,2} & (\mathbf{m_1})_{2,1} & (\mathbf{m_1})_{2,2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{m_2})_{1,1} & (\mathbf{m_2})_{1,2} & (\mathbf{m_2})_{2,1} & (\mathbf{m_2})_{2,2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{m_3})_{1,1} & (\mathbf{m_3})_{1,2} & (\mathbf{m_3})_{2,1} & (\mathbf{m_3})_{2,2} & 0 & 0 \\ \hdashline
(\mathbf{m_1})_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{m_1})_{1,1} & (\mathbf{m_1})_{1,2} & (\mathbf{m_1})_{2,1} \\
(\mathbf{m_2})_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{m_2})_{1,1} & (\mathbf{m_2})_{1,2} & (\mathbf{m_2})_{2,1} \\
(\mathbf{m_3})_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{m_3})_{1,1} & (\mathbf{m_3})_{1,2} & (\mathbf{m_3})_{2,1}
\end{array} \right).$$

We draw attention to the block circulant structure of M', composed of two blocks $M'_1, M'_2 \in \mathbb{C}^{3\times 3}$ highlighted using dashed lines. Much of the speed and elegance of the proposed algorithm arises from this block circulant structure: not only can such systems be inverted in essentially

linear time, they also lend themselves to easy analysis. As we will see in section 4.1, we can explicitly write out condition number bounds for M' resulting from certain classes of measurement masks.

Note that by solving this linear system, we automatically recover the magnitude of $\mathbf{x}$. In particular,

$$|x_1|^2 = y_1, \quad |x_2|^2 = y_4, \quad |x_3|^2 = y_7, \quad |x_4|^2 = y_{10}.$$

Additionally, after normalizing (to unit magnitude) the entries of $\mathbf{y}$, we also recover phase difference estimates,

$$\phi_{i,j} := \arg(x_j) - \arg(x_i), \quad i, j \in \{1, 2, 3, 4\}, |i - j \bmod 4| = 1.$$

For example, $\phi_{1,2} = \arg(x_2) - \arg(x_1) = \arg(y_2/|y_2|)$. We can now recover $\arg(\mathbf{x})$ as required by step (ii) of our two step recovery algorithm by using a greedy procedure.

Assume, without loss of generality, that $|x_1| \geq |x_i|$, $i \in \{2, 3, 4\}$. We start by setting $\arg(x_1) = 0.$[10] We may now set the phase of x_2 and x_4 using the estimated phase differences $\phi_{1,2}$ and $\phi_{1,4}$, respectively, i.e.,

$$\arg(x_2) = \arg(x_1) + \phi_{1,2}, \quad \arg(x_4) = \arg(x_1) + \phi_{1,4}.$$

Similarly, we next set $\arg(x_3) = \arg(x_2) + \phi_{2,3}$, thereby recovering all of the entries' unknown phases. We note that the computational cost of this procedure is essentially linear in the problem size d. Section 3 contains a more detailed description of the algorithm, while section 4 includes theoretical recovery guarantees.

2.2. General problem setup. Hereafter we will assume that our measurement matrix $M \in \mathbb{C}^{D \times d}$ has $D := (2\delta - 1)d$ rows corresponding to local correlation-based measurements, where $\delta \in \mathbb{N}$ represents the support size of the associated correlation masks. Furthermore, we will utilize the decomposition of M into its $(2\delta - 1)$ circulant blocks, $M_1, \ldots, M_{2\delta-1} \in \mathbb{C}^{d \times d}$, given by

$$(5) \qquad M = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_{2\delta-1} \end{pmatrix}.$$

Here each $M_l \in \mathbb{C}^{d \times d}$ will be both circulant, with

$$(6) \qquad (M_l)_{i,j} := (\mathbf{m}_l)_{(j-i) \bmod d+1}$$

for a mask $\mathbf{m}_l \in \mathbb{C}^d$, and banded, so that $(\mathbf{m}_l)_i = 0$ for all $i > \delta$ and $1 \leq l \leq 2\delta - 1.$[11]

---

[10]Recall that we can only recover $\mathbf{x}$ up to an unknown global phase factor which, in this case, will be the true phase of x_1.

[11]Every integer modulo d is considered to be an element of $\{0, \ldots, d-1\}$ in section 2.2. Furthermore, all indexes of vectors in $\mathbb{C}^d$ will be considered modulo d, $+ 1$, as per (6) for the remainder of section 2.2.

As a consequence of this structure, the squared magnitude measurements from the lth block, $|M_l\mathbf{x}|^2 \in \mathbb{R}^d$, can be rewritten as

$$(7) \qquad \left(|M_l\mathbf{x}|^2\right)_i = (M_l\mathbf{x})_i \, \overline{(M_l\mathbf{x})}_i = \sum_{j,k=1}^{\delta} (\mathbf{m}_l)_j \overline{(\mathbf{m}_l)}_k \, \overline{x}_{j+i-1} x_{k+i-1}.$$

Let $\mathbf{y} \in \mathbb{C}^D$ be defined by

$$(8) \qquad y_i := \overline{x}_{\left\lceil \frac{i+\delta-1}{2\delta-1} \right\rceil} x_{\left\lceil \frac{i+\delta-1}{2\delta-1} \right\rceil + ((i+\delta-2) \bmod (2\delta-1)) - \delta + 1}.$$

Furthermore, let $\mathbf{0}_\alpha \in \mathbb{R}^{1\times\alpha}$ be the row vector of α zeros for any given $\alpha \in \mathbb{N}$, and let $\tilde{\mathbf{m}}_{(l,j)} \in \mathbb{C}^{1\times\delta}$ be such that

$$(9) \qquad \left(\tilde{\mathbf{m}}_{(l,j)}\right)_k := (\mathbf{m}_l)_j \overline{(\mathbf{m}_l)}_k.$$

We can now reexpress $|M_l\mathbf{x}|^2 \in \mathbb{R}^d$ from (7) as $\tilde{M}_l\mathbf{y}$, where $\tilde{M}_l \in \mathbb{C}^{d\times D}$ is a $(2\delta-1)$-circulant matrix defined by

$$\begin{pmatrix} \tilde{\mathbf{m}}_{(l,1)} & \mathbf{0}_{\delta-2} & \tilde{\mathbf{m}}_{(l,2)} & \mathbf{0}_{\delta-2} & \tilde{\mathbf{m}}_{(l,3)} & \dots & \tilde{\mathbf{m}}_{(l,\delta)} & 0 & 0 & \dots & 0 \\ \mathbf{0}_{2\delta-1} & \tilde{\mathbf{m}}_{(l,1)} & \mathbf{0}_{\delta-2} & \tilde{\mathbf{m}}_{(l,2)} & \mathbf{0}_{\delta-2} & \tilde{\mathbf{m}}_{(l,3)} & \dots & \tilde{\mathbf{m}}_{(l,\delta)} & 0 & \dots & 0 \\ & & & & \ddots & & & & & & \\ \left(\tilde{\mathbf{m}}_{(l,2)}\right)_2 & \dots & \left(\tilde{\mathbf{m}}_{(l,2)}\right)_\delta & \mathbf{0}_{\delta-2} & \tilde{\mathbf{m}}_{(l,3)} & \mathbf{0}_{\delta-2} & \dots & 0 & \tilde{\mathbf{m}}_{(l,1)} & \mathbf{0}_{\delta-2} & \left(\tilde{\mathbf{m}}_{(l,2)}\right)_1 \end{pmatrix}.$$

Finally, after reordering the entries of $|M\mathbf{x}|^2$ via a permutation matrix $P \in \{0,1\}^{D\times D}$, we arrive at our final form

$$(10) \qquad P\,|M\mathbf{x}|^2 = M'\mathbf{y} = \begin{pmatrix} M'_1 & M'_2 & \dots & M'_\delta & 0 & 0 & \dots & 0 \\ 0 & M'_1 & M'_2 & \dots & M'_\delta & 0 & \dots & 0 \\ & & & \ddots & & & & \\ M'_2 & \dots & M'_\delta & 0 & \dots & 0 & \dots & M'_1 \end{pmatrix} \mathbf{y}.$$

Here $M' \in \mathbb{C}^{D\times D}$ is a block circulant matrix [50] whose blocks, $M'_1, \dots, M'_\delta \in \mathbb{C}^{(2\delta-1)\times(2\delta-1)}$, have entries

$$(11) \qquad (M'_l)_{i,j} := \begin{cases} (\mathbf{m}_i)_l \overline{(\mathbf{m}_i)}_{j+l-1} & \text{if } 1 \le j \le \delta - l + 1, \\ 0 & \text{if } \delta - l + 2 \le j \le 2\delta - l - 1, \\ (\mathbf{m}_i)_{l+1} \overline{(\mathbf{m}_i)}_{l+j-2\delta+1}, & \text{if } 2\delta - l \le j \le 2\delta - 1, \text{ and } l < \delta, \\ 0 & \text{if } j > 1, \text{ and } l = \delta. \end{cases}$$

Let I_α denote the $\alpha \times \alpha$ identity matrix. We now note that M' can be block diagonalized via the unitary block Fourier matrices $U_\alpha \in \mathbb{C}^{\alpha d\times\alpha d}$, with parameter $\alpha \in \mathbb{N}$, defined by

$$(12) \qquad U_\alpha := \frac{1}{\sqrt{d}} \begin{pmatrix} I_\alpha & I_\alpha & & \dots & I_\alpha \\ I_\alpha & I_\alpha e^{\frac{2\pi i}{d}} & & \dots & I_\alpha e^{\frac{2\pi i \cdot (d-1)}{d}} \\ & & \ddots & & \\ I_\alpha & I_\alpha e^{\frac{2\pi i \cdot (d-2)}{d}} & & \dots & I_\alpha e^{\frac{2\pi i \cdot (d-2) \cdot (d-1)}{d}} \\ I_\alpha & I_\alpha e^{\frac{2\pi i \cdot (d-1)}{d}} & & \dots & I_\alpha e^{\frac{2\pi i \cdot (d-1) \cdot (d-1)}{d}} \end{pmatrix}.$$

More precisely, one can see that we have

$$(13) \qquad U_{2\delta-1}^* \, M' \, U_{2\delta-1} = J := \begin{pmatrix} J_1 & 0 & 0 & \dots & 0 \\ 0 & J_2 & 0 & \dots & 0 \\ & & \ddots & & \\ 0 & 0 & 0 & J_{d-1} & 0 \\ 0 & 0 & 0 & 0 & J_d \end{pmatrix},$$

where $J \in \mathbb{C}^{D \times D}$ is block diagonal with blocks $J_1, \dots, J_d \in \mathbb{C}^{(2\delta-1) \times (2\delta-1)}$ given by

$$(14) \qquad J_k := \sum_{l=1}^{\delta} M_l' \cdot \mathbb{e}^{\frac{2\pi \mathrm{i} \cdot (k-1) \cdot (l-1)}{d}}.$$

Not so surprisingly, the fact that any block circulant matrix can be block diagonalized by block Fourier matrices will lead to more efficient computational techniques below.

2.3. Johnson–Lindenstrauss embeddings and restricted isometries. Below we will utilize results concerning *Johnson–Lindenstrauss embeddings* [32, 22, 1, 16, 3, 33] of a given finite set $\mathcal{S} \subset \mathbb{C}^d$ into $\mathbb{C}^m$ for $m < d$. These are defined as follows.

Definition 1. Let $\epsilon \in (0,1)$, and $\mathcal{S} \subset \mathbb{C}^d$ be finite. An $m \times d$ matrix A is a linear Johnson–Lindenstrauss embedding of $\mathcal{S}$ into $\mathbb{C}^m$ if

$$(1 - \epsilon) \| \mathbf{u} - \mathbf{v} \|_2^2 \leq \| A\mathbf{u} - A\mathbf{v} \|_2^2 \leq (1 + \epsilon) \| \mathbf{u} - \mathbf{v} \|_2^2$$

holds $\forall \mathbf{u}, \mathbf{v} \in \mathcal{S} \cup \{\mathbf{0}\}$. In this case we will say that A is a JL(m,d,ϵ)-embedding of $\mathcal{S}$ into $\mathbb{C}^m$.

Linear JL(m,d,ϵ)-embeddings are closely related to the *restricted isometry property* [8, 3, 21].

Definition 2. Let $s \in [d]$ and $\epsilon \in (0,1)$. The matrix $A \in \mathbb{C}^{m \times d}$ has the restricted isometry property if

$$(15) \qquad (1 - \epsilon) \| \mathbf{x} \|_2^2 \leq \| A\mathbf{x} \|_2^2 \leq (1 + \epsilon) \| \mathbf{x} \|_2^2$$

holds $\forall \mathbf{x} \in \mathbb{C}^d$ containing at most s nonzero coordinates. In this case we will say that A is RIP(s,ϵ).

In particular, the following theorem due to Krahmer and Ward [33, 21] demonstrates that a matrix with the restricted isometry property can be used to construct a Johnson–Lindenstrauss embedding matrix.

Theorem 1. Let $\mathcal{S} \subset \mathbb{C}^d$ be a finite point set with $|\mathcal{S}| = M$. For $\epsilon, p \in (0,1)$, let $A \in \mathbb{C}^{m \times d}$ be RIP($2s, \epsilon/C_1$) for some $s \geq C_2 \cdot \ln(4M/p)$.[12] Finally, let $B \in \{-1, 0, 1\}^{d \times d}$ be a random diagonal matrix with independent and identically distributed (i.i.d.) symmetric Bernoulli entries on its diagonal. Then, AB is a JL(m, d, ϵ)-embedding of $\mathcal{S}$ into $\mathbb{C}^m$ with probability at least $1 - p$.

---

[12] Here $C_1, C_2 \in (1, \infty)$ are both fixed absolute constants.

Below we will utilize Theorem 1 together with a result concerning the restricted isometry property for submatrices of a Fourier matrix. Let $F \in \mathbb{C}^{d \times d}$ be the unitary $d \times d$ discrete Fourier transform matrix. The *random sampling matrix*, $R' \in \mathbb{C}^{m \times d}$, for F is then

$$(16) \qquad R' := \sqrt{\frac{d}{m}} \cdot RF,$$

where $R \in \{0, 1\}^{m \times d}$ is a random matrix with exactly one nonzero entry per row (i.e., each entry's column position is drawn independently from $[d]$ uniformly at random with replacement). The following theorem is proven in [21].[13]

Theorem 2. *Let* $p \in (0, 1)$. *If the number of rows in the random sampling matrix* $R' \in \mathbb{C}^{m \times d}$ *satisfies both*

$$(17) \qquad \frac{m}{\ln(9m)} \geq C_3 \cdot \frac{s \ln^2(8s) \ln(8d)}{\epsilon^2}$$

and

$$(18) \qquad m \geq C_4 \cdot \frac{s \log(1/p)}{\epsilon^2},$$

then R' *will be* $RIP(2s, \epsilon/C_1)$ *with probability at least* $1 - p$.[14]

We are now prepared to present and analyze our phase retrieval method.

3. BlockPR: A fast phase retrieval algorithm. The proposed phase retrieval algorithm (*BlockPR*) works in two stages. In the first stage, the vector $\mathbf{y} \in \mathbb{C}^D$ from (8) of local entrywise products of $\mathbf{x} \in \mathbb{C}^d$ with its conjugate is approximated by solving the linear system (10). That is, we compute $(M')^{-1}P\mathbf{b}$, where $\mathbf{b} \in \mathbb{R}^D$ are our noisy local correlation-based measurements from (1), and M' and P are as in (10). This yields

$$(19) \qquad \tilde{\mathbf{y}} := (M')^{-1}P\mathbf{b} = (M')^{-1}P\,|M\mathbf{x}|^2 + (M')^{-1}P\mathbf{n} = \mathbf{y} + (M')^{-1}P\mathbf{n},$$

where $\mathbf{y}$ is as in (8).

Next, a greedy algorithm is used to recover the magnitudes and phases of each entry of $\mathbf{x}$ (up to a global phase factor) from our estimate of $\mathbf{y}$. To see how this works, note that $\mathbf{y}$ will contain all of the products $x_i \bar{x}_j$ for all $i, j \in [d]$ with $|(i - j) \bmod d| < \delta$.[15] As a result, the magnitude of each entry x_j can be estimated using the entry of $\mathbf{y}$ corresponding to $x_j \bar{x}_j = |x_j|^2$. Similarly, as long as both $x_j \bar{x}_j > 0$ and $x_i \bar{x}_i > 0$ hold, one can also compute the phase difference $\arg(x_i) - \arg(x_j)$ from $\arg(x_i \bar{x}_j)$. Thus, the phase of each x_i can be determined once $\arg(x_j)$ is established for a neighboring entry. Repeating this process allows one to determine a network of phase differences which all depend uniquely on the choice of a single entry's unknown phase. This entry's phase becomes the global phase factor $e^{i\theta}$ from (1.2). See Algorithm 1 for additional details.

---

[13]See Theorem 12.32 in Chapter 12.

[14]Here $C_3, C_4 \in (1, \infty)$ are both fixed absolute constants.

[15]In section 3 and below we always consider any integer modulo d to be in the set $\{-\lceil \frac{d}{2} \rceil + 1, \ldots, \lfloor \frac{d}{2} \rfloor\}$, for $\delta < \lfloor \frac{d}{2} \rfloor$.

Algorithm 1 *BlockPR.*

Input: Local Correlation Measurements $\mathbf{b} \in \mathbb{R}^D$ (Recall (1))
Output: $\tilde{\mathbf{x}} \in \mathbb{C}^d$ with $\tilde{\mathbf{x}} \approx \mathrm{e}^{-\mathrm{i}\theta}\mathbf{x}$ for some $\theta \in [0, 2\pi]$
 1: Compute $\tilde{\mathbf{y}} := (M')^{-1}P\mathbf{b}$ (see (19))
 2: Use Algorithm 2 with input $\tilde{\mathbf{y}} \in \mathbb{C}^D$ to compute $\tilde{\phi}_j \approx \phi_j := \arg(x_j)$ for all $j \in [d]$
 3: Set $\tilde{x}_j = \sqrt{|\tilde{y}_{j'}|}\mathrm{e}^{\mathrm{i}\tilde{\phi}_j} \; \forall j \in [d]$, where $\tilde{y}_{j'}$ (computed in line 1) is s.t. $\tilde{y}_{j'} = x_j\overline{x}_j + \left((M')^{-1}P\mathbf{n}\right)_{j'}$

Recall from (19) that we have access to

$$(20) \qquad \tilde{\mathbf{y}} = \mathbf{y} + \tilde{\mathbf{n}} \in \mathbb{C}^D,$$

where $\tilde{\mathbf{n}} := (M')^{-1}P\mathbf{n}$ results from the measurement noise, and $D = (2\delta - 1)d$. Thus, from (8) we have an index $k(i,j) \in [D]$ for all $i, j \in [d]$ with $|(i - j) \bmod d| < \delta$ such that

$$(21) \qquad \tilde{y}_{k(i,j)} = x_i\overline{x}_j + \tilde{n}_{k(i,j)}.$$

We will utilize the index function k for $\tilde{\mathbf{y}}$ defined by (21) in Algorithm 2.

It is important to note that Algorithm 1 assumes that the block circulant matrix M' arising from our choice of measurements, M, is invertible. As we shall see in sections 4 and 5, this is relatively easy to achieve. Similarly, Algorithm 2 implicitly assumes that $\tilde{\mathbf{y}}$ does not contain any strings of $\delta - 1$ consecutive zeros (or, more generally, $\delta - 1$ consecutive entires with "very small" magnitudes). This assumption will also be discussed in sections 4 and 5, and justified for measurements arising from arbitrary $\mathbf{x}$ by modifying the measurements M. For the time being, then, we are left free to consider the computational complexity of Algorithm 1.

3.1. Runtime analysis. We will begin our analysis of the runtime complexity of Algorithm 1 by considering the computation of $\tilde{\mathbf{y}} \in \mathbb{C}^D$ in line 1. Recalling section 2, we note that the permutation matrix P is based on a simple row reordering that clusters the first rows of $M_1, \ldots, M_{2\delta-1}$ into a contiguous block, the second rows of $M_1, \ldots, M_{2\delta-1}$ into a second contiguous block, etc. (see (5) and (6)). Thus, $P\mathbf{b}$ is simple to compute using only $\mathcal{O}(d \cdot \delta)$-operations. To finish calculating $\tilde{\mathbf{y}} = (M')^{-1}P\mathbf{b}$ we then use the decomposition of M' from (13) and compute $\tilde{\mathbf{y}} = U_{2\delta-1}J^{-1}U_{2\delta-1}^* P\mathbf{b}$.

Recalling the definition of $U_{2\delta-1}$ (12), one can see that both $U_{2\delta-1}$ and $U_{2\delta-1}^*$ have fast matrix-vector multiplies (i.e., because they can be computed by performing $2\delta - 1$ independent fast Fourier transforms on different subvectors of size d). Hence, matrix-vector multiplies with both of these matrices can be accomplished with $\mathcal{O}(\delta \cdot d\log d)$ operations. Finally, J is block diagonal with d blocks of size $(2\delta - 1) \times (2\delta - 1)$ (see (14)). Thus, J and J^{-1} can both be computed using $\mathcal{O}(d \cdot \delta^3)$ total operations. Putting everything together, we can now see that line 1 of Algorithm 1 requires only $\mathcal{O}(d \cdot \delta^3 + \delta \cdot d\log d)$ operations in general. Furthermore, these computations can easily benefit from parallelism due to the fact that the calculations above are all based on explicitly defined block decompositions.

The second line of Algorithm 1 calls Algorithm 2 whose runtime complexity is dominated by its main while-loop (lines 7 through 19). This loop will visit each entry of the input vector

---

Algorithm 2 Greedy angular synchronization.

Input: $\tilde{y}_{k(i,j)} = x_i \overline{x}_j + \tilde{n}_{k(i,j)} \quad \forall i, j \in [d]$ with $|(i - j) \mod d| < \delta$.

Output: Phase angles: $\tilde{\phi}_j \approx \phi_j := \arg(x_j)$ for all $j \in [d]$ (up to a global phase).

1: *% Estimate largest magnitude entry and set its phase to zero.*

$$a := \arg\max_j |\tilde{y}_{k(j,j)}| = \arg\max_j \left(\left| \, |x_j|^2 + \tilde{n}_{k(j,j)} \, \right| \right), \quad j = 0, \ldots, d - 1.$$

2: $\tilde{\phi}_a \leftarrow 0$ *% Note: We approximate unknown phases up to a global phase factor.*

3: *% Define a binary vector, phaseFlag $\in \{0,1\}^d$, to keep track of entries whose phase has*
    % already been set. This ensures that each phase is estimated once, and then not
    % changed again.

$$\text{phaseFlag}_i = \begin{cases} 0, & i = a, \\ 1, & \text{else.} \end{cases}$$

4: Initialize $j \leftarrow a$;

5: *% Estimate phases of all entries of* **x**

6: **while** $\displaystyle\sum_{i \in (j, j+\delta)} \text{phaseFlag}_i > 0$ **do**

7:    *% Estimate phases of $2\delta - 1$ entries nearest current x_j*

8:    **for** $i = 1 - \delta, 2 - \delta, \ldots, 0, \ldots, \delta - 1$ **do**

9:      *% Do not over-write previously estimated phases*

10:      **if** $\text{phaseFlag}_{j+i \mod d} == 1$ **then**

11:        *%Use current reference phase, $\tilde{\phi}_j$, and input phase difference estimates,*
        % $\arg\left(\tilde{y}_{k(j+i \mod d, \; j)}\right) \approx \arg\left(x_{j+i \mod d}\overline{x}_j\right)$ and $\arg\left(\tilde{y}_{k(j, \; j+i \mod d)}\right) \approx$
        % $\arg\left(x_j\overline{x}_{j+i \mod d}\right)$, to estimate the phase of the entry $x_{j+i \mod d}$.

$$\tilde{\phi}_{j+i \mod d} \leftarrow \tilde{\phi}_j + \frac{1}{2}\left(\arg\left(\tilde{y}_{k(j+i \mod d, \; j)}\right) - \arg\left(\tilde{y}_{k(j, \; j+i \mod d)}\right) \right);$$

12:        *% Remember that the phase of entry $x_{j+i \mod d}$ has now been estimated*

$$\text{phaseFlag}_{j+i \mod d} \leftarrow 0;$$

13:      **end if**

14:    **end for**

15:    *% Update reference entry to be the largest neighboring entry of the current x_j*

$$j \leftarrow \left(j + \arg\max_{0 < i < \delta} \left| \tilde{y}_{k(j+i \mod d, \; j+i \mod d)} \right| \right) \mod d;$$

16: **end while**

---

y at most a constant number of times. Hence, it requires $\mathcal{O}(\delta \cdot d)$ operations. Finally, the third line of Algorithm 1 uses only $\mathcal{O}(d)$ operations. Thus, the total runtime complexity of Algorithm 1 is $\mathcal{O}(d \cdot \delta^3 + \delta \cdot d \log d)$ in general.

4. Error analysis and recovery guarantees. In this section we analyze the performance of the proposed phase retrieval method (see Algorithm 1), and demonstrate measurement matrices which allow it to recover arbitrary vectors, up to an unknown phase factor, with high probability. Our analysis proceeds in two steps. First, in sections 4.1 and 4.2, we demonstrate the existence of a deterministic set of correlation-based measurements, $M \in \mathbb{C}^{D \times d}$, which allow Algorithm 1 to recover all relatively flat (i.e., nonsparse) vectors $\mathbf{x} \in \mathbb{C}^d$. Herein, "flat" simply means that there are no long strings of consecutive entries of $\mathbf{x}$ all of whose magnitudes are less than $\frac{\|\mathbf{x}\|_2}{2\sqrt{d}}$. The proposed measurements M are Fourier-like, roughly corresponding to a set of damped and windowed Fourier measurements of overlapping portions of $\mathbf{x}$. In addition to being well conditioned, these Fourier measurements also have fast inverse matrix-vector multiplies via (an additional usage of) the FFT. Hence, they confer additional computational advantages beyond those already enjoyed by our general block circulant measurement setup.

Next, in section 4.3, we extend our deterministic recovery guarantee for flat vectors to a nonuniform probabilistic recovery guarantee for arbitrary vectors. This is accomplished by right multiplying M with a concatenation of several Johnson–Lindenstrauss embedding matrices, each of which tends to "flatten out" vectors they are multiplied against. In particular, we construct a set of such matrices which are both (i) collectively unitary and (ii) rapidly invertible as a group via (yet another usage of) the FFT. The fact that this flattening matrix is unitary preserves the well-conditioned nature of our initial measurements, M. Furthermore, the fact that the flattening matrix enjoys a fast inverse matrix-vector multiply via the FFT allows us to maintain computational efficiency. Finally, the fact that the flattening matrix produces a flattened version of $\mathbf{x}$ with high probability allows us to apply our deterministic recovery guarantee for flat vectors to vectors which are not initially flat. The end result of this line of reasoning is the following recovery guarantee for noisy measurements.

Theorem 3. *Let* $\mathbf{x} \in \mathbb{C}^d$ *with d sufficiently large have* $\|\mathbf{x}\|_2^2 \geq C \ (d \ \ln d)^2 \ln^3(\ln d) \ \|\mathbf{n}\|_2.$[16] *Then, one can select a random measurement matrix* $\tilde{M} \in \mathbb{C}^{D \times d}$ *independently of both* $\mathbf{x}$ *and* $\mathbf{n}$ *such that the following holds with probability at least* $1 - \frac{1}{C' \cdot \ln^2(d) \cdot \ln^3(\ln d)}$*: Algorithm 1 will recover an* $\tilde{\mathbf{x}} \in \mathbb{C}^d$ *with*

$$(22) \qquad \min_{\theta \in [0,2\pi]} \left\| \mathbf{x} - e^{i\theta} \tilde{\mathbf{x}} \right\|_2^2 \leq C''(d \ \ln d)^2 \ln^3(\ln d) \|\mathbf{n}\|_2$$

when given arbitrarily noisy input measurements $\mathbf{b} = |M\mathbf{x}|^2 + \mathbf{n} \in \mathbb{R}^D$ *as per* (1). *Here D can be chosen to be* $\mathcal{O}(d \cdot \ln^2(d) \cdot \ln^3(\ln d))$. *Algorithm 1 runs in* $\mathcal{O}(d \cdot \ln^3(d) \cdot \ln^3(\ln d))$*-time.*

Note that the error bound in (22) is probably suboptimal due to, e.g., the appearance of the d^2 polylog d-term on its right-hand side. It is certainly the case, at least, that stronger error bounds are achievable using complex normal random measurements in combination

---

[16]Herein $C, C', C'' \in \mathbb{R}^+$ are all fixed and absolute constants.

with optimization techniques (see, e.g., [10, 29]). None the less, Theorem 3 provides the best existing error guarantee the authors are aware of for local correlation-based measurements, and computational experiments indicate that Algorithm 1 is highly robust to measurement noise in practice (see section 5). Furthermore, it is important to point out that Theorem 3 also guarantees exact recovery, up to a global phase multiple, of $\mathbf{x}$ by Algorithm 1 in the noiseless setting (i.e., when $\mathbf{n} = \mathbf{0}$). Finally, Algorithm 1 is fast, with a runtime complexity that is near linear in d. We are now ready to begin proving Theorem 3.

4.1. Well-conditioned measurements. In this section we develop a set of deterministic measurements $M \in \mathbb{C}^{D \times d}$ that lead to well-conditioned block circulant matrices $M' \in \mathbb{C}^{D \times D}$ in (10). To begin, we choose $a \in [4, \infty)$ and then set our local correlation masks to be

$$
(23) \qquad (\mathbf{m}_l)_i = \begin{cases} \dfrac{\mathrm{e}^{-i/a}}{\sqrt[4]{2\delta-1}} \cdot \mathrm{e}^{\frac{2\pi\mathrm{i}\cdot(i-1)\cdot(l-1)}{2\delta-1}} & \text{if } i \le \delta, \\ 0 & \text{if } i > \delta \end{cases}
$$

for $1 \le l \le 2\delta - 1$ and $1 \le i \le d$. This leads to blocks $M'_l \in \mathbb{C}^{(2\delta-1)\times(2\delta-1)}$ from (11) with entries given by

$$
(M'_l)_{i,j} := \begin{cases} (\mathbf{m}_i)_l\overline{(\mathbf{m}_i)}_{j+l-1} = \dfrac{\mathrm{e}^{-(2l+j-1)/a}}{\sqrt{2\delta-1}} \cdot \mathrm{e}^{-\frac{2\pi\mathrm{i}\cdot(i-1)\cdot(j-1)}{2\delta-1}} & \text{if } 1 \le j \le \delta-l+1, \\ 0 & \text{if } \delta-l+2 \le j \le 2\delta-l-1, \\ (\mathbf{m}_i)_{l+1}\overline{(\mathbf{m}_i)}_{l+j-2\delta+1} = \dfrac{\mathrm{e}^{-(2l+j-2(\delta-1))/a}}{\sqrt{2\delta-1}} \cdot \mathrm{e}^{-\frac{2\pi\mathrm{i}\cdot(i-1)\cdot(j-2\delta)}{2\delta-1}} & \text{if } 2\delta-l \le j \le 2\delta-1, \ l < \delta, \\ 0 & \text{if } j > 1, \text{ and } l = \delta. \end{cases}
$$

We will now begin to bound the condition number of this block circulant matrix, M', by block diagonalizing it via (13).

Considering the entries of each $J_k \in \mathbb{C}^{(2\delta-1)\times(2\delta-1)}$ from (14) results in two cases. First, suppose that $1 \le j \le \delta$. In this case one can see that

$$
(24) \qquad \begin{aligned} (J_k)_{i,j} &= \frac{\mathrm{e}^{(1-j)/a}}{\sqrt{2\delta-1}} \cdot \mathrm{e}^{\frac{-2\pi\mathrm{i}\cdot(i-1)\cdot(j-1)}{2\delta-1}} \cdot \sum_{l=1}^{\delta-j+1} \mathrm{e}^{-2l/a} \cdot \mathrm{e}^{\frac{2\pi\mathrm{i}\cdot(k-1)\cdot(l-1)}{d}} \\ &= \frac{\mathrm{e}^{-(j+1)/a}}{\sqrt{2\delta-1}} \cdot \mathrm{e}^{\frac{-2\pi\mathrm{i}\cdot(i-1)\cdot(j-1)}{2\delta-1}} \cdot \frac{1 - \mathrm{e}^{-2(\delta-j+1)/a} \cdot \mathrm{e}^{\frac{2\pi\mathrm{i}\cdot(k-1)\cdot(\delta-j+1)}{d}}}{1 - \mathrm{e}^{-2/a} \cdot \mathrm{e}^{\frac{2\pi\mathrm{i}\cdot(k-1)}{d}}}. \end{aligned}
$$

Second, suppose that $\delta + 1 \le j \le 2\delta - 1$. In this case one can see that

$$
\begin{aligned} (J_k)_{i,j} &= \frac{\mathrm{e}^{-(j-2(\delta-1))/a}}{\sqrt{2\delta-1}} \cdot \mathrm{e}^{\frac{-2\pi\mathrm{i}\cdot(i-1)\cdot(j-2\delta)}{2\delta-1}} \cdot \sum_{l=2\delta-j}^{\delta-1} \mathrm{e}^{-2l/a} \cdot \mathrm{e}^{\frac{2\pi\mathrm{i}\cdot(k-1)\cdot(l-1)}{d}} \\ &= \frac{\mathrm{e}^{-(2(\delta+1)-j)/a}}{\sqrt{2\delta-1}} \cdot \mathrm{e}^{\frac{-2\pi\mathrm{i}\cdot(i-1)\cdot(j-1)}{2\delta-1}} \cdot \mathrm{e}^{\frac{2\pi\mathrm{i}\cdot(k-1)(2\delta-j-1)}{d}} \cdot \frac{1 - \mathrm{e}^{-2(j-\delta)/a} \cdot \mathrm{e}^{\frac{2\pi\mathrm{i}\cdot(k-1)\cdot(j-\delta)}{d}}}{1 - \mathrm{e}^{-2/a} \cdot \mathrm{e}^{\frac{2\pi\mathrm{i}\cdot(k-1)}{d}}}. \end{aligned}
$$

Let $F_\alpha \in \mathbb{C}^{\alpha \times \alpha}$ be the unitary $\alpha \times \alpha$ discrete Fourier transform matrix. Defining

$$s_{k,j} := \begin{cases} \mathrm{e}^{-(j+1)/a} \cdot \dfrac{1 - \mathrm{e}^{-2(\delta-j+1)/a} \cdot \mathrm{e}^{2\pi\mathrm{i}\cdot(k-1)\cdot(\delta-j+1)/d}}{1 - \mathrm{e}^{-2/a} \cdot \mathrm{e}^{2\pi\mathrm{i}\cdot(k-1)/d}} & \text{if } 1 \leq j \leq \delta, \\[3mm] \mathrm{e}^{-(2(\delta+1)-j)/a} \cdot \mathrm{e}^{2\pi\mathrm{i}\cdot(k-1)(2\delta-j-1)/d} \cdot \dfrac{1 - \mathrm{e}^{-2(j-\delta)/a} \cdot \mathrm{e}^{2\pi\mathrm{i}\cdot(k-1)\cdot(j-\delta)/d}}{1 - \mathrm{e}^{-2/a} \cdot \mathrm{e}^{2\pi\mathrm{i}\cdot(k-1)/d}} & \text{if } \delta+1 \leq j \leq 2\delta-1, \end{cases}$$

we now have that

$$(25) \qquad J_k = F_{2\delta-1} \begin{pmatrix} s_{k,1} & 0 & \dots & 0 \\ 0 & s_{k,2} & 0 & \dots \\ 0 & 0 & \ddots & 0 \\ 0 & \dots & 0 & s_{k,2\delta-1} \end{pmatrix}.$$

Note that the condition number of J, and therefore of M', will be dictated by the singular values of these J_k matrices. Thus, we will continue by developing bounds for the singular values of each $J_k \in \mathbb{C}^{(2\delta-1)\times(2\delta-1)}$.

The fact that $F_{2\delta-1}$ is unitary implies that

$$(26) \qquad \min_{j\in[2\delta-1]} |s_{k,j}| \leq \sigma_{2\delta-1}(J_k) \leq \sigma_1(J_k) \leq \max_{j\in[2\delta-1]} |s_{k,j}|$$

for all $k \in [d]$. Thus, we will now devote ourselves to bounding the maximum and minimum values of $|s_{k,j}|$ from above and below, respectively, over all $k \in [d]$ and $j \in [2\delta - 1]$. These bounds will then collectively yield an upper bound on the condition number of our block circulant measurement matrix M'. The following simple technical lemmas will be useful.

Lemma 1. *Let $x \in [2, \infty)$. Then, $1 - \mathrm{e}^{-1/x} > \frac{2-\mathrm{e}^{1/x}}{x} \geq \frac{2-\sqrt{\mathrm{e}}}{x} > \frac{7}{20 \cdot x}$.*

Proof. Note that $1 - \mathrm{e}^{-1/x} = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{x^n n!} > \frac{1}{x} \cdot (2 - \sum_{n=0}^{\infty} \frac{1}{x^n(n+1)!}) > \frac{2-\mathrm{e}^{1/x}}{x}$. Furthermore, the numerator is a monotonically increasing function of x. ∎

Lemma 2. *Let $a, b, c \in \mathbb{R}^+$ and $f : \mathbb{R} \to \mathbb{R}$ below. Then,*
1. *$f(x) = b \cdot \mathrm{e}^{-x/a}(1 + c \cdot \mathrm{e}^{2x/a})$ has a unique global minimum at $x = -\frac{a}{2}\ln(c)$ and*
2. *$f(x) = b \cdot \mathrm{e}^{-x/a}(1 - c \cdot \mathrm{e}^{2x/a})$ is monotonically decreasing.*

Proof. In either case we have that $f'(x) = -\frac{b}{a} \cdot \mathrm{e}^{-x/a} \pm \frac{bc}{a} \cdot \mathrm{e}^{x/a}$ and $f''(x) = \frac{b}{a^2} \cdot \mathrm{e}^{-x/a} \pm \frac{bc}{a^2} \cdot \mathrm{e}^{x/a}$. For (1) we have a single critical point at $x = -\frac{a}{2}\ln(c)$, which is a global minimum since $f''(x) > 0 \; \forall x \in \mathbb{R}$. For (2) we have $f'(x) < 0$ for all $x \in \mathbb{R}$. ∎

Note that

$$(27) \quad |s_{k,j}| = \begin{cases} \mathrm{e}^{-(j+1)/a} \cdot \sqrt{\dfrac{1 + \mathrm{e}^{-4(\delta-j+1)/a} - 2\mathrm{e}^{-2(\delta-j+1)/a}\cos\left(2\pi \cdot [\delta-j+1]\cdot(k-1)/d\right)}{1 + \mathrm{e}^{-4/a} - 2\mathrm{e}^{-2/a}\cos\left(2\pi(k-1)/d\right)}} & \text{if } 1 \leq j \leq \delta, \\[3mm] \mathrm{e}^{-(2(\delta+1)-j)/a} \cdot \sqrt{\dfrac{1 + \mathrm{e}^{-4(j-\delta)/a} - 2\mathrm{e}^{-2(j-\delta)/a}\cos\left(2\pi \cdot [j-\delta]\cdot(k-1)/d\right)}{1 + \mathrm{e}^{-4/a} - 2\mathrm{e}^{-2/a}\cos\left(2\pi(k-1)/d\right)}} & \text{if } \delta+1 \leq j \leq 2\delta-1. \end{cases}$$

Fix $k \in [d]$. When $1 \leq j \leq \delta$ we have

$$(28) \qquad \max_{j\in[\delta]} |s_{k,j}| \leq \max_{j\in[\delta]} \left(\mathrm{e}^{-(j+1)/a} \cdot \frac{1 + \mathrm{e}^{-2(\delta+1-j)/a}}{1 - \mathrm{e}^{-2/a}} \right) \leq \frac{\mathrm{e}^{-2/a}(1 + \mathrm{e}^{-2\delta/a})}{1 - \mathrm{e}^{-2/a}},$$

where the second inequality follows from part one of Lemma 2. When $\delta + 1 \leq j \leq 2\delta - 1$ we have
(29)
$$\max_{j \in [2\delta-1] \backslash [\delta]} |s_{k,j}| \leq \max_{j \in [2\delta-1] \backslash [\delta]} \left(\mathbb{e}^{-(2(\delta+1)-j)/a} \cdot \frac{1 + \mathbb{e}^{-2(j-\delta)/a}}{1 - \mathbb{e}^{-2/a}} \right) \leq \frac{\mathbb{e}^{-3/a}(1 + \mathbb{e}^{-2(\delta-1)/a})}{1 - \mathbb{e}^{-2/a}},$$

where the second inequality again follows from part one of Lemma 2. Finally, combining (28) and (29) one can see that

$$(30) \quad \sigma_1(J_k) \leq \frac{\mathbb{e}^{-2/a}(1 + \mathbb{e}^{-2\delta/a})}{1 - \mathbb{e}^{-2/a}} \quad < \quad a \cdot \frac{\mathbb{e}^{-2/a}(1 + \mathbb{e}^{-2\delta/a})}{2(2 - \mathbb{e}^{2/a})} \quad < \quad a \cdot \frac{20\mathbb{e}^{-2/a}}{7} \quad < \quad 3a \cdot \mathbb{e}^{-2/a},$$

where the second inequality follows from Lemma 1 with $a \in [4, \infty)$.

Turning our attention to the lower bound, we note that part two of Lemma 2 implies that

$$(31) \qquad \min_{j \in [\delta]} |s_{k,j}| \geq \min_{j \in [\delta]} \left(\mathbb{e}^{-(j+1)/a} \cdot \frac{1 - \mathbb{e}^{-2(\delta+1-j)/a}}{1 + \mathbb{e}^{-2/a}} \right) \geq \frac{\mathbb{e}^{-(\delta+1)/a}(1 - \mathbb{e}^{-2/a})}{1 + \mathbb{e}^{-2/a}}.$$

Similarly, part two of Lemma 2 also ensures that

$$(32) \quad \min_{j \in [2\delta-1] \backslash [\delta]} |s_{k,j}| \geq \min_{j \in [2\delta-1] \backslash [\delta]} \left(\mathbb{e}^{-(2(\delta+1)-j)/a} \cdot \frac{1 - \mathbb{e}^{-2(j-\delta)/a}}{1 + \mathbb{e}^{-2/a}} \right) \geq \frac{\mathbb{e}^{-(\delta+1)/a}(1 - \mathbb{e}^{-2/a})}{1 + \mathbb{e}^{-2/a}}.$$

Combining (31) and (32) we see that

$$(33) \qquad \sigma_{2\delta-1}(J_k) \geq \frac{\mathbb{e}^{-(\delta+1)/a}(1 - \mathbb{e}^{-2/a})}{1 + \mathbb{e}^{-2/a}} \quad > \quad \frac{7}{20a} \cdot \mathbb{e}^{-(\delta+1)/a},$$

where the second inequality follows from Lemma 1 with $a \in [4, \infty)$. We are now equipped to prove the main theorem of this section.

Theorem 4. *Define $M' \in \mathbb{C}^{D \times D}$ via (23) with $a := \max\{4, \frac{\delta-1}{2}\}$. Then,*

$$\kappa(M') < \max \left\{ 144\mathbb{e}^2, \ \frac{9\mathbb{e}^2}{4} \cdot (\delta - 1)^2 \right\}.$$

Proof. We have from (30) and (33) that

$$(34) \qquad \kappa(M') = \frac{\sigma_1(M')}{\sigma_D(M')} = \frac{\sigma_1(J)}{\sigma_D(J)} \leq \frac{\max_{k \in [d]} \sigma_1(J_k)}{\min_{k \in [d]} \sigma_{2\delta-1}(J_k)} < 9a^2 \cdot \mathbb{e}^{(\delta-1)/a}.$$

Minimizing the rightmost upper bound as a function of a yields the stated result. ∎

Theorem 4 guarantees the existence of measurements which allow for the approximation of the phase difference vector $\mathbf{y} \in \mathbb{C}^D$ defined in (8). In the next three subsections we analyze the approximation of $\mathbf{x} \in \mathbb{C}^d$ from $\tilde{\mathbf{y}} \approx \mathbf{y}$ via the techniques discussed in section 3.

4.2. A recovery guarantee for flat vectors. As mentioned above, Algorithm 2 implicitly assumes that $\tilde{\mathbf{y}} \in \mathbb{C}^D$ does not contain any strings of $\delta - 1$ consecutive entries (mod d) all with very small magnitudes. In this section we will demonstrate that a general class of nonsparse vectors $\mathbf{x} \in \mathbb{C}^d$ lead to such $\tilde{\mathbf{y}}$ whenever noise levels are low enough. As a result, we will prove deterministic approximation guarantees for all sufficiently nonsparse vectors $\mathbf{x} \in \mathbb{C}^d$ in the high signal-to-noise setting. More specifically, we will utilize the following concrete characterization of flatness (i.e., nonsparsity) for $\mathbf{x}$ hereafter.

Definition 3. *Let $m \in [d]$. A vector $\mathbf{x} \in \mathbb{C}^d$ will be called m-flat if can be partitioned into at least $\lfloor \frac{d}{m} \rfloor$ blocks of consecutive entries such that*
  1. *every block contains either m or $m + 1$ neighboring entries of $\mathbf{x}$ and*
  2. *every such block of entries contains at least one entry whose magnitude is $\geq \frac{\|\mathbf{x}\|_2}{2\sqrt{d}}$.*

The following simple lemma proves that Algorithm 1 accurately estimates the magnitude of every entry of $\mathbf{x}$. Both it and the next lemma use the notation from (20) and (21), and implicitly assume the invertibility of M'.

Lemma 3. *Let $|\tilde{x}_j| = \sqrt{|\tilde{y}_{k(j,j)}|} = \sqrt{|x_j \overline{x}_j + \tilde{n}_{k(j,j)}|}$ be the estimate of $|x_j|$ utilized in line 3 of Algorithm 1, where $\tilde{\mathbf{n}} := (M')^{-1} P \mathbf{n}$. Then $\||x_j| - |\tilde{x}_j|\|^2 \leq 3\|\tilde{\mathbf{n}}\|_\infty$.*

Proof. Let $a := |x_j| \in \mathbb{R}^+$ and $\epsilon := \tilde{n}_{k(j,j)} \in \mathbb{C}$. We upper bound the right-hand side of

$$\||x_j| - |\tilde{x}_j|\|^2 = a^2 + |a^2 + \epsilon| - 2a\sqrt{|a^2 + \epsilon|} \leq 2a^2 + |\epsilon| - 2a^2 \sqrt{\left|1 + \frac{\epsilon}{a^2}\right|}$$

by considering two cases: if $a^2 \leq |\epsilon|$ we may bound the rightmost negative term by zero in order to obtain the desired result. If $a^2 > |\epsilon|$ then

$$-2a^2 \sqrt{\left|1 + \frac{\epsilon}{a^2}\right|} \leq -2a^2 \sqrt{1 - \frac{|\epsilon|}{a^2}} \leq -2a^2 \left(1 - \frac{|\epsilon|}{a^2}\right)$$

so that

$$\||x_j| - |\tilde{x}_j|\|^2 \leq 2a^2 + |\epsilon| - 2a^2 \left(1 - \frac{|\epsilon|}{a^2}\right) = 3|\epsilon|$$

again holds as desired. ∎

The next lemma proves that Algorithm 2 will accurately estimate the phases of all entries x_j whose magnitudes are sufficiently large relative to the noise level. This lemma requires that the vector $\mathbf{x}$ be m-flat for a sufficiently small m.

Lemma 4. *Suppose that $\mathbf{x} \in \mathbb{C}^d$ is $\lfloor \frac{\delta-3}{2} \rfloor$-flat with $d > 2$ and $\|\mathbf{x}\|_2^2 \geq 26 d^2 \|\tilde{\mathbf{n}}\|_\infty$, where $\tilde{\mathbf{n}} = (M')^{-1} P \mathbf{n}$ as in (20). Let $k \in [d]$ be such that $|x_k|^2 \geq \frac{5}{2} d \|\tilde{\mathbf{n}}\|_\infty$, and $\phi_a \in [0, 2\pi]$ be the true phase of the entry x_a of $\mathbf{x}$ chosen in line 2 of Algorithm 2. Then, Algorithm 2 will produce an estimate $\tilde{\phi}_k$ of the true phase of x_k, $\phi_k := \arg(x_k)$, satisfying*

$$\left| e^{i\tilde{\phi}_k} - e^{i(\phi_k - \phi_a)} \right|^2 \leq \frac{10 \, d^2 \, \|\tilde{\mathbf{n}}\|_\infty}{\|\mathbf{x}\|_2^2}.$$

Proof. Let $\alpha := \frac{\|\mathbf{x}\|_2}{d\sqrt{10\|\tilde{\mathbf{n}}\|_\infty}}$ and $\beta := \frac{\|\mathbf{x}\|_2}{2\sqrt{d}}$. Note that $\alpha > 1$, $|x_k| \geq \frac{\beta}{\alpha}$, $\|\tilde{\mathbf{n}}\|_\infty < \frac{1}{4}\frac{\beta^2}{d\alpha}$, and $|x_a| > \frac{\sqrt{3}}{2}\beta$ all hold. Furthermore, the fact that $\mathbf{x}$ is $\lfloor\frac{\delta-3}{2}\rfloor$-flat guarantees the existence of a sequence of entries $x_{b_1}, \ldots, x_{b_q}$ such that

1. $0 < |(b_{l+1} - b_l) \bmod \mathrm{d}| \leq \delta - 1$ for all $l \in [q-1]$,
2. $0 \leq \max\{|(b_1 - a) \bmod \mathrm{d}|, |(k - b_q) \bmod \mathrm{d}|\} \leq \delta - 1$, and
3. $|x_{b_l}| \geq \beta$ for all $l \in [q]$.

Thus, line 18 of Algorithm 2 is guaranteed to identify a sequence of entries $x_{j_1}, \ldots, x_{j_p}$ such that

1. $0 < |(j_{l+1} - j_l) \bmod \mathrm{d}| \leq \delta - 1$ for all $l \in [q-1]$,
2. $0 \leq \max\{|(j_1 - a) \bmod \mathrm{d}|, |(k - j_p) \bmod \mathrm{d}|\} \leq \delta - 1$, and
3. $|x_{j_l}| > \frac{\sqrt{3}}{2}\beta$ for all $l \in [q]$.

These entries contribute to the estimate $\tilde{\phi}_k$ of $\phi_k - \phi_a$ given by Algorithm 2,

$$(35) \quad \tilde{\phi}_k = \tilde{\phi}_{k,j_p} + \left(\sum_{l=1}^{p-1} \tilde{\phi}_{j_{l+1},j_l}\right) + \tilde{\phi}_{j_1,a} \approx (\phi_k - \phi_{j_p}) + \left(\sum_{l=1}^{p-1} \phi_{j_{l+1}} - \phi_{j_l}\right) + (\phi_{j_1} - \phi_a) = \phi_k - \phi_a,$$

where $\tilde{\phi}_{j_{l+1},j_l} := \frac{1}{2}(\arg(\tilde{y}_{k(j_{l+1},j_l)}) - \arg(\tilde{y}_{k(j_l,j_{l+1})}))$ is used as an estimate in line 13 of $\phi_{j_{l+1},j_l} := \arg(y_{k(j_{l+1},j_l)}) = \phi_{j_{l+1}} - \phi_{j_l}$.

To bound the approximation error in (35) we note that the law of sines implies that

$$\sin\left(|\tilde{\phi}_{j_{l+1},j_l} - \phi_{j_{l+1},j_l}|\right) \leq \left|\sin\left(\tilde{\phi}_{j_{l+1},j_l} - \phi_{j_{l+1},j_l}\right)\right| \leq \frac{\|\tilde{\mathbf{n}}\|_\infty}{|\tilde{y}_{l'}|},$$

where $\tilde{y}_{l'} = x_{j_{l+1}}\overline{x}_{j_l} + \tilde{n}_{l'}$ is, without loss of generality, the smaller of the two measurements (in magnitude) that contribute to the estimate $\tilde{\phi}_{j_{l+1},j_l}$. Using the facts from the previous paragraph concerning $|x_a|$, $|x_k|$, and $|x_{j_l}|$ for all $l \in [p]$, one can show that every measurement $\tilde{y}_{l'}$ that contributes to an estimate $\tilde{\phi}_{j_{l+1},j_l}$ used in (35) will have $|\tilde{y}_{l'}| > \frac{\beta^2}{2\alpha}$. Hence, we have that

$$\sin\left(|\tilde{\phi}_{j_{l+1},j_l} - \phi_{j_{l+1},j_l}|\right) \leq \frac{2\alpha}{\beta^2}\|\tilde{\mathbf{n}}\|_\infty < \frac{1}{4}$$

holds for all $l \in [p]$. As a consequence, we can infer that

$$\left|\tilde{\phi}_{j_{l+1},j_l} - (\phi_{j_{l+1}} - \phi_{j_l})\right| = \left|\tilde{\phi}_{j_{l+1},j_l} - \phi_{j_{l+1},j_l}\right| \leq \frac{5}{4}\sin\left(|\tilde{\phi}_{j_{l+1},j_l} - \phi_{j_{l+1},j_l}|\right) \leq \frac{5}{2}\frac{\alpha}{\beta^2}\|\tilde{\mathbf{n}}\|_\infty$$

also holds for all all $l \in [p]$. Combining this with (35) we learn that

$$\left|\tilde{\phi}_k - \phi_k + \phi_a\right| \leq d\frac{5}{2}\frac{\alpha}{\beta^2}\|\tilde{\mathbf{n}}\|_\infty < \frac{5}{8}.$$

This, in turn, implies that

$$\left|\mathrm{e}^{\mathrm{i}\tilde{\phi}_k} - \mathrm{e}^{\mathrm{i}(\phi_k - \phi_a)}\right| = \left|\mathrm{e}^{\mathrm{i}(\tilde{\phi}_k - \phi_k + \phi_a)} - 1\right| = 2\sin\left(\frac{1}{2}\left|\tilde{\phi}_k - \phi_k + \phi_a\right|\right)$$

$$\leq \left|\tilde{\phi}_k - \phi_k + \phi_a\right| \leq \frac{5d\alpha}{2\beta^2}\|\tilde{\mathbf{n}}\|_\infty$$

proving the lemma. ∎

We are now prepared to prove the main result of this section.

Theorem 5. *There exist fixed universal constants $C, C' \in \mathbb{R}^+$ such that the following holds: Let $M \in \mathbb{C}^{D \times d}$ be defined as in section 4.1, and suppose that $\mathbf{x} \in \mathbb{C}^d$ is $\lfloor \frac{\delta-3}{2} \rfloor$-flat with $d > 2$ and $\|\mathbf{x}\|_2^2 \geq C (\delta - 1)d^2 \|\mathbf{n}\|_2$. Then, Algorithm 1 is guaranteed to recover an $\tilde{\mathbf{x}} \in \mathbb{C}^d$ with*

$$(36) \qquad \min_{\theta \in [0,2\pi]} \left\| \mathbf{x} - e^{i\theta}\tilde{\mathbf{x}} \right\|_2^2 \leq C'd^2(\delta - 1)\|\mathbf{n}\|_2$$

when given arbitrarily noisy input measurements $\mathbf{b} = |M\mathbf{x}|^2 + \mathbf{n} \in \mathbb{R}^D$ as per (1). Furthermore, Algorithm 1 requires just $\mathcal{O}(\delta \cdot d \log d)$ operations for this choice of $M \in \mathbb{C}^{D \times d}$.

Proof. First, note that M' will be invertible by Theorem 4. Thus, we may set $\tilde{\mathbf{n}} = (M')^{-1}P\mathbf{n}$. Furthermore, the proof of Theorem 4 tells us that there exists an explicit universal constant $C'' \in \mathbb{R}^+$ such that

$$(37) \qquad C''(\delta - 1)\|\mathbf{n}\|_2 \geq \frac{\|\mathbf{n}\|_2}{\sigma_D (M')} \geq \|\tilde{\mathbf{n}}\|_2 \geq \|\tilde{\mathbf{n}}\|_\infty.$$

Setting $C = 26C''$ now allows us to verify that

$$\|\mathbf{x}\|_2^2 \geq 26C'' (\delta - 1)d^2 \|\mathbf{n}\|_2 \geq 26 \, d^2 \|\tilde{\mathbf{n}}\|_\infty.$$

Hence, we will be able to apply both Lemmas 3 and 4 as desired.

Let $\phi_a \in [0, 2\pi]$ be as in Lemma 4, and let $\mathbf{p}, \tilde{\mathbf{p}} \in \mathbb{C}^d$ be the vectors of phases of the entries of $e^{-i\phi_a}\mathbf{x}$ and $\tilde{\mathbf{x}}$, respectively, so that $p_j = e^{i(\phi_j - \phi_a)}$ and $\tilde{p}_j = e^{i\tilde{\phi}_j}$ hold for all $j \in [d]$. Similarly, recall that $|\mathbf{x}|, |\tilde{\mathbf{x}}| \in \mathbb{R}^d$ denote the vectors of magnitudes of the entries of $\mathbf{x}$ and $\tilde{\mathbf{x}}$, respectively, so that $|x|_j = |x_j|$ and $|\tilde{x}|_j = \sqrt{|\tilde{y}_{k(j,j)}|} = \sqrt{|x_j\overline{x}_j + \tilde{n}_{k(j,j)}|}$ hold for all $j \in [d]$ (here we are again using the notation from (20) and (21)). Thus, $e^{-i\phi_a}\mathbf{x} = |\mathbf{x}| \circ \mathbf{p}$ and $\tilde{\mathbf{x}} = |\tilde{\mathbf{x}}| \circ \tilde{\mathbf{p}}$ both hold, where $\circ$ denotes the entrywise (Hadamard) product.

To obtain (36) we bound

$$\|e^{-i\phi_a}\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq \||\mathbf{x}| \circ \mathbf{p} - |\mathbf{x}| \circ \tilde{\mathbf{p}}\|_2 + \||\mathbf{x}| \circ \tilde{\mathbf{p}} - |\tilde{\mathbf{x}}| \circ \tilde{\mathbf{p}}\|_2$$

$$= \sqrt{\sum_{j=1}^d |x_j|^2 \left| e^{i\tilde{\phi}_j} - e^{i(\phi_j - \phi_a)} \right|^2} + \sqrt{\sum_{j=1}^d \left| |x_j| - |\tilde{x}_j| \right|^2}$$

$$(38) \qquad \leq \sqrt{\sum_{j \in S} |x_j|^2 \left| e^{i\tilde{\phi}_j} - e^{i(\phi_j - \phi_a)} \right|^2}$$

$$+ \sqrt{\sum_{j \in [d] \setminus S} |x_j|^2 \left| e^{i\tilde{\phi}_j} - e^{i(\phi_j - \phi_a)} \right|^2} + \sqrt{3d\|\tilde{\mathbf{n}}\|_\infty},$$

where $S := \{k \in [d] \mid |x_k|^2 \geq \frac{5}{2} d \|\tilde{\mathbf{n}}\|_\infty\}$, and the last inequality results from an application of Lemma 3. Bounding the first two terms of (6) using Lemma 4 and the properties of S we

can see that

$$\|\mathbb{e}^{-\mathrm{i}\phi_a}\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq \sqrt{\sum_{j\in S}|x_j|^2\left(\frac{10\ d^2\ \|\tilde{\mathbf{n}}\|_\infty}{\|\mathbf{x}\|_2^2}\right)} + \sqrt{10d^2\|\tilde{\mathbf{n}}\|_\infty} + \sqrt{3d\|\tilde{\mathbf{n}}\|_\infty}$$

$$\leq 2d\sqrt{10\|\tilde{\mathbf{n}}\|_\infty} + \sqrt{3d\|\tilde{\mathbf{n}}\|_\infty} \leq 3d\sqrt{10\|\tilde{\mathbf{n}}\|_\infty}.$$

Using (37) now allows us to establish that

$$\|\mathbb{e}^{-\mathrm{i}\phi_a}\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq 3d\sqrt{10\|\tilde{\mathbf{n}}\|_\infty} \leq 3d\sqrt{10C''(\delta-1)\|\mathbf{n}\|_2}$$

which implies (36).

We finish by noting that the runtime complexity of Algorithm 1 simplifies to $\mathcal{O}(\delta \cdot d\log d)$ operations when using the measurements defined in section 4.1 because the matrix J also has a simple block-diagonal factorization in this case (recall (13) and (25) in light of section 3.1). ∎

The following corollary easily follows from Theorem 5.

Corollary 1. *Let $M \in \mathbb{C}^{D\times d}$ be defined as in section 4.1, and suppose that $\mathbf{x} \in \mathbb{C}^d$ is m-flat for some $m \leq \lfloor\frac{\delta-3}{2}\rfloor$. Then, Algorithm 1 will recover an $\tilde{\mathbf{x}} \in \mathbb{C}^d$ with $\tilde{\mathbf{x}} = \mathbb{e}^{\mathrm{i}\theta}\mathbf{x}$ for some $\theta \in [0, 2\pi]$ when given noiseless input measurements $\mathbf{b} = |M\mathbf{x}|^2 \in \mathbb{R}^D$. Furthermore, Algorithm 1 requires just $\mathcal{O}(\delta \cdot d\log d)$ operations in this case.*

Of course, not all vectors are m-flat for a suitably small value of m. We will generalize our results to arbitrary vectors $\mathbf{x}$ in the next section. This will be accomplished by showing that a well-chosen random unitary matrix W will have the property that $W\mathbf{x}$ is m-flat with high probability.

4.3. Flattening arbitrary vectors with high probability. Let $W \in \mathbb{C}^{d\times d}$ be the random unitary matrix

$$(39) \qquad\qquad\qquad W := PFB,$$

where $P \in \{0,1\}^{d\times d}$ is a permutation matrix selected uniformly at random from the set of all $d \times d$ permutation matrices, F is the unitary $d \times d$ discrete Fourier transform matrix, and $B \in \{-1, 0, 1\}^{d\times d}$ is a random diagonal matrix with i.i.d. symmetric Bernoulli entries on its diagonal. For any given $m \in [d]$, one can naturally partition W into $\lfloor\frac{d}{m}\rfloor$ blocks of contiguous rows, each of cardinality either m or $m+1$. This defines the $\lfloor\frac{d}{m}\rfloor$ submatrices of W, $W_1, \ldots, W_{d-m\lfloor\frac{d}{m}\rfloor} \in \mathbb{C}^{(m+1)\times d}$ and $W_{d-m\lfloor\frac{d}{m}\rfloor+1}, \ldots, W_{\lfloor\frac{d}{m}\rfloor} \in \mathbb{C}^{m\times d}$, by

$$(40) \qquad\qquad\qquad W = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_{\lfloor\frac{d}{m}\rfloor} \end{pmatrix}.$$

Note that each renormalized submatrix of W, $\sqrt{\frac{d}{m}} \cdot W_j$ for $j \in [\lfloor\frac{d}{m}\rfloor]$, is almost a random sampling matrix (16) times a random diagonal Bernoulli matrix. As a result, Theorems 1 and 2 suggest that each $\sqrt{\frac{d}{m}} \cdot W_j$ should behave like a JL(m,d,ϵ)-embedding of our signal $\mathbf{x}$ into $\mathbb{C}^m$ (or $\mathbb{C}^{m+1}$). If true, it would then be reasonable to expect that each block of m

consecutive entries of $W\mathbf{x}$ should have roughly the same ℓ_2-norm as one another. This, in turn, suggests that the random unitary matrix W should effectively flatten $\mathbf{x}$ with high probability, especially when m is small.

Of course, there are several small difficulties that must be addressed before the argument above can be made rigorous. First, the rows of F contributing to $\sqrt{\frac{d}{m}} \cdot W_j$ are effectively independently sampled uniformly *without replacement* from the set of all rows of F by our choice of P.[17] This means that Theorem 2 does not strictly apply in our situation since we cannot select any row of F more than once. Second, some care must be taken in order to select the smallest value of m possible in (40), since $W\mathbf{x}$ will become flatter, or less sparse, as m decreases. As a result, m will effectively provide a theoretical lower bound on the size of δ that one can utilize and still be guaranteed to accurately recover $W\mathbf{x}$ via our section 3 techniques (recall also section 4.2 above). We are now ready to begin proving our main result concerning W.

The following simple lemma will be used in order to help adapt Theorem 2 to the situation where the rows of F are sampled uniformly without replacement.

Lemma 5. *Let $m \in \mathbb{N}$ with $m \leq \sqrt{d}$. Independently draw $x_1, \ldots, x_m$ from $[d]$ uniformly at random with replacement. Then, $\mathbb{P}[|\{x_1, \ldots, x_m\}| = m] \geq 1/2$.*

Proof. A short induction argument establishes that

$$(41) \qquad \mathbb{P}\left[|\{x_1, \ldots, x_m\}| = m\right] = \prod_{j=1}^{m-1} \left(1 - \frac{j}{d}\right) \geq 1 - \sum_{j=1}^{m-1} \frac{j}{d} = 1 - \frac{m^2 - m}{2d}.$$

The result now follows easily via algebraic manipulation. ∎

The following corollary of Theorem 2 now demonstrates that a random sampling matrix R' formed by sampling a subset of rows of size m uniformly at random from F will still be RIP($2s,\epsilon/C_1$) with high probability.

Corollary 2. *Let $p \in (0,1)$. Form a random sampling matrix $R' \in \mathbb{C}^{m \times d}$ by independently sampling m rows from F uniformly without replacement. If the number of rows, m, satisfies both*

$$(42) \qquad \sqrt{d} \geq m \geq C_3 \cdot \frac{s \ln^2(8s) \ln(8d) \ln(9m)}{\epsilon^2}$$

and

$$(43) \qquad \sqrt{d} \geq m \geq C_4 \cdot \frac{s \log(2/p)}{\epsilon^2},$$

then R' will be RIP($2s,\epsilon/C_1$) with probability at least $1 - p$.

---

[17]Note that W being unitary helps us to be able to guarantee both exact recovery of $\mathbf{x}$ in the noiseless setting, and well-behaved approximation of $\mathbf{x}$ in the noisy setting. If we had chosen rows of F with replacement instead of without replacement in (39) we would not have a unitary (or even invertible) square matrix W with probability $\to 1/\mathrm{e}$ as $d \to \infty$. If one decides to make W rectangular instead of square simply in order to allow sampling from the rows of F with replacement, then many other small difficulties and inefficiencies result. Thus, we let P be a randomly selected permutation matrix instead of creating it by putting a one in each row independently at random.

Proof. Let $\mathcal{S} := \{x_1, \ldots, x_m\}$, where each $x_j \in [d]$ is selected independently and uniformly at random from $[d]$ (with replacement). Similarly, let $\mathcal{S}' \subset [d]$ be a subset of $[d]$ chosen uniformly at random from all subsets of $[d]$ with cardinality m (i.e., let $\mathcal{S}'$ contain m elements sampled independently and uniformly from $[d]$ without replacement). Furthermore, let E denote the event that the random sampling matrix whose rows from F are $x_1, \ldots, x_m$ is not RIP($2s, \epsilon/C_1$). Finally, let E' denote the event that the random sampling matrix whose rows from F are the elements of $\mathcal{S}'$ is not RIP($2s, \epsilon/C_1$). Applying Lemma 5 we can now see that

$$(44) \qquad \mathbb{P}\left[E\right] \geq \mathbb{P}\left[E \mid |\mathcal{S}| = m\right] \cdot \mathbb{P}\left[|\mathcal{S}| = m\right] = \mathbb{P}\left[E'\right] \cdot \mathbb{P}\left[|\mathcal{S}| = m\right] \geq \frac{1}{2} \cdot \mathbb{P}\left[E'\right].$$

The stated result now follows from Theorem 2. ∎

We are now ready to prove that W will flatten the signal $\mathbf{x} \in \mathbb{C}^d$ with high probability provided that m can be chosen appropriately. We have the following theorem.

Theorem 6. *Let $W \in \mathbb{C}^{d \times d}$ be formed as per (39) for $d \geq 8$. Then, $W\mathbf{x} \in \mathbb{C}^d$ will be m-flat with probability at least $1 - \frac{1}{m}$ provided that $\sqrt{d} \geq m + 1 \geq C_5 \cdot \ln^2(d) \cdot \ln^3(\ln d)$.*[18]

Proof. Our first goal will be to show that each $W_1, \ldots, W_{\lfloor \frac{d}{m} \rfloor}$ from (40) is a rescaled JL($m, d, 1/2$)-embedding of $\{\mathbf{x}\}$ into $\mathbb{C}^m$ (or $\mathbb{C}^{m+1}$). This will guarantee that each consecutive block of m (or $m + 1$) entries of $W\mathbf{x}$ has roughly the same ℓ_2-norm.

To achieve this goal we will apply Theorem 1 to each $\sqrt{\frac{d}{m}} \cdot W_1, \ldots, \sqrt{\frac{d}{m}} \cdot W_{\lfloor \frac{d}{m} \rfloor}$ in order to show that each one embeds $\{\mathbf{x}\}$ into $\mathbb{C}^m$ (or $\mathbb{C}^{m+1}$) with probability at least $1 - \frac{1}{2d}$. The union bound will then imply that $\{\mathbf{x}\}$ is embedded by all the $\sqrt{\frac{d}{m}} \cdot W_j$ with probability at least $1 - \frac{1}{2m}$. This argument will go through as long as each $\sqrt{\frac{d}{m}} \cdot W_1 B^{-1}, \ldots, \sqrt{\frac{d}{m}} \cdot W_{\lfloor \frac{d}{m} \rfloor} B^{-1}$ is RIP($2s, 1/2C_1$) for some $s \geq C_2 \cdot \ln(8d)$. Hence, we will now focus on determining the range of m which guarantees that all $\lfloor \frac{d}{m} \rfloor$ of these matrices are RIP($\lceil 2C_2 \cdot \ln(8d) \rceil, 1/2C_1$).

To demonstrate that each $\sqrt{\frac{d}{m}} \cdot W_j B^{-1}$ is RIP($\lceil 2C_2 \cdot \ln(8d) \rceil, 1/2C_1$) with probability at least $1 - \frac{1}{2d}$ one may apply Corollary 2 with m (or $m + 1$) chosen as above (assuming $d \geq 8$). Another application of the union bound then establishes that all of $\sqrt{\frac{d}{m}} \cdot W_1 B^{-1}, \ldots, \sqrt{\frac{d}{m}} \cdot W_{\lfloor \frac{d}{m} \rfloor} B^{-1}$ will be RIP($\lceil 2C_2 \cdot \ln(8d) \rceil, 1/2C_1$) with probability at least $1 - \frac{1}{2m}$. One final application of the union bound then establishes our first goal: all of $\sqrt{\frac{d}{m}} \cdot W_1, \ldots, \sqrt{\frac{d}{m}} \cdot W_{\lfloor \frac{d}{m} \rfloor}$ will be JL($m, d, 1/2$)-embeddings of $\{\mathbf{x}\}$ with probability at least $1 - \frac{1}{m}$.

To finish the proof, we now note that $W\mathbf{x}$ will be m-flat whenever all $\lfloor \frac{d}{m} \rfloor$ of the $\sqrt{\frac{d}{m}} \cdot W_j$ matrices are JL($m, d, 1/2$)-embeddings of $\{\mathbf{x}\}$. To see why, suppose that

$$\frac{1}{2}\| \mathbf{x} \|_2^2 \leq \frac{d}{m}\|W_j\mathbf{x}\|_2^2 \leq \frac{3}{2}\| \mathbf{x} \|_2^2.$$

This implies that $\frac{3m}{2d}\| \mathbf{x} \|_2^2 \geq \|W_j\mathbf{x}\|_2^2 \geq \frac{m}{2d}\|\mathbf{x}\|_2^2$, which can only happen if both of the following hold: (i) at least one entry of $W_j\mathbf{x}$ has magnitude at least $\frac{\|\mathbf{x}\|_2}{2\sqrt{d}} = \frac{\|W\mathbf{x}\|_2}{2\sqrt{d}}$ and (ii)

---

[18] Here $C_5 \in \mathbb{R}^+$ is a fixed absolute constant.

all entires of $W_j\mathbf{x}$ have magnitude less than $\sqrt{\frac{3m+3}{2d}}\|\mathbf{x}\|_2 = \sqrt{\frac{3m+3}{2d}}\|W\mathbf{x}\|_2$. This proves the theorem. ∎

Theorem 6 now allows us to alter our measurements so that we can recover arbitrary vectors. We are now ready to prove Theorem 3.

4.4. Proof of Theorem 3. We set our measurement matrix $\tilde{M} \in \mathbb{C}^{D \times d}$ to be $\tilde{M} := MW$, where $M \in \mathbb{C}^{D \times d}$ is defined as in section 4.1, and $W \in \mathbb{C}^{d \times d}$ is as defined as in (39). Theorem 6 guarantees that $W\mathbf{x}$ will be $m = \mathcal{O}(\ln^2(d) \cdot \ln^3(\ln d))$-flat with probability at least $1 - \frac{1}{C_5 \cdot \ln^2(d) \cdot \ln^3(\ln d)}$ provided that d is sufficiently large. Furthermore, if $W\mathbf{x}$ is m-flat and $\delta \geq 2m + 3$, then Theorem 5 guarantees that Algorithm 1 will recover an $\mathbf{x}' \in \mathbb{C}^d$ satisfying

$$(45) \qquad \min_{\theta \in [0,2\pi]} \left\| W\mathbf{x} - e^{i\theta}\mathbf{x}' \right\|_2 \leq C'd^2(\delta-1)\|\mathbf{n}\|_2$$

when given the noisy input measurements $\mathbf{b} = |MW\mathbf{x}|^2 + \mathbf{n} \in \mathbb{R}^D$. Hence, choosing $\delta = \mathcal{O}(\ln^2(d) \cdot \ln^3(\ln d))$ allows us to recover $\mathbf{x}' \approx W\left(e^{i\phi}\mathbf{x}\right)$ for some unknown phase $\phi \in [0, 2\pi]$ with probability at least $1 - \frac{1}{C_5 \cdot \ln^2(d) \cdot \ln^3(\ln d)}$.[19] Setting $\tilde{\mathbf{x}} = W^*\mathbf{x}'$ we can see that

$$\min_{\theta \in [0,2\pi]} \left\| \mathbf{x} - e^{i\theta}\tilde{\mathbf{x}} \right\|_2 = \min_{\theta \in [0,2\pi]} \left\| W^* \left(W\mathbf{x} - e^{i\theta}\mathbf{x}' \right) \right\|_2$$
$$= \min_{\theta \in [0,2\pi]} \left\| W\mathbf{x} - e^{i\theta}\mathbf{x}' \right\|_2 \leq C'd^2(\delta-1)\|\mathbf{n}\|_2$$

by (45) since W is always unitary.

Considering the runtime complexity, we note that $\mathbf{x}'$ can be obtained in $\mathcal{O}(\delta \cdot d \log d) = \mathcal{O}(d \cdot \ln^3(d) \cdot \ln^3(\ln d))$ operations by Theorem 5. Computing $W^*\mathbf{x}'$ can then be done in $\mathcal{O}(d \log d)$ operations via an inverse fast Fourier transform. The stated runtime complexity follows.

It is interesting to note that alternate constructions of flattening matrices, W, with fast inverse matrix-vector multiplies can also be created by using sparse Johnson–Lindenstrauss embedding matrices in the place of our Fourier-based matrices (see, e.g., [7]). Thus, one has several choices of matrices W to use in concert with a given block circulant measurement matrix M in principle.

5. Empirical evaluation. We now present numerical results demonstrating the efficiency and robustness of the *BlockPR* algorithm. We test our algorithm on i.i.d. zero-mean complex random Gaussian test signals. To test noise robustness, we add i.i.d. random Gaussian noise to the squared magnitude measurements at desired signal to noise ratios (SNRs). In particular, the noise vector $\mathbf{n} \in \mathbb{R}^D$ in (1) is chosen to be i.i.d. $\mathcal{N}(\mathbf{0}, \sigma^2\mathcal{I}_D)$. The variance σ^2 is chosen such that

$$\text{SNR (dB)} = 10 \log_{10} \left(\frac{\|M\mathbf{x}\|_2^2}{D\,\sigma^2} \right).$$

Errors in the recovered signal are also reported in dB with

$$\text{Error (dB)} = 10 \log_{10} \left(\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} \right),$$

---

[19]The probability estimate in Theorem 3 follows immediately with $C = C_5$.

(a) Execution time—phase retrieval from $D = 7d$ measurements.

(b) Execution time—phase retrieval from $D = \lceil 2d \log_2 d \rceil$ measurements.
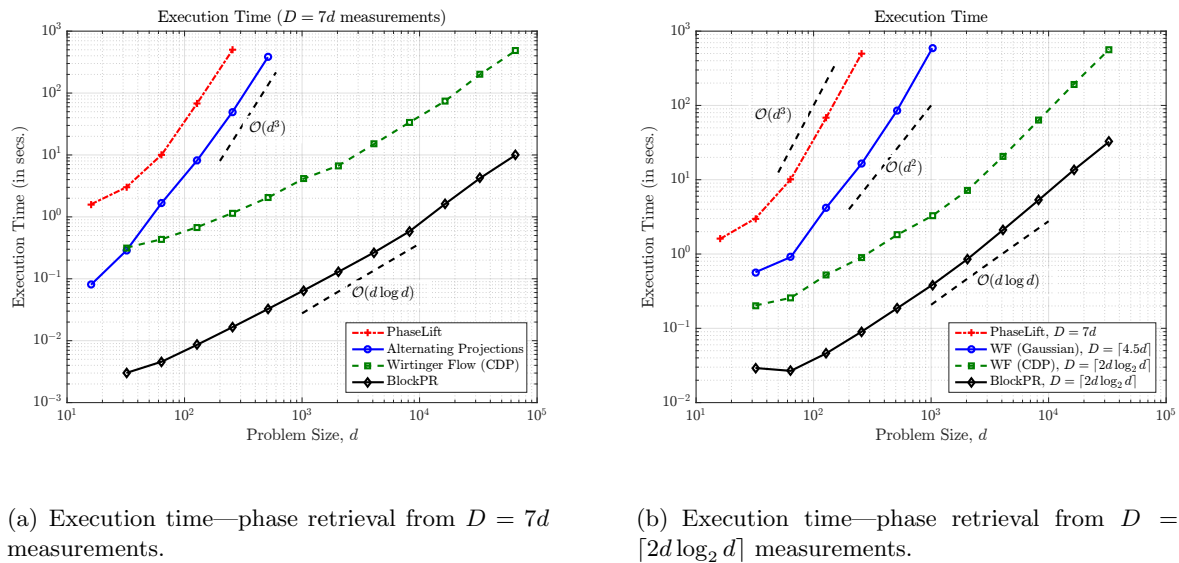
Figure 2. *Computational efficiency of the* BlockPR *phase retrieval algorithm (WF denotes the* Wirtinger flow *algorithm; parentheses are used to denote the type of measurements used).*

where $\tilde{\mathbf{x}}$ denotes the recovered signal. Matlab code used to generate the numerical results is freely available at [31].

We start by presenting numerical simulations demonstrating the efficiency of the *BlockPR* algorithm. In particular, we plot the execution time for solving the phase retrieval problem (averaged over 100 trials) from perfect (noiseless) measurements in Figure 2. Simulations were performed on a laptop computer with an Intel® Core™i3-2350M processor, 6 GB RAM and Matlab R2015a. For comparison, we also plot execution times for the *Gerchberg–Saxton* alternating projection algorithm, semidefinite-programming-based *PhaseLift* algorithm,[20] and the recently introduced *Wirtinger flow* algorithm. Simulation results with *PhaseLift* and the *Gerchberg–Saxton* alternating projection algorithm use random complex Gaussian measurements. We note that even though alternating projection algorithms are known to empirically work with certain FFT time measurement constructions, recovery guarantees are only available for random Gaussian measurements (see, for example, [42]). Simulation results for the *Wirtinger flow* algorithm are generated using either random complex Gaussian measurements or CDPs. We remark here that the motivating application typically determines the type of measurements used. For example, random Gaussian measurements or CDPs—such as those employed by *PhaseLift* and *Wirtinger flow*—are not directly applicable in local correlation-based applications, which are of primary interest in this discussion. Nevertheless, Figure 2 provides a useful comparison for evaluating the efficiency of different phase retrieval algorithms. Figure 2(a) plots the execution time for solving the phase retrieval problem using $7d$ measurements. We observe that *PhaseLift* and alternating projections have computational

---

[20]The PhaseLift algorithm was implemented as a trace-regularized least-squares problem using CVX [25, 24], a package for specifying and solving convex programs in Matlab.
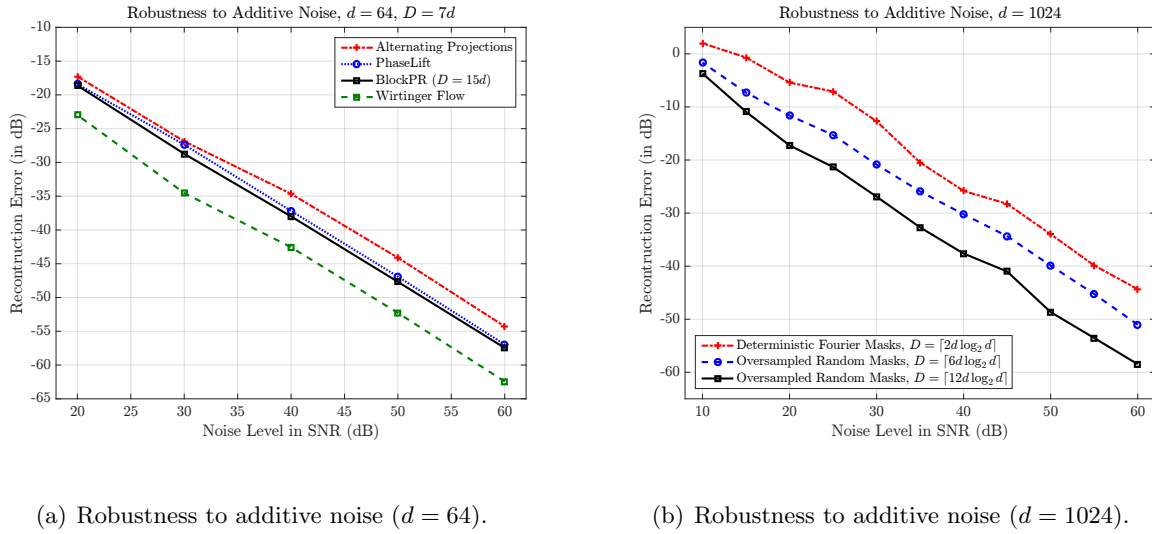
(a) Robustness to additive noise ($d = 64$).

(b) Robustness to additive noise ($d = 1024$).

Figure 3. *Robustness of the proposed* BlockPR *phase retrieval algorithm to additive noise.*

complexities which scale cubically with the problem dimension, thereby limiting their applicability to small-scale problems. On the other hand, both *BlockPR* and *Wirtinger flow* (with CDP measurements) have essentially FFT time computational complexities, with *BlockPR* observed to have a smaller constant than *Wirtinger flow*. Similar speedup is observed in Figure 2(b), where the number of measurements used by *BlockPR* and *Wirtinger flow* grows log linearly in the problem size d. For reference, execution times for phase retrieval from $\mathcal{O}(d)$ random complex Gaussian magnitude measurements using *PhaseLift* and *Wirtinger flow* are also provided. These plots confirm the significant speedup offered by the proposed *BlockPR* algorithm over other comparable methods and corroborate the efficiency of the proposed computational framework for solving large-scale phase retrieval problems.

We next demonstrate robustness of the proposed framework to additive noise. Figure 3(a) plots the reconstruction error in recovering a $d = 64$ complex random Gaussian signal at different SNRs, with each data point computed as the average of 100 trials.[21] We include reconstruction results using *Gerchberg–Saxton*, *PhaseLift*, and *Wirtinger flow* algorithms for comparison. In all cases, the algorithms recover the unknown signal $\mathbf{x}$ up to (or slightly better than) the added noise level. As opposed to the other algorithms in Figure 3(a), *BlockPR* uses *local* measurements as prescribed in (23), with each measurement only providing information about a corresponding *local* region of the underlying signal. Hence, we expect *BlockPR* to require more measurements than the other algorithms in Figure 3(a) to achieve the same noise robustness. Indeed, this is confirmed in the figure with *BlockPR* requiring roughly twice the number of measurements to achieve the same noise robustness. We note, however, that the number of additional measurements is typically a small constant irrespective of the problem size. Similar results are provided in Figure 3(b) for a larger problem size ($d = 1024$).

---

[21] A few iterations of the *Gerchberg–Saxton* alternating projection algorithm were used to postprocess the reconstructions.
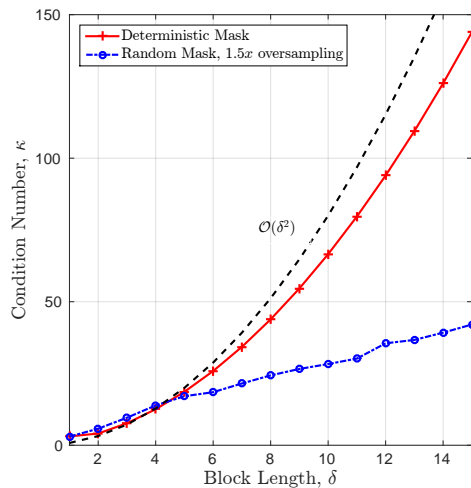
Figure 4. *Well-conditioned measurements—condition number as a function of the block length δ.*

Until now, the *deterministic* Fourier-based measurement masks of (23) have been utilized to generate simulation results. In Figure 3(b), we additionally provide robustness results using oversampled random local masks. Specifically, we choose the nonzero entries of each of the ℓ measurement masks $\mathbf{m}_\ell \in \mathbb{C}^d$ in (6) to be i.i.d. complex random Gaussian entries for $\ell = 1, 2, \ldots, \gamma \cdot (2\delta - 1)$, where $\gamma \geq 1$ is an oversampling factor. The dashed line in Figure 3(b) was generated using $\gamma = 1.5$, while the solid line was generated using $\gamma = 2$. In either case, the block length δ was chosen so as to achieve a total of $D = \lceil 6d \log_2 d \rceil$ and $D = \lceil 12d \log_2 d \rceil$ measurements, respectively. This plot confirms the graceful degradation of reconstruction error with added noise. Moreover, the block length δ and oversampling factor γ may be suitably chosen to achieve desired noise robustness.

Finally, Figure 4 plots the condition number of the system matrix used to solve for the phase differences (matrix M' in section 2). The figure plots the condition number as a function of the block length δ for $d = 64$.[22] It confirms that the condition number scales as a small multiple of δ^2. The figure also includes a plot of the condition number when using random masks at an oversampling factor of 1.5. Empirical simulations suggest that the use of oversampling can lead to essentially linear growth in the condition number $\kappa(M')$ with block length δ. Analyzing the performance of such an oversampled measurement construction and explicitly writing down associated condition number bounds would be an interesting avenue for future research.

We conclude this section by presenting representative numerical results highlighting the superior performance of the proposed *BlockPR* algorithm for *local* measurements. Specifically, we use the windowed Fourier-like local correlation measurements described in section 4.1 with the *BlockPR*, *PhaseLift*, and *Gerchberg–Saxton* alternating projection algorithms. We note that there are no theoretical recovery guarantees with *PhaseLift* and the *Gerchberg–Saxton* algorithm for such measurements; however, empirical simulations suggest that for sufficiently

---

[22]The condition number is independent of the problem dimension d and depends only on the block length δ.

large numbers of measurements, both algorithms recover the unknown signal. Indeed, this is illustrated in Figure 5(b), where the reconstruction error in recovering a $d = 64$ length complex vector is plotted as a function of the number of measurements required. For global (random Gaussian) measurements, it is known that *PhaseLift* recovers signals up to the noise level with about $6d$ measurements (see, for example, [14]). Yet, for the local measurements studied here, significantly larger numbers of measurements are necessary for successful phase retrieval by *PhaseLift* (and indeed the *Gerchberg–Saxton* alternating projection algorithm). Figure 5(b) shows that the proposed *BlockPR* algorithm compares well with *PhaseLift* and almost always outperforms alternating projections. We note that the *PhaseLift* error plots are only shown for up to $D = 15d$ measurements due to memory issues in implementing the algorithm for larger D with CVX on a laptop computer.

The performance of the proposed *BlockPR* algorithm is particularly noteworthy since it has an essentially FFT time computational cost as illustrated in Figure 5(a). Unsurprisingly, the computational cost of *PhaseLift* scales as $\mathcal{O}(d^3)$. Note that we had to use the Matlab software package TFOCS [4, 5] to implement *PhaseLift* for large problem sizes. TFOCS implements fast first-order conic solvers which trade off some accuracy for efficiency gains, thereby enabling the solution of large problems. We also note that the computational cost of the *Gerchberg–Saxton* alternating projections algorithm scales quadratically despite the use of FFT-based linear system solvers in each iteration. Numerical experiments suggest that the number of alternating projection iterations grows with the problem size, thereby resulting in the quadratic scaling.

For completeness, Figures 5(c) and 5(d) plot the reconstruction error versus added noise level using $D = 7d$ and $D = 15d$ measurements, respectively, for all three algorithms using local correlation measurements. In each plot, we recover a $d = 64$ length complex vector from i.i.d. Gaussian noise corrupted phaseless measurements. With $D = 15d$ measurements, both *BlockPR* and *PhaseLift* recover the unknown signal to the level of noise, while for $D = 7d$ measurements, there is a roughly 10 dB error in the reconstructed signal. Note the increased number of measurements necessary for *PhaseLift* to achieve the same error performance as in Figure 3(a)—this highlights the challenging nature of reconstructing signals from *local* phaseless measurements.

6. Extension: Sublinear-time phase retrieval for compressible signals. In this section we briefly focus on the compressive phase retrieval setting, (see, e.g., [43, 47, 34, 53, 18, 48]), where one aims to approximate a sparse or compressible $\mathbf{x} \in \mathbb{C}^d$ using fewer magnitude measurements than required for the recovery of general $\mathbf{x}$. It is known that robust compressive phase retrieval for s-sparse vectors is possible using only $\mathcal{O}(s\log(d/s))$ magnitude measurements [18, 30]. In this section we prove that it is also possible to recover s-sparse vectors $\mathbf{x} \in \mathbb{C}^d$ up to an unknown phase factor in only $\mathcal{O}(s\log^6 d)$-time using $\mathcal{O}(s\log^5 d)$ magnitude measurements. Thus, we establish the first known nearly runtime-optimal (i.e., essentially linear time in s) compressive phase retrieval recovery result. In particular, we prove the following theorem.

Theorem 7. *There exists a deterministic algorithm $\mathcal{A} : \mathbb{R}^D \to \mathbb{C}^d$ for which the following holds: let $\epsilon \in (0, 1]$, $\mathbf{x} \in \mathbb{C}^d$ with d sufficiently large and $s \in [d]$. Then, one can select a*
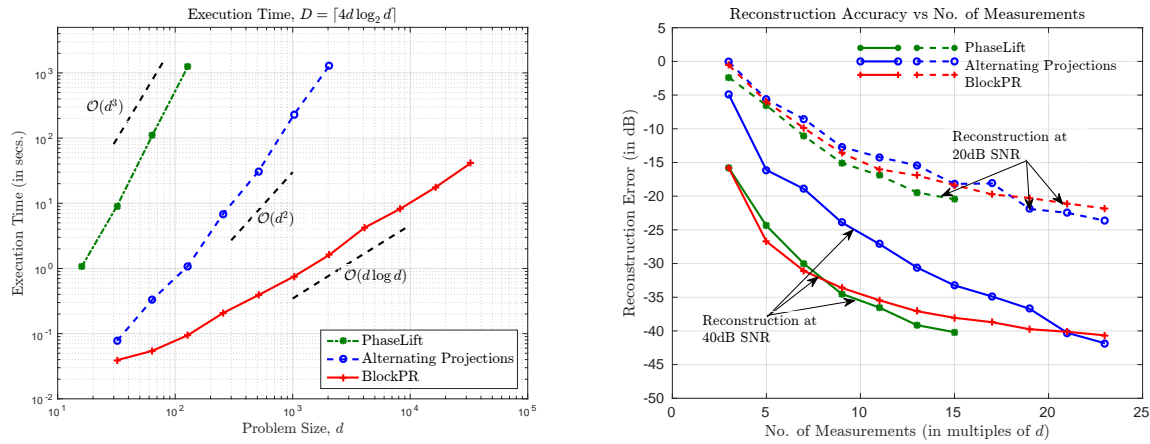
(a) Execution time versus problem size (local correlation measurements).

(b) Reconstruction error versus measurements ($d = 64$, i.i.d. Gaussian noise).
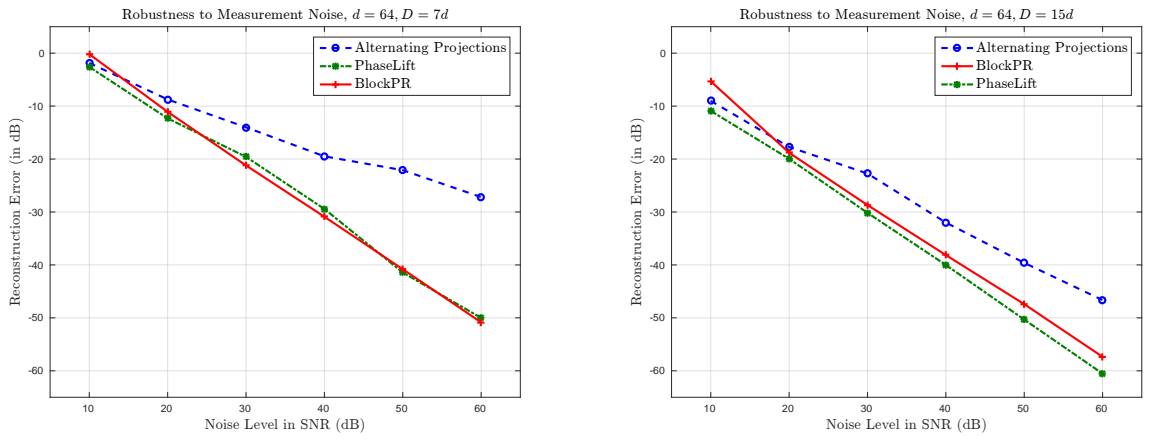
(c) Robustness to additive noise ($d = 64, D = 7d$, (Fourier-like) local correlation measurements)

(d) Robustness to additive noise ($d = 64, D = 15d$, (Fourier-like) local correlation measurements)

Figure 5. *Evaluation of phase retrieval algorithms for* local *correlation measurements.*

random measurement matrix $\tilde{M} \in \mathbb{C}^{D \times d}$ *such that*

$$(46) \qquad \min_{\theta \in [0,2\pi]} \left\| \mathrm{e}^{\mathrm{i}\theta}\mathbf{x} - \mathcal{A}\left(|\tilde{M}\mathbf{x}|^2\right) \right\|_2 \leq \left\| \mathbf{x} - \mathbf{x}_s^{\mathrm{opt}} \right\|_2 + \frac{22\epsilon \left\| \mathbf{x} - \mathbf{x}_{(s/\epsilon)}^{\mathrm{opt}} \right\|_1}{\sqrt{s}}$$

is true with probability at least $1 - \frac{1}{C \cdot \ln^2(d) \cdot \ln^3(\ln d)}$.[23]   *Here D can be chosen to be*

---

[23]Here $C \in \mathbb{R}^+$ is a fixed absolute constant.

$\mathcal{O}(\frac{s}{\epsilon} \cdot \ln^3(\frac{s}{\epsilon}) \cdot \ln^3(\ln\frac{s}{\epsilon}) \cdot \ln d)$. *Furthermore, the algorithm will run in* $\mathcal{O}(\frac{s}{\epsilon} \cdot \ln^4(\frac{s}{\epsilon}) \cdot \ln^3(\ln\frac{s}{\epsilon}) \cdot \ln d)$- *time in that case.*[24]

We prove Theorem 7 by following the generic compressive phase retrieval recipe presented in [30]. Let $C \in \mathbb{C}^{m \times d}$ be any compressive sensing matrix with an associated sparse approximation algorithm $\Delta : \mathbb{C}^m \to \mathbb{C}^d$ (see, e.g., [13, 9, 51, 39, 6, 40, 41]), and let $P \in \mathbb{C}^{D \times m}$ be any phase retrieval matrix with an associated recovery algorithm $\Phi : \mathbb{R}^D \to \mathbb{C}^m$. Then, $\Delta \circ \Phi : \mathbb{R}^D \to \mathbb{C}^d$ will approximately recover compressible vectors $\mathbf{x} \in \mathbb{C}^d$ up to an unknown phase factor when provided with the magnitude measurements $|PC\mathbf{x}|$. That is, one may first use Φ to recover $\mathbb{e}^{\mathbb{i}\phi}(C\mathbf{x}) = C(\mathbb{e}^{\mathbb{i}\phi}\mathbf{x})$ for some unknown $\phi \in [0, 2\pi]$ from $|PC\mathbf{x}|$, and then use Δ to recover $\mathbb{e}^{\mathbb{i}\phi}\mathbf{x}$ from $C(\mathbb{e}^{\mathbb{i}\phi}\mathbf{x})$. If both Φ and Δ are efficient, the result will be an efficient sparse phase retrieval method.

Herein we will utilize Algorithm 1 as our phase retrieval method. Note that it's runtime is only $\mathcal{O}(m \log^4 m)$, making it optimal up to log factors (recall Theorem 3). For the compressive sensing method we will utilize the following algorithmic result from [28].

Theorem 8. *Let* $\epsilon \in (0, 1]$, $\sigma \in [2/3, 1)$, $\mathbf{x} \in \mathbb{C}^d$, *and* $s \in [d]$. *With probability at least* σ *the deterministic compressive sensing algorithm from* [28] *will output a vector* $\mathbf{z} \in \mathbb{C}^d$ *satisfying*

$$(47) \qquad \|\mathbf{x} - \mathbf{z}\|_2 \le \left\|\mathbf{x} - \mathbf{x}_s^{\mathrm{opt}}\right\|_2 + \frac{22\epsilon \left\|\mathbf{x} - \mathbf{x}_{(s/\epsilon)}^{\mathrm{opt}}\right\|_1}{\sqrt{s}}$$

when executed with random linear input measurements $\mathcal{M}\mathbf{x} \in \mathbb{C}^m$. *It suffices to choose* $m = \mathcal{O}(\frac{s}{\epsilon} \cdot \ln(\frac{s/\epsilon}{1-\sigma}) \ln d)$. *The required runtime of the algorithm is* $\mathcal{O}(\frac{s}{\epsilon} \cdot \ln(\frac{s/\epsilon}{1-\sigma}) \ln(\frac{d}{1-\sigma}))$ *in this case.*[25]

Theorem 7 now follows easily from Theorem 3 with $\mathbf{n} = \mathbf{0}$, and Theorem 8.

7. Future work. This paper provides the first known global robust recovery guarantees for ptychographic phase retrieval problems, as well as a new numerical lifting + angular synchronization solution approach which is applicable with correlation-type measurements resulting from a wide class of locally supported masks $\mathbf{m}_\omega$. In doing so, it opens up many additional avenues of research. Examples include, e.g., questions associated with the use of different boundary conditions for $\mathbf{x}$. Herein, $\mathbf{x}$ is assumed to be periodic, which leads to the lifted linear system matrix M' in (10) being block circulant. It would be interesting to derive condition number bounds, as well as specialized numerical solution techniques, for the lifted linear system matrices that result from alternate boundary condition assumptions. More generally, it would be interesting to derive theoretical condition number bounds for the types of lifted linear systems that arise from more general classes of masks $\mathbf{m}_\omega$ under various sets of boundary condition assumptions for $\mathbf{x}$.

It would also be interesting to explore alternate approaches, besides Algorithm 2, for solving the angular synchronization problem via local phase differences that appears in the

---

[24] For the sake of simplicity, we assume $s = \Omega(\log d)$ when stating the measurement and runtime bounds above.

[25] For the sake of simplicity, we assume $s = \Omega(\log d)$ when stating the measurement and runtime bounds above.

second stage of our proposed phase retrieval approach. In particular, it would be interesting to develop theoretical error bounds for other solution methods (e.g., [49]) in the presence of noise for these types of local angular synchronization problems. Finally, it would also be instructive to carefully consider the optimality (or lack thereof) of the robustness guarantees derived herein for m-flat signals (i.e., Theorem 5) under various noise models.

Acknowledgment. The authors would like to thank the anonymous reviewers for many helpful suggestions and interesting ideas regarding possible directions for future work.

REFERENCES

[1] D. ACHLIOPTAS, *Database-friendly random projections*, in Proceedings of the ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, ACM, New York, 2001, pp. 274–281.

[2] B. ALEXEEV, A. S. BANDEIRA, M. FICKUS, AND D. G. MIXON, *Phase retrieval with polarization*, SIAM J. Imaging Sci., 7 (2014), pp. 35–66.

[3] R. BARANIUK, M. DAVENPORT, R. DEVORE, AND M. WAKIN, *A simple proof of the restricted isometry property for random matrices*, Constr. Approx., 28 (2008), pp. 253–263.

[4] S. BECKER, E. J. CANDES, AND M. GRANT, *Templates for convex cone problems with applications to sparse signal recovery*, Math. Program. Comput., 3 (2011), pp. 165–218.

[5] S. BECKER, E. J. CANDES, AND M. GRANT, *TFOCS: Templates for First-Order Conic Solvers, Version 1.4*. http://cvxr.com/tfocs, 2016.

[6] T. BLUMENSATH AND M. E. DAVIES, *Iterative hard thresholding for compressed sensing*, Appl. Comput. Harmon. Anal., 27 (2009), pp. 265–274.

[7] J. BOURGAIN AND J. NELSON, *Toward a unified theory of sparse dimensionality reduction in Euclidean space*, Geom. Funct. Anal., 25 (2015), pp. 1009–1088.

[8] E. J. CANDES AND T. TAO, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.

[9] E. J. CANDES AND T. TAO, *Near optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.

[10] E. J. CANDES AND X. LI, *Solving quadratic equations via PhaseLift when there are about as many equations as unknowns*, Found. Comput. Math., 14 (2014), pp. 1017–1026.

[11] E. J. CANDES, X. LI, AND M. SOLTANOLKOTABI, *Phase retrieval from coded diffraction patterns*, Appl. Comput. Harmon. Anal., 39 (2015), pp. 277–299.

[12] E. J. CANDES, X. LI, AND M. SOLTANOLKOTABI, *Phase retrieval via Wirtinger flow: Theory and algorithms*, IEEE Trans. Inform. Theory, 61 (2015), pp. 1985–2007.

[13] E. J. CANDES, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[14] E. J. CANDES, T. STROHMER, AND V. VORONINSKI, *Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming*, Comm. Pure Appl. Math., 66 (2013), pp. 1241–1274.

[15] Y. CHEN AND E. J. CANDES, *Solving random quadratic systems of equations is nearly as easy as solving linear systems*, Comm. Pure Appl. Math., doi:10.1002/cpa.21638.

[16] S. DASGUPTA AND A. GUPTA, *An elementary proof of a theorem of Johnson and Lindenstrauss*, Random Structures Algorithms, 22 (2003), pp. 60–65.

[17] L. DEMANET AND P. HAND, *Stable optimizationless recovery from phaseless linear measurements*, J. Fourier Anal. Appl., 20 (2014), pp. 199–221.

[18] Y. C. ELDAR AND S. MENDELSON, *Phase retrieval: Stability and recovery guarantees*, Appl. Comput. Harmon. Anal., 36 (2014), pp. 473–494.

[19] A. FANNJIANG AND W. LIAO, *Phase retrieval with random phase illumination*, J. Opt. Soc. Amer. A, 29 (2012), pp. 1847–1859.

[20] J. R. FIENUP, *Reconstruction of an object from the modulus of its Fourier transform*, Opt. Lett., 3 (1978), pp. 27–29.

[21] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, Basel, 2013.

[22] P. Frankl and H. Maehara, *The Johnson–Lindenstrauss lemma and the sphericity of some graphs*, J. Combin. Theory Ser. B, 44 (1988), pp. 355–362.

[23] R. Gerchberg and W. Saxton, *A practical algorithm for the determination of phase from image and diffraction plane pictures*, Optik, 35 (1972), pp. 237–246.

[24] M. Grant and S. Boyd, *Graph implementations for nonsmooth convex programs*, in Recent Advances in Learning and Control, V. Blondel, S. Boyd, and H. Kimura, eds., Lecture Notes in Control and Inform. Sci. 371, Springer-Verlag, London, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[25] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming, Version* 2.1, http://cvxr.com/cvx, 2014.

[26] D. Gross, F. Krahmer, and R. Kueng, *Improved Recovery Guarantees for Phase Retrieval from Coded Diffraction Patterns*, preprint, arXiv:1402.6286; available online from doi.org/10.1016/j.acha.2015.05.004 (2014).

[27] R. W. Harrison, *Phase problem in crystallography*, J. Opt. Soc. Amer. A, 10 (1993), pp. 1046–1055.

[28] M. Iwen, *Compressed sensing with sparse binary matrices: Instance optimal error guarantees in near-optimal time*, J. Complexity, 30 (2014), pp. 1–15.

[29] M. Iwen, F. Krahmer, and A. Viswanathan, *Technical Note: A Minor Correction of Theorem* 1.3 *from* [1], http://users.math.msu.edu/users/markiwen/Papers/PhaseLiftproof.pdf (2015).

[30] M. Iwen, A. Viswanathan, and Y. Wang, *Robust sparse phase retrieval made easy*, Appl. Comput. Harmon. Anal., doi:10.1016/j.acha.2015.06.007.

[31] M. Iwen, Y. Wang, and A. Viswanathan, *BlockPR: Matlab Software for Phase Retrieval from Local Correlation Measurements, Version* 0.2, https://bitbucket.org/charms/blockpr (2016).

[32] W. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Contemp. Math., 26 (1984), pp. 189–206.

[33] F. Krahmer and R. Ward, *New and improved Johnson–Lindenstrauss embeddings via the restricted isometry property*, SIAM J. Math. Anal., 43 (2011), pp. 1269–1281.

[34] X. Li and V. Voroninski, *Sparse signal recovery from quadratic measurements via convex programming*, SIAM J. Math. Anal., 45 (2013), pp. 3019–3033.

[35] A. M. Maiden and J. M. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*, Ultramicroscopy, 109 (2009), pp. 1256–1262.

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

[37] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, *Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes*, Annu. Rev. Phys. Chem., 59 (2008), pp. 387–410.

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

[39] D. Needell and J. A. Tropp, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 301–321.

[40] D. Needell and R. Vershynin, *Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit*, Found. Comput. Math., 9 (2009), pp. 317–334.

[41] D. Needell and R. Vershynin, *Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit*, IEEE J. Sel. Top. Signal Process., 4 (2010), pp. 310–316.

[42] P. Netrapalli, P. Jain, and S. Sanghavi, *Phase retrieval using alternating minimization*, IEEE Trans. Signal Process., 63 (2015), pp. 4814–4826.

[43] H. Ohlsson, A. Yang, R. Dong, and S. Sastry, *CPRL: An extension of compressive sensing to the phase retrieval problem*, in Proceedings of Advances in Neural Information Processing Systems, Neural Information Processing Systems, La Jolla, CA, 2012, pp. 1376–1384.

[44] J. Qian, C. Yang, A. Schirotzek, F. Maia, and S. Marchesini, *Efficient algorithms for ptychographic phase retrieval*, Inverse Probl. Appl., Contemp. Math, 615 (2014), pp. 261–280.

[45] J. M. Rodenburg, *Ptychography and related diffractive imaging methods*, Adv. Imaging Electron Phys., 150 (2008), pp. 87–184.

[46] J. M. Rodenburg and H. M. Faulkner, *A phase retrieval algorithm for shifting illumination*, Appl. Phys. Lett., 85 (2004), pp. 4795–4797.

[47] P. SCHNITER AND S. RANGAN, *Compressive phase retrieval via generalized approximate message passing*, IEEE Trans. Signal Process., 63 (2014), pp. 1043–1055.

[48] Y. SHECHTMAN, A. BECK, AND Y. C. ELDAR, *GESPAR: Efficient phase retrieval of sparse signals*, IEEE Trans. Signal Process., 62 (2014), pp. 928–938.

[49] A. SINGER, *Angular synchronization by eigenvectors and semidefinite programming*, Appl. Comput. Harmon. Anal., 30 (2011), pp. 20–36.

[50] G. J. TEE, *Eigenvectors of block circulant and alternating circulant matrices*, New Zealand J. Math., 36 (2007), pp. 195–211.

[51] J. TROPP AND A. GILBERT, *Signal recovery from partial information via orthogonal matching pursuit*, IEEE Trans. Inform. Theory, 53 (2007), pp. 4655–4666.

[52] A. WALTHER, *The question of phase retrieval in optics*, J. Modern Opt., 10 (1963), pp. 41–49.

[53] Y. WANG AND Z. XU, *Phase retrieval for sparse signals*, Appl. Comput. Harmon. Anal., 37 (2014), pp. 531–544.