INVERTING SPECTROGRAM MEASUREMENTS VIA ALIASED WIGNER DISTRIBUTION DECONVOLUTION AND ANGULAR SYNCHRONIZATION

MICHAEL PERLMUTTER, SAMI MERHI, ADITYA VISWANATHAN, MARK IWEN

Abstract. We propose a two-step approach for reconstructing a signal $\mathbf{x} \in \mathbb{C}^d$ from subsampled short-time Fourier transform magnitude (spectogram) measurements: First, we use an aliased Wigner distribution deconvolution approach to solve for a portion of the rank-one matrix $\widehat{\mathbf{x}}\widehat{\mathbf{x}}^*$. Second, we use angular syncrhonization to solve for $\widehat{\mathbf{x}}$ (and then for $\mathbf{x}$ by Fourier inversion). Using this method, we produce two new efficient phase retrieval algorithms that perform well numerically in comparison to standard approaches and also prove two theorems, one which guarantees the recovery of discrete, bandlimited signals $\mathbf{x} \in \mathbb{C}^d$ from fewer than d STFT magnitude measurements and another which establishes a new class of deterministic coded diffraction pattern measurements which are guaranteed to allow efficient and noise robust recovery.

1. Introduction

The *phase retrieval problem*, i.e., reconstructing a signal from phaseless measurements, is at the core of many scientific breakthroughs related to the imaging of cells [46], viruses [44], and nanocrystals [15], and also advances in crystallographic imaging [28], optics [48], astronomy [18], quantum mechanics [16], and speech signal processing [2, 24]. As a result, many sophisticated algorithms, which achieve great empircal success, have been developed for solving this problem in applications throughout science and engineering (see [19, 21, 24] for widely used examples). Motivated by the success of these methods, the mathematical community has recently began to study the challenging problem of designing measurement masks and corresponding reconstruction algorithms with rigorous convergence guarantees and noise robustness properties (see, e.g., the work of Balan, Candès, Strohmer, and others [1, 2, 13, 25]). In this paper, we aim to extend the mathematical analysis of phaseless measurement maps and noise-robust reconstruction algorithms to include a broad class of phaseless Gabor measurements such as those that are utilized in, e.g., ptychographic imaging [14, 17, 40, 41].

Specifically, we will develop and analyze several algorithms for recovering (up to a global phase) a signal $\mathbf{x} \in \mathbb{C}^d$ from the magnitudes of its inner products with shifts of masks that are *locally supported* in either physical space or Fourier space. The local support of these masks in physical space corresponds to the use of concentrated beams in ptychographic imaging to measure small portions of a large sample, whereas the local support of these masks in Fourier space simulates the recovery of samples belonging to a special class of deterministic coded diffraction patterns (CDP).

Following [33, 34], we will assume that we have a family of measurement masks, or windows, $\mathbf{m_0}, \mathbf{m_1}, \ldots, \mathbf{m_{K-1}} \in \mathbb{C}^d$ such that for all k, the nonzero entries of either $\mathbf{m_k}$ or $\widehat{\mathbf{m}}_\mathbf{k}$ are contained in the set $[\delta]_0$ for some fixed $\delta < \frac{d}{2}$, where for any integer $n \geq 0$, we let

$$[n]_0 = \{0, 1, \ldots, n-1\}$$

denote the set of the first n nonnegative integers. Let L be an integer which divides d, and let $Y' : \mathbb{C}^d \to [0,\infty)^{K \times L}$ be the matrix-valued measurement map defined by its coordinate functions

$$Y'_{k,\ell} \coloneqq Y'_{k,\ell}(\mathbf{x}) \coloneqq |\langle S_{\ell a}\mathbf{m_k}, \mathbf{x}\rangle|^2 + N'_{k,\ell}, \tag{1.1}$$

for $k \in [K]_0$ and $\ell \in [L]_0$, where $a \coloneqq \frac{d}{L}$, S_ℓ is the circular shift operator on $\mathbb{C}^d$ defined for $\mathbf{y} \in \mathbb{C}^d$ and $\ell \in \mathbb{Z}$ by

$$(S_\ell \mathbf{y})_j \coloneqq y_{((j+\ell) \bmod d)}, \tag{1.2}$$

Key words and phrases. Phase Retrieval, Spectrogram Measurements, Short-Time Fourier Transform (STFT), Wigner Distribution Deconvolution, Angular Synchronization, Ptychography.

and $N' = \left(N'_{k,\ell}\right)_{k\in[K]_0,\ell\in[L]_0} \in \mathbb{R}^{K\times L}$ represents an arbitrary perturbation due to, e.g., measurement noise or imperfect knowledge of the masks $\mathbf{m_k}$.

Our goal is to reconstruct $\mathbf{x}$ from these measurements. It is clear that $Y'(\mathbf{x}) = Y'(\mathrm{e}^{\mathrm{i}\phi}\mathbf{x})$ for all $\phi \in \mathbb{R}$, so at best we can hope to reconstruct $\mathbf{x}$ up to a global phase, i.e., up to the equivalence relation

$$\mathbf{x} \sim \mathbf{x}' \text{ if } \mathbf{x} = \mathrm{e}^{\mathrm{i}\phi}\mathbf{x}' \text{ for some } \phi \in \mathbb{R}.$$

Algorithms 1 and 2, presented in Section 4, will accomplish this goal in the special case where the masks $\mathbf{m_k}$ are obtained by modulating a single mask $\mathbf{m}$. Specifically, we let $\mathbf{m} \in \mathbb{C}^d$, and for $k \in [d]_0$ we let

$$\mathbf{m_k} = W_k\mathbf{m} \tag{1.3}$$

where W_k is the modulation operator given by

$$(W_k\mathbf{m})_j \coloneqq \mathrm{e}^{\frac{2\pi\mathrm{i}jk}{d}} m_j. \tag{1.4}$$

As we will see, assuming that our masks have this form will allow us to recover $\mathbf{x}$, even when the shift size a is strictly greater than one. Towards this end, we let $Y : \mathbb{C}^d \to [0,\infty)^{d\times d}$ be the matrix-valued measurement map defined by its coordinate functions

$$Y_{k,\ell} \coloneqq Y_{k,\ell}(\mathbf{x}) \coloneqq |\langle S_\ell W_k\mathbf{m}, \mathbf{x}\rangle|^2 + N_{k,\ell}, \tag{1.5}$$

where analogously to (1.1), $N = (N_{k,\ell})_{0\le k,\ell\le d-1}$ represents an arbirtrary perturbation. We note that Y is the special case of Y' where $K = L = d$ and the masks $\mathbf{m_k}$ have the form (1.3). For positve integers, K and L which divide d, we let

$$\frac{d}{K}[K]_0 \coloneqq \left\{0, \frac{d}{K}, \frac{2d}{K}, \ldots, d - \frac{d}{K}\right\}, \quad \text{and} \quad \frac{d}{L}[L]_0 \coloneqq \left\{0, \frac{d}{L}, \frac{2d}{L}, \ldots, d - \frac{d}{L}\right\},$$

and we let $Y_{K,L}$ be the $K \times L$ partial measurement matrix obtained by restricting Y to rows in $\frac{d}{K}[K]_0$ and columns in $\frac{d}{L}[L]_0$ so that the (k,ℓ)-th entry of $Y_{K,L}$ is given by

$$(Y_{K,L})_{k,\ell} = Y_{\frac{kd}{K},\frac{\ell d}{L}}. \tag{1.6}$$

Similarly, we let $N_{K,L}$ be the $K \times L$ matrix obtained by restricting N to rows and columns in $\frac{d}{K}[K]_0$ and $\frac{d}{L}[L]_0$.

Letting $\omega_k = k\frac{d}{K}$ so that $(Y_{K,L})_{k,\ell} = Y_{\omega_k,\ell a}$, we note that

$$(Y_{K,L})_{k,\ell} = |\langle \mathbf{x}, S_{\ell a}W_{\omega_k}\mathbf{m}\rangle|^2 + N_{\omega_k,\ell a} = \left|\left\langle \mathbf{x}, \mathrm{e}^{\frac{2\pi\mathrm{i}\ell a\omega_k}{d}} W_{\omega_k}S_{\ell a}\mathbf{m}\right\rangle\right|^2 + N_{\omega_k,\ell a} = |\langle \mathbf{x}, W_{\omega_k}S_{\ell a}\mathbf{m}\rangle|^2 + N_{\omega_k,\ell a}. \tag{1.7}$$

Therefore, $Y_{K,L}$ forms a matrix of STFT magnitude measurements. Furthermore, when $K = d$ and $\omega_k = k$, (1.5) also encompasses a large class of masked Fourier magnitude measurements (i.e., CDP measurements) of the form

$$(Y_{K,L})_{k,\ell} = |\langle \mathbf{x}, W_{\omega_k}S_{\ell a}\mathbf{m}\rangle|^2 + N_{k,\ell} = |\left(F_d\ \mathrm{Diag}(\mathbf{m'}_\ell)\ \mathbf{x}\right)_k|^2 + (N_{K,L})_{k,\ell}, \tag{1.8}$$

where $\mathbf{m'}_\ell \coloneqq S_{\ell a}\overline{\mathbf{m}}$ and F_d is the $d \times d$ discrete Fourier transform matrix whose entries are defined by

$$(F_d)_{j,k} \coloneqq \mathrm{e}^{\frac{-2\pi\mathrm{i}jk}{d}}. \tag{1.9}$$

Measurements similar to (1.8) are considered in, e.g., the recent works by Candès and others [3, 10, 11, 25]. However, their masks are usually generated randomly, whereas we will consider deterministically designed mask constructed as shifts of a single base mask $\mathbf{m}$.

Our method for recovering $\mathbf{x}$ is based on a two-step approach. Following the example of, e.g., [2, 13], we can lift the nonlinear, phaseless measurements (1.1) to linear measurements of the Hermitian rank-one matrix $\mathbf{x}\mathbf{x}^*$. Specifically, it can be shown that

$$Y'_{k,\ell}(\mathbf{x}) = \langle \mathbf{x}\mathbf{x}^*, S_{\ell a}\mathbf{m_k}\mathbf{m_k}^*S_{\ell a}^*\rangle + N'_{k,\ell},$$

where the inner product above is the Hilbert-Schmidt inner product. Restricting, for the moment, to the case $a = 1$ (i.e. $L = d$) and assuming that the nonzero entries of $\mathbf{m_k}$ are contained in the set $[\delta]_0$ for all k, one can see that every matrix $G \in \mathrm{span}\left(\{S_\ell\mathbf{m_k}\mathbf{m_k}^*S_\ell^* : \ell \in [d]_0, k \in [K]_0\}\right)$ will have all of its nonzero entries

concentrated near the main diagonal. Specifically, we have $G_{ij} = 0$ unless either $|i-j| < \delta$ or $|i-j| > d-\delta$ Therefore, letting $T_\delta : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ be the restriction operator given by

$$T_\delta(G)_{ij} = \begin{cases} G_{ij} & \text{if } |i-j| < \delta \text{ or } |i-j| > d-\delta \\ 0 & \text{otherwise} \end{cases},$$

we see

$$Y'_{k,\ell}(\mathbf{x}) = \langle \mathbf{x}\mathbf{x}^*, S_\ell \mathbf{m_k} \mathbf{m_k}^* S_\ell^* \rangle + N'_{k,\ell} = \langle T_\delta(\mathbf{x}\mathbf{x}^*), S_\ell \mathbf{m_k} \mathbf{m_k}^* S_\ell^* \rangle + N'_{k,\ell}, \quad (k,\ell) \in [K]_0 \times [d]_0.$$

Our lifted, linearized measurements are therefore given by $Y'_{k,\ell}(\mathbf{x}) = \mathcal{A}\big(T_\delta(\mathbf{x}\mathbf{x}^*)\big)_{(k,\ell)} + N'_{k,\ell}$ where $\mathcal{A} : T_\delta(\mathbb{C}^{d\times d}) \to \mathbb{C}^{K\times d}$ is defined by

$$(\mathcal{A}(X))_{(k,\ell)} = \langle X, S_\ell \mathbf{m_k} \mathbf{m_k}^* S_\ell^* \rangle \quad \text{for} \quad (k,\ell) \in [K]_0 \times [d]_0 \quad \text{and} \quad X \in T_\delta(\mathbb{C}^{d\times d}). \tag{1.10}$$

As a result, one can approximately solve for $\mathbf{x}$ up to a global phase factor by (i) evaluating $\mathcal{A}^{-1}$ on $Y'_{k,\ell}(\mathbf{x})$ in order to recover a Hermitian approximation X_e to $T_\delta(\mathbf{x}\mathbf{x}^*)$, and then (ii) applying a noise robust angular synchronization method (e.g., see [45, 47]) to obtain an estimate of $\mathbf{x}$ from X_e. See [33, 34] for further details.

The following theorem summarizes previous work using this two-stage approach for the case where the nonzero entries of the masks $\mathbf{m_k}$ are contained in the set $[\delta]_0$ for all $k \in [K]_0$.

Theorem 1 (See [33, 34]). *For $\mathbf{x} \in \mathbb{C}^d$, let $\min|\mathbf{x}| \coloneqq \min_{0\le j\le d}|x_j|$, and set $K = 2\delta - 1$ and $L = d$ so that $a = 1$ in (1.1). There exists a practical nonlinear reconstruction algorithm that takes in measurements $Y'_{k,\ell}(\mathbf{x})$ for all $(k,\ell) \in [K]\times[d]$ and outputs an estimate $\mathbf{x}_e \in \mathbb{C}^d$ that always satisfies*

$$\min_{\phi\in[0,2\pi]} \left\|\mathbf{x} - \mathrm{e}^{\mathrm{i}\phi}\mathbf{x}_e\right\|_2 \le C\left(\frac{\|\mathbf{x}\|_\infty}{\min|\mathbf{x}|^2}\right)\left(\frac{d}{\delta}\right)^2 \kappa\|N'\|_F + Cd^{\frac{1}{4}}\sqrt{\kappa\|N'\|_F}. \tag{1.11}$$

Here $\kappa > 0$ is the condition number of the linear map $\mathcal{A}$ in (1.10) and $C \in \mathbb{R}^+$ is an absolute universal constant.

Furthermore, it is possible to choose masks $\mathbf{m_0}, \mathbf{m_1}, \ldots, \mathbf{m_{2\delta-2}}$ such that $\kappa < 4\delta$ (see [33]), and it is also possible to construct a single mask $\mathbf{m} \in \mathbb{C}^d$ such that if $\mathbf{m_k} = W_{k\frac{d}{K}}\mathbf{m}$, then $\kappa = \mathcal{O}(\delta^2)$ for the measurements that appear in (1.5) (see [34]). Also, if $\|N'\|_F$ is sufficiently smalll, the algorithm mentioned above is guaranteed to require just $\mathcal{O}\left(\delta^2 d\log d + \delta^3 d\right)$ total flops to achieve (1.11) up to machine precision.

Note that (1.11) guarantees that the algorithm in [33] referred to by Theorem 1 *exactly inverts* (up to a global phase) the measurement map in (1.1) for all nonvanishing $\mathbf{x}$ in the noiseless setting (i.e., when $\|N'\|_F = 0$). Furthermore, the error between the recovered and original signal degrades gracefully with small amounts of arbitrary additive noise, and when $\delta \ll d$ the algorithm runs in essentially FFT-time. Indeed, a thorough numerical evaluation of this method has demonstrated it to be significantly more computationally efficient than competing techniques when, e.g., $\delta = \mathcal{O}(\log d)$ (see [33]). While the the first term of the error bound obtained in Theorem 1 exhibits quadratic dependence in d, we note that the main results of [32] imply, at least heuristically, that polynomial dependencies on d are actually unavoidable in any upper bound like (1.11) when the masks are locally supported. As a result, both (1.11) as well as the new error bounds developed below generally must exhibit such polynomial dependences on d.

1.1. **Main Results.** One of the main drawbacks of Theorem 1 is that it only holds for shifts of size $a = 1$, i.e., when $L = d$. In real ptychographic imaging applications, however, the equivalent of our parameter a will in fact often at least 0.4δ, with δ being moderately large. Therefore, we will consider recovery scenarios where both K and L are strictly less than d in (1.6), and consider classes of $\mathbf{x}$ and $\mathbf{m}$ for which we can still guarantee noise-robust recovery results. This motivates our first new result, which allows us to recover bandlimited signals.

Theorem 2 (Convergence Gaurantees for Algorithm 2). *Let $\mathbf{x}, \mathbf{m} \in \mathbb{C}^d$ with $\mathrm{supp}\,(\widehat{\mathbf{x}}) \subseteq [\gamma]_0$ and $\mathrm{supp}\,(\mathbf{m}) \subseteq [\delta]_0$, and let*

$$\mu_2 = \min_{\substack{|p|\le\gamma-1,\\ |q|\le\delta-1}} \left| F_d\left(\widehat{\mathbf{m}} \circ S_p\overline{\widehat{\mathbf{m}}}\right)_q \right| \; > \; 0. \tag{1.12}$$

Assume that $\gamma \leq 2\delta - 1 < d$, $L = 2\gamma - 1$, $K = 2\delta - 1$, *and also that* K *and* L *divide* d. *Furthermore, suppose that the phaseless measurements* (1.6) *have noise dominated by the norm of* $\mathbf{x}$ *so that*

$$\|N_{K,L}\|_F \leq \beta \|\mathbf{x}\|_2^2 \tag{1.13}$$

for some $\beta \geq 0$. *Then Algorithm 2 in Section 4 outputs an estimate* $\mathbf{x}_e$ *to* $\mathbf{x}$ *with relative error*

$$\min_{\phi \in [0,2\pi]} \frac{\left\|\mathbf{x} - \mathbb{e}^{\mathbb{i}\phi}\mathbf{x}_e\right\|_2}{\|\mathbf{x}\|_2} \leq \frac{\left(1+2\sqrt{2}\right)\beta}{\sigma_\gamma(W)} \frac{d^2}{\sqrt{KL}\mu_2}, \tag{1.14}$$

where $W \in \mathbb{C}^{2\delta-1\times\gamma}$ *is the partial Fourier matrix with entries* $W_{j,k} = \mathbb{e}^{-\frac{2\pi\mathbb{i}(j-\delta+1)k}{d}}$ *and* γ^{th} *singular value* $\sigma_\gamma(W)$. *Furthermore, if* $\|N_{K,L}\|_F$ *is sufficently small, then Algorithm 2 is always guaranteed to require at most* $\mathcal{O}\left(KL\log(KL) + \delta^3 + \log(\|\widehat{\mathbf{x}}\|_\infty)\gamma^2\right)$ *total flops to achieve* (1.14) *up to machine precision.*

As mentioned earlier, Theorem 2 allows us to recover $\mathbf{x}$ even when K and L are both strictly less than d. Indeed, the total number of measurements, $KL = \mathcal{O}(\gamma\delta)$ is independent of the sample size d (though it does exhibit dependence on the parameters, γ and δ, and it also requires that $\delta > \gamma/2$). Nonetheless, the fact that Algorithm 2 exhibits robustness to arbitrary noise indicates that one can use it to quickly obtain a low-pass approximation to a sufficiently smooth $\mathbf{x} \in \mathbb{C}^d$ using fewer than d STFT magnitude measurements. We also note that Proposition 2, stated in Section 4, shows that locally supported masks $\mathbf{m}$ with $\mu_2 > 0$ are relatively simple to construct.

Our second result utilizes the connection between CPD measurements and STFT magnitude measurements (see (1.7) and (1.8)) to provide a new class of deterministic CPD measurement constructions along with an associated noise-robust recovery algorithm. Unlike previously existing deterministic constructions (see, e.g., Theorem 3.1 in [10]) the following result presents a general means of constructing deterministic CDP masks using shifts of a single bandlimited mask $\mathbf{m}$.

Theorem 3 (Convergence Gaurantees for Algorithm 1)**.** *Let* $\mathbf{x}, \mathbf{m} \in \mathbb{C}^d$ *with* $supp\left(\widehat{\mathbf{m}}\right) \subseteq [\rho]_0$, *for some* $\rho < d/2$. *Let* $\min|\widehat{\mathbf{x}}| \coloneqq \min_{0\leq n\leq d-1}|\widehat{x}_n| > 0$, *and let*

$$\mu_1 \coloneqq \min_{\substack{|p|\leq\gamma-1\\|q|\leq\rho-1}} \left| F_d\left(\widehat{\mathbf{m}} \circ S_p\overline{\widehat{\mathbf{m}}}\right)_q \right| \ > \ 0. \tag{1.15}$$

Fix an integer $\kappa \in [2, \rho]$ *and assume that* $L = \rho + \kappa - 1$ *divides* d. *Then, when* $K = d$, *Algorithm 1 in Section 4 will output* $\mathbf{x}_e$, *an estimate of* $\mathbf{x}$, *such that*

$$\min_{\phi\in[0,2\pi]} \left\|\mathbf{x} - \mathbb{e}^{\mathbb{i}\phi}\mathbf{x}_e\right\|_2 \leq C\frac{d^{7/2}\|\widehat{\mathbf{x}}\|_\infty \|N_{d,L}\|_F}{L^{\frac{1}{2}}\mu_1\kappa^{\frac{5}{2}} \cdot \min|\widehat{\mathbf{x}}|^2} + C'\frac{d^{\frac{3}{2}}}{L^{\frac{1}{4}}}\sqrt{\frac{\|N_{d,L}\|_F}{\mu_1}} \tag{1.16}$$

for some absolute constants $C, C' \in \mathbb{R}^+$. *Furthermore, if* $\|N_{d,L}\|_F$ *is sufficently small, then Algorithm 1 is always guaranteed to require just* $\mathcal{O}\left(d(\rho+\kappa^2)\log d\right)$ *total flops to achieve* (1.16) *up to machine precision.*

When $\kappa = \rho = 2$, Theorem 3 guarantees that $3d$ CDP measurements suffice in order to recover any signal $\mathbf{x}$ with a nonvanishing discrete Fourier transform in the noiseless setting as long as $\mu_1 > 0$. Analogously to Proposition 2, Proposition 1, also stated in Section 4, provides a straightforward way to construct masks with $\mu_1 > 0$. For general κ and ρ, Theorem 3 shows that one can reconstruct signals $\mathbf{x}$ using $\mathcal{O}(d\rho)$ CPD measurements based on windows with Fourier support ρ in just $\mathcal{O}\left(\rho^2 d\log d\right)$-time when $\|N\|_F$ is sufficiently small. We also note that in addition to the theoretical guarantees provided by Theorems 2 and 3, Section 5 demonstrates that both Algorithm 1 and 2 are fast, accurate, and robust to noise in practice as well.

1.2. **Related Work.** The connections between theoretical time-frequency analysis and phaseless imaging (e.g., ptychography) have been touched on in the physics community many times over the past several decades. As noted well over two decades ago in [41] and later in [14], continuous spectrogram measurements can be written as the convolution of the Wigner distribution functions of the specimen $\mathbf{x}$ and the probe $\mathbf{m}$. Furthermore, [17] has pointed out that this allows one to recover the specimen of interest if enough samples are drawn so that the Heisenberg boxes sufficiently cover the time-frequency plane. In this work, we use similar ideas formulated in the discrete setting to efficiently invert the types of structured lifted linear maps $\mathcal{A}$ as per (1.10) that appear in [33, 34], and use angular synchronization approaches to recover the signal $\mathbf{x}$ up

to a global phase. Specifically, we produce two new, efficient algorithms for inverting discrete spectrogram measurements that are provably accurate and robust to arbitrary additive measurement errors.

In [7], Bendory and Eldar prove results similar to some of those summarized in Theorem 1 in the case there $a = 1$ and $\|N'\|_F = 0$, and they also demonstrate numerically that their algorithms are robust to noise. In this paper, we prove noise-robust recovery results, where we allow $a > 1$. However, we make additional assumptions about either the support of $\widehat{\mathbf{m}}$ or the supports of $\mathbf{m}$ and $\widehat{\mathbf{x}}$. We also note the very recent and excellent work of Rayan Saab and Brian Preskitt [39] as well as that of Melnyk, Filbir, and Krahmer [35] which both prove results similar to Theorem 2. As in Theorem 2, the results of [35, 39] can guarantee recovery with shift sizes $a > 1$. Their results primarily differ from Theorem 2 in that they don't used Wigner Distribution Deconvolution (WDD) based methods. As a result, they consider different classes of masks and signals than we do.

Other related work includes that of Salanevich and Pfander [38, 42] which builds upon the work of Alexeev et al. [1] to establish noise robust recovery results for Gabor frame-based measurements. Their noise robust approach has similar characteristics to the approach taken here with the primary differences being that they require additional measurements beyond those provided by shifts and modulations of a single mask (see, e.g., equation (8) in [38]), and in some sense utilize the reverse of the approach taken here: Instead of first solving a linear system to obtain an approximation of (a portion of) $\mathbf{x}\mathbf{x}^*$, and then using angular synchronization to obtain an approximation to $\mathbf{x}$, the methods of [38, 42] instead first use angular synchronization methods to obtain frame coefficients of $\mathbf{x}$, and then reconstruct $\mathbf{x}$ using the recovered frame coefficients.

The rest of the paper is organized as follows. In Section 2, we establish necessary notation and state a number of preliminary lemmas. Then, in Section 3, we establish several discrete and aliased variants of WDD, some of which can be used when the mask $\mathbf{m}$ is locally supported in physical space, and others for when $\mathbf{m}$ is locally supported in Fourier space. In Section 4, we prove Theorems 2 and 3 which provide recovery guarantees for our proposed methods and also state propositions which describe ways to design masks so that the assumptions of these theorems are valid. Finally, in Section 5, we evaluate our algorithms numerically and show that they are fast and robust to additive measurement noise.

2. Notation and Preliminary Results

For $\mathbf{x} \coloneqq (x_0, \dots, x_{d-1})^T \in \mathbb{C}^d$, we let

$$\operatorname{supp}(\mathbf{x}) \coloneqq \{n \in [d]_0 : x_n \neq 0\}$$

denote the support of $\mathbf{x}$, where, as in Section 1, $[d]_0 = \{0, 1, \dots, d-1\}$. We let $R\mathbf{x} \coloneqq \widetilde{\mathbf{x}}$ denote the reversal of $\mathbf{x}$ about its first entry, i.e.,

$$(R\mathbf{x})_n = \widetilde{x}_n \coloneqq x_{-n \bmod d} \quad \text{for } 0 \leq n \leq d-1,$$

and we recall from (1.2) and (1.4) the circular shift and modulation operators given by $(S_\ell \mathbf{x})_n = x_{(\ell+n) \bmod d}$ and $(W_k \mathbf{x})_n = x_n \mathbb{e}^{\frac{2\pi \mathbb{i} kn}{d}}$. In order to avoid cumbersome notation, if n is not an element of $[d]_0$, we will write x_n in place of $x_{n \bmod d}$. For $\mathbf{x} \in \mathbb{C}^d$, we define the Fourier transform of $\mathbf{x}$ by

$$\widehat{x}_k \coloneqq (F_d \mathbf{x})_k = \sum_{n=0}^{d-1} x_n \mathbb{e}^{-\frac{2\pi \mathbb{i} nk}{d}},$$

where as in (1.9), $F_d \in \mathbb{C}^{d \times d}$ denotes the $d \times d$ discrete Fourier transform matrix with entries $(F_d)_{j,k} = \mathbb{e}^{-\frac{2\pi \mathbb{i} jk}{d}}$ for $0 \leq j, k \leq d-1$. For $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$ and $\ell \in [d]_0$, we define circular convolution and Hadamard (pointwise) multiplication by

$$(\mathbf{x} *_d \mathbf{y})_\ell \coloneqq \sum_{n=0}^{d-1} x_n y_{\ell-n}, \quad \text{and} \quad (\mathbf{x} \circ \mathbf{y})_\ell \coloneqq x_\ell y_\ell,$$

and we define their componentwise quotient $\frac{\mathbf{x}}{\mathbf{y}}$ and componentwise absolute value $|\mathbf{x}|$ by

$$\left(\frac{\mathbf{x}}{\mathbf{y}}\right)_n = \frac{x_n}{y_n} \quad \text{and} \quad |\mathbf{x}|_n = |x_n|.$$

For a matrix M, we let M_k denote its k-th column, and let $\|M\|_F$ denote its Frobenius norm. When proving the convergence of our algorithms, we will use the fact that, up to a reorganization of the terms, a

banded $d \times d$ matrix, whose nonzero entries are contained within κ entries of the main diagonal is equivalent to a $(2\kappa-1) \times d$ matrix whose columns are the diagonal bands of the square, banded matrix. Towards this end, if $2\kappa - 1 \leq d$ and $M = (M_{1-\kappa}, \ldots, M_0, \ldots M_{\kappa-1})$ is a $(2\kappa-1) \times d$ matrix with columns indexed from $1-\kappa$ to $\kappa - 1$ so that column zero is the middle column, we let $C_{2\kappa-1}(M)$ be the banded $d \times d$ matrix with entries given by

$$\left(C_{2\kappa-1}\left(M\right)\right)_{j,k} = \begin{cases} M_{j,k-j} & \text{if } |j-k| < \kappa \text{ or } |j-k| > d-\kappa \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

for $j, k \in [d]_0$. By construction, the columns of M are the diagonal bands of $C_{2\kappa-1}(M)$ with the middle column M_0 lying on the main diagonal. For example, in the case where $\kappa = 2$,

$$C_3\left(\begin{bmatrix} a_{0,-1} & a_{0,0} & a_{0,1} \\ a_{1,-1} & a_{1,0} & a_{1,1} \\ \vdots & \vdots & \vdots \\ a_{d-2,-1} & a_{d-2,0} & a_{d-2,1} \\ a_{d-1,-1} & a_{d-1,0} & a_{d-1,1} \end{bmatrix}\right) = \begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & 0 & a_{0,-1} \\ a_{1,-1} & a_{1,0} & a_{1,1} & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & a_{d-2,-1} & a_{d-2,0} & a_{d-2,1} \\ a_{d-1,1} & 0 & \cdots & a_{d-1,-1} & a_{d-1,0} \end{bmatrix}.$$

Below, we will state a number of lemmas, some of which are well known, which we will use in the proofs of our main results. Proofs are provided in the appendix. Our first lemma summarizes a number of properties of the discrete Fourier transform and the operators above.

Lemma 1. *For all $\mathbf{x} \in \mathbb{C}^d$ and $\ell \in [d]_0$,*

(1) $F_d\widehat{\mathbf{x}} = d\widetilde{\mathbf{x}}$,
(2) $F_d\left(W_\ell \mathbf{x}\right) = S_{-\ell}\widehat{\mathbf{x}}$,
(3) $F_d\left(S_\ell \mathbf{x}\right) = W_\ell \widehat{\mathbf{x}}$,
(4) $W_{-\ell}F_d\left(S_\ell \overline{\widetilde{\mathbf{x}}}\right) = \overline{\widehat{\mathbf{x}}}$,
(5) $\widetilde{\overline{S_\ell \mathbf{x}}} = S_{-\ell}\overline{\widetilde{\mathbf{x}}}$,
(6) $F_d\overline{\mathbf{x}} = \overline{F_d\widetilde{\mathbf{x}}}$,
(7) $\widetilde{\widehat{\mathbf{x}}} = \widehat{\widetilde{\mathbf{x}}}$,
(8) $\left|F_d\mathbf{x}\right|^2 = F_d\left(\mathbf{x} *_d \overline{\widetilde{\mathbf{x}}}\right)$.

The following lemma is the discrete analogue of the convolution theorem.

Lemma 2. *(Convolution Theorem) For all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$,*

$$F_d^{-1}\left(\widehat{\mathbf{x}} \circ \widehat{\mathbf{y}}\right) = \mathbf{x} *_d \mathbf{y},$$

and

$$\left(F_d\mathbf{x}\right) *_d \left(F_d\mathbf{y}\right) = dF_d\left(\mathbf{x} \circ \mathbf{y}\right).$$

In much of our analysis, we will have to consider the Hadamard product of a vector with a shifted copy of itself. The next three lemmas will be useful when we need to manipulate terms of that form.

Lemma 3. *Let $\mathbf{x} \in \mathbb{C}^d$, and let $\alpha, \omega \in [d]_0$. Then,*

$$\left(F_d\left(\mathbf{x} \circ S_\omega \overline{\mathbf{x}}\right)\right)_\alpha = \frac{1}{d}\mathrm{e}^{\frac{2\pi \mathrm{i}\omega\alpha}{d}}\left(F_d\left(\widehat{\mathbf{x}} \circ S_{-\alpha}\overline{\widetilde{\mathbf{x}}}\right)\right)_\omega.$$

Lemma 4. *Let $\mathbf{x} \in \mathbb{C}^d$, and let $\alpha \in \mathbb{Z}$. Then,*

$$F_d\left(\widetilde{\mathbf{x}} \circ S_{-\alpha}\overline{\widetilde{\mathbf{x}}}\right) = R(F_d(\mathbf{x} \circ S_\alpha \overline{\mathbf{x}})).$$

Lemma 5. *Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$, and let $\ell, k \in [d]_0$. Then,*

$$\left((\mathbf{x} \circ S_{-\ell}\mathbf{y}) *_d \left(\overline{\widetilde{\mathbf{x}}} \circ S_\ell \overline{\widetilde{\mathbf{y}}}\right)\right)_k = \left((\mathbf{x} \circ S_{-k}\overline{\mathbf{x}}) *_d \left(\widetilde{\mathbf{y}} \circ S_k\overline{\widetilde{\mathbf{y}}}\right)\right)_\ell.$$

For a positive integer s which divides d, we introduce the subsampling operator

$$Z_s : \mathbb{C}^d \to \mathbb{C}^{\frac{d}{s}},$$

defined by

$$(Z_s\mathbf{x})_n \coloneqq x_{ns} \text{ for } n \in \left[\frac{d}{s}\right]_0.$$

The following lemma shows that taking the Fourier transform of a subsampled vector produses an aliasing effect.

Lemma 6. *(Aliasing) Let s be a positive integer which divides d. Then for $\mathbf{x} \in \mathbb{C}^d$ and $\omega \in \left[\frac{d}{s}\right]_0$,*

$$\left(F_{\frac{d}{s}}(Z_s\mathbf{x})\right)_\omega = \frac{1}{s}\sum_{r=0}^{s-1} \widehat{x}_{\omega - r\frac{d}{s}}.$$

3. Aliased Wigner Distribution Deconvolution for Fast Phase Retrieval

As in Section 1, we let $\mathbf{x} \in \mathbb{C}^d$ denote an unknown quantity of interest and let $\mathbf{m} \in \mathbb{C}^d$ denote a known measurement mask, and consider measurements $Y_{k,\ell}$ of the form (1.5). By (1.7), we see we may write $Y_{k,\ell}$ as a noisy windowed Fourier magnitude measurement of the form

$$Y_{k,\ell} = \left|\sum_{n=0}^{d-1} x_n m_{n-\ell} \mathfrak{e}^{-\frac{2\pi \mathfrak{i} nk}{d}}\right|^2 + N_{k,\ell}, \quad \text{for } 0 \le k, \ell \le d-1. \tag{3.1}$$

Let $\mathbf{y}_\ell$ and $\mathbf{n}_\ell$ denote the ℓ-th columns of the measurement matrix $Y = (Y_{k,\ell})_{0\le k,\ell\le d-1}$ and the noise matrix $N = (N_{k,\ell})_{0\le k,\ell\le d-1}$ respectively, and, as in Section 1, let $Y_{K,L}$ be the $K \times L$ partial measurement matrix obtained by restricting Y to rows in $\frac{d}{K}[K]_0$ and columns in $\frac{d}{L}[L]_0$ so the the entries of $Y_{K,L}$ are given by (1.6), and let $N_{K,L}$ be the analogous matrix obtained by restricting N to rows and columns in $\frac{d}{K}[K]_0$ and $\frac{d}{L}[L]_0$.

Our goal is to recover $\mathbf{x}$ (up to a global phase) from these measurements with an error that may be bounded in terms of the magnitude of the noise N. Our method will be based on the following result that is an aliased and discrete variant of the Wigner Distribution Deconvolution (WDD) approach presented in the continuous setting by Chapman in [14]. Together with Lemmas 9, 10, and 11, it will allow us to recover portions of the rank one matrices $\mathbf{x}\mathbf{x}^*$ and $\widehat{\mathbf{x}}\widehat{\mathbf{x}}^*$.

Theorem 4. *Let $Y_{K,L}$ be the $K \times L$ partial measurement matrix defined in* (1.6)*, and let $N_{K,L}$ be the corresponding partial noise matrix. Let $\widetilde{Y}$ and $\widetilde{N}$ be the $L \times K$ matrices defined by*

$$\widetilde{Y} \coloneqq F_L Y_{K,L}^T F_K^T \quad \text{and} \quad \widetilde{N} \coloneqq F_L N_{K,L}^T F_K^T.$$

Then for any $\omega \in [K]_0$ and $\alpha \in [L]_0$,

$$\widetilde{Y}_{\alpha,\omega} = \frac{KL}{d^3}\sum_{r=0}^{\frac{d}{K}-1}\sum_{\ell=0}^{\frac{d}{L}-1} \left(F_d\left(\widehat{\mathbf{x}} \circ S_{\ell L-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega-rK} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha-\ell L}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega-rK} + \widetilde{N}_{\alpha,\omega} \tag{3.2}$$

$$= \frac{KL}{d^2}\sum_{r=0}^{\frac{d}{K}-1}\sum_{\ell=0}^{\frac{d}{L}-1} \mathfrak{e}^{-2\pi\mathfrak{i}(\ell L-\alpha)(\omega-rK)/d} \left(F_d\left(\widehat{\mathbf{x}} \circ S_{\ell L-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega-rK} \left(F_d\left(\mathbf{m} \circ S_{\omega-rK}\overline{\mathbf{m}}\right)\right)_{\ell L-\alpha} + \widetilde{N}_{\alpha,\omega} \tag{3.3}$$

$$= \frac{KL}{d^2}\sum_{r=0}^{\frac{d}{K}-1}\sum_{\ell=0}^{\frac{d}{L}-1} \mathfrak{e}^{2\pi\mathfrak{i}(\ell L-\alpha)(\omega-rK)/d} \left(F_d\left(\mathbf{x} \circ S_{\omega-rK}\overline{\mathbf{x}}\right)\right)_{\alpha-\ell L} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha-\ell L}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega-rK} + \widetilde{N}_{\alpha,\omega} \tag{3.4}$$

$$= \frac{KL}{d}\sum_{r=0}^{\frac{d}{K}-1}\sum_{\ell=0}^{\frac{d}{L}-1} \left(F_d\left(\mathbf{x} \circ S_{\omega-rK}\overline{\mathbf{x}}\right)\right)_{\alpha-\ell L} \left(F_d\left(\mathbf{m} \circ S_{\omega-rK}\overline{\mathbf{m}}\right)\right)_{\ell L-\alpha} + \widetilde{N}_{\alpha,\omega},. \tag{3.5}$$

To aid in the readers understanding, before proving Theorem 4, we will first give a short proof of the following lemma which is the special case of (3.5) where $K = L = d$. It is the direct analogue of Chapman's WDD approach as formulated in the continuous setting in [14].

Lemma 7. *Let Y be the $d \times d$ measurement matrix with entries defined as in* (3.1) *and let N be the corresponding noise matrix. Then, the ω-th column of $\widetilde{Y} = F_d Y^T F_d^T$ is given by*

$$\widetilde{Y}_\omega = d \cdot F_d\left(\mathbf{x} \circ S_\omega \overline{\mathbf{x}}\right) \circ R\left(F_d\left(\mathbf{m} \circ S_\omega \overline{\mathbf{m}}\right)\right) + \widetilde{N}_\omega, \tag{3.6}$$

where $\widetilde{N} = F_d N^T F_d^T$.

The Proof of Lemma 7. As noted in (3.1), we may write $\mathbf{y}_\ell$ as the STFT of $\mathbf{x}$ with window $S_{-\ell}\mathbf{m}$. Therefore, by Lemma 1, parts 5 and 8, we see that for any $\ell \in [d]_0$,

$$\begin{aligned}\mathbf{y}_\ell &= \left|F_d\left(\mathbf{x} \circ S_{-\ell}\mathbf{m}\right)\right|^2 + \eta_\ell \\ &= F_d\left(\left(\mathbf{x} \circ S_{-\ell}\mathbf{m}\right) *_d \left(\widetilde{\overline{\mathbf{x}}} \circ S_\ell \widetilde{\overline{\mathbf{m}}}\right)\right) + \eta_\ell.\end{aligned} \tag{3.7}$$

Thus, taking a Fourier transform of $\mathbf{y}_\ell$ and applying Lemma 1, part 1, yields

$$\left(F_d \mathbf{y}_\ell\right)_\omega = d\left(\left(\mathbf{x} \circ S_{-\ell}\mathbf{m}\right) *_d \left(\widetilde{\overline{\mathbf{x}}} \circ S_\ell \widetilde{\overline{\mathbf{m}}}\right)\right)_{-\omega} + \left(F_d \eta_\ell\right)_\omega,$$

and so, by Lemma 5,

$$\left(F_d \mathbf{y}_\ell\right)_\omega = d\left(\left(\mathbf{x} \circ S_\omega \overline{\mathbf{x}}\right) *_d \left(\widetilde{\mathbf{m}} \circ S_{-\omega} \widetilde{\overline{\mathbf{m}}}\right)\right)_\ell + \left(F_d \eta_\ell\right)_\omega. \tag{3.8}$$

Since $\left(F_d \mathbf{y}_\ell\right)_\omega = \left(F_d Y\right)_{\omega,\ell}$, taking the transpose of the above equation implies

$$\left(Y^T F_d^T\right)_{\ell,\omega} = d\left(\left(\mathbf{x} \circ S_\omega \overline{\mathbf{x}}\right) *_d \left(\widetilde{\mathbf{m}} \circ S_{-\omega} \widetilde{\overline{\mathbf{m}}}\right)\right)_\ell + \left(N^T F_d^T\right)_{\ell,\omega},$$

Therefore, the ω-th columns of $Y^T F_d^T$ and $N^T F_d^T$ satisfy

$$\left(Y^T F_d^T\right)_\omega = d\left(\mathbf{x} \circ S_\omega \overline{\mathbf{x}}\right) *_d \left(\widetilde{\mathbf{m}} \circ S_{-\omega} \widetilde{\overline{\mathbf{m}}}\right) + \left(N^T F_d^T\right)_\omega,$$

so, taking the Fourier transform of both sides and applying Lemmas 2 and 4 yields

$$\begin{aligned}\left(F_d Y^T F_d^T\right)_\omega &= dF_d\left(\mathbf{x} \circ S_\omega \overline{\mathbf{x}}\right) \circ F_d\left(\widetilde{\mathbf{m}} \circ S_{-\omega} \widetilde{\overline{\mathbf{m}}}\right) + \left(F_d N^T F_d^T\right)_\omega \\ &= dF_d\left(\mathbf{x} \circ S_\omega \overline{\mathbf{x}}\right) \circ R\left(F_d\left(\mathbf{m} \circ S_\omega \overline{\mathbf{m}}\right)\right) + \left(F_d N^T F_d^T\right)_\omega.\end{aligned}$$

Recalling that $\widetilde{Y} = F_d Y^T F_d^T$ and $\widetilde{N} = F_d N^T F_d^T$ completes the proof. □

The following lemma applies analysis similar to the previous lemma to subsampled column vectors using Lemma 6.

Lemma 8. *For* $\ell \in [d]_0$ *and* $\omega \in [K]_0$,

$$\left(F_K Z_{\frac{d}{K}}\left(\mathbf{y}_\ell\right)\right)_\omega = K \sum_{r=0}^{\frac{d}{K}-1} \left(\left(\mathbf{x} \circ S_{\omega - rK} \overline{\mathbf{x}}\right) *_d \left(\widetilde{\mathbf{m}} \circ S_{rK-\omega} \widetilde{\overline{\mathbf{m}}}\right)\right)_\ell + \left(F_K Z_{\frac{d}{K}}\left(\eta_\ell\right)\right)_\omega.$$

Proof. As in the proof of Lemma 7, the ℓ^{th} columns of the Y and N satisfy

$$\begin{aligned}\mathbf{y}_\ell &= \left|F_d\left(\mathbf{x} \circ S_{-\ell}\mathbf{m}\right)\right|^2 \\ &= F_d\left(\left(\mathbf{x} \circ S_{-\ell}\mathbf{m}\right) *_d \left(\widetilde{\overline{\mathbf{x}}} \circ S_\ell \widetilde{\overline{\mathbf{m}}}\right)\right) + \eta_\ell.\end{aligned}$$

Therefore, subtracting η_ℓ from both sides, taking the Fourier transform, applying Lemma 6, and then using (3.8) we see

$$\begin{aligned}\left(F_K\left(Z_{\frac{d}{K}}\left(\mathbf{y}_\ell - \eta_\ell\right)\right)\right)_\omega &= \frac{K}{d} \sum_{r=0}^{\frac{d}{K}-1} \left(F_d(\mathbf{y}_\ell - \eta_\ell)\right)_{\omega - rK} \\ &= d \cdot \frac{K}{d} \sum_{r=0}^{\frac{d}{K}-1} \left(\left(\mathbf{x} \circ S_{\omega - rK} \overline{\mathbf{x}}\right) *_d \left(\widetilde{\mathbf{m}} \circ S_{rK-\omega} \widetilde{\overline{\mathbf{m}}}\right)\right)_\ell \\ &= K \sum_{r=0}^{\frac{d}{K}-1} \left(\left(\mathbf{x} \circ S_{\omega - rK} \overline{\mathbf{x}}\right) *_d \left(\widetilde{\mathbf{m}} \circ S_{rK-\omega} \widetilde{\overline{\mathbf{m}}}\right)\right)_\ell.\end{aligned}$$

The lemma follows from the linearity of the Fourier transform and of the subsampling operator $Z_{\frac{d}{K}}$. □

Now we shall prove Theorem 4.

The Proof of Theorem 4. Noting that $Y_{K,L}$ is obtained by subsampling the rows and columns of Y, we see that the ℓ-th column of $Y_{K,L} - N_{K,L}$ is given by

$$(Y_{K,L} - N_{K,L})_\ell = \left(Z_{\frac{d}{K}}\left(\mathbf{y}_{\ell\frac{d}{L}} - \eta_{\ell\frac{d}{L}}\right)\right).$$

Therefore, applying Lemma 8 we see

$$\begin{aligned}\left(\left(Y_{K,L} - N_{K,L}\right)^T F_K^T\right)_{\ell,\omega} &= \left(F_K\left(Y_{K,L} - N_{K,L}\right)_\ell\right)_\omega \\ &= \left(F_K\left(Z_{\frac{d}{K}}\left(\mathbf{y}_{\ell\frac{d}{L}} - \eta_{\ell\frac{d}{L}}\right)\right)\right)_\omega \\ &= K\sum_{r=0}^{\frac{d}{K}-1}\left(\left(\mathbf{x}\circ S_{\omega-rK}\overline{\mathbf{x}}\right) *_d \left(\widetilde{\mathbf{m}}\circ S_{rK-\omega}\overline{\widetilde{\mathbf{m}}}\right)\right)_{\ell\frac{d}{L}} \\ &= K\left(Z_{\frac{d}{L}}\left(\sum_{r=0}^{\frac{d}{K}-1}\left(\mathbf{x}\circ S_{\omega-rK}\overline{\mathbf{x}}\right) *_d \left(\widetilde{\mathbf{m}}\circ S_{rK-\omega}\overline{\widetilde{\mathbf{m}}}\right)\right)\right)_\ell.\end{aligned}$$

Thus, the ω-th column of $\left(Y_{K,L} - N_{K,L}\right)^T F_K^T$ is given by

$$\left(\left(Y_{K,L} - N_{K,L}\right)^T F_K^T\right)_\omega = K\left(Z_{\frac{d}{L}}\left(\sum_{r=0}^{\frac{d}{K}-1}\left(\mathbf{x}\circ S_{\omega-rK}\overline{\mathbf{x}}\right) *_d \left(\widetilde{\mathbf{m}}\circ S_{rK-\omega}\overline{\widetilde{\mathbf{m}}}\right)\right)\right).$$

Taking the Fourier transform of both sides and applying Lemmas 6, 2, and 4 we see that

$$\begin{aligned}\left(F_L\left(Y_{K,L} - N_{K,L}\right)F_K^T\right)_{\alpha,\omega} &= \left(F_L\left(\left(Y_{K,L} - N_{K,L}\right)^T F_K^T\right)_\omega\right)_\alpha \\ &= \frac{KL}{d}\sum_{r=0}^{\frac{d}{K}-1}\sum_{\ell=0}^{\frac{d}{L}-1}\left(F_d\left(\mathbf{x}\circ S_{\omega-rK}\overline{\mathbf{x}}\right)\right)_{\alpha-\ell L}\left(F_d\left(\widetilde{\mathbf{m}}\circ S_{rK-\omega}\overline{\widetilde{\mathbf{m}}}\right)\right)_{\alpha-\ell L} \\ &= \frac{KL}{d}\sum_{r=0}^{\frac{d}{K}-1}\sum_{\ell=0}^{\frac{d}{L}-1}\left(F_d\left(\mathbf{x}\circ S_{\omega-rK}\overline{\mathbf{x}}\right)\right)_{\alpha-\ell L}\left(F_d\left(\mathbf{m}\circ S_{\omega-rK}\overline{\mathbf{m}}\right)\right)_{\ell L-\alpha}\end{aligned}$$

for all $\alpha \in [L]_0$.

Using the linearity of the Fourier transform and the definitions of $\widetilde{Y}$ and $\widetilde{N}$ completes the proof of (3.5). (3.2), (3.3), and (3.4) follow by using Lemma 3 to see that

$$\left(F_d\left(\mathbf{x}\circ S_{\omega-rK}\overline{\mathbf{x}}\right)\right)_{\alpha-\ell L} = \frac{1}{d}\mathfrak{e}^{-2\pi\mathfrak{i}(\ell L-\alpha)(\omega-rK)/d}\left(F_d\left(\widehat{\mathbf{x}}\circ S_{\ell L-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega-rK},$$

and

$$\left(F_d\left(\mathbf{m}\circ S_{\omega-rK}\overline{\mathbf{m}}\right)\right)_{\ell L-\alpha} = \frac{1}{d}\mathfrak{e}^{2\pi\mathfrak{i}(\ell L-\alpha)(\omega-rK)/d}\left(F_d\left(\widehat{\mathbf{m}}\circ S_{\alpha-\ell L}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega-rK}.$$

□

3.1. **Solving for Diagonal Bands of the Rank-One Matrices.** We wish to use Theorem 4 to solve for diagonal bands of the rank-one matrix $\mathbf{x}\mathbf{x}^*$. In the case where $K = L = d$, one can use (3.6) to see that for $\omega \in [d]_0$

$$\mathbf{x}\circ S_\omega\overline{\mathbf{x}} = \frac{1}{d}F_d^{-1}\left(\frac{\left(F_d Y^T F_d^T\right)_\omega}{F_d\left(\widetilde{\mathbf{m}}\circ S_{-\omega}\overline{\widetilde{\mathbf{m}}}\right)}\right) - \frac{1}{d}F_d^{-1}\left(\frac{\left(F_d N^T F_d^T\right)_\omega}{F_d\left(\widetilde{\mathbf{m}}\circ S_{-\omega}\overline{\widetilde{\mathbf{m}}}\right)}\right).$$

However, in general, the right-hand side of (3.2)-(3.5) are linear combinations of multiple terms and therefore, it is not as straightforward to solve for these diagonal bands. In this subsection, we present several lemmas which make different assumptions on the spatial and frequency supports of $\mathbf{x}$, $\mathbf{m}$, and $\widehat{\mathbf{m}}$ and identify special cases where these sums reduce to a single nonzero term. In these cases, we will then be able to solve for diagonal bands of either $\mathbf{x}\mathbf{x}^*$ or $\widehat{\mathbf{x}}\widehat{\mathbf{x}}^*$ by formulas similar to the one above. We will use Lemmas 9 and 10 in the proofs of Theorems 2 and 3. We state Lemma 11 in order to demonstrate that Wigner deconvolution

approach can also be applied to the setting considered in [33]. We will provide the proof of Lemma 10. The proofs of Lemmas 9 and 11 are nearly identical.

The first lemma in this section assumes that $\mathbf{x}$ is bandlimited and the spatial support of $\mathbf{m}$ is contained in an interval of length δ. It allows us to recover diagonal bands of the rank-one matrix $\widehat{\mathbf{x}}\widehat{\mathbf{x}}^*$.

Lemma 9. *Let $\mathbf{x}, \mathbf{m} \in \mathbb{C}^d$ with $supp\,(\widehat{\mathbf{x}}) \subseteq [\gamma]_0$ and $supp\,(\mathbf{m}) \subseteq [\delta]_0$. Let K and L divide d, and let $Y_{K,L}$ be the $K \times L$ partial measurement matrix defined as in (1.6) and let $N_{K,L}$ be the corresponding subsampled noise matrix. As in the statement of Theorem 4, let*

$$\widetilde{Y} = F_L Y_{K,L} F_K^T \quad \text{and} \quad \widetilde{N} = F_L N_{K,L} F_K^T.$$

Then for any $\alpha \in [L]_0$ and $\omega \in [K]_0$,

$$\begin{aligned}\widetilde{Y}_{\alpha,\omega} &= \frac{KL}{d^2} \sum_{r=0}^{\frac{d}{K}-1} \sum_{\ell=0}^{\frac{d}{L}-1} \mathrm{e}^{\frac{2\pi \mathrm{i}}{d}(\omega - rK)(\alpha - \ell L)} \left(F_d\left(\widehat{\mathbf{x}} \circ S_{\ell L - \alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega - rK} \left(F_d\left(\mathbf{m} \circ S_{\omega - rK}\overline{\mathbf{m}}\right)\right)_{\ell L - \alpha} + \widetilde{N}_{\alpha,\omega} \\ &= \frac{KL}{d^3} \sum_{r=0}^{\frac{d}{K}-1} \sum_{\ell=0}^{\frac{d}{L}-1} \left(F_d\left(\widehat{\mathbf{x}} \circ S_{\ell L - \alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega - rK} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha - \ell L}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega - rK} + \widetilde{N}_{\alpha,\omega}.\end{aligned}$$

Moreover, if $K = \delta - 1 + \kappa$ for some $2 \le \kappa \le \delta$ and $L = \gamma - 1 + \xi$ for some $1 \le \xi \le \gamma$, and if $0 \le \omega \le \kappa - 1$ or $K - \kappa - 1 \le \omega \le K - 1$ and $0 \le \alpha \le \xi - 1$ or $L - \xi + 1 \le \alpha \le L - 1$, the sum above collapses to only one term, so that

$$\widetilde{Y}_{\alpha,\omega} = \begin{cases} \frac{KL}{d^3}\left(F_d\left(\widehat{\mathbf{x}} \circ S_{-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega} + \widetilde{N}_{\alpha,\omega} & \text{if } 0 \le \alpha \le \xi - 1 \text{ and } 0 \le \omega \le \kappa - 1 \\ \frac{KL}{d^3}\left(F_d\left(\widehat{\mathbf{x}} \circ S_{-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega - K} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega - K} + \widetilde{N}_{\alpha,\omega} & \text{if } 0 \le \alpha \le \xi - 1 \text{ and } \delta \le \omega \le K - 1 \\ \frac{KL}{d^3}\left(F_d\left(\widehat{\mathbf{x}} \circ S_{L-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha - L}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega} + \widetilde{N}_{\alpha,\omega} & \text{if } \gamma \le \alpha \le L - 1 \text{ and } 0 \le \omega \le \kappa - 1 \\ \frac{KL}{d^3}\left(F_d\left(\widehat{\mathbf{x}} \circ S_{L-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega - K} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha - L}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega - K} + \widetilde{N}_{\alpha,\omega} & \text{if } \gamma \le \alpha \le L - 1 \text{ and } \delta \le \omega \le K - 1 \end{cases}.$$

The next lemma is similar to the previous one, but replaces the assumptions that $\mathbf{m}$ has compact spatial support and that $\mathbf{x}$ is bandlimited and with the assumption that $\mathbf{m}$ is bandlimited. It allows us to recover diagonals of $\mathbf{x}\mathbf{x}^*$.

Lemma 10. *Let $\mathbf{x}, \mathbf{m} \in \mathbb{C}^d$ and assume $supp\,(\widehat{\mathbf{m}}) \subseteq [\rho]_0$. Let L divide d, let $Y_{d,L}$ be the $d \times L$ partial measurement matrix defined as in (1.6), and let $N_{d,L}$ be the corresponding partial noise matrix. As in the statement of Theorem 4, let*

$$\widetilde{Y} = F_L Y_{d,L} F_d^T \quad \text{and} \quad \widetilde{N} = F_L N_{d,L} F_d^T.$$

Then for any $\alpha \in [L]_0$ and $\omega \in [d]_0$,

$$\widetilde{Y}_{\alpha,\omega} = \frac{L}{d^2} \sum_{\ell=0}^{\frac{d}{L}-1} \left(F_d\left(\widehat{\mathbf{x}} \circ S_{\ell L - \alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha - \ell L}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega} + \widetilde{N}_{\alpha,\omega}. \tag{3.9}$$

Moreover, if $L = \rho + \kappa - 1$ for some $2 \le \kappa \le \rho$, then for all $\omega \in [d]_0$ and all α such that either $0 \le \alpha \le \kappa - 1$ or $\rho \le \alpha \le L - 1$, the sum above reduces to a single term and

$$\widetilde{Y}_{\alpha,\omega} = \begin{cases} \frac{L}{d^2}\left(F_d\left(\widehat{\mathbf{x}} \circ S_{-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega} + \widetilde{N}_{\alpha,\omega}, & \text{if } 0 \le \alpha \le \kappa - 1 \\ \frac{L}{d^2}\left(F_d\left(\widehat{\mathbf{x}} \circ S_{L-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha - L}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega} + \widetilde{N}_{\alpha,\omega}, & \text{if } \rho \le \alpha \le L - 1 \end{cases}.$$

Proof. (3.9) follows from Theorem 4 by setting $K = d$ in (3.2). To prove the second claim, we note that by the assumption that $\mathrm{supp}\,(\widehat{\mathbf{m}}) \subseteq [\rho]_0$, $\widehat{\mathbf{m}} \circ S_{\alpha - \ell L}\overline{\widehat{\mathbf{m}}} = \mathbf{0}$ unless

$$|\alpha - \ell L| < \rho. \tag{3.10}$$

If $L = \rho - 1 + \kappa$, and $0 \le \alpha \le \kappa - 1$, this can only occur if $\ell = 0$. Indeed, if $\ell \ge 1$ then

$$\alpha - \ell L \le \alpha - L \le \kappa - 1 - (\rho - 1 + \kappa) = -\rho,$$

and if $\ell \le -1$, then

$$\alpha - \ell L \ge \alpha + L \ge L = \rho - 1 + \kappa \ge \rho.$$

Therefore, all other terms in the above sum are zero, and the right-hand side of (3.9) reduces to the desired result. Likewise, if $\rho \leq \alpha \leq L-1$, then (3.10) can only hold when $\ell = 1$. □

As in Lemma 9, the following lemma assumes that the spatial support of $\mathbf{m}$ is contained in an interval of length δ and allows us to recover diagonals of $\widehat{\mathbf{x}}\widehat{\mathbf{x}}^*$. However, it differs in that it assumes that $L = d$, but does not assume that that $\mathbf{x}$ is γ-bandlimited.

Lemma 11. *Let $\mathbf{x}, \mathbf{m} \in \mathbb{C}^d$ with $\text{supp}\,(\mathbf{m}) \subseteq [\delta]_0$. Let K divide d, let $Y_{K,d}$ be the $K \times d$ partial measurement matrix defined as in (1.6), and let $N_{K,d}$ be the corresponding subsampled noise matrix. As in the statement of Theorem 4, let*

$$\widetilde{Y} = F_d Y_{K,d} F_K^T \quad \text{and} \quad \widetilde{N} = F_d N_{K,d} F_K^T.$$

Then for any $\alpha \in [d]_0$ and $\omega \in [K]_0$,

$$\begin{aligned}\widetilde{Y}_{\alpha,\omega} &= \frac{K}{d^2} \sum_{r=0}^{\frac{d}{K}-1} \left(F_d\left(\widehat{\mathbf{x}} \circ S_{-\alpha}\overline{\widehat{\mathbf{x}}}\right)\right)_{\omega - rK} \left(F_d\left(\widehat{\mathbf{m}} \circ S_{\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_{\omega - rK} + \widetilde{N}_{\alpha,\omega} \\ &= K \sum_{r=0}^{\frac{d}{K}-1} \left(F_d\left(\mathbf{x} \circ S_{\omega - rK}\overline{\mathbf{x}}\right)\right)_{\alpha} \left(F_d\left(\mathbf{m} \circ S_{\omega - rK}\overline{\mathbf{m}}\right)\right)_{-\alpha} + \widetilde{N}_{\alpha,\omega}.\end{aligned}$$

Moreover, if $K = \delta - 1 + \kappa$ for some $2 \leq \kappa \leq \delta$, and if either $0 \leq \omega \leq \kappa - 1$ or $\delta \leq \omega \leq K-1$, then for all $\alpha \in [d]_0$, the sum above reduces to only one term and

$$\widetilde{Y}_{\alpha,\omega} = \begin{cases} K\left(F_d\left(\mathbf{x} \circ S_{\omega}\overline{\mathbf{x}}\right)\right)_{\alpha} \left(F_d\left(\mathbf{m} \circ S_{\omega}\overline{\mathbf{m}}\right)\right)_{-\alpha} + \widetilde{N}_{\alpha,\omega}, & \text{if } 0 \leq \omega \leq \kappa - 1 \\ K\left(F_d\left(\mathbf{x} \circ S_{\omega - K}\overline{\mathbf{x}}\right)\right)_{\alpha} \left(F_d\left(\mathbf{m} \circ S_{\omega - K}\overline{\mathbf{m}}\right)\right)_{-\alpha} + \widetilde{N}_{\alpha,\omega}, & \text{if } \delta \leq \omega \leq K - 1 \end{cases}.$$

Remark 1. For convenience, in Lemmas 9, 10, and 11 we have assumed that the support of $\widehat{\mathbf{m}}, \mathbf{m}$ or $\widehat{\mathbf{x}}$, were contained in the first ρ, δ, or γ entries. However, inspecting the proofs we see these results remain valid if these intervals are replaced with any other intervals of the same length.

4. Recovery Guarantees

In this section, we will present two algorithms which allow us to reconstruct $\mathbf{x}$ from our matrix of noisy measurements $Y_{K,L}$ and prove Theorems 2 and 3, presented in the introduction, which guarantee that these algorithms converge. Before providing the proofs of these theorems, we will first state two propositions which show that it is possible to design masks in such a way that the mask dependent constants μ_1 and μ_2 are nonzero. For proofs of these propositions, please see the appendix.

Proposition 1. Let $\mathbf{m} \in \mathbb{C}^d$ be bandlimited with $\text{supp}\,(\widehat{\mathbf{m}}) \subseteq [\rho]_0$, so that its Fourier transform may be written as

$$\widehat{\mathbf{m}} = \left(a_0 \mathbb{e}^{\mathbb{i}\theta_0}, \ldots, a_{\rho-1}\mathbb{e}^{\mathbb{i}\theta_{\rho-1}}, 0, \ldots, 0\right)^T$$

for some real numbers $a_0, \ldots, a_{\rho-1}$. As in (1.15), let

$$\mu_1 = \min_{\substack{|p| \leq \kappa - 1 \\ q \in [d]_0}} \left| F_d\left(\widehat{\mathbf{m}} \circ S_p \overline{\widehat{\mathbf{m}}}\right)_q \right|,$$

for some $2 \leq \kappa \leq \rho$. If

$$|a_0| > (\rho - 1)\,|a_1|, \tag{4.1}$$

and

$$|a_1| \geq |a_2| \geq \cdots \geq |a_{\rho-1}| > 0, \tag{4.2}$$

then $\mu_1 > 0$.

Proposition 2. Let $\mathbf{m} \in \mathbb{C}^d$ be a compactly supported mask with $\text{supp}\,(\mathbf{m}) \subseteq [\delta]_0$, given by

$$\mathbf{m} = \left(a_0 \mathbb{e}^{\mathbb{i}\theta_0}, \ldots, a_{\delta-1}\mathbb{e}^{\mathbb{i}\theta_{\delta-1}}, 0, \ldots, 0\right)^T$$

Algorithm 1 Wigner Deconvolution and Angular Synchronization for Bandlimited Masks

Inputs

(1) $d \times L$ noisy measurement matrix $Y_{d,L} \in \mathbb{R}^{d\times L}$ with entries

$$(Y_{d,L})_{k,\ell} = \left|\sum_{n=0}^{d-1} x_n m_{n-\ell\frac{L}{d}} \mathfrak{e}^{-\frac{2\pi \mathfrak{i} nk}{d}}\right|^2 + (N_{d,L})_{k,\ell}\,,\ k \in [d]_0, \ell \in \frac{d}{L}\,[L]_0\,.$$

(2) Bandlimited mask $\mathbf{m} \in \mathbb{C}^d$ with $\operatorname{supp}(\widehat{\mathbf{m}}) \subseteq [\rho]_0$ for some $\rho < \frac{d}{2}$.

Steps

(1) Let $\kappa = L - \rho + 1$, and for $1-\kappa \le \alpha \le \kappa - 1$ estimate $F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right)$ by

$$F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right) \approx \begin{cases} \frac{d^2\left(F_d Y_{d,L} F_L^T\right)_{-\alpha}}{LF_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)} & \text{if } 1-\kappa \le \alpha \le 0 \\ \frac{d^2\left(F_d Y_{d,L} F_L^T\right)_{L-\alpha}}{LF_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)} & \text{if } 1 \le \alpha \le \kappa - 1 \end{cases}.$$

(2) Invert the Fourier transforms above to recover estimates of the $(2\kappa-1)$ vectors $\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}$.
(3) Organize these vectors into a banded matrix, $C_{2\kappa-1}(Y_{2\kappa-1})$ as described in (4.7) (see also (2.1)).
(4) Hermitianize the matrix above: $C_{2\kappa-1}(Y_{2\kappa-1}) \leftarrowtail \frac{1}{2}\left(C_{2\kappa-1}(Y_{2\kappa-1}) + C_{2\kappa-1}(Y_{2\kappa-1})^*\right)$.
(5) Estimate $|\widehat{\mathbf{x}}|$ from the main diagonal of $C_{2\kappa-1}(Y_{2\kappa-1})$.
(6) Normalize $C_{2\kappa-1}(Y_{2\kappa-1})$ componentwise to form $\widetilde{Y}_{2\kappa-1}$.
(7) Compute $\mathbf{v_1}$ the leading normalized eigenvector of $\widetilde{Y}_{2\kappa-1}$.

Output

$\mathbf{x}_e \coloneqq F_d^{-1}\widehat{\mathbf{x}}_e$, an estimate of $\mathbf{x}$, where $\widehat{\mathbf{x}}_e$ is given componentwise by

$$(\widehat{\mathbf{x}}_e)_j \coloneqq \sqrt{(C_{2\kappa-1}(Y_{2\kappa-1}))_{j,j}}\,(\mathbf{v_1})_j\,.$$

for some real numbers $a_0, \dots, a_{\delta-1}$. As in (1.12), let

$$\mu_2 = \min_{\substack{|p|\le\gamma-1 \\ |q|\le\delta-1}} \left|F_d\left(\widehat{\mathbf{m}} \circ S_p\overline{\widehat{\mathbf{m}}}\right)_q\right|$$

for some $1 \le \gamma \le 2\delta - 1$. If

$$|a_0| > (\delta - 1)\,|a_1|, \tag{4.3}$$

and

$$|a_1| \ge |a_2| \ge \cdots \ge |a_{\delta-1}| > 0, \tag{4.4}$$

then $\mu_2 > 0$.

We will now prove Theorem 3, which we restate below for the convenience of the reader.

Theorem 3 (Convergence Gaurantees for Algorithm 1)**.** *Let $\mathbf{x}, \mathbf{m} \in \mathbb{C}^d$ with $\operatorname{supp}(\widehat{\mathbf{m}}) \subseteq [\rho]_0$, for some $\rho < d/2$. Let $\min|\widehat{\mathbf{x}}| \coloneqq \min_{0\le n\le d-1}|\widehat{x}_n| > 0$, and let*

$$\mu_1 \coloneqq \min_{\substack{|p|\le\gamma-1 \\ |q|\le\rho-1}} \left|F_d\left(\widehat{\mathbf{m}} \circ S_p\overline{\widehat{\mathbf{m}}}\right)_q\right| \ > \ 0. \tag{1.15}$$

Fix an integer $\kappa \in [2, \rho]$ and assume that $L = \rho + \kappa - 1$ divides d. Then, when $K = d$, Algorithm 1 in Section 4 will output $\mathbf{x}_e$, an estimate of $\mathbf{x}$, such that

$$\min_{\phi\in[0,2\pi]} \left\|\mathbf{x} - \mathfrak{e}^{\mathfrak{i}\phi}\mathbf{x}_e\right\|_2 \le C\frac{d^{7/2}\,\|\widehat{\mathbf{x}}\|_\infty\,\|N_{d,L}\|_F}{L^{\frac{1}{2}}\mu_1\kappa^{\frac{5}{2}}\cdot\min|\widehat{\mathbf{x}}|^2} + C'\frac{d^{\frac{3}{2}}}{L^{\frac{1}{4}}}\sqrt{\frac{\|N_{d,L}\|_F}{\mu_1}} \tag{1.16}$$

for some absolute constants $C, C' \in \mathbb{R}^+$. Furthermore, if $\|N_{d,L}\|_F$ is sufficently small, then Algorithm 1 is always guaranteed to require just $\mathcal{O}\left(d(\rho+\kappa^2)\log d\right)$ total flops to achieve (1.16) up to machine precision.

The Proof of Theorem 3. Let $\mathbf{x}, \mathbf{m} \in \mathbb{C}^d$, $\operatorname{supp}(\widehat{\mathbf{m}}) = [\rho]_0$, and let $\kappa = L - \rho + 1$. Then by Lemma 10, if $0 \leq \beta \leq \kappa - 1$, then

$$F_d\left(\widehat{\mathbf{x}} \circ S_{-\beta}\overline{\widehat{\mathbf{x}}}\right) = \frac{d^2}{L}\frac{\left(F_d Y_{d,L} F_L^T\right)_\beta}{F_d\left(\widehat{\mathbf{m}} \circ S_\beta \overline{\widehat{\mathbf{m}}}\right)} - \frac{d^2}{L}\frac{\left(F_d N_{d,L} F_L^T\right)_\beta}{F_d\left(\widehat{\mathbf{m}} \circ S_\beta \overline{\widehat{\mathbf{m}}}\right)},$$

and therefore

$$\widehat{\mathbf{x}} \circ S_{-\beta}\overline{\widehat{\mathbf{x}}} + \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d N_{d,L} F_L^T\right)_\beta}{F_d\left(\widehat{\mathbf{m}} \circ S_\beta \overline{\widehat{\mathbf{m}}}\right)}\right) = \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d Y_{d,L} F_L^T\right)_\beta}{F_d\left(\widehat{\mathbf{m}} \circ S_\beta \overline{\widehat{\mathbf{m}}}\right)}\right).$$

Substituting $\alpha = -\beta$, we see

$$\widehat{\mathbf{x}} \circ S_{\alpha}\overline{\widehat{\mathbf{x}}} + \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d N_{d,L} F_L^T\right)_{-\alpha}}{F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha} \overline{\widehat{\mathbf{m}}}\right)}\right) = \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d Y_{d,L} F_L^T\right)_{-\alpha}}{F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha} \overline{\widehat{\mathbf{m}}}\right)}\right) \tag{4.5}$$

for all $1 - \kappa \leq \alpha \leq 0$. Likewise, for $\rho \leq \beta \leq L - 1$,

$$\widehat{\mathbf{x}} \circ S_{L-\beta}\overline{\widehat{\mathbf{x}}} + \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d N_{d,L} F_L^T\right)_\beta}{F_d\left(\widehat{\mathbf{m}} \circ S_{\beta-L} \overline{\widehat{\mathbf{m}}}\right)}\right) = \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d Y_{d,L} F_L^T\right)_\beta}{F_d\left(\widehat{\mathbf{m}} \circ S_{\beta-L} \overline{\widehat{\mathbf{m}}}\right)}\right),$$

so, since $L = \rho + \kappa - 1$, substituting $\alpha = L - \beta$ implies

$$\widehat{\mathbf{x}} \circ S_{\alpha}\overline{\widehat{\mathbf{x}}} + \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d N_{d,L} F_L^T\right)_{L-\alpha}}{F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha} \overline{\widehat{\mathbf{m}}}\right)}\right) = \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d Y_{d,L} F_L^T\right)_{L-\alpha}}{F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha} \overline{\widehat{\mathbf{m}}}\right)}\right) \tag{4.6}$$

for all $1 \leq \alpha \leq \kappa - 1$.

In order to write the equations above in a compact form, we will construct three $d \times 2\kappa - 1$ matrices, $X_{2\kappa-1}, N_{2\kappa-1}$, and $Y_{2\kappa-1}$. As in Section 2, for notatational convenience, we will index the columns of these matrices from $-\kappa + 1$ to $\kappa - 1$ so that column zero is the middle column. For $-\kappa + 1 \leq \alpha \leq \kappa - 1$, we let the α-th column of $X_{2\kappa-1}$ be the diagonal band of $\widehat{\mathbf{x}}\widehat{\mathbf{x}}^*$ which is α terms off of the main diagonal, i.e.

$$(X_{2\kappa-1})_\alpha = \widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}},$$

and we define the columns of $N_{2\kappa-1}$ and $Y_{2\kappa-1}$ by

$$(N_{2\kappa-1})_\alpha = \begin{cases} \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d N_{d,L} F_L^T\right)_{-\alpha}}{F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha} \overline{\widehat{\mathbf{m}}}\right)}\right) & \text{if } -\kappa + 1 \leq \alpha \leq 0 \\ \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d N_{d,L} F_L^T\right)_{L-\alpha}}{F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha} \overline{\widehat{\mathbf{m}}}\right)}\right) & \text{if } 1 \leq \alpha \leq \kappa - 1 \end{cases},$$

and

$$(Y_{2\kappa-1})_\alpha = \begin{cases} \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d Y_{d,L} F_L^T\right)_{-\alpha}}{F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha} \overline{\widehat{\mathbf{m}}}\right)}\right) & \text{if } -\kappa + 1 \leq \alpha \leq 0 \\ \frac{d^2}{L}F_d^{-1}\left(\frac{\left(F_d Y_{d,L} F_L^T\right)_{L-\alpha}}{F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha} \overline{\widehat{\mathbf{m}}}\right)}\right) & \text{if } 1 \leq \alpha \leq \kappa - 1 \end{cases}. \tag{4.7}$$

By construction, (4.5) and (4.6) imply

$$Y_{2\kappa-1} = X_{2\kappa-1} + N_{2\kappa-1}. \tag{4.8}$$

Using the fact $\frac{1}{\sqrt{d}}F_d$ is unitary, we see

$$\begin{aligned} \|N_{2\kappa-1}\|_F^2 &\leq \frac{d^4}{L^2} \cdot \frac{1}{d}\frac{\left\|F_d N_{d,L} F_L^T\right\|_F^2}{\mu_1^2} \\ &\leq \frac{d^4}{L\mu_1^2}\|N_{d,L}\|_F^2, \end{aligned} \tag{4.9}$$

where μ_1 is as in (1.15). Let $H : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ be the Hermitianizing operator

$$H(M) = \frac{M + M^*}{2}, \tag{4.10}$$

and note that $\|H(M)\|_F \leq \|M\|_F$. Since operator $C_{2\kappa-1}$ defined in (2.1) is linear and $C_{2\kappa-1}(X_{2\kappa-1})$ is Hermitian, (4.8) implies

$$C_{2\kappa-1}(X_{2\kappa-1}) = H(C_{2\kappa-1}(Y_{2\kappa-1})) - H(C_{2\kappa-1}(N_{2\kappa-1})).$$

Let $\operatorname{sgn}:\mathbb{C}\to\mathbb{C}$ be the signum function,

$$\operatorname{sgn}(z) \coloneqq \begin{cases} \frac{z}{|z|}, & \text{if } z\neq 0 \\ \text{if } 1, & z=0 \end{cases},$$

and let $\widetilde{X}_{2\kappa-1}$ and $\widetilde{Y}_{2\kappa-1}$, be the (componentwise) normalized versions of $C_{2\kappa-1}(X_{2\kappa-1})$ and $H(C_{2\kappa-1}(Y_{2\kappa-1}))$, respectively, i.e.,

$$\widetilde{X}_{2\kappa-1} \coloneqq \operatorname{sgn}(C_{2\kappa-1}((X_{2\kappa-1}))), \quad \text{and} \quad \widetilde{Y}_{2\kappa-1} \coloneqq \operatorname{sgn}(HC_{2\kappa-1}((Y_{2\kappa-1}))),$$

and note that

$$\left(\widetilde{Y}_{2\kappa-1}\right)_{j,k} = \operatorname{sgn}\left(\left(H(C_{2\kappa-1}(Y_{2\kappa-1}))\right)_{j,k}\right) = \operatorname{sgn}\left(\frac{\left(H\left(\left(C_{2\kappa-1}(Y_{2\kappa-1})\right)\right)\right)_{j,k}}{\left|\left(C_{2\kappa-1}(X_{2\kappa-1})\right)_{j,k}\right|}\right). \tag{4.11}$$

For all j and k, we have that

$$\frac{1}{\left|\left(C_{2\kappa-1}(X_{2\kappa-1})\right)_{j,k}\right|} \leq \frac{1}{\min|\widehat{\mathbf{x}}|^2}.$$

Therefore, we can apply (4.11) and the fact that for all $z_1, z_2 \in \mathbb{C}$,

$$\left|\frac{z_2}{|z_1|} - \operatorname{sgn}\left(\frac{z_2}{|z_1|}\right)\right| = \left|\frac{z_2}{|z_1|} - \operatorname{sgn}(z_2)\right| \leq \left|\operatorname{sgn}(z_1) - \frac{z_2}{|z_1|}\right|$$

to see

$$\begin{aligned}
&\left|\left(\widetilde{X}_{2\kappa-1}\right)_{j,k} - \left(\widetilde{Y}_{2\kappa-1}\right)_{j,k}\right| \\
=&\left|\left(\widetilde{X}_{2\kappa-1}\right)_{j,k} - \operatorname{sgn}\left(\frac{\left(H(C_{2\kappa-1}(Y_{2\kappa-1}))\right)_{j,k}}{\left|\left(C_{2\kappa-1}(X_{2\kappa-1})\right)_{j,k}\right|}\right)\right| \\
\leq&\left|\left(\widetilde{X}_{2\kappa-1}\right)_{j,k} - \frac{\left(H(C_{2\kappa-1}(Y_{2\kappa-1}))\right)_{j,k}}{\left|\left(C_{2\kappa-1}(X_{2\kappa-1})\right)_{j,k}\right|}\right| + \left|\frac{\left(H(C_{2\kappa-1}(Y_{2\kappa-1}))\right)_{j,k}}{\left|\left(C_{2\kappa-1}(X_{2\kappa-1})\right)_{j,k}\right|} - \operatorname{sgn}\left(\frac{\left(H(C_{2\kappa-1}(Y_{2\kappa-1}))\right)_{j,k}}{\left|\left(C_{2\kappa-1}(X_{2\kappa-1})\right)_{j,k}\right|}\right)\right| \\
\leq&2\left|\left(\widetilde{X}_{2\kappa-1}\right)_{j,k} - \frac{\left(H(C_{2\kappa-1}(Y_{2\kappa-1}))\right)_{j,k}}{\left|\left(C_{2\kappa-1}(X_{2\kappa-1})\right)_{j,k}\right|}\right| \\
=&2\frac{\left|\left(H(C_{2\kappa-1}(N_{2\kappa-1}))\right)_{j,k}\right|}{\left|\left(C_{2\kappa-1}(X_{2\kappa-1})\right)_{j,k}\right|} \\
\leq&\frac{2}{\min|\widehat{\mathbf{x}}|^2}\left|\left(H(C_{2\kappa-1}(N_{2\kappa-1}))\right)_{j,k}\right|,
\end{aligned} \tag{4.12}$$

for all $j,k\in[d]_0$. Thus, by (4.9) and the fact that $\left\|\widetilde{X}_{2\kappa-1}\right\|_F = \sqrt{(2\kappa-1)\,d}$, we see that

$$\begin{aligned}
\left\|\widetilde{Y}_{2\kappa-1} - \widetilde{X}_{2\kappa-1}\right\|_F &\leq \frac{2}{\min|\widehat{\mathbf{x}}|^2}\|H(C_{2\kappa-1}(N_{2\kappa-1}))\|_F \\
&\leq \frac{2}{\min|\widehat{\mathbf{x}}|^2}\|N_{2\kappa-1}\|_F \\
&\leq \frac{2}{\mu_1\min|\widehat{\mathbf{x}}|^2}\frac{d^2}{\sqrt{L}}\|N_{d,L}\|_F \\
&\leq C\frac{d^{3/2}\|N_{d,L}\|_F}{\mu_1\min|\widehat{\mathbf{x}}|^2\sqrt{\kappa L}}\|X_{2\kappa-1}\|_F.
\end{aligned}$$

Therefore, by Corollary 2 of [33], we have

$$\min_{\phi\in[0,2\pi]} \left\|\operatorname{sgn}(\widehat{\mathbf{x}}) - \mathrm{e}^{\mathrm{i}\phi}\operatorname{sgn}(\mathbf{v_1})\right\|_2 \leq C\frac{d^{3/2}\|N_{d,L}\|_F}{\mu_1 \min|\widehat{\mathbf{x}}|^2\sqrt{\kappa L}}\frac{d^{\frac{5}{2}}}{\kappa^2}$$
$$= C\frac{d^4\|N_{d,L}\|_F}{L^{1/2}\mu_1\kappa^{5/2}\min|\widehat{\mathbf{x}}|^2}, \tag{4.13}$$

where $\frac{\widehat{\mathbf{x}}}{|\widehat{\mathbf{x}}|}$ is the vector of true phases of $\widehat{\mathbf{x}}$, and $\mathbf{v_1}$ is the lead eigenvector of $\widetilde{Y}_{2\kappa-1}$.

As in Algorithm 1, define $\widehat{\mathbf{x}}_e$ by

$$(\widehat{\mathbf{x}}_e)_j = \sqrt{(C_{2\kappa-1}(Y_{2\kappa-1}))_{j,j}}\cdot\operatorname{sgn}(\mathbf{v_1})_j$$

for $j\in[d]_0$. Lemma 3 of [34] implies that

$$\left\|\,|\widehat{\mathbf{x}}| - |\widehat{\mathbf{x}}_e|\,\right\|_\infty^2 \leq C\|N_{2\kappa-1}\|_\infty,$$

and so

$$\left\|\,|\widehat{\mathbf{x}}| - |\widehat{\mathbf{x}}_e|\,\right\|_2 \leq C\sqrt{d\|N_{2\kappa-1}\|_\infty} \leq C\sqrt{d\|N_{2\kappa-1}\|_F}.$$

Therefore, by (4.9), we see

$$\begin{aligned}
\min_{\phi\in[0,2\pi]}\left\|\widehat{\mathbf{x}} - \mathrm{e}^{\mathrm{i}\phi}\widehat{\mathbf{x}}_e\right\|_2 &= \min_{\phi\in[0,2\pi]}\left\||\widehat{\mathbf{x}}|\circ\operatorname{sgn}(\widehat{\mathbf{x}}) - |\widehat{\mathbf{x}}_e|\circ\mathrm{e}^{\mathrm{i}\phi}\operatorname{sgn}(\widehat{\mathbf{x}}_e)\right\|_2 \\
&\leq \min_{\phi\in[0,2\pi]}\left(\left\||\widehat{\mathbf{x}}|\circ\operatorname{sgn}(\widehat{\mathbf{x}}) - |\widehat{\mathbf{x}}|\circ\mathrm{e}^{\mathrm{i}\phi}\operatorname{sgn}(\widehat{\mathbf{x}}_e)\right\|_2 + \left\||\widehat{\mathbf{x}}|\circ\mathrm{e}^{\mathrm{i}\phi}\operatorname{sgn}(\widehat{\mathbf{x}}_e) - |\widehat{\mathbf{x}}_e|\circ\mathrm{e}^{\mathrm{i}\phi}\operatorname{sgn}(\widehat{\mathbf{x}}_e)\right\|_2\right). \\
&= \min_{\phi\in[0,2\pi]}\left(\left\||\widehat{\mathbf{x}}|\circ\operatorname{sgn}(\widehat{\mathbf{x}}) - |\widehat{\mathbf{x}}|\circ\mathrm{e}^{\mathrm{i}\phi}\operatorname{sgn}(\widehat{\mathbf{x}}_e)\right\|_2\right) + \left\|\,|\widehat{\mathbf{x}}| - |\widehat{\mathbf{x}}_e|\,\right\|_2 \\
&\leq \|\widehat{\mathbf{x}}\|_\infty\left(\min_{\phi\in[0,2\pi]}\left\|\operatorname{sgn}(\widehat{\mathbf{x}}) - \mathrm{e}^{\mathrm{i}\phi}\operatorname{sgn}(\widehat{\mathbf{x}}_e)\right\|_2\right) + C\sqrt{d\,\|N_{2\kappa-1}\|_F} \\
&\leq \|\widehat{\mathbf{x}}\|_\infty\left(\min_{\phi\in[0,2\pi]}\left\|\operatorname{sgn}(\widehat{\mathbf{x}}) - \mathrm{e}^{\mathrm{i}\phi}\operatorname{sgn}(\widehat{\mathbf{x}}_e)\right\|_2\right) + C\sqrt{\frac{d^3}{\sqrt{L}\mu_1}\,\|N_{d,L}\|_F}.
\end{aligned}$$

Together with (4.13) this yields

$$\min_{\phi\in[0,2\pi]}\left\|\widehat{\mathbf{x}} - \mathrm{e}^{\mathrm{i}\phi}\widehat{\mathbf{x}}_e\right\|_2 \leq C\frac{d^4\,\|\widehat{\mathbf{x}}\|_\infty\,\|N_{d,L}\|_F}{L^{\frac{1}{2}}\mu_1\kappa^{\frac{5}{2}}\min|\widehat{\mathbf{x}}|^2} + C'\frac{d^{3/2}}{L^{1/4}}\sqrt{\frac{\|N_{d,L}\|_F}{\mu_1}}.$$

(1.16) now follows from the fact that $\|\widehat{\mathbf{x}}\|_2 = \sqrt{d}\|\mathbf{x}\|_2$ for all $\mathbf{x}\in\mathbb{C}^d$. □

Theorem 2, restated below, provides recovery guarantees for Algorithm 2 under the assumptions that $\mathbf{m}$ is compactly supported in space and that $\mathbf{x}$ is bandlimited. The proof is somewhat similar to the proof of Theorem 3 but uses Lemma 9 in place of Lemma 10 and uses the Lemma 8 of [33] during the angular synchronization step.

Theorem 2 (Convergence Gaurantees for Algorithm 2)**.** *Let $\mathbf{x},\mathbf{m}\in\mathbb{C}^d$ with $\operatorname{supp}(\widehat{\mathbf{x}})\subseteq[\gamma]_0$ and $\operatorname{supp}(\mathbf{m})\subseteq[\delta]_0$, and let*

$$\mu_2 = \min_{\substack{|p|\leq\gamma-1,\\ |q|\leq\delta-1}}\left|F_d\left(\widehat{\mathbf{m}}\circ S_p\overline{\widehat{\mathbf{m}}}\right)_q\right| \;>\; 0. \tag{1.12}$$

Assume that $\gamma\leq 2\delta-1<d$, $L=2\gamma-1$, $K=2\delta-1$, and also that K and L divide d. Furthermore, suppose that the phaseless measurements (1.6) *have noise dominated by the norm of $\mathbf{x}$ so that*

$$\|N_{K,L}\|_F \leq \beta\,\|\mathbf{x}\|_2^2 \tag{1.13}$$

for some $\beta\geq 0$. Then Algorithm 2 in Section 4 outputs an estimate $\mathbf{x}_e$ to $\mathbf{x}$ with relative error

$$\min_{\phi\in[0,2\pi]}\frac{\left\|\mathbf{x}-\mathrm{e}^{\mathrm{i}\phi}\mathbf{x}_e\right\|_2}{\|\mathbf{x}\|_2} \leq \frac{\left(1+2\sqrt{2}\right)\beta}{\sigma_\gamma(W)}\frac{d^2}{\sqrt{KL}\mu_2}, \tag{1.14}$$

Algorithm 2 Wigner Deconvolution and Angular Synchronization for Bandlimited Signals

Inputs

(1) $K \times L$ noisy measurement matrix, $Y_{K,L} \in \mathbb{R}^{K\times L}$, with entries

$$(Y_{K,L})_{k,\ell} = \left|\sum_{n=0}^{d-1} x_n m_{n-\ell\frac{L}{d}} \mathfrak{e}^{-\frac{2\pi \mathfrak{i} nkK}{d^2}}\right|^2 + (N_{K,L})_{k,\ell}\,,\ k \in \frac{d}{K}\left[K\right]_0, \ell \in \frac{d}{L}\left[L\right]_0.$$

(2) Compactly supported mask $\mathbf{m} \in \mathbb{C}^d$.

(3) Integers δ and γ, such that $\operatorname{supp}(\mathbf{m}) \subseteq [\delta]_0$, $\operatorname{supp}(\widehat{\mathbf{x}}) \subseteq [\gamma]_0$, and $\gamma \le 2\delta - 1 < d$.

Steps

(1) Ensure $L = 2\gamma - 1$ and $K = 2\delta - 1$.

(2) Estimate $\left(F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right)\right)_\beta$ for $|\sigma| \le \gamma - 1$ and $|\beta| \le \delta - 1$ by

$$\left(F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right)\right)_\omega \approx \begin{cases} \frac{d^3}{KL} \frac{\left(F_L Y_{K,L}^T F_K^T\right)_{-\alpha,\omega}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega} & \text{if } 1-\gamma \le \alpha \le 0 \text{ and } 1-\delta \le \omega \le -1 \\ \frac{d^3}{KL} \frac{\left(F_L Y_{K,L}^T F_K^T\right)_{-\alpha,\omega+K}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega} & \text{if } 1-\gamma \le \alpha \le 0 \text{ and } 0 \le \omega \le \delta - 1 \\ \frac{d^3}{KL} \frac{\left(F_L Y_{K,L}^T F_K^T\right)_{L-\alpha,\omega}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega} & \text{if } 1 \le \alpha \le \gamma - 1 \text{ and } 1-\delta \le \omega \le -1 \\ \frac{d^3}{KL} \frac{\left(F_L Y_{K,L}^T F_K^T\right)_{L-\alpha,\omega+K}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega} & \text{if } 1 \le \alpha \le \gamma - 1 \text{ and } 0 \le \omega \le \delta - 1 \end{cases}.$$

(3) Organize the $(2\delta-1)\cdot(2\gamma-1)$ values of $\left(F_d\left(\widehat{\mathbf{x}} \circ S_\sigma \overline{\widehat{\mathbf{x}}}\right)\right)_\beta$ for $|\sigma| \le \delta - 1$ and $|\beta| \le \gamma - 1$ in a matrix $V \in \mathbb{C}^{(2\delta-1)\times(2\gamma-1)}$ as specified in (4.15).

(4) Estimate $A \approx W^\dagger V \in \mathbb{C}^{\gamma\times(2\gamma-1)}$, where

$$W_{j,k} \coloneqq \mathfrak{e}^{-\frac{2\pi \mathfrak{i}(j-\delta+1)k}{d}},\ \text{for } j \in [2\delta-1]_0, k \in [\gamma]_0,$$
$$W^\dagger \coloneqq (W^*W)^{-1} W^* \in \mathbb{C}^{\gamma\times(2\delta-1)},$$
$$A \coloneqq \begin{bmatrix} 0 & \cdots & 0 & |\widehat{x}_0|^2 & \widehat{x}_0\overline{\widehat{x}_1} & \cdots & \widehat{x}_0\overline{\widehat{x}_{\gamma-1}} \\ 0 & \cdots & \widehat{x}_1\overline{\widehat{x}_0} & |\widehat{x}_1|^2 & \widehat{x}_1\overline{\widehat{x}_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \widehat{x}_{\gamma-2}\overline{\widehat{x}_{\gamma-3}} & |\widehat{x}_{\gamma-2}|^2 & \widehat{x}_{\gamma-2}\overline{\widehat{x}_{\gamma-1}} & \cdots & 0 \\ \widehat{x}_{\gamma-1}\overline{\widehat{x}_0} & \cdots & \widehat{x}_{\gamma-1}\overline{\widehat{x}_{\gamma-2}} & |\widehat{x}_{\gamma-1}|^2 & 0 & \cdots & 0 \end{bmatrix}.$$

(5) Reshape $W^\dagger V$, into an estimate $G \in \mathbb{C}^{\gamma\times\gamma}$ of the rank-one matrix

$$\widehat{\mathbf{x}}|_{[\gamma]_0}\, \widehat{\mathbf{x}}^*|_{[\gamma]_0} = \begin{bmatrix} |\widehat{x}_0|^2 & \widehat{x}_0\overline{\widehat{x}_1} & \widehat{x}_0\overline{\widehat{x}_2} & \cdots & \widehat{x}_0\overline{\widehat{x}_{\gamma-1}} \\ \widehat{x}_1\overline{\widehat{x}_0} & |\widehat{x}_1|^2 & \widehat{x}_1\overline{\widehat{x}_2} & \cdots & \widehat{x}_1\overline{\widehat{x}_{\gamma-1}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \widehat{x}_{\gamma-1}\overline{\widehat{x}_0} & \widehat{x}_{\gamma-1}\overline{\widehat{x}_1} & \widehat{x}_{\gamma-1}\overline{\widehat{x}_2} & \cdots & |\widehat{x}_{\gamma-1}|^2 \end{bmatrix}.$$

(6) Hermitianize the matrix G above: $G \leftarrow\!\shortmid \frac{1}{2}(G + G^*)$.

(7) Compute λ_1, the largest eigenvalue of G, and $\mathbf{v}_1$, its associated normalized eigenvector.

Output

$\mathbf{x}_e \coloneqq F_d^{-1}\widehat{\mathbf{x}}_e$, an estimate of $\mathbf{x}$, where $\widehat{\mathbf{x}}_e$ is given componentwise by

$$(\widehat{\mathbf{x}}_e)_j = \begin{cases} \sqrt{|\lambda_1|}\,(\mathbf{v}_1)_j, & j \in [\gamma]_0, \\ 0, & \text{otherwise.} \end{cases}$$

where $W \in \mathbb{C}^{2\delta-1\times\gamma}$ *is the partial Fourier matrix with entries* $W_{j,k} = \mathfrak{e}^{-\frac{2\pi\mathfrak{i}(j-\delta+1)k}{d}}$ *and* γ^{th} *singular value* $\sigma_\gamma(W)$*. Furthermore, if* $\|N_{K,L}\|_F$ *is sufficently small, then Algorithm 2 is always guaranteed to require at most* $\mathcal{O}\left(KL\log(KL) + \delta^3 + \log(\|\widehat{\mathbf{x}}\|_\infty)\gamma^2\right)$ *total flops to achieve* (1.14) *up to machine precision.*

The Proof of Theorem 2. Analogously to the proof of Theorem 3, we apply Lemma 9 with $\xi = \gamma$ and $\kappa = \delta$ to the cases where $\gamma \le \beta \le L-1$, $0 \le \beta \le \gamma - 1$, $\delta \le \nu \le K-1$, and $0 \le \nu \le \delta - 1$, and then substitute, $\alpha = L - \beta$, $\alpha = -\beta$, $\omega = \nu - K$, and $\omega = \nu$, to see that

(1) if $1-\gamma \le \alpha \le 0$ and $1-\delta \le \omega \le -1$, then

$$\left(F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right)\right)_\omega + \frac{d^3}{KL}\frac{\left(F_L N_{K,L}^T F_K^T\right)_{-\alpha,\omega+K}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega} = \frac{d^3}{KL}\frac{\left(F_L Y_{K,L}^T F_K^T\right)_{-\alpha,\omega+K}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega},$$

(2) if $1-\gamma \le \alpha \le 0$ and $0 \le \omega \le \delta - 1$, then

$$\left(F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right)\right)_\omega + \frac{d^3}{KL}\frac{\left(F_L N_{K,L}^T F_K^T\right)_{-\alpha,\omega}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega} = \frac{d^3}{KL}\frac{\left(F_L Y_{K,L}^T F_K^T\right)_{-\alpha,\omega}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega},$$

(3) if $1 \le \alpha \le \gamma - 1$ and $1-\delta \le \omega \le -1$, then

$$\left(F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right)\right)_\omega + \frac{d^3}{KL}\frac{\left(F_L N_{K,L}^T F_K^T\right)_{L-\alpha,\omega+K}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega} = \frac{d^3}{KL}\frac{\left(F_L Y_{K,L}^T F_K^T\right)_{L-\alpha,\omega+K}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega},$$

(4) if $1 \le \alpha \le \gamma - 1$ and $0 \le \omega \le \delta - 1$, then

$$\left(F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right)\right)_\omega + \frac{d^3}{KL}\frac{\left(F_L N_{K,L}^T F_K^T\right)_{L-\alpha,\omega}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{L-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega} = \frac{d^3}{KL}\frac{\left(F_L Y_{K,L}^T F_K^T\right)_{L-\alpha,\omega}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{L-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega}.$$

We will write the above equations in matrix form, with rows indexed from $1-\delta$ to $\delta-1$ and columns indexed from $1-\gamma$ to $\gamma-1$, as

$$T + U = V, \tag{4.14}$$

where T is $(2\delta-1)\times(2\gamma-1)$ matrix with entries defined by

$$T_{\alpha,\omega} = \left(F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right)\right)_\omega,$$

and the entries of U and V are given by

$$U_{\alpha,\omega} = \frac{d^3}{KL}\frac{\left(F_L N_{K,L}^T F_K^T\right)_{\beta(\alpha),\nu(\omega)}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega} \quad \text{and} \quad V_{\alpha,\omega} = \frac{d^3}{KL}\frac{\left(F_L N_{K,L}^T F_K^T\right)_{\beta(\alpha),\nu(\omega)}}{\left(F_d\left(\widehat{\mathbf{m}} \circ S_{-\alpha}\overline{\widehat{\mathbf{m}}}\right)\right)_\omega}, \tag{4.15}$$

with

$$\nu(\omega) = \begin{cases} \omega + K & \text{if } 1-\delta \le \omega \le -1 \\ \omega & \text{if } 0 \le \omega \le \delta - 1, \end{cases} \quad \text{and} \quad \beta(\alpha) = \begin{cases} -\alpha & \text{if } 1-\delta \le \omega \le -1 \\ L-\alpha & \text{if } 0 \le \omega \le \delta - 1, \end{cases}$$

for $1-\gamma \le \alpha \le \gamma - 1, 1-\delta \le \omega \le \delta - 1$. Let μ_2 be as in (2). Then, by the same reasoning as in (4.9), we see

$$\begin{aligned} \|U\|_F^2 &\le \frac{d^6}{K^2L^2}\frac{\left\|F_L N_{K,L} F_K^T\right\|_F^2}{\mu_2^2} \\ &= \frac{d^6}{K^2L^2\mu_2^2} LK \left\|N_{K,L}\right\|_F^2 \\ &= \frac{d^6}{KL\mu_2^2}\left\|N_{K,L}\right\|_F^2. \end{aligned} \tag{4.16}$$

Furthermore, for all α and ω such that $|\alpha| \le \gamma - 1$ and $|\omega| \le \delta - 1$,

$$T_{\omega,\alpha} = \left(F_d\left(\widehat{\mathbf{x}} \circ S_\alpha \overline{\widehat{\mathbf{x}}}\right)\right)_\omega = \sum_{n=0}^{\gamma-1} \mathfrak{e}^{-\frac{2\pi\mathfrak{i}\phi n}{d}} \widehat{x}_n \overline{\widehat{x}_{n+\alpha}}.$$

Therefore, we see $T = WA$, where $W \in \mathbb{C}^{(2\delta-1)\times\gamma}$ is the Vandermone matrix are given below by

$$W = \begin{bmatrix} 1 & \mathfrak{e}^{-\frac{2\pi\mathfrak{i}\cdot(-\delta+1)\cdot 1}{d}} & \cdots & \mathfrak{e}^{-\frac{2\pi\mathfrak{i}\cdot(-\delta+1)\cdot(\gamma-1)}{d}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathfrak{e}^{-\frac{2\pi\mathfrak{i}\cdot(-1)\cdot 1}{d}} & \cdots & \mathfrak{e}^{-\frac{2\pi\mathfrak{i}\cdot(-1)\cdot(\gamma-1)}{d}} \\ 1 & 1 & \cdots & 1 \\ 1 & \mathfrak{e}^{-\frac{2\pi\mathfrak{i}\cdot 1\cdot 1}{d}} & \cdots & \mathfrak{e}^{-\frac{2\pi\mathfrak{i}\cdot 1\cdot(\gamma-1)}{d}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathfrak{e}^{-\frac{2\pi\mathfrak{i}\cdot(\delta-1)\cdot 1}{d}} & \cdots & \mathfrak{e}^{-\frac{2\pi\mathfrak{i}\cdot(\delta-1)\cdot(\gamma-1)}{d}} \end{bmatrix}, \tag{4.17}$$

and $A \in \mathbb{C}^{\gamma\times(2\gamma-1)}$ is the partial autocorrelation type matrix

$$A = \begin{bmatrix} 0 & \cdots & 0 & |\widehat{x}_0|^2 & \widehat{x}_0\overline{\widehat{x}_1} & \cdots & \widehat{x}_0\overline{\widehat{x}_{\gamma-1}} \\ 0 & \cdots & \widehat{x}_1\overline{\widehat{x}_0} & |\widehat{x}_1|^2 & \widehat{x}_1\overline{\widehat{x}_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \widehat{x}_{\gamma-2}\overline{\widehat{x}_{\gamma-3}} & |\widehat{x}_{\gamma-2}|^2 & \widehat{x}_{\gamma-2}\overline{\widehat{x}_{\gamma-1}} & \cdots & 0 \\ \widehat{x}_{\gamma-1}\overline{\widehat{x}_0} & \cdots & \widehat{x}_{\gamma-1}\overline{\widehat{x}_{\gamma-2}} & |\widehat{x}_{\gamma-1}|^2 & 0 & \cdots & 0 \end{bmatrix}.$$

Therefore, by (4.14) we have

$$WA + U = V.$$

W has full rank since it is a Vandermonde matrix with distinct nodes, and therefore since $2\delta - 1 \geq \gamma$, it has a left inverse given by $W^\dagger = (W^*W)^{-1}W^*$. Thus,

$$W^\dagger V = A + W^\dagger U,$$

and so, by (4.16)

$$\begin{aligned} \left\|W^\dagger V - A\right\|_F &= \left\|W^\dagger U\right\|_F \\ &\leq \left\|W^\dagger\right\|_2 \|U\|_F \\ &\leq \frac{1}{\sigma_{min}(W)} \frac{d^3}{\sqrt{KL}\mu_2} \|N_{K,L}\|_F, \end{aligned} \tag{4.18}$$

where $\sigma_{min}(W)$ is the minimal singular value of W.

Let $P : \mathbb{C}^{\gamma\times(2\gamma-1)} \to \mathbb{C}^{\gamma\times\gamma}$ be a reshaping operator, such that if $M = (M_{1-\gamma}, \ldots, M_0, \ldots, M_{\gamma-1})$ is a $\gamma \times (2\gamma - 1)$ matrix,

$$(P(M))_{i,j} = M_{i,j-i} \tag{4.19}$$

for $0 \leq i, j \leq \gamma - 1$ so that

$$P(A) = \widehat{\mathbf{x}}|_{[\gamma]_0}\, \widehat{\mathbf{x}}^*|_{[\gamma]_0} = \begin{bmatrix} |\widehat{x}_0|^2 & \widehat{x}_0\overline{\widehat{x}_1} & \widehat{x}_0\overline{\widehat{x}_2} & \cdots & \widehat{x}_0\overline{\widehat{x}_{\gamma-1}} \\ \widehat{x}_1\overline{\widehat{x}_0} & |\widehat{x}_1|^2 & \widehat{x}_1\overline{\widehat{x}_2} & \cdots & \widehat{x}_1\overline{\widehat{x}_{\gamma-1}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \widehat{x}_{\gamma-1}\overline{\widehat{x}_0} & \widehat{x}_{\gamma-1}\overline{\widehat{x}_1} & \widehat{x}_{\gamma-1}\overline{\widehat{x}_2} & \cdots & |\widehat{x}_{\gamma-1}|^2 \end{bmatrix}.$$

and let $G = H(P(W^\dagger V))$, where H is the Hermitianizing operator defined in (4.10). (1.13) and the fact that $\|\widehat{\mathbf{x}}\|_2 = \sqrt{d}\|\mathbf{x}\|_2$ for all $\mathbf{x} \in \mathbb{C}^d$, imply that $\|N_{K,L}\|_F \leq \frac{\beta}{d}\|\widehat{\mathbf{x}}\|_2^2$, so the fact that $P(A) = \widehat{\mathbf{x}}|_{[\gamma]_0}\, \widehat{\mathbf{x}}^*|_{[\gamma]_0}$ is Hermitian, together with (4.18) implies

$$\begin{aligned} \left\|G - \widehat{\mathbf{x}}|_{[\gamma]_0}\, \widehat{\mathbf{x}}^*|_{[\gamma]_0}\right\|_F &= \left\|H(P(W^\dagger V)) - H(P(A))\right\|_F \\ &\leq \left\|W^\dagger V - A\right\|_F \\ &\leq \frac{1}{\sigma_{min}(W)} \frac{d^3}{\sqrt{KL}\mu_2} \|N_{K,L}\|_F \\ &\leq \frac{\beta}{\sigma_{min}(W)} \frac{d^2}{\sqrt{KL}\mu_2} \|\widehat{\mathbf{x}}\|_2^2. \end{aligned}$$

By Lemma 8 of [33], if λ_1 is the lead eigenvalue of G and $\mathbf{v}_1$ is an associated normalized eigenvector, then

$$\min_{\theta\in[0,2\pi]} \left\|\widehat{\mathbf{x}} - \mathbb{e}^{\mathbb{i}\theta}\widehat{\mathbf{x}}_e\right\|_2 = \min_{\theta\in[0,2\pi]} \left\|\mathbb{e}^{\mathbb{i}\theta}\ \widehat{\mathbf{x}}|_{[\gamma]_0} - \sqrt{|\lambda_1|}\mathbf{v}_1\right\|_2$$
$$\leq \frac{\left(1+2\sqrt{2}\right)\beta}{\sigma_{min}\left(W\right)} \frac{d^2}{\sqrt{KL}\mu_2} \left\|\widehat{\mathbf{x}}\right\|_2 .$$

(1.14) follows by taking the inverse Fourier transform of both sides. □

5. Numerical Experiments

We now present numerical experiments which demonstrate the robustness and efficiency of the proposed algorithms and provide comparisons to existing phase retrieval methods. These results were generated using the open source *BlockPR* MATLAB software package (freely available at [31]) on a desktop computer (iMac, 2017) with an Intel® Core™i7-7700 (7th generation, quad core) processor, 16GB RAM, and running macOS High Sierra and MATLAB R2018b. In all of our plots, each data point was obtained by averaging the results of 100 trials.

Unless otherwise stated, we used i.i.d. mean zero complex Gaussian random test signals with measurement errors modeled using a (real) i.i.d. Gaussian noise model. We will report both the signal to noise ratio (SNR) and reconstruction error in decibels (dB) with

$$\text{SNR (dB)} = 10\log_{10}\left(\frac{\sum_{k=1}^{K}\sum_{\ell=1}^{L}|\langle \mathbf{x}, S_\ell W_k \mathbf{m}\rangle|^4}{D\sigma^2}\right), \quad \text{Error (dB)} = 10\log_{10}\left(\frac{\min_\theta \|\mathbb{e}^{\mathbb{i}\theta}\mathbf{x}_e - \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}\right),$$

where $\mathbf{x}, \mathbf{x}_e, \sigma^2$ and $D \coloneqq KL$ denote the true signal, recovered signal, (Gaussian) noise variance, and number of measurements respectively.

We will present selected results comparing the proposed formulation against other popular phase retrieval algorithms such as *PhaseLift* [13] (implemented as a trace-regularized least-squares problem using the first order convex optimization package TFOCS [5, 6], *Hybrid Input-Output/Error Reduction (HIO+ER)* alternating projection algorithm [4, 20], and *Wirtinger Flow* [12]. We note that more accurate results using *PhaseLift* may be obtained using other solvers and software packages (such as CVX [22, 23]), albeit at a prohibitively expensive computational cost. For the HIO+ER algorithm, the following two projections were utilized: (i) projection onto the measured magnitudes, and (ii) projection onto the span of these measurement vectors. The initial guess was set to be the zero vector, although use of a random starting guess did not change the qualitative nature of the results. As is common practice, (see, for example, [20]) we implemented the HIO+ER algorithm in blocks of twenty-five HIO iterations followed by five ER iterations in order to accelerate the convergence of the algorithm. To minimize computational cost while ensuring convergence, the total number of HIO+ER iterations was limited to 600 (see Figure 5.1).

5.1. **Empirical Validation of Algorithm 1.** In Algorithm 1, whose convergence is guaranteed by Theorem 3, we assume that our measurements are obtained using a bandlimited mask with $\text{supp}\left(\widehat{\mathbf{m}}\right) \subseteq [\rho]_0$. To demonstrate the effictiveness of this algorithm, we performed numerical experiments on the following two types of masks:

$$\widehat{m}_k = \begin{cases} \left(1 + 0.5a_{\mathcal{U}}\right)\mathbb{e}^{2\pi\mathbb{i}a_{\mathcal{U}}} & \text{if } k \in [\rho]_0 \\ 0 & \text{otherwise} \end{cases} \quad a_{\mathcal{U}} \sim \mathcal{U}(0,1), \quad \text{(Random Mask)} \tag{5.1}$$

where $\mathcal{U}(0,1)$ denotes an i.i.d uniform random distribution on the interval $[0,1]$, and

$$\widehat{m}_k = \begin{cases} \frac{\mathbb{e}^{-k/a}}{\sqrt[4]{2\rho-1}} & \text{if } k \in [\rho]_0 \\ 0 & \text{otherwise} \end{cases} \quad a \coloneqq \max\left(4, (\rho-1)/2\right). \quad \text{(Exponential Mask)} \tag{5.2}$$

The exponential mask in (5.2) is closely related to the deterministic masks first introduced in [34]. The mask-dependent constant μ_1 (see (1.15) in Theorem 3) for the random mask, with $d = 60$ and $\rho = 8$ (and averaged over 50 trials), was 2.858×10^{-1}. The behavior for other choices of d and ρ was similar. For the exponential mask, this constant was 2.267×10^{-2}. The qualitative and quantitative performance of the algorithm was similar with both families of masks.

[1](HIO,ER) $= (x, y)$ indicates that every x iterations of the HIO algorithm was followed by y iterations of the ER algorithm.

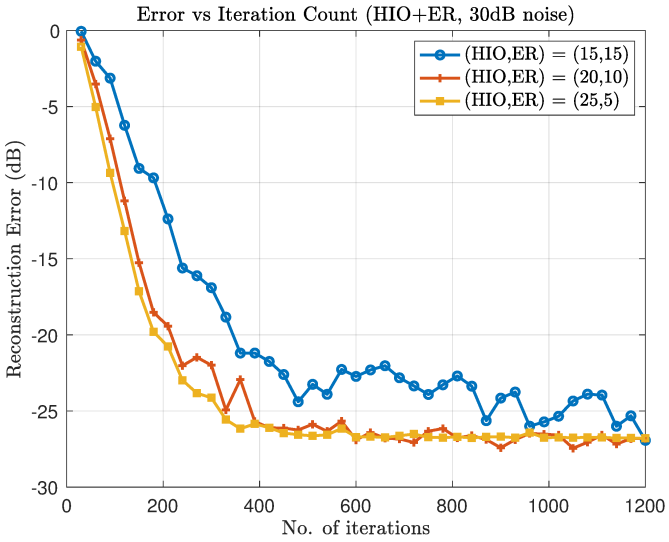
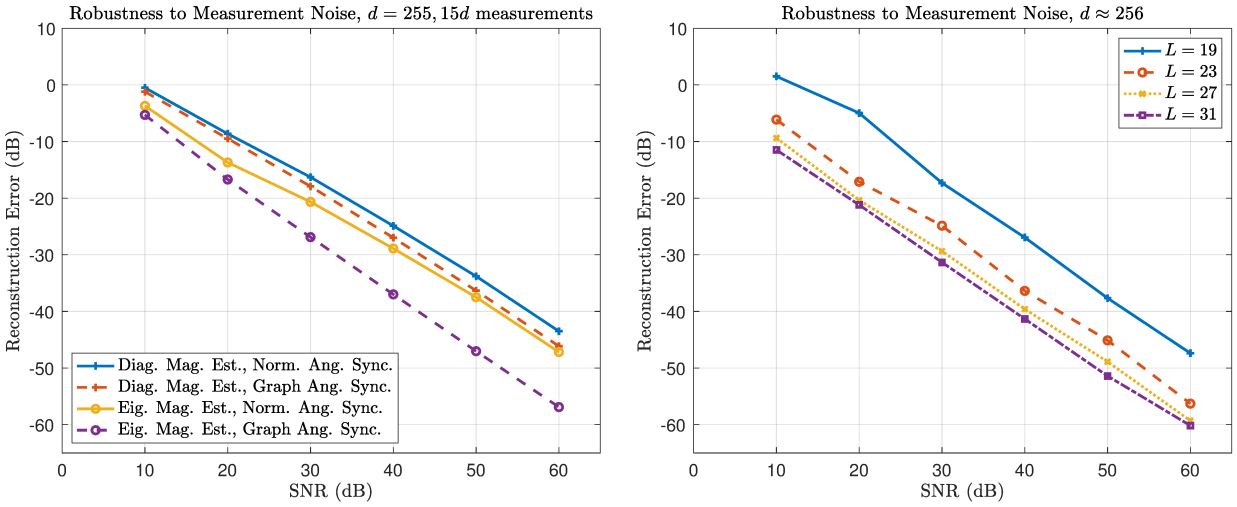


FIGURE 5.1. Selection of HIO+ER iteration parameters[1]

We performed experiments with both Algorithm 1, as presented in Section 4, and also with a modified version which uses a post-processing procedure to obtain improved accuracy. The modified algorithm replaces Steps (5) and (7) of Algorithm 1 (referred to as Diag. Mag. Est. and Norm. Ang. Sync. in Figure 5.2a) with the eigenvector based magnitude estimation procedure (Eig. Mag. Est.) in Section 6.1 of [33], and the graph Laplacian based angular synchronization method (Graph Ang. Sync.) described in Algorithm 3 of B. Preskitt's dissertation [39]. As seen in Figure 5.2a, which plots the reconstruction error at various noise levels with $d = 255$, $L = 15$, and a random mask constructed as in (5.1) with $\rho = 8$, these changes offered improved reconstruction accuracy.

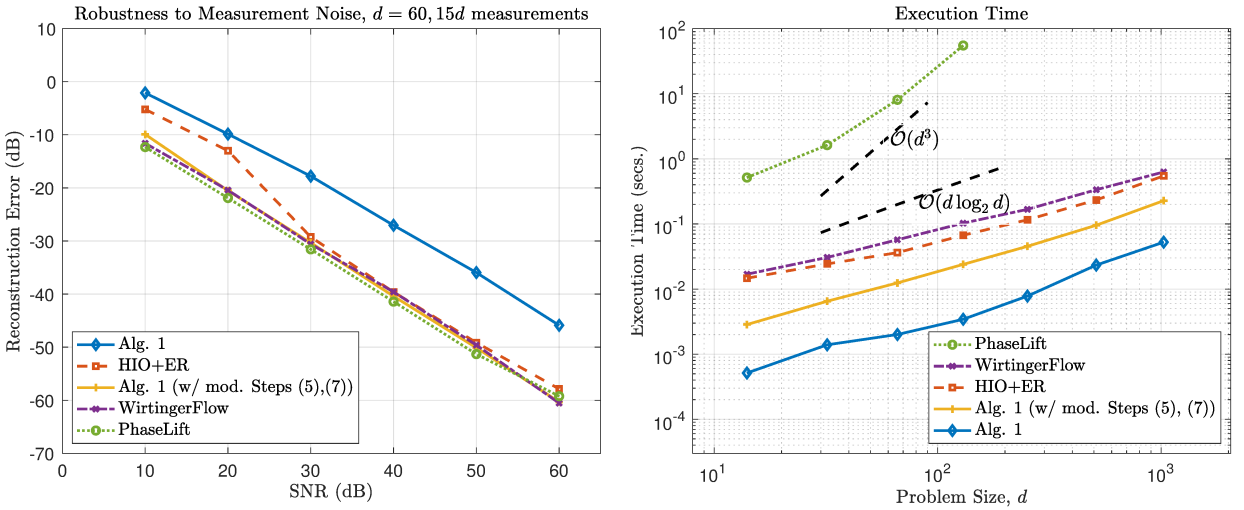


(A) Reconstruction accuracy for Algorithm 1 with and without modifications to Steps (5) and/or (7).

(B) Reconstruction accuracy vs. number of shifts L for Algorithm 1 (w/ mod. Steps (5),(7)).

FIGURE 5.2. Evaluating the performance of Algorithm 1 for various parameter choices

Figure 5.2b demonstrates the importance of the number of shifts L. As expected, the reconstructions using larger L (which entails using more measurements, each corresponding to greater overlap between successive masked regions of the specimen) offered improved accuracy. In order to ensure that L divides d, we varied the value of $d \approx 256$ slightly for different values of L. As in Figure 5.2a, we used random masks, constructed



(A) Reconstruction accuracy vs. added noise

(B) Execution time versus signal size d

FIGURE 5.3. Evaluating the robustness and efficiency of Algorithm 1 (and Theorem 3)

as in (5.1), with $\rho = 8$. We observe that for larger values of L, performance improved by about 10dB. We also note that, in practice, a suitable value of L can be chosen depending on whether the proposed method is used as a reconstruction procedure or as an initializer for another algorithm.

In Figure 5.3a, we compare the performance of the proposed method to other popular phase retrieval methods. Reconstruction errors for recovering a signal of length $d = 60$ using $L = 15$ shifts and a random mask with $\rho = 8$ are plotted for different levels of noise. We see that the proposed method performs well in comparison to the other algorithms, and even nearly matches the significantly more expensive algorithms such as *PhaseLift* which are based on semidefinite programming (SDP). We note that the Wirtinger Flow method is sensitive to the choice of parameters and iteration counts. We used fewer total iterations (150 at 10dB SNR) at higher noise levels and more iterations (4500 at 60dB) at lower levels in order to ensure that the algorithm converged to the level of noise. We are not aware of any methodical procedure for setting the various algorithmic parameters when utilizing the (local) measurement constructions considered in this paper. We also note that, of the algorithms considered, Algorithm 1 is the only one that has a theoretical convergence guarantee which applies to this class of spectrogram-type measurements.

Figure 5.3b plots the corresponding execution time for the various algorithms as a function of the problem size d. In this case, random masks were chosen with $\rho = \lceil 1.25 \log_2 d \rceil$ along with $L = \rho + \lceil \rho/2 \rceil - 1$ shifts. The figure confirms the essentially FFT–time computational cost of Algorithm 1. Furthermore, it also shows that while the post-processing procedure of modifying steps (5) and (7) does increase the computational cost of the algorithm, it does not increase it drastically.[2] In particular, even with these modifications, the proposed method provides best–in–class computational efficiency, and is significantly faster than the *HIO+ER*, *Wirtinger Flow*, and *PhaseLift* algorithms.

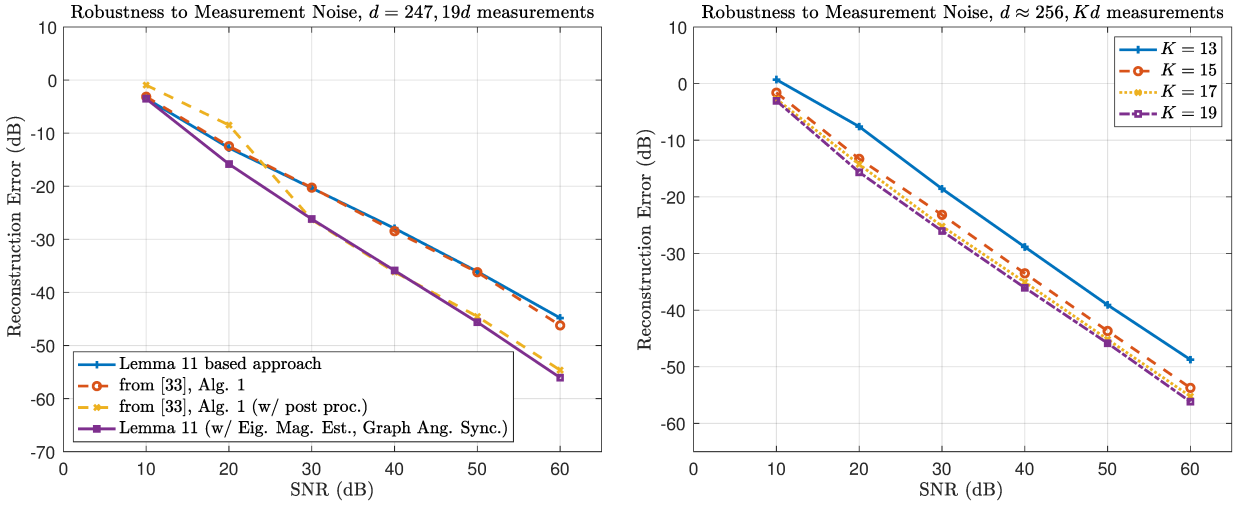
5.2. **Empirical Validation of a Lemma 11 Based Approach.** We next provide numerical results validating an approach based on Lemma 11 that applies the Wigner deconvolution method to the setting considered in [33]. As in Theorem 2, we assume that $\mathrm{supp}(\mathbf{m}) \subseteq [\delta]_0$, and we also add the assumption that $L = d$. In this setting, we may apply Lemma 11 and then solve for diagonal bands of $\mathbf{x}\mathbf{x}^*$ in a manner analogous to Algorithms 1 and 2. We then can recover $\mathbf{x}$ by applying the same angular synchronization procedure as in Algorithm 1. We note that, because we are using Lemma 11 rather than Lemma 9, we do not need to assume that $\mathbf{x}$ is bandlimited as we do in Theorem 2. As in Section 5.1, we conducted experiments with both deterministicly constructed and randomly constructed masks, and found that we obtained similar results for

[2]The modified Step (7) uses MATLAB's `eigs` command which can be computationally inefficient for this problem for large d; we defer a more detailed analysis and more efficient implementations to future work

both families of masks. The figures below use the exponential mask construction first introduced in [33],

$$m_k = \begin{cases} \frac{e^{-k/a}}{\sqrt[4]{2\delta-1}}, & \text{if } k \in [\delta]_0, \\ 0, & \text{otherwise,} \end{cases} \qquad a := \max(4, (\delta-1)/2), \tag{5.3}$$

and therefore allow us to directly compare the performance of the proposed method with the algorithm introduced in [33]. The mask-dependent constant μ_2 (see (1.12 in Theorem 2) for this mask, with $d = 247$ and $\delta = 10$, was 1.392×10^{-2}, with similar behavior for different choices of d and δ. Figure 5.4a plots the reconstruction error with $d = 247$, $K = 19$, $\delta = 10$, and $\mathbf{m}$ as in (5.3). Results with and without the post-processing modifications described in Section 5.1 are provided, along with results from [33] with and without the modified (see §6.1 in [33]) magnitude estimation and *HIO+ER* post-processing (60 iterations).



(A) Reconstruction accuracy with and without improved magnitude estimation/angular synchronization, and comparison with results from [33].

(B) Reconstruction accuracy vs. K, the number of Fourier modes.

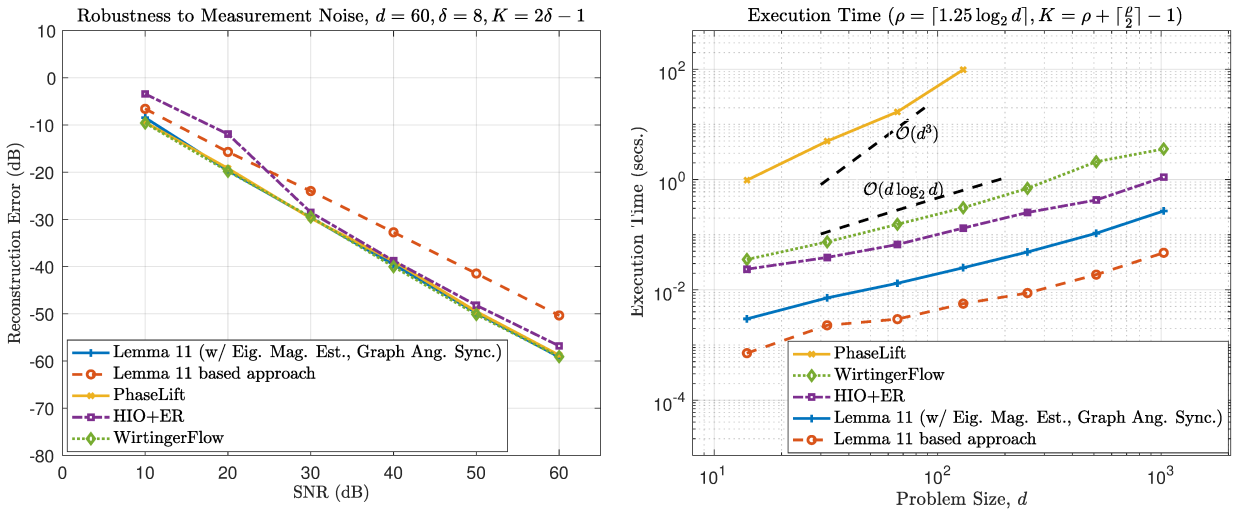
FIGURE 5.4. Evaluating the performance of the Lemma 11 based approach

As can be seen in Figure 5.4, the post-processing procedure yields a small improvement of about 5-10dB in the reconstruction error, especially at low noise levels. We observe that the Wigner deconvolution based approach yields numerical performance which is comparable to [33] in the settings where the theoretical guarantees of [33] are applicable, while also adding the additional flexibility of allowing shifts of length $a > 1$ under certain assumptions on either $\mathbf{m}$ or $\mathbf{x}$ as discussed in Theorems 2 and 3.

Next, we investigate the reconstruction accuracy as a function of K, the number of Fourier modes. Figure 5.4b plots reconstruction error in recovering a test signal for $K = 13, 15, 17$, and 19 respectively, with the exponential masks defined as in (5.3) with $\delta = 10$. As in Figure 5.2b, we vary the signal length d slighlty, in order to ensure that K divides d. As expected, the plot shows that reconstruction accuracy improves when K increases, i.e., when more measurements are acquired.

For completeness, we include noise robustness and execution time plots comparing the performance of the proposed method to the *HIO+ER*, *PhaseLift*, and *Wirtinger Flow* algorithms in Figures 5.5a and 5.5b respectively. From Figure 5.5a, we see that the proposed method (both with and without the modified magnitude estimation/angular synchronization procedures) performs well in comparison to *HIO+ER* and the other algorithms across a wide range of SNRs. Furthermore, Figure 5.5b demonstrates the essentially FFT–time computational cost of the method as well as the best-in-class computational efficiency when compared to other competing algorithms.

5.3. **Empirical Validation of Algorithm 2.** We now provide numerical results validating Algorithm 2, whose convergence is guaranteed by Theorem 2. We begin by noting that the Vandermonde matrix W defined in (4.17) often has a large condition number which poses a challenge in the accurate evaluation of Step (4) in



(A) Reconstruction accuracy vs. added noise

(B) Execution time versus signal size d

FIGURE 5.5. Evaluating the robustness and efficiency of the Lemma 11 based approach

Algorithm 2. One possible solution is to utilize the Tikhonov regularized solution $A = (W^*W + \sigma^2 I)^{-1} W^* V$ (see [27] for example), where the regularization parameter σ^2 is chosen using a procedure such as the L-curve method [26]. However, empirical simulations suggest that this procedure is not sufficiently robust to achieve reconstruction accuracy up to the level of added noise. Therefore, we replace Steps (4)–(6) in Algorithm 2 by a modified non-stationary iterated Tikhonov method inspired by the work of Buccini et al. in [9], which we detail in Algorithm 3. This procedure works by iteratively computing a Tikhonov regularized solution to the equation in Step (4) of Algorithm 2; however, at each step, the solution is applied to the residual of $WA = V$, with a geometrically decreasing regularization parameter. Buccini et al. showed that a similar iterative procedure has benefits over traditional Tikhonov regularization for more standard linear systems. While our problem setting is different, our empirical results suggest a similar benefit. We defer a more detailed theoretical analysis to future work.

Figure 5.6 presents empirical evaluation of the noise robustness and computational efficiency of Algorithm 2 with the Modified Iterated Tikhonov Method of Algorithm 3. Figure 5.6a plots the reconstruction error with signals of length $d = 190$, with frequency support of length $\gamma = 10$, using (complex random) masks with spatial support of length $\delta = 48$. We used $K = 2\delta - 1$ Fourier modes and $L = 2\gamma - 1$ shifts, and utilized the following iterated Tikhonov parameters: $q = 0.8$, $N = 20$, and α_0 chosen using the L-curve method. We note that using standard Tikhonov regularization (Alg. 2 in Figure 5.6a) yields rather poor results. An aggressive regularization parameter has to be chosen to surmount the ill-conditioning effects in Step (4) of Algorithm 2. Consequently, even a few iterations of the *HIO+ER* algorithm performs better than Algorithm 2. However, using the modified iterated Tikhonov procedure (Alg. 2 (w/ Alg. 3) in Figure 5.6a) yields significantly improved results, with a clear improvement in noise robustness over even the *HIO+ER* algorithm. Furthermore, Figure 5.6b plots the execution time as a function of the problem size for both Algorithms 2 and 3. The plot confirms that the modified iterative Tikhonov procedure of Algorithm 3 does not impose a significant computational burden.[3] Indeed, both Algorithms 2 and Algorithm 2 with the Modified Iterated Tikhonov Method of Algorithm 3 are faster than the *HIO+ER* algorithm. We note that more efficient implementations (involving fast computations of Vandermonde systems) of all the algorithms in Figure 5.6b may be possible; we defer this to future research.

6. Future Work

In future work, one might develop variants of the algorithms presented here for two-dimensional problems along the lines of [29]. Additionally, one might also develop variations of these algorithms for recovering

[3]We note that Step (1) of Algorithm 3 is computationally tractable since γ is typically small, and that the matrix $(W^*W + \alpha_0 q^k I)^{-1} W^*$ in Step 2(c) can be pre-computed.

Algorithm 3 Modified Non-Stationary Iterated Tikhonov Method with Geometrically Decaying Regularization Parameters

Inputs

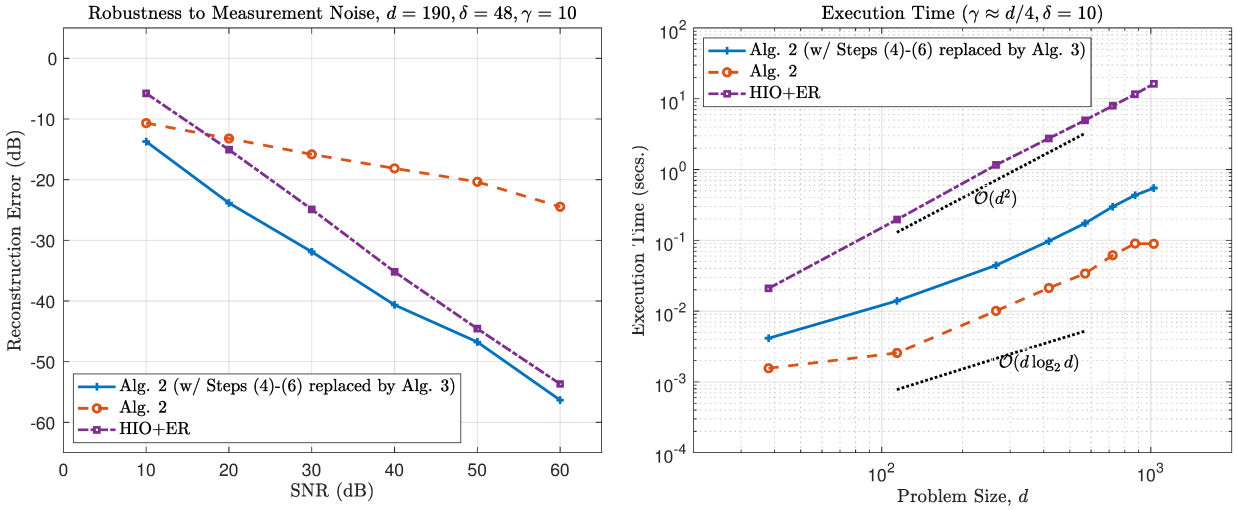
(1) Integers δ and γ, such that $\operatorname{supp}(\mathbf{m}) \subseteq [\delta]_0$ and $\operatorname{supp}(\widehat{\mathbf{x}}) \subseteq [\gamma]_0$.
(2) Vandermonde matrix $W \in \mathbb{C}^{(2\delta-1)\times\gamma}$ where $W_{j,k} = \mathrm{e}^{-\frac{2\pi \mathrm{i}(j-\delta+1)k}{d}}$ for $j \in [2\delta-1]_0, k \in [\gamma]_0$.
(3) Matrix $V \in \mathbb{C}^{(2\delta-1)\times(2\gamma-1)}$ from Step (3) of Algorithm 2 and as specified in (4.15).
(4) Non-stationary iterated Tikhonov parameters α_0 and q satisfying $\alpha_0 > 0$ and $0 < q < 1$.
(5) Iteration count N.

Steps

(1) Initialize $G \in \mathbb{C}^{\gamma\times\gamma}$ and $A \in \mathbb{C}^{\gamma\times(2\gamma-1)}$ to zero.
(2) For $k \leftarrow 1$ to N do
  (a) Compute a rank-one approximation $G_1 \in \mathbb{C}^{\gamma\times\gamma}$ of G:
$$G_1 = \tau_1 \mathbf{u}_1 \mathbf{v}_1^*,$$
  where τ_1 is the largest singular value of G and $\mathbf{u}_1$ and $\mathbf{v}_1$ are the corresponding left and right singular vectors respectively.
  (b) Let $A \in \mathbb{C}^{\gamma\times(2\gamma-1)}$ be the matrix such that $P(A) = G_1$ and $A_{i,j} = 0$ unless $j-i \le j \le j-i+\gamma-1$, where P is the reshaping operator defined in (4.19).
  (c) Apply Tikhonov regularization with decaying regularization parameter to the residual:
$$A \leftarrow\!\shortmid A + (W^*W + \alpha_0 q^k I)^{-1} W^* \underbrace{(V - WA)}_{\text{current residual}}$$
  (d) Obtain an updated estimate of G:
$$G = P(A),$$
  (e) Hermitianize the matrix G: $G \leftarrow\!\shortmid \frac{1}{2}(G + G^*)$.

Output

An estimate of the matrix $G \in \mathbb{C}^{\gamma\times\gamma}$ to be utilized in Step (5) of Algorithm 2.



(A) Reconstruction accuracy vs. added noise

(B) Execution time vs. signal size d

FIGURE 5.6. Empirical validation of Theorem 2 (Algorithm 2) and the Modified Iterated Tikhonov Method of Algorithm 3

compactly supported functions from sampled spectrogram measurements (see [36]) in the continuous setting. Furthermore, another, perhaps less direct, extension of these works would be to attempt to apply the Wigner distribution methods used here to the *sparse phase retrieval* problem. In, e.g., [30] it was shown that sparse vectors $\mathbf{x} \in \mathbb{C}^d$ with $\|\mathbf{x}\|_0 \leq s$ can be recovered up to a global phase from only $m = \mathcal{O}(s \log(d/s))$ magnitude measurements of the form $\left\{|\langle \mathbf{x}, \mathbf{a}_j \rangle|^2\right\}_{j=1}^m$. Thus, somewhat surprisingly, sparse phase retrieval problems generally do not require significantly more measurements to solve than compressive sensing problems. One may be able to generate new sparse phase retrieval methods for STFT magnitude measurements of the type considered here by replacing the standard Fourier techniques used in the methods above with sparse Fourier transform methods [8, 37, 43]. It has been shown that sparse phase retrieval problems can be solved in sublinear-time [47]. The further development of sublinear-time methods for solving sparse phase retrieval problems involving STFT magnitude measurements could prove valuable in the future for use in extremely large imaging scenarios.

Acknowledgements

Mark Iwen was supported in part by NSF DMS-1912706 and NSF CCF-1615489. Sami Merhi was supported in part by NSF CCF-1615489.

Appendix

In this section, we will prove the lemmas from Section 2 as well as Propositions 1 and 2.

The Proof of Lemma 1. Let $\mathbf{x} \in \mathbb{C}^d$, and let $\ell, \omega \in [d]_0$.

Part 1:

$$\begin{aligned}(F_d \widehat{\mathbf{x}})_\omega &= \sum_{k=0}^{d-1} \widehat{x}_k \mathrm{e}^{-\frac{2\pi \mathrm{i} k\omega}{d}} = \sum_{k=0}^{d-1}\sum_{n=0}^{d-1} x_n \mathrm{e}^{-\frac{2\pi \mathrm{i} nk}{d}} \mathrm{e}^{-\frac{2\pi \mathrm{i} k\omega}{d}} \\ &= \sum_{k=0}^{d-1}\sum_{n=0}^{d-1} x_{-n} \mathrm{e}^{\frac{2\pi \mathrm{i} k(n-\omega)}{d}} = d x_{-\omega} = d\widetilde{x}_\omega.\end{aligned}$$

Part 2:

$$\begin{aligned}\left(F_d \left(W_\ell \mathbf{x}\right)\right)_\omega &= \sum_{k=0}^{d-1} \left(x_k \mathrm{e}^{\frac{2\pi \mathrm{i} k\ell}{d}}\right) \mathrm{e}^{-\frac{2\pi \mathrm{i} k\omega}{d}} = \sum_{k=0}^{d-1} x_k \mathrm{e}^{-\frac{2\pi \mathrm{i} k(\omega-\ell)}{d}} \\ &= \widehat{x}_{\omega-\ell} = \left(S_{-\ell}\widehat{\mathbf{x}}\right)_\omega.\end{aligned}$$

Part 3:

$$\begin{aligned}\left(F_d \left(S_\ell \mathbf{x}\right)\right)_\omega &= \sum_{k=0}^{d-1} x_{k+\ell} \mathrm{e}^{-\frac{2\pi \mathrm{i} k\omega}{d}} = \sum_{k=0}^{d-1} x_{k+\ell} \mathrm{e}^{-\frac{2\pi \mathrm{i}(k+\ell)\omega}{d}} \mathrm{e}^{-\frac{2\pi \mathrm{i}(-\ell)\omega}{d}} \\ &= \mathrm{e}^{\frac{2\pi \mathrm{i} \ell\omega}{d}} \widehat{x}_\omega = \left(W_\ell \widehat{\mathbf{x}}\right)_\omega.\end{aligned}$$

Part 4:

$$\begin{aligned}\left(W_{-\ell} F_d \left(S_\ell \widetilde{\widehat{\mathbf{x}}}\right)\right)_\omega &= \mathrm{e}^{-\frac{2\pi \mathrm{i} \ell\omega}{d}} \left(W_\ell \widetilde{\widehat{\mathbf{x}}}\right)_\omega \qquad \text{(by part 3)} \\ &= \left(\widetilde{\widehat{\mathbf{x}}}\right)_\omega = \sum_{k=0}^{d-1} \widetilde{\widehat{x}}_k \mathrm{e}^{-\frac{2\pi \mathrm{i} k\omega}{d}} \\ &= \overline{\sum_{k=0}^{d-1} \widetilde{x}_k \mathrm{e}^{\frac{2\pi \mathrm{i} k\omega}{d}}} = \overline{\sum_{k=0}^{d-1} x_{-k} \mathrm{e}^{\frac{2\pi \mathrm{i} k\omega}{d}}} = \left(\widetilde{\widehat{\mathbf{x}}}\right)_\omega.\end{aligned}$$

Part 5:

$$\begin{aligned}\left(\widetilde{S_\ell \mathbf{x}}\right)_\omega &= \overline{\left(S_\ell \mathbf{x}\right)_\omega} \\ &= \overline{x_{-\omega+\ell}} = \widetilde{\widehat{x}}_{\ell-\omega} = \left(S_{-\ell}\widetilde{\widehat{\mathbf{x}}}\right)_\omega.\end{aligned}$$

Part 6:

$$\begin{aligned}(F_d\overline{\mathbf{x}})_\omega &= \sum_{k=0}^{d-1}\overline{x}_k \mathfrak{e}^{-\frac{2\pi \mathfrak{i} k\omega}{d}} = \overline{\sum_{k=0}^{d-1} x_k \mathfrak{e}^{\frac{2\pi \mathfrak{i} k\omega}{d}}} \\ &= \overline{\sum_{k=0}^{d-1} x_{-k} \mathfrak{e}^{-\frac{2\pi \mathfrak{i} k\omega}{d}}} = \overline{(F_d\widetilde{\mathbf{x}})_\omega}.\end{aligned}$$

Part 7:

$$\begin{aligned}\left(\widehat{\widetilde{\mathbf{x}}}\right)_\omega = \widehat{\mathbf{x}}_{-\omega} &= \sum_{k=0}^{d-1} x_k \mathfrak{e}^{\frac{2\pi \mathfrak{i} k\omega}{d}} \\ &= \sum_{k=0}^{d-1} x_{-k} \mathfrak{e}^{-\frac{2\pi \mathfrak{i} k\omega}{d}} = (F_d\widetilde{\mathbf{x}})_\omega .\end{aligned}$$

Part 8: For all $\mathbf{x} \in \mathbb{C}^d$,

$$\begin{aligned}|F_d\mathbf{x}|^2 &= (F_d\mathbf{x}) \circ \overline{(F_d\mathbf{x})} \\ &= (F_d\mathbf{x}) \circ \left(F_d\widetilde{\widehat{\mathbf{x}}}\right) && \text{(by Lemma 1, part 4, with } \ell = 0) \\ &= F_d\left(\mathbf{x} *_d \widetilde{\widehat{\mathbf{x}}}\right). && \text{(by Lemma 2)}\end{aligned}$$

□

The Proof of Lemma 2. For $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$, $k \in [d]_0$,

$$\begin{aligned}(F_d(\mathbf{x} *_d \mathbf{y}))_k &= \sum_{n=0}^{d-1}\sum_{\ell=0}^{d-1} x_\ell y_{n-\ell} \mathfrak{e}^{-\frac{2\pi \mathfrak{i} nk}{d}} \\ &= \sum_{\ell=0}^{d-1} x_\ell \mathfrak{e}^{-\frac{2\pi \mathfrak{i} \ell k}{d}} \sum_{n=0}^{d-1} y_{n-\ell} \mathfrak{e}^{-\frac{2\pi \mathfrak{i}(n-\ell)k}{d}} \\ &= \sum_{\ell=0}^{d-1} x_\ell \mathfrak{e}^{-\frac{2\pi \mathfrak{i} \ell k}{d}} \sum_{m=0}^{d-1} y_m \mathfrak{e}^{-\frac{2\pi \mathfrak{i} mk}{d}} \\ &= \widehat{x}_k \widehat{y}_k.\end{aligned}$$

Therefore, $F_d(\mathbf{x} *_d \mathbf{y}) = \widehat{\mathbf{x}} \circ \widehat{\mathbf{y}}$, so multiplying by F_d^{-1} proves the first claim. To verify the second claim, note that by Lemma 1 part 1,

$$\begin{aligned}F_d(F_d\mathbf{x} *_d F_d\mathbf{y}) &= F_dF_d\mathbf{x} \circ F_dF_d\mathbf{y} \\ &= d^2\widetilde{\mathbf{x}} \circ \widetilde{\mathbf{y}} \\ &= d^2\widetilde{\mathbf{x} \circ \mathbf{y}} \\ &= dF_d\left(dF_d(\mathbf{x} \circ \mathbf{y})\right).\end{aligned}$$

□

The Proof of Lemma 3. Let $\mathbf{x} \in \mathbb{C}^d$, and let $\alpha, \omega \in [d]_0$. Observe that

$$\begin{aligned}\left(F_d\left(\mathbf{x} \circ S_\omega\overline{\mathbf{x}}\right)\right)_\alpha &= \frac{1}{d}\left(\widehat{\mathbf{x}} *_d F_d\left(S_\omega\overline{\mathbf{x}}\right)\right)_\alpha && \text{(by Lemma 2)} \\ &= \frac{1}{d}\left(\widehat{\mathbf{x}} *_d \left(W_\omega\widehat{\overline{\mathbf{x}}}\right)\right)_\alpha && \text{(by Lemma 1, part 3)} \\ &= \frac{1}{d}\sum_{n=0}^{d-1}\widehat{x}_n\left(W_\omega\widehat{\overline{\mathbf{x}}}\right)_{\alpha-n} && \text{(by definition of } *_d) \\ &= \frac{1}{d}\sum_{n=0}^{d-1}\widehat{x}_n\widehat{\overline{x}}_{\alpha-n}\mathfrak{e}^{\frac{2\pi \mathfrak{i}\omega(\alpha-n)}{d}} && \text{(by definition of } W_\omega)\end{aligned}$$

$$= \frac{1}{d}\mathfrak{e}^{\frac{2\pi \mathfrak{i}\omega\alpha}{d}} \sum_{n=0}^{d-1} \widehat{x}_n \widetilde{\overline{x}}_{n-\alpha} \mathfrak{e}^{\frac{-2\pi \mathfrak{i}\omega n}{d}} \qquad \text{(by definition of } \widetilde{\cdot}\text{)}$$

$$= \frac{1}{d}\mathfrak{e}^{\frac{2\pi \mathfrak{i}\omega\alpha}{d}} \sum_{n=0}^{d-1} \widehat{x}_n \overline{\widetilde{x}}_{n-\alpha} \mathfrak{e}^{\frac{-2\pi \mathfrak{i}\omega n}{d}} \qquad \text{(by Lemma 1, parts 6 and 7)}$$

$$= \frac{1}{d}\mathfrak{e}^{\frac{2\pi \mathfrak{i}\omega\alpha}{d}} \left(F_d\left(\widehat{\mathbf{x}} \circ S_{-\alpha}\overline{\widetilde{\mathbf{x}}}\right)\right)_\omega .$$

□

The Proof of Lemma 4. For any $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$ and any $\alpha \in \mathbb{Z}$, it is straightforward to check that

$$R\overline{\mathbf{x}} = \overline{R\mathbf{x}}, \quad S_\alpha \overline{\mathbf{x}} = \overline{S_\alpha \mathbf{x}}, \quad \text{and} \quad R(\mathbf{x} \circ \mathbf{y}) = (R\mathbf{x}) \circ (R\mathbf{y}). \tag{6.1}$$

Therefore,

$$F_d\left(\widetilde{\mathbf{x}} \circ S_{-\alpha}\overline{\widetilde{\mathbf{x}}}\right) = F_d\left(R\mathbf{x} \circ S_{-\alpha}\overline{R\mathbf{x}}\right) \qquad \text{(by definition of R)}$$

$$= F_d\left(R\mathbf{x} \circ \overline{RS_\alpha \mathbf{x}}\right) \qquad \text{(by Lemma 1, part 5)}$$

$$= F_d\left(R\left(\mathbf{x} \circ S_\alpha \overline{\mathbf{x}}\right)\right) \qquad \text{(by (6.1))}$$

$$= R\left(F_d\left(\mathbf{x} \circ S_\alpha \overline{\mathbf{x}}\right)\right). \qquad \text{(by Lemma 1, part 7)}$$

□

The Proof of Lemma 5. Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$, and let $\ell, k \in [d]_0$. Then,

$$\left((\mathbf{x} \circ S_{-\ell}\mathbf{y}) *_d \left(\overline{\widetilde{\mathbf{x}}} \circ S_\ell \overline{\widetilde{\mathbf{y}}}\right)\right)_k = \sum_{n=0}^{d-1} (\mathbf{x} \circ S_{-\ell}\mathbf{y})_n \left(\overline{\widetilde{\mathbf{x}}} \circ S_\ell \overline{\widetilde{\mathbf{y}}}\right)_{k-n} \qquad \text{(by definition of } *_d\text{)}$$

$$= \sum_{n=0}^{d-1} x_n y_{n-\ell} \overline{\widetilde{x}}_{k-n} \overline{\widetilde{y}}_{\ell+k-n} \qquad \text{(by definition of } \circ\text{)}$$

$$= \sum_{n=0}^{d-1} x_n \overline{x}_{n-k} \widetilde{y}_{\ell-n} \overline{\widetilde{y}}_{\ell-n+k} \qquad \text{(by definition of } \widetilde{\cdot}\text{)}$$

$$= \left((\mathbf{x} \circ S_{-k}\overline{\mathbf{x}}) *_d \left(\widetilde{\mathbf{y}} \circ S_k \overline{\widetilde{\mathbf{y}}}\right)\right)_\ell . \qquad \text{(by definition of } *_d\text{)}$$

□

The Proof of Lemma 6. For $x \in \mathbb{C}^d$, the Fourier inversion formula states that

$$x_n = \left(F_d^{-1}\widehat{\mathbf{x}}\right)_n = \frac{1}{d}\sum_{k=0}^{d-1} \widehat{x}_k \mathfrak{e}^{\frac{2\pi \mathfrak{i}kn}{d}}.$$

Therefore, for all $\omega \in \left[\frac{d}{s}\right]_0$,

$$\left(F_{\frac{d}{s}}\left(Z_s\mathbf{x}\right)\right)_\omega = \sum_{n=0}^{\frac{d}{s}-1} \left(Z_s\mathbf{x}\right)_n \mathfrak{e}^{-\frac{2\pi \mathfrak{i}n\omega}{d/s}}$$

$$= \sum_{n=0}^{\frac{d}{s}-1} x_{ns} \mathfrak{e}^{-\frac{2\pi \mathfrak{i}n\omega}{d/s}}$$

$$= \frac{1}{d}\sum_{n=0}^{\frac{d}{s}-1} \left(\sum_{k=0}^{d-1} \widehat{x}_k \mathfrak{e}^{\frac{2\pi \mathfrak{i}kns}{d}}\right) \mathfrak{e}^{-\frac{2\pi \mathfrak{i}\omega ns}{d}}$$

$$= \frac{1}{d}\sum_{k=0}^{d-1} \widehat{x}_k \sum_{n=0}^{\frac{d}{s}-1} \mathfrak{e}^{\frac{2\pi \mathfrak{i}n(k-\omega)}{d/s}}$$

$$= \frac{1}{d}\frac{d}{s}\sum_{r=0}^{s-1}\widehat{x}_{\omega+r\frac{d}{s}} = \frac{1}{s}\sum_{r=0}^{s-1}\widehat{x}_{\omega-r\frac{d}{s}}.$$

□

The Proof of Proposition 1. Let $\mathbf{m}\in\mathbb{C}^d$ be a bandlimited mask, whose Fourier transform may be written as

$$\widehat{\mathbf{m}} = \left(a_0\mathbb{e}^{\mathbb{i}\theta_0},\dots,a_{\rho-1}\mathbb{e}^{\mathbb{i}\theta_{\rho-1}},0,\dots,0\right)^T$$

for some real numbers $a_0,\dots,a_{\rho-1}$, which satisfy (4.1) and (4.2). Let $2\le\kappa\le\rho$, and recall that μ_1 is defined by

$$\mu_1 = \min_{\substack{|p|\le\kappa-1\\ q\in[d]_0}}\left|F_d\left(\widehat{\mathbf{m}}\circ S_p\overline{\widehat{\mathbf{m}}}\right)_q\right|.$$

For $0\le p\le\kappa-1$, we have

$$\left(\widehat{\mathbf{m}}\circ S_p\overline{\widehat{\mathbf{m}}}\right)_n = \begin{cases} a_n a_{n+p}\mathbb{e}^{\mathbb{i}(\theta_n-\theta_{n+p})}, & \text{if } n\in[\rho-p]_0,\\ 0, & \text{otherwise,}\end{cases}$$

and for $-\kappa+1\le p<0$,

$$\left(\widehat{\mathbf{m}}\circ S_p\overline{\widehat{\mathbf{m}}}\right)_n = \begin{cases} a_{n-|p|}a_n\mathbb{e}^{\mathbb{i}\left(\theta_n-\theta_{n-|p|}\right)}, & \text{if } n\in\{|p|,|p|+1,\dots,\rho-1\},\\ 0, & \text{otherwise.}\end{cases}$$

Therefore, for any $q\in[d]_0$ and any $|p|\le\kappa-1$,

$$F_d\left(\widehat{\mathbf{m}}\circ S_p\overline{\widehat{\mathbf{m}}}\right)_q = \sum_{n=0}^{\rho-1-p} a_n a_{|p|+n}\mathbb{e}^{\mathbb{i}\phi_{n,p,q}},$$

where $\phi_{n,p,q}$ is some real number depending on n,p, and q. Using the assumptions (4.1) and (4.2) we see that

$$\left|\sum_{n=1}^{\rho-1-|p|} a_n a_{|p|+n}\mathbb{e}^{\mathbb{i}\phi_{n,p,q}}\right| \le (\rho-1)\left|a_1\right|\left|a_{1+|p|}\right| < \left|a_0\right|\left|a_{|p|}\right|. \tag{6.2}$$

With this,

$$\begin{aligned}\left|F_d\left(\widehat{\mathbf{m}}\circ S_p\overline{\widehat{\mathbf{m}}}\right)_q\right| &= \left|\sum_{n=0}^{\rho-1-|p|} a_n a_{|p|+n}\mathbb{e}^{\mathbb{i}\phi_{n,p,q}}\right|\\ &= \left|a_0 a_{|p|}\mathbb{e}^{\mathbb{i}\phi_{0,p,q}} + \sum_{n=1}^{\rho-1-|p|} a_n a_{|p|+n}\mathbb{e}^{\mathbb{i}\phi_{n,p,q}}\right|\\ &\ge \left|\left|a_0 a_{|p|}\mathbb{e}^{\mathbb{i}\phi_{0,p,q}}\right| - \left|\sum_{n=1}^{\rho-1-|p|} a_n a_{|p|+n}\mathbb{e}^{\mathbb{i}\phi_{n,p,q}}\right|\right|\\ &= \left|\left|a_0 a_{|p|}\right| - \left|\sum_{n=1}^{\rho-1-|p|} a_n a_{|p|+n}\mathbb{e}^{\mathbb{i}\phi_{n,p,q}}\right|\right|\\ &> 0,\end{aligned}$$

where the last inequality follows by 6.2. Therefore, $F_d\left(\widehat{\mathbf{m}}\circ S_p\overline{\widehat{\mathbf{m}}}\right)_q$ is nonzero for all p and q and so $\mu_1>0$.

□

The Proof of Proposition 2. Let

$$\mathbf{m} = \left(a_0\mathbb{e}^{\mathbb{i}\theta_0},\dots,a_{\delta-1}\mathbb{e}^{\mathbb{i}\theta_{\delta-1}},0,\dots,0\right)^T$$

be a compactly supported mask, where $a_0, \dots, a_{\delta-1}$, are real numbers which satisfy (4.3) and (4.4). Let $1 \leq \gamma \leq 2\delta - 1$, and recalll that μ_2 is defined by

$$\mu_2 = \min_{\substack{|p| \leq \gamma - 1 \\ |q| \leq \delta - 1}} \left| F_d \left(\widehat{\mathbf{m}} \circ S_p \overline{\widehat{\mathbf{m}}} \right)_q \right|.$$

By Lemma 3, it suffices to show that

$$F_d \left(\mathbf{m} \circ S_{-q} \overline{\mathbf{m}} \right)_p \neq 0$$

for all $|p| \leq \gamma - 1$ and all $|q| \leq \delta - 1$. If $-\delta + 1 \leq q < 0$, then

$$\left(\mathbf{m} \circ S_{-q} \overline{\mathbf{m}} \right)_n = \begin{cases} a_n a_{n+|q|} \mathbb{e}^{\mathbb{i}\left(\theta_n - \theta_{n+|q|}\right)}, & \text{if } 0 \leq n \leq \delta - |q| - 1, \\ 0, & \text{otherwise} \end{cases},$$

and if $0 \leq q \leq \delta - 1$, then

$$\left(\mathbf{m} \circ S_{-q} \overline{\mathbf{m}} \right)_n = \begin{cases} a_{n-q} a_n \mathbb{e}^{\mathbb{i}(\theta_n - \theta_{n-q})}, & \text{if } q \leq n \leq \delta - 1, \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, for all $|p| \leq \gamma - 1$ and all $|q| \leq \delta - 1$

$$F_d \left(\mathbf{m} \circ S_{-q} \overline{\mathbf{m}} \right)_p = \sum_{n=0}^{\delta - 1 - |q|} a_n a_{|q|+n} \mathbb{e}^{\mathbb{i}\phi_{n,p,q}},$$

where $\phi_{n,p,q}$ is some real number depending on n, p, and q. By the same reasoning as in the proof of Proposition 1, this combined with (4.3) and (4.4) implies that $F_d \left(\mathbf{m} \circ S_{-q} \overline{\mathbf{m}} \right)_p \neq 0$ for all $|p| \leq \gamma - 1$ and all $|q| \leq \delta - 1$.

□

References

[1] B. Alexeev, A. S. Bandeira, M. Fickus, and D. G. Mixon. Phase Retrieval with Polarization. *SIAM Journal on Imaging Sciences*, 7(1):35–66, 2014.

[2] R. Balan, P. Casazza, and D. Edidin. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345–356, 2006.

[3] A. S. Bandeira, Y. Chen, and D. G. Mixon. Phase retrieval from power spectra of masked signals. *Information and Inference: a Journal of the IMA*, 3(2):83–102, 2014.

[4] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization. *Journal of the Optical Society of America. A, Optics, Image science, and Vision*, 19(7):1334–1345, 2002.

[5] S. Becker, E. J. Candès, and M. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3(3):165–218, Aug. 2011.

[6] S. Becker, E. J. Candes, and M. Grant. TFOCS: Templates for first-order conic solvers, version 1.3.1. `http://cvxr.com/tfocs`, Sep. 2014.

[7] T. Bendory, Y. C. Eldar, and N. Boumal. Non-convex phase retrieval from stft measurements. *IEEE Transactions on Information Theory*, 64(1):467–484, 2017.

[8] S. Bittens, R. Zhang, and M. A. Iwen. A deterministic sparse fft for functions with structured fourier sparsity. *Advances in Computational Mathematics*, 45(2):519–561, Apr 2019.

[9] A. Buccini, M. Donatelli, and L. Reichel. Iterated tikhonov regularization with a general penalty term. *Numerical Linear Algebra with Applications*, 24(4):2089, 2017.

[10] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM review*, 57(2):225–251, 2015.

[11] E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, Sept. 2015.

[12] E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, April 2015.

[13] E. J. Candès, T. Strohmer, and V. Voroninski. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Commun. Pure Appl. Math.*, 66(8):1241–1274, 2013.

[14] H. N. Chapman. Phase-retrieval x-ray microscopy by Wigner-distribution deconvolution. *Ultramicroscopy*, 66(3):153 – 172, 1996.

[15] J. Clark, L. Beitra, G. Xiong, A. Higginbotham, D. Fritz, H. Lemke, D. Zhu, M. Chollet, G. Williams, and M. Messerschmidt. Ultrafast three-dimensional imaging of lattice dynamics in individual gold nanocrystals. *Science*, 341(6141):56–59, 2013.

[16] J. Corbett. The Pauli problem, state reconstruction and quantum-real numbers. *Reports on Mathematical Physics*, 57(1):53–68, 2006.
[17] J. C. da Silva and A. Menzel. Elementary signals in ptychography. *Opt. Express*, 23(26):33812–33821, Dec 2015.
[18] C. Fienup and J. Dainty. Phase retrieval and image reconstruction for astronomy. *Image Recovery: Theory and Application*, pages 231–275, 1987.
[19] J. R. Fienup. Reconstruction of an object from the modulus of its Fourier transform. *Opt. Lett.*, 3:27–29, 1978.
[20] J. R. Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.
[21] R. Gerchberg and W. Saxton. A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures. *Optik*, 35:237–246, 1972.
[22] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer–Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.
[23] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, Mar. 2014.
[24] D. Griffin and J. Lim. Signal estimation from modified short-time fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):236–243, 1984.
[25] D. Gross, F. Krahmer, and R. Kueng. Improved recovery guarantees for phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 42:37 – 64, 2017.
[26] P. C. Hansen. The L-Curve and its use in the numerical treatment of inverse problems. In *in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering*, pages 119–142. WIT Press, 2000.
[27] P. C. Hansen. *Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion*, volume 4. SIAM, 2005.
[28] R. W. Harrison. Phase problem in crystallography. *JOSA A*, 10(5):1046–1055, 1993.
[29] M. Iwen, B. Preskitt, R. Saab, and A. Viswanathan. Phase retrieval from local measurements in two dimensions. In *Wavelets and Sparsity XVII*, volume 10394, page 103940X. International Society for Optics and Photonics, 2017.
[30] M. Iwen, A. Viswanathan, and Y. Wang. Robust sparse phase retrieval made easy. *Applied and Computational Harmonic Analysis*, 42(1):135–142, 2017.
[31] M. Iwen, Y. Wang, and A. Viswanathan. BlockPR: Matlab software for phase retrieval using block circulant measurement constructions and angular synchronization, version 2.0. `https://bitbucket.org/charms/blockpr`, Apr. 2016.
[32] M. A. Iwen, S. Merhi, and M. Perlmutter. Lower Lipschitz bounds for phase retrieval from locally supported measurements. *Applied and Computational Harmonic Analysis*, 2019.
[33] M. A. Iwen, B. Preskitt, R. Saab, and A. Viswanathan. Phase retrieval from local measurements: Improved robustness via eigenvector-based angular synchronization. *Applied and Computational Harmonic Analysis*, 2018.
[34] M. A. Iwen, A. Viswanathan, and Y. Wang. Fast phase retrieval from local correlation measurements. *SIAM J. Imaging Sci.*, 9(4):1655–1688, 2016.
[35] O. Melnyk, F. Filbir, and F. Krahmer. Phase retrieval from local correlation measurements with fixed shift length. In *Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP)*, page MTu4D.3. Optical Society of America, 2019.
[36] S. Merhi, A. Viswanathan, and M. Iwen. Recovery of compactly supported functions from spectrogram measurements via lifting. In *Sampling Theory and Applications (SampTA), 2017 International Conference on*, pages 538–542. IEEE, 2017.
[37] S. Merhi, R. Zhang, M. A. Iwen, and A. Christlieb. A new class of fully discrete sparse Fourier transforms: Faster stable implementations with guarantees. *Journal of Fourier Analysis and Applications*, 25(3):751–784, Jun 2019.
[38] G. E. Pfander and P. Salanevich. Robust phase retrieval algorithm for time-frequency structured measurements. *SIAM Journal on Imaging Sciences*, 12(2):736–761, 2019.
[39] B. P. Preskitt. *Phase Retrieval from Locally Supported Measurements*. PhD thesis, University of California, San Diego, 2018.
[40] J. Rodenburg. Ptychography and related diffractive imaging methods. *Advances in Imaging and Electron Physics*, 150:87–184, 2008.
[41] J. M. Rodenburg and R. H. T. Bates. The theory of super-resolution electron microscopy via wigner-distribution deconvolution. *Philosophical Transactions: Physical Sciences and Engineering*, 339(1655):521–553, 1992.
[42] P. Salanevich and G. E. Pfander. Polarization based phase retrieval for time-frequency structured measurements. In *Proc. 2015 Int. Conf. Sampling Theory and Applications (SampTA)*, pages 187–191, 2015.
[43] B. Segal and M. A. Iwen. Improved sparse fourier approximation results: faster implementations and stronger guarantees. *Numerical Algorithms*, 63(2):239–263, Jun 2013.
[44] M. M. Seibert, T. Ekeberg, F. R. Maia, M. Svenda, J. Andreasson, O. Jönsson, D. Odić, B. Iwan, A. Rocker, and D. Westphal. Single mimivirus particles intercepted and imaged with an x-ray laser. *Nature*, 470(7332):78–81, 2011.
[45] A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36, 2011.
[46] G. Van Der Schot, M. Svenda, F. R. Maia, M. Hantke, D. P. DePonte, M. M. Seibert, A. Aquila, J. Schulz, R. Kirian, and M. Liang. Imaging single cells in a beam of live cyanobacteria with an x-ray laser. *Nature communications*, 6, 2015.
[47] A. Viswanathan and M. Iwen. Fast angular synchronization for phase retrieval via incomplete information. In *SPIE Optical Engineering+ Applications*, pages 959718–959718. International Society for Optics and Photonics, 2015.
[48] A. Walther. The question of phase retrieval in optics. *Optica Acta: International Journal of Optics*, 10(1):41–49, 1963.