UNIVERSITY OF CALIFORNIA, SAN DIEGO

General Phase Retrieval with Locally Supported Measurements

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctorate of Philosophy

in

Mathematics with Specialization in Computational Science

by

Brian P. Preskitt

Committee in charge:

Professor Rayan Saab, Chair
Professor Kamalika Chaudhuri
Professor Todd Kemp
Professor Jiawang Nie
Professor Yoav Freund

2018

The dissertation of Brian P. Preskitt is approved:

Chair

University of California, San Diego

2018

DEDICATION

This dissertation is lovingly dedicated to my brother, Charles Preskitt.

EPIGRAPH

For in much wisdom is much vexation, and he who increaseth knowledge increaseth sorrow. – Ecclesiastes 1:18

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Rayan Saab for all of his involvement in my academic career. In my first year of graduate school, it was his class on compressed sensing that convinced me to switch into the computational math track, which in turn is where I found the material and the problems that motivated me to continue in academic research. Since then, his support – in our many hours of teamwork at the whiteboard; his connections to problems, authors, conferences, and collaborators; and the simple things such as critiques of my writing and talks – was uniquely indispensable in shaping the environment in which I found my place as a researcher. I am immensely thankful for the inspiration of his talent and energetic leadership.

I would also like to thank Mark Iwen and Aditya Viswanathan, our two co-authors and collaborators on the subjects covered in this thesis. The ground that they broke on this new strain of phase retrieval proved to be incredibly fruitful, and I am thankful for being welcomed as a contributor to their ongoing research efforts. This partnership was foundational to my beginnings in research and much of my subsequent progress, and I am deeply indebted to their generous spirit of academic camaraderie. Chapter 3, sections 3.1–3.8, in full, is a reprint of material published with Iwen, Saab, and Viswanathan as published in Applied and Computational Harmonic Analysis 2018. Chapter 3 section **??**, in part, is a reprint of material published with these authors in the Proceedings of SPIE vol. 10394, 2017.

VITA

2013	Bachelor of Science, Texas Christian University
2013-2018	Teaching Assistant, University of California, San Diego
2015	Master of Science, University of California, San Diego
2016-2017	Associate Instructor, University of California, San Diego
2018	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

M. Iwen, B. Preskitt, R. Saab, and A. Viswanathan. *An Eigenvector-Based Angular Synchronization Method for Phase Retrieval from Local Correlation Measurements*. arXiv:1612.01182. Accepted for publication June 2018.

M. Iwen, B. Preskitt, R. Saab, and A. Viswanathan. *Phase retrieval from local measurements in two dimensions*. Proceedings of SPIE 10394, Wavelets and Sparsity XVII, San Diego, CA, 2017.

ABSTRACT

General Phase Retrieval with Locally Supported Measurements

by

Brian P. Preskitt

Doctorate of Philosophy in Mathematics with Specialization in Computational Science

University of California, San Diego, 2018

Professor Rayan Saab, Chair

In this dissertation, we study a new approach to the problem of phase retrieval, which is the task of reconstructing a complex-valued signal from magnitude-only measurements. This problem occurs naturally in several specialized imaging applications such as electron microscopy and X-ray crystallography. Although solutions were first proposed for this problem as early as the 1970s, these algorithms have lacked theoretical guarantees of success, and phase retrieval has suffered from a considerable gap between practice and theory for almost the entire history of its study.

A common technique in fields that use phase retrieval is that of *ptychography*, where

measurements are collected by only illuminating small sections of the sample at any time. We refer to measurements designed in this way as *local measurements*, and in this dissertation, we develop and expand the theory for solving phase retrieval in measurement regimes of this kind. Our first contribution is a basic model for this setup in the case of a one-dimensional signal, along with an algorithm that robustly solves phase retrieval under this model. This work is unique in many ways that represent substantial improvements over previously existing solutions: perhaps most significantly, many of the recovery guarantees in recent work rely on the measurements being generated by a random process, while we devise a class of measurements for which the conditioning of the system is known and quickly checkable (see section 4.2). These advantages constitute major progress towards producing theoretical results for phase retrieval that are directly usable in laboratory settings.

Chapter 1 conducts a survey of the history of phase retrieval and its applications. Chapter 2 reviews the mathematical literature on the subject, including the first solutions and the theoretical work of the last decade. Chapter 3 presents co-authored results defining and establishing the setting and solution of the base model explored in this dissertation. Chapter 4 expands the theory on what measurement schemes are admissible in our model, including an analysis of conditioning and runtime. Chapter 5 explores results that bring our model nearer to the actual practice of ptychography. Chapter 6 includes a few relevant results that may be used for future expansion on this topic.

Chapter 1

History of Phase Retrieval

1.1 Introduction

Phase retrieval is the problem of solving a system of equations of the form

$$y = |Ax_0|^2 + \eta, \tag{1.1}$$

where $x_0 \in \mathbb{C}^d$ is the objective signal, $A \in \mathbb{C}^{D\times d}$ is a known measurement matrix, $\eta \in \mathbb{R}^D$ is an unknown perturbation vector, and $y \in \mathbb{R}^D$ is the vector of measurement data. $|\cdot|^2$ represents the component-wise magnitude squared operation; i.e. for any $n \in \mathbb{N}$ we have $|\mathbf{v}|_j^2 = |\mathbf{v}_j|^2$ for all $\mathbf{v} \in \mathbb{C}^n$. In phase retrieval, the goal is to recover an estimate of x_0 from knowledge of y and A. We sometimes rephrase the system (1.1) as

$$y_j = |\langle a_j, x_0\rangle|^2 + \eta_j, \tag{1.2}$$

where a_j^* stand for the rows of A and are referred to as the measurement vectors. The name *phase retrieval* comes from viewing the $|\cdot|^2$ operation as erasing the phases of the complex-valued measurements $\langle a_j, x_0\rangle$ and leaving only their magnitudes; solving for x_0 may be considered as a way of retrieving this phase information. We immediately note that this problem contains an unavoidable phase

ambiguity, in the sense that, for any solution x and any $\theta \in [0, 2\pi)$, we will have that $e^{i\theta}x$ is also a solution.

The phase retrieval problem appears in a multitude of imaging systems, since most optical sensors – most significantly, charge-coupled devices and photographic film – do not respond to the phase of an incoming light wave. Rather they respond only to the number and energy of photons arriving at its surface, so they indicate only the intensity (absolute value squared), and not the phase, of the electromagnetic waves to which they are exposed. This corresponds to our model in (1.1) by imagining that the i^{th} entry $a_i^* x$ of $Ax \in \mathbb{C}^D$ corresponds to the magnitude and phase of the light arriving at the i^{th} pixel in an array of sensors. When such a sensor responds only to the amount of energy exciting it, we record $|a_i^* x|^2$ at each point and aggregate these data into the vector $|Ax|^2$. Areas of optics that encounter this problem include astronomy [78], diffraction imaging [**?**], and – among the earliest applications, and by far the most celebrated – x-ray crystallography [**?**]. Non-optical disciplines that can benefit from solutions to phase retrieval include speech recognition [4, 63], blind channel estimation [54, 56], and self-calibration [55].

The practice of these disciplines has produced many creative solutions to particular instances of the phase retrieval problem, and throughout the 20^{th} century the field largely evolved by the invention of *ad hoc* solutions that resolved the data at hand. Indeed, in their Nobel prize-winning work in 1915, William and Lawrence Bragg used their intensity-only measurements to deduce the crystal structures of sodium chloride, potassium chloride, and diamonds by largely geometrical analysis of their data based on strong prior knowledge of the atoms present in these materials [14, pp. 88-92, 102-105]. Attempts to systematize this process began in the 1930s with Arthur Lindo Patterson [61], with significant improvements coming with the work of Herbert Hauptman and Jerome Karle in the 1950s [42]. However, these solutions remained primarily non-algorithmic or made strong assumptions about

the crystals being solved, slowing the solution process and limiting the progress made in phase retrieval outside x-ray crystallography. The method proposed by R.W. Gerchberg and W.O. Saxton in 1971 [36] shifted this paradigm by providing an algorithm that can be applied to fairly general data, with remarkably minimal assumptions made on the structure of the object x being detected. This result inspired numerous variants (e.g., [10, 27, 30]), each of which empirically improved performance, but none of which produced a solid mathematical theory to explain why or when they would succeed. Physicists, chemists, and biologists made a great number of astounding scientific achievements in this fashion, but even with all this progress, the community remained largely in want of such a theoretical foundation that could offer reliable solutions in general settings until recent decades.

There are three main questions about phase retrieval problems that the scientific community would wish to answer theoretically: firstly, in an ideal, noiseless case where $\eta = 0$, for what matrices $A \in \mathbb{C}^{D\times d}$ does the system of equations (1.1) possess a unique solution (up to the known phase ambiguity)? Second, given such a case where a unique solution exists, is there an algorithm that can recover it? Third of all, when this recovery process exists, when is it stable in the sense that, in the presence of noise $\eta \neq 0$, the estimate x does not differ too much (or differs to a known degree, as a function of $\|\eta\|$) from x_0?

This dissertation expands upon the theory of phase retrieval by introducing a new class of matrices and an associated recovery algorithm that is proven to solve the system (1.1) with guaranteed stability to noise and with known, competitive computational cost.

Figure 1.1: Experimental setup for x-ray crystallography

1.2 X-ray Crystallography

1.2.1 Historical Preliminaries

The history of phase retrieval cannot be told without making mention of x-ray crystallography, the field that first brought scientific interest to this problem and by many metrics its most decorated and fruitful application. In x-ray crystallography, the goal is to gain an image of the positions of atoms within a molecule by illuminating a crystallized sample with x-rays. The molecular structure is deduced from the pattern of the radiation diffracted by the sample. A rough diagram of this setup is shown in figure 1.1.

This seemingly simple technique has been indispensable for the study of chemistry, biology, and physics, having been used to confirm or identify the arrangements of atoms in a wide variety of important compounds. Over a dozen discoveries made through x-ray crystallography – or made in developing the technique – have been recognized by Nobel Prizes in Physics, Chemistry, and Medicine or Physiology. Indeed, the first Nobel Prize in Physics was awarded to Wilhelm Röntgen in 1901 for his discovery of x-rays. The 1914 Prize in Physics was conferred upon Max von Laue for his discovery the diffraction of x-rays by atomic crystals, and in 1915 William and Lawrence Bragg earned the same distinction for performing the first complete characterizations of atomic crystal structures [35]. Since the time of these highly esteemed pioneering discoveries, x-ray crystallography has been used to produce accurate molecular models of a number of drugs (e.g., [16, 64, 67]), including penicillin in Dorothy Crowfoot Hodgkin's Nobel prize-winning work in 1963 [1]. It has elucidated several human biological compounds, including innumerable proteins [49, 76] and human DNA, for whose analysis in 1953 James

Watson, Maurice Wilkins, and Francis Crick were awarded the Nobel prize in 1962, relying on the crystallographic images of Rosalind Franklin [33, 79, 80, 82]. And this technology remains extremely relevant today, playing an active role in material sciences, where crystallography is being used to characterize the degradation of lithium-ion batteries [43, 69] and to study carbon nanostructures such as fullerenes [50, 51], whose analysis earned the 1996 Nobel Prize in Chemistry [35].

With such a history, x-ray crystallography is an essential application of phase retrieval, and it is informative to state its mathematical formulation in this dissertation.

1.2.2 Mathematical Model

1.2.3 Diffraction as a Fourier transform

We begin by considering the mechanism of diffraction of waves.

1.3 Notation

- For $k \in \mathbb{N}, n \in \mathbb{Z}, [k]_n = \{n, n+1, \ldots, n+k-1\}$ and $[k] = [k]_1$.
- Given $m, n, k \in \mathbb{N}$, $m \bmod_k n$ is the unique element of $[n]_k$ such that $n \mid (m-k)$. Without the subscript, we specify $m \bmod n := m \bmod_0 n \in \{0, \ldots, n-1\}$ to be the usual modulo operator.
- Indices of matrices in $\mathbb{C}^{d\times d}$ and vectors in $\mathbb{C}^d$ are always taken modulo d.
- $S_d \in \mathbb{R}^{d\times d}$ is the $d \times d$ shift operator, such that $(S_d x)_i = x_{i-1}$. Typically we imply the subscript by context, writing S.

- $R_d \in \mathbb{R}^{d\times d}$ is the operator that reverses a vector's entries, leaving the first entry fixed. Namely, $(R_d x)_i = x_{2-i}$. Typically, we imply the dimension d by context and write only R.

- For $i, n \in \mathbb{N}, e_i^n \in \mathbb{R}^n$ is the i^{th} column of the $n \times n$ identity matrix. When context permits, n is implied and we write e_i. In particular, whenever e_i is used in a matrix multiplication, n is taken to be appropriate so that the multiplication is valid; for $A \in \mathbb{C}^{m\times n}$, the "$e_i$" in Ae_i is assumed to be e_i^n.

- Given $x \in \mathbb{C}^d$ and $k \in [d], \text{circ}_k(x) \in \mathbb{C}^{d\times k}$ denotes the first k columns of the circulant matrix whose first column is x, defined by $\text{circ}_k(x)e_i = S^{i-1}x$ for $i \in [k]$. Alternatively,

$$\text{circ}_k(x) = \begin{bmatrix} x & Sx & \cdots & S^{k-1}x \end{bmatrix}.$$

  When the subscript is omitted, $\text{circ}(x) = \text{circ}_d(x)$.

- Given any $A \subseteq B$, we define the indicator function $\chi_A : B \to \{0, 1\}$ by

$$\chi_A(x) = \begin{cases} 1, & x \in A \\ 0, & \text{otherwise} \end{cases}.$$

- $\mathbb{1}_d \in \mathbb{C}^d$ is the vector of all 1's. When context makes the size clear, we write $\mathbb{1}$. Given a set $A \subseteq [d], \mathbb{1}_A^d \in \mathbb{C}^d$ has $(\mathbb{1}_A)_i = \chi_A(i)$.

- $\omega_d := \mathrm{e}^{\frac{2\pi \mathrm{i}}{d}}$ is the d^{th} root of unity. When context permits, d is implied and we use just ω.

- For $k \in \mathbb{Z}, F_k \in \mathbb{C}^{k\times k}$ is the $k \times k$ unitary Fourier matrix with $(F_k)_{ij} = \frac{1}{\sqrt{k}}\omega_k^{(i-1)(j-1)}$.

- For $m, n \in \mathbb{N}, f_n^m = F_m e_n$ is the n^{th} column of the $m \times m$ unitary Fourier matrix, where $e_n \in \mathbb{R}^m$ has its index taken modulo m.

- Given $x, y \in \mathbb{C}^d, x \circ y$ denotes the Hadamard/elementwise product of x and y; specifically $(x \circ y)_i = x_i y_i$.

- Given $A \in \mathbb{C}^{d\times d}, \mathrm{diag}(A, m) \in \mathbb{C}^d$ denotes the m^{th} circulant off-diagonal of A. That is, $\mathrm{diag}(A, m)_i = A_{i,i+m}$.

- Given $x \in \mathbb{C}^d, \mathrm{diag}(x) \in \mathbb{C}^{d\times d}$ is the diagonal matrix whose diagonal entries are the entries of x. Namely, $\mathrm{diag}(x)e_i = x_i e_i$. We may also write this as $\mathrm{diag}(x_j)_{j=1}^d$. When the intention is clear from context, we may write $D_x := \mathrm{diag}(x)$. Given matrices $V_j \in \mathbb{C}^{m_j \times n_j}$ for $j \in [n]$, we write

$$\mathrm{diag}(V_j)_{j=1}^n = \begin{bmatrix} V_1 & & \\ & \ddots & \\ & & V_n \end{bmatrix} \in \mathbb{C}^{\sum m_j \times \sum n_j}.$$

- $\mathcal{H}^d$ is the set of Hermitian matrices in $\mathbb{C}^{d\times d}$, to be viewed as a d^2-dimensional vector space over $\mathbb{R}$. $\mathcal{H}_+^d \subseteq \mathcal{H}^d$ is the complex Hermitian semi-definite cone, where $A \in \mathcal{H}_+^d$ if $A \succeq 0$.

- $\mathcal{R}_d : \bigcup_{k=1}^\infty \mathbb{C}^k \to \mathbb{C}^d$ is a resize mapping, where for $v \in \mathbb{C}^k$ and $i \in [d]$,

$$\mathcal{R}_d(v)_i = \begin{cases} v_i, & i \le k \\ 0, & \text{otherwise} \end{cases} \quad \text{for } i \in [d].$$

  Similarly, $\mathcal{R}_{m\times n} : \bigcup_{k_1,k_2}^\infty \mathbb{C}^{k_1\times k_2} \to \mathbb{C}^{m\times n}$ truncates or zero-pads matrices to size $m \times n$.

- Given $k, d \in \mathbb{N}$, we define the operator $T_k^d : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ by

$$T_k^d(A)_{ij} = \begin{cases} A_{ij}, & |i-j| \bmod d < k \\ 0, & \text{otherwise.} \end{cases}$$

  Note that T_k^d is simply the orthogonal projection operator onto its range $T_k^d(\mathbb{C}^{d\times d})$. We use T_k^d interchangeably to refer to both the operator and its range, and almost always exclude the dimension d by context, writing T_k.

- $\mathcal{S}^n = \mathbb{R}^{n\times n} \cap \mathcal{H}^n$ is the set of real, symmetric $n \times n$ matrices.
- We let vec : $\bigcup_{m,n\in\mathbb{N}} \mathbb{C}^{m\times n} \to \bigcup_{k\in\mathbb{N}} \mathbb{C}^k$ be the columnwise vectorization operator, such that for $A \in \mathbb{C}^{m\times n}, \mathrm{vec}(A) \in \mathbb{C}^{mn}$ and $\mathrm{vec}(A)_{(j-1)m+i} = A_{ij}$ for $i, j \in [m] \times [n]$.
- To invert vec, we use $\mathrm{mat}_{(m,n)} : \mathbb{C}^{mn} \to \mathbb{C}^{m\times n}$, such that $\mathrm{mat}_{(m,n)}(v)_{ij} = v_{(j-1)m+i}$.
- By a slight overloading of notation, by $\mathrm{vec}(a_j)_{j=1}^d$, we intend the vector in $v \in \mathbb{R}^d$ satisfying $v_j = a_j$. (This may come in handy to specify something such as $\begin{bmatrix}1 & 2 & 4 & 8\end{bmatrix}^T = \mathrm{vec}(2^{j-1})_{j=1}^4$).

Chapter 2

Applications

One common technique used to generate redundancy in phase retrieval-type measurements is to design a system that illuminates only a small part of the sample at a time. These "partial snapshots" are then positioned along an overlapping grid, which produces the redundancy. The overlap is necessary since, even if you could solve phase retrieval perfectly on each patch, each of these patches would have their own phase ambiguities; these would need to be synchronized to achieve a single, coherent image of the original sample. Usually, ptychography is performed by taking a single *mask* or *illumination function* with small supprt, say $\mathbf{m} \in \mathbb{C}^d$ with $\mathrm{supp}(m) \subseteq [\delta]$ where $\delta \ll d$ and shifting this mask to different positions relative to the sample. These measurements may then be modelled as

$$\mathbf{y}_{\ell,j} = |\mathcal{F}(S^{\ell}\mathbf{m} \circ \mathbf{x})_j|^2 + \eta_{\ell,j} = |\langle f_j \circ \mathbf{m}, \mathbf{x}\rangle|^2 + \eta_{\ell,j}. \tag{2.1}$$

This technique is the inspiration for our model, where we do not require our measurements to take this exact form.

Chapter 3

Phase retrieval from local correlation measurements

Here we consider the phase retrieval problem as modelled in (2.1).

3.1 Introduction

Consider the problem of recovering a vector $\mathbf{x}_0 \in \mathbb{C}^d$ from measurements $\mathbf{y} \in \mathbb{R}^D$ with entries y_j given by

$$y_j = |\langle \mathbf{a}_j, \mathbf{x}_0 \rangle|^2 + \eta_j, \qquad j = 1, \dots, D. \tag{3.1}$$

Here the measurement vectors $\mathbf{a}_j \in \mathbb{C}^d$ are known and the scalars $\eta_j \in \mathbb{R}$ denote noise terms. This problem is known as the *phase retrieval problem* (see, e.g., [59, 78]), as we may think of the $|\cdot|^2$ in (3.1) as erasing the phases of the measurements $\langle \mathbf{a}_j, \mathbf{x}_0 \rangle$ in an otherwise linear system of equations.

The phase retrieval problem arises in many important signal acquisition schemes, including crystallography and ptychography (e.g., [59], see Figure 3.1), diffrac-

tion imaging [36], and optics [59, 78], among many others. Due to the breadth and importance of the applications, there has been significant interest in developing efficient algorithms to solve this problem. Indeed, one of the first algorithms proposed came in the early 1970's with the work of Gerchberg and Saxton [36]. Since then many variations of their method have been proposed (e.g, [10, 11, 27, 30, 71, 73, 72]) and used widely in practice. On the other hand – until recently – there have not been theoretical guarantees concerning the conditions under which these algorithms recover the underlying signal and the extent to which they can tolerate measurement error. Nevertheless, starting in 2006 a growing body of work (e.g., [5, 4, 7, 13, 19, 26, 46, 53]) has emerged, proposing new methods with theoretical performance guarantees under various assumptions on the signal $\mathbf{x}_0$ and the measurement vectors $\mathbf{a}_j$. Unfortunately, the assumptions (especially on the measurement vectors) often do not correspond to the setups used in practice. In particular, the mathematical analysis often requires that the measurement vectors be random or generic (e.g., [4, 7, 19]) while in practice the measurement vectors are a deterministic aspect of the imaging apparatuses employed. A main contribution of this paper is analyzing a construction that more closely matches practicable and deterministic measurement schemes. We propose a two-stage algorithm for solving the phase retrieval problem in this setting and we analyze our method, providing upper bounds on the associated reconstruction error.

In short, we provide theoretical error guarantees for a numerically efficient reconstruction algorithm in a measurement setting that closely resembles measurements used in practice.

3.1.1 Local Correlation Measurements

Consider the case where the vectors $\mathbf{a}_j$ represent shifts of compactly-supported vectors $\mathbf{m}_j, j = 1, \dots, K$ for some $K \in \mathbb{N}$. Using the notation $[n]_k := \{k, \dots, k + n-1\} \subseteq \mathbb{N}$, and defining $[n] := [n]_1$ we take $\mathbf{x}_0, \mathbf{m}_j \in \mathbb{C}^d$ with $\operatorname{supp}(\mathbf{m}_j) \subseteq [\delta] \subseteq [d]$ for some $\delta \in \mathbb{N}$. We also denote the space of Hermitian matrices in $\mathbb{C}^{k\times k}$ by $\mathcal{H}^k$. Now we have measurements of the form

$$(\mathbf{y}_\ell)_j = |\langle \mathbf{x}_0, S_\ell^* \mathbf{m}_j \rangle|^2, \quad (j, \ell) \in [K] \times P, \tag{3.2}$$

where $P \subseteq [d]_0$ is arbitrary and $S_\ell : \mathbb{C}^d \to \mathbb{C}^d$ is the discrete circular shift operator, namely

$$(S_\ell \mathbf{x}_0)_j = (\mathbf{x}_0)_{\ell+j}.$$

One can see that (3.2) represents the modulus squared of the correlation between $\mathbf{x}_0$ and locally supported measurement vectors. Therefore, we refer to the entries of $\mathbf{y}$ as local correlation measurements. Following [4, 19, 46], the problem may be lifted to a linear system on the space of $\mathbb{C}^{d\times d}$ matrices. In particular, we observe that

$$\begin{aligned}(\mathbf{y}_\ell)_j &= |\langle S_\ell \mathbf{x}_0, \mathbf{m}_j \rangle|^2 = \mathbf{m}_j^* (S_\ell \mathbf{x}_0)(S_\ell \mathbf{x}_0)^* \mathbf{m}_j \\ &= \langle \mathbf{x}_0 \mathbf{x}_0^*, S_\ell^* \mathbf{m}_j \mathbf{m}_j^* S_\ell \rangle,\end{aligned}$$

where the inner product above is the Hilbert-Schmidt inner product. Restricting to the case $P = [d]_0$, for every matrix $A \in \operatorname{span}\{S_\ell^* \mathbf{m}_j \mathbf{m}_j^* S_\ell\}_{\ell,j}$ we have $A_{ij} = 0$ whenever $|i - j| \bmod d \geq \delta$. Therefore, we introduce the family of operators $T_k : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ given by

$$T_k(A)_{ij} = \begin{cases} A_{ij}, & |i-j| \bmod d < k \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

Note that T_δ is simply the orthogonal projection operator onto its range $T_\delta(\mathbb{C}^{d\times d}) \supseteq \operatorname{span}\{S_\ell^* \mathbf{m}_j \mathbf{m}_j^* S_\ell\}_{\ell,j}$; therefore,

$$(\mathbf{y}_\ell)_j = \langle \mathbf{x}_0 \mathbf{x}_0^*, S_\ell^* \mathbf{m}_j \mathbf{m}_j^* S_\ell \rangle = \langle T_\delta(\mathbf{x}_0 \mathbf{x}_0^*), S_\ell^* \mathbf{m}_j \mathbf{m}_j^* S_\ell \rangle, \quad (j, \ell) \in [K] \times P. \tag{3.4}$$

For convenience, we set $D := K|P|$ and define the map $\mathcal{A} : \mathbb{C}^{d\times d} \to \mathbb{C}^D$

$$\mathcal{A}(X) = [\langle X, S_\ell^* \mathbf{m}_j \mathbf{m}_j^* S_\ell \rangle]_{(\ell,j)}. \tag{3.5}$$

Sometimes, we consider $\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})}$, the restriction of $\mathcal{A}$ to the domain $T_\delta(\mathbb{C}^{d\times d})$; indeed, if this linear system is injective on $T_\delta(\mathbb{C}^{d\times d})$, then we can readily solve for

$$T_\delta(\mathbf{x}_0\mathbf{x}_0^*) =: X_0 \tag{3.6}$$

using our measurements $(\mathbf{y}_\ell)_j = (\mathcal{A}(\mathbf{x}_0\mathbf{x}_0^*))_{(\ell,j)}$. In [46], deterministic masks $\mathbf{m}_j$ were constructed for which (3.4) was indeed invertible for certain choices of K and P. An additional construction is given below in §3.2.

Improving on [46], we can further see that $\mathbf{x}_0$ can be deduced from X_0 up to a global phase in the noiseless case as follows: First, X_0 immediately gives the magnitudes of the entries of $\mathbf{x}_0$ since $(X_0)_{ii} = |(x_0)_i|^2$. The only challenge remaining, therefore, is to find $\arg((x_0)_i)$ up to a global phase. We proceed by defining $\tilde{\mathbf{x}}_0$ and $\widetilde{X}_0$ by

$$(\tilde{x}_0)_i = \operatorname{sgn}((x_0)_i)$$

$$(\widetilde{X}_0)_{ij} = \begin{cases} \operatorname{sgn}((X_0)_{ij}), & |i-j| \bmod d < \delta \\ 0, & \text{otherwise} \end{cases},$$

where $\operatorname{sgn} : \mathbb{C} \to \mathbb{C}$ is the usual normalization mapping

$$\operatorname{sgn}(z) = \begin{cases} \dfrac{z}{|z|}, & z \neq 0 \\ 1, & \text{otherwise} \end{cases}.$$

We emphasize that

$$\widetilde{X}_0 = \frac{X_0}{|X_0|} = \frac{T_\delta(\mathbf{x}_0\mathbf{x}_0^*)}{|T_\delta(\mathbf{x}_0\mathbf{x}_0^*)|} \text{ and } \widetilde{\mathbf{x}}_0 = \frac{\mathbf{x}_0}{|\mathbf{x}_0|}, \tag{3.7}$$

where the divisions are taken component-wise. Indeed, in [77], it was shown that the phases of the entries of $\mathbf{x}_0$ (up to a global phase) are given by the leading eigenvector of $\widetilde{X}_0$. Moreover, it was shown that this leading eigenvector is unique.

Lemma 2 of this paper improves in these results by giving a lower bound on the gap between the top two eigenvalues of $\widetilde{X}_0$. This better understanding of the spectrum of $\widetilde{X}_0$ is then leveraged to analyze the robustness of this eigenvector-based phase retrieval method to measurement noise.

3.1.2 Contributions

In this paper, we analyze a phase retrieval algorithm (Algorithm 1) for estimating a vector $\mathbf{x}_0$ from noisy localized measurements of the form

$$(\mathbf{y}_\ell)_j = |\langle \mathbf{x}_0, S_\ell^* \mathbf{m}_j \rangle|^2 + n_{j\ell}, \quad (j,\ell) \in [2\delta - 1] \times [d]_0. \tag{3.8}$$

This algorithm is composed of two main stages. First, we apply the inverse of the linear operator

$$\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})} : T_\delta(\mathbb{C}^{d\times d}) \to \mathbb{C}^{(2\delta-1)d}$$

defined immediately after (3.5), to obtain a Hermitian estimate X of $T_\delta(\mathbf{x}_0\mathbf{x}_0^*)$ given by

$$X = \left((\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})})^{-1}\mathbf{y}\right)/2 + \left((\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})})^{-1}\mathbf{y}\right)^*/2 \in T_\delta(\mathbb{C}^{d\times d}). \tag{3.9}$$

In particular, our choice of $\mathbf{m}_j$ as described in Section 3.2 ensures that $\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})}$ is both invertible and well conditioned. Next, once we have an approximation of $T_\delta(\mathbf{x}_0\mathbf{x}_0^*)$, we estimate the magnitudes and phases of the entries of $\mathbf{x}_0$ separately.

For the magnitudes, we simply use the square-roots of the diagonal entries of X. For the phases, we use the normalized eigenvector corresponding to the top eigenvalue of

$$\widetilde{X} := \frac{X}{|X|}, \tag{3.10}$$

where the operations are considered element-wise. The hope is that the leading eigenvector of $\widetilde{X}$ will serve as a good approximation to the leading eigenvector

of $\widetilde{X}_0$, which is seen in Section 3.3 (see also [77]) to indeed be a scaled version of the phase vector $\widetilde{\mathbf{x}}_0$ (up to a global phase ambiguity). The entire method is summarized in Algorithm 1, and its associated recovery guarantees are presented in Theorem 1, while its computational complexity is discussed after the theorem in §3.1.3. Here and throughout the paper, $\mathrm{e} = 2.71828\ldots$ refers to the base of the natural logarithm and i refers to the imaginary unit.

Algorithm 1 Fast Phase Retrieval from Local Correlation Measurements

Input: Measurements $\mathbf{y} \in \mathbb{R}^D$ as per (3.8)

Output: $\mathbf{x} \in \mathbb{C}^d$ with $\mathbf{x} \approx \mathrm{e}^{-\mathrm{i}\theta}\mathbf{x}_0$ for some $\theta \in [0, 2\pi]$

1: Compute the Hermitian matrix $X = \left((\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})})^{-1}\mathbf{y}\right)/2 + \left((\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})})^{-1}\mathbf{y}\right)^*/2 \in T_\delta(\mathbb{C}^{d\times d})$ as an estimate of $T_\delta(\mathbf{x}_0\mathbf{x}_0^*)$

2: Form the banded matrix of phases, $\tilde{X} \in T_\delta(\mathbb{C}^{d\times d})$, by normalizing the non-zero entries of X

3: Compute the top eigenvector $u \in \mathbb{C}^d$ of $\tilde{X}$ and set $\widetilde{\mathbf{x}} := \mathrm{sgn}(u)$.

4: Set $x_j = \sqrt{X_{j,j}} \cdot (\tilde{x})_j$ for all $j \in [d]$ to form $\mathbf{x} \in \mathbb{C}^d$

Theorem 1. *Suppose $\delta > 2$ and $d \geq 4\delta$. Let $(x_0)_{\min} := \min_j |(x_0)_j|$ be the smallest magnitude of any entry in $\mathbf{x}_0 \in \mathbb{C}^d$. Then, the estimate $\mathbf{x}$ produced in Algorithm 1 satisfies*

$$\min_{\theta\in[0,2\pi]} \left\|\mathbf{x}_0 - \mathrm{e}^{\mathrm{i}\theta}\mathbf{x}\right\|_2 \leq C\left(\frac{\|\mathbf{x}_0\|_\infty}{(x_0)_{\min}^2}\right)\left(\frac{d}{\delta}\right)^2 \kappa\|\mathbf{n}\|_2 + Cd^{\frac{1}{4}}\sqrt{\kappa\|\mathbf{n}\|_2},$$

where $\kappa > 0$ is the condition number of the system (3.9) *and $C \in \mathbb{R}^+$ is an absolute universal constant.*

Theorem 1, which deterministically depends on both the masks and the signal, provides improvements over the first deterministic theoretical robust recovery guarantees proven in [46] for a wide class of non-vanishing signals. Momentarily consider, e.g., the class of "flat" vectors $\mathbf{x}_0 \in \mathbb{C}^d$ for which both (i) $(x_0)_{\min} \geq \frac{\|\mathbf{x}_0\|_2}{2\sqrt{d}}$, and (ii) $\left(\frac{\|\mathbf{x}_0\|_\infty}{(x_0)_{\min}^2}\right) \leq \tilde{C}$ for some absolute constant $\tilde{C} \in \mathbb{R}^+$, hold. The main deterministic

result of [46] also applies to this class of vectors and states that an algorithm exists which can achieve the following robust recovery guarantee.

Theorem 2 (See Theorem 5 in [46])**.** *There exist fixed universal constants* $C, C' \in \mathbb{R}^+$ *such that the following holds for all* $\mathbf{x}_0 \in \mathbb{C}^d$ *of the class mentioned above: Let* $\|\mathbf{x}_0\|_2 \geq Cd\sqrt{(\delta-1)\|\mathbf{n}\|_2}$*. Then, the algorithm in [46], when provided with noisy measurements of* $\mathbf{x}_0 \in \mathbb{C}^d$ (3.8) *resulting from the masks discussed in Example 1 of §3.2, will output a vector* $\mathbf{x} \in \mathbb{C}^d$ *satisfying*

$$\min_{\theta\in[0,2\pi]} \left\|\mathbf{x}_0 - \mathbb{e}^{\mathbb{i}\theta}\mathbf{x}\right\|_2 \leq C'd\sqrt{(\delta-1)\|\mathbf{n}\|_2}.$$

Comparing the error bounds provided by Theorems 1 and 2 for the class of flat vectors $\mathbf{x}_0$ mentioned above when using measurements resulting from the masks discussed in Example 1 of §3.2,[1] we can see that Theorem 1 makes the following improvements over Theorem 2:

- Theorem 1 improves on the error bound of Theorem 2 for arbitrary small-norm noise $\mathbf{n}$ having $\|\mathbf{n}\|_2 = \mathcal{O}\left(\delta/d^2\right)$.
- Theorem 2's error bound breaks down entirely for noise $\mathbf{n}$ with ℓ^2-norm on the order of $\|\mathbf{n}\|_2 = \Theta\left(\frac{\|\mathbf{x}_0\|_2^2}{\delta d^2}\right)$. Theorem 1's error bound, on the other hand, still provides non-trivial error guarantees for such noise levels as long as $\|\mathbf{x}_0\|_2 = \mathcal{O}(\delta)$.[2]

In addition, Theorem 1 also applies to a more general set of masks and a larger class of signals $\mathbf{x}_0$ than Theorem 2 does. And, perhaps most importantly, Algorithm 1 generally outperforms the algorithm referred to by Theorem 2 numerically for all

[1] Note that the condition number κ mentioned in Theorem 1 for the masks discussed in Example 1 of §3.2 is $\mathcal{O}(\delta^2)$. See Theorem 3 below for a more exact statement. All asymptotic notation is with respect to $d \to \infty$. Below δ is always assumed to be independent of (and less than) d unless otherwise noted.

[2] This allows Theorem 1 to cover, e.g., the case of larger $\delta = \Omega\left(\sqrt{d}\right)$.

noise levels (see, e.g., Figures 3.2a and 3.6a in §3.6). Theorem 1 provides theoretical error guarantees for this numerically improved method.

We note that the $\mathcal{O}\left(\left(\frac{d}{\delta}\right)^2\right)$-factor in the first term of the error bound provided by Theorem 1 is probably suboptimal, especially in practice. Indeed, Theorem 1 provides a worst-case error guarantee that holds for any arbitrary (including worst-case/adversarial) perturbation $\mathbf{n}$ of the measurements (3.8). However, while the quadratic dependence on d/δ is probably suboptimal, *some* dependence on d/δ likely exists for worst-case additive-noise in our local measurement setting. There is, e.g., numerical evidence that the empirical noise robustness of many phase retrieval methods deteriorates as d grows for local measurements whose support size δ is held fixed. We will leave a rigorous theoretical investigation of the optimal scaling of such noise robustness guarantees with d/δ to future work. For the time being we will simply note here that the $\mathcal{O}\left(\left(\frac{d}{\delta}\right)^2\right)$-factor is mainly a product of the relatively small eigenvalue gap of the matrix $\widetilde{X}_0$ defined in (3.7) above. See §3.3 below for more details.

3.1.3 The Runtime Complexity of Algorithm 1

Consider now the computational complexity of Algorithm 1 (assuming, of course, that $\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})}$ is actually invertible). One can see that line 1 can always be done in at most $\mathcal{O}(d\cdot\delta^3+\delta\cdot d\log d)$ flops using a block circulant matrix factorization approach (see Section 3.1 in [46]). In certain cases one can improve on this; for example, the second (new) mask construction of Section 3.2 allows line 1 to be performed in only $\mathcal{O}(d\cdot\delta)$ flops. Even in the worst case, however, if one precomputes this block circulant matrix factorization in advance given the masks $\mathbf{m}_j$ then line 1 can always be done in $\mathcal{O}(d\cdot\delta^2+\delta\cdot d\log d)$ flops thereafter.

The top eigenvector $\tilde{\mathbf{x}}$ of $\tilde{X}$ is guaranteed to be found in line 3 of Algorithm 1 in the

low-noise (e.g., noiseless) setting via the shifted inverse power method with shift $\mu := 2\delta - 1$ and initial vector $\mathbf{e}_1$ (the first standard basis vector). More generally, one may utilize the Rayleigh quotient iteration with the initial eigenvalue estimate fixed to $2\delta - 1$ for the first few iterations. In either case, each iteration can be accomplished with $\mathcal{O}(d \cdot \delta^2)$ flops due to the banded structure of $\tilde{X}$ (see, e.g., [75]). In the low-noise setting the top eigenvector $\tilde{\mathbf{x}}$ can be computed to machine precision in $\mathcal{O}(\log d)$ such iterations,[3] for a total flop count of $\mathcal{O}(\delta^2 \cdot d \log d)$ for line 3 in that case. In total, then, one can see that Algorithm 1 will always require just $\mathcal{O}(\delta^2 \cdot d \log d + d \cdot \delta^3)$ total flops in low-noise settings. Furthermore, in all such settings a measurement mask support of size $\delta = \mathcal{O}(\log d)$ appears to suffice.

3.1.4 Connection to Ptychography

In ptychographic imaging (see Fig. 3.1), small regions of a specimen are illuminated one at a time and an intensity[4] detector captures each of the resulting diffraction patterns. Thus each of the ptychographic measurements is a local measurement, which under certain assumptions (e.g., appropriate wavelength of incident radiation, far-field Fraunhofer approximation), can be modeled as [24, 37]

$$y(t, \omega) = \left| \mathcal{F}[\widetilde{h} \cdot S_t f](\omega) \right|^2 + \eta(t, \omega). \tag{3.11}$$

[3] To see why $\mathcal{O}(\log d)$ iterations suffice one can appeal to lemmas 1 and 2 below. Let $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_d|$ be the eigenvalues of $\tilde{X}$ with associated orthonormal eigenvectors $\mathbf{u}_j \in \mathbb{C}^d$. Let $\delta := |\lambda_1| - |\lambda_2| > 0$. When the noise level is sufficiently low (so that $\tilde{X} \approx \tilde{X}_0$) one will have both (i) $|\mathbf{e}_1^* \mathbf{u}_j| = \Theta(1/\sqrt{d})$ $\forall j \in [d]$, and (ii) $\mu \in (\lambda_1 - \delta/4, \lambda_1 + \delta/4)$ be true. Thus, we will have that there exists some unit norm $\mathbf{r} \in \mathbb{C}^d$ such that

$$\frac{\left(\tilde{X} - \mu I\right)^{-k} \mathbf{e}_1}{\left\| \left(\tilde{X} - \mu I\right)^{-k} \mathbf{e}_1 \right\|_2} = \frac{\mathbf{u}_1 + \sum_{j=2}^{d} \mathcal{O}\left(\left| \frac{\lambda_1 - \mu}{\lambda_j - \mu} \right|^k \right) \mathbf{u}_j}{1 + \mathcal{O}\left(\frac{d}{9^k}\right)} = \mathbf{u}_1 + \mathcal{O}\left(\frac{d}{3^k}\right) \mathbf{r}$$

holds for any given integer $k = \Omega(\log_3 d)$.

[4] By intensity, we mean magnitude squared.

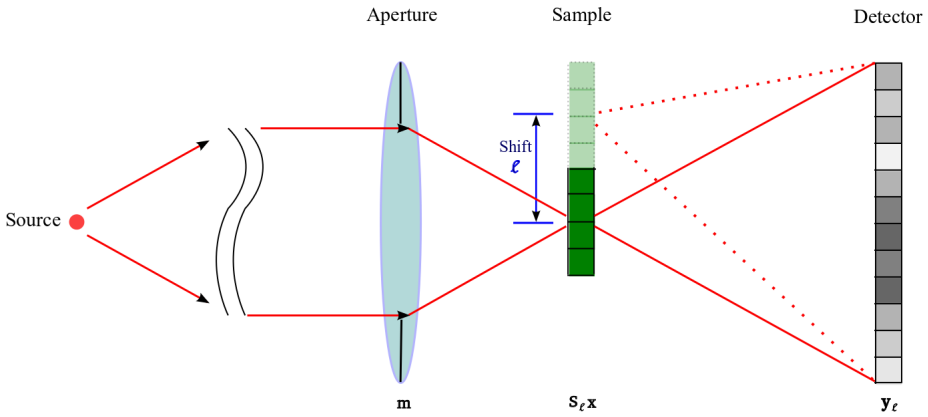


Figure 3.1: Illustration of one-dimensional ptychographic imaging (Adapted from "Fly-scan ptychography", Huang et al., Scientific Reports 5 (9074), 2015.)

Here, $\mathcal{F}$ denotes the Fourier transform, $f : [0,1] \to \mathbb{C}$ represents the unknown test specimen, S_t is the shift operator defined via

$$(S_t f)(s) := f(s+t),$$

and $\widetilde{h} : [0,1] \to \mathbb{C}$ is the so-called illumination function [83] of the imaging system. To account for the local nature of the measurements in (3.11), we assume that $\mathrm{supp}(\widetilde{h}) \subseteq \mathrm{supp}(f)$.

As the phase retrieval problem is inherently non-linear and requires sophisticated computer algorithms to solve, consider the discrete version of (3.11), with $\widetilde{\mathbf{m}}, \mathbf{x}_0 \in \mathbb{C}^d$ discretizing $\widetilde{h}$ and f. Thus (3.11), in the absence of noise, becomes

$$(\mathbf{y}_\ell)_j = \left| \sum_{n=1}^{d} \widetilde{m}_n \, (x_0)_{n+\ell} \, \mathrm{e}^{-\frac{2\pi \mathrm{i}(j-1)(n-1)}{d}} \right|^2, \quad (j,\ell) \in [d] \times [d]_0, \tag{3.12}$$

where indexing is considered modulo-d, so $(\mathbf{y}_\ell)_j$ is a diffraction measurement corresponding to the j^{th} Fourier mode of a circular ℓ-shift of the specimen. We use circular shifts for convenience and we remark that this is appropriate as one can zero-pad $\mathbf{x}_0$ and $\widetilde{\mathbf{m}}$ in (3.12) and obtain the same $(\mathbf{y}_\ell)_j$ as one would with non-circular shifts. In practice, one may not need to use all the shifts $\ell \in [d]_0$ as a

subset may suffice. Defining $\mathbf{m}_j \in \mathbb{C}^d$ by

$$(\mathbf{m}_j)_n = \overline{\widetilde{m}_n}\, \mathbb{e}^{\frac{2\pi \mathbb{i}(j-1)(n-1)}{d}} \tag{3.13}$$

and rearranging (3.12), we obtain

$$\begin{aligned}(\mathbf{y}_\ell)_j &= \left|\sum_{n=1}^{d}(x_0)_{n+\ell}\,\overline{(\mathbf{m}_j)_n}\right|^2 = \left|\sum_{n=1}^{\delta}(x_0)_{n+\ell}\,\overline{(\mathbf{m}_j)_n}\right|^2 \\ &= |\langle S_\ell \mathbf{x}_0, \mathbf{m}_j\rangle|^2 = \langle S_\ell \mathbf{x}_0\mathbf{x}_0^* S_\ell^*, \mathbf{m}_j\mathbf{m}_j^*\rangle \\ &= \langle T_\delta(\mathbf{x}_0\mathbf{x}_0^*), S_\ell^*\mathbf{m}_j\mathbf{m}_j^* S_\ell\rangle, \quad (j,\ell)\in[d]\times[d]_0\end{aligned} \tag{3.14}$$

where the second and last equalities follow from the fact that $\widetilde{\mathbf{m}}$ (and hence each $\mathbf{m}_j$) is locally supported. We note that (3.14) defines a correlation with local masks or window functions $\mathbf{m}_j$. More importantly, (3.14) shows that ptychography (with ℓ ranging over any subset of $[d]_0$) represents a case of the general system seen in (3.4).

3.1.5 Connections to Masked Fourier Measurements

Often, in imaging applications involving phase retrieval, a mask is placed either between the illumination source and the sample or between the sample and the sensor. Here, we will see that the mathematical setup that we consider is applicable in this scenario, albeit when the masks are band-limited. As before, let $\mathbf{x}_0, \mathbf{m} \in \mathbb{C}^d$ denote the unknown signal of interest, and a known mask (or window), respectively. Moreover, for a vector $\mathbf{x}_0 \in \mathbb{C}^d$ we denote its discrete Fourier transform $\widehat{\mathbf{x}_0} \in \mathbb{C}^d$ by

$$(\widehat{x_0})_k := \sum_{n=1}^{d}(x_0)_n e^{-2\pi\mathbb{i}(n-1)(k-1)/d}.$$

Here, we consider squared magnitude *windowed Fourier transform* measurements of the form

$$(\mathbf{y}_\ell)_k = \left|\sum_{n=1}^{d}(x_0)_n\, m_{n-\ell}\, \mathbb{e}^{-\frac{2\pi\mathbb{i}(k-1)(n-1)}{d}}\right|^2,\ k\in[d],\ \ell\in\{\ell_1,\dots,\ell_L\}\subseteq[d]_0. \tag{3.15}$$

As before, ℓ denotes a shift or translation of the mask/window, so $(\mathbf{y}_\ell)_k$ corresponds to the (squared magnitude of) the k^{th} Fourier mode associated with an ℓ-shift[5] of the mask $\mathbf{m}$. Defining the modulation operator, $W_k : \mathbb{C}^d \mapsto \mathbb{C}^d$, by its action $(W_k \mathbf{x}_0)_n = e^{2\pi \mathbb{i}(k-1)(n-1)/d} \, (x_0)_n$ and applying elementary Fourier transform properties[6] one has

$$\begin{aligned}
(\mathbf{y}_\ell)_k &= |\langle \mathbf{x}_0, S_{-\ell}(e^{2\pi \mathbb{i}(k-1)\ell/d} W_k \overline{\mathbf{m}})\rangle|^2 \\
&= |\langle \mathbf{x}_0, S_{-\ell}(W_k \overline{\mathbf{m}})\rangle|^2 = |\langle \widehat{\mathbf{x}_0}, \widehat{S_{-\ell}(W_k \overline{\mathbf{m}})}\rangle|^2 \\
&= |\langle \widehat{\mathbf{x}_0}, W_{-\ell+1}(S_{-k+1} \widehat{\overline{\mathbf{m}}})\rangle|^2 \\
&= |\langle \widehat{\mathbf{x}_0}, S_{-k+1}(W_{-\ell+1} \widehat{\overline{\mathbf{m}}})\rangle|^2. \qquad (3.16)
\end{aligned}$$

Defining $\widehat{\mathbf{m}}_\ell := W_{-\ell+1} \widehat{\overline{\mathbf{m}}}$ and assuming that $\mathrm{supp}(\widehat{\overline{\mathbf{m}}}) \subseteq [\delta]$ (e.g., assuming that $\mathbf{m}$ is real-valued and band-limited), we now have that

$$\begin{aligned}
(\mathbf{y}_\ell)_k &= \langle \widehat{\mathbf{x}}_0 \widehat{\mathbf{x}}_0^*, S_{-k+1} \widehat{\mathbf{m}}_\ell \widehat{\mathbf{m}}_\ell^* S_{-k+1}^* \rangle \\
&= \langle T_\delta(\widehat{\mathbf{x}}_0 \widehat{\mathbf{x}}_0^*), S_{-k+1} \widehat{\mathbf{m}}_\ell \widehat{\mathbf{m}}_\ell^* S_{-k+1}^* \rangle,
\end{aligned}$$

which again represents a case of the general system seen in (3.4). Moreover, our results all hold for this setting, albeit with the Fourier transforms of signals and conjugated masks.

3.1.6 Related Work

The first approach to the phase retrieval problem was proposed in the 1970's in [36] by Gerchberg and Saxton, where the measurement data corresponded to knowing the magnitude of both the image $\mathbf{x}_0$ and its Fourier transform. This result was famously expanded upon by Fienup [30] later that decade, one significant improvement being that only the magnitude of the Fourier transform of $\mathbf{x}_0$ must be known

[5] As above, all indexing and shifts are considered modulo-d.

[6] $\widehat{S_\ell \mathbf{x}_0} = W_{\ell+1} \widehat{\mathbf{x}_0}$, $\widehat{W_k \mathbf{x}_0} = S_{-k+1} \widehat{\mathbf{x}_0}$, and $W_k S_\ell \mathbf{x}_0 = e^{-2\pi \mathbb{i}(k-1)\ell/d} S_\ell W_k \mathbf{x}_0$.

in the case of a signal $\mathbf{x}_0$ belonging to some fixed convex set $\mathcal{C}$ (typically, $\mathcal{C}$ is the set of non-negative, real-valued signals restricted to a known domain). Though these techniques work well in practice and have been popular for decades, they are notoriously difficult to analyze (see, e.g., [10, 11, 27, 71, 73, 72]). These are iterative methods that work by improving an initial guess until they stagnate. Recently Marchesini et al. proved that alternating projection schemes using generic measurements are guaranteed to converge to the correct solution *if provided with a sufficiently accurate initial guess* and algorithms for ptychography were explored in particular [57]. However, no global recovery guarantees currently exist for alternating projection techniques using local measurements (i.e., finding a sufficiently accurate initial guess is not generally easy).

Other authors have taken to proving probabilistic recovery guarantees when provided with globally supported Gaussian measurements. Methods for which such results exist vary in their approach, and include convex relaxations [18, 19], gradient descent strategies [21], graph-theoretic [2] and frame-based approaches [6, 13], and variants on the alternating minimization (e.g., with resampling) [60].

Several recovery algorithms achieve theoretical recovery guarantees while using at most $D = \mathcal{O}(d \log^4 d)$ masked Fourier coded diffraction pattern measurements, including both *PhaseLift* [20, 41], and *Wirtinger Flow* [21]. However, these measurements are both randomized (which is crucial to the probabilistic recovery guarantees developed for both PhaseLift and Wirtinger Flow – deterministic recovery guarantees do not exist for either method in the noisy setting), and provide global information about $\mathbf{x}_0$ from each measurement (i.e., the measurements are not locally supported).

Among the first treatments of local measurements are [12, 25] and [48], in which it is shown that STFT measurements with specific properties can allow (sparse) phase retrieval in the noiseless setting, and several recovery methods are proposed. Sim-

ilarly, the phase retrieval approach from [2] was extended to STFT measurements in [66] in order to produce recovery guarantees in the noiseless setting. More recently, randomized robustness guarantees were developed for time-frequency measurements in [62]. However, no *deterministic* robust recovery guarantees have been proven in the noisy setting for any of these approaches. Furthermore, none of the algorithms developed in these papers are empirically demonstrated to be competitive numerically with standard alternating projection techniques for large signals when utilizing windowed Fourier and/or correlation-based measurements. In [46], the authors propose the measurement scheme developed in the current paper and prove the first deterministic robustness results for a different greedy recovery algorithm.

3.1.7 Organization

Section 3.2 discusses two collections of local correlation masks $\mathbf{m}_j$, one of which is novel and the other of which was originally studied in [46]. Most importantly, Section 3.2 shows that the recovery of $T_\delta(\mathbf{x}_0\mathbf{x}_0^*)$ from measurements associated with the proposed masks can be done stably in the presence of measurement noise. Moreover, since in the noisy regime, the leading eigenvector $\widetilde{\mathbf{x}}$ of $\widetilde{X}$ (associated with line 3 of Algorithm 1) will no longer correspond exactly to the true phases $\widetilde{\mathbf{x}}_0$, we are interested in a perturbation theory for the eigenvectors of $\widetilde{X}_0$. Intuitively, $\widetilde{\mathbf{x}}$ will be most accurate when the eigenvalue of $\widetilde{X}_0$ associated with $\widetilde{\mathbf{x}}_0$ is well separated from the rest of the eigenvalues and so, accordingly, Section 3.3 studies the spectrum of $\widetilde{X}_0$. Indeed, this eigenvalue is rigorously shown to control the stability of the top eigenvector of $\widetilde{X}_0$ with respect to noise, and Section 3.4 develops perturbation results concerning their top eigenvectors by adapting the spectral graph techniques used in [2]. Recovery guarantees for the proposed phase retrieval method are then compiled in Section 3.5. Numerical results demonstrating the

accuracy, efficiency, and robustness of the proposed methods are finally provided in Section 3.6[7], while Section 3.7 contains some concluding remarks and avenues for further research. In Appendix 3.8, we provide an alternate, weaker but easier to derive eigenvector perturbation result analogous to the one in Section 3.4 which may be of independent interest.

3.2 Well-conditioned measurement maps

Here, we present two example constructions for which the linear operator $\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})}$ used in Step 1 of Algorithm 1 is well conditioned. Such constructions are crucial for the stability of the method to additive noise.

Example 1:

In [46], a construction was proposed for the masks $\mathbf{m}_\ell$ in (3.2) that guarantees the stable invertibility of $\mathcal{A}$. This construction comprises windowed Fourier measurements with parameters $\delta \in \mathbb{Z}^+$ and $a \in [4,\infty)$ corresponding to the $2\delta-1$ masks $\mathbf{m}_j \in \mathbb{C}^d$, $j = 1,...,2\delta-1$ with entries given by

$$(\mathbf{m}_j)_n = \begin{cases} \frac{\mathrm{e}^{-n/a}}{\sqrt[4]{2\delta-1}} \cdot \mathrm{e}^{\frac{2\pi \mathrm{i}\cdot(n-1)\cdot(j-1)}{2\delta-1}} & \text{if } n \le \delta \\ 0 & \text{if } n > \delta \end{cases}. \tag{3.17}$$

Here, measurements using all shifts $\ell = 1,...,d$ of each mask are taken. In the notation of (3.4), this corresponds to $K = 2\delta-1$ and $P = [d]_0$, which yields $D = (2\delta-1)d$ total measurements. By considering the basis $\{E_{ij}\}$ for $T_\delta(\mathbb{C}^{d\times d})$ given by

$$E_{i,j}(s,t) = \begin{cases} 1, & (i,j) = (s,t) \\ 0, & \text{otherwise} \end{cases}$$

[7]MATLAB code to run the BlockPR algorithm is available online at [47].

it was shown in [46] that this system is both well conditioned and rapidly invertible. In particular, if M' is the matrix representing the measurement mapping $\mathcal{A} : T_\delta(\mathbb{C}^{d\times d}) \to T_\delta(\mathbb{C}^{d\times d})$ with respect to the basis $\{E_{ij}\}$, the following estimates of the condition number and cost of inversion hold.

Theorem 3 ([46]). *Consider measurements of the form* (3.17) *with* $a := \max\left\{4,\ \frac{\delta-1}{2}\right\}$. *Let* $M' \in \mathbb{C}^{D\times D}$ *be the matrix representing the measurement mapping* $\mathcal{A} : T_\delta(\mathbb{C}^{d\times d}) \to T_\delta(\mathbb{C}^{d\times d})$ *with respect to the basis* $\{E_{ij}\}$. *Then, the condition number of* M' *satisfies*

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

and the smallest singular value of M' *satisfies*

$$\sigma_{\min}\left(M'\right) > \frac{7}{20a}\cdot \mathrm{e}^{-(\delta+1)/a} > \frac{C}{\delta}$$

for an absolute constant $C \in \mathbb{R}^+$. *Furthermore,* M' *can be inverted in* $\mathcal{O}\left(\delta \cdot d\log d\right)$*-time.*

This theorem indicates that one can both efficiently and stably solve for $\mathbf{x}_0\mathbf{x}_0^*$ using (3.4) with the measurements given in (3.17). This measurement scheme is also interesting because it corresponds to a ptychography system if we take the illumination function (i.e. the physical mask) in (3.12) to be $\widetilde{m}_n = \frac{\mathrm{e}^{-n/a}}{\sqrt[4]{2\delta-1}}$ and assume that $d = k(2\delta-1)$ for some $k \in \mathbb{N}$; in practice, this may be achieved by zero-padding the specimen. Then we may take the subset of the measurements (3.13) given by $j = (p-1)k+1,\ p \in [2\delta-1]$ to obtain the masks specified in (3.17). We also remark that in this setup, only one physical mask is required, as the index j in (3.17) denotes the different frequencies observed in the Fourier domain at the sensor array.

Example 2:

We provide a second deterministic construction that improves on the condition number of the previous collection of measurement vectors. We merely set $\mathbf{m}_1 = e_1, \mathbf{m}_{2j} = e_1 + e_{j+1}$, and $\mathbf{m}_{2j+1} = e_1 + ie_{j+1}$ for $j = 1, \dots, \delta - 1$. A simple induction shows that $\{S_\ell \mathbf{m}_j \mathbf{m}_j^* S_\ell^*\}_{\ell \in [d]_0, j \in [2k-1]}$ is a basis for $T_k(\mathbb{C}^{d\times d})$, so if we take $\mathbf{m}_1, \dots, \mathbf{m}_{2\delta-1}$ for our masks we'll have a basis for $T_\delta(\mathbb{C}^{d\times d})$. Indeed, if we let

$$\mathcal{B} : T_k(\mathbb{C}^{d\times d}) \to \mathbb{C}^{\delta \times d}$$

be the measurement operator defined via

$$\big(\mathcal{B}(X)\big)_{\ell,j} = \langle S_\ell \mathbf{m}_j \mathbf{m}_j^* S_\ell^*, X\rangle, \quad (\ell, j) \in [d]_0 \times [2k-1]$$

we can immediately solve for the entries of $X \in T_k(\mathcal{H}^{d\times d})$ from $\mathcal{B}(X) =: B$ by observing that

$$\begin{aligned} X_{i,i} &= B_{i-1,1} \\ X_{i,i+k} &= \tfrac{1}{2}B_{i-1,2k} + \tfrac{i}{2}B_{i-1,2k+1} - \tfrac{1+i}{2}(B_{i-1,1} + B_{i+k-1,1}), \end{aligned}$$

where we naturally take the indices of B mod d. This leads to an upper triangular system if we enumerate X by its diagonals; namely we regard $T_\delta(\mathcal{H}^{d\times d})$ as a $d(2\delta - 1)$ dimensional vector space over $\mathbb{R}$ and set, for $i \in [d]$

$$z_{kd+i} = \begin{cases} \mathrm{Re}(X_{i,i+k}), & 0 \le k < \delta \\ \mathrm{Im}(X_{i,i+k-\delta+1}), & \delta \le k < 2\delta - 1 \end{cases}, \quad y_{kd+i} = \begin{cases} \mathcal{B}(X)_{i,1}, & k = 0 \\ \mathcal{B}(X)_{i,2k}, & 1 \le k < \delta \\ \mathcal{B}(X)_{i,2(k-\delta+1)+1}, & \delta \le k < 2\delta - 1 \end{cases}.$$

Then with $S = S_1 \in \mathbb{R}^{d\times d}$ representing the circular shift operator as before, we have

$$y = \begin{bmatrix} I_d & 0 & 0 \\ D & 2I_{d(\delta-1)} & 0 \\ D & 0 & 2I_{d(\delta-1)} \end{bmatrix} z =: Cz, \text{ where } D = \begin{bmatrix} I_d + S \\ I_d + S^2 \\ \vdots \\ I_d + S^{\delta-1} \end{bmatrix}.$$

Since the matrix C is upper triangular, its inverse is immediate:

$$C^{-1} = \begin{bmatrix} I_d & 0 & 0 \\ -D/2 & I_{d(\delta-1)}/2 & 0 \\ -D/2 & 0 & I_{d(\delta-1)}/2 \end{bmatrix}.$$

To ascertain the condition number of $\mathcal{B}$, then, all we need is the extremal singular values of C. We bound the top singular value by considering

$$\begin{aligned} \sigma_{\max}(C) = \max_{||w||^2+||v||^2=1} \left\| C \begin{bmatrix} w \\ v \end{bmatrix} \right\| &= \left\| \begin{bmatrix} w \\ Dw \\ Dw \end{bmatrix} + 2 \begin{bmatrix} 0 \\ v \end{bmatrix} \right\| \\ &\leq \sqrt{||w||^2 + 2||w+Sw||^2 + \cdots + 2||w+S^{\delta-1}w||^2} + ||2v|| \\ &\leq \sqrt{8(\delta-1)+1}||w|| + 2||v|| \leq \sqrt{8(\delta-1)+5} \leq 2\sqrt{2\delta}, \end{aligned}$$

where in the last line we have used $||w||^2 + ||v||^2 = 1$. By a nearly identical argument, we find

$$\frac{1}{\sigma_{\min}(C)} = \sigma_{\max}(C^{-1}) \leq \sqrt{2\delta}$$

so that the condition number is bounded by $\kappa(C) \leq 4\delta$.

3.3 The Spectrum of $\tilde{X}_0$

Consider line 3 of Algorithm 1, which shows that we are trying to recover $\widetilde{\mathbf{x}}_0 := \frac{\mathbf{x}_0}{|\mathbf{x}_0|}$ via an eigenvector method. Here, we show that $\widetilde{X}_0$ has $\widetilde{\mathbf{x}}_0$ as its top eigenvector and we investigate the spectral properties of $\widetilde{X}_0$ in this section, following the intuition that the eigenvalue gap $|\lambda_1 - \lambda_2|$ will affect the robustness of the spectral step in the algorithm.

For the remainder of the paper, we let $\mathbb{1}$ refer to a constant vector of all ones; its size will always be determined by context. To begin, consider $U = T_\delta(\mathbb{1}\mathbb{1}^*)$, i.e.,

$$U_{j,k} = \begin{cases} 1 & \text{if } |j-k| \bmod d < \delta \\ 0 & \text{otherwise} \end{cases}. \tag{3.18}$$

Observe that U is circulant for all δ, so its eigenvectors are always discrete Fourier vectors. Setting $\omega_j = \mathrm{e}^{2\pi \mathrm{i} \frac{j-1}{d}}$ for $j = 1, 2, \ldots, d$, one can also see that the eigenvalues of U are given by

$$\nu_j = \sum_{k=1}^{d} (U)_{1,k} \omega_j^{k-1} = 1 + \sum_{k=1}^{\delta-1} \omega_j^k + \omega_j^{-k} \;=\; 1 + 2 \sum_{k=1}^{\delta-1} \cos\left(\frac{2\pi(j-1)k}{d}\right), \quad (3.19)$$

for all $j = 1, \ldots, d$. In particular, $\nu_1 = 2\delta - 1$. Set $\Lambda = \mathrm{diag}\{\nu_1, \ldots, \nu_d\}$ and let F denote the unitary $d \times d$ discrete Fourier matrix with entries

$$F_{j,k} := \frac{1}{\sqrt{d}} \mathrm{e}^{2\pi \mathrm{i} \frac{(j-1)(k-1)}{d}},$$

then $U = F \Lambda F^*$.

We consider that $\widetilde{X}_0$ and U are similar; indeed $\widetilde{X}_0 = \widetilde{D}_0 U \widetilde{D}_0^*$, where $\widetilde{D}_0 = \mathrm{diag}\{(\widetilde{x}_0)_1, \ldots, (\widetilde{x}_0)_d\}$. Since $|(\widetilde{\mathbf{x}}_0)_j| = 1$ for each j, we have that $\widetilde{D}_0$ is unitary. Thus the eigenvalues of $\widetilde{X}_0$ are given by (3.19), and its eigenvectors are simply the discrete Fourier vectors modulated by the entries of $\widetilde{\mathbf{x}}_0$. We now have the following lemma.

Lemma 1. *Let $\widetilde{X}_0$ be defined as in* (3.7)*. Then*

$$\widetilde{X}_0 = \widetilde{D}_0 F \Lambda F^* \widetilde{D}_0^*$$

where F is the unitary $d \times d$ discrete Fourier transform matrix, $\widetilde{D}_0$ is the $d \times d$ diagonal matrix $\mathrm{diag}\{(\widetilde{x}_0)_1, \ldots, (\widetilde{x}_0)_d\}$, and Λ is the $d \times d$ diagonal matrix $\mathrm{diag}\{\nu_1, \ldots, \nu_d\}$ where

$$\nu_j := 1 + 2 \sum_{k=1}^{\delta-1} \cos\left(\frac{2\pi(j-1)k}{d}\right)$$

for $j = 1, \ldots, d$.

We next estimate the principal eigenvalue gap of $\widetilde{X}_0$. This information will be crucial to our understanding of the stability and robustness of Algorithm 1.

3.3.1 The Spectral Gap of $\widetilde{X}_0$

Set $\theta_j = \frac{2\pi j}{d}$ and begin by observing that, for any $\theta \in \mathbb{R}$,

$$\sum_{k=1}^{\delta-1} \cos(\theta k) \;=\; \frac{1}{2}\left(\frac{\sin(\theta(\delta-1/2))}{\sin(\theta/2)} - 1\right).$$

Accordingly, defining $l_\delta : \mathbb{R} \to \mathbb{R}$ by $l_\delta(\theta) := 1 + 2\sum_{k=1}^{\delta-1}\cos(\theta k)$ we have that

$$\nu_{j+1} \;=\; l_\delta(\theta_j) \;=\; \frac{\sin(\theta_j(\delta-1/2))}{\sin(\theta_j/2)}. \tag{3.20}$$

Thus, the eigenvalues of $\widetilde{X}_0$ are sampled from the $(\delta-1)^{\text{st}}$ Dirichlet kernel. Of course, $\nu_1 = 2\delta - 1$ is the largest of these in magnitude, so the eigenvalue gap $\min_j \nu_1 - |\nu_j|$ is at most equal to

$$\begin{aligned}
\nu_1 - \nu_2 &= (2\delta-1) - \frac{\sin(\pi/d(2\delta-1))}{\sin(\pi/d)} \\
&\leq (2\delta-1) - \frac{\pi/d(2\delta-1) - \frac{1}{6}(\pi/d(2\delta-1))^3}{\pi/d} \\
&= \frac{1}{6}\left(\frac{\pi}{d}\right)^2 (2\delta-1)^3 \leq \frac{4\pi^2}{3}\frac{\delta^3}{d^2}.
\end{aligned}$$

Thus, $\nu_1 - |\nu_2| \lesssim \frac{\delta^3}{d^2}$. However, a lower bound on the spectral gap is more useful. The following lemma establishes that the spectral gap is indeed $\sim \frac{\delta^3}{d^2}$ for most reasonable choices of $\delta < d$.

Lemma 2. *Let $\nu_1 = 2\delta - 1, \nu_2, \ldots, \nu_d$ be the eigenvalues of $\widetilde{X}_0$. Then*

$$\min_{j\in\{2,3,\ldots,d\}} (\nu_1 - |\nu_j|) \geq \frac{\pi^2}{3}\frac{\delta^3}{d^2}$$

whenever $d \geq 4\delta$ and $\delta \geq 3$.

Proof. Let $\theta_j = \frac{2\pi j}{d}$. We find the lower bound by considering that $\theta_j \in [\pi/d, 2\pi - \pi/d]$ for every $j > 0$, so

$$\nu_1 - \max|\nu_j| \geq \nu_1 - \max_{\theta\in[\pi/d,2\pi-\pi/d]} |l_\delta(\theta)| = (2\delta-1) - \max_{\theta\in[\pi/d,\pi]} |l_\delta(\theta)|,$$

where we have used our eigenvalue formula from (3.20), and the symmetry of l_δ about $\theta = \pi$.

We now show that l_δ is decreasing towards its first zero at $\theta = \frac{2\pi}{2\delta-1}$ by considering the derivative

$$l'_\delta(\theta) = \frac{(\delta - 1/2)\cos((\delta-1/2)\theta)\sin(\theta/2) - 1/2\sin((\delta-1/2)\theta)\cos(\theta/2)}{\sin(\theta/2)^2},$$

which is non-positive if and only if

$$(2\delta-1)\sin(\theta/2)\cos((\delta-1/2)\theta) \leq \sin((\delta-1/2)\theta)\cos(\theta/2).$$

Since $\tan(\cdot)$ is convex on $[0,\, \pi/2)$, this last inequality will hold for $\theta \in [0,\, \frac{\pi}{2\delta-1})$. For $\theta \in [\frac{\pi}{2\delta-1}, \frac{2\pi}{2\delta-1})$, $\cos((\delta-1/2)\theta) \leq 0$ while the remainder of the terms are non-negative, so the inequality also holds. Therefore,

$$\nu_1 - \max_{j>1}|\nu_j| \geq (2\delta-1) - \max\left\{\nu_2, \max_{\theta\in[\frac{2\pi}{2\delta-1},\pi]}|l_\delta(\theta)|\right\},$$

which permits us to bound $(2\delta-1)-\nu_2$ and $(2\delta-1) - \max\limits_{\theta\in[\frac{2\pi}{2\delta-1},\pi]}|l_\delta(\theta)|$ separately.

For $\max\limits_{\theta\in[\frac{2\pi}{2\delta-1},\pi]}|l_\delta(\theta)|$, we simply observe that

$$\max_{\theta\in[\frac{2\pi}{2\delta-1},\pi]}|l_\delta(\theta)| \leq \max_{\theta\in[\frac{2\pi}{2\delta-1},\pi]}\frac{1}{\sin(\theta/2)} = \left(\sin\left(\frac{\pi}{2\delta-1}\right)\right)^{-1} \leq \frac{2\delta-1}{2},$$

where the last line uses that $\frac{\pi}{2\delta-1} \leq \pi/2$ (since $\delta \geq 3$). This yields $\nu_1 - \max\limits_{\theta\in[\frac{2\pi}{2\delta-1},\pi]}|l_\delta(\theta)| \geq \frac{1}{2}(2\delta-1)$.

As for ν_2, we have $\theta_1\cdot(\delta-1) \leq \pi/2$ (since $4(\delta-1)\leq d$). Thus, $\cos(\cdot)$ will be concave on $[0,\, \theta_1(\delta-1)]$. Considering (3.19), this will give $\sum_{k=1}^{\delta-1}\cos(k\theta_1) \leq (\delta-1)\cos\left(\theta_1\frac{\delta}{2}\right)$, so

$$\begin{aligned}
\nu_1 - \nu_2 &\geq 2(\delta-1)\left(1-\cos\left(\pi\tfrac{\delta}{d}\right)\right) \\
&\geq 2(\delta-1)\left(\frac{(\pi\frac{\delta}{d})^2}{4}\right) \\
&\geq \frac{\pi^2}{3}\cdot\frac{\delta^3}{d^2}.
\end{aligned}$$

The stated result follows, since $d \geq 4\delta$ gives

$$\nu_1 - \max_{\theta \in [\frac{2\pi}{2\delta-1}, \pi]} |l_\delta(\theta)| \geq \frac{1}{2}(2\delta - 1) \geq \frac{1}{2}\delta \geq \frac{1}{3}\left(\frac{\pi\delta}{d}\right)^2 \delta = \frac{\pi^2}{3} \cdot \frac{\delta^3}{d^2}.$$

□

We are now sufficiently well informed about $\widetilde{X}_0$ to consider perturbation results for its leading eigenvector.

3.4 Perturbation Theory for $\tilde{X}_0$

In this section we will use spectral graph theoretic techniques to obtain a bound on the error associated with recovering phase information using our method. In particular, we will adapt the proof of Theorem 6.3 from [2] to develop a bound for $\min_{\theta \in [0,2\pi]} \|\widetilde{\mathbf{x}}_0 - \mathbb{e}^{\mathbb{i}\theta}\widetilde{\mathbf{x}}\|_2$. This approach involves considering both $\widetilde{X}$ from Algorithm 1 and $\widetilde{X}_0$ from (3.7) in the context of spectral graph theory, so we begin by defining essential terms. The idea is to consider a graph whose vertices correspond to the entries of $\widetilde{\mathbf{x}}_0$ from (3.7), and whose edges carry the relative phase data.[8]

We begin with an undirected graph $G = (V, E)$ with vertex set $V = \{1, 2, \ldots, d\}$ and weight mapping $w : V \times V \to \mathbb{R}^+$, where $w_{ij} = w_{ji}$ and $w_{ij} = 0$ iff $\{i, j\} \notin E$. The **degree** of a vertex i is

$$\deg(i) := \sum_{j \text{ s.t. } (i,j) \in E} w_{ij},$$

and we define the **degree matrix** and **weighted adjacency matrix** of G by

$$D := \operatorname{diag}(\deg(i)) \text{ and } W_{ij} := w_{ij},$$

[8]The interested reader is also referred to Appendix 3.8 where more standard perturbation theoretic techniques are utilized in order to obtain a weaker bound on the error associated with recovering phase information via the proposed approach.

respectively. The **volume** of G is

$$\mathrm{vol}(G) := \sum_{i \in V} \deg(i).$$

Finally, the **Laplacian** of G is the $d \times d$ real symmetric matrix

$$L := I - D^{-1/2} W D^{-1/2} = D^{-1/2}(D - W)D^{-1/2}, \tag{3.21}$$

where $I \in \{0,1\}^{d \times d}$ is the identity matrix.

When G is connected, Lemma 1.7 of [22] shows that the nullspace of $(D - W)$ is $\mathrm{span}(\mathbb{1})$, and the nullspace of L is $\mathrm{span}(D^{1/2}\mathbb{1})$. Observing that $D - W$ is diagonally semi-dominant, it follows from Gershgorin's disc theorem that $(D - W)$ and L are both positive semidefinite. Alternatively, one may also note that

$$\mathbf{v}^*(D - W)\mathbf{v} = \sum_{i \in V} \left(v_i^2 \deg(i) - \sum_{j \in V} v_i v_j w_{ij} \right) = \frac{1}{2} \sum_{i,j \in V} w_{ij}(v_i - v_j)^2 \geq 0$$

holds for all $\mathbf{v} \in \mathbb{R}^d$. Thus, we may order the eigenvalues of L in increasing order so that $0 = \lambda_1' < \lambda_2' \leq \cdots \leq \lambda_n'$. We then define the **spectral gap** of G to be $\tau = \lambda_2'$.

Herein, though we will state the main theorem of this section more generally, we will only be interested in the case where the graph $G = (V, E)$ is the simple unweighted graph whose adjacency matrix is $U = T_\delta(\mathbb{1}\mathbb{1}^*)$ as in (3.18). In this case we will have $W = U$ and $D = (2\delta - 1)I$. We also immediately obtain the following corollary of Lemmas 1 and 2.

Corollary 1. *Let G be the simple unweighted graph whose adjacency matrix is U from* (3.18)*. Let L be the Laplacian of G. Then, there exists a bijection $\sigma : [d] \to [d]$ such that*

$$\lambda'_{\sigma(j)} = 1 - \frac{1 + 2\sum_{k=1}^{\delta - 1} \cos\left(\frac{2\pi(j-1)k}{d}\right)}{2\delta - 1}$$

for $j = 1, \ldots, d$. In particular, if $d \geq 4\delta$ and $\delta \geq 3$ then

$$\tau = \lambda_2' > \frac{\pi^2}{6} \frac{\delta^2}{d^2}.$$

Using this graph G as a scaffold we can now represent our computed relative phase matrix $\widetilde{X}$ from Algorithm 1 by noting that for some (Hermitian) perturbations η_{ij} we will have

$$\widetilde{X}_{ij} = \frac{(x_0)_i(x_0)_j^* + \eta_{ij}}{|(x_0)_i(x_0)_j^* + \eta_{ij}|} \cdot w_{ij} = \frac{(x_0)_i(x_0)_j^* + \eta_{ij}}{|(x_0)_i(x_0)_j^* + \eta_{ij}|} \cdot \chi_{E(i,j)}. \tag{3.22}$$

Using this same notation we may also represent our original phase matrix $\widetilde{X}_0$ via G by noting that

$$(\widetilde{X}_0)_{ij} = \frac{(x_0)_i(x_0)_j^*}{|(x_0)_i(x_0)_j^*|} \cdot w_{ij} = \operatorname{sgn}\left((x_0)_i(x_0)_j^*\right) \cdot \chi_{E(i,j)}. \tag{3.23}$$

We may now define the **connection Laplacian** of the graph G associated with the Hermitian and entrywise normalized data given by $\widetilde{X}$ to be the matrix

$$L_1 = I - D^{-1/2}(\widetilde{X} \circ W)D^{-1/2}, \tag{3.24}$$

where $\circ$ denotes entrywise (Hadamard) multiplication. Following [8], given $\widetilde{X}$ and a vector $\mathbf{y} \in \mathbb{C}^d$, we define the **frustration of y with respect to** $\widetilde{X}$ by

$$\eta_{\widetilde{X}}(\mathbf{y}) := \frac{1}{2}\frac{\sum_{(i,j)\in E} w_{ij}|y_i - \widetilde{X}_{ij}y_j|^2}{\sum_{i\in V}\deg(i)|y_i|^2} = \frac{\mathbf{y}^*(D - (\widetilde{X} \circ W))\mathbf{y}}{\mathbf{y}^* D\mathbf{y}}. \tag{3.25}$$

We may consider $\eta_{\widetilde{X}}(\mathbf{y})$ to measure how well $\mathbf{y}$ (viewed as a map from V to $\mathbb{C}$) conforms to the computed relative phase differences $\widetilde{X}$ across the graph G.

In addition, we adapt a result from [8]:

Lemma 3 (Cheeger inequality for the connection Laplacian)**.** *Suppose that $G = (V = [d], E)$ is a connected graph with degree matrix $D \in [0,\infty)^{d\times d}$, weighted adjacency matrix $W \in [0,\infty)^{d\times d}$, and spectral gap $\tau > 0$, and that $\widetilde{X} \in \mathbb{C}^{d\times d}$ is Hermitian and entrywise normalized. Let $\mathbf{u} \in \mathbb{C}^d$ be an eigenvector of L_1 from*

(3.24) *corresponding to its smallest eigenvalue. Then,* $\mathbf{w} = \operatorname{sgn}(\mathbf{u}) = \operatorname{sgn}\left(D^{-1/2}\mathbf{u}\right)$ *satisfies*

$$\eta_{\widetilde{X}}(\mathbf{w}) \leq \frac{C'}{\tau} \cdot \min_{\mathbf{y}\in\mathbb{C}^d} \eta_{\widetilde{X}}(\operatorname{sgn}(\mathbf{y})),$$

where $C' \in \mathbb{R}^+$ *is a universal constant.*

Proof. One can see that

$$\begin{aligned}\inf_{\mathbf{v}\in\mathbb{C}^d\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^* L_1 \mathbf{v}}{\mathbf{v}^*\mathbf{v}} &= \inf_{\mathbf{y}\in\mathbb{C}^d\setminus\{\mathbf{0}\}} \frac{(D^{1/2}\mathbf{y})^* L_1 (D^{1/2}\mathbf{y})}{(D^{1/2}\mathbf{y})^*(D^{1/2}\mathbf{y})} = \inf_{\mathbf{y}\in\mathbb{C}^d\setminus\{\mathbf{0}\}} \frac{\mathbf{y}^*(D - (\widetilde{X}\circ W))\mathbf{y}}{\mathbf{y}^* D\mathbf{y}} \\ &= \inf_{y\in\mathbb{C}^d\setminus\{\mathbf{0}\}} \eta_{\widetilde{X}}(\mathbf{y}) \leq \min_{\mathbf{y}\in\mathbb{C}^d} \eta_{\widetilde{X}}(\operatorname{sgn}(\mathbf{y})).\end{aligned}$$

From here, Lemma 3.6 in [8] gives

$$\eta_{\widetilde{X}}(\mathbf{w}) \leq \frac{44}{\tau}\eta_{\widetilde{X}}\left(D^{-1/2}\mathbf{u}\right) = \frac{44}{\tau} \cdot \inf_{\mathbf{v}\in\mathbb{C}^d\setminus\{\mathbf{0}\}} \frac{\mathbf{v}^* L_1 \mathbf{v}}{\mathbf{v}^*\mathbf{v}} \leq \frac{44}{\tau} \cdot \min_{\mathbf{y}\in\mathbb{C}^d} \eta_{\widetilde{X}}(\operatorname{sgn}(\mathbf{y})).$$

□

We now state the main result of this section:

Theorem 4. *Suppose that* $G = (V = [d], E)$ *is an undirected, connected, and unweighted graph (so that* $W_{ij} = \chi_{E(i,j)}$*) with spectral gap* $\tau > 0$*. Let* $\mathbf{u} \in \mathbb{C}^d$ *be an eigenvector of* L_1 *from* (3.24) *corresponding to its smallest eigenvalue, and let*

$$\widetilde{\mathbf{x}} = \operatorname{sgn}(\mathbf{u}) \text{ and } \widetilde{\mathbf{x}}_0 = \operatorname{sgn}(\mathbf{x}_0).$$

Then for some universal constant $C \in \mathbb{R}^+$*,*

$$\min_{\theta\in[0,2\pi]} ||\widetilde{\mathbf{x}} - \mathrm{e}^{\mathrm{i}\theta}\widetilde{\mathbf{x}}_0||_2 \leq C\frac{\|\widetilde{X} - \widetilde{X}_0\|_F}{\tau \cdot \sqrt{\min_{i\in V}(\deg(i))}},$$

where $\widetilde{X}$ *and* $\widetilde{X}_0$ *are defined as per* (3.22) *and* (3.23)*, respectively.*

The proof follows by combining the two following lemmas, which share the hypotheses of the theorem. Additionally, we introduce the notation $\mathbf{g} \in \mathbb{C}^d$ and $\Lambda \in \mathbb{C}^{d\times d}$, where

$$g_i = (\widetilde{\mathbf{x}}_0)_i^* \widetilde{\mathbf{x}}_i \quad \text{and} \quad \Lambda_{ij} = (\widetilde{X}_0)_{ij}^* \widetilde{X}_{ij},$$

and observe that $|g_i| = |\Lambda_{ij}| = 1$ for each $(i,j) \in E$.

Lemma 4. *Under the hypotheses of Theorem 4, there exists an angle $\theta \in [0, 2\pi]$ such that*

$$\tau \sum_{i \in V} \deg(i) |g_i - \mathrm{e}^{\mathrm{i}\theta}|^2 \leq 2 \sum_{(i,j) \in E} |g_i - g_j|^2.$$

Lemma 5. *Under the hypotheses of Theorem 4, there exists an absolute constant C such that*

$$2 \sum_{(i,j) \in E} |g_i - g_j|^2 \leq \frac{C}{\tau} \|\widetilde{X} - \widetilde{X}_0\|_F^2.$$

From these lemmas, the theorem follows immediately by observing $\sum_{i \in V} |g_i - \mathrm{e}^{\mathrm{i}\theta}|^2 = \|\widetilde{\mathbf{x}} - \mathrm{e}^{\mathrm{i}\theta} \widetilde{\mathbf{x}}_0\|_2^2$.

Proof of Lemma 4. We set $\alpha = \dfrac{\sum_{i \in V} \deg(i) g_i}{\mathrm{vol}(G)}$ and $w_i = g_i - \alpha$. Then

$$\mathbb{1}^* D \mathbf{w} = \sum_{i \in V} \deg(i)(g_i - \alpha) = 0,$$

so $D^{1/2}\mathbf{w}$ is orthogonal to $D^{1/2}\mathbb{1}$. Noting that the null space of L is spanned by $D^{1/2}\mathbb{1}$ when $\tau > 0$, and recalling that $L \succeq 0$, we have

$$\frac{(D^{1/2}\mathbf{w})^* L (D^{1/2}\mathbf{w})}{\mathbf{w}^* D \mathbf{w}} \geq \min_{\mathbf{y}^* D^{1/2} \mathbb{1} = 0} \frac{\mathbf{y}^* L \mathbf{y}}{\mathbf{y}^* \mathbf{y}} = \tau.$$

Therefore,

$$\begin{aligned} \tau \mathbf{w}^* D \mathbf{w} &\leq \mathbf{w}^*(D - W)\mathbf{w} &&= \mathbf{g}^*(D - W)\mathbf{g} \\ &= \sum_{i \in V} \deg(i) |g_i|^2 - \sum_{i \in V} g_i^* \sum_{(i,j) \in E} g_j &&= \sum_{(i,j) \in E} (1 - g_i^* g_j) \\ &= \frac{1}{2} \sum_{(i,j) \in E} |g_i - g_j|^2. \end{aligned}$$

We note that $\tau \mathbf{w}^* D \mathbf{w} = \tau \sum_{i \in V} \deg(i) |g_i - \alpha|^2$, while we seek a bound on $\sum_{i \in V} \deg(i) |g_i - \mathrm{e}^{\mathrm{i}\theta}|^2$. To that end, we use the fact that $|g_i| = |\operatorname{sgn}(\alpha)| = 1$ to obtain

$$|g_i - \operatorname{sgn}(\alpha)| \leq |g_i - \alpha| + |\alpha - \operatorname{sgn}(\alpha)| \leq 2|g_i - \alpha|.$$

Setting $\theta := \arg \alpha$, we have the stated result. □

Proof of Lemma 5. Observe that for any two real numbers $a, b \in \mathbb{R}$, we have $\frac{1}{2}a^2 - b^2 \leq (a-b)^2$. Thus, by the reverse triangle inequality we have

$$\begin{aligned}\sum_{(i,j)\in E}\left(\frac{1}{2}|g_i - g_j|^2 - |\Lambda_{ij}-1|^2\right) &\leq \sum_{(i,j)\in E}\left(|g_i - g_j| - |\Lambda_{ij}-1|\right)^2 \\ &\leq \sum_{(i,j)\in E}|g_i - \Lambda_{ij}g_j|^2 \\ &= \sum_{(i,j)\in E}|\widetilde{\mathbf{x}}_i - \widetilde{X}_{ij}\widetilde{\mathbf{x}}_j|^2 \\ &= 2\,\mathrm{vol}(G)\cdot\eta_{\widetilde{X}}(\widetilde{\mathbf{x}}),\end{aligned} \tag{3.26}$$

as the denominator of (3.25) is $2\,\mathrm{vol}(G)$ whenever the entries of $\mathbf{y}$ all have unit modulus.

Lemma 3 now tells us that

$$\begin{aligned}\sum_{(i,j)\in E}\left(\frac{1}{2}|g_i - g_j|^2 - |\Lambda_{ij}-1|^2\right) &\leq \frac{2C'\,\mathrm{vol}(G)}{\tau}\min_{\mathbf{y}\in\mathbb{C}^d}\eta_{\widetilde{X}}(\mathrm{sgn}(\mathbf{y})) \\ &\leq \frac{2C'\,\mathrm{vol}(G)}{\tau}\eta_{\widetilde{X}}(\widetilde{\mathbf{x}}_0).\end{aligned} \tag{3.27}$$

Moreover,

$$\begin{aligned}\eta_{\widetilde{X}}(\widetilde{\mathbf{x}}_0) &= \frac{\displaystyle\sum_{(i,j)\in E}|(\widetilde{\mathbf{x}}_0)_i - \widetilde{X}_{ij}(\widetilde{\mathbf{x}}_0)_j|^2}{\displaystyle 2\sum_{i\in V}\deg(i)|(\widetilde{\mathbf{x}}_0)_i|^2} \\ &= \frac{\displaystyle\sum_{(i,j)\in E}|(\widetilde{\mathbf{x}}_0)_i(\widetilde{\mathbf{x}}_0)_j^* - \widetilde{X}_{ij}|^2}{2\,\mathrm{vol}(G)} \\ &= \frac{\|\widetilde{X}_0 - \widetilde{X}\|_F^2}{2\,\mathrm{vol}(G)},\end{aligned}$$

so that $\sum_{(i,j)\in E}\frac{1}{2}|g_i - g_j|^2 \leq \frac{C'}{\tau}\|X_0 - X\|_F^2 + \sum_{(i,j)\in E}|\Lambda_{ij}-1|^2$. Considering also that

$$\sum_{(i,j)\in E}|\Lambda_{ij}-1|^2 = \sum_{(i,j)\in E}\left|\widetilde{X}_{ij} - (\widetilde{X}_0)_{ij}\right|^2 = \left\|\widetilde{X} - \widetilde{X}_0\right\|_F^2 \tag{3.28}$$

and $\tau \leq 1$, this completes the proof. □

We may now use Theorem 4 to produce a perturbation bound for our banded matrix of phase differences $\widetilde{X}_0$.

Corollary 2. *Let $\widetilde{X}_0$ be the matrix in* (3.7)*, $\widetilde{\mathbf{x}}_0$ be the vector of true phases* (3.7)*, and $\widetilde{X}$ be as in line 3 of Algorithm 1 with $\widetilde{\mathbf{x}} = \operatorname{sgn}(\mathbf{u})$ where $\mathbf{u}$ is the top eigenvector of $\widetilde{X}$. Suppose that $\|\widetilde{X}_0 - \widetilde{X}\|_F \leq \eta \|\widetilde{X}_0\|_F$ for some $\eta > 0$. Then, there exists an absolute constant $C' \in \mathbb{R}^+$ such that*

$$\min_{\theta \in [0,2\pi]} \|\widetilde{\mathbf{x}}_0 - \mathbb{e}^{\mathbb{i}\theta}\widetilde{\mathbf{x}}\|_2 \leq C' \frac{\eta d^{\frac{5}{2}}}{\delta^2}.$$

Proof. We apply Theorem 4 with the unweighted and undirected graph $G = (V, E)$, where $V = [d]$ and $E = \{(i,j) : |i-j| \bmod d < \delta\}$. Observe that G is also connected and $(2\delta - 1)$-regular so that $\min_{i \in V}(\deg(i)) = 2\delta - 1$. The spectral gap of G is $\tau > \frac{\pi^2}{6}\delta^2/d^2 > 0$ by Corollary 1. We know that $\|\widetilde{X}_0\|_F = \sqrt{d(2\delta - 1)}$, so that $\|\widetilde{X}_0 - \widetilde{X}\|_F \leq \eta\sqrt{d(2\delta - 1)}$. Finally, if $\mathbf{u}$ is the top eigenvector of $\widetilde{X}$ then it will also be an eigenvector of L_1 corresponding to its smallest eigenvalue since, here, $L_1 = I - \frac{1}{2\delta - 1}\widetilde{X}$.

Combining these observations we have

$$\min_{\theta \in [0,2\pi]} \|\widetilde{\mathbf{x}}_0 - \mathbb{e}^{\mathbb{i}\theta}\widetilde{\mathbf{x}}\|_2 \leq C \frac{\eta(d(2\delta - 1))^{1/2}}{\pi^2/6 \cdot \delta^2/d^2 \cdot (2\delta - 1)^{1/2}} = C' \frac{\eta d^{5/2}}{\delta^2}.$$

□

We are now properly equipped to analyze the robustness of Algorithm 1 to noise.

3.5 Recovery Guarantees for the Proposed Method

Herein we will assume Algorithm 1 is provided with measurements $\mathbf{y}$ of the form (3.8) such that the linear operator (3.5) is invertible on $T_\delta(\mathbb{C}^{d \times d})$ with condition number $\kappa > 0$. Unless otherwise stated, we follow the notation of §3.1.1- §3.1.2; therefore, our assumptions imply that $\|X - X_0\|_F \leq \kappa \|\mathbf{n}\|_2$.

We now aim to bound the Frobenius norm of the perturbation error $(\tilde{X} - \tilde{X}_0)$ present in the matrix $\tilde{X}$ formed in line 2 of Algorithm 1. Toward this end we define the set of ρ-small indexes of $\mathbf{x}_0$ to be

$$S_\rho := \left\{ j \;\middle|\; |\,(x_0)_j\,| < \left(\frac{\kappa\|\mathbf{n}\|_2}{\rho}\right)^{\frac{1}{4}} \right\} \tag{3.29}$$

where $\rho \in \mathbb{R}^+$ is a free parameter. With the definition of S_ρ in hand we can bound the perturbation error $(\tilde{X} - \tilde{X}_0)$ using the next lemma.

Lemma 6. *Let $\tilde{X}$ be the matrix computed in line 2 of Algorithm 1. We have that*

$$\|\tilde{X} - \tilde{X}_0\|_{\mathrm{F}} \leq C\sqrt{\frac{\rho\frac{\kappa}{\delta}\|\mathbf{n}\|_2 + |S_\rho|}{d}} \cdot \|\tilde{X}_0\|_{\mathrm{F}}$$

holds for all $\rho \in \mathbb{R}^+$, where C is an absolute constant.

Proof. Set $N_{jk} = X_{jk} - (X_0)_{jk}$ and consider that, for any $j, k \in S_\rho^c$,

$$\begin{aligned} |(\widetilde{X}_0)_{jk} - \widetilde{X}_{jk}| &= \left|(\widetilde{X}_0)_{jk} - \operatorname{sgn}\left(\tfrac{X_{jk}}{|(X_0)_{jk}|}\right)\right| \leq \left|(\widetilde{X}_0)_{jk} - \tfrac{X_{jk}}{|(X_0)_{jk}|}\right| + \left|\tfrac{X_{jk}}{|(X_0)_{jk}|} - \operatorname{sgn}\left(\tfrac{X_{jk}}{|(X_0)_{jk}|}\right)\right| \\ &\leq 2\left|(\widetilde{X}_0)_{jk} - \tfrac{X_{jk}}{|(X_0)_{jk}|}\right| = 2\tfrac{|N_{jk}|}{|(X_0)_{jk}|} \leq 2\rho^{\frac{1}{2}}\tfrac{|N_{jk}|}{(\kappa\|\mathbf{n}\|_2)^{\frac{1}{2}}}. \end{aligned}$$

Thus, there exists an absolute constant $C' \in \mathbb{R}^+$ such that

$$\begin{aligned} \|\tilde{X} - \tilde{X}_0\|_{\mathrm{F}}^2 &\leq \sum_{j,k\in S_\rho^c} 4\rho\frac{|N_{jk}|^2}{\kappa\|\mathbf{n}\|_2} + \sum_{j\in S_\rho,\ \text{or}\ k\in S_\rho} |(\tilde{X}_0)_{jk} - \tilde{X}_{jk}|^2 \\ &\leq 4\rho\frac{\|N\|_{\mathrm{F}}^2}{\kappa\|\mathbf{n}\|_2} + \sum_{j\in S_\rho} 4\cdot(4\delta - 3) = 4\rho\frac{\|N\|_{\mathrm{F}}^2}{\kappa\|\mathbf{n}\|_2} + 4\cdot(4\delta-3)\,|S_\rho| \\ &\leq C'(\rho\kappa\|\mathbf{n}\|_2 + \delta\,|S_\rho|). \end{aligned}$$

The proof is completed by recalling that $\|\widetilde{X}_0\|_F = \sqrt{(2\delta-1)d}$. □

We are finally ready to prove a robustness result for Algorithm 1.

Theorem 5. *Suppose that $\tilde{X}$ and $\tilde{X}_0$ satisfy $\|\tilde{X} - \tilde{X}_0\|_{\mathrm{F}} \leq \eta\|\tilde{X}_0\|_F$ for some $\eta > 0$. Then, the estimate $\mathbf{x}$ produced by Algorithm 1 satisfies*

$$\min_{\theta\in[0,2\pi]} \left\|\mathbf{x}_0 - \mathrm{e}^{\mathbb{i}\theta}\mathbf{x}\right\|_2 \leq C\|\mathbf{x}_0\|_\infty\left(\frac{d^{5/2}}{\delta^2}\right)\eta + Cd^{\frac{1}{4}}\sqrt{\kappa\|\mathbf{n}\|_2},$$

where $C \in \mathbb{R}^+$ is an absolute universal constant. Alternatively, one can bound the error in terms of the size of the index set S_ρ from (3.29) *as*

$$\min_{\theta\in[0,2\pi]} \left\|\mathbf{x}_0 - \mathbb{e}^{\mathbb{i}\theta}\mathbf{x}\right\|_2 \leq C'\|\mathbf{x}_0\|_\infty \left(\frac{d}{\delta}\right)^2 \sqrt{\rho\frac{\kappa}{\delta}\|\mathbf{n}\|_2 + |S_\rho|} + C'd^{\frac{1}{4}}\sqrt{\kappa\|\mathbf{n}\|_2}, \quad (3.30)$$

for any desired $\rho \in \mathbb{R}^+$, where $C' \in \mathbb{R}^+$ is another absolute universal constant.

Proof. Let $\phi \in [0, 2\pi)$ be arbitrary; then $\mathbb{e}^{\mathbb{i}\phi}\mathbf{x} = |\mathbf{x}| \circ \mathbb{e}^{\mathbb{i}\phi}\tilde{\mathbf{x}}$ and $\mathbf{x}_0 = |\mathbf{x}_0| \circ \tilde{\mathbf{x}}_0$, where $\circ$ denotes the entrywise (Hadamard) product.

We see that

$$\begin{aligned}\min_{\phi\in[0,2\pi]} \|\mathbf{x}_0 - \mathbb{e}^{\mathbb{i}\phi}\mathbf{x}\|_2 &= \min_{\phi\in[0,2\pi]} \left\||\mathbf{x}_0| \circ \tilde{\mathbf{x}}_0 - |\mathbf{x}| \circ \mathbb{e}^{\mathbb{i}\phi}\tilde{\mathbf{x}}\right\|_2 \\ &\leq \min_{\phi\in[0,2\pi]} \left\||\mathbf{x}_0| \circ \tilde{\mathbf{x}}_0 - |\mathbf{x}_0| \circ \mathbb{e}^{\mathbb{i}\phi}\tilde{\mathbf{x}}\right\|_2 + \left\||\mathbf{x}_0| \circ \mathbb{e}^{\mathbb{i}\phi}\tilde{\mathbf{x}} - |\mathbf{x}| \circ \mathbb{e}^{\mathbb{i}\phi}\tilde{\mathbf{x}}\right\|_2\end{aligned}$$

where the second term is now independent of ϕ. As a result we have that

$$\min_{\phi\in[0,2\pi]} \|\mathbf{x}_0 - \mathbb{e}^{\mathbb{i}\phi}\mathbf{x}\|_2 \leq \|\mathbf{x}_0\|_\infty \left(\min_{\phi\in[0,2\pi]} \|\tilde{\mathbf{x}}_0 - \mathbb{e}^{\mathbb{i}\phi}\tilde{\mathbf{x}}\|_2\right) + C''\sqrt{\kappa\sqrt{d}\cdot\|\mathbf{n}\|_2}$$

for some absolute constant $C'' \in \mathbb{R}^+$. Here the bound on the second term follows from Lemma 3 of [46] and the Cauchy-Schwarz inequality. The first inequality of the theorem now results from an application of Corollary 2 to the first term. The second inequality then follows from Lemma 6. □

Looking at the second inequality (3.30) in Theorem 5 we can see that the error bound there will be vacuous in most settings unless $S_\rho = \emptyset$. Recalling (3.29), one can see that S_ρ will be empty as soon as $\rho = \kappa\delta\|\mathbf{n}\|_2 / |(x_0)_{\min}|^4$, where $(x_0)_{\min}$ is the smallest magnitude of any entry in $\mathbf{x}_0$. Utilizing this value of ρ in (3.30) leads to the following corollary of Theorem 5.

Corollary 3. *Let $(x_0)_{\min} := \min_j |(x_0)_j|$ be the smallest magnitude of any entry in $\mathbf{x}_0$. Then, the estimate $\mathbf{x}$ produced by Algorithm 1 satisfies*

$$\min_{\theta\in[0,2\pi]} \left\|\mathbf{x}_0 - \mathbb{e}^{\mathbb{i}\theta}\mathbf{x}\right\|_2 \leq C\left(\frac{\|\mathbf{x}_0\|_\infty}{(x_0)_{\min}^2}\right)\left(\frac{d}{\delta}\right)^2 \kappa\|\mathbf{n}\|_2 + Cd^{\frac{1}{4}}\sqrt{\kappa\|\mathbf{n}\|_2},$$

where $C \in \mathbb{R}^+$ is an absolute universal constant.

Corollary 3 yields a deterministic recovery result for any signal $\mathbf{x}_0$ which contains no zero entries. If desired, a randomized result can now be derived from Corollary 3 for arbitrary $\mathbf{x}_0$ by right multiplying the signal $\mathbf{x}_0$ with a random "flattening" matrix as done in [46]. Finally, we note that a trivial variant of Corollary 3 can also be combined with the discussion in §3.1.5 in order to generate recovery guarantees for the windowed Fourier measurements defined by (3.15). However, we will leave such variants and extensions to the interested reader.

3.6 Numerical Evaluation

We now present numerical simulations supporting the theoretical recovery guarantees in Section 3.5. In addition to illustrating the performance of our algorithm, we also compare it against other existing phase retrieval methods using *local* measurements. All the results presented here may be recreated using the open source *BlockPR* Matlab software package which is freely available at [47].

In Section 3.6.2 we utilize the sparse measurement masks of Example 2 (see §3.2 for details). For each choice of δ these measurements correspond to $2\delta - 1$ physical masks which are each shifted d times across the sample $\mathbf{x}_0$ by shifts of size 1. In Section 3.6.3, we then consider the Fourier measurement masks of Example 1 from §3.2. For each choice of δ these measurements correspond to $2\delta - 1$ Fourier measurements of *one* physical mask/illumination which is also shifted d times across the sample $\mathbf{x}_0$ by shifts of size 1. These Fourier measurements correspond particularly well to ptychographic measurements of a large sample in the small δ regime.

For completeness, we also present selected results comparing the proposed formulation against other well established phase retrieval algorithms (such as *Wirtinger Flow*) using *global* measurements such as coded diffraction patterns (CDPs) [20,

§1.5]. These measurements correspond to those one might obtain from the diffraction patterns of the sample $\mathbf{x}_0$ after it has been masked by several different global random windows. Unless otherwise stated, we use i.i.d. zero-mean complex Gaussian random test signals with measurement errors modeled using an additive Gaussian noise model. Applied measurement noise and reconstruction error are both reported in decibels (dB) in terms of signal to noise ratios (SNRs), with

$$\text{SNR (dB)} = 10\log_{10}\left(\frac{\sum_{j=1}^{D}|\langle \mathbf{a}_j, \mathbf{x}_0\rangle|^4}{D\sigma^2}\right),$$
$$\text{Error (dB)} = 10\log_{10}\left(\frac{\min_\theta \|\mathrm{e}^{\mathrm{i}\theta}\mathbf{x} - \mathbf{x}_0\|_2^2}{\|\mathbf{x}_0\|_2^2}\right),$$

where $\mathbf{a}_j, \mathbf{x}_0, \mathbf{x}, \sigma^2$ and D denote the measurement vectors, true signal, recovered signal, (Gaussian) noise variance and number of measurements respectively. All simulations were performed on a laptop computer running GNU/Linux (Ubuntu Linux 16.04 `x86_64`) with an Intel® Core™i3-3120M (2.5 GHz) processor, 4GB RAM and Matlab R2015b. Each data point in the timing and robustness plots was obtained as the average of 100 trials.

3.6.1 Numerical Improvements to Algorithm 1: Magnitude Estimation

Looking at the matrix X formed on line 1 of Algorithm 1 one can see that

$$X = X_0 + N',$$

where X_0 is the banded Hermitian matrix $T_\delta(\mathbf{x}_0\mathbf{x}_0^*)$ defined in (3.6), and N' contains arbitrary banded Hermitian noise. As stated and analyzed above, Algorithm 1 estimates the magnitude of each entry of $\mathbf{x}_0$ by observing that

$$X_{jj} = |(x_0)_j|^2 + N'_{jj}, \quad j \in [d].$$

Though this magnitude estimate suffices for our theoretical treatment above, it can be improved in practice by using slightly more general techniques.

Considering the component-wise magnitude of X, $|X| \in \mathbb{R}^{d\times d}$, one can see that its entries are

$$|X|_{jk} = \begin{cases} |(x_0)_j||(x_0)_k| + N''_{jk} & \text{if } |j-k| \bmod d < \delta \\ 0 & \text{otherwise} \end{cases},$$

where $N'' = |X| - |X_0|$ represents the changes in magnitude to the entries of $|X_0|$ due to noise. We may then let $D_j \in \mathbb{R}^{\delta\times\delta}$ denote the submatrix of $|X|$ given by

$$(D_j)_{kh} = |X|_{(j+k-1) \bmod d,\ (j+h-1) \bmod d},$$

for all $j \in [d]$; similarly we let N''_j denote the respective submatrices of N''. With this notation, it is clear that

$$D_j = |\mathbf{x}_0|^{(j)}(|\mathbf{x}_0|^{(j)})^* + N''_j,$$

where $|\mathbf{x}_0|^{(j)}_k = |\mathbf{x}_0|_{k+j-1}, k \in [\delta]$. This immediately suggests that we can estimate the magnitudes of the entries of $\mathbf{x}_0$ by calculating the top eigenvectors of these approximately rank one D_j matrices.

Indeed, if we do so for all of $D_1, \ldots, D_d \in \mathbb{R}^{\delta\times\delta}$, we will produce δ estimates of each $(x_0)_j$ entry's magnitude. A final estimate of each $|(x_0)_j|$ can then be computed by taking the average, median, etc. of the δ different estimates of $|(x_0)_j|$ provided by each of the leading eigenvectors of $D_{j-\delta+1}, \ldots, D_j$; in our experiments, we used the arithmetic mean. Of course, one need neither use all d possible D_j matrices, nor make them have size $\delta \times \delta$. More generally, to reduce computational complexity, one may instead use d/s matrices, $\tilde{D}_{j'} \in \mathbb{R}^{\gamma\times\gamma}$, of size $1 \le \gamma \le \delta$ and with shifts $s \le \gamma$ (dividing d), having entries

$$(\tilde{D}_{j'})_{k,h} = |X|_{(sj'+k-s) \bmod d,\ (sj'+h-s) \bmod d}.$$

Computing the leading eigenvectors of $\tilde{D}_{j'}$ for all $j' \in [d/s]$ will then produce (multiple) estimates of each magnitude $|(x_0)_j|$ which can then be combined as desired to produce our final magnitude estimates. As we shall see below, one can achieve better numerical robustness to noise using this technique than what can be achieved using the simpler magnitude estimation technique presented in line 4 of Algorithm 1.

3.6.2 Experiments with Measurements from Example 2 of §3.2

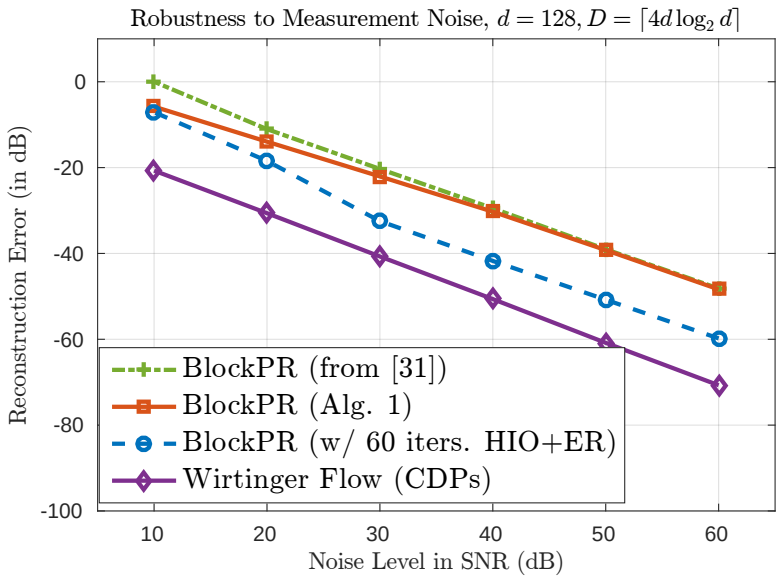
We begin by presenting results in Fig. 3.2a demonstrating the improved noise robustness of the proposed method over the formulation in [46]. Recall that [46] uses a greedy angular synchronization method instead of the eigenvector-based procedure analyzed in this paper. Fig. 3.2a plots the reconstruction error when recovering a $d = 128$ length complex Gaussian test signal using $D = \lceil 4d \log_2 d \rceil$ measurements at different added noise levels. As mentioned above, the local correlation measurements described in Example 2 of Section 3.2 are utilized in this plot and in all the ensuing experiments in this subsection unless otherwise indicated. Three variants of the proposed algorithm are plotted in Fig. 3.2a:

1. an implementation of Algorithm 1 (denoted by $\square$'s),

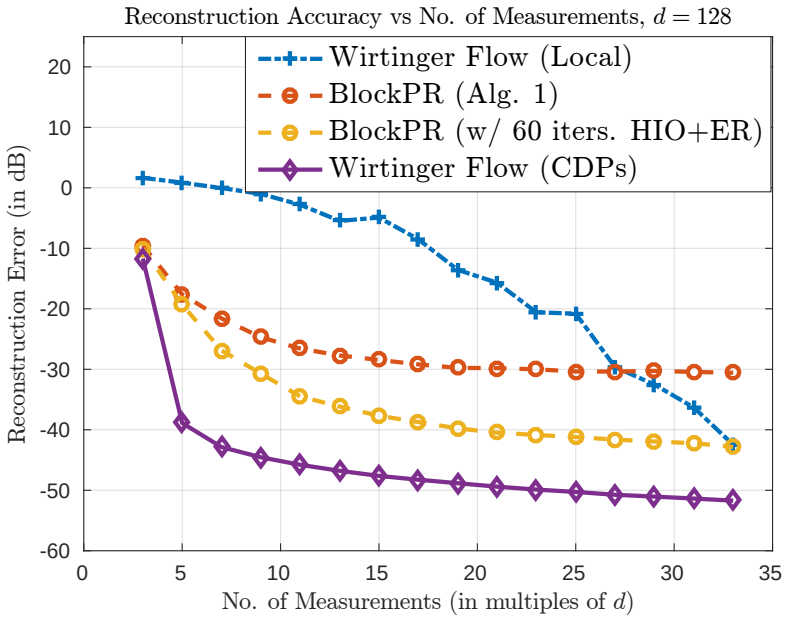
2. an implementation of Algorithm 1 post-processed using 60 iterations of the Hybrid Input–Output (HIO) and Error Reduction (ER) algorithms (implemented in two successive sets, with each set consisting of 20 iterations of HIO followed by 10 ER iterations; denoted by $\circ$'s), and

3. the algorithmic implementation from [46] (denoted by +'s).

We see that the eigenvector-based angular synchronization method proposed in



(a) Improved Robustness to Measurement Noise – Comparing Variants of the *BlockPR* algorithm

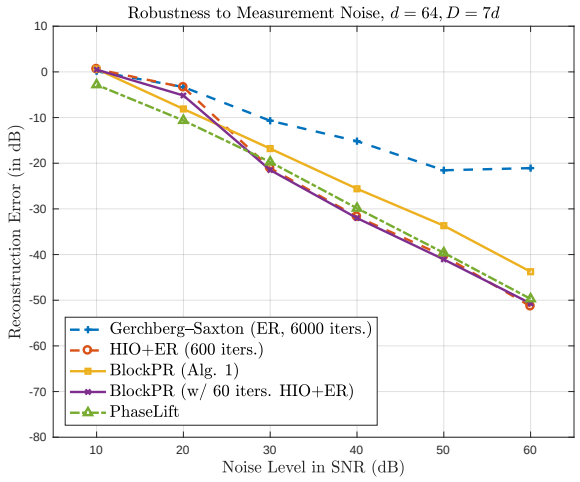


(b) Reconstruction Error vs. No. of Measurements; (Reconstruction at 40dB SNR)

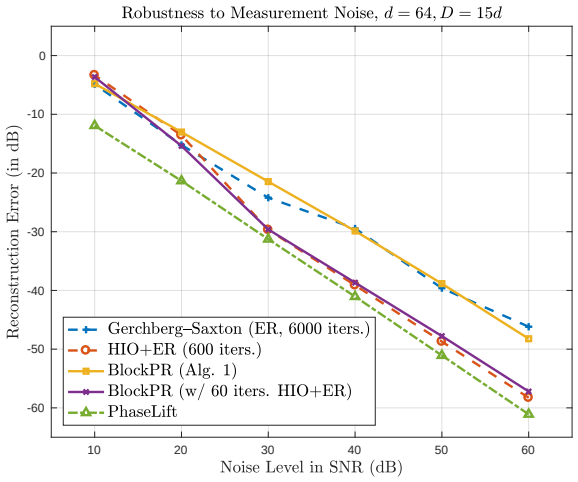
Figure 3.2: Robust Phase Retrieval – Local vs. Global Measurements

this paper provides more accurate reconstructions – especially at low SNRs – over the greedy angular synchronization of [46]. Moreover, post-processing using the HIO+ER algorithms as detailed above yields a significant improvement in reconstruction errors over the two other variants. For reference, we also include reconstruction errors with the *Wirtinger Flow* algorithm (denoted by $\diamondsuit$'s) when using (*global*) coded diffraction pattern (CDP) measurements. Specifically, we use $2\delta - 1$ (with $\delta = 2\log_2 d = 14$) octanary modulations/codes as described in [20, §1.5, (1.9)] to construct the CDP measurements. Clearly, using global measurements such as coded diffraction patterns provides superior noise tolerance; however, they are not applicable to the local measurement model considered here. Indeed, when the *Wirtinger Flow* algorithm is used with local measurements such as those described in this paper, the noise tolerance significantly deteriorates. Fig. 3.2b illustrates this phenomenon by plotting the reconstruction error in recovering a $d = 128$ length complex Gaussian test signal at 40 dB SNR when using different numbers of measurements, D. *Wirtinger flow*, for example, requires a large number of local measurements before returning accurate reconstructions. The wide disparity in reconstruction accuracy between local and global measurements for *Wirtinger Flow* illustrates the significant challenge in phase retrieval from local measurements. Furthermore, we see that the *BlockPR* method proposed in this paper is more noise tolerant than *Wirtinger Flow* for local measurements.

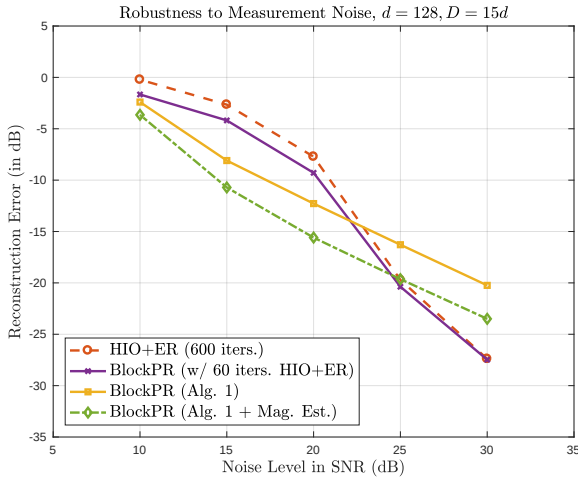
Given the weaker performance of *Wirtinger Flow* with local measurements, we now restrict our attention to the empirical evaluation of the proposed method (Alg. 1, as well as the post-processed variant (with HIO+ER iterations)) against *PhaseLift* and alternating projection algorithms. Although numerical simulations suggest that these methods work with local measurements, we note that (to the best of our knowledge) there are no theoretical recovery or robustness guarantees for these methods and measurements. The *PhaseLift* algorithm was implemented as a trace regularized least-squares problem using CVX [39, 40] – a package for specifying



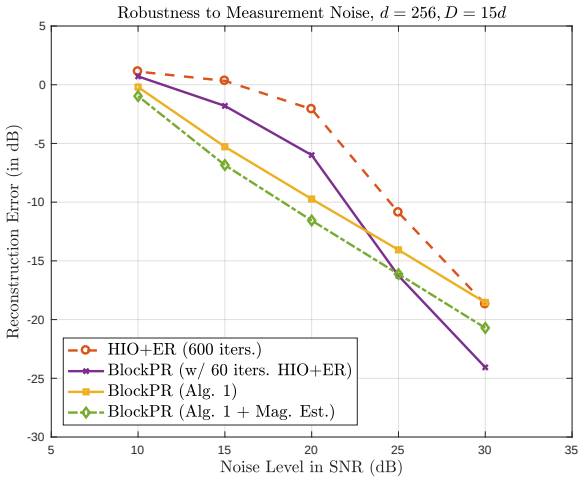
(a) Using $D = 7d$ measurements.



(b) Using $D = 15d$ measurements.



(c) Low SNR Simulations: Using $D = 15d$ measurements, Problem Size $d = 128$



(d) Low SNR Simulations: Using $D = 15d$ measurements, Problem Size $d = 256$

Figure 3.3: Robustness to measurement noise – Phase Retrieval from deterministic local correlation measurements.

and solving convex programs in Matlab. We consider two variants from the family of alternating projection methods – *Gerchberg–Saxton* (sometimes referred to as the Error Reduction (ER) algorithm) and Hybrid Input-Output (HIO). For both algorithms, the following two projections were utilized: (i) projection onto the measured magnitudes, and (ii) projection onto the span of the measurement vectors $\{\mathbf{a}_j\}_{j=1}^{D}$. This formulation as well as other details and connections to convex optimization theory can be found in [11]. For the ER and HIO implementations, the initial guess was set to be the all-zero vector.[9] For the HIO implementation, as is popular practice (see, for example, [31]) every few (20) HIO iterations were followed by a small number of (10) ER iterations, with the maximum number of HIO+ER iterations limited to 600 – this choice of iteration count ensures convergence of the algorithm (see Fig. 3.4) while comparing favorably with the computational cost (see Fig. 3.5b) of the proposed *BlockPR* method. For the ER implementation, $6,000$ iterations were necessary to ensure convergence.

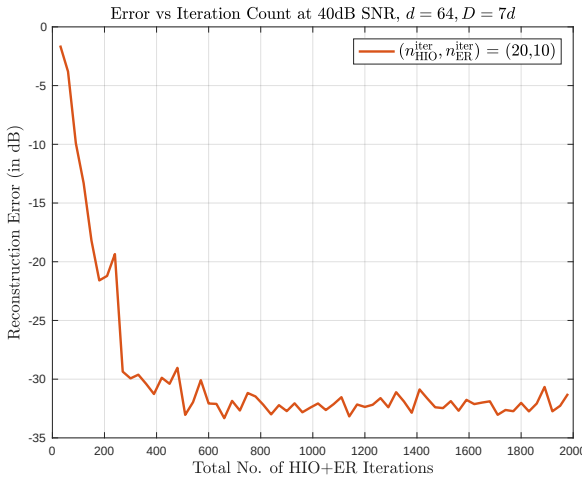


Figure 3.4: Reconstruction Error vs. Iteration Count for HIO+ER Implementation

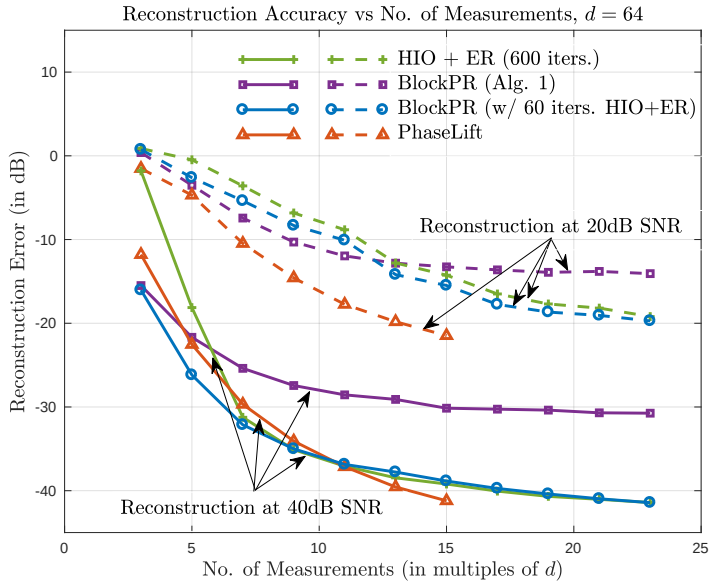
We begin by presenting numerical results evaluating the robustness to measurement noise. Figs. 3.3a and 3.3b plot the error in reconstructing a $d = 64$ length

[9] We note that using a random starting guess does not change the qualitative nature of the empirical results.

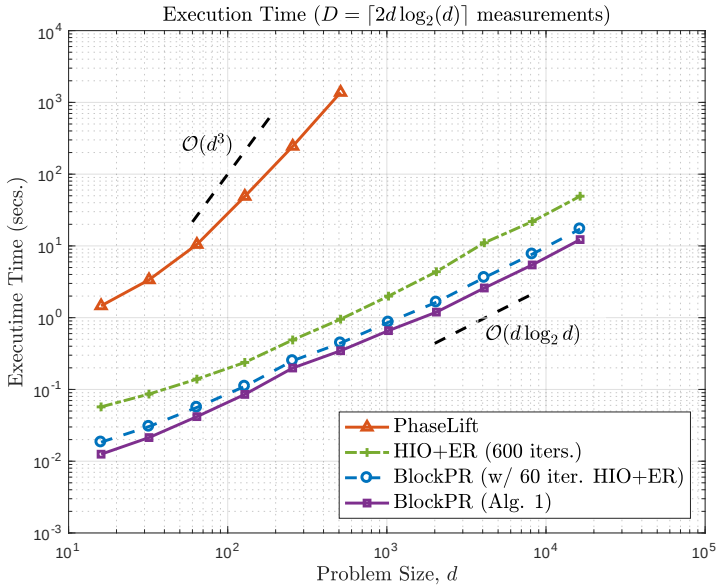
complex vector $\mathbf{x}_0$ using $D = 7d$ and $D = 15d$ local correlation-based phaseless measurements respectively. Note that this corresponds to using $\delta = 4$ (and the associated $2\delta - 1 = 7$ masks) and $\delta = 8$ (and the corresponding $2\delta - 1 = 15$ masks) respectively. Moreover, the well-conditioned *deterministic* (and sparse) measurement construction defined in Example 2 of Section 3.2 was utilized along with additive Gaussian measurement noise. We see from Fig. 3.3 that the method proposed in this paper (denoted *BlockPR* in the figure) performs reliably across a wide range of SNRs and compares favorably against existing popular phase retrieval algorithms. When using a small number of measurements (as in Fig. 3.3a, and modeling real-world situations), both variants of the proposed method – Alg. 1, as well as Alg. 1 post-processed using 60 HIO+ER iterations – outperform the ER algorithm by significant margins and compare well with the popular HIO+ER algorithm. When more measurements are available (as in Fig. 3.3b), the performance of the ER and HIO+ER algorithms approaches the performance of the *BlockPR* variants proposed in this paper. In addition, the proposed methods also compare well with the significantly more expensive *PhaseLift* reconstructions. We emphasize that the superior performance of the proposed methods demonstrated here comes with rigorous theoretical recovery guarantees for local measurements – something that cannot be said of any of the other methods in Fig. 3.3.

Additionally, Figs. 3.3c and 3.3d compare the performance of the HIO+ER algorithm at low SNRs for problems sizes $d = 128$ and $d = 256$ respectively, with the various *BlockPR* implementations – including one with the improved magnitude estimation procedure detailed in §3.6.1 (with $s = 1$ and using the average of the obtained $\tilde{D}_{j\prime}$ block magnitude estimates). These figures demonstrate the value of the magnitude estimation procedure from §3.6.1 at low SNRs over the HIO+ER post-processing method utilized in the other figures (and over the HIO+ER algorithm); we defer a more detailed study of this to future work.

Next, Fig. 3.5a plots the reconstruction error in recovering a $d = 64$-length complex



(a) Reconstruction Error vs. No. of Measurements

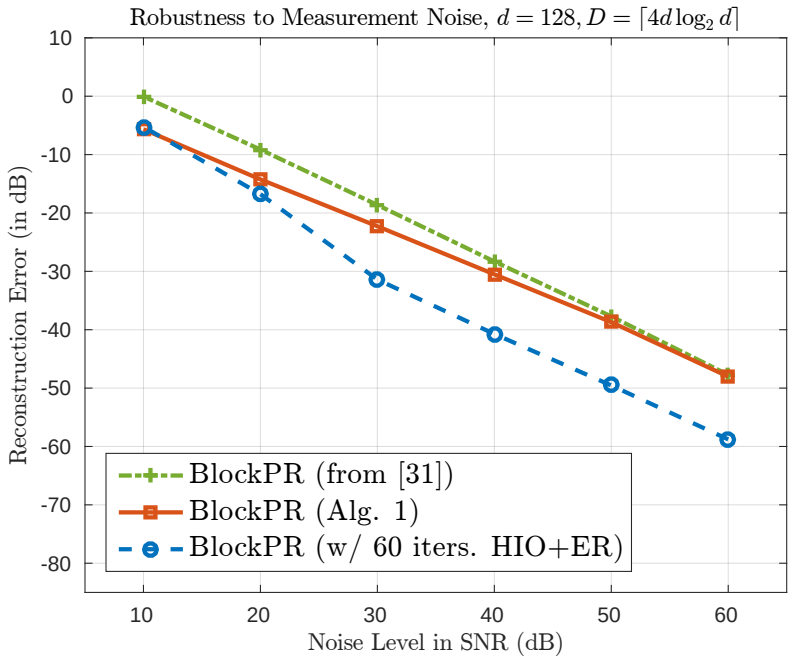


(b) Execution Time vs. Problem Size

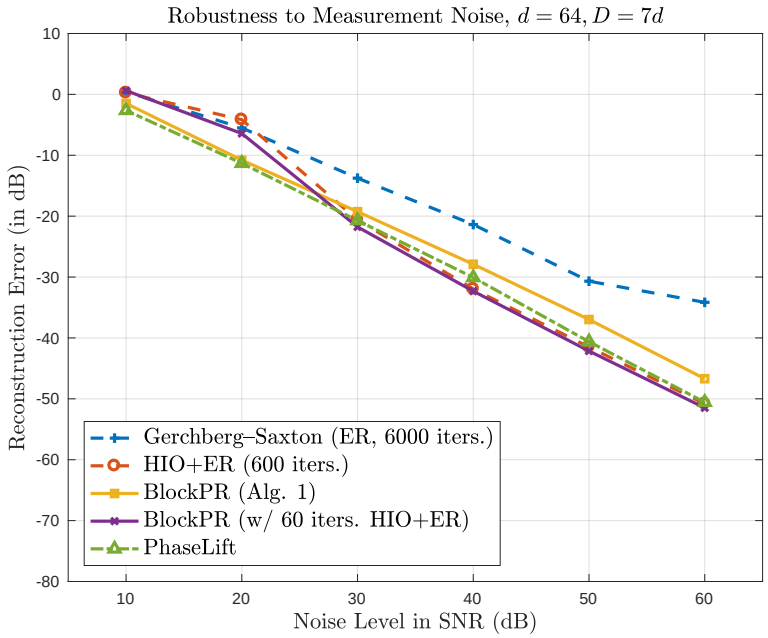
Figure 3.5: Performance Evaluation and Comparison of the Proposed Phase Retrieval Method (with Deterministic Local Correlation Measurements of Example 2, §3.2 and Additive Gaussian Noise)

vector as a function of the number of measurements used. This corresponds to using values of δ ranging from 2 to 12 (and the associated $2\delta - 1$ masks). As with Fig. 3.3, the deterministic correlation-based measurement constructions of Section 3.2 (Example construction 2) were utilized along with an additive Gaussian noise model. Plots are provided for simulations at two noise levels – 20 dB and 40 dB. Comparing the performance of the two *BlockPR* variants, we observe that the HIO+ER post-processing procedure provides improved reconstruction errors – with the margin of improvement increasing when more measurements are available. We also notice that both variants of *BlockPR* compare particularly well with the other algorithms (HIO+ER and *PhaseLift*) when small numbers of measurements are available.

Finally, Fig. 3.5b plots the average execution time (in seconds) required to solve the phase retrieval problem using $D = \lceil 2d \log_2 d \rceil$ measurements. For comparison, execution times for the *PhaseLift* and HIO+ER alternating projection algorithms are provided. We observe that both variants of the proposed method are several orders of magnitude faster than the *PhaseLift* algorithm,[10] and between 2–5 times faster than the HIO+ER implementation. Moreover, the plot illustrates the essentially FFT-time computational complexity (see §3.1.3) of the proposed method. Between the two *BlockPR* variants, we see that there is a small trade-off between reduced execution time and improved accuracy; one of the two variants may be more appropriate depending on the application requirements.



(a) Improved Robustness to Measurement Noise – Comparing Variants of the *BlockPR* algorithm



(b) Phase Retrieval from $D = 7d$ Measurements – Comparison with Other Phase Retrieval Algorithms

Figure 3.6: Numerical Evaluation of the Proposed Algorithm with the Ptychographic Measurements from Example 1 of §3.2

3.6.3 Experiments with Ptychographic Measurements from Example 1 of §3.2

We now present a selection of empirical results demonstrating the accuracy, robustness, and efficiency of the proposed method when using the ptychographic measurements from Example 1, §3.2 and observe that, in this case, the comparison between the proposed algorithm and existing methods is similar to that in section §3.6.2. Recall that (see Example 1 from §3.2 and §3.1.4 for details) this measurement construction corresponds to a discretization of the ptychographic measurements $\left|\mathcal{F}[\widetilde{h} \cdot S_l f](\omega)\right|^2$ as per (3.11), where f is the unknown specimen and $\widetilde{h}(t) = \frac{e^{-t/a}}{\sqrt[4]{2\delta-1}}$ denotes the *single, deterministic* illumination function (or mask). As before, δ defines the local support of the illumination function and we consider a discretization involving $2\delta - 1$ Fourier modes and sample/specimen shifts of 1 unit or pixel (yielding a total of d shifts in the discrete problem formulation).

We begin with Fig. 3.6a, which demonstrates the relative performance of the *BlockPR* variants described in [46] and this paper. More specifically, we plot the error in reconstructing a $d = 128$ length complex Gaussian test signal using $D = \lceil 4d \log_2 d \rceil$ measurements at different added noise levels. As with Fig. 3.2a, we plot results for three different variants of the *BlockPR* algorithm: (i) Algorithm 1, (ii) Algorithm 1 with the HIO+ER post-processing procedure as described in §3.6.2, and (iii) the implementation from [46]. We again observe (as in Fig. 3.2a) that the methods described in this paper (which use eigenvector-based angular synchronization) are more accurate than that detailed in [46] (which uses a greedy angular synchronization method), with the improvement in performance being especially significant at low SNRs. Next, Fig. 3.6b studies the performance of the proposed method(s) against three other popular phase retrieval algorithms –

[10] For computational efficiency and due to memory constraints, the *PhaseLift* plot in Fig. 3.5b was generated using the TFOCS software package (http://cvxr.com/tfocs/) instead of the more computationally expensive CVX software package.

PhaseLift, the *Gerchberg–Saxton* Error Reduction (ER) algorithm, and the Hybrid Input–Output (HIO) algorithm (with implementation parameters identical to those in §3.6.2). As with Fig. 3.3a, we consider the reconstruction of a $d = 64$ length complex vector x_0 using $D = 7d$ measurements in the presence of additive Gaussian noise. We observe that the methods proposed in this paper outperform the ER algorithm and compare very well with the HIO+ER and (the significantly more expensive) *PhaseLift* algorithms across a wide range of SNRs. Finally, we note that the execution time plot for this ptychographic measurement construction is qualitatively and quantitatively similar to Fig. 3.5b – the proposed methods are faster (by a factor of 2–5) than an equivalent HIO+ER implementation and orders of magnitude faster than convex optimization approaches such as *PhaseLift*.

3.7 Concluding Remarks

In this paper new and improved deterministic robust recovery guarantees are proven for the phase retrieval problem using local correlation measurements. In addition, a new practical phase retrieval algorithm is presented which is both faster and more noise robust than previously existing approaches (e.g., alternating projections) for such local measurements.

Future work might include the exploration of more general classes of measurements which are guaranteed to lead to well conditioned linear systems of the type used to reconstruct $X \approx X_0$ in line 1 of Algorithm 1. Currently two deterministic measurement constructions are known (recall, e.g., Section 3.2) – it should certainly be possible to construct more general families of such measurements.

Other interesting avenues of inquiry include the theoretical analysis of the magnitude estimate approach proposed in Section 3.6.1 in combination with the rest of Algorithm 1. Alternate phase retrieval approaches might also be developed by

using such local block eigenvector-based methods for estimating phases too, instead of just using the single global top eigenvector as currently done in line 3 of Algorithm 1.

Finally, more specific analysis of the performance of the proposed methods using masked/windowed Fourier measurements (recall Section 3.1.5) would also be interesting. In particular, an analysis of the performance of such approaches as a function of the bandwidth of the measurement mask/window could be particularly enlightening. One might also consider extending the discrete results of this paper to the analytic setting by, e.g., expanding on [58].

3.8 Alternate Perturbation Bounds

In this section we present a simpler (and easier to derive), albeit weaker, perturbation result in the spirit of Section 3.4, which is associated with the analysis of line 3 of Algorithm 1. Specifically, we will derive an upper bound on $\min_{\theta\in[0,2\pi]} \|\widetilde{\mathbf{x}}_0 - \mathrm{e}^{\mathrm{i}\theta}\widetilde{\mathbf{x}}\|_2$ (provided by Theorem 6), which scales like d^3. While this dependence is strictly worse than the one derived in Section 3.4, it is easier to obtain and the technique may be of independent interest.

We will begin with a result concerning the top eigenvector of any Hermitian matrix.

Lemma 1. Let $X_0 = \sum_{j=1}^{d} \nu_j \mathbf{x}_j \mathbf{x}_j^*$ be Hermitian with eigenvalues $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_d$ and orthonormal eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_d \in \mathbb{C}^d$. Suppose that $X = \sum_{j=1}^{d} \lambda_j \mathbf{v}_j \mathbf{v}_j^*$ is Hermitian with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_d \in \mathbb{C}^d$, and $\|X - X_0\|_{\mathrm{F}} \leq \eta \|X_0\|_F$ for some $\eta \geq 0$. Then,

$$\left(1 - |\langle \mathbf{x}_1, \mathbf{v}_1 \rangle|^2\right) \leq \frac{4\eta^2 \|X_0\|_{\mathrm{F}}^2}{(\nu_1 - \nu_2)^2}.$$

Proof. An application of the $\sin\theta$ theorem [23, 70] (see, e.g., the proof of Corollary 1 in [84]) tells us that

$$\sin\left(\arccos\left(|\langle \mathbf{x}_1, \mathbf{v}_1\rangle|\right)\right) \leq \frac{2\eta\|X_0\|_F}{|\nu_1 - \nu_2|}.$$

Squaring both sides we then learn that

$$\left(1 - |\langle \mathbf{x}_1, \mathbf{v}_1\rangle|^2\right) = \sin^2\left(\arccos\left(|\langle \mathbf{x}_1, \mathbf{v}_1\rangle|\right)\right) \leq \frac{4\eta^2\|X_0\|_F^2}{(\nu_1 - \nu_2)^2}, \tag{3.31}$$

giving us the desired inequality. □

The following variant of Lemma 1 concerning rank 1 matrices X_0 is of use in the analysis of many other phase retrieval methods, and can be used, e.g., to correct and simplify the proof of equation (1.8) in Theorem 1.3 of [18].

Lemma 2. Let $\mathbf{x}_0 \in \mathbb{C}^d$, set $X_0 = \mathbf{x}_0\mathbf{x}_0^*$, and let $X \in \mathbb{C}^{d\times d}$ be Hermitian with $\|X - X_0\|_F \leq \eta\|X_0\|_F = \eta\|\mathbf{x}_0\|_2^2$ for some $\eta \geq 0$. Furthermore, let λ_i be the i-th largest magnitude eigenvalue of X and $\mathbf{v}_i \in \mathbb{C}^d$ an associated eigenvector, such that the $\mathbf{v}_i$ form an orthonormal eigenbasis. Then

$$\min_{\theta\in[0,2\pi]} \|e^{i\theta}\mathbf{x}_0 - \sqrt{|\lambda_1|}\mathbf{v}_1\|_2 \leq (1 + 2\sqrt{2})\eta\|\mathbf{x}_0\|_2.^{11}$$

Proof. In this special case of Lemma 1 we have $\nu_1 = \|X_0\|_F = \|\mathbf{x}_0\|_2^2$ and $\mathbf{x}_1 := \mathbf{x}_0/\|\mathbf{x}_0\|$. Choose $\phi \in [0, 2\pi]$ such that $\langle e^{i\phi}\mathbf{x}_0, \mathbf{v}_1\rangle = |\langle \mathbf{x}_0, \mathbf{v}_1\rangle|$. Then,

$$\begin{aligned}
\|e^{i\phi}\mathbf{x}_0 - \sqrt{\nu_1}\mathbf{v}_1\|_2^2 &= 2\nu_1 - 2\nu_1 \cdot |\langle \mathbf{x}_0/\|\mathbf{x}_0\|, \mathbf{v}_1\rangle| = 2\nu_1 - 2\nu_1 \cdot |\langle \mathbf{x}_1, \mathbf{v}_1\rangle| \\
&\leq 2\nu_1\left(1 - |\langle \mathbf{x}_1, \mathbf{v}_1\rangle|\right)\left(1 + |\langle \mathbf{x}_1, \mathbf{v}_1\rangle|\right) \\
&= 2\nu_1\left(1 - |\langle \mathbf{x}_1, \mathbf{v}_1\rangle|^2\right) \leq 8\eta^2\|X_0\|_{\mathrm{F}}
\end{aligned} \tag{3.32}$$

[11] It is interesting to note that similar bounds can also be obtained using simpler techniques (see, e.g., [45]).

where the last inequality follows from Lemma 1 with $\nu_1 = \|X_0\|_F = \|\mathbf{x}_0\|_2^2$. Finally, by the triangle inequality, Weyl's inequality (see, e.g., [44]), and (3.32), we have

$$\begin{aligned}
\|\mathbb{e}^{\mathbb{i}\phi}\mathbf{x}_0 - \sqrt{|\lambda_1|}\mathbf{v}_1\|_2 &\leq \|\mathbb{e}^{\mathbb{i}\phi}\mathbf{x}_0 - \sqrt{\nu}_1\mathbf{v}_1\|_2 + \|\sqrt{\nu}_1\mathbf{v}_1 - \sqrt{|\lambda_1|}\mathbf{v}_1\|_2 \\
&\leq 2\sqrt{2}\cdot\eta\sqrt{\nu}_1 + \left|\sqrt{\nu}_1 - \sqrt{|\lambda_1|}\right| \\
&\leq 2\sqrt{2}\cdot\eta\sqrt{\nu}_1 + \frac{|\nu_1 - \lambda_1|}{\sqrt{\nu}_1 + \sqrt{|\lambda_1|}} \\
&\leq 2\sqrt{2}\cdot\eta\sqrt{\nu}_1 + \frac{\eta\nu_1}{\sqrt{\nu}_1 + \sqrt{|\lambda_1|}} \\
&\leq (1+2\sqrt{2})\eta\sqrt{\nu}_1.
\end{aligned}$$

The desired result now follows. □

We may now use Lemma 1 to produce a perturbation bound for our banded matrix of phase differences $\tilde{X}_0$ from (3.7).

Theorem 6. *Let $\widetilde{X}_0 = T_\delta(\widetilde{\mathbf{x}}_0\widetilde{\mathbf{x}}_0^*)$ where $|(\widetilde{\mathbf{x}}_0)_i| = 1$ for each i. Further suppose $\widetilde{X} \in T_\delta(\mathcal{H}^d)$ has $\widetilde{\mathbf{x}}$ as its top eigenvector, where $||\widetilde{\mathbf{x}}||_2 = \sqrt{d}$. Suppose that $\|\widetilde{X}_0 - \widetilde{X}\|_F \leq \eta\|\widetilde{X}_0\|_F$ for some $\eta > 0$. Then, there exists an absolute constant $C \in \mathbb{R}^+$ such that*

$$\min_{\theta\in[0,2\pi]} \|\widetilde{\mathbf{x}}_0 - \mathbb{e}^{\mathbb{i}\theta}\widetilde{\mathbf{x}}\|_2 \leq C\frac{\eta d^3}{\delta^{\frac{5}{2}}}.$$

Proof. Recall that the phase vectors $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}_0$ are normalized so that $\|\tilde{\mathbf{x}}\|_2 = \|\tilde{\mathbf{x}}_0\|_2 = \sqrt{d}$. Combining Lemmas 2 and 1 after noting that $\|\tilde{X}_0\|_{\mathrm{F}}^2 = d(2\delta - 1)$ we learn that

$$\left(1 - \frac{1}{d^2}|\langle\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}\rangle|^2\right) \leq C'\eta^2\left(\frac{d}{\delta}\right)^5 \tag{3.33}$$

for an absolute constant $C' \in \mathbb{R}^+$. Let $\phi \in [0, 2\pi)$ be such that $\mathrm{Re}\left(\langle\tilde{\mathbf{x}}_0, \mathbb{e}^{\mathbb{i}\phi}\tilde{\mathbf{x}}\rangle\right) = |\langle\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}\rangle|$. Then,

$$\begin{aligned}
\|\tilde{\mathbf{x}}_0 - \mathbb{e}^{\mathbb{i}\phi}\tilde{\mathbf{x}}\|_2^2 &= 2d - 2\,\mathrm{Re}\left(\langle\tilde{\mathbf{x}}_0, \mathbb{e}^{\mathbb{i}\phi}\tilde{\mathbf{x}}\rangle\right) \\
&= 2d\left(1 - \frac{1}{d}\left|\langle\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}\rangle\right|\right) \leq 2d\left(1 - \frac{1}{d^2}\left|\langle\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}\rangle\right|^2\right).
\end{aligned}$$

Combining this last inequality with (3.33) concludes the proof. □

Acknowledgements

The authors would like to thank Felix Krahmer for helpful discussions regarding Lemma 2. MI and RS would like to thank the Hausdorff Institute of Mathematics, Bonn for its hospitality during its Mathematics of Signal Processing Trimester Program. A portion of this work was completed during that time. In addition, the authors would like to thank and acknowledge the Institute for Mathematics and its Applications (IMA) for the workshop "Phaseless Imaging in Theory and Practice: Realistic Models, Fast Algorithms, and Recovery Guarantees" hosted there in a August of 2017. The revision of this paper benefitted greatly from many discussions with the participants there. In particular, conversations with James Fienup, Andreas Menzel, and Irene Waldspurger were exceedingly helpful. MI was supported in part by NSF DMS-1416752. RS was supported in part by a Hellman Fellowship and NSF DMS-1517204.

Acknowledgement of joint authorship

Chapter 3, in full, is a reprint of the material accepted for publication by Applied Computational and Harmonic Analysis. Iwen, Mark A.; Preskitt, Brian; Saab, Rayan; Viswanathan, Aditya. *Phase retrieval from local measurements: Improved robustness via eigenvector-based angular synchronization.* Available online as of 18 June 2018.

Chapter 4

Invertible Local Measurement Systems

In Chapter 3, we proposed the algorithm for robustly solving phase retrieval whose analysis and extension forms the basis of this dissertation. This algorithm, stated in Algorithm 1, operated in two steps: the first step is to solve a linear system, obtained by recasting the magnitude-only linear measurements $y_{(\ell,j)} = |\langle x_0, S^\ell m_j\rangle|^2 + \eta_{j\ell}$ as linear measurements on the space of Hermitian matrices, by rewriting

$$|\langle x_0, S^\ell m_j\rangle|^2 = \mathrm{Tr}\left(S^\ell m_j m_j^* S^{-\ell} x_0 x_0^*\right) = \langle S^\ell m_j m_j^* S^{-\ell}, x_0 x_0^*\rangle.$$

By design, the "vectors" $\{S^\ell m_j m_j^* S^{-\ell}\}$ are all contained in the subspace $T_\delta(\mathcal{H}^d)$ of $\mathcal{H}^d$ (defined in (3.3)), where d is the ambient dimension (meaning $x_0, m_j \in \mathbb{C}^d$) and δ is the support size of the masks (so $\mathrm{supp}(m_j) \subseteq [\delta]$). Hence, we used our measurements to solve for (an estimate of) the projection of $x_0 x_0^*$ onto $T_\delta(\mathcal{H}^d)$, namely $T_\delta(x_0 x_0^* + N) =: X$.

After the linear step, the second step is recovering x_0 from X. We accomplish this by inferring the magnitudes of x_0's entries from the main diagonal of X (as the

main diagonal is preserved by the projection T_δ) and finding their relative phases through an angular synchronization problem (see Section 3.4 and chapter 6).

The subject of this chapter is to study in more detail the linear system solved in the first step. Most crucially, we consider that, in the work of Chapter 3, we only produced two examples of collections of vectors $\{m_j\}_{j\in[2\delta-1]} \subseteq \mathbb{C}^d$ such that this linear system was invertible. Considering that one of the major contributions of our algorithm is that it admits a measurement model that somewhat replicates laboratory conditions, our knowledge of which vectors are compatible with our algorithm and the theory built for it is critical in promoting the applicability of this work. Therefore, in Section 4.1, we fully characterize a broad class of familes of masks – whose design is, notably, motivated by the application of ptychography, described in Section 3.1.4 – that are guaranteed to produce an invertible linear system of the form in (3.5). The basic idea of Section 4.1 is to generalize Example 1 of Section 3.2, stated in eq. (3.17), by abstracting out the term $\frac{\mathrm{e}^{-n/a}}{\sqrt[4]{2\delta-1}}$ and considering what "base vectors" $\gamma \in \mathbb{R}^d$ will produce spanning families of masks when m_j is taken to be $m_j = f_j^K \circ \gamma$, for some $K \in \mathbb{N}$. In Proposition 1, we describe conditions on γ and K which are both necessary and sufficient to produce an invertible linear system, providing a satisfying answer to this inquiry.

In Section 4.2, we analyze the conditioning of the linear systems produced by this class of mask families. We recall that the recovery guarantees of Section 3.5, most notably Corollary 3, rely on the condition number κ of this linear system. Indeed, any result guaranteeing the robustness of our algorithm must depend on κ, as the linear solve, along with whatever noise is inflated or compressed by it, feeds directly into the magnitude estimation and angular synchronization methods of the recovery step. Fortunately, in Proposition 2, we are able to calculate the condition number of this linear system exactly whenever the support size of the masks δ satisfies $2\delta - 1 \leq d$ (which includes the vast majority of cases; typically we assume $\delta \ll d$), and provide an estimate that covers all other cases.

We consider the act of inverting $\mathcal{A}$ from a practical perspective in Section 4.3. We write its inverse explicitly and analyze the runtime of calculating $\mathcal{A}^{-1}(y)$ in Section 4.3.1. In Section 4.3.3, we analyze the variance of the individual entries of $\mathcal{A}^{-1}(y)$ when the measurements are exposed to uniform, Gaussian noise. In Section 4.4, we provide a few examples of explicit $\gamma \in \mathbb{R}^d$ that are proven to satisfy the conditions to span $T_\delta(\mathcal{H}^d)$, and finally, we illustrate the impact these results with numerical studies in Section 4.5.

Before we begin these analyses, we introduce and unify some definitions and notational elements that will make discussion of the mathematical details more convenient.

Definition 1. We say that $\{m_j\}_{j=1}^D \subseteq \mathbb{C}^d$ is a *local measurement system* or *family of masks* of support δ if $1 \in \text{supp}(m_j)$ and $\text{supp}(m_j) \subseteq [\delta]$ for each j.

Definition 2. Let $\{m_j\}_{j=1}^D \subseteq \mathbb{C}^d$ be a local measurement system of support δ. If each m_j satisfies $m_j = \mathcal{R}_d(\sqrt{K} f_j^K) \circ \gamma$ for some $K \geq \max(\delta, D), \gamma \in \mathbb{C}^d$ satisfying $\text{supp}(\gamma) = [\delta]$, then we call $\{m_j\}_{j=1}^D$ a *local Fourier measurement system* of support δ with mask γ and modulation index K. If $K = D = 2\delta - 1$, then we simply refer to $\{m_j\}_{j=1}^D$ as a *local Fourier measurement system* of support δ with mask γ. We add that, if we say that $\{m_j\}_{j=1}^D$ is a local Fourier measurement system with support δ and mask γ, this implies an assertion that $\text{supp}(\gamma) = [\delta]$.

Definition 3. Given a local measurement system $\{m_j\}_{j=1}^D$ in $\mathbb{C}^d$, the associated *lifted measurement system* is the set $\mathcal{L}_{\{m_j\}} = \{S^\ell m_j m_j^* S^{-\ell}\}_{(\ell,j) \in [d]_0 \times [D]} \subseteq \mathbb{C}^{d \times d}$.

Definition 4. We say that a family of masks $\{m_j\}_{j=1}^D \subseteq \mathbb{C}^d$ of support δ is a *spanning family* if $\text{span}\, \mathcal{L}_{\{m_j\}} = T_\delta(\mathcal{H}^d)$.

Definition 5. Given a local measurement system $\{m_j\}_{j=1}^S$, the *measurement operator* associated with these vectors is the operator

$$\begin{aligned} \mathcal{A} &: T_\delta(\mathbb{C}^{d \times d}) \to \mathbb{C}^{[d]_0 \times [D]} \\ \mathcal{A}(X)_{(\ell,j)} &= \langle S^\ell m_j m_j^* S^{-\ell}, X \rangle. \end{aligned} \tag{4.1}$$

The *canonical matrix representation* of $\mathcal{A}$ is the matrix $A \in \mathbb{C}^{dD \times d(2\delta-1)}$, defined by

$$\left(A \begin{bmatrix} \operatorname{diag}(X, 1-\delta) \\ \vdots \\ \operatorname{diag}(X, \delta-1) \end{bmatrix} \right)_{(j-1)d+\ell} = \mathcal{A}(X)_{(\ell-1,j)}. \tag{4.2}$$

For convenience, we define the *diagonal vectorization operator* $\mathcal{D}_I : \mathbb{C}^{d\times d} \to \mathbb{C}^{|I|d}$ for any subset $\{m_i\}_{i=1}^{|I|} = I \subseteq [d]$ and $\mathcal{D}_k : \mathbb{C}^{d\times d} \to \mathbb{C}^{(2k-1)d}$ for any integer $k \leq \frac{d+1}{2}$ by

$$\mathcal{D}_I(X) = \begin{bmatrix} \operatorname{diag}(X, m_1) \\ \vdots \\ \operatorname{diag}(X, m_{|I|}) \end{bmatrix} \tag{4.3}$$

$$\mathcal{D}_k(X) = \mathcal{D}_{[2k-1]_{1-k}}(X) = \begin{bmatrix} \operatorname{diag}(X, 1-k) \\ \vdots \\ \operatorname{diag}(X, k-1) \end{bmatrix}, \tag{4.4}$$

so that (4.2) becomes $A\mathcal{D}_\delta(X)_{(j-1)d+\ell} = \mathcal{A}(X)_{(\ell,j)}$. We remark that $\mathcal{D}_k$ is invertible on $T_k(\mathbb{C}^{d\times d})$, and for $v \in \mathbb{C}^{d(2k-1)}$, we use $\mathcal{D}_k^{-1}(v)$ to represent the matrix in $T_k(\mathbb{C}^{d\times d})$ whose diagonals are given by the $2k-1$ distinct d-length blocks of v.

4.1 Conditions for a Spanning Family

Proposition 1. *Suppose that $\gamma \in \mathbb{R}^d$ has $1 \in \operatorname{supp}(\gamma) = [\delta]$. Set $D = \min\{2\delta - 1, d\}$, take $K \geq 2\delta - 1$ and let*

$$\begin{matrix} v_j = \sqrt{K}\mathcal{R}_d(F_K e_j) \\ v_j^D = \sqrt{K}\mathcal{R}_D(F_K e_j) \end{matrix}, \quad j \in [D],\ 2\delta - 1 \leq K.$$

Define a local measurement system $\{m_j\}_{j\in[D]}$ by setting $m_j = \gamma \circ v_j$. Then $\{m_j\}_{j\in[D]}$ is a spanning family if and only if all the sets $J_k := \{m \in [\delta]_0 :$

$(F_d(\gamma \circ S^{-m}\gamma))_k \neq 0\}$, *for all* $k \in [d]$ *satisfy*

$$\begin{cases} 2|J_k| - 1 \geq D, & 0 \in J_k \\ 2|J_k| \geq D, & \textit{otherwise} \end{cases}.$$

The proof will make use of the following lemmas.

Lemma 7. *Define* $w_j = \mathcal{R}_{N_1}(f_j^{N_2}), j \in [N_2]$ *and set*

$$\rho_j = \mathrm{Re}(w_j) \quad \textit{and} \quad \mu_j = \mathrm{Im}(w_j)$$

to be vectors containing the real and imaginary components of w_j*. Then for* $1 \leq \ell_1 < \cdots < \ell_k \leq \frac{N_2+1}{2}$ *with* $k \leq N_1$*, we have*

$$\begin{aligned} \dim \operatorname{span}\{w_{\ell_i}, w_{2-\ell_i}\}_{i=1}^k &= \dim \operatorname{span}\{\rho_{\ell_i}, \mu_{\ell_i}\}_{i=1}^k \\ &= \begin{cases} 2k-1, & \ell_1 = 1 \\ 2k, & \textit{otherwise} \end{cases}, \end{aligned}$$

where the indices are taken modulo N_2*.*

Proof of lemma 7. The first equality is clear by considering that $w_{2-i} = \overline{w_i}$, so $\rho_k = \frac{1}{2}(w_i + w_{2-i})$ and $\mu_i = -\frac{i}{2}(w_i - w_{2-i})$. We set $M = \dim \operatorname{span}\{w_{\ell_i}, w_{2-\ell_i}\}_{i=1}^k$ to be the common dimension of the two spaces under consideration.

We now divide into two cases: if $N_1 < N_2$, then $\{w_j\}_{j\in[N_2]}$ is full spark, as any $N_1 \times N_1$ submatrix of $\begin{bmatrix} w_1 & \cdots & w_{N_2} \end{bmatrix}$ will be a Vandermonde matrix of the form

$$V = \frac{1}{\sqrt{N_2}} \begin{bmatrix} w_{\ell_1} & \cdots & w_{\ell_{N_1}} \end{bmatrix}$$

with determinant

$$N_2^{-N_1/2} \prod_{1 \leq i < j \leq N_1} (\omega_{N_2}^{\ell_i - 1} - \omega_{N_2}^{\ell_j - 1}),$$

which is immediately non-zero since $\omega_{N_2}^{\ell_i-1} - \omega_{N_2}^{\ell_j-1} = 0$ only when $\ell_i - \ell_j = 0 \bmod N_2$, which cannot happen when $N_1 < N_2$.

When $N_1 \geq N_2, \{w_j\}_{j\in[N_2]}$ is linearly independent, since its members form the matrix $\begin{bmatrix} F_{N_2} \\ 0_{N_1-N_2\times N_2} \end{bmatrix}$.

In either case, M is equal to the cardinality of $\{\ell_i, 2-\ell_i\}_{i=1}^k$, which has $2k-1$ elements if and only if $\ell_1 = 1$; otherwise it has $2k$. We remark that a collision where $\ell_i = (2-\ell_i \bmod N_2) = N_2/2+1$ is precluded since we have asserted $\ell_i \leq \frac{N_2+1}{2}$.

□

Lemma 8. *For $v \in \mathbb{R}^d$, we have*

$$\operatorname{circ}(v)\rho_k^d = \frac{1}{2}\operatorname{Re}((Fv)_k f_k^d) \tag{4.5}$$

$$\operatorname{circ}(v)\mu_k^d = \frac{1}{2}\operatorname{Im}((Fv)_k f_k^d). \tag{4.6}$$

In particular, if $(Fv)_k \neq 0$ and $k \notin \{1, \frac{d}{2}+1\}$, then $\rho_k^d, \mu_k^d \notin \operatorname{Nul}(\operatorname{circ}(v))$; if $k \in \{1, \frac{d}{2}+1\}$, then $\rho_k^d \notin \operatorname{Nul}(\operatorname{circ}(v))$ and $\mu_k^d = 0$. On the other hand, if $(Fv)_k = 0$, then $\rho_k^d, \mu_k^d \in \operatorname{Nul}(\operatorname{circ}(v))$.

Proof of lemma 8. We set $\lambda_k^d = (Fv)_k$, and recalling that $\operatorname{circ}(v) = F\operatorname{diag}(Fv)F^*$, we observe that

$$\begin{aligned}\operatorname{circ}(v)\mu_k^d &= \operatorname{circ}(v)\frac{1}{2}(f_k^d + f_{2-k}^d) = \frac{1}{2}(\operatorname{circ}(v)f_k^d + \operatorname{circ}(v)f_{2-k}^d) \\ &= \frac{1}{2}(\lambda_k^d f_k^d + \lambda_{2-k}^d f_{2-k}^d).\end{aligned}$$

(4.5) follows immediately since $\lambda_k^d = \overline{\lambda_{2-k}^d}$ when $v \in \mathbb{R}^D$. (4.6) follows from an analogous calculation.

If $\lambda_k^d \neq 0$ and $k \notin \{1, \frac{d}{2}+1\}$, then ω_d^{k-1} is a non-real root of unity and there exists some j such that $\operatorname{Re}(\omega_d^{(j-1)(k-1)}\lambda_k^d) \neq 0$, and similarly for $\operatorname{Im}(\omega_d^{(j-1)(k-1)}\lambda_k^d) \neq 0$. When $k \in \{1, \frac{d}{2}+1\}, \omega_d^{(k-1)} \in \mathbb{R}$ so $\mu_k^d = 0$, but $\lambda_k^d \in \mathbb{R}$ in this case (because $v \in \mathbb{R}^d$), so $\operatorname{circ}(v)\rho_k^d = \lambda_k^d\rho_k^d \neq 0$. The claim concerning the case of $\lambda_k^d = 0$ is immediate from (4.5) and (4.6). □

Proof of proposition 1. For this proof, we set

$$(\rho_k^d, \mu_k^d) = (\mathrm{Re}(f_k^d), \mathrm{Im}(f_k^d))$$
$$(\rho_k, \mu_k) = (\mathrm{Re}(v_k), \mathrm{Im}(v_k))$$
$$(\rho_k^D, \mu_k^D) = (\mathrm{Re}(v_k^D), \mathrm{Im}(v_k^D))$$

In this case, we identify $\mathcal{L}_\gamma := \mathcal{L}_{\{m_j\}}$. By a basic dimension count, $\{m_j\}_{j\in[D]}$ is a spanning family if and only if $\mathcal{L}_\gamma$ is linearly independent, so we consider the conditions under which a linear combination of this lifted measurement system can be equal to zero. To this end, we define the operator $\mathcal{A}^* : \mathbb{R}^{d\times D} \to \mathbb{C}^{d\times d}$ by

$$\mathcal{A}^*(C) = \sum_{\ell\in[d], j\in[D]} C_{\ell,j} S^\ell m_j m_j^* S^{-\ell} \tag{4.7}$$

and begin with the observation that, for any $A \in \mathbb{C}^{d\times d}$ we have

$$\mathrm{diag}(S^\ell A S^{-\ell}, m) = S^\ell \,\mathrm{diag}(A, m).$$

We then have

$$\sum_{j\in[D],\ell\in[d]} C_{\ell,j} S^\ell m_j m_j^* S^{-\ell} = 0$$
$$\iff \mathrm{diag}\left(\sum_{j\in[D],\ell\in[d]} C_{\ell,j} S^\ell m_j m_j^* S^{-\ell}, m\right) = 0 \quad \text{for all } m \in [\delta]_0$$
$$\iff \sum_{j\in[D],\ell\in[d]} C_{\ell,j} \,\mathrm{diag}(S^\ell m_j m_j^* S^{-\ell}, m) = 0 \quad \text{for all } m \in [\delta]_0$$
$$\iff \sum_{j\in[D],\ell\in[d]} C_{\ell,j} S^\ell \,\mathrm{diag}(m_j m_j^*, m) = 0 \quad \text{for all } m \in [\delta]_0$$

At this point, we consider that

$$\mathrm{diag}(m_j m_j^*, m) = \mathrm{diag}((\gamma \circ v_j)(\gamma \circ v_j)^*, m) = \mathrm{diag}(D_{v_j}\gamma\gamma^* D_{\overline{v_j}}, m) \tag{4.8}$$
$$= \omega_K^{m(j-1)} \,\mathrm{diag}(\gamma\gamma^*, m). \tag{4.9}$$

We now set $g_m := \operatorname{diag}(\gamma\gamma^*, m) = \gamma \circ S^{-m}\gamma$ and proceed with the previous chain of implications:

$$\begin{aligned}
&\sum_{j\in[D],\ell\in[d]} C_{\ell,j} S^\ell \operatorname{diag}(m_j m_j^*, m) = 0 \quad \text{for all } m \in [\delta]_0 \\
\iff &\sum_{j\in[D],\ell\in[d]} C_{\ell,j} S^\ell (\omega_K^{m(j-1)} g_m) = 0 \quad \text{for all } m \in [\delta]_0 \\
\iff &\sum_{j\in[D],\ell\in[d]} C_{\ell,j} \omega_K^{m(j-1)} S^\ell g_m = 0 \quad \text{for all } m \in [\delta]_0 \\
\iff &\operatorname{circ}(g_m) C\, v_{m+1}^D = 0 \quad \text{for all } m \in [\delta]_0
\end{aligned}$$

We now recall that any circulant matrix $\operatorname{circ}(v)$ is diagonalized by the Discrete Fourier Matrix, such that, for $v \in \mathbb{C}^d$,

$$\operatorname{circ}(v) = F_d \operatorname{diag}(\sqrt{d} F_d v) F_d^* = \sqrt{d} \sum_{j=1}^{d} (F_d v)_j f_j^d (f_j^d)^*. \tag{4.10}$$

By writing $\lambda_k^m = \sqrt{d}(F g_m)_k$, we get a natural decoupling of the previous equations: for a fixed m, we have that $\operatorname{circ}(g_m) C f_{m+1} = 0$ if and only if

$$\sum_{k=1}^{d} \lambda_k^m f_k^d (f_k^d)^* C\, f_{m+1} = \sum_{k=1}^{d} (\lambda_k^m (f_k^d)^* C\, f_{m+1}) f_k^d = 0.$$

Since this last expression is a linear combination of an orthonormal basis, it occurs only when $\lambda_k^m (f_k^d)^* C\, f_{m+1} = 0$ for all $k \in [d]$. We collect these equations over $m \in [\delta]_0$, considering the definition of J_k and that $g_m \in \mathbb{R}^d$ implies $\lambda_k^m = 0 \iff \lambda_{2-k}^m = 0$ to restate this condition as $\begin{bmatrix} f_k^d & f_{2-k}^d \end{bmatrix}^* C v_{m+1}^D = 0$ for all $k \in [d], m \in J_k$. Since $\operatorname{span}\{f_k^d, f_{2-k}^d\} = \operatorname{span}\{\rho_k^d, \mu_k^d\}$, we further restate this as $\begin{bmatrix} \rho_k^d & \mu_k^d \end{bmatrix}^* C v_{m+1}^D = 0$ for all $k \in [d], m \in J_k$; setting $W_k = C^* \begin{bmatrix} \rho_k^d & \mu_k^d \end{bmatrix} \in \mathbb{R}^{D\times 2}$, we now get that $\mathcal{A}^*(C) = 0 \iff \operatorname{Col}(W_k) \subseteq \{v_{m+1}^D\}_{m\in J_k}^\perp \cap \mathbb{R}^D$ for all $k \in [d]$.

We now claim that $\mathcal{A}^*$ is invertible if and only if the subspaces $\{v_{m+1}^D\}_{m\in J_k}^\perp \cap \mathbb{R}^D$ are all trivial. Indeed, if we fix a k and have some non-zero $u \in \{v_{m+1}^D\}_{m\in J_k}^\perp \cap \mathbb{R}^D$, then we may set $C = \rho_k^d u^*$, such that

$$\operatorname{circ}(g_m) C v_{m+1}^D = (\operatorname{circ}(g_m)\rho_k^d)(u^* v_{m+1}^D).$$

For $m \in J_k, u^* v_{m+1}^D = 0$ by hypothesis on u, and for $m \notin J_k, \mathrm{circ}(g_m)\rho_k^d = 0$ by definition of J_k and lemma 8.

For the other direction, assume $\{v_{m+1}^D\}_{m\in J_k}^{\perp} \cap \mathbb{R}^D = 0$ for each $k \in [d]$. Then $\mathcal{A}^*(C) = 0 \iff \mathrm{Col}(W_k) = \{0\} \iff W_k = 0$ for all k. However, $\{\rho_k^d\}_{k\in[d]} \cup \{\mu_k^d\}_{k\in[d]\setminus\{1,\frac{d}{2}+1\}}$ is an orthogonal basis for $\mathbb{R}^d$, so

$$\begin{aligned} & W_k = 0 \quad \text{for all } k \in [d] \\ \iff \quad & C^*\rho_k^d = C^*\mu_k^d = 0 \quad \text{for all } k \in [d] \\ \iff \quad & C = 0 \end{aligned}$$

We complete the proof by considering that, for $u \in \mathbb{R}^D, \langle v_j^D, u\rangle = 0$ if and only if $\langle \rho_j^D, v\rangle = \langle \mu_j, v\rangle = 0$, so

$$\{v_{m+1}\}_{m\in J_k}^{\perp} \cap \mathbb{R}^D = \{\rho_{m+1}, \mu_{m+1}\}_{m\in J_k}^{\perp}$$

which has dimension $\max\{D - (2|J_k| - \mathbb{1}_{0\in J_k}), 0\}$ by lemma 7. Therefore, $\mathcal{A}^*$ is invertible if and only if $2|J_k| - \mathbb{1}_{0\in J_k} \leq D$ for all $k \in [d]$, as claimed. □

Remark. It turns out that this condition is generic, in the sense that it fails to hold only on a subset of $\mathbb{R}^d$ with Lebesgue measure zero. We consider that the set of $\gamma \in \mathbb{R}^d$ giving at least one zero in $F(\gamma \circ S^{-m}\gamma)$ is a finite union of zero sets of non-trivial quadratic polynomials (except when $2 \mid d, \delta \geq d/2$, and $m = d/2$, discussed below) and hence a set of zero measure; therefore, $J_k = [\delta]_0$ for all γ outside a set of measure zero and B_γ is linearly independent under generic conditions.

To address the case of $m = d/2$, we first remark that this is the only possible exception: indeed, when $m \neq d/2$, we have that

$$F((e_1 + e_{m+1}) \circ S^m(e_1 + e_{m+1}))_k = f_k^* e_{m+1} = \omega^{m(k-1)},$$

so $\gamma \to F(\gamma \circ S^m\gamma)_k$ is a non-zero, homogeneous quadratic polynomial and therefore has a zero locus of measure zero.

However, when $d = 2m$, then $\gamma \circ S^m \gamma$ is periodic with period m and $F(\gamma \circ S^m \gamma)_{2i} = 0$ for $i \in [m]_0$. In particular, if $\delta \geq m$, then $D = d$ and $m \notin J_{2i}$ for all $i \in [m]_0$ for any γ. In particular, $|J_2| \leq \delta - 1$ and $2|J_2| - \mathbb{1}_{0 \in J_2} \leq 2\delta - 3$, so if $\delta \in \{d/2, d/2+1\}$, all choices of γ automatically fail to produce a spanning family.

This exception is quite pathological, though: since our intention is to have $\delta \ll d$, this will rarely be an impediment. Nonetheless, in the case that you *do* want to have $\operatorname{span} B_\gamma = \mathcal{H}^d$, then taking $\delta > d/2 + 1$ gives some space for the condition $2|J_k| - \mathbb{1}_{0 \in J_k}$, and we again have that generic γ will produce spanning families.

4.2 Condition Number

Now that we have characterized this collection of spanning families, we are interested in the condition number for solving the linear system $y = \mathcal{A}(T_\delta(xx^*)) + \eta$ to estimate $T_\delta(xx^*)$. We begin by introducing the main result of this section:

Proposition 2. *Let $\{m_j\}_{j=1}^D \subseteq \mathbb{C}^d$ be a local Fourier measurement system with support δ, mask γ, and modulation index K, where $D = \min(d, 2\delta - 1)$. Let $\mathcal{A}$ be the associated measurement operator as in* (4.1)*, with canonical matrix representation A as in* (4.2)*.*

If we additionally assume that $2\delta - 1 \leq d$ and $K = 2\delta - 1$, then the condition number of $\mathcal{A}$ is

$$\kappa(\mathcal{A}) = \frac{\max\limits_{m \in [\delta]_0, j \in [d]} |F_d(\gamma \circ S^{-m}\gamma)_j|}{\min\limits_{m \in [\delta]_0, j \in [d]} |F_d(\gamma \circ S^{-m}\gamma)_j|}. \tag{4.11}$$

Otherwise, if $2\delta - 1 > d$ or $K > 2\delta - 1$, we may bound the condition number by

$$\kappa(\mathcal{A}) \leq \frac{\max\limits_{m \in [\delta]_0, j \in [d]} |F_d(\gamma \circ S^{-m}\gamma)_j|}{\min\limits_{m \in [\delta]_0, j \in [d]} |F_d(\gamma \circ S^{-m}\gamma)_j|} \kappa(\widetilde{F}_K), \tag{4.12}$$

where $\widetilde{F}_K \in \mathbb{C}^{D \times D}$ is the $D \times D$ principal submatrix of F_K.

4.2.1 Interleaving operators and circulant structure

To set the stage for the proof, we introduce a certain collection of permutation operators and study their interactions with circulant and block-circulant matrices. The structure we identify here will be of much use to us in unraveling the linear systems we encounter in our model for phase retrieval with local correlation measurements.

To this end, we introduce the *interleaving operators* $P^{(d,N)} : \mathbb{C}^{dN} \to \mathbb{C}^{dN}$ for any $d, N \in \mathbb{N}$, each of which is a permutation defined by

$$(P^{(d,N)}v)_{(i-1)N+j} = v_{(j-1)d+i}. \tag{4.13}$$

We can view this is beginning with $v \in C^{dN}$ written as N blocks of d entries, and interleaving them into d blocks each of N entries. Additionally, for $\ell, N_1, N_2 \in \mathbb{N}, v \in \mathbb{C}^{\ell N_1}, k \in [\ell]$, and $H \in \mathbb{C}^{\ell N_1 \times N_2}$, we define the block circulant operator circ^{N_1} by

$$\mathrm{circ}_k^{N_1}(v) = \begin{bmatrix} v & S^{N_1}v & \cdots & S^{(k-1)N_1}v \end{bmatrix}$$
$$\mathrm{circ}_k^{N_1}(H) = \begin{bmatrix} H & S^{N_1}H & \cdots & S^{(k-1)N_1}H \end{bmatrix},$$

where, as with $\mathrm{circ}(\cdot)$, when we omit the subscript we define $\mathrm{circ}^{N_1}(H) = \mathrm{circ}_\ell^{N_1}(H)$ and $\mathrm{circ}^{N_1}(v) = \mathrm{circ}_\ell^{N_1}(v)$. We now proceed with the following lemmas; the first establishes the inverse of $P^{(d,N)}$.

Lemma 9. *For $d, N \in \mathbb{N}$, we have*

$$(P^{(d,N)})^{-1} = P^{(d,N)*} = P^{(N,d)}.$$

Proof of lemma 9. To prove the statement, we simply take $v \in \mathbb{C}^{dN}$ and calculate,

for $i \in [d], j \in [N]$,

$$\begin{aligned}(P^{(d,N)}P^{(N,d)}v)_{(i-1)N+j} &= (P^{(d,N)}(P^{(N,d)}v))_{(i-1)N+j} \\ &= (P^{(N,d)}v)_{(j-1)d+i} \\ &= v_{(i-1)N+j},\end{aligned}$$

with these equalities coming from the definition in (4.13). □

We now observe some useful ways in which the interleaving operators commute with the construction of circulant matrices.

Lemma 10. *Suppose* $v_i, v_{ij} \in \mathbb{C}^k, w_j \in \mathbb{C}^{kN_1}$ *for* $i \in [N_1], j \in [N_2]$ *and*

$$M_1 = \begin{bmatrix} \operatorname{circ}(v_1) \\ \vdots \\ \operatorname{circ}(v_{N_1}) \end{bmatrix}, \quad M_2 = \begin{bmatrix} \operatorname{circ}^{N_1}(w_1) & \cdots & \operatorname{circ}^{N_1}(w_{N_2}) \end{bmatrix}, \text{ and}$$

$$M_3 = \begin{bmatrix} \operatorname{circ}(v_{11}) & \cdots & \operatorname{circ}(v_{1N_2}) \\ \vdots & \ddots & \vdots \\ \operatorname{circ}(v_{N_1 1}) & \cdots & \operatorname{circ}(v_{N_1 N_2}) \end{bmatrix}.$$

Then

$$P^{(k,N_1)}M_1 = \operatorname{circ}^{N_1}\left(P^{(k,N_1)}\begin{bmatrix} v_1 \\ \vdots \\ v_{N_1} \end{bmatrix}\right) \tag{4.14}$$

$$M_2 P^{(k,N_2)*} = \operatorname{circ}^{N_1}\left(\begin{bmatrix} w_1 & \cdots & w_{N_2} \end{bmatrix}\right) \tag{4.15}$$

$$P^{(k,N_1)}M_3P^{(k,N_2)*} = \operatorname{circ}^{N_1}\left(P^{(k,N_1)}\begin{bmatrix} v_{11} & \cdots & v_{1N_2} \\ \vdots & \ddots & \vdots \\ v_{N_1 1} & \cdots & v_{N_1 N_2} \end{bmatrix}\right). \tag{4.16}$$

Proof of lemma 10. We index the matrices to check the equalities. For (4.14), we

have

$$\begin{aligned}(P^{(k,N_1)}M_1)_{(a-1)N_1+b,j} &= (M_1)_{(b-1)k+a,j} \\ &= \begin{bmatrix} S^{j-1}v_1 \\ \vdots \\ S^{j-1}v_{N_1} \end{bmatrix}_{(b-1)k+a} \\ &= (S^{j-1}v_b)_a = (v_b)_{a+j-1}\end{aligned}$$

and

$$\begin{aligned}\mathrm{circ}^{N_1}\left(P^{(k,N_1)}\begin{bmatrix} v_1 \\ \vdots \\ v_{N_1} \end{bmatrix}\right)_{(a-1)N_1+b,j} &= \left(P^{(k,N_1)}\begin{bmatrix} v_1 \\ \vdots \\ v_{N_1} \end{bmatrix}\right)_{(a-1)N_1+b+(j-1)N_1} \\ &= (v_b)_{a+j-1}\end{aligned}$$

For (4.15), we have

$$(P^{(k,N_2)}M_2^*)_{(a-1)N_2+b,j} = (M_2)_{j,(b-1)k+a} = (w_b)_{j+(a-1)N_1}$$

and

$$\left(\mathrm{circ}^{N_1}\left(\begin{bmatrix} w_1 & \cdots & w_{N_2} \end{bmatrix}\right)\right)_{j,(a-1)N_2+b} = (S^{N_1(a-1)}w_b)_j = (w_b)_{j+N_1(a-1)}$$

(4.16) follows immediately by combining (4.14) and (4.15). □

Lemma 11 introduces a few useful identities relating the interleaving operators to kronecker products.

Lemma 11. *For* $v \in \mathbb{C}^N, V = \begin{bmatrix} V_1 & \cdots & V_\ell \end{bmatrix} \in \mathbb{C}^{N\times\ell}, A = \begin{bmatrix} A_1 & \cdots & A_m \end{bmatrix} \in \mathbb{C}^{d\times m}$,

and $B_i \in \mathbb{C}^{m\times k}, i \in [\ell]$*, we have*

$$P^{(d,N)}(v \otimes A) = A \otimes v \tag{4.17}$$

$$P^{(d,N)}(V \otimes A) = \begin{bmatrix} A \otimes V_1 & \cdots & A \otimes V_\ell \end{bmatrix} \tag{4.18}$$

$$P^{(d,N)}(V \otimes A)P^{(\ell,m)} = A \otimes V \tag{4.19}$$

$$(V \otimes A)\begin{bmatrix} B_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_\ell \end{bmatrix} = \begin{bmatrix} V_1 \otimes AB_1 & \cdots & V_\ell \otimes AB_\ell \end{bmatrix} \tag{4.20}$$

Proof of lemma 11. For (4.17), we see that, for $i, j, k \in [d] \times [N] \times [m]$, we have

$$\begin{aligned}(P^{(d,N)}v \otimes A)_{(i-1)N+j,k} &= (v \otimes A)_{(j-1)d+i,k} \\ &= v_j A_{ik}, \text{ while} \\ (A \otimes v)_{(i-1)N+j,k} &= A_{ik}v_j,\end{aligned}$$

and (4.18) follows by considering that $V \otimes A = \begin{bmatrix} V_1 \otimes A & \cdots & V_\ell \otimes A \end{bmatrix}$. To get (4.19), we trace the positions of columns, considering that $(V \otimes A)e_{(i-1)m+j} = V_j \otimes A_i$. From (4.18), we observe that $P^{(d,N)}(V \otimes A)e_{(i-1)m+j} = A_j \otimes V_i$, so

$$\begin{aligned}P^{(d,N)}(V \otimes A)P^{(m,\ell)}e_{(j-1)\ell+i} &= P^{(d,N)}(V \otimes A)e_{(i-1)m+j} \\ &= A_j \otimes V_i = (A \otimes V)e_{(j-1)\ell+i}.\end{aligned}$$

As for (4.20), we remark that

$$(V \otimes A)\begin{bmatrix} B_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_\ell \end{bmatrix} = (V \otimes A)\begin{bmatrix} e_1^\ell \otimes B_1 & \cdots & e_\ell^\ell \otimes B_\ell \end{bmatrix}$$

$$= \begin{bmatrix} (V \otimes A)(e_1^\ell \otimes B_1) & \cdots & (V \otimes A)(e_\ell^\ell \otimes B_\ell) \end{bmatrix} = \begin{bmatrix} V_1 \otimes AB_1 & \cdots & V_\ell \otimes AB_\ell \end{bmatrix},$$

as desired. □

The following lemma is a standard result (e.g., Theorem 13.26 in [52]) regarding the kronecker product.

Lemma 12. *We have* $\mathrm{vec}(ABC) = (C^T \otimes A)\,\mathrm{vec}(B)$ *for any* $A \in \mathbb{C}^{m\times n}, B \in \mathbb{C}^{n\times p}, C \in \mathbb{C}^{p\times k}$.

The next lemma generalizes the diagonalization of circulant matrices, stated in (4.10), for block circulant matrices.

Lemma 13. *Suppose* $V \in C^{kN\times m}$, *then* $\mathrm{circ}^N(V)$ *is block diagonalizable by*

$$\mathrm{circ}^N(V) = (F_k \otimes I_N)\,(\mathrm{diag}(M_1, \ldots, M_k))\,(F_k \otimes I_m)^*\,,$$

where

$$\sqrt{k}\,(F_k \otimes I_N)^*\,V = \begin{bmatrix} M_1 \\ \vdots \\ M_k \end{bmatrix}, \quad \textit{or} \quad M_j = \sqrt{k}(f_j^k \otimes I_N)^* V$$

Proof of lemma 13. We set V_i to be the $k \times m$ blocks of V such that $V^* = \begin{bmatrix} V_1^* & \cdots & V_k^* \end{bmatrix}$ and begin by observing that, for $u \in \mathbb{C}^k$ and $W \in \mathbb{C}^{m\times p}$, the ℓ^{th} $k \times p$ block of $\mathrm{circ}^N(V)(u \otimes W)$ is given by

$$\big(\mathrm{circ}^N(V)(u \otimes W)\big)_\ell = \sum_{i=1}^{k} u_i (S^{N(i-1)}V)_\ell W = \sum_{i=1}^{k} u_i V_{\ell-i+1} W.$$

Taking $u = f_j^k$ and $W = I_m$, this gives

$$\begin{aligned}
\big(\mathrm{circ}^N(V)(f_j^k \otimes I_m)\big)_\ell &= \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \omega_k^{(j-1)(i-1)} V_{\ell-i+1} I_m \\
&= \frac{1}{\sqrt{k}} \omega_k^{(j-1)(\ell-1)} \sum_{i=1}^{k} \omega_k^{-(j-1)(i-1)} V_i \\
&= (f_j^k)_\ell \left(\sqrt{k}(f_j^k \otimes I_N)^* V\right) = (f_j^k)_\ell M_j.
\end{aligned}$$

This relation is equivalent to having

$$\mathrm{circ}^N(V)(f_j^k \otimes I_m) = (f_j^k \otimes M_j) = (f_j^k \otimes I_N) M_j,$$

which is the statement of the lemma. □

Lemma 13 immediately gives the following corollary regarding the conditioning of $\mathrm{circ}^N(V)$, with which we return to considering spanning families of masks.

Corollary 4. *With notation as in lemma 13, the condition number of* $\mathrm{circ}^N(V)$ *is*

$$\frac{\max\limits_{i\in[k]}\sigma_{\max}(M_i)}{\min\limits_{i\in[k]}\sigma_{\min}(M_i)}.$$

4.2.2 Proof of proposition 2

To begin the proof of proposition 2, we apply the results of section 4.2.1 to the case of the measurement operator for a family of masks, as defined in section 4.1.

Proposition 3. *Given a family of masks* $\{m_j\}_{j\in[D]}$ *of support* $\delta \leq \frac{d+1}{2}$*, we define* $g_m^j = \mathrm{diag}(m_j m_j^*, m)$,

$$H = P^{(d,D)} \begin{bmatrix} Rg_{1-\delta}^1 & \cdots & Rg_{\delta-1}^1 \\ \vdots & \ddots & \vdots \\ Rg_{1-\delta}^D & \cdots & Rg_{\delta-1}^D \end{bmatrix},$$

and $M_j = \sqrt{d}\left(f_j^d \otimes I_D\right)^* H$*. Then the condition number of* $\mathcal{A}$ *as defined in* (4.1) *is*

$$\kappa(\mathcal{A}) = \frac{\max\limits_{i\in[d]}\sigma_{\max}(M_i)}{\min\limits_{i\in[d]}\sigma_{\min}(M_i)}.$$

Proof. We consider the rows of the matrix A representing the measurement operator $\mathcal{A}$, defined in (4.2) and (4.1). We vectorize $X \in T_\delta(\mathbb{C}^{d\times d})$ by its diagonals with $\mathcal{D}_\delta$, as in (4.4) and set $\chi_m = \mathrm{diag}(X, m), m = 1-\delta, \ldots, \delta-1$. Each measurement then looks like

$$\begin{aligned}\mathcal{A}(X)_{(\ell,j)} &= \langle S^\ell m_j m_j^* S^{-\ell}, X\rangle \\ &= \sum_{m=1-\delta}^{\delta-1} \langle S^\ell g_m^j, \chi_m\rangle,\end{aligned}$$

so that the definition of A in (4.2) immediately yields its $(j-1)d+\ell^{\text{th}}$ row as $\begin{bmatrix} g_{1-\delta}^{j*}S^{1-\ell} & \cdots & g_{\delta-1}^{j*}S^{1-\ell}\end{bmatrix}$. Transposing this expression, we see that the $(j-1)d+1^{\text{st}}$ through $(j-1)d+d^{\text{th}}$ rows of A compose $\begin{bmatrix}\text{circ}(g_{1-\delta}^{j})^* & \cdots & \text{circ}(g_{\delta-1}^{j})^*\end{bmatrix}$. Together with $\text{circ}(v)^* = \text{circ}(Rv)$, this determines A to be the block matrix given by

$$A = \begin{bmatrix}\text{circ}(g_1^{1-\delta})^* & \cdots & \text{circ}(g_1^{\delta-1})^* \\ \vdots & \ddots & \vdots \\ \text{circ}(g_D^{1-\delta})^* & \cdots & \text{circ}(g_D^{\delta-1})^*\end{bmatrix} = \begin{bmatrix}\text{circ}(Rg_1^{1-\delta}) & \cdots & \text{circ}(Rg_1^{\delta-1}) \\ \vdots & \ddots & \vdots \\ \text{circ}(Rg_D^{1-\delta}) & \cdots & \text{circ}(Rg_D^{\delta-1})\end{bmatrix},$$

which may be transformed, by (4.16) of lemma 10, to

$$P^{(d,D)}AP^{(d,2\delta-1)*} = \text{circ}^D\left(P^{(d,D)}\begin{bmatrix} Rg_{1-\delta}^1 & \cdots & Rg_{\delta-1}^1 \\ \vdots & \ddots & \vdots \\ Rg_{1-\delta}^D & \cdots & Rg_{\delta-1}^D\end{bmatrix}\right) = \text{circ}^D(H). \quad (4.21)$$

Quoting corollary 4 establishes the proposition. □

We are now able to prove proposition 2.

Proof of Proposition 2. For the moment, we assert that $D = 2\delta - 1 \leq d$ and set $\widetilde{F}_K \in \mathbb{C}^{2\delta-1\times 2\delta-1}, (\widetilde{F}_K)_{ij} = \frac{1}{\sqrt{K}}\omega_K^{(i-1)(j-\delta)}$ to be the principal submatrix of $\sqrt{K}\,\text{diag}(f_{2-\delta}^K)F_K$. In this case, $g_m^j = \text{diag}(m_j m_j^*, m) = \omega_K^{m(j-1)}g_m$, as in (4.9). Therefore, we label the $2\delta-1\times 2\delta-1$ blocks of H by $H^* = \begin{bmatrix} H_1^* & \cdots & H_d^*\end{bmatrix}$, so that

$$(H_\ell)_{ij} = (Rg_{j-\delta}^i)_\ell = \omega_K^{(i-1)(j-\delta)}(Rg_{j-\delta})_\ell$$

and $M_\ell = \sum_{k=1}^d \omega_d^{(\ell-1)(k-1)}H_k$, giving

$$\begin{aligned}(M_\ell)_{ij} &= \sum_{k=1}^d \omega_d^{(\ell-1)(k-1)}(H_k)_{ij} = \omega_K^{(i-1)(j-\delta)}\sum_{k=1}^d \omega_d^{(\ell-1)(k-1)}(Rg_{j-\delta})_k \\ &= \omega_K^{(i-1)(j-\delta)}(F_d^* g_{j-\delta})_\ell.\end{aligned}$$

In other words,

$$M_\ell = \sqrt{K}\widetilde{F}_K \operatorname{diag}(f_\ell^{d*} g_{1-\delta}, \ldots, f_\ell^{d*} g_{\delta-1}). \tag{4.22}$$

If $K = 2\delta - 1$, then $\widetilde{F}_K$ is unitary, and the singular values of M_ℓ are $\{\sqrt{K} f_\ell^{d*} g_j\}_{j=1-\delta}^{\delta-1}$. Recognizing that $S^j g_j = g_{-j}$, then proposition 3 takes us to (4.11).

If $D = 2\delta - 1 < K$, then the argument remains unchanged, except that the singular values of M_ℓ, instead of being known explicitly, are bounded above and below by $\max_{|j|<\delta} |f_\ell^{d*} g_j| \sigma_{\max}(\widetilde{F}_K)$ and $\min_{|j|<\delta} |f_\ell^{d*} g_j| \sigma_{\min}(\widetilde{F}_K)$ respectively, which gives the more general result of (4.12).

If $2\delta - 1 > d = D$, then, by considering that the argument of proposition 3 depended only on the indices of the diagonals under consideration. By a similar argument, however, we may obtain

$$A = \begin{bmatrix} \operatorname{circ}(Rg_0^1) & \cdots & \operatorname{circ}(Rg_0^d) \\ \vdots & \ddots & \vdots \\ \operatorname{circ}(Rg_{d-1}^1) & \cdots & \operatorname{circ}(Rg_{d-1}^d) \end{bmatrix},$$

and a similar application of (4.16) gives that

$$P^{(d,d)} A P^{(d,2\delta-1)*} = \operatorname{circ}^d \left(P^{(d,d)} \begin{bmatrix} Rg_0^1 & \cdots & Rg_0^d \\ \vdots & \ddots & \vdots \\ Rg_{d-1}^1 & \cdots & Rg_{d-1}^d \end{bmatrix} \right).$$

Setting

$$H = P^{(d,d)} \begin{bmatrix} Rg_0^1 & \cdots & Rg_0^d \\ \vdots & \ddots & \vdots \\ Rg_{d-1}^1 & \cdots & Rg_{d-1}^d \end{bmatrix},$$

and defining $H_\ell \in \mathbb{C}^{d\times d}$ by $H^* = \begin{bmatrix} H_1^* & \cdots & H_d^* \end{bmatrix}$, we have

$$(H_\ell)_{ij} = \omega_K^{(i-1)(j-1)} (Rg_{j-1})_\ell \quad \text{and} \quad (M_\ell)_{ij} = \omega_K^{(i-1)(j-1)} (F_d^* g_{j-1})_\ell,$$

giving $M_\ell = \sqrt{K}\mathcal{R}_{d\times d}(F_K) \operatorname{diag}(f_\ell^{d*} g_0, \ldots, f_\ell^{d*} g_{d-1})$, which immediately gives us (4.12). We remark that indexing only over the diagonals $m \in [\delta]_0$ in (4.12) suffices, again because $S^j g_j = g_{-j}$, so having $2\delta - 1 > d$ makes $1-\delta, \ldots, -1$ redundant.

□

4.3 Inverting $\mathcal{A}$

In this section, we use the results of section 4.2 to explicitly state the inverse of the measurement operator $\mathcal{A}$, as well as the computational complexity of calculating its inverse. Additionally, we calculate the variance in *each entry* of $\mathcal{A}^{-1}(y)$ when $y = \mathcal{A}(T_\delta(xx^*)) + \eta$ is produced under an i.i.d. Gaussian noise model, which will be useful to us in later analysis.

4.3.1 Explicit inverse of $\mathcal{A}$

We begin by fixing a local Fourier measurement system $\{m_j\}_{j=1}^{2\delta-1}$ with support $\delta \leq \frac{d+1}{2}$, mask γ, and modulation index $K = D = 2\delta - 1$. We take $\mathcal{A}$ to be the associated measurement operator and A its canonical matrix representation as in eqs. (4.1) and (4.2). We then remark from (4.21) and lemma 13 that

$$A = P^{(D,d)}(F_d \otimes I_D)\,\mathrm{diag}(M_\ell)_{\ell=1}^{d}(F_d \otimes I_D)^* P^{(d,D)},$$

where we recall M_ℓ from (4.22). Defining $Z \in \mathbb{C}^{D\times d}$ by

$$Z_{m\ell} = \sqrt{D} f_\ell^{d*} g_{m-\delta} \tag{4.23}$$

and setting $z_\ell = Ze_\ell$, we have

$$M_\ell = \widetilde{F}_D D_{z_\ell} \text{ and } \mathrm{diag}(M_\ell)_{\ell=1}^{d} = (I_d \otimes \widetilde{F}_D)\,\mathrm{diag}(\mathrm{vec}(Z)), \tag{4.24}$$

which gives

$$A = P^{(D,d)}(F_d \otimes I_D)(I_d \otimes \widetilde{F}_D)\,\mathrm{diag}(\mathrm{vec}(Z))(F_d \otimes I_D)^* P^{(d,D)}.$$

This immediately produces the inverse of A, which we state in proposition 4.

Proposition 4. *Let $A \in \mathbb{C}^{d\times 2\delta-1}$ be the canonical representation of the measurement operator $\mathcal{A}$ associated with a local Fourier measurement system $\{m_j\}_{j=1}^{d}$ of support $\delta \leq \frac{d+1}{2}$ with mask $\gamma \in \mathbb{R}^d$. Defining Z as in (4.23), we have*

$$A^{-1} = P^{(D,d)}(F_d \otimes I_D)(\mathrm{diag}(\mathrm{vec}(Z)))^{-1}(I_d \otimes \widetilde{F}_D^*)(F_d \otimes I_D)^* P^{(d,D)}. \tag{4.25}$$

This formulation makes it straightforward to deduce the computational complexity, as previously stated in section 3.1.3. Namely, each permutation $P^{(D,d)}$ requires $\mathcal{O}(dD)$ operations to run on a vector. Since, by (4.19) of lemma 11, we have that $F_d \otimes I_D = P^{(d,D)}(I_D \otimes F_d)P^{(D,d)}$, and considering also that $I_D \otimes F_d$ comprises D Fourier transforms of dimension d, multiplication by $F_d \otimes I_D$ costs $\mathcal{O}(dD + Dd\log d) = \mathcal{O}(dD\log d)$ operations. Since $\widetilde{F}_D = F_D S^{1-\delta}$, multiplying by $I_d \otimes \widetilde{F}_D^*$ takes $\mathcal{O}(dD\log D)$ operations. Finally, multiplying by $\mathrm{diag}(\mathrm{vec}(Z))^{-1}$ trivially has a cost of $\mathcal{O}(dD)$. Putting all these considerations together, and recalling that $d \geq D = 2\delta - 1$, the cost of inverting A comes out to

$$\mathcal{O}(dD)+\mathcal{O}(dD\log d)+\mathcal{O}(dD\log D)+\mathcal{O}(dD)+\mathcal{O}(dD\log d)+\mathcal{O}(dD) = \mathcal{O}(dD\log d),$$

or $\mathcal{O}(\delta\, d\log d)$, as concluded in section 3.1.3 and [46].

4.3.2 Preliminaries in Probability

Prior to discussing the variance of the images of vectors under A^{-1} or $\mathcal{A}^{-1}$, we review some preliminaries regarding the real and complex multivariate Gaussian distributions. These results and the notation with which we express them may be found in many standard texts, for example [74] for the real case and section 7.8.1 of [34] for the complex. For $\mu \in \mathbb{R}^n$ and $0 \prec \Sigma \in \mathcal{S}^n, \mathcal{N}(\mu, \Sigma)$ refers to the multivariate normal distribution on $\mathbb{R}^n$ with mean μ and covariance matrix

$\Sigma = \mathbb{E}_x[(x-\mu)(x-\mu)^T]$, and is determined by its probability density function

$$f_{\mathcal{N}}(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right).$$

For $\nu \in \mathbb{C}^n$ and $0 \prec \Xi \in \mathcal{H}^n, \mathcal{CN}(\nu,\Xi)$ is the (circularly symmetric about ν) multivariate complex normal distribution on $\mathbb{C}^n$ with mean ν, covariance $\Xi = \mathbb{E}_z[(z-\nu)(z-\nu)^*]$, and density function

$$f_{\mathcal{CN}}(z;\nu,\Xi) = \frac{1}{\pi^n \det(\Xi)} \exp(-(z-\nu)^*\Xi^{-1}(z-\nu))$$

We remark that circular symmetry is defined by having that the real and imaginary parts of $z \sim \mathcal{CN}(0,\Xi)$ be i.i.d., and is ensured by tacitly requiring, as we shall throughout this dissertation, that $\mathbb{E}[(z-\nu)(z-\nu)^T] = 0$ (see Theorem 7.8.1 of [34]).

We now relate a standard result from the literature concerning linear transformations of Gaussian random vectors.

Proposition 5 (Theorem 3.3.3 in [74] and Section 7.8.1 of [34])**.** *Suppose* $x \sim \mathcal{N}(\mu,\Sigma)$*. Then for* $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$*,*

$$Ax + b \sim \mathcal{N}(A\mu + b, A\Sigma A^T). \tag{4.26}$$

Suppose $z \sim \mathcal{CN}(\nu,\Xi)$*. Then for* $B \in \mathbb{C}^{m\times n}, c \in \mathbb{C}^m$*,*

$$Bz + c \sim \mathcal{CN}(B\nu + c, B\Xi B^*). \tag{4.27}$$

We remark that the result regarding complex Gaussian vectors implies that linear transformations preserve the property of circular symmetry.

4.3.3 Distribution of variance

In the interest of studying the propagation of error through our recovery algorithm, in this section, we describe the probability distribution of the noise on each

entry of $\mathcal{A}^{-1}(\mathcal{A}(xx^*)+\eta)$ as a function of the noise vector's distribution. To keep things tractable, we assume that the entries of η are identically and independently distributed; specifically, we will assume that $\eta_{(\ell,j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2)$ for some $\sigma \geq 0$.

Before we begin, we remark that the distribution of $\mathcal{A}^{-1}(\eta)_{i,i+m}$ will depend only on m. This follows intuitively by noting that $\mathcal{A}$ commutes nicely with "diagonal shifts" in the sense that

$$\mathcal{A}(S^k X S^{-k})_{(\ell,j)} = \langle S^\ell m_j m_j^* S^{-\ell}, S^k X S^{-k}\rangle = \langle S^{\ell-k} m_j m_j^* S^{-(\ell-k)}, X\rangle = \mathcal{A}(X)_{(\ell-k,j)},$$

so that if $y_1, y_2 \in \mathbb{R}^{[d]\times[D]}$ satisfy $(y_1)_{\ell,j} = (y_2)_{\ell-k,j}$ for some $k \in \mathbb{N}$, we will have $\mathcal{A}^{-1}(y_1) = S^k \mathcal{A}^{-1}(y_2) S^{-k}$. In particular, if the entries of y_1 are identically, independently distributed random variables and y_2 is *defined* by $(y_2)_{\ell-k,j} = (y_1)_{\ell,j}$, then y_1 and y_2 are drawn from the same distribution on $\mathbb{R}^{[d]\times[D]}$. This means that $\mathcal{A}^{-1}(y_1)$ and $\mathcal{A}^{-1}(y_2)$ are identically distributed, but $\mathcal{A}^{-1}(y_1) = S^k \mathcal{A}^{-1}(y_2) S^{-k}$, so the distributions of $\mathcal{A}^{-1}(y_1)$ and $S^k \mathcal{A}^{-1}(y_1) S^{-k}$ are identical. In particular, the distribution of the image of i.i.d. noise under $\mathcal{A}^{-1}$ is invariant under such diagonal shifts, so $\mathcal{A}^{-1}(\eta)_{i,i+m}$ is distributed identically to (though not necessarily independently from!) $\mathcal{A}^{-1}(\eta)_{1,1+m}$, and this conclusion holds for all i and m.

To make this precise, and to discover the distribution of $\mathcal{A}^{-1}(\eta)_{1,1+m}$ exactly, we will present another means by which $\mathcal{A}^{-1}(y)$ may be calculated from y. With this in mind, we remark that (4.18) of lemma 11 gives that

$$P^{(D,d)}(F_d \otimes I_D) = \begin{bmatrix} I_D \otimes f_1^d & \cdots & I_D \otimes f_d^d \end{bmatrix},$$

so A may be transformed by (4.20) to give

$$A = \begin{bmatrix} M_1 \otimes f_1^d & \cdots & M_d \otimes f_d^d \end{bmatrix} \begin{bmatrix} I_D \otimes f_1^d \\ \vdots \\ I_D \otimes f_d^d \end{bmatrix} = \sum_{j=1}^{d} M_j \otimes f_j^d f_j^{d*}.$$

Setting $X = \begin{bmatrix} \chi_{1-\delta} & \cdots & \chi_{\delta-1} \end{bmatrix} \in \mathbb{C}^{d \times 2\delta-1}$, lemma 12 gives us

$$A \begin{bmatrix} \chi_{1-\delta} \\ \vdots \\ \chi_{\delta-1} \end{bmatrix} = \operatorname{vec}\left(\sum_{j=1}^{d} f_j^d f_j^{d*} X M_j^T \right).$$

Recalling (4.24) along with $\widetilde{F}_D^{-T} = (\widetilde{F}_D^T)^* = \overline{\widetilde{F}}_D$, we have

$$f_j^{d*} \operatorname{mat}_{(d,D)} \left(A \begin{bmatrix} \chi_{1-\delta} \\ \vdots \\ \chi_{\delta-1} \end{bmatrix} \right) \overline{\widetilde{F}}_D = f_j^{d*} X D_{z_j},$$

so that, for $\ell \in [2\delta - 1]$, and recalling $\overline{\widetilde{F}}_D e_\ell = \overline{f}^D_{\ell+1-\delta} = f^D_{\delta+1-\ell}$, we have

$$f_j^{d*} \operatorname{mat}_{(d,D)}(A \operatorname{vec}(X)) f^D_{\delta+1-\ell} = f_j^{d*} \operatorname{mat}_{(d,D)}(A \operatorname{vec}(X)) \overline{\widetilde{F}}_D e_\ell = f_j^{d*} \chi_{\ell-\delta} Z_{\ell j}.$$

In this way, from $A \operatorname{vec}(X)$ we may recover

$$b_\ell := F_d^* \operatorname{mat}_{(d,D)}(A \operatorname{vec}(X)) f^D_{\delta+1-\ell} = \operatorname{vec}(f_j^{d*} Z_{\ell j} \chi_{\ell-\delta})_{j=1}^d = D_{Z^T e_\ell} F_d^* \chi_{\ell-\delta}$$

for each ℓ, from which $\chi_{\ell-\delta}$ is determined by taking $\chi_{\ell-\delta} = F_d D^{-1}_{Z^T e_\ell} b_\ell$. In other words, for $y \in \mathbb{R}^{dD}$, the m^{th} diagonal of $\mathcal{D}_\delta^{-1}(A^{-1}y)$, for $m = 1-\delta, \ldots, \delta - 1$, is given by

$$\chi_m = F_d D^{-1}_{Z^T e_{m+\delta}} F_d^* \operatorname{mat}_{(d,D)}(y) f^D_{1-m}, \tag{4.28}$$

and we use this expression to deduce the distribution of noise on the m^{th} diagonal when y is a random variable. For instance, if we consider that

$$\mathcal{D}_\delta^{-1} A^{-1}(A \mathcal{D}_\delta(T_\delta(xx^*)) + \eta) = T_\delta(xx^*) + \mathcal{D}_\delta^{-1}(A^{-1}\eta),$$

where $\eta_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ for $j \in [d(2\delta - 1)]$, knowing the distribution of $A^{-1}\eta$ will tell us the distribution of noise on our recovered estimate of $T_\delta(xx^*)$. We deduce this result directly from (4.28), and summarize it in proposition 6. We remark that, in the statement and proof of this proposition, exponentiation of vectors is entry-wise.

Proposition 6. *Suppose that* $\{m_j\}_{j=1}^{2\delta-1}$ *is a local Fourier measurement system with support* $\delta \leq \frac{d+1}{2}$*, mask* γ*, and modulation index* $K = D = 2\delta - 1$*, with associated measurement operator and representation matrix* $\mathcal{A}$ *and* A*. Suppose further that* $\eta \in \mathbb{R}^{d(2\delta-1)}$ *has entries that are i.i.d. Gaussian random variables, namely* $\eta_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ *for* $j \in [d(2\delta - 1)]$ *and some* $\sigma \geq 0$*. Then, setting* $N = \mathcal{D}_\delta^{-1} A^{-1} \eta \in T_\delta(\mathbb{C}^{d\times d})$*, then* N *is Hermitian, its* $0, \ldots, \delta - 1^{st}$ *diagonals are distributed independently from one another, and the* m^{th} *diagonal of* N *is distributed by*

$$\operatorname{diag}(N, m) \sim \mathcal{CN}(0, \sigma^2 \operatorname{circ}(s_m)), m \in [\delta - 1], \text{ and}$$
$$\operatorname{diag}(N, 0) \sim \mathcal{N}(0, \sigma^2 \operatorname{circ}(s_0)),$$

where $s_m = \frac{1}{D\sqrt{d}} F_d^* (F_d^* g_m)^{-2}$.

Proof of proposition 6. To establish that N is Hermitian, we remark that $\mathbb{C}^{d\times d} = \mathcal{H}^d \oplus \operatorname{Skew}(d)$, where

$$\operatorname{Skew}(d) = \{B \in \mathbb{C}^{d\times d} : B^* = -B\}$$

is the set of skew-Hermitian matrices and where $\oplus$ represents the direct product. In other words, given any $M \in \mathbb{C}^{d\times d}$, we have a unique $H \in \mathcal{H}^d, B \in \operatorname{Skew}(d)$ such that $M = H + B$. We now consider that, if $H \in \mathcal{H}^d$ and $B \in \operatorname{Skew}(d)$, we have that

$$\begin{aligned}\operatorname{Tr}(H^*B) &= \operatorname{Tr}(HB) = \operatorname{Tr}(BH) \\ &= -\operatorname{Tr}(B^*H) = -\overline{\operatorname{Tr}(H^*B)},\end{aligned}$$

meaning that $\operatorname{Tr}(H^*B) \in \mathrm{i}\mathbb{R}$. Additionally, we remark that the Hermitian matrices are a real Hilbert space, so for another $H' \in \mathcal{H}^d$, we have $\operatorname{Re}\langle H', M\rangle = \langle H', H\rangle$ and $\operatorname{Im}\langle H', M\rangle = \langle H', B\rangle/\mathrm{i}$. Therefore, since all the measurement matrices $S^\ell m_j m_j^* S^{-\ell}$ appearing in $\mathcal{A}$ are Hermitian, given $M \in \mathbb{C}^{d\times d}$ decomposed into its Hermitian and skew-Hermitian parts $M = H + B$, we have that $\operatorname{Re}\mathcal{A}(M) = \mathcal{A}(H)$ and

$\operatorname{Im}\mathcal{A}(M) = \mathcal{A}(B)/\mathrm{i}$. In particular, $\mathcal{A}^{-1}(\mathbb{R}^{[d]_0\times[D]}) \subseteq \mathcal{H}^d$, and N, being the inverse image of a real vector η, will be Hermitian.

For convenience, throughout the remainder of the proof we will set $\chi_m = \operatorname{diag}(N, m)$ for $m = 1-\delta, \ldots, \delta - 1$. To prove the independence of $\{\chi_m\}_{m\in[\delta]_0}$, we consider that (4.28) tells us that

$$\chi_m = F_d D^{-1}_{Z^T e_{m+\delta}} F_d^* \left(\operatorname{mat}_{(d,D)}(\eta) f^D_{1-m}\right),$$

and we focus on the term $\operatorname{mat}_{(d,D)}(\eta) f^D_{1-m}$. Considering that $\eta \sim \mathcal{N}(0, \sigma^2 I_{dD})$, we have that the rows $\begin{bmatrix} r_1 & \cdots & r_d \end{bmatrix}^T$ of $\operatorname{mat}_{(d,D)}(\eta)$ are distributed according to $r_i = \operatorname{mat}_{(d,D)}(\eta)^T e_i \sim \mathcal{N}(0, \sigma^2 I_D)$. At this point, it would be convenient if we could merely cite proposition 5 to establish the distribution of $r_i^T f^D_{1-m}$ or indeed $\operatorname{mat}_{(d,D)}(\eta) f^D_{1-m}$, but we remark that we don't have a result for the image of a real Gaussian vector under a complex linear transformation. Therefore, we consider the real and imaginary parts of $\operatorname{mat}_{(d,D)}(\eta) f^D_{1-m}$ separately, setting $v_m = \operatorname{mat}_{(d,D)}(\eta) f^D_{1-m}$ and seeing that

$$(v_m)_i = r_i^T \operatorname{Re}(f^D_{1-m}) + \mathrm{i}\, r_i^T \operatorname{Im}(f^D_{1-m}) = \operatorname{Re}(f^D_{1-m})^T r_i + \mathrm{i}\, \operatorname{Im}(f^D_{1-m})^T r_i.$$

Since $\{f_1^D\} \cup \{\sqrt{2}\operatorname{Re}(f^D_{1-m})\}_{m\in[\delta-1]} \cup \{\sqrt{2}\operatorname{Im}(f^D_{1-m})\}_{m\in[\delta-1]}$ is an orthonormal basis for $\mathbb{R}^D$, then compiling these vectors into a matrix Q, we have from proposition 5 that $Q^T r_i \sim \mathcal{N}(0, \sigma^2 Q^T Q) = \mathcal{N}(0, \sigma^2 I_D)$, meaning the real and imaginary components of $v_1, \ldots, v_{\delta-1}$ and v_0 are all independent, with $v_0 \sim \mathcal{N}(0, \sigma^2 I_d), \operatorname{Re}(v_m) \sim \mathcal{N}(0, \frac{\sigma^2}{2} I_d)$, and $\operatorname{Im}(v_m) \sim \mathcal{N}(0, \frac{\sigma^2}{2} I_d)$ for $m \in [\delta-1]$. Therefore, $v_m \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, \sigma^2 I_d)$ and $v_0 \sim \mathcal{N}(0, \sigma^2 I_d)$, independently from the other v_m. Since the v_m are independent, clearly their images under non-singular, fixed (independently of the random process) linear transformations will also be independent. In particular, the diagonals $\chi_m = F_d D^{-1}_{Z^T e_{m+\delta}} F_d^* v_m$ will be independent of one another for $m \in [\delta]_0$.

To get the actual distribution of the χ_m for $m \in [\delta - 1]$, we simply quote proposi-

tion 5 again to get that

$$\chi_m \sim \mathcal{CN}(0, \sigma^2 F_d D^{-1}_{Z^T e_{m+\delta}} F_d^* I_d F_d D^{-1}_{Z^T e_{m+\delta}} F_d^*) = \mathcal{CN}(0, \sigma^2 F_d D^{-2}_{Z^T e_{m+\delta}} F_d^*).$$

The covariance matrix becomes, by recalling (4.10) and the definition of Z in (4.23),

$$\sigma^2 F_d D^{-2}_{Z^T e_{m+\delta}} F_d^* = \operatorname{circ}\left(d^{-1/2} F_d^* (Z^T e_{m+\delta})^{-2}\right) = \operatorname{circ}(s_m).$$

The only distinction for χ_0 is that the distributions are all $\mathcal{N}$ instead of $\mathcal{CN}$; the same calculations as above give $\chi_0 \sim \mathcal{N}(0, \sigma^2 \operatorname{circ}(s_0))$. To verify that $s_0 \in \mathbb{R}^d$, we consider that $g_0 = \gamma \circ \gamma$, so $\tilde{g} := F_d^* g_0 = (F_d^* \gamma) * (F_d^* \gamma)$ satisfies $\tilde{g}_i = \tilde{g}_{2-i}$ (equivalently, $R\tilde{g} = \tilde{g}$). Similarly, $R(\tilde{g}^{-2}) = \tilde{g}^{-2}$, which guarantees that $s_0 = \frac{1}{D\sqrt{d}} F_d^* \tilde{g}^{-2} \in \mathbb{R}^d$. □

4.4 Examples of Spanning Families

Example

4.5 Numerical Analysis

Charts and crafts.

Chapter 5

Ptychographic Model

In our model for the ptychographic setup of (2.1), we assume that measurements are taken corresponding to all shifts $\ell \in [d]_0$. Unfortunately, in practice, this is usually an impossibility, since in many cases an illumination of the sample can cause damage to the sample, and applying the illumination beam (which can be highly irradiative) repeatedly at a single point can destory it. In usual ptychography, the beam is shifted by a far larger distance than the width of a single pixel – instead of overlapping on $\delta - 1$ of δ pixels, adjacent illumination regions will typically overlap on a percentage of their support on the order of 50% or even less. Considering the risks to the sample and the costs of operating the measurement equipment, there are strong incentives to reduce the number of illuminations applied to any object, and therefore our theory ought to address a model that reflects this concern.

In particular, instead of taking all d shifts in $[d]$, we hope to use only $\ell \in k[d/k]_0$, where k is an integer divisor of d.

5.1 Spanning Masks and Conditioning

In the case of ptychography, instead of using all shifts in our lifted measurement system, we instead fix a shift size $s \in \mathbb{N}$ where $d = \overline{d}s$ with $\overline{d} \in \mathbb{N}$ and use $S^{s\ell} m_j m_j^* S^{-s\ell}$ for $\ell \in [\overline{d}]$. Therefore, we introduce the following generalization of the lifted measurement system.

Definition 6. Given a family of masks in $\{m_j\}_{j\in[D]} \subseteq \mathbb{C}^d$ and $s, \overline{d} \in \mathbb{N}$ with $\overline{d} = d/s$, the associated *lifted measurement system of shift* s is the set $\mathcal{L}^s_{\{m_j\}} := \{S^{s\ell} m_j m_j^* S^{-s\ell}\}_{(\ell,j)\in[\overline{d}]\times[D]} \subseteq \mathbb{C}^{d\times d}$.

Of course, with a shift size $s > 1$, it is impossible for $\mathcal{L}^s$ to span $T_\delta(\mathbb{C}^{d\times d})$, so we consider the analagous subspace. We will define $\mathcal{J}_{\delta,s} = \bigcup_{\ell\in[\overline{d}]_0} \operatorname{supp}(S^{s\ell}\mathbb{1}\mathbb{1}^* S^{-s\ell})$ to be the set of indices "reached" by this system, and

$$T^s_\delta(X) = \begin{cases} X_{ij}, & (i,j) \in \mathcal{J}_{\delta,s} \\ 0, & \text{otherwise} \end{cases}$$

to be the projection onto the associated subspace of $\mathbb{C}^{d\times d}$. Namely, we observe that

$$\left(S^{s\ell} m_k m_k^* S^{-s\ell}\right)_{ij} = (S^{s\ell} m_k)_i (\overline{S^{s\ell} m_k})_j = (m_k)_{i-s\ell} (\overline{m_k})_{j-s\ell},$$

so $\left(S^{s\ell} m_k m_k^* S^{-s\ell}\right)_{ij} = 0$ when $(i - s\ell, j - s\ell) \notin [\delta]^2$, i.e. when $(i,j) \notin [\delta]^2_{s\ell+1}$. Hence the indices onto which we are projecting are those in $\bigcup_{\ell\in[\overline{d}]_0} [\delta]^2_{s\ell+1}$. This set may be revisualized by calculating which j's are admissible for each i; for a fixed i, we look at all shifts ℓ such that $i \in [\delta]_{s\ell+1}$, and j is allowed to be in their union.

In the (pathological) case where $s \geq \delta$, obviously any given index can only appear in one of the $[\delta]_{s\ell+1}$, namely $i \in [\delta]_{s\ell+1}$ iff $i \bmod s \leq \delta$ and $\lfloor i/s \rfloor = \ell$, so in this case we would have

$$\mathcal{J}_{\delta,s} = \{(i,j) : \lfloor i/s \rfloor = \lfloor j/s \rfloor \text{ and } i \bmod s, j \bmod s \leq \delta\}.$$

However, this case is not typical, since $T_{\delta,s}(\mathbb{1}\mathbb{1}^*)$ will be the adjacency matrix of an unconnected graph, and there will be groups of vertices whose relative phases

are completely undetermined by $T_{\delta,s}(xx^*)$. For example, for any $\theta \in \mathbb{R}$, we have

$$T_{2,2}\left(\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}^*\right) = T_{2,2}\left(\begin{bmatrix}1\\1\\\mathrm{e}^{\mathrm{i}\theta}\\\mathrm{e}^{\mathrm{i}\theta}\end{bmatrix}\begin{bmatrix}1\\1\\\mathrm{e}^{\mathrm{i}\theta}\\\mathrm{e}^{\mathrm{i}\theta}\end{bmatrix}^*\right) = \begin{bmatrix}1&1&0&0\\1&1&0&0\\0&0&1&1\\0&0&1&1\end{bmatrix}$$

In the ordinary case, where $s < \delta$, it is clear that we need only consider the first and last shifts that cover i, namely $i \in [\delta]_{1+s\ell}$ iff $\lceil \frac{i-\delta}{s} \rceil \le \ell \le \lceil \frac{i-s}{s} \rceil$, and therefore

$$\begin{aligned}\mathcal{J}_{\delta,s} &= \left\{(i,j) : j \in \{\left\lceil \frac{i-\delta}{s} \right\rceil s+1, \cdots, \left\lceil \frac{i-\delta}{s} \right\rceil + \delta\}\right.\\ &\qquad \left. \cup \{\left\lceil \frac{i-s}{s} \right\rceil s+1, \cdots, \left\lceil \frac{i-s}{s} \right\rceil s+\delta\}\right\}\\ &= \left\{(i,j) : j = \left\lceil \frac{i-\delta}{s} \right\rceil s+1, \ldots, \left\lceil \frac{i-s}{s} \right\rceil s+\delta\right\}\end{aligned}$$

Unfortunately, this formulation is not particularly transparent, but we mention an important special case. When s is also a divisor of δ, say $\delta = \overline{\delta}s$, then this condition becomes

$$\begin{aligned}(i,j) \in \mathcal{J}_{\delta,s} &\iff \left(\left\lceil \frac{i}{s} \right\rceil - \overline{\delta}\right) s+1 \le j \le \left(\left\lceil \frac{i}{s} \right\rceil - 1\right) s+\delta\\ &\iff \frac{1}{s} - \overline{\delta} \le \frac{j}{s} - \left\lceil \frac{i}{s} \right\rceil \le \overline{\delta} - 1\\ &\iff \left|\left\lceil \frac{j}{s} \right\rceil - \left\lceil \frac{i}{s} \right\rceil\right| < \overline{\delta}.\end{aligned}$$

Before addressing invertibility and conditioning of lifted measurement systems with shifts, for $N \in \mathbb{N}$, we introduce $\mathcal{T}_N : \bigcup_{\ell\in\mathbb{N}} \mathbb{C}^{\ell N\times m} \to \bigcup_{\ell\in\mathbb{N}} \mathbb{C}^{\ell m\times N}$, the blockwise transpose operator, defined by

$$\mathcal{T}_N\left(\begin{bmatrix}V_1\\\vdots\\V_\ell\end{bmatrix}\right) = \begin{bmatrix}V_1^*\\\vdots\\V_\ell^*\end{bmatrix}$$

for $V_1, \ldots, V_\ell \in \mathbb{C}^{N\times m}$. We also define, for $\{k_j\}_{j=1}^n$ and $V \in \mathbb{C}^{m\times n}$,

$$\mathcal{I}(V, (k_j)_{j=1}^n) = \begin{bmatrix} v_1 \otimes I_{k_1} & \cdots & v_n \otimes I_{k_n} \end{bmatrix},$$

where $v_j = Ve_j$ are the columns of V. This prepares us to prove the following lemmas.

Lemma 14. *Given $k, N, m \in \mathbb{N}$ and $V \in \mathbb{C}^{kN\times M}$, we have*

$$\operatorname{circ}^N(V)^* = \operatorname{circ}^m\left((R_k \otimes I_m)\mathcal{T}_N(V)\right).$$

Proof. Suppose V_i are the $N \times m$ blocks of V, such that $V = \begin{bmatrix} V_1^T \cdots V_k^T \end{bmatrix}^T$. Indexing blockwise, we have $\operatorname{circ}^N(V)_{[ij]} = V_{i-j+1}$, so that $\operatorname{circ}^N(V)^*_{[ij]} = V^*_{j-i+1}$. In other words,

$$\operatorname{circ}^N(V)^* = \begin{bmatrix} V_1^* & V_2^* & \cdots & V_N^* \\ V_N^* & V_1^* & \cdots & V_{N-1}^* \\ \vdots & & \ddots & \vdots \\ V_2^* & V_3^* & \cdots & V_1^* \end{bmatrix} = \operatorname{circ}^m((R_k \otimes I_m)\mathcal{T}_N(V))$$

as claimed.

□

Lemma 15. *Given $N_1, N_2, k, m \in \mathbb{N}$ and $V_i \in \mathbb{C}^{kN_1\times m}$ for $i \in [N_2]$, we have*

$$\begin{bmatrix} \operatorname{circ}^{N_1}(V_1) & \cdots & \operatorname{circ}^{N_1}(V_{N_2}) \end{bmatrix} (P^{(k,N_2)} \otimes I_m)^* = \operatorname{circ}^{N_1}(\begin{bmatrix} V_1 & \cdots & V_{N_2} \end{bmatrix}).$$

Proof. We quote (4.15) from lemma 10 and consider that $P^{(k,N_2)} \otimes I_m$ is a permutation that changes the blockwise indices of $m \times p$ blocks (or, acting from the right, $p \times m$ blocks) exactly the way that $P^{(k,N_2)}$ changes the indices of a vector. □

Lemma 16. *Given $k, n \in \mathbb{N}$ and $V_j \in \mathbb{C}^{m_j \times n_j}$, we have*

$$\operatorname{diag}(I_k \otimes V_j)_{j=1}^n = P_1(I_k \otimes \operatorname{diag}(V_j)_{j=1}^n)P_2^*$$

where $P_1 = \mathcal{I}(P^{(n,k)}, (m_j)_{j=1}^n)$ and $P_2 = \mathcal{I}(P^{(n,k)}, (n_j)_{j=1}^n)$.

Proof. We immediately reduce to the case $m_j = n_j = 1$ (and we replace V with $v \in \mathbb{C}^n$) for all j by observing that P_1 and P_2 will act on blockwise indices precisely as $P^{(n,k)}$ acts on individual indices. Hence, we need only remark that

$$(\operatorname{diag}(I_k \otimes v_\ell)_{\ell=1}^n)_{((i_1-1)k+i_2)((j_1-1)k+j_2)} = \begin{cases} v_{i_1}, & i_1 = j_1 \text{ and } i_2 = j_2 \\ 0, & \text{otherwise} \end{cases},$$

while

$$\begin{aligned} &(P^{(n,k)}(I_k \otimes \operatorname{diag}(v))P^{(n,k)*})_{((i_1-1)k+i_2)((j_1-1)k+j_2)} \\ &\quad = (I_k \otimes \operatorname{diag}(v))_{((i_2-1)n+i_1)((j_2-1)n+j_1)} \\ &\quad = \begin{cases} v_{i_1}, & i_1 = j_1 \text{ and } i_2 = j_2 \\ 0, & \text{otherwise} \end{cases}. \end{aligned}$$

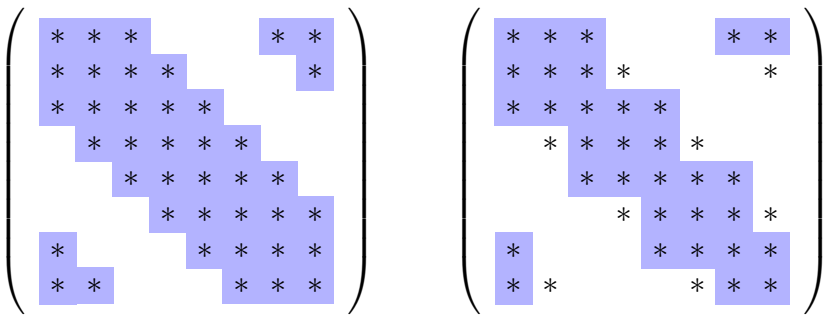
□

For the remainder of this section, we assume that $\delta > s$. We now consider the question of when $\operatorname{span} \mathcal{L}_{\{m_j\}}^s = T_{\delta,s}$ and what the condition number of $\mathcal{A}$ will be; naturally, this requires us to have redefined $\mathcal{A}$ by

$$\mathcal{A}(X)_{(\ell,j)} = \langle S^{s\ell} m_j m_j^* S^{-s\ell}, X \rangle, \quad (\ell, j) \in [\overline{d}]_0 \times [D].$$

As in (**??**), we vectorize X by its diagonals and write $A \in \mathbb{C}^{\overline{d}D \times (2\delta-1)d}$ such that

$$\left(A \begin{bmatrix} \chi_{1-\delta} \\ \vdots \\ \chi_{\delta-1} \end{bmatrix} \right)_{(j-1)\overline{d}+\ell} = \mathcal{A}(X)_{(\ell,j)},$$



(a) $T_3(\mathbb{C}^{8\times 8})$ (b) $T_{3,2}(\mathbb{C}^{8\times 8})$

Figure 5.1: $T_\delta(\mathbb{C}^{d\times d})$ vs. $T_{\delta,s}(\mathbb{C}^{d\times d})$ for $d = 8, \delta = 3, s = 2$

which gives the $(j-1)\overline{d} + \ell^{\text{th}}$ row of A as

$$\begin{bmatrix} S^{s(\ell-1)} g_{1-\delta}^j \\ \vdots \\ S^{s(\ell-1)} g_{\delta-1}^j \end{bmatrix}^*$$

so that, by lemma 14, we have

$$A = \begin{bmatrix} \operatorname{circ}^s(g_{1-\delta}^1) & \cdots & \operatorname{circ}^s(g_{1-\delta}^D) \\ \vdots & \ddots & \vdots \\ \operatorname{circ}^s(g_{\delta-1}^1) & \cdots & \operatorname{circ}^s(g_{\delta-1}^D) \end{bmatrix}^* = \begin{bmatrix} \operatorname{circ}(R_{\overline{d}} \mathcal{T}_s g_{1-\delta}^1) & \cdots & \operatorname{circ}(R_{\overline{d}} \mathcal{T}_s g_{\delta-1}^1) \\ \vdots & \ddots & \vdots \\ \operatorname{circ}(R_{\overline{d}} \mathcal{T}_s g_{1-\delta}^D) & \cdots & \operatorname{circ}(R_{\overline{d}} \mathcal{T}_s g_{\delta-1}^D) \end{bmatrix}. \tag{5.1}$$

However, because $T_{\delta,s} \subsetneq T_\delta$, this operator can never be invertible. Figure 5.1 shows this visually. Indeed, we consider that (restricting to $m \geq 0$), even if $\operatorname{supp}(m_k) = [\delta]$ for all k, $\operatorname{supp}(g_m^k) = [\delta - m]$, so when $\delta - m < s, \bigcup_{\ell=1}^{\overline{d}} \operatorname{supp}(S^{s(\ell-1)} g_m^k) \subsetneq [d]$. In particular, $\operatorname{circ}^s(g_m^k)_{ij} = 0$ for all j when $i \bmod_1 s > \delta - m$. By a similar argument, for $m < 0$ we have $\operatorname{circ}^s(g_m^k)_{ij} = 0$ when $i \bmod_1 s < s - (\delta - |m|)$. We remark that these inequalities can only be satisfied when $m > \delta - s$ or $m > \delta - s$, respectively.

By reference to (5.1), it is clear that each of these "missing indices" results in a column of all zeros in A; specifically, viewing $Ae_{(m+\delta-1)d+i}, (m, i) \in [2\delta - 1]_{1-\delta} \times [d]$

as the i^{th} column of the $m+\delta^{\text{th}}$ block of A, we see

$$Ae_{(m+\delta-1)d+i} = 0 \quad \text{if} \quad \begin{cases} i \bmod_1 s & > & \delta - m & , & m \geq 0 \\ i \bmod_1 s & < & s - (\delta + m), & & m \leq 0 \end{cases}.$$

Since $\delta > s$, we may reduce this condition to "$i \bmod s > \delta - m$ or $i \bmod s < s - (\delta + m)$," or further to $i \bmod s \notin [2\delta - s + 1]_{s-\delta-m}$. Therefore, the matrix representing $\mathcal{A}$ restricted to $T_{\delta,s}(\mathbb{C}^{d\times d})$ is found by right multiplying A by

$$N = \operatorname{diag}(I_{\overline{d}} \otimes N_{j-\delta})_{j=1}^{2\delta-1}, \text{ where } N_m = \begin{cases} \begin{bmatrix} 0_{\delta+m} \\ I_{s-(\delta+m)} \end{bmatrix}, & m < s - \delta \\ \begin{bmatrix} I_{s-(\delta-m)} \\ 0_{\delta-m} \end{bmatrix}, & m > \delta - s \\ I_s \quad , & \text{otherwise} \end{cases}. \tag{5.2}$$

But does this restriction commute well with the permutations used in the condition number analysis of section 4.2? Thankfully it does; following the intuition of (4.21) and making use of lemma 15, we can arrive at

$$\begin{aligned} A' := P^{(\overline{d},D)} A \left(P^{(\overline{d},2\delta-1)} \otimes I_s \right)^* &= \operatorname{circ}^D \left(P^{(\overline{d},D)} \begin{bmatrix} R_{\overline{d}}\mathcal{T}_s g^1_{1-\delta} & \cdots & R_{\overline{d}}\mathcal{T}_s g^1_{\delta-1} \\ \vdots & \ddots & \vdots \\ R_{\overline{d}}\mathcal{T}_s g^D_{1-\delta} & \cdots & R_{\overline{d}}\mathcal{T}_s g^D_{\delta-1} \end{bmatrix} \right). \\ &= \operatorname{circ}^D \left(P^{(\overline{d},D)} (I_D \otimes R_{\overline{d}}) \begin{bmatrix} \mathcal{T}_s g^1_{1-\delta} & \cdots & \mathcal{T}_s g^1_{\delta-1} \\ \vdots & \ddots & \vdots \\ \mathcal{T}_s g^D_{1-\delta} & \cdots & \mathcal{T}_s g^D_{\delta-1} \end{bmatrix} \right). \end{aligned}$$

In the interest of finding the locations of the zero columns after this permutation, we remark that the inner matrix is of size $\overline{d}D \times s(2\delta - 1)$, and that the circ^D

operation will therefore repeat it $\overline{d}$ times. It is then clear that

$$A'e_{(\ell-1)s(2\delta-1)+i} = 0 \iff \begin{bmatrix} \mathcal{T}_s g^1_{1-\delta} & \cdots & \mathcal{T}_s g^1_{\delta-1} \\ \vdots & \ddots & \vdots \\ \mathcal{T}_s g^D_{1-\delta} & \cdots & \mathcal{T}_s g^D_{\delta-1} \end{bmatrix} e_i = 0, \text{ and}$$

$$\begin{bmatrix} \mathcal{T}_s g^1_{1-\delta} & \cdots & \mathcal{T}_s g^1_{\delta-1} \\ \vdots & \ddots & \vdots \\ \mathcal{T}_s g^D_{1-\delta} & \cdots & \mathcal{T}_s g^D_{\delta-1} \end{bmatrix} e_{(m+\delta-1)s+i} = 0 \iff i \notin [2\delta - s + 1]_{s-\delta-m},$$

so we may remove the zero columns from A' by right multiplying the interior matrix by $N' = \operatorname{diag}(N_m)_{m=1-\delta}^{\delta-1}$. That is,

$$\begin{aligned} A'(I_{\overline{d}} \otimes N') &= \operatorname{circ}^D \left(P^{(\overline{d},D)}(I_D \otimes R_{\overline{d}}) \begin{bmatrix} \mathcal{T}_s g^1_{1-\delta} & \cdots & \mathcal{T}_s g^1_{\delta-1} \\ \vdots & \ddots & \vdots \\ \mathcal{T}_s g^D_{1-\delta} & \cdots & \mathcal{T}_s g^D_{\delta-1} \end{bmatrix} \begin{bmatrix} N_{1-\delta} & & \\ & \ddots & \\ & & N_{\delta-1} \end{bmatrix} \right) \\ &= P^{(\overline{d},D)} A N P', \end{aligned} \tag{5.3}$$

where the second equality comes from lemma 16 and

$$P' = \mathcal{I}(P^{(\overline{d},2\delta-1)}, (\min\{s, \delta - |m|\})_{m=1-\delta}^{\delta-1}).$$

This result, along with corollary 4, gives us the following proposition.

Proposition 7. *Taking A as in* (5.1)*, N and N_m as in* (5.2)*, and setting*

$$H = P^{(\overline{d},D)}(I_D \otimes R_{\overline{d}}) \begin{bmatrix} \mathcal{T}_s g^1_{1-\delta} & \cdots & \mathcal{T}_s g^1_{\delta-1} \\ \vdots & \ddots & \vdots \\ \mathcal{T}_s g^D_{1-\delta} & \cdots & \mathcal{T}_s g^D_{\delta-1} \end{bmatrix} \operatorname{diag}(N_m)_{m=1-\delta}^{\delta-1}$$

and $M_j = \sqrt{\overline{d}}(f_j^{\overline{d}} \otimes I_D)^* H$ for $j \in [\overline{d}]$, the condition number of AN is given by

$$\frac{\max\limits_{i \in [\overline{d}]} \sigma_{\max}(M_i)}{\min\limits_{i \in [\overline{d}]} \sigma_{\min}(M_i)}.$$

In particular, $\mathcal{A}|_{T_{\delta,s}(\mathbb{C}^{d\times d})}$ is invertible if and only if each of the M_i are of full rank.

5.2 Recovery Algorithm

In this section, we propose an algorithm by which we can recover an estimate of x_0 from $T_{\delta,s}(\mathcal{A}^{-1}(y))$.

Chapter 6

Angular Synchronization

6.1 Definition and Previous Work

In this section, we consider the problem of angular synchronization, which appears as a subproblem in many approaches to phase retrieval, including ours. In particular, we make a study of it in this section in order to improve the results in sections 3.4 and 3.5 and to supply a crucial lemma for section 5.2.

Angular synchronization is the problem of recovering a vector of complex units $x_i \in \mathbb{S}^1, i \in [n]$, or $x \in (\mathbb{S}^1)^n$ from estimates of their relative phases

$$\widetilde{X}_{ij} = x_j^* x_i \eta_{ij}, (i,j) \in E$$

where $\eta_{ij} \in \mathbb{S}^1$ are the "noise terms" and $E \subseteq [n]^2$ is a list of the pairs of indices for which we have such estimates. This immediately suggests associating a graph to this problem; namely, taking the vertex set to be $V = [n]$, we set $G = (V, E)$. We also remark that this problem invites the same global phase ambiguity as phase retrieval: indeed $(\mathrm{e}^{\mathrm{i}\theta} x_i)^* (\mathrm{e}^{\mathrm{i}\theta} x_j) \eta_{ij} = x_i^* x_j \eta_{ij}$ for any $\theta \in [0, 2\pi)$. Obviously, then, our goal is to get an estimate $\tilde{x}$ that is guaranteeably close to the ground truth x in

some chosen metric, say our usual $\min_{\theta\in\mathbb{R}} ||x - \mathrm{e}^{\mathrm{i}\theta}y||_2$, and to acquire this estimate with the least computational cost. Naturally, x is not known, so we instead attempt to minimize some cost function corresponding to how well our estimate explains the measurement data, usually taking a form similar to the frustration function stated in (3.25). Often, we more simply take only the numerator of this expression,

$$\sum_{(i,j)\in E} w_{ij}|x_i - X_{ij}x_j|^2 = x^*(D - W \circ X)x,$$

where D and W are the degree and weight matrices specified in (3.4). For convenience, we define

$$L = D - W \circ X,\ \underline{L} = D - W \circ xx^*,\ \text{and } L_G = D - W. \tag{6.1}$$

The approach to angular synchronization that we consider in this dissertation, and the approach most studied in the literature, then is to attempt the non-convex optimization problem

$$\begin{aligned} \min_{z\in\mathbb{C}^n} \quad & z^*Lz \\ \text{s.t.} \quad & |z_i| = 1 \end{aligned} \tag{6.2}$$

The modern study of eigenvector-based methods for angular synchronization appears to have begun with a 2011 paper by Amit Singer [68], in which he proposed two ways of solving this problem which remain as the basis of the state of the art. The first is almost identical to the eigenvector-based approach that we have used in the algorithm proposed in chapter 3, and the second is a semidefinite relaxation of the same. Namely, using the unweighted adjacency matrix $A_{ij} = X_{ij}\chi_E(i,j)$, his eigenvector method solves

$$\max_{||z||_2^2 = n} z^*Az$$

to find the largest eigenvector $\hat{z}$ of A and rounds to a vector of units by taking

$\widetilde{\mathbf{x}} = \operatorname{sgn}(\hat{z})$. The SDP method solves

$$\begin{aligned} \max_{Z \in \mathcal{H}^d} \quad & \operatorname{Tr}(AZ) \\ \text{s.t.} \quad & Z \succeq 0 \\ & Z_{ii} = 1 \end{aligned} \tag{6.3}$$

In this paper, he studies the problem under a noise model where the disturbances η_{ij} are distributed according to

$$\eta_{ij} = \begin{cases} 1, & \text{with probability } p \\ \operatorname{Unif}(\mathbb{S}^1), & \text{with probability } 1-p \end{cases},$$

such that the measurement $\widetilde{X}_{ij}$ is exact with probability p and is completely meaningless – being drawn from the uniform distribution on $\mathbb{S}^1$ – with probability $1-p$. He proves the robustness of this method in the sense that there is a probability p_c, dependent on the spectral gap and size of the graph G, for which parameter values $p > p_c$ guarantee "better than random" approximations of x with high probability. Moreover, he shows that, experimentally, both of these recovery algorithms work acceptably, if not extremely, well, with little to be gained by transferring from the eigenvector problem to the computationally more expensive semidefinite program.

The literature on angular synchronization since this paper has largely consisted of analyzing generalizations and variations of these methods. One major generalization has been to apply these methods to larger classes of group synchronization problems such as synchronization over the orthogonal groups $O(d)$ or the special Euclidean groups $SE(d)$ [8, 15, 65]. Naturally, much of the interest in this subject has been the treatment of synchronization over $SO(3)$ and $SE(3)$, as these correspond to pose estimation problems fundamental to computer vision, as in [28, 29, 32, 38]. Significant results giving guarantees of robustness, as well as proofs that these relaxations are solved *exactly* in certain cases may be found in [2, 9, 29, 65].

6.2 Tightness of SDP Relaxation

6.2.1 Introduction and Main Result

We have already presented one angular synchronization result in §3.4, which drew largely on [2]. We remark that this theorem [**BP Note: *rephrase: don't want to downplay your own result*] leaves something to be desired in that the graph G is not permitted to be weighted, which restricts us from applying some knowledge that we may have about the problem. For example, suppose our relative phase measurements $\{X_{ij}\}_{(i,j)\in E}$ are disturbed by noise drawn from a fixed, phase-invariant probability distribution, say

$$X_{ij} = \operatorname{sgn}(x_i^* x_j + \epsilon_{ij}), \epsilon_{ij} = a_{ij} + \mathrm{i}\, b_{ij} \quad \text{with} \quad a_{ij}, b_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2),$$

then we will have more confidence in the relative phases represented by the larger magnitude entries of $X = \mathcal{A}^{-1}(\mathbf{y})$. It would be intuitive to use this knowledge to privilege some edges of the graph over others in the frustration function (3.25) that we are trying to minimize, say by using $w_{ij} = |X_{ij}|$. Unfortunately, theorem 4 assumes an unweighted graph and its proof technique does not readily admit a satisfactory adjustment towards weighted edges, though we shall impose one later. Therefore, we will take a distinct approach, drawing upon recent results in the literature that consider certain convex relaxations of (6.2) [9, 17, 65].

To begin this discussion, we gather our notation: $G = (V = [n], E)$ is a connected graph with a weighted adjacency matrix $W = [w_{ij}] \in \mathcal{S}^n$ satisfying $w_{ij} \geq 0$ and $w_{ij} \neq 0$ only if $(i,j) \in E$. We take $D = \operatorname{diag}(W\mathbb{1})$ to be the degree matrix. $\underline{x} \in (\mathbb{S}^1)^n$ is the ground truth vector, which we attempt to recover, and $X, \underline{X} \in \mathcal{H}^n$ are our noisy and ground truth edge data matrices, satisfying

$$X_{ij} = \begin{cases} \eta_{ij} x_i x_j^*, & (i,j) \in E \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \underline{X}_{ij} = \begin{cases} x_i x_j^*, & (i,j) \in E \\ 0, & \text{otherwise} \end{cases},$$

where $\eta_{ij} = \eta_{ji}^* \in \mathbb{S}^1$ for each $(i,j) \in E$. We then define $L, \underline{L}$, and L_G as in (6.1).

Towards formulating the appropriate SDP relaxations, we recall the transformation $\mathfrak{R} : \mathbb{C} \to \mathbb{R}^{2\times 2}$ defined by

$$\mathfrak{R}(a + \mathrm{i}\, b) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

It is well-known that $\mathfrak{R}$ is the canonical isomorphism from $\mathbb{C}$ into $\mathbb{R}^{2\times 2}$, and indeed if we extend it to matrices by taking, for $A \in \mathbb{C}^{m\times n}$, $\mathfrak{R}(A) \in \mathbb{R}^{2m\times 2n}$ to be a block matrix with $\mathfrak{R}(A)_{ij} = \mathfrak{R}(A_{ij})$, then it remains an isomorphism on $\mathbb{C}^{m\times n}$, and indeed it preserves the eigenvalues and eigenvectors of Hermitian matrices (see, e.g. [81, p. 101]). In particular, a Hermitian matrix $A \in \mathcal{H}^n$ is semi-definite if and only if $\mathfrak{R}(A)$ is semi-definite; we notice that the multiplicities of its eigenvalues are all doubled, as $Av = \lambda v$ implies $\mathfrak{R}(A)\mathfrak{R}(v) = \lambda\mathfrak{R}(v)$, giving that the two columns of $\mathfrak{R}(v)$ are each eigenvectors of $\mathfrak{R}(A)$ with eigenvalue λ.

With this, we consider SDP relaxations of (6.2). Specifically, we observe that $z^*Lz = \mathrm{Tr}(Lzz^*)$, so an equivalent optimization problem will be

$$\begin{array}{ll} \min\limits_{Z\in\mathcal{H}^n} & \mathrm{Tr}(LZ) \\ \text{s.t.} & Z_{ii} = 1 \\ & \mathrm{rank}(Z) = 1 \\ & Z \succeq 0 \end{array}, \tag{6.4}$$

where the optimizer $\hat{z}$ of (6.2) is recovered from the optimal matrix $\hat{Z}$ by merely factoring $\hat{Z} = \hat{z}\hat{z}^*$. To get a convex relaxation, we simply omit the non-convex rank one constraint, yielding

$$\begin{array}{ll} \min\limits_{Z\in\mathcal{H}^n} & \mathrm{Tr}(LZ) \\ \text{s.t.} & Z_{ii} = 1 \\ & Z \succeq 0 \end{array}. \tag{6.5}$$

We remark that if an optimizer $\hat{Z}$ of (6.5) is rank one, then it is also an optimizer of (6.4) since the feasible set of (6.5) is strictly larger than that of (6.4); in this

case, then, factoring $\hat{Z}$ gives the global minimizer of (6.2). Considering that many results in the optimization literature are written for real-valued SDPs, we further relax the feasible set by casting into the real domain with

$$\begin{aligned} \min_{Z \in \mathcal{S}^{2n \times 2n}} \quad & \operatorname{Tr}(\mathfrak{R}(L)Z) \\ \text{s.t.} \quad & Z_{ii} = I_2 \\ & Z \succeq 0 \end{aligned} \quad , \tag{6.6}$$

where $Z_{ii} = [e_{2i-1}\ e_{2i}]^* Z [e_{2i-1}\ e_{2i}]$ in this case refers to the i^{th} 2×2 diagonal block of Z.

At this point, we recognize the previous work existing on this problem. Namely, in [9], Bandeira, Boumal, and Singer prove that the optimizer $\hat{Z}$ of (6.5) is rank one (and therefore yields a minimizer of (6.2)) when L is sufficiently close to $\underline{L}$. Unfortunately for our purposes, this paper only considers the case when $G = K_n$ is the complete graph and the weights $W = \mathbb{1}\mathbb{1}^* - I_n$ are constant. A more general result appears in [65], where Rosen, Carlone, Bandeira, and Leonard prove a similar result for synchronization over $SE(d)$, of which angular synchronization is a special case. Moreover, these results allow for a weighted graph, and include a bound on $\min_{\theta \in [0,2\pi)} \|\hat{z} - \mathrm{e}^{\mathrm{i}\theta}\underline{x}\|_2$ in terms of $\|L - \underline{L}\|_2$ and the spectral gap of the graph. Nonetheless, we find that narrowing to the case of $SO(2)$ (equivalent to angular synchronization) allows for a tighter error bound. In [17], Calafiore, Carlone, and Dellaert use methods similar to those in [65] to analyze $SE(2)$ synchronization. Furthermore, and pertinent to the present work, the authors exchange the rotational components in $SO(2)$ for complex units, but they do not admit weighted graphs, nor do they supply explicit bounds on the error of their estimate or on what level of noise may be tolerated and still guarantee that their convex relaxation solves (6.2) exactly. Significantly, all three of these works supply an *a posteriori*-certifiable condition that can verify whether the solution obtained is indeed optimal for eqs. (6.2) and (6.4)–(6.6).

Among the results just mentioned, of greatest interest to us are proposition 2 and Theorem 12 in [65], which may be restated as

Proposition 8 (Proposition 2 and Theorem 12 in [65])**.** *There exists a constant $\beta > 0$, depending on $\underline{L}$, such that, if $\|L-\underline{L}\|_2 < \beta$, then* (6.6) *has a unique solution $\hat{Z}$ which may be factored as $\hat{Z} = RR^*$, with $R = \mathfrak{R}(\hat{z})$ where $\hat{z} \in (\mathbb{S}^1)^n$ is a global optimizer of* (6.2)*. Furthermore,*

$$\min_{\theta\in[0,2\pi)} \|\hat{z} - \mathrm{e}^{\mathrm{i}\theta}\underline{x}\|_2 \leq 2\sqrt{\frac{n\|\underline{L} - L\|_2}{\lambda_2(L_G)}},$$

where $\lambda_2(L_G)$ is the second smallest eigenvalue of L_G.

We strengthen this result in theorem 7 by giving β explicitly and by increasing the exponent of $\|\underline{L} - L\|_2$ in the error bound, which improves the convergence rate as $L \to \underline{L}$.

Theorem 7. *Given a connected, weighted graph $G = (V = [n], E)$ with spectral gap $\tau = \lambda_2(D - W)$ and rotational data $X_{ij} \in \mathbb{S}^1$ for $(i, j) \in E$, suppose that $\hat{z}$ is a minimizer of* (6.2)*, where $L = D - W \circ X$. By $\underline{x}$ we denote the ground truth, and we take $\underline{L} = D - W \circ \underline{x}\underline{x}^*$ and $\hat{L} = D - W \circ \hat{z}\hat{z}^*$. Then if $\|L - \hat{L}\|_2 < \frac{\tau}{1+\sqrt{n}}$, $\hat{Z} = \hat{z}\hat{z}^*$ and $\mathfrak{R}(\hat{Z})$ are the unique minimizers of* (6.5) *and* (6.6) *and we have*

$$\min_{\theta\in[0,2\pi)} \|\hat{z} - \mathrm{e}^{\mathrm{i}\theta}\underline{x}\|_2 \leq \frac{2\sqrt{2n}\|\underline{L} - L\|_2}{\tau}. \tag{6.7}$$

6.2.2 Dual Problems

To prove theorem 7, we introduce dual problems for (6.2) and (6.6). Specifically, we give the Lagrangian function $\mathcal{L} : \mathbb{C}^n \times \mathbb{R}^n \to \mathbb{R}$ of (6.2),

$$\mathcal{L}(z, \lambda) = z^*Lz + \sum_{i=1}^{n} \lambda_i(1 - z_i^* z_i) = z^*(L - \mathrm{diag}(\lambda))z + \sum_{i=1}^{n} \lambda_i,$$

and the dual function $d : \mathbb{R}^n \to \mathbb{R}$

$$d(\lambda) = \inf_{z \in \mathbb{C}^n} \mathcal{L}(z, \lambda),$$

which have the properties that for any $\lambda \in \mathbb{R}^n, z \in (\mathbb{S}^1)^n$, we have

$$d(\lambda) \leq \mathcal{L}(z, \lambda) = z^* L z.$$

In particular, $\sup_{\lambda \in \mathbb{R}^n} d(\lambda) \leq \min_{z \in (\mathbb{S}^1)^n} z^* L z$. Additionally, for any $\lambda \in \mathbb{R}^n$ such that $L - \mathrm{diag}(\lambda) \not\succeq 0$, we have $d(\lambda) = -\infty$. Indeed, if $v^*(L - \mathrm{diag}(\lambda))v < 0$, then we may take

$$d(\lambda) \leq \lim_{t \to \infty} (tv)^*(L - \mathrm{diag}(\lambda))(tv) = \lim_{t \to \infty} t^2 v^* L v = -\infty.$$

Otherwise, in the case that $L - \mathrm{diag}(\lambda) \succeq 0$, the quadratic form $z^*(L - \mathrm{diag}(\lambda))z$ is minimized when $z = 0$, so

$$d(\lambda) = \mathcal{L}(0, \lambda) = \sum_{i=1}^{n} \lambda_i.$$

Together, this gives $d(\lambda)$ as

$$d(\lambda) = \begin{cases} \sum_{i=1}^n \lambda_i, & L - \mathrm{diag}(\lambda) \succeq 0 \\ -\infty, & L - \mathrm{diag}(\lambda) \not\succeq 0 \end{cases}$$

[**BP Note: *transition from d to opt. more gracefully*] Therefore, we may maximize d by considering the following optimization problem over the set of diagonal matrices where state the dual problem of (6.2) as

$$\begin{aligned} \max_{\Lambda \in \mathbb{R}^{n \times n}} \quad & \mathrm{Tr}(\Lambda) \\ \text{s.t.} \quad & L - \Lambda \succeq 0 \\ & \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \end{aligned} \quad , \tag{6.8}$$

in the sense that, if Λ^* optimizes (6.8) and $\hat{z}$ optimizes (6.2), then $\mathrm{Tr}(\Lambda^*) \leq \hat{z}^* L \hat{z}$.

To find the dual of (6.5), we define $\mathcal{L}_{SDP} : \mathcal{H}^n_+ \times \mathbb{R}^n \to \mathbb{R}$ and $d_{SDP} : \mathbb{R}^n \to \mathbb{R}$ by

$$\mathcal{L}_{SDP}(Z, \lambda) = \mathrm{Tr}(LZ) + \sum_{i=1}^{n} \lambda_i (1 - Z_{ii}) = \mathrm{Tr}((L - \mathrm{diag}(\lambda))Z) + \mathrm{Tr}(\mathrm{diag}(\lambda)), \text{ and}$$

$$d_{SDP}(\lambda) = \inf_{Z \in \mathcal{H}^n_+} \mathcal{L}_{SDP}(Z, \lambda) = \begin{cases} \mathrm{Tr}(\mathrm{diag}(\lambda)), & L - \mathrm{diag}(\lambda) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

As before, we see that $\sup_{\lambda \in \mathbb{R}^n} d_{SDP}(\lambda) \leq \mathcal{L}_{SDP}(Z, \lambda) = \mathrm{Tr}(LZ)$ for any Z which is feasible to (6.5); therefore, if Λ^* and $\hat{Z}$ optimize (6.8) and (6.5), we will have $\mathrm{Tr}(\Lambda^*) \leq \mathrm{Tr}(L\hat{Z})$.

To prove the uniqueness of the solution to (6.5), we will need to quote a result from [3]. They use the primal-dual format with the primal problem

$$\begin{aligned} \min_{X \in \mathcal{S}^n} \quad & \mathrm{Tr}(CX) \\ \text{s.t.} \quad & \mathrm{Tr}(A_k X) = b_k, \quad k \in [m] \\ & X \succeq 0 \end{aligned} \tag{6.9}$$

where $C, A_k \in \mathcal{S}^n$ and $b \in \mathbb{R}^m$ are fixed. The dual problem is

$$\begin{aligned} \max_{y \in \mathbb{R}^m} \quad & b^T y \\ \text{s.t.} \quad & C - \textstyle\sum_{k=1}^m y_k A_k \succeq 0 \end{aligned} \tag{6.10}$$

where $b \in \mathbb{R}^m$ and A_k are as in the primal. Among other results, they prove the following:

Proposition 9 (Theorems 9 and 10 in [3])**.** *Suppose that $y \in \mathbb{R}^m$ is dual feasible and optimal, with* $\mathrm{rank}(C - \sum_{k=1}^m y_k A_k) = s$*. Let the columns of $Q \in \mathbb{R}^{n \times n-s}$ be an orthonormal basis for* $\mathrm{Nul}(C - \sum_{k=1}^m y_k A_k)$*, such that*

$$\mathrm{Col}(Q) = \mathrm{Nul}(C - \sum_{k=1}^{m} y_k A_k) \quad \textit{and} \quad Q^T Q = I.$$

then if

$$\mathrm{span}\{Q^T A_k Q\}_{k \in [m]} = \mathcal{S}^{n-s}, \tag{6.11}$$

there is a unqiue optimal primal solution X.

In order to use this result, we will need the dual of (6.6). Following eqs. (6.9) and (6.10), this gives

$$\begin{aligned} \max_{\Lambda \in \mathcal{S}^{2n}} \quad & \mathrm{Tr}(\Lambda) \\ \text{s.t.} \quad & \mathfrak{R}(L) - \Lambda \succeq 0 \\ & \Lambda = \mathrm{diag}(\Lambda_1, \ldots, \Lambda_n) \\ & \Lambda_i \in \mathcal{S}^2 \end{aligned} \tag{6.12}$$

6.2.3 Proof of Theorem 7

With this in mind, we state a few lemmas whose proofs will constitute a proof of theorem 7. Each of these assumes the notation and hypotheses of theorem 7.

Lemma 17 (Sufficient conditions for strong duality)**.** *If $\hat{z} \in (\mathbb{S}^1)^n$ satisfies*

$$L - \mathrm{diag}\,\mathrm{Re}(\hat{z}\hat{z}^* L) \succeq 0, \text{ and} \tag{6.13}$$

$$\mathrm{Nul}(L - \mathrm{diag}\,\mathrm{Re}(\hat{z}\hat{z}^* L)) = \mathrm{span}(\hat{z}), \tag{6.14}$$

then $\hat{Z} = \hat{z}\hat{z}^*$ is the unique optimizer of (6.5) *and $\mathfrak{R}(\hat{Z})$ is the unique optimizer of* (6.6).

Lemma 18. *If $\|L - \hat{L}\|_2 < \frac{\tau}{1+\sqrt{n}}$, then $\hat{z}$ meets the conditions of lemma 17 and $\hat{Z} = \hat{z}\hat{z}^*$ is the unique optimizer of* (6.5) *and $\mathfrak{R}(\hat{Z})$ is the unique optimizer of* (6.6).

Lemma 19. *Suppose that $\hat{z}$ minimizes* (6.2)*. Then*

$$\min_{\theta \in [0, 2\pi)} \|\hat{z} - \mathrm{e}^{\mathrm{i}\theta} \underline{x}\|_2 \leq \frac{2\sqrt{2n}\|\underline{L} - L\|_2}{\tau}.$$

Before proving these, we begin with a further lemma that explains the recurrent $L - \mathrm{diag}\,\mathrm{Re}(\hat{z}\hat{z}^* L)$ term.

Lemma 20. *Suppose $A \succeq 0$. Then if $\hat{z}$ is an optimizer of*

$$\min_{z \in (\mathbb{S}^1)^n} \; z^* A z,$$

we have

$$\hat{z} \in \mathrm{Nul}(A - \operatorname{diag}\operatorname{Re}(\hat{z}\hat{z}^* A)).$$

Proof of lemma 20. We define $f : (\mathbb{S}^1)^n \to \mathbb{R}$ by $f(z) = z^* A z$. We observe that $(\mathbb{S}^1)^n$ is an n-dimensional manifold with charts given by

$$\phi_U : U \to (\mathbb{S}^1)^n, \quad \phi_U(\theta_1, \ldots, \theta_n)_i = \mathrm{e}^{\mathrm{i}\theta_i},$$

where $U \subseteq \mathbb{R}^n$ is any open set satisfying $\|x - y\|_\infty < 2\pi$ for all $x, y \in U$ (or, more generally, $x - y \notin 2\pi(\mathbb{Z}^n) \setminus \{0\}$ for all $x, y \in U$). Assume $\hat{z}$ is a minimizer of (6.2). Because $z^* A z$ is smooth on $(\mathbb{S}^1)^n$, for any chart ϕ_U with $\hat{z} \in \phi_U(U)$, we must have $\nabla(f \circ \phi_U)|_{\phi_U^{-1}(\hat{z})} = 0$. In particular, if $\theta \in \mathbb{R}^n$ is such that $\hat{z}_i = \mathrm{e}^{\mathrm{i}\theta_i}$, then

$$\nabla(f \circ \phi_U)|_{\phi_U^{-1}(\hat{z})} = \nabla_\theta \hat{z}^* A \hat{z}.$$

We remark that

$$\hat{z}^* A \hat{z} = \sum_{ij} \mathrm{e}^{\mathrm{i}(\theta_j - \theta_i)} A_{ij} = \sum_i \mathrm{e}^{\mathrm{i}\theta_i} \hat{z}^* A_i$$

where $A_i = A e_i$ and $A_{ij} = (A_i)_j$. This gives

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} \hat{z}^* A \hat{z} &= \mathrm{i}\mathrm{e}^{\mathrm{i}\theta_j} \hat{z}^* A_j + \sum_{i=1}^n \mathrm{e}^{\mathrm{i}\theta_i} \frac{\partial}{\partial \theta_j} \hat{z}^* A_i \\
&= \mathrm{i}\mathrm{e}^{\mathrm{i}\theta_j} \hat{z}^* A_j - \mathrm{i}\mathrm{e}^{\mathrm{i}\theta_j} \sum_{i=1}^n \mathrm{e}^{-\mathrm{i}\theta_i} A_{ji} \\
&= \mathrm{i}\mathrm{e}^{\mathrm{i}\theta_j} \hat{z}^* A_j - \mathrm{i}\mathrm{e}^{-\mathrm{i}\theta_j} A_j^* \hat{z} \\
&= -2 \operatorname{Im}(\mathrm{e}^{\mathrm{i}\theta_j} \hat{z}^* A_j) = -2 \operatorname{Im}(\hat{z}\hat{z}^* A)_{jj}
\end{aligned}$$

Therefore, $\nabla(f \circ \phi_U)|_{\phi_U^{-1}(\hat{z})} = 0$ if and only if $\operatorname{diag}\operatorname{Im}(\hat{z}\hat{z}^* A) = 0$. On the other hand, we observe that

$$\begin{aligned}
((A - \operatorname{diag}\operatorname{Re}(\hat{z}\hat{z}^* A))\hat{z})_i &= A_i^* \hat{z} - \operatorname{Re}(\hat{z}_i \hat{z}^* A_i)\hat{z}_i \\
&= A_i^* \hat{z} - \operatorname{Re}(\overline{\hat{z}}_i A_i^* \hat{z})\hat{z}_i,
\end{aligned}$$

such that $\hat{z} \in \mathrm{Nul}(A - \operatorname{diag}\mathrm{Re}(\hat{z}\hat{z}^* A))$ iff

$$\begin{aligned} & A_i^* \hat{z} &&= \mathrm{Re}(\overline{\hat{z}}_i A_i^* \hat{z})\hat{z}_i, && \text{for all } i \\ \iff \quad & \overline{\hat{z}}_i A_i^* \hat{z} &&= \mathrm{Re}(\overline{\hat{z}}_i A_i^* \hat{z}), && \text{for all } i \\ \iff \quad & \overline{\hat{z}}_i A_i^* \hat{z} &&\in \mathbb{R} && \text{for all } i \\ \iff \quad & \operatorname{diag}\mathrm{Im}(\hat{z}\hat{z}^* A) &&= 0 \end{aligned}$$

This may be summarized as

$$z \in \mathrm{Nul}(A - \operatorname{diag}\mathrm{Re}(\hat{z}\hat{z}^* A)) \iff \overline{\hat{z}}_i A_i^* \hat{z} \in \mathbb{R} \text{ for all } i \tag{6.15}$$

so that $\hat{z}$ is a minimizer of (6.2) only if $\hat{z} \in \mathrm{Nul}(A - \operatorname{diag}\mathrm{Re}(\hat{z}\hat{z}^* A))$. □

Beyond being a useful result, this suggests that the matrix $L - \operatorname{diag}\mathrm{Re}(\hat{z}\hat{z}^* L)$ can play a role in certifying minima of these optimization problems, as we shall see in the proofs of lemmas 17–19. We remark, however, that this condition is far from being *sufficient* to show that $\hat{z}$ is an optimizer of $\min_{z \in (\mathbb{S}^1)^n} z^* A z$. Indeed, even the stronger condition $\mathrm{Nul}(A - \operatorname{diag}\mathrm{Re}(\hat{z}\hat{z}^* A)) = \mathrm{span}(\hat{z})$ is insufficient. For example, if we take $A \in \mathcal{S}^2$ to be $A = \mathbb{1}\mathbb{1}^*$, then with $z = \mathbb{1}$ we have

$$A - \operatorname{diag}\mathrm{Re}(\mathbb{1}\mathbb{1}^* A) = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix},$$

so that $\mathrm{Nul}(A - \operatorname{diag}\mathrm{Re}(\mathbb{1}\mathbb{1}^* A)) = \mathrm{span}(\mathbb{1})$, but

$$\begin{bmatrix} 1 \\ \mathrm{e}^{\mathrm{i}\theta} \end{bmatrix}^* A \begin{bmatrix} 1 \\ \mathrm{e}^{\mathrm{i}\theta} \end{bmatrix} = 2 + 2\cos\theta,$$

which is *maximized* at $z = \mathbb{1}$ and minimized at $\hat{z} = (1, -1)^T$. We now prove lemmas 17–19.

Proof of lemma 17. We begin by showing that

$$\hat{z} \in \mathrm{Nul}(L - \operatorname{diag}\mathrm{Re}(\hat{z}\hat{z}^* L)) \text{ and } L - \operatorname{diag}\mathrm{Re}(\hat{z}\hat{z}^* L) \succeq 0$$

suffices to show that $\hat{Z} = \hat{z}\hat{z}^*$ and $\mathfrak{R}(\hat{Z})$ are optimizers of (6.5) and (6.6), respectively. We then show that, if $\mathrm{Nul}(L - \operatorname{diag}\operatorname{Re}(\hat{z}\hat{z}^*L)) = \operatorname{span}(\hat{z})$ holds in addition to these, then they are the *unique* optimizers of their problems.

Suppose that $\hat{z} \in (\mathbb{S}^1)^n$ satisfies $L - \operatorname{diag}\operatorname{Re}(\hat{z}\hat{z}^*L) \succeq 0$ and $(L - \operatorname{diag}\operatorname{Re}(\hat{z}\hat{z}^*L))\hat{z} = 0$, and set $\hat{Z} = \hat{z}\hat{z}^*$. This means $\operatorname{diag}\operatorname{Re}(\hat{z}\hat{z}^*L)$ is feasible for (6.8), so we set $\hat{\Lambda} = \operatorname{diag}\operatorname{Re}(\hat{z}\hat{z}^*L)$ to obtain

$$\begin{array}{rclcll} \operatorname{Tr}(\hat{\Lambda}) & = & \operatorname{Tr}(\operatorname{diag}\operatorname{Re}(\hat{z}\hat{z}^*L)) & = & \operatorname{Tr}(\operatorname{Re}(\hat{z}\hat{z}^*L)) & \\ & = & \sum_{i=1}^n \operatorname{Re}(\hat{z}_i\hat{z}^*L_i) & = & \sum_{i=1}^n \hat{z}_i\hat{z}^*L_i & \text{(from (6.15))} \\ & = & \hat{z}^*L\hat{z} & & & \end{array}$$

Therefore $(\hat{z}, \hat{\Lambda})$ is a primal-dual pair for (6.2) and (6.8) with equal objective function values, showing that they are both optimizers for their respective problems. Since (6.8) is also the dual problem for (6.5), and since $\operatorname{Tr}(L\hat{Z}) = \operatorname{Tr}(L\hat{z}\hat{z}^*) = \hat{z}^*L\hat{z}$, $\hat{\Lambda}$ also certifies optimality of $\hat{Z}$ for *its* problem, and similarly also for $\mathfrak{R}(\hat{Z})$ by observing that $\operatorname{Tr}(\mathfrak{R}(A)) = 2\operatorname{Tr}(A)$ for general $A \in \mathcal{H}^n$.

To show uniqueness, we will show that, under the additional assumption that $\mathrm{Nul}(L - \operatorname{diag}\operatorname{Re}(\hat{z}\hat{z}^*L)) = \operatorname{span}(\hat{z})$, then $\mathfrak{R}(\hat{\Lambda})$ satisfies the hypotheses of proposition 9. To this end, observe that $\mathrm{Nul}(L - \hat{\Lambda}) = \operatorname{span}(\hat{z})$ implies

$$\mathrm{Nul}(\mathfrak{R}(L) - \mathfrak{R}(\hat{\Lambda})) = \operatorname{Col}\mathfrak{R}(\hat{z}).$$

The two columns of $\mathfrak{R}(\hat{z})$ are trivially orthogonal, so we set $Z_i = \mathfrak{R}(\hat{z}_i)$ and

$$Q = \frac{1}{n}\mathfrak{R}(\hat{z}) = \frac{1}{n}\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}.$$

Towards proving that Q satisfies (6.11), we note that $Z_i^T Z_i = Z_i Z_i^T = I_2$ and claim that, for $A \in \mathcal{S}^2$, one block-diagonal pre-image of A is $n^2 \operatorname{diag}(Z_1 A Z_1^T, 0, \ldots, 0)$. Indeed,

$$Q^T n^2 \operatorname{diag}(Z_1 A Z_1^T, 0, \ldots, 0) Q = Z_1^T Z_1 A Z_1^T Z_1 = A.$$

This establishes that $\operatorname{span}\{Q^T \operatorname{diag}(\Lambda_i) Q\}_{\Lambda_i \in \mathcal{S}^2} = \mathcal{S}^2$, which, by proposition 9, gives us that (6.6) has a unique solution. Since we have already established optimality of $\mathfrak{R}(\hat{Z})$, this unique solution is $\mathfrak{R}(\hat{Z})$.

Label the feasible sets of (6.5) and (6.6) F_1 and F_2, respectively. Then $\mathfrak{R}(F_1) \subseteq F_2$, such that if $\mathfrak{R}(\hat{Z})$ is the unique minimizer of $\operatorname{Tr}(\mathfrak{R}(L)Z)$ over F_2, then $\hat{Z}$ is the unique minimizer of $\operatorname{Tr}(LZ)$ over F_1. This completes the proof. □

Proof of lemma 18. Accepting the notation of theorem 7, suppose that $\hat{z}$ is an optimizer of (6.2). For convenience, we set $Y = L - \operatorname{diag} \operatorname{Re}(\hat{z}\hat{z}^* L)$. Then, by lemma 20 we have $Y\hat{z} = 0$. Therefore, by lemma 17, to show that the solution to (6.5) is unique, it suffices to have $P^* Y P \succ 0$, where the columns of $P \in \mathbb{C}^{n \times n-1}$ form an orthonormal basis for $\hat{z}^\perp$.

To this end, we introduce $\hat{L} = D - W \circ \hat{z}\hat{z}^*$, which has that $\operatorname{Nul}(\hat{L}) = \operatorname{span}(\hat{z})$ and $P^* \hat{L} P \succ 0$ when G is connected (see, e.g., lemma 1.7 of [22]), so that

$$\hat{L} - \operatorname{diag} \operatorname{Re} \hat{z}\hat{z}^* \hat{L} = \hat{L}.$$

By Weyl's inequalities (Theorem 4.3.1 in [44]), the smallest eigenvalue $\lambda_1(P^* Y P)$ of $P^* Y P$ satisfies

$$\lambda_1(P^* Y P) \geq \lambda_1(P^* \hat{L} P) - \|P^*(Y - \hat{L})P\|_2 = \lambda_2(\hat{L}) - \|Y - \hat{L}\|_2,$$

so it suffices to have $\lambda_2(\hat{L}) = \tau > \|Y - \hat{L}\|_2$. The lemma follows by observing that

$$\begin{aligned}
\|Y - \hat{L}\|_2 &\leq \|L - \hat{L}\|_2 + \|\operatorname{diag} \operatorname{Re}(\hat{z}\hat{z}^* L)\|_2 \\
&\leq \|L - \hat{L}\|_2 + \|L\hat{z}\|_\infty \\
&\leq \|L - \hat{L}\|_2 + \|L\hat{z}\|_2 \\
&= \|L - \hat{L}\|_2 + \|(L - \hat{L})\hat{z}\|_2 \\
&\leq \|L - \hat{L}\|_2 + \sqrt{n}\|L - \hat{L}\|_2,
\end{aligned}$$

□

Proof of lemma 19. We begin by assuming that $\underline{x}^*\hat{z} = \hat{z}^*\underline{x} = |\underline{x}^*\hat{z}|$, which is accomplished by taking appropriate representatives of $\underline{x}$ and $\hat{z}$ in $(\mathbb{S}^1)^n/\mathbb{S}^1$. This gives that

$$\min_{\theta\in[0,2\pi)} \|\underline{x} - \mathrm{e}^{\mathrm{i}\theta}\hat{z}\|_2 = \|\underline{x} - \hat{z}\|_2.$$

Writing $\Delta L = L - \underline{L}$, we have, by optimality of $\hat{z}$, that $\hat{z}^*L\hat{z} \leq \underline{x}^*L\underline{x}$. Since $\underline{x} \in \mathrm{Nul}(\underline{L})$, this gives

$$\hat{z}^*\underline{L}\hat{z} + \hat{z}^*\Delta L\hat{z} = \hat{z}^*L\hat{z} \leq \underline{x}^*L\underline{x} = \underline{x}^*\underline{L}\underline{x} + \underline{x}^*\Delta L\underline{x} = \underline{x}^*\Delta L\underline{x},$$

which yields

$$\begin{aligned} \hat{z}^*\underline{L}\hat{z} &\leq \underline{x}^*\Delta L\underline{x} - \hat{z}^*\Delta L\hat{z} \\ &= (\underline{x} - \hat{z})^*\Delta L(\underline{x} + \hat{z}) \\ &\leq \|\underline{x} - \hat{z}\|_2\|\Delta L\|_2\sqrt{2n} \end{aligned} \tag{6.16}$$

We then lower-bound the left-hand side of (6.16) by setting

$$y = \mathrm{Proj}_{\underline{x}^\perp}\hat{z} = \hat{z} - \frac{1}{n}\underline{x}\underline{x}^*\hat{z},$$

so that

$$\hat{z}^*\underline{L}\hat{z} = y^*\underline{L}y.$$

At this point, we remark that $\|\underline{x}\underline{x}^* - \hat{z}\hat{z}^*\|_F^2 = 2n^2 - 2|\underline{x}^*\hat{z}|^2$ and $\|\underline{x} - \hat{z}\|_2^2 = 2n - 2|\underline{x}^*\hat{z}|$, such that

$$\begin{aligned} n\|\underline{x} - \hat{z}\|_2^2 \leq \|\underline{x}\underline{x}^* - \hat{z}\hat{z}^*\|_F^2 &= 2(n - |\underline{x}^*\hat{z}|)(n + |\underline{x}^*\hat{z}|) \\ &= \|\underline{x} - \hat{z}\|_2^2(n + |\underline{x}^*\hat{z}|) \leq 2n\|\underline{x} - \hat{z}\|_2^2. \end{aligned}$$

Therefore, we may observe that

$$\|y\|_2^2 = \|\hat{z}\|_2^2 - \frac{1}{n^2}|\underline{x}^*\hat{z}|^2\|\underline{x}\|_2^2 = n - \frac{1}{n}|\underline{x}^*\hat{z}|^2 = \frac{\|\underline{x}\underline{x}^* - \hat{z}\hat{z}^2\|_F^2}{2n},$$

giving $\frac{1}{2}\|\underline{x} - \hat{z}\|_2^2 \leq \|y\|_2^2 \leq \|\underline{x} - \hat{z}\|_2^2$. In this way, since y is orthogonal to the null space of $\underline{L}$, we have

$$\hat{z}^* \underline{L} \hat{z} = y^* \underline{L} y \geq \lambda_2(\underline{L}) \|y\|_2^2 \geq \frac{\lambda_2(\underline{L})}{2} \|\underline{x} - \hat{z}\|_2^2. \tag{6.17}$$

Combining this with (6.16) completes the proof. □

6.3 Refined Error Guarantees

6.3.1 Main Result

In this section, we consider how the theory developed in section 6.2 can improve the results of sections 3.4 and 3.5, specifically seeking improvements over the main error bounds proven in theorems 4 and 5 and corollaries 2 and 3. Since the main contribution of this dissertation is to provide theoretical upper bounds on the reconstruction error of the phase retrieval algorithm proposed in chapter 3 and its derivatives, these results are among the most significant of this chapter. In particular, we will find in corollary 6 that the main recovery results of section 3.5 may be reduced by as much as pulling a factor of d/δ out of the "phase error" terms by using exact angular synchronization through SDP. Considering that we tend to assume $\delta \ll d$, this constitutes a significant sharpening of these bounds.

We begin by noticing a few different strategies we could take to improve theorem 4. First of all, if the noise level is low enough, by theorem 7 we can use SDP to obtain the *actual* minimizer of $\min_{y \in \mathbb{C}^d} \eta_{\widetilde{X}}(\text{sgn}(y))$. This will spare us the inequalities of (3.27), which ultimately cost us a factor of τ in the denominator of the error bound. Doing so would immediately improve the guarantee of corollary 2. In more specific terms, we prove the following:

Theorem 8 (See theorem 4)**.** *Suppose that $G = (V = [d], E)$ is an undirected, connected, and unweighted graph (so that $W_{ij} = \chi_{E(i,j)}$) with $\tau_G = \lambda_2(D - W)$. Let*

$\underline{x} \in (\mathbb{S}^1)^d$ be the ground truth vector, and set $\underline{\widetilde{X}} = W \circ \underline{x}\underline{x}^*$. Let $L = D - W \circ \widetilde{X}$ with $\widetilde{X} \in \mathcal{H}^d, |\widetilde{X}_{ij}| = 1$ for all $(i,j) \in E$, and suppose $x \in (\mathbb{S}^1)^d$ is an optimizer of (6.2). *If $\|\widetilde{X} - \underline{\widetilde{X}}\|_2 < \frac{\tau_G}{1+\sqrt{d}}$, then xx^* is the unique solution to* (6.5). *In any case, x satisfies the following inequalities:*

$$\min_{\theta \in \mathbb{R}} \|x - \mathrm{e}^{\mathrm{i}\theta}\underline{x}\|_2 \leq \frac{2\sqrt{2}\|\widetilde{X} - \underline{\widetilde{X}}\|_F}{\sqrt{\tau_G}} \tag{6.18}$$

6.3.2 τ_G vs. $\tau_N \min_{i \in V} \deg(i)$

Before proving theorem 8, we remark that the spectral gap used here is different from the τ used in theorem 4. There, we used $\tau_N = \lambda_2(I - D^{-1/2}WD^{-1/2})$; if we were to merely "delete" a $\sqrt{\tau_N}$ factor from the bound given in theorem 4, we would get something slightly different from (6.18); namely, we would arrive at

$$\min_{\theta \in \mathbb{R}} \|x - \mathrm{e}^{\mathrm{i}\theta}\underline{x}\|_2 \leq \frac{C\|\widetilde{X} - \underline{\widetilde{X}}\|_F}{\sqrt{\tau_N \min_{i \in V}(\deg(i))}}. \tag{6.19}$$

We notice that this differs from (6.18) only on which "spectral gap" appears in the denominator: $\tau_N \min_{i \in V} \deg(i)$ or τ_G. By inspection, when G is k-regular, $D = kI$ and these two coincide, since $D - W = kI(I - D^{-1/2}WD^{-1/2})$, giving $\tau_G = k\tau_N = \min_{i \in V} \deg(i)\tau_N$. However, when the vertices of G do *not* have constant degree, it is possible that $\tau_N \min_{i \in V} \deg(i) < \tau_G$; this makes sense, since the $\min_{i \in V} \deg(i)$ factor comes from lower-bounding the left-hand side in lemma 4 quite wastefully by

$$\sum_{i \in V} \deg(i)|g_i - \mathrm{e}^{\mathrm{i}\theta}|^2 \geq \min_{i \in V} \deg(i) \sum_{i \in V} |g_i - \mathrm{e}^{\mathrm{i}\theta}|^2.$$

By contrast, using τ_G "evenly mixes" the potentially varying degrees into the eigenvector problem. To make this precise, proposition 10 establishes that (6.18) is never worse than (6.19), and we follow it with an example to show a case where the improvement is strict.

Proposition 10. *Given a weighted graph $G = (V = [d], E)$ with weight matrix $W \in \mathcal{S}^d$ satisfying $W_{ij} \geq 0$ and $W_{ij} = 0$ iff $(i,j) \notin E$ and degree matrix D, set*

$$\tau_G = \lambda_2(D - W) \text{ and } \tau_N = \lambda_2(I - D^{-1/2} W D^{-1/2}).$$

Then $\tau_G \geq \tau_N \min_{i \in V} \deg(i)$.

Proof of proposition 10. We rely multiple times on the identity, for $A \in \mathcal{H}^d$,

$$\inf_{v \perp \mathrm{Nul}(A)} \frac{v^* A v}{v^* v} = \inf_v \sup_{w \in \mathrm{Nul}(A)} \frac{v^* A v}{(v-w)^*(v-w)},$$

which holds since $v^* A v = (v-w)^* A (v-w)$ and the denominator is minimized by taking $w = \mathrm{Proj}_{\mathrm{Nul}(A)} v$. When $\mathrm{Nul}(A) = \mathrm{span}(w)$, this reduces to

$$\inf_{v \perp w} \frac{v^* A v}{v^* v} = \inf_v \sup_t \frac{v^* A v}{(v - tw)^*(v - tw)}.$$

With this in mind, we set $L = D - W$ and $\mathcal{L} = D^{-1/2} L D^{-1/2}$ and use a change of variables $w = D^{-1/2} v$ to obtain

$$\begin{aligned}
\tau_N &= \inf_{v \perp D^{1/2}\mathbb{1}} \frac{v^* \mathcal{L} v}{v^* v} \\
&= \inf_v \sup_t \frac{v^* \mathcal{L} v}{(v - tD^{1/2}\mathbb{1})(v - tD^{1/2})} \\
&= \inf_w \sup_t \frac{(D^{1/2} w)^* \mathcal{L} D^{1/2} w}{(D^{1/2} w - tD^{1/2}\mathbb{1})^*(D^{1/2} w - tD^{1/2}\mathbb{1})} \\
&= \inf_w \sup_t \frac{w^* L w}{(w - t\mathbb{1})^* D (w - t\mathbb{1})} \\
&\leq \frac{1}{\min_{i \in V} \deg(i)} \inf_w \sup_t \frac{w^* L w}{(w - t\mathbb{1})^*(w - t\mathbb{1})} \\
&= \frac{1}{\min_{i \in V} \deg(i)} \inf_{w \perp \mathbb{1}} \frac{w^* L w}{w^* w} \\
&= \frac{\tau_G}{\min_{i \in V} \deg(i)}
\end{aligned}$$

□

We now present a sequence of unweighted graphs G_n such that $\tau_G(G_n) > \tau_N(G_n)$. G_n will be the complete graph on n vertices, K_n, with an extra vertex having exactly one edge. Namely, we may set $G_n = (V_n, E_n)$ with

$$\begin{aligned} V_n &= [n+1] \text{ and} \\ E_n &= \{(x,y) \in [n]^2 : x \neq y\} \cup \{(n, n+1), (n+1, n)\}. \end{aligned} \tag{6.20}$$

Then, considering that $D = \operatorname{diag}((n-1)\mathbb{1}^*_{n-1}, n, 1)$, the graph Laplacians $L, \mathcal{L} \in \mathbb{R}^{(n+1)\times(n+1)}$ of G_n become

$$L = D - W = \begin{bmatrix} nI_{n-1} - \mathbb{1}_{n-1}\mathbb{1}^*_{n-1} & -\mathbb{1}_{n-1} & 0_{n-1} \\ -\mathbb{1}^*_{n-1} & n & -1 \\ 0^*_{n-1} & -1 & 1 \end{bmatrix} \quad \text{and}$$

$$\mathcal{L} = I - D^{-1/2}WD^{-1/2} = \begin{bmatrix} \frac{n}{n-1}I_{n-1} - \frac{1}{n-1}\mathbb{1}_{n-1}\mathbb{1}^*_{n-1} & -\frac{1}{\sqrt{n(n-1)}}\mathbb{1}_{n-1} & 0_{n-1} \\ -\frac{1}{\sqrt{n(n-1)}}\mathbb{1}^*_{n-1} & 1 & -\frac{1}{\sqrt{n}} \\ 0^*_{n-1} & -\frac{1}{\sqrt{n}} & 1 \end{bmatrix}$$

We finish the example by proving proposition 11.

Proposition 11. *With $n > 3$ and $G_n = (V_n, E_n)$ defined as in (6.20), $\tau_N = \lambda_2(\mathcal{L}) < 1$ and $\tau_G = \lambda_2(L) = 1$. In particular, since $\min_{i\in V} \deg(i) = \deg(n+1) = 1$, we have $\tau_G > \tau_N \min_{i\in V} \deg(i)$.*

Proof. Take $c_1, \ldots, c_{n-1} \in \mathbb{R}$ such that $\sum_{i=1}^{n-1} c_i = 0$ and set $c = \sum_{i=1}^{n-1} c_i e_i^{n-1} \in \mathbb{R}^{n-1}$. Then $c \perp \mathbb{1}_{n-1}$ and

$$L \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} (nI_{n-1} - \mathbb{1}_{n-1}\mathbb{1}^*_{n-1})c \\ -\mathbb{1}^*_{n-1}c \\ 0 \end{bmatrix} = \begin{bmatrix} nc \\ 0 \\ 0 \end{bmatrix} = n \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}, \text{ while}$$

$$\mathcal{L} \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} (\frac{n}{n-1}I_{n-1} - \frac{1}{n-1}\mathbb{1}_{n-1}\mathbb{1}^*_{n-1})c \\ -\frac{1}{\sqrt{n(n-1)}}\mathbb{1}^*_{n-1}c \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{n}{n-1}c \\ 0 \\ 0 \end{bmatrix} = \frac{n}{n-1} \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix},$$

which identifies $\{\mathbb{1}^{n+1}_{[n-1]}, e_n, e_{n+1}\}^\perp$ as an $n-2$-dimensional eigenspace of both L and $\mathcal{L}$, with eigenvalues n and $\frac{n}{n-1}$, respectively (recall $\mathbb{1}^{n+1}_{[n-1]} = (1,\dots,1,0,0)^T$). Therefore, the other three eigenvectors (including the nullspace vectors $\mathbb{1}_{n+1}$ and $D^{1/2}\mathbb{1}_{n+1}$ of L and $\mathcal{L}$) are in $W = \text{span}\{\mathbb{1}^{n+1}_{[n-1]}, e_n, e_{n+1}\}$.

We state the remaining eigenvalues of L directly by giving their eigenvectors. Clearly $L\mathbb{1}_{n+1} = 0_{n+1}$. We observe additionally that

$$L\left(\begin{bmatrix}\mathbb{1}_{n-1}\\0\\0\end{bmatrix} + \begin{bmatrix}0_{n-1}\\0\\-(n-1)\end{bmatrix}\right) = \begin{bmatrix}\mathbb{1}_{n-1}\\-(n-1)\\0\end{bmatrix} + \begin{bmatrix}0_{n-1}\\n-1\\1-n\end{bmatrix} = \begin{bmatrix}\mathbb{1}_{n-1}\\0\\-(n-1)\end{bmatrix},$$

such that $(1,\dots,1,0,1-n)^T$ has an eigenvalue of 1, and

$$L\begin{bmatrix}\mathbb{1}_{n-1}\\-n\\1\end{bmatrix} = L(\mathbb{1}_{n+1} - (n+1)e_n) = -(n+1)Le_n = (n+1)\begin{bmatrix}\mathbb{1}_{n-1}\\-n\\1\end{bmatrix},$$

such that $(1,\dots,1,-n,1)^T$ has an eigenvalue of $n+1$. Therefore, the spectrum of L is n with a multiplicity of $n-2$, and $0, 1$, and $n+1$, each with multiplicity 1, so the spectral gap is $\tau_G = 1$ as stated.

To get the last three eigenvalues of $\mathcal{L}$, we take

$$M = \begin{bmatrix}\frac{1}{\sqrt{n-1}}\mathbb{1}^{n+1}_{[n-1]} & e_n & e_{n+1}\end{bmatrix}$$

as an orthogonal basis of W, such that the remaining eigenvalues of $\mathcal{L}$ are also the eigenvalues of $M^*\mathcal{L}M$, which we calculate to be

$$M^*\mathcal{L}M = \begin{bmatrix}\frac{1}{n-1} & -\frac{1}{\sqrt{n}} & 0\\ -\frac{1}{\sqrt{n}} & 1 & -\frac{1}{\sqrt{n}}\\ 0 & -\frac{1}{\sqrt{n}} & 1\end{bmatrix}.$$

We calculate the characteristic polynomial of $M^*\mathcal{L}M$ directly:

$$\det(M^*\mathcal{L}M - \lambda I) = -\lambda\left(\lambda^2 - \left(\frac{2n-1}{n-1}\right)\lambda + \frac{n^2-n+2}{n(n-1)}\right).$$

The two remaining nonzero eigenvalues may be obtained by the quadratic formula:

$$\lambda = \frac{\frac{2n-1}{n-1} \pm \sqrt{\left(\frac{2n-1}{n-1}\right)^2 - 4\frac{n^2-n+2}{n(n-1)}}}{2}.$$

τ_N is obtained by taking the negative square root, and we reduce its argument to $\frac{4n^2-11n+8}{n(n-1)^2}$. Now, to prove $\tau_N < 1$, it suffices to show

$$\begin{aligned} & \frac{2n-1}{n-1} - \frac{1}{n-1}\left(\frac{4n^2-11n+8}{n}\right)^{1/2} &< 2 \\ \iff & 1 &< \left(\frac{4n^2-11n+8}{n}\right)^{1/2} \\ \iff & 0 &< n^2-3n+2, \end{aligned}$$

which factors into $(n-2)(n-1)$ and is trivially positive when $n \geq 3$. □

6.3.3 Proof of Theorem 8

We proceed to the proof of theorem 8, which uses the following variation of lemma 5.

Lemma 21 (See lemma 5)**.** *Under the hypotheses of theorem 8, set* $g \in \mathbb{C}^d, \Lambda \in \mathbb{C}^{d\times d}$ *by*

$$g_i = (\underline{x})_i^* x_i \quad \textit{and} \quad \Lambda_{ij} = (\underline{\widetilde{X}})_{ij}^* \widetilde{X}_{ij}.$$

Then

$$\sum_{(i,j)\in E} |g_i - g_j|^2 \leq 4\|\widetilde{X} - \underline{\widetilde{X}}\|_F^2.$$

Proof of lemma 21. By an argument identical to that used in (3.26) of lemma 5, we have

$$\begin{aligned} \sum_{(i,j)\in E}\left(\frac{1}{2}|g_i-g_j|^2 - |\Lambda_{ij}-1|^2\right) &\leq \sum_{(i,j)\in E} |x_i - \widetilde{X}_{ij}x_j|^2 = x^*Lx \\ &\leq \underline{x}^*L\underline{x} \;=\; \sum_{(i,j)\in E} |\underline{x}_i - \widetilde{X}_{ij}\underline{x}_j|^2, \end{aligned}$$

where the second inequality comes from optimality of x. Additionally, we observe that, just as in (3.28),

$$\sum_{(i,j)\in E} |\Lambda_{ij} - 1|^2 = \sum_{(i,j)\in E} |\underline{x}_i - \widetilde{X}_{ij}\underline{x}_j|^2 = \left\|\widetilde{X} - \underline{\widetilde{X}}\right\|_F^2.$$

Combining these two immediately gives the lemma. □

Proof of theorem 8. We take $\underline{L} = D - W \circ \underline{\widetilde{X}}$ as usual and recognize that, as in (6.17),

$$x^* \underline{L} x \geq \frac{\tau_G}{2} \min_{\theta\in\mathbb{R}} \|x - \mathrm{e}^{\mathrm{i}\theta}\underline{x}\|_2^2.$$

Since we additionally have, as in (3.25), that

$$x^* \underline{L} x = \sum_{(i,j)\in E} |x_i - \underline{x}_i \underline{x}_j^* x_j|^2 = \sum_{(i,j)\in E} |g_i - g_j|^2,$$

this gives

$$\frac{\tau_G}{2} \min_{\theta\in\mathbb{R}} \|x - \mathrm{e}^{\mathrm{i}\theta}\underline{x}\|_2^2 \leq \sum_{(i,j)\in E} |g_i - g_j|^2,$$

which suffices with lemma 21 to complete the proof. □

Substituting these results into the proof of corollary 2, we gain an immediate improvement.

Corollary 5 (See corollary 2)**.** *Let $\widetilde{X}_0$ be the matrix in* (3.7)*, $\widetilde{\mathbf{x}}_0$ be the vector of true phases* (3.7)*, and $\widetilde{X}$ be as in line 3 of Algorithm 1 with $\widetilde{\mathbf{x}}$ the optimizer of* (6.2) *using $L = I - \frac{1}{2\delta-1}\widetilde{X}$. Suppose that $\|\widetilde{X}_0 - \widetilde{X}\|_F \leq \eta \|\widetilde{X}_0\|_F$ for some $\eta > 0$. Then*

$$\min_{\theta\in[0,2\pi]} \|\widetilde{\mathbf{x}}_0 - \mathrm{e}^{\mathrm{i}\theta}\widetilde{\mathbf{x}}\|_2 \leq 3\frac{\eta d^{\frac{3}{2}}}{\delta}.$$

Proof of corollary 5. We apply (6.18) with

$$\tau_G = \tau_N(2\delta - 1) \geq \frac{\pi^2\delta^2}{6d^2}(2\delta - 1) \text{ and } \|X - X_0\|_F \leq \eta\sqrt{d(2\delta - 1)},$$

and reduce the constant by observing $2\sqrt{2}/(\pi/\sqrt{6}) \leq 3$. □

This leads us to construct a variant of algorithm 1, replacing the eigenvector-based angular synchronization stage with an exact solution of the angular synchronization problem; this variant is stated in algorithm 2. With this in hand, we replace corollary 2 with corollary 5 in the proofs of the final error bounds, theorem 5 and corollary 3, and immediately obtain the following improvement over corollary 3.

Algorithm 2 Phase Retrieval from Local Correlation Measurements, with SDP

Input: Measurements $\mathbf{y} \in \mathbb{R}^D$ as per (3.8)

Output: $\mathbf{x} \in \mathbb{C}^d$ with $\mathbf{x} \approx \mathbb{e}^{-\mathbb{i}\theta}\mathbf{x}_0$ for some $\theta \in [0, 2\pi]$

1: Compute the Hermitian matrix $X = \left((\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})})^{-1}\mathbf{y}\right)/2 + \left((\mathcal{A}|_{T_\delta(\mathbb{C}^{d\times d})})^{-1}\mathbf{y}\right)^*/2 \in T_\delta(\mathbb{C}^{d\times d})$ as an estimate of $T_\delta(\mathbf{x}_0\mathbf{x}_0^*)$

2: Form the banded matrix of phases, $\tilde{X} \in T_\delta(\mathbb{C}^{d\times d})$, by normalizing the non-zero entries of X

3: Compute $\hat{Z}$, the solution to (6.5) with $L = (2\delta - 1)I - \widetilde{X}$, and take $\widetilde{\mathbf{x}} = \mathrm{sgn}(u)$, where u is the top eigenvector of $\hat{Z}$.

4: Set $x_j = \sqrt{X_{j,j}} \cdot (\tilde{x})_j$ for all $j \in [d]$ to form $\mathbf{x} \in \mathbb{C}^d$

Corollary 6 (See corollary 3)**.** *Using the notation of corollary 5, let* $(x_0)_{\min} := \min_j |(x_0)_j|$ *be the smallest magnitude of any entry in* $\mathbf{x}_0$ *and suppose* $\|\widetilde{X} - \widetilde{X}_0\|_2 < \delta^2/d^{5/2}$*. Then the estimate* $\mathbf{x}$ *produced by Algorithm 2 satisfies*

$$\min_{\theta\in[0,2\pi]} \|\mathbf{x}_0 - \mathbb{e}^{\mathbb{i}\theta}\mathbf{x}\|_2 \leq C\left(\frac{\|\mathbf{x}_0\|_\infty}{(x_0)_{\min}^2}\right)\left(\frac{d}{\delta}\right)\kappa\|\mathbf{n}\|_2 + Cd^{\frac{1}{4}}\sqrt{\kappa\|\mathbf{n}\|_2}, \tag{6.21}$$

where $C \in \mathbb{R}^+$ *is an absolute universal constant.*

6.3.4 Results with Weighted Graphs

While the extrication of a d/δ factor from this inequality is considerable, we remark first of all that theorem 7, besides establishing a guaranteed basin wherein (6.2) may be solved exactly by semidefinite programming, plays a disappointing role in

these new bounds. Specifically, (6.7) is not quoted in this section at all, which means that we have no clear statement of what the admission of weighted graphs in theorem 7 accomplishes for us.

This defies intuition, since weighting our cost function should push the angular synchronization solver to focus on getting the phases of the large magnitude entries of x correct, and to trust their phase measurements more (as discussed in section 6.2). However, the proof techniques we have used so far are not sophisticated enough to quantify these benefits, which is not surprising, seeing that they derive primarily from model-specific intuition. In particular, bounding $\||\mathbf{x}_0| \circ (\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}_0)\|_2$ by $\|\mathbf{x}_0\|_\infty \|\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}_0\|_2$ wastes any potential "cooperation" between the magnitudes of $\mathbf{x}_0$ the accuracy of the phases in $\widetilde{\mathbf{x}}$, but it is not straightforward to show how the weight matrix W would guarantee such cooperation.

In [46], the authors approach this a bit differently by bounding the same "phase error" term with $\|\mathbf{x}_0\|_2 \|\widetilde{\mathbf{x}}_0 - \widetilde{\mathbf{x}}\|_\infty$. The angular synchronization algorithm they use involves finding a spanning tree of G, which admits a fairly clean analysis of $\widetilde{\mathbf{x}}$, but the assumptions required to bound $\|\widetilde{\mathbf{x}}_0 - \widetilde{\mathbf{x}}\|_\infty$ are heavy. In section 6.4, we will take a closer look at such spanning tree methods and make an attempt to improve upon the results of this type as found in [46].

For the time being, to satiate these intuitions, we state the results that may be compiled from the existing theory in corollary 7, and perform in section 6.5 a numerical study that compares weighted vs. unweighted angular synchronization.

Corollary 7 makes use of the following lemma.

Lemma 22. *Suppose* $X, \underline{X} \in \mathcal{H}^d$. *Define* $x, \underline{x} \in \mathbb{R}^d$ *by* $x_i = \sqrt{|X_{ii}|}$ *and* $\underline{x}_i = \sqrt{|\underline{X}_{ii}|}$. *Then*

$$\|x - \underline{x}\|_2 \leq d^{1/4} \sqrt{\|X - \underline{X}\|_F}.$$

Proof of lemma 22. We first claim that, for $f(x) = (1 - |1 - x|^{1/2})^2$, we have $f(x) \leq |x|$ for all $x \in \mathbb{R}$. To see this for $x \geq 1$, we set $t = |x-1|^{1/2} = (x-1)^{1/2}$ and observe

$$x = t^2 + 1 \geq t^2 - 2t + 1 = (t-1)^2 = f(x).$$

For $0 \leq x \leq 1$, we set $t = |1-x|^{1/2} = (1-x)^{1/2}$ and see

$$x = 1 - t^2 = (1-t)(1+t) \geq (1-t)(1-t) = (1-t)^2 = f(x).$$

For $x \leq 0$, we consider $g(t) = f(-t)$ for $t \geq 0$. Then

$$g(t) = (1 - (1+t)^{1/2})^2 = 2 + t - 2\sqrt{1+t} \leq t,$$

simply by bounding $\sqrt{1+t} \geq 1$. From this, it follows that

$$\left(a - \left|a^2 - b\right|^{1/2}\right)^2 = a^2 \left(1 - \left|1 - \frac{b}{a^2}\right|^{1/2}\right)^2 \leq |b| \tag{6.22}$$

for any $a, b \in \mathbb{R}$.

Setting $N = X - \underline{X}$, we may write $X_{ij} = \underline{X}_{ij} + N_{ij}$. In particular, $X_{ii} = \underline{X}_{ii} + N_{ii} = \underline{x}_i^2 + N_{ii}$, so that $x_i = \sqrt{|\underline{x}_i^2 + N_{ii}|}$. Setting $n_i = N_{ii}$, (6.22) gives

$$||x_i| - |\underline{x}_i||^2 \leq |n_i|.$$

Trivially, we have $\|n\|_2 \leq \|X - \underline{X}\|_F$, so

$$\|x - \underline{x}\|_2 \leq \sqrt{\sum_{i=1}^{d} |n_i|} \leq \sqrt{\sqrt{d}\|n\|_2} \leq d^{1/4}\sqrt{\|X - \underline{X}\|_F},$$

as desired. □

Corollary 7. *Given $\underline{x} \in \mathbb{C}^d$, suppose $\underline{X} = \underline{x}\underline{x}^* \circ (I + A_G)$, where A_G is the unweighted adjacency matrix of the connected graph $G = (V = [d], E)$. Suppose further that $X \in \mathcal{H}^d$ shares the sparsity structure of $\underline{X}$ (namely, $X = X \circ (I + A_G)$). We then define W, D, and L by*

$$W_{ij} = \begin{cases} 0, & i = j \\ |X_{ij}|, & \textit{otherwise} \end{cases}, \quad D = \operatorname{diag}(W\mathbb{1}),$$

$$L = D - W \circ \operatorname{sgn}(X), \quad \textit{and} \quad \underline{L} = D - W \circ \operatorname{sgn}(\underline{X}).$$

Then, setting $x_i = |X_{ii}|^{1/2}\hat{z}_i$, *where*

$$\hat{z} \in \operatorname*{argmin}_{z \in (\mathbb{S}^1)^d} z^* L z,$$

we have

$$\min_{\theta\in[0,2\pi]} \|x - \mathrm{e}^{\mathrm{i}\theta}\underline{x}\|_2 \leq \|\underline{x}\|_\infty \frac{4\sqrt{2d}\|X - \underline{X}\|_2}{\tau_W} + d^{1/4}\sqrt{\|X - \underline{X}\|_F}, \tag{6.23}$$

where $\tau_W = \lambda_2(D - W)$.

Proof of corollary 7. We split the norm into

$$\begin{aligned}
\min_{\theta\in[0,2\pi]} \|x - \mathrm{e}^{\mathrm{i}\theta}\underline{x}\|_2 &\leq \||\underline{x}| \circ (\operatorname{sgn}(x) - \operatorname{sgn}(\underline{x}))\|_2 + \||\underline{x}| - |x|\|_2 \\
&\leq \|\underline{x}\|_\infty \|\operatorname{sgn}(x) - \operatorname{sgn}(\underline{x})\|_2 + \||\underline{x}| - |x|\|_2 \\
&\leq \|\underline{x}\|_\infty \frac{2\sqrt{2d}\|L - \underline{L}\|_2}{\tau_W} + d^{1/4}\sqrt{\|X - \underline{X}\|_F},
\end{aligned}$$

where the last inequality comes from theorem 7 and lemma 22. To complete the proof, we bound $\|L - \underline{L}\|_2$ by setting

$$\underline{W} = \begin{cases} 0, & i = j \\ |\underline{X}_{ij}|, & \text{otherwise} \end{cases}$$

and taking

$$\begin{aligned}
\|L - \underline{L}\|_2 &= \|W \circ (\operatorname{sgn} X - \operatorname{sgn} \underline{X})\|_2 \\
&\leq \|W \circ \operatorname{sgn} X - \underline{W} \circ \operatorname{sgn} \underline{X}\|_2 + \|(W - \underline{W}) \circ \operatorname{sgn} \underline{X}\|_2 \\
&\leq \|X - \underline{X}\|_2 + \|W - \underline{W}\|_2 \\
&\leq 2\|X - \underline{X}\|_2
\end{aligned}$$

The second inequality in this series comes from considering that $W \circ \operatorname{sgn} X - \underline{W} \circ \operatorname{sgn} \underline{X}$ is simply $X - \underline{X}$ without its diagonal entries, and the third comes from $||a| - |b|| \leq |a - b|$ for any $a, b \in \mathbb{C}$. □

Comparing eqs. (6.21) and (6.23), we notice that (6.23) uses yet another variation on the spectral gap: $\tau_W = \lambda_2(D - W)$, the spectral gap of the graph weighted by the entries of X. Not only do we see τ_W in the denominator rather than $\sqrt{\tau_W}$ (we fear, but do not necessarily expect, spectral gaps to be small), but τ_W is a massively unpredictable quantity, possibly having very little to do with the unweighted graph sitting beneath it. Indeed, even if $G = K_n$ is a complete graph, by taking another graph $G' = ([n], E')$ and weighting G with

$$W_{ij} = \begin{cases} 1, & (i,j) \in E' \\ \epsilon, & \text{otherwise} \end{cases},$$

then $\lim_{\epsilon \to 0} \tau_W(\epsilon) = \tau_{G'}$, where $\tau_{G'} = D_{G'} - A_{G'}$ is the unweighted spectral gap of G'. In other words, weak entries (and, therefore, small weights) of X have the potential to effectively *disconnect* – meaning "drastically reduce the spectral gap of" – the graph on which we are depending for our angular synchronization. In principle, (6.21) is punished for small entries of X in the term $(x_0)^2_{\min}$ appearing in the denominator of the phase error bound, but it is hard to gain a theoretical statement telling us whether τ_W is a more efficient means of quantifying the extent to which noise and small entries of $\underline{x}$ "disconnect" our angular synchronization graph. For the moment, we relegate this question to the numerical study of section 6.5.

6.4 Spanning Tree Strategies

Triz

6.5 Numerical Experiments

Artzen und craftzen! Chartzen und graphzen!

Appendix A

Sample Appendix

If you seek a pleasant appendix, look no further.

References

[1] S. Abrahamsson, D. Hodgkin, and E. Maslen. The crystal structure of phenoxymethylpenicillin. *Biochemical Journal*, 86(3):514–535, 1963. ISSN 0264-6021. doi: 10.1042/bj0860514. URL http://www.biochemj.org/content/86/3/514.

[2] 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.

[3] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Complementarity and nondegeneracy in semidefinite programming. *Mathematical Programming*, 77 (1):111–128, Apr 1997. ISSN 1436-4646. doi: 10.1007/BF02614432. URL https://doi.org/10.1007/BF02614432.

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

[5] R. Balan, B. Bodmann, P. Casazza, and D. Edidin. Fast algorithms for signal reconstruction without phase. In *Optical Engineering+ Applications*, pages 67011L–67011L. International Society for Optics and Photonics, 2007.

[6] R. Balan, B. G. Bodmann, P. G. Casazza, and D. Edidin. Painless reconstruction from magnitudes of frame coefficients. *Journal of Fourier Analysis and Applications*, 15(4):488–501, 2009.

[7] A. S. Bandeira and D. G. Mixon. Near-optimal phase retrieval of sparse vectors. In *Wavelets and Sparsity XV*, volume 8858, page 88581O. International Society for Optics and Photonics, 2013.

[8] A. S. Bandeira, A. Singer, and D. A. Spielman. A Cheeger inequality for the graph connection Laplacian. *SIAM Journal on Matrix Analysis and Applications*, 34(4):1611–1630, 2013. doi: 10.1137/120875338.

[9] A. S. Bandeira, N. Boumal, and A. Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, 163(1):145–167, May 2017. ISSN 1436-4646. doi: 10.1007/s10107-016-1059-6. URL https://doi.org/10.1007/s10107-016-1059-6.

[10] H. Bauschke, P. Combettes, and D. Luke. Hybrid projection-reflection method for phase retrieval. *Journal of the Optical Society of America. A, Optics, Image science, and Vision*, 20(6):1025–1034, 2003.

[11] 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.

[12] T. Bendory, Y. C. Eldar, and N. Boumal. Non-convex phase retrieval from STFT measurements. *IEEE Transactions on Information Theory*, 64(1):467–484, 2018.

[13] B. G. Bodmann and N. Hammen. Stable phase retrieval with low-redundancy frames. *Advances in computational mathematics*, 41(2):317–331, 2015.

[14] W. H. Bragg and L. Bragg. *X-rays and crystal structure*. G. Bell and Sons, Ltd London, 1915.

[15] J. Briales and J. Gonzalez-Jimenez. Cartan-sync: Fast and global $SE(d)$-synchronization. *IEEE Robotics and Automation Letters*, 2(4):2127–2134, Oct 2017. ISSN 2377-3766. doi: 10.1109/LRA.2017.2718661.

[16] D. Brodersen, W. Clemons, A. P. Carter, R. J. Morgan-Warren, B. Wimberly, and V. Ramakrishnan. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin b on the 30s ribosomal subunit. *Cell*, 103:1143–54, 01 2001.

[17] G. C. Calafiore, L. Carlone, and F. Dellaert. *Lagrangian Duality in Complex Pose Graph Optimization*, pages 139–184. Springer International Publishing, Cham, 2016. ISBN 978-3-319-42056-1. doi: 10.1007/978-3-319-42056-1_5. URL https://doi.org/10.1007/978-3-319-42056-1_5.

[18] E. J. Candes and X. Li. Solving quadratic equations via Phaselift when there are about as many equations as unknowns. *Foundations of Computational Mathematics*, 14(5):1017–1026, 2014. ISSN 1615-3375.

[19] E. J. Candes, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

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

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

[22] F. R. Chung. *Spectral Graph Theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, 1997.

[23] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[24] M. Dierolf, O. Bunk, S. Kynde, P. Thibault, I. Johnson, A. Menzel, K. Jefimovs, C. David, O. Marti, and F. Pfeiffer. Ptychography & lensless X-ray imaging. *Europhysics News*, 39(1):22–24, 2008.

[25] Y. Eldar, P. Sidorenko, D. Mixon, S. Barel, and O. Cohen. Sparse phase retrieval from short-time fourier measurements. *IEEE Signal Proc. Letters*, 22(5), 2015.

[26] Y. C. Eldar and S. Mendelson. Phase retrieval: Stability and recovery guarantees. *Applied and Computational Harmonic Analysis*, 36(3):473–494, 2014.

[27] V. Elser. Phase retrieval by iterated projections. *Journal of the Optical Society of America. A, Optics, Image science, and Vision*, 20(1):40–55, 2003.

[28] O. Enqvist, F. Kahl, and C. Olsson. Non-sequential structure from motion. In *Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*, 2011.

[29] A. P. Eriksson, C. Olsson, F. Kahl, O. Enqvist, and T. Chin. Why rotation averaging is easy. *CoRR*, abs/1705.01362, 2017. URL http://arxiv.org/abs/1705.01362.

[30] J. R. Fienup. Reconstruction of an object from the modulus of its Fourier transform. *Optics letters*, 3(1):27–29, 1978.

[31] J. R. Fienup. Phase retrieval algorithms: A comparison. *Applied Optics*, 21 (15):2758–2769, 1982.

[32] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography.

Commun. ACM, 24(6):381–395, June 1981. ISSN 0001-0782. doi: 10.1145/358669.358692. URL http://doi.acm.org/10.1145/358669.358692.

[33] R. E. Franklin and R. G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, Apr 1953. URL http://dx.doi.org/10.1038/171740a0.

[34] R. G. Gallager. *Principles of Digital Communication*. Cambridge University Press, 2008. doi: 10.1017/CBO9780511813498.

[35] S. Galli. X-ray crystallography: One century of Nobel prizes. *Journal of Chemical Education*, 91(12):2009–2012, 2014. doi: 10.1021/ed500343x. URL https://doi.org/10.1021/ed500343x.

[36] R. Gerchberg and W. Saxton. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik*, 35:237246, 1972.

[37] J. W. Goodman. *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.

[38] V. M. Govindu. Robustness in motion averaging. In P. J. Narayanan, S. K. Nayar, and H. Shum, editors, *Computer Vision - ACCV 2006, 7th Asian Conference on Computer Vision, Hyderabad, India, January 13-16, 2006, Proceedings, Part II*, volume 3852 of *Lecture Notes in Computer Science*, pages 457–466. Springer, 2006. ISBN 3-540-31244-7. doi: 10.1007/11612704_46. URL https://doi.org/10.1007/11612704_46.

[39] 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, 2008. http://stanford.edu/~boyd/graph_dcp.html.

[40] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, Mar. 2014.

[41] D. Gross, F. Krahmer, and R. Kueng. Improved recovery guarantees for phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 2015.

[42] H. A. Hauptman and J. Karle. *Solution of the phase problem,*. American Crystallographic Association, Ann Arbor, Mich., 1953.

[43] R. Hausbrand, G. Cherkashinin, H. Ehrenberg, M. Grting, K. Albe, C. Hess, and W. Jaegermann. Fundamental degradation mechanisms of layered oxide Li-ion battery cathode materials: Methodology, insights and novel approaches. *Materials Science and Engineering: B*, 192:3 – 25, 2015. ISSN 0921-5107. doi: https://doi.org/10.1016/j.mseb.2014.11.014. URL http://www.sciencedirect.com/science/article/pii/S0921510714002657. Electrical Fatigue.

[44] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

[45] M. Iwen, F. Krahmer, and A. Viswanathan. Technical note: A minor correction of Theorem 1.3 from [1]. *Unpublished note available at http://users.math.msu.edu/users/markiwen/Papers/PhaseLiftproof.pdf*, April 2015.

[46] M. Iwen, A. Viswanathan, and Y. Wang. Fast phase retrieval from local correlation measurements. *SIAM Journal on Imaging Sciences*, 9(4):1655–1688, 2016.

[47] 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.

[48] K. Jaganathan, Y. C. Eldar, and B. Hassibi. STFT phase retrieval: Uniqueness guarantees and recovery algorithms. *IEEE Journal of selected topics in signal processing*, 10(4):770–781, 2016.

[49] M. S. Kimber, F. Vallee, S. Houston, A. Neakov, T. Skarina, E. Evdokimova, S. Beasley, D. Christendat, A. Savchenko, C. H. Arrowsmith, M. Vedadi, M. Gerstein, and A. M. Edwards. Data mining crystallization databases: Knowledge-based approaches to optimize protein crystal screens. *Proteins: Structure, Function, and Bioinformatics*, 51(4):562–568, 2003. doi: 10.1002/prot.10340. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.10340.

[50] W. Krätschmer, L. D. Lamb, K. Fostiropoulos, and D. R. Huffman. Solid C_{60}: a new form of carbon. *Nature*, 347:354 EP –, Sep 1990. URL http://dx.doi.org/10.1038/347354a0.

[51] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley. C_{60}: Buckminsterfullerene. *Nature*, 318:162 EP –, Nov 1985. URL http://dx.doi.org/10.1038/318162a0.

[52] A. J. Laub. *Matrix Analysis For Scientists And Engineers*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004. ISBN 0898715768.

[53] X. Li and V. Voroninski. Sparse signal recovery from quadratic measurements via convex programming. *SIAM Journal on Mathematical Analysis*, 45(5): 3019–3033, 2013.

[54] X. Li, S. Ling, T. Strohmer, and K. Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *Applied and Computational Harmonic Analysis*, 2018. ISSN 1063-5203. doi: https://doi.org/10.1016/

j.acha.2018.01.001. URL http://www.sciencedirect.com/science/article/pii/S1063520318300149.

[55] S. Ling and T. Strohmer. Self-calibration and biconvex compressive sensing. *Inverse Problems*, 31(11):115002, 2015. URL http://stacks.iop.org/0266-5611/31/i=11/a=115002.

[56] S. Ling and T. Strohmer. Regularized gradient descent: a non-convex recipe for fast joint blind deconvolution and demixing. *Information and Inference: A Journal of the IMA*, page iax022, 2018. doi: 10.1093/imaiai/iax022. URL http://dx.doi.org/10.1093/imaiai/iax022.

[57] S. Marchesini, Y.-C. Tu, and H.-t. Wu. Alternating projection, ptychographic imaging and phase synchronization. *Applied and Computational Harmonic Analysis*, 41(3):815–851, 2016.

[58] 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.

[59] R. Millane. Phase retrieval in crystallography and optics. *Journal of the Optical Society of America. A, Optics, Image science, and Vision*, 7(3):394–411, 1990.

[60] P. Netrapalli, P. Jain, and S. Sanghavi. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, pages 2796–2804, 2013.

[61] A. L. Patterson. A fourier series method for the determination of the components of interatomic distances in crystals. *Phys. Rev.*, 46:372–376, Sep 1934. doi: 10.1103/PhysRev.46.372. URL https://link.aps.org/doi/10.1103/PhysRev.46.372.

[62] G. E. Pfander and P. Salanevich. Robust phase retrieval algorithm for time-frequency structured measurements. *eprint arXiv:1611.02540*, 2016.

[63] L. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993. ISBN 0-13-015157-2.

[64] S. G. F. Rasmussen, H.-J. Choi, D. M. Rosenbaum, T. S. Kobilka, F. S. Thian, P. C. Edwards, M. Burghammer, V. R. P. Ratnala, R. Sanishvili, R. F. Fischetti, G. F. X. Schertler, W. I. Weis, and B. K. Kobilka. Crystal structure of the human β_2 adrenergic g-protein-coupled receptor. *Nature*, 450: 383 EP –, Oct 2007. URL http://dx.doi.org/10.1038/nature06325. Article.

[65] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard. *SE*-sync: A certifiably correct algorithm for synchronization over the special euclidean group. *CoRR*, abs/1611.00128, 2016. URL http://arxiv.org/abs/1611.00128.

[66] P. Salanevich and G. E. Pfander. Polarization based phase retrieval for time-frequency structured measurements. In *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pages 187–191. IEEE, 2015.

[67] T. Schindler, W. Bornmann, P. Pellicena, W. T. Miller, B. Clarkson, and J. Kuriyan. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science*, 289(5486):1938–1942, 2000. ISSN 0036-8075. doi: 10.1126/science.289.5486.1938. URL http://science.sciencemag.org/content/289/5486/1938.

[68] A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1):20 – 36, 2011. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2010.02.001. URL http://www.sciencedirect.com/science/article/pii/S1063520310000205.

[69] A. Singer, M. Zhang, S. Hy, D. Cela, C. Fang, T. A. Wynn, B. Qiu, Y. Xia, Z. Liu, A. Ulvestad, N. Hua, J. Wingert, H. Liu, M. Sprung, A. V. Zozulya,

E. Maxey, R. Harder, Y. S. Meng, and O. G. Shpyrko. Nucleation of dislocations and their dynamics in layered oxide cathode materials during battery charging. *Nature Energy*, 3(8):641–647, 2018. ISSN 2058-7546. doi: 10.1038/s41560-018-0184-2. URL https://doi.org/10.1038/s41560-018-0184-2.

[70] G. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, 1990.

[71] H. Takajo, T. Takahashi, H. Kawanami, and R. Ueda. Numerical investigation of the iterative phase-retrieval stagnation problem: Territories of convergence objects and holes in their boundaries. *Journal of the Optical Society of America. A, Optics, Image science, and Vision*, 14(12):3175–3187, 1997.

[72] H. Takajo, T. Takahashi, R. Ueda, and M. Taninaka. Study on the convergence property of the hybrid input–output algorithm used for phase retrieval. *Journal of the Optical Society of America. A, Optics, Image science, and Vision*, 15(11):2849–2861, 1998.

[73] H. Takajo, T. Takahashi, and T. Shizuma. Further study on the convergence property of the hybrid input–output algorithm used for phase retrieval. *Journal of the Optical Society of America. A, Optics, Image science, and Vision*, 16(9):2163–2168, 1999.

[74] Y. L. Tong. *The multivariate normal distribution*. Springer-Verlag, New York, 1990.

[75] L. N. Trefethen and D. Bau III. *Numerical Linear Algebra*, volume 50. SIAM, 1997.

[76] L. Varsani, T. Cui, M. Rangarajan, B. S. Hartley, J. Goldberg, C. Collyer, and D. M. Blow. Arthrobacter d-xylose isomerase: protein-engineered subunit interfaces. *Biochemical Journal*, 291(2):575–583, 1993. ISSN 0264-6021. doi: 10.1042/bj2910575. URL http://www.biochemj.org/content/291/2/575.

[77] A. Viswanathan and M. Iwen. Fast angular synchronization for phase retrieval via incomplete information. In *Wavelets and Sparsity XVI*, volume 9597, page 959718. International Society for Optics and Photonics, 2015.

[78] A. Walther. The Question of Phase Retrieval in Optics. *Optica Acta*, 10: 41–49, 1963. doi: 10.1080/713817747.

[79] J. D. Watson. The involvement of RNA in the synthesis of proteins. In *Nobel Lectures, Physiology or Medicine 1942-1962*. Elsevier Publishing Company, December 1962.

[80] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, Apr 1953. URL http://dx.doi.org/10.1038/171737a0.

[81] J. H. M. Wedderburn. *Lectures on matrices*. American Mathematical Society New York, 1934.

[82] M. H. F. Wilkins, A. R. Stokes, and H. R. Wilson. Molecular structure of nucleic acids: Molecular structure of deoxypentose nucleic acids. *Nature*, 171: 738–740, Apr 1953. URL http://dx.doi.org/10.1038/171738a0.

[83] C. Yang, J. Qian, A. Schirotzek, F. Maia, and S. Marchesini. Iterative Algorithms for Ptychographic Phase Retrieval. *ArXiv e-prints*, May 2011.

[84] Y. Yu, T. Wang, and R. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.