Multi-Frequency Phase Synchronization

Tingran Gao [* 1]   **Zhizhen Zhao** [* 2]

Abstract

We propose a novel formulation for *phase synchronization*—the statistical problem of jointly estimating alignment angles from noisy pairwise comparisons—as a nonconvex optimization problem that enforces consistency among the pairwise comparisons in multiple frequency channels. Inspired by *harmonic retrieval* in signal processing, we develop a simple yet efficient two-stage algorithm that leverages the multi-frequency information. We demonstrate in theory and practice that the proposed algorithm significantly outperforms state-of-the-art phase synchronization algorithms, at a mild computational costs incurred by using the extra frequency channels. We also extend our algorithmic framework to general synchronization problems over compact Lie groups.

1. Introduction

Angular or *phase synchronization* (Singer, 2011; Boumal, 2016) concerns estimating angles $\theta_1, \ldots, \theta_n$ in $[0, 2\pi)$ from a subset of possibly noise-contaminated relative offsets $(\theta_i - \theta_j) \bmod 2\pi$. An instance of phase synchronization can be encoded on an *observation graph* $G = (V, E)$, where each angle is assigned to a vertex $i \in V$ and relative offsets are measured between θ_i and θ_j if and only if there is an edge in G connecting vertices i and j. Equivalently, the angles can be encoded into a column *phase vector* $z = (\exp \iota\theta_1, \cdots, \exp \iota\theta_n)^\top$, and measurements constitute a Hermitian matrix

$$H = A \circ [zz^* + \Delta], \tag{1}$$

where A is the adjacency matrix of the observation graph G, $\circ$ is the entrywise product, and the Hermitian matrix

---

[*]Equal contribution [1]Committee on Computational and Applied Mathematics, Department of Statistics, University of Chicago, Chicago IL, USA [2]Department of Electrical and Computer Engineering, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana IL, USA. Correspondence to: Tingran Gao <tingrangao@galton.uchicago.edu>, Zhizhen Zhao <zhizhenz@illinois.edu>.

$\Delta \in \mathbb{C}^{n \times n}$ encodes measurement noise.

As a prototypical example of more general *synchronization problems* arising from many scientific fields concerning consistent pairwise comparisons within large collections of objects (e.g., cryogenic electron microscopy (Singer et al., 2011) and comparative biology (Gao et al., 2019)), phase synchronization attracted much attention due to its simple yet rich mathematical structure. One mathematical formulation is through nonconvex optimization

$$\max_{x \in \mathbb{C}_1^n} x^* H x \tag{2}$$

where $\mathbb{C}_1^n$ is the Cartesian product of n copies of U(1). Depending on the context of the scientific problem, H may be assumed to arise from an *additive Gaussian noise model* (Boumal, 2016; Bandeira et al., 2017), in which the Hermitian matrix Δ in (1) is a Wigner matrix with i.i.d. complex Gaussian entries above the diagonal, or from a *random corruption model* (Singer, 2011; Chen et al., 2016) that assumes

$$H_{ij} = \begin{cases} z_i \bar{z}_j & \text{with prob. } r \in [0, 1] \\ w \sim \text{Unif}\,(\text{U}(1)) & \text{with prob. } 1 - r \end{cases} \tag{3}$$

for each edge $(i, j) \in E$. Note that the random corruption model can also be cast in the form (1) after proper shifting and scaling. In general, the additive Gaussian noise model is more amenable to analysis, while the random corruption model is better at capturing the behavior of physical or imaging models where many outliers exist.

In this paper, we propose to tackle the phase synchronization problem by solving an alternative nonconvex optimization problem of "multi-frequency" nature, namely,

$$\max_{x \in \mathbb{C}_1^n} \sum_{k=1}^{k_{\max}} (x^k)^* H^{(k)} x^k \tag{4}$$

where $k_{\max}$ is the number of frequency channels, x^k is the entrywise kth power of x, and $H^{(k)} \in \mathbb{C}^{n \times n}$ is a Hermitian matrix containing information of the "true signal" z in the kth frequency component:

- For the random corruption model (3), we construct $H^{(k)}$ directly from H by entrywise power:

$$H_{ij}^{(k)} = H_{ij}^k, \quad k = 1, \ldots, k_{\max}; \ 1 \leq i, j \leq n. \tag{5}$$

- For the additive Gaussian noise model, following (Bandeira et al., 2015; Perry et al., 2018), we assume

$$H^{(k)} = A \circ [z^k (z^*)^k + \sigma_k \Delta^{(k)}], \qquad (6)$$

where each $\Delta^{(k)}$ is a complex Hermitian random matrix with independent upper diagonal entries, and the scaling σ_k is chosen such that the operator norm of $\Delta^{(k)}$ is upper bounded by $\sqrt{n}$. Unlike (Bandeira et al., 2015; Perry et al., 2018), we allow entries of $\Delta^{(k)}$ to be general sub-Gaussian random variables rather than restrictively complex Gaussian, and we do not assume independence of the $\Delta^{(k)}$'s across different k's.

We treat the two types of noise (5) (6) in a unified model, under which we design and analyze our *multi-frequency phase synchronization* algorithm. We demonstrate surprising theoretical and empirical results that drastically outperform all existing phase synchronization algorithms in their corresponding settings, measured in terms of the correlation between the output and the true phase vector z, at a mild increase in the computational cost incurred by parallelizing the computation in $k_{\max}$ frequency channels. As will be demonstrated in Section 4, in the noise regime where phase synchronization is tractable, the number of frequencies $k_{\max}$ needed to outperform single frequency algorithms is at most polylogarithmically dependent on the problem size n, while the estimation error decays polynomially in $k_{\max}$.

Motivation   The rationale behind the multi-frequency formulation (4) lies at the observation that statistical estimation can often benefit from higher moment estimates, even without introducing new measurements. As a motivating example, let G be a complete graph, and consider the following $k_{\max} = 2$ coupled problems:

$$\max_{x \in \mathbb{C}_1^n} x^* H^{(1)} x; \qquad \max_{x \in \mathbb{C}_1^n} (x^2)^* H^{(2)} x^2 \qquad (7)$$

where $H^{(1)} = H$, and $H^{(2)}$ is generated according to (5). Up to rescaling by a factor of $1/r$, $H^{(1)}$ and $H^{(2)}$ fit into model (6) with

$$\sigma_k \Delta_{ij}^{(k)} = \begin{cases} (1-r) z_i \bar{z}_j & \text{with prob. } r \in [0,1] \\ e^{\iota k \varphi_{ij}} - r z_i \bar{z}_j & \text{with prob. } 1-r \end{cases} \qquad (8)$$

where φ_{ij} are i.i.d. uniform on $\mathbb{R}/2\pi$ for $(i,j) \in E$, and $\varphi_{ji} = -\varphi_{ij}$. Note that $\sigma_1 \Delta^{(1)}$ and $\sigma_2 \Delta^{(2)}$ are by no means independent, but for all practical purposes satisfy the same sub-Gaussian bounds since $e^{\iota \varphi_{ij}}$ and $e^{\iota 2 \varphi_{ij}}$ are identically distributed; we thus assume without loss of generality that $\sigma_1 = \sigma_2$. If we can find $\hat{x} \in \mathbb{C}_1^n$ satisfying jointly

$$(\hat{x}^k)^* H^{(k)} \hat{x}^k \geq (z^k)^* H^{(k)} z^k \ k = 1, 2 \qquad (9)$$

then, by Lemma 1 of (Boumal, 2016) (assuming without loss of generality that $\hat{x}^* z = |\hat{x}^* z|$), we have for $k = 1, 2$

$$\left\| \hat{x}^k - z^k \right\|_2^2 = 2(n - |z^* \hat{x}|^k) \leq 16 \sigma^2 \|\Delta^{(k)}\|_2^2 / n, \quad (10)$$

which gives $|z^* \hat{x}| \geq \max_{k=1,2} \left\{ (n - 8\sigma^2 \|\Delta^{(k)}\|_2^2/n)^{\frac{1}{k}} \right\}$, a tighter bound than one could obtain from (9) with $k = 1$ alone, especially for large σ (with $n - 8\sigma^2 \|\Delta^{(k)}\|_2^2/n < 1$).

The lesson we learn from this motivating example is that statistical estimation can benefit from leveraging higher-order moment information, even when the moment measurements are not essentially independent of each other. This is particularly prominent for the random corruption model, where all the "higher-order trigonometric moments" in $H^{(k)}$, $k > 1$ come from the first moments in $H = H^{(1)}$ by taking entrywise powers. In drastic contrast is the message-passing algorithm in (Perry et al., 2018), for which independence of the complex Gaussian Wigner noise $\Delta^{(k)}$'s across the frequency channels play an essential role. The AMP approach was motivated by the *non-unique games* (NUG) framework in (Bandeira et al., 2015). Our algorithm follows an efficient two-stage paradigm (*initialization* and *iterative refinement*) popularized by recent progress in nonconvex optimization (see, e.g. (Candes et al., 2015; Chen & Candes, 2015)), and combines the trigonometric moments information across frequency channels in a manner akin to classical *harmonic retrieval* techniques in signal processing (Stoica & Moses, 1997; Tufts & Kumaresan, 1982; Bresler & Macovski, 1986; Ziskind & Wax, 1988; Schmidt, 1986; Roy & Kailath, 1989; Sorensen & De Lathauwer, 2017a;b) and the *generalized power method* (Boumal, 2016). This strategy easily extends to synchronization over general compact Lie groups, as illustrated in Section 5.

Notations   Upper case letters $A, B, C, \cdots$ and lower case letters $a, b, c, \cdots$ will be used to denote matrices and vectors, respectively. A^*, $A^\top$ are the transpose of A with or without conjugation, respectively. The entrywise (Hadamard) product of matrix A and B will be denoted as $A \circ B$. Graphs $G = (V, E)$ are always undirected and connected. Vertices of the graph will be denoted as integers $1, 2, \cdots, |V|$; pairs of integers (i, j) denote edges in E. For $n \in \mathbb{N}$ we write $[n] := \{1, \cdots, n\}$. Norms $\|\cdot\|_2$, $\|\cdot\|_\infty$ stand for matrix or vector norms, depending on the context; $\|\cdot\|_{\mathrm{op}}$, $\|\cdot\|_{\mathrm{F}}$ are matrix operator and Frobenius norms, respectively. The Cartesian product of n copies of U(1) is denoted as $\mathbb{C}_1^n$. The quotient space $\mathbb{R}/2\pi$ is identified with the unit circle.

2. Related Work

Phase synchronization   Directly solving (2) is NP-hard (Zhang & Huang, 2006), but many convex and nonconvex methods have been proposed to find high quality approximate solutions. These include spectral and semi-definite programming (SDP) relaxations (Singer, 2011; Cucuringu et al., 2012; Chaudhury et al., 2015; Bandeira et al., 2016; 2017). An alternative approach using *generalized power method* (GPM) is also studied (Boumal, 2016; Liu et al., 2017; Zhong & Boumal, 2018).

Phase synchronization in multiple frequency channels (Bandeira et al., 2015) proposed the *non-unique games* (NUG) SDP optimization framework for synchronization over compact Lie groups. The SDP is based on quadratically lifting the irreducible representations of the group elements, and imposing consistency among variables across frequency channels via a Féjer kernel; it is computationally expensive. (Perry et al., 2018) introduced an iterative approximate message passing (AMP) algorithm for noise model (6), assuming the noise are Gaussian and independent across frequency channels. Each iteration of the AMP performs matrix-vector multiplication and entrywise nonlinear transformation, followed by an extra Onsager correction term; it is conjectured to be asymptotically optimal.

3. Algorithm

In this section we formally state the two-stage multifrequency phase synchronization algorithmic paradigm. Stage One combines phase synchronization outcomes from individual frequency channels with harmonic retrieval, aiming at producing a high-quality initialization; Stage Two iteratively refines an input by an extended generalized power method that works concurrently in multiple frequency channels while striving to maintain entrywise consistency.

3.1. Stage One: Initialization Strategy

Our algorithm takes as input $k_{\max}$ Hermitian measurement matrices $H^{(k)}$, $k = 1, \ldots, k_{\max}$, arising from the general sub-Gaussian model (6) (which includes (5) as a special case). This stage can be divided into three steps.

Step 1. *Individual Frequency Synchronization*: Apply any phase synchronization algorithm (spectral/SDP relaxation or GPM) to get phase vector estimate $u^{(k)} \in \mathbb{C}^n$ from each $H^{(k)}$, $k = 1, \cdots, k_{\max}$, and form $W^{(k)} := u^{(k)}(u^{(k)})^*$;

Step 2. *Entrywise Harmonic Retrieval*: For each $(i, j) \in E$, use any harmonic retrieval technique to estimate $\theta_i - \theta_j$ from $W_{ij}^{(k)}$, $k = 1, 2, \cdots, k_{\max}$, call the estimators $\hat{\theta}_{ij}$;

Step 3. *Final Phase Synchronization*: Construct another Hermitian matrix $\widehat{H} \in \mathbb{C}^{n \times n}$ by $\widehat{H}_{ij} := e^{\iota \hat{\theta}_{ij}}$, and apply any phase synchronization algorithm to estimate the true phases $\{e^{\iota \theta_1}, \cdots, e^{\iota \theta_n}\}$ from matrix $\widehat{H}$.

The flexibility of the multi-frequency phase synchronization framework lies at the various choices to be made in each step. As a concrete example, we detail in Algorithm 1 a simple version that uses spectral relaxation for phase synchronization and periodogram-based harmonic retrieval. We will henceforth refer to Algorithm 1 as the *periodogram peak extraction with spectral methods* (PPE-SPC). If a different phase synchronization method is used, for instance, SDP relaxation, our nomenclature refers to it as PPE-SDP. We

---

Algorithm 1 Periodogram Peak Extraction with Spectral Methods (PPE-SPC)

Input: Hermitian matrices $\{H^{(k)} \mid 1 \leq k \leq k_{\max}\}$
Output: Initialization $\hat{x} \in \mathbb{C}_1^n$
Step 1: Individual Frequency Synchronization

---
for $k = 1$ **to** $k_{\max}$ **do**
  Extract the leading eigenvector $u^{(k)}$ of $H^{(k)}$ with scaling $\|u^{(k)}\|_2 = \sqrt{n}$
  $W^{(k)} \leftarrow u^{(k)}(u^{(k)})^*$
end for
Step 2: Entrywise Harmonic Retrieval

---
for $i = 1$ **to** n **do**

$$\hat{\theta}_{ij} \leftarrow \underset{\phi \in \mathbb{R}/2\pi}{\operatorname{argmax}} \left| \operatorname{Re}\left\{ \sum_{k=1}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi} \right\} \right|$$

end for
Step 3: Final Phase Synchronization

Construct Hermitian $\widehat{H} \in \mathbb{C}^{n \times n}$ by $\widehat{H}_{ij} = e^{\iota \hat{\theta}_{ij}}$
Extract the leading eigenvector $\hat{u} = (\hat{u}_1, \ldots, \hat{u}_n)^\top$ of $\widehat{H}$
$\hat{x} \leftarrow (\hat{u}_1/|\hat{u}_1|, \ldots, \hat{u}_n/|\hat{u}_n|)$

---

will focus on analyzing PPE-SPC in depth in Section 4, but the analysis strategy can be seamlessly carried in principle to other variants of this algorithmic paradigm.

We briefly motivate the argmax operation in Step 2 as follows. If our measurement matrices are noise-free, then the (i, j)th entry of $W^{(k)}$ from Step 1 should equal to $e^{\iota k(\theta_i - \theta_j)}$; in this case, the goal of Step 2 is to reconstruct $(\theta_i - \theta_j)$ from its "trigonometric moments," for which any harmonic retrieval technique can be applied; the periodogram method in Algorithm 1 is among the most naïve approach for this purpose. For clean signal, the periodogram $|\operatorname{Re}\{\sum_{k=1}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k \phi}\}|$ is equal to the modulus of the Dirichlet kernel

$$\begin{aligned} \operatorname{Dir}_{k_{\max}}(\theta_{ij} - \theta) &:= \sum_{k=-k_{\max}}^{k_{\max}} e^{\iota k(\theta_{ij} - \theta)} \\ &= \frac{\sin[(k_{\max} + 1/2)(\theta_{ij} - \theta)]}{\sin[(\theta_{ij} - \theta)/2]} \end{aligned} \quad (11)$$

which attains its maximum at $\theta_{ij} = \theta_i - \theta_j \pmod{2\pi}$. Since the peak of $\operatorname{Dir}_{k_{\max}}$ becomes sharper and sharper as $k_{\max}$ increases, we expect the periodogram peak identification step to be robust to noise, which will produce a very high quality estimate $\widehat{H}$ for Step 3. In fact, our analysis in Section 4 suggests that this initialization stage alone can produce highly accurate phase vectors for sufficiently large $k_{\max}$, and the estimation error drops inverse-polynomially in $k_{\max}$.

3.2. Stage Two: Iterative Refinement

In this stage, we use an iterative refinement scheme that takes an initial phase vector and enhances it successively.

In our implementation we warm-start this iterative algorithm with the $\hat{x}$ produced from the PPE-SPC Algorithm 1, but any initialization scheme can be applied in principle, including random initialization. This iterative refinement concurrently performs the generalized power method (GPM) (Boumal, 2016) in multiple frequency channels consistently: at each frequency k, we perform power iteration by multiplication with $H^{(k)}$; the results are combined across frequency channels to obtain one periodogram for each vertex i followed by a "soft harmonic retrieval" step that soft-thresholds (Donoho, 1995) the periodogram in frequency domain. We pick a relatively lower threshold at the beginning of this iterative scheme, but gradually raise the threshold over 0.99 to reveal the true peak that persists. Details can be found in Algorithm 2, henceforth referred to as *multi-frequency generalized power method* (MFGPM).

Algorithm 2 Multi-Frequency Generalized Power Method (MFGPM)

Input: Hermitian matrices $\left\{ H^{(k)} \mid 1 \leq k \leq k_{\max} \right\}$, initialization $\hat{x} \in \mathbb{C}_1^n$; threshold τ
Output: Phase vector $\hat{z} \in \mathbb{C}_1^n$
for $k = 1$ **to** $k_{\max}$ **do**
 $z^{(k,0)} \leftarrow \hat{x}^k$
end for
for $t = 1$ **to** T **do**
 for $k = 1$ **to** $k_{\max}$ **do**
  $u^{(k,t)} \leftarrow H^{(k)} z^{(k,t-1)}$.
 end for
 for $i = 1$ **to** n **do**
  $$h_i^{(t)}(\theta) = \operatorname{Re}\left\{ \sum_{k=1}^{k_{\max}} u_i^{(k,t)} e^{\imath k\theta} \right\}$$
  Soft-thresholding: $\hat{u}_i^{(k,t)} \leftarrow \int_0^{2\pi} \eta_\tau(h_i^{(t)}(\theta)) e^{-\imath k\theta} \mathrm{d}\theta$
  Rounding: $z_i^{(k,t)} \leftarrow \hat{u}_i^{(k,t)} / |\hat{u}_i^{(k,t)}|$
 end for
end for
$\hat{z} \leftarrow z^{(1,T)}$

MFGPM can be viewed as an iterative version of PPE-SPC, except that the stringent peak extraction step is replaced with the more malleable soft-thresholding. Periodograms $h_i^{(t)}$ are virtually the Dirichlet kernels, which truncate a Dirac delta function in the frequency domain; one can also take Cesáro means of these periodograms, or equivalently, work with the *Féjer kernels* that are known to converge faster to the Dirac delta function. We omit those results as no significant difference is observed in performance.

As an integral part of our two-stage algorithmic framework, MFGPM works most efficiently with initialization from PPE-SPC, but we also observed empirically that the MFGPM outperforms other methods given identical random initialization, illustrated in Figure 1, in the sense that
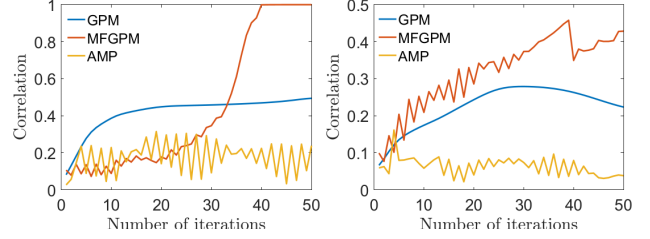

Figure 1. Evolution of correlations between iterates and the true phase vector $z \in \mathbb{C}_1^n$ for three iterative algorithms with the same random initialization. The three iterative algorithms in comparison are generalized power method (GPM, (Boumal, 2016)), multi-frequency generalized power method (MFGPM, Algorithm 2), and approximate message passing (AMP, (Perry et al., 2018)). The figure illustrates a typical success run (**left**) where MFGPM started with a random initialization of low correlation (around 0.1) but terminated with an output of correlation > 0.5, along with a failure run (**right**) started with another random initialization of comparable correlation but terminated with correlation < 0.5. In both circumstances, outputs from MFGPM attain higher correlation than GPM or AMP. Input data are generated from the random corruption model (5) with $r = 0.1$ and $n = 100$. See Figure 3 for more systematic comparison results under this noise model.

MFGPM often produces phase vectors that correlate more strongly with the true phase vector z. See Section 6 for more comprehensive comparisons results.

The computational complexities of PPE-SPC and MFGPM are $\mathcal{O}\left(k_{\max} n^3\right)$ and $\mathcal{O}\left(T k_{\max} n^2\right)$, respectively.

4. Analysis

In this section we analyze PPE-SPC in theory, under the general sub-Gaussian noise model (6). We assume the observation graph G is generated from a Erdős–Rényi model with edge connectivity $p \in [0, 1]$ independent of the $\Delta^{(k)}$'s.

Assumption 1 *For $\sigma > 0$ and each $k \in [k_{\max}]$, assume*

$$H^{(k)} = A \circ [z^k \left(z^k\right)^* + \sigma \Delta^{(k)}] \tag{12}$$

where $z^k \in \mathbb{C}_1^n$ is the entrywise kth power of z, and $\Delta^{(k)}$, $k = 1, \ldots, k_{\max}$ are complex random Wigner matrices satisfying the following assumptions:

(1) For any fixed $k \in [k_{\max}]$, $\{\operatorname{Re}(\Delta_{\ell j}^{(k)}), \operatorname{Im}(\Delta_{\ell j}^{(k)}) \mid 1 \leq \ell < j \leq n\}$ are jointly independent with zero mean, and unit sub-Gaussian norm (Vershynin, 2018);

(2) $\Delta_{ii}^{(k)} = 0$ for all $1 \leq k \leq k_{\max}$ and $1 \leq i \leq n$;

(3) $\Delta_{\ell j}^{(k)} = \overline{\Delta_{j\ell}^{(k)}}$ for all $k = 1, \ldots, k_{\max}$ and $\ell < j$.

Furthermore, assume A is the adjacency matrix of a Erdős–Rényi random graph independent of all the $\Delta^{(k)}$'s, with edge connecting probability $p \in [0, 1]$.

We emphasize again that Assumption 1 assumes no independence for the $\Delta^{(k)}$'s across frequency channels; only entries within the same $\Delta^{(k)}$ are assumed independent. As explained in Introduction, this enables us to unify our discussions on the random corruption model and additive Gaussian model in a single pass (see e.g., (8)). Another advantage for such generality is that we can focus on analyzing complete observation graphs, since

$$\mathbb{E}H^{(k)} = pz^k \left(z^k\right)^* - pI_n$$

where I_n is the identify matrix of dimension n-by-n, and thus we can apply the theoretical analysis in this section to $\frac{1}{p}\left(H^{(k)} + pI_n\right) = z^k \left(z^k\right)^* + E^{(k)}$ where

$$E^{(k)} = \frac{1}{p}\left\{A \circ [z^k \left(z^k\right)^* + \sigma\Delta^{(k)}]\right\} - z^k \left(z^k\right)^* + I_n$$

satisfies the same conditions as $\Delta^{(k)}$ in Assumption 1 with different absolute constants. Therefore, in the rest of this section we focus on complete observation graph G only, i.e.,

$$H^{(k)} = z^k \left(z^k\right)^* + \sigma\Delta^{(k)}, \quad 1 \le k \le k_{\max}. \quad (13)$$

Our first goal is to understand the spectral method in PPE-SPC Step 1 and Step 3. Since Step 2 is entrywise, it is crucial to bound the ℓ_∞ distance between z^k and the leading eigenvector $u^{(k)}$ (scaled to $\|u^{(k)}\|_2 = \sqrt{n}$). The proof of the following Lemma 1 uses recent ℓ_∞ perturbation results of eigenvectors of random matrices (Eldridge et al., 2017; Abbe et al., 2017; Fan et al., 2018; Zhong & Boumal, 2018) and can be found in the supplemental material.

Lemma 1 *Assume Assumption 1 is satisfied, and the observation graph G is a complete graph. Let $\epsilon \in (0, 2]$ be an arbitrarily chosen but fixed absolute constant. For any $k \in [k_{\max}]$, denote $u^{(k)}$ for the leading eigenvector of $H^{(k)}$ scaled such that $\left\|u^{(k)}\right\|_2 = \sqrt{n}$ and $(z^k)^* u^{(k)} = \left|(z^k)^* u^{(k)}\right|$. There exist absolute (in particular, independent of k and n) constants $c_0, C_0, C_2 > 0$ such that, if $\sigma < c_0 \sqrt{n/\log n}$, there holds with probability $1 - \mathcal{O}\left(n^{-(2+\epsilon)}\right)$*

$$\|u^{(k)} - z^k\|_\infty \le C_0 \sigma \sqrt{\log n/n}, \quad (14)$$

$$\left|W_{ij}^{(k)} - z_i^k \bar{z}_j^k\right| \le C_2 \sigma \sqrt{\log n/n}. \quad (15)$$

The inequality (15) is a direct consequence of (14), which is identical to Theorem 8 of (Zhong & Boumal, 2018), but we verify in the proof that the event probability $1 - \mathcal{O}(n^{-2})$ in (Zhong & Boumal, 2018) can be made slightly higher. This is necessary for taking the union bound across all $\mathcal{O}(n^2)$ entries in the main Theorem 2.

A quick consequence of Lemma 1 is the uniform proximity of the periodogram to a Dirichlet kernel up to constant scaling and shifts, with high probability. More specifically,

$$\left|\text{Re}\left\{\sum_{k=1}^{k_{\max}} W_{ij}^{(k)} e^{-\iota k\phi}\right\} - \frac{1}{2}\left[\text{Dir}_{k_{\max}}\left(\theta_i - \theta_j - \phi\right) - 1\right]\right|$$

$$\le \left|\sum_{k=1}^{k_{\max}}\left(W_{ij}^{(k)} - z_i^k \bar{z}_j^k\right) e^{-\iota k\phi}\right| \le 2C_2 k_{\max} \sigma \sqrt{\log n/n}$$

with probability $1 - \mathcal{O}\left(n^{-(2+\epsilon)}\right)$. Clearly, the maximum of $|\text{Dir}_{k_{\max}}\left(\theta_i - \theta_j - \phi\right) - 1|$ is attained at $\theta = \theta_i - \theta_j$. We thus expect the argmax operation in Step 2 of PPE-SPC to produce high accuracy estimates of $\theta_i - \theta_j$ as long as the difference between the "optimization landscape" of the periodogram and the Dirichlet kernel is small enough. This is formalized in the following lemma, which exploits the geometry of the Dirichlet kernel.

Lemma 2 *Under the same conditions as in Lemma 1, if*

$$[2k_{\max} \sin\left(\pi/\left(2k_{\max} + 1\right)\right)]^{-1} + 4C_2 \sigma \sqrt{\log n/n} < 1 \quad (16)$$

then with probability at least $1 - \mathcal{O}\left(n^{-(2+\epsilon)}\right)$

$$\left|\hat{\theta}_{ij} - (\theta_i - \theta_j)\right| \le 4\pi/\left(2k_{\max} + 1\right). \quad (17)$$

It is straightforward to check that (16) holds for sufficiently large $k_{\max}$ as long as $4C_2 \sigma \sqrt{\log n/n}$ is bounded from above by $1 - 1/\pi$. This can be seen by noticing that the function $[2x \sin\left(\pi/\left(2x + 1\right)\right)]^{-1}$ is differentiable and monotonically decreasing for all $x \ge 2$, and for sufficiently large $k_{\max}$ it infinitesimally approaches $1/\pi < 1$.

The most important message from Lemma 2 is the following: At the beginning of the Step 3 of PPE-SPC, the newly constructed Hermitian matrix $\widehat{H}$ is entrywise $\mathcal{O}\left(k_{\max}^{-1}\right)$–close to the ground truth rank-one matrix zz^*. We emphasize that this error incurred in $\widehat{H}$ is significantly smaller than the noise level σ in the raw input data, and can be made arbitrarily small by choosing large $k_{\max}$. We formalize this key observation in the main theorem below, for which the proof is deferred to the supplemental material.

Theorem 2 *Under the same conditions as Lemma 1 and Lemma 2, if (16) holds and $4c_0 C_2 < 1 - \sqrt{2}/\pi$, then there exists an absolute constant $C_3 > 0$ such that, with probability $1 - \mathcal{O}(n^{-\epsilon})$, the correlation between the true phase vector z and the leading eigenvector $\hat{u}$ (scaled to $\|\hat{u}\|_2 = \sqrt{n}$) of $\widehat{H}$ in PPE-SPC Step 3 is at least*

$$\text{Corr}\left(\hat{u}, z\right) \ge 1 - C_3/k_{\max}^2 \quad (18)$$

provided that

$$k_{\max} > \max\left\{5, \left(\sqrt{2}\pi\left(1 - 4C_2 \sigma \sqrt{\log n/n}\right) - 2\right)^{-1}\right\}.$$

Moreover, for the phase vector $\hat{x}$ output from PPE-SPC,

$$\text{Corr}\left(\hat{x}, z\right) \ge 1 - 4C_3/k_{\max}^2.$$

Following the discussion after Lemma 2, it is not surprising to see in Theorem 2 that the correlation can be

made arbitrarily close to 1 (or equivalently, the ℓ_2 distance between the estimated and true phase vectors can be made arbitrarily close to 0). Moreover, it doesn't take excessively large $k_{\max}$ for PPE-SPC to outperform all existing phase synchronization algorithms—in fact, for $\sigma \asymp \mathcal{O}(\sqrt{n/\log n})$ which is the highest level of noise tolerable to ensure the validity of Lemma 1, it suffices to take $k_{\max} = \mathcal{O}(\sqrt{n}/\sigma) \asymp \mathcal{O}(\sqrt{\log n})$ to suppress the ℓ_2 estimation error below the established near-optimal bound $\mathcal{O}(\sigma)$ for eigenvector based phase synchronization methods (Bandeira et al., 2017; Zhong & Boumal, 2018). We believe (18) can still be improved by a factor of $\sqrt{n}$ by leveraging the randomness in the residue error in (15), but such finer analysis relies on more detailed analysis on the ℓ_∞ perturbation and the change in the optimization landscape, which will be pursued in a future work.

5. Extension to General Synchronization

The algorithmic framework of multi-frequency phase synchronization proposed in this paper can be extended to synchronization over any compact Lie group $\mathcal{G}$, by the representation-theoretic analogue of Fourier series — the *Peter–Weyl decomposition*. In a nutshell, the *Peter–Weyl theorem* states that, for square integrable functions $f \in L^2(\mathcal{G})$, we have decomposition

$$f(g) = \sum_{k=0}^{\infty} d_k \mathrm{tr}\left(\hat{f}(k)\rho_k(g)\right) \tag{19}$$

where each $\rho_k : \mathcal{G} \to \mathbb{C}^{d_k \times d_k}$ is an irreducible, unitary representation of $\mathcal{G}$, and $\hat{f}(k)$ is the "Fourier coefficient"

$$\hat{f}(k) = \int_{\mathcal{G}} f(g)\rho_k(g)\,\mathrm{d}g, \tag{20}$$

where the integral is take with respect to the Haar measure.

On a connected observation graph G, the input data to a *synchronization problem over group* $\mathcal{G}$ are pairwise measurements $g_{ij} \in \mathcal{G}$ on edges $(i,j) \in E$ satisfying $g_{ij} = g_{ji}^{-1}$. The goal is to find n group elements $g_1, \ldots, g_n \in \mathcal{G}$, one for each vertex, that satisfy as many constraints $g_{ij} = g_i g_j^{-1}$ as possible. Mathematically, this type of problems can often be formulated as an optimization problem (Bandeira et al., 2015)

$$\min_{g_1,\ldots,g_n \in \mathcal{G}} \sum_{i,j=1}^{n} f_{ij}(g_i g_j^{-1}), \tag{21}$$

where each $f_{ij} \in L^2(\mathcal{G})$ measures the compatibility between the relative alignment $g_i g_j^{-1}$ and the observation data g_{ij} on edge $(i,j) \in E$. The f_{ij}'s are nonlinear and nonconvex in general. If f_{ij} are bandlimited, we can expand (21) using the Peter–Weyl decomposition

$$\sum_{i,j=1}^{n} f_{ij}(g_i g_j^{-1}) = \sum_{k=0}^{k_{\max}} \sum_{i,j=1}^{n} d_k \mathrm{tr}\left[\hat{f}_{ij}(k)\rho_k(g_i)\rho_k^*(g_j)\right]$$

which can be viewed as a generalization of the multi-frequency phase synchronization problem (4).

For simplicity of statement, we assume the observation graph G is complete in this section. Since ρ_k's are unitary representations, the matrices $\rho_k(g)$'s are unitary matrices for any $g \in \mathcal{G}$, and it is natural to solve for g_i from its irreducible representations $\rho_k(g_i)$. Vertically stacking the kth irreducible representations together, the variable can be organized in matrices $X^{(k)} \in \mathbb{C}^{nd_k \times d_k}$, $k \in \mathbb{Z}$ defined by

$$X^{(k)} = [\rho_k(g_1), \ldots, \rho_k(g_n)]^\top. \tag{22}$$

Analogies of the noise models also exist in this more general setting. The additive Gaussian noise model, following (Perry et al., 2018), amounts to

$$H^{(k)} = \frac{\lambda_k}{n} X^{(k)}(X^{(k)})^* + \frac{1}{\sqrt{nd_k}}\Delta^{(k)} \tag{23}$$

where the parameter $\lambda_k > 0$ stands for the signal-to-noise ratio (SNR) at "frequency k," $\Delta^{(k)} \in \mathbb{C}^{nd_k \times nd_k}$ is a Wigner matrix with i.i.d. standard complex Gaussian entries in the upper triangular part. For the random corruption model, let

$$g_{ij} = \begin{cases} g_i g_j^{-1}, & \text{with probability } r \\ \tilde{g} \sim \mathrm{Unif}(\mathcal{G}), & \text{with probability } 1 - r \end{cases} \tag{24}$$

and set the (i,j)th sub-block of $H^{(k)}$ to $\rho_k(g_{ij})$.

As we elaborate in the remainder of this section, all the key ingredients in PPE-SPC and MFGPM can be extended to this more general setting. We demonstrate the efficacy of this algorithm for SO(3) synchronization in Section 6.

Spectral relaxation: Compute the top d_k eigenvectors and stack them horizontally to form $U^{(k)} = [u_1^{(k)}, \ldots, u_{d_k}^{(k)}]$. Approximate $H^{(k)}$ with $\widehat{H}^{(k)} = U^{(k)}(U^{(k)})^*$.

Generalized harmonic retrieval: For each $(i,j) \in E$, set

$$\hat{g}_{ij} = \operatorname*{argmax}_{g \in \mathcal{G}} \sum_{k=1}^{k_{\max}} d_k \mathrm{tr}\left[\widehat{H}_{ij}^{(k)}\rho_k^*(g)\right]. \tag{25}$$

Based on these new estimates for the pairwise alignments, we build matrix $\widehat{H}$ with n^2 blocks with $\widehat{H}_{ij} = \rho_1(\hat{g}_{ij})$. We then extract the top d_1 eigenvectors of $\widehat{H}$, stack them horizontally to form $\widetilde{U} = [u_1, u_2, \ldots u_{d_1}]$, and project each of its n vertical blocks $\widetilde{U}_1 \ldots, \widetilde{U}_n \in \mathbb{C}^{d_1 \times d_1}$ to a unitary matrix through singular value decomposition (SVD)

$$\mathrm{Proj}(\widetilde{U}_i) = \Phi\Psi^* \quad \text{where} \quad \widetilde{U}_i = \Phi\Sigma\Psi^*. \tag{26}$$

Iterative refinement: At the tth iteration, denoting $X^{(k,t)}$ for the current stacked kth representations (22), we construct

$$Y^{(k)} = H^{(k)}X^{(k,t)},$$

(a) AMP   (b) PPE-SPC   (c) MFGPM

(d) PPE-SDP   (e) PPE-SDP   (f) PPE-SPC^3
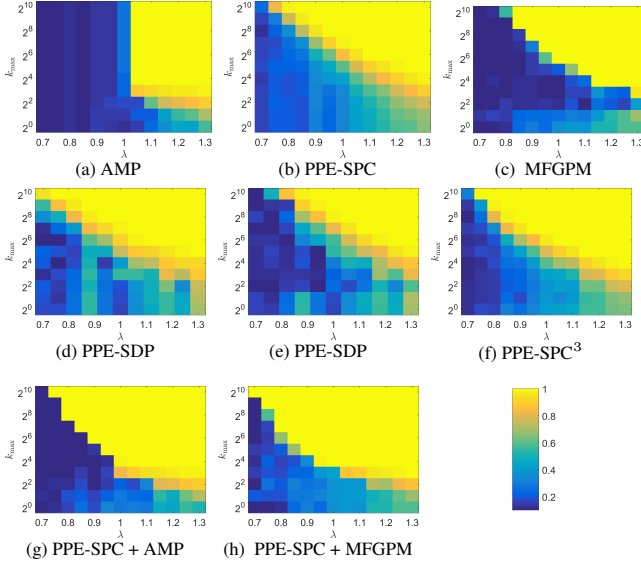
(g) PPE-SPC + AMP   (h) PPE-SPC + MFGPM

Figure 2. $U(1)$ synchronization under Gaussian noise model for $n = 100$. Here $\sigma = \sqrt{n}/\lambda$. Every data point is a median over 40 trials in (a)–(c) and (f)–(h), and over 5 trials in (d)–(e).

Figure 3. $U(1)$ synchronization under random corruption model for $n = 100$. Here $r = \lambda/\sqrt{n}$. Every data point is a median over 40 trials in (a)–(c) and (f)–(h), and over 5 trials in (d)–(e).

and compute the inverse Fourier transform for each of the n vertical sub-blocks $Y_1^{(k)}, \ldots, Y_n^{(k)} \in \mathbb{C}^{d_k \times d_k}$ of $Y^{(k)}$:

$$h_i(g) = \sum_{k=1}^{k_{\max}} \mathrm{Re}\left\{ d_k \mathrm{tr}\left[Y_i^{(k)} \rho_k^*(g) \right] \right\}, \quad i = 1, \ldots, n.$$

Note that we only need toe evaluate $C_i(g)$ on a finite number of uniformly sampled elements of $\mathcal{G}$, from which the "inverse Fourier transform" can be applied

$$U_i^{(k)} = \int_{\mathcal{G}} \eta_\tau \left(h_i(g) \right) \rho_k(g) \, dg \tag{27}$$

along with the soft-thresholding η_τ. We again project each $U_i^{(k)}$ to the closest unitary matrix by SVD (26), then form $X^{(k,t+1)}$ by vertically stacking the $\mathrm{Proj}(U_i^{(k)})$'s. The final outputs are $\widehat{X}^{(k)} = X^{(k,T)}$ for $k = 1, \ldots, k_{\max}$.

6. Numerical Experiments

This section contains detailed numerical results under both additive Gaussian noise and random corruption models, for both $U(1)$ and $SO(3)$. In all experiments with Gaussian noise, we keep $\sigma_k \equiv \sigma \equiv \sqrt{n}/\lambda$ where $\lambda > 0$ is the signal-to-noise ratio (SNR); for the random corruption model (3) we set $r \equiv \lambda/\sqrt{n}$. We fix $n = 100$ and vary λ and $k_{\max}$ to evaluate and compare the performance of different algorithms. When comparing iterative algorithms (AMP, GPM, MFGPM), within each random trial the random initialization is kept identical for all three algorithms and across frequency channels; between trials both data and initialization are redrawn. The remainder of the section contains
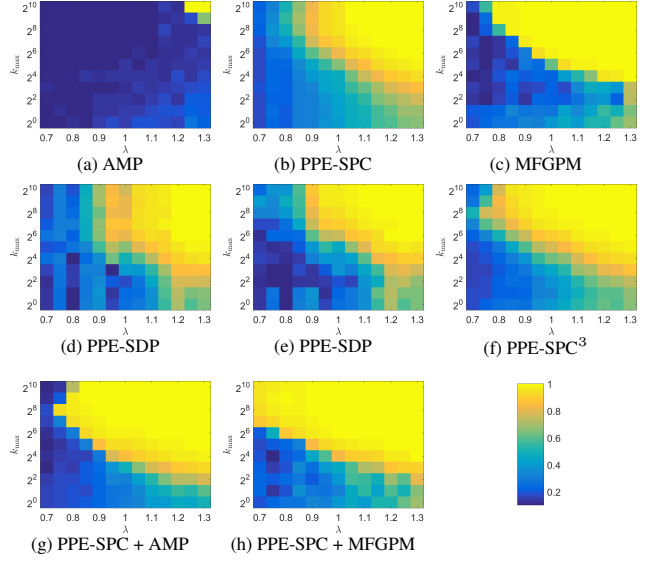
results for $U(1)$ and $SO(3)$ synchronization with complete observation graphs only; incomplete observation graph results are similar and included in the supplemental material.

$U(1)$ **synchronization:** In Figure 2 and Figure 3, we measure the correlation between the output and the truth phase vector for various single- and multi-frequency synchronization methods, under the additive Gaussian and random corruption noise model, respectively. The SNR λ varies between 0.7 and 1.3, which is in the extremely noisy regime: under the random corruption model, for instance, with $n = 100$, between 87% and 93% of the pairwise alignments are corrupted with random elements. In each subplot, the vertical axis varies $k_{\max}$ from 1 to 1024, and the horizontal axis marks the change in λ. The bottom row in each subplot thus represents the single-frequency ($k_{\max} = 1$) version of the algorithm. The methods under comparison are: (a) AMP (Perry et al., 2018) with random initialization; (b) PPE-SPC; (c) MFGPM with random initialization; (d) PPE-SDP (replacing the spectral methods in Algorithm 1 with SDP relaxation); (e) PPE-SDP with an additional projection to rank-one matrices in each iteration; (f) Iterating PPE-SPC three times; (g) AMP initialized with PPE-SPC; (h) MFGPM initialized with PPE-SPC.

It is clear from Figure 2 and Figure 3 that leveraging information in multiple frequency channels produces superior results than single-frequency approaches. Most shockingly, in Figure 3 our proposed PPE-SPC method and variants [subplots (b)–(h)] are capable of recovering the true phase vector when the SNR is well below the critical threshold $\lambda = 1$ (corresponding to $r < 1/\sqrt{n}$) determined in (Singer,
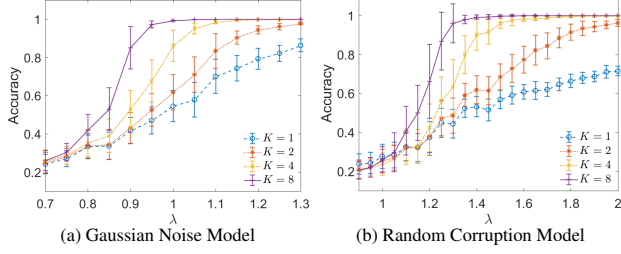
(a) Gaussian Noise Model   (b) Random Corruption Model

Figure 4. Accuracy and error-bars for $SO(3)$ synchronization with Gaussian noise (**left**) and random corruption (**right**) model with K frequencies, for various K. The noise levels are kept as $\lambda = \sqrt{n}/\sigma$ for Gaussian models and $r = \lambda/\sqrt{n}$ for random corruption models, where $n = 100$. Accuracy is measured by the correlation $\|(X^{(1)})^* \widehat{X}^{(1)}\|_F/(\sqrt{3}n)$ between estimates and the ground truth. Each data point is the median of 50 trials.

2011) by random matrix arguments. This is surprising because, according to (Singer, 2011), for single frequency phase synchronization one can not expect correlation to be much higher than $1/\sqrt{n}$, which is 0.1 in our experiments. This is confirmed by looking at the bottom row of each subplot of Figure 3, but with suitably large $k_{\max}$ this barrier no longer exists, even though in model (5) our high-frequency measurements are generated from the single frequency data.

In Figure 2 and Figure 3, (d) and (e) illustrates the performance of the SDP variant of PPE-SPC. The difference between (d) and (e) is the following: in (d) we use directly estimated $W^{(k)}$ by solving the SDP in (Singer, 2011), but in (e) we apply project the SDP solution to a rank-one matrix using eigen-decomposition. The results from these SDP variants are occasionally slightly better PPE-SPC, but the computational cost is expensive: the runtime is over 40 times longer, and a lot more memory is required. The SDP relaxation in (Bandeira et al., 2015) is even more demanding on computation resources so is not included here.

Figures 2f and 3f explore another possibility of extending PPE-SPC: After recovering $\widehat{H}$, take entrywise powers of $\widehat{H}$ and treat them as multi-frequency data input to another fresh run of PPE-SPC. Unlike the iterative refinement algorithm MFGPM, we observed empirically that the performance boost saturate quickly after just a couple of such repeated calls to PPE-SPC. The result in (f) from both figures are obtained from performing 3 such repetitions. Compared with (b), this strategy improves the estimation accuracy for smaller λ, but the performance gain is not as significant as using MFGPM for iterative refinements (h).

Initialization turns out to be important for AMP: As shown in Figure 2a, when the SNR is below the critical threshold predicted in (Perry et al., 2018) ($\lambda < 1$), increasing $k_{\max}$ does not lead to performance improvement; the critical threshold appears even higher for random corruption model (Figure 3a). In contrast, PPE-SPC and MFGPM can
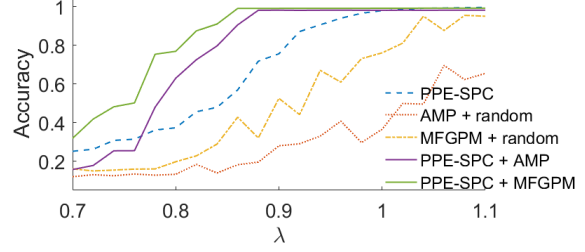

Figure 5. Comparison of random and spectral initialization for $SO(3)$ synchronization, under noise model (23) where $k_{\max} = 8, n = 100$. Each data point is averaged from 20 random trials.

always benefit from sufficiently larger $k_{\max}$.

$SO(3)$ **synchronization:** Comparison results for $SO(3)$ synchronization under Gaussian noise model and random corruption model are shown in Figure 4a and 4b, respectively. In all these experiments, the Fourier transform (27) is numerically evaluated using $m = 1000$ elements uniformly sampled in $SO(3)$. Clearly, the proposed method outperforms single frequency methods and achieve higher accuracy as $k_{\max}$ increases; moreover, the multi-frequency formulation and algorithm lead to drastic performance boost especially at the "low SNR regime."

In Figure 5 we compare AMP and MFGPM with different initialization strategies–PPE-SPC vs. random initialization–under the additive Gaussian noise model (23) with $k_{\max} = 8$. We plot the accuracy of using PPE-SPC alone without iterative refinement as a baseline. The results demonstrate the performance boost from using PPE-SPC for initialization, as well as improvements gain from using iterative refinements on top of the initialization PPE-SPC.

7. Conclusion

In this paper, we propose a novel, mult-frequency formulation for phase synchronization as a nonconvex optimization problem, for which we develop a two-stage algorithm inspired by harmonic retrieval and generalized power method that produces high accuracy approximate solutions. We demonstrate in theory and experiments that the new framework significantly outperform all existing phase synchronization algorithms.

There are many opportunities for future research. We are particularly interested in gaining deeper theoretical understandings for the multi-frequency GPM algorithm, especially its performance guarantees and behavior near local optimum. More general harmonic retrieval techniques can be potentially used in place of the periodogram-based peak extraction. We are also working on extending the algorithmic framework beyond compact Lie groups, such as Euclidean groups and symmetric groups, with applications to object matching (Shen et al., 2016; Pachauri et al., 2013).

Acknowledgements

References

Abbe, E., Fan, J., Wang, K., and Zhong, Y. Entrywise eigenvector analysis of random matrices with low expected rank. *arXiv preprint arXiv:1709.09565*, 2017.

Bandeira, A., Chen, Y., and Singer, A. Non-unique games over compact groups and orientation estimation in cryo-EM. *arXiv preprint arXiv:1505.03840*, 2015.

Bandeira, A. S., Kennedy, C., and Singer, A. Approximating the little Grothendieck problem over the orthogonal and unitary groups. *Mathematical Programming*, 160(1-2): 433–475, 2016.

Bandeira, A. S., Boumal, N., and Singer, A. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, 163(1-2): 145–167, 2017.

Boumal, N. Nonconvex Phase Synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.

Bresler, Y. and Macovski, A. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1081–1089, 1986.

Candes, E. J., Li, X., and Soltanolkotabi, M. Phase Retrieval via Wirtinger Flow: Theory and Algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

Chaudhury, K. N., Khoo, Y., and Singer, A. Global Registration of Multiple Point Clouds Using Semidefinite Programming. *SIAM Journal on Optimization*, 25(1): 468–501, 2015.

Chen, Y. and Candes, E. Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 28*, pp. 739–747. 2015.

Chen, Y., Suh, C., and Goldsmith, A. J. Information Recovery From Pairwise Measurements. *IEEE Transactions on Information Theory*, 62(10):5881–5905, 2016.

Cucuringu, M., Singer, A., and Cowburn, D. Eigenvector synchronization, graph rigidity and the molecule problem. *Information and inference : a journal of the IMA*, 1(1): 21, 2012.

Donoho, D. L. De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3):613–627, 1995.

Eldridge, J., Belkin, M., and Wang, Y. Unperturbed: spectral analysis beyond Davis-Kahan. *arxiv preprint. arxiv: 1706.06516*, 2017.

Fan, J., Wang, W., and Zhong, Y. An ℓ_∞ Eigenvector Perturbation Bound and Its Application. *Journal of Machine Learning Research*, 18(207):1–42, 2018.

Gao, T., Brodzki, J., and Mukherjee, S. The Geometry of Synchronization Problems and Learning Group Actions. *Discrete & Computational Geometry*, to appear 2019.

Liu, H., Yue, M.-C., and Man-Cho So, A. On the Estimation Performance and Convergence Rate of the Generalized Power Method for Phase Synchronization. *SIAM Journal on Optimization*, 27(4):2426–2446, 2017.

Pachauri, D., Kondor, R., and Singh, V. Solving the multiway matching problem by permutation synchronization. In *Advances in neural information processing systems*, pp. 1860–1868, 2013.

Perry, A., Wein, A. S., Bandeira, A. S., and Moitra, A. Message-Passing Algorithms for Synchronization Problems over Compact Groups. *Communications on Pure and Applied Mathematics*, 2018.

Roy, R. and Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7): 984–995, 1989.

Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.

Shen, Y., Huang, Q., Srebro, N., and Sanghavi, S. Normalized spectral map synchronization. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 29*, pp. 4925–4933. 2016.

Singer, A. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1):20–36, 2011.

Singer, A., Zhao, Z., Shkolnisky, Y., and Hadani, R. Viewing angle classification of cryo-electron microscopy images using eigenvectors. *SIAM Journal on Imaging Sciences*, 4(2):723–759, 2011.

Sorensen, M. and De Lathauwer, L. Multidimensional Harmonic Retrieval via Coupled Canonical Polyadic DecompositionPart I: Model and Identifiability. *IEEE Transactions on Signal Processing*, 65(2):517–527, 2017a.

Sorensen, M. and De Lathauwer, L. Multidimensional Harmonic Retrieval via Coupled Canonical Polyadic DecompositionPart II: Algorithm and Multirate Sampling. *IEEE Transactions on Signal Processing*, 65(2):528–539, 2017b.

Stoica, P. and Moses, R. L. *Introduction to spectral analysis*. Prentice Hall, 1997.

Tufts, D. W. and Kumaresan, R. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. *Proceedings of the IEEE*, 70 (9):975–989, 1982.

Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47. Cambridge University Press, 2018.

Zhang, S. and Huang, Y. Complex Quadratic Optimization and Semidefinite Programming. *SIAM Journal on Optimization*, 16(3):871–890, 2006.

Zhong, Y. and Boumal, N. Near-Optimal Bounds for Phase Synchronization. *SIAM Journal on Optimization*, 28(2): 989–1016, 2018.

Ziskind, I. and Wax, M. Maximum likelihood localization of multiple sources by alternating projection. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36 (10):1553–1560, 1988.