(Near-)optimal Results for Phase Synchronization

Yiqiao Zhong

Princeton University

with Nicolas Boumal (PACM)

SIAM AN18, Portland, July 10, 2018
Outline

1. Background
2. Main Results
3. Proof Ideas
4. Concluding Remarks
Phase (angular) synchronization

- Unknown parameters (angles): $\theta_1, \theta_2, \ldots, \theta_n \in [0, 2\pi)$. 

Goal: estimate these parameters from pairwise measurements (offsets):

$$y_{\ell k} = \text{noisy version of } \theta_\ell - \theta_k \mod 2\pi,$$

where $1 \leq \ell < k \leq n$. 


Phase (angular) synchronization

- Unknown parameters (angles): $\theta_1, \theta_2, \ldots, \theta_n \in [0, 2\pi)$.

- **Goal**: estimate these parameters from pairwise measurements (offsets):

  $$y_{\ell k} = \text{noisy version of } \theta_\ell - \theta_k \mod 2\pi,$$

where $1 \leq \ell < k \leq n$. 
Motivation

- Time synchronization.
Motivation

- Time synchronization.

- More generally, a group instead of $[0, 2\pi)$. Applications: Cryo-EM (Electron cryomicroscopy), calibration of cameras, robotics.
Motivation

- Re-formulate our problem:

\[ C_{\ell k} = \text{noisy version of } \bar{z}_\ell z_k, \]

where \( z_k = \exp(i\theta_k) \).
Motivation

- Re-formulate our problem:

\[ C_{\ell k} = \text{noisy version of } \bar{z}_\ell z_k, \]

where \( z_k = \exp(i\theta_k). \)

- The **model**:

\[ C_{\ell k} = \bar{z}_\ell z_k + \sigma W_{\ell k}, \quad \forall \ell > k \]

where \( W_{\ell k} \sim \mathcal{N}(0, 1). \) Assume all pairs of measurements.
Re-formulate our problem:

\[ C_{\ell k} = \text{noisy version of } \bar{z}_\ell z_k, \]

where \( z_k = \exp(i\theta_k) \).

The model:

\[ C_{\ell k} = \bar{z}_\ell z_k + \sigma W_{\ell k}, \quad \forall \ell > k \]

where \( W_{\ell k} \sim N_{\mathbb{C}}(0, 1) \). Assume all pairs of measurements.

The matrix form:

\[ C = zz^* + \sigma W, \]

where \( z \in \mathbb{C}^n \) with \( |z_k| = 1 \); \( W_{kk} = 0 \), \( W_{k\ell} = \bar{W}_{\ell k} \).
Motivation

- Deriving the MLE: minimize $\| C - xx^* \|^2_F$ over $x \in \mathbb{C}^n$ with $|x_k| = 1$. 

Information limit: $\sigma = \sqrt{n}$.

Our goal: under $\sigma = \tilde{O}(\sqrt{n})$, develop efficient algorithms that find $\hat{x}$; derive statistical guarantees.
Motivation

- Deriving the MLE: minimize $\| C - xx^* \|_F^2$ over $x \in \mathbb{C}^n$ with $|x_k| = 1$.

- Equivalently,

$$\max_{x \in \mathbb{C}^n} x^* C x \text{ subject to } |x_k| = 1 \quad \forall k \in [n]. \quad (P)$$
Motivation

- Deriving the MLE: minimize $\| C - xx^* \|_F^2$ over $x \in \mathbb{C}^n$ with $|x_k| = 1$.

- Equivalently,

$$\max_{x \in \mathbb{C}^n} x^* C x \text{ subject to } |x_k| = 1 \quad \forall k \in [n]. \quad \text{(P)}$$

- Denote the solution by $\hat{x}$. Up to a global phase.
Motivation

- Deriving the MLE: minimize $\| C - xx^* \|_F^2$ over $x \in \mathbb{C}^n$ with $|x_k| = 1$.

- Equivalently,

$$\max_{x \in \mathbb{C}^n} x^* C x \text{ subject to } |x_k| = 1 \quad \forall k \in [n].$$  \hspace{1cm} (P)

- Denote the solution by $\hat{x}$. Up to a global phase.

- Information limit: $\sigma = \sqrt{n}$. 

Deriving the MLE: minimize $||C - xx^*||_F^2$ over $x \in \mathbb{C}^n$ with $|x_k| = 1$.

Equivalently,

$$\max_{x \in \mathbb{C}^n} x^* C x \text{ subject to } |x_k| = 1 \quad \forall k \in [n]. \quad (P)$$

Denote the solution by $\hat{x}$. Up to a global phase.

Information limit: $\sigma = \sqrt{n}$.

Our goal: under $\sigma = \tilde{O}(\sqrt{n})$,

- Develop efficient algorithms that find $\hat{x}$;
- Derive statistical guarantees.
Recall the MLE $\hat{x}$ is a solution to:

$$\max_{x \in \mathbb{C}^n} x^* Cx \text{ subject to } |x_k| = 1 \quad \forall k \in [n]. \quad (P)$$
Recall the MLE $\hat{x}$ is a solution to:

$$\max_{x \in \mathbb{C}^n} x^* C x \quad \text{subject to} \quad |x_k| = 1 \quad \forall k \in [n]. \quad (P)$$

Trouble...nonconvexity!
Recall the MLE $\hat{x}$ is a solution to:

$$\max_{x \in \mathbb{C}^n} x^* C x \text{ subject to } |x_k| = 1 \quad \forall k \in [n]. \tag{P}$$

Trouble...nonconvexity!

Indeed, NP-hard \textit{in general}. Zhang and Huang [2006]
Standard recipe: semidefinite relaxation

- However...may be tractable under our model.

\[ X = x x^* \succeq 0 \]

Quadratic $\Rightarrow$ Linear:

\[ x^* C x \Rightarrow \text{Tr} (C X), \quad |x_k| = 1 \Rightarrow X_{kk} = 1 \]

Equivalently,

\[
\max_{X \in \mathbb{C}^{n \times n}, X = X^*, \text{Tr}(C X), \text{diag}(X) = 1, X \succeq 0, \text{rank}(X) = 1}
\]
Standard recipe: semidefinite relaxation

- However...may be tractable under our model.
- Lifting the problem to higher dimensional space:

\[ X = xx^* \succeq 0 \]

- Quadratic ⇒ Linear:

\[ x^* Cx \Rightarrow \text{Tr}(CX), \quad |x_k| = 1 \Rightarrow X_{kk} = 1 \]
Standard recipe: semidefinite relaxation

- However...may be tractable under our model.
- Lifting the problem to higher dimensional space:

\[ X = xx^* \succeq 0 \]

- Quadratic ⇒ Linear:

\[ x^* Cx \Rightarrow \text{Tr}(CX), \quad |x_k| = 1 \Rightarrow X_{kk} = 1 \]

- Equivalently,

\[
\max_{X \in \mathbb{C}^{n \times n}, X = X^*} \text{Tr}(CX) \quad \text{subject to } \text{diag}(X) = 1, X \succeq 0, \quad \text{rank}(X) = 1.
\]
Standard recipe: semidefinite relaxation

- However...may be tractable under our model.
- Lifting the problem to higher dimensional space:
  \[ X = xx^* \]

- Quadratic ⇒ Linear:
  \[ x^* Cx ⇒ \text{Tr}(CX), \quad |x_k| = 1 ⇒ X_{kk} = 1 \]

- semidefinite relaxation:
  \[
  \max_{X ∈ \mathbb{C}^{n×n}, X = X^*} \text{Tr}(CX) \quad \text{subject to} \quad \text{diag}(X) = 1, X ≥ 0.
  \]
  \[ \text{rank}(X) = 1 \quad \text{(SDP)} \]
Verify with dual certificate: find $\lambda$ such that $q(\lambda) = f(X)$.
Standard recipe: semidefinite relaxation

- Verify with dual certificate: find $\lambda$ such that $q(\lambda) = f(X)$.

- Widely studied: compressed sensing, matrix completion, robust PCA, Stochastic block model, etc.
Phase synchronization: why difficult?

Dual certificate:

\[ S = \text{Re}(\text{diag}(C^\top x^\top x^*)) - C. \]

Goal: to show \( S \succeq 0. \)

Complicated statistical dependence!

Previous analyses are sub-optimal, e.g., \( \sigma = O(n^{1/4}) \) in Bandeira, Boumal, and Singer [2016]. Simulations suggest success for \( \sigma = \tilde{O}(\sqrt{n}) \).
Standard recipe: semidefinite relaxation

- Phase synchronization: why difficult?
- Dual certificate:

\[ S = \text{Re}(\text{ddiag}(C\hat{x}\hat{x}^*)) - C. \]

Goal: to show \( S \succeq 0. \)
Phase synchronization: why difficult?

Dual certificate:

\[ S = \text{Re}(\text{ddiag}(C\hat{x}\hat{x}^*)) - C. \]

Goal: to show \( S \succeq 0. \)

Complicated statistical dependence!

Previous analyses are sub-optimal, e.g., \( \sigma = O(n^{1/4}) \) in Bandeira et al. [2016]. Simulations suggest success for \( \sigma = \tilde{O}(\sqrt{n}). \)
One of our main results:

**Theorem**

If \( \sigma = O\left(\sqrt{\frac{n}{\log n}}\right) \), with high probability for large \( n \), SDP admits a unique solution \( \hat{x}\hat{x}^* \), where \( \hat{x} \) is a global optimum of (P) (unique up to phase.)

‘With high probability’ is \( 1 - O(n^{-2}) \).
Faster approach: Generalized Power Method

- Beyond SDP?
Faster approach: Generalized Power Method

- Beyond SDP?

- Observe

\[
\max_{x \in \mathbb{C}^n} x^* C x \quad \text{subject to } |x_k| = 1 \quad \forall k \in [n].
\]
**Faster approach: Generalized Power Method**

- Beyond SDP?

- Similar to the **eigenvector** problem!

\[
\max_{x \in \mathbb{C}^n} x^* Cx \text{ subject to } |x_k| = 1 \quad \forall k \in [n].
\]

\[
\|x\| = 1
\]
Faster approach: Generalized Power Method

- Beyond SDP?

- Similar to the eigenvector problem!

\[
\max_{x \in \mathbb{C}^n} \ x^* C x \quad \text{subject to} \quad |x_k| = 1 \quad \forall k \in [n].
\]

\[
\|x\| = 1
\]
Faster approach: Generalized Power Method

- Beyond SDP?

- Similar to the \textit{eigenvector} problem!

\[
\max_{x \in \mathbb{C}^n} \quad x^* C x \quad \text{subject to} \quad |x_k| = 1 \quad \forall k \in [n].
\]

\[
\|x\| = 1
\]

\[
(x^{t+1})_k = \frac{(C x^t)_k}{|(C x^t)_k|} \quad \forall k \in [n]
\]

\textbf{Generalized Power method}
Generalized Power Method:

1. Set \( x^0 \) to be a leading eigenvector of \( C \) with \( \| x^0 \|_2 = \sqrt{n} \).

2. For \( t = 0, 1, \ldots \), update \( (x^{t+1})_k = \frac{(Cx^t)_k}{\| (Cx^t)_k \|} \).
Generalized Power Method:

1. Set $x^0$ to be a leading eigenvector of $C$ with $\|x^0\|_2 = \sqrt{n}$.

2. For $t = 0, 1, \ldots$, update $(x^{t+1})_k = \frac{(Cx^t)_k}{|(Cx^t)_k|}$.

**Theorem**

If $\sigma = O\left(\sqrt{\frac{n}{\log n}}\right)$, with high probability for large $n$, GPM converges linearly to the global optimum of (P) (unique up to phase.)
Fix (theoretically) the global phase such that $z^*\hat{x} = |z^*\hat{x}|$. 
Fix (theoretically) the global phase such that \( z^* \hat{x} = |z^* \hat{x}| \).

**Theorem**

If \( \sigma = O(\sqrt{n/\log n}) \), then w.h.p. for large \( n \),

\[
\| \hat{x} - z \|_2 = O(\sigma), \quad \text{and} \quad \| \hat{x} - z \|_\infty = O(\sigma \sqrt{\log n/n}).
\]
• Fix (theoretically) the global phase such that \( z^* \hat{x} = |z^* \hat{x}|. \)

**Theorem**

If \( \sigma = O(\sqrt{n/\log n}) \), then w.h.p. for large \( n \),

\[
\| \hat{x} - z \|_2 = O(\sigma), \text{ and} \\
\| \hat{x} - z \|_\infty = O(\sigma \sqrt{\log n/n}).
\]

• The eigenvector \( \tilde{x} \) has the same estimation error rate.
First analysis: eigenvector $\ell_\infty$ perturbation bound

- Low rank structure under our model:
  \[
  C = zz^* + \sigma W.
  \]

  Recall $\tilde{x}$ is the top eigenvector of $C$ with $\|\tilde{x}\|_2 = \sqrt{n}$. 

First analysis: eigenvector $\ell_\infty$ perturbation bound

- Low rank structure under our model:
  \[ C = zz^* + \sigma W. \]
  Recall $\tilde{x}$ is the top eigenvector of $C$ with $\|\tilde{x}\|_2 = \sqrt{n}$.

- The $\ell_2$ bound is easy: by Davis-Kahan, w.h.p.
  \[ \frac{1}{\sqrt{n}}\|\tilde{x} - z\| \leq \frac{\sigma\|W\|_{\text{op}}}{\lambda_1(zz^*)} = O\left(\frac{\sigma}{\sqrt{n}}\right) \]
First analysis: eigenvector $\ell_\infty$ perturbation bound

- Low rank structure under our model:
  \[ C = zz^* + \sigma W. \]
  Recall $\tilde{x}$ is the top eigenvector of $C$ with $\|\tilde{x}\|_2 = \sqrt{n}$.

- The $\ell_2$ bound is easy: by Davis-Kahan, w.h.p.
  \[
  \frac{1}{\sqrt{n}} \|\tilde{x} - z\| \leq \frac{\sigma \|W\|_{\text{op}}}{\lambda_1(zz^*)} = O\left(\frac{\sigma}{\sqrt{n}}\right)
  \]

- The $\ell_\infty$ bound is (a bit) hard:
  \[
  |\tilde{x}_m - z_m| = \left| \frac{(C\tilde{x})_m}{\lambda_1(C)} - z_m \right| \leq \left| \frac{|z^*\tilde{x}|}{\lambda_1(C)} - 1 \right| + \frac{\sigma |(W\tilde{x})_m|}{\lambda_1(C)}.
  \]
First analysis: eigenvector $\ell_\infty$ perturbation bound

- Low rank structure under our model:
  \[ C = zz^* + \sigma W. \]

  Recall $\tilde{x}$ is the top eigenvector of $C$ with $\|\tilde{x}\|_2 = \sqrt{n}$.

- The $\ell_2$ bound is easy: by Davis-Kahan, w.h.p.
  \[
  \frac{1}{\sqrt{n}} \|\tilde{x} - z\| \leq \frac{\sigma \|W\|_{\text{op}}}{\lambda_1(zz^*)} = O\left(\frac{\sigma}{\sqrt{n}}\right)
  \]

- The $\ell_\infty$ bound is (a bit) hard:
  \[
  |\tilde{x}_m - z_m| = \left| \frac{(C\tilde{x})_m}{\lambda_1(C)} - z_m \right| \leq \left| \frac{|z^*\tilde{x}|}{\lambda_1(C)} - 1 \right| + \frac{\sigma |(W\tilde{x})_m|}{\lambda_1(C)}.
  \]

- The goal: $\|W\tilde{x}\|_\infty = O(\sqrt{n \log n})$ w.h.p.

- Once this is proved, $\ell_\infty$ perturbation bound $\checkmark$. 

The idea: introduce auxiliary problems to decouple dependence (leave-one-out).

\[ C(m) = z z^* + \sigma W(m) \]

\[ W(m)_k^\ell = W_k^\ell \cdot \{ k \neq m \} \cdot \{ \ell \neq m \} \]

\[ \tilde{x}(m) = \text{leading eigenvector of } C(m) \]

Obs: \( C(m) \) is independent of \( m \)th row of \( W \), and w.h.p.

\[ |(W \tilde{x})_m| = |w^*_m \tilde{x}| \leq |w^*_m \tilde{x}(m)| + |w_m^*(\tilde{x} - \tilde{x}(m))| \leq |w^*_m \tilde{x}(m)| + \|w_m^*\| \cdot \|\tilde{x} - \tilde{x}(m)\| \leq O(\sqrt{n \log n}) + O(\sqrt{n}) \cdot \text{??} \]
First analysis: eigenvector $\ell_\infty$ perturbation bound

- The idea: introduce auxiliary problems to decouple dependence (leave-one-out).

- For each $m \in [n]$, define $C^{(m)} := zz^* + \sigma W^{(m)}$, with
  
  $$W_{k\ell}^{(m)} = W_{k\ell} 1\{k \neq m\} 1\{\ell \neq m\}, \quad \tilde{x}^{(m)} = \text{leading eigenvector of } C^{(m)}$$

\[
W^{(m)} = \begin{pmatrix}
W_{11} & W_{12} & 0 & W_{14} \\
W_{21} & W_{21} & 0 & W_{24} \\
0 & 0 & 0 & 0 \\
W_{41} & W_{42} & 0 & W_{44}
\end{pmatrix}
\]
First analysis: eigenvector $\ell_\infty$ perturbation bound

- The idea: introduce auxiliary problems to decouple dependence (leave-one-out).
- For each $m \in [n]$, define $C^{(m)} := zz^* + \sigma W^{(m)}$, with

$$W^{(m)}_{k\ell} = W_{k\ell} \mathbf{1}_{\{k \neq m\}} \mathbf{1}_{\{\ell \neq m\}}, \quad \tilde{x}^{(m)} = \text{leading eigenvector of } C^{(m)}$$

$$W^{(m)} = \begin{pmatrix}
W_{11} & W_{12} & 0 & W_{14} \\
W_{21} & W_{21} & 0 & W_{24} \\
0 & 0 & 0 & 0 \\
W_{41} & W_{42} & 0 & W_{44}
\end{pmatrix}$$

- Obs: $C^{(m)}$ is independent of $m$th row of $W$, and w.h.p.

$$| (W\tilde{x})_m | = | w_m^* \tilde{x} | \leq | w_m^* \tilde{x}^{(m)} | + | w_m^* (\tilde{x} - \tilde{x}^{(m)}) |$$

$$\leq | w_m^* \tilde{x}^{(m)} | + \| w_m \| \cdot \| \tilde{x} - \tilde{x}^{(m)} \|$$

$$\leq O(\sqrt{n\log n}) + O(\sqrt{n}) \cdot \text{???.}$$
To bound $\|\tilde{x} - \tilde{x}^{(m)}\|$, use a precise version of Davis-Kahan:

$$\frac{1}{\sqrt{n}} \|\tilde{x} - \tilde{x}^{(m)}\| = O\left(\frac{\sigma \| (W - W^{(m)}) \tilde{x}^{(m)} \|}{\sqrt{n}}\right) = O\left(\frac{\sqrt{\log n}}{n} \sigma\right) \text{ w.h.p.}$$

working! ✓
To bound $\|\tilde{x} - \tilde{x}^{(m)}\|$, use a precise version of Davis-Kahan:

$$\frac{1}{\sqrt{n}} \|\tilde{x} - \tilde{x}^{(m)}\| = O\left(\frac{\sigma\| (W - W^{(m)}) \tilde{x}^{(m)} \|}{\sqrt{n}}\right) = O\left(\frac{\sqrt{\log n}}{n} \sigma\right) \text{ w.h.p.}$$

working! ✓
Tracking $n$ Auxiliary Sequences

Introduce auxiliary sequences to analyze the MLE. Let $T$ be our GPM operator:

$$(T \times)^k = (C \times)^k | (C \times)^k.$$ Similarly, $$(T (m) \times)^k := (C (m) \times)^k | (C (m) \times)^k.$$ Define $n$ sequences:

$$\begin{align*}
# &= \%
\end{align*}$$

$$\begin{align*}
#(') &= \%,' *
\end{align*}$$

$$\begin{align*}
+,' &= \%,
\end{align*}$$

$$\begin{align*}
\ast(') &= \%\ast
\end{align*}$$

$$\begin{align*}
\ldots \ast\ast\ast(') &= \%\ast\ast\ast
\end{align*}$$

$$\begin{align*}
\ldots\ast\ast\ast\ast(') &= \%\ast\ast\ast\ast
\end{align*}$$

$$\begin{align*}
#(-) &= \%-+
\end{align*}$$

$$\begin{align*}
#(.) &= \%-+
\end{align*}$$

$$\begin{align*}
\ldots\ast\ast\ast\ast(-),\ast(.-) &= \%\ast\ast\ast\ast(.-)
\end{align*}$$

$$\begin{align*}
\ldots\ast\ast\ast\ast(.),\ast(.-) &= \%\ast\ast\ast\ast(.-)
\end{align*}$$
Introduce $n$ auxiliary sequences to analyze the MLE.
Introduce $n$ auxiliary sequences to analyze the MLE.

Let $\mathcal{T}$ be our GPM operator: $(\mathcal{T} x)_k = \frac{(Cx)_k}{|(Cx)_k|}$. Similarly, $(\mathcal{T}^{(m)} x)_k := \frac{(C^{(m)} x)_k}{|(C^{(m)} x)_k|}$. Define $n$ sequences:
Key: Contraction via induction.
\[ \Delta^{t+1,m} \leq \rho \Delta^{t,m} + \text{small discrepancy error (} \rho < 1). \]
Maintained throughout all iterates \( \Rightarrow \) guarantee for \( \hat{x} \).
A new method of analyzing nonconvex problems.
A new method of analyzing nonconvex problems.

*Key idea*: introducing auxiliary sequences to decouple + perturbation analysis
A new method of analyzing nonconvex problems.

**Key idea**: introducing auxiliary sequences to decouple + perturbation analysis

Can also analyze matrix completion, phase retrieval, blinded deconvolution, etc. [Chen et al., 2017].
Thank you!
