(Near-)optimal Results for Phase Synchronization

Yiqiao Zhong

Princeton University

with **Nicolas Boumal (PACM)**

SIAM AN18, Portland, July 10, 2018

Outline

Phase (angular) synchronization

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

Phase (angular) synchronization

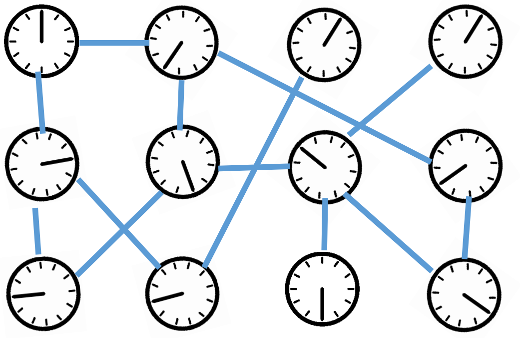
- Unknown parameters (angles): $\theta_1, \theta_2, \ldots, \theta_n \in [0, 2\pi)$.

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

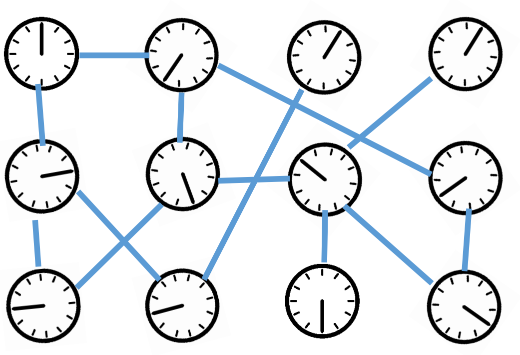
$$y_{\ell k} = \text{noisy version of } \theta_\ell - \theta_k \mod 2\pi,$$

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

- Time synchronization.



- Time synchronization.



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

- 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**:

$$C_{\ell k} = \bar{z}_\ell z_k + \sigma W_{\ell k}, \qquad \forall \ell > k$$

where $W_{\ell k} \sim N_{\mathbb{C}}(0,1)$. 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**:

$$C_{\ell k} = \bar{z}_\ell z_k + \sigma W_{\ell k}, \qquad \forall \ell > k$$

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 $|z_k| = 1$; $W_{kk} = 0$, $W_{k\ell} = \bar{W}_{\ell k}$.

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

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

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

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

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

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

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

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

- Denote the solution by $\widehat{x}$. 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,

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

- Denote the solution by $\widehat{x}$. Up to a global phase.
- Information limit: $\sigma = \sqrt{n}$.
- Our goal: under $\sigma = \tilde{O}(\sqrt{n})$,
  - Develop efficient algorithms that find $\widehat{x}$;
  - Derive statistical guarantees.

Standard recipe: semidefinite relaxation

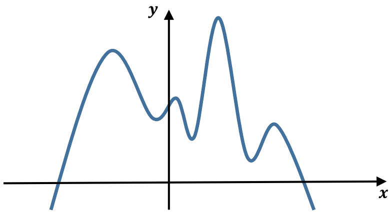
- Recall the MLE $\widehat{x}$ is a solution to:

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

- Recall the MLE $\widehat{x}$ is a solution to:

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

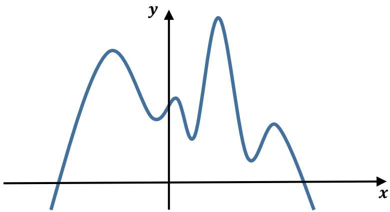
- Trouble...nonconvexity!



- Recall the MLE $\widehat{x}$ 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 in general. Zhang and Huang [2006]

Standard recipe: semidefinite relaxation

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

Standard recipe: semidefinite relaxation

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

$$X = xx^* \succeq 0$$

- Quadratic $\Rightarrow$ Linear:

$$x^* C x \Rightarrow \mathrm{Tr}(CX), \qquad |x_k| = 1 \Rightarrow X_{kk} = 1$$

Standard recipe: semidefinite relaxation

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

$$X = xx^* \succeq 0$$

- Quadratic $\Rightarrow$ Linear:

$$x^* C x \Rightarrow \mathrm{Tr}(CX), \qquad |x_k| = 1 \Rightarrow X_{kk} = 1$$

- Equivalently,

$$\max_{X \in \mathbb{C}^{n\times n}, X = X^*} \mathrm{Tr}(CX) \text{ subject to } \mathrm{diag}(X) = \mathbf{1}, X \succeq 0,$$

$$\mathrm{rank}(X) = 1.$$

Standard recipe: semidefinite relaxation

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

$$X = xx^*$$

- Quadratic $\Rightarrow$ Linear:

$$x^* C x \Rightarrow \mathrm{Tr}(CX), \qquad |x_k| = 1 \Rightarrow X_{kk} = 1$$

- semidefinite relaxation:

$$\max_{X \in \mathbb{C}^{n\times n}, X = X^*} \mathrm{Tr}(CX) \text{ subject to } \mathrm{diag}(X) = \mathbf{1}, X \succeq 0.$$

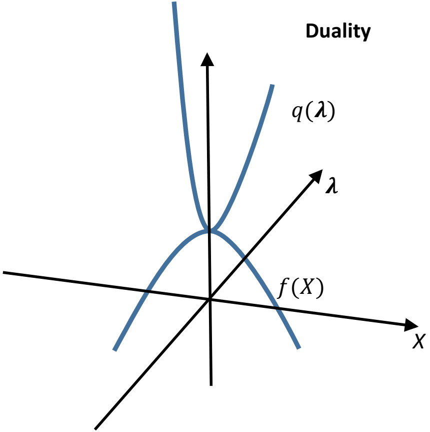
$$\sout{\mathrm{rank}(X) = 1} \qquad \text{(SDP)}$$

Standard recipe: semidefinite relaxation

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

Standard recipe: semidefinite relaxation

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



- Widely studied: compressed sensing, matrix completion, robust PCA, Stochastic block model, etc.

Standard recipe: semidefinite relaxation

- Phase sychronization: why difficult?

Standard recipe: semidefinite relaxation

- Phase sychronization: why difficult?
- Dual certificate:

$$S = \mathrm{Re}(\mathrm{ddiag}(C\widehat{x}\widehat{x}^*)) - C.$$

Goal: to show $S \succeq 0$.

Standard recipe: semidefinite relaxation

- Phase sychronization: why difficult?
- Dual certificate:

$$S = \mathrm{Re}(\mathrm{ddiag}(C\widehat{x}\widehat{x}^*)) - C.$$

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

'With high probability' is $1 - \mathcal{O}(n^{-2})$.

Faster approach: Generalized Power Method

- Beyond SDP?

Faster approach: Generalized Power Method

- Beyond SDP?
- Observe

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

Faster approach: Generalized Power Method

- Beyond SDP?

- Similar to the eigenvector problem!

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

$$\|x\| = 1$$

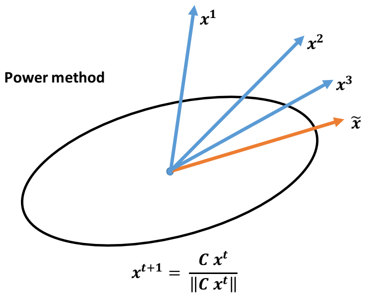
Faster approach: Generalized Power Method

- Beyond SDP?

- Similar to the eigenvector problem!

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

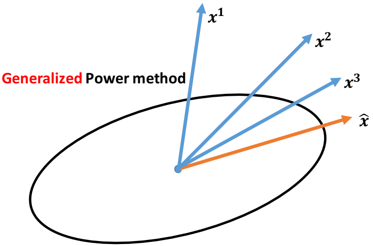
$$\|x\| = 1$$



- Beyond SDP?
- Similar to the eigenvector problem!

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

$$\|x\| = 1$$



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

Faster approach: 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 $(x^{t+1})_k = \frac{(Cx^t)_k}{|(Cx^t)_k|}$.

Faster approach: 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 $(x^{t+1})_k = \frac{(Cx^t)_k}{|(Cx^t)_k|}$.

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 $z^* \widehat{x} = |z^* \widehat{x}|$.

- Fix (theoretically) the global phase such that $z^* \widehat{x} = |z^* \widehat{x}|$.

Theorem

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

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

- Fix (theoretically) the global phase such that $z^*\widehat{x} = |z^*\widehat{x}|$.

Theorem

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

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

- The eigenvector $\tilde{x}$ has the same estimation error rate.

- 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 ℓ_∞ 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 ℓ_2 bound is easy: by Davis-Kahan, w.h.p.

$$\frac{1}{\sqrt{n}}\|\tilde{x} - z\| \leq \frac{\sigma\|W\|_{\mathrm{op}}}{\lambda_1(zz^*)} = O(\frac{\sigma}{\sqrt{n}})$$

First analysis: eigenvector ℓ_∞ 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 ℓ_2 bound is easy: by Davis-Kahan, w.h.p.

$$\frac{1}{\sqrt{n}}\|\tilde{x} - z\| \le \frac{\sigma\|W\|_{\mathrm{op}}}{\lambda_1(zz^*)} = O(\frac{\sigma}{\sqrt{n}})$$

- The ℓ_∞ bound is (a bit) hard:

$$|\tilde{x}_m - z_m| = \left|\frac{(C\tilde{x})_m}{\lambda_1(C)} - z_m\right| \le \left|\frac{|z^*\tilde{x}|}{\lambda_1(C)} - 1\right| + \frac{\sigma|(W\tilde{x})_m|}{\lambda_1(C)}.$$

First analysis: eigenvector ℓ_∞ 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 ℓ_2 bound is easy: by Davis-Kahan, w.h.p.

$$\frac{1}{\sqrt{n}}\|\tilde{x} - z\| \le \frac{\sigma\|W\|_{\mathrm{op}}}{\lambda_1(zz^*)} = O(\frac{\sigma}{\sqrt{n}})$$

- The ℓ_∞ bound is (a bit) hard:

$$|\tilde{x}_m - z_m| = \left|\frac{(C\tilde{x})_m}{\lambda_1(C)} - z_m\right| \le \left|\frac{|z^*\tilde{x}|}{\lambda_1(C)} - 1\right| + \frac{\sigma|(W\tilde{x})_m|}{\lambda_1(C)}.$$

- The goal: $\|W\tilde{x}\|_\infty = O(\sqrt{n\log n})$ w.h.p.
- Once this is proved, ℓ_∞ perturbation bound ✓.

First analysis: eigenvector ℓ_∞ perturbation bound

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

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

- For each $m \in [n]$, define $C^{(m)} := zz^* + \sigma W^{(m)}$, with

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

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

- 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 mth row of W, and w.h.p.

$$\begin{aligned} |(W\tilde{x})_m| = |w_m^*\tilde{x}| &\le |w_m^*\tilde{x}^{(m)}| + |w_m^*(\tilde{x} - \tilde{x}^{(m)})| \\ &\le |w_m^*\tilde{x}^{(m)}| + \|w_m\| \cdot \|\tilde{x} - \tilde{x}^{(m)}\| \\ &\le O(\sqrt{n\log n}) + O(\sqrt{n}) \cdot ???. \end{aligned}$$

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

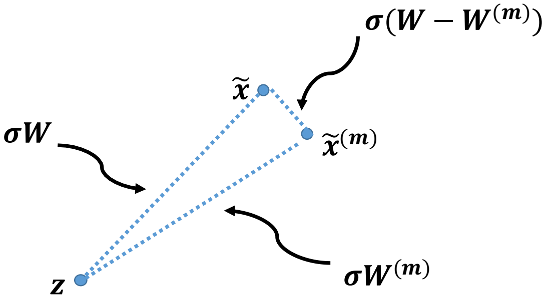
$$\frac{1}{\sqrt{n}}\|\tilde{x} - \tilde{x}^{(m)}\| = \mathcal{O}\Big(\frac{\sigma\|(W - W^{(m)})\frac{\tilde{x}^{(m)}}{\sqrt{n}}\|}{n}\Big) = O\big(\frac{\sqrt{\log n}}{n}\sigma\big) \quad w.h.p.$$

working! ✓

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

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

working! ✓

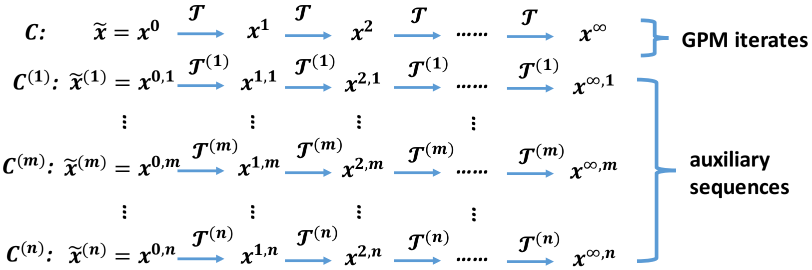


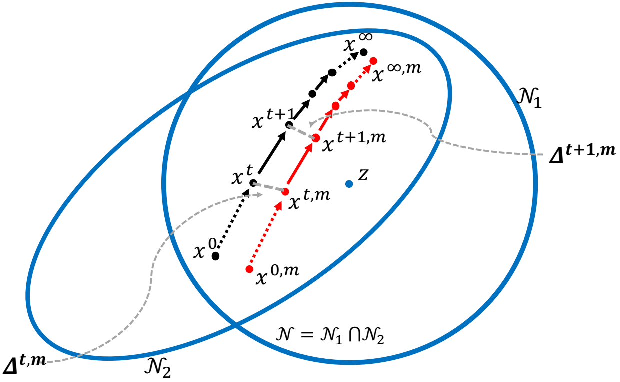
$$W = zz^* + \sigma W$$
$$W^{(m)} = zz^* + \sigma W^{(m)}$$

Tracking n Auxiliary Sequences

- 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.
- $\Delta^{t+1,m} \leq \rho \Delta^{t,m} +$ small discrepancy error ($\rho < 1$).
- Maintained throughout all iterates $\Rightarrow$ guarantee for $\widehat{x}$.

Take-away message

A new method of analyzing nonconvex problems.

Take-away message

A new method of analyzing nonconvex problems.

Key idea: introducing auxiliary sequences to decouple + perturbation analysis

Take-away message

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!

A.S. Bandeira, N. Boumal, and A. Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, pages 1–23, 2016. doi: 10.1007/s10107-016-1059-6.

Yuxin Chen, Jianqing Fan, Cong Ma, and Kaizheng Wang. Spectral method and regularized MLE are both optimal for top-K ranking. *arXiv preprint arXiv:1707.09971*, 2017.

S. Zhang and Y. Huang. Complex quadratic optimization and semidefinite programming. *SIAM Journal on Optimization*, 16(3):871–890, 2006.