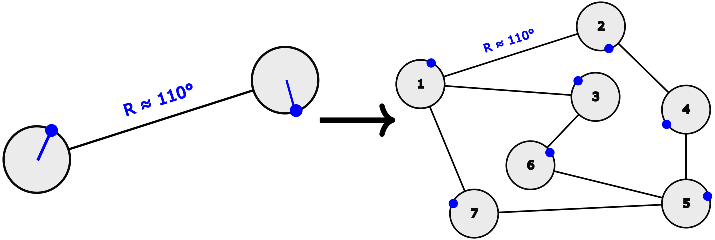
Benign nonconvexity of Burer-Monteiro SDP factorization for group synchronization

Andrew D. McRae

Institute of Mathematics, EPFL
Joint work with N. Boumal, P. Abdalla, A. Bandeira

ISMP
July 23, 2024

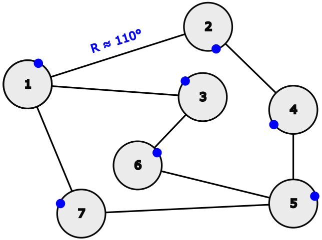
Motivation: the synchronization of rotations problem



Goal: estimate each node's angle from relative angles on edges

- Many applications in robotics, computer vision, signal processing...

General problem: orthogonal group synchronization on graph

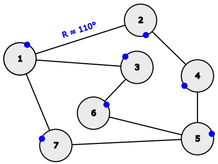


- Graph $G = (V, E)$ with vertices $V = \{1, \dots, n\}$
- Each node i has associated $r \times r$ orthogonal matrix Z_i $(Z_i Z_i^T = I_r)$
- Observed data: $R_{ij} \approx Z_i Z_j^T$ for $(i, j) \in E$
- Goal: estimate $Z_1, \dots, Z_n$

Optimization problem

Setup:

- Graph $G = (V, E)$ with vertices $V = \{1, \ldots, n\}$
- Want to estimate orthogonal matrices $Z_1, \ldots, Z_n$
- Observed data: $R_{ij} \approx Z_i Z_j^T$ for $(i, j) \in E$



Least-squares/maximum likelihood estimator:

$$\max_{Y_i \in \mathbf{R}^{r \times r}} \sum_{(i,j) \in E} \langle R_{ij}, Y_i Y_j^T \rangle \quad \text{s.t. } Y_i Y_i^T = I_r, i = 1, \ldots, n$$

QCQP form:

$$\max_{Y \in \mathbf{R}^{rn \times r}} \langle C, YY^T \rangle \quad \text{s.t. } \underbrace{\text{blkdiag}}_{n \text{ diag. } r \times r \text{ blks}}(YY^T) = I_{rn}$$

Nonconvex and in general has bad local optima.

- $r = 1$ is max-cut–type problem (NP-hard in general)
- What does **problem structure** buy us?

Relaxations

Original problem (can have **bad local minima**):

$$\max_{Y \in \mathbf{R}^{rn \times r}} \langle C, YY^T \rangle \text{ s.t. } \operatorname{blkdiag}(YY^T) = I_{rn}$$

Semidefinite relaxation (SDP):

$$\max_{X \in \mathbf{R}^{rn \times rn}} \langle C, X \rangle \text{ s.t. } \operatorname{blkdiag}(X) = I_{rn}, X \succeq 0.$$

Convex (no bad local minima) but **expensive** if n is large.

Our approach

Intermediate relaxation: for $p > r$:

$$\max_{Y \in \mathbf{R}^{rn \times p}} \langle C, YY^T \rangle \text{ s.t. } \operatorname{blkdiag}(YY^T) = I_{rn}$$

- Burer-Monteiro factorization of SDP
- Empirically successful in robotics literature (e.g., Rosen et al., 2019; Dellaert et al., 2020)
- Can we understand its performance theoretically?

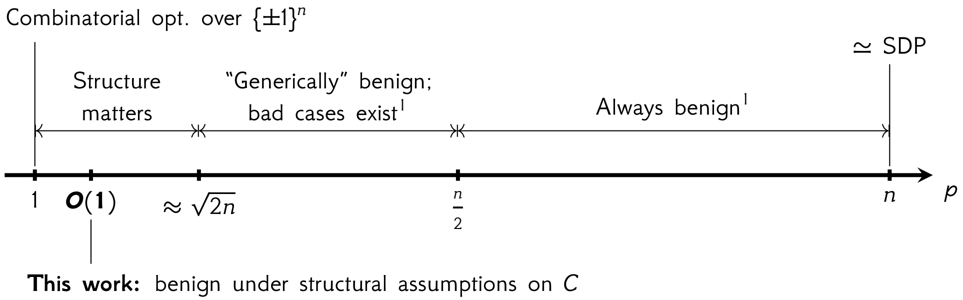
Landscapes landscape

Focus on simplest case $r = 1$: when does

$$\max_{Y \in \mathbf{R}^{n \times p}} \langle C, YY^T \rangle \text{ s.t. } \mathrm{diag}(YY^T) = I_n$$

(not) have bad local minima?

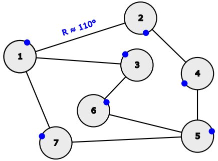
The answer depends on p and the cost matrix C:



[1] Boumal et al. (2019) and O'Carroll et al. (2022)

Main result

Theorem



Suppose

- *Connected measurement graph G on $1, \ldots, n$ with edges E.*
  - *Graph Laplacian matrix has second-smallest eigenvalue $\lambda_2 > 0$*
- *We observe $R_{ij} = Z_i Z_j^T + \Delta_{ij} \in \mathbf{R}^{r \times r}, (i,j) \in E$.*
- *$p \geq r + 3$, and we solve*

$$\max_{Y \in \mathbf{R}^{rn \times p}} \langle C, YY^T \rangle \text{ s.t. } \operatorname{blkdiag}(YY^T) = I_{rn}, \qquad C_{ij} = R_{ij} \text{ for } (i,j) \in E$$

*If $\|\boldsymbol{\Delta}\|_{\ell_2} \leq \boldsymbol{C_{p,r} \frac{\lambda_2}{\sqrt{n}}}$, any second-order critical point Y satisfies $YY^T = \widetilde{Z}\widetilde{Z}^T$, where $\widetilde{Z} \in \mathrm{O}(r)^n$ is a **global optimum to the original (unrelaxed) problem**.*

- Prior results for complete graph: we extend* to general graphs
- Noiseless case ("Kuramoto oscillator" sync.): $p \geq r + 2$ suffices
  - Optimal; previous best known condition $2p \geq 3(r+1)$ (Markdahl, 2021)

Proof ideas

Goal: show global optimality of critical points of

$$\max_{Y \in \mathbf{R}^{rn \times p}} \langle C, YY^T \rangle \text{ s.t. } \operatorname{blkdiag}(YY^T) = I_{rn} \tag{1}$$

Optimization over product of Stiefel manifolds.
Critical points Y of (1): setting $S(Y) = \operatorname{symblkdiag}(CYY^T) - C$

- First-order: $S(Y)Y = 0$ (Riemannian gradient is zero)
- Second-order: for any tangent vector $\dot{Y}$, $\langle S(Y), \dot{Y}\dot{Y}^T \rangle \geq 0$ (Riemannian Hessian $\preceq 0$)

Using these to prove optimality is an art and depends on the problem structure (particularly choosing $\dot{Y}$ in second-order condition)
Key innovations:

- A new **randomized** $\dot{Y}$ to plug into second-order inequality
- Study Laplacian of general graph (instead of complete-graph adjacency matrix)

Limitation: poor dimension scaling

Simple example

- Complete graph ($\lambda_2 = n$)
- Gaussian noise (entries of Δ are i.i.d. $\mathcal{N}(0, \sigma^2)$)

Our theorem requires

$$\sqrt{n}\sigma \approx \|\Delta\|_{\ell_2} \lesssim \frac{\lambda_2}{\sqrt{n}} = \sqrt{n} \Longleftrightarrow \sigma \lesssim 1$$

Previous results for complete-graph case:

- $\sigma \lesssim n^{1/4}$ suffices by similar but more specialized arguments (Ling, 2023b)
- $\sigma \lesssim \sqrt{\frac{n}{\log n}}$ suffices by involved “leave-one-out” analysis (Ling, 2022; Ling, 2023a)

Our results for general graphs **scale poorly** with the problem size n in some cases

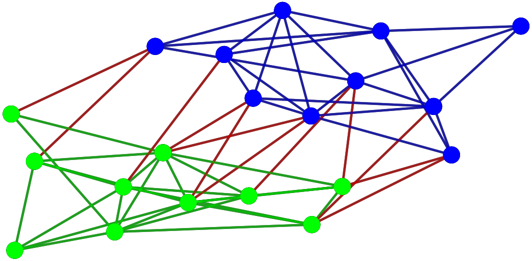
- **How can they be improved?**

Specialization to $r = 1$: $\mathbf{Z}_2$ synchronization

The orthogonal group $O(r)$ becomes **discrete** when $r = 1$:

$$O(1) = \{\pm 1\} = \mathbf{Z}_2$$

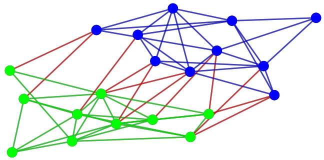
No longer "rotations" but closely linked to **graph clustering**



Even with noise, **exact recovery** becomes feasible

- Can we improve our results with this?

Stronger results for $\mathbf{Z}_2$ sync.



Theorem

Suppose $z_1, \ldots, z_n \in \{\pm 1\}$, and

- *We observe $R_{ij} = z_i z_j + \Delta_{ij} \in \mathbf{R}$, $(i, j) \in E$.*
- *$p \geq 4$, and we solve the rank-relaxed problem*

$$\max_{Y \in \mathbf{R}^{n \times p}} \langle C, YY^T \rangle \text{ s.t. } \operatorname{diag}(YY^T) = I_n, \qquad C_{ij} = R_{ij} \text{ for } (i, j) \in E.$$

Then, if

$$C_p \|\underset{\substack{\uparrow \\ \text{no } \sqrt{n}}}{\Delta}\|_{\ell_2} + \max_{1 \leq i \leq n} \left[-z_i \sum_{j \sim i} z_j \Delta_{ij} \right] \leq c_p \lambda_2,$$

*any second-order critical point Y satisfies $YY^T = zz^T$ (**exact recovery**)*

Back to complete graph with Gaussian noise

Our previous results required

$$\sqrt{n}\sigma \approx \|\Delta\|_{\ell_2} \lesssim \frac{\lambda_2}{\sqrt{n}} = \sqrt{n} \Leftrightarrow \sigma \lesssim 1$$

In the $\mathbf{Z}_2$ case, we need

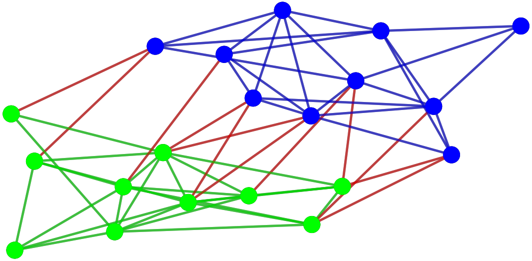
$$\begin{aligned} n = \lambda_2 &\gtrsim \|\Delta\|_{\ell_2} + \max_{1\le i\le n} \left[-z_i \sum_{j\sim i} z_j \Delta_{ij} \right] \\ &\approx \sqrt{n}\sigma + \sigma\sqrt{n\log n} \end{aligned} \quad \Leftrightarrow \quad \sigma \lesssim \sqrt{\frac{n}{\log n}}.$$

Recovers rate of Ling (2023a), however, working out the constants shows that

$$\sigma \le \sqrt{\frac{n}{(2+\epsilon)\log n}}$$

is the correct condition: this is **optimal** for exact recovery (Bandeira, 2018)

More optimal $\mathbf{Z}_2$ sync results

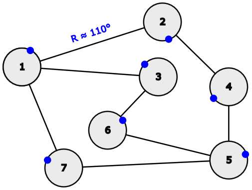


Our result also approaches optimal exact recovery thresholds for other benchmark problems:

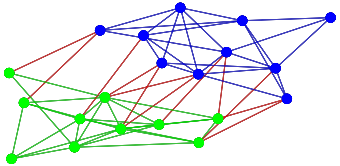
- Erdős–Rényi random graph with Bernoulli noise (“signed” or “correlation” clustering)
- Ordinary graph clustering with stochastic block model (Ling (2023a) again came within a constant)

Previously shown for SDP relaxation (Bandeira, 2018; Hajek et al., 2016a; Hajek et al., 2016b)

Summary



Group synchronization on graph



$\mathbf{Z}_2$ sync/graph clustering

We analyze the landscape of nonconvex QCQPs of the form

$$\max_{Y \in \mathbf{R}^{rn \times p}} \langle C, YY^T \rangle \text{ s.t. } \operatorname{blkdiag}(YY^T) = I_{rn}$$

with data from an O(r) synchronization problem.

- Benign nonconvex landscape results on a general graph
- Stronger (optimal) results and exact recovery in $\mathrm{O}(1) = \mathbf{Z}_2$ case

Future work: Improved conditions for general graphs and $r \geq 2$

- For example, by leave-one-out analysis like Ling (2022) and Ling (2023a)

Publications and acknowledgments

- Andrew D. McRae and Nicolas Boumal (2024). "Benign Landscapes of Low-Dimensional Relaxations for Orthogonal Synchronization on General Graphs". In: *SIAM J. Optim.* 34.2, pp. 1427–1454
- Andrew D. McRae, Pedro Abdalla, Afonso S. Bandeira, and Nicolas Boumal (2024). "Nonconvex landscapes for $\mathbf{Z}_2$ synchronization and graph clustering are benign near exact recovery thresholds". In: arXiv: 2407.13407 [math.OC]

Swiss National Science Foundation

References I

- Bandeira, Afonso S. (2018). "Random Laplacian Matrices and Convex Relaxations". In: *Found. Comput. Math.* 18, pp. 345–379.
- Boumal, Nicolas, Vladislav Voroninski, and Afonso S. Bandeira (2019). "Deterministic Guarantees for Burer-Monteiro Factorizations of Smooth Semidefinite Programs". In: *Commun. Pure Appl. Math.* 73.3, pp. 581–608.
- Dellaert, Frank, David M. Rosen, Jing Wu, Robert Mahony, and Luca Carlone (Aug. 2020). "Shonan Rotation Averaging: Global Optimality by Surfing $SO(p)^n$". In: *Proc. European Conf. Comput. Vis. (ECCV)*. Virtual conference, pp. 292–308.
- Hajek, Bruce, Yihong Wu, and Jiaming Xu (2016a). "Achieving Exact Cluster Recovery Threshold via Semidefinite Programming". In: *IEEE Trans. Inf. Theory* 62.5, pp. 2788–2797.
- — (2016b). "Achieving Exact Cluster Recovery Threshold via Semidefinite Programming: Extensions". In: *IEEE Trans. Inf. Theory* 62.10, pp. 5918–5937.

References II

- Ling, Shuyang (2022). "Improved Performance Guarantees for Orthogonal Group Synchronization via Generalized Power Method". In: *SIAM J. Optim.* 32.2, pp. 1018–1048.
- — (2023a). "Local Geometry Determines Global Landscape in Low-rank Factorization for Synchronization". In: arXiv: 2311.18670 [cs.IT].
- — (2023b). "Solving Orthogonal Group Synchronization via Convex and Low-Rank Optimization: Tightness and Landscape Analysis". In: *Math. Program.* 200, pp. 589–628.
- Markdahl, Johan (2021). "Synchronization on Riemannian Manifolds: Multiply Connected Implies Multistable". In: *IEEE Trans. Autom. Control* 66.9, pp. 4311–4318.
- McRae, Andrew D., Pedro Abdalla, Afonso S. Bandeira, and Nicolas Boumal (2024). "Nonconvex landscapes for $\mathbf{Z}_2$ synchronization and graph clustering are benign near exact recovery thresholds". In: arXiv: 2407.13407 [math.OC].
- McRae, Andrew D. and Nicolas Boumal (2024). "Benign Landscapes of Low-Dimensional Relaxations for Orthogonal Synchronization on General Graphs". In: *SIAM J. Optim.* 34.2, pp. 1427–1454.

References III

- O'Carroll, Liam, Vaidehi Srinivas, and Aravindan Vijayaraghavan (Dec. 2022). "The Burer-Monteiro SDP method can fail even above the Barvinok-Pataki bound". In: *Proc. Conf. Neural Inf. Process. Syst. (NeurIPS)*. New Orleans, Louisiana, pp. 31254–31264.
- Rosen, David M., Luca Carlone, Afonso S. Bandeira, and John J. Leonard (2019). "SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group". In: *Int. J. Robot. Res.* 38.2-3, pp. 95–125.