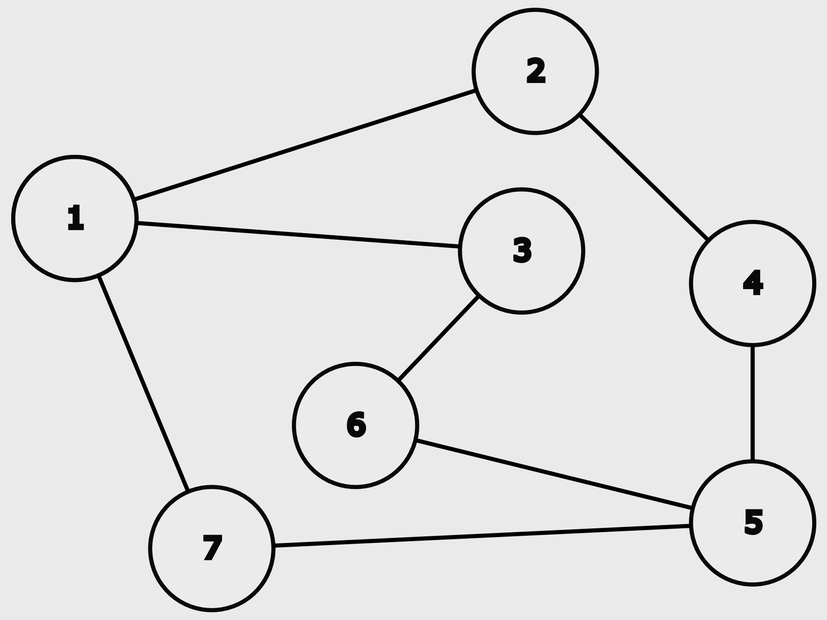
The rank-relaxed optimization landscape for orthogonal group synchronization on a general graph

Andrew D. McRae (joint work with Nicolas Boumal)

Institute of Mathematics, EPFL

An estimation problem



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

A first optimization approach

Least-squares optimization problem:

$$\min_{Y_i \in \mathbf{R}^{r \times r}} \sum_{(i,j) \in E} \|R_{ij} - Y_i Y_j^T\|_F^2 \text{ s.t. } Y_i Y_i^T = I_r, i = 1, \dots, n.$$

Equivalent problem (due to orthogonality of Z_i's):

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

Setting

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \in \mathbf{R}^{rn \times r},$$

we can rewrite this more compactly as

$$\max_{Y \in \mathbf{R}^{rn \times r}} \langle C, YY^T \rangle \text{ s.t. } Y_i Y_i^T = I_r, i = 1, \dots, n. \qquad \text{(LS)}$$

C is a symmetric $rn \times rn$ matrix formed from the observations $R_{ij}, (i, j) \in E$. This is well-known to have **bad local optima**.

SDP relaxation

The semidefinite relaxation of (LS) is

$$\max_{X \in \mathbf{R}^{rn \times rn}} \langle C, X \rangle \text{ s.t. } X_{ii} = I_r, 1, \dots, n, X \succeq 0. \qquad \text{(SDP)}$$

This is convex (no bad local optimum) but **expensive** if n is large.

Q: Is there an approach that

- is computationally efficient **and**
- has no bad local optimum?

Intermediate relaxation

The SDP relaxation replaces YY^T by a general PSD matrix $X \in \mathbf{R}^{rn \times rn}$ with potentially **much larger rank**:

- Original problem: for $Y \in \mathbf{R}^{rn \times r}$, YY^T has rank r
- SDP relaxation: matrix variable X can have rank as large as rn

Rank-p relaxation ($p \geq r$):

$$\max_{Y \in \mathbf{R}^{rn \times p}} \langle C, YY^T \rangle \text{ s.t. } Y_i Y_i^T = I_r, i = 1, \dots, n. \qquad \text{(Rank-}p\text{)}$$

This is a Burer–Monteiro factorization of (SDP).

Q: How big does p need to be to ensure no bad local optimum?

- For general C, we need $p = O(n)$, so there is no reduction in the number of variables over (SDP)
- Does our **problem structure** give a "benign optimization landscape" if p is much smaller?

Analysis ideas

The constraint set in (Rank-p) is a *Riemannian manifold* (specifically, a product of n Stiefel manifolds).

A **second-order critical point** Y of (Rank-p) satisfies, for every perturbation direction $\dot{Y}$ of Y,

$$\langle H(Y) \dot{Y}, \dot{Y} \rangle \leq 0,$$

where $H(Y)$ is the (Riemannian) Hessian that takes into account the constraints. Allowed perturbations are tangent vectors to the constraint manifold:

$$\dot{Y} = \begin{bmatrix} \dot{Y}_1 \\ \vdots \\ \dot{Y}_n \end{bmatrix} \text{ with } Y_i \dot{Y}_i^T + \dot{Y}_i Y_i^T = 0.$$

A general form for $\dot{Y}_i$ is

$$\dot{Y}_i = Y_i (Y_i \Gamma_i^T - \Gamma_i^T Y_i) \text{ for some } \Gamma_i \in \mathbf{R}^{r \times p}.$$

Key idea: We choose Γ_i *randomly* with

$$\Gamma_i = Z_i \Gamma, \text{ where } \Gamma \text{ has i.i.d. } \mathcal{N}(0, 1) \text{ entries.}$$

The rest of our proof is built on the inequality

$$0 \geq \langle H(Y) \dot{Y}, \dot{Y} \rangle = \langle H(Y), \mathbf{E}\, \dot{Y} \dot{Y}^T \rangle.$$

Other useful tools:

- Graph Laplacian matrix
- Dual analysis of (SDP)

Our proof heavily relies on the **measurement structure** $R_{ij} \approx Z_i Z_j^T$.

Theorem (Noiseless case)

Suppose $G = (V, E)$ is connected, and the measurements are exact, i.e., $R_{ij} = Z_i Z_j^T$ for $(i, j) \in E$. Then, if $\boldsymbol{p \geq r + 2}$, any second-order critical point Y of (Rank-p) *satisfies $YY^T = ZZ^T$. Thus* rank$(Y) = r$, *and Y exactly recovers the ground truth Z up to some global orthogonal transformation.*

Theorem (General noisy landscape)

Suppose $G = (V, E)$ is connected. Let $\Delta \in \mathbf{R}^{rn \times rn}$ be the portion of C corresponding to measurement error, i.e.,

$$\Delta_{ij} = \begin{cases} R_{ij} - Z_i Z_j^T & \text{if } (i, j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Let λ_2 be the Fiedler value of G, i.e., the second-smallest eigenvalue of the unnormalized graph Laplacian.

Then, if $\boldsymbol{p > r + 2}$

$$\|\Delta\|_{\ell_2} \leq \frac{c_{p,r}}{\sqrt{n}} \lambda_2,$$

then any second-order critical point Y of (Rank-p) *satisfies the following:*

- *Y is the unique solution to* (Rank-p) *up to a global orthogonal transformation.*
- *Y has rank r, and, if we write $Y = \hat{Z} U$ for some $\hat{Z} \in \mathbf{R}^{nr \times r}$ and $U \in \mathbf{R}^{r \times p}$ such that $UU^T = I_r$, $\hat{Z}$ is the unique solution of* (LS) *up to a global orthogonal transformation.*
- *YY^T is the unique solution to* (SDP).

Key takeaways

- Benign landscape for $p = O(1)$; we only need $O(n)$ variables in (Rank-p)
  - Good **computational efficiency**
- Noisy case: first such results for **general graph** G
  - Depends on "signal to noise ratio" $\lambda_2 / \|\Delta\|_{\ell_2}$
- In noiseless case, improves on previous best results requiring $2p \geq 3(r + 1)$.

Paper

A. D. McRae and N. Boumal, "Benign landscapes of low-dimensional nonconvex relaxations for orthogonal synchronization on general graphs," *Preprint*, 2023

go.epfl.ch/sync_landscape