Benign landscapes for synchronization on spheres via normalized Laplacian matrices

Andrew D. McRae

Institute of Mathematics
EPFL, Lausanne, Switzerland

EUROPT
July 2, 2025

The problem

We study a quadratic problem over n different $(r-1)$-dimensional spheres:

$$\max_{x_1,\ldots,x_n \in \mathbf{R}^r} \sum_{i,j=1}^{n} C_{ij} \langle x_i, x_j \rangle \text{ s.t. } \|x_i\| = 1 \ \forall i$$

QCQP form:

$$\max_{X \in \mathbf{R}^{n \times r}} \langle C, XX^T \rangle \text{ s.t. } \operatorname{diag}(XX^T) = \mathbf{1}$$

- **Nonconvex**, in general **NP-hard** (max-cut is one instance)
- Can we do better in some cases?
- In particular: what is the nonconvex **landscape**?
  - What can we say about (arbitrary) **local** optima?

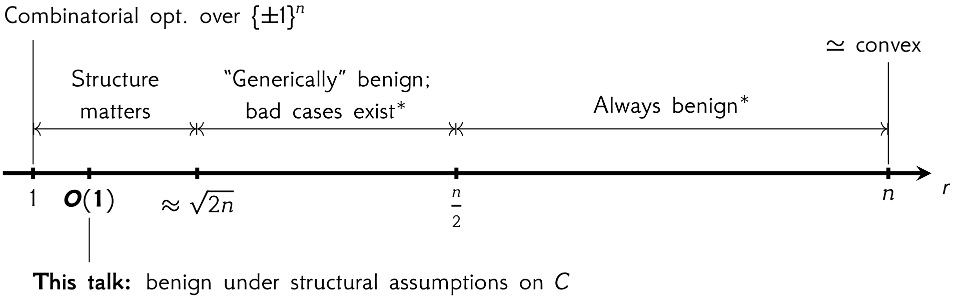
Landscapes landscape

When does

$$\max_{X \in \mathbf{R}^{n \times r}} \langle C, XX^T \rangle \text{ s.t. } \operatorname{diag}(XX^T) = \mathbf{1}$$

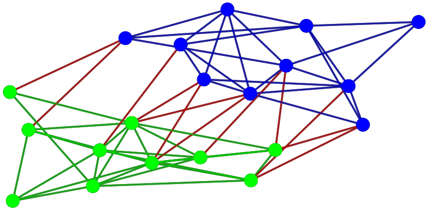
have spurious (non-global) local minima?

The answer depends on $\boldsymbol{r}$ and the cost matrix C:



*Boumal et al. (2019) and O'Carroll et al. (2022)

Example 1: graph clustering



- Graph $G = (V, E)$, $V = \{1, \ldots, n\}$
- We want to **label** the vertices in a way that corresponds to the edge information

Signed graph clustering

For simplicity, consider **signed** graph clustering

- (Unsigned clustering also works with some tweaks)

If $z_1, \dots, z_n \in \{\pm 1\}$ are "true" cluster labels, we (approximately) observe relative signs

$$R_{ij} \approx z_i z_j \qquad \text{for} \qquad (i,j) \in E$$

Estimate of clusters:

$$\arg\max_{x \in \{\pm 1\}^n} \underbrace{\sum_{(i,j)\in E} R_{ij} x_i x_j}_{= \langle C, xx^T \rangle}$$

This is a discrete problem

- **Q:** What is a good algorithm?

Continuous spherical relaxation

We are maximizing

$$\langle C, xx^T \rangle = \sum_{i,j} C_{ij} x_i x_j .$$

We can make this continuous and smooth by **relaxing**

$$x_i x_j, \quad x_1, \dots, x_n \in \{\pm 1\}$$

$$\Big\downarrow$$

$$\langle x_i, x_j \rangle, \quad x_1, \dots, x_n \in \mathbf{R}^r, \|x_i\| = 1, r \geq 2$$

We thus obtain our spherical QCQP:

$$\max_{X \in \mathbf{R}^{n \times r}} \; \langle C, XX^T \rangle \text{ s.t. } \operatorname{diag}(XX^T) = \mathbf{1}$$

(Another question: is this **tight**?)

Example 2: oscillator synchronization

Dynamical system of n angles $\theta_1, \dots, \theta_n$

$$\frac{d\theta_i}{dt} = \sum_{j=1}^{n} A_{ij} \sin(\theta_j - \theta_i), \qquad i = 1, \dots, n$$

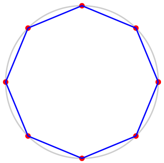
Simple version of the "Kuramoto model"—angles are phases of interacting oscillators

If A is the adjacency matrix of a connected graph, a family of stable equilibria is

$$\theta_1 = \cdots = \theta_n \mod 2\pi.$$

Q: are these the **only** stable equilibria?

- Model of "spontaneous synchronization" of natural systems
- Pendulums, fireflies, radio electronics...
- Answer depends on the coupling matrix A



(Counter)example

Connection to optimization

"Kuramoto oscillator" dynamical system:

$$\frac{d\theta_i}{dt} = \sum_{j=1}^{n} A_{ij} \sin(\theta_j - \theta_i), \qquad i = 1, \dots, n \tag{KUR}$$

This is (maximizing) gradient flow of the (negative) potential

$$\sum_{i,j} A_{ij} \cos(\theta_i - \theta_j)$$

By the angle parametrization of the unit circle, this is equivalent to

$$\max_{X \in \mathbf{R}^{n \times 2}} \langle A, XX^T \rangle \text{ s.t. } \operatorname{diag}(XX^T) = \mathbf{1} \tag{POT}$$

Why is this useful? (See, e.g., Ling et al., 2019)

(KUR) "generically" synchronizes $\iff^*$ only local optima of (POT) are $XX^T = \mathbf{1}\mathbf{1}^T$

How to analyze (non)convex landscape?

Nonconvex problem

$$\max_{X \in \mathbf{R}^{n\times r}} \langle C, XX^T \rangle \text{ s.t. } \operatorname{diag}(XX^T) = \mathbf{1}$$

For our applications, we hope local minima have the form $XX^T = zz^T$ for some $z \in \{\pm 1\}$.

Compare this to the **semidefinite** relaxation

$$\max_{Z \succeq 0} \langle C, Z \rangle \text{ s.t. } \operatorname{diag}(Z) = \mathbf{1}.$$

This is convex, so there is a well-developed theory of how to certify a given solution.

Convex dual certificate

How do we show that $Z_* \succeq 0$ is optimal for

$$\max_{Z \succeq 0} \ \langle C, Z \rangle \ \text{s.t.} \ \operatorname{diag}(Z) = \mathbf{1}?$$

Suppose there is a **diagonal** matrix Λ (Lagrange multipliers) such that

$$\begin{aligned} S := \Lambda - C \quad & \text{satisfies} \\ S \succeq 0, \quad & \text{and} \\ SZ_* = 0 & \end{aligned}$$

Then, for any feasible Z ($Z \succeq 0$ and $\operatorname{diag}(Z) = \mathbf{1}$),

$$\begin{aligned} \langle C, Z_* \rangle - \langle C, Z \rangle &= \underbrace{\langle \Lambda, Z_* - Z \rangle}_{=0} + \langle S, Z \rangle - \underbrace{\langle S, Z_* \rangle}_{=0} \\ &= \langle S, Z \rangle \\ &\geq 0 \end{aligned}$$

Convex dual certificate

Dual certificate of $Z_* \succeq 0$ for

$$\max_{Z \succeq 0} \langle C, Z \rangle \text{ s.t. } \operatorname{diag}(Z) = \mathbf{1}$$

is $S = \Lambda - C$ for diagonal Λ with

$$S \succeq 0, \quad \text{and}$$
$$SZ_* = 0.$$

If $Z_* = zz^*$ for $z \in \{\pm 1\}^n$, then

$$Szz^* = 0 \quad \Leftrightarrow \quad \Lambda z = Cz \quad \Leftrightarrow \quad \Lambda = \operatorname{diag}(z \circ (Cz)).$$

Hence

$$S = S(z) := \operatorname{diag}(z \circ (Cz)) - C.$$

Dual certificate eigenvalues

Dual certificate (we hope):

$$S = S(z) = \mathrm{diag}(z \circ (Cz)) - C$$

- We need $Sz = 0$ ✅ and $\boldsymbol{S} \succeq \mathbf{0}$
- Let $\lambda_1 \leq \cdots \leq \lambda_n$ be its (real) eigenvalues.
- $Sz = 0 \implies S$ has a zero eigenvalue
- If $\lambda_2 > 0$, indeed $S \succeq 0$
  - Furthermore, the solution zz^T is unique and the SDP is **tight**
- **Analysis recipe**: for some $z \in \{\pm 1\}$, **prove** that $\lambda_2 > 0$ (based on problem structure)

The eigenvalues turn out to be key in understanding the **nonconvex** landscape as well...

A nonconvex landscape result

Consider the nonconvex problem

$$\max_{X \in \mathbf{R}^{n \times r}} \langle C, XX^T \rangle \text{ s.t. } \operatorname{diag}(XX^T) = \mathbf{1} \qquad \text{(NCVX)}$$

Theorem (Rakoto Endor and Waldspurger (2024))

For $z \in \{\pm 1\}$, suppose $S := \operatorname{diag}(z \circ (Cz)) - C$ satisfies $\lambda_2(S) > 0$. Then, if

$$\frac{\lambda_n(S)}{\lambda_2(S)} < r,$$

every local optimum X of (NCVX) satisfies $XX^T = zz^T$.*

- Optimal in some cases

*Second-order (gradient and Hessian) optimality suffices.

Example: synchronization on Erdős–Rényi graph

Simple example:

- $z = \mathbf{1}$, $r = 2$
- Cost matrix $C = A$, where A is adjacency matrix of **Erdős–Rényi random graph** $\mathcal{G}(n, p)$:

$$A_{ij} = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

Oscillator network on graph synchronizes $\iff$ benign landscape of

$$\max_{X \in \mathbf{R}^{n \times 2}} \langle A, XX^T \rangle \text{ s.t. } \operatorname{diag}(XX^T) = \mathbf{1}.$$

$XX^T = \mathbf{1}\mathbf{1}^T$ is the unique **global** optimum when the graph is connected

- Happens with probability $\to 1$ as $n \to \infty$ if

$$p \geq (1 + \epsilon)\frac{\log n}{n}$$

- But we need to rule out non-synced **local** optima...

Landscape of Erdős–Rényi graph synchronization

- A adjacency matrix of Erdős–Rényi random graph $\mathcal{G}(n, p)$
- Dual certificate is just the graph Laplacian:

$$S = L = \operatorname{diag}(A\mathbf{1}) - A$$

For an Erdős–Rényi graph,

$$\mathbf{E}\, L = np\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \quad \Longrightarrow \quad \frac{\lambda_n(\mathbf{E}\, L)}{\lambda_2(\mathbf{E}\, L)} = 1.$$

However, with randomness, to have $\frac{\lambda_n(L)}{\lambda_2(L)} < 2$ requires

$$p \gg \frac{\log n}{n}$$ 😕

Can we do better?

Noise characteristics

- A adjacency matrix of Erdős–Rényi random graph $\mathcal{G}(n, p)$
- Dual certificate is the graph Laplacian:

$$S = L = \mathrm{diag}(A\mathbf{1}) - A$$

Write

$$A = \mathbf{E}A + \underbrace{(A - \mathbf{E}A)}_{=:\Delta_A} = p\mathbf{1}\mathbf{1}^T + \Delta_A$$
$$\Downarrow$$
$$L = np\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) + \underbrace{\mathrm{diag}(\Delta_A\mathbf{1}) - \Delta_A}_{=:\Delta_L}$$

For $p \approx \frac{\log n}{n}$, the noise spectral norm $\|\Delta_L\|_{\ell_2}$ is dominated by the **diagonal** $\mathrm{diag}(\Delta_A\mathbf{1})$.

Improved landscape result—diagonal preconditioning

Consider the nonconvex problem

$$\max_{X \in \mathbf{R}^{n \times r}} \langle C, XX^T \rangle \text{ s.t. } \operatorname{diag}(XX^T) = \mathbf{1} \tag{NCVX}$$

Theorem (McRae (2025))

For $z \in \{\pm 1\}$, *suppose* $S := \operatorname{diag}(z \circ (Cz)) - C$ *satisfies* $\lambda_2(S) > 0$.
Let D *be any* ***diagonal*** *matrix with* ***strictly positive*** *diagonal entries. If*

$$\frac{\lambda_n(D^{-1/2} S D^{-1/2})}{\lambda_2(D^{-1/2} S D^{-1/2})} < r,$$

every local optimum X *of* (NCVX) *satisfies* $XX^T = zz^T$.

Back to Erdős–Rényi graphs

For A adjacency matrix of Erdős–Rényi random graph $\mathcal{G}(n, p)$,

$$S = L = \operatorname{diag}(A\mathbf{1}) - A$$

Take the preconditioner to be the vertex degree matrix:

$$D = \operatorname{diag}(A\mathbf{1})$$

Then

$$D^{-1/2} L D^{-1/2} = I_n - D^{-1/2} A D^{-1/2} =: \mathcal{L}$$

is the (symmetric) **normalized graph Laplacian**.

- Much better spectral concentration than ordinary graph Laplacian
- As long as $p \geq (1 + \epsilon)\frac{\log n}{n}$ (connectivity threshold)

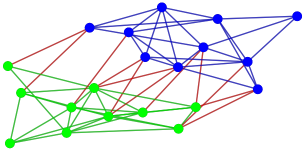
$$\frac{\lambda_n(\mathcal{L})}{\lambda_2(\mathcal{L})} \to 1 < 2 \quad \text{as} \quad n \to \infty$$

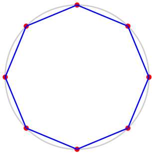
(Hoffman et al., 2021, for example)

Applications

We have synchronization of Kuramoto oscillator networks on Erdős–Rényi graphs $\mathcal{G}(n, p)$ for $p \geq (1+\epsilon)\frac{\log n}{n}$ (connectivity threshold):

- This result already shown by more specialized analysis (Abdalla et al., 2022)
  - Previous analysis is complicated and **fails for other problems**
- Our normalized condition number analysis is more general and applies to
  - Oscillator networks with negative edges (repulsion)
  - Graph clustering (with noise)
- Results are information-theoretically **optimal** for several popular random models
  - **Previous work** (McRae et al., 2025) required rank r **large near threshold**
  - New work only needs $r = 2$ (circles)





Conclusions

We looked at nonconvex landscape of spherical optimization problems

$$\max_{X \in \mathbf{R}^{n \times r}} \langle C, XX^T \rangle \text{ s.t. } \operatorname{diag}(XX^T) = \mathbf{1} \qquad (r \text{ is } \textbf{small})$$

Parting thoughts

- Structure (of C) coming from the **application** is critical
- Analysis depends on knowing the optimum **in advance**
  - Extensions to orthogonal/unitary groups possible...
  - $\{\pm 1\}$ special because **exact** recovery possible with noise

Preprint: Andrew D. McRae (2025). "Benign landscapes for synchronization on spheres via normalized Laplacian matrices". In: arXiv: 2503.18801 [math.OC]

Thanks!

References I

- Abdalla, Pedro, Afonso S. Bandeira, Martin Kassabov, Victor Souza, Steven H. Strogatz, and Alex Townsend (Oct. 23, 2022). "Expander graphs are globally synchronising". In: arXiv: 2210.12788 [math.CO].
- 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.
- Hoffman, Christopher, Matthew Kahle, and Elliot Paquette (2021). "Spectral Gaps of Random Graphs and Applications". In: *Int. Math. Res. Notices* 2021.11, pp. 8353–8404.
- Ling, Shuyang, Ruitu Xu, and Afonso S. Bandeira (2019). "On the Landscape of Synchronization Networks: A Perspective from Nonconvex Optimization". In: *SIAM J. Optim.* 29.3, pp. 1879–1907.
- McRae, Andrew D. (2025). "Benign landscapes for synchronization on spheres via normalized Laplacian matrices". In: arXiv: 2503.18801 [math.OC].
- McRae, Andrew D., Pedro Abdalla, Afonso S. Bandeira, and Nicolas Boumal (2025). "Nonconvex landscapes for $\mathbf{Z}_2$ synchronization and graph clustering are benign near exact recovery thresholds". In: *Inform. Inference.* 14.2.

References II

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.

Rakoto Endor, Faniriana and Irène Waldspurger (2024). "Benign landscape for Burer-Monteiro factorizations of MaxCut-type semidefinite programs". In: arXiv: 2411.03103 [math.OC].