Rotation Group Synchronization via Quotient Manifold

Linglingzhi Zhu

Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong (CUHK)

Joint work with Chong Li and Anthony Man-Cho So

SIAM Conference on Optimization 2023 (OP23)

June 3, 2023

Registration and travel support for this presentation was provided by the Society for Industrial and Applied Mathematics.

Rotation Group Synchronization

The rotation group elements (**Ground-Truth**)

$$G^\star = (G_1^\star, \ldots, G_n^\star) \in \mathcal{SO}(d)^n$$

is the target to be estimated, where

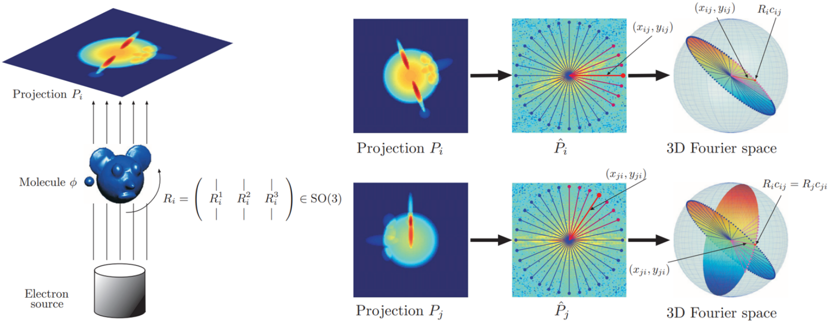
$$\mathcal{SO}(d) = \left\{ Q \in \mathbb{R}^{d\times d} : QQ^\top = Q^\top Q = I_d,\ \det(Q) = 1 \right\}.$$

Task: Recover $G^\star$ from $\left\{ C_{ij} \in \mathbb{R}^{d\times d} : 1 \leqslant i < j \leqslant n \right\}$

- C_{ij}: noisy measurement of relative transform $G_i^\star G_j^{\star\top}$;
- **(Generative Model)** $C_{ij} = G_i^\star G_j^{\star\top} + \Delta_{ij}$.

Examples of Applications

- Computer Vision
  - Cryo-Electron Microscopy [Singer, 2018, Singer and Shkolnisky, 2011]
  - Point Set Registration [Khoo and Kapoor, 2016]
  - Multiview Structure from Motion [Arie-Nachimson et al., 2012]
- Robotics
  - Simultaneous Localization and Mapping [Rosen et al., 2019]



Nonconvex Least Squares Formulation

Least squares estimator:

$$\min_{G_1,\ldots,G_n \in \mathcal{SO}(d)} \sum_{i<j} \| G_i G_j^\top - C_{ij} \|_F^2 \tag{LS}$$

$$\overset{G_i \in \mathcal{SO}(d)}{\Longleftrightarrow} \quad \max_{G \in \mathcal{SO}(d)^n} \operatorname{tr}(G^\top C G) \tag{QP-S}$$

where $G = (G_1, \ldots, G_n) \in \mathcal{SO}(d)^n$ and $C \in \mathbb{R}^{nd \times nd}$.

(QP-S) is **nonconvex** QP over $\mathcal{SO}(d)^n$

- Global optimum? C owns **generative model**;
- ($d = 2$) Phase synchronization (commutative group $\mathcal{SO}(2)$)
  [Boumal, 2016, Liu et al., 2017, Zhong and Boumal, 2018]

Existing Approaches for Solving (QP-S)

Step 1: Relax (QP-S) to

$$\max_{G \in \mathcal{O}(d)^n} \operatorname{tr}(G^\top C G) \tag{QP-O}$$

Step 2: Solve (QP-O) by **Generalized Power Method (GPM)** [Liu et al., 2020, Zhu et al., 2021, Ling, 2022a]:

$$G^{k+1} \in \operatorname{Proj}_{\mathcal{O}(d)^n}((C + \alpha I_{nd}) G^k).$$

Further relaxed form:

- **SDR** [Singer, 2011, Bandeira et al., 2017, Won et al., 2022]

$$\max_{X \in \mathbb{R}^{nd \times nd}} \operatorname{tr}(CX) \quad \text{s.t.} \quad X_{ii} = I_d, \; X \succeq 0$$

- **Burer-Monteiro** [Boumal et al., 2016, Ling, 2022c]

$$\max_{X \in \mathbb{R}^{nd \times p}} \operatorname{tr}(CXX^\top) \quad \text{s.t.} \quad X_i X_i^\top = I_d, \, X := [X_1; \ldots; X_n]$$

- **Spectral Relaxation** [Singer, 2011, Ling, 2022b]

$$\max_{X \in \mathbb{R}^{nd \times d}} \operatorname{tr}(CXX^\top) \quad \text{s.t.} \quad X^\top X = n \cdot I_d$$

Main Questions

Q1: Is the relaxation in **Step 1** reasonable?

$$\max_{G \in \mathcal{SO}(d)^n} \operatorname{tr}(G^\top C G) \quad \Longrightarrow \quad \max_{G \in \mathcal{O}(d)^n} \operatorname{tr}(G^\top C G)$$

Q2: For **Step 2**, whether we can design **simple and fast** algorithms utilizing **intrinsic** manifold structure?

Q3: Does (QP-S)/(QP-O) have **good landscape** that allows us to find a **global optimum** with fast convergence though it is **nonconvex**?

Main Questions

Q1: Is the relaxation in **Step 1** reasonable?

$$\max_{G\in\mathcal{SO}(d)^n} \operatorname{tr}(G^\top CG) \quad\Longrightarrow\quad \max_{G\in\mathcal{O}(d)^n} \operatorname{tr}(G^\top CG)$$

✓ Under generative model with deterministic noise when exact recovery.

Q2: For **Step 2**, whether we can design **simple and fast** algorithms utilizing **intrinsic** manifold structure?

Q3: Does (QP-S)/(QP-O) have **good landscape** that allows us to find a **global optimum** with fast convergence though it is **nonconvex**?

Main Questions

Q1: Is the relaxation in **Step 1** reasonable?

$$\max_{G \in \mathcal{SO}(d)^n} \operatorname{tr}(G^\top C G) \quad \Longrightarrow \quad \max_{G \in \mathcal{O}(d)^n} \operatorname{tr}(G^\top C G)$$

✓ Under generative model with deterministic noise when exact recovery.

Q2: For **Step 2**, whether we can design **simple and fast** algorithms utilizing **intrinsic** manifold structure?

✓ Riemannian algorithms stay on a connected component of orthogonal group naturally, e.g., Riemannian gradient method (RGM).

Q3: Does (QP-S)/(QP-O) have **good landscape** that allows us to find a **global optimum** with fast convergence though it is **nonconvex**?

Main Questions

Q1: Is the relaxation in **Step 1** reasonable?

$$\max_{G \in \mathcal{SO}(d)^n} \operatorname{tr}(G^\top C G) \quad \Longrightarrow \quad \max_{G \in \mathcal{O}(d)^n} \operatorname{tr}(G^\top C G)$$

✓ Under generative model with deterministic noise when exact recovery.

Q2: For **Step 2**, whether we can design **simple and fast** algorithms utilizing **intrinsic** manifold structure?

✓ Riemannian algorithms stay on a connected component of orthogonal group naturally, e.g., Riemannian gradient method (RGM).

Q3: Does (QP-S)/(QP-O) have **good landscape** that allows us to find a **global optimum** with fast convergence though it is **nonconvex**?

✓ Benefitted from the quotient geometric view.

Quotient View

For $\mathcal{G}^n = \mathcal{O}(d)^n$ or $\mathcal{SO}(d)^n$

$$\max_{G \in \mathcal{G}^n} \bar{f}(G) := \mathrm{tr}(G^\top C G) \tag{QP}$$

✗ **NP-hard** as QPQC (reduced to Max-Cut when $\mathcal{G}^n = \mathcal{O}(d)^n$, $d = 1$).

- **Generative model:**

$$C_{ij} = G_i^\star G_j^{\star\top} + \Delta_{ij}, \quad \Delta_{ij} : \text{deterministic noise}$$

- **Quotient equivalent form:**

$$\max_{[G] \in \mathcal{Q}} f([G]) := \mathrm{tr}(g^\top G^\top C G g) = \mathrm{tr}(G^\top C G) \tag{Q}$$

- $[G] := \{G' \in \mathcal{G}^n \mid G' = Gg,\ g \in \mathcal{G}\}$
- $\mathcal{Q} := \mathcal{G}^n / \mathcal{G}$

Improved Deterministic Estimation Performance

Lemma ([Zhu et al., 2021, Lemma 4.1])

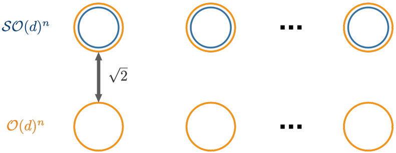
Let $\hat{G}$ *be an optimal solution of* (QP-O). *Then*[1] $\mathrm{d}_F([\hat{G}], [G^\star]) \lesssim \frac{\sqrt{d}\|\Delta\|}{\sqrt{n}}$.

✗ Gaussian random matrix $\|\Delta\| \lesssim \sqrt{nd} \Rightarrow$ constant noise level for exact recovery

Theorem (ℓ_∞ Estimation: from Average to Worst Case)

If $\|\Delta\| \lesssim \frac{n}{\sqrt{d}}$, *then*[1] $\mathrm{d}_\infty([\hat{G}], [G^\star]) \leqslant \|\hat{G}\hat{g}^\star - G^\star\|_\infty \lesssim \frac{\|\Delta\hat{G}\|_\infty}{n}$.

- $\|\Delta\| \lesssim \frac{n}{\sqrt{d}}$, $\|\Delta\hat{G}\|_\infty \lesssim n \Rightarrow \mathrm{d}_\infty([\hat{G}], [G^\star]) = \mathcal{O}(1)$;
- $\hat{G}$ in **same connected component** with $G^\star$ (o/w $\mathrm{d}_\infty([\hat{G}], [G^\star]) \geqslant \sqrt{2}$).



✓ Tightness of (QP-O) for (QP-S);

✓ GPM, SDR, BM, SpecR
for solving rotation synchronization.

(Quotient) Riemannian Algorithms

$$\max_{G \in \mathcal{G}^n} \bar{f}(G) := \mathrm{tr}(G^\top C G) \quad \text{and} \quad \max_{[G] \in \mathcal{Q}} f([G]) := \mathrm{tr}(G^\top C G)$$

Advantages:

- Keep on same connected component automatically
  - ✓ Naturally feasible for rotation group synchronization
  - ✓ Regardless of noise level
- Lower computational cost
  - ✓ Dimension reduction

	SDR	GPM	**Riemann**	**Quotient**
Dimension	$n^2 d^2$	nd^2	$\frac{1}{2}nd(d-1)$	$\frac{1}{2}(n-1)d(d-1)$
Dim ($d=3$)	$9n^2$	$9n$	$3n$	$3n-3$

  - ✓ SVD free: Projection $\Rightarrow$ Exponential map with explicit form

How can we design (quotient) Riemannian algorithms?

(Quotient) Riemannian Gradient Method

Algorithm 1 (Quotient) Riemannian gradient method

1: **Input:** The matrix C, the stepsize $t_k \geqslant 0$ and initial point $G^0 \in \mathcal{G}^n$.
2: **for** $k = 0, 1, \ldots$ **do**
3: Compute $[G^{k+1}] := \mathrm{Exp}_{[G^k]}(t_k \operatorname{grad} f([G^k]))$.
4: **end for**

Questions:

- How can we calculate "$\operatorname{grad} f([G^k])$"?
- Relationship to RGM: $G^{k+1} := \mathrm{Exp}_{G^k}(t_k \operatorname{grad} \bar{f}(G^k))$?

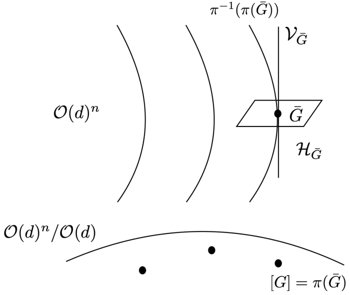
Quotient Manifold and Tangent Space

- Canonical projection $\pi : \mathcal{O}(d)^n \to \mathcal{Q}$, $\pi(G) := [G]$
- Vertical space $\mathcal{V}_{\bar{G}}$: $\mathrm{T}_{\bar{G}}(\pi^{-1}([G]))$
- **Horizontal space** $\mathcal{H}_{\bar{G}}$: $\mathcal{H}_{\bar{G}} \oplus \mathcal{V}_{\bar{G}} = \mathrm{T}_{\bar{G}}\, \mathcal{O}(d)^n$

Definition (Lifted Representation of $\mathrm{T}_{[G]}\, \mathcal{Q}$ on $\mathcal{O}(d)^n$)

The **horizontal lift** of $\xi_{[G]} \in \mathrm{T}_{[G]}\, \mathcal{Q}$ at $\bar{G} \in \pi^{-1}([G])$ is the unique vector

$$\bar{\xi}_{\bar{G}} \in \mathcal{H}_{\bar{G}} \text{ such that } \mathrm{D}\, \pi(\bar{G}) \left[\bar{\xi}_{\bar{G}}\right] = \xi_{[G]}.$$



Benefits: Well-defined Gradient

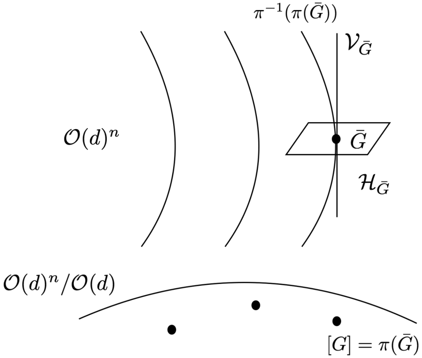
$$\begin{aligned} \mathrm{D}\, \bar{f}(\bar{G}) \left[\bar{\xi}_{\bar{G}}\right] &= \mathrm{D}\, f(\pi(\bar{G})) \left[\mathrm{D}\, \pi(\bar{G}) \left[\bar{\xi}_{\bar{G}}\right]\right] \\ &= \mathrm{D}\, f([G]) \left[\xi_{[G]}\right] \end{aligned}$$

$$\Rightarrow \overline{\operatorname{grad} f([G])}_{\bar{G}} = \operatorname{Proj}_{\mathcal{H}_{\bar{G}}}(\operatorname{grad} \bar{f}(\bar{G}))$$

Explicit Form of Horizontal Space

Proposition

- $\mathcal{V}_{\bar{G}} = \{\bar{G}E : E \in \mathrm{Skew}(d)\}$
- $\mathcal{H}_{\bar{G}} = \{(\bar{G}_1 E_1, \ldots, \bar{G}_n E_n),\ E_i \in \mathrm{Skew}(d) \text{ and } \sum_{i=1}^{n} E_i = 0\}$
- $\mathrm{Proj}_{\mathcal{H}_{\bar{G}}} = I_{nd} - \frac{1}{n}\bar{G}\bar{G}^\top$



Quotient Riemannian Gradient and Hessian

Proposition

Let $[G] \in \mathcal{Q}$ *and* $\bar{G} \in \pi^{-1}([G])$. *Then the unique horizontal lift of*

- *Riemannian gradient of* f *at* $\bar{G} \in \mathcal{O}(d)^n$ *is*

$$\boxed{\overline{\operatorname{grad} f([G])}_{\bar{G}} = \operatorname{grad} \bar{f}(\bar{G})} = -2S(\bar{G})\bar{G}.$$

- *Riemannian Hessian of* f *with direction* $H_{[G]}$ *at* $\bar{G} \in \mathcal{O}(d)^n$ *is*

$$\overline{\operatorname{Hess} f([G])\left[H_{[G]}\right]}_{\bar{G}} = \left(I_{nd} - \tfrac{1}{n}\bar{G}\bar{G}^\top\right)\left(\operatorname{Proj}_{\mathrm{T}_{\bar{G}}\mathcal{O}(d)^n}(-2S(\bar{G})\bar{H}_{\bar{G}})\right).$$

Here, $S(X) := \operatorname{symblockdiag}(CXX^\top) - C \in \mathbb{R}^{nd\times nd}$.

Quotient Riemannian gradient = Riemannian gradient:

$\bar{f}$ is **invariant** on equivalence class $\bar{G} \in \pi^{-1}([G])$

$\Rightarrow \mathrm{D}\,\bar{f}(\bar{G})\bar{\xi}_{\bar{G}} = \langle \operatorname{grad} \bar{f}(\bar{G}), \bar{\xi}_{\bar{G}}\rangle_{\bar{G}} = 0, \ \forall \bar{\xi}_{\bar{G}} \in \mathcal{V}_{\bar{G}}$

$\Rightarrow \operatorname{grad} \bar{f}(\bar{G}) \in (\mathcal{V}_{\bar{G}})^\perp = \mathcal{H}_{\bar{G}}$ is **horizontal lift** of $\operatorname{grad} f([G])$ at $\bar{G}$

Landscape on Quotient Manifold

Assumption: $\|\Delta\| \lesssim \frac{n^{3/4}}{\sqrt{d}}$, $\|\Delta G^\star\|_\infty \lesssim n \Rightarrow \|\Delta\| \lesssim \frac{n}{\sqrt{d}}$ (Leave-one-out)

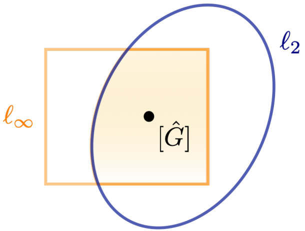
Theorem (Strong Concavity around Maximizers)

Suppose that

- $\mathrm{d}_F([G],[\hat{G}]) \lesssim \min\left\{\sqrt{n}, \frac{n}{\|\Delta\|}\right\}$, $\|G\hat{g} - \hat{G}\|_\infty \leqslant \frac{1}{4}$.

Then for all $H_{[G]} \in \mathrm{T}_{[G]}\,\mathcal{Q} \setminus \{0_{[G]}\}$,

$$-\left\langle \mathrm{Hess}\, f([G])[H_{[G]}], H_{[G]}\right\rangle \geqslant \frac{n}{5} \cdot \langle H_{[G]}, H_{[G]}\rangle > 0.$$



(Quotient) Riemannian Local Error Bound

Theorem ((Quotient) Riemannian Local Error Bound)

Suppose that

- $\mathrm{d}_F([G],[\hat{G}]) \lesssim \min\left\{\sqrt{n}, \frac{n}{\|\Delta\|}\right\}$, $\|G\hat{g} - \hat{G}\|_\infty \leqslant \frac{1}{4}$.

Then it follows that

$$\mathrm{d}_F([G],[\hat{G}]) \leqslant \mathrm{d}^{\mathcal{Q}}([G],[\hat{G}]) \leqslant \frac{10}{n} \cdot \|\operatorname{grad} f([G])\|_{[G]} \leqslant \frac{10}{n} \cdot \|\operatorname{grad} \bar{f}(\bar{G})\|_F.$$

- **FOCPs** $\Rightarrow$ **global maximizer** of (QP-S) with **quantitative** result.
- Theoretical motivation for using (Q)RGM to solve (QP-S)/(QP-O).

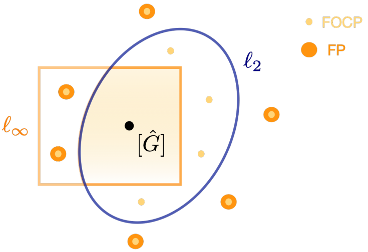
Comparison with Error Bound of GPM

Lemma (Error Bound of GPM [Zhu et al., 2021, Theorem 4.3])

Suppose that

- $\mathrm{d}_F([G],[G^\star]) \lesssim \sqrt{n}$ *and* $\alpha \lesssim n$.

Then it follows that $\mathrm{d}_F([G],[\hat{G}]) \leqslant 10d\|\tilde{C}\| \cdot \|G - T_\alpha(G)\|_F$.



- EB of GPM/Riemannian gradient:

$$\mathrm{d}_F([G],[\hat{G}]) = \mathcal{O}(\sqrt{n})$$
$$+\,\mathrm{d}_\infty([G],[\hat{G}]) = \mathcal{O}(1)$$

- ([Zhu et al., 2021])

Fixed points of GPM (FPs) $\subseteq$ FOCPs

Example: Necessity of ℓ_∞ Constraint

Example ($\mathrm{d}_\infty([G],[\hat{G}]) = \mathcal{O}(1)$ is Necessary)

Let $d = 2$ and $\Delta = 0$ (implying $G^\star = \hat{G}$). Let $G \in \mathcal{O}(d)^n$ satisfy

$$G_i = \begin{cases} -\hat{G}_i, & \text{if } i = 1, \\ \hat{G}_i, & \text{otherwise.} \end{cases}$$

- $\overline{\mathrm{grad}\, f([G])}_G = \mathrm{grad}\, \bar{f}(G) = S(G)G = 0$
- $\mathrm{d}_F([G],[\hat{G}]) = \sqrt{2}$

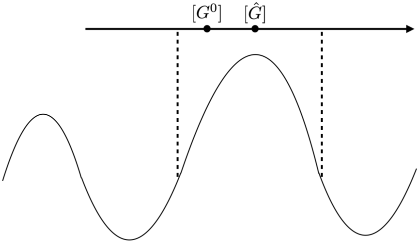
$\Rightarrow$ G is only a **FOCP**: global optimum $\hat{G}$ is unique (up to rotation)

Convergence of (Q)RGM: Initialization

Proposition (Spectral Initialization Estimation Error)

The spectral estimator $G^0 = \mathrm{Proj}_{\mathcal{G}^n}(\Phi) \in \mathcal{G}^n$ *(Φ is top d eigenvectors of C with $\Phi^\top \Phi = nI_d$) satisfies*

$$\mathrm{d}_F([G^0],[G^\star]) \lesssim \frac{\sqrt{d}\|\Delta\|}{\sqrt{n}} \quad \textit{and} \quad \|G^0 g_0^\star - G^\star\|_\infty \lesssim \frac{\|\Delta G^\star\|_\infty}{n} + \frac{\sqrt{d}\|\Delta\|}{n}.$$



✓ $\|\Delta\| \lesssim \frac{n^{3/4}}{d^{1/2}}$, $\|\Delta G^\star\|_\infty \lesssim n \Rightarrow \mathrm{d}_F([G^0],[G^\star]) \lesssim n^{1/4}$, $\|G^0 g_0^\star - G^\star\|_\infty \lesssim 1$

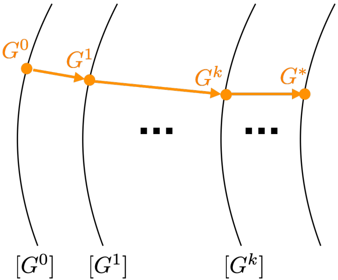
Convergence of (Q)RGM

Theorem (Sequential Linear Convergence)

The sequence $\{G^k\}_{k \geqslant 0}$ generated by (Q)RGM with spectral initialization converges to some $G^* \in [\hat{G}]$. Moreover, with $\lambda \in (0,1)$,

$$f([\hat{G}]) - f([G^{k+1}]) \leqslant \lambda \cdot (f([\hat{G}]) - f([G^k])),$$

$$\mathrm{d}_F([G^k], [\hat{G}]) \leqslant \|G^k - G^*\|_F \leqslant (f([\hat{G}]) - f([G^0]))^{\frac{1}{2}} \cdot \lambda^{\frac{k}{2}}.$$



Conclusion & Discussion

Conclusion:

- **(Landscape) Quotient geometric view** of least squares formulation of rotation/orthogonal group synchronization.
- **(Algorithm)** (Q)RGM: **simple and provably efficient** algorithm for rotation group synchronization.
- **(Tightness)** Improved **deterministic** estimation result $\Rightarrow$ guarantees for **various existing approaches** for rotation group sychronization.

Conclusion & Discussion

Conclusion:

- **(Landscape) Quotient geometric view** of least squares formulation of rotation/orthogonal group synchronization.
- **(Algorithm)** (Q)RGM: **simple and provably efficient** algorithm for rotation group synchronization.
- **(Tightness)** Improved **deterministic** estimation result $\Rightarrow$ guarantees for **various existing approaches** for rotation group sychronization.

? Other Riemannian algorithms: second-order/trust region method
- Iterative direction is different on **original and quotient** manifold.

? Landscape analysis from the quotient view for other problems.

Thank you!

References I

- Arie-Nachimson, M., Kovalsky, S. Z., Kemelmacher-Shlizerman, I., Singer, A., and Basri, R. (2012). Global motion estimation from point matches. In *2012 Second international conference on 3D imaging, modeling, processing, visualization & transmission*, pages 81–88. IEEE.
- Bandeira, A. S., Boumal, N., and Singer, A. (2017). Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, 163(1):145–167.
- Boumal, N. (2016). Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377.
- Boumal, N., Voroninski, V., and Bandeira, A. (2016). The non-convex burer-monteiro approach works on smooth semidefinite programs. *Advances in Neural Information Processing Systems*, 29.
- Khoo, Y. and Kapoor, A. (2016). Non-iterative rigid 2d/3d point-set registration using semidefinite programming. *IEEE Transactions on Image Processing*, 25(7):2956–2970.
- Ling, S. (2022a). Improved performance guarantees for orthogonal group synchronization via generalized power method. *SIAM Journal on Optimization*, 32(2):1018–1048.
- Ling, S. (2022b). Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *Applied and Computational Harmonic Analysis*.
- Ling, S. (2022c). Solving orthogonal group synchronization via convex and low-rank optimization: Tightness and landscape analysis. *Mathematical Programming*, pages 1–40.
- Liu, H., Yue, M.-C., and So, A. M.-C. (2017). On the estimation performance and convergence rate of the generalized power method for phase synchronization. *SIAM Journal on Optimization*, 27(4):2426–2446.
- Liu, H., Yue, M.-C., and So, A. M.-C. (2020). A unified approach to synchronization problems over subgroups of the orthogonal group. *arXiv preprint arXiv:2009.07514*.

References II

- Rosen, D. M., Carlone, L., Bandeira, A. S., and Leonard, J. J. (2019). SE-Sync: A certifiably correct algorithm for synchronization over the special euclidean group. *The International Journal of Robotics Research*, 38(2-3):95–125.
- Singer, A. (2011). Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36.
- Singer, A. (2018). Mathematics for cryo-electron microscopy. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3995–4014. World Scientific.
- Singer, A. and Shkolnisky, Y. (2011). Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. *SIAM Journal on Imaging Sciences*, 4(2):543–572.
- Won, J.-H., Zhang, T., and Zhou, H. (2022). Orthogonal trace-sum maximization: Tightness of the semidefinite relaxation and guarantee of locally optimal solutions. *SIAM Journal on Optimization*, 32(3):2180–2207.
- Zhong, Y. and Boumal, N. (2018). Near-optimal bounds for phase synchronization. *SIAM Journal on Optimization*, 28(2):989–1016.
- Zhu, L., Wang, J., and So, A. M.-C. (2021). Orthogonal group synchronization with incomplete measurements: Error bounds and linear convergence of the generalized power method. *arXiv preprint arXiv:2112.06556*.