Error Bounds for Orthogonal Group Synchronization & Convergence Analysis of the Generalized Power Method

Linglingzhi Zhu

Department of Systems Engineering & Engineering Management
The Chinese University of Hong Kong

December 5, 2021

Orthogonal Group Synchronization

Let the orthogonal group elements (**Ground-Truth**)

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

be the target to be estimated, where

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

Recover $G^\star$ from $\{C_{ij} : (i,j) \in E\}$, where

- $E \subseteq \{(i,j) : 1 \leq i < j \leq n\}$
- C_{ij} is the noisy measurement of the relative transform $G_i^\star G_j^{\star\top}$

Example of Applications

- Graph Realization
  - Sensor Network Localization [Cucuringu et al., 2012a]
  - Structural Biology [Cucuringu et al., 2012b]
- Computer Vision
  - 2D/3D Point Set Registration [Khoo et al., 2016]
  - Multiview Structure from Motion [Arie-Nachimson et al., 2012],
  - Common Lines in Cryo-Electron Microscopy [Singer et al., 2011],
- Robotics
  - Simultaneous Localization and Mapping (SLAM) [Rosen et al., 2019]

Nonconvex Least Squares Formulation

From the maximum likelihood estimator we formulate the problem:

$$\min_{G_1,\dots,G_n \in \mathcal{O}(d)} \sum_{(i,j)\in E} \|G_i G_j^\top - C_{ij}\|_F^2 \tag{MLE}$$

Since $G_1, \dots, G_n \in \mathcal{O}(d)$, Problem (MLE) is equivalent to

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

where $Q = (Q_1, \dots, Q_n) \in \mathcal{O}(d)^n \subseteq \mathbb{R}^{nd\times d}$ and $C \in \mathbb{R}^{nd\times nd}$.

Problem (QP) is nonconvex in general with structure:

- Quadratic objective function over orthogonal group constraint $\mathcal{O}(d)$
- The measurement matrix C usually owns a generative model

Approaches for Solving (QP)

- **Semidefinite Relaxation** [Ling, 2020a, Won et al., 2021]

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

  - strong recovery guarantees (under generative models) but not scale well with problem size

- **Burer-Monteiro** [Boumal, 2016, Ling, 2020a]

$$\max_{X \in \mathbb{R}^{nd \times p}} \operatorname{tr}(CXX^\top)$$

  where

$$p > d, \quad X := [X_1; \cdots; X_n] \in \mathbb{R}^{nd \times p}, \quad X_i X_i^\top = I_d$$

  - usually weak recovery guarantees

- **Spectral Relaxation** [Ling, 2020b]

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

  - simple but unsatisfactory estimation performance

Nonconvex Approach with Generative Model

Recall that

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

In general, Problem (QP) is **NP-hard** as a quadratic program problem with quadratic constraints (QPQC) (reduced to Max-Cut problem when $d = 1$).

Generative Model:

- The measurement matrix is the additive noise model

  $$C_{ij} = G_i^\star G_j^{\star\top} + \Delta_{ij}, \quad (i,j) \in E,$$

  where the measurement set E and noise matrices $\{\Delta_{ij} : (i,j) \in E\}$ possess certain statistical properties.

Generalized Power Method

The Generalized Power Method (GPM) is an efficient algorithm through the nonconvex approach [Journe et al., 2010, Boumal, 2016]. For

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

the method goes as follows:

Algorithm 1 GPM for Solving Problem (QP)

1: Input: the matrix C, stepsize $\alpha \geq 0$, initial point $G^0 \in \mathcal{O}(d)^n$.
2: **for** $k = 0, 1, \ldots$ **do**
3: $\quad G^{k+1} \in \mathrm{Proj}_{\mathcal{O}(d)^n}(\tilde{C} G^k)$, where $\tilde{C} := C + \alpha I_{nd}$.
4: **end for**

- The projection $\mathrm{Proj}_{\mathcal{O}(d)^n}(G) = (\mathrm{Proj}_{\mathcal{O}(d)}(G_1), \ldots, \mathrm{Proj}_{\mathcal{O}(d)}(G_n))$ has a closed-form solution by SVD.
- The GPM is actually the projected gradient method

$$G^{k+1} \in \mathrm{Proj}_{\mathcal{O}(d)^n}\left(G^k + \alpha^{-1} \nabla f(G^k)\right).$$

Existing Results

Theorem (Liu et al., 2020)

Let $\{G^k\}_{k\geq 0}$ *be the sequence generated by the GPM. Suppose that*

- *(Sampling) The measurement set* E *is sufficiently dense*
- *(Noise)* $\|\Delta\|_2$ *and* $\|\Delta G^\star\|_F$ *are sufficiently small*
- *(Initialization)* $\mathrm{d}(G^0, G^\star) := \min_{Q\in\mathcal{O}(d)} \|G^0 - G^\star Q\|_F$ *is sufficiently small*

Then for any $k \geq 1$*, there exists* $0 < \lambda < 1$ *and* $c > 0$ *such that*

$$\mathrm{d}(G^k, G^\star) \leq \lambda^{k+1}\mathrm{d}(G^0, G^\star) + c\,\|\Delta G^\star\|_F\,.$$

Existing Results

Theorem (Liu et al., 2020)

Let $\{G^k\}_{k\geq 0}$ *be the sequence generated by the GPM. Suppose that*

- *(Sampling) The measurement set* E *is sufficiently dense*
- *(Noise)* $\|\Delta\|_2$ *and* $\|\Delta G^\star\|_F$ *are sufficiently small*
- *(Initialization)* $\mathrm{d}(G^0, G^\star) := \min_{Q\in\mathcal{O}(d)} \|G^0 - G^\star Q\|_F$ *is sufficiently small*

Then for any $k \geq 1$*, there exists* $0 < \lambda < 1$ *and* $c > 0$ *such that*

$$\mathrm{d}(G^k, G^\star) \leq \lambda^{k+1}\mathrm{d}(G^0, G^\star) + c\,\|\Delta G^\star\|_F\,.$$

Question: which point does GPM converge to and at what rate?

Optimality Conditions

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

- **first-order critical point (FOCP)**: $S(G)G = 0$
- **second-order critical point (SOCP)**:

$$S(G)G = 0, \quad \langle H, S(G)H \rangle \geq 0$$

for all $H \in \left\{[X_1; \ldots; X_n] \in \mathbb{R}^{nd \times d} \mid X_i = E_i G_i,\ E_i = -E_i^\top,\ i \in [n]\right\}$.

Denote $S(G) := \mathrm{symblockdiag}(CGG^\top) - C$, where the linear operator symblockdiag: $\mathbb{R}^{nd \times nd} \to \mathbb{S}^{nd}$ is defined as

$$\mathrm{symblockdiag}(X)_{ij} = \begin{cases} \dfrac{X_{ii} + X_{ii}^\top}{2}, & \text{if} \quad i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Optimality Conditions

Let $\alpha \geq 0$. Denote the operator of the GPM by $\mathcal{T}_\alpha : \mathcal{O}(d)^n \rightrightarrows \mathcal{O}(d)^n$ for each $G \in \mathcal{O}(d)^n$ as follows:

$$\mathcal{T}_\alpha(G) := \mathrm{Proj}_{\mathcal{O}(d)^n}(\tilde{C}G), \text{ where } \tilde{C} := C + \alpha I_{nd}$$

We derive the following relationship **without any generative model**

- first-order critical points (FOCPs)
- second-order critical points (SOCPs)
- global maximizers (GMs)
- fixed points of $\mathcal{T}_\alpha(G)$ (i.e., $G \in \mathcal{T}_\alpha(G)$) (FPs)

$$\begin{tikzcd}
\boxed{\text{GMs of (QP)}} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \boxed{\text{SOCPs of (QP)}} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \boxed{\text{FOCPs of (QP)}} \\
\boxed{\bigcap_{\alpha \geq 0} \{\text{FPs of } \mathcal{T}_\alpha\}} \arrow[r, Rightarrow] & \boxed{\bigcap_{\alpha \geq \alpha'} \{\text{FPs of } \mathcal{T}_\alpha\}} \arrow[r, Rightarrow] & \boxed{\bigcup_{\alpha \geq 0} \{\text{FPs of } \mathcal{T}_\alpha\}} \arrow[u, Rightarrow]
\end{tikzcd}$$

Optimality Conditions

Let $\alpha \geq 0$. Denote the operator of the GPM by $\mathcal{T}_\alpha : \mathcal{O}(d)^n \rightrightarrows \mathcal{O}(d)^n$ for each $G \in \mathcal{O}(d)^n$ as follows:

$$\mathcal{T}_\alpha(G) := \mathrm{Proj}_{\mathcal{O}(d)^n}(\tilde{C}G), \text{ where } \tilde{C} := C + \alpha I_{nd}$$

We derive the following relationship **without any generative model**

- first-order critical points (FOCPs)
- second-order critical points (SOCPs)
- global maximizers (GMs)
- fixed points of $\mathcal{T}_\alpha(G)$ (i.e., $G \in \mathcal{T}_\alpha(G)$) (FPs)

$$\begin{tikzcd}
\boxed{\text{GMs of (QP)}} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \boxed{\text{SOCPs of (QP)}} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \boxed{\text{FOCPs of (QP)}} \\
\boxed{\bigcap_{\alpha \geq 0} \{\text{FPs of } \mathcal{T}_\alpha\}} \arrow[r, Rightarrow] & \boxed{\bigcap_{\alpha \geq \alpha'} \{\text{FPs of } \mathcal{T}_\alpha\}} \arrow[r, Rightarrow] & \boxed{\bigcup_{\alpha \geq 0} \{\text{FPs of } \mathcal{T}_\alpha\}} \arrow[u, Rightarrow]
\end{tikzcd}$$

Is the FP necessarily a GM? If so, any quantified result?
Relation between a GM and Ground-Truth $G^\star$?

Generative Model Setting

The noisy incomplete pairwise measurements

$$C_{ij} = \begin{cases} W_{ij} \cdot (G_i^\star G_j^{\star\top} + \Delta_{ij}), & \text{if } i \neq j, \\ W_{ii} \cdot I_d, & \text{otherwise,} \end{cases}$$

where

- $W \in \mathbb{R}^{n \times n}$ is the symmetric adjacency matrix of the measurement graph $\mathcal{G}([n], \Omega)$ with an edge set Ω
- $W_{ii} = \mu > 0$ for all $i \in [n]$

By defining $A := W \otimes (1_d 1_d^\top)$, we write

$$C = A \circ (G^\star G^{\star\top} + \Delta),$$

where "$\otimes$" is the Kronecker product and "$\circ$" is the Hadamard product.

Local Error Bound Property

Proposition (Distance between $\hat{G}$ and $G^\star$)

The global maximizer $\hat{G} \in \mathcal{O}(d)^n$ satisfies

$$\mathrm{d}(\hat{G}, G^\star) \leq 4\mu^{-1}\sqrt{n^{-1}d}\left(\left\|W - \mu \cdot 1_n 1_n^\top\right\| + \|A \circ \Delta\|\right)$$

Theorem (Local Error Bound)

Suppose that

- *(Sampling & Noise)*
  $\|W - \mu \cdot 1_n 1_n^\top\| + \|A \circ \Delta\| \leq \frac{n^{3/4}\mu}{40 d^{1/2}}, \quad \|(A \circ \Delta)G^\star\|_\infty \leq \frac{n\mu}{10},$
  $\max_{i\in[n]} \left\| \left((A - \mu \cdot 1_{nd}1_{nd}^\top) \circ G^\star G^{\star\top}\right)_i^\top \hat{G} \right\| \leq \frac{n\mu}{10}$
- $\alpha \leq \frac{n\mu}{20\sqrt{2}}$

Then for any $G \in \mathcal{O}(d)^n$ satisfying $\mathrm{d}(G, G^\star) \leq \frac{\sqrt{n}}{5}$ and any $\hat{G} \in \mathcal{O}(d)^n$,

$$\frac{n\mu}{10}\mathrm{d}(G, \hat{G}) \leq \rho_\alpha(G) \quad (\textit{residual function})$$

Residual Function

Recall that $\alpha \geq 0$ and $\tilde{C} = C + \alpha I_{nd}$. Let $D_\alpha : \mathcal{O}(d)^n \to \mathbb{S}^{nd}$ be defined as

$$D_\alpha(G) := \mathrm{Diag}\left(\left[U_{\tilde{C}_1^\top G}\Sigma_{\tilde{C}_1^\top G}U_{\tilde{C}_1^\top G}^\top; \ldots; U_{\tilde{C}_n^\top G}\Sigma_{\tilde{C}_n^\top G}U_{\tilde{C}_n^\top G}^\top\right]\right) - \tilde{C},$$

where $U_{\tilde{C}_i^\top G} \in \Xi(\tilde{C}_i^\top G)$ and

$$\Xi(Z) := \left\{U \in \mathcal{O}(m) \mid Z = U\Sigma(Z)V^\top \text{ for some } V \in \mathcal{O}(n)\right\}.$$

Then we define $\rho_\alpha : \mathcal{O}(d)^n \to \mathbb{R}_+$ as follows:

$$\rho_\alpha(G) := \|D_\alpha(G)G\|_F \tag{RES}$$

The operator D_α is a single-valued rather than set-valued mapping, since for any $U_X \in \Xi(X)$ there exists a unique positive semidefinite matrix

$$(XX^\top)^{1/2} = U_X\Sigma_X U_X^\top.$$

Relation to Fixed Points of the GPM

Recall the local error bound result:

$$\frac{n\mu}{10}\mathrm{d}(G, \hat{G}) \leq \rho_\alpha(G).$$

For any $G \in \mathcal{O}(d)^n$ and $T_\alpha(G) \in \mathcal{T}_\alpha(G)$

$$D_\alpha(G)G = \mathrm{Diag}(\tilde{C}G) \cdot \mathrm{Diag}\left(\left[(T_\alpha(G) - G)_1^\top; \ldots; (T_\alpha(G) - G)_n^\top\right]\right) G$$

$$\implies \quad \rho_\alpha(G) = \|D_\alpha(G)G\|_F \leq nd\|\tilde{C}\|\|G - T_\alpha(G)\|_F$$

- Answer the question that the fixed point (FP) of the GPM are the global maximizer (GM) of (QP) with the local quantitative result.
- Theoritical motivation for using the projected gradient method (i.e., the GPM) to solve Problem (QP).

GPM with Spectral Initialization

Algorithm 2 GPM with Spectral Initialization (GPM-Spec)

1: Input: the matrix C, stepsize $\alpha \geq 0$.
2: Compute the top d eigenvectors Φ of C with $\Phi^\top \Phi = nI_d$.
3: Compute $G^0 \in \mathrm{Proj}_{\mathcal{O}(d)^n}(\Phi)$ and generate $\{G^k\}$ by the GPM.

Proposition (Good Initialization & Stay in Ball)

The spectral estimator $G^0 \in \mathcal{O}(d)^n$ satisfies

$$\mathrm{d}(G^0, G^\star) \leq 8\mu^{-1}\sqrt{n^{-1}d}\left(\left\|W - \mu \cdot 1_n 1_n^\top\right\| + \|A \circ \Delta\|\right).$$

Suppose further that

- *(Sampling & Noise)* $\|W - \mu \cdot 1_n 1_n^\top\| + \|A \circ \Delta\| \leq \frac{n\mu}{60d^{1/2}}$
- *(Stepsize)* $\alpha \leq \frac{n\mu}{30\sqrt{2d}}$

Then $\{G^k\}_{k\geq 0}$ generated by the GPM-Spec satisfies $\mathrm{d}(G^k, G^\star) \leq \frac{\sqrt{n}}{5}$.

Convergence Analysis of GPM-Spec

Theorem (Linear convergence of the GPM-Spec)

Suppose that

- *(Sampling & Noise)*
  $\|W - \mu \cdot 1_n 1_n^\top\| + \|A \circ \Delta\| \leq \frac{n^{3/4}\mu}{60 d^{1/2}}, \quad \|(A \circ \Delta) G^\star\|_\infty \leq \frac{n\mu}{10},$
  $\max_{i \in [n]} \left\| \left((A - \mu \cdot 1_{nd} 1_{nd}^\top) \circ G^\star G^{\star\top} \right)_i^\top \hat{G} \right\| \leq \frac{n\mu}{10}$
- *(Stepsize)* $\|A \circ \Delta\| + \|W - \mu \cdot 1_n 1_n^\top\| < \alpha \leq \frac{n\mu}{30\sqrt{2d}}$

Then, the sequence $\{G^k\}_{k \geq 0}$ *generated by the GPM-Spec satisfies*

$$f(\hat{G}) - f(G^k) \leq (f(\hat{G}) - f(G^0))\lambda^k$$

and

$$\mathrm{d}(G^k, \hat{G}) \leq a \cdot (f(\hat{G}) - f(G^0))^{1/2} \lambda^{k/2},$$

where $a > 0$, $\lambda \in (0, 1)$ *are constants that depend only on* n, d, μ, α.

Erdös-Rényi Graph with Gaussian Noise Setting

Recall the noisy incomplete pairwise measurements

$$C_{ij} = \begin{cases} W_{ij} \cdot (G_i^\star G_j^{\star\top} + \Delta_{ij}), & \text{if } i \neq j, \\ W_{ii} \cdot I_d, & \text{otherwise.} \end{cases}$$

The Erdös-Rényi graph $\mathcal{G}([n], p)$ with Gaussian noise setting satisfying:

- W_{ij} are i.i.d. random variables following the Bernoulli distribution taking 1 with probability p (associated with n), otherwise being 0, and $W_{ji} = W_{ij}$ for each $i < j$
- $W_{ii} = \mu = \frac{\sum_{i<j} W_{ij}}{n(n-1)/2}$ for each $i \in [n]$
- $\Delta = \sigma Z$, where $\sigma > 0$, $Z \in \mathbb{S}^{nd}$ with $Z_{ii} = \mathbf{0}$ for $i \in [n]$ and Z_{ij} are i.i.d. standard Gaussian variables for $i \neq j$

Convergence Analysis under Gaussian Noise

Theorem (Linear convergence of the GPM under Gaussian noise)

Suppose that

- *(Sampling) the Erdös-Rényi graph* $\mathcal{G}([n], p)$ *satisfies* $p \geq \frac{\kappa_0 d}{\sqrt{n}}$
- *(Noise)* $\Delta = \sigma Z$*, where* $0 < \sigma \leq \frac{\kappa_1 n^{1/4} p^{1/2}}{d}$
- *(Stepsize)* $\frac{\kappa_0 n^{3/4} p}{d^{1/2}} \leq \alpha \leq \frac{\kappa_1 np}{d^{1/2}}$

where $\kappa_0, \kappa_1 > 0$ *are constants. Then for sufficiently large* $n \in \mathbb{N}$*, the sequence* $\{G^k\}_{k \geq 0}$ *generated by the GPM-Spec with high probability satisfies*

$$f(\hat{G}) - f(G^k) \leq (f(\hat{G}) - f(G^0))\lambda^k$$

and

$$\mathrm{d}(G^k, \hat{G}) \leq a \cdot (f(\hat{G}) - f(G^0))^{1/2} \lambda^{k/2},$$

where $a > 0$*,* $\lambda \in (0, 1)$ *are constants that depend only on* n*,* d*,* p*,* α*.*

Conclusion & Discussion

- The GPM (with good initialization) is a simple and provable effect algorithm for the orthogonal group synchronization problem, which is nonconvex but owns nice properties.
- The error bound result is motivated by the GPM but it is an algorithm-independent property.
- It will be intersting to investigate synchronization problems of other subgroups of orthogonal group, e.g. $\mathcal{SO}(d)$ as a generalization of the phase synchronization problem $\mathcal{SO}(2)$, where the noncommutative nature brings difficulty.

References

- M. Cucuringu, Y. Lipman, and A. Singer. Sensor network localization by eigenvector synchronization over the Euclidean group. ACM Transactions on Sensor Networks (TOSN), 8(3):1-42, 2012.
- M. Cucuringu, A. Singer, and D. Cowburn. Eigenvector synchronization, graph rigidity and the molecule problem. Information and Inference: A Journal of the IMA, 1(1):21 67, 2012.
- Y. Khoo and A. Kapoor. Non-iterative rigid 2d/3d point-set registration using semidefinite programming. IEEE Trans. Image Process., 25:2956-2970, 2016.
- M. Arie-Nachimson, S. Z. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri. Global motion estimation from point matches, in Proc. 2012 Int. Conf. 3D Imaging, Modeling, Processing, Visualization Transmission.
- A. Singer and Y. Shkolnisky. Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. SIAM J. Imaging Sci., 4:543-572, 2011.
- D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard. SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group. The International Journal of Robotics Research, 38(2-3):95-125, 2019.
- S. Ling. Solving orthogonal group synchronization via convex and low-rank optimization: tightness and landscape analysis. arXiv preprint arXiv:2006.00902, 2020.
- N. Boumal. A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints. arXiv preprint arXiv:1506.00575, 2016.
- S. Ling. Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. arXiv preprint arXiv:2008.05341, 2020.
- J. Won, T. Zhang, and H. Zhou. Orthogonal trace-sum maximization: tightness of the semidefinite relaxation and guarantee of locally optimal solutions. arXiv preprint arXiv:2110.05701, 2021.
- M. Journe, Y. Nesterov, P. Richtrik, and R. Sepulchre. Generalized power method for sparse principal component analysis. J. Mach. Learn. Res., 11:517553, 2010.
- H. Liu, M.-C. Yue, and A. M.-C. So. A unified approach to synchronization problems over subgroups of the orthogonal group. arXiv preprint arXiv:2009.07514, 2020.

Thank you!