

# Nonconvex Robust Synchronization of Rotations

**Huikang Liu\***

*Imperial College London*

HUIKANG.LIU@IMPERIAL.AC.UK

**Zengde Deng\***

*Cainiao AI*

DENGZENGDE@GMAIL.COM

**Xiao Li\***

*The Chinese University of Hong Kong, Shenzhen*

LIXIAO@CUHK.EDU.CN

**Shixiang Chen\***

*Texas A&M University*

SXCHEN@TAMU.EDU

**Anthony Man-Cho So**

*The Chinese University of Hong Kong*

MANCHOSO@SE.CUHK.EDU.HK

## Abstract

Synchronization of rotations finds many applications in computer vision, sensor network localization, and computational imaging. In this paper, we consider the problem of learning a set of rotation matrices on the manifold of the special orthogonal group from highly *incomplete* and *corrupted* relative rotation observations and solve a nonsmooth nonconvex optimization formulation of the problem. Under mild conditions, we provide a *global identifiability* guarantee, which asserts that solving our optimization formulation to global optimality will yield the underlying rotation matrices. Then, we design an efficient manifold subgradient method for addressing the optimization formulation, which can be shown to converge *linearly* to a globally optimal solution when properly initialized. We also present an initialization method and its theoretical guarantee. Finally, we conduct experiments to show the superiority of the nonconvex formulation over convex ones and demonstrate the efficacy of our algorithm.<sup>1</sup>

## 1. Introduction

The importance of synchronization of rotations is illustrated by applications arising from many engineering fields. For example, synchronization problems over  $\text{SO}(2)$  (also known as ‘phase synchronization’ [2]) has been applied to sensor network localization [5, 18], signal recovery from phaseless observations [1]. Synchronization on  $\text{SO}(3)$  is used for structuring from motion in computer vision [7, 12, 14].

Denote the special orthogonal group as  $\text{SO}(d) := \{\mathbf{R} \in \mathbb{R}^{d \times d} : \mathbf{R}^T \mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = 1\}$ . Robust synchronization of rotations amounts to find a set of rotation matrices in  $\text{SO}(d)$ ,

$$\mathbf{X}_1^*, \dots, \mathbf{X}_i^*, \dots, \mathbf{X}_n^* \in \text{SO}(d), \quad (1)$$

from *corrupted* measurements of relative rotations

$$\mathbf{Y}_{ij} = \begin{cases} \mathbf{X}_i^{*T} \mathbf{X}_j^*, & \text{with probability } p, \\ \mathbf{O}_{ij}, & \text{with probability } 1 - p, \end{cases} \quad (2)$$

---

1. \* indicates equal contribution

---

**Algorithm 1** ManSGM: Manifold Subgradient Method
 

---

- 1: **Input:** Initialization  $\mathbf{X}^0 = (\mathbf{X}_1^0, \dots, \mathbf{X}_n^0)$  and stepsize parameters  $\mu_0 > 0$  and  $\rho \in (0, 1)$ .
  - 2: **for**  $k = 0, 1, \dots$  **do**
  - 3:   Set the stepsize  $\mu_k = \mu_0 \cdot \rho^k$ .
  - 4:   Update  $\mathbf{X}^{k+1} = (\mathbf{X}_1^{k+1}, \dots, \mathbf{X}_n^{k+1})$  as
 
$$\mathbf{X}_i^{k+1} = \text{Retr}_{\mathbf{X}_i^k} \left( -\mu_k \tilde{\nabla}_{\mathcal{R}} f(\mathbf{X}_i^k) \right), \quad \forall 1 \leq i \leq n.$$
  - 5: **end for**
- 

where  $\mathbf{O}_{ij} \in \text{SO}(d)$  is outlying observation (outliers). Moreover, we assume the data is *incomplete*. Consider the synchronization graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with vertex set  $\mathcal{V} = \{1, \dots, n\}$  drawn from Erdős-Rényi model  $\mathcal{G}(n, q)$ , meaning that  $(i, j) \in \mathcal{E}$  with probability  $q$ , *the observation  $\mathbf{Y}_{ij}$  is available if and only if  $(i, j) \in \mathcal{E}$ .*

A large number of works focus on smooth least-squares optimization, which is, however, known to be sensitive to outliers. To robustify the solution in the presence of outliers, the works [15, 16] study semidefinite relaxations (SDR) method for solving nonsmooth least absolute deviation optimization. In addition to optimization methods, the work [9] considers a statistical method called message passing and provides theoretical guarantees.

**Our main results** Instead of using convex relaxation approach, we directly solve a *nonsmooth nonconvex* least absolute deviation formulation. In particular, we show that the aforementioned formulation has the *global identifiability* property and the so-called *sharpness property*. Consequently, a properly initialized Riemannian subgradient method can be shown to converge *linearly* to the underlying rotations.

## 2. Problem Formulation and Algorithm

In the sequel,  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n) \in \text{SO}(d)^n$  represents the Cartesian product of all the variables  $\mathbf{X}_i \in \text{SO}(d), 1 \leq i \leq n$  and the same applies to the ground truth rotations  $\mathbf{X}^* = (\mathbf{X}_1^*, \dots, \mathbf{X}_n^*)$ .

Unlike least-squares loss that is sensitive to outliers, the least absolute deviation is much more robust against outlying observations; see, e.g., [4, 6, 10]. Thus, to robustly synchronize the underlying rotations  $\mathbf{X}^* \in \text{SO}(d)^n$ , we consider the following nonsmooth nonconvex optimization problem.

$$\underset{\mathbf{X} \in \mathbb{R}^{d \times nd}}{\text{minimize}} \quad f(\mathbf{X}) := \sum_{(i,j) \in \mathcal{E}} \|\mathbf{X}_i^T \mathbf{X}_j - \mathbf{Y}_{ij}\|_F, \quad \text{s.t.} \quad \mathbf{X}_i \in \text{SO}(d), 1 \leq i \leq n, \quad (3)$$

which is also mentioned in [15], but they did not study this formulation directly.

To tackle our optimization problem, we propose a Manifold algorithm that utilizes the subgradient information of the objective function in (3). We first present a concise preliminary for manifold optimization. We impose the Euclidean inner product  $\langle \mathbf{R}_1, \mathbf{R}_2 \rangle = \text{trace}(\mathbf{R}_1^T \mathbf{R}_2)$  as the inherent Riemannian metric. Consequently, the tangent space to  $\text{SO}(d)$  at  $\mathbf{R} \in \text{SO}(d)$  is given by  $\text{T}_{\mathbf{R}} \text{SO} := \{\mathbf{R}\mathbf{S} : \mathbf{S} \in \mathbb{R}^{d \times d}, \mathbf{S} + \mathbf{S}^T = 0\}$ . The Riemannian subgradient  $\tilde{\nabla}_{\mathcal{R}} f(\mathbf{X}) = (\tilde{\nabla}_{\mathcal{R}} f(\mathbf{X}_1), \dots, \tilde{\nabla}_{\mathcal{R}} f(\mathbf{X}_n))$  can be computed as  $\tilde{\nabla}_{\mathcal{R}} f(\mathbf{X}_i) = \mathcal{P}_{\text{T}_{\mathbf{X}_i} \text{SO}}(\tilde{\nabla} f(\mathbf{X}_i)), 1 \leq i \leq n$ , where the projection  $\mathcal{P}_{\text{T}_{\mathbf{X}_i} \text{SO}}(\mathbf{B}) = \mathbf{X}_i(\mathbf{X}_i^T \mathbf{B} - \mathbf{B}^T \mathbf{X}_i)/2$  for any  $\mathbf{B} \in \mathbb{R}^{d \times d}$  and  $\tilde{\nabla} f(\mathbf{X}_i)$  is the Euclidean subgradient of  $f$  with respect to  $\mathbf{X}_i$  [17].

We employ a retraction operator to address the feasibility issue and our algorithm can be written as an iterative procedure,

$$\mathbf{X}_i^{k+1} = \text{Retr}_{\mathbf{X}_i^k}(\boldsymbol{\xi}_i^k), \quad 1 \leq i \leq n, \quad (4)$$

with  $\boldsymbol{\xi}_i^k = -\mu_k \tilde{\nabla}_{\mathcal{R}} f(\mathbf{X}_i)$  is the search direction. Note that the retraction can be implemented efficiently through  $\text{Retr}_{\mathbf{X}_i}(\boldsymbol{\xi}_i) = \text{qf}(\mathbf{X}_i + \boldsymbol{\xi}_i)$ , where  $\text{qf}(\mathbf{B})$  is the Q-factor in the thin QR factorization of  $\mathbf{B}$  (the diagonal of the R-factor is restricted to be positive) [3]. We display the pseudo code in Algorithm 1 and name our algorithm **ManSGM**.

### 3. Main Results

In this section, we present our main results stating that the optimization problem (3) *precisely locates* the true rotations  $\mathbf{X}^*$  (up to a global rotation) at its global minima and meanwhile our algorithm ManSGM is able to find a global minimum to problem (3) at a *linear rate*, under some mild conditions on the fraction of outliers and observation ratio.

The following theorem states the global identifiability and sharpness property of the optimization problem (3).

**Theorem 1 (global identifiability and sharpness property)** *Consider the measurement model (2), where the outliers  $\mathbf{O}_{ij}$  are sampled randomly on  $\text{SO}(d)$  following a uniform distribution for all  $(i, j) \in \Omega^c$ . Suppose*

$$p > \frac{2}{2 + \sqrt{2}} \quad \text{and} \quad q \geq \frac{24}{(2 + \sqrt{2})p - 2} \sqrt{\log n/n}.$$

*Then with probability at least  $1 - (4d + 3)/n$ , we have the following sharpness property,*

$$f(\mathbf{X}) - f(\mathbf{X}^*) \geq \alpha \text{dist}(\mathbf{X}, \mathbf{X}^*), \quad \forall \mathbf{X} \in \text{SO}(d)^n. \quad (5)$$

*where  $\alpha = \frac{((2 + \sqrt{2})p - 2) n q}{96} > 0$ . Consequently, the underlying rotation  $\mathbf{X}^*$  is precisely the ‘unique’ global minimum to problem (3) (up to a global rotation).*

**Remark 1** The model on the outliers— $\mathbf{O}_{ij}$  for all  $(i, j) \in \Omega^c$  is randomly and uniformly distributed on the manifold  $\text{SO}(d)$ —has been widely adopted for synchronization of rotations problems [15, 16] and robust PCA and robust subspace learning problems [8, 13, 19], etc.

**Remark 2** Sharpness property is a kind of regularity of optimization problems and plays a key role to the convergence analysis of our ManSGM algorithm. Currently, only a few applications are known to give rise to sharpness property, and most of them are *unconstrained* optimization problems. By contrast, we establish sharpness property for a *manifold-constrained* nonsmooth optimization problem, thus expanding the currently limited repertoire of sharp *constrained* nonsmooth nonconvex optimization problems.

**Remark 3** Theorem 1 proves the correctness of the formulation (3). It asserts that if one wants to compute the underlying rotations  $\mathbf{X}^*$ , one possibility is to solve problem (3) to its global optimality.

With the help of sharpness property, we establish the linear convergence guarantee for our ManSGM algorithm for solving problem (3), by means of generalizing the analysis framework developed in [11].

---

**Algorithm 2** Spectral Initialization
 

---

- 1: **Input:** Construct  $\mathbf{Y} \in \mathbb{R}^{nd \times nd}$  with its  $(i, j)$ -th block being  $\mathbf{Y}_{i,j} \in \mathbb{R}^{d \times d}$ .
  - 2: Compute the  $d$  (unit) leading eigenvectors of  $\mathbf{Y}$  to get  $\{\mathbf{u}_1, \dots, \mathbf{u}_d\}$ .
  - 3: Set  $\mathbf{W}_1 = \sqrt{n}[\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d] \in \mathbb{R}^{nd \times d}$  and  $\mathbf{W}_2 = \sqrt{n}[-\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d]$ .
  - 4: Set  $\widetilde{\mathbf{X}}^0 = \begin{cases} \mathbf{W}_1, & \text{if } \text{dist}(\text{SO}(d)^n, \mathbf{W}_1) \leq \text{dist}(\text{SO}(d)^n, \mathbf{W}_2), \\ \mathbf{W}_2, & \text{otherwise.} \end{cases}$
  - 5:  $\mathbf{X}^0 = \mathcal{P}_{\text{SO}(d)^n}(\widetilde{\mathbf{X}}^0)$ .
  - 6: **Output:** Initial point  $\mathbf{X}^0$ .
- 

**Theorem 2 (linear convergence)** *Under the condition of Theorem 1. Let  $\{\mathbf{X}^k\}_{k \geq 0}$  be generated by ManSGM (cf. Algorithm 1) with  $\text{dist}(\mathbf{X}^0, \mathbf{X}^*) < \frac{2\alpha}{\tau}$ . Suppose further the stepsize  $\mu_k$  is computed by  $\mu_k = \rho^k \mu_0$ , where  $\mu_0 < \min \left\{ \frac{2\alpha e_0 - \tau e_0^2}{4ML\alpha + L^2 + nM^2\tau^2}, \frac{e_0}{2\alpha - \tau e_0} \right\}$  and  $1 > \rho \geq \frac{\sqrt{1 + (\tau - 2\alpha/e_0)\mu_0 + (4ML\alpha + L^2 + nM^2\tau^2)\mu_0^2/e_0^2}}{e_0} > 0$  with  $e_0 = \max \left\{ \text{dist}(\mathbf{X}^0, \mathbf{X}^*), \frac{\alpha}{\tau} \right\}$ . Then, with probability  $1 - \mathcal{O}(1/n)$ , we have*

$$\text{dist}(\mathbf{X}^k, \mathbf{X}^*) \leq \rho^k \cdot e_0, \quad \forall k \geq 0.$$

Here,  $L = \sqrt{n}\tau$  with  $\tau = \frac{8}{3}nq$ ,  $\alpha$  is the sharpness parameter shown in (5),  $M > 0$  is a numerical constant.

The linear convergence result requires initializing our ManSGM close to the global minimum. We use a properly designed initialization method in Algorithm 2 to address this issue. The following proposition provides the upper bound on the estimation error of the initial point returned by Algorithm 2.

**Proposition 1 (good initialization)** *Under the condition of Theorem 1 and let  $\mathbf{X}^0$  be generated by Algorithm 2. Then, with probability at least  $1 - 2d/n$ , we have*

$$\text{dist}(\mathbf{X}^0, \mathbf{X}^*) \leq \mathcal{O}\left(1/\sqrt{q}\right). \quad (6)$$

## 4. Experiments

**Experiments on synthetic datasets.** We conduct a series of experiments on robust synchronization of rotations to demonstrate the performance of our ManSGM algorithm on synthetic datasets. We first generate the ground truth rotations  $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$  from a uniform distribution on  $\text{SO}(d)$ . Then, we generate a synchronization graph  $\mathcal{G}(\mathcal{V}, \mathcal{E})$  according to Erdős-Rényi model  $\mathcal{G}(n, q)$ . The available noise measurements  $\mathbf{Y}_{ij}$  with  $(i, j) \in \mathcal{E}$  are drawn according to (2) with co-outliers ratio  $p$ .

We compare the performance of our ManSGM with the SDR method developed in [15] which the authors used ADMM to solve the resultant SDP. We utilize spectral initialization Algorithm 2 to initialize ManSGM. For ADMM, we set the stepsize to be 1. For the following experiments, we assume an output  $\widehat{\mathbf{X}}$  returned by the algorithms to be a successful recovery if  $\text{dist}(\mathbf{X}, \mathbf{X}^*) < 10^{-4}$ .

- (a) **Convergence verification.** We evaluate the convergence performance of our ManSGM on  $\text{SO}(3)^{500}$  with different stepsize parameters  $\mu_0$  and  $\rho$ . We can observe from Figure 2a that our algorithm can converge linearly to the ground truth rotation  $\mathbf{X}^*$  for a wide range of  $\mu_0$  and  $\rho$ . However, it is noteworthy that too small a  $\rho$  (e.g.,  $\rho < 0.7$ ) results in the early termination before reaching the global minimum as in Figure 2b.

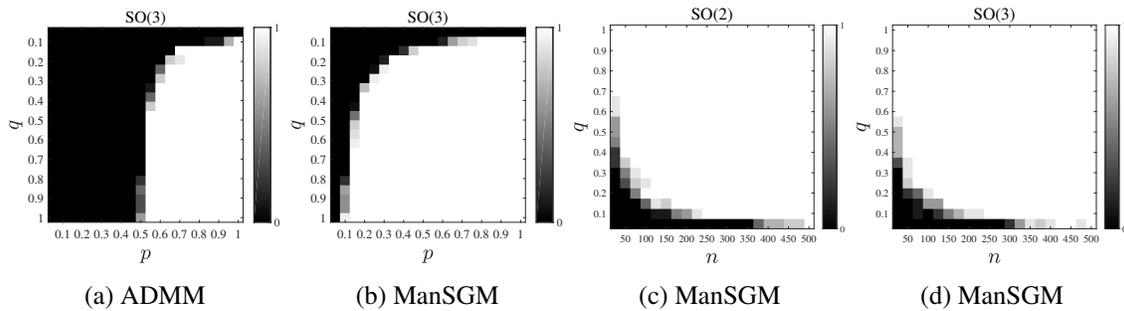


Figure 1: (a)-(b) Phase transition on  $(p, q)$  with fixed  $n = 500$  on  $\text{SO}(3)$  for ADMM and ManSGM. (c)-(d) Phase transition on  $(n, q)$  with fixed  $p = 0.8$  on  $\text{SO}(2)$  and  $\text{SO}(3)$  for ManSGM.

- (b) **Phase transition.** We plot the phase transition for  $(p, q)$  and  $(n, q)$  in Figure 1. For each  $(p, q)$  and  $(n, q)$ , we repeat the simulation for 10 times. Whiter pixel indicates higher successful recovery probability. As shown in Figure 1(a)-(b), in the  $\text{SO}(3)$  case, the white spaces generated via our ManSGM is much larger than that generated via ADMM, which indicates the nonsmooth nonconvex optimization (3) can handle much severe corruption and incompleteness than the SDR method.
- (c) **Running time comparison.** To compare the running time, we terminate both our ManSGM and ADMM when the Gram matrix  $\mathbf{G} := \mathbf{X}^T \mathbf{X}$  is close to  $\mathbf{G}^* := \mathbf{X}^{*T} \mathbf{X}^*$  (i.e.,  $\|\mathbf{G} - \mathbf{G}^*\|_F < 10^{-3}$ ). It is not surprising to observe that our ManSGM can be about  $3\times$  faster than ADMM as shown in Figure 2c due to the scalability issue of SDR method and the low computational load in each iteration of our ManSGM.
- (d) **Recovery with varying noise.** We compare the recovery probability between our method and ADMM with fixed  $n = 500$  and varying  $p$  and  $q$ . As can be observed from Figure 2d, solving the nonsmooth nonconvex optimization (3) enables successful recovery with much lower co-outlier ratio  $p$  and observation ratio  $q$  on both  $\text{SO}(2)$  and  $\text{SO}(3)$  scenarios.

**Experiment on real dataset.** We compare our ManSGM algorithm and ADMM algorithm on the Notre Dame dataset where the observations and ground truth are provided.<sup>4</sup> The results are in Table 1. We set  $\mu_0 = 0.002$ ,  $\rho = 0.999$  for ManSGM and stepsize to be 1 for ADMM. For both algorithms, we run 3000 iterations. From Table 1, we can observe that our ManSGM outperforms ADMM in terms of both running time and accuracy.

Table 1: Results on Notre Dame dataset with  $n = 715$  rotations and  $|\mathcal{E}| = 64678$  observations.

Method	iteration	time (s)	dist( $\mathbf{X}, \mathbf{X}^*$ )
ADMM	3000	3616	19.30
ManSGM	3000	427	16.52

4. Available on <https://github.com/RafaelMarinheiro/RotationAveraging/tree/master/tests/data>.

## References

- [1] Boris Alexeev, Afonso S Bandeira, Matthew Fickus, and Dustin G Mixon. Phase retrieval with polarization. *SIAM Journal on Imaging Sciences*, 7(1):35–66, 2014.

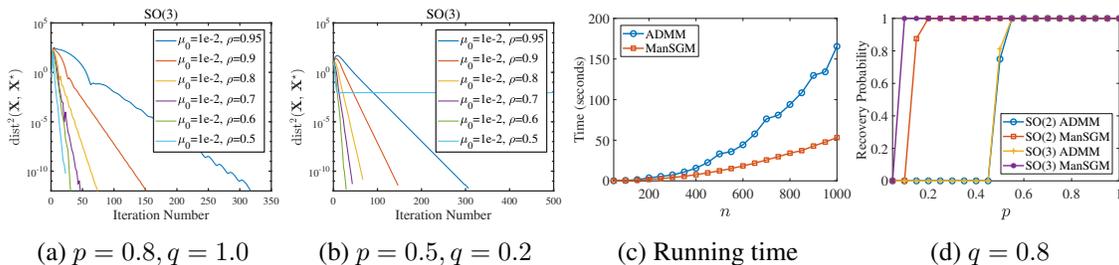


Figure 2: (a)-(b) Convergence of ManSGM on SO(3) for varying  $\mu_0$  and  $\rho$ , here we fix  $n = 500$ , (c) Running time (seconds) with varying  $n$  on SO(3), here  $p = 0.8, q = 1$ . (d) Recovery probability with fixed  $n = 500$  on both SO(2) and SO(3).

[2] N. Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.

[3] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *The Journal of Machine Learning Research*, 15(1):1455–1459, 2014.

[4] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11, 2011.

[5] Mihai Cucuringu, Yaron Lipman, and Amit Singer. Sensor network localization by eigenvector synchronization over the euclidean group. *ACM Transactions on Sensor Networks (TOSN)*, 8(3):1–42, 2012.

[6] John C Duchi and Feng Ruan. Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval. *Information and Inference: A Journal of the IMA*, 8(3):471–529, 2018.

[7] Richard Hartley, Jochen Trunpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *International journal of computer vision*, 103(3):267–305, 2013.

[8] Gilad Lerman and Tyler Maunu. Fast, robust and non-convex subspace recovery. *Information and Inference: A Journal of the IMA*, 7(2):277–336, 2018.

[9] Gilad Lerman and Yunpeng Shi. Robust group synchronization via cycle-edge message passing. *arXiv preprint arXiv:1912.11347*, 2019.

[10] Xiao Li, Zhihui Zhu, Anthony Man-Cho So, and Rene Vidal. Nonconvex robust low-rank matrix recovery. *arXiv preprint arXiv:1809.09237*, 2018.

[11] Xiao Li, Shixiang Chen, Zengde Deng, Qing Qu, Zhihui Zhu, and Anthony Man Cho So. Nonsmooth optimization over stiefel manifold: Riemannian subgradient methods. *arXiv preprint arXiv:1911.05047*, 2019.

[12] Daniel Martinec and Tomas Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

- [13] Mostafa Rahmani and George K Atia. Coherence pursuit: Fast, simple, and robust principal component analysis. *IEEE Transactions on Signal Processing*, 65(23):6260–6275, 2017.
- [14] Roberto Tron and René Vidal. Distributed image-based 3-d localization of camera sensor networks. In *Proceedings of the 48th IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 901–908. IEEE, 2009.
- [15] Lanhui Wang and Amit Singer. Exact and stable recovery of rotations for robust synchronization. *Information and Inference: A Journal of the IMA*, 2(2):145–193, 2013.
- [16] Lanhui Wang, Amit Singer, and Zaiwen Wen. Orientation determination of cryo-em images using least unsquared deviations. *SIAM journal on imaging sciences*, 6(4):2450–2483, 2013.
- [17] Wei Hong Yang, Lei-Hong Zhang, and Ruyi Song. Optimality conditions for the nonlinear programming problems on Riemannian manifolds. *Pacific Journal of Optimization*, 10(2): 415–434, 2014.
- [18] Stella Yu. Angular embedding: A robust quadratic criterion. *IEEE transactions on pattern analysis and machine intelligence*, 34(1):158–173, 2011.
- [19] Zhihui Zhu, Yifan Wang, Daniel Robinson, Daniel Naiman, Rene Vidal, and Manolis Tsakiris. Dual principal component pursuit: Improved analysis and efficient algorithms. In *Advances in Neural Information Processing Systems*, pages 2171–2181, 2018.