ANCHORED SPECTRAL ESTIMATOR FOR RIGID MOTION SYNCHRONIZATION

Ziyue Zhao[†] *Huikang Liu*[⋆] *Man-Chung Yue*[§]

[†]Department of Data and Systems Engineering, The University of Hong Kong
[⋆]Antai College of Economics and Management, Shanghai Jiao Tong University
[§]Department of Applied Mathematics, The Hong Kong Polytechnic University
zyuezhao@connect.hku.hk, hkl1u@sjtu.edu.cn, manchung.yue@polyu.edu.hk

ABSTRACT

A rigid motion in $\mathbb{R}^d$ consists of a proper rotation and a translation, and it can be represented as a matrix in $\mathbb{R}^{(d+1)\times(d+1)}$. The problem of rigid motion synchronization aims to estimate a collection of rigid motions $G_1^*, \ldots, G_n^*$ from noisy observations of their comparisons $G_i^{*-1}G_j^*$. Such problems naturally arise in diverse applications across signal processing, robotics, and computer vision, and have thus attracted intense research attention in recent years. Motivated by geometric considerations, this paper develops a novel spectral approach for rigid motion synchronization, called the anchored spectral estimator (ASE). Theoretically, we establish uniform estimation error bounds for the estimators produced by ASE. Empirically, we show that ASE outperforms the widely used two-stage approach, which first estimates the rotations and then the translations. Further numerical experiments on the multiple point-set registration problem are presented to demonstrate the superiority of ASE over state-of-the-art methods.

Index Terms— Special Euclidean group synchronization, Rigid motions, Spectral method, Estimation error bound

1. INTRODUCTION

A rigid motion in $\mathbb{R}^d$ consists of a proper rotation and a translation and can be represented as a $(d+1)\times(d+1)$ matrix. The problem of rigid motion synchronization is to recover a collection of rigid motions $G_1^*, \ldots, G_n^* \in \mathrm{SE}(d)$ from the noisy observations of their pairwise comparisons $G_i^{*-1}G_j^*$. Since the set of rigid motions forms the special Euclidean group $\mathrm{SE}(d)$, the problem is also called special Euclidean group synchronization. We denote it by SE-Sync for simplicity. SE-Sync finds a wide range of applications across many scientific and engineering domains, including simultaneous localization and mapping [1, 2], structure from motion [3, 4], multiple point-set registration [5], sensor network localization [6], and cryogenic electron microscopy [7].

SE-Sync is very challenging from a computational perspective, hindering its real-world applications. Indeed, it contains NP-hard problems as special cases [8]. Moreover, the natural approach via its least square estimator results in a non-convex and nonlinear constrained optimization problem. Existing approaches to SE-Sync and more generally other synchronization problems can roughly be divided into three categories: semidefinite relaxations (SDR), non-convex algorithms, and spectral estimators. The SDR approach relaxes the non-convex least square estimation problem to a convex semidefinite programming problem by lifting techniques [9–12]. Many SDRs for synchronization problems are proved to be tight (*i.e.*, the optimal solutions to the relaxed problem are also optimal to the original non-convex problem). However, the resulting semidefinite programming problem is of dimension $O(nd) \times O(nd)$ and hence prohibitively expensive to solve. Non-convex algorithms directly solve the non-convex optimization problem associated with the least squares estimator by iterative local search algorithms [8, 13–17]. These algorithms often enjoy strong statistical guarantees if they are suitably initialized [8, 13–17]. Finally, spectral estimators first compute a small number of eigenvectors of a data matrix constructed from the comparison observations and then generate the output by applying a certain rounding procedure to the eigenvectors [18–23]. Although spectral estimators are used as an initialization in many non-convex algorithms, they often enjoy an estimation error that is of the same order as those non-convex algorithms [24]. Furthermore, they are highly efficient and easy to implement. In view of these advantages, this paper focuses on spectral estimators.

In the context of SE-Sync, there have been a number of existing spectral estimators, and they can be subdivided into two types. The first type includes [25, 26] and adopts a divide-and-conquer, two-stage strategy that first estimates the rotations using only the rotation comparisons extracted from the full observations, but ignores the translation comparisons. The translations are then estimated by a separate procedure based on the rotations estimated in the first stage. In contrast, the second type jointly estimates these two constituents of the rigid motions. Examples of the second type include [21–23]. Since the translation comparisons are coupled with the rotations and hence can provide extra information for rotation estimation, the second type generally performs better than the first when the noise on the translation comparisons is not too large.

On the other hand, regardless of the type, the eigenvectors obtained from the first step of spectral estimators suffer from a symmetry issue; see Section 2.2 for details. If the issue is not properly handled, the performance of the spectral estimators can be significantly deteriorated.

1.1. Contributions

In this paper, we propose and analyze a new spectral estimator for SE-Sync. Our contributions are as follows.

- First, exploiting the special geometry of the set of rotations, we develop a novel spectral estimator called the anchored spectral estimator (ASE), which uses anchored projections for the rounding step to tackle the symmetry issue.
- Second, we establish a strong theoretical guarantee for ASE, which asserts that its estimators enjoy a uniform estimation error bound of the order $O((\sigma_1 + \sigma_2^2)(\sqrt{d} + \sqrt{\log n})d/\sqrt{n})$, where σ_1^2 and σ_2^2 are the variances of the noise on the rotation and translation comparisons, respectively.

- Third, we empirically verify the advantage of ASE over a two-stage approach for solving SE-Sync through synthetic data. We also demonstrate the superiority of ASE over state-of-the-art spectral estimators of the second type via a numerical experiment on the multiple point-set registration problem using real datasets.

1.2. Notation

For a vector x, $\|x\|_2$ denotes its 2-norm. For a matrix X, $X^\top$, $\|X\|$ and $\|X\|_{\mathrm{F}}$ denote its transpose, operator norm, and Frobenius norm, respectively. Also, $\mathrm{vec}(X)$ denotes the vector obtained by stacking its columns. Given a sequence $\{x_i\}_{i=1}^n \in \mathbb{R}^{a\times b}$ of vectors or matrices, $\mathrm{BlkDiag}(x_1,\dots,x_n) \in \mathbb{R}^{na\times nb}$ denotes the block matrix with its i-th diagonal block being x_i. Given two matrices X and Y, their Kronecker product is denoted by $X \otimes Y$. The $a\times a$ identity matrix and the all-one matrix are denoted by I_a and J_a. The sets of $a\times a$ orthogonal and special orthogonal matrices are denoted by $\mathrm{O}(a) = \{Q \in \mathbb{R}^{a\times a} \mid QQ^\top = I_a\}$ and $\mathrm{SO}(a) = \{Q \in \mathbb{R}^{a\times a} \mid QQ^\top = I_a,\ \det(Q) = 1\}$, respectively. We also denote by $[n]$ the set $\{1,\dots,n\}$.

2. A NEW SPECTRAL METHOD FOR SE-Sync

2.1. SE-Sync and Its Least Squares Estimator

This section formally introduces the special Euclidean group synchronization problem (SE-Sync) and our proposed spectral method ASE. To begin, recall that a rigid motion in $\mathbb{R}^d$ consists of a proper rotation $R \in \mathrm{SO}(d)$ and a translation $t \in \mathbb{R}^d$, and can be represented as a $(d+1)\times(d+1)$ matrix of the form

$$\begin{pmatrix} R^\top & t \\ 0 & 1 \end{pmatrix}.$$

The set of all such matrices is called the special Euclidean group and denoted by $\mathrm{SE}(d)$. In SE-Sync, we are interested in estimating a collection of rigid motions $G_1^*,\dots,G_n^* \in \mathrm{SE}(d)$ based on noisy observations of their comparisons $G_i^{*-1}G_j^*$. More precisely, we assume that we are given the following observations:

$$C_{ij} = G_i^{*-1}G_j^* + W_{ij}, \quad i,j \in [n].$$

In this observation model, the matrix $W_{ij} \in \mathbb{R}^{(d+1)\times(d+1)}$ represents the noise and takes the form

$$W_{ij} = \begin{pmatrix} W_{ij}^R & w_{ij}^t \\ 0 & 0 \end{pmatrix}, \quad i,j \in [n],\ i \neq j,$$

where for any $i \neq j$, $W_{ij}^R \in \mathbb{R}^{d\times d}$ has i.i.d. Gaussian entries with mean 0 and variance σ_1^2, $w_{ij}^t \in \mathbb{R}^d$ has i.i.d. Gaussian entries with mean 0 and variance σ_2^2, and they are statistically independent; and for any $i \in [n]$, $W_{ii} = 0$. For any $i \in [n]$, the rotation and translation associated with the ground truth G_i^* are denoted by R_i^* and t_i^*, respectively. Without loss of generality, we assume that $\sum_{i=1}^n t_i^* = 0$.

The least squares estimator, which also coincides with the maximum likelihood estimator under our noise assumption, is defined as any optimal solution to the problem

$$\min_{G_i\in\mathrm{SE}(d)} \sum_{i=1}^n\sum_{j=1}^n \|G_i^{-1}G_j - C_{ij}\|_{\mathrm{F}}^2 = \min_{\substack{R_i\in\mathrm{SO}(d),\\ t_i\in\mathbb{R}^d}} \sum_{i=1}^n\sum_{j=1}^n \left(\|R_iR_j^\top - S_{ij}\|_{\mathrm{F}}^2 + \|t_j - t_i - R_i^\top s_{ij}\|_2^2\right), \tag{2.1}$$

where for all $i \neq j$, $S_{ij} = R_i^*R_j^{*\top} + W_{ij}^R \in \mathbb{R}^{d\times d}$ and $s_{ij} = s_{ij}^* + w_{ij}^t \in \mathbb{R}^d$ with $s_{ij}^* = R_i^*(t_j^* - t_i^*) \in \mathbb{R}^d$.

Throughout the paper, we adopt the following convention to denote block matrices. For any sequence $\{x_i\}_{i=1}^n \subseteq \mathbb{R}^{a\times b}$ and two-dimensional sequence $\{y_{ij}\}_{i,j=1}^n \in \mathbb{R}^{a\times b}$ of vectors or matrices, we denote by $x \in \mathbb{R}^{na\times b}$ and $y \in \mathbb{R}^{na\times nb}$ the block matrices whose i-th and ij-th blocks are x_i and y_{ij}, respectively. For example, $s \in \mathbb{R}^{nd\times n}$ is a block matrix whose ij-th block is $s_{ij} \in \mathbb{R}^d$.

We then simplify the least squares estimator (2.1). Let $L = nI_n - J_n$ be the Laplacian matrix of a complete graph, $S \in \mathbb{R}^{nd\times nd}$ be the block matrix whose ij-th block is S_{ij}, $\hat{T} \in \mathbb{R}^{nd\times n}$ be the block matrix $\hat{T} = \mathrm{BlkDiag}(\sum_{j=1}^n s_{1j},\dots,\sum_{j=1}^n s_{nj}) - s$, and $\hat{\Sigma} \in \mathbb{R}^{nd\times nd}$ be the block matrix $\hat{\Sigma} = \mathrm{BlkDiag}(\sum_{j=1}^n s_{1j}s_{1j}^\top,\dots,\sum_{j=1}^n s_{nj}s_{nj}^\top)$. Then, we can equivalently reformulate problem (2.1) as

$$\min_{\substack{R\in\mathrm{SO}(d)^n\\ t\in\mathbb{R}^{nd}}} \mathrm{Tr}\Big(2R^\top(nI_{nd} - S)R + \hat{\Sigma} + 2t^\top(L\otimes I_d)t + 2\mathrm{vec}(R^\top\hat{T})^\top t\Big). \tag{2.2}$$

For any fixed R, problem (2.2) is an unconstrained quadratic optimization and can be solved analytically with optimal solution being $t = -\frac{1}{2n}\mathrm{vec}(R^\top\hat{T})$; see [11, Lemma 4]. We can then eliminate the variable t and further reformulate problem (2.1) as

$$\min_{R\in\mathrm{SO}(d)^n} \mathrm{Tr}(R^\top\Omega R), \tag{2.3}$$

where

$$\Omega = 2nI_{nd} - 2S + \hat{\Sigma} - \frac{1}{2n}\hat{T}\hat{T}^\top. \tag{2.4}$$

Since Ω is not positive semidefinite, the objective function $\mathrm{Tr}(R^\top\Omega R)$ is non-convex. Moreover, the block-wise $\mathrm{SO}(d)$ constraint is also non-convex and highly nonlinear. Problem (2.3) is therefore difficult to solve.

2.2. Anchored Spectral Estimator

We propose a new spectral estimator, called the anchored spectral estimator (ASE), for solving problem (2.3) and thus SE-Sync; see Algorithm 1. Roughly speaking, ASE first relaxes the block-wise special-orthogonality constraint $\mathrm{SO}(d)^n$ to the orthogonality constraint $\Phi^\top\Phi = nI_d$, turning our task into an eigenvalue problem. After computing the d eigenvectors $\Phi \in \mathbb{R}^{nd\times d}$ associated with the smallest eigenvalues of Ω, we round each block $\Phi_i \in \mathbb{R}^{d\times d}$ by projecting a right-rotated version $\Phi_i\Phi_1^\top$ back onto $\mathrm{SO}(d)$ using the projection operator $\Pi_{\mathrm{SO}(d)}$. This yields the estimators $\hat{R}_i$ to the target rotations R_i^* for $i \in [n]$. The projection $\Pi_{\mathrm{SO}(d)}$ can be computed efficiently via a singular value decomposition; see [16, Section 5.1.2]. Next, the translations are recovered as the optimal solutions to problem (2.2) with a fixed $R = \hat{R}$. In other words, $\hat{t}_i = -(\frac{1}{2n}(\hat{T}^\top \otimes I_d)\,\mathrm{vec}(R^\top))_i = -\frac{1}{2n}(\mathrm{vec}(\hat{R}^\top\hat{T}))_i$ for $i \in [n]$.

Algorithm 1: Anchored Spectral Estimator for $\mathrm{SE}(d)$

Input: The data matrix Ω defined in (2.4).

Output: Rigid motions $(\hat{R}_i, \hat{t}_i)$ for $i \in [n]$.

1 Compute the eigenvectors $\Phi \in \mathbb{R}^{nd\times d}$ associated with the d smallest eigenvalues of Ω;

2 For $i \in [n]$, compute

$$\hat{R}_i = \Pi_{\mathrm{SO}(d)}(\Phi_i \Phi_1^\top) \quad \text{and} \quad \hat{t}_i = -\frac{1}{2n}(\mathrm{vec}(\hat{R}^\top \hat{T}))_i.$$

A distinctive feature of ASE is that each block Φ_i is right-multiplied by $\Phi_1^\top$ before projecting onto $\mathrm{SO}(d)$. In the noiseless case, the block column matrix Φ consists of the true rotations up to right-multiplication by an unknown orthogonal matrix. The block Φ_1 therefore acts as an anchor, eliminating this global orthogonal ambiguity and correctly locating the blocks Φ_i within $\mathrm{SO}(d)$. In the presence of noise, Lemma 3.2 in Section 3 implies that Φ_1 remains close to $R_1^\star$ (up to an orthogonal right-multiplication), and therefore the anchoring mechanism continues to be effective.

2.3. Other Spectral Estimators

The spectral method that is closest to ASE is the one developed in [23], which uses the same matrix Ω for the eigenvalue problem in the first step. Nevertheless, unlike our anchored projection, their rounding procedure directly projects each block Φ_i onto $\mathrm{SO}(d)$. Since the blocks Φ_i may have determinant -1, such a rounding procedure suffers from the risk of distorting their relative positions.

There have also been other spectral methods developed for SE-Sync. The spectral method in [21] first computes the eigenvectors $\Phi' \in \mathbb{R}^{n(d+1)\times(d+1)}$ of a certain $n(d+1) \times n(d+1)$ data matrix and then projects the blocks Φ_i' onto $\mathrm{SE}(d)$. As another example, a spectral method for SE-Sync is developed in [22] based on the dual quaternion representation of rigid motions. The method seems to be specific to $d = 3$, and it is unclear how it can be generalized to other dimensions.

We have conducted an experiment on the multiple point-set registration problem to compare our ASE with the spectral methods in [21], [22], and [23]. The experiment results show that ASE performs the best in terms of estimation error.

3. ESTIMATION ERROR BOUND

The proposed ASE enjoys a strong theoretical guarantee on its estimation error. Specifically, we show that the (unsquared) ℓ_2 estimation error of the estimators $\hat{G}_1, \ldots, \hat{G}_n$ enjoys a uniform bound of the order $O((\sigma_1 + \sigma_2^2)(\sqrt{d} + \sqrt{\log n})d/\sqrt{n})$. To state the result, we let $M_t = \max_{i\in[n]} \|t_i^*\|_2$.

Theorem 3.1. Consider Algorithm 1. There exist absolute constants $c_0, c_1, c_2, c_3, c_4, c_5 > 0$ such that if $c_1\sigma_1 + c_2 M_t \sigma_2 + c_3\sigma_2^2 \le \frac{c_0\sqrt{n}}{\sqrt{d}+\sqrt{\log(n)}}$, then

$$\max_{i\in[n]} \min_{Q\in\mathrm{SE}(d)} \|\hat{G}_i - QG_i^*\|_\mathrm{F}$$
$$\le c_4 M_t (M_t^2+1)^2 (c_1\sigma_1 + c_2 M_t \sigma_2 + c_3\sigma_2^2)\frac{(d\sqrt{d} + d\sqrt{\log n})}{\sqrt{n}},$$

with probability at least $1 - c_5 n^{-1}$.

3.1. Proof Outline

We next outline the proof of Theorem 3.1, which mainly consists of the three lemmas below. The first one decomposes the (uniform) estimation error of estimated rigid motions into two parts: a (scaled) rotation error and a translation error. This lemma is proved based on [16, Lemma 2].

Lemma 3.1. Consider Algorithm 1. There exists an absolute constant $c > 0$ such that

$$\max_{i\in[n]} \min_{Q\in\mathrm{SE}(d)} \|\hat{G}_i - QG_i^*\|_\mathrm{F}$$
$$\le c \max_{i\in[n]} \|\Phi_i\| \underbrace{\max_{i\in[n]} \|\Phi_i - R_i^*\bar{Q}\|_\mathrm{F}}_{\text{rotation error}} + \underbrace{\max_{i\in[n]} \|\hat{t}_i - (\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top))^\top t_i^*\|_2}_{\text{translation error}},$$

where $\bar{Q} = \arg\min_{Z\in\mathrm{O}(d)} \|\Phi - R^* Z\|_\mathrm{F}$.

The second lemma establishes a bound on the (scaled) rotation error, which as discussed in Section 2.2, provides the theoretical underpinning for the anchoring mechanism of ASE.

Lemma 3.2. Consider Algorithm 1. There exist absolute constants $c_0, c_1, c_2, c_3, c_4, c_5 > 0$ such that if $c_1\sigma_1 + c_2 M_t\sigma_2 + c_3\sigma_2^2 \le \frac{c_0\sqrt{n}}{\sqrt{d}+\sqrt{\log(n)}}$, then

$$\max_{i\in[n]} \|\Phi_i\| \max_{i\in[n]} \|\Phi_i - R_i^*\bar{Q}\|_\mathrm{F}$$
$$\le c_4 (M_t^2+1)^2 (c_1\sigma_1 + c_2 M_t\sigma_2 + c_3\sigma_2^2)\frac{(d\sqrt{d} + d\sqrt{\log n})}{\sqrt{n}},$$

with probability at least $1 - c_5 n^{-1}$, where $\bar{Q}$ is defined in Lemma 3.1.

Lemma 3.2 is derived via the popular leave-one-out technique. To understand the technique, we note that the data matrix Ω in (2.4) can be decomposed as the sum of an informative part and a noise matrix Δ and that the proof of Lemma 3.2 involves bounding a certain product term between the eigenvectors Φ and the i-th block row Δ_i of the noise matrix Δ, which are statistically dependent. This obstructs controlling the product term by using concentration inequalities, which rely on statistical independence. To tackle this issue, the leave-one-out technique approximates Φ by the eigenvectors $\Phi^{(i)}$ of the matrix obtained from deleting the i-th block row and i-th block column from the matrix Δ. The upshot is that the approximate eigenvectors $\Phi^{(i)}$ are statistically independent of Δ_i, and concentration inequalities can therefore be applied. We emphasize that although the leave-one-out technique has been applied to study other group synchronization problems [27, 28], the analysis for SE-Sync is different and requires additional techniques due to the more complicated structure of the noise matrix Δ.

The final lemma provides a bound on the translation error, the proof of which is similarly based on the leave-one-out technique.

Lemma 3.3. Consider Algorithm 1. There exist absolute constants $c_0, c_1, c_2, c_3, c_4, c_5 > 0$ such that if $c_1\sigma_1 + c_2 M_t\sigma_2 + c_3\sigma_2^2 \le \frac{c_0\sqrt{n}}{\sqrt{d}+\sqrt{\log(n)}}$, then

$$\max_{i\in[n]} \|\hat{t}_i - (\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top))^\top t_i^*\|_2$$
$$\le c_4 M_t (M_t^2+1)^2 (c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2)\frac{(d\sqrt{d} + d\sqrt{\log n})}{\sqrt{n}},$$

with probability at least $1 - c_5 n^{-1}$, where $\bar{Q}$ is defined in Lemma 3.1.

Theorem 3.1 then follows by combining the Lemmas 3.1, 3.2, and 3.3.

4. NUMERICAL EXPERIMENTS

In this section, we conduct numerical experiments to investigate the practical performance of ASE. Our codes are available at `https://github.com/ziyuezhao1/GSP-SE_D`.

4.1. Comparison with the Two-Stage Approach

We first empirically study how the estimation error of ASE varies against the noise magnitudes σ_1 and σ_2, especially in comparison with the two-stage approach (see Section 1). To this end, we consider $d = 3$ and $n = 500$. The ground truth is generated as follows. The rotations R_i^* are generated independently according to the uniform distribution on $\mathrm{SO}(3)$. And the translations t_i^* are generated independently according to the 3-dimensional Gaussian distribution with zero-mean and covariance I_3. For each value of σ_1 and σ_2, the estimation errors for 25 independent realizations are recorded. Figure 1 shows the box-whisker plots of the maximum block-wise estimation error against the σ_2 (with $\sigma_1 = 1$) and σ_1 (with $\sigma_2 = 1$) on the left and right panels, respectively. The red plot corresponds to our ASE, whereas the blue plot corresponds to the two-stage approach.

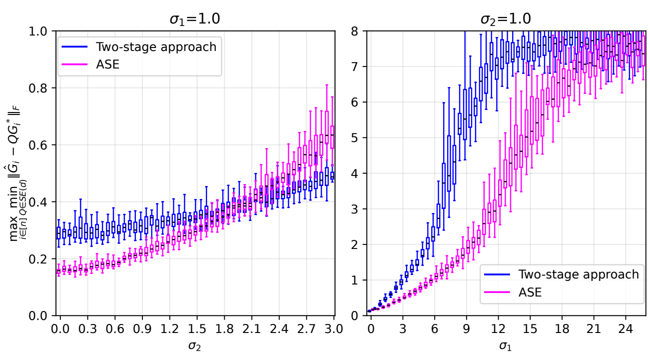


Fig. 1: Estimation errors against translation and rotation noise magnitudes.

From the left panel of Figure 1, we see that the estimation error of ASE is majorized by that of the two-stage approach when the translation error is not too large. This shows that a holistic approach that jointly estimates the translations and rotations is preferable over the divide-and-conquer two-stage approach when the translation noise is moderate. The advantage is even more apparent from the right panel of Figure 1 where we vary σ_1 with a fixed $\sigma_2 = 1$.

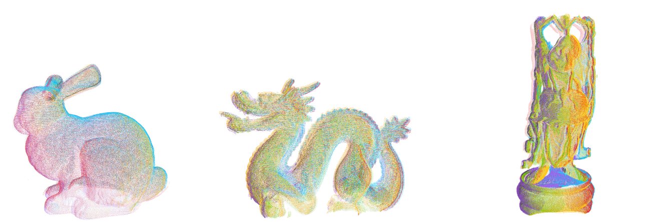
4.2. Multiple Point-Set Registration

In the next experiment, we compare the performance of ASE against other spectral methods on the multiple point-set registration problem, which aims to optimally align a collection of 3-dimensional point clouds by transforming them using rigid motions and is naturally an instance of SE-Sync. The experimental setup is similar to that in [21]. In particular, we use the Bunny, Dragon (standing), and Happy Buddha (standing) datasets from the Stanford 3D Scanning Repository [29]. The noisy comparison observations C_{ij} are obtained as follows. We first perturb each true relative rotation by rotating it by a random angle uniformly distributed in $[0, 8^\circ]$ about a uniformly random axis, and each true relative translation by a zero-mean Gaussian noise with a standard deviation of 0.8 millimeters. To refine the observations, we then apply the iterative closest point algorithm [30] to the observations C_{ij} as a pre-processing step. We compare our ASE with the spectral estimators developed in [21], [22], and [23]. The performance metrics are the average rotation error (*i.e.*, the average angles (in degrees) between the estimated rotations and the corresponding true rotations (up to a common rotation)) and the average translation error (*i.e.*, the average distance between the estimated translations and their corresponding true translations). The experiment results are shown in Table 1. We should point out that the approach by [23] is unsatisfactory if we implement it as described in the paper [23]. To improve its practical performance, we adopt a trick from [28] to suitably flip the signs of the eigenvectors.

Average Error	Method	Bunny	Dragon	Happy Buddha
	ASE	**0.76**	**1.62**	**1.25**
Rotation	[21]	1.00	2.01	1.50
	[22]	0.79	1.71	1.36
	[23]	1.08	14.28	1.32
	ASE	**2.59**	**3.82**	**1.58**
Translation	[21]	3.97	3.91	1.62
	[22]	8.14	5.42	2.23
	[23]	98.8	4.72	2.38

Table 1: Average rotation error (in degrees) and average translation error (in millimeters) of different spectral approaches for multiple point-set registration.

From Table 1, see that our proposed ASE achieves the smallest estimation error with respect to both rotation and translation. The 3D models constructed using our proposed ASE are shown in Figure 2.



(a) Bunny (b) Dragon (c) Happy Buddha

Fig. 2: 3D models constructed using ASE.

5. CONCLUSIONS

Motivated by the symmetry issue of existing spectral estimators, this paper develops a new spectral estimator, called the anchored spectral estimator (ASE), for rigid motion synchronization. ASE enjoys a strong theoretical guarantee on its estimation error. Numerically, we show that ASE outperforms a two-stage approach using synthetic data and several state-of-the-art spectral estimators through an experiment on the multiple point-set registration problem using real data.

6. ACKNOWLEDGMENT

Man-Chung Yue is supported in part by Hong Kong Research Grants Council under the GRF project 17309423.

7. REFERENCES

[1] Jesus Briales and Javier Gonzalez-Jimenez, "Cartan-Sync: Fast and global SE(d)-synchronization," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2127–2134, 2017.

[2] Xin Chen, Chunfeng Cui, Deren Han, and Liqun Qi, "Nonconvex pose graph optimization in slam via proximal linearized riemannian admm," *Journal of Optimization Theory and Applications*, vol. 206, no. 3, pp. 1–43, 2025.

[3] Anders Eriksson, Carl Olsson, Fredrik Kahl, and Tat-Jun Chin, "Rotation averaging with the chordal distance: Global minimizers and strong duality," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 256–268, 2019.

[4] Onur Özyeşil, Vladislav Voroninski, Ronen Basri, and Amit Singer, "A survey of structure from motion*.," *Acta Numerica*, vol. 26, pp. 305–364, 2017.

[5] Zi Jian Yew and Gim Hee Lee, "Learning iterative robust transformation synchronization," in *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021, pp. 1206–1215.

[6] Anthony Man-Cho So and Yinyu Ye, "Theory of semidefinite programming for sensor network localization," *Mathematical Programming*, vol. 109, no. 2, pp. 367–384, 2007.

[7] Amit Singer and Yoel Shkolnisky, "Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming," *SIAM Journal on Imaging Sciences*, vol. 4, no. 2, pp. 543–572, 2011.

[8] Huikang Liu, Man-Chung Yue, and Anthony Man-Cho So, "On the estimation performance and convergence rate of the generalized power method for phase synchronization," *SIAM Journal on Optimization*, vol. 27, no. 4, pp. 2426–2446, 2017.

[9] Afonso S Bandeira, Nicolas Boumal, and Amit Singer, "Tightness of the maximum likelihood semidefinite relaxation for angular synchronization," *Mathematical Programming*, vol. 163, pp. 145–167, 2017.

[10] Luca Carlone, David M Rosen, Giuseppe Calafiore, John J Leonard, and Frank Dellaert, "Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 125–132.

[11] David M Rosen, Luca Carlone, Afonso S Bandeira, and John J Leonard, "SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group," *The International Journal of Robotics Research*, vol. 38, no. 2-3, pp. 95–125, 2019.

[12] Shuyang Ling, "Solving orthogonal group synchronization via convex and low-rank optimization: Tightness and landscape analysis," *Mathematical Programming*, vol. 200, no. 1, pp. 589–628, 2023.

[13] Nicolas Boumal, "Nonconvex phase synchronization," *SIAM Journal on Optimization*, vol. 26, no. 4, pp. 2355–2377, 2016.

[14] Shuyang Ling, "Improved performance guarantees for orthogonal group synchronization via generalized power method," *SIAM Journal on Optimization*, vol. 32, no. 2, pp. 1018–1048, 2022.

[15] Chao Gao and Anderson Y Zhang, "Iterative algorithm for discrete structure recovery," *The Annals of Statistics*, vol. 50, no. 2, pp. 1066–1094, 2022.

[16] Huikang Liu, Man-Chung Yue, and Anthony Man-Cho So, "A unified approach to synchronization problems over subgroups of the orthogonal group," *Applied and Computational Harmonic Analysis*, 2023.

[17] Chao Gao and Anderson Y Zhang, "Optimal orthogonal group synchronization and rotation group synchronization," *Information and Inference: A Journal of the IMA*, vol. 12, no. 2, pp. 591–632, 2023.

[18] Amit Singer, "Angular synchronization by eigenvectors and semidefinite programming," *Applied and Computational Harmonic Analysis*, vol. 30, no. 1, pp. 20–36, 2011.

[19] Mihai Cucuringu, Yaron Lipman, and Amit Singer, "Sensor network localization by eigenvector synchronization over the Euclidean group," *ACM Transactions on Sensor Networks (TOSN)*, vol. 8, no. 3, pp. 1–42, 2012.

[20] Federica Arrigoni and Andrea Fusiello, "Synchronization problems in computer vision with closed-form solutions," *International Journal of Computer Vision*, vol. 128, no. 1, pp. 26–52, 2020.

[21] Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello, "Spectral synchronization of multiple views in SE(3)," *SIAM Journal on Imaging Sciences*, vol. 9, no. 4, pp. 1963–1990, 2016.

[22] Ido Hadi, Tamir Bendory, and Nir Sharon, "SE(3) synchronization by eigenvectors of dual quaternion matrices," *Information and Inference: A Journal of the IMA*, vol. 13, no. 3, pp. iaae014, 2024.

[23] Kevin J Doherty, David M Rosen, and John J Leonard, "Performance guarantees for spectral initialization in rotation averaging and pose-graph slam," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 5608–5614.

[24] Anderson Ye Zhang, "Exact minimax optimality of spectral methods in phase synchronization and orthogonal group synchronization," *The Annals of Statistics*, vol. 52, no. 5, pp. 2112–2138, 2024.

[25] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li, "Rotation averaging," *International Journal of Computer Vision*, vol. 103, no. 3, pp. 267–305, 2013.

[26] Venu Madhav Govindu, "Combining two-view constraints for motion estimation," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (CVPR)*. IEEE, 2001, vol. 2, pp. II–II.

[27] Shuyang Ling, "Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods," *Applied and Computational Harmonic Analysis*, vol. 60, pp. 20–52, 2022.

[28] Huikang Liu, Xiao Li, and Anthony Man-Cho So, "ReSync: Riemannian subgradient-based robust rotation synchronization," in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 2023, pp. 5236–5261.

[29] Stanford University Computer Graphics Laboratory, "The Stanford 3D Scanning Repository," `http://graphics.stanford.edu/data/3Dscanrep/`, 1996, Accessed: September 17, 2025.

[30] Paul J Besl and Neil D McKay, "Method for registration of 3-D shapes," in *Sensor fusion IV: Control Paradigms and Data Structures*. SPIE, 1992, vol. 1611, pp. 586–606.

[31] Roman Vershynin, *High-dimensional probability: An introduction with applications in data science*, vol. 47, Cambridge university press, 2018.

[32] Roman Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[33] Gilbert W Stewart, "Perturbation theory for the singular value decomposition," 1998.

8. APPENDIX

This appendix provides the proofs of Lemmas 3.1, 3.2, and 3.3. For any matrix $H \in \mathbb{R}^{na\times b}$, we denote by $H_i \in \mathbb{R}^{a\times b}$ its i-th block row. In particular, for $H \in \mathbb{R}^{n\times b}$, $H_i \in \mathbb{R}^{1\times b}$ denotes its i-th row. For eigenvectors Φ, we assume $\Phi^\top\Phi = nI_d$. Using the projection $\Pi_{\mathrm{SO}(d)}$, we can map any $nd\times d$ matrix to the feasible region $\mathrm{SO}(d)^n$ of the problem SE-Sync using the block-wise projection $\Pi^n_{\mathrm{SO}(d)}\colon \mathbb{R}^{nd\times d} \to \mathrm{SO}(d)^n$ given by $\Pi^n_{\mathrm{SO}(d)}(Y)_i = \Pi_{\mathrm{SO}(d)}([Y]_i)$, $i = 1,\dots,n$.

8.1. Proof of Lemma 3.1

Proof of Lemma 3.1. First, we define

$$\bar{Q}_E = \begin{pmatrix} \Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{SE}(d)$$

Then, for all $i \in [n]$,

$$\begin{aligned} &\min_{Q\in\mathrm{SE}(d)} \|\hat{G}_i - QG_i^*\|_\mathrm{F} \leqslant \|\hat{G}_i - \bar{Q}_E^\top G_i^*\|_\mathrm{F} \\ \leq& \|\hat{R}_i - R_i^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top)\|_\mathrm{F} + \|\hat{t}_i - (\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top))^\top t_i^*\|_2 \\ =& \|\Pi_{SO(d)}(\Phi_i\Phi_1^\top) - R_i^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top)\|_\mathrm{F} \\ &+ \|\hat{t}_i - (\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top))^\top t_i^*\|_2. \end{aligned} \tag{8.1}$$

We then bound the first term on the RHS of (8.1):

$$\begin{aligned} &\|\Pi_{\mathrm{SO}(d)}(\Phi_i\Phi_1^\top) - R_i^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top)\|_\mathrm{F} \\ \leq& 2\|\Phi_i\Phi_1^\top - R_i^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top)\|_\mathrm{F} \\ \leq& 2\|\Phi_i\Phi_1^\top - R_i^*\bar{Q}\Phi_1^\top\|_\mathrm{F} + 2\|R_i^*\bar{Q}\Phi_1^\top - R_i^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top)\|_\mathrm{F} \\ \leq& 2\|\Phi_1\|\|\Phi_i - R_i^*\bar{Q}\|_\mathrm{F} + 2\|\bar{Q}\Phi_1^\top - \Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top)\|_\mathrm{F} \\ \leq& 2\|\Phi_1\|\|\Phi_i - R_i^*\bar{Q}\|_\mathrm{F} + 2\|\bar{Q}\Phi_1^\top - R_1^{*\top}\|_\mathrm{F} \\ \leq& 2\max_i \|\Phi_i\|\|\Phi_i - R_i^*\bar{Q}\|_\mathrm{F} + 2\|\bar{Q}\Phi_1^\top - R_1^{*\top}\|_\mathrm{F}, \end{aligned} \tag{8.2}$$

where the first inequality follows from [16, Lemma 2], the second from the triangle inequality, and the third from the orthogonality of R_i^*. Note that

$$\sqrt{n} = \|\Phi\| = \sup_{\|x\|=1}\sqrt{\|\Phi x\|^2} = \sup_{\|x\|=1}\sqrt{(\Phi x)^\top(\Phi x)}$$

$$= \sup_{\|x\|=1}\sqrt{\sum_{i=1}^n(\Phi_i x)^2} \leq \sup_{\|x\|=1}\sqrt{\sum_{i=1}^n\|\Phi_i\|^2\|x\|^2} \leq \sqrt{n}\max_i\|\Phi_i\|.$$

We thus have $\max_i\|\Phi_i\| \geq 1$. Combining this fact with inequalities (8.1), (8.2), we obtain the desired result. □

8.2. Proof of Lemma 3.2

8.2.1. Structure of the Data Matrix

To analyze the rotation error, we decompose the measurement matrix Ω into ground-truth and noise components. The structure of the resulting matrix is described in Lemma 8.1, while the supporting estimates for T^*, E, and EM (for a fixed matrix M) are given in Lemma 8.2, Lemma 8.3, and Lemma 8.4, respectively. The constituent matrices are defined as follows:

$$\begin{aligned} T^* &:= \mathrm{BlkDiag}(\sum_{j=1}^n s_{1j}^*,\dots,\sum_{j=1}^n s_{nj}^*) - s^*, \\ \Sigma^* &:= \mathrm{BlkDiag}(\sum_j^n s_{1j}^*s_{1j}^{*\top},\dots,\sum_j^n s_{nj}^*s_{nj}^{*\top}), \\ E &:= \mathrm{BlkDiag}(\sum_{j=1}^n w_{1j}^t,\dots,\sum_{j=1}^n w_{nj}^t) - w^t, \\ \Delta &:= \mathrm{BlkDiag}(\sum_{j\neq 1}^n[s_{1j}^*w_{1j}^{t\top} + w_{1j}^t s_{1j}^{*\top} + (w_{1j}^t w_{1j}^{t\top} - \sigma_2^2 I_d)],\dots, \\ &\quad \sum_{j\neq n}^n[s_{nj}^*w_{nj}^{t\top} + w_{nj}^t s_{nj}^{*\top} + (w_{nj}^t w_{nj}^{t\top} - \sigma_2^2 I_d)]) \\ &\quad -\frac{1}{2n}(ET^{*\top} + T^*E^\top + EE^\top) - 2W^R, \end{aligned}$$

The matrices $T^*, E \in \mathbb{R}^{nd\times n}$ represent the ground-truth and translation noise components, respectively. Under these definitions, the measurement matrix Ω admits the following decomposition:

$$\begin{aligned} \Omega &= 2nI_{nd} - 2R^*R^{*\top} - 2W^R - \frac{1}{2n}(T^*+E)(T^*+E)^\top \\ &+ \Sigma^* + \underbrace{\mathrm{BlkDiag}(\sum_{j\neq i}^n w_{1j}^t w_{1j}^{t\top},\dots,\sum_{j\neq i}^n w_{nj}^t w_{nj}^{t\top})}_{\text{Quadratic term}} \\ &+ \underbrace{\mathrm{BlkDiag}(\sum_{j\neq 1}^n(s_{1j}^*w_{1j}^{t\top}),\dots,\sum_{j\neq n}^n(s_{nj}^*w_{nj}^{t\top}))}_{\text{Cross term}} \\ &+ \underbrace{\mathrm{BlkDiag}(\sum_{j\neq 1}^n(w_{1j}^t s_{1j}^{*\top}),\dots,\sum_{j\neq n}^n(w_{nj}^t s_{nj}^{*\top}))}_{\text{Cross term}}. \end{aligned}$$

The quadratic terms in w_{ij}^t introduce a bias. To center the noise matrix, we define

$$H := \Omega - \sigma_2^2(n-1)I_{nd} = 2nI_{nd} - 2R^*R^{*\top} + \Sigma^* - \frac{1}{2n}T^*T^{*\top} + \Delta.$$

Since Ω and H share the same eigenvectors Φ, analyzing the estimation error using H is equivalent. For the subsequent analysis, we define the following auxiliary matrices:

$$\begin{aligned}\Xi^* &:= \frac{n}{2}\operatorname{BlkDiag}(t_1^* t_1^{*\top}, \dots, t_n^* t_n^{*\top}) \\ &+ \operatorname{BlkDiag}(\sum_{j=1}^n t_j^* t_j^{*\top}, \dots, \sum_{j=1}^n t_j^* t_j^{*\top}), \\ \Upsilon^* &:= \frac{1}{2}\operatorname{BlkDiag}(t_1^* t_1^{*\top}, \dots, t_n^* t_n^{*\top})(J_n \otimes I_d) \\ &+ \frac{1}{2}(J_n \otimes I_d)\operatorname{BlkDiag}(t_1^* t_1^{*\top}, \dots, t_n^* t_n^{*\top}) \\ &+ \frac{1}{2n}(J_n \otimes I_d)\operatorname{BlkDiag}(t_1^* t_1^{*\top}, \dots, t_n^* t_n^{*\top})(J_n \otimes I_d) \\ &- \frac{1}{2}\operatorname{BlkDiag}(t_1^*, \dots, t_n^*) J_n \operatorname{BlkDiag}(t_1^*, \dots, t_n^*)^\top.\end{aligned}$$

The following lemma elucidates the structure of H.

Lemma 8.1. The matrix H admits the decomposition $H = 2nI_{nd} - 2R^*R^{*\top} + \operatorname{BlkDiag}(R_1^*, \dots, R_n^*)(\Xi^* - \Upsilon^*)\operatorname{BlkDiag}(R_1^*, \dots, R_n^*)^\top + \Delta$.

Proof. Let $BR = \operatorname{BlkDiag}(R_1^*, \dots, R_n^*)$. The ground-truth matrix T^* can be written as

$$\begin{aligned}T^* := &- BR\operatorname{BlkDiag}(t_1^*, \dots, t_n^*)L \\ &+ BR(J_n \otimes I_d)\operatorname{BlkDiag}(t_1^*, \dots, t_n^*).\end{aligned}$$

Moreover,

$$\begin{aligned}\Sigma^* = BR[&\operatorname{BlkDiag}(\sum_{j=1}^n t_j^* t_j^{*\top}, \dots, \sum_{j=1}^n t_j^* t_j^{*\top}) \\ &+ n\operatorname{BlkDiag}(t_1^* t_1^{*\top}, \dots, t_n^* t_n^{*\top})](BR)^\top.\end{aligned}$$

Consequently,

$$\begin{aligned}&\Sigma^* - \frac{1}{2n}T^*T^{*\top} \\ &= BR\{\underbrace{\operatorname{BlkDiag}(\sum_{j=1}^n t_j^* t_j^{*\top}, \dots, \sum_{j=1}^n t_j^* t_j^{*\top})}_{\Xi^*} \\ &+ \underbrace{\frac{n}{2}\operatorname{BlkDiag}(t_1^* t_1^{*\top}, \dots, t_n^* t_n^{*\top})}_{\Xi^*} \\ &\underbrace{+\frac{1}{2}\operatorname{BlkDiag}(t_1^*, \dots, t_n^*) J_n \operatorname{BlkDiag}(t_1^*, \dots, t_n^*)^\top}_{\Upsilon^*} \\ &\underbrace{-\frac{1}{2}(J_n \otimes I_d)\operatorname{BlkDiag}(t_1^* t_1^{*\top}, \dots, t_n^* t_n^{*\top})}_{\Upsilon^*} \\ &\underbrace{-\frac{1}{2}\operatorname{BlkDiag}(t_1^* t_1^{*\top}, \dots, t_n^* t_n^{*\top})(J_n \otimes I_d)}_{\Upsilon^*} \\ &\underbrace{-\frac{1}{2n}(J_n \otimes I_d)\operatorname{BlkDiag}(t_1^* t_1^{*\top}, \dots, t_n^* t_n^{*\top})(J_n \otimes I_d)}_{\Upsilon^*}\}(BR)^\top \\ &= BR(\Xi^* - \Upsilon^*)(BR)^\top.\end{aligned}$$

Noting that $(J_n \otimes I_d)\operatorname{BlkDiag}(t_1^*, \dots, t_n^*)J_n = 0$ and using the definitions of Ξ^*, Υ^* and Δ, we obtain the stated decomposition of H. □

The next lemma bounds the ground-truth translation matrix T^*.

Lemma 8.2. Let $T^* = T_D^* - s^*$, where $T_D^* := \operatorname{BlkDiag}(\sum_j s_{1j}^*, \dots, \sum_j s_{nj}^*)$. Then,

$$\begin{aligned}\|T_D^*\| \leqslant M_t n, \quad \|s^*\| \leqslant 2M_t n, \\ \|T^*\| \leqslant 3M_t n, \quad \|s_i^*\| \leqslant 2M_t\sqrt{n}.\end{aligned}$$

Proof. First, $\|T_D^*\| = \|n\operatorname{BlkDiag}(t_1^*, \dots, t_n^*)\| \leqslant nM_t$. For s^*, we have

$$\begin{aligned}&\|s^*\| \\ &= \|(J_n \otimes I_d)\operatorname{BlkDiag}(t_1^*, \dots, t_n^*) - \operatorname{BlkDiag}(t_1^*, \dots, t_n^*)J_n\| \\ &\leqslant \|(J_n \otimes I_d)\operatorname{BlkDiag}(t_1^*, \dots, t_n^*)\| + \|\operatorname{BlkDiag}(t_1^*, \dots, t_n^*)J_n\| \\ &\leqslant 2nM_t.\end{aligned}$$

Hence, $\|T^*\| \leqslant \|T_D^*\| + \|s^*\| \leqslant 3nM_t$. At the same time, $\|s_i^*\| \leqslant \|(t_1^*, \dots, t_n^*)\| + \|(t_1^*, \dots, t_1^*)\| \leqslant 2\sqrt{n}M_t$. □

The next lemma provides high-probability bounds for the noise matrix E, which is used frequently in Section 8.2.2.

Lemma 8.3. Under the Gaussian noise model, let $E = E_D - w^t$, where $E_D := \operatorname{BlkDiag}(\sum_j w_{1j}^t, \dots, \sum_j w_{nj}^t)$. Then, there exists an absolute constant $c > 0$ such that

$$\begin{aligned}\|E_D\| \leqslant c\sigma_2(\sqrt{nd} + \sqrt{n\log n}), \quad \|w^t\| \leqslant c\sigma_2\sqrt{nd}, \\ \|(w^t)_i\| \leqslant c\sigma_2\sqrt{n}, \quad \|E\| \leqslant c\sigma_2(\sqrt{nd} + \sqrt{n\log n}).\end{aligned}$$

with probability at least $1 - O(n^{-2})$.

Proof. We have $(E_D E_D^\top)_{ii} = (\sum_{j\neq i} w_{ij}^t)(\sum_{j\neq i} w_{ij}^t)^\top = Z_i Z_i^\top$, where $Z_i := \sum_{j\neq i} w_{ij}^t$ is a Gaussian random vector with variance $(n-1)\sigma_2^2 I_d$. For fixed i, by [31, Theorem 3.1.1], we have:

$$\left|\frac{\|Z_i\|_2}{\sigma_2\sqrt{n-1}} - \sqrt{d}\right| \leqslant c_1 t.$$

with probability at least $1 - 2e^{-c_2 t^2}$ for some constant $c_2 > 0$. This implies

$$\|Z_i\|_2 \leqslant \sigma_2\sqrt{n}\left(\sqrt{d} + c_1 t\right).$$

with probability at least $1 - 2e^{-c_2 t^2}$. Squaring both sides yields

$$\|Z_i\|_2^2 \leqslant n\sigma_2^2\left(\sqrt{d} + c_1 t\right)^2.$$

Choosing $t = c_3\sqrt{\log n}$ for some $c_3 > 0$, we obtain

$$\|Z_i\|_2^2 \leqslant n\sigma_2^2\left(\sqrt{d} + c_4\sqrt{\log n}\right)^2.$$

with probability at least $1 - O(n^{-3})$. Applying the union bound, we have $\|E_D\| \leqslant c_4\sigma_2(\sqrt{nd} + \sqrt{n\log n})$ with probability at least $1 - O(n^{-2})$.

Here $w^t \in \mathbb{R}^{nd\times n}$ is a block matrix whose (i,j)-th block is vector $w_{ij}^t \in \mathbb{R}^d$ for $i, j \in [n]$ where $w_{ii}^t = 0$ and $w_{ji}^t = -w_{ij}^t$ for $i \neq j$. For $i \leq j$, the random vectors w_{ij}^t is independent Gaussian with mean zero and covariance $\sigma_2^2 I_d$. Decompose $w^t = (w^t)^+ + (w^t)^-$, where $(w^t)^+$ contains w_{ij}^t for $i \leq j$ and zero otherwise, and $(w^t)^-$ contains w_{ij}^t for $i \geq j$ and zero otherwise. [31, Theorem 4.4.3] applies for each part $(w^t)^+$ and $(w^t)^-$ separately. By a union bound, we get $\|w^t\| \leqslant c_5\sqrt{nd}$ with probability at least $1 - O(n^{-2})$. And $(w^t)_i$ is the $d \times n$ independent Gaussian matrix. Using [32, Corollary 5.35], $\|(w^t)_i\| \leqslant c_6\sigma_2\sqrt{n}$ with probability at least $1 - O(n^{-2})$. Finally, since $E = E_D - w^t$, setting $c = c_4 + c_5 + c_6$ and applying a union bound yields the desired result. □

Lemma 8.4. Let $M_{1i} \in \mathbb{R}^{1\times d}$ and $M_{2i} \in \mathbb{R}^{d\times d}$ for $i = 1, \ldots, n$. Define the concatenated matrices $M_1^\top = [M_{11}^\top, M_{12}^\top, \ldots, M_{1n}^\top] \in \mathbb{R}^{d\times n}$ and $M_2^\top = [M_{21}^\top, M_{22}^\top, \ldots, M_{2n}^\top] \in \mathbb{R}^{d\times nd}$. Assume that M_1 is independent of E_i, and M_2 is independent of $(E^\top)_i$. Then, there exists an absolute constant $c > 0$ such that

$$\|E_i M_1\|_{\mathrm{F}} \leqslant c\sigma_2(\sqrt{d} + \sqrt{\log n})(\|M_1\|_{\mathrm{F}} + \sqrt{n}\|M_{1i}\|_{\mathrm{F}}), \quad (8.3)$$

$$\|(E^\top)_i M_2\|_2 \leqslant c\sigma_2(\sqrt{d} + \sqrt{\log n})(\|M_2\|_{\mathrm{F}} + \sqrt{n}\|M_{2i}\|_{\mathrm{F}}), \quad (8.4)$$

with probability at least $1 - O(n^{-3})$.

Proof. Recall that $(E_D)_i \in \mathbb{R}^{d\times n}$ is the i-the block row of $E_D \in \mathbb{R}^{nd\times n}$, we get $\|E_i M_1\|_{\mathrm{F}} \leqslant \|(E_D)_i M_1\|_{\mathrm{F}} + \|(w^t)_i M_1\|_{\mathrm{F}}$. To bound $\|w^t M_1\|_{\mathrm{F}}$, we consider the SVD $M_1 = U\Sigma V^\top$, where $U \in \mathbb{R}^{n\times d}$ with $U^\top U = I_d$, $\Sigma \in \mathbb{R}^{d\times d}$, and $V \in \mathbb{R}^{d\times d}$. Note that $(w^t)_i$ is a $d \times n$ Gaussian random matrix. Using $\|M_1\|_{\mathrm{F}} = \|\Sigma\|_{\mathrm{F}}$, we have:

$$\|(w^t)_i M_1\|_{\mathrm{F}} = \|(w^t)_i U\Sigma V^\top\|_{\mathrm{F}} \leqslant \|(w^t)_i U\|\|M_1\|_{\mathrm{F}}.$$

Since $U^\top U = I_d$, $(w^t)_i U$ is an asymmetric $d \times d$ Gaussian random matrix. [27, Theorem 5.4] guarantees that with probability at least $1 - O(n^{-3})$,

$$\|(w^t)_i M_1\|_{\mathrm{F}} \leqslant \|(w^t)_i U\|\|M_1\|_{\mathrm{F}} \leqslant c\sigma_2\|M_1\|_{\mathrm{F}}(\sqrt{d} + \sqrt{\log n}).$$

Moreover,

$$\|(E_D)_i M_1\|_{\mathrm{F}} \leqslant \|\sum_j w_{ij}^t M_{1i}\|_{\mathrm{F}} \leqslant c\sigma_2(\sqrt{nd}+\sqrt{n\log n})\|M_{1i}\|_2,$$

with probability at least $1 - O(n^{-3})$, since w_{ij}^t is Gaussian random vector. Combining these bounds yields inequality (8.3) with probability at least $1 - O(n^{-3})$. The proof for inequality (8.4) follows a similar argument and is therefore omitted. □

8.2.2. *Rotation error bound*

Proof of Lemma 3.2. We bound $\max_i \|\Phi_i - R_i^*\bar{Q}\|_{\mathrm{F}}$ similarly to [28]:

$$\begin{aligned}
&2n\|\Phi_i - R_i^*\bar{Q}\|_{\mathrm{F}} \leqslant (2n + \lambda_{\min}(R_i^*\Xi_{ii}^* R_i^{*\top}))\|\Phi_i - R_i^*\bar{Q}\|_{\mathrm{F}} \\
&\leqslant \|(2nI_d + R_i^*\Xi_{ii}^* R_i^{*\top})\Phi_i - (2nI_d + R_i^*\Xi_{ii}^* R_i^{*\top})R_i^*\bar{Q}\|_{\mathrm{F}} \\
&= \|2n\Phi_i + R_i^*\Xi_{ii}^* R_i^{*\top}\Phi_i + \Phi_i\Lambda - (H\Phi)_i - 2nR_i^*\bar{Q} \\
&\qquad - R_i^*\Xi_{ii}^*\bar{Q}\|_{\mathrm{F}} \\
&\leqslant \|2R^{*\top} + \Upsilon_i^* \operatorname{BlkDiag}(R_1^*, \ldots, R_n^*)^\top\|\|\Phi - R^*\bar{Q}\|_{\mathrm{F}} \\
&\quad + \|\Delta_i\Phi\|_{\mathrm{F}} + \|\Lambda\|\|\Phi_i\|_{\mathrm{F}} \\
&\leqslant 8(M_t^2 + 1)\sqrt{d}\|\Delta\| + \frac{4}{7n}\|\Delta_i\Phi\|_{\mathrm{F}}\|\Delta\| + \|\Delta_i\Phi\|_{\mathrm{F}},
\end{aligned}$$

where Λ is the $d \times d$ diagonal matrix whose diagonal entries are the d smallest eigenvalues of H, Ξ_{ii}^* denotes the i-th $d \times d$ block of the block-diagonal matrix Ξ^*, the equality is obtained by $\Phi_i\Lambda = (H\Phi)_i$, the third inequality follows from Lemma 8.1. To prove the last inequality, we note that by Lemmas 8.5-8.10 in proved Section 8.2.3, there exist constants $c_0, c_1, c_2, c_3 > 0$ such that if $(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2) \leqslant \frac{c_0\sqrt{n}}{\sqrt{d}+\sqrt{\log(n)}}$, then with probability at least $1 - O(n^{-2})$ the following inequalities hold simultaneously:

$$\|\Lambda\| \leqslant \|\Delta\|, \quad (8.5)$$

$$\|2R^{*\top} + \Upsilon_i^* \operatorname{BlkDiag}(R_1^*, \ldots, R_n^*)^\top\| \leqslant 2\sqrt{n}(M_t^2 + 1), \quad (8.6)$$

$$\|\Phi_i\|_{\mathrm{F}} \leqslant \frac{8(M_t^2 + 1)\sqrt{d}}{7} + \frac{4}{7n}\|\Delta_i\Phi\|_{\mathrm{F}}, \quad (8.7)$$

$$\|\Phi - R^*\bar{Q}\|_{\mathrm{F}} \leqslant \frac{c_0\sqrt{d}\|\Delta\|}{\sqrt{n}}, \quad (8.8)$$

$$\|\Delta\| \leqslant (c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2)(\sqrt{nd} + \sqrt{n\log n}), \quad (8.9)$$

$$\begin{aligned}&\|\Delta_i\Phi\|_{\mathrm{F}} \\ &\leqslant (M_t^2 + 1)(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2)(d\sqrt{n} + \sqrt{nd\log n}),\end{aligned} \quad (8.10)$$

$$\max_i \|\Phi_i\| \leqslant \max_i \|\Phi_i\|_{\mathrm{F}} \leqslant c_4(M_t^2 + 1)\sqrt{d}, \quad (8.11)$$

Finally, applying a union bound over $i = 1, \ldots, n$ for $\max_i \|\Phi_i - R_i^*\bar{Q}\|_{\mathrm{F}}$, we have:

$$\begin{aligned}&\max_{i\in[n]} \|\Phi_i - R_i^*\bar{Q}\|_{\mathrm{F}} \\ &\leq c_4(M_t^2 + 1)(c_1\sigma_1 + c_2 M_t\sigma_2 + c_3\sigma_2^2)\frac{(d + \sqrt{d\log n})}{\sqrt{n}}\end{aligned} \quad (8.12)$$

Combining inequality (8.12) and inequality (8.11) yields the desired result. □

8.2.3. *Analysis of Noise Matrix for* $\mathrm{SE}(d)$

The following six lemmas prove inequalities (8.5)-(8.11) used in the proof of Lemma 3.2 in Section 8.2.2.

Lemma 8.5 (Proof of inequality (8.5)). If $\|\Delta\| \leqslant n/4$, then the eigenvalues of the data matrix H are controlled by the noise matrix, i.e., $\|\Lambda\| \leqslant \|\Delta\|$ for all $1 \leqslant i \leqslant n$, where Λ is the $d \times d$ diagonal matrix whose diagonal entries are the d smallest eigenvalues of H.

Proof. By Weyl's theorem [33],

$$\max_i|\lambda_i(\Lambda) - \lambda_i(2nI_{nd} - 2R^*R^{*\top} + \Sigma^* - \frac{1}{2}T^*T^{*\top})|\leqslant \|\Delta\|.$$

Since the smallest d eigenvalues of $2nI_{nd} - 2R^*R^{*\top} + \Sigma^* - \frac{1}{2}T^*T^{*\top}$ are zero, we obtain $\max_i|\lambda_i(\Lambda) - 0|\leqslant \|\Delta\|$. □

Lemma 8.6 (Proof of inequality (8.6)). If $\max_i \|t_i^*\|_2 \leqslant M_t$, $\|2R^{*\top} + \Upsilon_i^* \operatorname{BlkDiag}(R_1^*, \ldots, R_n^*)^\top\| \leqslant 2\sqrt{n}(M_t^2 + 1)$.

Proof. From the definition of Υ_i^*,

$$\begin{aligned}\|\Upsilon_i^*\| &\leqslant \frac{1}{2}\max_i \|t_i^* t_i^{*\top}\|\sqrt{n} + \frac{1}{2}\|(t_i^* t_1^{*\top} \ldots t_i^* t_n^{*\top})\| \\ &\quad + \frac{1}{2}\|(t_1^* t_1^{*\top} \ldots t_n^* t_n^{*\top})\| + \frac{1}{2}\sqrt{n}\max_i \|t_i^* t_i^{*\top}\| \\ &\leqslant 2M_t^2\sqrt{n}.\end{aligned}$$

Therefore, $\|2R^{*\top} + \Upsilon_i^* \operatorname{BlkDiag}(R_1^*, \ldots, R_n^*)^\top\| \leqslant 2\|R^*\| + \|\Upsilon_i^*\| \leqslant 2\sqrt{n}(M_t^2 + 1)$. □

Lemma 8.7 (Proof of inequality (8.7)). If $\|\Delta\| \leqslant \frac{n}{4}$ and $\max_i \|t_i^*\|_2 \leqslant M_t$, then $\|\Phi_i\|_{\mathrm{F}} \leqslant \frac{8(M_t^2+1)\sqrt{d}}{7} + \frac{4}{7n}\|\Delta_i\Phi\|_{\mathrm{F}}$.

Proof. For $\|\Phi_i\|_{\mathrm{F}}$, we have

$$\begin{aligned}&(2n+\lambda_{\min}(R_i^*\Xi_{ii}^*R_i^{*\top}))\|\Phi_i\|_{\mathrm{F}}\\&\leqslant \|2n\Phi_i+R_i^*\Xi_{ii}^*R_i^{*\top}\Phi_i-(H\Phi)_i+\Phi_i\Lambda\|_{\mathrm{F}}\\&\leqslant \|2n\Phi_i+R_i^*\Xi_{ii}^*R_i^{*\top}\Phi_i-(H\Phi)_i\|_{\mathrm{F}}+\|\Phi_i\|_{\mathrm{F}}\|\Lambda\|.\end{aligned}$$

Using the inequality $\|\Lambda\|\leqslant\frac{n}{4}$ and $\lambda_{\min}(R_i^*\Xi_{ii}^*R_i^{*\top})\geqslant 0$ ($\Xi_{ii}^*=\frac{n}{2}t_i^*t_i^{*\top}+\sum_{j=1}^n t_j^*t_j^{*\top}$ are positive semi-definite), we obtain

$$\begin{aligned}&\frac{7n}{4}\|\Phi_i\|_{\mathrm{F}}\\&\leqslant \|2n\Phi_i+R_i^*\Xi_{ii}^*R_i^{*\top}\Phi_i-(H\Phi)_i\|_{\mathrm{F}}\\&\leqslant \|(2R^{*\top}+\Upsilon_i^*\,\mathrm{BlkDiag}(R_1^*,\ldots,R_n^*)^\top)\Phi\|_{\mathrm{F}}+\|\Delta_i\Phi\|_{\mathrm{F}}.\end{aligned}$$

By Lemma 8.6, we have $\|\Phi_i\|_{\mathrm{F}}\leqslant\frac{8(M_t^2+1)\sqrt{d}}{7}+\frac{4}{7n}\|\Delta_i\Phi\|_{\mathrm{F}}$. □

Lemma 8.8 (Proof of inequality (8.8)). $\|\Phi-R^*\bar{Q}\|_{\mathrm{F}}\leqslant\frac{c_0\sqrt{d}\|\Delta\|}{\sqrt{n}}$.

Proof. The smallest d eigenvalues of $2nI_{nd}-2R^*R^{*\top}+\Sigma^*-\frac{1}{2n}T^*T^{*\top}$ are 0. Each column of $\frac{1}{\sqrt{n}}R^*$ is an eigenvector. Moreover, $\|R^*\|_{\mathrm{F}}=\sqrt{nd}$. Furthermore, each column of $\frac{1}{\sqrt{n}}\Phi$ is a normalized eigenvector of H. Using the variant of the Davis-Kahan theorem [16, Theorem 5], we obtain $\frac{1}{\sqrt{n}}\|\Phi-R^*\bar{Q}\|_{\mathrm{F}}\leqslant\frac{c_0\sqrt{d}\|\Delta\|}{n}$. □

Lemmas 8.9 and 8.10 and Proposition 1 rely on a decomposition of the noise matrix Δ into three parts: $2W^R$, Δ_T, and Δ_Σ, where

$$\begin{aligned}&\Delta=2W^R+\Delta_\Sigma-\Delta_T,\\&\Delta_T:=\frac{1}{2n}(ET^{*\top}+T^*E^\top+EE^\top),\\&\Delta_\Sigma:=\mathrm{BlkDiag}(\sum_{j\neq 1}^n[s_{1j}^*w_{1j}^{t\top}+w_{1j}^ts_{1j}^{*\top}+(w_{1j}^tw_{1j}^{t\top}-\sigma_2^2I_d)],\ldots,\\&\qquad\sum_{j\neq n}^n[s_{nj}^*w_{nj}^{t\top}+w_{nj}^ts_{nj}^{*\top}+(w_{nj}^tw_{nj}^{t\top}-\sigma_2^2I_d)])\end{aligned}$$

Here $2W^R$ is the rotation noise in H, while Δ_T and Δ_Σ constitute the translation noise.

Lemma 8.9 (Proof of inequality (8.5)). There exist absolute positive constants c_1, c_2, c_3, c_4 such that, if $\sigma_2\leqslant\frac{M_t\sqrt{n}}{c_4(\sqrt{d}+\sqrt{\log n})}$, then,

$$\|\Delta\|\leqslant(c_1\sigma_1+c_2M_t\sigma_2+c_3\sigma_2^2)(\sqrt{nd}+\sqrt{n\log n}).$$

with probability at least $1-O(n^{-2})$.

Proof. Note that $\|\Delta\|\leqslant\|2W^R\|+\|\Delta_T\|+\|\Delta_\Sigma\|$. Since $2W^R$ is a standard Gaussian matrix, $\|2W^R\|\leqslant c_1\sigma_1\sqrt{nd}$ with probability at least $1-O(n^{-2})$. For Δ_T, we have

$$\begin{aligned}\|\Delta_T\|&=\|\frac{1}{2n}ET^{*\top}+\frac{1}{2n}T^*E^\top+\frac{1}{2n}EE^\top\|\\&\leqslant\frac{1}{2n}\|ET^{*\top}\|+\frac{1}{2n}\|T^*E^\top\|+\frac{1}{2n}\|EE^\top\|.\end{aligned}\tag{8.13}$$

By Lemma 8.2 and Lemma 8.3,

$$\begin{aligned}&\frac{1}{2n}\|ET^{*\top}\|=\frac{1}{2n}\|T^*E^\top\|\\&\leqslant\frac{1}{2n}\|T^*\|\|E^\top\|\leqslant\frac{1}{4}c_2M_t\sigma_2(\sqrt{nd}+\sqrt{n\log n}),\end{aligned}\tag{8.14}$$

with probability at least $1-O(n^{-2})$. For the higher-order term, using the condition $\sigma_2\leqslant\frac{M_t\sqrt{n}}{c_4(\sqrt{d}+\sqrt{\log n})}$,

$$\begin{aligned}\frac{1}{2n}\|EE^\top\|&\leqslant\frac{1}{2n}\|E\|\|E^\top\|\leqslant c_2^2\sigma_2^2(\sqrt{d}+\sqrt{\log n})^2\\&\leqslant\frac{c_2}{4}M_t\sigma_2^2(\sqrt{nd}+\sqrt{n\log n}),\end{aligned}\tag{8.15}$$

with probability at least $1-O(n^{-2})$. For Δ_Σ,

$$\begin{aligned}&\|\Delta_\Sigma\|\\&\leqslant 2\max_i\|\sum_{j\neq i}^n w_{ij}^ts_{ij}^{*\top}\|+\max_i\|\sum_{j\neq i}^n w_{ij}^tw_{ij}^{t\top}-\sigma_2^2(n-1)I_d\|.\end{aligned}$$

For $\|\sum_{j\neq i}^n w_{ij}^ts_{ij}^{*\top}\|$ where i is fixed, we apply Lemma 8.4 to get $\|\sum_{j\neq i}^n w_{ij}^ts_{ij}^{*\top}\|\leqslant\frac{c_2}{4}M_t\sigma_2(\sqrt{nd}+\sqrt{n\log n})$ with probability at least $1-O(n^{-3})$, since $\|(s_{i1}^*,s_{i2}^*,\ldots,s_{in}^*)^\top\|_2\leqslant 2\sqrt{n-1}M_t$.

For term $\|\sum_{j\neq i}^n w_{ij}^tw_{ij}^{t\top}-\sigma_2^2(n-1)I_d\|$ with fixed i, let A be the $(n-1)\times d$ matrix whose rows are $\sigma_2^{-1}w_{ij}^{t\top}$ (each $\sigma_2^{-1}w_{ij}^t$ is a column vector in $\mathbb{R}^d$, so $\sigma_2^{-1}w_{ij}^{t\top}$ is a row vector). Then $S:=A^\top A=\sum_{j\neq i}\sigma_2^{-2}w_{ij}^tw_{ij}^{t\top}$. The second moment matrix for each row $\sigma_2^{-1}w_{ij}^{t\top}$ is $\Sigma=\sigma_2^{-2}\mathbb{E}[w_{ij}^tw_{ij}^{t\top}]=I_d$. The sub-gaussian norm of $\sigma_2^{-1}w_{ij}^t$ is less than an absolute constant c (since $\sigma_2^{-1}w_{ij}^t$ is standard Gaussian). Applying [32, Remark 5.40], we obtain, with probability at least $1-2\exp(-ct^2)$,

$$\|S-(n-1)I_d\|\leqslant(n-1)\max(\delta,\delta^2)\sigma_2^2,$$

where $\delta=c_3\left(\sqrt{\frac{d}{n-1}}+\frac{t}{\sqrt{n-1}}\right)$. Setting $t=c'\sqrt{\log n}$ and assuming $n\geqslant d$, we obtain

$$\left\|\sum_{j\neq i}w_{ij}^tw_{ij}^{t\top}-\sigma_2^2(n-1)I_d\right\|\leqslant c_3\sigma_2^2\sqrt{n}(\sqrt{d}+\sqrt{\log n}),$$

with probability at least $1-O(n^{-3})$. Therefore, by a union bound,

$$\|\Delta_\Sigma\|\leqslant(\frac{c_2}{4}M_t\sigma_2+c_3\sigma_2^2)\sqrt{n}(\sqrt{d}+\sqrt{\log n}),\tag{8.16}$$

with probability at least $1-O(n^{-2})$. Combining inequalities (8.13)-(8.16) yields the desired result. □

Proposition 1. Let $M\in\mathbb{R}^{nd\times d}$ be a matrix independent of Δ_{Ri}, E_i, and $(E^\top)_i$. Assume $\sigma_2\leqslant\frac{M_t\sqrt{n}}{c_4(\sqrt{d}+\sqrt{\log n})}$. Then, for a fixed i, there exist absolute constants c_1,c_2,c_3,c_4 such that,

$$\begin{aligned}&\|\Delta_iM\|_{\mathrm{F}}\\&\leqslant c_1\sigma_1(\|M\|_{\mathrm{F}}+\sqrt{n}\|M_i\|_{\mathrm{F}})(\sqrt{d}+\sqrt{\log n})\\&\quad+(c_2M_t\sigma_2+c_3\sigma_2^2)(\|M\|_{\mathrm{F}}+\sqrt{n}\|M_i\|_{\mathrm{F}})(\sqrt{d}+\sqrt{\log n}),\end{aligned}$$

with probability at least $1-O(n^{-2})$.

Proof. We decompose the norm as $\|\Delta_iM\|_{\mathrm{F}}\leqslant 2\|W_i^RM\|_{\mathrm{F}}+\|\Delta_{Ti}M\|_{\mathrm{F}}+\|\Delta_{\Sigma i}M\|_{\mathrm{F}}$. For the first term, since W_i^R is a standard Gaussian block row, Lemma 5.8 of [27] gives

$$\|W_i^RM\|_{\mathrm{F}}\leqslant\sqrt{d}\|W_i^RM\|\leqslant c_1\sigma_1\|M\|(d+\sqrt{d\log n}),$$

with probability at least $1-O(n^{-2})$. We now bound $\|(\Delta_T)_i M\|_{\mathrm{F}}$. By definition,

$$\begin{aligned}\|(\Delta_T)_i M\|_{\mathrm{F}} &= \frac{1}{2n}\|(ET^{*\top})_i M + (T^* E^\top)_i M + (EE^\top)_i M\|_{\mathrm{F}} \\ &\leqslant \frac{1}{2n}(\|(ET^{*\top})_i M\|_{\mathrm{F}} + \|(T^* E^\top)_i M\|_{\mathrm{F}} + \|(EE^\top)_i M\|_{\mathrm{F}}).\end{aligned} \tag{8.17}$$

The term $\|(T^* E^\top)_i M\|_{\mathrm{F}}$ satisfies

$$\frac{1}{2n}\|T_i^* E^\top M\|_{\mathrm{F}} \leqslant \frac{1}{2n}\|(\sum_j s_{ij}^*)(E^\top)_i M\|_{\mathrm{F}} + \frac{1}{2n}\|s_i^* E M\|_{\mathrm{F}},$$

Using Lemma 8.2 and Lemma 8.3,

$$\begin{aligned}\frac{1}{2n}\|s_i^* E^\top M\|_{\mathrm{F}} &\leqslant \frac{1}{2n}\|s_i^*\|\|E^\top\|\|M\|_{\mathrm{F}} \\ &\leqslant \frac{c_2}{4}\sigma_2 M_t(\sqrt{d}+\sqrt{\log n})\|M\|_{\mathrm{F}},\end{aligned}$$

with probability $1-O(n^{-2})$. Since M is independent of $(E^\top)_i$, Lemma 8.4 yields

$$\begin{aligned}&\frac{1}{2n}\|(\sum_j s_{ij}^*)(E^\top)_i M\|_{\mathrm{F}} \\ &\leqslant \frac{c_2}{4}\sigma_2 M_t(\|M\|_{\mathrm{F}} + \sqrt{n}\|M_i\|_{\mathrm{F}})(\sqrt{d}+\sqrt{\log n}),\end{aligned}$$

with probability at least $1-O(n^{-3})$. Applying the bound for $\|E_i M\|_{\mathrm{F}}$ from Lemma 8.4 to $\|(ET^{*\top})_i M\|_{\mathrm{F}}$, we obtain

$$\begin{aligned}&\frac{1}{2n}\|E_i T^{*\top} M\|_{\mathrm{F}} \\ &\leqslant \frac{c_2\sigma_2(\sqrt{d}+\sqrt{\log n})}{4n}\|T^{*\top} M\|_{\mathrm{F}} \\ &+ \frac{c_2\sigma_2(\sqrt{d}+\sqrt{\log n})}{4\sqrt{n}}(\|(T_D^{*\top})_i M\|_{\mathrm{F}} + \|(s^{*\top})_i M\|_{\mathrm{F}}) \\ &\leqslant \frac{c_2\sigma_2(\sqrt{d}+\sqrt{\log n})}{4n}\|T^{*\top}\|\|M\|_{\mathrm{F}} \\ &+ \frac{c_2\sigma_2(\sqrt{d}+\sqrt{\log n})}{4\sqrt{n}}(\|\sum_j s_{ij}^*\|_2\|M_i\|_{\mathrm{F}} + \|(s^{*\top})_i\|_2\|M\|_{\mathrm{F}}) \\ &\leqslant \frac{c_2}{4}\sigma_2 M_t(\sqrt{d}+\sqrt{\log n})(\|M\|_{\mathrm{F}} + \sqrt{n}\|M_i\|_{\mathrm{F}}),\end{aligned}$$

with probability $1-O(n^{-3})$. The second inequality uses $T^* = T_D^* - s^*$, and the third uses $(T_D^{*\top})_i M = (\sum_j s_{ij}^{*\top})M_i$, where $T_D^{*\top}$ is block diagonal matrix. For $\frac{1}{2n}\|(EE^\top)_i M\|_{\mathrm{F}}$, using the condition $\sigma_2 \leqslant \frac{M_t\sqrt{n}}{c_4(\sqrt{d}+\sqrt{\log n})}$,

$$\begin{aligned}\frac{1}{2n}\|(EE^\top)_i M\|_{\mathrm{F}} &\leqslant \frac{1}{2n}\|E_i\|\|E^\top\|\|M\|_{\mathrm{F}} \\ &\leqslant \frac{c_2}{4}M_t\sigma_2(\sqrt{d}+\sqrt{\log n})\|M\|_{\mathrm{F}},\end{aligned}$$

with probability $1-O(n^{-2})$. Finally, for $\|(\Delta_\Sigma)_i M\|_{\mathrm{F}}$,

$$\begin{aligned}\|(\Delta_\Sigma)_i M\|_{\mathrm{F}} &\leqslant \|\Delta_{\Sigma i}\|\|M_i\|_{\mathrm{F}} \\ &\leqslant \sigma_2(\frac{c_2}{4}M_t + c_3\sigma_2)(\sqrt{nd}+\sqrt{n\log n})\|M_i\|_{\mathrm{F}},\end{aligned}$$

with probability $1-O(n^{-3})$. As a result, we obtain

$$\begin{aligned}&\|\Delta_i M\|_{\mathrm{F}} \\ &\leqslant c_1\sigma_1(\|M\|_{\mathrm{F}} + \sqrt{n}\|M_i\|_{\mathrm{F}})(\sqrt{d}+\sqrt{\log n}) \\ &+ (c_2 M_t\sigma_2 + c_3\sigma_2^2)(\|M\|_{\mathrm{F}} + \sqrt{n}\|M_i\|_{\mathrm{F}})(\sqrt{d}+\sqrt{\log n}).\end{aligned}$$

Collecting the bounds for W_i^R, $(\Delta_T)_i$, and $(\Delta_\Sigma)_i$, and adjusting the constants c_1, c_2, c_3, we obtain the desired inequality. ☐

Lemma 8.10 (Proof of inequalities (8.10) and (8.11)). Let $\Phi^{(i)}$ be the matrix whose columns are the d smallest eigenvectors of the data matrix $H^{(i)} := 2nI_{nd} - R^*R^{*\top} + \Sigma^* - \frac{1}{2n}T^*T^{*\top} + \Delta^{(i)}$, where $\Delta^{(i)}$ is obtained from Δ by removing the i-th block row and i-th block column of $2W^R$, the i-th block row and i-th column of E, and the i-th diagonal block of Δ_Σ. Assume $(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2) \leqslant \dfrac{c_0\sqrt{n}}{\sqrt{d}+\sqrt{\log n}}$ for some absolute positive constants c_0, c_1, c_2, c_3, c_4 such that

$$\begin{aligned}&\max_i \|\Delta_i\Phi\|_{\mathrm{F}} \\ &\leqslant (M_t^2+1)(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2)(d\sqrt{n}+\sqrt{nd\log n}),\end{aligned} \tag{8.18}$$

$$\max_i \|\Phi_i\|_{\mathrm{F}} \leqslant c_4(M_t^2+1)\sqrt{d}, \tag{8.19}$$

with probability at least $1-O(n^{-2})$.

Proof. We employ a leave-one-out argument. Note that the quantity $\|\Delta_i\Phi\|_{\mathrm{F}}$ is invariant under the orthogonal group $\mathrm{O}(d)$. Define

$$S^{(i)} := \arg\min_{S\in\mathrm{O}(d)} \|\Phi - \Phi^{(i)}S\|_{\mathrm{F}}, S^{(i)} = \Pi_{\mathrm{O}(d)}((\Phi^{(i)})^\top\Phi).$$

We decompose $\|\Delta_i\Phi\|$ into three terms and find an upper bound for each of them:

$$\begin{aligned}\|\Delta_i\Phi\|_{\mathrm{F}} &\leqslant \|\Delta_i(\Phi - \Phi^{(i)}S^{(i)} + \Phi^{(i)}S^{(i)})\|_{\mathrm{F}} \\ &\leqslant \|\Delta_i\|\|\Phi - \Phi^{(i)}S^{(i)}\|_{\mathrm{F}} + \|\Delta_i\Phi^{(i)}\|_{\mathrm{F}}.\end{aligned} \tag{8.20}$$

The term $\|\Delta_i\Phi^{(i)}\|_{\mathrm{F}}$ can be bounded by using Proposition 1. We now bound $\|\Phi - \Phi^{(i)}S^{(i)}\|_{\mathrm{F}}$ using Davis–Kahan theorem. The d-th smallest eigenvalue of H is at most $\|\Delta\|$, and the $(d+1)$-th smallest eigenvalue of H is at least $2n - \|\Delta\|$. Hence,

$$\begin{aligned}\|\Phi - \Phi^{(i)}S^{(i)}\|_{\mathrm{F}} &\leqslant \frac{\sqrt{2}\|(\Delta - \Delta^{(i)})\Phi^{(i)}\|_{\mathrm{F}}}{2n - 3\|\Delta\|} \\ &\leqslant \frac{4\sqrt{2}\|(\Delta - \Delta^{(i)})\Phi^{(i)}\|_{\mathrm{F}}}{5n},\end{aligned}$$

provided $\|\Delta\| \leqslant n/4$. Let $\Phi_i^{(i)}$ be the i-th $d\times d$ block of $\Phi^{(i)}$. Then

$$\begin{aligned}
&\|(\Delta-\Delta^{(i)})\Phi^{(i)}\|_{\mathrm{F}} \\
&\leqslant 2\|W_i^R\|\|\Phi_i^{(i)}\|_{\mathrm{F}} + 2\|W_i^R\Phi^{(i)}\|_{\mathrm{F}} + \|\Delta_{\Sigma i}\|\|\Phi_i^{(i)}\|_{\mathrm{F}} \\
&+ \frac{1}{2n}\|E_iT^{*\top}\Phi^{(i)}\|_{\mathrm{F}} + \frac{1}{2n}\|(E^\top)_i\|_2\|T^*\|\|\Phi_i^{(i)}\|_{\mathrm{F}} \\
&+ \frac{1}{2n}\|T^*\|\|(E_D)_i\|\|\Phi_i^{(i)}\|_{\mathrm{F}} + \frac{1}{2n}\|(E_D)_i\|\|T^*\|\|\Phi_i^{(i)}\|_{\mathrm{F}} \\
&+ \frac{1}{2n}\|T^*\|\|(E^\top)_i\Phi^{(i)}\|_{\mathrm{F}} + \frac{1}{2n}\|T^*\|\|E_i\|\|\Phi_i^{(i)}\|_{\mathrm{F}} \\
&+ \frac{1}{2n}\|E\|\|E_i\Phi^{(i)}\|_{\mathrm{F}} + \frac{1}{2n}\|E\|\|E_i\|\|\Phi_i^{(i)}\|_{\mathrm{F}} \\
&+ \frac{1}{2n}\|E\|\|(E_D)_i\|\|\Phi_i^{(i)}\|_{\mathrm{F}} + \frac{1}{2n}\|(E_D)_i\|\|E\|\|\Phi_i^{(i)}\|_{\mathrm{F}} \\
&+ \frac{1}{2n}\|E_iE^{(i)}\Phi^{(i)}\|_{\mathrm{F}} + \frac{1}{2n}\|(E^\top)_i\|_2\|E^{(i)}\|\|\Phi_i^{(i)}\|_{\mathrm{F}} \\
&\leqslant (c_1\sigma_1 + c_2\sigma_2M_t + c_3\sigma_2^2)(\sqrt{d}+\sqrt{\log n})\|\Phi\|_{\mathrm{F}} \\
&+ (c_1\sigma_1 + c_2\sigma_2M_t + c_3\sigma_2^2)(\sqrt{d}+\sqrt{\log n})\sqrt{n}\|\Phi_i^{(i)}\|_{\mathrm{F}}),
\end{aligned}$$

with probability at least $1-O(n^{-2})$. Here $E^{(i)}$ is obtained from E by deleting its i-th block row and i-th column. Moreover, $(E^\top)_i$ denotes the i-th row of $E^\top \in \mathbb{R}^{n\times nd}$. And then,

$$\begin{aligned}
&\|\Phi_i^{(i)}\|_{\mathrm{F}} - \|\Phi_i\|_{\mathrm{F}} &(8.21)\\
&\leqslant \|\Phi_i - \Phi_i^{(i)}S^{(i)}\|_{\mathrm{F}} \leqslant \|\Phi - \Phi^{(i)}S^{(i)}\|_{\mathrm{F}} \\
&\leqslant (c_1\sigma_1 + c_2\sigma_2M_t + c_3\sigma_2^2)\frac{(\sqrt{d}+\sqrt{\log n})}{\sqrt{n}}(\|\Phi_i^{(i)}\|_{\mathrm{F}} + \sqrt{d}). &(8.22)
\end{aligned}$$

Take $c_0 \le 1/4$. Since $(c_1\sigma_1 + c_2\sigma_2M_t + c_3\sigma_2^2)\frac{\sqrt{d}+\sqrt{\log n}}{\sqrt{n}} \leqslant c_0 \leqslant \frac{1}{4}$ and $\max_i \|\Phi_i\|_{\mathrm{F}} \geqslant \sqrt{d}$, we obtain

$$\begin{aligned}
&\|\Phi_i^{(i)}\|_{\mathrm{F}} - \|\Phi_i\|_{\mathrm{F}} \leqslant \|\Phi_i - \Phi_i^{(i)}S^{(i)}\|_{\mathrm{F}} \\
&\leqslant \frac{1}{4}(\|\Phi_i^{(i)}\|_{\mathrm{F}} + \sqrt{d}) \leqslant \frac{1}{4}(\|\Phi_i^{(i)}\|_{\mathrm{F}} + \max_j \|\Phi_j\|_{\mathrm{F}}). \qquad (8.23)
\end{aligned}$$

Hence, $\|\Phi_i^{(i)}\|_{\mathrm{F}} \leqslant 2\max_j \|\Phi_j\|_{\mathrm{F}}$. Combining inequalities (8.20) and (8.23), and assuming $\|\Delta\| \le n/4$,

$$\begin{aligned}
&\max_i \|\Delta_i\Phi\|_{\mathrm{F}} \leqslant \frac{3}{4}\max_i\|\Delta_i\|\max_i\|\Phi_i\|_{\mathrm{F}} + \max_i\|\Delta_i\Phi^{(i)}\|_{\mathrm{F}} \\
&\leqslant \frac{3n}{16}\max_i\|\Phi_i\|_{\mathrm{F}} + \max_i\|\Delta_i\Phi^{(i)}\|_{\mathrm{F}}. \qquad (8.24)
\end{aligned}$$

Combining Lemma 8.7, inequality (8.24), Proposition 1, and the assumption $(c_1\sigma_1 + c_2\sigma_2M_t + c_3\sigma_2^2)\frac{\sqrt{d}+\sqrt{\log n}}{\sqrt{n}} \leqslant c_0 \leqslant \frac{1}{4}$,

$$\begin{aligned}
&\max_i\|\Phi_i\|_{\mathrm{F}} \leqslant \frac{8(M_t^2+1)\sqrt{d}}{7} + \frac{4}{7n}\|\Delta_i\Phi\|_{\mathrm{F}} \\
&\leqslant 2(M_t^2+1)\sqrt{d} + \frac{3}{28}\max_i\|\Phi_i\|_{\mathrm{F}} + \frac{4}{7n}\max_i\|\Delta_i\Phi^{(i)}\|_{\mathrm{F}} \\
&\leqslant 3(M_t^2+1)\sqrt{d} + \frac{3}{28}\max_i\|\Phi_i\|_{\mathrm{F}} + \frac{1}{7}\max_i\|\Phi_i^{(i)}\|_{\mathrm{F}} \\
&\leqslant 3(M_t^2+1)\sqrt{d} + \frac{3}{28}\max_i\|\Phi_i\|_{\mathrm{F}} + \frac{2}{7}\max_i\|\Phi_i\|_{\mathrm{F}}.
\end{aligned}$$

Therefore, $\max_i\|\Phi_i\|_{\mathrm{F}} \leqslant c_4(M_t^2+1)\sqrt{d}$. Finally, using inequality (8.20), Proposition 1, and Lemma 8.9, $\|\Delta_i\Phi\|_{\mathrm{F}} \leqslant (M_t^2+1)(c_1\sigma_1 + c_2\sigma_2M_t + c_3\sigma_2^2)(d\sqrt{n}+\sqrt{nd\log n})$ with probability at least $1-O(n^{-2})$. ☐

8.3. Proof of Lemma 3.3

Before we prove Lemma 3.3, we need Lemma 8.11, which is a simple corollary of [16, Lemma 2].

Lemma 8.11. $\|\Pi^n_{\mathrm{SO}(d)}(Y) - X\|_F \le 2\|Y-X\|_{\mathrm{F}}$ for $Y \in \mathbb{R}^{nd\times d}$ and $X \in \mathbb{R}^{nd\times d}$ where each block $X_i \in \mathrm{SO}(d)$.

Proof. Let $B := \Pi^n_{\mathrm{SO}(d)}(Y) - X$, we have

$$\begin{aligned}
\|\Pi^n_{\mathrm{SO}(d)}(Y) - X\|_F^2 &= \mathrm{Tr}(B^\top B) = \sum_{i=1}^n(\mathrm{Tr}(B_i^\top B_i)) \\
&= \sum_{i=1}^n\|B_i\|_{\mathrm{F}}^2 \leqslant \sum_{i=1}^n 4\|Y_i - X_i\|_{\mathrm{F}}^2 = 4\|Y-X\|_{\mathrm{F}}^2.
\end{aligned}$$

where the inequality is due to [16, Lemma 2]. ☐

Here Lemma 8.12 is an operator norm error bound for $\|\Phi - R^*\bar{Q}\|$:

Lemma 8.12. There exist absolute constants $c_1, c_2, c_3, c_4 > 0$ such that if $\sigma_2 \leqslant \frac{M_t\sqrt{n}}{c_4(\sqrt{d}+\sqrt{\log n})}$, then $\|\Phi - R^*\bar{Q}\| \leqslant c_0(c_1\sigma_1 + c_2M_t\sigma_2 + c_3\sigma_2^2)(\sqrt{d}+\sqrt{\log n})$, with probability at least $1-O(n^{-2})$.

Proof. The smallest d eigenvalues of $2nI_{nd} - 2R^*R^{*\top} + \Sigma^* - \frac{1}{2n}T^*T^{*\top}$ are 0. Each column of $\frac{1}{\sqrt{n}}R^*$ is an eigenvector of $2nI_{nd} - 2R^*R^{*\top} + \Sigma^* - \frac{1}{2n}T^*T^{*\top}$. Moreover, $\|R^*\| = \sqrt{n}$. Furthermore, each column of $\frac{1}{\sqrt{n}}\Phi$ is a normalized eigenvector of H. We next apply the Davis-Kahan theorem [27, Theorem 6.2] and [27, Lemma 6.3]. Specifically, we flip the sign of the ground truth matrix $2nI_{nd} - 2R^*R^{*\top} + \Sigma^* - \frac{1}{2n}T^*T^{*\top}$ and the data matrix H, and then analyze the leading d eigenvectors of both matrices using the theorem and the Lemma. This results in

$$\begin{aligned}
\frac{1}{\sqrt{n}}\|\Phi - R^*\bar{Q}\| &\leqslant 2\frac{1}{\sqrt{n}}\|(I_n - \frac{1}{n}\Phi\Phi^\top)R^*\| \\
&\leqslant \frac{c_0\|\Delta\frac{1}{\sqrt{n}}R^*\|}{n} \leqslant \frac{c_0\|\Delta\|}{n}. \qquad (8.25)
\end{aligned}$$

According to the estimation in Lemma 8.9, we have $\|\Delta\| \leqslant (c_1\sigma_1 + c_2M_t\sigma_2 + c_3\sigma_2^2)(\sqrt{nd}+\sqrt{n\log n})$. Substituting this into (8.25) completes the proof. ☐

Proof of Lemma 3.3. From the structure of $\hat{T}$, for any fixed i,

$$\begin{aligned}
&\|\hat{t}_i - (\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top))^\top t_i^*\|_2 \\
&= \frac{1}{2n}\|(T^{*\top})_iR^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - (\hat{T}^\top)_i\Pi^n_{\mathrm{SO}(d)}(\Phi\Phi_1^\top)\|_2 \\
&\leqslant \frac{1}{2n}\|nt_i^{*\top}R_i^{*\top}R_i^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - (nt_i^{*\top}R_i^{*\top})\Pi^n_{\mathrm{SO}(d)}(\Phi\Phi_1^\top)_i\|_2 \\
&+ \frac{1}{2n}\|(s^{*\top})_iR^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - (s^{*\top})_i\Pi^n_{\mathrm{SO}(d)}(\Phi\Phi_1^\top)\|_2 \\
&+ \frac{1}{2n}\|(E_D)_i\Pi^n_{\mathrm{SO}(d)}(\Phi\Phi_1^\top)\|_2 + \frac{1}{2n}\|((w^t)^\top)_i\Pi^n_{\mathrm{SO}(d)}(\Phi\Phi_1^\top)\|_2.
\end{aligned}$$

The above inequality uses the decomposition $(\hat{T}^\top)_i = (\hat{T}_D^\top)_i - (s^\top)_i = (T_D^{*\top})_i + (E_D)_i - (s^{*\top})_i + ((w^t)^\top)_i$ and the triangle inequality. Moreover, we have

$$
\begin{aligned}
&\|\hat{t}_i - (\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top))^\top t_i^*\|_2 \\
&\leqslant M_t \underbrace{\|R_i^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - \Phi_i\Phi_1^\top\|_{\mathrm{F}}}_{①} + \underbrace{\frac{\sqrt{d}}{2n}\|\sum_{j\neq i}^n w_{ij}^\top\|_2}_{②} \\
&+ \underbrace{\frac{1}{2n}\|(s^{*\top})_i\|_2}_{③}\underbrace{\|R^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - \Phi\Phi_1^\top\|_{\mathrm{F}}}_{④} \\
&+ \underbrace{\frac{1}{2n}\|((w^t)^\top)_i\Pi_{\mathrm{SO}(d)}^n(\Phi\Phi_1^\top)\|_2}_{⑤}, \qquad (8.26)
\end{aligned}
$$

where we use the fact $M_t = \max_{i\in[n]}\|t_i^*\|_2$, R_i^* and $\Pi_{\mathrm{SO}(d)}(\Phi\Phi_1^\top)_i$ orthogonal matrices, and Lemma 8.11 to obtain term ④. Therefore, we bound each term in (8.26) separately. For term ① in (8.26),

$$
\begin{aligned}
① &\leqslant \|R_i^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - R_i^*\bar{Q}\Phi_1^\top\|_{\mathrm{F}} + \|R_i^*\bar{Q}\Phi_1^\top - \Phi_i\Phi_1^\top\|_{\mathrm{F}} \\
&\leqslant \|\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - \bar{Q}\Phi_1^\top\|_{\mathrm{F}} + \|R_i^*\bar{Q} - \Phi_i\|_{\mathrm{F}}\|\Phi_1\| \\
&\leqslant 2\|\Phi_1 - R_1^*\bar{Q}\|_{\mathrm{F}} + \|\Phi_i - R_i^*\bar{Q}\|_{\mathrm{F}}\|\Phi_1\| \\
&\leqslant c_4(M_t^2+1)^2(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2)\frac{(d\sqrt{d} + d\sqrt{\log n})}{\sqrt{n}}, \qquad (8.27)
\end{aligned}
$$

with probability at least $1 - O(n^{-2})$. The third inequality in (8.27) follows from [16, Lemma 2], as in the proof of Lemma 3.1. The last inequality in (8.27) uses (8.12) from the proof of Lemma 3.2, as we have the same conditions of Lemma 3.2, and $\max_i\|\Phi_i\| \leqslant \max_i\|\Phi_i\|_{\mathrm{F}} = O((M_t^2+1)\sqrt{d})$ in Lemma 8.10.

$$
② = \frac{\sqrt{d}}{2n}\|\sum_{j\neq i}^n w_{ij}^\top\|_2 \leqslant c_5\sigma_2\frac{(d + \sqrt{d\log n})}{\sqrt{n}}, \qquad (8.28)
$$

with probability at least $1 - O(n^{-2})$ by simply using Lemma 8.3. And we also have

$$
③ = \frac{1}{n}\|(s^{*\top})_i\|_2 = \frac{1}{n}\sqrt{\sum_j s_{ij}^{*\top}s_{ij}^*} \leqslant \frac{c_6 M_t}{\sqrt{n}}. \qquad (8.29)
$$

The quantity ④ can be bound similarly to the term ①:

$$
\begin{aligned}
④ &\leqslant \|R^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - R^*\bar{Q}\Phi_1^\top\|_{\mathrm{F}} + \|R^*\bar{Q}\Phi_1^\top - \Phi\Phi_1^\top\|_{\mathrm{F}} \\
&\leqslant \sqrt{n}\|\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - \bar{Q}\Phi_1^\top\|_{\mathrm{F}} + \|R^*\bar{Q} - \Phi\|_{\mathrm{F}}\|\Phi_1\| \\
&\leqslant 2\sqrt{n}\|\Phi_1 - R_1^*\bar{Q}\|_{\mathrm{F}} + \|\Phi - R^*\bar{Q}\|_{\mathrm{F}}\|\Phi_1\| \\
&\leqslant (M_t^2+1)^2(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2)(d\sqrt{d} + d\sqrt{\log n}), \qquad (8.30)
\end{aligned}
$$

with probability at least $1 - O(n^{-2})$. The last inequality in (8.30) follows from Lemma 8.8 and inequality (8.12) in Lemma 3.2, and $\max_i\|\Phi_i\| \leqslant \max_i\|\Phi_i\|_{\mathrm{F}} = O((M_t^2+1)\sqrt{d})$ in Lemma 8.10.

For term ⑤, note that $((w^t)^\top)_i$ appears in Δ_i. Similarly to bounding the rotation error, we apply a leave-one-out argument and borrow definitions $\Phi^{(i)}$ and $S^{(i)}$ in Lemma 8.10. First, we rewrite ⑤ as follows:

$$
\begin{aligned}
⑤ = \frac{1}{2n}\|((w^t)^\top)_i[\Pi_{\mathrm{SO}(d)}^n(\Phi\Phi_1^\top) - \Phi\Phi_1^\top + \Phi\Phi_1^\top \\
- \Phi^{(i)}S^{(i)}\Phi_1^\top + \Phi^{(i)}S^{(i)}\Phi_1^\top]\|_2.
\end{aligned}
$$

By the triangle inequality, we obtain

$$
\begin{aligned}
⑤ &\leqslant \frac{1}{2n}\|((w^t)^\top)_i\|_2\|\Pi_{\mathrm{SO}(d)}^n(\Phi\Phi_1^\top) - \Phi\Phi_1^\top\| \\
&+ \frac{1}{2n}\|((w^t)^\top)_i\|_2\|\Phi\Phi_1^\top - \Phi^{(i)}S^{(i)}\Phi_1^\top\|_{\mathrm{F}} \\
&+ \frac{1}{2n}\|((w^t)^\top)_i\Phi^{(i)}S^{(i)}\Phi_1^\top\|_2. \qquad (8.31)
\end{aligned}
$$

By Lemma 8.10, we get $\max_i\|\Phi_i\| = O((M_t^2+1)\sqrt{d})$, which together with (8.31), yields

$$
\begin{aligned}
⑤ &\leqslant \frac{1}{2n}\|((w^t)^\top)_i\|_2\|\Pi_{\mathrm{SO}(d)}^n(\Phi\Phi_1^\top) - \Phi\Phi_1^\top\| \\
&+ c_8\frac{(M_t^2+1)\sqrt{d}}{n}\|((w^t)^\top)_i\|_2\|\Phi - \Phi^{(i)}S^{(i)}\|_{\mathrm{F}} \\
&+ c_8\frac{(M_t^2+1)\sqrt{d}}{n}\|((w^t)^\top)_i\Phi^{(i)}\|_2. \qquad (8.32)
\end{aligned}
$$

Then, the first term in inequality (8.32) can be bounded as follows:

$$
\begin{aligned}
&\frac{1}{2n}\|((w^t)^\top)_i\|_2\|\Pi_{\mathrm{SO}(d)}^n(\Phi\Phi_1^\top) - \Phi\Phi_1^\top\| \\
&\qquad\leqslant c_9\frac{\sqrt{d}}{\sqrt{n}}\|\Pi_{\mathrm{SO}(d)}^n(\Phi\Phi_1^\top) - \Phi\Phi_1^\top\| \\
&\qquad\leqslant c_9\frac{\sqrt{d}}{\sqrt{n}}\|R^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - \Phi\Phi_1^\top\|, \qquad (8.33)
\end{aligned}
$$

where we use $\|((w^t)^\top)_i\|_2 \leqslant c\sigma_2\sqrt{nd}$ (due to Lemma 8.3). The last term in (8.33) can be bounded as follows:

$$
\begin{aligned}
&\frac{\sqrt{d}}{\sqrt{n}}\|R^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - \Phi\Phi_1^\top\| \\
&\leqslant \frac{\sqrt{d}}{\sqrt{n}}\|R^*\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - R^*\bar{Q}\Phi_1^\top\| + \frac{\sqrt{d}}{\sqrt{n}}\|R^*\bar{Q}\Phi_1^\top - \Phi\Phi_1^\top\| \\
&\leqslant \sqrt{d}\|\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top) - \bar{Q}\Phi_1^\top\| + \frac{\sqrt{d}}{\sqrt{n}}\|R^*\bar{Q} - \Phi\|\|\Phi_1\| \\
&\leqslant \sqrt{d}\|\Phi_1 - R_1^*\bar{Q}\| + \frac{\sqrt{d}}{\sqrt{n}}\|\Phi - R^*\bar{Q}\|\|\Phi_1\| \\
&\leqslant (M_t^2+1)(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2)\frac{(d\sqrt{d} + d\sqrt{\log n})}{\sqrt{n}}, \qquad (8.34)
\end{aligned}
$$

where the last inequality in (8.34) follows from Lemma 8.12, inequality (8.12), and the fact $\max_i\|\Phi_i\| = O((M_t^2+1)\sqrt{d})$.

Meanwhile, using (8.22), Lemma 8.3, Lemma 8.10, and the assumption $(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2) \leqslant \frac{c_0\sqrt{n}}{\sqrt{d}+\sqrt{\log n}}$, we have $\|\Phi - \Phi^{(i)}S^{(i)}\|_{\mathrm{F}} \leqslant (M_t^2+1)(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2)\frac{(d+\sqrt{d\log n})}{\sqrt{n}}$ with probability at least $1 - O(n^{-2})$. Substituting this into (8.32) yields

$$
\begin{aligned}
&\frac{(M_t^2+1)\sqrt{d}}{n}\|((w^t)^\top)_i\|_2\|\Phi - \Phi^{(i)}S^{(i)}\|_{\mathrm{F}} \\
&\leqslant c_{10}(M_t^2+1)^2(c_1\sigma_1 + c_2\sigma_2 M_t + c_3\sigma_2^2)\frac{(d\sqrt{d} + d\sqrt{\log n})}{\sqrt{n}}, \qquad (8.35)
\end{aligned}
$$

where we used the fact that $\|((w^t)^\top)_i\|_2 \leqslant c\sigma_2\sqrt{nd} \leqslant \frac{n}{4}$ (due to Lemma 8.3 and the assumption that $\sigma_2 \leqslant c'\frac{\sqrt{n}}{\sqrt{d}}$ for some constant c).

Recall the definition of $\Phi^{(i)}$ in Lemma 8.10. It is independent of of the i-th block row and i-th column of $E := \mathrm{BlkDiag}(\sum_{j=1}^{n} w_{1j}^t, \ldots, \sum_{j=1}^{n} w_{nj}^t) - w^t$. Therefore, $((w^t)^\top)_i$ in the third term in (8.32) is independent of $\Phi^{(i)}$. Furthermore, following the same argument for bounding the third term in (8.32), we obtain

$$\frac{(M_t^2+1)\sqrt{d}}{2n}\|((w^t)^\top)_i\Phi^{(i)}\|_2 \leqslant c_{11}\sigma_2(M_t^2+1)^2\frac{(d\sqrt{d}+d\sqrt{\log n})}{\sqrt{n}}, \quad (8.36)$$

with probability at least $1-O(n^{-2})$. Finally, substituting (8.34), (8.35) and (8.36) into (8.32), we obtain

$$⑤ \leqslant c_{12}(M_t^2+1)^2(c_1\sigma_1+c_2\sigma_2 M_t+c_3\sigma_2^2)\frac{(d\sqrt{d}+d\sqrt{\log n})}{\sqrt{n}}. \quad (8.37)$$

Under the same assumption on the noise magnitudes and using the bounds for ①-⑤, i.e., (8.27), (8.28), (8.29), (8.30) and (8.37) into (8.26), for a fixed i, we have

$$\|\hat{t}_i - (\Pi_{\mathrm{SO}(d)}(\bar{Q}\Phi_1^\top))^\top t_i^*\|_2 \leqslant c_{13}M_t(M_t^2+1)^2(c_1\sigma_1+c_2\sigma_2 M_t+c_3\sigma_2^2)\frac{(d\sqrt{d}+d\sqrt{\log n})}{\sqrt{n}},$$

with probability at least $1-O(n^{-2})$. An application of the union bound over $i=1,\ldots,n$ completes the proof. □