A Robust Translation Synchronization Algorithm

Zihang He*
UT Austin
Austin, TX 78712
zhhe@utexas.edu

Hang Ruan*
Shanghai Jiaotong University
Shanghai, 200240
ruanhang1123@gmail.com

Qixing Huang
UT Austin
Austin, TX 78712
huangqx@utexas.edu

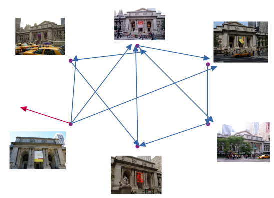
Abstract

This paper introduces a robust translation synchronization approach which takes relative directions between pairs of images as inputs and outputs absolute image locations. Our approach is based on a generalized eigenvalue formulation, which contains edge weights in relative directions and vertex weights in absolute image translations. We present a rigorous stability analysis to determine how to set these weights optimally. Specifically, optimal vertex weights are always identical, whereas optimal edge weights depend on the magnitudes of relative translations and variances of relative directions. These results lead to an iterative synchronization procedure, which progressively removes outliers in the inputs by adaptively adjusting the edge weights. Experimental results justify our theoretical results and show that our approach outperforms state-of-the-art baseline approaches on both synthetic and real datasets.

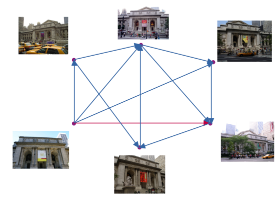
1. Introduction

This paper studies the classical pose synchronization problem of computing the absolute 6D poses of a collection of images from pairwise relative 6D poses along a graph of these images. A 6D pose has a rotation component and a translation component. Pose synchronization typically proceeds in two phases. The first phase performs rotation synchronization, which determines the absolute rotation of each image. The resulting rotations transform the relative translations expressed in local coordinate frames into relative translations in a global coordinate system. The second phase then determines the absolute translations of the input images.

As we cannot predict the absolute scale of the underlying objects, translation synchronization is much more challenging than rotation synchronization. Specifically, the relative translation associated with each pair of images is a direction that offers only two constraints. Therefore, even though a connected graph of image pairs is sufficient for rotation synchronization, translation synchronization needs more image pairs. Computationally, the fundamental challenge lies in jointly estimating the scales of relative translations and absolute translations. This challenge also makes it difficult to obtain closed-form solutions for translation synchronization.



(a) Input graph of relative directions. Purple points are the camera locations. The blue arrows are the input relative translations. The red arrow is an outlier relative prediction.



(b) Recovered global absolute translations. Purple points are the camera locations. Blue arrows are the predicted relative translations

Figure 1. (Left) Input graph of relative directions, in which the color of each edge indicates the directional error. (Right) Recovered global absolute translations, overlaid with the ground-truth absolute translations.

In this paper, we introduce a novel translation synchronization that admits a closed-form solution. Our approach solves a generalized eigenvalue problem. The formulation has a set of edge weights associated with the relative directions and a set of vertex weights associated with the input images. We present a rigorous stability analysis of this generalized eigenvalue formulation. We show that the optimal vertex weights are identical, whereas the optimal edge weights depend on the variances and magnitudes of relative translations. Based on these results, we then develop an iterative translation synchronization formulation that adaptively adjusts the edge weights, in which estimations of variances and magnitudes of relative translations are derived from current predicted absolute translations. We present theoretical results on the robust recovery conditions of our approach, showing the robustness of our approach against corrupted measurements.

We have evaluated our approach on both synthetic and real datasets. Experimental results on synthetic datasets justify our theoretical results. Experimental results on both synthetic and real datasets show that our approach outperforms state-of-the-art approaches. We also present an advanced analysis to show the effectiveness of our iterative scheme.

2. Related Work

We discuss related work in three categories, namely, map synchronization, translation synchronization, and spectral techniques.

2.1. Pose synchronization

Map synchronization, which generalizes pose synchronization, estimates consistent maps among a collection of objects from maps estimated between pairs of objects in isolation. Maps can take different forms, including transformations (the case of pose synchronization) [5, 28, 31], point-based maps [12, 17], and functional maps [13]. Map synchronization typically utilizes the generic constraint that composite maps along cycles should be equal to identify maps. In the case of point-based maps, Huang and Guibas [11] show the equivalence between the cycle-consistency constraint and the positive semidefiniteness of the matrix that encodes pairwise maps in blocks. In fact, many state-of-the-art map synchronization approaches solve low-rank matrix recovery problems to recover consistent maps. Optimization strategies include convex optimization [8, 11, 13, 24, 28, 31, 33], non-convex optimization [5, 7, 14, 18, 19, 34], and spectral techniques [1, 2, 4, 12, 16, 17, 22, 23, 26, 29, 30]. Our approach introduces a novel matrix encoding of the directional constraints for translation synchronization and falls into the categories of non-convex optimization and spectral techniques.

2.2. Translation synchronization

Translation synchronization requires optimizing both the scales of the directional constraints and the absolute translations. Because of this, few existing approaches admit closed-form solutions. Govindu [9] utilized the cross product and constructed a linear system of to estimate translations. Özyesil et al. [21] proposed iteratively reweighted least squares algorithm for the least squared deviations (LUD) solver in L^1 the norm of the relaxed displacement cost function. Hand et al. [10] proposed the ShapeFit algorithm by optimizing a convex program in the translation direction. Wilson proposed 1DSfm [32] to remove outliers by projecting from 3d to 1d and optimize the direction-based cost function on the remaining accurate matches. Zhuang et al. [35] analyzed LUD and ShapeFit and proposed the bilinear angle-based translation averaging algorithm (BATA) that optimizes a relaxed direction cost function. In particular, BATA shows that the loss based on the angle between the input direction and the optimized direction is superior to standard distance norms between directions. Our approach also uses angles to define the loss term but introduces a novel spectral formulation that admits a closed-form solution.

Combining displacement and direction cost functions, Lalit et al. [20] alternatively update translations by direction and displacement-based methods. In contrast, our approach is based on a spectral formulation that admits a closed-form solution. We analyze its stability to derive an iterative approach to robust translation synchronization.

2.3. Spectral techniques

Our approach is motivated by the stability of eigenvalues and eigenvectors of a symmetric matrix under perturbation. We refer to a survey article on this topic [6]. Compared to convex and non-convex optimization formulations, spectral techniques are simple and efficient. Unlike robust low-rank matrix recovery, in which outliers may be unbouned, outliers in pairwise maps have bounded norms. This fact improves the robustness of applying spectral techniques in map synchronization [3, 15, 22, 23, 25, 27]. Unlike most prior work that studies robust recovery under random and independent noise, this paper presents a robustness result under adversarial noise.

3. Problem Statement and Approach Overview

3.1. Problem Statement

Consider a connected graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$ of $n = |\mathcal{I}|$ image objects. Each edge $(i,j) \in \mathcal{E}$ is associated with a measurement $\boldsymbol{v}_{ij}^{\text{inp}}$ ($\|\boldsymbol{v}_{ij}^{\text{inp}}\| = 1$) of the normalized translation from I_i to I_j. Specifically, denote $\boldsymbol{t}_i^{gt}$ as the underlying ground-truth translation of I_i, we have

$$\boldsymbol{v}_{ij}^{\text{inp}} \approx \frac{\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}}{\|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|}. \quad (1)$$

In particular, we expect that for a subset of the edges $\{(i,j)\}$, $\boldsymbol{v}_{ij}^{\text{inp}}$ have large errors.

Translation synchronization aims at recovering $\boldsymbol{t}_i$ from $\boldsymbol{v}_{ij}^{\text{inp}}$ up to a global scale and a global translation.

3.2. Approach Overview

A fundamental challenge of translation synchronization is that (1) is non-linear. Our approach addresses this challenge using a generalized spectral formulation by introducing weights associated with edges $\mathcal{E}$ and vertices $\mathcal{I}$ (Section 4.1). When there is no input noise, this generalized formulation allows us to derive simple conditions on which the translation synchronization problem has a unique solution (Section 4.2). When the input has noise, the accuracy of the output depends on these weights. We present a rigorous stability analysis on the relations between output errors and these weights (Section 4.3). The results lead to an iterative approach that adaptively adjusts these weights to progressively eliminate outliers in the inputs and average inliers in the inputs (Section 4.4). Finally, we present a robust recovery condition for our approach under adversarial input noise (Section 4.5).

3.3. Notations

For each $(i, j) \in \mathcal{E}$, we define $\boldsymbol{e}_{ij} \in \mathbb{R}^n$ as the edge indicator function whose element $e_{ij}(k)$ satisfies $e_{ij}(i) = 1$, $e_{ij}(j) = -1$, and $e_{ij}(k) = 0$ otherwise.

We will also use several matrix norms. Given a block matrix $A \in \mathbb{R}^{3m\times 3n}$ and a block vector $\boldsymbol{v} \in \mathbb{R}^{3m}$, we define the L^1 norm of A and the L^∞ norm of $\boldsymbol{v}$ as

$$\|A\|_1 = \max_{1\leq i\leq m} \sum_{j=1}^{n} \|A_{ij}\|, \quad \|\boldsymbol{v}\|_\infty = \max_{1\leq i\leq m} \|\boldsymbol{v}_i\|. \quad (2)$$

Moreover, $\|A\|$ denotes the spectral norm of matrix A. $A^\dagger$ denotes the pseudo-inverse of A.

4. Approach

This section presents our technical approach.

4.1. Spectral Translation Synchronization

Let s_{ij} be the latent absolute scale parameter of edge $(i, j) \in \mathcal{E}$. We can write out the relation between $\boldsymbol{v}_{ij}$ and $\boldsymbol{t}_i$ and $\boldsymbol{t}_j$ as

$$s_{ij}\boldsymbol{v}_{ij}^{\mathrm{inp}} \approx (\boldsymbol{t}_i - \boldsymbol{t}_j). \quad (3)$$

When $\boldsymbol{t}_i$ and $\boldsymbol{t}_j$ are fixed, the optimal s_{ij} is given by

$$\begin{aligned} s_{ij} &= \underset{s}{\mathrm{argmin}} \quad \|s\boldsymbol{v}_{ij}^{\mathrm{inp}} - (\boldsymbol{t}_i - \boldsymbol{t}_j)\|^2 \\ &= \frac{\boldsymbol{v}_{ij}^{\mathrm{inp}^T}(\boldsymbol{t}_i - \boldsymbol{t}_j)}{\boldsymbol{v}_{ij}^{\mathrm{inp}^T}\boldsymbol{v}_{ij}^{\mathrm{inp}}} = \boldsymbol{v}_{ij}^{\mathrm{inp}^T}(\boldsymbol{t}_i - \boldsymbol{t}_j). \end{aligned} \quad (4)$$

Denote $\boldsymbol{t} = (\boldsymbol{t}_i) \in \mathbb{R}^{3n}$. Substituting Eq. (4) into Eq. (3) and introducing a weight $w_{ij} > 0$ for each edge $(i, j) \in \mathcal{E}$, we arrive at the following quadratic loss term among $\boldsymbol{t}$:

$$\begin{aligned} f(\boldsymbol{t}) &= \sum_{(i,j)\in\mathcal{E}} w_{ij}\|(I_3 - \boldsymbol{v}_{ij}^{\mathrm{inp}}\boldsymbol{v}_{ij}^{\mathrm{inp}^T})(\boldsymbol{t}_i - \boldsymbol{t}_j)\|^2 \\ &= \sum_{(i,j)\in\mathcal{E}} w_{ij}(\boldsymbol{t}_i - \boldsymbol{t}_j)^T(I_3 - \boldsymbol{v}_{ij}^{\mathrm{inp}}\boldsymbol{v}_{ij}^{\mathrm{inp}^T})(\boldsymbol{t}_i - \boldsymbol{t}_j). \end{aligned} \quad (5)$$

Remark 1. *Let θ_{ij} be the angle between $\boldsymbol{v}_{ij}^{\mathrm{inp}}$ and $\frac{\boldsymbol{t}_i-\boldsymbol{t}_j}{\|\boldsymbol{t}_i-\boldsymbol{t}_j\|}$. It is easy to see that*

$$\begin{aligned} &(\boldsymbol{t}_i - \boldsymbol{t}_j)^T(I_3 - \boldsymbol{v}_{ij}^{\mathrm{inp}}\boldsymbol{v}_{ij}^{\mathrm{inp}^T})(\boldsymbol{t}_i - \boldsymbol{t}_j) \\ =&\sin(\theta_{ij})^2\|\boldsymbol{t}_i - \boldsymbol{t}_j\|^2. \end{aligned}$$

Therefore, our formulation can be viewed as a variant of the angle-based loss proposed in [35].

Minimizing Eq. (5) directly has a trivial solution, in which $\boldsymbol{t}_i = 0$. We apply the standard approach of enforcing a quadratic normalization constraint on $\boldsymbol{t}_i$:

$$\sum_{i=1}^{n} s_i\|\boldsymbol{t}_i\|^2 = 1 \quad (6)$$

where $s_i > 0$ is the weight associated with image I_i.

Introduce a generalized connection Laplacian matrix $L \in \mathbb{R}^{3n\times 3n}$ whose blocks are given by

$$L_{ij} = \begin{cases} \sum\limits_{j\in\mathcal{N}_i} w_{ij}(I_3 - \boldsymbol{v}_{ij}^{\mathrm{inp}}\boldsymbol{v}_{ij}^{\mathrm{inp}^T}) & i = j \\ -w_{ij}(I_3 - \boldsymbol{v}_{ij}^{\mathrm{inp}}\boldsymbol{v}_{ij}^{\mathrm{inp}^T}) & (i, j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

It is clear that we can rewrite

$$L = \sum_{(i,j)\in\mathcal{E}} w_{ij}E_{ij}(I_3 - \boldsymbol{v}_{ij}^{\mathrm{inp}}\boldsymbol{v}_{ij}^{\mathrm{inp}^T})E_{ij}^T, \qquad E_{ij} = \boldsymbol{e}_{ij}\otimes I_3.$$

Denote $S = \mathrm{diag}(\{s_i\}) \otimes I_3$. We can then reformulate the optimization problem of minimizing Eq. (5) subject to the constraint Eq. (6) as

$$\min_{\boldsymbol{t}} \ \boldsymbol{t}^T L\boldsymbol{t} \qquad s.t. \quad \boldsymbol{t}^T S\boldsymbol{t} = 1 \quad (7)$$

It is easy to see that the optimal solution $\boldsymbol{t}^\star$ of Eq. (7) is given by the following generalized eigen-value problem:

$$L\boldsymbol{t}^\star = \lambda S\boldsymbol{t}^\star. \quad (8)$$

The following proposition describes some basic properties of L.

Proposition 1. *L is positive semidefinite. It has three zero eigenvalues. The corresponding eigenvectors are given by $\mathbf{1} \otimes I_3$.*

Proof: See Section B.1. □

Therefore, we look for $\boldsymbol{t}^\star$ as the eigenvector that corresponds to the fourth-smallest eigenvalue of Eq. (8) and satisfies $(\mathbf{1} \otimes I_3)^T \boldsymbol{t}^\star = 0$.

4.2. Uniqueness of Translation Synchronization

This section studies the uniqueness problem of translation synchronization, that is, under what conditions of $\boldsymbol{t}_i^{gt}$ and $\mathcal{E}$, one can recover $\boldsymbol{t}_i^{gt}$ from $\boldsymbol{v}_{ij}^{gt} = \frac{\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}}{\|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|}, \forall (i,j) \in \mathcal{E}$. Although there are a lot of translation synchronization approaches, this uniqueness problem is underexplored. We show a simple necessary and sufficient uniqueness condition using Eq. (8).

Without losing generality, we assume that $\boldsymbol{t}^{gt}$ is normalized so that

$$\sum_{i=1}^{n} \boldsymbol{t}_i^{gt} = 0, \qquad \sum_{i=1}^{n} s_i \|\boldsymbol{t}_i^{gt}\|^2 = 1.$$

Proposition 2. *Suppose that the input directions are exact, that is, $\boldsymbol{v}_{ij} = \frac{\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}}{\|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|}$, then the fourth eigenvalue of Eq. (8) $\lambda_4 = 0$. If and only if the fifth eigenvalue $\lambda_5 > 0$, $\boldsymbol{t}^{gt}$ is the unique fourth eigenvector of Eq. (8) and translation synchronization has a unique solution.*

Proof: See Section B.2. □

It is difficult to develop explicit conditions of $\boldsymbol{t}_i^{gt}$ on which $\lambda_5^{gt} > 0$. However, in many practical settings, $\boldsymbol{t}_i^{gt}$ usually lie close to a plane, e.g., images captured by a moving vehicle. Therefore, we study the properties of λ_5^{gt} when projecting $\boldsymbol{v}_{ij}^{gt}$ onto a plane.

First, we have the relation between the uniqueness of the 2D projection problem and the uniqueness of the original problem.

Proposition 3. *Let $\boldsymbol{n}$ be the normal to the plane of consideration. Denote $\boldsymbol{v}_{ij}^{gt,2D} = (I_3 - \boldsymbol{n}\boldsymbol{n}^T)\boldsymbol{v}_{ij}^{gt} / \|(I_3 - \boldsymbol{n}\boldsymbol{n}^T)\boldsymbol{v}_{ij}^{gt}\|$ as the projected direction along (i,j). Introduce L^{2D} as the connection Laplacian derived from $\boldsymbol{v}_{ij}^{gt,2D}$. With $\lambda_5^{gt,2D}$ we denote the fifth smallest eigenvalue of L^{2D}. We have, if $\lambda_5^{gt,2D} > 0$, then $\lambda_5^{gt} > 0$.*

Proof: See Section B.3. □

We can also understand why 2D uniqueness is more difficult than 3D uniqueness as follows. We have $3n-4$ variables in 3D and each edge offers two constraints. Therefore, we need $|\mathcal{E}| \geq \frac{3n-4}{2}$ to have a unique solution. We have $2n-3$ variables in 2D and each edge offers one constraint. Therefore, we need more edges, i.e., $|\mathcal{E}| \geq 2n-3$, to have a unique solution. This also means that if the $\mathcal{G}$ is a 2D planar graph, then we do not have unique solutions.

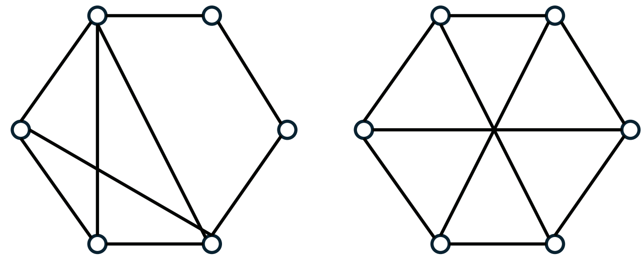


Figure 2. Connected graphs with six vertices. (Left) An example graph that does not have unique solutions. (Right) An example graph that has unique solutions.

In Section C of the supp. material, we present a necessary uniqueness condition on the topological structure of $\mathcal{E}$. It is also a sufficient condition in a probabilistic sense. Figure 2 shows examples of topological graphs that admit unique solutions and do not admit unique solutions.

4.3. Local Stability Analysis

When $\boldsymbol{v}_{ij}^{\text{inp}}$ are not exact, the fourth eigenvector $\boldsymbol{u}_4$ of generalized eigen-decomposition probelm in (8) will deviate from $\boldsymbol{t}^{gt}$. The deviations depend on w_{ij}, s_i, and the edge set $\mathcal{E}$. This section presents local stability results that provide insights into how to set w_{ij} and s_i to minimize the deviations of $\boldsymbol{u}_4$.

We begin by describing a noise model for $\boldsymbol{v}_{ij}^{\text{inp}}$. Denote $\boldsymbol{v}_{ij}^{gt}$ as the ground truth of $\boldsymbol{v}_{ij}^{\text{inp}}$. Let θ_{ij} be the angle between $\boldsymbol{v}_{ij}^{\text{inp}}$ and $\boldsymbol{v}_{ij}^{gt}$. We assume $\boldsymbol{v}_{ij}^{\text{inp}}$ is a random perturbation of $\boldsymbol{v}_{ij}^{gt}$ such that the perturbation is circular symmetric and $\mathbb{E}(\sin^2(\theta_{ij})) = \sigma_{ij}^2$.

Given $\{w_{ij}, s_i, \sigma_{ij}\}$, we define a stability score of $\boldsymbol{u}_4$ as

$$f(\{w_{ij}\}, \{s_i\}, \{\sigma_{ij}\}) = \mathop{\mathbb{E}}_{\{\boldsymbol{v}_{ij}^{\text{inp}}\}} \frac{\|\boldsymbol{u}_4 - \boldsymbol{t}^{gt}\|^2}{\|\boldsymbol{t}^{gt}\|^2}. \qquad (9)$$

Our analysis focuses on the regime in which σ_{ij} is small, i.e., $\boldsymbol{v}_{ij}^{gt\,T}(\boldsymbol{v}_{ij}^{\text{inp}} - \boldsymbol{v}_{ij}^{gt}) \approx 0$. As we shall discuss next, we have simple and analytical results on the optimal w_{ij} and s_i in this regime. We will use these results to derive our algorithm, as we find them to be effective on arbitrary inputs.

The following theorem shows that when $\{\sigma_{ij}, w_{ij}\}$ are fixed, the optimal values of s_i to minimize $f(\{w_{ij}\}, \{s_i\}, \{\sigma_{ij}\})$ are identical.

Theorem 1. *Suppose* $\sum_{i=1}^{n} s_i = n$*. Then*

$$\{1\} = \underset{\{s_i\}}{\text{argmin}} \quad f(\{w_{ij}\}, \{s_i\}, \{\sigma_{ij}\}).$$

Proof: See Section D.1 and Section D.2. □

The following theorem characterizes the optimal values of $\{w_{ij}\}$ given σ_{ij}.

Theorem 2. *Factoring out the global scale of* w_{ij}*, we have*

$$\{\frac{1}{\sigma_{ij}^2 \|\boldsymbol{t}_{ij}^{gt}\|^2}\} = \underset{\{w_{ij}\}}{\text{argmin}} \quad f(\{w_{ij}\}, \{1\}, \{\sigma_{ij}\}).$$

where $\boldsymbol{t}_{ij}^{gt} = \boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}$.

Proof: See Section D.1 and Section D.3. □

From the analysis above, we can always set $s_i = 1$. The values of w_{ij} depend on the estimations of σ_{ij} and $\|\boldsymbol{t}_{ij}^{gt}\|$, which can be determined in an iterative manner. This leads to our iterative spectral translation synchronization algorithm, which is introduced next.

4.4. Iterative Spectral Translation Synchronization

Our algorithm consists of an initialization phase and an alternating phase of location optimization and weight update.

Initialization phase. We set $w_{ij} = 1, \forall(i,j) \in \mathcal{E}$ at iteration $k = 1$. On real datasets where graphs may contain weakly connected components, we find that taking the whole graph as the input can lead to degenerate solutions, i.e., $\|\boldsymbol{t}_i\| \to 1$ for some i. To address this issue, we iteratively prune vertices until $\min_i \|\boldsymbol{t}_i\| < \delta$ where δ is a user specified threshold. Translation synchronization is performed on the remaining sub-graph. Please refer to Section A in the supp. material for more details.

Alternating step II: location optimization. Given the current edge weights w_{ij} and vertex weights $s_i = 1$ at the current iteration k, we solve (8) to obtain an estimation of the image locations $\boldsymbol{t}_i$.

Alternating step III: weight update. Given the current image locations $\boldsymbol{t}_i, 1 \leq i \leq n$ at iteration k, we estimate

$$\sigma_{ij}^2 = \|\boldsymbol{v}_{ij}^{\text{inp}} - \frac{\boldsymbol{t}_{ij}}{\|\boldsymbol{t}_{ij}\|}\|^2$$

where $\boldsymbol{t}_{ij} = \boldsymbol{t}_i - \boldsymbol{t}_j$. Similarly, we predict $\boldsymbol{t}_{ij}^{gt}$ using $\boldsymbol{t}_{ij}$

Based on these two estimations, we set

$$w_{ij} = \frac{\sigma_k^2}{\sigma_k^2 + \|\sigma_{ij}^2\| \|\boldsymbol{t}_{ij}\|^2}$$

where σ_k is introduced to avoid overconfident estimations of σ_{ij} and $\boldsymbol{t}_{ij}^{gt}$. We set σ_k using the geometric decaying scheme, i.e.,

$$\sigma_k = \sigma_{\max}(\frac{\sigma_{\min}}{\sigma_{\max}})^{s_k}, \quad s_k = \frac{k-1}{k_{\max}-1}.$$

where $k_{\max}$ is the maximum number of iterations. In all of our experiments, we set $\sigma_{\max} = 1$, $\sigma_{\min} = 10^{-3}$, $k_{\max} = 30$. During optimization, we also truncate $w_{ij} = 0$ when $w_{ij} \leq 0.01$. This operation removes outlier contributions and improves performance.

4.5. Robust Recovery Conditions

In this section, we present a robust recovery condition on our approach. We will use the ground-truth connection Laplacian matrix in which all input directions are exact.

$$L^{gt}(\boldsymbol{w}) = \sum_{(i,j)\in\mathcal{E}} w_{ij} E_{ij} (I_3 - \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt\,T}) E_{ij}^T. \qquad (10)$$

where $\boldsymbol{w} = (w_{ij}) \in \mathbb{R}^{|\mathcal{E}|}$ collects all edge vectors. In the following, we first present a noise model. We then present the robust recovery result.

Noise model. Most previous work performs analysis by assuming that the noisy observations are independent. However, this assumption does not hold as the pairwise matches are computed from the input images, which exhibit correlations. In this paper, we present a worst-case analysis where the noisy observations can be adversarial.

Our noise model considers a mixture of inliners and outliers. Specifically, given a fixed observation graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$. The edge set $\mathcal{E} = \mathcal{E}^{\text{in}} \cup \mathcal{E}^{\text{out}}$ decomposes into an inlier set $\mathcal{E}^{\text{in}}$ and an outlier set $\mathcal{E}^{\text{out}}$. With $\mathcal{N}_i^{\text{in}} = \{j|(i,j) \in \mathcal{E}^{\text{in}}\}$ and $\mathcal{N}_i^{\text{out}} = \{j|(i.j) \in \mathcal{E}^{\text{out}}\}$ we denote the neighbors of i that form inlier and outlier edges.

For each $(i,j) \in \mathcal{E}_{\text{in}}$, we assume $\boldsymbol{v}_{ij}$ is an arbitrary unit length vector in the neighborhood of $\boldsymbol{v}_{ij}^{gt}$ that satisfies

$$\|(I_3 - \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt\,T}) \boldsymbol{v}_{ij}\| \leq \epsilon \qquad (11)$$

where ϵ is a sufficiently small constant. For each edge $(i,j) \in \mathcal{E}_{\text{out}}$, $\boldsymbol{v}_{ij}$ can be an arbitrary unit vector. Based on this noise model, we proceed to present the following robust recovery result.

Denote $\boldsymbol{t}_{ij}^\star = \boldsymbol{t}_i^\star - \boldsymbol{t}_j^\star$ and $\boldsymbol{t}_{ij}^{gt} = \boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}$, where $\boldsymbol{t}_i^\star$ and $\boldsymbol{t}_i^{gt}$ as the output translation of our approach and the ground-truth translation of image i.

Theorem 3. *Consider three universal small constants*

$c_1, c_2, c_3 \geq 1$. *Denote*

$$\delta_1(\sigma) = \max_{1\leq i\leq n} \Big(\epsilon|\mathcal{N}_i^{\text{in}}| + c_1\sigma^2|\mathcal{N}_i^{\text{out}}|\Big),$$

$$\delta_2(\sigma) = \max_{1\leq i\leq n} \Big(\epsilon \sum_{j\in\mathcal{N}_i^{\text{in}}} \|\boldsymbol{t}_{ij}^{gt}\| + c_1\sigma^2 \sum_{j\in\mathcal{N}_i^{\text{out}}} \|\boldsymbol{t}_{ij}^{gt}\|\Big),$$

$$\delta_3(\sigma) = \epsilon^2 \sum_{(i,j)\in\mathcal{E}^{\text{in}}} \|\boldsymbol{t}_{ij}^{gt}\|^2 + c_1\sigma^2 \sum_{(i,j)\in\mathcal{E}^{\text{out}}} \|\boldsymbol{t}_{ij}^{gt}\|^2.$$

Suppose $\delta_1(1) \leq \frac{1}{3}\lambda_5^{gt}$ *and*

$$c_2(\delta_1(1) + \delta_3(1))\|L^{gt+}(\mathbf{1})\|_1 < 1$$

Then under mild conditions in $\sigma_{\min}$*,* $\sigma_{\max}$*, and* $k_{\max}$*, our interactive spectral translation synchronization approach converges to a solution* $\boldsymbol{t}_i^\star$ *that satisfies*

$$\|\boldsymbol{t}_{ij}^\star - \boldsymbol{t}_{ij}^{gt}\| \leq (1-\beta)\frac{c_3\delta_2^{\min}\|E_{ij}^T L^{gt}(\mathbf{1})^\dagger\|_1}{1 - c_3(\delta_1^{\min} + \delta_3^{\min})\|L^{gt}(\mathbf{1})^\dagger\|_1} + \beta\|\boldsymbol{t}_{ij}^{gt}\|, \qquad \forall (i,j)\in\mathcal{E}, \tag{12}$$

and

$$\|\boldsymbol{t}^\star - \boldsymbol{t}^{gt}\|_\infty \leq (1-\beta)\frac{c_3\delta_2^{\min}\|L^{gt}(\mathbf{1})^\dagger\|_1}{1 - c_3(\delta_1^{\min} + \delta_3^{\min})\|L^{gt}(\mathbf{1})^\dagger\|_1} + \beta\|\boldsymbol{t}_{ij}^{gt}\|, \tag{13}$$

where

$$\beta \leq \frac{{\delta_1^{\min}}^2}{2(\lambda_5^{gt} - \delta_1^{\min} - \delta_3^{\min})^2} \tag{14}$$

and $\delta_i^{\min} = \delta_i(\sigma_{\min})$

Proof: See Section E. □

Theorem 3 shows that our approach can effectively remove input outliers whenever their noise level is upper-bounded. Note that we do not place any assumptions on their correlations.

5. Experimental Results

We begin with the experimental setup in Section 5.1. We then present results on synthetic and real datasets in Section 5.2 and Section 5.3, respectively. Finally, we present an analysis of our approach in Section 5.4.

5.1. Experimental Setup

Synthetic datasets. We describe a synthetic dataset as $\mathcal{D}(n, p_{\text{edge}}, t, p_{\text{noise}}, \sigma)$. Here n denotes the number of images that are randomly sampled from the unit sphere. $p_{\text{edge}} \in [0,1]$ denotes the percentage of edges. When $t = r$, these edges follow the Erdős–Rényi model. When $t = g$, these edges connect nearest neighbors, meaning $\mathcal{G}$ is a geometric graph. p_{noise} denotes the percentage of edges, in which the associated directions are random outliers. σ is the variance of the inliers. We fix $n = 100$ and use two values for each other hyper-parameter, i.e., $p_{\text{edge}} \in \{0.3, 0.7\}$, $t \in \{'r', 'g'\}$, $p_{\text{noise}} \in \{0.1, 0.4\}$, $\sigma \in \{0.01, 0.03\}$. For each configuration of hyper-parameters, we sample 20 times. In total, we have 16 synthetic datasets. As $n = 100$, we simplify the notation as $\mathcal{D}(p_{\text{edge}}, t, p_{\text{noise}}, \sigma)$.

	BATA	Fused_TA	1dSFM	LUD	ShapeFit	TranSync
$\mathcal{D}(0.7, r, 0.1, 0.01)$	3.87	<u>3.16</u>	30.6	4.15	2.86	**1.53**
$\mathcal{D}(0.7, g, 0.1, 0.01)$	3.73	<u>2.98</u>	29.1	3.38	2.43	**1.32**
$\mathcal{D}(0.7, r, 0.1, 0.03)$	5.68	<u>5.49</u>	34.3	11.91	8.30	**5.31**
$\mathcal{D}(0.7, g, 0.1, 0.03)$	4.94	<u>4.83</u>	30.5	10.03	7.14	**4.49**
$\mathcal{D}(0.7, r, 0.4, 0.01)$	<u>13.15</u>	14.96	106	26.26	102	**1.93**
$\mathcal{D}(0.7, g, 0.4, 0.01)$	<u>10.82</u>	13.89	100	15.28	17.4	**1.70**
$\mathcal{D}(0.7, r, 0.4, 0.03)$	<u>13.93</u>	17.45	108	44.19	228	**6.75**
$\mathcal{D}(0.7, g, 0.4, 0.03)$	<u>11.59</u>	16.00	96.3	28.90	78.6	**5.79**
$\mathcal{D}(0.3, r, 0.1, 0.01)$	6.93	5.44	70.8	6.57	<u>4.55</u>	**2.58**
$\mathcal{D}(0.3, g, 0.1, 0.01)$	4.37	4.34	55.5	<u>3.97</u>	21.4	**1.61**
$\mathcal{D}(0.3, r, 0.1, 0.03)$	10.12	<u>9.46</u>	69.6	18.93	14.0	**8.97**
$\mathcal{D}(0.3, g, 0.1, 0.03)$	6.21	<u>6.18</u>	50.9	11.11	8.64	**5.54**
$\mathcal{D}(0.3, r, 0.4, 0.01)$	<u>32.66</u>	57.87	226	86.14	592	**9.19**
$\mathcal{D}(0.3, g, 0.4, 0.01)$	<u>16.02</u>	20.83	219	38.36	267	**2.22**
$\mathcal{D}(0.3, r, 0.4, 0.03)$	<u>29.54</u>	40.80	232	96.23	508	**18.29**
$\mathcal{D}(0.3, g, 0.4, 0.03)$	<u>16.70</u>	23.46	234	46.64	297	**7.28**

Table 1. Translation errors($\times 10^{-3}$) on synthetic datasets. The top-performing approach is bold-faced. The second best is underlined.

Real datasets. We perform evaluations on the dataset provided by [32]. To ensure uniqueness, we remove the images whose degree is less than 3.

Baseline techniques. The baselines include five state-of-the-art methods: LUD [21], 1DSfm [32], BATA [35], ShapeFit [10], Fused- TA [20].

Evaluation protocol. We compare our method with the baseline methods using the protocol in [32], which is based on estimating the optimal scaling and global translation to align the output of an algorithm with the ground-truth. On synthetic datasets, we report the mean translation error of each algorithm. On real datasets, we report the median translation error of each algorithm. This is because our approach may remove a small number of images. We include these images when calculating the median. Note that we find these images have large errors among baseline approaches as well and do not affect the median value.

5.2. Results on Synthetic Datasets

Table 1 presents the quantitative results of our approach and the baseline approaches in synthetic data sets. Overall, our approach outperforms all baseline approaches by salient margins. The top baseline approaches are BATA and FusedTA, which show different behaviors in different data sets.

BATA shows a strong performance among baseline approaches in data sets with high noise levels, that is, $p_{\text{noise}} = 0.4$. The error reductions of our approach on these datasets range 38.1% and 86.1%. The greatest improvement is in $\mathcal{D}(0.3, g, 0.4, 0.01)$. The improvements are great when $\sigma = 0.01$. The smallest improvement is in $\mathcal{D}(0.3, r, 0.4, 0.03)$. We observe similar behaviors when $\sigma = 0.03$. These salient improvements show the strength of our approach in handling outliers in the inputs. In particular, our approach excels when there are clear separations between inliers and outliers, i.e., there are large variations in magnitudes of σ_{ij}. This shows the strength of our local analysis.

Fused_TA (LUD and ShapeFit) show good performance in the small outlier regime or large σ. The error reductions of our approach range from 3.3% to 59.4%. The greatest improvement is in $\mathcal{D}(0.3, g, 0.1, 0.01)$. We obtain similar improvements when $\sigma = 0.01$. The smallest improvement is in $\mathcal{D}(0.7, g, 0.1, 0.03)$. We observe a similar relative performance when $\sigma = 0.03$. Moreover, the relative improvements when $p_{\text{edge}} = 0.3$ are greater than when $p_{\text{edge}} = 0.7$. Those results are consistent with our discussion that our formulation is very effective in pruning outliers. When p_{noise} and the variations in σ_{ij} are small, our approach is restricted to the linear approximation of $\boldsymbol{u}_4$ used to determine w_{ij} and s_i. This shows certain limitations when $\sigma = 0.03$ (See Figure 4).

Our approach and baseline approaches perform better on geometric graphs than random graphs. This is counter-intuitive because the spectral gap, which is closely related to stability of eigenvectors under noisy observations, is smaller on geometric graphs than on random graphs. However, in Appendix D.3 we show that for our approach the difference between $\boldsymbol{u}_4$ and $\boldsymbol{u}_4^{gt}$ is related to $\text{Tr}(L^{gt^\dagger})$. In geometric graphs, this quantity is smaller than their random counterparts. This explains why geometric graphs have smaller translation errors than random graphs.

5.3. Results on Real Datasets

Table 2 shows the quantitative results in the 1DSFM benchmark datasets. In general, our approach still outperforms all baseline approaches. To better understand the relative performance of our approach and baseline approaches, we divide the fifteen datasets into four groups: (ds) that consists of Alamo, Montreal-Notre Dame, Notre Dame, Piazza del Popolo, which are dense graphs with small noise ratios, (dl) that consists of Ellis Island, Madrid Metropolis, and NYC Library, which are dense graphs with large noise ratios, (ss) that consists of Roman Forum, Tower of London, Vienna Cathedral, Yorkminster, which are sparse graphs with small noise ratios, and (sl) that consists of Gandarmenmarkt, Piccadilly, Trafalgar, Union Square, which are sparse graphs with large noise ratios.

	BATA	Fused_TA	1dSFM	LUD	ShapeFit	TranSync
Alamo	0.87	1.03	<u>0.86</u>	2.77	2.56	**0.82**
Montreal Notre Dame	<u>0.73</u>	2.06	1.23	1.20	1.96	**0.68**
Notre Dame	1.28	1.49	**0.92**	1.67	1.32	**0.92**
Piazza del Popolo	1.76	1.79	2.80	1.77	<u>1.66</u>	**1.65**
Roman Forum	6.04	42.94	**5.40**	24.77	46.01	5.65
Tower of London	<u>2.86</u>	4.52	9.40	8.93	47.49	**2.46**
Vienna Cathedral	<u>2.95</u>	5.38	3.90	5.91	11.23	**2.54**
Yorkminster	<u>1.73</u>	2.66	6.36	5.34	8.09	**1.56**
Ellis Island	4.09	7.70	<u>3.46</u>	8.02	16.24	**3.56**
Madrid Metropolis	<u>3.78</u>	7.07	8.12	8.84	28.80	**2.57**
NYC Library	<u>1.06</u>	1.90	2.00	2.22	9.83	**0.92**
Gendarmenmarkt	58.30	31.48	46.72	<u>30.19</u>	33.18	**27.81**
Piccadilly	<u>1.74</u>	7.25	2.45	4.00	14.59	**1.56**
Trafalgar	<u>5.90</u>	15.35	6.66	13.26	53.38	**5.12**
Union Square	5.48	7.10	**4.82**	7.32	10.60	6.12

Table 2. Translation errors for real dataset from 1dsfm. Units are specified by the ground-truth.

In (ds), which are easier datasets, our approach is comparable to the top performing baselines. In these data sets, all approaches can successfully remove small fractions of outliers in the inputs and perform a prediction from the remaining inputs.

In (ss), which are harder than (ds) because of the sparsity of the graph, our approach outperforms the baseline approaches in average. This is attributed to our interactive scheme and the local analysis that shows how to properly reweight the measurements.

In (ds) and (dl), which have large noise ratios, our approach outperforms baseline approaches except Union Square. This again shows the advantage of our approach, which is based on principled analysis to reweight inputs for translation synchronization.

Note that the relative improvements of our approach in real data are smaller than the relative improvements in synthetic data. One reason is that there is no clear separation of inliers and outliers in real data. However, our approach can still remove outliers and properly reweight the inputs to balance their errors to obtain considerably improved results.

5.4. Advanced Analysis

We proceed to analyze our iterative spectral translation synchronization approach. For this analysis, we show average results on the synthetic datasets, where we know the distributions of input errors.

Effects of the alternating minimization. As shown in Figure 3(a), the iterative procedure can effectively reduce the prediction error. The relative improvements depend on the type of observations. On dense graphs with small noise ratio, where the initialization

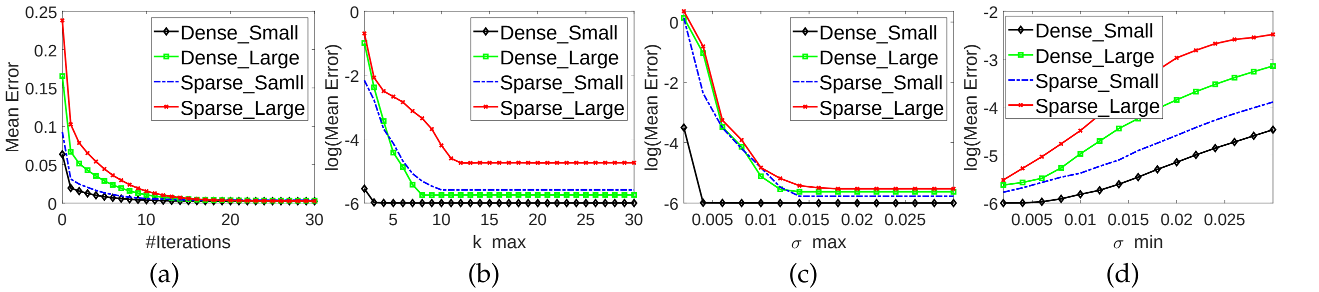


Figure 3. Advanced analysis of our approach. We show average translation error among all synthetic datasets. Above are on dense graphs with small noise ratio and bottom are on sparse graphs with large noise ratio (a) Effects of alternating minimization. (b) Effects of varying $k_{\max}$. (c) Effects of varying $\sigma_{\max}$. (d) Effects of varying $\sigma_{\min}$.

stage already returns a good solution, the improvement obtained from the iterative procedure is not significant. However, on sparse graphs with large noise ratio, the initial solution has large error as outliers participate in the estimation. Our reweighted scheme can substantially reduce the prediction error as the outliers have small weights in the end.

Effects of varying k_{max}. Figure 3(b) shows the results when varying $k_{\max}$. We can see that when $k_{\max}$ is above 10 across all datasets, the output of our algorithm becomes steady. In addition, on more challenging datasets, i.e., sparse graphs with large noise ratios, $k_{\max}$ that is required to reach a steady state is larger. This is expected, as when using a small $k_{\max}$, the weights of inliers may not differentiate from those of outliers during the alternating procedure due to inaccurate initial solution.

Effects of varying $\sigma_{\max}$. Figure 3(c) shows the effects when varying $\sigma_{\max}$. We can see that for easy datasets, there is not much difference when choosing a small $\sigma_{\max}$. This is because the initial solution is close to the underlying ground truth. The weighting scheme with a small $\sigma_{\max}$ can differentiate inliers and outliers. However, for hard data sets, that is, sparse graphs with large noise ratios, it is important to start from a large $\sigma_{\max}$. This allows us to gradually remove outliers at different noise levels.

Effects when varying $\sigma_{\min}$. Figure 3(d) shows the effects when varying $\sigma_{\min}$. We can see that decreasing the value of $\sigma_{\min}$ improves the prediction accuracy. However, $\sigma_{\min}$ can not be too small, i.e., smaller than variances of inliers. In this case, the weighting scheme cannot differentiate inliers and outliers, and we see that the prediction error drastically goes up. This effect is particularly salient on hard datasets, i.e., sparse graphs with large noise ratios.

6. Conclusions and Limitations

In this paper, we have introduced a robust translation synchronization approach that iteratively solves generalized eigenvalue problems. We present a rigorous stability analysis of this spectral formulation. The results are used to determine how to adaptively set the edge weights to progressively remove input outliers and average inliers of the input. Experimental results justify the effectiveness of our approach.

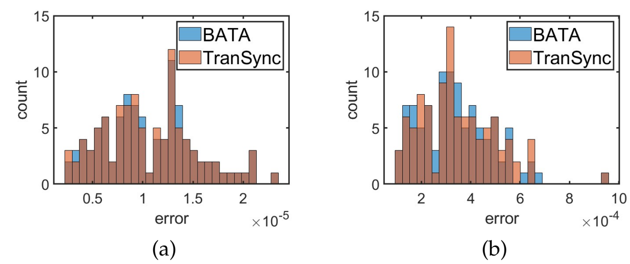


Figure 4. Comparison with our approach and BATA under $\mathcal{N}(0.7, g, 0, \sigma)$. (Left) $\sigma = 10^{-3}$. (Right) $\sigma = 0.03$.

One limitation of our approach is that we derive the optimal values of s_i and σ_{ij} in the infinitesimal regime, in which we have simple expressions of $\boldsymbol{v}_{ij}^{\text{inp}}$ and $\boldsymbol{u}_4$. The optimal w_{ij} and s_i are effective in removing outliers in $\boldsymbol{v}_{ij}^{\text{inp}}$. On the other hand, when all $\boldsymbol{v}_{ij}^{\text{inp}}$ are inliers with nonnegligible variance, we find that the linear approximation we used is sub-optimal when σ is large. As shown in Figure 4, our approach slightly outperforms BATA when $\sigma = 10^{-3}$. However, BATA slightly outperforms our approach when $\sigma = 0.03$. In the future, we plan to address this issue using second-order approximations of $\boldsymbol{u}_4$.

In the future, we also plan to integrate translation synchronization and rotation synchronization into a single optimization formulation, since the constraints for translation synchronization depend on rotations. Another direction is to explore learning-based approaches to set the edge weights, which have been shown to be effective for pose synchronization of RGB-D scans.

Acknowledgment. Q. H. acknowledges the support from NSF IIS-2047677, IIS-2413161, and GIFTs from Adobe and Google.

References

[1] Federica Arrigoni, Andrea Fusiello, and Beatrice Rossi. Camera motion from group synchronization. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 546–555. IEEE, 2016. 2

[2] Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello. Spectral synchronization of multiple views in se (3). *SIAM Journal on Imaging Sciences*, 9(4):1963–1990, 2016. 2

[3] Chandrajit Bajaj, Tingran Gao, Zihang He, Qixing Huang, and Zhenxiao Liang. SMAC: simultaneous mapping and clustering using spectral decompositions. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 334–343. PMLR, 2018. 2

[4] Florian Bernard, Johan Thunberg, Peter Gemmar, Frank Hertel, Andreas Husch, and Jorge Goncalves. A solution for multi-alignment by transformation synchronisation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2161–2169, 2015. 2

[5] Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In *ICCV*, pages 521–528. IEEE Computer Society, 2013. 2

[6] Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.*, 14(5):566–806, 2021. 2

[7] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *CVPR*, pages 5556–5565. IEEE Computer Society, 2015. 2

[8] Frank Dellaert, David M Rosen, Jing Wu, Robert Mahony, and Luca Carlone. Shonan rotation averaging: Global optimality by surfing $so(n)$. In *European Conference on Computer Vision*, pages 292–308. Springer, 2020. 2

[9] 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 2001*, pages II–II. IEEE, 2001. 2

[10] Paul Hand, Choongbum Lee, and Vladislav Voroninski. Shapefit: Exact location recovery from corrupted pairwise directions. *Communications on Pure and Applied Mathematics*, 71(1):3–50, 2018. 2, 6

[11] Qixing Huang and Leonidas Guibas. Consistent shape maps via semidefinite programming. In *Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing*, pages 177–186, 2013. 2

[12] Qi-Xing Huang, Guo-Xin Zhang, Lin Gao, Shi-Min Hu, Adrian Butscher, and Leonidas J. Guibas. An optimization approach for extracting and encoding consistent maps in a shape collection. *ACM Trans. Graph.*, 31(6): 167:1–167:11, 2012. 2

[13] Qixing Huang, Fan Wang, and Leonidas J. Guibas. Functional map networks for analyzing and exploring large shape collections. *ACM Trans. Graph.*, 33(4):36:1–36:11, 2014. 2

[14] Xiangru Huang, Zhenxiao Liang, Chandrajit Bajaj, and Qixing Huang. Translation synchronization via truncated least squares. In *NIPS*, 2017. 2

[15] Xiangru Huang, Zhenxiao Liang, Xiaowei Zhou, Yao Xie, Leonidas J. Guibas, and Qixing Huang. Learning transformation synchronization. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 8082–8091. Computer Vision Foundation / IEEE, 2019. 2

[16] Xiangru Huang, Zhenxiao Liang, Xiaowei Zhou, Yao Xie, Leonidas J Guibas, and Qixing Huang. Learning transformation synchronization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8082–8091, 2019. 2

[17] Vladimir G. Kim, Wilmot Li, Niloy J. Mitra, Stephen DiVerdi, and Thomas Funkhouser. Exploring collections of 3d models using fuzzy correspondences. *ACM Trans. Graph.*, 31(4):54:1–54:11, 2012. 2

[18] Spyridon Leonardos, Xiaowei Zhou, and Kostas Daniilidis. Distributed consistent data association via permutation synchronization. In *ICRA*, pages 2645–2652. IEEE, 2017. 2

[19] Shaohan Li, Yunpeng Shi, and Gilad Lerman. Fast, accurate and memory-efficient partial permutation synchronization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 15714–15722. IEEE, 2022. 2

[20] Lalit Manam and Venu Madhav Govindu. Fusing directions and displacements in translation averaging. In *International Conference on 3D Vision (3DV)*, 2024. 2, 6

[21] Onur Özyesil and Amit Singer. Robust camera location estimation by convex programming. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 2674–2683. IEEE Computer Society, 2015. 2, 6

[22] Deepti Pachauri, Risi Kondor, and Vikas Singh. Solving the multi-way matching problem by permutation synchronization. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 1860–1868, 2013. 2

[23] Deepti Pachauri, Risi Kondor, Gautam Sargur, and Vikas Singh. Permutation diffusion maps (PDM) with application to the image association problem in computer vision. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 541–549, 2014. 2

[24] D.M. Rosen, L. Carlone, A.S. Bandeira, and J.J. Leonard. SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group. *Intl. J. of Robotics Research*, 38(2–3):95–125, 2019. 2

[25] Yanyao Shen, Qixing Huang, Nati Srebro, and Sujay Sanghavi. Normalized spectral map synchronization.

In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 4925–4933, 2016. 2

[26] Yanyao Shen, Qixing Huang, Nati Srebro, and Sujay Sanghavi. Normalized spectral map synchronization. In *NIPS*, pages 4925–4933, 2016. 2

[27] Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30:20–36, 2011. 2

[28] Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30:20–36, 2011. 2

[29] Amit Singer and Hau tieng Wu. Vector diffusion maps and the connection laplacian. *Communications on Pure and Applied Mathematics*, 65:1067–1144, 2012. 2

[30] Yifan Sun, Zhenxiao Liang, Xiangru Huang, and Qixing Huang. Joint map and symmetry synchronization. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part V*, pages 257–275, 2018. 2

[31] Lanhui Wang and Amit Singer. Exact and stable recovery of rotations for robust synchronization. *Information and Inference: A Journal of the IMA*, 2:145–193, 2013. 2

[32] Kyle Wilson and Noah Snavely. Robust global translations with 1dsfm. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III 13*, pages 61–75. Springer, 2014. 2, 6

[33] Yuxin Chen, Leonidas J. Guibas, and Qi-Xing Huang. Near-optimal joint object matching via convex relaxation. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 100–108, 2014. 2

[34] Xiaowei Zhou, Menglong Zhu, and Kostas Daniilidis. Multi-image matching via fast alternating minimization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4032–4040, 2015. 2

[35] Bingbing Zhuang, Loong-Fah Cheong, and Gim Hee Lee. Baseline desensitizing in translation averaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4539–4547, 2018. 2, 3, 6

A Robust Translation Synchronization Algorithm

Supplementary Material

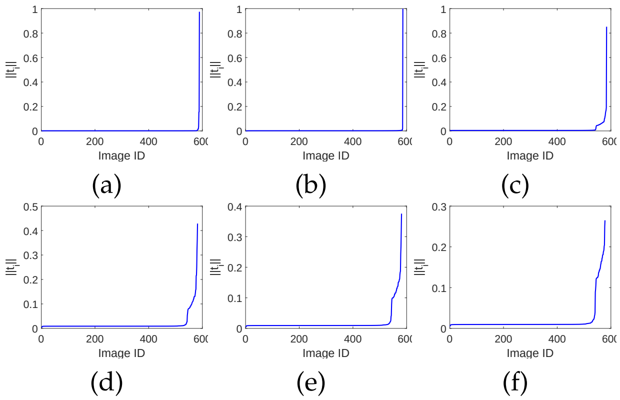


Figure 5. This figure shows $\|\boldsymbol{t}_i\|$ sorted by their magnitudes at each iteration on the Alamo dataset. (a) Iteration 1. (b) Iteration 2. (c) Iteration 3. (d) Iteration 4. (e) Iteration 5. (f) Iteration 6.

A. Details on Graph Pruning on Real Data

In this section, we provide more details on the pruning procedure which obtains a sub-graph to perform translation synchronization. Specifically, given the current sub-graph, which is initialized as the original graph, we perform translation synchronization with edge weights $w_{ij} = 1$. Let $\boldsymbol{t}_i$ be the resulting translations. If $\max \boldsymbol{t}_i \leq \delta$, we use the corresponding sub-graph to perform the alternating procedure of updating the edge weights and optimizing the image translations. Otherwise, we remove the image with the largest value in $\|\boldsymbol{t}_i\|$ and detect the maximum connected component within the remaining vertices to iterate this procedure. We set $\delta = \frac{10}{\sqrt{n}}$ in our experiments.

Figure 5 shows the iterative procedure on the Alamo dataset in 1DSFM, which removes x images. The numbers of images removed from all the datasets in 1DSFM are less than 8 images.

B. Spectral Translation Synchronization Properties

B.1. Proof of Proposition 1

Denote

$$E_{ij} = \boldsymbol{e}_{ij} \otimes I_3. \tag{15}$$

Introduce $\boldsymbol{v}_{ij,1}$ and $\boldsymbol{v}_{ij,2}$, so that $(\boldsymbol{v}_{ij}, \boldsymbol{v}_{ij,1}, \boldsymbol{v}_{ij,2})$ forms an orthonormal basis. Then we have

$$\begin{aligned} L &= \sum_{(i,j)\in\mathcal{E}} w_{ij} E_{ij}(I_3 - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^T)E_{ij}^T \\ &= \sum_{(i,j)\in\mathcal{E}} w_{ij} E_{ij}(\boldsymbol{v}_{ij,1}\boldsymbol{v}_{ij,1}^T + \boldsymbol{v}_{ij,2}\boldsymbol{v}_{ij,2}^T)E_{ij}^T \succeq 0. \end{aligned}$$

Consider any vector $\boldsymbol{c} \in \mathbb{R}^3$. Introduce $\boldsymbol{u} = \boldsymbol{1} \otimes \boldsymbol{c}$. The i-th block of $L\boldsymbol{u}$ is

$$\begin{aligned} &(L\boldsymbol{u})_i \\ =& \sum_{j\in\mathcal{N}_i} w_{ij}(I_3 - v_{ij}v_{ij}^T)\boldsymbol{u}_i + \sum_{j\in\mathcal{N}_i} -w_{ij}(I_3 - v_{ij}v_{ij}^T)\boldsymbol{u}_j \\ =& \sum_{j\in\mathcal{N}_i} w_{ij}(I_3 - v_{ij}v_{ij}^T)\boldsymbol{c} - \sum_{j\in\mathcal{N}_i} w_{ij}(I_3 - v_{ij}v_{ij}^T)\boldsymbol{c} = 0. \end{aligned}$$

Therefore, the first three eigenvalues of L are zero, and the eigenvectors are $\boldsymbol{1} \otimes I_3$. It is easy to check that this also applies to generalized eigen-values and eigen-vectors. □

B.2. Proof of Proposition 2

The i-th block of $L\boldsymbol{t}^{gt}$ is

$$\begin{aligned} &(L\boldsymbol{t}^{gt})_i \\ =& \sum_{j\in\mathcal{N}_i} w_{ij}(I_3 - v_{ij}^{gt}{v_{ij}^{gt}}^T)(\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}) \\ =& \sum_{j\in\mathcal{N}_i} w_{ij}\|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|(I_3 - v_{ij}^{gt}{v_{ij}^{gt}}^T)\boldsymbol{v}_{ij}^{gt} = 0. \end{aligned}$$

Therefore, $\boldsymbol{t}^{gt}$ is a fourth eigenvector L. If $\lambda_5 > 0$. Then $\boldsymbol{t}^{gt}$ is the unique fourth eigenvector.

If $\lambda_5 = 0$. Then there exists a different vector $\boldsymbol{u}_5 \neq \boldsymbol{t}^{gt}$ that is a slight pertubration of $\boldsymbol{t}^{gt}$, so that $L\boldsymbol{u}_5 = 0$. Note that

$$\begin{aligned} 0 &= \boldsymbol{u}_5^T L \boldsymbol{u}_5 \\ &= \sum_{(i,j)\in\mathcal{E}} \boldsymbol{u}_5^T E_{ij}(I_3 - \boldsymbol{v}_{ij}^{gt}{\boldsymbol{v}_{ij}^{gt}}^T)E_{ij}\boldsymbol{u}_5 \\ &= \sum_{(i,j)\in\mathcal{E}} \Big(\|\boldsymbol{u}_{5i} - \boldsymbol{u}_{5j}\|^2 - \big((\boldsymbol{u}_{5i} - \boldsymbol{u}_{5j})^T\boldsymbol{v}_{ij}^{gt}\big)^2\Big). \end{aligned}$$

Therefore, $\forall (i,j) \in \mathcal{E}$,

$$\|\boldsymbol{u}_{5i} - \boldsymbol{u}_{5j}\|^2 = \big((\boldsymbol{u}_{5i} - \boldsymbol{u}_{5j})^T\boldsymbol{v}_{ij}^{gt}\big)^2.$$

This means

$$\frac{\boldsymbol{u}_{5i} - \boldsymbol{u}_{5j}}{\|\boldsymbol{u}_{5i} - \boldsymbol{u}_{5j}\|} = s_{ij}\boldsymbol{v}_{ij}^{gt},$$

where $s_{ij} \in \{-1, 1\}$. As $\boldsymbol{u}_5$ is a slight perturbation of $\boldsymbol{t}^{gt}$, we have $s_{ij} = 1, \forall (i,j) \in \mathcal{E}$. This ends the proof. □

B.3. Proof of Proposition 3

As $\lambda_4^{gt,2D} > 0$, we can determine the $\boldsymbol{t}_i^{gt,2D}$ on the plane with normal $\boldsymbol{n}$, up to a global scale s and translation. In addition, we have $\boldsymbol{t}_i^{gt,2D} \neq \boldsymbol{t}_j^{gt,2D}, \forall (i,j) \in \mathcal{E}$. Consider a spanning tree of $\mathcal{G}$. Without losing generality, we assume for the root r of this tree, we have $\boldsymbol{n}^T\boldsymbol{t}_r^{gt} = 0$. It is easy to see that starting the root r, we can determine $\boldsymbol{n}^T\boldsymbol{t}_i^{gt}$ iteratively. This is because, $\boldsymbol{t}_i^{gt,2D} - \boldsymbol{t}_j^{gt,2D}$ is given, and $\boldsymbol{n}^T(\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt})$ can be recovered from $\boldsymbol{v}_{ij}^{gt}$. □

C. Topological Uniqueness Condition

In the following, we present a necessary uniqueness condition on the topological structure of $\mathcal{E}$. It is also a sufficient condition in a probabilistic sense.

Definition 1. *Introduce a rigidity matrix* $A = (\mathbf{1} \otimes I_2; \boldsymbol{v}; B) \in \mathbb{R}^{(|\mathcal{E}|+3)\times 2n}$ *of a graph* $\mathcal{G}$ *with edge set* $\mathcal{E}$*. The elements of* $\boldsymbol{v}$ *are* $v_{2i-1} = \cos(i\theta)$ *and* $v_{2i} = \sin(i\theta)$ *where* $\theta = \frac{2\pi}{n}$*. The elements of* B *are zero except* $B_{(i,j),2i-1} = \cos(\frac{(i+j)\theta}{2}$*,* $B_{(i,j),2i} = \sin(\frac{(i+j)\theta}{2}$*,*$B_{(i,j),2j-1} = -\cos(\frac{(i+j)\theta}{2}$*, and* $B_{(i,j),2j} = -\sin(\frac{(i+j)\theta}{2}$*. We say* $\mathcal{G}$ *is rigid if* $\mathrm{rank}(B) = 2n$*.*

Theorem 4. *$\mathcal{G}$ is rigid is a necessary condition for 2D uniqueness of translation synchronization. It is a sufficient condition in the sense that 2D uniqueness holds with probability 1 if we randomly sample* $\boldsymbol{t}_i^{gt,2D}$*.*

Proof: Consider an arbitrary set of n vertices $\boldsymbol{p}_i = (x_i, y_i)^T, 1 \le i \le n$ with edge set $\mathcal{E}$.

$$\boldsymbol{v}_{ij} = \frac{(x_i - x_j, y_i - y_j)}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}},$$
$$\boldsymbol{v}_{ij}^{\perp} = \frac{(-(y_i - y_j), x_i - x_j)}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}.$$

Define $J \in \mathbb{R}^{\mathcal{E}\times 2n}$, where $J((i,j),(2i-1):(2i)) = \boldsymbol{v}_{ij}^{\perp}$ and $J((i,j),(2j-1):(2j)) = -\boldsymbol{v}_{ij}^{\perp}$. When $x_i = \cos(\frac{\pi i}{n})$ and $y_i = \sin(\frac{\pi i}{n})$, it is clear that $\mathrm{rank}(J) = 2n-3$ if and only if $\mathrm{rank}(B) = 2n$. When $\mathrm{rank}(A) < 2n$. Then $\mathrm{rank}(J) < 2n-3$. This means the rank of $L^{gt} = J' * J$ is smaller than $2n-4$. In this case, $\lambda_i^{4,gt} = 0$, and 2D uniqueness does not hold.

Suppose $\mathrm{rank}(B) = 2n$. We show that $\mathrm{rank}(J) = 2n-3$ when x_i, y_i are random samples. It is sufficient to show that for an edge subset $\mathcal{E}' \subset \mathcal{E}$ where $|\mathcal{E}'| = 2n-3$ and $\mathrm{rank}(B') = 2n$, the corresponding Jacobian matrix J has rank $\mathrm{rank}(J') = 2n-3$. Denote

$$J' = \mathrm{diag}(\frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}})\overline{J}'$$

where

$$\overline{J}'((i,j),(2i-1):(2i)) = (-(y_i - y_j), (x_i - x_j)),$$
$$\overline{J}'((i,j),(2j-1):(2j)) = ((y_i - y_j), -(x_i - x_j)).$$

Then it is clear that $\mathrm{rank}(J') = 2n-3$ when $x_i = \cos(\frac{i\pi}{n})$ and $y_i = \sin(\frac{i\pi}{n})$. We show that $\mathrm{rank}(J') = 2n-3$ when x_i and y_i are arbitrary. Suppose this is not true, there exists non-zero coefficients $c_i, 1 \le i \le 2n-3$, so that $\sum\limits_{i=1}^{2n-3} c_i\overline{J}'(i,:) = 0$ for any x_i and y_i almost surely. As $\overline{J}'(i,:)$ are linear in x_i and y_i, it means that $\sum\limits_{i=1}^{2n-3} c_i\overline{J}'(i,:) = 0$ holds for all x_i and y_i, including $x_i = \cos(\frac{i\pi}{n})$ and $y_i = \sin(\frac{i\pi}{n})$. □

D. Proofs of Theorems 1 and Theorems 2

We begin with an analytic expression of the objective function $f(\{w_{ij}\},\{s_i\},\{\sigma_{ij}\})$ in Section D.1. Section D.2 and Section D.3 complete the proofs of Theorem 1 and Theorem 2. Section D.4 presents proofs of the propositions in Section D.1.

D.1. Analytical Expression of $f(\{w_{ij}\},\{s_i\},\{\sigma_{ij}\})$

Denote $A_{ij} = I_3 - \boldsymbol{v}_{ij}^{\mathrm{inp}}\boldsymbol{v}_{ij}^{\mathrm{inp}\,T}$ and $A_{ij}^{gt} = I_3 - \boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T}$. Introduce

$$dA_{ij} = A_{ij} - A_{ij}^{gt}$$

The following proposition characterizes an important property regarding dA_{ij}.

Proposition 4. *Consider a symmetric matrix* $F \in \mathbb{R}^{3\times 3}$*. Then by dropping third-and-higher order terms, we have*

$$\mathop{\mathbb{E}}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} dA_{ij} = \frac{\sigma_{ij}^2}{2}(3\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T} - I_3),$$

and

$$\mathop{\mathbb{E}}_{\{\epsilon_{ij,k}\}} dA_{ij}FdA_{ij} = \frac{\sigma_{ij}^2}{2}\Big(\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T}(\mathrm{Tr}(F) - 4\boldsymbol{v}_{ij}^{gt\,T}F\boldsymbol{v}_{ij}^{gt}) + \boldsymbol{v}_{ij}^{gt\,T}F\boldsymbol{v}_{ij}^{gt}I_3 + \boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T}F + F\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T}\Big) \quad (16)$$

Proof. See Section D.4.1. □

The blocks of the perturbation connection Laplacian matrix $dL = L - L^{gt}$ is given by

$$dL_{ij} = \begin{cases} \sum\limits_{k\in\mathcal{N}_i} w_{ik} dA_{ik} & i = j \\ -w_{ij} dA_{ij} & (i,j)\in\mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

The following proposition characterize an expectation of dL when dA_{ij} follows the distribution described above.

Proposition 5. *Consider a positive semidefinite matrix $B \in \mathbb{R}^{3n\times 3n}$. Let $\boldsymbol{t}^{gt} \in \mathbb{R}^{3n}$ collect $\boldsymbol{t}_i^{gt}$ in its blocks. Then*

$$\begin{aligned} &\underset{\{\boldsymbol{v}_{ij}^{\text{inp}}\}}{\mathbb{E}} {\boldsymbol{t}^{gt}}^T dLBdL\boldsymbol{t}^{gt} = \sum_{(i,j)\in\mathcal{E}} \frac{\sigma_{ij}^2}{2} w_{ij}^2 \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2 \\ &\Big(\mathrm{Tr}(E_{ij}BE_{ij}^T) - {\boldsymbol{v}_{ij}^{gt}}^T E_{ij}BE_{ij}^T \boldsymbol{v}_{ij}^{gt}\Big) \\ &\quad + \Big(\sum_{(i,j)\in\mathcal{E}} w_{ij}\sigma_{ij}^2 \|t_i^{gt} - t_j^{gt}\| E_{ij}^T v_{ij}^{gt}\Big)^T B \\ &\Big(\sum_{(i,j)\in\mathcal{E}} w_{ij}\sigma_{ij}^2 \|t_i^{gt} - t_j^{gt}\| E_{ij}^T v_{ij}^{gt}\Big). \end{aligned} \tag{17}$$

where E_{ij} is defined in Eq. (15).

Proof. See Section D.4.2. □

We proceed to analyze the stability of $\boldsymbol{u}_4$. First of all, all $3n$ generalized eigenvectors $\boldsymbol{u}_i, 1 \le i \le 3n$ satisfy that

$$\boldsymbol{u}_i^T W \boldsymbol{u}_j = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

Note that $\sum\limits_{i=1}^n \boldsymbol{t}_i^{gt} = 0$. Without losing generalization, we assume

$$\sum_{i=1}^n s_i = 1, \qquad \sum_{i=1}^n s_i \boldsymbol{t}_i^{gt} = 0. \tag{18}$$

The first constraint in (18) normalizes the scale of s_i. The second equality in (18) places an additional constraint on s_i. The following proposition characterizes the top four generalized eigenvectors of L.

Proposition 6. *Under the assumptions in (18), we have*

$$(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3) = \boldsymbol{1} \otimes I_3, \qquad \boldsymbol{u}_4 = \frac{\boldsymbol{t}^{gt}}{\sqrt{\sum\limits_{i=1}^n s_i \|\boldsymbol{t}_i^{gt}\|^2}} \tag{19}$$

The next proposition describes the perturbation in $\boldsymbol{u}_4$ with respect to the perturbation dL in L.

Proposition 7. *Suppose $\lambda_5(L^{gt}) > 0$. Then*

$$d\boldsymbol{u}_4 = -\Big(I_{3n} - U_4 U_4^T S\Big) L^{\dagger} (I_{3n} - S\boldsymbol{u}_4 \boldsymbol{u}_4^T) dL \boldsymbol{u}_4. \tag{20}$$

where $U_4 = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4)$.

Proof: See Section D.5. □

Denote $\boldsymbol{s} = (s_i) \in \mathbb{R}^n$. When σ_{ij} and w_{ij} are fixed and applying (20), we can rewrite the objective function f as

$$f(\boldsymbol{s}) = \underset{\{\epsilon_{ij,k}\}}{\mathbb{E}} {\boldsymbol{t}^{gt}}^T dLB(\boldsymbol{s}) dL \boldsymbol{t}^{gt} \tag{21}$$

where

$$B(\boldsymbol{s}) = (I_{3n} - \frac{\boldsymbol{t}^{gt}(S\boldsymbol{t}^{gt})^T}{\sum\limits_{i=1}^n s_i \|\boldsymbol{t}_i^{gt}\|^2}) L^{\dagger} C(\boldsymbol{s}) L^{\dagger} (I_{3n} - \frac{S\boldsymbol{t}^{gt}{\boldsymbol{t}^{gt}}^T}{\sum\limits_{i=1}^n s_i \|\boldsymbol{t}_i^{gt}\|^2})$$

where

$$\begin{aligned} C(\boldsymbol{s}) &= I_{3n} - (\boldsymbol{1}\boldsymbol{s}^T + \boldsymbol{s}\boldsymbol{1}^T) \otimes I_3 + n\boldsymbol{s}\boldsymbol{s}^T \otimes I_3 \\ &\quad - \frac{\boldsymbol{t}^{gt}(S\boldsymbol{t}^{gt})^T + S\boldsymbol{t}^{gt}{\boldsymbol{t}^{gt}}^T}{\sum\limits_{i=1}^n s_i \|\boldsymbol{t}_i^{gt}\|^2} + \frac{\|\boldsymbol{t}^{gt}\|^2 S\boldsymbol{t}^{gt}(S\boldsymbol{t}^{gt})^T}{(\sum\limits_{i=1}^n s_i \|\boldsymbol{t}_i^{gt}\|^2)^2} \end{aligned}$$

Note that $L(\boldsymbol{1} \otimes I_3) = 0$ and $L\boldsymbol{t}^{gt} = 0$. We can simplify

$$\begin{aligned} L^{\dagger} C(\boldsymbol{s}) L^{\dagger} &= {L^{\dagger}}^2 + nL^{\dagger}(\boldsymbol{s}\boldsymbol{s}^T \otimes I_3) L^{\dagger} \\ &\quad + \frac{\|\boldsymbol{t}^{gt}\|^2}{(\sum\limits_{i=1}^n s_i \|\boldsymbol{t}_i^{gt}\|^2)^2} (L^{\dagger} S\boldsymbol{t}^{gt})(L^{\dagger} S\boldsymbol{t}^{gt})^T. \end{aligned} \tag{22}$$

To apply Prop. 5 to derive a closed-form expression of $f(\boldsymbol{w})$, we compute

$$\begin{aligned} E_{ij} B(\boldsymbol{s}) E_{ij}^T &= (E_{ij} - \frac{(\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt})(S\boldsymbol{t}^{gt})^T}{\sum\limits_{i=1}^n s_i \|\boldsymbol{t}_i^{gt}\|^2}) \\ L^{\dagger} C(\boldsymbol{w}) L^{\dagger} &(E_{ij}^T - \frac{S\boldsymbol{t}^{gt}(\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt})^T}{\sum\limits_{i=1}^n s_i \|\boldsymbol{t}_i^{gt}\|^2}) \end{aligned} \tag{23}$$

To simplify $f(\boldsymbol{s})$, we note that for any vector $\boldsymbol{h} \in \mathbb{R}^3$,

$$\begin{aligned} &\mathrm{Tr}((\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt})\boldsymbol{h}^T) - {\boldsymbol{v}_{ij}^{gt}}^T (\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt})\boldsymbol{h}^T) \boldsymbol{v}_{ij}^{gt} \\ =&\mathrm{Tr}((\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt})\boldsymbol{h}^T) - \mathrm{Tr}((\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt})\boldsymbol{h}^T) \\ =&0. \end{aligned}$$

It follows that we can simplify the objective function as

$$\begin{aligned} f(\boldsymbol{s}) = \sum_{(i,j)\in\mathcal{E}} &\frac{\sigma_{ij}^2}{2} w_{ij}^2 \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2 \Big(\mathrm{Tr}(E_{ij} L^{\dagger} C(\boldsymbol{s}) L^{\dagger} E_{ij}^T) \\ &- {\boldsymbol{v}_{ij}^{gt}}^T E_{ij} L^{\dagger} C(\boldsymbol{w}) L^{\dagger} E_{ij}^T \boldsymbol{v}_{ij}^{gt}\Big). \end{aligned} \tag{24}$$

D.2. Proof of Theorem 1

According Eq. (22), it is clear that

$$L^{\dagger}C(\boldsymbol{s})L^{\dagger} \succeq {L^{\dagger}}^2,$$

and equality holds if and only if $\boldsymbol{s} = \mathbf{1}$. Therefore,

$$f(\boldsymbol{s}) \geq \sum_{(i,j)\in\mathcal{E}} \frac{\sigma_{ij}^2}{2} w_{ij}^2 \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2 \Big(\mathrm{Tr}(E_{ij}{L^{\dagger}}^2 E_{ij}^T) - {\boldsymbol{v}_{ij}^{gt}}^T E_{ij}{L^{\dagger}}^2 E_{ij}^T \boldsymbol{v}_{ij}^{gt}\Big)$$

and equality holds when $\boldsymbol{s} = \mathbf{1}$. This ends the proof of Theorem 1. □

D.3. Proof of Theorem 2

Suppose $s_i = 1$. In this case, we optimize w_{ij} to minimize $f(\{w_{ij}\}, \{s_i\}, \{\sigma_{ij}\})$. L becomes a function of $\{w_{ij}\}$. Define

$$\overline{L} = \sum_{(i,j)\in\mathcal{E}} \sigma_{ij}^2 w_{ij}^2 \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2 E_{ij}^T (I_3 - \boldsymbol{v}^{gt}{\boldsymbol{v}^{gt}}^T) E_{ij},$$

and

$$\overline{\boldsymbol{g}} = \sum_{(i,j)\in\mathcal{E}} \sigma_{ij}^2 w_{ij} \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\| E_{ij}^T \boldsymbol{v}_{ij}^{gt}.$$

The objective function to be minimized is given by

$$\begin{aligned} f(\{w_{ij}^{\star}\}) &= \frac{1}{2}\sum_{(i,j)\in\mathcal{E}} \sigma_{ij}^2 w_{ij}^2 \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2 \Big(\mathrm{Tr}(E_{ij}{L^{\dagger}}^2 E_{ij}^T) - {\boldsymbol{v}_{ij}^{gt}}^T E_{ij}{L^{\dagger}}^2 E_{ij}^T \boldsymbol{v}_{ij}^{gt}\Big) \\ &= \frac{1}{2}\mathrm{Tr}\Big(L^{\dagger}\overline{L}L^{\dagger}\Big). \end{aligned}$$

Apply the chain rule, we have

$$\frac{\partial f}{\partial w_{ij}} = \frac{1}{2}\mathrm{Tr}\Big(L^{\dagger}\frac{\partial \overline{L}}{\partial w_{ij}}L^{\dagger} - L^{\dagger}\overline{L}L^{\dagger}\frac{\partial L}{\partial w_{ij}}L^{\dagger} - L^{\dagger}\frac{\partial L}{\partial w_{ij}}L^{\dagger}\overline{L}L^{\dagger}\Big). \tag{25}$$

To simplify (25), we introduce $G = (G_1; G_2) \in \mathbb{R}^{2|\mathcal{E}|\times 3n}$ where

$$G_1 = (\sqrt{w_{ij}}{\boldsymbol{v}_{ij,1}^{gt}}^T E_{ij}), \quad G_2 = (\sqrt{w_{ij}}{\boldsymbol{v}_{ij,2}^{gt}}^T E_{ij}).$$

It is easy to check that

$$L = G^T G, \quad L^{\dagger} = G^{\dagger}{G^{\dagger}}^T,$$

and

$$\overline{L} = G^T D_4 G, \quad D_4 = I_2 \otimes \mathrm{diag}(w_{ij}\sigma_{ij}^2 \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2)$$

Moreover,

$$\frac{\partial L}{\partial w_{ij}} = \frac{1}{w_{ij}} G^T (I_2 \otimes \boldsymbol{e}_{ij}\boldsymbol{e}_{ij}^T) G \tag{26}$$

$$\frac{\partial \overline{L}}{\partial w_{ij}} = \frac{2}{w_{ij}} G^T D_4 (I_2 \otimes \boldsymbol{e}_{ij}\boldsymbol{e}_{ij}^T) G \tag{27}$$

Substituting Eq. (26) and Eq. (27) into Eq. (25), we have

$$\frac{\partial f}{\partial w_{ij}} = \frac{2}{w_{ij}}\mathrm{Tr}\Big((I_2 \otimes \boldsymbol{e}_{ij})^T {G^{\dagger}}^T G^{\dagger} D_4 (I - G^{\dagger}G^T)(I_2 \otimes \boldsymbol{e}_{ij})\Big). \tag{28}$$

When $w_{ij}^{\star} = \frac{1}{\sigma_{ij}^2\|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2}$, we have $D_4 = I_{2|\mathcal{E}|}$. As

$${G^{\dagger}}^T G^{\dagger} = {G^{\dagger}}^T G^{\dagger} G^{\dagger} G^T,$$

we have $\frac{\partial f}{\partial w_{ij}} = 0$. This means $\{w_{ij}^{\star}\}$ is a critical point. Next, we show that $\{w_{ij}^{\star}\}$ is a local minimum.

Proposition 8. *Consider any vector $\boldsymbol{x} \neq \mathbf{0} \in \mathbb{R}^{|\mathcal{E}|}$, where $\sum\limits_{(i,j)\in\mathcal{E}} x_{ij}w_{ij}^{\star} = 0$. We have*

$$\sum_{(i,j)\in\mathcal{E}}\sum_{(i',j')\in\mathcal{E}} x_{ij}x_{i'j'}\frac{\partial^2 f}{\partial w_{ij}\partial w_{i'j'}} > 0.$$

Proof. See Section D.5.1. □

Next, we show that $\{w_{ij}^{\star}\}$ is the only critical point of f.

Proposition 9. *$\{sw_{ij}^{\star}\}$ is the only solution to $\forall (i,j) \in \mathcal{E}$,*

$$\mathrm{Tr}\Big((I_2 \otimes \boldsymbol{e}_{ij})^T {G^{\dagger}}^T G^{\dagger} D_4 (I - G^{\dagger}G^T)(I_2 \otimes \boldsymbol{e}_{ij})\Big) = 0. \tag{29}$$

Proof. See Section D.5.2. □

This ends the proof of Theorem 2. □

D.4. Proofs of Propositions

D.4.1 Proof of Prop. 4

Note that We can decompose $\boldsymbol{v}_{ij}^{\mathrm{inp}}$ to the orthogonal vectors $\boldsymbol{v}_{ij}^{gt}, \boldsymbol{v}_{ij}^{\perp}$. $\boldsymbol{v}_{ij}^{\mathrm{inp}} = \cos\theta_{ij}\boldsymbol{v}_{ij}^{gt} + \sin\theta_{ij}\boldsymbol{v}_{ij}^{\perp}$

$$\mathbb{E}_{\boldsymbol{v}_{ij}^{\perp}} \boldsymbol{v}_{ij}^{\perp T} F \boldsymbol{v}_{ij}^{\perp} = \frac{1}{2}\Big(\mathrm{Tr}(F) - \boldsymbol{v}_{ij}^{gt T} F \boldsymbol{v}_{ij}^{gt}\Big),$$

$$\mathbb{E}_{\boldsymbol{v}_{ij}^{\perp}} \boldsymbol{v}_{ij}^{\perp} \boldsymbol{v}_{ij}^{\perp T} = \frac{1}{2}\Big(I_3 - \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T}\Big).$$

$$\mathbb{E}_{\boldsymbol{v}_{ij}^{\perp}} \boldsymbol{v}_{ij}^{\perp} F \boldsymbol{v}_{ij}^{gt T} = 0.$$

$$\mathbb{E}_{d\boldsymbol{v}_{ij}} \boldsymbol{v}_{ij}^{\mathrm{inp}} \boldsymbol{v}_{ij}^{\mathrm{inp} T} = \mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} (\cos^2\theta_{ij} \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T} + \sin^2\theta_{ij} \boldsymbol{v}_{ij}^{\perp} \boldsymbol{v}_{ij}^{\perp T}).$$

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} dA_{ij} &= \mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T} - \boldsymbol{v}_{ij}^{\mathrm{inp}} \boldsymbol{v}_{ij}^{\mathrm{inp} T} \\
&= - \mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} (\sin^2\theta_{ij}) \boldsymbol{v}_{ij}^{\perp} \boldsymbol{v}_{ij}^{\perp T} + \mathbb{E}_{\theta_{ij}} (\sin^2\theta_{ij}) \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T} \\
&= \frac{\sigma_{ij}^2}{2} (3 \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T} - I_3).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{E}_{d\boldsymbol{v}_{ij}} dA_{ij} F dA_{ij} \\
=& \mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} \big(\boldsymbol{v}_{ij}^{gt} d\boldsymbol{v}_{ij}^T + d\boldsymbol{v}_{ij} \boldsymbol{v}_{ij}^{gt T} + d\boldsymbol{v}_{ij} d\boldsymbol{v}_{ij}^T - \|d\boldsymbol{v}_{ij}\|^2 \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T}\big) \\
=& \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T} \mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} \sin^2\theta_{ij} \cos^2\theta_{ij} \boldsymbol{v}_{ij}^{\perp T} F \boldsymbol{v}_{ij}^{\perp} \\
&+ \boldsymbol{v}_{ij}^{gt T} F \boldsymbol{v}_{ij}^{gt} \mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} \sin^2\theta_{ij} \cos^2\theta_{ij} \boldsymbol{v}_{ij}^{\perp T} \boldsymbol{v}_{ij}^{\perp} \\
&+ \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T} F \mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} \sin^2\theta_{ij} \cos^2\theta_{ij} \boldsymbol{v}_{ij}^{\perp} \boldsymbol{v}_{ij}^{\perp T} \\
&+ \mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} \sin^2\theta_{ij} \cos^2\theta_{ij} \boldsymbol{v}_{ij}^{\perp} \boldsymbol{v}_{ij}^{\perp T} F \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T} \\
=& \frac{\sigma_{ij}^2}{2} \Big(\boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T} (\mathrm{Tr}(F) - 4 \boldsymbol{v}_{ij}^{gt T} F \boldsymbol{v}_{ij}^{gt}) + \boldsymbol{v}_{ij}^{gt T} F \boldsymbol{v}_{ij}^{gt} I_3 \\
&\quad + \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T} F + F \boldsymbol{v}_{ij}^{gt} \boldsymbol{v}_{ij}^{gt T}\Big).
\end{aligned}$$

□

D.4.2 Proof of Prop. 5

First of all, we have

$$\boldsymbol{t}^{gt T} dLBdL \boldsymbol{t}^{gt} = \sum_{1 \le i,j \le n} \sum_{k \in \mathcal{N}_i} \sum_{l \in \mathcal{N}_j} w_{ik} w_{jl} (\boldsymbol{t}_i^{gt} - \boldsymbol{t}_k^{gt})^T dA_{ik} B_{ij} dA_{jl} (\boldsymbol{t}_j^{gt} - \boldsymbol{t}_l^{gt}).$$

Note that $dA_{ij}, (i,j) \in \mathcal{E}$ are independent. Therefore

$$\begin{aligned}
&\mathbb{E}_{\{\boldsymbol{v}_{ij}^{\mathrm{inp}}\}} \boldsymbol{t}^{gt T} dLBdL \boldsymbol{t}^{gt} = \\
&\mathbb{E}_{\{\boldsymbol{v}_{ij}^{\mathrm{inp}}\}} \Big(\sum_{1 \le i \le n} \sum_{k \in \mathcal{N}_i} (\boldsymbol{t}_i^{gt} - \boldsymbol{t}_k^{gt})^T w_{ik}^2 dA_{ik} B_{ii} dA_{ik} (\boldsymbol{t}_i^{gt} - \boldsymbol{t}_k^{gt}) \\
&- \sum_{1 \le i \le n} \sum_{k \in \mathcal{N}_i} w_{ik}^2 (\boldsymbol{t}_i^{gt} - \boldsymbol{t}_k^{gt})^T dA_{ik} B_{ik} dA_{ik} (\boldsymbol{t}_i^{gt} - \boldsymbol{t}_k^{gt})\Big) \\
&+ \mathbb{E}_{\{\boldsymbol{v}_{ij}^{\mathrm{inp}}\}} \sum_{1 \le i,j \le n} \sum_{k \in \mathcal{N}_i} \sum_{l \in \mathcal{N}_j} 1(i \ne j \vee k \ne l)(w_{ik} w_{jl}) \\
&(t_i^{gt} - t_k^{gt})^T dA_{ik} B_{ij} dA_{jl} (t_j^{gt} - t_l^{gt})\Big).
\end{aligned}$$

We have

$$\begin{aligned}
&\mathbb{E}_{\{\boldsymbol{v}_{ij}^{\mathrm{inp}}\}} \boldsymbol{t}^{gt T} dLBdL \boldsymbol{t}^{gt} = \sum_{(i,j) \in \mathcal{E}} w_{ij}^2 (\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt})^T \\
&\mathbb{E}_{\boldsymbol{v}_{ij}^{\mathrm{inp}}} dA_{ij} (B_{ii} + B_{jj} - B_{ij} - B_{ij}^T) dA_{ij} (\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}) \\
&+ \sum_{(i,k),(j,l) \in \mathcal{E}} 1(i \ne j \vee k \ne l) w_{ik} w_{jl} (t_i^{gt} - t_k^{gt})^T \\
&\mathbb{E}_{\boldsymbol{v}_{ik}^{\mathrm{inp}}} dA_{ik} (B_{ij} + B_{kl} - B_{il} - B_{jk}) \mathbb{E}_{\boldsymbol{v}_{jl}^{\mathrm{inp}}} dA_{jl} (t_j^{gt} - t_l^{gt}).
\end{aligned} \tag{30}$$

As $\boldsymbol{v}_{ij}^{gt} = \frac{\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}}{\|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|}$. Applying Prop. 4, we have

$$\begin{aligned}
&\mathbb{E}_{\{\epsilon_{ij,k}\}} \boldsymbol{t}^T dLBdL \boldsymbol{t} = \sum_{(i,j) \in \mathcal{E}} \frac{\sigma_{ij}^2}{2} w_{ij}^2 \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2 \Big(\mathrm{Tr}(B_{ii} + B_{jj} \\
&\quad - B_{ij} - B_{ij}^T) - \boldsymbol{v}_{ij}^{gt T} (B_{ii} + B_{jj} - B_{ij} - B_{ij}^T) \boldsymbol{v}_{ij}^{gt}\Big) \\
&+ \sum_{(i,k),(j,l) \in \mathcal{E}} 1(i \ne j \vee k \ne l)(w_{ik} w_{jl} \sigma_{ik}^2 \sigma_{jl}^2) \\
&(t_i^{gt} - t_k^{gt})^T ((B_{ij} + B_{kl} - B_{il} - B_{jk}))(t_j^{gt} - t_l^{gt})\Big) \\
&= \sum_{(i,j) \in \mathcal{E}} \frac{\sigma_{ij}^2}{2} w_{ij}^2 \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2 \Big(\mathrm{Tr}(E_{ij} B E_{ij}^T) \\
&\quad - \boldsymbol{v}_{ij}^{gt T} E_{ij} B E_{ij}^T \boldsymbol{v}_{ij}^{gt}\Big) \\
&\quad + \sum_{(i,k),(j,l) \in \mathcal{E}} w_{ik} w_{jl} \sigma_{ik}^2 \sigma_{jl}^2 (t_i^{gt} - t_k^{gt})^T E_{ik} B E_{jl}^T (t_j^{gt} - t_l^{gt})
\end{aligned} \tag{31}$$

□

D.5. Proof of Prop. 7

Note that

$$\frac{\partial \boldsymbol{u}_i^T W \boldsymbol{u}_4}{\partial \boldsymbol{v}} = 0.$$

Since $\boldsymbol{u}_i, 1 \le i \le 3$ do not depend on $\boldsymbol{v}$, we have

$$U_4^T W d\boldsymbol{u}_4 = 0. \quad (32)$$

Let the columns of $\overline{U}_4 \in \mathbb{R}^{3n\times(3n-4)}$ collect bases of vectors that orthogonal to U_4, i.e., $\overline{U}_4^T U_4 = 0$. Express

$$d\boldsymbol{u}_4 = \overline{U}_4 \boldsymbol{y} + U_4 \boldsymbol{x} \quad (33)$$

where $\boldsymbol{y} \in \mathbb{R}^{3n-4}$ and $\boldsymbol{x} \in \mathbb{R}^4$. Combining Eq. (33) and Eq. (32), we have

$$\boldsymbol{x} = -(U_4^T W U_4)^{-1} U_4^T W \overline{U}_4 \boldsymbol{y} = -U_4^T W \overline{U}_4 \boldsymbol{y}. \quad (34)$$

This means

$$d\boldsymbol{u}_4 = (I_{3n} - U_4 U_4^T W)\overline{U}_4 \boldsymbol{y}. \quad (35)$$

Consider the equality

$$L\boldsymbol{u}_4 = \lambda_4 W \boldsymbol{u}_4.$$

Compute the derivatives of both sides with respect to $\boldsymbol{v}$, we have

$$Ld\boldsymbol{u}_4 + dL\boldsymbol{u}_4 = d\lambda_4 W \boldsymbol{u}_4. \quad (36)$$

The derivative of the eigen-value is given by

$$d\lambda_4 = \boldsymbol{u}_4^T dL \boldsymbol{u}_4. \quad (37)$$

Substituting Eq. (37) into Eq. (36), we have

$$Ld\boldsymbol{u}_4 + dL\boldsymbol{u}_4 = W\boldsymbol{u}_4\boldsymbol{u}_4^T dL\boldsymbol{u}_4 \quad (38)$$

Multiply both sides of Eq. (38) by $\overline{U}_4^T$ and combine Eq. (33), we arrive at

$$\overline{U}_4^T L \overline{U}_4 \boldsymbol{y} + \overline{U}_4^T dL \boldsymbol{u}_4 = 0. \quad (39)$$

Substituting Eq. (39) into Eq. (35), we have

$$d\boldsymbol{u}_4 = -(I_{3n} - U_4 U_4^T W)\overline{U}_4 \big(\overline{U}_4^T L \overline{U}_4\big)^{-1} \overline{U}_4^T dL\boldsymbol{u}_4. \quad (40)$$

Note that

$$L^{\dagger} = \overline{U}_4 \big(\overline{U}_4^T L \overline{U}_4\big)^{-1} \overline{U}_4^T.$$

□

D.5.1 Proof of Prop. 8

We can expand

$$\sum_{(i,j)\in\mathcal{E}} \sum_{(i',j')\in\mathcal{E}} x_{ij} x_{i'j'} \frac{\partial^2 f}{\partial w_{ij}\partial w_{i'j'}} = H_1 - H_2 + H_3 - H_4 + H_5 + H_6 \quad (41)$$

where

$$H_1 = \sum_{(i,j)\in\mathcal{E}} \sum_{(i',j')\in\mathcal{E}} x_{ij}^2 \mathrm{Tr}\big(L^{\dagger} \frac{\partial^2 \overline{L}}{\partial^2 w_{ij}} L^{\dagger}\big),$$

$$H_2 = \sum_{(i,j)\in\mathcal{E}} \sum_{(i',j')\in\mathcal{E}} x_{ij} x_{i'j'} \mathrm{Tr}\big(L^{\dagger} \frac{\partial L}{\partial w_{i'j'}} L^{\dagger} \frac{\partial \overline{L}}{\partial w_{ij}} L^{\dagger} + L^{\dagger} \frac{\partial \overline{L}}{\partial w_{ij}} L^{\dagger} \frac{\partial L}{\partial w_{i'j'}} L^{\dagger}\big),$$

$$H_3 = \sum_{(i,j)\in\mathcal{E}} \sum_{(i',j')\in\mathcal{E}} x_{ij} x_{i'j'} \mathrm{Tr}\big(L^{\dagger} \frac{\partial L}{\partial w_{i'j'}} L^{\dagger}\overline{L}L^{\dagger} \frac{\partial L}{\partial w_{ij}} L^{\dagger} + L^{\dagger} \frac{\partial L}{\partial w_{ij}} L^{\dagger}\overline{L}L^{\dagger} \frac{\partial L}{\partial w_{i'j'}} L^{\dagger}\big),$$

$$H_4 = \sum_{(i,j)\in\mathcal{E}} \sum_{(i',j')\in\mathcal{E}} x_{ij} x_{i'j'} \mathrm{Tr}\big(L^{\dagger} \frac{\partial \overline{L}}{\partial w_{i'j'}} L^{\dagger} \frac{\partial L}{\partial w_{ij}} L^{\dagger} + L^{\dagger} \frac{\partial L}{\partial w_{ij}} L^{\dagger} \frac{\partial \overline{L}}{\partial w_{i'j'}} L^{\dagger}\big),$$

$$H_5 = \sum_{(i,j)\in\mathcal{E}} \sum_{(i',j')\in\mathcal{E}} x_{ij} x_{i'j'} \mathrm{Tr}\Big(L^{\dagger}\overline{L}L^{\dagger} \frac{\partial L}{\partial w_{i'j'}} L^{\dagger} \frac{\partial L}{\partial w_{ij}} L^{\dagger} + L^{\dagger} \frac{\partial L}{\partial w_{ij}} L^{\dagger} \frac{\partial L}{\partial w_{i'j'}} L^{\dagger}\overline{L}L^{\dagger}\big),$$

$$H_6 = \sum_{(i,j)\in\mathcal{E}} \sum_{(i',j')\in\mathcal{E}} x_{ij} x_{i'j'} \mathrm{Tr}\big(L^{\dagger}\overline{L}L^{\dagger} \frac{\partial L}{\partial w_{ij}} L^{\dagger} \frac{\partial L}{\partial w_{i'j'}} L^{\dagger} + L^{\dagger} \frac{\partial L}{\partial w_{i'j'}} L^{\dagger} \frac{\partial L}{\partial w_{ij}} L^{\dagger}\overline{L}L^{\dagger}\big).$$

Introduce

$$D_2 = I_2 \otimes \mathrm{diag}(\frac{x_{ij}}{w_{ij}}).$$

We have

$$H_1 = 2\sum_{(i,j)\in\mathcal{E}} \frac{x_{ij}^2}{w_{ij}} \sigma_{ij}^2 \|\boldsymbol{t}_i^{gt} - \boldsymbol{t}_j^{gt}\|^2 \mathrm{Tr}\big(L^{\dagger} E_{ij} w_{ij} (\boldsymbol{v}_{ij,1}^{gt} {\boldsymbol{v}_{ij,1}^{gt}}^T + \boldsymbol{v}_{ij,2}^{gt} {\boldsymbol{v}_{ij,2}^{gt}}^T) E_{ij}^T L^{\dagger}\big) = 2\mathrm{Tr}\big(L^{\dagger} G^T D_2^2 D_4 G L^{\dagger}\big) = 2\mathrm{Tr}\big(G^{\dagger} D_2^2 D_4 {G^{\dagger}}^T\big). \quad (42)$$

Through similar calculations, we have

$$H_2 = H_4 = 4\mathrm{Tr}\big(G^{\dagger} D_2 G G^{\dagger} D_2 D_4 {G^{\dagger}}^T\big), \quad (43)$$

and

$$H_3 = 2\text{Tr}\big(G^\dagger D_2 {G^\dagger}^T G^T D_4 = GG^\dagger D_2 {G^\dagger}^T\big), \quad (44)$$

and

$$H_5 = H_6 = 2\text{Tr}\big(G^\dagger D_4 G G^\dagger D_2 {G^\dagger}^T G^T D_2 {G^\dagger}^T\big). \quad (45)$$

Introduce orthonormal matrix $\overline{U} \in \mathbb{R}^{2|\mathcal{E}|\times(2|\mathcal{E}|-3n+4)}$ where

$$I - \overline{U}\overline{U}^T = GG^\dagger.$$

Substituting Eq. (42), Eq. (43), Eq. (44), and Eq. (45) into Eq. (41), we have

$$\sum_{(i,j)\in\mathcal{E}}\sum_{(i',j')\in\mathcal{E}} x_{ij}x_{i'j'}\frac{\partial^2 f}{\partial w_{ij}\partial w_{i'j'}} = 2\text{Tr}\big(G^\dagger F {G^\dagger}^T\big) \quad (46)$$

where

$$\begin{aligned}
F &= D_2^2 D_4 - 2D_2(I-\overline{U}\overline{U}^T)D_2D_4 - 2D_2D_4(I-\overline{U}\overline{U}^T)D_2 \\
&+ D_2(I-\overline{U}\overline{U}^T)D_4(I-\overline{U}\overline{U}^T)D_2 \\
&+ D_4(I-\overline{U}\overline{U}^T)D_2(I-\overline{U}\overline{U}^T)D_2 \\
&+ D_2(I-\overline{U}\overline{U}^T)D_2(I-\overline{U}\overline{U}^T)D_4 \\
&= D_2\overline{U}\overline{U}^T D_4 \overline{U}\overline{U}^T D_2 - D_2(I-\overline{U}\overline{U}^T)D_2\overline{U}\overline{U}^T D_4 \\
&- D_4\overline{U}\overline{U}^T D_2(I-\overline{U}\overline{U}^T)D_2.
\end{aligned}$$

When $w_{ij} = \frac{1}{\sigma_{ij}^2\|\boldsymbol{t}_i^{gt}-\boldsymbol{t}_j^{gt}\|^2}$, we have D_4. As $\overline{U}^T G^\dagger = 0$, we have

$$\text{Tr}\big(G^\dagger F {G^\dagger}^T\big) = \|G^\dagger D_2 \overline{U}\overline{U}^T\|_{\mathcal{F}}^2 = \|G^\dagger D_2 \overline{U}\|_{\mathcal{F}}^2$$

It is clear that $\text{Tr}\big(G^\dagger F {G^\dagger}^T\big) \geq 0$ and equality holds if and only if $D_2 = sI_{2|\mathcal{E}|}$ □

D.5.2 Proof Prop. 9

We prove a stronger result.

Lemma 1. *Consider a unitary matrix $U \in \mathbb{R}^{2m\times n}$ where $2m > n$. Let $\Sigma = \text{diag}(\sigma_i) \in R^{n\times n}$ be a diagonal matrix with all positive values. Consider a vector $\boldsymbol{x} \in \mathbb{R}^m$. Suppose $\forall 1 \leq i \leq m$,*

$$Tr\Big((I_2\otimes \boldsymbol{e}_i)^T U\Sigma U^T (I_2\otimes \text{diag}(\boldsymbol{x}))(I-UU^T)(I_2\otimes \boldsymbol{e}_i\Big) = 0, \quad (47)$$

then $\boldsymbol{x} = s\mathbf{1}$.

Proof. Denote $U = (U_1; U_2)$. Then Eq. (47) is equivalent to

$$\Big((U_1\Sigma U_1^T).(I_m - U_1U_1^T) + (U_2\Sigma U_2^T).(I_m - U_2U_2^T) - (U_1\Sigma U_2^T).(U_1U_2^T) - (U_2\Sigma U_1^T).(U_2U_1^T)\Big)\boldsymbol{x} = 0. \quad (48)$$

where $A.B$ is the element-wise matrix multiplication operation.

Denote

$$\begin{aligned}
A =& (U_1\Sigma U_1^T).(I_m - U_1U_1^T) + (U_2\Sigma U_2^T).(I_m - U_2U_2^T) \\
-& \quad (U_1\Sigma U_2^T).(U_1U_2^T) - (U_2\Sigma U_1^T).(U_2U_1^T).
\end{aligned}$$

We show that $\boldsymbol{x}^T A\boldsymbol{x} \geq 0$ and equality holds if and only if $\boldsymbol{x} = s\mathbf{1}$. Let i-th row of U_1 and U_2 as $\boldsymbol{u}_{1i}$ and $\boldsymbol{u}_{2i}$. It follows that

$$\begin{aligned}
\boldsymbol{x}^T A\boldsymbol{x} &= \sum_{i=1}^m x_i^2\big(\boldsymbol{u}_{1i}\Sigma\boldsymbol{u}_{1i}^T + \boldsymbol{u}_{2i}\Sigma\boldsymbol{u}_{2i}^T\big) - \sum_{i=1}^m\sum_{j=1}^m \\
& x_ix_j\Big((\boldsymbol{u}_{1i}\Sigma\boldsymbol{u}_{1j}^T)(\boldsymbol{u}_{1i}\boldsymbol{u}_{1j}^T) + (\boldsymbol{u}_{2i}\Sigma\boldsymbol{u}_{2j}^T)(\boldsymbol{u}_{2i}\boldsymbol{u}_{2j}^T) \\
&+ (\boldsymbol{u}_{1i}\Sigma\boldsymbol{u}_{2j}^T)(\boldsymbol{u}_{1i}\boldsymbol{u}_{2j}^T) + (\boldsymbol{u}_{2i}\Sigma\boldsymbol{u}_{1j}^T)(\boldsymbol{u}_{2i}\boldsymbol{u}_{1j}^T)\Big) \\
&\geq \sum_{i=1}^m x_i^2\big(\boldsymbol{u}_{1i}\Sigma\boldsymbol{u}_{1i}^T + \boldsymbol{u}_{2i}\Sigma\boldsymbol{u}_{2i}^T\big) - \sum_{i=1}^m\sum_{j=1}^m \\
& \frac{x_i^2+x_j^2}{2}\Big((\boldsymbol{u}_{1i}\Sigma\boldsymbol{u}_{1j}^T)(\boldsymbol{u}_{1i}\boldsymbol{u}_{1j}^T) + (\boldsymbol{u}_{2i}\Sigma\boldsymbol{u}_{2j}^T)(\boldsymbol{u}_{2i}\boldsymbol{u}_{2j}^T) \\
&+ (\boldsymbol{u}_{1i}\Sigma\boldsymbol{u}_{2j}^T)(\boldsymbol{u}_{1i}\boldsymbol{u}_{2j}^T) + (\boldsymbol{u}_{2i}\Sigma\boldsymbol{u}_{1j}^T)(\boldsymbol{u}_{2i}\boldsymbol{u}_{1j}^T)\Big)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^m x_i^2\big(\boldsymbol{u}_{1i}\Sigma\boldsymbol{u}_{1i}^T + \boldsymbol{u}_{2i}\Sigma\boldsymbol{u}_{2i}^T\big) - \sum_{i=1}^m x_i^2\Big(\boldsymbol{u}_{1i}\Sigma \\
&(\sum_{j=1}^m \boldsymbol{u}_{1j}^T\boldsymbol{u}_{1j})\boldsymbol{u}_{1i}^T + \boldsymbol{u}_{2i}\Sigma(\sum_{j=1}^m \boldsymbol{u}_{2j}^T\boldsymbol{u}_{2j})\boldsymbol{u}_{2i}^T \\
&+ \boldsymbol{u}_{1i}\Sigma(\sum_{j=1}^m \boldsymbol{u}_{2j}^T\boldsymbol{u}_{2j})\boldsymbol{u}_{1i}^T + \boldsymbol{u}_{2i}\Sigma(\sum_{j=1}^m \boldsymbol{u}_{1j}^T\boldsymbol{u}_{jj})\boldsymbol{u}_{2i}^T\Big) \\
&= \sum_{i=1}^m x_i^2\big(\boldsymbol{u}_{1i}\Sigma\boldsymbol{u}_{1i}^T + \boldsymbol{u}_{2i}\Sigma\boldsymbol{u}_{2i}^T\big) - \sum_{i=1}^m x_i^2\big(\boldsymbol{u}_{1i}\Sigma \\
&(U_1^TU_1 + U_2^TU_2)\boldsymbol{u}_{1i}^T + \boldsymbol{u}_{2i}\Sigma(U_1^TU_1 + U_2^TU_2)\boldsymbol{u}_{2i}^T\big) = 0.
\end{aligned}$$

and equality holds if and only if $x_i = x_j, \forall i \neq j$. □

E. Proof of Theorem 3

We begin with key lemmas regarding general-purpose stability results of eigen-values and eigen-vectors and matrix-norms in Section E.1. Section E.2 complete the proof of Theorem 3. Section E.3 presents proofs of the lemmas in Section E.1.

E.1. Key Lemmas

We first present two lemmas regarding the stability of eigen-values and eigen-vectors. Suppose that the measurement with edge $(i,j) \in \mathcal{E}$ is $\boldsymbol{v}_{ij}^{\text{inp}}$, and the underlying ground truth is $\boldsymbol{v}_{ij}^{gt}$. The edge weight is $w_{ij} \in [0,1]$. Let $\boldsymbol{w} = (w_{ij})$. We define

$$L(\boldsymbol{w}) = \sum_{(i,j)\in\mathcal{E}} w_{ij} E_{ij}(I_3 - \boldsymbol{v}_{ij}^{\text{inp}} {\boldsymbol{v}_{ij}^{\text{inp}}}^T) E_{ij}^T,$$
$$L^{gt}(\boldsymbol{w}) = \sum_{(i,j)\in\mathcal{E}} w_{ij} E_{ij}(I_3 - \boldsymbol{v}_{ij}^{gt} {\boldsymbol{v}_{ij}^{gt}}^T) E_{ij}^T,$$
$$dL(\boldsymbol{w}) = \sum_{(i,j)\in\mathcal{E}} w_{ij} E_{ij}(\boldsymbol{v}_{ij}^{gt} {\boldsymbol{v}_{ij}^{gt}}^T - \boldsymbol{v}_{ij}^{\text{inp}} {\boldsymbol{v}_{ij}^{\text{inp}}}^T) E_{ij}^T.$$

It is clear that $L(\boldsymbol{w}) = L^{gt}(\boldsymbol{w}) + dL(\boldsymbol{w})$. Consider

$$L(\boldsymbol{w})\boldsymbol{u}_4 = \lambda_4 \boldsymbol{u}_4, \qquad L^{gt}(\boldsymbol{w})\boldsymbol{u}_4^{gt} = \lambda_4^{gt}\boldsymbol{u}^{gt}.$$

The first lemma characterizes an upper bound of λ_4 using $dL(\boldsymbol{w})$

Lemma 2. *We have,*

$$\lambda_4 \leq {\boldsymbol{u}^{gt}}^T dL(\boldsymbol{w})\boldsymbol{u}^{gt}. \tag{49}$$

Proof: See Section E.3.1. □

Denote $\boldsymbol{t}_{ij} = \boldsymbol{u}_{4i} - \boldsymbol{u}_{4j}$ and $\boldsymbol{t}_{ij}^{gt} = \boldsymbol{u}_{4i}^{gt} - \boldsymbol{u}_{4j}^{gt}$. The next lemma provides an upper bound on $\|\boldsymbol{t}_{ij} - \boldsymbol{t}_{ij}^{gt}\|$.

Lemma 3. *Suppose* $\|dL(\boldsymbol{w})\| \leq \frac{\lambda_5^{gt}}{3}$ *and*

$$\|L^{gt}(\boldsymbol{w})^{\dagger} dL(\boldsymbol{w})\|_1 + \|L^{gt}(\boldsymbol{w})^{\dagger}\|_1 {\boldsymbol{t}^{gt}}^T dL(\boldsymbol{w})\boldsymbol{t}^{gt} < 1.$$

Then

$$\|\boldsymbol{t}_{ij} - \boldsymbol{t}_{ij}^{gt}\| \leq (1-\beta)\frac{\|E_{ij}^T L^{gt}(\boldsymbol{w})^{\dagger}\|_1 \|dL(\boldsymbol{w})\boldsymbol{t}^{gt}\|_{\infty}}{\alpha} + \beta\|\boldsymbol{t}_{ij}^{gt}\|, \tag{50}$$

and

$$\|\boldsymbol{t} - \boldsymbol{t}^{gt}\|_{\infty} \leq (1-\beta)\frac{\|L^{gt}(\boldsymbol{w})^{\dagger}\|_1 \|dL(\boldsymbol{w})\boldsymbol{t}^{gt}\|_{\infty}}{\alpha} + \beta\|\boldsymbol{t}^{gt}\|_{\infty}, \tag{51}$$

where

$$\beta \leq \frac{\|dL(\boldsymbol{w})\|^2}{2(\lambda_5^{gt} - {\boldsymbol{u}_4^{gt}}^T dL(\boldsymbol{w})\boldsymbol{u}_4^{gt} - \|dL(\boldsymbol{w})\|)^2},$$
$$\alpha = 1 - \left({\boldsymbol{u}_4^{gt}}^T dL(\boldsymbol{w})\boldsymbol{u}_4^{gt} + \|dL(\boldsymbol{w})\|_1\right)\|L^{gt}(\boldsymbol{w})^{\dagger}\|_1.$$

Proof: See Section E.3.2. □

We proceed to bound $\|L^{gt}(\boldsymbol{w})^{\dagger}\|_1$ and $\|E_{ij}^T L^{gt}(\boldsymbol{w})^{\dagger}\|_1$ with respect to references $\|L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1$ and $\|E_{ij}^T L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1$ where $\overline{\boldsymbol{w}}$ is some reference edge vector.

Lemma 4. *Suppose* $\lambda_5(L^{gt}(\boldsymbol{w})) > 0$ *and* $\lambda_5(L^{gt}(\overline{\boldsymbol{w}})) > 0$. *Then*

$$\|L^{gt}(\boldsymbol{w})^{\dagger}\|_1 \leq \frac{\|L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1}{1 - \|L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1 c(d\boldsymbol{w})} \tag{52}$$
$$\|E_{ij}^T L^{gt}(\boldsymbol{w})^{\dagger}\|_1 \leq \frac{\|E_{ij}^T L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1}{1 - \|L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1 c(d\boldsymbol{w})} \tag{53}$$

where

$$c(d\boldsymbol{w}) = \max_{1\leq i\leq n} \sum_{j\in\mathcal{N}_i} |w_{ij} - \overline{w}_{ij}|.$$

Proof: See Section E.3.3. □

We then provide two L^{∞} bounds on $dL(\boldsymbol{w})$.

Lemma 5.

$$\|dL(\boldsymbol{w})\boldsymbol{t}^{gt}\|_{\infty} \leq \max_{1\leq i\leq n} \left(\epsilon\delta_i^{\text{in}}(\boldsymbol{w}) + \delta_i^{\text{out}}(\boldsymbol{w})\right) \tag{54}$$
$$\|dL(\boldsymbol{w})\|_1 \leq \max_{1\leq i\leq n} \Big(\epsilon \sum_{j\in\mathcal{N}_i^{\text{in}}} w_{ij} + \sum_{j\in\mathcal{N}_i^{\text{out}}} w_{ij} \|\boldsymbol{v}_{ij}^{\text{inp}} {\boldsymbol{v}_{ij}^{\text{inp}}}^T - \boldsymbol{v}_{ij}^{gt} {\boldsymbol{v}_{ij}^{gt}}^T\|\Big) \tag{55}$$

where

$$\delta_i^{\text{in}}(\boldsymbol{w}) = \sum_{j\in\mathcal{N}_i^{\text{in}}} w_{ij}\|\boldsymbol{t}_{ij}^{gt}\|$$
$$\delta_i^{\text{out}}(\boldsymbol{w}) = \sum_{j\in\mathcal{N}_i^{\text{out}}} w_{ij}\|(\boldsymbol{v}_{ij}^{\text{inp}} {\boldsymbol{v}_{ij}^{\text{inp}}}^T - \boldsymbol{v}_{ij}^{gt} {\boldsymbol{v}_{ij}^{gt}}^T)\boldsymbol{t}_{ij}^{gt}\|$$

Proof: See Section E.3.4. □

Next, we provide two spectral norms of $dL(\boldsymbol{w})$.

Lemma 6.

$${\boldsymbol{t}^{gt}}^T dL(\boldsymbol{w})\boldsymbol{t}^{gt} \leq \epsilon^2 \sum_{(i,j)\in\mathcal{E}^{\text{in}}} w_{ij}\|\boldsymbol{t}_{ij}^{gt}\|^2 + \sum_{(i,j)\in\mathcal{E}^{\text{out}}} w_{ij}\|\boldsymbol{t}_{ij}^{gt}\|^2 \|\boldsymbol{v}_{ij}^{\text{inp}} {\boldsymbol{v}_{ij}^{\text{inp}}}^T - \boldsymbol{v}_{ij}^{gt} {\boldsymbol{v}_{ij}^{gt}}^T\|, \tag{56}$$

and

$$\|dL(\boldsymbol{w})\| \leq \max_{1\leq i\leq n} \Big(\epsilon \sum_{j\in\mathcal{N}_i^{\text{in}}} w_{ij} + \sum_{j\in\mathcal{N}_i^{\text{out}}} w_{ij} \|\boldsymbol{v}_{ij}^{\text{inp}} {\boldsymbol{v}_{ij}^{\text{inp}}}^T - \boldsymbol{v}_{ij}^{gt} {\boldsymbol{v}_{ij}^{gt}}^T\|\Big) \tag{57}$$

Proof: See Section E.3.5. □

Finally, we present two lemmas which are used to control w_{ij} during the alternating procedure.

Lemma 7. *Consider a hyper-parameters $\eta < 1$. Define*

$$b_l(\eta, \boldsymbol{t}_{ij}^{gt}, \boldsymbol{v}_{ij}^{\text{inp}}) := \min_{\|\boldsymbol{t}_{ij}-\boldsymbol{t}_{ij}^{gt}\|\leq\eta\|\boldsymbol{t}_{ij}^{gt}\|} \|\boldsymbol{v}_{ij}^{\text{inp}} - \boldsymbol{v}_{ij}\|^2\|\boldsymbol{t}_{ij}\|^2.$$

Then $b_l(\eta, \boldsymbol{t}_{ij}^{gt}, \boldsymbol{v}_{ij}^{\text{inp}}) = 0$ when

$$\|(I_3 - \boldsymbol{v}_{ij}^{\text{inp}}\boldsymbol{v}_{ij}^{\text{inp}T})\boldsymbol{v}_{ij}^{gt}\| \leq \eta.$$

Otherwise,

$$b_l(\eta, \boldsymbol{t}_{ij}^{gt}, \boldsymbol{v}_{ij}^{\text{inp}}) \geq 4(1-\eta)^2\|\boldsymbol{t}_{ij}^{gt}\|^2 \sin^2(\frac{\phi_1 - \phi_2}{2})$$

where

$$\phi_1 = \text{acos}(\boldsymbol{v}_{ij}^{\text{inp}T}\boldsymbol{v}_{ij}^{gt}), \quad \phi_2 = \text{asin}(\eta).$$

Proof: See Section E.4. □

Lemma 8. *Consider a hyper-parameter $\eta \leq 1$. Define*

$$b_u(\eta, \boldsymbol{t}_{ij}^{gt}, \boldsymbol{v}_{ij}^{\text{inp}}) := \max_{\|\boldsymbol{t}_{ij}-\boldsymbol{t}_{ij}^{gt}\|\leq\eta\|\boldsymbol{t}_{ij}^{gt}\|} \|\boldsymbol{v}_{ij}^{\text{inp}} - \boldsymbol{v}_{ij}\|^2\|\boldsymbol{t}_{ij}\|^2.$$

Then

$$\begin{aligned} b_u(\eta, \boldsymbol{t}_{ij}^{gt}, \boldsymbol{v}_{ij}^{\text{inp}}) \leq & 4\sin^2(\min(\frac{\pi}{2}, \frac{\phi_1+\phi_2}{2})) \\ & (1+\eta)^2\|\boldsymbol{t}_{ij}^{gt}\|)^2 \end{aligned}$$

Proof: See Section E.5. □

E.2. Complete the Proof of Theorem 3

Our proof is based on the eigen stability results in Lemma 2 and Lemma 3 and the bounds in Lemma 4 to Lemma 8.

As our goal is to show the robustness of our algorithm against outliers, our proof do not aim to provide tight values of c_1, c_2, c_3. Define

$$r = \max_{(i,j)\in\mathcal{E}} \|\boldsymbol{t}_{ij}^{gt}\| / \min_{(i,j)\in\mathcal{E}} \|\boldsymbol{t}_{ij}^{gt}\|.$$

We assume that the value of r is not super big.

We show that the iterative procedure converges to a local minimum that is sufficiently close to $\boldsymbol{t}^{gt}$. Denote

$$\begin{aligned} \delta(\epsilon, \boldsymbol{w}) := \max_{1\leq i\leq n} \Big(\epsilon \sum_{j\in\mathcal{N}_i^{\text{in}}} w_{ij} + \\ \sum_{j\in\mathcal{N}_i^{\text{out}}} w_{ij}\|\boldsymbol{v}_{ij}^{\text{inp}}\boldsymbol{v}_{ij}^{\text{inp}T} - \boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T}\|\Big)\|. \end{aligned}$$

Applying Lemma 5, we have $\forall(i,j) \in \mathcal{E}$,

$$\|dL(\boldsymbol{w})\boldsymbol{t}^{gt}\|_\infty \leq r\delta(\epsilon, \boldsymbol{w})\|\boldsymbol{t}_{ij}^{gt}\|,$$

and

$$\|dL(\boldsymbol{w})\| \leq \|dL(\boldsymbol{w})\boldsymbol{t}^{gt}\|_1 \leq \delta(\epsilon, \boldsymbol{w}).$$

This means α and β in Lemma 3 satisfy

$$\alpha \geq 1 - 2\delta(\epsilon, \boldsymbol{w})\|L^{gt}(\boldsymbol{w})^\dagger\|_1, \tag{58}$$

$$\beta \leq \frac{\delta(\epsilon, \boldsymbol{w})^2}{2(1 - 2\delta(\epsilon, \boldsymbol{w}))^2}. \tag{59}$$

Applying Lemma 3, we have

$$\frac{\|\boldsymbol{t}_{ij} - \boldsymbol{t}_{ij}^{gt}\|}{\|\boldsymbol{t}_{ij}^{gt}\|} \leq \beta + \frac{1-\beta}{\alpha}\|E_{ij}^T L^{gt}(\boldsymbol{w})^+\|_1\delta(\epsilon, \boldsymbol{w})$$

It is clear that we can choose c_1, c_2, c_3 so that the output of the first iteration of our algorithm satisfies

$$\|\boldsymbol{t}_{ij} - \boldsymbol{t}_{ij}^{gt}\| \leq \frac{1}{4}\|\boldsymbol{t}_{ij}^{gt}\|, \quad \alpha \geq \frac{7}{8}.$$

Applying Lemma 7 and Lemma 8 and after some calculation, we have that the edge weights w_{ij} converge to the neighborhood of

$$\overline{w}_{ij} = \frac{\sigma_{\min}^2}{\sigma_{\min}^2 + \|\boldsymbol{v}_{ij}^{\text{inp}} - \boldsymbol{v}_{ij}^{gt}\|^2\|\boldsymbol{t}_{ij}^{gt}\|^2}.$$

Theorem 3 then follows Lemma 3 and Lemma 4.

□

E.3. Proof of Lemmas in Section E.1

E.3.1 Proof of Lemma 2

Note the following variational definition of

$$\lambda_4 = \min_{\boldsymbol{x}\in\mathbb{R}^{3n}, \|\boldsymbol{x}\|=1, (\boldsymbol{1}\otimes I_3)^T\boldsymbol{x}=0} \boldsymbol{x}^T L\boldsymbol{x}. \tag{60}$$

As $(\boldsymbol{1} \otimes I_3)^T\boldsymbol{u}_4^{gt} = 0$, we have

$$\begin{aligned} \lambda_4 \leq & \boldsymbol{u}_4^{gt\,T} L\boldsymbol{u}_4^{gt} \\ = & \boldsymbol{u}_4^{gt\,T}(L^{gt} + dL)\boldsymbol{u}_4^{gt} = \boldsymbol{u}_4^{gt\,T} dL\boldsymbol{u}_4^{gt}. \end{aligned}$$

□

E.3.2 Proof of Lemma 3

Let $\overline{U}_4^{gt} \in \mathbb{R}^{3n\times(3n-4)}$ collect the 5-th to $3n$-th eigenvectors L^{gt}. As $(\boldsymbol{1} \otimes I_3)^T\boldsymbol{u}_4 = (\boldsymbol{1} \otimes I_3)^T\boldsymbol{u}_4^{gt} = \boldsymbol{0}$, we can express

$$d\boldsymbol{u} = \boldsymbol{u}_4 - \boldsymbol{u}_4^{gt} = -x\boldsymbol{u}_4^{gt} + \overline{U}_4^{gt}\boldsymbol{y}. \tag{61}$$

As $\|\boldsymbol{u}_4\|=1$, we have

$$(1-x)^2+\|\boldsymbol{y}\|^2=1. \quad (62)$$

The following proposition describes the formula for $\overline{U}_4^{gt}\boldsymbol{y}$

Proposition 10. *Suppose* $\|dL\|+\boldsymbol{u}_4^{gt^T}dL\boldsymbol{u}_4^{gt}<\lambda_5^{gt}$. *Then*

$$\overline{U}_4^{gt}\boldsymbol{y}=-(1-x)\big(I+L^\dagger(\lambda_4)dL\big)^{-1}L^\dagger(\lambda_4)dL\boldsymbol{u}_4^{gt}. \quad (63)$$

where

$$L^\dagger(\lambda_4)=\overline{U}_4^{gt}(\Lambda-\lambda_4 I)^{-1}\overline{U}_4^{gt^T}.$$

Proof:
First of all, from $(L^{gt}+dL)(\boldsymbol{u}_4^{gt}+d\boldsymbol{u})=\lambda_4(\boldsymbol{u}_4^{gt}+d\boldsymbol{u})$, we have

$$(L^{gt}+dL-\lambda_4 I)d\boldsymbol{u}=(\lambda_4 I-dL)\boldsymbol{u}_4^{gt}. \quad (64)$$

Substituting Eq. (61) into Eq. (64) and multiplying both sides by $\overline{U}_4^{gt^T}$, we arrive at

$$\big(\Lambda-\lambda_4 I+\overline{U}_4^{gt^T}dL\overline{U}_4^{gt}\big)\boldsymbol{y}=-(1-x)\overline{U}_4^{gt^T}dL\boldsymbol{u}_4^{gt} \quad (65)$$

which means

$$\boldsymbol{y}=-(1-x)\big(\Lambda-\lambda_4 I+\overline{U}_4^{gt^T}dL\overline{U}_4^{gt}\big)^{-1}\overline{U}_4^{gt^T}dL\boldsymbol{u}_4^{gt}.$$

Note that

$$\begin{aligned}&\overline{U}_4^{gt}\big(\Lambda-\lambda_4 I+\overline{U}_4^{gt^T}dL\overline{U}_4^{gt}\big)^{-1}\overline{U}_4^{gt^T}dL\boldsymbol{u}_4^{gt}\\=&\overline{U}_4^{gt}(\Lambda-\lambda_4 I)^{-\frac{1}{2}}\big(I+(\Lambda-\lambda_4 I)^{-\frac{1}{2}}\overline{U}_4^{gt^T}dL\overline{U}_4^{gt}\\&(\Lambda-\lambda_4 I)^{-\frac{1}{2}}\big)^{-1}(\Lambda-\lambda_4 I)^{-\frac{1}{2}}\overline{U}_4^{gt^T}dL\boldsymbol{u}_4^{gt}\\=&(I+L^\dagger(\lambda_4)dL)^{-1}L^\dagger(\lambda_4)dL\boldsymbol{u}_4^{gt}\end{aligned}$$

□

The following proposition provides an upper bound on x.

Proposition 11. *Suppose* $\|dL\|\leq\frac{\lambda_5^{gt}}{3}$, *we have*

$$x\leq\frac{\|dL\|^2}{2(\lambda_5^{gt}-\lambda_4-\|dL\|)^2}. \quad (66)$$

Proof: Denote

$$\alpha=\|\big(I+L^\dagger(\lambda_4)dL\big)^{-1}L^\dagger(\lambda_4)dL\boldsymbol{u}_4^{gt}\|,$$

then

$$(1-x)^2=\frac{1}{1+\alpha^2}. \quad (67)$$

Note that

$$\|L^\dagger(\lambda_4)dL\|\leq\frac{\|dL\|}{\lambda_5^{gt}-\lambda_4}.$$

It follows that

$$\alpha\leq\frac{\|dL\|}{\lambda_5^{gt}-\lambda_4-\|dL\|}. \quad (68)$$

As $\|dL\|\frac{1}{3}\lambda_5^{gt}$, we have

$$\alpha\leq\frac{1}{2}.$$

Substituting (68) into (67), we have

$$\begin{aligned}x\leq&1-\frac{1}{\sqrt{1+\alpha^2}}\leq\frac{\alpha^2}{2}\\\leq&\frac{\|dL\|^2}{2(\lambda_5^{gt}-\lambda_4-\|dL\|)^2}\end{aligned} \quad (69)$$

□

We have

$$\begin{aligned}\|\boldsymbol{t}_{ij}-\boldsymbol{t}_{ij}^{gt}\|=&\|E_{ij}^T d\boldsymbol{u}\|\leq x\|E_{ij}^T\boldsymbol{u}_4^{gt}\|+(1-x)\\&\|E_{ij}^T L^\dagger(\lambda_4)dL(I+L^\dagger(\lambda_4)dL)^{-1}\boldsymbol{u}_4^{gt}\|\\\leq&x\|\boldsymbol{t}_{ij}^{gt}\|+(1-x)\|E_{ij}^T L^\dagger(\lambda_4)\\&\sum_{i=0}^{+\infty}(-dLL^\dagger(\lambda_4))^i dL\boldsymbol{u}_4^{gt}\|_1\\\leq&x\|\boldsymbol{t}_{ij}^{gt}\|+(1-x)\frac{\|E_{ij}^T L^\dagger(\lambda_4)\|_1\|dL\boldsymbol{u}_4^{gt}\|_\infty}{1-\|L^\dagger(\lambda_4)dL\|_1}\end{aligned} \quad (70)$$

Note that

$$\begin{aligned}\|E_{ij}^T L^\dagger(\lambda_4)\|_1=&\|\sum_{i=0}^{+\infty}E_{ij}^T\lambda_4^i(L^\dagger)^{i+1}\|_1\\\leq&\|E_{ij}^T L^\dagger\|_1\sum_{i=0}^{+\infty}(\lambda_4\|L^\dagger\|_1)^i\\=&\frac{\|E_{ij}^T L^\dagger\|_1}{1-\lambda_4\|L^\dagger\|_1}\end{aligned} \quad (71)$$

Similarly,

$$\begin{aligned}\|L^\dagger(\lambda_4)dL\|_1=&\|\sum_{i=0}^{+\infty}\lambda_4^i L^\dagger)^{i+1}dL\|_1\\\leq&\frac{\|L^\dagger dL\|_1}{1-\lambda_4\|L^\dagger\|_1}\end{aligned} \quad (72)$$

Substituting (71) and (72) into (70), we obtain

$$\|\boldsymbol{t}_{ij}-\boldsymbol{t}_{ij}^{gt}\|\leq x\|\boldsymbol{t}_{ij}^{gt}\|+(1-x)\frac{\|E_{ij}^T L^\dagger\|_1\|dL\boldsymbol{u}_4^{gt}\|_\infty}{1-\lambda_4\|L^\dagger\|_1-\|L^\dagger dL\|_1} \quad (73)$$

Similarly, we have

$$\|\boldsymbol{t}-\boldsymbol{t}^{gt}\| \leq x\|\boldsymbol{t}^{gt}\| + (1-x)\frac{\|L^{\dagger}\|_1 \|dL\boldsymbol{u}_4^{gt}\|_{\infty}}{1-\lambda_4\|L^{\dagger}\|_1 - \|L^{\dagger}dL\|_1}. \quad (74)$$

We now end the proof because

$$\lambda_4 \leq \boldsymbol{u}_4^{gt\,T} dL \boldsymbol{u}_4^{gt}.$$

□

E.3.3 Proof Lemma 4

We only prove Eq. (52) as the proof of Eq. (53) is very similar.

Note that both $L^{gt}(\boldsymbol{w})$ and $L^{gt}(\overline{\boldsymbol{w}})$ share the same non-trivial eigenvectors denoted as $\overline{U}_4^{gt}$. This means,

$$L^{gt}(\boldsymbol{w})^{\dagger} = \overline{U}_4^{gt}\Big(\overline{U}_4^{gt\,T} L^{gt}(\boldsymbol{w})\overline{U}_4^{gt}\Big)^{-1}\overline{U}_4^{gt\,T},$$
$$L^{gt}(\overline{\boldsymbol{w}})^{\dagger} = \overline{U}_4^{gt}\Big(\overline{U}_4^{gt\,T} L^{gt}(\overline{\boldsymbol{w}})\overline{U}_4^{gt}\Big)^{-1}\overline{U}_4^{gt\,T}.$$

It follows that,

$$\begin{aligned}
&L^{gt}(\boldsymbol{w})^{\dagger}\\
=&\overline{U}_4^{gt}\Big(\overline{U}_4^{gt\,T} L^{gt}(\overline{\boldsymbol{w}})\overline{U}_4^{gt} + \overline{U}_4^{gt\,T} L^{gt}(d\boldsymbol{w})\overline{U}_4^{gt}\Big)^{-1}\overline{U}_4^{gt\,T}\\
=&\overline{U}_4^{gt}(\overline{U}_4^{gt\,T} L^{gt}(\overline{\boldsymbol{w}})\overline{U}_4^{gt})^{-1}\sum_{i=0}^{\infty}\Big(-(\overline{U}_4^{gt\,T} L^{gt}(\overline{\boldsymbol{w}})\overline{U}_4^{gt})^{-1}\\
&\overline{U}_4^{gt\,T} L^{gt}(\overline{d\boldsymbol{w}})\overline{U}_4^{gt}\Big)^{i}\overline{U}_4^{gt\,T}\\
=&L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\sum_{i=0}^{n}\Big(-L^{gt}(\overline{\boldsymbol{w}})^{\dagger}L^{gt}(d\boldsymbol{w})\Big)^{i}.
\end{aligned}$$

Applying triangle inequality, we arrive at

$$\begin{aligned}
\|L^{gt}(\boldsymbol{w})^{\dagger}\|_1 &\leq \|L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1\sum_{i=0}^{\infty}\big(\|L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1\|L^{gt}(d\boldsymbol{w})\|_1\big)^{i}\\
&= \frac{\|L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1}{1-\|L^{gt}(\overline{\boldsymbol{w}})^{\dagger}\|_1\|L^{gt}(d\boldsymbol{w})\|_1}.
\end{aligned}$$

□

E.3.4 Proof of Lemma 5

We first describe two propositions regarding the non-empty blocks of $dL(\boldsymbol{w})$.

Proposition 12. *$\forall (i,j)\in\mathcal{E}^{\text{in}}$, we have*

$$\|(\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^{T})\boldsymbol{t}_{ij}^{gt}\| \leq \epsilon\|\boldsymbol{t}_{ij}^{gt}\|, \quad (75)$$
$$\|\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^{T}\| \leq \epsilon. \quad (76)$$

Moreover, $\forall (i,j)\in\mathcal{E}^{\text{out}}$, we have

$$\|(\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^{T})\boldsymbol{t}_{ij}^{gt}\| \leq \|\boldsymbol{t}_{ij}^{gt}\|, \quad (77)$$
$$\|\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^{T}\| \leq 1, \quad (78)$$

Proof: Express $\boldsymbol{v}_{ij}$ as

$$\boldsymbol{v}_{ij} = \cos(\theta)\boldsymbol{v}_{ij}^{gt} + \sin(\theta)\boldsymbol{v}_{ij}^{gt\perp},$$
$$\boldsymbol{v}_{ij}^{gt\perp} = \frac{(I_3 - \boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T})\boldsymbol{v}_{ij}}{\|(I_3 - \boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T})\boldsymbol{v}_{ij}\|}.$$

Then

$$\begin{aligned}
&\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^{T}\\
=&(\boldsymbol{v}_{ij}^{gt},\boldsymbol{v}_{ij}^{gt\perp})\sin(\theta)\begin{pmatrix}\sin(\theta) & -\cos(\theta)\\ -\cos(\theta) & -\sin(\theta)\end{pmatrix}(\boldsymbol{v}_{ij}^{gt},\boldsymbol{v}_{ij}^{gt\perp})^{T}.
\end{aligned} \quad (79)$$

This means

$$\begin{aligned}
&\|\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^{T}\|\\
=&|\sin(\theta)|\|\begin{pmatrix}\sin(\theta) & -\cos(\theta)\\ -\cos(\theta) & -\sin(\theta)\end{pmatrix}\|\\
=&|\sin(\theta)|.
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\|(\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^{T})\boldsymbol{t}_{ij}^{gt}\|\\
=&|\sin(\theta)|\|\begin{pmatrix}\sin(\theta) & -\cos(\theta)\\ -\cos(\theta) & -\sin(\theta)\end{pmatrix}(\|\boldsymbol{t}_{ij}^{gt}\|,0)^{T}\|\\
=&|\sin(\theta)|\|\boldsymbol{t}_{ij}^{gt}\|.
\end{aligned}$$

We complete the proof by noting that When $(i,j)\in\mathcal{E}^{\text{in}}$, we have $|\sin(\theta)|\leq\epsilon$. □

We now complete the proof of Lemma 5. Applying Eq. (75) and Eq. (77), Eq. (54) is true because

$$\|dL(\boldsymbol{w})\boldsymbol{t}^{gt}\|_{\infty} \leq \|dL(\overline{\boldsymbol{w}})\boldsymbol{t}^{gt}\|_{\infty} + \|dL(d\boldsymbol{w})\boldsymbol{t}^{gt}\|_{\infty}$$

where

$$\begin{aligned}
&\|dL(d\boldsymbol{w})\boldsymbol{t}^{gt}\|_{\infty}\\
\leq&\max_{1\leq i\leq n}\Big(\sum_{j\in\mathcal{N}_i}|dw_{ij}|\|(\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt\,T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^{T})\boldsymbol{t}_{ij}^{gt}\|\Big)\\
\leq&\max_{1\leq i\leq n}\Big(\sum_{j\in\mathcal{N}_i^{\text{in}}}\epsilon|dw_{ij}|\|\boldsymbol{t}_{ij}^{gt}\| + \sum_{j\in\mathcal{N}_i^{\text{out}}}|dw_{ij}|\|\boldsymbol{t}_{ij}^{gt}\|\Big)\\
=&\max_{1\leq i\leq n}\Big(\epsilon\delta_i^{\text{in}}(d\boldsymbol{w}) + \delta_i^{\text{out}}(d\boldsymbol{w})\Big).
\end{aligned}$$

Eq. (55) can be proven in a similar fashion. □

E.3.5 Proof of Lemma 6

First of all, we have

$$\boldsymbol{t}^{gt^T} dL(\boldsymbol{w})\boldsymbol{t}^{gt} = \boldsymbol{t}^{gt^T} dL(\overline{\boldsymbol{w}})\boldsymbol{t}^{gt} + \boldsymbol{t}^{gt^T} dL(d\boldsymbol{w})\boldsymbol{t}^{gt}$$

where

$$\begin{aligned} & \boldsymbol{t}^{gt^T} dL(d\boldsymbol{w})\boldsymbol{t}^{gt} \\ = & \sum_{(i,j)\in\mathcal{E}} (w_{ij} - \overline{w}_{ij})\boldsymbol{t}_{ij}^{gt^T}(\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt^T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^T)\boldsymbol{t}_{ij}^{gt} \end{aligned}$$

Applying Eq. (79), we have

$$\begin{aligned} & \boldsymbol{t}_{ij}^{gt^T}(\boldsymbol{v}_{ij}^{gt}\boldsymbol{v}_{ij}^{gt^T} - \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^T)\boldsymbol{t}_{ij}^{gt} \\ = & \sin^2(\theta)\|\boldsymbol{t}_{ij}^{gt}\|^2 \end{aligned}$$

Therefore,

$$\begin{aligned} \boldsymbol{t}^{gt^T} dL(d\boldsymbol{w})\boldsymbol{t}^{gt} \leq \epsilon^2 & \sum_{(i,j)\in\mathcal{E}^{\text{in}}} |w_{ij} - \overline{w}_{ij}|\|\boldsymbol{t}_{ij}^{gt}\|^2 \\ + & \sum_{(i,j)\in\mathcal{E}^{\text{out}}} |w_{ij} - \overline{w}_{ij}|\|\boldsymbol{t}_{ij}^{gt}\|^2, \end{aligned}$$

which proves Eq. (56).

Moreover,

$$\begin{aligned} \|dL(\boldsymbol{w})\| & \leq \|dL(\overline{\boldsymbol{w}}\| + \|dL(d\boldsymbol{w})\| \\ & \leq \|dL(\overline{\boldsymbol{w}}\| + \|dL(d\boldsymbol{w})\|_1. \end{aligned}$$

The rest of the proof follows that of Lemma 5 in Section E.3.4. □

E.4. Proof of Lemma 7

It is clear that the minimum value of $\|\boldsymbol{t}_{ij}\|$ and the minimum value of $\|\boldsymbol{v}_{ij}^{\text{inp}} - \boldsymbol{v}_{ij}\|$ can be obtained in isolation. The minimum value of $\|\boldsymbol{t}_{ij}\|$ is $(1-\eta)\|\boldsymbol{t}_{ij}^{gt}\|$. When $\boldsymbol{v}_{ij}^{\text{inp}}$ is in the cone specified by $\|\boldsymbol{t}_{ij} - \boldsymbol{t}_{ij}^{gt}\| \leq \eta\|\boldsymbol{t}_{ij}^{gt}\|$, the minimum value is given by $\|\boldsymbol{v}_{ij}^{\text{inp}} - \boldsymbol{v}_{ij}\|$. Otherwise, $\|\boldsymbol{v}_{ij}^{\text{inp}} - \boldsymbol{v}_{ij}\|$ is given by the difference between the angle between $\boldsymbol{v}_{ij}^{\text{inp}}$ and $\boldsymbol{v}_{ij}^{gt}$ and half-angle of the cone. This ends the proof.

□

E.5. Proof of Lemma 8

The proof is very similar to that of Lemma 7. The only difference is that the maximum value of $\|\boldsymbol{v}_{ij}^{\text{inp}} - \boldsymbol{v}_{ij}\|$ is 2.

□