ShapeFit: Exact location recovery from corrupted pairwise directions

Paul Hand*, Choongbum Lee and Vladislav Voroninski

Department of Mathematics, Massachusetts Institute of Technology
*Department of Computational and Applied Mathematics, Rice University

June 4, 2015. Revised July 4, 2015

Abstract

Let $t_1, \ldots, t_n \in \mathbb{R}^d$ and consider the location recovery problem: given a subset of pairwise direction observations $\{(t_i - t_j)/\|t_i - t_j\|_2\}_{i<j\in[n]\times[n]}$, where a constant fraction of these observations are arbitrarily corrupted, find $\{t_i\}_{i=1}^n$ up to a global translation and scale. We propose a novel algorithm for the location recovery problem, which consists of a simple convex program over dn real variables. We prove that this program recovers a set of n i.i.d. Gaussian locations exactly and with high probability if the observations are given by an Erdős-Rényi graph, d is large enough, and provided that at most a constant fraction of observations involving any particular location are adversarially corrupted. We also prove that the program exactly recovers Gaussian locations for $d = 3$ if the fraction of corrupted observations at each location is, up to poly-logarithmic factors, at most a constant. Both of these recovery theorems are based on a set of deterministic conditions that we prove are sufficient for exact recovery.

1 Introduction

Let T be a collection of n distinct vectors $t_1^{(0)}, t_2^{(0)}, \ldots, t_n^{(0)} \in \mathbb{R}^d$, and let $G = ([n], E)$ be a graph, where $[n] = \{1, 2 \ldots, n\}$, and $E = E_g \sqcup E_b$, with E_b and E_g corresponding to pairwise direction observations that are respectively corrupted and uncorrupted. That is, for each $ij \in E$, we are given a vector v_{ij}, where

$$v_{ij} = \frac{t_i^{(0)} - t_j^{(0)}}{\|t_i^{(0)} - t_j^{(0)}\|_2} \text{ for } ij \in E_g, \qquad v_{ij} \in \mathbb{S}^{d-1} \text{ for } ij \in E_b. \tag{1}$$

Thus, an uncorrupted observation v_{ij} is exactly the direction of $t_i^{(0)}$ relative to $t_j^{(0)}$, and a corrupted observation is an arbitrary unit vector. Consider the task of recovering the locations T up to a global translation and scale, from only the observations $\{v_{ij}\}_{ij\in E}$, and without any knowledge about the decomposition $E = E_g \sqcup E_b$, nor the nature of the pairwise direction corruptions.

A special case of this problem, with $d = 3$, is a necessary subtask in the Structure from Motion (SfM) pipeline for 3D structure recovery from a collection of images taken from different vantage points, a vital aspect of modern computer vision. In the SfM problem, camera locations and orientations are represented as vectors and rotations in $\mathbb{R}^3$, with respect to some global reference frame. Given a collection of images, and for any point in $\mathbb{R}^3$, there is a unique perspective projection of it onto each imaging plane. By building local coordinate frames around salient points in the given images, based entirely on photometric information, and comparing them across images, one obtains an estimate of a set of point correspondences. That is, one obtains a set of equivalence classes, where each class corresponds to a physical point in 3D space. Given sufficiently many such sets of point correspondences, epipolar geometry and physical constraints yield estimates of the relative directions and orientations between pairs of cameras. Noise in these estimates is inherent to any real-world application, and worse yet, due to intrinsic challenges arising from the image formation process and properties of man-made scenes (illumination changes, specularities, occlusions, shadows, duplicate structures etc), severe outliers in estimated point correspondences and hence relative camera poses are unavoidable.

Once camera locations and orientations are estimated, 3D structure can then be recovered by a process called bundle adjustment [24], which is a simultaneous nonlinear refinement of 3D structure, camera locations, and camera orientations. Bundle-adjustment is a local method, which generally works well when started close to an optimum. Thus, it is critical to obtain accurate camera location and rotation estimates for initialization. SfM therefore consists of three steps: 1) estimating relative camera pose from point correspondences, 2) recovering camera locations and orientations in a global coordinate framework, and 3) bundle adjustment. While the first and third steps have well-founded theories and algorithms, methods for the second step are mostly heuristically motivated.

Several efficient and stable algorithms exist for estimating global camera orientations [9, 6, 2, 18, 7, 22, 11, 4, 8, 4, 10, 17, 20]. Hence, it is standard to recover locations separately based on estimates of the orientations.

There have been many different approaches to location recovery from relative directions, such as least squares [9, 2, 3, 17], second-order cone programs and l_∞ methods [13, 17, 18, 14, 21], spectral methods [3], similarity transformations for pair alignment [22], Lie-algebraic averaging [10], markov random fields [5], and several others [22, 25, 20, 12]. Unfortunately, most location recovery algorithms either lack robustness to correspondence errors (which are unavoidable in large unordered datasets), at times produce illegitimate collapsed solutions, or suffer from convergence to local minima, in sum causing large errors in or a complete degradation of, the recovered locations.

There are some recent notable exceptions to the above limitations. An algorithm called 1dSfM [28] focuses on removing outliers by examining inconsistencies along one-dimensional projections, before attempting to recover camera locations. However, one drawback of this method is that it does not reason about self-consistent outliers, which occur due to repetitive structures, commonly found in man-made scenes. Also, Özyeşil and Singer propose a convex program over $dn + |E|$ variables for location recovery and empirically demonstrate its robustness to outliers [19]. While both of these methods exhibit favorable empirical performance, they lack theoretical guarantees of robustness to outliers.

In this paper, we propose a novel convex program for location recovery from pairwise direction observations, and prove that this method recovers locations exactly, in the face of adversarial corruptions, and under rather broad technical assumptions. To the best of our knowledge, this is the first theoretical result guaranteeing location recovery in the challenging case of corrupted pairwise direction observations. We also demonstrate that this method performs well empirically, recovering locations exactly under severe corruptions of relative directions, and is stable to the simultaneous presence of noise on all the observations, as well as a fraction of arbitrary corruptions.

1.1 Problem formulation

The location recovery problem is to recover a set of points in $\mathbb{R}^d$ from observations of pairwise directions between those points. Since relative direction observations are invariant under a global translation and scaling, one can at best hope to recover the locations $T^{(0)} = \{t_1^{(0)}, \dots, t_n^{(0)}\}$ up to such a transformation. That is, successful recovery from $\{v_{ij}\}_{(i,j)\in E}$ is finding a set of vectors $\{\alpha(t_i^{(0)}+w)\}_{i\in[n]}$ for some $w \in \mathbb{R}^d$ and $\alpha > 0$. We will say that two sets of n vectors $T = \{t_1, \dots, t_n\}$ and $T^{(0)}$ are equal up to global translation and scale if there exists a vector w and a scalar $\alpha > 0$ such that $t_i = \alpha(t_i^{(0)} + w)$ for all $i \in [n]$. In this case, we will say that T and $T^{(0)}$ have the same 'shape,' and we will denote this property as $T \sim T^{(0)}$. The location recovery problem is then stated as:

$$\begin{aligned} &\text{Given:} \quad G([n], E), \quad \{v_{ij}\}_{ij\in E} \;\text{ satisfying (1)} \\ &\text{Find:} \quad T = \{t_1, \dots, t_n\} \in \mathbb{R}^{d\times n}, \quad \text{such that} \quad T \sim T^{(0)} \end{aligned} \tag{2}$$

For this problem to be information theoretically well-posed under arbitrary corruptions, the maximum number of corrupted observations affecting any particular location must be at most $\frac{n}{2}$. Otherwise, suppose that for some location $t_i^{(0)}$, half of its associated observations v_{ij} are consistent with $t_i^{(0)}$ and the other half are corrupted so as to be consistent with some arbitrary alternative location w. Distinguishing between $t_i^{(0)}$ and w is then impossible in general. Formally, let $\deg_b(i)$ be the degree of location i in the graph $([n], E_b)$. Then well-posedness under adversarial corruption requires that $\max_i \deg_b(i) \le \gamma n$ for some $\gamma < 1/2$.

Beyond the above necessary degree condition on E_g for well-posedness of recovery, we do not assume anything else about the nature of corruptions. That is, we work with adversarially chosen corrupted edges E_b and arbitrary corruptions of observations associated to those edges. To solve the location recovery problem in this challenging setting, we introduce a simple convex program called ShapeFit:

$$\min_{\{t_i\}\in\mathbb{R}^d, i\in[n]} \sum_{ij\in E} \|P_{v_{ij}^\perp}(t_i - t_j)\|_2 \quad \text{subject to} \quad \sum_{ij\in E} \langle t_i - t_j, v_{ij}\rangle = 1, \;\; \sum_{i=1}^n t_i = 0 \tag{3}$$

where $P_{v_{ij}^\perp}$ is the projector onto the orthogonal complement of the span of v_{ij}.

This convex program is a second order cone problem with dn variables and two constraints. Hence, the search space has dimension $dn - 2$, which is minimal due to the dn degrees of freedom in the locations $\{t_i\}$ and the two inherent degeneracies of translation and scale.

1.2 Main results

In this paper, we consider the model where pairwise direction observations about n i.i.d. Gaussian locations are given according to an Erdős-Rényi random graph. We start by showing that in a high-dimensional setting, ShapeFit *exactly* recovers the locations with high probability, provided that there are fewer than an exponential number of locations, and provided that at most a fixed fraction of observations are *adversarially* corrupted.

Theorem 1. *Let $G([n], E)$ be drawn from $G(n,p)$ for some $p = \Omega(n^{-1/4})$. Take $t_1^{(0)}, \dots t_n^{(0)} \sim \mathcal{N}(0, I_{d\times d})$ to be i.i.d., independent from G. There exists an absolute constant $c > 0$ and a $\gamma = \Omega(p^4)$ not depending on n, such that if $\max(\frac{2^6}{c^6}, \frac{4^3}{c^3}\log^3 n) \le n \le e^{\frac{1}{6}cd}$ and $d = \Omega(1)$, then there exists an event with probability at least $1 - e^{-n^{1/6}} - 13e^{-\frac{1}{2}cd}$, on which the following holds:*

For arbitrary subgraphs E_b satisfying $\max_i \deg_b(i) \le \gamma n$ and arbitrary pairwise direction corruptions $v_{ij} \in \mathbb{S}^{d-1}$ for $ij \in E_b$, the convex program (3) has a unique minimizer equal to $\left\{\alpha\Big(t_i^{(0)} - \bar{t}^{(0)}\Big)\right\}_{i\in[n]}$ for some positive α and for $\bar{t}^{(0)} = \frac{1}{n}\sum_{i\in[n]} t_i^{(0)}$.

This probabilistic recovery theorem is based on a set of deterministic conditions that we prove are sufficient to guarantee exact recovery. These conditions are satisfied with high probability in the model described above. See Section 2.1 for the deterministic conditions.

This recovery theorem is high-dimensional in the sense that the probability estimate and the exponential upper bound on n are only meaningful for $d = \Omega(1)$. Concentration of measure in high dimensions and the upper bound on n ensure control over the angles and distances between random points. As a result, lower dimensional spaces are a more challenging regime for recovery.

Our other main result is in the physically relevant setting of three-dimensional Euclidean space, where for instance the locations correspond to camera locations. In this setting, we prove that exact recovery holds for any sufficiently large number of locations, provided that a poly-logarithmically small fraction of observations are adversarially corrupted.

Theorem 2. *There exists $n_0 \in \mathbb{N}$ and $c \in \mathbb{R}$ such that the following holds for all $n \geq n_0$. Let $G([n], E)$ be drawn from $G(n, p)$ for some $p = \Omega(n^{-1/5} \log^{3/5} n)$. Take $t_1^{(0)}, \dots t_n^{(0)} \in \mathbb{R}^3$, where $t_i^{(0)} \sim \mathcal{N}(0, I_{3\times 3})$ are i.i.d., independent from G. There exists $\gamma = \Omega(p^5/\log^3 n)$ and an event of probability at least $1 - \frac{1}{n^4}$ on which the following holds:*

For arbitrary subgraphs E_b satisfying $\max_i \deg_b(i) \leq \gamma n$ and arbitrary pairwise direction corruptions $v_{ij} \in \mathbb{S}^2$ for $ij \in E_b$, the convex program (3) has a unique minimizer equal to $\left\{\alpha\left(t_i^{(0)} - \bar{t}^{(0)}\right)\right\}_{i\in[n]}$ for some positive α and for $\bar{t}^{(0)} = \frac{1}{n}\sum_{i\in[n]} t_i^{(0)}$.

Numerical simulations empirically verify the main message of these recovery theorems: ShapeFit recovers a set of locations exactly from corrupted direction observations, provided that up to a constant fraction of the observations at each location are corrupted. We present numerical studies in the setting of locations in $\mathbb{R}^3$, with an underlying random Erdős-Rényi graph of observations. Further numerical simulations show that recovery is stable to the additional presence of noise on the uncorrupted measurements. That is, locations are recovered approximately under such conditions, with a favorable dependence of the estimation error on the measurement noise.

1.3 Intuition.

ShapeFit is a convex program that seeks a set of points whose pairwise directions agree with as many of the corresponding observations as possible. The objective, $\sum_{ij\in E} \|P_{v_{ij}^\perp}(t_i - t_j)\|_2$, incentivizes the correct shape, while permitting translation and a possibly-negative global scale. Each term $\|P_{v_{ij}^\perp}(t_i - t_j)\|_2$ is a length-scaled notion for how rotated $t_i - t_j$ is relative to $\pm v_{ij}$. The objective is in this sense a measure of how much total rotation is needed to deform all $\{t_i - t_j\}_{ij\in E}$ into the observed directions of $\{\pm v_{ij}\}$. Successful recovery would mean that $\{\|P_{v_{ij}^\perp}(t_i - t_j)\|_2\}_{ij\in E}$ is sparse. Motivated by the sparsity promoting properties of ℓ_1-minimization, the objective in ShapeFit is precisely the ℓ_1 norm over the edges $E(G)$ of these ℓ_2 lengths.

The first constraint in ShapeFit, $\sum_{ij\in E}\langle t_i - t_j, v_{ij}\rangle = 1$, requires that the recovered locations correlate with the provided observations by a strictly positive amount. It prevents the trivial solution and resolves the global scale ambiguity. As opposed to the objective, this constraint forbids negative scalings of $\{t_i^{(0)}\}_{i\in[n]}$. The second constraint, $\sum_{i=1}^n t_i = 0$, resolves the global translation ambiguity.

1.4 Organization of the paper

Section 1.5 presents the notation used throughout the rest of the paper. Section 2 presents the proof of Theorem 1. Section 3 presents the proof of Theorem 2. Section 4 presents results from numerical simulations.

1.5 Notation

Let $[n] = \{1, \dots, n\}$. Let e_i be the ith standard basis element. Let K_n be the complete graph on n vertices. Let $E(K_n)$ be the set of edges in K_n. Let $\|\cdot\|_2$ be the standard ℓ_2 norm on a vector. For any nonzero vector v, let $\hat{v} = v/\|v\|_2$. For a subspace W, let P_W be the orthogonal projector onto W. For a vector v, let $P_{v^\perp}$ be the orthogonal projector onto the orthogonal complement of the span of $\{v\}$.

Let T denote the set $T = \{t_i\}_{i\in[n]}$, for $t_i \in \mathbb{R}^d$. Define $t_{ij} = t_i - t_j$ for all distinct $i, j \in [n]$. We define $\mu_\infty = \max_{i\neq j} \|t_{ij}^{(0)}\|_2$, and we define $\mu = \frac{1}{|E(G)|}\sum_{ij\in E(G)} \|t_{ij}^{(0)}\|_2$. Define $\bar{t} = \frac{1}{n}\sum_{i\in[n]} t_i$. Define $t_{ij}^{(0)}$, $T^{(0)}$, and $\bar{t}^{(0)}$ similarly. For a scalar c, let $cT = \{ct_i\}_{i\in[n]}$. For a given $G = G([n], E)$ and $\{v_{ij}\}_{ij\in E}$, where $v_{ij} \in \mathbb{R}^d$ have unit norm, let $R(T) = \sum_{ij\in E} \|P_{v_{ij}^\perp}(t_i - t_j)\|_2$. Let $L(T) = \sum_{ij\in E}\langle t_i - t_j, v_{ij}\rangle$. Let $\ell_{ij} = \langle t_i - t_j, v_{ij}\rangle$, and similarly for $\ell_{ij}^{(0)}$. In this notation, ShapeFit is

$$\min_T R(T) \quad \text{subject to} \quad L(T) = 1, \quad \bar{t} = 0$$

For vectors $v_1, \dots, v_k$, let $S(v_1, \dots, v_k) = \operatorname{span}(v_1, \dots, v_k)$ be the vector space spanned by these vectors. Given t_{ij} and $t_{ij}^{(0)}$, define δ_{ij}, η_{ij}, and s_{ij} such that

$$t_{ij} = (1 + \delta_{ij})t_{ij}^{(0)} + \eta_{ij}s_{ij}$$

where s_{ij} is a unit vector orthogonal to $t_{ij}^{(0)}$ and $\eta_{ij} = \|P_{t_{ij}^{(0)\perp}} t_{ij}\|_2$. Note that $\eta_{ij} \geq 0$.

2 Proof of high dimensional recovery

The proof of Theorem 1 can be separated into two parts: a recovery guarantee under a set of deterministic conditions, and a proof that the random model meets these conditions with high probability. These sufficient deterministic conditions, roughly speaking, are (1) that the graph is connected and the nodes have tightly controlled degrees; (2) that the angles between pairs of locations is uniformly bounded away from 0 and π; (3) that all pairwise distances are within a constant factor of each other; (4) that there are not too many corruptions affecting any single location; and (5) that the locations are 'well-distributed' relative to each other in a sense we will make precise. Theorem 3 in Section 2.1 states these deterministic conditions formally.

We will prove the deterministic recovery theorem directly, using several geometric properties concerning how deformations of a set of points induce rotations. Note that an infinitesimal rigid rotation of two points $\{t_i, t_j\}$ about their midpoint to $\{t_i + h_i, t_j + h_j\}$ is such that $h_i - h_j$ is orthogonal to $t_{ij} = t_i - t_j$. We will abuse terminology and say that $\|P_{t_{ij}^\perp}(h_i - h_j)\|$ is a measure of the rotation in a finite deformation $\{h_i, h_j\}$, and we say that $\langle h_i - h_j, t_i - t_j\rangle$ is the amount of stretching in that deformation. Using this terminology, the geometric properties we establish are:

- If a deformation stretches two adjacent sides of a triangle at different rates, then that induces a rotation in some edge of the triangle (Lemma 2).

- If a deformation stretches two nonadjacent sides of a tetrahedron at different rates, then that induces a rotation in some edge of the tetrahedron (Lemma 3).
- If a deformation rotates one edge shared by many triangles, then it induces a rotation over many of those triangles, provided the opposite points of those triangles are 'well-distributed' (Lemma 4).
- A deformation that rotates bad edges, must also rotate good edges (Lemma 5).
- For any deformation, some fraction of the sum of all rotations must affect the good edges (Lemma 6).

By using these geometric properties, we show that all nonzero feasible deformations induce a large amount of total rotation. Since some fraction of the total rotation must be on the good edges, the objective must increase.

In Section 2.1, we present the deterministic recovery theorem. In Section 2.2, we present and prove Lemmas 2–3. In Section 2.3, we present and prove Lemmas 4–6. In Section 2.4, we prove the deterministic recovery theorem. In Section 2.5, we prove that Gaussians satisfy several properties, including well-distributedness, with high probability. In Section 2.6, we prove that Erdős-Rényi graphs are connected and have controlled degrees and codegrees with high probability. Finally, in Section 2.7, we prove Theorem 1.

2.1 Deterministic recovery theorem in high dimensions

To state the deterministic recovery theorem, we need two definitions.

Definition 1. *We say that a graph $G([n], E)$ is* p-typical *if it satisfies the following properties:*

1. *G is connected,*
2. *each vertex has degree between $\frac{1}{2}np$ and $2np$, and*
3. *each pair of vertices has codegree between $\frac{1}{2}np^2$ and $2np^2$, where the codegree of a pair of vertices i, j is defined as $|\{k \in [n] : ik, jk \in E(G)\}|$.*

Note that if G is p-typical, then its number of edges is between $\frac{1}{4}n^2p$ and n^2p.

Definition 2. *Let $T = \{t_i\}_{i\in[n]} \subseteq \mathbb{R}^d$ be a set of n vectors. Let G be a graph with vertex set $[n]$.*

(i) For a pair of vectors $x, y \in \mathbb{R}^d$ and a positive real number c, we say that T is c-well-distributed with respect to (x, y) *if the following holds for all $h \in \mathbb{R}^d$:*

$$\sum_{t\in T} \|P_{\mathrm{span}\{t-x,t-y\}^\perp}(h)\|_2 \geq c|T| \cdot \|P_{(x-y)^\perp}(h)\|_2.$$

(ii) We say that T is c-well-distributed along G *if for all distinct $i, j \in [n]$, the set $S_{ij} = \{t_k : ik, jk \in E(G)\}$ is c-well-distributed with respect to (t_i, t_j).*

We now state sufficient deterministic recovery conditions on the graph G, the subgraph E_b corresponding to corrupted observations, and the locations $T^{(0)}$.

Theorem 3. *Suppose $T^{(0)}, E_b, G$ satisfy the conditions*

1. *The underlying graph G is p-typical,*
2. *For all distinct $i, j, k \in [n]$, we have $\sqrt{1 - \langle \hat{t}^{(0)}_{ij}, \hat{t}^{(0)}_{ik}\rangle^2} \geq \beta$,*

3. *For all i, j, k, ℓ with $i \neq j$ and $k \neq l$, we have $c_0 \|t_{k\ell}^{(0)}\|_2 \leq \|t_{ij}^{(0)}\|_2$,*
4. *Each vertex has at most εn edges in E_b incident to it,*
5. *The set $\{t_i^{(0)}\}_{i \in [n]}$ is c_1-well-distributed along G,*
6. *All vectors $t_i^{(0)}$ are distinct,*

for constants $0 < p, \beta, c_0, \varepsilon, c_1 \leq 1$. If $\varepsilon \leq \frac{\beta c_0 c_1^2 p^4}{3 \cdot 256 \cdot 64 \cdot 32}$, then $L(T^{(0)}) \neq 0$ and $T^{(0)}/L(T^{(0)})$ is the unique optimizer of ShapeFit.

Note that Condition 3 implies that for $\mu_\infty = \max_{i \neq j} \|t_{ij}^{(0)}\|_2$, we have $c_0 \mu_\infty \leq \|t_{ij}^{(0)}\|_2 \leq \mu_\infty$ for all distinct $i, j \in [n]$. Also note that Conditions 1–6 are invariant under translation and non-zero scalings of $T^{(0)}$.

Before we prove the theorem, we establish that $L(T^{(0)}) \neq 0$ when ε is small enough. This property guarantees that some scaling of $T^{(0)}$ is feasible and occurs, roughly speaking, when $|E_b| < |E_g|$.

Lemma 1. *If $\varepsilon < \frac{c_0 p}{8}$, then $L(T^{(0)}) \neq 0$.*

Proof. Since $v_{ij} = t_{ij}^{(0)}$ for all $ij \in E_g$, we have

$$L(T) = \sum_{ij \in E(G)} \langle t_{ij}^{(0)}, v_{ij} \rangle \geq \sum_{ij \in E_g} \|t_{ij}^{(0)}\|_2 - \sum_{ij \in E_b} \|t_{ij}^{(0)}\|_2.$$

By Condition 3, $c_0 \mu_\infty |E_g| \leq \sum_{ij \in E_g} \|t_{ij}^{(0)}\|_2$ and $\mu_\infty |E_b| \geq \sum_{ij \in E_b} \|t_{ij}^{(0)}\|_2$. Thus it suffices to prove that $c_0 |E_g| > |E_b|$. As $\varepsilon < \frac{p}{8}$, Condition 1 and 4 gives $|E_g| \geq \frac{1}{4} n^2 p - \varepsilon n^2 \geq \frac{1}{8} n^2 p$. Since $|E_b| \leq \varepsilon n^2$, if $\varepsilon < \frac{c_0 p}{8}$, then we have $c_0 |E_g| > |E_b|$. □

The proof of Theorem 3 appears in Section 2.4.

2.2 Unbalanced parallel motions induce rotation

Lemma 2. *Let $d \geq 2$. Let $t_1, t_2, t_3 \in \mathbb{R}^d$ be distinct. Let $v_1, v_2, v_3 \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}$. Let $\{\tilde{\delta}_{ij}\}$ be such that $\langle v_i - v_j - \alpha t_{ij}, \hat{t}_{ij} \rangle = \tilde{\delta}_{ij} \|t_{ij}\|_2$ for each distinct $i, j \in [3]$. Then*

$$\sum_{\substack{i,j \in [3] \\ i<j}} \|P_{t_{ij}^\perp}(v_i - v_j)\|_2 \geq \sqrt{1 - \langle \hat{t}_{12}, \hat{t}_{23} \rangle^2} \|t_{12}\|_2 \left| \tilde{\delta}_{12} - \tilde{\delta}_{13} \right|.$$

Proof. Note that $t_{ij} = -t_{ji}$ and $\tilde{\delta}_{ij} = \tilde{\delta}_{ji}$ for each distinct $i, j \in [3]$. Define $W = \operatorname{span}(\hat{t}_{12}, \hat{t}_{23}, \hat{t}_{31})$ and define $w_i = P_W v_i$ for each i. Note that

$$\sum_{i<j} \|P_{t_{ij}^\perp}(v_i - v_j)\|_2 \geq \sum_{i<j} \|P_{t_{ij}^\perp}(w_i - w_j)\|_2.$$

The given condition implies $P_{t_{ij}^\perp}(w_i - w_j) = w_i - w_j - (\alpha + \tilde{\delta}_{ij}) t_{ij}$ for each distinct $i, j \in [3]$. Therefore,

$$\begin{aligned}
\sum_{i<j} \|P_{t_{ij}^\perp}(w_i - w_j)\|_2 &= \sum_{i<j} \left\| w_i - w_j - \left(\alpha + \tilde{\delta}_{ij}\right) t_{ij} \right\|_2 \\
&\geq \left\| \sum_{(i,j)=(1,2),(2,3),(3,1)} w_i - w_j - \left(\alpha + \tilde{\delta}_{ij}\right) t_{ij} \right\|_2 \\
&= \|\tilde{\delta}_{12} t_{12} + \tilde{\delta}_{23} t_{23} + \tilde{\delta}_{31} t_{31}\|_2.
\end{aligned}$$

Since $\tilde{\delta}_{13}(t_{12}+t_{23}+t_{31})=0$, the right-hand-side above equals $\|(\tilde{\delta}_{12}-\tilde{\delta}_{13})t_{12}+(\tilde{\delta}_{23}-\tilde{\delta}_{13})t_{23}\|_2$. Furthermore,

$$\begin{aligned}
\left\|(\tilde{\delta}_{12}-\tilde{\delta}_{13})t_{12}+(\tilde{\delta}_{23}-\tilde{\delta}_{13})t_{23}\right\|_2 &\geq \min_{s\in\mathbb{R}}\|(\tilde{\delta}_{12}-\tilde{\delta}_{13})t_{12}-st_{23}\|_2 \\
&= \left\|P_{t_{23}^{\perp}}(\tilde{\delta}_{12}-\tilde{\delta}_{13})t_{12}\right\|_2 \\
&\geq \left|\tilde{\delta}_{12}-\tilde{\delta}_{13}\right|\|t_{12}\|_2\sqrt{1-\langle\hat{t}_{12},\hat{t}_{23}\rangle^2}. \quad \square
\end{aligned}$$

The previous lemma is applicable only when two disproportionally scaled edges are incident to each other. The following lemma shows how to apply the lemma above to the case when we have two vertex-disjoint edges that are disproportionally scaled.

Lemma 3. *Let $d\geq 2$. Let $t_1,t_2,t_3,t_4\in\mathbb{R}^d$ be distinct. Let $v_1,v_2,v_3,v_4\in\mathbb{R}^d$ and $\alpha\in\mathbb{R}$. Let $\{\tilde{\delta}_{ij}\}$ be such that $\langle v_i-v_j-\alpha t_{ij},\hat{t}_{ij}\rangle=\tilde{\delta}_{ij}\|t_{ij}\|_2$ for each distinct $i,j\in[4]$. Define $\beta=\min\sqrt{1-\langle\hat{t}_{ij},\hat{t}_{ik}\rangle^2}$ where the minimum is taken over all distinct $i,j,k\in[4]$ except for the cases when $\{j,k\}=\{1,2\}$. Then*

$$\sum_{\substack{i,j\in[4]\\ i<j}}\|P_{t_{ij}^{\perp}}(v_i-v_j)\|_2\geq\frac{\beta}{4}\|t_{12}\|_2\left|\tilde{\delta}_{12}-\tilde{\delta}_{34}\right|.$$

Proof. Note that $t_{ij}=-t_{ji}$ and $\tilde{\delta}_{ij}=\tilde{\delta}_{ji}$ for each distinct $i,j\in[4]$. Since the given conditions are symmetric under re-labelling of (1 and 2), and of (3 and 4), we may re-label if necessary and assume that $\|t_{13}\|_2\geq\max\{\|t_{14}\|_2,\|t_{23}\|_2,\|t_{24}\|_2\}$. By the triangle inequality, we have $2\|t_{13}\|_2\geq\|t_{13}\|_2+\|t_{23}\|_2\geq\|t_{12}\|_2$. Apply Lemma 2 to the triangle $\{1,2,3\}$ to obtain

$$\begin{aligned}
\sum_{i<j,\ i,j\in\{1,2,3\}}\|P_{t_{ij}^{\perp}}(v_i-v_j)\|_2 &\geq \sqrt{1-\langle\hat{t}_{12},\hat{t}_{23}\rangle^2}\|t_{12}\|_2\left|\tilde{\delta}_{12}-\tilde{\delta}_{13}\right| \\
&\geq \beta\|t_{12}\|_2\left|\tilde{\delta}_{12}-\tilde{\delta}_{13}\right|, \qquad (4)
\end{aligned}$$

and similarly apply the lemma to the triangle $\{3,1,4\}$ to obtain

$$\begin{aligned}
\sum_{i<j,\ i,j\in\{1,3,4\}}\|P_{t_{ij}^{\perp}}(v_i-v_j)\|_2 &\geq \sqrt{1-\langle\hat{t}_{13},\hat{t}_{14}\rangle^2}\|t_{13}\|_2\left|\tilde{\delta}_{13}-\tilde{\delta}_{34}\right| \\
&\geq \beta\|t_{13}\|_2|\tilde{\delta}_{13}-\tilde{\delta}_{34}|\geq\frac{\beta}{2}\|t_{12}\|_2|\tilde{\delta}_{13}-\tilde{\delta}_{34}|. \qquad (5)
\end{aligned}$$

By adding (4) and (5), we see that

$$\begin{aligned}
&\sum_{i<j,\ i,j\in\{1,2,3\}}\|P_{t_{ij}^{\perp}}(v_i-v_j)\|_2+\sum_{i<j,\ i,j\in\{1,3,4\}}\|P_{t_{ij}^{\perp}}(v_i-v_j)\|_2 \\
&\geq \beta\|t_{12}\|_2\left|\tilde{\delta}_{12}-\tilde{\delta}_{13}\right|+\frac{\beta}{2}\|t_{12}\|_2\left|\tilde{\delta}_{13}-\tilde{\delta}_{34}\right|\geq\frac{\beta}{2}\|t_{12}\|_2\left|\tilde{\delta}_{12}-\tilde{\delta}_{34}\right|.
\end{aligned}$$

The lemma follows since the left-hand-side is bounded from above by $2\sum_{\substack{i,j\in[4]\\ i<j}}\|P_{t_{ij}^{\perp}}(v_i-v_j)\|_2$. $\square$

2.3 Triangles inequality and rotation propagation

Lemma 4 (Triangles Inequality). *Let $d \geq 3$; $x, y, t_1, t_2, \cdots, t_k \in \mathbb{R}^d$. If $T = \{t_1, \cdots, t_k\}$ is c-well-distributed with respect to (x, y), then for all vectors $h_x, h_y, h_1, \cdots, h_k \in \mathbb{R}^d$ and sets $X \subseteq [k]$, we have*

$$\sum_{i \in [k] \setminus X} \|P_{(x-t_i)^\perp}(h_x - h_i)\|_2 + \|P_{(t_i - y)^\perp}(h_i - h_y)\|_2 \geq (ck - |X|) \cdot \|P_{(x-y)^\perp}(h_x - h_y)\|_2.$$

Proof. For each $i \in [k]$, define $W_i = \mathrm{span}\langle x - t_i\,,\, t_i - y\rangle$. Define P as the projection map to the space of vectors orthogonal to $x - y$, and define P_i for each $i \in [k]$ as the projection map to $W_i^\perp$. Since $(x - t_i)^\perp \supseteq W_i^\perp$ and $(t_i - y)^\perp \supseteq W_i^\perp$, it follows that

$$\begin{aligned} &\sum_{i \in [k] \setminus X} \|P_{(x-t_i)^\perp}(h_x - h_i)\|_2 + \|P_{(t_i - y)^\perp}(h_i - h_y)\|_2 \\ \geq &\sum_{i \in [k] \setminus X} \|P_i(h_x - h_i)\|_2 + \|P_i(h_i - h_y)\|_2 \geq \sum_{i \in [k] \setminus X} \|P_i(h_x - h_y)\|_2. \end{aligned}$$

Since $t_1, \cdots, t_k$ are well-distributed with respect to (x, y), we have

$$\sum_{i \in [k]} \|P_i(h_x - h_y)\|_2 \geq ck \cdot \|P(h_x - h_y)\|_2. \tag{6}$$

Since $\|P_i(h_x - h_y)\|_2 \leq \|P(h_x - h_y)\|_2$ holds for all i, it follows that

$$\sum_{i \in [k] \setminus X} \|P_i(h_x - h_y)\|_2 \geq (ck - |X|) \cdot \|P(h_x - h_y)\|_2,$$

proving the lemma. □

The proof of Theorem 3 will rely on the following two lemmas, which state that rotational motions on some parts of the graph bound rotational motions on other parts. The following lemma relates the rotational motions on bad edges to the rotational motions on good edges. Recall the notation $t_{ij} = (1+\delta_{ij})t_{ij}^{(0)} + \eta_{ij}s_{ij}$ where s_{ij} is a unit vector orthogonal to $t_{ij}^{(0)}$ and $\eta_{ij} = \|P_{t_{ij}^{(0)\perp}} t_{ij}\|_2$.

Lemma 5. *Fix T. If $\varepsilon_0 \leq \frac{c_1 p^2}{8}$, then $\sum_{ij \in E_g} \eta_{ij} \geq \frac{c_1 p^2}{8\varepsilon_0} \sum_{ij \in E_b} \eta_{ij}$.*

Proof. For each edge $ij \in E(K_n)$, by Conditions 1, 4, 5; Lemma 4 and $\varepsilon_0 \leq \frac{c_1 p^2}{8}$, we have

$$\sum_{\substack{k \neq i,j \\ ik, jk \in E_g}} (\eta_{ik} + \eta_{jk}) \geq \left(c_1 \cdot \frac{1}{2} np^2 - 2\varepsilon_0 n\right) \cdot \eta_{ij} \geq \frac{c_1}{4} np^2 \cdot \eta_{ij}.$$

Therefore, if we sum the inequality above for all bad edges $ij \in E_b$, then

$$\sum_{ij \in E_b} \sum_{\substack{k \neq i,j \\ ik, jk \in E_g}} (\eta_{ik} + \eta_{jk}) \geq \frac{c_1}{4} np^2 \cdot \sum_{ij \in E_b} \eta_{ij}.$$

For fixed $ik \in E_g$, the left-hand-side may sum η_{ik} as many times as the number of bad edges incident to the edge ik. Hence by Condition 4, the left-hand-side of above is at most

$$\sum_{ij \in E_b} \sum_{\substack{k \neq i,j \\ ik, jk \in E_g}} (\eta_{ik} + \eta_{jk}) \leq 2\varepsilon_0 n \cdot \sum_{ij \in E_g} \eta_{ij}.$$

Therefore by combining the two inequalities above, we obtain

$$\sum_{ij \in E_b} \eta_{ij} \leq \frac{8\varepsilon_0}{c_1 p^2} \sum_{ij \in E_g} \eta_{ij}. \qquad \square$$

The following lemma relates the rotational motions over the good graph E_g to rotational motions over the complete graph K_n.

Lemma 6. *Fix T. If $\varepsilon_0 \leq \frac{c_1 p^2}{8}$, then $\sum_{ij \in E_g} \eta_{ij} \geq \frac{c_1 p}{16} \sum_{ij \in E(K_n)} \eta_{ij}$.*

Proof. For each $ij \in E(K_n)$, since $\{t_i^{(0)}\}_{i=1}^n$ is c_1-well-distributed along G and G is p-typical, we have as in the proof of Lemma 5,

$$\sum_{\substack{k \neq i,j \\ ik,jk \in E_g}} (\eta_{ik} + \eta_{jk}) \geq \left(c_1 \cdot \frac{1}{2} np^2 - 2\varepsilon_0 n \right) \cdot \eta_{ij} \geq \frac{c_1}{4} np^2 \cdot \eta_{ij}.$$

If we sum the above over all $ij \in E(K_n)$, we obtain

$$\sum_{ij \in E(K_n)} \sum_{\substack{k \neq i,j \\ ik,jk \in E_g}} (\eta_{ik} + \eta_{jk}) \geq \frac{c_1}{4} np^2 \cdot \sum_{ij \in E(K_n)} \eta_{ij}.$$

For a fixed $ik \in E_g$, the left-hand-side may sum η_{ik} as many as times as the number of edges of G incident to ik. Therefore since G is p-typical, we see that

$$\sum_{\substack{k \neq i,j \\ ik,jk \in E_g}} (\eta_{ik} + \eta_{jk}) \leq 2 \cdot 2np \sum_{ij \in E_g} \eta_{ij}.$$

By combining the two inequalities, we obtain

$$\frac{c_1}{4} np^2 \sum_{ij \in E(K_n)} \eta_{ij} \leq 4np \sum_{ij \in E_g} \eta_{ij},$$

and thus $\frac{c_1 p}{16} \sum_{ij \in E(K_n)} \eta_{ij} \leq \sum_{ij \in E_g} \eta_{ij}$. $\square$

2.4 Proof of Theorem 3

We now prove the deterministic recovery theorem.

Proof of Theorem 3. By Lemma 1 and the fact that Conditions 1–6 are invariant under global translation and nonzero scaling, we can take $\overline{t}^{(0)} = 0$ and $L(T^{(0)}) = 1$ without loss of generality. The variable μ_∞ is to be understood accordingly.

We will directly prove that $R(T) > R(T^{(0)})$ for all $T \neq T^{(0)}$ such that $L(T) = 1$ and $\bar{t} = 0$. Consider an arbitrary feasible T and recall the notation $t_{ij} = (1 + \delta_{ij}) t_{ij}^{(0)} + \eta_{ij} s_{ij}$ where s_{ij} is a unit vector orthogonal to $t_{ij}^{(0)}$ and $\eta_{ij} = \|P_{t_{ij}^{(0)\perp}} t_{ij}\|_2$. A useful lower bound for the objective $R(T)$ is given by

$$\begin{aligned} R(T) = \sum_{ij} \|P_{v_{ij}^\perp} t_{ij}\|_2 &= \sum_{ij \in E_g} \eta_{ij} + \sum_{ij \in E_b} \|P_{v_{ij}^\perp} t_{ij}\|_2 \\ &\geq \sum_{ij \in E_g} \eta_{ij} + \sum_{ij \in E_b} \left(\|P_{v_{ij}^\perp} t_{ij}^{(0)}\|_2 - |\delta_{ij}| \|t_{ij}^{(0)}\|_2 - \eta_{ij} \right) \\ &\geq R(T^{(0)}) + \sum_{ij \in E_g} \eta_{ij} - \sum_{ij \in E_b} (|\delta_{ij}| \|t_{ij}^{(0)}\|_2 + \eta_{ij}). \end{aligned} \tag{7}$$

Suppose that $\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2 < \sum_{ij\in E_b}\eta_{ij}$. Since Lemma 5 for $\varepsilon \le \frac{c_1p^2}{16}$ implies $\sum_{ij\in E_b}\eta_{ij} \le \frac{1}{2}\sum_{ij\in E_g}\eta_{ij}$, by (7), we have

$$\begin{aligned} R(T) &\ge R(T^{(0)}) + \sum_{ij\in E_g}\eta_{ij} - \sum_{ij\in E_b}(|\delta_{ij}|\|t_{ij}^{(0)}\|_2 + \eta_{ij}) \\ &> R(T^{(0)}) + \sum_{ij\in E_g}\eta_{ij} - \sum_{ij\in E_b}2\eta_{ij} \ge R(T^{(0)}). \end{aligned}$$

Hence we may assume

$$\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2 \ge \sum_{ij\in E_b}\eta_{ij}. \tag{8}$$

In the case $|E_b| \ne 0$, define $\overline{\delta} = \frac{1}{|E_b|}\sum_{ij\in E_b}|\delta_{ij}|$ as the average 'relative parallel motion' on the bad edges. For distinct edges $ij, k\ell \in E(K_n)$, if $\{i,j\}\cap\{k,\ell\} = \emptyset$, then define $\eta(ij,k\ell) = \eta_{ij}+\eta_{ik}+\eta_{i\ell}+\eta_{jk}+\eta_{j\ell}+\eta_{k\ell}$, and if $\{i,j\}\cap\{k,\ell\}\ne\emptyset$ (without loss of generality, assume $\ell = i$), then define $\eta(ij,k\ell) = \eta_{ij}+\eta_{ik}+\eta_{jk}$.

Case 0. $\overline{\delta} = 0$ or $|E_b| = 0$.

Note that $\overline{\delta} = 0$ implies $\delta_{ij} = 0$ for all $ij \in E_b$, which by (8) implies $\eta_{ij} = 0$ for all $ij \in E_b$. Therefore by (7), we have

$$R(T) \ge R(T^{(0)}) + \sum_{ij\in E_g}\eta_{ij}.$$

If $\sum_{ij\in E_g}\eta_{ij} > 0$, then we have $R(T) > R(T^{(0)})$. Thus we may assume that $\eta_{ij} = 0$ for all $ij \in E_g$. In this case, we will show that $T = T^{(0)}$.

By Lemma 6, if $\varepsilon_0 \le \frac{c_1p^2}{8}$, then $\eta_{ij} = 0$ for all $ij \in E(G)$ implies that $\eta_{ij} = 0$ for all $ij \in E(K_n)$. For $ij \in E_b$, since $\delta_{ij} = \eta_{ij} = 0$, it follows that $\ell_{ij} = \ell_{ij}^{(0)}$. Since $\delta_{ij}\|t_{ij}^{(0)}\|_2 = \ell_{ij} - \ell_{ij}^{(0)}$ for $ij \in E_g$, we have

$$0 = \sum_{ij\in E(G)}(\ell_{ij}-\ell_{ij}^{(0)}) = \sum_{ij\in E_b}(\ell_{ij}-\ell_{ij}^{(0)}) + \sum_{ij\in E_g}(\ell_{ij}-\ell_{ij}^{(0)}) = \sum_{ij\in E_g}(\ell_{ij}-\ell_{ij}^{(0)}) = \sum_{ij\in E_g}\delta_{ij}\|t_{ij}^{(0)}\|_2,$$

where the first equality is because $L(T) = L(T^{(0)}) = 1$. By Condition 6, $\|t_{ij}^{(0)}\|_2 \ne 0$ for all $i \ne j$. Therefore, if $\delta_{ij} \ne 0$ for some $ij \in E_g$, then there exists $ab, cd \in E_g$ such that $\delta_{ab} > 0$ and $\delta_{cd} < 0$. By Lemma 2 or 3 and Condition 2, this forces $\eta(ab,cd) > 0$, contradicting the fact that $\eta_{ij} = 0$ for all $ij \in E(K_n)$. Therefore $\delta_{ij} = 0$ for all $ij \in E_g$, and hence $\delta_{ij} = 0$ for all $ij \in E(G)$.

Define $t_i = t_i^{(0)} + h_i$ for each $i \in [n]$. Because $\eta_{ij} = \delta_{ij} = 0$ for all $ij \in E(G)$, we have $h_i = h_j$ for all $ij \in E(G)$. Since G is connected, this implies $h_i = h_j$ for all $i \in [n]$. Then by the constraint $\sum_{i\in[n]}t_i = \sum_{i\in[n]}t_i^{(0)} = 0$, we get $h_i = 0$ for all $i \in [n]$. Therefore $T = T^{(0)}$.

Case 1. $\overline{\delta} \ne 0$ and $\sum_{ij\in E_g}|\delta_{ij}| < \frac{1}{8}\overline{\delta}|E_g|$ and $|E_b| \ne 0$.

Define $L_b = \{ij \in E_b : |\delta_{ij}| \ge \frac{1}{2}\overline{\delta}\}$. Note that $\sum_{ij\in E_b\setminus L_b}|\delta_{ij}| < \frac{1}{2}\overline{\delta}|E_b|$ and therefore

$$\sum_{ij\in L_b}|\delta_{ij}| = \sum_{ij\in E_b}|\delta_{ij}| - \sum_{ij\in E_b\setminus L_b}|\delta_{ij}| > \sum_{ij\in E_b}|\delta_{ij}| - \frac{1}{2}\overline{\delta}|E_b| = \frac{1}{2}\overline{\delta}|E_b|. \tag{9}$$

Define $F_g = \{ij \in E_g : |\delta_{ij}| < \frac{1}{4}\bar{\delta}\}$. Then by the condition of Case 1,

$$\frac{1}{8}\bar{\delta}|E_g| > \sum_{ij \in E_g} |\delta_{ij}| \geq \sum_{ij \in E_g \setminus F_g} |\delta_{ij}| \geq \frac{1}{4}\bar{\delta}|E_g \setminus F_g|,$$

and therefore $|E_g \setminus F_g| < \frac{1}{2}|E_g|$, or equivalently, $|F_g| > \frac{1}{2}|E_g|$.

For each $ij \in L_b$ and $k\ell \in F_g$, by Lemmas 2, 3, and Condition 3, we have $\eta(ij, k\ell) \geq \frac{\beta}{4}|\delta_{ij} - \delta_{k\ell}| \cdot \|t_{ij}\|_2 \geq \frac{\beta}{4} \cdot \frac{1}{2}|\delta_{ij}| \cdot \|t_{ij}\|_2 \geq \frac{\beta c_0 \mu_\infty}{8}|\delta_{ij}|$. Therefore by Condition 1,

$$\begin{aligned}\sum_{ij \in E_b}\sum_{k\ell \in E_g} \eta(ij, k\ell) &\geq \sum_{ij \in L_b}\sum_{k\ell \in F_g} \frac{\beta c_0 \mu_\infty}{8}|\delta_{ij}| = \sum_{ij \in L_b} |F_g| \cdot \frac{\beta c_0 \mu_\infty}{8}|\delta_{ij}| \\ &> \sum_{ij \in L_b} \frac{\beta c_0 \mu_\infty}{16}|E_g||\delta_{ij}| \geq \frac{\beta c_0 \mu_\infty}{16}|E_g| \cdot \frac{1}{2}\bar{\delta}|E_b|,\end{aligned}$$

where the last inequality follows from (9). For each $ij \in E(K_n)$, we would like to count how many times each η_{ij} appear on the left hand side. If $ij \in E_b$, then there are at most $\binom{n}{2}$ K_4s and n K_3s containing ij; hence η_{ij} may appear at most $6\binom{n}{2} + 3n = 3n^2$ times. If $ij \notin E_b$, then η_{ij} appears when there is a K_4 or a K_3 containing ij and some bad edge. By Condition 4, there are at most $2\varepsilon n$ such bad K_3s. If the bad edge in K_4 is incident to ij, then there are at most $2\varepsilon n \cdot (n-3)$ such K_4s, and if the bad edge is not incident to ij, then there are at most $|E_b| \leq \varepsilon n^2$ such K_4. Thus η_{ij} may appear at most $3 \cdot 2\varepsilon n + 6 \cdot (2\varepsilon n(n-3) + \varepsilon n^2) \leq 18\varepsilon n^2$ times. Therefore

$$\sum_{ij \in E_b}\sum_{k\ell \in E_g} \eta(ij, k\ell) \leq \sum_{ij \in E_b} 3n^2 \cdot \eta_{ij} + \sum_{ij \in E(K_n)} 18\varepsilon n^2 \cdot \eta_{ij}.$$

By Lemma 5, if $\varepsilon < \frac{c_1 p^2}{8}$, we have

$$\sum_{ij \in E_b}\sum_{k\ell \in E_g} \eta(ij, k\ell) \leq \frac{24\varepsilon}{c_1 p^2} n^2 \sum_{ij \in E_g} \eta_{ij} + \sum_{ij \in E(K_n)} 18\varepsilon n^2 \cdot \eta_{ij} \leq \frac{42\varepsilon}{c_1 p^2} n^2 \sum_{ij \in E(K_n)} \eta_{ij}.$$

Hence

$$\frac{42\varepsilon}{c_1 p^2} n^2 \sum_{ij \in E(K_n)} \eta_{ij} \geq \frac{\beta c_0 \mu_\infty}{32}|E_g| \cdot \bar{\delta}|E_b|.$$

If $\varepsilon < \frac{p}{8}$, then $|E_g| \geq \frac{n^2 p}{4} - |E_b| \geq \frac{n^2 p}{8}$. Further, if $\varepsilon < \frac{\beta c_0 c_1^2 p^4}{32 \cdot 42 \cdot 32 \cdot 8}$, then by Condition 3, $\bar{\delta} \neq 0$, and $|E_b| \neq 0$, the above implies

$$\begin{aligned}\sum_{ij \in E(K_n)} \eta_{ij} &\geq \frac{\beta c_0 c_1 p^2}{42 \cdot 32 \varepsilon n^2} \mu_\infty |E_g| \cdot \bar{\delta}|E_b| \geq \frac{\beta c_0 c_1 p^3}{42 \cdot 32 \cdot 8} \cdot \frac{1}{\varepsilon} \cdot \mu_\infty \bar{\delta}|E_b| \\ &> \frac{32}{c_1 p}\mu_\infty \cdot \bar{\delta}|E_b| \geq \frac{32}{c_1 p} \sum_{ij \in E_b} |\delta_{ij}|\|t_{ij}^{(0)}\|_2.\end{aligned}$$

Lemma 6 implies

$$\sum_{ij \in E_g} \eta_{ij} \geq \frac{c_1 p}{16} \sum_{ij \in E(K_n)} \eta_{ij} > 2 \sum_{ij \in E_b} |\delta_{ij}|\|t_{ij}^{(0)}\|_2.$$

Therefore by (8),we have $\sum_{ij \in E_g} \eta_{ij} > \sum_{ij \in E_b}(|\delta_{ij}|\|t_{ij}^{(0)}\|_2 + \eta_{ij})$ if $\varepsilon \leq \min\{\frac{c_1 p^2}{8}, \frac{p}{8}, \frac{\beta c_0 c_1^2 p^4}{32 \cdot 42 \cdot 32 \cdot 8}\}$. By (7), this shows $R(T) > R(T^{(0)})$. This condition on ε is satisfied under the assumption $\varepsilon \leq \frac{\beta c_0 c_1^2 p^4}{3 \cdot 256 \cdot 64 \cdot 32}$.

Case 2. $\bar{\delta} \neq 0$ and $\sum_{ij \in E_g} |\delta_{ij}| \geq \frac{1}{8}\overline{\delta}|E_g|$ and $|E_b| \neq 0$.

Define $E_+ = \{ij \in E_g : \delta_{ij} \geq 0\}$ and $E_- = \{ij \in E_g : \delta_{ij} < 0\}$. Since $\ell_{ij} - \ell_{ij}^{(0)} = \delta_{ij}\|t_{ij}^{(0)}\|_2$ for $ij \in E_g$, we have

$$0 = \sum_{ij \in E(G)} (\ell_{ij} - \ell_{ij}^{(0)}) \quad = \quad \sum_{ij \in E_b} (\ell_{ij} - \ell_{ij}^{(0)}) + \sum_{ij \in E_g} \delta_{ij}\|t_{ij}^{(0)}\|_2.$$

where the first equality follows from $L(T) = L(T^{(0)})$. Therefore,

$$\left| \sum_{ij \in E_g} \delta_{ij}\|t_{ij}^{(0)}\|_2 \right| \leq \left| \sum_{ij \in E_b} (\ell_{ij} - \ell_{ij}^{(0)}) \right| \leq \sum_{ij \in E_b} (|\delta_{ij}|\|t_{ij}^{(0)}\|_2 + \eta_{ij}) \leq 2\mu_\infty\overline{\delta}|E_b|,$$

where the last inequality follows from (8), Condition 3, and the definition of $\overline{\delta}$. On the other hand, the condition of Case 2 and Condition 3 implies $\sum_{ij \in E_g} |\delta_{ij}|\|t_{ij}^{(0)}\|_2 \geq \frac{1}{8}c_0\mu_\infty\overline{\delta}|E_g|$. Therefore

$$\sum_{ij \in E_-} (-\delta_{ij})\|t_{ij}^{(0)}\|_2 = \frac{1}{2}\left(-\sum_{ij \in E_g} \delta_{ij}\|t_{ij}^{(0)}\|_2 + \sum_{ij \in E_g} |\delta_{ij}|\|t_{ij}^{(0)}\|_2 \right) \geq \frac{1}{2}\left(\frac{1}{8}c_0\mu_\infty\overline{\delta}|E_g| - 2\mu_\infty\overline{\delta}|E_b| \right).$$

If $\varepsilon \leq \frac{1}{256}c_0 p$, then since $|E_b| \leq \varepsilon n^2$ and $|E_g| \geq \frac{1}{4}n^2 p - |E_b| \geq \frac{1}{8}n^2 p$, we see that $\frac{1}{8}c_0\mu_\infty\overline{\delta}|E_g| - 2\mu_\infty\overline{\delta}|E_b| \geq \frac{1}{16}c_0\mu_\infty\overline{\delta}|E_g|$. Therefore $\sum_{ij \in E_-} (-\delta_{ij})\|t_{ij}^{(0)}\|_2 \geq \frac{1}{32}c_0\mu_\infty\overline{\delta}|E_g|$. Similarly, $\sum_{ij \in E_+} \delta_{ij}\|t_{ij}^{(0)}\|_2 \geq \frac{1}{32}c_0\mu_\infty\overline{\delta}|E_g|$.

If $|E_+| \geq \frac{1}{2}|E_g|$, then by Lemmas 2, 3, and Condition 3, we have

$$\begin{aligned} \sum_{ij \in E_-} \sum_{k\ell \in E_+} \eta(ij, k\ell) &\geq \sum_{ij \in E_-} \sum_{k\ell \in E_+} \frac{\beta}{4}(-\delta_{ij})\|t_{ij}^{(0)}\|_2 \\ &\geq \sum_{ij \in E_-} (-\delta_{ij})\|t_{ij}^{(0)}\|_2 \cdot \frac{\beta}{4}|E_+| \geq \frac{\beta}{4}|E_+| \cdot \frac{1}{32}c_0\mu_\infty\overline{\delta}|E_g| \\ &\geq \frac{\beta}{256}c_0\mu_\infty\overline{\delta}|E_g|^2. \end{aligned}$$

Similarly, if $|E_-| \geq \frac{1}{2}|E_g|$, then we can switch the order of summation and consider $\sum_{ij \in E_+} \sum_{k\ell \in E_-} \eta(ij, k\ell)$ to obtain the same conclusion.

Since each edge is contained in at most $\frac{n(n-1)}{2}$ copies of K_4 and n copies of K_3 (and there are 6 edges in a K_4, 3 edges in a K_3), we have

$$\sum_{ij \in E_-} \sum_{k\ell \in E_+} \eta(ij, k\ell) \leq \left(6\frac{n(n-1)}{2} + 3n \right) \sum_{ij \in E(K_n)} \eta_{ij} \leq 3n^2 \sum_{ij \in E(K_n)} \eta_{ij}.$$

If $\varepsilon \leq \frac{p}{8}$, then $|E_g| \geq \frac{1}{4}n^2 p - |E_b| \geq \frac{1}{8}n^2 p$. Further, if $\varepsilon < \frac{\beta c_0 c_1 p^3}{3 \cdot 256 \cdot 64 \cdot 32}$, then since $\bar{\delta} \neq 0$ and $|E_b| \leq \varepsilon n^2$, we have

$$\sum_{ij \in E(K_n)} \eta_{ij} \geq \frac{1}{3n^2} \cdot \frac{\beta c_0\mu_\infty\overline{\delta}}{256}|E_g|^2 \geq \frac{\beta c_0 p^2}{3 \cdot 256 \cdot 64}\mu_\infty\overline{\delta}n^2 > \frac{32}{c_1 p}\mu_\infty\overline{\delta}|E_b|.$$

By Lemma 6, if $\varepsilon < \frac{c_1 p^2}{8}$, then this implies

$$\sum_{ij \in E_g} \eta_{ij} \geq \frac{c_1 p}{16} \sum_{ij \in E(K_n)} \eta_{ij} > 2\mu_\infty\overline{\delta}|E_b|.$$

Therefore from (7), (8), and Condition 3, if $\varepsilon \leq \min\{\frac{c_0 p}{256}, \frac{c_1 p^2}{8}, \frac{p}{8}, \frac{\beta c_0 c_1 p^3}{3\cdot 256\cdot 64\cdot 32}\}$, then

$$\begin{aligned} R(T) &\geq R(T^{(0)}) + \sum_{ij\in E_g} \eta_{ij} - \sum_{ij\in E_b} (|\delta_{ij}|\|t_{ij}^{(0)}\|_2 + \eta_{ij}) \\ &> R(T^{(0)}) + 2\mu_\infty \overline{\delta}|E_b| - \sum_{ij\in E_b} 2|\delta_{ij}|\|t_{ij}^{(0)}\|_2 \geq R(T^{(0)}). \end{aligned}$$

This condition on ε is satisfied under the assumption $\varepsilon \leq \frac{\beta c_0 c_1^2 p^4}{3\cdot 256\cdot 64\cdot 32}$. ☐

2.5 Properties of Gaussians in high dimensions

In this section, we prove that i.i.d. Gaussians satisfy properties needed to establish Conditions 2, 3, and 5 in Theorem 3. We begin by recording some useful facts regarding concentration of random Gaussian vectors:

Lemma 7. *Let x, y be i.i.d. $\mathcal{N}(0, I_{d\times d})$, and $\epsilon \leq 1$, then*

$$\mathbb{P}\left(d(1-\epsilon) \leq \|x\|_2^2 \leq d(1+\epsilon)\right) \geq 1 - e^{-c\epsilon^2 d}$$

and

$$\mathbb{P}\left(|\langle x, y\rangle| \geq d\epsilon\right) \leq e^{-c\epsilon^2 d}$$

where $c > 0$ is an absolute constant.

Proof. Both statements follow from Corollary 5.17 in [27], concerning concentration of sub-exponential random variables. ☐

Lemma 8 ([27] Corollary 5.35)**.** *Let A be an $n \times d$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries. Then for any $t \geq 0$,*

$$\mathbb{P}\left(\sigma_{\max}(A) \geq \sqrt{n} + \sqrt{d} + t\right) \leq 2e^{-\frac{t^2}{2}}$$

where $\sigma_{\max}(A)$ is the largest singular value of A.

Lemma 9. *Let $t_i^{(0)}, i \in [n]$ be i.i.d. $\mathcal{N}(0, I_{d\times d})$. Then, there exists an event E, such that on E, we have for all $i, j, k, l \in [n], i \neq j, k \neq l$,*

$$\frac{\|t_{ij}^{(0)}\|_2}{\|t_{kl}^{(0)}\|_2} \geq \frac{9}{10}$$

and for all distinct $i, j, k \in [n]$,

$$\langle \hat{t}_{ij}^{(0)}, \hat{t}_{ik}^{(0)}\rangle^2 \leq 1/3$$

and $\mathbb{P}(E^c) \leq 3n^2 e^{-cd}$, where $c > 0$ is an absolute constant.

Proof. This follows from repeated application of Lemma 7 with $\epsilon = 1/100$ and a union bound. ☐

We can now show that gaussian vectors have the well-distributed property with high probability. Recall that $S(x, y) = \operatorname{span}(x, y)$.

Lemma 10. *Let $t_1, \dots t_n \in \mathbb{R}^d$ be i.i.d. $\mathcal{N}(0, I_{d\times d})$, and let $n \geq 16$ and $d \geq 3$. For a fixed $k \neq l$, the inequality*

$$\sum_{i\in[n], i\neq l,k} \|P_{S(t_l - t_i, t_k - t_i)^\perp}(h)\|_2 \geq \frac{1}{5}(n-2)\|P_{S(t_l - t_k)^\perp}(h)\|_2$$

holds for all $h \in \mathbb{R}^d$ with probability of failure at most $5ne^{-cd}$, where $c > 0$ is an absolute constant.

Proof. Throughout the proof, constants named c may be different from line to line, but are always bounded below by a positive absolute constant. For a fixed (l,k), let $x = t_l, y = t_k$. We would like to show

$$\sum_{i=1}^{n} \|P_{S(x-t_i,y-t_i)^\perp}(h)\|_2 \geq \frac{1}{5}n\|P_{S(x-y)^\perp}(h)\|_2$$

We note that $S(x-t_i, y-t_i) = S(x-y, x+y-2t_i)$. Thus,

$$P_{S(x-t_i,y-t_i)^\perp}(h) = P_{S(x-y,x+y-2t_i)^\perp}(h) = P_{S(x-y,x+y-2t_i)^\perp}(P_{S(x-y)^\perp}(h))$$

Thus, it's enough to show

$$\sum_{i=1}^{n} \|P_{S(x-y,x+y-2t_i)^\perp}(h)\|_2 \geq \frac{1}{5}n\|h\|_2$$

for $h \perp (x-y)$.

Now, for any vectors v, w, we have

$$S(v,w) = S(v, w_{v^\perp})$$

where $w_{v^\perp} = w - \langle w, \hat{v}\rangle\hat{v}$. If $h \perp v$, we have

$$\begin{aligned}
P_{S(v,w)^\perp}(h) &= P_{S(v,w_{v^\perp})^\perp}(h) \\
&= h - \langle h, \hat{v}\rangle\hat{v} - \langle h, \hat{w}_{v^\perp}\rangle\hat{w}_{v^\perp} \\
&= h - \left\langle h, \frac{w}{\|w_{v^\perp}\|_2}\right\rangle \frac{w_{v^\perp}}{\|w_{v^\perp}\|_2} \\
&= h - \langle h, \hat{w}\rangle\hat{w} + \langle h, \hat{w}\rangle\left[\hat{w} - \frac{\|w\|_2}{\|w_{v^\perp}\|_2}\frac{w_{v^\perp}}{\|w_{v^\perp}\|_2}\right] \\
&= P_{S(w)^\perp}(h) + \langle h, \hat{w}\rangle z
\end{aligned}$$

Where $z = \hat{w} - \frac{\|w\|_2}{\|w_{v^\perp}\|_2}\frac{w_{v^\perp}}{\|w_{v^\perp}\|_2}$. Now, assuming that $|\langle \hat{v}, \hat{w}\rangle| < 1/2$ and using that

$$\|w_{v^\perp}\|_2^2 = \|w - \langle w, \hat{v}\rangle\hat{v}\|_2^2 = \|w\|_2^2 - 2\|w\|_2^2\langle \hat{w}, \hat{v}\rangle + \|w\|_2^2|\langle \hat{w}, \hat{v}\rangle|^2 \geq \|w\|_2^2(1 - 2|\langle \hat{w}, \hat{v}\rangle|)$$

we have

$$\begin{aligned}
\|z\|_2 &= \left\|\hat{w} - \frac{\|w\|_2}{\|w_{v^\perp}\|_2^2}(w - \langle w, \hat{v}\rangle\hat{v})\right\|_2 \\
&= \left\|\hat{w}\left[1 - \frac{\|w\|_2^2}{\|w_{v^\perp}\|_2^2}\right] + \frac{\|w\|_2^2}{\|w_{v^\perp}\|_2^2}\langle \hat{w}, \hat{v}\rangle\hat{v}\right\|_2 \\
&\leq \left|1 - \frac{\|w\|_2^2}{\|w_{v^\perp}\|_2^2}\right| + \frac{\|w\|_2^2}{\|w_{v^\perp}\|_2^2}|\langle \hat{w}, \hat{v}\rangle| = \epsilon(\langle \hat{w}, \hat{v}\rangle) \\
&= \frac{\|w\|_2^2}{\|w_{v^\perp}\|_2^2}(1 + |\langle \hat{w}, \hat{v}\rangle|) - 1 \\
&\leq \frac{3|\langle \hat{w}, \hat{v}\rangle|}{1 - 2|\langle \hat{w}, \hat{v}\rangle|} \triangleq \zeta(\langle \hat{w}, \hat{v}\rangle)
\end{aligned}$$

Thus, we have

$$\|P_{S(v,w)^\perp}(h)\|_2 \geq \|P_{S(w)^\perp}(h)\|_2 - \zeta(\langle \hat{w}, \hat{v}\rangle)\|h\|_2$$

Therefore, by taking $v = x-y$ and $w = x+y-2t_i$, to conclude the desired statement of the present Lemma, it suffices to show that

$$\sum_{i=1}^{n} \|P_{S(x+y-2t_i)^{\perp}}(h)\|_2 \geq \gamma n \|h\|_2$$

where $\gamma > 1/5 + \zeta\left(\frac{\langle x-y, x+y+2t_i\rangle}{\|x-y\|_2\|x+y-2t_i\|_2}\right)$. Note that $x-y$ and $x+y-2t_i$ are independent, and $\frac{1}{2}(x-y) =^d \frac{1}{6}(x+y-2t_i) =^d \mathcal{N}(0, I_{d\times d})$. Applying Lemma 7 to $x-y$ and $x+y-2t_i$ with a small enough value of ϵ to ensure $\zeta\left(\frac{\langle x-y, x+y+2t_i\rangle}{\|x-y\|_2\|x+y-2t_i\|_2}\right) < 1/20$, we get

$$\mathbb{P}\left(\zeta\left(\frac{\langle x-y, x+y+2t_i\rangle}{\|x-y\|_2\|x+y-2t_i\|_2}\right) > \frac{1}{20}\right) \leq 3e^{-cd}$$

Thus, it suffices to show with high probability, that

$$\sum_{i=1}^{n} \|P_{S(x+y-2t_i)^{\perp}}(h)\|_2 \geq 0.3n\|h\|_2,$$

which we proceed to establish below.

To begin, redefine v, w as $v = x+y$ and $w = -2t_i$ and consider

$$\begin{aligned}
\sum_{i=1}^{n} \left\|P_{S(v+w_i)^{\perp}}(h)\right\|_2 &\geq \left\|\sum_{i=1}^{n} P_{S(v+w_i)^{\perp}}(h)\right\|_2 \\
&= \left\|\sum_{i=1}^{n}\left(h - \frac{1}{\|v+w_i\|_2^2}\langle h, v+w_i\rangle(v+w_i)\right)\right\|_2 \\
&\geq n\|h\|_2 - \left\|\sum_{i=1}^{n} \frac{1}{\|v+w_i\|_2^2}(v+w_i)(v+w_i)^* h\right\|_2 \\
&\geq n\|h\|_2 - \left\|\sum_{i=1}^{n} \frac{1}{\|v+w_i\|_2^2}(v+w_i)(v+w_i)^*\right\|_{\text{op}} \|h\|_2 \\
&\geq \|h\|_2 \left[n - \frac{1}{\min_i \|v+w_i\|_2^2}\left\|\sum_{i=1}^{n}(v+w_i)(v+w_i)^*\right\|_{\text{op}}\right]
\end{aligned}$$

where in the last inequality we used

$$\sum_{i=1}^{n} \frac{1}{\|v+w_i\|_2^2}(v+w_i)(v+w_i)^* \preceq \frac{1}{\min_i \|v+w_i\|_2^2}\sum_{i=1}^{n}(v+w_i)(v+w_i)^*$$

Now, let $A = \sum_{i=1}^n e_i w_i^* \in \mathbb{R}^{n\times d}$. We have

$$\begin{aligned}\left\|\sum_{i=1}^n (v+w_i)(v+w_i)^*\right\|_{\text{op}} &= \left\|\sum_{i=1}^n (vv^* + vw_i^* + w_i v^* + w_i w_i^*)\right\|_{\text{op}} \\ &\le n\|vv^*\|_{\text{op}} + \left\|v\left(\sum_{i=1}^n w_i\right)^* + \left(\sum_{i=1}^n w_i\right)v^*\right\|_{\text{op}} + \left\|\sum_{i=1}^n w_i w_i^*\right\|_{\text{op}} \\ &\le n\|v\|_2^2 + 2\|v\|_2\left\|\sum_{i=1}^n w_i\right\|_2 + \left\|\sum_{i=1}^n w_i w_i^*\right\|_{\text{op}} \\ &= n\|v\|_2^2 + 2\|v\|_2\left\|\sum_{i=1}^n w_i\right\|_2 + \sigma_{\max}(A)^2\end{aligned}$$

Thus,

$$\sum_{i=1}^n \left\|P_{S(v+w_i)^\perp}(h)\right\|_2 \ge \|h\|_2\left[n - \frac{n\|v\|_2^2 + 2\|v\|_2\left\|\sum_{i=1}^n w_i\right\|_2 + \sigma_{\max}(A)^2}{\min_i \|v+w_i\|_2^2}\right]$$

Now, consider the event

$$E = \left\{\min_i \|v+w_i\|_2^2 \ge 6d\beta_1, \quad \|v\|_2^2 \le 2d\beta_2, \quad \left\|\sum_{i=1}^n w_i\right\|_2^2 \le 4nd\beta_3, \quad \sigma_{\max}(A)^2 \le n\beta_4\right\}$$

On E, we have

$$\begin{aligned}\sum_{i=1}^n \left\|P_{S(v+w_i)^\perp}(h)\right\|_2 &\ge \|h\|_2\left[n - \frac{1}{6d\beta_1}\left(2nd\beta_2 + 2\sqrt{2d\beta_2}2\sqrt{nd}\sqrt{\beta_3} + n\beta_4\right)\right] \\ &= \|h\|_2\left[n - \frac{1}{3}n\frac{\beta_2}{\beta_1} - \frac{4\sqrt{2}d\sqrt{n}\sqrt{\beta_2\beta_3}}{6d\beta_1} - \frac{\beta_4}{6d\beta_1}n\right] \\ &= \|h\|_2\left[n\left(1 - \frac{1}{3}\frac{\beta_2}{\beta_1} - \frac{\beta_4}{6d\beta_1} - \frac{1}{\sqrt{n}}\frac{4\sqrt{2}\sqrt{\beta_2\beta_3}}{6\beta_1}\right)\right]\end{aligned}$$

Now, note that $\frac{1}{6}\|v+w_i\|_2^2 =^d \frac{1}{2}\|v\|_2^2 =^d \frac{1}{4n}\left\|\sum_{i=1}^n w_i\right\|_2^2 =^d \chi^2(d)$ and A is a random $n\times d$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries.

Thus by applying Lemma 7 we have

$$\mathbb{P}\left(6d(1-\epsilon) \le \|v+w_i\|_2^2 \le 6d(1+\epsilon)\right) \ge 1 - e^{-c\epsilon^2 d}$$

$$\mathbb{P}\left(2d(1-\epsilon) \le \|v\|_2^2 \le 2d(1+\epsilon)\right) \ge 1 - e^{-c\epsilon^2 d}$$

$$\mathbb{P}\left(4nd(1-\epsilon) \le \left\|\sum_{i=1}^n w_i\right\|_2^2 \le 4nd(1+\epsilon)\right) \ge 1 - e^{-c\epsilon^2 d}$$

where $c > 0$ is a universal constant. Also by taking $t = \sqrt{2d}$ in Lemma 8 we get

$$\mathbb{P}\left(\sigma_{\max}(A) \geq \sqrt{n} + 2\sqrt{d}\right) \leq 2e^{-d}$$

Now, let $\beta_1 = 1 - \frac{1}{100}, \quad \beta_2 = \beta_3 = 1 + \frac{1}{100}, \quad \beta_4 = \frac{d}{2}$, which gives

$$\frac{1}{3}\frac{\beta_2}{\beta_1} \leq 1/3 + 1/99, \quad \frac{1}{\sqrt{n}}\frac{4\sqrt{2}\sqrt{\beta_2\beta_3}}{6\beta_1} < \frac{1}{\sqrt{n}}, \quad \frac{\beta_4}{5d} = 1/10$$

We have

$$\mathbb{P}\left(\sigma_{\max}(A) \geq \sqrt{n\beta_4}\right) \leq \mathbb{P}\left(\sigma_{\max}(A) \geq \sqrt{n} + 2\sqrt{d}\right) \leq 2e^{-d}$$

whenever $\sqrt{n} + 2\sqrt{d} \leq \sqrt{n}\sqrt{d/2}$, which holds whenever

$$n \geq \left(\frac{2\sqrt{d}}{\sqrt{d/2} - 1}\right)^2$$

which holds for $n \geq 16$ when $d \geq 3$. Since for $n \geq 16$, $\frac{1}{\sqrt{n}} \leq 1/4$, we have on E

$$\sum_{i=1}^{n} \left\| P_{S(v+w_i)^\perp}(h) \right\|_2 \geq 0.3n\|h\|_2$$

Thus,

$$\mathbb{P}\left(\sum_{i=1}^{n} \left\| P_{S(v+w_i)^\perp}(h) \right\|_2 < 0.3n\|h\|_2\right) \leq \mathbb{P}(E^c) \leq (n+3)e^{-cd}$$

where $c > 0$ is an absolute constant.

Combining all of the above, we get

$$\sum_{i=1}^{n} \|P_{S(x-t_i, y-t_i)^\perp}(h)\|_2 \geq \frac{1}{5}n\|P_{S(x-y)^\perp}(h)\|_2$$

with probability of failure at most $5ne^{-cd}$.

□

Lemma 11. *Let $G([n], E)$ be p-typical, and $t_1, \ldots t_n \sim \mathcal{N}(0, I_{d\times d})$ be i.i.d. Then $T = \{t_i\}_{i\in[n]}$ is $\frac{1}{5}$-well distributed along G with probability at least $1 - 10n^3e^{-cd}$, where $c > 0$ is an absolute constant.*

Proof. For each $ij \in E$, let $S_{ij} = \{k \in [n]; ik, jk \in E(G)\}$ and note that $|S_{ij}| \leq 2np^2$. Now apply Lemma 10 to the set of vectors $\{t_i, t_j\} \bigcup \{t_k\}_{k\in\mathcal{I}_{ij}}$, with the distinguished vectors being $\{t_i, t_j\}$, which gives the desired property for the pair (i, j) with probability of failure at most $5(|S_{ij}|)e^{-cd} \leq 5(2np^2)e^{-cd}$, where $c > 0$ is an absolute constant. Taking the union bound over pairs of distinct integers $i, j \in [n]$, we get the desired property simultaneously for all pairs with probability at least $1 - n^2 \cdot 5(2np^2)e^{-cd} = 1 - 10n^3p^2e^{-cd} \geq 1 - 10n^3e^{-cd}$. □

2.6 Random graphs are p-typical with high probability

Lemma 12. *There exists an absolute constant $c > 0$ such that for all positive real numbers $p \leq 1$, $G(n,p)$ is p-typical with probability at least $1 - n^2 e^{-cnp^2}$ if $np \geq 4 \log n$.*

Proof. A graph is not connected only if there exists a partition $V_1 \cup V_2$ of its vertex set for which there are no edges between V_1 and V_2. Without loss of generality, we may assume that $|V_1| \leq \lfloor \frac{n}{2} \rfloor$. Since the number of ways to choose a set of size k from a set of size n is $\binom{n}{k}$, the probability that $G(n,p)$ is not connected is at most

$$\sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^{k(n-k)} \leq \sum_{k=1}^{\lfloor n/2 \rfloor} \left(\frac{en}{k}\right)^k e^{-pk(n-k)} < \sum_{k=1}^{\lfloor n/2 \rfloor} \left(ne^{1-p(n-k)}\right)^k.$$

Since $k \leq \lfloor \frac{n}{2} \rfloor$, we have $ne^{1-p(n-k)} \leq ne^{1-pn/2} < 1$ (since $np \geq 4 \log n$). Therefore the summand on the right-hand-side is at most $(ne^{1-pn/2})^k$, which is maximized at $k = 1$. This shows that the probability that $G(n,p)$ is not connected is at most $n^2 e^{1-pn/2}$.

In $G(n,p)$, for a fixed vertex v, the expected value of $\deg(v)$is $(n-1)p$, and for a pair of vertices v, w, the expected value of the codegree of v and w is $(n-2)p^2$. Therefore the lemma follows from Chernoff's inequality — see Fact 4 from [1] — and a union bound. □

2.7 Proof of Theorem 1

We can now prove the high dimensional recovery theorem, which we state here again for convenience:

Theorem 1. *Let $G([n], E)$ be drawn from $G(n,p)$ for some $p = \Omega(n^{-1/4})$. Take $t_1^{(0)}, \ldots t_n^{(0)} \sim \mathcal{N}(0, I_{d\times d})$ to be i.i.d., independent from G. There exists an absolute constant $c > 0$ and a $\gamma = \Omega(p^4)$ not depending on n, such that if $\max(\frac{2^6}{c^6}, \frac{4^3}{c^3} \log^3 n) \leq n \leq e^{\frac{1}{6}cd}$ and $d = \Omega(1)$, then there exists an event with probability at least $1 - e^{-n^{1/6}} - 13e^{-\frac{1}{2}cd}$, on which the following holds:*

For arbitrary subgraphs E_b satisfying $\max_i \deg_b(i) \leq \gamma n$ and arbitrary pairwise direction corruptions $v_{ij} \in \mathbb{S}^{d-1}$ for $ij \in E_b$, the convex program (3) has a unique minimizer equal to $\left\{\alpha\left(t_i^{(0)} - \bar{t}^{(0)}\right)\right\}_{i\in[n]}$ for some positive α and for $\bar{t}^{(0)} = \frac{1}{n}\sum_{i\in[n]} t_i^{(0)}$.

Proof. It is enough to verify that G, T and E_b in the assumption of the present theorem satisfy the deterministic conditions 1–6 in Theorem 2, with appropriate constants $p, \beta, c_0, \epsilon, c_1$, and with the purported probability. By Lemma 12, Lemma 9, and Lemma 11, we have that Condition 1 holds with value p, Condition 2 holds with $\beta = \sqrt{\frac{2}{3}}$, Condition 3 holds with $c_0 = \frac{9}{10}$, and Condition 5 holds with $c_1 = \frac{1}{5}$, with probability at least

$$1 - n^2 e^{-cnp^2} - 3n^2 e^{-cd} - 10n^3 e^{-cd}$$

where $c > 0$ is an absolute constant.

Thus, taking any E_b, which satisfies Condition 4 with $\gamma = \frac{p^4}{10^7} \leq \frac{\beta c_0 c_1^2 p^4}{256\cdot 32\cdot 64\cdot 3}$, we get that recovery via ShapeFit is guaranteed. Note that the condition $\max \deg_b(i) \leq \gamma n$ is nontrivial when $p = \Omega(n^{-1/4})$. Using the requirements on n and p, we have $n^2 e^{-cnp^2} \leq n^2 e^{-cn^{1/3}} \leq e^{-\frac{1}{6}n}$ and $13n^3 e^{-cd} \leq 13(e^{\frac{1}{6}cd})^3 e^{-cd} \leq 13e^{-\frac{1}{2}cd}$. Thus, the probability of exact recovery via ShapeFit, uniformly in E_b and v_{ij} satisfying the assumptions of the theorem, is at least

$$1 - e^{-n^{1/6}} - 13e^{-\frac{1}{2}cd}.$$

□

3 Proof of three-dimensional recovery

The proof of recovery in three dimensions parallels the proof in high dimensions, but it is more technical because it can not capitalize on the concentration of measure phenomenon in high dimensions. Specifically, the additional technicality in three dimensions comes from the fact that for large n, there exist pairs of locations $t_i^{(0)}$, $t_j^{(0)}$ that are close to each other, i.e., $\|t_{ij}^{(0)}\|_2$ is small. For such pairs of vectors, with high probability, for all $k \neq i, j$ the value of $1 - \langle \hat{t}_{ik}^{(0)}, \hat{t}_{jk}^{(0)} \rangle^2$ will be small. This fact introduces the following two main obstacles in carrying out the same analysis:

1. There is no uniform lower bound on $1 - \langle \hat{t}_{ik}^{(0)}, \hat{t}_{jk}^{(0)} \rangle^2$. Hence Condition 2 in Theorem 3 fails.
2. There is no uniform lower bound on $\|t_{ij}^{(0)}\|_2$. Hence Condition 3 in Theorem 3 fails.

These are indeed obstacles since the gains in rotational motions coming from Lemmas 2 and 3 are proportional to $\sqrt{1 - \langle \hat{t}_{ik}^{(0)}, \hat{t}_{jk}^{(0)} \rangle^2}$ and $\|t_{ij}^{(0)}\|_2$. We avoid these difficulties and prove the three-dimensional analogue of Theorem 3 by weakening Conditions 2 and 3. Roughly speaking, in $\mathbb{R}^3$, Condition 2 holds for most triples $i, j, k \in [n]$ (instead of all triples) and Condition 3 gets replaced by a one-sided version where we only have a uniform upper bound on the lengths $\|t_{ij}^{(0)}\|_2$.

Unlike in the high-dimensional case where we allowed a constant fraction of edges incident to each vertex to be corrupted, the three-dimensional case requires the fraction of corrupted edges incident to each vertex to be at most $O(\frac{1}{\log^3 n})$. This additional poly-logarithmic factor is due to the fact that our well-distributedness proof in three dimensions hinges on the maximum ℓ_2 norm of locations, which is $\Omega(\sqrt{\log n})$ with high probability. It can be removed for a distribution of locations that has a uniform constant upper bound on $\|t_i^{(0)}\|_2$.

3.1 Deterministic recovery theorem in three dimensions

We now state deterministic conditions on the graph G, the corrupted observations E_b, and the locations $T^{(0)}$ that guarantee recovery. Recall the definition $\mu = \frac{1}{|E(G)|} \sum_{ij \in E(G)} \|t_{ij}^{(0)}\|_2$.

Theorem 4. *Suppose $T^{(0)}, E_b, G$ satisfy the conditions*

1. *The underlying graph G is p-typical,*
2. *For all distinct $i, j \in [n]$, for all but at most $\varepsilon_1 n$ indices $k \in [n]$ satisfying $k \neq i, j$, we have $1 - \langle \hat{t}_{ij}, \hat{t}_{ik} \rangle^2 \geq \beta^2$ and $1 - \langle \hat{t}_{ij}, \hat{t}_{jk} \rangle^2 \geq \beta^2$,*
3. *For all distinct $i, j \in [n]$, we have $\|t_{ij}^{(0)}\|_2 \leq c_0 \mu$,*
4. *Each vertex has at most $\varepsilon_0 n$ edges in E_b incident to it,*
5. *The set $\{t_i^{(0)}\}_{i \in [n]}$ is c_1-well-distributed along G,*
6. *No three vectors $t_i^{(0)}, t_j^{(0)}, t_k^{(0)}$ are collinear for distinct i, j, k.*

for constants $0 < p, \beta, \varepsilon_0, \varepsilon_1, c_1 \leq 1 \leq c_0$. If $\varepsilon_0 \leq \frac{\beta c_1^2 p^4}{32 \cdot 3 \cdot 64 \cdot 1024 c_0^2}$ and $\varepsilon_1 \leq \frac{p}{192 c_0}$, then $L(T^{(0)}) \neq 0$ and $T^{(0)}/L(T^{(0)})$ is the unique optimizer of ShapeFit.

Note that all six conditions are invariant under translation and non-zero scalings of $T^{(0)}$ (Condition 3 is invariant since both $t_{ij}^{(0)}$ and μ scale together and are invariant under translation). Before we prove the theorem, we establish that $L(T^{(0)}) \neq 0$ when ε_0 is small. This non-equality guarantees that some scaling of $T^{(0)}$ is feasible whenever, roughly speaking, $|E_b| < |E_g|$.

Lemma 13. *If $\varepsilon_0 < \frac{p}{8c_0}$, then $L(T^{(0)}) \neq 0$.*

Proof. Since $v_{ij} = \hat{t}_{ij}^{(0)}$ for all $ij \in E_g$, we have

$$L(T^{(0)}) = \sum_{ij \in E(G)} \langle t_{ij}^{(0)}, v_{ij} \rangle \geq \sum_{ij \in E_g} \|t_{ij}^{(0)}\|_2 - \sum_{ij \in E_b} \|t_{ij}^{(0)}\|_2 = \sum_{ij \in E(G)} \|t_{ij}^{(0)}\|_2 - 2 \sum_{ij \in E_b} \|t_{ij}^{(0)}\|_2.$$

By Condition 3, $\sum_{ij \in E_b} \|t_{ij}^{(0)}\|_2 \leq c_0 \mu |E_b| \leq c_0 \mu \cdot \varepsilon_0 n^2 < \frac{1}{8} n^2 p \mu$. Since Condition 1 implies $|E(G)| \geq \frac{1}{4} n^2 p$, we have $\sum_{ij \in E(G)} \|t_{ij}^{(0)}\|_2 \geq \frac{1}{4} n^2 p \mu$. Therefore it follows that $L(T^{(0)}) > 0$. □

3.2 Proof of Theorem 4

Lemmas 2 and 3 will be repeatedly used throughout the proof. Note that these lemmas can be used only if the given set of vectors satisfies a certain condition on the angles between them. For each distinct $ij \in E(K_n)$, define $B(ij)$ as the set of edges $k\ell \in E(K_n)$ such that $\sqrt{1 - \langle \hat{t}_{ac}^{(0)}, \hat{t}_{bc}^{(0)} \rangle^2} < \beta$ holds for some distinct $a, b, c \in \{i, j, k, \ell\}$ satisfying $(a, b) \neq (i, j)$. Note that Lemmas 2 and 3 can be applied to the set of indices $\{i, j, k, \ell\}$ (having size either 3 or 4) for all $k\ell \notin B(ij)$. The following lemma shows that $B(ij)$ is small for each ij.

Lemma 14. *For each $ij \in E(K_n)$, we have $|B(ij)| \leq 6\varepsilon_1 n^2$.*

Proof. For each $ab \in E(K_n)$, define $B_3(ab)$ as the set of indices $c \in [n]$ distinct from a, b for which $\sqrt{1 - \langle \hat{t}_{ab}^{(0)}, \hat{t}_{ac}^{(0)} \rangle^2} < \beta$ or $\sqrt{1 - \langle \hat{t}_{ab}^{(0)}, \hat{t}_{bc}^{(0)} \rangle^2} < \beta$ holds. Condition 2 implies $|B_3(ab)| \leq \varepsilon_1 n$ for all $ab \in E(K_n)$. One can check that $k\ell \in B(ij)$ if and only if one of the following events hold: $k \in B_3(ij)$, $\ell \in B_3(ij)$, $k \in B_3(i\ell) \cup B_3(j\ell)$, $\ell \in B_3(ik) \cup B_3(jk)$. Therefore

$$\begin{aligned} |B(ij)| &\leq 2|B_3(ij)| \cdot n + \sum_{\ell \neq i,j} \Big(|B_3(i\ell)| + |B_3(j\ell)| \Big) + \sum_{k \neq i,j} \Big(|B_3(ik)| + |B_3(jk)| \Big) \\ &\leq 2\varepsilon_1 n^2 + n \cdot 2\varepsilon_1 n + n \cdot 2\varepsilon_1 n = 6\varepsilon_1 n^2. \end{aligned}$$ □

We now prove the deterministic recovery theorem in three dimensions.

Proof of Theorem 4. By Lemma 13 and the fact that Conditions 1–6 are invariant under global translation and nonzero scaling, we can take $\bar{t}^{(0)} = 0$ and $L(T^{(0)}) = 1$ without loss of generality. The variable μ is to be understood accordingly.

We will directly prove that $R(T) > R(T^{(0)})$ for all $T \neq T^{(0)}$ such that $L(T) = 1$ and $\bar{t} = 0$. Consider an arbitrary feasible T and recall the notation $t_{ij} = (1 + \delta_{ij}) t_{ij}^{(0)} + \eta_{ij} s_{ij}$ where s_{ij} is a unit vector orthogonal to $t_{ij}^{(0)}$ and $\eta_{ij} = \|P_{t_{ij}^{(0)\perp}} t_{ij}\|_2$. A useful lower bound for the objective $R(T)$ is given by

$$\begin{aligned} R(T) = \sum_{ij \in E(G)} \|P_{v_{ij}^\perp} t_{ij}\|_2 &= \sum_{ij \in E_g} \eta_{ij} + \sum_{ij \in E_b} \|P_{v_{ij}^\perp} t_{ij}\|_2 \\ &\geq \sum_{ij \in E_g} \eta_{ij} + \sum_{ij \in E_b} \Big(\|P_{v_{ij}^\perp} t_{ij}^{(0)}\|_2 - |\delta_{ij}| \|t_{ij}^{(0)}\|_2 - \eta_{ij} \Big) \\ &= R(T^{(0)}) + \sum_{ij \in E_g} \eta_{ij} - \sum_{ij \in E_b} (|\delta_{ij}| \|t_{ij}^{(0)}\|_2 + \eta_{ij}). \end{aligned} \quad (10)$$

Suppose that $\sum_{ij\in E_b} |\delta_{ij}|\|t_{ij}^{(0)}\|_2 < \sum_{ij\in E_b} \eta_{ij}$. Since $\varepsilon_0 \le \frac{c_1p^2}{16}$, Lemma 5 implies $\sum_{ij\in E_b} \eta_{ij} \le \frac{1}{2}\sum_{ij\in E_g} \eta_{ij}$. Therefore by (10), we have

$$\begin{aligned} R(T) &\ge R(T_0) + \sum_{ij\in E_g} \eta_{ij} - \sum_{ij\in E_b} (|\delta_{ij}|\|t_{ij}^{(0)}\|_2 + \eta_{ij}) \\ &> R(T_0) + \sum_{ij\in E_g} \eta_{ij} - \sum_{ij\in E_b} 2\eta_{ij} \ge R(T_0). \end{aligned}$$

Hence we may assume

$$\sum_{ij\in E_b} |\delta_{ij}|\|t_{ij}^{(0)}\|_2 \ge \sum_{ij\in E_b} \eta_{ij}. \tag{11}$$

In other words, the total parallel motion is larger than the total rotational motions on the bad edges. The key idea of the proof is to show that parallel motions on bad edges induce a large amount of rotational motions on good edges.

In the case $|E_b| \ne 0$, define $\overline{\delta} := \frac{\sum_{ij\in E_b} |\delta_{ij}|\|t_{ij}^{(0)}\|_2}{\sum_{ij\in E_b} \|t_{ij}^{(0)}\|_2}$ as the average 'relative parallel motion' on the bad edges. For distinct $ij, k\ell \in E(K_n)$, if $\{i,j\}\cap\{k,\ell\} = \emptyset$, then define $\eta(ij,k\ell) = \eta_{ij} + \eta_{ik} + \eta_{i\ell} + \eta_{jk} + \eta_{j\ell} + \eta_{k\ell}$, and if $\{i,j\}\cap\{k,\ell\} \ne \emptyset$ (without loss of generality, assume $\ell = i$), then define $\eta(ij,k\ell) = \eta_{ij} + \eta_{ik} + \eta_{jk}$.

Case 0. $\bar{\delta} = 0$ or $|E_b| = 0$.

Note that $\bar{\delta} = 0$ implies $\delta_{ij} = 0$ for all $ij \in E_b$, which by (8) implies $\eta_{ij} = 0$ for all $ij \in E_b$. Therefore by (7), we have

$$R(T) \ge R(T^{(0)}) + \sum_{ij\in E_g} \eta_{ij}.$$

If $\sum_{ij\in E_g} \eta_{ij} > 0$, then we have $R(T) > R(T^{(0)})$. Thus we may assume that $\eta_{ij} = 0$ for all $ij \in E_g$. In this case, we will show that $T = T^{(0)}$.

By Lemma 6, if $\varepsilon_0 \le \frac{c_1p^2}{8}$, then $\eta_{ij} = 0$ for all $ij \in E(G)$ implies that $\eta_{ij} = 0$ for all $ij \in E(K_n)$. For $ij \in E_b$, since $\delta_{ij} = \eta_{ij} = 0$, it follows that $\ell_{ij} = \ell_{ij}^{(0)}$. Since $\delta_{ij} = \ell_{ij} - \ell_{ij}^{(0)}$ for $ij \in E_g$, we have

$$0 = \sum_{ij\in E(G)} (\ell_{ij} - \ell_{ij}^{(0)}) = \sum_{ij\in E_b} (\ell_{ij} - \ell_{ij}^{(0)}) + \sum_{ij\in E_g} (\ell_{ij} - \ell_{ij}^{(0)}) = \sum_{ij\in E_g} (\ell_{ij} - \ell_{ij}^{(0)}) = \sum_{ij\in E_g} \delta_{ij}\|t_{ij}^{(0)}\|_2,$$

where the first equality is because $L(T) = L(T^{(0)}) = 1$. By Condition 0, we have $\|t_{ij}^{(0)}\| \ne 0$ for all $ij \in E_g$. Hence if $\delta_{ij} \ne 0$ for some $ij \in E_g$, then there exists $ab, cd \in E_g$ such that $\delta_{ab} > 0$ and $\delta_{cd} < 0$. By Lemma 2 or 3 and Condition 6, this forces $\eta(ab,cd) > 0$, contradicting the fact that $\eta_{ij} = 0$ for all $ij \in E(K_n)$. Therefore $\delta_{ij} = 0$ for all $ij \in E_g$, and hence $\delta_{ij} = 0$ for all $ij \in E(G)$.

Define $t_i = t_i^{(0)} + h_i$ for each $i \in [n]$. Because $\eta_{ij} = \delta_{ij} = 0$ for all $ij \in E(G)$, we have $h_i = h_j$ for all $ij \in E(G)$. Since G is connected (by Condition 1), this implies $h_i = h_j$ for all $i \in [n]$. Then by the constraint $\sum_{i\in[n]} t_i = \sum_{i\in[n]} t_i^{(0)} = 0$, we get $h_i = 0$ for all $i \in [n]$. Therefore $T = T^{(0)}$. This proves Case 0.

We may now assume that $\overline{\delta} \ne 0$. Since $\ell_{ij} - \ell_{ij}^{(0)} = \delta_{ij}\|t_{ij}^{(0)}\|_2$ for $ij \in E_g$, we have

$$0 = \sum_{ij\in E(G)} (\ell_{ij} - \ell_{ij}^{(0)}) \;=\; \sum_{ij\in E_b} (\ell_{ij} - \ell_{ij}^{(0)}) + \sum_{ij\in E_g} \delta_{ij}\|t_{ij}^{(0)}\|_2.$$

Therefore

$$\left|\sum_{ij\in E_g}\delta_{ij}\|t_{ij}^{(0)}\|_2\right| \le \left|\sum_{ij\in E_b}(\ell_{ij}-\ell_{ij}^{(0)})\right| \le \sum_{ij\in E_b}(|\delta_{ij}|\|t_{ij}^{(0)}\|_2+\eta_{ij}) \le 2\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2. \tag{12}$$

where the final inequality follows from (8).

Define $E_g' = \{ij \in E_g : \|t_{ij}^{(0)}\|_2 \ge \frac{1}{2}\mu\}$ as the set of 'long' good edges. Since $\sum_{ij\in E_g\setminus E_g'}\|t_{ij}^{(0)}\|_2 < \frac{1}{2}\mu|E_g|$, we have

$$\sum_{ij\in E_g'}\|t_{ij}^{(0)}\|_2 = \sum_{ij\in E(G)}\|t_{ij}^{(0)}\|_2 - \sum_{ij\in E_b}\|t_{ij}^{(0)}\|_2 - \sum_{ij\in E_g\setminus E_g'}\|t_{ij}^{(0)}\|_2 > \mu|E(G)| - c_0\mu\cdot|E_b| - \frac{1}{2}\mu|E_g| \ge \mu\cdot\frac{1}{16}n^2p.$$

where the last inequality uses $|E(G)| \ge \frac{n^2p}{4}$, $|E_b| < \varepsilon_0 n^2$, $|E_g| \le |E(G)|$, $\varepsilon_0 < \frac{p}{16c_0}$. By Condition 3, we have $\|t_{ij}^{(0)}\|_2 \le c_0\mu$ for all ij, and thus it follows that

$$|E_g'| \ge \frac{1}{16c_0}n^2p. \tag{13}$$

Case 1. $\overline{\delta} \neq 0$ and $\sum_{ij\in E_g'}|\delta_{ij}| < \frac{1}{8}\overline{\delta}|E_g'|$ and $|E_b| \neq 0$.

In this case, we will exploit the fact that there is a difference between average relative parallel motions on long good edges and that on bad edges, to show that there is a large amount of rotational motion on the K_4s of the form $\{i,j,k,\ell\}$ where $ij \in E_b$ and $k\ell \in E_g'$. Define $L_b = \{ij \in E_b : |\delta_{ij}| \ge \frac{1}{2}\overline{\delta}\}$. Note that $\sum_{ij\in E_b\setminus L_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2 < \frac{1}{2}\overline{\delta}\sum_{ij\in E_b}\|t_{ij}^{(0)}\|_2 = \frac{1}{2}\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2$. Therefore

$$\sum_{ij\in L_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2 = \sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2 - \sum_{ij\in E_b\setminus L_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2 > \frac{1}{2}\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2. \tag{14}$$

Define $F_g = \{ij \in E_g : |\delta_{ij}| < \frac{1}{4}\overline{\delta}\}$. Then by the condition of Case 1,

$$\frac{1}{8}\overline{\delta}|E_g'| > \sum_{ij\in E_g'}|\delta_{ij}| \ge \sum_{ij\in E_g'\setminus F_g}|\delta_{ij}| \ge \frac{1}{4}\overline{\delta}|E_g'\setminus F_g|,$$

and therefore $|E_g'\setminus F_g| < \frac{1}{2}|E_g'|$, or equivalently, $|F_g| > \frac{1}{2}|E_g'| \ge \frac{1}{32c_0}n^2p$ (where the second inequality comes from (13)).

For each $ij \in L_b$ and $k\ell \in F_g\setminus B(ij)$, by Lemmas 2 and 3, we have $\eta(ij,k\ell) \ge \frac{\beta}{4}|\delta_{k\ell}-\delta_{ij}|\|t_{ij}^{(0)}\|_2 \ge \frac{\beta}{4}\cdot\frac{1}{2}|\delta_{ij}|\|t_{ij}^{(0)}\|_2$. Therefore,

$$\sum_{ij\in E_b}\sum_{k\ell\in E_g}\eta(ij,k\ell) \ge \sum_{ij\in L_b}\sum_{k\ell\in F_g\setminus B(ij)}\frac{\beta}{8}|\delta_{ij}|\|t_{ij}^{(0)}\|_2 = \sum_{ij\in L_b}|F_g\setminus B(ij)|\cdot\frac{\beta}{8}|\delta_{ij}|\|t_{ij}^{(0)}\|_2.$$

By Lemma 14, we know that $|B(ij)| < 6\varepsilon_1 n^2$ holds for all $ij \in E(K_n)$. For $\varepsilon_1 \le \frac{p}{192c_0}$, we have

$$|F_g\setminus B(ij)| > \frac{1}{32c_0}n^2p - 6\varepsilon_1 n^2 \ge \frac{1}{64c_0}n^2p.$$

Therefore

$$\sum_{ij\in E_b}\sum_{k\ell\in E_g}\eta(ij,k\ell) > \frac{\beta}{8}\cdot\frac{1}{64c_0}n^2p\cdot\sum_{ij\in L_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2 \geq \frac{\beta}{1024c_0}n^2p\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2,$$

where the second inequality comes from (14).

For each $ij\in E(K_n)$, we would like to count how many times each η_{ij} appear on the left hand side. If $ij\in E_b$, then there are at most $\binom{n}{2}$ K_4s and n K_3s containing ij; hence η_{ij} may appear at most $6\binom{n}{2}+3n=3n^2$ times. If $ij\notin E_b$, then η_{ij} appears when there is a K_4 or a K_3 containing ij and some bad edge. By Condition 4, there are at most $2\varepsilon_0 n$ such bad K_3s. If the bad edge in K_4 is incident to ij, then there are at most $2\varepsilon_0 n\cdot(n-3)$ such K_4s, and if the bad edge is not incident to ij, then there are at most $|E_b|\leq\varepsilon_0 n^2$ such K_4s. Thus η_{ij} may appear at most $3\cdot 2\varepsilon_0 n+6\cdot(2\varepsilon_0 n(n-3)+\varepsilon_0 n^2)\leq 18\varepsilon_0 n^2$ times. Therefore

$$\sum_{ij\in E_b}\sum_{k\ell\in E_g}\eta(ij,k\ell) \quad\leq\quad \sum_{ij\in E_b}3n^2\cdot\eta_{ij}+\sum_{ij\in E(K_n)}18\varepsilon_0 n^2\cdot\eta_{ij}.$$

If $\varepsilon_0\leq\frac{c_1p^2}{8}$, then by Lemma 5, we thus have

$$\sum_{ij\in E_b}\sum_{k\ell\in E_g}\eta(ij,k\ell)\leq\frac{24\varepsilon_0}{c_1p^2}n^2\sum_{ij\in E_g}\eta_{ij}+\sum_{ij\in E(K_n)}18\varepsilon_0 n^2\cdot\eta_{ij}\leq\frac{42\varepsilon_0}{c_1p^2}n^2\sum_{ij\in E(K_n)}\eta_{ij}.$$

Hence

$$\frac{42\varepsilon_0}{c_1p^2}n^2\sum_{ij\in E(K_n)}\eta_{ij}\geq\sum_{ij\in E_b}\sum_{k\ell\in E_g}\eta(ij,k\ell)>\frac{\beta}{1024c_0}n^2p\cdot\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2.$$

If $\varepsilon_0\leq\frac{\beta c_1^2p^4}{16\cdot 42\cdot 1024c_0}$, then

$$\sum_{ij\in E(K_n)}\eta_{ij}>\frac{\beta c_1p^3}{42\cdot 1024c_0\varepsilon_0}\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2>\frac{16}{c_1p}\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2.$$

If $\varepsilon_0\leq\frac{c_1p^2}{8}$, then by Lemma 6, this gives

$$\sum_{ij\in E_g}\eta_{ij}\geq\frac{c_1p}{8}\sum_{ij\in E(K_n)}\eta_{ij}>2\sum_{ij\in E_b}|\delta_{ij}|\|t_{ij}^{(0)}\|_2.$$

Since Lemma 5 implies $\sum_{ij\in E_g}\eta_{ij}\geq 2\sum_{ij\in E_b}\eta_{ij}$ (given $\varepsilon_0\leq\frac{c_1p^2}{16}$), together with the inequality above, we have $\sum_{ij\in E_g}\eta_{ij}>\sum_{ij\in E_b}(|\delta_{ij}|\|t_{ij}^{(0)}\|_2+\eta_{ij})$. By (10), this shows that $R(T)>R(T_0)$. The parameters must satisfy $\varepsilon_0\leq\min\{\frac{c_1p^2}{8},\frac{\beta c_1^2p^4}{16\cdot 42\cdot 1024c_0},\frac{c_1p^2}{16}\}$ and $\varepsilon_1\leq\frac{p}{192c_0}$.

Case 2. $\overline{\delta}\neq 0$ and $\sum_{ij\in E_g'}|\delta_{ij}|\geq\frac{1}{8}\overline{\delta}|E_g'|$ and $|E_b|\neq 0$.

In this case, we first show that there are large amount of positive and negative parallel motions on the good edges. This will imply that there is a large amount of rotational motions on the K_4s of the form $\{i,j,k,\ell\}$ where $ij,k\ell\in E_g$ and $\delta_{ij}\geq 0$, $\delta_{k\ell}<0$. Since $\|t_{ij}^{(0)}\|_2\geq\frac{1}{2}\mu$ for all $ij\in E_g$, Case 2 implies

$$\sum_{ij\in E_g}|\delta_{ij}|\|t_{ij}^{(0)}\|_2\geq\sum_{ij\in E_g'}|\delta_{ij}|\|t_{ij}^{(0)}\|_2\geq\frac{1}{2}\mu\cdot\sum_{ij\in E_g'}|\delta_{ij}|\geq\frac{1}{16}\mu\overline{\delta}|E_g'|.$$

Define $E_+ = \{ij \in E_g : \delta_{ij} \geq 0\}$ and $E_- = \{ij \in E_g : \delta_{ij} < 0\}$. The inequality above and (12) implies

$$\sum_{ij \in E_+} \delta_{ij} \|t_{ij}^{(0)}\|_2 = \frac{1}{2} \sum_{ij \in E_g} (|\delta_{ij}| + \delta_{ij}) \|t_{ij}^{(0)}\|$$

$$\geq \frac{1}{2} \left(\frac{1}{16} \mu \bar{\delta} |E'_g| - 2 \sum_{ij \in E_b} |\delta_{ij}| \|t_{ij}^{(0)}\|_2 \right) \geq \frac{1}{2} \left(\frac{1}{16} \mu \bar{\delta} |E'_g| - 2c_0 \mu \bar{\delta} |E_b| \right).$$

From (13), we have $|E_{g'}| \geq \frac{1}{16c_0} n^2 p$. Therefore if $\varepsilon_0 \leq \frac{p}{1024c_0^2}$, then

$$\sum_{ij \in E_+} \delta_{ij} \|t_{ij}^{(0)}\|_2 \geq \frac{1}{32} \mu \bar{\delta} \left(\frac{1}{16c_0} n^2 p - 32 c_0 \varepsilon_0 n^2 \right) \geq \frac{1}{1024 c_0} \mu \bar{\delta} n^2 p.$$

Similarly $\sum_{ij \in E_-} (-\delta_{ij}) \|t_{ij}^{(0)}\|_2 \geq \frac{1}{1024c_0} \mu \bar{\delta} n^2 p$.

We either have $|E_+| \geq \frac{1}{2}|E_g|$ or $|E_-| > \frac{1}{2}|E_g|$. If the former holds, then by Lemmas 2 and 3,

$$\begin{aligned} \sum_{ij \in E_-} \sum_{k\ell \in E_+} \eta(ij, k\ell) &\geq \sum_{ij \in E_-} \sum_{k\ell \in E_+ \setminus B(ij)} \frac{\beta}{4} (-\delta_{ij}) \|t_{ij}^{(0)}\|_2 \\ &\geq \frac{\beta}{4} \cdot \sum_{ij \in E_-} (-\delta_{ij}) \|t_{ij}^{(0)}\|_2 (|E_+| - |B(ij)|). \end{aligned}$$

By Lemma 14, we have $|B(ij)| \leq 6\varepsilon_1 n^2$, and thus $|E_+| - |B(ij)| \geq \frac{1}{2}|E_g| - 6\varepsilon_1 n^2 \geq \frac{1}{2}(\frac{1}{4}n^2 p - \varepsilon_0 n^2) - 6\varepsilon_1 n^2$. If $\varepsilon_0 < \frac{1}{16}p$ and $\varepsilon_1 \leq \frac{1}{192}p$, then $|E_+| - |B(ij)| \geq \frac{1}{16}n^2 p$, and the above gives

$$\sum_{ij \in E_-} \sum_{k\ell \in E_+} \eta(ij, k\ell) \geq \frac{\beta}{4} \cdot \frac{1}{16} n^2 p \cdot \sum_{ij \in E_-} (-\delta_{ij}) \|t_{ij}^{(0)}\|_2 \geq \frac{\beta}{64} \cdot \frac{1}{1024 c_0} \mu \bar{\delta} n^4 p^2.$$

Similarly, if $|E_-| > \frac{1}{2}|E_g|$, then $\sum_{ij \in E_+} \sum_{k\ell \in E_-} \eta(ij, k\ell) \geq \frac{\beta}{64 \cdot 1024 c_0} \mu \bar{\delta} n^4 p^2$.

On the other hand since each edge is contained in at most $\frac{n(n-1)}{2}$ copies of K_4 and n copies of K_3 (and there are 6 edges in a K_4), we have

$$\sum_{ij \in E_-} \sum_{k\ell \in E_+} \eta(ij, k\ell) \leq \left(6 \frac{n(n-1)}{2} + 3n \right) \sum_{ij \in E(K_n)} \eta_{ij} \leq 3n^2 \sum_{ij \in E(K_n)} \eta_{ij}.$$

If $\varepsilon_0 \leq \frac{\beta c_1 p^3}{32 \cdot 3 \cdot 64 \cdot 1024 \cdot c_0^2}$, then

$$\begin{aligned} \sum_{ij \in E(K_n)} \eta_{ij} &\geq \frac{1}{3n^2} \cdot \frac{\beta}{64 \cdot 1024 c_0} \mu \bar{\delta} n^4 p^2 = \frac{\beta p^2}{3 \cdot 64 \cdot 1024 c_0} \mu \bar{\delta} n^2 \\ &> \frac{32}{c_1 p} c_0 \mu \bar{\delta} |E_b| \geq \frac{32}{c_1 p} \sum_{ij \in E_b} |\delta_{ij}| \|t_{ij}^{(0)}\|_2, \end{aligned}$$

where the last inequality follows from Condition 3. If $\varepsilon_0 \leq \frac{c_1 p^2}{8}$, then by Lemma 6, this implies

$$\sum_{ij \in E_g} \eta_{ij} \geq \frac{c_1 p}{16} \sum_{ij \in E(K_n)} \eta_{ij} > 2 \sum_{ij \in E_b} |\delta_{ij}| \|t_{ij}^{(0)}\|_2,$$

Therefore from (10) and (11),

$$\begin{aligned} R(T) &\geq R(T_0) + \sum_{ij \in E_g} \eta_{ij} - \sum_{ij \in E_b} (|\delta_{ij}| \|t_{ij}^{(0)}\|_2 + \eta_{ij}) \\ &> R(T_0) + 2 \sum_{ij \in E_b} |\delta_{ij}| \|t_{ij}^{(0)}\|_2 - \sum_{ij \in E_b} 2|\delta_{ij}| \|t_{ij}^{(0)}\|_2 = R(T_0). \end{aligned}$$

The parameters must satisfy $\varepsilon_0 \leq \min\{\frac{p}{1024c_0^2}, \frac{1}{16}p, \frac{\beta c_1 p^3}{32\cdot 3\cdot 64\cdot 1024\cdot c_0^2}, \frac{c_1 p^2}{8}\}$ and $\varepsilon_1 \leq \frac{1}{192}p$. ☐

3.3 Properties of Gaussians in three dimensions

The first lemma establishes a bound on the average distance between random Gaussian vectors.

Lemma 15. *There exists a positive constant c such that if G is a p-typical graph with vertex set $[n]$, then with probability at least $1 - 3ne^{-cnp/2}$,*

$$\sum_{ij \in E(G)} \|t_i - t_j\|_2 \geq \frac{1}{8} n^2 p.$$

Proof. Let $v \in \mathbb{R}^3$ be a fixed vector. Note that for all $j \in [n]$, we have

$$\|v - t_j\|_2^2 = \|v\|^2 + \|t_j\|^2 - 2\langle v, t_j \rangle.$$

Therefore if $\langle v, t_j \rangle \leq 0$, then $\|v - t_j\|_2 \geq \|t_j\|_2$. Further, by the symmetry of Gaussian random variables, we know that the distribution of $\|t_j\|_2$ remains the same even after conditioning on the event $\langle v, t_j \rangle \leq 0$. Therefore

$$\mathbb{E}[\|v - t_j\|_2] \geq \mathbb{P}\Big(\langle v, t_j \rangle \leq 0\Big) \cdot \mathbb{E}\Big[\|t_j\|_2 \,\Big|\, \langle v, t_j \rangle \leq 0\Big] = \frac{1}{2}\mathbb{E}[\|t_j\|_2] = \sqrt{\frac{2}{\pi}},$$

where the final equality holds since each $\|t_j\|_2$ is subgaussian with mean $\sqrt{8/\pi}$. Fix an index $i \in [n]$ and let N_i be the neighborhood of i in G. Since G is p-typical, we have $|N_i| \geq \frac{1}{2}np$. By the analysis above, we see that

$$\mathbb{E}\left[\sum_{j \in N_i} \|t_i - t_j\|_2\right] \geq |N_i|\sqrt{\frac{2}{\pi}}.$$

By Proposition 5.10 in Vershynin [27] on the concentration of subgaussians, there is a constant c such that with probability at least $1 - e^{1-c|N_i|}$, we have $\sum_{j \in N_i} \|t_i - t_j\|_2 \geq \frac{1}{2}|N_i| \geq \frac{1}{4}np$. Therefore by taking the union bound over all indices $i \in [n]$, we see that with probability at least $1 - 3ne^{-cnp/2}$,

$$\sum_{ij \in E(G)} \|t_i - t_j\|_2 \geq \frac{1}{2}\sum_{i \in [n]}\sum_{j \in N_i} \|t_i - t_j\|_2 \geq \frac{n}{2} \cdot \frac{1}{4}np = \frac{1}{8}n^2 p.$$ ☐

The second lemma establishes a bound on the angle between random Gaussian vectors.

Lemma 16. *Let $x, y \in \mathbb{R}^3$ be linearly independent vectors. If $t_1, t_2, \cdots, t_n \in \mathbb{R}^3$ are independent random Gaussian vecotrs, then with probability $1 - e^{-\Omega(\beta n)}$, for all but at most βn vectors t_i, we have*

$$1 - \left\langle \frac{t_i - x}{\|t_i - x\|_2}, \frac{y - x}{\|y - x\|_2} \right\rangle^2 \geq \frac{\beta^2}{2(\|t_i\|_2^2 + \|x\|_2^2)}.$$

Proof. Fix an index $i \in [n]$. Note that

$$1 - \left\langle \frac{t_i - x}{\|t_i - x\|_2}, \frac{y - x}{\|y - x\|_2} \right\rangle^2 = \left\| P_{(y-x)^\perp} \frac{t_i - x}{\|t_i - x\|_2} \right\|_2^2 = \frac{\|P_{(y-x)^\perp}(t_i - x)\|_2^2}{\|t_i - x\|_2^2}$$
$$\geq \frac{\|P_{\{x,y\}^\perp} t_i\|_2^2}{\|t_i - x\|_2^2} \geq \frac{\|P_{\{x,y\}^\perp} t_i\|_2^2}{2(\|t_i\|_2^2 + \|x\|_2^2)}. \tag{15}$$

Since t_i is a random Gaussian vector and x, y are linearly independent, the distribution of $\|P_{\{x,y\}^\perp} t_i\|_2$ is that of the absolute value of a standard normal distribution. Therefore $\mathbb{P}(\|P_{\{x,y\}^\perp} t_i\|_2 < \beta) \leq \frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\beta} e^{-x^2/2} dx \leq \sqrt{\frac{2}{\pi}}\beta$. Let $\mathbf{1}_i$ be the indicator random variable of the event that $\|P_{\{x,y\}^\perp} t_i\|_2 < \beta$. We seen above that $\mathbb{E}[\mathbf{1}_i] < \sqrt{\frac{2}{\pi}}\beta$. Further, since $\{t_i\}_{i \in [n]}$ are independent, it follows that $\{\mathbf{1}_i\}_{i \in [n]}$ are independent. Therefore by Chernoff's inequality,

$$\mathbb{P}\left(\sum_{i \in [n]} \mathbf{1}_i - \sqrt{\frac{2}{\pi}}\beta n > \frac{1}{5}\beta n \right) \leq e^{-\Omega(\beta n)}.$$

Hence with probability $1 - e^{-\Omega(\beta n)}$, there are at most $\sqrt{\frac{2}{\pi}}\beta n + \frac{1}{5}\beta n < \beta n$ vectors t_i for which $\|P_{\{x,y\}^\perp} t_i\|_2 < \beta$. The lemma now follows from (15). ☐

The next lemma shows that random Gaussian vectors are well-distributed with respect to a fixed pair of vectors.

Lemma 17. *There exists a positive real number c such that the following holds for all pairs of linearly independent vectors $x, y \in \mathbb{R}^3$. If $t_1, \cdots, t_n \in \mathbb{R}^3$ are independent random Gaussian vectors, then with probability $1 - e^{-\Omega(n)}$, the set of vectors $\{t_1, \cdots, t_n\}$ are $\frac{c}{\max\{1, \|x+y\|_2\}}$-well-distributed with respect to (x, y).*

Proof. Let c be a positive real number to be chosen later. We may rotate the vectors so that $x = (\ell, 0, x_3)$ and $y = (\ell, 0, y_3)$ for some $x_3, y_3, \ell \in \mathbb{R}$ where $\ell \geq 0$. Note that $\|x + y\|_2 \geq 2\ell$. Define $\ell_0 = \max\{1, \ell\}$. It suffices to give an estimate on the probability that

$$\sum_{i=1}^{n} \|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 \geq \frac{cn}{\ell_0}$$

holds for all vectors $h \in (x - y)^\perp = \{(a, b, 0) : a, b \in \mathbb{R}\}$ satisfying $\|h\|_2 = 1$.

Fix a vector $h = (h_1, h_2, 0)$ satisfying $\|h\|_2 = 1$. For $t_i = (t_{i,1}, t_{i,2}, t_{i,3})$, we have $span\{t_i - x, t_i - y\} = span\{(0, 0, 1), x + y - 2t_i\} = span\{(0, 0, 1), (2\ell - 2t_{i,1}, -2t_{i,2}, 0)\}$. Hence $s = (t_{i,2}, \ell - t_{i,1}, 0) \in span\{t_i - x, t_i - y\}^\perp$, and

$$\|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 = \frac{|\langle s, h \rangle|}{\|s\|_2} = \left| \frac{(t_{i,2}, \ell - t_{i,1}, 0) \cdot h}{\sqrt{t_{i,2}^2 + (\ell - t_{i,1})^2}} \right|$$
$$= \frac{|h_1 t_{i,2} + (\ell - t_{i,1}) h_2|}{\sqrt{t_{i,2}^2 + (\ell - t_{i,1})^2}}.$$

Assume that $h_1 \geq h_2 \geq 0$, which implies $h_1 \geq \frac{1}{\sqrt{2}}$. Since $t_{i,1}$ is normally distributed with variance 1, the probability that $-1 \leq t_{i,1} \leq 0$ is p for some fixed postive real number p. Conditioned on this event and the event that $t_{i,2} \geq 0$ (note that $t_{i,1}$ and $t_{i,2}$ are independent), we have

$$\|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 = \frac{|h_1 t_{i,2} + (\ell - t_{i,1})h_2|}{\sqrt{t_{i,2}^2 + (\ell - t_{i,1})^2}} \geq \frac{h_1 t_{i,2}}{\sqrt{t_{i,2}^2 + 4\ell_0^2}}.$$

Therefore

$$\mathbb{P}\left(\|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 > \frac{1}{2\ell_0}\right) \geq \mathbb{P}\left(\frac{h_1 t_{i,2}}{\sqrt{t_{i,2}^2 + 4\ell_0^2}} > \frac{1}{2\ell_0} \,\Big|\, t_{i,2} \geq 0\right) \cdot \frac{1}{2}p.$$

Note that for $t_{i,2} \geq 0$, the inequality $\frac{h_1 t_{i,2}}{\sqrt{t_{i,2}^2 + 4\ell_0^2}} > \frac{1}{2\ell_0}$ is equivalent to $h_1^2 t_{i,2}^2 > \frac{t_{i,2}^2}{4\ell_0^2} + 1$, which is equivalent to $t_{i,2}^2(h_1^2 - \frac{1}{4\ell_0^2}) > 1$. Since $h_1^2 \geq \frac{1}{2}$ and $\ell_0 \geq 1$, we have

$$\mathbb{P}\left(\|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 > \frac{1}{2\ell_0}\right) \geq \mathbb{P}(t_{i,2}^2 > 4 \,|\, t_{i,2} \geq 0) \cdot \frac{1}{2}p = q$$

for some fixed positive real number q. By considering the indicator random variable of the events $\|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 > \frac{1}{2\ell_0}$, we see by Chernoff's inequality that with probability $1 - e^{-\Omega(n)}$, there are at least $\frac{qn}{2}$ indices $i \in [n]$ such that $\|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 > \frac{1}{2\ell_0}$. Note that this implies

$$\sum_{i=1}^{n} \|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 \geq \frac{qn}{2} \cdot \frac{1}{2\ell_0} = \frac{qn}{4\ell_0}.$$

To handle the case of $h_2 \geq h_1 \geq 0$, note that if $t_{i,1} \leq 0$ and $0 \leq t_{i,2} \leq 1$, then

$$\|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 = \frac{|h_1 t_{i,2} + (\ell - t_{i,1})h_2|}{\sqrt{t_{i,2}^2 + (\ell - t_{i,1})^2}} \geq \frac{(\ell - t_{i,1})h_2}{\sqrt{1 + (\ell - t_{i,1})^2}} \geq \frac{1}{\sqrt{2}} \cdot \frac{\ell - t_{i,1}}{\sqrt{1 + (\ell - t_{i,1})^2}}.$$

Since $\frac{x}{\sqrt{1+x^2}}$ is decreasing in the range $x \geq 0$, if $t_{i,1} \leq -1$, then $\|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 \geq \frac{1}{2}$. Therefore we see as in above that with probability $1 - e^{-\Omega(n)}$, there are at least $\frac{qn}{2}$ indices $i \in [n]$ such that $\|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 > \frac{1}{2} \geq \frac{1}{2\ell_0}$. All the remaining cases can be handled analogously.

Let H be a set of $\lceil 2\pi \cdot \frac{8\ell_0}{q} \rceil \leq \frac{60\ell_0}{q}$ vectors uniformly distributed along the circle $S_2 = \{(x, y, 0) : x^2 + y^2 = 1\}$. Apply the analysis above to each vector in H and take the union bound to conclude that with probability $1 - \ell_0 e^{-\Omega(n)}$, for all $h \in H$,

$$\sum_{i=1}^{n} \|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 \geq \frac{q}{4} \frac{n}{\ell_0}.$$

Let $h' \in S_2$ be an arbitrary vector and let $h \in H$ be the vector closest to h'. The distance from h to h' along the circle S_2 is at most $2\pi \cdot \frac{1}{|H|} \leq \frac{q}{8\ell_0}$, and hence $\|h - h'\|_2 \leq \frac{q}{8\ell_0}$. Thus for all i, we have

$$\begin{aligned} \|P_{span\{t_i - x, t_i - y\}^\perp}(h')\|_2 &\geq \|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 - \|h - h'\|_2 \\ &\geq \|P_{span\{t_i - x, t_i - y\}^\perp}(h)\|_2 - \frac{q}{8\ell_0} \end{aligned}$$

Therefore

$$\sum_{i=1}^{n} \|P_{span\{t_i-x,t_i-y\}^\perp}(h')\|_2 \geq \frac{q}{4}\frac{n}{\ell_0} - \frac{q}{8}\frac{n}{\ell_0} \geq \frac{q}{8}\frac{n}{\ell_0}.$$

Since $\ell_0 = \max\{1, \ell\} \leq \max\{1, \|x+y\|_2\}$, this implies the lemma for $c = \frac{q}{8}$. □

By applying the union bound together with the three lemmas above, we obtain the following lemma.

Lemma 18. *There exists $c, \zeta \in \mathbb{R}$ and $n_0 \in \mathbb{N}$ such that the following holds for all positive real numbers ε and natural numbers $n \geq n_0$. Let G be a p-typical graph with vertex set $[n]$ for some p satisfying $np^2 \geq \zeta \log n$. If $t_1, \cdots, t_n \in \mathbb{R}^3$ are independent random Gaussian vectors, then the following holds with probability $1 - n^{-5}$,*

1. *For each distinct $i, j \in [n]$, for all but at most εn indices $k \in [n]$, we have $1 - \langle \frac{t_k - t_i}{\|t_k - t_i\|_2}, \frac{t_j - t_i}{\|t_j - t_i\|_2} \rangle^2 \geq \frac{\varepsilon^2}{64 \log n}$,*
2. *for all distinct $i, j \in [n]$, we have $\|t_i - t_j\|_2 \leq 40\sqrt{\log n} \cdot \mu$, where $\mu = \frac{1}{|E(G)|}\sum_{ij \in E(G)} \|t_i - t_j\|_2$, and*
3. *the set $\{t_i\}_{i \in [n]}$ is $\frac{c}{\sqrt{\log n}}$-well-distributed along G.*

Proof. Let c be eight times the constant coming from Lemma 17. For each distinct $i, j \in [n]$, define $S_{ij} = \{t_k : ik, jk \in E(G)\}$. Consider the following events:

(i) for all $i \in [n]$, we have $\|t_i\|_2 \leq 4\sqrt{\log n}$,

(ii) $\sum_{ij \in E(G)} \|t_i - t_j\|_2 \geq \frac{1}{8} n^2 p$,

(iii) for each distinct $i, j \in [n]$, for all but at most εn integers $k \in [n]$, we have $1 - \langle \frac{t_k - t_i}{\|t_k - t_i\|_2}, \frac{t_j - t_i}{\|t_j - t_i\|_2} \rangle^2 \geq \frac{\varepsilon^2}{2(\|t_k\|_2^2 + \|t_i\|_2^2)}$,

(iv) for each distinct $i, j \in [n]$, S_{ij} is $\frac{8c}{\max\{1, \|t_i + t_j\|_2\}}$-well-distributed with respect to (t_i, t_j).

For a fixed $i \in [n]$, since $\|t_i\|_2^2$ follows a χ^2 distribution with 3 degrees of freedom, standard estimates on Chi-squared random variables, such as Lemma 1 in [15], give

$$\mathbb{P}(\|t_i\|_2^2 \geq 3 + 2\sqrt{3}t + 2t^2) \leq e^{-t^2}.$$

Let $t = \sqrt{7 \log n}$. If n is sufficiently large, then $2t^2 + 2\sqrt{3}t + 3 < 16 \log n$. As a result, $\mathbb{P}(\|t_i\|_2^2 \geq 16 \log n) \leq e^{-7 \log n}$. Hence Property (i) holds with probability $1 - n^{-6}$ by taking the union bound over all $i \in [n]$. Property (ii) holds with probability $1 - e^{-\Omega(n)}$ by Lemma 15. For a fixed pair $i, j \in [n]$, Property (iii) holds with probability $1 - e^{-\Omega(\varepsilon n)}$ by Lemma 16. Hence by taking the union bound, we see that Property (iii) holds with probability $1 - n^2 e^{-\Omega(\varepsilon n)}$. For a fixed pair $i, j \in [n]$, by Lemma 17 and the fact that each pair is contained in at least $\frac{1}{2}np^2$ triangles, we have Property (iv) for the pair i, j with probability $1 - e^{-\Omega(np^2)}$. Hence by taking the union bound, we see that Property (iv) holds with probability $1 - n^2 e^{-\Omega(np^2)}$. Thus we see that all four events (i)-(iv) simultaneously hold with at least probability $1 - n^{-5}$ for sufficiently large n, provided that $np^2 \geq \zeta \log n$ for sufficiently large ζ.

We now show that Properties (i)-(iv) imply Properties 1-3. Note that Properties 1 and 3 immediately follow from Properties (i), (iii), and (iv). Further, since $|E(G)| \leq n^2 p$, Property (ii) implies

$$\mu = \frac{1}{|E(G)|} \sum_{ij \in E(G)} \|t_i - t_j\|_2 \geq \frac{1}{n^2 p} \cdot \frac{1}{8} n^2 p = \frac{1}{8}.$$

Hence by Property (i), we have for all $i, j \in [n]$,

$$\|t_i - t_j\|_2 \le \|t_i\|_2 + \|t_j\|_2 \le 8\sqrt{\log n} \le 64\mu\sqrt{\log n}. \qquad \square$$

3.4 Proof of Theorem 2

We can now prove the three-dimensional recovery theorem, which we state here again for convenience:

Theorem 2. *There exists $n_0 \in \mathbb{N}$ and $c \in \mathbb{R}$ such that the following holds for all $n \ge n_0$. Let $G([n], E)$ be drawn from $G(n,p)$ for some $p = \Omega(n^{-1/5}\log^{3/5} n)$. Take $t_1^{(0)}, \dots t_n^{(0)} \in \mathbb{R}^3$, where $t_i^{(0)} \sim \mathcal{N}(0, I_{3\times 3})$ are i.i.d., independent from G. There exists $\gamma = \Omega(p^5/\log^3 n)$ and an event of probability at least $1 - \frac{1}{n^4}$ on which the following holds:*

For arbitrary subgraphs E_b satisfying $\max_i \deg_b(i) \le \gamma n$ and arbitrary pairwise direction corruptions $v_{ij} \in \mathbb{S}^2$ for $ij \in E_b$, the convex program (3) has a unique minimizer equal to $\left\{\alpha\left(t_i^{(0)} - \bar{t}^{(0)}\right)\right\}_{i\in[n]}$ for some positive α and for $\bar{t}^{(0)} = \frac{1}{n}\sum_{i\in[n]} t_i^{(0)}$.

Proof. Let n_0 be a sufficiently large natural number larger than that coming from Lemma 18. Lemma 12 implies G is p-typical with probability $1 - n^2 e^{-\Omega(np^2)}$. Condition on G being p-typical. Let c be the constant from Lemma 18. By applying Lemma 18 with $\varepsilon = \frac{p}{2^{15}\sqrt{\log n}}$, with probability at least $1 - n^{-5}$, we have

1. For each distinct $i, j \in [n]$ satisfying $i < j$, for all but at most $2\varepsilon n = \frac{p}{2^{14}\sqrt{\log n}}$ integers $k \in [n]$, we have $1 - \langle \frac{t_k - t_i}{\|t_k - t_i\|}, \frac{t_j - t_i}{\|t_j - t_i\|}\rangle^2 \ge \frac{p^2}{2^{36}\log^2 n}$ and $1 - \langle \frac{t_k - t_j}{\|t_k - t_j\|}, \frac{t_i - t_j}{\|t_i - t_j\|}\rangle^2 \ge \frac{p^2}{2^{36}\log^2 n}$,
2. for all distinct $i, j \in [n]$, we have $\|t_i - t_j\| \le 64\sqrt{\log n}\cdot\mu$, where $\mu = \frac{1}{|E(G)|}\sum_{ij\in E(G)} \|t_i - t_j\|$, and
3. the set $\{t_i\}_{i\in[n]}$ is $\frac{c}{\sqrt{\log n}}$-well-distributed along G.

Thus the probability that G is p-typical and Properties 1-3 listed above holds is at least $1 - n^{-4}$. Hence we may apply Theorem 4 with with $c_0 = 64\sqrt{\log n}$, $\varepsilon_1 = \frac{p}{2^{14}\sqrt{\log n}} \le \frac{p}{192 c_0}$, $\beta = \sqrt{\frac{p^2}{2^{36}\log^2 n}} = \frac{p}{2^{18}\log n}$, and $c_1 = \frac{c}{\sqrt{\log n}}$. The theorem holds if

$$\varepsilon_0 \le \frac{c^2 p^5}{2^{53}\log^3 n} \le \frac{p}{2^{18}\log n}\cdot\frac{c^2}{\log n}p^4 \cdot \frac{1}{32\cdot 3\cdot 64\cdot 1024\cdot 64^2 \log n} = \frac{\beta c_1^2 p^4}{32\cdot 3\cdot 64\cdot 1024\ c_0^2}.$$

Letting γ from the theorem statement be ε_0, note that the condition $\max_i \deg_b(i) \le \gamma n$ is nontrivial when $p = \Omega(n^{-1/5}\log^{3/5} n)$. $\square$

4 Numerical simulations

In this section, we use numerical simulation to verify that ShapeFit recovers locations in $\mathbb{R}^3$ in the presence of corrupted pairwise direction measurements. Further, we empirically demonstrate that ShapeFit is robust to noise in the uncorrupted measurements.

Let the graph of observations be an Erdős-Rényi graph $G(n,p)$ for $p = 1/2$. Let $\tilde{t}_i^{(0)} \in \mathbb{R}^3$ be independent $\mathcal{N}(0, I_{3\times 3})$ random variables for $i = 1 \dots n$. Let $t_i^{(0)} = \tilde{t}_i^{(0)} - \frac{1}{n}\sum_j \tilde{t}_j^{(0)}$. For $ij \in E(G)$,

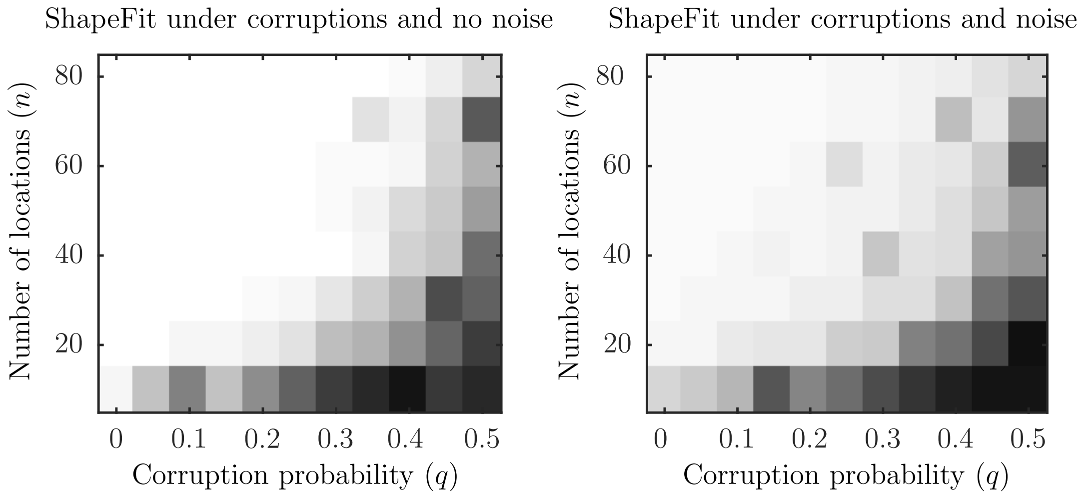

Figure 1: Average recovery error of ShapeFit as a function of the number of locations n and the corruption probability q. The data model has n 3d Gaussian locations whose pairwise directions are observed in accordance with an Erdős-Rényi graph $G(n, 1/2)$ and are corrupted with probability q. White blocks represent an average recovery error of zero over 10 independently generated problems. Black blocks represent an average recovery error of 100%. The left panel corresponds to the noiseless case $\sigma = 0$, and the right panel corresponds to the noisy case $\sigma = 0.05$.

let

$$\tilde{v}_{ij} = \begin{cases} z_{ij} & \text{with probability } q \\ \frac{t_i^{(0)} - t_j^{(0)}}{\|t_i^{(0)} - t_j^{(0)}\|_2} + \sigma z_{ij} & \text{with probability } 1-q \end{cases}$$

where z_{ij} are independent and uniform over $\mathbb{S}^2$. Let $v_{ij} = \tilde{v}_{ij}/\|\tilde{v}_{ij}\|_2$. That is, each observation is corrupted with probability q, and each corruption is in a random direction. In the noiseless case, with $\sigma = 0$, each observation is exact with probability $1-q$.

We solved ShapeFit using the SDPT3 solver [23, 26] and YALMIP [16]. For output $T = \{t_i\}_{i\in[n]}$, define its relative error with respect to $T^{(0)} = \{t_i^{(0)}\}_{i\in[n]}$ as

$$\left\| \frac{T}{\|T\|_F} - \frac{T^{(0)}}{\|T^{(0)}\|_F} \right\|_F$$

where $\|T\|_F$ is the Frobenius norm of the matrix whose column are $\{t_i\}$. This error metric amounts to an ℓ_2 norm after rescaling.

Figure 1 shows the average residual of the output of ShapeFit over 10 independent trials for locations in $\mathbb{R}^3$ generated by $p = 1/2$, $\sigma \in \{0, 0.05\}$, and a range of values $10 \leq n \leq 80$ and $0 \leq q \leq 0.5$. White blocks represent zero average residual, and black blocks represent an average residual of 1 or higher. Average residuals between 0 and 1 are represented by the appropriate shade of gray. The figure shows that ShapeFit successfully recovers 3d locations in the presence of a surprisingly large probability of corruption, provided n is big enough. For example, if $n \geq 50$, recovery succeeds even when around 25% of all measurements are randomly corrupted. Further, successful recovery occurs both in the noiseless case, and in the noisy case with $\sigma = 0.05$.

Figure 2 shows the average residual over 10 independent trials for locations in $\mathbb{R}^3$ generated by $p = 1/2$, $n = 40$, $q = 0.2$ and a range of values of $10^{-6} \leq \sigma \leq 10^0$. We see that ShapeFit

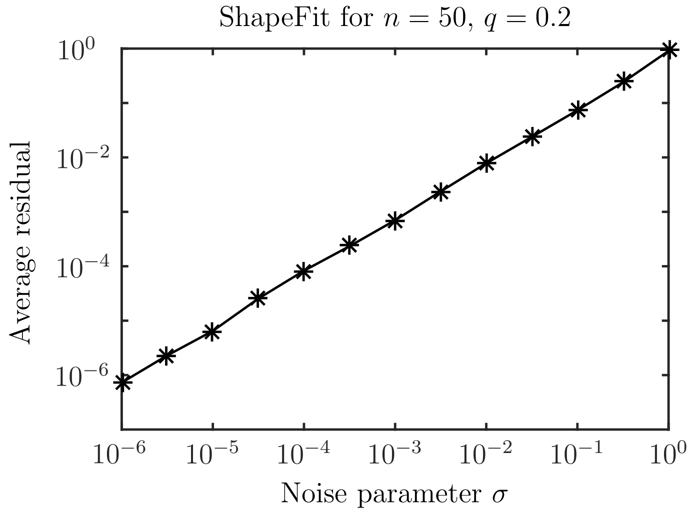

Figure 2: Average recovery error of ShapeFit versus the noise parameter σ. These simulations are based on $n = 50$ Gaussian locations in $\mathbb{R}^3$ whose pairwise directions are observed in accordance with an Erdős-Rényi graph $G(n, 1/2)$ and are corrupted with probability $q = 0.2$. The average is based on 10 independently generated problems.

is empirically stable to noise, with average residuals that are approximately linear in the noise parameter σ.

Acknowledgements

VV acknowledges discussions with Tomasz Malisiewicz, Stefano Soatto, and Ram Sripracha. VV is partially supported by the Office of Naval Research. CL is partially supported by the National Science Foundation Grant DMS-1362326. PH is partially supported by the National Science Foundation Grant DMS-1418971.

References

[1] Dana Angluin and Leslie G. Valiant. Fast probabilistic algorithms for hamiltonian circuits and matchings. In *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing*, STOC '77, pages 30–41, New York, NY, USA, 1977. ACM.

[2] Mica Arie-Nachimson, Shahar Z Kovalsky, Ira Kemelmacher-Shlizerman, Amit Singer, and Ronen Basri. Global motion estimation from point matches. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 81–88. IEEE, 2012.

[3] Matthew Brand, Matthew Antone, and Seth Teller. Spectral solution of large-scale extrinsic camera calibration as a graph embedding problem. In *Computer Vision-ECCV 2004*, pages 262–273. Springer, 2004.

[4] Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 521–528. IEEE, 2013.

[5] David Crandall, Andrew Owens, Noah Snavely, and Dan Huttenlocher. Discrete-continuous optimization for large-scale structure from motion. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3001–3008. IEEE, 2011.

[6] Peter Eades, Xuemin Lin, and William F Smyth. A fast and effective heuristic for the feedback arc set problem. *Information Processing Letters*, 47(6):319–323, 1993.

[7] Olof Enqvist, Fredrik Kahl, and Carl Olsson. Non-sequential structure from motion. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 264–271. IEEE, 2011.

[8] Johan Fredriksson and Carl Olsson. Simultaneous multiple rotation averaging using lagrangian duality. In *Computer Vision–ACCV 2012*, pages 245–258. Springer, 2013.

[9] Venu Madhav Govindu. Combining two-view constraints for motion estimation. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–218. IEEE, 2001.

[10] Venu Madhav Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–684. IEEE, 2004.

[11] Richard Hartley, Khurrum Aftab, and Jochen Trumpf. L1 rotation averaging using the weiszfeld algorithm. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3041–3048. IEEE, 2011.

[12] Nianjuan Jiang, Zhaopeng Cui, and Ping Tan. A global linear method for camera pose registration. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 481–488. IEEE, 2013.

[13] Fredrik Kahl. Multiple view geometry and the l^∞-norm. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1002–1009. IEEE, 2005.

[14] Fredrik Kahl and Richard Hartley. Multiple-view geometry under the l_∞-norm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(9):1603–1617, 2008.

[15] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338, 10 2000.

[16] J. Löfberg. Yalmip : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[17] Daniel Martinec and Tomas Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[18] Pierre Moulon, Pascal Monasse, and Renaud Marlet. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3248–3255. IEEE, 2013.

[19] Onur Özyeşil and Amit Singer. Robust camera location estimation by convex programming. *CoRR*, abs/1412.0165, 2014.

[20] Onur Özyeşil, Amit Singer, and Ronen Basri. Camera motion estimation by convex programming. *CoRR*, abs/1312.5047, 2013.

[21] Kristy Sim and Richard Hartley. Recovering camera motion using l^∞ minimization. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 1230–1237. IEEE, 2006.

[22] Sudipta N Sinha, Drew Steedly, and Richard Szeliski. A multi-stage linear approach to structure from motion. In *Trends and Topics in Computer Vision*, pages 267–281. Springer, 2012.

[23] K. C. Toh, M.J. Todd, and R. H. Tutuncu. Sdpt3 - a matlab software package for semidefinite programming. *Optimization Methods and Software*, 11:545–581, 1998.

[24] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustmenta modern synthesis. In *Vision algorithms: theory and practice*, pages 298–372. Springer, 2000.

[25] Roberto Tron and René Vidal. Distributed image-based 3-d localization of camera sensor networks. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 901–908. IEEE, 2009.

[26] R.H. Tutuncu, K.C. Toh, and M.J. Todd. Solving semidefinite-quadratic-linear programs using sdpt3. *Mathematical Programming Ser. B*, 95:189–217, 2003.

[27] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y.C. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[28] Kyle Wilson and Noah Snavely. Robust global translations with 1dsfm. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014.