GLOBAL REGISTRATION OF MULTIPLE POINT CLOUDS USING SEMIDEFINITE PROGRAMMING

KUNAL N. CHAUDHURY, YUEHAW KHOO, AND AMIT SINGER

ABSTRACT. Consider N points in $\mathbb{R}^d$ and M local coordinate systems that are related through unknown rigid transforms. For each point we are given (possibly noisy) measurements of its local coordinates in some of the coordinate systems. Alternatively, for each coordinate system, we observe the coordinates of a subset of the points. The problem of estimating the global coordinates of the N points (up to a rigid transform) from such measurements comes up in distributed approaches to molecular conformation and sensor network localization, and also in computer vision and graphics.

The least-squares formulation, though non-convex, has a well known closed-form solution for the case $M = 2$ (based on the singular value decomposition). However, no closed form solution is known for $M \geq 3$.

In this paper, we propose a semidefinite relaxation of the least-squares formulation, and prove conditions for exact and stable recovery for both this relaxation and for a previously proposed spectral relaxation. In particular, using results from rigidity theory and the theory of semidefinite programming, we prove that the semidefinite relaxation can guarantee recovery under more adversarial measurements compared to the spectral counterpart.

We perform numerical experiments on simulated data to confirm the theoretical findings. We empirically demonstrate that (a) unlike the spectral relaxation, the relaxation gap is mostly zero for the semidefinite program (i.e., we are able to solve the original non-convex problem) up to a certain noise threshold, and (b) the semidefinite program performs significantly better than spectral and manifold-optimization methods, particularly at large noise levels.

Keywords: Global registration, rigid transforms, rigidity theory, spectral relaxation, spectral gap, convex relaxation, semidefinite program (SDP), exact recovery, noise stability.

1. INTRODUCTION

The problem of point-cloud registration comes up in computer vision and graphics [50, 57, 63], and in distributed approaches to molecular conformation [19, 16] and sensor network localization [15, 9]. The registration problem in question is one of determining the coordinates of a point cloud P from the knowledge of (possibly noisy) coordinates of smaller point cloud subsets (called *patches*) $P_1, \ldots, P_M$ that are derived from P through some general transformation. In certain applications [43, 57, 38], one is often interested in finding the optimal transforms (one for each patch) that consistently align $P_1, \ldots, P_M$. This can be seen as a sub-problem in the determination of the coordinates of P [15, 49].

In this paper, we consider the problem of *rigid registration* in which the points within a given P_i are (ideally) obtained from P through an unknown rigid transform. Moreover, we assume that the correspondence between the local patches and the original point cloud is known, that is, we know beforehand as to which points from P are contained in a given P_i. In fact, one has a control on the correspondence in distributed approaches to molecular conformation [16] and sensor network localization [9, 66, 15]. While this correspondence is not directly available for certain graphics and vision problems, such as multiview registration [47], it is in principle possible to estimate the correspondence by aligning pairs of patches, e.g., using the ICP (Iterative Closest Point) algorithm [6, 49, 33].

1.1. **Two-patch registration.** The particular problem of two-patch registration has been well-studied [20, 31, 2]. In the noiseless setting, we are given two point clouds $\{x_1,\ldots,x_N\}$ and $\{y_1,\ldots,y_N\}$ in $\mathbb{R}^d$, where the latter is obtained through some rigid transform of the former. Namely,

$$y_k = Ox_k + t \qquad (k=1,\ldots,N), \tag{1}$$

where O is some unknown $d\times d$ orthogonal matrix (that satisfies $O^TO = I_d$) and $t\in\mathbb{R}^d$ is some unknown translation.

The problem is to infer O and t from the above equations. To uniquely determine O and t, one must have at least $N \geq d+1$ non-degenerate points[1]. In this case, O can be determined simply by fixing the first equation in (1) and subtracting (to eliminate t) any of the remaining d equations from it. Say, we subtract the next d equations:

$$[y_2 - y_1 \ \cdots \ y_{d+1} - y_1] = O[x_2 - x_1 \ \cdots \ x_{d+1} - x_1].$$

By the non-degeneracy assumption, the matrix on the right of O is invertible, and this gives us O. Plugging O into any of the equations in (1), we get t.

In practical settings, (1) would hold only approximately, say, due to noise or model imperfections. A particular approach then would be to determine the optimal O and t by considering the following least-squares program:

$$\min_{O\in\mathbb{O}(d),\ t\in\mathbb{R}^d} \quad \sum_{k=1}^{N} \|y_k - Ox_k - t\|_2^2. \tag{2}$$

Note that the problem looks difficult a priori since the domain of optimization is $\mathbb{O}(d)\times\mathbb{R}^d$, which is non-convex. Remarkably, the global minimizer of this non-convex problem can be found exactly, and has a simple closed-form expression [18, 36, 29, 20, 31, 2]. More precisely, the optimal $O^\star$ is given by VU^T, where $U\Sigma V^T$ is the singular value decomposition (SVD) of

$$\sum_{k=1}^{N}(x_k - x_c)(y_k - y_c)^T,$$

in which $x_c = (x_1+\cdots+x_N)/N$ and $y_c = (y_1+\cdots+y_N)/N$ are the centroids of the respective point clouds. The optimal translation is $t^\star = y_c - O^\star x_c$.

The fact that two-patch registration has a closed-form solution is used in the so-called incremental (sequential) approaches for registering multiple patches [6]. The most well-known method is the ICP algorithm [49] (note that ICP uses other heuristics and refinements besides registering corresponding points). Roughly, the idea in sequential registration is to register two overlapping patches at a time, and then integrate the estimated pairwise transforms using some means. The integration can be achieved either locally (on a patch-by-patch basis), or using global cycle-based methods such as synchronization [50, 32, 51, 57, 61]. More recently, it was demonstrated that, by locally registering overlapping patches and then integrating the pairwise transforms using synchronization, one can design efficient and robust methods for distributed sensor network localization [15] and molecular conformation [16]. Note that, while the registration phase is local, the synchronization method integrates the local transforms in a globally consistent manner. This makes it robust to error propagation that often plague local integration methods [32, 61].

1.2. **Multi-patch registration.** To describe the multi-patch registration problem, we first introduce some notations. Suppose $x_1, x_2, \ldots, x_N$ are the unknown global coordinates of a point cloud in $\mathbb{R}^d$. The point cloud is divided into patches $P_1, P_2, \ldots, P_M$, where each P_i is a subset of $\{x_1, x_2, \ldots, x_N\}$. The patches are in general overlapping, whereby a given point can belong to multiple patches. We represent this membership using an undirected bipartite graph $\Gamma = (V_x \cup V_P, E)$. The set of vertices $V_x = \{x_1,\ldots,x_N\}$ represents the point cloud, while

[1]By non-degenerate, we mean that the affine span of the points is d dimensional.

$V_P = \{P_1, \ldots, P_M\}$ represents the patches. The edge set $E = E(\Gamma)$ connects V_x and V_P, and is given by the requirement that $(k,i) \in E$ if and only if $x_k \in P_i$. We will henceforth refer to Γ as the *membership graph*.

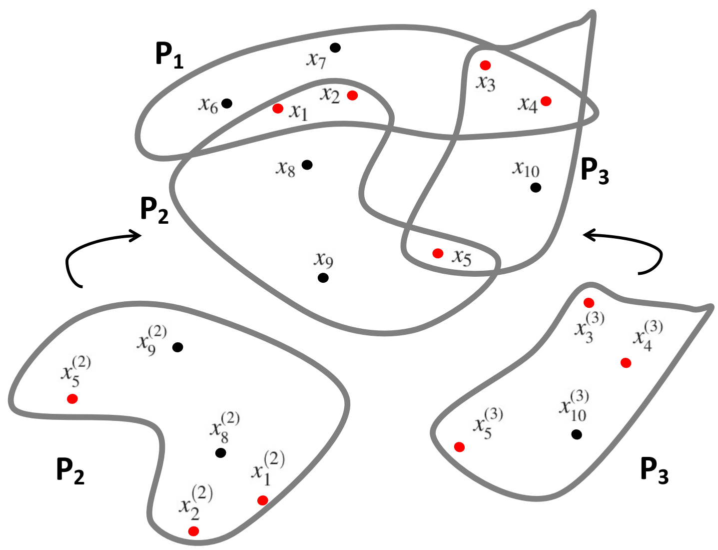

FIGURE 1. The problem of registering 3 patches on $\mathbb{R}^2$. One is required to find the global coordinates of the points from the corresponding local patch coordinates. The local coordinates of the points in patches P_2 and P_3 are shown (see (5) for the notation of local coordinates). It is only the common points (belonging to two or more patches, marked in red) that contribute to the registration. Note that sequential or pairwise registration would fail in this case. This is because no pair of patches can be registered as they have less than 3 points in common (at least 3 points are required to fix rotations, reflections, and translations in $\mathbb{R}^2$). The SDP-based algorithm proposed in this paper does a global registration, and is able to recover the exact global coordinates for this example.

In this paper, we assume that the local coordinates of a given patch can (ideally) be related to the global coordinates through a single rigid transform, that is, through some rotation, reflection, and translation. More precisely, with every patch P_i we associate some (unknown) orthogonal transform O_i and translation t_i. If point x_k belongs to patch P_i, then its representation in P_i is given by (cf. (1) and Figure 1)

$$x_k^{(i)} = O_i^T(x_k - t_i) \qquad (k,i) \in E(\Gamma). \tag{3}$$

Alternatively, if we fix a particular patch P_i, then for every point belonging to that patch,

$$x_k = O_i x_k^{(i)} + t_i \qquad (k,i) \in E(\Gamma). \tag{4}$$

In particular, a given point can belong to multiple patches, and will have a different representation in the coordinate system of each patch.

The premise of this paper is that we are given the membership graph and the local coordinates (referred to as measurements), namely

$$\Gamma \quad \text{and} \quad \{x_k^{(i)}, (k,i) \in E(\Gamma)\}, \tag{5}$$

and the goal is to recover the coordinates $x_1, \ldots, x_N$, and in the process the unknown rigid transforms $(O_1, t_1), \ldots, (O_M, t_M)$, from (5). Note that the global coordinates are determined up to a global rotation, reflection, and translation. We say that two points clouds (also called *configurations*) are *congruent* if one is obtained through a rigid transformation of the other. We will always identify two congruent configurations as being a single configuration.

Under appropriate non-degeneracy assumptions on the measurements, one task would be to specify appropriate conditions on Γ under which the global coordinates can be uniquely determined. Intuitively, it is clear that the patches must have enough points in common for the registration problem to have an unique solution. For example, it is clear that the global coordinates cannot be uniquely recovered if Γ is disconnected.

In practical applications, we are confronted with noisy settings where (4) holds only approximately. In such cases, we would like to determine the global coordinates and the rigid transforms such that the discrepancy in (4) is minimal. In particular, we consider the following quadratic loss:

$$\phi = \sum_{(k,i) \in E(\Gamma)} \|x_k - O_i x_k^{(i)} - t_i\|^2, \tag{6}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$. The optimization problem is to minimize ϕ with respect to the following variables:

$$x_1, x_2, \ldots, x_N \in \mathbb{R}^d, \quad O_1, \ldots, O_M \in \mathbb{O}(d), \quad t_1, \ldots, t_M \in \mathbb{R}^d.$$

The input to the problem are the measurements in (5). Note that our ultimate goal is to determine $x_1, x_2, \ldots, x_N$; the rigid transforms can be seen as latent variables.

The problem of multipatch registration is intrinsically non-convex since one is required to optimize over the non-convex domain of orthogonal transforms. Different ideas from the optimization literature have been deployed to attack this problem, including Lagrangian optimization and projection methods. In the Lagrangian setup, the orthogonality constraints are incorporated into the objective; in the projection method, the constraints are forced after every step of the optimization [47]. Following the observation that the registration problem can be viewed as an optimization on the Grassmanian and Stiefel manifolds, researchers have proposed algorithms using ideas from the theory and practice of manifold optimization [38]. A detailed review of these methods is beyond the scope of this paper, and instead we refer the interested reader to these excellent reviews [17, 1]. Manifold-based methods are, however, local in nature, and are not guaranteed to find the global minimizer. Moreover, it is rather difficult to certify the noise stability of such methods.

1.3. **Contributions.** In Section 2, we demonstrate that the non-convex least-squares formulation can be relaxed into a semidefinite program, and that the global coordinates can be computed from the solution of this program. The registration algorithm based on this relaxation is described in Algorithm 2. The corresponding algorithm for the spectral relaxation, that was already considered in [38], is described in Algorithm 1 for reference.

In Section 3, we prove different conditions on the membership graph Γ for exact recovery using Algorithms 1 and 2. In particular, we present different *admissibility* conditions on Γ under which the spectral and semidefinite programs are guaranteed to have an unique solution. We identify a particular combinatorial structure on Γ, called lateration, that is sufficient to make Γ admissible for both the programs (cf. Propositions 5 and 6, and Theorem 8). We describe here how it is possible to construct patch systems in certain applications that are tailored to satisfy the lateration condition. In Section 4, we demonstrate that the semidefinite relaxation is tighter than its spectral counterpart in that the former is able to guarantee exact recovery under weaker assumptions on Γ. More precisely, in Theorem 12, we prove that the rigidity properties (namely, it unique localizability [54]) of a certain backbone graph derived from Γ are both necessary and sufficient for exact recovery by the semidefinite program. This result also shows us that the global rigidity [24] of the backbone graph is a necessary and sufficient condition for uniqueness. We present a registration example in Remark 13 for which the spectral relaxation fails, but for which the solution of the semidefinite relaxation is unique. The proof of Theorem 12 is based on results from the theory of semidefinite programming and the theory of graph rigidity [54]. In Section 5, we present an efficient randomized test for certifying admissibility.

In Section 6, we study the stability of Algorithms 1 and 2 for the noise model in which the patch coordinates are perturbed using noise of bounded size (note that the stability of the spectral relaxation was not investigated in [38]). Our main result here is Theorem 17 which states that, if Γ is admissible in a certain sense, then the registration error for the semidefinite relaxation is within a constant factor of the noise level. To the best of our knowledge, there is no existing algorithm for multipatch registration that comes with a similar stability guarantee.

In Section 7, we present numerical results on simulated data to numerically verify the exact recovery and noise stability properties of Algorithms 1 and 2. Our main empirical findings are the following:
(a) The semidefinite relaxation performs significantly better than spectral and manifold-based optimization (say, with the spectral solution as initialization) in terms of the reconstruction quality (see first plot in Figure 8).
(b) The relaxation gap is mostly zero for the semidefinite program (we are able to solve the original non-convex problem) up to a certain noise threshold (see second plot in Figure 8).

1.4. **Broader context and related work.** The objective (6) is a straightforward extension of the objective for two-patches [18, 20, 31, 2]. In fact, this objective was earlier considered by Zhang et al. for distributed sensor localization [66]. The present work is also closely tied to the work of Cucuringu et al. on distributed localization [15, 16], where a similar objective is implicitly optimized. The common theme in these works is that some form of optimization is used to globally register the patches, once their local coordinates have been determined by some means. There is however a fundamental difference between the algorithms used to perform the optimization. Zhang et al. [66] use alternating least-squares to iteratively optimize over the global coordinates and the transforms, which to the best of our knowledge has no convergence guarantee. On the other hand, Cucuringu et al. [15, 16] first optimize over the orthogonal transforms (using synchronization [51]), and then solve for the translations (in effect, the global coordinates) using least-squares fitting. In this work, we combine these different ideas into a single framework. While our objective is similar to the one used in [66], we jointly optimize the rigid transforms and positions. In particular, the algorithms considered in Section 2 avoid the convergence issues associated with alternating least-squares in [66], and is able to register patch systems that cannot be registered using the approach in [15, 16].

Another closely related work is the paper by Krishnan et al. on global registration [38], where the optimal transforms (rotations to be specific) are computed by extending the objective in (1) to the multipatch case. The subsequent mathematical formulation has strong resemblance with our formulation, and, in fact, leads to a subproblem that is equivalent to the following:

$$\max_{O_1,\ldots,O_M} \sum_{i,j=1}^{M} \mathrm{Tr}(O_i Q_{ij} O_j^T) \quad \text{subject to} \quad O_i \in \mathbb{O}(d)\ (1 \le i \le M), \tag{7}$$

where the block matrix $Q \in \mathbb{R}^{Md\times Md}$ is positive semidefinite whose (i,j)-th block is Q_{ij} .

Krishnan et al. [38] propose the use of manifold optimization to solve (7), where the manifold is the product manifold of rotations. However, as mentioned earlier, manifold methods generally do not offer guarantees on convergence (to the global minimum) and stability. Moreover, the manifold in (7) is not connected. Therefore, any local method cannot solve (7) if the initial guess is on the wrong component of the manifold. It is exactly at this point that we depart from [38], namely, we propose to relax (7) into a tractable semidefinite program (SDP). This was motivated by a long line of work on the use of SDP relaxations for non-convex (particularly NP-hard) problems. See, for example, [41, 22, 64, 45, 12, 39], and these reviews [58, 46, 67]. Note that for $d = 1$, (7) is a quadratic Boolean optimization, similar to the `MAX-CUT` problem. An SDP-based algorithm with randomized rounding for solving `MAX-CUT` was proposed in the seminal work of Goemans and Williamson [22]. The semidefinite relaxation that we consider in Section 2 is motivated by this work. In connection with the present work, we note that provably stable SDP algorithms have been considered for low rank matrix completion [12], phase retrieval [13, 60], and graph localization [34].

We note that a special case of the registration problem considered here is the so-called generalized Procrustes problem [25]. Within the point-patch framework just introduced, the goal in Procrustes analysis is to find $O_1, \ldots, O_M \in \mathbb{O}(d)$ that minimizes

$$\sum_{k=1}^{N} \sum_{i,j=1}^{M} \|O_i x_k^{(i)} - O_j x_k^{(j)}\|^2. \tag{8}$$

In other words, the goal is to achieve the best possible alignment of the M patches through orthogonal transforms. This can be seen as an instance of the global registration problem without the translations ($t_1 = \cdots = t_M = 0$), and in which Γ is complete. It is not difficult to see that (8) can be reduced to (7). On the other hand, using the analysis in Section 2, it can be shown that (6) is equivalent to (8) in this case. While the Procrustes problem is known to be NP-hard, several polynomial-time approximations with guarantees have been proposed. In particular, SDP relaxations of (8) have been considered in [45, 53, 44], and more recently in [4]. We use the relaxation of (7) considered in [4] for reasons to be made precise in Section 2.

1.5. **Notations.** We use upper case letters such as O to denote matrices, and lower case letters such as t for vectors. We use I_d to denote the identity matrix of size $d \times d$. We denote the diagonal matrix of size $n \times n$ with diagonal elements $c_1, \ldots, c_n$ as $\mathrm{diag}(c_1, \ldots, c_n)$. We will frequently use block matrices built from smaller matrices, typically of size $d \times d$, where d is the dimension of the ambient space. For some block matrix A, we will use A_{ij} to denote its (i, j)-th block, and $A(p, q)$ to denote its (p, q)-th entry. In particular, if each block has size $d \times d$, then

$$A_{ij}(p, q) = A\big((i-1)d + p, (j-1)d + q\big) \qquad (1 \le p, q \le d).$$

We use $A \succeq 0$ to mean that A is positive semidefinite, that is, $u^T A u \ge 0$ for all u. We use $\mathbb{O}(d)$ to denote the group of orthogonal transforms (matrices) acting on $\mathbb{R}^d$, and $\mathbb{O}(d)^M$ to denote the M-fold product of $\mathbb{O}(d)$ with itself. We will also conveniently identify the matrix $[O_1 \cdots O_M]$ with an element of $\mathbb{O}(d)^M$ where each $O_i \in \mathbb{O}(d)$. We use $\|x\|$ to denote the Euclidean norm of $x \in \mathbb{R}^n$ (n will usually be clear from the context, and will be pointed out if this is not so). We denote the trace of a square matrix A by $\mathrm{Tr}(A)$. The Frobenius and spectral norms are defined as

$$\|A\|_F = \mathrm{Tr}(A^T A)^{1/2} \quad \text{and} \quad \|A\|_{\mathrm{sp}} = \max_{\|x\| \le 1} \|Ax\|.$$

The Kronecker product between matrices A and B is denoted by $A \otimes B$ [23]. The all-ones vector is denoted by e (the dimension will be obvious from the context), and e_i^N denotes the all-zero vector of length N with 1 at the i-th position.

2. Spectral and semidefinite relaxations

The minimization of (6) involves unconstrained variables (global coordinates and patch translations) and constrained variables (the orthogonal transformations). We first solve for the unconstrained variables in terms of the unknown orthogonal transformations, representing the former as linear combinations of the latter. This reduces (6) to a quadratic optimization problem over the orthogonal transforms of the form (7).

In particular, we combine the global coordinates and the translations into a single matrix:

$$Z = \begin{bmatrix} x_1 & \cdots & x_N & t_1 & \cdots & t_M \end{bmatrix} \in \mathbb{R}^{d\times(N+M)}. \tag{9}$$

Similarly, we combine the orthogonal transforms into a single matrix,

$$O = [O_1 \ \cdots \ O_M] \in \mathbb{R}^{d\times Md}. \tag{10}$$

Recall that we will conveniently identify O with an element of $\mathbb{O}(d)^M$.

To express (6) in terms of Z and O, we write $x_k - t_i = Ze_{ki}$, where

$$e_{ki} = e_k^{N+M} - e_{N+i}^{N+M}.$$

Similarly, we write $O_i = O(e_i^M \otimes I_d)$. This gives us

$$\phi(Z,O) = \sum_{(k,i)\in E(\Gamma)} \|Ze_{ki} - O(e_i^M \otimes I_d)x_k^{(i)}\|^2.$$

Using $\|x\|^2 = \mathrm{Tr}(xx^T)$, and properties of the trace, we obtain

$$\phi(Z,O) = \mathrm{Tr}\left([Z \ \ O]\begin{bmatrix} L & -B^T \\ -B & D \end{bmatrix}\begin{bmatrix} Z^T \\ O^T \end{bmatrix}\right), \tag{11}$$

where

$$\begin{aligned} L &= \sum_{(k,i)\in E} e_{ki}e_{ki}^T, \quad B = \sum_{(k,i)\in E} (e_i^M \otimes I_d)x_k^{(i)}e_{ki}^T, \text{ and} \\ D &= \sum_{(k,i)\in E} (e_i^M \otimes I_d)x_k^{(i)}{x_k^{(i)}}^T(e_i^M \otimes I_d)^T. \end{aligned} \tag{12}$$

The matrix L is the combinatorial graph Laplacian of Γ [14], and is of size $(N+M)\times(N+M)$. The matrix B is of size $Md\times(N+M)$, and the size of the block diagonal matrix D is $Md\times Md$.

The optimization program now reads

$$\text{(P)} \qquad \min_{Z,O} \ \phi(Z,O) \quad \text{subject to} \quad Z \in \mathbb{R}^{d\times(N+M)}, \ O \in \mathbb{O}(d)^M.$$

The fact that $\mathbb{O}(d)^M$ is non-convex makes (P) non-convex. In the next few Sections, we will show how this non-convex program can be approximated by tractable spectral and convex programs.

2.1. **Optimization over translations.** Note that we can write (P) as

$$\min_{O\in\mathbb{O}(d)^M} \Big[\min_{Z\in\mathbb{R}^{d\times(N+M)}} \ \phi(Z,O)\Big].$$

That is, we first minimize over the free variable Z for some fixed $O \in \mathbb{O}(d)^M$, and then we minimize with respect to O.

Fix some arbitrary $O \in \mathbb{O}(d)^M$, and set $\psi(Z) = \phi(Z,O)$. It is clear from (11) that $\psi(Z)$ is quadratic in Z. In particular, the stationary points $Z^\star = Z^\star(O)$ of $\psi(Z)$ satisfy

$$\nabla\psi(Z^\star) = 0 \quad \Rightarrow \quad Z^\star L = OB. \tag{13}$$

Note that the Hessian of $\psi(Z)$ equals $2L$, and it is clear from (12) that $L \succeq 0$. Therefore, $Z^\star$ is a minimizer of $\psi(Z)$.

If Γ is connected, then e is the only vector in the null space of L [14]. Let $L^\dagger$ be the pseudoinverse of L, which is again positive semidefinite [23]. It can be verified that

$$LL^\dagger = L^\dagger L = I_{N+M} - (N+M)^{-1}ee^T. \tag{14}$$

We right multiply (13) by $L^\dagger$ and use $Be = 0$, to get

$$Z^\star = OBL^\dagger + te^T, \tag{15}$$

where $t \in \mathbb{R}^d$ is some global translation. Conversely, if we right multiply (15) by L and use $e^T L = 0$, we get (13). That is, every solution of (13) is of the form (15).

Substituting (15) into (11), we get

$$\psi(Z^\star) = \phi(Z^\star, O) = \mathrm{Tr}(CO^T O) = \sum_{i,j=1}^{M} \mathrm{Tr}(O_i C_{ij} O_j^T), \tag{16}$$

where

$$C = \begin{bmatrix} BL^\dagger & I_{Md} \end{bmatrix} \begin{bmatrix} L & -B^T \\ -B & D \end{bmatrix} \begin{bmatrix} L^\dagger B^T \\ I_{Md} \end{bmatrix} = D - BL^\dagger B^T. \tag{17}$$

Note that (16) has the global translation t taken out. This is not a surprise since ϕ is invariant to global translations. Moreover, note that we have not forced the orthogonal constraints on O as yet. Since $\phi(Z, O) \geq 0$ for any Z and O, it necessarily follows from (16) that $C \succeq 0$. We will see in the sequel how the spectrum of C dictates the performance of the convex relaxation of (16).

In analogy with the notion of stress in rigidity theory [24], we can consider (6) as a sum of the "stress" between pairs of patches when we try to register them using rigid transforms. In particular, the (i, j)-th term in (16) can be regarded as the stress between the (centered) i-th and j-th patches generated by the orthogonal transforms. Keeping this analogy in mind, we will henceforth refer to C as the *patch-stress matrix*.

2.2. **Optimization over orthogonal transforms.** The goal now is to optimize (16) with respect to the orthogonal transforms, that is, we have reduced (P) to the following problem:

$$(\mathrm{P}_0) \qquad \min_{O \in \mathbb{R}^{d \times Md}} \ \mathrm{Tr}(CO^T O) \quad \text{subject to} \quad (O^T O)_{ii} = I_d \ (1 \leq i \leq M).$$

This is a non-convex problem since O lives on a non-convex (disconnected) manifold [1]. We will generally refer to any method which uses manifold optimization to solve (P_0) and then computes the coordinates using (15) as "Global Registration over Euclidean Transforms using Manifold Optimization" (`GRET-MANOPT`).

2.3. **Spectral relaxation and rounding.** Following the quadratic nature of the objective in (P_0), it is possible to relax it into a spectral problem. More precisely, consider the domain

$$\mathcal{S} = \{O \in \mathbb{R}^{d \times Md} \ : \ \text{rows of } O \text{ are orthogonal and each row has norm } \sqrt{M}\}.$$

That is, we do not require the $d \times d$ blocks in $O \in \mathcal{S}$ to be orthogonal. Instead, we only require the rows of O to form an orthogonal system, and each row to have the same norm. It is clear that $\mathcal{S}$ is a larger domain than that determined by the constraints in (P_0). In particular, we consider the following relaxation of (P_0):

$$(\mathrm{P}_1) \qquad \min_{O \in \mathcal{S}} \mathrm{Tr}(CO^T O).$$

This is precisely a spectral problem in that the global minimizers are determined from the spectral decomposition of C. More precisely, let $\mu_1 \leq \ldots \leq \mu_{Md}$ be eigenvalues of C, and let $r_1, \ldots, r_{Md}$ be the corresponding eigenvectors. Define

$$W^\star \stackrel{\text{def}}{=} \begin{bmatrix} r_1 \cdots r_d \end{bmatrix}^T \in \mathbb{R}^{d \times Md}. \tag{18}$$

Then

$$\mathrm{Tr}(CW^{\star T} W^\star) = \min_{O \in \mathcal{S}} \mathrm{Tr}(CO^T O) = M(\mu_1 + \cdots + \mu_d). \tag{19}$$

Due to the relaxation, the blocks of $W^\star$ are not guaranteed to be in $\mathbb{O}(d)$. We round each $d \times d$ block of $W^\star$ to its "closest" orthogonal matrix. More precisely, let $W^\star = [W_1^\star \cdots W_M^\star]$. For every $1 \leq i \leq M$, we find $O_i^\star \in \mathbb{O}(d)$ such that

$$\|O_i^\star - W_i^\star\|_F = \min_{O \in \mathbb{O}(d)} \|O - W_i^\star\|_F.$$

As noted earlier, this has a closed-form solution, namely $O_i^\star = UV^T$, where $U\Sigma V^T$ is the SVD of $W_i^\star$. We now put the rounded blocks back into place and define

$$O^\star \stackrel{\text{def}}{=} \left[O_1^\star \dots O_M^\star\right] \in \mathbb{O}(d)^M. \tag{20}$$

In the final step, following (15), we define

$$Z^\star \stackrel{\text{def}}{=} O^\star B L^\dagger \in \mathbb{R}^{d\times(N+M)}. \tag{21}$$

The first N columns of $Z^\star$ are taken to be the reconstructed global coordinates.

We will refer to this spectral method as the "Global Registration over Euclidean Transforms using Spectral Relaxation" (`GRET-SPEC`). The main steps of `GRET-SPEC` are summarized in Algorithm 1. We note that a similar spectral algorithm was proposed for angular synchronization by Bandeira et al. [3], and by Krishnan et al. [38] for initializing the manifold optimization.

Algorithm 1 `GRET-SPEC`

Require: Membership graph Γ, local coordinates $\{x_k^{(i)}, (k,i) \in E(\Gamma)\}$, dimension d.
Ensure: Global coordinates $x_1, \dots, x_N$ in $\mathbb{R}^d$.

Build B, L and D in (12) using Γ.
Compute $L^\dagger$ and $C = D - BL^\dagger B^T$.
Compute bottom d eigenvectors of C, and set $W^\star$ as in (18).
for $i = 1$ to M **do**
  if $W_i^\star \in \mathbb{O}(d)$ **then**
    $O_i^\star \leftarrow W_i^\star$.
  else
    Compute $W_i^\star = U_i \Sigma_i V_i^T$.
    $O_i^\star \leftarrow U_i V_i^T$.
  end if
end for
$Z^\star \leftarrow \left[O_1^\star \cdots O_M^\star\right] BL^\dagger$.
Return first N columns of $Z^\star$.

The question at this point is how are the quantities $O^\star$ and $Z^\star$ obtained from `GRET-SPEC` related to the original problem (P)? Since (P_1) is obtained by relaxing the block-orthogonality constraint in (P_0), it is clear that if the blocks of $W^\star$ are orthogonal, then $O^\star$ and $Z^\star$ are solutions of (P), that is,

$$\phi(Z^\star, O^\star) \le \phi(Z, O) \qquad \text{for all } Z \in \mathbb{R}^{d\times(N+M)},\ O \in \mathbb{O}(d)^M.$$

We have actually found the global minimizer of the original non-convex problem (P) in this case.

Observation 1 (Tight relaxation using `GRET-SPEC`)**.** *If the $d \times d$ blocks of the solution of (P_1) are orthogonal, then the coordinates and transforms computed by `GRET-SPEC` are the global minimizers of (P).*

If some the blocks are not orthogonal, the rounded quantities $O^\star$ and $Z^\star$ are only an approximation of the solution of (P).

2.4. **Semidefinite relaxation and rounding.** We now explain how we can obtain a tighter relaxation of (P_0) using a semidefinite program, for which the global minimizer can be computed efficiently. Our semidefinite program was motivated by the line of works on the semidefinite relaxation of non-convex problems [41, 22, 58, 12].

Consider the domain

$$\mathcal{C} = \{O \in \mathbb{R}^{Md\times Md} \ :\ (O^TO)_{11} = \cdots = (O^TO)_{MM} = I_d\}.$$

That is, while we require the columns of each $Md \times d$ block of $O \in \mathcal{C}$ to be orthogonal, we do not force the non-convex rank constraint $\mathrm{rank}(O) = d$. This gives us the following relaxation

$$\min_{O \in \mathcal{C}} \mathrm{Tr}(CO^T O). \tag{22}$$

Introducing the variable $G = O^T O$, (22) is equivalent to

$$(\mathrm{P}_2) \qquad \min_{G \in \mathbb{R}^{Md \times Md}} \mathrm{Tr}(CG) \quad \text{subject to} \quad G \succeq 0,\ G_{ii} = I_d \ (1 \le i \le M).$$

This is a standard semidefinite program [58] which can be solved using software packages such as `SDPT3` [56] and `CVX` [26]. We provide details about SDP solvers and their computational complexity later in Section 2.5.

Let us denote the solution of (P_2) by $G^\star$, that is,

$$\mathrm{Tr}(CG^\star) = \min_{G \in \mathbb{R}^{Md \times Md}} \{\mathrm{Tr}(CG) : G \succeq 0,\ G_{11} = \cdots = G_{MM} = I_d\}. \tag{23}$$

By the linear constraints in (P_2), it follows that $\mathrm{rank}(G^\star) \ge d$. If $\mathrm{rank}(G^\star) > d$, we need to round (approximate) it by a rank-d matrix. That is, we need to project it onto the domain of (P_0). One possibility would be to use random rounding that come with approximation guarantees; for example, see [22, 4]. In this work, we use deterministic rounding, namely the eigenvector rounding which retains the top d eigenvalues and discards the remaining. In particular, let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{Md}$ be the eigenvalues of $G^\star$, and $q_1, \ldots, q_d$ be the corresponding eigenvectors. Let

$$W^\star \overset{\text{def}}{=} \left[\sqrt{\lambda_1} q_1 \ \cdots \ \sqrt{\lambda_d} q_d\right]^T \in \mathbb{R}^{d \times Md}. \tag{24}$$

We now proceed as in the GRET-SPEC, namely, we define $O^\star$ and $Z^\star$ from $W^\star$ as in (20) and (21). We refer to the complete algorithm as "Global Registration over Euclidean Transforms using SDP" (`GRET-SDP`). The main steps of `GRET-SDP` are summarized in Algorithm 2.

Algorithm 2 `GRET-SDP`

Require: Membership graph Γ, local coordinates $\{x_k^{(i)}, (k,i) \in E(\Gamma)\}$, dimension d.
Ensure: Global coordinates $x_1, \ldots, x_N$ in $\mathbb{R}^d$.
Build B, L and D in (12) using Γ.
Compute $L^\dagger$ and $C = D - BL^\dagger B^T$.
$G^\star \leftarrow$ Solve the SDP (P_2) using C.
Compute top d eigenvectors of $G^\star$, and set $W^\star$ using (24).
if $\mathrm{rank}(G^\star) = d$ **then**
  $O^\star \leftarrow W^\star$.
else
  for $i = 1$ to M **do**
    Compute $W_i^\star = U_i \Sigma_i V_i^T$.
    $O_i^\star \leftarrow U_i V_i^T$.
  end for
  $O^\star \leftarrow \left[O_1^\star \cdots O_M^\star\right]$
end if
$Z^\star \leftarrow O^\star B L^\dagger$.
Return first N columns of $Z^\star$.

Similar to Observation 1, we note the following for `GRET-SDP`.

Observation 2 (Tight relaxation using `GRET-SDP`)**.** *If the rank of the solution of* (P_2) *is exactly* d*, then the coordinates and transforms computed by* `GRET-SDP` *are the global minimizers of* (P)*.*

If $\mathrm{rank}(G^\star) > d$, the output of `GRET-SDP` can only be considered as an approximation of the solution of (P). The quality of the approximation for (P_2) can be quantified using, for example, the randomized rounding in [4]. More precisely, note that since D is block-diagonal, (22) is equivalent (up to a constant term) to

$$\max_{O \in \mathcal{C}} \ \mathrm{Tr}(QO^T O)$$

where $Q = BL^\dagger B^T \succeq 0$. Bandeira et al. [4] show that the orthogonal transforms (which we continue to denote by $O^\star$) obtained by a certain random rounding of $G^\star$ satisfy

$$\mathbb{E}\left[\mathrm{Tr}(Q\, O^{\star T} O^\star)\right] \geq \alpha_d^2 \cdot \mathrm{OPT},$$

where OPT is the optimum of the unrelaxed problem (7) with $Q = BL^\dagger B^T$, and α_d is the expected average of the singular values of a $d \times d$ random matrix with entries iid $\mathcal{N}(0, 1/d)$. It was conjectured in [4] that α_d is monotonically increasing, and the boundary values were computed to be $\alpha_1 = \sqrt{2/\pi}$ (α_1 was also reported here [46]) and $\alpha_\infty = 8/3\pi$. We refer the reader to [4] for further details on the rounding procedure, and its relation to previous work in terms of the approximation ratio. Empirical results, however, suggest that the difference between deterministic and randomized rounding is small as far as the final reconstruction is concerned. We will therefore simply use the deterministic rounding.

2.5. **Computational complexity.** The main computations in `GRET-SPEC` are the Laplacian inversion, the eigenvector computation, and the orthogonal rounding. The cost of inverting L when Γ is dense is $O((N+M)^3)$. However, for most practical applications, we expect Γ to be sparse since every point would typically be contained in a small number of patches. In this case, it is known that the linear system $Lx = b$ can be solved in time almost linear in the number of edges in Γ [55, 59]. Applied to (14), this means that we can compute $L^\dagger$ in $O((N+M)|E(\Gamma)|)$ time (up to logarithmic factors). Note that, even if L is dense, it is still possible to speed up the inversion (say, compared to a direct Gaussian elimination) using the formula [30, 48]:

$$L^\dagger = [L + (N+M)^{-1} ee^T]^{-1} - (N+M)^{-1} ee^T.$$

The speed up in this case is however in terms of the absolute run time. The overall complexity is still $O((N+M)^3)$, but with smaller constants. We note that it is also possible to speed up the inversion by exploiting the bipartite nature of Γ [30], although we have not used this in our implementation.

The complexity of the eigenvector computation is $O(M^3 d^3)$, while that of the orthogonal rounding is $O(Md^3)$. The total complexity of `GRET-SPEC`, say, using a linear-time Laplacian inversion, is (up to logarithmic factors)

$$O\left(|E(\Gamma)|(N+M) + (Md)^3\right).$$

The main computational blocks in `GRET-SDP` are identical to that in `GRET-SPEC`, plus the SDP computation. The SDP solution can be computed in polynomial time using interior-point programming [65]. In particular, the complexity of computing an ε-accurate solution using interior-point solvers such as `SDPT3` [56] is $O((Md)^{4.5}\log(1/\varepsilon))$. It is possible to lower this complexity by exploiting the particular structure of (P_2). For example, notice that the constraint matrices in (P_2) have at most one non-zero coefficient. Using the algorithm in [27], one can then bring down the complexity of the SDP to $O((Md)^{3.5}\log(1/\varepsilon))$. By considering a penalized version of the SDP, we can use first-order solvers such as `TFOCS` [5] to further cut down the dependence on M and d to $O((Md)^3\varepsilon^{-1})$, but at the cost of a stronger dependence on the accuracy. The quest for efficient SDP solvers is currently an active area of research. Fast SDP solvers have been proposed that exploit either the low-rank structure of the SDP solution [11, 35] or the simple form of the linearity constraints in (P_2) [62]. More recently, a sublinear time approximation algorithm for SDP was proposed in [21]. The complexity of GRET-SDP using a linear-time Laplacian inversion and an interior-point SDP solver is thus

$$O\left(|E(\Gamma)|(N+M)+(Md)^{4.5}\log(1/\varepsilon)+(Md)^3\right).$$

For problems where the size of the SDP variable is within 150, we can solve (P_2) in reasonable time on a standard PC using SDPT3 [56] or CVX [26]. We use CVX for the numerical experiments in Section 7 that involve small-to-moderate sized SDP variables. For larger SDP variables, one can use the low-rank structure of (P_2) to speed up the computation. In particular, we were able to solve for SDP variables of size up to 2000×2000 using SDPLR [11]. We would, however, note that SDPLR uses low-rank based heuristics, and is not guaranteed to converge to the SDP solution.

3. Exact recovery

We now demonstrate that, under certain conditions on the membership graph Γ, the proposed spectral and convex relaxations can exactly reconstruct the global coordinates from the knowledge of the clean local coordinates. More precisely, let $\{\bar{x}_1,\ldots,\bar{x}_N\}$ be the true coordinates a point cloud in $\mathbb{R}^d$. Suppose that the point cloud is divided into patches whose membership graph is Γ, and that we are provided the measurements

$$x_k^{(i)}=\bar{O}_i^T(\bar{x}_k-\bar{t}_i) \qquad (k,i)\in E(\Gamma), \tag{25}$$

where $\bar{O}_i\in\mathbb{O}(d)$ and $\bar{t}_i\in\mathbb{R}^d$ are unknown. The patch-stress matrix C is constructed from Γ and the clean measurements (25).

For future use, we define the variables

$$\bar{Z}=\left[\bar{x}_1\ \cdots\ \bar{x}_N\ \bar{t}_1\ \cdots\ \bar{t}_M\right]\in\mathbb{R}^{d\times(N+M)},$$

and

$$\bar{O}=[\bar{O}_1\ \cdots\ \bar{O}_M]\in\mathbb{R}^{d\times Md} \quad\text{and}\quad \bar{G}=\bar{O}^T\bar{O}.$$

We will show that under precise conditions on Γ, we have

$$Z^\star=\Omega\bar{Z}+te^T \qquad (\Omega\in\mathbb{O}(d),\ t\in\mathbb{R}^d)$$

for both GRET-SPEC and GRET-SDP. We will refer to this as *exact recovery*. Henceforth, we will always assume that Γ is connected (clearly one cannot have exact recovery otherwise).

From (25), we can write $\bar{Z}L=\bar{O}B$. Since Γ is connected,

$$\bar{Z}=\bar{O}BL^\dagger+te^T \qquad (t\in\mathbb{R}^d). \tag{26}$$

Using (26), it is not difficult to verify that $\phi(\bar{Z},\bar{O})=\mathrm{Tr}(C\bar{G})$. Moreover, it follows from (25) that $\phi(\bar{Z},\bar{O})=0$. Therefore,

$$\mathrm{Tr}(C\bar{G})=\mathrm{Tr}(C\bar{O}^T\bar{O})=0. \tag{27}$$

Now, since the objectives in (P_1) and (P_2) are non-negative, this means that $\bar{O}$ and $\bar{G}$ are the solutions of (P_1) and (P_2). Notice that $\{\Omega\bar{O}:\Omega\in\mathbb{O}(d)\}$ are also solutions of (P_1). It is not difficult to show that we have exact recovery for either relaxation if we can guarantee these to be the *only* solutions of the respective programs.

Observation 3 (Exact recovery)**.** *If $W^\star=\Omega\bar{O},\Omega\in\mathbb{O}(d)$, then we have exact recovery for GRET-SPEC. Similarly, we have exact recovery for GRET-SDP if $\bar{G}$ is the unique solution of (P_2).*

Remark 4 (Exact recovery and rank of SDP solution)**.** *The condition $W^\star=\Omega\bar{O},\Omega\in\mathbb{O}(d)$, is necessary for exact recovery using GRET-SPEC. However, it is possible to have exact recovery using GRET-SDP even if (P_2) has high rank solutions. For example, consider two patches in $\mathbb{R}^d$ that share d overlapping points (each patch has d points). In this case, one can find a reflection (about the hyperplane containing the d points) that fixes the overlapping points. By applying this reflection to a rank-d solution of (P_2), one can obtain a solution of rank $d+1$ that results in exact recovery. However, such examples* *are rather pathological, and one cannot have exact recovery if* $\mathrm{rank}(G^\star) > d$ *provided the patches have sufficient non-overlapping points.*

The uniqueness of solutions for either relaxation can be resolved by examining the spectrum of the patch-stress matrix.

Proposition 5 (Uniqueness of solutions of GRET-SPEC)**.** *We have* $W^\star = \Omega\bar{O}, \Omega \in \mathbb{O}(d)$, *if and only if* $\mathrm{rank}(C) = (M-1)d$.

Proof. Indeed, it follows from (18) that $W^{\star T}$ is in the nullspace of C if $\mathrm{rank}(C) \leq (M-1)d$. Since $\bar{O}^T$ is also in the nullspace of C, and moreover $W^\star, \bar{O} \in \mathcal{S}$, it is not difficult to verify that $W^\star = \Omega\bar{O}, \Omega \in \mathbb{O}(d)$, if and only if $\mathrm{rank}(C) = (M-1)d$. □

Since relaxation (P_2) is tighter than (P_1), we automatically have that $G^\star = \bar{G}$ if $\mathrm{rank}(C) = (M-1)d$.

Corollary 6 (Strict convexity of GRET-SDP)**.** *If* $\mathrm{rank}(C) = (M-1)d$, *then* (P_2) *is strictly convex.*

We will henceforth say that the patch-stress C is *admissible* if $\mathrm{rank}(C) = (M-1)d$.

The previous discussion leads to the question as to under what conditions on Γ can we guarantee the patch-stress matrix to be admissible? To address this question, we introduce the following construct for Γ.

Definition 7 (Lateration)**.** Γ *is said to be laterated in* $\mathbb{R}^d$ *if there exists a reordering of the patch indices such that, for every* $2 < i \leq M$, P_i *and* $P_1 \cup \cdots \cup P_{i-1}$ *have at least* $d+1$ *non-degenerate points in common.*

Note that if Γ is laterated, then it is automatically connected, which is required to ensure exact recovery.

Theorem 8 (Lateration and Rank)**.** *If* Γ *is laterated in* $\mathbb{R}^d$, *then* C *is admissible.*

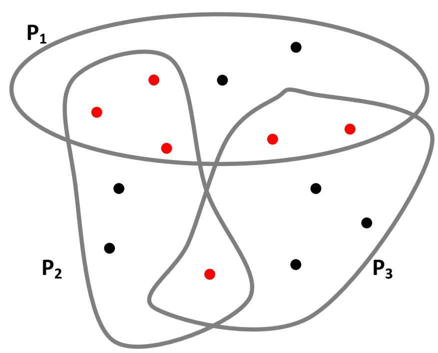

FIGURE 2. Instance of three overlapping patches, where the overlapping points are shown in red. In this case, P_3 cannot be registered with either P_1 or P_2 due to insufficient overlap. Therefore, the patches cannot be localized in two dimension using an algorithm (e.g., [66, 16]) that works by registering pairs of patches. The patches can however be registered globally using GRET-SPEC and GRET-SDP since the ordered patches P_1, P_2, P_3 form a lateration in $\mathbb{R}^2$ (cf. Definition 7).

Remark 9 (Pairwise vs. global registration)**.** *Note that any distributed algorithm (for example,* [66, 16]*) that proceeds by registering pairs of overlapping patches must consider pairs that have at least $d+1$ points in common; this is the minimum one needs to successfully register a pair of patches in $\mathbb{R}^d$. In particular, define the graph $\Gamma_P = (V_P, E_P)$, where V_P represents the patch indices, and $(i,j) \in E_P$ if and only if $|P_i \cap P_j| \geq d+1$. In order to globally register all patches using pairwise alignment, it is necessary that Γ_P be connected. Now, if Γ_P is connected, then it is easy to see that Γ is laterated. This means that, in the noiseless case, whenever a pairwise-alignment based algorithm provides an unique solution,* `GRET-SPEC` *and* `GRET-SDP` *are also guaranteed to have an unique solution.*

What is perhaps interesting is that the proposed relaxations also provide an unique solution for instances where Γ_P is not even connected. For example, consider the 3-patch system in $\mathbb{R}^2$ shown in Figure 2. Here, $|P_1 \cap P_2| = 3, |P_1 \cap P_3| = 2$, and $|P_2 \cap P_3| = 1$. Clearly, Γ_P is disconnected in this case. In particular, while it possible to register P_1 and P_2, P_3 cannot be registered with either P_1 or P_2. Therefore, a distributed algorithm which works with pairs of patches cannot recover the global coordinates of the points. However, note that the patch system is laterated in this case, and as predicted by Theorem 8, `GRET-SPEC` *and* `GRET-SDP` *can exactly recover the global coordinates in this case.*

We will now prove Theorem 8. To do so, we will need the following lemma (which will also be used later). The proof is provided in Section 9.1.

Lemma 10. *Suppose we have points $x_1, \ldots, x_N$, matrices $F_1, \ldots, F_M$, and translations $t_1, \ldots, t_M$, such that*

$$x_k = F_i x_k^{(i)} + t_i \qquad (k,i) \in E(\Gamma), \tag{28}$$

for some membership graph Γ. Suppose that Γ is connected, and each patch in Γ contains $d+1$ or more points. Then the span of $[F_1 \cdots F_M]$ is identical to the affine span of $\{x_1, \ldots, x_N\}$.

We recall that the affine span of of a set of points P is defined as the linear subspace of smallest dimension whose translate contains P. The dimension of this subspace is the affine rank of P.

Proof of Theorem 8. We know that the nullity of C is at least d (the rows of $\bar{O}$ are in its null space), so that $\mathrm{rank}(C) \leq (M-1)d$. We need to show that the rank is exactly $(M-1)d$ under the given assumptions.

We claim that if $\mathrm{Tr}(CX) = 0$ implies $\mathrm{rank}(X) \leq d$ for any $X \succeq 0$, then $\mathrm{rank}(C) = (M-1)d$. Suppose the former is true, but $\mathrm{rank}(C) < (M-1)d$. We will show that $\mathrm{Tr}(CY) = 0$ for some $Y \succeq 0$ with $\mathrm{rank}(Y) > d$, which will contradict our hypothesis. Indeed, since the nullity of C is $d+1$ or more, we can find some null vector u that does not belong to the span of the rows of $\bar{O}$. Take $Y = \bar{G} + uu^T \succeq 0$ whose rank is $d+1$, and note that $\mathrm{Tr}(CY) = 0$.

It remains to show that under the assumption of Theorem 8, $\mathrm{Tr}(CX) = 0$ implies $\mathrm{rank}(X) \leq d$ for any $X \succeq 0$. Fix the order of patches so that $P_1, \cdots, P_M$ satisfy the lateration condition.

Take $X \succeq 0$ and write $X = F^T F$, where $F \in \mathbb{R}^{Md \times Md}$. We are done if we can show that $\mathrm{rank}(F) \leq d$. Substituting for C, we can write

$$0 = \mathrm{Tr}(FCF^T) = \sum_{(k,i) \in E(\Gamma)} \|Ze_{ki} - F(e_i \otimes I_d)x_k^{(i)}\|^2,$$

where $Z \stackrel{\text{def}}{=} FBL^{\dagger}$. Write $F = [F_1, \ldots, F_M]$ where each $F_i \in R^{Md \times d}$. Then we have the equations

$$Ze_{ki} = F_i x_k^{(i)} \qquad (k,i) \in E(\Gamma).$$

Further, define x_k to the k-th column of Z for $1 \leq k \leq M$, and t_i to be the $(N+i)$-th column of Z for $1 \leq i \leq M$. Then

$$x_k = F_i x_k^{(i)} + t_i \qquad (k,i) \in E. \tag{29}$$

Clearly, the rank of each F_i is at most d. We are done if we can show that the columns of $F_1, \ldots, F_M$ span the same space.

We proceed by inducting on M. When $M = 2$, we know that there are $d+1$ points in common between P_1 and P_2, say, $x_1, \ldots, x_{d+1}$. From (29), we have

$$x_k = F_1 x_k^{(1)} + t_1 = F_2 x_k^{(2)} + t_2 \qquad (k = 1, 2, \ldots, d+1).$$

We now take out the translation by subtracting pairs of equations. In particular, we fix the first equation and subtract the remaining d equations from it. This gives us the equation

$$F_1 X_1 = F_2 X_2, \tag{30}$$

where

$$X_i = [x_2^{(i)} - x_1^{(i)} \ \cdots \ x_{d+1}^{(i)} - x_1^{(i)}] \in \mathbb{R}^{d \times d} \qquad (i = 1, 2).$$

Since the common points are assumed to be non-degenerate, it follows that the X_1 and X_2 are of full rank. Therefore, F_1 and F_2 span the same space.

Assume, by the induction hypothesis, that the result holds for the first $2 \leq m < M$ patches. Namely, that the columns of $F_1, \ldots, F_m$ span the same space. Let $x_1, \ldots, x_{d+1}$ (after relabeling) be the points belonging to the union of $P_1, \ldots, P_m$ that are also in P_{m+1}. It trivially follows from the lateration assumption that each patch contains $d+1$ or more non-degenerate points, and that Γ is connected. Therefore, by Lemma 10, the affine span of $\{x_1, \cdots, x_{d+1}\}$ is contained in the span of $[F_1 \cdots F_m]$.

On the other hand, from (36), we have

$$x_k = F_{m+1} x_k^{(m+1)} + t_{m+1} \qquad (k = 1, 2, \ldots, d+1).$$

After subtraction, we get

$$F_{m+1}\,[x_2^{(m+1)} - x_1^{(m+1)} \cdots x_{d+1}^{(m+1)} - x_1^{(m+1)}] = [x_2 - x_1 \cdots x_{d+1} - x_1].$$

Now, since the matrix on the left is invertible (by the non-degeneracy assumption), it follows that the span of the columns of F_{m+1} is identical to the affine span of $\{x_1, \ldots, x_{d+1}\}$. Therefore, the columns of F_{m+1} have the same span as $[F_1 \cdots F_m]$. This completes the induction. □

The previous discussion leads to the following questions:

- Is lateration also necessary for exact rank recovery?
- For what class of Γ would `GRET-SPEC` fail to recover the exact coordinates, while the tighter relaxation `GRET-SDP` would succeed?
- Is there a necessary and sufficient condition for exact recovery using `GRET-SDP`?

The answer to the first question is in the negative, and as a counterexample, we again consider a 3-patch system shown in Figure 1. This patch system is not laterated in $\mathbb{R}^2$, since no two patches have three points in common. However, numerical experiments (cf. Section 7) show that we have exact recovery for this example using `GRET-SDP`. On the other hand, if the points are chosen to be non-degenerate, the rank of the patch-stress matrix turns out to be less than $(M-1)d = 4$ for this example. As a result, following Proposition 5, `GRET-SPEC` would fail to recover the exact coordinates in this case. To explain why `GRET-SDP` succeeds while `GRET-SPEC` fails for this patch system, and to resolve the final question, we will now take a different route.

4. Unique localizability and rank recovery

Corollary 6 tells us that (P_2) is strictly convex if C is admissible. We will now show that the strict convexity of (P_2) can be established under weaker conditions. To do so, we will use a standard result in the theory of semidefinite programming: If the solution of SDP's such as (P_2) is guaranteed to be of fixed rank, then the solution is necessarily unique [37, pg. 36-39]. That is, if we can guarantee the rank of $G^\star$ to be exactly d, then $G^\star = \bar{G}$. We will refer to this as *rank recovery*.

To establish conditions for rank recovery, we will need a particular notion of graph rigidity called *unique localizability* due to So and Ye [54]. The setting here is that we are given a graph

(V, E) and distances $\{d_{kl} : (k,l) \in E\}$. A set of points $(x_k)_{k \in V}$ in $\mathbb{R}^d$ is said to be a *realization* of $\{d_{kl} : (k,l) \in E\}$ in $\mathbb{R}^d$ if $d_{kl} = ||x_k - x_l||$ for $(k,l) \in E$.

Definition 11 (Unique localizability, [54])**.** *We say that a graph (V, E), along with distances $\{d_{kl} : (k,l) \in E\}$, is uniquely localizable in $\mathbb{R}^d$ if*

(a) *There exists a realization of $\{d_{kl} : (k,l) \in E\}$ in $\mathbb{R}^d$.*

(b) *One cannot find a realization of $\{d_{kl} : (k,l) \in E\}$ whose affine rank is larger than d.*

We will show that rank recovery for `GRET-SDP` is equivalent to the unique localizability of a graph (derived from the membership graph) together with distances (defined on this new graph) computed from the local coordinates. In particular, consider the graph $\Gamma_B = (V_B, E_B)$, where $V_B = \{1, 2, \ldots, N\}$ and $(k,l) \in E_B$ if and only if x_k and x_l belong to the same patch. We call Γ_B the *backbone* graph (see Figure 3). We associate the following distances with Γ_B:

$$d_{kl} = \|x_k^{(i)} - x_l^{(i)}\| \qquad (k,l) \in E_B, \tag{31}$$

where $x_k, x_l \in P_i$, say. Note that the above assignment is independent of the choice of patch.

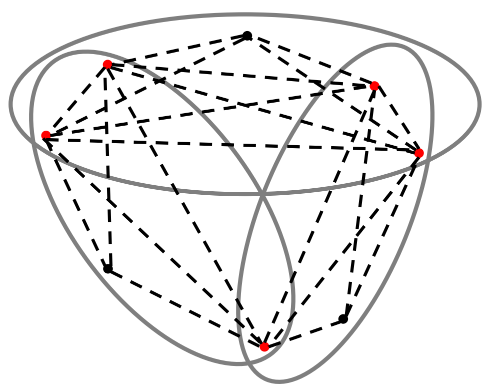

FIGURE 3. This shows the backbone graph for a 3-patch system. The edges of the backbone graph are obtained by connecting points that belong to the same patch. The edges within a given patch are marked with the same color. We prove in the text that `GRET-SDP` can successfully register all the patches if the backbone graph is rigid in a certain sense.

It is clear that Γ_B, together with the distances (31), has a realization in $\mathbb{R}^d$, namely the one arising from the clean configuration. What is less obvious is whether it also has a realization in some higher dimension. The resolution of this question tells us when $\mathrm{rank}(G^\star) = d$. Intuitively, there cannot be exact recovery if the backbone is realizable in a space whose affine dimension is larger than d.

Theorem 12 (Rank recovery and uniquely localizability)**.** *If every patch contains at least $d+1$ non-degenerate points, then the following are equivalent:*

(1) *The rank of $G^\star$ is d.*

(2) Γ_B *along with the distances* (31) *is uniquely localizable in $\mathbb{R}^d$.*

Before turning to the proof of Theorem (12), we make some observations.

Remark 13 (`GRET-SPEC` fails, but `GRET-SDP` succeeds)**.** *The significance of Theorem 12 is that it offers an explanation as to why `GRET-SDP` works perfectly for the patch system in Figure 1, while `GRET-SPEC` fails. Note that the backbone graph in this case is trilaterated (we recall that a graph is a* *trilateration if there exists an ordering of the vertices such that the first three vertices form a triangle, and every vertex after that is connected to at least three vertices earlier in the order). Now it is well-known that trilaterated graphs are uniquely localizable in* $\mathbb{R}^2$ [54], *which explains why* `GRET-SDP` *is able to recover the exact coordinates for the patch system in this example.*

Notice that, following Theorem 12, it is necessary that the backbone graph in Figure 2 is uniquely localizable. In fact, the backbone graph in this case is also trilaterated.

A precise explanation as to why `GRET-SPEC` *fails in this example is as follows. As explained (we skip the details) in Figure 4, we can construct a matrix* $F = [F_1, F_2, F_3]$, *where each* F_i *is a* 3×2 *matrix, such that*

$$\mathrm{rank}(F) > 2 \quad \textit{and} \quad \mathrm{Tr}(FCF^T) = 0.$$

In other words, the patch-stress matrix C *has a larger nullspace than what is required to make it admissible. The larger nullspace is precisely due to the additional degree of freedom coming from the affine constraints as against what we would have if we restricted the transforms to be orthogonal. Indeed, it is the diagonal (orthogonality) constraints in* (P_1) *that eliminates the extra degree of freedom coming from the affine transforms.*

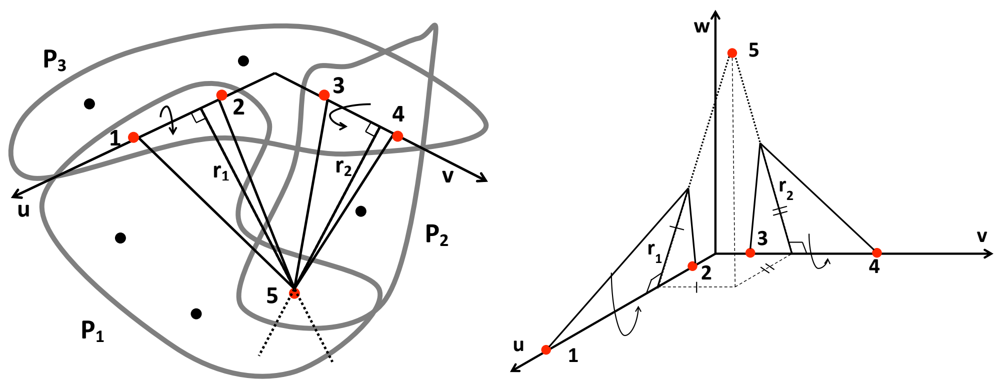

FIGURE 4. We give an explanation as to why the point-patch system in Figure 1 cannot be registered using `GRET-SPEC`, and in particular, why the patch-stress matrix in this example fails to be admissible. Recall that the backbone graph in this case is uniquely localization (forms a trilateration), and is thus uniquely localizable in $\mathbb{R}^2$. **Left**: Unique configuration in $\mathbb{R}^2$ (up to a rigid transform) that satisfies the measurements (25). **Right**: Configuration in $\mathbb{R}^3$ obtained from the one on the left through affine transforms applied on (the points in) patches P_1 and P_2. In particular, we first scale P_1 along r_1, and then rotate it about the u axis. We then scale P_2 along r_2, and rotate it about the v axis. It can be shown the combined affine transform is in the null space of the patch-stress matrix, and that the rank of the transform is greater than 2 (extra degree of freedom obtained by replacing the orthogonal constraints by affine constraints). This explains why the patch-stress matrix is rank deficient (see Remark 13 for a precise description).

We now give the proof Theorem 12. To do so, we introduce the following notations. Let $\lambda_1, \cdots, \lambda_{Md}$ be the full set of eigenvalues of $G^\star$ sorted in non-increasing order, and $q_1, \ldots, q_{Md}$ be the corresponding eigenvectors. Define

$$O^{\star\star} \stackrel{\text{def}}{=} \left[\sqrt{\lambda_1} q_1 \ \cdots \ \sqrt{\lambda_{Md}} q_{Md}\right]^T \in \mathbb{R}^{Md \times Md}, \tag{32}$$

and $O_i^{\star\star}$ to be the i-th $Md \times d$ block of $O^{\star\star}$, that is,

$$O^{\star\star} \stackrel{\text{def}}{=} [O_1^{\star\star} \ \cdots \ O_M^{\star\star}]. \tag{33}$$

This should not lead to confusion since we will no longer refer to (18) in the rest of the paper.

By construction, $G^\star = {O^{\star\star}}^T O^{\star\star}$. In particular, the rank of $G^\star$ is identical to that of $O^{\star\star}$. This reduces the problem to one of understanding when the rank of the latter is exactly d. Moreover, note that by feasibility,

$$G^\star_{ii} = {O^{\star\star}_i}^T O^{\star\star}_i = I_d \quad (1 \le i \le M).$$

Thus the d columns of $O^{\star\star}_i$ form an orthonormal system in $\mathbb{R}^{Md}$.

Now define

$$Z^{\star\star} \stackrel{\text{def}}{=} O^{\star\star} B L^\dagger \in \mathbb{R}^{Md\times(N+M)}, \tag{34}$$

and, denoting the p-th column of $Z^{\star\star}$ by $Z^{\star\star}_p$, define

$$x^{\star\star}_k \stackrel{\text{def}}{=} Z^{\star\star}_k \ (1 \le k \le N) \quad \text{and} \quad t^{\star\star}_i \stackrel{\text{def}}{=} Z^{\star\star}_{N+i} \ (1 \le i \le M). \tag{35}$$

The configuration $\{x^{\star\star}_1, \ldots, x^{\star\star}_N\}$ and the translations $\{t^{\star\star}_1, \ldots, t^{\star\star}_M\}$ live in $\mathbb{R}^{Md}$. Roughly speaking, the desired configuration $Z^\star$ is obtained by projecting $\{x^{\star\star}_1, \ldots, x^{\star\star}_N\}$ onto a d-dimensional subspace of $\mathbb{R}^{Md}$. If $\mathrm{rank}(G^\star) = d$, then the projection is trivial, $Z^{\star\star} = Z^\star$. In this case, we have seen that the configuration and the transforms defined in (35), and (33) satisfy the measurement model in (25). More generally, even if the rank is larger than d, we have the following relation.

Proposition 14 (Consistency)**.**

$$x^{\star\star}_k = O^{\star\star}_i x^{(i)}_k + t^{\star\star}_i \qquad (k, i) \in E(\Gamma). \tag{36}$$

In other words, the configuration and the transforms that are obtained by solving (P_2) are consistent with the measurements in $\mathbb{R}^{Md}$.

Proof. Note that from (34), (35), and (33), we can write

$$\sum_{(k,i)\in E(\Gamma)} \|x^{\star\star}_k - O^{\star\star}_i x^{(i)}_k - t^{\star\star}_i\|^2 = \phi(Z^{\star\star}, O^{\star\star}) = \phi(O^{\star\star} B L^\dagger, O^{\star\star}) = \mathrm{Tr}(C G^\star),$$

where the Euclidean norm on the left is on $\mathbb{R}^{Md}$. Then, from (23) and (27), we conclude that $\mathrm{Tr}(CG^\star) = 0$. This gives us the consistency relations. $\square$

We now show how the above consistency relations along with Lemma 10 can be used to infer when the rank of $O^{\star\star}$ is exactly d.

Proof of Theorem (12). Note that if either (1) or (2) holds, then the membership graph Γ must be connected.

(2) ⇒ (1): If Γ_B is unique localizable, then, as a realization, $\{x^{\star\star}_1, x^{\star\star}_2, \ldots, x^{\star\star}_N\}$ must have affine rank d. However, Lemma 10 tells us that the affine rank of $\{x^{\star\star}_1, x^{\star\star}_2, \ldots, x^{\star\star}_N\}$ equals that of $O^\star$.

(1) ⇒ (2): We need to show that any realization of Γ_B (along with the distances (31)) has affine rank d. Let $\{y_1, \ldots, y_N\}$ be such a realization. Then

$$\|y_k - y_l\| = \|x^{(i)}_k - x^{(i)}_l\| \qquad (k, l) \in E_B.$$

For a fixed patch P_i, consider the realizations $\Pi_1 = \{y_k : (k, i) \in E(\Gamma)\}$ and $\Pi_2 = \{x^{\star\star}_k : (k, i) \in E(\Gamma)\}$. It follows from (36) that

$$\|x^{\star\star}_k - x^{\star\star}_l\| = \|x^{(i)}_k - x^{(i)}_l\| \qquad (k, l) \in E_B.$$

Therefore, the distance between a pair of points in Π_1 is equal to that of a pair of corresponding points in Π_2. Since P_i contains $d + 1$ or more non-degenerate points, it must be that Π_1 is obtained through some rigid transform of Π_2. Combining this with (36), it is easy to see that

$$y_k = O_i x^i_k + t_i \qquad (k, i) \in E(\Gamma), \tag{37}$$

where the columns of $O_i \in \mathbb{R}^{Md\times d}$ form an orthonormal system in $\mathbb{R}^{Md}$.

Let $O = [O_1 \cdots O_M]$. It follows from (37) and Lemma 10 that the affine rank of $\{y_1, \ldots, y_N\}$ is identical to $\mathrm{rank}(O)$. Moreover, it is not difficult to show using (37) that $\mathrm{Tr}(CO^T O) = 0$, so that $O^T O$ is optimal for (P_2). Since SDP solvers return the maximum rank solution, it follows that $\mathrm{rank}(O) \leq \mathrm{rank}(O^{\star\star}) = d$ This establishes our claim. □

5. Randomized rank test

In this Section, we propose an efficient randomized test for exact recovery. That is, given some membership graph Γ and clean measurements defined on Γ, we describe a test that takes Γ and returns certificates of exact recovery for `GRET-SPEC` and `GRET-SDP`. First, let us consider the various conditions of exact recovery described in the previous Sections:

- A necessary and sufficient condition for exact recovery using `GRET-SPEC` is that the patch-stress matrix constructed from the clean measurements is admissible.
- A sufficient condition for the patch-stress matrix to be admissible is that Γ is laterated.
- This rank condition is sufficient for exact recovery using `GRET-SDP`, but is not necessary.
- A necessary and sufficient condition for exact recovery using `GRET-SDP` is that the backbone graph of Γ is unique localizability.

To the best of our knowledge, there is no known polynomial-time algorithm for testing lateration for a given patch system. As described earlier, it is however possible to construct patch systems in certain applications that are tailored to satisfy the lateration condition. Thus, while the lateration condition is useful in practice when one has some control on Γ, it cannot be used as a test for either relaxation.

On the other hand, it is known that unique localizability can be certified in polynomial time using an SDP-based test [54]. This test can, in principle, be used to certify exact recovery for `GRET-SDP`. However, the test requires one to run an SDP program where the size of the variable scales linearly with the size of the backbone graph (cf. [54] for details). Since the size of the backbone graph is comparable to the size Γ, namely $N + M$, the complexity of this test would far exceed the complexity of `GRET-SDP` itself. In short, the SDP-based algorithm in [54] cannot be used as an efficient test for `GRET-SDP`.

This leaves us with the rank condition, which can of course be tested efficiently. However, we note that the rank condition is only a sufficient condition for `GRET-SDP`. In particular, if the patch-stress matrix is not admissible the test is inconclusive. That is, it is possible that the given Γ is good enough to guarantee exact recovery (e.g., the backbone is uniquely localizable) even though the patch-stress matrix fails to be admissible.

As we will see in Section 6, the rank condition also plays a role in determining the performance of `GRET-SDP` when the measurements are corrupted. What we can hope for in this case is that the membership graph Γ should at least guarantee exact recovery if clean measurements were provided instead. The point is that it would not make sense to register the patches with noisy measurements if exact recovery cannot be guaranteed with clean measurements. An efficient test then would be to test the rank of the patch-stress matrix constructed from the clean measurements. The problem, however, is that we do not have access to the clean measurements (else, there would be noting to solve for). To bypass this problem, we propose a procedure similar to the randomized tests proposed for local rigidity by Hendrickson [28], for generic global rigidity by Gortler et al. [24], and for matrix completion by Singer and Cucuringu [52].

Before we do so, we need a result. Let us continue to denote the patch-stress matrix obtained from Γ and the measurements (25) by C. We introduce the notation C_0 to denote the patch-stress matrix obtained from the same graph Γ, but using the (unknown) original coordinates as measurements, namely,

$$x_k^{(i)} = \bar{x}_k \qquad (k, i) \in \Gamma. \tag{38}$$

Algorithm 3 RRT

Require: Membership graph Γ, and dimension d.
Ensure: Exact recovery certificate for GRET-SDP.

Build L using Γ, and compute $L^{\dagger}$.
Randomly pick $\{x_1, \ldots, x_N\}$ from the unit cube in $\mathbb{R}^d$, where $N = |V_x(\Gamma)|$.
$x_k^{(i)} \leftarrow x_k$ for every $(k,i) \in E(\Gamma)$.
$C_0 \leftarrow D - BL^{\dagger}B^T$.
if $\mathrm{rank}(C_0) = (M-1)d$ **then**
    Positive certificate for GRET-SPEC and GRET-SDP.
else
    Negative certificate for GRET-SPEC.
    GRET-SDP cannot be certified.
end if

The advantage of working with C_0 over C is that the former can be computed using just the global coordinates, while the latter requires the knowledge of the global coordinates as well as the clean transforms. In particular, this only requires us to simulate the global coordinates. Since the coordinates of points in a given patch are determined up to a rigid transform, we claim the following (cf. Section 9.2 for a proof).

Proposition 15 (Rank equivalence). *For a fixed Γ, C and C_0 have the same rank.*

In other words, the rank of C_0 can be used to certify exact recovery. The proposed test is based on Proposition 9.2, and the fact that if two different *generic* configurations are used as input in (38) (for the same Γ), then the patch-stress matrices they produce would have the same rank. By generic, we mean that the coordinates of the configuration do not satisfy any non-trivial algebraic equation with rational coefficients [24]. It is not difficult to reason that if the points in the configuration are drawn randomly (and independently) from a non-singular distribution (say, the uniform distribution over the unit cube in $\mathbb{R}^d$), then the configuration is generic with probability one. The idea then is to randomly pick $x_1, \ldots, x_N$ from this distribution, and use it in place of the unknown (38) to compute the patch-stress matrix and its rank. If the rank is $(M-1)d$, we return a positive certificate for GRET-SPEC and GRET-SDP. On the other hand, if the rank is less than $(M-1)d$, we return a negative certificate for GRET-SPEC. However, we are not able to certify GRET-SDP in this case. The complete test is described in Algorithm 3. We will henceforth refer to this as the "Randomized Rank Test" (RRT). Note that the main computations in RRT are the Laplacian inversion (which is also required for the registration algorithm), and the rank determination.

6. Stability analysis

We have so far studied the problem of exact recovery from noiseless measurements. In practice, however, the measurements are invariably noisy. This brings us to the question of stability, namely how stable are GRET-SPEC and GRET-SDP to perturbations in the measurements? Numerical results (to be presented in the next Section) show that both the relaxations are indeed quite stable to perturbations. In particular, the reconstruction error degrades quite gracefully with the increase in noise (reconstruction error is the gap between the outputs with clean and noisy measurements). In this Section, we try to quantify these empirical observations. In particular, we show that, for a specific noise model, the reconstruction error grows at most linearly with the level of noise.

The noise model we consider is the "bounded" noise model. Namely, we assume that that the measurements are obtained through bounded perturbations of the clean measurements in (25). More precisely, we suppose that we have a membership graph Γ, and that the observed local coordinates are of the form

$$x_k^{(i)} = \bar{O}_i^T(\bar{x}_k - \bar{t}_i) + \epsilon_{k,i}, \qquad \|\epsilon_{k,i}\| \leq \varepsilon \quad (k,i) \in E(\Gamma). \tag{39}$$

In other words, every coordinate measurement is offset within a ball of radius ε around the clean measurements. Here, ε is a measure of the noise level per measurement. In particular, $\varepsilon = 0$ corresponds to the case where we have the clean measurements (25).

Since the coordinates of points in a given patch are determined up to a rigid transform, it is clear that the above problem is equivalent to the one where the measurements are

$$x_k^{(i)} = \bar{x}_k + \epsilon_{k,i}, \qquad \|\epsilon_{k,i}\| \leq \varepsilon \quad (k,i) \in E(\Gamma). \tag{40}$$

By equivalent, we mean that the reconstruction errors obtained using either (39) or (40) are equal. The reason we use the latter measurements is that the analysis in this case is much more simple.

The reconstruction error is defined as follows. Generally, let $Z^\star$ be the output of Algorithms 1 and 2 using (40) as input, and let

$$Z_0 \stackrel{\text{def}}{=} [\bar{x}_1 \cdots \bar{x}_N \; 0 \cdots 0] \in \mathbb{R}^{d \times (N+M)}, \tag{41}$$

where we assume that the centroid of $\{\bar{x}_1, \cdots, \bar{x}_N\}$ is at the origin.

Ideally, we would require that $Z^\star = Z_0$ (up to a rigid transformation) when there is no noise, that is, when $\varepsilon = 0$. This is the exact recovery phenomena that we considered earlier. In general, the gap between Z_0 and $Z^\star$ is a measure of the reconstruction quality. Therefore, we define the reconstruction error to be

$$\eta = \min_{\Theta \in \mathbb{O}(d)} \|Z^\star - \Theta Z_0\|_F.$$

Note that we are not required to factor out the translation since Z_0 is centered by construction.

Our main results are the following.

Theorem 16 (Stability of `GRET-SPEC`). *Assume that R is the radius of the smallest Euclidean ball that encloses the clean configuration $\{\bar{x}_1, \ldots, \bar{x}_N\}$. For fixed noise level $\varepsilon \geq 0$ and membership graph Γ, suppose we input the noisy measurements* (40) *to `GRET-SPEC`. If C_0 is admissible, then we have the following bound for `GRET-SPEC`:*

$$\eta \leq \frac{|E(\Gamma)|^{1/2}}{\lambda_2(L)}(K_1 \varepsilon + K_2 \varepsilon^2),$$

where

$$K_1 = \frac{8\pi R}{\mu_{d+1}(C)}\sqrt{2MN|E(\Gamma)|(2+N)d(d+1)}\left(4R\frac{\sqrt{N|E(\Gamma)|}}{\lambda_2(L)} + 1\right) + \sqrt{2+N+M}.$$

and

$$K_2 = \frac{8\pi R}{\mu_{d+1}(C)}\sqrt{2MN|E(\Gamma)|(2+N)d(d+1)}\left(2\frac{\sqrt{N|E(\Gamma)|}}{\lambda_2(L)} + 1\right).$$

Here $\lambda_2(L)$ is the second smallest eigenvalue of L.

We assume here that $\mu_{d+1}(C)$ is non-zero[2]. The bounds here are in fact quite loose. Note that when $\varepsilon = 0$, then by the admissibility assumption $\mu_{d+1}(C) > 0$, and we recover the perfect reconstruction results for `GRET-SPEC`.

[2]Numerical experiments suggest that this is indeed the case if C_0 is admissible. In fact, we notice a growth in the eigenvalue with the increase in noise level. We have however not been able to prove this fact.

Theorem 17 (Stability of `GRET-SDP`)**.** *Under the conditions of Theorem 16, we have the following for GRET-SDP:*

$$\eta \leq \frac{|E(\Gamma)|^{1/2}}{\lambda_2(L)} \Big[32\sqrt{2d(d+1)(2+N)|E(\Gamma)|}\mu_{d+1}^{-1/2}(C_0)R + \sqrt{2+N+M} \Big] \varepsilon.$$

The bounds are again quite loose. The main point is that the reconstruction error for `GRET-SDP` is within a constant factor of the noise level. In particular, when $\varepsilon = 0$ (measurements are clean), we recover the perfect reconstruction results.

The rest of this Section is devoted to the proofs of Theorem 16 and 17. First, we introduce some notations.

Notations. We note that the patch-stress matrix in (P_1) is computed from the noisy measurements (40), and the same patch-stress matrix is used in (P_2). The quantities $G^\star, W^\star, O^\star$, and $Z^\star$ are as defined in Algorithms 1 and 2 . We continue to denote the clean patch-stress matrix by C_0. Define

$$O_0 \overset{\text{def}}{=} [I_d \cdots I_d] \quad \text{and} \quad G_0 \overset{\text{def}}{=} O_0^T O_0.$$

Let $e_1, \ldots, e_d$ be the standard basis vectors of $\mathbb{R}^d$, and let e be the all-ones vector of length M. Define

$$s_i \overset{\text{def}}{=} \frac{1}{\sqrt{M}} e \otimes e_i \in \mathbb{R}^{Md} \qquad (1 \leq i \leq d). \tag{42}$$

Note that every $d \times d$ block of G_0 is I_d, and that we can write

$$G_0 = \sum_{i=1}^{d} M s_i s_i^T. \tag{43}$$

We first present an estimate that applies generally to both algorithms. The proof is provided in Section 9.3.

Proposition 18 (Basic estimate)**.** *Let R be the radius of the smallest Euclidean ball that encloses the clean configuration. Then, for any arbitrary Θ,*

$$\|Z^\star - \Theta Z_0\|_F \leq \frac{|E(\Gamma)|^{1/2}}{\lambda_2(L)} \Big[R(2+N)^{1/2} \|O^\star - \Theta O_0\|_F + \varepsilon(2+N+M)^{1/2} \Big]. \tag{44}$$

In other words, the reconstruction error in either case is controlled by the rounding error:

$$\delta = \min_{\Theta \in \mathbb{O}(d)} \|O^\star - \Theta O_0\|_F. \tag{45}$$

The rest of this Section is devoted to obtaining a bound on δ for `GRET-SPEC` and `GRET-SDP`. In particular, we will show that δ is of the order of ε in either case. Note that the key difference between the two algorithms arises from the eigenvector rounding, namely the assignment of the "unrounded" orthogonal transform $W^\star$ (respectively from the patch-stress matrix and the optimal Gram matrix). The analysis in going from $W^\star$ to the rounded orthogonal transform $O^\star$, and subsequently to $Z^\star$, is however common to both algorithms.

We now bound the error in (45) for both algorithms. Note that we can generally write

$$W^\star = \big[\sqrt{\alpha_1} u_1 \cdots \sqrt{\alpha_d} u_d \big]^T,$$

where $u_1, \ldots, u_d$ are orthonormal. In `GRET-SPEC`, each $\alpha_i = M$, while in `GRET-SDP` we set α_i using the eigenvalues of $G^\star$.

Our first result gives a control on the quantities obtained using eigenvector rounding in terms of their Gram matrices.

Lemma 19 (Eigenvector rounding)**.** *There exist $\Theta \in \mathbb{O}(d)$ such that*

$$\|W^\star - \Theta O_0\|_F \leq \frac{4}{\sqrt{M}} \|W^{\star T} W^\star - G_0\|_F.$$

Next, we use a result by Li [40] to get a bound on the error after orthogonal rounding.

Lemma 20 (Orthogonal rounding). *For arbitrary $\Theta \in \mathbb{O}(d)$,*

$$\|O^\star - \Theta O_0\|_F \leq 2\sqrt{d+1}\, \|W^\star - \Theta O_0\|_F.$$

The proofs of Lemma 19 and 20 are provided in Appendices 9.4 and 9.5. At this point, we record a result from [42] which is repeatedly used in the proof of these lemmas and elsewhere.

Lemma 21 (Mirsky, [42]). *Let $|||\cdot|||$ be some unitarily invariant norm, and let $A, B \in \mathbb{R}^{n\times n}$. Then*

$$|||\, \mathrm{diag}(\sigma_1(A) - \sigma_1(B), \cdots, \sigma_n(A) - \sigma_n(B))\, ||| \leq |||A - B|||.$$

In particular, the above result holds for the Frobenius and spectral norms.

By combining Lemma 19 and 20, we have the following bound for (45):

$$\delta \leq 8\sqrt{\frac{d+1}{M}}\, \|W^{\star T} W^\star - G_0\|_F. \tag{46}$$

We now bound the quantity on the right in (46) for GRET−SPEC and GRET−SDP.

6.1. **Bound for GRET−SPEC.** For the spectral relaxation, this can be done using the Davis-Kahan theorem [7]. Note that from (18), we can write

$$\frac{1}{M}(W^{\star T} W^\star - G_0) = \sum_{i=1}^{d} r_i r_i^T - \sum_{j=1}^{d} s_j s_j^T. \tag{47}$$

Following [7, Ch. 7], let A be some symmetric matrix and S be some subset of the real line. Denote $P_A(S)$ to be the orthogonal projection onto the subspace spanned by the eigenvectors of A whose eigenvalues are in S. A particular implication of the Davis-Kahan theorem is that

$$\|\, P_A(S_1) - P_B(S_2)\, \|_{\mathrm{sp}} \leq \frac{\pi}{2\rho(S_1^c, S_2)} \|A - B\|_{\mathrm{sp}}, \tag{48}$$

where S_1^c is the complement of S_1, and $\rho(S_1, S_2) = \min\{|u - v| : u \in S_1, v \in S_2\}$.
In order to apply (48) to (47), set $A = C, B = C_0, S_1 = [\mu_1(C),\ \mu_d(C)]$, and $S_2 = \{0\}$. If C_0 is admissible, then $P_B(S_2) = \sum_{j=1}^d s_j s_j^T$. Applying (48), we get

$$\|W^{\star T} W^\star - G_0\|_{\mathrm{sp}} \leq \frac{M\pi}{2\mu_{d+1}(C)} \|C - C_0\|_F. \tag{49}$$

Now, it is not difficult to verify that for the noise model (40),

$$\|C - C_0\|_F \leq 2\sqrt{N|E(\Gamma)|} \left[\Big(4R\frac{\sqrt{N|E(\Gamma)|}}{\lambda_2(L)} + 1\Big)\varepsilon + \Big(2\frac{\sqrt{N|E(\Gamma)|}}{\lambda_2(L)} + 1\Big)\varepsilon^2 \right]. \tag{50}$$

Combining Proposition 18 with (46),(49), and (50), we arrive at Theorem 16.

6.2. **Bound for GRET−SDP.** To analyze the bound for GRET−SDP, we require further notations. Recall (42), and let S be the space spanned by $\{s_1, \ldots, s_d\} \subset \mathbb{R}^{Md}$, and let $\bar{S}$ be the orthogonal complement of S in $\mathbb{R}^{Md}$. In the sequel, we will be required to use matrix spaces arising from tensor products of vector spaces. In particular, given two subspaces U and V of $\mathbb{R}^{Md}$, denote by $U \otimes V$ the space spanned by the rank-one matrices $\{uv^T : u \in U, v \in V\}$. In particular, note that G_0 is in $S \otimes S$.
Let $A \in \mathbb{R}^{Md\times Md}$ be some arbitrary matrix. We can decompose it into

$$A = P + Q + T \tag{51}$$

where

$$P \in S \otimes S,\ Q \in (S \otimes \overline{S}) \cup (\overline{S} \otimes S),\ \text{and } T \in \overline{S} \otimes \overline{S}.$$

We record a result about this decomposition from Wang and Singer [61].

Lemma 22 ([61], pg. 7)**.** *Suppose* $G_0 + \Delta \succeq 0$ *and* $\Delta_{ii} = 0$ $(1 \leq i \leq M)$*. Let* $\Delta = P + Q + T$ *as in* (51)*. Then*

$$T \succeq 0, \quad \textit{and} \quad P_{ij} = -\frac{1}{M}\sum_{l=1}^{M} T_{ll} \quad (1 \leq i, j \leq M).$$

It is not difficult to verify that $\mathrm{Tr}(C_0 G_0) = 0$ and that $C_0 \succeq 0$. From (43), we have

$$0 = \mathrm{Tr}(C_0 G_0) = \sum_{i=1}^{d} s_i^T C_0 s_i \geq 0.$$

Since each term in the above sum is non-negative, $C_0 s_i = 0$ for $1 \leq i \leq d$. In other words, S is contained in the null space of C_0. Moreover, if C_0 is admissible, then S is exactly be the null space of C_0. Based on this observation, we give a bound on the residual T.

Proposition 23 (Bound on the residual)**.** *Suppose that* C_0 *is admissible. Decompose* $\Delta = P + Q + T$ *as in* (51)*. Then*

$$\mathrm{Tr}(T) \leq 4\mu_{d+1}^{-1}(C_0)|E(\Gamma)|\varepsilon^2. \tag{52}$$

Proof. The main idea here is to compare the objective in (P_0) with the trace of T. To do so, we will use the unrounded Z and O defined in (32) and (34). In particular, we will use the fact that $(Z^{\star\star}, O^{\star\star})$ are the minimizers of the unconstrained program

$$\min_{(Z,O)} \sum_{(k,i)\in E(\Gamma)} \|Z e_{ki} - O_i x_k^{(i)}\|^2 \quad \text{s.t.} \quad Z \in \mathbb{R}^{Md\times(N+M)},\ O \in \mathbb{R}^{Md\times Md}. \tag{53}$$

Note that $\mathrm{Tr}(C_0 G^\star) = \mathrm{Tr}(C_0(G_0 + \Delta)) = \mathrm{Tr}(C_0 T)$. Now, from Lemma (22), $T \succeq 0$. Therefore, writing

$$T = \sum_i v_i v_i^T \quad (v_i \in \bar{S}),$$

we get

$$\mathrm{Tr}(C_0 T) = \sum_i v_i^T C_0 v_i \geq \mu_{d+1}(C_0) \sum_i v_i^T v_i = \mu_{d+1}(C_0)\mathrm{Tr}(T).$$

Therefore,

$$\mathrm{Tr}(T) \leq \mu_{d+1}^{-1}(C_0)\ \mathrm{Tr}(C_0 G^\star). \tag{54}$$

We are done if we can bound the term on the right. To do so, we note that

$$\mathrm{Tr}(C_0 G^\star) = \mathrm{Tr}(C_0 O^{\star\star T} O^{\star\star}) = \min_{Z\in\mathbb{R}^{Md\times N+M}} \sum_{(k,i)\in E(\Gamma)} \|Z e_{ki} - O_i^{\star\star}\bar{x}_k\|^2.$$

Therefore,

$$\mathrm{Tr}(C_0 G^\star) \leq \sum_{(k,i)\in E(\Gamma)} \|Z^{\star\star} e_{ki} - O_i^{\star\star}\bar{x}_k\|^2.$$

To bring in the error term, we write

$$Z^{\star\star} e_{ki} - O_i^{\star\star}\bar{x}_k = Z^{\star\star} e_{ki} - O_i^{\star\star} x_k^{(i)} + O_i^{\star\star}\epsilon_{k,i},$$

and use $\|x + y\|^2 \leq 2(\|x\|^2 + \|y\|^2)$ to get

$$\mathrm{Tr}(C_0 G^\star) \leq 2 \sum_{(k,i)\in E} \|Z^{\star\star} e_{ki} - O_i^{\star\star} x_k^{(i)}\|^2 + 2|E(\Gamma)|\varepsilon^2. \tag{55}$$

Finally, using the optimality of $(Z^{\star\star}, O^{\star\star})$ for (53), we have

$$\sum_{(k,i)\in E(\Gamma)} \|Z^{\star\star} e_{ki} - O_i^{\star\star} x_k^{(i)}\|^2 \leq \sum_{(k,i)\in E(\Gamma)} \|Z_0 e_{ki} - I_d x_k^{(i)}\|^2 \leq |E(\Gamma)|\varepsilon^2. \tag{56}$$

The desired result follows from (54), (55), and (56). $\square$

Finally, we note that $\mathrm{Tr}(T)$ can be used to bound the difference between the Gram matrices.

Proposition 24 (Trace bound)**.** $\|W^{\star T}W^\star - G_0\|_F \leq 2\sqrt{2Md\mathrm{Tr}(T)}$.

Proof. We will heavily use decomposition (51) and its properties. Let $G^\star = G_0 + \Delta$. By triangle inequality,

$$\begin{aligned}\|W^{\star T}W^\star - G_0\|_F &\leq \|\sum_{i=d+1}^{Md} \lambda_i(G^\star)\, u_i u_i^T \|_F + \|\Delta\|_F \\ &= \|\,\mathrm{diag}(\lambda_{d+1}(G^\star), \ldots, \lambda_{Md}(G^\star))\,\|_F + \|\Delta\|_F.\end{aligned}$$

Moreover, since the bottom eigenvalues of G_0 are zero, it follows from Lemma 21 that the norm of the diagonal matrix is bounded by $\|\Delta\|_F$. Therefore,

$$\|W^{\star T}W^\star - G_0\|_F \leq 2\|\Delta\|_F. \tag{57}$$

Fix $\{s_{d+1}, \ldots, s_{Md}\}$ to be some orthonormal basis of $\bar{S}$. For arbitrary $A \in \mathbb{R}^{Md}$, let

$$A(p,q) = s_p^T A s_q \qquad (1 \leq p, q \leq Md).$$

That is, $(A(p,q))$ are the coordinates of A in the basis $\{s_1, ..., s_d\} \cup \{s_{d+1}, \ldots, s_{Md}\}$.
Decompose $\Delta = P + Q + T$ as in (51). Note that P, Q, and T are represented in the above basis as follows: P is supported on the upper $d \times d$ diagonal block, T is supported on the lower $(M-1)d \times (M-1)d$ diagonal block, and Q on the off-diagonal blocks. The matrix G_0 is diagonal in this representation.
We can bound $\|P\|_F$ using Lemma 22,

$$\|P\|_F^2 = M^2\|P_{11}\|_F^2 = \|\sum_{l=1}^{M} T_{ll}\|_F^2 \leq \Big[\mathrm{Tr}\,\Big(\sum_{l=1}^{M} T_{ll}\Big)\Big]^2 = \mathrm{Tr}(T)^2, \tag{58}$$

where we have used the properties $T \succeq 0$ and $T_{ll} \succeq 0$ $(1 \leq l \leq M)$. In particular,

$$\|T\|_F \leq \mathrm{Tr}(T). \tag{59}$$

On the other hand, since $G_0 + \Delta \succeq 0$, we have $(G_0+\Delta)(p,q)^2 \leq (G_0+\Delta)(p,p)(G_0+\Delta)(q,q)$. Therefore,

$$\|Q\|_F^2 = 2\sum_{p=1}^{d}\sum_{q=d+1}^{Md} Q(p,q)^2 \leq 2\sum_{p=1}^{d}(G_0+\Delta)(p,p)\sum_{q=d+1}^{Md} T(q,q).$$

Notice that $0 = \mathrm{Tr}(\Delta) = \mathrm{Tr}(T) + \mathrm{Tr}(P)$. Therefore,

$$\|Q\|_F^2 \leq 2Md\mathrm{Tr}(T) - 2\mathrm{Tr}(T)^2. \tag{60}$$

Combining (57), (58), (60), and (59), we get the desired bound. □

Putting together (46) with Propositions (18),(23), and (24), we arrive at Theorem (17).

7. Numerical experiments

We now present some numerical results on multipatch registration using `GRET-SPEC` and `GRET-SDP`. In particular, we study the exact recovery and stability properties of the algorithm. We define the reconstruction error in terms of the root-mean-square deviation (RMSD) given by

$$\mathrm{RMSD} = \min_{\Omega \in \mathbb{O}(d), t \in \mathbb{R}^d} \left[\frac{1}{N}\sum_{k=1}^{N} \|Z_k^\star - \Omega\bar{x}_k - t\|^2\right]^{1/2}. \tag{61}$$

In other words, the RMSD is calculated after registering (aligning) the original and the reconstructed configurations. We use the SVD-based algorithm [2] for this purpose.

We first consider a few examples concerning the registration of three patches in $\mathbb{R}^2$, where we vary Γ by controlling the number of points in the intersection of the patches. We work with the clean data model in (25) and demonstrate exact recovery for different Γ.

In the left plot in Figure 5, we consider the patch system that was considered earlier in Figure 2. In the present case, we have $N = 10$ points. The points that belong to two or more patches are marked red, while the rest are marked black. The patches taken in the order P_1, P_2, P_3 form a lateration in this case. As predicted by Theorem 8, the rank of the patch-stress matrix C_0 for this system must be $2(3-1) = 4$. This is indeed confirmed by our experiment. We expect GRET-SPEC and GRET-SDP to recover the exact configuration. Indeed, we get a very small RMSD of the order of 1e-7 in this case. As shown in the figure, the reconstructed coordinates obtained using GRET-SDP perfectly match the original ones after alignment.

We next consider the example shown in the center plot in Figure 5. The patch system is not laterated in this case, but the rank of C_0 is 4. Again we obtain a very small RMSD of the order 1e-7 for this example. This example demonstrates that lateration is not necessary for exact recovery.

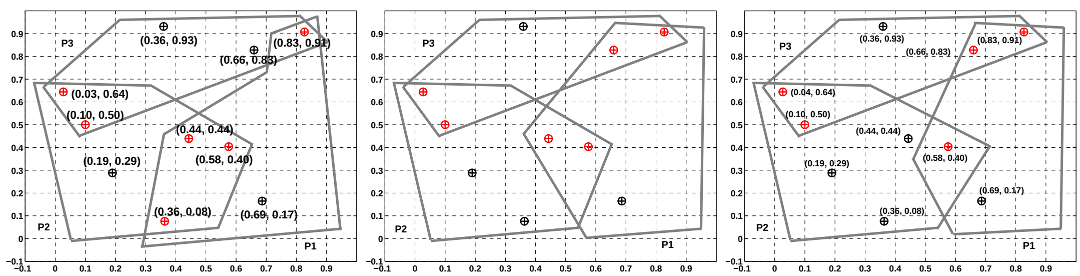

FIGURE 5. Instances of a three-patch systems in $\mathbb{R}^2$. Left: Patch system is laterated. Center: Patch system is not laterated but for which C_0 has rank 4. Right: The backbone graph is uniquely localizable but for which $\mathrm{rank}(C_0) = 3$. The original coordinates are marked with $\circ$, and the coordinates reconstructed by GRET-SDP with $+$.

In the next example, we show that the condition $\mathrm{rank}(C_0) = (M-1)d$ is not necessary for exact recovery using GRET-SDP. To do so, we use Theorem 12 which tells us that the unique localizability of the backbone graph is both necessary and sufficient for exact recovery. Consider the example shown in the right plot in Figure 5. This has barely enough points in the patch intersections to make the backbone graph uniquely localizable. Experiments confirm that we have exact recovery in this case. However, it can be shown that $\mathrm{rank}(C_0) < (M-1)d = 4$.

We now consider the structured PACM data in $\mathbb{R}^3$ shown in Figure 6. The are a total of 799 points in this example that are obtained by sampling the 3-dimensional PACM logo [16, 19]. To begin with, we divide the point cloud into $M = 30$ disjoint pieces (clusters) as shown in the figure. We augment each cluster into a patch by adding points from neighboring clusters. We ensure that there are sufficient common points in the patch system so that C_0 has rank $(M-1)d = 87$. We generate the measurements using the bounded noise model in (40). In particular, we perturb the clean coordinates using uniform noise over the hypercube $[-\varepsilon, \varepsilon]^d$. For the noiseless setting, the RMSD's obtained using using GRET-SPEC and GRET-SDP are 3.3e-11 and 1e-6. The respective RMSD's when $\varepsilon = 0.5$ are 1.4743 and 0.3823. The results are shown in Figure 7.

In the final experiment, we demonstrate the stability of GRET-SDP and GRET-SPEC by plotting the RMSD against the noise level for the PACM data. We use the noise model in (40) and vary ε from 0 to 2 in steps of 0.1. For a fixed noise level, we average the RMSD over 20 noise realizations. The results are reported in the bottom plot in Figure 8. We see that the RMSD increases gracefully with the noise level. The result also shows that the semidefinite relaxation

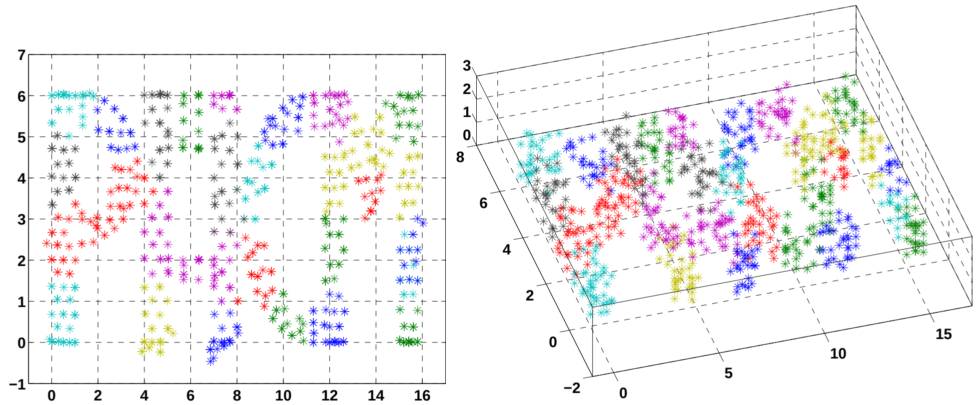

FIGURE 6. Disjoint clusters for the PACM point cloud. Each cluster is marked with a different color. The clusters are augmented to form overlapping patches which are then registered using `GRET-SDP`.

is more stable than spectral relaxation, particularly at large noise levels. Also shown in the figure are the RMSD obtained using `GRET-MANOPT` with the solutions of `GRET-SPEC` and `GRET-SDP` as initialization. In particular, we used the trust region method provided in the `Manopt` toolbox [10] for solving the manifold optimization (P_0). For either initialization, we notice some improvement from the plots. It is clear that the manifold method relies heavily on the initialization, which is not surprising.

Finally, we plot the rank of the SDP solution $G^\star$ and notice an interesting phenomenon. Up to a certain noise level, $G^\star$ has the desired rank and rounding is not required. This means that the relaxation gap is zero for the semidefinite relaxation, and that we can solve the original non-convex problem using `GRET-SDP` up to a certain noise threshold. It is therefore not surprising that the RMSD shows no improvement after we refine the SDP solution using manifold optimization. We have noticed that the rank of the SDP solution is stable with respect to noise for other numerical experiments as well (not reported here).

8. DISCUSSION

There are several directions along which the present work could be extended and refined. We summarize some of these below.

Rank recovery. Exhaustive numerical simulations (see, for example, Figure 8) show us that the proposed program is quite stable as far as rank recovery is concerned. By rank recovery, we mean that $\mathrm{rank}(G^\star) = d$. In this case, the relaxation gap is zero – we have actually solved the original non-convex problem. We have performed numerical experiments in which we fix some admissible Γ, and gradually increase the noise in the measurements as per the model in (40). When the noise is zero, we recover the exact Gram matrix that has rank d. What is interesting is that the program keeps returning a rank-d solution up to a certain noise level. In other words, we observe a *phase transition* phenomenon in which $\mathrm{rank}(G^\star)$ is consistently d up to a certain noise threshold. This threshold seems to depend on on the number of points in the intersection of the patches, which is perhaps not surprising. A precise understanding of this phase transition in terms of the properties of Γ would be interesting study.

Conditions on Γ. We established that the unique localizability of the backbone graph is both necessary and sufficient for admissibility. However, to test unique localizability, we need to run a semidefinite program [54]. The complexity of this program would, however, be much more than the program used for the registration itself. This led us to propose the rank test for admissibility that could be tested efficiently. The rank test is nonetheless not necessary

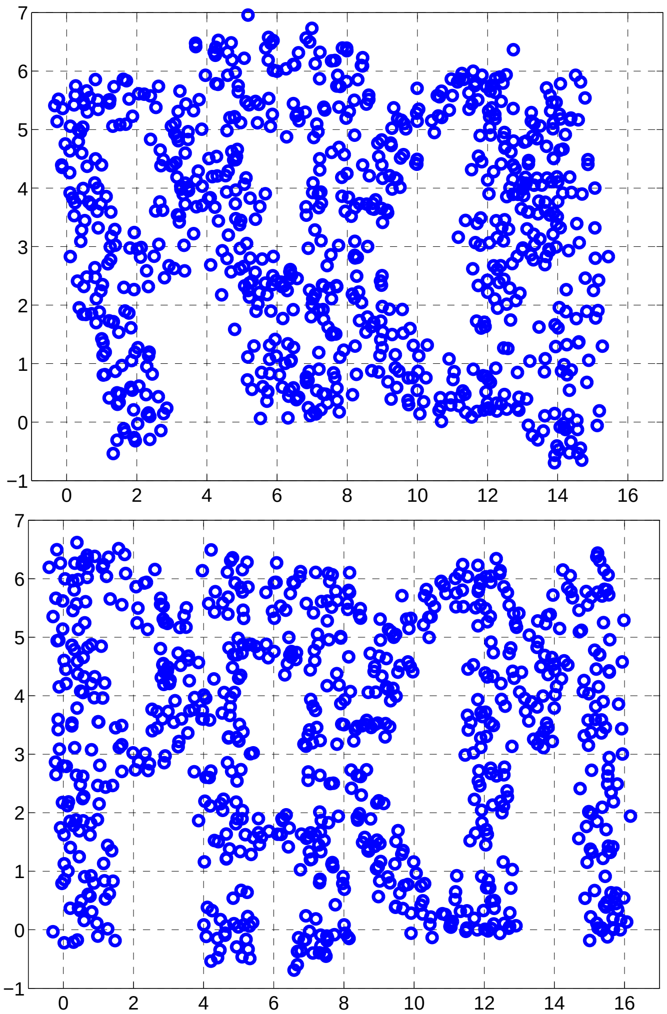

FIGURE 7. Reconstruction of the PACM data from corrupted patch coordinates ($\varepsilon = 0.5$). **Left**: GRET-SPEC, RMSD = 1.4743. **Right**: GRET-SDP, RMSD = 0.3823. The measurements were generated using the noise model in (40).

for admissibility, and weaker admissibility conditions on Γ can be found. In particular, an interesting question is whether we could find an efficiently-testable admissibility condition that holds true for the extreme example in Figure 5, in which Γ fails the rank test?

Tighter bounds. The stability in Theorem 17 was for the bounded noise model, which made the subsequent analysis quite straightforward. The goal was to establish that the reconstruction error is within $C\varepsilon$ for some constant C independent of the noise. In particular, the bounds in Theorem 17 are quite loose. One possible direction would be to consider a stochastic noise model with statistically independent perturbations to tighten the bound.

Anchor points. In sensor network localization, one has to infer the coordinates of sensors from the knowledge of distances between sensors and its geometric neighbors. In distributed approaches to sensor localization [15, 9], one is faced exactly with the multipatch registration problem described in this paper. Besides the distance information, one often has the added knowledge of the precise positions of selected sensors known as *anchors* [8]. This is often by design and is used to improve the localization accuracy. The question is can we incorporate the anchor constraints into the present registration algorithm? One possible way of leveraging the

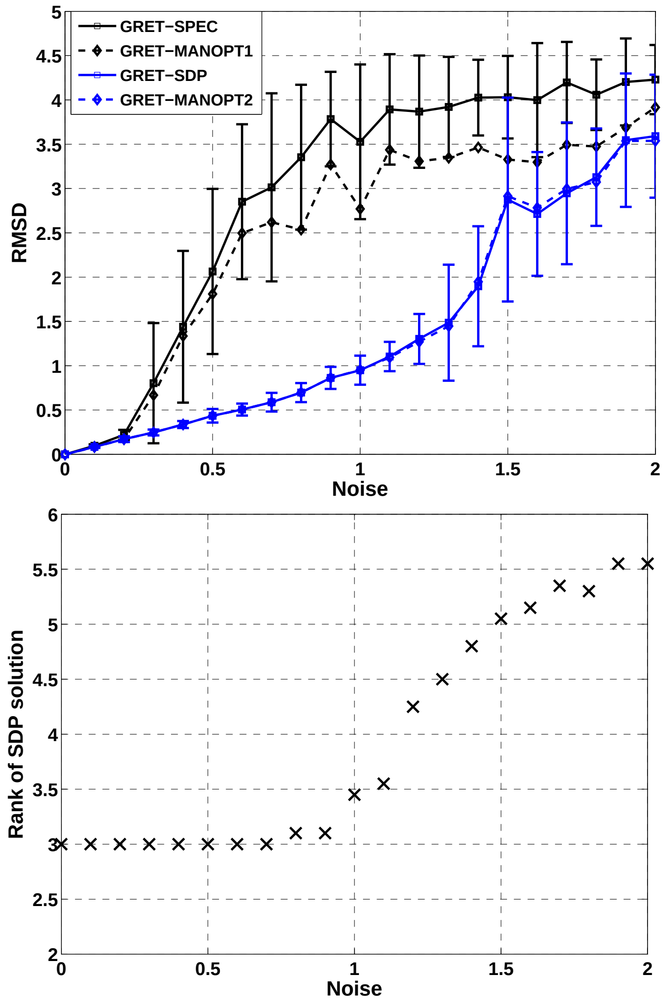

FIGURE 8. **Left**: RMSD versus noise level. `GRET-MANOPT1` (resp. `GRET-MANOPT2`) is the result obtained by refining the output of `GRET-SPEC` (resp. `GRET-SDP`) using manifold optimization. **Right**: Rank of $G^\star$ in `GRET-SDP`.

existing framework is to introduce an additional patch (called *anchor patch*) for the anchor points. The anchor coordinates are assigned to the points in the anchor patch (treating them as local coordinates). This gives us an augmented bipartite graph Γ_a which has one more patch vertex than Γ, and extra edges connecting the anchor patch to the anchor vertices. We then proceed exactly as before, that is, we solve for the global coordinates of both the anchor and non-anchor points given the measurements on Γ_a.

ACKNOWLEDGEMENT

Kunal Chaudhury was supported by the Swiss National Science Foundation under Grant PBELP2-135867, and by Award Number R01GM090200 from the National Institute of General Medical Sciences.

Yuehaw Khoo was partially supported by Award Number R01 GM090200-01 from the National Institute for Health.

Amit Singer was partially supported by Award Numbers FA9550-13-1-0076 and FA9550-12-1-0317 from AFOSR, Award Number R01GM090200 from the National Institute of General Medical Sciences, and Award Number LTR DTD 06-05-2012 from the Simons Foundation.

The authors wish to thank Mihai Cucuringu, Lanhui Wang, and Afonso Bandeira for useful discussions, and Nicolas Boumal for advice on the usage of the `Manopt` toolbox.

References

1. P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2009.
2. K. S. Arun, T. S. Huang, and S. D. Blostein, *Least-squares fitting of two 3d point sets*, IEEE Transactions on Pattern Analysis and Machine Intelligence (1987), no. 5, 698–700.
3. A. S. Bandeira, , A. Singer, and D. A. Spielman, *A cheeger inequality for the graph connection laplacian*, arXiv:1204.3873 (2012).
4. A. S. Bandeira, C. Kennedy, and A. Singer, *Approximating the little Grothendieck problem over the orthogonal group*, arXiv:1308.5207 (2013).
5. S. R. Becker, E. J. Candès, and M. C. Grant, *Templates for convex cone problems with applications to sparse signal recovery*, Mathematical Programming Computation **3** (2011), no. 3, 165–218.
6. P. J. Besl and N. D. McKay, *A method for registration of 3d shapes*, IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992), no. 2, 239–256.
7. R. Bhatia, *Matrix Analysis*, vol. 169, Springer, 1997.
8. P. Biswas, T.-C. Liang, K.-C. Toh, Y. Ye, and T.-C. Wang, *Semidefinite programming approaches for sensor network localization with noisy distance measurements*, IEEE Transactions on Automation Science and Engineering **3** (2006), no. 4, 360–371.
9. P. Biswas, K.-C. Toh, and Y. Ye, *A distributed SDP approach for large-scale noisy anchor-free graph realization with applications to molecular conformation*, SIAM Journal on Scientific Computing **30** (2008), no. 3, 1251–1277.
10. N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, *Manopt: a matlab toolbox for optimization on manifolds*, arXiv:1308.5200 (2013).
11. S. Burer and R. D. C. Monteiro, *A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, Mathematical Programming **95** (2003), no. 2, 329–357.
12. E. Candes and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics **9** (2009), no. 6, 717–772.
13. E. J. Candes, T. Strohmer, and V. Voroninski, *Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming*, Communications on Pure and Applied Mathematics **66** (2013), no. 8, 1241–1274.
14. F. R. K. Chung, *Spectral Graph Theory*, vol. 92, American Mathematical Society, 1997.
15. M. Cucuringu, Y. Lipman, and A. Singer, *Sensor network localization by eigenvector synchronization over the euclidean group*, ACM Transactions on Sensor Networks **8** (2012), no. 3, 19.
16. M. Cucuringu, A. Singer, and D. Cowburn, *Eigenvector synchronization, graph rigidity and the molecule problem*, Information and Inference **1** (2012), no. 1, 21–67.
17. A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM Journal on Matrix Analysis and Applications **20** (1998), no. 2, 303–353.
18. K. Fan and A. J. Hoffman, *Some metric inequalities in the space of matrices*, Proceedings of the American Mathematical Society **6** (1955), no. 1, 111–116.
19. X. Fang and K.-C. Toh, *Using a distributed SDP approach to solve simulated protein molecular conformation problems*, Distance Geometry, Springer, 2013, pp. 351–376.
20. O. D. Faugeras and M. Hebert, *The representation, recognition, and locating of 3d objects*, International Journal of Robotics Research **5** (1986), no. 3, 27–52.
21. D. Garber and E. Hazan, *Approximating semidefinite programs in sublinear time*, Advances in Neural Information Processing Systems, 2011, pp. 1080–1088.
22. M. X. Goemans and D. P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, Journal of the ACM **42** (1995), no. 6, 1115–1145.
23. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1996.
24. S. J. Gortler, A. D. Healy, and D. P. Thurston, *Characterizing generic global rigidity*, American Journal of Mathematics **132** (2010), no. 4, 897–939.
25. J. C. Gower and G. B. Dijksterhuis, *Procrustes Problems*, vol. 3, Oxford University Press Oxford, 2004.
26. M. Grant, S. Boyd, and Y. Ye, *CVX: Matlab software for disciplined convex programming*, 2008.
27. C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz, *An interior-point method for semidefinite programming*, SIAM Journal on Optimization **6** (1996), no. 2, 342–361.
28. B. Hendrickson, *Conditions for unique graph realizations*, SIAM Journal on Computing **21** (1992), no. 1, 65–84.
29. J. N. Higham, *Computing the polar decomposition - with applications*, SIAM Journal on Scientific and Statistical Computing **7** (1986), no. 4, 1160–1174.

30. N.-D. Ho and P. Van Dooren, *On the pseudo-inverse of the Laplacian of a bipartite graph*, Applied Mathematics Letters **18** (2005), no. 8, 917–922.
31. B. K. P. Horn, *Closed-form solution of absolute orientation using unit quaternions*, JOSA A **4** (1987), no. 4, 629–642.
32. S. D. Howard, D. Cochran, W. Moran, and F. R. Cohen, *Estimation and registration on graphs*, arXiv:1010.2983 (2010).
33. Qi-Xing Huang and Leonidas Guibas, *Consistent shape maps via semidefinite programming*, Eurographics Symposium on Geometry Processing, vol. 32, 2013, pp. 177–186.
34. A. Javanmard and A. Montanari, *Localization from incomplete noisy distance measurements*, Foundations of Computational Mathematics **13** (2013), no. 3, 297–345.
35. M. Journee, F. Bach, P.-A. Absil, and R. Sepulchre, *Low-rank optimization on the cone of positive semidefinite matrices*, SIAM Journal on Optimization **20** (2010), no. 5, 2327–2351.
36. J. B. Keller, *Closest unitary, orthogonal and hermitian operators to a given operator*, Mathematics Magazine **48** (1975), no. 4, 192–197.
37. E. De Klerk, *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*, vol. 65, Kluwer Academic Publications, 2002.
38. S. Krishnan, P. Y. Lee, J. B. Moore, and S. Venkatasubramanian, *Global registration of multiple 3d point sets via optimization-on-a-manifold*, Eurographics Symposium on Geometry Processing (2005), 187–197.
39. G. Lerman, M. McCoy, J. A. Tropp, and T. Zhang, *Robust computation of linear models, or how to find a needle in a haystack*, arXiv:1202.4044 (2012).
40. R.-C. Li, *New perturbation bounds for the unitary polar factor*, SIAM Journal on Matrix Analysis and Applications **16** (1995), no. 1, 327–332.
41. L. Lovász and A. Schrijver, *Cones of matrices and set-functions and 0-1 optimization*, SIAM Journal on Optimization **1** (1991), no. 2, 166–190.
42. L. Mirsky, *Symmetric gauge functions and unitarily invariant norms*, The Quarterly Journal of Mathematics **11** (1960), no. 1, 50–59.
43. N. J. Mitra, N. Gelfand, H. Pottmann, and L. Guibas, *Registration of point cloud data from a geometric optimization perspective*, Eurographics Symposium on Geometry Processing, 2004, pp. 22–31.
44. A. Naor, O. Regev, and T. Vidick, *Efficient rounding for the noncommutative rothendieck inequality*, Proceedings of the 45th Annual ACM Symposium on Theory of Computing, 2013, pp. 71–80.
45. A. Nemirovski, *Sums of random symmetric matrices and quadratic optimization under orthogonality constraints*, Mathematical Programming **109** (2007), no. 2-3, 283–317.
46. Y. Nesterov, *Semidefinite relaxation and nonconvex quadratic optimization*, Optimization methods and software **9** (1998), no. 1-3, 141–160.
47. H. Pottmann, Q.-X. Huang, Y.-L. Yang, and S.-M. Hu, *Geometry and convergence analysis of algorithms for registration of 3d shapes*, International Journal of Computer Vision **67** (2006), no. 3, 277–296.
48. G. Ranjan, Z.-L. Zhang, and D. Boley, *Incremental computation of pseudo-inverse of Laplacian: Theory and applications*, arXiv:1304.2300 (2013).
49. S. Rusinkiewicz and M. Levoy, *Efficient variants of the ICP algorithm*, Proceedings of the Third International Conference on 3d Digital Imaging and Modeling, 2001, pp. 145–152.
50. G. C. Sharp, S. W. Lee, and D. K. Wehe, *Multiview registration of 3d scenes by minimizing error between coordinate frames*, Proceedings of the 7th European Conference on Computer Vision-Part II, Springer-Verlag, 2002, pp. 587–597.
51. A. Singer, *Angular synchronization by eigenvectors and semidefinite programming*, Applied and Computational Harmonic Analysis **30** (2011), no. 1, 20–36.
52. A. Singer and M. Cucuringu, *Uniqueness of low-rank matrix completion by rigidity theory*, SIAM Journal on Matrix Analysis and Applications **31(4)** (2010), 1621–1641.
53. A. M.-C. So, *Moment inequalities for sums of random matrices and their applications in optimization*, Mathematical Programming **130** (2011), no. 1, 125–151.
54. A. M.-C. So and Y. Ye, *Theory of semidefinite programming for sensor network localization*, Mathematical Programming **109** (2007), no. 2-3, 367–384.
55. D. A. Spielman and S.-H. Teng, *Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems*, Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing, 2004, pp. 81–90.
56. K.-C. Toh, M. J. Todd, and R. H. Tutuncu, *SDPT3 – A matlab software package for semidefinite programming, version 1.3*, Optimization Methods and Software **11** (1999), no. 1-4, 545–581.
57. T. Tzeneva, *Global alignment of multiple 3d scans using eigevector synchronization*, Senior Thesis, Princeton University (supervised by S. Rusinkiewicz and A. Singer) (2011).
58. L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM review **38** (1996), no. 1, 49–95.
59. N. K. Vishnoi, *Lx=b*, Foundations and Trends in Theoretical Computer Science **8** (2012), no. 1-2, 1–141.
60. Irène Waldspurger, Alexandre d'Aspremont, and Stéphane Mallat, *Phase recovery, maxcut and complex semidefinite programming*, arXiv:1206.0102 (2012).
61. L. Wang and A. Singer, *Exact and stable recovery of rotations for robust synchronization*, Information and Inference: A Journal of the IMA, accepted for publication (2013).
62. Z. Wen, D. Goldfarb, S. Ma, and K. Scheinberg, *Block coordinate descent methods for semidefinite programming*, Handbook on Semidefinite, Conic and Polynomial Optimization, Springer, 2012, pp. 533–564.

63. J. A. Williams and M. Bennamoun, *Simultaneous registration of multiple point sets using orthonormal matrices*, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 6, 2000, pp. 2199–2202.
64. H. Wolkowicz and M. F. Anjos, *Semidefinite programming for discrete optimization and matrix completion problems*, Discrete Applied Mathematics **123** (2002), no. 1, 513–577.
65. H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, vol. 27, Kluwer Academic Pub, 2000.
66. L. Zhang, L. Liu, C. Gotsman, and S. Gortler, *An As-Rigid-As-Possible approach to sensor network localization*, ACM Transactions on Sensor Networks **6** (2010), no. 4, 35.
67. L. Zhi-Quan, M. Wing-Kin, A.M.-C. So, Y. Ye, and S. Zhang, *Semidefinite relaxation of quadratic optimization problems*, IEEE Signal Processing Magazine **27** (2010), no. 3, 20–34.

9. TECHNICAL PROOFS

In this Section, we give the proof of Lemmas 10, 19, and 20, and Propositions 15 and 18.

9.1. **Proof of Lemma 10.** Fix patch P_i and (without loss of generality) assume that the labels of the $d+1$ non-degenerate points in P_i are $\{1, 2, \ldots, d+1\}$. From (28),

$$x_k = F_i x_k^{(i)} + t_i \qquad (k = 1, 2, \ldots, d+1).$$

We fix $k = 1$ and subtract the remaining d equations from it, to get

$$F_i \big[x_2^{(i)} - x_1^{(i)} \ \ldots \ x_{d+1}^{(i)} - x_1^{(i)} \big] = \big[x_2 - x_1 \ \ldots \ x_{d+1} - x_1 \big].$$

From the non-degeneracy assumption, it follows that the matrix on the left is invertible, so that the span of F_i is identical to the affine span of $\{x_1, \ldots, x_{d+1}\}$. Applying this observation to every patch $P_1, \ldots, P_M$, we conclude that the span of $[F_1 \ \cdots \ F_M]$ is contained in the affine span of $\{x_1, \ldots, x_N\}$. Note that we did not require the connectivity of Γ for this direction.

The reverse inclusion can be be deduced from the assumed connectivity of Γ. Namely, for every pair of points taken from different patches, there exists a path in Γ connecting these points. Note that it suffices to show that for arbitrary x_k and x_l, the difference $x_k - x_l$ is in the span of $[F_1 \cdots F_M]$. To avoid notational complications, we consider the case where we have three patches P_1, P_2 and P_3. Consider points x_k and x_l belonging respectively to P_1 and P_3. If P_1 and P_3 have a common point, say x_p, then we write

$$x_k - x_l = (x_k - x_p) + (x_p - x_l)$$

From (28), it follows that $x_k - x_p$ is in the span of F_1, and $x_p - x_l$ is in the span of F_3. The desired conclusion then follows from the above decomposition.

On the other hand, it is possible that P_1 and P_3 have no points in common. However, since Γ is connected, there must be points in P_2, say x_p and x_q, such that

$$x_k, x_p \in P_1 \quad \text{and} \quad x_q, x_l \in P_3.$$

Pick some point x_r in P_2 different from x_p and x_q, and write

$$x_k - x_l = (x_k - x_p) + (x_p - x_r) + (x_r - x_q) + (x_q - x_l).$$

From (28), the first term is in the span of F_1, the middle terms are in the span of F_2, while the last term is in the span of F_3. Therefore, $x_k - x_l$ is in the span of $[F_1 \ F_2 \ F_3]$.

The strategy for the general proof is now clear, namely, we have to use the connectivity assumption to write $x_k - x_l$ as a chain of differences, where each term in the chain belongs to a single patch.

9.2. **Proof of Proposition 15.** We are done if we can show that there exists a bijection between the nullspace of C and that of C_0. To do so, we note that the associated quadratic forms can be expressed as

$$u^T C u = \min_{z \in \mathbb{R}^{1\times N+M}} \sum_{(k,i)\in E(\Gamma)} \|z e_{ki} - u_i^T x_k^{(i)}\|^2,$$

and

$$v^T C_0 v = \min_{z \in \mathbb{R}^{1\times N+M}} \sum_{(k,i)\in E(\Gamma)} \|z e_{ki} - v_i^T \bar{x}_k\|^2.$$

Here $u_1, \ldots, u_M$ are the $d \times 1$ blocks of the vector $u \in \mathbb{R}^{Md\times 1}$.

Now, it follows from (25) that there is a one-to-one correspondence between u and v, namely

$$u_i = \bar{O}_i v_i \qquad (1 \leq i \leq M),$$

such that $u^T C u = v^T C_0 v$. In other words, the null space of C is related to the null space of C_0 through an orthogonal transform, as was required to be shown.

9.3. **Proof of Proposition 18.** Without loss of generality, we assume that the smallest Euclidean ball that encloses the clean configuration $\{\bar{x}_1, \ldots, \bar{x}_N\}$ is centered at the origin, that is,

$$\|\bar{x}_k\| \leq R \qquad (1 \leq k \leq N). \tag{62}$$

Let B_0 be the matrix B in (12) computed from the clean measurements, i.e., from (40) with $\varepsilon = 0$. Let $B_0 + H$ be the same matrix obtained from (40) for some $\varepsilon > 0$.

Recall that $Z_0 = O_0 B_0 L^\dagger$ (by the centering assumption in (41)). Therefore,

$$\|Z^\star - \Theta Z_0\|_F = \|O^\star (B_0 + H) L^\dagger - \Theta O_0 B_0 L^\dagger\|_F = \|(O^\star - \Theta O_0) B_0 L^\dagger + O^\star H L^\dagger\|_F.$$

By triangle inequality,

$$\|Z^\star - \Theta Z_0\|_F \leq \|O^\star - \Theta O_0\|_F \|B_0 L^\dagger\|_F + \|O^\star H L^\dagger\|_F, \tag{63}$$

Now

$$\|B_0 L^\dagger\|_F \leq \|L^\dagger\|_{\mathrm{sp}} \|B_0\|_F = \frac{1}{\lambda_2(L)} \|B_0\|_F,$$

where $\lambda_2(L)$ is the smallest non-zero eigenvalue of L. On the other hand,

$$B_0 = \sum_{(k,i)\in E(\Gamma)} (e_i^M \otimes I_d) \bar{x}_k e_{ki}^T.$$

Using Cauchy-Schwarz and (62), we get

$$\begin{aligned}
\|B_0\|_F^2 &= \sum_{(k,i)\in E(\Gamma)} \sum_{(l,j)\in E(\Gamma)} \mathrm{Tr}\left(e_{ki} \bar{x}_k^T (e_i^M \otimes I_d)^T (e_j^M \otimes I_d) \bar{x}_l e_{lj}^T\right) \\
&= \sum_{(k,i)\in E(\Gamma)} \sum_{(l,i)\in E(\Gamma)} \bar{x}_k^T \bar{x}_l \; e_{ki}^T e_{li}. \\
&\leq \sum_{(k,i)\in E(\Gamma)} 2R^2 + \sum_{(k,i)\in E(\Gamma)} \sum_{(l,i)\in E(\Gamma)} R^2.
\end{aligned}$$

Therefore,

$$\|B_0 L^\dagger\|_F \leq \lambda_2(L)^{-1} \sqrt{2+N} |E(\Gamma)|^{1/2} R. \tag{64}$$

As for the other term in (63), we can write

$$\|O^\star H L^\dagger\|_F \leq \|L^\dagger\|_2 \|O^\star H\|_F \leq \lambda_2(L)^{-1} \|O^\star H\|_F.$$

Now

$$O^\star H = O^\star (B - B_0) = \sum_{(k,i)\in E(\Gamma)} O_i^\star \epsilon_{k,i} e_{ki}^T.$$

Therefore, using Cauchy-Schwarz, the orthonormality of the columns of $O_i^\star$'s, and the noise model (40), we get

$$\begin{aligned}\|O^\star H\|_F^2 &= \sum_{(k,i)\in E(\Gamma)} \sum_{(l,j)\in E(\Gamma)} (O_i^\star \epsilon_{k,i})^T (O_j^\star \epsilon_{l,j}) e_{ki}^T e_{li} \\ &\leq \sum_{(k,i)\in E(\Gamma)} 2\varepsilon^2 + \sum_{(k,i)\in E(\Gamma)} \sum_{(l,i)\in E(\Gamma)} \varepsilon^2 + \sum_{(k,i)\in E(\Gamma)} \sum_{(k,j)\in E(\Gamma)} \varepsilon^2.\end{aligned}$$

This gives us

$$\|O^\star H L^\dagger\|_F \leq \sqrt{2+N+M}\,|E(\Gamma)|^{1/2} \lambda_2(L)^{-1} \varepsilon. \tag{65}$$

Combining (63),(64), and (65), we get the desired estimate.

9.4. **Proof of Lemma 19.** The proof is mainly based on the observation that if u and v are unit vectors and $0 \leq u^T v \leq 1$, then

$$\|u - v\| \leq \|uu^T - vv^T\|_F. \tag{66}$$

Indeed,

$$\|uu^T - vv^T\|_F^2 = \mathrm{Tr}\left(uu^T + vv^T - 2(u^T v)^2\right) \geq \mathrm{Tr}(uu^T + vv^T - 2u^T v) = \|u - v\|^2.$$

To use this result in the present setting, we use the theory of principal angles [7, Ch. 7.1]. This tells us that, for the orthonormal systems $\{u_1, \ldots, u_d\}$ and $\{s_1, \ldots, s_d\}$, we can find $\Omega_1, \Omega_2 \in \mathbb{O}(Md)$ such that

(1) $\Omega_1[u_1 \cdots u_d] = [u_1 \cdots u_d]\Theta_1^T$ where $\Theta_1 \in \mathbb{O}(d)$,
(2) $\Omega_2[s_1 \cdots s_d] = [s_1 \cdots s_d]\Theta_2^T$ where $\Theta_2 \in \mathbb{O}(d)$,
(3) $(\Omega_1 s_i)^T(\Omega_2 u_j) = 0$ for $i \neq j$, and $0 \leq (\Omega_1 s_i)^T(\Omega_2 u_i) \leq 1$ for $1 \leq i \leq d$.

Here Θ_1 and Θ_2 are the orthogonal transforms that map $\{u_1, \ldots, u_d\}$ and $\{s_1, \ldots, s_d\}$ into the corresponding principal vectors.

Using properties 1 and 2 and the fact[3] that $\alpha_i \leq M$, we can write

$$\sqrt{M}\ \|\Theta_1 W^\star - \Theta_2 O_0\|_F \leq \|\Omega_1[\alpha_1 u_1 \cdots \alpha_d u_d] - M\Omega_2[s_1 \cdots s_d]\|_F + \Big[\sum_{i=1}^d (M - \alpha_i)^2\Big]^{1/2}.$$

Moreover, by triangle inequality,

$$\|\Omega_1[\alpha_1 u_1 \cdots \alpha_d u_d] - M\Omega_2[s_1 \cdots s_d]\|_F \leq M\|\Omega_1[u_1 \cdots u_d] - \Omega_2[s_1 \cdots s_d]\|_F + \Big[\sum_{i=1}^d (M - \alpha_i)^2\Big]^{1/2}.$$

Therefore,

$$\sqrt{M}\ \|\Theta_1 W^\star - \Theta_2 O_0\|_F \leq M\|\Omega_1[u_1 \cdots u_d] - \Omega_2[s_1 \cdots s_d]\|_F + \Big[\sum_{i=1}^d (M - \alpha_i)^2\Big]^{1/2}. \tag{67}$$

Now, using (66) and the principal angle property 3, we get

$$\|\Omega_1[u_1 \cdots u_d] - \Omega_2[s_1 \cdots s_d]\|_F \leq \|\sum_{i=1}^d \Omega_1 u_i(\Omega_1 u_i)^T - \sum_{i=1}^d \Omega_2 s_i(\Omega_2 s_i)^T\|_F.$$

[3]To see why the eigenvalues of $G^\star$ are at most M (the authors thank Afonso Bandeira for suggesting this), note that by the SDP constraints, for every block G_{ij},

$$u^T G_{ij} v \leq (\|u\|^2 + \|v\|^2)/2 \quad (u, v \in \mathbb{R}^d).$$

Let $x = (x_1, \ldots, x_M)$ where each $x_i \in \mathbb{R}^d$. Then

$$x^T G x = \sum_{i,j} x_i^T G_{ij} x_j \leq \sum_{i,j} (\|x_i\|^2 + \|x_j\|^2)/2 = M\|x\|^2.$$

Moreover, using triangle inequality and properties 1 and 2, we have

$$M\| \sum_{i=1}^d \Omega_1 u_i(\Omega_1 u_i)^T - \sum_{i=1}^d \Omega_2 s_i(\Omega_2 s_i)^T\|_F \leq \|W^{\star T}W^\star - G_0\|_F + \Big[\sum_{i=1}^d (M-\alpha_i)^2\Big]^{1/2}.$$

That is,

$$\|\Omega_1[u_1\cdots u_d] - \Omega_2[s_1\cdots s_d]\|_F \leq \|W^{\star T}W^\star - G_0\|_F + \Big[\sum_{i=1}^d (M-\alpha_i)^2\Big]^{1/2}. \tag{68}$$

Finally, note that by Lemma 21,

$$\Big[\sum_{i=1}^d (M-\alpha_i)^2\Big]^{1/2} \leq \|W^{\star T}W^\star - G_0\|_F. \tag{69}$$

Combining (67), (68), and (69), and setting $\Theta = \Theta_1^T\Theta_2$, we arrive at Lemma 19.

9.5. **Proof of Lemma 20.** This is done by adapting the following result by Li [40]: If A, B are square and non-singular, and if $\mathcal{R}(A)$ and $\mathcal{R}(B)$ are their orthogonal rounding (obtained from their polar decompositions [29]), then

$$\|\mathcal{R}(A) - \mathcal{R}(B)\|_F \leq \frac{2}{\sigma_{\min}(A) + \sigma_{\min}(B)}\|A - B\|_F. \tag{70}$$

We recall that if $A = U\Sigma V^T$ is the SVD of A, then $\mathcal{R}(A) = UV^T$.

Note that it is possible that some of the blocks of $W^\star$ are singular, for which the above result does not hold. However, the number of such blocks can be controlled by the global error. More precisely, let $\mathcal{B} \subset \{1, 2, \ldots, M\}$ be the index set such that, for $i \in \mathcal{B}$, $\|W_i^\star - \Theta\|_F \geq \beta$. Then

$$\|W^\star - \Theta O_0\|_F^2 \geq \sum_{i\in\mathcal{B}} \|W_i^\star - \Theta\|_F^2 = |\mathcal{B}|\beta^2.$$

This gives a bound on the size of $\mathcal{B}$. In particular, the rounding error for this set can trivially be bounded as

$$\sum_{i\in\mathcal{B}} \|O_i^\star - \Theta\|_F^2 \leq \sum_{i\in\mathcal{B}} 2d = \frac{2d}{\beta^2}\|W^\star - \Theta O_0\|_F^2. \tag{71}$$

On the other hand, we known that, for $i \in \mathcal{B}^c$, $\|W_i^\star - \Theta\|_F < \beta$. From Lemma 21, it follows that

$$|1 - \sigma_{\min}(W_i^\star)| \leq \|W_i^\star - \Theta\|_{\mathrm{sp}} < \beta.$$

Fix $\beta \leq 1$. Then $\sigma_{\min}(W_i^\star) > 1 - \beta$, and we have from (70),

$$\|O_i^\star - \Theta\|_F \leq \frac{2}{2-\beta}\|W_i^\star - \Theta\|_F \qquad (i \in \mathcal{B}^c) \tag{72}$$

Fixing $\beta = 1/\sqrt{2}$ and combining (71) and (72), we get the desired bound.

PROGRAM IN APPLIED AND COMPUTATIONAL MATHEMATICS (PACM), PRINCETON UNIVERSITY, PRINCETON, NJ 08540

E-mail address: kchaudhu@math.princeton.edu

DEPARTMENT OF PHYSICS, PRINCETON UNIVERSITY, PRINCETON, NJ 08540

E-mail address: ykhoo@princeton.edu

DEPARTMENT OF MATHEMATICS AND PACM, PRINCETON UNIVERSITY, PRINCETON, NJ 08540

E-mail address: amits@math.princeton.edu