Synchronization over Cartan motion groups via contraction

Onur Özyeşil∗, Nir Sharon†, and Amit Singer‡1,2

1Program in Applied and Computational Mathematics, Princeton University
2Department of Mathematics, Princeton University

Abstract

Group contraction is an algebraic map that relates two classes of Lie groups by a limiting process. We utilize this notion for the compactification of the class of Cartan motion groups. The compactification process is then applied to reduce a non-compact synchronization problem to a problem where the solution can be obtained by means of a unitary, faithful representation. We describe this method of synchronization via contraction in detail and analyze several important aspects of this application. One important special case of Cartan motion groups is the group of rigid motions, also called the special Euclidean group. We thoroughly discuss the synchronization over this group and show numerically the advantages of our approach compared to some current state-of-the-art synchronization methods on both synthetic and real data.

Key Words. Synchronization, Cartan motion groups, group contraction, special Euclidean group

1 Introduction

The synchronization (also known as registration or averaging) problem over a group is to estimate \( n \) unknown group elements \( \{g_i\}_{i=1}^n \) from a set of measurements \( g_{ij} \), potentially partial and noisy, of their ratios \( g_i g_j^{-1} \), \( 1 \leq i < j \leq n \). The solution of this problem is not unique as the data is invariant to right action of an arbitrary group element \( g \) (also termed global alignment) since \( g_i g_j^{-1} = (g_i g)(g_j g)^{-1} \).

When all measurements are exact and noiseless, it is possible to solve the above synchronization problem as follows. Define an undirected graph \( G = (V, E) \) with the set of vertices \( V \) corresponding to group elements \( \{g_i\}_{i=1}^n \) and the edges representing the measurements. Then, if the graph is connected the elements are determined by traversing a spanning tree, up to an arbitrary group element in the root of the tree. Nevertheless, in presence of noise, this solution accumulates errors while traversing the spanning tree. Since measurement errors naturally arise in real world problems, such as in dynamical system data [53], in networks time distribution [23], or in Cryo-EM imaging [51], we henceforth focus on addressing the synchronization problem with noisy measurements.

We roughly divide the synchronization problem into two main cases, corresponding to two types of groups: compact groups and non-compact groups. In the case of compact group, there exist a faithful,
finite, and orthogonal representation of the group (that is a injective map to orthogonal elements of a finite dimension linear space). Therefore, the problem can be reduced to one in linear algebra. For example, in [50], two main approaches are suggested: to define a proper (kernel) matrix and extract its (top) eigenvectors, or to solve an appropriate optimization problem. An extensive survey is given in Subsection [2.1]. While the study of synchronization over compact group has resulted in many solutions, with a special attention to the special orthogonal groups, e.g., [8], the case of non-compact groups has not received as much attention. A fundamental difficulty in the non-compact case is that many of the methods for compact groups are no more valid in their original form as there is no faithful, orthogonal representation available in a finite linear space. In addition, the Riemannian structure is more complicated.

One important class of non-compact groups is the Cartan motion groups. These groups have the structure of semidirect product $K \ltimes V$, where $K$ is a compact group and $V$ is a linear space (and thus the non-compactness). One important special case from the application standpoint is the special Euclidean group, denoted by SE$(d)$, and consists of oriented rotations and translations in $d$-dimensional Euclidean space, that is $\text{SE}(d) = \text{SO}(d) \ltimes \mathbb{R}^d$. This group, also known as the group of rigid motions, is ubiquitous in various research fields, and problems related to synchronization over this group appear in many areas, for example in sensor network localization [15, 42], object alignment in computer vision [6], estimation of location and orientation of mobile robots [44], the fresco reconstruction problem [10] and more.

Synchronization over $\text{SE}(d)$ has gained considerable attention lately, see e.g., [42, 44, 56]. Most work is concentrated on applications in $d = 2$ and $d = 3$, but some also refer to general $d$. One such general approach is to form a solution separately in the rotational part and the translational part of the data, see e.g., [32]. The linear part in those approaches is solved by least squares or consistency search. The consistency of the data in synchronization is the relation $g_{ij}g_{j\ell} = g_{i\ell}$ for noiseless measurements. Other solutions for arbitrary $d$ generalize the approaches in [50], either directly based on spectral approach [3, 6] or by using modified optimization formulations and different relaxations to solve them, e.g., [40]. There is a variety of solutions for the cases $d = 2$ and $d = 3$, as rigid motion synchronization can facilitate solving several classical problems such as Structure from Motion (SfM) in computer vision and pose graph optimization in robotics. Furthermore, some mathematical relations are only valid for these cases of low dimension, for example the tight connection between SE$(3)$ and dual quaternions, which is used for synchronization in [53] to solve registration problems in computer graphics. A more comprehensive survey, including these cases and more, is given in Subsection [2.2]. Our approach to address the problem of synchronization over SE$(d)$ stems from the algebraic structure of Cartan motion groups and their relation to compact groups. Specifically, we use the notion of group contraction.

The seminal paper [31] introduced the notion of contraction of a Lie group. The motivation was to link two theories in physics which are related via a limiting process. The results show algebraically how the associated groups (and their representations) are linked, for example, the Galiliean group and the Poincaré group. We use a specific family of contractions, presented in [16], which is based upon Cartan decomposition of groups, and relates between a group and its associated Cartan motion group. For example, in the case of the special Euclidean group, this contraction maps $\text{SE}(d)$ to $\text{SO}(d + 1)$. These maps facilitate rewriting the synchronization over the Cartan motion group in terms of synchronization problem in its associated group. When the associated group is compact it means one can approximate the solutions of the original synchronization by solving synchronization over the compact group using
any existing tools available. The procedure of reducing the problem by means of compact groups is also termed compactification.

The compactification implies a solution in the compact domain, which needs then to be mapped back to the original domain. The invertibility is not the only requirement on the map as we also want to guarantee the quality of the solution. To this end we impose an additional algebraic condition on the maps which we term approximated homomorphism. This condition enables to relate the two problems, the compact and the non-compact one. Also, it gives rise for other variants for compactification; we provide one such example for the special case of SE(d), where we use an invertible map for projecting data from SE(d) to SO(d + 1). In contrary to contractions which are based on group level decomposition, this variant is based upon matrix decomposition over the group representations.

Our main focus is on the application of the compactification methods to synchronization. Thus, we present a full, detailed algorithm of synchronization via contraction and analyze several important aspects of it. One such aspect is the invariance of synchronization to global alignment. Apparently, this invariant does not directly pass from the compact synchronization to the non-compact problem. We accurately characterize the correct invariance that the two problems do share, and based on that suggest a method to choose the right global alignment in the solution of the compact synchronization for maximizing the quality of solution in the non-compact side, after back mapping the solution. We also provide some analysis for the case of noisy data. In particular, we show a condition that guarantees, under certain assumptions on the noise, that meaningful information can be extracted to form a solution for synchronization.

We conclude the paper with a numerical part, where we compare numerically several methods for synchronization over SE(d) including our own. In the first part we provide numerical examples for synthetic data where the noise model consists of inliers, outliers, and missing data. The results show that the contraction approach, which comes together with some flexibility in choosing the solver for synchronization over the compact group of rotations lead to superior performance in almost any tested scenario. The second part is devoted to real data, coming from the problem of 3D registration, that is to estimate a 3D object from raw partial scans of it. This problem presents a realistic mix of noise and missing data and provides a solution that can be evaluated visually. For two different objects, in three different scenarios the synchronization via contraction shows its power in presenting excellent solutions. Furthermore, these results demonstrate that contraction in the group level outperforms contraction in the matrix level, and so justifies the use of algebraic structures.

The paper is organized as follows. In Section 2 we provide a survey on synchronization over groups, including the case of SE(d). Then, in Section 3 we introduce the Cartan motion groups and the notion of group contraction. We also provide a geometrical interpretation in low dimension and the linear variant of compactification on the matrix level. Section 4 presents how to apply compactification for solving synchronization and its analysis. In Section 5 we show numerical examples, starting with conventions about measuring error and noise and how to address several implementation issues such as parameter selection. Finally, we numerically compare the performance of various algorithms for synchronization using synthetic and real data.
2 Synchronization over groups – background

We provide some background on the synchronization problem over groups. This problem is to estimate \( n \) unknown elements \( \{ g_i \}_{i=1}^n \) from a group \( G \), given a set of ratio measurements \( \{ g_{ij} \}_{1 \leq i < j \leq n}, g_{ij} \approx g_i g_j^{-1} \). This set might include only partial collection of measurements, outliers, and noisy ratio samples. As above, we denote the undirected graph of measurements by \( G = (V, E) \), where any edge \((i, j)\) in \( E \) corresponds to \( g_{ij} \) and the set of vertices \( V \) corresponds to the unknown group elements \( \{ g_i \}_{i=1}^n \). We also assign weights \( w_{ij} \geq 0 \) to edges, where in a naive model these weights indicate whether \((i, j) \in E\), that is 1 if such edge exists and 0 otherwise. In other more advanced models, these weights can be correlated with the noise or suggest our confidence in the measurements.

2.1 Synchronization over compact groups: a short survey and an example

The basic assumption in this part is that \( G \) is a compact group, and thus having a faithful, finite dimensional and orthogonal (or unitary, for the complex case) representation \( \rho: G \to GL(d, \mathbb{R}) \). Note that henceforth we focus on the real case, because the complex case (with representations in \( GL(d, \mathbb{C}) \)) is essentially similar. Having such \( \rho \) means that solving synchronization over \( G \) is equivalent to solving synchronization over the group of orthogonal transformations \( O(d) \), and we specifically focus on the case of \( SO(d) \).

The problem of synchronization of rotations is widely studied, since it naturally arises in many robotics (motion) applications and as a crucial step in solving the classical computer vision problem of Structure from Motion (SfM). A comprehensive survey about robust rotation optimization in SfM applications is done in [58]. Another extensive survey is done in [11] within the context of pose graph optimization (estimating the positions and orientations of a mobile robot). Since we ultimately plan to solve synchronization over some non-compact groups by mapping it into compact groups and solve it there, we shortly state several approaches for directly solving the synchronization of rotations.

We begin with an example of a method that globally integrates the data of synchronization to obtain a solution. The full analysis of this method is given in [7, 50].

Example 2.1 (The spectral method). \textit{The solution is constructed based upon the spectral structure of a specially designed (kernel) matrix} \( H \), \textit{defined as follows}. Let \( M \) be a \( dn \times dn \) block matrix, \textit{consists of the} \( d \times d \) blocks

\[
M_{ij} = \begin{cases} 
    w_{ij} \rho(g_{ij}) & \text{if } (i, j) \in E \\
    0_{d \times d} & \text{if } (i, j) \notin E.
\end{cases}
\]

Set \( w_{ii} = 1 \) and \( M_{ii} = I_d \), for all \( 1 \leq i \leq n \) and for consistency \( M_{ji} = M_{ij}^T \) since \( \rho \) is orthogonal and so

\[
(g_i g_j^{-1})^{-1} = (g_j g_j^T)^T = g_j g_j^T = g_j g_j^{-1}.
\]

\textit{Namely, the orthogonality of our representation} \( \rho \) and the fact that \( w_{ij} = w_{ji} \) \textit{make} \( M \) \textit{a self adjoint matrix}. Denote by \( \deg(i) = \sum_{j: (i,j) \in E} w_{ij} \) \textit{the total sum of edge weights from the} \( i \text{th} \) \textit{vertex} (the degree in} \( G \) \textit{and define a diagonal} \( dn \times dn \) \textit{matrix} \( D \) \textit{where each} \( d \times d \) \textit{diagonal block} \( D_i \) \textit{is given by} \( D_i = \deg(i) I_d \). \textit{The kernel matrix is} \( H = D^{-1} M \).
The fundamental observation of the spectral method is that in the noise-free case we get

\[ M\mu = D\mu, \quad \mu = \begin{bmatrix} \rho(g_1) \\ \vdots \\ \rho(g_n) \end{bmatrix}. \] (2)

Namely, the columns of \( \mu \) are the top \( d \) eigenvectors of \( H \), corresponding to the eigenvalue 1. In addition, based on the non-negativity of the weights \( w_{ij} \), the matrix \( H \) is similar to the kronecker product of the normalized Laplacian matrix of the graph and the identity matrix \( I_d \), using the conjugate matrix \( D^{-1/2}A \), where \( A \) is a block diagonal matrix with the block \( \rho(g_i) \), \( i = 1, \ldots, n \). The last step in the spectral approach is a rounding procedure, since in the noisy case there is no guarantee that each \( \mu_i \) is in the image of \( \rho \). We summarize this example into a pseudocode in Algorithm 2 on Appendix B.

As seen in Example 2.1, the structure of the data implies a spectral solution. Yet a different way to integrate the data globally is by solving semidefinite programming (SDP) problem, e.g., \[25, 46, 47, 50\]. Global integration can also be exploited for the case of outliers in the data, where the cost function is tailored to be more robust for inconsistent data \[29, 60\]. It is also possible to attain robustness using a maximum likelihood \[8\]. Another closely related approach is to use the Riemannian structure of the manifold of rotations to define iterative procedures, e.g., \[13, 30\]. A different approach is related to the cyclic features of the data. Specifically, the consistency of the data states that for noiseless measurements \( g_{ij}g_{j\ell} = g_{i\ell} \). This information can be used to either form a solution or to discard outliers, see e.g., \[48, 61\]. Many other approaches have been designed for low dimensional rotations, for example with \( d = 3 \) there is a unique relation between rotations and quaternions, e.g.,\[25\]. These low dimensional cases are important as they appear in many applications. A further popular approach for practitioners is to mix between several methods to have a superior unified algorithm. These topics are well covered in the abovementioned surveys \[11, 58\].

### 2.2 An overview of synchronization over the special Euclidean group

In the main focus of this paper is the class of Cartan motion groups. This class of groups, which will be precisely defined in the sequel section, includes the important group of rigid motions, that is the special Euclidean group \( G = \text{SE}(d) \) of isometries that preserve orientation. We naturally identify \( \text{SE}(d) \) by the semidirect product \( \text{SO}(d) \ltimes \mathbb{R}^d \) and multiplication is given by

\[ (\mu_1, b_1)(\mu_2, b_2) = (\mu_1\mu_2, b_1 + \mu_1 b_2), \quad (\mu_1, b_1), (\mu_2, b_2) \in \text{SO}(d) \ltimes \mathbb{R}^d. \] (3)

The above description is also reflected by the \( d + 1 \) dimensional representation \( \rho : \text{SE}(d) \to \text{GL}(d + 1, \mathbb{R}) \),

\[ \rho(g) = \begin{bmatrix} \mu \\ b \end{bmatrix}, \quad g = (\mu, b) \in \text{SO}(d) \ltimes \mathbb{R}^d. \] (4)

Henceforth, we will identify the group elements with their matrix representation \([4]\).

Next we briefly survey some of the main approaches used today to solve the so called rigid motion synchronization. One fundamental issue is the more intricate Riemannian structure. This leads, for example, to extra effort in working with the exponential and logarithm maps. One other issue with synchronization over non-compact group, and in particular in rigid motion synchronization, is that there
is no faithful, finite dimensional and orthogonal representation. This fact complicates most methods for rotation synchronization. For example, for the spectral method of Example 2.1, the affinity matrix $M$ defined using (4) is no longer self adjoint, which makes the spectral structure of the kernel matrix less accessible. Nevertheless, there is a way to address this difficulty, by considering a Laplacian operator and then to reveal its null space using the SVD decomposition. This spectral method is done both in [3] and in [6], where it was shown to be an efficient and accurate method.

Optimization also plays a significant role in solving rigid motion synchronization. Indeed, the search space is non-compact and exponential in $n$, and thus it requires either relaxing or reducing the original problem into a more tractable one. One way to do so is by separating the optimization over the compact and non-compact parts of the group, e.g., [32]. One different motivation to use separation appears when considering quadratic loss function, taken from the compact group case [50]

$$\minimize \left\{ \mu_i \right\}_{i=1}^{n} \sum_{(i,j) \in E} w_{ij} \left\| \mu_i \mu_j - \mu_{ij} \right\|_F^2 + \sum_{(i,j) \in E} w_{ij} \left\| b_i - \mu_i \mu_j b_j - b_{ij} \right\|_2^2$$

subject to $\mu_i^T \mu_i = I_d$, $\det(\mu_i) = 1$, $b_i \in \mathbb{R}^d$, $i = 1, \ldots, n$. (6)

The general approach of separation suggests to first minimize the first sum, using for example the spectral method, and then solve the least squares problem of the second sum with the estimated $\left\{ \mu_i \right\}_{i=1}^{n}$, as done in [15]. Since the representation of SE(d) is not orthogonal, there is no canonical way of fixing the cost function in the rotations case. In fact, changing the cost function of (5) leads to several interesting solutions. One naive change is to consider

$$\sum_{(i,j) \in E} w_{ij} \left\| \rho(g_i^{-1} g_{ij} g_j) - 1_{d+1} \right\|_F^2,$$

which yields to the completely separated cost function

$$\sum_{(i,j) \in E} w_{ij} \left\| \mu_i \mu_j - \mu_{ij} \right\|_F^2 + \sum_{(i,j) \in E} w_{ij} \left\| b_i - \mu_i b_j - b_{ij} \right\|_2^2.$$ (8)

In this case the separated cost function has an undesirable property – it is not invariant to right multiplication by an arbitrary group element. The loss of invariance is clear directly from (7) but also in (8) as the translational part of such global alignment results in additional terms in the summand. However, this solution is potentially more suitable for problems where we need to introduce extra constraints for translation parts (the $b_i$-s), e.g. in the fresco reconstruction problem, where one can introduce extra constraints to prevent overlapping of the fresco fragments [20]. Another variant is the least unsquared deviations (LUD) which uses

$$\sum_{(i,j) \in E} w_{ij} \left\| \rho(g_i^{-1} g_{ij}) - \rho(g_{ij}) \right\|_F.$$
This is done in [30] to face the problem of outliers. Yet another cost function is

$$\sum_{(i,j) \in E} w_{ij} \| \rho(g_i) - \rho(g_{ij}g_j) \|^2_F = \text{tr} \left( \mu^T L \mu \right),$$

where $\mu$ is the vectors of group elements (matrices) as in [2], and $L = (I - M)(I - M^T)$ where $M$ is the measurements matrix as defined in [1]. The matrix $L$ can be also considered as an extension of the “twisted Laplacian”, as defined in [36]. This cost functions brings us back to the spectral method, as the operator $L$ is is related to the Laplacian operator in [3] and [6]. Note that both papers [3, 6] focus on the case $d = 3$ and relax the constraint $\mu_i \in \rho(G)$ by forcing the solution $\mu$ to coincide with $(0, 0, 0, 1)$ in each fourth row and then project the upper left block of each $\mu_i$ on $\text{SO}(3)$.

As in rotation synchronization, there are a few additional approaches, motivated by the SfM problem. Similar to the above spectral and optimization methods, the following solutions do not use separation, and integrate the data concurrent in both parts of the rigid motion. One method solves the optimization problem related to the SfM with rigid motions by using semidefinite relaxation (SDR) [40]. Relaxing the constraints of $\text{SE}(3)$ by its convex hull was considered in [44]. Note that some of the solutions related to the SfM problem are specifically designed to operate under the assumption that $d = 3$, and their generalizations to higher dimensions are not always trivial, if possible; one example of such case is in [54] where the authors use the unique relation between $\text{SE}(3)$ and quaternions, in order to use diffusion tools on the graph. The explicit structure of the exponential map and its inverse over the case $\text{SE}(3)$ is also utilized in [26] to define iterative averaging algorithm. The same explicit form of the exponential map helps in defining the cyclic constraints that leads to an optimized solution in [42]. Iterative optimization that also exploits the graph structure is done in [50] by a distributed algorithm for sensor network orientation. Riemannian gradient descent adapted for $\text{SE}(3)$ is also used for rigid motion synchronization under two different variants in [55, 57].

3 Compactification of Cartan motion groups

We introduce the class of Cartan motion groups and then describe the notion of group contraction and compactification, including two specific examples of compactifications in the group level and in the matrix level.

3.1 Cartan motion groups

We start with a brief introduction to Cartan decomposition and its induced Cartan motion groups. This prefatory description is tailored for our ultimate goal of addressing the problem of synchronization for such groups via contraction maps.

Let $G_0$ be a semisimple compact Lie group. Its worth mentioning that algebraically, most of the following can be carried for non-compact groups with finite center instead of $G_0$, but this case is not much useful for us as will be clear soon. Let $g$ be the corresponding Lie algebra of $G_0$. For $g$, there is a Cartan decomposition $g = t \oplus p$ such that for an associated Cartan involution, $t$ and $p$ are the eigenspaces corresponding to eigenvalues of $+1$ and $-1$, respectively. As such, the linear space $t$ satisfies $[t, t] \subset t$ and thus is a Lie subalgebra, where in general $p$ is just a linear space. We denote by $K$ the Lie subgroup
Thus, we end up with the Cartan motion group,  
\[ G = K \ltimes V. \]

The action of the group is

\[ (k_1, v_1)(k_2, v_2) = (k_1 k_2, v_1 + \text{Ad}_{k_1}(v_2)), \tag{9} \]

where Ad(·) refers to the adjoint representation on \( G_0 \). Note that \( p \) is closed under the adjoint, that is \( \text{Ad}_k(v) \in p \) for all \( k \in K \) and \( v \in p \). Before proceeding, we provide some examples for the above settings.

**Example 3.1** (Euclidean group). Let \( G_0 = \text{SO}(d+1) \) be the group of all orthogonal matrices of order \( d+1 \) with positive determinant. This Lie group has the Lie algebra \( g \) consists of all real square matrices of size \( d+1 \), which are skew symmetric, \( X + X^T = 0 \). To form the decomposition above, consider the matrices \( E(i,j) \), having 1 on the \((i,j)\) entry and zero otherwise. Thus, \( \{E(i,j) - E(j,i)\}_{1 \leq i < j \leq d+1} \) is a basis of \( g \). The span of \( \{E(i,j) - E(j,i)\}_{1 \leq i < j \leq d} \) forms the Lie algebra \( t \), which is isomorphic to the Lie algebra of \( \text{SO}(d) \). Indeed, we can identify \( \text{SO}(d) \) with \( K \), the subgroup of \( G_0 \) of all matrices having \( e_{d+1} \) as their last column, where \( e_i \) is the unit vector that has one on the \( i \)th component. The complement of \( t \) is \( p \), which is spanned by \( \{E(i,d+1) - E(d+1,i)\}_{1 \leq i \leq d} \). The linear space \( p \) is easily identified with \( \mathbb{R}^d \) via the map

\[ E(i,d+1) - E(d+1,i) \mapsto e_i. \tag{10} \]

Thus, we end up with the Cartan motion group \( \text{SO}(d) \ltimes \mathbb{R}^d \), that is \( \text{SE}(d) \). More on this example can be found in [16] and in the next section.

**Example 3.2.** Consider \( G_0 = \text{SU}(d) \), the group of all unitary matrices with determinant equal to one. Although its definition relies on complex matrices, it is common to consider it as a real Lie group. There are three main types of Cartan decompositions, we mention one that corresponds with the Cartan involution of complex conjugation. As such, it divides the Lie algebra \( \{X \mid X + X^* = 0, \ \text{tr}(X) = 0\} \) to a real part, which is the same Lie algebra of \( \text{SO}(d) \), consisting of all skew symmetric matrices. Its orthogonal complement, denoted by \( W = \text{SO}(d)^\perp \), can be described as the direct sum of a \( d-1 \) dimensional subspace of pure complex, diagonal matrices with zero trace, and a \( \frac{d(d-1)}{2} \) dimensional subspace of pure complex matrices of the form \( X_{i,j} = X_{j,i}, \) with zero diagonal. Then, the Cartan Motion group in this case is \( \text{SO}(d) \ltimes W \).

**Example 3.3** (matrix motion group). Let \( G_0 = \text{O}(d+\ell) \) be the group of all real orthogonal matrices of order \( d+\ell \). Denote by \( M(d,\ell) \) the space of all real matrices of order \( d \times \ell \). Then, one Cartan decomposition of \( g \) can be identified with the product of the Lie algebras of \( \text{O}(d) \) and \( \text{O}(\ell) \) (namely \( t \)) and the space \( p = M(d,\ell) \). This decomposition yields the so called matrix motion group, \( G = (\text{O}(d) \times \text{O}(\ell)) \ltimes M(d,\ell) \). This particular example draws a lot of attention since this Cartan motion group is also associated with the quotient space \( G/K \) which in this case is the Grassmannian manifold, consisting of all \( d \) dimensional linear subspaces of \( \mathbb{R}^{d+\ell} \). For more details see [38, Pages 107–128].

### 3.2 Group contraction map

Lie group contraction is a process of obtaining one Lie group from a series of other Lie groups. This concept was developed primary in the seminal paper [31] to describe the relation between groups of two
physical theories linked by a limiting process (for example, the inhomogeneous Lorentz group is a limiting case of the de Sitter groups, more details on the theory of contraction can be found in e.g., [22]). In this paper we focus on one family of contractions, established in [16, 17], for semisimple Lie groups.

Following [16] we concentrate on the family of contractions \( \{\Psi_\lambda\}_{\lambda>0} \), which are smooth maps \( \Psi_\lambda : G \mapsto G_0 \) and defined by

\[
\Psi_\lambda(k,v) = \exp(v/\lambda)k, \tag{11}
\]

where the exponential is the Lie group exponential map of \( G \) (note that indeed \( v/\lambda \in \mathfrak{g} \)). The group \( G \) is commonly referred to as the contraction group of \( G_0 \).

The parameter \( \lambda \) in (11) has a major role as it reflects, in a sense, how much significant information from \( p \) passes to \( G_0 \). In the extreme case of \( \lambda \to \infty \), the contraction map \( \Psi_\lambda \) simply maps the group \( K \) to itself as a subgroup of \( G_0 \). In the opposite scale of \( \lambda \) where it is closer to zero, the map becomes less practical for two main reasons. First, the vector \( v/\lambda \) is likely to be very large and lies outside the injectivity radius of the exponential map of \( G_0 \), namely the exponential map is no longer a diffeomorphism. This means, for example, that we cannot apply the (correct) inverse of \( \Psi_\lambda \). For synchronization, the ability to apply the inverse is crucial. Second, the exponential map depends on the curvature of the Lie group manifold (given algebraically by a commutator term). Thus, small \( \lambda \) values lead to larger vectors which are more exposed to the effects of curvature and result in large distortions of the mapped elements. Subsequently, and without loss of generality we consider \( \lambda \) to be restricted in \([1, \infty)\). Note that the global Cartan decomposition of the group \( G_0 \) is defined by (11) with \( \lambda = 1 \).

The first two algebraic properties which we require from the contractions are

\[
\Psi_\lambda(g^{-1}) = (\Psi_\lambda(g))^{-1}, \quad g \in G, \\
\|\Psi_\lambda(g_1 g_2) - \Psi_\lambda(g_1)\Psi_\lambda(g_2)\|_F = O\left(\frac{1}{\lambda^2}\right), \quad g_1, g_2 \in G, 
\]

where \( O(\cdot) \) means that \( O(x)/x \) is bounded for small enough \( x \), and the group elements of \( G_0 \) are identified with their orthogonal matrix representation. We term the properties in (12) “approximated homomorphism”. Note that while we allow some approximation on the homomorphism of the action of the group, we require exactness in the inverse. For our family of contractions, these requirements are satisfied.

**Proposition 3.1.** The contractions defined in (11) are approximated homomorphisms.

**Proof.** Let \( g = (k,v) \in G \), then by (9) we have \( g^{-1} = (k^{-1}, \text{Ad}_{k^{-1}}(-v)) \). Therefore, using the linearity of the adjoint and its commutativity with the exponent we have

\[
\Psi_\lambda\left(g^{-1}\right) = \exp(\text{Ad}_{k^{-1}}(-v)/\lambda)k^{-1} = \text{Ad}_{k^{-1}}(\exp(-v/\lambda))k^{-1} = k^{-1}\exp(-v/\lambda) = (\exp(v/\lambda)k)^{-1} = (\Psi_\lambda(g))^{-1}. 
\]

For the second part, let \( g_1 = (k_1, v_1), g_2 = (k_2, v_2) \in G \). By (9)

\[
\|\Psi_\lambda(g_1 g_2) - \Psi_\lambda(g_1)\Psi_\lambda(g_2)\|_F = \|\exp(v_1 + \text{Ad}_{k_1}(v_2)/\lambda))k_1k_2 - \exp(v_1/\lambda)k_1 \exp(v_2/\lambda)k_2\|_F. 
\]
Since we identify each component with its orthogonal representation in $G_0$, we can easily cancel $k_2$ from the right hand side. Moreover, by multiply $k_1^{-1}$ from the right and using again the adjoint commutativity with the exponential we get

$$\exp(v_1/\lambda)k_1 \exp(v_2/\lambda)k_1^{-1} = \exp(v_1/\lambda)\text{Ad}_{k_1} \left( \exp(v_2/\lambda) \right) = \exp(v_1/\lambda) \exp(\text{Ad}_{k_1} (v_2/\lambda)).$$

The Zassenhaus formula (see e.g., [12]) indicates that

$$\exp((v_1 + \text{Ad}_{k_1} (v_2))/\lambda) = \exp(v_1/\lambda) \exp(\text{Ad}_{k_1} (v_2)/\lambda) \exp(-1/2 \left[ v_1/\lambda, \text{Ad}_{k_1} (v_2)/\lambda \right]) \left( 1 + O(1/\lambda^3) \right).$$

For simplicity denote $c = [v_1/\lambda, \text{Ad}_{k_1} (v_2)/\lambda]$. Then, since the norm is invariant to the common exponential terms we are left with

$$\left\| I - \exp(-1/2) \exp(1/6 \left( 2 [\text{Ad}_{k_1} (v_2)/\lambda, c] - [v_1/\lambda, c] \right)) \left( 1 + O(1/\lambda^3) \right) \right\|.$$

Consider the Taylor expansion of the matrix exponential map $\exp(A) = I + A + A^2/2 + \ldots$, and recall that $c = \lambda^{-2} (v_1 \text{Ad}_{k_1} (v_2) - \text{Ad}_{k_1} (v_2)v_1)$, to conclude that the leading order in (13) is indeed $\lambda^{-2}$, as required.

Note that the convergence of the above product of the Zassenhaus formula is guaranteed when $\|v_1\| + \|\text{Ad}_{k_1} (v_2)\| \leq \lambda r$, with $r \approx 0.59$ (see e.g., [12]), which is always true for large enough $\lambda$.

### 3.3 The Euclidean group case

The case of $G = \text{SE}(d)$ has already been shown in Example 3.1 to be a special case of Cartan motion groups. Note that the general action (9), where $\text{Ad}(\cdot)$ is taken as similarity transformation in the matrix representation of $G_0 = \text{SO}(d + 1)$, is identical to the explicit action of the group as given in (3), under the standard matrix representation in $G$.

The contraction maps in $\text{SE}(d)$ take the form

$$\Psi_\lambda (\mu, b) = \exp(\frac{1}{\lambda} b)\mu, \quad \lambda \geq 1,$$

where $\mu \in \text{SO}(d)$ and $\text{SO}(d)$ is identified as a subgroup of $G_0$ given by $\{Q \in \text{SO}(d + 1) \mid Q e_{d+1} = e_{d+1}\}$. In addition, the exponential map is in the group $G_0$ and thus $b \in \mathbb{R}^d$ is understood in this context as a matrix. To be specific, recall [10], then we refer to the map $b \mapsto \begin{bmatrix} a_{d \times d} & b \\ -b^T & 0 \end{bmatrix}$. Further examination of the latter matrix reveals that for $b \neq 0$ its rank is 2, that is the kernel is of dimension $d - 1$. In other words, the spectrum consists of $d - 1$ zero eigenvalues. The remaining two eigenvalues are pure complex (and conjugated as the matrix is real skew symmetric) and equal to $\pm \sqrt{-b^T b}$ since

$$\begin{bmatrix} a_{d \times d} & b \\ -b^T & 0 \end{bmatrix} \begin{bmatrix} b \\ \pm \sqrt{-b^T b} \end{bmatrix} = \pm \sqrt{-b^T b} \begin{bmatrix} b \\ \pm \sqrt{-b^T b} \end{bmatrix}.$$  

The spectrum is important since the exponential function as a matrix function can be interpreted as a scalar function on the spectrum. Therefore, injectivity of the exponential is guaranteed when $\|b\| = \sqrt{b^T b}$.
To have some intuition regarding the nature of our contraction map $\Psi_\lambda$ in $\text{SE}(d)$, it is interesting to consider the geometrical interpretation of the special cases $\text{SE}(1)$ and $\text{SE}(2)$. The following interpretation is accompanied by a visual illustration given in Figure 1.

The group $\text{SE}(1)$ is isomorphic to the real line as the rotations are only the scalar 1 and the translational part is of dimension 1. Therefore, $\Psi_\lambda$ is just $\exp\left(\begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}\right)$, $b \in \mathbb{R}$ which basically maps the scaled number $b/\lambda$ to a point on the unit sphere in $\mathbb{R}^2$ by an angular distance of $b/\lambda$ from the “east pole” $(0, 1)$. Note that (16) ensures that we cannot approach the west pole and thus we remain in the injectivity domain on the sphere. This interpretation is illustrated in Figure 1 (left panel).

For $\text{SE}(2)$, $\Psi_\lambda$ maps any $g = (\mu, b)$ to an element of $\text{SO}(3)$. This element is equivalent to first applying the rotation $\mu$ around the $z$-axis (recall the subgroup $\text{SO}(2) \cong \{Q \in \text{SO}(3) \mid Qe_3 = e_3\}$) followed by rotation in the “direction” of the matrix $b$ with a geodesic distance of $\|b\|/\lambda$. This characterization, also known as the misorientation axis and angle, helps us interpret geometrically this exponential map as mapping a point from a tangent plane $\{(x, y, 1) : x, y \in \mathbb{R}\}$ on the “north pole” $(0, 0, 1)$ of the sphere $S^2$ into $S^2$. Note that while the usual representation of a rotation in $\text{SO}(3)$ requires a unit vector in $\mathbb{R}^4$ (unit quaternion), in this case since $\exp(b/\lambda)$ is determined by only 2 coordinates we can identify any such matrix with a unit vector in $\mathbb{R}^3$. Therefore, we can think about the mapping $\Psi_\lambda(g)$ as sending the sufficiently small, scaled translational part $b/\lambda$, that is a point close to the origin in $\mathbb{R}^2$, to a corresponding point on the sphere $S^2$ close to its north pole. See also Figure 1 (right panel).

### 3.4 A linear algebra variant – matrix level projections

In previous parts we described how one can exploit the group structure using a group decomposition (Cartan) for mapping $\text{SE}(d)$ to $\text{SO}(d+1)$. In this part, we aim to use the same mapping notion, but on the matrix level and by a matrix decomposition. Precisely, our second compactification method (valid for the group of special Euclidean) is based upon the approximation of the matrices representing elements of
SE(d) by the orthogonal part of their polar decomposition (PD).

Approximating elements from SE(d) by elements from SO(d + 1) is not new. In the context of motion synthesis and the design of approximate bi-invariant metrics on SE(d), the cases \(d = 2\) and \(d = 3\) were studied in [37] using Euler angles and geometrically intuitive ideas. The related following works [19, 33] use Taylor expansions. These methods are also related to the contraction maps discussed above, however since they are based on geometrical notions and Taylor series of special functions, it is difficult to generalize them further to other groups. The projections based on PD were introduced in [34, 35], and can be considered as alternative generalizations to the approximations in [37], for the study of approximate invariant metrics in SE(d).

To be aligned with the contraction maps, we define a family of PD-based projections (we also refer to them broadly as contractions) \(\{\Phi_\lambda\}_{\lambda \geq 1}\), which are maps \(\Phi_\lambda: \text{SE}(d) \rightarrow \text{SO}(d + 1)\) that are defined as

\[
\Phi_\lambda \left( \begin{bmatrix} \mu & b \\ 0_{1 \times d} & 1 \end{bmatrix} \right) = U_\lambda, \tag{17}
\]

where \(U_\lambda\) is the orthogonal part of the PD decomposition \(g_\lambda = Q_\lambda H_\lambda\) (\(H_\lambda\) is symmetric positive definite) of

\[
g_\lambda = \begin{bmatrix} \mu & b/\lambda \\ 0_{1 \times d} & 1 \end{bmatrix}. \]

As in the contraction maps of groups, \(\lambda\) is used as a parameter determining the scale of the translational part of \(b\) and controlling the effect of the translational part in the mapping. In addition, for large \(\lambda\), \(\Phi_\lambda(g)\) tends to \(\begin{bmatrix} \mu & 0_{d \times 1} \\ 0_{1 \times d} & 1 \end{bmatrix}\) which means SE(d) is mapped into SO(d) (as a subgroup of SO(d + 1)).

The term projection is used in the context of \(\Phi_\lambda\) since in the space of square matrices of size \(d + 1\), under the metric induced from the Frobenius norm, we have that \(\Phi_\lambda\left( \begin{bmatrix} \mu & b/\lambda \\ 0_{1 \times d} & 1 \end{bmatrix} \right)\) is the closest element in SO\((d + 1)\) to \(g_\lambda\). This is true since the closest orthogonal matrix is given by \(V_\lambda U_\lambda^T\) where \(g_\lambda = V_\lambda \Sigma_\lambda U_\lambda^T\) is the SVD decomposition, but this is exactly \(Q_\lambda\) from the PD decomposition. Next, we describe \(\Phi_\lambda\) in more detail.

**Proposition 3.2.** Let \(0 \neq b \in \mathbb{R}^d\), and let \(P\) be an orthogonal projection matrix for the subspace

\[
\{v \in \mathbb{R}^d \mid v^T b = 0\}. \tag{18}
\]

Then, the maps \(\Phi_\lambda\) of (17) are in SO\((d + 1)\) and of the form

\[
\Phi_\lambda \left( \begin{bmatrix} \mu & b \\ 0_{1 \times d} & 1 \end{bmatrix} \right) = \begin{bmatrix} 2\tau_\lambda \hat{b}b^T + P \mu & \frac{\tau_\lambda}{\lambda} b \\ -\frac{\tau_\lambda}{\lambda} b^T \mu & 2\tau_\lambda \end{bmatrix},
\]

where \(\hat{b} = b/\|b\|\) and \(\tau_\lambda = \left(4 + \|b\|^2/\lambda^2\right)^{-\frac{1}{2}}\).

The proof is given in Appendix A.1

The explicit form of the projections allows as to deduce some important properties. First, we do not have to fully compute the SVD of the representing matrix of \(g_\lambda\), we just need to construct \(P = P(b)\) (by finding a basis for the subspace (18)). Second, it shows how to invert \(\Phi_\lambda\), which is crucial for its use as a compactification method for synchronization, as we will see in the next section. Thus, we conclude
Corollary 3.3. The maps $\Phi_\lambda$ of (17) are injective.

Proof. The corollary holds since Proposition 3.2 implies that for any $g_1 = \begin{bmatrix} \mu_1 & b_1 \\ \mathbf{0}_{1 \times d} & 1 \end{bmatrix}$ and $g_2 = \begin{bmatrix} \mu_2 & b_2 \\ \mathbf{0}_{1 \times d} & 1 \end{bmatrix}$ in SE($d$), $\Phi_\lambda(\gamma_1) = \Phi_\lambda(\gamma_2)$ entails $b_1 = b_2$ and $\mu_T^1 b_1 = \mu_T^2 b_2$. Namely, $g_1 = g_2$.

As in the case of group contraction maps, the justification for using $\Phi_\lambda$ is as follows,

Proposition 3.4. The maps $\Phi_\lambda$ of (17) are approximated homomorphisms, as defined in (12).

The proof is given in Appendix A.2.

4 Application of compactification to synchronization

We now focus on the application of the above algebraic tools of compactification to synchronization over groups. In light of current solutions for synchronization, described in Subsection 2.2, the use of contractions for synchronization offers several important advantages. To begin with, it uses full integration of the available data (in contrary to methods that separate between the two components of the group). It is valid for a wide class of non-compact Cartan motion groups, which includes the important special case of SE($d$). One important feature when operating on SE($d$) data is the ability to handle different scales of the two components of the data. Namely, the rotational and translational parts are naturally having different scales, and this might be crucial when applying any scheme on data from SE($d$), see e.g., [41, 62] and Section 5. Using contraction, we have an inherent mechanism to handle such differences in scaling. At last, as reviewed in Subsection 2.1 many efficient methods were designed and suggested for solving rotation synchronization. By applying contraction, we facilitate the use of almost any such method in order to gain extra robustness (for example, dealing with outliers or noise) for the case of synchronization over Cartan motion groups as well.

4.1 Synchronization: from Cartan motion to compact and back

We describe the use of compactification for solving the synchronization problem over Cartan motion groups. The compactification methods are exploited to reduce this problem into the well-studied rotation synchronization. Using the notation of the previous section, $G_0$ is a compact group and $G$ is its contraction, a Cartan motion group. We solve the synchronization over $G$ as follows. First, we apply the contraction map $\Psi_\lambda$ to $\{g_{ij}\}$, forming a new set of measurements in $G_0$. Then, we solve the corresponding synchronization problem, having the same data graph $\mathcal{G}$ under the new set of measurements. The constraint of the new synchronization problem is now $\{\Psi_\lambda(g_i)\}_{i=1}^n \subseteq \rho(G_0)$. Then, having a solution to the problem in $G_0$, we map it back to $G$ by using the inverse map $\Psi_\lambda^{-1}$. This procedure is summarized in Algorithm 1.

The back mapping from $G_0$ to $G$ after the compact synchronization step leads to the following observation; while compact synchronization offers a non-unique solution, which is determined up to a global alignment (right multiplication by a fixed group element), the choice of the certain solution which we actually map back to $G$ effects the solution there. Namely, the solution in $G$ is still non-unique, but for any two different solutions (equal up to global alignment) in $G_0$ we get two different solutions in $G$ which are not necessary differ only by global alignment. This is a result of the fact that we cannot guarantee
the back mapping to be a homomorphism. Nevertheless, we can characterize when global alignment in $G_0$ remains a global alignment in $G$ after the back mapping.

**Proposition 4.1.** Let $\{Q_i\}_{i=1}^n$ be a solution to the compact synchronization step in Line 7 of Algorithm 7. Denote the Cartan decomposition in $G_0$ of each of its elements by $Q_i = \exp(b_i)\mu_i$. Then, the output of Algorithm 7 is invariant to any global alignment $Q$ with the Cartan decomposition $Q = \exp(b)\mu$ such that $\text{Ad}_{\mu_i}(b)$ commutes with $b_i$, $i = 1, \ldots, n$.

Note that $b = 0$ always satisfies the conditions in the above proposition.

**Proof.**

$$Q_iQ = \exp(b_i)\mu_i \exp(b)\mu = \exp(b_i) \exp(\text{Ad}_{\mu_i}(b))\mu_i\mu.$$  

However, $[\text{Ad}_{\mu_i}(b), b_i] = 0$ and therefore $\exp(b_i) \exp(\text{Ad}_{\mu_i}(b)) = \exp(b_i + \text{Ad}_{\mu_i}(b))$ with both $b_i + \text{Ad}_{\mu_i}(b) \in p$ and $\mu_i\mu \in t$. Namely,

$$\Psi^{-1}_\lambda(Q_iQ) = \begin{bmatrix} \mu_i \mu & \lambda(b_i + \mu_i b) \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \mu_i \lambda b_i & \mu \lambda \mu_i b \\ 0 & 1 \end{bmatrix} = \Psi^{-1}_\lambda(Q_i) \Psi^{-1}_\lambda(Q).$$

Since the inverse is invariant for approximated homomorphism maps, so is the case for the inverse map, and overall we conclude that

$$\Psi^{-1}_\lambda(Q_iQ) \Psi^{-1}_\lambda((Q_jQ)^{-1}) = \Psi^{-1}_\lambda(Q_i) \Psi^{-1}_\lambda(Q_j)^{-1},$$

In other words, any cost function of synchronization over $G$ that is based on the distance between $g_{ij}$ and $\Psi^{-1}_\lambda(Q_i) \Psi^{-1}_\lambda(Q_j)^{-1}$ remains unchanged after the global alignment $Q_i \mapsto Q_iQ$.

Proposition 4.1 basically states that any global alignment that preserves the structure of the Cartan decomposition in $G_0$, does not affect the back mapping to $G$. In other words, this result opens the door for further potential improvement in synchronization, done by carefully choosing a global alignment in $G_0$. 

---

**Algorithm 1 Synchronization via group contraction on Cartan motion groups**

**Require:** Ratio measurements $\{g_{ij}\}_{(i,j) \in E} \subset G$, and corresponding weights $\{w_{ij}\}_{(i,j) \in E}$.

Non-trivial orthogonal, faithful representation $\rho$ of $G_0$.

**Ensure:** An approximated solution $\{\hat{g}_i\}_{i=1}^n$.

1: Choose $\lambda = \lambda \left( \{g_{ij}\}_{(i,j) \in E} \right) \geq 1$. {See Section 5}

2: for $(i,j) \in E$ do

3: $\mu_{ij} \leftarrow \Psi_\lambda(g_{ij})$ {Mapping to $G_0$ (can be done with $\Phi_\lambda$ for the SE(3) case)}

4: end for

5: $\{\mu_i\}_{i=1}^n \leftarrow$ Solve synchronization over $G_0$ with $\{\mu_{ij}\}_{(i,j) \in E}$, $\{w_{ij}\}_{(i,j) \in E}$, and $\rho$. {For example, using Algorithm 2}

6: for $i = 1, \ldots, n$ do

7: $\hat{g}_i \leftarrow \Psi^{-1}_\lambda(\mu_i)$ {Back mapping to $G$}

8: end for

9: return $\{\hat{g}_i\}_{i=1}^n$. 

---

14
that minimizes the error after mapping back to \( G \). Next in Subsection 5.1.3 we describe how to exploit
this idea in the case of \( G = \text{SE}(d) \).

Many of current solutions for synchronization are based on minimizing a cost function that depends on the Euclidean distance between the ratios of estimated elements and the given measurements. One such example is given in (5) for the case of \( \text{SE}(d) \). In terms of a general Cartan motion group \( G = K \ltimes V \) we can rewrite this problem as

\[
\begin{align*}
\minimize_{\{g_i\}_{i=1}^n} & \sum_{(i,j) \in E} w_{ij} d_H \left( g_i g_j^{-1}, g_{ij} \right)^2 \\
\text{subject to} & \quad \{g_i\}_{i=1}^n \subseteq G,
\end{align*}
\]

with the hybrid metric \( d_H(g_1, g_2) = \sqrt{\|k_1 - k_2\|^2_K + \|v_1 - v_2\|^2_V} \) for \( g_1 = (k_1, v_1), g_2 = (k_2, v_2) \in G \). The norm on \( K \), as a (compact) subgroup of the compact group \( G_0 \), is defined as the Frobenius norm on \( \rho \), its faithful, orthogonal matrix representation, and the norm on \( V \) is the Euclidean norm.

For the special case \( G = \text{SE}(d) \) the above synchronization problem (19) gets the following form after applying the contraction \( \Psi_\lambda \)

\[
\begin{align*}
\minimize_{\{Q_i\}_{i=1}^n \subset \text{SO}(d+1)} & \sum_{(i,j) \in E} w_{ij} \|Q_i Q_j^T - \Psi_\lambda(g_{ij})\|_F^2 \\
\text{subject to} & \quad Q_i = \Psi_\lambda(g_i), \quad g_i \in G, \quad i = 1, \ldots, n.
\end{align*}
\]

Following Algorithm 1 we firstly approach (20) by solving synchronization with \( \{\Psi_\lambda(g_{ij})\}_{i,j \in E} \subset \text{SO}(d+1) \). For that, we can utilize one of the methods mentioned in Subsection 2.1. For example, following the spectral method of Example 2.1 we construct the measurement matrix \( M \), given by

\[
M_{ij} = \begin{cases}
  w_{ij} \Psi_\lambda(g_{ij}) & \text{if } (i, j) \in E, \\
  0_{(d+1) \times (d+1)} & \text{if } (i, j) \notin E,
\end{cases}
\]

and extract its leading eigenvectors to have our rotations estimation. We then calculate each inverse \( \Psi_\lambda^{-1}(Q_i) \) to recover a solution, back in \( \text{SE}(d) \). This step requires the calculation of the Cartan decomposition for every other element (see Example 3.1 for this Cartan decomposition). We elaborate more on the calculations involved for obtaining the Cartan decomposition in Appendix B.2.

In the above example, we see a typical difference between an original synchronization problem in \( G \) and its reduced corresponding problem in \( G_0 \). The source of difference in the cost functions of (5) and (19) are the norms. Since these Euclidean distances play a key role in many instances of cost functions of synchronization, e.g., in LUD methods \([39, 60]\) or in least squares methods \([50]\), we prove that there exists a factor, quadratic in \( \lambda \), that links them.

**Proposition 4.2.** Let \( \Psi_\lambda \) be a contraction of the form (11). Assume that \( g_1, g_2, g \in G \) are such that \( g_1 g_2^{-1} \) and \( g \) are close enough in the sense of Lemma A.2, that is their \( p \) components after scaling by \( \lambda \)
are inside the injectivity radius around the identity. Then,

\[ \|\Psi_\lambda(g_1)\Psi_\lambda(g_2)^{-1} - \Psi_\lambda(g)\|_F^2 \leq \left( d_H (g_1g_2^{-1}, g) \right)^2 + O(\lambda^{-2}). \]

where the constant in \( O(\lambda^{-2}) \) depends on \([g_1, g_2]\) but not on \( \lambda \).

Note that the condition above on sufficiently closeness of elements is always true for large enough \( \lambda \). The proof of the proposition is given in Appendix A.3.

4.2 Additional analysis of synchronization via contractions

One feature which is in the backbone of many synchronization methods, such as the spectral method, is the exactness of solution in the case of no noise (where sufficiently data is available). Note that for the noiseless case, the naive solution of constructing spanning tree is sufficient, so from a design point of view we aim to apply synchronization for cases of corrupted (either by noise, outliers or missing data) sets of measurements. Nevertheless, it is interesting from the analysis point of view to understand the contraction based synchronization for this case. We start with a lemma.

**Lemma 4.3.** Let \( g_1 = (\mu_1, b_1) \) and \( g_2 = (\mu_2, b_2) \) be element of a Cartan motion group \( G \). Then,

\[ \Psi_\lambda \left( g_1 g_2^{-1} \right) = \Psi_\lambda (g_1) Q (\Psi_\lambda (g_2))^T, \]

where

\[ Q = \prod_{m=2}^\infty \exp \left( Z_m(\text{Ad}_{\mu_1 T} (b_1 / \lambda), \text{Ad}_{\mu_2 T} (b_2 / \lambda)) \right), \]

and \( Z_m \) is a homogeneous Lie polynomials (nested commutators) of degree \( m \).

**Proof.** By definition (recall that we identify \( \Psi_\lambda(\cdot) \) with its orthogonal representation)

\[ Q = \Psi_\lambda (g_1)^T \Psi_\lambda \left( g_1 g_2^{-1} \right) \Psi_\lambda (g_2) \]
\[ = \mu_1^T \exp(\lambda^{-1}(-b_1)) \exp(\lambda^{-1}(b_1 - \mu_1 \mu_1^T b_2)) \mu_1 \mu_2^T \exp(\lambda^{-1}(b_2)) \mu_2 \]
\[ = \exp(\lambda^{-1}(\text{Ad}_{\mu_1 T} (-b_1))) \exp(\lambda^{-1}(\text{Ad}_{\mu_1 T} (b_1) - \text{Ad}_{\mu_2 T} (b_2))) \exp(\lambda^{-1}(\text{Ad}_{\mu_2 T} (b_2))) \]

The last equation is of the form \( \exp(-A) \exp(A + B) \exp(B) \), so applying the Zassenhaus formula on the most right product yields \( \exp(A) \) multiply by an infinite product of exponentials of nested commutators. Note that \([A + B, B] = [A, B] \). \( \Box \)

As a conclusion from Lemma 4.3 we learn that only by applying the contraction map we suffer from some distortion of the data (by a multiplicative factor).

**Corollary 4.4.** In the case of clean measurements \( g_{ij} = g_i g_j^{-1} \), the compact synchronization (Line 5 of Algorithm 1) is done for data \( \Psi_\lambda (g_i) Q_{ij} \Psi_\lambda (g_j) \) distorted by an orthogonal matrix \( Q_{ij} \), in a distance that is bounded by \( O(\lambda^{-2} [g_i, g_j]) \) from the identity matrix.
Bounding $Q_{ij}$ can be done as in the proof of Lemma A.2 by either rewriting the leading term of the Taylor expansion or by considering the leading matrix in the product and its geodesic distance to the identity.

The latter corollary gives an explanation for a typical numerical lost of a digit in synchronization error when acting on a full set of pure clean samples, see also Section 5. While it is true that in the noiseless case one should use an unbiased method, such as the spanning tree method (see Section 1), for noisy measurements, it might be beneficial to estimate the group elements with a biased estimator that incurs a smaller variance error. Synchronization via contraction, which does not reconstruct the solution in case of clean measurements, does lead to significant improvements in many more realistic scenarios, as we will see later in Section 5.

4.3 Notes on noisy synchronization

The analysis of synchronization in presence of noise, mostly in the case of the rotation group, has received a lot of attention in the past few years, e.g., [7, 50, 59, 60]. One common approach for analyzing global methods for synchronization is to study the spectrum of the measurement matrix, that is $M$ in $|1|$. The largest eigenvalues are associated with the synchronization solution, and are ideally separated from the rest of the spectrum which includes the noise. In other words, a gap between the largest eigenvalues and the rest of the spectrum means that a meaningful information is dominating the noise and a solution can be extracted. The eigenvalue separation fades away as noise increases and so is the accuracy of the solution. Other approaches for analysis include adapting tools from information theory, such as the analysis done in [50], or the construction of lower bounds such as Cramér-Rao [9]. Since the analysis of synchronization over compact groups has already been established, the key step in our analysis of synchronization via contraction lies in the compactification stage.

Let $g_{ij}$ be a noisy measurement, contaminated with a multiplicative noise matrix $N_{ij} \in G$,

$$g_{ij} = g_i N_{ij} g_j^{-1}. \quad (21)$$

Such a noise model is used in [9] and is often adapted by practitioners as well, for example over $G = \text{SE}(3)$ see [5]. In this multiplicative model we actually assume that the graph of the data $G$ is contaminated with noise on its edges (this is not always the case in applications, e.g., [52] but it facilitates the analysis comparing to the case of noisy vertices).

A spectral analysis of rotation synchronization, for the case of noise on edges, is given in [7, 50, Chapter 5]. The essence is as follows. Let $R$ be a block diagonal matrix, with its diagonal blocks $R_i$ are the unknown rotations. Then, each noisy measurement $g_{ij}$ can be viewed as the $ij$ block of $M = RWR^T$, where $W$ is a block matrix with the noise blocks $W_{ij}$. Since $M$ and $W$ are similar, the spectrum of $M$ is the same as $W$, which justifies studying $W$ for deriving conclusions on the spectrum of $M$. The goal is to deduce a condition that guarantees a separation of leading eigenvalues, as described above (meaning that the largest eigenvalues are not dominated by the noise). It is shown in [7, Chapter 5] that if $W_{ij}$ is drawn from a distribution that has a density function which is both a class function and centered around
Let Corollary 4.5.

Next, we study some analogue requirements on the noise block $N_{ij}$ to satisfy similar conditions after applying the contraction mapping $\Psi_\lambda$.

Denote the components of each group element by $g_i = (\mu_i, b_i)$ and $N_{ij} = (v_{ij}, a_{ij})$. Also, for simplicity we denote $\exp_\lambda(\cdot) = \exp(\cdot/\lambda)$. Then, a direct calculation of (21) by (9) yields

$$g_{ij} = \left(\mu_i v_{ij} \mu_j^T, b_i + \text{Ad}_{\mu_i} (a_{ij}) - \text{Ad}_{\mu_i v_{ij} \mu_j^T} (b_j)\right).$$

Some simplification using the commutativity of adjoint operator and the exponential map, and the definition of $\Psi_\lambda$ mean we can rewrite $r_{ij} = \Psi_\lambda(g_{ij}) \in G_0$ as

$$r_{ij} = \mu_i \left[\exp_\lambda(\text{Ad}_{\mu_i}^T(b_i) + a_{ij} - \text{Ad}_{v_{ij} \mu_j^T} (b_j))\right] \mu_j^T.$$ 

One interpretation as a noisy data matrix for synchronization is to consider $\hat{W}_{ij} = \exp_\lambda(\text{Ad}_{\mu_i}^T(b_i) + a_{ij} - \text{Ad}_{v_{ij} \mu_j^T} (b_j))v_{ij} \in G_0$ with the unknown $\{\mu_i\}_{i=1}^{n}$. However, further examination reveals that considering a slightly different matrix

$$W_{ij} = \exp_\lambda(-\text{Ad}_{\mu_i}^T(b_i))\exp_\lambda(\text{Ad}_{\mu_i}^T(b_i) + a_{ij} - \text{Ad}_{v_{ij} \mu_j^T} (b_j))\exp_\lambda(\text{Ad}_{v_{ij} \mu_j^T} (b_j))v_{ij}$$

leads to the more natural form $r_{ij} = \Psi_\lambda(g_{ij}) W_{ij} \Psi_\lambda(g_j)^T$.

To simplify $W_{ij}$ of (23), we apply the Baker-Campbell-Hausdorff (BCH) formula to have

$$\exp_\lambda(\text{Ad}_{\mu_i}^T(b_i) + a_{ij} - \text{Ad}_{v_{ij} \mu_j^T} (b_j))\exp_\lambda(\text{Ad}_{v_{ij} \mu_j^T} (b_j))v_{ij} = \exp((\text{Ad}_{\mu_i}^T(b_i) + a_{ij})/\lambda + X),$$

where $X = \frac{1}{2} \left[ (\text{Ad}_{\mu_i}^T(b_i) + a_{ij})/\lambda, (\text{Ad}_{v_{ij} \mu_j^T} (b_j))/\lambda \right] + \cdots$ is the series of nested commutators, with the leading term of order $\lambda^{-2}$

$$\lambda^{-2} \left( [\text{Ad}_{\mu_i}^T(b_i), \text{Ad}_{v_{ij} \mu_j^T} (b_j)] + [a_{ij}, \text{Ad}_{v_{ij} \mu_j^T} (b_j)] \right).$$

Then, a second application of the BCH formula implies that

$$W_{ij} = \exp(a_{ij}/\lambda + Y)v_{ij},$$

where $Y = \frac{1}{2} \left[ -\text{Ad}_{\mu_i}^T(b_i)/\lambda, (\text{Ad}_{\mu_i}^T(b_i) + a_{ij})/\lambda + X \right] + \cdots$ is the corresponding series of nested commutators. Combine $X$ with the above gives the explicit form of the commutator as

$$\lambda^{-2} \left( [a_{ij}, \text{Ad}_{\mu_i}^T(b_i)] \right) + \lambda^{-3} \left( [\text{Ad}_{\mu_i}^T(b_i), \text{Ad}_{v_{ij} \mu_j^T} (b_j)] + [a_{ij}, \text{Ad}_{v_{ij} \mu_j^T} (b_j)], \text{Ad}_{\mu_j}^T(b_i) \right).$$

The next order nested commutator term in $Y$ corresponds to $\lambda^{-3}$ and $\lambda^{-4}$ and so on for higher order terms. One conclusion is that if the translational part of the noise is of zero expectation, then (24) is in expectation a third order approximation to the Cartan decomposition of $W_{ij}$.

**Corollary 4.5.** Let $a_{ij}$ be a random variable independent of $\mu_i, \mu_j$ and $v_{ij}$, and satisfies $E[a_{ij}] = 0$. 

18
Then,
\[ \mathbb{E}[a_{ij} + Y] + \mathcal{O}(\lambda^{-4}) \in p. \]

**Proof.** The proof follows from two arguments regarding the series \( Y \). First, for any independent factor \( b \in g \), the commutator \([a_{ij}, b]\) is linear in \( a_{ij} \) at each entry. So, \( \mathbb{E}[a_{ij}] = 0 \) implies \( \mathbb{E}\left([a_{ij}, b]\right) = 0 \).

Second, we know that \([p, p] \subset t\) and \([p, t] \subset p\), meaning that the remaining nonzero third order nested commutators like \( [\text{Ad}_{\mu_i^T}(b_i), \text{Ad}_{\alpha_{ij}\mu_j^T}(b_j)], \text{Ad}_{\mu_i^T}(b_i) \] are in \( p \). Namely, the leading order outside \( p \) is of order \( \lambda^{-4} \).

The proof of Corollary 4.5 shows that if \( a_{ij} \) has a zero expectation, mapping the noise matrix \( \Psi_\lambda(N_{ij}) = \exp(a_{ij}/\lambda) v_{ij} \), which is already a matrix in its Cartan form, is not too far away from the Cartan form, in expectation, of \( W_{ij} \) of (24). For this case, which is a first order perturbation in \( \lambda \) to \( W_{ij} \), we can adapt conditions like (22) from rotation synchronization to synchronization over the Euclidean group. This case refers to a model where all samples are available (full graph of data). We use \( \circ \) for Hadamard (entrywise) matrix product.

**Proposition 4.6.** Denote by \( W_{ij} = \Psi_\lambda(N_{ij}) = \exp(a_{ij}/\lambda) v_{ij} \) the \( i,j \)th block of the noise matrix \( W \) in the rotations synchronization problem \( RWRR^T \). Assume all \( a_{ij} \), and \( v_{ij} \) are independent random variables over \( V = \mathbb{R}^d \) and \( K = SO(d) \) with probability density functions (PDF) \( f_V \) and \( f_K \), respectively. Furthermore, we assume each PDF is smooth and non-negative such that \( \mathbb{E}[v_{ij}] = \beta I, 0 \leq \beta \leq 1 \) and \( \mathbb{E}[a_{ij}] = 0 \) with even \( f_V \).

Then, the following condition guarantees the synchronization solution dominates the noise,

\[ \min\{(1 - \alpha_1)\beta, 1 - \alpha_2\} > \frac{1}{\sqrt{n}}. \]

Here \( \alpha_1 > 0 \) is \( \alpha_1 I = \int_{\|\mathbf{v}\| \leq \pi} \left(I \circ \mathbf{v}^T\right) h(\mathbf{v})d\mathbf{v} \), and \( \alpha_2 = \int_{\|\mathbf{v}\| \leq \pi} \|\mathbf{v}\|^2 h(\mathbf{v})d\mathbf{v} > 0 \), with respect to the density function \( h(x) = (1 - \cos(\|x\|))f_V(x)\left(\|I - xx^T\|^{-\frac{1}{2}}\right) \).

The proof is given in Appendix A.4.

The new condition of Proposition 4.6 shows a tradeoff between the amount of noise in the two parts of the data over \( SE(d) \) versus the sample size. In addition, it encodes the effect of geometry of the induced homogeneous space \( G_0/K \) as it includes two additional factors that depends on the new, modified density function \( h \). The condition shows that robustness to noise increases as the number of group elements is growing, as it allows more variance around the identity. The positivity of \( \alpha_1 \) and \( \alpha_2 \), implies the expectation values on (the diagonal parts of) \( \mathbb{E}[W_{ij}] \) are bounded by 1. Indeed, as \( f_V \) tends to be concentrated around 0, which means lower variance on the translational parts, \( \alpha_1 \) and \( \alpha_2 \) approach to zero and have a lesser effect on the condition. The expectation term \( \beta \) affects the upper \( d \times d \) block of each \( \mathbb{E}[W_{ij}] \), that is associated with the subgroup \( K \). As \( f_K \) gets more concentrated around the identity, \( \beta \) becomes closer to one. High variance on the element parts of \( K \) (in case of \( f_K \) that concentrated around the identity) implies smaller \( \beta \) values and more data is needed to guarantee valid information.

As a final note, one can apply the same proof technique as we use in Proposition 4.6 to prove a similar result for more general Cartan motion groups. Nevertheless, such a result might have a less explicit form, as both the Jacobian of local coordinates in \( G_0/K \) as well as the exponential map there can be less accessible or lead to more complicated expressions.
5 Numerical examples over special Euclidean groups

In this section we demonstrate numerically the method of contraction for rigid-motion synchronization.

5.1 Preliminaries

We firstly state how we address three preliminary issues: measuring the error, choosing the parameter $\lambda$ of the contraction, and optimizing the global rotations alignment. All algorithms and examples were implemented in Matlab and are available online in

https://github.com/nirsharon/Synchronization-over-SE

The numerical examples were executed on a Springdale Linux desktop equipped with 3.2 GHz i7-Intel Core™ processor with 16 GB of memory.

5.1.1 Measuring errors

In our examples we choose to evaluate the performance of the algorithms in terms of mean squared error (MSE). However, since the solution of synchronization is non-unique, we explicitly show how we measure this error. Consider $X = \{\hat{\gamma}_i\}_{i=1}^n$ to be the solution of one algorithm. Then, the measure of its performance is given by

$$\text{MSE}(X) = \frac{1}{n} \min_{\gamma \in \text{SE}(d)} \sum_{i=1}^{n} \|\hat{\gamma}_i \gamma - \gamma_i\|_F^2.$$  (25)

To calculate (25), let $\{\hat{\gamma}_i = (\hat{\mu}_i, \hat{b}_i)\}_{i=1}^n$ and $\{\gamma_i = (\mu_i, b_i)\}_{i=1}^n$ be the two compared sets. Then,

$$\sum_{i=1}^{n} \|\hat{\gamma}_i \gamma - \gamma_i\|_F^2 = \sum_{i=1}^{n} \|\hat{\mu}_i \mu - \mu_i\|_F^2 + \|b - \hat{\mu}_i^T (b_i - \hat{b}_i)\|_2^2,$$

$$= \left( nb^T b - 2b^T \sum_{i=1}^{n} \hat{\mu}_i^T (b_i - \hat{b}_i) \right) - 2 \text{tr} \left( \mu^T \left( \sum_{i=1}^{n} \hat{\mu}_i^T \mu_i \right) \right) + C,$$

where $C = \sum_{i=1}^{n} \|b_i - \hat{b}_i\|_2^2 + 4n$ is independent of $\gamma = (b, \mu)$. Therefore, since the cost function is separable, it is minimized by $b^* = \frac{1}{n} \sum_i \hat{\mu}_i^T (b_i - \hat{b}_i)$ and $\mu^* = V \begin{bmatrix} 1 & 0 \\ 0 & \text{det}(VU^T) \end{bmatrix} U^T$, where $\sum_i \hat{\mu}_i^T \mu_i = U \Sigma V^T$ is the SVD of $\sum_i \hat{\mu}_i^T \mu_i$ with $\Sigma_{1,1} \geq \Sigma_{2,2}$, see e.g., [28]. In other words, the error measure (25) is feasible to calculate.

5.1.2 Choosing $\lambda$

The basic tradeoff of varying $\lambda$ is as follows. For large parameter values ($\lambda \gg 1$), the translational part has smaller effect on the data after applying $\Psi_\lambda$ (In fact, by looking at the explicit form of the exponential map, as given in Appendix B.2, the entries of the exponential roughly perturbs the identity quadratically with respect the norm of the translational parts). On the other hand, in such cases the mapped measurements $\Psi_\lambda(g_{ij})$ are more consistent as data for synchronization in terms of satisfying the triplet relation $\Psi_\lambda(g_{ij})\Psi_\lambda(g_{j\ell}) \approx \Psi_\lambda(g_{i\ell})$. In contrary, small $\lambda$ values increase the information we gain from the translational part of the data but enforce solving non-consistent synchronization problem which often results in large errors. Given this tradeoff, it is clear that in ideal conditions, where data is free of
noise, the optimized value of \( \lambda \) lies in a certain segment inside \([1, \infty)\) and not too close to the boundaries. Two additional issues we would like to address in choosing the parameter of contraction are to fulfill \( (16) \) (that is keeping the data inside the injectivity domain of the exponential map) and coping with bad scaling effects of the translational parts of the data, if any. The latter two can be partially addressed by collecting fundamental descriptive statistics on the translational parts of the data.

We tested numerically the dependency between varying \( \lambda \) values and the synchronization error, resulted by our synchronization via contraction. As expected, this dependency is highly related to the different properties of the data (noise level, sparsity of the graph of data, etc.) as well as to the specific method one uses for the synchronization of rotations. To demonstrate the latter, we compare the optimal \( \lambda \) value, given as a function of the noise level, between two different rotation synchronization algorithms. One algorithm is the spectral method (see Algorithm 2) and the other is synchronization based on maximum-likelihood estimation (MLE) \([8]\). The ground truth were 100 elements, collected randomly on \( \text{SE}(3) \). We formed the data by considering a full data graph and contaminated the samples with a multiplicative noise, as in \( (21) \) (we elaborate more on that type of noise in the next subsection). The results are given in Figure 2a, where a clear difference is seen between the two cases. As an example of how the error changes in the vicinity of the optimal \( \lambda \) value, we present in Figure 2b the error as a function of \( \lambda \), for the case of MLE rotations estimator. One observed behavior of the error as a function of varying \( \lambda \) is that the optimal \( \lambda \) tends to be found closer to 1 as the noise increases. Limiting the search segment together with an error estimation based on an appropriate cost function (such as \( (5) \)) can be used to establish an efficient ad hoc optimization algorithm for finding the optimal \( \lambda \) in an arbitrary data scenario.

### 5.1.3 Optimize the global alignment in the inner rotations synchronization

Proposition 4.1 characterizes the conditions on \( Q \in G_0 \) such that

\[
\Psi^{-1}_\lambda(Q_i Q) \Psi^{-1}_\lambda((Q_j Q)^{-1}) = \Psi^{-1}_\lambda(Q_i) \Psi^{-1}_\lambda(Q_j)^{-1}.
\]

This means that a global alignment \( Q \) on \( G_0 \) also serves as a global alignment in \( G \), after the back mapping \( \Psi^{-1}_\lambda \). The conditions on \( Q \) means that it does not break the Cartan decomposition of \( Q_i \). Namely, if \( Q_i = \exp(b_i)\mu_i \) and \( Q = \exp(b)\mu \) are the global Cartan decompositions, their product, under
the conditions of Proposition 4.1 has the form \( \exp(b + \text{Ad}_\mu(b))\mu \), which is already in its Cartan form since \( \mu \in K \) and \( b + \text{Ad}_\mu(b) \in p \). Therefore, since multiplying from the right by any element from \( K \) is invariant for the back mapping, it is clear that for an arbitrary global alignment \( Q = \exp(b)\mu \in G_0 \), the problem is reduced to finding an optimal \( b \in p \) that minimizes some error after applying \( \Psi^{-1}_\lambda \).

In this numerical section we focus on the special case \( G = \text{SE}(d) \). For this case, the explicit form of the exponential map, as given in Appendix B.2 makes the back mapping accessible for fast applications. In particular, we can efficiently apply the back mapping to have

\[
\Psi^{-1}_\lambda (Q_i \exp(b)) = \Psi^{-1}_\lambda (\exp(b_i) \exp(\text{Ad}_\mu(b))\mu) = \begin{bmatrix} \tilde{\mu}_i & \tilde{b}_i \\
0_{1 \times d} & 1 \end{bmatrix},
\]

with the Cartan decomposition

\[
\exp(b_i) \exp(\text{Ad}_\mu(b)) = \exp(\tilde{b}_i)\tilde{\mu}_i, \quad \tilde{b}_i \in p, \quad \tilde{\mu}_i \in K.
\]

Then, the minimization is done only on \( \mathbb{R}^d \) and can be realized by a standard unconstrained optimization, by defining

\[
\min_{b \in \mathbb{R}^d} \sum_{(i,j) \in E} w_{ij} d\left( \begin{bmatrix} \tilde{\mu}_i & \tilde{b}_i \\
0_{1 \times d} & 1 \end{bmatrix}, \begin{bmatrix} \mu_j & b_j \\
0_{1 \times d} & 1 \end{bmatrix} \right),
\]

where \( d(\cdot, \cdot) \) is an appropriate metric on \( \text{SE}(d) \) and \((\mu_{ij}, b_{ij})\) are samples, that is the synchronization data. To keep the implementation of synchronization efficient, we limit this procedure to a fixed, small amount of iterations. For cases of large database of samples, a further time improvement can be done by adopting a random sampling strategy ("bootstrapping") to approximate the cost function by averaging the results of several partial sum calculations and avoid a large number of computations of Cartan decomposition needed for each iteration.

### 5.2 Synthetic Data

We study the performances of three main methods for synchronization over \( \text{SE}(d) \); our method of synchronization via contraction, the spectral method as done in [3, 6], and separation-based method, similar to the approach in [15] (solving first the rotational parts of synchronization and then use the result for estimating the translational parts). To make the comparison more realistic, we strengthen the separation-based algorithm by allowing it to solve the rotational part in the most suitable way, exactly as we do for our method after applying the mapping to rotations. We elaborate more on the choice of rotations synchronization when describing the specific examples. One common feature to those three methods is their availability for data with any given dimensionality \( d \). For more details on the abovementioned approaches see Subsection 2.2.

In the sequel, we generate synthetic random data as our ground truth elements, to be estimated by the synchronization algorithms. Every element is generated by projecting to \( \text{SO}(d) \) a \( d \times d \) matrix with entries chosen uniformly over \([0, 1]\), and a translation part with entries uniformly picked from \([0, 2]\). Having the ground truth data, we generate the relative relations \( g_ig_j^{-1} \) to serve as data for the synchronization problem. The three obstacles that we artificially create are missing data, added noise, and outliers in data. In the different scenarios we mix these conditions in several ways to imitate real world conditions.
In the first set of examples we examine the effect of missing data, added noise, and their combination on the performance of the synchronization algorithms. A common choice for characterizing of noise is to assume its Gaussian distribution. However, there is more than one way to define such distribution over a given Lie group. One approach is to define the PDF using the Euclidean multivariate Gaussian PDF on the Lie algebra (in terms of manifolds, on the tangent plane of the identity). This is done by clipping the tail of the Gaussian function or by folding the tails, also known as “wrapped Gaussian”, for more details see e.g., [14, Chapter 2]. This distribution, which is closely related to the von Mises distribution on the circle, facilitates the implementation of normal noise and therefore we adopt it on our experiments. To evaluate the noise we use Signal-to-Noise Ratio (SNR), which is a standard measure in signal processing. Given a data sample, the SNR is defined as ratio between deviations in data and deviations in noise. Also, it is usually measured in logarithmic scale that is decibels (dB) units. Since we use Gaussian noise over the Lie algebra, we measure the noise level there too. To be specific, recall our noisy measurements $g_{ij}$ are of the form (21), we use

$$\text{SNR} \left( \{g_{ij}\}_{(i,j) \in E} \right) = \frac{20}{|E|} \sum_{(i,j) \in E} \log_{10} \left( \frac{\| \log(g_{ij}^{-1}) \|}{\| \log(N_{ij}) \|} \right),$$

where $|E|$ stands for the number of edges in $E$. Measuring norms in the Lie algebra is also justified as it stands for the geodesic distance from the identity.

The other difficulty that we artificially create in our database is formed by making some of the measurements non-available. Potentially, for a set of $n$ unknown group elements, one has $\binom{n}{2}$ different meaningful measurements $g_{ij}$. However, it is often the case in applications that not all measurements are available. In the first set of examples, we consider cases with only a fraction of $p\binom{n}{2}$, $p \in [0,05, .5]$ available measurements. Theses examples are given in Figure 3 and consist of four examples correspondence to two different data sets. The ground truth are elements over the group $\text{SE}(4)$ ($d=4$, rigid motions over $\mathbb{R}^4$), and we generated two sets of lengths $n=100$ and $n=300$. For the rotation synchronization in both the contraction and separation algorithms we used the spectral method (see Algorithm 2). Figure 3a depicts the error as a function of the fraction of available data, contaminated with wrapped Gaussian noise of $\approx 12dB$. In this first example we use the data set of $n=100$ group elements. We see that the performance of the three methods is almost the same till approaching the level of 20% of available data. Then, the contraction based method shows superior error rates, where the separation method exhibits most inferior results. Now, focusing on the case of a fixed 25% of available data, we increase the noise level, down to almost 4dB (recall that lower dB means higher noise level). The error rates of this scenario are shown in Figure 3b, where we can clearly see the difference between the method of separation and the two other approaches that integrate the data as a whole. The contraction based method still achieves the best error rates. We repeated the above but with one major change, where we consider the data set of $n=300$ group elements. In this case, more information is obtained for the same percentage of available data compared to the case of $n=100$, as the amount of measurements grows quadratically with $n$. Indeed, as seen in Figure 3c, having the same level of noise of $\approx 12dB$, the error rates are far lower than the case $n=100$ in Figure 3a. As one approaches the level of 50%, the available data leads to very accurate estimations, where the spectral method is slightly better then the other two methods. On the other side, with small fractions of data, the contraction presents better error rates. As the last scenario on this set, we focus on
the 25% case and increase the noise level even higher, up to almost 2dB. The spectral and contraction methods present very similar results, up to around 4dB. Then, in higher levels of noise, the contraction shows much better error rates.

The second set of examples deals with the case of outliers in data. Each outlier is a randomly chosen SE(d) element, that replaces a certain measurement $g_{ij}$. We generate the outlier to be distinguished from the other measurements, specifically the rotational part is a projection of a square matrix of order $d$ with standard normal distributed entries to SO($d$), and a $d$ size vector of uniform entries over $[0, 1]$ for the translational part. These examples are done for $d = 2$, that is rigid motions in $\mathbb{R}^2$. In order to efficiently address cases of outliers we use for rotations synchronization the LUD algorithm [60], minimizing $\ell_1$ cost function. This is true for the contraction method as well as for the separation-based method. The results are given in Figure 4. We examine the error rates as a function of outliers, where the $x$-axis in each plot is the non-outliers rate, that is the fraction of available genuine data. The range of outliers is 10%−60%, meaning 40% to 90% of good measurements. As in the first set of examples, we used two sets of data, with $n = 100$ and $n = 300$ group elements in each. Starting with $n = 100$ in Figure 4a we observe significant advantage to the contraction method for values of around 40% of outliers. As outliers rate decreases both methods reaches almost zero error rates. On the other hand, when approaching 60% of outliers rates the spectral method, although its inferior results on other values, closes the gap and shows lowest error rates. The effect of using the LUD is significant also for the separation method, for the outliers rates of around
30% – 40%. We repeat the example, now with noise of $\approx 9dB$, that is added to the measurements. The results, depicted in Figure 4b show slightly higher error rates (as expected) compared to the previous example in Figure 4a. In addition, the gap between the performance of the spectral method and that of the separation method is closed and is almost the same for most values of outliers. Nevertheless, the contraction still shows superior performance in almost all values expect near the boundary values (almost no outliers or most outliers rates). Repeating the above examples with $n = 300$, the results are given in both Figure 4c (just outliers) and Figure 4d (outliers with added noise) and show the clear advantage of using contraction combined with LUD, where its performance is superior for any tested value of outliers rate.

5.3 Examples of real data — 3D point cloud registration

We further demonstrate the performance of our synchronization method for real data. The problem we address is multiple point-set registration in $\mathbb{R}^3$. In this problem we are given point clouds that represent raw scans of a three dimensional object. We aim to recover the transformations required to bring each range scan into a single coordinate system, in order to form an estimation to the original object. In our examples we have the original model object so we can calculate the ground truth transformations and compare them to the results we get from the different algorithms being tested.
The recovery process of the 3D object begins with estimating the relative motions between pairs of scans, computed by the Iterative Closest Point Algorithm (ICP), see e.g., [45]. In particular, for any two sets of point clouds we apply the ICP algorithm for several initial guesses, as done e.g., in [27, 54]. The ICP algorithm has two outputs, a rigid motion (an element from SE(3)) that transforms one point cloud to the other, and an estimated error of the mutual registration after applying the transform. We use the latter as an indicator for determining whether to include a given measurement or relative motion or not. This should be done carefully in order to eliminate both measurements of large deviations as well as outliers measurements. Since most of raw scans have small part of intersection, if any, just adding measurements would not improve the registration results but the opposite.

In general, there are two main approaches for solving the multiple point-set registration. First is to optimize a cost function that depends on bringing together corresponding points, e.g., [43]. The other approach, also known as frame space approach, is to solve the relations (rotations and translations) between mutual scans, e.g., [48]. We follow the second approach in order to highlight the synchronization algorithms and their performance for addressing the 3D registration problem. We compare the performance of six algorithms, including two variants of our contraction algorithm. Apart from the synchronization via contraction, the algorithms participants in the examples are: the spectral method and separation-based method, as done in the previous subsection. Another method is the diffusion method in [54] that is based upon the relation of SE(3) and dual quaternions, we termed this method QD in the tables. We also use the linear variant of contraction, done in the matrix level and based on polar decomposition, see Subsection 3.4. Both contraction methods as well as the separation methods use for rotations synchronization the maximum likelihood estimator (MLE) from [8]. At last we also include synchronization via contraction with the spectral method (Algorithm 2).

For our first example we use the frog model taken from the shape Repository of the Visualisation Virtual Services [423]. This model consists of 24 scans, each is a point cloud consists of between 24,000 to 36,000 points. The full model of the frog is presented in Figures 5a–5b, where the entire set of scans, each under different color, is depicted in Figure 5c. We repeat the comparison for two different sets of measurements; the first one has 80 pairwise measurements (from a total of potential 276, about 29%) and the second has 163 pairwise measurements. These measurements were chosen to minimize the errors according to the estimated error of each measurements, as generated by the ICP algorithm. This also means that adding more measurements increases the noise level. The results are given both numerically in Table 1 and visually in Figures 6–7. In Table 1 we notice that the three leading methods are the two contraction methods (both the group level contraction and the matrix level contraction) and the separation method. All three methods use MLE rotations synchronization. The method based on group contraction shows a slightly better numerical error, for both sets of measurements, however the visual differences in shape reconstruction is minor. The reconstructions (except of the one of QD which was too erroneous) are given in Figures 6a–6e, where the colors correspond to different scans, as shown in Figure 5c. The error of the synchronization via contraction, that is based on EIG, shows larger errors but still leads to a good visual reconstruction. However, both the spectral method and the diffusion method (dual quaternions) yield higher errors which were also very significant visually and do not provide good

\[^{1}\text{The code is courtesy of the authors of [54], see https://vision.in.tum.de/members/rodola/code}\]
reconstructions, see Figure 7. Last interesting result is that the additional measurements of the second set of data did not lead to better results, probably due to the extra noise it adds to the data. This is true for all methods except for the method based on group contraction combined with the spectral method for rotation synchronization which shows improvement in error.

The second example is registration over the “Skeleton Hand” model of the Georgia Tech large geometric models archive, [http://www.cc.gatech.edu/projects/large_models/hand.html](http://www.cc.gatech.edu/projects/large_models/hand.html). In this example we have nine scans, each of between 33,000 to 55,000 points, where the complete model holds 327,323 points. We use the same technique of deriving pairwise measurements, as done in previous example. Due to the small amount of scans, we use one optimal set of measurements which consists of 9 out of the 36 possible
Table 1: Numerical results over the "Frog" model. Set 1 includes 29% of available measurements, set 2 consists of 59% of available measurement.

![Image](image_url)

(a) Original object  (b) Raw scans  (c) Reconstruction

Figure 8: The “Skeleton Hand” model, its raw scans and their registration by synchronization via contraction.

pairwise measurements. The raw scans in this example are somehow more similar than the ones in the frog model, which leads to less variant in error. However, a very similar hierarchy of performance was observed in this example as well, as seen in Table 2. The synchronization via contraction, based on MLE rotations synchronization presents the least error result. Contrary to the previous example, the spectral method shows very good results as well. The reconstruction of the synchronization via contraction is given visually in Figure 8.

6 Conclusion

This paper focuses on applying a compactification process, that is, a mapping from a non-compact domain into a compact one, for solving synchronization problem. We illustrate this idea using contraction, a tool for compactification drawn from group theory. The contraction mappings enable to transform measurements from a Cartan motion group to an associated compact group. We then efficiently solve the reduced synchronization problem over the compact group, which ultimately provides a solution for synchronization in its original domain. We analyze the properties of synchronization via contraction and show its numerical efficiency in several different data scenarios.

As an extension to this work, it will be interesting to further consider contractions (or any alternative compactification methods) to solve approximation problems on non-compact groups other than synchronization, such as non-unique games [4], and to form a broader base of knowledge regarding the power of

Table 2: Numerical results over the "Hand" model.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Contraction (MLE)</th>
<th>Separation (MLE)</th>
<th>PD (MLE)</th>
<th>Spectral</th>
<th>Contraction (EIG)</th>
<th>QD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error on set 1</td>
<td>.001757</td>
<td>.00176</td>
<td>.001763</td>
<td>&gt; 0.01</td>
<td>.003891</td>
<td>&gt; 0.01</td>
</tr>
<tr>
<td>Error on set 2</td>
<td>.001758</td>
<td>.001759</td>
<td>.001816</td>
<td>&gt; 0.01</td>
<td>.002807</td>
<td>&gt; 0.01</td>
</tr>
</tbody>
</table>

![Image](image_url)
these methods. Furthermore, since compactification methods exist for general Lie groups, these methods might find use for addressing different classes of problems. As another perspective, we hope in future work to extend our analysis to more refined models of noise and data; for example, by considering other statistical measures in SE(d) or by changing the assumptions on the graph structure of the data.

Acknowledgement. The authors thank Amit Bermano for many useful discussions regarding 3D registration. The authors were partially supported by Award Number R01GM090200 from the NIGMS, FA9550-12-1-0317 and FA9550-13-1-0076 from AFOSR, Simons Foundation Investigator Award and Simons Collaborations on Algorithms and Geometry, and the Moore Foundation Data-Driven Discovery Investigator Award.

References


A Complementary proofs

A.1 Proof of Proposition 3.2

Proof. Let $T_\lambda = \begin{bmatrix} I_d & \frac{b/\lambda}{1} \\ 0_{1 \times d} & 1 \end{bmatrix}$, so that $g_\lambda = T_\lambda \begin{bmatrix} \mu & 0_{d \times 1} \\ 0_{1 \times d} & 1 \end{bmatrix}$. We rewrite the SVD of $g_\lambda$ in terms of the SVD of $T_\lambda = U_\lambda \Sigma_\lambda V_\lambda^T$, i.e., $g_\lambda = U_\lambda \Sigma_\lambda V_\lambda^T \begin{bmatrix} \mu & 0_{d \times 1} \\ 0_{1 \times d} & 1 \end{bmatrix}$. Here, the columns of $U_\lambda$ and $V_\lambda$ are the eigenvectors of $T_\lambda^T T_\lambda = I_{d+1} + \begin{bmatrix} \frac{b^T/\lambda^2}{b^T/\lambda} & b/\lambda \\ b^T/\lambda & 0 \end{bmatrix}$ and $T_\lambda^T T_\lambda = I_{d+1} + \begin{bmatrix} 0_{d \times d} & b/\lambda \\ b^T/\lambda & b^T b/\lambda^2 \end{bmatrix}$, respectively.
We observe that the subspace of vectors orthogonal to \( b \) is of dimension \( d - 1 \), and is the nullspace of both
\[
\begin{bmatrix}
\frac{b}{\lambda} b^T \lambda^2 & b/\lambda \\
\frac{b}{\lambda} b^T & 0
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
0_{d \times d} & b/\lambda \\
\frac{b}{\lambda} b^T / \lambda & 0
\end{bmatrix},
\]
Therefore, any set of vectors \( \left\{ \nu_i \right\}_{i=1}^{d-1} \) that forms an orthogonal basis for this subspace is also a set of orthogonal eigenvectors, with the eigenvalue \( 1 \), for \( T_\lambda T_\lambda^T \) and \( T_\lambda^T T_\lambda \).

To find the two remaining columns of \( U_\lambda \) and \( V_\lambda \), we make the ansatz that \( \begin{bmatrix} b/\lambda \alpha \\ \sqrt{\|b\|^2 / \lambda^2} \end{bmatrix} \) and \( \begin{bmatrix} b/\lambda \beta \\ \sqrt{\|b\|^2 / \lambda^2} \end{bmatrix} \) are the two remaining eigenvectors of \( T_\lambda T_\lambda^T \) and \( T_\lambda^T T_\lambda \), respectively. We also remember that the spectrum of the matrices are identical. Indeed, we get
\[
T_\lambda T_\lambda^T \begin{bmatrix} b/\lambda \\ \alpha \end{bmatrix} = \begin{bmatrix} (1+\alpha+\|b\|^2 / \lambda^2) b/\lambda \\ \alpha+\|b\|^2 / \lambda^2 \end{bmatrix} = \frac{1}{\lambda^2} \begin{bmatrix} (1+\alpha+\|b\|^2 / \lambda^2) b/\lambda \end{bmatrix},
\]
with the eigenvalues
\[
\alpha_{1,2} = -\frac{\|b\|^2 / \lambda^2}{\lambda^2} \pm \frac{\sqrt{\|b\|^4 / \lambda^4 + 4 \|b\|^2 / \lambda^2}}{2}.
\]
And also,
\[
T_\lambda^T T_\lambda \begin{bmatrix} b/\lambda \\ \beta \end{bmatrix} = \begin{bmatrix} (\beta+1) b/\lambda \\ (\|b\|^2 / \lambda^2 + 1) \beta+\|b\|^2 / \lambda^2 \end{bmatrix} = (\beta+1) \begin{bmatrix} b/\lambda \end{bmatrix},
\]
with the eigenvalues
\[
\beta_{1,2} = \frac{\|b\|^2 / \lambda^2}{\lambda^2} \pm \frac{\sqrt{\|b\|^4 / \lambda^4 + 4 \|b\|^2 / \lambda^2}}{2} = -\alpha_{2,1}.
\]
Denote the normalized remaining eigenvectors of \( U_\lambda \) as
\[
u_1 = \frac{1}{\sqrt{\|b\|^2 / \lambda^2 + \alpha_+^2}} \begin{bmatrix} b/\lambda \\ \alpha_+ \end{bmatrix}, \quad u_2 = \frac{1}{\sqrt{\|b\|^2 / \lambda^2 + \alpha_-^2}} \begin{bmatrix} b/\lambda \\ \alpha_- \end{bmatrix},
\]
and those of \( V_\lambda \) as
\[
u_1 = \frac{1}{\sqrt{\|b\|^2 / \lambda^2 + \beta_+^2}} \begin{bmatrix} b/\lambda \\ \beta_+ \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{\|b\|^2 / \lambda^2 + \beta_-^2}} \begin{bmatrix} b/\lambda \\ \beta_- \end{bmatrix}.
\]
As a result we get
\[
\Phi_\lambda(g) = U_\lambda V_\lambda^T \begin{bmatrix}
\mu & 0_{d \times 1} \\
0_{1 \times d} & 1
\end{bmatrix}
\]
\[
= \begin{bmatrix}
\sum_{i=1}^{d-1} \nu_i v_i^T & 0_{d \times 1} \\
0_{1 \times d} & 0
\end{bmatrix} + v_1 r_1^T + v_2 r_2^T \begin{bmatrix}
\mu & 0_{d \times 1} \\
0_{1 \times d} & 1
\end{bmatrix}
\]
\[
= \begin{bmatrix}
2\tau \lambda b b^T + P & \mu \\
-\frac{\mu}{\lambda} b^T & 2\tau \lambda
\end{bmatrix}.
\]
Note that a direct calculation shows that the columns of \( \Phi_\lambda(g) \) are indeed orthonormal. The explicit
form of $\Phi_\lambda$ also implies positive determinant of $\Phi_\lambda (g)$ since for blocks $A, B, C, D$ it holds that

$$\det( \begin{bmatrix} A & B \\ C & D \end{bmatrix} ) = \det(A - BD^{-1}C) \det(D).$$

In addition, $\lambda, \tau_\lambda > 0$, $\mu \in \text{SO}(d)$, and the spectrum of $P$, as a projection operator, is non-negative.

A.2 Proof of Proposition 3.4

Proof. Starting with the inverse claim of the approximated homomorphism, we have that for $\gamma_\lambda = V_\lambda \Sigma_\lambda W_\lambda^T$ we also get $\gamma_\lambda^{-1} = W_\lambda \Sigma_\lambda^{-1} V_\lambda^T$. Therefore,

$$\Phi_\lambda \left( \gamma^{-1} \right) = W_\lambda V_\lambda^T = \left( V_\lambda W_\lambda^T \right)^{-1} = (\Phi_\lambda(\gamma))^{-1}.$$

For the second part of the approximated homomorphism, observe that $\tau_\lambda$ of Proposition 3.2 satisfies $\tau_\lambda = \frac{1}{2} + O\left( \frac{1}{\lambda^2} \right)$. Thus,

$$\left\| 2\tau_\lambda \hat{b} \hat{b}^T + \sum_{i=1}^{d-1} \nu_i \nu_i^T - I_k \right\|_F = O\left( \frac{1}{\lambda^2} \right).$$

As a consequence, we get that $\left\| \Phi_\lambda(\gamma_1 \gamma_2) - \Phi_\lambda(\gamma_1) \Phi_\lambda(\gamma_2) \right\|_F$ is equal to

$$\left\| \begin{bmatrix} \mu_1 \mu_2 - \frac{b_1 b_2^T}{2\lambda} & \mu_1 b_2 + b_1 \\ \frac{b_1^T \mu_1 + b_2^T \mu_2}{2\lambda} & 1 - \frac{b_1^T \mu_1 + b_2^T \mu_2}{2\lambda} \end{bmatrix} - \begin{bmatrix} \mu_1 \mu_2 & \frac{b_1}{2\lambda} \\ -\frac{b_1^T \mu_1}{2\lambda} & 1 \end{bmatrix} \begin{bmatrix} \mu_2 & \frac{b_2}{2\lambda} \\ -\frac{b_2^T \mu_2}{2\lambda} & 1 \end{bmatrix} \right\|_F + O\left( \frac{1}{\lambda^2} \right),$$

which also has $\frac{1}{\lambda^2}$ as a leading term in $\lambda$, as required. 

A.3 Proof of Proposition 4.2

To prove Proposition 4.2 we use two auxiliary Lemmas.

Lemma A.1. Suppose $Q_1, Q_2$ and $P_1, P_2$ are orthogonal matrices, such that

$$\|Q_1 - Q_2\|_F \leq \varepsilon_1, \quad \text{and}, \quad \|P_1 - P_2\|_F \leq \varepsilon_1.$$

Then,

$$\|Q_1 P_1 - Q_2 P_2\|_F \leq \varepsilon_1 + \varepsilon_2.$$
Proof. We use the orthogonality invariance of the Frobenius norm and the triangle inequality to get
\[
\|Q_1P_1 - Q_2P_2\|_F = \left\| I - P_1^TQ_1^TQ_2P_2 \right\|_F = \left\| I + P_1^T(I - Q_1^TQ_2)P_2 - P_1^TP_2 \right\|_F 
\leq \left\| I - P_1^TP_2 \right\| + \left\| P_1^T(I - Q_1^TQ_2)P_2 \right\|_F = \|P_1 - P_2\|_F + \|Q_1 - Q_2\|_F \leq \epsilon_1 + \epsilon_2.
\]

The next lemma bounds the distance between two exponential maps with respect to the original distance in the algebra. The two distances are linked by a term involves the commutator (which also related to the curvature of the Lie group manifold).

**Lemma A.2.** Let \( p_1, p_2 \in g \), and let \( \lambda \) be sufficiently large such that
\[
\exp(p_1/\lambda), \exp(p_2/\lambda), \text{ and } \exp(p_1/\lambda)\left(\exp(p_2/\lambda)\right)^{-1}
\]
are within the injectivity radius around the identity. Then,
\[
\|\exp(p_1/\lambda) - \exp(p_2/\lambda)\| \leq \|p_1 - p_2\| + C(\lambda^{-2}),
\]
where the constant \( C \) depends on the commutator \([p_1, p_2]\) (and thus can be bound by \( \max\|p_1\|, \|p_2\| \)) but independent on \( \lambda \).

**Proof.** Since \( G_0 \) is compact, the Lie group exponential map is surjective and there is \( q \in g \) such that \( \exp(q) = \exp(p_1/\lambda)\left(\exp(p_2/\lambda)\right)^{-1}. \) Now, the compactness also implies that \( G_0 \) admits a bi-invariant metric, see e.g., [1, Chapter 2], and thus the induced Riemannian metric \( d_R(\cdot, \cdot) \) satisfies
\[
d_R(\exp(p_1/\lambda), \exp(p_2/\lambda)) = d_R(\exp(q), I) = \|q\|.
\]
However, from Baker-Campbell-Hausdorff (BCH) formula we get
\[
q = \frac{1}{\lambda}(p_1 - p_2) + O\left(\left[\frac{p_1}{\lambda}, \frac{p_2}{\lambda}\right]\right).
\]
The result of the lemma follows since the geodesic Riemannian distance bounds the Euclidean distance and since \( \frac{1}{\lambda} \leq 1 \), that is
\[
\|\exp(p_1/\lambda) - \exp(p_2/\lambda)\| \leq d_R(\exp(p_1/\lambda), \exp(p_2/\lambda)) \leq \|p_1 - p_2\| + \frac{1}{\lambda^2}O\left(\|\left[ p_1, p_2 \right]\|\right).
\]

We now turn to prove the proposition.
Proof of Proposition 4.4. By the approximated homomorphism (12) we have
\[ \|\Psi_\lambda(g_1)\Psi_\lambda(g_2)^{-1} - \Psi_\lambda(g)\|_F = \|\Psi_\lambda(g_1)\Psi_\lambda(g_2)^{-1} - \Psi_\lambda(g)\|_F. \]
Then, using the second part of (12) and triangle inequality we get
\[ \|\Psi_\lambda(g_1)\Psi_\lambda(g_2)^{-1} - \Psi_\lambda(g)\|_F = \|\Psi_\lambda(g_1)\Psi_\lambda(g_2)^{-1} - \Psi_\lambda(g_1g_2^{-1}) + \Psi_\lambda(g_1g_2^{-1}) - \Psi_\lambda(g)\|_F \]
\[ \leq O(\lambda^{-2}) + \|\Psi_\lambda(g_1g_2^{-1}) - \Psi_\lambda(g)\|_F. \]
Denote the decompositions \( g_1g_2^{-1} = (k, v), g = (\bar{k}, \bar{v}) \in G \) and by (11)
\[ \|\Psi_\lambda(g_1g_2^{-1}) - \Psi_\lambda(g)\|_F = \|k \cdot \exp(\lambda v) - \bar{k} \cdot \exp(\lambda \bar{v})\|_F. \]
Now, recall that \( k, \exp(\lambda v), \bar{k}, \exp(\lambda \bar{v}) \in G_0 \) and thus they are represented by orthogonal matrices. Apply Lemma A.1
\[ \|\Psi_\lambda(g_1g_2^{-1}) - \Psi_\lambda(g)\|_F \leq \|k - \bar{k}\|_F + \|\exp(\lambda v) - \exp(\lambda \bar{v})\|_F. \]
Assume the elements are close enough so we can apply Lemma A.2 to finally obtain
\[ \|\Psi_\lambda(g_1)\Psi_\lambda(g_2)^{-1} - \Psi_\lambda(g)\|_F \leq \|k - \bar{k}\|_F + \|v - \bar{v}\|_F + O(\lambda^{-2}). \]

\[ \square \]

A.4 Proof of Proposition 4.6

Proof. We follow the analysis of the phase transition point in the spectral method, as done in [7, Chapter 5.3] for the case of a full set of measurements. This proof uses the estimation of maximal eigenvalue of random block matrix in [24]. In essence, \( M = RWRT \) and \( W \) share the same full rank spectrum. Therefore, it is easier to approach the largest eigenvalues of the data matrix \( M \) by analyze \( W \). In particular, setting \( W_{ii} = 0 \) and since \( M \) is a symmetric matrix, we also have that \( Y = W - \mathbb{E}[W] \) is a symmetric block matrix, with zero mean and \( Y_{ii} = 0 \). These conditions together with the fact that \( W \) has a fixed Frobenius norm (\( W_{ij} \) are orthogonal matrices) give the general structure needed to apply the results in [24] for estimating the largest eigenvalue in \( Y \). There are two specific conditions to show. First is the independence between the blocks which we have from the independence assumptions on \( a_{ij} \) and \( v_{ij} \). Second is the expectation over \( W_{ij} \). To calculate this expectation, we use \( f_V \) and \( f_K \) and define a joint PDF \( f \) based upon the global Cartan decomposition \( G_0 \cong P \cdot K \),
\[ f(g) = f_V(\log(p))f_K(k), \quad g = p \cdot k \in G_0. \]
Note that \( K \) acts transitively on \( P \) so \( f \) is well-defined. The coset \( X = G_0/K \) is a symmetric homogeneous space equipped with a Riemannian metric so integration over \( X \) is well-understood, and we can use it to separate the integration on \( G_0 \). Also, \( f \) is separable by definition and smooth. We use the Haar measure.
and the fact that $G_0$ is compact to have

$$\mathbb{E} [W_{ij}] = \int_{G_0} g f(g) d\mu(g) = \int_X p f_V(\log(p)) d\mu(p) \int_K k f_K(k) d\mu(k).$$

The most right integral is equal by assumption to $\left[ \mathbb{E}[^{\nu_{ij}}_{\nu}] 0_{d \times 1} \right]$. For the second integral, we change coordinates to have

$$\int_X p f_V(\log(p)) d\mu(p) = \int_{\|v\| \leq \pi} \exp(v) f_V(v) \|J(v)\|^{-1} d\mu(v).$$

Here $v$ in $f_V(v)$ is understood as a vector and inside the exponential as a matrix. For the case of infinitesimal displacement on the sphere, since it is a rank one coset we have an explicit form for the Jacobian, $(I - vv^t)\frac{1}{2}$, see e.g., [22, Chapter 9, p. 391]. To simplify the exponential we use the result obtained in [B.2] to deduce that only three terms are involved in $\exp(v)$, that are $I$, $v$ and $v^2$ (here, in its matrix form). In addition, the function $h(v) = f_V(v) \| (I - vv^t)^{-\frac{1}{2}} \|$ is even, which leaves us with only two terms, the identity and the diagonal part of $v^2$ (the parts that include the quadratic terms) as all other terms lead to odd Integrands over a symmetric domain around the identity. To conclude, the expectation of each noise block is a diagonal matrix of the form

$$\mathbb{E} [W_{ij}] = \left[ \begin{array}{cc}
(1 - \alpha_1) & 0_d \\
0_d & \alpha_2
\end{array} \right].$$

where $\alpha_1$ and $\alpha_2$ as described in the claim of the proposition. Note that although $\mathbb{E} [W_{ij}]$ not a scalar matrix (the last entry on the diagonal is not necessary equals to the others), the conclusions about the largest eigenvalues $Y$ from [7, Chapter 5.3] are still valid. \hfill \square

### B Supplements

#### B.1 The spectral method – pseudocode algorithm

The spectral method is given as a pseudocode in Algorithm [2]. For more details on this algorithm see [50].

#### B.2 Calculating the Cartan decomposition in special orthogonal group

To calculate $\Psi^{-1}_\lambda$ we have to reveal the Cartan decomposition of an element from $G_0$. In the algebra level, the Cartan decomposition is straightforward since it induced from a direct sum of linear spaces. However, in the group level this calculation is less trivial. To make the discussion concrete, consider $G_0 = \text{SO}(d+1)$ and denote by $Q$ the element in $G_0$ to be decomposed. Then, the Cartan decomposition consists of two matrices $P, R \in \text{SO}(d+1)$ such that

$$Q = PR, \quad P = \exp(p), \quad p \in p, \quad R = \left[ \begin{array}{cc}
\mu & 0_d \\
0_d & 1
\end{array} \right], \quad \mu \in \text{SO}(d).$$

(26)
Algorithm 2 Spectral method for data on a semisimple compact group $G$

**Require:** Ratio measurements $\{g_{ij}\}_{(i,j) \in E}$.
Measurement confidence weights $\{w_{ij}\}_{(i,j) \in E}$.
Non-trivial orthogonal faithful representation $\rho$ of order $d$ over $G$.

**Ensure:** An approximated solution for $\{g_i\}_{i=1}^n$.

1: Initialize $M$ and $D$ as $dn \times dn$ zeros matrices.
2: for $(i,j) \in E$ do
3: Define block $M_{ij} \leftarrow w_{ij} \rho(g_{ij})$.
4: for $\{\ell \mid (i,\ell) \in E\}$ do
5: Define block $D_{ii} \leftarrow D_{ii} + w_{i\ell}I_d$.
6: end for
7: end for
8: $H \leftarrow D^{-1}M$.
9: $\mu \leftarrow \text{EIG}(H, d)$. \{Extract top $d$ eigenvectors\}
10: for $i = 1, \ldots, n$ do
11: $b_i \leftarrow \text{BLOCK}(\mu, d, i)$. \{The $i$th block\}
12: $\mu_i \leftarrow \text{ROUND}(b_i)$. \{Rounding procedure\}
13: end for
14: return $\{\rho^{-1}(\mu_i)\}, i = 1, \ldots, n$.

Here $p$ is the subspace given from the Cartan decomposition in the algebra level. Let $q = \log(Q)$ be the correspondence Lie algebra element of $Q$, and denote its Cartan decomposition by

$$q = q_t + q_p, \quad q_t \in t, \quad q_p \in p.$$  

The difficulty in decompose at group level is that in general, $q_t \neq r = \log(R)$ and similarly $q_p \neq p$, since by BCH

$$q_t + q_p = p + r + \frac{1}{2}[p,r] + \frac{1}{12}([p,[p,r]] + [r,[r,p]]) + \ldots. \quad (27)$$  

From the other hand, it is not enough to work only in the group level and decompose $Q$ as $\hat{P}[\hat{\mu}]_{0 \times d}$ (this can be done, for example, by applying Givens rotations to $Q$) does not guarantee that $\log \hat{P} \in p$.

One approach in literature is to approximate $p$ and $r$ using the observation that the right hand side of (27) can be divided into two types of terms: ones from $p$, that are $p, \frac{1}{2}[p,r], \frac{1}{12}[r,[r,p]], \ldots$ and the other terms from $t$. Then, truncated BCH series can be applied which also requires solving a non-linear polynomial equation in either $r$ or $p$, see e.g., [18]. Another approach is to optimize the decomposition. In details, there are only $d$ degrees of freedom in determine the desired element of the linear subspace $p$, and there is a unique solution in the neighbourhood of zero there. Therefore, one can use zero as initial guess in the unconstrained optimization

$$\min_{p \in p} \|U \exp(-p)Qu\|_F^2 \quad (28)$$

where $U = \begin{bmatrix} I_{d \times d} & 0 \\ 0_{1 \times d} & 0 \end{bmatrix}$ and $u$ is the unit vector in $\mathbb{R}^{d+1}$ with one in its last entry. By extracting the gradient of the cost function in (28), there are efficient gradient-based methods, such as the trust region algorithm, to efficiently solve (28). Testing this approach numerically, the solution $p^*$ is typically found, to double
precision, in about 20 iterations.

The most efficient approach we have for the case of Cartan decomposition of rotations is based upon generalizing the Rodrigues formula. In particular, we can exploit the special structure of

\[ p = \begin{bmatrix} 0 & -b^T \\ b & 0 \end{bmatrix} \in \mathbb{R}^d, \]

where \( b \in \mathbb{R}^d \), to have the Rodrigues formula for any \( d \) (and not just for \( P = \exp(p) \) of order 3). By \([15]\) we learn that the only nonzero eigenvalues of \( p \) are \( \pm i \theta \) with \( \theta = \|b\| \). Therefore, inspired from the proof of the Rodrigues formula in \([21]\), the key observation is that the normalized \( p_{\theta} = \frac{1}{\theta} p \) satisfies \( p_{\theta}^3 = -p_{\theta} \). This leads to the more general relations \( p_{\theta}^{2m+j} = (-1)^m p_{\theta}^j \) for \( j = 1, 2 \) and \( m \in \mathbb{N} \), which imply the Rodrigues formula

\[
\exp(p) = \sum_{n=0}^{\infty} \frac{p^n}{n!} = I + \frac{\theta p_{\theta}}{1!} + \frac{\theta^2 p_{\theta}^2}{2!} + \ldots
\]

\[
= I + \left( \frac{\theta}{1!} - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \ldots \right) p_{\theta} + \left( \frac{\theta^2}{2!} - \frac{\theta^4}{4!} + \ldots \right) p_{\theta}^2
\]

\[
= I + \sin(\theta)p_{\theta} + (1 - \cos(\theta))p_{\theta}^2.
\]

The block structure of \( p \) indicates that \( p_{\theta} \) and \( p_{\theta}^2 \) do not have any common places of nonzero blocks. Therefore, the first \( d \) entries in the last column of \( \exp(p) \) are exactly as in \( \sin(\theta)p_{\theta} \) and the last entry (on the diagonal) is affected by \( I \) and \( p_{\theta}^2 \) and equals to \( 1 + (1 - \cos(\theta))(p_{\theta}^2)_{d+1,d+1} \), which is actually just \( \cos(\theta) \) since the normalization implies \( (p_{\theta}^2)_{d+1,d+1} = -1 \). The last column is of particular interest as by \([26]\) we deduce that the last columns of \( \exp(p) \) and \( Q \) are identical.

We summarize the above in the following procedure of Cartan decomposition, given \( Q \) of \([26]\):

1. Calculate \( \theta = \cos^{-1}(Q_{d+1,d+1}) \).

2. For \( 0 < \theta < \pi \), denote by \( Q_{1:d,d+1} \) the first \( d \) entries of the last column of \( Q \), and derive \( b \) using

\[
\frac{\sin(\theta)}{\theta} b/\lambda = Q_{1:d,d+1}.
\]

3. Extract \( R = \exp(-b/\lambda)Q \). The exponential can be efficiently calculated by reusing the above Rodrigues formula for \( \exp(-p) \).

4. Special cases: \( \theta = 0 \) means \( b = 0 \) and \( R = Q \). On the other hand, \( \theta = \pi \) implies we are on the boundary of the diffeomorphism domain of the exponential map and there is no unique decomposition (which indicates that \( \lambda \) was wrongly chosen initially).