Article

The International Journal of
Robotics Research
2019, Vol. 38(2-3) 95–125

DOI: 10.1177/0278364918784361
journals.sagepub.com/home/ijr
SAGE

SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group

David M Rosen[1], Luca Carlone[2], Afonso S Bandeira[3], and John J Leonard[4]

Abstract

Many important geometric estimation problems naturally take the form of synchronization over the special Euclidean group*: estimate the values of a set of unknown group elements $x_1, \ldots, x_n \in \mathrm{SE}(d)$ given noisy measurements of a subset of their pairwise relative transforms $x_i^{-1}x_j$. Examples of this class include the foundational problems of pose-graph simultaneous localization and mapping (SLAM) (in robotics), camera motion estimation (in computer vision), and sensor network localization (in distributed sensing), among others. This inference problem is typically formulated as a non-convex maximum-likelihood estimation that is computationally hard to solve in general. Nevertheless, in this paper we present an algorithm that is able to efficiently recover* certifiably globally optimal *solutions of the special Euclidean synchronization problem in a non-adversarial noise regime. The crux of our approach is the development of a semidefinite relaxation of the maximum-likelihood estimation (MLE) whose minimizer provides an* exact *maximum-likelihood estimate so long as the magnitude of the noise corrupting the available measurements falls below a certain critical threshold; furthermore, whenever exactness obtains, it is possible to* verify *this fact* a posteriori*, thereby* certifying *the optimality of the recovered estimate. We develop a specialized optimization scheme for solving large-scale instances of this semidefinite relaxation by exploiting its low-rank, geometric, and graph-theoretic structure to reduce it to an equivalent optimization problem defined on a low-dimensional Riemannian manifold, and then design a Riemannian truncated-Newton trust-region method to solve this reduction efficiently. Finally, we combine this fast optimization approach with a simple rounding procedure to produce our algorithm,* SE-Sync*. Experimental evaluation on a variety of simulated and real-world pose-graph SLAM datasets shows that SE-Sync is capable of recovering certifiably globally optimal solutions when the available measurements are corrupted by noise up to an order of magnitude greater than that typically encountered in robotics and computer vision applications, and does so significantly faster than the Gauss–Newton-based approach that forms the basis of current state-of-the-art techniques.*

1. Introduction

The preceding decades have witnessed remarkable advances in both the capability and reliability of autonomous robots. As a result, robotics is now on the cusp of transitioning from an academic research project to a ubiquitous technology in industry and in everyday life, with tremendous potential in applications such as transportation (Leonard et al., 2008; Thrun et al., 2006; Urmson et al., 2008), medicine (Burschka et al., 2005; Taylor, 2006; Taylor et al., 2008), and disaster response (Fallon et al., 2015; Johnson et al., 2015; Pratt and Manzo, 2013) to increase productivity, alleviate suffering, and preserve life.

At the same time, however, the very applications for which robotics is poised to realize the greatest benefit typically carry a correspondingly high cost of poor performance. For example, in the case of transportation, the failure of an autonomous vehicle to function correctly may lead to destruction of property, severe injury, or even loss of life. This high cost of poor performance presents a serious barrier to the widespread adoption of robotic technology in the high-impact but safety-critical applications that we would most like to address, absent some guarantee of "good behavior" on the part of the autonomous agent(s).

[1]Oculus Research, Redmond, WA, USA
[2]Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Department of Mathematics and Center for Data Science, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA
[4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Corresponding author:
David M Rosen, Oculus Research, Redmond, WA 91809, USA.
Email: david.rosen@oculus.com

While such guarantees (namely of correctness, optimality, bounded suboptimality, etc.) have long been a feature of algorithms for planning (Russel and Norvig, 2010) and control (Åström and Murray, 2008; Stengel, 1994), to date it appears that the development of practical algorithms with clearly delineated performance guarantees for robotic *perception* has to a great extent remained an open problem.

This paper presents one such method, *SE-Sync*, for solving the fundamental perceptual problem of *pose estimation*. Formally, we address the problem of *synchronization over the special Euclidean group*: estimate the values of a set of unknown group elements $x_1, \ldots, x_n \in \mathrm{SE}(d)$ given noisy measurements of a subset of their pairwise relative transforms $x_i^{-1}x_j$.[1] This estimation problem lies at the core of many important geometric perception tasks in robotics and computer vision, including pose-graph simultaneous localization and mapping (SLAM) (Grisetti et al., 2010; Stachniss et al., 2016), camera motion and/or orientation estimation (Arrigoni et al., 2016; Hartley et al., 2013; Tron et al., 2016), and sensor network localization (Peters et al., 2015), among others.

In the context of robotics, current state-of-the-art algorithms for solving the special Euclidean synchronization problem as it arises in pose-graph SLAM formulate the problem as an instance of maximum-likelihood estimation under an assumed probability distribution for the measurement noise, and then apply general first- or second-order smooth numerical optimization methods (Nocedal and Wright, 2006) to estimate a critical point of the objective function (Grisetti et al., 2010; Stachniss et al., 2016). While this approach has been crucial in enabling the development of fast and scalable SLAM inference methods (thereby reducing SLAM to a practically solvable problem), its reliance upon *local* search techniques leaves the resulting algorithms vulnerable to convergence to significantly suboptimal critical points. Indeed, it is not difficult to find even fairly simple real-world examples where estimates recovered by local search methods can be so poor as to be effectively unusable (Figure 1), even for relatively low levels of measurement noise (Carlone et al., 2015a). Given the crucial role that the estimates supplied by SLAM systems play in enabling the basic functions of mobile robots, this lack of reliability in existing SLAM inference methods represents a serious barrier to the development of robust autonomous agents generally.

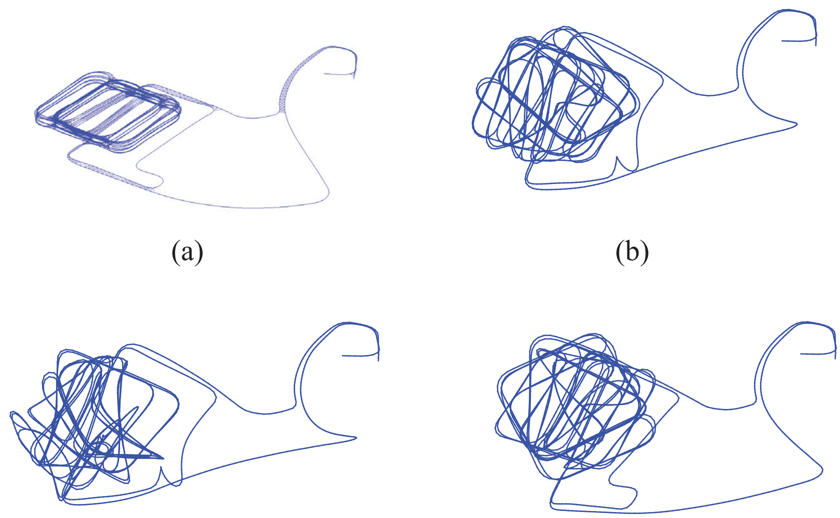
SE-Sync ameliorates this lack of reliability by enabling the efficient recovery of *provably globally optimal* pose-graph SLAM solutions under realistic operating conditions. Our algorithm exploits a convex *semidefinite relaxation* of the special Euclidean synchronization problem whose minimizer provides an *exact* maximum-likelihood estimate so long as the magnitude of the noise corrupting the available measurements falls below a certain critical threshold; furthermore, whenever exactness obtains, it is possible to *verify* this fact *a posteriori*, thereby *certifying* the optimality of the recovered estimate. SE-Sync thus belongs to the class of *certifiably correct* algorithms (Bandeira, 2016), meaning that it is capable of efficiently solving a generally intractable problem within a restricted (but still practically relevant) operational regime, and of *certifying* the correctness of the solutions that it recovers. In the case of our algorithm, experimental evaluation on a variety of simulated and real-world pose-graph SLAM datasets shows that SE-Sync is capable of recovering certifiably globally optimal pose-graph SLAM estimates when the available measurements are corrupted by noise up to an order of magnitude greater than that typically encountered in robotics and computer vision applications, and does so significantly faster than the Gauss–Newton-based approach that forms the basis of current state-of-the-art (*local*) search techniques.

The rest of this paper is organized as follows. In the next section we provide an overview of prior work on the special Euclidean synchronization problem, with a particular emphasis on its instantiations in robotics and computer vision applications. Section 3 introduces the specific formulation of the special Euclidean synchronization problem that we will address in the remainder of the paper. In Section 4, we derive the semidefinite relaxation that will form the basis of our approach, and prove that its minimizer provides an *exact* maximum-likelihood estimate for sufficiently small measurement noise. Section 5 develops a fast optimization scheme for solving large-scale instances of the semidefinite relaxation efficiently. Section 6 provides experimental results demonstrating the advantages of our algorithm in comparison with the Gauss–Newton-based approach that forms the basis of current state-of-the-art pose-graph SLAM methods. Finally, Section 7 concludes with a summary of this paper's contributions and a discussion of future research directions.

2. Related work

In this section we provide an overview of prior work on the special Euclidean synchronization problem (and its relatives), focusing in particular on its instantiations in robotics and computer vision applications.

The standard formulation of the general group synchronization problem formalizes it as an instance of maximum-likelihood estimation under an assumed probability distribution for the measurement noise (Singer, 2011). This formulation is attractive from a theoretical standpoint, due to the powerful analytical framework and strong performance guarantees that maximum-likelihood estimation affords (Ferguson, 1996). However, this formal rigor often comes at the expense of computational tractability, as it is frequently the case that the optimization problem underlying an instance of maximum-likelihood estimation is non-convex, and therefore computationally hard to solve in general. Unfortunately, this turns out to be the case for the special Euclidean synchronization problem in particular, due to the non-convexity of $\mathrm{SE}(d)$ itself.



(a) (b)

(c) (d)

Fig. 1. Examples of suboptimal estimates in pose-graph SLAM. Several estimates are shown for the trajectory of a robot as it enters and explores a multi-level parking garage, obtained as critical points of the corresponding pose-graph SLAM maximum-likelihood estimation: (a) globally optimal trajectory estimate, obtained using SE-Sync; (b), (c), and (d) suboptimal critical points of the MLE obtained using local search.

Given the fundamental computational hardness of nonconvex optimization, prior work on special Euclidean synchronization (and related estimation problems) has predominantly focused on the development of *approximate* algorithms that can efficiently compute high-quality estimates in practice. These approximate algorithms can be broadly categorized into two classes.

The first class consists of algorithms that are based upon the (heuristic) application of fast *local* search techniques to identify promising estimates. This approach has proven particularly attractive in robotics and computer vision applications, as the high computational speed of first- and second-order smooth nonlinear programming methods (Nocedal and Wright, 2006), together with their ability to exploit the measurement sparsity that typically occurs in naturalistic problem instances (Dellaert and Kaess, 2006), enables these techniques to scale effectively to large problems while maintaining real-time operation with limited computational resources. Indeed, there now exist a variety of mature algorithms and software libraries implementing this approach that are able to process special Euclidean synchronization problems involving tens to hundreds of thousands of latent states in real time using only a single thread on a commodity processor (Dellaert et al., 2010; Grisetti et al., 2009; Kaess et al., 2012; Konolige, 2010; Kümmerle et al., 2011; Lourakis and Argyros, 2009; Olson et al., 2006; Rosen et al., 2014). However, the restriction to *local* search leaves these methods vulnerable to convergence to significantly suboptimal critical points, even for relatively low levels of measurement noise (Carlone et al., 2015a). Some recent research has attempted to more explicitly tackle the problem of suboptimal convergence, including the pursuit of methods with larger basins of attraction to favorable solutions (Grisetti et al., 2009; Olson et al., 2006), advanced initialization techniques (Carlone et al., 2015b; Martinec and Pajdla, 2007; Rosen et al., 2015), and the elucidation of deeper insights into the global structure of the SLAM optimization problem (Huang et al., 2010, 2012; Wang et al., 2012). Although all of the aforementioned efforts have led to substantial improvements in the robustness of current practical SLAM techniques, ultimately they are still unable to provide any guarantees on the correctness of the solutions that they recover.

As an alternative to local search, a second class of algorithms employs *convex relaxation*: in this approach, one modifies the original estimation problem so as to obtain a *convex approximation* that can be (*globally*) solved efficiently. Recent work has proposed a wide variety of convex relaxations for special Euclidean synchronization and related estimation problems, including linear (Carlone et al., 2015b; Martinec and Pajdla, 2007), spectral (Arrigoni et al., 2016; Bandeira et al., 2013; Cucuringu et al., 2012; Singer, 2011) and semidefinite (Özyeşil et al., 2015; Rosen et al., 2015; Wang and Singer, 2013; Singer, 2011) formulations, among others. The advantage of these techniques is that the convex surrogates they employ generally capture the global structure of the original problem well enough that their solutions lie near high-quality regions of the search space for the original estimation problem. However, as these surrogates are typically obtained from the original problem by *relaxing constraints*, their minimizers are generally infeasible for the original estimation problem, and therefore must be (potentially suboptimally) reprojected onto the original problem's feasible set.[2]

Although SE-Sync is also (strictly speaking) a fast approximation technique based upon convex relaxation, it differs from prior art in that (as we prove in Section 4.2) it admits a (fairly large) regime of operation in which it is able to recover the *exact* optimal solution, together with a *certificate* of that solution's correctness.

3. Problem formulation

3.1. Notation and mathematical preliminaries

3.1.1. Miscellaneous sets. The symbols $\mathbb{N}$ and $\mathbb{R}$ denote the non-negative integers and the real numbers, respectively, and we write $[n] \triangleq \{1, \ldots, n\}$ for $n > 0$ as a convenient shorthand notation for sets of indexing integers. Here $|S|$ denotes the cardinality of a set S.

3.1.2. Differential geometry and Lie groups. We will encounter several smooth manifolds and Lie groups over the course of this paper, and will often make use of both the *intrinsic* and *extrinsic* formulations of the same manifold as convenience dictates; our notation will generally be consistent with that of Warner (1983) in the former case and (Guillemin and Pollack, 1974) in the latter. When considering an extrinsic realization $\mathcal{M} \subseteq \mathbb{R}^d$ of a manifold $\mathcal{M}$ as an embedded submanifold of a Euclidean space and a function $f: \mathbb{R}^d \to \mathbb{R}$, it will occasionally be important for us to distinguish the notions of f considered as a function on $\mathbb{R}^d$ and f considered as a function on the submanifold $\mathcal{M}$; in these cases, to avoid confusion we will always reserve ∇f and $\nabla^2 f$ for the gradient and Hessian of f with respect to the usual metric on $\mathbb{R}^d$, and write $\operatorname{grad} f$ and $\operatorname{Hess} f$ to refer to the Riemannian gradient and Hessian of f considered as a function on $\mathcal{M}$ (equipped with the metric inherited from its embedding) (Boothby, 2003; Kobayashi and Nomizu, 1996).

We let $\mathrm{O}(d)$, $\mathrm{SO}(d)$, and $\mathrm{SE}(d)$ denote the orthogonal, special orthogonal, and special Euclidean groups in dimension d, respectively. For computational convenience we will often identify the (abstract) Lie groups $\mathrm{O}(d)$ and $\mathrm{SO}(d)$ with their realizations as the matrix groups:

$$\mathrm{O}(d) \cong \{R \in \mathbb{R}^{d\times d} \mid R^\mathsf{T} R = R R^\mathsf{T} = I_d\} \tag{1a}$$

$$\mathrm{SO}(d) \cong \{R \in \mathbb{R}^{d\times d} \mid R^\mathsf{T} R = R R^\mathsf{T} = I_d,\ \det(R) = +1\} \tag{1b}$$

and $\mathrm{SE}(d)$ with the semidirect product $\mathbb{R}^d \rtimes \mathrm{SO}(d)$ with group operations:

$$(t_1, R_1) \cdot (t_2, R_2) = (t_1 + R_1 t_2, R_1 R_2) \tag{2a}$$

$$(t, R)^{-1} = (-R^{-1} t, R^{-1}) \tag{2b}$$

The set of orthonormal k-frames in $\mathbb{R}^n$ ($k \leq n$):

$$\mathrm{St}(k, n) \triangleq \left\{Y \in \mathbb{R}^{n\times k} \mid Y^\mathsf{T} Y = I_k\right\} \tag{3}$$

is also a smooth compact matrix manifold, called the *Stiefel manifold*, and we equip $\mathrm{St}(k,n)$ with the Riemannian metric induced by its embedding into $\mathbb{R}^{n\times k}$ (Absil et al., 2008, Section 3.3.2).

3.1.3. Linear algebra. In addition to the matrix groups defined above, we write $\mathrm{Sym}(d)$ for the set of real $d \times d$ symmetric matrices; $A \succeq 0$ and $A \succ 0$ indicate that $A \in \mathrm{Sym}(d)$ is positive semidefinite and positive definite, respectively. For general matrices A and B, $A \otimes B$ indicates the Kronecker (matrix tensor) product, $A^\dagger$ the Moore–Penrose pseudoinverse, and $\mathrm{vec}(A)$ the vectorization operator that concatenates the columns of A (Horn and Johnson, 1991, Section 4.2). We write $e_i \in \mathbb{R}^d$ and $E_{ij} \in \mathbb{R}^{m\times n}$ for the ith unit coordinate vector and (i,j)th unit coordinate matrix, respectively, and $\mathbf{1}_d \in \mathbb{R}^d$ for the all-ones vector. Finally, $\|\cdot\|_2$, $\|\cdot\|_F$, and $\|\cdot\|_*$ denote the spectral, Frobenius, and nuclear matrix norms, respectively.

We also frequently consider various $(d \times d)$-block-structured matrices, and it will be useful to have specialized operators for them. To that end, given matrices $A_i \in \mathbb{R}^{d\times d}$ for $i \in [n]$, we let $\mathrm{Diag}(A_1, \ldots, A_n)$ denote their matrix direct sum. Conversely, given a $(d \times d)$-block-structured matrix $M \in \mathbb{R}^{dn\times dn}$ with ij-block $M_{ij} \in \mathbb{R}^{d\times d}$, we let $\mathrm{BlockDiag}_d(M)$ denote the linear operator that extracts the $(d \times d)$-block diagonal of M:

$$\mathrm{BlockDiag}_d(M) \triangleq \begin{pmatrix} M_{11} & & \\ & \ddots & \\ & & M_{nn} \end{pmatrix} \tag{4}$$

and $\mathrm{SymBlockDiag}_d$ its corresponding symmetrized form:

$$\mathrm{SymBlockDiag}_d(M) \triangleq \frac{1}{2} \mathrm{BlockDiag}_d\left(M + M^\mathsf{T}\right) \tag{5}$$

Finally, we let $\mathrm{SBD}(d, n)$ denote the set of symmetric $(d \times d)$-block-diagonal matrices in $\mathbb{R}^{dn\times dn}$:

$$\mathrm{SBD}(d, n) \triangleq \{\mathrm{Diag}(S_1, \ldots, S_n) \mid S_1, \ldots, S_n \in \mathrm{Sym}(d)\} \tag{6}$$

3.1.4. Graph theory. An *undirected graph* is a pair $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a finite set and $\mathcal{E}$ is a set of unordered pairs $\{i, j\}$ with $i, j \in \mathcal{V}$ and $i \neq j$. Elements of $\mathcal{V}$ are called *vertices* or *nodes*, and elements of $\mathcal{E}$ are called *edges*. An edge $e = \{i, j\} \in \mathcal{E}$ is said to be *incident* to the vertices i and j; i and j are called the *endpoints* of e. We write $\delta(v)$ for the set of edges incident to a vertex v.

A *directed graph* is a pair $\vec{G} = (\mathcal{V}, \vec{\mathcal{E}})$, where $\mathcal{V}$ is a finite set and $\vec{\mathcal{E}} \subset \mathcal{V} \times \mathcal{V}$ is a set of *ordered* pairs (i, j) with $i \neq j$.[3] As before, elements of $\mathcal{V}$ are called *vertices* or *nodes*, and elements of $\vec{\mathcal{E}}$ are called (*directed*) *edges* or *arcs*. Vertices i and j are called the *tail* and *head* of the directed edge $e = (i, j)$, respectively; e is said to *leave* i and *enter* j (we also say that e is *incident* to i and j, and that i and j are the *endpoints* of e, as in the case of undirected graphs). We let $t, h: \vec{\mathcal{E}} \to \mathcal{V}$ denote the functions mapping each edge to its tail and head, respectively, so that $t(e) = i$ and $h(e) = j$ for all $e = (i, j) \in \vec{\mathcal{E}}$. Finally, we again let $\delta(v)$ denote the set of directed edges incident to v, and $\delta^-(v)$ and $\delta^+(v)$ denote the sets of edges leaving and entering vertex v, respectively.

Given an undirected graph $G = (\mathcal{V}, \mathcal{E})$, we can construct a directed graph $\vec{G} = (\mathcal{V}, \vec{\mathcal{E}})$ from it by arbitrarily ordering the elements of each pair $\{i,j\} \in \mathcal{E}$; the graph $\vec{G}$ so obtained is called an *orientation* of G.

A *weighted graph* is a triplet $G = (\mathcal{V}, \mathcal{E}, w)$ comprising a graph $(\mathcal{V}, \mathcal{E})$ and a weight function $w\colon \mathcal{E} \to \mathbb{R}$ defined on its edges. As $\mathcal{E}$ is finite, we can specify the weight function w by simply listing its values $\{w_e\}_{e\in\mathcal{E}}$ on each edge. Any unweighted graph can be interpreted as a weighted graph equipped with the *uniform weight function* $w \equiv 1$.

We can associate to a directed graph $\vec{G} = (\mathcal{V}, \vec{\mathcal{E}})$ with $n = |\mathcal{V}|$ and $m = |\vec{\mathcal{E}}|$ the *incidence matrix* $A(\vec{G}) \in \mathbb{R}^{n\times m}$ whose rows and columns are indexed by $i \in \mathcal{V}$ and $e \in \vec{\mathcal{E}}$, respectively, and whose elements are determined by

$$A(\vec{G})_{ie} = \begin{cases} +1, & e \in \delta^+(i) \\ -1, & e \in \delta^-(i) \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

Similarly, we can associate to an undirected graph G an *oriented incidence matrix* $A(G)$ obtained as the incidence matrix of any of its orientations $\vec{G}$. We obtain a *reduced* (*oriented*) *incidence matrix* $\underline{A}(G)$ by removing the final row from the (oriented) incidence matrix $A(G)$.

Finally, we can associate to a weighted undirected graph $G = (\mathcal{V}, \mathcal{E}, w)$ with $n = |\mathcal{V}|$ the *Laplacian matrix* $L(G) \in \mathrm{Sym}(n)$ whose rows and columns are indexed by $i \in \mathcal{V}$, and whose elements are determined by

$$L(G)_{ij} = \begin{cases} \sum_{e\in\delta(i)} w(e), & i = j \\ -w(e), & i \neq j \text{ and } e = \{i,j\} \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

A straightforward computation shows that the Laplacian of a weighted graph G and the incidence matrix of one of its orientations $\vec{G}$ are related by

$$L(G) = A(\vec{G})\, W A(\vec{G})^{\mathsf{T}} \tag{9}$$

where $W \triangleq \mathrm{Diag}(w(e_1), \dots, w(e_m))$ is the diagonal matrix containing the weights of the edges of G.

3.1.5. Probability and statistics. We write $\mathcal{N}(\mu, \Sigma)$ for the multivariate Gaussian distribution with mean $\mu \in \mathbb{R}^d$ and covariance matrix $0 \preceq \Sigma \in \mathrm{Sym}(d)$, and $\mathrm{Langevin}(M, \kappa)$ for the isotropic Langevin distribution on $\mathrm{SO}(d)$ with mode $M \in \mathrm{SO}(d)$ and concentration parameter $\kappa \geq 0$ (cf. Appendix A). With reference to a hidden parameter X whose value we wish to infer, we will write $\underline{X}$ for its true (latent) value, $\tilde{X}$ to denote a noisy observation of $\underline{X}$, and $\hat{X}$ to denote an estimate of $\underline{X}$.

3.2. The special Euclidean synchronization problem

The $\mathrm{SE}(d)$ synchronization problem consists of estimating the values of a set of n unknown group elements $x_1, \dots, x_n \in \mathrm{SE}(d)$ given noisy measurements of m of their pairwise relative transforms $x_{ij} \triangleq x_i^{-1} x_j$ $(i \neq j)$. We model the set of available measurements using an undirected graph $G = (\mathcal{V}, \mathcal{E})$ in which the nodes $i \in \mathcal{V}$ are in one-to-one correspondence with the unknown states x_i and the edges $\{i,j\} \in \mathcal{E}$ are in one-to-one correspondence with the set of available measurements, and we assume without loss of generality that G is connected.[4] We let $\vec{G} = (\mathcal{V}, \vec{\mathcal{E}})$ be the directed graph obtained from G by fixing an orientation, and assume that a noisy measurement $\tilde{x}_{ij}$ of the relative transform x_{ij} is obtained by sampling from the following probabilistic generative model:

$$\begin{aligned} \tilde{t}_{ij} &= \underline{t}_{ij} + t_{ij}^{\epsilon}, \qquad & t_{ij}^{\epsilon} &\sim \mathcal{N}\left(0, \tau_{ij}^{-1} I_d\right) \\ \tilde{R}_{ij} &= \underline{R}_{ij} R_{ij}^{\epsilon}, \qquad & R_{ij}^{\epsilon} &\sim \mathrm{Langevin}\left(I_d, \kappa_{ij}\right) \end{aligned} \tag{10}$$

for all $(i,j) \in \vec{\mathcal{E}}$, where $\underline{x}_{ij} = (\underline{t}_{ij}, \underline{R}_{ij})$ is the true (latent) value of x_{ij}.[5] Finally, we define $\tilde{x}_{ji} \triangleq \tilde{x}_{ij}^{-1}$, $\kappa_{ji} \triangleq \kappa_{ij}$, $\tau_{ji} \triangleq \tau_{ij}$, and $\tilde{R}_{ji} \triangleq \tilde{R}_{ij}^{\mathsf{T}}$ for all $(i,j) \in \vec{\mathcal{E}}$.

Given a set of noisy measurements $\tilde{x}_{ij}$ sampled from the generative model (10), applying (2) to express the measurement noise t_{ij}^{ϵ} and R_{ij}^{ϵ} in terms of x_i, x_j, and $\tilde{x}_{ij}$ and substituting into the probability density functions for the Gaussian and Langevin distributions (cf. (52)) appearing in (10) produces the following likelihood function for the states x_i given the data $\tilde{x}_{ij}$:

$$\begin{aligned} &p\left(\{\tilde{x}_{ij}\}_{(i,j)\in\vec{\mathcal{E}}} \mid \{x_i\}_{i=1}^n\right) \\ &= \prod_{(i,j)\in\vec{\mathcal{E}}} \left\{ \begin{array}{l} (2\pi)^{-\frac{d}{2}} \tau_{ij}^{\frac{d}{2}} \exp\left(-\frac{\tau_{ij}}{2} \left\| t_j - t_i - R_i \tilde{t}_{ij} \right\|_2^2\right) \\ \times \dfrac{1}{c_d(\kappa_{ij})} \exp\left(\kappa_{ij} \operatorname{tr}\left(R_j^{\mathsf{T}} R_i \tilde{R}_{ij}\right)\right) \end{array} \right\} \end{aligned} \tag{11}$$

Taking negative logarithms of (11) and observing that $\operatorname{tr}(R_j^{\mathsf{T}} R_i \tilde{R}_{ij}) = d - \frac{1}{2}\|\tilde{R}_{ij} - R_i^{\mathsf{T}} R_j\|_F^2$ (as $\tilde{R}_{ij}, R_i, R_j \in \mathrm{SO}(d)$), it follows that a maximum-likelihood estimate $\hat{x}_{\mathrm{MLE}} \in \mathrm{SE}(d)^n$ for $x_1, \dots, x_n$ is obtained as a minimizer of the following problem:[6]

Problem 1 (Maximum-likelihood estimation for $\mathrm{SE}(d)$ synchronization)**.**

$$p_{\mathrm{MLE}}^* = \min_{\substack{t_i \in \mathbb{R}^d \\ R_i \in \mathrm{SO}(d)}} \sum_{(i,j)\in\vec{\mathcal{E}}} \left\{ \begin{array}{l} \kappa_{ij} \|R_j - R_i \tilde{R}_{ij}\|_F^2 \\ + \tau_{ij} \left\| t_j - t_i - R_i \tilde{t}_{ij} \right\|_2^2 \end{array} \right\} \tag{12}$$

Unfortunately, Problem 1 is a high-dimensional nonconvex nonlinear program, and is therefore computationally hard to solve in general. Consequently, in this paper our strategy will be to replace this problem with a (convex) *semidefinite relaxation* (Vandenberghe and Boyd, 1996), and then exploit this relaxation to search for solutions of Problem 1.

4. Forming the semidefinite relaxation

In this section, we develop the semidefinite relaxation that we will solve in place of the maximum-likelihood estimation Problem 1. Our approach proceeds in two stages. We begin in Section 4.1 by developing a sequence of simplified but equivalent reformulations of Problem 1 with the twofold goal of simplifying its analysis and elucidating some of the structural correspondences between the optimization (12) and several simple graph-theoretic objects that can be constructed from the set of available measurements $\tilde{x}_{ij}$ and the graphs G and $\vec{G}$. We then exploit the simplified versions of Problem 1 so obtained to derive the semidefinite relaxation in Section 4.2.

4.1. Simplifying the maximum-likelihood estimation

Our first step is to rewrite Problem 1 in a more standard form for quadratic programs. First, define the *translational* and *rotational weight graphs* $W^\tau \triangleq (\mathcal{V}, \mathcal{E}, \{\tau_{ij}\})$ and $W^\rho \triangleq (\mathcal{V}, \mathcal{E}, \{\kappa_{ij}\})$ to be the weighted undirected graphs with node set $\mathcal{V}$, edge set $\mathcal{E}$, and edge weights τ_{ij} and κ_{ij} for $\{i,j\} \in \mathcal{E}$, respectively. The Laplacians of W^τ and W^ρ are then:

$$L(W^\tau)_{ij} = \begin{cases} \sum_{e\in\delta(i)} \tau_e, & i=j \\ -\tau_{ij}, & \{i,j\} \in \mathcal{E} \\ 0, & \{i,j\} \notin \mathcal{E} \end{cases} \tag{13a}$$

$$L(W^\rho)_{ij} = \begin{cases} \sum_{e\in\delta(i)} \kappa_e, & i=j \\ -\kappa_{ij}, & \{i,j\} \in \mathcal{E} \\ 0, & \{i,j\} \notin \mathcal{E} \end{cases} \tag{13b}$$

Similarly, let $L(\tilde{G}^\rho)$ denote the *connection Laplacian* for the rotational synchronization problem determined by the measurements $\tilde{R}_{ij}$ and measurement weights κ_{ij} for $(i,j) \in \vec{\mathcal{E}}$; this is the symmetric $(d \times d)$-block-structured matrix determined by (cf. Singer and Wu, 2012; Wang and Singer, 2013)):

$$L(\tilde{G}^\rho) \in \mathrm{Sym}(dn)$$
$$L(\tilde{G}^\rho)_{ij} \triangleq \begin{cases} d_i^\rho I_d, & i=j \\ -\kappa_{ij}\tilde{R}_{ij}, & \{i,j\} \in \mathcal{E} \\ 0_{d\times d}, & \{i,j\} \notin \mathcal{E} \end{cases} \tag{14a}$$

$$d_i^\rho \triangleq \sum_{e\in\delta(i)} \kappa_e \tag{14b}$$

Finally, let $\tilde{V} \in \mathbb{R}^{n\times dn}$ be the $(1 \times d)$-block-structured matrix with (i,j)-blocks:

$$\tilde{V}_{ij} \triangleq \begin{cases} \sum_{e\in\delta^-(j)} \tau_e \tilde{t}_e^\mathsf{T}, & i=j \\ -\tau_{ji}\tilde{t}_{ji}^\mathsf{T}, & (j,i) \in \vec{\mathcal{E}} \\ 0_{1\times d}, & \text{otherwise} \end{cases} \tag{15}$$

and $\tilde{\Sigma}$ the $(d \times d)$-block-structured block-diagonal matrix determined by

$$\tilde{\Sigma} \triangleq \mathrm{Diag}(\tilde{\Sigma}_1, \ldots, \tilde{\Sigma}_n) \in \mathrm{SBD}(d,n)$$
$$\tilde{\Sigma}_i \triangleq \sum_{e\in\delta^-(i)} \tau_e \tilde{t}_e \tilde{t}_e^\mathsf{T} \tag{16}$$

Aggregating the rotational and translational states into the block matrices:

$$R \triangleq \begin{pmatrix} R_1 & \cdots & R_n \end{pmatrix} \in \mathrm{SO}(d)^n \subset \mathbb{R}^{d\times dn} \tag{17a}$$

$$t \triangleq \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \in \mathbb{R}^{dn} \tag{17b}$$

and exploiting definitions (13a)–(16), Problem 1 can be rewritten more compactly in the following standard form:

Problem 2 (Maximum-likelihood estimation, QP form)**.**

$$p^*_{\mathrm{MLE}} = \min_{\substack{t\in\mathbb{R}^{dn} \\ R\in\mathrm{SO}(d)^n}} \begin{pmatrix} t \\ \mathrm{vec}(R) \end{pmatrix}^\mathsf{T} (M \otimes I_d) \begin{pmatrix} t \\ \mathrm{vec}(R) \end{pmatrix} \tag{18a}$$

$$M \triangleq \begin{pmatrix} L(W^\tau) & \tilde{V} \\ \tilde{V}^\mathsf{T} & L(\tilde{G}^\rho) + \tilde{\Sigma} \end{pmatrix} \tag{18b}$$

Problem 2 is obtained from Problem 1 through a straightforward (although somewhat tedious) manipulation of the objective function (Appendix B.1).

Expanding the quadratic form in (18), we obtain

$$p^*_{\mathrm{MLE}} = \min_{\substack{t\in\mathbb{R}^{dn} \\ R\in\mathrm{SO}(d)^n}} \left\{ \begin{array}{l} t^\mathsf{T}(L(W^\tau)\otimes I_d)\,t + 2t^\mathsf{T}\left(\tilde{V}\otimes I_d\right)\mathrm{vec}(R) \\ + \mathrm{vec}(R)^\mathsf{T}\left(\left(L(\tilde{G}^\rho)+\tilde{\Sigma}\right)\otimes I_d\right)\mathrm{vec}(R) \end{array} \right\} \tag{19}$$

Now observe that for a fixed value of R, Equation (19) reduces to the *unconstrained* minimization of a quadratic form in the translational variable t, for which we can find a closed-form solution. This enables us to analytically eliminate t from the optimization problem (19), thereby obtaining the following:

Problem 3 (Rotation-only maximum-likelihood estimation)**.**

$$p^*_{\mathrm{MLE}} = \min_{R\in\mathrm{SO}(d)^n} \mathrm{tr}(\tilde{Q}R^\mathsf{T}R) \tag{20a}$$

$$\tilde{Q} \triangleq L(\tilde{G}^\rho) + \underbrace{\tilde{\Sigma} - \tilde{V}^\mathsf{T} L(W^\tau)^\dagger \tilde{V}}_{\tilde{Q}^\tau} \tag{20b}$$

Furthermore, given any minimizer R^* of (20), we can recover a corresponding optimal value t^* for t via:

$$t^* = -\,\mathrm{vec}\left(R^*\tilde{V}^\mathsf{T} L(W^\tau)^\dagger\right) \tag{21}$$

The derivation of (20) and (21) from (19) is given in Appendix B.2.

Finally, we derive a simplified expression for the translational data matrix $\tilde{Q}^{\tau}$ appearing in (20b). Let

$$\Omega \triangleq \mathrm{Diag}(\tau_{e_1}, \ldots, \tau_{e_m}) \in \mathrm{Sym}(m) \tag{22}$$

denote the diagonal matrix whose rows and columns are indexed by the directed edges $e \in \vec{\mathcal{E}}$ and whose eth diagonal element gives the precision of the translational observation corresponding to that edge. Similarly, let $\tilde{T} \in \mathbb{R}^{m \times dn}$ denote the $(1 \times d)$-block-structured matrix with rows and columns indexed by $e \in \vec{\mathcal{E}}$ and $k \in \mathcal{V}$, respectively, and whose (e, k)-block is given by

$$\tilde{T}_{ek} \triangleq \begin{cases} -\tilde{t}_{kj}^{\mathsf{T}}, & e = (k,j) \in \vec{\mathcal{E}} \\ 0_{1\times d}, & \text{otherwise} \end{cases} \tag{23}$$

Then Problem 3 can be rewritten as:

Problem 4 (Simplified maximum-likelihood estimation)**.**

$$p^*_{\mathrm{MLE}} = \min_{R \in \mathrm{SO}(d)^n} \mathrm{tr}(\tilde{Q} R^{\mathsf{T}} R) \tag{24a}$$

$$\tilde{Q} = L(\tilde{G}^{\rho}) + \tilde{Q}^{\tau} \tag{24b}$$

$$\tilde{Q}^{\tau} = \tilde{T}^{\mathsf{T}} \Omega^{\frac{1}{2}} \Pi \Omega^{\frac{1}{2}} \tilde{T} \tag{24c}$$

Here $\Pi \in \mathbb{R}^{m \times m}$ is the matrix of the orthogonal projection operator $\pi : \mathbb{R}^m \to \ker(A(\vec{G})\Omega^{\frac{1}{2}})$ onto the kernel of the weighted incidence matrix $A(\vec{G})\Omega^{\frac{1}{2}}$ of $\vec{G}$. The derivation of (24c) from (20b) is presented in Appendix B.3.

The advantage of expression (24c) for $\tilde{Q}^{\tau}$ versus the original formulation given in (20b) is that the constituent matrices Π, Ω, and $\tilde{T}$ in (24c) each admit particularly simple interpretations in terms of the underlying directed graph $\vec{G}$ and the translational data $(\tau_{ij}, \tilde{t}_{ij})$ attached to each edge $(i,j) \in \vec{\mathcal{E}}$; our subsequent development will heavily exploit this structure.

4.2. Relaxing the maximum-likelihood estimation

In this subsection, we turn our attention to the development of a convex relaxation that will enable us to recover a global minimizer of Problem 4 in practice. We begin by relaxing the condition that $R \in \mathrm{SO}(d)^n$, obtaining:

Problem 5 (Orthogonal relaxation of the maximum-likelihood estimation)**.**

$$p^*_{\mathrm{O}} = \min_{R \in \mathrm{O}(d)^n} \mathrm{tr}(\tilde{Q} R^{\mathsf{T}} R) \tag{25}$$

We immediately have that $p^*_{\mathrm{O}} \le p^*_{\mathrm{MLE}}$ since $\mathrm{SO}(d)^n \subset \mathrm{O}(d)^n$. However, we expect that this relaxation will often be *exact* in practice: as $\mathrm{O}(d)$ is a disjoint union of two components separated by a distance of $\sqrt{2}$ under the Frobenius norm, and the values $\underline{R}_i$ that we wish to estimate all lie in $\mathrm{SO}(d)$, the elements R^*_i of an estimate R^* obtained as a minimizer of (25) will still all lie in the $+1$ component of $\mathrm{O}(d)$ so long as the elementwise estimation error in R^* satisfies $\|R^*_i - \underline{R}_i\|_F < \sqrt{2}$ for all $i \in [n]$. This latter condition will hold so long as the noise perturbing the data matrix $\tilde{Q}$ is not too large (cf. Appendix C.4).[7]

Now we derive the Lagrangian dual of Problem 5, using its *extrinsic* formulation:

$$\begin{aligned} p^*_{\mathrm{O}} = \min_{R \in \mathbb{R}^{d \times dn}} \ & \mathrm{tr}(\tilde{Q} R^{\mathsf{T}} R) \\ \text{s.t.} \ & R_i^{\mathsf{T}} R_i = I_d \quad \forall i = 1, \ldots, n \end{aligned} \tag{26}$$

The Lagrangian corresponding to (26) is

$$\begin{aligned} \mathcal{L} &: \mathbb{R}^{d \times dn} \times \mathrm{Sym}(d)^n \to \mathbb{R} \\ \mathcal{L}(R, \Lambda_1, \ldots, \Lambda_n) &= \mathrm{tr}(\tilde{Q} R^{\mathsf{T}} R) - \sum_{i=1}^{n} \mathrm{tr}\left(\Lambda_i (R_i^{\mathsf{T}} R_i - I_d)\right) \\ &= \mathrm{tr}(\tilde{Q} R^{\mathsf{T}} R) + \sum_{i=1}^{n} \mathrm{tr}(\Lambda_i) - \mathrm{tr}\left(\Lambda_i R_i^{\mathsf{T}} R_i\right) \end{aligned} \tag{27}$$

where $\Lambda_i \in \mathrm{Sym}(d)$ are symmetric matrices of Lagrange multipliers for the (symmetric) matrix orthonormality constraints in (26). We can simplify (27) by aggregating the Lagrange multipliers Λ_i into a single direct sum matrix $\Lambda \triangleq \mathrm{Diag}(\Lambda_1, \ldots, \Lambda_n) \in \mathrm{SBD}(d, n)$ to yield

$$\begin{aligned} \mathcal{L} &: \mathbb{R}^{d \times dn} \times \mathrm{SBD}(d,n) \to \mathbb{R} \\ \mathcal{L}(R, \Lambda) &= \mathrm{tr}\left((\tilde{Q} - \Lambda) R^{\mathsf{T}} R\right) + \mathrm{tr}(\Lambda) \end{aligned} \tag{28}$$

The Lagrangian dual problem for (26) is thus:

$$\max_{\Lambda \in \mathrm{SBD}(d,n)} \left\{ \inf_{R \in \mathbb{R}^{d \times dn}} \mathrm{tr}\left((\tilde{Q} - \Lambda) R^{\mathsf{T}} R\right) + \mathrm{tr}(\Lambda) \right\} \tag{29}$$

with corresponding dual function:

$$d(\Lambda) \triangleq \inf_{R \in \mathbb{R}^{d \times dn}} \mathrm{tr}\left((\tilde{Q} - \Lambda) R^{\mathsf{T}} R\right) + \mathrm{tr}(\Lambda) \tag{30}$$

However, we observe that as

$$\mathrm{tr}\left((\tilde{Q} - \Lambda) R^{\mathsf{T}} R\right) = \mathrm{vec}(R)^{\mathsf{T}} \left((\tilde{Q} - \Lambda) \otimes I_d\right) \mathrm{vec}(R) \tag{31}$$

then $d(\Lambda) = -\infty$ in (30) unless $(\tilde{Q} - \Lambda) \otimes I_d \succeq 0$, in which case the infimum is attained for $R = 0$. Furthermore, we have $(\tilde{Q} - \Lambda) \otimes I_d \succeq 0$ if and only if $\tilde{Q} - \Lambda \succeq 0$. Therefore, the dual problem (29) is equivalent to the following semidefinite program.

Problem 6 (Primal semidefinite relaxation for SE(d) synchronization)**.**

$$\begin{aligned} p^*_{\mathrm{SDP}} = \max_{\Lambda \in \mathrm{SBD}(d,n)} \ & \mathrm{tr}(\Lambda) \\ \text{s.t.} \ & \tilde{Q} - \Lambda \succeq 0 \end{aligned} \tag{32}$$

Finally, a straightforward application of the duality theory for semidefinite programs (see Appendix B.4 for details) shows that the dual of Problem 6 is:

Problem 7 (Dual semidefinite relaxation for SE(d) synchronization)**.**

$$p^*_{\text{SDP}} = \min_{Z \in \text{Sym}(dn)} \text{tr}(\tilde{Q}Z)$$
$$\text{s.t. } Z = \begin{pmatrix} I_d & * & * & \cdots & * \\ * & I_d & * & \cdots & * \\ * & * & I_d & & * \\ \vdots & \vdots & & \ddots & \vdots \\ * & * & * & \cdots & I_d \end{pmatrix} \succeq 0 \tag{33}$$

At this point, it is instructive to compare the dual semidefinite relaxation (33) with the simplified maximum-likelihood estimation (24). For any $R \in \text{SO}(d)^n$, the product $Z = R^{\mathsf{T}}R$ is positive semidefinite and has identity matrices along its $(d \times d)$-block-diagonal, and so is a feasible point of (33); in other words, Equation (33) can be regarded as a relaxation of the maximum-likelihood estimation obtained by *expanding the feasible set of* (24). Consequently, if it so happens that a minimizer Z^* of Problem 7 admits a decomposition of the form $Z^* = R^{*\mathsf{T}}R^*$ for some $R^* \in \text{SO}(d)^n$, then it is straightforward to verify that this R^* is also a minimizer of Problem 4. More precisely, we have the following:

Theorem 1. *Let Z^* be a minimizer of the semidefinite relaxation Problem 7. If Z^* factors as*

$$Z^* = R^{*\mathsf{T}}R^*, \quad R^* \in \text{O}(d)^n \tag{34}$$

then R^* is a minimizer of Problem 5. If, in addition, $R^* \in \text{SO}(d)^n$, then R^* is also a minimizer of Problem 4, and $x^* = (t^*, R^*)$ (with t^* given by Equation (21)) is an optimal solution of the maximum-likelihood estimation Problem 1.

Proof. Weak Lagrangian duality implies that the optimal values of Problems 5 and 6 satisfy $p^*_{\text{SDP}} \leq p^*_{\text{O}}$. But if Z^* admits the factorization (34), then R^* is also a feasible point of (25), and so we must have that $p^*_{\text{O}} \leq \text{tr}(\tilde{Q}R^{*\mathsf{T}}R^*) = p^*_{\text{SDP}}$. This shows that $p^*_{\text{O}} = p^*_{\text{SDP}}$, and consequently that R^* is a minimizer of Problem 5, since it attains the optimal value.

Similarly, we have already established that the optimal values of Problems 4 and 5 satisfy $p^*_{\text{O}} \leq p^*_{\text{MLE}}$. But if additionally $R^* \in \text{SO}(d)^n$, then R^* is feasible for Problem 4, and so by the same logic as before we have that $p^*_{\text{O}} = p^*_{\text{MLE}}$ and R^* is a minimizer of Problem 4. The final claim now follows from the optimality of R^* for Problem 4 together with equation (21). □

From a practical standpoint, Theorem 1 serves to identify a class of solutions of the (convex) semidefinite relaxation Problem 7 that correspond to *global minimizers* of the nonconvex maximum-likelihood estimation Problem 1. This naturally leads us to consider the following two questions: Under what conditions does Problem 7 admit a solution belonging to this class? And if such a solution exists, can we guarantee that we will be able to *recover* it by solving Problem 7 using a numerical optimization method?[8] These questions are addressed by the following.

Proposition 2 (Exact recovery via the semidefinite relaxation Problem 7)**.** *Let $\underline{Q}$ be the matrix of the form* (24b) *constructed using the true (latent) relative transforms $\underline{x}_{ij} = (\underline{t}_{ij}, \underline{R}_{ij})$ in* (10)*. There exists a constant $\beta \triangleq \beta(\underline{Q}) > 0$ (depending upon $\underline{Q}$) such that, if $\|\tilde{Q} - \underline{Q}\|_2 < \beta$, then:*

(i) the dual semidefinite relaxation Problem 7 has a unique solution Z^*; and
(ii) $Z^* = R^{*\mathsf{T}}R^*$, where $R^* \in \text{SO}(d)^n$ is a minimizer of the simplified maximum-likelihood estimation Problem 4.

This result is proved in Appendix C, using an approach adapted from Bandeira et al. (2017).

In short, Proposition 2 guarantees that as long as the noise corrupting the available measurements $\tilde{x}_{ij}$ in (10) is not too large (as measured by the spectral norm of the deviation of the data matrix $\tilde{Q}$ from its exact latent value $\underline{Q}$),[9] we can recover a global minimizer R^* of Problem 4 (and, hence, also a global minimizer $x^* = (t^*, R^*)$ of the maximum-likelihood estimation Problem 1 via (21)) by solving Problem 7 using *any* numerical optimization method.

5. The SE-Sync algorithm

In light of Proposition 2, our overall strategy in this paper will be to search for exact solutions of the (hard) maximum-likelihood estimation Problem 1 by solving the (convex) semidefinite relaxation Problem 7. To realize this strategy as a practical algorithm, we therefore require (i) a method that is able to solve Problem 7 effectively in large-scale real-world problems, and (ii) a rounding procedure that recovers an *optimal* solution of Problem 1 from a solution of Problem 7 when exactness obtains, and a *feasible approximate solution* otherwise. In this section, we develop a pair of algorithms that fulfill these requirements. Together, these procedures comprise *SE-Sync*, our proposed algorithm for synchronization over the special Euclidean group.

5.1. Solving the semidefinite relaxation

As a semidefinite program, Problem 7 can, in principle, be solved in polynomial time using interior-point methods (Todd, 2001; Vandenberghe and Boyd, 1996). In practice, however, the high computational cost of general-purpose semidefinite programming algorithms prevents these methods from scaling effectively to problems in which the dimension of the decision variable Z is greater than a few thousand (Todd, 2001). Unfortunately, typical instances of Problem 7 arising in (for example) robotics and computer vision applications are one to two orders of magnitude larger than this maximum effective problem size, and are therefore well beyond the reach of these general-purpose methods. To overcome this limitation, in this subsection we develop a specialized optimization procedure for solving large-scale instances of Problem 7 efficiently. We first exploit this problem's low-rank, geometric, and graph-theoretic structure to reduce it to an equivalent optimization problem defined on a low-dimensional Riemannian manifold (Boothby, 2003; Kobayashi and Nomizu, 1996), and then design a fast Riemannian optimization method to solve this reduction efficiently.

5.1.1. *Simplifying Problem 7.*

Exploiting low-rank structure: The dominant computational cost when applying general-purpose semidefinite programming methods to solve Problem 7 is the need to store and manipulate expressions involving the (large, dense) matrix variable Z. In particular, the $O(n^3)$ computational cost of multiplying and factoring such expressions quickly becomes intractable as the problem size n increases. On the other hand, in the case that exactness holds, we know that the actual *solution* Z^* of Problem 7 that we seek has a very concise description in the factored form $Z^* = R^{*\mathsf{T}} R^*$ for $R^* \in \mathrm{SO}(d)^n$. More generally, even in those cases where exactness fails, minimizers Z^* of Problem 7 typically have a rank r not much greater than d, and therefore admit a symmetric rank decomposition $Z^* = Y^{*\mathsf{T}} Y^*$ for $Y^* \in \mathbb{R}^{r\times dn}$ with $r \ll dn$.

In a pair of papers, Burer and Monteiro (2003, 2005) proposed an elegant general approach to exploit the fact that large-scale semidefinite programs often admit such low-rank solutions: simply replace every instance of the decision variable Z with a rank-r product of the form $Y^\mathsf{T} Y$ to produce a *rank-restricted* version of the original problem. This substitution has the two-fold effect of (i) dramatically reducing the size of the search space and (ii) rendering the positive semidefiniteness constraint *redundant*, since $Y^\mathsf{T} Y \succeq 0$ for *any* choice of Y. The resulting rank-restricted form of the problem is thus a low-dimensional *nonlinear* program, rather than a *semidefinite* program. In the specific case of Problem 7, this produces:

Problem 8 (Rank-restricted semidefinite relaxation, NLP form)**.**

$$
\begin{aligned}
p^*_{\mathrm{SDPLR}} = \min_{Y\in\mathbb{R}^{r\times dn}} \ & \operatorname{tr}(\tilde{Q} Y^\mathsf{T} Y) \\
\text{s.t.} \ \ & \operatorname{BlockDiag}_d(Y^\mathsf{T} Y) = \operatorname{Diag}(I_d, \ldots, I_d).
\end{aligned}
\tag{35}
$$

Provided that Problem 7 admits a solution Z^* with $\operatorname{rank}(Z^*) \le r$, we can *recover* such a solution from an optimal solution Y^* of Problem 8 according to $Z^* = Y^{*\mathsf{T}} Y^*$.

Exploiting geometric structure: In addition to exploiting Problem 7's low-rank structure, following Boumal (2015) we also observe that the specific form of the constraints appearing in Problems 7 and 8 (i.e. that the $d \times d$ block-diagonals of Z and $Y^\mathsf{T} Y$ must be I_d) admits a nice geometric interpretation that can be exploited to further simplify Problem 8. Introducing the block decomposition:

$$
Y \triangleq \begin{pmatrix} Y_1 & \cdots & Y_n \end{pmatrix} \in \mathbb{R}^{r\times dn} \tag{36}
$$

the block-diagonal constraints appearing in (35) are equivalent to

$$
Y_i^\mathsf{T} Y_i = I_d, \quad Y_i \in \mathbb{R}^{r\times d} \tag{37}
$$

i.e. they require that each Y_i be an element of the Stiefel manifold $\mathrm{St}(d, r)$ in (3). Consequently, Problem 8 can be equivalently formulated as an *unconstrained* Riemannian optimization problem on a product of on a product of Stiefel manifolds:

Problem 9 (Rank-restricted semidefinite relaxation, Riemannian optimization form)**.**

$$
p^*_{\mathrm{SDPLR}} = \min_{Y\in\mathrm{St}(d,r)^n} \operatorname{tr}(\tilde{Q} Y^\mathsf{T} Y) \tag{38}
$$

This is the optimization problem that we will actually solve in practice.

Exploiting graph-theoretic structure: The reduction of Problem 7 to Problem 9 obviates the need to form or manipulate the large, dense matrix variable Z directly. However, the data matrix $\tilde{Q}$ that parameterizes each of Problems 4–9 is also dense and of the same order as Z, and so presents a similar computational difficulty. Accordingly, here we develop an analogous concise description of $\tilde{Q}$ in terms of sparse matrices (and their inverses) associated with the graph $\vec{G}$.

Equations (24b) and (24c) provide a decomposition of $\tilde{Q}$ in terms of the sparse matrices $L(\tilde{G}^\rho)$, $\tilde{T}$, and Ω, and the dense orthogonal projection matrix Π. However, as Π is also a matrix derived from a sparse graph, we might suspect that it too should admit some kind of sparse description. Indeed, it turns out that Π admits a sparse decomposition as

$$
\underline{A}(\vec{G})\,\Omega^{\frac{1}{2}} = LQ_1 \tag{39a}
$$

$$
\Pi = I_m - \Omega^{\frac{1}{2}} \underline{A}(\vec{G})^\mathsf{T} L^{-\mathsf{T}} L^{-1} \underline{A}(\vec{G})\,\Omega^{\frac{1}{2}} \tag{39b}
$$

where (39a) is a thin LQ decomposition[10] of the weighted reduced incidence matrix $\underline{A}(\vec{G})\,\Omega^{\frac{1}{2}}$ of $\vec{G}$. This result is derived in Appendix B.3. Note that expression (39b) for Π requires only the sparse lower-triangular factor L from (39a), which can be easily and efficiently obtained (e.g. by applying successive Givens rotations (Golub and Loan, 1996, Section 5.2.1) directly to $\underline{A}(\vec{G})\,\Omega^{\frac{1}{2}}$ itself).

Together, Equations (24b), (24c), and (39b) provide a concise description of $\tilde{Q}$ in terms of sparse matrices, as desired. We exploit this decomposition in Section 5.1.3 to design a fast Riemannian optimization method for solving Problem 9.

5.1.2. The Riemannian staircase. At this point, it is again instructive to compare Problem 9 with the simplified maximum-likelihood estimation Problem 4 and its relaxation Problem 7. As the (special) orthogonal matrices satisfy condition (3) with $k = n = d$, we have the set of inclusions

$$\mathrm{SO}(d) \subset \mathrm{O}(d) = \mathrm{St}(d, d) \subset \mathrm{St}(d, d+1) \subset \cdots \tag{40}$$

and we can therefore view the set of rank-restricted Riemannian optimization problems (38) as comprising a *hierarchy* of relaxations of the maximum-likelihood estimation (24) that are intermediate between Problem 5 and Problem 7 for $d < r < dn$. However, unlike Problem 7, the various instantiations of Problem 9 are *non-convex* owing to the (re)introduction of the quadratic orthonormality constraints (3). It may therefore not be clear whether anything has really been gained by relaxing Problem 4 to Problem 9, since it appears that we may have simply replaced one difficult non-convex optimization problem with another. The following remarkable result (Boumal et al., 2016b, Corollary 8) justifies this approach:

Proposition 3 (A sufficient condition for global optimality in Problem 9)**.** *If $Y \in \mathrm{St}(d, r)^n$ is a (row) rank-deficient second-order critical point*[11] *of Problem 9, then Y is a global minimizer of Problem 9 and $Z^* = Y^\mathsf{T} Y$ is a solution of the dual semidefinite relaxation Problem 7.*

Proposition 3 immediately suggests an algorithm for recovering solutions Z^* of Problem 7 from Problem 9: simply apply a second-order Riemannian optimization method to search successively higher levels of the hierarchy of relaxations (38) until a *rank-deficient* second-order critical point is obtained.[12] This algorithm, the *Riemannian staircase* (Boumal, 2015; Boumal et al., 2016b), is summarized as Algorithm 1. We emphasize that while Algorithm 1 may require searching up to $O(dn)$ levels of the hierarchy (38) in the worst case, in practice this a gross overestimate; typically, one or two "stairs" suffice.

Algorithm 1 The Riemannian staircase

Input: An initial point $Y \in \mathrm{St}(d, r_0)^n$, $r_0 \geq d + 1$.
Output: A minimizer Y^* of Problem 9 corresponding to a solution $Z^* = Y^{*\mathsf{T}} Y^*$ of Problem 7.

function RiemannianStaircase(Y)
  for $r = r_0, \ldots, dn + 1$ **do**
    Starting at Y, apply a Riemannian optimization method[1] to identify a second-order critical point $Y^* \in \mathrm{St}(d, r)^n$ of Problem 9.
    if $\mathrm{rank}(Y^*) < r$ **then**
      return Y^*
    else
      Set $Y \leftarrow \begin{pmatrix} Y^* \\ 0_{1 \times dn} \end{pmatrix}$.
    end if
  end for
end function

5.1.3. A Riemannian optimization method for Problem 9. Proposition 3 and the Riemannian staircase (Algorithm 1) provide a means of obtaining *global* minimizers of Problem 7 by *locally* searching for second-order critical points of Problem 9. In this subsection, we design a Riemannian optimization method that will enable us to rapidly identify these critical points in practice.

Equations (24b), (24c), and (39b) provide an efficient means of computing products with $\tilde{Q}$ *without* the need to form $\tilde{Q}$ explicitly by performing a sequence of sparse matrix multiplications and sparse triangular solves. This operation is sufficient to evaluate the objective appearing in Problem 9, as well as its gradient and Hessian-vector products when it is considered as a function on the ambient Euclidean space $\mathbb{R}^{r \times dn}$:

$$F(Y) \triangleq \mathrm{tr}(\tilde{Q} Y^\mathsf{T} Y) \tag{41a}$$

$$\nabla F(Y) = 2Y\tilde{Q} \tag{41b}$$

$$\nabla^2 F(Y)[\dot{Y}] = 2\dot{Y}\tilde{Q} \tag{41c}$$

Furthermore, there are simple relations between the ambient Euclidean gradient and Hessian-vector products in (41b) and (41c) and their corresponding Riemannian counterparts when $F(\cdot)$ is viewed as a function restricted to the embedded submanifold $\mathrm{St}(d, r)^n \subset \mathbb{R}^{r \times dn}$. With reference to the orthogonal projection operator onto the tangent space of $\mathrm{St}(d, r)^n$ at Y (Edelman et al., 1998, equation (2.3)):

$$\begin{aligned} \mathrm{Proj}_Y &: T_Y\left(\mathbb{R}^{r \times dn}\right) \to T_Y\left(\mathrm{St}(d, r)^n\right) \\ \mathrm{Proj}_Y(X) &= X - Y\,\mathrm{SymBlockDiag}_d(Y^\mathsf{T} X) \end{aligned} \tag{42}$$

the Riemannian gradient $\mathrm{grad}\, F(Y)$ is simply the orthogonal projection of the ambient Euclidean gradient $\nabla F(Y)$ (cf. Absil et al., 2008, equation (3.37)):

$$\mathrm{grad}\, F(Y) = \mathrm{Proj}_Y \nabla F(Y) \tag{43}$$

Similarly, the Riemannian Hessian-vector product $\mathrm{Hess}\, F(Y)[\dot{Y}]$ can be obtained as the orthogonal projection of the ambient directional derivative of the gradient vector field $\mathrm{grad}\, F(Y)$ in the direction of $\dot{Y}$ (cf. Absil et al., 2008, equation (5.15)). A straightforward computation shows that this is given by[13]

$$\begin{aligned} \mathrm{Hess}\, F(Y)[\dot{Y}] &= \mathrm{Proj}_Y\left(\mathrm{D}\left[\mathrm{grad}\, F(Y)\right][\dot{Y}]\right) \\ &= \mathrm{Proj}_Y \begin{pmatrix} \nabla^2 F(Y)[\dot{Y}] \\ \quad - \dot{Y}\,\mathrm{SymBlockDiag}_d\left(Y^\mathsf{T} \nabla F(Y)\right) \end{pmatrix} \end{aligned} \tag{44}$$

Equations (24b), (24c), and (39)–(44) provide an efficient means of computing $F(Y)$, $\mathrm{grad}\, F(Y)$, and $\mathrm{Hess}\, F(Y)[\dot{Y}]$. Consequently, we propose to employ a *truncated-Newton trust-region* optimization method (Dembo and Steihaug,

1. For example, the second-order Riemannian trust-region method (Boumal et al., 2016a, Algorithm 3).

1983; Nash, 2000; Steihaug, 1983) to solve Problem 9; this approach will enable us to exploit the availability of an efficient routine for computing Hessian-vector products $\operatorname{Hess} F(Y)[\dot{Y}]$ to implement a second-order optimization method *without* the need to explicitly form or factor the dense matrix $\operatorname{Hess} F(Y)$. Moreover, truncated-Newton methods comprise the current state of the art for superlinear large-scale unconstrained nonlinear programming (Nocedal and Wright, 2006, Section 7.1), and are therefore ideally suited for solving large-scale instances of (38). Accordingly, we will apply the truncated-Newton *Riemannian trust-region* (RTR) method (Absil et al., 2007; Boumal et al., 2016a) to efficiently compute high-precision[14] estimates of second-order critical points of Problem 9.

Algorithm 2 Rounding procedure for solutions of Problem 9

Input: A minimizer $Y^* \in \mathrm{St}(d,r)^n$ of Problem 9.
Output: A feasible point $\hat{R} \in \mathrm{SO}(d)^n$.

```
function RoundSolution(Y*)
    Compute a rank-d truncated singular value
    [1] decomposition U_d Ξ_d V_d^T for Y* and assign
    [1] R̂ ← Ξ_d V_d^T.
    Set N+ ← |{R̂_i | det(R̂_i) > 0}|.
    if N+ < ⌈n/2⌉ then
        R̂ ← Diag(1, ..., 1, −1) R̂.
    end if
    for i = 1, ..., n do
        Set R̂_i ← NearestRotation(R̂_i).
    end for
    return R̂
end function
```

5.2. Rounding the solution

In the previous subsection, we described an efficient algorithmic approach for computing minimizers $Y^* \in \mathrm{St}(d,r)^n$ of Problem 9 that correspond to solutions $Z^* = Y^{*\mathsf{T}}Y^*$ of Problem 7. However, our ultimate goal is to extract an optimal solution $R^* \in \mathrm{SO}(d)^n$ of Problem 4 from Z^* whenever exactness holds, and a *feasible approximate solution* $\hat{R} \in \mathrm{SO}(d)^n$ otherwise. In this subsection, we develop a rounding procedure satisfying these criteria. To begin, let us consider the case in which exactness obtains; here

$$Y^{*\mathsf{T}}Y^* = Z^* = R^{*\mathsf{T}}R^* \tag{45}$$

for some optimal solution $R^* \in \mathrm{SO}(d)^n$ of Problem 4. As $\operatorname{rank}(R^*) = d$, this implies that $\operatorname{rank}(Y^*) = d$ as well. Consequently, letting

$$Y^* = U_d \Xi_d V_d^{\mathsf{T}} \tag{46}$$

denote a (rank-d) thin singular value decomposition (Golub and Loan, 1996, Section 2.5.3) of Y^*, and defining

$$\bar{Y} \triangleq \Xi_d V_d^{\mathsf{T}} \in \mathbb{R}^{d\times dn} \tag{47}$$

it follows from substituting (46) into (45) that

$$\bar{Y}^{\mathsf{T}}\bar{Y} = Z^* = R^{*\mathsf{T}}R^* \tag{48}$$

Equation (48) implies that the $d\times d$ block-diagonal of $\bar{Y}^{\mathsf{T}}\bar{Y}$ satisfies $\bar{Y}_i^{\mathsf{T}}\bar{Y}_i = I_d$ for all $i \in [n]$, i.e. $\bar{Y} \in \mathrm{O}(d)^n$. Similarly, comparing the elements of the first block rows of $\bar{Y}^{\mathsf{T}}\bar{Y}$ and $R^{*\mathsf{T}}R^*$ in (48) shows that $\bar{Y}_1^{\mathsf{T}}\bar{Y}_j = R_1^* R_j^*$ for all $j \in [n]$. Left-multiplying this set of equalities by $\bar{Y}_1$ and letting $A = \bar{Y}_1 R_1^*$ then gives

$$\bar{Y} = AR^*, \quad A \in \mathrm{O}(d) \tag{49}$$

As any product of the form AR^* with $A \in \mathrm{SO}(d)$ is *also* an optimal solution of Problem 4 (by gauge symmetry), Equation (49) shows that $\bar{Y}$ as defined in (47) is optimal provided that $\bar{Y} \in \mathrm{SO}(d)^n$ specifically. Furthermore, if this is not the case, we can always make it so by left-multiplying $\bar{Y}$ by any orientation-reversing element of $\mathrm{O}(d)$, for example $\operatorname{Diag}(1,\ldots,1,-1)$. Thus, Equations (46)–(49) give a straightforward means of recovering an optimal solution of Problem 4 from Y^* whenever exactness holds.

Moreover, this procedure can be straightforwardly generalized to the case that exactness fails, thereby producing a convenient rounding scheme. Specifically, we can consider the right-hand side of (47) as taken from a rank-d *truncated* singular value decomposition of Y^* (so that $\bar{Y}$ is an orthogonal transform of the best rank-d approximation of Y^*), multiply $\bar{Y}$ by an orientation-reversing element of $\mathrm{O}(d)$ according to whether a majority of its block elements have positive or negative determinant, and then project each of the blocks of $\bar{Y}$ to the nearest rotation matrix.[15] This generalized rounding scheme is formalized as Algorithm 2.

5.3. The complete algorithm

Combining the efficient optimization approach of Section 5.1 with the rounding procedure of Section 5.2 produces *SE-Sync* (Algorithm 3), our proposed algorithm for synchronization over the special Euclidean group.

Algorithm 3 The SE-Sync algorithm

Input: An initial point $Y \in \mathrm{St}(d,r_0)^n$, $r_0 \geq d+1$.
Output: A feasible estimate $\hat{x} \in \mathrm{SE}(d)^n$ for the maximum-likelihood estimation Problem 1 and the lower bound p^*_{SDP} for the optimal value of Problem 1.

```
function SE-Sync(Y)
    Set Y* ← RiemannianStaircase(Y).
    Set p*_SDP ← F(Q̃ Y*^T Y*).
    Set R̂ ← RoundSolution(Y*).
    Recover the optimal translational estimates t̂
    [1] corresponding to R̂ via (21).
    Set x̂ ← (t̂, R̂).
    return {x̂, p*_SDP}
end function
```

When applied to an instance of $\mathrm{SE}(d)$ synchronization, SE-Sync returns a feasible point $\hat{x} \in \mathrm{SE}(d)^n$ for the maximum-likelihood estimation Problem 1 and the lower bound $p^*_{\mathrm{SDP}} \leq p^*_{\mathrm{MLE}}$ for the optimal value of Problem 1. This lower bound provides an *upper* bound on the suboptimality of *any* feasible point $x = (t, R) \in \mathrm{SE}(d)^n$ as a solution of Problem 1 according to

$$F(\tilde{Q}R^{\mathsf{T}}R) - p^*_{\mathrm{SDP}} \geq F(\tilde{Q}R^{\mathsf{T}}R) - p^*_{\mathrm{MLE}} \qquad (50)$$

Furthermore, in the case that Problem 7 is exact, the estimate $\hat{x} = (\hat{t}, \hat{R}) \in \mathrm{SE}(d)^n$ returned by Algorithm 3 *attains* this lower bound:

$$F(\tilde{Q}\hat{R}^{\mathsf{T}}\hat{R}) = p^*_{\mathrm{SDP}} \qquad (51)$$

Consequently, verifying *a posteriori* that (51) holds provides a *computational certificate* of $\hat{x}$'s correctness as a solution of the maximum-likelihood estimation Problem 1. SE-Sync is thus a *certifiably correct* algorithm for SE(d) synchronization, as claimed.

6. Experimental results

In this section, we evaluate the performance of SE-Sync on a variety of special Euclidean synchronization problems drawn from the motivating application of pose-graph SLAM. As a basis for comparison, we also evaluate the performance of the Powell's Dog-Leg optimization method (Powell, 1970) using the Gauss–Newton local quadratic model (PDL-GN), a state-of-the-art approach for solving large-scale instances of the special Euclidean synchronization problem in robotics and computer vision applications (Rosen et al., 2014).

All of the following experiments are performed on a Dell Precision 5510 laptop with an Intel Xeon E3-1505M 2.80 GHz processor and 16 GB of RAM running Ubuntu 16.04. Our experimental implementation of SE-Sync [16] is written in C++ (Rosen and Carlone, 2017), and we compare it against the Powell's Dog-Leg implementation supplied by GTSAM,[17] a highly optimized, state-of-the-art software library specifically designed for large-scale SLAM and bundle adjustment applications (Dellaert, 2012). Each optimization algorithm is limited to a maximum of 1,000 (outer) iterations, and each outer iteration of the RTR algorithm employed in SE-Sync is limited to a maximum of 10,000 Hessian-vector product operations; convergence is declared whenever the relative decrease in function value between two subsequent (accepted) iterations is less than 10^{-6}.[18] The Powell's Dog-Leg method is initialized using the *chordal initialization*, a state-of-the-art method for bootstrapping an initial solution in SLAM and bundle adjustment problems (Carlone et al., 2015b; Martinec and Pajdla, 2007), and we set $r_0 = 5$ in the Riemannian staircase (Algorithm 1). Finally, as SE-Sync is based upon solving the (convex) semidefinite relaxation Problem 7, it does not require a high-quality initialization to reach a globally optimal solution; nevertheless, it can still benefit (in terms of reduced computation time) from being supplied with one. Consequently, in the following experiments we employ two initialization procedures in conjunction with SE-Sync: the first (random) simply samples a point uniformly randomly from $\mathrm{St}(d, r_0)^n$, while the second (chordal) supplies the same chordal initialization that the Powell's Dog-Leg method receives, to enable a fair comparison of the algorithms' computational speeds.

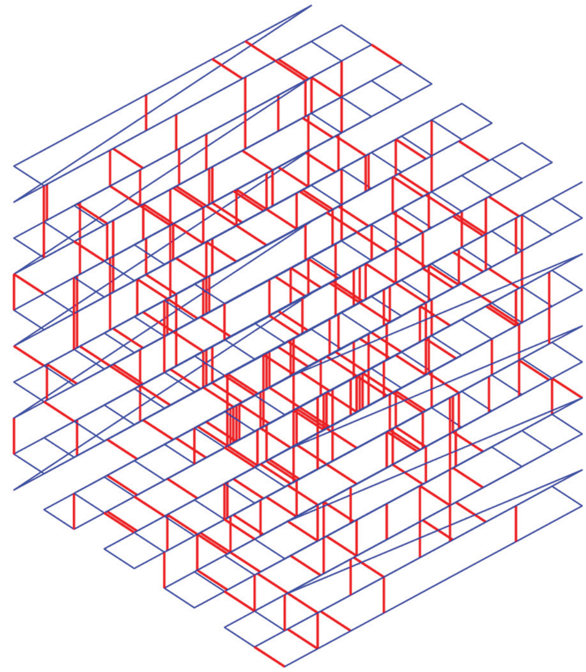


Fig. 2. The ground-truth configuration of an instance of the cube dataset with $s = 10$ and $p_{LC} = 0.1$. The robot's trajectory (with associated odometric measurements) is drawn in blue, and loop closure observations in red.

6.1. *Cube experiments*

In this first set of experiments, we are interested in investigating how the performance of SE-Sync is affected by factors such as measurement noise, measurement density, and problem size. To that end, we conduct a set of simulation studies that enable us to interrogate each of these factors individually. Concretely, we revisit the cube experiments considered in our previous work (Carlone et al., 2015a); this scenario simulates a robot traveling along a rectilinear path through a regular cubical lattice with a side length of s poses (Figure 2). An odometric measurement is available between each pair of sequential poses, and measurements between nearby non-sequential poses are available with probability p_{LC}; the measurement values $\tilde{x}_{ij} = (\tilde{t}_{ij}, \tilde{R}_{ij})$ themselves are sampled according to (10). We fix default values for these parameters at $\kappa = 16.67$ (corresponding to an expected angular root-mean-squared (RMS) error of 10° for the rotational measurements $\tilde{R}_{ij}$, cf. (63) in Appendix A), $\tau = 75$ (corresponding to an expected RMS error of 0.20 m for the translational measurements $\tilde{t}_{ij}$), $p_{LC} = 0.1$, and $s = 10$ (corresponding to a default problem size of 1,000 poses), and consider the effect of varying each of them individually; our complete dataset consists of 50 realizations of the cube sampled from the generative model just described for *each* joint setting of the parameters κ, τ, p_{LC}, and s. Results for these experiments are shown in Figure 3.

Consistent with our previous findings (Carlone et al., 2015a), these results suggest that the exactness of the semidefinite relaxation (33) depends primarily upon the level of noise corrupting the rotational observations $\tilde{R}_{ij}$ in (10). Furthermore, we see from Figure 3(a) that in these experiments, exactness obtains for rotational noise with an RMS angular error up to about 20°; this is roughly an order of magnitude greater than the level of noise affecting sensors typically deployed in robotics and computer vision applications, which provides strong empirical evidence that SE-Sync is capable of recovering certifiably globally optimal solutions of pose-graph SLAM problems under "reasonable" operating conditions.

In addition to its ability to recover certifiably optimal solutions, examining the center column of Figure 3 reveals that SE-Sync is also significantly faster than the Gauss–Newton-based approach that underpins current state-of-the-art pose-graph SLAM algorithms (Grisetti et al., 2010; Kaess et al., 2012; Kümmerle et al., 2011; Rosen et al., 2014). Given that SE-Sync performs direct *global* optimization, whereas these latter methods are purely *local* search techniques, this observation may at first seem somewhat counterintuitive. However, we can attribute SE-Sync's good computational performance to two key design decisions that distinguish SE-Sync from more traditional Gauss–Newton-based techniques. First, SE-Sync makes use of the *exact* Hessian (cf. Section 5.1.3), whereas the Gauss–Newton model, by construction, uses an approximation whose quality degrades in the presence of either large measurement residuals *or* strong nonlinearities in the underlying objective function (cf. e.g. Rosen et al., 2014, Section III-B), both of which are typical features of SLAM problems. This implies that the quadratic model function (cf. e.g. (Nocedal and Wright, 2006, Chp. 2)) that SE-Sync employs better captures the shape of the underlying objective than the one used by Gauss–Newton, so that the former is capable of computing higher-quality update steps. Second, and more significantly, SE-Sync makes use of a truncated-Newton method (RTR), which avoids the need to explicitly form or factor the (Riemannian) Hessian $\operatorname{Hess} F(Y)$; instead, at each iteration this approach *approximately* solves the Newton equations using a truncated conjugate gradient algorithm (Golub and Loan, 1996, Chapter 10),[19] and only computes this approximate solution as accurately as is necessary to ensure adequate progress towards a critical point with each applied update. The result is that RTR requires only a few sparse matrix–vector multiplications to obtain each update step; moreover, Equations (24b), (24c), (39), and (41)–(44) show that the constituent matrices involved in the computation of these products are *constant*, and can therefore be precomputed and cached at the beginning of the SE-Sync algorithm. In contrast, standard Gauss–Newton-based methods must recompute and refactor the Jacobian at *each* iteration. As the linearization of the objective function typically comprises the majority of the computational effort required in each iteration of a nonlinear optimizer, SE-Sync's exploitation of the particular geometry of the special Euclidean synchronization problem to avoid this expensive step enables it to realize substantial computational savings versus a standard Gauss–Newton-based approach.

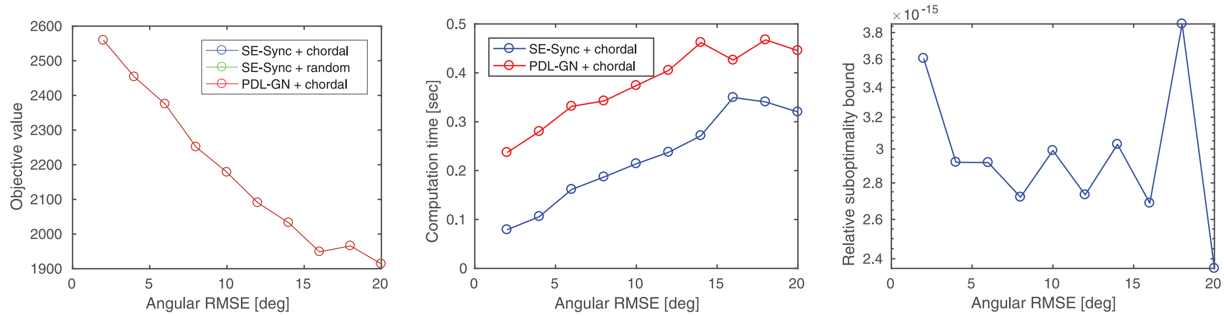
6.2. SLAM benchmark datasets

The experiments in the previous section made extensive use of simulated cube datasets to investigate the effects of measurement noise, measurement density, and problem size on SE-Sync's performance. In this next set of experiments, we evaluate SE-Sync on a suite of larger and more heterogeneous 2D and 3D pose-graph SLAM benchmarks that better represent the distribution of problems encountered in real-world SLAM applications. Five of these (the manhattan, city, sphere, torus, and grid datasets) are also synthetic (although generated using an observation model different from (10)), while the remainder (the csail, intel, ais2klinik, garage, cubicle , and rim datasets) are real-world examples (Figures 4 and 5). For the purpose of these experiments, we restrict attention to the case in which SE-Sync is supplied with the chordal initialization, and once again compare it with the Powell's Dog-Leg method. Results for these experiments are listed in Tables 1 and 2.

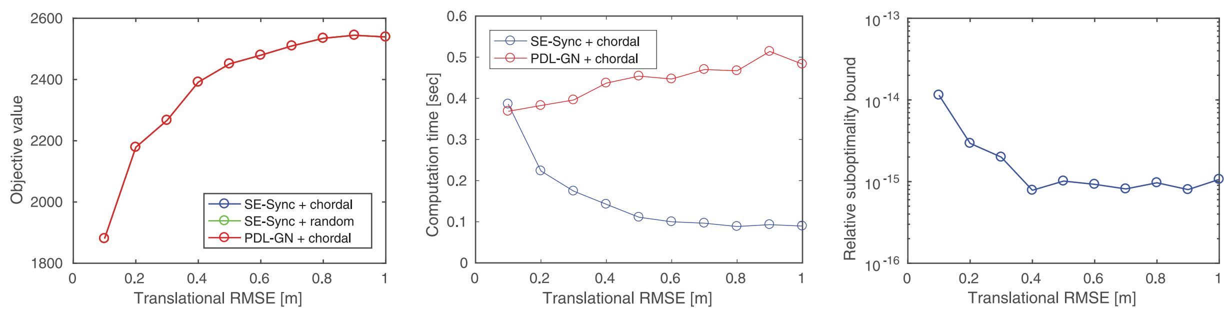
On each of these examples, both SE-Sync and Powell's Dog-Leg converged to the same (globally optimal) solution. However, consistent with our findings in Section 6.1, SE-Sync did so considerably faster, outperforming Powell's Dog-Leg on all but one of these examples (the garage dataset), by a factor of 3.05 on average (excluding the extreme case of the grid dataset, where SE-Sync outperformed Powell's Dog-Leg by an impressive factor of 27). These results further support our claim that SE-Sync provides an effective means of recovering certifiably globally optimal pose-graph SLAM solutions under real-world operating conditions, and does so significantly faster than current state-of-the-art Gauss–Newton-based alternatives.

7. Conclusion

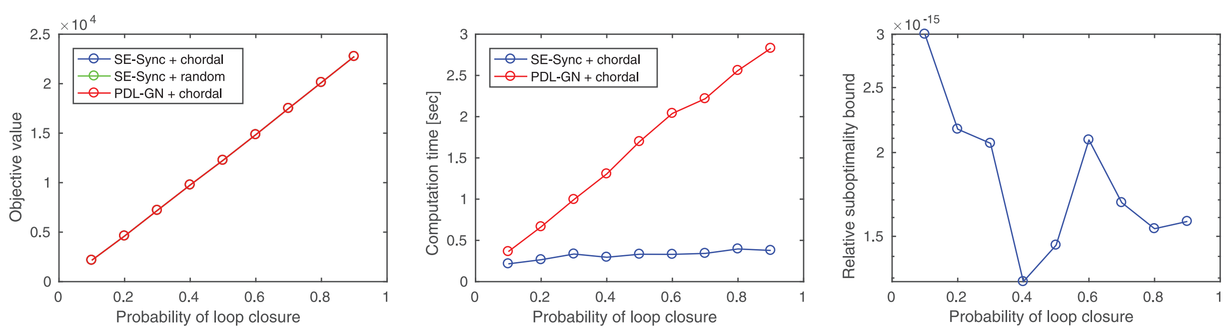
In this paper, we presented SE-Sync, a certifiably correct algorithm for synchronization over the special Euclidean group. Our algorithm is based upon the development of a novel semidefinite relaxation of the special Euclidean synchronization problem whose minimizer provides an *exact*, *globally optimal* solution so long as the magnitude of the noise corrupting the available measurements falls below a certain critical threshold, and employs a specialized, structure-exploiting Riemannian optimization method to solve large-scale instances of this semidefinite relaxation



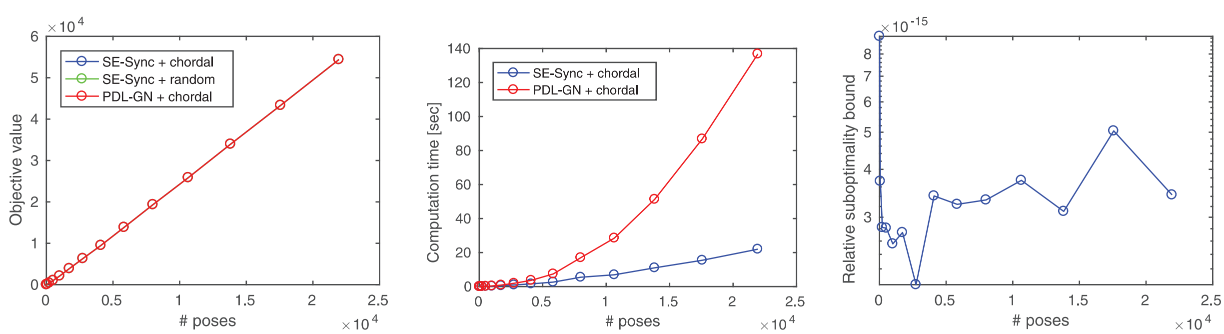
(a) Varying κ while holding $\tau = 75$, $p_{LC} = 0.1$, $s = 10$.



(b) Varying τ while holding $\kappa = 16.67$, $p_{LC} = 0.1$, $s = 10$.



(c) Varying p_{LC} while holding $\kappa = 16.67$, $\tau = 75$, $s = 10$.



(d) Varying s while holding $\kappa = 16.67$, $\tau = 75$, $p_{LC} = 0.1$.

Fig. 3. Results for the cube experiments. These figures plot the median of the objective values (left column) and elapsed computation times (center column) attained by the Powell's Dog-Leg and SE-Sync algorithms, as well as the upper bound $(F(\tilde{Q}\hat{R}^{\mathsf{T}}\hat{R}) - p^*_{\text{SDP}})/p^*_{\text{SDP}}$ for the relative suboptimality of the solution recovered by SE-Sync (right column), for 50 realizations of the cube dataset as functions of the measurement precisions κ (first row) and τ (second row), the loop closure probability p_{LC} (third row), and the problem size (fourth row).

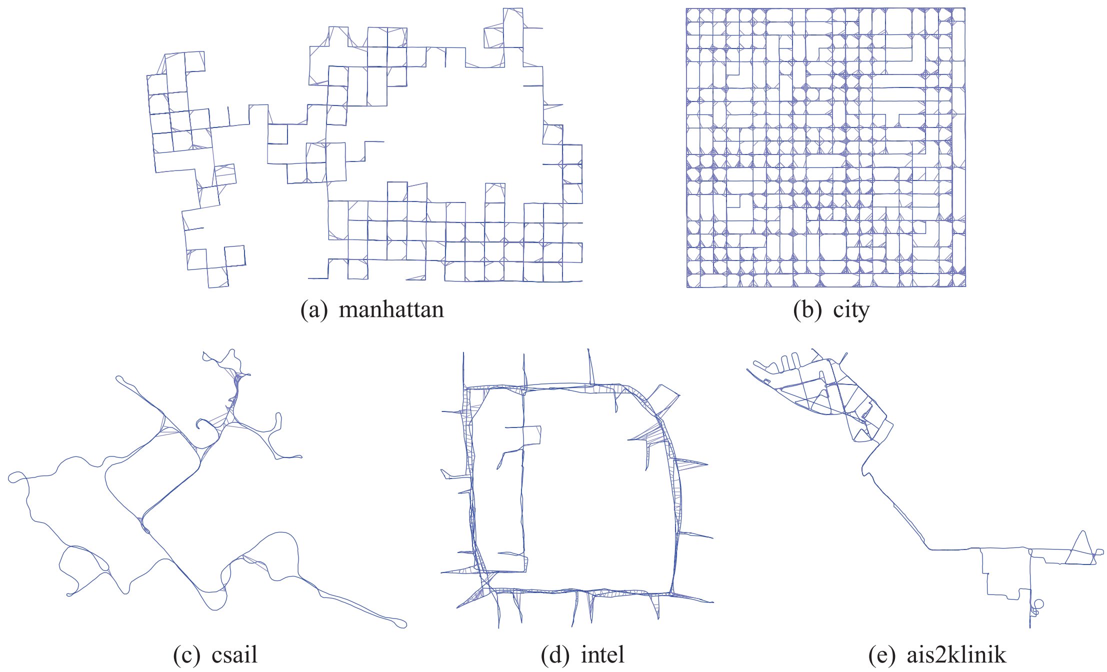


Fig. 4. Globally optimal solutions for the 2D SLAM benchmark datasets listed in Table 1.

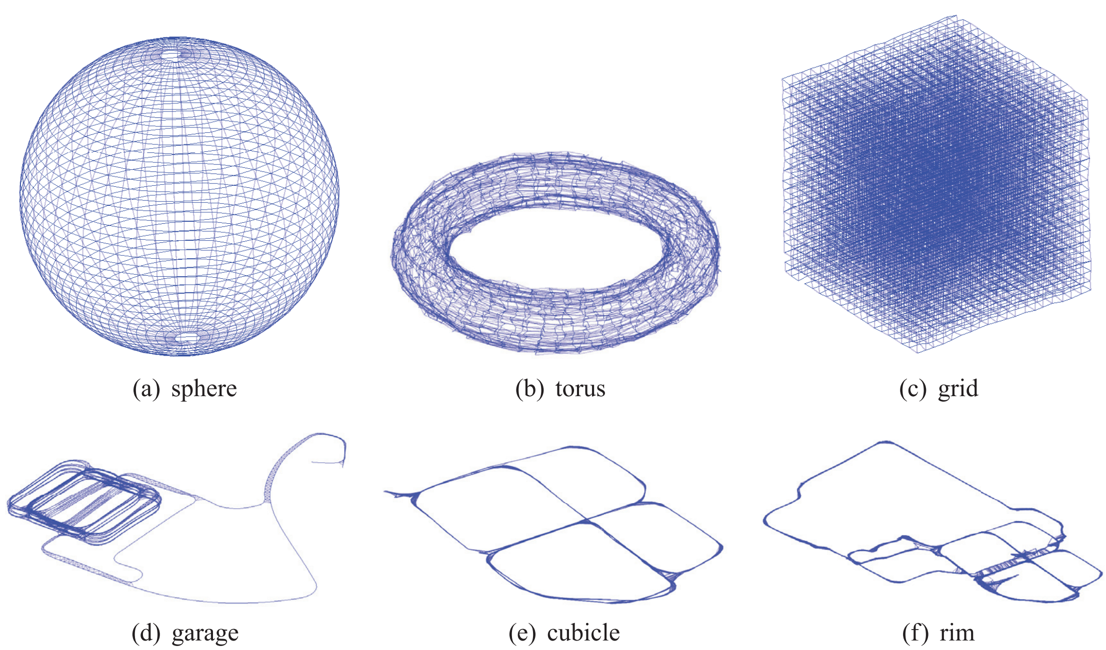


Fig. 5. Globally optimal solutions for the 3D SLAM benchmark datasets listed in Table 2.

efficiently. Experimental evaluation on a variety of simulated and real-world pose-graph SLAM datasets shows that SE-Sync is capable of recovering globally optimal solutions when the available measurements are corrupted by noise up to an order of magnitude greater than that typically encountered in robotics and computer vision applications, and does so significantly faster than the Gauss–Newton-based approach that forms the basis of current state-of-the-art techniques.

In addition to enabling the computation of certifiably correct solutions under *nominal* operating conditions, we believe that SE-Sync may also be extended to support provably robust and statistically efficient estimation in the case that some fraction of the available measurements $\tilde{x}_{ij}$ in (10) are contaminated by outliers. Our basis for this belief is the observation that Proposition 2 together with the experimental results of Section 6 imply that, under typical operating conditions, the maximum-likelihood estimation Problem 1

Table 1. Results for the 2D SLAM benchmark datasets.

	# Poses	# Measurements	PDL-GN		SE-Sync		
			Objective value	Time (s)	Objective value	Time (s)	Rel. suboptimality
manhattan	3,500	5,453	6.432×10^3	0.268	6.432×10^3	0.080	3.677×10^{-15}
city	10,000	20,687	6.386×10^2	1.670	6.386×10^2	1.576	1.015×10^{-14}
csail	1,045	1,172	3.170×10^1	0.029	3.170×10^1	0.010	7.844×10^{-16}
intel	1,728	2,512	5.235×10^1	0.120	5.235×10^1	0.071	1.357×10^{-16}
ais2klinik	15,115	16,727	1.885×10^2	12.472	1.885×10^2	1.981	2.412×10^{-15}

Table 2. Results for the 3D SLAM benchmark datasets.

	# Poses	# Measurements	PDL-GN		SE-Sync		
			Objective value	Time (s)	Objective value	Time (s)	Rel. suboptimality
sphere	2,500	4,949	1.687×10^3	0.704	1.687×10^3	0.580	1.890×10^{-15}
torus	5,000	9,048	2.423×10^4	1.963	2.423×10^4	0.284	5.256×10^{-15}
grid	8,000	22,236	8.432×10^4	46.343	8.432×10^4	1.717	2.934×10^{-15}
garage	1,661	6,275	1.263×10^0	0.415	1.263×10^0	0.468	1.618×10^{-14}
cubicle	5,750	16,869	7.171×10^2	2.456	7.171×10^2	0.754	2.061×10^{-15}
rim	10,195	29,743	5.461×10^3	6.803	5.461×10^3	2.256	5.663×10^{-15}

is equivalent to a *low-rank* convex program with a *linear* observation model and a *compact* feasible set (Problem 7); in contrast to general nonlinear estimation, this class of problems enjoys a beautiful geometric structure (Chandrasekaran et al., 2012) that has already been shown to enable remarkably robust recovery, even in the presence of gross contamination (Candès et al., 2011; Zhou et al., 2010). We intend to investigate this possibility in future research.

Finally, although the specific relaxation (33) underpinning SE-Sync was obtained by exploiting the well-known Lagrangian duality between quadratically constrained quadratic programs and semidefinite programs (Luo et al., 2010), recent work in real algebraic geometry has revealed the remarkable fact that the much broader class of (*rational*) *polynomial optimization problems*[20] also admits a hierarchy of semidefinite relaxations that is likewise frequently exact (or can be made arbitrarily sharp) when applied to real-world problem instances (Bugarin et al., 2016; Lasserre, 2001, 2006; Laurent, 2009; Nie and Demmel, 2008; Nie et al., 2006; Parrilo, 2003; Waki et al., 2006). Given the broad generality of this latter class of models, SE-Sync's demonstration that it is indeed possible to solve surprisingly large (but suitably structured) semidefinite programs with the temporal and computational resources typically available on mobile autonomous platforms suggests the further possibility of designing a broad class of practically-effective certifiably correct algorithms for robust machine perception based upon structured semidefinite programming relaxations. It is our hope that this report will encourage further investigation of this exciting possibility for machine perception.[21]

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the Office of Naval Research (grant numbers N00014-11-1-0688 and N00014-16-1-2628) and by the National Science Foundation (grant numbers IIS-1318392, DMS-1317308, DMS-1712730, and DMS-1719545).

Notes

1. The "synchronization" nomenclature originates with the prototypical example of this class of problems: synchronization of clocks over a communications network (Karp et al., 2003; Giridhar and Kumar, 2006) (corresponding to synchronization over the additive group $\mathbb{R}$).
2. This reprojection operation is often referred to as *rounding*.
3. Note that our definitions of directed and undirected graphs exclude loops and parallel edges. Although all of our results can be straightforwardly generalized to admit parallel edges (and indeed our experimental implementation of SE-Sync supports them), we have adopted this restriction to simplify the following presentation.
4. If G is not connected, then the problem of estimating the unknown states $x_1, \ldots, x_n$ decomposes into a set of independent estimation problems that are in one-to-one correspondence with the connected components of G; thus, the general case is always reducible to the case of connected graphs.
5. We use a directed graph to model the measurements $\tilde{x}_{ij}$ sampled from (10) because the distribution of the noise corrupting the latent values $\underline{x}_{ij}$ is not invariant under the group inverse operation of SE(d), as can be seen by composing (10) with (2b). Consequently, we must keep track of which state x_i was the "base frame" for each measurement.

6. Note that a minimizer of Problem 1 is *a* maximum-likelihood estimate (rather than *the* maximum-likelihood estimate) because Problem 1 always has multiple (in fact, infinitely many) solutions: as the objective function in (12) is constructed from *relative* measurements of the form $x_i^{-1}x_j$, if $x^* = (x_1^*, \ldots, x_n^*) \in \mathrm{SE}(d)^n$ minimizes (12), then $g \bullet x^* \triangleq (g \cdot x_1^*, \ldots, g \cdot x_n^*)$ also minimizes (12) for all $g \in \mathrm{SE}(d)$. Consequently, the solution set of Problem 1 is organized into orbits of the diagonal action $\bullet$ of $\mathrm{SE}(d)$ on $\mathrm{SE}(d)^n$. This gauge symmetry simply corresponds to the fact that *relative* measurements $x_i^{-1}x_j$ provide no information about the *absolute* values of the states x_i.
7. There is also some empirical evidence that the relaxation from $\mathrm{SO}(d)$ to $\mathrm{O}(d)$ is not the limiting factor in the exactness of our approach. In our prior work (Tron et al., 2015), we observed that in the specific case $d = 3$ it is possible to replace the (cubic) determinantal constraint in (1b) with an equivalent *quadratic* constraint by using the cross-product operation on the columns of each R_i to enforce the correct orientation; this leads to an equivalent formulation of Problem 1 as a quadratically constrained quadratic program that can be relaxed directly to a semidefinite program (Luo et al., 2010) *without* the intermediate relaxation through $\mathrm{O}(d)$. We found no significant difference between the sharpness of the relaxation incorporating the determinantal constraint and the relaxation without (Problem 7).
8. Note that Problem 7 could conceivably have multiple solutions, only *some* of which belong to the class specified in Theorem 1; in that case, it is possible that a numerical optimization method might converge to a minimizer of Problem 7 that does *not* correspond to a solution of Problem 1.
9. Ideally, one would like to have both (i) an explicit (i.e. closed-form) expression that lower-bounds the magnitude β of the admissible deviation of the data matrix $\tilde{Q}$ from its exact value $\underline{Q}$ (as measured in some suitable norm) and (ii) a concentration inequality (Tropp, 2015) (or several) that upper-bounds the probability $p(\|\tilde{Q} - \underline{Q}\| > \delta)$ of large deviations; together, these would enable the derivation of a lower bound on the probability that a given realization of Problem 4 sampled from the generative model (10) admits an exact semidefinite relaxation (33). Although it is possible (with a bit more effort) to derive such lower bounds on β using straightforward (although somewhat tedious) quantitative refinements of the continuity argument given in Appendix C, to date the sharpest concentration inequalities that we have been able to derive appear to be significantly suboptimal, and therefore lead to estimates for the probability of exactness that are grossly conservative versus what we observe empirically (cf. also the discussion in Remark 4.6 and Section 5 of (Bandeira et al., 2017)). Consequently, we have chosen to state Proposition 2 as a simple existence result for β to simplify its presentation and proof, while still providing some rigorous justification for our convex relaxation approach. We remark that as a practical matter, we have already shown in our previous work (Carlone et al., 2015a) (and do so again here in Section 6) that Problem 7 in fact remains exact with high probability when the measurements $\tilde{x}_{ij}$ in (10) are corrupted with noise up to an order of magnitude greater than what is encountered in typical robotics and computer vision applications; consequently, we leave the derivation of sharper concentration inequalities and explicit lower bounds on the probability of exactness to future research.
10. This is the transpose of a QR decomposition (Golub and Loan, 1996, Section 5.2).
11. That is, a point satisfying $\operatorname{grad} F(Y) = 0$ and $\operatorname{Hess} F(Y) \succeq 0$ (cf. (41)–(44)).
12. Note that because *every* $Y \in \mathrm{St}(d, r)^n$ is row rank-deficient for $r > dn$, this procedure is guaranteed to recover an optimal solution after searching at most $dn + 1$ levels of the hierarchy (38).
13. We point out that Equations (42), (43), and (44) correspond to equations (7), (8), and (9) in Boumal (2015), with the caveat that Boumal's definition of Y is the *transpose* of ours. Our notation follows the more common convention (cf. e.g. Edelman et al., 1998) that elements of a Stiefel manifold are matrices with orthonormal *columns*, rather than *rows*.
14. The requirement of high precision here is not superfluous: because Proposition 3 requires the identification of *rank-deficient* second-order critical points, whatever local search technique we apply to Problem 9 must be capable of numerically approximating a critical point precisely enough that its rank can be correctly determined.
15. This is the *special orthogonal Procrustes* problem, which admits a simple closed-form solution based upon the singular value decomposition (Hanson and Norris, 1981; Umeyama, 1991).
16. Available at https://github.com/david-m-rosen/SE-Sync.
17. Version 4.0, available at https://bitbucket.org/gtborg/gtsam.
18. Ordinarily, SE-Sync employs (preconditioned) gradient-based stopping criteria to directly control the precision of the estimate obtained for a first-order critical point. However, GTSAM does not implement a gradient-based stopping condition, so for the purpose of these experiments, we have disabled SE-Sync's gradient-based stopping criteria, and used a relative-decrease-based stopping condition for both algorithms (with a sufficiently tight tolerance to ensure that each obtains a high-precision solution) to enable a fair comparison of their computational speeds.
19. Hence the nomenclature.
20. These are nonlinear programs in which the feasible set is a *semialgebraic set* (the set of real solutions of a system of polynomial (in)equalities) and the objective is a (rational) polynomial function.
21. This report is an extended version of a paper originally presented at the 12th International Workshop on the Algorithmic Foundations of Robotics (Rosen et al., 2016). An earlier draft appeared as a technical report issued by the Computer Science and Artificial Intelligence Laboratory of the Massachusetts Institute of Technology (Rosen et al., 2017).

References

Åström K and Murray R (2008) *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton, NJ: Princeton University Press.

Absil PA, Baker C and Gallivan K (2007) Trust-region methods on Riemannian manifolds. *Foundations of Computational Mathematics* 7(3): 303–330.

Absil PA, Mahony R and Sepulchre R (2008) *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press.

Alizadeh F, Haeberly JPA and Overton M (1997) Complementarity and nondegeneracy in semidefinite programming. *Mathematical Programming* 77(1): 111–128.

Arrigoni F, Rossi B and Fusiello A (2016) Spectral synchronization of multiple views in SE(3). *SIAM Journal on Imaging Sciences* 9(4): 1963–1990.

Bandeira A (2016) A note on probably certifiably correct algorithms. *Comptes Rendus Mathematique* 354(3): 329–333.

Bandeira A, Boumal N and Singer A (2017) Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming* 163(1–2): 145–167.

Bandeira A, Singer A and Spielman D (2013) A Cheeger inequality for the graph connection Laplacian. *SIAM Journal on Matrix Analysis and Applications* 34(4): 1611–1630.

Best D and Fisher N (1979) Efficient simulation of the von Mises distribution. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 28(2): 152–157.

Biggs N (1997) Algebraic potential theory on graphs. *Bulletin of London Mathematical Society* 29: 641–682.

Boothby W (2003) *An Introduction to Differentiable Manifolds and Riemannian Geometry* (2nd edn). London: Academic Press.

Boumal N (2015) A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints. Preprint: arXiv:1506.00575v2.

Boumal N, Absil PA and Cartis C (2016a) Global rates of convergence for nonconvex optimization on manifolds. Preprint: arXiv:1605.08101v1.

Boumal N, Singer A, Absil PA and Blondel V (2014) Cramér–Rao bounds for synchronization of rotations. *Information and Inference* 3: 1–39.

Boumal N, Voroninski V and Bandeira A (2016b) The non-convex Burer-Monteiro approach works on smooth semidefinite programs. Preprint arXiv:1606.04970v1.

Boyd S and Vandenberghe L (2004) *Convex Optimization*. Cambridge: Cambridge University Press.

Bugarin F, Henrion D and Lasserre J (2016) Minimizing the sum of many rational functions. *Mathematical Programming Computation* 8: 83–111.

Burer S and Monteiro R (2003) A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming* 95: 329–357.

Burer S and Monteiro R (2005) Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming* 103: 427–444.

Burschka D, Corso J, Dewan M, et al. (2005) Navigating inner space: 3-D assistance for minimally invasive surgery. *Robotics and Autonomous Systems* 52(1): 5–26.

Candès E, Li X, Ma Y and Wright J (2011) Robust principal component analysis? *Journal of the ACM* 58(3): Article 11.

Carlone L, Rosen D, Calafiore G, Leonard J and Dellaert F (2015a) Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany.

Carlone L, Tron R, Daniilidis K and Dellaert F (2015b) Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization. In: *IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, WA, pp. 4597–4605.

Chandrasekaran V, Recht B, Parrilo P and Willsky A (2012) The convex geometry of linear inverse problems. *Foundations of Computational Mathematics* 12: 805–849.

Chiuso A, Picci G and Soatto S (2008) Wide-sense estimation on the special orthogonal group. *Communications in Information and Systems* 8(3): 185–200.

Cucuringu M, Lipman Y and Singer A (2012) Sensor network localization by eigenvector synchronization over the special Euclidean group. *ACM Transactions on Sensor Networks* 8(3): 19:1–19:42.

Dellaert F (2012) Factor graphs and GTSAM: A hands-on introduction. Technical Report GT-RIM-CP&R-2012-002, Institute for Robotics & Intelligent Machines, Georgia Institute of Technology.

Dellaert F, Carlson J, Ila V, Ni K and Thorpe C (2010) Subgraph-preconditioned conjugate gradients for large scale SLAM. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, pp. 2566–2571.

Dellaert F and Kaess M (2006) Square Root SAM: Simultaneous localization and mapping via square root information smoothing. *The International Journal of Robotics Research* 25(12): 1181–1203.

Dembo R and Steihaug T (1983) Truncated-Newton algorithms for large-scale unconstrained optimization. *Mathematical Programming* 26: 190–212.

Edelman A, Arias T and Smith S (1998) The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications* 20(2): 303–353.

Fallon M, Kuindersma S, Karumanchi S, et al. (2015) An architecture for online affordance-based perception and whole-body planning. *The International Journal of Robotics Research* 32(2): 229–254.

Ferguson T (1996) *A Course in Large Sample Theory*. Boca Raton, FL: Chapman & Hall/CRC.

Fisher R (1953) Dispersion on a sphere. *Proceedings of the Royal Society* 217: 295–305.

Gallier J (2010) The Schur complement and symmetric positive semidefinite (and definite) matrices. Unpublished note, available online: http://www.cis.upenn.edu/ jean/schur-comp.pdf.

Giridhar A and Kumar P (2006) Distributed clock synchronization over wireless networks: Algorithms and analysis. In: *IEEE Conference on Decision and Control*, San Diego, CA, pp. 4915–4920.

Golub G and Loan CV (1996) *Matrix Computations* (3rd edn). Baltimore, MD: Johns Hopkins University Press.

Grisetti G, Kümmerle R, Stachniss C and Burgard W (2010) A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine* 2(4): 31–43.

Grisetti G, Stachniss C and Burgard W (2009) Nonlinear constraint network optimization for efficient map learning. *IEEE Transactions on Intelligent Transportation Systems* 10(3): 428–439.

Guillemin V and Pollack A (1974) *Differential Topology*. Englewood Cliffs, NJ: Prentice-Hall.

Hanson R and Norris M (1981) Analysis of measurements based on the singular value decomposition. *SIAM Journal on Scientific and Statistical Computing* 2(3): 363–373.

Hartley R, Trumpf J, Dai Y and Li H (2013) Rotation averaging. *International Journal of Computer Vision* 103(3): 267–305.

Horn R and Johnson C (1991) *Topics in Matrix Analysis*. Cambridge: Cambridge University Press.

Huang S, Lai Y, Frese U and Dissanayake G (2010) How far is SLAM from a linear least squares problem? In: *IEEE/RSJ*

International Conference on Intelligent Robots and Systems (IROS), Taipei, Taiwan, pp. 3011–3016.

Huang S, Wang H, Frese U and Dissanayake G (2012) On the number of local minima to the point feature based SLAM problem. In: *IEEE International Conference on Robotics and Automation (ICRA)*, St. Paul, MN, pp. 2074–2079.

Johnson M, Shrewsbury B, Bertrand S, et al. (2015) Team IHMC's lessons learned from the DARPA Robotics Challenge Trials. *The International Journal of Robotics Research* 32(2): 192–208.

Kaess M, Johannsson H, Roberts R, Ila V, Leonard J and Dellaert F (2012) iSAM2: Incremental smoothing and mapping using the Bayes tree. *The International Journal of Robotics Research* 31(2): 216–235.

Karp R, Elson J, Estrin D and Shenker S (2003) Optimal and global time synchronization in sensornets. Technical report, Center for Embedded Network Sensing, University of California.

Kobayashi S and Nomizu K (1996) *Foundations of Differential Geometry*, vol. 1 (*Wiley Classics*). Hoboken, NJ: John Wiley & Sons, Inc.

Konolige K (2010) Sparse sparse bundle adjustment. In: *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 1–11.

Kümmerle R, Grisetti G, Strasdat H, Konolige K and Burgard W (2011) g2o: A general framework for graph optimization. In: *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, pp. 3607–3613.

Lasserre J (2001) Global optimization and the problem of moments. *SIAM Journal on Optimization* 11(3): 796–817.

Lasserre J (2006) Convergent SDP-relaxations in polynomial optimization with sparsity. *SIAM Journal on Optimization* 17(3): 822–843.

Laurent M (2009) Sums of squares, moment matrices and optimization over polynomials. In: Putinar M and Sullivant S (eds.) *Emerging Applications of Algebraic Geometry* (*The IMA Volumes in Mathematics and Its Applications*, vol. 149). New York: Springer, pp. 157–270.

Leonard J, How J, Teller S, et al. (2008) A perception-driven autonomous urban vehicle. *The International Journal of Robotics Research* 25(10): 727–774.

Lourakis M and Argyros A (2009) SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software* 36(1): 1–30.

Luo ZQ, Ma WK, So A, Ye Y and Zhang S (2010) Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine* 27(3): 20–34.

Martinec D and Pajdla T (2007) Robust rotation and translation estimation in multiview reconstruction. In: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, MN, pp. 1–8.

Meyer C (2000) *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA: The Society for Industrial and Applied Mathematics (SIAM).

Nash S (2000) A survey of truncated-Newton methods. *Journal of Computational and Applied Mathematics* 124: 45–59.

Nie J and Demmel J (2008) Sparse SOS relaxations for minimizing functions that are summations of small polynomials. *SIAM Journal on Optimization* 19(4): 1534–1558.

Nie J, Demmel J and Sturmfels B (2006) Minimizing polynomials via sum of squares over the gradient ideal. *Mathematical Programming* 106: 587–606.

Nocedal J and Wright S (2006) *Numerical Optimization* (2nd edn). New York: Springer Science+Business Media.

Olson E, Leonard J and Teller S (2006) Fast iterative alignment of pose graphs with poor initial estimates. In: *IEEE International Conference on Robotics and Automation (ICRA)*, Orlando, FL, pp. 2262–2269.

Olver F, Olde Daalhuis A, Lozier D, et al. (eds.) (2016) *NIST Digital Library of Mathematical Functions*, Release 1.0.13. Available online: http://dlmf.nist.gov/.

Özyeşil O, Singer A and Basri R (2015) Stable camera motion estimation using convex programming. *SIAM Journal on Imaging Sciences* 8(2): 1220–1262.

Parrilo P (2003) Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming* 96(2): 293–320.

Peters J, Borra D, Paden B and Bullo F (2015) Sensor network localization on the group of three-dimensional displacements. *SIAM Journal on Control and Optimization* 53(6): 3534–3561.

Powell M (1970) A new algorithm for unconstrained optimization. In: Rosen J, Mangasarian O and Ritter K (eds.) *Nonlinear Programming*. London: Academic Press, pp. 31–65.

Pratt G and Manzo J (2013) The DARPA Robotics Challenge. *IEEE Robotics & Automation Magazine* 20(2): 10–12.

Rosen D and Carlone L (2017) Computational enhancements for certifiably correct SLAM. Presented at the *International Conference on Intelligent Robots and Systems (IROS) in the workshop "Introspective Methods for Reliable Autonomy"*.

Rosen D, Carlone L, Bandeira A and Leonard J (2016) A certifiably correct algorithm for synchronization over the special Euclidean group. In: *International Workshop on the Algorithmic Foundations of Robotics (WAFR)*, San Francisco, CA.

Rosen D, Carlone L, Bandeira A and Leonard J (2017) SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group. Technical Report MIT-CSAIL-TR-2017-002, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA.

Rosen D, DuHadway C and Leonard J (2015) A convex relaxation for approximate global optimization in simultaneous localization and mapping. In: *IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, WA, pp. 5822–5829.

Rosen D, Kaess M and Leonard J (2014) RISE: An incremental trust-region method for robust online sparse least-squares estimation. *IEEE Transactions on Robotics* 30(5): 1091–1108.

Russell S and Norvig P (2010) *Artificial Intelligence: A Modern Approach* (3rd edn). Upper Saddle River, NJ: Prentice Hall.

Singer A (2011) Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis* 30(1): 20–36.

Singer A and Wu HT (2012) Vector diffusion maps and the connection Laplacian. *Communications on Pure and Applied Mathematics* 65: 1067–1144.

Stachniss C, Leonard J and Thrun S (2016) Simultaneous localization and mapping. In: *Springer Handbook of Robotics*. New York: Springer International Publishing, pp. 1153–1176.

Steihaug T (1983) The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal on Numerical Analysis* 20(3): 626–637.

Stengel R (1994) *Optimal Control and Estimation*. New York: Dover Publications.

Taylor R (2006) A perspective on medical robotics. *Proceedings of the IEEE* 94(9): 1652–1664.

Taylor R, Menciassi A, Fichtinger G and Dario P (2008) Medical robotics and computer-integrated surgery. In: *Springer Handbook of Robotics*. New York: Springer, pp. 1199–1222.

Thrun S, Montemerlo M, Dahlkamp H, et al. (2006) Stanley: The robot that won the DARPA Grand Challenge. *The International Journal of Robotics Research* 23(9): 661–692.

Todd M (2001) Semidefinite optimization. *Acta Numerica* 10: 515–560.

Tron R, Rosen D and Carlone L (2015) On the inclusion of determinant constraints in Lagrangian duality for 3D SLAM. Presented at *Robotics: Science and Systems (RSS) in the workshop "The Problem of Mobile Sensors"*.

Tron R, Zhou X and Daniilidis K (2016) A survey on rotation optimization in structure from motion. In: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Las Vegas, NV, pp. 77–85.

Tropp J (2015) An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning* 8(1–2): 1–230.

Umeyama S (1991) Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(4): 376–380.

Urmson C, Anhalt J, Bagnell D, et al. (2008) Autonomous driving in urban environments: Boss and the Urban Challenge. *The International Journal of Robotics Research* 25(8): 425–466.

Vandenberghe L and Boyd S (1996) Semidefinite programming. *SIAM Review* 38(1): 49–95.

Waki H, Kim S, Kojima M and Muramatsu M (2006) Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM Journal on Optimization* 17(1): 218–242.

Wang H, Hu G, Huang S and Dissanayake G (2012) On the structure of nonlinearities in pose graph SLAM. In: *Robotics: Science and Systems (RSS)*, Sydney, Australia.

Wang L and Singer A (2013) Exact and stable recovery of rotations for robust synchronization. *Information and Inference* 2(2): 145–193.

Warner F (1983) *Foundations of Differentiable Manifolds and Lie Groups* (*Springer Graduate Texts in Mathematics*). New York: Springer.

Zhou Z, Li X, Wright J, Candès E and Ma Y (2010) Stable principal component pursuit. In: *IEEE International Symposium on Information Theory (ISIT)*, Austin, TX, pp. 1518–1522.

Appendix A. The isotropic Langevin distribution

In this appendix, we provide a brief overview of the isotropic Langevin distribution on $\mathrm{SO}(d)$, with a particular emphasis on the important special cases $d = 2, 3$.

The *isotropic Langevin distribution* on $\mathrm{SO}(d)$ with mode $M \in \mathrm{SO}(d)$ and concentration parameter $\kappa \geq 0$, denoted $\mathrm{Langevin}(M, \kappa)$, is the distribution determined by the following probability density function (with respect to the Haar measure on $\mathrm{SO}(d)$ (Warner, 1983)):

$$p(X; M, \kappa) = \frac{1}{c_d(\kappa)} \exp\left(\kappa \operatorname{tr}(M^\mathsf{T} X)\right) \tag{52}$$

where $c_d(\kappa)$ is a normalization constant (Boumal et al., 2014; Chiuso et al., 2008). Note that the product $M^\mathsf{T} X = M^{-1} X \in \mathrm{SO}(d)$ appearing in (52) is the relative rotation sending M to X.

In general, given any $Z \in \mathrm{SO}(d)$, there exists some $U \in \mathrm{O}(d)$ such that

$$U^\mathsf{T} Z U = \begin{cases} \begin{pmatrix} R(\theta_1) & & \\ & \ddots & \\ & & R(\theta_k) \end{pmatrix}, & d \bmod 2 = 0 \\ \begin{pmatrix} R(\theta_1) & & & \\ & \ddots & & \\ & & R(\theta_k) & \\ & & & 1 \end{pmatrix}, & d \bmod 2 = 1 \end{cases} \tag{53}$$

where $k \triangleq \lfloor d/2 \rfloor$, $\theta_i \in [-\pi, \pi)$ for all $i \in [k]$, and

$$R(\theta) \triangleq \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \in \mathrm{SO}(2) \tag{54}$$

This *canonical decomposition* corresponds to the fact that every rotation $Z \in \mathrm{SO}(d)$ acts on $\mathbb{R}^d$ as a set of elementary rotations (54) of mutually orthogonal 2D subspaces. As $\operatorname{tr}(\cdot)$ is a class function, it follows from (53) and (54) that the trace appearing in the isotropic Langevin density (52) can be equivalently expressed as

$$\operatorname{tr}(M^\mathsf{T} X) = d \bmod 2 + 2 \sum_{i=1}^{k} \cos(\theta_i) \tag{55}$$

where θ_i are the rotation angles for each of the elementary rotations of $M^\mathsf{T} X$. Note, however, that although the right-hand side of (55) depends upon the *magnitudes* of these elementary rotations, it does *not* depend upon the *orientation* of their corresponding subspaces; this is the sense in which the Langevin density (52) is "isotropic."

For the special cases $d = 2, 3$, the normalization constant $c_d(\kappa)$ appearing in (52) admits the following simple closed forms (cf. Boumal et al., 2014, equations (4.6) and (4.7)):

$$c_2(\kappa) = I_0(2\kappa) \tag{56a}$$

$$c_3(\kappa) = \exp(\kappa)\left(I_0(2\kappa) - I_1(2\kappa)\right) \tag{56b}$$

where $I_n(z)$ denotes the *modified Bessel function of the first kind* (Olver et al., 2016, equation (10.32.3)):

$$I_n(z) \triangleq \frac{1}{\pi} \int_0^\pi e^{z \cos(\theta)} \cos(n\theta)\, d\theta, \quad n \in \mathbb{N} \tag{57}$$

Furthermore, in these dimensions every rotation acts on a single 2D subspace, and hence is described by a single rotation angle. Letting $\theta \triangleq \angle(M^\mathsf{T} X)$, it follows from (52) and (55) that the distribution over θ induced by $\mathrm{Langevin}(M, \kappa)$ has a density satisfying

$$p(\theta; \kappa) \propto \exp(2\kappa \cos(\theta)), \quad \theta \in [-\pi, \pi) \tag{58}$$

Now recall that the *von Mises distribution* on the circle (here identified with $[-\pi, \pi)$) with mode $\mu \in [-\pi, \pi)$ and

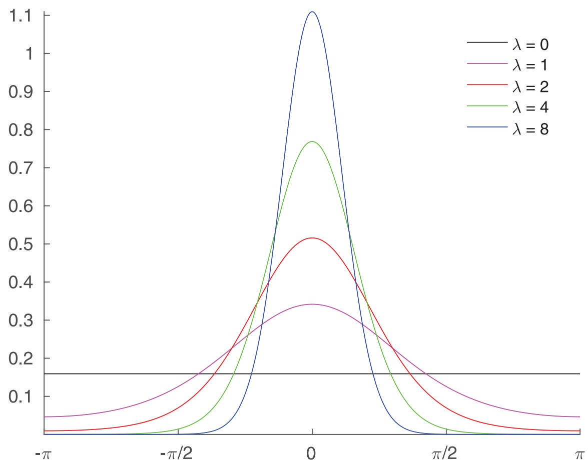


Fig. 6. The von Mises distribution. This plot shows the densities (59) of several von Mises distributions on the circle (here identified with $[-\pi, \pi)$) with common mode $\mu = 0$ and varying concentrations λ. For $\lambda = 0$ the von Mises distribution reduces to the uniform distribution, and its density asymptotically converges to the Gaussian $\mathcal{N}(0, \lambda^{-1})$ as $\lambda \to \infty$.

concentration parameter $\lambda \geq 0$, denoted by $\text{vM}(\mu, \lambda)$, is the distribution determined by the following probability density function:

$$p(\theta; \mu, \lambda) = \frac{\exp(\lambda \cos(\theta - \mu))}{2\pi I_0(\lambda)} \tag{59}$$

This distribution plays a role in circular statistics analogous to that of the Gaussian distribution on Euclidean space (Fisher, 1953). For $\lambda = 0$ it reduces to the uniform distribution on S^1, and becomes increasingly concentrated at the mode μ as $\lambda \to \infty$ (cf. Figure 6). In fact, considering the asymptotic expansion (Olver et al., 2016, 10.40.1):

$$I_0(\lambda) \sim \frac{\exp(\lambda)}{\sqrt{2\pi\lambda}}, \quad \lambda \to \infty \tag{60}$$

and the second-order Taylor series expansion of the cosine function:

$$\cos(\theta - \mu) \sim 1 - \frac{1}{2}(\theta - \mu)^2, \quad \theta \to \mu \tag{61}$$

it follows from (59)–(61) that

$$p(\theta; \mu, \lambda) \sim \frac{1}{\sqrt{2\pi\lambda^{-1}}} \exp\left(-\frac{(\theta - \mu)^2}{2\lambda^{-1}}\right), \quad \lambda \to \infty \tag{62}$$

which we recognize as the density for the Gaussian distribution $\mathcal{N}(\mu, \lambda^{-1})$.

These observations lead to a particularly convenient generative description of the isotropic Langevin distribution in dimensions 2 and 3: namely, a realization of $X \sim \text{Langevin}(M, \kappa)$ is obtained by perturbing the mode M by a rotation through an angle $\theta \sim \text{vM}(0, 2\kappa)$ about a uniformly distributed axis (Algorithm 4). Furthermore, it follows from (59) that the standard deviation of the angle θ of the relative rotation between M and X is given by

Algorithm 4 A sampler for the isotropic Langevin distribution on $\text{SO}(d)$ in dimensions 2 and 3

Input: Mode $M \in \text{SO}(d)$ with $d \in \{2, 3\}$, concentration parameter $\kappa \geq 0$.
Output: A realization of $X \sim \text{Langevin}(M, \kappa)$.

```
function SampleIsotropicLangevin(M, κ)
    Sample a rotation angle θ ∼ vM(0, 2κ).¹
    if d = 2 then
        Set perturbation matrix P ← R(θ).
    else                                        ▷ d = 3
        Sample an axis of rotation v̂ ∼ U(S²).
        Set perturbation matrix P ← exp(θ[v̂]×).
    end if
    return MP
end function
```

$$\text{SD}[\theta] = \sqrt{\int_{-\pi}^{\pi} \theta^2 \frac{\exp(2\kappa \cos(\theta))}{2\pi I_0(2\kappa)} \, d\theta} \tag{63}$$

which provides a convenient and intuitive measure of the dispersion of $\text{Langevin}(M, \kappa)$. The right-hand side of (63) can be efficiently evaluated to high precision via numerical quadrature for values of κ less than 150 (corresponding to an angular standard deviation of 3.31°). For $\kappa > 150$, one can alternatively use the estimate of the angular standard deviation coming from the asymptotic Gaussian approximation (62) for the von Mises distribution $\text{vM}(0, 2\kappa)$:

$$\text{SD}[\theta] \sim \frac{1}{\sqrt{2\kappa}}, \quad \kappa \to \infty \tag{64}$$

This approximation is accurate to within 1% for $\kappa > 12.87$ (corresponding to an angular standard deviation of 11.41°) and to within .1% for $\kappa > 125.3$ (corresponding to an angular standard deviation of 3.62°).

Appendix B. Reformulating the estimation problem

B.1. Deriving Problem 2 from Problem 1

In this section, we show how to derive Problem 2 from Problem 1. Using the fact that $\text{vec}(v) = v$ for any $v \in \mathbb{R}^{d \times 1}$ and the fact that $\text{vec}(XY) = (Y^T \otimes I_k)\,\text{vec}(X)$ for all $X \in \mathbb{R}^{m \times k}$ and $Y \in \mathbb{R}^{k \times l}$ (Horn and Johnson, 1991, Lemma 4.3.1), we

1. See e.g. (Best and Fisher, 1979)

can write each summand of (12) in a vectorized form as

$$\tau_{ij} \left\| t_j - t_i - R_i \tilde{t}_{ij} \right\|_2^2 = \tau_{ij} \left\| t_j - t_i - \left(\tilde{t}_{ij}^\mathsf{T} \otimes I_d \right) \operatorname{vec}(R_i) \right\|_2^2 = \left\| (\sqrt{\tau_{ij}} I_d)\, t_j - (\sqrt{\tau_{ij}} I_d)\, t_i - \sqrt{\tau_{ij}} \left(\tilde{t}_{ij}^\mathsf{T} \otimes I_d \right) \operatorname{vec}(R_i) \right\|_2^2 \tag{65}$$

and

$$\kappa_{ij} \left\| R_j - R_i \tilde{R}_{ij} \right\|_F^2 = \kappa_{ij} \left\| \operatorname{vec}(R_j) - \operatorname{vec}(R_i \tilde{R}_{ij}) \right\|_2^2 = \left\| \sqrt{\kappa_{ij}} (I_d \otimes I_d) \operatorname{vec}(R_j) - \sqrt{\kappa_{ij}} \left(\tilde{R}_{ij}^\mathsf{T} \otimes I_d \right) \operatorname{vec}(R_i) \right\|_2^2 \tag{66}$$

Letting $t \in \mathbb{R}^{dn}$ and $R \in \mathbb{R}^{d \times dn}$ denote the concatenations of the t_i and R_i as defined in (17), (65) and (66) imply that (12) can be rewritten in a vectorized form as

$$p^*_{\text{MLE}} = \min_{\substack{t \in \mathbb{R}^{dn} \\ R \in \mathrm{SO}(d)^n}} \left\| B \begin{pmatrix} t \\ \operatorname{vec}(R) \end{pmatrix} \right\|_2^2 \tag{67}$$

where the coefficient matrix $B \in \mathbb{R}^{(d+d^2)m \times (d+d^2)n}$ (with $m = |\vec{\mathcal{E}}|$) has the block decomposition

$$B \triangleq \begin{pmatrix} B_1 & B_2 \\ 0 & B_3 \end{pmatrix} \tag{68}$$

and $B_1 \in \mathbb{R}^{dm \times dn}$, $B_2 \in \mathbb{R}^{dm \times d^2 n}$, and $B_3 \in \mathbb{R}^{d^2 m \times d^2 n}$ are block-structured matrices whose block rows and columns are indexed by the elements of $\vec{\mathcal{E}}$ and $\mathcal{V}$, respectively, and whose (e, k)-block elements are given by

$$(B_1)_{ek} = \begin{cases} -\sqrt{\tau_{kj}} I_d, & e = (k,j) \in \vec{\mathcal{E}} \\ \sqrt{\tau_{ik}} I_d, & e = (i,k) \in \vec{\mathcal{E}} \\ 0_{d \times d}, & \text{otherwise} \end{cases} \tag{69a}$$

$$(B_2)_{ek} = \begin{cases} -\sqrt{\tau_{kj}} \left(\tilde{t}_{kj}^\mathsf{T} \otimes I_d \right), & e = (k,j) \in \vec{\mathcal{E}} \\ 0_{d \times d^2}, & \text{otherwise} \end{cases} \tag{69b}$$

$$(B_3)_{ek} = \begin{cases} -\sqrt{\kappa_{kj}} \left(\tilde{R}_{kj}^\mathsf{T} \otimes I_d \right), & e = (k,j) \in \vec{\mathcal{E}} \\ \sqrt{\kappa_{ik}} (I_d \otimes I_d), & e = (i,k) \in \vec{\mathcal{E}} \\ 0_{d \times d}, & \text{otherwise} \end{cases} \tag{69c}$$

We can further expand the squared ℓ_2-norm objective in (67) to obtain:

$$p^*_{\text{MLE}} = \min_{\substack{t \in \mathbb{R}^{dn} \\ R \in \mathrm{SO}(d)^n}} \begin{pmatrix} t \\ \operatorname{vec}(R) \end{pmatrix}^\mathsf{T} B^\mathsf{T} B \begin{pmatrix} t \\ \operatorname{vec}(R) \end{pmatrix} \tag{70}$$

with

$$B^\mathsf{T} B = \begin{pmatrix} B_1^\mathsf{T} B_1 & B_1^\mathsf{T} B_2 \\ B_2^\mathsf{T} B_1 & B_2^\mathsf{T} B_2 + B_3^\mathsf{T} B_3 \end{pmatrix} \tag{71}$$

Computing each of the constituent products in (71) blockwise using (69) (cf. Rosen et al., 2017, equations (72)–(84)), we obtain

$$B_1^\mathsf{T} B_1 = L(W^\tau) \otimes I_d \tag{72a}$$

$$B_1^\mathsf{T} B_2 = \tilde{V} \otimes I_d \tag{72b}$$

$$B_2^\mathsf{T} B_2 = \tilde{\Sigma} \otimes I_d \tag{72c}$$

$$B_3^\mathsf{T} B_3 = L(\tilde{G}^\rho) \otimes I_d \tag{72d}$$

where $L(W^\tau)$, $\tilde{V}$, $\tilde{\Sigma}$, and $L(\tilde{G}^\rho)$ are as defined in Equations (13a), (15), (16), and (14), respectively. This establishes that $B^\mathsf{T} B = M \otimes I_d$, where M is the matrix defined in (18b), and consequently that Problem 1 is equivalent to Problem 2.

B.2. Deriving Problem 3 from Problem 2

In this section, we show how to analytically eliminate the translations t appearing in Problem 2 to obtain the simplified form of Problem 3. We make use of the following lemma (cf. Boyd and Vandenberghe (2004, Appendix A.5.5) and Gallier (2010, Proposition 4.2)).

Lemma 4. *Given $A \in \mathrm{Sym}(p)$ and $b \in \mathbb{R}^p$, the function*

$$f(x) = x^\mathsf{T} A x + 2 b^\mathsf{T} x \tag{73}$$

attains a minimum if and only if $A \succeq 0$ and $(I_d - AA^\dagger)\, b = 0$, in which case

$$\min_{x \in \mathbb{R}^p} f(x) = -b^\mathsf{T} A^\dagger b \tag{74}$$

and

$$\operatorname*{argmin}_{x \in \mathbb{R}^p} f(x) = \left\{ -A^\dagger b + U \begin{pmatrix} 0 \\ z \end{pmatrix} \mid z \in \mathbb{R}^{p-r} \right\} \tag{75}$$

where $A = U \Lambda U^\mathsf{T}$ is an eigendecomposition of A and $r =$ rank(A).

Now $L(W^\tau) \otimes I_d \succeq 0$ as $L(W^\tau)$ is a (necessarily positive semidefinite) graph Laplacian, and $I_{dn} - (L(W^\tau) \otimes I_d)(L(W^\tau) \otimes I_d)^\dagger$ is the orthogonal projection operator onto $\ker\left((L(W^\tau) \otimes I_d)^\mathsf{T} \right) = \ker(L(W^\tau) \otimes I_d)$ (Meyer, 2000, equation (5.13.12)). Using the relation $\operatorname{vec}(AYC) = (C^\mathsf{T} \otimes A) \operatorname{vec}(Y)$ (Horn and Johnson, 1991, Lemma 4.3.1), we find that $y \in \ker(L(W^\tau) \otimes I_d)$ if and only if $y = \operatorname{vec}(Y)$ for some $Y \in \mathbb{R}^{d \times n}$ satisfying $YL(W^\tau) = 0$, or equivalently $L(W^\tau)\, Y^\mathsf{T} = 0$. As G is connected, then $\ker(L(W^\tau)) = \operatorname{span}\{\mathbf{1}_n\}$, and therefore we must have $Y^\mathsf{T} = \mathbf{1}_n c^\mathsf{T} \in \mathbb{R}^{n \times d}$ for some $c \in \mathbb{R}^d$. Altogether, this establishes that

$$\ker(L(W^\tau) \otimes I_d) = \left\{ \operatorname{vec}\left(c \mathbf{1}_n^\mathsf{T} \right) \mid c \in \mathbb{R}^d \right\} \tag{76}$$

Now let $b = (\tilde{V} \otimes I_d) \operatorname{vec}(R)$; we claim that $b \perp y$ for all $y = \operatorname{vec}(Y) \in \ker(L(W^\tau) \otimes I_d)$. To see this, observe that $b =$

$\mathrm{vec}(R\tilde{V}^\mathsf{T})$ and, therefore, $\langle b, y\rangle_2 = \langle R\tilde{V}^\mathsf{T}, Y\rangle_F = \mathrm{tr}(Y\tilde{V}R^\mathsf{T})$. However, $\mathbf{1}_n^\mathsf{T}\tilde{V} = 0$ because the sum down each column of $\tilde{V}$ is identically 0 by (15), and therefore $Y\tilde{V} = 0$ by (76). This establishes that $b \perp \ker(L(W^\tau)\otimes I_d)$ for *any* value of R.

Consequently, if we fix $R \in \mathrm{SO}(d)^n$ and consider performing the optimization in (19) over the decision variables t *only*, we can apply Lemma 4 to compute the optimal value of the objective and a minimizing value t^* of t as functions of R. This in turn enables us to analytically eliminate t from (19), producing the equivalent optimization problem (77),

$$p^*_{\mathrm{MLE}} = \min_{R\in\mathrm{SO}(d)} \mathrm{vec}(R)^\mathsf{T}\left(\left(L(\tilde{G}^\rho)+\tilde{\Sigma} - \tilde{V}^\mathsf{T} L(W^\tau)^\dagger \tilde{V}\right)\otimes I_d\right)\mathrm{vec}(R) \tag{77}$$

with a corresponding optimal estimate t^* given by

$$t^* = -\left(L(W^\tau)^\dagger\,\tilde{V}\otimes I_d\right)\mathrm{vec}(R^*) \tag{78}$$

Rewriting (77) and (78) in a more compact matricized form gives (20) and (21), respectively.

B.3. Simplifying the translational data matrix

In this section, we derive the simplified form of the translational data matrix $\tilde{Q}^\tau$ given in (24c) from that originally presented in (20b). To begin, recall from Appendix B.1 that

$$\begin{aligned}\tilde{Q}^\tau\otimes I_d &= \left(\tilde{\Sigma} - \tilde{V}^\mathsf{T}L(W^\tau)^\dagger\,\tilde{V}\right)\otimes I_d\\ &= B_2^\mathsf{T}B_2 - B_2^\mathsf{T}B_1\left(B_1^\mathsf{T}B_1\right)^\dagger B_1^\mathsf{T}B_2\\ &= B_2^\mathsf{T}\left(I_{dm} - B_1\left(B_1^\mathsf{T}B_1\right)^\dagger B_1^\mathsf{T}\right)B_2\end{aligned} \tag{79}$$

where B_1 and B_2 are the matrices defined in (69a) and (69b), respectively. Using Ω and $\tilde{T}$ as defined in (22) and (23), respectively, we may write B_1 and B_2 alternatively as

$$B_1 = \Omega^{\frac{1}{2}}A^\mathsf{T}\otimes I_d, \qquad B_2 = \Omega^{\frac{1}{2}}\tilde{T}\otimes I_d \tag{80}$$

where $A \triangleq A(\vec{G})$ is the incidence matrix of $\vec{G}$. Substituting (80) into (79), we derive as shown in (81),

$$\begin{aligned}\tilde{Q}^\tau\otimes I_d &= B_2^\mathsf{T}\left(I_{dm} - B_1(B_1^\mathsf{T}B_1)^\dagger B_1^\mathsf{T}\right)B_2\\ &= B_2^\mathsf{T}\left(I_{dm} - \left(\Omega^{\frac{1}{2}}A^\mathsf{T}\otimes I_d\right)\left(\left(\Omega^{\frac{1}{2}}A^\mathsf{T}\otimes I_d\right)^\mathsf{T}\left(\Omega^{\frac{1}{2}}A^\mathsf{T}\otimes I_d\right)\right)^\dagger\left(\Omega^{\frac{1}{2}}A^\mathsf{T}\otimes I_d\right)^\mathsf{T}\right)B_2\\ &= B_2^\mathsf{T}\left(I_{dm} - \left(\Omega^{\frac{1}{2}}A^\mathsf{T}\otimes I_d\right)\left(\left(A\Omega A^\mathsf{T}\right)^\dagger\otimes I_d\right)\left(\Omega^{\frac{1}{2}}A^\mathsf{T}\otimes I_d\right)^\mathsf{T}\right)B_2\\ &= B_2^\mathsf{T}\left(I_{dm} - \left(\Omega^{\frac{1}{2}}A^\mathsf{T}\left(A\Omega A^\mathsf{T}\right)^\dagger A\Omega^{\frac{1}{2}}\right)\otimes I_d\right)B_2\\ &= B_2^\mathsf{T}\left[\left(I_m - \Omega^{\frac{1}{2}}A^\mathsf{T}\left(A\Omega A^\mathsf{T}\right)^\dagger A\Omega^{\frac{1}{2}}\right)\otimes I_d\right]B_2\\ &= \left(\tilde{T}^\mathsf{T}\Omega^{\frac{1}{2}}\left(I_m - \Omega^{\frac{1}{2}}A^\mathsf{T}\left(A\Omega A^\mathsf{T}\right)^\dagger A\Omega^{\frac{1}{2}}\right)\Omega^{\frac{1}{2}}\tilde{T}\right)\otimes I_d\end{aligned} \tag{81}$$

or, equivalently,

$$\tilde{Q}^\tau = \tilde{T}^\mathsf{T}\Omega^{\frac{1}{2}}\underbrace{\left(I_m - \Omega^{\frac{1}{2}}A^\mathsf{T}\left(A\Omega A^\mathsf{T}\right)^\dagger A\Omega^{\frac{1}{2}}\right)}_{\Pi}\Omega^{\frac{1}{2}}\tilde{T} \tag{82}$$

Now let us develop the term Π appearing in (82):

$$\begin{aligned}\Pi &= I_m - \Omega^{\frac{1}{2}}A^\mathsf{T}\left(A\Omega A^\mathsf{T}\right)^\dagger A\Omega^{\frac{1}{2}}\\ &= I_m - \left(A\Omega^{\frac{1}{2}}\right)^\mathsf{T}\left(\left(A\Omega^{\frac{1}{2}}\right)\left(A\Omega^{\frac{1}{2}}\right)^\mathsf{T}\right)^\dagger\left(A\Omega^{\frac{1}{2}}\right)\\ &= I_m - \left(A\Omega^{\frac{1}{2}}\right)^\dagger\left(A\Omega^{\frac{1}{2}}\right)\end{aligned} \tag{83}$$

where we have used the fact that $X^\mathsf{T}(XX^\mathsf{T})^\dagger = X^\dagger$ for any matrix X in passing from line 2 to line 3 above. We may now recognize the final line of (83) as the matrix of the orthogonal projection operator $\pi : \mathbb{R}^m \to \ker(A\Omega^{\frac{1}{2}})$ onto the kernel of the weighted incidence matrix $A\Omega^{\frac{1}{2}}$ (Meyer, 2000, equation (5.13.12)). Equation (24c) thus follows from (82) and (83).

Finally, although Π is generically dense, we now show that it admits a computationally convenient decomposition in terms of sparse matrices and their inverses. By the Fundamental Theorem of Linear Algebra, $\ker(A\Omega^{\frac{1}{2}})^\perp = \mathrm{image}(\Omega^{\frac{1}{2}}A^\mathsf{T})$, and therefore every vector $v \in \mathbb{R}^m$ admits the orthogonal decomposition:

$$v = \pi(v) + c, \qquad \pi(v)\in\ker(A\Omega^{\frac{1}{2}}),\ c\in\mathrm{image}(\Omega^{\frac{1}{2}}A^\mathsf{T}) \tag{84}$$

Now $\mathrm{rank}(A) = n-1$ because A is the incidence matrix of the weakly connected directed graph $\vec{G}$; it follows that $\mathrm{image}(\Omega^{\frac{1}{2}}A^\mathsf{T}) = \mathrm{image}(\Omega^{\frac{1}{2}}\underline{A}^\mathsf{T})$, where $\underline{A}$ is the reduced incidence matrix of $\vec{G}$ formed by removing the final row of A. Furthermore, as c is the complement of $\pi(v)$ in the orthogonal decomposition (84), it is itself the orthogonal projection of v onto $\mathrm{image}(\Omega^{\frac{1}{2}}A^\mathsf{T}) = \mathrm{image}(\Omega^{\frac{1}{2}}\underline{A}^\mathsf{T})$, and is therefore the value of the product realizing the minimum norm in

$$\min_{w\in\mathbb{R}^{n-1}}\|v - \Omega^{\frac{1}{2}}\underline{A}^\mathsf{T}w\|_2 \tag{85}$$

Consequently, it follows from (84) and (85) that

$$\begin{aligned}\pi(v) &= v - \Omega^{\frac{1}{2}}\underline{A}^\mathsf{T}w^*\\ w^* &= \operatorname*{argmin}_{w\in\mathbb{R}^{n-1}}\|v - \Omega^{\frac{1}{2}}\underline{A}^\mathsf{T}w\|_2\end{aligned} \tag{86}$$

As $\underline{A}$ is full-rank, we can solve for the minimizer w^* in (86) via the normal equations, obtaining

$$w^* = \left(\underline{A}\Omega\underline{A}^\top\right)^{-1}\underline{A}\Omega^{\frac{1}{2}}v = L^{-\top}L^{-1}\underline{A}\Omega^{\frac{1}{2}}v \tag{87}$$

where $\underline{A}\Omega^{\frac{1}{2}} = LQ_1$ is a thin LQ decomposition of the weighted reduced incidence matrix $\underline{A}\Omega^{\frac{1}{2}}$. Substituting (87) into (86), we obtain

$$\begin{aligned}\pi(v) &= v - \Omega^{\frac{1}{2}}\underline{A}^\top L^{-\top}L^{-1}\underline{A}\Omega^{\frac{1}{2}}v \\ &= \left(I_m - \Omega^{\frac{1}{2}}\underline{A}^\top L^{-\top}L^{-1}\underline{A}\Omega^{\frac{1}{2}}\right)v\end{aligned} \tag{88}$$

As v was arbitrary, we conclude that the matrix in parentheses on the right-hand side of (88) is Π, which gives (39).

B.4. Deriving Problem 7 from Problem 6

This derivation is a straightforward application of the duality theory for semidefinite programs (Vandenberghe and Boyd, 1996). Letting $\lambda_{iuv} \in \mathbb{R}$ denote the (u, v)-element of the ith diagonal block of Λ for $i = 1, \dots, n$ and $1 \leq u \leq v \leq d$, we can rewrite Problem 6 in the primal standard form of Vandenberghe and Boyd (1996, equation (1)) as

$$\begin{aligned}-p^*_{\text{SDP}} = &\min_{\lambda \in \mathbb{R}^{\frac{d(d+1)}{2}n}} c^\top \lambda \\ &\text{s.t.} \quad F(\lambda) \succeq 0\end{aligned} \tag{89}$$

with the problem data in (89) given explicitly by

$$\begin{aligned}F(\lambda) &\triangleq F_0 + \sum_{i=1}^{n}\sum_{1\leq u\leq v\leq d}\lambda_{iuv}F_{iuv} \\ F_0 &= Q \\ F_{iuv} &= \begin{cases}-\operatorname{Diag}(e_i)\otimes E_{uu}, & u = v \\ -\operatorname{Diag}(e_i)\otimes(E_{uv}+E_{vu}), & u \neq v\end{cases} \\ c_{iuv} &= \begin{cases}-1, & u = v \\ 0, & u \neq v\end{cases}\end{aligned} \tag{90}$$

The standard dual problem for (89) is then (cf. Vandenberghe and Boyd, 1996, equations (1) and (27)):

$$\begin{aligned}-d^*_{\text{SDP}} = &\max_{Z \in \text{Sym}(dn)} -\operatorname{tr}(F_0 Z) \\ &\text{s.t.} \quad \operatorname{tr}(F_{iuv}Z) = c_{iuv} \\ &\qquad Z \succeq 0\end{aligned} \tag{91}$$

for all $i \in \{1, \dots, n\}$ and $1 \leq u \leq v \leq d$. Comparing (91) with (90) reveals that the equality constraints are satisfied precisely when the $(d \times d)$-block-diagonal of Z is composed of identity matrices, which gives the form of Problem 7. Furthermore, as $\tilde{Q} - \Lambda \succ 0$ for any $\Lambda = -sI_{dn}$ with $s > \|\tilde{Q}\|_2$ is strictly feasible for Problem 6 (hence, also (89)) and $I_{dn} \succ 0$ is strictly feasible for (91), Theorem 3.1 of Vandenberghe and Boyd (1996) implies that the optimal sets of (89) and (91) are non-empty, and that strong duality holds between them (so that the optimal value of Problem 7 is p^*_{SDP}, as claimed).

Appendix C. Proof of Proposition 2

In this section, we prove Proposition 2, following the general roadmap of the proof of a similar result for the special case of *angular synchronization* due to Bandeira et al. (2017). At a high level, our approach is based upon exploiting the Lagrangian duality between Problems 5 and 7 to identify a matrix C (constructed from an optimal solution R^* of Problem 5) with the property that $C \succeq 0$ and $\operatorname{rank}(C) = dn - d$ imply that $Z^* = R^{*\top}R^*$ is the unique optimal solution of Problem 7; we then show that these conditions can be assured by controlling the magnitude of the deviation $\Delta Q \triangleq \tilde{Q} - \underline{Q}$ of the observed data matrix $\tilde{Q}$ from its exact latent value $\underline{Q}$. Specifically, our proof proceeds according to the following chain of reasoning:

1. We begin by deriving the first-order necessary optimality conditions for the extrinsic formulation of Problem 5; these take the form $(\tilde{Q} - \Lambda^*)R^{*\top} = 0$, where $R^* \in \mathrm{O}(d)^n$ is a minimizer of Problem 5 and $\Lambda^* = \operatorname{SymBlockDiag}_d(\tilde{Q}R^{*\top}R^*)$ is a symmetric block-diagonal matrix of Lagrange multipliers corresponding to the orthogonality constraints in (26).
2. Exploiting the Lagrangian duality between Problems 5 and 7 together with non-degeneracy results for semidefinite programs (Alizadeh et al., 1997), we identify a matrix $C \triangleq \tilde{Q} - \Lambda^*$ with the property that $C \succeq 0$ and $\operatorname{rank}(C) = dn - d$ imply that $Z^* = R^{*\top}R^*$ is the unique optimal solution of Problem 7 (Theorem 7).
3. We next observe that for the idealized case in which the measurements $\tilde{x}_{ij}$ of the relative transforms are *noiseless*, the true latent rotations $\underline{R}$ comprise a minimizer of Problem 5 with corresponding certificate $\underline{C} = \underline{Q} \succeq L(\underline{G}^\rho)$. We then show (by means of similarity) that the spectrum of $L(\underline{G}^\rho)$ consists of d copies of the spectrum of the rotational weight graph W^ρ; in particular, $L(\underline{G}^\rho)$ has d eigenvalues equal to 0, and the remaining $d(n-1)$ are lower-bounded by the algebraic connectivity $\lambda_2(L(W^\rho)) > 0$ of W^ρ. Consequently, $\underline{C} \succeq 0$, and $\operatorname{rank}(\underline{C}) = dn - d$ as $\underline{R} \in \ker \underline{Q} = \ker \underline{C}$. It follows that Problem 7 is *always* exact in the absence of measurement noise (Theorem 10).
4. In the presence of noise, the minimizer $R^* \in \mathrm{O}(d)^n$ of Problem 5 will generally not coincide with the true latent rotations $\underline{R} \in \mathrm{SO}(d)^n$. Nevertheless, we can still derive an upper bound for the error in the estimate R^* in terms of the magnitude of the deviation $\Delta Q \triangleq \tilde{Q} - \underline{Q}$ of the data matrix $\tilde{Q}$ from the true latent value $\underline{Q}$ (Theorem 12).
5. The first-order necessary optimality condition for Problem 5 (point 1 above) can alternatively be read as $CR^{*\top} = 0$, which shows that d eigenvalues of C are always 0. Since in general the eigenvalues of a matrix X are continuous functions of X, it follows from points 3 and 4 above and the definition of C that the other $d(n-1)$ eigenvalues of C can be controlled into remaining non-negative by controlling the norm of ΔQ. In light

of point 2, this establishes the existence of a constant $\beta_1 > 0$ such that, if $\|\tilde{Q} - \underline{Q}\|_2 < \beta_1$, $Z^* = R^{*\mathsf{T}} R^*$ is the unique solution of Problem 7.

6. Finally, we observe that because $\mathrm{O}(d)$ is the disjoint union of two components separated by a distance $\sqrt{2}$ in the Frobenius norm, and the true latent rotations $\underline{R}_i$ all lie in the $+1$ component, it follows from point 4 that there exists a constant $\beta_2 > 0$ such that, if $\|\tilde{Q} - \underline{Q}\|_2 < \beta_2$, a minimizer $R^* \in \mathrm{O}(d)^n$ of Problem 5 must in fact lie in $\mathrm{SO}(d)^n$, and is therefore also a minimizer of Problem 4.
7. Taking $\beta \triangleq \min\{\beta_1, \beta_2\}$, Proposition 2 follows from points 5 and 6 and Theorem 1.

The remainder of this appendix is devoted to rigorously establishing each of claims 1–7 above.

C.1. Gauge symmetry and invariant metrics for Problems 4 and 5

A critical element of the proof of Proposition 2 is the derivation of an upper bound for the estimation error of a minimizer R^* of Problem 5 as a function of ΔQ (point 4). However, we have observed previously that Problems 1–5 always admit *multiple* (in fact, infinitely many) solutions for dimensions $d \geq 2$ due to gauge symmetry.[2] In consequence, it may not be immediately clear how we should quantify the estimation error of a *particular* point estimate R^* obtained as a minimizer of Problem 4 or Problem 5, since R^* is an arbitrary representative of an infinite set of *distinct* but *equivalent* minimizers that are related by gauge transforms. To address this complication, in this section we study the gauge symmetries of Problems 4 and 5, and then develop a pair of *gauge-invariant* metrics suitable for quantifying estimation error in these problems in a consistent, "symmetry-aware" manner.

Recall from Section 3.2 (cf. note 6) that solutions of Problem 1 are determined only up to a global gauge symmetry (corresponding to the diagonal left-action of $\mathrm{SE}(d)$ on $\mathrm{SE}(d)^n$). Similarly, it is straightforward to verify that if R^* is any minimizer of (24) (respectively (25)), then $G \bullet R^*$ also minimizes (24) (respectively (25)) for any choice of $G \in \mathrm{SO}(d)$ (respectively $G \in \mathrm{O}(d)$), where $\bullet$ is the diagonal left-action of $\mathrm{O}(d)$ on $\mathrm{O}(d)^n$:

$$G \bullet R \triangleq (GR_1, \ldots, GR_n) \tag{92}$$

It follows that the sets of minimizers of Problems 4 and 5 are partitioned into *orbits* of the form

$$\mathcal{S}(R) \triangleq \{G \bullet R \mid G \in \mathrm{SO}(d)\} \subset \mathrm{SO}(d)^n \tag{93a}$$

$$\mathcal{O}(R) \triangleq \{G \bullet R \mid G \in \mathrm{O}(d)\} \subset \mathrm{O}(d)^n \tag{93b}$$

each of which comprise a set of point estimates that are *equivalent* to R for the estimation problems (24) and (25), respectively. Consequently, when quantifying the dissimilarity between a pair of feasible points X, Y for Problem 4 or 5, we are interested in measuring the distance between the *orbits* determined by these two points, *not* the distance between the *specific representatives* X and Y themselves (which may, in fact, represent *equivalent* solutions, but differ in their assignments of coordinates by a global gauge symmetry). We therefore introduce the following *orbit distances*:

$$d_{\mathcal{S}}(X, Y) \triangleq \min_{G \in \mathrm{SO}(d)} \|X - G \bullet Y\|_F, \quad X, Y \in \mathrm{SO}(d)^n \tag{94a}$$

$$d_{\mathcal{O}}(X, Y) \triangleq \min_{G \in \mathrm{O}(d)} \|X - G \bullet Y\|_F, \quad X, Y \in \mathrm{O}(d)^n \tag{94b}$$

these functions report the Frobenius norm distance between the two closest representatives of the orbits (93) in $\mathrm{SO}(d)^n$ and $\mathrm{O}(d)^n$ determined by X and Y, respectively. Using the orthogonal invariance of the Frobenius norm, it is straightforward to verify that these functions satisfy

$$d_{\mathcal{S}}(X, Y) = d_{\mathcal{S}}(G_1 \bullet X, G_2 \bullet Y) \, \forall X, Y \in \mathrm{SO}(d)^n, \quad G_1, G_2 \in \mathrm{SO}(d) \tag{95a}$$

$$d_{\mathcal{O}}(X, Y) = d_{\mathcal{O}}(G_1 \bullet X, G_2 \bullet Y) \, \forall X, Y \in \mathrm{O}(d)^n, \quad G_1, G_2 \in \mathrm{O}(d) \tag{95b}$$

i.e. they define notions of dissimilarity between feasible points of Problems 4 and 5, respectively, that are *invariant* with respect to the action of the gauge symmetries for these problems, and so provide a consistent, gauge-invariant means of quantifying the estimation error.[3]

The following result enables us to compute these distances easily in practice.

2. Recall that $\mathrm{SO}(1) = \{+1\}$, so for $d = 1$ Problems 3 and 4 admit only the (trivial) solution $R^* = (1, \ldots, 1)$. Similarly, $\mathrm{O}(1) = \{\pm 1\}$, so for $d = 1$ Problem 5 admits pairs of solutions related by multiplication by -1.

3. We remark that although the formulation of the distance functions presented in (94) may at first appear somewhat *ad hoc*, one can justify this choice rigorously using the language of Riemannian geometry (Boothby, 2003; Kobayashi and Nomizu, 1996). As the Frobenius norm distance is orthogonally invariant, the diagonal left-actions (92) of $\mathrm{SO}(d)$ on $\mathrm{SO}(d)^n$ and $\mathrm{O}(d)$ on $\mathrm{O}(d)^n$ are *isometries*, and are trivially free and proper; consequently, the quotient spaces $\mathcal{M}_{\mathcal{S}} \triangleq \mathrm{SO}(d)^n / \mathrm{SO}(d)$ and $\mathcal{M}_{\mathcal{O}} \triangleq \mathrm{O}(d)^n / \mathrm{O}(d)$ obtained by *identifying* the elements of the orbits (93) are manifolds, and the projections $\pi_{\mathcal{S}} \colon \mathrm{SO}(d)^n \to \mathcal{M}_{\mathcal{S}}$ and $\pi_{\mathcal{O}} \colon \mathrm{O}(d)^n \to \mathcal{M}_{\mathcal{O}}$ are submersions. Furthermore, it is straightforward to verify that the restrictions of the derivative maps $d(\pi_{\mathcal{S}})_R : T_R(\mathrm{SO}(d)^n) \to T_{[R]}(\mathcal{M}_{\mathcal{S}})$ and $d(\pi_{\mathcal{O}})_R : T_R(\mathrm{O}(d)^n) \to T_{[R]}(\mathcal{M}_{\mathcal{O}})$ to the horizontal spaces $H_R(\mathrm{SO}(d)^n) \triangleq \ker(d(\pi_{\mathcal{S}})_R)^\perp$ and $H_R(\mathrm{O}(d)^n) \triangleq \ker(d(\pi_{\mathcal{O}})_R)^\perp$ are linear isometries onto $T_{[R]}(\mathcal{M}_{\mathcal{S}})$ and $T_{[R]}(\mathcal{M}_{\mathcal{O}})$, respectively, and therefore induce well-defined Riemannian metrics on the quotient spaces $\mathcal{M}_{\mathcal{S}}$ and $\mathcal{M}_{\mathcal{O}}$ from the Riemannian metrics on the total spaces $\mathrm{SO}(d)^n$ and $\mathrm{O}(d)^n$, with corresponding distance functions $d_{\mathcal{M}_{\mathcal{S}}}(\cdot,\cdot)$, $d_{\mathcal{M}_{\mathcal{O}}}(\cdot,\cdot)$, respectively. The functions $d_{\mathcal{S}}(\cdot,\cdot)$ and $d_{\mathcal{O}}(\cdot,\cdot)$ defined in (94) are then simply the functions that report the distances between the images of X and Y after projecting them to these quotient spaces: $d_{\mathcal{S}}(X, Y) = d_{\mathcal{M}_{\mathcal{S}}}(\pi_{\mathcal{S}}(X), \pi_{\mathcal{S}}(Y))$ and $d_{\mathcal{O}}(X, Y) = d_{\mathcal{M}_{\mathcal{O}}}(\pi_{\mathcal{O}}(X), \pi_{\mathcal{O}}(Y))$. Consequently, these are, in fact, the canonical distance functions for comparing points in $\mathrm{SO}(d)^n$ and $\mathrm{O}(d)^n$ while accounting for the gauge symmetry (92).

Theorem 5 (Computing the orbit distance)**.** *Given $X, Y \in \mathrm{O}(d)^n$, let*

$$XY^{\mathsf{T}} = U\Sigma V^{\mathsf{T}} \tag{96}$$

be a singular value decomposition of XY^{T} with $\Sigma = \mathrm{Diag}(\sigma_1, \dots, \sigma_d)$ and $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$. Then the orbit distance $d_{\mathcal{O}}(X, Y)$ between $\mathcal{O}(X)$ and $\mathcal{O}(Y)$ in $\mathrm{O}(d)^n$ is

$$d_{\mathcal{O}}(X, Y) = \sqrt{2dn - 2\|XY^{\mathsf{T}}\|_*} \tag{97}$$

and a minimizer $G^*_{\mathcal{O}} \in \mathrm{O}(d)$ realizing this optimal value in (94b) *is given by*

$$G^*_{\mathcal{O}} = UV^{\mathsf{T}}. \tag{98}$$

If, in addition, $X, Y \in \mathrm{SO}(d)$, then the orbit distance $d_{\mathcal{S}}(X, Y)$ between $\mathcal{S}(X)$ and $\mathcal{S}(Y)$ in $\mathrm{SO}(d)^n$ is given by

$$d_{\mathcal{S}}(X, Y) = \sqrt{2dn - 2\,\mathrm{tr}(\Xi\Sigma)} \tag{99}$$

where Ξ is the matrix

$$\Xi = \mathrm{Diag}\left(1, \dots, 1, \det(UV^{\mathsf{T}})\right) \in \mathbb{R}^{d\times d} \tag{100}$$

and a minimizer $G^*_{\mathcal{S}} \in \mathrm{SO}(d)$ realizing this optimal value in (94a) *is given by*

$$G^*_{\mathcal{S}} = U\Xi V^{\mathsf{T}} \tag{101}$$

Proof. Observe that

$$\begin{aligned}
\|X - G\bullet Y\|_F^2 &= \sum_{i=1}^n \|X_i - GY_i\|_F^2 \\
&= 2dn - 2\sum_{i=1}^n \langle X_i, GY_i\rangle_F \\
&= 2dn - 2\left\langle G, \sum_{i=1}^n X_i Y_i^{\mathsf{T}}\right\rangle_F \\
&= 2dn - 2\langle G, XY^{\mathsf{T}}\rangle_F
\end{aligned} \tag{102}$$

Consequently, a minimizer $G^*_{\mathcal{O}} \in \mathrm{O}(d)$ attaining the optimal value in (94b) is determined by

$$G^*_{\mathcal{O}} \in \underset{G\in\mathrm{O}(d)}{\mathrm{argmax}}\,\langle G, XY^{\mathsf{T}}\rangle_F \tag{103}$$

However, we may now recognize (103) as an instance of the orthogonal Procrustes problem, with maximizer $G^*_{\mathcal{O}}$ given by (98) (Hanson and Norris, 1981; Umeyama, 1991). Substituting (96) and (98) into (102) and simplifying the resulting expression using the orthogonal invariance of the Frobenius inner product then shows that the orbit distance $d_{\mathcal{O}}(X, Y)$ is

$$d_{\mathcal{O}}(X, Y) = \sqrt{2dn - 2\,\mathrm{tr}(\Sigma)} = \sqrt{2dn - 2\|XY^{\mathsf{T}}\|_*} \tag{104}$$

which is (97). If, in addition, $X, Y \in \mathrm{SO}(d)$, then (102) implies that a minimizer $G^*_{\mathcal{S}} \in \mathrm{SO}(d)$ attaining the optimal value in (94a) is determined by

$$G^*_{\mathcal{S}} \in \underset{G\in\mathrm{SO}(d)}{\mathrm{argmax}}\,\langle G, XY^{\mathsf{T}}\rangle_F \tag{105}$$

Equation (105) is an instance of the special orthogonal Procrustes problem, with a maximizer $G^*_{\mathcal{S}} \in \mathrm{SO}(d)$ given by (101) (Hanson and Norris, 1981; Umeyama, 1991). Substituting (96) and (101) into (102) and once again simplifying the resulting expression using the orthogonal invariance of the Frobenius inner product then shows that the orbit distance $d_{\mathcal{S}}(X, Y)$ is given by (99). □

C.2. A sufficient condition for exact recovery in Problem 5

In this section, we address points 1 and 2 in our roadmap, deriving sufficient conditions to ensure the recovery of a minimizer $R^* \in \mathrm{O}(d)^n$ of Problem 5 by means of solving the dual semidefinite relaxation Problem 7. Our approach is based upon exploiting the Lagrangian duality between Problem 5 and Problems 6 and 7 to construct a matrix C whose positive semidefiniteness serves as a *certificate of optimality* for $Z^* = R^{*\mathsf{T}}R^*$ as a solution of Problem 7.

We begin by deriving the first-order necessary optimality conditions for (25).

Lemma 6 (First-order necessary optimality conditions for Problem 5)**.** *If $R^* \in \mathrm{O}(d)^n$ is a minimizer of Problem 5, then there exists a matrix $\Lambda^* \in \mathrm{SBD}(d, n)$ such that*

$$(\tilde{Q} - \Lambda^*)R^{*\mathsf{T}} = 0 \tag{106}$$

Furthermore, Λ^* can be computed in closed-form according to

$$\Lambda^* = \mathrm{SymBlockDiag}_d\left(\tilde{Q}R^{*\mathsf{T}}R^*\right) \tag{107}$$

Proof. If we consider (25) as an *unconstrained* minimization of the objective $F(R) \triangleq \mathrm{tr}(\tilde{Q}R^{\mathsf{T}}R)$ over the Riemannian manifold $\mathrm{O}(d)^n$, then the first-order necessary optimality condition is simply that the Riemannian gradient $\mathrm{grad}\,F$ must vanish at the minimizer R^*:

$$\mathrm{grad}\,F(R^*) = 0 \tag{108}$$

Furthermore, if we consider $\mathrm{O}(d)^n$ as an embedded submanifold of $\mathbb{R}^{d\times dn}$, then this embedding induces a simple relation between the Riemannian gradient $\mathrm{grad}\,F$ of F viewed as a function restricted to $\mathrm{O}(d)^n \subset \mathbb{R}^{d\times dn}$ and ∇F, the gradient of F considered as a function on the ambient Euclidean space $\mathbb{R}^{d\times dn}$. Specifically, we have

$$\mathrm{grad}\,F(R) = \mathrm{Proj}_R\,\nabla F(R) \tag{109}$$

where $\mathrm{Proj}_R\colon T_R(\mathbb{R}^{d\times dn}) \to T_R(\mathrm{O}(d)^n)$ is the orthogonal projection operator onto the tangent space of $\mathrm{O}(d)^n$ at R (Absil et al., 2008, equation (3.37)); this latter is given explicitly by

$$\begin{aligned}
\mathrm{Proj}_R\colon T_R\left(\mathbb{R}^{d\times dn}\right) &\to T_R\left(\mathrm{O}(d)^n\right) \\
\mathrm{Proj}_R(X) &= X - R\,\mathrm{SymBlockDiag}_d(R^{\mathsf{T}}X)
\end{aligned} \tag{110}$$

Straightforward differentiation shows that the Euclidean gradient is $\nabla F(R) = 2R\tilde{Q}$, and consequently (108)–(110) imply that

$$\begin{aligned} 0 &= \operatorname{Proj}_{R^*} \nabla F(R^*) \\ &= 2R^*\tilde{Q} - 2R^* \operatorname{SymBlockDiag}_d\left(R^{*\mathsf{T}} R^* \tilde{Q}\right) \end{aligned} \tag{111}$$

Dividing both sides of (111) by two and taking the transpose produces (106), using the definition of Λ^* given in (107). □

Despite its simplicity, it turns out that Lemma 6 is actually already enough to enable the derivation of sufficient conditions to ensure the exactness of the semidefinite relaxation Problem 7 with respect to Problem 5. Comparing (106) with the *extrinsic* formulation (26) of Problem 5, we may recognize Λ^* as nothing more than a matrix of Lagrange multipliers corresponding to the orthogonality constraints $R_i^{\mathsf{T}} R_i = I_d$. Consequently, in the case that exactness holds between Problems 5 and 7 (i.e. that $Z^* = R^{*\mathsf{T}} R^*$ is a minimizer of Problem 7), strong duality obtains *a fortiori* between Problems 5 and 6, and therefore Λ^* also comprises an optimal solution for the Lagrangian relaxation (32) (cf. e.g. Boyd and Vandenberghe, 2004, Section 5.5.3). It follows that $\tilde{Q} - \Lambda^* \succeq 0$ (as Λ^* is *a fortiori* feasible for Problem 6 if it is optimal), and $(\tilde{Q} - \Lambda^*) Z^* = 0$ (from the definition of Z^* and (106)). However, observe now that these last two conditions are precisely the first-order necessary and sufficient conditions for Z^* to be an optimal solution of Problem 7 (cf. Vandenberghe and Boyd, 1996, equation (33)); furthermore, they provide a *closed-form expression* for a KKT certificate for Z^* (namely, $\tilde{Q} - \Lambda^*$) in terms of a minimizer R^* of Problem 5 using (107). The utility of this expression is that, although it was originally derived under the *assumption* of exactness, it can also be exploited to derive a *sufficient condition* for same, as shown in the next theorem.

Theorem 7 (A sufficient condition for exact recovery in Problem 5). *Let $R^* \in \mathrm{O}(d)^n$ be a minimizer of Problem 5 with corresponding Lagrange multiplier matrix $\Lambda^* \in$ $\mathrm{SBD}(d,n)$ as in Lemma 6, and define*

$$C \triangleq \tilde{Q} - \Lambda^* \tag{112}$$

If $C \succeq 0$, then $Z^* = R^{*\mathsf{T}} R^*$ is a minimizer of Problem 7. If, in addition, $\operatorname{rank}(C) = dn - d$, *then Z^* is the* unique *minimizer of Problem 7.*

Proof. Since $C = \tilde{Q} - \Lambda^* \succeq 0$ by hypothesis, and Equation (106) of Lemma 6 implies that

$$(\tilde{Q} - \Lambda^*) Z^* = (\tilde{Q} - \Lambda^*)\left(R^{*\mathsf{T}} R^*\right) = 0, \tag{113}$$

Λ^* and Z^* satisfy the necessary and sufficient conditions characterizing primal–dual pairs of optimal solutions for the strictly feasible primal–dual pair of semidefinite programs (32) and (33) (cf. Vandenberghe and Boyd, 1996, Appendix B.4 and Theorem 3.1). In other words, $C \succeq 0$ *certifies* the optimality of Z^* as a solution of Problem 7.

Now suppose further that $\operatorname{rank}(C) = dn - d$; we establish that Z^* is the *unique* solution of Problem 7 using nondegeneracy results from Alizadeh et al. (1997). Specifically, we observe that the equivalent formulations (91) and (89) of Problems 7 and 6 derived in Appendix B.4 match the forms of the primal and dual semidefinite programs given in equations (2) and (3) of Alizadeh et al. (1997), respectively. Consequently, Theorem 10 of Alizadeh et al. (1997) guarantees that Problem 7 has a unique solution provided that we can exhibit a dual non-degenerate solution of Problem 6. As we have already identified Λ^* as a solution of Problem 6 (via (113)), it suffices to show that Λ^* is dual non-degenerate. To that end, let

$$\tilde{Q} - \Lambda^* = \begin{pmatrix} U & V \end{pmatrix} \operatorname{Diag}(0, 0, 0, \sigma_{d+1}, \ldots, \sigma_{dn}) \begin{pmatrix} U & V \end{pmatrix}^{\mathsf{T}} \tag{114}$$

be an eigendecomposition of $\tilde{Q} - \Lambda^*$ as in equation (16) of Alizadeh et al. (1997) (with $\sigma_k > 0$ for $k \geq d+1$, $U \in \mathbb{R}^{dn \times d}$, and $V \in \mathbb{R}^{dn \times (dn-d)}$). Theorem 9 of Alizadeh et al. (1997) states that Λ^* is dual non-degenerate if and only if

$$\left\{ U^{\mathsf{T}} \Xi U \mid \Xi \in \mathrm{SBD}(d,n) \right\} = \mathrm{Sym}(d) \tag{115}$$

Now the matrix U appearing in (114) can be characterized as a matrix whose columns form an orthonormal basis for the d-dimensional null space of $\tilde{Q} - \Lambda^*$. However, Equation (106) shows that the columns of $R^{*\mathsf{T}}$ span this same subspace, and are pairwise orthogonal since $R^{*\mathsf{T}}$ is composed of orthogonal blocks. Consequently, without loss of generality we may take $U = \frac{1}{\sqrt{n}} R^{*\mathsf{T}}$ in (114). Now we can write the left-hand side of (115) more explicitly as

$$\begin{aligned} &\left\{ U^{\mathsf{T}} \Xi U \mid \Xi \in \mathrm{SBD}(d,n) \right\} \\ &= \left\{ \frac{1}{n} \sum_{i=1}^{n} R_i^* \Xi_i R_i^{*\mathsf{T}} \mid \Xi \in \mathrm{SBD}(d,n) \right\} \end{aligned} \tag{116}$$

and it is immediate from (116) that given any $S \in \mathrm{Sym}(d)$, we have $S = U^{\mathsf{T}} \Xi U$ for $\Xi = \operatorname{Diag}(n R_1^{*\mathsf{T}} S R_1^*, 0, \ldots, 0)$. This shows that (115) holds, so Λ^* is a dual non-degenerate solution of Problem 6, and we conclude that Z^* is indeed the *unique* minimizer of Problem 7, as claimed. □

In short, Theorem 7 enables us to reduce the question of the exactness of Problem 7 to the problem of verifying the positive semidefiniteness of a certain matrix C that can be constructed from an optimal solution of Problem 5. The remainder of this appendix is devoted to establishing conditions that are sufficient to guarantee that this latter (much simpler) criterion is satisfied.

C.3. The noiseless case

As our first application of Theorem 7, in this section we prove that the semidefinite relaxation Problem 7 is always exact in the (highly idealized) case that the measurements $\tilde{x}_{ij}$ in (10) are noiseless. In addition to providing a baseline sanity check for the feasibility of our overall strategy, our analysis of this idealized case will also turn out to admit a straightforward generalization that suffices to prove Proposition 2.

To that end, let $L(\underline{G}^\rho)$ and $\underline{Q}^\tau$ denote the rotational and translational data matrices of the form appearing in Problem 4 constructed using the true (latent) relative transforms $\underline{x}_{ij}$ appearing in (10). The following pair of lemmas characterize several important properties of these matrices.

Lemma 8 (Exact rotational connection Laplacians). *Let $L(\underline{G}^\rho)$ be the rotational connection Laplacian constructed using the true (latent) relative rotations $\underline{R}_{ij} \triangleq \underline{R}_i^{-1}\underline{R}_j$ in* (10)*, W^ρ the corresponding rotational weight graph, and $\underline{R} \in \mathrm{SO}(d)^n$ the matrix of true (latent) rotational states. Then the following hold:*

(i) $L(W^\rho)\otimes I_d = SL(\underline{G}^\rho)S^{-1}$ for $S = \mathrm{Diag}(\underline{R}_1, \ldots, \underline{R}_n) \in \mathbb{R}^{dn\times dn}$;
(ii) $\lambda_{d+1}(L(\underline{G}^\rho)) = \lambda_2(L(W^\rho))$;
(iii)$\ker(L(\underline{G}^\rho)) = \{\underline{R}^\mathsf{T} v \mid v \in \mathbb{R}^d\}$.

Proof. A direct computation using the definitions of S in claim (i) and of $L(\underline{G}^\rho)$ in (14) shows that the (i,j)-block of the product $SL(\underline{G}^\rho)S^{-1}$ is given by

$$\left(SL(\underline{G}^\rho)S^{-1}\right)_{ij} = \begin{cases} d_i^\rho I_d, & i = j \\ -\kappa_{ij} I_d, & \{i,j\} \in \mathcal{E} \\ 0_{d\times d}, & \{i,j\} \notin \mathcal{E} \end{cases} \tag{117}$$

which we recognize as the $(d \times d)$-block description of $L(W^\rho)\otimes I_d$; this proves claim (i). For claim (ii), we observe that $L(W^\rho)\otimes I_d$ and $L(\underline{G}^\rho)$ have the same spectrum (since claim (i) shows that they are similar), and the spectrum of $L(W^\rho)\otimes I_d$ consists of d copies of the spectrum of $L(W^\rho)$ (this follows from the fact that the spectra of $A \in \mathbb{R}^{d_1\times d_1}$, $B \in \mathbb{R}^{d_2\times d_2}$ and $A\otimes B$ are related by $\lambda(A\otimes B) = \{\lambda_i(A)\lambda_j(B) \mid i \in [d_1], j \in [d_2]\}$ (see e.g. Horn and Johnson, 1991, Theorem 4.2.12)). For claim (iii), another direct computation using definition (14) shows that $L(\underline{G}^\rho)\underline{R}^\mathsf{T} = 0$, and therefore that $\mathrm{image}(\underline{R}^\mathsf{T}) = \{\underline{R}^\mathsf{T} v \mid v \in \mathbb{R}^d\} \subseteq \ker(L(\underline{G}^\rho))$; furthermore, $\dim(\mathrm{image}(\underline{R}^\mathsf{T})) = d$ as $\mathrm{rank}(\underline{R}^\mathsf{T}) = d$ (as it has d orthonormal columns). On the other hand, claim (ii) shows that $\lambda_{d+1}(L(\underline{G}^\rho)) = \lambda_2(L(W^\rho)) > 0$ as G is connected, and therefore $\dim(\ker(L(\underline{G}^\rho))) \leq d$; consequently, $\mathrm{image}(R^\mathsf{T})$ is *all* of $\ker(L(\underline{G}^\rho))$. □

Lemma 9 (Orthogonal projections of exact measurements). *Let $\underline{T} \in \mathbb{R}^{m\times dn}$ denote the data matrix of the form* (23) *constructed using the true (latent) values of the translational measurements $\underline{t}_{ij}$ in* (10) *and $\underline{R} \in \mathrm{SO}(d)^n$ the matrix of true (latent) rotational states. Then $\Omega^{\frac{1}{2}}\underline{TR}^\mathsf{T} \in \ker \Pi$.*

Proof. It follows from (2) and the definitions of $\underline{t}_{ij}$ in (10) and $\underline{T}$ in (23) that the product $\underline{TR}^\mathsf{T} \in \mathbb{R}^{m\times d}$ is a $(1 \times d)$-block structured matrix with rows indexed by the edges $(i,j) \in \vec{\mathcal{E}}$ and whose (i,j)th row is given by

$$\left(\underline{TR}^\mathsf{T}\right)_{(i,j)} = -\underline{t}_{ij}^\mathsf{T}\underline{R}_i^\mathsf{T} = -\left(\underline{R}_i\underline{t}_{ij}\right)^\mathsf{T} = \underline{t}_i^\mathsf{T} - \underline{t}_j^\mathsf{T} \tag{118}$$

Now observe that the quantities $\underline{t}_i^\mathsf{T} - \underline{t}_j^\mathsf{T}$ associated with each edge $(i,j) \in \vec{\mathcal{E}}$ in the product $\underline{TR}^\mathsf{T}$ are realizable as differences of values $\underline{t}_i, \underline{t}_j$ assigned to the endpoints of (i,j), i.e. the columns of $\underline{TR}^\mathsf{T}$ are realizable as *potential differences* associated with the *potential function* assigning $\underline{t}_i^\mathsf{T}$ to vertex $i \in \mathcal{V}$ for all i (Biggs, 1997). Formally, we have from (118) that

$$\underline{TR}^\mathsf{T} = A(\vec{G})^\mathsf{T}\begin{pmatrix} -\underline{t}_1^\mathsf{T} \\ \vdots \\ -\underline{t}_n^\mathsf{T} \end{pmatrix} \tag{119}$$

so that the columns of $\underline{TR}^\mathsf{T}$ lie in $\mathrm{image}(A(\vec{G})^\mathsf{T})$. It follows that $\Omega^{\frac{1}{2}}\underline{TR}^\mathsf{T} \in \mathrm{image}(\Omega^{\frac{1}{2}}A(\vec{G})^\mathsf{T})$. However, $\mathrm{image}(\Omega^{\frac{1}{2}}A(\vec{G})^\mathsf{T}) \perp \ker(A(\vec{G})\Omega^{\frac{1}{2}})$ by the Fundamental Theorem of Linear Algebra, and therefore $\Omega^{\frac{1}{2}}\underline{TR}^\mathsf{T}$ lies in the kernel of the orthogonal projector Π, as claimed. □

With the aid of Lemmas 8 and 9, it is now straightforward to show that Problem 7 is always exact in the noiseless case:

Theorem 10 (Problem 7 is exact in the noiseless case). *Let $\underline{Q}$ be the data matrix of the form* (24b) *constructed using the true (latent) relative transforms $\underline{x}_{ij}$ in* (10)*. Then $Z^* = \underline{R}^\mathsf{T}\underline{R}$ is the unique solution of the instance of Problem 7 parameterized by $\underline{Q}$.*

Proof. As $L(\underline{G}^\rho) \succeq 0$ by Lemma 8(i) and $\underline{Q}^\tau \succeq 0$ (immediate from the definition (24c)), $\underline{Q} = L(\underline{G}^\rho) + \underline{Q}^\tau \succeq 0$ as well, and therefore the optimal value of Problem 5 satisfies $p_{\mathrm{O}}^* \geq 0$. Furthermore, $\underline{R}^\mathsf{T} \in \ker(L(\underline{G}^\rho)), \ker(\underline{Q}^\tau)$ by Lemmas 8(iii) and 9, respectively, so $\underline{R}^\mathsf{T} \in \ker(\underline{Q})$ as well. This implies that $\mathrm{tr}(\underline{Q}\underline{R}^\mathsf{T}\underline{R}) = 0$, and we conclude that $\underline{R}$ is an optimal solution of the noiseless version of Problem 5 (as it is a feasible point that attains the lower bound of 0 for the optimal value of Problem 5 p_{O}^*). This also implies $\mathrm{rank}(\underline{Q}) = dn - d$, since $\underline{Q} \succeq L(\underline{G}^\rho)$ (and $L(\underline{G}^\rho)$ has $dn - d$ positive eigenvalues by Lemma 8(i)) and $\mathrm{image}(\underline{R}^\mathsf{T}) \subseteq \ker(\underline{Q})$ with $\dim(\mathrm{image}(\underline{R}^\mathsf{T})) = d$. Finally, a straightforward computation using Equations (107) and (112) shows that the candidate certificate matrix corresponding to the optimal solution $\underline{R}$ is $\underline{C} = \underline{Q}$. The claim then follows from an application of Theorem 7. □

In addition to providing a useful sanity check on the feasibility of our overall strategy by showing that it will at least succeed under ideal conditions, the proof of Theorem 10 also points the way towards a proof of the more general Proposition 2, as we now describe. Observe that in the noiseless case, the certificate matrix $C = \underline{Q} = L(\underline{G}^\rho) + \underline{Q}^\tau$

corresponding to the optimal solution $\underline{R}$ has a spectrum consisting of d copies of 0 and $dn-d$ strictly positive eigenvalues that are lower-bounded by $\lambda_2(L(W^\rho)) > 0$. Now, in the more general (noisy) case, both the data matrix $\tilde{Q}$ *and* the minimizer R^* of Problem 5 will vary as a function of the noise added to the measurements $\tilde{x}_{ij}$ in (10), and in consequence so will the matrix C. However, the first-order condition (106) appearing in Lemma 6 can alternatively be read as $CR^{*\mathsf{T}} = 0$, which guarantees that C always has at least d eigenvalues fixed to 0; furthermore, in general the eigenvalues of a matrix X are continuous functions of X, and Equations (107) and (112) show that C is a continuous function of $\tilde{Q}$ and R^*. Consequently, if we can bound the magnitude of the estimation error $d_{\mathcal{O}}(\underline{R}, R^*)$ for a minimizer R^* of Problem 5 as a function of the magnitude of the noise $\Delta Q = \tilde{Q} - \underline{Q}$ corrupting the data matrix $\tilde{Q}$, then by controlling ΔQ we can in turn ensure (via continuity) that the eigenvalues of the matrix C constructed at the minimizer R^* remain nonnegative, and hence Problem 7 will remain exact.

C.4. An upper bound for the estimation error in Problem 5

In this section, we derive an upper bound on the estimation error $d_{\mathcal{O}}(\underline{R}, R^*)$ of a minimizer R^* of Problem 5 as a function of the noise $\Delta Q \triangleq \tilde{Q} - \underline{Q}$ corrupting the data matrix $\tilde{Q}$. To simplify the derivation, in the following we assume (without loss of generality) that R^* is an element of its orbit (93b) attaining the orbit distance $d_{\mathcal{O}}(\underline{R}, R^*)$ defined in (94b).

To begin, the optimality of R^* implies that

$$\begin{aligned} \operatorname{tr}(\tilde{Q}\underline{R}^\mathsf{T}\underline{R}) &= \operatorname{tr}\left(\Delta Q\underline{R}^\mathsf{T}\underline{R}\right) + \operatorname{tr}\left(\underline{Q}\underline{R}^\mathsf{T}\underline{R}\right) \\ &\geq \operatorname{tr}\left(\Delta Q R^{*\mathsf{T}}R^*\right) + \operatorname{tr}\left(\underline{Q}R^{*\mathsf{T}}R^*\right) \\ &= \operatorname{tr}\left(\tilde{Q}R^{*\mathsf{T}}R^*\right) \end{aligned} \tag{120}$$

Now $\operatorname{tr}(\underline{Q}\underline{R}^\mathsf{T}\underline{R}) = 0$ because we showed in Appendix C.3 that $\operatorname{image}(\underline{R}^\mathsf{T}) = \ker(\underline{Q})$, and the identity $\operatorname{tr}(\Delta Q\underline{R}^\mathsf{T}\underline{R}) = \operatorname{vec}(\underline{R})^\mathsf{T}(\Delta Q \otimes I_d)\operatorname{vec}(\underline{R})$ together with the submultiplicativity of the spectral norm shows that

$$\begin{aligned} |\operatorname{tr}(\Delta Q\underline{R}^\mathsf{T}\underline{R})| &\leq \|\Delta Q \otimes I_d\|_2 \left\|\operatorname{vec}(\underline{R})\right\|_2^2 = \|\Delta Q\|_2 \left\|\underline{R}\right\|_F^2 \\ &= dn\|\Delta Q\|_2 \end{aligned} \tag{121}$$

(and similarly for $|\operatorname{tr}(\Delta QR^{*\mathsf{T}}R^*)|$); consequently, (120) in turn implies

$$2dn\|\Delta Q\|_2 \geq \operatorname{tr}\left(\underline{Q}R^{*\mathsf{T}}R^*\right) \tag{122}$$

We now lower-bound the right-hand side of (122) as a function of the estimation error $d_{\mathcal{O}}(\underline{R}, R^*)$, thereby enabling us to upper-bound this error by controlling $\|\Delta Q\|_2$. To do so, we make use of the following result.

Lemma 11. *Fix $R \in \mathrm{O}(d)^n \subset \mathbb{R}^{d\times dn}$, and let $M = \{WR \mid W \in \mathbb{R}^{d\times d}\} \subset \mathbb{R}^{d\times dn}$ denote the subspace of matrices whose rows are contained in* $\operatorname{image}(R^\mathsf{T})$. *Then*

$$\begin{aligned} \operatorname{Proj}_V &: \mathbb{R}^{dn} \to \operatorname{image}(R^\mathsf{T}) \\ \operatorname{Proj}_V(x) &= \frac{1}{n}R^\mathsf{T}Rx \end{aligned} \tag{123}$$

is the orthogonal projection operator onto $\operatorname{image}(R^\mathsf{T})$ *with respect to the usual ℓ_2 inner product on $\mathbb{R}^{dn}$, and the mapping*

$$\begin{aligned} \operatorname{Proj}_M &: \mathbb{R}^{d\times dn} \to M \\ \operatorname{Proj}_M(X) &= \frac{1}{n}XR^\mathsf{T}R \end{aligned} \tag{124}$$

that applies Proj_V *to each row of X is the orthogonal projection operator onto M with respect to the Frobenius inner product on $\mathbb{R}^{d\times dn}$.*

Proof. If $x \in \operatorname{image}(R^\mathsf{T})$, then $x = R^\mathsf{T}v$ for some $v \in \mathbb{R}^d$, and

$$\operatorname{Proj}_V(x) = \frac{1}{n}R^\mathsf{T}R(R^\mathsf{T}v) = R^\mathsf{T}v = x \tag{125}$$

because $RR^\mathsf{T} = nI_d$ as $R \in \mathrm{O}(d)^n$ by hypothesis; this shows that Proj_V is a projection onto $\operatorname{image}(R^\mathsf{T})$. To show that Proj_V is *orthogonal* projection with respect to the ℓ_2 inner product on $\mathbb{R}^{dn}$, it suffices to show that $\operatorname{image}(\operatorname{Proj}_V) \perp \ker(\operatorname{Proj}_V)$. To that end, let $x, y \in \mathbb{R}^{dn}$, and observe that

$$\begin{aligned} &\langle \operatorname{Proj}_V(x), y - \operatorname{Proj}_V(y) \rangle \\ &= \left\langle \frac{1}{n}R^\mathsf{T}Rx, y - \frac{1}{n}R^\mathsf{T}Ry \right\rangle_2 \\ &= \frac{1}{n}\left\langle R^\mathsf{T}Rx, y\right\rangle_2 - \frac{1}{n^2}\left\langle R^\mathsf{T}Rx, R^\mathsf{T}Ry\right\rangle_2 \\ &= \frac{1}{n}x^\mathsf{T}R^\mathsf{T}Ry - \frac{1}{n^2}x^\mathsf{T}R^\mathsf{T}RR^\mathsf{T}Ry \\ &= 0 \end{aligned} \tag{126}$$

Next, let $X \in \mathbb{R}^{d\times dn}$ and observe that $\operatorname{Proj}_M(X) = \frac{1}{n}XR^\mathsf{T}R$ is the matrix obtained by applying the projection Proj_V to each row of X; this immediately implies that Proj_M is itself a projection onto M. Furthermore, given $X, Y \in \mathbb{R}^{d\times dn}$, we observe that

$$\begin{aligned} &\left\langle \operatorname{Proj}_M(X), Y - \operatorname{Proj}_M(Y) \right\rangle_F \\ &= \left\langle \operatorname{Proj}_M(X)^\mathsf{T}, Y^\mathsf{T} - \operatorname{Proj}_M(Y)^\mathsf{T} \right\rangle_F \\ &= \left\langle \operatorname{vec}\left(\operatorname{Proj}_M(X)^\mathsf{T}\right), \operatorname{vec}\left(Y^\mathsf{T} - \operatorname{Proj}_M(Y)^\mathsf{T}\right)\right\rangle_2 \\ &= 0 \end{aligned} \tag{127}$$

as we have already established that Proj_M acts row-wise by Proj_V, which is orthogonal projection with respect to the ℓ_2 inner product. □

As $\ker(\underline{Q}) = \operatorname{image}(\underline{R}^\mathsf{T})$ and $\dim(\operatorname{image}(\underline{R}^\mathsf{T})) = d$, it follows from Lemma 11 that

$$\operatorname{tr}\left(\underline{Q}R^{*\mathsf{T}}R^*\right) \geq \lambda_{d+1}(\underline{Q})\,\|P\|_F^2 \tag{128}$$

where

$$R^* = K + P$$
$$K = \operatorname{Proj}_M(R^*) = \frac{1}{n} R^* \underline{R}^{\mathsf{T}} \underline{R} \tag{129}$$
$$P = R^* - \operatorname{Proj}_M(R^*) = R^* - \frac{1}{n} R^* \underline{R}^{\mathsf{T}} \underline{R}$$

is an orthogonal decomposition of R^* with respect to the Frobenius inner product on $\mathbb{R}^{d\times dn}$, and the rows of P are contained in image$(\underline{R}^{\mathsf{T}})^{\perp} =$ ker$(\underline{Q})^{\perp}$. Using (129), we compute

$$\|K\|_F^2 = \frac{1}{n^2} \operatorname{tr}\left(\underline{R}^{\mathsf{T}} \underline{R} R^{*\mathsf{T}} R^* \underline{R}^{\mathsf{T}} \underline{R}\right) = \frac{1}{n} \operatorname{tr}\left(\underline{R} R^{*\mathsf{T}} R^* \underline{R}^{\mathsf{T}}\right) = \frac{1}{n} \left\| \underline{R} R^{*\mathsf{T}} \right\|_F^2 \tag{130}$$

where we have used the cyclic property of the trace and the fact that $\underline{R}\underline{R}^{\mathsf{T}} = nI_d$. As (129) is an orthogonal decomposition, it follows that

$$\|P\|_F^2 = \left\|R^*\right\|_F^2 - \|K\|_F^2 = dn - \frac{1}{n}\left\|\underline{R}R^{*\mathsf{T}}\right\|_F^2 \tag{131}$$

We may therefore lower-bound $\|P\|_F^2$ by upper-bounding $\|\underline{R}R^{*\mathsf{T}}\|_F^2$ as functions of $d_{\mathcal{O}}(\underline{R}, R^*)$. To that end, recall that R^* is by hypothesis a representative of its orbit (93b) that attains the orbit distance (94b); Theorem 5 then implies that

$$d_{\mathcal{O}}(\underline{R}, R^*)^2 = \|\underline{R} - R^*\|_F^2 = 2dn - 2\sum_{i=1}^{d} \sigma_i \tag{132}$$

where

$$\underline{R}R^{*\mathsf{T}} = U \operatorname{Diag}(\sigma_1, \ldots, \sigma_d)\, V^{\mathsf{T}} \tag{133}$$

is a singular value decomposition of $\underline{R}R^{*\mathsf{T}}$. It follows from (133) and the orthogonal invariance of the Frobenius inner product that

$$\left\|\underline{R}R^{*\mathsf{T}}\right\|_F^2 = \|\operatorname{Diag}(\sigma_1, \ldots, \sigma_d)\|_F^2 = \sum_{i=1}^{d} \sigma_i^2 \tag{134}$$

and, therefore, (132) and (134) imply that we may obtain an upper bound ϵ^2 for $\|\underline{R}R^{*\mathsf{T}}\|_F^2$ in terms of $\delta^2 = d_{\mathcal{O}}(\underline{R}, R^*)^2$ as the optimal value of

$$\begin{aligned} \epsilon^2 = \max_{\sigma_i \geq 0} \sum_{i=1}^{d} \sigma_i^2 \\ \text{s.t. } 2dn - 2\sum_{i=1}^{d} \sigma_i = \delta^2 \end{aligned} \tag{135}$$

The first-order necessary optimality condition for (135) is

$$2\sigma_i = -2\lambda \tag{136}$$

for all $i \in [d]$, where $\lambda \in \mathbb{R}$ is a Lagrange multiplier, and therefore $\sigma_1 = \cdots = \sigma_d = \sigma$ for some $\sigma \in \mathbb{R}$. Solving the constraint in (135) for σ shows that

$$\sigma = n - \frac{\delta^2}{2d} \tag{137}$$

and therefore the optimal value of the objective in (135) is

$$\epsilon^2 = d\left(n - \frac{\delta^2}{2d}\right)^2. \tag{138}$$

Recalling the original definitions of ϵ^2 and δ^2, we conclude from (138) and (131) that

$$\begin{aligned} \|P\|_F^2 &\geq dn - \frac{d}{n}\left(n - \frac{d_{\mathcal{O}}(\underline{R}, R^*)^2}{2d}\right)^2 \\ &= d_{\mathcal{O}}(\underline{R}, R^*)^2 - \frac{d_{\mathcal{O}}(\underline{R}, R^*)^4}{4dn} \end{aligned} \tag{139}$$

Applying the inequality $d_{\mathcal{O}}(\underline{R}, R^*)^2 \leq 2dn$ (which follows immediately from the non-negativity of the nuclear norm in (97)), we may in turn lower-bound the right-hand side of (139) as

$$\begin{aligned} & d_{\mathcal{O}}(\underline{R}, R^*)^2 - \frac{d_{\mathcal{O}}(\underline{R}, R^*)^4}{4dn} \\ &= \left(1 - \frac{d_{\mathcal{O}}(\underline{R}, R^*)^2}{4dn}\right) d_{\mathcal{O}}(\underline{R}, R^*)^2 \\ &\geq \frac{1}{2} d_{\mathcal{O}}(\underline{R}, R^*)^2 \end{aligned} \tag{140}$$

Finally, combining inequalities (122), (128), (139), and (140), we obtain the following theorem.

Theorem 12 (An upper bound for the estimation error in Problem 5)**.** *Let $\underline{Q}$ be the data matrix of the form* (24b) *constructed using the true (latent) relative transforms $\underline{x}_{ij} = (\underline{t}_{ij}, \underline{R}_{ij})$ in* (10)*, $\underline{R} \in \mathrm{SO}(d)^n$ the matrix composed of the true (latent) rotational states, and $R^* \in \mathrm{O}(d)^n$ an estimate of $\underline{R}$ obtained as a minimizer of Problem 5. Then the estimation error $d_{\mathcal{O}}(\underline{R}, R^*)$ admits the following upper bound:*

$$\sqrt{\frac{4dn\|\tilde{Q} - \underline{Q}\|_2}{\lambda_{d+1}(\underline{Q})}} \geq d_{\mathcal{O}}(\underline{R}, R^*) \tag{141}$$

C.5. Finishing the proof

Finally, we complete the proof of Proposition 2 with the aid of Theorems 7 and 12.

Proof of Proposition 2. Let $\underline{R} \in \mathrm{SO}(d)^n$ be the matrix of true (latent) rotations, $R^* \in \mathrm{O}(d)^n$ an estimate of $\underline{R}$ obtained as a minimizer of Problem 5, and assume without loss of generality that R^* is an element of its orbit (93b) attaining the orbit distance $d_{\mathcal{O}}(\underline{R}, R^*)$ defined in (94b). Set $\Delta Q \triangleq \tilde{Q} - \underline{Q}$ and $\Delta R \triangleq R^* - \underline{R}$, and consider the decomposition

$$
\begin{aligned}
C &= \tilde{Q} - \mathrm{SymBlockDiag}_d\left(\tilde{Q}R^{*\mathsf{T}}R^*\right) \\
&= \left(\underline{Q} + \Delta Q\right) - \mathrm{SymBlockDiag}_d \\
&\quad \left(\left(\underline{Q} + \Delta Q\right)(\underline{R} + \Delta R)^{\mathsf{T}}(\underline{R} + \Delta R)\right) \\
&= \underline{Q} + \Delta Q - \mathrm{SymBlockDiag}_d \\
&\quad \begin{pmatrix} \underline{Q}\underline{R}^{\mathsf{T}}\underline{R} + \Delta Q\underline{R}^{\mathsf{T}}\underline{R} + \underline{Q}\underline{R}^{\mathsf{T}}\Delta R + \underline{Q}\Delta R^{\mathsf{T}}\underline{R} \\ + \Delta Q\underline{R}^{\mathsf{T}}\Delta R + \Delta Q\Delta R^{\mathsf{T}}\underline{R} + \underline{Q}\Delta R^{\mathsf{T}}\Delta R + \Delta Q\Delta R^{\mathsf{T}}\Delta R \end{pmatrix} \\
&= \underline{Q} + \underbrace{\Delta Q - \mathrm{SymBlockDiag}_d \begin{pmatrix} \Delta Q\underline{R}^{\mathsf{T}}\underline{R} + \underline{Q}\Delta R^{\mathsf{T}}\underline{R} + \Delta Q\underline{R}^{\mathsf{T}}\Delta R \\ + \Delta Q\Delta R^{\mathsf{T}}\underline{R} + \underline{Q}\Delta R^{\mathsf{T}}\Delta R + \Delta Q\Delta R^{\mathsf{T}}\Delta R \end{pmatrix}}_{\Delta C}
\end{aligned}
\tag{142}
$$

of the certificate matrix C defined in (112) and (107), where we have used the fact that $\mathrm{image}(\underline{R}^{\mathsf{T}}) = \ker(\underline{Q})$ in passing from lines 2 to 3 above (cf. Lemmas 8 and 9). Observe that the term labeled ΔC in (142) depends continuously upon ΔQ and ΔR, with $\Delta C = 0$ for $(\Delta Q, \Delta R) = (0, 0)$; furthermore, $\Delta Q \to 0$ implies $\Delta R \to 0$ by Theorem 12. It therefore follows from continuity that there exists some $\beta_1 > 0$ such that $\|\Delta C\|_2 < \lambda_{d+1}(\underline{Q})$ for all $\|\Delta Q\|_2 < \beta_1$. Moreover, if $\|\Delta C\|_2 < \lambda_{d+1}(\underline{Q})$, it follows from (142) that

$$
\lambda_i(C) \geq \lambda_i(\underline{Q}) - \|\Delta C\|_2 > \lambda_i(\underline{Q}) - \lambda_{d+1}(\underline{Q}) \tag{143}
$$

and, therefore, $\lambda_i(C) > 0$ for $i \geq d + 1$; i.e. C has at least $dn - d$ strictly positive eigenvalues. Furthermore, Lemma 6 shows that $CR^{*\mathsf{T}} = 0$, which implies that $\ker(C) \supseteq \mathrm{image}(R^{*\mathsf{T}})$; as $\dim(\mathrm{image}(R^{*\mathsf{T}})) = d$, this in turn implies that C has at least d eigenvalues equal to 0. As this exhausts the dn eigenvalues of C, we conclude that $C \succeq 0$ and $\mathrm{rank}(C) = dn - d$, and consequently Theorem 7 guarantees that $Z^* = R^{*\mathsf{T}}R^*$ is the unique minimizer of Problem 7.

Now suppose further that $\|\Delta Q\|_2 < \beta_2$ with $\beta_2 \triangleq \lambda_{d+1}(\underline{Q})/2dn$. Then Theorem 12 implies $d_{\mathcal{O}}(\underline{R}, R^*) = \|\underline{R} - R^*\|_F < \sqrt{2}$, and therefore in particular that $\|\underline{R}_i - R_i^*\|_F < \sqrt{2}$ for all $i \in [n]$. However, the $+1$ and -1 components of $\mathrm{O}(d)$ are separated by a distance $\sqrt{2}$ under the Frobenius norm, so $\underline{R}_i \in \mathrm{SO}(d)$ and $\|\underline{R}_i - R_i^*\|_F < \sqrt{2}$ for all $i \in [n]$ together imply that $R^* \in \mathrm{SO}(d)^n$, and therefore that R^* is in fact an optimal solution of Problem 4 as well.

Proposition 2 then follows from the preceding paragraphs by taking $\beta \triangleq \min\{\beta_1, \beta_2\} > 0$. □