The Riemannian Elevator for certifiable distance-based localization

Trevor Halsted[1], Mac Schwager[2]

Abstract—We introduce a novel and exact reformulation of the distance-based localization problem, which involves an optimization over edge directions—unit vectors associated with the set of distance measurements. Our approach for solving this problem, which we call the Riemannian Elevator, recovers a solution to the distance-based localization problem by first solving a higher dimensional relaxation of the original problem. Solving this relaxation yields a non-trivial lower bound (the "certificate") as well as a useful initialization for the original problem. We subsequently solve on the oblique manifold (the manifold of unit vectors). The certificate provides a numerical bound on the worst-case suboptimality of this solution, which is particularly useful when measurements are noisy. We report empirical studies exhibiting the performance of our algorithm compared to prior methods across a range of graph sizes and connectivities.

Index Terms—Sensor network localization, range measurements, convex relaxations, certifiable algorithms

I. INTRODUCTION

DISTANCE-based localization—the task of determining a set of positions that best agree with measured distances between pairs of positions—arises in a variety of applications. In robotics, deploying mobile robots to GPS-denied environments necessitates the use of alternative sensors for localization, such as by acquiring robot-to-robot distances using ultra-wideband radios [1] or acoustic transponders [2]. Similar challenges arise in other types of wireless sensor networks, such as in environmental monitoring or inventory tracking [3]. The molecular conformation problem in structural biology involves determining the shape of a molecule from inter-atomic distances measured by nuclear magnetic resonance [4]. Phase retrieval is also a distance-based localization problem: in this case, we model components of an unknown signal as points that we wish to reconstruct based on intensity measurements [5], [6]. In natural language processing, finding word embeddings involves mapping word similarity to semantic space [7]. In each of these settings, we seek the most probable embedding of points in Euclidean space given measurements of the proximity between the points. We generally do not assume the availability of distance measurements between all pairs of points, and the distance measurements that we do have are often noisy.

[1] Mechanical Engineering Department, Stanford University
halsted@stanford.edu
[2] Department of Aeronautics & Astronautics, Stanford University
schwager@stanford.edu

Despite the ubiquity of distance-based localization across several domains, it remains a challenging computational problem. The standard formulation of the problem, known as the stress function, is neither convex nor globally differentiable. Alternative formulations of the problem achieve global differentiability at the expense of inexactly approximating the original problem. Existing methods reliably solve the fully-connected version of the problem, in which we have measurements between all pairs of points. However, performance of these methods typically decays given sparser measurements. It is known that the graph embedding problem, in which the distance measurements are treated as constraints, is strongly NP-hard [8], [9].

Exact embedding (an objective value of zero) is generally not attainable when the measurements between points are noisy; in this case, we cannot discern the absolute quality of a solution from its objective value alone. Therefore, we consider two main objectives in distance-based localization:

1) Recover a graph embedding that minimizes error between estimated and measured distances.
2) When the measured distances are not exactly feasible, generate a non-trivial lower bound for objective value of the solution.

In this paper, we introduce a novel reformulation of the distance-based localization problem that incorporates unit vectors associated with each measured distance as additional variables. Our approach of optimizing this alternative form of the distance-based localization problem addresses the two objectives outlined above. First, initializing the solution from a higher-dimensional relaxation helps to avoid spurious local minima, enabling our method to recover better solutions than alternative approaches. Second, we compute a numerical lower bound (a "certificate") that provides provable guarantees about the quality of a given solution.

A. Related work

Distance-based localization methods

Prior work in distance-based localization encompasses a variety of applications, but many methods originate from the topic of multidimensional scaling (MDS). MDS is the problem of finding a low-dimensional embedding that best represents samples from a higher dimensional space [12]. Kruskal proposed the *stress* function as a way to assess goodness of fit [13]. Minimizing the stress corresponds to generating an undistorted visualization of a set of samples. While early uses of the stress function focused on nonmetric dissimilarity measurements, *e.g.* in psychometrics, these techniques also
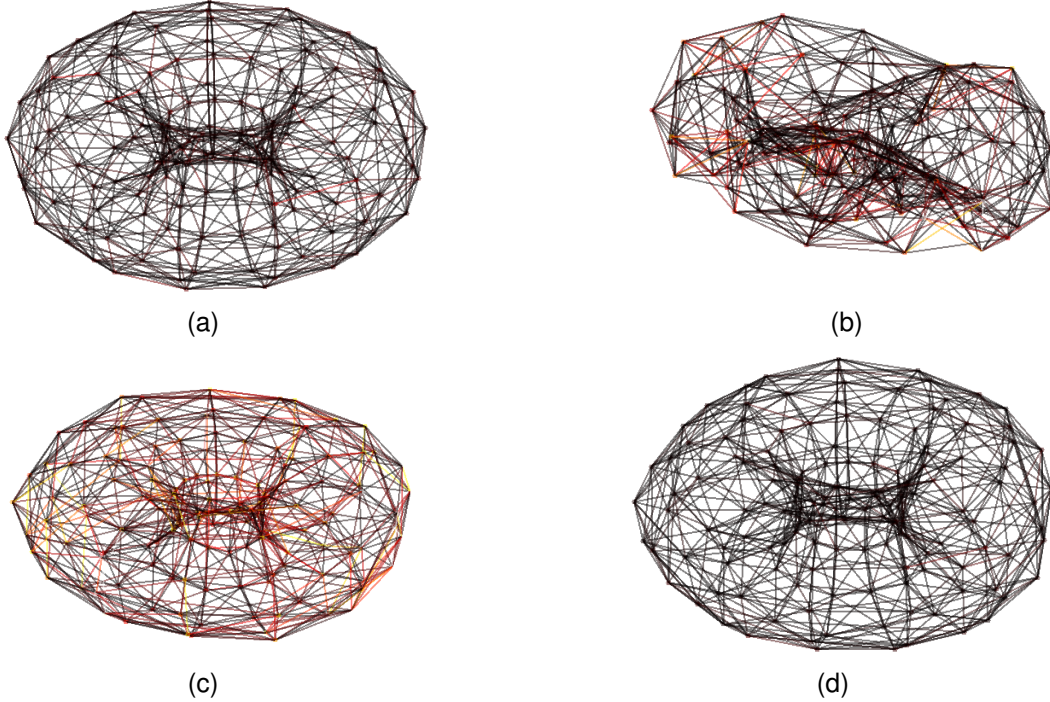
Fig. 1: A visualization of distance-based localization methods on `rigid-torus`, shown in (a), for which noisy measurements are generated from the ground truth configuration. The SMACOF method is an example of a non-global method that recovers a local minimum of the stress function, *e.g.* in (b). Approximate methods like the trace-regularized semidefinite relaxation [10], [11] find global solutions at the expense of modifying the objective function (notice the flatness of the solution in (c)). Our Riemannian Elevator method recovers a good initialization with high probability, enabling convergence to the global optimum as shown in (d). The Riemannian Elevator also generates a numerical bound on the suboptimality of the solution. Colors indicate error (stress) on respective edges (black indicates low error).

apply to physical distances. In this paper, we refer to the general weighted version of the stress function as the maximum likelihood estimate (MLE) problem, described further in Section II.

As a non-convex function, the stress function has many local minima. Consequently, Kruskal's proposed method of minimizing the stress function using gradient descent is highly reliant on a good initialization (perhaps using prior knowledge) [13]. A more widely-used approach is stress majorization, also known as SMACOF, proposed in [14]. SMACOF iteratively minimizes a majorizing function [15] of the stress function to achieve better results than gradient descent, though it too is sensitive to initialization, as demonstrated in Figure 1b. In practice, performance improves by choosing the best solution from multiple trials with different random initializations [16]. In contrast, our method initializes the solution from a projection of a global optimum to a relaxed problem. This initialization requires no prior knowledge or application-specific heuristics, and it substantially improves the probability of avoiding suboptimal local minima.

The SMACOF method provides a basic framework for a variety of applications ranging from graph drawing [17], [18] to target tracking [19], [20]. In the topic of molecular conformation [21]–[24], application-specific methods address a similar problem. More recent works have proposed machine learning approaches to the molecular conformation problem

[25]–[27]. These methods generally do not provide guarantees of global optimality even if the solutions are practically useful in the proposed applications. Particularly in the case of noisy distance measurements, for which the globally optimal solution likely has a nonzero objective value, quantifying the suboptimality of a solution is an important challenge. Our proposed method provides a numerical bound on the suboptimality of a given estimate.

Matrix completion

Also arising from the MDS literature [28], the *s-stress* function penalizes the difference between estimated and measured *squared* distances. Squaring the distance terms ensures global differentiability at the expense of distorting the objective function[1]. Although we emphasize solving the unsquared distance-based localization problem in this paper, the proposed algorithm more closely resembles methods addressing this alternative problem.

---

[1]Minimizing errors between squared distances alters the optimal solution, as penalizing the differences between estimated and measured squared distances makes the objective function more sensitive to errors for larger measured distances than for smaller measured distances, even if terms are weighted proportionally. We also note that many physical sensors measure distances rather than squared distances. The solution to the MLE problem is the set of feasible distances that best approximates a set of measured distances in the sense of least-squares.

The squared distance-based localization can be formulated as a Euclidean distance matrix (EDM) completion problem. The EDM—the matrix of all pairwise squared distances—is an essential tool for many localization algorithms, and [11], [29]–[32] provide thorough overviews on the properties of EDMs. The EDM completion problem comprises a rank-constrained (and thus non-convex) semidefinite program (SDP), for which solution methods generally belong to one of two perspectives.

One approach to solving rank-constrained SDPs is to use the Burer-Monteiro factorization [33], [34], which ensures feasible solutions albeit with a non-convex objective function. Consequently, iterative methods such as gradient descent, majorization-minimization [15], or Newton's method [35] are susceptible to local minima whether the distance measurements are squared or not. More recently, quartic first-order methods [36] promise a substantial improvement in efficiency over previous iterative methods. The Dyn-NoLips algorithm proposed in [37] applies quartic first-order methods to the EDM completion problem (see also [38]), though it does not eliminate the need for a good initialization.

An alternative approach to low-rank matrix completion is to replace the rank constraint with a convex approximation. Heuristics such as regularizing the objective function with a penalty on the trace of the estimated EDM encourage low-rank solutions [10]. Other SDP-based approaches for distance-based localization include [39]–[41]. While convex relaxations of the EDM completion problem enable global optimization, the resulting solutions only approximate the global optima of the original problem, as illustrated in Figure 1b. Additional matrix completion approaches include those in [42], [43]. Riemannian methods for matrix completion are explored in [44]–[46]. Several noteworthy papers consider application-focused modifications to EDM completion: [47], extends EDMs to trajectories over multiple timesteps; [48] uses a majorization-minimization scheme for optimization with TDOA signals; [49]–[51] address inverse kinematics in robotics with a distance-based formulation.

In this paper, we introduce an exact reformulation of the MLE problem that yields a rank-constrained SDP. As with the EDM completion problem, we must reconcile the guaranteed convexity of an SDP relaxation with the guaranteed feasibility of an iterative method. By using the relaxed SDP solution to initialize a fast iterative solver, our method provides both a numerical lower bound for the globally optimal solution and a feasible exact solution, which empirically outperforms EDM completion methods.

Manifold optimization and low-rank SDPs

Our approach draws on a rich set of tools adressing quadratically constrained quadratic programs (QCQPs). In particular, we build on the Boumal's Riemannian Staircase method [52], which finds the lowest rank Burer-Monteiro factorization that yields a globally optimal solution. The Riemannian Staircase is fundamental to SE-Sync [53], an algorithm that provides certifiable results for the pose graph optimization problem, and it has also found applications in other related problems [54], [55]. As in these approaches, our algorithm uses the solution of a dual problem to provide a lower bound that certifies the optimality of our estimate. We also employ the Manopt solver [56], which uses a Riemannian trust-regions approach (see [57]–[60]) to solve the primal MLE problem. The tightness of the convex relaxation is further explored in [61] and [62].

While our formulation of the MLE problem is suitable for direct application of the Riemannian Staircase, several problem-specific considerations make this approach inefficient. In particular, we observe that the global optimum occurs at a higher rank than in other applications such as SE-Sync (where it is often sufficient to lift the rank by only one). In order to circumvent many successive optimization steps, our method, which we term the "Riemannian Elevator", ascends directly to the desired rank by first solving the dual problem.

B. Contributions

The contributions of this paper include

- A reformulation of the distance-based localization problem, for which we prove equivalence to the original problem in Section II
- A relaxation of the problem that yields a provable lower bound for the solution to the original problem (Section III)
- Closed-form bounds for the dual of the relaxation as well as upper bounds on the rank of the solution to the relaxation, also in Section III
- A proof that the relaxed primal problem is equivalent to a feasibility problem (Section IV)
- The Riemannian Elevator algorithm for distance-based localization, described in Section IV
- Numerical comparisons of the Riemannian Elevator to other distance-based localization methods in Section V.

II. PROBLEM FORMULATION

A. Notation

1) Linear algebra

In general, we denote vectors indexed $1, \ldots, n$ as $x_1, \ldots, x_n$. We refer to these individual vectors collectively as $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$, *i.e.* each $x_i^\top$ corresponds to the ith *row* of X. For any matrix A, $[A]_{ij}$ refers to its (i,j)-th element. The notation $\mathrm{vec}(X)$ indicates the vertical concatenation of the columns of X.

For a matrix $X \in \mathbb{R}^{n \times n}$, $\mathrm{diag}(X)$ refers to a vector of its diagonal elements. For a vector $x \in \mathbb{R}^{n \times 1}$, $\mathrm{Diag}(x) \in \mathbb{R}^{n \times n}$ contains the elements of x on the diagonal and zeros elsewhere, such that $\mathrm{diag}(\mathrm{Diag}(x)) = x$.

Unless otherwise indicated, $\|x\|$ refers to the 2-norm of vector x. For a square matrix X, $\|X\|_F$ refers to the Frobenius norm, *i.e.* the 2-norm of $\mathrm{vec}(X)$. We also use the notation $\|x\|_W$ to indicate a W-weighted norm, such that $\|x\|_W^2 = x^\top W x$. Similarly, $\|X\|_W$ refers to the Frobenius norm of the matrix X weighted by W as $\mathrm{tr}\left(X^\top W X\right)^{\frac{1}{2}}$.

Of particular interest to this paper is the space of unit vectors. The set of d-dimensional unit vectors is the unit sphere $\mathcal{S}^{d-1} \subseteq \mathbb{R}^d$:

$$\mathcal{S}^{d-1} = \{y \in \mathbb{R}^d \mid y^\top y = 1\}.$$

$\mathcal{S}^{d-1}$ is a $(d-1)$-dimensional submanifold of $\mathbb{R}^d$ [58]. We use the notation

$$(\mathcal{S}^{d-1})^n = \{Y \in \mathbb{R}^{n \times (d-1)} \mid y_i^\top y_i = 1, \ i = 1, \ldots, n\}$$

in reference to the space of n d-dimensional unit vectors, also known as the *oblique manifold*. In general, we use X in reference to matrices in Euclidean space while reserving the variable Y for matrices on the oblique manifold specifically.

We refer to symmetric positive-semidefinite matrices using the notation $\mathbb{S}_+^n \subseteq \mathbb{R}^{n \times n}$, while $\succ$ and $\prec$ refer to matrix inequalities in the sense of definiteness. The operator $\circ$ refers to the Hadamard (elementwise) product. We refer to the Moore-Penrose inverse of matrix A as $A^\dagger$.

2) Graph theory

The graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of n vertices $\mathcal{V} = \{1, \ldots, n\}$ and m edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The incidence matrix is $C(\mathcal{G}) \in \mathbb{R}^{m \times n}$ such that $[C]_{ki} = -1$ and $[C]_{kj} = 1$ for each edge $e_k = (i,j)$ for $k = 1, \ldots, m$. The graph Laplacian is $L(\mathcal{G}) \in \mathbb{R}^{n \times n}$ and $L = C^\top C$.

In the context of distance-only localization, we are particularly concerned with the embedding of a graph, where vertices refer to points in Euclidean space and the edges indicate distance measurements between pairs of points. The terms *vertex positions*, *edge distances*, and *edge directions* refer to those quantities for a particular graph embedding.

B. The distance-only localization problem

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a set of measured edge distances $\tilde{d}_{ij}$ for all $(i,j) \in \mathcal{E}$, the distance-only localization problem consists of finding vertex positions $x_1, \ldots, x_n$, such that the estimated edge distances are as close as possible to the measured edge distances:

Problem 1 (MLE for distance-based localization)**.**

$$p_{\text{MLE}}^* = \min_{X \in \mathbb{R}^{d \times n}} \sum_{(i,j) \in \mathcal{E}} \frac{w_{ij}}{2} \left(\|x_j - x_i\| - \tilde{d}_{ij} \right)^2. \quad (1)$$

We refer to the solution of Problem 1 as the maximum likelihood estimate (MLE) under the assumption that measurement noise arises under the following model:

$$\tilde{d}_{ij} = \|x_j - x_i\| + v_{ij}, \quad v_{ij} \sim \mathcal{N}\left(0, \sigma_{ij}^2\right), \quad (2)$$

where σ_{ij}^2 is the variance of the noise. The weights w_{ij} are proportional to the inverse variance of the distance measurements (*i.e.*, $w_{ij} = \frac{\alpha}{\sigma_{ij}^2}$ for some $\alpha \in \mathbb{R}$ for all $(i,j) \in \mathcal{E}$). The objective function in (1) also resembles the stress function introduced in [13]. An analogous interpretation of Problem 1 is as the minimization of the total energy in a network of linear springs, each with spring constant w_{ij} and rest length $\tilde{d}_{ij}$.

As described in Section I-A, many distance-only localization methods instead address the similar but globally differentiable problem of EDM completion:

Problem 2 (EDM completion)**.**

$$p_{\text{EDM}}^* = \min_{X \in \mathbb{R}^{d \times n}} \sum_{(i,j) \in \mathcal{E}} \frac{w_{ij}}{2} \left(\|x_j - x_i\|^2 - \tilde{d}_{ij}^2 \right)^2 \quad (3)$$

We can also write Problem 2 with the EDM as the decision variable:

$$\min_{D \in \mathcal{D}} \left\| \bar{W} \circ (D - \tilde{D}) \right\|_F^2, \quad (4)$$

where $\bar{W}_{ij} = w_{ij}$ if $(i,j) \in \mathcal{E}$ and 0 otherwise and $\mathcal{D} \subseteq \mathbb{S}_+^n$ consists of all feasible EDMs for n points in $\mathbb{R}^d$. Given an EDM D, we can recover a corresponding set of positions $X \in \mathbb{R}^{n \times n}$ using linear transformations and an eigenvalue decomposition (see [11] for a thorough description). However, if we relax the non-convex rank constraint, the D^* that is closest to the measured $\tilde{D}$ does not correspond exactly to some X in the desired embedding dimension d, and in the worst case $\text{rank}(X) = \text{rank}(D) - 2$ [11]. Therefore, a key challenge in solving the EDM completion problem is in constraining the rank of D so that it maps to a realizable solution.

Problem 2 is the MLE under the assumption of additive Gaussian noise on the *squared* distances rather than on the raw distances as in (2). This presupposition of sensors that directly measure squared distances is not a reasonable assumption in most applications. In general the solution to Problem 2 is not equivalent to the original MLE, with the notable exception of the noiseless case (*i.e.* the solution to both Problems 1 and 2 has an objective value of exactly zero when $\sigma_{ij}^2 = 0$ for all $(i,j) = \mathcal{E}$).

Rather than solve Problem 2 as an approximation of Problem 1, we instead propose a novel reformulation of the original MLE objective function. Introducing an additional variable $y_{ij} \in \mathcal{S}^{d-1}$ corresponding to the direction of each edge (i,j), we consider the following problem:

Problem 3 (MLE over edge directions)**.**

$$p_{\text{MLE}^+}^* = \min_{\substack{X \in \mathbb{R}^{n \times d}, \\ Y \in (\mathcal{S}^{d-1})^m}} \sum_{(i,j) \in \mathcal{E}} \frac{w_{ij}}{2} \left\| x_j - x_i - \tilde{d}_{ij} y_{ij} \right\|^2. \quad (5)$$

The inclusion of edge directions as additional variables makes the objective function convex at the expense of introducing the additional non-convex constraints $y_{ij}^\top y_{ij} = 1$. Problem 3 is a homogeneous quadratically-constrained quadratic program (QCQP) and as such is known to be NP-hard. Nevertheless, this reformulation enables us to construct a series of relaxations that yield a provable lower bound to $p_{\text{MLE}^+}^*$. Furthermore, Problem 3 is a *tight* relaxation of Problem 1: the positions $x_1^*, \ldots, x_n^*$ that minimize the objective in (5) also minimize the objective in (1). Consequently, the lower bound that we derive in Section III also applies to the original problem.

Lemma 1 (Equivalence of Problems 1 and 3)**.** *The positions X^* minimize Problem 1 if and only if there exist edge directions Y^* such that (X^*, Y^*) minimize Problem 3, and $p_{\text{MLE}}^* = p_{\text{MLE}^+}^*$.*

Proof. First, we show that for X^* fixed, minimizing the objective of Problem 3 over $Y \in (\mathcal{S}^{d-1})^m$ has the solution

$$y_{ij}^* = \frac{x_j^* - x_i^*}{\|x_i^* - x_j^*\|}. \tag{6}$$

Given X^*, the optimization over edge directions can be separated into m independent problems, which we can solve geometrically. Letting $\delta_{ij} = x_j^* - x_i^* - \tilde{d}_{ij} y_{ij}$, we consider the triangle in $\mathbb{R}^d$ with edges of lengths $\|\delta_{ij}\|$, $\|x_j - x_i\|$, and $\tilde{d}_{ij}$. By the law of cosines,

$$\|\delta\|_{ij}^2 = \|x_j - x_i\|^2 + \tilde{d}_{ij}^2 - 2\|x_j - x_i\|\tilde{d}_{ij}\cos\theta,$$

where $\cos\theta = y_{ij}^\top \frac{(x_j - x_i)}{\|x_j - x_i\|}$. Clearly $\|\delta_{ij}\|^2$ attains a minimum when $\cos\theta = 1$, which only occurs at the value of y_{ij}^* specified in (6).

Since we have a closed-form solution for y_{ij} given x_i and x_j, we make the following substitution:

$$
\begin{aligned}
p_{\text{MLE}^+}^* &= \min_{X \in \mathbb{R}^{d \times n}} \sum_{(i,j)} \frac{w_{ij}}{2} \left\| x_j - x_i - \tilde{d}_{ij} \frac{(x_j - x_i)}{\|x_j - x_i\|} \right\|^2 \\
&= \min_{X \in \mathbb{R}^{d \times n}} \sum_{(i,j)} \frac{w_{ij}}{2} \left\| \frac{(x_j - x_i)\left(\|x_j - x_i\| - \tilde{d}_{ij}\right)}{\|x_j - x_i\|} \right\|^2 \\
&= \min_{X \in \mathbb{R}^{d \times n}} \sum_{(i,j)} \frac{w_{ij}}{2} \left\| \frac{x_j - x_i}{\|x_j - x_i\|} \right\|^2 \left(\|x_j - x_i\| - \tilde{d}_{ij}\right)^2 \\
&= \min_{X \in \mathbb{R}^{d \times n}} \sum_{(i,j)} \frac{w_{ij}}{2} \left(\|x_j - x_i\| - \tilde{d}_{ij}\right)^2 \\
&= p_{\text{MLE}}^*
\end{aligned}
$$

To summarize, if (X^*, Y^*) minimize Problem 3, then Y^* relates to X^* by (6), and X^* minimizes Problem 1 with the same objective value. Similarly, if X^* minimizes Problem 1, then (X^*, Y^*) minimize Problem 3 when we choose Y^* according to (6). $\square$

This result suggests that solving the QCQP in Problem 3 also solves Problem 1. In practice, we first solve a further relaxation of Problem 3, which provides a provable lower bound as well as a useful initialization.

III. CERTIFIABLE OPTIMIZATION

We generate a certificate for Problem 3 using a relaxation (Section III-A) for which the Lagrangian dual is a convex SDP (III-B). The solution to the dual problem provides three important quantities:

1) a numerical lower bound for Problem 3 (the certificate),
2) the minimum rank of the relaxed problem that achieves this lower bound,
3) a set of orthonormal vectors that form a basis for the solution to the relaxed problem.

The certificate bounds the suboptimality of the primal solution, and the rank and basis substantially improve the efficiency of finding the solution. In contrast to the Riemannian Staircase method [52], which successively lifts the problem to higher dimensions until finding a certificate of optimality, we proceed directly to the dimension of the globally optimal solution by first solving the dual problem. In this section, we address this dual problem, which aids in solving the original problem in Section IV.

A. Relaxing Problem 3

As a preliminary step, we recast Problem 3 as an optimization over edge direction only, rather than over both edge directions and vertex positions. For simplicity, we rewrite (5) as

$$p_{\text{MLE}^+}^* = \min_{\substack{X \in \mathbb{R}^{n \times d}, \\ Y \in (\mathcal{S}^{d-1})^m}} \left\| CX - \tilde{D}Y \right\|_W^2, \tag{7}$$

where $C = C(\mathcal{G}) \in \mathbb{R}^{m \times n}$ is the incidence matrix of the underlying graph $\mathcal{G}$, $\tilde{D} \in \text{Diag}(\mathbb{R}^m)$ is a diagonal matrix consisting of the measured distances along each edge, and $W \in \text{Diag}(\mathbb{R}^m)$ is contains the weights of each measurement agreement term. We apply the Schur complement to factor out the vertex positions from the problem. Given a fixed value of Y^*, we have an unconstrained minimization over the vertex positions X with the solution

< Remove first C

$$X^* = C(C^\top W C)^\dagger C^\top W \tilde{D} Y^*. \tag{8}$$

Using the closed-form solution for X^* in terms of Y^* gives $p_{\text{MLE}^+}^*$ in terms of Y^* only:

$$p_{\text{MLE}^+}^* = \min_{Y \in (\mathcal{S}^{d-1})^m} \text{tr}\left(QYY^\top\right), \tag{9}$$

where $Q = \tilde{D}^2(W - WC(C^\top W C)^\dagger C^\top W)$.

Remark 2. *If $\mathcal{G}$ is connected and $W \succ 0$, then $Q \succeq 0$ with rank $m - n + 1$.*

Proof. Let

$$M = \begin{bmatrix} C^\top \\ \tilde{D} \end{bmatrix} W \begin{bmatrix} C & \tilde{D} \end{bmatrix} = \begin{bmatrix} C^\top W C & C^\top W \tilde{D} \\ \tilde{D} W C & W \tilde{D}^2 \end{bmatrix}.$$

Then $Q = \tilde{D}^2(W - WC(C^\top W C)^\dagger C^\top W)$ is the generalized Schur complement of the block $W\tilde{D}^2$. As shown in [63], $\text{rank}(Q) = \text{rank}(M) - \text{rank}(C^\top W C)$. Since $\tilde{D}$ is a diagonal matrix with nonzero diagonal entries, $\text{rank}(M) = m$. Noting that $C^\top W C$ is the weighted graph Laplacian of the underlying graph $\mathcal{G}$, its rank is $n - 1$ if $\mathcal{G}$ is connected. Therefore, $\text{rank}(Q) = m - n + 1$. Furthermore, we note that if $W \succeq 0$, then $M \succeq 0$ and $C^\top W C \succeq 0$. As shown in [64, Theorem 4.3], if $M \succeq 0$, then $Q \succeq 0$. $\square$

Noting that Q is the sum of a diagonal matrix and a dense low-rank matrix of rank $n - 1$, we use the decomposition $Q = Q_1 + Q_2$, where $Q_1 = \tilde{D}^2 W$ and $Q_2 = -\tilde{D}^2 WC(C^\top W C)^\dagger C^\top W$. Given that each column of Y has unit norm, we observe that $\text{tr}(Q_1 Y^\top Y) = \text{tr}(Q_1)$. Therefore, (9) is equivalent to

$$p_{\text{MLE}^+}^* = \min_{Y \in (\mathcal{S}^{d-1})^m} \text{tr}(Q_1) + \text{tr}\left(Q_2 YY^\top\right). \tag{10}$$

Next, we consider a relaxation of (9) in which we optimize over r-dimensional unit vectors for some integer $r \geq d$:

Problem 4 (Relaxation of Problem 3).

$$p_r^* = \min_{Y_r \in (\mathcal{S}^{r-1})^m} \operatorname{tr}(Q_1) + \operatorname{tr}\left(Q_2 Y_r Y_r^\top\right). \qquad (11)$$

Problem 4 expands the search space of Problem 3. Any p_{MLE}^* must also be feasible for the relaxed problem, meaning that $p_r^* \leq p_{\mathrm{MLE}}^*$ for all $r \geq d$. By the same logic, $p_{r_2}^* \leq p_{r_1}^*$ for any $r_1 < r_2$.

In the extreme of $r = m$ (and possibly for $r \leq m$ as well), Problem 4 is equivalent to the following SDP:

Problem 5 (SDP relaxation of 3).

$$p_{\mathrm{SDP}}^* = \min_{Z \in \mathbb{S}_+^m} \operatorname{tr}(Q_1) + \operatorname{tr}(Q_2 Z)$$
$$\text{subject to} \quad \operatorname{diag}(Z) = \mathbf{1}. \qquad (12)$$

Solving this SDP relaxation gives a lower bound for the original problem (*i.e.* $p_{\mathrm{SDP}}^* \leq p_r^* \leq p_{\mathrm{MLE}}^*$ for $m \leq r \leq d-1$). Furthermore, the solution admits a factorization $Z^* = Y_m^* Y_m^{*\top}$, where $Y_m^* \in (\mathcal{S}^{m-1})^m$. Projecting Y_m^* to $(\mathcal{S}^{d-1})^m$ provides a (not necessarily optimal) solution to the original problem. While p_{SDP}^* has the advantage of being convex, the downside is that we have expanded the search space from $Y \in (\mathcal{S}^{d-1})^m \subseteq \mathbb{R}^{d \times m}$ to $Z \in \mathbb{S}_+^m$. Not only does this increase computation time, but also it may be difficult to project a solution in $(\mathcal{S}^{m-1})^m$ to $(\mathcal{S}^{d-1})^m$ without a significant increase in the objective function. Therefore, we seek the smallest r for which $p_{\mathrm{SDP}}^* = p_r^*$. In fact, it is always the case that a low-rank solution exists, as proven by Barvinok [65] (in a paper addressing low-rank solutions for a relaxation of Problem 2) and Pataki [66] (see also [67] for a thorough discussion of this result):

Theorem 3 (Existence of a low-rank solution [65], [66]). *Problem 5 is a tight relaxation of Problem 4 (i.e. $p_{SDP}^* = p_r^*$) for all r satisfying*

$$r \geq \left\lfloor \frac{-1 + \sqrt{8m+1}}{2} \right\rfloor.$$

For example, given a measurement graph with 500 edges, there must exist some $Y_r^* \in (\mathcal{S}^{r-1})^m$ for $r \leq 31$ such that $Z^* = Y_r^* Y_r^{*\top}$ minimizes Problem 5, regardless of the number of vertices. While it may not yet be obvious that a 31-dimensional embedding aids in our search for the 2- or 3-dimensional embeddings we typically seek for Problem 1, this insight nonetheless reduces the search space for p_{SDP}^*, which we know is a lower bound for p_{MLE}^*.

B. Solving the dual SDP

The Lagrangian form of the SDP relaxation in (12) is

$$\mathcal{L}(Z, \Lambda) = \operatorname{tr}(Q_1) + \operatorname{tr}(Q_2 Z) + \sum_{i=1}^m \lambda_i \left(1 - e_i^\top Z e_i\right) \quad (13)$$
$$= \operatorname{tr}(Q_1) + \operatorname{tr}(Q_2 Z) + \operatorname{tr}(\Lambda(I - Z)) \qquad (14)$$
$$= \operatorname{tr}(Q_1) + \operatorname{tr}((Q_2 - \Lambda) Z + \Lambda), \qquad (15)$$

where $Z \in \mathbb{S}_+^m$ and $\Lambda = \operatorname{Diag}(\lambda_1, \ldots, \lambda_m) \in \operatorname{Diag}(\mathbb{R}^m)$. We note that $\mathcal{L}(Z, \Lambda)$ is unbounded from below with respect

to Z if $Q - \Lambda \nsucceq 0$. Consequently, we express the positive-definiteness of $Q - \Lambda$ as a constraint, and pose the Lagrangian dual problem in the following form:

Problem 6 (Dual problem of MLE over edge directions).

$$p_{\mathrm{DSDP}}^* = \max_{\Lambda \in \operatorname{diag}(\mathbb{R}^m)} \operatorname{tr}(Q_1) + \operatorname{tr}(\Lambda)$$
$$\text{subject to} \quad R = Q_2 - \Lambda \qquad (16)$$
$$R \succeq 0$$

Noting that $p_{\mathrm{DSDP}}^* \leq p_{\mathrm{SDP}}^* \leq p_r^*$, if we have Λ^* and Y_r^* such that strong duality holds, then Λ^* certifies Y_r^* as globally optimal. Furthermore, solving the dual problem reveals the rank r for which we can find a certifiably optimal Y_r^*. The following theorem shows that the rank of the null space of $R^* = Q_2 - \Lambda^*$ provides an upper bound for r:

Theorem 4 (Upper bound on rank of low-rank solution [67]). *Let Z^* and R^* solve Problems 5 and 6, respectively, such that $p_{\mathrm{SDP}}^* = p_{\mathrm{DSDP}}^*$. Then*

$$\operatorname{rank}(Z^*) + \operatorname{rank}(R^*) \leq m.$$

This result follows from complementary slackness (see [67]–[69]). If $p_{\mathrm{DSDP}}^* = p_{\mathrm{SDP}}^*$, then $(Q_2 - \Lambda^*) Z^* = 0$, *i.e.* the ranges of $R^* \in \mathbb{R}^{m \times m}$ and $Z^* \in \mathbb{R}^{m \times m}$ are orthogonal. Consequently, given Λ^* that solves Problem 6, we choose $r = m - \operatorname{rank}(R^*)$ and expect to find a solution of p_r^* with zero duality gap.

We expand on this prior result to establish additional bounds on the solution to p_{SDP}^* and the minimum rank r for which we obtain $p_r^* = p_{\mathrm{SDP}}^*$.

Lemma 5 (Negative definiteness of the optimal dual). *Given $Q_2 \preceq 0$ with $\operatorname{diag}(Q_2) < 0$ and Λ^* which solves Problem 6, then $\Lambda^* \preceq Q2 \circ I \prec 0$.*

Proof. If Λ^* solves Problem 6, then $R^* = Q_2 - \Lambda^* \succeq 0$. If $R^* \succeq 0$, all of its diagonal elements must be non-negative. Therefore, we require $\lambda_i \leq [Q_2]_{ii} < 0$, which implies the diagonal matrix Λ^* is negative definite. $\square$

Corollary 6 (Upper bound on optimal objective value). *Given $Q_2 \preceq 0$ with $\operatorname{diag}(Q_2) < 0$ and Λ^* which solves Problem 6, then $p_{\mathrm{SDP}}^* \leq \operatorname{tr}(Q)$.*

Proof. If Λ^* solves Problem 6, then we have $p_{\mathrm{SDP}}^* = p_{\mathrm{DSDP}}^* = \operatorname{tr}(Q_1) + \operatorname{tr}(\Lambda) \leq \operatorname{tr}(Q_1) + \operatorname{tr}(Q_2 \circ I) = \operatorname{tr}(Q)$. $\square$

Theorem 7 (Upper bound on minimum rank). *If a solution exists to Problem 5, where $Q_2 \preceq 0$, $\operatorname{diag}(Q_2) < 0$, and $\operatorname{rank}(Q_2) = r$, then there exists a solution Z^* such that $\operatorname{rank}(Z^*) \leq r$.*

Proof. Lemma 5 shows that if $Q_2 \preceq 0$ with $\operatorname{diag}(Q_2) < 0$, then $\Lambda^* \prec 0$. We then refer to [70, Theorem 1], which establishes that for matrix $R = Q_2 + \Lambda$ with $\operatorname{rank}(Q_2) = r$, all but r eigenvalues of R lie on the interval $[\min_i \lambda_i, \max_i \lambda_i]$. As $\max_i \lambda_i^* < 0$, at least $m - r$ eigenvalues of R^* must be strictly negative, implying that $\operatorname{rank}(R^*) \geq m - r$. By Theorem 4, $\operatorname{rank}(Z^*) \leq m - \operatorname{rank}(R^*) = r$. Remark 2 establishes conditions for when $r = n - 1$. $\square$

Compiling the results of Theorems 3, 4, and 7, we conclude that for a connected measurement graph, the minimum rank r for which $p_r^* = p_{\text{SDP}}^* = p_{\text{DSDP}}^*$ has the following upper bound:

$$r \leq \min\left(\left\lfloor \frac{-1 + \sqrt{8m+1}}{2} \right\rfloor, m - \text{rank}(R^*), n - 1\right). \tag{17}$$

The first term depends only on the number of edges of the measurement graph and is therefore quickly computed. The second term comes as an added benefit of numerically solving the dual problem and typically provides the tightest bound of the three terms. The third term, while readily available, is only tighter than the first term for fully connected graphs. However, Theorem 7 provides an additional insight: besides the lower bound for p_{MLE}^* that we compute from solving Problem 6, we can establish a provable upper bound for p_{MLE}^*.

Let $\tilde{Q}_2$ be a rank d projection of Q_2 such that $Q_2 \preceq \tilde{Q}_2$. Solving Problem 6 with $\tilde{Q}_2$ replacing Q_2 (the result of which we will refer to as $\tilde{p}_{\text{DSDP}}$) yields the optimal dual $\tilde{\Lambda}$. Theorem 7 implies that there exists a corresponding primal solution $\tilde{Z}$ such that $\text{rank}(\tilde{Z}) \leq d$. This primal solution yields an exact decomposition $\tilde{Z} = \tilde{Y}\tilde{Y}^\top$ such that $\tilde{Y}$ is feasible for Problem 9. Therefore,

$$p_{\text{MLE}}^* \leq \tilde{p}_{\text{DSDP}} = \text{tr}(Q_1) + \text{tr}(\tilde{\Lambda}), \tag{18}$$

which provides an upper bound for the optimal solution without directly solving the primal problem.

A question for further study is how to find $\tilde{Q}_2$ that provides the tightest upper bound to the original problem. An obvious choice is to use an eigenvalue decomposition; retaining the d most negative eigenvalues of Q_2 while setting the rest to zero provides $\tilde{Q}_2$ of appropriate rank. On the other hand, given that the diagonal terms of Q_2 do not affect the minimizing solution Z^*, we could instead consider perturbing Q_2 on its diagonal in order to generate a low-rank $\tilde{Q}_2$. This is an example of an inverse eigenvalue problem [71].

In practice, we solve Problem 6 directly. The optimal dual Λ^* bounds the suboptimality of some solution p_{MLE}^* according to the inequality

$$\text{tr}(Q_1) + \text{tr}(\Lambda^*) \leq p_{\text{MLE}}^*. \tag{19}$$

Furthermore, $R^* = Q_2 - \Lambda^*$ provides the minimum rank r for which $p_{\text{DSDP}}^* = p_r^*$ according to (17). As we discuss in Section IV, R^* also narrows the search space for this solution.

IV. THE RIEMANNIAN ELEVATOR

Solving Problem 6 provides a rank r for which we can expect to find a solution to 4 as well as the value of that solution. There are several additional steps to finding a solution to Problem 3. The full Riemannian Elevator algorithm consists of three main phases:

1) solving the dual problem (Problem 6),
2) solving the high-dimensional relaxation of the primal problem (Problem 4)
3) solving the original problem (Problem 3).

The remainder of this section provides the implementation details for applying the Riemannian Elevator to the distance-

---

Algorithm 1 The Riemannian Elevator

Inputs: Incidence matrix C, measurements $\tilde{D}$, weights W
Outputs: Estimate X^*, cost p_{MLE}^*, lower bound p_{SDP}^*

1: **function** RIEMANNIANELEVATOR
2: $\quad Q_1 \leftarrow \tilde{D}^2 W$
3: $\quad Q_2 \leftarrow -\tilde{D}^2 W C (C^\top W C)^\dagger C^\top W$
$\quad$ Solve the dual SDP:
4: $\quad \Lambda^* \leftarrow \arg\max_{\Lambda \in \text{Diag}(\mathbb{R}^m)} \text{tr}(\Lambda)$ s.t. $Q_2 - \Lambda \succeq 0$
5: $\quad p_{\text{SDP}}^* \leftarrow \text{tr}(Q_1) + \text{tr}(\Lambda^*) \qquad \triangleright$ lower bound
$\quad$ Find a basis for Y_r:
6: $\quad R^* \leftarrow Q_2 - \Lambda^*$
7: $\quad V \leftarrow \text{null}(R^*) \qquad \triangleright$ any basis for $\text{null}(R^*)$
$\quad$ Solve relaxed primal problem:
8: $\quad \bar{Z} \leftarrow \arg\min_{\bar{Z} \in \mathbb{S}_+^r} \left\| \text{diag}(V\bar{Z}V^\top) - \mathbf{1} \right\|_2^2$
9: $\quad Y_r^* \leftarrow \text{chol}(\bar{Z})^\top V$
$\quad$ Solve MLE problem:
10: $\quad Y^0 \leftarrow \text{proj}(Y_r^*) \qquad \triangleright$ initialization for Y^*
11: $\quad Y^* \leftarrow \arg\min_{Y \in (\mathcal{S}^{d-1})^m} \text{tr}(Q_2 Y Y^\top)$
12: $\quad X^* \leftarrow C(C^\top W C)^\dagger C^\top W \tilde{D} Y^*$
13: $\quad$ **return** X^*, p_{MLE}^*, p_{SDP}^*
14: **end function**

---

based localization problem, including solving Problem 4 using the results of Problem 6, projecting the solution of Problem 4 from rank r to rank d, and solving Problem 3 on the manifold of unit vectors.

A. Initialization from the dual solution

Solving Problem 6 provides a numerical lower bound for the solution to Problem 4 as well as an upper bound on the rank for which the bound is tight. Additionally, R^* provides a basis for Y_r^*. We introduce a stronger claim than that in Theorem 4: not only must Y_r^* lie in $\text{null}(R^*)$, but if we find *any* feasible Y_r that is also orthogonal to R^*, then that Y_r minimizes Problem 4.

Lemma 8 (Optimality conditions for Problem 4). *Given* $r \geq m - \text{rank}(R^*)$, *where* $R^* = Q_2 - \Lambda^*$, Y_r^* *minimizes Problem 4 if and only if* $Y_r^* \in \mathcal{S}(r-1)^m \cap \text{null}(R^*)$.

Proof. We already know that Y_r^* minimizing Problem 4 implies $Y_r^* \in \text{null}(R^*)$ from Theorems 4 and 7. Theorem 7 guarantees that if $r \geq m - \text{rank}(R^*)$, then Z^* with rank r minimizes Problem 5. Choosing Y_r^* such that $Z^* = Y_r^* Y_r^{*\top}$ gives $p_r^* = p_{\text{SDP}}^*$.

Next, we assume for the sake of contradiction that there exists some $Y_r \in \mathcal{S}(r-1)^m \cap \text{null}(R^*)$ such that

$$\text{tr}(Q_1) + \text{tr}\left(Q_2 Y_r Y_r^\top\right) \neq \text{tr}(Q_1) + \text{tr}\left(Q_2 Y_r^* Y_r^{*\top}\right). \tag{20}$$

If $R^* = Q_2 - \Lambda^*$, then it must also be true that

$$\text{tr}\left((R^* + \Lambda^*)Y_r Y_r^\top\right) \neq \text{tr}\left((R^* + \Lambda^*)Y_r^* Y_r^{*\top}\right). \tag{21}$$

We note that both Y_r and Y_r^* lie in $(\mathcal{S}^{r-1})^m \cap \text{null}(R^*)$, so

$$R^* Y_r = R^* Y_r^* = \mathbf{0}. \tag{22}$$

Furthermore, since $\mathrm{diag}(Y_r Y_r^\top) = \mathrm{diag}(Y_r^* Y_r^{*\top}) = \mathbf{1}$ and Λ^* is a diagonal matrix, it follows that

$$\mathrm{tr}(\Lambda^* Y_r Y_r^\top) = \mathrm{tr}(\Lambda^* Y_r^* Y_r^{*\top}) = \mathrm{tr}(\Lambda^*). \tag{23}$$

Therefore, we can simplify (21) as

$$\mathrm{tr}(\Lambda^*) \neq \mathrm{tr}(\Lambda^*). \tag{24}$$

We conclude that $Y_r \in (\mathcal{S}^{r-1})^m \cap \mathrm{null}(R^*)$ is a necessary and sufficient condition for Y_r to minimize Problem 4. However, Y_r^* is not necessarily a unique solution to Problem 4. $\qquad\square$

Lemma 8 suggests that we can substantially narrow the search space for Y_r^* from $(\mathcal{S}^{r-1})^m$ to $(\mathcal{S}^{r-1})^m \cap \mathrm{null}(R^*)$. Let $V \in \mathbb{R}^{m \times r}$ denote some basis for $\mathrm{null}(R^*)$. Lemma 8 suggests that finding a rank-r Z that minimizes Problem 5 is equivalent to finding $\bar{Z} \in \mathbb{S}_+^r$ such that $Z = V\bar{Z}V^\top$ is feasible, *i.e.* solving

$$\min_{\bar{Z} \in \mathbb{S}_+^r} \left\| \mathrm{diag}(V\bar{Z}V^\top) - \mathbf{1} \right\|_2^2. \tag{25}$$

This minimization is equivalent to fitting a hyperellipsoid to m points in $\mathbb{R}^r$.

Given $\bar{Z}^*$ that minimizes (25), $Z^* = V\bar{Z}^*V^\top$ is rank r by construction. The rank r solution to Problem 4 is then Y_r^* such that $Y_r^* Y_r^{*\top} = Z_r^*$. We use a projection of Y_r^* to $(\mathcal{S}^{d-1})^m$ as an initial guess for solving Problem 3.

B. Projecting and refining the solution

Projecting the solution to the high-dimensional relaxation Y_r^* from $(\mathcal{S}^{r-1})^m$ to $(\mathcal{S}^{d-1})^m$ provides a useful initialization for solving Problem 3. This approach is similar to rank-constrained SDP algorithms, including SE-Sync [53] and the Goemans-Williamson algorithm for MaxCut [72]. While these algorithms typically treat the projection of the high-dimensional solution as optimal and do not refine further, this assumption is generally insufficient in the case of the distance-based localization problem. Because the relaxed solution Y_r^* usually does not project to a local minimizer of (5), we include a subsequent refinement step that finds a local minimum Y^*.

Given Y^*, if $p_{\mathrm{MLE}}^* = p_r^* = p_{\mathrm{SDP}}^*$, then Y^* is certainly the global optimum to Problem 3. The converse is not necessarily true: if $p_{\mathrm{MLE}}^* > p_{\mathrm{SDP}}^*$ the solution Y^* may or may not be the global optimum of the rank-constrained problem. Either way, the duality gap $p_{\mathrm{MLE}}^* - p_{\mathrm{SDP}}^*$ provides an estimate of the worst-case suboptimality of the solution.

In the refinement step, we use a Riemannian Trust Region (RTR) algorithm [60] available in the Manopt toolbox [56] to recover a local second-order critical point of Problem 3. Finally, given the local optimum Y^*, we recover the corresponding set of positions according to (8).

C. Practical considerations

Figure 2 depicts the evolution of the Riemannian Elevator solving an example problem. We also consider two alternative versions of our algorithm. The first simply solves Problem 3 directly using the Riemannian Trust Region method [56], [60].
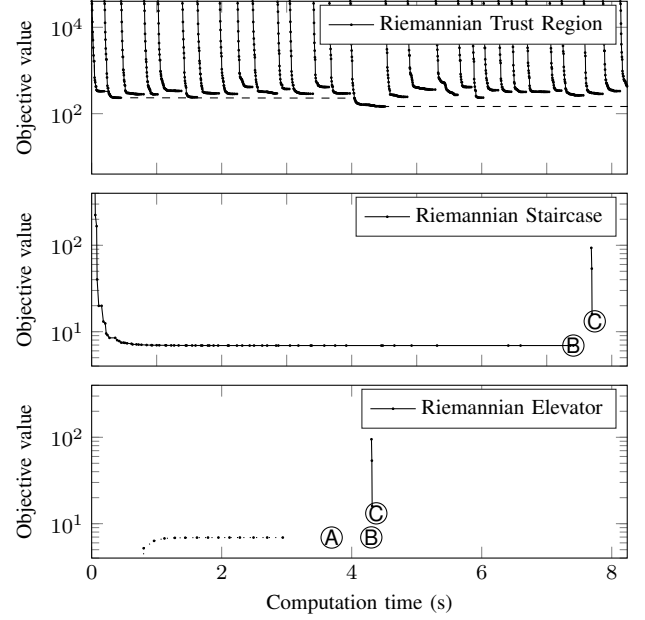

Fig. 2: Milestones in solving a `rigid-torus` problem using the Riemannian Trust Region method with successive random initializations (top), the Riemannian Staircase with *known* final rank r (middle), and the Riemannian Elevator (our method, bottom). Each trial of the randomized Riemannian Trust Region method is fast but converges to a spurious local minimum (note the vertical scaling differs between plots). The Riemannian Staircase, which in this example has the unrealistic advantage of starting at rank r specified by Theorem 4, converges to close to the globally optimal objective value quickly, but requires a strict convergence tolerance to find Y_r^* (B) that certifies the solution Y^* (C). Our method, the Riemannian Elevator, generates a certificate by solving the dual problem (A) and then quickly recovers the rank-r solution Y_r^* (B) and rank-d solution Y^* (C).

As this approach is susceptible to local minima, we consider a randomized version of the algorithm, in which the optimization is randomly initialized repeatedly and the best solution is retained. As shown in Figure 2, each optimization can be performed quickly but converges to spurious local minima with high probability. Repeating the optimization provides a modest performance benefit at a significantly increased computational cost. Furthermore, like other iterative methods, this approach provides no estimate of how close or far the best solution is to the global optimum.

The second method shown in Figure 2 is a simple version of the Riemannian Staircase in which we find Y_r^* using the Riemannian Trust Region method from a random initialization on the oblique manifold. Supposing that we do not solve the dual problem but guess the lowest rank r for which $p_r^* = p_{\mathrm{SDP}}^*$, we can generate a certificate by choosing $\Lambda^* = \mathrm{diag}(Q_2 Y_r^* Y_r^{*\top})$ and showing that $Q_2 - \Lambda^* \succeq 0$ [52]. (For the full implementation of the Riemannian Staircase in which r is unknown *a priori*, the rank is lifted successively until a certificate is found.) A relatively tight convergence

criterion is necessary to generate a certificate. As demonstrated in Figure 2, our approach of first solving Problem 6 and then solving (25) often outperforms this alternative approach, especially for smaller problems.

The Riemannian Elevator approach generates the certificate by solving the dual problem, so the high-dimensional solution Y_r^* is only needed to initialize the final step of finding Y^*. As an added benefit, the decision variable for (25) is $\bar{Z} \in \mathbb{S}_+^r \subseteq \mathbb{R}^{r \times r}$ rather than $Y_r \in (\mathcal{S}^{r-1})^m \subseteq \mathbb{R}^{m \times r}$. Given Λ^*, Problem 4 is solved quickly. The key milestones in this process are labeled in Figure 2 as follows:

Ⓐ the optimal dual value, Λ^*
Ⓑ the optimal high-dimensional solution, Y_r^*
Ⓒ the solution to the original problem, Y^*

The Riemannian Elevator's innovations on the Riemannian Staircase address the inefficiencies of visiting every intermediate rank before finding Y_r^* that is certifiably the global optimum of Problem 4. In other implementations of the Riemannian Staircase such as SE-Sync [53], this is not a significant consideration as certifiable solutions are found at a relatively low rank. Although the Riemannian Staircase finds the globally optimal solution to the relaxed problem regardless of initialization, a good initialization (for example, as odometry in pose-graph optimization) is advantageous in computing the optimization quickly. The Riemannian Elevator has the advantage of not depending on an initial guess, which can be beneficial in cases where there is no prior knowledge. The downside of the Riemannian Elevator comes in the computation cost of solving the dual problem.

V. NUMERICAL RESULTS

In this section, we assess the performance and computation time of the Riemannian Elevator (RE) algorithm for distance-based localization problems on geometric random graphs in $\mathbb{R}^3$. We implement the Riemannian Elevator in Matlab, using CVX [73] with the MOSEK solver [74] to solve Problem 6 and (18). We also use the Manopt toolbox [56] implementation of the RTR method [60] to solve Problem 3. All experiments use a MacBook Pro with M1 Pro chip and 16GB unified memory. We compare our approach to the following algorithms:

- Riemannian Trust Region (RTR): an ablation of the RE method, which includes only the final step—optimizing Problem 3 on the oblique manifold using the Manopt toolbox (initialized randomly).
- Trace-regularized SDP (EDM): an SDP method that solves Problem 2 with an additional penalty term which encourages a low-rank solution (proposed in [10] and implemented in [11]).
- Dyn-NoLips (QFO): a quartic first-order method that solves Problem 2 [37].
- SMACOF (SMA): the classic stress majorization algorithm for solving Problem 1 (implementation adapted from [16]).

A. Noisy measurements on geometric random graphs

In the first numerical study, we solve the distance-based localization problem with noisy measurements. We compare
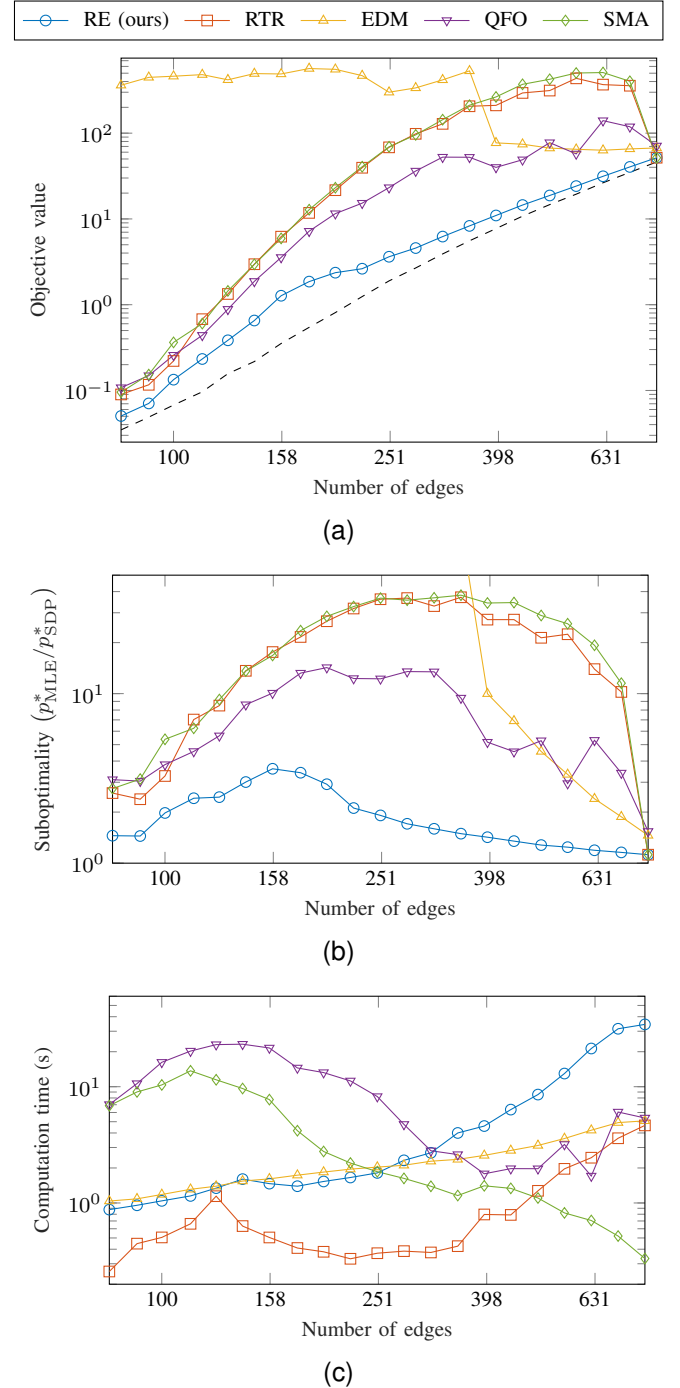

Fig. 3: Objective value (a), suboptimality (b), and computation time (c) for five distance-based localization methods for geometric random graphs with $n = 40$ vertices and $m \in [80, 780]$ edges. For each problem, vertices are uniformly distributed in a unit cube and measurement noise has a variance of 1×10^{-4}. RE provides a lower bound (dashed line) that bounds the suboptimality of the solution. Results are averaged over 50 trials.

the algorithms' performance and computation time along two dimensions: graph connectivity and graph size. To scale the graph connectivity, we fix the number of vertices as $n = 40$ and generate random geometric graphs with varying numbers
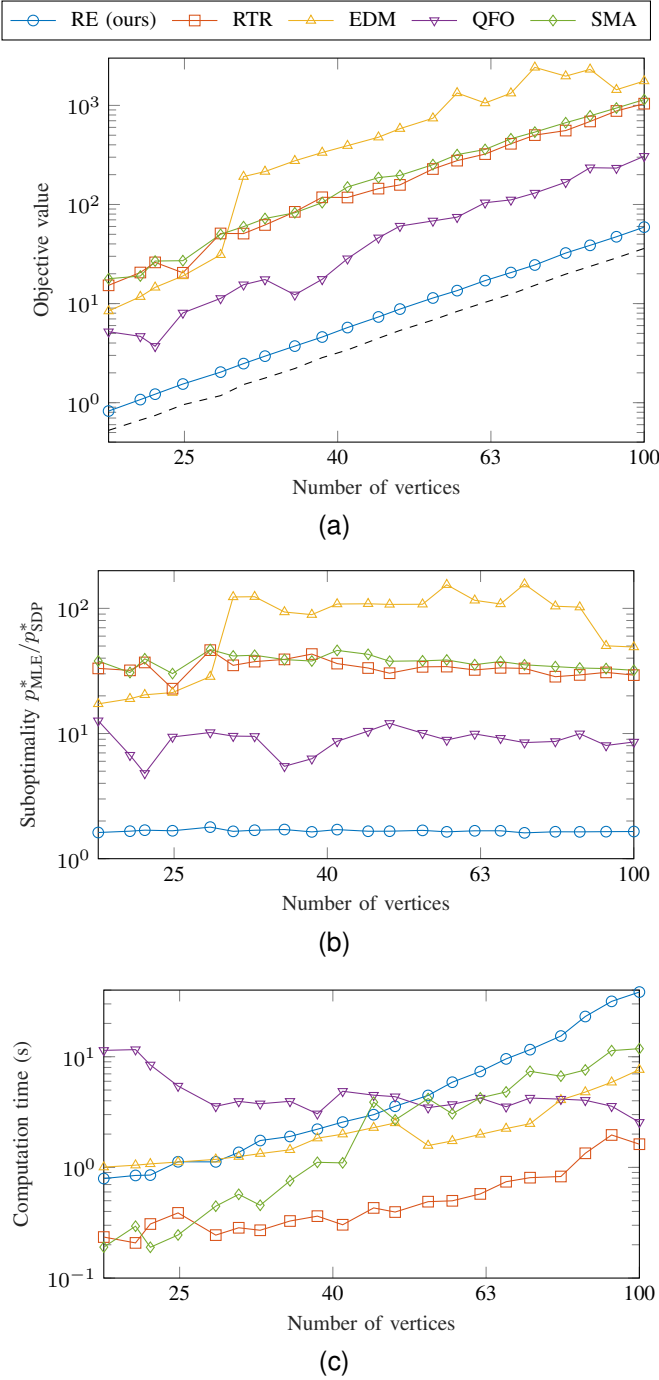
(a)

(b)

(c)

Fig. 4: Objective value (a), suboptimality (b), and computation time (c) as a function of graph size. Each ground truth configuration is a geometric random graph on the unit cube with $n \in [20, 100]$, $m = \lfloor 2n \log(n) \rfloor$, and $\sigma^2 = 1 \times 10^{-4}$. The dashed line indicates the lower bound generated by RE. Results are averaged over 50 trials.
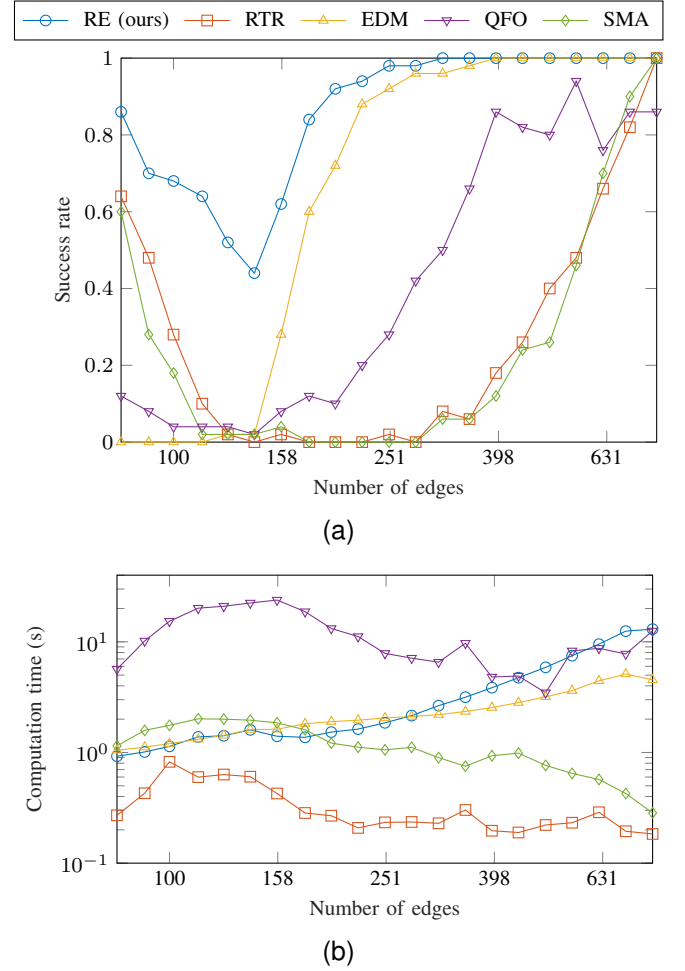

(a)

(b)

Fig. 5: Success rate (a) and computation time (b) for five distance-based localization methods with noise-free measurements on geometric random graphs with $n = 40$ vertices $m \in [80, 780]$ edges. A solution is considered correct if $p^*_{\mathrm{MLE}} < 1 \times 10^{-4}$. Results are averaged over 50 trials.

experiment, we generate a geometric random graph in $\mathbb{R}^3$ (*i.e.* n vertices embedded uniformly at random in the unit cube with edges corresponding to the m shortest true pairwise distances).

We report the performances with respect to graph connectivity in Figure 3. In addition to the objective value and computation time, we also show the maximum suboptimality of each performance, which is the ratio of the objective value and the numerical lower bound generated by RE.

Figure 4 shows the effect of the size of the graph on objective value and computation time. Given the choice of $m = \lfloor 2n \log(n) \rfloor$, the suboptimalities of the solutions are roughly constant across graph sizes, especially for RE.

B. Noise-free measurements on geometric random graphs

In the second numerical study, we consider the case of noise-free measurements. In this case, there must exist an exact feasible realization of the measurements. Therefore, the optimal objective value is always zero, whether we consider the MLE problem or the EDM completion problem. Instead of reporting the objective value of each solution, which

of edges, from sparsely connected ($m = 80$) to fully connected ($m = 780$). To scale the graph size, we consider graphs ranging from $n = 20$ to $n = 100$ vertices, with the number of edges scaling[2] with n as $m = \lfloor 2n \log(n) \rfloor$. In each

---

[2]This choice corresponds to $m = 295$ when $n = 40$. The minimum degrees of the graphs are roughly constant across graph sizes with this choice of m.

should be close to zero depending on numerical precision and the stopping conditions of each algorithm, we measure performance using "success rate". We define success rate as the fraction of solutions that achieve an objective value under a common tolerance. Figure 5 shows the success rate as a function of connectivity, replicating the conditions in Figure 3 with the exception of measurement noise. Performance is roughly constant with number of vertices, so we do not include a comparison of the methods along that dimension.

C. Discussion

The numerical results suggest that RE outperforms the other methods across a range of problems. Using RE is particularly advantageous when solving the MLE problem with noisy measurements on sparsely-connected graphs. While the reformulation of Problem 1 as Problem 3 is essential to the RE algorithm, the improved performance is not simply a result of optimizing with respect to edge directions rather than vertex positions (note that RTR and SMA perform similarly even though RTR optimizes edge directions while SMA optimizes vertex positions directly).

In scaling the connectivity of the underlying graphs, Figure 5a suggests three regimes in the distance-based localization problem. For low graph connectivity, iterative methods like SMA and RTR are likely to recover the global solution due to the non-rigidity of the graph. RE recovers the solutions with high probability as well, though a good initialization is less beneficial when the lack of rigidity makes many initializations suitable.

For problems with high connectivity, in which the underlying graph is nearly fully connected, the EDM completion methods tend to work well. In this case, the measured incomplete distance matrix is close to a feasible distance matrix. When the underlying graph is fully connected case, the vertex positions can be extracted from the EDM in closed form [11]. Distance-based localization on a fully connected graph with exact distance measurements is an easy problem, and all methods perform well, as expected.

The most challenging regime lies in the middle, for which many local minima exist. In this case, RE recovers a high-dimensional solution to the relaxed problem, but the projection operation initializes Y^0 close to a spurious local minimum. Despite the challenge of solving the problem for moderately-connected graphs, RE still substantially outperforms the other methods. This trend is also seen in Figure 3b, where the suboptimality ratios of the RE solutions are orders of magnitude smaller than the alternatives. We do not anticipate that any computationally efficient algorithm always performs well in this regime due to the NP-hardness of Problem 1.

While RE finds better solutions than the alternative algorithms in all of the problems that we tested, several algorithms are consistently faster. We expect RTR to be faster than RE, as it consists of only the final step of RE. Other methods are also faster than RE depending on the connectivity and size of the problem (interestingly, no method besides RTR dominates RE across all problems in our implementation). We note that a principal difference of RE and RTR from the other

methods is that the size of the decision variable scales with m rather than n. A question for future work is whether RE can be parallelized in a way that improves computational efficiency for larger problems. Alternatively, removing the certification steps of RE would also decrease computation time.

One advantage of implementing RE is its low sensitivity to parameter choices. In fact, the only parameters in our implementation of RE are the precision of the CVX solver and the stopping condition in Manopt. On the other hand, as a first-order method, the performance of QFO (Dyn-NoLips) depends substantially on parameter choices. Similarly, EDM (trace-regularized SDP) requires the selection of a regularization constant, for which the optimal value will depend on the particular problem. In our comparisons, we hand-tuned QFO and EDM to fixed settings that achieve relatively strong performance across the range of problems, though problem-specific parameter tuning would likely improve their results.

VI. CONCLUSION

We proposed the Riemannian Elevator method, which we use to solve the distance-based localization problem. Our novel reformulation of the MLE problem facilitates what is to our knowledge the first SDP approach applied to the *unsquared* distance-based localization problem (as opposed to EDM completion, which penalizes errors between squared distances and for which many SDP methods exist). This approach addresses several challenges in distance-based localization by recovering a good solution with high probability and by additionally generating a numerical bound on the solution's maximum suboptimality. The numerical results in Section V demonstrate the suitability of this approach across a range of problems.

While we specifically address distance-based localization in this paper, the Riemannian Elevator is suitable for a broader class of problems, namely homogeneous quadratic programs constrained to the oblique manifold. Under minor modificiations, this approach should also apply to general quadratic programs on the Steifel manifold, as does the Riemannian Staircase method that inspired our approach. The Riemannian Elevator circumvents the successive lifting steps of the Riemannian Staircase at the expense of solving the dual problem once. Of general interest is the insight that given the dual, finding a primal solution to the SDP is a feasibility problem (Lemma 8). While the Riemannian Elevator outperforms other methods for a range of problems, it does not avoid the limitations of solving large-scale SDPs. A topic of future study will be in distributing the approach for improved computational efficiency.

REFERENCES

[1] W. Shule, C. M. Almansa, J. P. Queralta, Z. Zou, and T. Westerlund, "UWB-based localization for multi-UAV systems and collaborative heterogeneous multi-robot systems," *Procedia Computer Science*, vol. 175, pp. 357–364, 2020.

[2] G. Han, C. Zhang, L. Shu, and J. J. Rodrigues, "Impacts of deployment strategies on localization performance in underwater acoustic sensor networks," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1725–1733, 2014.

[3] A. K. Paul and T. Sato, "Localization in wireless sensor networks: A survey on algorithms, measurement techniques, applications and challenges," *Journal of Sensor and Actuator Networks*, vol. 6, no. 4, p. 24, 2017.

[4] C. Lavor, L. Liberti, N. Maculan, and A. Mucherino, "The discretizable molecular distance geometry problem," *Computational Optimization and Applications*, vol. 52, no. 1, pp. 115–146, 2012.

[5] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Convex algorithms and limitations," in *2013 IEEE International Symposium on Information Theory*, IEEE, 2013, pp. 1022–1026.

[6] S. X.-Y. Ni, M.-C. Yue, K.-F. Cheung, and A. M.-C. So, "Phase retrieval via sensor network localization," *Journal of the Operations Research Society of China*, vol. 7, no. 1, pp. 127–146, 2019.

[7] L. Liberti, "Distance geometry and data science," *Top*, vol. 28, no. 2, pp. 271–339, 2020.

[8] J. B. Saxe, "Embeddability of weighted graphs in k-space is strongly np-hard," in *Proc. of 17th Allerton Conference in Communications, Control and Computing, Monticello, IL*, 1979, pp. 480–489.

[9] J. Aspnes, T. Eren, D. K. Goldenberg, *et al.*, "A theory of network localization," *IEEE Transactions on Mobile Computing*, vol. 5, no. 12, pp. 1663–1678, 2006.

[10] P. Biswas, T.-C. Lian, T.-C. Wang, and Y. Ye, "Semidefinite programming based algorithms for sensor network localization," *ACM Transactions on Sensor Networks (TOSN)*, vol. 2, no. 2, pp. 188–220, 2006.

[11] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: Essential theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 12–30, 2015.

[12] W. S. Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

[13] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[14] J. De Leeuw, "Applications of convex analysis to multidimensional scaling," *Recent Developments in Statistics*, 1977.

[15] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2016.

[16] J. De Leeuw and P. Mair, "Multidimensional scaling using majorization: Smacof in r," *Journal of Statistical Software*, vol. 31, pp. 1–30, 2009.

[17] E. R. Gansner, Y. Koren, and S. North, "Graph drawing by stress majorization," in *International Symposium on Graph Drawing*, Springer, 2004, pp. 239–250.

[18] Y. Wang, Y. Wang, Y. Sun, *et al.*, "Revisiting stress majorization as a unified framework for interactive constrained graph visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 489–499, 2017.

[19] P. Oguz-Ekim, J. P. Gomes, J. Xavier, and P. Oliveira, "Robust localization of nodes and time-recursive tracking in sensor networks using noisy range measurements," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3930–3942, 2011.

[20] S. Korkmaz and A.-J. van der Veen, "Robust localization in sensor networkswith iterative majorization techniques," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2009, pp. 2049–2052.

[21] G. M. Crippen, T. F. Havel, *et al.*, *Distance geometry and molecular conformation*. Research Studies Press Taunton, 1988, vol. 74.

[22] W. Glunt, T. L. Hayden, and M. Raydan, "Molecular conformations from distance matrices," *Journal of Computational Chemistry*, vol. 14, no. 1, pp. 114–120, 1993.

[23] M. W. Trosset, "Applications of multidimensional scaling to molecular conformation," 1997.

[24] C.-s. Wang, "Determining molecular conformation from distance or density data," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.

[25] G. N. Simm and J. M. Hernández-Lobato, "A generative model for molecular distance geometry," *arXiv preprint arXiv:1909.11459*, 2019.

[26] G. Simm, R. Pinsler, and J. M. Hernández-Lobato, "Reinforcement learning for molecular design guided by quantum mechanics," in *International Conference on Machine Learning*, PMLR, 2020, pp. 8959–8969.

[27] M. Xu, S. Luo, Y. Bengio, J. Peng, and J. Tang, "Learning neural generative dynamics for molecular conformation generation," *arXiv preprint arXiv:2102.10240*, 2021.

[28] Y. Takane, F. W. Young, and J. De Leeuw, "Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features," *Psychometrika*, vol. 42, no. 1, pp. 7–67, 1977.

[29] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, "Euclidean distance geometry and applications," *SIAM review*, vol. 56, no. 1, pp. 3–69, 2014.

[30] L. Liberti and C. Lavor, "Six mathematical gems from the history of distance geometry," *International Transactions in Operational Research*, vol. 23, no. 5, pp. 897–920, 2016.

[31] L. Liberti and C. Lavor, *Euclidean Distance Geometry*. Springer, 2017.

[32] A. Y. Alfakih, *Euclidean distance matrices and their applications in rigidity theory*. Springer, 2018.

[33] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Mathematical Programming*, vol. 95, no. 2, pp. 329–357, 2003.

[34] S. Burer and R. D. Monteiro, "Local minima and convergence in low-rank semidefinite programming," *Mathematical Programming*, vol. 103, no. 3, pp. 427–444, 2005.

[35] A. J. Kearsley, R. A. Tapia, and M. W. Trosset, "The solution of the metric stress and sstress problems in multidimensional scaling using newton's method," Tech. Rep., 1995.

[36] J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan, and B. Recht, "First-order methods almost always avoid strict saddle points," *Mathematical Programming*, vol. 176, no. 1, pp. 311–337, 2019.

[37] R.-A. Dragomir, A. d'Aspremont, and J. Bolte, "Quartic first-order methods for low-rank minimization," *Journal of Optimization Theory and Applications*, vol. 189, no. 2, pp. 341–363, 2021.

[38] R.-A. Dragomir, A. B. Taylor, A. d'Aspremont, and J. Bolte, "Optimal complexity and certification of Bregman first-order methods," *Mathematical Programming*, pp. 1–43, 2021.

[39] A. M.-C. So and Y. Ye, "Theory of semidefinite programming for sensor network localization," *Mathematical Programming*, vol. 109, no. 2, pp. 367–384, 2007.

[40] P. Tseng, "Second-order cone programming relaxation of sensor network localization," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 156–185, 2007.

[41] R. M. Vaghefi, M. R. Gholami, R. M. Buehrer, and E. G. Strom, "Cooperative received signal strength-based sensor localization with unknown transmit powers," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1389–1403, 2012.

[42] B. Recht, "A simpler approach to matrix completion.," *Journal of Machine Learning Research*, vol. 12, no. 12, 2011.

[43] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[44] B. Vandereycken, "Low-rank matrix completion by Riemannian optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1214–1236, 2013.

[45] K. Wei, J.-F. Cai, T. F. Chan, and S. Leung, "Guarantees of Riemannian optimization for low rank matrix recovery," *SIAM Journal on Matrix Analysis and Applications*, vol. 37, no. 3, pp. 1198–1222, 2016.

[46] B. Mishra, G. Meyer, and R. Sepulchre, "Low-rank optimization for distance matrix completion," in *2011 50th IEEE Conference on Decision and Control and European Control Conference*, IEEE, 2011, pp. 4455–4460.

[47] P. Tabaghi, I. Dokmanić, and M. Vetterli, "Kinetic Euclidean distance matrices," *IEEE Transactions on Signal Processing*, vol. 68, pp. 452–465, 2019.

[48] S. Li, H. Sun, and H. Esmaiel, "Underwater TDOA acoustical location based on majorization-minimization optimization," *Sensors*, vol. 20, no. 16, p. 4457, 2020.

[49] F. Marić, M. Giamou, I. Petrović, and J. Kelly, "Inverse kinematics as low-rank Euclidean distance matrix completion," *arXiv preprint arXiv:2011.04850*, 2020.

[50] F. Marić, M. Giamou, A. W. Hall, S. Khoubyarian, I. Petrović, and J. Kelly, "Riemannian optimization for distance-geometric inverse kinematics," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1703–1722, 2021.

[51] M. Giamou, F. Marić, D. M. Rosen, *et al.*, "Convex iteration for distance-geometric inverse kinematics," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1952–1959, 2022.

[52] N. Boumal, "A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints," *arXiv preprint arXiv:1506.00575*, 2015.

[53] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, "SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group," *The International Journal of Robotics Research*, vol. 38, no. 2-3, pp. 95–125, 2019.

[54] A. S. Bandeira, N. Boumal, and V. Voroninski, "On the low-rank approach for semidefinite programs arising in synchronization and community detection," in *Conference on Learning Theory*, PMLR, 2016, pp. 361–382.

[55] J. Briales and J. Gonzalez-Jimenez, "Cartan-sync: Fast and global SE(d)-synchronization," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2127–2134, 2017.

[56] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1455–1459, 2014.

[57] N. Boumal, V. Voroninski, and A. Bandeira, "The non-convex Burer-Monteiro approach works on smooth semidefinite programs," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[58] N. Boumal, "An introduction to optimization on smooth manifolds," *Available online*, vol. 3, 2020.

[59] N. Boumal and P.-A. Absil, "RTRMC: A Riemannian trust-region method for low-rank matrix completion," *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[60] P.-A. Absil, C. G. Baker, and K. A. Gallivan, "Trust-region methods on Riemannian manifolds," *Foundations of Computational Mathematics*, vol. 7, no. 3, pp. 303–330, 2007.

[61] Y. Chen, Y. Chi, J. Fan, C. Ma, and Y. Yan, "Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization," *SIAM Journal on Optimization*, vol. 30, no. 4, pp. 3098–3121, 2020.

[62] A. S. Bandeira, N. Boumal, and A. Singer, "Tightness of the maximum likelihood semidefinite relaxation for angular synchronization," *Mathematical Programming*, vol. 163, no. 1, pp. 145–167, 2017.

[63] R. A. Horn and F. Zhang, "Basic properties of the Schur complement," in *The Schur Complement and Its Applications*, Springer, 2005, pp. 17–46.

[64] J. Gallier, "The Schur complement and symmetric positive semidefinite (and definite) matrices," 2020.

[65] A. I. Barvinok, "Problems of distance geometry and convex properties of quadratic maps," *Discrete & Computational Geometry*, vol. 13, no. 2, pp. 189–202, 1995.

[66] G. Pataki, "On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues," *Mathematics of Operations Research*, vol. 23, no. 2, pp. 339–358, 1998.

[67] A. Lemon, A. M.-C. So, Y. Ye, *et al.*, "Low-rank semidefinite programming: Theory and applications," *Foundations and Trends® in Optimization*, vol. 2, no. 1-2, pp. 1–156, 2016.

[68] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, 1996.

[69] D. Hong, H. Lee, and A. Wei, "Optimal solutions and ranks in the max-cut SDP," *arXiv preprint arXiv:2109.02238*, 2021.

[70] W. W. Barrett, D. Olesky, and P. Van den Driessche, "A note on eigenvalues of fixed rank perturbations of diagonal matrices," *Linear and Multilinear Algebra*, vol. 30, no. 1-2, pp. 13–16, 1991.

[71] M. T. Chu, "Inverse eigenvalue problems," *SIAM Review*, vol. 40, no. 1, pp. 1–39, 1998.

[72] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the ACM (JACM)*, vol. 42, no. 6, pp. 1115–1145, 1995.

[73] M. Grant and S. Boyd, *Cvx: Matlab software for disciplined convex programming, version 2.1*, 2014.

[74] MOSEK ApS, *The MOSEK optimization toolbox for Matlab manual. version 9.0.* 2019. [Online]. Available: http://docs.mosek.com/9.0/toolbox/index.html.