Sensor Network Localization from Local Connectivity : Performance Analysis for the MDS-MAP Algorithm

Sewoong Oh, Amin Karbasi, and Andrea Montanari

August 27, 2009

Abstract

Sensor localization from only connectivity information is a highly challenging problem. To this end, our result for the first time establishes an analytic bound on the performance of the popular MDS-MAP algorithm based on multidimensional scaling. For a network consisting of n sensors positioned randomly on a unit square and a given radio range $r = o(1)$, we show that resulting error is bounded, decreasing at a rate that is inversely proportional to r, when only connectivity information is given. The same bound holds for the *range-based model*, when we have an approximate measurements for the distances, and the same algorithm can be applied without any modification.

1   Introduction

Localization plays an important role in wireless sensor networks when the positions of the nodes are not provided in advance. One way to acquire the positions is by equipping all the sensors with a global positioning system (GPS). Clearly, this adds considerable cost to the system. As an alternative, many research efforts have focused on creating algorithms that can derive positions of sensors based on basic local information such as proximity (which nodes are within communication range of each others) or local distances (pairwise distances between neighboring sensors).

The problem is solvable, meaning that it has a unique set of coordinates satisfying the given local information, only if there are enough constraints. Note that based on local information any solution is unique only up to rigid transformations (rotations, reflections and translations).

The simplest of such algorithms assumes that all pairwise distances are known. However, it is well known that in practical scenarios such information is unavailable and far away sensors (the ones outside the communication range) cannot obtain their pairwise distances. Many algorithms have been proposed to resolve this issue by using heuristic approximations to the missing distances where the success of them has mostly been measured experimentally.

Of the same interest and complementary in nature, however, are the theoretical guarantees associated with the performance of the existing methods. Such analytical bounds on the performance of localization algorithms can provide answers to practical questions: for example, how large should the radio range be in order to get the error within a threshold? With this motivation in mind, our work takes a forward step in this direction. In particular, our analysis focuses on providing a bound on the performance of the popular MDS-MAP [SRZF03] algorithm when applied to sensor localization from only connectivity information, which is a highly challenging problem. We prove for the first time that using MDS-MAP, we are able to localize sensors up to a bounded error in a connected network where most of the pairwise distances are missing and only local connectivity information is given.

1

1.1 Related work

A great deal of work over the past few years has focused on the sensor localization problem. A detailed survey of this area is provided in [JH01].

In the case when all pairwise distances are known, the coordinates can be derived by using a classical method known as multidimensional scaling (MDS) [SRZF03, JZ04]. The underlying principle of the MDS is to convert distances into an inner product matrix, whose eigenvectors are the unknown coordinates. In practice, MDS tolerates errors gracefully due to the overdetermined nature of the solution. However, when most pairwise distances are missing, the problem of finding the unknown coordinates becomes more challenging. Two types of practical solutions to this problem have been proposed. The first group consists of algorithms who try first to estimate the missing entries of the distance matrix and then apply MDS to the reconstructed distance matrix to find the coordinates of the sensors. MDS-MAP [SRZF03, SRZF04] can be mentioned as a well known example of this class where it computes the shortest paths between all pairs of nodes in order to approximate the missing entries of the distance matrix. The algorithms in the second group mainly consider the sensor localization as a non-convex optimization problem and directly estimate the coordinates of sensors. A famous example of this type is a relaxation to semidefinite programming (SDP)[BY04].

The performance of these practical algorithms are invariably measured through simulations and little is known about the theoretical analysis supporting their results. One exception is the work done in [DJMI$^+$] where they use matrix completion methods as a mean to reconstruct the distance matrix. The main contribution of their paper is that they are able to provably localize the sensors up to a bounded error. However, their analysis is based on a number of strong assumptions. First, they assume that even far away sensors have non-zero probability of detecting their distances. Second, the algorithm explicitly requires the knowledge of detection probabilities between all pairs. Third, their theorem only works when the average degree of the network (i.e., the average number of nodes detected by each sensor) grows linearly with the number of sensors in the network.

Our work has a similar flavor as [DJMI$^+$], namely we provide theoretical guarantee that backs up experimental results. We use shortest paths (the same way it is used in MDS-MAP) as our primary guess for the missing entries in the distance matrix and apply MDS to find the topology of the network. In contrast to [DJMI$^+$], we require weaker assumptions for our result. More specifically, we assume that only neighboring sensors have information about each other and that only connectivity information is known. Furthermore, the knowledge of detection probabilities plays no role in our analysis or the algorithm. And last, in our analysis we assume that the average degree grows logarithmically with the number of sensors not linearly, which results in needing much less revealed entries in the distance matrix. In particular, the last condition is quite realistic: if the average degree grows any slower then the network is not even connected. Due to the fact that the shortest paths algorithm works for both rage-free and range-aware cases, our analysis includes both and provides the first error bounds on the performance of MDS-MAP.

The organization of this paper is as follows. Section 2 describes the state of the art algorithm known as MDS-MAP and states the main result on the performance of MDS-MAP. In Section 3, we provide detailed proof of the main theorem.

2   Localization using multidimensional scaling

In this section, we first define the basic mathematical model for the sensors and the measurements and state preliminary facts on multidimensional scaling, a technique to derive sensor configuration from pairwise distances [SRZF03]. We will then describe the algorithm and give a bound on the error between X and $\widehat{X}$, the estimate returned by the algorithm.

2.1 Model definition

For the sensor localization problem, we consider the random geometric graph model $G(n, r) = (V, E, P)$ where V is a set of n vertices that correspond to n sensor nodes which are distributed uniformly at random within the d-dimensional hypercube $[0, 1]^d$, $E \subseteq V \times V$ is a set of undirected edges that connect pairs of sensors which are close to each other, and $P : E \to \mathbb{R}^+$ is a non-negative real-valued function. We consider the function P as a mapping from a pair of connected nodes (i, j) to the approximate measurement for the distance between i and j, which we call the proximity measurement. Let $|| \cdot ||$ be the Euclidean norm on $\mathbb{R}^d$ and let r be some positive parameter. Given a set of random positions of n sensors $\mathcal{X} = \{x_1, \ldots, x_n\}$, where x_i is the position of sensor i in d-dimensional space, a common model for the random geometric graph is the disc model where node i and j are connected if the Euclidean distance $d_{i,j} = ||x_i - x_j||$ is less than r. More formally,

$$(i, j) \in E \Leftrightarrow d_{i,j} \leq r .$$

To each edge $(i, j) \in E$, we associate the proximity measurement $P_{i,j}$ between sensors i and j, which is a function of the distance $d_{i,j}$ and random noise. In an ideal case when our measurements are exact, we have $P_{i,j} = d_{i,j}$. On the other extreme, when we are given only network connectivity information and no distance information, we have $P_{i,j} = r$ for all $(i, j) \in E$.

The sensor localization algorithms can be classified into two different categories. In the first, only connectivity information is available (connectivity-based). In the second category, however, approximate measurement for the distance between neighboring sensors are available, perhaps with limited accuracy (range-based). For the *connectivity-based model*, which is alternatively also known as the range-free model, only the connectivity information is available. Formally,

$$P_{i,j} = \begin{cases} r & \text{if } (i, j) \in E, \\ * & \text{otherwise,} \end{cases}$$

where a $*$ denotes that $d_{i,j} > r$.

For the *range-based model*, which is also known as the range-aware model, the distance measurement is available but may be corrupted by noise or have limited accuracy.

$$P_{i,j} = \begin{cases} [d_{i,j} + z_{i,j}]_+ & \text{if } (i, j) \in E, \\ * & \text{otherwise,} \end{cases}$$

where $z_{i,j}$ models the measurement noise (in the noiseless case $z_{i,j} = 0$), possibly a function of the distance $d_{i,j}$, and $[a]_+ \equiv \max\{0, a\}$. Common examples are the additive Gaussian noise model, where the $z_{i,j}$'s are i.i.d. Gaussian random variables, and the multiplicative noise model, where $P_{i,j} = [(1 + N_{i,j})d_{i,j}]_+$, for independent Gaussian random variables $N_{i,j}$'s.

In the following, let D denote the $n \times n$ squared distance matrix where $D_{ij} = d_{i,j}^2$ and X denote the $n \times d$ position matrix where the ith row corresponds to x_i, the d-dimensional position vector of sensor i. Given the graph $G(n, r) = (V, E, P)$ with associated proximity measurements for each edges in E, the goal of *sensor network localization* is to find an embedding of the nodes that best fits the measured proximity between all the pairs in E. The notion of embedding, or configuration, will be made clear in the following sections.

The main contribution of this paper is that we provide, for the first time, the performance bound for the MDS-MAP algorithm when only the connectivity information is available as in the *connectivity-based model*. However, the algorithm can be readily applied to the *range-based model* without any modification. Further, given $G(V, E, P)$ from from the *range-based model* we can easily produce $G'(V, E, P')$ where $P'_{i,j} = r$ for all $(i, j) \in E$. This implies that the performance of MDS-MAP under the *range-based model* is also bounded by the same bound.

2.2 Multidimensional scaling (MDS)

Multidimensional scaling (MDS) is a technique used in finding the configuration of objects in a low dimensional space such that the measured pairwise distances are preserved. If all the pairwise distances are measured without error then a naive application of MDS exactly recovers the configuration of sensors [SRZF03].

Algorithm : Classical Metric MDS [SRZF03]
Input: dimension d, estimated distance matrix M
Output: estimated positions $\mathsf{MDS}_d(M)$
1:   Compute $(-1/2)LML$, where $L = \mathbb{I}_n - (1/n)\mathbb{1}_n\mathbb{1}_n^T$;
2:   Compute the best rank-d approximation $U_d\Sigma_d U_d^T$ of $(-1/2)LML$;
3:   Return $\mathsf{MDS}_d(M) \equiv U_d\Sigma_d^{1/2}$.

There are many types of MDS techniques, but, throughout this paper, whenever we say MDS we refer to the classical metric MDS, which is defined as follows. Let L be an $n \times n$ symmetric matrix such that $L = \mathbb{I}_n - (1/n)\mathbb{1}_n\mathbb{1}_n^T$, where $\mathbb{1}_n \in \mathbb{R}^n$ is the all ones vector and $\mathbb{I}_n$ is the $n \times n$ identity matrix. Let $\mathsf{MDS}_d(D)$ denote the $n \times d$ matrix returned by MDS when applied to the squared distance matrix D. Then, in formula, given the singular value decomposition (SVD) of a symmetric and positive definite matrix $(-1/2)LDL$ as $(-1/2)LDL = U\Sigma U^T$,

$$\mathsf{MDS}_d(D) \equiv U_d\Sigma_d^{1/2} \ ,$$

where U_d denotes the $n \times d$ left singular matrix corresponding to the d largest singular values and Σ_d denotes the $d \times d$ diagonal matrix with d largest singular values in the diagonal. This is also known as the MDSLOCALIZE algorithm in [DJMI+]. Note that since the columns of U are orthogonal to $\mathbb{1}_n$ by construction, it follow that $L \cdot \mathsf{MDS}_d(D) = \mathsf{MDS}_d(D)$.

It can be easily shown that when MDS is applied to the correct squared distance matrix without noise, the configuration of sensors are exactly recovered [DJMI+]. This follows from the following equality

$$-\frac{1}{2}LDL = LXX^TL \ . \tag{1}$$

Note that we only get the configuration and not the absolute positions, in the sense that $\mathsf{MDS}_d(D)$ is one version of infinitely many solutions that matches the distance measurements D. Intuitively, it is clear that the pairwise distances are invariant to a rigid transformation (a combination of rotation, reflection and translation) of the positions X, and therefore there are multiple incidents of X that result in the same D. For future use, we introduce a formal definition of rigid transformation and related terms.

Denote by $\mathsf{O}(d)$ the orthogonal group of $d \times d$ matrices. A set of sensor positions $Y \in \mathbb{R}^{n \times d}$ is a rigid transform of X, if there exists a d-dimensional shift vector s and an orthogonal matrix $Q \in \mathsf{O}(d)$ such that

$$Y = XQ + \mathbb{1}_n s^T \ .$$

Y should be interpreted as a result of first rotating (and/or reflecting) sensors in position X by Q and then adding a shift by s. Similarly, when we say two position matrices X and Y are equal up to a rigid transformation, we mean that there exists a rotation Q and a shift s such that $Y = XQ + \mathbb{1}_n s^T$. Also, we say a function $f(X)$ is invariant under rigid transformation if and only if for all X and Y that are

equal up to a rigid transformation we have $f(X) = f(Y)$. Under these definitions, it is clear that D is invariant under rigid transformation, since for all (i, j), $D_{ij} = ||x_i - x_j||^2 = ||(x_i Q + s^T) - (x_j Q + s^T)||^2$, for any $Q \in \mathsf{O}(d)$ and $s \in \mathbb{R}^d$.

Note that when solving a localization problem, there are two possible outputs : a relative map and an absolute map. The task of finding a relative map is to find a configuration of sensors that have the same neighbor relationships as the underlying graph G. In the following we use the terms configuration, embedding, and relative map interchangeably. The task of finding an absolute map is to determine the absolute geographic coordinates of all sensors. In this paper, we are interested in presenting an algorithm MDS-MAP, that finds a configuration that best fits the proximity measurements, and providing an analytical bound on the error between the estimated configuration and the correct configuration.

While MDS works perfectly when D is available, in practice not all proximity measurements are available because of the limited radio range r. The next section describes the MDS-MAP algorithm introduced in [SRZF03], where we apply the shortest paths algorithm on the graph $G = (V, E, P)$ to get a good estimate of the unknown proximity measurements.

2.3   Algorithm

Based on MDS, MDS-MAP consists of two steps :

Algorithm : MDS-MAP [SRZF03]
Input: dimension d, graph $G = (V, E, P)$
Output: estimation $\widehat{X}$
1:   Compute the shortest paths, and let $\widehat{D}$ be the squared shortest paths matrix;
2:   Apply MDS to $\widehat{D}$, and let $\widehat{X}$ be the output.

The original MDS-MAP algorithm introduced in [SRZF03] has additional post-processing step to fit the given configuration to an absolute map using a small number (typically $d+1$) of special nodes with known positions, which are called anchors. However, the focus of this paper is to give a bound on the error between the relative map found by the algorithm and the correct configuration. Hence, the last step, which does not improve the performance, is omitted here.

Shortest paths. The shortest path between node i and j in graph $G = (V, E, P)$ is defined as a path between two nodes such that the sum of the proximity measures of its constituent edges is minimized. Let $\hat{d}_{i,j}$ be the computed shortest path between node i and j. Then, the squared shortest paths matrix $\widehat{D} \in \mathbb{R}^{n \times n}$ is defined as

$$\widehat{D}_{ij} = \begin{cases} 0 & \text{if } i = j \text{ ,} \\ \hat{d}_{i,j}^2 & \text{otherwise .} \end{cases}$$

Note that $\widehat{D}$ is well defined only if the graph G is connected. If G is not connected, there are multiple configurations resulting in the same observed proximity measures and global localization is not possible. In the unit square, assuming sensor positions are drawn uniformly at random as define in the previous section, the graph is connected, with high probability, if $\pi r^2 > (\log n + c_n)/n$ for $c_n \to \infty$ [GK98]. A similar condition can be derived for generic d-dimensions as $C_d r^d > (\log n + c_n)/n$, where $C_d \leq \pi$ is a constant that depends on d. Hence, we focus in the regime where $r = (\alpha \log n/n)^{1/d}$ for some positive constant α.

As will be shown in Lemma 3.1, the key observation of the shortest paths step is that the estimation is guaranteed to be arbitrarily close to the correct distance for large enough radio range r

and large enough n. Moreover, the all-pairs shortest paths problem has an efficient algorithm whose complexity is $O(n^2 \log n + n|E|)$ [Joh77]. For $r = (\alpha \log n/n)^{1/d}$ with constant α, the graph is sparse with $|E| = O(n \log n)$, whence the complexity is $O(n^2 \log n)$.

Multidimensional scaling. In step 2, we apply the MDS to $\widehat{D}$ to get a good estimate of X. As explained in 2.2, we compute $\widehat{X} = \mathsf{MDS}_d(\widehat{D})$. The main step in MDS is singular value decomposition of a dense matrix $\widehat{D}$, which has complexity of $O(n^3)$.

2.4 Main results

Our main result establishes that MDS-MAP achieves an arbitrarily small error from a radio range $r = (\alpha \log n/n)^{1/d}$ with a large enough constant α, when we have only the connectivity information as in the case of the *connectivity-based model*. The same bound holds immediately for the *range-based model*, when we have an approximate measurements for the distances, and the same algorithm can be applied without any modification.

Let $\widehat{X}$ denote an $n \times d$ estimation for X with estimated position for node i in the ith row. Then, we need to define a metric for the distance between the original position matrix X and the estimation $\widehat{X}$ which is invariant under rigid transformation of X or $\widehat{X}$.

Define $L \equiv \mathbb{I}_n - (1/n)\mathbb{1}_n\mathbb{1}_n^T$ as in the MDS algorithm. L is an $n \times n$ rank $n-1$ symmetric matrix which eliminates the contributions of the translation, in the sense that $LX = L(X + \mathbb{1}s^T)$ for all $s \in R^d$. Note that L has the following nice properties : (1) LXX^TL is invariant under rigid transformation. (2) $LXX^TL = L\widehat{X}\widehat{X}^TL$ implies that X and $\widehat{X}$ are equal up to a rigid transformation. This naturally defines following distance between X and $\widehat{X}$.

$$d_1(X, \widehat{X}) = \frac{1}{n}\big|\big|LXX^TL - L\widehat{X}\widehat{X}^TL\big|\big|_F \,, \tag{2}$$

where $||\cdot||_F$ denotes the Frobenius norm. Notice that the factor $(1/n)$ corresponds to the usual normalization by the number of entries in the summation, and $(1/\sqrt{d})$ corresponds to the maximum distance between any two nodes in the $[0, 1]^d$ hypercube. Indeed this distance is invariant to rigid transformation of X and $\widehat{X}$. Furthermore, $d_1(X, \widehat{X}) = 0$ implies that X and $\widehat{X}$ are equal up to a rigid transformation.

Define

$$r_0 \equiv \frac{24\,d^2}{\sqrt{10}}\left(\frac{\log n}{n}\right)^{\frac{1}{d}} . \tag{3}$$

Theorem 2.1. *Assume n sensors distributed uniformly at random in the $[0, 1]^d$ hypercube, for a bounded dimension $d = O(1)$. For a given radio range r, a corresponding graph $G(V, E, P)$ is defined according to the* connectivity-based model. *Then, with high probability, the distance between the estimate $\widehat{X}$ given by* MDS-MAP *and the correct position matrix X is bounded by*

$$d_1(X, \widehat{X}) \leq \frac{r_0}{r} + O(r) \,, \tag{4}$$

for $r > r_0$, where $d_1(\cdot)$ is defined in Eq. (2) and r_0 in Eq. (3).

The proof is provided in Section 3. As described in the previous section, we are interested in the regime where $r = (\alpha \log n/n)^{1/d}$ for some constant α. Given a small positive constant δ, this implies that MDS-MAP is guaranteed to produce estimated positions that satisfy $d_1(X, \widehat{X}) \leq \delta$ with a large enough constant α and large enough n.

2.5 Some technical remark and notations

It is natural to view a set of positions X as a point on the manifold $\mathsf{M}(n,d)$, where a point on $\mathsf{M}(n,d)$ is defined as the equivalence class of an $n \times d$ matrix A :

$$[A] = \{AQ + \mathbb{1}_n s^T \ : \ Q \in \mathsf{O}(d) \,, s \in \mathbb{R}^d\} \,,$$

which represent a set of all matrices that are equal up to a rigid transformation. To represent a point in $\mathsf{M}(n,d)$, we will use an explicit representative of this form.

Under this definition, it is clear that the distance $d_1(X, \widehat{X})$ should be viewed as defined on a manifold $\mathsf{M}(n,d)$, since $d_1(X, \widehat{X})$ depends on the matrices $(X, \widehat{X})$ only through their equivalence classes $([X], [\widehat{X}])$. Therefore, we interpret it as a function defined on the manifold $\mathsf{M}(n,d) \times \mathsf{M}(n,d)$.

$$
\begin{aligned}
d_1 \ : \ \mathsf{M}(n,d) \times \mathsf{M}(n,d) \ &\to \ \mathbb{R} \,, \\
([X], [Y]) \ &\mapsto \ d_1(X, Y) \,.
\end{aligned}
$$

In the following, whenever we write that a property $\mathcal{A}$ holds with high probability (w.h.p.), we mean that $\mathbb{P}(\mathcal{A})$ approaches 1 as the number of sensors n goes to infinity. Given a matrix $A \in \mathbb{R}^{m \times n}$, the spectral norm of A is denoted by $||A||_2$, and the Frobenius norm is denoted by $||A||_F$. For a vector $a \in \mathbb{R}^n$, $||a||$ denotes the Euclidean norm. Finally, we use C to denote generic constants that do not depend on n.

3 Proof of theorem 2.1

We start by bounding the distance $d_1(X, \widehat{X})$, as defined in Eq. (2), in terms of D and $\widehat{D}$. Since $L(XX^T - \widehat{X}\widehat{X}^T)L$ has rank at most $2d$, it follows that

$$||L(XX^T - \widehat{X}\widehat{X}^T)L||_F \leq \sqrt{2d}||L(XX^T - \widehat{X}\widehat{X}^T)L||_2 \,, \tag{5}$$

where we used the fact that, for any rank $2d$ matrix A with singular values $\{\sigma_1, \ldots, \sigma_{2d}\}$, $||A||_F = \sqrt{\sum_{i=1}^{2d} \sigma_i^2} \leq \sqrt{2d}||A||_2$. The spectral norm can be bounded in terms of D and $\widehat{D}$.

$$
\begin{aligned}
&||L(XX^T - \widehat{X}\widehat{X}^T)L||_2 \\
&\quad \leq ||L(XX^T + (1/2)\widehat{D})L||_2 + || - (1/2)L\widehat{D}L - \widehat{X}\widehat{X}^T||_2 \tag{6} \\
&\quad \leq (1/2)||L(-D + \widehat{D})L||_2 + (1/2)||L(-\widehat{D} + D)L||_2 \tag{7} \\
&\quad \leq ||\widehat{D} - D||_2 \,, \tag{8}
\end{aligned}
$$

where in (6) we used the triangular inequality and the fact that $\widehat{X} = L\widehat{X}$ as shown in Section 2.2. In (7) we used Eq. (1) and the fact that

$$|| - (1/2)L\widehat{D}L - \widehat{X}\widehat{X}^T||_2 \leq || - (1/2)L\widehat{D}L - A||_2 \,, \tag{9}$$

for any rank-d matrix A. Since the rank of $-(1/2)LDL$ is d, the second term in (7) follows by setting $A = -(1/2)LDL$. The inequality in (8) follows trivially from the observation that $||L||_2 = 1$.

Eq. (9) can be proved as follows. From the definition of $\widehat{X} = \mathsf{MDS}_d(\widehat{D})$ in Section 2.2, we know that $\widehat{X}\widehat{X}^T$ is the best rank-d approximation to $-(1/2)L\widehat{D}L$. Hence, $\widehat{X}\widehat{X}^T$ minimizes $|| - (1/2)L\widehat{D}L - A||_2$ for any rank-d matrix A, and this proves Eq. (9).

To bound $||\widehat{D} - D||_2$, we use the following key result on the shortest paths. The main idea is that, for sensors with uniformly random positions, shortest paths provide arbitrarily close estimation to the correct distances for large enough radio range r. Define

$$\tilde{r}_0(\beta) \equiv 24\, d \left(\frac{(1+\beta)\log n}{n} \right)^{\frac{1}{d}} . \tag{10}$$

For simplicity we denote $\tilde{r}_0(0)$ by $\tilde{r}_0$.

Lemma 3.1 (Bound on the shortest paths). *Under the hypothesis of Theorem 2.1, w.h.p., the shortest paths between all pairs of nodes in the graph $G(V, E, P)$ are uniformly bounded by*

$$\hat{d}_{i,j}^2 \leq (1 + \frac{\tilde{r}_0}{r})d_{i,j}^2 + O(r) ,$$

for $r > r_0$, where $\tilde{r}_0$ is defined as in Eq. (10) and r_0 as in Eq. (3).

The proof of this lemma is given in Section 3.1. Define an error matrix $Z = \widehat{D} - D$. Then by Lemma 3.1, Z is element-wise bounded by $0 \leq Z \leq (\tilde{r}_0/r)D + O(r)$. Since all the entries are non-negative, we can bound the spectral norm of Z. Let u and v be the left and right singular vectors of Z respectively. Then by Perron-Frobenius theorem, u and v are non-negative. It follow that

$$
\begin{aligned}
||\widehat{D} - D||_2 &= u^T Z v \\
&\leq (\tilde{r}_0/r)u^T D v + O(rn) \\
&\leq (\tilde{r}_0/r)||D||_2 + O(rn) .
\end{aligned} \tag{11}
$$

The first inequality follows from the element-wise bound on Z and the non-negativity of u and v. It is simple to show that $||D||_2 = \Theta(n)$. Typically, we are interested in the region where $r = o(1)$, which implies that the first term in Eq. (11) dominates the error.

Now we are left with the task of bounding $||D||_2$. Using similar argument, we can prove a slightly loose bound : $||D||_2 \leq dn$. Note that D is non-negative and element-wise bounded by d. Let u and v be the left and right singular vectors of D, then again by Perron-Frobenius theorem, $||D||_2 = uDv^T \leq u(d\mathbb{1}_n\mathbb{1}_n^T)v^T \leq dn$.

We provide a tighter bound in the following lemma, and the proof, using concentration of measure inequalities, is provided in Section 3.2.

Lemma 3.2.

$$||D||_2 \leq \frac{d}{\sqrt{20}} n + o(n) ,$$

with high probability.

Applying Lemma 3.2 to Eq. (11) and substituting $||\widehat{D} - D||_2$ in Eq. (8), we get the desired bound :

$$d_1(X, \widehat{X}) \leq \frac{d\,\tilde{r}_0}{\sqrt{10}\,r} + O(r) ,$$

and this finishes the proof of Theorem 2.1.

3.1 Proof of the bound on the shortest paths

In this section, we prove a slightly stronger version of Lemma 3.1. We will show that under the hypothesis of Theorem 2.1, for any $\beta \geq 0$ and all sensors $i \neq j$, there exists a constant C such that, with probability larger than $1 - \frac{Cn^{-\beta}}{(1+\beta)\log n}$, the shortest paths between all the pairs of nodes are uniformly bounded by

$$\hat{d}_{i,j}^2 \leq (1 + \frac{\tilde{r}_0(\beta)}{r})d_{i,j}^2 + O(r) , \tag{12}$$

for $r > r_0$, where $\tilde{r}_0(\beta)$ is defined as in Eq. (10) and r_0 as in Eq. (3). Especially, Lemma 3.1 follows if we set $\beta = 0$.

We start by applying bin-covering technique to the random geometric points in $[0,1]^d$ in a similar way as in [?]. We cover the space $[0,1]^d$ with a set of non-overlapping hypercubes whose edge lengths are δ. Thus there are total $\lceil 1/\delta \rceil^d$ bins, each of volume δ^d. In formula, bin $(i_1, \ldots, i_d)$ is the hypercube $[(i_1 - 1)\delta, i_1\delta) \times \cdots \times [(i_d - 1)\delta, i_d\delta)$, for $i_k \in \{1, \ldots, \lceil 1/\delta \rceil\}$ and $k \in \{1, \ldots, d\}$. When n nodes are deployed in $[0,1]^d$ uniformly at random, we say a bin is empty if there is no node inside the bin. We want to ensure that, with high probability, there are no empty bins.

For a given δ, define a parameter $\beta \equiv (\delta^d n/\log n) - 1$. Since a bin is empty with probability $(1 - \delta^d)^n$, we apply union bound over all the $\lceil 1/\delta \rceil^d$ bins to get,

$$\mathbb{P}(\text{no bins are empty}) \geq 1 - \left\lceil \frac{1}{\delta} \right\rceil^d (1 - \delta^d)^n \tag{13}$$

$$\geq 1 - \frac{Cn}{(1+\beta)\log n}\left(1 - \frac{(1+\beta)\log n}{n}\right)^n \tag{14}$$

$$\geq 1 - \frac{Cn^{-\beta}}{(1+\beta)\log n} , \tag{15}$$

where in (14) we used the fact that there exists a constant C such that $\lceil 1/\delta \rceil^d \leq C/\delta^d$, and in (15) we used $(1 - 1/n)^n \leq e^{-1}$, which is true for any positive n.

Assuming that, with high probability, no bins are empty, we first show that the shortest paths is bounded by a function $F(d_{i,j})$ that only depends on the distance between the two nodes. Let $\hat{d}_{i,j}$ denote the shortest path between nodes i and j and $d_{i,j}$ denote the Euclidean distance. Define a function $F : \mathbb{R}^+ \to \mathbb{R}^+$ as

$$F(y) = \begin{cases} r & \text{if } y \leq r , \\ (k+1)r & \text{if } y \in \mathcal{L}_k \text{ for } k \in \{1, 2, \ldots\} , \end{cases}$$

where $\mathcal{L}_k$ denotes the interval $(r + (k-1)(r - 2\sqrt{d})\delta, r + k(r - 2\sqrt{d})\delta]$.

We will show, by induction, that for all pairs (i, j),

$$\hat{d}_{i,j} \leq F(d_{i,j}) . \tag{16}$$

First, in the case when $d_{i,j} \leq r$, by definition of *connectivity based model*, nodes i and j are connected by an edge in the graph G, whence $\hat{d}_{i,j} = r$.

Next, assume that the bound in Eq. (16) is true for all pairs (l, m) with $d_{l,m} \leq r + (k-1)(r - 2\sqrt{d})\delta$. We consider two nodes i and j at distance $d_{i,j} \in \mathcal{L}_k$. Consider a line segment $l_{i,j}$ in a d-dimensional space with one end at x_i and the other at x_j, where x_i and x_j denote the positions of nodes i and j, respectively. Let p be the point in the line $l_{i,j}$ which is at distance $r - \sqrt{d}\delta$ from x_i. Then, we focus on the bin that contains p. By assumption that no bins are empty, we know that we can always find
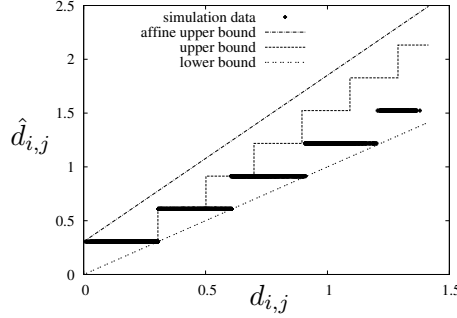
9

Figure 1: comparison of upper and lower bound of shortest paths $\{\hat{d}_{i,j}\}$ with respect to the correct distance $\{d_{i,j}\}$ computed for $n = 6000$ sensors in 2-dimensional square $[0,1]^2$ under *connectivity based model*.

at least one node in the bin. Let k denote any one of those nodes in the bin. Then we use following inequality which is true for all nodes (i, k, j).

$$\hat{d}_{i,j} \leq \hat{d}_{i,k} + \hat{d}_{k,j} .$$

Each of these two terms can be bounded using triangular inequality. To bound the first term, note that two nodes in the same bin are at most distance $\sqrt{d}\delta$ apart. Since p and x_k are in the same bin and p is at distance $r - \sqrt{d}\delta$ from node x_i by construction, we know that $d_{i,k} \leq ||x_i - p|| + ||p - x_k|| \leq r$, whence $\hat{d}_{i,k} = r$. Analogously for the second term, $d_{k,j} \leq ||x_j - p|| + ||p - x_k|| \leq r + (k-1)(r - 2\sqrt{d})\delta$, which implies that $\hat{d}_{k,j} \leq F(d_{k,j}) = kr$. Hence, we proved that if Eq. (16) is true for pairs (i, j) such that $d_{i,j} \leq r + (k-1)(r - 2\sqrt{d})\delta$, then $\hat{d}_{i,j} \leq (k+1)r$ for pairs (i, j) such that $d_{i,j} \in \mathcal{L}_k$. By induction, this proves that the bound in Eq. (16) is true for all pairs (i, j).

Let $\mu = (r/2\sqrt{d})\big(n/((1+\beta)\log n)\big)^{1/d}$. Then, the function $F(y)$ can be easily upper bounded by an affine function $F_a(y) = (1 + 1/(\mu - 1))y + r$. Hence we have following bound on the shortest paths between any two nodes i and j.

$$\hat{d}_{i,j} \leq \left(1 + \frac{1}{\frac{r}{2\sqrt{d}}\left(\frac{n}{(1+\beta)\log n}\right)^{1/d} - 1}\right) d_{i,j} + r . \tag{17}$$

Figure 1 illustrates the comparison of the upper bounds $F(d_{i,j})$ and $F_a(d_{i,j})$, and the trivial lower bound $\hat{d}_{i,j} \geq d_{i,j}$ in a simulation with parameters $d = 2$, $n = 6000$ and $r = \sqrt{64 \log n / n}$. The simulation data shows the distribution of shortest paths between all pairs of nodes with respect to the actual pairwise distances, which confirms that shortest paths lie between the analytical upper and lower bounds. While the gap between the upper and lower bound is seemingly large, in the regime where $r = \sqrt{\alpha \log n / n}$ with a constant α, the vertical gap r vanishes as n goes to infinity and the slope of the affine upper bound can be made arbitrarily small by increasing the radio range r or equivalently taking large enough α.

The bound on the squared shortest paths $\hat{d}_{i,j}^2$ can be derived from the bound on the shortest

paths in Eq. (17) after some calculus.

$$\hat{d}_{i,j}^2 \;\leq\; \left\{ \frac{\mu}{\mu-1} d_{i,j} + r \right\}^2 \tag{18}$$

$$= \; \frac{\mu^2}{(\mu-1)^2} d_{i,j}^2 + r^2 + 2\frac{\mu}{\mu-1} d_{i,j}\, r \tag{19}$$

$$= \; \left(1 + \frac{2\mu-1}{(\mu-1)^2}\right) d_{i,j}^2 + O(r) \tag{20}$$

$$\leq \; \left(1 + \frac{12}{\mu}\right) d_{i,j}^2 + O(r)\,. \tag{21}$$

where in (20) we used the fact that $(\mu/(\mu-1)) d_{i,j} = O(1)$, which follows from the assumptions $(r/4\sqrt{d})^d > (1+\beta)\log n/n$ and $d = O(1)$. In (21), we used the inequality $(2\mu-1)/(\mu-1)^2 \leq 12/\mu$, which is true for $\mu \geq 3/2$. Substituting μ in the formula, this finishes the proof of the desired bound in Eq. (12).

Note that although for the sake of simplicity, we focus on $[0,1]^d$ hypercube, our analysis easily generalizes to any bounded convex set and homogeneous Poisson process model with density $\rho = n$. The homogeneous Poisson process model is characterized by the probability that there are exactly k nodes appearing in any region with volume A : $\mathbb{P}(k_A = k) = \frac{(\rho A)^k}{k!} e^{-\rho A}$. Here, k_A is a random variable defined as the number of nodes in a region of volume A.

3.2   Proof of Lemma 3.2

In this section, we prove a slightly stronger version of Lemma 3.2. We will show that for any $\epsilon \in (0.5, 1)$, there exists a constant C such that

$$||D||_2 \;\leq\; \frac{d}{\sqrt{20}} n + C d\, n^\epsilon\,, \tag{22}$$

with probability larger than $1 - 8e^{-3n^{(2\epsilon-1)}}$.

First, define a set of i.i.d. random matrices $\{D^{(1)}, \ldots, D^{(d)}\}$ as $D_{ij}^{(k)} = (X_{ik} - X_{jk})^2$ such that $D = \sum_{k=1}^{d} D^{(k)}$. Then, by the triangular inequality, the task of bounding $||D||_2$ is divided into simpler tasks of bounding $||D^{(k)}||_2$.

$$||D||_2 \leq \sum_{k=1}^{d} ||D^{(k)}||_2\,. \tag{23}$$

To bound $||D^{(k)}||_2$, we introduce the following notations. Let $y_k \in \mathbb{R}^n$ denote the kth column of X and $\tilde{y}_k \in \mathbb{R}^n$ the vector with $\tilde{y}_{k,i} = y_{k,i}^2$. In the following, it is more convenient to work with a model in which the entries of y_k are drawn independently from uniform distribution over $[-0.5, 0.5]$. Since pairwise distances are invariant to any translation of X, any line segment of length one is equivalent.

Note that there is an alternate representation of $D^{(k)}$ in terms of y_k and $\tilde{y}_k$, which clearly shows the rank-3 structure.

$$D^{(k)} = \mathbb{1}\tilde{y}_k^T + \tilde{y}_k \mathbb{1}^T - 2 y_k y_k^T\,.$$

Let $\{u_1, u_2, u_3\}$ be the orthogonal basis for subspace spanned by the columns of $D^{(k)}$ defined as follows : $u_1 = \mathbb{1}$, $u_2 = \tilde{y}_k - (\mathbb{1}^T \tilde{y}_k/n)\mathbb{1}$, and $u_3 = x - (\mathbb{1}^T y_k/n)\mathbb{1} - (u_2^T y_k/||u_2||^2) u_2$. Then

$$||D^{(k)}||_2 \leq ||A||_2 + ||D^{(k)} - A||_2\,,$$

11

where we used the triangular inequality and defined $A = \mathbb{1}\tilde{y}_k^T + \tilde{y}_k\mathbb{1}^T - 2u_3u_3^T$. We will show that $||A||_2 \leq n/\sqrt{20} + Cn^\epsilon$ and $||D^{(k)} - A||_2 \leq Cn^\epsilon$, whence the bound in Eq. (22) follows from the union bound over $D^{(k)}$ for $k \in \{1, \ldots, d\}$ and Eq. (23).

The key technical step in proving the bounds for $||A||_2$ and $||D^{(k)} - A||_2$ is the following Remark on the concentration of measure around the mean. For the proof we refer to Appendix A.

Remark 3.3. *For any positive constant $\epsilon \in (0.5, 1)$, define a set of events $\{E_1, E_2, E_3, E_4\}$ as $E_1 \equiv \{|\mathbb{1}^T y_k| \leq n^\epsilon\}$, $E_2 \equiv \{\left|||y_k||^2 - \frac{1}{12}n\right| \leq n^\epsilon\}$, $E_3 \equiv \{|\tilde{y}_k^T y_k| \leq n^\epsilon\}$, and $E_4 \equiv \{\left|||\tilde{y}_k||^2 - \frac{1}{80}n\right| \leq n^\epsilon\}$. Then*

$$\mathbb{P}(E_i) \geq 1 - 2e^{-3n^{(2\epsilon-1)}},$$

for all $i \in \{1, 2, 3, 4\}$.

Assuming the events E_1, E_2, E_3 and E_4 are true, we can prove the bounds on $||A||_2$ and $||D^{(k)} - A||_2$.

To upper bound $||A||_2$, we use the fact that u_3 is orthogonal to the subspace spanned by $\{\mathbb{1}, \tilde{y}_k\}$, whence

$$||A||_2 \leq \max\left\{\left||\mathbb{1}\tilde{y}_k^T + \tilde{y}_k\mathbb{1}^T\right||_2, \ 2\left||u_3u_3^T\right||_2\right\}.$$

Applying the bound in E_4, the first term can be bounded by $||\mathbb{1}\tilde{y}_k^T + \tilde{y}_k\mathbb{1}^T||_2 \leq 2||\mathbb{1}||\,||\tilde{y}_k|| \leq (1/\sqrt{20})n + Cn^\epsilon$. Since u_3 is an orthogonal projection of x, the second term is bounded by $||u_3u_3^T||_2 \leq ||y_k||^2 \leq (1/12)n + Cn^\epsilon$, using the bound in E_2. This proves the desired bound on $||A||_2$.

Similarly, we use following inequality to show $||D^{(k)} - A||_2 \leq Cn^\epsilon$.

$$||D^{(k)} - A||_2 = 2||y_k y_k^T - u_3 u_3^T||_2$$
$$\leq 2\left\{\left||\frac{(\mathbb{1}^T y_k)^2}{n^2}\mathbb{1}\mathbb{1}^T\right||_2 + \left||\frac{\mathbb{1}^T y_k}{n}(\mathbb{1}y_k^T + y_k\mathbb{1}^T)\right||_2 + \left||\frac{(u_2^T y_k)^2}{||u_2||^4}u_2 u_2^T\right||_2 + \left||\frac{u_2^T y_k}{||u_2||^2}(u_2 y_k^T + y_k u_2^T)\right||_2\right\}$$

Using the bound in E_1 and E_2, the first term is bounded by $n^{(2\epsilon-1)}$ and the second term is bounded by Cn^ϵ. Note that Remark 3.3 implies that $\left|||u_2||^2 - (59/720)n\right| \leq Cn^\epsilon$ and $|u_2^T y_k| \leq n^\epsilon$ with high probability. Then analogously, the third term is bounded by $Cn^{(2\epsilon-1)}$ and the fourth term is bounded by Cn^ϵ.

Applying union bound over the set of events $\{E_i^c\}$, we proved that Eq. (22) is true with probability larger than $1 - 8e^{-3n^{(2\epsilon-1)}}$.

A   Proof of Remark 3.3

Define a random variable $Y_p = \sum_{i=1}^n y_{k,i}^p$. For $p \in \{1, 2, 3, 4\}$ we start by proving following bound :

$$\mathbb{P}\left(Y_p - \mathbb{E}[Y_p] > n^\epsilon\right) \leq \exp\left\{-(2p+1)2^{(2p-2)}n^{(2\epsilon-1)}\right\},$$

for any $\epsilon \in (0.5, 1)$. Assuming n large enough, fix $\lambda = (2p+1)2^{2p-2}n^{-(1-\epsilon)}$ such that $|\lambda y_{k,i}| \leq 1/2$, whence $e^{\lambda y_{k,i}^p} \leq 1 + \lambda y_{k,i}^p + 2(\lambda y_{k,i}^p)^2$. It follows that

$$\mathbb{E}[e^{\lambda Y_p}] \leq \mathbb{E}[1 + \lambda y_{k,i}^p + 2\lambda^2 y_{k,i}^{2p}]^n$$
$$= \exp\left\{n\log\left(1 + \lambda\mathbb{E}[y_{k,i}^p] + \lambda^2\frac{1}{(2p+1)2^{(2p-1)}}\right)\right\} \tag{24}$$
$$\leq \exp\left\{n\left(\lambda\mathbb{E}[y_{k,i}^p] + \lambda^2\frac{1}{(2p+1)2^{(2p-1)}}\right)\right\}, \tag{25}$$

where Eq. (24) follows from $\mathbb{E}[y_{k,i}^{2p}] = 1/(2p+1)2^{2p}$ and Eq. (25) from the inequality $\log(1+a) \le a$ for all $a > -1$. Substituting $\lambda = (2p+1)2^{2p-2}n^{-(1-\epsilon)}$, the statement follows from Chernoff bound $\mathbb{P}(Y_p > \mathbb{E}[Y_p] + a) \le e^{-\lambda(\mathbb{E}[Y_p]+a)}\mathbb{E}[e^{\lambda Y_p}]$ after simple calculus.

Applying the same argument to $-Y_p + \mathbb{E}[Y_p]$, we get

$$\mathbb{P}\Big(|Y_p - \mathbb{E}[Y_p]| > n^\epsilon\Big) \le e^{-(2p+1)2^{(2p-2)}n^{(2\epsilon-1)}} \le 2e^{-3n^{(2\epsilon-1)}} , \tag{26}$$

for $p \in \{1, 2, 3, 4\}$.

By definition, we have $\mathbb{1}^T y_k = Y_1$, $||y_k||^2 = Y_2$, $\tilde{y}_k^T y_k = Y_3$, $||\tilde{y}_k||^2 = Y_4$. Applying Eq. (26) with $\mathbb{E}[Y_1] = 0, \mathbb{E}[Y_2] = \frac{1}{12}n, \mathbb{E}[Y_3] = 0$ and $\mathbb{E}[Y_4] = \frac{1}{80}n$, we proved the remark.

Acknowledgment

References

[BY04] P. Biswas and Y. Ye, *Semidefinite programming for ad hoc wireless sensor network localization*, IPSN '04: Proceedings of the 3rd international symposium on Information processing in sensor networks (New York, NY, USA), ACM, 2004, pp. 46–54.

[DJMI+] P. Drineas, A. Javed, M. Magdon-Ismail, G. Pandurangant, R. Virrankoski, and A. Savvides, *Distance matrix reconstruction from incomplete distance information for sensor network localization*, Proc. IEEE SECON 2006.

[GK98] P. Gupta and P.R. Kumar, *Critical power for asymptotic connectivity*, Decision and Control, 1998. Proceedings of the 37th IEEE Conference on, vol. 1, 1998, pp. 1106–1110 vol.1.

[JH01] G. Boriello J. Hightower, *Locations systems for ubiquitous computing*, IEEE Computer **34** (2001), 57–66.

[Joh77] D. B. Johnson, *Efficient algorithms for shortest paths in sparse networks*, J. ACM **24** (1977), no. 1, 1–13.

[JZ04] Xiang Ji and Hongyuan Zha, *Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling*, 2004.

[SRZF03] Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz, *Localization from mere connectivity*, MobiHoc '03: Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing (New York, NY, USA), ACM, 2003, pp. 201–212.

[SRZF04] ———, *Localization from connectivity in sensor networks*, IEEE Trans. Parallel Distrib. Syst. **15** (2004), no. 11, 961–974.