Localization from Incomplete Noisy Distance Measurements

Adel Javanmard
Department of Electrical Engineering
Stanford University

Andrea Montanari
Department of Electrical Engineering and
Department of Statistics
Stanford University

***Abstract*—We consider the problem of positioning a cloud of points in the Euclidean space $\mathbb{R}^d$, from noisy measurements of a subset of pairwise distances. This task has applications in various areas, such as sensor network localizations, NMR spectroscopy of proteins, and molecular conformation. Also, it is closely related to dimensionality reduction problems and manifold learning, where the goal is to learn the underlying global geometry of a data set using measured local (or partial) metric information. Here we propose a reconstruction algorithm based on a semidefinite programming approach. For a random geometric graph model and uniformly bounded noise, we provide a precise characterization of the algorithm's performance: In the noiseless case, we find a radius r_0 beyond which the algorithm reconstructs the exact positions (up to rigid transformations). In the presence of noise, we obtain upper and lower bounds on the reconstruction error that match up to a factor that depends only on the dimension d, and the average degree of the nodes in the graph.**

I. Introduction

A. Problem Statement

Consider the random geometric graph model $G(n,r) = (V,E)$ where V is a set of n nodes distributed uniformly at random in the d-dimensional hypercube $[-0.5, 0.5]^d$, and $E \in V \times V$ is a set of edges that connect the nodes which are close to each other; i.e., $(i,j) \in E \Leftrightarrow d_{ij} = \|x_i - x_j\| \le r$. For each edge $(i,j) \in E$, $\tilde{d}_{ij}$ denotes the measured distance between nodes i and j. Denoting by $z_{ij} \equiv \tilde{d}_{ij}^2 - d_{ij}^2$ the measurement error, we consider a *"worst case model"*, in which the errors $\{z_{ij}\}_{(i,j)\in E}$ are arbitrary but uniformly bounded $|z_{ij}| \le \Delta$.

Given the graph $G(n,r)$ and its associated proximity distance measurements, $\tilde{d}_{ij}$, the *localization* problem is to reconstruct the positions of the nodes. In this paper, we propose an algorithm for this problem based on semidefinite programming and provide a rigorous analysis of its performance.

Notice that the positions of the nodes can only be determined up to rigid transformations (a combination of rotation, reflection and translation) of the nodes, because the inter point distances are invariant to rigid transformations. Therefore, we use the following metric, defined in [8], to evaluate the distance between the original position matrix $X \in \mathbb{R}^{n\times d}$ and the estimation $\widehat{X} \in \mathbb{R}^{n\times d}$. Let $L = I - (1/n)uu^T$, where $u \in \mathbb{R}^n$ is the all-ones vector. It is easy to see that LXX^TL is invariant under rigid transformations of X. The metric is defined as $d(X,\widehat{X}) = 1/n\|LXX^TL - L\widehat{X}\widehat{X}^TL\|_F$. This is a measure of the average reconstruction error per point, when X and $\widehat{X}$ are aligned optimally. More specifically, there exists a rotation of $\hat{X}$, call it $\hat{Y}$, such that $d(X,\hat{X})$ measures the mean square error between the positions in X and in $\hat{Y}$ [7].

Remark. Clearly, connectivity of G is a necessary assumption for the localization problem to be solvable. It is a well known result that the graph $G(n,r)$ is connected with high probability if $K_d r^d > (\log n + c_n)/n$, where K_d is the volume of the $d-$dimensional unit ball and $c_n \to \infty$ [9]. Viceversa, the graph is with positive probability disconnected if $K_d r^d \le (\log n + C)/n$ for some constant C. Hence, we focus on the regime where $r = \alpha(\log n/n)^{\frac{1}{d}}$ for some constant α. We further notice that, under the random geometric graph model, the configuration of the points is almost surely *generic*, in the sense that the coordinates do not satisfy any nonzero polynomial equation with integer coefficients.

B. Algorithm and main results

The following algorithm uses semidefinite programming (SDP) to solve the localization problem.

Algorithm SDP-based Algorithm for Localization

Input: dimension d, distance measurements $\tilde{d}_{ij}$ for $(i,j) \in E$, bound on the measurement noise Δ
Output: estimated coordinates in $\mathbb{R}^d$
1: Solve the following SDP problem:
$$\begin{aligned} &\text{minimize} && \mathrm{Tr}(Q) \\ &\text{s.t.} && \left|\langle M_{ij}, Q\rangle - \tilde{d_{ij}}^2\right| \le \Delta, \qquad (i,j) \in E \\ &&& Q \succeq \mathbf{0}. \end{aligned}$$
2: Compute the best rank-d approximation $U_d\Sigma_dU_d^T$ of Q
3: Return $\widehat{X} = U_d\Sigma_d^{1/2}$.

Here $M_{ij} = e_{ij}e_{ij}^T \in \mathbb{R}^{n\times n}$, where $e_{ij} \in \mathbb{R}^n$ is the vector with $+1$ at the i^{th} position, -1 at the j^{th} position and zero everywhere else. Also, $\langle A, B\rangle \equiv \mathrm{tr}(A^TB)$. Note that with a slight abuse of notation, the solution of the SDP problem in the first step is denoted by Q.

Let $Q_0 := XX^T$ be Gram matrix of the node positions, namely $Q_{0,ij} = x_i \cdot x_j$. A key observation is that Q_0 is a low rank matrix: $\mathrm{rank}(Q_0) \le d$, and obeys the constraints of the SDP problem. By minimizing $\mathrm{Tr}(Q)$ in the first step, we promote low-rank solutions Q (since $\mathrm{Tr}(Q)$ is the sum of the eigenvalues of Q). Alternatively, this minimization can be interpreted as setting the center of gravity of $\{x_1,\dots,x_n\}$ to coincide with the origin, thus removes the degeneracy due to translational invariance.

In step 2, the algorithm computes the eigendecomposition of Q and retains the d largest eigenvalues. This is equivalent to computing the best rank-d approximation of Q in Frobenius norm. The gravity center of the reconstructed points remains at the origin by this operation.

Our main result provides a complete characterization of the robustness properties of the SDP-based algorithm. Here and below 'with high probability' means with probability converging to 1 as $n \to \infty$ for d fixed.

Theorem I.1 *Let $\{x_1, \dots, x_n\}$ be n nodes distributed uniformly at random in the hypercube $[-0.5, 0.5]^d$. Further, assume connectivity radius $r \geq \alpha(\log n/n)^{\frac{1}{d}}$, with $\alpha > 16 \cdot (8/K_d)^{1/d}$, and K_d the volume of $d-$dimensional unit ball. Then with high probability, the error distance between the estimate $\widehat{X}$ returned by the SDP-based algorithm and the correct coordinate matrix X is upper bounded as*

$$d(X, \widehat{X}) \leq C_1 (nr^d) \frac{\Delta}{r^4}. \tag{1}$$

Conversely, with high probability, there exist adversarial measurement errors $\{z_{ij}\}_{(i,j)\in E}$ such that

$$d(X, \widehat{X}) \geq C_2 \frac{\Delta}{r^4}\,, \tag{2}$$

Here, C_1 and C_2 denote constants that depend only on d.

A special case of this theorem concerns the case of exact measurements.

Corollary I.1. *Let $\{x_1, \dots, x_n\}$ be n nodes distributed uniformly at random in the hypercube $[-0.5, 0.5]^d$. If $r \geq 16 \cdot (8/K_d)^{1/d} (\log n/n)^{\frac{1}{d}}$, and the distance measurements are exact, then with high probability, the SDP-based algorithm recovers the exact positions (up to rigid transformations).*

C. Related work

The localization problem and its variants have attracted significant interest over the past years due to their applications in numerous areas, such as sensor network localization [2], NMR spectroscopy [5], and manifold learning [10], [12]; to name a few.

Of particular interest to our work are the algorithms proposed for the localization problem [8], [11], [2]. In general, few analytical results are known about the performance of these algorithms, particularly in the presence of noise.

The existing algorithms can be categorized in to two groups. The first group consists of algorithms who try first to estimate the missing distances and then use MDS to find the positions from the reconstructed distance matrix [8], [3]. The algorithms in the second group formulates the localization problem as a non-convex optimization problem and then use different relaxation schemes to solve it. A recent example of this type is relaxation to an SDP [2]. A crucial assumption in these works is the existence of some anchors among the nodes whose exact positions are known. The SDP is then used to efficiently check whether the graph is uniquely d-localizable and to find its unique realization.

II. Preliminaries

A. Rigidity Theory

This section is a very brief overview of definitions and results in rigidity theory which will be useful in this paper. We refer the interested reader to [6], [1], for a thorough discussion.

A *framework* G_X is an undirected graph $G = (V, E)$ along with a *configuration* $X \in \mathbb{R}^{n \times d}$ whose i^{th} row $x_i^T \in \mathbb{R}^d$ is the position of node i in the graph. The edges of G correspond to the distance constraints.

Rigidity matrix. Consider a motion of the framework with $x_i(t)$ being the position vector of point i at time t. Any smooth motion that instantaneously preserves the distance d_{ij} must satisfy $\frac{d}{dt}\|x_i - x_j\|^2 = 0$ for all edges (i,j). Equivalently,

$$(x_i - x_j)^T (\dot{x_i} - \dot{x_j}) = 0 \quad \forall (i,j) \in E, \tag{3}$$

where $\dot{x_i}$ is the velocity of the i^{th} point. Given a framework $G_X \in \mathbb{R}^d$, a solution $\dot{X} = [\dot{x}_1^T\ \dot{x}_2^T\ \cdots\ \dot{x}_n^T]^T$, with $\dot{x}_i \in \mathbb{R}^d$, for the linear system of equations (3) is called an *infinitesimal motion* of the framework G_X. This linear system of equations consists of $|E|$ equations in dn unknowns and can be written in the matrix form $R_G(X)\dot{X} = 0$, where $R_G(X)$ is called the $|E| \times dn$ *rigidity matrix*.

It can be seen that for every skew symmetric matrix $A \in \mathbb{R}^{d \times d}$ and for every vector $b \in \mathbb{R}^d$, $\dot{x}_i = Ax_i + b$ is an infinitesimal motion. Notice that these motions span a $d(d+1)/2$ dimensional space, accounting $d(d-1)/2$ degrees of freedom for orthogonal transformations, A, and d degrees of freedom for translations, b. Hence, dim Ker$(R_G(X)) \geq d(d+1)/2$. A framework is said to be *infinitesimally rigid* if dim Ker$(R_G(X)) = d(d+1)/2$.

Stress matrix. A *stress* for a framework G_X is an assignment of scalars ω_{ij} to the edges such that for each $i \in V$,

$$\sum_{j:(i,j)\in E} \omega_{ij}(x_i - x_j) = (\sum_{j:j\neq i} \omega_{ij})x_i - \sum_{j:j\neq i} \omega_{ij} x_j = 0.$$

A stress vector can be rearranged into an $n \times n$ symmetric matrix Ω , known as the *stress matrix*, such that for $i \neq j$, the (i,j) entry of Ω is $\Omega_{ij} = -\omega_{ij}$, and the diagonal entries for (i,i) are $\Omega_{ii} = \sum_{j:j\neq i} \omega_{ij}$. Since all the coordinate vectors of the configuration as well as the all-ones vector are in the null space of Ω, the rank of the stress matrix for generic configurations is at most $n - d - 1$.

B. Notations

For a vector $v \in \mathbb{R}^n$, and a subset $T \subseteq \{1, \cdots, n\}$, v_T is the restriction of v to indices in T. We use the notation $\langle v_1, \cdots, v_n \rangle$ to represent the subspace spanned by vectors v_i, $1 \leq i \leq n$. The orthogonal projections onto subspaces V and $V^\perp$ are respectively denoted by P_V and $P_V^\perp$. Throughout this paper, $u \in \mathbb{R}^n$ is the all-ones vector and C is a constant depending only on the dimension d, whose value may change from case to case.

Given a matrix A, we denote its operator norm by $\|A\|_2$, its Frobenius norm by $\|A\|_F$ and its nuclear norm by $\|A\|_*$. (the latter is simply the sum of the singular values of A).

Finally, we denote by $x^{(i)} \in \mathbb{R}^n$, $i \in \{1,\dots,d\}$ the i^{th} column of the positions matrix X. In other words $x^{(i)}$ is the vector containing the i^{th} coordinate of points $x_1,\dots,x_n$.

Throughout the proof we shall adopt the convention of using the notations X, $\{x_j\}_{j\in[n]}$, and $\{x^{(i)}\}_{i\in[d]}$ to denote the centered positions. In other words $X = LX'$ where the rows of X' are i.i.d. uniform in $[-0.5, 0.5]^d$.

III. Proof of Theorem I.1

Let $V = \langle u, x^{(1)}, \cdots, x^{(d)}\rangle$ and for any matrix $S \in \mathbb{R}^{n\times n}$, define

$$\tilde{S} = P_V S P_V + P_V S P_V^{\perp} + P_V^{\perp} S P_V, \qquad S^{\perp} = P_V^{\perp} S P_V^{\perp}.$$

Thus $S = \tilde{S} + S^{\perp}$. Also, denote by R the difference between the optimum solution Q and the actual Gram matrix Q_0, i.e., $R = Q - Q_0$. The proof of Theorem I.1 is based on the following key lemmas that bound $R^{\perp}$ and $\tilde{R}$ separately.

Lemma III.1. *There exists a numerical constant $C = C(d)$, such that, with high probability,*

$$\|R^{\perp}\|_* \le C\frac{n}{r^4}(nr^d)\Delta\,. \tag{4}$$

Lemma III.2. *There exists a numerical constant $C = C(d)$, such that, with high probability,*

$$\|\tilde{R}\|_F \le C\frac{n}{r^4}(nr^d)\Delta \tag{5}$$

We defer the proof of lemmas III.1 and III.2 to the next section.

Proof (Theorem I.1): Let $Q = \sum_{i=1}^n \sigma_i u_i u_i^T$, where $\|u_i\| = 1$, $u_i^T u_j = 0$ for $i \neq j$ and $\sigma_1 \ge \sigma_2 \ge \cdots \ge 0$. In the second step of algorithm, Q is projected onto subspace $\langle u_1, \cdots, u_d\rangle$. Denote the result by $P_d(Q)$. As pointed out before, $P_d(Q)u = 0$ and $Q_0 u = 0$. This implies that $P_d(Q) = LP_d(Q)L$ and $Q_0 = LQ_0L$. By triangle inequality,

$$\begin{aligned}\|LP_d(Q)L - LQ_0L\|_F &= \|P_d(Q) - Q_0\|_F \\ &\le \|P_d(Q) - \tilde{Q}\|_F + \|\tilde{Q} - Q_0\|_F.\end{aligned} \tag{6}$$

Observe that, $\tilde{Q} = Q_0 + \tilde{R}$ and $Q^{\perp} = R^{\perp}$. Since $P_d(Q) - \tilde{Q}$ has rank at most $3d$, it follows that $\|P_d(Q) - \tilde{Q}\|_F \le \sqrt{3d}\|P_d(Q) - \tilde{Q}\|_2$ (for any matrix A, $\|A\|_F \le \text{rank}(A)\|A\|_2$). By triangle inequality, we have

$$\|P_d(Q) - \tilde{Q}\|_2 \le \|P_d(Q) - Q\|_2 + \|\underbrace{Q - \tilde{Q}}_{R^{\perp}}\|_2 \tag{7}$$

Note that $\|P_d(Q) - Q\|_2 = \sigma_{d+1}$. Recall the variational principle for the eigenvalues.

$$\sigma_q = \min_{H, dim(H)=n-q+1}\ \max_{y\in H, \|y\|=1} y^T Q y$$

Taking $H = \langle x^{(1)}, \cdots, x^{(d)}\rangle^{\perp}$, for any $y \in H$, $y^TQy = y^T P_V^{\perp} Q P_V^{\perp} y = y^T Q^{\perp} y = y^T R^{\perp} y$, where we used the fact $Qu = 0$ in the first equality (recall that $Qu = 0$ because Q minimizes $\text{Tr}(Q)$). Therefore, $\sigma_{d+1} \le \max_{\|y\|=1} y^T R^{\perp} y = \|R^{\perp}\|_2$ It follows from Eqs. (6) and (7) that

$$\|LP_d(Q)L - LQ_0L\|_F \le 2\sqrt{3d}\|R^{\perp}\|_2 + \|\tilde{R}\|_F$$

Using Lemma III.1 and III.2, we obtain

$$d(X, X') = \frac{1}{n}\|LP_d(Q)L - LQ_0L\|_F \le C(nr^d)\frac{\Delta}{r^4}$$

which proves the thesis.

A proof sketch of the converse part is provided in Appendix B. ∎

IV. Proofs of the lemmas

In this section we provide the proofs of lemmas III.1 and III.2. Due to space limitations, we will omit the proofs of several technical steps, and defer them to [7].

A. Proof of Lemma III.1

The proof is based on the following three steps: (i) Construct a stress matrix Ω of rank $n-d-1$ for the framework; (ii) Upper bound the quantity $\langle \Omega, R^{\perp}\rangle$; (iii) Lower bound the minimum nonzero eigenvalue of Ω.

For the graph G, we define $\text{cliq}(G) := \{\mathcal{C}_1, \cdots, \mathcal{C}_n\}$, where $\mathcal{C}_i = \{j \in V(G) : d_{ij} \le r/2\}$. (Note that the nodes in each $\mathcal{C}_i$ form a clique in G). Our first lemma estabilishes a simple property of $\text{cliq}(G)$. Its proof is immediate and deferred to a journal version of this paper [7].

Proposition IV.1. *If $r \ge \alpha(\log n/n)^{1/d}$ with $\alpha > 16 \cdot (8/K_d)^{1/d}$, the following is true with high probability. For any two nodes i and j, such that $\|x_i - x_j\| \le r/2$, $|\mathcal{C}_i \cap \mathcal{C}_j| \ge d+1$.*

A crucial role in the proof is played by the stress matrix of G_X. A special construction of such a matrix is obtained as follows

$$\Omega = \sum_{\mathcal{C}_i \in \text{cliq}(G)} P^{\perp}_{\langle u_{\mathcal{C}_i}, x^{(1)}_{\mathcal{C}_i}, \cdots, x^{(d)}_{\mathcal{C}_i}\rangle}.$$

The proof of the next statement is again immediate and omitted from this version of the paper.

Proposition IV.2. *The matrix Ω defined above is a positive semidefinite (PSD) stress matrix of rank $n-d-1$ for the framework G_X.*

Proposition IV.3. *There exists a constant $C = C(d)$, such that, with high probability,*

$$\langle \Omega, R^{\perp}\rangle \le Cn(nr^d)^2\Delta$$

Proof: Note that the matrix $\Omega = [\omega_{ij}]$ can be written as $\Omega = \sum_{(i,j)\in E} \omega_{ij} M_{ij}$. Define $\omega_{\max} = \max_{i\neq j}|\omega_{ij}|$. Then,

$$\begin{aligned}\langle \Omega, R^{\perp}\rangle &\overset{(a)}{=} \langle \Omega, R\rangle = \sum_{(i,j)\in E} \omega_{ij}\langle M_{ij}, R\rangle \\ &\le \sum_{(i,j)\in E} \omega_{\max}|\langle M_{ij}, Q - Q_0\rangle| \\ &\le \sum_{(i,j)\in E} \omega_{\max}(|\langle M_{ij}, Q\rangle - \tilde{d_{ij}}^2| + |\underbrace{\tilde{d_{ij}}^2 - d_{ij}^2}_{z_{ij}}|) \\ &\le 2\omega_{\max}|E|\Delta,\end{aligned}$$

where (a) follows from the fact $\Omega X = 0$. Note that the expected number of edges in G is at most $\frac{1}{2}n^2 K_d r^d$, and the number of edges is concentrated around its mean (Chernoff bounds). Hence, with high probability, $|E| \le n^2 K_d r^d$. Setting $C = 2K_d$, we get $\langle \Omega, R^\perp \rangle \le C\, \omega_{\max} n (nr^d)\Delta$.

Since $\Omega \succeq \mathbf{0}$, for any $1 \le i,j \le n$, $\omega_{ij}^2 \le \omega_{ii}\omega_{jj} = (e_i^T \Omega e_i)(e_j^T \Omega e_j) \le \sigma_{\max}^2(\Omega)$, where $\sigma_{\max}(\Omega)$ is the largest eigenvalue of Ω.

Hence, $\omega_{\max} \le \sigma_{\max}(\Omega) \le Cnr^d$ whereby the last step is is proved in the Claim below. ■

Claim IV.1. *There exists a constant $C = C(d)$, such that, with high probability,*

$$\sigma_{\max}(\Omega) \le Cnr^d$$

Proof: For any vector $v \in \mathbb{R}^n$,

$$v^T \Omega v = \| \sum_{\mathcal{C}_i \in cliq(G)} P^\perp_{\langle u_{\mathcal{C}_i}, x^{(1)}_{\mathcal{C}_i}, \cdots, x^{(d)}_{\mathcal{C}_i} \rangle} v \|^2 \le \sum_{i=1}^n \|v_{\mathcal{C}_j}\|^2$$
$$= \sum_{j=1}^n v_j^2 \sum_{i: j \in \mathcal{C}_i} 1 = \sum_{j=1}^n |\mathcal{C}_j| v_j^2 \le Cnr^d \|v\|^2.$$

The last inequality follows from the fact that, with high probability, $|\mathcal{C}_j| \le Cnr^d$ for all j and some constant C. ■

We now pass to lower bounding the smallest non-zero singular value of Ω, $\sigma_{\min}(\Omega)$. To prove such an estimate, recall that the laplacian $\mathcal{L}$ of the graph G is the symmetric matrix indexed by the vertices V, such that $\mathcal{L}_{ij} = -1$ if $(i,j) \in E$, $\mathcal{L}_{ii}$ =degree(i) and $\mathcal{L}_{ij} = 0$ otherwise. It is useful to recall a basic estimate on the laplacian of random geometric graphs.

Remark IV.1. Let $\mathcal{L}_{\text{sym}}$ denote the normalized laplacian of the random geometric graph $G(n,r)$, defined as $\mathcal{L}_{\text{sym}} = D^{-1/2}\mathcal{L}D^{-1/2}$, where D is the diagonal matrix with degrees of the nodes on diagonal. Then, with high probability, $\lambda_2(\mathcal{L}_{\text{sym}})$, the second smallest eigenvalue of $\mathcal{L}_{\text{sym}}$, is at least Cr^2. Therefore, $\lambda_2(\mathcal{L}) \ge C(nr^d)r^2$.

Proposition IV.4. *There exists a constant $C = C(d)$, such that, with high probability, $\Omega \succeq Cr^2\mathcal{L}$ on the space $V^\perp$.*

Proof: Due to space limitations, we present the proof for the case $d = 1$. The general argument proceeds along the same lines, and we defer it to [7].

Let $v \in V^\perp$ be an arbitrary vector. Decompose v locally as $v_{\mathcal{C}_i} = \beta_i \tilde{x}_{\mathcal{C}_i} + \gamma_i u_{\mathcal{C}_i} + w^{(i)}$, where $\tilde{x}_{\mathcal{C}_i} = P^\perp_{u_{\mathcal{C}_i}} x$ and $w^{(i)} \in \langle x_{\mathcal{C}_i}, u_{\mathcal{C}_i} \rangle^\perp$. Hence, $v^T \Omega v = \sum_{i=1}^n \|w^{(i)}\|^2$. Note that $v_{\mathcal{C}_i \cap \mathcal{C}_j}$ has two representations, whence we obtain

$$w^{(i)}_{\mathcal{C}_i \cap \mathcal{C}_j} - w^{(j)}_{\mathcal{C}_i \cap \mathcal{C}_j} = (\beta_j - \beta_i)\tilde{x}_{\mathcal{C}_i \cap \mathcal{C}_j} + \tilde{\gamma}_{i,j} u_{\mathcal{C}_i \cap \mathcal{C}_j}. \tag{8}$$

Here, $\tilde{x}_{\mathcal{C}_i \cap \mathcal{C}_j} = P^\perp_{u_{\mathcal{C}_i \cap \mathcal{C}_j}} x_{\mathcal{C}_i \cap \mathcal{C}_j}$. The value of $\gamma_{i,j}$ does not matter to our argument; however it can be given explicitly.

Claim IV.2. *There exists a constant $C = C(d)$, such that, with high probability,*

$$\mathcal{L} \preceq C \sum_{i=1}^n P^\perp_{u_{\mathcal{C}_i}}.$$

We omit the proof of this claim due to space constraint. The argument is closely related to the Markov chain comparison technique [4].

Using Claim IV.2, $v^T \mathcal{L} v \le C(\beta_i^2 \|\tilde{x}_{\mathcal{C}_i}\|^2 + \|w^{(i)}\|^2)$. Hence, we only need to show

$$\sum_{i=1}^n \|w^{(i)}\|^2 \ge Cr^2 \sum_{i=1}^n \beta_i^2 \|\tilde{x}_{\mathcal{C}_i}\|^2, \tag{9}$$

Since the degree of each node is bound by Cnr^d (with high probability), we have

$$\begin{aligned}\sum_{i=1}^n \|w^{(i)}\|^2 &\ge (Cnr^d)^{-1} \sum_{(i,j)\in E} (\|w^{(i)}\|^2 + \|w^{(j)}\|^2) \\ &\ge (Cnr^d)^{-1} \sum_{(i,j)\in E} (\|w^{(i)}_{\mathcal{C}_i \cap \mathcal{C}_j} - w^{(j)}_{\mathcal{C}_i \cap \mathcal{C}_j}\|^2) \\ &\overset{(8)}{\ge} (Cnr^d)^{-1} \sum_{(i,j)\in E} (\beta_j - \beta_i)^2 \|\tilde{x}_{\mathcal{C}_i \cap \mathcal{C}_j}\|^2\end{aligned}$$

Applying Chernoff bounds, there exists constants C_1 and C_2, such that, with high probability, $\|\tilde{x}_{\mathcal{C}_i \cap \mathcal{C}_j}\|^2 \ge C_1 (nr^d) r^2$ and $\|\tilde{x}_{\mathcal{C}_i}\|^2 \le C_2 (nr^d) r^2$ for all i and j. Thus, in order to prove (9), we need to show $\sum_{(i,j)\in E}(\beta_j - \beta_i)^2 \ge C(nr^d)r^2 \sum_{i=1}^n \beta_i^2$.

Define $\underline{\beta} = (\beta_1, \cdots, \beta_n)$. Observe that $\sum_{(i,j)\in E}(\beta_j - \beta_i)^2 = \underline{\beta}^T \mathcal{L} \underline{\beta} \ge \sigma_{\min}(\mathcal{L}) \|P_u^\perp \underline{\beta}\|^2$, where $\sigma_{\min}(\mathcal{L})$ denotes the minimum nonzero eigenvalue of $\mathcal{L}$. Since $v \perp x$, it can be shown that $\|P_u^\perp \underline{\beta}\|^2 \ge C\|\underline{\beta}\|^2$ (we omit the details). The proof is completed by using Remark IV.1. ■

We are finally in position to prove Lemma III.1.

Proof (Lemma III.1): Note that $R^\perp = Q^\perp = P_V^\perp Q P_V^\perp \succeq \mathbf{0}$. Write $R^\perp = \sum_{i=1}^{n-d-1} \lambda_i u_i u_i^T$, where $\|u_i\| = 1$, $u_i^T u_j = 0$ for $i \neq j$ and $\lambda_1 \ge \lambda_2 \ge \cdots \lambda_{n-d-1} \ge 0$. Therefore, $\langle \Omega, R^\perp \rangle = \langle \Omega, \sum_{i=1}^{n-d-1} \lambda_i \quad u_i u_i^T \rangle = \sum_{i=1}^{n-d-1} \lambda_i u_i^T \Omega u_i \ge \sigma_{\min}(\Omega)\|R^\perp\|_*$, where $\sigma_{\min}(\Omega)$ denotes the minimum nonzero eigenvalue of Ω. Here, we used the fact that $u_i \in V^\perp = \text{Ker}^\perp(\Omega)$.

As a direct consequence of proposition IV.4 and Remark IV.1, $\sigma_{\min}(\Omega) \ge C(nr^d)r^4$. The result follows. ■

B. *Proof of Lemma III.2*

Recall that $\tilde{R} = P_V R P_V + P_V R P_{V^\perp} + P_{V^\perp} R P_V$. Therefore, there exist a matrix $Y \in \mathbb{R}^{n\times d}$ and a vector $a \in \mathbb{R}^n$ such that $\tilde{R} = XY^T + YX^T + ua^T + au^T$. Denote by $y_i^T \in \mathbb{R}^d$, $i \in [n]$, the i^{th} row of the matrix Y.

The following proposition plays a key role in the proof. Its proof is deferred to the next subsection.

Proposition IV.5. *There exists a constant $C = C(d)$, such that, with high probability,*

$$\sum_{i,j} |\langle x_i - x_j, y_i - y_j \rangle| \le Cr^{-d-2} \sum_{(l,k)\in E} |\langle x_l - x_k, y_l - y_k \rangle|$$

The next statement provides an upper bound on $\|\tilde{R}\|_F$. We defer its proof to [7].

Proposition IV.6. *There exists a constant $C = C(d)$, such that, with high probability,*

$$\|\tilde{R}\|_F \le C\,\frac{1}{n}\sum_{i,j}|\langle x_i - x_j, y_i - y_j\rangle|.$$

Now we have in place all we need to prove lemma III.2.

Proof (Lemma III.2): Define the operator $\mathcal{A}_G : \mathbb{R}^{n\times n} \to \mathbb{R}^{|E|}$ as $\mathcal{A}_G(S) = [\langle M_{ij}, S\rangle]_{(i,j)\in E}$. By our assumptions,

$$\begin{aligned}&|\langle M_{ij}, \tilde{R}\rangle + \langle M_{ij}, R^\perp\rangle| = |\langle M_{ij}, Q\rangle - \langle M_{ij}, Q_0\rangle|\\ &\le |\langle M_{ij}, Q\rangle - \tilde{d}_{ij}^2| + \underbrace{|\tilde{d}_{ij}^2 - \langle M_{ij}, Q_0\rangle|}_{|z_{ij}|} \le 2\Delta.\end{aligned}$$

Therefore, $\|\mathcal{A}_G(\tilde{R})\|_1 \le 2|E|\Delta + \|\mathcal{A}_G(R^\perp)\|_1$. Write the laplacian matrix L as $L = \sum_{(i,j)\in E} M_{ij}$. Then, $\langle L, R^\perp\rangle = \sum_{(i,j)\in E}\langle M_{ij}, R^\perp\rangle = \|\mathcal{A}_G(R^\perp)\|_1$. Here, we used the fact that $\langle M_{ij}, R^\perp\rangle \ge 0$, since $M_{ij} \succeq \mathbf{0}$ and $R^\perp \succeq \mathbf{0}$. Hence, $\|\mathcal{A}_G(\tilde{R})\|_1 \le 2|E|\Delta + \langle L, R^\perp\rangle$.

Applying propositions IV.4 and IV.3, $\langle L, R^\perp\rangle \le Cr^{-2}\langle \Omega, R^\perp\rangle \le Cnr^{-2}(nr^d)^2\Delta$, whence we obtain $\|\mathcal{A}_G(\tilde{R})\|_1 \le Cnr^{-2}(nr^d)^2\Delta$.

The last step is to write $\|\mathcal{A}_G(\tilde{R})\|_1$ more explicitly. Notice that, $\|\mathcal{A}_G(\tilde{R})\|_1 = \sum_{(l,k)\in E}|\langle M_{lk}, XY^T + YX^T + ua^T + au^T\rangle| = 2\sum_{(l,k)\in E}|\langle x_l - x_k, y_l - y_k\rangle|$.

The result follows as a direct consequence of propositions IV.5 and IV.6. ∎

C. Proof of Proposition IV.5

We will focus here on the case $d = 2$. The general argument proceeds along the same lines and is deferred to [7].

We first need to establish the following definition.

Definition 1. A chain G_{ij} is a sequence of subgraphs $H_1, H_2, \cdots, H_k$ along with the vertices i and j, such that, each H_p is isomorphic to K_4 and each two successive H_p share one side. Further, i (resp. j) is connected to the two vertices in $V(H_1)\setminus V(H_2)$ (resp. $V(H_k)\setminus V(H_{k-1})$).

See Fig. 1 in the Appendix A for an illustration of a chain.

Proposition IV.7. *For any two nodes i and j in our random geometric graph G, there exists a chain $G_{ij} \subseteq G$.*

Proposition IV.8. *For any two nodes i and j, there exists a constant $C = C(d)$, such that,*

$$|\langle x_i - x_j, y_i - y_j\rangle| \le Cr^{-1}\sum_{(l,k)\in E(G_{ij})}|\langle x_l - x_k, y_l - y_k\rangle|$$

Proof: Assume that $|V(G_{ij})| = m + 1$. Relabel the vertices in the chain such that the nodes i and j have labels 0 and m respectively. Since both sides of the desired inequality are invariant to translations, without loss of generality we assume that $x_0 = y_0 = 0$. For a fixed vector y_m consider the following optimization problem.

$$\Theta = \min_{y_1,\cdots,y_{m-1}\in\mathbb{R}^d}\sum_{(l,k)\in E(G_{ij})}|\langle x_l - x_k, y_l - y_k\rangle|\,.$$

To each edge $(l,k)\in E(G_{ij})$, assign a number λ_{lk}. For any assignment with $\max|\lambda_{lk}| \le 1$, we have

$$\begin{aligned}\Theta &\ge \min_{y_1,\cdots,y_{m-1}\in\mathbb{R}^d}\sum_{(l,k)\in E(G^*)}\lambda_{lk}\langle x_l - x_k, y_l - y_k\rangle\\ &= \min_{y_1,\cdots,y_{m-1}\in\mathbb{R}^d}\sum_{\substack{l\in G^*\\ l\neq 0}}\sum_{k\in\partial l}\lambda_{lk}\langle y_l, x_l - x_k\rangle\\ &= \min_{y_1,\cdots,y_{m-1}\in\mathbb{R}^d}\sum_{\substack{l\in G^*\\ l\neq 0}}\langle y_l, \sum_{k\in\partial l}\lambda_{lk}(x_l - x_k)\rangle,\end{aligned}$$

where ∂l denotes the set of adjacent vertices to l in G_{ij}. The numbers λ_{lk} that maximize the right hand side should satisfy $\sum_{k\in\partial l}\lambda_{lk}(x_l - x_k) = 0, \forall l \neq 0, m$. Thus, $\Theta \ge \langle y_m, \sum_{k\in\partial m}\lambda_{mk}(x_m - x_k)\rangle$. The result follows as a direct consequence of the following Claim whose proof is deferred to Appendix A.

Claim IV.3. *There exist numbers λ_{lk} that satisfy the following three conditions*

$$\begin{gathered}\sum_{k\in\partial l}\lambda_{lk}(x_l - x_k) = 0 \quad \forall l \neq 0, m,\\ \sum_{k\in\partial m}\lambda_{mk}(x_m - x_k) = x_m, \qquad (10)\\ \max|\lambda_{lk}| \le Cr^{-1}.\end{gathered}$$

The proof is completed by the following proposition, whose proof we omit due to space constraints.

Proposition IV.9. *Define the "congestion number" of the graph G as $b(G) = \max_{e\in E(G)}\sharp\{G_{i,j}\subseteq G : e \in E(G_{ij})\}$. Then, $b(G) \le Cr^{-d-1}$, for some constant $C = C(d)$.* ∎

Acknowledgment. A. Javanmard is supported by Caroline and Fabian Pease Stanford Graduate Fellowship. This work was partially supported by the NSF CAREER award CCF-0743978, the NSF grant DMS-0806211, and the AFOSR grant FA9550-10-1-0360.

References

[1] L. Asimow and B. Roth. The rigidity of graphs. *Trans. Amer. Math. Soc.*, 245:279–289, 1978.
[2] P. Biswas and Y. Ye. Semidefinite programming for ad hoc wireless sensor network localization. In *IPSN '04*.
[3] T. Cox and M. Cox. *Multidimensional Scaling. Monographs on Statistics and Applied Probability 88.* Chapman and Hall, 2001.
[4] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible markov chains. *Annals of Applied Probability*, 3(3):696–730, 1993.
[5] L. Fan, S. Keles, S. J. Wright, and G. Wahba. Framework for kernel regularization with application to protein clustering. In *In Proc. National Academic Science*, volume 102, pages 12332–12337, USA, 2005.
[6] S. J. Gortler, A. Healy, and D. Thurston. Characterizing generic global rigidity. *American Journal of Mathematics*, 132(4):897–939, 2010.
[7] A. Javanmard and A. Montanari. Localization from incomplete noisy distance measurements. *journal version, in preparation.*
[8] S. Oh, A. Karbasi, and A. Montanari. Sensor network localization from local connectivity: Performance analysis for the mds-map algorithm. In *ITW*, 2010.
[9] M. Penrose. *Random Geometric Graphs.* OUP Inc., 2003.
[10] L. Saul and S. Roweis. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *JMLR*, 4:119–155, 2003.
[11] A. Singer. A remark on global positioning from local distances. In *PNAS*, volume 105, pages 9507–9511, 2008.
[12] J. Tenenbaum, V. Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

APPENDIX

A. *Proof of Claim IV.3*

Proof: Notice that for any values λ_{lk} satisfying (10), we have $\sum_{k\in\partial 0}\lambda_{0k}(x_0 - x_k) = -x_m$. As a generalization, consider the following linear system of equations with unknown variables λ_{lk}.

$$\sum_{k\in\partial l}\lambda_{lk}(x_l - x_k) = u_l, \qquad \text{for } l = 0, \cdots, m \tag{11}$$

Writing Eqs. (11) in terms of the rigidity matrix of G_{ij}, and using the characterization of its null space, as discussed in section II-A, it follows that Eqs. (11) have a solution if and only if

$$\sum_{i=0}^{m} u_i = 0, \quad \sum_{i=0}^{m} u_i^T A x_i = 0, \tag{12}$$

where $A \in \mathbb{R}^{d\times d}$ is an arbitrary skew symmetric matrix.

A mechanical interpretation. If we think of each u_i as a force imposed on the node i, then the first constraint in Eq. (12) states that the net force on G_{ij} is zero (*force equilibrium*), while the second condition states that the net torque is zero (*torque equilibrium*).

With this interpretation in mind, we propose a two-stage procedure to find the values λ_{lk} that obey the constraints in (10).

Stage (i): Let $\mathcal{F}_p$ denote the common side of H_p and H_{p+1}. Without loss of generality, assume $V(\mathcal{F}_p) = \{1, 2\}$. Find the forces f_1, f_2 such that

$$\begin{gathered} f_1 + f_2 = x_m, \quad f_1 \wedge x_1 + f_2 \wedge x_2 = 0, \\ \|f_1\|^2 + \|f_2\|^2 \le C\|x_m\|^2. \end{gathered} \tag{13}$$

To this end, we solve the following optimization problem.

$$\begin{aligned} &\text{minimize} \quad 1/2(\|f_1\|^2 + \|f_2\|^2) \\ &s.t. \qquad\quad f_1 + f_2 = x_m, \quad f_1 \wedge x_1 + f_2 \wedge x_2 = 0 \end{aligned} \tag{14}$$

It is easy to see that the solutions of (14), given by

$$\begin{cases} f_1 = \frac{1}{2}x_m + \frac{1}{2}\gamma A(x_1 - x_2) \\ f_2 = \frac{1}{2}x_m - \frac{1}{2}\gamma A(x_1 - x_2) \end{cases}$$

$$\gamma = -\frac{1}{\|x_1 - x_2\|^2} x_m^T A(x_1 + x_2), \quad A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

satisfy the constraints in (13).

Stage (ii): For each H_p consider the following set of forces

$$\tilde{f}_i = \begin{cases} f_i & \text{if } i \in V(\mathcal{F}_p) \\ -f_i & \text{if } i \in V(\mathcal{F}_{p-1}) \end{cases}, \ \tilde{f}_m = x_m, \ \tilde{f}_0 = -x_m \tag{15}$$

See Fig. 2 for an illustration. Notice that $\sum_{i\in V(H_p)}\tilde{f}_i = 0$, $\sum_{i\in V(H_p)}\tilde{f}_i \wedge x_i = 0$, and thus by our previous discussion, there exist values $\lambda_{lk}^{(H_p)}$, such that, $\sum_{k:(l,k)\in E(H_p)}\lambda_{lk}^{(H_p)}(x_l - x_k) = \tilde{f}_l, \ \forall l \in V(H_p)$. Writing this in terms of $R^{(H_p)}$, the rigidity matrix of H_p, we have $R^{(H_p)}\underline{\lambda}^{(H_p)} = \underline{\tilde{f}}$. Therefore, $\sigma_{\min}(R^{(H_p)})\|\lambda^{(H_p)}\|_\infty \le \sigma_{\min}(R^{H_p})\|\lambda^{(H_p)}\|_2 \le \|\tilde{f}\| \le C\|x_m\|$. It can be shown that $\sigma_{\min}(R^{H_p}) \ge Cr$ (we omit the proof). Also, $\|x_m\| = O(1)$. Hence, $\|\lambda^{(H_p)}\|_\infty \le Cr^{-1}$.

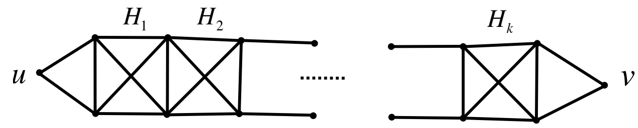

Fig. 1. An illustration of a chain G_{uv}

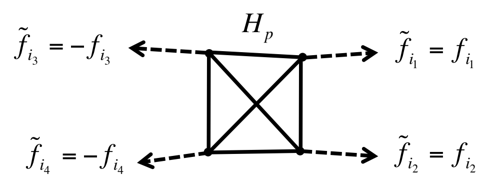

Fig. 2. H_p and the set of forces in Stage (ii)

Now define $\lambda_{lk} = \sum_{H_p:(l,k)\in E(H_p)}\lambda_{lk}^{(H_p)}$ for every $(l,k) \in E(G_{ij})$. We claim that the values λ_{lk} satisfy the constraints in (10). First, note that in the summation $\sum_{k\in\partial l}\lambda_{lk}(x_l - x_k)$, all the internal forces cancel each other and the sum is zero at the internal nodes l. At the extreme nodes 0 and m, this sum would be equal to $-x_m$ and x_m respectively. In addition, since each edge participates in at most two H_p, we have $|\lambda_{lk}| \le Cr^{-1}$. ■

B. *Proof Sketch of the Converse to Theorem I.1*

Proof: Consider the '*bending*' map $T : [-0.5, 0.5]^d \to \mathbb{R}^{d+1}$, defined as

$$T(t_1, t_2, \cdots, t_d) = (R\sin(t_1/R), R(1-\cos(t_1/R)), t_2, \cdots, t_d)$$

This map bends the hypercube in the $d+1$ dimensional space. Here, R is the curvature radius of the embedding (for instance, $R \gg 1$ corresponds to slightly bending the hypercube).

Now for a given Δ, let $R = r^2\Delta^{-1/2}$ and give the distances $\tilde{d}_{ij} = \|T(x_i) - T(x_j)\|$ as the proximity measurements to the algorithm. First we show that these adversarial measurements satisfy the noise constraint, $\|\tilde{d}_{ij}^2 - d_{ij}^2\| \le \Delta$.

$$\begin{aligned} d_{ij}^2 - \tilde{d}_{ij}^2 &= (x_i^{(1)} - x_j^{(1)})^2 - R^2\left(\sin(\frac{x_i^{(1)}}{R}) - \sin(\frac{x_j^{(1)}}{R})\right)^2 \\ &\quad - R^2\left(\cos(\frac{x_i^{(1)}}{R}) - \cos(\frac{x_j^{(1)}}{R})\right)^2 \\ &= (x_i^{(1)} - x_j^{(1)})^2 - R^2\left(2 - 2\cos(\frac{x_i^{(1)} - x_j^{(1)}}{R})\right) \\ &\le \frac{(x_i^{(1)} - x_j^{(1)})^4}{2R^2} \le \frac{r^4}{2R^2} \le \Delta. \end{aligned}$$

Also, $\tilde{d}_{ij} \le d_{ij}$. Therefore, $|z_{ij}| = |\tilde{d}_{ij}^2 - d_{ij}^2| \le \Delta$.

The crucial point is that given the measurements $\tilde{d}_{ij}$ as the input to the algorithm, the SDP in the first step will return the positions $\tilde{x}_i = LT(x_i)$, since it is oblivious of dimension d.

Let Q be Gram matrix of the positions $\{\tilde{x}_i\}_{i\in[n]}$, namely $Q_{ij} = \tilde{x}_i \cdot \tilde{x}_j$. Denote by $\{u_1, \cdots, u_d\}$, the eigenvectors of Q corresponding to the d largest eigenvalues. In the second step, the positions $\{\tilde{x}_i\}_{i\in[n]}$ are projected onto the space $U = \langle u_1, \cdots, u_d\rangle$ and the results are returned as the estimated positions in $\mathbb{R}^d$, i. e., $\hat{x}_i = P_U \tilde{x}_i$. It can be shown that the space U is almost aligned with the space $W = \langle e_1, e_3, \cdots, e_{d+1}\rangle$, where e_i refers to the i^{th} standard basis element, e.g., $e_1 = (1, 0, \cdots, 0)$. More specifically, there exists a constant C, such that, with high probability,

$$\|P_U \tilde{x}_i - x_i\| \geq C \|P_W \tilde{x}_i - x_i\| \qquad \forall i \in [n].$$

(we defer the details to [7]). Therefore,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \|\hat{x}_i - x_i\| &= \frac{1}{n}\sum_{i=1}^{n} \|P_U \tilde{x}_i - x_i\| \\
&\geq C\frac{1}{n}\sum_{i=1}^{n} \|P_W \tilde{x}_i - x_i\| \\
&\geq C\frac{1}{n}\sum_{i=1}^{n} |R\sin(\tilde{x}_i^{(1)}/R) - x_i^{(1)}| \\
&\geq \frac{C}{R^2} = C\frac{\Delta}{r^4}
\end{aligned}$$

Since $d(X, \hat{X})$ is a measure of the average reconstruction error per point (we omit the details), we obtain

$$d(X, \hat{X}) \geq \frac{C}{n}\sum_{i=1}^{n} \|\hat{x}_i - x_i\| \geq C\frac{\Delta}{r^4}$$

which proves the thesis. ■